



## Wprowadzenie

Klasa blokowych modeli nieliniowych obejmuje złożone modele, które składają się z liniowych systemów dynamicznych i nieliniowych elementów statycznych. Modele Wienera i Hammersteina są najbardziej znanymi i najczęściej wdrażanymi przedstawicielami tej klasy. Model nazywany jest modelem Wienera, jeśli liniowy blok (element) dynamiczny poprzedza nieliniowy element statyczny. W modelu Hammersteina kolejność połączeń jest odwrócona.

Modele nieliniowych elementów statycznych mogą być realizowane w różnych formach, takich jak wielomiany, splajny, funkcje bazowe, falki, sieci neuronowe, tablice przeglądowe i modele rozmyte. Modele odpowiedzi impulsowej, modele przenoszenia impulsów i modele przestrzeni stanów są powszechnymi reprezentacjami liniowych systemów dynamicznych. W zależności od form realizacji obu tych elementów można uzyskać różne struktury modeli Wienera i Hammersteina. Aby je ocenić i porównać, powszechnie bierze się pod uwagę następujące właściwości: dokładność aproksymacji, zachowanie ekstrapolacyjne, zachowanie interpolacyjne, gładkość, wrażliwość na szum, dostępne metody optymalizacji parametrów i dostępne metody optymalizacji struktury.

Z twierdzenia Weierstrassa o aproksymacji wynika, że każda funkcja ciągła zdefiniowana na przedziale  $[a, b]$  może być dowolnie ściśle aproksymowana wielomianem. Modele wielomianowe są szeroko stosowane jako modele elementów nieliniowych. Dużą zaletą modeli wielomianowych jest efektywna optymalizacja parametrów, którą można przeprowadzić off-line metodą najmniejszych kwadratów lub on-line jej rekurencyjną wersją. Co więcej, wybór struktury można również skutecznie przeprowadzić za pomocą algorytmu ortogonalnych najmniejszych kwadratów, w którym zbiór regresorów jest przekształcany w zbiór ortogonalnych wektorów bazowych. Dzięki temu algorytmowi możliwe jest obliczenie indywidualnego wkładu każdego wektora bazowego w wariancję wyjściową.

Wielomiany mają również pewne fundamentalne wady. Po p i e r w s z e , chociaż każda funkcja ciągła może być dowolnie blisko przybliżona wielomianem, niektóre funkcje nieliniowe wymagają bardzo wysokiego rzędu wielomianu. W modelach z wieloma wejściami i wieloma wyjściami liczba parametrów rośnie silnie wraz ze wzrostem liczby wejść. Zwiększa to niepewność modelu.

A. Janczak: Identification of Nonlinear Systems, LNCIS 310, pp. 1-30, 2005.  
© Springer-Verlag Berlin Heidelberg 2005



Wadą wielomianów jest ich oscylacyjna interpolacja i ekstrapolacja. Inną wadą wielomianów są ich oscylacyjne właściwości interpolacyjne i ekstrapolacyjne. Dlatego w praktyce stosowanie modeli wielomianowych jest zalecane tylko w niektórych szczególnych przypadkach, w których można założyć, że struktura systemu jest w przybliżeniu typu wielomianowego.

Alternatywą dla modeli wielomianowych są modele sieci neuronowych o architekturze perceptronu wielowarstwowego. Perceptrony wielowarstwowe są sieciami neuronowymi typu feedforward zawierającymi jedną lub więcej ukrytych warstw elementów nieliniowych, ale w praktyce najczęściej wybierana jest jedna warstwa ukryta. Zastosowanie perceptronów wielowarstwowych w problemach aproksymacyjnych jest uzasadnione ich uniwersalną właściwością aproksymacyjną, która stwierdza, że pojedyncza warstwa ukryta jest wystarczająca do równomiernej aproksymacji dowolnej funkcji ciągłej ze wsparciem w jednostkowym sześcianie [31]. Perceptrony wielowarstwowe, ze względu na swoje zalety, takie jak wysoka dokładność aproksymacji, mniejsza liczba węzłów i wag w porównaniu z innymi architekturami modeli, zdolność do generowania szerokiej gamy funkcji, są najczęściej stosowanymi sieciami neuronowymi. W przeciwieństwie do wielomianów, wielowarstwowe modele perceptronowe nie wykazują oscylacyjnych zachowań interpolacyjnych i ekstrapolacyjnych. Wykazują one tendencję do interpolacji monotonicznej. Zachowanie ekstrapolacji jest również płynne, ale w długim zakresie odpowiedzi sieci ma tendencję do stałej wartości ze względu na nasycenie powszechnie stosowanych funkcji sigmoidalnych. Wielowarstwowe modele perceptronowe są również bardzo przydatne w przypadku problemów o dużej wymiarowości. Wynika to z faktu, że liczba wag w wielowarstwowym modelu perceptronowym jest proporcjonalna do liczby wejść. W przeciwieństwie do modeli wielomianowych, wielowarstwowe modele perceptronowe mogą być z powodzeniem stosowane nie tylko do reprezentowania systemów opisanych wielomianami, ale także nieskończonymi rozwinięciami szeregów potęgowych. Podsumowując, wielowarstwowe modele perceptronowe są uzupełnieniem modeli wielomianowych, ponieważ oferują użytkownikowi kilka różnych interesujących funkcji.

Oczywiście wielowarstwowe modele perceptronowe nie są wolne od wad. Najważniejsze z nich to wykorzystanie lokalnych metod optymalizacji do aktualizacji wag oraz ryzyko uwięzienia w płytkim minimum lokalnym. Prowadzi to często do konieczności powtarzania treningu z różnymi inicjalizacjami wag. Co więcej, techniki prób i błędów muszą być stosowane dla niektórych parametrów, takich jak wagi początkowe, szybkość uczenia się itp. Ponadto dostępne metody optymalizacji struktury modelu są dość intensywnie obliczeniowo.

Obecnie, wraz z rozwojem identyfikacji systemów Wienera i Hammersteina, możliwe jest usystematyzowanie różnych technik i przedstawienie ich w ujednoliconych ramach. Próbowujemy dokonać takiej prezentacji w niniejszej monografii poprzez przegląd istniejących podejść wraz z prezentacją oryginalnych prac badawczych i niektórych wyników, które nie zostały jeszcze opublikowane.

Modele sieci neuronowych systemów Wienera i Hammersteina rozważane w rozdziałach 2 i 3 składają się z wielowarstwowego modelu perceptronowego

elementu nieliniowego i jednego lub więcej węzłów liniowych z zaczepionymi liniami opóźniającymi stanowiący model liniowego układu dynamicznego. Szeregowo-równoległe modele Wienera zawierają również inny wielowarstwowy model perceptronowy odwrotnego elementu nieliniowego.

Omówiono dwie podstawowe konfiguracje modeli, tj. modele szeregowo-równoległe i równoległe. W modelach szeregowo-równoległych gradient można obliczyć za pomocą dobrze znanej metody propagacji wstecznej. W modelach równoległych za pomocą metody propagacji wstecznej można uzyskać jedynie przybliżoną wartość gradientu. W związku z tym należy wziąć pod uwagę dwie inne metody, określane jako metoda wrażliwości i metoda propagacji wstecznej w czasie, które zapewniają dokładną wartość gradientu lub jego dokładniejsze przybliżenie. Wszystkie te metody obliczania gradientu zostały wyprowadzone w ujednolicony sposób, zarówno dla przypadków SISO, jak i MIMO. Złożoność obliczeniowa metod jest analizowana i wyrażana w postaci rzędów wielomianów oraz liczby kroków czasowych w przypadku metody obciążonej propagacji wstecznej w czasie. Przeanalizowano również dokładność obliczania gradientu metodą skróconej propagacji wstecznej w czasie. Wykazano, że dokładność obliczania gradientu zależy od liczby dyskretnych kroków czasowych niezbędnych do zmniejszenia odpowiedzi impulsowych modeli wrażliwości do nieistotnie małych wartości. Na podstawie tego wyniku zaproponowano adaptacyjne procedury dostosowywania liczby rozwiniętych kroków czasowych w celu spełnienia określonych stopni dokładności. Oryginalny wkład książki obejmuje nowe podejścia do identyfikacji systemów Wienera i Hammersteina oraz pewne istotne wyniki teoretyczne. Dla modeli sieci neuronowych SISO i MIMO wyprowadzono i przeanalizowano następujące metody obliczania gradientu:

- Propagacja wsteczna dla szeregowo-równoległych modeli Wienera.
- Propagacja wsteczna dla równoległych modeli Wienera.
- Metoda wrażliwości dla równoległych modeli Wienera.
- Obciążona propagacja wsteczna w czasie dla równoległych modeli Wienera.
- Propagacja wsteczna dla szeregowo-równoległych modeli Hammersteina.
- Propagacja wsteczna dla równoległych modeli Hammersteina.
- Metoda wrażliwości dla równoległych modeli Hammersteina.
- Obciążona propagacja wsteczna w czasie dla równoległych modeli Hammersteina.

Mając wyprowadzone zasady obliczania gradientu, można łatwo zaimplementować różne algorytmy uczenia oparte na gradiencie. Sekwencyjne wersje algorytmu stromego zejścia, błędu predykcji oraz połączonego algorytmu stromego zejścia i rekurencyjnego algorytmu najmniejszych kwadratów zostały szczegółowo omówione w rozdziałach 2 i 3.

Inna grupa metod identyfikacji, wyprowadzona i omówiona w książce, wykorzystuje wielomianową reprezentację systemów Wienera:

- Metoda najmniejszych kwadratów dla wielomianowych systemów Wienera z członem liniowym.
- Metoda najmniejszych kwadratów dla wielomianowych systemów Wienera bez członu liniowego.
- Połączona metoda najmniejszych kwadratów i zmiennych instrumentalnych dla wielomianowych systemów Wienera z członem liniowym.
- Połączona metoda najmniejszych kwadratów i zmiennych instrumentalnych dla wielomianowych systemów Wienera bez członu liniowego.

- Rekursywna metoda błędu predykcji.
- Rekursywna metoda regresji pseudolinearnej.

Wszystkie te metody identyfikacji oparte na wielomianach wykorzystują reprezentację dynamiki systemu za pomocą funkcji transferu impulsów oraz wielomianowe modele elementu nieliniowego lub jego odwrotności. Pomimo faktu, że wielomianowe modele systemów Wienera i Hammersteina można wyrazić w postaci liniowej w parametrach, takie przekształcenia prowadzą do nadmiarowości parametrów, ponieważ przekształcone modele mają większą liczbę parametrów niż modele oryginalne. Liczba parametrów przekształconych modeli silnie rośnie wraz ze wzrostem rzędu modelu. W rezultacie wzrasta błąd wariancji i mogą wystąpić pewne problemy numeryczne. Ponadto, jak pokazano w rozdziale 4, aby przekształcić system Wienera do postaci liniowej w parametrach, nieliniowość musi być odwracalna, tj. odwzorowanie nieliniowe musi być ściśle monotoniczne. Wykazano również, że oszacowania parametrów metodą najmniejszych kwadratów nie są spójne. Aby uzyskać spójne oszacowania parametrów, zaproponowano połączoną metodę najmniejszych kwadratów i zmiennych instrumentalnych. Restrykcyjne założenie odwracalności elementu nieliniowego nie jest już konieczne zarówno w metodzie błędu predykcji, jak i metodzie regresji pseudoliniowej. W tym przypadku jednak model Wienera jest nieliniowy w parametrach, a estymacja parametrów staje się zadaniem optymalizacji nieliniowej.

Modele Wienera i Hammersteina znalazły liczne zastosowania przemysłowe do modelowania systemów, sterowania, wykrywania i izolowania błędów. Rozdział 6 zawiera krótki przegląd zastosowań, które obejmują następujące systemy i procesy:

- Proces neutralizacji pH,
- wymienniki ciepła,
- kolumny destylacyjne,
- proces separacji chromatograficznej,
- reaktor polimeryzacji,
- Czujniki z mikrowagą kwarcową z powłoką polimerową,
- instalacje hydrauliczne,
- elektrohydrauliczne serwosystemy,
- zawory pneumatyczne,
- systemy pompa-zawór,
- elektrooptyczne systemy dynamiczne,
- wzmacniacze dużej mocy w kanałach komunikacji satelitarnej,
- głośniki,
- aktywne systemy redukcji szumów,
- proces ładowania w silnikach wysokoprężnych,
- proces chłodzenia granulatu tlenku żelaza,
- parownik cukru.

Modele Wienera i Hammersteina ujawniają zdolność do opisywania szerokiej klasy różnych systemów i oprócz przykładów przemysłowych, istnieje wiele innych zastosowań w biologii i medycynie.





## 1.1 Modele systemów dynamicznych

Niniejsza sekcja zawiera krótki przegląd modeli dyskretnych systemów dynamicznych niezmiennych w czasie. Tytułem wprowadzenia, sekcja rozpoczyna się od struktur modeli liniowych i pokazuje modele nieliniowe jako uogólnienia modeli liniowych.

### 1.1.1 Modele liniowe

Według Ljunga [112] model niezmienniczy w czasie może być określony przez odpowiedź impulsową  $h(n)$ , widmo  $\Phi(\omega)$  zakłócenia addytywnego  $H(q^{-1})\varepsilon(n)$  oraz funkcję gęstości prawdopodobieństwa  $f_\varepsilon(\cdot)$  zakłócenia  $\varepsilon(n)$ . Wyjście  $y(n)$  dyskretnego modelu liniowego wzbudzonego przez wejście  $u(n)$  i zaburzonego addytywnie przez  $\varepsilon(n)$  wynosi

$$y(n) = G(q^{-1})u(n) + H(q^{-1})\varepsilon(n), \quad (1.1)$$

gdzie  $G(q^{-1})$  jest nazywane funkcją przenoszenia impulsu wejściowego,  $H(q^{-1})$  jest funkcją przenoszenia impulsu szumu, a  $q^{-1}$  oznacza operator przesunięcia wstecznego. O ile nie określono dokładniej, zakładamy, że  $\varepsilon(n)$  jest stacjonarnym, niezależnym procesem stochastycznym o zerowej średniej. Przedstawienie  $G(q^{-1})$  i  $H(q^{-1})$  jako funkcji racjonalnych prowadzi do ogólnej struktury modelu o następującej postaci [112, 149]:

$$A(q^{-1})y(n) = \frac{B(q)^{-1}}{F(q^{-1})} u(n) + \frac{C(q)^{-1}}{D(q^{-1})} \varepsilon(n), \quad (1.2)$$

gdzie

$$A(q^{-1}) = 1 + a_1 q_1^{-1} + \dots + a_{na} q_{na}^{-na}, \quad (1.3)$$

$$B(q^{-1}) = b_1 q_1^{-1} + \dots + b_{nb} q_{nb}^{-nb}, \quad (1.4)$$

$$C(q^{-1}) = 1 + c_1 q_1^{-1} + \dots + c_{nc} q_{nc}^{-nc}, \quad (1.5)$$

$$D(q^{-1}) = 1 + d_1 q_1^{-1} + \dots + d_{nd} q_{nd}^{-nd}, \quad (1.6)$$

$$F(q^{-1}) = 1 + f_1 q_1^{-1} + \dots + f_{nf} q_{nf}^{-nf}. \quad (1.7)$$

W praktyce struktura (1.2) jest zazwyczaj zbyt ogólna. W zależności od tego, które z wielomianów (1.3) - (1.7) są używane, można wyróżnić 32 różne zestawy modeli. Poniżej wymieniono kilka powszechnie stosowanych struktur, które należą do ogólnej rodziny struktur.

Struktura skończonej odpowiedzi impulsowej (FIR). Wybór  $A(q^{-1}) = C(q^{-1}) = D(q^{-1}) = F(q^{-1}) = 1$  skutkuje najprostszą strukturą modelu znaną jako model o skończonej odpowiedzi impulsowej.

$$y(n) = B(q^{-1})u(n) + \varepsilon(n). \quad (1.8)$$

Wynik modelu (1.8) jest ważoną sumą  $nb$  przeszłych danych wejściowych  $u(n-1)$ ,  
...,  $u(n-nb)$ :

$$y(n) = b_1 u(n-1) + \dots + b_{nb} u(n-nb) + \varepsilon(n). \quad (1.9)$$

Optymalny predyktor z jednokrokovym wyprzedzeniem, tj. predyktor, który minimalizuje wariancję błędu predykcji, to

$$\hat{y}(n|n-1) = B(q^{-1})u(n) = b_1 u(n-1) + \dots + b_{nb} u(n-nb). \quad (1.10)$$

Wprowadzenie wektora parametrów  $\theta$ ,

$$\theta = \begin{bmatrix} b_1 \dots b_{nb} \end{bmatrix}^T, \quad (1.11)$$

i wektor regresji  $x(n)$ ,

$$x(n) = \begin{bmatrix} u(n-1) \dots u(n-nb) \end{bmatrix}^T, \quad (1.12)$$

(1.10) może być równoważnie wyrażona w postaci regresji:

$$\hat{y}(n|n-1) = x^T(n)\theta. \quad (1.13)$$

Model FIR jest w stanie dość dobrze aproksymować asymptotycznie stabilne systemy dynamiczne, jeśli ich odpowiedzi impulsowe zanikają dość szybko.

Struktura autoregresyjna (AR). Struktura modelu AR jest zdefiniowana poprzez wybór  $B(q^{-1}) = 0$ , oraz  $C(q^{-1}) = D(q^{-1}) = F(q^{-1}) = 1$ :

$$y(n) = \frac{1}{A(q^{-1})} \varepsilon(n). \quad (1.14)$$

W tym przypadku wektor parametrów  $\theta$  i wektor regresji  $x(n)$  to

$$\theta = \begin{bmatrix} a_1 \dots a_{na} \end{bmatrix}^T, \quad (1.15)$$

$$x(n) = \begin{bmatrix} -y(n-1) \dots -y(n-na) \end{bmatrix}^T. \quad (1.16)$$

Struktura średniej ruchomej (MA). Struktura modelu MA odpowiada wyborowi  $B(q^{-1}) = 0$  i  $A(q^{-1}) = D(q^{-1}) = F(q^{-1}) = 1$ :

$$y(n) = C(q^{-1})\varepsilon(n). \quad (1.17)$$

Wektor parametrów  $\theta$  i wektor regresji  $x(n)$  są następujące

$$\theta = \begin{bmatrix} c_1 \dots c_{nc} \end{bmatrix}^T, \quad (1.18)$$

$$x(n) = \begin{bmatrix} \varepsilon(n-1) \dots \varepsilon(n-nc) \end{bmatrix}^T. \quad (1.19)$$

Struktura autoregresyjna z egzogenicznymi danymi wejściowymi (ARX). Strukturę modelu ARX można uzyskać, wybierając  $C(q^{-1}) = D(q^{-1}) = F(q^{-1}) = 1$ :

$$y(n) = \frac{B(q)^{-1}}{A(q^{-1})} u(n) + \frac{1}{A(q^{-1})} \varepsilon(n). \quad (1.20)$$

Dla modelu ARX wektor parametrów  $\theta$  i wektor regresji  $x(n)$  mają postać

$$\theta = a_1 \dots a_{na} \ b_1 \dots b_{nb}^T, \quad (1.21)$$

$$x(n) = -y(n-1) \dots -y(n-na) u(n-1) \dots u(n-nb)^T. \quad (1.22)$$

Struktura autoregresyjnej średniej ruchomej (ARMA). Połączenie modelu autoregresyjnego i modelu średniej ruchomej daje model ARMA. Można go uzyskać przy wyborze  $B(q^{-1}) = 0$  i  $D(q^{-1}) = F(q^{-1}) = 1$ :

$$y(n) = \frac{C(q)^{-1}}{A(q^{-1})} \varepsilon(n). \quad (1.23)$$

W tym przypadku wektor parametrów  $\theta$  i wektor regresji  $x(n)$  to

$$\theta = a_1 \dots a_{na} \ c_1 \dots c_{nc}^T, \quad (1.24)$$

$$x(n) = -y(n-1) \dots -y(n-na) \varepsilon(n-1) \dots \varepsilon(n-nc)^T. \quad (1.25)$$

Autoregresyjna średnia ruchoma z egzogenicznymi danymi wejściowymi (ARMAX). Model ARMAX jest najbardziej ogólną strukturą ze wszystkich rozważanych do tej pory, ponieważ zawiera wszystkie z nich jako przypadki szczególne. Aby uzyskać model ARMAX, wybieramy  $D(q^{-1}) = F(q^{-1}) = 1$ :

$$y(n) = \frac{B(q)^{-1}}{A(q^{-1})} u(n) + \frac{C(q)^{-1}}{A(q^{-1})} \varepsilon(n). \quad (1.26)$$

Dla modelu ARMAX, wektor parametrów  $\theta$  i wektor regresji  $x(n)$  są zdefiniowane w następujący sposób:

$$\theta = a_1 \dots a_{na} \ b_1 \dots b_{nb} \ c_1 \dots c_{nc}^T, \quad (1.27)$$

$$x(n) = -y(n-1) \dots -y(n-na) u(n-1) \dots u(n-nb) \varepsilon(n-1) \dots \varepsilon(n-nc)^T. \quad (1.28)$$

Struktura błędu wyjściowego (OE). Model OE można uzyskać, jeśli wybierzemy

$A(q^{-1}) = C(q^{-1}) = D(q^{-1}) = 1$ :

$$y(n) = \frac{B(q)^{-1}}{F(q^{-1})} u(n) + \varepsilon(n). \quad (1.29)$$

W tym przypadku wektor parametrów  $\theta$  i wektor regresji  $x(n)$  są zdefiniowane jako

$$\theta = f_1 \dots f_{nf} \ b_1 \dots b_{nb}^T, \quad (1.30)$$

$$x(n) = -\hat{y}(n-1) \dots -\hat{y}(n-nf) u(n-1) \dots u(n-nb)^T, \quad (1.31)$$

gdzie

$$\hat{y}(n) = \frac{B(q)^{-1}}{F(q^{-1})} u(n). \quad (1.32)$$

Struktura Box-Jenkins (BJ). Struktura, która jest bardziej ogólnym rozwinięciem modelu OE, nazywana jest modelem Boxa-Jenkinsa. Aby uzyskać model BJ, należy wybrać  $A(q^{-1}) = 1$ :

$$y(n) = \frac{B(q)^{-1}}{F(q^{-1})} u(n) + \frac{C(q)^{-1}}{D(q^{-1})} \varepsilon(n). \quad (1.33)$$

Predyktor z jednokrokovym wyprzedzeniem dla modelu BJ ma postać [112]:

$$\hat{y}(n/n-1) = \frac{D(q^{-1})B(q)^{-1}}{C(q^{-1})F(q^{-1})} u(n) + \frac{C(q^{-1}) - D(q^{-1})}{C(q^{-1})} y(n). \quad (1.34)$$

### 1.1.2 Modele nieliniowe

Nieliniowe odpowiedniki liniowych struktur modeli można zdefiniować przy założeniu, że istnieje nieliniowa zależność między rzeczywistym wyjściem systemu a przeszłymi wejściami systemu, przeszłymi wyjściami systemu lub modelu oraz rzeczywistymi i przeszłymi zakłóceniami addytywnymi. Dla nieliniowych modeli wejścia-wyjścia mamy

$$y(n) = g(x(n), \theta) + \varepsilon(n) \quad (1.35)$$

lub w formie predyktora,

$$\hat{y}(n/n-1) = g(x(n), \theta). \quad (1.36)$$

W zależności od postaci  $x(n)$  można zdefiniować nieliniowe struktury modelu znane jako NFIR, NAR, NMA, NARX, NARMAX, NOE, NBJ. The Funkcja  $g(\cdot)$  jest nieliniowym odwzorowaniem, które dla danego  $\theta$  odwzorowuje  $\mathbb{R}^d$  na  $\mathbb{R}^p$ , gdzie  $d$  jest liczbą regresorów (elementów wektora  $x(n)$ ), a  $p$  jest liczbą wyników modelu. W podejściu parametrycznym,  $g(\cdot)$  jest wyrażone przez rozwinięcie funkcji:

$$g(x(n), \theta) = \sum_{m=1}^{\infty} \theta_m g_{mm}(x(n)), \quad (1.37)$$

gdzie  $g_m(\cdot)$  nazywa się funkcją bazową, a  $\theta = [\theta_1 \dots \theta_{ng}]^T$ . Istnieją różne formy ekspansji funkcji, stosowane do reprezentacji systemów nieliniowych, które są oparte na wielomianach, jądrach Volterry, szeregach Fouriera, funkcjach jednostajnie stałych, radialnych funkcjach bazowych, falkach, estymatorach jądra, sieciach neuronowych i modelach rozmytych [112].

Modele Volterry czasu dyskretnego. Jeśli funkcja  $g(\cdot)$  jest analityczna, wówczas odpowiedź systemu może być reprezentowana przez szereg Volterry [109]:

$$y(n) = \sum_{i_1=1}^n h_1(i_1)u(n-i_1) + \sum_{i_1=1}^n \sum_{i_2=1}^n h_2(i_1, i_2)u(n-i_1)u(n-i_2) + \dots + \varepsilon(n), \quad (1.38)$$

gdzie funkcje jądra  $h_j(i_1, \dots, i_j)$ ,  $j = 1, 2, \dots$ , opisują dynamiki systemu. Dwa podstawowe problemy związane z praktycznym zastosowaniem szeregów Volterry to trudności związane z pomiarem funkcji jądra Volterry oraz zbieżność szeregów Volterry [145]. Reprezentacja szeregu Volterry jest tak naprawdę szeregiem Taylora z pamięcią. W związku z tym problem zbieżności jest taki sam jak w przypadku reprezentacji funkcji za pomocą szeregu Taylora, tzn. reprezentacja układu za pomocą szeregu Volterry może być zbieżna tylko dla ograniczonego zakresu amplitudy wejściowej układu. Aby obejść ten problem, Wiener utworzył nowy zestaw funkcjonałów z funkcjonałów Volterry - G-funkcjonały, które mają właściwość ortogonalności, gdy ich wejście jest procesem gaussowskim.

Szeroko zakrojone badania nad modelami Volterry pokazują ich skuteczne zastosowanie do systemów niskiego rzędu. Przyczyną tego ograniczenia rzędu systemu są praktyczne trudności w rozszerzeniu estymacji jądra na rzędy wyższe niż trzeci [117].

Modele Kołmogorowa-Gabora. Zastosowanie uogólnionych modeli wielomianowych do reprezentacji nieliniowych systemów dynamicznych

$$y(n) = f(u(n-1), \dots, u(n-nb), y(n-1), \dots, y(n-na)) + \varepsilon(n) \quad (1.39)$$

daje w wyniku modele Kołmogorowa-Gabora:

$$y(n) = a_0 + \sum_{i_1=1}^{nab} a_{i_1} x_{i_1}(n) + \sum_{i_1=1}^{nab} \sum_{i_2=1}^{nab} a_{i_1 i_2} x_{i_1}(n) x_{i_2}(n) + \dots + \sum_{i_1=1}^{nab} \sum_{i_2=1}^{nab} \dots \sum_{i_l=1}^{nab} a_{i_1 i_2 \dots i_l} x_{i_1}(n) x_{i_2}(n) \dots x_{i_l}(n) + \varepsilon(n), \quad (1.40)$$

gdzie  $nab = na + nb$ ,

$$x_j(n) = \begin{cases} u(n-j) & \text{jeśli } 1 \leq j \\ y(n-j+nb), & \text{jeśli } nb < j \leq nab. \end{cases} \quad (1.41)$$

Liczba parametrów  $M$  w (1.40) silnie wzrasta wraz ze wzrostem  $nab$  lub  $l$  [123]:

$$M = \frac{(l+nab)!}{l!nab!}. \quad (1.42)$$

Duża złożoność modelu Kołmogorowa-Gabora ogranicza jego praktyczną stosowalność i prowadzi do zredukowanych modeli wielomianowych zawierających tylko wybrane wyrażenia (1.40). Należy zauważyć, że (1.40) ma strukturę NARX. Inne struktury nieliniowe, takie jak NFIR, NAR, NMA, NARMA, NARX, NARMAX, NBJ można uzyskać poprzez odpowiednią redefinicję  $x_j(n)$ .

Nieliniowe modele ortonormalnych funkcji bazowych (NOBF). Główną wadą zarówno modeli FIR, jak i NFIR jest to, że wiele parametrów może być potrzebnych do odpowiedniego opisanie systemu, jeśli jego odpowiedź impulsowa zanika powoli. Wadę tę można zmniejszyć poprzez wprowadzenie filtrów liniowych, które uwzględniają wcześniejszą wiedzę na temat dynamiki procesu. Powszechnie stosowane są ortonormalne filtry Laguerre'a i Kautza, które mają ortonormalne odpowiedzi impulsowe. Filtr Laguerre'a można opisać za pomocą tylko jednego parametru  $\alpha$ , rzeczywistego bieguna

$$L_k(q) = \frac{1}{q - \alpha} \frac{1 - \alpha q}{q - \alpha} \dots \frac{1 - \alpha q}{q - \alpha} \quad (1.43)$$

Dlatego ten rodzaj filtrów jest odpowiedni do modelowania systemów o dobrze tłumionym zachowaniu. W przypadku systemów z zachowaniem rezonansowym filtry Kautza są odpowiednie, ponieważ mają złożoną parę biegunów. W praktyce stosuje się oszacowania dominującego bieguna systemu lub dominujących sprzężonych biegunów złożonych. Wektor regresji dla modelu NOBF ma postać

$$x(n) = L(q) L(q) \dots L(q)^T. \quad (1.44)$$

### 1.1.3 Modele szeregowo-równoległe i równoległe

Modele systemów dynamicznych mogą być używane w dwóch podstawowych konfiguracjach: konfiguracji przewidywania lub konfiguracji symulacji. Konfiguracja predykcji pozwala na przewidywanie przyszłych wyników systemu w oparciu o przeszłe dane wejściowe i wyjściowe systemu.

Przykładami konfiguracji predykcji są modele FIR, AR, ARX w przypadku liniowym oraz modele NFIR, NOBF, NAR, NARX w przypadku nieliniowym. W konfiguracji symulacji przyszłe wyniki systemu są również przewidywane, ale tylko na podstawie przeszłych danych wejściowych systemu bez wykorzystywania przeszłych wyników systemu. Modele OE, BJ i NOE, NBJ są przykładami konfiguracji symulacyjnej. W literaturze dotyczącej identyfikacji systemów, konfiguracja jednoetapowego przewidywania nazywana jest modelem szeregowo-równoległym, a konfiguracja symulacji nazywana jest modelem równoległym [121, 123]. Modele równoległe, same w sobie będące systemami dynamicznymi, są również nazywane modelami rekurencyjnymi, ponieważ ich opis matematyczny ma postać równań różniczkowych. W przeciwieństwie do modeli równoległych, modele szeregowo-równoległe są opisywane równaniami algebraicznymi. W kontekście sieci neuronowych modele te nazywane są modelami sprzężenia zwrotnego. W procesie identyfikacji parametry modelu są obliczane w taki sposób, aby zminimalizować wybraną funkcję kosztu zależną od błędu identyfikacji

$e(n)$ . Dwie powyższe konfiguracje modelu pociągają za sobą dwie różne definicje błędu identyfikacji. W przypadku modelu szeregowo-równoległego błąd identyfikacji nazywany jest błędem równania, a w przypadku modelu równoległego - błędem wyjścia.

### 1.1.4 Modele przestrzeni stanów

Rozszerzenie liniowego modelu przestrzeni stanów na przypadek nieliniowy daje następujący nieliniowy model przestrzeni stanów:

$$x(n+1) = h(x(n), u(n)) + u(n), \quad (1.45)$$

$$y(n) = g(x(n)) + \varepsilon(n), \quad (1.46)$$

gdzie  $x(n) \in \mathbb{R}^m$ ,  $u(n) \in \mathbb{R}^m$ ,  $y(n) \in \mathbb{R}^m$ ,  $\varepsilon(n) \in \mathbb{R}^m$ . Jeśli stan systemu  $x(n)$  jest dostępny do pomiaru, identyfikacja jest równoważna określeniu funkcji wektorowych  $h(\cdot)$  i  $g(\cdot)$ :

$$\hat{x}(n+1) = h(x(n), u(n)), \quad (1.47)$$

$$\hat{y}(n) = g(x(n)). \quad (1.48)$$

Model (1.47), (1.48) jest typu szeregowo-paralelnego. W praktyce, przynajmniej niektóre zmienne stanu są nieznane i muszą być oszacowane. Jeśli żadne zmienne stanu nie są mierzone, wymagana jest jednoczesna estymacja stanów układu i wyznaczenie funkcji  $h(\cdot)$  i  $g(\cdot)$ . Wysoka złożoność takiego zadania jest głównym powodem dominacji znacznie prostszych metod wejścia-wyjścia [123].

### 1.1.5 Modele nieliniowe składające się z podmodeli

Z wyjątkiem sztucznych "zjawisk" tworzonych przez równania matematyczne, modele zawsze różnią się od modelowanych zjawisk [131]. Modele matematyczne są po prostu równaniami wyprowadzonymi na podstawie pierwszych zasad i/lub danych eksperymentalnych i różnią się od leżących u ich podstaw zjawisk o charakterze niematematycznym. Ogólnie rzecz biorąc, modele można scharakteryzować za pomocą ich rozdzielczości czasowej i ziarnistości, gdzie ziarnistość określa poziom szczegółowości zawarty w modelu.

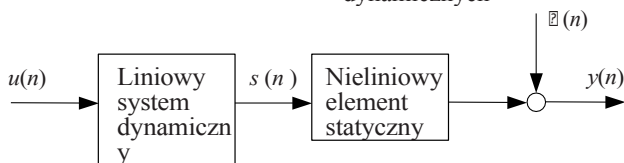
Do tej pory rozważaliśmy modele, które są klasyfikowane jako modele czarnej skrzynki [148]. Modele te opierają się wyłącznie na danych pomiarowych, tj. ich parametry i struktura są określane na podstawie eksperymentów. Zazwyczaj parametry modeli czarnej skrzynki nie mają interpretacji w kategoriach praw fizycznych, chemicznych, biologicznych, ekonomicznych i innych.

Inną klasą modeli są modele białoskrzynkowe, ponieważ są one całkowicie wyprowadzone z podstawowych praw. Nawet jeśli niektóre parametry są szacowane na podstawie danych, model jest również zaliczany do tej klasy. W przeciwieństwie do modeli czarnej skrzynki, parametry modeli białej skrzynki mają jasną interpretację [94].

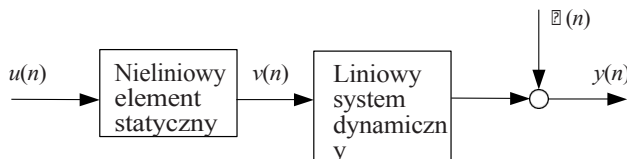
Modele szarej skrzynki łączą w sobie cechy modeli białej i czarnej skrzynki. Są one wyprowadzane zarówno na podstawie podstawowych praw, jak i danych pomiarowych. Na przykład, struktura systemu może być określona z wykorzystaniem wiedzy a priori o naturze systemu, a parametry systemu mogą być oszacowane na podstawie danych.

Nieliniowe modele o danej strukturze wewnętrznej składające się z podmodeli, nazywane również modelami blokowymi, należą do klasy modeli szarych pudełek. Modele Wienera i Hammersteina, pokazane na rysunkach 1.1 - 1.2, są dwoma modelami nieliniowymi.





Rys. 1.1. System SISO Wiener



Rys. 1.2. System SISO Hammerstein

well-known examples of such models, composed of sub-models [61, 108, 109]. Both of them contain a linear dynamic system and a nonlinear static element in a cascade. While in the Wiener system the nonlinear element follows the linear dynamic system, in the Hammerstein model, both of these sub-models are connected in reverse order. The Wiener model is given by

$$y(n) = f \frac{B(q)^{-1}}{A(q^{-1})} u(n) + \varepsilon(n), \quad (1.49)$$

gdzie  $f(\cdot)$  oznacza funkcję nieliniową opisującą element nieliniowy, a  $B(q^{-1})/A(q^{-1})$  jest funkcją przenoszenia impulsów liniowego układu dynamicznego.

Stosując tę samą notację, model Hammersteina można wyrazić jako

$$y(n) = \frac{B(q^{-1})}{A(q^{-1})} f u(n) + \varepsilon(n). \quad (1.50)$$

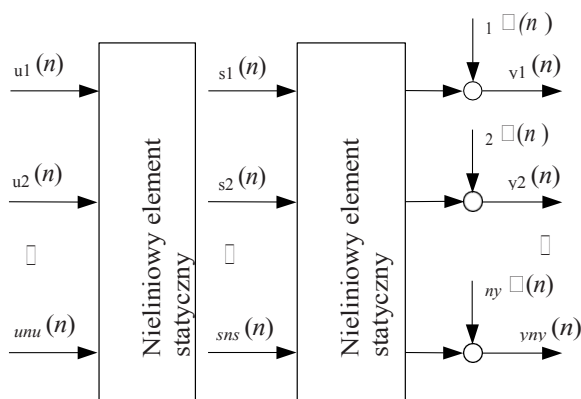
Model Wienera z wieloma wejściami i wieloma wyjściami (MIMO) można opisać następującym równaniem:

$$y(n) = f s(n) + \varepsilon(n), \quad (1.51)$$

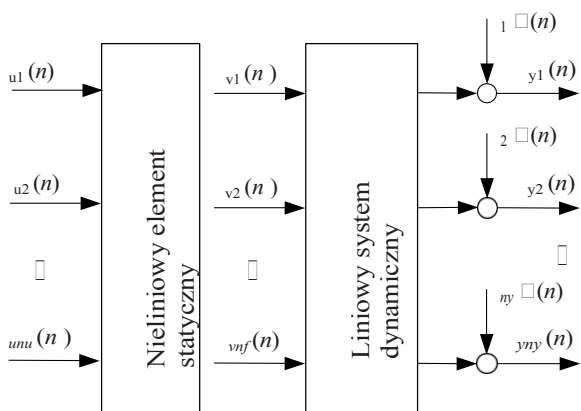
gdzie  $f(\cdot) : \mathbb{R}^{n_u} \rightarrow \mathbb{R}^{n_y}$  jest niezerową funkcją wektorową,  $y(n) \in \mathbb{R}^{n_y}$ ,  $\varepsilon(n) \in \mathbb{R}^{n_y}$ . Wyjście  $s(n)$ ,  $s(n) \in \mathbb{R}^{n_s}$ , liniowego systemu dynamicznego MIMO wynosi

$$s(n) = - \sum_{m=1}^{na} A^{(m)} s(n-m) + \sum_{m=1}^{nb} B^{(m)} u(n-m), \quad (1.52)$$

gdzie  $u(n) \in \mathbb{R}^{n_u}$ ,  $A^{(m)} \in \mathbb{R}^{n_s \times n_s}$ ,  $B^{(m)} \in \mathbb{R}^{n_s \times n_u}$ . W podobny sposób można uzyskać model MIMO Hammersteina, łącząc liniowy układ dynamiczny MIMO ze statycznym elementem nieliniowym MIMO - rys. 1.4. Dane wyjściowe modelu MIMO Hammersteina to



Rys. 1.3. System MIMO Wienera



Rys. 1.4. System MIMO Hammerstein

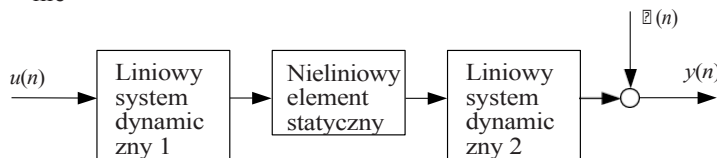
$$y(n) = - \sum_{m=1}^{na} A^{(m)} y(n-m) + \sum_{m=1}^{nb} B^{(m)} f(u(n-m)) + \varepsilon(n), \quad (1.53)$$

gdzie  $A \in \mathbb{R}^{ny \times ny}$ ,  $B \in \mathbb{R}^{ny \times nf}$ ,  $f(\cdot) : \mathbb{R}^{nu} \rightarrow \mathbb{R}^{nf}$  jest niezerową funkcją wektorową,  $u(n) \in \mathbb{R}^{nu}$ .

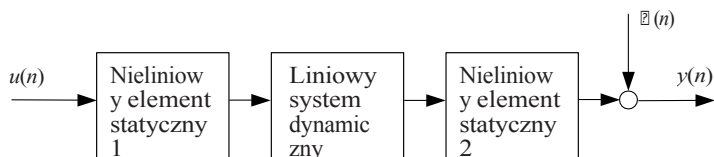
Zauważmy, że definicje (1.51), (1.52) i (1.53) opisują modele ze sprzężoną częścią statyczną i sprzężoną dynamiką. Możemy również rozważyć modele z niesprzężoną częścią statyczną i sprzężoną dynamiką lub sprzężoną częścią statyczną i niesprzężoną dynamiką jako szczególne przypadki tych ogólnych form.

Ogólny model Wienera rozważany przez Siebena [147] to model SISO (single-input single-output), w którym nieliniowy statyczny element MISO (multi-input single-output) podąża za liniowym dynamicznym elementem SIMO (single-input multiple-output).





Rys. 1.5. System SISO Wiener-Hammerstein



Rys. 1.6. System SISO Hammerstein-Wiener

system. Inna struktura, znana jako model Urysona [40], składa się z kilku równoległych modeli Hammersteina, przy czym każda ścieżka ma to samo wejście i kilka zsumowanych wyjść.

More complicated structures arise through the interconnection of three sub-models in a cascade. In this way, structures called Wiener-Hammerstein (Fig. 1.5) and Hammerstein-Wiener models (Fig. 1.6) can be obtained. The Wiener-Hammerstein structure is given by

$$y(n) = \frac{B_2(q^{-1})}{A_2(q^{-1})} f \left( \frac{B_1(q^{-1})}{A_1(q^{-1})} u(n) \right) + \varepsilon(n), \quad (1.54)$$

gdzie  $B_1(q^{-1})/A_1(q^{-1})$  i  $B_2(q^{-1})/A_2(q^{-1})$  są funkcjami przenoszenia impulsów odpowiednio pierwszego i drugiego liniowego układu dynamicznego. Identyfikacja układów Wienera-Hammersteina za pomocą metod korelacyjnych była badana przez Billingsa i Fakhouriego [17] oraz Huntera i Korenberga [71]. Rekursywna metoda identyfikacji dla modelu MISO Wiener-Hammerstein została zaproponowana przez Boutayeb i Darouach [21]. Bloemen *et al.* [19] rozważali zastosowanie modeli Hammersteina-Wienera do problemu sterowania predykcyjnego.

Bai [7] opracował dwuetapowy algorytm identyfikacji dla systemów Hammersteina-Wienera, w którym optymalne oszacowania parametrów zarówno elementów nieliniowych, jak i liniowego systemu dynamicznego są uzyskiwane przy użyciu algorytmu RLS, po którym następuje dekompozycja wartości osobliwych dwóch macierzy. Algorytm jest zbieżny przy braku szumu i zbieżny z prawdopodobieństwem 1 w obecności białego szumu.

Recently, the identification of Hammerstein-Wiener systems was also studied by Bai [9]. In the Hammerstein-Wiener model, two nonlinear blocks, described by the functions  $f_1(\cdot)$  and  $f_2(\cdot)$ , are separated by a linear dynamic system:

$$y(n) = f_2 \left( \frac{B(q^{-1})}{A(q^{-1})} f_1(u(n)) \right) + \varepsilon(n). \quad (1.55)$$

## 1.1.6 Modele Wienera w przestrzeni stanów

Opis przestrzeni stanów systemów MIMO Wienera z  $nu$  wejściami,  $ny$  wyjściami,  $nx$  zmiennymi stanu i  $ns$  zmiennymi wewnętrznymi ma postać [113, 158, 164]

$$\mathbf{x}(n+1) = \mathbf{A}\mathbf{x}(n) + \mathbf{B}\mathbf{u}(n), \quad (1.56)$$

$$\mathbf{s}(n) = \mathbf{C}\mathbf{x}(n) + \mathbf{D}\mathbf{u}(n), \quad (1.57)$$

$$\mathbf{y}(n) = \mathbf{f}(\mathbf{s}(n)) + \boldsymbol{\varepsilon}(n), \quad (1.58)$$

gdzie  $\mathbf{x}(n) \in \mathbb{R}^{nx}$ ,  $\mathbf{u}(n) \in \mathbb{R}^{nu}$ ,  $\mathbf{s}(n) \in \mathbb{R}^{ns}$ ,  $\mathbf{y}(n) \in \mathbb{R}^{ny}$ , oraz  $\mathbf{f}(\cdot) : \mathbb{R}^{ns} \rightarrow \mathbb{R}^{ny}$

jest niezerową funkcją wektorową,  $\boldsymbol{\varepsilon}(n) \in \mathbb{R}^{ny}$  jest procesem stochastycznym o zerowej średniej.

Równoległa reprezentacja modelu Wienera z przestrzenią stanów ma postać

$$\hat{\mathbf{x}}(n+1) = \mathbf{A}\hat{\mathbf{x}}(n) + \mathbf{B}\mathbf{u}(n), \quad (1.59)$$

$$\hat{\mathbf{s}}(n) = \mathbf{C}\hat{\mathbf{x}}(n) + \mathbf{D}\mathbf{u}(n), \quad (1.60)$$

$$\hat{\mathbf{y}}(n) = \mathbf{f}(\hat{\mathbf{s}}(n)). \quad (1.61)$$

Zastąpienie stanu modelu  $\hat{\mathbf{x}}(n)$  w (1.59) i (1.60) stanem systemu  $\mathbf{x}(n)$  skutkuje szeregowo-równoległą reprezentacją modelu Wienera przestrzeni stanów:

$$\hat{\mathbf{x}}(n+1) = \mathbf{A}\mathbf{x}(n) + \mathbf{B}\mathbf{u}(n), \quad (1.62)$$

$$\hat{\mathbf{s}}(n) = \mathbf{C}\mathbf{x}(n) + \mathbf{D}\mathbf{u}(n), \quad (1.63)$$

$$\hat{\mathbf{y}}(n) = \mathbf{f}(\hat{\mathbf{s}}(n)). \quad (1.64)$$

## 1.1.7 Modele Hammersteina w przestrzeni stanów

Model MIMO Hammersteina w przestrzeni stanów można opisać następującymi równaniami [157]:

$$\mathbf{x}(n+1) = \mathbf{A}\mathbf{x}(n) + \mathbf{B}\mathbf{v}(n), \quad (1.65)$$

$$\mathbf{y}(n) = \mathbf{C}\mathbf{x}(n) + \mathbf{D}\mathbf{v}(n) + \boldsymbol{\varepsilon}(n), \quad (1.66)$$

$$\mathbf{v}(n) = \mathbf{f}(\mathbf{u}(n)), \quad (1.67)$$

gdzie  $\mathbf{x}(n) \in \mathbb{R}^{nx}$ ,  $\mathbf{u}(n) \in \mathbb{R}^{nu}$ ,  $\mathbf{y}(n) \in \mathbb{R}^{ny}$ ,  $\boldsymbol{\varepsilon}(n) \in \mathbb{R}^{ny}$  jest procesem stochastycznym o zerowej średniej,  $\mathbf{v}(n) \in \mathbb{R}^{nv}$ , a  $\mathbf{f}(\cdot) : \mathbb{R}^{nu} \rightarrow \mathbb{R}^{nv}$  jest niezerową funkcją wektorową. Model MIMO Hammersteina w równoległej reprezentacji przestrzeni stanów ma postać

$$\hat{\mathbf{x}}(n+1) = \mathbf{A}\hat{\mathbf{x}}(n) + \mathbf{B}\hat{\mathbf{v}}(n), \quad (1.68)$$

$$\hat{\mathbf{y}}(n) = \mathbf{C}\hat{\mathbf{x}}(n) + \mathbf{D}\hat{\mathbf{v}}(n), \quad (1.69)$$

$$\hat{\mathbf{v}}(n) = \mathbf{f}(\mathbf{u}(n)). \quad (1.70)$$

Zastępując stan modelu  $\hat{x}(n)$  w (1.68) i (1.69) stanem systemu  $x(n)$ , otrzymujemy szeregowo-równoległą reprezentację modelu Hammersteina z przestrzenią stanów:

$$\hat{x}(n+1) = Ax(n) + Bv^{\wedge}(n), \quad (1.71)$$

$$\hat{y}(n) = Cx(n) + Dv^{\wedge}(n), \quad (1.72)$$

$$\hat{v}(n) = f(u(n)) \quad (1.73)$$

## 1.2 Perceptron wielowarstwowy

Wielowarstwowe sieci neuronowe typu feedforward, zwane również perceptronami wielowarstwowymi, są najbardziej znanymi i najczęściej stosowanymi sieciami neuronowymi [60, 68, 69, 127, 144, 169]. W perceptronie wielowarstwowym (MLP) neurony są uporządkowane w jedną lub więcej warstw ukrytych i połączone z warstwą wyjściową. Ten typ sieci neuronowej jest intensywnie wykorzystywany w rozdziałach 2 i 3 do modelowania zarówno elementu nieliniowego, jak i jego odwrotności.

### 1.2.1 Architektura MLP

$i$ -ty sygnał wyjściowy  $x_i(n)$  pierwszej warstwy ukrytej (rys. 1.7) to

$$x_i(n) = \phi \left( \sum_{j=1}^{nu} w_{ij}^{(1)} u_j^{(1)}(n) + w_{i0}^{(1)} \right), \quad (1.74)$$

gdzie  $w_{ij}^{(1)}$  jest  $j$ -tą wagą  $i$ -tego neuronu,  $w_{i0}^{(1)}$  jest biasem  $i$ -tego neuronu,

$\phi(\cdot)$  jest funkcją aktywacji neuronów warstwy ukrytej,  $nu$  jest liczbą wejść, a  $u_j(n)$  jest  $j$ -tym wejściem. Często wybieranymi funkcjami aktywacji są funkcje sigmoidalne, takie jak funkcja logistyczna

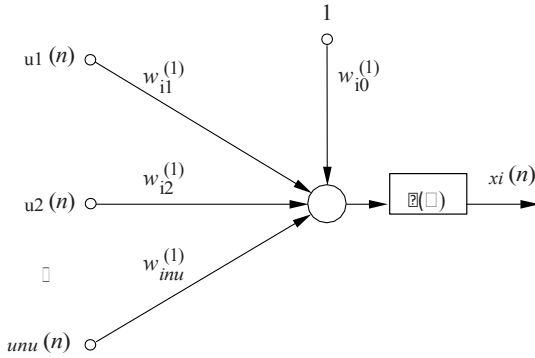
$$\phi(x) = \frac{1}{1 + \exp(-x)} \quad (1.75)$$

i hiperboliczna funkcja styyczna

$$\phi(x) = \tanh(x) = \frac{1 - \exp(-2x)}{1 + \exp(-2x)}. \quad (1.76)$$

Wyjścia pierwszej warstwy ukrytej mogą być połączone z kolejnymi warstwami ukrytymi i ostatecznie z warstwą wyjściową. W powszechnie stosowanej sieci neuronowej MLP z jedną warstwą ukrytą, dane wyjściowe  $x_i(n)$ ,  $i = 1, \dots, M$ , gdzie  $M$  jest liczbą neuronów warstwy ukrytej, są przekształcane przez warstwę wyjściową w dane wyjściowe  $y_k(n)$ :

$$y_k(n) = \phi \left( \sum_{i=1}^M w_{ki}^{(2)} x_i^{(2)}(n) + w_{k0}^{(2)} \right), \quad (1.77)$$



Rys. 1.7. Neuron i-tej warstwy ukrytej

gdzie  $w_{ki}^{(2)}$  jest  $i$ -tą wagą  $k$ -tego neuronu wyjściowego,  $w_{k0}^{(2)}$  jest biasem  $k$ -tego neuronu wyjściowego.

$k$ -tego neuronu wyjściowego, a  $\varphi(\cdot)$  jest funkcją aktywacji neuronów warstwy wyjściowej. Chociaż  $\varphi(\cdot)$  może być funkcją nieliniową, typowym wyborem jest liniowa funkcja aktywacji  $\varphi(x) = x$ .

### 1.2.2 Algorytmy uczenia się

Nieliniowa optymalizacja wag sieci neuronowej jest najczęściej stosowaną techniką uczenia MLP. Korzystając z metod uczenia opartych na gradiencie, funkcja kosztu  $J$ , zazwyczaj suma kwadratów błędów między wyjściem systemu a wyjściem modelu sieci neuronowej, jest minimalizowana. W rezultacie wagi sieci neuronowej są dostosowywane wzdłuż ujemnego gradientu funkcji kosztu. Algorytm uczenia wstecznego (BP) jest implementacją metody optymalizacji gradientowej dla aktualizacji wag. Algorytm uczenia wstecznego wykorzystuje algorytm wstecznej propagacji jako technikę obliczania gradientu MLP względem jego wag [68, 123, 132]. Pomimo swojej prostoty obliczeniowej, trenowanie MLP za pomocą algorytmu BP może powodować kilka problemów, takich jak bardzo powolna zbieżność, oscylacje, rozbieżność i efekt "zygzakowania". W celu obejścia tego problemu opracowano wiele ulepszeń, rozszerzeń i modyfikacji podstawowego algorytmu uczenia BP [60, 68, 123].

Powodem powolnej zbieżności algorytmu uczenia BP jest to, że działa on w oparciu o liniowe przybliżenie funkcji kosztu. Aby osiągnąć znacznie wyższy współczynnik zbieżności, należy zastosować przybliżenia funkcji kosztu wyższego rzędu. Przykładami takich technik uczenia są metoda Levenberga-Marquarda, metody quasi-Newtona, metody gradientu sprzężonego [68, 123] oraz algorytmy uczenia RLS [144].

Aby wydobyć informacje z danych treningowych i zwiększyć skuteczność uczenia się, zwykle konieczne jest wstępne przetwarzanie danych, takie jak filtrowanie, usuwanie nadmiaru i usuwanie wartości odstających.

Ponadto skalowanie danych jest niezbędne, aby algorytmy uczenia się były niezawodne i skracaly czas uczenia. Zalecana technika skalowania opiera się na usunięciu średniej i skalowaniu sygnałów do tej samej wariancji [127]. Alternatywnie, średnia może być usunięta z sygnałów, a sygnały o zerowej średniej skalowane w odniesieniu do ich maksymalnych wartości bezwzględnych, aby uzyskać wartości w określonym przedziale, np.

Ogólnie rzecz biorąc, aby zminimalizować czas uczenia, stosowanie sygnałów wejściowych o niezerowej średniej

należy unikać. Wynika to z faktu, że czas uczenia dla algorytmu stromego zejścia jest wrażliwy na zmiany liczby warunkowej  $\lambda / \lambda_{maxmin}$ , gdzie  $\lambda_{max}$  jest największą wartością własną Hessjana funkcji kosztu, a  $\lambda_{min}$  jest jej najmniejszą niezerową wartością własną. Wyniki eksperymentalne pokazują, że dla sygnałów wejściowych o niezerowej średniej liczba warunków  $\lambda / \lambda_{maxmin}$  jest większa niż dla sygnałów wejściowych o zerowej średniej [68]. Należy zauważyć, że wybór asymetrycznej funkcji aktywacji, np. funkcji logistycznej, wprowadza systematyczne odchylenie dla neuronów warstwy ukrytej. Ma to podobny wpływ na liczbę warunków  $\lambda / \lambda_{maxmin}$  jak niezerowe wartości średnie. Dlatego, aby zwiększyć szybkość zbieżności algorytmów uczenia opartych na gradiencie, zaleca się wybór antysymetrycznych funkcji aktywacji, takich jak tangens hiperboliczny.

### 1.2.3 Optymalizacja architektury modelu

W praktyce modele MLP rzeczywistych procesów mają dość duże rozmiary. Powszechnie wiadomo, że modele nie powinny być zbyt złożone, ponieważ uczyłyby się szumu, a tym samym źle uogólniałyby się na nowe dane. Z drugiej strony, nie powinny być zbyt proste, ponieważ nie byłyby w stanie uchwycić zachowania procesu. W przypadku MLP z jedną warstwą ukrytą, problem optymalizacji architektury sprowadza się do wyboru liczby węzłów warstwy ukrytej i eliminacji nieistotnych wag.

Ogólny błąd modelu składa się z dwóch komponentów - błędu systematycznego, który wyraża błąd systematyczny spowodowany ograniczoną elastycznością modelu oraz błędu wariancji, będącego błędem stochastycznym wynikającym z ograniczonej dokładności szacunków parametrów. Oba te składniki błędu modelu są w konflikcie, znanym jako dylemat odchylenia/wariancji, ponieważ błąd odchylenia maleje, a błąd wariancji rośnie wraz ze wzrostem złożoności modelu. Dlatego konieczne jest znalezienie kompromisu - tak zwanego kompromisu błędu systematycznego i błędu wariancji. Aby to osiągnąć, konieczna jest optymalizacja architektury modelu.

Istnieją dwie grupy metod wykorzystywanych do optymalizacji architektury sieci neuronowych, znane jako rozrastanie się sieci (ang. network growing) i przycinanie sieci (ang. network pruning). W metodach wzrostu sieci nowe węzły lub warstwy są dodawane począwszy od sieci o małym rozmiarze, aż powiększona struktura spełni założone wymagania. Dobrze znanym przykładem metody rosnącej jest algorytm kaskadowej korelacji [36]. W metodach przycinania początkowa struktura jest duża, a następnie przycinamy ją, osłabiając lub eliminując niektóre wybrane wagi.

Idea przycinania opiera się na założeniu, że w pełni połączonych sieci MLP przechowywana jest duża ilość nadmiarowych informacji. Przycinanie sieci



jest zwykle osiągane za pomocą dwóch podejść - jednego opartego na ponownej gularyzacji złożoności i drugiego opartego na usuwaniu niektórych wag przy użyciu informacji o pochodnych drugiego rzędu funkcji kosztu.

W metodach regularyzacji złożoności termin kary złożoności jest dodawany do funkcji kosztu. Standardową funkcją kosztu w uczeniu metodą propagacji wstecznej jest błąd średniokwadratowy. W zależności od formy terminu kary złożoności można zdefiniować różne techniki regularyzacji, takie jak usuwanie wagi, eliminacja wagi, przybliżone wygładzanie, podejście kary Chauvina [60, 123]. W rzeczywistości metody regularyzacji nie zmieniają struktury modelu, ale zmniejszają jego elastyczność, utrzymując niektóre wagi na ich początkowych wartościach lub ograniczając ich wartości. W ten sposób można osiągnąć zmniejszenie błędu wariancji za cenę błędu odchylenia.

Optymalne uszkodzenie mózgu (OBD) [107] i optymalny chirurg mózgu (OBS) [67] są najbardziej znanymi i stosowanymi metodami opartymi na wykorzystaniu informacji o pochodnych drugiego rzędu funkcji kosztu. Obie są używane do zmniejszania rozmiaru sieci poprzez selektywne usuwanie wag.

Obie wykorzystują rozszerzenie Taylora drugiego rzędu funkcji kosztu o punkt operacyjny -  $n$ -wymiarowy wektor wagowy  $w^* = [w^* \dots w^*]^T$   $nw$

dla których funkcja kosztu ma lokalne minimum. Ich celem jest znalezienie zestawu wag, których usunięcie powoduje najmniejszą zmianę funkcji kosztu. Aby osiągnąć rozsądnie niską złożoność obliczeniową, tylko diagonalne wyrazy rozwinięcia Taylora drugiego rzędu są uwzględniane w definicji parametrów w metodzie OBD. Odpowiada to założeniu, że macierz Hessjana jest macierzą diagonalną. Metoda OBD jest procedurą iteracyjną o następującej postaci:

1. Wytrenuj MLP do minimalnego błędu średniokwadratowego funkcji kosztu.
2. Obliczyć pochodne diagonalne drugiego rzędu  $h_{ii}$ ,  $i = 1, \dots, nw$ , funkcji kosztu.
3. Obliczenie salw dla wag:  $S_i = h_{ii}(w)^{*2} / 2$ .  $i$
4. Usuń niektóre wagi, które mają małe zasolenie.
5. Powrót do kroku 1.

W metodzie OBS nie przyjmuje się takiego założenia dotyczącego macierzy Hessjana, a OBD można uznać za szczególny przypadek OBS.

### 1.3 Identyfikacja systemów Wienera

Zaproponowano wiele różnych podejść do identyfikacji systemu Wienera opartych na analizie korelacji, optymalizacji liniowej, regresji nieparametrycznej, optymalizacji nieliniowej z różnymi modelami nieliniowymi, takimi jak wielomiany, sieci neuronowe, falki, funkcje ortogonalne i modele zbiorów rozmytych.

Metody korelacyjne. Billings i Fakhouri wykorzystali analizę korelacji do identyfikacji systemów zorientowanych blokowo w oparciu o teorię

procesów rozłącznych [14]. Gdy wejście jest białym sygnałem gaussowskim, możliwe jest oddzielenie identyfikacji liniowego układu dynamicznego od identyfikacji elementu nieliniowego [15, 17]. W przypadku systemów Wienera funkcja korelacji pierwszego rzędu  $R_{uy}(k)$  wejścia systemu  $u(n)$  i wyjścia systemu  $y(n)$  jest wprost proporcjonalna do odpowiedzi impulsowej  $h(k)$  systemu liniowego, a funkcja korelacji drugiego rzędu  $R_{u^2y}(k)$  jest wprost proporcjonalna do odpowiedzi impulsowej  $h(k)$  systemu liniowego.

do kwadratu  $h(k)$ . Zatem, jeśli  $R^2(k)$  i  $R_{u^2y}(k)$  są równe, z wyjątkiem dla stałej proporcjonalności, system ma strukturę typu Wienera. Podejście korelacyjne do identyfikacji systemów dynamicznych zależnych od kierunku za pomocą modeli Wienera zostało zastosowane przez Barkera i in. [12]. Rozważali oni układy Wienera zawierające liniową część dynamiczną z różnymi funkcjami transferu dla rosnącej i malejącej mocy wyjściowej układu. System Wienera był wzbudzany pseudolosowymi lub odwrotnymi pseudolosowymi sygnałami binarnymi o maksymalnej długości. Określenie parametrów modelu Wienera przeprowadzono poprzez dopasowanie funkcji korelacji systemu i modelu, wyjść i dyskretnych transformat Fouriera wyjść.

Metody optymalizacji liniowej. W metodach optymalizacji liniowej przyjmuje się, że model można sparametryzować za pomocą skończonego zestawu parametrów. Element nieliniowy jest zwykle modelowany za pomocą wielomianu wraz z modelem funkcji przenoszenia impulsów liniowego układu dynamicznego. Sparametryzowany w ten sposób model Wienera jest nieliniowy w parametrach, a estymacja parametrów staje się nieliniowym problemem optymalizacyjnym. Należy zauważyć, że inwersja modelu Wienera jest modelem Hammersteina, którego część liniowa jest opisana przez odwrotną funkcję transferu, a element nieliniowy jest opisany przez odwrotną funkcję nieliniową. Warunkiem koniecznym takiego przekształcenia jest odwracalność funkcji  $f(\cdot)$  opisującej nieliniową część modelu Wienera. Aby uzyskać asymptotycznie stabilny model odwrotny, model Wienera powinien mieć minimalną fazę. Odwrotny model Wienera jest nadal nieliniowy w parametrach, ale jest znacznie wygodniejszy do estymacji parametrów, ponieważ można go przekształcić do postaci liniowej w parametrach MISO za pomocą metody zaproponowanej przez Changa i Luusa [27] dla systemów Hammersteina. Parametry przekształconego modelu można obliczyć metodą najmniejszych kwadratów. Zakładając znajomość  $f(\cdot)$ , identyfikacja odwrotnego systemu Wienera była rozważana przez Pearsona i Pottmana [138]. Aby przezwyciężyć praktyczną trudność związaną z faktem, że podejście do identyfikacji systemu odwrotnego penalizuje błędy predykcji sygnału wejściowego  $u(n)$  zamiast sygnału wyjściowego  $y(n)$ , zastosowali oni ważoną metodę najmniejszych kwadratów z parametrami ważącymi, które ważą względną wagę błędu.

Wielomianowy model odwrotny elementu nieliniowego i model liniowego układu dynamicznego z filtrem próbkującym częstotliwość

$$\hat{s}(n) = \sum_{m=0}^{N-1} G(j\omega_m) F_m(n), \quad (1.78)$$

gdzie  $\hat{s}(n)$  jest wyjściem liniowego modelu dynamicznego,  $G(jw_m)$ ,  $m = 0, \dots, n_f - 1$ , jest dyskretną odpowiedzią częstotliwościową układu liniowego przy  $w_m = 2\pi m/n_f$ ,  $F_m(n)$  jest wyjściem  $m$ -tego filtra próbkowania częstotliwości zdefiniowanego jako

$$F_m(n) = \frac{1 - q^{-n_f}}{n_f (1 - e^{j(2\pi m/n_f)q-1})} u(n), \quad (1.79)$$

gdzie  $u(n)$  jest wejściem systemu, a  $q^{-1}$  jest operatorem przesunięcia wstecznego, został użyty przez Kalafatisa *i in.* [95]. Wyjście modelu  $\hat{y}(n)$  można następnie wyrazić w postaci liniowej w parametrach:

$$\hat{y}(n) = \sum_{m=0}^{n_f-1} G(jw_m) F_m(n) - \sum_{k=2}^r \gamma \gamma_k^k(n), \quad (1.80)$$

gdzie  $y(n)$  jest wyjściem systemu, a  $\gamma_k, k = 2, \dots, r$ , są parametrami wielomianowego modelu odwrotnego elementu nieliniowego. Parametry  $G(jw_m)$  i  $\gamma_k$  można obliczyć metodą najmniejszych kwadratów. Niestety, takie podejście prowadzi do niespójnych oszacowań parametrów dla systemów Wienera z addytywnymi zakłóceniami wyjściowymi, ponieważ wektor regresji jest skorelowany z zakłóceniami. Aby przezwyciężyć ten problem, zaproponowano algorytm iteracyjny [96], który składa się z następujących trzech kroków: Po pierwsze, oszacowania parametrów są obliczane przy użyciu metody najmniejszych kwadratów. Następnie, wykorzystując uzyskane estymaty, obliczane są przewidywane wyjścia systemu. Wreszcie, aby obliczyć skorygowane oszacowania parametrów, przewidywane wyjścia systemu są wykorzystywane w kolejnym kroku estymacji przy użyciu metody najmniejszych kwadratów. Powyższa procedura estymacji jest powtarzana do momentu, aż oszacowania parametrów zbiegną się do stałych wartości. Schemat rekurencyjnej metody najmniejszych kwadratów jest stosowany w metodzie identyfikacji opartej na ortonormalnych funkcjach bazowych zaproponowanej przez Marciaka *i in.* Używają oni nieodwróconego modelu układu liniowego złożonego z filtrów Laguerre'a i odwrotnego modelu wielomianowego elementu nieliniowego.

W podejściu częstotliwościowym do identyfikacji systemów Wienera Bai [10], estymata fazy wyjścia liniowego systemu dynamicznego jest określana na podstawie dyskretnej transformaty Fouriera (DFT) przefiltrowanego wyjścia systemu. Mając oszacowanie fazy, strukturę nieliniowości można określić na podstawie wykresu wyjścia systemu w stosunku do oszacowanego wyjścia liniowego systemu dynamicznego i aproksymować wielomianem.

Identyfikację układów Wienera opartą na zmodyfikowanym szeregowo-równoległym modelu Wienera, zdefiniowanym przez nieodwrócony model przenoszenia impulsów elementu liniowego i odwrotny model wielomianowy elementu nieliniowego, można przeprowadzić metodą zaproponowaną przez Janczaka [80, 83]. Zmodyfikowany model szeregowo-równoległy jest liniowy w parametrach, a jego parametry są obliczane metodą najmniejszych kwadratów. Metoda ta wymaga, aby nieliniowa funkcja  $f(\cdot)$  była odwracalna, a liniowy człon modelu wielomianowego był niezerowy. Jednak w przypadku addytywnego szumu wyjściowego bezpośrednie zastosowanie tego podejścia skutkuje niespójnymi oszacowaniami parametrów. Jako remedium na taką sytuację stosuje się połączoną procedurę estymacji metodą najmniejszych kwadratów i zmiennych instrumentalnych.

zaproponowana. Szczegółowy opis tej metody, wraz z jej rozszerzeniem na identyfikację systemów Wienera bez liniowego członu charakterystyki nieliniowej, podano w rozdziale 4.

Estymacja parametrów systemów Wienera składających się z modelu skończonej odpowiedzi impulsowej (FIR) systemu liniowego i odwrotnego modelu wielomianowego elementu nieliniowego była rozważana przez Mzyka [119]. Aby uzyskać spójne oszacowania parametrów, zastosowano metodę zmiennych instrumentalnych ze zmiennymi instrumentalnymi zdefiniowanymi jako suma zasilanych sygnałów wejściowych i niektórych stałych strojenia wybranych przez użytkownika.

Identyfikacja systemów MIMO Wienera z wykorzystaniem funkcji bazowych do reprezentacji zarówno liniowego systemu dynamicznego, jak i elementu nieliniowego została zaproponowana przez Go'meza i Baeyensa [44]. Zakłada się, że element nieliniowy jest odwracalny i może być opisany nieliniowymi funkcjami bazowymi. Liniowy system dynamiczny MIMO jest reprezentowany przez racjonalne ortonormalne funkcje bazowe ze stałymi biegunami (OBFP). Szczególnymi przypadkami OBFP są bazy FIR, Laguerre i Kautz. Zgodnie z powyższymi założeniami, model MIMO Wienera można przekształcić w postać liniową w parametrach, a estymaty parametrów przekształconego modelu, które minimalizują kwadratową funkcję kosztu błędów predykcji, są obliczane metodą najmniejszych kwadratów. Aby obliczyć parametry macierzowe elementu nieliniowego na podstawie oszacowań parametrów przekształconego modelu, stosuje się technikę dekompozycji wartości pojedynczej.

Metody regresji nieparametrycznej. Metody regresji parametrycznej opierają się na restrykcyjnym założeniu dotyczącym klasy funkcji nieliniowych. Podejście regresji nieparametrycznej z jądrem, które znacznie rozszerza klasę nieliniowości zidentyfikowanych w systemach Wienera, zostało wprowadzone przez Greblickiego [46] i badane dalej w [48]. Następnie idea regresji nieparametrycznej została rozwinięta poprzez zastosowanie szeregów ortogonalnych do odzyskiwania odwrotności charakterystyki nieliniowej [47]. Dla algorytmów regresji nieparametrycznej wykorzystujących funkcje ortogonalne trygonometryczne, Legendre'a i Hermite'a wykazano spójność punktową i podano współczynniki zbieżności. Greblicki proponował również i przeanalizował rekurencyjne algorytmy identyfikacji oparte na regresji jądra zarówno dla dyskretnych [51], jak i ciągłych systemów Wienera [50, 52].

Metody optymalizacji nieliniowej. Identyfikacja systemów Wienera przy użyciu metod błędu predykcji została omówiona w [85, 88, 128, 165, 166]. Wigren

[165] przeanalizował rekurencyjne algorytmy identyfikacji Gaussa-Newtona i gradientu stochastycznego przy założeniu, że nieliniowość systemu jest znana a priori. Określił on warunki lokalnej i globalnej zbieżności oszacowań parametrów do rzeczywistych parametrów systemu przy skorelowanych zakłóceniach pomiarowych. Inna rekurencyjna metoda błędu predykcji, zaproponowana przez Wigrena [166], szacuje parametry modelu funkcji przenoszenia impulsu systemu liniowego i fragmentarycznego modelu

liniowego elementu nieliniowego. Dzięki technice zlinearyzowanego równania różniczkowego, lokalna zbieżność estymat do parametrów systemu nieliniowego nie jest już istotna.

metrów. Wykazano również, że sygnał wejściowy powinien być taki, aby istniała energia w całym zakresie aproksymacji liniowej.

Vooros [162] zastosował model parametryczny do opisu układów Wienera ze szczególnym rodzajem nieciągłego elementu nieliniowego - funkcją fragmentarycznie liniową z obciążeniem wstępnym i strefą martwą. Czyste obciążenie wstępne, strefa martwa i dwusegmentowa asymetryczna nieliniowość są szczególnymi przypadkami ogólnej nieciągłej nieliniowości. Zaproponował on metodę identyfikacji, która wykorzystuje opis elementu nieliniowego oparty na zasadzie separacji kluczy. Takie sformułowanie problemu umożliwia przekształcenie modelu elementu nieliniowego do postaci pseudolinearnej w parametrach. Estymacja parametrów jest przeprowadzana iteracyjnie, ponieważ model zawiera trzy zmienne wewnętrzne, które są niemierzalne. Procedura iteracyjna opiera się na wykorzystaniu oszacowań parametrów z poprzedniego kroku do oszacowania tych niemierzalnych zmiennych wewnętrznych.

Inne podejście do identyfikacji układów Wienera z założonymi formami nieliniowości typu twardego sparametryzowanymi przez pojedynczy nieznany parametr  $a$  zostało zaproponowane przez Bai [8]. Przykładami nieliniowości tego typu są nasycenie, obciążenie wstępne, przekątnik, strefa martwa, histereza-przekątnik i histereza. Aby znaleźć nieznany parametr  $a$ , stosuje się metodę rozdzielnych najmniejszych kwadratów, w której problem identyfikacji jest przekształcany w jednowymiarowy problem minimalizacji. Ponieważ funkcja kosztu jest jednowymiarowa, do znalezienia jej minimum można zastosować metody wyszukiwania globalnego. Alternatywnie, możliwe jest również znalezienie minimum bezpośrednio z wykresu funkcji kosztu względem  $a$ . Po znalezieniu optymalnego oszacowania  $a$ , parametry liniowego układu dynamicznego można oszacować metodą najmniejszych kwadratów. Dla tego podejścia podano warunki, w których można osiągnąć silną spójność oszacowań parametrów. Chociaż separowalną metodę najmniejszych kwadratów można łatwo rozszerzyć na przypadek dwuwymiarowy, jej rozszerzenie na przypadek niecałkowitości sparametryzowanych większą liczbą parametrów jest bardziej skomplikowane. Iteracyjny schemat identyfikacji układów Wienera metodą błędu przewidywania został zaproponowany przez Norquay *et al.* [128]. W tym podejściu element nieliniowy jest modelowany przez wielomian. Do obliczenia przybliżonego Hessjana stosuje się metodę Levenberga-Marquardta wraz z obliczeniem gradientu poprzez symulację modeli wrażliwości. W [85, 88] zastosowano rekurencyjną wersję tej metody.

Algorytm regresji pseudoliniowej, opisany przez Janczaka [86, 88], można uzyskać ze schematu błędu predykcji, jeśli zignoruje się zależność zarówno de- layowanych, jak i zasilanych sygnałów wyjściowych modelu liniowego od parametrów modelu. W ten sposób model jest traktowany jako liniowy w parametrach. Takie uproszczenie zmniejsza złożoność obliczeniową algorytmu za cenę dokładności aproksymacji gradientowej.

Metoda oparta na sieci neuronowej do identyfikacji systemów Wienera została opracowana przez Al-Duwaish [3]. W metodzie tej liniowy układ dynamiczny i element nieliniowy są identyfikowane oddzielnie. Po pierwsze, parametry systemu liniowego, opisanego liniowym równaniem różniczkowym, są szacowane za pomocą

rekurencyjny algorytm najmniejszych kwadratów (RLS) oparty na odpowiedzi na mały sygnał wejściowy, który zapewnia liniową perturbację elementu nieliniowego. Następnie, po zidentyfikowaniu układu liniowego, algorytm uczenia wstecznego jest stosowany do trenowania wielowarstwowego modelu sieci neuronowej elementu nieliniowego. Ten krok jest wykonywany z kolejną odpowiedzią na zwiększony sygnał wejściowy, który nieliniowo zaburza element nieliniowy.

Identyfikacja systemu MISO Wienera była badana przez Ikonena i Najima [72]. W oparciu o dane wejściowo-wyjściowe z systemu pilotowego pompawódź, model Wienera sieci neuronowej został wytrenowany metodą Levenberga-Marquardta. Visala *et al.* [159] wykorzystali model MIMO Wienera do modelowania procesu separacji chromatograficznej. Liniowa część dynamiczna ich modelu składa się z filtrów Laguerre'a, podczas gdy sieć neuronowa MIMO feedforward jest wykorzystywana jako statyczne mapowanie nieliniowe. Zakłada się, że liniowy model dynamiczny MIMO jest oddzielony. Szacowanie parametrów sprowadza się do trenowania nieliniowej części modelu za pomocą metody Levenberga-Marquardta, ponieważ zakłada się, że dynamika liniowa jest znana. Oznacza to, że odpowiednie wartości parametrów Laguerre'a są wybierane na podstawie informacji a priori.

Identyfikacja nieliniowego systemu dynamicznego przy użyciu modelu składającego się z systemu dynamicznego SIMO, a następnie elementu nieliniowego MISO, została zbadana przez Alatarisa *i in* [1]. Ich metodologia wykorzystuje perceptron wielowarstwowy z pojedynczą warstwą ukrytą i wielomianowymi funkcjami aktywacji jako model nieliniowości statycznej. Wejścia do tego modelu są wyjściami banku filtrów Laguerre'a używanego jako model liniowego systemu dynamicznego. Parametry modelu sieci neuronowej są dostosowywane za pomocą metody zejścia gradientowego. Wybór rzeczywistego bieguna, będącego jedynym stopniem swobody filtrów Laguerre'a, opiera się na zasadzie prób i błędów. Zastosowanie wielomianowej funkcji aktywacji pozwala na łatwe przejście między siecią neuronową a modelami szeregowymi Volterra.

Model sieci neuronowej zarówno liniowego systemu dynamicznego, jak i elementu nieliniowego jest wykorzystywany do opisu systemów SISO Wienera w jednoetapowej procedurze sekwencyjnej zaproponowanej przez Janczaka [89]. Więcej szczegółów na temat tego podejścia i jego rozszerzenia na przypadek MIMO można znaleźć w rozdziale 2.

Identyfikacja systemów Wienera przy użyciu metody genetycznej, ujawniającej globalną właściwość optymalizacji, była badana przez Al-Duwaish [5]. W tym podejściu obliczane są parametry modelu Wienera składającego się z modelu funkcji przenoszenia impulsu części liniowej i modelu wielomianowego elementu nieliniowego. Model funkcji przenoszenia impulsów jest zdefiniowany w postaci biegunowo-zerowej. Z funkcją fitness zdefiniowaną jako suma kwadratów błędów między wyjściami systemu i modelu oraz wykonując operacje genetyczne krzyżowania, mutacji i selekcji, generowane są nowe generacje rozwiązań i oceniane aż do osiągnięcia określonej dokładności aproksymacji. Równoległy model sieci neuronowej Wienera został również wytrenowany przy użyciu rekurencyjnego programowania ewolucyjnego z zależną od czasu szybkością uczenia przez Janczaka i Mrugałskiego [93].







## 1.4 Identyfikacja systemów Hammersteina

Podobnie jak w przypadku systemów Wienera, omówionych w sekcji 1.3, dokonamy krótkiego przeglądu różnych metod identyfikacji systemów Hammersteina.

Metody korelacyjne. W przypadku białych gaussowskich danych wejściowych obliczenie funkcji korelacji krzyżowej umożliwia oddzielenie identyfikacji układów Hammersteina i oddzielną identyfikację liniowego układu dynamicznego i elementu nieliniowego [14, 15, 16, 17]. Po pierwsze, odpowiedź impulsowa  $h(k)$  liniowego układu dynamicznego jest szacowana przy użyciu techniki korelacji. Funkcja korelacji pierwszego rzędu  $R_{uy}(k)$  jest wprost proporcjonalna do odpowiedzi impulsowej,  $R_{uy}(k) = \alpha h(k)$ . W Zazie potrzeby parametry funkcji przenoszenia impulsu  $[\alpha h(k)] = B(q^{-1})/A(q^{-1})$  można łatwo obliczyć na podstawie odpowiedzi impulsowej. Mając dostępne oszacowanie  $\alpha h(k)$ , parametry modelu wielomianowego elementu nieliniowego można obliczyć za pomocą algorytmu najmniejszych kwadratów [16].

W przypadku systemów Hammersteina funkcja korelacji drugiego rzędu  $R_{u_2y}(k)$  jest wprost proporcjonalna do odpowiedzi impulsowej elementu liniowego i zapewnia wygodny test struktury systemu. Jeśli funkcje korelacji pierwszego i drugiego rzędu są równe, z wyjątkiem stałej proporcjonalności, system musi mieć strukturę modelu Hammersteina.

Metody optymalizacji liniowej. W przeciwieństwie do metod korelacyjnych, które wykorzystują nieparametryczny model liniowego układu dynamicznego i parametryczny model elementu nieliniowego, metody optymalizacji liniowej wykorzystują parametryczne reprezentacje dla obu części układu Hammersteina.

Parametry reprezentacji funkcji przenoszenia impulsu elementu liniowego i modelu wielomianowego elementu nieliniowego można oszacować za pomocą iteracyjnej metody najmniejszych kwadratów zaproponowanej przez Narendrę i Gallmana [120]. Metoda ta opiera się na naprzemiennym dostosowywaniu parametrów liniowej i nieliniowej części modelu.

Inne podejście iteracyjne zostało zaproponowane przez Haist *et al.* [63] dla układów Hammersteina ze skorelowanymi addytywnymi zakłóceniami wyjściowymi. Metoda ta, będąca rozszerzeniem metody Narendry i Gallmana, przewyższa jej oczywistą wadę polegającą na tendencyjnych oszacowaniach poprzez estymację zarówno parametrów modelu Hammersteina, jak i liniowego modelu szumu.

Transformacja modelu SISO Hammersteina do postaci MISO umożliwia nieiteracyjną estymację parametrów przekształconego modelu metodą najmniejszych kwadratów [27]. Pomimo swojej prostoty i eleganckiej formy jednoetapowego rozwiązania, metoda ta ma niedogodność w postaci nadmiarowości w obliczaniu parametrów elementu nieliniowego. Dokładniej mówiąc, *nb* różnych zestawów tych parametrów można obliczyć na podstawie parametrów przekształconego modelu MISO, gdzie *nb* jest nominalnym rzędem funkcji przenoszenia impulsu.

Ze zmiennymi instrumentalnymi zdefiniowanymi jako wyjścia filtrów liniowych i moce

metody zmiennych instrumentalnych były badane przez Stoicę i Söderströma [152]. Udowodnili oni, że podejście zmiennych instrumentalnych daje spójne oszacowania parametrów w łagodnych warunkach.

Algorytm wykorzystujący model OBF liniowego systemu dynamicznego MIMO i nieliniowy model funkcji bazowej elementu nieliniowego MIMO został zaproponowany przez Gómeza i Baeyensa [44]. W tym podejściu model MIMO Hammersteina jest przekształcany do postaci liniowej w parametrach, a jego parametry są obliczane metodą najmniejszych kwadratów. Aby obliczyć parametry macierzy elementu nieliniowego na podstawie oszacowań parametrów przekształconego modelu, stosuje się technikę dekompozycji wartości pojedynczej. Algorytm zapewnia spójne oszacowania parametrów przy słabych założeniach dotyczących trwałości wzbudzenia wejść systemu.

Bai [11] zaproponował dwuetapowy algorytm, który oddziela identyfikację liniowego systemu dynamicznego od identyfikacji elementu nieliniowego. W pierwszym kroku algorytm wykorzystuje pseudolosową sekwencję binarną (PRBS) do identyfikacji liniowego systemu dynamicznego. Dzięki sygnałowi PRBS można wyeliminować efekt nieliniowości, ponieważ każda statyczna funkcja nieliniowa może być całkowicie scharakteryzowana przez funkcję liniową na wejściu PRBS. W związku z tym można zastosować dowolną metodę identyfikacji systemów liniowych w celu uzyskania liniowego modelu dynamicznego. Identyfikacja elementu nieliniowego odbywa się w drugim kroku. Ponieważ sygnał PRBS przyjmuje tylko dwie wartości, do identyfikacji elementu nieliniowego wykorzystywany jest nowy sygnał wejściowy, który jest wystarczająco bogaty, np. sekwencja pseudolosowa o równomiernym rozkładzie. Ważną zaletą tej techniki oddzielania jest to, że asymptotyczna wariancja oszacowania funkcji transferu systemu Hammersteina jest równa asymptotycznej wariancji oszacowania funkcji transferu systemu w przypadku liniowym [126].

W schemacie identyfikacji zaproponowanym przez Giri et al. [43] zastosowano fragmentaryczny model liniowy elementu nieliniowego oraz model funkcji przenoszenia impulsów liniowego układu dynamicznego. Zdefiniowany w ten sposób model Hammersteina jest przekształcany do postaci liniowej w parametrach, a estymacja parametrów jest przeprowadzana przy użyciu rekurencyjnego algorytmu gradientowego rozszerzonego o projekcję parametrów. Aby zapewnić zbieżność modelu do prawdziwego systemu, generowana jest stale ekscytująca sekwencja wejściowa.

Metody regresji nieparametrycznej. W metodach regresji parametrycznej zakłada się, że charakterystyka nieliniowa należy do klasy, którą można sparametryzować za pomocą skończonego zestawu parametrów. Oczywiście takie założenie jest bardzo restrykcyjne. Na przykład, powszechnie stosowana wielomianowa reprezentacja elementu nieliniowego wyklucza typowe nieciągłe charakterystyki, takie jak ograniczniki martwej strefy, ograniczniki twarde i kwantyzatory. Bardzo dużą klasę funkcji nieliniowych, w tym wszystkie mierzalne  $L_2$  funkcje, można zidentyfikować za pomocą regresji nieparametrycznej. Podejście regresji nieparametrycznej do identyfikacji układów Hammersteina zostało pierwotnie zaproponowane przez Greblickiego i Pawlaka [55] i badane

dalej w [56, 57, 103]. Procedury identyfikacji za pomocą regresji jądrowej dla różnych układów blokowych, w tym układów Hammersteina.

mersteina, były również badane przez Krzyżaka i Partykę [105]. Metody regresji nieparametrycznej składają się z dwóch oddzielnych etapów. Po pierwsze, funkcja odpowiedzi impulsowej systemu dynamicznego jest szacowana za pomocą standardowej metody korelacji. Następnie charakterystyka elementu nieliniowego  $f(\cdot)$  jest szacowana jako estymata regresji jądrowej

$$\hat{f}(u(n)) = \frac{\sum_{k=0}^{N-1} y(k+1)K\left(\frac{u(n)-u(k)}{l(N)}\right)}{\sum_{k=0}^{N-1} K\left(\frac{u(n)-u(k)}{l(N)}\right)}, \quad (1.81)$$

gdzie  $N$  jest liczbą pomiarów,  $K(\cdot)$  jest funkcją jądra,  $l(\cdot)$  jest ciągiem liczb dodatnich. W tej definicji  $0/0$  jest rozumiane jako zero.

Można wykazać, że szacunki regresji nieparametrycznej są zbieżne do prawdziwej charakterystyki elementu nieliniowego, gdy liczba pomiarów  $N$  dąży do nieskończoności. Greblicki i Pawlak wykazali również, że dla dostatecznie gładkich charakterystyk szybkość zbieżności wynosi  $O(N^{-2/5})$ . Wszystkie powyższe algorytmy identyfikacji odzyskują charakterystykę nieliniową w sposób nerekursywny. Dwa inne algorytmy oparte na oszacowaniach regresji jądra, zaproponowane przez Greblickiego i Pawlaka [58], pozwalają na rekurencyjną identyfikację charakterystyki nieliniowej.

Inna klasa metod nieparametrycznych wykorzystuje ortogonalne rozwinięcia funkcji nieliniowej. Greblicki [45] rozważał system Hammersteina napędzany losowym białym sygnałem wejściowym, z wyjściem systemu zakłóconym losowym białym szumem, i zaproponował dwie procedury nieparametryczne oparte na ortogonalnych rozwinięciach trygonometrycznych i hermitowskich. Oba te algorytmy zbiegają do nieliniowej charakterystyki systemu w sposób punktowy, a błąd całkowity zbiega do zera. Współczynnik zbieżności punktowej wynosi  $O(N^{-(2q-1)/4q})$ , gdzie  $q$  jest liczbą pochodnych charakterystyki nieliniowej. Identyfikacja systemów Hammersteina za pomocą algorytmów opartych na rozwinięciu szeregu Hermite'a z liczbą członów zależną nieliniowo od pomiarów wejścia-wyjścia była rozważana przez Krzyżaka i innych [106]. W tym podejściu system jest sterowany przez stacjonarny biały szum, a składowe liniowe i nieliniowe są estymowane jednocześnie. Zastosowanie estymacji szeregu Fouriera do identyfikacji nieliniowości w układach blokowych, w tym w układach Hammersteina, było badane przez Krzyżaka [102, 104]. W przypadku funkcji nieliniowych i gęstości sygnałów wejściowych mających skończone rozwinięcia w szereg Fouriera, takie podejście ma dwie zalety w porównaniu z regresją jądrową - wyższą wydajność obliczeniową i wyższe współczynniki zbieżności. Odzyskiwanie funkcji nieliniowej za pomocą metody opartej na wielomianach Legendre'a z odpowiednio dobraną liczbą wyrazów zostało zbadane przez Pawlaka [136]. Wykazano, że oszacowanie  $f(\cdot)$  jest globalnie spójne i określono współczynniki zbieżności. Greblicki i Pawlak [59] rozważali również identyfikację układów Hammersteina za pomocą wielomianów Laguerre'a. Alternatywa dla metod nieparametrycznych opartych na rozszerzeniach ortogonalnych

to algorytmy wykorzystujące aproksymację wielorozdzielczą. W kontekście identyfikacji systemów Hammersteina analiza wielorozdzielcza Haara została po raz pierwszy zastosowana przez Pawlaka i Hasiewicza [137]. Pomysł ten był dalej badany przez Hasiewicza [64, 65, 66]. Algorytmy aproksymacji wielorozdzielczej Haara są zbieżne punktowo. Ich formy i warunki zbieżności są takie same dla białego i skorelowanego addytywnego szumu wyjściowego. Ich kolejną zaletą jest szybsza zbieżność w porównaniu z innymi nieparametrycznymi algorytmami identyfikacji, które wykorzystują ortogonalne rozszerzenia szeregowo.

Idea identyfikacji nieliniowości w systemach zorientowanych blokowo za pomocą falek Daubechiesia była badana przez Śliwińskiego i Hasiewicza [155]. Ponieważ brak

W związku z tym, że nie ma zamkniętej postaci analitycznej, co sprawia, że falki Daubechies są praktycznie nie do zastosowania, zastosowano procedurę estymacji, która wykorzystuje przybliżenia, które są łatwe do obliczenia.

Metody optymalizacji nieliniowej. Podejście błędu predykcji do identyfikacji systemów Hammersteina zostało omówione przez Eskinat *et al.* [34]. W przeciwieństwie do metody najmniejszych kwadratów stosowanej przez Changa i Luusa [27], metody błędu predykcji umożliwiają oszacowanie parametrów funkcji przenoszenia impulsu układu liniowego i wielomianowego nieliniowego elementu charakterystycznego bezpośrednio, bez żadnej transformacji parametrów i nie ma problemu nadmiarowości parametrów. Metoda błędu predykcji wykorzystuje metodę Levenberga-Marquardta do aproksymacji Hessjana funkcji kosztu sumy kwadratów. Zaniedbanie faktu, że model jest nieliniowy w parametrach i traktowanie go jako liniowego prowadzi do metod regresji pseudoliniowej. Przykładem takiego podejścia jest metoda zaproponowana przez Boutayeb i Daro- uach [21], w której parametry systemu MISO Hammersteina z wyjściem zakłóconym addytywnie przez skorelowany szum są szacowane rekurencyjnie.

Metoda identyfikacji układów Hammersteina, która wykorzystuje dwusegmentowy model elementu nieliniowego, składający się z oddzielnych map wielomianowych dla wejść dodatnich i ujemnych, została zaproponowana przez Vořr̃osa [161]. Metoda ta również wykorzystuje modele w postaci pseudolinearnej do iteracyjnego obliczania parametrów. Idea metody opiera się na podziale charakterystyki nieliniowej, którą można dokładnie opisać wielomianem wysokiego rzędu, na dwa segmenty, które można aproksymować wielomianami znacznie niższego rzędu. Podobną technikę zastosował również Vořr̃os [160] do identyfikacji układów Hammersteina z elementem nieliniowym opisanym funkcją nieciągłą.

Podejście oparte na sieciach neuronowych można zastosować do identyfikacji układów Hammersteina z elementami nieliniowymi opisanymi funkcją ciągłą. Identyfikacja układów Hammersteina za pomocą modeli sieci neuronowych została rozważona przez Su i McAvoy [153]. Wykorzystali oni dane w stanie ustalonym i przejściowym do wytrenowania modelu Hammersteina w sieci neuronowej. W tym podejściu model sieci neuronowej elementu nieliniowego i model liniowego układu dynamicznego są trenowane oddzielnie. Najpierw sieć neuronowa jest trenowana za pomocą algorytmu uczenia wstecznego (BP) na danych w stanie ustalonym. Po

Sieć neuronowa służy jako operator nieliniowy. Liniowy model dynamiczny jest następnie trenowany na zbiorze danych przejściowych wykorzystujących zmienne wejściowe systemu przekształcone przez operator nieliniowy jako dane wejściowe. Rozważając dwie podstawowe konfiguracje liniowego modelu dynamicznego, tj. szeregowo-równoległą i równoległą, pokazano, że gradienty funkcji kosztu można uzyskać za pomocą metody propagacji wstecznej dla modelu szeregowo-równoległego i metody propagacji wstecznej w czasie dla modelu równoległego. Po obliczeniu gradientu sugeruje się zastosowanie algorytmu stromego zejścia lub gradientu sprzężonego w celu dostosowania parametrów modelu.

Identyfikacja systemów SISO Hammerstein za pomocą wielowarstwowych sieci neuronowych typu feedforward była również badana przez Al-Duwaish *et al.* [3]. Rozważali oni równoległy model sieci neuronowej Hammersteina trenowany rekurencyjnie przy użyciu algorytmu uczenia BP do trenowania modelu elementu nieliniowego i algorytmu najmniejszych kwadratów (RLS) do trenowania liniowego modelu dynamicznego. W tym podejściu częściowe pochodne kwadratowej funkcji kosztu są obliczane w sposób przybliżony, bez uwzględniania zależności przeszłych wyników modelu od jego parametrów. Odpowiada to równaniom (3.36) - (3.38). Ponadto nie uwzględnia się faktu, że częściowe pochodne wyników modelu względem wag modelu elementu nieliniowego zależą nie tylko od rzeczywistych, ale także od przeszłych wyników modelu elementu nieliniowego. Oba te uproszczenia obliczania gradientu zmniejszają złożoność obliczeniową uczenia za cenę zmniejszonej dokładności obliczania gradientu i mogą skutkować zmniejszonym współczynnikiem zbieżności. Rozszerzenie algorytmu RLS/BP na przypadek MIMO omówiono w [4].

Podejście genetyczne do identyfikacji systemów Hammersteina jest rozważane w [5]. W podejściu tym identyfikowana jest postać biegunowo-zerowa liniowego układu dynamicznego i nieliniowość znanej struktury, ale nieznanymi parametrów. Ogólnie rzecz biorąc, modele sieci neuronowej SISO Hammersteina mogą być reprezentowane przez wielowarstwowy model perceptronowy elementu nieliniowego i węzeł liniowy z dwoma stożkowymi liniami opóźniającymi używanymi jako model układu liniowego [73]. Można rozważyć zarówno modele szeregowo-równoległe, jak i równoległe. Mają one podobną strukturę, a jedyną różnicą jest połączenie sprzężenia zwrotnego w modelu równoległym [100]. W przypadku modelu szeregowo-równoległego gradient jego wyjścia w odniesieniu do parametrów modelu można uzyskać za pomocą efektywnego obliczeniowo algorytmu propagacji wstecznej. Obliczenia gradientu w modelach równoległych można dokonać za pomocą metody wrażliwości lub metody propagacji wstecznej w czasie [75, 77, 90]. Ponieważ model Hammersteina zawiera część liniową, możliwe jest również zastosowanie algorytmów niejednorodnych, które łączą rekurencyjne algorytmy najmniejszych kwadratów lub rekurencyjne algorytmy regresji pseudolinearnej ze wszystkimi wyżej wymienionymi metodami [75]. Modele sieci neuronowych Hammersteina mają prostą architekturę, zwykle z kilkudziesięcioma węzłami przetwarzania i adekwatnymi wagami, a wytrenowane modele są łatwe do zastosowania w praktyce. Ze względu na prostą architekturę modeli sieci neuronowych Hammersteina, ich algorytmy uczenia mają niską złożoność obliczeniową. Więcej szczegółów na temat

Podejście sieci neuronowej do identyfikacji systemu Hammersteina można znaleźć w rozdziale 3.

## 1.5 Podsumowanie

Materiał przedstawiony w tym rozdziale rozpoczyna się od zwięzłego wprowadzenia wyjaśniającego motywację do napisania tej książki. Następnie, sekcja 1.1 zawiera przegląd dyskretnych modeli czasowych systemów dynamicznych. Zaczyna się od dobrze znanych struktur modeli liniowych i pokazuje modele nieliniowe jako generalizacje modeli liniowych. Sekcja 1.2 przedstawia perceptron wielowarstwowy, architekturę sieci neuronowej, która jest używana zarówno w modelach sieci neuronowej Wienera, jak i Hammersteina. Następnie, w sekcjach 1.3 i 1.4, krótko omówiono różne dostępne metody identyfikacji systemów Wienera i Hammersteina i podzielono je na następujące cztery grupy: metody korelacyjne, metody optymalizacji liniowej, metody regresji nieparametrycznej i metody optymalizacji nieliniowej.