

Politechnika Wrocławska
Wydział Elektroniki
Wydziałowy Zakład Systemów i Sieci Komputerowych

Raport Serii Preprinty nr 1/00

JACEK CICHOSZ

Komputerowe algorytmy
złożonej identyfikacji rekurencyjnej

ROZPRAWA DOKTORSKA

PROMOTOR

Prof. dr hab. inż. Marek Kurzyński

Wrocław 2000

Spis treści

Wykaz oznaczeń	4
1. Wstęp	5
1.1. Wprowadzenie	5
1.2. Zarys tematyki pracy	7
1.3. Rezultaty	10
2. Matematyczny opis zadania identyfikacji	13
2.1. Pełna informacja probabilistyczna	13
2.2. Niepełna informacja probabilistyczna	14
2.2.1. Identyfikacja parametryczna	16
2.2.2. Modele liniowe	19
2.3. Dyskusja przypadków doboru modelu i zbieżność	20
2.3.1. Interpretacja zbieżności	24
3. Algorytmy o strojonej strukturze modelu	27
3.1. Wprowadzenie	27
3.2. Podstawowa wersja algorytmu	29
3.3. Kryteria rekonstrukcji struktury modelu	31
3.4. Przypadek szczególny dla modelu regresji liniowej	35
3.5. Algorytm ze stałą pamięcią	42
4. Dwustopniowe algorytmy rekurencyjne	48
4.1. Zagadnienia wstępne	48
4.2. Opis matematyczny i sformułowanie problemu	52
4.3. Przypadek skalarny, model liniowy	54
4.4. Dwustopniowa identyfikacja obiektu opisanego równaniami stanu	59
4.5. Konstrukcja algorytmu dla skalarnego obiektu dwustopniowego	63
4.6. Uwagi o identyfikacji obiektów wielowymiarowych	70
5. Własności złożonych algorytmów identyfikacji	72
5.1. Przypadek deterministyczny	72
5.2. Identyfikacja w obecności zakłóceń	75
6. Zastosowanie złożonych algorytmów rekurencyjnych	77
6.1. Identyfikacja procesu farmakokinetycznego	77

6.1.1. Sformułowanie zadania	78
6.1.2. Dane pomiarowe	79
6.1.3. Wyniki identyfikacji	80
6.2. Modelowanie i identyfikacja stężenia mocznika w procesie dializy.....	80
6.2.1. Charakterystyka problemu	81
6.2.2. Model poziomu mocznika w procesie hemodializy	83
6.2.2.1. Model ze stałą objętością kompartmentu	85
6.2.2.2. Model ze zmienną objętością kompartmentu	86
6.2.3. Wyniki identyfikacji	88
6.3. Identyfikacja procesu odsiarczania spalin	89
6.3.1. Charakterystyka problemu	89
6.3.2. Model matematyczny procesu	92
6.3.3. Materiał pomiarowy i opis eksperymentu	94
6.3.4. Wyniki identyfikacji	95
7. Uwagi końcowe	97
Literatura	99

Wykaz oznaczeń

a	wektor nieznanych parametrów modelu
A^T	macierz transponowana
A^{-T}	transponowana macierz odwrotna, tzn. $(A^{-1})^T$
A^{-1}	macierz odwrotna
$\arg \min f(u)$	wartość u , która minimalizuje funkcję $f(u)$
D_r	zbiór dopuszczalnych wartości r -wymiarowego wektora parametrów
$\dim(A)$	rozmiar macierzy (wektora)
Eu	wartość oczekiwana zmiennej losowej u
$E(y/u)$	warunkowa wartość oczekiwana
$\{e_t\}$	ciąg zmiennych losowych
$\hat{f}$	estymator funkcji gęstości prawdopodobieństwa
g^0	teoretyczny model systemu (model użyty do generacji danych)
$\hat{g}_N$	estymator modelu systemu otrzymany na podstawie N obserwacji
I_n	macierz jednostkowa stopnia n
$Var(u)$	wariancja zmiennej losowej u
$Q(S_N, a)$	kryterium identyfikacji
$\hat{\varepsilon}_t$	błąd resztkowy modelu
$\varphi, \varphi^{(k)}$	wektor regresji, k -ta składowa tego wektora
$\varphi_{r,t}$	wektor regresji o r składowych wyznaczony w momencie t .
R^d	d -wymiarowa przestrzeń Euklidesowa
P	macierz kowariancji
$P(A)$	prawdopodobieństwo zdarzenia A
$P(A B)$	prawdopodobieństwo warunkowe
S_N	ciąg uczący
u_t, y_t	wejście i wyjście systemu w momencie t
U_t, Y_t	szeregi czasowe obserwacji wejściowych i wyjściowych od momentu 1 do t
$U_{t+N,t+1},$	
$Y_{t+N,t+1}$	j.w., lecz odpowiednio od momentu $t + 1$ do momentu $t + N$

1. Wstęp

1.1. Wprowadzenie

Identyfikacja zajmuje się konstruowaniem modelu matematycznego systemu na podstawie danych eksperymentalnych. Przez model matematyczny rozumiemy tutaj związek pomiędzy wyróżnionymi wielkościami systemu wyrażony najczęściej w postaci zestawu formuł matematycznych, wykresów lub tabel wraz z ograniczeniami i założeniami upraszczającymi. Takie podejście do konstruowania modeli matematycznych znalazło wiele praktycznych zastosowań w różnych dziedzinach technicznych, a także ekonomii, biologii, medycynie i innych dyscyplinach. Na tak liczne zastosowania praktyczne, a częściowo także na rozwój teorii identyfikacji, miało wpływ rozpowszechnienie technologii mikroprocesorowej i wzrost mocy obliczeniowej komputerów. Pomimo to, stosowane metody są wciąż niedoskonałe i często niewystarczające do rozwiązywania rzeczywistych problemów. Jedną z przyczyn tej niedoskonałości jest trudność ścisłego określenia schematu postępowania, który prowadziłby w każdym przypadku do otrzymania modelu o jak najlepszej jakości. W konsekwencji wzrastają koszty i czas fazy projektowania algorytmów, w której powtarzane są wielokrotnie eksperymenty z różnymi strukturami modeli, by po weryfikacji wybrana została ta z nich, która daje najlepsze, w przyjętym sensie, dopasowanie do danych pomiarowych. Problemy tego rodzaju próbuje się przezwyciężać uwzględniając, na ile to możliwe, wiedzę wstępną o systemie [7, 42, 43], a w przypadku znacznego jej ograniczenia lub braku, stosując struktury modeli charakteryzujące się dużą elastycznością i zdolnością do generalizacji jak. np. sieci neuronowe [54, 32, 61], czy modele falkowe [30]. Inne spotykane w literaturze podejście pozwalające usprawnić iteracyjny proces eksperymentalnego otrzymywania modelu polega na budowaniu uniwersalnych komputerowych systemów wspomagających identyfikację [45] lub ekspertowych systemów identyfikacji [22]. Próby takie podejmowano również i w Polsce [55], choć dotyczyły one doboru parametrów szczególnej klasy modeli liniowych – modeli autoregresyjnych typu ARX. Rozwój powyższych systemów przyczynił się do zautomatyzowania i usprawnienia złożonego procesu identyfikacji będąc czymś

więcej niż pakietem standardowych procedur numerycznych dzięki wbudowanemu mechanizmowi wnioskującemu, opartemu na zbiorze reguł danych przez eksperta. Mechanizm ten, oparty na metodach logiki formalnej i automatycznego dowodzenia twierdzeń, w istocie naśladuje sposób rozumowania eksperta, ułatwiając wybór metody identyfikacji bądź ocenę i weryfikację wynikowego modelu. Znane z literatury podejścia ograniczają się zwykle do wykorzystywania jednej struktury modelu w czasie trwania całego eksperymentu. Uwaga ta dotyczy również algorytmów rekurencyjnych, w których parametry modelu systemu podlegają dostrajaniu w miarę napływania poszczególnych obserwacji.

Można wymienić takie zastosowania identyfikacji jak: sterowanie procesami przemysłowymi [2, 73, 56], filtracja sygnałów [1], prognozowanie [8], wykrywanie uszkodzeń w systemach [3, 32], opracowywanie danych pomiarowych [24, 60, 52], wspomaganie diagnostyki medycznej [21, 17, 37, 38, 71, 11, 12, 15, 29].

Identyfikacja funkcjonuje na pograniczu automatyki i informatyki i jest już dziś dziedziną o ugruntowanej metodologii, której poświęcono całościowe monografie [44, 69, 24].

Dla ilustracji praktycznych zastosowań identyfikacji rozważmy system złożony ze zbiornika z cieczą dostarczaną za pośrednictwem pompy napędzanej silnikiem elektrycznym [42]. Interesuje nas zależność poziomu cieczy w zbiorniku od prądu płynącego w uzwojeniach silnika. Taki model mógłby być następnie wykorzystany do zaprojektowania regulatora poziomu cieczy w zbiorniku.

Inny przykład, z dziedziny medycyny, dotyczy modelowania zmian poziomu glukozy we krwi u osób chorych na cukrzycę [30]. U ludzi zdrowych poziom glukozy jest regulowany za pośrednictwem insuliny – hormonu wydzielanego przez trzustkę. Diabetycy natomiast muszą otrzymywać insulinę z zewnątrz bądź to w formie zastrzyków, bądź ze specjalnej pompy dozującej. Trudność polega na podjęciu decyzji kiedy i jaką dawkę insuliny należy podać danemu pacjentowi. Przydatny byłby więc model ilościowy poziomu glukozy w zależności od rodzaju diety, czasu jaki upłynął od ostatniego posiłku, stresów i emocji, aktywności fizycznej i bieżącego poziomu glukozy.

Ważnym obszarem zastosowań identyfikacji jest budowa modeli wykorzystywanych do otrzymywania optymalnych prognoz wyróżnionych wielkości systemu [8]. Przykładowo, obserwując zbył pewnego towaru w kolejnych miesiącach można za pomocą odpowiedniego modelu przewidzieć jego wielkość z zadaniem wyprzedzeniem. Zbudowany model można wykorzystać jako narzędzie wspomagające podejmowanie efektywnych decyzji ekonomicznych. Podane wstępnie przykłady mają charakter ilustracyjny i nie wyczerpują w pełni bogactwa możliwości zastosowań identyfikacji.

1.2. *Zarys tematyki pracy*

Tematem rozprawy są komputerowe algorytmy złożonej identyfikacji rekurencyjnej. Dane wejściowe rozpatrywanych w pracy algorytmów mają postać dogodną do przetwarzania komputerowego tzn. tworzą ciągi liczb reprezentujące wielkości pomiarowe. Złożoność rekurencyjnych metod identyfikacji badanych w pracy leży w koncepcji algorytmu lub w dwustopniowej strukturze modelu identyfikowanego procesu. Pozwala to wyróżnić dwa zasadniczo odmienne typy algorytmów będących przedmiotem rozprawy. W pierwszym z nich dostrajane są na podstawie przychodzących obserwacji nie tylko wartości parametrów, lecz również postać modelu. Drugi rodzaj algorytmów dotyczy rekurencyjnej identyfikacji dwustopniowej, w której charakter złożoności procedury wynika z komplikacji założonej postaci modelu mającego strukturę dwustopniową.

Znane z literatury podejścia dotyczą identyfikacji ustalonej struktury modelu. W przypadku znikomej wiedzy o obiekcie lub zupełnym jej braku postać modelu często przyjmowana jest ad hoc i podlega weryfikacji w końcowym etapie identyfikacji decydując o przyjęciu bądź odrzuceniu wybranej a priori struktury. Podejście opisane w pracy opiera się na idei sekwencyjnej weryfikacji modelu i w razie potrzeby zwiększania jego komplikacji w miarę pozyskiwania informacji pomiarowej. Inspiracją, a zarazem intuicyjną przesłanką przemawiającą na korzyść prezentowanego podejścia do problemu jest z jednej strony „zasada oszczędności” [69, 44], czyli stosowania jak najprostszego modelu, a z drugiej pewna analogia z algorytmami nieparametrycznymi [26, 36], w których wynik jest silnie uzależniony od napływających danych. Można się spodziewać, że stosowanie jak

najprostszej formy modelu powinno również przynieść w pewnych przypadkach poprawę własności numerycznych algorytmu. Modele przeparametryzowane mogą bowiem prowadzić do gorszego uwarunkowania numerycznego procedur identyfikacji o czym wspomina wielu autorów [44, 69, 5]. We wstępie scharakteryzowanym typie algorytmu, który zostanie opisany w pracy, mieliśmy do czynienia ze złożonością koncepcyjną metody zastosowanej do klasycznych struktur modeli.

Drugi wątek rozprawy dotyczy dwustopniowej identyfikacji rekurencyjnej. Prace badawcze nad identyfikacją dwustopniową i jej zastosowaniami były prowadzone przez polskich autorów [71, 35, 37, 38]. Badania te dotyczyły głównie identyfikacji charakterystyk statycznych i omawiały wyłącznie algorytmy bezpośrednie*. W monografii [71], poświęconej w całości problematyce identyfikacji dwustopniowej, zaproponowano jako jeden z kierunków dalszych badań opracowanie dwustopniowych algorytmów rekurencyjnych. W niniejszej pracy podano metodykę projektowania algorytmów identyfikacji na bieżąco, gdy model ma strukturą dwustopniową. Zaproponowano odmienne podejścia dla modeli wejściowo-wyjściowych i dla modeli w przestrzeni stanu. Uzasadniono potrzebę zastosowania takich algorytmów w adaptacyjnych systemach podejmowania decyzji i sterowania.

W obydwu odrębnych koncepcyjnie wątkach tematycznych poruszanych w pracy ważną rolę odgrywa wybór klasy modeli. O ile w podejściu pierwszym postać modelu kształtuje się w miarę pozyskiwania danych przy niewielkim udziale wiedzy wstępnej, to w algorytmach dwustopniowych postać ta jest mocno zdeterminowana przez wiedzę a priori o modelowanym procesie.

W literaturze wyróżniono trzy główne stopnie znajomości informacji a priori i związane z nimi podejścia do identyfikacji:

- Modele teoretyczne, których postać jest dokładnie znana i wyznaczona w oparciu o wiedzę wstępną (np. prawa fizyczne, zasady zachowania, równania bilansu itp.) [51, 66, 4].

* Algorytmy, w których oceny nieznanymi parametrów otrzymuje się po zgromadzeniu całej serii pomiarowej.

- Modele, których postać jest znana z dokładnością do pewnego zbioru parametrów lub w których wiedza wstępna stanowi wskazówkę do konstrukcji pewnych składników struktury modelu mających określoną interpretację fizyczną [7, 11, 42].
- Modele typu „czarna skrzynka”, które w bardzo niewielkim stopniu lub w ogóle nie wykorzystują wiedzy wstępnej. Klasa modeli ma wówczas na tyle uniwersalny charakter by pozwalala na modelowanie szerokiego spektrum obiektów [30, 41, 62].

Większość badań z dziedziny identyfikacji dotyczy ostatniego z wymienionych podejść. Najwięcej prac z tego zakresu poświęconych jest teorii modeli liniowych [97, 44, 53, 24, 18, 10]. Algorytmy o strojonej strukturze modelu, analizowane w pracy, wykorzystują ostatni z wymienionych typ modelu, natomiast algorytmy dwustopniowe mają ściślejsze powiązanie z fizyczną analizą systemu i z tej przyczyny są w nich częściej stosowane modele z wiedzą wstępną. Uwidocznimy to w przykładach rzeczywistych obiektów analizowanych w pracy.

W wielu praktycznych zastosowaniach klasyczna teoria modeli liniowych okazała się niewystarczająca. Rozwinął się nurt badań modelowania systemów nieliniowych będących rozszerzeniem lub modyfikacją klasycznych liniowych struktur modeli [6, 41]. Równolegle rozwijane były metody nieparametryczne oparte głównie na wygładzaniu histogramu poprzez jądrowe estymatory funkcji gęstości prawdopodobieństwa [26]. W pracy ograniczono się do przypadków liniowych względem parametrów struktur modeli zakładając, że ewentualne nieliniowości da się ująć poprzez odpowiednie związki wejściowych zmiennych systemu. W przypadku rekurencyjnych algorytmów dwustopniowych zastosowano linearyzację.

Kolejną trudnością występującą w modelowaniu rzeczywistych systemów jest nieprecyzyjna informacja obserwacyjna lub pomiarowa. Z problemem takim spotykamy się bardzo często w systemach biomedycznych, jak w podanym wcześniej przykładzie modelowania poziomu insuliny, w którym wielkości takie jak rodzaj diety, poziom stresów czy emocji trudno jest wyrazić w precyzyjny ilościowy sposób. Do budowy modeli takich systemów wykorzystano zbiory

i logikę rozmytą [42, 46, 70]. Powyższe systemy wykorzystują reguły, które są podawane przez eksperta (np. doświadczonego lekarza) znającego dobrze zjawiska zachodzące w modelowanym procesie. Ten rodzaj opisu nieprecyzyjnej informacji nie będzie rozważany w pracy, choć należy o nim wspomnieć w kontekście sposobu reprezentacji informacji opisującej obiekt.

W odróżnieniu od podejścia analitycznego do modelowania [47] opartego na prawach fizycznych (zasadach zachowania, równaniach bilansu), w których związki pomiędzy wielkościami systemu są ściśle zdeterminowane, w identyfikacji chodzi nam o znalezienie jak najlepszego przybliżenia zależności między zmiennymi na podstawie danych pomiarowych, a więc konstruowany model rzeczywistego systemu zawsze obarczony jest jakąś dozą niepewności, która wynika co najmniej z dwóch powodów: po pierwsze z rozbieżności między rzeczywistym systemem, a zaproponowaną strukturą modelu; po drugie z obecności w systemie sygnałów zakłócających. Wybór klasy modeli jest podyktowany wiedzą wstępną o systemie pozwalającą raczej na postawienie hipotezy o postaci modelu, która jest weryfikowana eksperymentalnie w wyniku identyfikacji. Istnienie w obiekcie losowych zakłóceń powoduje podstawową trudność w podejściu do konstrukcji modeli, zmuszając do wykorzystywania aparatu rachunku prawdopodobieństwa i statystyki do analizy ogromnego bogactwa metod, które proponuje identyfikacja. W pracy stosujemy probabilistyczną miarę niepewności. Przyjęto więc, że sygnały zakłócające mają charakter dyskretnych szeregów czasowych, w których wartości w dowolnych momentach są zmiennymi losowymi o jednakowych rozkładach. W pewnych przypadkach rozważamy przypadki czysto deterministyczne w celu uproszczenia analizy.

1.3. **Rezultaty**

Rezultaty pracy mają charakter zarówno metodologiczny jak i praktyczny. Do rezultatów metodologicznych należą:

- opracowanie algorytmów identyfikacji rekurencyjnej z adaptacyjną zmianą struktury modelu z rozważeniem przypadku algorytmu wymagającego stałej ilości pamięci,

- propozycje kryteriów zmiany struktury modelu dla wymienionych przypadków algorytmów,
- opracowanie algorytmów rekurencyjnej identyfikacji dwustopniowej dla modeli wejściowo-wyjściowych oraz w przestrzeni stanu.

Rezultaty praktyczne obejmują przykładowe zastosowania opracowanych algorytmów w biomedycynie i technice:

- zastosowanie algorytmu z adaptacyjnym doбором struktury modelu do identyfikacji dystrybucji propranololu,
- identyfikację procesu odsiarczania węgla przy pomocy algorytmu dwustopniowego,
- wykorzystanie dwustopniowej identyfikacji na bieżąco jako elementu wspomagającego terapię hemodializy.

W niniejszym rozdziale przedstawiono przedmiot pracy, jej podstawowe założenia i cele wraz z krótkim wprowadzeniem do zagadnienia identyfikacji. Drugi rozdział podaje dokładny opis zadania identyfikacji, wprowadza potrzebną terminologię i oznaczenia. Przeprowadzono tam również dyskusję doboru modelu i podano interpretację zbieżności modelu do idealnej charakterystyki systemu. W rozważaniach teoretycznych przyjęto probabilistyczny opis sygnałów zakłócających. Rozdział trzeci zawiera omówienie algorytmów z dostrajaną na bieżąco strukturą modelu wraz z propozycjami odpowiednich kryteriów przebudowy postaci modelu. Rozpatrzono szczegółowo przypadek szczególny dla modelu w postaci regresji liniowej względem parametrów. Rekurencyjna identyfikacja modeli dwustopniowych jest przedmiotem rozdziału czwartego. Zbadano takie algorytmy dla modeli wejściowo-wyjściowych i w formie równań stanu. Pokazano, że nawet proste skalarne przypadki, dokładnie analizowane w pracy, prowadzą do problemów nieliniowych. Dla modelu wejściowo-wyjściowego zastosowano podejście bayesowskie, a identyfikację dwustopniowego modelu stanowego sprowadzono do uogólnionego filtru Kalmana. Piąty rozdział traktuje o własnościach zbadanych w pracy algorytmów i zawiera również badania symulacyjne. Zastosowania biomedyczne i techniczne złożonych algorytmów identyfikacji rekurencyjnej są przedmiotem rozdziału szóstego. Rozważamy

zastosowanie algorytmu o dostrajanej strukturze modelu do modelowania farmakokinetycznego dystrybucji propranololu oraz algorytmu dwustopniowego do identyfikacji parametrów modelu mocznikowego, który może stanowić narzędzie wspomagające terapii hemodializy. Trzecie zastosowanie dotyczy identyfikacji procesu eliminacji zanieczyszczeń pochodzenia siarkowego z gazów spalinowych bloku energetycznego opalanego węglem brunatnym. Podsumowanie pracy i kierunki dalszych badań zawarto w rozdziale siódmym.

2. Matematyczny opis zadania identyfikacji

Celem niniejszego rozdziału jest podanie formalnego opisu matematycznego rozpatrywanego zadania identyfikacji, której przedmiotem jest przewidywanie wartości pewnej zmiennej $y \in R$ na podstawie informacji niesionej przez inną zmienną $u \in R^d$. Umownie nadaje się znaczenie zmiennym u i y przyjmując, że wielkości te są odpowiednio wejściem i wyjściem systemu.

2.1. Pełna informacja probabilistyczna

Przyjmijmy model probabilistyczny zakładając, że (u, y) jest parą zmiennych losowych* o pewnym łącznym rozkładzie. Celem identyfikacji jest znalezienie funkcji $g(u)$ takiej, by różnica

$$\varepsilon = y - g(u) \quad (2.1)$$

między zmienną wyjściową a wartością funkcji była jak najmniejsza w przyjętym sensie. W terminologii statystycznej powyższy problem nazywa się problemem wyznaczania *funkcji regresji pierwszego rodzaju*. Miarę jakości identyfikacji pozwalającą porównywać ze sobą różne modele nazywamy *kryterium identyfikacji* i oznaczamy symbolem Q . Jeśli przyjmujemy kryterium średniokwadratowe

$$Q(g) = E[y - g(y)]^2, \quad (2.2)$$

to problem regresji sprowadza się do znalezienia takiej funkcji $\hat{g}$, która minimalizuje kryterium (2.2), czyli

$$Q(\hat{g}) = \min_g Q(g). \quad (2.3)$$

Otrzymujemy następujący wynik

$$\hat{g}(u) = E(y/u). \quad (2.4)$$

Jeśli zmienne losowe u i y są typu ciągłego, to zależność (2.4) możemy rozwinąć do postaci

* W celu uproszczenia zapisów nie wprowadzono notacyjnego rozróżnienia zmiennych losowych i wartości tych zmiennych. Kontekst wyrażenia lub objaśnienie w tekście wskazuje, o którą wielkość chodzi

$$\hat{g}(u) = f(u, y) / f_u(u), \quad (2.5)$$

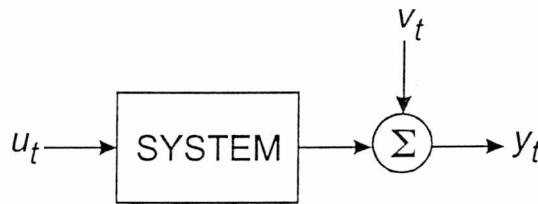
gdzie $f(u, y)$ jest gęstością łączną pary zmiennych losowych (u, y) , a $f_u(u)$ jest gęstością brzegową zmiennej losowej u . Rezultat (2.4) stanowi rozwiązanie zadania wyznaczania funkcji regresji w przypadku pełnej informacji probabilistycznej. Przypadek ten rzadko występuje w praktyce, ponieważ na ogół funkcje gęstości f i f_u nie są znane, stanowi jednak podstawę wyprowadzenia wielu algorytmów identyfikacji. Przykładowo, w metodach nieparametrycznych, opisywanych w [26] w miejsce nieznanych funkcji gęstości w (2.5) wstawiane są ich estymatory $\hat{f}$ i $\hat{f}_u$ konstruowane na podstawie danych pomiarowych.

2.2. Niepełna informacja probabilistyczna

W praktyce nie znamy rozkładów zmiennych losowych u i y lecz dysponujemy informacją pomiarową w postaci *ciągu uczącego*

$$S_N = \{(u_1, y_1), \dots, (u_N, y_N)\}, \quad (2.6)$$

który stanowi zbiór N par pomiarów wejścia i wyjścia systemu uzyskanych w drodze eksperymentu. Eksperyment ten może być *aktywny*, gdy mamy wpływ na wybór sygnału wejściowego lub *pasywny*, gdy odpowiednie wielkości mierzone są w rzeczywistych warunkach funkcjonowania systemu. Dobór sygnału wejściowego, nie rozpatrywany w pracy, jest ważnym elementem wstępnej fazy procesu



Rys. 2.1. Wpływ zakłóceń pomiarowych na układ identyfikacji

identyfikacji [60, 9]. Ciąg S_N jest pewną realizacją pary zmiennych losowych (u, y) . Związek pomiędzy zmiennymi u i y wyraża następujące równanie

$$y_t = g^0(u_t) + v_t \text{ dla } t = 1, \dots, N. \quad (2.7)$$

W zależności (2.7) g^0 jest teoretyczną charakterystyką* (idealnym modelem) systemu. Składnik v_t na rys. 2.1 reprezentuje zakłócenia addytywne spowodowane np. ograniczoną dokładnością czujników pomiarowych, szumami cieplnymi itp. Przyjęcie założenia addytywności ułatwia analizę choć rozważa się również inny charakter oddziaływań zakłóceń w systemie. Można interpretować v_t jako tą składową sygnału wyjściowego, której nie da się przewidzieć na podstawie obserwacji uzyskanych z systemu.

Zakładamy, że ciąg zmiennych losowych $\{v_t\}$ jest niezależny od sygnału wejściowego $\{u_t\}$. Podkreślamy, że w identyfikacji nie chodzi o znalezienie dokładnej charakterystyki modelu g^0 lecz o jak najdokładniejsze jej przybliżenie. Konstruowany model rzeczywistego systemu zawsze obarczony jest błędem.

Dla $N \geq 1$ definiujemy ocenę funkcji g^0 następująco

$$\hat{g}_N : S_N \rightarrow \hat{g}_N(S_N, \cdot). \quad (2.8)$$

Dla ustalonego ciągu uczącego S_N funkcja

$$u \rightarrow \hat{g}(S_N, u) \quad (2.9)$$

jest estymatorem funkcji regresji. Przedstawmy ciąg uczący (2.5) w postaci odpowiednio wejściowego i wyjściowego szeregu czasowego. Zapiszmy szeregi w zwężłej macierzowej postaci jako:

$$U_N = [u_1 \dots u_N], \quad (2.10a)$$

$$Y_N = [y_1 \dots y_N]. \quad (2.10b)$$

Szeregi te reprezentują obserwacje systemu, którymi dysponujemy w chwili $t = N$. Naszym celem jest skonstruowanie modelu systemu w postaci:

$$\hat{y}_t = \hat{g}_N(\varphi_t), \quad (2.11)$$

gdzie $\hat{y}_t$ oznacza wyjście modelu, a φ_t nazywamy *wektorem regresji*

$$\varphi_t = [\varphi_t^{(1)} \dots \varphi_t^{(r)}]^T. \quad (2.12)$$

* W eksperymentach symulacyjnych g^0 oznacza charakterystykę systemu, przy pomocy której wygenerowano dane pomiarowe.

Składowe wektora (2.12) są funkcjami pomiarów w bieżącym i poprzednich momentach czasu zawartych w ciągu (2.10). Jeśli wszystkie składowe wektora (2.12) zależą od pomiaru w chwili bieżącej, czyli gdy

$$\varphi_t = \varphi(u_t), \quad (2.13a)$$

to takie zadanie nazywamy *identyfikacją charakterystyk statycznych*. Jeśli natomiast wektor regresji zależy od pomiarów z przeszłości, tj. gdy

$$\varphi_t = \varphi(U_{t-1}, Y_{t-1}), \quad (2.13b)$$

to taki przypadek nazywamy *identyfikacją charakterystyk dynamicznych*. Zadanie estymacji funkcji regresji na podstawie obserwacji znane jest w terminologii statystycznej jako problem wyznaczania *funkcji regresji drugiego rodzaju* i może być wykonane na dwa sposoby. Z pierwszym z nich, zwanym *estymacją parametryczną*, mamy do czynienia wtedy, gdy postać modelu g znana jest z dokładnością do skończonego zbioru nieznanych parametrów. Przypadek drugi, określany mianem *estymacji nieparametrycznej* występuje wtedy, gdy postać modelu nie zawiera explicite nieznanych parametrów.

2.2.1. Identyfikacja parametryczna

W przypadku identyfikacji parametrycznej postać modelu (*klasa modeli*) jest dokładnie określona i zależy od skończonego zbioru nieznanych parametrów, co zapisujemy

$$g(\varphi_t, a), \quad a \in D_r \subset R^r, \quad (2.14)$$

gdzie D_r jest zbiorem dopuszczalnym wartości parametrów, a zapisanych jako wektor kolumnowy

$$a = [a^{(1)} \dots a^{(r)}]^T. \quad (2.15)$$

Przy tak określonej klasie modeli zadanie identyfikacji polega na wyznaczeniu oceny nieznanych parametrów, którą oznaczamy przez $\hat{a}_N$, wykorzystując informacje zawarte w ciągu uczącym, czyli

$$\hat{g}_N(\varphi_t) = g(\varphi_t, \hat{a}_N). \quad (2.16)$$

2.1.1. Identyfikacja parametryczna

Estymator nieznanego parametru jest wyznaczany w oparciu o kryterium identyfikacji $Q(S_N, a)$, które wprowadza uporządkowanie w klasie modeli (2.15). Dla ustalonego S_N i funkcji g kryterium $Q(S_N, a)$ przyjmuje różne wartości dla różnych ocen a dając możliwość wyboru optymalnego oszacowania nieznanego parametru $\hat{a}_N$. Kryteria identyfikacji mogą mieć różne postacie [44, 9], jednak najczęściej jest stosowane kryterium średniokwadratowe wyrażone zależnością:

$$Q(S_N, a) = \sum_{t=1}^N \|y_t - g(\varphi_t, a)\|^2, \quad (2.17)$$

przy czym $\|\cdot\|$ oznacza szeroko rozumianą normę. Najlepsza ocena parametru w sensie średniokwadratowym minimalizuje (2.17), co zapisujemy

$$Q(S_N, \hat{a}_N) = \min_a Q(S_N, a). \quad (2.18)$$

Z dotychczasowej dyskusji wynika, że nie znając dokładnej charakterystyki obiektu g^0 , wybierana jest pewna aproksymacja tej charakterystyki g , dzięki czemu problem wyznaczania modelu zostaje zredukowany do wyestymowania nieznanymi parametrów. Zauważmy, że g jest złożeniem dwóch odwzorowań:

- φ (zależność 2.13) odwzorowuje wejściową przestrzeń obserwacji w przestrzeń regresji. Dobór φ wynika najczęściej z wiedzy wstępnej na temat systemu. Składowe wektora regresji mogą być nieliniowymi funkcjami wielkości wejściowych [42, 7].
- $g(\varphi_t)$ (zależność 2.14) odwzorowuje przestrzeń regresji w przestrzeń wyjściową. Odwzorowanie to w przypadku braku wiedzy a priori o systemie opiera się na założeniu regularności charakterystyki obiektu nie mając żadnej interpretacji fizycznej.

Założmy, że idealną charakterystykę obiektu można przedstawić w postaci rozwinięcia w szereg

$$g^0(\varphi_t) = \sum_{j=0}^{\infty} a^{(j)} g_j(\varphi_t), \quad (2.19a)$$

gdzie $\{g_j(\cdot)\}$, $j = 1, \dots$ jest znaną a priori rodziną *funkcji bazowych*. Wówczas wspomniany wyżej warunek regularności polega na zanikaniu w pewien sposób współczynników $a^{(j)}$ w rozwinięciu (2.19), gdy $j \rightarrow \infty$. Taki warunek jest spełniony dla bardzo szerokiej klasy funkcji. Po nieznacznej modyfikacji, charakterystyka (2.19) obejmuje cały szereg powszechnie znanych, uniwersalnych struktur modeli, które zebrano w tabeli 2.1. Modyfikacja ta polega na uogólnieniu klasy modeli (2.19) do postaci

$$\sum_{j=1}^{\infty} a^{(j)} g_j(\varphi_t, \beta_j, \gamma_j), \quad (2.19b)$$

przy czym wektor parametrów ma postać $[a^T \quad \beta^T \quad \gamma^T]^T$. Wyznaczenie modelu polega na zastąpieniu nieskończonego rozwinięcia (2.19) jego aproksymacją zawierającą skończony zbiór współczynników, których oceny wyznacza się minimalizując kryterium

$$\sum_{t=0}^N \|y_t - a^T g(\varphi_t)\|^2, \quad (2.20)$$

gdzie $a = [a^{(1)} \dots a^{(r)}]^T$, a $g(\varphi_t) = [g_1(\varphi_t) \dots g_r(\varphi_t)]^T$.

Klasa modeli	$g_j(\varphi_t, \beta_j, \gamma_j)$
Szereg trygonometryczny	$\sin(\beta_j^T \varphi_t + \gamma_j)$
Regresja liniowa	$\varphi_t^{(j)}$
Sieci neuronowe	$\sigma(\beta_j^T \varphi_t + \gamma_j),$ gdzie $\sigma(u) = 1/(1 - e^{-u})$
Sieci neuronowe o podstawie radialnej (RBFN*)	$\kappa(\ \varphi_t - \gamma_j\ _{\beta_j}),$ gdzie $\kappa(u) = \exp(-u^2)$

Tab. 2.1. Przykłady struktur modeli objęte zależnością (2.19b)

* ang. Radial Basis Function Network

Zdolności aproksymacyjne powyższego modelu zależą od doboru funkcji bazowych $\{g_j(\cdot)\}$ oraz od liczby parametrów określanej również mianem *rzędu modelu*, czyli $\dim(a) = r$. Omówiona klasa modeli zawiera wiele szczególnych przypadków struktur modeli zarówno liniowych jak i nieliniowych. W pracy będą rozpatrywane klasy liniowe względem parametrów o założonej z góry regularności w sensie omówionym w niniejszym rozdziale.

2.2.2. Modele liniowe

Identyfikacja modeli liniowych jest opisywana bardzo obszernie w literaturze [69, 44, 53, 77, 18, 9]. Wszystkie stacjonarne modele liniowe można ująć za pomocą następującej struktury modelu

$$A(q)y_t = \frac{B(q)}{F(q)}u_{t-nk} + \frac{C(q)}{D(q)}e_t, \quad (2.21)$$

gdzie $nk > 0$ jest opóźnieniem systemu, natomiast A, B, C, D, F są wielomianami operatora opóźnienia q ($q^{-1}u_t = u_{t-1}$) danymi poniżej

$$\begin{aligned} A(q) &= 1 + a_1q^{-1} + \dots + a_{na}q^{-na}, \\ B(q) &= b_1 + b_2q^{-1} + \dots + b_{nb}q^{-nb+1}, \\ C(q) &= 1 + c_1q^{-1} + \dots + c_{nc}q^{-nc}, \\ D(q) &= 1 + d_1q^{-1} + \dots + d_{nd}q^{-nd}, \\ F(q) &= 1 + f_1q^{-1} + \dots + f_{nf}q^{-nf}. \end{aligned} \quad (2.22)$$

Ciąg $\{e_t\}$ jest sekwencją niezależnych zmiennych losowych o jednakowych rozkładach prawdopodobieństwa oraz spełniający warunki $Ee_t = 0$, $Var(e_t) = \lambda < \infty$. Zapis operatorowy przy pomocy wielomianów zmiennej q jest wygodną i zwartą reprezentacją równań różnicowych – alternatywnej formy opisu systemów liniowych [2]. Zauważmy, że zakłócenie v_t w ogólnej zależności (2.6) jest reprezentowane jako

$$v_t = \frac{C(q)}{D(q)}e_t, \quad (2.23)$$

a więc jako wynik przekształcenia *białego szumu* e_t przez dowolny układ liniowy o transmitancji dyskretniej $C(q)/D(q)$. Klasa modeli (2.22) jest bardzo ogólna. Znaczenie praktyczne mają jej szczególne przypadki, z których najważniejsze to: model ARX (A, B), ARMA (A, C), model błędu wyjściowego (B, F) i model Boxa-Jenkinsa (B, F, C, D). W nawiasach podano wielomiany użyte w danym modelu. Model ARX można również wyrazić w formie regresji liniowej (2.26), co umożliwia zastosowanie wielu znanych metod estymacji parametrów.

2.3. Dyskusja przypadków doboru modelu i zbieżność

W klasycznym zadaniu identyfikacji parametrycznej struktura modelu, a więc i liczba nieznanych parametrów zostaje arbitralnie ustalona przed eksperymentem. Założenie, że postać modelu znana jest przed doświadczeniem, a specyfikacja modelu jest dokładna, ma poważne ograniczenia. Często nasza wiedza o obiekcie przed eksperymentem jest niewielka i pozwala co najwyżej na sformułowanie jakościowe pewnych ogólnych cech charakterystyki obiektu. Elastyczność modelu sformułowanego w postaci (2.20) zależy od typu funkcji bazowych $\{g_j\}_{j=1}^r$ oraz ich ilości r . Problem pracy dotyczy szczególnego przypadku wspomnianego modelu, w którym ciąg funkcji $\{g_j\}_{j=1}^r$ jest znany, natomiast liczba funkcji bazowych, a zatem i rozmiar wektora parametrów modelu może ulegać zmianie w trakcie identyfikacji. W celu formalnego opisu tego zagadnienia wprowadzimy nieco ściślej pojęcie *rodziny struktur modeli*, rozpatrzmy także przypadki doboru postaci modelu i określimy wymagania zbieżności odpowiadających im estymatorów parametrów.

Def. 2.1. Niech r oznacza rząd modelu, a N zbiór liczb naturalnych. *Rodzinę klas modeli* M^* określamy jako zbiór parametrycznych postaci modeli indeksowany liczbą nieznanych parametrów, co zapisujemy następująco:

$$M^* = \{g_r(\varphi_t, a_r) : r \in N\}. \quad (2.24)$$

Def. 2.2. Niech D_r oznacza zbiór wartości, jakie może przyjmować wektor nieznanych parametrów. Strukturą modelu rzędu r , którą oznaczamy przez M_r , nazywamy odwzorowanie D_r w zbiór wartości modelu

$$M_r : D_r \ni a_r \rightarrow M_r(a_r) = g_r(\varphi_t, a_r) \in M^*. \quad (2.25)$$

Def. 2.3. Mówimy, że struktura M_i zawiera się w strukturze M_j ($i < j$), co zapisujemy: $M_i \subset M_j$, jeśli $D_i \subset D_j$ i odwzorowanie M_i jest otrzymywane z M_j przez ograniczenie $a_j \subset D_i$.

W skrajnym przypadku można rozumieć M_i jako otrzymane z M_j przez wyzerowanie wszystkich składowych wektora parametrów $a_j = [a^{(1)} \dots a^{(j)}]^T$, których indeks jest większy niż i , tzn. $a^{(l)} = 0$ dla $l > i$.

Przykład 3.1.

Rozważmy rodzinę klas modeli liniowych

$$L^* = \{g_r(u_t, a_r) = a_r^T \varphi(u_t), r \in N\}. \quad (2.26)$$

Założmy, że u_t jest wielkością skalarną, $a_j = [a^{(1)} \dots a^{(j)}]^T$, $\varphi(u_t) = [\varphi^{(1)}(u_t) \dots \varphi^{(r)}(u_t)]^T$, funkcje bazowe niech będą postaci jednomianów o coraz wyższych stopniach $\varphi^{(k)}(u_t) = u_t^{k-1}$, gdzie $k = 1, \dots, r$. Zwróćmy uwagę, że poszczególne struktury modelu są liniowe względem parametru, lecz nie koniecznie względem zmiennej wejściowej u_t . Pierwsze trzy struktury mają następującą postać:

$$\begin{aligned} L_1(a_1) &= a^{(1)}, \\ L_2(a_2) &= a^{(1)} + a^{(2)}u_t, \\ L_3(a_3) &= a^{(1)} + a^{(2)}u_t + a^{(3)}u_t^2. \end{aligned}$$

W przytoczonym przykładzie zachodzi $L_i \subset L_{i+1}$ dla dowolnego $i < r$. Wobec liniowej niezależności kolejno dodawanych funkcji jednomianowych coraz wyższych stopni zauważmy, że daną rodzinę możemy uważać jako superpozycję kolejnych struktur modeli

$$L^* = \bigcup_{k=1}^r L_k. \quad (2.27)$$

Po wprowadzeniu potrzebnych określeń sformułujemy podstawowe postacie algorytmów identyfikacji ze zmienną liczbą parametrów oraz wymagania asymptotyczne, jakie te algorytmy powinny spełniać. Przyjmując założenie

o istnieniu idealnej charakterystyki obiektu g^0 możemy wyróżnić następujące przypadki.

1. Obiekt należy do rodziny klas modeli $g^0 \in M^*$, a rząd modelu r jest a priori znany.

Rodzina struktur M^* została tak dobrana, że zawiera teoretyczną charakterystykę obiektu. W tym przypadku istnieje taka wartość parametru a_r^0 , dla której model idealnie przybliża teoretyczną charakterystykę obiektu, co można zapisać formalnie w następujący sposób:

$$\exists a_r^0 \forall \varphi_t \ g_r(\varphi_t, a_r^0) = g^0(\varphi_t). \quad (2.28)$$

Spodziewamy się zgodnego oszacowania parametru modelu:

$$\lim_{n \rightarrow \infty} \hat{a}_{r,n} = a_r^0, \quad (2.29)$$

przy czym $\hat{a}_{r,n}$ jest optymalną, w przyjętym sensie, oceną nieznanego parametru. Estymator parametru w (2.29) jest indeksowany dwiema zmiennymi r i n , z których pierwsza oznacza ilość składowych wektora parametrów, a druga jest bieżącą ilością obserwacji, na podstawie których wyznaczono estymator. Opisana sytuacja jest daleką idealizacją i ma małe znaczenie praktyczne, ponieważ na ogół nie znamy dokładnie klasy modeli, do której należy dany obiekt.

2. Obiekt należy do rodziny struktur $g^0 \in M^*$, a rząd modelu r jest ustalony arbitralnie przed eksperymentem.

W tym przypadku istnieje optymalnie dobrany w sensie kryterium (2.17) parametr a_r^* taki, że struktura (2.14) przybliża idealną charakterystykę obiektu:

$$\exists a_r^* \forall \varphi_t \ g_r(\varphi_t, a_r^*) \approx g^0(\varphi_t). \quad (2.30)$$

Znak równości przybliżonej wyraża fakt, iż nawet w granicznym przypadku przy bardzo dużej liczbie obserwacji postać modelu będzie się różnić od teoretycznej charakterystyki obiektu g^0 . Błąd przybliżenia charakterystyki założoną postacią modelu wynika z nieprawidłowego doboru struktury modelu. Błąd ten można nazwać, nadużywając nieco terminologii statystycznej, błędem obciążenia. W praktyce aby zmniejszyć wpływ takiego błędu na jakość otrzymywanego

W praktyce aby zmniejszyć wpływ takiego błędu na jakość otrzymywanego modelu, trzeba poszerzyć rodzinę klas o dodatkowe struktury. Przy dostatecznie dużej liczbie obserwacji otrzymamy obciążone oszacowanie nieznanego parametru:

$$\lim_{n \rightarrow \infty} \hat{a}_{r,n} = a_r^*. \quad (2.31)$$

3. Obiekt należy do rodziny struktur $g^0 \in M^*$, natomiast r nie jest znane przed eksperymentem.

W ogólnym przypadku obiekt może być tak złożony, że jego charakterystyki nie da się zamodelować przy pomocy skończonej liczby parametrów. Klasa modeli może być wówczas rozumiana jako złożenie postaci modeli należących do M^*

$$g^0(\varphi_t) = \bigcup_{r=1}^{\infty} M_r, \quad (2.32)$$

gdzie $M_r = g_r(\varphi_t, a_r) \in M^*$. Nieznajomość postaci modelu przed eksperymentem skłania nas do stopniowego zwiększania komplikacji struktury modelu w miarę pozyskiwania wiedzy o obiekcie. Ze wzrostem ilości obserwacji n może następować wzrost komplikacji struktury modelu wyrażający się wzrostem ilości jego parametrów. W konsekwencji może ulegać zmianie struktura modelu g_r . Rząd modelu jest pewną funkcją n . Zakładamy że $r(n)$ zdąża dostatecznie wolno do nieskończoności przy wzroście liczby obserwacji n , a postać modelu w granicy dąży do idealnej charakterystyki obiektu. Zapiszmy formalnie powyższe wymagania:

$$\lim_{n \rightarrow \infty} g_{r(n)}(\varphi_t, \hat{a}_{r(n),n}) = g^0(\varphi_t), \quad (2.33)$$

$$\lim_{n \rightarrow \infty} r(n) = \infty. \quad (2.34)$$

Funkcja $r(n)$ nie może być podana w sposób jawny; jest uwarunkowana napływającymi obserwacjami. Mamy tu analogię z *metodami nieparametrycznymi* [26], w których wobec dużego stopnia nieznajomości systemu konstruujemy np. estymator gęstości prawdopodobieństwa opierając się na serii pomiarowej, i których jakość jest również w dużym stopniu uzależniona od napływających danych.

Podstawowy problem pracy koncentruje się głównie na ostatnim z omówionych przypadków. Dwa poprzednie można uważać za jego szczególne przypadki. Opierając się na powyższych rozważaniach możemy w sposób opisowy przedstawić ogólne strategie algorytmów analizowanych w pracy. Startując od pewnego prostego modelu wyjściowego estymujemy jego parametry w miarę przychodzenia obserwacji i staramy się ocenić na bieżąco na ile nasz model odzwierciedla charakterystykę obiektu. Jeśli przyjęty model nie jest zadowalający, wówczas bieżąca struktura jest przebudowywana na bardziej złożoną oraz dokonywana jest korekta bieżącej oceny parametrów. Ten proces w miarę potrzeby jest powtarzany sekwencyjnie, dopóki nie uzyskamy zadowalającego przybliżenia. Można podać kilka motywacji dla omówionego postępowania. Po pierwsze, wobec niewielkiej wiedzy o modelowanym systemie trudno jest z góry wyznaczyć właściwą strukturę. Zbyt prosty model może być niedokładny. Zbyt skomplikowany, z kolei, może powodować złe uwarunkowanie algorytmu obliczeniowego oraz większą wariancję błędu [63, 44], co analizujemy w rozdziale 3. Po drugie, zgodnie z *zasadą oszczędności* [68], w procedurze identyfikacji należy najpierw stosować najprostszą z możliwych struktur modelu, w szczególności gdy nie posiadamy wiedzy wstępnej o systemie. Po trzecie, zastosowanie jednoczesnej aktualizacji wartości parametrów i postaci modelu powinno wykazywać w wielu przypadkach lepsze własności śledzące w rekurencyjnych algorytmach identyfikacji.

2.3.1. Interpretacja zbieżności

Rozpatrzmy najpierw idealny przypadek deterministyczny, gdy na układ nie działają zakłócenia. Wówczas niedokładność modelu jest spowodowana nieprawidłowym doбором struktury modelu. Potraktujmy postacie modeli w zbiorze (2.24) jako ciąg funkcyjny

$$\{\hat{g}_{r(n)}(\varphi_t, \hat{a}_{r,n})\}. \quad (2.35)$$

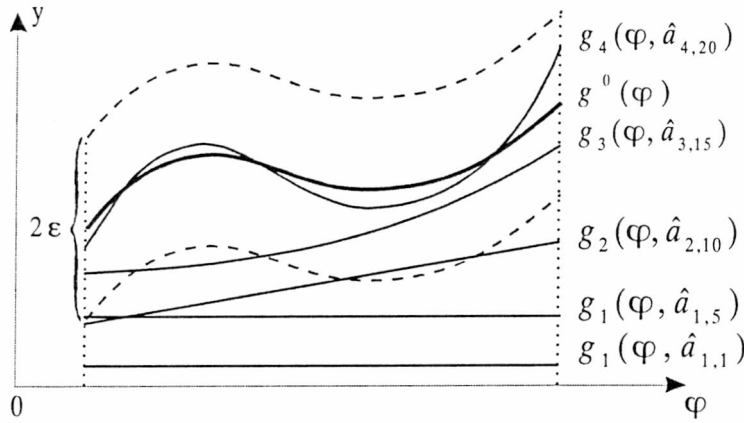
Wymagamy, aby ten ciąg postaci modeli był zbieżny jednostajnie do teoretycznej charakterystyki obiektu g^0 . Zachodzi więc następujący warunek:

$$\begin{aligned} \hat{g}_{r(n)}(\varphi_t, \hat{a}_{r,n}) \rightarrow g^0(\varphi_t) \Leftrightarrow \\ \forall \varepsilon > 0 \exists \delta \forall \varphi_t \exists n > \delta \left| \hat{g}_{r(n)}(\varphi_t, \hat{a}_{r(n),n}) - g^0(\varphi_t) \right| < \varepsilon \end{aligned} \quad (2.36)$$

Z warunku (2.36) wynika, że dla każdej niewielkiej liczby ε istnieje liczba δ o pożądanych własnościach wspólna dla całego zbioru. Jeśli tak, to i dla każdego $\varphi_t, t = 1, \dots, n$ także istnieje taka liczba.

Z warunku (2.36) wynika, że dla każdej niewielkiej liczby ε istnieje liczba δ o pożądanych własnościach wspólna dla całego zbioru. Jeśli tak, to i dla każdego $\varphi_t, t = 1, \dots, n$ także istnieje taka liczba.

Geometryczną interpretację zbieżności jednostajnej modelu przedstawiono na rys. 2.2. Pogrubiona linia reprezentuje teoretyczną charakterystykę systemu



Rys. 2.2. Interpretacja geometryczna zbieżności modelu

biegnącą pośrodku epsilonowego obszaru, którego granice zaznaczono dwoma przerywanymi liniami. Cienkie ciągłe linie odpowiadają kilku wybranym wyrazom ciągu (2.35). W omawianym deterministycznym przypadku poszczególne oceny modeli nie wykraczają, po odpowiednio dużej liczbie obserwacji, poza epsilonowy obszar określający margines dokładności modelu. Własność (2.36) gwarantuje spełnienie warunku Cauchy'ego dla każdego φ_t z rozpatrywanego przedziału i dla każdych dwóch liczb naturalnych $r(n), s(n)$ spełniających warunek $\min(r(n), s(n)) > \delta$

$$\left| \hat{g}_{r(n)}(\varphi_t, \hat{a}_{r(n),n}) - \hat{g}_{s(n)}(\varphi_t, \hat{a}_{s(n),n}) \right| < \varepsilon. \quad (2.37)$$

W szczególności nierówność (2.37) zachodzi, gdy $r(n)$, $s(n)$ są sąsiednimi liczbami naturalnymi. W dalszej części pracy często będziemy upraszczać zapis, pomijając uzależnienie r od n . Zakładamy ponadto, że w kryterium (2.37) indeksy funkcji nie będą się różnić o więcej niż 1, tzn. $|s - r| \leq 1$. Różnica w (2.37) jest miarą bezwzględną zbieżności ciągu (2.35). Możemy, uwzględniając powyższe uwagi, sformułować w oparciu o nierówność (2.37) względne kryterium dokładności modelu w momencie t dla przypadku deterministycznego.

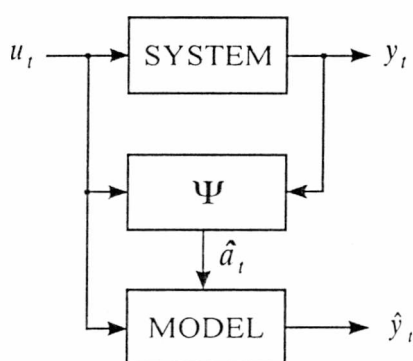
$$\left| g_r(\varphi_t, \hat{a}_{r,t}) - g^0(\varphi_t) \right| < \varepsilon \left| g^0(\varphi_t) \right|. \quad (2.38)$$

Nierówność (2.38) dotyczy idealnego, deterministycznego przypadku i ma charakter punktowy, tzn. wyznaczana jest dla danego φ_t . Wielkość progowa ε jest zadany parametrem i ma istotny wpływ na działanie algorytmu identyfikacji. Wykorzystanie (2.38) w charakterze warunku logicznego determinującego przebudowę modelu zawarto w rozdziale trzecim.

3. Algorytmy o strojonej strukturze modelu

3.1. Wprowadzenie

W wielu sytuacjach spotykanych w praktyce model systemu musi być wyznaczany on-line w trakcie normalnego funkcjonowania systemu. Jest tak wówczas, gdy decyzję o stanie systemu trzeba podejmować na bieżąco, w czasie rzeczywistym. W sterowaniu komputerowym jest to decyzja, jaką wartość sygnału zadać na wejściu w kolejnym takcie próbkowania. Problem filtracji wymaga dostrojenia parametrów filtru w sposób zapewniający pożądane własności sygnału wyjściowego. Prognoza pewnych wielkości charakteryzujących stan pacjenta wspomaga efektywną decyzję o dalszej terapii. Z kolei w systemach nadzoru i monitorujących podejmuje się decyzję, czy wystąpił błąd w systemie, a jeśli tak, to jakiego rodzaju. Powyższa lista problemów nie



Rys. 3.1. Ogólna struktura adaptacyjnego układu identyfikacji

jest wyczerpująca, można ją rozszerzać o przykłady z innych dziedzin. Wspólną cechą wymienionych zagadnień technicznych jest stosowanie modeli matematycznych identyfikowanych w sposób adaptacyjny.

Rekurencyjne algorytmy identyfikacji, zwane także algorytmami adaptacyjnymi, strojonego modelu lub estymacji sekwencyjnej mają ogólną strukturę przedstawioną na rys. 3.1. Ψ oznacza algorytm identyfikacji, który pozwala na wyznaczenie wartości estymatora nieznanego parametru na podstawie poprzednio wyznaczonego oszacowania oraz aktualnych pomiarów wejścia i wyjścia obiektu oznaczonych odpowiednio jako u_t i y_t . Ogólna formuła algorytmu z rys. 3.1 ma następującą postać

$$\hat{a}_{t+1} = \Psi(\hat{a}_t, u_{t+1}, y_{t+1}). \quad (3.1)$$

Startując od pewnej arbitralnej oceny początkowej wektora parametrów $\hat{a}_0$ wyznacza się kolejne $\hat{a}_t$, $t = 1, 2, \dots$ na podstawie pomiarów wejścia i wyjścia obiektu w kolejnych chwilach czasu.

Algorytmy rekurencyjne otrzymuje się różnymi metodami [50], wśród których najważniejszymi są:

- Modyfikacja algorytmów bezpośrednich
- Filtracja nieliniowa (podejście bayesowskie)
- Zastosowanie podejścia zwanego aproksymacją stochastyczną
- Techniki oparte na modelu odniesienia

Metody rekurencyjne są bardzo szeroko opisywane w literaturze [2, 24, 50, 61, 53, 69]. Tematyka ta jest na tyle szeroka i ważna, że poświęcono jej oddzielne monografie [50, 5] jak również wiele artykułów przeglądowych [28, 48]. W odróżnieniu od algorytmów bezpośrednich, w których nieznane parametry modelu wyznacza się po wykonaniu całej serii pomiarów, algorytmy rekurencyjne pozwalają na wyznaczanie kolejnych oszacowań parametrów w miarę przychodzenia kolejnych, skąd wynika ich duże znaczenie praktyczne. W przypadku gdy prognoza wielkości wyjściowej systemu musi być wyznaczona na bieżąco, z aktualizacją stanu modelu w kolejnych momentach czasu, należy zastosować pewną wersję algorytmu (3.1).

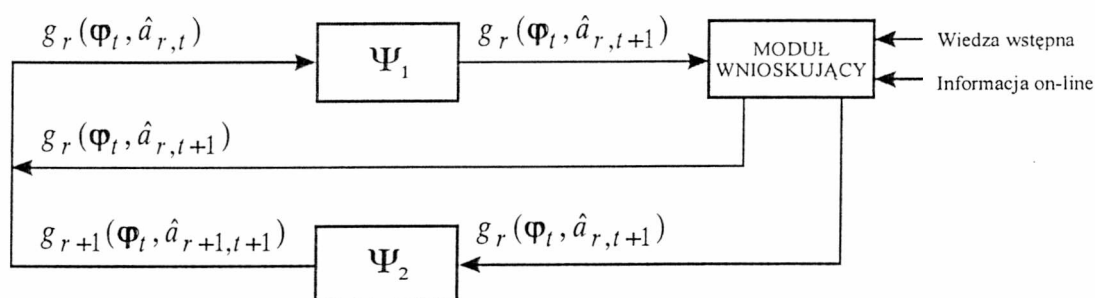
Główne zastosowania rekurencyjnych metod identyfikacji obejmują sterowanie adaptacyjne [2, 56], przetwarzanie sygnałów [1], detekcję uszkodzeń i gwałtownych zmian sygnału [3].

Algorytmy w postaci (3.1) są w swej istocie swoistymi metodami kompresji danych. Do wyznaczenia kolejnej oceny parametru nie jest na ogół potrzebne pamiętanie całej serii pomiarów, a jedynie pewnych wielkości reprezentujących bieżący stan modelu, których nie uwzględniono explicite w zależności 3.1 dla uproszczenia notacji. Większość algorytmów rekurencyjnych wymaga stałej wielkości pamięci w każdej iteracji, mimo że długość ciągu uczącego wzrasta z nadejściem kolejnych danych pomiarowych. Ten ważny aspekt praktyczny powoduje, że metody sekwencyjne bywają stosowane także w wypadkach, gdzie wystarczyłyby metody *off-line*.

Rekurencyjne algorytmy identyfikacji rozpatrywano w literaturze w wielu wariantach oraz w systematyczny ujednolicony sposób [50]. Ich wspólną cechą jest a priori ustalona struktura modelu. W tym rozdziale zbadamy pewną klasę metod rekurencyjnych, które łączą zdolność adaptacyjnego dostrajania parametrów z możliwością przebudowy struktury modelu w trakcie procedury identyfikacji. Stawiamy tezę, że zaproponowane algorytmy mają większą elastyczność w dostrajaniu się do zmian charakterystyki obiektu. Po drugie algorytmy te mają własności zbliżone do nieparametrycznych procedur identyfikacji, o ile przyjęta rodzina klas modeli jest na tyle uniwersalna, że pozwala na modelowanie szerokiego spektrum obiektów. Nazwa: *złożone algorytmy rekurencyjne* odnosi się do wieloetapowego charakteru procedur, co zostanie omówione bardziej szczegółowo w następnych rozdziałach.

3.2. Podstawowa wersja algorytmu

W niniejszym podrozdziale zaproponowano strukturę rekurencyjnego algorytmu identyfikacji, w którym powiązано etap estymacji parametrów i przebudowy struktury modelu. W klasycznych algorytmach wyboru struktury modelu dokonuje się a priori i nie ulega ona zmianie w czasie eksperymentu. Obecnie dopuszczamy możliwość zmiany postaci modelu bez konieczności powtarzania od nowa całej procedury



Rys. 3.2. Schemat blokowy podstawowej wersji algorytmu

identyfikacji. Wersja podstawowa złożonego algorytmu identyfikacji przedstawiona jest na rys. 3.2. Struktura algorytmu zawiera dwie pętle i związane z nimi dwa główne człony algorytmu Ψ_1 i Ψ_2 . Wewnętrzna pętla służy do wyznaczania kolejnych ocen wektora parametrów za pomocą pewnego sekwencyjnego algorytmu estymacji Ψ_1 , co zapisujemy w postaci ogólnej formuły

$$\hat{a}_{r,t+1} = \Psi_1(\hat{a}_{r,t}, \varphi_{r,t+1}), \quad (3.2)$$

gdzie

$$\varphi_{r,t+1} = [\varphi_{t+1}^{(1)} \dots \varphi_{t+1}^{(r)}]^T \quad (3.3)$$

jest wartością wektora regresji po nadejściu nowej obserwacji. Pętla wewnętrzna z procedurą Ψ_1 wykonuje się po każdym pomiarze wartości wejścia i wyjścia. Pętla zewnętrzna, wykonująca się rzadziej i nie koniecznie regularnie, służy do przebudowy struktury modelu, realizowanej przez ogólną procedurę Ψ_2 o postaci

$$\hat{a}_{r+1,t} = \Psi_2(\hat{a}_{r,t}, \phi_{r+1,t}), \quad (3.4)$$

gdzie

$$\phi_{r+1,t} = [\varphi_1^{(r+1)} \dots \varphi_t^{(r+1)}] \quad (3.5)$$

jest wektorem wierszowym odpowiadającym dodaniu nowej składowej wektora regresji $\varphi^{(r+1)}$. Oprócz poszerzenia wektora parametrów modelu jest wektorem wierszowym odpowiadającym dodaniu nowej składowej wektora regresji $\varphi^{(r+1)}$. Oprócz poszerzenia wektora parametrów modelu w członie Ψ_2 dokonywana jest odpowiednia korekcja jego wszystkich składowych. Widoczny na rys. 3.2 moduł wnioskujący podejmuje decyzję, czy struktura modelu powinna zostać przebudowana, czy też można poprzestać na skorygowaniu oceny parametrów bez wchodzenia w zewnętrzną pętlę algorytmu. Moduł ten ma postać predykatu, którego wartość logiczna jest uzależniona od oceny bieżącej jakości modelu. Ocena ta jest wyznaczana w oparciu o dane pomiarowe uzyskane do momentu t . Wiedza wstępna może być użyta do ustalenia warunków początkowych algorytmu oraz wyboru funkcji bazowych, na podstawie których obliczane są składowe wektora (3.3). Działanie algorytmu z rys. 3.2 istotnie zależy od sformułowania warunku decydującego o przebudowie postaci modelu. W kolejnym podrozdziale analizujemy kryteria determinujące przebudowę struktury modelu.

3.3. Kryteria rekonstrukcji struktury modelu

W najprostszym, deterministycznym przypadku, proponujemy zastosowanie kryterium opartego na nierówności (2.38) sformułowanej w oparciu o pojęcie zbieżności jednostajnej. Gdy ta nierówność jest spełniona, oznacza to, że dokładność modelu mieści się w dopuszczalnych granicach określonych wielkością ε i algorytm identyfikacji pozostaje w swojej wewnętrznej pętli, działając jak klasyczna procedura rekurencyjna. Niespełnienie (3.28) oznacza konieczność przebudowy struktury modelu realizowanej przez procedurę (3.4) w zewnętrznej pętli. W warunku (2.38) wartość progowa ε jest zadawanym parametrem liczbowym decydującym o jakości algorytmu. Zbyt duża wartość ε będzie powodować małą wrażliwość na punktowe różnice między wyjściem modelu i obiektu. Odwrotnie, mała wartość ε da dużą czułość algorytmu na punktowe niedokładności. Widzimy więc, że przy takim sformułowaniu warunkowego kryterium przebudowy struktury modelu cała procedura byłaby wrażliwa na zakłócenia. Zastosowanie (2.38) jest ograniczone do warunków deterministycznych lub do przypadku, w którym poziom zakłóceń jest znacznie mniejszy niż wartość parametru ε . Formalnie oznacza to, że w zależności (2.7) mamy

$$|v_t| \leq C_v, \quad (3.6)$$

$$C_v \ll \varepsilon,$$

gdzie C_v jest pewną stałą. Ostatecznie kryterium oparte na nierówności (2.38) można zapisać w następujący sposób:

$$|y_t - \hat{y}_t| < \varepsilon |y_t|. \quad (3.7)$$

W przypadku, gdy w systemie występują zakłócenia losowe, ocena jakości zastosowanego modelu jest szczególnie utrudniona, gdyż na rozbieżności między wyjściem obiektu i systemu ma wpływ zarówno dobór struktury modelu jak i zakłócenia losowe. W warunkach probabilistycznych jako miarę służącą do oceny jakości modeli stosowana jest przeciętna wartość odchylenia wyjścia modelu i obiektu:

$$\begin{aligned} Q(a_r) &= E\|y_t - g_r(\varphi_t, a_r)\|^2 = \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \|g^0(\varphi_t) - g_r(\varphi_t, a_r)\|^2. \end{aligned} \quad (3.8)$$

Zakładamy, że φ_t jest sygnałem deterministycznym lub stacjonarnym procesem losowym oraz przyjmujemy, że zakłócenia w (2.7) mają zerową średnią i skończoną wariancję tzn.

$$Ev_t = 0, \quad (3.9a)$$

$$Var(v_t) = \lambda < \infty. \quad (3.9b)$$

Przy założeniu, że granica (3.8) istnieje możemy zapisać to wyrażenie w postaci:

$$Q(a_r) = \lambda + E\|g^0(\varphi_t) - g_r(\varphi_t, a_r)\|^2. \quad (3.10)$$

Zależność (3.10) pokazuje, że wartość odchylenia zależy od postaci modelu g_r jak również od wartości parametrów a_r . Konkretna wartość a_r jest estymowana na podstawie pomiarów, a więc jest zmienną losową $\hat{a}_{r,n}$. Za miarę dokładności modelu przyjmujemy wielkość:

$$Q_r^* = E_{\hat{a}_{r,n}} Q(\hat{a}_{r,n}), \quad (3.11)$$

którą możemy rozwinąć otrzymując wyrażenie:

$$Q_r^* \approx \lambda + E\|g^0(\varphi_t) - g_r(\varphi_t, a_r^*)\|^2 + E\|g_r(\varphi_t, a_r^*) - g_r(\varphi_t, \hat{a}_{r,n})\|^2, \quad (3.12)$$

gdzie a_r^* zostało dane wzorem (2.31). Miara jakości (3.12) zawiera trzy zasadnicze składniki:

$$Q_r^* = \lambda + Q_B^* + Q_R^*. \quad (3.13)$$

Pierwszy z nich zdeterminowany jest właściwościami niemierzalnych zakłóceń losowych, na które nie mamy wpływu. Składnik Q_B^* zależy od doboru struktury modelu i można by go nazwać, nadużywając terminologii statystycznej, błędem obciążenia. Ostatni składnik Q_R^* opisuje błąd losowy. Im bardziej skomplikowana

postać modelu, tym mniejszy błąd obciążenia, ale zarazem tym większa wariancja estymatora. W pracy [44] wykazano, że składnik wariancji jest w przybliżeniu proporcjonalny do rzędu modelu i jednocześnie odwrotnie proporcjonalny do liczby obserwacji n

$$E\left\|g_r(\varphi_t, a_r^*) - g_r(\varphi_t, \hat{a}_{r,n})\right\|^2 \approx \lambda \frac{r}{n}. \quad (3.14)$$

Powyższa dyskusja stanowi podstawę do pewnych wniosków natury heurystycznej, które jednocześnie stanowią motywację do badań rozważanych w pracy algorytmów. Mianowicie, nawet w przypadku dostatecznie dużej liczby pomiarów rozbieżność między wyjściem obiektu i modelu może być spowodowana przez nieodpowiedni dobór struktury modelu. Można więc mieć nadzieję, że poprawę jakości modelu uzyskamy stosując bardziej skomplikowany model. Jeśli niedobór strukturalny modelu został trafnie oceniony, to poprawa jakości powinna być zauważalna, objawiając się zmniejszeniem składnika obciążenia. Jeśli natomiast model został rozbudowany „na wyrost”, to składnik obciążenia nie zmniejszy się istotnie, co sugerowałoby że rekonstrukcja modelu okazała się niepotrzebna, nastąpiło przeparametryzowanie objawiające się zwiększeniem wariancji zgodnie z równaniem (3.14).

W przypadku gdy w systemie występują zakłócenia losowe, kryterium (3.7) jest niewystarczające i dlatego sięga się do statystycznych miar jakości modelu. Statystyki te są na ogół funkcjami błędu resztkowego zdefiniowanego jako:

$$\varepsilon_t = y_t - \hat{y}_t \quad (3.15)$$

Wymieńmy najczęściej używane statystyki do oceny jakości modelu.

- Maksymalna wartość bezwzględna błędu resztkowego modelu

$$M_t^\varepsilon = \max_{1 \leq k \leq t} |\varepsilon_k|. \quad (3.16)$$

- Wartość średnia błędu resztkowego

$$m_t^\varepsilon = \frac{1}{t} \sum_{k=1}^t \varepsilon_k. \quad (3.17)$$

3.3. Kryteria rekonstrukcji struktury modelu

- Wariancja

$$V_t^\varepsilon = \frac{1}{t-1} \sum_{k=1}^t (\varepsilon_k - m_t^\varepsilon)^2. \quad (3.18)$$

- Miara średniokwadratowa

$$S_t^\varepsilon = \frac{1}{t} \sum_{k=1}^t \varepsilon_k^2 = (m_t^\varepsilon)^2 + \frac{t-1}{t} V_t^\varepsilon. \quad (3.19)$$

Wielkości te mogą być wyznaczone w sposób sekwencyjny, czyli do ich obliczenia wystarczy aktualna wartość i bieżące oszacowanie błędu resztkowego. Rekurencyjny odpowiednik (3.16) ma postać

$$M_{t+1}^\varepsilon = \max(M_t^\varepsilon, \varepsilon_{t+1}). \quad (3.20)$$

Sekwencyjne wersje statystyk (3.17), (3.18), (3.19) otrzymuje się po wykonaniu przekształceń algebraicznych. Dla średniej (3.17) mamy

$$m_{t+1}^\varepsilon = m_t^\varepsilon + \frac{1}{t+1} (\varepsilon_{t+1} - m_t^\varepsilon). \quad (3.21)$$

Wariancja (3.18) w wersji rekurencyjnej ma postać

$$V_{t+1}^\varepsilon = \frac{t-1}{t} V_t^\varepsilon + \frac{1}{t+1} (\varepsilon_{t+1} - m_t^\varepsilon)^2. \quad (3.22)$$

Zależność (3.22) otrzymujemy z (3.18)

$$V_{t+1}^\varepsilon = \frac{1}{t} \sum_{k=1}^t (\varepsilon_k - m_{t+1}^\varepsilon)^2 + \frac{1}{t} (\varepsilon_{t+1} - m_{t+1}^\varepsilon)^2. \quad (3.23)$$

Podstawiając do (3.23) m_{t+1}^ε wyznaczoną z wzoru (3.21) otrzymujemy

$$\begin{aligned} V_{t+1}^\varepsilon &= \frac{1}{t} \sum_{k=1}^t (\varepsilon_k - m_t^\varepsilon)^2 - \frac{2}{t(t+1)} (\varepsilon_{t+1} - m_t^\varepsilon) \sum_{k=1}^t (\varepsilon_k - m_t^\varepsilon) + \\ &+ \frac{1}{(t+1)^2} \sum_{k=1}^t (\varepsilon_{t+1} - m_t^\varepsilon)^2 + \frac{t}{(t+1)^2} \sum_{k=1}^t (\varepsilon_{t+1} - m_t^\varepsilon)^2. \end{aligned}$$

Zauważając, że drugi składnik w powyższym wyrażeniu jest równy zero otrzymujemy po prostych przekształceniach (3.23). W podobny sposób możemy otrzymać wyrażenie dla miary średniokwadratowej

$$S_{t+1}^{\varepsilon} = S_t^{\varepsilon} + \frac{1}{t+1}(\varepsilon_{t+1}^2 - S_t^{\varepsilon}) \quad (3.24)$$

lub wyrażając przy pomocy (3.20) i (3.21)

$$S_{t+1}^{\varepsilon} = (m_{t+1}^{\varepsilon})^2 + \frac{t}{t+1}V_{t+1}^{\varepsilon}. \quad (3.25)$$

W kolejnym rozdziale wyprowadzimy konkretną postać algorytmu dla modelu w postaci regresji liniowej.

3.4. Przypadek szczególny dla modelu regresji liniowej

Rozważmy szczególny przypadek klasy modeli (2.14) w postaci kombinacji liniowej pewnych funkcji wejścia i wyjścia systemu

$$y_t = a_r^T \varphi_{r,t}, \quad (3.25)$$

gdzie a_r jest opisane przez (2.15), a $\varphi_{r,t}$ jest wektorem regresji zdefiniowanym przez (2.12). W wektorze nieznanych parametrów i regresji wprowadzono dodatkowy indeks r oznaczający rząd modelu, który jest zmienną algorytmu. Składowe wektora (2.12) nazywamy *funkcjami bazowymi* lub *rozpinającymi regresję*. Są to funkcje, których ogólną postać dano wzorami (2.13). Kryterium średniokwadratowe (2.17) dla tego szczególnego przypadku można zapisać jako:

$$Q(S_N, a_r) = (Y_t - a_r^T \Phi_{r,t})(Y_t - a_r^T \Phi_{r,t})^T, \quad (3.26)$$

gdzie $\Phi_{r,t}$ jest macierzą wyznaczoną na podstawie obserwacji pomiarowych

Do wyznaczenia estymatora parametrów modelu na podstawie pomiarów w momentach od 1 do t stosujemy metodę najmniejszych kwadratów [9] uzyskując następujące oszacowanie:

3.4. Przypadek szczególny dla modelu regresji liniowej

$$\Phi_{r,t} = \begin{bmatrix} \varphi_1^{(1)} & \cdots & \varphi_t^{(1)} \\ \vdots & \ddots & \vdots \\ \varphi_1^{(r)} & \cdots & \varphi_t^{(r)} \end{bmatrix}. \quad (3.27)$$

Do wyznaczenia estymatora parametrów modelu na podstawie pomiarów w momentach od 1 do t stosujemy metodę najmniejszych kwadratów [9] uzyskując następujące oszacowanie:

$$\hat{a}_{r,t} = (\Phi_{r,t} \Phi_{r,t}^T)^{-1} \Phi_{r,t}^T Y_t^T. \quad (3.28)$$

Wynik (3.28) stanowi podstawę do wyprowadzenia rekurencyjnej metody najmniejszych kwadratów opisywanej wyczerpująco w [69, 44, 2, 18] i dlatego poprzestaniemy na podaniu końcowego rezultatu bez szczegółowego wyprowadzania. Po wprowadzeniu oznaczenia

$$P_{r,t} = (\Phi_{r,t} \Phi_{r,t}^{-1})^{-1} \quad (3.29)$$

otrzymujemy następujące zależności do sekwencyjnego wyznaczania estymatora nieznanych parametrów modelu zgodnie z ogólną zależnością (3.2).

Procedura Ψ_1 :

$$\hat{a}_{r,t+1} = \hat{a}_{r,t} + K_{r,t+1}(y_{t+1} - \hat{a}_{r,t}^T \varphi_{r,t+1}), \quad (3.30a)$$

$$K_{r,t+1} = P_{r,t} \varphi_{r,t+1} / (1 + \varphi_{r,t+1}^T P_{r,t} \varphi_{r,t+1}), \quad (3.30b)$$

$$P_{r,t+1} = (I_r - K_{r,t+1} \varphi_{r,t+1}^T) P_{r,t}. \quad (3.30c)$$

Poprawka wartości estymatora w zależności (3.11a) zależy od błędu reszkowego modelu

$$\hat{\varepsilon}_{r,t} = y_{t+1} - \hat{a}_{r,t}^T \varphi_{r,t}, \quad (3.31)$$

który jest różnicą rzeczywistej i prognozowanej wartości wyjścia. Wartość (3.31) jest mnożona przez współczynnik wzmocnienia (3.30b), który z kolei zależy od bieżącej wartości macierzy kowariancji $P_{r,t}$ będącej miarą niepewności stanu modelu w bieżącej chwili czasu. I_r oznacza macierz jednostkową stopnia r . Zauważmy, że

3.4. Przypadek szczególny dla modelu regresji liniowej

algorytm komputerowy oparty na zależnościach (3.30) musi przechowywać bieżące wartości $\hat{a}_{r,t}$, $K_{r,t}$ oraz $P_{r,t}$, nie musi natomiast gromadzić poszczególnych rekordów danych z ciągu uczącego (2.6). Zwróćmy uwagę, że aktualizacja estymatora parametrów (3.30a) nie wymaga odwracania macierzy, w przeciwieństwie do algorytmu bezpośredniego (3.28).

Człon Ψ_2 algorytmu służy do przebudowy struktury modelu oraz odpowiedniej aktualizacji wektora parametrów. Podstawą do wyprowadzenia tej procedury jest algorytm bezpośredni (3.28), w którym dodajemy kolejną składową wektora parametrów, co wiąże się ze zwiększeniem rozmiarów macierzy w odpowiednich zależnościach:

$$\hat{a}_{r+1,t} = (\Phi_{r+1,t} \Phi_{r+1,t}^T)^{-1} \Phi_{r+1,t} Y_t^T \quad (3.32)$$

macierz $\Phi_{r+1,t}$ ma dodatkowy wiersz odpowiadający uwzględnieniu następnej składowej w wektorze regresji

$$\Phi_{r+1,t} = \begin{bmatrix} \varphi_1^{(1)} & \dots & \varphi_t^{(1)} \\ \vdots & \ddots & \vdots \\ \varphi_1^{(r)} & \dots & \varphi_t^{(r)} \\ \varphi_1^{(r+1)} & \dots & \varphi_t^{(r+1)} \end{bmatrix}. \quad (3.33)$$

Zapiszmy macierz (3.33) w postaci blokowej uwzględniając (3.5) i (3.27)

$$\Phi_{r+1,t} = \begin{bmatrix} \Phi_{r,t} \\ \phi_{r+1,t} \end{bmatrix}, \quad (3.34)$$

gdzie $\phi_{r+1,t}$ jest wektorem wierszowym zawierającym dane pomiarowe dla $r+1$ -ej składowej wektora regresji zdefiniowanej przez (3.5).

Macierz

$$P_{r+1,t} = (\Phi_{r+1,t} \Phi_{r+1,t}^T)^{-1} \quad (3.35)$$

zapiszmy w postaci blokowej korzystając z oznaczeń (3.5), (3.29) oraz (3.34)

$$P_{r+1,t} = \begin{bmatrix} P_{r,t}^{-1} & \Phi_{r,t} \phi_{r+1,t}^T \\ \phi_{r+1,t} \Phi_{r,t}^T & \phi_{r+1,t} \phi_{r+1,t}^T \end{bmatrix}. \quad (3.36)$$

3.4. Przypadek szczególny dla modelu regresji liniowej

W celu wyznaczenia macierzy odwrotnej do (3.36) skorzystamy z poniższego lematu, którego dowód można znaleźć w [31, 27].

Lemat 3.1.

Jeśli macierze P_{11} i P_{22} są kwadratowe, a macierze P_{11} i $P_\alpha = P_{22} - P_{21}P_{11}^{-1}P_{12}$ są nieosobliwe, to zachodzi następująca tożsamość macierzowa:

$$\begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}^{-1} = \begin{bmatrix} P_{11}^{-1} + P_{11}^{-1}P_{12}P_\alpha^{-1}P_{21}P_{11}^{-1} & -P_{11}^{-1}P_{12}P_\alpha^{-1} \\ -P_\alpha^{-1}P_{21}P_{11}^{-1} & P_\alpha^{-1} \end{bmatrix}.$$

Przyjmując $P_{11} = P_{r,t}^{-1}$, $P_{12} = \Phi_{r,t}\phi_{r+1,t}^T$, $P_{21} = \phi_{r+1,t}\Phi_{r,t}^T$, $P_{22} = \phi_{r+1,t}\phi_{r+1,t}^T$ oraz uwzględniając lemat 3.1 dostajemy po przekształceniach algebraicznych

$$P_{r+1,t} = \begin{bmatrix} P_{r,t} + \alpha_{r+1,t}b_{r+1,t}b_{r+1,t}^T & -\alpha_{r+1,t}b_{r+1,t} \\ -\alpha_{r+1,t}b_{r+1,t}^T & \alpha_{r+1,t} \end{bmatrix}, \quad (3.37)$$

gdzie dla zwięzłości zapisu wprowadzono następujące oznaczenia:

$$b_{r+1,t} = P_{r,t}\Phi_{r,t}\phi_{r+1,t}, \quad (3.38)$$

$$\alpha_{r+1,t} = [\phi_{r+1,t}(I_t - \Phi_{r,t}^T P_{r,t}\Phi_{r,t})\phi_{r+1,t}^T]^{-1}. \quad (3.38b)$$

Z kolei biorąc pod uwagę (3.34) i (3.35) w zależności (3.32) oraz wykorzystując wyznaczoną macierz (3.37) i wielkości pomocnicze (3.38) dostajemy zależność na korektę parametrów modelu po poszerzeniu wektora regresji o dodatkową składową

$$\begin{aligned} \hat{a}_{r+1,t} &= P_{r+1,t}\Phi_{r+1,t}Y_t^T = \\ &= \begin{bmatrix} \hat{a}_{r,t} + \beta_{r+1,t}b_{r+1,t} \\ -\beta_{r+1,t} \end{bmatrix}, \end{aligned} \quad (3.39)$$

gdzie

$$\beta_{r+1,t} = \alpha_{r+1,t}(b_{r+1,t}\Phi_{r,t} - \phi_{r+1,t})Y_t^T. \quad (3.40)$$

Z postaci wektora (2.12) wynika zapis blokowy

$$\varphi_{r+1,t} = [\varphi_{r,t}^T \quad \varphi_t^{(r+1)T}]^T. \quad (3.41)$$

3.4. Przypadek szczególny dla modelu regresji liniowej

DANE WEJŚCIOWE: $U_n, Y_n, \varepsilon, \mu, N_V, rmax, \varphi_{rmax}$

WARUNKI POCZĄTKOWE: $t = 1, N_V = 10, r = 1, \hat{a}_{r,0} = 0,$
 $\Phi_{r,n} = [1 \dots 1]_{1 \times n}, K_{r,0} = 0, P_{r,0} = \mu I_r$

KROK 1: Pobranie obserwacji $\{u_t, y_t\}$

KROK 2: Wyznaczenie $\hat{a}_{r,t}, K_{r,t}, P_{r,t}$ (3.30)

KROK 3: Ocena jakości modelu (3.7) lub (3.20-24)

KROK 4: Jeśli jakość modelu zadowalająca skok do kroku 11.

KROK 5: Jeśli $r > rmax$, skok do kroku 11.

KROK 6: $r = r + 1$

KROK 7: Wyznacz $\phi_{r,t}$ (3.5)

KROK 8: Wyznacz $\alpha_{r,t}, b_{r,t}$ (3.38), $\gamma_{r,t}, \delta_{r,t}$ (3.44), $\beta_{r,t}$ (3.40)

KROK 9: Wyznacz $\hat{a}_{r,t}$ (3.39)

KROK 10: Aktualizuj $K_{r,t}$ (3.43), $P_{r,t}$ (3.37), $\Phi_{r,t}$ (3.34)

KROK 11: jeśli $t < n$, to $t = t + 1$, skok do kroku 1

WYJŚCIE: $\hat{a}_{r,t}$

Algorytm 3.1

Zależność (3.30b) dla $r + 1$ ma postać

$$K_{r+1,t} = P_{r+1,t} \varphi_{r+1,t} / (1 + \varphi_{r+1,t}^T P_{r+1,t} \varphi_{r+1,t}). \quad (3.42)$$

Podstawiając do wyrażenia (3.42) macierz (3.37) oraz wektor kolumnowy (3.41) dostajemy po przekształceniach następujące wyrażenie:

$$K_{r+1,t} = \frac{1}{1 - \delta_{r+1,t} (\varphi_{r+1,t}^T b_{r+1,t} - \varphi_t^{(r+1)})} \begin{bmatrix} K_{r,t} - \delta_{r+1,t} b_{r+1,t} \\ \delta_{r+1,t} \end{bmatrix}, \quad (3.43)$$

gdzie wprowadzono dla uproszczenia zapisów poniższe wielkości:

$$\delta_{r+1,t} = \frac{\gamma_{r+1,t}}{1 + \varphi_{r,t}^T P_{r,t} \varphi_{r,t}}, \quad (3.44a)$$

3.4. Przypadek szczególny dla modelu regresji liniowej

$$\gamma_{r+1,t} = \alpha_{r+1,t}(\varphi_t^{(r+1)} - b_{r+1,t}^T \varphi_{r,t}). \quad (3.44b)$$

Zależności (3.39) , (3.37) i (3.43) stanowią główne formuły członu Ψ_2 algorytmu będąc skorygowanymi odpowiednikami wzorów (3.30) w członie Ψ_1 . Widać, że nawet w przypadku prostego modelu regresji liniowej algorytm jest dość skomplikowany obliczeniowo. Zwracamy również uwagę, że ze wzrostem rzędu modelu r rosną wymagania pamięciowe, gdyż zwiększeniu ulegają wymiary odpowiednich macierzy. Zasadniczą niekorzystną własnością algorytmu w tej prostej postaci jest konieczność przechowywania ciągu uczącego w celu poprawnej korekcji parametrów po przebudowie modelu. Dotychczas dyskutowaliśmy algorytmy, w których rekonstrukcja modelu polegała na komplikacji jego struktury. Naturalnym uzupełnieniem byłby algorytm, w którym możliwe jest również upraszczanie postaci modelu. Dla przypadku regresji liniowej algorytm ma strukturę przedstawioną na rys. 3.3. Różnica polega na dodaniu jeszcze jednej pętli zewnętrznej z członem Ψ_{2b} odpowiedzialnym za upraszczanie modelu. Procedury Ψ_1 i Ψ_2 pełnią analogiczne funkcje jak odpowiednio Ψ_1 i Ψ_2 w algorytmie 3.1. Analityczną postać procedury Ψ_{2b} otrzymujemy postępując odwrotnie niż w procedurze Ψ_{2a} . Wyrażmy (3.28) w postaci blokowej

$$\hat{a}_{r,t} = \begin{bmatrix} \tilde{a}_{r-1,t} \\ \beta_{r,t} \end{bmatrix}, \quad (3.45)$$

gdzie $\tilde{a}_{r-1,t}$ oznacza $r-1$ górnych składowych wektora kolumnowego (3.28). Wówczas, w oparciu o (3.39) dostajemy

$$\tilde{a}_{r-1,t} = \hat{a}_{r-1,t} + \beta_{r,t} b_{r,t}. \quad (3.46)$$

Podobnie, korzystając z postaci blokowej macierzy (3.37) mamy

$$P_{r,t} = \begin{bmatrix} \tilde{P}_{r-1,t} & -\alpha_{r,t} b_{r,t} \\ -\alpha_{r,t} b_{r,t} & \alpha_{r,t} \end{bmatrix}. \quad (3.47)$$

Symbol $\tilde{P}_{r-1,t}$ oznacza macierz kwadratową będącą klatką skorygowanej macierzy $P_{r,t}$, natomiast $P_{r-1,t}$ oznacza odpowiadającą jej macierz o rzędzie niższym o jeden. Analizując (3.37) i (3.47) dostajemy

3.4. Przypadek szczególny dla modelu regresji liniowej

$$P_{r-1,t} = \tilde{P}_{r-1,t} - b_{r,t} b_{r,t}^T \alpha_{r,t}. \quad (3.48)$$

Podobnie, wyrażając $K_{r,t}$ jako

$$K_{r,t} = \begin{bmatrix} \tilde{K}_{r-1,t} \\ K_t^{(r)} \end{bmatrix} \quad (3.49)$$

otrzymujemy

$$K_{r-1,t} = [1 - \delta_{r,t} (\varphi_r^T b_{r,t} - \varphi_t^{(r)})] \tilde{K}_{r-1,t} + \delta_{r,t} b_{r,t}. \quad (3.50)$$

Odpowiednie wielkości pomocnicze algorytmu wyrażamy następującymi zależnościami

$$\alpha_{r,t} = [\phi_{r,t} (I_t - \Phi_{r-1,t}^T P_{r-1,t} \Phi_{r-1,t}) \phi_{r,t}^T]^{-1}, \quad (3.51a)$$

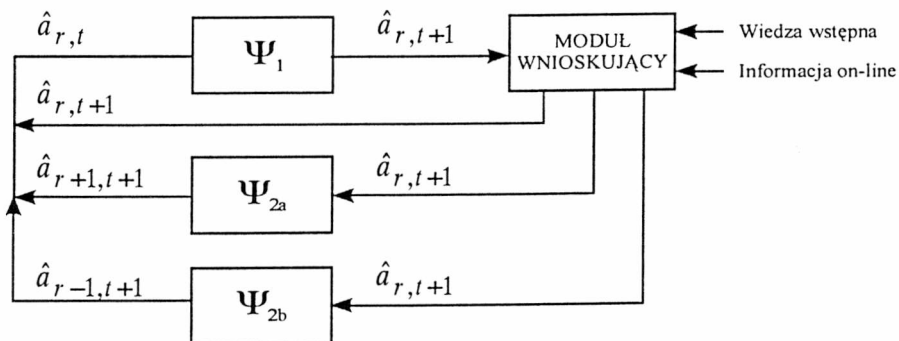
$$\beta_{r,t} = \alpha_{r,t} (b_{r,t} \Phi_{r-1,t} - \phi_{r,t}) Y_t^T, \quad (3.51b)$$

$$\gamma_{r,t} = \alpha_{r,t} (\varphi_t^{(r)} - b_{r,t}^T \varphi_{r,t}), \quad (3.51c)$$

$$\delta_{r,t} = \frac{\gamma_{r,t}}{1 + \varphi_{r-1,t}^T P_{r-1,t} \varphi_{r-1,t}}, \quad (3.51d)$$

$$b_{r,t} = P_{r-1,t} \Phi_{r-1,t} \phi_{r,t}. \quad (3.51e)$$

W następnym rozdziale opiszemy wersje algorytmu wykorzystujące techniki, w których przechowywana jest tylko część informacji z ciągu uczącego. Jednak, jak



Rys. 3.3. Algorytm o zmiennym rzędzie modelu

pokażemy, algorytm taki jest jeszcze bardziej pracochłonny obliczeniowo, niż obecna postać.

3.5. Algorytm ze stałą pamięcią

Algorytm prezentowany w rozdziale 3.4 wymaga przechowywania całego ciągu uczącego do przebudowy struktury modelu. Tę niedogodność można wyeliminować stosując techniki opisane w [24, 53], a polegające na odrzuceniu najstarszego z N przechowywanych rekordów danych pomiarowych po nadejściu bieżącej obserwacji. W ten sposób algorytm bazuje na N ostatnich obserwacjach. Wprowadźmy notację uwzględniającą stałą liczbę pomiarów. Przez $\hat{a}_{r,t+N,t+1}$ oznaczmy estymator r -elementowego wektora parametrów modelu, którego wartość wyznaczana jest w oparciu o N obserwacji wejścia i wyjścia dokonanych w chwilach od $t + 1$ do $t + N$ zapisanych analogicznie do (2.10) jako wektory wierszowe:

$$U_{t+N,t+1} = [u_{t+1} \cdots u_{t+N}], \quad (3.52a)$$

$$Y_{t+N,t+1} = [y_{t+1} \cdots y_{t+N}]. \quad (3.52b)$$

Zmodyfikowaną wersję algorytmu przedstawiono na rys. 3.4. Człon Ψ_1 wyznacza estymator kolejnej obserwacji w momencie $t + N$ i jest ogólnie określony następującą formułą

$$\hat{a}_{r,t+N,t} = \Psi_1(\hat{a}_{r,t+N-1,t}, \varphi_{r,t+N}), \quad (3.53)$$

przy czym

$$\varphi_{r,t+N} = [\varphi_{t+N}^{(1)} \cdots \varphi_{t+N}^{(r)}]^T. \quad (3.54)$$

Człon Ψ_{1a} koryguje wartość estymatora parametrów podczas usunięcia najstarszego pomiaru z przechowywanej N -elementowej części ciągu uczącego. Ogólna postać tej procedury opisana jest następującym wzorem:

$$\hat{a}_{r,t+N,t+1} = \Psi_{1a}(\hat{a}_{r,t+N,t}, \varphi_{r,t}). \quad (3.55)$$

przy czym $\varphi_{r,t}$ określone podobnie jak (2.12) odpowiada usuwanemu składnikowi.

Procedura Ψ_2 jest związana z przebudową modelu i ma następującą postać:

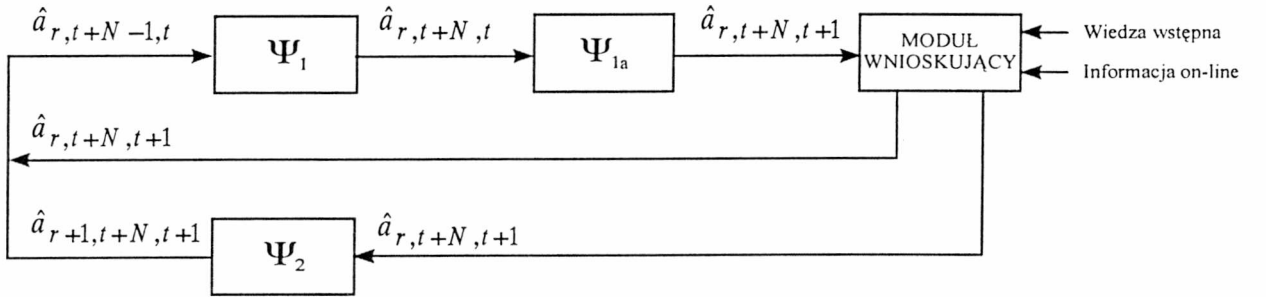
3.5. Algorytm ze stałą pamięcią

$$\hat{a}_{r+1,t+N,t+1} = \Psi_2(\hat{a}_{r,t+N,t+1}, \phi_{r+1,t+N,t+1}), \quad (3.56)$$

gdzie

$$\phi_{r+1,t+N,t+1} = [\varphi_{t+1}^{(r+1)} \dots \varphi_{t+N}^{(r+1)}] \quad (3.57)$$

odpowiada dodatkowej składowej wprowadzonej do wektora regresji. Dokładne analityczne postacie odpowiednich części algorytmu otrzymuje się poprzez



Rys. 3.4. Algorytm ze stałą pamięcią – wersja 1.

modyfikacji podstawowej wersji prezentowanej dla przypadku liniowego w rozdziale 3.4. Szczegółowe zależności dla bloku Ψ_1 na rys. 3.4. można otrzymać przez modyfikację wzorów (3.30). Indeksy odpowiednich macierzy ulegają zmianie wskazując, do których podzbiorów ciągu uczącego się odnoszą.

Procedura Ψ_1

$$\hat{a}_{r,t+N,t} = \hat{a}_{r,t+N-1,t} + K_{r,t+N,t} (y_{t+N} - \hat{a}_{r,t+N-1,t}^T \varphi_{r,t+N}), \quad (3.58a)$$

$$P_{r,t+N,t} = (I_r - K_{r,t+N,t} \varphi_{r,t+N}^T) P_{r,t+N-1,t}, \quad (3.58b)$$

$$K_{r,t+N,t} = \frac{P_{r,t+N-1,t} \varphi_{r,t+N}}{1 + \varphi_{r,t+N}^T P_{r,t+N-1,t} \varphi_{r,t+N}}. \quad (3.58c)$$

W zależnościach (3.58) $\varphi_{r,t+N}$ wyznaczone jest na podstawie obserwacji w momencie $t+N$ na podstawie wzoru analogicznego do (2.13), a macierz $P_{r,t+N,t+1}$ jest określona analogicznie do (3.29) wzorem

$$P_{r,t+N,t+1} = (\Phi_{r,t+N,t+1} \Phi_{r,t+N,t+1}^{-1})^{-1}, \quad (3.59)$$

w którym macierz pomiarowa $\Phi_{r,t+N,t+1}$ zależy teraz od N ostatnich obserwacji

$$\Phi_{r,t+N,t+1} = \begin{bmatrix} \varphi_{t+1}^{(1)} & \cdots & \varphi_{t+N}^{(1)} \\ \vdots & \ddots & \vdots \\ \varphi_{t+1}^{(r)} & \cdots & \varphi_{t+N}^{(r)} \end{bmatrix}, \quad (3.60)$$

a jej rozmiar wynosi $\dim(\Phi_{r,t+N,t+1}) = r \times N$.

Procedura Ψ_{1a}

Człon Ψ_{1a} z rys. 3.4 jest dodatkowym członem wprowadzonym do podstawowej wersji algorytmu rozważanej w rozdz. 3.4. Odpowiednie szczegółowe zależności dane są następującymi wzorami

$$\hat{a}_{r,t+N,t+1} = \hat{a}_{r,t+N,t} + K_{r,t+N,t+1}(y_t - \hat{a}_{r,t+N,t}^T \varphi_{r,t}), \quad (3.61a)$$

$$P_{r,t+N,t+1} = (I_r - K_{r,t+N,t+1} \varphi_{r,t}^T) P_{r,t+N,t}, \quad (3.61b)$$

$$K_{r,t+N,t+1} = \frac{P_{r,t+N,t} \varphi_{r,t}}{1 + \varphi_{r,t}^T P_{r,t+N,t} \varphi_{r,t}}. \quad (3.61c)$$

Procedura Ψ_2

Człon Ψ_2 uzyskuje się zmieniając indeksy w zależnościach (3.39), (3.37), (3.43), (3.38), (3.40) oraz (3.44). Otrzymujemy następujące zależności algorytmu

$$\hat{a}_{r+1,t+N,t+1} = \begin{bmatrix} \hat{a}_{r,t+N,t+1} + \beta_{r+1,t+N,t+1} b_{r+1,t+N,t+1} \\ -\beta_{r+1,t+N,t+1} \end{bmatrix}, \quad (3.62a)$$

$$P_{r+1,t+N,t+1} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}, \quad (3.62b)$$

gdzie $P_{11} = P_{r,t+N,t+1} + \alpha_{r+1,t+N,t+1} b_{r+1,t+N,t+1} b_{r+1,t+N,t+1}^T$,

$P_{12} = -\alpha_{r+1,t+N,t+1} b_{r+1,t+N,t+1}$, $P_{21} = P_{12}^T$, $P_{22} = \alpha_{r+1,t+N,t+1}$.

$$K_{r+1,t+N,t+1} = \frac{1}{1 - \delta_{r+1,t+N,t+1} (\varphi_{r+1,t+N,t+1}^T b_{r+1,t+N,t+1} - \varphi_t^{(r+1)})} \times \begin{bmatrix} K_{r,t+N,t+1} - \delta_{r+1,t+N,t+1} b_{r+1,t+N,t+1} \\ \delta_{r+1,t+N,t+1} \end{bmatrix}. \quad (3.62c)$$

3.5. Algorytm ze stałą pamięcią

Wartości pomocnicze dane są zależnościami

$$b_{r+1,t+N,t+1} = P_{r,t+N,t+1} \Phi_{r,t+N,t+1} \phi_{r+1,t+N,t+1}^T, \quad (3.63a)$$

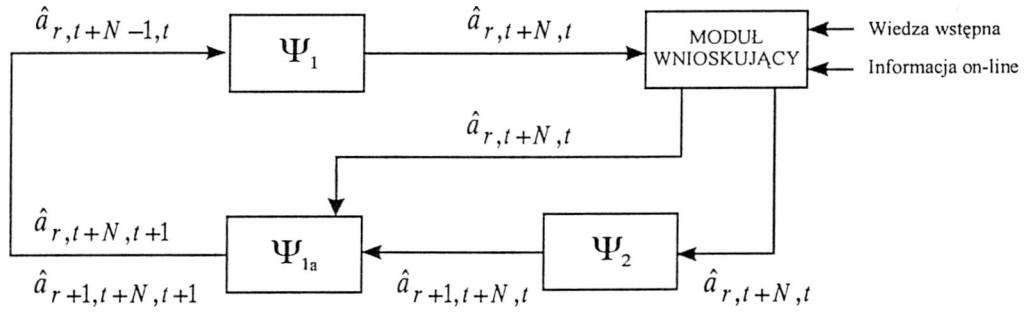
$$\alpha_{r+1,t+N,t+1} = [\phi_{r+1,t+N,t+1} (I_N - \Phi_{r,t+N,t+1}^T P_{r,t+N,t+1} \Phi_{r,t+N,t+1}) \phi_{r+1,t+N,t+1}^T]^{-1}, \quad (3.63b)$$

$$\beta_{r+1,t+N,t+1} = \alpha_{r+1,t+N,t+1} (b_{r+1,t+N,t+1} \Phi_{r,t+N,t+1} - \phi_{r+1,t+N,t+1}) Y_{t+N,t+1}^T, \quad (3.63c)$$

$$\gamma_{r+1,t+N,t+1} = \alpha_{r+1,t+N,t+1} (\varphi_{t+N}^{(r+1)} - b_{r+1,t+N,t+1}^T \varphi_{r,t+N}), \quad (3.63d)$$

$$\delta_{r+1,t+N,t+1} = \frac{\gamma_{r+1,t+N,t+1}}{1 + \varphi_{r,t+N}^T P_{r,t+N} \varphi_{r,t+N}}. \quad (3.63e)$$

Kolejność procedur Ψ_{1a} oraz Ψ_2 w algorytmie na rys. 3.4. może być zmieniona, co prowadzi do algorytmu na rys. 3.5. Opisy na rys. 3.5. nad linią odnoszą się do sytuacji,



Rys.3.5. Algorytm ze stałą pamięcią – wersja 2.

gdy algorytm wykonuje tylko wewnętrzną pętlę. (tj. nie wykonuje przebudowy modelu). Procedura Ψ_1 jest taka sama w obydwu wariantach algorytmu. Nieznacznej zmianie ulegną człony Ψ_{1a} i Ψ_2 . Usunięcie wpływu pomiaru pobranego w chwili t ma teraz następującą postać.

Procedura Ψ_{1a} (wariant 2.)

$$\hat{a}_{r+1,t+N,t+1} = \hat{a}_{r+1,t+N,t} - K_{r+1,t+N,t+1} (y_t - \hat{a}_{r+1,t+N,t}^T \varphi_{r+1,t}), \quad (3.64a)$$

$$P_{r+1,t+N,t+1} = (I_{r+1} - K_{r+1,t+N,t+1} \varphi_{r+1,t}^T) P_{r+1,t+N,t}, \quad (3.64b)$$

3.5. Algorytm ze stałą pamięcią

$$K_{r+1,t+N,t+1} = \frac{P_{r+1,t+N,t+1}\varphi_{r+1,t}}{1 - \varphi_{r+1,t}^T P_{r+1,t+N,t} \varphi_{r+1,t}}, \quad (3.64c)$$

Procedura Ψ_2 (wariant 2.)

$$\hat{a}_{r+1,t+N,t} = \begin{bmatrix} \hat{a}_{r,t+N,t} + \beta_{r+1,t+N,t} b_{r+1,t+N,t} \\ -\beta_{r+1,t+N,t} \end{bmatrix}, \quad (3.65a)$$

$$P_{r+1,t+N,t+1} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}, \quad (3.65b)$$

gdzie $P_{11} = P_{r,t+N,t} + \alpha_{r+1,t+N,t} b_{r+1,t+N,t}^T$, $P_{12} = -\alpha_{r+1,t+N,t} b_{r+1,t+N,t}$, $P_{21} = P_{12}^T$, $P_{22} = \alpha_{r+1,t+N,t}$.

$$K_{r+1,t+N,t} = \frac{1}{1 - \delta_{r+1,t+N,t} (\varphi_{r+1,t+N}^T b_{r+1,t+N,t} - \varphi_t^{(r+1)})} \times \begin{bmatrix} K_{r,t+N,t} - \delta_{r+1,t+N,t} b_{r+1,t+N,t} \\ \delta_{r+1,t+N,t} \end{bmatrix}. \quad (3.65c)$$

Wielkości pomocnicze algorytmu obliczane są według poniższych wzorów

$$\alpha_{r+1,t+N,t} = [\phi_{r+1,t+N,t} (I_{N+1} - \Phi_{r,t+N,t}^T P_{r,t+N,t} \Phi_{r,t+N,t}) \phi_{r+1,t+N,t}^T]^{-1}, \quad (3.66a)$$

$$\beta_{r+1,t+N,t} = \alpha_{r+1,t+N,t} (b_{r+1,t+N,t} \Phi_{r,t+N,t} - \phi_{r,t+N,t}) Y_{t+N,t}^T, \quad (3.66b)$$

$$\gamma_{r+1,t+N,t} = \alpha_{r+1,t+N,t} (\varphi_{t+N}^{(r+1)} - b_{r+1,t+N,t}^T \varphi_{r,t+N,t}), \quad (3.66c)$$

$$\delta_{r+1,t+N,t} = \frac{\gamma_{r+1,t+N,t}}{1 + \varphi_{r,t+N,t}^T P_{r,t+N,t} \varphi_{r,t+N,t}}, \quad (3.66d)$$

$$b_{r+1,t+N,t} = P_{r,t+N,t} \Phi_{r,t+N,t} \phi_{r+1,t+N,t}^T. \quad (3.66e)$$

Rekurencyjne kryteria rekonstrukcji modelu rozpatrywane w rozdz. 3.3 dadzą się uogólnić dla algorytmów ze stałą pamięcią. Ich wyznaczanie jest wówczas oparte na stałej liczbie pomiarów N . Wyznaczenie ocen dla algorytmów rozpatrywanych w bieżącym rozdziale wymaga dwóch kroków obliczeniowych. Pierwszy uwzględnia

nowy pomiar, a drugi wynika z odrzucenia najstarszej obserwacji. Wzór (3.16) daje następujący odpowiednik

$$M_{t+N,t}^{\varepsilon} = \max(M_{t+N-1,t}^{\varepsilon}, \varepsilon_{t+N}), \quad (3.67a)$$

$$M_{t+N,t+1}^{\varepsilon} = M_{t+N,t}^{\varepsilon}. \quad (3.67b)$$

Podamy wzory dla ocen wyrażonych zależnościami (3.21-24)

- Wartość średnia błędu resztkowego

$$m_{t+N,t}^{\varepsilon} = m_{t+N-1,t}^{\varepsilon} + \frac{1}{N+1}(\varepsilon_{t+N} - m_{t+N-1,t}^{\varepsilon}), \quad (3.68a)$$

$$m_{t+N,t+1}^{\varepsilon} = m_{t+N,t}^{\varepsilon} + \frac{1}{N}(m_{t+N,t}^{\varepsilon} - \varepsilon_t). \quad (3.68b)$$

- Wariancja błędu resztkowego

$$V_{t+N,t}^{\varepsilon} = \frac{N-1}{N}V_{t+N-1,t}^{\varepsilon} + \frac{1}{N+1}(\varepsilon_{t+N} - m_{t+N-1,t}^{\varepsilon})^2, \quad (3.69a)$$

$$V_{t+N,t+1}^{\varepsilon} = \frac{N}{N-1}V_{t+N,t}^{\varepsilon} - \frac{N}{N^2+1}(\varepsilon_t - m_{t+N,t}^{\varepsilon})^2. \quad (3.69b)$$

- Średniokwadratowa suma błędów resztkowych

$$S_{t+N,t}^{\varepsilon} = S_{t+N-1,t}^{\varepsilon} + \frac{1}{N+1}(\varepsilon_{t+N}^2 - S_{t+N-1,t}^{\varepsilon}), \quad (3.70a)$$

$$S_{t+N,t+1}^{\varepsilon} = S_{t+N,t}^{\varepsilon} + \frac{1}{N}(S_{t+N,t}^{\varepsilon} - \varepsilon_t^2). \quad (3.70b)$$

lub innym sposobem

$$S_{t+N,t}^{\varepsilon} = (m_{t+N,t}^{\varepsilon})^2 + \frac{N}{N+1}V_{t+N,t}^{\varepsilon}, \quad (3.71a)$$

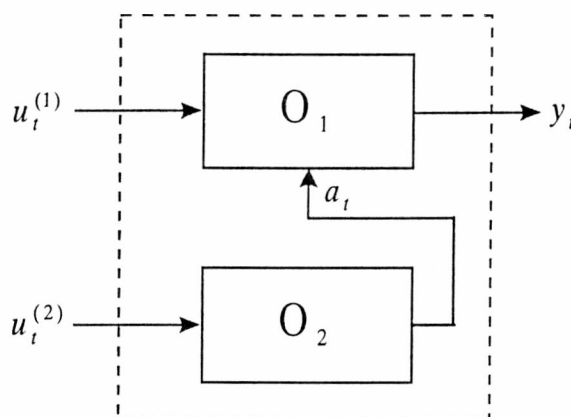
$$S_{t+N,t+1}^{\varepsilon} = (m_{t+N,t+1}^{\varepsilon})^2 + \frac{N-1}{N}V_{t+N,t+1}^{\varepsilon}. \quad (3.71b)$$

4. Dwustopniowe algorytmy rekurencyjne

Identyfikacja dwustopniowa oraz jej zastosowania techniczne i biomedyczne jest przedmiotem monografii [71]. W pracy tej wskazano opracowanie rekurencyjnych algorytmów identyfikacji dwustopniowej jako jeden z kierunków dalszych badań. Potrzeba zastosowania takich algorytmów pojawia się w wielu praktycznych problemach podejmowania decyzji lub sterowania, w których używana jest dwustopniowa struktura modelu danego zjawiska bądź procesu. Wybór takiej struktury podyktowany jest najczęściej wiedzą wstępną, jaką dysponujemy po analizie systemu, którego model zamierzamy konstruować. Komputerowe procedury wyznaczania na bieżąco ocen parametrów w klasach modeli dwustopniowych są przedmiotem niniejszego rozdziału, który jest oryginalnym uzupełnieniem badań nad identyfikacją dwustopniową zawartych w monografii [71].

4.1. Zagadnienia wstępne

W identyfikacji dwustopniowej przyjmuje się, że model analizowanego zjawiska ma złożoną hierarchiczną strukturę przedstawioną ogólnie na rys. 4.1. Dla pewnego obiektu O_1 na pierwszym stopniu konstruujemy model na podstawie obserwacji wejść



Rys. 4.1. Dwustopniowy obiekt identyfikacji

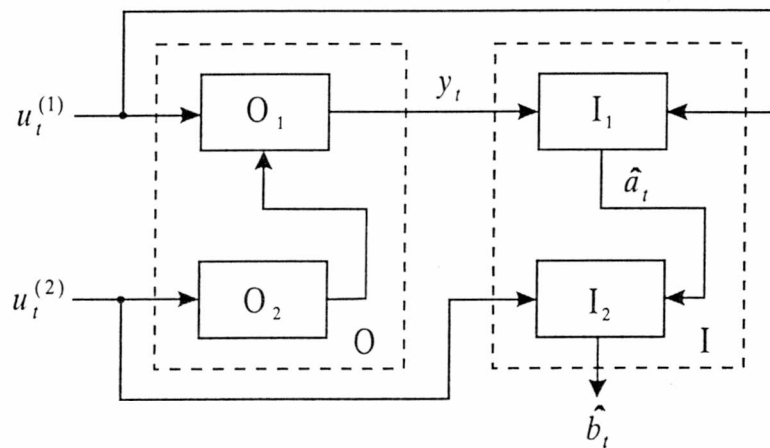
$u_t^{(1)}$ i wyjść y_t w dyskretnych momentach czasu $t = 1, \dots, N$ wyznaczając parametr a_t w opisie tej zależności. Jest to identyfikacja na pierwszym stopniu. Następnie zwracamy uwagę na pewną wielkość $u_t^{(2)}$, która ma wpływ na wartość parametru a_t . Dla różnych wartości zmiennej $u_t^{(2)}$ wyniki identyfikacji na pierwszym stopniu będą

się na ogół różnić. Możemy badać zależność parametru a_t od zmiennej $u_t^{(2)}$, a dokładniej, wyznaczyć parametr b w opisie tej zależności – jest to identyfikacja na drugim stopniu. W schemacie przedstawionym na rys. 4.1 parametr modelu pierwszego stopnia sam jest wyjściem obiektu, przy czym nie można go bezpośrednio zmierzyć ani zaobserwować. Wiadomo tylko, że parametr ten zależy od pewnej wielkości wejściowej na drugim stopniu $u_t^{(2)}$ i wpływ ten potrafimy opisać w formie modelu o nieznanym parametrze b , który należy wyznaczyć na podstawie obserwacji wejść i wyjść złożonego systemu dwustopniowego. Zauważmy, że parametr b jest wyznaczany w sposób pośredni, poprzez identyfikację na pierwszym stopniu. W ujęciu identyfikacji dwustopniowej opisanym w monografii [71] zakłada się, że podczas identyfikacji na pierwszym stopniu wielkość $u_t^{(2)}$ jest stała podczas całego procesu wyznaczania modelu dla pierwszego stopnia. Oznacza to, że $u_t^{(2)} = \text{const}$ dla $t = 1 \dots N$. Przyjmuje się tam również, że eksperyment jest czynny co najmniej dla drugiego stopnia, a parametr b wyznacza się powtarzając wielokrotnie identyfikację na pierwszym stopniu dla różnych wartości $u_t^{(2)}$. W rzeczywistych problemach założenia takie nie zawsze mogą być spełnione. Dotyczy to szczególnie przypadku identyfikacji dynamicznego związku na drugim stopniu. Zdarza się, że sygnał wejściowy drugiego stopnia jest procesem losowym, który choć mierzalny, może nie zależeć od naszej woli. Po drugie, czasem nie jesteśmy w stanie zapewnić warunku stałości sygnału $u_t^{(2)}$, gdyż taki eksperyment jest niemożliwy z fizycznego punktu widzenia (mógłby narazić obiekt na uszkodzenie). Możemy również brać pod uwagę przypadki, w których w ogóle nie występuje sygnał $u_t^{(2)}$, mając do czynienia z analizą szeregu czasowego na drugim stopniu [8]. Ta sama uwaga dotyczy również pierwszego stopnia.

Uzasadnienia zastosowania identyfikacji dwustopniowej mogą być wielorakie. Rozważa się identyfikację oddzielnych realnych obiektów na pierwszym stopniu charakteryzujących się różnymi stałymi wartościami wielkości $u_t^{(2)}$. Wielkości $u_t^{(1)}$ i y_t są takie same dla wszystkich obiektów. Poszczególne obiekty mogą być nawet badane w różnych miejscach (np. laboratoriach). Jako przykład mogą służyć medyczne badania populacyjne, gdzie przedmiot analizy dotyczy wpływu cechy wyróżniającej obiekt z populacji (np. wieku) na badany proces (np. funkcje

mechaniczne układu oddechowego). Powyższy przypadek określono jako *dekompozycję przestrzenną* [71]. Zwróćmy uwagę, że zmiana $u_t^{(2)}$ dla pojedynczego obiektu jest niewygodna lub wręcz niemożliwa. W pracy tej nie będziemy zajmować się dwustopniową identyfikacją obiektów rozproszonych przestrzennie, ponieważ rekurencyjne podejście nie znajduje tu uzasadnienia praktycznego.

Drugi przypadek, określany mianem *dekompozycji czasowej*, dotyczy identyfikacji pojedynczego obiektu i związany jest z odpowiednią organizacją eksperymentu. Można rozpatrywać równoczesną zmianę $u_t^{(1)}$ i $u_t^{(2)}$ jak i szybką zmianę $u_t^{(1)}$ oraz wolną $u_t^{(2)}$. Dla obiektów dyskretnych, rozważanych w pracy, ma to konsekwencje

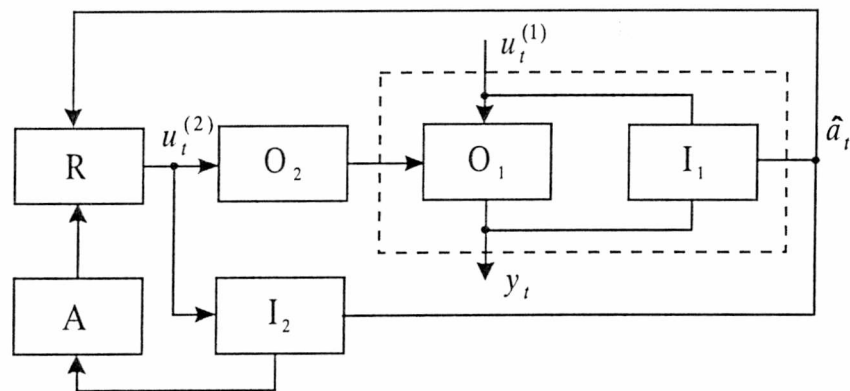


Rys. 4.2. Układ identyfikacji pośredniej parametru modelu drugiego stopnia

w postaci różnych okresów próbkowania sygnałów pomiarowych dla obydwu stopni. W pracy [71] przyjęto, że eksperyment zorganizowany jest tak, że utrzymywana jest stała wartość $u_t^{(2)}$ w kolejnych cyklach identyfikacji na pierwszym stopniu, w których zmieniają się $u_t^{(1)}$ i y_t . Wyodrębnienie dwóch stopni dla pojedynczego obiektu może mieć samoistne znaczenie praktyczne i wynikać z wiedzy wstępnej i analizy modelowanego zjawiska. Drugi stopień reprezentuje np. oddziaływanie pewnych czynników zewnętrznych na podstawowy proces na pierwszym stopniu. Konieczność wyodrębnienia drugiego stopnia wynika również z niemożności bezpośredniego pomiaru parametru a i wówczas identyfikacja na pierwszym stopniu pełni rolę pomocniczą pewnego rodzaju układu pomiarowego, w którym mierzy się dodatkowe wielkości $u_t^{(1)}$ i y_t w celu pośredniego wyznaczenia b . Sytuacja ta została

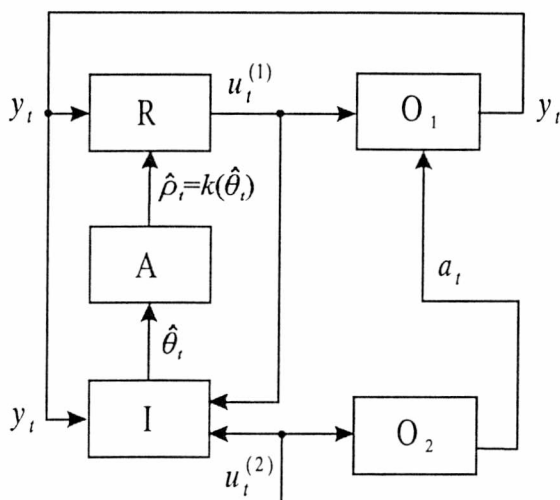
zilustrowana na rys. 4.2. Obszary otoczone liniami przerywanymi, oznaczone jako O i I, reprezentują odpowiednio złożony dwustopniowy obiekt i identyfikator. Identyfikator tak jak i obiekt ma strukturę dwustopniową złożoną z bloków I_1 oraz I_2 . Do wyznaczenia oceny $\hat{b}_t$ używany jest bieżący estymator parametru a , który oznaczono przez $\hat{a}_t$. Omawiany przypadek wiąże się z zastosowaniem identyfikacji dwustopniowej w sterowaniu adaptacyjnym z bieżącą identyfikacją. Taki system sterowania przedstawiono ogólnie na rys. 4.3. Obiekt sterowania znajduje się na drugim stopniu. Wielkością sterującą jest $u_t^{(2)}$, a wielkością sterowaną a_t . Algorytm sterowania R wymaga określenia modelu obiektu (zależność a_t od $u_t^{(2)}$). Wyjście obiektu nie może być mierzone bezpośrednio lecz jest wynikiem pomocniczej identyfikacji na pierwszym stopniu.

Podana na rys. 4.3. konfiguracja systemu sterowania nie jest jedyną z możliwych. Sterowaniu może podlegać także proces na pierwszym stopniu, natomiast drugi



Rys. 4.3. Układ sterowania adaptacyjnego z identyfikacją dwustopniową

stopień dodany jest w celu jawnego określenia wpływu pewnej zmiennej na parametr głównego procesu na pierwszym stopniu. Oddziaływanie zmiennej wejściowej na parametr procesu jest na tyle istotne, że nie można go zaniedbać. Określamy je w postaci pewnego modelu o nieznanym parametrze. Opisaną sytuację odpowiada system sterowania przedstawiony na rys. 4.4. Identyfikator I wyznacza ocenę łączną parametrów procesu $\hat{\theta}_t = [\hat{a}_t^T \ \hat{b}_t^T]^T$, które służą do wyznaczenia parametrów regulatora R przez układ adaptacji A. Parametry regulatora $\hat{\rho}_t$ wyznaczane są przez układ A według pewnej reguły k.



Rys. 4.4. Inna konfiguracja systemu sterowania z obiektem dwustopniowym

Wyodrębnienie oddzielnego stopnia może być również uzasadnione względami obliczeniowymi. Wiadomo z literatury [69, 24, 60], jak ważną rolę w identyfikacji spełnia etap planowania eksperymentu. Sygnały użyte do identyfikacji powinny zapewniać trwałe pobudzenie sytemu. Ze względu na pewne ograniczenia natury fizycznej, sygnał $u_t^{(2)}$ może nie spełniać tego wymogu. Spełnia go natomiast sygnał $u_t^{(1)}$. Jeśli przykładowo $u_t^{(2)}$ oznacza podaż leku w formie jednorazowej dawki, to sygnał taki ma charakter pojedynczego impulsu, a więc nie jest dogodny do celów identyfikacji. Jeśli interesuje nas model wpływu leku na pewien proces zachodzący w organizmie, to próbujemy wyznaczyć parametry tego modelu poprzez identyfikację procesu na pierwszym stopniu.

Wstępne rozważania i przykłady przedstawione powyżej służyły uzasadnieniu celowości zastosowania bieżącej identyfikacji dwustopniowej. Stąd wynika również potrzeba opracowania algorytmów takiej identyfikacji dogodnych dla realizacji komputerowej.

4.2. Opis matematyczny i sformułowanie problemu

Obecnie nieco dokładniej niż w poprzednim rozdziale sformułujemy problem identyfikacji obiektu dwustopniowego. Rozważać będziemy na początku skalarne obiekty dyskretne, a więc sygnały uzyskane z pomiarów bądź obserwacji mają charakter ciągów czasowych. Dwa z nich (4.1a,b) odpowiadają zmiennym

wejściowym dla obu stopni, a trzeci (4.1c) opisuje wyjście obiektu na pierwszym stopniu. Zapiszmy dane pomiarowe w postaci odpowiednich wektorów wierszowych.

$$U_t^{(1)} = [u_1^{(1)} \dots u_t^{(1)}], \quad (4.1a)$$

$$U_t^{(2)} = [u_1^{(2)} \dots u_t^{(2)}], \quad (4.1b)$$

$$Y_t = [y_1 \dots y_t]. \quad (4.1c)$$

Interesuje nas wyznaczenie pewnej funkcji

$$\hat{y}_t = \hat{g}_t(\varphi_t^{(1)}, \varphi_t^{(2)}), \quad (4.2)$$

gdzie $\varphi_t^{(1)}$ i $\varphi_t^{(2)}$ są wektorami regresji odpowiednio na pierwszym i drugim stopniu, przy czym w przypadku identyfikacji dynamicznej wektor regresji na pierwszym stopniu ma postać

$$\varphi_t^{(1)} = \varphi^{(1)}(U_{t-1}^{(1)}, Y_{t-1}) \quad (4.3)$$

i zależy od przeszłych obserwacji dla pierwszego stopnia. Z kolei na drugim stopniu mamy ogólną zależność

$$\varphi_t^{(2)} = \varphi^{(2)}(U_{t-1}^{(2)}). \quad (4.4)$$

Rozważamy przypadek parametryczny i dla uproszczenia założymy, że dane pomiarowe z obydwu stopni pobieramy w tym samym cyklu próbkowania. Przyjmujemy, że oba stopnie są sparametryzowane oddzielnie, przy czym parametr modelu na pierwszym stopniu nie jest wyznaczany bezpośrednio, lecz jest wynikiem pewnego przetwarzania (np. filtracji), na który wpływ ma wielkość $u_t^{(2)}$. Można także rozważać statyczne uzależnienie parametru a_t od tej zmiennej. Wówczas otrzymamy zamiast zależności (4.4) następujący związek

$$\varphi_t^{(2)} = \varphi^{(2)}(u_t^{(2)}) \quad (4.5)$$

Zależność (4.2) możemy zdekomponować na równania odpowiadające pierwszemu i drugiemu stopniowi.

$$\hat{y}_t = g_1(\varphi_t^{(1)}, \hat{a}_t), \quad (4.6)$$

$$\hat{a}_t = g_2(\varphi_t^{(2)}, \hat{b}_t), \quad (4.7)$$

gdzie g_1 i g_2 oznaczają postacie modelu na pierwszym i drugim stopniu, natomiast $\hat{a}_t$ i $\hat{b}_t$ oznaczają oceny parametrów dla pierwszego i drugiego stopnia wyznaczane w procesie identyfikacji. Rozważmy przypadek, gdy identyfikacja na pierwszym stopniu ma charakter pomocniczy, a istotne znaczenie ma drugi stopień, w którym modelujemy wpływ zmiennej $u_t^{(2)}$ na parametr a . Wstawiamy zależność (4.7) do (4.6), otrzymując

$$\hat{y}_t = g_1(\varphi_t^{(1)}, g_2(\varphi_t^{(2)}, \hat{b}_t)). \quad (4.8)$$

Parametr b jest wynikiem identyfikacji złożonego obiektu, dla którego przyjęto strukturę modelu postaci:

$$g(\varphi_t^{(1)}, \varphi_t^{(2)}, b) = g_1(\varphi_t^{(1)}, g_2(\varphi_t^{(2)}, b)). \quad (4.9)$$

Wyznaczenie modelu w sensie zależności (4.2) jest rozumiane jako estymacja parametru b , czyli biorąc pod uwagę (4.2) i (4.9) otrzymujemy

$$\hat{g}_t(\varphi_t^{(1)}, \varphi_t^{(2)}) = g(\varphi_t^{(1)}, \varphi_t^{(2)}, \hat{b}_t). \quad (4.10)$$

Równania (4.6) i (4.7) wskazują na niestacjonarny charakter parametru a . Stąd wynika jeszcze jedna interpretacja drugiego stopnia jako próba określenia modelu niestacjonarności dla parametru a pierwszego stopnia identyfikacji. Jak zobaczymy dalej, taki model zmian parametru a , poprzez wprowadzenie drugiego stopnia spowoduje duże trudności obliczeniowe nawet dla prostych przypadków.

4.3. Przypadek skalarny, model liniowy

Kontynuując rozważania z poprzedniego podrozdziału zajmiemy się bardziej szczegółowo przypadkiem identyfikacji dwustopniowej, w którym dla obydwu stopni zastosowano liniowe klasy modeli. Przyjmujemy również, że wielkości $u_t^{(1)}$, $u_t^{(2)}$, y_t oraz parametry modeli obu stopni a i b są skalarnie. Dla pierwszego stopnia mamy

$$g_1(\varphi_t^{(1)}, a_t) = a_t \varphi_t^{(1)}, \quad (4.11)$$

gdzie $\varphi_t^{(1)}$ określono wzorem (4.3), a a_t jest parametrem zależnym liniowo od zmiennej $u_t^{(2)}$ w następujący sposób

$$a_t = g_2(\varphi_t^{(2)}, b) = b \varphi_t^{(2)}, \quad (4.12)$$

gdzie $\varphi_t^{(2)}$ określono wzorem (4.4). Naszym celem jest wyznaczenie rekurencyjnego algorytmu Ψ , który znajduje kolejne oceny parametru b , w oparciu o oszacowania z poprzedniej chwili oraz nadchodzące obserwacje wejść i wyjść złożonego obiektu

$$\hat{b}_{t+1} = \Psi(\varphi_{t+1}^{(1)}, \varphi_{t+1}^{(2)}, \hat{b}_t). \quad (4.13)$$

Zastosujemy podejście bayesowskie, w którym zakładamy, że parametr b jest zmienną losową o pewnej funkcji gęstości znanej a priori. Zakładamy, że wielkości $u_t^{(1)}$ i $u_t^{(2)}$ są deterministyczne. Estymacja metodą Bayesa polega na wyznaczeniu w każdej chwili t rozkładu a posteriori parametru b czyli

$$p(b | Y_t, U_t^{(1)}, U_t^{(2)}), \quad (4.14)$$

z którego możemy na różne sposoby wyznaczyć estymator nieznanego parametru b . Jedną z metod jest *metoda średniej a posteriori* [9]

$$\hat{b}_t = E(b | Y_t, U_t^{(1)}, U_t^{(2)}). \quad (4.15)$$

Drugą powszechnie stosowaną metodą jest *metoda maksymalnego prawdopodobieństwa a posteriori*, w której estymator nieznanego parametru wyznaczamy w oparciu o zależność

$$\hat{b}_t = \arg \max_b p(b | Y_t, U_t^{(1)}, U_t^{(2)}). \quad (4.16)$$

Obie metody dają takie same estymatory, jeśli rozkład (4.14) jest symetryczny. Znalezienie rekurencyjnych związków dla estymatorów (4.15) i (4.16) jest ogólnie bardzo trudne. Rozwiązanie analityczne jest możliwe w przypadku modelu liniowego i gaussowskiego rozkładu zakłóceń. W naszym przypadku, dla klas modeli

zdefiniowanych według wzorów (4.11) oraz (4.12) równania opisujące dwustopniowy obiekt mają następującą postać

$$y_t = a_t \varphi_t^{(1)} + e_t, \quad (4.17)$$

$$a_t = b \varphi_t^{(2)} + w_t, \quad (4.18)$$

gdzie $\{e_t\}$ i $\{w_t\}$ są z założenia ciągami zmiennych losowych o jednakowych rozkładach prawdopodobieństwa dla każdego momentu t , skończonych wariancjach równych odpowiednio λ_e i λ_w , i średnich równych zero, tzn. $Ee_t = Ew_t = 0$. Zakładamy również brak korelacji tych sygnałów zakłócających z sygnałami wejściowymi $u_t^{(1)}$ i $u_t^{(2)}$. Przyjmiemy, że rozkłady zakłóceń są normalne tzn. $e_t \sim N(0, \lambda_e)$ oraz $w_t \sim N(0, \lambda_w)$, a także założymy, że rozkład a priori parametru b jest również gaussowski

$$b \sim N(b^0, P^0). \quad (4.19)$$

Zakłócenia w równaniach (4.17), (4.18) reprezentują nie tylko rzeczywiste szумы, czy błędy pomiarowe lecz, co również jest ważne, wyrażają niepewność w doborze struktur modeli obu stopni systemu. Przyjęte klasy modeli (4.11), (4.12) są w istocie pewnymi hipotezami. Stąd, model dolnego stopnia jest obarczony błędem mimo tego, że jego wyjście nie jest mierzone, czy obserwowane. Do układu (4.17), (4.18) stosujemy przekształcenie wyrażone ogólnie przez (4.8). Wstawiając (4.18) do (4.17) dostajemy równoważny model wypadkowy

$$y_t = b \varphi_t^{(1)} \varphi_t^{(2)} + \varphi_t^{(1)} w_t + e_t, \quad (4.20)$$

który można zapisać również jako

$$y_t = b \varphi_t + v_t, \quad (4.21)$$

gdzie $\varphi_t = \varphi_t^{(1)} \varphi_t^{(2)}$ jest wektorem regresji złożonego modelu, a $v_t = \varphi_t^{(1)} w_t + e_t$ jest wypadkowym zakłóceniem. Zauważamy, że z determinizmu $\varphi_t^{(1)}$ oraz niezależności e_t i w_t , ponieważ v_t jest liniową kombinacją niezależnych zmiennych losowych, więc ma ono również rozkład normalny $N(0, \lambda_v^v)$, gdzie

$$\lambda_t^v = (\varphi_t^{(1)})^2 \lambda_w + \lambda_e. \quad (4.22)$$

Przyjęcie normalnego rozkładu prawdopodobieństwa a priori parametru b w postaci (4.19) oraz normalny rozkład v_t w równaniu (4.21) prowadzą do wniosku, że rozkład a posteriori szacowanego parametru jest również normalny

$$b \sim N(\hat{b}_t, P_t), \quad (4.23)$$

przy czym postulujemy, że $\hat{b}_t$ i P_t są wyznaczone w oparciu o następujące zależności rekurencyjne:

$$\hat{b}_t = \hat{b}_{t-1} + K_t(y_t - \hat{b}_{t-1}\varphi_t), \quad (4.24a)$$

$$K_t = \frac{P_{t-1}\varphi_t}{\lambda_t^v + \varphi_t^2 P_{t-1}}, \quad (4.24b)$$

$$P_t = (1 - K_t\varphi_t)P_{t-1}, \quad (4.24c)$$

przy warunkach początkowych

$$\hat{b}_0 = b_0, \quad (4.25a)$$

$$P_0 = P^0, \quad (4.25b)$$

wynikających z (4.19). Dowód przeprowadzimy w oparciu o zasadę indukcji oraz skorzystamy z wzoru Bayesa w postaci

$$P(A|B, C) = \frac{P(B|A, C)P(A|C)}{P(B|C)}. \quad (4.26)$$

Zauważamy, że

$$p(b|Y_t, U_t^{(1)}, U_t^{(2)}) = p(b|y_t, Y_{t-1}, U_t^{(1)}, U_t^{(2)}). \quad (4.27)$$

Z (4.26) otrzymujemy wyrażenie na rozkład a posteriori parametru b

$$p(b|y_t, Y_{t-1}, U_t^{(1)}, U_t^{(2)}) = \frac{p(y_t|b, Y_{t-1}, U_t^{(1)}, U_t^{(2)})P(b|Y_{t-1}, U_t^{(1)}, U_t^{(2)})}{p(y_t|Y_{t-1}, U_t^{(1)}, U_t^{(2)})}. \quad (4.28)$$

Krok 1. Z warunków początkowych wynika, że

$$p(b | Y_0, U_0^{(1)}, U_0^{(2)}) = (2\pi P_0)^{-1/2} \exp[-(b - \hat{b}_0)^2 / (2P_0)]. \quad (4.29)$$

Krok 2. Zakładamy, że

$$p(b | Y_{t-1}, U_t^{(1)}, U_t^{(2)}) = (2\pi P_{t-1})^{-1/2} \exp[-(b - \hat{b}_{t-1})^2 / (2P_{t-1})]. \quad (4.30)$$

Krok 3. Wyznaczamy $p(b | Y_t, U_t^{(1)}, U_t^{(2)})$.

Korzystając z (4.21) i z faktu, że $\{v_t\}$ jest ciągiem niezależnych zmiennych losowych takich, że $Ev_t = 0$ i $Ev_t^2 = \lambda_t^v$ otrzymujemy

$$p(y_t | b, Y_{t-1}, U_t^{(1)}, U_t^{(2)}) = (2\pi\lambda_t^v)^{-1/2} \exp[-(y_t - b\varphi_t)^2 / (2\lambda_t^v)]. \quad (4.31)$$

Wstawiamy wyznaczone prawdopodobieństwa (4.30) i (4.31) do równości (4.28) otrzymując wynik kroku 3.

$$p(b | Y_t, U_t^{(1)}, U_t^{(2)}) = C_t \exp[-(y_t - b\varphi_t)^2 / (2\lambda_t^v) - (b - \hat{b}_{t-1})^2 / (2P_{t-1})]. \quad (4.32)$$

gdzie $\{C_t\}$ jest ciągiem stałych normalizujących rozkłady (4.32) i niezależnych od parametru b . Z (4.28) i wyliczonych prawdopodobieństw wynika że $C_t > 0$ dla każdego t . Wyrażenie (4.32) wygodnie jest rozpatrywać w postaci zlogarytmowanej

$$-2 \ln p(b | Y_t, U_t^{(1)}, U_t^{(2)}) = C_t' + (y_t - b\varphi_t)^2 / \lambda_t^v + (b - \hat{b}_{t-1})^2 / P_{t-1}, \quad (4.33)$$

gdzie C_t' jest nową stałą niezależną od b . Zdefiniujemy wielkość $\bar{P}^{-1}$ jako

$$\bar{P}^{-1} = P_{t-1}^{-1} + \varphi_t^2 / \lambda_t^v. \quad (4.34)$$

Rozwijamy wyrażenie (4.33) wykorzystując wielkość zdefiniowaną w (4.34). Dostajemy po przekształceniach następującą równość

$$-2 \ln p(b | Y_t, U_t^{(1)}, U_t^{(2)}) = C_t'' + (b - \bar{P}\varphi_t y_t / \lambda_t^v - \bar{P}P_{t-1}^{-1}\hat{b}_{t-1})^2 / \bar{P}, \quad (4.35)$$

4.3. Przypadek skalarny, model liniowy

gdzie C_t'' jak i poprzednio nie zależy od b . Z (4.35) wynika wniosek, że gęstość rozkładu a posteriori ma postać rozkładu normalnego $N(\bar{b}, \bar{P})$. Z definicji (4.34) wynika związek

$$\bar{P}P_{t-1}^{-1} = 1 - \frac{\bar{P}\varphi_t^2}{\lambda_t^v}, \quad (4.36)$$

który po uwzględnieniu w (4.35) i (4.34) pozwala zapisać parametry rozkładu w sposób rekurencyjny jako

$$\bar{b} = \hat{b}_{t-1} + \frac{1}{\lambda_t^v} \bar{P} \varphi_t (y_t - \varphi_t \hat{b}_{t-1}), \quad (4.37)$$

$$\bar{P} = P_{t-1} - \frac{P_{t-1}^2 \varphi_t^2}{\lambda_t^v + \varphi_t^2 P_{t-1}}, \quad (4.38)$$

a stąd po wprowadzeniu oznaczenia

$$K_t = \frac{1}{\lambda_t^v} \bar{P} \varphi_t \quad (4.39)$$

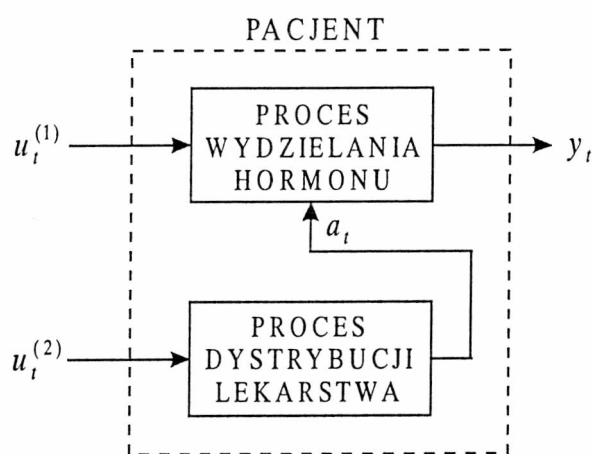
i uwzględnieniu go w (4.38) otrzymujemy tezę twierdzenia w postaci wzorów (4.24). Zależności (4.24) przypominają rekurencyjną metodę najmniejszych kwadratów. Zwróćmy uwagę na ciąg $\{\lambda_t^v\}$ w zależności (4.24) mający wpływ na własności śledzące algorytmu [48], czyli własność nadążania algorytmu za zmieniającą się dynamiką obiektu. Dotyczy to w szczególności obiektów niestacjonarnych.

4.4. Dwustopniowa identyfikacja obiektu opisanego równaniami stanu

W poprzednim rozdziale rozważaliśmy rekurencyjne algorytmy identyfikacji dyskretnego obiektu dwustopniowego, w którym na obu stopniach przyjęto modele wejściowo-wyjściowe. W wielu zagadnieniach praktycznych wygodną i zarazem najbardziej uniwersalną formą opisu matematycznego obiektów są równania stanu [33]. Ta postać opisu jest również dogodna do implementacji komputerowej. Z tego powodu nawet wtedy, gdy pierwotny model obiektu jest wyrażony za pomocą np. równania różniczkowego lub różnicowego, jest on często transformowany do opisu

4.4. Dwustopniowa identyfikacja obiektu opisanego równaniami stanu

w przestrzeni stanu w celu umożliwienia implementacji komputerowej np. do przeprowadzenia symulacji, sterowania czy podejmowania decyzji. Z drugiej strony, równania stanu są często modelem wyjściowym systemu, wynikającym z analitycznego opisu danego zjawiska, praw fizyki itp. Przykładem mogą być kompartmentowe modele systemów farmakokinetycznych [4], w których modele wyznacza się z praw bilansu masy. W takim przypadku współczynniki równań modelu wyjściowego mają często określoną interpretację fizyczną. Część z nich może być zresztą znana i fakt ten powinien być uwzględniony przy konstrukcji modelu matematycznego. Zarówno postać równań stanu, jak też znajomość pewnych ich parametrów składają się na wiedzę wstępną o modelowanym procesie. Pozostałe nieznane parametry są wyznaczane metodą eksperymentalną. Podane tu wstępne rozważania uzasadniają potrzebę opracowania identyfikacji dla obiektów dwustopniowych opisanych równaniami stanu. Obecny rozdział dotyczy identyfikacji dwustopniowego obiektu, w którym opis obiektu pierwszego stopnia przyjmuje formę równań stanu o pewnych nieznanych współczynnikach, natomiast pewna liczba tych współczynników jest zależna od innego procesu, którego opis dany jest również w postaci równań stanu. Dynamika procesu na drugim stopniu ma wpływ na pewne parametry procesu na pierwszym stopniu. Naszym celem jest identyfikacja parametrów procesów obu stopni złożonego obiektu. Wprowadzenie dwóch stopni

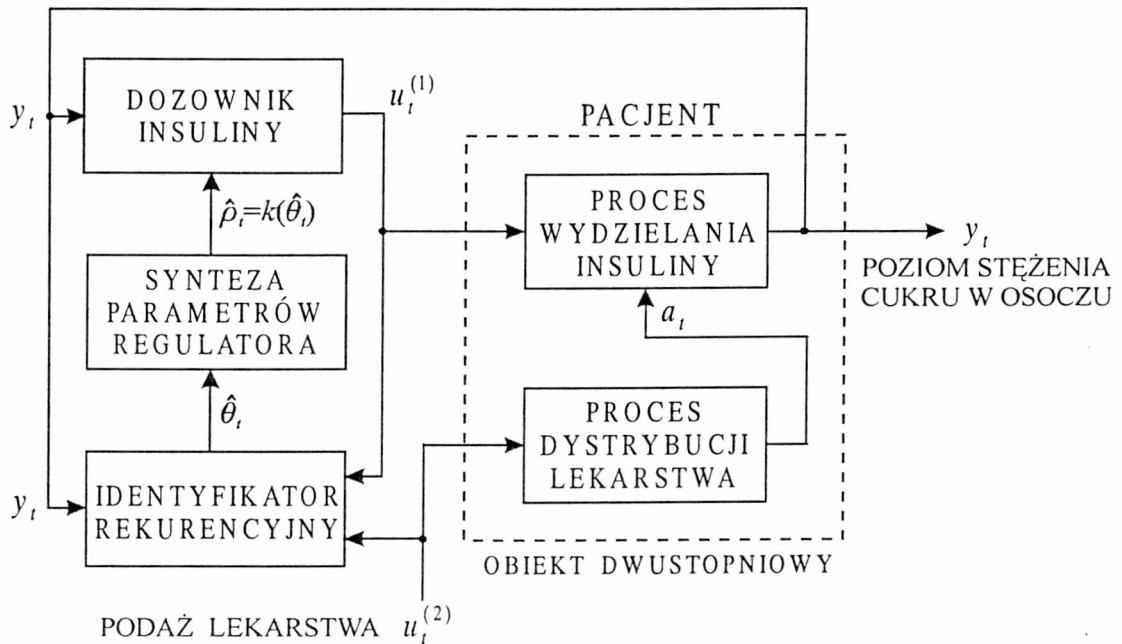


Rys. 4.5. Dwustopniowy obiekt biomedyczny (objaśnienia w tekście)

uzasadnimy na przykładzie z dziedziny biomedycyny. Chcemy zbudować

matematyczny opis pewnego procesu, którego przebieg zależy od dawki leku. Zauważamy, że lekarstwo może być podawane w różny sposób przyjmując postać jednorazowego wstrzyknięcia, wielu zastrzyków podawanych w całym cyklu leczenia, kroplówki, czy wreszcie może być podawane doustnie. Jest oczywiste, że wchłanianie leku, jego metabolizm i oddziaływanie zależą od sposobu podaży i własności osobniczych organizmu. Dystrybucja leku w organizmie jest złożonym procesem dynamicznym, którego przebieg ma wpływ na zasadniczy proces np. wydzielania pewnego hormonu regulującego jakąś funkcję w organizmie. Opisaną powyżej sytuację przedstawiono na rys. 4.5. Pierwszy stopień reprezentuje proces wydzielania hormonu insuliny w organizmie, który jest odpowiedzialny za regulację poziomu stężenia cukru we krwi y_t . W przypadku zaburzeń produkcji insuliny przez trzustkę, konieczne staje się podawanie hormonu z zewnątrz w ilości oznaczonej przez $u_t^{(1)}$. Jednocześnie pacjent otrzymuje pewne dawki leku, którego podaż oznaczono przez $u_t^{(2)}$. Podawanie tego leku ma wpływ na podstawowy proces wydzielania hormonu. Wpływ ten ma charakter dynamiczny z powodu podaży leku i jego dystrybucji, a w jego wyniku zmianom podlega pewien parametr a_t podstawowego procesu. Pozostałe parametry obu procesów zależą od indywidualnych cech organizmu. Przedmiotem zainteresowania może być określenie w jakim stopniu lekarstwo oddziałuje na proces wydzielania hormonu. Jeśli przyjąć, że zmienne modelu dwustopniowego są mierzalne, to byłoby pożądane prognozowanie poziomu stężenia cukru w osoczu z pewnym, zazwyczaj dość krótkim wyprzedzeniem, aby podjąć decyzję sterującą o dodatkowej podaży hormonu w celu kompensacji niedomagań trzustki przy jednoczesnym i koniecznym zażywaniu lekarstwa. Model taki, a więc i prognoza, powinien być wyznaczany na bieżąco, aby odpowiednio dostroić parametry regulatora w automatycznym dozowniku insuliny. Występowanie zakłóceń w pomiarach oraz skomplikowanych powiązań i współzależności między procesami zachodzącymi w organizmie dodatkowo uzasadnia potrzebę zastosowania sterowania adaptacyjnego, w którym parametry sterownika są dostrajane stosownie do niestacjonarności i odchyłeń od nominalnej dynamiki sterowanego procesu. Na rys. 4.6. przedstawiono opisany wyżej dwustopniowy obiekt biomedyczny w adaptacyjnym systemie sterowania w jednej z możliwych konfiguracji. Celem

sterowania jest utrzymanie stężenia cukru w osoczu na zadanym poziomie. Parametry regulatora $\hat{\rho}_t$ wyznaczone są w oparciu o parametry modelu $\hat{\theta}_t$ według pewnej reguły k . Parametry modelu, z kolei, estymowane są na bieżąco przez rekurencyjny



Rys. 4.6. System sterowania z dwustopniowym obiektem identyfikacji

identyfikator na podstawie obserwacji wejść i wyjść złożonego obiektu identyfikacji. Przedstawiony na rys. 4.6. system sterowania jest wysoce uproszczony i ilustruje ogólną koncepcję podejmowania decyzji z użyciem dwustopniowego modelu. W rzeczywistym układzie pojawiłoby się wiele technicznych problemów (np. związanych z pomiarami odpowiednich wielkości), a schemat ogólny z rys. 4.6. musiałby ulec znacznej rozbudowie. Przedstawiony tu przykład wskazuje na możliwość praktycznego zastosowania i celowość badania dwustopniowych algorytmów identyfikacji rekurencyjnej.

W następnym rozdziale przedstawimy ogólną metodykę konstrukcji sekwencyjnych algorytmów identyfikacji dwustopniowego obiektu opisanego w przestrzeni stanu.

4.5. Konstrukcja algorytmu dla skalarne go obiektu dwustopniowego

Przedmiotem rozważań tego rozdziału jest sformalizowane podejście do komputerowej identyfikacji na bieżąco dwustopniowego obiektu, przy czym opisy obu stopni mają postać równań stanu w postaci dyskretniej. Zakładamy, że serie pomiarowe są szeregami czasowymi w postaci (4.1) uzyskanymi w wyniku aktywnego eksperymentu w otwartym układzie identyfikacji (tzn. zakłada się brak jakichkolwiek pętli sprzężenia zwrotnego). Przyjmujemy, że proces na pierwszym stopniu opisany jest następującym ogólnym równaniem stanu

$$x_{t+1} = f_1(t, x_t, u_t^{(1)}, a_t) + w_t \quad (4.40)$$

oraz równaniem wyjścia

$$y_t = g_1(t, x_t, u_t^{(1)}, a_t) + e_t, \quad (4.41)$$

gdzie x_t jest zmienną stanu, a_t parametrem procesu, a $\{e_t\}$ i $\{w_t\}$ stanowią ciągi zakłóceń będące dyskretnymi procesami losowymi scharakteryzowanymi następującymi parametrami

$$Ew_t = 0, Ee_t = 0, \quad (4.42)$$

$$Ew_t w_t^T = r_1(a_t), \quad (4.43)$$

$$Ee_t e_t^T = r(a_t). \quad (4.44)$$

W ogólnym przypadku, jak to uwzględniono w (4.43) i (4.44), macierze kowariancji zakłóceń mogą zależeć od parametru procesu. W myśl podstawowej zasady identyfikacji dwustopniowej, parametr procesu a_t sam jest wynikiem pewnego procesu opisanego w postaci równań stanu

$$a_{t+1} = f_2(t, a_t, u_t^{(2)}, b) + v_t, \quad (4.45)$$

gdzie $u_t^{(2)}$ jest pewną wielkością wejściową procesu na drugim stopniu mającą wpływ na dynamikę zmian parametru a_t . Wielkość b jest nieznanym parametrem równania (4.45). Wyznaczenie tego parametru jest równoznaczne z określeniem wpływu $u_t^{(2)}$ na

parametr a_t . Wielkość $\{v_t\}$ jest ciągiem zakłóceń (błędem równania na drugim stopniu). Kolejne zmienne losowe ciągu są niezależne i mają jednakowe rozkłady dla każdego t . Zakładamy, że

$$Ev_t = 0, \quad (4.46)$$

$$Ev_tv_t^T = r_2(b). \quad (4.47)$$

Przyjmujemy, że analityczna postać związków funkcyjnych f_1 , f_2 i g_1 jest znana. Założmy ponadto, że znane są również parametry statystyczne sygnałów zakłócających, określonych zależnościami (4.42-44) i (4.46, 47). Naszym celem jest wyznaczenie estymatorów parametrów $\hat{a}_t$ i $\hat{b}_t$ na bieżąco dysponując danymi pomiarowymi w formie (4.1). Ogólne wyrażenia określające szukane estymatory mają następującą postać

$$\hat{a}_{t+1} = \Psi_1(\hat{a}_t, \hat{b}_t, U_t^{(1)}, U_t^{(2)}, Y_t), \quad (4.48)$$

$$\hat{b}_{t+1} = \Psi_2(\hat{a}_t, \hat{b}_t, U_t^{(2)}). \quad (4.49)$$

W zależnościach (4.48) i (4.49) występują całe serie pomiarowe, co oddaje ogólny charakter procedur. W praktyce jednak nie ma potrzeby pamiętania wszystkich pomiarów w kolejnych krokach procedury estymacji. Cała historia pomiarów jest zawarta w pewnej zmiennej pomocniczej, a do estymacji wykorzystuje się co najwyżej krótką sekwencję przeszłych pomiarów. Parametr zależności na drugim stopniu jest, jak i poprzednio, wyrażony pośrednio poprzez identyfikację na pierwszym stopniu. Jest jednak różnica między obecnym podejściem, a stosowanym w przykładzie z poprzedniego rozdziału, w którym parametr pierwszego stopnia był rugowany z równań w sposób analityczny. Podejście takie zastosowano również w monografii [71]. Obecnie estymator parametru pierwszego stopnia $\hat{a}_t$ będzie wyznaczany i występuje explicite w formule (4.49). Choć rozróżniliśmy zależności na estymatory parametrów obu stopni jako procedury Ψ_1 i Ψ_2 , to obydwie estymatory będą wyznaczone jednocześnie. Parametry obu stopni zapiszmy w postaci wektora

$$\hat{\theta}_t = \begin{bmatrix} \hat{a}_t \\ \hat{b}_t \end{bmatrix}. \quad (4.50)$$

Będziemy poszukiwać pewnej procedury identyfikacji Ψ w postaci rekurencyjnej, która wyznacza łącznie estymatory parametrów obydwu stopni. Potraktujemy parametr a_t jako zmienną stanu, a równania obu stopni (4.40) i (4.45) zapiszmy w formie następujących zależności

$$\begin{bmatrix} x_{t+1} \\ a_{t+1} \end{bmatrix} = \begin{bmatrix} f_1(t, x_t, u_t^{(1)}, a_t) \\ f_2(t, a_t, u_t^{(2)}, b) \end{bmatrix} + \begin{bmatrix} w_t \\ v_t \end{bmatrix}. \quad (4.51)$$

Otrzymaliśmy nowe równania stanu, które zapiszmy w zwartej postaci

$$\tilde{x}_{t+1} = f(t, \tilde{x}_t, u_t, b) + \tilde{w}_t, \quad (4.52)$$

gdzie

$$\tilde{x}_t = \begin{bmatrix} x_t \\ a_t \end{bmatrix}, u_t = \begin{bmatrix} u_t^{(1)} \\ u_t^{(2)} \end{bmatrix}, \tilde{w}_t = \begin{bmatrix} w_t \\ v_t \end{bmatrix}, f = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}.$$

Aby otrzymać rekurencyjny estymator nieznanego parametru b , zdefiniujmy nowy, rozszerzony wektor stanu

$$X_t = \begin{bmatrix} \tilde{x}_t \\ b \end{bmatrix} \quad (4.53)$$

i rozważmy następujące równania stanu i wyjściowe

$$X_{t+1} = \bar{F}(X_t, u_t) + \bar{w}_t, \quad (4.54)$$

$$y_t = \bar{H}(X_t, u_t) + e_t, \quad (4.55)$$

gdzie

$$\bar{F}(X_t, u_t) = \begin{bmatrix} f(t, \tilde{x}_t, u_t, b) \\ b \end{bmatrix}, \bar{w}_t = \begin{bmatrix} \tilde{w}_t \\ 0 \end{bmatrix},$$

$$\bar{H}(X_t, u_t) = g_1(t, x_t, u_t^{(1)}, a_t).$$

Rozważmy problem estymacji stanu dla systemu opisanego równaniami (4.54) i (4.55). Zagadnienie to polega na sekwencyjnym wyznaczaniu oceny stanu systemu na podstawie obserwacji wyjścia systemu

$$\hat{X}_{t+1} = E(X_{t+1} | Y_t). \quad (4.56)$$

Wielkości składowe wektora stanu X_t nie muszą być mierzalne, ani posiadać jakiegś interpretacji fizycznej. Istotny jest fakt ich korelacji z wielkością wyjściową, co daje podstawę do oszacowania bieżącego stanu układu na podstawie (4.56). Formalnie taka korelacja jest związana z pojęciem obserwowalności systemu [33]. Zakładając obserwowalność wszystkich składowych rozszerzonego wektora stanu (4.53), możemy do wyznaczenia estymatora stanu (4.56) zastosować koncepcję filtracji rekursywnej (Kalmana) [1, 58]. Zauważmy, że wyznaczając estymator $\hat{X}_t$ dany wzorem (4.56) otrzymujemy jednocześnie następujące oszacowania

$$\hat{x}_{t+1} = E(x_{t+1} | Y_t), \quad (4.57)$$

$$\hat{a}_{t+1} = E(a_{t+1} | Y_t), \quad (4.58)$$

$$\hat{b}_{t+1} = E(b_{t+1} | Y_t). \quad (4.59)$$

Problem identyfikacji dwustopniowego obiektu sprowadzono zatem do rekurencyjnej estymacji stanu (filtracji). Równania stanu (4.53) są na ogół nieliniowe i to nawet w przypadkach bardzo prostych obiektów dwustopniowych, jak to będzie widoczne w przykładzie zamieszczonym dalej. Najpowszechniej stosowanym podejściem do estymacji stanu w przypadku równań nieliniowych jest tzw. uogólniony filtr Kalmana [1], w którym linearyzuje się równania stanu w sąsiedztwie bieżącej estymaty stanu i następnie do zlinearyzowanych równań stosuje się zwykły filtr Kalmana. Równania takiego rozszerzonego filtru dane są w postaci następujących zależności rekurencyjnych, przy czym ich teoretyczne uzasadnienie można znaleźć np. w [1].

$$\hat{X}_{t+1} = \bar{F}(\hat{X}_t, u_t) + \bar{K}_t[y_t - \bar{H}(\hat{X}_t, u_t)], \quad (4.60a)$$

$$\bar{K}_t = F(\hat{X}_t, u_t) \bar{P}_t H^T (\hat{X}_t, u_t) [r_2 + H(\hat{X}_t, u_t) \bar{P}_t H^T (\hat{X}_t, u_t)]^{-1}, \quad (4.60b)$$

$$\begin{aligned} \bar{P}_{t+1} = & F(\hat{X}_t, u_t) \bar{P}_t F^T (\hat{X}_t, u_t) + R_{1,t} - F(\hat{X}_t, u_t) \bar{P}_t H^T (\hat{X}_t, u_t) \times \\ & [r_2 + H(\hat{X}_t, u_t) \bar{P}_t H^T (\hat{X}_t, u_t)]^{-1} H(\hat{X}_t, u_t) \bar{P}_t F^T (\hat{X}_t, u_t) \end{aligned}, \quad (4.60c)$$

gdzie

$$F(\hat{X}_t, u_t) = \frac{\partial}{\partial X} \bar{F}(X, u_t) \Big|_{X=\hat{X}_t}, \quad (4.61a)$$

$$H(\hat{X}_t, u_t) = \frac{\partial}{\partial X} \bar{H}(X, u_t) \Big|_{X=\hat{X}_t}, \quad (4.61b)$$

$$R_{1,t} = \bar{w}_t \bar{w}_t^T. \quad (4.62)$$

Równania (4.60) są skomplikowane, a konkretna analityczna postać algorytmu zależy od struktur modeli na pierwszym i drugim stopniu i jest konsekwencją pochodnych macierzowych określonych wzorami (4.61). Zaproponowane powyżej ogólne podejście do rekurencyjnej identyfikacji dwustopniowej zilustrujemy konkretnym przykładem. Rozpatrzmy obiekt dwustopniowy, dla którego procesy obydwu stopni są skalarne. Opis pierwszego stopnia dany jest w formie następujących równań

$$x_{t+1} = a_t x_t + w_t, \quad (4.63)$$

$$y_t = x_t + e_t. \quad (4.64)$$

Parametr równania pierwszego stopnia a_t sam jest wynikiem pewnego procesu opisanego skalarным równaniem stanu z parametrem b

$$a_{t+1} = b a_t + v_t. \quad (4.65)$$

Wielkości w_t , e_t i v_t reprezentują skalarne sygnały zakłócające o charakterze losowym tzn. $\{w_t\}$, $\{e_t\}$ i $\{v_t\}$ są ciągami zmiennych losowych o jednakowych rozkładach dla każdego t . Załóżmy, że znane są następujące charakterystyki probabilistyczne tych sygnałów

$$Ew_t = Ee_t = Ev_t = 0,$$

$$Ew_t^2 = r_1, Ev_t^2 = r_2, Ee_t^2 = r.$$

Zgodnie z przyjętą wcześniej metodyką postępowania definiujemy równania złożonego systemu dwustopniowego z rozszerzonym wektorem stanu

$$\begin{bmatrix} x_{t+1} \\ a_{t+1} \\ b_t \end{bmatrix} = \begin{bmatrix} a_t x_t \\ ba_t \\ b_t \end{bmatrix} + \begin{bmatrix} w_t \\ v_t \\ 0 \end{bmatrix}, \quad (4.66)$$

w którym jako rozszerzony wektor stanu przyjęto $X = [x \ a \ b]^T$, a odpowiedni estymator przyjmuje postać

$$\hat{X}_t = [\hat{x}_t \ \hat{a}_t \ \hat{b}_{t-1}]^T. \quad (4.67)$$

Wyznamy $F(\hat{X}_t, u_t)$ i $H(\hat{X}_t, u_t)$ dla rozpatrywanego szczególnego przypadku

$$F(\hat{X}_t, u_t) = \frac{\partial}{\partial X} \begin{bmatrix} ax \\ ba \\ b \end{bmatrix} \Big|_{X=\hat{X}_t} = \begin{bmatrix} a & x & 0 \\ 0 & b & a \\ 0 & 0 & 1 \end{bmatrix} \Big|_{X=\hat{X}_t} = \begin{bmatrix} \hat{a}_t & \hat{x}_t & 0 \\ 0 & \hat{b}_{t-1} & \hat{a}_t \\ 0 & 0 & 1 \end{bmatrix}, \quad (4.68)$$

$$H(\hat{X}_t, u_t) = \frac{\partial}{\partial X} x \Big|_{X=\hat{X}_t} = [1 \ 0 \ 0]. \quad (4.69)$$

Macierz $\bar{P}_t$ w równaniach (4.60) jest dodatnio określona i symetryczna z definicji.

Zapiszmy tą macierz w rozwiniętej postaci uwzględniając jej symetrię

$$\bar{P}_t = \begin{bmatrix} p_t^{(11)} & p_t^{(12)} & p_t^{(13)} \\ p_t^{(12)} & p_t^{(22)} & p_t^{(23)} \\ p_t^{(13)} & p_t^{(23)} & p_t^{(33)} \end{bmatrix}. \quad (4.70)$$

Przedstawmy podobnie wektor współczynników wzmocnień w równaniu (4.60a)

$$\bar{K}_t = \begin{bmatrix} k_t^{(1)} \\ k_t^{(2)} \\ k_t^{(3)} \end{bmatrix}. \quad (4.71)$$

Uwzględniając (4.60a) , (4.66) i (4.69) dostajemy zależność, według której będzie wyznaczana aktualna wartość estymaty stanu złożonego systemu

$$\hat{X}_{t+1} = \begin{bmatrix} \hat{a}_t \hat{x}_t \\ \hat{b}_{t-1} \hat{a}_t \\ \hat{b}_{t-1} \end{bmatrix} + \bar{K}_t (y_t - \hat{x}_t). \quad (4.72)$$

Wyznaczamy składowe wektora (4.71) podstawiając pochodne macierzowe (4.68) i (4.69) do zależności (4.60b) otrzymując

$$k_t^{(1)} = \frac{p_t^{(11)} \hat{a}_t + p_t^{(12)} \hat{x}_t}{r + p_t^{(11)}}, \quad (4.73a)$$

$$k_t^{(2)} = \frac{p_t^{(12)} \hat{b}_{t-1} + p_t^{(13)} \hat{a}_t}{r + p_t^{(11)}}, \quad (4.73b)$$

$$k_t^{(3)} = \frac{p_t^{(13)}}{r + p_t^{(11)}}. \quad (4.73c)$$

Następnie wyznaczamy elementy macierzy $\bar{P}_{t+1}$ korzystając z jej symetrii oraz symetrii pewnych członów zależności (4.60c). Dostajemy po uwzględnieniu wzorów (4.68) i (4.69) w zależności (4.60c) następujące formuły rekurencyjne

$$p_{t+1}^{(11)} = p_t^{(11)} \hat{a}_t^2 + 2p_t^{(12)} \hat{x}_t \hat{a}_t + p_t^{(22)} \hat{x}_t^2 + r_1 - k_t^{(1)} (p_t^{(11)} \hat{a}_t + p_t^{(12)} \hat{x}_t), \quad (4.74a)$$

$$\begin{aligned} p_{t+1}^{(12)} &= p_t^{(12)} \hat{a}_t \hat{b}_{t-1} + p_t^{(22)} \hat{x}_t \hat{b}_{t-1} + p_t^{(13)} \hat{a}_t^2 + \\ &+ p_t^{(23)} \hat{a}_t \hat{x}_t - k_t^{(1)} (p_t^{(12)} \hat{b}_{t-1} + p_t^{(13)} \hat{a}_t), \end{aligned} \quad (4.74b)$$

$$p_{t+1}^{(13)} = p_t^{(13)} \hat{a}_t + p_t^{(23)} \hat{x}_t - k_t^{(1)} p_t^{(13)}, \quad (4.74c)$$

$$p_{t+1}^{(22)} = p_t^{(22)} \hat{b}_{t-1}^2 + 2p_t^{(23)} \hat{a}_t \hat{b}_{t-1} + p_t^{(33)} \hat{a}_t^2 + r_2 - k_t^{(2)} (p_t^{(12)} \hat{b}_{t-1} + p_t^{(13)} \hat{a}_t), \quad (4.74d)$$

4.5. Konstrukcja algorytmu dla skalarnego obiektu dwustopniowego

$$p_{t+1}^{(23)} = p_t^{(23)} \hat{b}_{t-1} + p_t^{(33)} \hat{a}_t - k_t^{(2)} p_t^{(13)}, \quad (4.74e)$$

$$p_{t+1}^{(33)} = p_t^{(33)} - k_t^{(3)} p_t^{(13)}. \quad (4.74f)$$

Rozwijając wzór (4.72) dostajemy rekurencyjne formuły pozwalające na estymację stanu systemu oraz parametrów modeli na obu stopniach

$$\hat{x}_{t+1} = \hat{a}_t \hat{x}_t + k_t^{(1)} (y_t - \hat{x}_t), \quad (4.75a)$$

$$\hat{a}_{t+1} = \hat{b}_{t-1} \hat{a}_t + k_t^{(2)} (y_t - \hat{x}_t), \quad (4.75b)$$

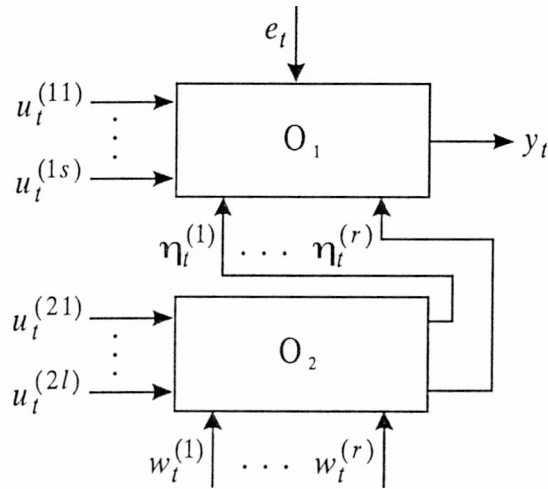
$$\hat{b}_t = \hat{b}_{t-1} + k_t^{(3)} (y_t - \hat{x}_t). \quad (4.75c)$$

W zaproponowanym algorytmie identyfikacji, na który składają się zależności (4.73), (4.74) i (4.75), dzięki zastosowaniu filtracji rekursywnej, możliwa jest zatem jednoczesna estymacja stanu i parametrów na obydwu stopniach obiektu. Nawet w tak prostym problemie jak przedstawiony w analizowanym powyżej przykładzie problem staje się nieliniowy i trzeba było zastosować aproksymację wynikającą z linearyzacji równań, która oczywiście ma wpływ na dokładność estymacji nieznanymi parametrów. Zaproponowane podejście jest cenne, ze względu na dużą uniwersalność i możliwość zastosowania do różnych struktur modeli. Jest ono ponadto rozszerzeniem zagadnień poruszanych w pracy [71], w której w ogóle nie rozważano modeli stanowych w odniesieniu do identyfikacji dwustopniowej.

4.6. Uwagi o identyfikacji obiektów wielowymiarowych

Rozważmy obecnie ogólniejszy, niż w rozdziale poprzednim, przypadek dwustopniowej identyfikacji, w której pewna grupa parametrów na pierwszym stopniu jest wynikiem identyfikacji procesu na drugim stopniu. Obiekt na drugim stopniu jest w pełni wielowymiarowy, natomiast obiekt na pierwszym stopniu jest systemem o wielu wejściach i jednym wyjściu, jak to zostało przedstawione na rys. 4.7. Zauważmy na wstępie, że próba ogólnego ujęcia algorytmów identyfikacji dla wielowymiarowych obiektów dwustopniowych niesie cały szereg problemów związanych z bogactwem możliwych struktur modeli na pierwszym i drugim stopniu. Jeśli ograniczymy się tylko do przypadku liniowych klas modeli na obu stopniach, to

i tak łączna liczba kombinacji jest bardzo liczna. Z tego powodu, celem tego rozdziału jest raczej podanie ogólnej metodyki postępowania, zamiast rozważania możliwych przypadków wynikających z kombinacji różnych modeli na obu stopniach. Dla obiektów złożonych bowiem jeszcze jaskrawiej, niż to miało miejsce dla prostych obiektów jednowymiarowych, ujawnia się trudność podejścia algorytmicznego



Rys. 4.7. Dwustopniowy obiekt wielowejściowy

i stosowania ściśle określonych, utartych schematów postępowania. By uzyskać rekurencyjną procedurę identyfikacji dwustopniowej rozpatrujemy łącznie zestaw równań opisujących łącznie systemy obu stopni i staramy się na ich podstawie uzyskać strukturę modelu dla całego złożonego obiektu w postaci dogodnej do zastosowania którejś ze znanych metod sekwencyjnej estymacji parametrów. Celowe jest wykorzystanie jednej z opisanych wcześniej metod. Jeśli równania modelu są uwikłane uniemożliwiając wyznaczenie parametrów η_t , to rozsądnym rozwiązaniem jest zamiana równań modelu z postaci wejściowo-wyjściowej do stanowej i postępowanie analogiczne do opisanego w rozdziale 4.4. W prostszych przypadkach postępuje się zgodnie z metodą opisaną w rozdziale 4.3. Nie jest możliwe podanie ogólnej metody postępowania dla dwustopniowych obiektów wielowymiarowych.

5. Własności złożonych algorytmów identyfikacji

Obecnie zajmujemy się badaniami symulacyjnymi algorytmów omawianych w poprzednich rozdziałach. Ograniczymy się do dyskretnych obiektów jednowymiarowych o zadanych charakterystykach, które zbadamy w warunkach deterministycznych, a następnie w przypadku działania addytywnych zakłóceń opisanych przez model z rys. 2.1. W eksperymentach wzięto pod uwagę różne wartości progowe kryterium rekonstrukcji modelu opisanego zależnością (3.7). W symulacjach komputerowych wykorzystano pakiet MATLAB.

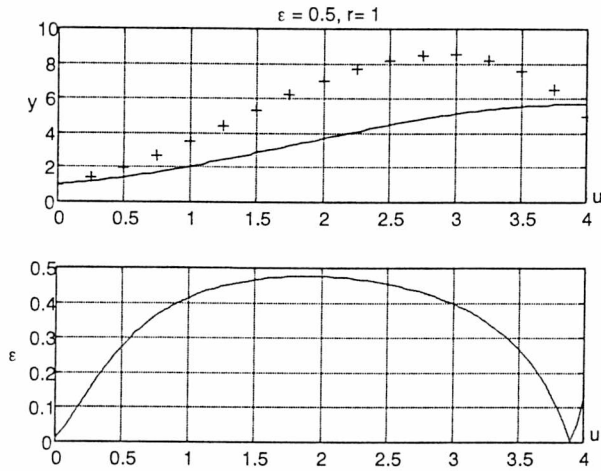
5.1. Przypadek deterministyczny

Eksperyment dotyczy algorytmu podstawowego, o strukturze przedstawionej na rys. 3.1., którego podstawy metodologiczne opisano w rozdz. (3.4), w przypadku gdy na układ nie działają zakłócenia losowe. Opis poszczególnych kroków procedury opisano w algorytmie 3.1. Podstawowym celem badań była eksperymentalna weryfikacja zdolności aproksymacyjnych algorytmu 3.1. oraz jego wrażliwość na parametry wejściowe decydujące o rekonstrukcji modelu. Streścimy w punktach podstawowe warunki eksperymentu.

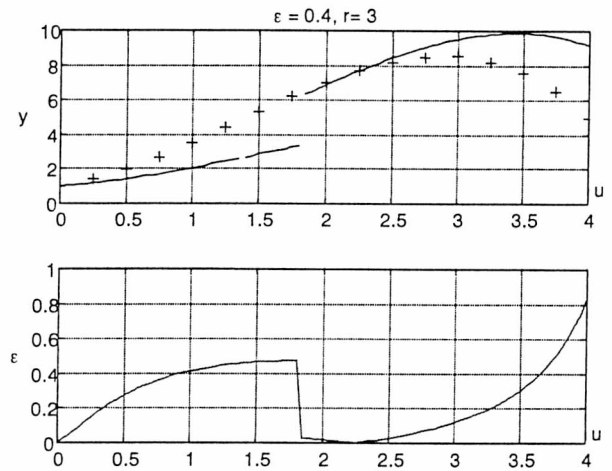
- Obiekt dyskretny, jednowymiarowy o zadanej charakterystyce statycznej
- Długość ciągu uczącego $n = 81$
- Próbkowanie jednostajne, tzn. pomiary w stałych odstępach czasu
- Postać modelu: regresja liniowa o wielomianowych, liniowo niezależnych funkcjach rozpinających, tzn. składowe wektora (212) są postaci $\varphi_t^{(k)} = u_t^{(k-1)}$, $k = 1, \dots, r$
- Maksymalny rząd modelu $r_{max} = 5$
- Kryterium identyfikacji – średniokwadratowe, wzór (3.26)
- Warunkowe kryterium rekonstrukcji modelu oparte na pojęciu zbieżności jednostajnej, zależność (3.7)
- Wartości progowe kryterium (3.7) wybrano ze zbioru $\varepsilon = \{ 0,5; 0,1; 0,05; 0,01 \}$
- Sprawdzanie kryterium rekonstrukcji modelu co $N_V = 10$ próbek (patrz krok 3. algorytmu 3.1.).

5.1. Przypadek deterministyczny

Wyniki badań przedstawiono na wykresach rys. 5.1-4. Górne wykresy przedstawiają dane pomiarowe (linia w postaci krzyżyków), oraz uzyskany model

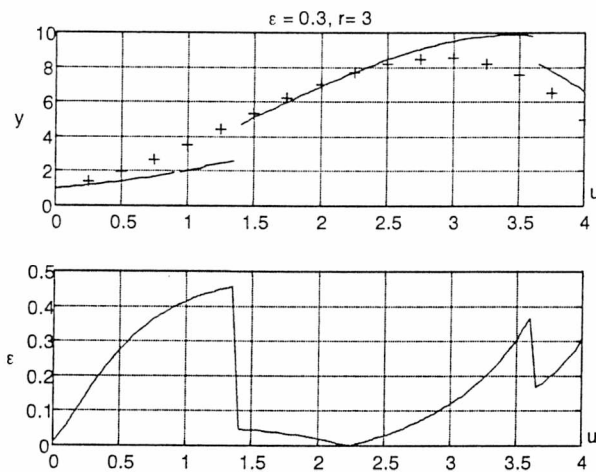


Rys. 5.1

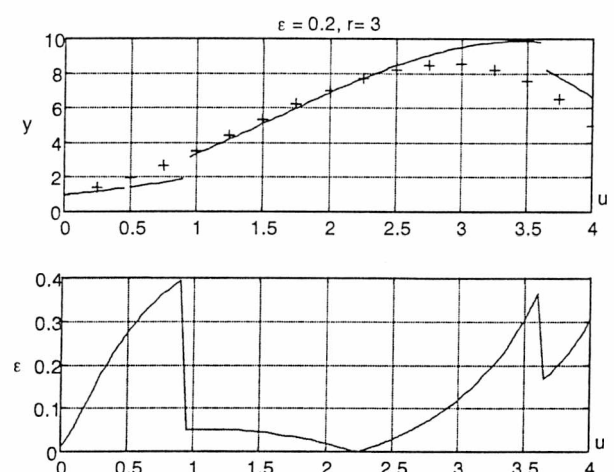


Rys. 5.2

(linia ciągła). Dolne wykresy obrazują wartość błędu bezwzględnego uzyskanego modelu równą liczbowo wartości kryterium (3.7). By rysunki były bardziej



Rys. 5.3

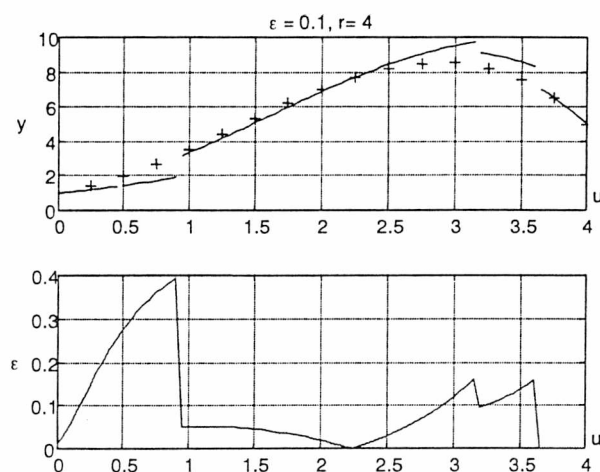


Rys. 5.4

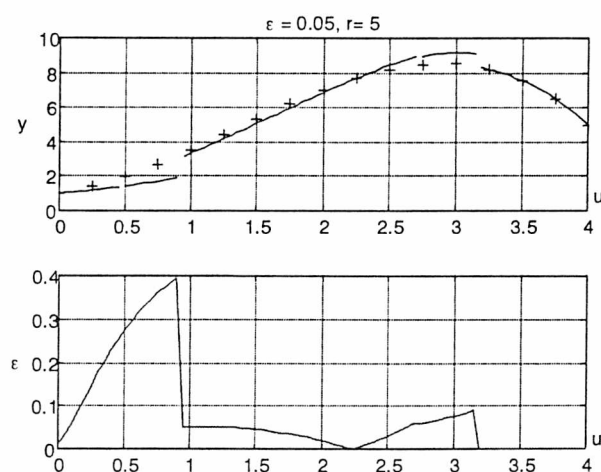
przejrzyste, zamieszczono tylko niektóre pomiary z ciągu uczącego. Kryterium (3.7), mające charakter predykatu logicznego wprowadza nieliniowość do algorytmu, co uwidacznia się w skokowym charakterze modelu (linia ciągła na wykresach) i przebiegu błędu (dolne wykresy).

5.1. Przypadek deterministyczny

Charakterystyczne nieciągłości na wykresach występują w momentach przebudowy struktury modelu. W przypadkach przedstawionych na rys. 5.1–7 w punktach tych

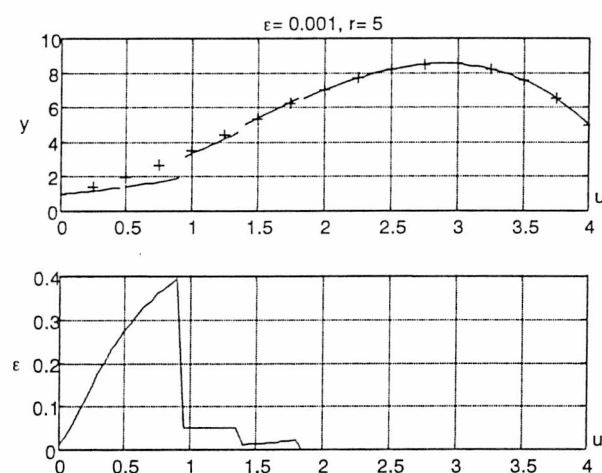


Rys. 5.5.



Rys. 5.6

występuje na ogół skokowa poprawa jakości, co widoczne jest na dolnych częściach rysunków obrazujących błąd względny modelu. Z budowy algorytmu wynika



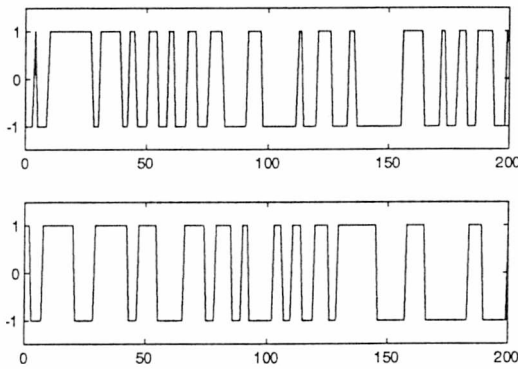
Rys. 5.7

wrażliwość metody na zakłócenia o charakterze losowym. W przypadku silnie zakłóconych obserwacji zbyt niska wartość progowa spowoduje nadmiernie częste rekonstrukcje modelu. Dobór poziomu progu przed eksperymentem jest jednak utrudniony, gdyż natura zakłóceń nie jest na ogół znana dostatecznie precyzyjnie. Do zmniejszenia wrażliwości algorytmu na zakłócenia można wykorzystać filtrację

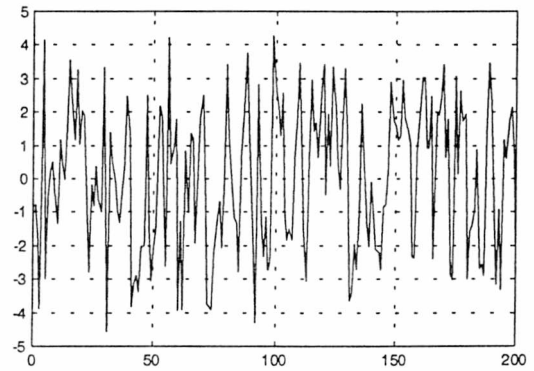
sygnałów pomiarowych, którą można z kolei uzyskać stosując kryteria dane zależnościami 3.16-24.

5.2. Identyfikacja w obecności zakłóceń

Zbadamy obecnie metodą symulacji komputerowej algorytm dwustopniowy w formie wejściowo-wyjściowej omówiony szczegółowo w rozdziale 4.3. Analizie poddano model o następującej strukturze danej wzorem (4.17) i (4.18), przy czym $\varphi_t^{(1)} = u_t^{(1)}$, a $\varphi_t^{(2)} = u_t^{(2)}$. Celem eksperymentu była pośrednia identyfikacja parametru b , przy założeniu, że nie można bezpośrednio zmierzyć przebiegu wyjścia obiektu na drugim stopniu. Identyfikację przeprowadzono generując sygnały wejściowe dla pierwszego i drugiego stopnia w postaci losowych ciągów



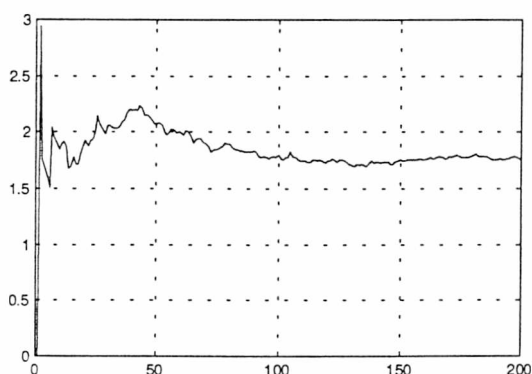
Rys. 5.8. Sygnały wejściowe



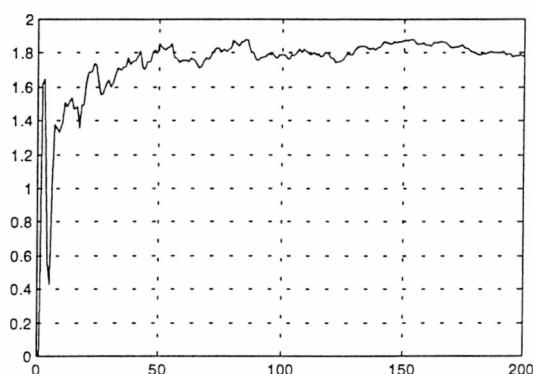
Rys. 5.9. Sygnał wyjściowy

binarnych. Liczność serii wynosiła 200. Sygnał zakłócający miał postać losowego sygnału o rozkładzie normalnym o zerowej średniej i wariancji jednostkowej. Przebiegi wejściowe przedstawiono na rys. 5.8, a przebieg sygnału wyjściowego na rys. 5.9. Do wyznaczenia oceny parametru b zastosowano algorytm rekurencyjny wyprowadzony w rozdziale (4.3) i określony zależnościami (4.24). Wynik jego działania w formie przebiegu parametru b przedstawiono na rys. (5.10,11). Przyjęto, że wielkość λ_t^v we wzorze (4.24) jest wielkością stałą i bliską jedności. Pokazano przebiegi dla $\lambda_t^v = 1$ (rys. 5.10) oraz dla $\lambda_t^v = 0,98$ (rys. 5.11). Przy malejących wartościach λ_t^v model staje się bardziej podatny na zmiany parametrów na pierwszym stopniu (niestacjonarność). Znajduje to odzwierciedlenie na rys. 5.11, który w porównaniu na rys. 5.10 jest silniej zakłócony – jest to wpływ zmian

5.2. Identyfikacja w obecności zakłóceń



Rys. 5.10. Adaptacja parametru ($\lambda = 1$)



Rys. 5.11. Adaptacja parametru ($\lambda = 0,98$)

w czasie parametru a . Wnioski wynikające z wcześniejszych rozważań teoretycznych znalazły potwierdzenie w badaniach symulacyjnych. Przedstawiony model jest bardzo prosty lecz pozwala przypuszczać, że w bardziej złożonych przypadkach metoda ta może dać obiecujące rezultaty. Trudno jest jednak podać ogólnie analityczne zależności dla bardziej złożonych modeli, ponieważ zależność między parametrami obydwu stopni może być uwikłana w sposób uniemożliwiający wyrugowanie wyjść obiektu drugiego stopnia z zależności na pierwszym stopniu, jak to uczyniliśmy dla naszego prostego modelu. Pewnym wyjściem jest sprowadzenie równań wejściowo-wyjściowych do równań w przestrzeni stanu, co doprowadza do większej wymiarowości problemu i na ogół większej liczby parametrów. Ogólnie zarysowano ten problem w rozdziale 4.6.

6. Zastosowania złożonych algorytmów rekurencyjnych

Przedstawimy zastosowania praktyczne zaproponowanych wcześniej algorytmów identyfikacji. W rozdziale 6.1 omawiamy zastosowanie algorytmu w identyfikacji przykładowego procesu farmakokinetycznego. Rozdział 6.2 omawia wykorzystanie algorytmu w modelowaniu procesu hemodializy. Jako przykład techniczny, rozważamy w rozdziale 6.3 proces odsiarczania węgla.

6.1. Identyfikacja procesu farmakokinetycznego

Jednym z bardziej interesujących z praktycznego punktu widzenia zadań modelowania obiektów biomedycznych jest modelowanie procesów farmakokinetycznych. Modele farmakokinetyczne bowiem, opisujące metodami matematycznymi zjawiska rozprzestrzeniania się leku w organizmie, umożliwiają obliczanie i analizowanie czasowych przebiegów stężeń leku w wybranych obszarach i tym samym określenie właściwego dozowania leku. W wielu sytuacjach proste modele kinetyki leków, opisane jednym lub kilkoma równaniami różniczkowymi, okazały się bardzo przydatne w farmakoterapii i są od dawna stosowane w praktyce leczniczej. Jest tak na przykład w przypadku procesów dystrybucji niektórych cytostatyków i leków nasercowych [4], jak również – sięgając do trochę odmiennych zagadnień – przy opisie rozprzestrzeniania się toksyn mocznicowych (mocznik, kwas moczowy, kreatynina) [64], przydatnym między innymi do komputerowego nadzoru procesu dializy [38]. Jednak również w wielu przypadkach proste modele, nie ujmujące w pełni złożoności procesów uwalniania i rozpuszczania leków w płynach ustrojowych, wchłaniania, przechodzenia leku do krwiobiegu i jego transportu, czy wreszcie procesów jego eliminacji i wiązania z białkami, są zbyt daleko idącym kompromisem i uproszczeniem, nie znajdującym wymaganej akceptacji klinicznej.

Przedstawiony tu przykład, omawiany w pracy [1], dotyczy identyfikacji procesu rozprzestrzeniania propranololu podanego donaczyniowo i może być interesującą alternatywą przede wszystkim dla przypadków złożonych (modele wielokompartментowe, modele perfuzyjne, uwzględnienie nieliniowości, itp.), w których tradycyjne metody identyfikacji obiektów napotykają na poważne trudności koncepcyjne i obliczeniowe [21].

Propranolol stosowany jest głównie w leczeniu nadciśnienia tętniczego, zaburzeń rytmu serca i dusznicy bolesnej.

6.1.1. Sformułowanie zadania

Model procesu farmakokinetycznego, niezależnie od sposobu jego uzyskania, ma postać równań stanu:

$$\dot{m}(t) = f(m(t), d(t), \alpha), \quad (6.1a)$$

$$y(t) = h(m(t), \beta), \quad (6.1b)$$

w których wektor stanu $m(t)$ oznacza masy leku w poszczególnych przedziałach, sterowanie $d(t)$ ma sens strumienia dozowania leku, zależnego od sposobu jego podaży, l – wymiarowy wektor wielkości wyjściowych $y(t)$ stanowi stężenia leku w poszczególnych kompartmentach, natomiast α i β oznaczają wektory nieznanych parametrów, związane z osobniczymi właściwościami procesu dystrybucji (np. objętości kompartmentów, stałe eliminacji, stałe wymiany pomiędzy przedziałami, stała wchłaniania leku).

W klasycznej procedurze identyfikacji parametrów α i β [9] należy w pierwszej kolejności rozwiązać równania (6.1) dla pewnego sterowania (strumienia podaży leku) $d^*(t)$ i warunków początkowych $m(0)$

$$y(t) = \varphi(d^*(t), m(0), t, \alpha, \beta), \quad (6.2)$$

a następnie dla tych samych $d^*(t)$ i $m(0)$ zmierzyć wyjścia obiektu w określonych chwilach:

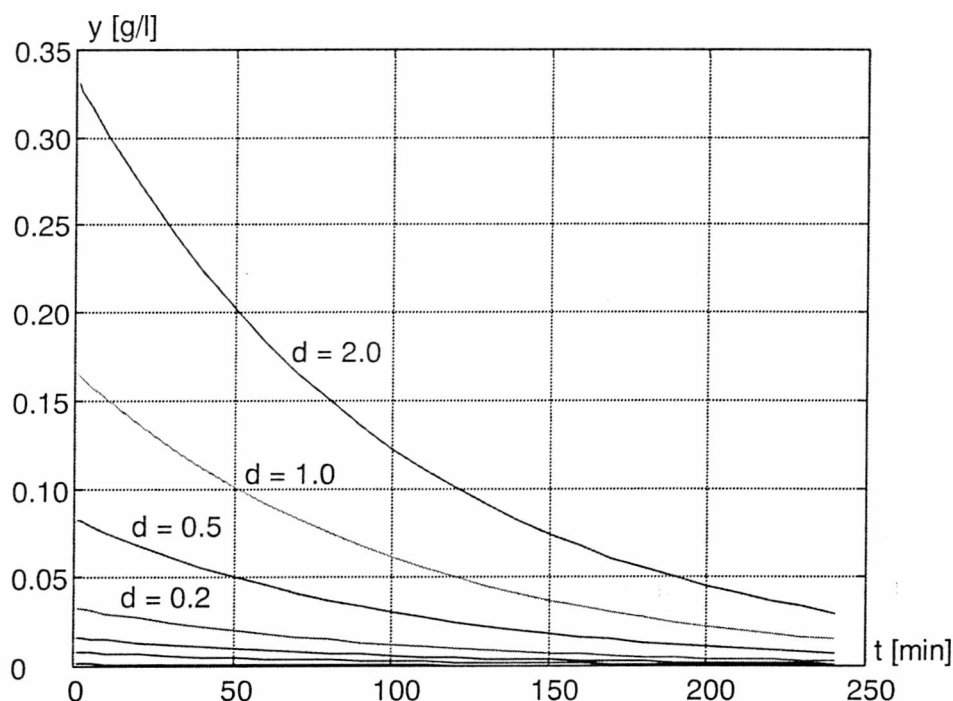
$$y(t_1), y(t_2), \dots, y(t_n). \quad (6.3)$$

Poszukiwany rezultat otrzymamy rozwiązując odpowiedni układ równań wynikający z przyrównania (6.2) i (6.3) dla tylu kolejnych chwil czasu, ile jest nieznanych parametrów. Zasadnicza trudność tej koncepcyjnie prostej procedury polega na doborze takiego sygnału wejściowego $d^*(t)$, który umożliwiłby analityczne rozwiązanie równań (6.1), a ponadto był fizycznie realizowalny, tak by można było podać go na wejście obiektu. W zadaniu identyfikacji procesów farmakokinetycznych trudność ta jest dodatkowo spotęgowana narzuconą formą strumienia dozowania,

wynikającą z formalizacji konkretnych sposobów podaży leku. Z drugiej jednak strony, konkretna forma strumienia dozowania pozwala sparametryzować sygnał $d^*(t)$, to znaczy uznać iż stężenie leku – dla danego sposobu podaży – zależy od parametrów tej podaży (np. dla wstrzyknięcia od dawki leku, dla wlewu od prędkości wlewu, dla podaży wielokrotnej od czasu pomiędzy kolejnymi zastrzykami itp.), a tym samym sprowadzić zadanie identyfikacji do przypadku statycznego.

6.1.2. Dane pomiarowe

Przeprowadzono pomiary stężenia propranololu w czasie 250 min dla siedmiu różnych dawek początkowych leku uzyskując siedem serii pomiarowych. Każda seria liczyła łącznie 27 pomiarów wykonanych dla momentów czasu $t = 1, 2, 5, 10, 20, \dots, 240$. Dane pomiarowe przedstawiono na rys. 6.1. w postaci rodziny wykresów.



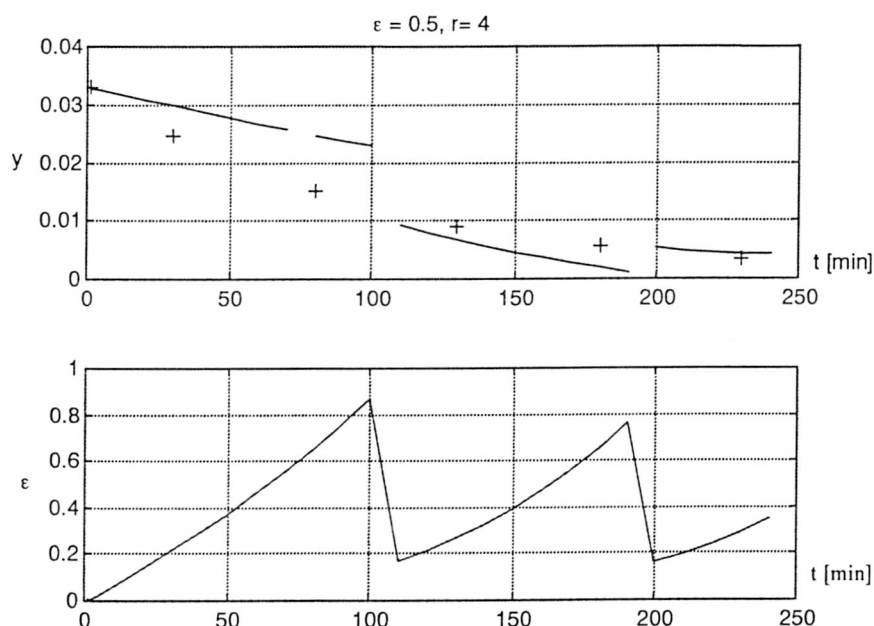
Rys. 6.1. Dane pomiarowe

Każda linia na rys 6.1. oznacza stężenie leku w osoczu w g/l w zależności od czasu. Dawki początkowe propranololu d zaznaczono nad pierwszymi czterema liniami wykresów. Pełny zbiór dawek początkowych zawierał następujące wartości: $D = \{ 0.01, 0.05, 0.1, 0.2, 0.5, 1.0, 2.0 \}$.

Dla każdej serii pomiarowej przeprowadzono identyfikację obiektu przyjmując różne parametry wejściowe algorytmu. Wobec założenia o braku znajomości kształtu charakterystyki obiektu posłużono się modelem regresji liniowej z funkcjami rozpinającymi w postaci jednomianów coraz wyższych stopni. Przyjęty model ma dobre własności aproksymujące w szerokiej klasie charakterystyk.

6.1.3. Wyniki identyfikacji

Na rys. 6.2 przedstawiono wyniki identyfikacji dla wybranej charakterystyki stężenia propranololu. Na górnych wykresach poszczególnych rysunków przedstawiono wyjście modelu (linia ciągła) oraz kilka punktów pomiarowych (linia w postaci krzyżyków). Dolna część rysunku ilustruje błąd względny modelu



Rys. 6.2. Wyniki identyfikacji stężenia propranololu

w stosunku do danych pomiarowych. Wszystkie eksperyment przeprowadzono badając kryterium (3.7) co $N_V = 4$ pomiary. Tak niewielką liczbę przyjęto wobec krótkich serii pomiarowych, jakimi dysponowano. Wejściowe parametry algorytmu ϵ i N_V mają decydujący wpływ na jakość modelu.

6.2. Modelowanie i identyfikacja stężenia mocznika w procesie dializy

Rozdział stanowi ilustrację praktycznego zastosowania rekurencyjnej identyfikacji dwustopniowej do opisu procesu usuwania mocznika z organizmu w czasie trwania

dializy. W literaturze możemy spotkać rozmaite modele analityczne w formie równań różniczkowych, które formalnie są opisem odpowiednich modeli kompartmentowych dla różnych toksyn, które podlegają usuwaniu w procesie dializy [64, 65]. Proponowany w pracy system identyfikacji, zbudowany w oparciu o analityczne opisy znane z literatury, uwzględnia bieżące dane pomiarowe, powinien więc dać dokładniejszy model eliminacji mocznika z organizmu.

Od czasu pojawienia się pierwszych propozycji modeli stosowanych przy opisie dializy nastąpił znaczny postęp w technice pomiarowej i mikroprocesorowej, co rokuje nadzieję na zastosowanie praktyczne bardziej wyrafinowanych algorytmów wspomagających terapię hemodializy.

6.2.1. Charakterystyka problemu

Dializa jest procesem pozaustrojowego oczyszczania krwi z różnych toksycznych substancji, które gromadzą się w osoczu w stanach patologicznych, gdy zawodzą naturalne fizjologiczne mechanizmy usuwania z organizmu produktów przemiany materii. Proces dializy zachodzi w urządzeniu zwanym dializatorem lub sztuczną nerką*, której głównym elementem jest wymiennik masy w postaci syntetycznej błony o odpowiednio dużej powierzchni aktywnej. Po jednej stronie tej błony przepływa krew pacjenta, a po drugiej płyn dializujący o odpowiednim składzie. Następuje transport szkodliwych substancji z krwi do dializatu przez błonę. Wielkość porów w błonie decyduje o rodzaju i liczbie cząsteczek przechodzących do płynu dializacyjnego. W membranowych wymiennikach masy w transporcie cząsteczek biorą udział różne rodzaje zjawisk, lecz najistotniejszą rolę odgrywa transport dyfuzyjny i konwekcyjny. Czynnikiem dyfuzji jest różnica stężeń substancji, natomiast transport konwekcyjny wynika z różnicy ciśnień po obu stronach membrany. W hemodializie, która jest przedmiotem naszych rozważań, główną rolę odgrywa dyfuzja. W jej wyniku usuwane są z osocza niepożądane toksyny jak: mocznik, kreatynina oraz uzupełniane są braki niektórych związków (jonów potasu, wapnia, wodorowęglanów). Zabiegowi temu towarzyszy tzw. ultrafiltracja wywołana wymuszoną różnicą ciśnień, która sprzyja usuwaniu nadmiaru wody z ustroju

* Przeprowadza się również dializę wewnątrzuustrojową, w której jako błony dializacyjnej używa się otrzewnej, a płyn dializujący wlewa się poprzez ściany brzucha do jamy otrzewnej.

powodując zmianę objętości kompartmentu procesu. Charakterystyka aspektów medycznych terapii hemodializy została przedstawiona w bardzo ogólnym zarysie. Bardziej szczegółowy opis można znaleźć w pracach [57,39, 40].

Wskazaniem do leczenia dializą są głównie stany ostrej i przewlekłej niewydolności nerek. W wielu wypadkach leczenie to ma intensywny charakter, co oznacza kilkukrotne dializowanie w ciągu tygodnia, a pojedynczy zabieg może trwać od czterech do pięciu godzin. Związane z tym wysokie koszty leczenia powodują dążenie do optymalnego wykorzystania stacji dializ poprzez indywidualny dobór parametrów dializy dla poszczególnych pacjentów tak, by zarówno czas trwania dializy jak i okres między zabiegami były dobrane optymalnie dla każdego pacjenta. Stąd wynika potrzeba ilościowego ujęcia procesu dializy w formie modelu matematycznego, który umożliwiłby bardziej precyzyjne określenie tych parametrów dializoterapii, na które bezpośredni wpływ ma lekarz. Pierwsze modele matematyczne przebiegu dializy pojawiły się w latach sześćdziesiątych. Od tego czasu powstało wiele różnych modeli opisujących kinetykę takich substancji jak mocznik, kreatynina, sód, potas, dwuwęglany i heparyna. Modele te miały charakter wielokompartментowy. Ze względu na dużą liczbę parametrów trudnych do wyznaczenia u indywidualnego pacjenta nie znalazły one szerszego zastosowania w praktyce klinicznej, umożliwiły natomiast analizę procesów w czasie prowadzenia dializy [57]. Poszukiwania modeli nadających się do praktycznego stosowania doprowadziły do uproszczenia istniejących rozwiązań i przyjęcia jednego kompartmentu reprezentującego całą wodę organizmu* [64]. Szczególnie duże zastosowanie uzyskał jednokompartментowy model mocznikowy [66, 67, 51]. Mocznik jest jedną z najważniejszych toksyn, której poziom w osoczu decyduje o harmonogramie prowadzenia dializy u indywidualnego pacjenta. W kolejnym rozdziale zbudujemy hierarchiczny dwustopniowy model parametryczny dla mocznika przyjmując jako podstawę klasyczny model jednokompartментowy.

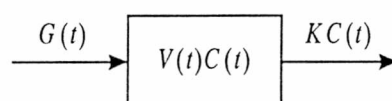
* ang. Total Body Water

6.2.2. Model poziomu mocznika w procesie hemodializy

Ogólna formuła dla stężenia mocznika może być wyrażona zależnością

$$C(t) = f(G, V, t, K_r, K_d), \quad (6.4)$$

gdzie $C(t)$ [mg/ml] jest stężeniem mocznika w momencie t , G [mg/min] oznacza szybkość generacji mocznika w organizmie, $V(t)$ [ml] objętość kompartmentu (całkowita ilość wody w ustroju), K_r [ml/min] klirens resztkowy nerek, K_d [ml/min] klirens dializatora. Klirens oznacza objętość, która jest całkowicie oczyszczona z danej



Rys. 6.3. Jednokompartментowy model mocznikowy

substancji w jednostce czasu. Jest on ważnym parametrem używanym przez lekarzy zarówno do oceny pracy nerek (K_r) jak i dializatora (K_d). Zauważmy, że w zależności (6.4) parametry G , V , K_r charakteryzują pacjenta, natomiast K_d i t podlegają wyborowi przez lekarza. Proces dializy przedstawiono w formie modelu kompartmentowego na rys. 6.3. Z bilansu masy wynika następujący model w postaci równania różniczkowego

$$\frac{d}{dt}(VC) = G - KC, \quad (6.5)$$

gdzie całkowity klirens $K = K_r + K_d$ wynika z resztkowej funkcji nerek i działania dializatora. Model (6.5) opisuje stężenie mocznika zarówno w czasie trwania dializy, jak i w okresie między dializami (wówczas $K_d = 0$) oraz w stanie ustalonym ($G = KC$). W dalszej dyskusji ograniczymy się do analizy przypadku opisującego proces dializy (tzn. $K_d \neq 0$). Objętość kompartmentu jak to wynika ze wzoru (6.5) może się zmieniać w czasie w wyniku wspomnianej wstępnie ultrafiltracji, która powoduje odprowadzenie zbędnej wody z organizmu. W modelach rozpatrywanych w literaturze przyjmuje się liniową zmienność objętości kompartmentu według zależności

$$V(t) = V_0 + \beta t, \quad (6.6)$$

gdzie V_0 jest początkową objętością kompartmentu, a β jest pewnym współczynnikiem, którego wartość jest stała w czasie trwania dializy. Jeśli przyjmiemy, że objętość kompartmentu zmienia się w czasie zgodnie z zależnością (6.6), to równanie (6.5) stanie się niejednorodne, lecz wciąż możliwe do rozwiązania analitycznego, które ma w takim przypadku postać

$$C(t) = C_0 \left(\frac{V_0 + \beta t}{V_0} \right)^{-\frac{K+\beta}{\beta}} + \frac{G}{K + \beta} \left[1 - \left(\frac{V_0 + \beta t}{V_0} \right)^{-\frac{K+\beta}{\beta}} \right]. \quad (6.7)$$

Wyrażenie (6.7) jest ogólną zależnością na stężenie mocznika w przypadku, gdy obecna jest pewna reszkowa funkcja nerek z jednoczesnym pozaustrojowym usuwaniem tej toksyny. Jeśli znane są parametry tego modelu, to można oszacować np. czas trwania zabiegu lub długość okresu między dializami tak, by poziom stężenia mocznika nie przekroczył zadanej wartości. Z zależności (6.7) wynika, że czas dializy Δt wymagany do zmniejszenia poziomu mocznika do wartości $C_{\Delta t}$ można wyrazić następująco

$$\Delta t = \frac{V_0}{\beta} \left[\left(\frac{C_{\Delta t} - G/(\beta + K)}{C_0 - G/(\beta + K)} \right)^{-\frac{\beta}{\beta+K}} - 1 \right]. \quad (6.8)$$

Model mocznikowy (6.7) był stosowany w praktyce klinicznej [64,65,66], więc w dalszych rozważaniach przyjmujemy, iż oddaje on z dobrym przybliżeniem charakter dynamiki procesu dializy. Nie wykorzystuje on jednak danych pomiarowych napływających w czasie trwania zabiegu. W pracy [11] zaproponowano podejście do modelowania procesu hemodializy, które można określić jako identyfikację z wiedzą wstępną. Polega ono na przyjęciu poprawności modelu analitycznego stosowanego do tej pory w literaturze i takim jego przekształceniu, by nadawał się do celów identyfikacji. Rozpatrzono dwa przypadki: ze stałą i zmieniającą się objętością kompartmentu. W drugim przypadku zaproponowano zastosowanie identyfikacji dwustopniowej.

6.2.2.1. Model ze stałą objętością kompartmentu

Zakładamy, że $V(t) = \text{const}$. Równanie (6.5) stanowi podstawę do zbudowania opisu systemu w formie równań stanu w postaci dyskretnej

$$C_{t+1} = A_T(a)C_t + B_T(a)G_t, \quad (6.9a)$$

$$y_t = C_T(a)C_t + e_t, \quad (6.9b)$$

gdzie y_t oznacza wyjściową koncentrację mocznika, e_t jest sygnałem zakłócającym, a T oznacza okres próbkowania. Wielkości $A_T(a)$, $B_T(a)$, $C_T(a)$ zależą od wektora parametrów a i oznaczają odpowiednie macierze równań stanu. Wielkości te wyznaczymy na podstawie równania (6.5) w następujący sposób:

$$A_T(a) = e^{-KT/V}, \quad (6.10a)$$

$$B_T(a) = \int_0^T e^{-KT\tau/V} d\tau = -(1 - e^{-KT/V}), \quad (6.10b)$$

$$C_T(a) = 1. \quad (6.10c)$$

Szczegółowe wzory zamiany z postaci ciągłej (6.5) na dyskretną (6.9) można znaleźć w [2]. Dyskretna postać równań stanu (6.9) pozwala na przejście do formy wejściowo-wyjściowej

$$y_t = G_T(q, a)G_t + e_t, \quad (6.11)$$

przy czym $G_T(q, a)$ jest dyskretną funkcją transmitancji, która jest z kolei wymierną funkcją operatora q (np. $qG_t = G_{t+1}$). Funkcję tę możemy wyrazić poprzez wielkości (6.10) następująco

$$\begin{aligned} G_T(q, a) &= C_T(a)[q - A_T(a)]^{-1}B_T(a) = \\ &= \frac{1 - e^{-KT/V}}{K(q - e^{-KT/V})} \end{aligned} \quad (6.12)$$

Uwzględniając (6.12) w zależności (6.11) dostajemy

$$y_t - e^{-KT/V} y_{t-1} = K^{-1}(1 - e^{-KT/V})G_{t-1} + e_t - e^{-KT/V} e_{t-1}. \quad (6.13)$$

Parametry modelu w równaniu (6.13) są powiązane nieliniowymi zależnościami i z tego powodu trudno jest wyznaczyć bezpośrednio ich oceny. Po reparametryzacji w równaniu (6.13) dostaniemy strukturę modelu dogodną do celów identyfikacji

$$\hat{y}_{t|a} = a_1 y_{t-1} + a_2 G_{t-1} + a_3 e_{t-1}. \quad (6.14)$$

Parametry modelu (6.14) nie mają interpretacji fizycznej lecz ich wyznaczenie służy wyłącznie jak najlepszemu dopasowaniu modelu (6.14) do danych pomiarowych. Liniowy charakter modelu (6.14) pozwala zastosować wiele standardowych metod estymacji włącznie z metodą najmniejszych kwadratów. Otrzymana postać modelu nie została zaproponowana arbitralnie, lecz wyprowadzona w oparciu o pewne omówione wcześniej przesłanki teoretyczne i praktyczne.

6.2.2.2. Model ze zmienną objętością kompartmentu

W omówionym poprzednio wstępnym przykładzie założono stałą objętość całkowitej wody organizmu. Wiadomo jednak, że podczas zabiegu hemodializy wskutek ultrafiltracji następuje ubytek wody z ustroju, co jest równoznaczne ze zmianą objętości kompartmentu modelu. W celu uzyskania dyskretnego modelu zastosujemy przybliżenie Eulera pochodnej w równaniu (6.5)

$$\frac{C(t+T)V(t+T) - C(t)V(t)}{T} = G(t) - KC(t) + e(t), \quad (6.15)$$

przy czym $t = kT$, ($k = 0, 1, \dots$). Zakłócenie e_t jest błędem równania i zawiera w sobie nie tylko składową losową, lecz również ewentualne błędy aproksymacji pochodnej oraz błędy pomiarowe. Przyjmując T za jednostkę czasu otrzymujemy równanie w postaci dyskretniej

$$C_t = \frac{V_{t-1}C_{t-1}}{V_t} - TK \frac{C_{t-1}}{V_t} + T \frac{G_{t-1}}{V_t} + T e_{t-1}. \quad (6.16)$$

Z równania (6.16) wynika struktura modelu

$$\hat{y}_{t|a} = a_1 \frac{V_{t-1}y_{t-1}}{V_t} + a_2 \frac{y_{t-1}}{V_t} + a_3 \frac{G_{t-1}}{V_t} + a_4 e_{t-1}, \quad (6.17)$$

która ma, jak łatwo zauważyć, postać regresji liniowej $\hat{y}_{t|a} = a^T \varphi_t$, przy czym wektor regresji φ_t wiąże w sposób nieliniowy pierwotne wielkości systemu. Można również widzieć strukturę (6.17) jako liniowy model o parametrach zmieniających się w czasie

$$\hat{y}_{t|a} = a_1(t)y_{t-1} + a_2(t)y_{t-1} + a_3G_{t-1} + a_4(t)e_{t-1}. \quad (6.18)$$

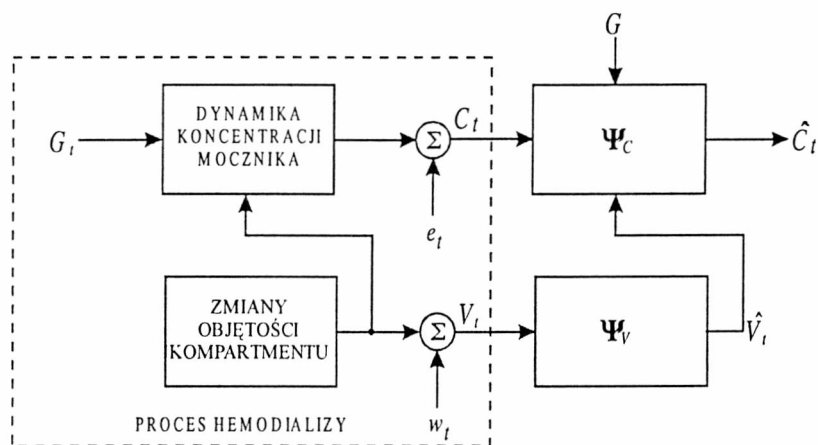
Z drugiej strony, uwzględniając zależność zmiany objętości (6.6) możemy przyjąć, że z dostatecznie dobrym przybliżeniem opisuje ją proces pierwszego rzędu

$$V_t = b_1V_{t-1} + w_t, \quad (6.19)$$

gdzie b_1 jest parametrem modelu, a w_t oznacza zakłócenie w procesie zmian objętości. Objętość kompartmentu (całkowita woda organizmu) jest mierzona za pomocą wagi łóżkowej. Przyjmując równanie opisu tego procesu w postaci (6.19) naturalnym oszacowaniem objętości z wyprzedzeniem jednostkowym jest wielkość

$$\hat{V}_{t|b} = b_1V_{t-1}. \quad (6.20)$$

Analizując wzór (6.17) zauważamy, że zmienna V_t jest otrzymywana jako rezultat identyfikacji innego procesu, mianowicie (6.19). Do równania (6.17) w miejsce zmiennej V_t trafia jej ocena na podstawie modelu (6.20). Powyższa dyskusja pozwala



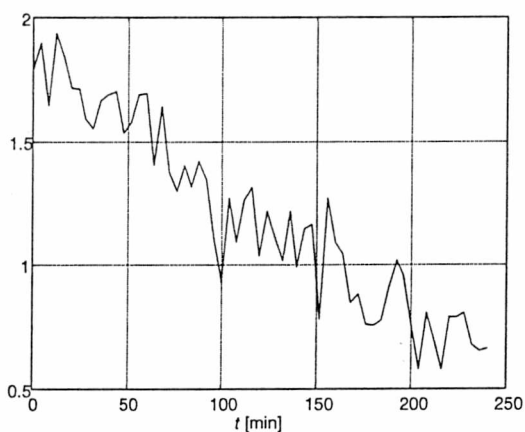
Rys. 6.4. Struktura dwustopniowego systemu identyfikacji modelu mocznikowego

na skonstruowanie systemu identyfikacji dwustopniowej przedstawionej na rys 6.4. Linia przerywaną otoczono hipotetyczną strukturę procesu złożonego z dwóch części.

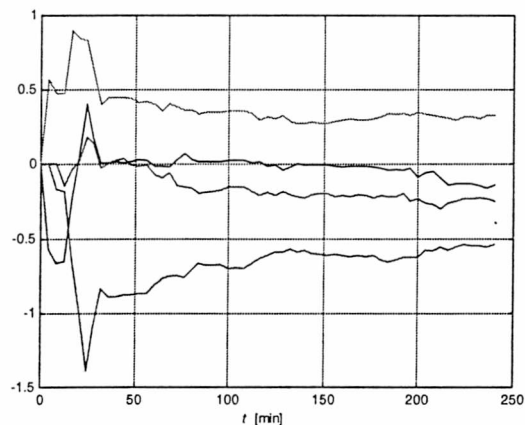
Pierwsza obejmuje dynamikę zmian poziomu stężenia mocznika, druga natomiast dynamikę zmian objętości kompartmentu. Odpowiednie algorytmy identyfikacji oznaczono przez Ψ_C i Ψ_V . Przedstawiony na rys. 6.4 system identyfikacji nie odpowiada dokładnie przedstawionym w rozdziale 4. algorytmom. W procesie na pierwszym stopniu wykorzystujemy raczej prognozowaną wartość zmiennej V_t , a nie parametr. Ta modyfikacja została przyjęta na podstawie wstępnej wiedzy o systemie zawartej w równaniach opisu.

6.2.3. Wyniki identyfikacji

Przeprowadzono eksperyment, w którym oszacowano następujące początkowe wartości parametrów: $C_0 = 1,9$ mg/ml, $K_r = 3$ ml/min, $K_d = 130$ ml/min, $V = 30$ l. Wartość pomiarowa generacji mocznika wynosiła $G = 6$ mg/min. Czas trwania dializy ustalono na 4 h, a wartości potrzebne do estymacji parametrów modelu mierzono co 4 min. Ta wartość okresu próbkowania została użyta do odpowiedniego wyskalowania wykresów. Do estymacji parametrów zastosowano rekurencyjny algorytm z wykładniczym zapominaniem [24], przy czym zastosowano współczynnik zapominania równy $\lambda = 0.98$. Dane pomiarowe przedstawiono na rys. 6.5,



Rys. 6.5. Przebieg stężenia mocznika



Rys. 6.6. Przebieg adaptacji parametrów

a adaptacyjną estymację parametrów na rys. 6.6. Błąd średniokwadratowy osiągnął wartość 0.14. Czynnikiem utrudniającym identyfikację parametrów były silnie zakłócone wartości pomiarowe (rys. 6.5). Uzyskane wyniki można uznać za zadowalające, choć zaprezentowany tu model ma raczej charakter ilustracyjny

i oddaje charakter podstawowych własności dynamicznych procesu. Pokazano tu również pewien sposób budowania modelu dogodnego do celów identyfikacji na podstawie sprawdzonego klinicznie modelu wyprowadzonego na podstawie własności fizycznych. Jego niedokładność i wrażliwość na zakłócenia jest charakterystyczna dla modeli analitycznych. Wydaje się, że poprawę własności modelu można by uzyskać stosując opis dogodny dla reprezentowania informacji nieprecyzyjnej (np. stosując logikę rozmytą). Wówczas model taki uwzględniał by więcej czynników i mógłby być stosowany także w okresie między dializami.

6.3. Identyfikacja procesu odsiarczania spalin

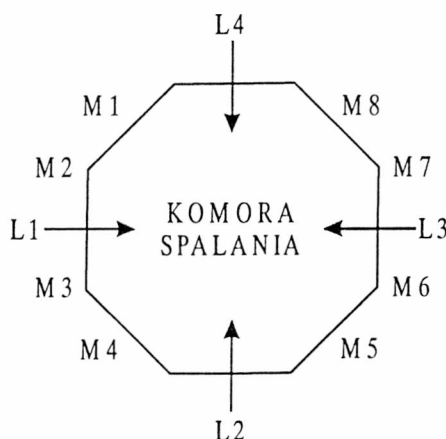
Wiele procesów przemysłowych podlegających sterowaniu ma naturę dwustopniową (lub wielostopniową). Taki złożony charakter posiadają w szczególności niektóre procesy chemiczne [71]. Znajduje to odbicie w odpowiedniej strukturze modelu, która zazwyczaj w sposób pełniejszy i dokładniejszy opisuje analizowany układ reakcji, których przebieg jest wzajemnie uzależniony. Spodziewamy się, że uwzględniając w strukturze modelu wpływ przebiegu jednej reakcji na drugą w sposób jawny, powinniśmy otrzymać lepszy opis rzeczywistości, niż gdybyśmy wiedzę o złożonej naturze procesu zaniedbali. W charakterze przykładu ilustrującego zastosowanie bieżącej identyfikacji dwustopniowej wybrano modelowanie procesu odsiarczania węgla, którego celem jest zmniejszenie zawartości szkodliwych związków siarki w gazach spalinowych, co ma duże znaczenie dla ochrony środowiska naturalnego.

6.3.1. Charakterystyka problemu

Wśród wielu elektrowni w Polsce znaczna ich liczba wykorzystuje węgiel brunatny jako podstawowe paliwo bloków energetycznych. Surowiec ten często zawiera znaczne ilości zanieczyszczeń, z których szczególnie toksyczne i szkodliwe ekologicznie są związki siarki emitowane w procesie spalania w gazach spalinowych w postaci dwutlenku siarki SO_2 . Stanowi to powód stosowania różnych technologii mających na celu zredukowanie emisji tego związku do atmosfery. Jedną z nich jest technologia wtrysku kamienia wapiennego¹ (WKW). Proces WKW polega na

¹ ang. Furnace Limestone Injection - FLI

dozowaniu mielonego kamienia wapiennego do komory paleniskowej każdego bloku.



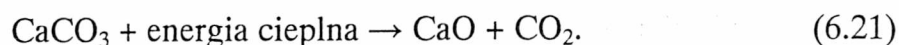
Rys. 6.7. Komora spalania w przekroju

Wtryskiwany miął wapienny reaguje z lotnym składnikiem SO_2 , a produkt odsiarczania oraz niewykorzystany miął zbiera się razem z popiołem w zainstalowanych odpylaczach elektrycznych. Wtrysk miału wapiennego do komory spalania następuje za pomocą dmuchaw transportowych (4 na każdy kocioł). Miął wapienny jest dawkowany przez zasilacz dozujący o regulowanej liczbie obrotów, dzięki czemu jego ilość może być korygowana w zależności od obciążenia kotła lub zawartości siarki w węglu brunatnym. Każda linia transportowa L_1 – L_4 rozgałęzia się na dwa punkty wtryskowe wprowadzone do komory spalania bloku. Opisaną powyżej sytuację ilustruje rys. 6.7, który przedstawia przekrój komory paleniskowej wraz z usytuowaniem punktów wtrysku miału wapiennego L_1 – L_4 i miejsc wlotu pyłu węglowego wytwarzanego w młynach M_1 – M_8 .

W komorze paleniskowej zachodzą dwie reakcje chemiczne:

- reakcja kalcynacji,
- reakcja wiązania SO_2 .

Proces kalcynacji, w wyniku którego powstaje aktywny CaO opisany jest następującą reakcją chemiczną



Do podstawowych czynników przebiegu tej reakcji zaliczamy:

- wartość temperatury, która musi być wystarczająco wysoka, by zaszła reakcja, lecz nie może być zbyt wysoka, by nie następowało „spalanie” cząsteczek absorbentu,
- CaO powinien być produkowany w odpowiedniej ilości,
- CaO powinien być wytwarzany w odpowiednich miejscach komory paleniskowej z uwagi na temperaturę i czas pozostawania w komorze,
- Cząsteczki CaO powinny być odpowiednio małe i mieć porowatą strukturę, co daje dużą powierzchnię aktywną i sprzyja wiązaniu SO₂.

Reakcja wiązania SO₂ ma następującą postać



W wyniku tej egzotermicznej reakcji powstaje siarczan wapniowy CaSO₄, który jest stałym produktem spalania usuwanym z popiołem. W powyższej reakcji istotną rolę odgrywają następujące czynniki:

- czas reakcji, który z kolei jest uzależniony od czasu zużytego na kalcynację i ogólnego czasu pozostawania w komorze reagujących materiałów,
- optymalny dobór temperatury, zależny od rodzaju absorbentu, wielkości cząstek i innych warunków, w których ma przebiegać reakcja jak. np. zawartość SO₂ i O₂,
- rodzaj absorbentu: małe cząstki lub większe, ale o porowatej strukturze, dając większą powierzchnię aktywną wspomagając reakcję wiązania SO₂.

Pomiary temperatury w komorze spalania pozwoliły na stwierdzenie, że właściwym punktem wtrysku kamienia wapiennego jest poziom doprowadzania tzw. powietrza dodatkowego górnego². Powietrze, wdmuchiwane z dysz ma dodatkowo wpływ na przemieszanie absorbentu z SO₂, co przyspiesza reakcję.

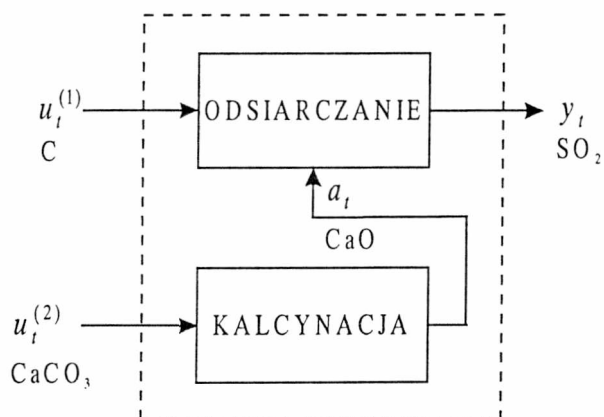
Podstawowym problemem w sterowaniu poziomem SO₂ jest duża zmienność zasiarczenia węgla dostarczanego jako surowiec opałowy bloków energetycznych. Dużą trudność stwarza także odpowiednio szybkie określenie poziomu tych zanieczyszczeń drogą badań laboratoryjnych. Od stopnia zasiarczenia węgla zależy oczywiście ilość SO₂ emitowanego w gazach spalinowych. W następnym podrozdziale

² ang. Over Fire Air (OFA).

przedstawimy uproszczony model opisanego powyżej procesu, traktując zasiarczenie jako niemierzalną wielkość wejściową.

6.3.2. Model matematyczny procesu

Charakter reakcji chemicznych (6.21,22) opisujących proces eliminacji SO_2 z gazów spalinowych pozwala na zastosowanie modelu dwustopniowego, którego strukturę przedstawiono na rys. 6.8. Parametrem wpływającym na podstawowy proces odsiarczania na pierwszym stopniu jest ilość CaO , który powstaje w wyniku procesu kalcynacji na drugim stopniu. Parametr ten oznaczono przez a_t . Wielkością wejściową pierwszego stopnia $u_t^{(1)}$ jest ilość węgla podawanego do komory spalania zawierającego zanieczyszczenia siarkowe będące źródłem SO_2 emitowanego w gazach



Rys. 6.8. Dwustopniowy proces redukcji SO_2 w gazach spalinowych

spalinowych. Ilość emitowanego SO_2 stanowi zmienną wyjściową modelu na pierwszym stopniu y_t . Drugi stopień stanowi opis procesu kalcynacji opisanego reakcją (6.21). Zmienna wejściowa $u_t^{(2)}$ oznacza łączną ilość sorbentu CaCO_3 podawaną do komory spalania. Z równania reakcji (6.21) wynika, że ilość CaO biorącego udział w odsiarczaniu jest proporcjonalna do ilości podawanego sorbentu CaCO_3 przy zachowaniu stałej temperatury. Z kolei emitowany SO_2 zależy od ilości węgla i od ilości CaO , co opisuje reakcja (6.22). Spostrzeżenia te pozwalają na przyjęcie nieco tylko zmodyfikowanego modelu opisanego równaniami (4.17) i (4.18). Przyjmujemy następujące równania modelu:

$$y_t + \alpha_1 y_{t-1} = a_t u_{t-n1}^{(1)} + \beta_1 u_{t-n1}^{(1)} + e_t, \quad (6.23)$$

$$a_t = b u_{t-n2}^{(2)} + w_t. \quad (6.24)$$

Literami greckimi oznaczono parametry modelu pierwszego stopnia, b oznacza jedyny parametr drugiego stopnia, e_t i w_t zakłócenia, a $n1$ oraz $n2$ reprezentują opóźnienia procesów odpowiednio pierwszego i drugiego stopnia. Składnik $a_t u_{t-n1}^{(1)}$ występujący we wzorze (6.23) pokazuje wzajemne oddziaływanie reakcji procesu. W skrajnym przypadku (odcięcie dopływu sorbentu) $u_t^{(2)} = 0$ i na emisję SO_2 będzie miał wpływ jedynie strumień dopływu węgla reprezentowany przez składnik $\beta_1 u_{t-n1}^{(1)}$. Na podstawie równań (6.23) i (6.24) możemy wyznaczyć rekurencyjny algorytm identyfikacji w sposób pokazany w rozdziale 4.3, gdy zauważymy, że model wypadkowy procesu ma postać

$$y_t = \theta^T \varphi_t + v_t, \quad (6.25)$$

gdzie $\theta = [\alpha_1 \ \beta_1 \ b]^T$ jest wektorem estymowanych parametrów, $\varphi_t = [-y_{t-1} \ u_{t-n1}^{(1)} \ u_{t-n1}^{(1)} u_{t-n2}^{(2)}]^T$ oznacza wektor regresji wypadkowego modelu, a $v_t = u_{t-n1}^{(1)} w_t + e_t$ jest sygnałem zakłócającym. Na podstawie rozważań z rozdziału 4.3 możemy wyprowadzić formuły algorytmu rekurencyjnego

$$\hat{\theta}_t = \hat{\theta}_{t-1} + K_t (y_t - \hat{\theta}_{t-1}^T \varphi_t), \quad (6.26a)$$

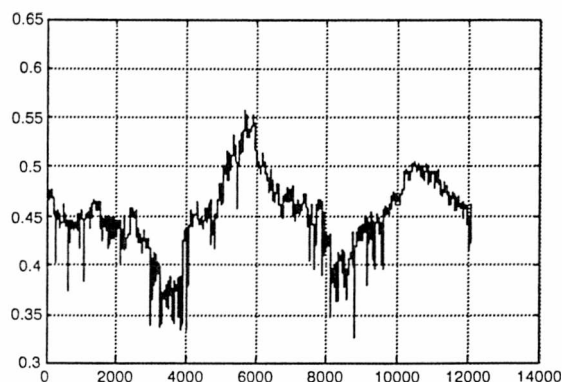
$$K_t = P_{t-1} \varphi_t / (\lambda_t^v + \varphi_t^T P_{t-1} \varphi_t), \quad (6.26b)$$

$$P_t = (I_3 - K_t \varphi_t^T) P_{t-1}. \quad (6.26c)$$

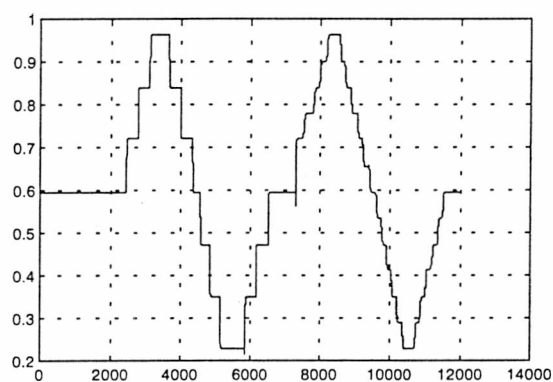
Zależności (6.26) zostały wyprowadzone w rozdziale (4.3) przy założeniu normalności rozkładów sygnałów zakłócających obu procesów. Wielkość λ_t^v ma sens wariancji sygnału zakłócającego wypadkowego modelu. W praktyce przyjmujemy, że jest ona niewielką liczbą bliską jedności. Określony wzorami (6.26) algorytm posłuży do estymacji na bieżąco parametrów modelu (6.23) i (6.24). W kolejnym rozdziale opisujemy eksperyment i dane pomiarowe.

6.3.3. Materiał pomiarowy i opis eksperymentu

Wielkości modelu z rys. 6.8 rejestrowano na rzeczywistym obiekcie co 1 s przez okres ponad 3 h 20 min zadając zmiany sumarycznego przepływu sorbentu i obserwując zmiany zawartości SO_2 w spalinach. Wyniki pomiarów z bloku

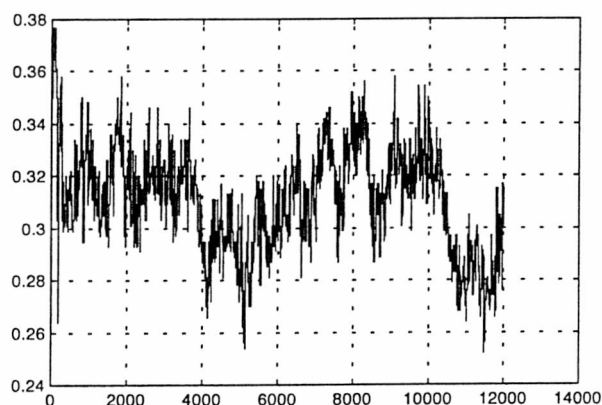


Rys. 6.9. Pomiary stężenia SO_2



Rys. 6.10. Sumaryczny strumień CaCO_3

energetycznego zbierano na dyskietki wykorzystując istniejący na tym bloku system zbierania i archiwizacji danych. Zebrany materiał pomiarowy przedstawiono na rys. 6.9-11. Przebieg zawartości SO_2 w spalinach mierzony za elektrofiltrem pokazano na rys. 6.9. Ograniczono się do dwóch sygnałów wyjściowych: sumarycznego strumienia sorbentu (rys. 6.10) i strumienia pyłu węglowego zasilającego palenisko (rys. 6.11)



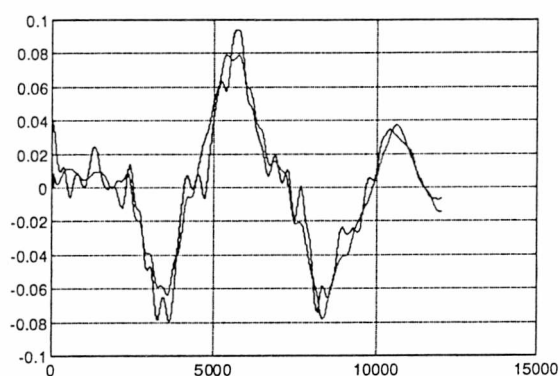
Rys. 6.11. Strumień dopływu paliwa C

przy pozostałych wielkościach ustalonych (temperatura, położenie dysz OFA,

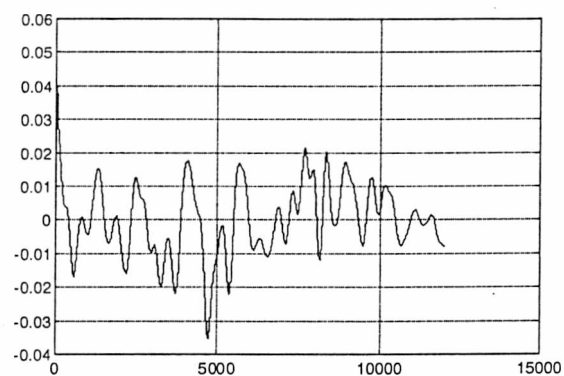
konfiguracja młynów M_1 – M_8). Przebieg strumienia sorbentu pokazany na rys. 6.10 ma charakter zbliżony do wolnozmiennnej sinusoidy i stanowi sumę pomiarów z linii L_1 – L_3 (linia L_4 była nieczynna podczas eksperymentu). Rys. 6.11. obrazuje silnie zakłócony pomiar strumienia węgla podawanego do komory spalania. Dane wejściowe poddano standardowej procedurze usuwania trendów. Ze względu na silne zakłócenia poddano je również filtracji dolnoprzepustowej. Wartości opóźnień $n_1 = n_2 = 30s$ przyjęto na podstawie pracy [59], w której podobny problem rozwiązywano za pomocą rozmytej sieci neuronowej i pakietu ANFIS³. Rozważany w tym rozdziale algorytm stanowi interesującą alternatywę dla procedury w pracy przytoczonej powyżej. Do obróbki danych i prezentacji wyników wykorzystano procedury pakietu MATLAB [45].

6.3.4. Wyniki identyfikacji

Stosując rekurencyjną procedurę 6.26 wyznaczono parametry modelu opisanego zależnościami (6.23) i (6.24). Na rys. 6.12 przedstawiono przebiegi wyjścia obiektu (po filtracji i usunięciu trendów) i modelu. Rys. 6.13. przedstawia wykres błędu



Rys. 6.12. Wyjście obiektu i modelu



Rys. 6.13. Przebieg błędu modelu

modelu. Adaptację wartości parametrów modelu obrazuje rys. 6.14. Zaproponowany model dość dobrze odzwierciedla dynamikę identyfikowanego procesu. Jego zaletą jest stosunkowo niewielka liczba parametrów. Rozbieżności między zmierzoną

³ ang. Adaptive Neuro Fuzzy Inference System

6.3. Identyfikacja procesu odsiarczania spalin

i symulowaną wartością wyjścia (rys. 6.12) pojawiają się na skutek bardzo silnie zakłóconych obserwacji, szczególnie ilości podawanego paliwa (rys. 6.11). Zakłócenia te wynikają zarówno z właściwości dostarczanego paliwa węglowego (zanieczyszczenia siarką) jak i z trudności pomiarowych.

7. Uwagi końcowe

Podstawowym celem pracy było opracowanie i badania dwóch grup komputerowych algorytmów złożonej identyfikacji rekurencyjnej. Pierwszą z nich stanowiły algorytmy ze strojoną strukturą modelu mające zastosowanie do identyfikacji obiektów w przypadku, gdy stopień wiedzy o obiekcie jest niewielki lub gdy w ogóle nie dysponujemy wiedzą wstępną. Do drugiej grupy należały algorytmy rekurencyjnej identyfikacji dwustopniowej, w których w odróżnieniu od algorytmów pierwszej grupy struktura modelu była ściśle związana z naturą identyfikowanego procesu. Dwie odmienne formy złożoności: koncepcyjna i strukturalna reprezentowane odpowiednio przez wspomniane rodzaje algorytmów wymagały zastosowania odmiennych podejść. Przy opracowywaniu pierwszej zastosowano odpowiednią modyfikację algorytmu bezpośredniego i wprowadzając dodatkowe kryterium przebudowy struktury modelu, otrzymano algorytm rekurencyjny. W przypadku algorytmów dwustopniowych zastosowano podejście bayesowskie do modelu wejściowo-wyjściowego, a dla modelu stanowego wykazano, że problem da się sprowadzić do uogólnionej filtracji Kalmana. Okazało się, że nawet w prostych skalarnych przypadkach problem identyfikacji dwustopniowej staje się nieliniowy i dość złożony obliczeniowo.

Modele uzyskane za pomocą procedur o strojonej strukturze mogą charakteryzować się mniejszą liczbą parametrów, a więc posiadają cechę ze wszech miar pożądaną z punktu widzenia identyfikacji. Z drugiej strony są to algorytmy elastyczniejsze w dopasowywaniu się do danych pomiarowych niż klasyczne znane z literatury algorytmy rekurencyjne. Dzięki temu w pewnych przypadkach algorytmy te cechuje większa odporność na nieprawidłową specyfikację struktury modelu. Dla obiektów niestacjonarnych korzystną właściwością są lepsze własności śledzące tych procedur. Ceną za te korzystne własności są znacznie większe nakłady obliczeniowe. Osiągnięte w badaniach symulacyjnych wyniki znalazły potwierdzenie w praktycznym przykładzie dotyczącym identyfikacji propanololu.

Rozważania przedstawione w rozprawie stanowiły wybrane zagadnienia dotyczące złożonych algorytmów rekurencyjnych mając ograniczony zakres i nie wyczerpując

w pełni tej problematyki. Jako kierunki dalszych badań w tej dziedzinie można wymienić:

- analizę porównawczą algorytmów ze strojoną strukturą modelu z klasycznymi nieparametrycznymi procedurami identyfikacji,
- pogłębienie badań nad kryteriami zmiany struktury polegające na zastosowaniu bardziej wyrafinowanej logiki decyzyjnej w module wnioskującym, wykorzystanie zbioru reguł zamiast jednego predykatu z jednoczesną możliwością lepszego doboru funkcji bazowych,
- badania nad identyfikacją rekurencyjną wielowymiarowych obiektów dwustopniowych,
- wykorzystanie algorytmów prezentowanych w pracy w zagadnieniach sterowania i podejmowania decyzji,
- rekurencyjne planowanie eksperymentów dla algorytmów złożonych, a w tym dobór okresów próbkowania dla złożonych procedur dwustopniowych.

Niniejszym chciałbym wyrazić serdeczne podziękowanie promotorowi pracy – Panu prof. Markowi Kurzyńskiemu za jej prowadzenie, szereg cennych wskazówek naukowych oraz życzliwe wsparcie. Dziękuję Panu prof. Andrzejowi Kasprzakowi za umożliwienie mi prowadzenia pracy naukowej w zakładzie pod jego kierunkiem. Dziękuję także Panu dr Januszowi Pollakowi i Pani mgr Teresie Lipanowicz z Instytutu Automatyki Systemów Energetycznych we Wrocławiu za udostępnienie danych pomiarowych wykorzystanych w przykładach praktycznych, które wzbogaciły moją pracę. Wyrażam również podziękowania wszystkim kolegom z Wydziałowego Zakładu Systemów i Sieci Komputerowych Politechniki Wrocławskiej za uwagi i dyskusje, które wpłynęły na ostateczną postać pracy.

Jacek Cichosz

Literatura

- [1] ANDERSON B.D.O., MOORE J.B., *Filtracja optymalna*. WNT, Warszawa, 1984.
- [2] ÅSTRÖM K. J., WITTENMARK B., *Computer Controlled Systems*. Prentice Hall Inc., Englewood Cliffs, N.J., 1984.
- [3] BASSEVILLE M., *Detecting Changes in Signals and Systems – a Survey*. Automatica, **24**, 309-326, 1988.
- [4] BIELAWSKI S., *Modele farmakokinetyczne*. WKiŁ, Warszawa, 1989.
- [5] BIERMAN G. J., *Factorization Methods for Discrete Sequential Estimation*. Academic Press, Inc., New York, 1977.
- [6] BILLINGS S.A., VOON S.F., *Least Squares Parameter Estimation Algorithms for Nonlinear Systems*. Int. J. Systems Science, **15**, no. 6, pp. 601-615, 1984.
- [7] BOHLIN T., *A Case Study of Grey Box Identification*. Automatica, **30**, no. 2, 307-318, 1994.
- [8] BOX G.E.P., JENKINS G.M., *Analiza szeregów czasowych: prognozowanie, i sterowanie*. PWN, Warszawa, 1983.
- [9] BUBNICKI Z., *Identyfikacja obiektów sterowania*. PWN, Warszawa, 1974.
- [10] BUBNICKI Z., KURZYŃSKI M., SAS J., ŚWIĄTEK J., ŻOŁNIEREK A., *Algorytmy i programy komputerowo wspomaganą diagnostyki medycznej z zastosowaniem metod modelowania i identyfikacji dla wybranych przykładów. Etap I. Rap. Inst. Ster. i Techn. Syst. P. Wr., ser. SPR nr 4/87, Wrocław 1987.*
- [11] CICHOSZ J., *Grey-Box Computer Identification of Biomedical Systems*. Int. Conference Computer Based Experiments, Learning and Teaching, COMBELET '99, Szklarska Poręba, 1999.
- [12] CICHOSZ J., *Simulation Model of Dialysis Process*. XXIst Int. Autumn Colloquium-Workshop, Advanced Simulation of Systems, ASIS '99, Krnov, Czech Republic, 1999.
- [13] CICHOSZ J., *Identification Algorithm for Nonlinear Medical Systems with On-line Model Order Selection*. Proc. of Int. Workshop on Biomedical Engineering and Medical Informatics, Gliwice, Poland, 1997.
- [14] CICHOSZ J., *Recursive Identification Algorithm for Nonlinear Systems with On-line Model Order Selection*. Proc. of 9th Int. Symposium on System Modeling Control, Zakopane, Poland, 1998.
- [15] CICHOSZ J., KURZYŃSKI M., *Rekurencyjny algorytm identyfikacji obiektów biomedycznych z adaptacyjną zmianą rzędu modelu*. III Krajowa Konferencja Naukowa: Techniki Informatyczne w Medycynie, Jaszowiec 1998.
- [16] COVER T.M., *Estimation by the Nearest Neighbor Rule*. IEEE Trans. Information Theory, **IT-14**, no. 1, pp. 50-55, 1968.
- [17] EYKHOFF P. ed., *Identification and System Parameter Estimation*, Part 1, Proc. of the 3rd IFAC Symp., The Hague/Delft, North-Holland Publ. Comp., Amsterdam-London, 1973.
- [18] EYKHOFF P., *System Identification, Parameter and State Estimation*. Wiley, New York, London, Sydney, 1974.

- [19] EYKHOFF P., PARKS P.C., *Identification and System Parameter Estimation, Where Do We Stand Now ?* Automatica, **26**, no. 1, pp. 3-5, 1990.
- [20] FISZ M., *Rachunek prawdopodobieństwa i statystyka matematyczna*. PWN, Warszawa, 1969.
- [21] FLOOD R.L., CARSON E.R., *The Future Role of Mathematical Models in Medicine: A Case Study*. Proc. of IFAC Symposium: Modeling and Control in Biomedical Systems, Venice, Italy, Pergamon Press, str. 99-102, 1988.
- [22] GENTIL S., BARRAUD A.Y., SZAFNICKI K., *SEXI: an Expert Identification Package*. Automatica, **26**, 1990.
- [23] GOLUB G. H., VAN LOAN C. F., *Matrix Computations*. 2nd ed., John Hopkins University Press, 1989.
- [24] GOODWIN G. C., PAYNE R. L., *Dynamic System Identification: Experiment Design and Data Analysis*. Academic Press, Inc., New York, 1977.
- [25] GRAUPE D., JAIN V. K., SALAH J., *A Comparative Analysis of Various Least-Squares Identification Algorithms*. Automatica, **16**, 663-681, 1980.
- [26] GREBLICKI W. *Asymptotycznie optymalne algorytmy rozpoznawania i identyfikacji w warunkach probabilistycznych*. Prace Naukowe Instytutu Cybernetyki Technicznej Politechniki Wrocławskiej, Wrocław, 1974.
- [27] HORN R. A., JOHNSON C. R., *Matrix Analysis*. Cambridge University Press, Cambridge, 1986.
- [28] HUNT J.K. *A Survey of Recursive Identification Algorithms*. Trans. Inst. MC, **8**, no. 5, 273-278, 1996.
- [29] JAGIELSKI J., ZMYSŁOWSKI W. (red.), *Problemy biocybernetyki i inżynierii biomedycznej*. t 1., Biosystemy, WKiŁ, Warszawa 1991.
- [30] JUDITSKY A., et al., *Nonlinear Black-Box Models in System Identification: Mathematical Foundations*. Automatica, **31**, no. 12, 1725-1750, 1995.
- [31] KACZOREK T., *Macierze w automatyce i elektrotechnice*. WNT, Warszawa, 1984.
- [32] KORBICZ J., OBUCHOWICZ A., UCIŃSKI D., *Sztuczne sieci neuronowe. Podstawy i zastosowania*. Akademicka Oficyna Wydawnicza PLJ, Warszawa 1994, s. 159-205.
- [33] KUO B. C., *Automatic Control Systems*. Prentice Hall, Englewood Cliffs, New Jersey, 1991.
- [34] KURZYŃSKI M., *Złożone metody rozpoznawania i identyfikacji w zastosowaniach medycznych*. Biocybernetyka i inżynieria biomedyczna - stan badań w Polsce, tom 2, 689-693, Warszawa, 1994.
- [35] KURZYŃSKI M., *Wielostopniowa identyfikacja obiektów statycznych w warunkach probabilistycznych*. Prace nauk. Inst. Cybern. Tech. P.Wr. **46**, 47-58, Wrocław 1975.
- [36] KURZYŃSKI M. *Rozpoznawanie obiektów, metody statystyczne*. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław, 1997.
- [37] KURZYŃSKI M., SAS J., ŚWIĄTEK J., PUCHAŁA E., WĄSIEWICZ E., *Algorytmy i programy komputerowo wspomaganey diagnostyki medycznej z zastosowaniem metod modelowania i identyfikacji dla wybranych przykładów. Etap II*. Rap. Inst. Ster. i Techn. Syst. P. Wr., ser. SPR nr 13/87, Wrocław 1987.

- [38] KURZYŃSKI M., SAS J., ŚWIĄTEK J., PUCHAŁA E., WĄSIEWICZ E., *Algorytmy i programy komputerowo wspomaganie diagnostyki medycznej z zastosowaniem metod modelowania i identyfikacji dla wybranych przykładów. Etap III. Rap. Inst. Ster. i Techn. Syst. P. Wr., ser. SPR nr 16/88, Wrocław 1988.*
- [39] KURZYŃSKI M., SAS J., WIKIERA I., *Komputerowe wspomaganie diagnozowania ostrej niewydolności nerek u dzieci. Polski Tygodnik Lekarski* **45**, nr 404, 820-824, 1990.
- [40] KURZYŃSKI M., WĄSIEWICZ E., WIKIERA I., *Komputerowy system wspomagania procesu dializy otrzewnowej w ostrej niewydolności nerek u dzieci. Polski Tygodnik Lekarski*, **44**, nr 1819, 436-440, 1989.
- [41] LEONARITIS J.J., BILLINGS S.A., *Input-Output Parametric Models for Non-linear Systems, Part I – Deterministic Nonlinear Systems, Part II – Stochastic Nonlinear Systems. Int. J. Control*, **41**, no. 2, pp. 303-344, 1985.
- [42] LINDSKOG P., *Methods, Algorithms and Tools for System Identification Based on Prior Knowledge. Phd thesis 436, Dept. of Electrical Eng., Linköping University, Linköping, Sweden, 1996.*
- [43] LINDSKOG P., LJUNG L., *Tools for Semiphsical Modeling. Int. J. Adaptive Control Signal Proc.*, **9**, no. 6, 509-523.
- [44] LJUNG L., *System Identification - Theory for the User. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1987.*
- [45] LJUNG L., *System Identification Toolbox for Use with MATLAB, Users Guide. The Math Works Inc., 1997.*
- [46] LJUNG L., *A Discussion of Model Accuracy in System Identification. Int. J. Adaptive Control and Signal Processing*, **6**, 161-171, 1992.
- [47] LJUNG L., GLAD T., *Modeling of Dynamic Systems. Prentice Hall, Englewood Cliffs, N.J. 1994.*
- [48] LJUNG L., GUNNARSSON S., *Adaptation and Tracking in System Identification - a Survey. Automatica*, **26**, no. 1, 7-21, 1990.
- [49] LJUNG L., GUO L., *The Role of Model Validation for Assesing Unmodeled Dynamics. Tech. Report, Linköping University, Sweden, <ftp.control.isy.liu.se>, 1995*
- [50] LJUNG L., SÖDERSTRÖM T., *Theory and Practice of Recursive Identification. MIT Press, Cambridge Mass., 1987.*
- [51] LOWRIE G.E., SARGENT J.A., *Clinical Example of Pharmacokinetic and Metabolic Modeling: Quantitative and Individualized Prescription of Dialysis Therapy. Kidney International*, **18**, suppl. 10, S-11—S-16, 1980.
- [52] MAŃCZAK K., *Metody identyfikacji wielowymiarowych obiektów sterowania. WNT, Warszawa 1979.*
- [53] MAŃCZAK K., NAHORSKI Z., *Komputerowa identyfikacja obiektów dynamicznych. PWN, Warszawa, 1983.*
- [54] NARENDRA K.S., PARTHASARATHY K., *Identification and Control of Dynamical Systems Using Neural Networks. IEEE Trans. Neural Networks*, **1**, no. 1, pp. 4-27, 1990.

- [55] NIEDERLIŃSKI A., KASPRZYK J., FIGWER J., *Inteligentna identyfikacja dynamicznych modeli matematycznych ARX – drzewo decyzyjne lub system ekspertowy ?* Problemy Sztucznej Inteligencji, red. W. Traczyk, Wiedza i Życie, Warszawa 1995.
- [56] NIEDERLIŃSKI A., MOŚCIŃSKI J., OGONOWSKI Z. *Regulacja adaptacyjna*. PWN, Warszawa, 1995.
- [57] ORŁOWSKI T., WERYŃSKI A. (red.), *Problemy biocybernetyki i inżynierii biomedycznej*. t. 3., *Sztuczne narządy*. str. 101-180, WKiŁ, Warszawa 1990.
- [58] PAPOULIS A., *Prawdopodobieństwo, zmienne losowe i procesy stochastyczne*. WNT, Warszawa, 1972.
- [59] POLLAK J., LIPANOWICZ T., GODOWICZ J., DYNIA Z., *Opracowanie systemu automatycznej regulacji i optymalizacji pracy instalacji odsiarczania spalin*. Raport techn. nr 774, Instytut Automatyki Systemów Energetycznych, Wrocław 1999.
- [60] RAFAJŁOWICZ E., *Algorytmy planowania eksperymentu z implementacjami w środowisku MATHEMATICA*. Akademicka Oficyna Wydawnicza PLJ, Warszawa, 1996.
- [61] RUTKOWSKA D., PILIŃSKI M., RUTKOWSKI L., *Sieci neuronowe, algorytmy genetyczne i systemy rozmyte*. PWN, Warszawa, 1997.
- [62] SJÖBERG J. et al., *Nonlinear Black-Box Modeling in System Identification: a Unified Overview*. Automatica **31**, no. 12, 1691-1724, 1995.
- [63] SJÖBERG J., HJALMARSSON H., LJUNG L., *Neural Networks in System Identification*. In Preprints 10th IFAC Symp. on System Identification, Copenhagen, vol. 2, 49-72, anonymous ftp: 130.236.24.1.
- [64] SARGENT J.A., *Control of Dialysis by a Single-Pool Urea Model*. The National Cooperative Dialysis Study, Kidney International, **23**, 35-42, 1983.
- [65] SARGENT J.A., GOTCH F. A., *The Analysis of Concentration Dependence of Uremic Lesions in Clinical Studies*. Kidney International, **7**, suppl. 2, S-35—S-42, 1977.
- [66] SARGENT J. A., GOTCH F. A., *Mathematic Modeling of Dialysis Therapy*. Kidney International, **18**, suppl. 10, S-2—S-10, 1980.
- [67] SARGENT J. A., LOWRIE E. G., *Which Mathematical Model to Study Uremic Toxicity ?* Clinical Nephrology, **17**, no. 6, 303-314, 1982.
- [68] SÖDERSTRÖM T., *On Model Structure Testing in System Identification*. Int. J. Control, **26**, no. 1, 1-18, 1977.
- [69] SÖDERSTRÖM T., STOICA P., *Identyfikacja systemów*. PWN, Warszawa, 1997.
- [70] SUGENO M., YASUKAWA T., *A Fuzzy Logic Based Approach to Qualitative Modeling*. IEEE Trans. Fuzzy Syst., **1**, 1-31, 1993.
- [71] ŚWIĄTEK J., *Identyfikacja dwustopniowa oraz jej zastosowania techniczne i biomedyczne*. Prace naukowe Instytutu Sterowania i Techniki Systemów, Wrocław, 1987.
- [72] UNBEHAUEN H., GÖHRING B., *Tests for Determining Model Order in Parameter Estimation*. Automatica, **10**, 233-244, 1974.
- [73] WITTENMARK B., ÅSTRÖM K. J., *Practical Issues in the Implementation of Self-Tuning Control*. Automatica, **20**, no. 5, pp. 595-605, 1984.

Niniejszy raport otrzymują:

1. Recenzenci	2 egz.
2. Promotor	1 egz.
3. Biblioteka Międzyinstytutowa i OINT	1 egz.
4. Autor	6 egz.
10 egz.	