

## MODELE OPTYMALNEGO PRZETWARZANIA SYGNAŁÓW LOSOWYCH

*Estymacja sygnałów losowych jest istotnym zagadnieniem matematycznym, mającym zastosowanie w wielu dziedzinach nauki i techniki. Celem niniejszego artykułu jest porównanie procedur estymacji sygnałów losowych z wykorzystaniem liniowych estymatorów MMSE (Minimum Mean-Squared Error – minimum błędu średniokwadratowego).*

**Słowa kluczowe:** optymalne przetwarzanie sygnałów, estymacja sygnałów losowych.

### WSTĘP

Przez optymalne przetwarzanie sygnałów losowych należy rozumieć wydobywanie informacji z danych dyskretnych (danych czasu dyskretnego), jak np. z danych pomiarowych skalarnych lub wektorowych wielkości fizycznych, w sposób „optymalny”, przy czym kryteria optymalności mogą być różne. Zakłada się zatem, że modelem matematycznym takich sygnałów są dyskretne procesy stochastyczne. Warto tutaj przypomnieć, że tego typu procesy stochastyczne są zdefiniowane przez rodziny  $\{x(n)\}$  zmiennych losowych, generowane przez dyskretne momenty czasu lub jako zbiory deterministycznych ciągów liczbowych, stanowiących tzw. realizacje procesu stochastycznego. Na ogół w praktyce nie dysponuje się pełnym opisem procesu stochastycznego, a więc znajomością funkcji gęstości prawdopodobieństwa każdej próbki-zmiennej losowej,  $x(n)$ . Co więcej, znajomość procesu ograniczona jest jedynie do  $M < \infty$  próbek lub  $M$  punktów tylko jednej realizacji. Jeżeli przykładowo informację stanowi deterministyczny parametr  $\Theta$ , to znajomość  $M$  próbek procesu pozwala na wyznaczenie zmiennej losowej  $\hat{\Theta} = T(x(n))$ , będącej estymatorem parametru  $\Theta$ . Każda szczególna wartość  $\hat{\Theta}$  jest estymatą parametru  $\Theta$ .  $T(\cdot)$  jest symbolem pewnego odwzorowania funkcjonalnego. Obok estymacji momentów i widm procesów stochastycznych procedury estymacji są powszechnie wykorzystywane do estymacji parametrów modeli sygnałów losowych w przetwarzaniu, obejmującym wygładzanie, filtrację i predykcję. Celem niniejszego artykułu jest przegląd i porównanie procedur estymacji sygnałów losowych z wykorzystaniem liniowych estymatorów MMSE (*Minimum Mean-Squared Error* – minimum błędu średniokwadratowego).

## 1. PROBLEM ESTYMACJI SYGNAŁU LOSOWEGO

Problem estymacji sygnału losowego można sprowadzić do następującego zagadnienia [3]:

Zadany jest ciąg  $x(n)$ ;  $n = 0, 1, \dots, N-1$  będący zbiorem próbek z obserwowanej (pomiarowej) realizacji procesu stochastycznego. Ciąg  $d(n)$ ;  $n = 0, 1, \dots, N-1$  jest informacją zawartą w ciągu  $x(n)$ , tzn. istnieje funkcjonalne odwzorowanie  $T(\cdot)$  takie, że:

$$\hat{d}(n) = T(x(n)) \quad (1)$$

gdzie  $\hat{d}(n)$  jest estymatorem sygnału  $d(n)$ .

Liniowy estymator MMSE określony jest przez zależność:

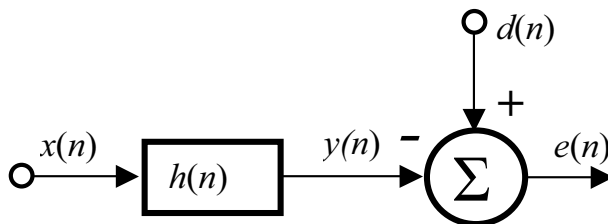
$$\hat{d}(n) = \sum_{i=0}^{N-1} h_i x(i), \quad n = 0, 1, \dots, N-1 \quad (2)$$

gdzie  $h_i$  – parametry estymatora.

Można pokazać, że estymator MMSE jest realizowalny z użyciem filtrów liniowych (rozwiązanie klasyczne – optymalna filtracja LS), a także sztucznych sieci neuronowych (LS – *Least Squares*).

## 2. OPTYMALNE FILTRY NAJMNIEJSZYCH KWADRATÓW (LS – LEAST SQUARES)

Optymalna filtracja z zastosowaniem liniowych filtrów LS jest jednym z podstawowych narzędzi przetwarzania sygnałów losowych. Podstawowy schemat takiej filtracji przedstawiono na rysunku 1 [3].



**Rys. 1.** Blokowy schemat filtracji optymalnej  
**Fig. 1.** The block diagram of the optimal filtering

Parametry filtra o odpowiedzi impulsowej  $h(n)$  należy tak dobrać, aby odpowiedź  $y(n)$  była „bliska” zadanemu ciągowi  $d(n)$ ,  $y(n) \approx d(n)$ . Stąd, definiując funkcję kosztów (celu) jako:

$$J = E \{ e^2(n) \} \quad (3)$$

w przypadku, gdy  $x(n)$  jest ciągiem losowym lub

$$J = \sum_n e^2 \{ n \} \quad (4)$$

dla sygnałów deterministycznych, realizację zależności  $y(n) \approx d(n)$  uzyskuje się, wyznaczając

$$\min J \{ e(n) \} \quad (5)$$

Ponieważ:

$$e(n) = d(n) - \sum_i h(i)x(n-i) \quad (6)$$

gdzie wskaźnik sumowania  $i$  przybiera wartości zależne od długości odpowiedzi impulsowej filtra, to warunki na minimum funkcji celu:

$$\min J \{ e^2(n) \} = \min J \left\{ \left( d(n) - \sum_i h(i)x(n-i) \right)^2 \right\} \quad (7)$$

ze względu na wartości odpowiedzi impulsowej  $h(i)$  prowadzą do układu równań liniowych znanych jako **równania normalne**, o postaci:

$$\sum_i R(n-i, n-j)h(i) = g(n-j, n) \quad (8)$$

gdzie:

$R(n-i, n-j) = E \{ x(n-i)x(n-j) \}$  – funkcja autokorelacji ciągu wejściowego,

$g(n-i, n-j) = E \{ x(n-i)d(n) \}$  – funkcja korelacji skrośnej  $x(n)$  oraz  $d(n)$ .

Zakres wartości wskaźnika  $j$  zależy od długości odpowiedzi impulsowej. Rozwiązanie równań normalnych ustala parametry  $h(i)$  filtra optymalnego.

Należy zwrócić uwagę na następujące cechy filtracji LS:

- funkcja  $J \{ e(n) \}$  posiada globalne minimum;
- filtracja optymalna LS zapewnia ortogonalność ciągów  $e(n)$  i  $y(n)$ :

$$E \{ e(n), y(n) \} = 0; \quad (9)$$

- filtracja optymalna LS dostarcza estymatora zadanego ciągu  $d(n)$ :

$$\hat{d}(n) = y(n) = T \{ x(n) \} \quad (10)$$

Ze względów praktycznych zakłada się stacjonarność w szerokim sensie ciągów wejściowych  $x(n)$ , co prowadzi do zmiany formy równań normalnych (8):

$$\sum_i R(j-i)h(i) = g(j) \quad (11)$$

Drugim istotnym ze względów praktycznych założeniem jest przyjęcie skończonej długości odpowiedzi impulsowej  $h(n) = \{h(0), h(1), \dots, h(L-1)\}$ .

Odpowiedź  $y(n)$  przyczynowego filtra LS można przedstawić w postaci:

$$y(n) = \mathbf{h}^T(n) \cdot \mathbf{x}(n) \quad (12)$$

gdzie:

$\mathbf{h}^T(n) = [h(0), h(1), \dots, h(L-1)]$  – odpowiedź impulsowa,

$\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-(L-1))]^T$  – wektor regresji.

Stąd równania normalne (11) otrzymują postać:

$$\mathbf{R} \cdot \mathbf{h} = \mathbf{g} \quad (13)$$

gdzie  $\mathbf{R}$  – macierz autokorelacji:

$$\mathbf{R} = \begin{bmatrix} R(0) & R(1) & R(2) & \dots & R(L-1) \\ R(1) & R(0) & R(2) & \cdot & \cdot \\ \vdots & \cdot & \cdot & \cdot & \cdot \\ R(L-1) & \cdot & \cdot & \cdot & R(0) \end{bmatrix} \quad (14)$$

jest macierzą Toeplitza.

$$\mathbf{g} = [g(0), g(1), \dots, g(L-1)]^T \text{ – wektor korelacji skrośnej} \quad (15)$$

Optymalne rozwiązanie równania (13) istnieje dla nieosobliwej macierzy autokorelacji, tzn.:

$$\mathbf{h} = \mathbf{R}^{-1} \mathbf{g} \quad (16)$$

Należy jednak zauważyć, że macierz Toeplitza jest w ogólności dodatnio pół-określona (*positive semi-definite*), co nie zapewnia istnienia  $\mathbf{R}^{-1}$ .

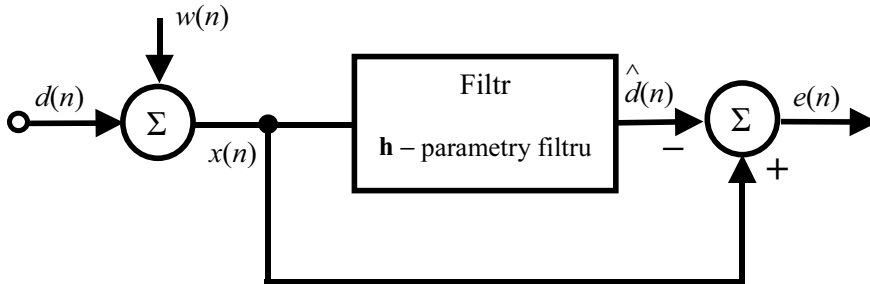
W przypadku ciągów  $x(n)$  deterministycznych lub ergodycznych dane są jedynie ciągi liczbowe  $x(n)$  oraz  $d(n)$  dla  $n = 0, 1, \dots, M-1$ . Operację uśredniania po zbiorze  $E\{\cdot\}$  należy zastąpić uśrednianiem po czasie. Stąd:

$$R(i) = \frac{1}{M} \sum_{n=0}^{M-i-1} x(n)x(n+i); \quad i = 0, 1, \dots, L-1 \quad (17)$$

$$g(i) = \frac{1}{M} \sum_{n=0}^{M-i-1} x(n)d(n+i); \quad i = 0, 1, \dots, L-1 \quad (18)$$

$$\mathbf{R} = \{R(i)\} \quad (19)$$

Warto zwrócić uwagę, że liczba operacji dla rozwiązania równania (16) może być numerycznie znacznie zredukowana przez zastosowanie algorytmu Levinsona. Pewną modyfikację struktury optymalnej filtracji z rysunku 1 można wprowadzić w przypadku, gdy sygnał wejściowy  $x(n)$  zawiera addytywny szum, w szczególności biały,  $w(n)$ . Schemat takiej struktury przedstawiono na rysunku 2.



**Rys. 2.** Blokowy schemat filtracji optymalnej – sygnał zawiera addytywny szum biały  
**Fig. 2.** The block diagram of the optimal filtering – signal contains additive white noise

Funkcja celu struktury dana jest poprzez błąd średniokwadratowy (MSE), jako:

$$J\{e(n)\} = E\{e^2(n)\} = E\left\{\left(d(n) - \hat{d}(n)\right)^2\right\} + E\{w^2(n)\} + 2E\{d(n)w(n)\} - 2E\{\hat{d}(n)w(n)\} \quad (20)$$

Zakładając, że ciągi  $d(n)$ ,  $\hat{d}(n)$  i  $w(n)$ , a więc sygnał zadany i szum są statystycznie niezależne, tzn.

$$E\{d(n)w(n)\} = 0, \quad E\{\hat{d}(n)w(n)\} = 0 \quad (21)$$

wartość funkcji celu może być zmniejszona przez odpowiedni dobór (adaptację) parametrów filtru. Minimum tej funkcji osiąga się, spełniając warunki:

$$\begin{aligned} \hat{d}(n) &= d(n) \\ e(n) &= w(n) \end{aligned} \quad (22)$$

Stąd  $\hat{d}(n)$  jest estymatorem MSE sygnału  $d(n)$ .

Przy spełnieniu warunków (21) funkcja celu (20) ma postać:

$$J\{e(n)\} = E\{e^2(n)\} = E\left\{\left(d(n) - \hat{d}(n)\right)^2\right\} + E\{w^2(n)\} \quad (23)$$

Ponieważ związek sygnału wyjściowego  $\hat{d}(n)$  oraz wejściowego  $x(n)$  filtru dyskretnego można zapisać w postaci równania różnicowego:

$$\begin{aligned} \hat{d}(n) + c_1 \hat{d}(n-1) + c_2 \hat{d}(n-2) + \dots + c_N \hat{d}(n-N) = \\ = b_0 x(n) + b_1 x(n-1) + \dots + b_M x(n-M), N \geq M \end{aligned} \quad (24)$$

lub

$$\hat{d}(n) = \mathbf{x}^T(n) \mathbf{h} \quad (25)$$

gdzie  $\mathbf{x}^T(n) = [-\hat{d}(n-1) - \hat{d}(n-2), \dots, -\hat{d}(n-1), x(n), x(n-1), \dots, x(n-M)]$  jest

wektorem regresji (regresor),

natomiast  $\mathbf{h} = [c_1, c_2, \dots, c_N, b_0, b_1, \dots, b_M]^T$  – wektor parametrów,

to funkcja celu (23) jest funkcją wektora  $\mathbf{h}$  (wielu zmiennych):

$$J\{e(n)\} = J(\mathbf{h}) = E\{d^2(n)\} + \mathbf{h}^T E\{\mathbf{x}(n)\mathbf{x}^T(n)\} \mathbf{h} - 2E\{d(n)\mathbf{x}^T(n)\} \mathbf{h} + E\{w^2(n)\} \quad (26)$$

Spełnienie warunku koniecznego na minimum funkcji  $J(\mathbf{h})$ , tzn.

$$\nabla J\{\mathbf{h}\} = \mathbf{0}, \quad (\nabla - \text{gradient}) \quad (27)$$

prowadzi do zbioru równań liniowych:

$$\mathbf{R}_{\mathbf{xx}} \cdot \mathbf{h} = \mathbf{R}_{\mathbf{xd}} \quad (28)$$

znanych jako równania Wienera-Hopfa.

Macierz  $\mathbf{R}_{\mathbf{xx}} = E\{\mathbf{x}(n) \cdot \mathbf{x}^T(n)\}$  jest macierzą autokorelacji wektora regresji, natomiast  $\mathbf{R}_{\mathbf{xd}} = E\{\mathbf{x}(n) \cdot d(n)\}$  jest wektorem korelacji skrośnej wektora regresji oraz zadanego sygnału. Rozwiązanie równania (28) istnieje dla macierzy nieosobliwej  $\mathbf{R}_{\mathbf{xx}}$ :

$$\mathbf{h} = \mathbf{R}_{\mathbf{xx}}^{-1} \mathbf{R}_{\mathbf{xd}} \quad (29)$$

Co więcej  $\mathbf{h}$  jest punktem minimum  $\mathbf{h}^*$  funkcji  $J(\mathbf{h})$ , gdy  $\mathbf{R}_{\mathbf{xx}}$  jest dodatnio określona ( $\mathbf{R}_{\mathbf{xx}}$  – hessian funkcji  $J(\mathbf{h})$ ). Jeśli ponadto  $\mathbf{R}_{\mathbf{xx}}$  nie zależy od  $\mathbf{h}$ , tzn.  $\mathbf{R}_{\mathbf{xx}}$  jest macierzą liczbową, rozwiązanie (29) jest punktem minimum globalnego  $\mathbf{h}^{**}$  funkcji  $J(\mathbf{h})$ . Minimum globalne  $\mathbf{h}^{**}$  można otrzymać dla kwadratowej funkcji  $J(\mathbf{h})$ , a więc dla wektora parametrów spełniających  $c_i = 0$ ,  $i = 1, \dots, N$ . W przypadku ergodycznych wektorów regresji operator uśredniania po zbiorze  $E\{\cdot\}$  zastępuje się uśrednianiem po czasie. Stąd równanie Wienera-Hopfa upraszcza się do postaci:

$$\mathbf{h} = \mathbf{R}_{\mathbf{xx}}^{-1} \mathbf{R}_{\mathbf{xd}} = \lim_{k \rightarrow \infty} \left\{ \sum_{i=1}^k \{\mathbf{x}(i)\mathbf{x}^T(i)\} \right\}^{-1} \sum_{i=1}^k \{\mathbf{x}(i)d(i)\} \quad (30)$$

### 3. MODELE LINIOWYCH OBSERWATORÓW

Problem liniowej obserwacji dany jest przez zbiór równań algebraicznych o postaci [1]:

$$\mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{w} \quad (31)$$

gdzie:

$\mathbf{x}$  – nieznanany wektor,  $\mathbf{x} \in \mathbb{R}^n$ ,

$\mathbf{y}$  – zawiera znane dane pomiarowe (obserwacje),  $\mathbf{y} \in \mathbb{R}^r$ ,

$\mathbf{D} \in \mathbb{R}^{r \times n}$  – dana macierz liczbowa,

$\mathbf{w} \in \mathbb{R}^r$  – wektor zawierający nieznanne błędy pomiarowe lub składowe szumu o różnych właściwościach widmowych (np. szum biały).

Wiadomo, że rozwiązanie równania (31) można otrzymać metodą najmniejszych kwadratów (LS), wyznaczając minimum funkcji kwadratowej:

$$F(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 \quad (32)$$

więc:

$$\hat{\mathbf{x}}_{LS} = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \left\{ F(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 \right\}, \quad (33)$$

$\hat{\mathbf{x}}_{LS}$  – estymator LS rozwiązania równania obserwacji (31).

Warto zwrócić uwagę, że jeden z modeli generujących wielkie zbiory danych (*Big Data*) [2] ma formę równania (31). W przypadku  $r > n$ , pomijając właściwości szumu, analityczne rozwiązanie równania (33) dane jest przez znany estymator Gaussa:

$$\hat{\mathbf{x}} = (\mathbf{D}^T \mathbf{D})^{-1} \cdot \mathbf{D}^T \mathbf{y} \quad (34)$$

Macierz  $\mathbf{D}^T \mathbf{D}$  znana jest jako macierz normalna Gaussa.

Estymator  $\hat{\mathbf{x}}$  uwzględniający właściwości szumu, można uzyskać za pomocą algorytmu MMSE, przy czym  $r \geq n$  w równaniu (31). Traktując wektory  $\mathbf{x}$  i  $\mathbf{w}$  jako próbki (zmiennne losowe) procesu stochastycznego o następujących właściwościach:

$$E\{\mathbf{x}\} = \mathbf{0}, \quad E\{\mathbf{x} \cdot \mathbf{x}^T\} = \mathbf{P} \quad (35)$$

gdzie  $\mathbf{P}$  – macierz kowariancji zmiennej losowej  $\mathbf{x}$

oraz

$$E\{\mathbf{w}\} = \mathbf{0}, \quad E\{\mathbf{w} \cdot \mathbf{w}^T\} = \mathbf{N} \quad (36)$$

gdzie  $\mathbf{N}$  – macierz kowariancji zmiennej losowej  $\mathbf{w}$  (szumu),

zmienne losowe  $\mathbf{x}$  i  $\mathbf{w}$  spełniają z założenia:

$$E\{\mathbf{x} \cdot \mathbf{w}^T\} = \mathbf{0} \quad (37)$$

a więc są nieskorelowane.

Funkcja celu  $J\{\cdot\}$ , pozwalająca na znalezienie struktury liniowego obserwatora,

$$\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}, \mathbf{G} \in R^{n \times r} \quad (38)$$

ma postać:

$$J\{\|\mathbf{e}\|_2^2\} = E\{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2\} \quad (39)$$

Ponieważ błąd estymacji:

$$\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}} = [(x_1 - \hat{x}_1), \dots, (x_n - \hat{x}_n)]^T \quad (40)$$

oraz macierz wag obserwatora

$$\mathbf{G} = \{\mathbf{g}_1, \dots, \mathbf{g}_n\} \quad (41)$$

to minimum funkcji celu w równaniu (39) oznacza spełnienie:

$$\min_{\mathbf{g}_i} E\{(x_i - \hat{x}_i)^2\}, i = 1, \dots, n \quad (42)$$

gdzie  $\mathbf{g}_i$  oznacza  $i$ -ty wiersz macierzy wag  $\mathbf{G}$ .

Spełnienie warunków koniecznych dla równań (42) prowadzi do równania macierzowego:

$$E\{\mathbf{xy}^T\} - \mathbf{G}E\{\mathbf{yy}^T\} = \mathbf{0} \quad (43)$$

Tak więc warunek wystarczający dla istnienia minimum funkcji  $J\{\cdot\}$ :

$$E\{\mathbf{yy}^T\} > 0$$

tzn. macierz (Hessa)  $E\{\mathbf{yy}^T\}$  musi być dodatnio określona.

Ponieważ zachodzi:

$$E\{\mathbf{xy}^T\} = \mathbf{P} \cdot \mathbf{D}^T \quad (44)$$

oraz:

$$E\{\mathbf{yy}^T\} = \mathbf{D} \cdot \mathbf{P} \cdot \mathbf{D}^T + \mathbf{N}, \quad (45)$$

to rozwiązanie równania (43) otrzymuje się jako:

$$\mathbf{G} = \mathbf{P}\mathbf{D}^T (\mathbf{D}\mathbf{P}\mathbf{D}^T + \mathbf{N})^{-1} \quad (46)$$



Stąd zależność (38) określa pierwszą formę estymatora MMSE wektora  $\mathbf{x}$ :

$$\hat{\mathbf{x}}_{MSE} = \mathbf{P}\mathbf{D}^T (\mathbf{D}\mathbf{P}\mathbf{D}^T + \mathbf{N})^{-1} \mathbf{y} \quad (47)$$

gdzie  $(\mathbf{D}\mathbf{P}\mathbf{D}^T + \mathbf{N}) \in R^{r \times r}$ .

Taka forma estymatora jest szczególnie korzystna dla  $r < n$  (liczba obserwacji  $r$  mniejsza niż liczba nieznanymi zmiennymi  $x_i$ ,  $i = 1, \dots, n$ ). W przypadku, gdy macierze kowariancji:  $\mathbf{P}$  oraz  $\mathbf{N}$  są nieosobliwe, proste przekształcenia macierzowe prowadzą do tzw. drugiej formy estymatora:

$$\hat{\mathbf{x}}_{MSE} = (\mathbf{D}^T \mathbf{N}^{-1} \mathbf{D} + \mathbf{P}^{-1})^{-1} \mathbf{D}^T \mathbf{N}^{-1} \mathbf{y} \quad (48)$$

gdzie  $(\mathbf{D}^T \mathbf{N}^{-1} \mathbf{D} + \mathbf{P}^{-1}) \in R^{n \times n}$ .

Ta forma estymatora jest zatem korzystna dla  $r \gg n$ . Co więcej, przy założeniu  $\mathbf{P}^{-1} = \mathbf{0}$ , tzn. przy braku jakiegokolwiek wiedzy a priori o mocy średniej zmiennymi  $x_i$ , oraz przy warunku rząd  $\mathbf{D} = n$ , estymator  $\hat{\mathbf{x}}$  równania (48) przyjmuje postać:

$$\hat{\mathbf{x}} = (\mathbf{D}^T \mathbf{N}^{-1} \mathbf{D})^{-1} \mathbf{D}^T \mathbf{N}^{-1} \mathbf{y} \quad (49)$$

znaną jako estymator Gaussa-Markowa. Nietrudno zauważyć, że estymator Gaussa-Markowa jest uogólnioną postacią estymatora Gaussa (estymator Gaussa  $\rightarrow$  estymator Gaussa-Markowa dla  $\mathbf{N} = \mathbf{1}$ ).

Rozwiązanie problemów obserwacji dane przez estymator (47) oraz (48) wymaga macierzowych operacji, takich jak mnożenie i odwracanie. Według niektórych autorów są zatem nieodpowiednie dla rozwiązywania problemów obserwacji w przypadku wielkich zbiorów danych (*Big Data*) [2]. Powyżej opisane procesy rozwiązywania problemów obserwacji można bezpośrednio wykorzystywać do uzyskania struktur optymalnych liniowych filtrów Kalmana. Warto zauważyć, że w literaturze przedmiotu można znaleźć modyfikacje form estymatorów opisanych powyżej. W szczególności estymator określony zależnością (49) z uwzględnieniem regularyzatora Tichonowa przyjmuje postać [5]:

$$\hat{\mathbf{x}}_R = (\mathbf{D}^T \mathbf{N}^{-1} \mathbf{D} + \delta \mathbf{T})^{-1} \mathbf{D}^T \mathbf{N}^{-1} \mathbf{y} \quad (50)$$

gdzie  $\mathbf{T} > 0$  (macierz dodatnio określona) oraz  $\delta$  parametr regularyzacyjny (określa normę operatora  $\hat{\mathbf{x}}_R$ ).

Także skalowalna wersja estymatora (49) o postaci:

$$\hat{\mathbf{x}}_s = \chi \hat{\mathbf{x}} = \chi (\mathbf{D}^T \mathbf{N}^{-1} \mathbf{D})^{-1} \mathbf{D}^T \mathbf{N}^{-1} \mathbf{y} \quad (51)$$

zawiera parametr regularyzacyjny  $\chi$ . Założenie  $\chi = \chi(\mathbf{y})$  prowadzi do znanego estymatora James-Steina [5].

#### 4. MODEL PRZETWARZANIA SYGNAŁÓW W UCZENIU MASZYNOWYM (MACHINE LEARNING)

Uczenie maszynowe jest jednym z ważnych zagadnień teoretycznych i praktycznych, rozważanych w ramach inteligencji obliczeniowej [9]. Danymi wejściowymi w uczeniu maszynowym są tzw. zbiory treningowe  $T = \{\mathbf{x}_i, d_i\}_{i=1}^N$ , par  $(\mathbf{x}_i, d_i)$ , gdzie zadane wektory  $\mathbf{x}_i$  są punktami pewnej przestrzeni  $X$  ( $\mathbf{x}_i \in X$ ), natomiast  $d_i$  – zadanymi wielkościami skalarnymi ( $d_i \in \mathbb{R}$ ). Zbiory treningowe są źródłami danych dla konstrukcji modeli przetwarzania sygnałów występujących m.in. w zagadnieniach rozpoznawania wzorców, klasyfikacji, filtracji, predykcji i korekcji. Nietrudno zauważyć, że wyżej wspomniane modele można sformułować jako aproksymację funkcji  $f(\cdot)$  wielu zmiennych, spełniających warunki zgodności z danymi pomiarowymi  $d_i$ , tzn.  $f: X \rightarrow \mathbb{R}$ ,  $f(\mathbf{x}_i) \approx d_i$ . Ważną cechą takiej aproksymacji i opartych na niej modeli jest cecha predykcyjności. Jak wiadomo, aproksymacja funkcji jest istotnym zagadnieniem rozważanym w analizie matematycznej. Ze względu na przedmiot niniejszego artykułu, poniżej rozważa się taką aproksymację jako problem estymacji sygnałów losowych.

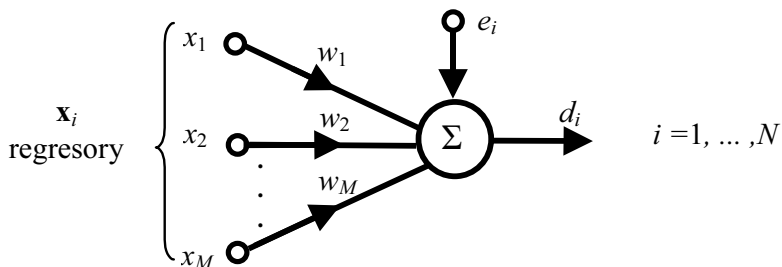
##### 4.1. Aproksymacja funkcji liniowej jako estymacja LS

Zakładając, że punkty  $\mathbf{x}_i = [x_1, x_2, \dots, x_M]^T$  są wektorami regresji (próbkami) wymuszającymi odpowiedzi  $d_i$ , zbiór treningowy  $T = \{\mathbf{x}_i, d_i\}_{i=1}^N$  jako wynik pewnego doświadczenia losowego, generowany jest przez liniowe odwzorowania [7]:

$$d_i = \mathbf{w}^T \mathbf{x}_i + e_i, \quad i = 1, 2, \dots, N \quad (52)$$

gdzie  $\mathbf{w}^T = [w_1, w_2, \dots, w_M]$  – wektor parametrów.

W analogii do filtracji LS zilustrowanej na rysunku 1, schemat prezentujący zależność (52) przedstawiono na rysunku 3.



**Rys. 3.** Estymator sygnałów losowych

**Fig. 3.** The random signals estimator

Parametry w estymatora z rysunku 3 można otrzymać przez minimalizację funkcji kosztów:

$$J_0(\mathbf{w}) = \sum e_i^2(\mathbf{w}) = \sum_{i=1}^N (d_i - \mathbf{w}^T \mathbf{x}_i)^2 \quad (53)$$

$J_0(\mathbf{w})$  jest funkcją kwadratową i jej minimalizacja prowadzi do równań normalnych (por. zależność (28)):

$$\mathbf{w}_{LS} = \mathbf{R}_{xx}^{-1}(N) \cdot \mathbf{r}_{dx}(N) \quad (54)$$

gdzie:

$$\mathbf{R}_{xx}(N) = \sum_{i=1}^N \sum_{j=1}^N \mathbf{x}_i \mathbf{x}_j^T,$$

$\dim \mathbf{R}_{xx}(N) = M \times M$  – macierz autokorelacji wektorów regresji,

$$\mathbf{r}_{dx}(N) = \sum_{i=1}^N \mathbf{x}_i d_i \text{ – wektor korelacji skrośnej.}$$

Należy pamiętać, że w rozwiązaniu (54) zakłada się nieosobliwość macierzy  $\mathbf{R}_{xx}(N)$ . Nieosobliwość takiej symetrycznej macierzy dla dowolnych wektorów regresji realizuje się poprzez tzw. strukturalną regularyzację. W przypadku macierzy  $\mathbf{R}_{xx}(N)$  zmienia się jej przekątną diagonalną i stąd równania normalne otrzymują postać:

$$\mathbf{w}_{RLS} = (\mathbf{R}_{xx}(N) + \gamma \mathbf{I})^{-1} \mathbf{r}_{dx}(N) \quad (55)$$

gdzie:

$\gamma$  – parametr regularyzacji,

$\mathbf{I}$  – macierz jednostkowa,

$\mathbf{w}_{RLS}$  – regularyzowane rozwiązanie najmniejszych kwadratów.

Warto zwrócić uwagę, że tego typu regularyzację stosuje się także w problemach optymalizacyjnych przy obliczaniu odwrotnej macierzy Hessa.

## 4.2. Aproksymacja funkcji nieliniowej jako problem nieliniowej regresji

Jak wspomniano powyżej, zbiór treningowy  $T = \{\mathbf{x}_i, d_i\}_{i=1}^N$ , generowany przez pewien fizyczny proces, może być reprezentowany przez następujący model regresji [9]:

$$d_i = f(\mathbf{x}_i) + e_i, \quad i = 1, 2, \dots, N \quad (56)$$

gdzie:

$\mathbf{x}_i$  – wektor regresji (wymuszenie),

$d_i$  – odpowiedź.

Strukturę takiego modelu można otrzymać przez minimalizację następującej funkcji kosztów:

$$J(e_i^2) = \sum_{i=1}^N (d_i - f(\mathbf{x}_i))^2 + \lambda \|f\|^2 \quad (57)$$

gdzie:

- $\lambda > 0$  – parametr regularyzacyjny,
- $\| \cdot \|$  – norma (w przestrzeni Hilberta).

Ponieważ zdefiniowana powyżej funkcja kosztów ma postać regularyzowanego funkcjonału kwadratowego, minimalizację funkcjonału  $J(\cdot)$  można przeprowadzić, wykorzystując klasyczną teorię regularyzacji Tichonowa. Można więc pokazać, że funkcja  $f(\mathbf{x})$  minimalizująca funkcjonał  $J(\cdot)$  ma postać:

$$f(\mathbf{x}) = \sum_{i=1}^N w_i K(\mathbf{x}, \mathbf{x}_i) \quad (58)$$

gdzie:

- $\mathbf{w}^T = [w_1, w_2, \dots, w_M]$  – wektor parametrów, współczynniki superpozycji,
- $K(\mathbf{x}, \mathbf{x}_i)$  – funkcja jądra (RKHS – *Reproducing Kernel Hilbert Space*).

Należy zauważyć, że aproksymacja funkcji  $f(\mathbf{x})$  w równaniu (58) zadana jest przez superpozycję symetrycznych dodatnio określonych funkcji  $K(\mathbf{x}, \mathbf{x}_i)$ , ciągłych w przestrzeni  $X \times X$ .

Minimalizacja funkcjonału (57) prowadzi do zbioru równań liniowych, których rozwiązanie ustala wektor współczynników  $\mathbf{w}$ , minimalizujących błędy danych treningowych:

$$(\mathbf{K} + \lambda \mathbf{I}) \mathbf{w} = \mathbf{d} \quad (59)$$

gdzie:

- $\mathbf{K}$  – symetryczna macierz kwadratowa ( $N \times N$ ) (tzw. macierz jąder lub Gram),
- $\mathbf{I}$  – macierz jednostkowa,
- $\mathbf{d} = [d_1, d_2, \dots, d_N]^T$ .

$$K_{ij} = K(\mathbf{x}_i, \mathbf{x}_j) = K_{ji} \quad (60)$$

Nietrudno zauważyć, że zawsze można znaleźć taką wartość  $\lambda > 0$ , dla której istnieje stabilne rozwiązanie równania (61):

$$\mathbf{w} = (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{d} \quad (61)$$

Dobór odpowiednich funkcji jąder oraz wartości parametru  $\lambda$  do zadanego zbioru treningowego może być przedmiotem osobnych badań. Warto jednak zauważyć, że jednymi z istotnych form takich funkcji są funkcje Gaussa:

$$K(\mathbf{x}, \mathbf{x}_i) = e^{-\|\mathbf{x}_i - \mathbf{x}\|^2 / 2\sigma^2} \quad (62)$$

W tym przypadku zależność (58) prowadzi do struktury sieci RBF (*Radial Basis Functions*).

### 4.3. Aproksymacja z zastosowaniem przekształceń ortogonalnych

Jak wykazano powyżej, dobór odpowiednich funkcji jąder oraz wartości parametru regularyzującego  $\lambda$  jest przedmiotem badań. Jedną z możliwości doboru funkcji  $K(\mathbf{x}, \mathbf{x}_i)$  polega na wykorzystaniu przekształceń ortogonalnych, co prowadzi do możliwości implementacji aproksymatora z zastosowaniem struktur sztucznych sieci neuronowych. Definicja takiej funkcji dana jest jako [10]:

$$K(\mathbf{x}_i, \mathbf{x}) = \Theta(\mathbf{x}_i^T \mathbf{H}_M \mathbf{x}) \quad (63)$$

gdzie:

$\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{iM}]$ ,  $\mathbf{x}_i \in R^M$  – wektor regresji,

$\mathbf{H}_M$  – macierz skośnie-symetryczna,

$\Theta(\cdot)$  – funkcja nieparzysta.

Ponieważ

$$\Theta(\mathbf{x}_i^T \mathbf{H}_M \mathbf{x}_i) = 0, \forall \mathbf{x}_i \quad (64)$$

oraz

$$\Theta(\mathbf{x}_i^T \mathbf{H}_M \mathbf{x}_j) = -\Theta(\mathbf{x}_j^T \mathbf{H}_M \mathbf{x}_i) \quad (65)$$

to macierz  $\mathbf{K} = \{K_{ij}\} = \{K(\mathbf{x}_i, \mathbf{x}_j)\}$  jest skośnie-symetryczna. Nieosobliwość macierzy  $\mathbf{K}$  można uzyskać poprzez następującą zmianę funkcji  $\Theta(\cdot)$ :

$$\Theta(\cdot) \rightarrow \Theta(\cdot) + \lambda \delta(\cdot) \quad (66)$$

gdzie  $\delta(\cdot)$  jest funkcją Kroneckera.

Stąd równanie (59) przybiera postać:

$$(\mathbf{K} + \lambda \mathbf{I}) \mathbf{w} = \mathbf{d} \quad (67)$$

gdzie:

$\mathbf{K}$  – macierz skośnie-symetryczna,

$\mathbf{I}$  – macierz jednostkowa.

Przekształcenie (66) można zinterpretować jako formę regularyzacji zadania aproksymacji z zastosowaniem macierzy  $\mathbf{H}_M$ , której cechą jest ortogonalizacja przekształconych wektorów  $\mathbf{x}$ , tzn.:  $\mathbf{y} = \mathbf{H}_M \mathbf{x}$ ,  $\langle \mathbf{x}, \mathbf{y} \rangle = 0, \forall \mathbf{x}$ .

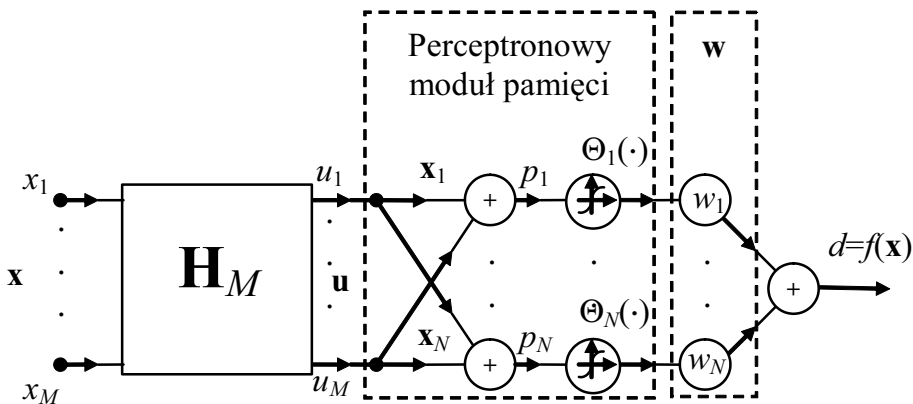
Należy zauważyć, że regularyzacja wykorzystywana w równaniu (57) z zastosowaniem funkcji jąder jest nieskuteczna dla funkcji (63), ponieważ

$$\begin{aligned} \|f\|_{\mathbf{K}}^2 &= \left( \sum_{i=1}^m c_i K(\mathbf{x}_i, \cdot), \sum_{j=1}^m c_j K(\mathbf{x}_j, \cdot) \right) = \sum_{i=1}^m \sum_{j=1}^m c_i c_j (K(\mathbf{x}_i, \cdot), K(\mathbf{x}_j, \cdot)) = \\ &= \sum_{i=1}^m \sum_{j=1}^m c_i c_j (K(\mathbf{x}_i, \mathbf{x}_j)) = \mathbf{c}^T \mathbf{K} \mathbf{c} = 0 \end{aligned} \quad (68)$$

jest seminormą. Tym niemniej rozwiązanie równania (67) jest numerycznie stabilne dla  $\lambda \geq 0$ :

$$\mathbf{w} = (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{d} \quad (69)$$

a formule aproksymacji (58) można nadać formę schematu jak na rysunku 4.



**Rys. 4.** Układowa realizacja formuły aproksymacji  
**Fig. 4.** The implementation of approximation formulas

W układowej realizacji aproksymatora z rysunku 4 zastosowano  $N$  perceptronów jako elementów pamięci zadanych wektorów  $\mathbf{x}_i$ ;  $i = 1, \dots, N$ .

## 5. REALIZACJA UCZENIA MASZYNOWEGO PRZEZ ZASTOSOWANIE ALGEBRY MACIERZY HURWITZA-RADONA

Jak wspomiano powyżej, uczenie maszynowe jest jedynym z uniwersalnych modeli wykorzystywanym w obszarze inteligencji obliczeniowej. Celem przedstawionych rozważań jest pokazanie możliwości rozwiązywania zagadnień uczenia maszynowego poprzez zastosowanie algebry macierzy Hurwitza-Radona (HR) [8, 10]. Co więcej, rodzinę macierzy HR można wykorzystać do tworzenia baz

ortogonalnych najlepiej dopasowanych do zadanych zbiorów sygnałów występujących w różnych zagadnieniach systemów przetwarzania sygnałów. Warto zauważyć, że problemy projektowania takich baz mogą prowadzić do dziedziny, zwanej kwantowym przetwarzaniem sygnałów (QSP – *Quantum Signal Processing* – algorytmiczna emulacja zasad mechaniki kwantowej) [5, 6]. Zbiór rzeczywistych  $N \times N$  macierzy  $\mathbf{A}_j$  spełniających następujące równanie macierzowe:

$$\mathbf{A}_j^2 = -\mathbf{1}, \quad \mathbf{A}_j \mathbf{A}_k + \mathbf{A}_k \mathbf{A}_j = \mathbf{0} \quad (70)$$

dla  $k \neq j, k = 1, \dots, s$

tworzy więc rodzinę HR. Macierze  $\mathbf{A}_j$  są ortogonalne i skośnie-symetryczne, więc:  $\mathbf{A}_j = -\mathbf{A}_j^T, \quad \mathbf{A}_j^{-1} = \mathbf{A}_j^T; j = 1, \dots, s$

Maksymalna liczba macierzy  $s_{\max}$  w rodzinie (70) określona jest przez liczbę Radona  $\rho(N)$ .

Można pokazać, że

$$s_{\max} = \rho(N) - 1 \quad (71)$$

gdzie  $\rho(N) \leq N$  i jedynie dla  $N = 2, 4, 8$

$$\rho(N) = N \quad (72)$$

stąd  $s_{\max} = N - 1; N = 2, 4, 8$

Niech  $\mathbf{A}_1, \dots, \mathbf{A}_s$  będzie rodziną HR macierzy  $\{-1, 0, 1\}$  o wymiarze  $N \times N$ . Wtedy macierz:

$$\mathbf{A}(\mathbf{a}) = \sum_{i=1}^s a_i \mathbf{A}_i \quad (73)$$

gdzie  $\sum_{i=1}^s a_i^2 = 1$

jest macierzą ortogonalną i zależność (73) można traktować jako odwzorowanie kuli  $S^s$  na grupę ortogonalną  $O(N)$ .

Rodzina HR macierzy o wymiarze 8 ( $\dim N = 8$ ) zawiera 7 macierzy. Wszystkie macierze ortogonalne skośnie-symetryczne dla  $N = 8$  mają zgodnie z zależnością (73) postać:

$$\mathbf{H}_8 = \sum_{i=1}^7 h_i \mathbf{A}_i = \begin{bmatrix} 0 & h_1 & h_2 & h_3 & h_4 & h_5 & h_6 & h_7 \\ -h_1 & 0 & h_3 & -h_2 & h_5 & -h_4 & -h_7 & h_6 \\ -h_2 & -h_3 & 0 & h_1 & h_6 & h_7 & -h_4 & -h_5 \\ -h_3 & h_2 & -h_1 & 0 & h_7 & -h_6 & h_5 & -h_4 \\ -h_4 & -h_5 & -h_6 & -h_7 & 0 & h_1 & h_2 & h_3 \\ -h_5 & h_4 & -h_7 & h_6 & -h_1 & 0 & -h_3 & h_2 \\ -h_6 & h_7 & h_4 & -h_5 & -h_2 & h_3 & 0 & -h_1 \\ -h_7 & -h_6 & h_5 & h_4 & -h_3 & -h_2 & h_1 & 0 \end{bmatrix} \quad (74)$$

gdzie  $h_i \in \mathbb{R}$ ,  $i = 1, \dots, 7$ .

Podobnie dla  $N = 16$  istnieją 8 macierzy ( $s_{\max} = 8$ ) w rodzinie HR. Stąd

$$\mathbf{H}_{16} = \begin{bmatrix} & & & & h_8 & & & \mathbf{0} \\ & & & & & \ddots & & \\ & \mathbf{H}_8 & & & & & & \\ & & & & \mathbf{0} & & & h_8 \\ -h_8 & & & & & & & \\ & & \ddots & & & & & \\ \mathbf{0} & & & & & & \mathbf{H}_8^T & \\ & & & & & & & -h_8 \end{bmatrix} = \begin{bmatrix} \mathbf{H}_8 & \mathbf{0} \\ \mathbf{0} & -\mathbf{H}_8 \end{bmatrix} + h_8 \begin{bmatrix} \mathbf{0} & \mathbf{1} \\ -\mathbf{1} & \mathbf{0} \end{bmatrix} \quad (75)$$

gdzie  $h_i \in \mathbb{R}$ ,  $i = 1, \dots, 7$ .

Dla wymiaru  $N = 32$ ,  $\rho(N) = 10$ , stąd  $s_{\max} = 9$ .

W związku z tym

$$\mathbf{H}_{32} = \begin{bmatrix} \mathbf{H}_{16} & \mathbf{0} \\ \mathbf{0} & -\mathbf{H}_{16} \end{bmatrix} + h_9 \begin{bmatrix} \mathbf{0} & \mathbf{1} \\ -\mathbf{1} & \mathbf{0} \end{bmatrix} \quad (76)$$

Rodzina macierzy HR o wymiarach  $N = 2^k$ ,  $k = 6, 7, \dots$  generuje macierz ortogonalną o strukturze:

$$\mathbf{H}_{2^k} = \begin{bmatrix} \mathbf{H}_{2^{k-1}} & \mathbf{0} \\ \mathbf{0} & -\mathbf{H}_{2^{k-1}} \end{bmatrix} + h_K \begin{bmatrix} \mathbf{0} & \mathbf{1} \\ -\mathbf{1} & \mathbf{0} \end{bmatrix} \quad (77)$$

gdzie  $h_K \in \mathbb{R}$ .



Warto zauważyć, że macierze  $\mathbf{H}_{2^k}$  są szczególnie interesujące dla zastosowań w przetwarzaniu sygnałów, ponieważ ich struktura jest kompatybilnym złożeniem macierzy  $\mathbf{H}_8$ . Wykorzystując struktury macierzy  $\mathbf{H}_{2^k}$ , można stworzyć rodzinę analizatorów widma Haara:

$$\mathbf{y} = \frac{1}{1+h_0^2}(\mathbf{H}_{2^k} + h_0\mathbf{1})\mathbf{x} \quad (78)$$

gdzie  $h_0 \geq 0$ ,  $\mathbf{1}$  – macierz jednostkowa.

Rzeczywiście, jeżeli  $\mathbf{x}$  oznacza wektor danych wejściowych, to rezultatem transformacji (78) jest widmo  $\mathbf{y}$  wektora  $\mathbf{x}$  w bazie ortogonalnej, tworzonej przez kolumny macierzy  $(\mathbf{H}_{2^k} + h_0\mathbf{1})$ . Szczególnie interesująca jest możliwość zaprojektowania bazy  $\mathbf{H}_8$  tak, aby dany wektor wejściowy  $\mathbf{x}$  posiadał zadane widmo  $\mathbf{y}$ , tzn:

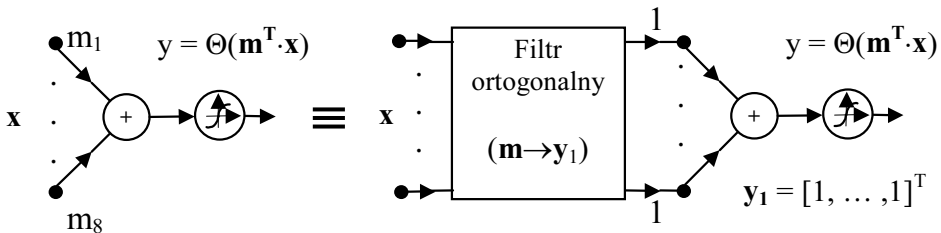
$$\mathbf{y} = \frac{1}{1+h_0^2}(\mathbf{H}_8 + h_0\mathbf{1})\mathbf{x} \quad (79)$$

gdzie  $\mathbf{x} = [x_1, x_2, \dots, x_8]^T$  i  $\mathbf{y} = [y_1, y_2, \dots, y_8]^T$  są ustalone.

Struktura macierzy  $\mathbf{H}_8$ , tworzącej taką bazę dla zależności (79), jest następująca:

$$\begin{bmatrix} h_0 \\ h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \\ h_7 \end{bmatrix} = \frac{1}{\sum_{i=1}^8 y_i^2} \begin{bmatrix} y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8 \\ -y_2 & y_1 & -y_4 & y_3 & -y_6 & y_5 & y_8 & -y_7 \\ -y_3 & y_4 & y_1 & -y_2 & -y_7 & -y_8 & y_5 & y_6 \\ -y_4 & -y_3 & y_2 & y_1 & -y_8 & y_7 & -y_6 & y_5 \\ -y_5 & y_6 & y_7 & y_8 & y_1 & -y_2 & -y_3 & -y_4 \\ -y_6 & -y_5 & y_8 & -y_7 & y_2 & y_1 & y_4 & -y_3 \\ -y_7 & -y_8 & -y_5 & y_6 & y_3 & -y_4 & y_1 & y_2 \\ -y_8 & y_7 & -y_6 & -y_5 & y_4 & y_3 & -y_2 & y_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{bmatrix} \quad (80)$$

Wykorzystując zależność (79) i (80), można podać struktury perceptronu lub korelatora, jak pokazano na rysunku 5.



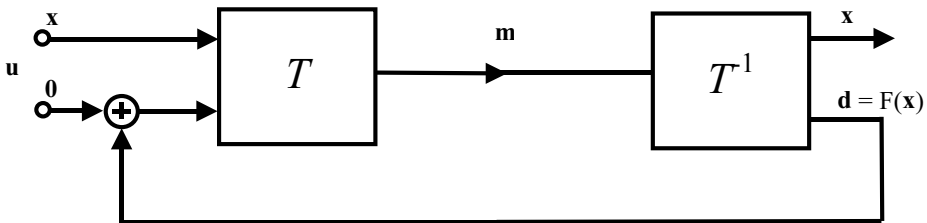
**Rys. 5.** Struktura perceptronu/korelatora  
**Fig. 5.** The structure of perceptron/correlator

Należy zauważyć, że wektory kierunkowe  $\mathbf{m}$  perceptronu/korelatora z rysunku 5 posiadają płaskie widmo w dopasowanej bazie, tzn. z równania (81) [10]:

$$\mathbf{y}_1 = \frac{1}{1 + h_0^2} (\mathbf{H}_8 + h_0 \mathbf{1}) \mathbf{m} \quad (81)$$

gdzie  $\mathbf{y}_1 = [1, 1, \dots, 1]^T$ .

Jako przykład zastosowania algorytmów przetwarzania sygnałów wykorzystujących macierze rodziny HR rozważono model pewnego odwzorowania nieliniowego  $\mathbf{d} = F(\mathbf{x})$ , generowanego przez zbiór treningowy  $\mathcal{S} = \{\mathbf{x}_i, \mathbf{d}_i\}_{i=1}^n$  tzn.  $\mathbf{d}_i = F(\mathbf{x}_i)_{i=1}^n$ . Model taki posiada strukturę pokazaną na rysunku 6, gdzie wykorzystano analizę widmową opartą na macierzach z rodziny HR.



**Rys. 6.** Model odwzorowania nieliniowego danego przez zbiór treningowy

**Fig. 6.** The model of the nonlinear mapping given by training set

Oznaczając wektor treningowy jako:

$$\mathbf{u}_i = \begin{bmatrix} \mathbf{x}_i \\ \mathbf{d}_i \end{bmatrix}, i = 1, \dots, n \quad (82)$$

otrzymuje się widmo Haara wektorów  $\mathbf{u}_i$ :

$$\mathbf{m}_i = \frac{1}{2} \left( \mathbf{H}_{2^k} + \mathbf{1} \right) \mathbf{u}_i, \quad (83)$$

gdzie  $\mathbf{H}_{2^k}$  – macierz skośnie-symetryczna, ortonormalna  $\mathbf{H}_{2^k}^2 = -\mathbf{1}$

oraz macierz widmową:

$$\mathbf{M} = \{\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_n\} \quad (84)$$

Transformacja  $T$ :  $\mathbf{m} = T(\mathbf{u})$  dana jako:

$$\mathbf{m} = \left( 2 \cdot \mathbf{1} - \mathbf{M}_s - \mathbf{H}_{2^k} \right)^{-1} \mathbf{u} \quad (85)$$

gdzie:

$$\mathbf{M}_s = \mathbf{M}(\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \quad (86)$$

$(\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T$  – macierz pseudoodwrotna macierzy  $\mathbf{M}$

oraz transformacja odwrotna  $T^{-1}$ :  $\mathbf{u} = T^{-1}(\mathbf{m})$ :

$$\mathbf{u} = \begin{bmatrix} \mathbf{x}_i \\ \mathbf{d}_i \end{bmatrix} = (-\mathbf{H}_{2^k} + \mathbf{1}) \mathbf{m} \quad (87)$$

Należy zauważyć, że z równania (85) otrzymuje się:

$$\hat{\mathbf{m}}_i = T \left( \begin{bmatrix} \mathbf{x}_i \\ \mathbf{0} \end{bmatrix} \right) - \text{estymator widma } \mathbf{m}_i, i = 1, 2, \dots, n \quad (88)$$

oraz

$$\hat{\mathbf{d}}_i = F(\mathbf{x}_i) - \text{estymator odpowiedzi } \mathbf{d}_i$$

W procesie rekurencyjnym (zamknięta pętla sprzężenia zwrotnego) dostaje się:

$$\hat{\mathbf{d}}_i \rightarrow \mathbf{d}_i = F(\mathbf{x}_i) \quad (89)$$

## PODSUMOWANIE

W artykule dokonano przeglądu wybranych metod estymacji sygnałów losowych z wykorzystaniem estymatorów MMSE. Oprócz klasycznych metod związanych z estymatorem Gaussa oraz metodami Wienera, omówiono także nowsze metody, wynikające z zastosowania metod uczenia maszynowego (*machine learning*), które prowadzą do struktur sztucznych sieci neuronowych. Na szczególną uwagę zasługują tutaj struktury, oparte na wykorzystaniu rodziny macierzy Hurwitza-Radona. Struktury takie można traktować jako implementację metod kwantowego przetwarzania sygnałów (*Quantum Signal Processing*). Warto zwrócić uwagę, że ze względu na charakter artykułu nie poruszono tutaj szeregu aktualnych zagadnień z dziedziny przetwarzania optymalnego sygnałów, takich jak metody tensorowe, ślepe (*blind*) oraz przetwarzanie wielkich zbiorów danych (*Big Data*).

## LITERATURA

1. Brammer K., Siffling G., *Kalman-Bucy-Filter*, R. Oldenbourg Verlag, 1975.
2. Cevher V., Becker S., Schmidt N., *Convex Optimization for Big Data*, Signal Processing Magazine, September 2014, vol. 31, no. 5.
3. Clarkson P.M., *Optimal and Adaptive Signal Processing*, CRC Press, Inc., 1993.
4. Eckmann B., *Topology, Algebra, Analysis-Relations and Missing Links*, Notices of the AMS, 1999, vol. 46, no. 5.
5. Eldar Y.C., *Quantum Signal Processing*, Ph. D. Dissertation, MIT, 2001.
6. Eldar Y.C., Oppenheim A.V., *Quantum Signal Processing*, IEEE Signal Proc. Magazine, 2002, vol. 19, no. 6.
7. Haykin S., *Neural Networks and Learning Machines*, Pearson Education, Inc., 2009.
8. Jakóbczak D., *Zastosowanie dyskretnego operatora Hurwitza-Radona*, rozprawa doktorska, Polsko-Japońska Wyższa Szkoła Technik Komputerowych, Warszawa 2006.
9. Poggio T., Smale S., *The Mathematics of Learning: Dealing with Data*, Notices of the AMS, 2003, vol. 50, no. 5, s. 537–544.
10. Sieńko W., Citko W., *Hamiltonian Neural Networks Based Networks for Learning*, [w:] *Machine Learning*, red. A. Mellouk, A. Chebira, I-Tech, Vienna 2009, s. 75–92.

## MODELS OF RANDOM SIGNALS OPTIMAL PROCESSING

### Summary

*Estimation of random signals is a very important tool for design of different systems. The goal of this article is a review of Classical methods and newer, e.g. machine learning, structures of MMSE estimators.*

**Keywords:** *optimal signal processing, estimation of random signals.*