

KU 0197 pozycja wydawnictw naukowych
Akademii Górnictwo-Hutniczej im. Stanisława Staszica w Krakowie

© Wydawnictwa AGH, Kraków 2007
ISBN 978-83-7464-046-6

Redaktor Naczelny Uczelnianych Wydawnictw
Naukowo-Dydaktycznych: *Jan Sas*

Z-ca Redaktora Naczelnego: *Beata Barszczewska-Wojda*

Komitek Naukowy UWND AGH:
Kazimierz Jeleni (przewodniczący),
Edward Fraś,
Tadeusz Sawik,
Ryszard Überman,
Adam Paweł Wojda,
Mariusz Ziółko

Recenzenci: *prof. dr hab. Henryk Górecki*
prof. dr hab. inż. Jerzy Klamka



W 75477 | 7

Redakcja: *Małgorzata Koch*

Redakcja techniczna: *Beata Barszczewska-Wojda*

Projekt okładki i stron tytułowych: *Wojciech J. Steifer*

Skład komputerowy: „Andre”, tel. 012 423 10 10

Druk i oprawa: „DjaF”, tel. 012 636 32 40, www.djaf.pl

WYDANIE I

Redakcja Uczelnianych Wydawnictw Naukowo-Dydaktycznych AGH
al. Mickiewicza 30, 30-059 Kraków
tel. 012 617 32 28, tel./fax 012 636 40 38
e-mail: redakcja@wydawnictwoagh.pl
www.WydawnictwoAGH.pl

BIBLIOTEKA GŁÓWNA AGH



1000214331

Spis treści

Przedmowa	1
Preface	1
Wstęp	1
1. Problemy automatyki	2
2. Modelowanie procesów dla celów sterowania	3
2.1. Wstęp	3
2.2. Modele stanu ustalonego obiektów liniowych i nieliniowych	4
2.3. Modele procesów dynamicznych	4
2.3.1. Modele sygnałów	4
2.3.2. Modelowanie układów ciągłych	4
2.3.3. Proste modele dynamiczne liniowe i analogie równań	5
2.3.4. Proste modele dynamiczne nieliniowe	5
2.3.5. Standardowe nieliniowości w równaniach różniczkowych	6
3. Analiza układów liniowych ciągłych w dziedzinie czasu	65
3.1. Równania różniczkowe	65
3.2. Równania różniczkowe stanu	66
3.3. Liniowe niestacjonarne równanie stanu	69
3.4. Liniowe stacjonarne równanie stanu	71
3.5. Transformacje liniowe zmiennych stanu	74
3.5.1. Transformacje macierzy stanu A układu MIMO do postaci Jordana	74
3.5.2. Równanie stanu dla układu SISO n -tego rzędu	75
3.5.3. Inna postać równania stanu	77
3.5.4. Transformacje dowolnej macierzy stanu A układu SISO do postaci Frobeniusa F	79
3.6. Splot i jego zastosowanie	82
4. Analiza układów liniowych ciągłych w dziedzinie częstotliwości	87
4.1. Transformata Laplace'a i jej cechy	87

4.1.1. Uwagi ogólne	87
4.1.2. Własności i przykłady transformat Laplace'a	88
4.2. Odwrotna transformata Laplace'a i jej zastosowanie	91
4.2.1. Odwrotna transformata Laplace'a	91
4.2.2. Rozwiązywanie równań różniczkowych techniką operatorową	92
4.3. Pojęcie transmitancji operatorowej układu liniowego	95
4.3.1. Własności transmitancji. Realizowalność fizyczna	96
4.3.2. Algebra schematów blokowych	99
4.3.3. Naturalne zmienne stanu dla obiektów typu SISO opisanych transmitancją	101
4.3.4. Transmitancja układów wielowymiarowych	104
4.3.5. Przykłady transmitancji układów skalarnych	106
4.3.6. Elektroniczna realizacja transmitancji	112
4.4. Związki transmitancji z odpowiedzią układu na wymuszenia harmoniczne	115
4.4.1. Odpowiedź układu I rzędu na sygnał sinusoidalny	115
4.4.2. Przechodzenie sygnałów wykładniczych przez układy liniowe	116
4.4.3. Charakterystyki częstotliwościowe układów dynamicznych	118
4.4.4. Układy nieminimalnofazowe	122
4.5. Przekształcenie Fouriera	123
4.5.1. Szereg Fouriera dla funkcji okresowej	123
4.5.2. Wzór całkowy Fouriera dla funkcji nieokresowej	124
4.5.3. Niektóre własności transformaty Fouriera	125
4.5.4. Związki pomiędzy charakterystykami częstotliwościowymi i impulsowymi	127
4.5.5. Częstotliwościowe kryterium realizowalności fizycznej	128
4.6. Analiza własności obiektu II rzędu	130
4.6.1. Odpowiedzi skokowe	130
4.6.2. Charakterystyka częstotliwościowa obiektu II rzędu	134
4.6.3. Analiza dynamiki układu II rzędu na płaszczyźnie fazowej	137
4.7. Układ III rzędu	139
4.8. Przykład analizy obiektu IV rzędu	144
5. Analiza układów liniowych dyskretnych	147
5.1. Podstawy równań różnicowych	148
5.2. Charakterystyczne postacie równań różnicowych	153
5.3. Analiza układów dyskretnych w dziedzinie czasu	155
5.3.1. Równanie stanu	155
5.3.2. Wykorzystanie splotu dyskretnego	159
5.4. Analiza układów dyskretnych w dziedzinie zmiennej zespolonej	162
5.4.1. Transformata Z	162
5.4.2. Aproxymacje transformat dyskretnych	169
5.4.3. Obiekty o nieskończonej i skończonej odpowiedzi impulsowej	171
5.4.4. Wybór okresu dyskretyzacji	174

6. Struktury systemów sterowania	177
6.1. Systemy automatyki procesowej i zabezpieczeniowej	177
6.2. Ogólny schemat układu regulacji ze sprzężeniem zwrotnym	180
6.3. Przykładowe realizacje jednowymiarowych układów regulacji SISO	181
6.3.1. Układ klasyczny	181
6.3.2. Układ kaskadowy	182
6.3.3. Układ adaptacyjny	183
6.4. Inne dwupiętlowe struktury regulacji	186
6.4.1. Struktura ze sprzężeniem zwrotnym i sprzężeniem do przodu	186
6.4.2. Struktura z modelem wewnętrznym (IMC)	188
6.4.3. Zmodyfikowana struktura z modelem wewnętrznym (MFC)	190
6.5. Elementy, układy i systemy automatyki	191
6.6. Algorytmy ciągłej regulacji PID	193
6.7. Analogowe struktury regulatorów ciągłych	196
6.8. Algorytmy dyskretnie regulatorów PID	198
6.9. Programowalne sterowniki logiczne PLC i nowe ich wersje PAC	201
6.10. Regulatory przekaźnikowe stosowane w układzie regulacji	206
7. Stabilność układów dynamicznych	209
7.1. Stabilność układów liniowych ciągłych	209
7.1.1. Stabilność względem warunków początkowych	209
7.1.2. Stabilność w sensie „ograniczone wejście-ograniczone wyjście” (BIBO)	211
7.1.3. Kryteria algebraiczne stabilności	214
7.1.4. Metody częstotliwościowe badania stabilności	223
7.1.5. Stan ustalony odpowiedzi niestabilnego układu liniowego	225
7.2. Stabilność układów liniowych dyskretnych	229
7.3. Stabilność układów nieliniowych ciągłych	234
7.3.1. Definicje stabilności rozwiązania	234
7.3.2. Punkty równowagi	236
7.3.3. Pierwsza metoda Lapunowa	237
7.3.4. Metoda Popowa	241
7.3.5. Metoda funkcji opisującej	245
8. Sterowalność i obserwowałość	253
8.1. Sterowalność	253
8.2. Stabilizowalność	262
8.3. Obserwowałość	266
8.4. Niezmienność sterowalności i obserwowałości	271
8.5. Zachowanie sterowalności i obserwowałości po dyskretyzacji układu ciągłego	272
8.6. Asymptotyczne różniczkowe estymatory stanu	273
9. Kryteria i dobór optymalnych nastaw regulatorów PID	277
9.1. Podstawowe kryteria jakości	277

9.2. Całkowe kryteria jakości	280
9.3. Formuły na liczenie całki z kwadratu błędu	283
9.4. Uogólnione kwadratowe kryterium całkowe	288
9.5. Całkowe kryterium z modułu błędu	291
9.6. Liniowe kryterium jakości	294
9.7. Inżynierskie metody strojenia regulatorów	295
9.7.1. Pierwsza metoda Zieglera–Nicholsa dla układów ciągłych	296
9.7.2. Druga metoda Zieglera–Nicholsa dla układów ciągłych	299
9.7.3. Metody Zieglera–Nicholsa dla układów dyskretnych	300
9.8. Strojenie regulatorów w zadaniach z niepewnością	301
9.8.1. Zakłócenia najbardziej niebezpieczne	301
9.8.2. Oszacowania maksymalnych odchyłek wyjścia	302
10. Przykład optymalizacji parametrycznej dla regulatora PI	309
10.1. Transmitancja wyparki	310
10.2. Identyfikacja wyparki	310
10.3. Stabilność układu zamkniętego z regulatorem PI dla toru zakłócenia	312
10.4. Sformułowanie zadania optymalizacji parametrycznej	314
10.5. Analityczne formuły dla kwadratowych wskaźników jakości	314
10.5.1. Optymalny regulator PI według całkowego wskaźnika kwadratu błędu	314
10.5.2. Optymalny regulator P według całkowego wskaźnika kwadratu błędu	318
10.5.3. Optymalny regulator PI według wskaźnika z kwadratu pochodnej błędu	319
10.6. Numeryczne wyniki symulacji układu optymalnego dla kryterium kwadratowego	324
10.7. Optymalizacja parametryczna dla innych wskaźników jakości	330
11. Metody projektowania regulatorów stanu	335
11.1. Metoda lokowania zer i biegunów	335
11.2. Minimalnonormowe sterowania programowe	340
11.3. Sterowanie modalne	342
11.4. Optymalny regulator liniowo-kwadratowy LQR	343
11.5. Nieliniowy czasooptymalny regulator od stanu	351
12. Klasyczne metody obserwacji stanu	359
12.1. Obserwatory asymptotyczne	359
12.1.1. Rola obserwatorów	359
12.1.2. Obserwatory asymptotyczne w układzie ze sprzężeniem zwrotnym	361
12.2. Obserwator Luenbergera	362
12.3. Optymalny filtr Kalmana	367

12.3.1. Ciągły filtr Kalmana	367
12.3.2. Dyskretny filtr Kalmana	370
12.3.3. Rozszerzony dyskretny filtr Kalmana	372
13. Całkowe metody dokładnej obserwacji stanu	375
13.1. Obserwatory dokładne w przestrzeni $L^2[0, T]$	375
13.1.1. Przykład obserwatora całkowego odtwarzającego dokładnie stan końcowy	376
13.1.2. Ogólne warunki istnienia obserwatora całkowego	378
13.1.3. Optymalna postać obserwatora całkowego z minimalną normą	379
13.1.4. Optymalna postać obserwatora dla wybranych wag	383
13.2. Ogólna teoria obserwacji w przestrzeniach Hilberta	392
13.2.1. Warunki istnienia obserwatora	392
13.2.2. Optymalny obserwator skończonej liczby parametrów układu liniowego	395
13.2.3. Minimalnonormowa postać obserwatora	396
13.3. Przykład obserwacji całkowej w układzie o parametrach rozłożonych	398
13.3.1. Problemy obserwacji w układach o parametrach rozłożonych	398
13.3.2. Model procesu cieploprzewodnictwa	399
13.3.3. Obserwacja punktowa i uśredniona wyjścia	402
13.3.4. Obserwator dokładny. Przykład numeryczny	403
13.4. Ogólniejsza wersja optymalnego obserwatora całkowego w $L^2[0, T]$ dla zakłóceń należących do elipsoid	406
14. Obserwatory całkowe w układzie on-line	409
14.1. Obserwatory całkowe z przesuwanym i rozszerzanym oknem obserwacji	409
14.2. Analiza stabilności obserwatorów całkowych w układzie ze sprzężeniem zwrotnym	411
15. Dyskretne wersje obserwatorów dokładnych	417
15.1. Prosta wersja obserwatora przy dyskretnych pomiarach wyjścia	417
15.2. Uproszczona wersja minimalnonormowa obserwatora przy dyskretnych pomiarach wyjścia	419
15.3. Pełna wersja minimalnonormowa obserwatora przy dyskretnych pomiarach wyjścia i wejścia	420
16. Różniczkowe wersje obserwatorów dokładnych	425
16.1. Różniczkowa wersja obserwatora z rozszerzanym oknem obserwacji EWO	425
16.2. Różniczkowa wersja obserwatora EWO w ogólniejszym problemie z zakłóceniami niejednostkowymi	428

16.3. Różniczkowe wersje obserwatora z przesuwanym oknem obserwacji MWO	430
16.3.1. Różniczkowa wersja obserwatora z opóźnieniem	430
16.3.2. Różniczkowa wersja obserwatora różnicowego bez opóźnienia	431
17. Adaptacyjne algorytmy obserwatorów całkowych	435
17.1. Sekwencyjna współpraca obserwatora całkowego MWO i filtru Kalmana	435
17.2. Równoległa współpraca obserwatorów całkowych o różnych długościach okna	441
18. Modelowanie, obserwacja i sterowanie kolumną destylacyjną	445
18.1. Kolumna jako obiekt sterowania	445
18.2. Stacyczny model kolumny destylacyjnej dwuskładnikowej	448
18.3. Dynamiczny nieliniowy model kolumny destylacyjnej dwuskładnikowej	453
18.4. Zlinearyzowany model kolumny destylacyjnej	456
18.5. Eksperymenty symulacyjne z wykorzystaniem rozszerzonego filtru Kalmana	459
18.6. Eksperymenty symulacyjne z wykorzystaniem obserwatora całkowego	460
19. Aneks 1. Elementy algebry liniowej	465
19.1. Macierz i jej zagadnienie własne	465
19.2. Przekształcenie podobieństwa i jego związek z bazą przestrzeni	466
19.3. Macierze wielomianowe	468
19.4. Macierz charakterystyczna i jej własności	469
19.5. Sprowadzanie macierzy A do postaci Jordana	470
19.6. Własności wektorów własnych prawo- i lewostronnych	474
19.7. Ogólne własności macierzy	476
19.8. Rozwiązywanie algebraicznego układu liniowego	478
19.9. Własności formy kwadratowej	481
20. Aneks 2. Podstawowe pojęcia analizy funkcjonalnej	485
21. Zakończenie	493
Literatura	495
Skorowidz	511

Przedmowa

Treścią książki są metody analizy i syntezy układów obserwacji i sterowania dynamicznymi procesami występującymi w ciągłych technologiach. Takie układy składają się na komputerowe systemy automatyki procesowej. Główny nacisk położono na prezentację nowych metod obserwacji związkanych z odtwarzaniem niemierzalnych zmiennych stanu, czyli z techniką syntezy i konstrukcją obserwatorów stanu stosowanych w systemach stabilizacji układu. Od kilku lat przedmiotem badań naukowych autora są niestandardowe metody rekonstrukcji stanu opierające się na obserwatorach całkowych, które, w odróżnieniu od asymptotycznych obserwatorów różniczkowych typu filtru Kalmana, odtwarzają stan dokładnie w zadany i skończony czasie. W zastosowaniach on-line całkowe obserwatory stanu realizują zaawansowane przetwarzanie sygnałów wejścia i wyjścia w przesuwanych oknach pomiarowych o zadanej długości T – MWO (*Moving Windows Observer*). Obecne moce obliczeniowe systemów mikroprocesorowych są wystarczające dla stosowania prezentowanych algorytmów on-line, gwarantując skończony, zadany z góry czas wykonania zadania odtworzenia stanu. Ma to ważne znaczenie w komputerowych systemach czasu rzeczywistego. Klasyczne algorytmy różniczkowe dostarczają tylko wartość estymaty stanu z gwarancją jej asymptotycznej zbieżności do stanu rzeczywistego, ale w niesprecyzowanym bliżej czasie i bez możliwości podania aktualnego błędu estymacji. Do lat 90. tylko takie rekurencyjne algorytmy mogły być realizowane, ze względu na ograniczone moce komputerów.

W zakresie klasycznych i znanych metod modelowania i sterowania położony został nacisk na skrótną, ale wszechstronną ich prezentację, oraz możliwe dokładny ich opis matematyczny. Dla ilustracji problematyki w tekście zamieszczono przykłady obliczeniowe i numeryczne, a w aneksach zebrano podstawowe wiadomości z algebry liniowej, rachunku macierzowego i analizy funkcjonalnej. W książce autor starał się przedstawić całościową problematykę obserwacji skończenie wymiarowego wektora stanu w układach liniowych oraz zebrał swój dorobek naukowy w dziedzinie nowych metod obserwacji, może więc ona stanowić źródło do tej problematyki dla pracowników naukowych zainteresowanych tymi metodami. Książka napisana jest tak, by mogła być też wykorzystana przez studentów studiów doktoranckich kierunku automatyka i robotyka.

Preface

This book is about analysis and synthesis of observation and control methods designed for dynamic systems, which one can find in continuous processes technologies. These methods are implemented in control process and automation systems. The main emphasis was placed on presentation of new observation methods connected with reconstruction of inaccessible to measurement state variables, as well as on synthesis and designing of state observers used in computer stabilization systems. Since many years the author's research subject are nonstandard methods of state reconstruction based on integral observers, which distinct from asymptotic differential observers like Kalman Filter. The main difference is that the integral observers reconstruct the exact value of the state in finite time interval. In on-line mode these observers perform advanced processing of input/output signals on moving measurement windows with given length T – MWO (Moving Windows Observer).

Computational power of nowadays microprocessor systems is sufficient for on-line execution of presented algorithms and guarantee finishing of state reconstruction task in presumed time. It's essential property of each task in real-time multitasking computer systems. Classical differential algorithms supply only the value of state estimate and guarantee of its asymptotic convergence to real state, but in unknown time and with unknown error. Up to ninetieths only such recursive algorithms could be realized in on-line mode, because of computer possibilities.

In the field of known and classic methods of modeling and control, shortened but wide-range presentation was given with their precise mathematical description. For illustration of problems many numerical examples were placed in the text. In the end of the book two Appendixes were added in which basic formulas and theorems of linear algebra, matrix calculus and functional analysis were collected. The author tried to present in the book the broad scope of finite dimension state vector observation problems for linear systems as well as new results of his researches. Therefore this book can be treated as the source of this issues useful for scientific researchers interested in these topics. The book may also be suitable for PhD students in the field of Automatics and Robotics.

Wstęp

Navigare necesse est...
Pompejusz Wielki, według Plutarcha

Grecki wyraz *autos* oznacza – własny (przymiotnik) lub sam (zaimek). Stąd autobiografia – to własny życiorys, autograf własny podpis, autonomia to samodzielność, a autor to samodzielny twórca. Automat to urządzenie, które działa samoczynnie i wykonuje samo-powtarzaną akcję (np. samoprzeładowująca się broń automatyczna).

Układ regulacji automatycznej to układ, który będzie samodzielnie sterował jakimś procesem. Obecność regulatorów w otaczających nas układach technicznych jest tak powszechna, że przestajemy ją postrzegać jako cechę nadzwyczajną i zaczynamy ją traktować jako nieodzowny składnik zapewniający wysoką jakość urządzenia. Konieczność wyposażenia dowolnego procesu fizycznego, którym chcemy sterować, w sprzęt automatyki wynika z potrzeby wykonania tego zadania lepiej, niż uzyskano by w wyniku sterowania ręcznego, a w końcu z chęci uwolnienia człowieka od bezpośredniego nadzoru nad tym procesem. Jakkolwiek pojęcie sterowania może być rozszerzone na każde zamierzone oddziaływanie na otaczające nas zjawiska i systemy – od zarządzania firmą poprzez sterowanie klimatyzacją, do lądowania na Księżycu, to głównym polem zastosowań dla inżyniera automatyka będą produkcyjne procesy technologiczne lub techniczne urządzenia użytkowe. Każde z nich musi być dziś wyposażone w sprzęt automatyki. Automatyczna skrzynia biegów, pralka automatyczna, pilot automatyczny – te nazwy są już właściwie nazwami własnymi. Nie przyjęły się nazwy „pilot elektroniczny” czy „pilot komputerowy”, choć elektronika i informatyka stanowią „ciało” tego urządzenia samolotowego. Automatyka jest jednak jego istotą. Elektronika tworzy uniwersalną platformę sprzętową wykorzystywaną przez automatykę, na przykład w postaci kart pomiarowych czy komputera przetwarzającego dane, który tak jak obrabiarka jest cyfrową wersją precyzyjnego narzędzia. Informatycznym medium aktywującym to narzędzie i niosącym polecenia jego pracy jest program napisany w dowolnie wybranym języku i zawierający w sobie zakodowane reguły poszczególnych czynności. Sterowanie, jako zespół tych czynności, powinno charakteryzować się celowością, sensownością, sprawnością, niezawodnością, precyją, a niekiedy samodzielną inteli-

gencją i powinno prowadzić do wykonania zaplanowanego zadania, czyli wykonania jakiegoś dzieła, mając dodatkowo na uwadze koszt jego tworzenia. Musi więc wcześniej powstać plan tego postępowania, czyli zostać stworzona strategia sterowania w odniesieniu do pojedynczego narzędzia, złożonego układu, czy w końcu całego systemu złożonego z układów.

Nauka, zajmująca się planowaniem i inżynierską implementacją takiej strategii sterowania, która przeprowadza obiekt sterowany z jednego stanu do innego, nazywana jest automatyką.

Gdy – pomimo nadciągającej burzy – wódz rzymski Pompejusz (w 50 r. p.n.e) musiał wypływać z portu, wypowiedział do marynarzy słowa, które znane są do dzisiaj: *Navigare necesse est, vivere non est necesse* (łac.) (Żeglowanie jest koniecznością, życie koniecznym nie jest). Żeglarze i podróżnicy przyjęli je jako swoje motto, ale niektórzy doszukują się w nich głębszych treści, parafrując, że wytyczanie celów i ich osiąganie jest istotą życia ludzkiego. Wymaga to aktywnego oddziaływania na rzeczywistość w celu dokonywania zmian, czyli sterowania (*navigator* (łac.) – sternik).

Pierwsze urządzenia automatyczne powstawały już przed naszą erą, ale ich rozwój nastąpił w XVIII wieku (przykłady przedstawiono w rozdziale 1).

Dzisiejsi inżynierowie automatycy projektują i budują urządzenia sterujące, które przeprowadzają obiekty wzdłuż trajektorii ich ruchu. Trajektoria ta musi być wcześniej opracowana we współpracy z technologami danego procesu. Takimi podstawowymi urządzeniami sterującymi są regulatory, sterowniki cyfrowe, komputery sterujące oraz czlonki wykonawcze.

Użytkownika urządzeń technicznych nie dziwi obecność prostych regulatorów w żelazku, pralce, lodówce, klimatyzacji czy w spłuczce WC. Przyzwyczailiśmy się też do bardziej zaawansowanych układów regulacji, np. we współczesnych aparatach fotograficznych, w których w czasie pojedynczego naciśnięcia na spust migawki zdąży zadziałać kolejno kilka układów automatyki:

- nastawiania ostrości obiektywu (autofocus),
- pomiaru światła i ustawiania przesłony/migawki,
- sterowania długością błysku flesza na podstawie pomiaru odbitego od obiektu światła,
- stabilizacji obrazu (w droższych modelach),
- przewinięcia klatki filmu lub przeładowania pamięci EPROM.

O istnieniu tych układów użytkownik wie, bo informują o tym instrukcje obsługi sprzętu. W wielu jednak urządzeniach jest automatyka, ale spełnia swoją rolę dyskretnie (nie wspominają o niej instrukcje użytkowania). Przeciętni użytkownicy telewizorów nie wiedzą więc o istnieniu układu automatycznej regulacji luminancji, układu regulacji wzmacnienia sygnału wizji ARW czy układu regulacji częstotliwości odchylenia stabilizującego obraz. Nie każdy użytkownik komputera wie, że wewnętrzny obudowy dysku twardego HD musi być układ regulatora do znalezienia ścieżki i stabilizacji położenia odczytująco-piszącej głowicy magnetycznej oraz układ stabilizacji prędkości obrotowej talerzy. W napędzie CD-ROM nad bezdotykowym śledzeniem przez głowicę laserową ślimakowo wypalonej ścieżki pracują cztery układy regulacji. Wszystkie zegary mechaniczne konstruowane od

1793 roku (A. Breguet) do dziś działają dokładnie, bo zawierają w sobie sprzężenie zwrotne stabilizujące częstotliwość drgań.

Liczba mikroprocesorowych układów automatyki w nowoczesnym samochodzie liczy się już w dziesiątki [0JUR], m.in. są to:

- 1) automatyczna skrzynia biegów;
- 2) automatyka kąta wyprzedzenia zapłonu i wtrysku paliwa;
- 3) automatyka czystości spalin (stabilizacja składu mieszanek lambda równe 1);
- 4) automatyczne ssanie (obroty biegu jałowego przy zimnym silniku);
- 5) stabilizacja temperatury płynu chłodzącego;
- 6) układ samoregulujący luz zaworowy;
- 7) system ABS – antypoślizgowy, pulsacyjnie hamujący auto na granicy poślizgu (1976 r.);
- 8) system ASR – kontroli trakcji, sterujący momentem napędowym przy przyśpieszaniu;
- 9) DSC lub ESP – system stabilizacji toru jazdy przez hamowanie wybranych kół;
- 10) AMVAR – nastawiania twardości zawieszenia hydraulicznego i stabilizacja przechylew, zwany też CDC (*Continuous Damping Control*) lub ABC (*Active Body Control*),
- 11) automatyczne dostosowanie prześwitu zawieszenia do szybkości auta i terenu;
- 12) DAV – system inteligentnego wspomagania kierownicy (przełożenie zależne od szybkości);
- 13) TEMPOMAT – układ utrzymujący zadaną prędkość jazdy bez względu na kształt terenu;
- 14) ACC (*Adaptive Cruise Control*) – układ kontrolujący odległość od poprzednika w kolumnie (inna firmowa nazwa DISTRONIC);
- 15) HHC (*Hill Hold Control*) – układ automatyki hamulca ręcznego, samoczynnie zwalniającego przy ruszaniu pod góru (też HSA – *Hill Start Assist*);
- 16) EBP (*Electronic Brake Prefill*) – układ minimalizujący odstęp szczęk hamulcowych (wybierania luzu), gdy kierowca gwałtownie zdejmie nogę z pedalu gazu;
- 17) układ realizujący skracanie i napinanie pasów bezpieczeństwa w razie kolizji;
- 18) układ uruchamiający poduszkę automatyczne (czas reakcji 20 ms);
- 19) automatyczna klimatyzacja, osobna dla kierowcy i pasażera;
- 20) samoczynnie przełączające się światła z drogowych na mijania;
- 21) samoczynnie włączające się wycieraczki, których częstotliwość pracy jest uzależniona od szybkości auta;
- 22) inne mniej ważne układy działające automatycznie takie jak:
zmiana pozycji lusterek przy cofaniu, zapamiętywanie pozycji foteli, aktywne zagłówki przesuwające się do przodu w razie kolizji, alarm w razie kradzieży + immobilizer, zamkane automatycznie drzwi po uruchomieniu silnika, zmienny poziom głosu radia w zależności od szybkości i hałasu, samogaszące się z opóźnieniem światła mijania (po zaparkowaniu oświetlają jeszcze drogę kierowcy) i światła pod sufitem (oświetlają wnętrze jeszcze po zamknięciu drzwi).

Ciekawym układem automatyki samochodowej wykorzystującym sprzężenie zwrotne jest aktywny układ redukcji hałasu. Częstotliwościowa analiza szumu on-line, generacja szumu o przeciwnej fazie i podanie takiego dźwięku do słuchawek ma podnieść komfort pracy kierowców i operatorów ciężkiego sprzętu [0MET; 0HAE].

Pod kątem struktury sterowania i inteligencji, największym i najbardziej rozbudowanym z wymienionych systemów jest system stabilizacji toru jazdy zwany niekiedy popularnie ESP (*Electronic Stabilization Program*). Firmowe nazwy bardziej akcentują jego funkcje sterowania dynamicznego, np. DSC (*Dynamic Stabilization Control* – BMW), VDC (*Vehicle Dynamic Control* – Subaru, Alfa Romeo) lub VSC (*Vehicle Stabilization Control* – Toyota). System ten stabilizuje tor jazdy i podnosi poziom bezpieczeństwa aktywnego. Wcześniej zostały opracowane systemy, które zapobiegają poślizgom wzdułu kierunku jazdy: system ABS do kontroli hamowania i ASR do kontroli trakcji (zwany też TCS-Seat lub TRC-Toyota). System ESP został opracowany w 1995 r. przez firmy Bosch i Mercedes. Współdziałając z ABS i ASR, przeciwdziała on poślizgom bocznym i obrotom wokół pionowej osi pojazdu. System musi rozpoznać różnicę między kierunkiem pożdanym a rzeczywistym. W tym celu analizuje prędkości poszczególnych kół oraz wykorzystuje informację z czujnika prędkości kątowej auta, bezwładnościowego czujnika przyspieszenia poprzecznego, kąta skrętu kierownicy, ciśnienia w układzie hamulcowym i urządzenia wtryskowego (obciążenie silnika). System wybiera koła, które powinny być przyhamowane. Dodatkowo system może redukować moment napędowy. Ta akcja wyprowadza samochód z poślizgu. Sterownik mikroprocesorowy pełni rolę wielowymiarowego regulatora cyfrowego obsługującego proces o wielu wejściach (cztery hamulce i pompa wtryskowa) i wielu wyjściach (prędkość, przyśpieszenie poprzeczne i moment obrotowy). Algorytm oprogramowania czasu rzeczywistego musi zawierać elementy inteligencji w ocenie faktu wystąpienia poślizgu bocznego (brak bezpośredniego czujnika np. wizyjnego wskazującego na zerwanie bocznej przyczepności opon) i wykazuje się bardzo szybką reakcją przeciwdziałającą temu poślizgowi.

Osobnym działem techniki, gdzie automatyka odgrywa decydującą rolę, są zastosowania lotnicze i rakietowe. Pierwsze wdrożone układy automatycznego sterowania i stabilizacji lotu zastosowano w pociskach V1 (1943–45 r.). Wyrzucane pod odpowiednim kątem, między innymi w stronę Londynu, utrzymywały zadany kierunek i wysokość dzięki wykorzystaniu układu pilota automatycznego działającego w oparciu o żyroskopy (zamknięta pętla regulacji), choć czas lotu był sterowany programowo przez zegar mechaniczny (tylko w pętli otwartej). Po upływie zadanego czasu zamknięty był zawór paliwa i przestawiany ster wysokości. Pocisk spadał swobodnie i jeśli nie było wcześniej zakłóceń lotu, często trafiał w cel.

Pierwsze metody stabilizacji pionowego startu rakiet były testowane od 1935 r. w USA (Robert H. Goddard). Jednak pierwsze wdrożone rozwiązania pojawiły się w Niemczech (Wernher von Braun) pod koniec drugiej wojny światowej w rakietach V2. Zasięg tych rakiet nie przekraczał 300 km.

Po opanowaniu najważniejszych problemów inżynierskich związanych przede wszystkim z konstrukcją silników dużej mocy i opracowaniem wydajniejszych układów automatyki, systemy rakietowe pozwoliły na pierwsze krótkie lotyokoziemskie (*Sputnik* – 1957 r.). Od tego czasu sukcesywnie podnoszono niezawodność i precyzję sterowania układami latającymi. Celność dzisiejszych ziemskich rakiet typu *Cruise* i *Tomahawk* jest stuprocentowa. Wspomagane systemem GPS (*General Positioning System*) mają dodatkowo wizyjny układ

śledzenia terenu on-line i wybierania drogi zgodnie z mapą cyfrową umieszczoną w pamięci [0WWW].

Również sterowanie trajektorią lotu sond międzyplanetarnych osiąga dziś wysoką precyzję. Pierwszą sondą kosmiczną, która miała za zadanie trafić w mały obiekt oddalony o miliony kilometrów od Ziemi, była sonda *Near Shoemaker* opracowana przez zespół z Uniwersytetu Johna Hopkinaa. Sonda wystartowała w lutym 1996 r. i po przebyciu 4,7 mld kilometrów osiadła na podłużnej asteroidzie 433 Eros w lutym 2001 r., przesyłając do końca zdjęcia powierzchni (ostatnie z wysokości 120 m). Trafienie w obiekt o długości 40 km było możliwe poprzez wysoką jakość wizyjne układy samonaprowadzające z dokładnością 50 mradianów/s i 16 silników korekcyjnych [0NEAR].

W październiku 1997 NASA Jet Propulsion Laboratory i European Space Agency wysłały z Florydy rakietą Tytan-4-Centaur, sondę kosmiczną Cassini (5,7 tony). Po 7 latach lotu (3,5 mld km), w styczniu 2005 sonda Cassini znalazła się w pobliżu Saturna i uwolniła próbnik lądujący Huygens, który wylądował na Tytanie – największym księżyku Saturna (średnica 5200 km) i zaczął przesyłać zdjęcia i pomiary.

W lutym 1999 r. wystrzelono sondę kosmiczną Stardust (380 kg), która w styczniu 2006 powróciła na Ziemię. Po siedmioletniej podróży i przebyciu 3,2 mld km sonda, przewiozła próbki cząsteczek z ogona komety Wild2, do której jądra zbliżyła się na odległość 250 km. Są to pierwsze w historii próbki dostarczone na Ziemię z takiej odległości.

Jeszcze większa precyzja sterowania musiała być zagwarantowana w ostatnim przedsięwzięciu. W styczniu 2005 NASA wystartowała sondę Deep Impact, która po przebyciu 431 mln km znalazła się w pobliżu komety Temple 1 (średnica jądra 6 km). W lipcu 2005, w odległości 864 tys. km od komety, sonda uwolniła autonomiczny próbnik (*Impact*). Próbnik o wadze 370 kg, wyposażony w optyczną aparaturę do samonaprowadzania z dokładnością 1 μradiana i w komputery sterujące gwarantujące precyzję sterowania z dokładnością 1 mm/s, wykonał manewr trafienia w jądro komety, wykonując po drodze serię zdjęć i pomiarów.

Szczegóły techniczne wszystkich misji i opisy układów sterujących są umieszczone na stronach WWW agencji NASA. Biorąc pod uwagę odległość i wielkość celów, należy z podziwem oceniać coraz większą precyzję aparatury pozycjonującej, komputerów sterujących oraz jakość oprogramowania sterującego.

Niestabilne zachowanie startującej pionowo rakiety może być zamodelowane zachowaniem się wahadła odwróconego, w którym środek ciężkości jest powyżej ruchomego punktu podparcia. Ruchomość punktu podparcia zapewnia w rakiecie odchylana główna dysza wylotu gazów, lub przy dyszach wielokrotnych – dozowanie paliwa do odpowiednich dysz, a w modelu wahadła np. ruchomy wózek.

Rozwój automatyki pilotażu w układach latających ma swoją długą historię [0PEM]. Pierwszy pełny przelot Atlantyku z użyciem pilota automatycznego miał miejsce w 1947 r. Wdrożenie w Europie w lotnictwie pasażerskim pilotażu automatycznego obejmującego fazę lądowania nastąpiło w 1969 r.

W klasycznych rozwiązaniach konstrukcyjnych samolotów, środek jego masy był położony przed usytuowanym na skrzydłach środkiem wyporu aerodynamicznego. Stąd dwie

siły skierowane w dół – ciężar i siła nacisku na duże tylne stateczniki (ustawione końcami do góry pod „ujemnym” kątem), równoważą centralną siłę wyporu skierowaną do góry. W locie poziomym samolot jest stabilny, bo małe wychylenia w pionie osi podłużnej samolotu są samoczynnie stabilizowane, bez ingerencji pilota, przez zanik lub wzrost siły nacisku pionowego na stateczniki tylne. Gdy dziób pochyla się w dół, siła nacisku na tylne stateczniki rośnie, gdy dziób odchyla się do góry siła nacisku na stateczniki maleje. Widoczny jest stabilny efekt dźwigni dwustronnej podpartej w punkcie wyporu. Ma on swój analog w wiszącej stabilnej pozycji wahadła pionowego [0GUB].

W nowoczesnych rozwiązańach (Airbus A330, F16, Mirage 2000) środek wyporu jest wysunięty przed środek masy. Stąd dwie siły skierowane do góry – wypór skrzydeł i wypór na tylnych małych statecznikach (ustawionych w dół pod „dodatnim” kątem) powinny równoważyć centralną siłę ciężkości skierowaną w dół. Nie występuje tu samoczynny efekt wzrostu lub maleńcia siły wyporu na statecznikach tylnych w takiej dźwigni dwustronnej z punktem podparcia w środku ciężkości. Do stabilizacji poziomego lotu musi być zrealizowana aktywna funkcja reakcji tylnymi statecznikami i dodatkowo ciągiem silnika, wykonywana przez pilota lub pilota automatycznego. Taki układ odzwierciedla niestabilną reakcję wahadła odwróconego, w którym ruch wózka jest analogiem aktywnej reakcji stateczników. Umieszczenie w nowoczesnych konstrukcjach odrzutowych punktu przyłożenia siły wyporu przed środkiem ciężkości (blżej dzioba) kwalifikuje je jako „statycznie niestabilne”, co można interpretować jak wahadło odwrócone ułożone poziomo. Dodatkowo zastosowanie zmiennej geometrii dysz wylotowych daje efekt „wektorowania ciągu”, co zwiększa zwrotność samolotu i zmniejsza zużycie paliwa, ale wymaga precyzyjnej stabilizacji sterami i ciągiem silnika. Bez pracującego silnika, dzisiejsze myśliwskie samoloty odrzutowe nie mogą latać. Dawne samoloty śmigłowe z „napędem przednim” były stabilniejsze w locie i mogły wylądować lotem ślizgowym [0STB]. Szczytem osiągnięć automatyki komputerowej był samolot Concorde, w którym zastosowano dynamiczne przesuwanie środka ciężkości dla prędkości poddźwiękowych i naddźwiękowych. Zmienną statykę uzyskiwano w czasie lotu poprzez przepompowywanie paliwa ze zbiorników przednich do tylnych.

W wielu procesach technologicznych obserwuje się zachowania niestabilne (np. reakcja łańcuchowa w reaktorze jądrowym). Procesy te wymagają niezawodnych algorytmów sterownia i pewnych układów automatyki. Niezawodność sprzętową uzyskuje się np. poprzez redundancję układów (nawet 3-krotną) [0STG]. Obowiązują tu dwie „szkoły”: trzy takie same komputery z tym samym oprogramowaniem, które wyłączane są kolejno w razie ich awarii, lub trzy komputery różnych typów i trzy oprogramowania napisane niezależnie przez trzy różne zespoły, co ma wyeliminować ukryte „fabryczne” błędy sprzętu i oprogramowania.

Ważnym działem zastosowań automatyki, o którym należy wspomnieć, to robotyka. Oprócz dobrze poznanych i wdrożonych zadań sterowania ramionami manipulacyjnymi, inżynierskim celem jest zbudowanie robota kroczącego. Jest to z natury układ niestabilny. Wahadło odwrócone, w którym środek ciężkości jest powyżej ruchomego punktu podparcia, pokazuje problem stabilizacji takiego robota. Jego budowa to trudne zadanie – nie do pokonania bez wielu układów automatyki. Pierwowzorem jest żywy człowiek, którego cia-

ło ma 200 stopni swobody i nad którego stabilizacją chodu czuwają dziesiątki mięśni [0ABJ]. Nawet dla człowieka nauka chodzenia trwa rok, nauka jazdy na łyżwach, nartach czy rowerze też zabiera czas. Ryby kostnoskieletowe również uzyskały ewolucyjnie zdolność do wykorzystania grawitacji, do zwiększenia zwrotności w czasie ucieczki. Ich środek ciężkości jest powyżej pęcherza pławnego (środka wyporu), stąd muszą, nawet w czasie spoczynku, utrzymywać cały czas równowagę ruchami płetw.

Jednym z pierwszych przykładów rozwiązania dwunożnego robota humanoidnego jest robot ASIMO (*Advanced Step in Innovative Mobility*) japońskiej firmy Honda, o 26 stopniach swobody, którego opis można znaleźć na stronie internetowej [0ASIM]. W zewnętrznym komputerze sterującym, komunikującym się bezprzewodowo z robotem, znajdują się programy dla stabilizacji i płynności chodu, ruszania z miejsca, zatrzymywania się i wchodzenia po schodach. Dla płynnej zmiany kierunku, wymagało to np. opracowania sterowania opartego na predykcji ruchu, poprzez wcześniejsze przesunięcie w bok głównego środka ciężkości i pozycji stopy, zanim ruch do przodu zostanie wykonany (*Intelligent Walking Technology*). Wyzwaniem dla automatyki są więc samoczynne roboty, zdolne do podejmowania właściwej decyzji ze zbioru możliwości, przygotowanych przez projektanta.

Zastosowanie zaawansowanych systemów elektronicznych w rozbudowanych układach mechanicznych, ich wzajemna współpraca i przenikanie spowodowało, że na początku lat 90. w inżynierii mechanicznej powstał nowy termin na określenie takich technologii – mechatronika. Jej naturalne związki z robotyką i dyskusja o szerszym zakresie badawczym tej dyscypliny są tematem całego numeru miesięcznika *IEEE Robotics & Automation Magazine* [0ALD].

Wiele układów technologicznych zachowuje się jednak stabilnie. Ich praca może być modelowana przez zachowanie się wahadła prostego w polu grawitacyjnym, wiszącego w swym stabilnym punkcie pracy, podczas gdy ruchomy punkt zaczepienia lub wyporu (którym można sterować) jest powyżej punktu ciężkości. Zachowanie takie pojawia się w układzie pracującej suwnicy, w zjawisku kołysania statku czy w locie balonu. Takie stabilne układy są łatwiejsze i bezpieczniejsze do sterowania. Łatwością sterowania można więc tłumaczyć wyższość przedniego napędu samochodu, gdy siła ciągu zaczepiona jest „powyżej” środka bezwładności („grawitacji”) w takim poziomym wahadle. Popularnie mówi się, że taki podsterowany samochód nie wykazuje tendencji do zarzucania przy przyspieszaniu. Tylny napęd samochodu, barki rzecznej czy łodzi motorowej wykazuje analogie do wahadła odwróconego. Jak wspomniano, takie układy strukturalnie niestabilne są jednak szybsze i przy precyzyjnym sterowaniu – wydajniejsze. Stąd w samochodach wyścigowych stosuje się ciągle tylny napęd. Mówiąc, że przy przyspieszaniu taki samochód wykazuje nadsterowność, czyli samorzutnie pogłębia zakręt, zarzucając tyłem.

Podobnie najnowsze rozwiązania napędu dużych statków pasażerskich stosują komputerowo sterowane, ruchome układy dwóch lub czterech pędników śrubowych podwieszanych obrotowo na rufie, co pozwoliło na całkowite zrezygnowanie z klasycznego steru kierunku, jednocześnie wielokrotnie podnosząc zwrotność statku. Dodatkowo wysuwane z kadłuba boczne płaszczyzny stabilizatorów o zmiennym kącie natarcia, zmniejszają kołysanie statku na fali.

Inne znane zastosowanie automatyki w wojsku to stabilizacja położenia lufy czołgu, co pozwala na prowadzenie ognia w czasie jazdy bez względu na nierówności terenu. Przykładem automatyki jest również konstrukcja helikopterowych układów naprowadzania na cel działa za pomocą ruchów głowy pilota (specjalny hełm), konstrukcja samonaprowadzających się pocisków, śledzących w podczerwieni fragmenty celu o podwyższonej temperaturze i w końcu samoloty pionowego startu.

Automatyka nie ominęła urządzeń medycznych, znajdująca zastosowanie w automatycznych dozownikach insuliny, rozrusznikach serca [0MAE] czy w zaawansowanych narzędziach chirurgicznych. Do tych ostatnich należy np. laserowe urządzenie do operacji oka. Mimowolne ruchy gałki ocznej uniemożliwiłyby ręczne pozycjonowanie i prowadzenie promienia lasera przez chirurga. Kamera śledzi położenie żrenicy, zgodnie z jej ruchami układ automatyki natychmiast koryguje i dodatkowo przesuwa rzeczywisty promień lasera. Układ wizji koryguje też położenie obrazu oka na monitorze. Chirurg na ekranie widzi tylko nieruchomą żrenicę i miejsce operacji, odnosząc wrażenie nieruchomości pola operacyjnego oraz jedynej i pełnej kontroli nad promieniem lasera.

Jednym z nowszych rozwiązań automatyki, stosowanym od 1992 r. w nietypowej dziedzinie technologii, jest zastosowanie układów sterowania i stabilizacji w architekturze przy konstrukcji wieżowców i mostów wiszących. Specjalne aktywne tłumiki i duże masy (od 20–60 ton) umieszczone na ostatnich piętrach wieżowców i poruszane po szynach za pomocą siłowników hydraulicznych zmieniają środek ciężkości wieżowca (wieżowiec też reprezentuje model wahadła odwróconego). Zapobiega to dużym wychyleniom konstrukcji w czasie wichur i trzęsień ziemi [0SPB]. Również automatyka zwykłych budynków w zakresie klimatyzacji, centralnego ogrzewania, sterowania oświetleniem, instalacją przeciwpożarową i alarmową wchodzi coraz bardziej do użytku, tworząc tzw. budynek inteligentny.

W najdłuższym moście świata Honshu-Shikoku w cieśninie Akashi (Japonia) dwie 283-metrowe wieże oddalone od siebie o 2 km podtrzymują dwie czterokilometrowej długości liny nośne, na których podwieszona jest sześciopasmowa jezdnia. Konstrukcja mostu gwarantuje wytrzymanie efektów trzęsienia ziemi o sile 8,5 stopnia Richtera. We wnętrzu każdej ze stalowych wież znajduje się 20 aktywnych stabilizatorów masy, które są poruszane wahadłowo przez układy sterujące dla zapobiegania wychyleniom konstrukcji. Zdalny nadzór nad zachowaniem mostu prowadzony jest przez komputerowe centrum monitoringu mostu. Taka technologia staje się standardem. Widać ją również w oddanym do użytku w 2004 roku 2,25-kilometrowym moście wiszącym w Atenach. Umieszczono w nim 300 czujników naprężeń lin i trójkierunkowych czujników przyśpieszenia ruchu jezdni w celu monitorowania i ostrzegania przed drganiami i kołysaniami konstrukcji w czasie trzęsień ziemi [0REU]. Pomiary zbierane są przez cztery komputerowe jednostki akwizycji danych, które decydują o ostrzeganiu i wyłączaniu mostu z ruchu. Niedocenianie zagadnienia drgan konstrukcji mostowych było powodem poważnych katastrof. Na przykład w roku 1940 zerwaniu uległ długi most wiszący Narrows Bridge (853 m × 12 m) w Tacoma (USA), gdy częstotliwość podmuchów wiatru zbliżyła się do częstotliwości drgan rezonansowych.

Również należy wspomnieć o nowoczesnych systemach produkcyjnych w rolnictwie i w przemyśle produkcji żywności. Automatyka znajduje tam zastosowanie w sterowaniu

mikroklimatem i biosferą szkarni i ferm (*greenhouses*) do podniesienia wydajności upraw. Metodologia modelowania i prognozowania matematycznego wywodząca się z automatyki, stosowana jest też do prognozowania ilości opadów i stanu pogody. Nadzorowanie zarówno wilgotności gleby pod uprawą, jak i wilgotności ziarna w elewatorach, pozwala na automatyczne uruchomienie systemów nawadniająco-zraszających lub suszących. Zagadnienia te są na tyle ważne, że został im poświęcony cały numer miesięcznika „IEEE Control Systems” [0SIG].

Aby zamknąć paletę różnorodności dziedzin stosowania automatyki, należy też wymienić nowoczesne systemy komputerowego sterowania systemami transportowymi. Dotyczy to np. systemów sterowania komunikacją publiczną miast, które pomagają w uniknięciu powstawania zatorów, zapewniają optymalną obsługę pasażerów i optymalne wykorzystanie taboru w sensie zużycia paliwa i minimalizacji zanieczyszczeń [0ADA]. Również w systemach kontroli ruchu lotniczego wykorzystuje się komputerowe metody wspomagania decyzji, jednak z uwagi na wymogi bezpieczeństwa, układy te wciąż nie doczekały się pełnej automatyzacji. Ale już największy na świecie lotniskowy system automatycznego transportu, śledzenia, identyfikacji i komputerowej dystrybucji bagażu pasażerskiego pomiędzy terminalami w Hongkongu, obejmujący 20 km linii taśmociągowych, jest w pełni zautomatyzowany i bezobsługowy.

Jednak oprócz tych spektakularnych dziedzin zastosowań automatyki, wciąż największe obszar zajmują procesy przemysłowe o technologii ciągłej i dyskretnej. Produkcja wyrobów hutniczych stali i szkła, chemicznych, paliwowych, budowlanych, wytwarzania energii oraz innych wyrobów przemysłu lekkiego to podstawowe pole aplikacji automatyki. Zwłaszcza rozbudowane są komputerowe systemy sterowania w kompleksowych procesach technologicznych. Zaangażowanie do tych systemów sprzętu programowalnego, który zrealizowałby skomplikowane procedury nadzoru i sterowania on-line setek zmiennych procesowych, jest tak duże, że dziś inżynierowie automatyki musieli opanować też dziedzinę **informatyki technicznej**. Nie wystarcza już znajomość reguł i metod sterowania, potrzebna jest do tego celu umiejętność tworzenia profesjonalnego oprogramowania czasu rzeczywistego. Szerze spojrzenie na **informatykę techniczną** wskazuje na jej powiązania z **informatyką zarządzania** zajmującą się zbieraniem i przetwarzaniem informacji na cele zarządzania. Bardzo często te dwa wspomniane profile występują nierożelcznie w dużych nowoczesnych wielozakładowych przedsiębiorstwach, w postaci komputerowych systemów zintegrowanych CIM (*Computer Integrated Systems*), definiując hierarchiczną wielopoziomową i wielowarstwową strukturę zarządzania i sterowania. Na każdym poziomie występuje zróżnicowanie tych systemów zarówno w warstwie sprzętowej, jak i programowej. Poziom najwyższy (*Primary Control*) dotyczy sterowania procesem technologicznym i zawiera problematykę czujników, cyfrowych przetworników, programowalnych sterowników, interfejsów przemysłowych i komputerowych systemów sterowania bezpośredniego (*Direct Control*) oraz nadzędnego (*Supervisory Control*). Wiąże się z tym zagadnienie oprogramowania tych systemów, z reguły z wykorzystaniem systemów operacyjnych czasu rzeczywistego i systemów wizualizacji typu SCADA (*Supervisory Control and Data Acquisition*). Systemy te wyposażone w bazy wiedzy i reguły eksperckie wykazują cechy

inteligentnego operatora, np. w sytuacjach krytycznych zapewniając automatyczne zamknięcie awaryjnych cykli produkcyjnych ESS (*Emergency Shutdown Systems*).

Oprogramowanie poziomów wyższych przedsiębiorstwa wiąże się głównie z oprogramowaniem systemów MRP (*Manufacturing Resource Planning*) i dotyczy systemu informatycznego zarządzania przedsiębiorstwem w działach zaopatrzenia, produkcji, marketingu, sprzedaży, finansów i kadr.

W przedstawionym skrótnym zestawieniu zadań automatyki omawiano głównie problemy, które można ogólnie ująć jako problemy automatyki procesowej (*Control Process Systems*). Zadaniem tych systemów jest najlepsze przeprowadzanie procesu technologicznego pomiędzy kolejnymi punktami pracy, a w końcu prowadzenie procesu w pobliżu wybranego optymalnego punktu tak, aby bez względu na zakłócenia otrzymywać produkt o wysokiej jakości, w sposób wydajny i oszczędny.

Inżynierowie wdrażający układy automatyki, mają jednak też na uwadze inny typ układów, które mają też do spełnienia inne funkcje – pilnują granic bezpieczeństwa obszaru pracy. Są to układy automatyki zabezpieczeniowej SIS (*Safety Instrumented Systems*). Mają one za zadanie pilnowanie bezpiecznej odległości zmiennych procesowych od ich ograniczeń – co ma gwarantować pozostawanie punktu pracy w obszarze stabilności. Układy te są odpowiedzialne za rozpoznawanie zakłóceń najbardziej niebezpiecznych i sytuacji awaryjnych. Do ich zadań należy alarmowanie o przekroczeniu ograniczeń technicznych i technologicznych, wyłączanie zepsutych i włączanie zapasowych urządzeń pomiarowo-kontrolnych (które pracują na zasadach gorącej rezerwy albo oczekiwania) takich, jak czujniki, przetworniki, regulatory, nastawniki, tory pomiarowo-komunikacyjne, zamykanie lokalnych niebezpiecznych cykli produkcyjnych, włączanie sekwencji blokad i awaryjne odstawianie instalacji.

W książkach poświęconych teorii technik sterowania dynamicznego mało miejsca przeznacza się dla układów SIS, wychodząc z założenia, że są to głównie problemy sekwencyjno-logiczne. Należy jednak zwrócić uwagę, że te zagadnienia mogą również stawać trudne teoretyczne zadania do rozwiązania, takie jak projektowanie układów diagnostyki i układów odpornych na awarie [0BLM]. Łączenie układów automatyki procesowej i zabezpieczeniowej w jednych urządzeniach (np. w jednym sterowniku PLC) jest zabronione, zwłaszcza dla procesów, w których występuje ryzyko katastrofalnej awarii. W celu podniesienia niezawodności układu automatyki, międzynarodowe unormowania zalecają dla instalacji przemysłowych, rozdzielenie funkcji sterujących od funkcji zabezpieczeń. Błędy automatyki zabezpieczeniowej (awarie sprzętowe i błędy oprogramowania) wraz z błędami operatorów mogą doprowadzić do poważnych katastrof, np. elektrowni jądrowych (Czarnobyl) czy instalacji chemicznych (Bhopal – Indie).

W niniejszej książce główną uwagę skupiono na automatyce procesowej, jakkolwiek problemy automatyki zabezpieczeniowej omówiono szerzej w rozdziale 6.

Z tego krótkiego wstępu ma wynikać ważna rola automatyki i konieczność obecności jej układów, we wszystkich urządzeniach technicznych, w których zachodzą jakiekolwiek zmiany wielkości fizycznych, nad którymi chceliśmy panować. Nieraz użytkownik chciałby, aby pomimo pojawienia się zewnętrznych zakłóceń nie obserwowano zmian

zmiennych procesowych, np. zmiany temperatury w akwarium, zmiany częstotliwości natrojonej automatycznie stacji radiowej (co eliminuje zanikanie głosu), ustalonej pozycji satelity na orbicie geostacjonarnej. Nieraz natomiast pożądane jest, aby zmiany pewnych wielkości mogły być duże i mogły zachodzić tak szybko, jak sobie tego życzymy, np. zmiana temperatury stopki żelazka przy przestawieniu pozycji zadajnika temperatury z „nylon” na „bawełna”, przejście do nowego punktu pracy instalacji chemicznej czy pościg za uciekającym celem przez rakietę „ziemia-powietrze”.

Urządzenia automatyki można więc znaleźć wszędzie. Podnoszą one standard życia, jak również wydajność i bezpieczeństwo procesów technologicznych.

Czy nauka zwana automatyką, teorią sterowania lub cybernetyką techniczną odkryła już wszystkie metody i algorytmy związane z techniką sterowania? Z jednej strony wiadomo, że poziom teorii i całej wiedzy na ten temat oraz poziom dzisiejszej elektroniki wystarczają do budowy bardzo wydajnych procesorów sterujących, zapewniających precyzyjną pracę urządzeń technicznych, których celem jest albo wykonanie produktu o gwarantowanej jakości (stężenie 0,9999), albo wykonanie czynności z gwarantowaną precyzją (samoczynne lądowanie samolotu). W wielu jednak zadaniach sterowania nie wiadomo, jak budować sterowniki, zwłaszcza dla układów nielinowych, wielowymiarowych czy wysokiego rzędu, których inteligencja wykazywałaby dodatkowo cechy adaptacyjności do zmiennych warunków. Inżynieria, na przykład, nie umie budować małych urządzeń latających MAV (*Micro Air Vehicles*), które potrafiłyby latać tak precyjnie jak mucha [0ZBR]. Wydaje się, że dzisiejsze algorytmy sterowania polegające na wykorzystaniu pomiarów z kilkunastu czy kilkudziesięciu czujników i przeliczaniu skomplikowanego zestawu równań różniczkowych modelu (np. 1,1 mln linii kodu dla modelu samolotu LMC F35) przeciągają z systemem sterującym muchy, która używa pomiarów z 80 000 czujników rozrzuconych na ciele. Mózg muchy zbudowany jest zaledwie z 338 000 neuronów. Gdyby traktować neuron jako przerzutnik w układzie cyfrowym, to układ mózgu posiadałby niewielką moc małego procesora. W mózgu główną rolę odgrywa jednak też ilość połączeń, a neuron pracuje analogowo, a nie binarnie. Pomimo to szacuje się, że mózg muchy 98% mocy poświęca przetwarzaniu sygnałów, a tylko 2% rozpoznawaniu niebezpieczeństwa i algorytmom sterowania [0ZBR1; 0SZJ].

Również bardzo charakterystyczną cechą obwodowego układu nerwowego człowieka jest duża przewaga nerwów czuciowych – aferentnych (95%), doprowadzających sygnały sensoryczne słuchowe, wzrokowe, węchowe, skórne, trzewne i inne do mózgu, w stosunku do nerwów ruchowych – eferentnych (5%) odprowadzających sygnały z mózgu do niewielu narządów wykonawczych (mięśnie, gruczoły). Efektywne sterowanie opiera się więc na możliwie pełnej informacji pomiarowej, przetwarzanej dla prostego modelu.

W układach technologicznych montaż wielu czujników fizycznych zmiennych procesowych może być nieopłacalny. Przyszłośćą mogą więc być czujniki programowe (*softsensors*), zbudowane dla niemierzalnych zmiennych stanu przy wykorzystaniu zmiennych mierzalnych i obserwatorów stanu. Obecność tych ostatnich komplikuje budowę warstwy przetwarzania sygnałów (*preprocessing*), ale upraszcza warstwę sterowania (np. regulatory proporcjonalne LQR – *Linear Quadratic Regulator*).

W niniejszej monografii położono duży nacisk na nowe metody obserwacji stanu.

W kraju i zagranicą co roku ukazują się różne monograficzne opracowania z dziedziny automatyki. Niektóre oferują szersze spojrzenie na całokształt tematyki, np. [0CAF; 0FRB; 0GOG; 0KAT; 0MAJ; 0RUW; 0SOE; 0ŽAS]. Inne koncentrują się na wybranych aspektach dziedziny, np. [0FRG; 0GOH1; 0GOH2; 0SOT; 0WEP; 0WAK].

Celem niniejszej monografii jest przedstawienie całokształtu nowych wyników uzyskanych przez autora w tematyce obserwacji stanu. Odtwarzanie niemierzalnego wektora stanu na podstawie pomiarów wyjścia procesu znajduje zastosowanie zarówno w zagadniach sterowania, jak i diagnostyki (*soft sensing*). W zakresie klasycznych metod modelowania i sterowania autor dokonał wyboru i prezentacji tych zagadnień, które mogły być pomocne w omówieniu zagadnień sterowania od stanu, akcentując na opracowanych przez siebie przykładach zakres ich stosowności. Niektóre przykłady są bardziej rozbudowane i prezentują tzw. *Case Study* (podrozdz. 2.3.4, rozdz. 10 i 18).

W pierwszych pięciu rozdziałach monografii przedstawiono skrótną prezentację zagadnień modelowania i analizy dynamiki systemów skończenie wymiarowych. Zaakcentowano fenomen zjawiska sprzężenia zwrotnego, przedstawiając możliwość jego wykorzystania np. do identyfikacji stanów ustalonych i charakterystyk układów niestabilnych. Omówiono własności układów sterowania ze sprzężeniem zwrotnym. W rozdziale 6 przedstawiono klasyczne i nietypowe struktury układów regulacji. W następnych rozdziałach przedstawiono metody projektowania regulatorów od wyjścia i od stanu, również w obecności zakłóceń najbardziej niebezpiecznych. Główna problematyka monografii dotycząca obserwacji i odtwarzania niemierzalnego wektora stanu obejmuje podrozdział 8.3 oraz rozdziały 12, 13, 14, 15, 16, 17 i 18. W dwóch aneksach w sposób zwarty umieszczono najważniejsze zagadnienia algebry macierzowej i analizy funkcjonalnej.

Monografia opiera się między innymi na wieloletnich wynikach badań i doświadczeniach zdobytych przez autora przy tworzeniu i wdrażaniu komputerowych systemów sterowania [18BYW1; 18BYW2; 18BYW3; 18BYW4; 18BYW5; 18BYW6; 18BYW7; 18BYW8].

1. Problemy automatyki

Automatyka to nauka o metodach poszukiwania właściwych sterowań i o sposobach ich realizacji. Steruje się pojedynczymi obiektami, układami obiektów i całymi systemami. Metodologia poszukiwań sterowania opiera się na matematycznych metodach modelowania statyki i dynamiki procesów oraz na algorytmach optymalizacji wybranych wskaźników jakości zachowania się procesu. Po znalezieniu teoretycznie najlepszego kształtu sterowania, następuje proces jego implementacji – z reguły w układzie ze sprzężeniem zwrotnym, w celu ciągłej kontroli poprawności zachowania się regulowanego procesu. Implementacja od strony sprzętowej sprowadza się do konfiguracji odpowiedniej struktury układu sterowania złożonej z elektronicznych, cyfrowych urządzeń automatyki, sterowników, komputerów i najważniejsze – algorytmów numerycznych realizujących on-line rozbudowaną sekwencję ciągłych sterowań, akcji logicznych, zabezpieczeń, samostrojenia regulatorów i raportowania. W obszarze automatyki, oprócz specyficznych zagadnień sterowania, znajdują się więc aż trzy inne dyscypliny naukowe: matematyka, elektronika i informatyka stosowana. Stąd obszar badawczy zespół zajmujących się automatyką jest szeroki i rozciąga się od matematycznej teorii sterowania układami abstrakcyjnymi nieskończonymi wymiarowymi, poprzez komputerowe algorytmy i programy przetwarzające sygnały pomiarowe i realizujące strategię sterowania (on-line), aż do konstrukcji elektronicznej kart pomiarowych i ich oprogramowania oraz budowy sterowników i członów wykonawczych. Obszerność tematyki jest widoczna w oficjalnych zestawach słów kluczowych, jakie dopuszcza się na największych konferencjach naukowych dotyczących tej dziedziny. Zaprezentujemy je pod koniec tego rozdziału.

W automatyce, jak w każdej innej dziedzinie nauki, można wydzielić zagadnienia wchodzące w zakres analizy i w zakres syntezy. Celem analizy jest zrozumienie podstawowych cech sygnałów i cech obiektów, poznanie metod ich opisu i reguł składania tych obiektów w większe układy oraz badanie własności tak uzyskanych systemów. W zakres syntezy wchodzi problem postawienia celu sterowania, metodyka wyliczania tego sterowania i zaprojektowanie regulatora realizującego to sterowanie, jak również konstruowanie innych elementów układu sterowania takich, jak np.: korektory dynamiki, człony odsprzęgające czy obserwatory stanu i estymatory parametrów.

U podstaw tematyki przedmiotu leżą dwie różne idee:

- 1) idea sprzężenia zwrotnego SZ (*feedback*),
- 2) idea sterowania optymalnego (*optimal control*).

Na ideę sprzężenia zwrotnego, wykorzystywaną w automatyce, można popatrzyć z szerszej perspektywy.

Spotyka się opinię, że występowanie sprzężenia zwrotnego w fizyce materii nieozwionej jest tak powszechnie, jak powszechna jest III zasada dynamiki Newtona. W układzie inercjalnym stwierdzenie, że siła oddziaływania ciała A na ciało B spotyka się z taką samą reakcją od ciała B na A, jest stwierdzeniem istnienia naturalnego sprzężenia zwrotnego.

Sprzężenie występuje też na przykład w polu grawitacyjnym, przy interakcji dwóch zbiorników połączonych na jednym poziomie, w których występuje samowyrównywanie się poziomów cieczy (przykład 10 w podrozdz. 4.3.5). Sprzężenie nie występuje przy połączeniu kaskadowym zbiorników bez oddziaływania wstecznego, np. przy wypływie swobodnym z jednego zbiornika do drugiego, położonego niżej. W takim wypadku, człowiek może takie sprzężenie stworzyć sztucznie i za pomocą pomp uzależnić poziomy od siebie.

Fenomen naturalnego sprzężenia zwrotnego występujący w przyrodzie ujawnił się również w zachowaniu materii żywnej. Stał się podstawą powstania i rozwoju życia na Ziemi. Wykorzystywany był przez organizmy rozwijające swój mózg i coraz obszerniejszą pamięć, zdolną do przechowywania obrazów z coraz głębszej przeszłości ich życia. Sprawny mechanizm przeglądania tej pamięci (szynkość kojarzenia) musiał zawierać procedury porównywania z wzorcem. Sprzężenie zwrotne stało się więc podstawą rozwoju procesu uczenia i rozwoju inteligencji oraz sposobem na dostosowywanie się organizmów do zmiennych warunków otoczenia, w celu przetrwania – czyli na ich ewolucję. Mechanizm sprzężenia zwrotnego umożliwił bowiem realizację prostych zasad (Darwin): szukaj i znajdź to, co jest ci potrzebne; unikaj tego, co przynosi szkodę; rób to wszystko szybciej niż konkurencja. W tych trzech regułach podstawą była wielokrotna sekwencja akcji, obserwacji efektów i korekcji akcji.

Sprzężenie zwrotne zaadoptowane praktycznie w układach technicznych dało początek nauce zwanej automatyką. Jej związki z regułami sterowania w organizmach żywych wykazywał Norbert Wiener, wprowadzając dla nauki o sterowaniu ogólny termin cybernetyka (od greckiego słowa ster, sternik) [1WIN1; 1WIN2]. Stąd teoria sterowania systemami przemysłowymi bywa nazywana cybernetyką techniczną. Odpowiednio w zakres cybernetyki ekonomicznej wchodzą metody zarządzania i sterowania w systemach ekonomicznych, a w zakres biocybernetyki – badanie sygnałów sterujących, jakie generuje mózg i realizują układy wewnętrzne organizmu.

Idea sztucznego sprzężenia zwrotnego stosowana w inżynierskich rozwiązaniach automatyki, różni się jednak w sposób zasadniczy od sprzężeń naturalnych. Sprzężenia naturalne pobierają i przekazują informację za pomocą oddziaływań fizycznych takich, jak: ciśnienia, przepływy, napięcia, prądy, temperatury, wykorzystując istniejącą energię wewnętrzną układu (np. różnicę potencjałów czy poziomów cieczy).

W rozwiązaniach technicznych przyjmuje się, że sprzężenia sztuczne są wyłącznie typu informacyjnego. Czujnik przetwarza sygnał pomiarowy, nie obciążając samego procesu mierzonego (pomijalne błędy). Stąd wymóg, aby woltomierz cyfrowy miał wysoką oporność wejściową, a turbinkowy czujnik przepływu miał zerowy opór hydrodynamiczny. Przetworzony sygnał wyjścia, przenoszony jest przez tor sprzężenia (zwykle jako sygnał elektryczny) i po wzmacnieniu służy jako sygnał wejściowy do regulatora. W tym celu wszystkie elementy toru sprzężenia wymagają zewnętrznych źródeł energii (wyjątek stan-

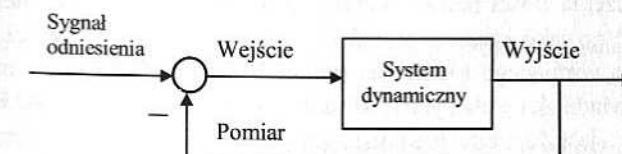
wią regulatorów bezpośredniego działania, np. regulatory ciśnieniowe w butlach tlenowych lub płynakowe regulatory poziomu). Przy zewnętrznych zasilaniach elementów toru SZ, zachowana jest ważna ich cecha – brak wzajemnej obciążalności. Pozwala ona na separację dynamiki tych elementów, stosowanie dla każdego z nich pojęcia transmitancji i stosowanie algorytmów schematów blokowych. Graficzne połączenia na schematach blokowych nie są połączaniami galwanicznymi, a tylko informacyjnymi. Ta cecha nie zawsze jest podkreślana w opracowaniach z podstaw automatyki. Problem ten zostanie też omówiony w dalszych rozdziałach.

Automatyka jest więc nauką, która bada wszelkie problemy wynikające z istnienia i wykorzystania sprzężenia zwrotnego dla celów obserwacji i sterowania. Właśnościom układów ze sprzężeniem poświęcono wiele monografii, np. [1 CRJ; 6LUB].

Przyjmuje się, że sprzężenie zwrotne SZ jest przede wszystkim podstawową metodą sterowania systemami niestabilnymi. Systemy stabilne – lepiej lub gorzej, mogą być sterowane w układzie otwartym. Ogólnie, SZ stosuje się zawsze w warunkach niepewności i niekompletności informacji o systemie i zakłócenach. Sprzężenie reprezentuje łańcuch przyczynowo-skutkowy, przedstawiający zależności:

$$\begin{aligned} \text{Wejście} &\rightarrow \text{System Dynamiczny} \rightarrow \text{Wyjście} \rightarrow \text{Pomiar} \rightarrow \\ &\rightarrow \text{Porównanie z Sygnałem Odniesienia} \rightarrow \text{Wejście} \end{aligned}$$

Łańcuch ten formuje pętlę sprzężenia zwrotnego (rys. 1.1), tworząc nowy zamknięty układ, który wewnątrz zawiera system początkowy. Podstawowymi zagadnieniami powstającymi są odpowiedzi dynamiczne i stabilność tego nowego układu. Koncepcja sprzężenia zwrotnego zawiera bogatą różnorodność matematycznych zagadnień takich, jak teoria stabilności Lapunowa czy metoda Nyquista na płaszczyźnie zespolonej. Komputerowe sterowanie odrzutowcami (*fly-by-wire*), rakiet samonaprowadzających i robotów kroczących jest szczytowym osiągnięciem koncepcji sprzężenia zwrotnego.



Rys. 1.1. Układ ze sprzężeniem zwrotnym

Drugi główny temat jest związany z ideą sterowania optymalnego w układzie otwartym, w oparciu o dobrze poznany model. Poszukuje się off-line sterowania ekstremalizującego wybrany wskaźnik jakości. Każdemu przebiegowi sterowania odpowiada przebieg zmiennych stanu na zadanym odcinku czasu. Wskaźnik jakości określany na funkcji stanu jest funkcją charakteryzującą różnicę między pożądanym a realizowanym stanem systemu na pewnym przedziale czasu. Jego wartość jest liczbą rzeczywistą należącą do R^1 (niekiedy tylko rozuwa się wskaźniki wektorowe, gdy poszukuje się kompromisu między

dwoema wskaźnikami skalarnymi różnego typu lub większą ich liczbą – jest to optymalizacja wielokryterialna). Rozwiążanie problemu minimalizacji funkcjonału umożliwia znalezienie i zaprogramowanie sterowania dla całego przedziału czasu i jest nazywane niekiedy optymalizacją trajektorii. Korzenie tematu należą do rachunku wariacyjnego. Bywa, że bieżące sterowanie optymalne może być uzależnione od bieżących wartości zmiennych stanu. Z powodu istnienia niemierzalnych zakłóceń, korzystne jest wtedy nieroziwiązywanie zadania optymalizacji trajektorii do końca, ale korzystając z pomiaru stanu i sprzężenia zwrotnego, w naturalny sposób stworzyć układy zamknięte.

Główną przyczyną użycia sprzężenia zwrotnego jest potrzeba osiągnięcia celu bez względu na zakłócenia zewnętrzne, zakłócenia pomiarowe i niepewność modelu. Sprzężenie redukuje te efekty, kompensując wszystkie błędy bez względu na ich pochodzenie. W zależności od typu dodatkowych urządzeń dołączonych w sprzężeniu, mówi się o sprzężeniu nieliniowym, wielowymiarowym, adaptacyjnym, intelligentnym.

Proces projektowania układu sterowania rozdziela się na szereg następujących zagadnień:

- modelowanie obiektów na podstawie praw fizycznych i analizę ich własności;
- przetwarzanie sygnałów pomiarowych We/Wy metodami filtracji, predykcji, estymacji;
- dobór i identyfikację modelu obiektu w oparciu o eksperyment;
- syntezę obserwatora stanu;
- syntezę sterowań;
- implementację sterowania w sterownikach cyfrowych lub komputerach sterujących.

Omówimy skrótnie niektóre z tych zagadnień.

MODELOWANIE

Rzeczywisty system fizyczny może być opisany wieloma różnymi zastępczymi modelami matematycznymi. Wszystko zależy od przeznaczenia budowanego modelu. Na przykład w celu wyliczenia ilości paliwa wymaganego do posadowienia satelity na orbicie, można modelować go jako cząsteczkę o danej masie, pozycji i prędkości. Do szybkiego sterowania pozycją wirującego satelity na orbicie będzie już wymagany model ujmujący jego moment bezwładności i elastyczną strukturę konstrukcji anten. Tak więc podstawowym stwierdzeniem jest, że model nie musi reprezentować generalnie systemu rzeczywistego. Najważniejsze jest, aby dla konkretnego zadania model dostarczał wiarygodnych predykcji jego rozwiązania przy rozsądnych kosztach. Z reguły nie szukamy jak najwierniejszego odbicia całego systemu w modelu, ale najpierw formułujemy dla konkretnego procesu fizycznego zadanie, dla którego dobieramy odpowiedni model tego procesu.

I tak, dla problemu sterowania, wybór właściwego modelu matematycznego danego układu ma decydujące znaczenie. Bez znajomości modelu procesu, dobór układu regulacji musiałby odbywać się metodą prób i błędów. Taka metodologia projektowania może być stosowana przy doborze termostatu żelazka, ale jest wykluczona przy budowie systemu sterowania w procesach technologicznych. Trudności w projektowaniu pojawiają się więc, gdy nie znamy systemu lub gdy dokładne prawa dynamiki i równania je opisujące mogą być

zbyt skomplikowane do wykorzystania w syntezie sterowań. Bywa, że w pewnych zastosowaniach skomplikowany system może być sterowany przez sprzężenie zwrotne przy wykorzystaniu całkiem prostego modelu. Według tego podejścia niedoskonałości modelu nie są istotnie ważne, ponieważ sprzężenie redukuje efekty niepewności, włączając w to błędy modelowania. Potrzebna jest raczej efektywna metodologia projektowania sprzężeń prowadząca do adaptacyjnego, odpornego na błędy (pewnego) systemu, czyli bardziej niezawodna metoda sterowania, a nie lepsze modele. Poszukiwanie adekwatnego, ale uproszczonego modelu wymaga jednak również pewnego wysiłku badawczego.

Inne, krańcowe podejście do tego zagadnienia, spotykane niekiedy w literaturze, kładzie nacisk na wykorzystanie dokładnych modeli otrzymanych z praw fizyki i zastosowanie prostych zasad sterowania. Wymagane moce obliczeniowe i kłopoty numeryczne w zastosowaniach on-line, oraz niekorzystanie z dorobku teorii sterowania składają się na wady takiego rozwiązania.

W rzeczywistych zastosowaniach, jeden lub drugi punkt widzenia może być mniej lub bardziej użyteczny. Wiadomo, że modele używane do projektowania konstrukcji technologicznych oraz symulacji punktów pracy i reżimów technologicznych mogą być nieprzydatne dla sterowania. Z drugiej strony zagadnienia takie, jak reżimy czasowe obliczeń w czasie rzeczywistym czy liczba i lokalizacja czujników i przetworników, nie mające znaczenia dla symulacji, mogą mieć zasadnicze znaczenie dla celów sterowania.

Powszechnie stosowany model ciągłego systemu dynamicznego o parametrach skupionych jest dany przez skończony zbiór zwyczajnych równań różniczkowych i algebraicznych:

$$\begin{aligned}\dot{x}(t) &= f_1(t, x(t), u(t)), & x(t_0) &= x_0, \\ y(t) &= f_2(t, x(t), u(t)),\end{aligned}$$

gdzie wektor stanu $x(t) \in R^n$, wektor sterowania $u(t) \in R^r$, wektor wyjścia $y(t) \in R^m$, dla $\forall t \geq t_0$, a wektor-funkcje $f_1: R \times R^n \times R^r \rightarrow R^n$, $f_2: R \times R^n \times R^r \rightarrow R^m$.

Dla układów z dyskretnym czasem jego analogiem jest zbiór równań różnicowych.

FILTRACJA

Problematyka przetwarzania sygnałów istnieje niezależnie od automatyki, ale wkład teorii sterowania do przetwarzania sygnałów jest duży, szczególnie w zagadnieniach filtracji i estymacji, tzn. w rekonstrukcji sygnałów przy obserwacjach zaburzonych szumem. Problematyka ta znajduje zastosowanie w identyfikacji, obserwacji stanu i zaawansowanych algorytmach sterowania adaptacyjnego, w zagadnieniach komunikacji, odbierania danych z satelitów oraz przetwarzania glosu i obrazu. Ogólnie filtracja dyskretnych próbek sygnału dotyczy zagadnień wyznaczania warunkowej wartości oczekiwanej \bar{y}_n na podstawie ciągu wartości $\{y_m, y_{m-1}, \dots\}$. Z reguły, jeżeli $n < m$, mówi się o wygładzaniu wartości y_n , jeżeli $n = m$ mówi się o filtracji y_n , jeżeli $n > m$ mówi się o predykcji z wyprzedzeniem $(n-m)$ -krokowym. Przykładowe zastosowanie filtracji w identyfikacji zostanie dalej omówione.

IDENTYFIKACJA

Następnym zagadnieniem, którym zajmuje się automatyka, jest identyfikacja. Identyfikację układu można zdefiniować jako proces określenia modelu układu statycznego lub dynamicznego na podstawie obserwacji jego danych wejściowych i wyjściowych. Gdy struktura modelu jest dana, a nieznane są parametry, identyfikacja sprawdza się do estymacji parametrów. Stosuje się czynne metody inżynierskie oparte na aktywnym testowaniu sygnałami standardowymi obiektu i pomiarze jego wyjścia. Znajomość analitycznych charakterystyk tego obiektu pozwala na wyliczenie jego wybranych parametrów. W systemach komputerowych działających on-line stosuje się również metody bierne, polegające tylko na obserwacji aktualnie przychodzącego sterowania i odpowiadającego mu wyjścia z obiektu. Wtedy stosuje się rekurencyjne algorytmy identyfikacji wykazujące najlepsze własności w obecności szumów (z reguły minimalno-kwadratowe) tzn. wyliczające takie parametry modelu, które minimalizują błąd wyjścia – metoda OE (*Output Error* – błąd pomiędzy wyjściem modelu i obiektu) lub minimalizujące błąd równania – metoda EE (*Equation Error* – błąd pomiędzy prawą i lewą stroną równania modelu).

W latach 60. główny wysiłek badawczy koncentrował się na opracowaniu metod identyfikacji parametrycznej dla układów dyskretnych w wersji rekurencyjnej [1EYP; 1SOT], [1MAK; 1SIN]. Miało to uzasadnienie z jednej strony w fakcie dyskretyzacji pomiarów i używania modeli dyskretnych, a z drugiej strony w możliwościach obliczeniowych komputerów, które realizowały wersje numeryczne algorytmów on-line. Identyfikacja parametrów ciągłych modeli dynamicznych wysokiego rzędu musiała więc być robiona trójetapowo:

1. zamieniano model ciągły na dyskretny,
2. wykonywano identyfikację parametrów modelu dyskretnego,
3. starano się odtworzyć odpowiadające im parametry modelu ciągłego.

Etap trzeci był trudny i zawodny, gdyż często z powodu szumów pomiarowych znalezione parametry dyskretne (pozwalające nawet na dobrą symulację ciągu czasowego dyskretnego) dawały duże różnice w odpowiadających im parametrach ciągłych lub w ogóle nie posiadały swoich rzeczywistych odpowiedników ciągłych. Wyjściem z tej sytuacji są metody identyfikacji, które mogą być stosowane bezpośrednio do modeli ciągłych. Główną metodą jest metoda funkcji modulujących (filtrujących) i zastosowanie filtrów splotowych [1MAV; 1BYW1]. Filtracja splotowa posiada cenną własność symetrii między funkcją filtrowaną i filtrującą, stąd splot nieznanej n -tej pochodnej funkcji filtrowanej i znanej funkcji filtrującej równa się splotowi znanej funkcji filtrowanej i znanej n -tej pochodnej funkcji filtrującej. Funkcję filtrowaną często znamy, bo ją mierzymy, ale nie znamy jej n -tej pochodnej, bo ze względu na szумy jest niemierzalna. Powyższa operacja umożliwia więc zamianę równania różniczkowego n -tego rzędu o nieznanych parametrach na równanie algebraiczne – funkcyjne o tych samych parametrach. Pomiary zmiennych procesowych i przetwarzanie splotowe dostarczają danych do następnego etapu. W etapie tym można użyć metody optymalnej identyfikacji do tych parametrów, które są bezpośrednio parametrami układu ciągłego [1BYW2; 1BYW3]. Dzisiejsze komputery posiadają moce obliczeniowe wystarczające do wykonywania na bieżąco operacji splatania funkcji na wystarczająco długich oknach obserwacyjnych, a więc do stosowania tych bardziej zaawansowanych całkowych metod identyfikacyjnych on-line.

OBSERWACJA STANU

W klasycznej teorii sterowania i jej zastosowaniach regulatory działające w sprzężeniu zwrotnym korzystają z pomiaru interesujących technologa procesowych zmiennych wyjściowych. Korzystniejsze efekty stabilizacji całego obiektu uzyskuje się jednak, stosując wiele petli sprzężenia, teoretycznie od wszystkich zmiennych procesowych (stanu obiektu), których liczba równa się rzędowi dynamiki obiektu. Nie wszystkie zmienne stanu są dostępne do pomiaru. W celu ich odtworzenia, wykorzystując pomiar wyjść obiektu (których jest zwykle mniej niż zmiennych stanu) i znajomość sterowania, stosuje się urządzenia lub ogólniej algorytmy do obserwacji stanu. Można je traktować jak niesprzętowe czujniki pomiarowe (softsensory).

W latach 60. R. Kalman i D. Luenberger opracowali rekurencyjne struktury obserwatorów stanu opisywane równaniem różniczkowym. Takie obserwatory dawały asymptotyczne nadążanie estymaty stanu za jego rzeczywistą wartością. Autor w swoich badaniach [1BYW4], zajmował się innymi – całkowitymi strukturami obserwatorów (rozdz. 13). Takie algorytmy odtwarzają stan dokładnie (nieasymptotycznie) na podstawie skończonego okna obserwacyjnego przesuwanego na tle pomiarów i wykonującego operację iloczynu skalarnego pomiarów i specjalnej funkcji obserwacyjnej. Problemem jest znalezienie najlepszej funkcji obserwacyjnej ze względu na zakłócenia pomiarowe. Moce obliczeniowe dzisiejszych komputerów umożliwiają na bieżąco wykonywanie operacji całkowania (iloczyny i sumy) wewnątrz skończonego i ruchomego okna obserwacyjnego. Główne rezultaty badań były zebrane w pracy [1BYW]. Od tego czasu uzyskano nowe wyniki, które również ujęto w tej monografii.

SYNTEZA REGULATORA

Syntezą sterowania zależy od poprzednich punktów jak również od celu sterowania i przyjętego zbioru ograniczeń na zmienne procesowe. Końcowym etapem projektowania jest dobór gotowego sterownika PLC i jego oprogramowanie lub konstrukcja mikrosterownika specjalizowanego. Najczęściej stosowane struktury dają: regulator PID (proporcjonalno-całkowo-różniczkujący), regulator proporcjonalno-macierzowy realizujący wielowymiarowy regulator od stanu, regulator predykcyjny, regulator przekaźnikowy (dwu- lub trójpołożeniowy) lub inny regulator nieliniowy (np. czasooptymalny). Osobny problem to prawidłowe strojenie regulatora, czyli najlepszy dobór jego nastawialnych parametrów ze względu na przyjęte kryterium jakości.

Poniżej przedstawione są bardziej szczegółowe zagadnienia, jakimi zajmuje się automatyka procesowa w działach analizy i syntezы:

- Analiza obiektów, układów i systemów sterowania obejmująca:
 - 1) analizę w dziedzinie czasu:
 - modeli matematycznych elementarnych procesów fizycznych,
 - dynamiki złożonych systemów liniowych otwartych i zamkniętych,
 - dynamiki obiektów nieliniowych,
 - stabilności obiektów i układów (typy stabilności i kryteria jej oceny),
 - sterowalności, obserwacyjności;

- 2) analizę w dziedzinie częstotliwości w tym:
 - przekształcenie Fouriera i Laplace'a,
 - opis dynamiki za pomocą transmitancji,
 - algebrę schematów blokowych,
 - sygnały harmoniczne i analizę charakterystyk częstotliwościowych.
- Synteza układów regulacji i sterowania obejmująca:
 - 1) wybór struktury systemu sterowania (otwarty, zamknięty, kaskadowy);
 - 2) wybór globalnej strategii (metody) sterowania systemem i metod sterowania w lokalnych podsystemach; rozwiązywanie zadania syntezy ma na celu wyprowadzenie równania regulatora lub podanie algorytmu jego działania (programu); przy zadanej strukturze regulatora trzeba określić jego najlepsze parametry.

Do rozwiązania zadania syntezy potrzebne są:

- opis matematyczny obiektu,
- sformułowanie zadania sterowania,
- podanie charakterystyk zakłóceń,
- wybór wskaźnika jakości sterowania,
- ograniczenia nakładane na zmienne sterowania i procesowe,
- założenia co do struktury regulatora.

Bieżące badania w dziedzinie automatyki i teorii sterowania koncentrują się w kilku kierunkach. Wymienione są one poniżej.

- Teoria sterowania odpornego (*robust*), której celem jest znalezienie sprzężenia zwrotnego gwarantującego szczególną niezależność cech systemu zamkniętego (np. stabilności) od niepewności strukturalnej modelu, od błędów pomiarowych i od zewnętrznych zakłóceń. Dla podejmowania decyzji i sterowania w warunkach dużej niepewności zakłóceń stosuje się np. podejście mini-max, tzn. przygotowuje się układ sterowania dla hipotetycznych zakłóceń najbardziej niebezpiecznych, wybierając jego parametry tak, by minimalizowały maksymalne uchyby sterowania, jakie mogłyby wystąpić w niekorzystnych warunkach. Z matematycznego punktu widzenia strategia taka sprowadza się do minimalizacji normy systemu w torze zakłócenia (rozdz. 9.8). Inne podejścia do sterowania odpornego to opisy probabilistyczne lub poprzez zbiory rozmyte (*fuzzy sets*) [1BUZ; 1KOJ1]. Problematyka procesów stochastycznych jest obecnie przedstawiona w dwóch monografiach [1PAA; 1SCF].
- Zastosowanie teorii sterowania w detekcji awarii procesów i systemów sterowania (*failure detection*), w diagnostyce błędów torów pomiarowo-kontrolnych (*fault diagnosis*) i w końcu w projektowaniu sterowania odpornego na błędy torów i awarie sprzętowych, np. czujników (*fault tolerant control*). Metody detekcji i diagnostyki z wykorzystaniem algorytmów eksperckich, obserwacji stanu czy metod statystycznych pozwalają wcześniej wykryć, a nawet przewidywać, awarie niż metody oparte na sprzężeniu zwrotnym [1KOJ; 1KOJ2; 1ISR1].
- Teoria sterowania adaptacyjnego, zajmująca się automatyczną zmianą nastaw regulatora lub doborem funkcji sterującej i reguł sterowania w trakcie procesu sterowania i poznawania modelu obiektu. Stosuje się jednoczesną identyfikację parametrów, obserwację stanu oraz metody samostrojenia regulatorów w sterowaniu *on-line*. Problematyka ta od lat 60. była szczególnym obiektem badań i tematem szeregu monografii [1ASK; 1CHV; 1ISR; 1KRM; 1LAY; 1MIE; 1NIA; 1SAA; 1UZH; 1WEP].

- Metody sterowania predykcyjnego. W ciągu ostatnich 20 lat rozwinięta została metoda sterowania predykcyjnego. Algorytmy sterowania systemami w obecności zakłóceń i niepewności modelu wykorzystujące koncepcję przesuwanego horyzontu sterowania i predykcji kształtu zakłóceń, dają dobre rezultaty zarówno dla systemów liniowych, jak i nieliniowych. Główne publikacje można znaleźć w [1CAE; 1CLD; 1DEH; 1DUJ].
- Sterowanie systemami nieliniowymi, o parametrach rozłożonych i z opóźnieniami.
- Sterowanie systemami wielowymiarowymi ze sprzężeniami skośnymi.
- Nowe metody identyfikacji parametrów.
- Nowe metody obserwacji stanu.
- Sterowanie z wykorzystaniem sieci neuronowych [1TAR; 1WAK].
- Teoretyczne i techniczne aspekty sterowania cyfrowego i modele dyskretnie.

Pole działania współczesnej teorii sterowania jest niezwykle obszerne. Obejmuje ono nowoczesne komputerowe systemy automatyki przemysłowej, procesy biotechnologiczne i biomedyczne, procesy diagnostyczne, identyfikacyjne, adaptacyjne i samouczące, robotykę, przetwarzanie sygnałów pomiarowych i wzajemnych, systemy transportowe i komunikacyjne – naziemne, wodne oraz powietrzne, procesy stochastyczne i inne.

Oszczerny zestaw szczegółowych problemów współczesnej automatyki charakteryzuje oficjalna lista słów kluczowych (*keywords*), jaka zwykle obowiązuje przy okazji większych konferencji naukowych.

Przykładowy alfabetyczny zestaw takich słów sugerowany przez organizatorów jednej z Europejskich Konferencji Automatyki ECC (European Control Conference), przedstawiony jest poniżej.

Automation

*Factory Modelling and Simulation,
Instrumentation Systems,
Man-machine Interactions,
Manufacturing Systems,
Process Automation.*

Biotechnological Systems

*Agricultural Processes,
Bio-technical and Biomedical Processes,
Environmental and Safety Systems.*

Complex Systems

*Bifurcation and Chaos,
Decentralized Systems,
Distributed Systems,
Infinite Dimensional Systems,
Large Scale Systems,
Power Systems and Power Plants,
Vehicles and Transportation Systems.*

Computer Control

*Algorithms for Real-time Control,
Fault-Tolerant Control,
Hardware Architectures for Control,
Parallel and Distributed Control Systems,
Software Design Methodologies.*

Control Education

*Continuing Education,
Control Education on the Web,
Control Education in Industry.*

Control Applications

*Control of Structures and Space Stations,
Active Noise Cancellation,
Process Control,
Industrial Processes,
Marine/offshore Systems,
Micro Systems,
Oceanic Engineering.*

Design Methodologies

Adaptive Control,
Control and Optimization,
Dissipative and Energy-based Design,
Fuzzy Systems,
Learning Systems.

Discrete Event and Hybrid Systems

Applications,
Design Tools, Petri Nets,
Modelling, Control and Verification,
Scheduling,
Supervisory Control.

Fault Diagnosis and Supervision

Methods of Fault Diagnosis,
Statistical Methods of Fault Diagnosis,
Fault and Uncertainty Modelling,
Fault Detection and Diagnosis in Bilinear and Non-linear Systems,
Process Supervision.

Linear Systems

PID Control,
Sampled Data Systems,
Singular Systems,
Stabilization,
Stochastic Systems,
Time Delay Systems,
Time Varying and Periodic Systems.

Non-linear Systems

Bilinear Systems,
Control of Non-linear Discrete Systems,
Controllability and Observability of Non-linear Systems,
Feedback Linearization of Non-linear Systems,
Flatness,
Geometry and Structure of Systems,
Non-linear H-Infinity,
Non-linear Observers,

Non-linear Oscillations and Chaos,
Stability Analysis of Non-linear Systems,
Stabilization of Non-linear Systems,
Stochastic Systems,
Trajectory Tracking in Non-linear Systems.

Robotics and Mechatronics

Flexible Robots,
Learning Robot Control,
Manipulators, Mechatronic Systems,
Micro-robotics,
Mobile Robots,
Redundant Manipulators,
Robot Adaptive Control,
Robot Applications,
Robot Force and Motion Control,
Task and Motion Execution Control,
Tele-robotics, Vision Systems
Undersea Robotics.

System Identification and Signal Processing

Bounded Uncertainty and Errors,
Estimation,
Frequency Domain Identification Methods,
Identification for Control,
Identification Methods,
Identification of Non-linear Systems,
Observers,
Robust Filtering,
Set-membership Estimation and Identification,
Signal Processing and Systems,
Subspace Methods.

Vehicles and Transportation Systems

Aerospace Systems,
Automotive Systems,
Autonomous Vehicles,
Hybrid and Electric Vehicles,
Marine Systems, Rail Systems,
Traffic Control Automation,
Transportation Systems.

Powyższe słowa kluczowe pokazują na obszerność tematyki wchodzącej w zakres automatyki. Ze względu na tę obszerność, poszczególne ośrodki naukowo-badawcze koncentrują się z reguły na kilku wybranych dziedzinach, w których opracowują i testują nowe metodologie. Akademickie programy nauczania na kierunku automatyka i robotyka, muszą jednak obejmować wszystkie przedmioty, które dadzą wystarczającą wstępna bazę do

ewentualnych pogłębionych studiów na dowolnej wybranej specjalności, wynikającej z przedstawionego zestawu.

Dla zaprojektowania, zbadania i uruchomienia konkretnych instalacji przemysłowych, w których systemy automatyki odgrywają decydującą rolę, potrzebne są duże zespoły inżynierskie obejmujące różnych specjalistów i technologów z danej dziedziny. Obecność inżynierów automatyków jest w nich obowiązkowa.

Historia powstawania i rozwoju automatyki jest przedstawiona między innymi w dwóch książkach [1BES1; 1BES2], i w szeregu artykułów [1BES3; 1BES4; 1BED; 1LAC]. Zagadniemu temu poświęcono też dwa numery czasopisma „IEEE Control Systems Magazine”, Vol. 15, No 3, 1995 i Vol. 16, No 3, 1996.

Na zakończenie tego rozdziału podamy chronologiczną listę najważniejszych dat, nazwisk i wydarzeń składających się na historyczny rozwój automatyki. Autor uzupełnił ją o nazwiska matematyków, których prace zostały wykorzystane w teorii sterowania.

260 p.n.e.	Ktesibios z Aleksandrii	Zegar wodny ze stabilizacją poziomu cieczy
100	Heron z Aleksandrii	Pierwsze pomysły i urządzenia automatyczne
860	Heron z Bizancjum	Automaty w pałacowej sali tronowej
1620	C. Drebbel	Regulacja temperatury w piecu hutniczym i inkubatorze
1745	E. Lee	Samoczyny układ ustalania młyna wiatrakowego pod wiatr
1755	L. Euler	Dwutomowa praca nt. rachunku różniczkowego i całkowego
1758	J. Bindley	Patent na płynakowy regulator poziomu cieczy
1787	M. Boulton, T. Mead	Ustawianie odległości kamieni młynskich regulatorem odśrodkowym
1788	J. Watt	Nowy regulator odśrodkowy do regulacji prędkości turbiny parowej
1807	J. Fourier	Szeregi i przekształcenie całkowe (opublikowane w 1822)
1809	K. Gauss	Metoda najmniejszych kwadratów
1812	P. Laplace	Przekształcenie całkowe Laplace'a
1868	J.C. Maxwell	Pierwsza praca o stabilności układu z regulatorem Watta
1877	J. Wyszniewicz	Stabilność systemu regulacji na płaszczyźnie parametrów
1877	E.J. Routh	Rozwiążany problem stabilności układów liniowych

1892	O. Heaviside	Rachunek operatorowy
1892	A. Lapunow	Stabilność rozwiązań równań różniczkowych nieliniowych
1895	A. Hurwitz	Kryteria algebraiczne stabilności układów liniowych
1910	E. Sperry	Wynalezienie żyroskopu dla stabilizacji kierunku
1912	E. Armstrong	Użycie dodatniego sprzężenia zwrotnego w układzie triody dla budowy oscylatora harmonicznego i superheterodyn
1922	N. Minorski	Pierwsza koncepcja regulatora PID dla sterowania statkiem
1932	H. Nyquist	Kryteria stabilności układów zamkniętych
1932	Foxboro Instr. Comp.	Pierwsze przemysłowe regulatory PID „Stabilog”
1940	H. Bode	Metody częstotliwościowe w analizie układów
1942	J. Ziegler-N. Nichols	Optymalizacja parametryczna regulatorów PID
1942	N. Wiener	Teoria filtracji optymalnej
1948	J. Mauchly, J. Eckert	Elektroniczny komputer ENIAC (I generacja – lampy)
1948	N. Wiener	Podstawy cybernetyki
1956	L. Pontriagin	Matematyczna teoria sterowania optymalnego. Zasada Maksimum
1957	R. Belman	Programowanie dynamiczne
1958	F. Rosenblatt	Sieci neuronowe – Perceptron
1959	W.M. Popow	Stabilność absolutna układów zamkniętych nielinowych
1960	R. Kalman	Ogólna teoria liniowych systemów sterowania
1965	L. Zadeh	Zbiory rozmyte
1973	K. Åström	Regulatory samostrojące
1982	J. Hopfield-T. Kohonen	Sieci neuronowe ze sprzężeniem zwrotnym i samouczące

2. Modelowanie procesów dla celów sterowania

2.1. Wstęp

Matematyczne modelowanie dynamiki procesów fizycznych ma długą historię. Siega ona teoretycznie końca wieku XVII, kiedy to w pracach Newtona, Leibniza i Bernoullich pojawia się pojęcie różniczki. Całościowe, systematyczne i twórcze ujęcie zagadnienia rachunku różniczkowego przypisuje się jednak dopiero L. Eulerowi, który na ten temat opublikował w 1755 r. dwutomową pracę [2JUA].

Rozwój metod modelowania w wieku XX zawiązuje się przede wszystkim rozwojowi elektroniki i powstaniu maszyn analogowych. Za ich pomocą przez wiele lat rozwiązywane były zadania symulacji dynamiki systemów liniowych i nieliniowych. Modelowanie polegało na fizycznym konfigurowaniu zestawu modułów wchodzących w skład maszyny (głównie analogowych wzmacniaczy operacyjnych i członów całkujących) w cały system modelujący równanie n -tego rzędu. Dobór współczynników pozwalał na skalowanie czasu, czyli modelowanie rozwiązań w skali czasowej dostosowanej do potrzeb projektanta, zwykle znacznie krótszej niż przebieg modelowanego procesu rzeczywistego. Sam proces rozwiązywania przebiegał z szybkością równą 0,8 prędkości światła (szybkość prądu w miedzi). Początek ciągłego rozwiązywania na wyjściu każdego integratora pojawiał się na ekranie oscyloskopu w czasie rzeczywistym natychmiast po starcie i był niezależny od typu maszyny.

Jednoprocessorowe komputery cyfrowe, w których wszystkie operacje przebiegają sekwencyjnie, zostały wprowadzone do powszechniej eksploatacji w latach 60. XX w. Wykorzystując inną technikę modelowania dynamiki niż w maszynach analogowych, opartą na numerycznych, rekurencyjnych algorytmach rozwiązywania równań różniczkowych, wypisały one sukcesywnie maszyny analogowe zdobywając nad nimi przewagę głównie w uniwersalności, łatwości i niezawodności przy oprogramowaniu problemów, przy jednociennym praktycznym braku ograniczenia na wielkość tego problemu (rzędu równania). Jednak przez następne 25 lat, szybkość uzyskiwania rozwiązania, głównie z powodu stosowania algorytmów rekurencyjnych, i możliwości sprzętowych nie była ich głównym atutem w stosunku do maszyn analogowych. Dopiero podniesienie skali integracji, wprowadzenie procesorów 32-bitowych i podniesienie częstotliwości zegara pozwoliło na osiągnięcie szybkości rozwiązywania spełniającej wymogi projektanta i umożliwiającej zastosowanie komputerów w sterowaniu on-line. Ogólnie czas rozwiązywania zestawu iteracji zależał od szybkości zegara i typu komputera i nie miał bezpośredniego przełożenia na czas rzeczywisty.

Ostatecznie duża szybkość obliczeń pozwoliła na wprowadzenie komunikacyjnych interfejsów graficznych użytkownika, co w konsekwencji umożliwiło powrót do wygodnych algorytmów programowania (modelowania) znanych z techniki analogowej. Programowanie konfiguracji modułów (integratorów i wzmacniaczy) odbywa się obecnie graficznie na ekranie. Na tej zasadzie działa np. pakiet do wspomagania projektowania systemów sterowania typu CADCS (*Computer Aided Design of Control Systems*) – Matlab/Simulink, w którym z bogatej, graficznej biblioteki modułów reprezentujących bloki statyczne i dynamiczne oraz liniowe i nieliniowe wybiera się właściwe i łączy w cały system. Po starcie symulacji następuje programowa interpretacja bloków i ich połączeń, a następnie poprzez numeryczne algorytmy (z reguły oparte na schemacie Rungego–Kutty) jest rozwiązywany rekurencyjnie zadany układ równań różniczkowych, a wyniki wyświetlane są na ekranie. Najnowsze wersje tego pakietu pracują już w czasie rzeczywistym.

W automatyce przez model będziemy rozumieli matematyczną abstrakcję o znanych własnościach, odwzorowującą nie do końca poznany proces rzeczywisty. Zbiór równań reprezentujący model daje więc w najlepszym przypadku tylko aproksymację rzeczywistości. Model nie może zawierać wszystkich makroskopowych i mikroskopowych cech procesu. Stąd, przy określaniu modelu inżynier powinien brać pod uwagę koszty jego budowy, czyli czas i wysiłek poświęcony na jego otrzymanie i weryfikację. Wiąże się to z poziomem szczegółowości modelu (np. praw fizykochemicznych) i oczekiwanych stąd zysków.

Można stwierdzić, że przy tworzeniu właściwego modelu należy wykazać się nie tylko wiedzą i umiejętnościami, ale również twórczą kreatywnością. Wiedza potrzebna jest do uwzględnienia wszystkich związków przyczynowych i stworzenia zbioru równań, które gwarantują istnienie jednoznacznego rozwiązania, co wiąże się z zerową liczbą stopni swobody modelu, tzn. z warunkiem, aby liczba niezależnych równań była równa liczbie niezależnych zmiennych procesowych. Natomiast twórcze aspekty budowy modelu są związane z umiejętnością przyjęcia zbioru założeń upraszczających takich, które z jednej strony nie pozbawią modelu najważniejszych cech dynamicznych, a z drugiej zachowana zostanie rozsądna wymiarowość zadania.

W automatyce prawidłowo zbudowane i zidentyfikowane modele obiektów sterowania służą do:

- wyboru właściwego regulatora i jego nastaw bądź przez analizę teoretyczną dynamiki układu zamkniętego bądź przez symulację komputerową;
- syntez nowoczesnych algorytmów sterowania i obserwacji, które w swoich regułach zawierają model procesu, np. sterowanie liniowo-kwadratowe, predykcyjne, filtracja Kalmana;
- statycznej optymalizacji punktu pracy przy wykorzystaniu modeli stanu ustalonego uwzględniające koszt surowców i energii, ceny produktów oraz wymagania technologiczne i rynkowe; znalezionej najlepszy statyczny punkt pracy instalacji jest ustawiany i nadzorowany przez system sterowania nadzawanego, który przesyła wartości zadane dla regulatorów lokalnych;
- zwiększenia wiedzy na temat procesu, bez potrzeby eksperymentów na obiekcie rzeczywistym;

2.1. Wstęp

- wyboru i testowania różnych algorytmów sterowania procesem, co wiąże się z testowaniem różnych lokalizacji czujników, czyli wyborem pomiarowych zmiennych procesowych, a w razie konieczności – projektowaniem czujników wirtualnych (softsensorytów);
- bezpiecznego i taniego szkolenia operatorów procesu poprzez symulację zachowania instalacji w wybranych punktach pracy oraz w sytuacjach nietypowych i awaryjnych.

Ogólnie do zagadnienia modelowania danego procesu można podejść trzema metodami:

- 1) przez budowę matematycznego modelu teoretycznego wykorzystującego tylko prawa fizykochemii:
 - zasadę zachowania pędu (mechanika),
 - zasadę zachowania masy i ładunku (technologia chemiczna, ekonomia, elektrotechnika),
 - zasadę zachowania energii (technologie grzewcze i chemiczne);
- 2) przez budowę modelu empirycznego, wybierając z danej klasy funkcji matematycznych funkcję opisującą dane zjawisko w sposób najlepszy, w sensie zgodności z zebrańymi w eksperymencie danymi wejściowymi i wyjściowymi; używa się do tego celu metod regresji liniowej i nieliniowej, oraz optymalnych metod programowania nielinowego; tak uzyskana funkcja aproksymacyjna nie musi posiadać interpretacji fizycznej i obowiązuje ona w ograniczonym zakresie zmienności wejść;
- 3) przez budowę modeli półempirycznych (będących kompromisem pomiędzy 1) i 2)), w których pewne parametry, jak np. współczynniki sprawności czy współczynniki przenikania cieplnego, są określone na podstawie danych pomiarowych.

Generalnie w automatyce rozróżniamy następujące typy modeli procesów fizycznych:

- statyczne – dynamiczne,
- liniowe – nieliniowe,
- o parametrach skupionych – o parametrach rozłożonych,
- stacjonarne – niestacjonarne,
- ciągłe – dyskretnie,
- deterministyczne – losowe (stochastyczne).
- chaotyczne.

Do opisu tych modeli używa się odpowiednio:

- równań algebraicznych – równań różniczkowych,
- równań liniowych – równań nieliniowych,
- równań różniczkowych zwyczajnych – równań różniczkowych cząstkowych,
- równań o parametrach stałych w czasie – równań o parametrach zmiennych w czasie,
- równań różniczkowych – równań różnicowych,
- równań deterministycznych – równań z teorii procesów stochastycznych.
- równań różniczkowych nieliniowych.

2. Modelowanie procesów dla celów sterowania

Najprostsze modele dynamiki opisywane są przez liniowe, stacjonarne równania różniczkowe zwyczajne, a najtrudniejsze – przez nieliniowe, niestacjonarne równania czasowe.

Najważniejszy koncepcyjny podział modeli, to podział na modele statyczne i dynamiczne. Modele statyczne są tworzone głównie dla celów poszukiwania optymalnego punktu pracy instalacji technologicznej. Dla instalacji złożonej z wielu nieliniowych modeli połączonych szeregowo, równolegle i poprzez sprzężenia zwrotne, można wyszczególnić zmienne będące zmiennymi sterującymi i zmienne procesowe będące zmiennymi wyjściowymi. Najczęstszy pożądanym reżimem pracy instalacji – to praca w stanie ustalonym. Ustabilizowane wartości sterowań u i wyjścia określają tzw. punkt pracy instalacji, gwarantujący przede wszystkim spełnienie wszystkich ograniczeń technicznych i technologicznych narzuconych na proces produkcji, jak również ograniczeń na jakość produktu narzuconych przez normy produkcji (wymagania klienta). Z reguły spełnienie wszystkich tych ograniczeń jest zagwarantowane w wielu różnych punktach pracy. Jednak prowadzenie produkcji w każdym z tych punktów różni się kosztami produkcji, o których można wnioskować na podstawie cen surowców i energii dostarczonej oraz cen otrzymanych produktów, rozważanych w przyjętej jednostce czasu.

Celem nowoczesnych kompleksowych systemów sterowania procesami ciągłymi jest określenie ustalonego (statycznego) punktu pracy całej instalacji, w którym koszty produkcji będą najwyższe, a zysk największy. Poszukiwanie tego punktu sprowadza się do poszukiwania r -wymiarowego wektora liczb rzeczywistych reprezentującego r niezależnych zmiennych sterujących decydujących o położeniu punktu $u \in R^r$. W trakcie przeszukiwania wektorowej przestrzeni R^r , w ramach narzuconych ograniczeń, testuje się symulacyjnie, opierając się na zestawie modeli statycznych, jakość i koszty każdego znalezionej punktu pracy. W końcu wybiera się punkt najlepszy, czyli optymalny. Do poszukiwań stosuje się numeryczne procedury programowania liniowego lub gradientowe i bezgradientowe metody nieliniowej optymalizacji statycznej. Znaleziony optymalny wektor liczbowy $u = [u_1, u_2, \dots, u_r]^T$ reprezentuje sterowania optymalne, np. przepływy. Wartości takich przepływów nie przesyła się jednak bezpośrednio na zawory, ale realizuje się je z reguły w systemie sterowania nadzawanego, jako wartości zadawane dla regulatorów pracujących w lokalnych pętlach regulacji o szybkiej dynamice, pomijalnej w stosunku do dynamiki procesu głównego. Lokalne regulatorы, eliminując wpływ zakłóceń, stabilizują zmienne procesowe (np. przepływy) na żądanych wielkościach i funkcjonalnie w odniesieniu do nich mogą być traktowane jako beziniceryjne inteligentne zawory [2DUJ; 2TAP].

W odróżnieniu od modeli statycznych reprezentujących zachowanie się systemu w stanach ustalonych, modele dynamiczne odwzorowują zachowanie się systemu w stanach przejściowych.

Stany przejściowe wywołane są:

- planowymi zmianami punktów pracy instalacji przez personel operatorski lub komputer sterujący,
- nieprzewidzianymi zakłóceniami oddziaływanymi na zmienne sterujące, zmienne wyjściowe i parametry procesu.

2.1. Wstęp

Zadaniem układów regulacji ze sprzężeniem zwrotnym jest jak najlepsze nadążanie wyjściowych zmiennych procesowych za zmianami planowymi i jak najlepsza likwidacja zakłóceń powodujących niepożądane zmiany zmiennych procesowych. W obu przypadkach regulator ma za zadanie likwidację dynamicznego błędu regulacji. Różne regulatory wykonują to zadanie z różną efektywnością. Można sprecyzować wskaźnik jakości regulacji i poszukiwać regulatora, który przeprowadzi likwidację błędu, minimalizując ten wskaźnik jakości. O takim regulatorze mówi się, że jest optymalnym. Dobór regulatora optymalnego przeprowadza się z reguły, opierając się na modelu dynamicznym obiektu sterowanego. Stąd wynika potrzeba określenia modeli dynamicznych tych obiektów i ich ważna rola w procesie syntezy regulatora.

Dodatkowych kilka uwag należy poświęcić modelom z czasem dyskretnym. Nabraly one znaczenia wraz z rozwojem komputerowych technik pomiarowo sterujących. Modele wykorzystujące równania różnicowe stosuje się do opisu dynamiki dwóch niżej wymienionych klas procesów:

- 1) Procesów, o których zakłada się, że mają tylko rejestrówalne wyjścia i nie mają sterowanych wejść. Są to więc źródła sygnałów dyskretnych generujące ciągi czasowe. Przykładem takich procesów mogą być losowo zmieniające się ceny akcji giełdowych, czy zjawiska meteorologiczne. Zwykle przyjmuje się, że wejściem do takiego procesu jest szum biały o zerowej wartości oczekiwanej i zadanej wariancji.
- 2) Procesów ciągłych będących obiektami sterowania, których wyjścia możemy zmieniać tylko poprzez okresową zmianę sygnału sterującego na wyspecyfikowanym wejściu.

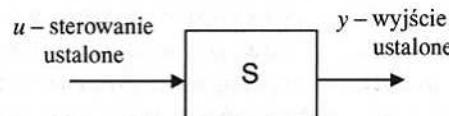
Jak wynika z listy słów kluczowych, w automatyce występuje również specjalny dział systemów sterowania – sterowanie procesami sekwencyjnymi zwany również procesami produkcyjnymi dyskretnymi. Nazwa wynika z rozproszenia gniazd produkcyjnych, w których występuje sekwencja zdarzeń dyskretnych w czasie (*Discrete Event*). Dominuje w nich typ sterowania dwupołożeniowego (sygnały dwustanowe) załączaniem i wyłącza niem zasilania w zależności od zaprogramowanego cyku i osiągnięcia progowych wartości przez wybrane zmienne procesowe. Główną teorią wykorzystywaną w tym dziale jest logika dwuwartościowa Boole'a i szereg metod projektowania sekwencyjnych układów logicznych sterowania. Podstawowymi typami stosowanych sterowników są programowalne sterowniki logiczne PLC (*Programmable Logical Controllers*). W ostatnim czasie technologia projektowania bardziej złożonych logik jest wspierana komputerowymi programami optymalizującymi (minimalizującymi) funkcje logiczne i tworzącymi automatycznie projekt elektroniczny procesora logicznego realizujące te funkcje. W dzisiejszych rozwiązaniach projektowych cyfrowych układów VLSI (*Very Large Scale Integration*) stosuje się do tego celu specjalny język opisu sprzętu (standard powstał w 1987 r.) zwany HDL (*Hardware Description Language*). Program napisany w tym języku i skompilowany umożliwia symulację działania projektowanego układu [2WRW]. Inne oprogramowanie we współpracy ze specjalizowanym sprzętem na podstawie takiego programu pozwala wygenerować sprzętową strukturę w uniwersalnej matrycy FPGA (*Field Programmable Gate Array*) lub CPLD (*Complex Programmable Logic Device*), tworząc np. gotowy procesor (rozwiązania firm Xilinx, Altera, Atmel). W rozwiązaniach tych można nawet obecnie zmieniać ich strukturę programowo i dynamicznie, a nawet przemieszczać całą strukturę na powierzchni FPGA.

Specjalizowane sterowniki logiczne znajdują zastosowanie w wielu układach, od prawie automatycznych do wind osobowych, w których logika sterownika musi zagwarantować np. minimalny czas oczekiwania pasażerów na piętrach. Taśmowe linie produkcyjne wymagają już obszerniejszego oprogramowania realizowanego za pomocą komputerów.

2.2. Modele stanu ustalonego obiektów liniowych i nieliniowych

Uznając jakiś system fizyczny za statyczny (pozbawiony dynamiki), przyjmujemy, że wielkość wyjściowa $y(t_1)$ dla każdej chwili t_1 zależy tylko od wejścia $u(t_1)$ i nie zależy od wartości $u(t)$ dla $t \neq t_1$. Typowym przykładem jest rezistor, w którym zakłada się, że przy skokowym źródle napięcia wystąpi skokowe zbroce prądowe. W rzeczywistości idealne systemy statyczne nie istnieją, gdyż przy ograniczonych mocach źródeł każda zmiana stanu energetycznego wymaga czasu różnego od zera. Niemniej modele procesów, dla których można z zadanym błędem pominąć istnienie ich własnej dynamiki (pomijalne stany przejściowe), nazywa się modelami statycznymi.

Modele statyczne (rys. 2.1) można też przypisać wielu procesom i obiektom automatyki, mimo ich istniejącej dynamiki. Mogą być one rozważane jako charakterystyki ich stanów ustalonych, czyli zależności wartości ustalonych wyjść od wartości ustalonej wejścia dla czasu $t \rightarrow \infty$ (jeśli takie ustalone wartości istnieją).



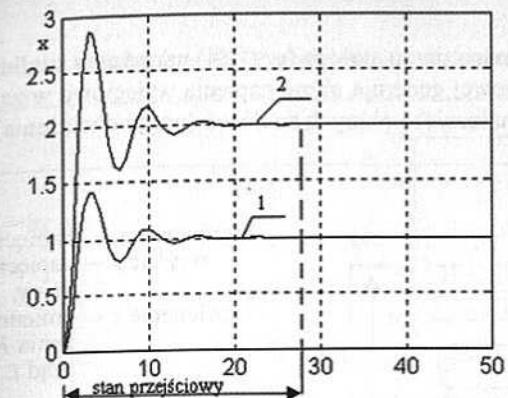
Rys. 2.1. Model statyczny

Wartości te można uzyskać z równań różniczkowych zwyczajnych, opisujących dynamikę procesu, przez przyrównanie do zera wszystkich pochodnych zmiennych procesowych po czasie, otrzymując tylko zależności algebraiczne. W obiektach o parametrach rozłożonych opisywanych równaniami częstotliwościowymi pozostały oczywiście dalej pochodne po zmiennych przestrzennych i formalnie równanie stanu ustalonego, choć zredukowane, pozostało równaniem różniczkowym.

Na rysunku 2.2 widać przykładowe przebiegi przejściowe i ustalone sygnału wyjściowego dla obiektu liniowego danego równaniem

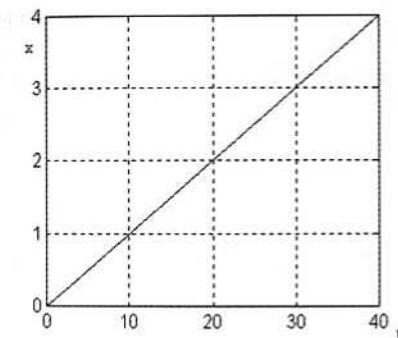
$$\ddot{x}(t) + 0,1\dot{x}(t) + x(t) = 0,1u(t)$$

i dla sterowań $u_1(t) = 10 \cdot 1(t)$ i $u_2(t) = 20 \cdot 1(t)$.



Rys. 2.2. Odpowiedzi modelu II rzędu dla sterowań: 1 – $u_1(t) = 10 \cdot 1(t)$ i 2 – $u_2(t) = 20 \cdot 1(t)$

Model statyczny tego procesu jest liniowy (rys. 2.3) i ma postać $x = 0,1u$.



Rys. 2.3. Charakterystyka statyczna modelu II rzędu

Przykładem obiektu, dla którego nie istnieją stany ustalone $x(\infty)$ dla skokowego sterowania, są np. obiekty całkujące

$$\ddot{x}(t) + 0,1\dot{x}(t) = 0,1u(t).$$

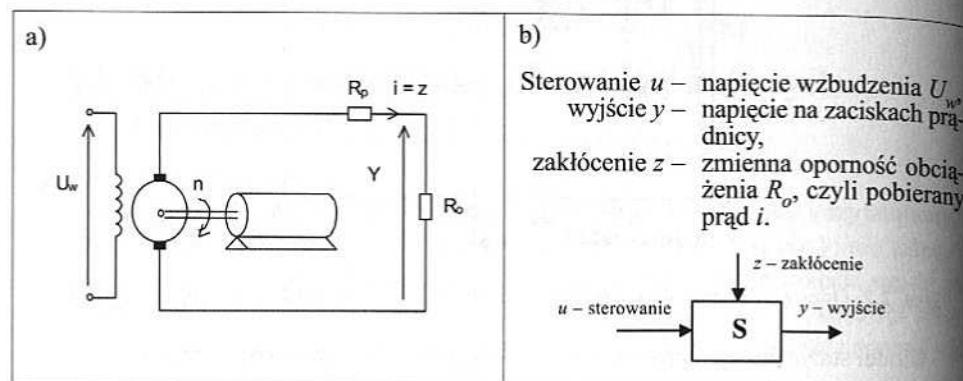
Dla nieliniowych dynamicznych obiektów sterowania opisywanych równaniem różniczkowym nieliniowym, model stanu ustalonego może być:

- liniowy, np. dla $[\ddot{x}(t)]^3 + 0,1\dot{x}(t) + x(t) = 10 \cdot u(t) \Rightarrow x = 10 \cdot u$,
- nieliniowy, np. dla $[\ddot{x}(t)]^3 + 0,1\dot{x}(t) + [x(t)]^2 = 10 \cdot u(t) \Rightarrow x^2 = 10 \cdot u$.

Wiele jest obiektów mniej lub bardziej złożonych, których charakterystyki statyczne pełnią ważną rolę w układach automatyki. Część z tych obiektów ma również dynamikę, która może być uwzględniona w ich modelu dynamicznym (np. generator), dla innych dynamika może być pominięta (np. w przekaźnikach).

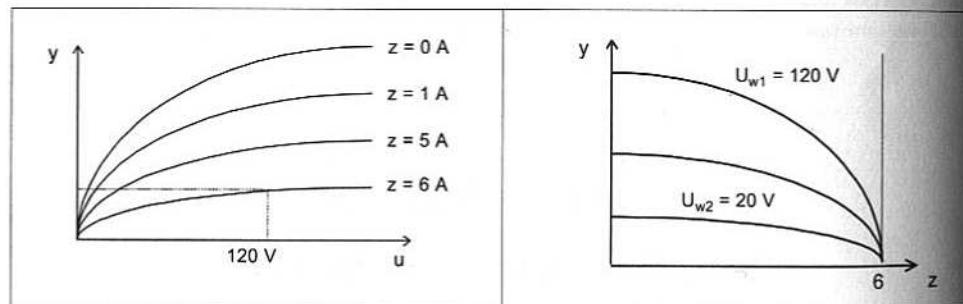
Przykład 2.1

Prądnica obcowzbudna prądu stałego (rys. 2.4) napędzana silnikiem synchronicznym o stałej prędkości obrotowej generuje różne napięcia wyjściowe w zależności od różnych sterowań (napięcie wzbudzenia) i różnych zakłóceń (prady obciążenia).



Rys. 2.4. Techniczny (a) i blokowy (b) schemat prądnicy

Na rysunkach 2.5 i 2.6 pokazano charakterystyki statyczne pradnicy.



Rys. 2.5. Charakterystyki statyczne toru sterowania przy różnych obciążeniach

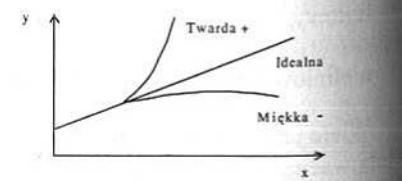
Przykład 2.2

Sprzęzyna o charakterystyce nieliniowej twardej i miękkiej (rys. 2.7).

$$y = a + b \cdot x \pm c \cdot x^3,$$

gdzie:

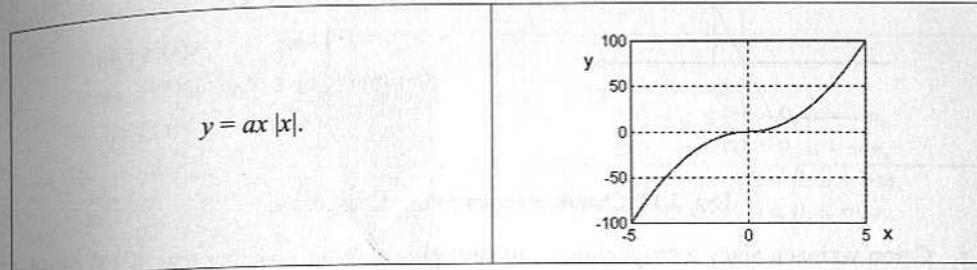
y – siła sprężystości
 x – przesunięcie.



Rys. 2.7. Charakterystyki statyczne

Przykład 2.3

Charakterystyka kwadratowa (rys. 2.8) ze zmiennym znakiem (wentylatorowa)

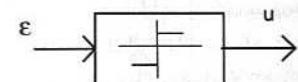


Rys. 2.8. Charakterystyka kwadratowa

Przykład 2.4

Na rysunkach 2.9–2.16 przedstawiono wybrane charakterystyki przekaźników.

Sygnalem wejściowym jest sygnał błędu regulacji ϵ , sygnałem wyjściowym z prze-kaźnika (rys. 2.9) jest sygnał sterujący dla obiektu u .

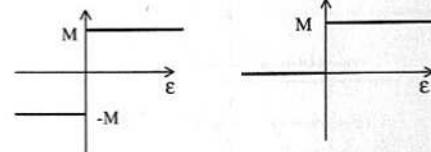


Rys. 2.9. Blok przekaźnika

- Przekaźnik idealny dwupołożeniowy (rys. 2.10)

Przykładowy opis charakterystyki prze-kaźnika:

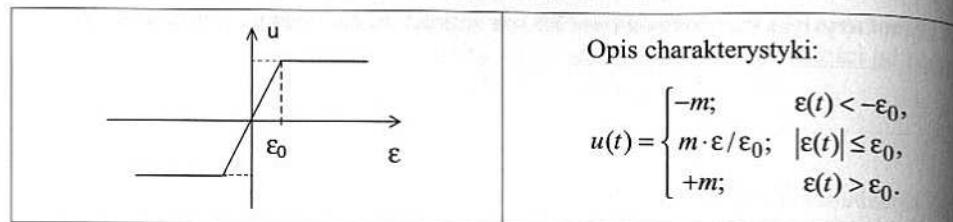
$$u(t) = \begin{cases} -M & \text{lub } 0 \quad \varepsilon(t) < 0, \\ 0; & \varepsilon(t) = 0, \\ +M; & \varepsilon(t) > 0. \end{cases}$$



Rys. 2.10. Charakterystyki przekaźników dwupołożeniowych

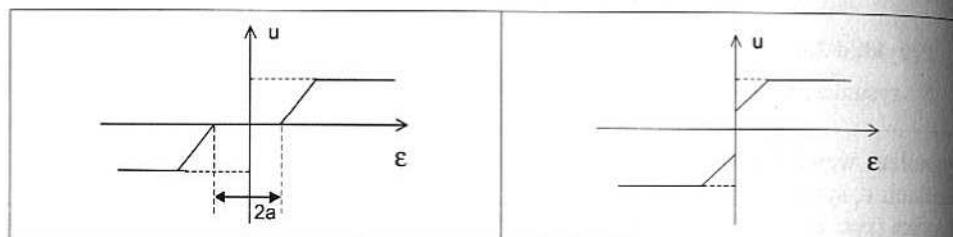
- Przekaźnik trójpołożeniowy ze strefą nieczułości (martwa) (rys. 2.11)

- Człon wzmacniający z nasyceniem (rys. 2.12)



Rys. 2.12. Charakterystyka członu z nasyceniem

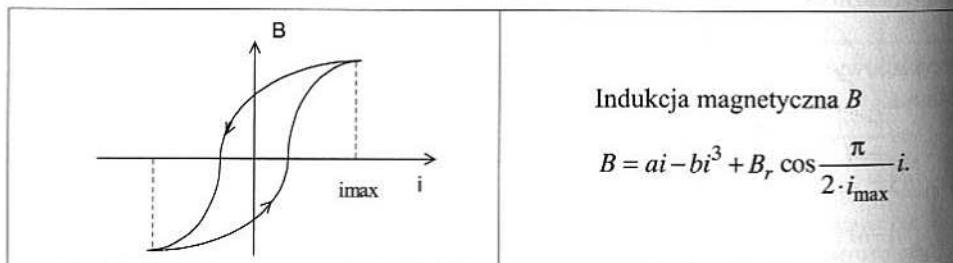
- Człon wzmacniający z nasyceniem i ze strefą nieczułości dodatnią (rys. 2.13) (model tarcia suchego), ze strefą ujemną – wstępne napięcie sprężyny (rys. 2.14).



Rys. 2.13. Charakterystyka ze strefą nieczułości

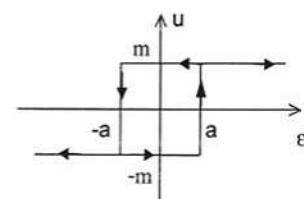
Rys. 2.14. Charakterystyka z ujemną strefą nieczułości

- Charakterystyka z histerezą magnetyczną (strefą niejednoznaczności) (rys. 2.15).



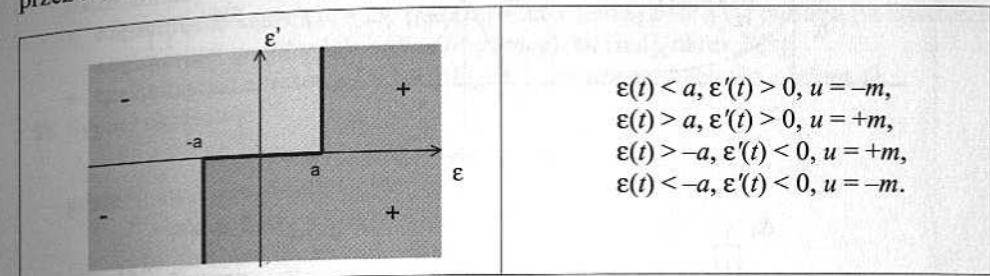
Rys. 2.15. Histerea magnetyczna

- Przekaźnik dwupołożeniowy z histerezą (rys. 2.16).



Rys. 2.16. Histerea przekaźnikowa

Krzywa przełączania tego przekaźnika na płaszczyźnie fazowej $\varepsilon(t)$, $d\varepsilon/dt$ separuje obszary, w których wyjściem przekaźnika jest sygnał dodatni i ujemny. Można ją uzyskać przez analizę znaków sygnałów przedstawioną na rysunku 2.17.



Rys. 2.17. Obszary przełączania przekaźnika

2.3. Modele procesów dynamicznych

Przed przystąpieniem do opisu modeli procesów dynamicznych zostaną modele najprostszych standardowych sygnałów sterujących. Z sygnałami tymi na nieskończonym odcinku czasu związana bywa nieskończona amplituda lub energia, są więc one modelami abstrakcyjnymi w odróżnieniu od sygnałów rzeczywistych o skończonej energii.

2.3.1. Modele sygnałów

Poniżej pokazano 10 modeli sygnałów.

- Skok jednostkowy $1(t)$ (rys. 2.18).



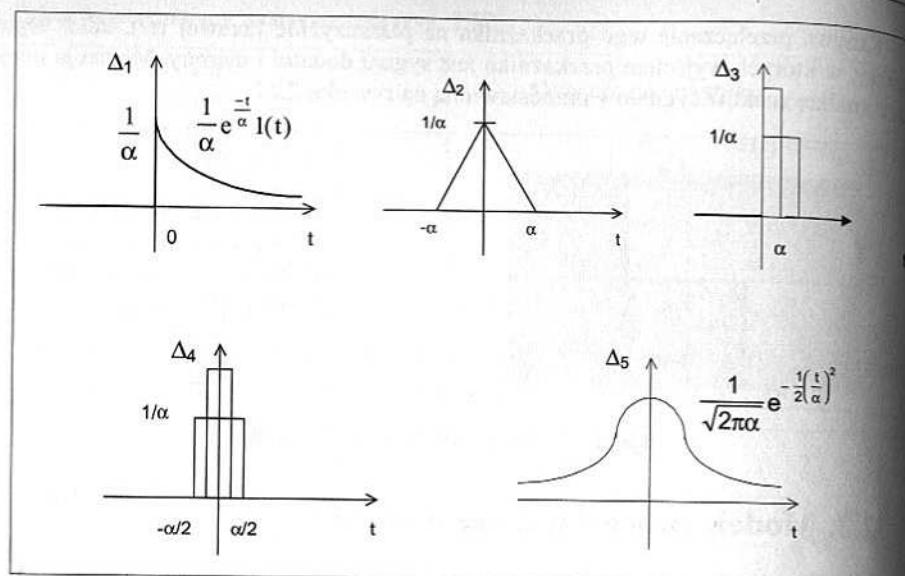
Rys. 2.18. Skok jednostkowy

- Impuls jednostkowy z parametrem α , $\Delta(t, \alpha)$.
Własności:

a) $\Delta(t, \alpha) : \int_{-\infty}^{\infty} \Delta(t, \alpha) dt = 1,$

b) $\lim_{\alpha \rightarrow 0} \Delta(t, \alpha) = \begin{cases} 0 & t \neq 0, \\ \infty & t = 0. \end{cases}$

Różne realizacje impulsu jednostkowego przedstawiono na rysunku 2.19.

Rys. 2.19. Impulsy jednostkowe z parametrem α 3) Delta Diraca $\delta(t)$.**Def. 2.1**

$$\delta(t) = \lim_{\alpha \rightarrow 0} \Delta(t, \alpha) \Rightarrow \int_{-\infty}^{\infty} \delta(t) dt = 1$$

Dla malejącego parametru α , impuls jednostkowy coraz lepiej aproksymuje sygnał $\delta(t)$.

Def. 2.2

$$\delta(t) = \frac{d}{dt} l(t) \Rightarrow \int_{-\infty}^t \delta(\tau) d\tau = l(t)$$

Delta Diraca ma jednostkę t^{-1} [1/s].

4) Impuls Diraca w chwili t_i

$$u(t) = a_i \delta(t - t_i) = \begin{cases} 0 & t \neq t_i \\ \infty & t = t_i \end{cases} \Rightarrow \int_{-\infty}^{\infty} a_i \delta(t - t_i) dt = a_i.$$

Impuls Diraca ma jednostkę taką jak sygnał $u(t)$. Współczynnik a_i zwany jest mocą impulsu i ma jednostkę $a_i [u \cdot t]$, np. u [V] stąd a_i [V·s].

Własność filtrująca delty Diraca pojawia się przy splotie funkcji $x(t)$ i $\delta(t)$ w chwili t_i :

$$\int_{-\infty}^{\infty} x(t) \delta(t - t_i) dt = \int_a^b x(t) \delta(t - t_i) dt = \lim_{\alpha \rightarrow 0} \int_a^b x(t) \Delta(t - t_i) dt = x(t_i).$$

Operacja splotu funkcji z deltą Diraca $\delta(t - t_i)$ zwraca wartość funkcji w punkcie t_i .

5) Sygnał sinusoidalny o amplitudzie A i częstotliwości ω :

$$u(t) = A \sin(\omega t + \varphi), \quad -\infty < t < \infty.$$

Rozróżniamy:

- częstotliwość kątową $\omega = 2\pi f$ [rad·Hz] = $2\pi/T$ [rad/s] zwana też pulsacją (w literaturze anglosaskiej *angular* lub *radian frequency*), ω [rad], okres T [s];
- częstotliwość okresową $f = 1/T$ [Hz = s⁻¹], zwana też częstością (okres/s).

6) Sygnał okresowy:

$$u(t) = u(t + mT), \quad -\infty < t < \infty,$$

gdzie:

$$\begin{aligned} T &- \text{okres funkcji } u(t), \\ m &- \text{dowolna liczba całkowita.} \end{aligned}$$

7) Sygnał zmodulowany: $u(t) = U_m(t) \cos[\varphi(t)]$. Zmieniając chwilową amplitudę $U_m(t)$ otrzymuje się modulację amplitudową, a zmieniając $\varphi(t)$ – modulację fazową

8) Sygnał liniowy

$$u(t) = (at + b) l(t).$$

9) Sygnał wykładniczy

$$u(t) = A e^{-\lambda t}.$$

Stała czasowa eksponenta $T = 1/\lambda$. W każdym przedziale o długości nT amplituda sygnału wykładniczego zmniejsza się o czynnik e^{-n} , tzn.

$$u(t + nT) = e^{-n} u(t).$$

10) Jedną z najważniejszych klasyfikacji sygnałów jest ich teoretyczny podział na:

- sygnały o skończonej energii

$$\lim_{T \rightarrow \infty} \int_{-T}^T [x(t)]^2 dt < \infty,$$

- sygnały o skończonej mocy

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T [y(t)]^2 dt < \infty.$$

Dla sygnałów $x(t)$ o ograniczonej energii ich średnia moc musi być równa zero, np. dla $Ae^{-t/a}$ lub $A[x(t) - x(t-\tau)]$, $\tau < \infty$.

Przyjmuje się też istnienie sygnałów $y(t)$ o nieograniczonej energii, ale o ograniczonej i różnej od zera mocy, np. $1(t)$, $\sin(\omega t + \Theta)$, $1(t) \cdot t$.

2.3.2. Modelowanie układów ciągłych

W modelach o parametrach skupionych nie uwzględnia się wymiarów geometrycznych obiektów, np. mas, sprężyn, rezystorów czy kondensatorów, i pomija wpływ tych wymiarów na rozwiązanie układu, przyjmując, że są one skupione w jednym punkcie przestrzeni, a jedyną zmienną niezależną jest czas. Rząd układu (równania różniczkowego) pokrywa się z reguły z ilością niezależnych magazynów energii, pomiędzy którymi może występować jej wymiana.

Trzy niżej wymienione podstawowe zasady ciągłości [2LUW] stosowane są do budowy modeli dynamicznych.

1) Zasada ciągłości pędu (bilans sił)

$$\frac{d[m(t) \cdot v_i(t)]}{dt} = \frac{dm(t)}{dt} v_i(t) + \frac{dv_i(t)}{dt} m(t) = \sum_{j=1}^N F_{ij}(t),$$

gdzie:

$v_i(t)$ – prędkość w i -tym kierunku,
 F_{ij} – j -ta składowa siły w i -tym kierunku.

- 2) **Zasada ciągłości masy** (bilans materiałowy); dla układu dynamicznego szybkość zmiany magazynowanej masy równa się różnicy dopływu i odpływu masy do układu.
- 3) **Zasada ciągłości energii** (bilans cieplny); szybkość zmian energii układu równa się różnicy dopływu i odpływu energii układu przez konwekcję, dyfuzję, przewodzenie, promieniowanie plus ciepło wydzielone przez reakcję chemiczną, minus praca wykonyana przez układ na otoczeniu.

Podczas modelowania wielu procesów nie zawsze zachodzi potrzeba uwzględniania jednocześnie wszystkich trzech zasad. Do opisu dynamiki prostych modeli mechanicznych wykorzystuje się wspomnianą podstawową zasadę ciągłości pędu, z której wynika również prawo równowagi sił. Z zasady najmniejszego działania uzyskuje się równania Eulera-Lagrange'a dla opisu modelu dynamicznego, a z tych równań – II prawo Newtona.

W elementarnych obiektach mechanicznych, przyjmuje się, że siła zewnętrzna wymuszająca ich ruch może być proporcjonalna do ich:

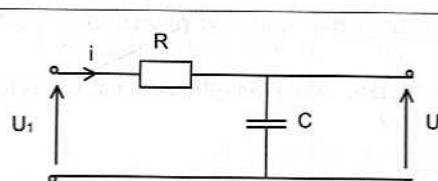
- przesunięcia (prawo Hooke'a dla sprężyny),
- prędkości (zjawisko tarcia i tłumienia),
- przyspieszenia (siła bezwładności masy).

Z zasady ciągłości wynikają również prawa Kirchhoffa o sumowaniu się napięć w oczku i prądów w węźle.

2.3.3. Proste modele dynamiczne liniowe i analogie równań

Przykład 2.5

Układ I rzędu. Człon inercyjny elektryczny i mechaniczny (rys. 2.20 i 2.21).

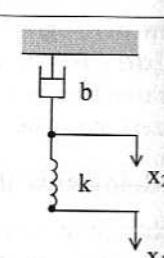


Rys. 2.20. Układ RC

$$i(t) = C \frac{du_2(t)}{dt} = \frac{1}{R} (u_1(t) - u_2(t)).$$

Stąd ładunek: $q(t) = C u_2(t)$,

$$R \frac{dq(t)}{dt} + \frac{1}{C} q(t) = u_1(t).$$



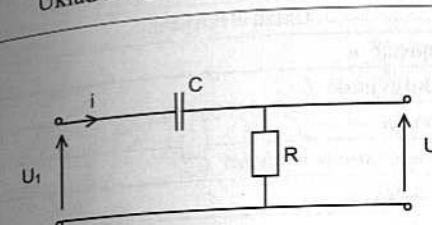
Rys. 2.21. Układ mechaniczny

$$b \frac{dx_2(t)}{dt} + k(x_2(t) - x_1(t)) = 0,$$

$$b \frac{dx_2(t)}{dt} + k x_2(t) = k x_1(t).$$

Przykład 2.6

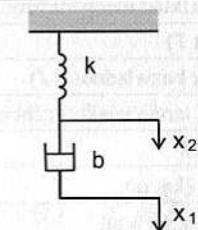
Układ I rzędu. Człon różniczkujący z inercją (rys. 2.22 i 2.23).



Rys. 2.22. Układ RC

$$C \frac{d(u_2(t) - u_1(t))}{dt} = -\frac{u_2(t)}{R},$$

$$R \frac{du_2(t)}{dt} + \frac{u_2(t)}{C} = R \frac{du_1(t)}{dt}.$$



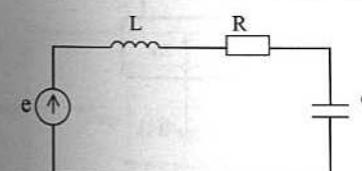
Rys. 2.23. Układ mechaniczny

$$b \frac{d(x_2(t) - x_1(t))}{dt} + kx_2(t) = 0,$$

$$b \frac{dx_2(t)}{dt} + k x_2(t) = b \frac{dx_1(t)}{dt}.$$

Przykład 2.7

Układ II rzędu (oszczędzający) (rys. 2.24 i 2.25).

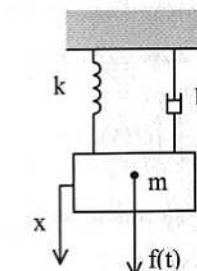


Rys. 2.24. Szeregowy układ RLC

$$L \frac{di(t)}{dt} + Ri(t) + \frac{1}{C} \int_0^t i(\tau) d\tau = e(t).$$

Przyjmując, że $q(t)$, to ładunek elektryczny,

$$L \ddot{q}(t) + R \dot{q}(t) + \frac{1}{C} q(t) = e(t).$$



Rys. 2.25. Układ mechaniczny II rzędu

Równanie równowagi sił
 $m\ddot{x}(t) + b\dot{x}(t) + kx(t) = f(t)$.

Zastosowanie zasady zachowania pędu do prostego układu mechanicznego II rzędu prowadzi do bilansu sił bezwładności, tłumienia i sprężystości oraz sił wymuszających. Zastosowanie zasady zachowania ładunku (praw Kirchhoffa) do prostego obwodu elektrycznego RLC prowadzi do bilansu napięć, który daje w konsekwencji równanie podobne do równania układu mechanicznego. Widać pełne podobieństwo postaci równań. Stąd wynika analogia (tab. 2.1) pomiędzy ich parametrami.

Tabela 2.1

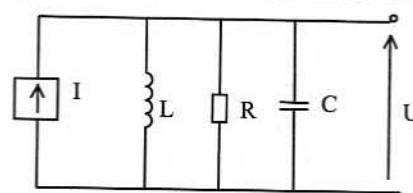
Analogie występujące w układach mechanicznych i elektrycznych

Układ mechaniczny	Układ elektryczny
Sила f (moment T)	Napięcie u
Masa (moment bezwładności J)	Indukcyjność L
Współczynnik tarcia wiskotycznego b	Opór R
Stała sprężyny k	Odwrotność pojemności C^{-1}
Przesunięcie x (kąt α)	Ładunek q
Szybkość $\dot{x} = v, (\dot{\alpha} = \omega)$	Prąd i
Moc $p = fv$	Moc $p = ui$

Omówione analogie nie są jedynymi możliwymi. Rozważając układ dualny do przykładu 2.7, można zauważyć inne możliwości. Ujęto to w tabeli 2.2.

Przykład 2.8

Układ II rzędu (oscylacyjny) (rys. 2.26 i 2.27).

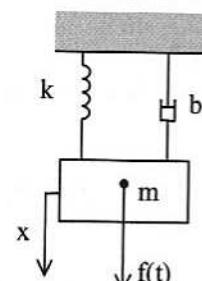


Rys. 2.26. Układ RLC

$$C \frac{dU}{dt} + \frac{U}{R} + \frac{1}{L} \int_0^t U(\tau) d\tau = I,$$

 $d\psi/dt = U$ – strumień magnetyczny,

$$C\ddot{\psi} + \dot{\psi}/R + \psi/L = I.$$



Rys. 2.27. Układ mechaniczny

$$m\ddot{x}(t) + b\dot{x}(t) + kx(t) = f(t).$$

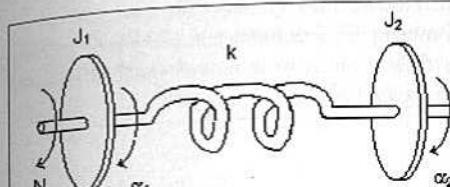
Tabela 2.2

Dualne analogie występujące w układach mechanicznych i elektrycznych

Układ mechaniczny	Układ elektryczny
Sila f (moment T)	Prąd i
Masa (moment bezwładności J)	Pojemność C
Współczynnik tarcia wiskotycznego b	Odwrotność oporu R^{-1}
Stała sprężyny k	Odwrotność Indukcyjności L^{-1}
Przesunięcie x (kąt α)	Strumień ψ
Szybkość $\dot{x} = v, (\dot{\alpha} = \omega)$	Napięcie U
Moc $p = fv$	Moc $p = ui$

Przykład 2.9

Model mechaniczny w ruchu obrotowym (rys. 2.28).



Rys. 2.28. Układ mechaniczny

$$J \frac{d\omega}{dt} = J \frac{d^2\alpha}{dt^2} \text{ - moment mechaniczny,}$$

$$J_1 \frac{d^2\alpha_1}{dt^2} + k(\alpha_1 - \alpha_2) = N,$$

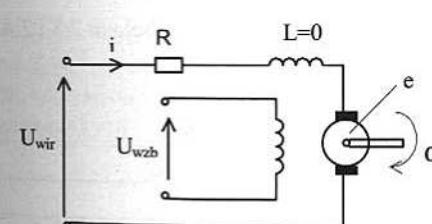
$$J_2 \frac{d^2\alpha_2}{dt^2} + k(\alpha_2 - \alpha_1) = 0,$$

$$J_1 J_2 \alpha_2^{(4)}(t) + k(J_1 + J_2) \alpha_2^{(2)}(t) = \\ = k \cdot N(t)$$

Układ opisywany dwoma równaniami II rzędu lub jednym równaniem IV rzędu ze względu na wybrany kąt $\alpha_2(t)$.

Przykład 2.10

Model elektryczny uproszczonego silnika prądu stałego (obcowzbudnego) (rys. 2.29).



Rys. 2.29. Układ silnika

$$i_{wzb} = \text{const},$$

$$U_{wzb}(t) = u(t) \text{ - sterowanie (napięcie),}$$

$$\alpha(t) = y(t) \text{ - wyjście (kąt obrotu),}$$

$$m_m = J\ddot{\alpha}(t) \text{ - moment bezwładności,}$$

$$m_o = 0 \text{ - moment obciążenia,}$$

$$m_e(t) = k_1 i_{wzb} i(t) \text{ - moment elektryczny,}$$

$$e(t) = k_2 \dot{\alpha}(t) \text{ - siła elektromotoryczna indukcji.}$$

Prawo zachowania ładunku

$$U_{wir}(t) = Ri(t) + e(t) = Ri(t) + k_2 \dot{\alpha}(t),$$

$$i(t) = \frac{U_{wir}(t) - k_2 \dot{\alpha}(t)}{R}, \quad m_m = m_e,$$

Prawo zachowania pędu

$$J\ddot{\alpha}(t) = k_1 i_{wzb} \frac{U_{wir}(t) - k_2 \dot{\alpha}(t)}{R} = \\ = k_3 U_{wir}(t) - k_4 \dot{\alpha}(t),$$

$$T\ddot{\alpha}(t) + \dot{\alpha}(t) = kU_{wir}(t).$$

Model silnika ze względu na kąt obrotu, to człon II rzędu (całkujący z inercją).

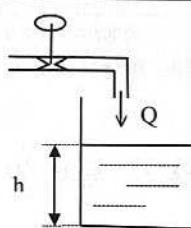
Jeśli prędkość obrotowa $\omega(t)$ będzie przyjęta jako wielkość wyjściowa, to dynamika obiektu opisywana będzie przez równanie

$$T\ddot{\omega}(t) + \dot{\omega}(t) = kU_{wir}(t),$$

a więc reprezentuje człon inercyjny I rzędu

Przykład 2.11

Model hydraulicznego członu całkującego (zbiornik liniowy bezodpływowy) (rys. 2.30 i 2.31).

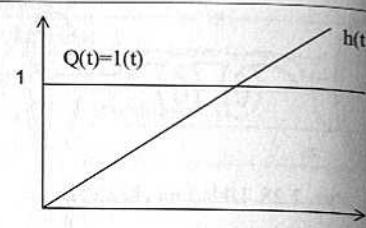


Rys. 2.30. Zbiornik prostopadłościenny

$Q(t)$ – masowe natężenie przepływu,
 ρ – gęstość,

P – pole powierzchni (const),
 objętość $V(t) = Ph(t)$

$$\rho P \frac{dh(t)}{dt} = Q(t) \Rightarrow Th(t) = Q(t).$$

Rys. 2.31. Liniowa zmiana $h(t)$ w czasie

Stąd przy zerowym warunku początkowym

$$h(t) = \frac{1}{T} \int_0^t Q(\tau) d\tau.$$

Zbiornik prostopadłościenny jest dla zmiennej $h(t)$ członem całkującym.

Dodatkowe analogie pomiędzy wielkościami fizycznymi pokazano w tabelach 2.3 i 2.4.

Tabela 2.3

Analogie występujące w układach cieplnych i elektrycznych

Układ cieplny	Układ elektryczny
Różnica temperatur ΔT	Napięcie U
Przepływ ciepła q	Prąd I
Pojemność cieplna (c – ciepło właściwe) $c_T = m \cdot c$	Pojemność C
Odwrotność przewodności cieplnej $R_T = dT/dq$	Opór R
Moc $p = dT \cdot q$	Moc $p = U \cdot I$

Tabela 2.4

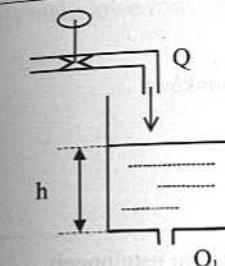
Analogie występujące w układach hydraulicznych i elektrycznych

Układ hydrauliczny	Układ elektryczny
Różnica ciśnień ΔP	Napięcie U
Przepływ cieczy Q	Prąd I
Objętość V	Pojemność C
Opór $R = dP/dQ$	Opór R
Moc $p = \Delta P \cdot Q$	Moc $p = U \cdot I$

Przykład 2.12

Model hydraulicznego obiektu inercyjnego (zbiornik liniowy z odpływem swobodnym).

Zastosowanie zasady zachowania masy i energii oraz równania Bernoulliego w przypadku zbiornika z rysunku 2.32 prowadzi do równania nieliniowego [2GUJ]. Można go jednak zlinearyzować w otoczeniu wybranego punktu pracy. Otrzymuje się w wyniku tego liniową zależność wypływu grawitacyjnego od wysokości cieczy.



Rys. 2.32. Zbiornik z odpływem

$$Q_1(t) = \frac{\sqrt{h(t)}}{R} \approx \frac{h(t)}{R},$$

$$\rho P \dot{h}(t) = Q(t) - \frac{h(t)}{R},$$

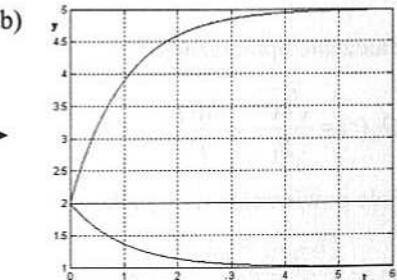
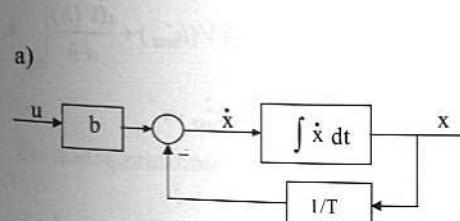
$$\dot{h}(t) = -\frac{1}{\rho PR} h(t) + \frac{1}{\rho P} Q(t).$$

Zbiornik z odpływem posiada cechę samowrównywania poziomu tak jak obiekt inercyjny.

Warto zwrócić uwagę na fakt, że równania dla układu inercyjnego, takie jak równania opisujące układ elektryczny i mechaniczny w przykładzie 2.5 i ostatnie równanie z przykładu 2.12 o postaci

$$\dot{x}(t) = -x(t)/T + bu(t),$$

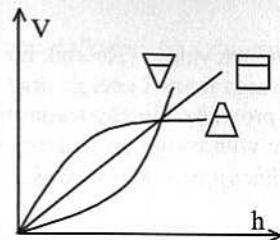
na skok jednostkowy ma rozwiązanie $x(t) = b + e^{-t/T}[x_0 - b]$ i można go przedstawić przy użyciu członu całkującego i sprzężenia zwrotnego (rys. 2.33a). Na rysunku 2.33b przedstawiono również przebiegi zmiennej $x(t)$ dla $x(0) = 2$ i współczynników $b = 5$ oraz $b = 1$.

Rys. 2.33. Struktura członu inercyjnego (a) i odpowiedź $x(t)$ (b)

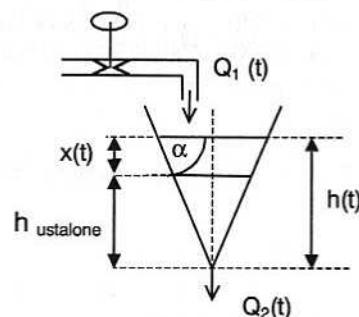
Ten typ naturalnego sprzężenia, niewidoczny w strukturze układu fizycznego, reprezentuje ukryte zasady współoddziałania zmiennych procesowych i obiektów w obecności pola grawitacyjnego, według zasad zachowania pędów, ładunków i mas. Problem sprzężeń omówiono również w podrozdziale 4.3.3 i przykładzie 4.14.

2.3.4. Proste modele dynamiczne nieliniowe

W zależności od kształtu zbiornika charakterystyki statyczne będą opisywane różnymi funkcjami (rys. 2.34).

Rys. 2.34. Charakterystyki $V(h)$ dla zbiorników**Przykład 2.13**

Model zbiornika stożkowego z odpływem (rys. 2.35).



Rys. 2.35. Zbiornik stożkowy

Założenie upraszczające:

$$Q_2(t) = \frac{\sqrt{h(t)}}{R_1} \approx \frac{h(t)}{R}.$$

Pole powierzchni na wysokości h :

$$P_h = \pi \cdot r_h^2 = \pi \frac{h^2}{\tan^2 \alpha} = k_1 h^2,$$

$$V(h) = \int_0^h P_h(x) dx = \frac{k_1}{3} h^3 = f(h).$$

Stan ustalony:

$$h_{ust}, Q_{1ust}, Q_{2ust}$$

$$Q_{1ust} = h_{ust} / R.$$

Odchyłki od stanu ustalonego

$$h(t) = h_{ust} + x(t),$$

$$Q_1(t) = Q_{1ust} + u(t).$$

Bilans masowy daje równanie różniczkowe nieliniowe ze względu na wysokość $h(t)$

$$\frac{dV(t)}{dt} = Q_1(t) - Q_2(t) = Q_1(t) - \frac{h(t)}{R}.$$

Rozkładamy nieliniowość $V(h)$ w szereg Taylora:

$$V(h) = V(h_u + x) \equiv V(h_{ust}) + \left. \frac{dV(h)}{dh} \right|_{h_u} x,$$

$$V(h) \approx \frac{k_1}{3} h_{ust}^3 + k_1 h_{ust}^2 \cdot x,$$

stąd

$$k_1 h_{ust}^2 \frac{dx(t)}{dt} = Q_{1ust} + u(t) - \frac{h_{ust}}{R} - \frac{1}{R} x(t).$$

Zbiornik jest modelowany jako liniowy człon inercyjny I rzędu. Nieliniowość przejawia się w stałej czasowej

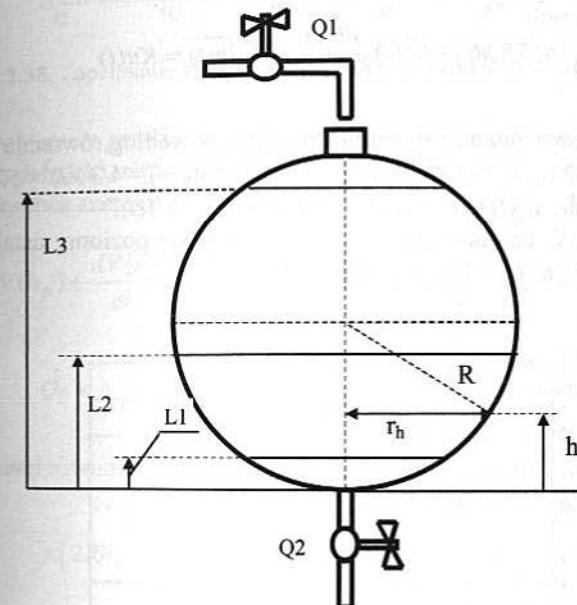
$$T(h_{ust}) = R k_1 h_{ust}^2$$

– różnej w różnych punktach pracy h_{ust}

$$T(h_{ust}) \dot{x}(t) + x(t) = R u(t).$$

Przykład 2.14

Model zbiornika kulistego z odpływem.

Przeprowadzona zostanie dokładniejsza analiza zapełniania cieczą kulistego zbiornika o promieniu R , przy stałym zasilaniu objętościowym Q_1 . W związku z nieliniową zależnością objętości zapełnienia od poziomu cieczy oraz nieliniową zależnością natężenia wypływu od poziomu cieczy (pod ciśnieniem hydrostatycznym), dynamikę zmian wysokości cieczy opisuje nieliniowe równanie różniczkowe I rzędu.

Rys. 2.36. Zbiornik kulisty

Zachodzą zależności (rys. 2.36)

$$(R-h)^2 + r_h^2 = R^2 \Rightarrow r_h^2 = 2Rh - h^2 \quad \text{dla } h \in [0, 2R].$$

Pole powierzchni lustra cieczy $P(h) = \pi r_h^2 = \pi(2Rh - h^2)$.

$$\text{Objętość zapełnienia: } V(h) = \int_0^h P(s) ds = \pi \left(2R \frac{h^2}{2} - \frac{h^3}{3} \right).$$

Stąd

$$\frac{dV(h)}{dt} = \pi \left(2Rh(t) \frac{dh(t)}{dt} - h^2(t) \frac{dh(t)}{dt} \right).$$

Z prawa ciągłości przepływu

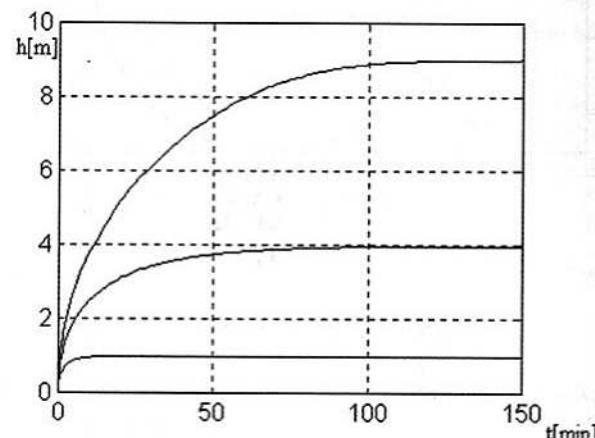
$$\frac{dV(t)}{dt} = Q_1(t) - Q_2(t),$$

przy założeniu przepływu zasilającego $Q_1(t) = Ku(t)$ i turbulentnego przepływu wyjściowego $Q_2(t) = b\sqrt{h(t)}$, (gdzie współczynnik b zależy od zastosowanego zaworu i rodzaju cieczy), otrzymuje się nieliniowe równanie różniczkowe poziomu

$$\pi(2Rh(t) - h^2(t)) \frac{dh(t)}{dt} + b\sqrt{h(t)} = Ku(t) \quad (2.1)$$

Przeprowadzono symulację dynamiki zapełniania według równania (2.1) dla parametrów zbiornika: $R = 5 \text{ m}$, $K = 12 \text{ m}^3/\text{min}$, $b = 12 \text{ m}^{2.5}/\text{min}$, $V_k = 4/3 \pi R^3 = 523.6 \text{ m}^3$, dla trzech sterowań skokowych: $u_1(t) = 1(t)$, $u_2(t) = 2 \cdot 1(t)$, $u_3(t) = 3 \cdot 1(t)$.

Łatwo zauważyc, że dla czasu $t \rightarrow \infty$ otrzymuje się poziomy ustalone w zbiorniku odpowiednio: $L_1 = 1 \text{ m}$, $L_2 = 4 \text{ m}$, $L_3 = 9 \text{ m}$ (rys. 2.37).

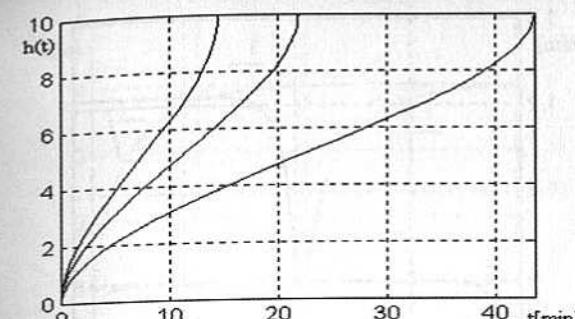


Rys. 2.37. Zapełnianie zbiornika kulistego

Wynikają one z nieliniowego algebraicznego równania stanu ustalonego

$$\sqrt{h(\infty)} = u(\infty) \quad (2.2)$$

Przy zamkniętym zaworze wypływu ($Q_2 = 0$) czasy t_k całkowitego zapełnienia zbiornika, tzn. $h(t_k) = 10 \text{ m}$ wynoszą dla trzech sterowań, odpowiednio: $t_{k1} = \pi R^3 / 9 = 43,63 \text{ min}$, $t_{k2} = \pi R^3 / 18 = 21,81 \text{ min}$, $t_{k3} = \pi R^3 / 27 = 14,54 \text{ min}$ (rys. 2.38), a dynamika zapełniania przedstawiona jest na rysunku 2.37. Przebiegi nie są liniowe w czasie, w odróżnieniu od przypadku zbiornika prostopadłościennego z rysunku 2.31 (np. walcowego).



Rys. 2.38. Zapełnianie zbiornika kulistego przy zamkniętym odpływie

Osobnym problemem jest linearyzacja równania (2.1) wokół wybranego punktu pracy h_p dla niewielkich odchyłek: sterowania i zmiennej stanu, tzn. $u(t) = u_p + \Delta u(t)$; $h(t) = h_p + x(t)$. Linearyzacja za pomocą szeregu Taylora daje wzory:

$$V(h_p + x) \equiv V(h_p) + \left. \frac{dV(h)}{dh} \right|_{h_p} \cdot x = \pi \left(2R \frac{h_p^2}{2} - \frac{h_p^3}{3} \right) + \pi \left(2Rh_p - h_p^2 \right) \cdot x,$$

$$Q_2 = b\sqrt{h_p + x} \equiv b\sqrt{h_p} + \frac{b}{2}(h_p)^{-0.5} \cdot x.$$

Przy wymaganej równości $Ku_p = b\sqrt{h_p}$ otrzymuje się równanie różniczkowe liniowe

$$\pi(2Rh_p - h_p^2) \cdot \frac{dx(t)}{dt} = Ku_p + K \Delta u(t) - b\sqrt{h_p} - b\frac{1}{2}(h_p)^{-0.5} \cdot x(t).$$

Po uproszczeniach liniowy model odchyłkowy ma postać obiektu inercyjnego

$$\pi(2Rh_p - h_p^2) \cdot \frac{dx(t)}{dt} + b\frac{1}{2}(h_p)^{-0.5} \cdot x(t) = K \Delta u(t).$$

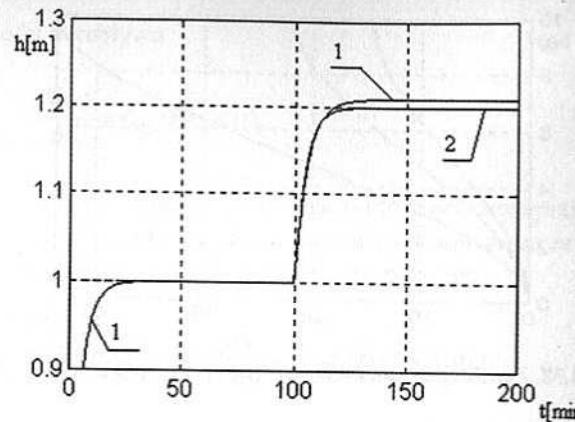
Dla punktu pracy $h_p = 1 \text{ m}$, $u_p = 1$ otrzymuje się więc równanie liniowe

$$9\pi \dot{x}(t) + 6x(t) = 12 \Delta u(t) \quad (2.3)$$

Dla czasu $t \rightarrow \infty$ upraszcza się ono do równania algebraicznego

$$6x(\infty) = 12 \Delta u(\infty) \quad (2.4)$$

Dla przykładowej odchyłki sterowania $\Delta u(t) = 0,1 \cdot 1(t) \text{ m}^3/\text{min}$ daje to przyrost stanu ustalonego o $x(\infty) = 0,2 \text{ m}$, czyli $h(\infty) = 1,2 \text{ m}$ (rys. 2.39, krzywa 2). Rozwiązanie dokładne równania (2.1) dla $u(t) = 1,1 \cdot 1(t)$ daje stan ustalony $h(\infty) = 1,21 \text{ m}$, bo $u(\infty) = 1,1 = \sqrt{1,21} = \sqrt{h(\infty)}$ (rys. 2.39, krzywa 1). Błąd pomiędzy modelem liniowym i nielinijkim jest niewielki.



Rys. 2.39. Charakterystyki zbiornika
Objaśnienia w tekście

Dla dużych odchyłek sterowania, np. $\Delta u(t) = 1(t) \text{ m}^3/\text{min}$, rozwiązanie ustalone odchyłkowe (2.4) dla otrzymanego modelu liniowego (2.3) daje $x(\infty) = 2 \text{ m}$, czyli $h(\infty) = 3 \text{ m}$, a rozwiązanie dokładne równania nieliniowego (2.1) dla $u(t) = 2 \cdot 1(t)$ daje stan ustalony $h(\infty) = 4 \text{ m}$, bo z (2.2) $u(\infty) = 2 = \sqrt{4} = \sqrt{h(\infty)}$ (krzywa 1, rys. 2.40).

Błąd ustalony wynikający z posługiwania się modelem liniowym (2.3) jest duży (1 m) i może być usunięty poprzez modyfikację wykorzystywanego modelu liniowego (2.3) na inny model liniowy dostosowany do maksymalnych amplitud sterowania.

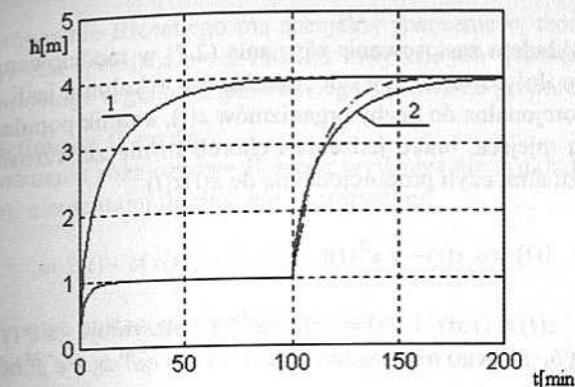
Proste rachunki prowadzą do modelu

$$17,44\pi \dot{x}(t) + 4x(t) = 12\Delta u(t) \quad (2.5)$$

Gwarantuje on dokładniejsze odwzorowanie odpowiedzi układu nieliniowego wokół punktu pracy $h_{p1} = 1 \text{ m}$ na odchyłki sterowania $\Delta u(t) = 1(t) \text{ m}^3/\text{min}$, dając wartość w stanie ustalonym $x(\infty) = 12/4 = 3 \text{ m}$, czyli poziom $h(\infty) = 1 \text{ m} + 3 \text{ m} = 4 \text{ m}$ – taki jak w układzie nieliniowym. Model ten otrzymuje się linearyzując model nieliniowy wokół punktu $h_{p2} = 2,25 \text{ m}$.

Na rysunku 2.40 przedstawiono:

- rozwiązanie dokładne równania nieliniowego (2.1) dla sterowania $u(t) = 2 \cdot 1(t)$, dochodzące do $h(\infty) = 4 \text{ m}$ (krzywa 1);
- rozwiązanie dokładne równania nieliniowego startujące z poziomu $h_{p1} = 1 \text{ m}$ na dodatkowe odchylenie sterowania $\Delta u(t) = 1(t) \text{ m}^3/\text{min}$, dochodzące również do poziomu $h(\infty) = 4 \text{ m}$ (krzywa 2);
- rozwiązanie równania liniowego otrzymane z modelu (2.5), zlinearyzowanego w $h_{p2} = 2,25 \text{ m}$, na skok $\Delta u(t) = 1(t) \text{ m}^3/\text{min}$, dające $x(\infty) = 3 \text{ m}$; startując z początkowego poziomu $h_{p1} = 1 \text{ m}$, dochodzi ono również do poziomu $h(\infty) = 4 \text{ m}$ (krzywa przerywana).



Rys. 2.40. Zapełnianie zbiornika
Objaśnienia w tekście

Z przykładu widać, że przy odchyłkach sterowania typu skokowego o znanej amplitudzie, zmniejszenie błędu modelowania w stanie ustalonym, wynikającego z linearyzacji, może być uzyskane poprzez dobór teoretycznego punktu linearyzacji. Jest to istotne przy projektowaniu układów stabilizacji instalacji przemysłowych w ich punktach pracy.

2.3.5. Standardowe nieliniowości w równaniach różniczkowych

Poniżej przedstawiono trzy typy standardowych nieliniowości.

1) Równanie Bernoulliego I rzędu

Równanie Bernoulliego I rzędu jest tworzone przez standardową nieliniowość typu wykładniczego:

$$\dot{x}(t) + f(t) \cdot x(t) - g(t) \cdot x^\alpha(t) = 0; \quad \alpha \neq 0, \quad \alpha \neq 1, \quad f, g \in C[a, b] \quad (2.6)$$

To pozornie nieliniowe równanie można zamienić na równanie liniowe przez wprowadzenie nowej zmiennej

$$z(t) = \frac{1}{x^{\alpha-1}(t)}, \quad \text{stąd } \dot{z}(t) = \frac{(1-\alpha)\dot{x}(t)}{x^\alpha(t)}.$$

Ponieważ równanie (2.6) ma postać

$$\frac{\dot{x}(t)}{x^\alpha(t)} + \frac{f(t)}{x^{\alpha-1}(t)} - g(t) = 0$$

to można go zapisać jako równanie liniowe zmiennej $z(t)$,

$$\dot{z}(t) + (1-\alpha)f(t) \cdot z(t) - (1-\alpha) \cdot g(t) = 0 \quad (2.7)$$

Przykład 2.15

Typowym przykładem zastosowania równania (2.7) w modelowaniu jest prawo rozwoju (przyrostu ilościowego) żywych organizmów w kolonii, jeśli szybkość rozmania jest proporcjonalna do liczby organizmów $x(t)$, a zanik populacji z powodu konkurencji, braku miejsca, braku jedzenia i chorób rośnie ze wzrostem ilości spotkań między organizmami, czyli proporcjonalnie do $x(t)x(t)$.

$$\dot{x}(t) = a x(t) - b x^2(t).$$

Podstawiając: $z(t) = 1/x(t)$ i $\dot{z}(t) = -\dot{x}(t)/x^2(t)$ otrzymuje się równanie liniowe $\dot{z}(t) = -a z(t) + b$, którego rozwiązaniem jest $z(t) = e^{-at} z_0 + e^{-at} b(e^{at} - 1)/a$. Stąd ostatecznie

$$x(t) = \frac{1}{z(t)} = \frac{a x_0}{(a - b x_0) e^{-at} + b x_0}.$$

Liczba osobników zmienia się w czasie od wartości początkowej x_0 do wartości końcowej a/b .

Przykład 2.16

Inny przykład równania Bernoulliego otrzymuje się, modelując układ regulacji z nieliniowym dodatnim sprzężeniem zwrotnym

$$\dot{x}(t) = -x(t) + x^3(t).$$

Podstawiając: $z(t) = \frac{1}{x^2(t)}$, $\dot{z}(t) = -\frac{2\dot{x}(t)}{x^3(t)}$, otrzymuje się równanie liniowe $\dot{z}(t) = 2z(t) - 2$, którego rozwiązaniem jest $z(t) = e^{2t}(z_0 - 1) + 1$.

Ostatecznie rozwiązanie $x(t)$ zanika do zera dla warunków początkowych z pasem

$$|x(t)| < 1 \text{ według wzoru } x(t) = \left[e^{2t} (x_0^{-2} - 1) + 1 \right]^{-0.5}.$$

2) Równanie Riccatiego I rzędu

Nieliniowość wielomianowa występuje również w prawej stronie równania Riccatiego

$$\dot{x}(t) = ax^2(t) + bx(t) + c.$$

Ogólna postać jego rozwiązania ma wtedy formę

$$x(t) = \frac{\gamma + \delta e^{at}}{1 - \omega e^{at}}.$$

2.3. Modele procesów dynamicznych

Macierzowe równanie Riccatiego ma specjalne znaczenie w teorii sterowania optymalnego, gdyż jest spełniane przez macierz optymalnych współczynników wielowymiarowego liniowego regulatora generującego sterowanie proporcjonalne do stanu.

3) Równania nieliniowe II rzędu

Nieliniowe równania różniczkowe II rzędu wywodzą się z układów mechanicznych z nieliniowymi elementami tłumiącymi i sprężystymi

$$m \ddot{x}(t) + b[\dot{x}(t)] + g[x(t)] = 0.$$

Najbardziej typowe postaci wymieniono poniżej:

- układ z twardą sprężyną $\ddot{x}(t) + a \dot{x}(t) + b x(t) + c x^2(t) = 0$,
- układ wahadła $\ddot{x}(t) + a \dot{x}(t) + b \sin[x(t)] = 0$,
- układ z nieliniowym tłumieniem $\ddot{x}(t) + a \dot{x}(t) - b \dot{x}^3(t) + x(t) = 0$,
- równanie Duffinga $\ddot{x}(t) + b x(t) + c x^3(t) = 0$,
- równanie Lienarda $\ddot{x}(t) + b[x(t)] \cdot \dot{x}(t) + g[x(t)] = 0$,
- równanie Lienarda (szczególna postać) $\ddot{x}(t) - a x(t) \cdot \dot{x}(t) - b x^3(t) = 0$,
- równanie Van der Pola (szczególna postać Lienarda) $\ddot{x}(t) - \xi [1 - x^2(t)] \cdot \dot{x}(t) + x(t) = 0$,
- równania Lotki–Volterry (opisujące populację układu dwóch przeciwników: drapieżcy i ofiary) $\dot{x}_1(t) = x_1(t) [a - b x_2(t)],$
 $\dot{x}_2(t) = x_2(t) [c x_1(t) - d],$
- równania z nieliniowościami nieciągłyimi $\ddot{x}(t) + M \operatorname{sgn}[x(t) + k \dot{x}(t)] = 0$.

Takie równania powstają przy opisie układów sterowania z regulatorami przekaźnikowymi (dwu- i trójpołożeniowymi). Na przykład równanie opisujące stabilizację pozycji wirnika (kąt $x(t)$), w układzie regulacji z uproszczonym modelem silnika (człon podwójnie całkujący $\ddot{x}(t) = u(t)$) i z regulatorem przekaźnikowym realizującym przełączane sterowanie $\pm M$ (quasi-optymalnoczasowe $u^{opt}(t)$) na podstawie pomiaru położenia i prędkości wirnika,

$$u^{opt}(t) = -M \operatorname{sgn}[x(t) + k \dot{x}(t)]$$

ma postać równania nieliniowego $\ddot{x}(t) + M \operatorname{sgn}[x(t) + k \dot{x}(t)] = 0$.

3. Analiza układów liniowych ciągłych w dziedzinie czasu

3.1. Równania różniczkowe

Def. 3.1. Równanie zawierające pochodne lub różniczki jednej lub wielu zmiennych zależnych, będących funkcjami jednej lub wielu zmiennych niezależnych, nazywa się równaniem różniczkowym.

Def. 3.2. Stopień najwyższej pochodnej jest zwany rzędem równania.

Def. 3.3. Jeśli w równaniu występują tylko różniczki zwyczajne zmiennych zależnych tylko od jednej zmiennej niezależnej, to równanie różniczkowe nazywa się zwykłym

$$x(t) \frac{d^2 x(t)}{dt^2} + 2 \frac{dx(t)}{dt} + x^2(t) = 0.$$

Def. 3.4. Jeśli w równaniu występują różniczki cząstkowe jednej lub wielu zmiennych zależnych od jednej lub wielu zmiennych niezależnych, to równanie różniczkowe nazywa się cząstkowym.

$$x(z_1, z_2, t) \frac{\partial^2 x(z_1, z_2, t)}{\partial z_1^2} + \frac{\partial^2 x(z_1, z_2, t)}{\partial z_2^2} + 2 \frac{\partial x(z_1, z_2, t)}{\partial t} + x^2(z_1, z_2, t) = 0.$$

Def. 3.5. Równanie różniczkowe nazywa się liniowym jeśli zmienne zależne i ich pochodne występują w pierwszej potędze, a współczynniki są stałe lub zależą tylko od zmiennej niezależnej

$$a_n(t) \frac{d^n x(t)}{dt^n} + a_{n-1}(t) \frac{d^{n-1} x(t)}{dt^{n-1}} + \dots + a_1(t) \frac{dx(t)}{dt} + a_0(t)x(t) = 0.$$

W pozostałych przypadkach mówi się o równaniach nieliniowych.

Def. 3.6. Rozwiązańem na pewnym przedziale, równania różniczkowego n -tego rzędu, nazywamy każdą funkcję posiadającą co najmniej n pochodnych, która podstawiona do równania zamienia go w identyczność. Rozwiązanie, które byłoby funkcją równą zero w tym przedziale, nazywa się trywialnym.

Można wymienić kilka charakterystycznych stwierdzeń dotyczących równań różniczkowych:

- Zwykle istnieje nieskończenie wiele funkcji $x(t)$, które spełniają dane równanie różniczkowe zwyczajne

$$\frac{dx(t)}{dt} = f(t).$$

- W większości przypadków daje się utworzyć n -parametrową rodzinę rozwiązań, z której poprzez dobór parametrów wynikają wszystkie inne rozwiązania,

$$X(t) = F(x) + C, \quad \text{gdzie } C \text{ oznacza stałą.}$$

Taki zbiór funkcji nazywamy rozwiązań ogólnym lub całką ogólną równania. Dla konkretnych parametrów otrzymujemy tzw. rozwiązanie szczególne, które charakteryzuje się tym, że przyjmuje dla zadanej t_0 z góry zadaną wartość $x(t_0) = x_0$.

Czasami rozwiązaniem jest też funkcja, której nie da się uzyskać z rodziny przez dobranie jakichkolwiek wartości parametrów. Nazywa się ono rozwiązaniem osobliwym.

- Parametry mogą być dobierane na podstawie szczegółów sformułowania zadania.
- Zadanie z warunkami początkowymi otrzymuje się, zadając wartości zmiennych zależnych i ich pochodnych dla jednej wybranej wartości zmiennej niezależnej.
- Zadania brzegowe otrzymuje się, zadając wartości zmiennej zależnej dla dwóch lub więcej wartości zmiennych niezależnych.

3.2. Równania różniczkowe stanu

Pojęcie stanu układu wprowadzone zostanie na przykładzie obiektów o parametrach skupionych z nieodchylonym, argumentem (tzn. np. bez opóźnień transportowych).

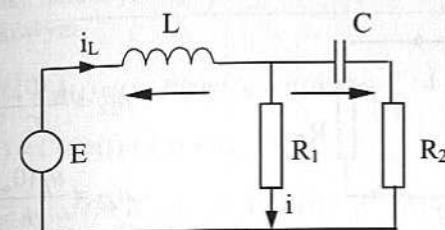
Def. 3.7. Stan $x(\cdot)$ układu o parametrach skupionych, jest to najmniejsza liczba wielkości, które trzeba określić w chwili $t = t_0$, aby można było przewidzieć jednoznacznie zachowanie się tego układu, czyli przyszłą wartość stanu dla każdej chwili czasu $t_1 > t_0$, przy znajomości modelu systemu i przebiegu sterowania $u(t)$ w przedziale $[t_0, t_1]$. Dla dowolnej chwili t wielkości te można przedstawić jako wektor liczbowy $x(t) = [x_1(t), x_2(t), \dots, x_n(t)]^T$. Wielkości te niekoniecznie muszą być związane z rzeczywistymi zmiennymi procesowymi. Często mogą być wybrane umownie i nie posiadać interpretacji fizycznej.

Def. 3.8. Wyjściem $y(\cdot)$ układu będziemy nazywać wybrany zbiór wielkości procesowych, które są szczególnie interesujące z punktu widzenia modelowania lub sterowania (np. należy je stabilizować na pewnym poziomie z użyciem sprzężenia zwrotnego). Praktycznie wielkości te powinny być dostępne bezpośrednio do pomiaru, choć teoretycznie nie jest to wymagane. Zbiór ten oznacza się $y(t) = [y_1(t), y_2(t), \dots, y_m(t)]^T$, gdzie $m \leq n$. Wyjście y układu jest z reguły funkcją algebraiczną zmiennych stanu $x(\cdot)$ i ewentualnie bezpośrednio sterowania.

Dynamika układu, będzie więc opisywana przez równanie różniczkowe stanu i równanie algebraiczne wyjścia.

Przykład. 3.1

Przykład zapisu równania stanu i wyjścia dla prostego układu elektrycznego RLC z rysunku 3.1.



Rys. 3.1. Układ RLC

Metoda wyprowadzania równania zmiennych stanu dla liniowego układu elektrycznego oparta jest na twierdzeniach Kirchhoffa dla oczek i węzłów. W celu szybkiego określenia parametrów równań stanu, można wykorzystać mnemotechniczną metodę rozcinania obwodów w chwili $t = 0$ i wprowadzenia idealnych jednostkowych źródeł napięcia E i prądu I . Jako zmienne stanu przyjmuje się napięcia na pojemnościach C (kondensatorach) i prądy w indukcyjnościach L . Wyjściem jest prąd i .

W układzie liniowym II rzędu (2 równania I rzędu) należy określić 9 parametrów a, b, c, d :

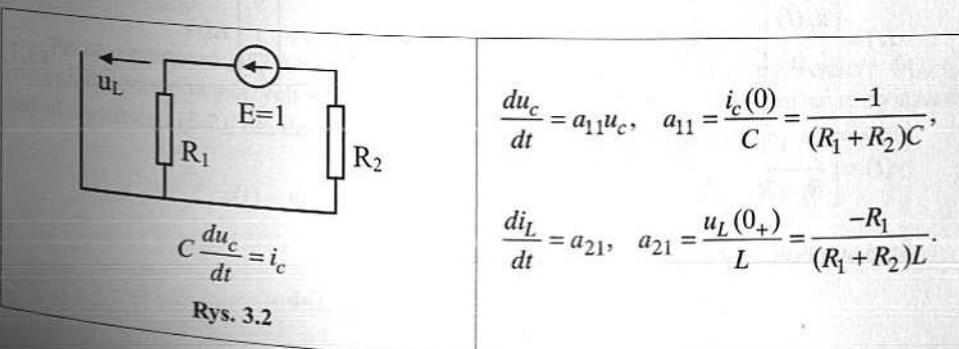
$$\frac{du_C}{dt} = a_{11}u_C + a_{12}i_L + b_1E,$$

$$\frac{di_L}{dt} = a_{21}u_C + a_{22}i_L + b_2E,$$

$$y = c_1u_C + c_2i_L + dE.$$

- Obliczenie a_{11} i a_{21}

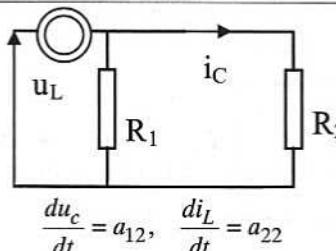
W chwili $t = 0$ do obwodu biernego (źródła napięcia zwarte, źródła prądu rozwarłe) włączamy szeregowo z pojemnością źródło napięcia o wartościowej $u_C = E = 1$ (rys. 3.2). W chwili $t = 0_+$ pojemności C stanowią zwarcie, a indukcyjności L – rozwarcie, tzn. $i_L = 0$.



Rys. 3.2

- Obliczenie a_{12} i a_{22}

W chwili $t = 0_+$ włączamy do obwodu biernego równolegle do L idealne jednostkowe źródło prądu. $E = 0$, $i_L = i = 1$ (rys. 3.3)



Rys. 3.3

$$a_{12} = \frac{i_c(0_+)}{C} = \frac{R_1}{(R_1 + R_2)C},$$

$$a_{22} = \frac{u_L(0_+)}{L} = \frac{-R_1 R_2}{(R_1 + R_2)L}.$$

$$\frac{du_c}{dt} = a_{12}, \quad \frac{di_L}{dt} = a_{22}$$

- Wyliczanie macierzy $B = [b_1, b_2]^T$

Zakładamy:

$$u_C(0_+) = 0, \quad i_L(0_+) = 0, \quad E(0_+) = 1,$$

$$\frac{du_c}{dt} = b_1 = 0, \quad \frac{di_L}{dt} = b_2 \cdot 1, \quad b_2 = \frac{1}{L}.$$

- Wyliczanie elementów macierzy $C = [c_1, c_2]$:

$$u_C = 1, \quad i = y = c_1 u_C, \quad i_L = 0, \quad E = 0, \quad i = c_1 = \frac{1}{R_1 + R_2},$$

$$i = c_2 i_L, \quad u_C = 0, \quad E = 0, \quad i = \frac{R_2}{R_1 + R_2} = c_2.$$

- Wyliczanie elementów macierzy $D = d$:

$$i_L = 0, \quad u_C = 0, \quad E = 1, \quad i = dE = d = 0.$$

Końcowa postać równania stanu i wyjścia:

$$x(t) = \begin{bmatrix} u_C(t) \\ i_L(t) \end{bmatrix}, \quad \dot{x}(t) = \begin{bmatrix} -1/(R_1 + R_2)C & R_1/(R_1 + R_2)C \\ -R_1/(R_1 + R_2)L & -R_1 R_2/(R_1 + R_2)L \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1/L \end{bmatrix} E(t) \quad (3.1)$$

$$y(t) = \begin{bmatrix} 1/(R_1 + R_2) & R_2/(R_1 + R_2) \end{bmatrix} x(t)$$

Stąd symboliczny zapis:

$$\dot{x}(t) = A x(t) + B u(t), \quad x_0 \in R^n,$$

$$y(t) = C x(t).$$

3.3. Liniowe niestacjonarne równanie stanu

Załóżmy, że dynamika układu jest opisywana liniowym, macierzowym i niestacjonarnym równaniem różniczkowym

$$\frac{dx(t)}{dt} = A(t)x(t) + B(t)u(t), \quad x(0) = x_0 \quad (3.2)$$

$$y(t) = C(t)x(t) + D(t)u(t)$$

$$\forall t \geq t_0 : x(t) \in R^n, u(t) \in R^r, y(t) \in R^m, A(t) \in PC(T : L(R^n)),$$

$$B(t) \in PC(T : L(R^r : R^n)), C(t) \in L(R^n : R^m), D(t) \in L(R^r : R^m),$$

gdzie:

$PC(T : L)$ – przestrzeń funkcji przedziałami ciągły,

$L(\cdot, \cdot)$ – przestrzeń funkcji całkowalnych w dziedzinie czasu.

Rozwiążanie równania (3.2) bazuje na tzw. macierzowym rozwiązyaniu fundamentalnym $\Phi(t, t_0)$ równania jednorodnego

$$\frac{dx(t)}{dt} = A(t)x(t) \quad (3.3)$$

Macierz fundamentalna $\Phi(t, t_0)$ zwana też macierzą tranzycyjną układu, zależy od dwóch parametrów t i t_0 i składa się z kolumn $\phi_i(t, t_0) = [\phi_{1i}(\cdot), \dots, \phi_{ni}(\cdot)]^T$, $i \in [1, n]$, otrzymywanych jako rozwiązanie wektorowe $x_i(t) = \phi_i(t, t_0)$ równania (3.3) dla specjalnych warunków początkowych $x_i(t_0) = e_i$, gdzie $e_i = [0, 0, \dots, 0, 1, 0, \dots, 0]^T$ (jedynka na i -tej pozycji).

Tak utworzona macierz fundamentalna posiada pewne właściwości [3MIW]:

- 1) $\frac{d\Phi(t, t_0)}{dt} = A(t)\Phi(t, t_0)$
- 2) $\Phi(t_1, t_2)\Phi(t_2, t_3) = \Phi(t_1, t_3)$
- 3) $\Phi(t, t_0)\Phi(t_0, t) = \Phi(t, t) = I \Rightarrow \Phi(t, t_0) = \Phi^{-1}(t_0, t)$

$$4) \quad \Phi(t, t_0) = \sum_{n=0}^{\infty} \Phi_n(t, t_0), \quad \text{gdzie } \Phi_{n+1}(t, t_0) = \int_{t_0}^t A(\tau)\Phi_n(\tau, t_0)d\tau; \quad \Phi_0(t, t_0) = I$$

Po obustronnym wymnożeniu równania różniczkowego (3.2) przez macierz $\Phi(t, t_0)$ i uwzględnieniu powyższych własności, można otrzymać wzór na pełną postać rozwiązania niejednorodnego (3.2), dla dowolnego wektora warunków początkowych i sterowań

$$x(t) = \Phi^{-1}(t_0, t) x(t_0) + \int_{t_0}^t \Phi^{-1}(\tau, t) B(\tau) u(\tau) d\tau =$$

$$= \Phi(t, t_0) x(t_0) + \int_{t_0}^t \Phi(t, \tau) B(\tau) u(\tau) d\tau.$$

3. Analiza układów liniowych ciągły w dziedzinie czasu

W rozwiązyaniu zwraca uwagę wzajemna niezależność części rozwiązyania pochodzącej od warunków początkowych i od sterowania (własność superpozycji).

Dla przypadku równania jednorodnego (3.3) rozwiązanie ma postać: $x(t) = \Phi(t, t_0)x_0$.

W przypadku równania niestacjonarnego, skalarnego, $x(t) \in R$:

$$\dot{x}(t) = a(t)x(t) + b(t)u(t), \quad x(t_0) = x_0$$

Rozwiązanie ma również postać

$$x(t) = \varphi(t, t_0)x_0 + \int_{t_0}^t \varphi(t, \tau)b(\tau)u(\tau)d\tau,$$

A skalarne rozwiązanie fundamentalne jest funkcją wykładniczą

$$\varphi(t, t_0) = e^{\int_{t_0}^t a(\tau)d\tau}$$

Zachodzi pytanie, czy dla równania macierzowego (3.2) lub (3.3) macierz fundamentalna może być również wyrażona wzorem wykładniczym

$$\Phi(t, t_0) = e^{\int_{t_0}^t A(\tau)d\tau}$$

Odpowiedź jest twierdząca, tylko dla szczególnych przypadków niestacjonarnej macierzy $A(t)$. Z własności (3.4) wynika, bowiem, że powinno zachodzić

$$\frac{d}{dt} \left[e^{\int_{t_0}^t A(\tau)d\tau} \right] = A(t) e^{\int_{t_0}^t A(\tau)d\tau}$$

Jak można wykazać (np. [3CSF]), zachodzi to tylko wtedy, gdy mnożenie $A(t)$ i jej całki $\int_{t_0}^t A(\tau)d\tau$ jest przemienne lub gdy dla $\forall t_1$ i t_2 przemienne jest mnożenie macierzy $A(t)$, tzn. $A(t_1)^*A(t_2) = A(t_2)^*A(t_1)$.

Te dwa równoważne warunki zachodzą zawsze, np. dla szczególnych postaci $A(t)$, gdy:

- 1) $A(t)$ jest macierzą diagonalną $J(t)$;
- 2) $A(t) = A$, tzn. jest macierzą stałą (system stacjonarny).

Dla dowolnej macierzy $A(t)$ zachodzi natomiast wzór Ostrogradskiego–Liouville'a [3ANJ].

$$\det[\Phi(t, t_0)] = e^{\int_{t_0}^t \text{tr}[A(s)]ds}$$

3.4. Liniowe stacjonarne równanie stanu

3.4. Liniowe stacjonarne równanie stanu

Standardowa forma liniowego równania stanu dla układu stacjonarnego ma postać:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \quad x(t_0) = x_0 \\ y(t) &= Cx(t) + Du(t) \end{aligned} \quad (3.5)$$

dla

$$\forall t \geq t_0 : x(t) \in R^n, u(t) \in R^r, y(t) \in R^m.$$

Macierze stanu A , sterowania B , obserwacji C i wyjścia D są stałe i mają wymiary: $[n \times n]$, $[n \times r]$, $[m \times n]$, $[m \times r]$, a fundamentalna macierz rozwiązań zależy tylko od jednego parametru, tzn. różnicy $\Delta = t - t_0$ i ma postać wykładniczą

$$\Phi(t, t_0) = \Phi(t - t_0) = e^{A(t-t_0)}.$$

Rozwiązanie $x(t)$ ma postać

$$x(t) = e^{A(t-t_0)}x(t_0) + \int_{t_0}^t e^{A(t-\tau)}B(\tau)u(\tau)d\tau \quad (3.6)$$

a funkcja wyjścia dana jest wzorem

$$y(t) = Ce^{A(t-t_0)}x(t_0) + C \int_{t_0}^t e^{A(t-\tau)}B(\tau)u(\tau)d\tau + Du(t) \quad (3.7)$$

Rozwiązanie nie zależy więc od bezwzględnych wartości chwili początkowej t_0 i końcowej t , ale od względnego parametru, jakim jest ich różnica $\Delta = t - t_0$. Stąd dla układu stacjonarnego (3.5) wyniki eksperymentów po czasie Δ będą zawsze takie same, bez względu na początkowy moment t_0 jego rozpoczęcia. Bez straty ogólności dla wyników, przyjmujemy więc, że $t_0 = 0$.

Rozwiązanie macierzowego równania różniczkowego (3.5) ma wtedy postać

$$x(t) = e^{At}x_0 + \int_0^t e^{A(t-\tau)}Bu(\tau)d\tau \quad (3.8)$$

Przykładowo dla $u(t) = 1(t)$ i przy założeniu niosobliwości A rozwiązanie ma postać

$$x(t) = e^{At}x_0 + A^{-1}[e^{At} - I]B = e^{At}[x_0 + A^{-1}B] - A^{-1}B.$$

Zasadniczym problemem obliczeniowym jest algorytm wyliczania wykładniczej postaci macierzy funkcyjnej e^{At} .

Można podać kilka metod wyliczenia macierzy e^{At} ; poniżej wymieniono pięć metod.

1) Z definicji szeregu wykładniczego:

$$e^{At} = \sum_{k=0}^{\infty} \frac{(At)^k}{k!} = I + At + \frac{A^2 t^2}{2} + \frac{A^3 t^3}{6} + \dots,$$

$$\text{np.: } A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}; e^{At} = \begin{bmatrix} e^t & 0 \\ 0 & e^{2t} \end{bmatrix}.$$

2) Z twierdzenia Sylwestera, wykorzystując wielomian interpolacyjny Lagrange'a $f(A) = e^{At}$.

$$\lambda_1 \neq \lambda_2 \neq \dots \neq \lambda_n, \quad f(A) = \sum_{i=1}^n f(\lambda_i) \prod_{j=1, j \neq i}^n \left[\frac{A - \lambda_j I}{\lambda_i - \lambda_j} \right] = \sum_{i=1}^n \alpha_i(t) \cdot A^{i-1},$$

$$\text{np.: } A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad \lambda_1 = 1, \lambda_2 = 3,$$

$$e^{At} = e^{3t} \frac{A - 1I}{3-1} + e^{1t} \frac{A - 3I}{1-3} = e^{2t} \begin{bmatrix} \text{cht} & \text{sht} \\ \text{sht} & \text{cht} \end{bmatrix} = e^{2t} \text{sht} \cdot A + e^{2t} (2e^{-t} - \text{cht}) \cdot I.$$

3) Jako odwrotne przekształcenie Laplace'a: (podrozdz. 4.2.1), np.:

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad e^{At} = \mathcal{L}^{-1}[(sI - A)^{-1}],$$

$$\begin{bmatrix} s-1 & 0 \\ 0 & s-2 \end{bmatrix}^{-1} = \frac{1}{(s-1)(s-2)} \begin{bmatrix} s-2 & 0 \\ 0 & s-1 \end{bmatrix} = \begin{bmatrix} \frac{1}{s-1} & 0 \\ 0 & \frac{1}{s-2} \end{bmatrix} \xrightarrow{\mathcal{L}^{-1}} \begin{bmatrix} e^t & 0 \\ 0 & e^{2t} \end{bmatrix}.$$

4) Przez transformację A do jej diagonalnej postaci Jordana J (rozdz. 19):

$$e^{At} = T e^{Jt} T^{-1}, \quad J = T^{-1} A T,$$

gdzie:

λ_i – wartości własne,

T_i – odpowiadające im wektory własne (główne) macierzy A ,

$T = [T_1 \ T_2 \ \dots \ T_n]$: modalna macierz transformacji A do postaci Jordana J .

Na przykład:

$$A = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}: \quad |sI - A| = \begin{bmatrix} \lambda & -1 \\ 2 & \lambda + 3 \end{bmatrix} = \lambda^2 + 3\lambda + 2,$$

$$\Delta = 9 - 8 = 1 \Rightarrow \lambda_{1,2} = \frac{-3 \pm 1}{2} = -2, -1, \quad (\text{wartości własne } A).$$

Wektor własny $T_1 = \begin{pmatrix} T_{11} \\ T_{12} \end{pmatrix}$ obliczamy z równania $|\lambda_1 I - A| T_1 = 0$.

$$\left[\begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} \right] \begin{pmatrix} T_{11} \\ T_{12} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \begin{cases} -2T_{11} - T_{12} = 0 \\ 2T_{11} + T_{12} = 0 \end{cases} \Rightarrow T_1 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}.$$

Wektor własny $T_2 = \begin{pmatrix} T_{21} \\ T_{22} \end{pmatrix}$ obliczamy z równania $|\lambda_2 I - A| T_2 = 0$.

$$\begin{bmatrix} -1 & -1 \\ 2 & 2 \end{bmatrix} \begin{pmatrix} T_{21} \\ T_{22} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow T_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad T = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix}, \quad T^{-1} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix},$$

$$T^{-1} A T = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} \cdot \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix} = \begin{bmatrix} -2 & 0 \\ 0 & -1 \end{bmatrix} = J,$$

$$e^{At} = T \cdot e^{Jt} \cdot T^{-1} = T \cdot \begin{bmatrix} e^{-2t} & 0 \\ 0 & e^{-t} \end{bmatrix} T^{-1} = \begin{bmatrix} -e^{-2t} - 2e^{-t}, & -e^{-2t} - e^{-t} \\ 2e^{-2t} - 2e^{-t}, & 2e^{-2t} - e^{-t} \end{bmatrix}.$$

5) Numeryczne wyliczanie e^{At} .

Jedną z szybszych i dokładnych metod numerycznego obliczania wartości macierzy e^{Ah} dla niewielkich wartości kroku h jest metoda aproksymacyjna wynikająca z metod Padégo (podrozdz. 4.3.5) [3DAE]

$$e^{Ah} \cong \left[I - Ah/2 + (Ah)^2/12 \right]^{-1} \cdot \left[I + Ah/2 + (Ah)^2/12 \right] \quad (3.9)$$

$$e^{At} \approx e^{Aih} = \left[e^{Ah} \right]^i = e^{Ah} \cdot e^{Ah} \cdot e^{Ah} \cdot \dots \cdot e^{Ah}$$

Dla asymptotycznie stabilnej macierzy A , mającej wszystkie wartości własne o ujemnych częściach rzeczywistych, macierz $[I - Ah/2 + (Ah)^2/12]$ dla każdej wartości kroku h ma wszystkie wartości własne dodatnie, a stąd jest nieosobliwa. Dla niestabilnej macierzy A zawsze można dobrą dostatecznie mały krok h tak, żeby zachodziła odwracalność macierzy, gdyż $\lim[I - Ah/2 + (Ah)^2/12] = I$ dla $h \rightarrow 0$. Dla asymptotycznie stabilnej macierzy A wszystkie wartości własne λ_i macierzy (3.9) leżą w kole jednostkowym $|\lambda_i| < 1$.

3.5. Transformacje liniowe zmiennych stanu

3.5.1. Transformacje macierzy stanu A układu MIMO do postaci Jordana

Dany jest układ wielowymiarowy (*Multi Input-Multi Output*):

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t), \quad x_0 = x(0) - \text{warunki początkowe} \\ x(t) &\in R^n, \quad u(t) \in R^r, \quad y(t) \in R^m, \quad \forall t \geq 0\end{aligned}\tag{3.10}$$

Jak pokazano w podrozdziale 3.4 rozwiązywanie równania (3.10) bywa łatwiejsze po jego liniowym przekształceniu do innej przestrzeni stanu. W tym celu można znaleźć nieosobliwą macierz transformacji podobieństwa T zbudowaną z wektorów własnych i głównych (macierz modalną), przekształcającą zmienne stanu x w inne zmienne stanu z

$$x(t) = Tz(t).$$

Wtedy:

$$\begin{aligned}\dot{z}(t) &= T^{-1}ATz(t) + T^{-1}Bu(t) = Jz(t) + \hat{B}u(t) \\ y(t) &= CTz(t) = \hat{C}z(t) \\ z_0 &= T^{-1}x_0\end{aligned}\tag{3.11}$$

Po podstawieniu $T^{-1}AT = J$, rozwiązanie (3.11) ma postać

$$z(t) = e^{Jt}z_0 + \int_0^t e^{J(t-\tau)}\hat{B}u(\tau)d\tau,$$

a rozwiązanie systemu (3.10)

$$x(t) = Te^{Jt}T^{-1}x_0 + \int_0^t Te^{J(t-\tau)}T^{-1}Bu(\tau)d\tau\tag{3.12}$$

Macierz J jest postacią diagonalną Jordana macierzy A , z jej wartościami własnymi na głównej diagonali. W przypadku wielokrotnej wartości własnej λ_i w macierzy J pojawia się blok odpowiadający wymiarowo krotności tej λ_i . W strukturze bloku na przekątnej mogą pojawić się klatki Jordana z jedynkami nad diagonalą. Macierz transformacji T składa się wtedy z wektorów własnych i głównych stwarzyszonych z daną wielokrotną wartością własną.

Zaletą wzoru (3.12) jest możliwość wykorzystania łatwiejszej do policzenia postaci macierzy e^{Jt} . Z wzorów (3.12) i (3.6) wynika zależność

$$e^{At} = Te^{Jt}T^{-1}\tag{3.13}$$

3.5. Transformacje liniowe zmiennych stanu

Wyjście $y(t)$ może być wyrażone wzorem

$$y(t) = CTe^{Jt}T^{-1}x_0 + C \int_0^t Te^{J(t-\tau)}T^{-1}Bu(\tau)d\tau\tag{3.14}$$

3.5.2. Równanie stanu dla układu SISO n -tego rzędu

Dane skalarne równanie różniczkowe n -tego rzędu na zmienną wyjściową $y(t)$ dla układu jednowymiarowego (*Single Input-Single Output*)

$$y^{(n)}(t) + a_{n-1}y^{(n-1)}(t) + \dots + a_1\dot{y}(t) + a_0y(t) = b_nu^{(n)}(t) + \dots + b_1\dot{u}(t) + b_0u(t)\tag{3.15}$$

z warunkami początkowymi:

$$\begin{aligned}y(0) &= y_0, & u(0) &= u_0, \\ \dot{y}(0) &= \dot{y}_0, & \dot{u}(0) &= \dot{u}_0, \\ & \vdots & & \vdots \\ y^{(n-1)}(0) &= y_0^{(n-1)}, & u^{(n-1)}(0) &= u_0^{(n-1)}.\end{aligned}$$

Równanie to można zamienić na n równań różniczkowych I rzędu. Wprowadza się wektor n liniowo niezależnych zmiennych stanu, które jednoznacznie opisują zachowanie układu w przyszłości i przechodzi się na macierzowy zapis układu za pomocą równań stanu i wyjścia:

$$x(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix} \in R^n, \quad y(t) \in R, \quad u(t) \in R\tag{3.16}$$

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t)\end{aligned}$$

gdzie:

- A – macierz stanu,
- B – macierz sterowania,
- C – macierz obserwacji,
- D – macierz wyjścia.

Istnieje nieskończoność wiele różnych zestawów zmiennych stanów dających opis tego samego układu. Powszechnie przyjmuje się naturalne zmienne stanu reprezentujące w układach mechanicznych przesunięcie, prędkość, przyspieszenie itd. Stąd kolejna zmienna stanu jest pochodną poprzedniej. Z takiego zapisu wynika charakterystyczna postać macierzy stanu A typu Frobeniusa z jedynkami nad główną (zerową) diagonalą. Dla takiej macierzy, modalną macierzą transformacji do postaci Jordana jest macierz Vandermonda.

Dalej przedstawiono sześć standardowych postaci naturalnych zmiennych stanu [3OGK], których strukturalne uzasadnienie można znaleźć w podrozdziale 4.3.3 (przy wykorzystaniu pojęcia transmitancji operatorowej).

W tym celu wprowadzamy oznaczenia:

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_n \end{bmatrix} = \begin{bmatrix} b_{n-1} - a_{n-1}b_n \\ b_{n-2} - a_{n-1}\beta_1 - a_{n-2}b_n \\ \dots \\ b_0 - a_{n-1}\beta_{n-1} - a_{n-2}\beta_{n-2} - \dots - a_1\beta_1 - a_0b_n \end{bmatrix},$$

$$\gamma = [\gamma_1 \ \gamma_2 \ \dots \ \gamma_n] = [b_0 - a_0b_n \ b_1 - a_1b_n \ \dots \ b_{n-1} - a_{n-1}b_n],$$

dla

$$b_n = 0 \Rightarrow \gamma = [b_0 \ b_1 \ \dots \ b_{n-1}].$$

Postacie I typu (tzw. X -sterowalne):

a)

$$A = \begin{bmatrix} 0 & 1 & \dots & 0 \\ 0 & 0 & 1 & \dots \\ \dots & 0 & 0 & 1 \\ -a_0 & -a_1 & \dots & -a_{n-1} \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 1 \end{bmatrix} \quad (3.17)$$

$$C = [\gamma_1 \ \gamma_2 \ \dots \ \gamma_n], \quad D = b_n$$

b)

$$A = \begin{bmatrix} 0 & 0 & 0 & -a_0 \\ 1 & 0 & 0 & -a_1 \\ \dots & 1 & \dots & \dots \\ \dots & \dots & \dots & -a_{n-1} \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 0 \\ \dots \\ 0 \end{bmatrix} \quad (3.18)$$

$$C = [\beta_1 \ \beta_2 \ \dots \ \beta_n], \quad D = b_n$$

c)

$$A = \begin{bmatrix} -a_{n-1} & -a_{n-2} & \dots & -a_0 \\ 1 & \dots & \dots & \dots \\ \dots & 1 & \dots & \dots \\ \dots & \dots & \dots & 1, 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 0 \\ \dots \\ 0 \end{bmatrix} \quad (3.19)$$

$$C = [\gamma_n \ \gamma_{n-1} \ \dots \ \gamma_1], \quad D = b_n$$

Postacie II typu (tzw. obserwowańe):

a)

$$A = \begin{bmatrix} 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ \dots & \dots & \dots & 1 \\ -a_0 & -a_2 & \dots & -a_{n-1} \end{bmatrix}, \quad B = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_n \end{bmatrix} \quad (3.20)$$

$$C = [1 \ 0 \ \dots \ 0], \quad D = b_n$$

b)

$$A = \begin{bmatrix} -a_{n-1} & 1 & 0 & \dots \\ -a_{n-2} & 0 & 1 & \dots \\ \dots & \dots & \dots & 1 \\ -a_0 & \dots & \dots & 0 \end{bmatrix}, \quad B = \begin{bmatrix} \gamma_n \\ \gamma_{n-1} \\ \dots \\ \gamma_1 \end{bmatrix} \quad (3.21)$$

$$C = [1 \ 0 \ \dots \ 0], \quad D = b_n$$

c)

$$A = \begin{bmatrix} 0 & 0 & 0 & -a_0 \\ 1 & 0 & 0 & -a_1 \\ \dots & 1 & \dots & \dots \\ \dots & \dots & 1 & -a_{n-1} \end{bmatrix}, \quad B = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \dots \\ \gamma_n \end{bmatrix} \quad (3.22)$$

$$C = [0 \ 0 \ \dots \ 1], \quad D = b_n$$

Przechodząc ze skalarnego równania n -tego rzędu (3.15) na układ równań (3.16), należy dla każdej postaci zmiennych stanu dodatkowo wyliczyć odpowiadające jej warunki początkowe.

Przykładowo dla postaci IIa):

$$x_1(0) = y(0) - b_n u(0),$$

$$x_2(0) = \dot{y}(0) - b_n \dot{u}(0) - \beta_1 u(0),$$

$$\vdots$$

$$x_n(0) = y^{(n-1)}(0) - b_n u^{(n-1)}(0) - \dots - \beta_{n-1} u(0).$$

3.5.3. Inna postać równania stanu

Przedstawiona zostanie jeszcze jedna postać macierzowego równania stanu dla skalarnego równania (3.15). Jest to postać nietypowa i stąd rzadko używana (stosowana np. w metodach sterowania z odwrotną dynamiką – *inverse dynamic* [3LAS]). Autor wykorzystywał ją w zagadnieniach identyfikacji parametrycznej układów ciągłych [3BYW1] oraz w zagadnieniach obserwacji stanu [3BYW2].

Dla równania stanu (3.16) można wykazać, że poprzez wielokrotne różniczkowanie wzoru na wyjście układu y i wykorzystanie równania stanu otrzymuje się nowe formuły na stan x , liczone za pomocą pochodnych wyjścia i wejścia:

$$\begin{aligned}y &= Cx + Du, \\ \dot{y} &= C\dot{x} = CAx + CBu + D\dot{u}, \\ \ddot{y} &= CA\dot{x} + CB\dot{u} = CA^2x + CABu + CB\dot{u} + D\ddot{u}, \\ &\vdots\end{aligned}$$

Grupując macierzowo powyższe wzory, otrzymuje się np. dla $n = 4$ postać

$$\tilde{y} = \begin{bmatrix} y \\ \dot{y} \\ \ddot{y} \\ \dddot{y} \end{bmatrix} = \begin{bmatrix} C \\ CA \\ CA^2 \\ CA^3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} D & 0 & 0 & 0 \\ CB & D & 0 & 0 \\ CAB & CB & D & 0 \\ CA^2B & CAB & CB & D \end{bmatrix} \begin{bmatrix} u \\ \dot{u} \\ \ddot{u} \\ \dddot{u} \end{bmatrix} = [Q_0]x + [H_0]\tilde{u}.$$

Stąd dla systemów obserwacyjnych otrzymuje się za pomocą macierzy Q_0 (obserwacyjności) i macierzy trójkątnej H_0 liniową postać równania na stan x , równoważną w sensie dystrybucyjnym postaci (3.16).

$$x = [Q_0^{-1}] \tilde{y} - [Q_0^{-1}] [H_0] \tilde{u} = [H_1] \tilde{y} - [H_2] \tilde{u}.$$

Wektory \tilde{y} i \tilde{u} zawierają kolejne pochodne sygnału wyjścia y i wejścia u . Aby wykorzystać taką postać równania stanu, pochodne powinny być dostępne do pomiaru. W praktyce wykorzystuje się jednak filtry splotowe z nośnikami zwartymi, które pozwalają otrzymać transformacje kolejnych pochodnych y i u z bezpośrednich pomiarów funkcji y i u . Równanie powyższe przekształca się wówczas w równanie algebraiczne [3BYW2]. W zależności od przyjętej postaci macierzy A, B, C, D (wzory (3.20), (3.21), (3.22)) otrzymuje się różne końcowe postacie H_1, H_2 (H_2 – macierz prostokątna).

Przykładowo można pokazać, że dla obserwacyjnej postaci (3.22) i $D = b_n = 0$ otrzymuje się

$$x(t) = H_1 \begin{bmatrix} y(t) \\ \dot{y}(t) \\ \vdots \\ y^{(n-1)}(t) \end{bmatrix} - H_2 \begin{bmatrix} u(t) \\ \dot{u}(t) \\ \vdots \\ u^{(n-2)}(t) \end{bmatrix},$$

gdzie macierze H_1 i H_2 przyjmują charakterystyczne postacie macierzy Toepliza:

$$H_1 = \begin{bmatrix} a_1 & \dots & a_{n-1} & 1 \\ a_2 & \dots & 1 & 0 \\ \vdots & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \quad H_2 = \begin{bmatrix} b_1 & \dots & b_{n-2} & b_{n-1} \\ b_2 & \dots & b_{n-1} & 0 \\ \vdots & b_{n-1} & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

3.5.4. Transformacje dowolnej macierzy stanu A układu SISO do postaci Frobeniusa F

Poniżej przedstawiono cztery charakterystyczne zestawy tych przekształceń.

- I) Dla systemu (3.7) typu SISO z dowolną macierzą stanu A (para (A, B) sterowalna), może istnieć nieosobliwa macierz transformacji T przekształcająca tę macierz A w macierz o postaci Frobeniusa, a cały system w system typu Ia) (3.17):

$$x(t) = Tz(t),$$

$$\dot{z}(t) = T^{-1}ATz(t) + T^{-1}Bu(t) = Fz(t) + B_Fu(t),$$

$$y(t) = CTz(t) = C_Fz(t),$$

$$F = T^{-1}AT = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \vdots \\ \vdots & 0 & \ddots & & \vdots \\ & & & 1 & \\ -a_0 & -a_1 & \dots & & -a_{n-1} \end{bmatrix},$$

$$\det[A - sI] = \det[F - sI] = s^n + a_{n-1}s^{n-1} + \dots + a_1s + a_0,$$

$$B_F = T^{-1}B = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}, \quad C_F = [b_0, b_1 \dots b_{n-1}].$$

Macierz transformacji ma postać

$$T = Q_s M,$$

gdzie:

$$\text{macierz } Q_s = [B \ AB \ A^2B \ \dots \ A^{n-1}B],$$

M – macierz Toeplitza zbudowana ze współczynników równania charakterystycznego macierzy A

$$M = \begin{bmatrix} a_1 & a_2 & \dots & a_{n-1} & 1 \\ a_2 & a_3 & \dots & 1 & 0 \\ \vdots & & \dots & & \vdots \\ a_{n-1} & 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & & 0 \end{bmatrix}.$$

Nieosobliwość macierzy Q_s , a więc odwracalność macierzy T jest zawsze zagwarantowana dla tzw. sterowej pary (A, B) (rozdz. 8).

- 2) Dla systemu (3.7) typu SISO z dowolną macierzą stanu A , może istnieć nieosobliwa macierz transformacji T przekształcająca tą macierz A w macierz o transponowanej postaci Frobeniusa, a cały system w system typu IIc) (3.22):

$$x(t) = Tz(t),$$

$$\dot{z}(t) = T^{-1}ATz(t) + T^{-1}Bu(t) = F^T z(t) + B_F u(t),$$

$$y(t) = CTz(t) = C_F z(t),$$

$$F^T = T^{-1}AT = \begin{bmatrix} 0 & 0 & 0 & \dots & -a_0 \\ 1 & 0 & 0 & 0 & -a_1 \\ \vdots & 1 & \ddots & & \vdots \\ \vdots & & & & \\ 0 & \dots & 1 & & -a_{n-1} \end{bmatrix},$$

$$\det[A - sI] = \det[F^T - sI] = s^n + a_{n-1}s^{n-1} + \dots + a_1s + a_0,$$

$$B_F = T^{-1}B = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{n-1} \end{bmatrix}, \quad C_F = CT = [0, 0, \dots, 0, 1].$$

Macierz transformacji ma postać

$$T = (MQ_0)^{-1} \quad \text{lub} \quad T^{-1} = MQ_0,$$

gdzie:

$$Q_0 = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix},$$

M – macierz Toeplitza.

Nieosobliwość macierzy Q_0 , a więc odwracalność macierzy T jest zawsze zagwarantowana dla tzw. obserwowlanej pary (A, C) (rozdz. 8).

- 3) Jeżeli para (A, B) jest sterowalna, to stosując do systemu (3.16) prostszą transformację $T = Q_s$, otrzymuje się sterowalną postać typu Ib) (3.18):

$$x(t) = Tz(t),$$

$$\dot{z}(t) = T^{-1}ATz(t) + T^{-1}Bu(t) = F^T z(t) + B_F u(t),$$

$$y(t) = CTz(t) = C_F z(t).$$

Macierz transformacji $T = Q_s = [B, AB, \dots, A^{n-1}B]$,

$$F^T = Q_s^{-1}AQ_s = \begin{bmatrix} 0 & 0 & 0 & \dots & -a_0 \\ 1 & 0 & 0 & 0 & -a_1 \\ \vdots & 1 & \ddots & & \vdots \\ \vdots & & & & \\ 0 & \dots & 1 & & -a_{n-1} \end{bmatrix}, \quad B_F = T^{-1}B = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = Q_s^{-1}B,$$

$$C_F = CT = CQ_s = [\beta_1, \beta_2, \dots, \beta_n], \quad \beta_i - \text{ze wzorów (3.16).}$$

- 4) Jeżeli para (A, C) jest obserwowlana, to stosując dla systemu (3.10) transformację $T = Q_0^{-1}$, otrzymuje się obserwowlaną postać typu IIa) (3.20) z macierzą Frobeniusa F .

$$x(t) = Tz(t),$$

$$\dot{z}(t) = T^{-1}ATz(t) + T^{-1}Bu(t) = Fz(t) + B_F u(t),$$

$$y(t) = CTz(t) = C_F z(t),$$

$$F = T^{-1}AT = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & & 1 \\ -a_0 & -a_1 & \dots & & -a_{n-1} \end{bmatrix} = Q_0AQ_0^{-1},$$

$$B_F = T^{-1}B = Q_0B = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}, \quad C_F = CT = CQ_0^{-1} = [1, 0, 0, \dots, 0].$$

Macierz transformacji T ma postać

$$T = Q_0^{-1} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}^{-1}.$$

Z postaci 3) i 4) wynika przy okazji zależność $CQ_s = [Q_0B]^T$.

Przykład 3.2

Należy przetransformować podany poniżej system do postaci IIc):

$$A = \begin{bmatrix} -2 & 1 \\ -2 & -5 \end{bmatrix}, \quad B = \begin{bmatrix} 4 \\ 3 \end{bmatrix}, \quad C = [2 \ 3],$$

$$|\lambda I - A| = \begin{vmatrix} \lambda + 2 & -1 \\ 2 & \lambda + 5 \end{vmatrix} = \lambda^2 + 7\lambda + 2 = 0,$$

$$Q_0 = \begin{bmatrix} 2 & 3 \\ -10 & -13 \end{bmatrix}, \quad M = \begin{bmatrix} 7 & 1 \\ 1 & 0 \end{bmatrix},$$

$$T^{-1} = MQ_0 = \begin{bmatrix} 4 & 8 \\ 2 & 3 \end{bmatrix}, \quad T = (MQ_0)^{-1} = \frac{1}{4} \begin{bmatrix} -3 & 8 \\ 2 & -4 \end{bmatrix},$$

$$T^{-1}AT = \begin{bmatrix} 0 & -12 \\ 1 & -7 \end{bmatrix}, \quad B = \begin{bmatrix} 4 & 8 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 40 \\ 17 \end{bmatrix},$$

$$CT = \frac{1}{4} [2 \ 3] T = [0 \ 1].$$

3.6. Splot i jego zastosowanie

Znane z matematyki pojęcie splotu znajduje zastosowanie w wyliczaniu odpowiedzi układu liniowego i stacjonarnego na dowolne wymuszenie. Od układu wymaga się cechy przyczynowości lub inaczej realizowalności. Cecha ta gwarantuje, że odpowiedź systemu na dane sterowanie nie wyprzedzi momentu wystąpienia samego sterowania.

Odpowiedź $y(t)$ takiego systemu na specjalny sygnał sterujący będący impulsem Diraca $u(t) = \delta(t)$ oznacza się

$$y(t) = a \cdot g(t).$$

Funkcja $g(t)$ nazywa się **impulsową funkcją przejścia** układu liniowego. Zastosowanie impulsu Diraca, a nie delty Diraca, pozwala na bardziej realistyczne rozważania i dostosowanie za pomocą współczynnika a jednostek, np. $u(t)$ [V], $y(t)$ [mA], t [s], $g(t)$ [mA/(V·s)], a [V·s], $\delta(t)$ [1/s].

Wykorzystując własność filtrującą delty Diraca i założenie, że sygnał wejściowy $u(t)$ można przedstawić jako sumę kolejno występujących impulsów Diraca, oraz zasadę super-

3.6. Splot i jego zastosowanie

pozycji, można wykazać, że wyjście $y(t)$ z dowolnego liniowego, stacjonarnego układu na sterowanie $u(t)$ jest opisywane za pomocą całki splotowej

$$y(t) = \int_{-\infty}^{\infty} g(t-\tau) u(\tau) d\tau = g * u \quad (3.23)$$

Liniowy i stacjonarny układ jest realizowalny wtedy i tylko wtedy, gdy impulsowa funkcja $g(t)$ równa się zero dla $t < 0$, co ma uzasadnienie w przyczynowości odpowiedzi na $\delta(t)$. Wtedy $g(t-\tau) = 0$ dla $\tau > t$ i całka splotowa upraszcza się do

$$y(t) = \int_{-\infty}^t g(t-\tau) u(\tau) d\tau.$$

Zamiana zmiennych pokazuje też inną postać całki splotowej

$$y(t) = \int_{-\infty}^t g(t-\tau) u(\tau) d\tau = \int_0^{\infty} g(\tau) u(t-\tau) d\tau.$$

Jeśli przyjęte zostanie dodatkowe założenie, że sygnał sterujący $u(t) = 0$ dla $t < 0$, to całka splotowa uzyska jeszcze prostszą postać

$$y(t) = \int_0^t g(t-\tau) u(\tau) d\tau = \int_0^t g(\tau) u(t-\tau) d\tau \quad (3.24)$$

Jeśli powyższe założenie o sterowaniu nie może być przyjęte, to postać ogólną (3.23), ze względu na superpozycję, można zawsze przedstawić jako

$$y(t) = \int_{-\infty}^0 g(t-\tau) u(\tau) d\tau + \int_0^t g(t-\tau) u(\tau) d\tau = y(0) + \int_0^t g(t-\tau) u(\tau) d\tau.$$

Stąd rozwiązanie układu może być rozważane tylko dla czasów dodatnich $t \geq 0$ od umownego momentu $t = 0$ pod warunkiem, że w chwili $t = 0$ będzie uwzględniony pierwszy człon zwany warunkiem początkowym $y(0)$.

Dla zerowych warunków początkowych stanu $x_0 = 0$, rozwiązanie na zmienną wyjściową $y(t)$ systemu (3.10) ma postać

$$y(t) = \int_0^t CT e^{J(t-\tau)} T^{-1} B u(\tau) d\tau = \int_0^t C e^{A(t-\tau)} B u(\tau) d\tau = \int_0^t g(t-\tau) u(\tau) d\tau,$$

gdzie impulsowa funkcja przejścia systemu (3.10) jest dana wzorem

$$g(t) = C e^{At} B.$$

Dla systemów jednowymiarowych ($r = 1, m = 1$) $g(t)$ jest funkcją skalarną, a dla systemów wielowymiarowych jest funkcją macierzową o wymiarach $[m \times r]$.

Zastosowanie w chwili $t = 0$, dla systemu (3.10) z warunkami początkowymi $x_0 = 0$, teoretycznego sterowania w postaci delty Diraca $u(t) = \delta(t)$ daje jego odpowiedź w postaci

$$y(t) = Ce^{At}B = g(t) \quad (3.25)$$

Wynika to z filtrującej własności delty Diraca i ze wzoru (3.23).

Wzór (3.24) przedstawia specjalny typ operacji na dwóch funkcjach $g(t)$ i $u(t)$, zwany splotaniem tych funkcji. Wynika z niego, że odpowiedź systemu dynamicznego typu (3.5) i (3.10) na dowolny sygnał sterujący $u(\tau)$ można wyliczyć, znając to sterowanie $u(\tau)$ i impulsową funkcję systemu $g(\tau)$ w przedziale $[0, t]$. Funkcje te należy w odpowiedni sposób pomnożyć między sobą i całkować w tym przedziale.

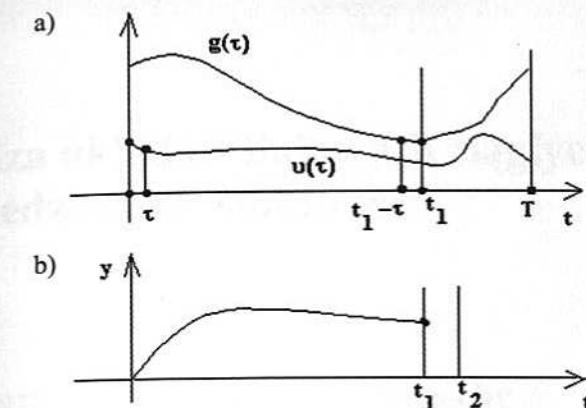
Ze wzoru $g(t) = Ce^{At}B$ wynika, że impulsowa funkcja przejścia może być policzona analitycznie ze znajomości modelu stanu (3.6) lub może być zjęta eksperymentalnie jako wyjście obiektu rzeczywistego $y(t)$ na sygnał $u(t) = a\delta(t)$. W obu przypadkach przebieg $g(t)$ można zapamiętać w komputerze na odpowiednio długim odcinku.

Zastosowanie wzoru (3.24) do numerycznego obliczania odpowiedzi systemu $y(t_1)$ w chwili $t_1 \in [0, T]$, gdy cały przebieg $g(t)$ dla $t \in [0, T]$ jest zarejestrowany jako zbiór dyskretnych wartości (w ilości gwarantującej żądaną dokładność) i znany jest zbiór dyskretnych sterowań $u(t)$ dla $t \in [0, T]$, wiąże się z numerycznym całkowaniem w bieżącym przedziale $[0, t_1]$ iloczynu dwóch funkcji $g(\cdot)$ i $u(\cdot)$. Otrzymuje się go w specjalny sposób, tzn. jako iloczyn początkowej wartości sterowania z końcową wartością funkcji impulsowej $u(0) \cdot g(t_1)$, a następnie $u(\tau) \cdot g(t_1 - \tau), \dots, u(t_1) \cdot g(0)$, dla $0 \leq \tau \leq t_1$, dając w efekcie wartość $y(t_1)$ (rys. 3.4). Zwiększenie bieżącego przedziału do $t_2 > t_1$ ($t_2 < T$) wymaga powtórzenia całej procedury iloczynów splotowych $u(0) \cdot g(t_2)$, a następnie $u(\tau) \cdot g(t_2 - \tau), \dots, u(t_2) \cdot g(0)$ dla $0 \leq \tau \leq t_2$ i całkowania wyniku w przedziale $[0, t_2]$, dając w efekcie wartość $y(t_2)$. Stosowanie tej procedury w czasie rzeczywistym, przy *on-line* rejestracyjnych próbkach pomiarowych sterowania, wymaga komputera, który zdąży wykonywać na bieżąco rosnącą w jednostce czasu liczbę obliczeń (mnożeń i sumowań) (rys. 3.4).

Głębokie uzasadnienie użycia zapisu splotowego dla liczenia wyjścia systemu liniowego stacjonarnego wynika z ważnego w matematyce twierdzenia Riesza o reprezentacji w przestrzeni C_0 , mówiącego, że każdy funkcjonal (operator) liniowy, przyczynowy i stacjonarny przekształcający przestrzeń funkcji ciągłych, ograniczonych $C_0[0, \infty)$ i równych tożsamościowo zero dla $t \leq 0$, w siebie (wejście i wyjście systemu należy do C_0), musi mieć postać całki Stieltjesa

$$y(t) = F(u) = \int_0^t u(t - \tau) dh(\tau),$$

gdzie $h(\tau)$ jest odpowiedzią układu na skok jednostkowy i jest funkcją o wahaniu ograniczonym (rozdz. 20) [3KUJ].



Rys. 3.4. Przebieg funkcji g, u (a); wynik splotu (b)

W innych przestrzeniach wzór (3.24) określający splot ma następujące własności:

- jeśli $g, u \in L^1(0, \infty)$, to $(g^*u) \in L^1(0, \infty)$;
- jeśli $g \in L^1(0, \infty)$, $u \in L^2(0, \infty)$, to $(g^*u) \in L^2(0, \infty)$;
- jeśli $g \in L^1(0, \infty)$, $u \in L^p(0, \infty)$, to $(g^*u) \in L^p(0, \infty)$.

4. Analiza układów liniowych ciągłych w dziedzinie częstotliwości

4.1. Transformata Laplace'a i jej cechy

4.1.1. Uwagi ogólne

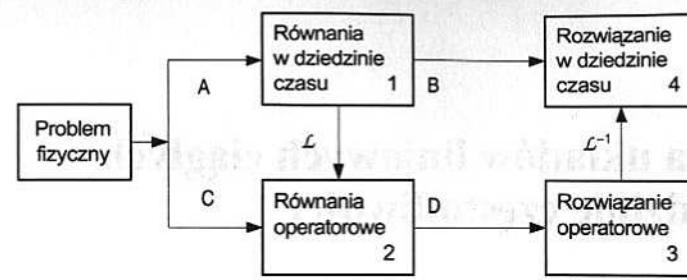
Modele liniowe opisywane stacjonarnymi równaniami różniczkowymi mogą być rozwiązywane różnymi technikami. Jedną z nich jest rachunek operatorowy [4OSJ]. Wprowadził go O. Heaviside w roku 1899 w swych pracach na temat teorii elektrotechniki, bez głębszych uzasadnień matematycznych. W rachunku tym z reguły różniczka jest traktowana jako specjalny operator D działający na funkcję $x(t)$. Jego n -ta potęga reprezentuje n -ty rząd różniczki, a n -krotny operator odwrotny reprezentuje n -krotną całkę:

$$D^n = \frac{d^n}{dt^n}, \quad \frac{1}{D^n} = \int \int \int.$$

Takie podstawienie zamienia równanie różniczkowe w równanie algebraiczne, które można rozwiązać prostymi przekształceniami. Ostatecznie otrzymuje się operację wynikową na funkcji wymuszającej $u(t)$, w postaci ułamka wymiernego zmiennej D . Rozwiniecie ułamka w szereg potęgowy (geometryczny) umożliwia znalezienie rozwiązania. Powiązanie metody operatorowej z przekształceniem całkowym Laplace'a, znany w matematyce od 1812 r., pozwoliło w latach 1920–1930 na budowę precyzyjnego aparatu metody operatorów Laplace'a. Powstały obszerne tablice podające inżynierom gotowe transformaty Laplace'a dla wielu typów funkcji czasu. Umożliwiają one szybką zamianę równania różniczkowego na algebraiczne, jego rozwiązanie w dziedzinie zmiennej zespolonej s i ponowną zamianę tego rozwiązania na postać czasową funkcji oryginalnej. Dokonuje się tego np. poprzez rozkład operacji wynikowej zmiennej s na sumę ułamków prostych i odnalezienie w tablicach gotowych wzorów na ich odwzorowania odwrotne. Daje to w konsekwencji pełną postać rozwiązania w dziedzinie czasu.

Miejsce techniki operatorowej w procedurze uzyskiwania rozwiązania (odpowiedź w funkcji sterowania) widać na rysunku 4.1.

Przez \mathcal{L} oznacza się transformację Laplace'a, a \mathcal{L}^{-1} odwrotną transformację Laplace'a.



Rys. 4.1. Schematy rozwiązań

Dla procesu fizycznego PF możliwa jest różna metodyka budowy modelu liniowego i różne drogi dojścia do postaci rozwiązań z PF → 4.

- 1) PF $\xrightarrow{A} 1 \xrightarrow{B} 4$ (metoda bezpośrednią),
- 2) PF $\xrightarrow{A} 1 \xrightarrow{\mathcal{L}} 2 \xrightarrow{D} 3 \xrightarrow{\mathcal{L}^{-1}} 4$ (metoda pośrednia),
- 3) PF $\xrightarrow{C} 2 \xrightarrow{D} 3 \xrightarrow{\mathcal{L}^{-1}} 4$ (metoda pośrednia skrócona).

4.1.2. Własności i przykłady transformat Laplace'a

Dla funkcji czasu $f(t)$ definiuje się jednostronną całkową transformację $\mathcal{L}[f(t)]$

$$\mathcal{L}[f(t)] = F(s) = \int_0^\infty f(\tau) e^{-st} d\tau,$$

która danej funkcji $f(t)$ zmiennej rzeczywistej t przyporządkowuje inną funkcję $F(s)$ zmiennej zespolonej $s = \alpha + j\omega$, pełniącej przy całkowaniu rolę parametru. Aby transformata $F(s)$ istniała, musi istnieć dla danej funkcji $f(t)$ przynajmniej jedno s (lub zbiór wielu s), dla których taka całka istnieje, tzn. jest mniejsza od ∞ .

$$\left| \int_0^\infty e^{-st} f(t) dt \right| < \infty.$$

Dzięki wprowadzeniu eksponenty powyższa całka istnieje dla szerokiej klasy funkcji, np. dla funkcji $f(t)$ przedziałami ciągłymi dla $0 < t < \infty$ oraz ogólnie – dla funkcji wykładniczych, czyli takich, dla których istnieją ρ i $M > 0$ takie, że dla $t \geq 0$ $|f(t)| \leq M e^{\rho t}$. Stąd np. dla funkcji $f(t) = e^{t^2}$, która rośnie szybciej niż e^{-st} maleje, całka nie jest zbieżna.

Przykłady obliczonych transformat pokazano w tabeli 4.1.

Tabela 4.1

Przykładowe \mathcal{L} -transformaty

Oryginal $f(t)$	Transformata $F(s)$
$1(t)$	$1/s$
$1(t) e^{-at}$	$1/(s + a)$
$\frac{1}{a}(1 - e^{-at})$	$\frac{1}{s(s + a)}$
$1(t) \frac{t^{n-1}}{(n-1)!}$	$\frac{1}{s^n}$
$1(t) t e^{-at}$	$1/(s + a)^2$
$1(t) t^n e^{-at}$	$\frac{n!}{(s + a)^{n+1}}$
$1(t) [1 - e^{-t}]^n$	$\frac{n!}{s(s+1)(s+2)\dots(s+n)}$
$1(t) \sin(\omega t)$	$\frac{\omega}{s^2 + \omega^2}$
$1(t) \cos(\omega t)$	$\frac{s}{s^2 + \omega^2}$
$1(t) e^{-at} \sin(\omega t)$	$\frac{\omega}{(s + a)^2 + \omega^2}$
$e^{-bt} \sinh(at)$	$\frac{a}{(s + b)^2 - a^2}$
$1 + \frac{1}{T_2 - T_1} \left(T_1 e^{-\frac{t}{T_1}} - T_2 e^{-\frac{t}{T_2}} \right)$	$\frac{1}{s(T_1 s + 1)(T_2 s + 1)}$
	$\frac{a}{s} \operatorname{tgh}\left(\frac{sl}{2}\right)$
	$\frac{a}{s} \left\{ \frac{1}{sl} - \frac{1}{\sinh(sl)} \right\}$

Przykłady obliczania transformat Laplace'a

Poniżej pokazane są trzy przykłady obliczania transformat Laplace'a:

1) skok jednostkowy

$$\mathcal{L}[1(t)] = F(s) = \int_{0^-}^{\infty} e^{-st} 1(t) dt = \frac{e^{-st}}{-s} \Big|_{0^-}^{\infty} = \frac{1}{s},$$

2) impuls Diraca

$$\mathcal{L}[\delta(t)] = \int_{0^-}^{\infty} \delta(t) e^{-st} dt = 1,$$

3) transformata pochodnej

$$\mathcal{L}\left[\frac{d}{dt} f(t)\right] = \int_{0^-}^{\infty} \dot{f}(t) e^{-st} dt = f(t) e^{-st} \Big|_{0^-}^{\infty} + s \int_{0^-}^{\infty} f(t) e^{-st} dt = -f(0^-) + sF(s).$$

Przekształcenie Laplace'a posiada wszystkie cechy przekształcenia liniowego. Jeśli funkcji $f_1(t)$ odpowiada transformata $F_1(s)$, a funkcji $f_2(t)$ odpowiada transformata $F_2(s)$, to własności przekształcenia Laplace'a można przedstawić w punktach tabeli 4.2.

Ze wszystkich zebranych w tej tabeli własności zwłaszcza dwie są szczególnie ważne – nr 5 i nr 8.

Tabela 4.2

Własności transformaty Laplace'a

1	Liniowość	$af_1(t) + bf_2(t) \Leftrightarrow aF_1(s) + bF_2(s)$
2	Zmiana skali	$f(a \cdot t) \Leftrightarrow \frac{1}{ a } F\left(\frac{s}{a}\right)$
3	Przesunięcie w czasie	$f(t - \tau) \Leftrightarrow F(s)e^{-s\tau}$
4	Przesunięcie w dziedzinie częstotliwości	$e^{-at} f(t) \Leftrightarrow F(s+a)$
5	Splot w dziedzinie czasu	$f_1(t) * f_2(t) = \int_{-\infty}^{\infty} f_1(\tau) \cdot f_2(t-\tau) d\tau$ $\Leftrightarrow F_1(s) \cdot F_2(s)$
6	Splot w dziedzinie częstotliwości	$f_1(t) \cdot f_2(t) \Leftrightarrow \frac{1}{2\pi j} \int_{-j\infty}^{+j\infty} F_1(u) \cdot F_2(s-u) du$
7	Jednostronne różniczkowanie w dziedzinie czasu	$\frac{df(t)}{dt} \Leftrightarrow sF(s) - f(0^-)$

Tabela 4.2 cd.

8	Jednostronne różniczkowanie w dziedzinie czasu dla zerowych warunków początkowych i dla niezerowych warunków początkowych	$\frac{d^n f(t)}{dt^n} \Leftrightarrow s^n F(s)$
9	Różniczkowanie w dziedzinie częstotliwości	$(-t^n) \cdot f(t) \Leftrightarrow \frac{d^n F(s)}{ds^n}$
10	Całkowanie w dziedzinie czasu	$\int_{-\infty}^t f(\tau) d\tau \Leftrightarrow \frac{F(s)}{s}$
11	Całkowanie jednostronne	$\int_{0^-}^t f(\tau) d\tau \Leftrightarrow \frac{F(s)}{s} + \frac{1}{s} \left[\int f(t) dt \right]_{t=0^-}$
12	Całka oznaczona	$\int_{-\infty}^{\infty} f(t) dt = \lim_{s \rightarrow 0} F(s)$
13	Jeśli istnieje stała wartość graniczna $f(\infty)$, to istnieje też granica $\lim_{s \rightarrow 0} sF(s)$ i zachodzi równość	$\lim_{s \rightarrow 0} sF(s) = f(\infty)$
	Twierdzenie odwrotne jest nieprawdziwe, bo np. istnieje granica dla $F(s) = \frac{1}{s^2 + 1}$; $\lim_{s \rightarrow 0} sF(s) = 0$, a $f(t) = \sin(t)$ i wartość ustalona $f(\infty)$ nie istnieje	
14	Jeśli istnieje początkowa wartość graniczna oryginału dla $t \rightarrow 0^+$, to zachodzi równość	$\lim_{s \rightarrow \infty} sF(s) = f(0^+)$ $\lim_{s \rightarrow \infty} s^2 F(s) = \dot{f}(0^+)$
15	Twierdzenie Rayleigha	$\int_{-\infty}^{+\infty} f^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(j\omega) ^2 d\omega$

4.2. Odwrotna transformata Laplace'a i jej zastosowanie**4.2.1. Odwrotna transformata Laplace'a**

W podrozdziałach poprzednich omawiano własności i obliczanie transformat Laplace'a dla danej funkcji czasu. Zasadnicze znaczenie w zastosowaniach ma również zagadnienie odwrotne: jak przy danej funkcji zmiennej zespolonej $F(s)$ wyznaczyć $f(t)$, dla której $F(s)$ jest \mathcal{L} -transformata.

Problem transformacji odwrotnej sprowadza się do rozwiązyania równania całkowego Laplace'a

$$F(s) = \int_0^\infty f(\tau) e^{-s\tau} d\tau,$$

a poszukiwana funkcja $f(t)$ będzie rozwiązaniem transformacji odwrotnej Laplace'a

$$f(t) = \mathcal{L}^{-1}[F(s)].$$

Formalnie istotną sprawą jest jednoznaczność rozwiązyania zagadnienia odwrotnego. Jak można udowodnić [4OSJ], rozwiązanie to jest jednoznaczne z dokładnością do funkcji zerowej, czyli takiej funkcji, która może mieć wartość niezerową tylko w swoich punktach nieciągłości. W praktyce ciągłe funkcje czasu różniące się funkcją zerową, czyli równe sobie prawie wszędzie, w wielu zagadnieniach fizyki są sobie równoważne i mogą być traktowane jako ta sama funkcja. Przy takiej umowie odwrotne przekształcenie Laplace'a staje się jednoznaczne. Postać analityczną dla operatora przekształcenia odwrotnego podaje wtedy wzór Riemanna–Mellina

$$f(t) = \mathcal{L}^{-1}[F(s)] = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} F(s) e^{st} ds.$$

Jest on prawdziwy prawie wszędzie dla $t > 0$ i ogólnie dla funkcji $f(t)$ o ograniczonym wahaniu (zmienności) w każdym przedziale $[0, T]$ dla $T > 0$, a liczba c związana jest z obszarem zbieżności całki Laplace'a.

4.2.2. Rozwiązywanie równań różniczkowych techniką operatorową

Praktyczne obliczanie transformacji odwrotnej przeprowadza się, korzystając z rachunku residiów dla wyliczenia całki we wzorze Riemanna–Mellina.

Dla funkcji wymiernej $F(s)$ o współczynnikach rzeczywistych i stopniu licznika mniejszym od stopnia mianownika, prostszą metodą jest jednak rozkład na ułamki proste i zastosowanie przekształcenia odwrotnego do każdego ułamka z osobna. Zastosowanie tej metody przedstawiono na przykładach.

Przykład 4.1

Znaleźć rozwiązanie równania trzeciego rzędu

$$\ddot{y}(t) + 3\dot{y}(t) + 6\dot{y}(t) + 4y(t) = \dot{u}(t) + 3u(t) \quad (4.1)$$

Warunki początkowe zerowe: $y(0) = 0, \dot{y}(0) = 0, \ddot{y}(0) = 0$.

Funkcja wymuszająca $u(t) = \delta(t)$ dla $t \geq 0$.

Po zastosowaniu do równania (4.1) transformacji Laplace'a otrzymuje się

$$Y(s)(s^3 + 3s^2 + 6s + 4) = (s+3)U(s) \quad (4.2)$$

gdzie $U(s) = 1$.

Dla rozwiązyania równania (4.2) należy zastosować rozkład na ułamki proste:

$$Y(s) = \frac{s+3}{(s^2 + 2s + 4) \cdot (s+1)} = \frac{s+3}{[(s+1)^2 + 3] \cdot (s+1)} = \frac{Bs+C}{s^2 + 2s + 4} + \frac{A}{s+1},$$

$$s+3 \equiv (s+1)(Bs+C) + A(s^2 + 2s + 4),$$

$$s+3 \equiv Bs^2 + Cs + Bs + C + As^2 + 2As + 4A,$$

$$s+3 \equiv s^2(A+B) + s(2A+B+C) + 4A+C,$$

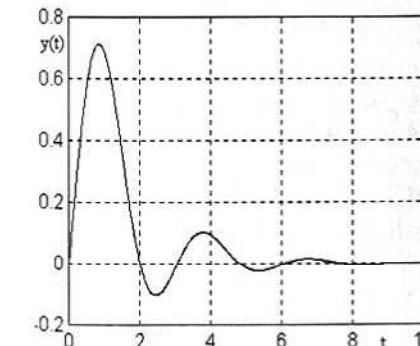
a następnie porównać współczynniki lewej i prawej strony przy jednakowych potęgach s :

$$\begin{cases} A+B=0 \\ 2A+B+C=1 \Rightarrow A=\frac{2}{3} \text{ i } B=-\frac{2}{3} \text{ i } C=\frac{1}{3}, \\ 4A+C=3 \end{cases}$$

$$Y(s) = \frac{2}{3} \frac{1}{s+1} - \frac{2}{3} \frac{s+1}{(s+1)^2 + 3} + \frac{1}{3} \frac{3}{(s+1)^2 + 3}.$$

Z tabeli 4.1 wynika postać rozwiązania (4.3), którego przebieg przedstawiono na rysunku 4.2.

$$y(t) = \left[\frac{2}{3} e^{-t} - \frac{2}{3} e^{-t} \cos \sqrt{3}t + \frac{1}{\sqrt{3}} e^{-t} \sin \sqrt{3}t \right] \cdot 1(t) \quad (4.3)$$



Rys. 4.2. Kształt rozwiązania (4.3)

Przykład 4.2

Znaleźć rozwiązanie równania trzeciego rzędu

$$\ddot{y}(t) + 9\dot{y}(t) + 27\dot{y}(t) + 27y(t) = 2\ddot{u}(t) + 3\dot{u}(t) + 3u(t) \quad (4.4)$$

Warunki początkowe zerowe: $y(0) = 0$, $\dot{y}(0) = 0$, $\ddot{y}(0) = 0$.

Wykładnicza funkcja wymuszająca $u(t) = e^{-t}1(t)$ dla $t \geq 0$.

Po zastosowaniu do równania (4.4) transformacji Laplace'a otrzymuje się

$$Y(s)(s+3)^3 = (2s^2 + 3s + 3)U(s) \quad (4.5)$$

$$\text{gdzie } U(s) = \frac{1}{s+1}.$$

Ze względu na wielokrotność pierwiastka mianownika dla rozwiązywania należy zastosować inny rozkład na ułamki proste i inną metodologię wyliczenia współczynników

$$Y(s) = \frac{2s^2 + 3s + 3}{(s+3)^3 \cdot (s+1)} = \frac{A}{s+1} + \frac{B}{s+3} + \frac{C}{(s+3)^2} + \frac{D}{(s+3)^3}.$$

Po obustronnym wymnożeniu tego równania przez kolejne mianowniki prawej strony, otrzymuje się równoważności, które muszą być spełnione dla każdego s (dla poszczególnych pierwiastków)

$$(s+1)Y(s)|_{s=-1} = A = 0,25, \quad (s+3)^3 Y(s)|_{s=-3} = D = -6,$$

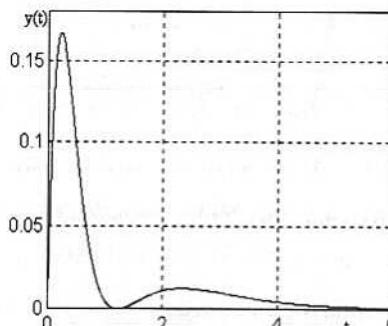
$$\frac{d}{ds} [(s+3)^3 Y(s)]|_{s=-3} = \frac{2s^2 + 4s}{(s+1)^2} \Big|_{s=-3} = C = \frac{3}{2},$$

$$\frac{d^2}{2!ds^2} [(s+3)^3 Y(s)]|_{s=-3} = \frac{2}{(s+1)^3} \Big|_{s=-3} = B = -0,25,$$

$$Y(s) = \frac{0,25}{s+1} + \frac{-0,25}{s+3} + \frac{1,5}{(s+3)^2} + \frac{-6}{(s+3)^3} \quad (4.6)$$

Z tabeli 4.1 wynika końcowa postać rozwiązania, którego przebieg przedstawiono na rysunku 4.3

$$y(t) = \left[\frac{1}{4}e^{-t} - \frac{1}{4}e^{-3t} + \frac{3}{2}t e^{-3t} - 3t^2 e^{-3t} \right] \cdot 1(t) \quad (4.7)$$



Rys. 4.3. Kształt rozwiązania (4.7)

4.3. Pojęcie transmitancji operatorowej układu liniowego

Rozważać będziemy liniowe układy dynamiczne o stałych współczynnikach, opisywane stacjonarnymi zwyczajnymi równaniami różniczkowymi. Jak wspomniano w podrozdziale 3.3 na rozwiązywanie takiego równania składa się część zależna tylko od warunków początkowych i część zależna tylko od zewnętrznej funkcji wymuszającej (sterowania). Przy założeniu zerowych warunków początkowych można zdefiniować pewną operatorową funkcję opisującą dynamikę toru sterowania systemu, tzn. dynamikę pomiędzy miejscem wprowadzenia sterowania i punktem pomiarowym charakteryzującym wyjście systemu.

Z równań (4.2) i (4.5) wynika, że równanie operatorowe systemu może być zawsze przedstawione w postaci

$$Y(s)M(s) = L(s)U(s),$$

gdzie $M(s)$ i $L(s)$ są wielomianami zmiennej s .

Def. 4.1. Stosunek transformaty Laplace'a funkcji wyjścia systemu do transformaty Laplace'a wejścia systemu (przy zerowych warunkach początkowych) tworzy ułamek wymierny zmiennej s i nazywa się **transmitancją operatorową systemu $G(s)$** .

$$\frac{Y(s)}{U(s)} = \frac{L(s)}{M(s)} = \frac{b_m s^m + b_{m-1} s^{m-1} + \dots + b_1 s + b_0}{a_n s^n + a_{n-1} s^{n-1} + \dots + a_1 s + a_0} = G(s) \quad (4.8)$$

Pierwiastki wielomianu licznika $L(s)$ nazywane są zerami transmitancji, a pierwiastki mianownika $M(s)$ – biegunami.

Transmitancja jest zależna tylko od struktury konstrukcyjnej układu i wybranego toru wyjście-wyjście. Zbudowany układ elektryczny RLC ma swoją transmitancję, niezależnie od kształtu funkcji, która w przyszłości będzie sterowaniem. Transmitancję można więc policzyć bezpośrednio po ukończeniu konstrukcji, przy znanych wartościach parametrów RLC . Należy tylko zadecydować o zmiennych wejście-wyjście. Dobrze znana z elektrotechniki impedancja $Z(s)$, rozumiana jako stosunek transformaty Laplace'a napięcia do transformaty prądu w wybranym dwójniku (prawo Ohma), jest przykładem transmitancji operatorowej. Transmitancja może być też stosunkiem transformat dwóch napięć, na co nie ma już specjalnej nazwy w terminologii elektrotechniki.

Wcześniej znajomość transmitancji pozwala na teoretyczne obliczenie przyszłego wyjścia $y(t)$ dla zadanego wejścia $u(t)$ ze wzoru

$$Y(s) = G(s)U(s).$$

Stosowanie transmitancji przy analizie dynamiki wnosi wiele korzyści poprzez algebraizację obliczeń. Konsekwencją tego jest również prosta algebra schematów blokowych i związki z transmitancją widmową, czyli analizą odpowiedzi układu na wymuszenia harmoniczne.

Postać $Y(s) = G(s)U(s)$ wiąże się ściśle z tematyką splotu z podrozdziału 3.7 i własnością nr 5 transformaty Laplace'a z tabeli 4.2. Odpowiedź systemu liniowego w dziedzinie czasu jest bowiem splotem impulsowej funkcji przejścia, czyli oryginału transmitancji i sterowania.

Iloczyn splotowy dwóch funkcji w dziedzinie czasu oznacza się znakiem „*”

$$y(t) = g(t)*u(t).$$

W dziedzinie zmiennej s odpowiada to algebraicznemu iloczynowi ich dwóch transformat

$$Y(s) = G(s) \cdot U(s).$$

Ponieważ transformata delty Diraca $u(t) = \delta(t)$, $U(s) = 1$, to odpowiedź obiektu na takie sterowanie równa się:

$$Y(s) = G(s), \text{ czyli } y(t) = g(t),$$

co zgadza się ze wzorem (3.25). Potwierdza to fakt, że impulsowa funkcja przejścia $g(t)$ jest oryginałem transmitancji $G(s)$.

4.3.1. Własności transmitancji. Realizowalność fizyczna

Istotnym warunkiem, jaki powinna spełniać transmitancja układu ciągłego, jest tzw. warunek realizowalności. Problem ten był już omawiany w podrozdziale 3.6. Dla transmitancji właściwej ten warunek oznacza $m \leq n$ (stopień licznika nie może być większy od stopnia mianownika). W przeciwnym wypadku, $m > n$, transmitancja opisywałaby nieobserwowane w rzeczywistości zachowanie układów fizycznych. Na przykład dla teoretycznego systemu zerowego rzędu

$$a_0 y(t) = b_1 \dot{u}(t) + b_0 u(t),$$

wyjście dla sterowania skokowego, opisywałoby wzór operatorowy

$$Y(s) = \left[\frac{b_1 s + b_0}{a_0} \right] \cdot \frac{1}{s} = \frac{b_1}{a_0} + \frac{b_0}{a_0} \frac{1}{s}.$$

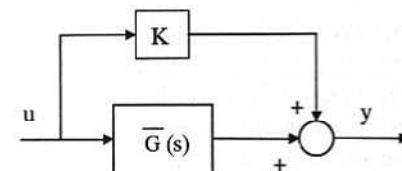
Oznacza to, że dla takiego sterowania w chwili $t = 0$ w odpowiedzi powinna być zaobserwowana delta Diraca, nierealizowalna w rzeczywistych układach.

$$y(t) = \frac{b_1}{a_0} \delta(t) + \frac{b_0}{a_0}.$$

System, dla którego $m = n$, jest już w praktyce realizowalny. We wzorze (4.8) można bowiem podzielić licznik przez mianownik i w wyniku otrzymuje się

$$G(s) = \frac{b_n}{a_n} + \frac{\bar{b}_{n-1}s^{n-1} + \dots + \bar{b}_1s + \bar{b}_0}{a_ns^n + a_{n-1}s^{n-1} + \dots + a_1s + a_0}.$$

Obok transmitancji ściśle właściwej ($m < n$), występuje tutaj wyraz stały, który wskazuje na istnienie w modelu dynamicznym również toru statycznego. Tor ten omija człony dynamiczne i przenosi sterowanie bezpośrednio na wyjście z pewnym współczynnikiem wzmacnienia $K = b_n/a_n$, (rys. 4.4).



Rys. 4.4. Statyczny tor sterowania

Sytuacja taka ma swoje odbicie w zapisie równań stanu. Przyjmuję one wtedy ogólniejszą niż w np. w równaniu (3.10) postać z niezerową macierzą wyjścia D (3.5).

Podczas korzystania z opisu za pomocą transmitancji należy zawsze pamiętać o dwóch jej cechach:

- 1) transmitancja nie uwzględnia wpływu warunków początkowych na rozwiązanie;
- 2) transmitancja obejmuje swoim opisem dynamikę systemu związaną bezpośrednio z torem wejście-wyjście; niekiedy tor ten reprezentuje cały układ, ale niekiedy tylko jego część.

Cechy te interpretowane są niekiedy jako cechy niekorzystne, choć jako takie zostały z góry przyjęte przy definiowaniu transmitancji. Omówimy bliżej te dwie własności transmitancji.

Ad 1)

Jeśli w chwili startu układu (zadziałania sygnału sterującego) do wyjściowej odpowiedzi policzonej ze znanego sterowania i transmitancji dołączy się dodatkowa odpowiedź od nieznanych warunków początkowych (naładowane kondensatory, poziomy cieczy w zbiornikach, prędkości mas), to sumaryczne wyjście będzie różne od wyliczonego. Tę część odpowiedzi związaną z warunkami początkowymi należy więc liczyć osobno i uwzględnić jako superpozycję z odpowiedzią na sterowanie.

Modele liniowe otrzymane są często jako linearyzacja układu nieliniowego wokół danego punktu pracy. Niezerowy punkt pracy może być traktowany jako warunki początkowe dla modelu liniowego. Model liniowy dany transmitancją traktuje się zwykle jako model ze

zmiennymi **odchyłkowymi**, dla których początek układu współrzędnych czasoprzestrzennych został umieszczony w punkcie linearyzacji. W punkcie linearyzacji, dla liniowego modelu odchyłkowego, odchyłkowe warunki początkowe są więc zerowe. To uzasadnia użycie transmitancji również w niezerowych punktach pracy.

Ad 2)

Typowym przykładem nireprezentatywności całego układu przez transmitancję jest przypadek obiektu, w którym występuje równość zer i biegunów i ich skracanie, np.

$$G(s) = \frac{s^2 + 3s + 2}{s^3 + 8s^2 + 19s + 12} = \frac{(s+1)(s+2)}{(s+1)(s^2 + 7s + 12)} = \frac{s+2}{s^2 + 7s + 12}.$$

Oryginalny układ rzędu trzeciego, mający trzy zmienne procesowe, w torze sterowania jest tylko układem rzędu drugiego. Jedna ze zmiennych może być obserwowały, ale niesterowalna (część obiektu reprezentowana przez blok G_{ns0}) lub sterowalna, ale nieobserwowały G_{sno} . Poglądowo przedstawiają to rysunek 4.5 i przykład 4.3. Transmitancja zawsze reprezentuje część toru związaną tylko z częścią sterowalną i obserwowały G_{so} .

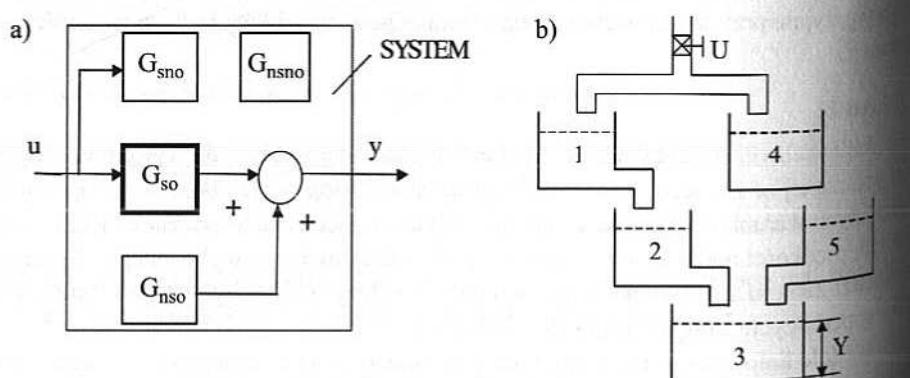
Dla jednowymiarowego systemu liniowego danego wzorem (3.16), $u(t) \in R$ i $y(t) \in R$

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t) + Du(t).\end{aligned}$$

Transmitancja dana jest wzorem $G(s) = C[I\delta - A]^{-1}B + D$.

Przykład 4.3

W poniższym układzie pięciu zbiorników (rys. 4.5), tor sterowania tworzony przez zbiorniki 1, 2 i 3 (poziom x_3 jest wyjściem) reprezentuje podsystem sterowalny i obserwowały. Zbiornik nr 4 (poziom x_4) jest sterowalny, ale nieobserwowały, zbiornik nr 5 (poziom x_5) jest obserwowały, ale niesterowalny. Pominięto podsystem niesterowalny i nieobserwowały.



Rys. 4.5. Cztery podsystemy (a) i przykład realizacji (b)

Dla układu zbiorników obowiązują równania:

$$\left. \begin{array}{l} \dot{x}_1(t) = -a_1 x_1(t) + b_1 u(t) \\ \dot{x}_2(t) = -a_2 x_2(t) + a_3 x_1(t) \\ \dot{x}_3(t) = +a_4 x_2(t) + a_5 x_5(t) \\ \dot{x}_4(t) = b_4 u(t) \\ \dot{x}_5(t) = -a_5 x_5(t) \\ y(t) = x_3(t) \end{array} \right\} \dot{\mathbf{x}}(\mathbf{t}) = \begin{bmatrix} -a_1 & 0 & 0 & 0 & 0 \\ a_3 & -a_2 & 0 & 0 & 0 \\ 0 & a_4 & 0 & 0 & a_5 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -a_5 \end{bmatrix} \mathbf{x}(\mathbf{t}) + \begin{bmatrix} b_1 \\ 0 \\ 0 \\ b_4 \\ 0 \end{bmatrix} u(t), \\ y(t) = [0 \ 0 \ 1 \ 0 \ 0] \mathbf{x}(\mathbf{t}).$$

Przyjmując przykładowe współczynniki $a_1 = 1$, $a_2 = 2$, $a_3 = 3$, $a_4 = 4$, $a_5 = 1$, $b_1 = 1$, i $b_4 = 1$, można policzyć transmitancję Y/U

$$G(s) = C[I\delta - A]^{-1}B + D = \frac{12s^2 + 12s}{s^5 + 4s^4 + 5s^3 + 2s^2} = \frac{12s(s+1)}{s^2(s+2)(s+1)^2} = \frac{12}{s(s+2)(s+1)}.$$

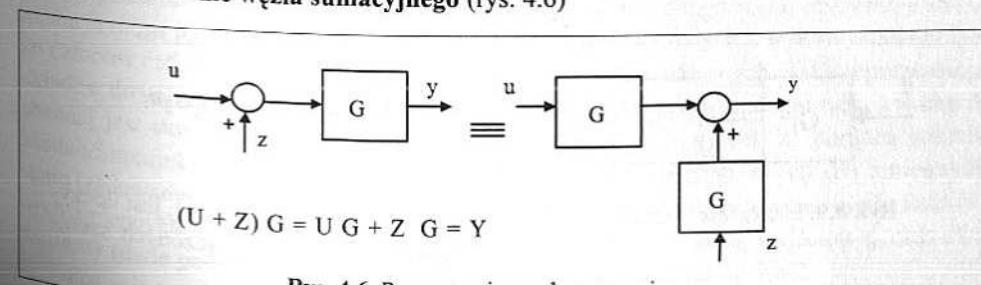
Widać, że transmitancja toru sterowania (bez skróceń) reprezentuje układ piątego rzędu. Przy zerowych warunkach początkowych zachowanie się toru sterowania dla zmiennej x_3 odpowiada jednak układowi trzeciego rzędu (całkowo-inercyjnemu z inercją drugiego rzędu). Obecność w sygnale wyjściowym części niesterowalnej (zbiornik 5) ujawnia się tylko przy niezerowych warunkach początkowych i może być uwzględniana w opisie dynamiki tylko za pomocą równań stanu. W tym sensie opis dynamiki systemu w dziedzinie czasu, za pomocą równania stanu jest ogólniejszy od opisu transmitancyjnego.

4.3.2. Algebra schematów blokowych

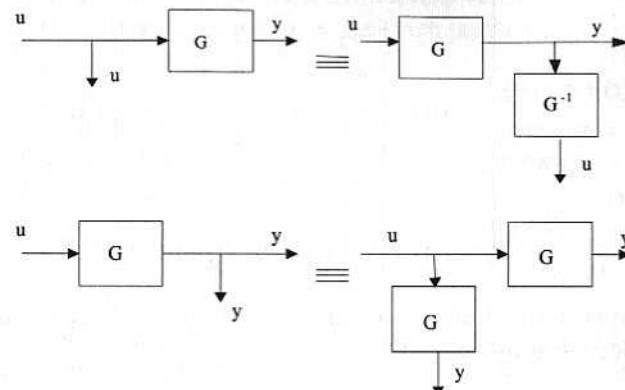
Złożone układy można przedstawić w postaci zbioru połączonych schematów blokowych reprezentujących podsystemy. Należy zwrócić uwagę, że połączenia bloków reprezentują jedynie połączenia informacyjne i nie mogą być utożsamiane z połączeniami galwanicznymi przenoszącymi moc. Połączone transmitancje nie obciążają się wzajemnie (przykład 4.14).

Schematy blokowe można zwijać do jednego zastępczego bloku na podstawie wymienionych poniżej pięciu reguł.

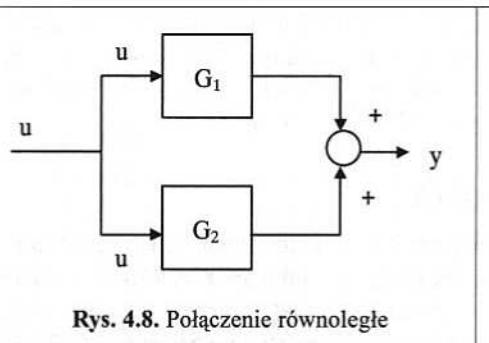
1) Przesuwanie węzła sumacyjnego (rys. 4.6)



Rys. 4.6. Przesuwanie węzła sumacyjnego

2) Przesuwanie węzła zaczepowego (rys. 4.7)

Rys. 4.7. Przesuwanie węzła zaczepowego

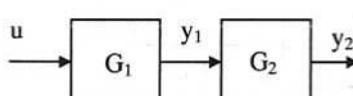
3) Połączenie równolegle transmitancji (rys. 4.8)

$$y = G_1 u + G_2 u = (G_1 + G_2) \cdot u,$$

$$G_Z = G_1 + G_2.$$

Transmitancja zastępcza G_z jest sumą transmitancji równoległych.

Rys. 4.8. Połączenie równolegle

4) Połączenie szeregowe transmitancji (rys. 4.9)

$$G_1 = \frac{y_1}{u} \Rightarrow y_1 = G_1 \cdot u,$$

$$y_2 = G_2 \cdot y_1 = G_1 G_2 u,$$

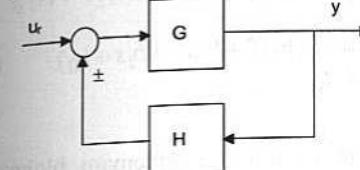
$$G_Z = G_1 \cdot G_2.$$

Transmitancja zastępcza G_z jest iloczynem transmitancji połączonych szeregowo.

Rys. 4.9. Połączenie szeregowo

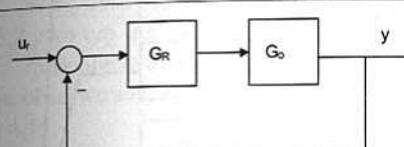
5) Sprzężenie zwrotne

- Schemat ogólny układu zamkniętego sprzeżeniem zwrotnym (rys. 4.10)



Rys. 4.10. Układ ze sprzężeniem zwrotnym

- Schemat układu nadążania za sygnałem referencyjnym u_r (rys. 4.11)

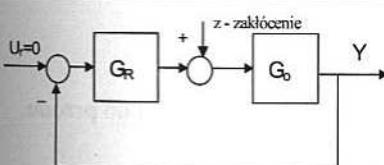


Transmitancja toru sterowania

$$G_z = \frac{G_R G_0}{1 + G_R G_0} = \frac{Y}{U_r}.$$

Rys. 4.11. Układ ze sprzężeniem zwrotnym

- Schemat układu stabilizacji w punkcie pracy w obecności zakłóceń (rys. 4.12)



Transmitancja toru zakłócenia

$$G_{zakl} = \frac{G_0}{1 + G_R G_0} = \frac{Y}{Z}.$$

Rys. 4.12. Układ ze sprzężeniem zwrotnym

4.3.3. Naturalne zmienne stanu dla obiektów typu SISO opisanych transmitancją

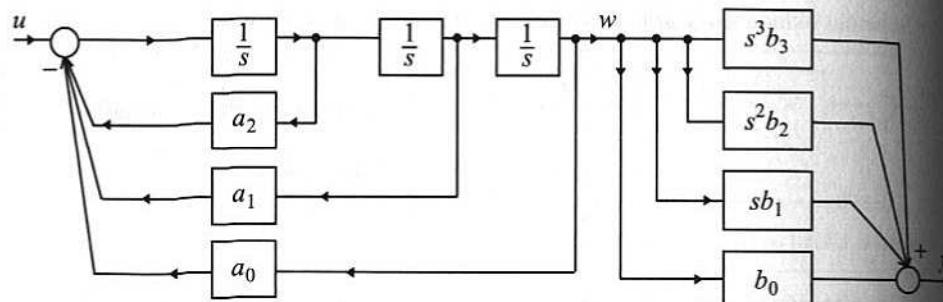
W podrozdziale zostanie wykazane, że układy liniowe n -tego rzędu opisywane transmitancją mogą być przedstawione w postaci schematów blokowych z wykorzystaniem tylko członów całkujących pierwszego rzędu i członów wzmacniających. Taka reprezentacja układów dowolnego rzędu była podstawą programowania maszyn analogowych, a w chwili obecnej jest stosowana w projektowaniu schematów symulacyjnych za pomocą pakietu Matlab/Simulink, umożliwiając uwzględnienie warunków początkowych dla zmiennych stanu (zadawanych dla członów całkujących). Reprezentacja ta wprowadza pojęcie naturalnych zmiennych stanu, którymi mogą być kolejne pochodne zmiennej wyjściowej, oraz dwa warianty (dwie postacie) przestrzeni stanu, znane już z podrozdziału 3.6.2. Postacie te zostaną wyprowadzone na przykładzie układu jednowymiarowego III rzędu.

Przykład 4.4

Dana jest transmitancja

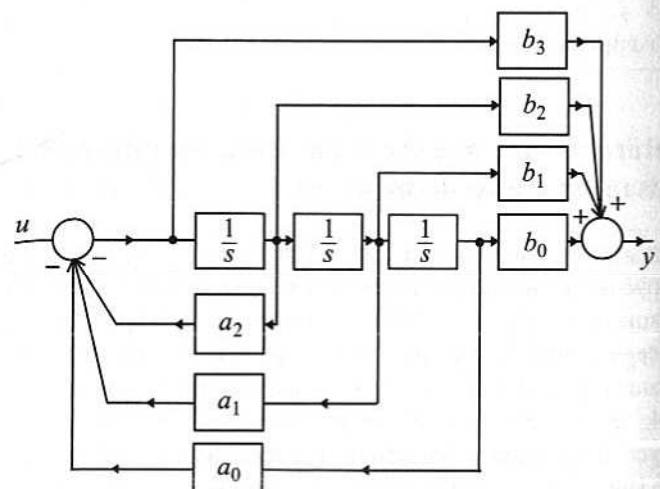
$$G(s) = \frac{b_3 s^3 + b_2 s^2 + b_1 s + b_0}{s^3 + a_2 s^2 + a_1 s + a_0} = \frac{1}{s^3 + a_2 s^2 + a_1 s + a_0} (b_3 s^3 + b_2 s^2 + b_1 s + b_0) \quad (4.9)$$

Transmitancję tę można formalnie przedstawić za pomocą schematu blokowego (rys. 4.13 i 4.14):



Rys. 4.13. Schemat blokowy

W schemacie tym po przekształceniu występuje sprzężenie do tyłu i do przodu.



Rys. 4.14. Schemat blokowy

4.3. Pojęcie transmitancji operatorowej układu liniowego

Wprowadzając wektor naturalnych zmiennych stanu $x(t) \in R^3$, w którym składowymi są wyjścia z członów całkujących, otrzymuje się:

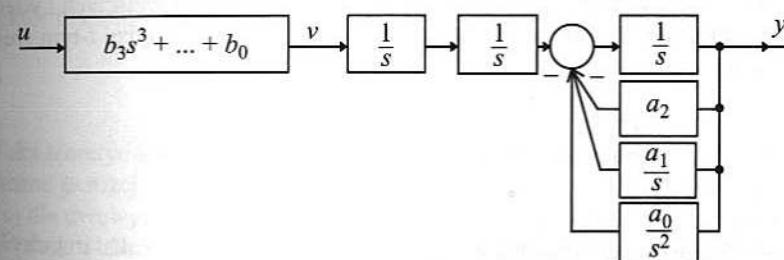
$$\dot{x}_1(t) = x_2(t), \quad \dot{x}_2(t) = x_3(t), \quad \dot{x}_3(t) = -a_0 x_1(t) - a_1 x_2(t) - a_2 x_3(t) + u(t),$$

$$\dot{x} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_0 & -a_1 & -a_2 \end{bmatrix} x + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u,$$

$$y = [b_0, b_1, b_2] x - [b_3 a_0, b_3 a_1, b_3 a_2] x + b_3 u = \\ = [b_0 - a_0 b_3, b_1 - a_1 b_3, b_2 - a_2 b_3] x + b_3 u.$$

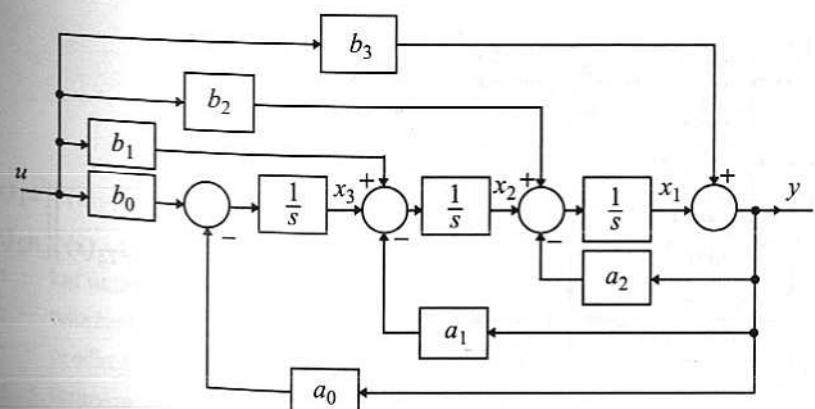
W wyniku otrzymuje się tzw. sterowaną postać przestrzeni stanu znaną z podrozdziału 3.5.2 jako typ Ia).

Inna algebra dla schematu blokowego opisującego transmitancję (4.9) prowadzi do tzw. obserwowlanej przestrzeni stanu. Wynika ona ze schematu przedstawionego na rysunku 4.15.



Rys. 4.15. Schemat blokowy

Schemat ten po przekształceniu ma postać przedstawioną na rysunku 4.16.



Rys. 4.16. Schemat blokowy

Analiza schematu z rysunku 4.16 prowadzi do równania stanu znanego z podrozdziału 3.5.2 jako typ IIb):

$$\dot{x} = \begin{bmatrix} -a_2 & 1 & 0 \\ -a_1 & 0 & 1 \\ -a_0 & 0 & 0 \end{bmatrix} x + \begin{bmatrix} b_2 - a_2 b_3 \\ b_1 - a_1 b_3 \\ b_0 - a_0 b_3 \end{bmatrix} u,$$

$$y = [1 \ 0 \ 0] x + b_3 u.$$

Jak wspomniano, zaletą powyższych struktur budowanych za pomocą pakietu Matlab/Simulink jest możliwość zadawania warunków początkowych dla członów całkujących, czyli na zmienne stanu, oraz dostępność na schemacie każdej zmiennej stanu.

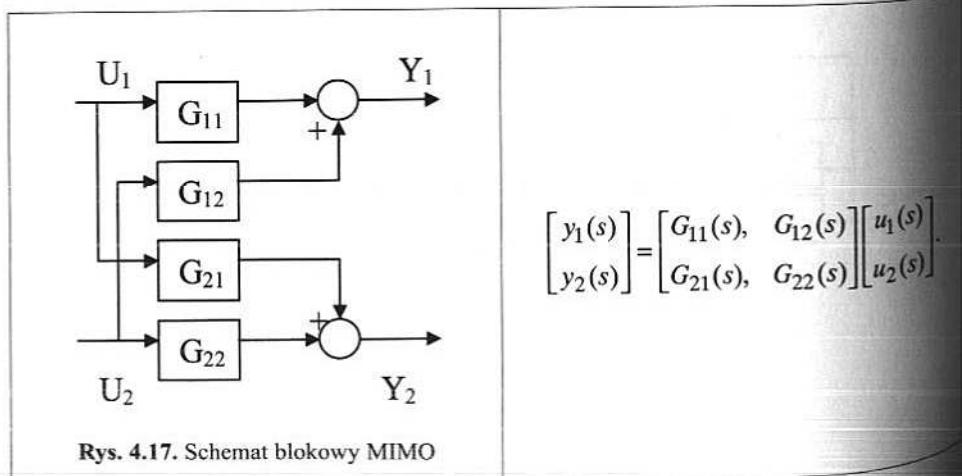
4.3.4. Transmitancja układów wielowymiarowych

W liniowych układach wielowymiarowych przy wielu różnych wejściach i wielu różnych wyjściach można określić transmitancje torów skrośnych pomiędzy i -tym wejściem i j -tym wyjściem

$$\frac{Y_j(s)}{U_i(s)} = G_{ij}(s).$$

Transmitancja $G(s)$ tworzona jest więc jako macierz transmitancji skalarnych $G_{ij}(s)$.

Dla układu o wymiarze $[2 \times 2]$ można to przedstawić na schemacie blokowym (rys. 4.17).



Rys. 4.17. Schemat blokowy MIMO

4.3. Pojęcie transmitancji operatorowej układu liniowego

Dla układów wielowymiarowych opisywanych macierzowym równaniem różniczkowym (3.5) z zerowymi warunkami początkowymi $x_0 = 0$, po zastosowaniu transformaty Laplace'a można otrzymać wzór na macierzową transmitancję $Y(s) = G(s)U(s)$ w funkcji macierzy A, B, C, D

$$G(s) = C [sI - A]^{-1} B + D.$$

Układy wielowymiarowe stanowią pewną trudność dla zadań sterowania, gdyż zmiana każdego z osobna sterowania powoduje jednocześnie zmiany wielu wyjść. Aby rozseparować wpływ sterowań na wyjścia, stosuje się sprzężenie zwrotne i specjalne układy odsprzęgające dołączone do systemu sterowanego. Mają one za zadanie uzależnienie każdego wyjścia tylko od jednego sterowania.

Dla układu z ujemnym sprzężeniem zwrotnym (rys. 4.10, podrozdz. 4.3.2) łatwo wykazać, że ponieważ blokowa transmitancja zastępcza $G_z(s)$ wyraża się wzorem

$$G_z(s) = [I + G(s)H(s)]^{-1} G(s),$$

to korektor realizujący żądaną postać $G_z(s)$ powinien spełniać równanie

$$H(s) = G_z^{-1}(s) - G^{-1}(s).$$

Taka teoretyczna reguła nie zawsze jest realizowalna fizycznie. Jej zastosowanie można pokazać poniżej na prostym przykładzie syntezy korektora diagonalizującego: policzymy $H(s)$ dla dwuwymiarowego, uproszczonego modelu silnika turboodrzutowego o dwóch sterowaniach i dwóch zmiennych wyjściowych, opisanego przykładowym układem równań stanu i wyjścia [4GIJ]:

$$\dot{x}_1(t) = -x_1(t) + u_1(t),$$

$$\dot{x}_2(t) = -x_2(t) + u_2(t),$$

$$y_1(t) = -2x_1(t) + 3x_2(t),$$

$$y_2(t) = 4x_1(t) - 6x_2(t) + 8u_2(t),$$

gdzie:

u_1 – kąt natarcia śmigła,

u_2 – natężenie dopływu paliwa,

y_1 – prędkość obrotowa silnika,

y_2 – temperatura na wlocie turbiny,

x – wewnętrzne zmienne stanu.

Stąd macierze i transmitancja blokowa mają postać:

$$A = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} -2 & 3 \\ 4 & -6 \end{bmatrix}, \quad D = \begin{bmatrix} 0 & 0 \\ 0 & 8 \end{bmatrix},$$

$$G(s) = \begin{bmatrix} \frac{-2}{s+1} & \frac{3}{s+1} \\ \frac{4}{s+1} & \frac{8s+2}{s+1} \end{bmatrix}.$$

Aby otrzymać przykładową, diagonalną macierz układu zamkniętego $G_z(s)$

$$G_z(s) = \begin{bmatrix} \frac{-2}{s+1} & 0 \\ 0 & \frac{8s+2}{s+1} \end{bmatrix},$$

korektor powinien mieć transmitancję

$$H(s) = \begin{bmatrix} \frac{-(s+1)}{2} & 0 \\ 0 & \frac{s+1}{8s+2} \end{bmatrix} - \begin{bmatrix} \frac{-(8s+2)}{16} & \frac{3}{16} \\ \frac{1}{4} & \frac{1}{8} \end{bmatrix} = \begin{bmatrix} \frac{-3}{8} & \frac{-3}{16} \\ \frac{-1}{4} & \frac{3}{4(8s+2)} \end{bmatrix}.$$

Kompensacja sprzężeń skrośnych poprzez wyzerowanie odpowiednich elementów transmitancji w ogólności realizowana jest przez skracanie biegunków i zer i prowadzi do utraty sterowności od stanu dla całego układu zamkniętego [4TAY]. Pełne odsprzężenie (*decoupling*) jest możliwe tylko dla wybranego przedziału częstotliwości [4GOG].

4.3.5. Przykłady transmitancji układów skalarnych

W niniejszym rozdziale podano przykłady (4.5–4.17) prostych obiektów prezentowanych w podrozdziale 2.3.3 z odpowiadającymi im transmitancjami (rys. 4.18–4.30).

Przykład 4.5

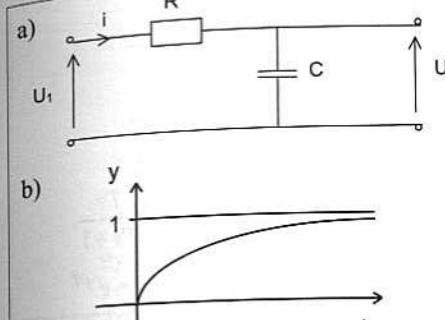
Człon całkujący

Dla zerowego warunku początkowego napięcia $u(0) = 0$ transmitancja pojemności (impedancia) wynosi:

$$C \frac{du(t)}{dt} = i(t), \quad G(s) = \frac{U(s)}{I(s)} = \frac{1}{Cs}.$$

Przykład 4.6

Stabilny układ inercyjny rzędu I

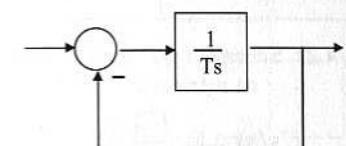


Rys. 4.18. Schemat (a) i odpowiedź (b) $y = u_2$ na skok u_1

$$U_2(s) = \frac{1}{Cs} \cdot \frac{U_1(s)}{R + 1/Cs},$$

$$G(s) = \frac{U_2(s)}{U_1(s)} = \frac{1}{RCs + 1} = \frac{1}{Ts + 1}.$$

Transmitancję taką można otrzymać również z członu całkującego objętego ujemnym sprzężeniem zwrotnym



Rys. 4.19. Układ inercyjny

$$G(s) = \frac{1/Ts}{1 + 1/Ts} = \frac{1}{Ts + 1}$$

Odpowiedź na skok jednostkowy ma postać jak na rysunku 4.18

$$u_1(t) = K1(t);$$

$$y(t) = u_2(t) = K[1 - e^{-t/T}].$$

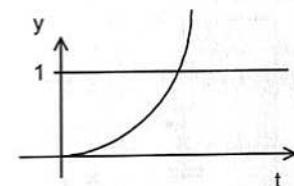
Przykład 4.7

Niestabilny układ inercyjny rzędu I

Transmitancję niestabilną można otrzymać z członu całkującego, objętego dodatnim sprzężeniem zwrotnym (por. rys. 4.19).

$$G(s) = \frac{1/Ts}{1 - 1/Ts} = \frac{1}{Ts - 1}.$$

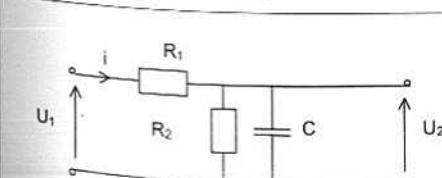
Dodatni biegum i sterowanie skokowe $u_1(t) = 1(t)$ dają odpowiedź: $y(t) = e^{t/T} - 1$.



Rys. 4.20. Odpowiedź narastająca na $u(t) = 1(t)$

Przykład 4.8

Układ inercyjny I rzędu



Rys. 4.21. Układ inercyjny

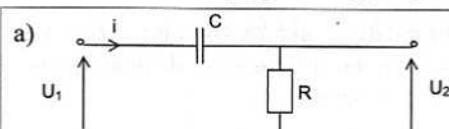
$$G(s) = \frac{K}{1 + KTs},$$

$$T = R_1 C,$$

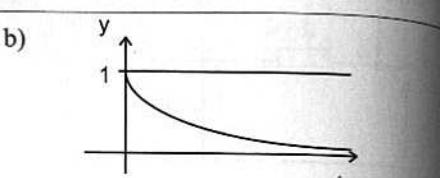
$$K = \frac{R_2}{R_1 + R_2}.$$

Przykład 4.9

Układ różniczkujący z inercją rzędu I

Rys. 4.22. Schemat (a) i odpowiedź (b) na $u_1(t)=1(t)$

$$U_2(s) = R \cdot \frac{U_1(s)}{R + \frac{1}{Cs}}$$

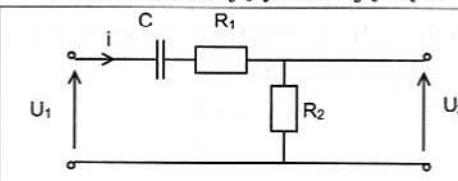


$$G(s) = \frac{U_2(s)}{U_1(s)} = \frac{RCs}{RCs+1} = \frac{Ts}{Ts+1}$$

Odpowiedź na $u_1(t) = 1(t)$: $y(t) = e^{-t/T}$

Przykład 4.10

Układ różniczkujący z inercją rzędu I



Rys. 4.23. Układ różniczkujący z inercją

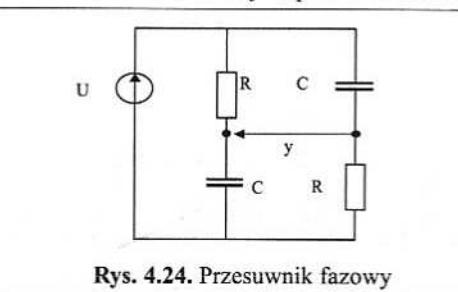
$$G(s) = \frac{T_2 s}{1 + (T_1 + T_2)s},$$

$$T_1 = R_1 C,$$

$$T_2 = R_2 C.$$

Przykład 4.11

Przesuwnik fazowy w postaci mostka



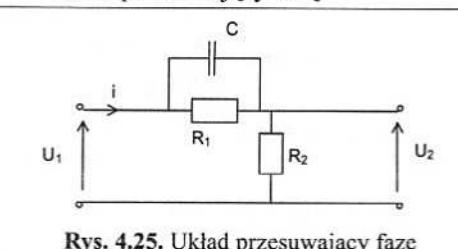
Rys. 4.24. Przesuwnik fazowy

$$G(s) = \frac{1-Ts}{1+Ts} = \frac{Y(s)}{U(s)},$$

$$T = RC.$$

Przykład 4.12

Układ przesuwający fazę



Rys. 4.25. Układ przesuwający fazę

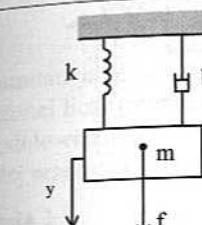
$$G(s) = k \frac{1+Ts}{1+kTs},$$

$$k = \frac{R_1}{R_1 + R_2},$$

$$T = R_1 C.$$

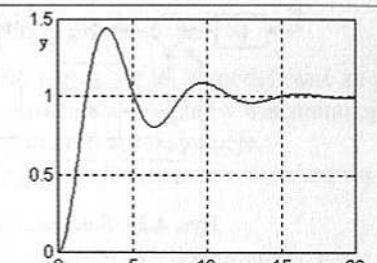
Przykład 4.13

Układ mechaniczny II rzędu

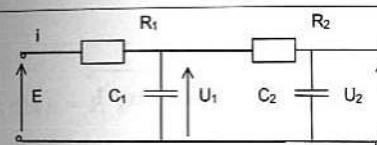


Rys. 4.26. Układ mechaniczny

$$G(s) = \frac{Y(s)}{F(s)} = \frac{1}{ms^2 + bs + k}.$$

Rys. 4.27. Przykładowa odpowiedź na $f(t) = 1(t)$
 $m\ddot{y}(t) + b\dot{y}(t) + ky(t) = f(t)$.**Przykład 4.14**

Układ II rzędu inercyjny RC



Rys. 4.28. Układ II rzędu RC

Transmitancje podukładów pracujących osobno:

$$G_1(s) = \frac{U_1(s)}{E(s)} = \frac{1}{T_1 s + 1},$$

$$G_2(s) = \frac{U_2(s)}{U_1(s)} = \frac{1}{T_2 s + 1}.$$

Z powodu wzajemnej obciążalności układów transmitancja zastępcza nie jest iloczynem G_1 i G_2 .

$$\frac{U_2(s)}{E(s)} = G_Z(s) \neq G_1(s) \cdot G_2(s),$$

$$G_Z(s) = \frac{1}{T_1 T_2 s^2 + (T_1 + T_2 + R_1 C_2)s + 1},$$

$$T_1 = R_1 C_1, T_2 = R_2 C_2.$$

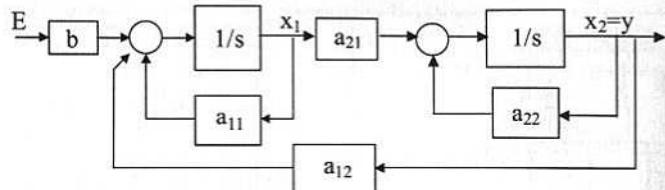
Oznaczając napięcia na kondensatorach jako zmienne stanu $u_1(t) = x_1(t)$, $u_2(t) = x_2(t) = y(t)$, i $u(t) = E(t)$ otrzymujemy się dla przykładu 4.14 równanie stanu:

$$\dot{x}(t) = \begin{bmatrix} -\left(\frac{1}{T_1} + \frac{1}{C_1 R_2}\right) & \frac{1}{C_1 R_2} \\ \frac{1}{T_2} & -\frac{1}{T_2} \end{bmatrix} x(t) + \begin{bmatrix} \frac{1}{T_1} \\ 0 \end{bmatrix} u(t),$$

$$y(t) = [0 \ 1] x(t),$$

$$G_z(s) = C[SI - A]^{-1}B = \frac{1}{T_1 T_2 s^2 + (T_1 + T_2 + R_1 C_2)s + 1}.$$

Schemat blokowy budowany na podstawie równania stanu jest pokazany na rysunku 4.29.



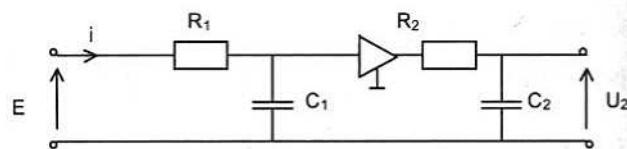
Rys. 4.29. Schemat blokowy układu z rysunku 4.28

Na schemacie występuje sprzężenie zwrotne ze współczynnikiem a_{12} , choć w obiekcie fizycznym (przykład 4.14) nie jest ono widoczne. Potwierdza to wspomniany już w rozdziale 1 fakt, że naturalne sprzężenia zwrotne są formalnie obecne we wszystkich układach wzajemnie oddziałujących na siebie.

Przykład 4.15

Układ II rzędu z operacyjnym wzmacniaczem separującym jak na rysunku 4.30 ma transmitancję

$$\frac{U_2(s)}{E(s)} = G_Z(s) = G_1(s) \cdot G_2(s) = \frac{1}{(T_1 s + 1)(T_2 s + 1)} = \frac{1}{T_1 T_2 s^2 + (T_1 + T_2)s + 1}.$$



Rys. 4.30. Układ II rzędu rozseparowany

Przykład 4.16

Człon inercyjny n -tego rzędu [4GOH1]

Transmitancja i odpowiedź tego członu na skok jednostkowy wyrażają się wzorami:

$$G_n(s) = \frac{K}{\left(\frac{\tau}{n}s + 1\right)^n},$$

$$y(t) = K \left[1 - e^{-\frac{nt}{\tau}} \sum_{i=0}^{n-1} \left(\frac{nt}{\tau} \right)^i \cdot \frac{1}{i!} \right].$$

Zwiększenie liczby członów inercyjnych daje w granicy transmitancję członu z opóźnieniem

$$\lim_{n \rightarrow \infty} G_n(s) = e^{-\tau s}.$$

Przykład 4.17

Model opóźnienia transportowego

$$y(t) = u(t - \tau) \Rightarrow Y(s) = e^{-\tau s} U(s) \Rightarrow G(s) = e^{-\tau s}.$$

Transmitancja ta ma postać niewymierną, tzn. nie da się jej zamodelować za pomocą np. skończonej liczby elementów RLC (nieskończona liczba biegunów transmitancji). Często dla modelowania tego członu stosuje się jej wymiarne aproksymacje. Poniżej przedstawimy trzy główne metody aproksymacji skończenie wymiarowej dla $e^{-\tau s}$.

Metoda 1

Wykorzystując aproksymację nieskończonym ciągiem układów inercyjnych z malejącymi stałymi czasowymi (jak w przykładzie 4.16)

$$e^{-\tau s} = \lim_{n \rightarrow \infty} [1/(1 + \tau s/n)^n],$$

można uzyskać skończenie wymiarową aproksymację członu opóźniającego

$$e^{-\tau s} = \frac{e^{-\tau s/2}}{e^{\tau s/2}} \approx \frac{(1 - \tau s/(2n))^n}{(1 + \tau s/(2n))^n},$$

$$\text{np. dla } n=1, e^{-\tau s} \approx \frac{2 - \tau s}{2 + \tau s}, \text{ a dla } n=2, e^{-\tau s} \approx \frac{\tau^2 s^2 - 8\tau s + 16}{\tau^2 s^2 + 8\tau s + 16}.$$

Metoda 2

Trzy poniższe struktury a, b, c wykorzystują definicję eksponenty.

$$e^{-\tau s} \stackrel{\text{def}}{=} \sum_{i=0}^{\infty} \frac{(-\tau s)^i}{i!} = 1 - \tau s + \frac{(\tau s)^2}{2!} - \frac{(\tau s)^3}{3!} + \dots$$

a) Ponieważ zachodzi zależność

$$e^{-\tau s} = \frac{1}{e^{\tau s}} \approx \frac{1}{1 + \tau s + \frac{(\tau s)^2}{2!} + \frac{(\tau s)^3}{3!} + \dots},$$

$$\text{to np. dla } n=1, e^{-\tau s} \approx \frac{1}{1 + \tau s}, \text{ a dla } n=2, e^{-\tau s} \approx \frac{2}{\tau^2 s^2 + 2\tau s + 2}.$$

b) Ponieważ zachodzi zależność

$$e^{-\tau s} = \frac{e^{-\tau s/2}}{e^{\tau s/2}} \approx \frac{1 - \tau s/2 + (\tau s/2)^2 / 2! - (\tau s/2)^3 / 3! \dots}{1 + \tau s/2 + (\tau s/2)^2 / 2! + (\tau s/2)^3 / 3! \dots},$$

$$\text{to np. dla } n=1, e^{-\tau s} \approx \frac{2 - \tau s}{2 + \tau s}, \text{ a dla } n=2, e^{-\tau s} \approx \frac{\tau^2 s^2 - 4\tau s + 8}{\tau^2 s^2 + 4\tau s + 8}.$$

c) Ponieważ zachodzi zależność

$$1 + e^{-s\tau} = (1 + e^{s\tau})e^{-s\tau} \Rightarrow e^{-s\tau} = \frac{1 + e^{-s\tau}}{1 + e^{s\tau}} \approx \frac{2 - \tau s + (\tau s)^2 / 2! - (\tau s)^3 / 3! \dots}{2 + \tau s + (\tau s)^2 / 2! + (\tau s)^3 / 3! \dots},$$

to np. dla $n = 1$, $e^{-s\tau} \approx \frac{2 - \tau s}{2 + \tau s}$, a dla $n = 2$, $e^{-s\tau} \approx \frac{\tau^2 s^2 - 2\tau s + 4}{\tau^2 s^2 + 2\tau s + 4}$.

Metoda 3

Metoda wykorzystuje aproksymację Padégo

$$e^{-s\tau} = \frac{L_n(s)}{M_n(s)},$$

gdzie $L_n(s\tau) = M_n(-s\tau)$

$$\begin{aligned} M_n(s\tau) = & 1 + \frac{n}{2n \cdot 1!} s\tau + \frac{n(n-1)}{2n(2n-1) \cdot 2!} (s\tau)^2 + \\ & + \frac{n(n-1)(n-2)}{2n(2n-1)(2n-2) \cdot 3!} (s\tau)^3 + \dots + \frac{n!}{2n! 2!} (s\tau)^n, \end{aligned}$$

np. dla $n = 1$, $e^{-s\tau} \approx \frac{2 - \tau s}{2 + \tau s}$, a dla $n = 2$, $e^{-s\tau} \approx \frac{\tau^2 s^2 - 6\tau s + 12}{\tau^2 s^2 + 6\tau s + 12}$.

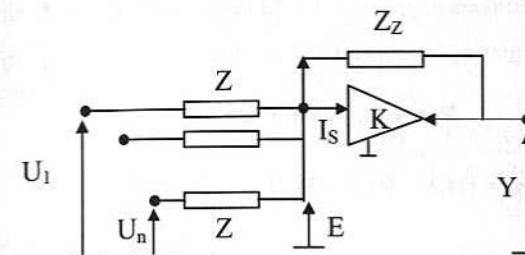
Można zauważyć, że w większości aproksymacji skończenie wymiarowy model członu opóźniającego dla $n = 1$ jest taki sam, ale dla $n = 2$ jest już inny. Zauważmy również, że w metodzie 3 aproksymacja drugiego stopnia daje wzór, którego analog macierzowy został zaproponowany w podrozdziale 3.4 do obliczeń numerycznych wykładniczej funkcji macierzowej e^{At} (3.9).

4.3.6. Elektroniczna realizacja transmitancji

Do budowy analogowych, ciągłych dynamicznych elementów w układach automatyki takich, jak człony korekcyjne, regulatory, obserwatory, które realizowałyby zadaną transmitancję, wykorzystywana jest elektroniczna technika analogowa [4SZZ]. Jednak w tych strukturach tylko częściowo są używane bierne układy RC przedstawione w poprzednim podrozdziale. Budowane układy korekcyjne automatyki, oprócz pożądanych własności dynamicznych powinny, posiadać jeszcze dodatkową cechę – wzajemnej małej obciążalności (dużą oporność wejściową). Gwarantowałoby to zachowanie reguł algebry schematów blokowych, a stąd bezproblemowe ich łączenie (przykłady 4.14 i 4.15 w podrozdz. 4.3.5). W realizacjach praktycznych wykorzystywane są więc zawsze elektroniczne konstrukcje aktywne ze wzmacniaczami operacyjnymi o wzmacnieniu rzędu 10^5 i oporności wejściowej rzędu $M\Omega$. Poniżej przedstawimy opis takiej uniwersalnej konstrukcji członu dynamicznego z wieloma wejściami napięciowymi poprzez impedancje Z_i i ze sprzężeniem zwrotnym o impedancji Z_Z obejmującym wzmacniacz.

Przy założeniu, że prąd wejściowy wzmacniacza jest mały $I_S \approx 0$, otrzymuje się równanie prądowe w węźle (rys. 4.31).

$$\sum_{i=1}^n \frac{U_i - E}{Z_i} = \frac{E - Y}{Z_Z}.$$



Rys. 4.31. Schemat zastępczy wzmacniacza operacyjnego

Wykorzystując równanie wzmacniacza (z zachowaniem odwracania fazy napięcia) $Y = -KE$, można otrzymać ostatecznie transformatę Laplace'a równania wyjścia

$$Y = -\frac{Z_Z}{M} \sum_{i=1}^n \frac{U_i}{Z_i},$$

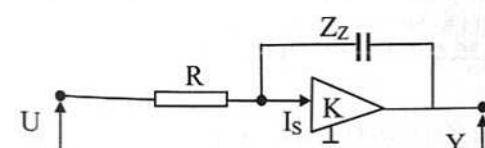
$$\text{gdzie } M = \frac{1}{1 + \frac{1}{K} \left[1 + Z_Z \sum_{i=1}^n \frac{1}{Z_i} \right]}.$$

Widac, że dla $K \rightarrow \infty$, współczynnik $M \rightarrow 1$, a stąd równanie wyjścia upraszcza się do

$$Y = -\sum_{i=1}^n \frac{Z_Z}{Z_i} U_i = -\sum_{i=1}^n \alpha_i U_i.$$

Współczynniki α_i wynikające ze stosunku impedancji są funkcjami operatora Laplace'a s.

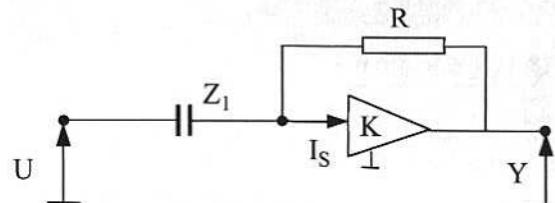
- Poniżej przedstawiono kilka (pięć) podstawowych realizacji transmitancji.
1) Człon całkujący (rys. 4.32)



Rys. 4.32. Człon całkujący

$$Z_Z(s) = 1/Cs, \quad G(s) = Y(s)/U(s) = -1/(RC \cdot s).$$

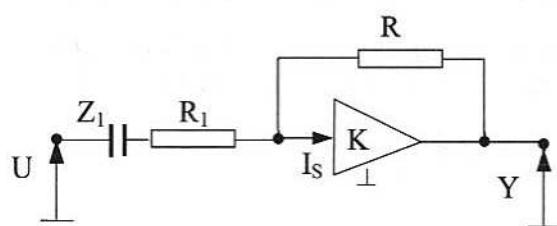
2) Czół różniczkujący idealny (rys. 4.33)



Rys. 4.33. Czół różniczkujący

$$Z_1(s) = 1/Cs, \quad G(s) = Y(s)/U(s) = -RC \cdot s.$$

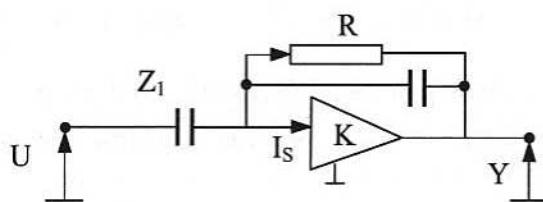
3) Czół różniczkujący z inercją (rzeczywisty) (rys. 4.34)



Rys. 4.34. Czół różniczkujący z inercją (rzeczywisty)

$$Z_1(s) = 1/Cs + R_1, \quad G(s) = Y(s)/U(s) = -RC s / (1 + R_1 C s), \quad \text{zwykle } R_1 \ll R.$$

4) Czół różniczkujący z inercją (rzeczywisty) (rys. 4.35)



Rys. 4.35. Czół różniczkujący z inercją (rzeczywisty)

$$Z_1(s) = 1/C_1 s, \quad G(s) = Y(s)/U(s) = -RC_1 s / (1 + RC_1 s), \quad \text{zwykle } C \ll C_1.$$

5) Do budowy czólu proporcjonalnego (wzmacniającego) jako impedancje Z_i i Z_Z stosuje się rezistory. Czóły wyższych rzędów buduje się na podstawie czółów całkowych (patrz podrozdz. 4.3.3).

4.4. Związki transmitancji z odpowiedzią układu na wymuszenia harmoniczne

4.4.1. Odpowiedź układu I rzędu na sygnał sinusoidalny

Dla sterowania $u(t) = A \cdot \sin(\omega t) \Rightarrow U(s) = \frac{A\omega}{s^2 + \omega^2}$ i transmitancji $G(s) = \frac{K}{1+sT}$ transformata wyjścia (przy zerowych warunkach początkowych) może być rozłożona na ułamki proste. Po obliczeniach otrzymujemy

$$Y(s) = \frac{AK \omega}{(Ts+1)(s^2 + \omega^2)} = \frac{AK}{T^2 \omega^2 + 1} \left(\frac{T^2 \omega}{Ts+1} + \frac{\omega}{s^2 + \omega^2} - \frac{sT \omega}{s^2 + \omega^2} \right).$$

Stąd oryginał transformaty $Y(s)$ ma postać:

$$\begin{aligned} y(t) &= \frac{AK}{T^2 \omega^2 + 1} \left(T \omega e^{\frac{-t}{T}} + \sin(\omega t) - T \omega \cos(\omega t) \right) = \\ &= \frac{AKT \omega}{T^2 \omega^2 + 1} e^{\frac{-t}{T}} + \frac{AK}{\sqrt{T^2 \omega^2 + 1}} \sin(\omega t - \varphi) \end{aligned} \quad (4.10)$$

gdzie $\operatorname{tg}(\varphi) = \omega T$, co oznacza, że sygnał na wyjściu jest zawsze opóźniony.

Traktując transmitancję jako liczbę zespoloną ($s = j\omega$), można policzyć jej moduł w funkcji parametru częstotliwości ω :

$$G(j\omega) = \frac{K}{1 + Tj\omega} = \frac{K(1 - Tj\omega)}{1 + T^2 \omega^2} = \frac{K}{1 + T^2 \omega^2} - j \frac{KT\omega}{1 + T^2 \omega^2} = |G(j\omega)| e^{-j\varphi(\omega)},$$

gdzie:

$$|G(j\omega)| = \frac{K}{\sqrt{1 + T^2 \omega^2}},$$

$$\operatorname{tg}[\varphi(\omega)] = \omega T.$$

Uwzględniając we wzorze (4.10) zanikanie pierwszej składowej do zera dla $t \rightarrow \infty$, otrzymuje się w stanie ustalonym na wyjściu drgania sinusoidalne o tej samej częstotliwości ω , amplitudzie równej iloczynowi amplitudy wejściowej A i modułu transmitancji $|G(\omega)|$

$$y_u(t) = A \cdot |G(\omega)| \cdot \sin(\omega t - \varphi(\omega)).$$

Otrzymany wynik może być uogólniony na układy liniowe wyższego rzędu.

W podrozdziale następnym omówimy tę ważną własność układów liniowych o stałych współczynnikach, z której wynika, że sygnały harmoniczne o częstotliwościach ω przechodzą przez układy liniowe z tą samą częstotliwością, ich amplituda jest przemnożona przez moduł transmitancji $|G(\omega)|$, a faza ulega przesunięciu o kąt $\varphi(\omega)$.

4.4.2. Przechodzenie sygnałów wykładniczych przez układy liniowe

Sygnal rzeczywisty $u(t) = A \cos(\omega t + \psi)$ można interpretować ogólniej jako sygnal będący sumą dwóch sygnałów wykładniczych, zespółonych sprzężonymi o połówkowymi amplitudach, których części urojone się znoszą, a części rzeczywiste dodają

$$u(t) = A \cos(\omega t + \psi) = 0,5 A e^{j(\omega t + \psi)} + 0,5 A e^{-j(\omega t + \psi)} \quad (4.1)$$

Sygnal ten jest przykładem pewnej dużej klasy sygnałów rzeczywistych, które mają specjalne znaczenie dla układów liniowych. Chodzi o sygnały rzeczywiste, które mogą być przedstawione jako suma skończonej liczby funkcji wykładniczych zespółonych

$$u(t) = \sum_{i=1}^N A_i e^{j\psi_i} e^{j\omega_i t} = \sum_{i=1}^N \alpha_i e^{j\omega_i t} \stackrel{s_i=j\omega_i}{=} \sum_{i=1}^N \alpha_i e^{s_i t}.$$

Sygnały wykładnicze jako jedyne nie zmieniają swojego kształtu (charakteru) po przejściu przez liniowy, stacjonarny układ dynamiczny, tzn. sygnal wyjściowy jest również wykładniczy. Wynika to z własności układu liniowego i z postaci całki splotowej. Dla zerowych warunków początkowych przy takim sygnale wejściowym, wyjście układu równe

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} g(\tau) u(t-\tau) d\tau = \int_{-\infty}^{\infty} g(\tau) \sum_{i=1}^N \alpha_i e^{s_i t} e^{-s_i \tau} d\tau = \\ &= \sum_{i=1}^N \left[\int_{-\infty}^{\infty} g(\tau) e^{-s_i \tau} d\tau \right] \alpha_i e^{s_i t} = \sum_{i=1}^N \left[G(s_i) \alpha_i e^{s_i t} \right] = \sum_{i=1}^N \beta_i(s_i) e^{s_i t} \end{aligned} \quad (4.12)$$

Widac, że wszystkie składowe wykładnicze sygnału wejściowego pojawiły się na wyjściu układu. Zmianie uległy tylko współczynniki w liniowej kombinacji tych eksponent.

Przy okazji zdefiniowana została tzw. transmitancja widmowa

$$G(j\omega) = \int_{-\infty}^{\infty} g(\tau) e^{-j\omega\tau} d\tau.$$

Def. 4.2. Transmitancją widmową $G(j\omega)$ liniowego układu stacjonarnego nazywa się stosunek wartości zespółonej składowej odpowiedzi $Y(j\omega)$ wywołanej wymuszeniem sinusoidalnym $U(j\omega)$ do wartości zespółonej tego wymuszenia. Otrzymuje się ją formalnie bezpośrednio z transmitancji operatorowej przez podstawienie $s = j\omega$, co związane jest z faktem, że transformata Fouriera (podrozdz. 4.5.2) jest szczególnym przypadkiem transformaty Laplace'a.

Transmitancja widmowa może być stosowana wtedy, gdy sygnal wejściowy jest funkcją mającą transformatę Fouriera [4KLJ].

Transmitancja widmowa jest funkcją zespółoną zmiennej ω

$$G(j\omega) = P(\omega) + jQ(\omega),$$

4.4. Związki transmitancji z odpowiedzią układu na wymuszenia harmoniczne

przy czym część rzeczywista jest funkcją parzystą $P(\omega) = P(-\omega)$, a część urojona funkcją nieparzystą $Q(\omega) = -Q(-\omega)$.

W szczególnym przypadku, gdy wejściem jest funkcja (4.11); $u(t) = A \cos(\omega t + \psi)$, to ze wzoru (4.12) otrzymuje się wyjście

$$y(t) = G(j\omega) 0,5 A e^{j\psi} e^{j\omega t} + G(-j\omega) 0,5 A e^{-j\psi} e^{-j\omega t}.$$

Jeśli impulsowa funkcja przejścia $g(\tau)$ jest rzeczywista, to $G(j\omega)$ i $G(-j\omega)$ są liczbami zespółonymi sprzężonymi $G(\pm j\omega) = P(\omega) \pm jQ(\omega)$, i wyjście jest równe:

$$\begin{aligned} y(t) &= 0,5A(P + jQ)[\cos(\omega t + \psi) + j \sin(\omega t + \psi)] + \\ &+ 0,5A(P - jQ)[\cos(\omega t + \psi) - j \sin(\omega t + \psi)] = \\ &= PA \cos(\omega t + \psi) - QA \sin(\omega t + \psi) = \\ &= \sqrt{P^2 + Q^2} A \cos(\omega t + \psi + \varphi) = B(\omega) \cos(\omega t + \psi + \varphi). \end{aligned}$$

Stąd wyjście ma ten sam charakter, co wejście. Amplituda uległa zmianie. Faza jest z reguły opóźniona $\varphi(\omega) = \arctg(Q(\omega)/P(\omega))$. Moduł transmitancji $|G(j\omega)| = \sqrt{P^2(\omega) + Q^2(\omega)}$ i faza $\varphi(\omega)$ są funkcjami częstotliwości sinusoidy wejściowej ω i w zależności od układu zmieniają się w różny sposób przy zmianach ω .

Transmitancja $G(j\omega) = P(\omega) + jQ(\omega)$, jako liczba zespółona, może być formalnie wyrysowana graficznie na płaszczyźnie zespółonej Gaussa, zwanej w tym przypadku płaszczyzną Nyquista. Dla układów, dla których jest możliwa realizacja ustalonego stanu sinusoidalnego (zachodzi to np. dla wszystkich układów asymptotycznie stabilnych), moduł transmitancji widmowej można otrzymać eksperymentalnie, jako stosunek amplitudy sygnału sinusoidalnego pojawiającego się na wyjściu do amplitudy sygnału sinusoidalnego działającego na wejściu układu.

Jeśli sygnal wejściowy składa się z kilku harmonicznych, to wykres modułu transmitancji w funkcji częstotliwości ω (wykres Bodego) daje informację – jak mocno są tłumione lub wzmacniane amplitudy poszczególnych harmonicznych, a wykres fazy – o ile stopni te harmoniczne są przesunięte po przejściu przez układ.

Uogólnienie skalarnego sinusoidalnego sygnału sterującego

$$u(t) = \alpha \sin \omega t + \beta \cos \omega t$$

na przypadek wielowymiarowy można zapisać jako

$$u(t) = [u_1, u_2] \cdot \begin{bmatrix} \sin \omega t \\ \cos \omega t \end{bmatrix},$$

gdzie:

$$u_1 \in R^r,$$

$$u_2 \in R^r,$$

$$u \in R^r.$$

Odpowiedź ustalona układu macierzowego z zerowymi warunkami początkowymi

$$\dot{x}(t) = Ax(t) + Bu(t)$$

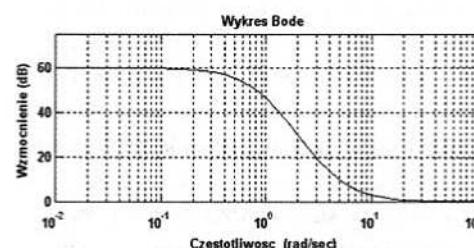
na to sterowanie, dana jest wzorem [4TAY]

$$x(t) = -(\omega^2 I - A^2)^{-1} [(ABu_1 - \omega Bu_2)(ABu_2 + \omega Bu_1)] \begin{bmatrix} \sin \omega t \\ \cos \omega t \end{bmatrix}.$$

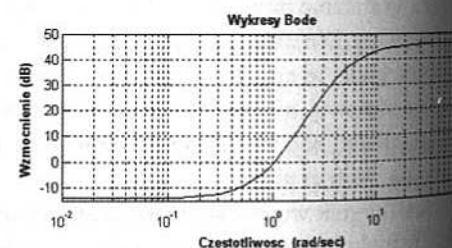
4.4.3. Charakterystyki częstotliwościowe układów dynamicznych

Transmitancja widmowa w torze wybranych sygnałów harmonicznych wejście/wyjście jest uogólnioną wersją elektrycznej impedancji widmowej dla sinusoidalnych sygnałów prąd/napięcie. Dla celów analizy i syntezy układów sterowania przydatnym jest graficzna reprezentacja modułu i fazy transmitancji w funkcji częstotliwości. Charakterystyki amplitudowo-częstotliwościowe pomagają w projektowaniu filtrów przeciwickłoceniowych, dolno-przepustowych, górnoprzepustowych, środkowo-zaporowych lub środkowo-przepustowych (rys. 4.36 i 4.37).

Logarytmiczne charakterystyki wzmacnienia (charakterystyki Bodego) filtrów pokazano na rysunkach 4.36 i 4.37.

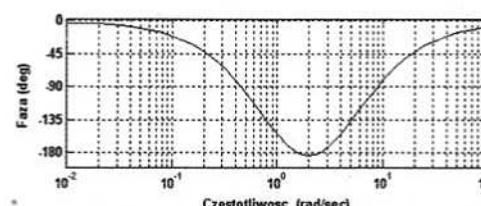


Rys. 4.36. Charakterystyka wzmacnienia dla $G(s) = [(s + 4)/(s + 1)]^5$



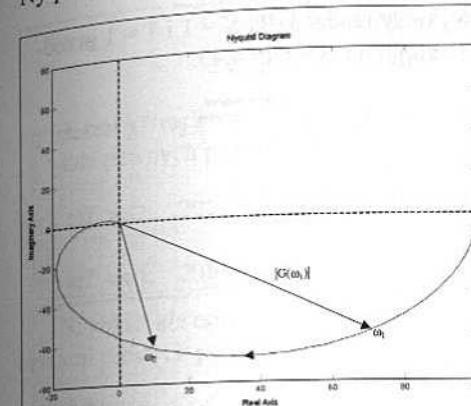
Rys. 4.37. Charakterystyka wzmacnienia dla $G(s) = 200[(s + 1)/(s + 4)]^5$

Wykresy fazy w funkcji częstotliwości dla filtrów z rysunków 4.36 i 4.37 pokazano na rysunku 4.38.



Rys. 4.38. Charakterystyki fazowe

Jak wspomniano, znaną transmitancję widmową możemy narysować na płaszczyźnie Nyquista jako liczbę zespoloną w funkcji parametru częstotliwości ω (rys. 4.39).



Rys. 4.39. Charakterystyka Nyquista

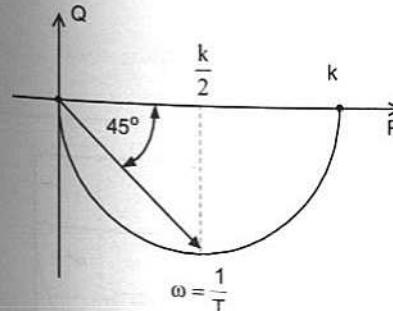
$$|G(\omega)| = \sqrt{P^2(\omega) + Q^2(\omega)},$$

$$\operatorname{tg}[\varphi(\omega)] = \frac{Q(\omega)}{P(\omega)}.$$

Przykład 4.18

Człon inercyjny I rzędu (rys. 4.40)

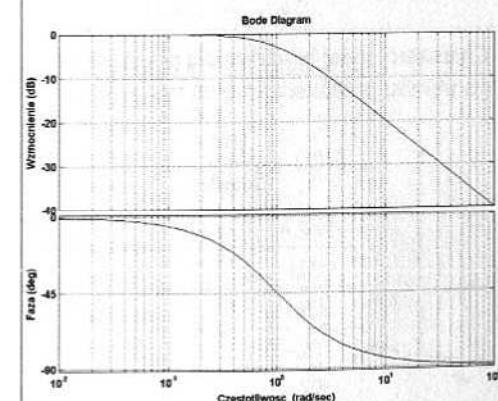
$$G(s) = \frac{K}{Ts + 1}, \quad \operatorname{tg} \varphi(\omega) = \omega T$$



Rys. 4.40. Charakterystyka Nyquista układu I rzędu

Wykresy Bodego dla $K = 1$ i $T = 1$ przedstawiono obok na rysunku 4.41.

$$G(j\omega) = \frac{K}{1 + \omega^2 T^2} - j \frac{K\omega T}{1 + \omega^2 T^2} = \frac{K}{\sqrt{1 + \omega^2 T^2}} e^{-j\varphi} = |G| e^{-j\varphi}.$$



Rys. 4.41. Wykres $|G(\omega)|$ i $\varphi(\omega)$

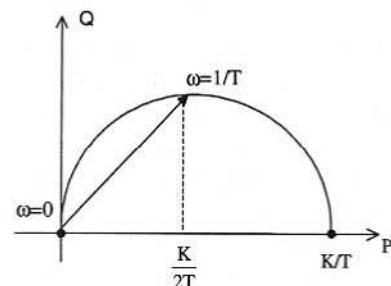
Przykład 4.19

Człon różniczkującoinercyjny (rys. 4.42)

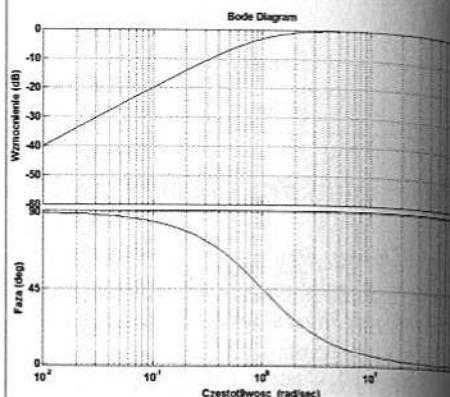
$$G(s) = \frac{Ks}{Ts+1}, \quad \operatorname{tg} \varphi(\omega) = 1/(\omega T),$$

$$G(j\omega) = \frac{K\omega^2 T}{1+\omega^2 T^2} + j \frac{K\omega}{1+\omega^2 T^2} =$$

$$= \frac{K\omega}{\sqrt{1+\omega^2 T^2}} e^{j\varphi} = |G(\omega)| e^{j\varphi}.$$



Rys. 4.42. Wykres charakterystyki układowego dla członu różniczkującoinercyjnego

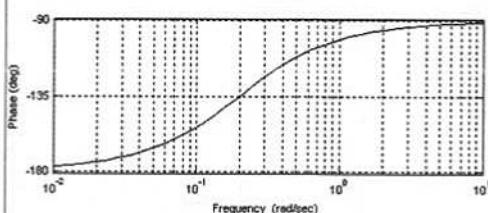
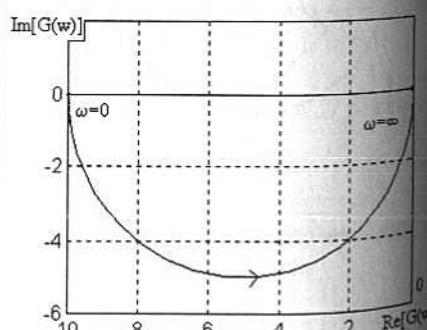
Wykresy Bodego dla $K = 1$ i $T = 1$ przedstawiono na rysunku 4.43.Rys. 4.43. Wykres $|G(\omega)|$ i $\varphi(\omega)$ Dla $T \ll K$ ten człon daje dobrą aproksymację idealnego członu różniczkującego $G(s) = Ks$.**Przykład 4.20**

Dany jest układ inercyjny niestabilny (rys. 4.44)

$$G(s) = \frac{K}{Ts-1} \stackrel{s=j\omega}{=} \frac{-K}{1-jT\omega} = -\frac{K}{1+\omega^2 T^2} - j \frac{K\omega T}{1+\omega^2 T^2}.$$

Charakterystykę Nyquista dla przykładowego układu przedstawiono na rysunku 4.45.

$$G(s) = \frac{10}{5s-1}$$

Rys. 4.44. Charakterystyka Bodego obiektu niestabilnego dla $\omega \in [0,1, 10]$ Rys. 4.45. Charakterystyka Nyquista obiektu niestabilnego dla $\omega \in [0, \infty)$ **Charakterystyka logarytmiczna wzmacnienia**

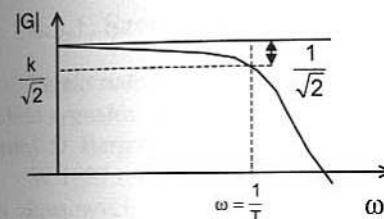
W praktyce używane są charakterystyki logarytmiczne liczone w decybelach [dB]:

$$\log G(j\omega) = \log |G(j\omega)| + j\varphi \log e,$$

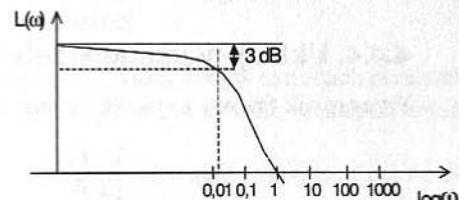
$$L(\omega) = 20 \log |G(j\omega)| \text{ [dB].}$$

Jeżeli $|G(\omega_2)| = 1,12 \cdot |G(\omega_1)|$, to mówi się, że różnica między nimi wynosi 1 dB. Dla pewnych przedziałów ωT można posługiwać się uproszczonym zapisem:

- $\omega T \gg 1$; $20 \log \sqrt{1+\omega^2 T^2} \cong 20 \log \omega T = 20 \log \omega + 20 \log T$,
- $\omega T \ll 1$; $20 \log \sqrt{1+\omega^2 T^2} \cong 20 \log 1 = 0$.

Szybkie, ale uproszczone, rysowanie charakterystyk logarytmicznych realizuje się odcinkami prostymi pomiędzy częstotliwościami granicznymi $\omega_g = 1/T$ (rys. 4.46 i 4.47).

Rys. 4.46. Charakterystyka Bodego



Rys. 4.47. Logarytmiczna charakterystyka Bodego

Przykład 4.21

$$G(s) = \frac{5(0,5s+1)}{(s+1)(0,25s+1)}, \quad T_1 = 1, \quad \omega_{g1} = \frac{1}{T} = 1,$$

$$T_2 = 0,5, \quad \omega_{g2} = 2,$$

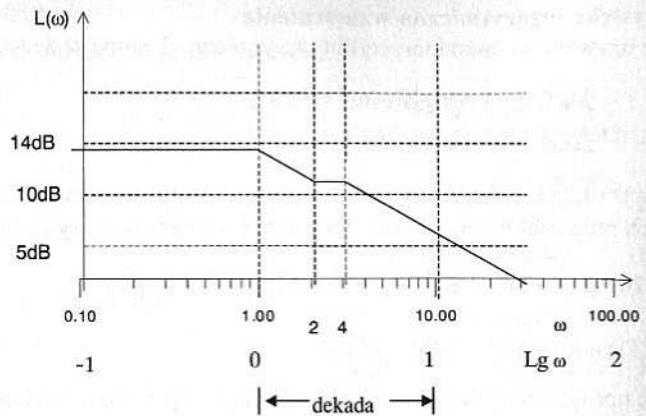
$$L_0(\omega) = 20 \log 5 = 14 \text{ dB}, \quad T_3 = 0,25, \quad \omega_{g3} = 4.$$

Tabela 4.3

Częstotliwości graniczne i spadki charakterystyki

Częstotliwość ω_g	0,1÷1	1÷2	2÷4	4÷∞
Różniczkowanie	0	0	+20	+20
Inercja 1	0	-20	-20	-20
Inercja 2	0	0	0	-20
Suma	0	-20	0	-20

Dla członów inercyjnych zachodzi spadek charakterystyki o 20 dB/dek., a dla członów różniczkujących przyrost o 20 dB/dek., (tab. 4.3 i rys. 4.48).



Rys. 4.48. Charakterystyka logarytmiczna uproszczona

4.4.4. Układy nieminimalnofazowe

Przesuwnik fazowy to układ o transmitancji

$$G_{PF}(s) = \frac{1-Ts}{1+Ts}.$$

Moduł jego transmitancji w całym zakresie częstotliwości jest stały i równy jeden. Faza zmienia się jednak od $0 \div 180^\circ$ dla $\omega \in [0, \infty)$ (patrz przykład 4.11)

$$|G_{PF}(\omega)| = \frac{1+\omega^2 T^2}{1+\omega^2 T^2} = 1.$$

Układy, których transmitancje zawierają w sobie jedną lub kilka transmitancji przesuwników fazowych, nazywają się układami nieminimalnofazowymi $G_{NF}(s)$.

$$G_{NF}(s) = G_{MF}(s) \cdot G_{PF}(s),$$

$$|G_{NF}(\omega)| = |G_{MF}(\omega)|, \quad \arg(G_{NF}(\omega)) \neq \arg(G_{MF}(\omega)).$$

Typowymi układami nieminimalnofazowymi są układy z dodatnimi biegunami (niestabilne) lub z dodatnimi zerami. Te ostatnie często występują w strukturach równoległych technologii przemysłowych, gdy wyjście jest wynikiem odejmowania się wyjść dwóch innych układów dynamicznych.

Poniższa transmitancja zastępcza $G(s)$ wykazuje cechę nieminimalnofazowości

$$G(s) = \frac{3}{8s^3 + 14s^2 + 7s + 1} - \frac{4s}{8s^2 + 6s + 1} = \frac{(3+2s)(1-2s)}{(1+4s)(1+s)(1+2s)} = G_{MF}(s) \cdot G_{PF}(s).$$

Ponieważ nieminimalnofazowość zależy od parametrów układu, to zmieniając parametry, można zmieniać tę cechę.

Na przykład układ

$$G(s) = \frac{K}{1+Ts} - \frac{1}{s+1} = \frac{K-1+s(K-T)}{(1+Ts)(1+s)}$$

jest nieminimalnofazowy tylko wtedy, jeśli: $T > 1$, a $1 < K < T$, oraz jeśli $T < 1$, a $T < K < 1$.

Odpowiedź na skok jednostkowy układów nieminimalnofazowych może wykazywać charakterystyczny przebieg: w pierwszej fazie procesu przejściowego odpowiedź ma znak przeciwny do znaku sterowania. Stąd układy takie należą do trudniejszych obiektów sterowania niż układy minimalnofazowe.

4.5. Przekształcenie Fouriera

4.5.1. Szereg Fouriera dla funkcji okresowej

Przekształcenie Fouriera i szereg Fouriera odgrywają ważną rolę w metodach przetwarzania sygnałów i w automatyce. Zagadnienie te zapoczątkował w swoich badaniach drgan struny D. Bernoulli w 1753 r.

W 1807 r. J. Fourier przedstawił pierwsze twierdzenia związane z rozkładaniem funkcji okresowych w szeregi trygonometryczne. Zagadnienia te badał również w swoich pracach Dirichlet (1829 r.) [4BRR].

Jeśli funkcja $f(t)$ jest okresowa o okresie T (tzn. powtarza się z częstotliwością $\omega = 2\pi/T$) i w przedziale o długości T ma ograniczone wahanie, to można ją przedstawić nieskończonym szeregiem trygonometrycznym Fouriera

$$f(t) = 0,5 a_0 + \sum_{i=1}^{\infty} [a_i \cos(i\omega t) + b_i \sin(i\omega t)] = 0,5 a_0 + \sum_{i=1}^{\infty} A_i \cos(i\omega t + \varphi_i).$$

Dla $\omega = 2\pi/T$ poszczególne współczynniki rozkładu dane są wzorami:

$$a_i(T) = \frac{2}{T} \int_0^T f(\tau) \cos(i \frac{2\pi}{T} \tau) d\tau, \quad b_i(T) = \frac{2}{T} \int_0^T f(\tau) \sin(i \frac{2\pi}{T} \tau) d\tau \quad (4.13)$$

$$A_i = \sqrt{a_i^2 + b_i^2}, \quad \cos(\varphi_i) = \frac{a_i}{A_i}$$

Dla funkcji $f(t)$ parzystej współczynniki $b_i = 0$, a dla funkcji nieparzystej $a_i = 0$. Wprowadzając nowy zespolony współczynnik $c_i(T)$

$$c_i = \frac{a_i - jb_i}{2} = \frac{1}{T} \int_0^T f(\tau) e^{-j\omega \tau} d\tau,$$

można szereg Fouriera przedstawić w uogólnionej postaci wykładniczej

$$f(t) = \sum_{i=-\infty}^{\infty} c_i(T) e^{j i \omega t} = \sum_{i=-\infty}^{\infty} \left(\frac{1}{T} \int_0^T f(\tau) e^{-j i \omega \tau} d\tau \right) e^{j i \omega t}.$$

Współczynniki szeregu Fouriera są zwane widmem sygnału i mogą być wyrysowane na osi częstotliwości, tworząc charakterystyczny dyskretny wykres nieskończonej liczby prążków. Wynika z tego, że tylko dla funkcji z widmem ograniczonym (o skończonej liczbie prążków) istnieje pewność, że znajomość wszystkich prążków pozwoli na odtworzenie sygnału oryginalnego z całkowitą dokładnością. Z reguły sygnały fizyczne mają widmo nieskończone, stąd dla celów praktycznych pomija się prążki, poczawszy od pewnej częstotliwości. Ze wzorów widać, że gdy zwiększy się okres T funkcji n razy, to prążki w widmie wystąpią n -krotnie gęściej i będą one n -krotnie mniejsze. Ma to ostatecznie związek z przeaniem widma dyskretnego w widmo ciągłe oraz zamiany sum występujących we wzorach w całki, a funkcji okresowej $f(t)$ w funkcję nieokresową.

4.5.2. Wzór całkowy Fouriera dla funkcji nieokresowej

Jeśli nieokresowa funkcja $f(t)$ jest bezwzględnie całkowalna w przedziale $(-\infty, +\infty)$, tzn.

$$\int_{-\infty}^{\infty} |f(t)| dt < \infty$$

i jest rozwijalna w szereg Fouriera w każdym przedziale ograniczonym, tzn. ma na nim wahanie ograniczone, to funkcję można przedstawić jako

$$f(t) = \int_0^{\infty} [a(\omega) \cos(\omega t) + b(\omega) \sin(\omega t)] d\omega = \int_0^{\infty} A(\omega) \cos[\omega t + \phi(\omega)] d\omega$$

gdzie współczynniki rozkładu dane są wzorami:

$$a(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} f(\tau) \cos(\omega \tau) d\tau, \quad b(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} f(\tau) \sin(\omega \tau) d\tau,$$

$$A(\omega) = \sqrt{a^2(\omega) + b^2(\omega)}, \quad \cos(\phi(\omega)) = \frac{a(\omega)}{A(\omega)}.$$

Ogólniejsza postać zespolona definiuje dwa przekształcenia całkowe dla ω [rad/s] i f [Hz]

$$\begin{aligned} f(t) &= \int_{-\infty}^{\infty} \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} f(\tau) e^{-j\omega\tau} d\tau \cdot e^{j\omega t} \right] d\omega = \int_{-\infty}^{\infty} F(j2\pi f) e^{j2\pi f t} df = \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} F(j\omega) e^{j\omega t} d\omega \end{aligned}$$

Funkcja $F(j\omega)$ nazywa się transformatą Fouriera funkcji $f(t)$ lub jej gęstością widmową.

Jeżeli funkcja $f(t)$ jest bezwzględnie całkowalna w $(-\infty, +\infty)$, to istnieje jej dwustronna transformata Fouriera $F(j\omega)$ dla wszystkich ω . Dokładniej – muszą być spełnione warunki Dirichleta, tzn. wartości maksymalne i minimalne funkcji muszą być skończone i liczba punktów nieciągłości w skończonym przedziale musi być skończona. Jeżeli $f(t) = 0$ dla $t < 0$, można stosować transformatę jednostronną, dającą równanie analizy częstotliwościowej dla funkcji $f(t)$.

$$F(j\omega) = \int_0^{\infty} f(\tau) e^{-j\omega\tau} d\tau \quad (4.14)$$

Jeżeli dodatkowo $f(t)$ ma ograniczone wahanie, to dla $F(j\omega)$ istnieje transformata odwrotna dającą $f(t)$. Wiele prostych funkcji takich, jak $\delta(t)$, $1(t)$, $\sin(t)$, e^t nie ma transformaty Fouriera, bo ich całki są rozbieżne. Jest to istotna różnica w stosunku do transformaty Laplace'a.

Interpretacja

Dla funkcji okresowej współczynniki A_i szeregu Fouriera reprezentują amplitudy kolejnych harmonicznych. Na osi dyskretnych częstotliwości można zaznaczyć nieskończoną liczbę tych amplitud tworzących widmo. Z reguły dla rosnących częstotliwości amplitudy są malejące i praktycznie od pewnej częstotliwości są do pominięcia. W takich przypadkach znaczące harmoniczne obejmują tylko skończone pasmo na osi częstotliwości. Fizyczna interpretacja mówi, że każdy mierzony sygnał dociera do obserwatora poprzez pewien system dynamiczny, który ma skończone pasmo przepustowe (*bandwidth*). Dla sygnałów okresowych zbudowanych ze skończonej liczby sinusoid, widmo z definicji jest skończone. Dla sygnałów nieokresowych rolę zespolonego widma dyskretnego przejmuje transformata Fouriera, która reprezentuje ciągłe widmo zwane gęstością widmową (spektralną) sygnału $f(t)$. Jednostką gęstości widmowej $F(j\omega)$ jest $t \cdot f(t)$, np. $f(t)$ [V], $F(j\omega)$ [V·s] lub [V/Hz]. Stąd nazwa gęstość widmowa, a nie widmo.

4.5.3. Niektóre własności transformaty Fouriera

Poniżej przedstawiamy kilka własności transformaty Fouriera.

- 1) Jeżeli funkcje $f_1(t)$ i $f_2(t)$ są całkowalne z kwadratem na $(-\infty, +\infty)$, to transformata \mathcal{F} splotu tych funkcji równa się iloczynowi ich transformat

$$\mathcal{F} \left[\int_{-\infty}^{\infty} f_1(\tau) f_2(t - \tau) d\tau \right] = F_1(j\omega) \cdot F_2(j\omega).$$

Stąd:

$$\mathcal{F}[y(t)] = \mathcal{F}[g(t)] \mathcal{F}[u(t)] \quad \text{lub} \quad y(t) = \mathcal{F}^{-1}[G(j\omega) \mathcal{F}(u(t))].$$

- 2) Transformata funkcji przesuniętej w dziedzinie czasu

$$\mathcal{F}[f(t - \tau)] = e^{-j\omega\tau} \cdot F(j\omega).$$

- 3) Transformata funkcji mnożonej przez eksponentę

$$\mathcal{F}[f(t) \cdot e^{-j\hat{\omega}t}] = F[j(\omega - \hat{\omega})].$$

Stąd

$$\frac{F[j(\omega - \hat{\omega})] + F[j(\omega + \hat{\omega})]}{2} = \mathcal{F}[f(t) \cdot \cos(\hat{\omega}t)].$$

Własność ta ma zastosowanie w teorii modulacji amplitudowej sygnału $\cos(\omega t)$ przenoszonego nieokresowym $f(t)$.

- 4) Wzór Rayleigha dla sygnałów o skończonej energii wynika z własności splotu

$$\int_{-\infty}^{\infty} [f(t)]^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(j\omega)|^2 d\omega.$$

Dla takich rzeczywistych, fizycznych sygnałów o skończonej energii na nieskończonym przedziale czasu $(-\infty, +\infty)$ znormalizowana moc równa się zero

$$\text{Energia: } \int_{-\infty}^{\infty} f^2(t) dt < \infty \Rightarrow \text{Moc: } M_Z = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f^2(t) dt = 0.$$

- 5) Dla sygnałów $f(t)$ o nieskończonej energii, ale o skończonej i różnej od zera mocy

$$M_Z = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f^2(t) dt \neq 0,$$

widmowa gęstość mocy wyraża się wzorem

$$\mathcal{F}[f] = \lim_{T \rightarrow \infty} \frac{1}{T} 2 |\mathcal{F}_r(j2\pi f)|^2,$$

$$\text{gdzie } \mathcal{F}_r(j\omega) = \int_{-T/2}^{T/2} f(t) e^{-j\omega t} dt.$$

- 6) Całkowita moc sygnału okresowego $f(t)$ wynika z wzoru Parsevala dla szeregu Fouriera

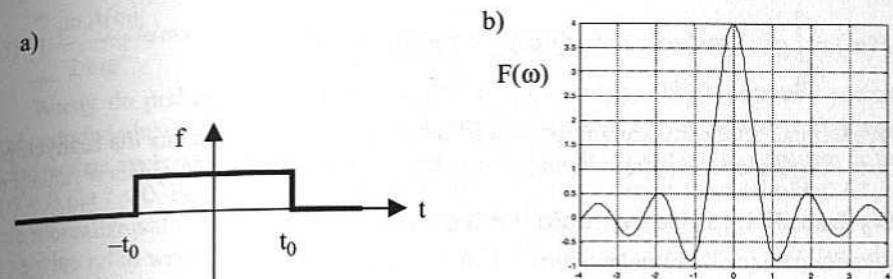
$$M = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f^2(t) dt = A_0^2 + 0,5 \sum_{i=1}^{\infty} A_i^2.$$

- 7) Jeżeli $f(t) = 0$ dla $t < 0$ i $\int_0^{\infty} |f(t)| dt < \infty$, to transformata Fouriera przechodzi w transformatę Laplace'a przez podstawienie $j\omega = s$; $\mathcal{F}(j\omega) = \mathcal{F}(s)$.

Przykład 4.22

Czasowe okno prostokątne (rys. 4.49)

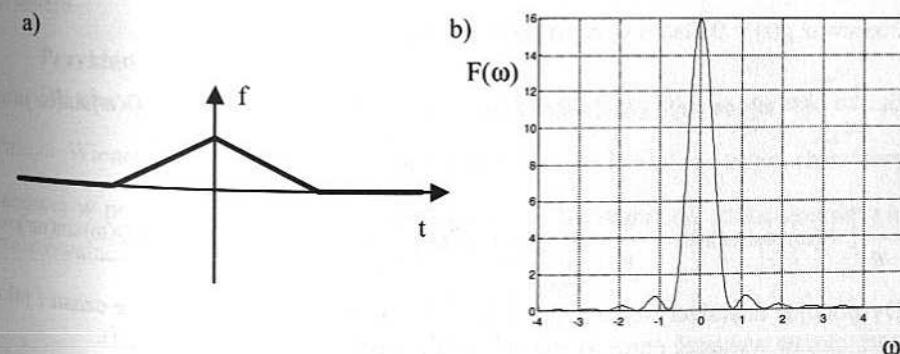
$$f(t) = \begin{cases} 1 & |t| < t_0 \\ 0 & |t| > t_0 \end{cases}, \quad F(j\omega) = 2 \frac{\sin(\omega t_0)}{\omega}.$$



Rys. 4.49. Funkcje f (a) i ich transformaty F (b)

Czasowe okno trójkątne (rys. 4.50)

$$f(t) = \begin{cases} t + 2t_0 & -2t_0 < t < 0 \\ 2t_0 - t & 0 \leq t \leq 2t_0 \\ 0 & 2t_0 < |t| \end{cases}, \quad F(j\omega) = 4 \left[\frac{\sin(\omega t_0)}{\omega} \right]^2.$$



Rys. 4.50. Funkcje f (a) i ich transformaty F (b)

4.5.4. Związki pomiędzy charakterystykami częstotliwościowymi i impulsowymi

Wyliczanie charakterystyk częstotliwościowych

Z transmitancji operatorowej można otrzymać transmitancję widmową przez podstawienie $s = j\omega$. Jest to formalne przejście z transformaty Laplace'a na transformatę Fouriera:

$$G(j\omega) = P(\omega) + jQ(\omega) = \int_0^\infty g(\tau) e^{-j\omega\tau} d\tau =$$

$$= \int_0^\infty g(\tau) \cos(\omega\tau) d\tau - j \int_0^\infty g(\tau) \sin(\omega\tau) d\tau \quad (4.15)$$

$$P(\omega) = \int_0^\infty g(\tau) \cos(\omega\tau) d\tau; \quad Q(\omega) = - \int_0^\infty g(\tau) \sin(\omega\tau) d\tau$$

Dysponując zarejestrowaną impulsową funkcją przejścia $g(t)$, można dla różnych parametrów ω wykonać numeryczne obliczenia całek i otrzymać wykres Nyquista w funkcji $g(t)$.

Wyliczanie impulsowej funkcji przejścia

Wykorzystując odwrotne przekształcenie Fouriera, otrzymuje się

$$g(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(j\omega) e^{j\omega t} d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} [P(\omega) + jQ(\omega)][\cos(\omega t) + j\sin(\omega t)] d\omega.$$

Po wymnożeniu funkcji podcałkowych i uwzględnieniu parzystości funkcji $P(\omega)$ i $\cos(\omega t)$ oraz nieparzystości $Q(\omega)$ i $\sin(\omega t)$ powyższy wzór upraszcza się do

$$g(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} [P(\omega) \cos(\omega t) - Q(\omega) \sin(\omega t)] d\omega \quad (4.16)$$

Ponieważ $g(t) = 0$ dla $t < 0$, czyli $g(-t) = 0$ dla $t > 0$:

$$g(-t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} [P(\omega) \cos(\omega t) + Q(\omega) \sin(\omega t)] d\omega = 0 \Rightarrow \int_{-\infty}^{\infty} P(\omega) \cos(\omega t) d\omega = - \int_{-\infty}^{\infty} Q(\omega) \sin(\omega t) d\omega$$

$$g(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} P(\omega) \cos(\omega t) d\omega = - \frac{1}{\pi} \int_{-\infty}^{\infty} Q(\omega) \sin(\omega t) d\omega = \frac{2}{\pi} \int_0^{\infty} P(\omega) \cos(\omega t) d\omega = - \frac{2}{\pi} \int_0^{\infty} Q(\omega) \sin(\omega t) d\omega$$

Dysponując charakterystykami częstotliwościowymi i wybierając chwilę czasu t jako parametr, można wykonać numeryczne obliczenia całek i otrzymać wykres $g(t)$.

4.5.5. Częstotliwościowe kryterium realizowalności fizycznej

W podrozdziałach 3.6 i 4.3.1 omawiano już warunki fizycznej realizowalności układu liniowego. Warunki te sprawdzają się do zerowania impulsowej funkcji przejścia dla czasów ujemnych $g(t) = 0$ dla $t < 0$. Teraz podane zostanie dodatkowe twierdzenie precyzujące w dziedzinie częstotliwości warunki, jakie musi spełniać amplitudowa charakterystyka częstotliwościowa, aby mogła reprezentować fizycznie realizowalny liniowy układ dynamiczny

Twierdzenie Paleya–Wienera [4GOH1, 4OSJ]

Układ liniowy stacjonarny o zadanej charakterystyce amplitudowej $A(\omega)$ może być zrealizowany wtedy i tylko wtedy, gdy zachodzą trzy warunki:

- 1) parzysta $A(\omega) \geq 0$,
- 2) $A(\omega) \in L^2(-\infty, \infty)$,
- 3) $\int_{-\infty}^{\infty} \frac{|\ln A(\omega)|}{1+\omega^2} d\omega < \infty$.

Wtedy do praktycznej realizacji układu $G(j\omega) = A(\omega) e^{-j\phi(\omega)}$ potrzeba jeszcze dobrać odpowiednią charakterystykę fazową $\phi(\omega)$, ale jest to na pewno możliwe. Taki układ będzie spełniał warunek realizowalności w dziedzinie czasu, tzn. impulsowa funkcja $g(t) = 0$ dla $t < 0$. Przy doborze nieodpowiedniej charakterystyki fazowej układ $G(j\omega)$ będzie nierealizowalny, pomimo spełniania warunków twierdzenia Paleya–Wienera. Warunki te są więc tylko warunkami koniecznymi realizowalności, ale nie wystarczającymi do tego, aby $\mathcal{F}^{-1}[G(j\omega)] = 0$ dla $t < 0$.

Ważny wniosek: Nie jest realizowalny żaden układ o charakterystyce amplitudowej stale równej zero $A(\omega) = 0$ w pewnym przedziale częstotliwości $[\omega_1, \omega_2]$, bo wtedy $|\ln A(\omega)| = \infty$ i cała w trzecim warunku Paleya–Wienera jest rozbieżna. Nie można więc zbudować idealnego ciągłego (analogowego) filtra dolnoprzepustowego o prostokątnej charakterystyce amplitudowej na pewnym zakresie częstotliwości i zerującej się poza tym zakresem. Zerowanie się charakterystyki amplitudowej w odsobnionych punktach ω_i jest już dopuszczalne, bo nie powoduje rozbieżności całki.

Z powyższych rozważań wynika prawdziwość twierdzenia odwrotnego: charakterystyka amplitudowa każdego układu realizowalnego spełnia warunki twierdzenia Paleya–Wienera.

Przykład 4.23 [4OSJ]

Dla charakterystyki amplitudowej $A(\omega) = a/(\omega^2+1)$, która spełnia warunki twierdzenia Paleya–Wienera, bo $\int_{-\infty}^{\infty} \left(|\ln A(\omega)| / (1+\omega^2) \right) d\omega = \pi(|\ln a| - \ln 4) < \infty$, dobór charakterystyki fazowej w postaci $\phi_1(\omega) = 2 \operatorname{arctg}(\omega)$ prowadzi do transmitancji, która spełnia warunek realizowalności, bo: $G_1(j\omega) = A(\omega) e^{-j\phi_1(\omega)} = \frac{a}{(1+j\omega)^2}$,

$$g_1(t) = \mathcal{F}^{-1}[G_1(j\omega)] = at \cdot e^{-t} \cdot 1(t) = 0 \quad \text{dla } t < 0.$$

Dla innej charakterystyki fazowej, np. w postaci $\phi_2(\omega) = 0$ dla każdego ω , prowadzi to do transmitancji $G_2(j\omega) = A(\omega) \cdot e^0 = A(\omega)$. Odwrotna transformata Fouriera (podrozdz. 4.5.2) ma wtedy postać, która nie spełnia warunku realizowalności:

$$g_2(t) = \mathcal{F}^{-1}[G_2(j\omega)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{a e^{j\omega t}}{\omega^2 + 1} d\omega = \frac{a}{2} e^{-|t|} \neq 0 \quad \text{dla } t < 0.$$

4.6. Analiza własności obiektu II rzędu

4.6.1. Odpowiedzi skokowe

Dane jest równanie różniczkowe opisujące dynamikę układu II rzędu

$$a_2 \ddot{y}(t) + a_1 \dot{y}(t) + a_0 y(t) = b_1 \dot{u}(t) + b_0 u(t)$$

z zerowymi warunkami początkowymi.

Dokonując przekształcenia Laplace'a, otrzymuje się równanie operatorowe

$$a_2 s^2 Y(s) + a_1 s Y(s) + a_0 Y(s) = b_1 s U(s) + b_0 U(s)$$

oraz transmitancję w znormalizowanej postaci

$$G(s) = \frac{Y(s)}{U(s)} = \frac{b_1 s + b_0}{a_2 s^2 + a_1 s + a_0} = \frac{\omega^2 b_1}{a_0} \left[s + \frac{1}{\alpha} \right] \quad (4.17)$$

gdzie: $\alpha = \frac{b_1}{b_0}$, $0 \leq \alpha \leq \infty$,

$\omega = \sqrt{\frac{a_0}{a_2}}$ – częstotliwość drgań własnych,

$\xi = \frac{1}{2} \frac{a_1}{\sqrt{a_0 a_2}}$ – współczynnik tłumienia.

Dla różnych wartości współczynnika tłumienia układ wykazuje różne cechy. Można rozróżnić tutaj cztery przypadki:

- 1) $0 \leq \xi < 1$ – układ oscylacyjny tłumiony,
- 2) $\xi = 0$ – układ oscylacyjny,
- 3) $\xi = 1$ – układ z tłumieniem krytycznym,
- 4) $\xi > 1$ – układ przetłumiony.

Przypadek 1)

Dla współczynnika $0 \leq \xi < 1$ pierwiastki mianownika są liczbami zespolonymi z ujemną częścią rzeczywistą i reprezentują biegury transmitancji dane wzorami $s_{1,2} = -\xi \omega \pm j\omega \sqrt{1-\xi^2}$. Wprowadza się następujące oznaczenia na katy:

$$\cos \Phi = \xi, \quad \sin \Phi = \sqrt{1-\xi^2} \quad \rightarrow \quad \operatorname{tg} \Phi = \frac{\sqrt{1-\xi^2}}{\xi},$$

$$\operatorname{tg} \Omega = \frac{\sqrt{1-\xi^2}}{\alpha \omega - \xi}, \quad \operatorname{tg} \theta = \frac{\omega \sqrt{1-\xi^2}}{\xi \omega - \frac{1}{\alpha}}, \quad \theta = \Omega + \Phi.$$

4.6. Analiza własności obiektu II rzędu

Wartość ustaloną odpowiedzi $y(t)$ dla sygnału wejściowego $u(t) = 1(t)$ przy $t \rightarrow \infty$ oblicza się z twierdzenia granicznego

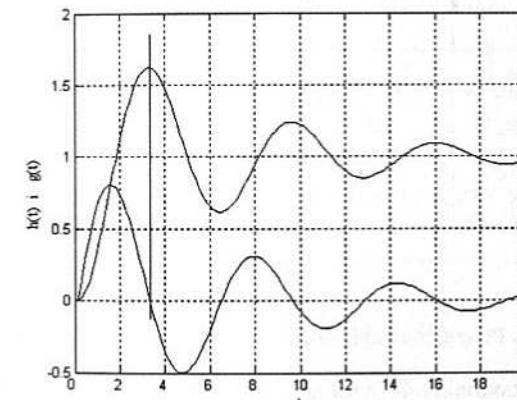
$$\lim_{t \rightarrow \infty} y(t) = \lim_{s \rightarrow 0} s \frac{1}{s} \frac{b_1 s + b_0}{a_2 s^2 + a_1 s + a_0} = \frac{b_0}{a_0} = h_\infty.$$

Odpowiedź $y(t)$ na skok jednostkowy (oznaczana przez $h(t)$) dana jest wzorem

$$h(t) = h_\infty \left[1 + \frac{\sqrt{\alpha^2 \omega^2 - 2\alpha \omega \xi + 1}}{\sqrt{1-\xi^2}} \cdot e^{-\xi \omega t} \cdot \sin(\omega \sqrt{1-\xi^2} t - \Omega) \right]$$

lub

$$h(t) = h_\infty \left[1 + \frac{e^{-\xi \omega t}}{\sin \Omega} \cdot \sin(\omega \sqrt{1-\xi^2} t - \Omega) \right].$$



Rys. 4.51. Jakościowe przebiegi odpowiedzi skokowej $h(t)$ i impulsowej funkcji przejścia $g(t)$

Odpowiedź układu na sygnał $u(t) = \delta(t)$ daje postać impulsowej funkcji przejścia $y(t) = g(t)$ (rys. 4.51):

$$g(t) = \mathcal{L}^{-1}[G(s)],$$

$$g(t) = h(t) = \frac{dh(t)}{dt} = -A e^{-\xi \omega t} \omega \sin(\beta t - \theta),$$

gdzie: $A = h_\infty \sqrt{\frac{\alpha^2 \omega^2 - 2\alpha \omega \xi + 1}{1-\xi^2}}$, $\beta = \omega \sqrt{1-\xi^2}$ – częstotliwość drgań swobodnych.

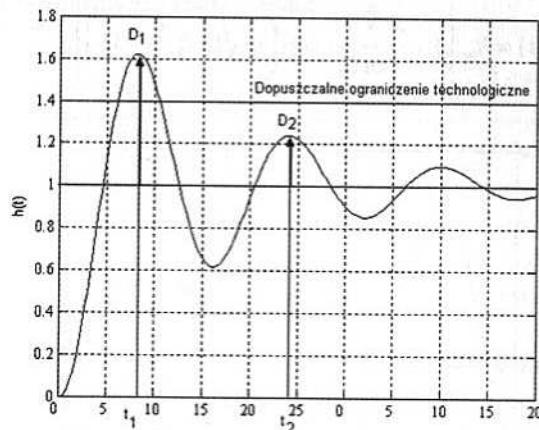
W szczególnym przypadku, jeśli: $b_1 = 0 \rightarrow \alpha = 0 \rightarrow G(s) = \frac{b_0}{a_2 s^2 + a_1 s + a_0}$,

wtedy: $\theta = \pi$, $\Omega = \pi - \Phi$, $\operatorname{tg} \Phi = \frac{\sqrt{1-\xi^2}}{\xi} = -\operatorname{tg} \Omega$,

$$h(t) = h_{\infty} \left[1 - \frac{1}{\sqrt{1-\xi^2}} e^{-\xi \omega t} \cdot \sin(\omega \sqrt{1-\xi^2} t + \Phi) \right]$$

$$g(t) = h_{\infty} \left[\frac{\omega}{\omega \sqrt{1-\xi^2}} e^{-\xi \omega t} \cdot \sin \beta t \right] \quad (4.18)$$

Warto zwrócić uwagę, że częstotliwość β zanikających drgań swobodnych impulsowej funkcji przejścia $g(t)$ (4.18) nie równa się częstotliwości drgań własnych ω . Równość zachodzi tylko dla obiektu ze współczynnikiem tłumienia $\xi = 0$.



Rys. 4.52. Przeregulowania odpowiedzi skokowej układu II rzędu

Dla określenia maksymalnych przeregulowań, poszukuje się zer impulsowej funkcji przejścia

$$g(t) = 0 \Rightarrow t_1, t_2, t_3, \dots,$$

gdzie:

$$t_1 = \frac{\pi}{\omega \sqrt{1-\xi^2}},$$

$$t_2 = \frac{3\pi}{\omega \sqrt{1-\xi^2}}.$$

Maksymalna wartość odpowiedzi skokowej i pierwsze przeregulowanie (rys. 4.52) wynoszą:

$$h(t_1) = h_{\infty} \left(1 + e^{-\frac{\xi}{\sqrt{1-\xi^2}} \cdot \pi} \right) \Rightarrow D_1 = h_{\infty} e^{-\frac{\xi}{\sqrt{1-\xi^2}} \cdot \pi} \Rightarrow \begin{cases} \xi = 0: & D_1 = h_{\infty} \\ \xi = \frac{1}{\sqrt{2}}: & D_1 = h_{\infty} e^{-\pi} \\ \xi = 1: & D_1 = 0 \end{cases}$$

Drugie przeregulowanie

$$D_2 = h_{\infty} e^{-\frac{\xi}{\sqrt{1-\xi^2}} \cdot 3\pi}$$

Praktycznym kryterium jakości przebiegu odpowiedzi układu na skok jednostkowy bywa kryterium „tłumienia do 1/4”, tzn. taki współczynnik tłumienia, dla którego $D_2 = D_1/4$

$$e^{-\frac{\xi}{\sqrt{1-\xi^2}} \cdot 3\pi} = 0,25 \cdot e^{-\frac{\xi}{\sqrt{1-\xi^2}} \cdot \pi}$$

Po zlogarytmowaniu obu stron otrzymuje się równanie

$$\frac{\xi}{\sqrt{1-\xi^2}} = \frac{\ln 4}{2\pi} = 0,22 = m.$$

Stąd wynika, że dający „tłumienie do 1/4” współczynnik $\xi = \frac{m}{\sqrt{1+m^2}} = 0,2154$.

Przypadek 2)

W szczególnym przypadku oscylatora harmonicznego, gdy współczynnik tłumienia $\xi = 0$, wystąpią drgania harmoniczne nietłumione spowodowane biegunami urojonymi $s_{1,2} = \pm j\omega$:

$$h(t) = h_{\infty} \left[1 + \sqrt{\alpha^2 \omega^2 + 1} \sin(\omega t - \Omega) \right],$$

$$g(t) = -h_{\infty} \sqrt{\alpha^2 \omega^2 + 1} \omega \sin(\omega t - \Theta).$$

Przypadek 3)

Dla krytycznego współczynnika tłumienia $\xi = 1$ występuje podwójny biegun rzeczywisty $s_{1,2} = -\omega$. Odpowiedzi układu II rzędu są dane wzorami:

$$h(t) = h_{\infty} \left[1 + e^{-\omega t} (t\omega(\alpha\omega - 1) - 1) \right]$$

$$g(t) = h_{\infty} e^{-\omega t} [(1 - \alpha\omega)t + \alpha] \omega^2 \quad (4.19)$$

Przypadek 4)

Dla $\xi > 1$ występują dwa pierwiastki rzeczywiste różne, zawsze ujemne

$$s_{1,2} = -\xi\omega \pm \omega\sqrt{\xi^2 - 1}.$$

Odpowiedzi na sterowania $u(t) = 1(t)$ i $u(t) = \delta(t)$ są dane wzorami:

$$\begin{aligned} h(t) &= h_\infty \left[1 + e^{-\xi \omega t} \cdot \left(\frac{\alpha \omega - \xi}{\sqrt{\xi^2 - 1}} \operatorname{sh}(\beta t) - \operatorname{ch}(\beta t) \right) \right], \quad \beta = \omega \sqrt{\xi^2 - 1} \\ g(t) &= h_\infty \cdot \frac{\omega e^{-\xi \omega t}}{\sqrt{\xi^2 - 1}} \left[(1 - \xi \omega \alpha) \operatorname{sh}(\beta t) + \alpha \beta \operatorname{ch}(\beta t) \right] \end{aligned} \quad (4.20)$$

W rozważanym przypadku, przy innym niż we wzorze (4.17) zapisie transmitancji przedstawiającym człon różniczkujący z podwójną inercją i stałymi czasowymi T_1 i T_2

$$G(s) = \frac{K(T_0 s + 1)}{(T_1 s + 1)(T_2 s + 1)},$$

odpowiedź na skok jednostkowy dana jest wzorem

$$y(t) = h(t) = K \left(1 + \frac{T_0 - T_1}{T_1 - T_2} e^{-t/T_1} + \frac{T_0 - T_2}{T_2 - T_1} e^{-t/T_2} \right) \quad (4.21)$$

Jest to inna postać wzoru (4.20).

4.6.2. Charakterystyka częstotliwościowa obiektu II rzędu

Dana jest transmitancja II rzędu (wzór (4.17) dla $\alpha = 0$)

$$G(s) = \frac{K \omega_0^2}{s^2 + 2\xi \omega_0 s + \omega_0^2}.$$

We wzorze tym przez ω_0 oznaczono częstotliwość drgań własnych, a przez ω – częstotliwość sygnału wejściowego $u(t) = A \sin(\omega t)$. Odpowiedź ustalona sinusoidalna na takie wymuszenie ma postać:

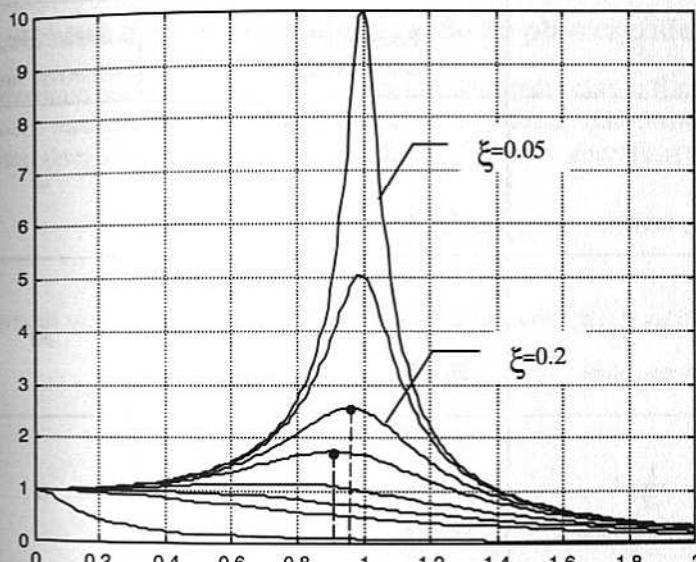
$$y(t) = |G(\omega)| \cdot A \sin(\omega t + \varphi(\omega)),$$

$$|G(\omega)| = \frac{K}{\sqrt{\left[1 - (\omega/\omega_0)^2\right]^2 + (2\xi\omega/\omega_0)^2}} = \frac{K}{\sqrt{\bar{\omega}^4 - 2\bar{\omega}^2(1 - 2\xi^2) + 1}}$$

$$\bar{\omega} = \frac{\omega}{\omega_0}, \quad \varphi = -\operatorname{tg}^{-1} \left[\frac{2\xi\bar{\omega}}{1 - \bar{\omega}^2} \right].$$

Maksimum modułu $|G(\omega)|$ w funkcji ω występuje przy pewnej wartości częstotliwości, zwanej częstotliwością rezonansową ω_r . Jeśli więc sygnał sinusoidalny na wejściu będzie miał tę częstotliwość, to zostanie on wzmacniony maksymalnie. Tę częstotliwość można wyliczyć, przyrównując do zera pochodną modułu po częstotliwości względnej $\bar{\omega} = \omega/\omega_0$.

Tylko dla $0 < \xi < 1/\sqrt{2}$ otrzymuje się różną od zera częstotliwość rezonansową, daną wzorem $\omega_r = \omega_0 \sqrt{1 - 2\xi^2}$. Jest ona mniejsza od częstotliwości drgań swobodnych $\beta = \omega_0 \sqrt{1 - \xi^2}$ i mniejsza od częstotliwości drgań własnych ω_0 ($\omega_r < \beta < \omega_0$).



Rys. 4.53. Charakterystyki modułu kolejno dla $\xi = 0,05; 0,1; 0,2; 0,3; 0,5; 0,707; 1,0; 5,0$ i $K = 1$

Dla częstotliwości rezonansowej na wejściu $\omega = \omega_r$, i dla $0 < \xi < 1/\sqrt{2} \approx 0,707$ występuje więc maksimum charakterystyki amplitudowej, czyli maksymalny efekt wzmacniania modułu sygnału wejściowego (rys. 4.53).

Wtedy

$$|G|_{\max} = |G(\omega_r)| = \frac{K}{2\xi\sqrt{1-\xi^2}} \Rightarrow \text{dla } 0 < \xi < 1/\sqrt{2} \Rightarrow |G|_{\max} > K.$$

Tylko dla obiektu z $\xi = 0$ częstotliwość rezonansowa jest równa częstotliwości drgań własnych i swobodnych $\omega_r = \omega_0 = \beta$. Wzmocnienie sygnału wejściowego o takiej częstotliwości $\omega = \omega_r$ wynosi wtedy $|G(\omega_r)| = \infty$. Dla takiego obiektu ($\xi = 0$) charakterystyka Nyquistowa jest zdegenerowana tylko do części rzeczywistej $G(\bar{\omega}) = K/(1 - \bar{\omega}^2)$.

Na charakterystyce częstotliwościowej w zakresie $1/\sqrt{2} < \xi < 1$, pomimo jeszcze biegunków zespolonych, maksimum modułu występuje tylko dla $\omega = 0$. Jak wynika z po przednich rozważań (podrozdz. 4.6.1), odpowiedź czasowa obiektu II rzędu na wymuszenie skokowe wykazuje przeregulowania dla wszystkich $\xi < 1$ (bieguny zespolone), np. dla $\xi = 1/\sqrt{2}$, $D_1 = h_\infty e^{-\pi} = h_\infty 0,0432$.

Cechy te ujęto w tabeli 4.4.

Tabela 4.4

Przedziały oscylacyjności dla układu II rzędu

ξ	Przeregulowanie D_1 dla wejścia $u = 1(t)$	$\max G(\omega_r) $ dla $u(t) = \sin(\omega t)$
0	$D_1 = 1$	∞
Biegunki urojone		
$0 + \frac{1}{\sqrt{2}}$	$D_1 \neq 0$	$\neq 0$
Biegunki zespolone		
$\frac{1}{\sqrt{2}} + 1$	$D_1 \neq 0$	0
Biegunki zespolone		
$1 + \infty$	$D_1 = 0$	0
Biegunki rzeczywiste		

Analiza problemu przeregulowań i występowania drgań rezonansowych w układach rzeczywistych jest ważnym problemem technicznym decydującym o jakości tych układów. Zagadnienie to dotyczy zarówno pozornie statycznych obiektów budowlanych, jak i maszyn obrotowych czy np. obiektów latających. Nieprzebadanie przez konstruktorów tego zagadnienia do końca było powodem poważnych katastrof. Na przykład w 1940 r. zerwanie uległ długi most wiszący Narrows Bridge (853×12 m) w Tacoma (USA), gdy częstotliwość podmuchów wiatru zbliżała się do częstotliwości drgań rezonansowych konstrukcji. Podobnie w początku lat 60. XX w. szczegółowe badania przyczyn dwóch katastrof samolotowych wykazały, że przy przekraczaniu szybkości 650 km/h zawieszone na skrzydłach silników przechodziły przez krótką fazę drgań o częstotliwościach zbliżonych do częstotliwości

rezonansowej drgań skrzydła. Po pewnym czasie efekt rezonansu i zmęcenie materiału powodowały urywanie skrzydła.

Efekt wzbudzania drgań wokół częstotliwości rezonansowych bywa też niekiedy wykorzystywany z pozytywnym efektem, np. w konstrukcji generatora drgań sinusoidalnych, w którym zastosowano dodatnie sprzężenie zwrotne we wzmacniaczu, jakim jest trioda lub tranzystor, lub w układach strojenia radioodbiorników czy w pewnych konstrukcjach czujników pomiarowych (np. w czujniku gęstości cieczy, w którym mierzy się położenie maksimum rezonansowego U-rurki wypełnionej badaną ciecą, przy płynnej zmianie częstotliwości wymuszanych drgań).

4.6.3. Analiza dynamiki układu II rzędu na płaszczyźnie fazowej

W poprzednim podrozdziale podano wzory na odpowiedzi układu II rzędu: $h(t)$ i $g(t)$. Uzupełnieniem analizy tego układu będzie analiza na płaszczyźnie fazowej odpowiedzi układu jednorodnego, tzn. bez wymuszenia, ale z warunków początkowych różnych od zera:

$$\ddot{y}(t) + a_1 \dot{y}(t) + a_0 y(t) = 0, \quad y(0) \neq 0, \dot{y}(0) \neq 0.$$

Wprowadza się wektor stanu $x(t)$, dokonując podstawienia, $y(t) = x_1(t)$:

$$\dot{x}_1(t) = x_2(t),$$

$$\ddot{x}_1(t) = \dot{x}_2(t) = -a_0 x_1(t) - a_1 x_2(t),$$

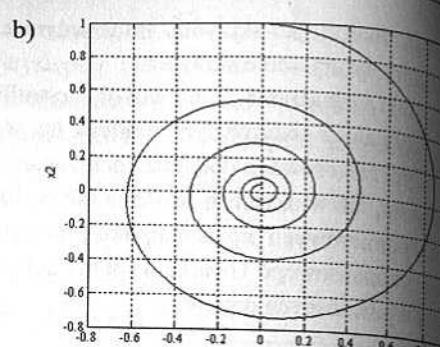
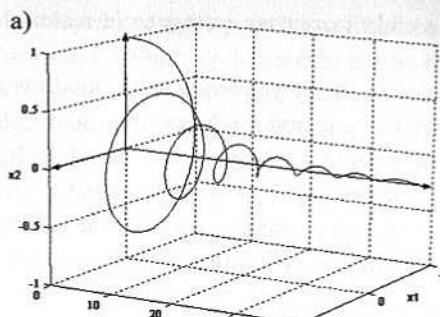
$$x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}, \quad x(0) = \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix}.$$

W zapisie macierzowym otrzymuje się zapis

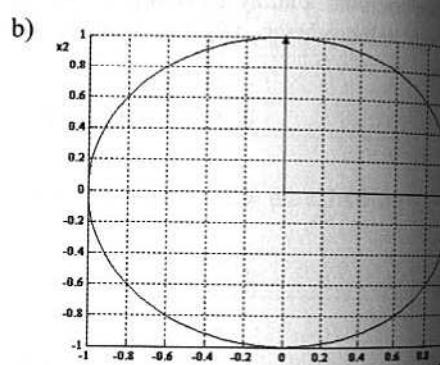
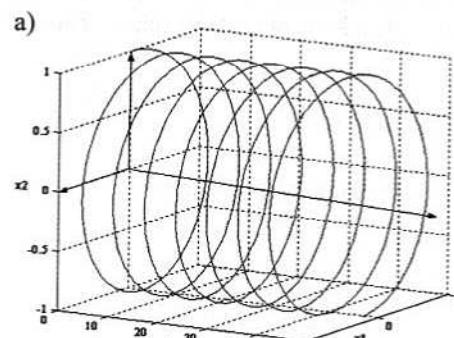
$$\dot{x} = A \cdot x = \begin{bmatrix} 0 & 1 \\ -a_0 & -a_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Dla analizy zachowania się trajektorii wokół początku układu (stabilności systemu) przydatne są wykresy przebiegu zmiennych stanu w czasoprzestrzeni stanu $[x_1(t), x_2(t), t]$.

W zależności od wartości własnych macierzy A (biegunów transmitancji) otrzymuje się różne rzuty trajektorii odpowiedzi czasowej na płaszczyznę fazową, czyli różne „portrety fazowe” (oś czasu prostopadła do powierzchni rys. 4.54). Czas jest parametrem na trajektoriach fazowych (rys. 4.55).



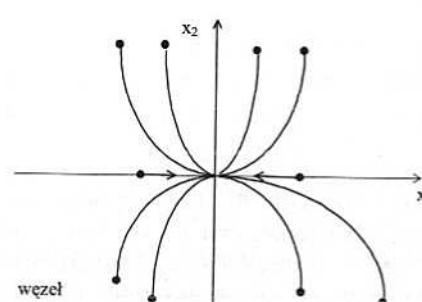
Rys. 4.54. Trajektorie fazowe rozwiązań tłumionego (a)
i ich rzut na płaszczyznę fazową (b)



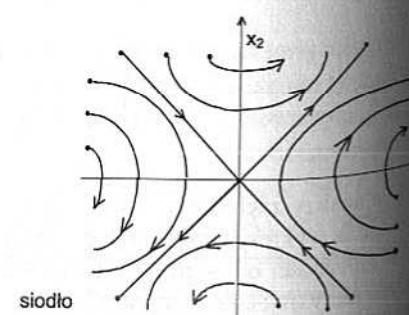
Rys. 4.55. Trajektorie fazowe rozwiązań nietłumionego (a)
i ich rzut na płaszczyznę fazową (b)

Cztery charakterystyczne portrety fazowe przedstawione na rysunkach 4.56–4.59.

- 1) Rzeczywiste wartości własne λ_1, λ_2 o jednakowych znakach – środek układu współrzędnych typu „węzeł” (rys. 4.56).



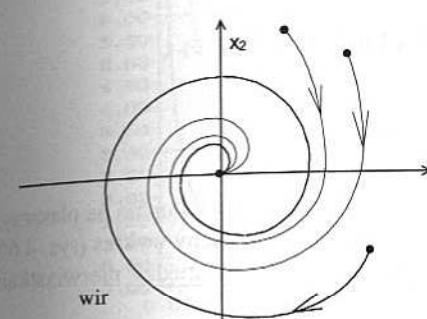
Rys. 4.56. Rodzina trajektorii typu „węzeł”



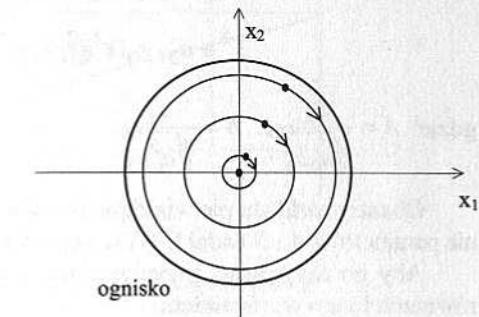
Rys. 4.57. Rodzina trajektorii typu „siodło”

4.7. Układ III rzędu

- 2) Rzeczywiste wartości własne λ_1, λ_2 o różnych znakach – środek układu współrzędnych typu „siodło” niestabilne (rys. 4.57)
- 3) Zespolone wartości własne λ_1, λ_2 – środek układu współrzędnych typu „wir” stabilny (rys. 4.58).
- 4) Urojone wartości własne λ_1, λ_2 – środek układu współrzędnych typu „ognisko” (rys. 4.59).



Rys. 4.58. Rodzina trajektorii typu „wir-spirala”



Rys. 4.59. Rodzina trajektorii typu „ognisko”

4.7. Układ III rzędu

Dane jest równanie różniczkowe opisujące dynamikę układu III rzędu

$$a_3 \ddot{y}(t) + a_2 \dot{y}(t) + a_1 y(t) + a_0 y(t) = b_0 u(t) \quad (4.22)$$

przy zerowych warunkach początkowych.

Dokonując przekształcenia Laplace'a

$$a_3 s^3 Y(s) + a_2 s^2 Y(s) + a_1 s Y(s) + a_0 Y(s) = b_0 U(s),$$

otrzymuje się transmitancję

$$\begin{aligned} G(s) &= \frac{Y(s)}{U(s)} = \frac{b_0}{a_3 s^3 + a_2 s^2 + a_1 s + a_0} = \\ &= \frac{b_0 / a_0}{a_3 / a_0 s^3 + a_2 / a_0 s^2 + a_1 / a_0 s + 1} \end{aligned} \quad (4.23)$$

Analiza własności modelu III rzędu przeprowadzona zostanie dla transmitancji znormalizowanej w mianowniku do dwóch parametrów. W tym celu zostanie dokonane przeskalowanie w dziedzinie zmiennej zespolonej operatora s do bezwymiarowego operatora \bar{s} , co w dziedzinie czasu będzie równoważne przeskalowaniu osi czasu t do bezwymiarowej osi \bar{t} (dla $a_0 < a_3$ oś czasu \bar{t} ulega skróceniu).

$$\frac{a_3}{a_0} s^3 = \bar{s}^3 \Rightarrow s^3 = \frac{a_0}{a_3} \bar{s}^3 = \Omega^3 \bar{s}^3 \Rightarrow s = \Omega \bar{s}$$
(4.24)

gdzie: $\Omega = \sqrt[3]{\frac{a_0}{a_3}}$, $t = \frac{1}{\Omega} \bar{t}$.

Mianownik transmitancji przyjmie postać dwuparametrową

$$\bar{s}^3 + a_2/a_0 \Omega^2 \bar{s}^2 + a_1/a_0 \Omega \bar{s} + 1 = \bar{s}^3 + A \bar{s}^2 + B \bar{s} + 1,$$

gdzie: $A = \frac{a_2}{\sqrt[3]{a_0 a_3^2}}$, $B = \frac{a_1}{\sqrt[3]{a_0^2 a_3}}$.

Obszary rozkładu pierwiastków powyższego wielomianu trzeciego stopnia na płaszczyźnie parametrów A i B badał I. Wyszniewogradski (1877 r.), uzyskując znany wykres (rys. 4.60).

Aby go narysować, przypomnijmy podstawowe zależności pomiędzy pierwiastkiem równania i jego wyróżnikiem.

Znormalizowany wielomian trzeciego stopnia

$$\bar{s}^3 + A \bar{s}^2 + B \bar{s} + 1 = 0$$
(4.25)

przez podstawienie $\bar{s} = \lambda - \frac{A}{3}$, sprowadza się do postaci

$$\lambda^3 + 3p\lambda^2 + 2p = 0$$
(4.26)

gdzie: $p = \frac{3B - A^2}{9}$, $\rho = \frac{2A^3 - 9AB + 27}{54}$

(4.27)

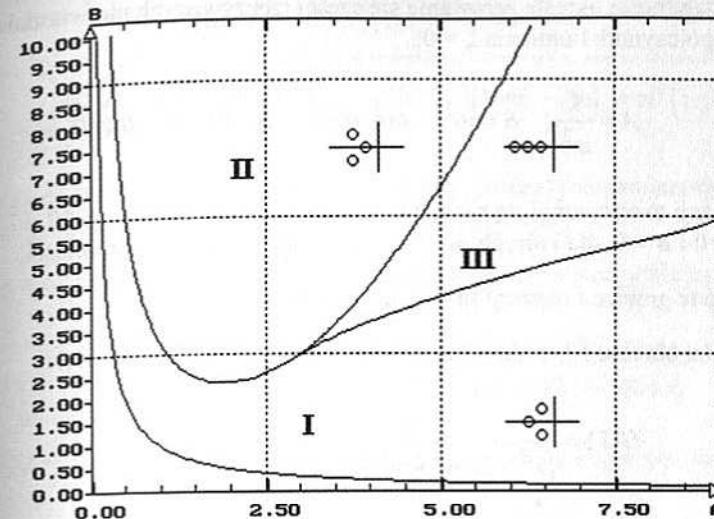
Różne typy pierwiastków tego wielomianu warunkuje znak jego wyróżnika $D = -p^3 - \rho^2$:

- 1) ujemność wyróżnika, $D < 0$, gwarantuje istnienie pary pierwiastków zespolonych sprzężonych;
- 2) zerowanie wyróżnika, $D = 0$, gwarantuje istnienie trzykrotnego pierwiastka zerowego $\lambda = 0$;
- 3) dodatniość wyróżnika, $D > 0$, gwarantuje istnienie trzech różnych pierwiastków rzeczywistych.

Trzy powyższe obszary są na płaszczyźnie Wyszniewogradskiego $[A, B]$ rozdzielone krzywymi, których równania parametryczne wynikają z zapisu (4.23) w postaci

$$\begin{aligned} (\bar{s}^2 + 2\xi\omega\bar{s} + \omega^2) \left(\bar{s} + \frac{1}{\omega^2} \right) &= \bar{s}^3 + \left(2\xi\omega + \frac{1}{\omega^2} \right) \bar{s}^2 + \left(\frac{2\xi}{\omega} + \omega^2 \right) \bar{s} + 1 = \\ &= \bar{s}^3 + A \bar{s}^2 + B \bar{s} + 1, \end{aligned}$$

gdzie trzy pierwiastki wyrażają się wzorami: $\bar{s}_{1,2} = -\xi\omega \pm j\omega\sqrt{1-\xi^2}$, $\bar{s}_3 = -1/\omega^2$.



Rys. 4.60. Wykres Wyszniewogradskiego
Obajśnienia w tekście

Pierwsza krzywa (rys. 4.60) rozdziela obszar pierwiastków zespolonych na dwa podobszary. W pierwszym podobszarze pierwiastek rzeczywisty jest mniejszy (bardziej na lewo) niż część rzeczywista pierwiastków zespolonych.

W drugim podobszarze pierwiastek rzeczywisty jest większy (bardziej na prawo) niż część rzeczywista pierwiastków zespolonych. Podobszar rozdziela granica obliczana z przyrównania ich części rzeczywistych $\operatorname{Re}[s_{1,2}] = \operatorname{Re}[s_3]$.

$$-\xi\omega = -\frac{1}{\omega^2} \Rightarrow \xi = \frac{1}{\omega^3} \Rightarrow A = \frac{3}{\omega^2}, \quad B = \frac{2}{\omega^4} + \omega^2, \quad \omega \in [1, \infty).$$

Stąd po eliminacji ω otrzymujemy się równanie brzegu obowiązujące dla $A < 3$

$$2A^3 - 9AB + 27 = 0.$$

Obszar trzeci – pierwiastków rzeczywistych – jest ograniczony krzywymi otrzymanymi z przyrównania współczynnika tłumienia do jedności, $\xi = 1$, i obowiązuje dla $A > 3$.

$$A = 2\omega + \frac{1}{\omega^2}, \quad B = \frac{2}{\omega} + \omega^2, \quad \omega \in [0, \infty).$$

Stąd po eliminacji ω otrzymujemy się równanie granicy

$$A^2 B^2 - 4(A^3 + B^3) + 18AB - 27 = D = 0.$$

Granicę stabilności określa zerowanie się części rzeczywistych pierwiastków zespolonych, czyli współczynnik tłumienia $\xi = 0$:

$$A = \frac{1}{\omega^2}, \quad B = \omega^2, \quad \omega \in [0, \infty), \quad A \in [0, \infty), \quad B \in [0, \infty).$$

Stąd po eliminacji ω otrzymuje się równanie granicy w postaci hiperboli $AB = 1$. Parametry $A > 0$ i $B > 0$, dla których $AB < 1$, wyznaczają obszar niestabilności.

Wszystkie te granice i obszary są widoczne na rysunku 4.60.

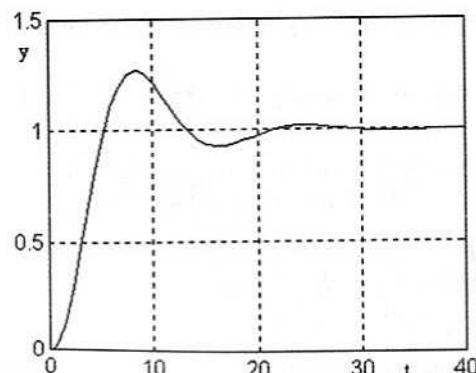
Odpowiedzi obiektu III rzędu o transmitancji z biegunami zespolonymi

$$G(\bar{s}) = \frac{K}{(\bar{s}^2 + 2\xi\omega\bar{s} + \omega^2)(\bar{s} + 1/\omega^2)}$$

na sterowania w postaci skoku jednostkowego $u(t) = 1(t)$ i $u(t) = \delta(t)$ mają postać:

$$h(t) = K \left[1 - \frac{\omega^2 e^{-t/\omega^2}}{\left(\frac{1}{\omega^2} - \xi\omega\right)^2 + \omega^4(1-\xi^2)} - \frac{e^{-\xi\omega t} \cdot \sin\left(\omega\sqrt{1-\xi^2}t + \Psi - \Phi\right)}{\sqrt{(1-\xi^2)\left(\frac{1}{\omega^2} - \xi\omega\right)^2 + \omega^2(1-\xi^2)^2}} \right] \quad (4.28)$$

$$\operatorname{tg}\Psi = \frac{\sqrt{1-\xi^2}}{\xi}, \quad \operatorname{tg}\Phi = \frac{\omega\sqrt{1-\xi^2}}{1/\omega^2 - \xi\omega}, \quad \lim_{t \rightarrow \infty} h(t) = h_\infty = K.$$



Rys. 4.61. Odpowiedź obiektu oscylacyjnego $A = 6, B = 2$

$$\dot{h}(t) = g(t) = K \left[\frac{e^{-t/\omega^2}}{\left(\frac{1}{\omega^2} - \xi\omega\right)^2 + \omega^4(1-\xi^2)} + \frac{e^{-\xi\omega t} \cdot \sin\left(\omega\sqrt{1-\xi^2}t - \Phi\right)}{\sqrt{(1-\xi^2)\left(\frac{1}{\omega^2} - \xi\omega\right)^2 + \omega^2(1-\xi^2)^2}} \right] \quad (4.29)$$

Odpowiedź obiektu na skok jednostkowy dla przykładowych parametrów $K = 1, A = 6, B = 2$ wybranych wewnątrz obszaru I ma postać jak na rysunku 4.61.

Odpowiedź obiektu III rzędu o transmitancji z biegunami rzeczywistymi

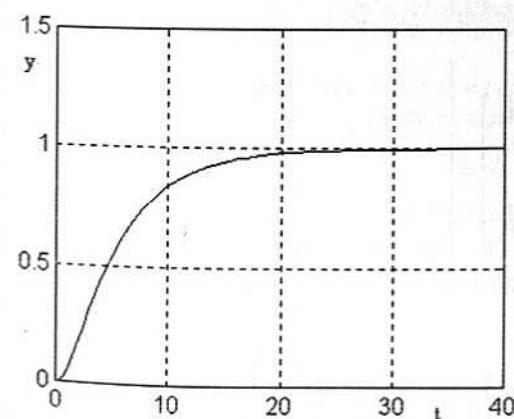
$$G(\bar{s}) = \frac{K}{(\bar{s} + \gamma)(\bar{s} + \delta)(\bar{s} + \omega)} = \frac{K}{(\bar{s}^3 + A\bar{s}^2 + B\bar{s} + 1)},$$

gdzie: $A = \gamma + \delta + \omega, B = \gamma\delta + \gamma\omega + \delta\omega, 1 = \gamma\delta\omega$, na sterowania w postaci skoku jednostkowego $u(t) = 1(t)$ i $u(t) = \delta(t)$ ma postać:

$$h(t) = K \left[1 - \frac{e^{-\gamma t}}{\gamma(\gamma - \delta)(\gamma - \omega)} - \frac{e^{-\delta t}}{\delta(\delta - \gamma)(\delta - \omega)} - \frac{e^{-\omega t}}{\omega(\omega - \gamma)(\omega - \delta)} \right] \quad (4.30)$$

$$\dot{h}(t) = g(t) = K \left[\frac{e^{-\gamma t}}{(\gamma - \delta)(\gamma - \omega)} + \frac{e^{-\delta t}}{(\delta - \gamma)(\delta - \omega)} + \frac{e^{-\omega t}}{(\omega - \gamma)(\omega - \delta)} \right] \quad (4.31)$$

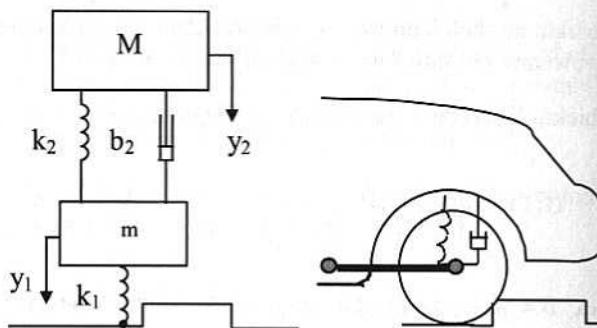
Dla przykładowych parametrów $K = 1, A = 6, B = 6$ przebieg odpowiedzi na skok jednostkowy jest dany na rysunku 4.62. W tym przypadku bieguny są liczbami rzeczywistymi.



Rys. 4.62. Odpowiedź układu przethumionego

4.8. Przykład analizy obiektu IV rzędu

Jako przykład obiektu opisywanego równaniem różniczkowym IV rzędu rozważmy uproszczony fragment układu zawieszenia w samochodzie (ćwiartkę) (rys. 4.63).



Rys. 4.63. Model układu zawieszenia

Sterowanie u w postaci skoku przesunięcia dolnego końca sprężyny k_1 (opony) przynosi się jako siła $k_1 u$ na masę m i na karoserię M . Układ zawieszenia powinien łagodzić efekt pojedynczego skoku (krawężnik), jak również zmniejszać amplitudę ciągłych drgań sinusoidalnych karoserii w stosunku do amplitudy wybojów na drodze.

Punkt równowagi reprezentuje wstępne napięcia sprężyn pod wpływem sił ciężkości. Dynamikę modelu opisują odchyłkowe zmienne stanu:

$$y_1 = x_1, \quad \dot{x}_1 = x_2, \quad y_2 = x_3, \quad \dot{x}_3 = x_4,$$

$$M\ddot{x}_4 + b_2(x_4 - x_2) + k_2(x_3 - x_1) = 0,$$

$$m\ddot{x}_2 + b_2(x_2 - x_4) + k_2(x_1 - x_3) + k_1x_1 = k_1u,$$

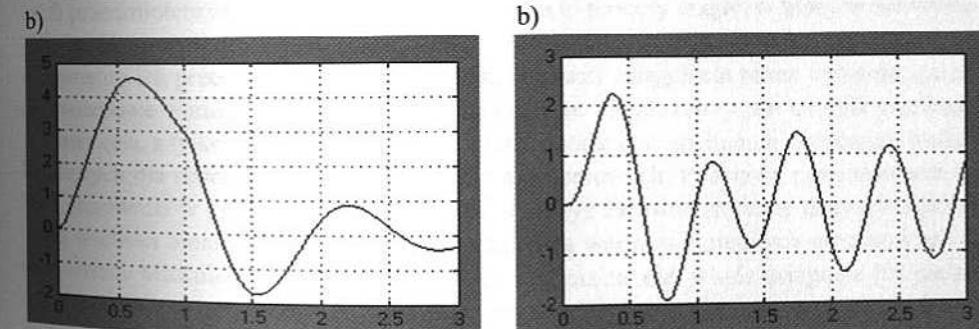
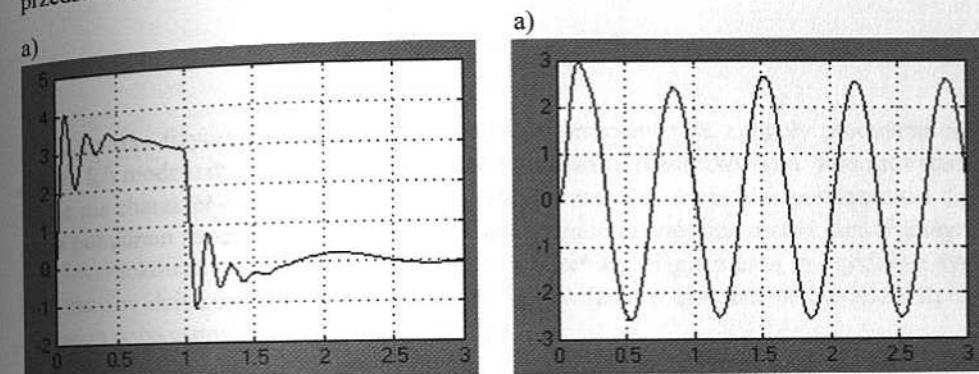
$$\begin{bmatrix} \dot{x}_1 \\ \cdot \\ \cdot \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -\frac{k_1+k_2}{m} & -\frac{b_2}{m} & \frac{k_2}{m} & \frac{b_2}{m} \\ 0 & 0 & 0 & 1 \\ \frac{k_2}{M} & \frac{b_2}{M} & -\frac{k_2}{M} & -\frac{b_2}{M} \end{bmatrix} \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ x_4 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{k_1}{m}u \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ x_4 \end{bmatrix}.$$

4.8. Przykład analizy obiektu IV rzędu

Przymijmy średnie (fikcyjne) masy: $M = 350 \text{ kg}$ (1/4 karoserii), $m = 35 \text{ kg}$ (koło). Dobrane eksperymentalnie współczynniki przyjmują wartości:
 $k_1 = 40\,000 \text{ kg/s}^2$ (współczynnik sprężystości opony),
 $k_2 = 10\,000 \text{ kg/s}^2$ (współczynnik sprężystości resor),
 $b_2 = 500 \text{ kg/s}$ (współczynnik tłumienia).

Wyniki symulacji przedstawiono dla dwóch typów sterowań: funkcji skoku i sinus przedstawiają rysunki 4.64 i 4.65.



Rys. 4.64. Przesunięcie środka ciężkości koła (a) i karoserii (b) na wymuszenie skokowe opony
 $u_1(t) = 3 \cdot 1(t) - 3 \cdot 1(t-1)$

Rys. 4.65. Przesunięcie środka ciężkości koła (a) i karoserii (b) na wymuszenie sinusoidalne opony $u_2(t) = 3 \cdot 1(t) \cdot \sin(3\pi \cdot t)$

Dobrane eksperymentalnie współczynniki pokazują, że możliwe jest zapewnienie wymagań funkcjonalnych zawieszenia, tzn. układ zawieszenia łagodzi efekt pojedynczego skoku (krawężnik), jak również zmniejsza amplitudę ciągłych drgań sinusoidalnych karoserii w stosunku do amplitudy wybojów na drodze.

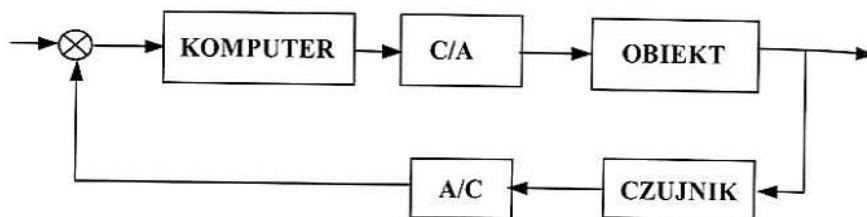
5. Analiza układów liniowych dyskretnych

Zjawiska fizyczne występujące w technologii procesowej są z reguły procesami ciągłymi. Ich modele dynamiczne są opisywane równaniami różniczkowymi. Również sterowanie ma charakter ciągle zmieniającej się funkcji czasu i to z reguły w powiązaniu z ciągłym pomiarem zmiennych procesowych. Jednak z punktu widzenia metod numerycznych oraz uwarunkowań techniki cyfrowej, wszystkie przebiegi ciągłe muszą być poddane dyskretyzacji w dziedzinie czasu. Mamy wtedy do czynienia z dyskretnym procesem dynamicznym opisywanym równaniem różnicowym.

Należy tu wspomnieć o innej odmianie dyskretnych procesów produkcyjnych, będących przedmiotem zainteresowania automatyki. Są to procesy ciągłe, w których nie występuje problem syntezy kształtu ciągłej funkcji sterowania. Sterowania są stałwartościowe, a momenty ich przełączeń są wybierane bądź w chwili osiągnięcia przez wybrane zmienne procesowe wartości granicznej (stąd przykładowo niepotrzebny jest czujnik wielkości przesunięcia, a tylko ogranicznik brzegowy) lub na podstawie spełnienia ustalonych funkcji logicznych dla stałwartościowych zmiennych procesowych. Podstawą projektowania takich sterowników logicznych jest algebra Boola, gdyż zarówno zmienne decyzyjne przyjmują wartości binarne, jak i interesujące technologa wartości zmiennych procesowych są pojmowane binarnie (zbiornik jest napełniony lub jeszcze nie, winda osiągnęła już piętro, element na taśmie produkcyjnej dotarł do automatu mocującego śrubę) – stąd ich angielska nazwa *Discrete Event Systems*. Do takich procesów należą więc wspomniane dyskretne procesy produkcyjne z gniazdami produkcji i taśmowym rytmem produkcji. Problem dynamiki sprowadza się tu do czasu opóźnienia transportowego i czasu przebywania obiektu w gnieździe. Takie dyskretne procesy produkcyjne nie będą dokładniej omawiane w tej monografii.

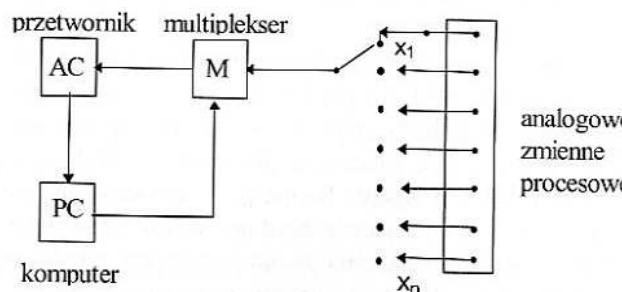
Rozwój cyfrowych układów sterowania procesami ciągłymi dokonywał się na podstawie matematycznych podstaw opisu ich dynamiki, w sytuacji gdy informacja o ich zachowaniu napływa nie w sposób ciągły, ale w dyskretnych chwilach czasu. Układy cyfrowe stały się dziś standardowym wyposażeniem układów automatyki. Cechą tych układów jest zmienne oddziaływanie na ciągły obiekt sterowania w dyskretnych chwilach czasu za pomocą dyskretnego pomiaru ich wyjścia. Taki dyskretny charakter niektórych pomiarów występował w wielu klasycznych rozwiązaniach sterowania układami ciągłymi. Do typowych należały chromatograficzne pomiary składu chemicznego, w których próbka pomiarowa przepływała przez kolumnę chromatograficzną. Wynik generowany był raz na

kilka minut. Dyskretność związana więc była z czasem całkowania urządzenia pomiarowego. Również dzisiejsze woltomierze cyfrowe (przetworniki A/C i C/A) potrzebują niewielkiego czasu integracji dla odczytu wielkości analogowej i jej zamiany na postać cyfrową (rys. 5.1).



Rys. 5.1. Schemat układu regulacji cyfrowej

Jednak dyskretyzacja sterowania i pomiaru wyjścia wiąże się nie tylko z technologią budowy urządzeń sterujących, ale też z możliwością obsługi wielowymiarowego obiektu sterowania przez jeden komputer. W cyklu o wybranej długości T komputer poświęca kolejne ułamki czasu tego cyklu na obsługę kolejnych torów sterowania. Każdy tor jest obsługiwany co okres T (rys. 5.2). Okres ten zwany jest okresem próbkowania toru.



Rys. 5.2. Multiplekser w układzie regulacji cyfrowej

Jak wspomniano, innym źródłem dyskretyzacji równań różniczkowych są numeryczne metody ich rozwiązywania. Dla potrzeb wszystkich wymienionych sytuacji dobierany okres dyskretyzacji będzie inny i stąd inne modele dyskretne.

5.1. Podstawy równań różnicowych

Przechowywane w pamięci komputera wartości dyskretne zmiennej tworzą ciąg pomiarowy, którego cechą charakterystyczną jest fakt, że nie istnieje informacja na temat wartości tej zmiennej między chwilami impulsowania. Należało więc stworzyć aparat matematyczny, który korzystając z niepełnej informacji precyzyjnie, opisywałby dynamikę obiek-

5.1. Podstawy równań różnicowych

tów, posługując się szeregiem liczbowymi, a nie ciągłymi funkcjami czasu. Takim aparatem w dziedzinie czasu dyskretnego są równania różnicowe. Formalnie mogą one być wyrowadzone poprzez np. lewostronną (*backward*) aproksymację pochodnych w równaniu różniczkowym.

$$\frac{dy(t)}{dt} = \lim_{\Delta t \rightarrow 0} \frac{y(t) - y(t - \Delta t)}{\Delta t} = \frac{y(kT) - y(kT - T)}{T},$$

$$\frac{d^2 y(t)}{dt^2} = \lim_{\Delta t \rightarrow 0} \frac{\frac{dy(t)}{dt} - \frac{dy(t - \Delta t)}{dt}}{\Delta t} = \frac{y(kT) - 2y(kT - T) + y(kT - 2T)}{T^2},$$

$$\frac{d^3 y(t)}{dt^3} = \lim_{\Delta t \rightarrow 0} \frac{\frac{d^2 y(t)}{dt^2} - \frac{d^2 y(t - \Delta t)}{dt^2}}{\Delta t} = \frac{y(kT) - 3y(kT - T) + 3y(kT - 2T) + y(kT - 3T)}{T^3}.$$

W praktyce pomija się w oznaczaniu czasu dyskretnego $t = kT$ symbol okresu T , przypisując do próbki pomiarowej tylko jej wskaźnik:

$$y(kT) \equiv y(k) \equiv y_k, \quad y(kT - T) \equiv y(k-1) \equiv y_{k-1}.$$

Przykładowe zastosowanie aproksymacji lewostronnej do równania różniczkowego II rzędu

$$a_{c2}\ddot{y}(t) + a_{c1}\dot{y}(t) + y(t) = b_0u(t) \quad (5.1)$$

da wynik

$$a_{d0}y(k) + a_{d1}y(k-1) + a_{d2}y(k-2) = b_0u(k) \quad (5.2)$$

gdzie współczynniki dane są wzorami:

$$a_{d0} = \frac{a_{c2}}{T^2} + \frac{a_{c1}}{T} + 1, \quad a_{d1} = -\frac{a_{c1}}{T} - \frac{2a_{c2}}{T^2}, \quad a_{d2} = \frac{a_{c2}}{T^2}.$$

Ważne jest spostrzeżenie, że współczynniki równania różnicowego są funkcjami nie tylko współczynników równania różniczkowego, ale również czasu dyskretyzacji (okresu próbkowania) T .

Ponieważ można przyjmować różne inne aproksymacje pochodnej, np. aproksymacje prawostronną (*forward*)

$$\frac{dy(t)}{dt} = \lim_{\Delta t \rightarrow 0} \frac{y(t + \Delta t) - y(t)}{\Delta t} \approx \frac{y(kT + T) - y(kT)}{T}$$

lub aproksymację symetryczną

$$\frac{dy(t)}{dt} = \lim_{\Delta t \rightarrow 0} \frac{y(t + \Delta t) - y(t - \Delta t)}{2\Delta t} \approx \frac{y(kT + T) - y(kT - T)}{2T},$$

to w praktyce dyskretyzacja równania różniczkowego może prowadzić do różnych wartości współczynników równania różnicowego.

Na przykład zastosowanie aproksymacji prawostronnych do równania (5.1) da wynik

$$a_{2d}y(k+2) + a_{1d}y(k+1) + a_{0d}y(k) = b_0u(k) \quad (5.3)$$

gdzie współczynniki dane są wzorami:

$$a_{2d} = \frac{a_{c2}}{T^2}, \quad a_{1d} = \frac{a_{c1}}{T} - \frac{2a_{c2}}{T^2}, \quad a_{0d} = \frac{a_{c2}}{T^2} - \frac{a_{c1}}{T} + 1.$$

Stąd równania (5.2) i (5.3) można traktować jako różne aproksymacje równania różniczkowego (5.1).

Można badać dokładność zastosowanych aproksymacji pochodnych. I tak aproksymacja prawostronna daje dokładność I rzędu i jest mniej dokładna niż aproksymacja symetryczna, która jest II rzędu. Wynika to z rozwinięcia w szereg Taylora:

$$y(t + \Delta t) = y(t) + \dot{y}(t)\Delta t + \ddot{y}(t)\frac{(\Delta t)^2}{2} + \dddot{y}(t)\frac{(\Delta t)^3}{6} + O((\Delta t)^4),$$

$$y(t - \Delta t) = y(t) - \dot{y}(t)\Delta t + \ddot{y}(t)\frac{(\Delta t)^2}{2} - \dddot{y}(t)\frac{(\Delta t)^3}{6} + O((\Delta t)^4).$$

Stąd aproksymacja prawostronna daje wzór

$$\frac{y(t + \Delta t) - y(t)}{\Delta t} = \dot{y}(t) + \ddot{y}(t)\frac{\Delta t}{2} + O((\Delta t)^2),$$

a aproksymacja symetryczna wzór

$$\frac{y(t + \Delta t) - y(t - \Delta t)}{2\Delta t} = \dot{y}(t) + \ddot{y}(t)\frac{(\Delta t)^2}{6} + O((\Delta t)^4).$$

Innym ważnym spostrzeżeniem jest fakt niejednoznaczności operacji odwrotnej, tzn. przejścia z danego równania różnicowego na równanie różniczkowe. W modelu dyskretnym brak jest bowiem informacji o typie zastosowanej aproksymacji. Bywa też, że przejście takie jest w ogóle niemożliwe z powodu braku rozwiązania na parametry a_{ci} .

Na przykład dla prostego równania różnicowego

$$y(k) - y(k-1) - y(k-2) = 0,$$

z warunkiem początkowym $y(-2) = 0, y(-1) = 1$, generującego dla kolejnych k narastający ciąg Fibonaciego liczb: $0, 1, 1, 2, 3, 5, 8, \dots$ nie istnieją rozwiązania na parametry odpowied-

5.1. Podstawy równań różnicowych

niego równania różniczkowego (5.1), ani w wersji aproksymacji lewostronnej, ani w prawostronnej.

Nazwa równanie różnicowe wywodzi się z faktu, że każde równanie dyskretnie może być przedstawione za pomocą samych różnic próbek pomiarowych.

Oznaczmy pierwszą, drugą i n -tą różnicę lewostronną w chwili k jako:

$$\nabla y_k = y_k - y_{k-1}, \quad \nabla^2 y_k = \nabla y_k - \nabla y_{k-1}, \quad \nabla^n y_k = \nabla^{n-1} y_k - \nabla^{n-1} y_{k-1},$$

$$\nabla^2 y_k = \nabla y_k - \nabla y_{k-1} = y_k - 2y_{k-1} + y_{k-2}.$$

Można łatwo wykazać, że równanie (5.2) wyraża się jako:

$$a_{d2}\nabla^2 y(k) - (a_{d1} + 2a_{d2})\nabla y(k) + (a_{d2} + a_{d1} + a_{d0})y(k) = b_0u(k) \quad (5.4)$$

lub po podstawieniu a_{di} z (5.2)

$$\frac{a_{c2}}{T^2} \nabla^2 y(k) + \frac{a_{c1}}{T} \nabla y(k) + y(k) = b_0u(k).$$

Widać wyraźnie analogię do wyjściowego równania różniczkowego (5.1), w którym odpowiednie pochodne zastąpiono ilorazami różnicowymi. Forma rekurencyjna (5.2) bywa jednak łatwiejsza do obliczeń.

Analogicznie do różnic lewostronnych ∇ zwanych w języku angielskim *backward difference operator* można zdefiniować różnice prawostronne Δ : zwane *forward difference operator*:

$$\Delta y_k = y_{k+1} - y_k, \quad \Delta^2 y_k = \Delta y_{k+1} - \Delta y_k, \quad \Delta^n y_k = \Delta^{n-1} y_{k+1} - \Delta^{n-1} y_k,$$

$$\Delta^2 y_k = y_{k+2} - 2y_{k+1} + y_k.$$

Postacie równań różnicowych mogą być również wyprowadzane z numerycznych metod całkowania. Przykładowo rozwiązywanie równania różniczkowego z warunkiem początkowym

$$\dot{y}(t) = u(t), \quad y(0) = y_0 \quad (5.5)$$

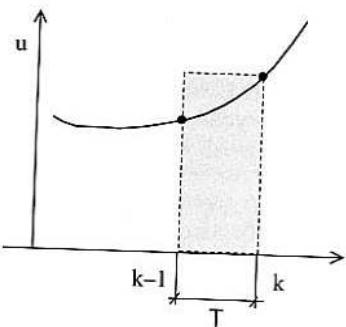
ma postać

$$y(t) = y_0 + \int_0^t u(\tau) d\tau.$$

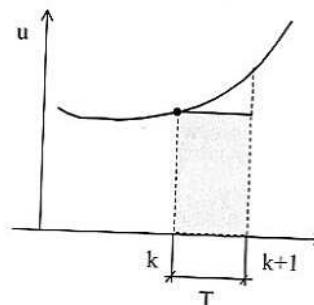
Standardowe reguły całkowania:

- Reguła całkowania prostokątów wstecz (*backward*, rys. 5.3) ma uzasadnienie w prawostronnej aproksymacji pochodnej, która zastosowana do (5.5) (przy założeniu, że y_{k-1} jest wartością całki w chwili $k-1$) daje:

$$y_k = y_{k-1} + T u_k.$$



Rys. 5.3. Reguła całkowania wstecz



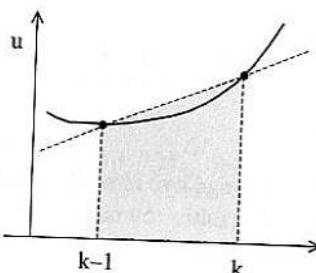
Rys. 5.4. Reguła całkowania w przód

- Reguła całkowania prostokątów w przód (*forward*, rys. 5.4) ma uzasadnienie w prawostronnej aproksymacji pochodnej, która zastosowana do (5.5) daje:

$$y_{k+1} = y_k + Tu_k \quad \text{lub} \quad y_k = y_{k-1} + Tu_{k-1}.$$

- Wykorzystanie reguły całkowania trapezów (Tustina, rys. 5.5) daje jeszcze inną aproksymację równania różniczkowego (5.5) (metoda Heuna)

$$y_k = y_{k-1} + \frac{u_k + u_{k-1}}{2} \cdot T \Rightarrow y_k = y_{k-1} + 0,5 \cdot Tu_k + 0,5 \cdot Tu_{k-1}.$$



Rys. 5.5. Reguła całkowania trapezów

Na zakończenie zwróćmy uwagę na zasadniczą różnicę pomiędzy równaniem różniczkowym i różnicowym (np. (5.1) i (5.2)).

Równość stron w równaniu różniczkowym oznacza jego spełnianie dla każdej chwili t . Dla weryfikacji tej równości oprócz stałych parametrów równania (stacjonarność) konieczna jest znajomość wartości wszystkich pochodnych wyjścia i wejścia do rzędu n w każdej wybranej chwili czasu. Zajomość wartości funkcji wejścia, wyjścia i ich pochodnych w tej samej chwili czasu niesie więc całą informację o modelu ciągłym. Im obiekt „trudniejszy”, tym w modelu trzeba uwzględnić informację o większej liczbie pochodnych w danej chwili, czyli stosować wyższy rzad równania różniczkowego. Weryfikacja modelu ciągłego

na podstawie danych pomiarowych może być utrudniona, w związku z koniecznością posiadania wielu przyrządów pomiarowych, każdy mierzący wartość kolejnej pochodnej wyjścia i wejścia w danej chwili czasu.

Do budowy modelu dyskretnego używa się dwóch ruchomych okien pomiarowych (rejestrów przesuwnych) o większej lub mniejszej długości, wypełnionych kolejnymi dyskretnymi próbками pomiarowymi samych sygnałów wejścia i wyjścia, a nie ich pochodnych. Im obiekt „trudniejszy”, tym należy zastosować dłuższe okno pomiarowe (wyższy rzad równania dyskretnego, ściśle związane z rzędem równania ciągłego). Zajomość współczynników równania i zbiór kolejnych próbek pomiarowych w różnych chwilach czasu (historia) niesie więc całą informację o modelu. Model może być łatwo weryfikowany poprzez pomiar tylko jednym typem przyrządu.

5.2. Charakterystyczne postacie równań różnicowych

Liniowe równanie różnicowe opisujące układ n -tego rzędu typu SISO z wejściami deterministycznymi $u(k)$ i stochastycznymi $w(k)$ ma ogólną postać

$$\sum_{i=0}^n a_i y(k-i) = \sum_{i=0}^m b_i u(k-i) + \sum_{i=0}^r c_i w(k-i) \quad (5.6)$$

Wprowadza się operator przesunięcia pomiaru o n próbek:

$$q^{-n} [y(k)] \equiv y(k-n) \quad \text{operator opóźnienia},$$

$$q^n [y(k)] \equiv y(k+n) \quad \text{operator wyprzedzenia}.$$

Równanie (5.6) może być formalnie zapisane z użyciem tych operatorów jako

$$\mathcal{A}(q^{-1})[y(k)] = \mathcal{B}(q^{-1})[u(k)] + \mathcal{C}(q^{-1})[w(k)] \quad (5.7)$$

gdzie przez operatory \mathcal{A} , \mathcal{B} , \mathcal{C} , rozumie się wielomiany operatora q^{-1} :

$$\mathcal{A}(q^{-1}) = a_0 + a_1 q^{-1} + \dots + a_n q^{-n},$$

$$\mathcal{B}(q^{-1}) = b_0 + b_1 q^{-1} + \dots + b_m q^{-m},$$

$$\mathcal{C}(q^{-1}) = c_0 + c_1 q^{-1} + \dots + c_r q^{-r}.$$

Część związaną z operatorem \mathcal{A} i wyjściem nazywa się częścią autoregresyjną AR modelu, część związaną z operatorem \mathcal{B} i wejściem deterministycznym nazywa się częścią X modelu, część związaną z operatorem \mathcal{C} i wejściem stochastycznym nazywa się średnią ruchomą MA. Cały model (5.7) oznacza się ARMAX(n, m, r) ([5NIA1]).

Wyszczególnia się jako charakterystyczne następujące odmiany modelu ARMAX (tab. 5.1).

Tabela 5.1

Odmiany modelu ARMAX

ARMAX ($n, m, 0$)	$\mathcal{A}(q^{-1})[y(k)] = \mathcal{B}(q^{-1})[u(k)]$ <i>(Infinite Impulse Response)</i>	model IIR
ARMAX ($1, m, 0$)	$y(k) = \mathcal{B}(q^{-1})[u(k)]$ <i>(Finite Impulse Response)</i>	model FIR
ARMAX ($n, 0, 1$)	$\mathcal{A}(q^{-1})[y(k)] = w(k)$ <i>(AutoRegressive)</i>	model AR
ARMAX ($1, 0, r$)	$y(k) = \mathcal{C}(q^{-1})[w(k)]$ <i>(Moving Average)</i>	model MA
ARMAX ($n, 0, r$)	$\mathcal{A}(q^{-1})[y(k)] = \mathcal{C}(q^{-1})[w(k)]$ <i>(AutoRegr-MovingAverage)</i>	model ARMA
ARMAX ($n, m, 1$)	$\mathcal{A}(q^{-1})[y(k)] = \mathcal{B}(q^{-1})[u(k)] + w(k)$ <i>(with eXogenous input)</i>	model ARX
ARMAX (n, m, r, τ)	$\mathcal{A}(q^{-1})[y(k)] = \mathcal{B}(q^{-1})[u(k-\tau)] + \mathcal{C}(q^{-1})[w(k)]$ <i>(with delay in control)</i>	model ARMAX
(Box-Jenkins)	$y(k) = \frac{\mathcal{B}(q^{-1})}{\mathcal{F}(q^{-1})}[u(k-\tau)] + \frac{\mathcal{C}(q^{-1})}{\mathcal{D}(q^{-1})}[w(k)]$	model BJ
(Output Error)	$y(k) = \mathcal{B}(q^{-1})/\mathcal{F}(q^{-1})[u(k-\tau)] + w(k)$	model OE

Model mieszany ARMAX – deterministyczno-stochastyczny można sprowadzić do modelu stochastycznego ARMA, usuwając wartości średnie (trendy) $m_y(k)$, $m_w(k)$, $u(k)$.
Dla modelu

$$\mathcal{A}(q^{-1})[y(k)] = \mathcal{B}(q^{-1})[u(k)] + \mathcal{C}(q^{-1})[w(k)],$$

można zapisać odpowiedni model dla wartości oczekiwanych jako

$$\mathcal{A}(q^{-1})[m_y(k)] = \mathcal{B}(q^{-1})[u(k)] + \mathcal{C}(q^{-1})[m_w(k)].$$

Odejmując je stronami, otrzyma się model ARMA

$$\mathcal{A}(q^{-1})[y(k) - m_y(k)] = \mathcal{C}(q^{-1})[w(k) - m_w(k)].$$

Przez wartość średnią arytmetyczną dla szeregu czasowego o N próbkach rozumie się

$$m_y = \frac{1}{N} \sum_{i=1}^N y_i = \hat{y}.$$

Wartość średniokwadratowa

$$\hat{y}^2 = \frac{1}{N} \sum_{i=1}^N y_i^2.$$

Wariancja – miara zmienności sygnału wokół wartości średniej (średnia arytmetyczna z kwadratów)

$$\text{var } y = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2.$$

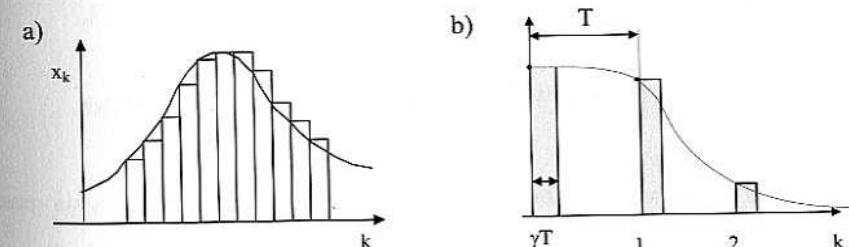
Odchylenie standardowe

$$\sigma_y = \sqrt{\text{var } y} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2}.$$

5.3. Analiza układów dyskretnych w dziedzinie czasu

5.3.1. Równanie stanu

Sygnal ciągły można aproksymować ciągiem sygnałów prostokątnych o różnych amplitudach x_k i różnym wypełnieniu $0 < \gamma \leq 1$ okresu dyskretyzacji T (rys. 5.6).

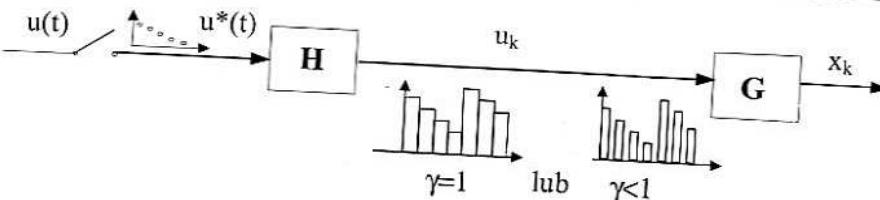


Rys. 5.6. Sygnał ciągły (a) i jego aproksymacja (b)

Fakt ten został wykorzystany w praktycznych rozwiązaniach cyfrowych układów sterowania, w których występuje człon zawierający idealny impulsator i człon podtrzymujący wartość sygnału zmierzzonego, przez cały okres dyskretyzacji lub przez jego część. Sygnał powstający w procesie modulowania za impulsatorem wyraża się formalnie ciągiem deltów Diraca, a za członem podtrzymującym (rys. 5.7) przyjmuje wartości stałe

$$u^*(t) = \sum_{k=0}^{\infty} u(kT) \cdot \delta(t - kT),$$

gdzie $u_k = u(kT)$.



Rys. 5.7. Czlon podtrzymujacy sygnał

Czlon formujacy sygnał $u(t)$ na zasadzie jego podtrzymania pomiędzy chwilą kT i $kT + \tau$ na poziomie stałym zwany jest czlonem podtrzymujacym zerowego rzędu ZOH (Zero Order Hold)

$$u(t) = \begin{cases} u(kT): & kT \leq t \leq kT + \tau, \\ 0: & (kT + \tau) < t \leq (k+1)T. \end{cases}$$

Wzór na wyliczanie dyskretnych wartości odpowiedzi $x(kT)$, można otrzymać z analizy rozwiązań równania układu ciągłego

$$x(t) = e^{A(t-t_0)}x(t_0) + \int_{t_0}^t e^{A(t-s)}Bu(s)ds \quad (5.8)$$

Zakładając, że chwila początkowa t_0 będzie k -tym momentem próbkowania $t_0 = kT$, można obliczyć dokładną wartość $x((k+1)T)$

$$x[(k+1)T] = e^{A((k+1)T-kT)}x(kT) + \int_{kT}^{kT+\tau} e^{A((k+1)T-s)}Bu(s)ds.$$

Po dokonaniu podstawienia: $[(k+1)T - s] = \xi \Rightarrow dt = -d\xi$; zmianie ulegają granice całki:

$$s = kT \rightarrow \xi = T, \quad s = kT + \tau \rightarrow \xi = T - \tau.$$

W przedziale całkowania sterowanie jest stałe, stąd można go wyciągnąć przed całkę

$$x[(k+1)T] = e^{AT}x(kT) + \int_{T-\tau}^T e^{A\xi}Bd\xi \cdot u(kT) \quad (5.9)$$

Jeżeli $\tau = T$, to współczynnik wypełnienia $\gamma = \tau/T = 1$ (całkowe wypełnienie okresu impulsowania sterowaniem) i całkowanie obejmuje przedział $[0, T]$. Równanie macierzowe n -tego rzędu:

$$x_{k+1} = A_D x_k + B_D u_k$$

$$\begin{bmatrix} x_{k+1}^1 \\ x_{k+1}^2 \\ \vdots \\ x_{k+1}^n \end{bmatrix} = \begin{bmatrix} A_D & & & \\ & A_D & & \\ & & A_D & \\ & & & A_D \end{bmatrix} \begin{bmatrix} x_k^1 \\ x_k^2 \\ \vdots \\ x_k^n \end{bmatrix} + \begin{bmatrix} B_D & & & \\ & B_D & & \\ & & B_D & \\ & & & B_D \end{bmatrix} \begin{bmatrix} u_k^1 \\ u_k^2 \\ \vdots \\ u_k^n \end{bmatrix} \quad (5.10)$$

gdzie macierze stałe A_D, B_D (dla $\tau = T$) mają postać:

$$A_D = e^{AT}, \quad B_D = \int_0^T e^{A\tau} B d\tau \quad (5.11)$$

Dla danego systemu liniowego ciągłego z macierzami A, B, C można policzyć odpowiadający mu system dyskretny z macierzami $A_D, B_D, C_D = C$ i czasem dyskretyzacji T .

Do policzenia macierzy wykładniczej można użyć trzech formuł:

$$1) \quad e^{AT} = \mathcal{L}^{-1} \left[(sI - A)^{-1} \right]_{t=T}.$$

$$2) \quad e^{AT} = P e^{JT} P^{-1},$$

gdzie:

P – macierz wektorów własnych związkanych z wartościami własnymi λ_i macierzy A ,

J – diagonalna macierz Jordana:

$$J = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}, \quad e^{JT} = \begin{bmatrix} e^{\lambda_1 T} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_n T} \end{bmatrix}.$$

$$3) \quad e^{AT} = I + AT + \frac{A^2 T^2}{2} + \frac{A^3 T^3}{3!} + \dots = I + AT \cdot \left[I + \frac{AT}{2!} + \frac{A^2 T^2}{3!} + \dots \right] = I + AT \cdot \Psi.$$

Dla obliczania macierzy Ψ można wykorzystać formuły rekurencyjne

$$\Psi = I + \frac{AT}{2} \left[I + \frac{AT}{3} \left[\dots \frac{AT}{N-1} \left(I + \frac{AT}{N} \right) \right] \dots \right].$$

Gdy istnieje macierz A^{-1} , to macierz sterowania B_D wyraża się wzorem

$$B_D = \int_0^T e^{A\tau} B d\tau = A^{-1} (e^{AT} - I) B$$

lub

$$B_D = \sum_{i=0}^{\infty} \frac{A^i T^i}{(i+1)!} \cdot T \cdot B = \Psi \cdot T \cdot B.$$

Liniowe, macierzowe równanie dyskretne stanu z warunkami początkowymi x_0

$$x_{k+1} = A_D x_k + B_D u_k, \quad x(0) = x_0 \quad (5.12)$$

ma rozwiązanie otrzymane z rekurencji:

$$\begin{aligned} x_1 &= A_D x_0 + B_D u_0, \\ x_2 &= A_D x_1 + B_D u_1 = A_D^2 x_0 + A_D B_D u_0 + B_D u_1. \end{aligned}$$

Ostateczne rozwiązanie może być dane różnymi wzorami zawierającymi splot dyskretny

$$\begin{aligned} x_k &= A_D^k x_0 + \sum_{i=0}^{k-1} A_D^{k-i-1} B_D u_i = A_D^k x_0 + \sum_{i=0}^{k-1} A_D^i B_D u_{k-i-1} = \\ &= A_D^k x_0 + \sum_{i=1}^k A_D^{k-i} B_D u_{i-1} = A_D^k x_0 + \sum_{i=1}^k A_D^{i-1} B_D u_{k-i} \end{aligned} \quad (5.13)$$

Dla układu SISO

$$x_k = A_D^k x_0 + \begin{bmatrix} B_D & A_D B_D & \dots & A_D^{k-1} B_D \end{bmatrix} \cdot \begin{bmatrix} u_{k-1} \\ \vdots \\ u_0 \end{bmatrix}.$$

Równanie (5.12) wraz z algebraicznym równaniem wyjścia

$$y_k = C_D x_k + D_D u_k$$

daje pełny opis liniowego systemu dyskretnego, analogiczny do wzoru (3.6) dla układu ciągłego. Zamiast całki splotowej występuje w nim suma splotowa (splot dyskretny).

Przykład 5.1

Dla członu całkującąco-inercyjnego wyprowadzić model dyskretny.

$$G(s) = \frac{1}{s(s+1)} \Rightarrow \ddot{y}(t) + \dot{y}(t) = u(t),$$

$$\begin{aligned} y &= x_1, \\ \dot{x}_1 &= x_2, \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \Rightarrow \dot{x} = \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u, \quad y = [1 \ 0] x, \\ \dot{x}_2 &= -x_2 + u, \end{aligned}$$

$$A_D = e^{AT} = \mathcal{L}^{-1} \left[(sI - A)^{-1} \right] = \mathcal{L}^{-1} \left\{ \begin{bmatrix} \frac{1}{s} & \frac{1}{s(s+1)} \\ 0 & \frac{1}{s+1} \end{bmatrix} \right\} = \begin{bmatrix} 1 & 1-e^{-T} \\ 0 & e^{-T} \end{bmatrix},$$

$$B_D = \int_0^T e^{A\tau} B d\tau = \begin{bmatrix} T + e^{-T} - 1 \\ 1 - e^{-T} \end{bmatrix}, \quad A_D^k = e^{AkT} = \begin{bmatrix} 1 & 1-e^{-kT} \\ 0 & e^{-kT} \end{bmatrix}.$$

Przykład 5.2

Dla członu oscylacyjnego z tłumieniem wyprowadzić model dyskretny.

$$G(s) = \frac{1}{s^2 + 2\alpha s + \omega^2} = \frac{1}{(s+\alpha)^2 + \beta^2}, \quad \alpha^2 + \beta^2 = \omega^2, \quad s_{1,2} = -\alpha \pm j\beta,$$

$$A = \begin{bmatrix} 0 & 1 \\ -\omega^2 & -2\alpha \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

$$A_D = e^{AT} = e^{-\alpha T} \begin{bmatrix} \cos(\beta T) + (\alpha/\beta)\sin(\beta T) & (1/\beta)\sin(\beta T) \\ -(\omega^2/\beta)\sin(\beta T) & \cos(\beta T) - (\alpha/\beta)\sin(\beta T) \end{bmatrix},$$

$$B_D = \int_0^T e^{A\tau} B d\tau = \begin{bmatrix} (1/\omega^2)[1 - e^{-\alpha T}(\cos(\beta T) + (\alpha/\beta)\sin(\beta T))] \\ (1/\beta)e^{-\alpha T}\sin(\beta T) \end{bmatrix}.$$

5.3.2. Wykorzystanie splotu dyskretnego

Z rozdziału poprzedniego wynika, że podobnie jak dla układów ciągłych (podrozdz. 3.6) w układach dyskretnych splot dowolnego ciągu sterującego i impulsowej funkcji przejścia służy do wyliczenia odpowiedzi systemu. Przy zerowych warunkach początkowych, $x_0 = 0$, można pokazać, że dysponując pomiarami dyskretnymi wejścia i wyjścia na skończonym odcinku czasu, można odtworzyć w tym przedziale kształt impulsowej funkcji przejścia układu, czyli dokonać identyfikacji nieparametrycznej obiektu.

Wzór na odpowiedź y_k , dla $k \geq 0$, ma według (5.13) postać ($C = C_D$)

$$y_k = \sum_{i=0}^{k-1} C A_D^{k-1-i} B_D u_i + D_D u_k = \sum_{i=0}^k g_{k-i} u_i = \sum_{i=0}^k g_i u_{k-i}.$$

Powyżej przyjęte oznaczenie na impulsową funkcję przejścia dla $i \geq 1$

$$g_i = C A_D^{i-1} B_D,$$

potwierdza, że odpowiedź systemu na sterowanie impulsem $u_0 = \delta(0) = 1$ jest równa impulsowej funkcji przejścia

$$y_0 = D_D = g_0, \quad y_k = g_k \quad \text{dla } k \geq 1.$$

Ze wzoru (8.13) wynika, że dla $k \geq 1$, niezerowych warunków początkowych i $D_D = 0$ kolejne wartości wyjścia y_k systemu SISO mogą być przedstawione w postaci wektora:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} = \begin{bmatrix} CA_D \\ CA_D^2 \\ \vdots \\ CA_D^k \end{bmatrix} \begin{bmatrix} x_{10} \\ x_{20} \\ \vdots \\ x_{n0} \end{bmatrix} + \begin{bmatrix} u_0 & 0 & \dots & 0 \\ u_1 & u_0 & 0 & 0 \\ \vdots & \vdots & \ddots & 0 \\ u_{k-1} & u_{k-2} & \dots & u_0 \end{bmatrix} \begin{bmatrix} CB_D \\ CA_DB_D \\ \vdots \\ CA_D^{k-1}B_D \end{bmatrix}.$$

Ostatecznie, przy zerowych warunkach początkowych przebieg dowolnie długiego fragmentu impulsowej funkcji przejścia może być wyliczony przez odwrócenie macierzy sterowań U (gdy macierz U jest nieosobliwa):

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} = \begin{bmatrix} u_0 & 0 & \dots & 0 \\ u_1 & u_0 & 0 & 0 \\ \vdots & \vdots & \ddots & 0 \\ u_{k-1} & u_{k-2} & \dots & u_0 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_k \end{bmatrix} = U \cdot g \quad \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_k \end{bmatrix} = U^{-1} \cdot \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}.$$

Ze względu na trójkątną postać U kolejne wartości impulsowej funkcji przejścia mogą być szybciej obliczone bez odwracania U , z równania rekurencyjnego, co jest przydatne w zastosowaniach on-line:

$$g_1 = \frac{y_1}{u_0}, \quad g_k = \frac{1}{u_0} \left[y_k - \sum_{i=1}^{k-1} u_i \cdot g_{k-i} \right], \quad \text{dla } k = 2, 3, \dots$$

Szereg g_0, g_1, g_2, \dots zwany jest niekiedy szeregiem parametrów Markowa.

Z powyższych wzorów wynika również, że odpowiedź układu dyskretnego $y \in R^k$ na dowolne wymuszenie, może być przedstawiona za pomocą macierzy $G[k \times k]$, której kolejne kolumny są utworzone przez próbki impulsowej funkcji przejścia, kolejno poprzesuwane.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} = \begin{bmatrix} g_1 & 0 & \dots & 0 \\ g_2 & g_1 & 0 & 0 \\ \vdots & \vdots & \ddots & 0 \\ g_k & g_{k-1} & \dots & g_1 \end{bmatrix} \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_{k-1} \end{bmatrix} = G \cdot u.$$

Jeśli $g_0 = D_D \neq 0$, to $y \in R^{k+1}$, $G[(k+1) \times (k+1)]$

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_k \end{bmatrix} = \begin{bmatrix} g_0 & 0 & \dots & 0 \\ g_1 & g_0 & 0 & 0 \\ \vdots & \vdots & \ddots & 0 \\ g_k & g_{k-1} & \dots & g_0 \end{bmatrix} \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_k \end{bmatrix} = G \cdot u.$$

Podanie sterowania w postaci odpowiednika delty Diraca w chwili $k = 0$, $u = [1, 0, 0, \dots, 0]$, zrealizuje pierwszą kolumnę macierzy G , czyli impulsową funkcję przejścia.

Podanie sterowania w postaci delty Diraca w chwili $k = 1$ (przesuniętej o jedną próbke), $u = [0, 1, 0, \dots, 0]$, zrealizuje drugą kolumnę macierzy G , czyli też impulsową funkcję przejścia, ale przesuniętą o jeden okres. Jest to oczywista własność stacjonarności modelu.

Jeśli sterowanie jest skokiem jednostkowym $u_i = 1$ dla $\forall i \geq 0$, to jak było to podane wyżej

$$g_k = y_k - \sum_{i=1}^{k-1} g_{k-i} = y_k - \sum_{i=1}^{k-1} g_i.$$

Definiując ogólnie odpowiedź na skok jednostkowy dla $k \geq 0$ jako:

$$h_k = \sum_{i=0}^{k-1} CA_D^{k-i-1} B_D + D_D = \sum_{i=0}^{k-1} g_{k-i} + D_D = \sum_{i=0}^k g_i, \quad \text{gdzie } g_0 = D_D,$$

otrzymuje się znaną zależność pomiędzy impulsową funkcją przejścia i odpowiedzią na skok jednostkowy:

$$g_k = h_k - h_{k-1}, \quad g_0 = h_0 = D_D.$$

Wykorzystując ten wzór, łatwo zauważyc, że wyprowadzony wcześniej wzór na wyjście przyjmuje również postać

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} = \begin{bmatrix} g_1 & 0 & \dots & 0 \\ g_2 & g_1 & 0 & 0 \\ \vdots & \vdots & \ddots & 0 \\ g_k & g_{k-1} & \dots & g_1 \end{bmatrix} \cdot \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_{k-1} \end{bmatrix} = \begin{bmatrix} h_1 - h_0 & 0 & \dots & 0 \\ h_2 - h_1 & h_1 - h_0 & 0 & 0 \\ \vdots & \vdots & \ddots & 0 \\ h_k - h_{k-1} & h_{k-1} - h_{k-2} & \dots & h_1 - h_0 \end{bmatrix} \cdot \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_{k-1} \end{bmatrix}.$$

Zapis tej postaci na przyrostach sterowania $\Delta u_i = u_i - u_{i-1}$ pozwala operować próbками odpowiedzi na skok jednostkowy h_i , a nie impulsową funkcję przejścia g_i .

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} = \begin{bmatrix} h_1 & 0 & \dots & 0 \\ h_2 & h_1 & 0 & 0 \\ \vdots & \vdots & \ddots & 0 \\ h_k & h_{k-1} & \dots & h_1 \end{bmatrix} \cdot \begin{bmatrix} \Delta u_0 \\ \Delta u_1 \\ \vdots \\ \Delta u_{k-1} \end{bmatrix} - \begin{bmatrix} h_0 & 0 & \dots & 0 \\ h_0 & h_0 & 0 & 0 \\ \vdots & \vdots & \ddots & 0 \\ h_0 & h_0 & \dots & h_0 \end{bmatrix} \cdot \begin{bmatrix} \Delta u_0 \\ \Delta u_1 \\ \vdots \\ \Delta u_{k-1} \end{bmatrix} = H \Delta u - H_0 \Delta u.$$

Splot dyskretny dwóch sygnałów jest również podstawowym narzędziem w teorii liniowej filtracji i przy przetwarzaniu sygnałów, np. dźwięku lub obrazu (splot dwuwymiarowy).

5.4. Analiza układów dyskretnych w dziedzinie zmiennej zespolonej

5.4.1. Transformata Z

Jak wspomniano, w procesie modulowania za idealnym impulsatorem powstaje sygnał (rys. 5.7).

$$u^*(t) = \sum_{k=0}^{\infty} u(kT) \cdot \delta(t - kT).$$

Jego transformata Laplace'a wyraża się wzorem

$$U^*(s) = \int_0^{\infty} \left[\sum_{k=0}^{\infty} u(kT) \cdot \delta(t - kT) \right] e^{-st} dt = \sum_{k=0}^{\infty} u(kT) \cdot \int_0^{\infty} \delta(t - kT) e^{-st} dt = \sum_{k=0}^{\infty} u(kT) \cdot e^{-kTs}.$$

Przyjmując podstawienie (operator) $z = e^{Ts}$, wzór powyższy można przedstawić jako

$$U^*(s) \Big|_{e^{Ts}=z} = \sum_{k=0}^{\infty} u(kT) \cdot z^{-k}.$$

Wzór ten daje zależność pomiędzy przekształceniem Laplace'a a tzw. przekształceniem Z (oznaczenie \mathcal{Z}).

Poniżej przedstawione zostaną bardziej formalne własności tego przekształcenia, które będzie analogiem rachunku operatorowego Laplace'a dla układów dyskretnych.

Def. 5.1. Transformata Z funkcji impulsowej $x(n)$ dana jest wzorem:

$$\mathcal{Z}\{x(n)\} = X(z) = \sum_{n=-\infty}^{\infty} x(n) \cdot z^{-n} \quad (5.14)$$

$$\mathcal{Z}\{x(n)\} = \sum_{n=0}^{\infty} x(n) \cdot z^{-n},$$

gdy $x(n) = 0$ dla $n < 0$, gdzie z – bezwymiarowa zmienna zespolona na płaszczyźnie zespolonej.

Wymaga się, aby szereg (5.14) był zbieżny dla co najmniej jednego z . Przykładowo, oryginalny sygnał impulsowy o wartościach dyskretnych:

$x(-1) = 5, x(0) = 10, x(1) = 1, x(2) = 0, x(3) = 2$ oraz $x(n) = 0$, dla każdego innego n , ma zgodnie z (5.14) transformatę $X(z)$ o postaci

$$X(z) = 5z - 10 - z^{-1} + 2z^{-3} = 5z - 10 - \frac{1}{z} + 2 \cdot \frac{1}{z^3} = \frac{5z^4 - 10z^3 - z^2 + 2}{z^3}.$$

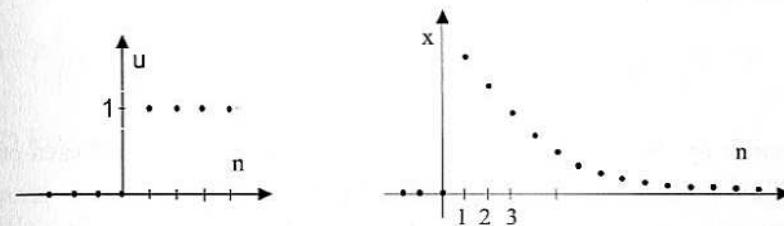
Problem zbieżności prześledzmy na innym przykładzie.

Niech sygnał dyskretny będzie dany wzorem

$$x(n) = \lambda^n \cdot u(n), \text{ gdzie } u(n) = \begin{cases} 0, & n \leq 0, \\ 1, & n > 0. \end{cases}$$

Transformata Z sygnału $x(n)$ z wykorzystaniem sumy szeregu geometrycznego jest równa

$$X(z) = \sum_{n=1}^{\infty} (\lambda \cdot z^{-1})^n = \lambda z^{-1} \cdot [1 + \lambda z^{-1} + \lambda z^{-2} + \dots] = \lambda z^{-1} \frac{1}{1 - \lambda z^{-1}} = \frac{\lambda}{z - \lambda} \Leftrightarrow |\lambda z^{-1}| < 1.$$

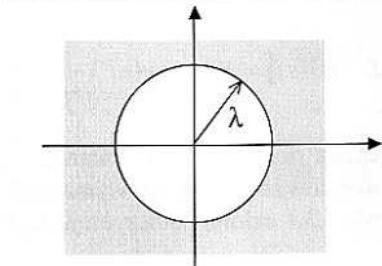


Rys. 5.8. Szereg czasowy

Sygnały $u(n)$ i $x(n)$ przedstawiono na rysunku 5.8.

$$X(z) = \frac{\lambda}{z - \lambda}, \text{ dla } |z| > |\lambda|.$$

Obszar zbieżności istnieje i znajduje się na zewnątrz koła o promieniu λ (rys. 5.9).



Rys. 5.9. Obszar zbieżności

Słuszne jest stwierdzenie, że transformata Z istnieje np. dla funkcji dyskretnych $x(k)$, które nie rosną szybciej niż funkcja wykładnicza.

Niektóre własności transformaty Z:

- Transformata sumy dwóch szeregów jest sumą ich transformat

$$y(n) = w(n) + x(n) \longrightarrow Y(z) = W(z) + X(z).$$

- Przesunięcie sygnału o k -próbek odpowiada przemnożeniu transformaty przez z^{-k} (opóźnienie)

$$\mathcal{Z}\{x(n-k)\} = z^{-k} \mathcal{Z}\{x(n)\} = z^{-k} X(z).$$

- Transformata splotu dwóch funkcji jest iloczynem transformat tych funkcji:

$$\mathcal{Z} \left\{ y(n) = \sum_{m=-\infty}^{\infty} x(m) \cdot h(n-m) \right\} = \mathcal{Z}\{x(n)\} \cdot \mathcal{Z}\{h(n)\},$$

$$\begin{aligned} \mathcal{Z}\{y(n)\} &= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} x(m) \cdot h(n-m) \cdot z^{-n} = \\ &= \sum_{m=-\infty}^{\infty} x(m) \cdot z^{-m} \cdot \mathcal{Z}\{h(n)\} = \mathcal{Z}\{x(n)\} \mathcal{Z}\{h(n)\}. \end{aligned}$$

- Wartość końcowa szeregu

$$\lim_{n \rightarrow \infty} y(n) = \lim_{z \rightarrow 1} (1 - z^{-1}) \mathcal{Z}[y(n)].$$

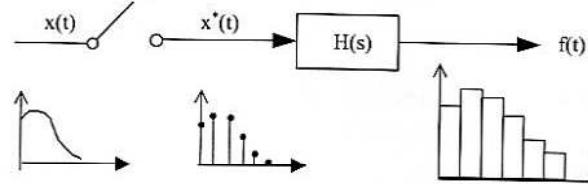
Rachunek operatorowy ze zmienną z pozwala na algebraizację obliczeń przebiegów dynamicznych z użyciem transmitancji dyskretnej $G(z)$

$$Y(z) = G(z) U(z).$$

Z pomocą próbki impulsowej funkcji przejścia układu ciągłego $g(t)$ teoretycznie można wyliczyć postać dyskretną transmitancji $G(z)$.

$$\mathcal{L}^{-1}[G(s)] = g(t) \rightarrow g(nT) \rightarrow \mathcal{Z}[g(nT)] = \sum_{n=0}^{\infty} g(nT) z^{-n} = G(z) \quad (5.15)$$

Jednak obliczanie transmitancji obiektu sterowanego $G(z)$ w praktycznych realizacjach układów impulsowych wymaga dla sterowania uwzględnienia wspomnianego w podrozdziale 5.3.1 członu podtrzymującego (rys. 5.10) zerowego rzędu (*Zero Order Hold*).



Rys. 5.10. Czon podtrzymujący

Transmitancja członu podtrzymującego sterowanie wynika z kształtu jego odpowiedzi na sygnał wejściowy, którym jest delta Diraca. Na wyjściu jest formowany sygnał prostokątny:

$$f(t) = 1(t) - 1(t - \gamma T), \quad F(s) = \frac{1}{s} - e^{-\gamma Ts} \cdot \frac{1}{s}.$$

C

Stąd transmitancja $H(s)$ jest otrzymywana ze wzoru

$$F(s) = H(s) \cdot 1 = \frac{1 - e^{-\gamma Ts}}{s}.$$

Jeżeli współczynnik wypełnienia γ lub okres T jest mały, to:

$$1 - e^{-\gamma Ts} = 1 - [1 - \gamma Ts + (\gamma Ts)^2 / 2 - \dots] \approx \gamma Ts, \text{ stąd } H(s) = \gamma T.$$

Dla układu ciągłego $G(s)$ pracującego szeregowo z członem podtrzymującym $H(s)$ obowiązuje ogólna zasada wyliczania wspólnej transmitancji dyskretnej w oparciu o wzór:

$$G(z) = \mathcal{Z} \left\{ \mathcal{L}^{-1} \left[G(s) \frac{1 - e^{-sT}}{s} \right] \right\} = (1 - z^{-1}) \mathcal{Z} \left\{ \mathcal{L}^{-1} \left[\frac{G(s)}{s} \right] \right\} = \frac{z-1}{z} Z\{h(kT)\} \quad (5.16)$$

gdzie $h(kT)$ jest ciągiem próbek odpowiedzi skokowej układu ciągłego.

Również, wyprowadzone poprzednio wzory (5.11) obowiązują przy założeniu, że w przedziale dyskretyzacji sterowanie jest stałe, a więc że w układzie zastosowany jest człon podtrzymujący.

Tabela 5.2 przedstawia przykładowe transformaty wybranych sygnałów czasowych bez członu ZOH i z członem ZOH.

Tabela 5.2
Transformaty z wybranych sygnałów

Oryginał	Transformata \mathcal{L}	Transformata \mathcal{Z} bez członu ZOH	Transformata \mathcal{Z} z członem ZOH
$1(t)$	$\frac{1}{s}$	$\frac{z}{z-1}$	$\frac{T}{z-1}$
t	$\frac{1}{s^2}$	$\frac{Tz}{(z-1)^2}$	$\frac{T^2(z+1)}{2(z-1)^2}$
$1(t) \frac{t^2}{2}$	$\frac{1}{s^3}$	$\frac{T^2 z(z+1)}{2(z-1)^3}$	$\frac{T^3(z^2+4z+1)}{3(z-1)^3}$
$1(t) e^{-at}$	$\frac{1}{s+a}$	$\frac{z}{z-e^{-aT}}$	$\frac{1-e^{-aT}}{a(z-e^{-aT})}$
$1(t) \sin(\omega t)$	$\frac{\omega^2}{s^2 + \omega^2}$	$\frac{z\omega \sin \omega T}{z^2 - 2z \cos \omega T + 1}$	$\frac{(z+1)(1-\cos \omega T)}{z^2 - 2z \cos \omega T + 1}$

Związki pomiędzy opisem w dziedzinie czasu i częstotliwości wynikają z własności splotu dyskretnego. Wzór na odpowiedź $y(k)$ ma postać

$$y(k) = \sum_{i=0}^{k-1} g(k-i)u(i).$$

Z porównania tego wzoru z ogólną postacią rozwiązania macierzowego równania stanu (5.13), przy zerowych warunkach początkowych,

$$y_k = C_D \sum_{i=0}^{k-1} A_D^{k-i-1} B_D u_i + D_D u_k$$

wynika, że impulsowa funkcja przejścia ma postać

$$g_k = \begin{cases} C_D A_D^{k-1} B_D & \text{dla } k \geq 1 \\ D_D & \text{dla } k = 0 \end{cases} \quad (5.17)$$

Odpowiedź na skok jednostkowy ma postać

$$h_k = C_D \sum_{i=0}^{k-1} A_D^{k-i-1} B_D + D_D = C_D \sum_{i=0}^{k-1} A_D^i B_D + D_D.$$

Zachodzi, więc wykazana w podrozdziale 5.3.2 zależność

$$h_{k+1} - h_k = C_D \sum_{i=0}^k A_D^i B_D - C_D \sum_{i=0}^{k-1} A_D^i B_D = C_D A_D^k B_D = g_{k+1}.$$

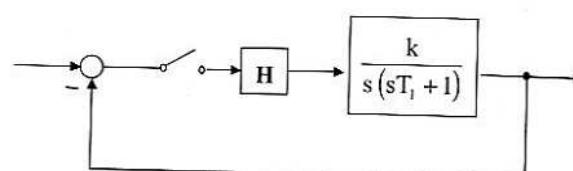
Odpowiadająca impulsowej funkcji (5.17) transmitancja $G(z)$ dana jest wzorem

$$G(z) = C_D [zI - A_D]^{-1} B_D + D_D \quad (5.18)$$

Jest to również postać odpowiadająca formule (5.16) i różna od wzoru (5.15).

Przykład 5.3

Wyznaczyć transmitancję układu zamkniętego z interpolatorem zerowego rzędu $H(s) = \gamma T$ (rys. 5.11).



Rys. 5.11. Schemat układu

$$G_0(s) = \frac{k}{s(sT_1 + 1)} = \frac{k}{s} - \frac{kT_1}{sT_1 + 1}, \quad g_0(t) = k - k \cdot e^{-\frac{t}{T_1}}$$

Dla $t = nT$:

$$G_0(z) = \sum_{n=0}^{\infty} kz^{-n} - k \sum_{n=0}^{\infty} e^{-\frac{nT}{T_1}} \cdot z^{-n}.$$

Oznaczając $\alpha = e^{-\frac{T}{T_1}}$, otrzymamy transmitancję układu otwartego:

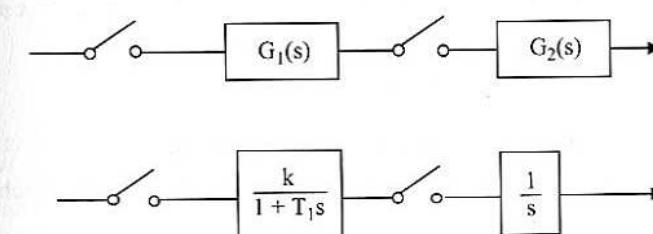
$$G_0(z) = k \cdot \frac{z}{z-1} - k \sum_{n=0}^{\infty} \left(\frac{1}{\alpha} \cdot z \right)^{-n} = k \cdot \frac{z}{z-1} - k \cdot \frac{z}{z-\alpha} = \frac{k(1-\alpha)z}{(z-1)(z-\alpha)},$$

i zamkniętego

$$G_Z(z) = \frac{G_0(z) \cdot \gamma T}{1 + G_0(z) \cdot \gamma T} = \frac{\gamma kT(1-\alpha)z}{z^2 - [1 + \alpha - kT(1-\alpha)z + \alpha]}.$$

Wyliczanie transmitancji zastępczych

Szeregowo połączenie transmitancji ciągłych (rys. 5.12) daje iloczyn transmitancji dyskretnych, tylko gdy między transmitancjami ciągłymi są impulsatory.



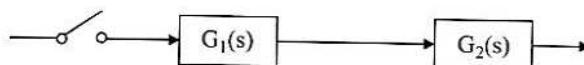
Rys. 5.12. Schemat szeregowy z impulsatorami

$$G(z) = G_1(z) \cdot G_2(z),$$

$$G(z) = \sum_{n=0}^{\infty} \left(\frac{k}{T_1} e^{\frac{nT}{T_1}} \cdot z^{-n} \right) \cdot \sum_{n=0}^{\infty} (1 \cdot z^{-n}) = \frac{z}{z-1} \cdot \frac{kz}{T_1(z-\alpha)}.$$

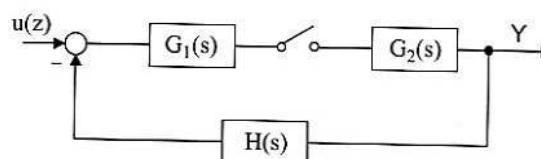
W przeciwnym wypadku transmitancja zastępcza $G(z)$ (rys. 5.13):

$$G(s) = G_1(s) \cdot G_2(s) \rightarrow g(t) \rightarrow G(z) = Z[g(nT)] = \frac{1-\alpha}{(z-1)} \cdot \frac{kz}{(z-\alpha)}.$$



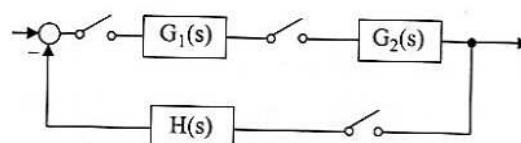
Rys. 5.13. Schemat szeregowy

Połączenia układów ze sprzężeniem zwrotnym przedstawiono na rysunkach 5.14 i 5.15.



Rys. 5.14. Schemat ze sprzężeniem zwrotnym

$$Y(z) = \frac{G_2(z) \cdot (G_1 \cdot u)(z)}{1 + (G_1 G_2 H)(z)}$$



Rys. 5.15. Schemat ze sprzężeniem zwrotnym

$$Y(z) = \frac{G_1(z) \cdot G_2(z) \cdot u(z)}{1 + G_1(z) \cdot G_2(z) \cdot U(z)}$$

Pewne własności odpowiedzi układów impulsowych

Poniżej wymienione są trzy własności transmitancji układów impulsowych:

- Dla równania różnicowego

$$y_k + a_1 y_{k-1} + \dots + a_n y_{k-n} = b_0 u_k + b_1 u_{k-1} + \dots + b_m u_{k-m},$$

któremu odpowiada transmitancja

$$G(z) = \frac{b_0 + b_1 z^{-1} + \dots + b_m z^{-m}}{1 + a_1 z^{-1} + \dots + a_n z^{-n}},$$

amplituda odpowiedzi w stanie ustalonym (jeśli taki istnieje) wynika z zależności:

$$y(k \rightarrow \infty) = G_\infty u(k \rightarrow \infty),$$

gdzie G_∞ wylicza się z twierdzenia granicznego

$$G_\infty = \lim_{z \rightarrow 1} \frac{y(z)}{u(z)} = \lim_{z \rightarrow 1} G(z) = \frac{b_0 + b_1 + \dots + b_m}{1 + a_1 + \dots + a_n}.$$

- W układzie ciągłym uwzględnienie opóźnienia transportowego zmienia transmitancję $G(s)$ z funkcji wymiernej na niewymierną funkcję s .

$$y(t) = u(t - \tau_o) \rightarrow G(s) = e^{-\tau_o s}.$$

W układach impulsowych, jeśli opóźnienie transportowe wyraża się jako całkowita wielokrotność czasu dyskretyzacji $d = \tau_o/T = 1, 2, 3, \dots$, to zgodnie z pierwszą własnością transformaty Z (podrozdz. 5.4.1) transmitancja nie zmienia swego zasadniczego charakteru.

$$G(z) = z^{-d}.$$

- W układach dyskretnych również występuje dyskretna transmitancja widmowa (stosunek amplitudy dyskretnej sinusoidy wyjściowej do amplitudy dyskretnej sinusoidy wejściowej) i istnieją analogi charakterystyk częstotliwościowych [SNIA2]. Z tymi charakterystykami wiąże się zjawisko stroboskopowe. Ponieważ $e^{j(\omega T)} = e^{j(\omega T + 2\pi n)}$ dla $n = 0, 1, 2, \dots$, to charakterystyki częstotliwościowe dyskretnych obiektów są określonymi funkcjami częstotliwości ωT , tzn. przyjmują jednakowe wartości dla $\omega = (2\pi/T)n$ i są jednoznaczne tylko wewnątrz przedziału $0 \leq \omega \leq 2\pi/T$. Stąd istnieje nieskończenie wiele sinusoid ciągłych, dla których to samo próbkowanie, da taką samą sinusoidę dyskretną (niejednoznaczność przejścia). Powyższy zakres częstotliwości jest podstawą w kształtowaniu (syntezie) filtrów cyfrowych.

5.4.2. Aproksymacje transformat dyskretnych

Omówienie problemu aproksymacji transformat zaczniemy od przykładu układu I rzędu. Zgodnie z tabelą 5.1 transmitancji członu inercyjnego bez interpolatora odpowiada $G(z)$.

$$G(s) = \frac{a}{s+a} \rightarrow G(z) = \frac{az}{z-e^{-aT}} \quad (5.19)$$

Jak przedstawiono w podrozdziale 5.4.1 pomiędzy operatorami z i s zachodzi zależność $z = e^{sT}$. Przyjmując różne aproksymacje funkcji e^{sT} , można otrzymać różne postacie uproszczonej transmitancji dyskretnej. Aproksymacje te mają swoją interpretację w przyjętej technice całkowania.

- Pierwsza aproksymacja związana jest z techniką całkowania „do przodu” (*forward*) lub pochodną prawostronną i wynika ze skróconego wzoru na z

$$z = e^{sT} = 1 + sT + s^2 T^2 / 2 + \dots \approx 1 + sT \Rightarrow s = \frac{z-1}{T}.$$

Podstawiając takie s do $G(s)$ ze wzoru (5.19), otrzymuje się

$$G_F(z) = \frac{a}{\frac{z-1}{T} + a} = \frac{aT}{z-1+aT}.$$

- 2) Druga aproksymacja związana jest z techniką całkowania „do tyłu” (*backward*) lub pochodną lewostronną i wynika ze skróconego wzoru na z

$$z = e^{sT} = \frac{1}{e^{-sT}} \approx \frac{1}{1-sT} \Rightarrow s = \frac{z-1}{Tz}.$$

Podstawiając takie s do $G(s)$ ze wzoru (5.19), otrzymuje się

$$G_B(z) = \frac{a}{\frac{z-1}{Tz} + a} = \frac{aTz}{z-1+aTz}.$$

- 3) Trzeci typ aproksymacji związany z techniką całkowania „regułą trapezów” (Tustina), zwany jest też transformacją biliniową i wynika ze skróconego wzoru (podrozdz. 4.3.5)

$$z = e^{sT} = \frac{1+e^{sT}}{1+e^{-sT}} \approx \frac{1+1+sT}{1+1-sT} = \frac{2+sT}{2-sT} \Rightarrow s = \frac{2}{T} \cdot \frac{z-1}{z+1}.$$

Podstawiając takie s do $G(s)$ z (5.19), otrzymuje się

$$G_T(z) = \frac{aT}{2\left[\frac{z-1}{z+1}\right] + aT} = \frac{aT(z+1)}{(2+aT)z+aT-2}.$$

Te reguły mnemotechniczne pozwalają na szybkie przejście od transmitancji ciągłej do dyskretnej poprzez zastąpienie operatora s przez odpowiednie wyrażenie z w zależności od metody aproksymacji:

$$1) \text{ forward } s = \frac{z-1}{T},$$

$$2) \text{ backward } s = \frac{z-1}{Tz},$$

$$3) \text{ Tustina } s = \frac{2}{T} \left[\frac{z-1}{z+1} \right].$$

Na przykład zastosowanie aproksymacji Tustina do transmitancji $G(s) = \frac{b_1s+b_0}{a_2s^2+a_1s+a_0}$ daje

$$G(z) = \frac{(2Tb_1+T^2b_0)z^2 + 2T^2b_0z + T^2b_0 - 2Tb_1}{(4a_2+2Ta_1+T^2a_0)z^2 + (2T^2a_0-8a_2)z + 4a_2 - 2Ta_1 + T^2a_0}.$$

Zastosowanie omawianych aproksymacji do wzoru na równanie stanu n -tego rzędu:

$$\dot{x}(t) = Ax(t) + Bu(t),$$

$$y(t) = Cx(t) + Du(t),$$

pozwoli skompletować postacie macierzy dla odpowiednich aproksymacji układów dyskretnych. Jak pokazano w podrozdziale 5.3 (wzór (5.11)), zastosowanie członu formującego zerowego rzędu prowadzi do dokładnej formuły równania dyskretnego stanu (gdy $\exists A^{-1}$):

$$x_{k+1} = \left[e^{AT} \right] x_k + \left[A^{-1} (e^{AT} - I) B \right] \cdot u_k = A_D x_k + B_D \cdot u_k,$$

$$y_k = C_D x_k + D_D u_k.$$

Natomiast zastosowanie kolejnych aproksymacji do transformaty Laplace'a równania stanu prowadzi do wzorów:

$$sX(s) = AX(s) + BU(s),$$

1) *forward*:

$$x_{k+1} = (I + AT)x_k + BTu_k,$$

$$y_k = Cx_k + Du_k,$$

2) *backward*:

$$x_k = (I - AT)^{-1} x_{k-1} + (I - AT)^{-1} BT \cdot u_k,$$

$$y_k = C(I - AT)^{-1} x_{k-1} + \left(D + C(I - AT)^{-1} BT \right) \cdot u_k,$$

3) Tustina:

$$w_{k+1} = (I + AT/2)(I - AT/2)^{-1} \cdot w_k + (I - AT/2)^{-1} B \cdot \sqrt{T} u_k,$$

$$y_k = \sqrt{T} \cdot C(I - AT/2)^{-1} \cdot w_k + \left(D + C(I - AT/2)^{-1} \cdot BT/2 \right) u_k,$$

$$w_{k+1} = \frac{1}{\sqrt{T}} \left(I + \frac{T}{2} A \right) x_k + \frac{1}{\sqrt{T}} \frac{BT}{2} u_k.$$

5.4.3. Obiekty o nieskończonej i skończonej odpowiedzi impulsowej

Sterowanie to aktywne oddziaływanie na obiekt w celu uzyskania pożdanego przebiegu zmiennej wyjściowej. Z reguły zadanie syntezy sterowania sprowadza się do zadania konstrukcji regulatora. W problematyce przetwarzania sygnałów w dziedzinie częstotliwości spotyka się również zadania syntezy innych liniowych układów, zwanych filtrami. Zakładając, że w sygnale wejściowym zawarte będą składowe harmoniczne, można postawić problem syntezy struktury pewnego obiektu, którego zadaniem byłoby wyłumienie w sygnale wyjściowym niepożdanego pasma, obecnego w sygnale wejściowym, czyli jego filtracja. Ciągłe filtry analogowe, tłumiące składowe zbędne i przepuszczające składowe pożądane stosowane były na dłużej przed filtrami cyfrowymi. Teoretycznie, charakteryzowały

się nieskończenie dugo trwającą postacią odpowiedzi impulsowej. Rozwój urządzeń cyfrowych i specjalne własności modeli dyskretnych spowodowały ich wykorzystanie w zadańach filtracji cyfrowej.

Filtry cyfrowe można podzielić na dwie główne grupy [5ZIT]:

- 1) filtry mające nieskończoną odpowiedź impulsową NOI (*Infinite Impulse Response – IIR*), zwane filtrami rekursywnymi;
- 2) filtry mające skończoną odpowiedź impulsową SOI (*Finite Impulse Response – FIR*).

Filtry NOI (IIR)

Ogólne równanie filtru IIR, jako podprzypadek modelu ARMAX($n, m, 0$) pokazane już zostało w podrozdziale 5.2 (wzór (5.6))

$$y_k = -\sum_{i=1}^n a_i y_{k-i} + \sum_{i=0}^m b_i u_{k-i}.$$

Bieżąca próbka wyjścia jest sumą ważoną n poprzednich próbek sygnału wyjściowego oraz sterowania bieżącego i jego m poprzednich próbek (dwa okna pomiarowe).

Odpowiedź impulsowa będzie określona na przedziele nieskończonym czasu, tzn. dla wszystkich $k \rightarrow \infty$ (może być malejąca, ale zawsze różna od zera).

Wykorzystując treść podrozdziałów 5.4.1 i 5.4.2 można wskazać metodkę syntezы filtru cyfrowego IIR. Wychodząc od transmitancji np. dolnoprzepustowego filtru ciągłego o częstotliwości odcięcia ω_c , można zaprojektować dolnoprzepustowy filtr cyfrowy o zbliżonej charakterystyce i innej częstotliwości odcięcia ω_d . Należy policzyć okres próbkowania T [s], przy którym dla obu częstotliwości odcięcia moduł obu transmitancji będzie ten sam.

Dyskretną aproksymację transmitancji ciągłej można otrzymać na podstawie aproksymacji Tustina

$$s = 2[(z-1)/(z+1)]/T \quad (5.20)$$

Przyjmując, że $z = e^{j\omega T}$, to moduł wyrażenia $(z-1)/(z+1)$ ma wartość

$$\begin{aligned} \frac{|e^{j\omega_d T} - 1|}{|e^{j\omega_d T} + 1|} &= \frac{|\cos(\omega_d T) - 1 + j \sin(\omega_d T)|}{|\cos(\omega_d T) + 1 + j \sin(\omega_d T)|} = \sqrt{\frac{2(1-\cos(\omega_d T))}{2(1+\cos(\omega_d T))}} = \\ &= \sqrt{\tan^2\left[\frac{\omega_d T}{2}\right]} = \tan(\omega_d T / 2). \end{aligned}$$

Zakładając, że dla częstotliwości odcięcia moduł zmiennej $s = j\omega_c$ i moduł aproksymacji Tustina s (5.20) powinien być taki sam, uzyskuje się związek pomiędzy częstotliwościami odcięcia i okresem próbkowania T

$$\omega_c = 2 \tan(\omega_d T / 2) / T \Rightarrow \omega_c T / 2 = \tan(\omega_d T / 2).$$

Rozwiązywanie powyższego równania ze względu na T może być znalezione numerycznie metodą odwzorowań zwężających dla funkcji odwrotnej (lepsza zależność) w przedziale $[\pi, 1,5\pi]$ radianów

$$\pi + \arctan(\omega_c T / 2) = \omega_d T / 2 \Rightarrow T = 2 \cdot (\pi + \arctan(\omega_c T / 2)) / \omega_d \quad (5.21)$$

Przykład 5.4

Dla ciągłego filtru

$$G(s) = \frac{1}{s^2 + \sqrt{2}s + 1} \quad (5.22)$$

o częstotliwości odcięcia $\omega_c = 1$ rad/s można podać trzy różne czasy dyskretyzacji gwarantujące różne założone częstotliwości odcięcia filtru dyskretnego otrzymane z (5.21), np.:

- 1) $\omega_d = 10$ rad/s $\Rightarrow T_1 = 0,69523,$
- 2) $\omega_d = 50$ rad/s $\Rightarrow T_2 = 0,12822,$
- 3) $\omega_d = 100$ rad/s $\Rightarrow T_3 = 0,06346.$

Wybierając odpowiednie T_i i podstawiając odpowiadające mu s ze wzoru (5.20) do wzoru (5.22), otrzymuje się odpowiednią transmitancję dyskretną filtru, np.:

$$G_1(z) = \frac{1 + 2z^{-1} + z^{-2}}{13,344 - 14,551z^{-1} + 5,207z^{-2}},$$

$$G_3(z) = \frac{1 + 2z^{-1} + z^{-2}}{1038,621 - 1984,111z^{-1} + 949,489z^{-2}}.$$

Odpowiadające np. transmitancji $G_1(z)$ równanie różnicowe wyjścia

$$13,344y_k - 14,551y_{k-1} + 5,207y_{k-2} = u_k + 2u_{k-1} + u_{k-2},$$

przyjmuje wspomnianą na początku rozdziału rekursywną postać filtru

$$y_k = 1,09y_{k-1} - 0,39y_{k-2} + 0,075u_k + 0,15u_{k-1} + 0,075u_{k-2}.$$

Filtry SOI (FIR)

Filtr SOI jest podprzypadkiem filtru NOI, uzyskanym dla współczynników $a_i = 0$. Ogólne równanie filtru SOI ma więc postać

$$y_k = \sum_{i=0}^m b_i u_{k-i} \quad (5.23)$$

Transmitancja $G(z)$ dla takiego filtra wyrazi się wzorem $G(z) = \sum_{i=0}^m b_i z^{-i}$.

Transmitancję widmową uzyska się podstawiając za $z^{-1} = e^{-j\omega T}$

$$\hat{G}(j\omega) = \sum_{i=0}^m b_i e^{-i j\omega T}.$$

Warto zauważyć, że ponieważ odpowiedź układu liniowego dyskretnego wyraża się splotem wejścia i impulsowej funkcji przejścia układu, to przekształcony wzór (5.23)

$$y_k = \sum_{i=0}^m b_i u_{k-i} = \sum_{i=k-m}^k b_{k-i} u_i \quad (5.24)$$

może być interpretowany jako splot dyskretny sygnału wejściowego u_i i specjalnej impulsowej funkcji przejścia o skończonej długości (okno pomiarowe o długości m próbek). Współczynniki b_i stanowią obwiednię tego okna.

Wzór (5.23) może również reprezentować cyfrową wersję całki

$$y(t) = \frac{1}{T} \int_{-\infty}^{\infty} u(\tau) g(t-\tau) d\tau.$$

Okna pomiarowe $g(t)$ są najczęściej funkcjami parzystymi lub nieparzystymi.

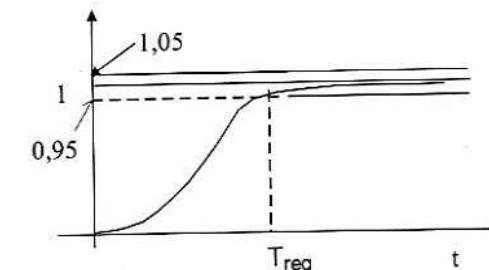
5.4.4. Wybór okresu dyskretyzacji

Wybór okresu próbkowania sygnału analogowego zależy od:

- pożąданej jakości sterowania,
- dynamiki procesu (najkrótsza stała czasowa),
- widma zakłóceń,
- możliwości silników i nastawników,
- aparatury pomiarowej,
- wymagań operatora (technologii procesu),
- możliwości sprzętu komputerowego,
- zadań sterowania i identyfikacji procesu.

Istnieją różne zalecenia praktyczne doboru czasu próbkowania. Można wyliczyć cztery najbardziej popularne:

- 1) Okres próbkowania może być przyjmowany w granicach $[0,07-0,2]T_{reg}$, gdzie T_{reg} jest czasem regulacji (rys. 5.16), czyli momentem wejścia odpowiedzi procesu w 5-procentowe otoczenie stanu ustalonego, po którym wyjście nie opuści tego otoczenia (czas T95).



Rys. 5.16. Czas regulacji

- 2) Ważnym wynikiem w teorii przetwarzania sygnałów jest twierdzenie Shannona, określające warunki konieczne dla częstotliwości próbkowania sygnału, gwarantujące jednoznaczna rekonstrukcję tego sygnału z próbek. Twierdzenie to mówi, że dla sygnałów ciągłych o skończonym paśmie, czyli mających skończoną maksymalną częstotliwość widma ω_{max} , częstotliwość próbkowania ω_0 musi być większa niż podwójna częstotliwość maksymalna występująca w sygnale.

$$\omega_0 > 2 \omega_{max} \text{ lub } T < \pi/\omega_{max}.$$

W praktyce, gdy występują wysokoczestotliwościowe zakłócenia, zaleca się mocniejsze ograniczenie $T < \pi/(a \cdot \omega_{max})$, gdzie $a \in [10 \div 20]$. Przykładowo dla cyfrowych nagrań muzycznych częstotliwość próbkowania przyjęto na poziomie 44 kHz ze względu na maksymalną słyszalną częstotliwość 20 kHz.

- 3) Dla obiektu wysokiego rzędu z wieloma stałymi czasowymi, częstotliwość próbkowania musi być większa niż odwrotność najwyższej stałej czasowej.
- 4) W procesach wolno zmiennych (chemia) czasy próbkowania mogą wynosić:
dla przepływów: $T = 1$ s, dla ciśnień: $T = 5$ s,
dla poziomów: $T = 10$ s, dla temperatur: $T = 20$ s.

6. Struktury systemów sterowania

6.1. Systemy automatyki procesowej i zabezpieczeniowej

Kompleksowy system automatyki, sterujący instalacją technologiczną, obejmuje różne układy automatyki, które można podzielić na dwie grupy:

- 1) układy automatyki procesowej DCS (*Distributed Control Systems*),
- 2) układy automatyki zabezpieczeniowej SIS (*Safety Instrumented Systems*).

Układy DCS pracują z reguły w konfiguracji systemu sterowania nadrzędnego, którego zadaniem jest znalezienie optymalnego punktu pracy ustalonej dla instalacji, doprowadzenie instalacji do tego punktu i prowadzenie instalacji w pobliżu tego punktu, co ma gwarantować uzyskanie produktu o wymaganych normach, przy jednoczesnej wysokiej wydajności i oszczędności procesu.

Układy SIS mają za zadanie pilnowanie bezpiecznej odległości zmiennych procesowych od ich ograniczeń, co ma gwarantować pozostawanie punktu pracy w obszarze stabilności. Układy te są odpowiedzialne za rozpoznawanie zakłóceń najbardziej niebezpiecznych i sytuacji awaryjnych. Do ich zadań należy alarmowanie o przekroczeniu ograniczeń technicznych i technologicznych, wyłączenie zepsutych i włączanie zapasowych urządzeń pomiarowo-kontrolnych takich, jak: czujniki, przetworniki, regulatory, nastawniki, tory pomiarowo-komunikacyjne (które pracują na zasadach gorącej rezerwy albo oczekiwania), zamykanie lokalnych niebezpiecznych cykli produkcyjnych, włączanie sekwencji blokad i awaryjne odstawianie instalacji.

Łączenie tych dwóch klas układów automatyki w jednych urządzeniach (np. w jednym sterowniku PLC – *Programmable Logic Controllers*) jest zabronione, zwłaszcza dla procesów, w których występuje ryzyko katastrofalnej awarii. Międzynarodowe unormowania zalecają dla instalacji przemysłowych rozdzielenie funkcji sterujących od funkcji zabezpieczeń. Funkcje sterujące realizując bowiem wiele obliczeń na podstawie ciągłyzych zależności dynamicznych, potrzebują cyfrowych układów liczących o wysokiej wydajności. Układy zabezpieczeń operując głównie na zależnościach statycznych i posługując się logiką binarną i sygnałami dwustanowymi, wykorzystują binarne układy I/O (ograniczniki, przekaźniki). Pomimo takich argumentów, jak: redukcja kosztów, wspólne zasilania, wspólna rejestracja danych, jedno stanowisko operatorskie, wysoka niezawodność układów elektronicznych –

Łączenie funkcji sterujących z funkcjami zabezpieczania w jednym urządzeniu nie jest prawidłowe. Unormowania dotyczące bezpieczeństwa, odnoszą się do trzech poziomów wymaganego bezpieczeństwa SIL (Safety Integrity Level):

- poziom 1 – zagrożenie małe,
- poziom 2 – zagrożenie średnie,
- poziom 3 – zagrożenie wysokie.

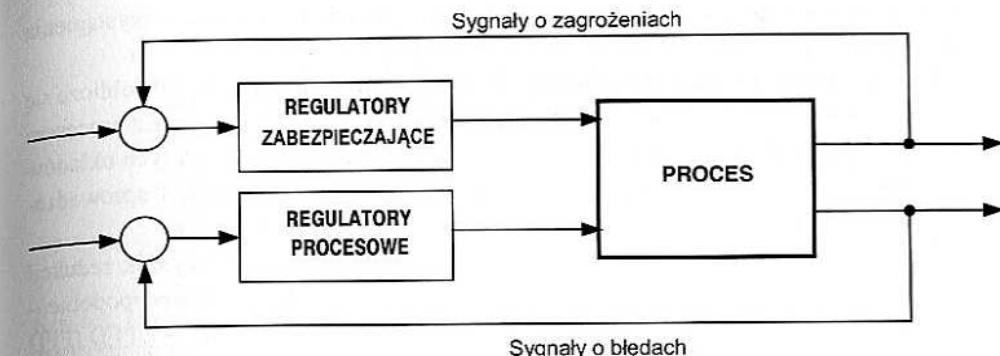
Jak wspomniano, w wielu krajach różne organizacje opublikowały zbiory dyrektyw i norm, w których stanowczo zaleca się pełne oddzielenie od siebie systemów automatyki procesowej i systemów zabezpieczeń.

Według opracowania [6GRP] należą do nich:

- Komisja IEC (International Electrotechnical Commision), która opublikowała w 1995 roku normę IEC61508/61511 dla wszystkich przemysłów używających w układach zabezpieczeń, urządzeń elektrycznych oraz programowalnych urządzeń elektronicznych „Bezpieczeństwo pracy – systemy związane z bezpieczeństwem”.
- Stowarzyszenie ISA (International Society for Measurement and Control), które opublikowało w 1996 roku dokument ISA S84 zatytułowany „Zastosowanie systemów zabezpieczających w przemyśle przetwórczym”.
- Brytyjska Komisja Zdrowia i Bezpieczeństwa UK HSE (UK Health & Safety Executive), która w 1987 r. opublikowała PES (*Programmable Electronic Systems in Safety Related Applications*), gdzie postuluje się fizyczną i funkcjonalną separację wszystkich elementów obu systemów: czujników, elementów wykonawczych CPU, modułów I/O oraz obudów [06UKH].
- Amerykański Instytut AIChE (American Institute of Chemical Engineers), który stworzył po wielkiej katastrofie chemicznych zakładów w Bhopalu, w Indiach, Centrum bezpieczeństwa procesów chemicznych (Center for Chemical Process Safety, 1993 r.).
- Stowarzyszenie NFPA (National Fire Protection Association), które opublikowało w 1995 r. normy projektowe NFPA 8502, dla układów sterowania kotłami oraz palnikami „Zapobieganie eksplozjom oraz implozjom w kotłach wielopalnikowych”.
- Amerykański Instytut API (American Petroleum Institute), który opublikował w 1994 roku zalecenia API RP14C dotyczące analizy, projektowania, montażu oraz testów podstawowych układów zabezpieczających dla instalacji na platformach wydobywczych.
- Amerykański Instytut IEEE (Institute of Electric and Electronic Engineers), który opublikował normę IEEE Std 603-19807 dotyczącą parametrów układów zabezpieczających w elektrowniach nuklearnych.

Podstawowa funkcjonalna struktura układów automatyki powinna więc mieć postać jak na rysunku 6.1.

W literaturze przedmiotu od lat 70. XX w. prowadzone są też prace nad komputerowymi systemami sterowania do zastosowań krytycznych, odpornymi na awarie (*fault-tolerant control*) [6BLM, 6LEN, 6NEP]. Badania doprowadziły do stworzenia systemów typu TMR (*Triple Modular Redundancy*), czyli o potrójnych układach wewnętrznych, stosowanych głównie w przemyśle lotniczym i kosmicznym.



Rys. 6.1. Schemat rozdzielnych systemów DCS i SIS

Systemy te zapewniają prawidłową pracę nawet przy awarii jednego z komputerów. Układ generujący nieprawidłowe sygnały jest wykrywany i odłączany. W zastosowaniach technologii przemysłowych takie rozwiązania mają sens tylko wtedy, gdy zredundowane są również pozostałe składowe układów sterowania takie, jak: czujniki, przetworniki, sieci, monitory. Obecnie produkowane są sterowniki PLC z podwójnymi procesorami. Znajazły one zastosowanie przy automatyzacji procesów wymagających wysokiej niezawodności, których awarie wiążą się z dużym ryzykiem. Standardy międzynarodowe określają takie układy jako **1oo2D**. Według niemieckiej normy TUV, takie systemy odpowiadają układom TMR. Rozwiązania mogą posiadać zarówno architekturę z podwójnymi procesorami i pojedynczymi układami I/O, jak i architekturę z podwójnymi procesorami i podwójnymi modułami I/O. Niska awaryjność to wynik rozwiązań konstrukcyjnych zakładających wytrzymałość na wysokie temperatury, wibracje, korozję, wpływy środowiska itp. Przemysłowe konstrukcje oparte są na kasetach i wymienne modułach – każdy przeznaczony do odrębnych zadań.

Z reguły sterowniki do zastosowań ogólnych nie gwarantują spełnienia standardów układów zabezpieczeń, które wymagają np. intensywnego samodiagnozowania sprawności procesora, pamięci i modułów wejść i wyjść, poziomów napięć i temperatur oraz testów sprawdzających okablowanie obiektowe pod kątem zwarć, rozwarć i komunikacji. Standardowe elektroniczne sterowniki programowalne, nie gwarantują bezpiecznego zachowania w czasie wewnętrznej awarii elektroniki – stosuje się więc zabezpieczanie wyjść modułu dodatkowym przekaźnikiem elektromechanicznym, który sprowadza wyjście do stanu bezpiecznego (*fail-safe*).

Uszkodzenia w systemie mogą być klasyfikowane jako „awarie bezpieczne” i „awarie groźne”. Awarie bezpieczne powodują tylko straty w produkcji, ale nie pogarszają bezpieczeństwa układów, stąd niepotrzebne jest wyłączenie instalacji. Awarie groźne blokują wyjścia w stanie zwarcia i nie pozwalają na zadziałanie blokad i awaryjnego wyłączenia instalacji. W pełni zredundowane systemy oferują najwyższe bezpieczeństwo, tzn. małe praw-

dopodobieństwo wystąpienia awarii groźnych i niskie prawdopodobieństwo wystąpienia awarii bezpiecznych.

Przy pracy wielu instalacji przemysłowych, szczególnie chemicznych, stale oblicza się prawdopodobieństwo awarii i koszt ewentualnych szkód. Gdy ryzyko związane z procesem staje się za duże, włączają się układy automatyki zabezpieczeniowej. Praca tych układów zawiera monitorowanie procesu, wykrywanie stanów groźnych i inicjację akcji spowodujących proces do stanu bezpiecznego (otwarcie zaworu, wyłączenie zasilania).

Najważniejszym parametrem układu zabezpieczającego jest współczynnik redukcji ryzyka RRF (*Risk Reduction Factor*), który definiuje się jako odwrotność prawdopodobieństwa, że układ zabezpieczający nie zadziała, gdy zajdzie tego potrzeba. $RRF = 1/PFD$ (*PFD – Probability of Failure on Demand*). Wszystkie elementy systemu mają swoje własne prawdopodobieństwa awarii – PFD. Wynoszą one od 0,05 do 0,0001. Współczynnik PFD całego systemu składa się z kombinacji prawdopodobieństw awarii wszystkich jego elementów składowych. W przeszłości najbardziej zawodnymi elementami układów zabezpieczających były sterowniki PLC, wyposażone w elektronikę nieprzystosowaną do wymagań stawianych przez układy zabezpieczeń. Obecnie najsłabszymi ogniwami obwodów zabezpieczających pozostają czujniki oraz elementy wykonawcze. Wiele firm podkreśla w swojej ofercie obecność specjalizowanych układów typu SIS. Są to na przykład układy ProSafe firmy Yokogawa [6YOK] czy QUADLOG firmy Moore [6MOO].

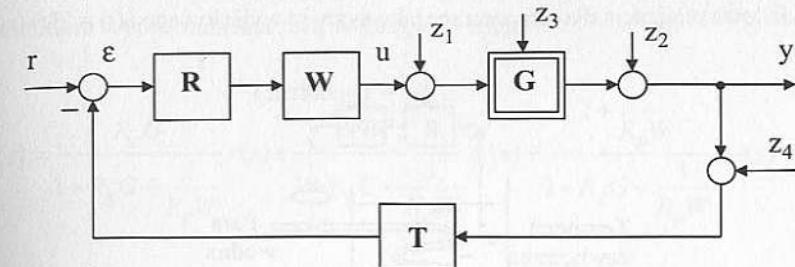
W niniejszej książce główny nacisk położono na problemy automatyki procesowej. W dalszych podrozdziałach omówione zostaną podstawowe struktury tych układów.

6.2. Ogólny schemat układu regulacji ze sprzężeniem zwrotnym

Podstawową przyczyną użycia sprzężenia zwrotnego jest umożliwienie osiągnięcia celu sterowania bez względu na zakłócenia sterowania, zakłócenie pomiaru wyjścia i zakłócenia parametrów obiektu sterowania (niepewność modelu). W wyniku sprzężenia zwrotnego powstaje sygnał błędu ε , który pobudza do pracy regulatora, a akcja regulatora redukuje efekt tego błędu, mimo stale działających zakłóceń (rys. 6.2).

Klasyczny układ regulacji kompensuje wpływ wszystkich zakłóceń, ale dla niektórych ich lokalizacji, np. dla zakłóceń z_1, z_3 , może to być układ mało efektywny. Główna pętla regulacji szybciej likwiduje wpływ zakłócenia z_2 niż z_1 . Strukturę regulatorów i ich nastawy dobiera się tak, aby jak najlepiej sprowadzały do zera błąd regulacji ε , który powstaje w czasie realizacji dwóch różnych zadań:

- 1) nadążania za planowaną zmianą sygnału referencyjnego (przejście do nowego punktu pracy),
- 2) likwidacji wpływu zakłóceń na pracę w wybranym punkcie (stabilizacja punktu pracy).



Rys. 6.2. Miejsce zakłóceń w układzie regulacji. r – wartość referencyjna dla regulatora (zadawana przez operatora), ε – błąd regulacji, u – sterowanie generowane przez regulator, y – wyjście z obiektu (wielkość regulowana), z_1 – zakłócenie sterowania, z_2 – zakłócenie wyjścia, z_3 – zakłócenie oddziałujące na parametry obiektu, z_4 – szumy pomiarowe, R – regulator, W – człon wykonawczy, np.: przetwornik, silownik, zawór, G – obiekt sterowania, T – tor pomiarowy (czujnik, przetwornik, transmisja)

Dla zakłóceń z_1 i z_2 i $T(s) = 1$ otrzymuje się transformatę sygnału wyjściowego $y(s)$

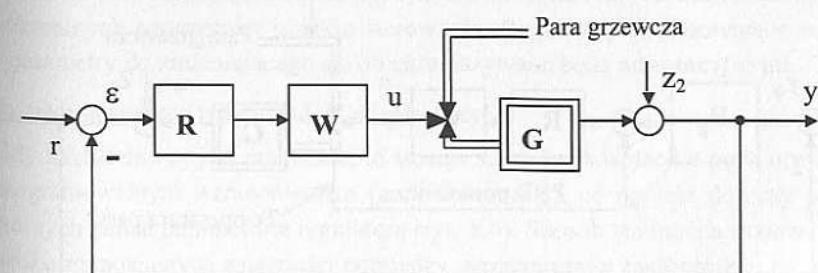
$$y(s) = \frac{RWG}{1+RWG} r(s) + \frac{G}{1+RWG} z_1(s) + \frac{1}{1+RWG} z_2(s) \quad (6.1)$$

z której widać transmitancje toru nadążania i torów zakłóceń.

6.3. Przykładowe realizacje jednowymiarowych układów regulacji SISO

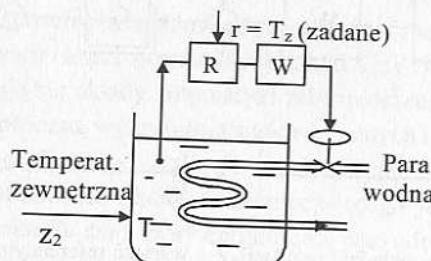
6.3.1. Układ klasyczny

Klasyczny układ stabilizacji sygnału wyjściowego y dla eliminacji wpływu zakłócenia z_2 ma postać jak na rysunku 6.3.



Rys. 6.3. Schemat regulacji

Przykładowa realizacja dla temperatury jako sygnału wyjściowego $y(t) = T(t)$ (rys. 6.4).



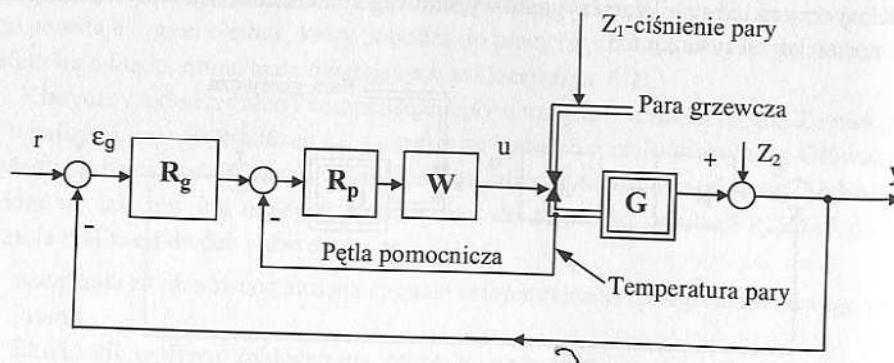
Rys. 6.4. Układ regulacji temperatury

Transformata wyjścia dana jest wzorem

$$y(s) = \frac{RWG}{1+RWG} r(s) + \frac{1}{1+RWG} z_2(s) \quad (6.2)$$

6.3.2. Układ kaskadowy

Gdy oprócz zakłócenia z_2 występują duże zakłócenia z_1 zmiennej sterującej u (np. wahania ciśnienia pary grzewczej, czyli jej temperatury), to w układzie klasycznym dodatkowo stosuje się układ dwuobwodowy – kaskadowy z regulatorem głównym i pomocniczym. Pomocnicza pętla sprzężenia zwrotnego o szybkiej dynamice (małej stałej czasowej) stabilizuje zmienną pomocniczą jaką jest np. temperatura pary grzewczej. Regulator główny, stabilizujący główną zmienną wyjściową, np. temperaturę w reaktorze, wypracowuje wartość zadawaną dla regulatora pomocniczego stabilizującego temperaturę pary grzewczej (rys. 6.5).



Rys. 6.5. Schemat i układ regulacji temperatury

Po krótkim wyprowadzeniu można otrzymać wzór

$$y(s) = \frac{R_g G}{1+R_g G + \frac{1}{R_p W}} r(s) + \frac{\frac{1}{R_p W} G}{1+R_g G + \frac{1}{R_p W}} z_1(s) + \frac{1 + \frac{1}{R_p W}}{1+R_g G + \frac{1}{R_p W}} z_2(s) \quad (6.3)$$

Jeżeli wzmacnienie R_p jest duże, to praktycznie eliminuje to wpływ zakłócenia z_1 na wyjście y .

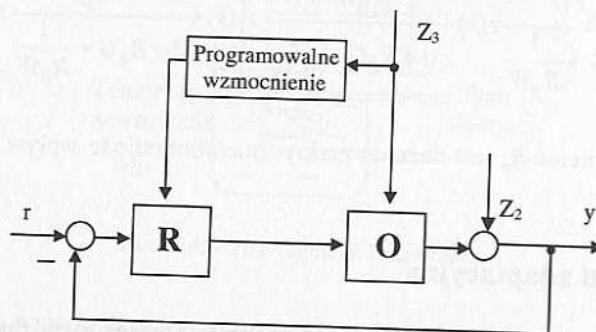
6.3.3. Układ adaptacyjny

Pojęcie adaptacji wiąże się zawsze z elementami procesu rozpoznawania i uczenia. W automatyce te zagadnienia były rozważane już w latach 60. XX w. R. Bellman akcentował w pozycji [6BER] niepewność modeli i związaną z tym konieczność wieloetapowości decyzji (programowanie dynamiczne). Z systemem adaptacyjnym sterowania będziemy więc mieli do czynienia, gdy np. regulator będzie musiał zmieniać optymalną strategię sterowania na skutek rozpoznanych zakłóceń zagrażających procesowi (bezpieczne zamiany cykli procesowych). Interpretowane jest to jako konieczność zmiany typu wskaźnika jakości, a stąd zmiany sterowania optymalnego w trakcie procesu sterowania. Jednak obecnie powszechnie przyjmuje się, że podstawowy typ układów adaptacyjnych wiąże się z faktem nieznajomości obiektu sterowania, lub choćby jego parametrów [6ASK, 6NIA]. Początkowa nieznajomość parametrów obiektu sterowania lub późniejsze ich zmiany w trakcie procesu sterowania utrudniają bowiem prawidłowe strojenie regulatora, a stąd prawidłowy proces stabilizacji. Można przyjąć, że taka zmienność (lub niepewność) parametrów modelu jest spowodowana jeszcze jednym typem zakłócenia z_3 w układzie regulacji. Bywa, że zmiany parametrów obiektu można skojarzyć ze zmianami dostępnych do pomiaru zmiennych zewnętrznych, co interpretowane jest jako możliwość pomiaru przyczyny zmian parametrów, czyli zakłócenia z_3 i dostosowywania regulatora do tego pomiaru. Przyczyny zmian czasem jednak nie są mierzalne i jedynym rozwiązaniem jest bezpośrednia identyfikacja aktualnych parametrów obiektu sterowania. Regulatory dostosowujące na bieżąco swoje parametry do zmieniającego się obiektu nazywane będą **adaptacyjnymi**.

Rozróżnia się dwa typy układów adaptacyjnych.

- 1) Gdy zakłócenie z_3 jest mierzalne, to stosuje się układ adaptacji z pętlą otwartą i tzw. programowalnym wzmacnieniem (*gain scheduling*), co ogólnie dotyczy programowalnych zmian parametrów regulatora (rys. 6.6). Sposób ten można stosować w przypadku rozpoznanych zależności pomiędzy parametrami a zakłóceniem, np. pomiędzy oporami przepływu w rurociągu a porą roku. Również różny punkt pracy układu nieliniowego generuje zmiany parametrów modelu zlinearyzowanego, np. parametry aerodynamiczne samochodu.

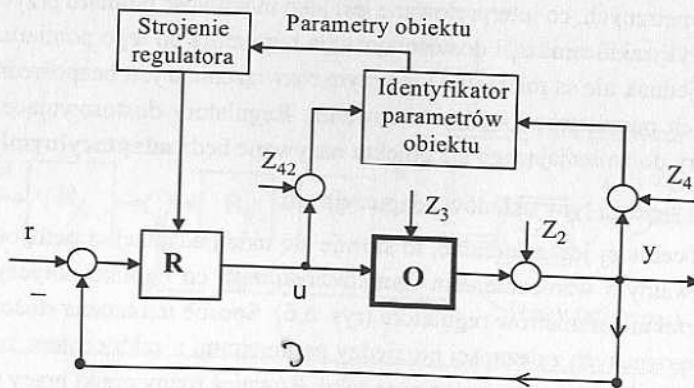
dynamiczne samolotu w funkcji wysokości lotu i prędkości samolotu. Gdy brak bezpośredniego pomiaru sygnału zakłócającego dany parametr, to jako sygnał z_3 , można wybrać inną mierzoną zmienną wiodącą, np. szybkość samolotu.



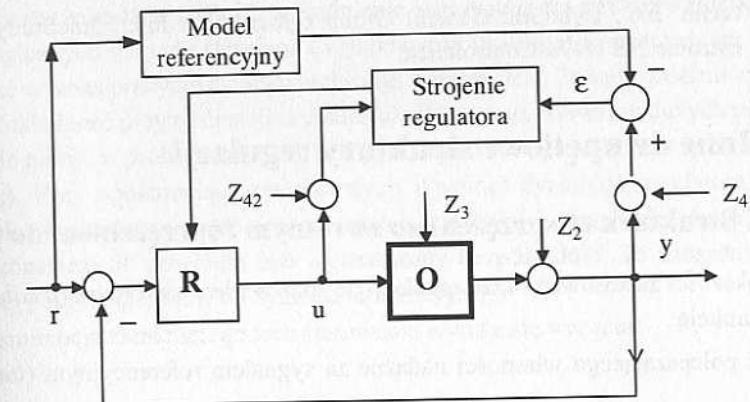
Rys. 6.6. Schemat adaptacyjnego układu z programowalnym wzmacnieniem

W zależności od zmierzonej wartości zakłócenia Z_3 układ programowo zmienia nastawy regulatora przy założeniu, że zależności pomiędzy zakłóceniami a charakterystykami procesu są dobrze znane. Nastawy regulatorów mogą być ujęte w tablicy. Cechą pozytywną jest prostota i szybkość reakcji na zakłócenia. Cechą ujemną jest otwartość pętli, czyli brak strojenia regulatora przy niemierzalnych zakłócenach.

- 2) Gdy zakłócenia Z_3 są niemierzalne, stosuje się układ adaptacyjny z pętlą zamkniętą – tzw. układ samonastrajalnego regulatora zwanego STR (Self Tuning Regulator). W regulatorze samonastrajalnym STR występuje blok, który automatycznie stroi parametry regulatora w zewnętrznej pętli albo pośrednio na podstawie estymowania wartości parametrów procesu (rys. 6.7) lub bezpośrednio za pomocą modelu referencyjnego układu zamkniętego MRAC (Model Reference Adaptive Control) (rys. 6.8).



Rys. 6.7. Schemat pośredniej adaptacji z identyfikatorem



Rys. 6.8. Schemat bezpośredni adaptacji z modelem odniesienia

Wybierając różne metody estymacji parametrów i różne algorytmy nastawiania, można otrzymać różne regulatory adaptacyjne.

W adaptacyjnych systemach sterowania on-line jako metody identyfikacji są stosowane:

- inżynierskie metody aktywnej identyfikacji parametrycznej z wykorzystaniem standardowych sygnałów skoku, impulsu Diraca lub sygnału sinusoidalnego, stosowane do ciągłych modeli I i II rzędu i z opóźnieniem;
- metody z funkcjami modulującymi, dla modeli ciągłych wyższych rzędów, wykorzystujące filtry splotowe, przy aktywnym lub pasywnym trybie wymuszeń;
- rekursywne metody pasywnej estymacji parametrów wykorzystywane on-line dla obiektów dyskretnych:
 - najmniejszych kwadratów RLS (Recursive Least Squares),
 - zmiennej instrumentalnej RIV (Recursive Instrumental Variable),
 - największej wiarygodności RML (Recursive Maximum Likelihood).

Dla najlepszego nastawiania regulatorów stosuje się metody:

- inżynierskie, np. Zieglera–Nicholsa;
- lokowania zer i biegunów transmitancji układu zamkniętego (z modelem odniesienia);
- metody minimalno-wariancyjne;
- metody liniowo-kwadratowe;
- metody predykcyjne;
- optymalne – dające minimum innych wybranych wskaźników jakości;
- oparte na zbiorach rozmytych lub sieciach neuronowych.

Wśród przemysłowych rozwiązań, znymi regulatorami adaptacyjnymi są regulatory Sipart DR 19, 21, 22, 24-Siemens, Exact-MV-Invensys Foxboro, BrainWave-Universal

Dynamics Techn. Inc., CyboSoft-General Group Cybernation Inc., QuicStudy-Adaptive Resources, Eftronic XS U498-Controlatica.

6.4. Inne dwupętlowe struktury regulacji

6.4.1. Struktura ze sprzężeniem zwrotnym i sprzężeniem do przodu

W większości zastosowań układ regulacji ze sprzężeniem zwrotnym (*feedback*) pełni podwójną funkcję:

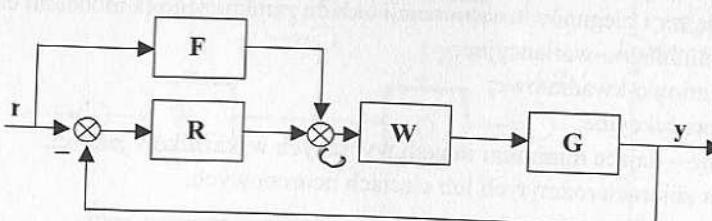
- 1) układu polepszającego własności nadążania za sygnałem referencyjnym (tor sterowania),
- 2) układu likwidującego wpływ zakłócenia wyjścia (tor zakłócienny).

Obie te własności mogą być dodatkowo wzmacnione poprzez zastosowanie drugich pętli tworzących tak zwane sprzężenie do przodu (*feedforward*). Pętle te dostarczają sterowania bezpośrednio, omijając główny regulator. Przedstawimy dwie odmiany takich struktur.

Ad 1)

Struktura ze sprzężeniem zwrotnym i sprzężeniem do przodu w torze nadążania [6LUB]

Struktura ze sprzężeniem do przodu (*feedforward*, rys. 6.9) ma podstawowe znaczenie przy budowie uniwersalnego schematu regulacji pełniącego rolę szybkiego układu nadążania i dokładnego układu stabilizacyjnego. Szybka zmiana punktu pracy jest przenoszona bezpośrednio na człon wykonawczy z pominięciem dynamiki regulatora. Po dojściu do punktu pracy (gdy błąd jest równy zero) regulator proporcjonalny dający sterowanie $u = K_F e$ przestaje być aktywny, jednak stały sygnał sterujący $u_b = rF$ jest konieczny dla utrzymania stanu ustalonego. Wszystkie cyfrowe regulatory PLC mają możliwość sumowania na swoim wyjściu sygnału sterującego u i stałego zewnętrznego sygnału u_b (*bias*). Dobór transmitancji F jako odwrotności wzmacnienia obiektu, gwarantuje również błąd ustalony równy zero. Błąd ustalony może mieć miejsce przy regulatorach proporcjonalnych i obiektach inertycznych.



Rys. 6.9. Schemat ze sprzężeniem do przodu

W literaturze można znaleźć uzasadnienie stosowania tej struktury również w układach biologicznych. Układy sterujące i wykonawcze (mięśnie) w wolnych akcjach muszą wykazywać wysoką precyzję działania (chirurg, zegarmistrz). Te same mięśnie mogą wykazać małą dokładność przy realizacji szybkich akcji i przy uzyskiwaniu dużych przyspieszeń (np. łapanie piłki). W wolnych akcjach precyzja uzyskiwana jest poprzez sprężenie zwrotne (wzrok). Przy opóźnieniach pomiarowych i wolnej dynamice regulatora sprężenie zwrotne nie gwarantuje jednak wystarczającej szybkości i sprawności działania. Wtedy człon wykonawczy W powinien być wysterowany bezpośrednio. To uzasadnia obecność dodatkowej pętli *feedforward* od sygnału referencyjnego.

Transmitancja zamkniętego toru sterowania wyraża się wzorem

$$G_z = \frac{y}{r} = \frac{WG(F + R)}{1 + WGR} \quad (6.4)$$

Transmitancja $G_z = 1$ dla:

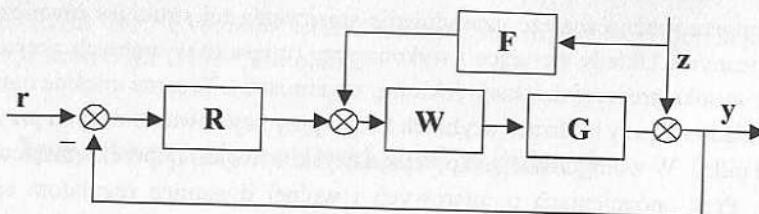
- $R \rightarrow \infty$,
- transmitancji $F = (WG)^{-1}$ (równość przynajmniej w zakresie wysokich częstotliwości).

Przy niskich częstotliwościach (wolny ruch) wzmacnienie regulatora R jest wysokie i system głównie wykorzystuje sprzężenie zwrotne. Przy wysokich częstotliwościach (szybki ruch) wzmacnienie regulatora R jest niskie i dominujące staje się sprzężenie do przodu F . Przedstawiona metodologia budowy regulatorów wiąże się z teoretycznym wyliczeniem transmitancji odwrotnej do znanej transmitancji obiektu. W układach ciągłych może to być nierealizowalne w sposób dokładny dla wszystkich częstotliwości (stopień licznika większy od stopnia mianownika). W układach dyskretnych teoretycznie dłużość okna pomiarowego na sygnale błędu ε (prawa strona równania różnicowego regulatora) może być dłuższa niż na sygnale sterującym u (lewa strona równania regulatora) (podrozdz. 6.7). Zakłada się tylko aby licznik transmitancji obiektu nie posiadał niestabilnych zer (obiekt minimalnofazowy). Często, więc w przykładach sterowania położeniem wału silnika (G) z użyciem regulatora położenia (R) zaleca się, aby transmitancja (F) reprezentowała różniczkowanie. Wtedy regulator prędkości (W) otrzymuje dodatkowy sygnał przyspieszający akcję.

Ad 2)

Struktura ze sprzężeniem zwrotnym i sprzężeniem do przodu w torze zakłócenia [6MAJ]

Bywa, że zakłócenie wyjścia z jest mierzalne (rys. 6.10). Można wtedy kompensować jego wpływ poprzez dodatkową pętlę sprzężenia do przodu od zakłócenia [6MAJ].



Rys. 6.10. Schemat ze sprzężeniem do przodu

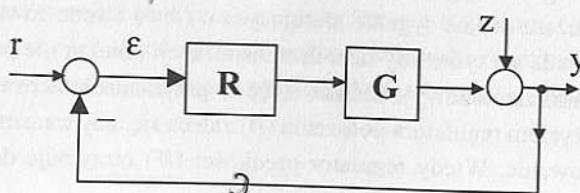
Transmitancja zamkniętego toru zakłócenia ma postać

$$G_z = \frac{y}{z} = \frac{1 + FWG}{1 + RWG} \quad (6.5)$$

Jeśli udałoby się aproksymować transmitancję kompensatora F , jako ujemną odwrotność transmitancji procesu $F = -(WG)^{-1}$, to transmitancja układu zamkniętego byłaby zero-wa $G_z = 0$ i otrzymano by całkowitą nieczułość na działanie zakłócenia z . Dla układu dyskretnego również tutaj obowiązuje powyższa uwaga o obiekcie minimalnofazowym.

6.4.2. Struktura z modelem wewnętrznym (IMC)

W 1989 r. M. Morari i E. Zafiriou [6MOM] opublikowali wyniki badań nad nową strukturą układu regulacji, nazwaną układem sterowania według modelu wewnętrznego IMC (Internal Model Control), która zawierała dwie pętle i gwarantowała odporność układu regulacji ze względu na niepewność modelu procesu. Przy wykorzystaniu tej metodologii zostały również podane reguły konstrukcji i strojenia regulatora [6SED]. Poniżej przedstawimy własności tej struktury. W tym celu przypomnijmy jeszcze raz jednopętlową strukturę podstawową (rys. 6.11).

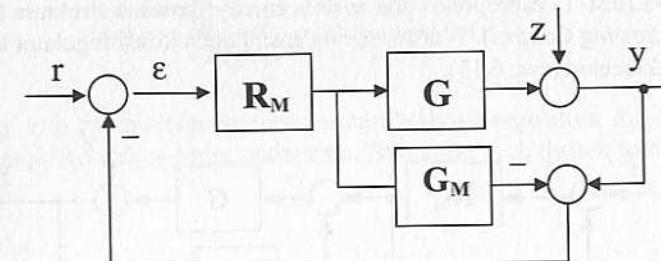


Rys. 6.11. Schemat regulacji

Transformata wyjścia wyraża się wzorem

$$y(s) = \frac{R \cdot G}{1 + RG} r(s) + \frac{1}{1 + RG} z(s).$$

Dwupętlowy układ IMC z wewnętrznym modelem obiektu i z nowym regulatorem R_M ma strukturę jak na rysunku 6.12.



Rys. 6.12. Schemat IMC

W tym układzie problem podstawowy to identyfikacja najlepszego modelu G_M odpowiadającego G . Wtedy na regulator R_M można podać odpowiednie wzory.

Nietrudno wyprowadzić zależność na transformatę wyjścia

$$y(s) = \frac{R_M \cdot G}{1 + R_M(G - G_M)} r(s) + \frac{1 - R_M G_M}{1 + R_M(G - G_M)} z(s) \quad (6.6)$$

Jeśli zachodzi równość transmitancji modelu i obiektu sterowania $G_m = G$, powyższa transformata redukuje się do

$$y(s) = R_M \cdot G \cdot r(s) + (1 - R_M G_M) z(s).$$

Teoretycznie gdyby udało się wtedy zaprojektować regulator $R_M = 1/G = 1/G_M$, to uzyskano by idealny tor nadążania całkowicie odporny na zakłócenie, w którym $y(s) = r(s)$. Powyższą zależność utrzymuje się nawet przy dowolnej nierówności modelu i obiektu $G_m \neq G$, jeśli tylko przyjmie się transmitancję regulatora $R_M = 1/G_M$ [6SKS].

Z drugiej strony, aby przynajmniej w torze głównym nadążania, transmitancje $y(s)/r(s)$ dla struktury podstawowej i struktury IMC były takie same, należy przyrównać do siebie

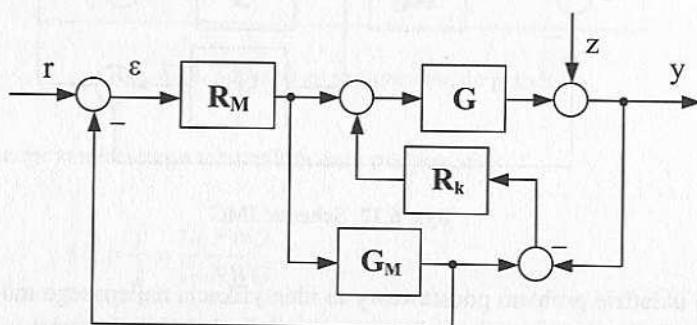
$$\frac{RG}{1 + RG} = \frac{R_M \cdot G}{1 + R_M(G - G_M)}.$$

Wynika stąd, że wtedy między regulatorami powinny zachodzić zależności:

$$R_M = \frac{R}{1 + RG_M} \quad \text{lub} \quad R = \frac{R_M}{1 - R_M G_M}.$$

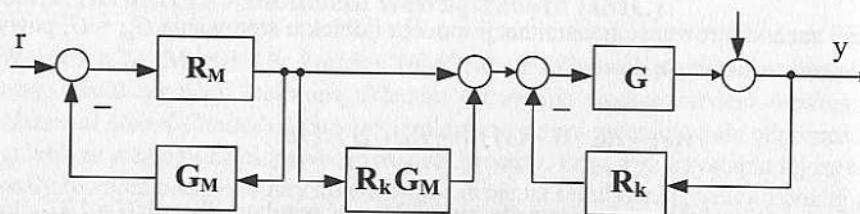
6.4.3. Zmodyfikowana struktura z modelem wewnętrzny (MFC)

W roku 1993 [6SUT] zaproponowana została zmodyfikowana struktura IMC nazwana MFC (*Model Following Control*). Wprowadzony został dodatkowy regulator korekcyjny R_k dla likwidacji zakłócenia (rys. 6.13).



Rys. 6.13. Schemat MFC

Aby uzyskać wyraźne zależności, przedstawimy prostszy schemat tej struktury (rys. 6.14).



Rys. 6.14. Schemat MFC

Transformata wyjścia spełnia zależność

$$y(s) = \frac{R_M \cdot (1 + R_k G_M) \cdot G}{(1 + R_M G_M) \cdot (1 + R_k G)} r(s) + \frac{1}{1 + R_k G} z(s) \quad (6.7)$$

Mogą zajść dwa przypadki:

- 1) Jeżeli $R_k = R_M$, to struktura traci swoje specjalne własności i staje się strukturą prostą (jednopiętową), w której jeden regulator $R = R_m = R_k$ wpływa na dwa tory regulacji.

$$y(s) = \frac{R \cdot G}{1 + R \cdot G} r(s) + \frac{1}{1 + RG} z(s).$$

- 2) Jeżeli uda się uzyskać równość modeli $G_M = G$, to uzyskujemy wzór

$$y(s) = \frac{R_M \cdot G_M}{(1 + R_M G_M)} r(s) + \frac{1}{1 + R_k G} z(s) \quad (6.8)$$

Świadczy to o całkowitym rozseparowaniu wpływu regulatora R_M dla nadajania za sygnałem r i regulatora R_k w torze zakłócenia. Dynamikę tych dwóch torów można kształtać osobno.

6.5. Elementy, układy i systemy automatyki

Kompleksowy system automatyki składa się z elementów, urządzeń i układów wymienionych poniżej.

Elementy automatyki są to członki spełniające w urządzeniu lub układzie proste funkcje takie jak np. wzmacnianie sygnału. Elementami automatyki będą więc: wzmacniacze, przekaźniki, zawory, silniki, czujniki, przetworniki elektropneumatyczne lub analogowo-cyfrowe.

Urządzenia automatyki to członki spełniające bardziej złożone funkcje takie, jak:

- elementy nastawcze z elementami napędowymi (całe siłowniki),
- urządzenia pomiarowe (czujnik, przetwornik, wskaźnik),
- urządzenia regulacyjne,
- urządzenia rejestrujące,
- urządzenia zasilające.

Układ automatyki to połączenie elementów i urządzeń w zespół wykonujący pewne zadanie:

- układ regulacji stałwartościowej,
- układ regulacji kaskadowej,
- układ regulacji nadążnej,
- układ centralnej rejestracji i wizualizacji danych,
- układ sterowania nadzawanego,
- układ sterowania programowego,
- układ sterowania optymalnego,
- układy zabezpieczeń.

System sterowania obejmuje wiele układów automatyki zainstalowanych na konkretnym obiekcie, np. w instalacji przemysłowej lub w samolocie.

Elementy i urządzenia automatyki klasyfikuje się:

- według rodzaju energii zasilającej na:
 - pneumatyczne,
 - elektryczne,
 - hydrauliczne,

- według sposobu działania na:
 - analogowe-ciągłe, gdy wielkość sygnału jest określona przez wartość wielkości fizycznej,
 - cyfrowe-dyskretne, gdy wartość sygnału jest określona liczbą,
- według funkcji spełnianych w układzie na:
 - pomiarowe,
 - wykonawcze,
 - korekcyjne,
 - regulacyjne,
 - filtracyjne.

W chwili obecnej większość układów i systemów automatyki jest zbudowana przy wykorzystaniu techniki cyfrowej i mikrokomputerów pracujących w sieciowych, rozproszonych systemach sterowania [6GRW].

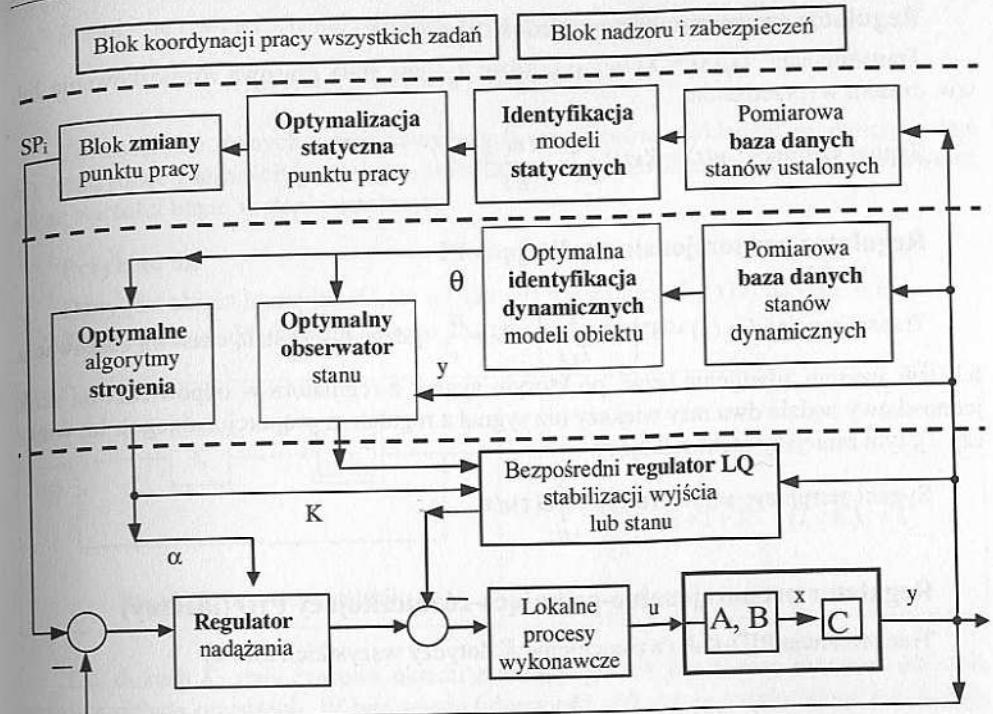
Przyczyny stosowania komputerowych systemów sterowania to:

- oszczędność w procesie produkcyjnym i łatwość serwisowania (modułowość),
- oszczędność w sprzęcie kontrolno-pomiarowym,
- niezawodność i odporność na zakłócenia zewnętrzne,
- większe bezpieczeństwo pracy,
- większa wydajność instalacji,
- podwyższenie kwalifikacji personelu operatorskiego,
- niska cena mikrokomputerów,
- małe wymiary i mały pobór mocy.

Celem systemu sterowania jest:

- stopniowe osiągnięcie optymalnego punktu pracy ustalonej, gwarantującego wymaganą jakość produkcji i maksymalny zysk,
- utrzymanie procesu w tym punkcie, pomimo działających zakłóceń (stabilizacja).

Cele te realizuje się, wprowadzając warstwy sterowania. W dolnej warstwie umieszczone są regulatory nadążające i stabilizacyjne, realizowane sprzętowo najczęściej za pomocą PLC. Zaawansowane algorytmy programujące i strojące te regulatory oraz algorytmy identyfikacji dynamicznej parametrów obiektu i obserwacji stanu umieszczają się w warstwie adaptacji, ulokowanej w komputerze nadzorującym. Następną warstwę stanowi właściwa warstwa sterowania nadzorującego, w której muszą być zaangażowane algorytmy akwizycji kolejnych wektorów stanów ustalonych dla identyfikacji charakterystyk modeli statycznych poszczególnych podsystemów technologicznych, składających się na cały proces. Za pomocą tych charakterystyk zadanie optymalizacji statycznej działające w przestrzeni zmiennych decyzyjnych poszukuje najlepszego punktu pracy (wektor liczbowy), dającego minimum wskaźnika jakości, przy uwzględnieniu powiązań między charakterystykami i ograniczeń na zmienne procesowe. Znaleziony punkt pracy, reprezentujący nowe wartości zadane SP (Set Point) dla regulatorów nadążających, z reguły nie powinien być przesyłany do nich w jednym cyklu. Ze względów bezpieczeństwa, zmiana realizowana jest stopniowo, poprzez blok zmiany punktu pracy (tranzycji) (rys. 6.15).



Rys. 6.15. Warstwowa dekompozycja systemu sterowania

6.6. Algorytmy ciągłej regulacji PID

Różne typy algorytmów regulacji wykazują różne własności, pracując w układzie zamkniętym. Do klasycznych algorytmów regulacji ciągłej należy liniowa kombinacja algorytmu proporcjonalnego, różniczkującego i całkującego, zaproponowana w 1922 roku przez N. Minorskiego. Zastosowanie zasady, że aktualna wielkość sterowania generowana w danej chwili przez regulator jest proporcjonalna do aktualnej wartości błędu, proporcjonalna do pola powierzchni błędu od momentu startu do chwili obecnej i proporcjonalna do aktualnej pochodnej błędu, dało na tyle dobre rezultaty, że powyższy algorytm w postaci sumy ważonej trzech równoległych akcji zastosowany w pierwszych przemysłowych regulatorach mechaniczno-pneumatycznych z lat 30. XX w., przetrwał do dziś i znajduje zastosowanie w cyfrowych regulatorach PLC.

Regulator proporcjonalny P

Transmitancja: $G_R(s) = K$, gdzie K jest współczynnikiem wzmacnienia.

Sygnał sterujący: $u(t) = K\varepsilon(t)$, $1/K \cdot 100\%$ zwany jest zakresem proporcjonalności.

Regulator proporcjonalno-różniczkujący PD

Transmitancja: $G_R(s) = K(1+T_D s)$, gdzie T_D jest stałą czasową różniczkowania lub tzw. czasem wyprzedzenia.

$$\text{Sygnał sterujący: } u(t) = K\varepsilon(t) + T_D \frac{d\varepsilon(t)}{dt}.$$

Regulator proporcjonalno-całkujący PI

Transmitancja: $G_R(s) = K \left(1 + \frac{1}{T_I s}\right) = K + \frac{\alpha}{s}$, gdzie T_I jest stałą czasową całkowania lub tzw. czasem zdwojenia (czas, po którym sygnał z regulatora w odpowiedzi na skok jednostkowy będzie dwa razy większy niż sygnał z regulatora proporcjonalnego). Im większa T_I , tym mniejszy efekt całkujący.

$$\text{Sygnał sterujący: } u(t) = K\varepsilon(t) + \alpha \int_0^t \varepsilon(\tau) d\tau.$$

Regulator proporcjonalno-całkująco-różniczkujący PID (idealny)

Transmitancja PID-ISA (wzmocnienie K dotyczy wszystkich torów)

$$G_R(s) = K \left(1 + \frac{1}{T_I s} + T_D s\right)^C \quad (6.9)$$

Transmitancja PID-IND (indywidualne wzmocnienie torów)

$$G_R(s) = K + \frac{\alpha}{s} + \beta s = \frac{\beta s^2 + Ks + \alpha}{s}.$$

Widac, że regulator PID reprezentuje człon nieprzyczynowy.

$$\text{Sygnał sterujący: } u(t) = K\varepsilon(t) + \alpha \int_0^t \varepsilon(\tau) d\tau + \beta \frac{d\varepsilon(t)}{dt}.$$

Regulator proporcjonalno-całkująco-różniczkujący PID z inercją (rzeczywisty)

$$\text{Transmitancja: } G_R(s) = K \left(1 + \frac{1}{T_I s} + \frac{T_D s}{Ts+1}\right) = K + \frac{\alpha}{s} + \frac{\beta s}{Ts+1} : T \ll T_D.$$

W przemysłowych rozwiązańach spotyka się również inne realizacje transmitancji PID, np. w adaptacyjnym regulatorze ECA600 firmy ABB Satt AB jest zastosowana struktura szeregowa z filtrem dolnoprzepustowym o stałej czasowej $T = 0,125 T_D$

$$G_R(s) = K \left(1 + \frac{1}{T_I s}\right) \left(1 + \frac{T_D s}{Ts+1}\right).$$

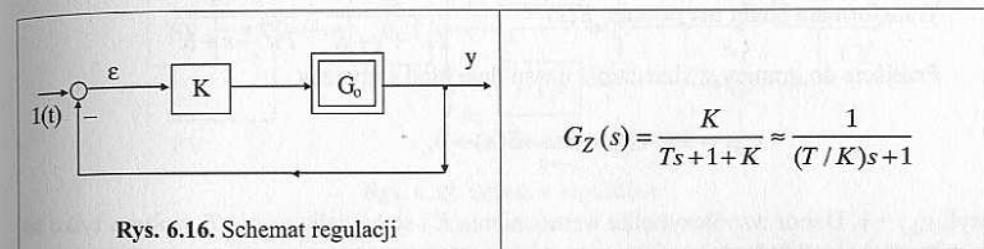
Stosuje się również równoległą strukturę gotowych członów PI i PD

$$G_R(s) = K_1 (1 + 1/T_I s) + K_2 (1 + T_D s).$$

Zastosowanie różnych wersji powyższych regulatorów w układzie zamkniętym daje nie tylko różne właściwości dynamiczne stabilizowanego przebiegu wyjścia $y(t)$, ale również różne wartości błędu w stanie ustalonym.

Przykład 6.1

Dany jest obiekt inercyjny $G_o(s) = 1/(Ts+1)$ i regulator $G_R(s) = K$ (rys. 6.16). Transmitancja układu zamkniętego jest rzedu pierwszego.



Rys. 6.16. Schemat regulacji

Dla dużych K , stała czasowa układu zamkniętego T/K jest K razy mniejsza niż stała czasowa układu otwartego. W tym sensie odpowiedź $y(t)$ układu zamkniętego jest K razy szybsza.

Na sygnał zadany $r(t) = 1(t)$, $y(t)$ ma przebieg wynikający ze wzoru

$$y(t) = \frac{K}{1+K} \left[1 - e^{-(1+K)t/T}\right],$$

a przebieg błędu regulacji jest dany wzorem

$$\varepsilon(t) = 1(t) - \frac{K}{1+K} \left[1 - e^{-(1+K)t/T}\right] = \frac{1}{1+K} + \frac{K}{1+K} e^{-(1+K)t/T} = \varepsilon_{ust} + \varepsilon_p(t).$$

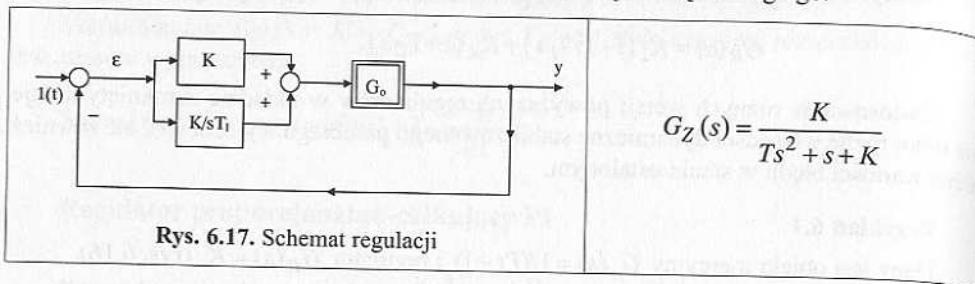
Błąd składa się z ustalonej wartości błędu statycznego ε_{ust} i zanikającego (dynamicznego) błędu przejściowego $\varepsilon_p(t)$. Błąd statyczny zeruje się w przypadku współczynnika wzmocnienia $K = \infty$. Dla K o skończonej wartości, wyjście ustalone równa się $y(\infty) = K/(K+1)$ i nie osiąga poziomu $r = 1$. Dobór współczynnika wzmocnienia K wpływa więc na wielkość błędu statycznego i kształt przebiegu błędu dynamicznego.

Przykład 6.2

Zastosowanie do obiektu inercyjnego regulatora PI o transmitancji $G_R(s)$ zmniejsza błąd statyczny do zera.

$$G_R(s) = K + K/(sT_I).$$

Transmitancja układu zamkniętego z rysunku 6.17 jest rzędu drugiego.



Rys. 6.17. Schemat regulacji

$$\text{Transformata błędu ma postać: } E(s) = \frac{1}{s} - \frac{1}{s} \cdot \frac{K}{Ts^2 + s + K} = \frac{Ts + 1}{Ts^2 + s + K}.$$

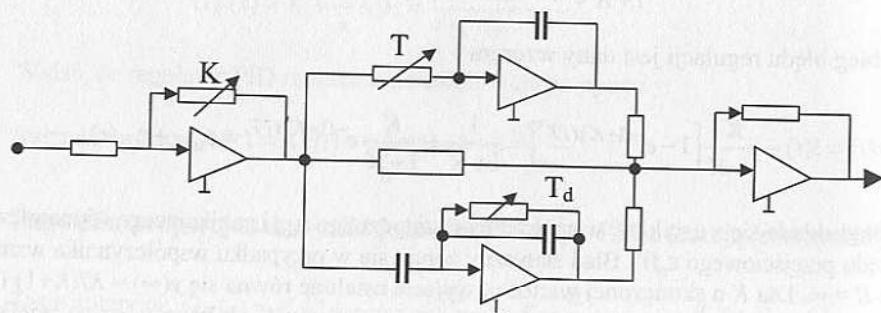
Przejście do granicy w dziedzinie czasu daje błąd statyczny

$$\epsilon_{ust} = \lim_{t \rightarrow \infty} \epsilon(t) = \lim_{s \rightarrow 0} sE(s) = 0,$$

czyli $y_{ust} = 1$. Dobór współczynnika wzmacniania K i stałej całkowania T_1 wpływa tylko na kształt przebiegu błędu dynamicznego.

6.7. Analogowe struktury regulatorów ciągłych

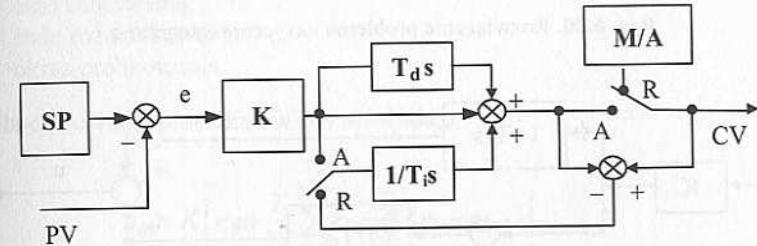
Dysponując elektronicznymi członami PID (podrozdz. 4.3.6), można zestawić je w strukturę regulatora jak na rysunku 6.18.



Rys. 6.18. Struktura elektroniczna PID

W rzeczywistych realizacjach występuje jeszcze dodatkowy moduł stacyjki manipulacyjnej zawierający nadajnik wartości zadanej (referencyjnej) tzw. *SET_POINT* (SP), przełącznik sterowania ręczne/automatyczne *MANUAL_AUTO* (M/A), nadajnik sterowania ręcznego, którym można generować sterowanie w układzie otwartym, wskaźniki pomiarowe wielkości wyjścia obiektu *PROCESS_VALUE* (PV) i sterowania *CONTROL_VALUE*

(CV). Regulator zawiera też inne bloki takie jak ograniczniki min-max sygnału wyjściowego z regulatora (czyli sterowania), przetworniki napięcie/prąd oraz ważny układ śledzący wartość sterowania ręcznego w układzie otwartym i generujący jednocześnie tą samą wartość sterowania przez regulator. Ten ostatni układ ma za zadanie zlikwidowanie różnic między tymi sygnałami w chwili przełączenia z ręcznego na automatyczne R/A i bez zakłóceń włączenie regulatora (zamknięcie pętli sprzężenia zwrotnego). Na rysunku 6.19 przedstawiono ideę działania tego układu. Główną rolę odgrywa tu człon całkujący i zespolony klucz R/A likwidujący różnice sterowań w modach R/A.



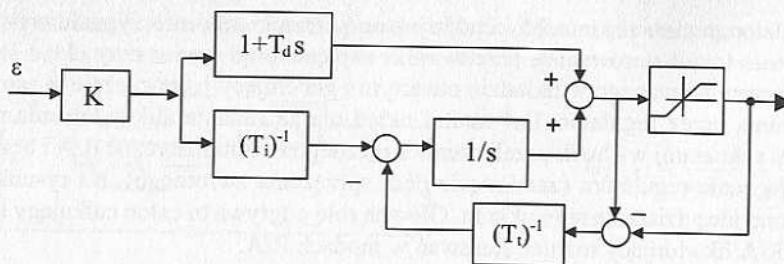
Rys. 6.19. Schemat regulatora

Często zdarza się, że zastosowany za regulatorem człon wykonawczy (nastawnik, żarówka) jest obiektem proporcjonalnym i pracuje w liniowym zakresie swojej charakterystyki statycznej. Gdy człon ten wchodzi w swoją strefę nasycenia, generuje stałą wartość wyjścia bez względu na wartość sygnału wchodzącego. Mimo że regulator dostarcza coraz większego sygnału sterującego, człon wykonawczy nie może go zrealizować, błąd regulacji się pogłębia i część całkująca regulatora PID generuje jeszcze większe sterowanie bez widocznych efektów. Dynamika procesu regulacji spada, a cały system zamknięty staje się procesem nieliniowym. Zjawisko to jest więc związane z nasyceniem działania całkującego i zwane jest zjawiskiem *windup* regulatora. Aby temu zapobiec, stosuje się różne modyfikacje w algorytmie regulatora zwane rozwiązaniami *anti-windup*. W pozycji literaturowej [6KUJ] wymienia się sześć różnych rozwiązań firmowych na przykładzie Siemens Sipart, ZPDA Ostrów Wielkopolski, Eftronik. Ogólnie stosuje się albo bezpośrednie ograniczenie sygnału wyjściowego z regulatora, albo wyłączenie jego akcji całkowania, bądź obie te akcje równocześnie.

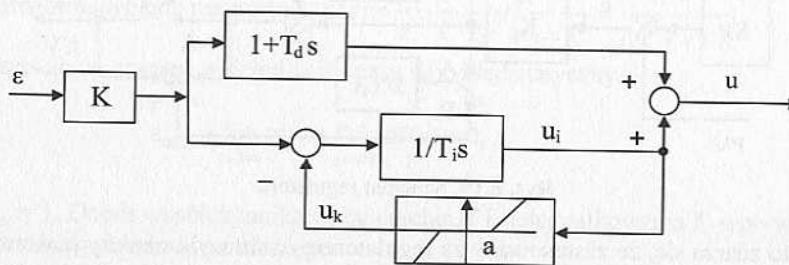
Poniżej przedstawiono dwie metody likwidacji zjawiska *windup*.

W pierwszej metodzie podejmuje się akcję gdy wartość sterowania przekroczy strefę proporcjonalności członu wykonawczego. Wtedy generowany jest sygnał różnicy pomiędzy poziomem nasycenia a poziomem sterowania i ta różnica odejmowana jest od wejścia na człon całkujący (rys. 6.20). Parametrem sterującym jest czas T_f .

W drugiej metodzie (rys. 6.21) stosowany jest dodatkowy człon ze strefą martwą o szerokości a równej szerokości strefy proporcjonalności członu wykonawczego. Dopóki sygnał sterujący u jest wewnątrz strefy proporcjonalności członu wykonawczego (o szerokości a), nic nie jest korygowane w pracy członu całkującego. Jeśli sygnał sterujący przekroczy szerokość charakterystyki liniowej członu wykonawczego a i wejdzie w strefę nasycenia, to sygnał wejściowy na człon całkujący jest zmniejszany o różnicę $(u_i - a)$ ze współczynnikiem wzmacniania b , tzn $u_k = b \cdot [u_i(t) - a]$.



Rys. 6.20. Rozwiązywanie problemu nasycenia integratora

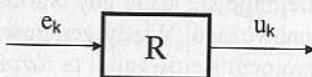


Rys. 6.21. Rozwiązywanie problemu nasycenia integratora

Porównanie jakości stabilizacji z zastosowaniem czterech różnych metod unikania efektu *windup* dla układów ciągłych omówiono w pracy [6BOC], a dla układów dyskretnych w pracy [6DEA].

6.8. Algorytmy dyskretne regulatorów PID

Sygnałem wejściowym do algorytmu regulatora jest sygnał błędu oznaczany $e(t)$, a sygnałem wyjściowym jest sygnał sterowania $u(t)$.



Rys. 6.22. Blok regulatora

W cyfrowych wersjach regulatorów (rys. 6.22) właściwym opisem algorytmu jest równanie różnicowe o postaci

$$u_k + a_1 u_{k-1} + a_2 u_{k-2} + \dots + a_n u_{k-n} = b_0 e_k + b_1 e_{k-1} + \dots + b_m e_{k-m}.$$

Poniżej rozważane zostaną trzy wersje dyskretne ciągłego równania regulatora PID

$$u(t) = K \left[e(t) + \frac{1}{T_I} \int_0^t e(\tau) d\tau + T_d \frac{de(t)}{dt} \right] \quad (6.10)$$

gdzie:

K – współczynnik wzmacniania regulatora,

T_I – czas całkowania,

T_d – stała różniczkowania,

T – okres próbkowania.

1) Reguła całkowania prostokątów (*forward*) daje

$$u_k = K \left[e_k + \frac{T}{T_I} \sum_{i=0}^{k-1} e_i + \frac{T_d}{T} [e_k - e_{k-1}] \right] \quad (6.11)$$

Reguła ta pozwala policzyć u_k od razu dla dowolnego k , przy czym zapamiętany jest przebieg błędu e_i w przedziale $[0, k-1]$. Algorytm powyższy zwany jest algorytmem **pozycyjnym**, i jak wspomniano, wymaga pamiętania wszystkich próbek błędu lub ich bieżącej sumy.

Częściej stosuje się algorytm **przyrostowy** zwany **prędkościowym**, który wykorzystuje policzoną w chwili poprzedniej wartość sterowania u_{k-1} i trzy ostatnie próbki pomiarowe błędu. Ponieważ zachodzi

$$u_{k-1} = K \left[e_{k-1} + \frac{T}{T_I} \sum_{i=0}^{k-2} e_i + \frac{T_d}{T} [e_{k-1} - e_{k-2}] \right],$$

stąd wzór (6.11) można wyrazić jako

$$u_k = u_{k-1} + b_0 e_k + b_1 e_{k-1} + b_2 e_{k-2} \quad (6.12)$$

$$\text{gdzie: } b_0 = K \left(1 + \frac{T_d}{T} \right), \quad b_1 = -K \left(1 + 2 \frac{T_d}{T} - \frac{T}{T_I} \right), \quad b_2 = K \cdot \frac{T_d}{T}.$$

2) Aproxymacja wsteczna *backward* (podrozdz. 5.4.2), dla której można przyjąć aproksymację $s = \frac{z-1}{Tz}$, zastosowaną do (6.9) daje popularną wersję transmitancji PID

$$G_R(z) = K \left[1 + \frac{Tz}{T_I(z-1)} + T_d \frac{z-1}{Tz} \right] = K \left[1 + \frac{T}{T_I(1-z^{-1})} + \frac{T_d}{T} (1-z^{-1}) \right] \quad (6.13)$$

czyli współczynniki w (6.12) wynoszą:

$$b_0 = K \left(1 + \frac{T_d}{T} + \frac{T}{T_I} \right), \quad b_1 = K \left(1 + 2 \frac{T_d}{T} \right), \quad b_2 = K \cdot \frac{T_d}{T}.$$

- 3) Reguła całkowania „trapezów” zastosowana do (6.9) prowadzi do innych współczynników:

$$b_0 = K \left(1 + \frac{T}{2T_I} + \frac{T_d}{T} \right), \quad b_1 = -K \left(1 + 2 \frac{T_d}{T} - \frac{T}{2T_I} \right), \quad b_2 = K \cdot \frac{T_d}{T}.$$

We wszystkich przypadkach transmitancja dyskretna PID wyraża się więc wzorem

$$G(z) = \frac{U(z)}{E(z)} = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 - z^{-1}}.$$

W ogólniejszym przypadku: $G(z) = \frac{U(z)}{E(z)} = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots + b_r z^{-m}}{1 - z^{-1}}$.

Dla różnych wykładników m transmitancja taka reprezentuje różne regulatory:

dla $m = 0$ – regulator P, dla $m = 1$ – regulator PI, dla $m = 2$ – regulator PID, dla $m = 3$ – regulator PIDD. Są one opisane równaniem różnicowym

$$u_k = u_{k-1} + b_0 e_k + b_1 e_{k-1} + \dots + b_m e_{k-m}.$$

Jeszcze ogólniejszy przypadek można uzyskać, biorąc pod uwagę efekt inercji w praktycznych realizacjach członu różniczkującego (6.9)

$$G(s) = K \left(1 + \frac{1}{T_I s} + \frac{T_d s}{1 + \hat{T}_d s} \right).$$

Dyskretna i dokładna transmitancja regulatora PID (z uwzględnieniem członu podtrzymującego ZOH) ma wtedy postać [6ISR]:

$$G_R(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 - a_1 z^{-1} + a_2 z^{-2}} = \frac{z-1}{z} Z \left[\frac{G_R(s)}{s} \right],$$

$$b_0 = K \cdot \left[1 + \frac{T_d}{\hat{T}_d} \right], \quad b_1 = -K \left[1 - \gamma + 2 \frac{T_d}{\hat{T}_d} - \frac{T}{T_I} \right], \quad b_2 = K \left[\frac{T_d}{\hat{T}_d} + \left(\frac{T}{T_I} - 1 \right) \gamma \right],$$

gdzie: $\gamma = -e^{-T/T_I}$, $a_1 = \gamma - 1$, $a_2 = -\gamma$; zwykle przyjmuje się $T_d / \hat{T}_d \in [3:10]$.

Odpowiednie równanie różnicowe ma postać:

$$\begin{aligned} u_k &= (\gamma - 1)u_{k-1} + \gamma u_{k-2} + b_0 e_k + b_1 e_{k-1} + b_2 e_{k-2} = \\ &= -u_{k-1} - \gamma(u_{k-1} + u_{k-2}) + b_0 e_k + b_1 e_{k-1} + b_2 e_{k-2}. \end{aligned}$$

W literaturze funkcjonuje wiele wzorów na dyskretny algorytm regulatora PID, ze względu na dużą liczbę możliwości:

- człon podtrzymujący uwzględniony lub nie,
- trzy różne aproksymacje,
- człon różniczkujący idealny lub rzeczywisty,
- człon całkujący idealny lub rzeczywisty,
- wersja pozycyjna lub przyrostowa.

6.9. Programowalne sterowniki logiczne PLC i nowe ich wersje PAC

Pierwsze cyfrowe sterowniki programowalne PLC powstały w roku 1968 w firmie General Motors i reprezentowały tylko elektroniczne, kompaktowe wersje sekwencyjnych logicznych układów przekaźnikowych z wejściami i wyjściami dwustanowymi. Znalały one głównie zastosowanie w systemach automatyki zabezpieczeniowej. Późniejsze rozwiązania rekonfigurowalne, które umożliwiły dołączenie do płyty głównej, modułów przetworników pomiarowych A/C i C/A oraz modułów kontrolnych PID, pozwoliły na szerokie ich stosowanie również w automatyce procesowej.

W chwili obecnej produkcją sterowników zajmuje się ponad 30 większych lub mniejszych firm takich, jak: Allen Bradley, Analog Devices, ABB (ASEA-Brown Bovery) AEG Schneider, Burr Brown, Cegelec, General-Electric Fanuc, Hartman Brown, Hewlett Packard, Honeywell, Modicon MODBUS, National Instruments, Omron PLC, Rockwell, SAIA-Burgess Electronic, Siemens, Schlumberger, Texas Instruments, Telemecanique, Westinghouse, Yokogawa i inne.

Od roku 1977 (Allen Bradley) w sterownikach PLC wykorzystuje się mikroprocesory.

Często wymieniane własności PLC [6LET] to:

- łatwość programowania i przeprogramowania stosownie do zmieniających się warunków,
- łatwość utrzymania w ruchu z możliwością napraw przez szybką wymianę modułów,
- większa niezawodność w warunkach przemysłowych,
- koszty porównywalne ze stosowanymi panelami przekaźnikowymi,
- komunikacja przez moduły portów Ethernet, TCP/IP, Profibus, USB (Universal Serial Bus).

Jakkolwiek struktura produkowanych obecnie przez różnych producentów sterowników jest ujednolicona, to z reguły sterowniki te nie są ze sobą kompatybilne. Producenci trzymają się zasad, aby wprowadzać na rynek całe rodziny swoich sterowników, co pozwala na zaspokojenie potrzeb użytkownika przy różnych wielkościach projektowanych instalacji.

Dla takiej rodziny sterowników charakterystyczne jest to, że poszczególne modele:

- mogą być programowane w tym samym języku za pomocą tego samego programu,
- posiadają takie same zmienne programowe oraz tę samą strukturę modułów I/O,
- istnieje możliwość przenoszenia programów między modelami.

Do zadań PLC należy:

- zbieranie danych pomiarowych za pośrednictwem modułów wejściowych z analogowymi i dyskretnymi czujników oraz urządzeń pomiarowych,
- transmisja danych za pomocą modułów i łącz komunikacyjnych,
- wykonywanie programów aplikacyjnych,
- generacja sygnału sterującego zgodnie z wynikami obliczeń i przekazanie go poprzez analogowe i dyskretne moduły wyjściowe do elementów wykonawczych,
- realizacja funkcji samodiagnozowania programowego i sprzętowego.

Sterowniki PLC są właściwie komputerami przemysłowymi, gdyż posiadają procesor, pamięć, układy I/O, własny system operacyjny czasu rzeczywistego oraz połączenie do urządzenia programującego i monitorującego. Pamięć dzieli się na pamięć programu zawartą w układach EPROM/ROM/FLASH i pamięć operacyjną RAM podtrzymywana baterijnie, wykorzystywaną do wykonywania programu i chwilowego pamiętania danych wejściowych i wyjściowych. Modularna budowa PLC umożliwia łatwą wymianę i dodawanie modułów na płycie (kasecie) głównej.

Na płycie znajdują się gniazda do podłączania od 4 do 10 modułów [6TRL]:

- modułu zasilacza i modułu procesora o różnej częstotliwości zegara ($10\text{--}64\text{ MHz}$);
- wejść i wyjść binarnych (dwustanowych) z optoizolacją, np. 32 kanały;
- wejść i wyjść analogowych, np. 8 kanałów;
- szybkiego licznika do zliczania impulsów (do 100 kHz), np. z enkoderów;
- komunikacji poprzez łączą szeregowe RS-232 (115 kbit/s) lub RS422/485 (187.5 kbit/s); poprzez sieć Ethernet, protokół Profibus, TCP/IP (100 Mbit/s), Web Server, USB, GPS;
- moduły rozszerzeń do 8 płyt dodatkowych.

Programowanie PLC odbywa się z wykorzystaniem języka programowania. Zagadniecie to było przedmiotem normy PN-EN61131-3/1998 i międzynarodowej normy IEC61131-3/1993.

Wyróżnia się dwie grupy języków:

- 1) języki tekstowe typu *Instruction List* lub *Structured Text*,
- 2) języki graficzne typu *Ladder Diagram* lub *Function Block Diagram*.

Zwłaszcza popularny jest język typu schematów drabinkowych *Ladder Diagram*, dający analogie do schematów stykowo-przełącznikowych..

Jak wspomniano, metodologia programowania sterowników została unormowana w 1993 r., przez Międzynarodową Komisję Elektroniki (IEC) normą IEC 61131, zatytułowaną *Programmable Controllers*, opartą na takich normach, jak: IEC 848 – Grafset (Francja), DIN 40719, DIN 19239, VDI 2880 (Niemcy), NEMA ICS-3-304 (USA). Programowanie odbywa się z panelu sterowania lub częściej z komputera PC, w którym należy zainstalować pakiet narzędziowy (np. dla sterowników GE-Fanuc – pakiet LOGICMASTER 90). Pakiet narzędziowy służy do stworzenia i przetestowania programu, a następnie do przesłania go poprzez interfejs szeregowy do pamięci PLC.

Cykl pracy sterownika

Część logiczna programu sterującego w sterowniku wykonywana jest w sposób powtarzalny, do momentu zatrzymania za pośrednictwem instrukcji z komputera nadziednego, programatora lub innego urządzenia zewnętrznego. Ciąg operacji koniecznych do jednorazowego wykonania części logicznej programu sterującego nazywany jest cyklem pracy sterownika i w zależności od sterownika i od wykonywanego programu trwa od 6 ms do np. 20 ms [06OPK]. Oprócz wykonywania części logicznej, na cykl pracy sterownika składają się również fazy: gromadzenia danych z urządzeń wejściowych, wysyłania informacji do urządzeń wyjściowych, wewnętrznej inicjalizacji sterownika, obsługi programatora oraz komunikacji z innymi urządzeniami.

Standardowy cykl pracy sterownika składa się z siedmiu faz, które przedstawia schemat blokowy przedstawiony na rysunku 6.23.

1. Inicjacja cyklu sterownika

Inicjacja cyklu sterownika polega na wykonaniu wszystkich operacji koniecznych do rozpoczęcia cyklu. Jeśli sterownik znajduje się w trybie pracy ze stałym czasem cyklu, rozpoczęcie nowego cyklu jest opóźniane do momentu, aż upłynie czas przewidziany na wykonanie poprzedniego cyklu.

2. Obsługa wejść

W czasie przeznaczonym na obsługę wejść dokonywane jest odczytywanie stanu wejść sterownika, bezpośrednio przed wykonaniem kolejnej fazy czyli części logicznej programu sterującego. W celu przyspieszenia wymiany informacji z periferiami w pamięci systemowej sterownika istnieją wydzielone obszary odwzorowujące stany wejść PII (*Process Image Input*) oraz wyjścia PIQ (*Process Image Output*). Zastosowanie obszaru odwzorowania stanu wejść znacznie skraca czas wykonywania programu, ponieważ czas dostępu do komórki pamięci jest krótszy niż czas potrzebny na odczytanie wejścia fizycznego, gdzie konieczna jest konwersja poziomu sygnału wejściowego na poziom cyfrowy. Dodatkową zaletą stosowania obszaru PII jest to, że w trakcie wykonywania całego programu stany wejść „widziane” przez poszczególne rozkazy będą takie same w całym cyklu obiegu pętli.

3. Wykonanie części logicznej programu sterującego

Program użytkownika zawsze zapisywany jest w postaci bloków programowych OB_x, gdzie „x” oznacza numer bloku. Każdy z programów powinien zawierać blok OB1 – jest to specjalny blok, który zawsze wywoływany jest automatycznie przez systemem operacyjnym sterownika. Blok ten stanowi „pomost” pomiędzy systemem operacyjnym sterownika a programem sterującym. Program sterujący wykonywany jest zawsze w kolejności instrukcji od pierwszej do ostatniej i zawiera z reguły logiczną sekwencję załączania i wyłączania dwustanowych przekaźników zestawionych zgodnie z regulami funkcji logicznych Boola, funkcji arytmetycznych lub funkcji sterujących (np. dyskretną wersja algorytmu sterowania PID).

4. Obsługa wyjść

Po zakończeniu pętli programowej – wykonaniu ostatniej instrukcji bloku OB1 system operacyjny sterownika uaktualnia stan wyjść poprzez przepisanie obszaru PIQ do fizycznych wyjść. Zakończenie fazy obsługi wyjść następuje w momencie, gdy wartości zmiennych przypisanych wyjściom zostaną przesłane do wszystkich modułów sterownika.

5. Komunikacja z programatorem

Faza ta jest wykonywana, gdy do sterownika podłączony jest programator lub moduł wymagający konfiguracji. Jeśli żaden z wymienionych warunków nie jest spełniony, faza komunikacji z programatorem nie jest wykonywana. Podczas jednego cyku może zostać skonfigurowany tylko jeden moduł.

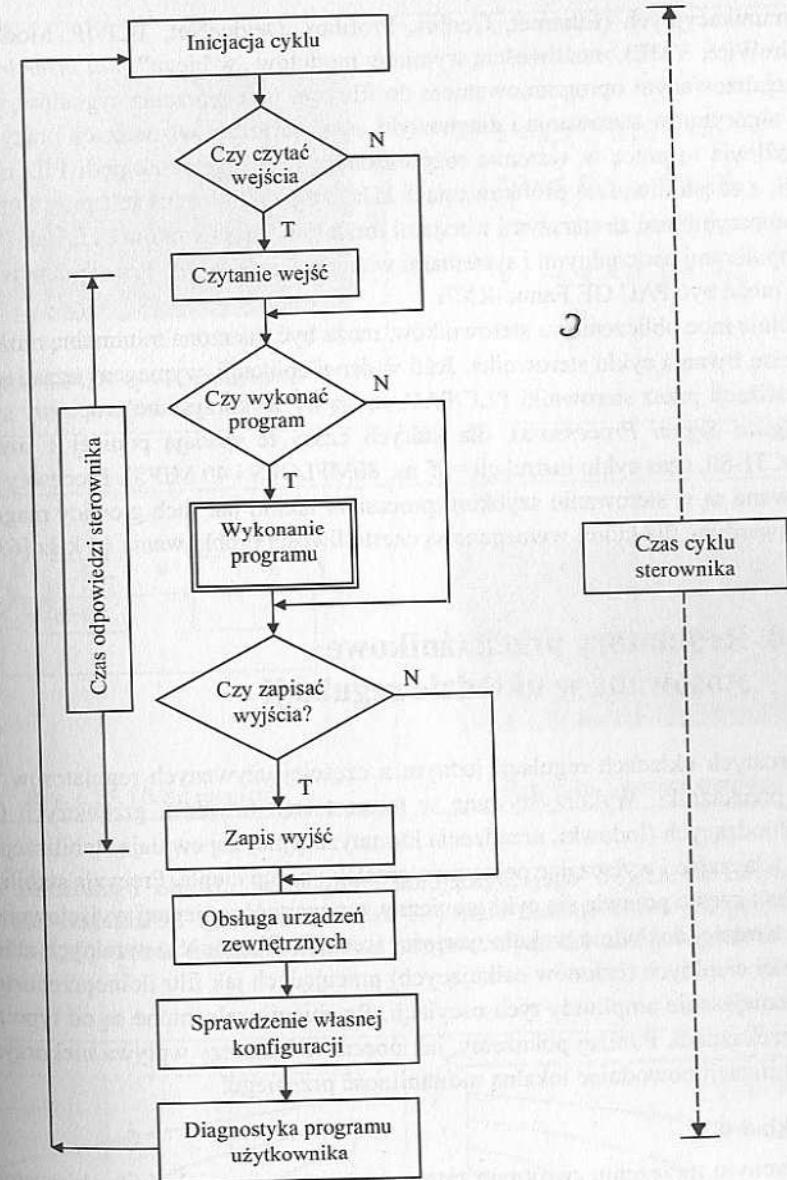
6. Komunikacja systemowa (z innymi urządzeniami)

Podczas fazy komunikacji systemowej realizowane są żądania komunikacji z modułami urządzeń dodatkowych, np. modułem szybkiego licznika, modułem komunikacyjnym itp. Żądania komunikacji są obsługiwane w kolejności napływu. Należy podkreślić, że żaden z modułów nie jest uprzywilejowany w stosunku do innych.

7. Diagnostyka

Jak wspomniano, cykl programowy sterownika PLC przedstawiony jest na rysunku 6.23. Na końcu każdego cyku pracy sterownika następuje obliczenie sumy kontrolnej programu sterującego. W przypadku gdy suma nie jest zgodna z wartością porównawczą, zgłoszany jest błąd, w wyniku którego następuje zatrzymanie pracy sterownika.

Na koniec należy zaznaczyć, że jakkolwiek sterowniki PLC mogą pracować samodzielnie jako wielowymiarowe regulatory, to obecnie stosuje się je najczęściej w warstwie najniższej DDC (*Direct Digital Control*) wielowarstwowych systemów sterowania nadzawanego. W warstwie wyższej znajduje się wtedy specjalizowany komputer przemysłowy PC, też z systemem czasu rzeczywistego, który wykonuje bardziej zaawansowane obliczenia algorytmiczne, np. optymalizacyjne. Wylicza wtedy okresowo nowy punkt pracy instalacji i poprzez kanał RS przesyła do PLC wartości zadawane do poszczególnych pętli regulacji PID.



Rys. 6.23. Cykl programowy sterownika PLC

Obecnie w automatyce procesowej wchodzi do użycia nowa generacja sterowników, która ma znacznie zwiększone możliwości działania. Dla tych sterowników pojawiła się też nowa nazwa – PAC (*Programmable Automation Controllers*). Charakteryzuje się one szybkimi zmiennoprzecinkowymi procesorami dużej mocy (np. Pentium 3/300), szybką magistralą PCI, dużą pamięcią operacyjną, jednoczesną obsługą wielu połączeń i wielu protoko-

łów komunikacyjnych (Ethernet, Genius, Profibus, DeviceNet, TCP/IP, Modbus, UDP, OPC, FireWire, VME), możliwością wymiany modułów „w biegu” (*hot insertion*) i bogatym specjalizowanym oprogramowaniem do filtracji, przetwarzania sygnałów, zaawansowanych algorytmów sterowania i diagnostyki oraz graficznej wizualizacji pracy sterownika. Umożliwia to pracę w systemie rozproszonym i obsługę setek pętli PID na odległej instalacji, z częstotliwością próbkowania 1 kHz. Zagwarantowana jest przy tym zarówno pełna kompatybilność ze starszymi wersjami modułów i sterowników PLC, jak i współpracą z komputerami nadzorującymi i systemami wizualizacji SCADA. Przykładem takiego sterownika może być PAC GE Fanuc-RX7i.

Ogólnie moc obliczeniowa sterowników, może być mierzoną minimalną możliwą wartością czasu trwania cyklu sterownika. Jeśli w danej aplikacji, wymagany czas jest za krótki dla realizacji przez sterowniki PLC/PAC, mogą być wykorzystane procesory sygnałowe DSP (*Digital Signal Processors*), dla których czasy te spadają poniżej 1 ms (np. dla TMS320C31-80, czas cyklu instrukcji =25 ns, 80MFLOPS i 40 MIPS). Procesory DSP wykorzystywane są w sterowaniu szybkimi procesami takimi jak ruch głowicy magnetycznej w dysku twardym, dla której wymagane są częstotliwości próbkowania 14 kHz [6ABD].

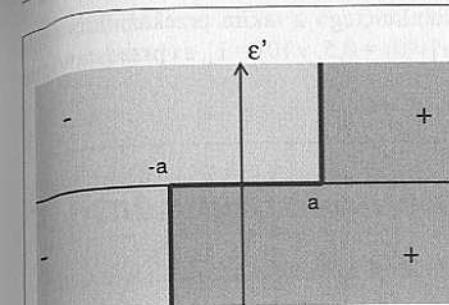
6.10. Regulator przekaźnikowe stosowane w układzie regulacji

W prostych układach regulacji jednym z częściej używanych regulatorów nielinowych są przekaźniki. Wykorzystywane w termostatach urządzeń grzewczych (żelazka, pralki) i chłodzących (lodówka, urządzenia klimatyzacyjne) zapewniają stabilizację temperaturową, włączając i wyłączając pełną moc grzałek i pomp ciepła. Precyzyja stabilizacji nie jest wysoka i często pojawia się cykl graniczny, tzn wartość zmiennej wyjściowej oscyluje mniej lub bardziej dokładnie wokół wartości średniej. Obecność naturalnych zbiorników i pojemności cieplnych (członów całkujących) pracujących jak filtr dolnoprzepustowy powoduje zmniejszenie amplitudy tych oscylacji. Przebiegi uzależnione są od typu zastosowanego przekaźnika. Poniżej pokażemy, jak obecność histerezy wpływa niekorzystnie na proces stabilizacji powodując lokalną niestabilność przebiegu.

Przykład 6.3

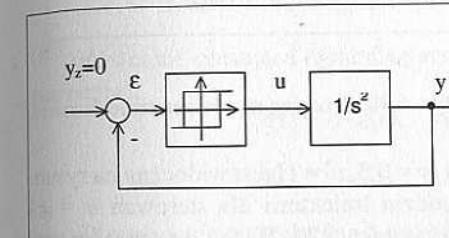
W ujemnym sprzężeniu zwrotnym zastosowany zostanie przekaźnik z histerezą (strefa $\pm a$). Krzywą przełączania tego przekaźnika na płaszczyźnie fazowej $\epsilon, d\epsilon/dt$, uzyskuje się, analizując obszary, w których wyjściem przekaźnika jest sygnał dodatni i ujemny (podrozdz. 2.2).

Krzywa przełączania daje nam pełną информацию o efektach zastosowania przekaźnika w układzie sterowania ze sprzężeniem zwrotnym. Ponieważ $\epsilon = -y$, a $d\epsilon/dt = -dy/dt$, to krzywą przełączania (rys. 6.24) można też przedstawić we współrzędnych „fazowych” $[y, dy/dt]$ z odwróceniem obszarów sterowań (rys. 6.25 i 6.26).

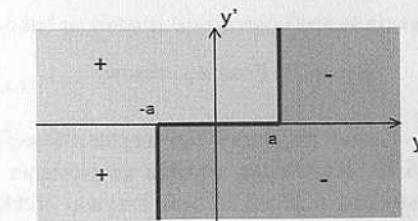


$$\begin{aligned}\epsilon < a, \epsilon' > 0, u &= -u_m \\ \epsilon > -a, \epsilon' < 0, u &= +u_m \\ \epsilon > a, \epsilon' > 0, u &= +u_m \\ \epsilon < -a, \epsilon' < 0, u &= -u_m\end{aligned}$$

Rys. 6.24. Krzywa przełączania przekaźnika

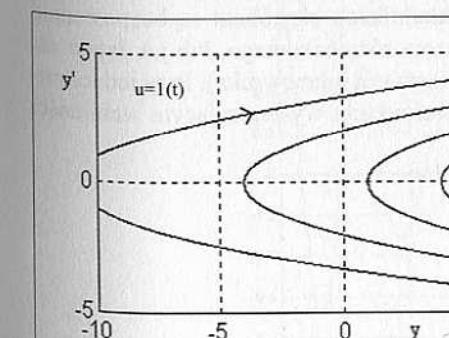


Rys. 6.25. Układ regulacji

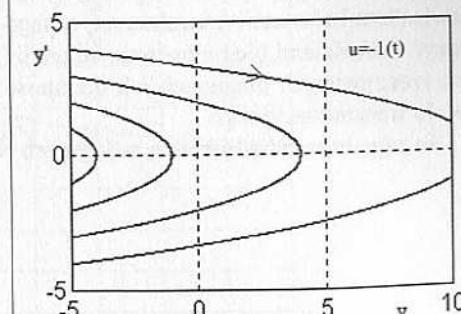


Rys. 6.26. Obszary sterowań

Dla założonego obiektu podwójnie całkującego, jego trajektorie fazowe tworzą dla sterowań dodatnich i ujemnych $u = \pm M \cdot 1(t)$ dwie rodziny krzywych parabolicznych (rys. 6.27 i 6.28). Każda trajektoria fazowa otrzymana jest dla innego warunku początkowego.

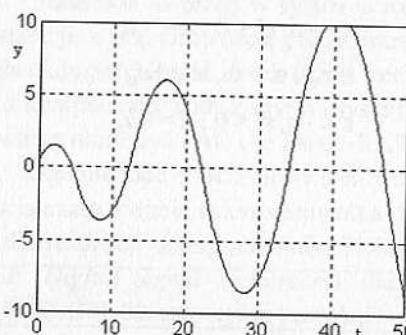


Rys. 6.27. Rodzina trajektorii dla sterowań $+M$

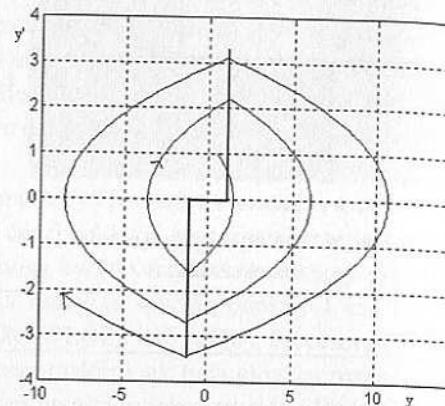


Rys. 6.28. Rodzina trajektorii dla sterowań $-M$

Przebiegi czasowe odpowiedzi $y(t)$ układu zamkniętego z takim przekaźnikiem (że strefą $a = 1$), startujące z warunków początkowych $[y(0) = 0,5, y'(0) = 1]$, są przedstawione na rysunku 6.29.



Rys. 6.29. Przebieg czasowy



Rys. 6.30. Przebieg na płaszczyźnie fazowej

Kształt trajektorii fazowej startującej z punktu $[y = 0,5, y' = 1]$ jest widoczny na rysunku 6.30. Kształt ten wynika ze złożenia dwóch rodzin trajektorii dla sterowań $u = \pm 1$ z rysunków 6.27 i 6.28 oraz krzywej przełączania z rysunku 6.24. Widać, że przebieg rozwiązań w układzie zamkniętym rozbiega się. Fakt niepożądanego narastania odpowiedzi układu wiąże się z ważnym problemem jakim jest stabilność układu sterowania. Stabilność jest własnością każdego obiektu dynamicznego, którą dla obiektów liniowych można odczytać z postaci ich transmitancji. Nie wnikając w szczegóły, można powiedzieć, że pojęcie to wiąże się między innymi z możliwością uzyskania ograniczonej odpowiedzi obiektu na ograniczone sterowanie (np. na skok jednostkowy). Można pokazać (np. podrozdz. 4.2, przykład 4.2, wzory (4.6), (4.7)), że nienarastanie w nieskończoność odpowiedzi skokowej układu związane jest z ujemnością współczynników w wykładnikach eksponent, które występują jako składowe tej odpowiedzi. Wartości eksponent w takim przypadku maleją do zera przy rosnącym czasie. Współczynniki w wykładnikach eksponent są bezpośrednio pierwiastkami równania charakterystycznego równania różniczkowego, lub jak łatwo zauważać wartościami bieguna transmitancji (pierwiastkami mianownika). Przy jednokrotnych rzeczywistych biegunach ich ujemność jest warunkiem wystarczającym stabilności modelu transmitancyjnego.

Te zagadnienia będą treścią następnych rozdziałów.

7. Stabilność układów dynamicznych

7.1. Stabilność układów liniowych ciągłych

7.1.1. Stabilność względem warunków początkowych

Równanie opisujące dynamikę stanu swobodnego układu liniowego (bez sterowania) ma postać:

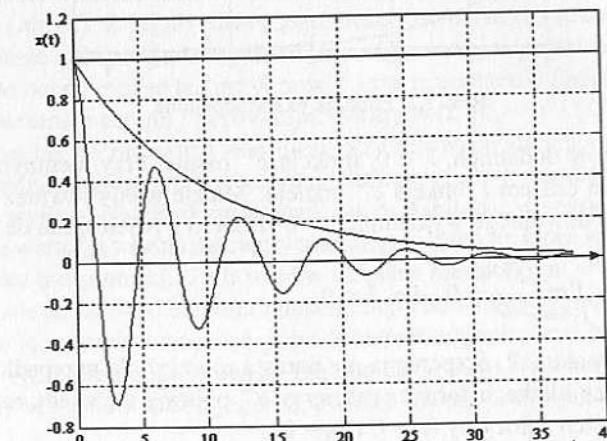
$$\dot{x}(t) = A \cdot x(t), \quad x(0) = x_0 \neq 0 \quad (7.1)$$

$$x(t) = [x_1(t), \dots, x_n(t)]^T \in R^n, \quad \forall t \geq 0$$

Rozwiązywanie tego równania wyraża się wzorem

$$x(t) = e^{At} x_0.$$

Def. 7.1. Mówimy, że układ (7.1) jest stabilny w sensie Lapunowa, jeśli dla dowolnych warunków początkowych x_0 , każda składowa jego rozwiązania $x(t)$ dąży do zera lub przynajmniej nie narasta dla $t > 0$ (rys. 7.1).



Rys. 7.1. Odpowiedź swobodna

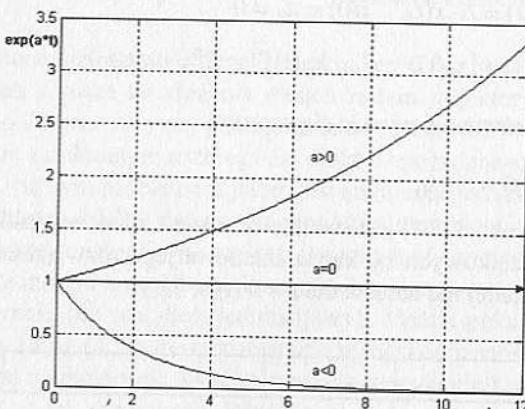
Aby łatwo sformułować odpowiedź na pytanie, co jest warunkiem koniecznym i dostatecznym stabilności układu swobodnego (7.1), założymy diagonalną postać Jordana dla macierzy A . Wnioski pozostaną słuszne dla dowolnej innej postaci podobnej do A .

Założymy, że w macierzy A wystąpią trzy wartości własne: jednokrotna λ_1 , jednokrotna λ_2 i czterokrotna λ_3 .

Poniżej zapisane są przykładowe postaci macierzy A i macierzy wykładniczej e^{At} :

$$A = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_3 & 1 & 0 \\ 0 & 0 & 0 & 0 & \lambda_3 & 1 \\ 0 & 0 & 0 & 0 & 0 & \lambda_3 \end{bmatrix}; \quad e^{At} = \begin{bmatrix} e^{\lambda_1 t} & 0 & 0 & 0 & 0 & 0 \\ 0 & e^{\lambda_2 t} & 0 & 0 & 0 & 0 \\ 0 & 0 & e^{\lambda_3 t} & 0 & 0 & 0 \\ 0 & 0 & 0 & e^{\lambda_3 t} & te^{\lambda_3 t} & 0,5t^2 e^{\lambda_3 t} \\ 0 & 0 & 0 & 0 & e^{\lambda_3 t} & te^{\lambda_3 t} \\ 0 & 0 & 0 & 0 & 0 & e^{\lambda_3 t} \end{bmatrix}$$

W rozwiążaniu równania (7.1) pojawią się, więc kombinacje liniowe powyższych elementów macierzy e^{At} . Przebieg eksponencjalnej funkcji czasu $e^{\lambda t}$ przedstawia rysunek 7.2.



Rys. 7.2. Funkcja eksponentalna

Dla wykładników dodatnich, $\lambda > 0$, funkcja $e^{\lambda t}$ rośnie. Przy ujemnym wykładniku λ wraz z wzrastającym czasem t funkcja $e^{\lambda t}$ maleje. Maleje wtedy również funkcja będąca iloczynem $t^n e^{\lambda t}$ dla dowolnego wykładnika n . Wynika to z twierdzenia de l'Hospitala

$$\lim_{t \rightarrow \infty} \frac{t^n}{e^{-\lambda t}} = 0 \quad \text{dla } \lambda < 0.$$

Jak widać z rysunku 7.2, eksponenta nie narasta również dla przypadku $\lambda = 0$. Jednak w wielowymiarowych klatkach Jordana macierzy e^{At} pojawią się wtedy rosnące z czasem funkcje. Widać to, jeżeli założymy $\lambda_2 = 0$ i $\lambda_3 = 0$.

Dla wielokrotnej wartości własnej równej zero $\lambda_3 = 0$, mogą wystąpić już wyrazy narastające i układ będzie niestabilny.

$$e^{At} = \begin{bmatrix} e^{\lambda_1 t} & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & t & t^2 / 2 \\ 0 & 0 & 0 & 0 & 1 & t \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Twierdzenie 7.1

Układ liniowy, stacjonarny (7.1) $\dot{x} = Ax$ jest **stabilny globalnie** (dla dowolnych warunków początkowych) wtedy i tylko wtedy, gdy wszystkie wartości własne macierzy A mają niedodatnie części rzeczywiste, a każda wartość własna o zerowej części rzeczywistej jest pierwiastkiem jednokrotnym wielomianu minimalnego macierzy A .

Twierdzenie 7.2

Układ liniowy, stacjonarny $\dot{x} = Ax$ jest **globalnie asymptotycznie stabilny** wtedy i tylko wtedy, gdy wszystkie wartości własne macierzy A mają ujemne części rzeczywiste.

7.1.2. Stabilność w sensie „ograniczone wejście-ograniczone wyjście” (BIBO)

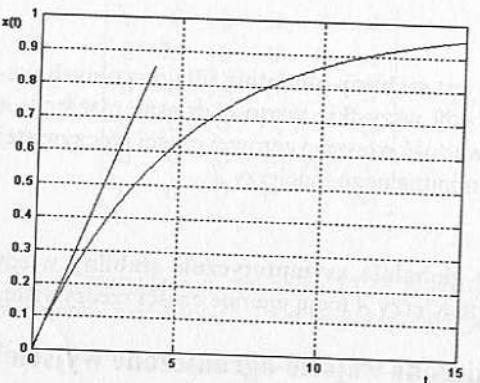
Poprzedni podrozdział poświęcono stabilności układu liniowego w sensie możliwości powrotu stanu x układu niesterowanego do rozwiązania zerowego, z dowolnych niezerowych warunków początkowych stanu układu (tzw. stabilność w sensie Lapunowa). W tych rozważaniach sygnał sterujący i tor sterowania nie odgrywały roli. W zastosowaniach technicznych, inżynierskich, często pojawia się jednak inny problem związany z pytaniem o warunki gwarantujące uzyskanie ograniczonej odpowiedzi wyjścia y układu liniowego na każdy ograniczony sygnał sterujący u . Tak zdefiniowana stabilność nosi nazwę „technicznej” lub stabilności w sensie „ograniczone wejście-ograniczone wyjście” BIBO (*Bounded Input-Bounded Output*). Z reguły słowo ograniczoność oznacza ograniczoność chwilowej amplitudy, co może być sprowadzone do normy wektorowej w przestrzeni R' i R'' dla funkcji $u(t)$ i $y(t)$. Do oceny można też przyjmować inne przestrzenie funkcyjne, np. analizując ograniczoność w sensie energii i porównując normy w $(L^2[0, \infty))^r$ i $(L^2[0, \infty))^m$.

Widać, że za tak zdefiniowaną stabilność przy zerowych warunkach początkowych odpowiedzialna będzie struktura toru sterowania, która może być przedstawiona transmitancją lub równaniem wyjścia. Należy przypuszczać, że za stabilność w sensie BIBO będą również odpowiedzialne wartości własne macierzy stanu A , ale tylko te, które wystąpią w uproszczonym mianowniku transmitancji. Jeśli więc w układzie niestabilnym w sensie Lapunowa nastąpi uproszczenie dodatniego bieguna transmitancji i dodatniego zera, to odpowiedź układu na ograniczone sterowanie może dać już odpowiedź ograniczoną. Na przykład poniższy układ II rzędu $G_1(s)$ niestabilny w sensie Lapunowa, w torze sterowania reprezentuje układ I rzędu, stabilny w sensie BIBO, a jego odpowiedź $h(t)$ ma postać jak na rysunku 7.3

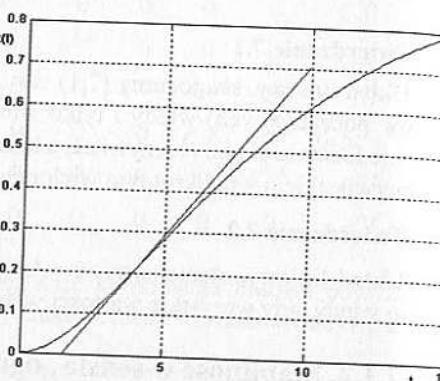
$$G_1(s) = \frac{s-2}{s^2 - s - 2} = \frac{s-2}{(s+1)(s-2)} = \frac{1}{s+1}.$$

Układ $G_2(s)$ jest stabilny zarówno w sensie Lapunowa, jak i BIBO, a jego odpowiedź $h(t)$ ma postać jak na rysunku 7.4

$$G_2(s) = \frac{2}{(s+1)(s+2)} = \frac{2}{s^2 + 3s + 2}.$$



Rys. 7.3. Odpowiedź układu I rzędu



Rys. 7.4. Odpowiedź układu II rzędu

Poniżej omówimy relacje, jakie występują pomiędzy stabilnością względem warunków początkowych (analiza wartości własnych macierzy A) a stabilnością względem sterowania (analiza biegunów transmitancji przy zerowych warunkach początkowych). Z równania stanu i wyjścia wynika postać jego rozwiązania na $y(t)$:

$$\begin{cases} \dot{x} = Ax + Bu \\ y = Cx \end{cases} \Rightarrow G(s) = \frac{Y(s)}{U(s)} = C[sI - A]^{-1}B$$

$$y(t) = \int_0^t Ce^{A(t-\tau)}Bu(\tau)d\tau = \int_0^t g(t-\tau)u(\tau)d\tau \quad (7.2)$$

Widać, że za uproszczenia w transmitancji i za znikanie narastających niestabilnych modów macierzy e^{At} w macierzy impulsowej funkcji przejścia $Ce^{At}B$ odpowiedzialne są dodatkowo macierze B i C .

Ponieważ o stabilności w sensie Lapunowa decyduje wyrażenie e^{At} , a o stabilności w sensie BIBO $g(t) = Ce^{At}B$, to stabilność względem warunków początkowych jest pełniej i pierwotniejszym pojęciem stabilności.

Dla układów liniowych można podać trzy poniższe stwierdzenia:

- 1) Każdy układ stabilny asymptotycznie w sensie Lapunowa jest stabilny w sensie BIBO.

7.1. Stabilność układów liniowych ciągłych

2) Układy stabilne nieasymptotycznie w sensie Lapunowa mogą być:

- niestabilne w sensie BIBO, np.
 - człon całkujący $G(s) = 1/s$, stabilny w sensie Lapunowa – daje odpowiedź narastającą na skok jednostkowy;
 - oscylator harmoniczny $G(s) = 1/(s^2 + 1)$, stabilny w sensie Lapunowa – wykazuje nieskończone wzmacnianie dla ograniczonego sygnału sinusoidalnego o częstotliwości rezonansowej;
- stabilne w sensie BIBO, np.
 - układ $G(s) = s/(s^2 + s)$.

3) Układ stabilny w sensie BIBO nie musi być stabilny w sensie Lapunowa.

Można wykazać [7DIS], że jeżeli wymuszenie $u(t)$ systemu SISO typu (7.2) jest funkcją przedziałami ciągłą o ograniczonej amplitudzie $|u(t)| \leq M$, to warunkiem koniecznym i wystarczającym stabilności typu BIBO jest, aby

$$\int_0^\infty |g(\tau)| d\tau \leq \infty.$$

Przy powyższym warunku na dowolnym odcinku czasu $[0, t]$ nawet najgorszy sygnał typu $u(t) = M \operatorname{sgn}[g(t - \tau)]$ spowoduje tylko ograniczoną reakcję systemu równą y_{\max}

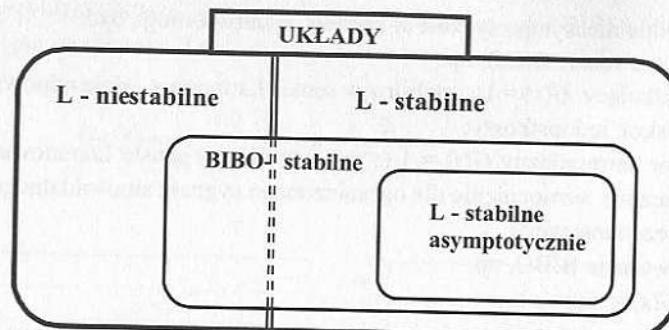
$$y_{\max}(t) = M \cdot \int_0^t |g(\tau)| d\tau \leq \infty.$$

Teoria stabilności BIBO łączy się ściśle ze wspomnianym w podrozdziale 3.6 twierdzeniem Riesza o reprezentacji każdego funkcjonału liniowego, ograniczonego, określonego na przestrzeni funkcji ciągłych ograniczonych $C_0[a, b]$, poprzez całkę Stieltjesa. Pojawia się tu również pojęcie unormowanej przestrzeni funkcji o wahaniu ograniczonym NBV[a, b] (sprzężonej do $C_0[a, b]$), gdzie wahanie spełnia aksjomaty normy. Przykładowo, wahanie (pełna wariacja) funkcji $h(t)$ będącej odpowiedzią systemu na skok jednostkowy jest dana wzorem

$$\operatorname{Var}_{(0, \infty)} h = \int_0^\infty |g(\tau)| d\tau.$$

Problematyka ta jest prezentowana w podręcznikach analizy funkcjonalnej i innych opracowaniach [7KOA, 7KUJ1, 7KUJ2]. Teorię tę wykorzystuje się przy projektowaniu regulatorów dla układów pracujących w warunkach najbardziej niebezpiecznych zakłóceń, generujących maksymalne możliwe błędy regulacji [7BYW]. Regulator ma za zadanie minimalizację błędu maksymalnego (podrozdz. 9.8).

Na rysunku 7.5 przedstawiono przenikanie się różnych typów stabilności.



Rys. 7.5. Zbiory układów stabilnych

7.1.3. Kryteria algebraiczne stabilności

Widać, że do oceny stabilności potrzebna jest analiza pierwiastków równania charakterystycznego lub mianownika transmitancji. Dzisiejsze narzędzia komputerowe pozwalają na szybkie obliczanie numeryczne pierwiastków wielomianu dowolnego rzędu. W początkach automatyki nie było to jednak łatwe zadanie dla wielomianów stopnia większego niż 4. Ponieważ do oceny stabilności potrzebna jest jedynie informacja, ile z tych pierwiastków leży po lewej i prawej stronie płaszczyzny zespolonej, wykorzystane zostały kryteria stabilności dla układów liniowych, które bez liczenia pierwiastków dają odpowiedź o ich lokalizacji. W ten sposób numeryczne zadanie o nieskończonej liczbie operacji (jaką jest dokładne znalezienie pierwiastka) zostało zastąpione zadaniem o skończonej liczbie iteracji, ale dającym tylko odpowiedź o jego lokalizacji. Ponieważ z kryteriów można uzyskać wzory analityczne dające warunki pożąданej lokalizacji pierwiastków, kryteria te można używać nie tylko do analizy, ale i syntezy stabilnych układów regulacji.

Do kryteriów tych zaliczamy:

- kryterium Hurwitzza,
- Meerowa,
- Michajłowa,
- Nyquista,
- Gerszgorina.

KRYTERIUM HURWITZZA

Dany jest wielomian zmiennej s (wielomian charakterystyczny macierzy A lub mianownik transmitancji)

$$a_n s^n + a_{n-1} s^{n-1} + \dots + a_1 s + a_0 = 0.$$

Twierdzenie 7.3

Warunkiem koniecznym stabilności (ujemności części rzeczywistych pierwiastków) jest, aby wszystkie współczynniki a_i były niezerowe i miały ten sam znak.

Twierdzenie 7.4 [7AJM]

Warunkiem koniecznym i wystarczającym stabilności asymptotycznej jest, aby w wyznaczniku Hurwitza wyznacznik główny i każdy jego podwyznacznik były dodatnie (dla $a_n > 0$).

$$\begin{vmatrix} a_1 & a_0 & 0 & \cdots & 0 & 0 \\ a_3 & a_2 & a_1 & a_0 & \cdots & 0 \\ a_5 & a_4 & a_3 & a_2 & \cdots & 0 \\ \vdots & \vdots & & & & \vdots \\ 0 & & & a_{n-2} & a_{n-3} \\ 0 & 0 & \cdots & 0 & a_n & a_{n-1} \end{vmatrix}_{(n-1) \times (n-1)}$$

Twierdzenie 7.5 (Lienarda–Chiparta)

Dla wielomianu rzeczywistego o dodatnich współczynnikach z dodatniością wyznaczników Hurwitza o wskaźnikach parzystych wynika dodatniość wyznaczników o wskaźnikach nieparzystych i na odwrót.

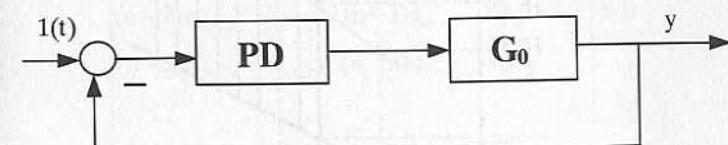
Wniosek

Wystarczy zbadać dodatniość co drugiego podwyznacznika Hurwitza, czyli przy n nieparzystym $(n-1)/2$ wyznaczników.

Przykład 7.1

Podać zakres zmienności parametrów regulatora PD, aby zapewnić stabilność układu i jednocześnie błąd w stanie ustalonym $\varepsilon_u \leq 10\% \cdot y(\infty)$, przy wymuszeniu skokowym i obiekcie o transmitancji (rys. 7.6):

$$G_0 = \frac{1}{s^3 + 8,5s^2 + s + 1}, \quad G_R = \alpha + \beta s.$$



Rys. 7.6. Schemat regulacji

Transmitancja układu zamkniętego

$$G_Z = \frac{G_R G_0}{1 + G_R G_0} = \frac{\alpha + \beta s}{s^3 + 8,5s^2 + (1+\beta)s + 1 + \alpha}.$$

Przy założeniu $\alpha > 0$, $\beta > 0$, dla układu zamkniętego otrzymuje się warunki:

$$\begin{vmatrix} 1+\beta & 1+\alpha \\ 1 & 8,5 \end{vmatrix} > 0$$

1) $1 + \beta > 0 \Rightarrow \beta > -1$,

2) $7,5 + 8,5\beta - \alpha > 0 \Rightarrow \beta > \frac{\alpha - 7,5}{8,5}$.

Analiza błędu prowadzi do wzorów:

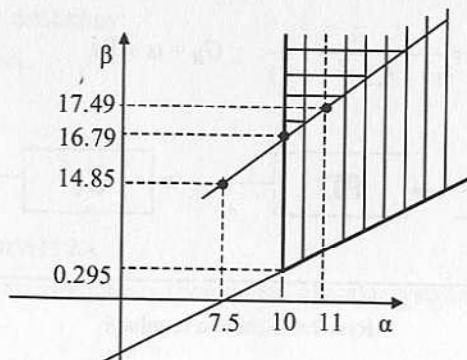
$$\varepsilon(t) = 1(t) - y(t) \Rightarrow \text{dla } t \rightarrow \infty, \varepsilon_u = 1 - y(\infty),$$

$$\lim_{t \rightarrow \infty} y(t) = \lim_{s \rightarrow 0} sY(s) = \frac{\alpha}{1 + \alpha},$$

$$\varepsilon_u = 1 - y(\infty) = 1 - \frac{\alpha}{1 + \alpha} = \frac{1}{1 + \alpha} \quad \Rightarrow \quad \text{jeśli } \alpha \text{ rośnie, to błąd } \varepsilon_u \text{ maleje,}$$

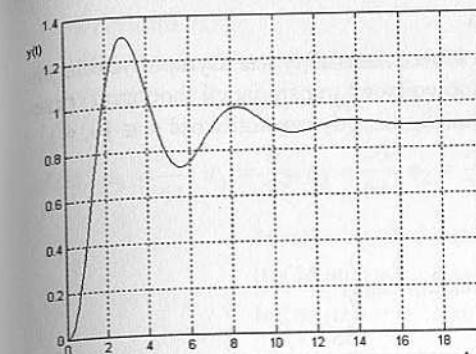
$$\varepsilon_u = \frac{1}{1 + \alpha} \leq 0,1 \cdot y(\infty) = 0,1 \frac{\alpha}{1 + \alpha} \quad \Rightarrow \quad \alpha \text{ powinno spełniać warunek } \alpha \geq 10.$$

Na rysunku 7.7 zaznaczono obszary stabilności i stabilności aperiodycznej (podwójne kreskowanie).

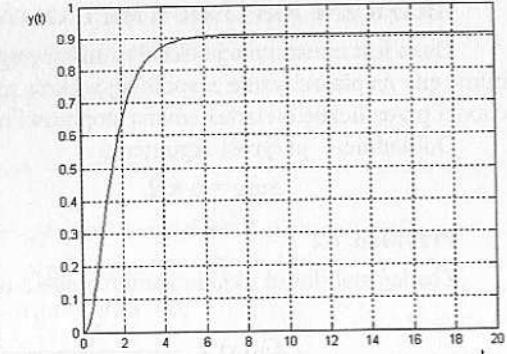


Rys. 7.7. Obszary stabilności i stabilności aperiodycznej dla $\alpha \geq 10$

Na rysunkach 7.8 i 7.9 widać zadany błąd statyczny i charakter przebiegów dla różnych β .



Rys. 7.8. Odpowiedź układu dla $\alpha = 10, \beta = 5$



Rys. 7.9. Odpowiedź układu dla $\alpha = 10, \beta = 17$

KRYTERIUM STABILNOŚCI APERIODYCZNEJ

Jeżeli pierwiastki wielomianu charakterystycznego lub pierwiastki mianownika transmitancji są rzeczywiste i ujemne, to przebiegi będą bez oscylacji harmonicznych. W zależności od postaci licznika, może wtedy pojawić się tylko skończona liczba przeregulowań (oscylacji nieharmonicznych). Jeżeli pierwiastki mianownika są liczbami zespolonymi o częściach rzeczywistych ujemnych, to wystąpią drgania harmoniczne (teoretycznie niekończące się).

Twierdzenie 7.6 (Meerowa)

Warunkiem koniecznym i wystarczającym, aby wszystkie pierwiastki wielomianu charakterystycznego były liczbami rzeczywistymi ujemnymi, jest, by wszystkie podwyznaczniki poniższego wyznacznika głównego $2n \times 2n$ były dodatnie.

$$\begin{vmatrix} na_n & a_n & 0 & 0 & 0 \\ (n-1)a_{n-1} & a_{n-1} & na_n & a_n & \\ (n-2)a_{n-2} & a_{n-2} & (n-1)a_{n-1} & a_n & \vdots \\ (n-3)a_{n-3} & a_{n-3} & (n-2)a_{n-2} & a_{n-2} & \\ & & \dots & & a_0 \end{vmatrix} > 0$$

Obszar stabilności aperiodycznej dla transmitancji z przykładu 7.1

$$G_Z = \frac{G_R G_0}{1 + G_R G_0} = \frac{\alpha + \beta s}{s^3 + 8,5s^2 + (1 + \beta)s + 1 + \alpha}$$

i dla $\alpha \geq 10$ zaznaczono dodatkowo na rysunku 7.7 (podwójne kreskowanie).

GRAFICZNE KRYTERIUM MICHAJŁOWA

Dana jest transmitancja układu zamkniętego. Układ zamknięty jest asymptotycznie stabilny, gdy na płaszczyźnie zespolonej wykres mianownika tej transmitancji (hodograf) przechodzi przez liczbę ćwiartek równą stopniowi mianownika, gdy częstotliwość $\omega \in [0, \infty)$.

Dokładniej – przyrost argumentu

$$\Delta \arg = n \cdot \pi / 2.$$

Przykład 7.2

Zbadać stabilność układu zamkniętego z obiektem $G_0(s)$

$$G_0(s) = \frac{K}{s(1+sT_1)(1+sT_2)} = \frac{58}{s(1+0,01s)(1+0,57s)}.$$

Po podstawieniu $s = j\omega$ mianownik ma postać liczby zespolonej:

$$\begin{aligned} M &= j\omega(1 + 0,01j\omega)(1 + 0,57j\omega) = \\ &= j(1 + j\omega 0,58 + 0,0057j^2\omega^2) = j\omega(1 - 0,0057\omega^2 + j\omega 0,58) = \\ &= -\omega^2 0,58 + j(\omega - 0,0057\omega^3), \end{aligned}$$

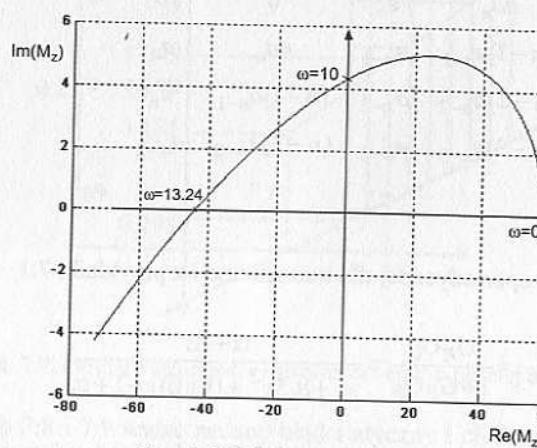
$$G_Z = \frac{G_0}{1+G_0} = \frac{L/M}{1+L/M} = \frac{L}{M+L},$$

$$M_Z = 58 - 0,58\omega^2 + j(\omega - 0,0057\omega^3),$$

$$\operatorname{Re}\{M_Z\} = 0 \Rightarrow \omega = 10,$$

$$\operatorname{Im}\{M_Z\} = 0 \Rightarrow \omega - 0,0057\omega^3 = 0 \Rightarrow \omega = 0, \text{ i } \omega = 13,24.$$

Otrzymuje się wykres, który przechodzi przez trzy ćwiartki, czyli układ zamknięty będzie stabilny (rys. 7.10).



Rys. 7.10. Wykres Michajłowa

Przykład 7.3

Mianownik układu zamkniętego jest dany w postaci

$$a_5 s^5 + a_4 s^4 + a_3 s^3 + a_2 s^2 + a_1 s + a_0 = 0,$$

$$\text{gdzie: } a_5 = \frac{1}{64}, a_4 = \frac{8}{72}, a_3 = \frac{20}{64}, a_2 = \frac{80}{72}, a_1 = 1, a_0 = 1,$$

$$M(j\omega) = a_5 j\omega^5 + a_4 \omega^4 + a_3 (-j)\omega^3 + a_2 (-1)\omega^2 + a_1 j\omega + a_0 = 0,$$

$$\operatorname{Re}(M(j\omega)) = 0: a_4 \omega^4 - a_2 \omega^2 + a_0 = 0,$$

$$\operatorname{Im}(M(j\omega)) = 0: a_5 \omega^5 - a_3 \omega^3 + a_1 \omega = 0,$$

$$\frac{8}{72} \bar{\omega}^2 - \frac{80}{72} \bar{\omega} + 1 = 0 \Rightarrow 8\bar{\omega}^2 - 80\bar{\omega} + 72 = 0 \Rightarrow \bar{\omega}^2 - 10\bar{\omega} + 9 = 0,$$

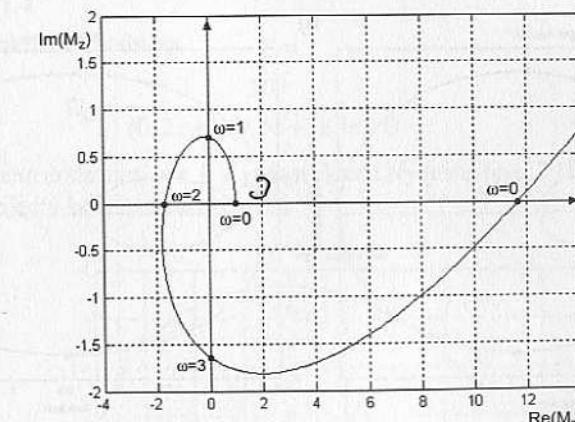
$$\bar{\omega}_1 = 1 \Rightarrow \omega_1 = 1, \quad \bar{\omega}_2 = 9 \Rightarrow \omega_2 = 3,$$

$$\frac{1}{64} \omega^5 - \frac{20}{64} \omega^3 + 1\omega = 0, \quad \omega_3 = 0,$$

$$\bar{\omega}^2 - 20\bar{\omega} + 64 = 0, \quad \bar{\omega}_4 = 4 \Rightarrow \omega_4 = 2, \quad \bar{\omega}_5 = 16 \Rightarrow \omega_5 = 4.$$

Z przyrównania części rzeczywistej do zera otrzymuje się częstotliwości $\omega_1 = 1, \omega_2 = 3$.

Z przyrównania części urojonej do zera otrzymuje się częstotliwości $\omega_3 = 0, \omega_4 = 2, \omega_5 = 4$.

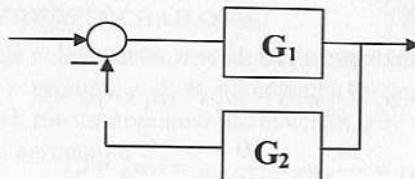


Rys. 7.11. Wykres Michajłowa

Kolejne częstotliwości zerujące część rzeczywistą i urojoną mianownika występują na przemian, co świadczy o kolejnym przecinaniu osi urojonej i rzeczywistej przez wykres (rys. 7.11), czyli o przyroście argumentu $\Delta \arg = 5 \cdot \pi / 2$. Gwarantuje to stabilność układu zamkniętego.

KRYTERIUM NYQUISTA

Kryterium Nyquista (1932 r.) pozwala ocenić stabilność układu zamkniętego na podstawie amplitudowo-fazowego wykresu transmitancji układu otwartego $G_0 = G_1 G_2$.



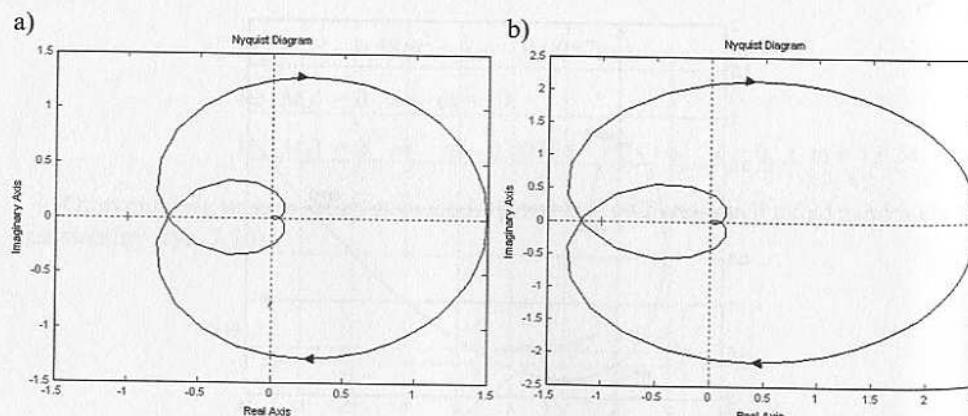
Rys. 7.12. Schemat blokowy

Nie trzeba specjalnie analizować mianownika transmitancji zastępczej układu zamkniętego (rys. 7.12)

$$G_Z = \frac{G_1}{1 + G_1 G_2}.$$

Twierdzenie 7.7

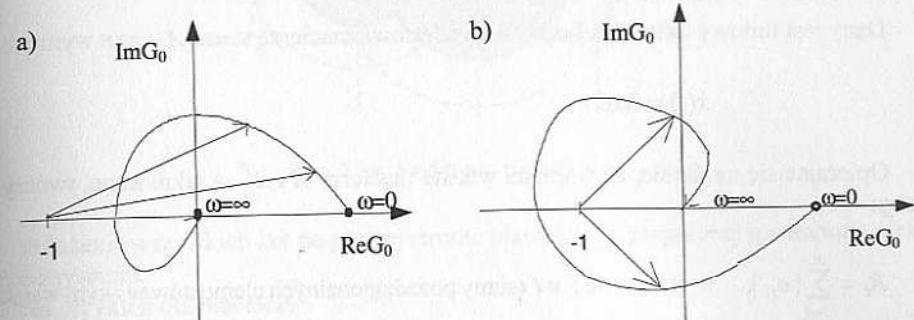
- 1) Jeżeli układ otwarty jest stabilny i przyrost argumentu ($\Delta\arg$) transmitancji otwartej $G_0(j\omega)$ liczony na płaszczyźnie zespolonej od $\omega = 0$ do $\omega = \infty$ widziany z punktu $(-1, 0)$ jest równy zero, to układ zamknięty będzie też stabilny (wykres nie obejmuje punktu $(-1, 0)$). Jeżeli $\Delta\arg = 0$ \Rightarrow układ po zamknięciu będzie dalej stabilny (rys. 7.13a), jeśli np. $\Delta\arg = 2\pi$ \Rightarrow układ po zamknięciu będzie niestabilny (rys. 7.13b).

Rys. 7.13. Wykresy Nyquista dla a) $G(s) = 1,5/(s + 1)^7$; b) $G(s) = 2,5/(s + 1)^7$
Objaśnienia w tekście

- 2) Jeżeli układ otwarty jest niestabilny (tzn. mianownik $G_0(j\omega)$ ma m pierwiastków z prawej strony płaszczyzny), to układ zamknięty jednostkowym sprzężeniem zwrotnym jest stabilny, jeżeli przyrost argumentu $\Delta\arg = m\pi$ lub $\Delta\arg = (m/2)\cdot 2\pi$, przy zmianie ω od 0 do ∞ .

Na przykład układ G_0 jest niestabilny i $m = 2$, jeżeli $\Delta\arg = 0$, to po zamknięciu układ będzie dalej niestabilny (rys. 7.14a). Jeżeli $\Delta\arg = 2\pi$, to układ po zamknięciu będzie stabilny (rys. 7.15).

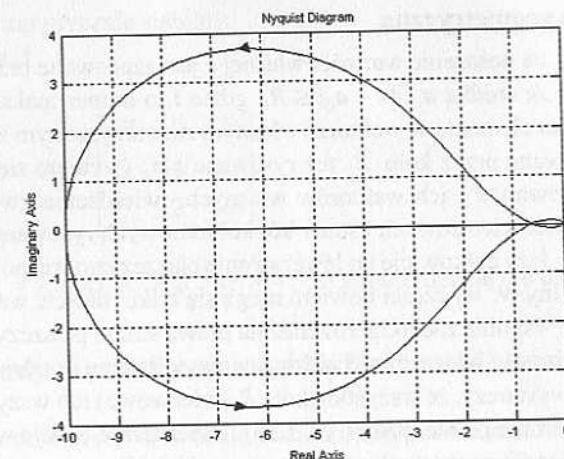
- 3) Jeżeli w transmitancji występują bieguna zerowe i wykres amplitudowo-fazowy dla $\omega = 0$ ma punkt nieciągłości, tzn. $\text{Im}G_0(0) = \pm\infty$ lub $\text{Re}G_0(0) = \pm\infty$, należy hipotetycznie uzupełnić ten wykres łukiem o promieniu nieskończonym, sięgającym punktu $\text{Re}G = \infty$ i traktować ten punkt jako formalny początek wykresu dla $\text{Re}G_0(0)$. Następnie należy zastosować reguły punktów 1) lub 2). Jest to związane z obchodzeniem bieguna zerowego, przy poruszaniu się wzdłuż osi urojonej, po nieskończenie małym półokręgu, po jego prawej stronie.

Rys. 7.14. Wykres Nyquista
Objaśnienia w tekście**Przykład 7.4**

Dany jest układ niestabilny

$$G_0 = \frac{10}{(0,2s+1)(0,5s+1)(3s-1)}$$

z jednym biegunem dodatnim, $m = 1$. Z jego wykresu Nyquista (rys. 7.15) widać, że $\Delta\arg = \pi \Rightarrow$ układ po zamknięciu będzie stabilny.



Rys. 7.15. Wykres Nyquista

Fakt ten można sprawdzić, licząc pierwiastki mianownika transmitancji zamkniętej

$$(0,2s + 1)(0,5s + 1)(3s - 1) + 10 = 0,3s^3 + 2s^2 + 2,3s + 9 = 0.$$

Wynoszą one: $s_1 = -6,21$, $s_{2,3} = -0,228 \pm j2,186$, czyli układ po zamknięciu jest stabilny.

METODA GERSZGORINA DO SZYBKIEJ OCENY STABILNOŚCI LAPUNOWA UKŁADÓW LINIOWYCH

Dany jest liniowy układ swobodny z kwadratową macierzą stanu $A - n \times n$ wymiarową

$$\dot{x}(t) = Ax(t).$$

Opierając się na fakcie, że wartości własne macierzy A i A^T są takie same, tworzy się liczby:

$$R_i = \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|, \quad i = 1, 2, \dots, n; j \neq i \text{ (sumy pozadiagonalnych elementów w } i\text{-tym wierszu);}$$

$$\bar{R}_j = \sum_{\substack{i=1 \\ i \neq j}}^n |a_{ij}|, \quad j = 1, 2, \dots, n; i \neq j \text{ (sumy pozadiagonalnych elementów w } j\text{-tej kolumnie).}$$

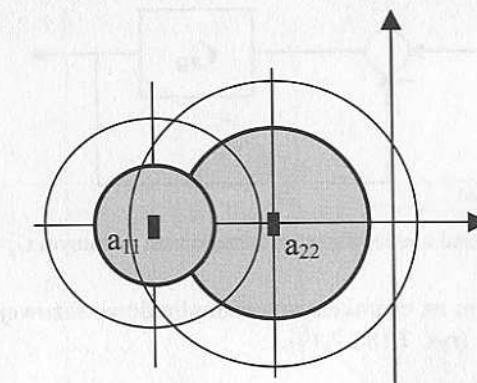
Twierdzenie 7.8

Jeżeli $a_{ii} < 0$ oraz $R_i < |a_{ii}|$ dla każdego i lub $\bar{R}_j < |a_{jj}|$ dla każdego j , to układ jest asymptotycznie stabilny.

Interpretacja geometryczna

Można pokazać, że położenie wartości własne s jest szacowane przez koło i -te o promieniu R_i , rysowane ze środka a_{ii} , $|s - a_{ii}| \leq R_i$, gdzie i to numer maksymalnego – co do bezwzględnej wartości elementu w wektorze własnym stwarzyszonym z wartością s , i jednocześnie jest szacowane przez koło \bar{R}_i też rysowane z a_{ii} (wybiera się więc koło mniejsze). Ponieważ nie znamy s i ich wektorów własnych, twierdzenie gwarantuje, że jeżeli wspólna część sumy kół wierszowych i sumy kół kolumnowych rysowanych ze środków a_{ii} o promieniach R_i i \bar{R}_i leży całkowicie po lewej stronie płaszczyzny zespolonej, to układ jest asymptotycznie stabilny. W tej części bowiem mogą się tylko mieścić wartości własne macierzy A . Jeżeli część wspólna zachodzi również na prawą stronę płaszczyzny, to nie ma już gwarancji, że jakąś wartość własną nie znalazła się też po stronie prawej.

Dla stabilności wystarczy, że wszystkie koła R_i (wierszowe) lub wszystkie \bar{R}_j (kolumnowe) leżą po lewej stronie płaszczyzny (rys. 7.16). Twierdzenie to daje więc tylko warunki wystarczające stabilności asymptotycznej, co oznacza, że jeżeli te warunki nie są spełnione, to układ może być stabilny lub niestabilny.



Rys. 7.16. Kryterium Gerszgorina

Położenie wszystkich kół po prawej stronie płaszczyzny zespolonej gwarantuje niestabilność systemu.

Na przykład dla macierzy

$$A_1 = \begin{bmatrix} -7 & -3 & 5 \\ 2 & -5 & -2 \\ 4 & 1 & -8 \end{bmatrix},$$

$$R_1 = 8 > |a_{11}| = 7, \text{ ale } \bar{R}_1 = 6 < |a_{11}| = 7,$$

$$\bar{R}_2 = 4 < |a_{22}| = 5, \quad \bar{R}_3 = 7 < |a_{33}| = 8,$$

stąd układ jest asymptotycznie stabilny.

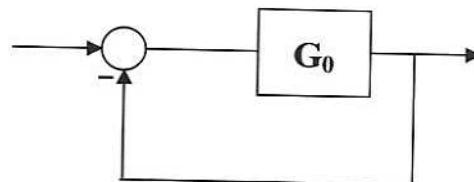
Dla macierzy A

$$A_2 = \begin{bmatrix} -7 & -3 & 5 \\ 3 & -5 & 1 \\ 5 & 2 & -8 \end{bmatrix},$$

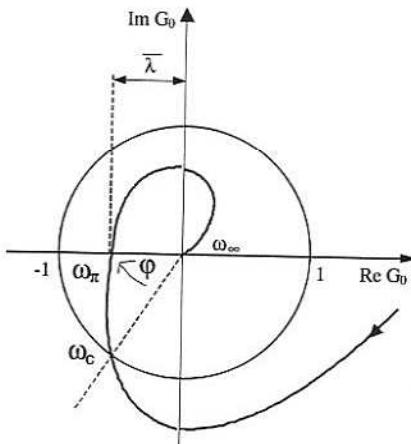
$$R_1 = 8 > |a_{11}| = 7, \text{ oraz } \bar{R}_1 = 8 > |a_{11}| = 7, \text{ stąd twierdzenie nie daje wyniku.}$$

7.1.4. Metody częstotliwościowe badania stabilności

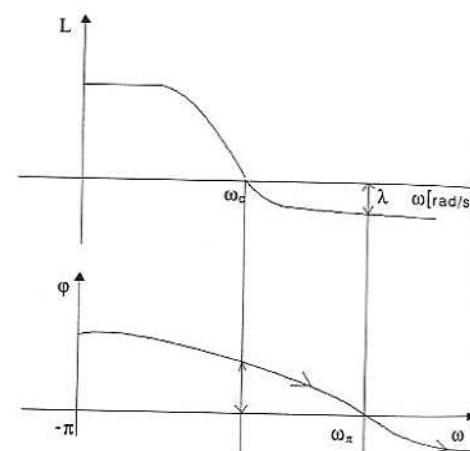
Metody częstotliwościowe badania stabilności układu zamkniętego z jednostkowym sprzężeniem zwrotnym (rys. 7.17) są oparte na analizie częstotliwościowych charakterystyk logarytmicznych (Bodego) układu otwartego.

Rys. 7.17. Układ z jednostkowym sprzężeniem zwrotnym $G_z = G_0/(1+G_0)$

Punktem krytycznym na charakterystyce amplitudowo-fazowej (Nyquista) jest punkt $(-1, 0)$, tzn. $\text{Re}[G_0] = -1$ (rys. 7.18 i 7.19).



Rys. 7.18. Płaszczyzna Nyquista



Rys. 7.19. Charakterystyki częstotliwościowe

Oznaczenia występujące na rysunkach 7.18 i 7.19:

kąt $\bar{\phi} = 180 - \phi(\omega_c)$ – zapas fazowy,

$\bar{\lambda}$ – zapas modułu, $\bar{\lambda} = 20 \log |G_0(j\omega_\pi)|$,

$L = 20 \log |G_0|$,

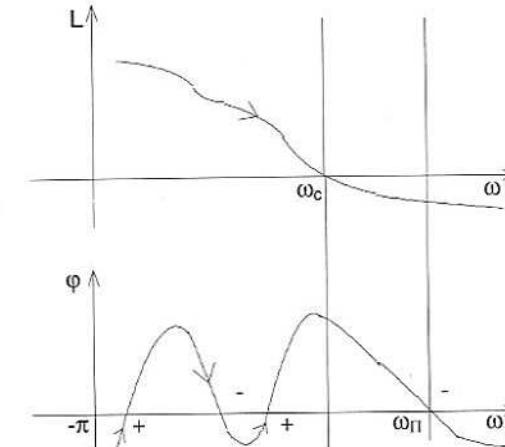
ω_c – częstotliwość odcięcia (gdy $|G_0(j\omega_c)| = 1$),

ω_π – częstotliwość graniczna fazy.

Warunki stabilności układu zamkniętego

Podamy dwa warunki stabilności dla układu zamkniętego [7AJM]:

- 1) aby dla układu otwartego stabilnego układ zamknięty jednostkowym sprzężeniem zwrotnym był stabilny, to częstotliwość $\omega_c < \omega_\pi$;
- 2) liczba przejść charakterystyki fazowej przez linię $-\pi$ musi być parzysta w zakresie $0 \div \omega_c$ (rys. 7.20).



Rys. 7.20. Charakterystyki częstotliwościowe

Przy układzie otwartym G_0 o transmitancji niestabilnej (m biegumów w prawej pół-płaszczyźnie), warunkiem stabilności układu zamkniętego jest, aby liczba przejść dodatnich charakterystyki fazowej przez linię $-\pi$ przewyższała o $m/2$ liczbę przejść ujemnych w zakresie $0 \div \omega_c$, np. dla $m = 2$ powinny być dwa przejścia dodatnie i jedno ujemne.

7.1.5. Stan ustalony odpowiedzi niestabilnego układu liniowego

Charakterystyki stanu ustalonego obiektów niestabilnych

Często odpowiedź układu w dziedzinie czasu dla $t \rightarrow \infty$ ma wartość ustaloną. Wtedy jego wartość określa się z twierdzenia o wartości granicznej

$$\lim_{t \rightarrow \infty} y(t) = y(\infty) = \lim_{s \rightarrow 0} [sY(s)] = \lim_{s \rightarrow 0} [sG(s)U(s)].$$

Dla wymuszenia skokowego: $u(t) = 1(t) \Rightarrow U(s) = 1/s$ wartość ustalona odpowiedzi określana jest stosunkiem wyrazów wolnych licznika i mianownika transmitancji

$$y(\infty) = \lim_{s \rightarrow 0} [G(s)] = \lim_{s \rightarrow 0} \left[\frac{b_m s^m + \dots + b_1 s + b_0}{a_n s^n + \dots + a_1 s + a_0} \right] = \frac{b_0}{a_0}.$$

Twierdzenie o wartości granicznej obowiązuje również dla układów z dodatnimi biegumami. Intuicyjnie wydaje się, że nie jest możliwe eksperymentalne uzyskanie w układzie niestabilnym stanu ustalonego, w którym dla $t = \infty$ przy ustalonym sterowaniu, odpowiedź przyjmie wartość ustaloną. W literaturze przedmiotu nie mówi się więc o identyfikacji charakterystyk statycznych układów niestabilnych. Wynikałoby to też z przykładu 4.7, podroz-

dział 4.3.5, w którym pokazano, że dla niestabilnej transmitancji I rzędu, odpowiedź na skok jednostkowy $u_1(t) = 1(t)$ narasta w nieskończoność.

$$G(s) = \frac{1}{Ts-1} \Rightarrow Y(s) = \frac{1}{s(Ts-1)} \Rightarrow y(t) = e^{t/T} - 1.$$

Jednak podanie na wejście tego obiektu innego sygnału, np. eksponencjalnie zanikającego

$$u(t) = [(T+1)e^{-t} - 1] \cdot 1(t),$$

w chwili początkowej $t = 0$ przyjmującego wartość $u(0) = T$, a w $t = \infty$ wartość ustaloną $u(\infty) = -1$, i którego transformata Laplace'a ma postać

$$U(s) = \frac{T+1}{s+1} - \frac{1}{s} = \frac{Ts-1}{s(s+1)},$$

spowoduje, że transformata odpowiedzi układu ma postać:

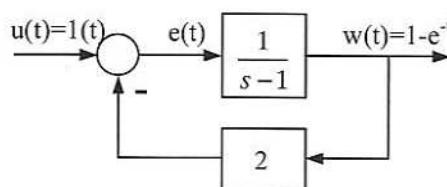
$$Y(s) = \frac{1}{Ts-1} \cdot \frac{Ts-1}{s(s+1)} = \frac{1}{s(s+1)}, \text{ a stąd } \lim_{t \rightarrow \infty} [y(t)] = \lim_{s \rightarrow 0} [s Y(s)] = 1.$$

To implikuje dla układu odpowiedź o postaci $y(t) = 1 - e^{-t}$. W stanie ustalonym sterowania, wyjście tego układu niestabilnego ma również wartość ustaloną, $y(\infty) = 1$, która wynika ze stosunku wyrazów wolnych tej transmitancji.

Poniżej zostanie pokazane, jak użycie sprzężenia zwrotnego pozwala zrealizować praktycznie powyższy układ, w którym na wejściu i wyjściu układu niestabilnego wystąpi stan ustalony, a do eksperymentu zastosowany będzie skok jednostkowy.

Przykład 7.5

Rozważmy obiekt niestabilny I rzędu objęty ujemnym sprzężeniem zwrotnym, z rysunku 7.21.



Rys. 7.21. Układ zamknięty

Transmitancja układu zamkniętego przedstawia już obiekt stabilny

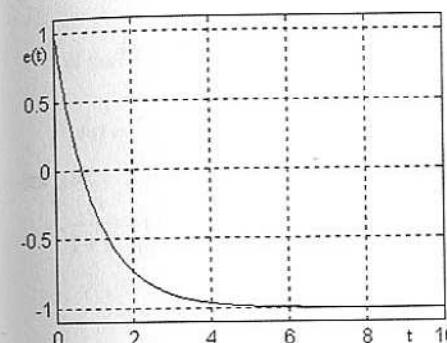
$$G_z(s) = \frac{1/(s-1)}{1 + 2/(s-1)} = \frac{1}{s+1}.$$

Stąd wyjście układu dla $u(t) = 1(t)$ (podrozdz. 4.3.5) dane jest funkcją: $y(t) = 1 - e^{-t}$.

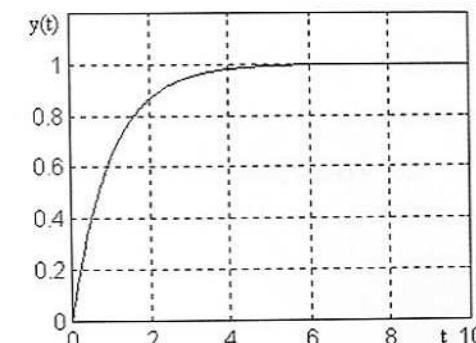
$$\text{Transformata błędu ma wartość } E(s) = \frac{1}{s} - \frac{2}{s(s+1)} = \frac{s-1}{s(s+1)}.$$

Stąd sygnał błędu $e(t)$ ma postać $e(t) = 1(t) - 2[1 - e^{-t}] = 2e^{-t} - 1(t)$.

Kształty sygnałów $e(t)$ i $y(t)$ widać na rysunkach 7.22 i 7.23.



Rys. 7.22. Przebieg $e(t)$



Rys. 7.23. Przebieg $y(t)$

Odpowiedź układu dana jest wzorem

$$Y(s) = G(s)E(s) = \frac{1}{s-1} \left[\frac{2}{s+1} - \frac{1}{s} \right] = \frac{s-1}{s(s-1)(s+1)} = \frac{1}{s(s+1)}.$$

Redukcja niestabilnego bieguna w $Y(s)$ wynika z powyższych transformat $G(s)$ i $E(s)$.

W stanie ustalonym na wejściu układu niestabilnego będzie wartość $e(\infty) = -1$, a na jego wyjściu $y(\infty) = 1$. Układ ze sprzężeniem zwrotnym może więc ustabilizować obiekty niestabilne i umożliwić tym samym identyfikację ich stanu ustalonego, a stąd charakterystyki statycznej.

Charakterystyki częstotliwościowe obiektów niestabilnych

W podrozdziale 4.4.2 wspomniano, że charakterystyki częstotliwościowe Nyquista lub Bodego można otrzymać eksperymentalnie dla układów, dla których można zarejestrować stan drgań ustalonych przy wymuszeniu sinusoidalnym. Formalnie dotyczy to również układów niestabilnych takich jak np. pokazane w przykładzie 4.7.

Dalej zostanie wykazane, że taki układ objęty sprzężeniem zwrotnym umożliwia również zdejmowanie eksperymentalne charakterystyk częstotliwościowych układów niestabilnych.

Przykład 7.6

Podobnie jak w podrozdziale 4.4.1 można wyprowadzić wzór na narastającą odpowiedź $y(t)$ dla obiektu niestabilnego $G(s) = \frac{1}{s-1}$ przy wymuszeniu sinusoidalnym $u(t) = \sin(\omega t)$,

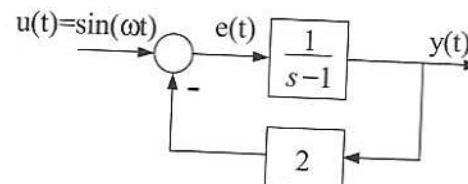
$$y(t) = \frac{\omega}{\omega^2 + 1} e^t - \frac{1}{\omega^2 + 1} \sin(\omega t) - \frac{\omega}{\omega^2 + 1} \cos(\omega t).$$

To samo wymuszenie sinusoidalne $u(t) = \sin(\omega t)$ dla obiektu stabilnego $G(s) = \frac{1}{s+1}$ daje odpowiedź oscylacyjną z zanikającym przebiegiem przejściowym $y(t)$ o postaci (4.10):

$$y(t) = \frac{\omega}{\omega^2 + 1} e^{-t} + \frac{1}{\omega^2 + 1} \sin(\omega t) - \frac{\omega}{\omega^2 + 1} \cos(\omega t) = \frac{\omega}{\omega^2 + 1} e^{-t} + \sin(\omega t - \varphi),$$

gdzie $\operatorname{tg}(\varphi) = \omega$.

Taki właśnie przebieg musi pojawić się na wyjściu układu zamkniętego (rys. 7.24), który jest inercyjnym układem stabilnym I rzędu o zastępczej transmitancji $G_z(s) = \frac{1}{s+1}$.



Rys. 7.24. Układ zamknięty

Przebieg błędu $e(t)$, który jest sygnałem wejściowym na obiekt niestabilny, ma postać:

$$\begin{aligned} e(t) &= \sin(\omega t) - \frac{2\omega}{\omega^2 + 1} e^{-t} - \frac{2}{\omega^2 + 1} \sin(\omega t) + \frac{2\omega}{\omega^2 + 1} \cos(\omega t) = \\ &= \sin(\omega t + \psi) - \frac{2\omega}{\omega^2 + 1} e^{-t} = e_1(t) + e_2(t), \end{aligned}$$

$$\text{gdzie } \operatorname{tg}(\psi) = \frac{2\omega}{\omega^2 - 1}.$$

W sygnale $e(t)$ występują dwie składowe: sinusoidalna $e_1(t)$ i eksponencjalna $e_2(t)$. Odpowiedź $y_1(t)$ obiektu niestabilnego na składową sinusoidalną $e_1(t)$ ma postać:

$$y_1(t) = \frac{\omega}{\omega^2 + 1} e^t - \frac{1}{\omega^2 + 1} \sin(\omega t + \psi) - \frac{\omega}{\omega^2 + 1} \cos(\omega t + \psi).$$

Odpowiedź $y_2(t)$ obiektu niestabilnego na składową $e_2(t)$, można otrzymać z przekształcenia Laplace'a:

$$\mathcal{L}[e_2(t)] = \mathcal{L}\left[-\frac{2\omega}{\omega^2 + 1} e^{-t}\right] = -\frac{2\omega}{\omega^2 + 1} \cdot \frac{1}{s+1} \Rightarrow$$

$$Y_2(s) = -\frac{2\omega}{\omega^2 + 1} \cdot \frac{1}{s+1} \cdot \frac{1}{s-1} = -\frac{2\omega}{\omega^2 + 1} \cdot \frac{1}{s^2 - 1},$$

$$y_2(t) = \mathcal{L}^{-1}[Y_2(s)] = -\frac{2\omega}{\omega^2 + 1} \cdot \operatorname{sh}(t) = \frac{\omega}{\omega^2 + 1} \cdot (e^{-t} - e^t).$$

W sumarycznym wyjściu $y(t) = y_1(t) + y_2(t)$ redukcji ulegną więc niestabilne eksponenty e^t

$$\begin{aligned} y(t) &= \frac{\omega}{\omega^2 + 1} e^t - \frac{1}{\omega^2 + 1} \sin(\omega t + \psi) - \frac{\omega}{\omega^2 + 1} \cos(\omega t + \psi) + \frac{\omega}{\omega^2 + 1} \cdot (e^{-t} - e^t) = \\ &= \frac{\omega}{\omega^2 + 1} e^{-t} - \frac{1}{\omega^2 + 1} \sin(\omega t + \psi) - \frac{\omega}{\omega^2 + 1} \cos(\omega t + \psi) = \\ &= \frac{\omega}{\omega^2 + 1} e^{-t} + \sin(\omega t + \psi + \varphi). \end{aligned}$$

Otrzymany powyżej wzór zgadza się ze wzorem (4.10), ponieważ $\psi = -2\varphi$, a $\operatorname{tg}(\varphi) = \omega$. Przyjmując praktyczną definicję stanu ustalonego, jako stanu, w którym składowe eksponencjalne zanikną, można powiedzieć, że na wejściu i wyjściu obiektu niestabilnego pozostały tylko sygnały sinusoidalne.

Powysze wyprowadzenia przedstawiono, aby uzasadnić fakt, że dla obiektów niestabilnych charakterystykę Nyquista można zawsze teoretycznie wyliczyć z postaci transmitancji, ale można ją również zdjąć eksperymentalnie, wykorzystując własność sprzężenia zwrotnego.

7.2. Stabilność układów liniowych dyskretnych

Liniowy swobodny układ dyskretny:

$$x_{k+1} = A_D x_k, \quad x_0 \neq 0, \quad x_k \in R^n, \quad \forall k \geq 0,$$

$$\begin{bmatrix} x_1(k+1) \\ \vdots \\ x_n(k+1) \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1(k) \\ \vdots \\ x_n(k) \end{bmatrix}$$

nazywamy asymptotycznie stabilnym (w sensie Lapunowa), gdy dyskretne wartości odpowiedzi swobodnej maleją do zera dla $k \rightarrow \infty$.

Twierdzenie 7.9 (O stabilności)

Warunkiem koniecznym i wystarczającym asymptotycznej stabilności układu liniowego, stacjonarnego, dyskretnego jest, aby wszystkie pierwiastki równania charakterystycznego macierzy stanu A_D leżały wewnątrz koła jednostkowego, tzn. $|z_i| < 1$ dla $i = 1, \dots, n$.

System będzie stabilny, jeśli na okręgu jednostkowym, będą leżały tylko jednokrotne pierwiastki wielomianu minimalnego. Pierwiastki te można wyliczyć z wyznacznika macierzy charakterystycznej $|zI - A_D| = 0$.

Dla udowodnienia twierdzenia stosuje się liniowe przekształcenie stanu srowadzące macierz A_D do jej postaci diagonalnej Jordana poprzez nieosobliwą macierz kwadratową T .

$$x_k = T\hat{x}_k, \quad T\hat{x}_{k+1} = A_D T\hat{x}_k, \quad \hat{x}_{k+1} = T^{-1} A_D T\hat{x}_k,$$

$$\hat{x}_{k+1} = J_D \hat{x}_k = \begin{bmatrix} z_1 & 0 & \cdots & 0 \\ 0 & z_2 & 0 & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & z_n \end{bmatrix} \hat{x}_k.$$

Rozwiązanie powyższego równania ma postać:

$$\hat{x}_k = J_D^k \hat{x}_0 = \begin{bmatrix} z_1^k & 0 & \cdots & 0 \\ 0 & z_2^k & 0 & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & z_n^k \end{bmatrix} \hat{x}_0,$$

$$x_k = A_D^k x_0 = T J_D^k T^{-1} x_0.$$

Widać, że rozwiązanie dla wszystkich składowych stanu i dla dowolnych warunków początkowych będzie dążyło do zera tylko wtedy, gdy wartości własne z_i będą ułamkami (ogólnie ich moduł będzie mniejszy od jedena). Dla przypadku granicznego $|z_i| = 1$ i pojętych wartości własnych układ będzie pozostawał na granicy stabilności.

Tak jak i w przypadku układu ciągłego, dla określenia stabilności nie trzeba liczyć wartości własnych. Można zastosować różne kryteria oceniające rozkład pierwiastków na płaszczyźnie zespolonej; poniżej omówiono pięć takich kryteriów.

1) Różne warunki wystarczające**Twierdzenie 7.10a (O niestabilności)**

Niestabilność systemu dyskretnego przy zadanej macierzy stanu A_D można szybko sprawdzić, wykorzystując własność śladu macierzy

$$\text{ślad}(A_D) = \sum_{i=1}^n a_{ii} = \sum_{i=1}^n z_i.$$

Warunkiem wystarczającym niestabilności układu dyskretnego jest, aby wartość modyfikacji ze śladu macierzy A_D była większa od wymiaru macierzy n : $|\text{ślad}(A_D)| > n$. Wtedy przynajmniej jeden z pierwiastków musi mieć moduł większy od 1.

Twierdzenie 7.10b (O niestabilności)

Dany jest wielomian charakterystyczny macierzy A_D

$$w(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1z + a_0.$$

Warunkiem wystarczającym niestabilności układu dyskretnego jest, aby $|a_{n-1}| > n$ lub $|a_0| > 1$. Dowód wynika ze wzorów Viete'a, gdyż wiadomo, że współczynniki:

$$a_{n-1} = -\sum_{i=1}^n z_i, \quad a_0 = \prod_{i=1}^n z_i.$$

Jeżeli $|a_{n-1}| > n$ lub $|a_0| > 1$, gdzie $z_i, i = 1, \dots, n$, są pierwiastkami $w(z)$, to przynajmniej jeden z pierwiastków musi mieć moduł większy od 1, $|z_i| > 1$.

Twierdzenie 7.11 (O stabilności)

Dany jest wielomian charakterystyczny macierzy A_D

$$w(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1z + a_0.$$

Jeśli kolejne współczynniki a_i wielomianu $w(z)$ maleją, czyli spełniają warunek

$$1 > a_{n-1} > \cdots > a_1 > a_0,$$

to wszystkie pierwiastki leżą wewnątrz koła jednostkowego i system jest asymptotycznie stabilny. Wynika to również ze wzorów Viete'a.

2) Kryterium Michajłowa (wersja dyskretna)

Ze znanych zależności można trzymać równanie mianownika transmitancji $M(z)$:

$$z = e^{j\omega}: \quad \omega \in [-\pi, \pi] \quad \text{lub} \quad \omega \in [0, 2\pi], \quad z = \cos \omega + j \sin \omega,$$

$$z^n = \cos n\omega + j \sin n\omega, \quad M(j\omega) = P(\omega) + jQ(\omega) = \operatorname{Re} M(\omega) + j \operatorname{Im} M(\omega).$$

Twierdzenie 7.12

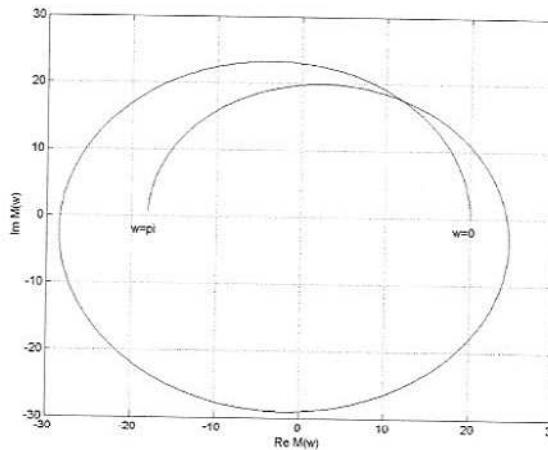
Równanie $M(z)$ ma $m \leq k$ pierwiastków wewnątrz okręgu jednostkowego wtedy i tylko wtedy, gdy przyrost argumentu $M(j\omega)$ przy zmianie ω od $[0, \pi]$ wynosi:

$$\Delta \arg(M(j\omega)) = m\pi \quad \text{lub} \quad \Delta \arg(M(j\omega)) = 2m\pi. \\ 0 \leq \omega \leq \pi \quad -\pi \leq \omega \leq \pi$$

Aby układ dyskretny był asymptotycznie stabilny, to wszystkie n pierwiastków musi leżeć wewnątrz okręgu jednostkowego, czyli przyrost argumentu musi być $n\pi$.

$$\Delta \arg(M(j\omega)) = n\pi = 2n \frac{\pi}{2}. \\ 0 \leq \omega \leq \pi$$

Przykład przebiegu wykresu mianownika transmitancji dla stabilnego układu trzeciego rzędu widać na rysunku 7.25.



Rys. 7.25. Wykres mianownika
 $M(\omega) = 24\omega^3 + 2\omega^2 - 5\omega - 1 = 24(\omega + 0,25)(\omega + 0,33)(\omega - 0,5)$

Inna interpretacja mówi, że dla wzrastającego ω , zera wielomianów $P(\omega) = 0$ i $Q(\omega) = 0$ powinny występować naprzemiennie na osiach P i Q (rys. 7.25).

3) Kryterium Nyquista (wersja dyskretna)

Analog twierdzenia Nyquista może być stosowany bezpośrednio do układu dyskretnego.

Twierdzenie 7.13

Jeżeli układ dyskretny otwarty o transmitancji $W(j\omega)$ jest niestabilny tzn. m pierwiastków znajduje się poza kołem jednostkowym, to układ zamknięty będzie stabilny wtedy i tylko wtedy, gdy przyrost argumentu ($\Delta \arg$) dyskretnej charakterystyki amplitudowo-fazowej układu otwartego spełnia warunki:

$$\begin{aligned} \Delta \arg [1 + W(j\omega)] &= 2m\pi \quad \text{lub} \\ -\pi \leq \omega \leq \pi & \quad \Delta \arg [1 + W(j\omega)] = m\pi. \\ 0 \leq \omega \leq \pi & \end{aligned}$$

4) Przekształcenie w

Przekształcenie geometryczne w przekształca okrąg jednostkowy na płaszczyźnie z w lewą półpłaszczyznę na płaszczyźnie w . Do tak przekształconego równania można stosować znane kryteria dla układów ciągłych

$$z = \frac{w+1}{w-1}.$$

Dany jest wielomian charakterystyczny

$$M(z) = a_k z^k + a_{k-1} z^{k-1} + \dots + a_1 z + a_0 = 0,$$

i dwie liczby zespolone: $w = a + jb$, $z = \bar{a} + j\bar{b}$.

Po zastosowaniu przekształcenia w , $z = \frac{w+1}{w-1}$, otrzymuje się:

$$N(w) = b_k w^k + b_{k-1} w^{k-1} + \dots + b_1 w + b_0 = 0.$$

Do wielomianu $N(w)$ można już stosować znane kryteria stabilności układów ciągłych. Na przykład:

$$M(z) = a_3 z^3 + a_2 z^2 + a_1 z + a_0 = 0 \Rightarrow z = \frac{w+1}{w-1},$$

$$N(w) = a_3 \frac{(w+1)^3}{(w-1)^3} + a_2 \frac{(w+1)^2}{(w-1)^2} + a_1 \frac{w+1}{w-1} + a_0 = 0,$$

$$a_3 (w+1)^3 + a_2 (w+1)^2 (w-1) + a_1 (w+1)(w-1)^2 + a_0 (w-1)^3 = 0,$$

$$b_3 w^3 + b_2 w^2 + b_1 w + b_0 = 0,$$

$$b_3 = a_3 + a_2 + a_1 + a_0, \quad b_1 = 3a_3 - a_2 - a_1 + 3a_0,$$

$$b_2 = 3a_3 + a_2 - a_1 - 3a_0, \quad b_0 = a_3 - a_2 + a_1 - a_0.$$

Do wielomianu $b_3 w^3 + b_2 w^2 + b_1 w + b_0 = 0$ na płaszczyźnie w można już stosować np. kryterium Hurwitz'a.

5) Kryterium Gerszgorina (wersja dyskretna)

Twierdzenie 7.14

Twierdzenie Gerszgorina nie może być stosowane bezpośrednio do układu dyskretnego:

$$x_{k+1} = A_D x_k, \quad x_0 \neq 0, \quad x_k \in R^n \quad \forall k \geq 0.$$

Równanie charakterystyczne ma postać $\det |zI - A_D| = 0$.

Wykorzystamy transformację w

$$\begin{aligned} \det \left| \frac{w+1}{w-1} I - A_D \right| &= \det \left(\frac{1}{w-1} (w(I - A_D) + I + A_D) \right) = \\ &= \det \left(\frac{1}{w-1} (I - A_D)(wI - (A_D - I)^{-1}(I + A_D)) \right). \end{aligned}$$

Ponieważ $\det\left(\frac{1}{w-1}(I - A_D)\right) \neq 0$, to ostatecznie $\det(wI - A_G) = 0$,

$$A_G = (A_D - I)^{-1}(I + A_D) = (A_D - I)^{-1}(2I + A_D - I) = 2(A_D - I)^{-1} + I,$$

gdzie w – jest zmienną na płaszczyźnie odpowiadającej układom ciągłym, tzn. odpowiada to

$$\det(sI - A) = 0.$$

Stąd, twierdzenie Gerszgorina można stosować do układów dyskretnych, ale nie bezpośrednio do macierzy A_D , tylko do macierzy A_G : $A_G = 2(A_D - I)^{-1} + I$.

7.3. Stabilność układów nieliniowych ciągłych

7.3.1. Definicje stabilności rozwiązania

Układ liniowy, posiadający konkretne parametry fizyczne, jest albo globalnie stabilny, albo globalnie niestabilny. Globalność dotyczy dowolności wyboru warunków początkowych. Dla układów nieliniowych podstawowe wyniki matematyczne dotyczące stabilności zostały opracowane przez A.M. Lapunowa. Wykazał on, że dla układów dynamicznych opisujących nieliniowymi równaniami różniczkowymi, cechę stabilności nie zawsze można przypisać bezpośrednio do układu, gdyż dany układ nieliniowy może być stabilny w jednym punkcie pracy, a niestabilny w innym. Należy więc raczej mówić o stabilności konkretnego wybranego rozwiązania równania różniczkowego względem zmian warunków początkowych lub zmian wymuszenia.

Stabilność rozwiązania względem zmian warunków początkowych

Jest dany system nieliniowy opisany wektorowym równaniem różniczkowym nieliniowym

$$\begin{aligned}\dot{x}(t) &= f(x(t), t), \\ x(t) &= [x_1(t), \dots, x_n(t)]^T \in R^n\end{aligned}\tag{7.3}$$

Oznaczmy interesujące nas rozwiązanie równania (7.3) przez $\hat{x}(t)$, $\hat{x}(t_0) = \hat{x}_0$.

Def. 7.2. (A.M. Lapunow) Rozwiązanie $\hat{x}(t)$ równania (7.3) nazywamy stabilnym wtedy i tylko wtedy, gdy dla każdego $\varepsilon > 0$ istnieje taka liczba $\delta(\varepsilon) > 0$, że implikacja

$$\|x(t_0) - \hat{x}(t_0)\| < \delta \Rightarrow \|x(t) - \hat{x}(t)\| < \varepsilon \quad \text{dla } \forall t \geq t_0,$$

zachodzi dla każdego rozwiązania $x(t)$ sąsiadującego z $\hat{x}(t)$ [7SAJ; 7SKB; 7GRP1].

Definicja ta mówi, że interesujące rozwiązanie $\hat{x}(t)$ będzie stabilne jeżeli dla dowolnie wybranej dla niego odległości normowej $\varepsilon > 0$ (ε promień rurki otaczający $\hat{x}(t)$ dla $t > t_0$) można dobrąć (istnieje) otoczenie kołowe o promieniu $\delta > 0$ warunku początkowego $\hat{x}(t_0)$ takie, że jeżeli warunki początkowe dla wszystkich innych rozwiązań $x_i(t)$ będą z wnętrza tego koła $x_i(t_0)$, to te rozwiązania $x_i(t)$ nie oddalą się od rozwiązania $\hat{x}(t)$ poza rurkę o wybranym promieniu ε dla $t > t_0$. Otoczenie początkowe δ może być przy tym bardzo małe – byleby istniało. Widać wyraźnie, że z jednej strony można taką własność stabilności rozwiązania $\hat{x}(t)$ interpretować jako jego zdolność do przyciągania do siebie sąsiednich rozwiązań, które wystartowały z pobliskich warunków początkowych (punkt atrakcji), lub jako własną cechę nieczułości na małe zmiany własnych warunków początkowych – mała zmiana warunków początkowych \Rightarrow mała zmiana rozwiązania. Z punktu widzenia matematyki powyższa definicja pokrywa się z definicją ciągłości takiego rozwiązania $\hat{x}(t)$ ze względu na zmiany jego warunku początkowego. Zwraca się uwagę, że nie istotna jest ograniczoność rozwiązania. Rozwiązanie może rosnąć nieograniczenie i być stabilne w sensie Lapunowa.

Definicja powyższa znajduje prostą interpretację w nieliniowym modelu wahadła. Wahadło w swoim dolnym położeniu posiada w stanie ustalonym rozwiązanie stałe $\hat{x}(t)$ dla kąta odchylenia od pionu $\hat{x}(t) = \alpha(t) = 0$. Wszystkie inne swobodne rozwiązania startujące w dużym zakresie wychyleń początkowych zawsze wróć do rozwiązania $\alpha(t) = 0$ (można wybrać dowolne ε i dobrą δ). To rozwiązanie zerowe jest stabilne. Inne rozwiązanie stałe (pionowe górne) $\hat{x}(t) = \alpha(t) = \pi$ jest niestabilne. Teoretycznie takie rozwiązanie istnieje – odwrócone wahadło może stać pionowo i nieruchomo. Nie można jednak dobrą dla takiego warunku początkowego $\hat{x}(t_0) = \alpha(t_0) = \pi$, żadnego otoczenia początkowego δ, z którego wahadło powróci do górnego pionu. Parametr δ nie istnieje.

Stabilność rozwiązania względem warunku początkowego i wymuszenia

Rozważmy układ nieliniowy z zaburzeniem prawej strony

$$\dot{x}_Z(t) = f(x_Z, t) + r(x_Z, t),\tag{7.4}$$

Def. 7.3. (I.G. Małkin) Rozwiązanie $\hat{x}(t)$ równania bez zaburzenia (7.3) nazywamy stabilnym w sensie Małkinia wtedy i tylko wtedy, gdy dla każdego $\varepsilon > 0$ istnieją takie liczby $\delta_1(\varepsilon) > 0$ i $\delta_2(\varepsilon) > 0$, że dla każdego rozwiązania $x_Z(t)$ układu z dowolnym ograniczonym zaburzeniem (7.4) zachodzi implikacja

$$\|x_Z(t_0) - \hat{x}(t_0)\| < \delta_1 \quad i \quad \|r(x_Z, t)\| < \delta_2 \Rightarrow \|x_Z(t) - \hat{x}(t)\| < \varepsilon \quad \text{dla } \forall t \geq t_0.$$

Małkin [7MAI1; 7MAI2] udowodnił twierdzenie, że jeżeli rozwiązanie systemu (7.3) jest asymptotycznie stabilne w sensie Lapunowa, to jest ono również stabilne przy stale działającym ograniczonym zaburzeniu (w sensie Def. 7.3). Powyżej omówiona stabilność (w sensie Małkinia) prezentuje rozszerzone pojęcie stabilności rozwiązania przy zmiennych warunkach początkowych i stale działającym ograniczonym zaburzeniu. Jest więc to połączenie pojęcia stabilności w sensie Lapunowa i stabilności w sensie BIBO przy zerowych warunkach początkowych (podrozdz. 7.1.2). Obszerną dyskusję i wyniki naukowe dotyczące tych problemów można znaleźć w pozycjach [7MAI; 7HSJ] i [7BYW].

7.3.2. Punkty równowagi

Omówione w poprzednim podrozdziale definicje stabilności mogą dotyczyć dowolnego rozwiązania nieliniowego równania (7.3). W zastosowaniach inżynierskich nie zachodzi potrzeba sprawdzania stabilności wszystkich rozwiązań danego równania różniczkowego. Rozważa się tylko stabilność pewnych szczególnych rozwiązań – takich, które dla wszystkich czasów $t \geq 0$ wykazują wartość ustaloną $\hat{x}(t) = \text{const}$, czyli ich pochodna się zeruje, $\dot{\hat{x}}(t) = 0$ dla $\forall t \geq 0$. Takie rozwiązania nazywa się punktami równowagi systemu (7.3).

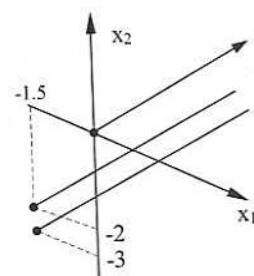
$$\begin{aligned}\dot{\hat{x}}(t) &= \text{const}, \quad \forall t \geq 0, \\ \dot{\hat{x}}(t) &= 0, \quad \forall t \geq 0,\end{aligned}\quad \text{stąd } 0 = f(\hat{x}(t), t) \quad (7.5)$$

Z (7.5) można znaleźć punkty równowagi systemu (7.3).

Przykład 7.7

Dane jest równanie nieliniowe Lotki – Volterry

$$\begin{cases} \dot{x}_1(t) = -x_1(t)x_2(t) - \frac{3}{2}x_2(t), \\ \dot{x}_2(t) = [x_2(t) + 2] \cdot [2x_1(t) - x_2(t)]. \end{cases}$$



Rys. 7.26. Punkty równowagi

Punktami równowagi (rys. 7.26) są trzy punkty płaszczyzny fazowej, wynikające z równań:

$$-x_1x_2 - \frac{3}{2}x_2 = 0, \quad (x_2 + 2)(2x_1 - x_2) = 0,$$

$$x_{1r} = \begin{cases} x_1 = 0, \\ x_2 = 0, \end{cases} \quad x_{2r} = \begin{cases} x_1 = -1.5, \\ x_2 = -3, \end{cases} \quad x_{3r} = \begin{cases} x_1 = -1.5, \\ x_2 = -2. \end{cases}$$

Trudno bez obliczeń ocenić, który z tych punktów równowagi jest stabilny, a które nie.

Można pokazać, że te ogólne definicje stabilności obejmują również układy liniowe.

$$\dot{x}(t) = Ax(t), \quad x(0) = x_0.$$

Jeżeli macierz A jest nieosobliwa, to jedynym rozwiązaniem układu $Ax = 0$ jest punkt równowagi $x = 0$. Do tego też punktu (rozwiązania) odnoszą się wszystkie rozwiązania o stabilności przedstawione w podrozdziale 7.1. Jeśli macierz stanu A jest osobliwa, np. dla układu całkującąco-inercyjnego, to takich punktów równowagi jest nieskończono wiele. Jeżeli jakiś punkt równowagi układu liniowego jest stabilny, to cały układ jest stabilny.

Dla układów liniowych sterowanych, z nieosobliwą macierzą A i ustaloną sterowaniem u_u

$$\dot{x}(t) = Ax(t) + Bu(t) \quad (7.6)$$

odpowiednim punktem równowagi będzie

$$Ax_u + Bu_u = 0 \Rightarrow x_u = -A^{-1}Bu_u.$$

Wprowadzając zmienne odchyłkowe od tego stanu równowagi: $x(t) = x_u + \Delta x(t)$, $u(t) = u_u + \Delta u(t)$, otrzymuje się równanie stanu dla zmiennych odchyłkowych

$$\dot{\Delta x}(t) = A \Delta x(t) + Ax_u + B \Delta u(t) + Bu_u = A \Delta x(t) + B \Delta u(t).$$

Jest to takie samo równanie jak (7.6).

Bez straty ogólności można więc przyjąć $u_u = 0$ i $x_u = 0$ i w zagadnieniach stabilności względem warunków początkowych rozważać stabilność rozwiązania zerowego układu jednorodnego:

$$\dot{x}(t) = Ax(t), \quad x(0) = 0.$$

W układach nieliniowych mogą zachodzić różne sytuacje: może być jeden punkt równowagi, może ich być kilka lub nieskończono wiele. Tylko w tym pierwszym przypadku można utożsamiać stabilność punktu równowagi z globalną stabilnością całego układu. Jeżeli układ nieliniowy ma wiele punktów równowagi, to z reguły obok stabilnego punktu równowagi występuje niestabilny punkt równowagi i nic da się powiedzieć, czy globalnie układ jest stabilny, czy niestabilny.

Z przedstawionej powyżej definicji stabilności w sensie Lapunowa nie wynika metoda badania i praktycznego oceniania stabilności poszczególnych punktów równowagi.

7.3.3. Pierwsza metoda Lapunowa

Pierwsza metoda Lapunowa oceny stabilności względem warunków początkowych polega na linearizacji układu nieliniowego wokół wybranego punktu równowagi x_r i ocenie stabilności tak uzyskanego modelu liniowego.

Dany jest układ nieliniowy (7.3) i jego punkty równowagi

$$\dot{x}(t) = f(x(t), t) \Rightarrow x_r.$$

Do równania wprowadzamy zmienne odchyłkowe Δx , oznaczamy $x(t) = x_r + \Delta x(t)$ oraz rozwijamy prawą stronę w szereg Taylora wokół punktu równowagi. Po pominięciu wyrazów wyższych od pierwszego stopnia, otrzymujemy

$$\dot{\Delta x}(t) = f(x_r) + \frac{\partial f}{\partial x} \Big|_{x_r} \Delta x + \dots$$

Oznaczmy jakobianową macierz pierwszych pochodnych (jakobian) przez $J(x)$. Macierz $J(x)$ w wybranym punkcie równowagi x_r staje się macierzą liczbową kwadratową J .

$$\frac{\partial f}{\partial x} = J(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \dots & \dots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}.$$

Biorąc pod uwagę zależność $f(x_r) = 0$, otrzymuje się znaną liniową postać równania stanu

$$\dot{\Delta x}(t) = J \Delta x(t) \quad (7.7)$$

Twierdzenie 7.15 (warunek wystarczający [7KAT])

- Jeżeli model liniowy (7.7) otrzymany po linearyzacji wokół wybranego punktu równowagi x_r jest asymptotycznie stabilny, to model nieliniowy (7.3), będzie również lokalnie asymptotycznie stabilny wokół tego punktu w sensie Lapunowa.
- Jeżeli model liniowy (7.7) jest niestabilny, to rozwiązanie zerowe układu nieliniowego jest też niestabilne.
- Jeżeli model liniowy jest tylko stabilny, to nic o stabilności tego punktu x_r dla modelu (7.3) powiedzieć nie można.

Przykład 7.8

Dany jest mechaniczny model II rzędu z nieliniową sprężyną:

$$\ddot{x}(t) + a\dot{x}(t) + bx(t) + cx^2(t) = 0, \quad a > 0, \quad b > 0, \quad c > 0.$$

Należy znaleźć punkty równowagi i ocenić ich stabilność.

$$\dot{x}_1(t) = f_1(x) = x_2(t),$$

$$\dot{x}_2(t) = f_2(x) = -bx_1(t) - ax_2(t) - cx_1^2(t).$$

Punkty równowagi: $f_1 = 0 \Rightarrow x_2 = 0$,

$$f_2 = 0 \Rightarrow -bx_1 - cx_1^2 = 0, \quad x_1(-b - cx_1) = 0 \Rightarrow x_{11} = 0 \text{ lub } x_{12} = -b/c.$$

Na płaszczyźnie zmiennych stanu (przesunięcie, szybkość) otrzymuje się dwa punkty równowagi:

$$x_{1r} = [0, 0]^T, \quad x_{2r} = [-b/c, 0]^T.$$

Po linearyzacji macierz jakobianowa ma postać

$$J(x) = \begin{bmatrix} 0 & 1 \\ -b - 2cx_1 & -a \end{bmatrix}.$$

W pierwszym punkcie równowagi otrzymamy macierz liczbową

$$J_1 = \begin{bmatrix} 0 & 1 \\ -b - 2c \cdot 0 & -a \end{bmatrix}_{x_{1r}} = \begin{bmatrix} 0 & 1 \\ -b & -a \end{bmatrix} \Rightarrow \begin{vmatrix} 0-\lambda & 1 \\ -b & -a-\lambda \end{vmatrix} = 0.$$

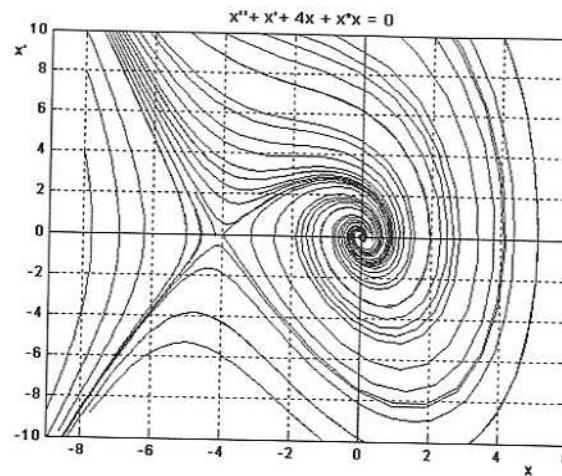
Wielomian charakterystyczny $\lambda^2 + a\lambda + b = 0$, $\Delta = a^2 - 4b$, $\lambda_{1,2} = \frac{-a \pm \sqrt{a^2 - 4b}}{2}$ posiada dla:

- $a > 2\sqrt{b}$ – dwie wartości własne rzeczywiste, ujemne; stąd układ liniowy jest asymptotycznie stabilny, a obrazem jest węzeł;
- $a < 2\sqrt{b}$ – dwie wartości własne zespolone o ujemnych częściach rzeczywistych; stąd układ liniowy jest asymptotycznie stabilny, a obrazem jestognisko (wir lub spirala).

Dla drugiego punktu równowagi x_{2r} macierz jakobianowa ma wartość

$$J_2 = \begin{bmatrix} 0 & 1 \\ b & -a \end{bmatrix} \Rightarrow \begin{vmatrix} 0-\lambda & 1 \\ b & -a-\lambda \end{vmatrix} = 0,$$

^a wielomian charakterystyczny $\lambda^2 + a\lambda - b = 0$, $\Delta = a^2 + 4b$, $\lambda_{1,2} = \frac{-a \pm \sqrt{a^2 + 4b}}{2}$.



Rys. 7.27. Płaszczyzna fazowa

W drugim punkcie równowagi występują dwa pierwiastki rzeczywiste o różnych znakach – układ jest niestabilny (obrazem jest siodło). Złożenie tych dwóch obrazów pokazuje całą rodzinę trajektorii na płaszczyźnie fazowej (rys. 7.27). Technikami cyfrowymi można uzyskać dokładny obszar warunków początkowych gwarantujących stabilność, czyli obszar atrakcji punktu $(0, 0)$.

Przykład 7.9

Przykładem układu posiadającego nieskończoną liczbę punktów równowagi jest model wahadła matematycznego (z tarciem). Zmienna $x(t)$ oznacza kąt odchylenia od pionu z pozycji dolnej.

$$\ddot{x}(t) + a\dot{x}(t) + \sin[x(t)] = 0, \Rightarrow \begin{cases} \dot{x}_1(t) = x_2(t), \\ \dot{x}_2(t) = -ax_1(t) - \sin[x_1(t)]. \end{cases}$$

Punkty równowagi: $x_{2r} = 0, \sin x_1 = 0 \Rightarrow x_{1r} = \pm n\pi$. Macierz jacobianowa

$$J(x) = \begin{bmatrix} 0 & 1 \\ -\cos(x_1) & -a \end{bmatrix}.$$

Zachodzą tutaj dwa przypadki:

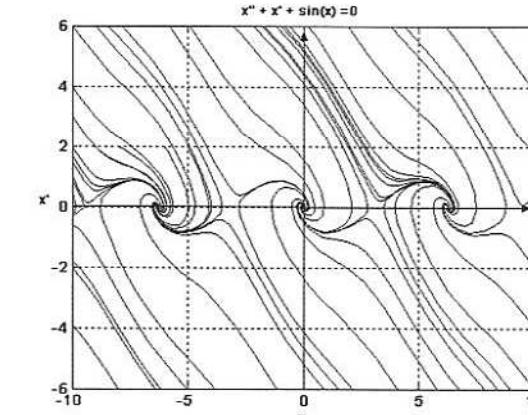
1) Punkty równowagi dla parzystych n dają macierz $J = \begin{bmatrix} 0 & 1 \\ -1 & -a \end{bmatrix}$ i wielomian charakterystyczny $\lambda^2 + \lambda a + 1 = 0, \Delta = a^2 - 4$, a stąd dla:

• $0 < a < 2 \Rightarrow \Delta < 0, \lambda_{1,2} = \left(-a \pm \sqrt{a^2 - 4} \right)/2 = \left(-a \pm j\sqrt{4 - a^2} \right)/2$ mamy dwie wartości własne zespolone o ujemnych częściach rzeczywistych; układ liniowy jest asymptotycznie stabilny, a obrazem jest ognisko zwane wirem lub spiralą;

• $a > 2 \Rightarrow \Delta > 0, \lambda_{1,2} = \left(-a \pm \sqrt{a^2 - 4} \right)/2$ mamy dwie wartości własne rzeczywiste ujemne; stąd układ liniowy jest asymptotycznie stabilny, a obrazem jest węzeł.

2) Punkty równowagi dla nieparzystych n dają macierz $J = \begin{bmatrix} 0 & 1 \\ 1 & -a \end{bmatrix}$ i wielomian charakterystyczny $\lambda^2 + \lambda a - 1 = 0, \Delta = a^2 + 4$, a stąd warunek $\Delta > 0$ i $\lambda_{1,2} = \left(-a \pm \sqrt{a^2 + 4} \right)/2$ mamy dwie wartości własne rzeczywiste o różnych znakach; układ liniowy jest niestabilny, a obrazem jest siodło.

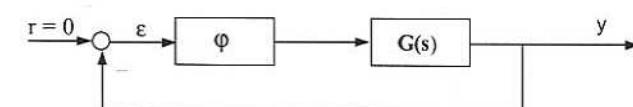
Zbiorczy obraz trajektorii fazowych (x_1, x_2) przedstawia rysunek 7.28.



Rys. 7.28. Płaszczyzna fazowa

7.3.4. Metoda Popowa

Do oceny stabilności specjalnej klasy nieliniowych układów regulacji służy metoda Popowa [7LES]. Bada ona globalną asymptotyczną stabilność rozwiązania zerowego, układu z jednostkowym sprzężeniem zwojnym, przy nieliniowym regulatorze statycznym i liniowym obiekcie dynamicznym (rys. 7.29).

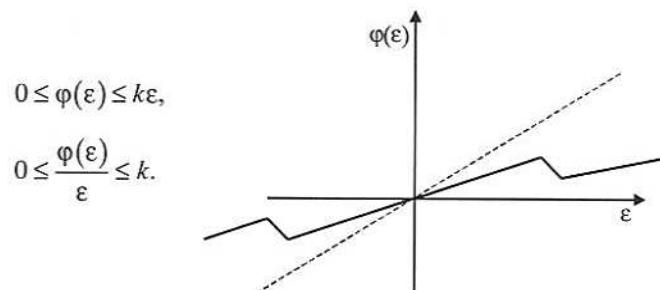


Rys. 7.29. Schemat blokowy

Transmitancja $G(s)$ reprezentuje obiekt ciągły liniowy, stacjonarny i stabilny, dla którego znana jest częstotliwościowa charakterystyka Nyquista.

$$G(s) = C[sI - A]^{-1}B \Rightarrow \begin{cases} \dot{x}(t) = Ax(t) + Bu(t), \\ y(t) = Cx(t). \end{cases}$$

Funkcja $\varphi(\varepsilon)$, to nieliniowa, statyczna charakterystyka regulatora, mieszcząca się w stożku $k\varepsilon$, gdzie k – wybrany współczynnik (rys. 7.30). Dodatkowe, poniżej wymienione założenia, gwarantują istnienie tylko jednego punktu równowagi w „zerze” układu:



Rys. 7.30. Stożek Popowa

Problem: Jakie warunki gwarantują globalną, asymptotyczną stabilność rozwiązania zerowego układu zamkniętego (rys. 7.29), tzn. stabilność dla dowolnych warunków początkowych i dowolnych nieliniowości zawartych pomiędzy osią ε , a prostą $k\varepsilon$?

Twierdzenie 7.16 (W. Popow)

Warunkiem wystarczającym globalnej asymptotycznej stabilności układu automatycznej regulacji jest, aby dla zadanego k ($0 < k < \infty$) istniała liczba rzeczywista q ($-\infty < q < \infty$) taka, że dla każdej częstotliwości $\omega \in (0, \infty)$, będzie zachodzić nierówność

$$\frac{1}{k} + \operatorname{Re}[(1+j\omega q)G(j\omega)] > 0 \quad (7.8)$$

Ponieważ transmitancja widmowa jest liczbą zespoloną: $G(j\omega) = \operatorname{Re}[G(j\omega)] + j\operatorname{Im}[G(j\omega)]$, to nierówność (7.8) ma postać:

$$\frac{1}{k} + \operatorname{Re}[\operatorname{Re}[G] + j\operatorname{Im}[G] + j\omega q \operatorname{Re}[G] - \omega q \operatorname{Im}[G]] > 0,$$

$$\frac{1}{k} + \operatorname{Re}[G(j\omega)] - \omega q \operatorname{Im}[G(j\omega)] > 0.$$

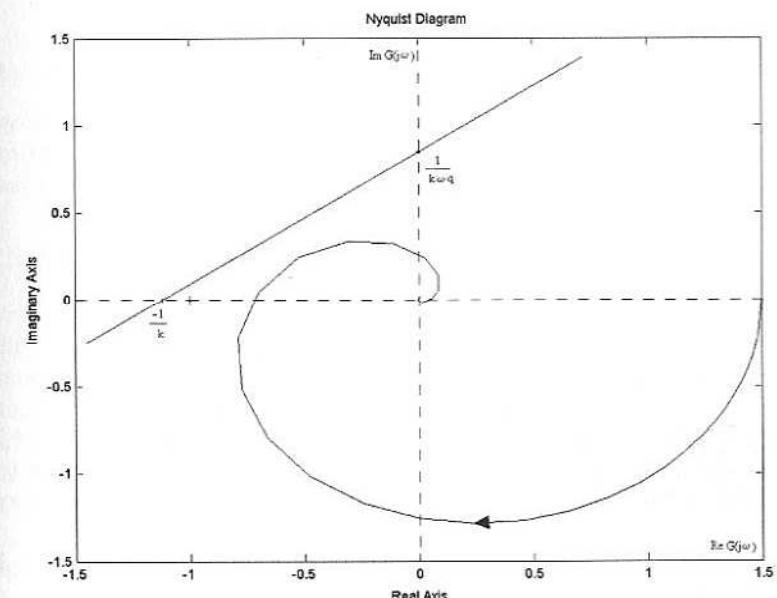
Ostatecznie powinien zachodzić warunek Popowa $\operatorname{Re}[G(j\omega)] > \omega q \operatorname{Im}[G(j\omega)] - \frac{1}{k}$. Nierówność ta może być zapisana w innej formie

$$\operatorname{Im}[G(j\omega)] < \frac{1}{\omega q} \operatorname{Re}[G(j\omega)] + \frac{1}{k\omega q} \quad (7.9)$$

Prawa strona przedstawia tzw. prostą Popowa

$$\operatorname{Im}[G] = \frac{1}{\omega q} \operatorname{Re}[G] + \frac{1}{k\omega q},$$

zaczeplioną w punkcie na osi rzeczywistej $(-1/k)$ i przecinającą oś pionową w punkcie $(1/(k\omega q))$. Ten ostatni punkt zmienia się wraz ze zmieniającą się częstotliwością ω .



Rys. 7.31. Kryterium Popowa

Główny warunek stabilności (7.9) mówi, że dla zadanego stożka nieliniowości k powinna istnieć taka liczba q , że wykres charakterystyki Nyquista transmitancji $G(s)$ powinien mieścić się pod prostą Popowa dla wszystkich ω (rys. 7.31). Sprawdzanie graficzne tego warunku i poszukiwanie takiej liczby q dla wszystkich ω przy ciągle zmieniającym się położeniu prostej Popowa mogło być uciążliwe. Popow zaproponował specjalną transformację unieruchamiającą omawianą prostą. W tym celu wprowadził zmodyfikowaną transmitancję widmową i zmodyfikowany układ współrzędnych

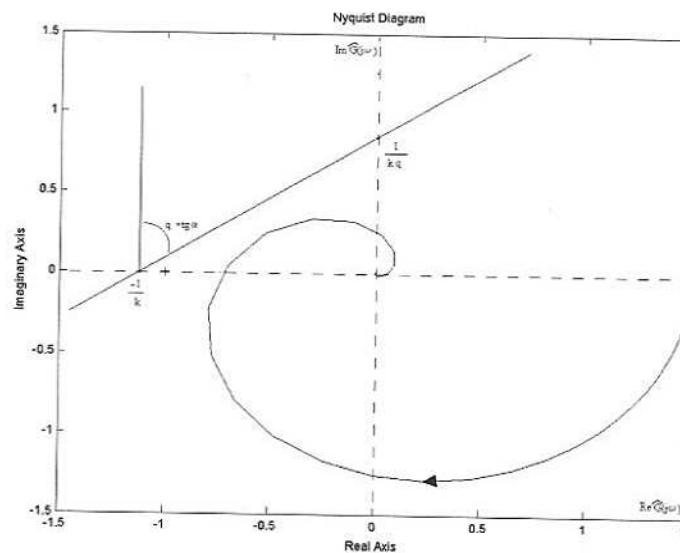
$$\hat{G}(j\omega) = \operatorname{Re}[G(j\omega)] + j\operatorname{Im}[G(j\omega)].$$

W zmodyfikowanym układzie współrzędnych warunek Popowa (7.9) ma postać:

$$\operatorname{Re}[\hat{G}(j\omega)] > -\frac{1}{k} + q \operatorname{Im}[\hat{G}(j\omega)]$$

lub

$$\operatorname{Im}[\hat{G}(j\omega)] < \frac{1}{q} \operatorname{Re}[\hat{G}(j\omega)] + \frac{1}{qk}.$$



Rys. 7.32. Prosta Popowa i wykres zmodyfikowanej transmitancji

Teraz już łatwiej odpowiedzieć na pytanie: czy istnieje taka liczba q , że wykres Nyquista transmitancji zmodyfikowanej $\hat{G}(j\omega)$ dla wszystkich ω mieści się całkowicie pod prostą Popowa, która nie zmienia położenia przy zadanym współczynniku k (rys. 7.32).

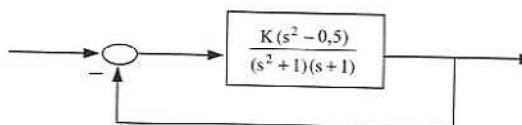
Ogólne zalecenie: współczynnik k powinien być możliwie jak największy, czyli $1/k$ jak najmniejsze. Kryterium Popowa daje warunki wystarczające stabilności.

Przykład 7.10 ([7GRP2])

Dana jest transmitancja:

$$G(s) = \frac{s^2 - 0,5}{(s^2 + 1)(s + 1)}; \quad 0 \leq \frac{\Phi}{\varepsilon} \leq k.$$

1) Dla jakiego liniowego regulatora (proporcjonalnego) $\varphi = K$ układ jest stabilny?



Rys. 7.33. Schemat blokowy

Z warunku Hurwitz wynika, że powyższy układ liniowy (rys. 7.33) jest stabilny dla $0 < K < 2$.

2) Podać warunek na k , aby dla dowolnego φ takiego, że $\frac{\Phi}{\varepsilon} \leq k$, układ nieliniowy był stabilny.

Z warunku (7.9) otrzymuje się:

$$\operatorname{Re}[\hat{G}] - q \operatorname{Im}[\hat{G}] + \frac{1}{k} > 0,$$

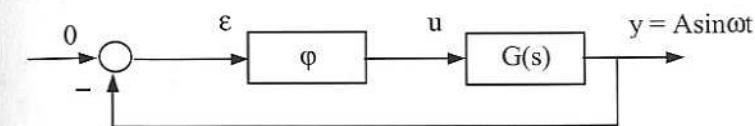
$$\frac{1+q\omega^2}{\omega^2-1} \cdot \frac{\omega^2+0,5}{\omega^2+1} + \frac{1}{k} > 0.$$

Można zauważyć, że dla $q = -1$ otrzymamy $-1 + \frac{0,5}{\omega^2+1} + \frac{1}{k} > 0$.

Powyższa nierówność zachodzi tylko dla $k < 1$ (dla przyjętej $q = -1$ i wszystkich ω). Dopuszczalne nachylenie stożka Popowa jest więc dwa razy mniejsze niż dopuszczalne wzmacnienie regulatora liniowego K .

7.3.5. Metoda funkcji opisującej

Dla systemów nieliniowych, których dynamika opisywana jest równaniem różniczkowym nieliniowym, nie istnieje pojęcie transmitancji. Jedną z przybliżonych metod badania ich stabilności jest częstotliwościowa metoda „funkcji opisującej”, która wykorzystuje analogie metody Nyquista dla transmitancji układu liniowego pracującego w jednostkowej pętli sprzężenia zwrotnego, przy nieliniowości statycznej i dynamice liniowej niskiego rzędu. Obecność układu niskiego rzędu (hipoteza filtra dolnoprzepustowego) ma gwarantować obecność tylko pierwszych harmonicznych w układzie zamkniętym. Metoda ta, zwana też metodą Goldfarba (1947 r.) [7GOL] lub Cypkina (1955 r.) [7CYJ], pozwala ustalić istnienie stabilnych drgań samowzbudnych o prawie sinusoidalnej postaci (określenie ich amplitudy i częstotliwości).



Rys. 7.34. Schemat blokowy

Dla schematu blokowego z rysunku 7.34 można podać trzy następujące założenia:

- 1) $\varphi(\varepsilon)$ – nieliniowość statyczna;
- 2) transmitancja $G(s)$ jest modelem członu filtrującego niskiego rzędu, przepuszczającym tylko główną składową sygnału (jego pierwszą harmoniczną) $-\varepsilon = y = A \sin \omega t$; oznacza to, że w układzie ustalił się stan ustalony w postaci drgań harmonicznych (hipoteza filtra dolnoprzepustowego);
- 3) sygnał sterujący (na wyjściu członu nieliniowego) $u = f(y)$ jest periodyczny z okresem T , można go więc rozwiniąć w szereg Fouriera.

$$u(t) = D_0 + \sum_{k=1}^{\infty} [C_k \cos(k\omega t) + B_k \sin(k\omega t)] = D_0 + \sum_{k=1}^{\infty} D_k (\sin(k\omega t) + \Phi_k).$$

Odpowiedź ustalona członu liniowego będzie zawierała wszystkie harmoniczne wywołane przez człon nieliniowy, z różnymi amplitudami

$$y(t) = -\varepsilon(t) = A_0 + \sum_{k=1}^n A_k [\sin(k\omega t) + \Psi_k],$$

gdzie: $A_k = D_k |G(jk\omega)|$, $A_0 = D_0 |G(0)|$, $\Psi_k = \Phi_k + \arg G(jk\omega)$.

Jeśli częstotliwości wszystkich harmonicznych od $k = 2$, leżą w takim zakresie, że

$$|G(j\omega)| \gg |G(jk\omega)| \text{ dla } k = 2, 3, \dots$$

to prawdziwa pozostaje hipoteza filtra dolnoprzepustowego i sygnał zawiera tylko pierwsze harmoniczne, ze współczynnikami Fouriera B_1 i C_1

$$u(t) = B_1 \sin \omega t + C_1 \cos \omega t.$$

Dla statycznego członu nieliniowego wprowadza się pojęcie tzw. funkcji opisującej pełniącej rolę transmitancji, zależnej od amplitudy, ale niewrażliwej na częstotliwość.

$$J(A) = \frac{\text{amplituda 1. harmonicznej } u}{\text{amplituda } \varepsilon},$$

$$J(A) = \frac{B_1 + jC_1}{A} \text{ (w ogólności liczba zespolona),}$$

$$B_1 = \frac{1}{\pi} \int_0^{2\pi} [\varphi(A \sin \psi)] \sin \psi d\psi, \quad C_1 = \frac{1}{\pi} \int_0^{2\pi} [\varphi(A \sin \psi)] \cos \psi d\psi, \quad \psi = \omega t,$$

$$\text{lub } C_1 = \frac{1}{\pi A} \int_{-E}^E \varphi(\varepsilon) d\varepsilon = \frac{1}{\pi A} \text{ (pole powierzchni pod krzywą } \varphi).$$

Warunkiem powstania drgań w układzie jest, aby $G(j\omega) \cdot J(A) + 1 = 0$.
Musí zachodzić więc równość

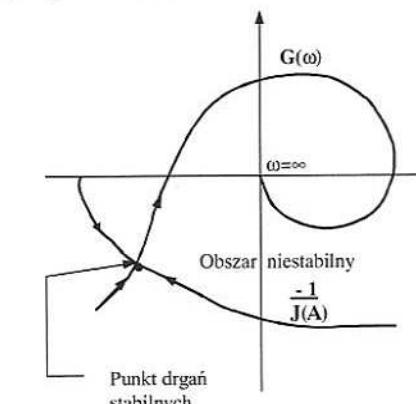
$$G(j\omega) = -1/J(A) \tag{7.10}$$

Stąd można wyznaczyć amplitudę i częstotliwość drgań, czyli tzw. cykl graniczny.

Można wyróżnić dwa kryteria stabilności punktu pracy (rys. 7.35):

- 1) Jeżeli wykres $-1/J(A)$ wchodzi (przy wzrastającej amplitudzie A – czyli drgania narastające) do obszaru stabilności (zdefiniowanego przez transmitancję G), przecinając wykres transmitancji, to w punkcie przecięcia powstają drgania stabilne (stabilny cykl graniczny).

- 2) Jeżeli wykres $-1/J(A)$ wchodzi (przy wzrastającej amplitudzie) do obszaru niestabilności, przecinając wykres transmitancji, to w punkcie przecięcia powstają drgania niestabilne (niestabilny cykl graniczny).



Rys. 7.35. Funkcja opisująca i charakterystyka Nyquista

Przykład 7.11

Wyprowadzić wzór na funkcję opisującą dla przekaźnika za strefę martwą (rys. 7.36):

$$u = 0, \quad -a < \varepsilon < a,$$

$$u = B \frac{\varepsilon}{|\varepsilon|}, \quad \begin{cases} \varepsilon \leq -a, \\ \varepsilon \geq a. \end{cases}$$

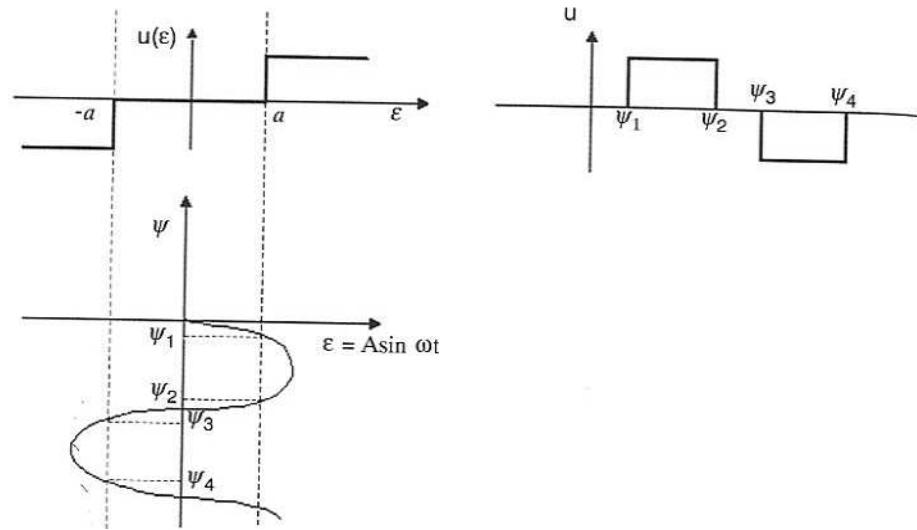
$$B_1 = 4 \frac{B}{\pi} \int_{\psi_1}^{\frac{\pi}{2}} \sin \psi d\psi = \frac{4B}{\pi} \left[-\cos \frac{\pi}{2} + \cos \psi_1 \right] = \frac{4B}{\pi} \cos \psi_1,$$

$$A \sin \psi_1 = a, \quad \sin \psi_1 = \frac{a}{A}, \quad \cos \psi_1 = \frac{A}{\sqrt{1-(a/A)^2}}, \quad B_1 = (4B/\pi) \sqrt{1-(a/A)^2},$$

$$C_1 = 0, \quad J(A) = (4B/\pi A) \sqrt{1-(a/A)^2}, \quad A \geq a \Rightarrow A/a = \alpha \geq 1,$$

$$-\frac{1}{J(A)} = -\frac{-\pi A}{4B a / A \cdot \sqrt{A^2 / a^2 - 1}} = \frac{-\pi A^2 a}{4Ba^2 \sqrt{\alpha^2 - 1}} = \frac{-A\pi}{4B} \frac{\alpha}{\sqrt{\alpha^2 - 1}} = \frac{-a\pi}{4B} \frac{\alpha^2}{\sqrt{\alpha^2 - 1}}.$$

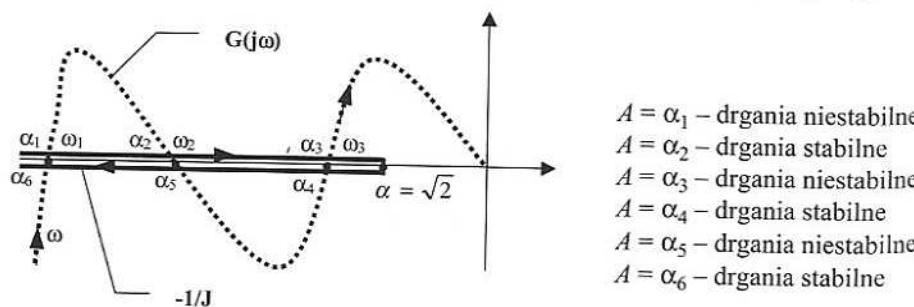
Funkcja ta przebiega wzdłuż ujemnej osi rzeczywistej od wartości $-\infty$ dla $\alpha = 1$ do maksymalnej wartości $(-\pi/2B)$ dla $\alpha = \sqrt{2}$ i ponownie do $-\infty$ dla $\alpha > \sqrt{2}$.



Rys. 7.36. Charakterystyki przekaźnika

Obszar stabilny, to obszar na lewo od wykresu Nyquista $G(j\omega)$ przy zwiększającym się ω , obszar niestabilny to obszar na prawo.

Nakładając otrzymany wykres $-1/J$ na przykładową charakterystykę Nyquista $G(j\omega)$ (rys. 7.37) otrzymuje się w punktach z częstotliwością $\omega_1 < \omega_2 < \omega_3$ drgania o różnych amplituach α , np. dla ω_1 i amplitudy $A = \alpha_1$ amplituda drgań narasta i stabilizuje się przy tej samej częstotliwości ω_1 i większej amplitudzie $A = \alpha_6$ ($\alpha_1 < \alpha_2 < \alpha_3 < \alpha_4 < \alpha_5 < \alpha_6$).



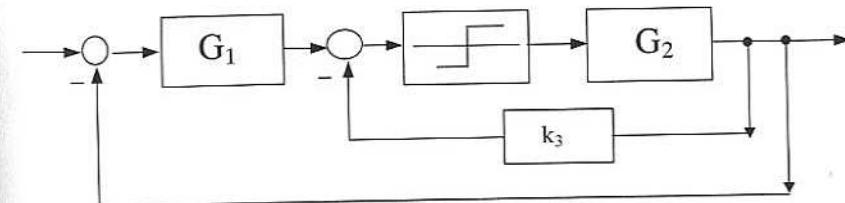
Rys. 7.37. Funkcja opisująca

Przykład 7.12

Znaleźć częstotliwość i amplitudę drgań ustalonych w układzie (rys. 7.38).

$$G_1 = \frac{k_1}{1+sT_1}, \quad G_2 = \frac{k_2}{s(1+sT_1)},$$

$$T_1 = 0,1; \quad T_2 = 0,5; \quad k_1 = 10; \quad k_2 = 20; \quad k_3 = 5; \quad B = 10.$$



Rys. 7.38. Schemat blokowy

Układ spełnia warunek filtra dolnoprzepustowego – stosujemy metodę funkcji opisujących.

Dla przekaźnika idealnego parametr $a = 0$. Funkcja opisująca $J(A) = 4B/\pi A$. „Transmitancja zastępcza” i równanie charakterystyczne:

$$\frac{J(A)k_2}{(1+T_2s)s+k_2k_3J(A)} \cdot \frac{k_1}{1+T_1s} + 1 = 0,$$

$$T_1T_2s^3 + (T_1+T_2)s^2 + [1+T_1k_2k_3J(A)]s + (k_1+k_3)k_2J(A) = 0.$$

Problem można rozwiązać na dwa sposoby.

- 1) Zakładając, że w układzie powstaną drgania ustalone, szukamy rozwiązań okresowego $\varepsilon = A \sin \omega t$. Jeżeli to rozwiązanie istnieje, to układ ma parę pierwiastków czysto urojonych $s = \pm j\omega$. Podstawiamy je do równania charakterystycznego i przerównujemy do zera część Re i Im:

$$P(A, \omega) = (k_1 + k_3)k_2J(A) - (T_1 + T_2)\omega^2 = 0,$$

$$Q(A, \omega) = (1 + T_1k_1k_2J(A)) - T_1T_2\omega^2 = 0,$$

$$\omega^2 = \frac{(k_1+k_3)k_2J(A)}{T_1+T_2}, \quad (1+T_1k_2k_3J(A)) = T_1T_2 \frac{(k_1+k_3)k_2J(A)}{T_1+T_2},$$

$$1 = \left[\frac{T_1T_2}{T_1+T_2} (k_1+k_3)k_2 - T_1k_2k_3 \right] J(A), \quad J(A) = 4B/(\pi A),$$

$$A^o = \frac{4B}{\pi} \left[\frac{T_1T_2}{T_1+T_2} (k_1+k_3)k_2 - T_1k_2k_3 \right] = \frac{4B}{\pi} k_2 T_1 \left[\frac{k_1 T_2 - T_1 k_3}{T_1+T_2} \right] = 191,$$

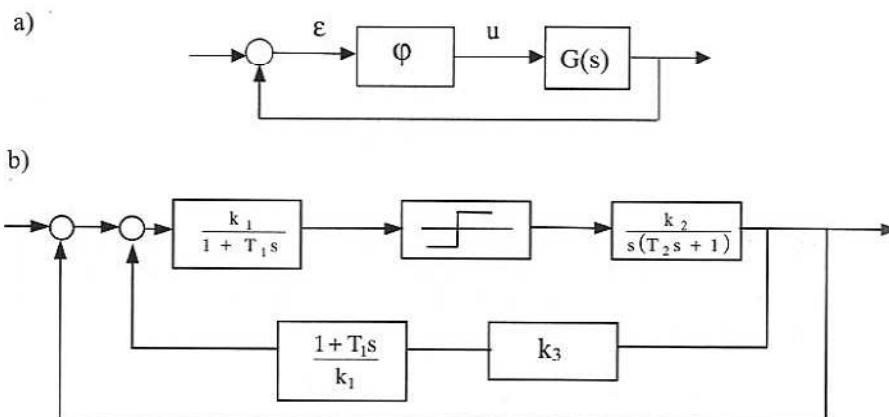
$$\omega^o = \sqrt{\frac{(k_1+k_3)k_2 \frac{4B}{\pi A}}{T_1+T_2}} = \sqrt{\frac{(k_1+k_3)}{k_1+k_2-T_1k_3(T_1+T_2)}} = 5,781 \text{ s}^{-1}$$

Warunek konieczny istnienia drgań, $A > 0$ i $\omega > 0$, jest spełniony, bo zachodzi $k_1 T_2 > T_1 k_3$ ($5 > 0,5$). Istnienie rozwiązania okresowego nie oznacza istnienia drgań stabilnych. Można pokazać, że dla wyliczonej amplitudы A^o spełniony jest warunek Hurwitz'a.

$$\left| \frac{1+T_1k_2k_3J(A^o)}{T_1T_2} \frac{(k_1+k_3)k_3J(A^o)}{T_1+T_2} \right| = 0,75 > 0.$$

Na wyjściu przekaźnika powstaną drgania o stałej amplitudzie A^o i częstotliwości ω^o .

- 2) Graficzny sposób rozwiązywania tego zadania poprzez sprowadzenie układu do schematu podstawowego i zastosowanie warunku Nyquista (rys. 7.39).



Rys. 7.39. Schematy blokowe: a) ogólny; b) szczegółowy

Transmitancja zastępcza układu liniowego

$$\begin{aligned} G(s) &= \frac{k_1}{1+sT_1} \cdot \frac{k_2}{s(T_2s+1)} \left[\frac{T_1s+1}{k_1} \cdot k_3 + 1 \right] = \\ &= \frac{k_2 [k_3 T_1 s + k_3 + k_1]}{s(1+T_1s)(1+T_2s)} = \frac{10s + 300}{s(0,05s^2 + 0,6s + 1)} = \\ &= \frac{-68\,000 - 200\omega^2}{(\omega^2 + 4)(\omega^2 + 100)} + j \frac{400 \cdot (9\omega^2 - 300)}{\omega(\omega^2 + 4)(\omega^2 + 100)}. \end{aligned}$$

Ponieważ funkcja opisująca $J(A) = 4B/(\pi A)$ jest położona tylko na osi rzeczywistej, to przecięcie wykresów $G(j\omega)$ i $-1/J(A)$ według (7.10), też wystąpi tylko na osi rzeczywistej. Stąd zerowanie części urojonej $G(j\omega)$ wystąpi dla częstotliwości:

$$\omega^o = \sqrt{\frac{300}{9}} = 5,77 \Rightarrow \text{Im}[G(\omega^o)] = 0, \quad \text{Re}[G(\omega^o)] = -15,$$

$$-\frac{1}{J(A^o)} = \frac{-\pi A^o}{4B} = -0,0785 A^o = -15 \Rightarrow A^o = 191.$$

Otrzymano wynik identyczny z wynikiem uzyskanym metodą analityczną.

Obszerne tablice z wyprowadzonymi funkcjami opisującymi dla nieliniowości jednoznacznych i wieloznacznych (z histerezą) są zamieszczone między innymi w [7KUB] i [7GIJ].

8. Sterowalność i obserwowałość

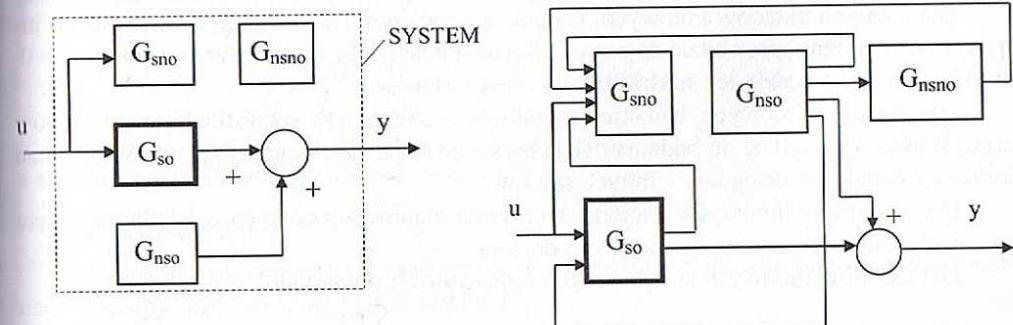
8.1. Sterowalność

Własność sterowalności i obserwowałości stanu obiektu dynamicznego została zdefiniowana i przebadana przez R. Kalmana w latach 50. XX w. Pojęcia te pojawiły się w związku z badaniami nad możliwością konstrukcji kompensatora stabilizującego niestabilny obiekt dynamiczny, za pomocą redukcji niestabilnych biegunków przez takie same niestabilne zera. Otrzymana transmitancja niższego rzędu była wprawdzie stabilna, ale niestabilne zmienne stanu były albo niedostępne dla zmiennej sterującej (niesterowalne), albo niewidoczne w zmiennej wyjściowej (nieobserwowało). R. Kalman w pracy [8KAR] wykazał, że każdy system zapisany równaniami (3.16) można przetransformować, wydzielając wyraźnie cztery bloki:

- 1) sterowalny i obserwowały G_{so} ,
- 2) sterowalny i nieobserwowały G_{sno} ,
- 3) niesterowalny i obserwowały G_{nso} ,
- 4) niesterowalny i nieobserwowały G_{nsno} .

Strukturę połączeń można przedstawić schematycznie (rys. 8.1) lub bardziej dokładnie jak na rysunku 8.2.

Zagadnieniom sterowalności poświęcona jest np. monografia [8KLJ].



Rys. 8.1. Struktura systemu

Rys. 8.2. Schemat możliwych połączeń bloków

Sterowalność wektora stanu wiąże się z możliwością przeprowadzenia wszystkich zmiennych stanu z dowolnego punktu przestrzeni stanu, w dowolny inny punkt, w skończonym czasie. Widać, że niesterowalność wektora stanu wiąże się z niemożliwością znalezienia sterowania, które to zrealizuje. Ten fakt trzeba brać pod uwagę przy projektowaniu układu regulacji (stabilizacji). Przyczyna wystąpienia niesterowalności może mieć swoje źródło w kilku uwarunkowaniach. Pierwszym jest niezamierzona nadmiarowość wektora stanu w budowanym modelu procesu. Taki wektor nie jest rzeczywistym stanem reprezentującym minimalną wymiarowość procesu, co może prowadzić do niesterowalności w całym wymiarze (wystąpi sterowalność tylko w podprzestrzeni). Druga przyczyna niesterowalności względem stanu może tkwić w specyfice doboru parametrów systemu dających za dużo symetrii. Ma to miejsce, np. gdy dwie takie same siły (sterowania) o przeciwnych zwrotach oddziałują w tym samym kierunku na dwie takie same masy połączone sprężyną. Nie ma możliwości zmiany położenia ich środka ciężkości. Również symetria, do której dąży się, równoważąca mostki pomiarowe, powoduje, że w równowadze napięcie wynikowe w jednej przekatnej jest niezależne od napięcia zasilającego w drugiej przekątnej. Stan naładowania baterii zasilającej, np. w układzie pomiarowym światłomierza, nie wpływa więc na poprawność pomiaru. Jest to przypadek wykorzystania zjawiska niesterowalności w praktyce (przykład 8.3). Poniżej przedstawimy bardziej formalne sformułowania, które w odniesieniu do układów liniowych pozwolą na podanie warunków sterowalności i kryteriów jej badania.

Dla systemu nieliniowego $\dot{x}(t) = f(x, u)$ ze zmienną stanu $x(t)$ i sterowaniem $u(t)$ należącym do przestrzeni funkcji całkowalnych $U = L^1[0, \infty]$ wprowadza się trzy definicje:

Def. 8.1. Stan $x_0(t_0)$ nazywa się sterowalnym w danej klasie sterowań U i w danej chwili t_0 , jeżeli dla

$$\forall x_1, \exists t_1 < \infty \text{ oraz } \exists u[t_0, t_1] \in U \text{ takie, że } x_1(t_1) = x(t_1, t_0, x_0(t_0), u[t_0, t_1]).$$

Def. 8.2. Cały system nazywa się w pełni sterowalnym względem stanu w danej klasie sterowań U i w danej chwili t_0 , jeżeli każdy stan $x_0(t_0)$ jest sterowany.

Def. 8.3. Stan $x_1(t_1)$ nazywa się osiągalnym w danej klasie sterowań U i w danej chwili t_1 , jeżeli dla

$$\forall x_0, \exists t_0 < \infty \text{ oraz } \exists u[t_0, t_1] \in U \text{ takie, że } x_1(t_1) = x(t_1, t_0, x_0(t_0), u[t_0, t_1]).$$

Jeżeli dany stan jest sterowany, to nie znaczy, że jest on równocześnie osiągalny.

Dla ciągłych układów liniowych, badanie sterowalności stanu $x_0(t_0)$ do wszystkich innych stanów można sprowadzić do badania sterowalności tylko do stanu zerowego $x_1(t_1) = 0$. Wtedy stan $x_0(t_0)$ będzie też sterowany do innych stanów.

Dla układów liniowych, badanie osiągalności stanu $x_1(t_1)$ z wszystkich innych stanów $x_0(t_0)$ można sprowadzić do badania osiągalności ze stanu zerowego $x_0(t_0) = 0$. Wtedy stan końcowy będzie też osiągalny z innych stanów.

Dla systemów liniowych stacjonarnych zbiór stanów końcowych osiągalnych z zera równa się zbiorowi stanów sterowalnych do zera.

Dla układów liniowych stacjonarnych opisywanych równaniem

$$\dot{x}(t) = Ax(t) + Bu(t),$$

można sformułować trzy równoważne warunki X -sterowalności przedstawione dalej.

Twierdzenie 8.1

Warunkiem koniecznym i dostatecznym X -sterowalności układu liniowego stacjonarnego, gwarantującym istnienie sterowania przeprowadzającego dowolny stan x_0 w dowolny inny stan $x(T)$, jest nieosobliwość kwadratowej $[n \times n]$, symetrycznej macierzy Grama G

$$G(T) = \int_0^T e^{-A\tau} BB' e^{-A'\tau} d\tau = G \quad (8.1)$$

Dowód [8KAR, 8KLJ]

Dla wykazania dostateczności tego warunku wystarczy pokazać, że np. sterowanie $u(t)$ realizujące to zadanie może mieć postać:

$$u(t) = B'e^{-A't}G^{-1}[e^{-AT}x(T) - x_0].$$

Wykorzystując to sterowanie i wzór na rozwiązywanie równania stanu

$$x(T) = e^{AT}x_0 + \int_0^T e^{A(T-\tau)}Bu(\tau)d\tau,$$

otrzymujemy się dla dowolnego czasu T tożsamość

$$x(T) = e^{AT}x_0 + e^{AT}\int_0^T e^{-A\tau}BB'e^{-A'\tau}d\tau \cdot G^{-1}[e^{-AT}x(T) - x_0] = e^{AT}x_0 + x(T) - e^{AT}x_0.$$

Warunek konieczny dowodzi się nie wprost, zakładając, że dla układu sterowanego macierz G jest osobliwa, czyli że wiersze macierzy $e^{-At}B$ są liniowo zależne w przedziale $[0, T]$.

Istnieje więc pewien niezerowy wektor $a \in R^n$ i chwila $t \in [0, T]$ taka, że $a'e^{-At}B = 0$. Z drugiej strony, gdy przyjmie się, np. $x_0 = e^{-AT}a$ i $x(T) = 0$, we wzorze na rozwiązywanie, drugi wyraz (całka) się zeruje i otrzymuje się sprzeczność, gdyż

$$0 = a'e^{AT}e^{-AT}a + a'\int_0^T e^{A(T-\tau)}Bu(\tau)d\tau,$$

a stąd nie jest możliwe, aby $a'a = 0$, przy $a \neq 0$.

Twierdzenie 8.2

Warunkiem koniecznym i dostatecznym X -sterowalności układu liniowego, stacjonarnego jest, aby rząd macierzy Q_c był równy n .

$$Q_c = \begin{bmatrix} B & AB & A^2B & \dots & A^{n-1}B \end{bmatrix}.$$

Dowód

Ponieważ $x(T) - e^{AT}x_0 = \int_0^T e^{A(T-\tau)}Bu(\tau)d\tau$, a ze wzoru Sylwestera $e^{A\tau} = \sum_{i=0}^{n-1} \alpha_i(\tau) \cdot A^i$,

$$\eta(T) = x(T) - e^{AT}x_0 = \int_0^T \sum_{i=0}^{n-1} \alpha_i(T-\tau)A^i Bu(\tau)d\tau =$$

$$= \sum_{i=0}^{n-1} A^i B \int_0^T \alpha_i(T-\tau)u(\tau)d\tau = \sum_{i=0}^{n-1} A^i B \cdot \xi_i(T),$$

$$\eta(T) = \begin{bmatrix} B & AB & A^2B & \dots & A^{n-1}B \end{bmatrix} \begin{bmatrix} \xi_0(T) \\ \vdots \\ \xi_{n-1}(T) \end{bmatrix}.$$

Aby wektor $\eta(T)$ mógł być dowolnym wektorem realizowanym za pomocą odpowiedniego $\xi(T)$, to macierz Q_c musi być odwracalna

$$Q_c = \begin{bmatrix} B & AB & A^2B & \dots & A^{n-1}B \end{bmatrix} \quad (8.2)$$

Kolumny macierzy Q_c muszą tworzyć bazę w R^n , czyli jej rząd musi być równy n , rząd $Q_c = n$.

Jeżeli rząd $B = r$, to warunkiem koniecznym i wystarczającym sterowalności jest, aby

$$\text{rząd } \begin{bmatrix} B & AB & A^2B & \dots & A^{n-r}B \end{bmatrix} = n.$$

Twierdzenie 8.3

Jeżeli macierz charakterystyczna utworzona dla jakiejś wartości własnej λ_k i uzupełniona macierzą B ma rzząd mniejszy od n

$$\text{rząd } [\lambda_k I - A : B] < n, \text{ dla } k \in [1 \dots n],$$

to dla tej wartości własnejej istnieje niesterowalny wektor własne (mod) (dowód [8CAF]).

Niesterowalny w całej przestrzeni stanu $X = R^n$ system liniowy może być sterowany w pewnej podprzestrzeni. Ta liniowa podprzestrzeń sterowalności jest generowana przez liniowo niezależne kolumny macierzy Q_c .

Def. 8.4. System wielowymiarowy nazywamy mocno sterowalnym wtedy i tylko wtedy, gdy dla $u(t) \in R^r$ i macierzy B zbudowanej z kolumnami b_i , $B = [b_1 \ b_2 \ \dots \ b_r]_{n \times r}$, odpowiednie macierze $Q_{Ci} = [b_i \ Ab_i \ A^2b_i \ \dots \ A^{n-1}b_i]$ mają rzząd n . Czyli system jest sterowany każdym sterowaniem z osobna. Każdy system mocno sterowany jest też w pełni sterowany.

Przykład 8.1

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad Q_C = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \Rightarrow \text{rząd } Q_C = 2. \text{ System jest sterowany w pełni.}$$

$$Q_{C1} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \quad Q_{C2} = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} \Rightarrow \text{rząd } Q_{Ci} = 1. \text{ System nie jest mocno sterowany.}$$

Warunkiem koniecznym mocnej X -sterowalności jest, aby wielomian minimalny macierzy A był równy wielomianowi charakterystycznemu.

Warunek Y -sterowalności względem wyjścia

Warunkiem koniecznym i wystarczającym sterowalności systemu względem wyjścia $y(t) \in R^m$ jest warunek, aby rzząd macierzy Q_{Cy} był równy m

$$\text{rząd } \begin{bmatrix} CB & CAB & CA^2B & \dots & CA^{n-1}B \end{bmatrix} = m.$$

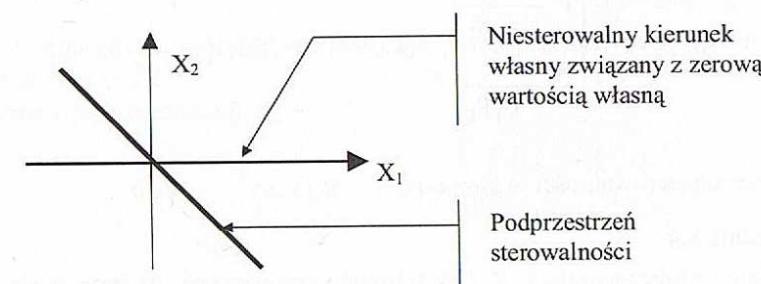
Przykład 8.2

$$A = \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad Q_{Cx} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \Rightarrow \text{system } X\text{-niesterowany},$$

$$C = [1, 0], \quad Q_{Cy} = [1, -1] \Rightarrow \text{system } Y\text{-sterowany.}$$

$$G(s) = \frac{s}{s(s+1)}. \text{ Dla wartości } \lambda_1 = 0, \text{ rząd } [\lambda_1 I - A : B] = \text{rząd } \begin{bmatrix} 0 & -1 & 1 \\ 0 & 1 & -1 \end{bmatrix} = 1 < 2.$$

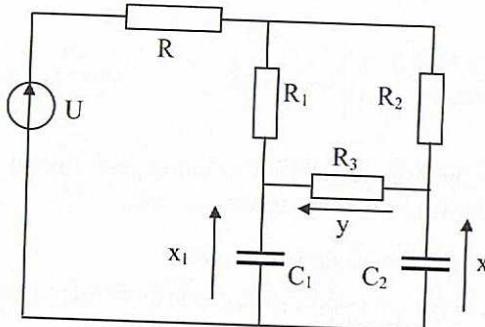
Stąd kierunek własne $[1, 0]^T$ jest niesterowany, a kierunek własne $[1, -1]^T$ pokazuje podprzestrzeń sterowalności (rys. 8.3).



Rys. 8.3. Płaszczyzna stanu

Przykład 8.3

Dany jest mostek RC [8FRB] (rys. 8.4). Należy wykazać jego X -niesterowalność i Y -niesterowalność w przypadku jego zrównoważenia, czyli niezależność napięcia $y(t)$ od źródła zasilania $U(t)$.



Rys. 8.4. Układ mostka RC

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} U, \quad a_{11} = -\frac{1}{C_1} \left(\frac{1}{R_1} + \frac{1}{R_3} \right), \quad a_{12} = \frac{1}{C_1 R_3},$$

$$y = [c_1 \ c_2] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad a_{21} = \frac{1}{C_2 R_3}, \quad a_{22} = -\frac{1}{C_2} \left(\frac{1}{R_2} + \frac{1}{R_3} \right),$$

$$c_1 = 1, \quad c_2 = -1, \quad b_1 = \frac{1}{C_1 R_1}, \quad b_2 = \frac{1}{C_2 R_2}.$$

Warunkiem zrównoważenia mostka jest $R_1 C_1 = R_2 C_2$. Łatwo sprawdzić, że właśnie dla takiego warunku macierze sterowalności są osobliwe. System nie jest sterowalny.

$$Q_{Cx} = [B \ AB] = \frac{1}{C_1 R_1} \begin{bmatrix} 1 & -\frac{1}{C_1 R_1} \\ 1 & -\frac{1}{C_1 R_1} \end{bmatrix}, \quad Q_{Cy} = [CB \ CAB] = \frac{1}{C_1 R_1} [0 \ 0].$$

Podprzestrzeń sterowalności to $x_1 = x_2$.

Twierdzenie 8.4

Dla systemu z macierzami A, B, C warunkiem koniecznym i wystarczającym pełnej sterowalności względem stanu jest, aby transmitancje wchodzące w skład macierzy transmitacji $\bar{G}(s) = (sI - A)^{-1} B$ nie miały skróceń, czyli zera licznika \neq zer mianownika.

Przykład 8.4

Dla równania stanu:

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix} x(t) + \begin{bmatrix} 1 \\ -1 \end{bmatrix} u(t),$$

$$y(t) = [1 \ 0] x(t).$$

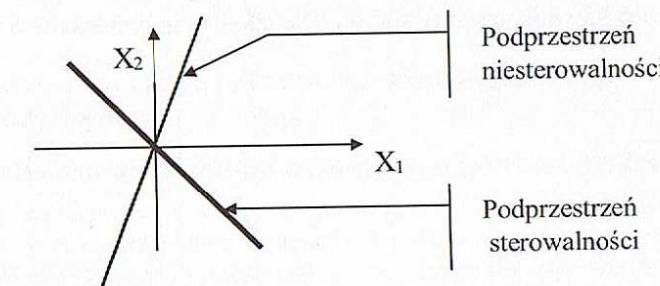
Transmitancja ma postać

$$\bar{G}(s) = \frac{1}{s^2 - s - 2} \begin{bmatrix} s-1 & 1 \\ 2 & s \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{(s-2)(s+1)} \begin{bmatrix} s-2 \\ 2-s \end{bmatrix} = \frac{1}{(s+1)} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Widać, że w transmitancji występują skrócenia. Zgodnie z twierdzeniem układ nie powinien być w pełni sterowalny. Można to stwierdzić, korzystając z warunku (8.2).

$$Q_C = [B \ AB] = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad \text{rząd } Q_C = 1.$$

Podprzestrzeń sterowalności jest pokazana na rysunku 8.5.



Rys. 8.5. Płaszczyzna stanu

$$\text{Transmitancja całkowita ma postać } G(s) = C[sI - A]^{-1} B = \frac{s-2}{(s-2)(s+1)} = \frac{1}{s+1}.$$

Dla tej transmitancji, wyjściowe równanie stanu przedstawia obserwowałość postać według (3.20).

Zgodnie z twierdzeniem 8.3:

$$\text{dla } \lambda_1 = 2, \text{ rząd } [2I - A; B] = \text{rząd } \begin{bmatrix} 2 & -1 & 1 \\ -2 & 1 & -1 \end{bmatrix} = 1,$$

$$\text{dla } \lambda_1 = -1, \text{ rząd } [-I - A; B] = \text{rząd } \begin{bmatrix} -1 & -1 & 1 \\ -2 & -2 & -1 \end{bmatrix} = 2.$$

Niesterowalny jest mod (wektor własny) związany z wartością własną $\lambda_1 = 2$. Ponieważ A jest postaci Frobeniusa, to wektory własne są postaci

$$T = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}.$$

Kierunek $[1 \ -1]^T$ według Q_c , daje podprzestrzeń sterowalności.

Sterowalność w układach wielowymiarowych diagonalnych

Dla standardowej postaci wielowymiarowego równania stanu z blokową macierzą Jordana można sformułować proste kryteria sterowalności.

Układ jest w pełni X -sterowalny, jeżeli:

- w każdym wierszu macierzy B , co najmniej jeden element jest różny od zera;
- jeśli występuje blok Jordana związany z wielokrotną wartością własną i w jego ramach występują klatki Jordana, to wymagana jest liniowa niezależność wierszy macierzy B odpowiadających końcom klatek.

Zaznaczono to w poniższym przykładzie, podkreślając wiersze macierzy B , których liniowa niezależność decyduje o sterowalności. Dla tak dobranej macierzy B układ jest sterowalny.

$$\dot{x}(t) = \begin{bmatrix} \lambda_1 & 0 & & & & 0 \\ 0 & \lambda_1 & 0 & & & \\ & 0 & \lambda_1 & 1 & & 0 \\ & 0 & \lambda_1 & 0 & & \\ & & 0 & \lambda_1 & 1 & \\ & & 0 & \lambda_1 & 1 & \\ & & 0 & \lambda_1 & 0 & 0 \\ & 0 & & & & \lambda_2 & 1 \\ & 0 & & & & 0 & \lambda_2 \end{bmatrix} x(t) + \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ \hline 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ \hline 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} u(t)$$

Wniosek

Układ jednowymiarowy SISO (macierz B jest kolumną), posiadający wielokrotne wartości własne, ale takie, że dają pełny spadek rzędu macierzy $[sI - A]$ (tzn. blok Jordana składa się tylko z jednowymiarowych klatek), jest na pewno niesterowalny. Można to wykazać też z analizy macierzy Q_c . Taki model reprezentuje bowiem równoległe połączenie tych samych podobiektów, sterowanych tym samym sygnałem.

Sterowalność układów liniowych dyskretnych

Warunek X -sterowalności dla układów dyskretnych

$$x_{k+1} = A_D x_k + B_D u_k$$

ma taką samą postać jak w (8.2)

$$\text{rząd } \begin{bmatrix} B_D & A_D B_D & A_D^2 B_D & \dots & A_D^{n-1} B_D \end{bmatrix} = n.$$

Postać tę można wyprowadzić z rozwiązania równania, wykazując dodatkową cechę układów dyskretnych – jeśli układ jest sterowalny, to stan dowolny x_0 można sprowadzić do początku układu w n lub mniej krokach (patrz również (11.8)).

$$x_n = A_D^n x_0 + \sum_{i=0}^{n-1} A_D^{n-1-i} B_D u_i = A_D^n x_0 + \begin{bmatrix} B_D & A_D B_D & \dots & A_D^{n-1} B_D \end{bmatrix} \begin{bmatrix} u_{n-1} \\ \vdots \\ u_0 \end{bmatrix} = A_D^n x_0 + Q_c \cdot u.$$

Dla układu jednowymiarowego sterowanie stabilizujące do $x_n = 0$ ma więc postać wektora $u \in R^n$ (próbki w odwrotnej kolejności)

$$u = -Q_c^{-1} \cdot A_D^n x_0 \quad (8.3)$$

Jest to jednocześnie przepis na kształt sterowania programowego w układzie otwartym (sterownik programowalny).

Sterowalność praktyczna dla sterowań z przestrzeni funkcji ograniczonych

Powyżej przedstawiono własność sterowalności. Z rozważań wynika, że cecha ta związana jest ze strukturą układu (macierze A i B), a nie z klasą sterowań. Jest to więc własność strukturalna, którą system nabywa w momencie jego konstrukcji. W praktyce może okazać się, że mimo posiadania cechy sterowalności, w danym układzie nie można osiągnąć pewnych stanów, gdyż nie mamy do dyspozycji wystarczająco dużych sygnałów sterujących. Poniżej pokażemy, jak wygląda obszar stanów początkowych sterowalnych do „zera” dla systemu X -sterowalnego, gdy na sygnały sterujące narzucone jest ograniczenie na amplitudę.

Dany jest X -sterowalny system z macierzą stanu typu Jordana ($\lambda_i \neq \lambda_j$) i $b_i \neq 0$ oraz sterowaniem skalarnym

$$\dot{x}(t) = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{bmatrix} x(t) + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} u(t).$$

Przestrzeń sterowań ograniczonych $U = \{u(t) : |u(t)| \leq M; t \in (0, \infty)\}$.

Rozwiążanie dla każdej składowej stanu ma postać

$$x_i(t) = e^{\lambda_i t} \left[x_{i0} + \int_0^t e^{-\lambda_i \tau} b_i u(\tau) d\tau \right].$$

Założymy, że istnieje sterowanie przeprowadzające stan x_0 do $x(T) = 0$ w czasie T , wtedy

$$x_{i0} = - \int_0^T e^{-\lambda_i \tau} b_i u(\tau) d\tau.$$

Zbiór takich warunków początkowych, które można sprowadzić do zera, można oszacować

$$\begin{aligned} |x_{i0}| &= \left| \int_0^T e^{-\lambda_i \tau} b_i u(\tau) d\tau \right| \leq \int_0^T |e^{-\lambda_i \tau} b_i u(\tau)| d\tau \leq \\ &\leq \int_0^T e^{-\lambda_i \tau} |b_i| |u(\tau)| d\tau \leq M \int_0^T e^{-\lambda_i \tau} |b_i| d\tau = \frac{|b_i| M}{\lambda_i} (1 - e^{-\lambda_i T}). \end{aligned}$$

Czy powyższa nierówność może być spełniona dla dowolnie dużych $|x_0|$?

Pojawiają się dwa przypadki związane ze znakiem λ_i :

- 1) gdy $\lambda_i \leq 0$ (system stabilny), to dla dowolnie dużych $|x_0|$ zawsze istnieje odpowiednie $T < \infty$, że nierówność będzie spełniona;
- 2) gdy $\lambda_i > 0$ (system niestabilny), to dla $T \rightarrow \infty$ otrzymuje się maksymalne oszacowanie zbioru warunków początkowych sprowadzalnych do zera, $|x_{i0}| < \frac{|b_i| M}{\lambda_i}$.

A więc mimo teoretycznej X -sterowalności, system niestabilny można doprowadzić do zera układu tylko dla ograniczonego zbioru warunków początkowych. Obszar ten można powiększyć, zwiększając dopuszczalną amplitudę sterowania lub zmieniając parametry obiektu [8KLJ1, 8KLJ2, 8KAT].

8.2. Stabilizowalność

Właściwości sterowalności i stabilności układu dynamicznego składają się na jego nową cechę zwaną stabilizowalnością układu. Ponieważ podstawowym celem układu regulacji jest stabilizacja położenia wektora stanu w punkcie równowagi, to należy rozstrzygnąć, czy jest to możliwe bądź przez aktywne sterowanie, bądź przez swobodne zachowanie się systemu z powodu naturalnej asymptotycznej stabilności rozwiązania. W tym celu dokonuje się dekompozycji przestrzeni stanu X na sumę prostą dwóch podprzestrzeni: asymptotycznej stabilności X_{as} i jej dopełnienia tworzonego przez podprzestrzeń niestabilności i podprzestrzeń stabilności nieasymptotycznej. Takie dopełnienie można nazwać podprzestrzenią, która nie jest asymptotycznie stabilna X_{nas} . Przestrzeń asymptotycznej stabilności jest generowana przez wektory własne stwarzające z wartościami własnymi o ujemnych czę-

ciach rzeczywistych, a przestrzeń X_{nas} przez wektory własne stwarzające z wartościami własnymi o nieujemnych częściach rzeczywistych.

$$X_{as} \oplus X_{nas} = R^n.$$

Def. 8.5. System nazywamy stabilizowalnym, jeżeli zachodzi inkluza $X_{nas} \subset X_{ster}$, tzn. podprzestrzeń, która nie jest asymptotycznie stabilna zawiera się w podprzestrzeni sterowalności, lub $X_{niester} \subset X_{as}$, tzn. podprzestrzeń niesterowalności zawiera się w podprzestrzeni asymptotycznej stabilności.

Z tej definicji wynikają dwa twierdzenia:

Twierdzenie 8.5

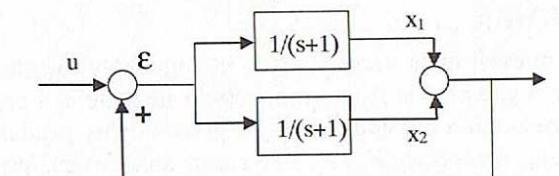
Każdy system liniowy stacjonarny asymptotycznie stabilny jest stabilizowalny.

Twierdzenie 8.6

Każdy system liniowy stacjonarny w pełni sterowny jest stabilizowalny.

Przykład 8.5

Dany jest układ przedstawiony na rysunku 8.6.



Rys. 8.6. Schemat blokowy

Równanie stanu

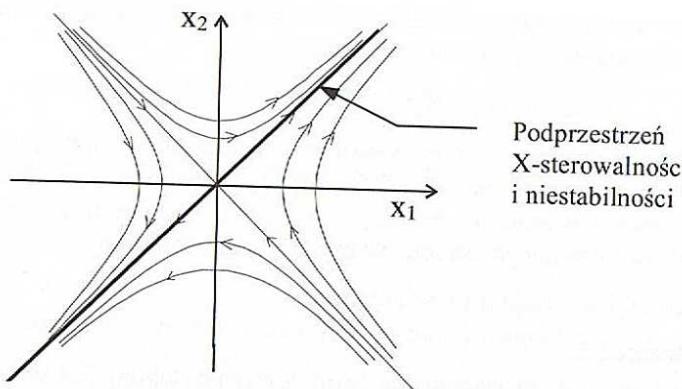
$$\begin{cases} \dot{x}_1(t) = -x_1(t) + \varepsilon(t) \\ \dot{x}_2(t) = -x_2(t) + \varepsilon(t) \\ \varepsilon(t) = x_1(t) + x_2(t) + u(t) \end{cases} \Rightarrow \begin{cases} \dot{x}_1(t) = x_2(t) + u(t), \\ \dot{x}_2(t) = x_1(t) + u(t), \\ y(t) = x_1(t) + x_2(t). \end{cases}$$

Macierze A, B, C :

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}; \quad B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}; \quad C = \begin{bmatrix} 1 & 1 \end{bmatrix}; \quad \Rightarrow \quad Q_C = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Ponieważ rząd macierzy Q_C jest równy jeden, to układ jest X -niesterowalny i podprzestrzeń sterowalności jest generowana przez wektor $w = [1 \ 1]^T$. Jednocześnie układ jest Y -sterowny.

Wartości własne macierzy A : $\lambda_1 = -1, \lambda_2 = 1$ (rzeczywiste o różnych znakach) dają obraz siodła. Odpowiadające im wektory własne: $w_1 = [1 \ -1]'$ i $w_2 = [1 \ 1]'$ generują odpowiednio przestrzeń stabilności i niestabilności. Widać to na rysunku 8.7. Podprzestrzeń niestabilności pokrywa się z podprzestrzenią sterowalności, czyli system jest stabilizowalny.

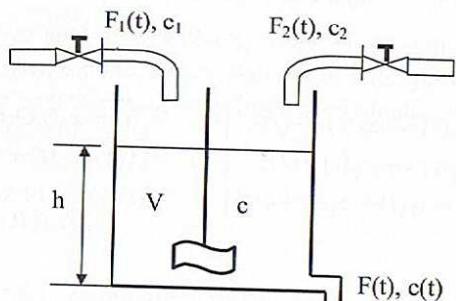


Rys. 8.7. Płaszczyzna stanu

W transmitancji $G(s)$ występuje skrócenie ujemnego zera i bieguna $G(s) = \frac{2(s+1)}{(s+1)(s-1)}$ stąd niesterowalność tego modu nie wpływa na stabilizowalność.

Przykład 8.6 ([8KWH])

Dany jest reaktor mieszalnikowy (rys. 8.8), do którego wpływają dwie mieszanki. Jedna o stałym stężeniu c_1 i przepływie $F_1(t)$ oraz druga o stężeniu c_2 i przepływie $F_2(t)$. Po natychmiastowym wymieszaniu następuje wypływ grawitacyjny produktu o stężeniu $c(t)$ i natężeniu $F(t)$. Traktując przepływy F_1 i F_2 jako sterowania, a $V(t)$, lub $h(t)$ oraz $c(t)$ jako zmienne stanu, napisać równanie stanu i zbadać warunki sterowalności.



Rys. 8.8. Schemat reaktora mieszalnikowego

Równania ciągłości całej masy składników i składnika lżejszego

$$\dot{V}(t) = F_1(t) + F_2(t) - F(t),$$

$$\frac{d}{dt}[V(t)c(t)] = F_1(t)c_1 + F_2(t)c_2 - F(t)c(t).$$

Wpływ zmiennych stanu w drugim równaniu powoduje, że równanie jest nieliniowe. Wypływ grawitacyjny daje dodatkową nieliniowość $F(t) = k\sqrt{h(t)}$. Zależności w stanie ustalonym: $h = \text{const}$, $c = \text{const}$, definiują punkt równowagi

$$F_{10} + F_{20} - F_0 = 0; \quad F_0 = k\sqrt{\frac{V_0}{s}}, \quad \frac{V_0}{F_0} = \Theta, \quad c_1 F_1 - F_2 c_2 - c_0 F = 0.$$

Model liniowy ze zmiennymi odchyłkowymi:

$$F_1(t) = F_{10} + u_1(t), \quad c(t) = c_0 + x_2(t),$$

$$F_2(t) = F_{20} + u_2(t), \quad y_1(t) = F(t) - F_0,$$

$$V(t) = V_0 + x_1(t), \quad y_2(t) = c(t) - c_0,$$

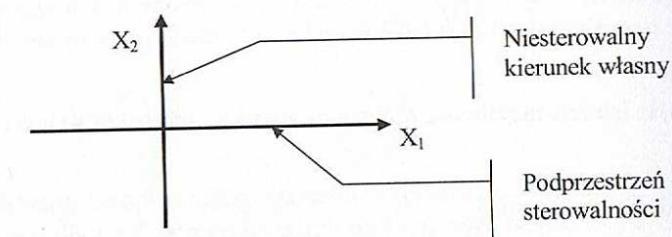
otrzymuje się, dokonując linearyzacji szeregiem Taylora wokół punktu równowagi.

$$\dot{x} = \begin{bmatrix} -\frac{1}{2\Theta} & 0 \\ 0 & -\frac{1}{\Theta} \end{bmatrix} x + \begin{bmatrix} 1 & 1 \\ \frac{c_1 - c_0}{V_0} & \frac{c_2 - c_0}{V_0} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}; \quad y = \begin{bmatrix} \frac{1}{2\Theta} & 0 \\ 0 & 1 \end{bmatrix} x.$$

Sprawdzamy sterowalność

$$Q_c = [B \ AB] = \begin{bmatrix} 1 & 1 & -\frac{1}{2\Theta} & -\frac{1}{2\Theta} \\ \frac{c_1 - c_0}{V_0} & \frac{c_2 - c_0}{V_0} & -\frac{c_1 - c_0}{\Theta V_0} & -\frac{c_2 - c_0}{\Theta V_0} \end{bmatrix}.$$

Widać, że system jest niesterowalny, tylko gdy zachodzi symetria stężeń: $c_1 = c_2 = c_0$. Wtedy $Q_c = 1$. Podprzestrzeń sterowalności (rys. 8.9) jest generowana przez kolumnę macierzy $Q_c - [1, 0]^T$. Sterowaniem można wtedy wpływać tylko na poziom $h(t)$. System jest stabilizowalny.



Rys. 8.9. Płaszczyzna stanu

8.3. Obserwowałośc

Zwykle do pomiaru dostępny jest wektor wyjścia i na podstawie jego wartości buduje się układy regulacji. Bardziej wydajne wielowymiarowe regulatory stabilizujące pracują opierając się na pomiarze wektora stanu. Pełny wektor stanu nie zawsze jest dostępny do pomiaru. Zachodzi pytanie, czy jest możliwe odtworzenie wektora stanu na podstawie pomiarów wyjścia i ewentualnie wejścia układu sterowanego. Jeżeli taka możliwość istnieje, to mówimy, że układ jest układem obserwowały w względem stanu. W przypadku liniowych układów dynamicznych przy badaniu podstawowego warunku obserwowałości można bez straty ogólności założyć, że sterowania są zerowe. Warunek ten zaprezentujemy więc dla znanego ciągłego, liniowego, stacjonarnego układu swobodnego (bez sterowania)

$$\begin{aligned}\dot{x}(t) &= A x(t) \\ y(t) &= C x(t)\end{aligned}\quad (8.4)$$

gdzie dla $\forall t \geq 0 : x(t) \in R^n$, $y(t) \in R^m$.

Trudności z odtworzeniem nieznanego bieżącego stanu $x(t)$ na podstawie bieżących pomiarów wyjścia $y(t)$ pojawią się w przypadku, gdy wymiar znanego wektora $y \in R^m$ jest mniejszy niż wymiar nieznanego wektora $x \in R^n$, $m < n$. Nie można wtedy odwrócić prostokątnej macierzy C i rozwiązywać na bieżąco równania $x(t) = C^{-1}y(t)$. Niektórzy autorzy sugerują użycie w takim przypadku rozwiązania minimalno-normowego (czyli składowej prostopadłej do tzw. podprzestrzeni nieobserwowałości) wyliczanej jako prawa pseudoodwrotność macierzy C , tzn. $x(t) = C^T [CC^T]^{-1}y(t)$, (podrozdz. 19.8 – aneks 1).

Taka metodyka wyboru estymaty stanu – jako najkrótszego wektora x , nie ma jednak uzasadnienia, gdyż taki wektor może znacznie się różnić od rzeczywistego stanu, a nawet dawać jego niedopuszczalne wartości, np. ujemne poziomy czy nierealizowalne przepływy (przykład 8.7). Jedyną drogą jest pozyskanie większej ilości informacji z dodatkowych pomiarów. W wersji ciągłej sprowadza się to do pomiaru ciągłego na odcinku czasu T i wykonaniu dodatkowych obliczeń.

Twierdzenie 8.7 (Warunek obserwowałości)

Jeżeli liczbową kwadratową ($n \times n$) i symetryczną macierz Grama M_0

$$M_0 = \int_0^T [Ce^{At}]^T Ce^{At} dt \quad (8.5)$$

jest nieosobliwa, to istnieje możliwość obliczenia stanu x_0 , a układ (8.4) jest obserwowy.

Dowód

Rozwiązanie (8.4) ma postać

$$y(t) = Ce^{At} x_0 \quad (8.6)$$

Mnożąc lewostronnie (8.6) przez macierz transponowaną $[C \exp(At)]'$, otrzymamy

$$[Ce^{At}]' y(t) = [Ce^{At}]' Ce^{At} x_0.$$

Tak otrzymana macierz funkcyjna $[[Ce^{At}]' Ce^{At}]$ ($n \times n$ -wymiarowa) jest kwadratowa i symetryczna, ale jej rząd jest mniejszy lub równy m , dla dowolnej wartości czasu t , jest więc też osobliwa. Można dokonać na obu stronach równania operacji całkowania w wybranych granicach $[0, T]$. Całkowanie reprezentuje dodatkową operację sumacyjną na oknie pomiarowym sygnału $y(t)$. Długość okna dla przypadku ciągłego, może być teoretycznie dowolna $T > 0$.

$$\int_0^T [Ce^{At}]' y(t) dt = \int_0^T [Ce^{At}]' Ce^{At} dt \cdot x_0 \quad (8.7)$$

Jeśli rząd $M_0 = n$, to z (8.7) można wyliczyć x_0

$$x_0 = M_0^{-1} \int_0^T [Ce^{At}]' y(t) dt = \int_0^T M_0^{-1} [Ce^{At}]' y(t) dt = \int_0^T G(T, t) y(t) dt \quad (8.8)$$

Jeśli macierz obserwowałości Grama M_0 jest osobliwa, to system jest nieobserwowany i nie istnieje możliwość odtworzenia dokładnej wartości pełnego wektora stanu x_0 .

Otrzymana funkcyjna macierz podcałkowa $G(T, t)$ ma wymiary $(n \times m)$ i może być polinomiona off-line w przedziale $[0, T]$, a następnie użyta do mnożenia z przebiegiem wyjścia $y(t)$ i całkowania.

$$\begin{bmatrix} x_{01} \\ \vdots \\ x_{0n} \end{bmatrix} = \int_0^T G(T, t) y(t) dt = \int_0^T \begin{bmatrix} G_{11}(t) & \cdots & G_{1m}(t) \\ \vdots & \ddots & \vdots \\ G_{n1}(t) & \cdots & G_{nm}(t) \end{bmatrix} \begin{bmatrix} y_1(t) \\ \vdots \\ y_m(t) \end{bmatrix} dt.$$

Macierz $G(T, t)$ jest funkcją czasu t i przyjętego parametru, jakim jest szerokość przedziału obserwacji T . Widać, że możliwość odtworzenia dokładnej wartości x_0 nie zależy od przyjętej wartości T . Teoretycznie może być ona więc dowolnie mała, co ma znaczenie przy dążeniu do zmniejszenia ilości obliczeń numerycznych związanych z przedziałem całkowania.

Po obliczeniu stanu x_0 można go użyć do obliczania wartości stanu $x(t)$ dla $t > T$ ze wzoru

$$x(t) = e^{At} x_0 \quad (8.9)$$

Taki obserwator po odtworzeniu wartości x_0 (po czasie T) nie śledzi już dalszego przebiegu wyjścia $y(t)$ dla $t > T$ i generuje dalszy stan $x(t)$ tylko na podstawie modelu (8.9). Nie gwarantuje to jego wiarygodności, gdyż stan rzeczywisty z powodu zakłóceń może być już inny.

Lepszą wersją do odtwarzania stanu on-line dla $t > T$ jest obserwator całkowy w postaci ruchomego okna o szerokości T przesuwającego się na tle pomiarów $y(t)$ i wykonującego szybką akcję całkowania na bieżąco iloczynu $G(t)y(t)$ w oknie pomiarowym $[t-T, t]$, gdzie t jest czasem bieżącym, a $t-T$ czasem początkowym (wymagane szybkie techniki cyfrowe).

$$x(t) = e^{A(t-t_0)} x(t_0) = e^{AT} x(t-T) = e^{AT} \int_{t-T}^t G(T, \tau) y(\tau) d\tau \quad (8.10)$$

Prostsze niż warunek obserwowałości (8.5) jest algebraiczne kryterium obserwowałości układu (8.4), oparte na analizie rzędu macierzy obserwowałości Q_o .

Twierdzenie 8.8

Dla zapewnienia obserwowałości (8.4) rzad prostokątnej macierzy Q_o , o $m \times n$ wierszach, który równa się rzędowi macierzy Grama M_0 , musi też równać się n , gdzie m – liczba wyjść y .

$$Q_o = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} \quad (8.11)$$

Dla $m > 1$ dotyczy to również macierzy $Q_o = [C \quad CA \quad \dots \quad CA^{n-m}]'$.

Dowód

Warunek rzad $Q_o = n$ gwarantuje, że zerowanie się wyjścia $y(t)$ układu, na skończonym odcinku czasu (a więc też jego pochodnych $y^{(i)}(t)$), jest spowodowane zerowymi warunkami początkowymi stanu $x_0 = 0$. Na przykład dla układu z $m = 1$, zachodzi bowiem:

$$\left. \begin{array}{l} y(t) = Ce^{At} x_0 \\ \dot{y}(t) = CAe^{At} x_0 \\ \ddot{y}(t) = CA^2 e^{At} x_0 \\ \vdots \\ y^{(n-1)}(t) = CA^{n-1} e^{At} x_0 \end{array} \right\} \begin{array}{l} \text{Ponieważ } e^{At} \neq 0; \text{ to z warunku} \\ 0 = Q_o e^{At} x_0 \text{ wynika, że tylko, jeśli} \\ \text{rzad } Q_o = n, \text{ to jest gwarancja, że } x_0 = 0. \end{array}$$

Potwierdza to uwagę, że obserwowałość jest cechą strukturalną (decydującą tylko macierze A i C) i nie zależy od sterowania, ani od długości przyjętego przedziału obserwacji T .

Odwzorowanie $x_0 \rightarrow y$, zapisane ogólnie jako $\Psi: R^n \rightarrow F(R, R^m)$, wynika ze wzoru (8.6) $y = Ce^{At} x_0 = \Psi x_0$. Aby było możliwe odwzorowanie x_0 z odcinka funkcji y , to jądro (kernel) odwzorowania musi być zerowe $\ker \Psi = 0$.

$$\ker \Psi = \ker C \cap \ker CA \cap \dots \cap \ker CA^{n-1} = \ker [C \quad CA \quad \dots \quad CA^{n-1}]'$$

Podprzestrzeń zerowa odwzorowania Ψ generowana przez ker Ψ jest nazywana podprzestrzenią stanów nieobserwowań i jest oznaczana $N(M_0)$, gdyż pokrywa się z podprzestrzenią zerową macierzy Grama M_0 (8.5).

Stan jest nieobserwalny, gdy należy do podprzestrzeni zerowej $N(M_0)$. System jest obserwowywalny, jeśli oprócz „zerowego” stanu nie istnieją w nim inne stany nieobserwowań.

W literaturze przedmiotu przyjęto się nazywać podprzestrzenią stanów obserwowań, ortogonalną do podprzestrzeni stanów nieobserwowań, podprzestrzeń dopełniającą do góry R^n , $N(M_0)^\perp = R(M_0)$. Dekompozycja całej przestrzeni stanu $R^n = N(M_0) \oplus R(M_0)$ oznacza, że każdy stan $x \in R^n$ może być jednoznacznie przedstawiony przez składową obserwowań x_{ob} i nieobserwowań x_{nob} , która jest \perp do x_{ob} . Odpowiednie przetworzenie informacji na wyjściu y pozwala określić tylko umowną składową „obserwowań” stanu początkowego. Umowność dotyczy wyboru przestrzeni dopełniającej. Ortogonalna „obserwowań” podprzestrzeń dopełniająca jest łatwo identyfikowalna, gdyż jest generowana przez liniowo niezależne wektory wierszowe macierzy obserwowałości Q_o .

Twierdzenie 8.9

Jeżeli macierz charakterystyczna utworzona dla jakiejś wartości własnej λ_k i uzupełniona macierzą C , ma rzad mniejszy od n

$$\text{rzad } \begin{bmatrix} \lambda_k I - A \\ C \end{bmatrix} < n, \text{ dla } k \in [1, \dots, n],$$

to dla tej wartości własnej λ_k istnieje nieobserwowywany wektor własny (mod).

Dowód twierdzenia w pracy [8CAF].

Przykład 8.7

Dany jest układ o transmitancji jak w przykładzie 8.4, ale w sterowej wersji przejęteli stanu.

$$\begin{aligned} \dot{x}(t) &= \begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t) \Rightarrow G(s) = \frac{s-2}{(s-2)(s+1)} = \frac{1}{s+1}. \\ y(t) &= [-2 \quad 1] x(t) \end{aligned}$$

Widac, że w transmitancji występują skrócenia. Układ nie powinien być obserwowywalny (8.11).

$$Q_o = \begin{bmatrix} C \\ CA \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 2 & -1 \end{bmatrix}, \text{ rzad } Q_o = 1.$$

Zgodnie z twierdzeniem 8.9:

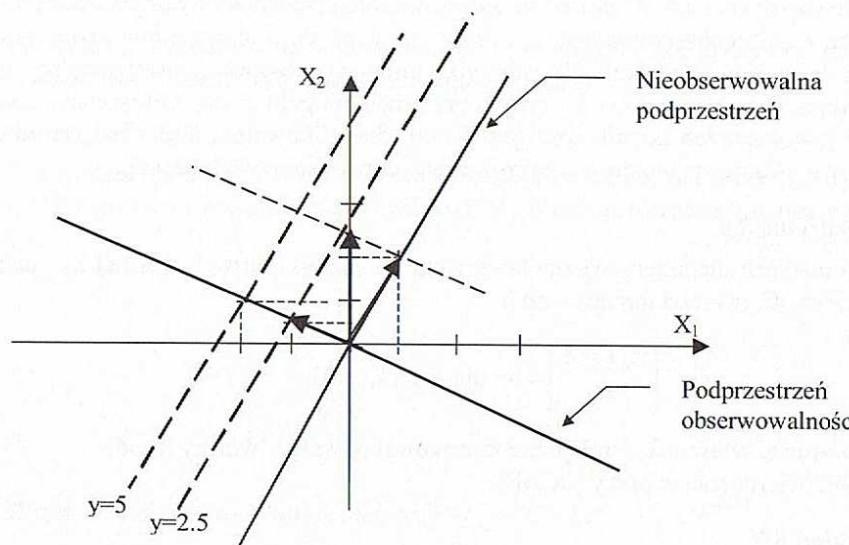
$$\text{dla } \lambda_1 = 2, \text{ rzad } \begin{bmatrix} 2 & -1 \\ -2 & 1 \end{bmatrix} = 1, \text{ a dla } \lambda_2 = -1, \text{ rzad } \begin{bmatrix} -1 & -1 \\ -2 & -2 \\ -2 & 1 \end{bmatrix} = 2.$$

Nieobserwowywany jest mod (wektor własny) związany z wartością własną $\lambda_1 = 2$.

Ponieważ A jest postaci Frobeniusa, to wektory własne są postaci

$$T = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}.$$

Kierunek $[-2, 1]'$ według Q_o , daje podprzestrzeń „obserwowałości”. A kierunek prostopadły $[1, 2]'$ daje podprzestrzeń „nieobserwowałości” (rys. 8.10).



Rys. 8.10. Płaszczyzna stanu

„Obserwowało” rozwiążanie minimalnonormowe $x_o = C^T [CC^T]^{-1}y$, np. dla $y = 2,5$ równa się $x_o = [-1, 0,5]^T$. Jeśli przykładowo składowa wektora stanu nieobserwowało będzie miała wartość $x_n = [1, 2]^T$ to rzeczywisty wektor stanu ma wartość $x = [-1+1, 0,5+2]^T = [0, 2,5]^T$ i różni się znacznie od składowej obserwowałojej x_o . Łatwo zauważyć, że dla $y = 5$ minimalnonormowa składowa „obserwowało” równa się $x_o = [-2, 1]^T$ i jest to wiersz macierzy obserwowałości Q_o , z którego buduje się podprzestrzeń obserwowałości.

Przykład 8.8

Rozważmy układ oscylatora harmonicznego:

$$\ddot{y}(t) + y(t) = 0,$$

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}, \quad y = [1, 0] \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}.$$

Postawmy zadanie: należy odtworzyć stan początkowy x_0 z chwilą $t = 0$, znając odpowiedź w przedziale $t \in [0, \pi/2]$ o postaci $y(t) = \cos t + \sin t$.

System jest obserwowały, bo rząd macierzy Q_o jest 2. Należy obliczyć M_0 i x_0 dla $T = \pi/2$.

$$Q_o = \begin{bmatrix} C \\ CA \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

$$e^{At} = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix}, \quad (Ce^{At})' Ce^{At} = \begin{bmatrix} \cos^2 t & \cos t \sin t \\ \cos t \sin t & \sin^2 t \end{bmatrix},$$

$$M_o = \frac{1}{2} \begin{bmatrix} \pi/2 & 1 \\ 1 & \pi/2 \end{bmatrix}, \quad M_0^{-1} = \frac{16}{\pi^2 - 4} \begin{bmatrix} \pi/4 & -1/2 \\ -1/2 & \pi/4 \end{bmatrix},$$

$$x_0 = M_0^{-1} \int_0^{\pi/2} [Ce^{At}]' y(t) dt =$$

$$= M_0^{-1} \int_0^{\pi/2} \begin{bmatrix} \cos t \\ \sin t \end{bmatrix} (\cos t + \sin t) dt = M_0^{-1} \begin{bmatrix} \pi/4 + 0,5 \\ \pi/4 + 0,5 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Łatwo wyliczyć, że tak znaleziony stan początkowy sprawdza założone równanie wyjścia

$$y(t) = Ce^{At} x(0) = [1, 0] e^{At} [1, 1]' = \cos t + \sin t.$$

8.4. Niezmienność sterowalności i obserwowałości

Jak to przedstawiono na rysunkach 8.1 i 8.2, system będzie w pełni sterowalny względem stanu, gdy nie wystąpią bloki niesterowalne G_{nso} , G_{nsno} i będzie obserwowały, gdy nie wystąpią bloki nieobserwowały G_{sno} , G_{nsno} . Gdy w systemie występuje tor wejście-wyjście, to na pewno system posiada transmitancję reprezentującą blok sterowalny i obserwowały. W praktyce więc głównie spotyka się trzy typy obiektów regulacji, o których można ogólnie powiedzieć, że są:

- 1) sterowalne i obserwowały,
- 2) sterowalne ale nieobserwowały (może wystąpić obserwowałość w podprzestrzeni),
- 3) obserwowały ale niesterowalne (może wystąpić częściowa sterowalność w podprzestrzeni),

W kanonicznej strukturze każdego z trzech typów można wyodrębnić odpowiednio podsystemy:

- 1) podsystem tylko sterowalny i obserwowały,
- 2) podsystem sterowalny i obserwowały + sterowalny ale nieobserwowały,
- 3) podsystem sterowalny i obserwowały + obserwowały ale niesterowalny.

W jednowymiarowych transmitancjach opisujących dwa ostatnie systemy na pewno wystąpią skrócenia. Wtedy projektant ma do wyboru taką postać przestrzeni stanu, która jest wygodniejsza dla celów projektu, albo typu 2) – wtedy układ będzie X -sterowalny, ale nieobserwowały, albo typu 3) – wtedy układ będzie X -niesterowalny, ale obserwowały.

Na fakt ten była już zwracana uwaga w rozdziale 3, gdzie pogrupowano wzory na równanie stanu dla sterowalnych przestrzeni stanu (3.17), (3.18) i (3.19) i obserwowańnych przestrzeni stanu (3.20)–(3.22). Jeśli w transmitancji nie występują skrócenia i cały układ jest sterowany i obserwowywalny, to zarówno zestawy macierzy A , B , C dane wzorami (3.17)–(3.19), jak i wzorami (3.20)–(3.22) będą sterowalne i obserwowańne. Warto również zwrócić uwagę, że decydując się na któryś z przestrzeni stanu typu 2 lub 3) nie można już zmienić jej własności sterowalności i obserwowańności poprzez liniowe nieosobliwe przekształcenie podobieństwa tych zmiennych stanu $x(t)$ na inne zmienne $z(t)$, tzn. $x(t) = Tz(t)$. Sterowalność i obserwowałość są bowiem tak jak i stabilność niezmiennikami przekształceń liniowych.

8.5. Zachowanie sterowalności i obserwowańności po dyskretyzacji układu ciągłego

Twierdzenie 8.10 [3OGK]

System jednowymiarowy liniowy, ciągły zachowuje swoją sterowalność i obserwowałość po dyskretyzacji z okresem dyskretyzacji T wtedy i tylko wtedy, gdy dla każdej pary pierwiastków zespolonych takich, że $\operatorname{Re}(s_i) = \operatorname{Re}(s_j)$, zachodzi nierówność

$$\operatorname{Im}(s_i - s_j) \neq m2\pi/T, \quad m = \pm 1, \pm 2, \dots$$

Fakt ten wiąże się ze skracaniem biegunów i zer transmitancji dyskretnej, którą otrzymuje się z transmitancji ciągłej zawierającej biegony zespolone. Dla biegonów rzeczywistych s_i , problem ten nie występuje.

Dla układów wielowymiarowych spełnienie tej nierówności jest warunkiem koniecznym zachowania sterowalności.

Przykład 8.9

Zbadać sterowalność i obserwowałość układu oscylatora harmonicznego ciągłego z $\omega = 50^\circ = (2\pi \cdot 50/360) = \pi/3,6$, przed i po jego dyskretyzacji z krokiem $T = 7,2$.

$$G(s) = \frac{\omega^2}{s^2 + \omega^2}, \quad s_1 = j\omega, \quad s_2 = -j\omega,$$

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ -\omega^2 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ \omega^2 \end{bmatrix} u(t),$$

$$y(t) = [1 \ 0] x(t).$$

Łatwo sprawdzić, że rzad $Q_c = 2$ i rzad $Q_o = 2$, stąd system ciągły jest sterowany i obserwowany.

Dla zachowania sterowalności po dyskretyzacji musi zachodzić warunek

$$\operatorname{Im}(s_i - s_j) = 2\omega \neq m2\pi/T, \quad \text{czyli } T \neq m\pi/\omega.$$

Nie zachodzi to dla przyjętych parametrów i $m = 2$, bo $7,2 = m \cdot \pi / (2\pi \cdot 50/360) = m \cdot 3,6 = 2 \cdot 3,6$.

Można to również sprawdzić, licząc model dyskretny. Po dyskretyzacji według wzoru z podrozdziału 5.3.1, dla układu z przykładu 5.2. otrzymuje się:

$$A_D = e^{AT} = \begin{bmatrix} \cos(\omega T) & \sin(\omega T)/\omega \\ -\omega \sin(\omega T) & \cos(\omega T) \end{bmatrix}, \quad B_D = \begin{bmatrix} 1 - \cos(\omega T) \\ \omega \sin(\omega T) \end{bmatrix}.$$

Macierz sterowalności układu dyskretnego

$$Q_{DC} = \begin{bmatrix} 1 - \cos(\omega T) & -2\cos^2(\omega T) + \cos(\omega T) + 1 \\ \omega \sin(\omega T) & 2\omega \cos(\omega T) \sin(\omega T) - \omega \sin(\omega T) \end{bmatrix}.$$

Przyrównanie wyznacznika do zera daje po obliczeniach warunek: $\sin(\omega T)[\cos(\omega T) - 1] = 0$. Stąd $\Rightarrow \omega T = m\pi$, $m = \pm 1, \pm 2, \dots$ Czas dyskretyzacji T powinien być różny od

$$T \neq m\pi/\omega \neq m\pi \cdot 3,6/\pi = 3,6, \quad 7,2, \quad 10,8 \quad \text{itd.}$$

W rozważanym przypadku $T = 7,2$, stąd układ po dyskretyzacji będzie niesterowany. Ten sam warunek otrzymuje się z analizy transmitancji. Omawiany oscylator harmoniczny z członem ZOH posiada transmitancję dyskretną $G(z)$ (tab. 5.1)

$$G(z) = C_D [zI - A_D]^{-1} B_D = \frac{(z+1)(1-\cos\omega T)}{z^2 - 2z\cos\omega T + 1}$$

$$\text{Biegony mianownika } z_{1,2} = \cos\omega T \pm \sqrt{(\cos^2\omega T - 1)}.$$

Aby nie zachodziło skracanie zera i bieguna, musi zachodzić warunek

$$\cos\omega T \pm \sqrt{\cos^2\omega T - 1} \neq -1 \Rightarrow (\cos^2\omega T - 1) \neq (1 + \cos\omega T)^2 \Rightarrow \cos\omega T \neq -1 \Rightarrow \omega T \neq (2m+1)\pi, \quad m = 0, \pm 1, \pm 2, \dots$$

Również z licznika transmitancji wynika, że powinno zachodzić

$$\cos\omega T \neq 1 \Rightarrow \omega T \neq 2m\pi, \quad m = 0, \pm 1, \pm 2, \dots$$

Ostatecznie otrzymuje się ten sam warunek: $\omega T \neq m\pi$, $m = 0, \pm 1, \pm 2, \dots$

8.6. Asymptotyczne różniczkowe estymatory stanu

Idea obserwatora całkowego odtwarzającego dokładnie stan $x(T)$, rozbudowana o po-miar sterowania daje ogólną strukturę obserwatora o postaci

$$x(T) = \int_0^T G_1(\tau) y(\tau) d\tau + \int_0^T G_2(\tau) u(\tau) d\tau.$$

Wzór ten może być zmodyfikowany do zastosowań on-line podobnie jak w (8.10).

Obserwator taki dla czasów $t \geq T$ można potraktować jako przesuwane na tle bieżących pomiarów okno o stałej długości T

$$x(t) = \int_0^T G_1(\tau) y(t-T+\tau) d\tau + \int_0^T G_2(\tau) u(t-T+\tau) d\tau \quad (8.12)$$

Formuła (8.12) ma głębsze uzasadnienie, które pokazuje, że obustronne całkowanie we wzorze (8.6) nie było przypadkowe. Konieczność takiej postaci ogólnego obserwatora dokładnego wynika z faktu, że należy odtworzyć wektor liczbowy $x \in R^n$ reprezentujący stan, na podstawie dostępnych odcinków funkcji y i u . Taka operacja musi być funkcjonalnym liniowym i ciągłym. Z twierdzenia Frécheta–Riesza o reprezentacji w przestrzeni L^2 (aneks 2) wynika, że w przestrzeniach funkcyjnych $L^2[0, T]$ każdy funkcjonal liniowy i ciągły (a więc ograniczony) operujący na jakimś elemencie tej przestrzeni np. $y \in L^2[0, T]$ jest przedstawialny jako iloczyn skalarny tego elementu i specjalnie dobranego innego elementu tej przestrzeni, np. $G_1 \in L^2[0, T]$. Iloczyn skalarny w L^2 jest reprezentowany przez całkę z iloczywu tych funkcji. Wynika z tego, że w przestrzeni L^2 nie istnieje inna niż całkowa postać obserwatora dokładnego. Dla spełnienia (8.12), można użyć wielu różnych funkcji $G(t)$. Można więc znaleźć optymalne w jakimś sensie macierze obserwacji $G_1(\tau), G_2(\tau)$ [8BYW]. Realizacja algorytmu obserwacji on-line na obecnym poziomie techniki komputerowej nie sprawia trudności.

W latach 60. XX w. algorytmy te nie mogły być jednak stosowane i dlatego do celów odtwarzania stanu on-line w tym czasie została zaproponowana struktura asymptotycznego obserwatora (estymatora stanu), oparta na liniowych zwyczajnych równaniach różniczkowych, które można było szybko rozwiązywać rekurencyjnie.

Gdy znany jest system ciągły (macierze A, B, C) i macierz A jest asymptotycznie stabilna, strukturę obserwatora asymptotycznego można łatwo otrzymać przez symulację zachowania się modelu systemu, pobudzanego tym samym sterowaniem co system rzeczywisty.

Oznaczając stan rzeczywisty przez $x(t)$ a jego estymatę przez $\bar{x}(t)$, otrzymamy równania:

modelu: $\dot{x}(t) = Ax(t) + Bu(t), \quad x(0);$

estymatora: $\dot{\bar{x}}(t) = A\bar{x}(t) + Bu(t), \quad \bar{x}(0).$

Podstawowym problemem dla odtwarzania wektora $x(t)$ przez $\bar{x}(t)$ jest nieznajomość rzeczywistych warunków początkowych $x(0)$. Dla przyjętej początkowej estymaty $\bar{x}(0) \neq x(0)$ dynamika błędu obserwacji (szybkość zbieżności) zależy od rozkładu wartości własnych macierzy A , bo od eksponenty $\exp(At)$, zależy szybkość zanikania części odpowiedzi zależnej od warunków początkowych. Część zależna od sterowania jest dla obu równań taka sama. Dla obiektów niestabilnych ta metoda zawodzi.

Dla rozwiązania tego problemu i zwiększenia szybkości zbieżności można zastosować dodatkowy pomiar wyjścia $y(t)$ i macierz wzmacnianie G . Aby zagwarantować asymptotycz-

zbieżność odtwarzania stanu należy określić warunki na zmodyfikowaną macierz stanu obserwatora F .

Oznaczając błąd obserwacji przez $\varepsilon(t)$, otrzymujemy zależności na obserwator Luenbergera:

$$\varepsilon(t) = x(t) - \bar{x}(t),$$

$$\dot{\bar{x}}(t) = F \bar{x}(t) + B u(t) + G y(t), \quad \bar{x}(0) = \bar{x}_0,$$

$$\dot{x}(t) - \dot{\bar{x}}(t) = (A - GC)x(t) - F\bar{x}(t),$$

$$\dot{\varepsilon}(t) = F\varepsilon(t) \iff F = A - GC,$$

$$\lim_{t \rightarrow \infty} \varepsilon(t) \rightarrow 0 \iff F \text{ będzie asymptotycznie stabilna.}$$

Ostatecznie obserwator ma postać

$$\dot{\bar{x}}(t) = (A - GC)\bar{x}(t) + Bu(t) + Gy(t), \quad \bar{x}(0) = \bar{x}_0 \quad (8.13)$$

Wynika z tego, że jeśli uda się dobrać warunki początkowe estymacji równe rzeczywistym $\bar{x}(0) = x(0)$, to obserwator (8.13) dokładnie symuluje zachowanie się systemu rzeczywistego. W przeciwnym wypadku obserwatory różniczkowe generują estymatę $\bar{x}(t)$, która dopiero w granicy ($t = \infty$) zapewnia równość $\bar{x}(t) = x(t)$.

Jeśli system (A, C) jest obserwowałny, wtedy wartości własne macierzy F mogą być ulokowane w dowolnym miejscu płaszczyzny zespolonej. Oznacza to, że przez wybór macierzy G można zmieniać dynamikę nadążania obserwatora Luenbergera i w praktyce przepieszać zbieżność estymaty stanu, choć teoretycznie równość stanu i estymaty zawsze wystąpi dopiero w nieskończoności. Stąd dla obserwatora typu Luenbergera stosuje się określenie „obserwatory asymptotyczne”. Macierz wzmacnianie G dobiera się przez lokowanie wartości własnych macierzy F , tzn. dla zadanych wartości własnych macierzy F otrzymuje się liczbowe współczynniki wielomianu charakterystycznego, które równocześnie są funkcjami współczynników macierzy G . Rozwiązuując odpowiednią liczbę równań, znajduje się G . Nie zmienia to faktu, że gdy stosuje się obserwator Luebergera, teoretycznie nigdy nie można stwierdzić, jaka jest wartość wektora stanu i jak daleko od niego jest jego bieżąca estymata.

Warto odnotować wyniki badań w dziedzinie dokładnej rekonstrukcji stanu, również dla układów singularnych z osobliwą macierzą E .

$$E\dot{x}(t) = Ax(t) + Bu(t), \quad x(0) = x_0,$$

$$y(t) = Cx(t),$$

gdzie macierze $E, A \in R^{n \times n}$ i $E \neq 0, \det(E) = 0$.

Twierdzenie 8.11 [8KAT]

Warunkiem koniecznym i wystarczającym na to, aby obserwator singularny o postaci różniczkowej

$$E\dot{\bar{x}}(t) = (A + GC)\bar{x}(t) + Bu(t) - Gy(t), \quad \bar{x}(0) = \bar{x}_0,$$

odtwarzał stan dokładnie

$$\bar{x}(t) = x(t), \quad t > 0,$$

jest, aby

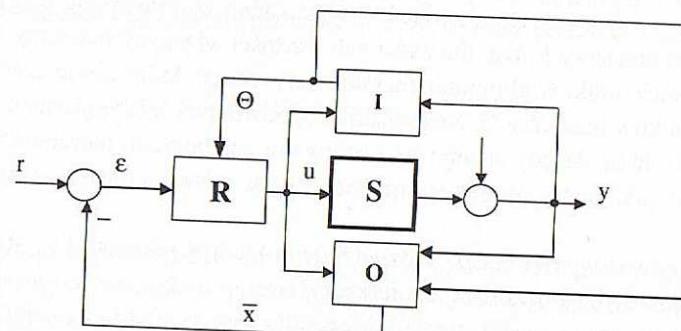
$$\text{rząd} \begin{bmatrix} E \\ C \end{bmatrix} = n, \quad \text{rząd} \begin{bmatrix} Es - A \\ C \end{bmatrix} = n \quad \text{dla } \forall s.$$

Wtedy bowiem tylko zachodzi specjalna redukcja wielomianu charakterystycznego odpowiedniej macierzy wielomianowej, do liczby α .

$$\det[Es - (A + GC)] = \alpha \neq 0.$$

Obserwatory stanu i identyfikatory parametrów tworzą nową klasę urządzeń (algorytmów), które obok regulatorów, korektorów dynamiki, filtrów sygnałów, członów wykonawczych, przetworników sygnałów i czujników składają się na aparaturę automatyki. Zostańą one szczegółowo omówione w rozdziale 12 i w następnych. Wymagają one dla swoich obliczeń dostarczenia wartości wejść i wyjść obiektu sterowanego.

Umiejscowienie dokładnego obserwatora stanu 0 lub estymatora stanu oraz identyfikatora I parametrów Θ obiektu S w układzie sterowania z regułatorem od stanu R , nastrajanym na podstawie parametrów Θ pokazano na rysunku 8.11. Taki zestaw stanowi jedną ze struktur pełnego regulatora adaptacyjnego (podrozdz. 6.3.3).

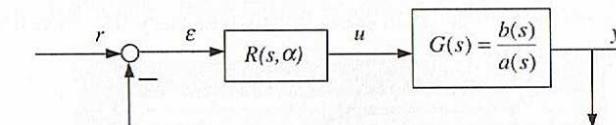


Rys. 8.11. Schemat blokowy adaptacyjnego układu sterowania od stanu

9. Kryteria i dobór optymalnych nastaw regulatorów PID

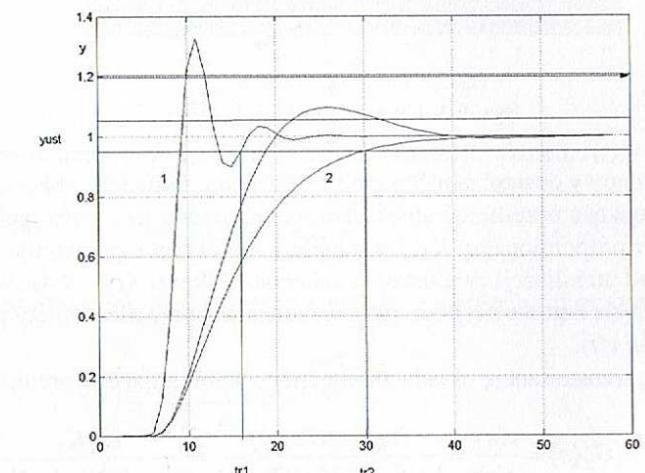
9.1. Podstawowe kryteria jakości

W zależności od dobranych parametrów α regułatora R (rys. 9.1) uzyskuje się różny kształt sygnału sterującego u , który jest sygnałem wejściowym do obiektu sterowanego G , a w następstwie różny kształt przebiegu czasowego regulowanej zmiennej wyjściowej procesu y .



Rys. 9.1. Schemat blokowy

Uzyskana dynamika odpowiedzi może być więc różna – szybka lub wolna (rys. 9.2). Aby ją ocenić jakościowo i ilościowo, należy wprowadzić pewne kryteria zwane w automatyce kryteriami jakości przebiegów regulacji.



Rys. 9.2. Przebiegi odpowiedzi $y(t)$

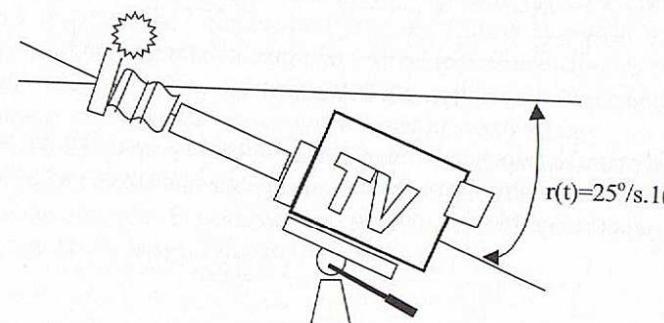
Objaśnienia w tekście

Typowymi wskaźnikami regulacji mogą być tzw.: „tłumienie do 1/4” (podrozdz. 4.6.1) lub czas regulacji t_r zwany też czasem T_{95} – gdyż jest to chwila czasu, w której przebieg $y(t)$ wchodzi w 5-procentowe otoczenie poziomu ustalonego $y_{ust} = y(\infty)$ i już go nie opuszcza $[0,95y_{ust} \div 1,05y_{ust}]$. Czas ten powinien być jak najkrótszy. W tym sensie dla pewnych procesów najlepsze będą parametry regulatora gwarantujące przebieg $y_1(t)$ na rysunku 9.2. Jeśli jednak dodatkowo na zmienną procesową będą narzucone ograniczenia technologiczne (np. maksymalna wartość temperatury, powyżej której nastąpi niepożądana polimeryzacja produktu reaktora chemicznego), to przebieg $y_1(t)$ będzie niedopuszczalny i najlepszym regulatorem gwarantującym minimum t_r będzie przebieg $y_2(t)$. Przebieg $y_3(t)$, jakkolwiek dopuszczalny, nie jest optymalny w tym sensie.

Omawiane kryterium t_r nie prowadzi jednak do formuł analitycznych na parametry regulatora. Już w najprostszym przypadkach muszą być stosowane metody numeryczne. Pokazane jest to na przykładzie sygnalizowanym w [9DOR], a rozwiązane do końca poniżej.

Przykład 9.1

Dla stabilizacji obrazu w ruchomych kamerach telewizyjnych i filmowych stosuje się układ elektryczno-mechaniczny ruchomego obiektywu Dynalens korygującego pionowe przesunięcia kadru wskutek wstrząsów kamery. Układ regulacji ma gwarantować nadążanie szybkości przesunięć obiektywu za szybkością korpusu kamery tak, by mimo drgań kamery obraz w kadrze był nieruchomy (rys. 9.3).



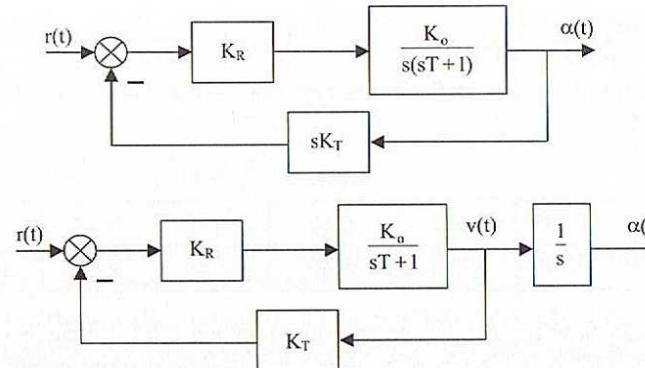
Rys. 9.3. Układ stabilizacji kamery

Układ żyroskopowy dostarcza informacji o skokowej zmianie szybkości ruchu korpusu kamery $r(t)$. Prędkość przesunięć obiektywu $v(t)$ mierzona jest przez tachometr. Stosowany jest regulator proporcjonalny K_R . Układ silnik–obiektyw reprezentuje czon całkująco-inercyjny. Układ stabilizacji przedstawia schemat blokowy (rys. 9.4), w którym wyjściem jest przesunięcie kątowe $\alpha(t)$, lub (gdy węzeł zaczepowy przesunięty jest poza człon całkujący) prędkość $v(t)$.

Otrzymuje się transmitancję układu zamkniętego stabilizującego prędkość.

$$G_z(s) = \frac{v(s)}{r(s)} = \frac{K_R K_o / (Ts + 1)}{1 + K_R K_o K_T / (Ts + 1)} = \frac{K_R K_o}{Ts + 1 + K_R K_o K_T}.$$

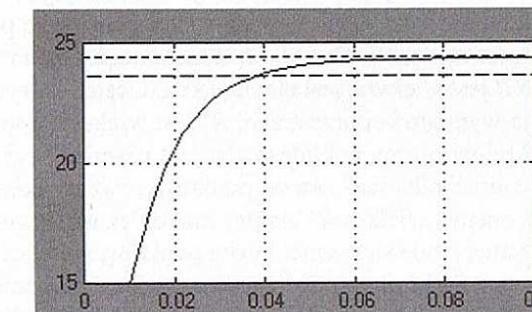
Przyjmuje się, że: $K_T = 1$, $r(t) = 25^\circ/\text{s}$ $1(t)$, $T = 0,4 \text{ s}$, $K_R K_o = K$.



Rys. 9.4. Schematy blokowe

Należy dobrąć współczynnik wzmacnienia K regulatora, który zagwarantuje błąd prędkości w stanie ustalonym $\epsilon_u \leq 1^\circ/\text{s}$ i czas regulacji $t_r \leq 0,04 \text{ s}$. Pierwszy warunek daje $\ell_u = 25 - \frac{25K}{1+K} < 1 \Rightarrow \text{dla } K \geq 24, \epsilon_u \leq 1$.

Pięcioprocentowe odchylenie determinujące czas regulacji wynosi $\pm 0,05 \cdot 25 = \pm 1,25$.



Rys. 9.5. Odpowiedź na $1(t)$

Odpowiedź członu inercyjnego na skok (rys. 9.5) wyraża się wzorem $v(t) = \frac{25 \cdot K}{1+K} \left[1 - e^{-\frac{1+K}{T} t} \right]$.

Stąd widać, że musi zachodzić nierówność $\frac{25 \cdot K}{1+K} \left[1 - e^{-\frac{1+K}{0,4} t_r} \right] \geq 23,75$.

Przyjmując graniczny czas regulacji $t_r = 0,04$, z nierówności otrzymuje się

$$1 - \frac{23,75}{25} \cdot \frac{1+K}{K} \geq e^{-\frac{1+K}{0,4} 0,04},$$

$$\frac{1,25K - 23,75}{25K} \geq e^{-\frac{1+K}{0,4} 0,04}.$$

Dla przypadku granicznego otrzymuje się $10 \cdot \ln \left(\frac{25K}{1,25K - 23,75} \right) - 1 = K$.

Rozwiążanie na K można znaleźć numeryczną metodą odwzorowań zwężających dla $f(K) = K$ (tab. 9.1).

Tabela 9.1

K	20	59,00	32,84	37,60	35,60	36,46	36,32	36,35
$f(K)$	59	32,84	37,60	35,60	36,46	36,32	36,35	36,35

Dla $K \geq 36,35$ czas regulacji t_r będzie mniejszy niż 0,04, a błąd w stanie ustalonym odpowiednio $e_u \leq 0,67^\circ/\text{s}$. Układ zamknięty jest oczywiście asymptotycznie stabilny.

9.2. Całkowe kryteria jakości

Można wprowadzić inne kryteria oceniające jakość przebiegu przejściowego, np. maksymalną amplitudę przeregulowania lub pole powierzchni pod funkcją opisującą przebieg przejściowego błędu regulacji $\varepsilon_p(t) = y_{ust} - y(t)$. W zadaniach nadążania przez zmienną wyjściową y za sygnałem referencyjnym r chodzi o jednoznaczna ocenę różnicę między dwoma funkcjami i minimalizację tej różnicy. Nadaje się do tego celu pojęcie normy i metryki wprowadzane w przestrzeniach funkcyjnych. Prosta ocena, która z dwóch liczb rzeczywistych w przestrzeni R jest większa, jest elementarna. Ocena, który z dwóch wektorów liczbowych w skończenie wymiarowej przestrzeni R^n jest większy, poprzez policzenie ich długości za pomocą naturalnej normy euklidesowej, jest również oczywista. Taka norma daje bowiem odległość wierzchołka wektora od punktu zerowego układu współrzędnych. Użycie jednej liczby dla oceny „wielkości” ciągłej funkcji czasu, określonej na skończonym odcinku, ale zbudowanej z nieskończonych liczb punktów, wymaga zdefiniowania pojęcia normy w przestrzeniach funkcyjnych. Taka norma, będąca nieujemną liczbą i posiadająca pewne dodatkowe cechy, powinna określić jednoznacznie stopień oddalenia funkcji od funkcji „zerowej” (osi czasu) nawet wtedy, gdy ta funkcja wielokrotnie przecina tę os.

Przykładowe normy przedstawiono poniżej.

- W przestrzeni funkcji jednostajnie ciągłych i ograniczonych, określonych na dodatniej półosi czasu $C[0, \infty)$, normą pojedynczego elementu należącego do tej przestrzeni (jakim jest każda funkcja ciągła i ograniczona) jest jego maksymalna amplituda.

$$\|f\|_C = \sup_{t \in [0, \infty)} |f(t)|.$$

- W przestrzeni funkcji absolutnie całkowalnych na dodatniej półosi czasu $L^1[0, \infty)$ (tzn. takich funkcji, że ich całka z modułu jest liczbą skończoną) normą jest całka z ich bezwzględnej wartości.

$$\|f\|_L = \int_0^\infty |f(t)| dt.$$

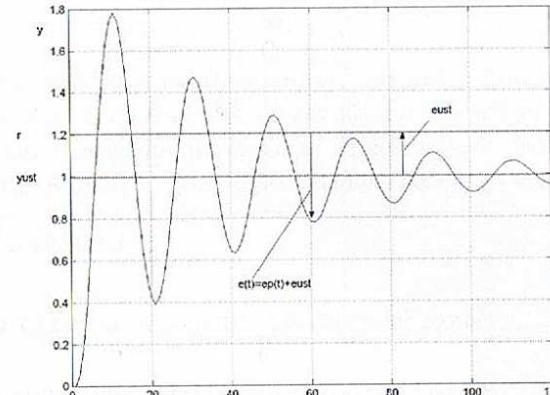
9.2. Całkowe kryteria jakości

W przestrzeni funkcji całkowalnych z kwadratem na dodatniej półosi czasu $L^2[0, \infty)$ (tzn. takich funkcji, że całka z ich kwadratu jest liczbą skończoną) normą każdego elementu jest pierwiastek z całki z kwadratu funkcji.

$$\|f\|_{L^2} = \sqrt{\int_0^\infty (f(t))^2 dt}.$$

Jedna i ta sama funkcja rozważana w różnych przestrzeniach funkcyjnych może mieć przypisane do siebie różne normy odpowiednie do danej przestrzeni.

Widać, że kryterium, które opierałoby się na liczeniu zwykłego pola powierzchni pod funkcją (całka z funkcji), nie stanowi normy, bo nie jest ono jednoznaczne (pole zerowe ma równo funkcja zerowa, jak i np. sinus). Wielkość normy liczonej w sposób „punktowy” (maksymalna amplituda) lub „powierzchniowy” (całka) reprezentuje ocenę błędu przejściowego $\varepsilon_p(t)$, a więc szybkość i dokładność nadążania sygnału wyjścia $y(t)$ za sygnałem referencyjnym np. $r = 1,2 \cdot 1(t)$ (rys. 9.6).

Rys. 9.6. Odpowiedzi na sygnał $1(t)$. Błąd ustalony $e_{ust} = \varepsilon_{ust}$

Zwykle parametry regulatora α mają wpływ na wielkość tej normy. Zwraca się tu zawsze uwagę, że chodzi o błąd przejściowy $\varepsilon_p(t)$, a nie błąd regulacji $\varepsilon(t)$, który składa się z błędu przejściowego $\varepsilon_p(t)$ (rys. 9.6) i błędem ustalonego ε_{ust} (jeśli taki w ogólnym przypadku występuje). Z uwagi na możliwość istnienia stałego błędu ustalonego całkowitego błędu regulacji $\varepsilon(t)$ mogłyby nie istnieć. Zachodzą podstawowe wzory:

$$\varepsilon(t) = r(t) - y(t), \quad \varepsilon(t) = \varepsilon_{ust} + \varepsilon_p(t), \quad \varepsilon_p(t) = y_\infty - y(t).$$

Odpowiednio do definicji norm stosuje się więc minimalizację wskaźników opartych na błędach przejściowych $\varepsilon_p(t)$:

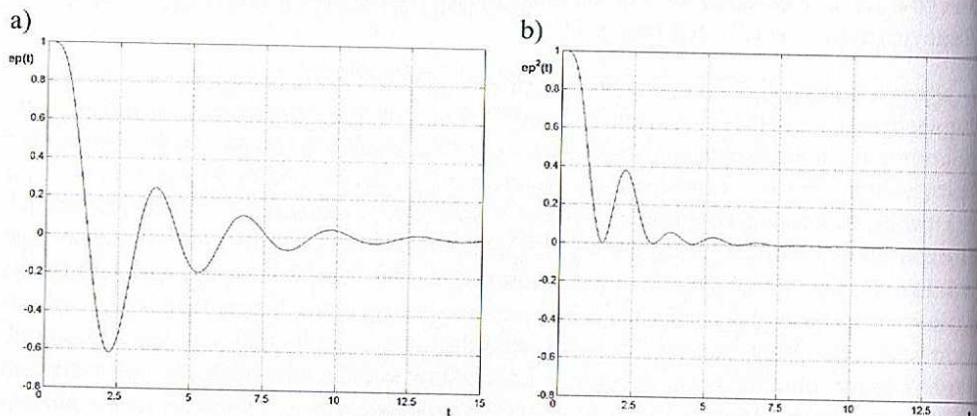
$$\min_{\alpha} J_1 = \min_{\alpha} \int_0^\infty |y_\infty - y(t)| dt,$$

$$\min_{\alpha} J_2 = \min_{\alpha} \int_0^\infty (y_\infty - y(t))^2 dt,$$

$$\min J_{21} = \min_{\alpha} \int_0^{\infty} [(y_{\infty} - y(t))^2 + (\dot{y}(t))^2] dt.$$

Kryterium J_{21} reprezentuje normę w tzw. przestrzeni H^1 , czyli w przestrzeni funkcji dla których zarówno funkcja, jak i jej pierwsza pochodna należą do $L^2[0, \infty)$.

Kryterium J_2 reprezentujące pole powierzchni pod kwadratem funkcji, jest proporcjonalne do energii, jaką niesie ten sygnał, stąd taka norma ma konkretną interpretację w zadaniach inżynierskich. Poza tym, z uwagi na cytowane w podrozdziale 4.1.2 twierdzenie Parsevala, jest ono dla sygnałów będących wyjściami obiektów liniowych obliczalne analitycznie w funkcji parametrów obiektu i regulatora. Wartość tej całki dla konkretnego regulatora i obiektu może być również policzona numerycznie z eksperymentu trwającego wystarczająco długo, aby przebiegi się ustaliły i pole powierzchni już nie przyrastało (rys. 9.7). Stąd norma J_2 jest szczególnie często stosowana w zagadnieniach sterowania.



Rys. 9.7. Przebiegi błędu (a) i kwadratu błędu przejściowego (b)

Jak wspomniano, kryterium minimalizacji całki z kwadratu błędu przejściowego $\epsilon_p(t)$ jest wygodne, bo dla układów liniowych istnieją wzory analityczne do wyrażenia wartości tej całki w funkcji parametrów obiektu i regulatora.

$$J_2 = \int_0^{\infty} \epsilon_p^2(t) dt = f(a, b, \alpha).$$

Aby znaleźć minimum całki po parametrach α :

$$\min J_2(\alpha) = \min_{\alpha} f(a, b, \alpha),$$

trzeba z warunku koniecznego istnienia ekstremum

$$\frac{\partial J_2(\alpha)}{\partial \alpha} = 0$$

znać optymalne parametry α^* i podstać je do wzoru na $J_2 = f(a, b, \alpha)$.

Pierwsze wzory na wartość całki J_2 dla obiektów do siódmego rzędu zostały wyprodukowane w 1947 r. [9JAH]. Uniwersalne formuły zostały podane przez A. Krasowskiego w 1948 r.

Na podstawie tych ostatnich prac podamy poniżej przepis, jak wykorzystując twierdzenie Parsevala, wyliczać funkcję $J_2 = f(a, b, a)$.

9.3. Formuły na liczenie całki z kwadratu błędu

Dana jest transformata odpowiedzi układu, której przebieg chcemy optymalizować, w postaci

$$Y(s) = \frac{b_m s^m + b_{m-1} s^{m-1} + \dots + b_1 s + b_0}{a_n s^n + a_{n-1} s^{n-1} + \dots + a_1 s + a_0} \cdot \frac{1}{s}, \quad (m \leq n) \quad (9.1)$$

Transformata wyjścia (z reguły interesuje nas układ zamknięty) powinna mieć powyższą postać i mimo że sugeruje ona odpowiadź na skok jednostkowy, to może opisywać odpowiedzi na inne sterowania, np. na deltę Diraca, sygnał narastający lub inny. Warunkiem stosowalności wzoru (9.1) będzie istnienie skończonego stanu ustalonego $y_{\infty} = b_0/a_0$.

Przykładowo dla układu:

- inercyjnego $G(s) = \frac{1}{s+1}$ wzór (9.1) będzie miał postać:
– dla sterowania deltą Diraca: $Y(s) = \frac{1}{s+1} \cdot 1 = \frac{s}{s+1} \cdot \frac{1}{s}, \quad y_{\infty} = 0,$
- dla sterowania skokiem jednostkowym: $Y(s) = \frac{1}{s+1} \cdot \frac{1}{s}, \quad y_{\infty} = 1,$
- różniczkującą-inercyjnego $G(s) = \frac{s}{s+1}$ wzór (9.1) będzie miał postać:
– dla sterowania skokiem jednostkowym: $Y(s) = \frac{s}{s+1} \cdot \frac{1}{s}, \quad y_{\infty} = 0,$
- sterowania sygnałem liniowym $u(t) = t$: $Y(s) = \frac{s}{s+1} \cdot \frac{1}{s^2} = \frac{1}{s+1} \cdot \frac{1}{s}, \quad y_{\infty} = 1.$

Formuła (9.1) jest uniwersalna. Zmianie ulegają tylko współczynniki a_i, b_i w zależności od przypadku.

Dla przypadku $m \leq n - 1$ całkę z kwadratu błędu $\varepsilon_p(t)$ dla (9.1) można obliczyć za pomocą wzoru [9KRA]

$$J_2 = \int_0^\infty \varepsilon_p^2(t) dt = \frac{1}{2a_0^2 \Delta} \cdot [B_0 \Delta_0 + B_1 \Delta_1 + \dots + B_m \Delta_m - 2b_0 b_1 \Delta] \quad (9.1)$$

Wyznaczniki Δ i Δ_k mają wymiar $(n \times n)$. Wyznaczniki Δ_k dla $k = 0, 1, \dots, m$ otrzymujemy się z wyznacznika Δ , zmieniając jego $(k+1)$ kolumnę na kolumnę $[a_1 \ a_0 \ 0 \ 0 \ \dots \ 0]^T$.

$$\Delta = \begin{bmatrix} a_0 & -a_2 & a_4 & -a_6 & \dots & \dots \\ 0 & a_1 & -a_3 & a_5 & \dots & \dots \\ 0 & -a_0 & a_2 & -a_4 & \dots & \dots \\ 0 & 0 & -a_1 & a_3 & \dots & \dots \\ 0 & 0 & a_0 & -a_2 & \dots & \dots \\ \vdots & & \ddots & & & \\ \vdots & & & a_{n-1} & & \end{bmatrix}, \quad \begin{aligned} B_0 &= b_0^2 \\ B_1 &= b_1^2 - 2b_0 b_2 \\ B_2 &= b_2^2 - 2b_1 b_3 + 2b_0 b_4 \\ &\vdots \\ B_k &= b_k^2 - 2b_{k-1} b_{k+1} + \dots + 2(-1)^k b_0 b_{2k} \\ &\vdots \\ B_m &= b_m^2 \end{aligned} \quad (9.1)$$

Dla przypadku $m = n$ i $b_0 = 0$ całkę z kwadratu błędu $\varepsilon_p(t)$ można obliczyć za pomocą podobnego wzoru:

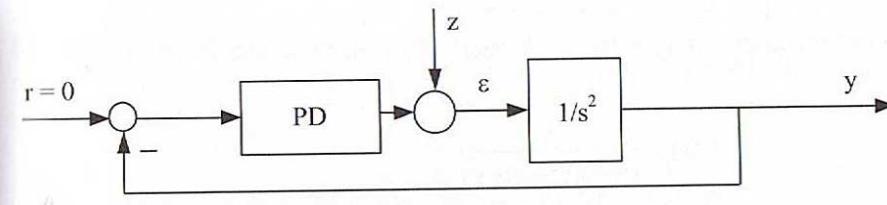
$$J_2 = \int_0^\infty \varepsilon_p^2(t) dt = \frac{1}{2a_0^2 \Delta} \cdot [\bar{B}_0 \Delta_0 + \bar{B}_1 \Delta_1 + \dots + \bar{B}_{n-1} \Delta_{n-1} - 2\bar{b}_0 \bar{b}_1 \Delta],$$

$$\Delta = \begin{bmatrix} a_0 & -a_2 & a_4 & -a_6 & \dots & \dots \\ 0 & a_1 & -a_3 & a_5 & \dots & \dots \\ 0 & -a_0 & a_2 & -a_4 & \dots & \dots \\ 0 & 0 & -a_1 & a_3 & \dots & \dots \\ 0 & 0 & a_0 & -a_2 & \dots & \dots \\ \vdots & & \ddots & & & \\ \vdots & & & a_{n-1} & & \end{bmatrix}, \quad \begin{aligned} \bar{b}_i &= b_n \left(\frac{b_i}{b_n} - \frac{a_i}{a_n} \right), \quad \text{przy } b_0 = 0 \\ \bar{B}_0 &= \bar{b}_0^2 \\ \bar{B}_1 &= \bar{b}_1^2 - 2\bar{b}_0 \bar{b}_2 \\ \bar{B}_2 &= \bar{b}_2^2 - 2\bar{b}_1 \bar{b}_3 + 2\bar{b}_0 \bar{b}_4 \\ &\vdots \\ \bar{B}_k &= \bar{b}_k^2 - 2\bar{b}_{k-1} \bar{b}_{k+1} + \dots + 2(-1)^k \bar{b}_0 \bar{b}_{2k} \\ &\vdots \\ \bar{B}_{n-1} &= \bar{b}_{n-1}^2 \end{aligned} \quad (9.4)$$

Wyznaczniki Δ_k dla $k = 0, 1, \dots, m$ otrzymuje się z wyznacznika Δ przez zmianę jego $(k+1)$ kolumny na kolumnę $[a_1 \ a_0 \ 0 \ 0 \ \dots \ 0]^T$.

Przykład 9.2

Dany jest układ stabilizacji (rys. 9.8).



Rys. 9.8. Schemat stabilizacji

Regulator PD ma transmitancję $G_R(s) = K + 2\xi s$. Transmitancja toru zakłócenioowego

$$G(s) = \frac{1}{s^2 + 2s\xi + K} = \frac{Y(s)}{Z(s)}.$$

Jaka wartość parametru ξ daje najmniejszy wskaźnik jakości $J_2 = \int_0^\infty \varepsilon_p^2(t) dt$ przy $\varepsilon(t) = 1(t)$?

$$Y(s) = \frac{1}{s^2 + 2s\xi + K} \cdot \frac{1}{s}.$$

Współczynniki we wzorze (9.1) wynoszą: $b_0 = 1$, $a_2 = 1$, $a_1 = 2\xi$ i $a_0 = K$.

$$J_2 = \int_0^\infty \varepsilon_p^2(t) dt = \frac{1}{2a_0^2 \cdot \Delta} [B_0 \Delta_0], \quad \Delta = \begin{vmatrix} a_0 & -a_2 \\ 0 & a_1 \end{vmatrix} = a_0 a_1 = 2\xi K,$$

$$B_0 = 1, \quad \Delta_0 = \begin{vmatrix} a_1 & -a_2 \\ a_0 & a_1 \end{vmatrix} = a_1^2 + a_0 a_2 = 4\xi^2 + K, \quad J_2 = \frac{4\xi^2 + K}{4\xi K^3},$$

$$\frac{\partial J_2}{\partial \xi} = \frac{8\xi \cdot 4\xi K^2 - (4\xi^2 + K)4K^3}{16\xi^2 K^6} = 0, \quad 32\xi^2 - 16\xi^2 - 4K = 0,$$

$$\xi^2 = K/4, \quad \xi^o = 0,5\sqrt{K}, \quad G_R^o = K + s\sqrt{K}, \quad J_2^o = 1/(K\sqrt{K}).$$

Dla $K = 1$ optymalna transmitancja ma postać $G_R^o(s) = \frac{1}{s^2 + s + 1}$, $J_2^o = 1$.

Przykład 9.3

Dany jest obiekt $G(s) = \frac{1}{s^3 + As^2 + Bs + 1}$.

Jakie wartości parametrów A, B dają najmniejszy wskaźnik $J_2 = \int_0^\infty \varepsilon_p^2(t)dt$ przy $u(t) = 1(t)$?

$$Y(s) = \frac{1}{s^3 + As^2 + Bs + 1} \cdot \frac{1}{s}.$$

Współczynniki we wzorze (9.1) wynoszą: $b_0 = 1, a_2 = A, a_1 = B, a_0 = 1$.

$$J_2 = \int_0^\infty \varepsilon_p^2(t)dt = \frac{1}{2 \cdot \Delta} [B_0 \Delta_0], \quad \Delta = \begin{vmatrix} 1 & -A & 0 \\ 0 & B & -1 \\ 0 & -1 & A \end{vmatrix} = AB - 1,$$

$$B_0 = 1, \quad \Delta_0 = \begin{vmatrix} B & -A & 0 \\ 1 & B & -1 \\ 0 & -1 & A \end{vmatrix} = B(AB - 1) + A^2, \quad J_2 = \frac{B}{2} + \frac{A^2}{2(AB - 1)},$$

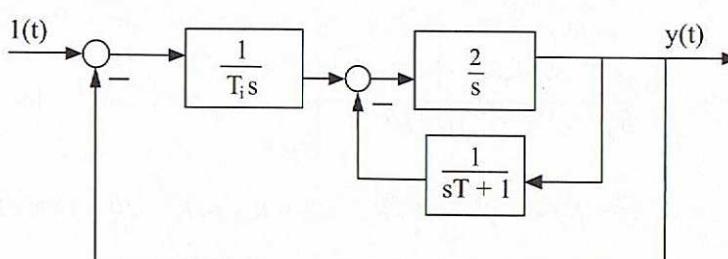
$$\frac{\partial J}{\partial A} = \frac{2A(AB - 1) - A^2 B}{2(AB - 1)^2} = 0, \Rightarrow B = 2/A, \quad \frac{\partial J}{\partial B} = \frac{1}{2} \left[1 - \frac{A^3}{(AB - 1)^2} \right] = 0, \Rightarrow A = 1,$$

$$B = 2, \quad A = 1.$$

Optymalna transmitancja ma postać $G^o(s) = \frac{1}{s^3 + s^2 + 2s + 1}, \quad J_2^0 = 1,5$.

Przykład 9.4 [9AMK]

Dla układu nadążania (rys. 9.9) podać optymalną, według kryterium J_2 , stałą regulatora całkującego T_i .



Rys. 9.9. Schemat blokowy

Transmitancja układu zamkniętego w torze sterowania wynosi:

$$G(s) = \frac{\frac{1}{T_i s} \cdot \frac{2}{s}}{1 + \frac{1}{T_i s} \cdot \frac{2}{s} \cdot \frac{k}{sT+1}} = \frac{\frac{2}{s}}{sT_i \left(1 + \frac{2}{s} \cdot \frac{k}{sT+1} \right) + \frac{2}{s}} = \frac{2(sT+1)}{s^3 T_i T + s^2 T_i + 2kT_i s + 2Ts + 2},$$

$$b_1 = 2T, \quad a_3 = T_i T, \quad a_1 = 2kT_i + 2T, \quad b_0 = 2, \quad a_2 = T_i, \quad a_0 = 2,$$

$$\Delta = \begin{vmatrix} 2 & -T_i & 0 \\ 0 & 2kT_i + 2T & -T_i T \\ 0 & -2 & T_i \end{vmatrix} = 2(2(kT_i + T)T_i - 2T_i T) = 4kT_i^2, \quad B_0 = 4, \quad B_1 = 4T^2,$$

$$\Delta_0 = \begin{vmatrix} 2kT_i + 2T & -T_i & 0 \\ 0 & 2 & -T_i T \\ 0 & -2 & T_i \end{vmatrix} = (2kT_i + 2T)(2kT_i^2) + 2T_i^2 = 2T_i^2(2k^2 T_i + 2kT + 1),$$

$$\Delta_1 = \begin{vmatrix} 2 & 2kT_i + 2T & 0 \\ 0 & 2 & -T_i T \\ 0 & 0 & T_i \end{vmatrix} = 4T_i,$$

$$\begin{aligned} J &= \frac{1}{8 \cdot 4kT_i^2} \left[4 \cdot 2T_i^2 (2k^2 T_i + 2kT + 1) + 4T^2 \cdot 4T_i - 2 \cdot 2 \cdot 2T \cdot 4kT_i^2 \right] = \\ &= \frac{2k^2 T_i^3 - 2kT_i^2 T + T_i^2 + 2T^2 T_i}{4kT_i^2} = \frac{2k^2 T_i^2 - T_i (2kT - 1) + 2T^2}{4kT_i^2}, \end{aligned}$$

$$\frac{\partial J}{\partial T_i} = 0 \Rightarrow [4k^2 T_i - (2kT - 1)]4kT_i - [2k^2 T_i - T_i (2kT - 1) + 2T^2]4k = 0.$$

$$16k^3 T_i - 4kT_i (2kT - 1) - 8k^3 T_i^2 + 4kT_i (2kT - 1) - 8kT^2 = 0,$$

$$8k^3 T_i^2 - 8kT^2 = 0; \quad k^2 T_i^2 = T^2,$$

$$T_i^o = \frac{T}{k}, \quad G_R^o(s) = \frac{k}{Ts}, \quad G^o(s) = \frac{2k(1+Ts)}{s^3 T^2 + s^2 T + 4Tks + 2k}, \quad J^o = 2k + \frac{1}{T}.$$

9.4. Uogólnione kwadratowe kryterium całkowe

Jak stwierdzono eksperymentalnie, przebiegi przejściowe otrzymane w wyniku zastosowania nastaw regulatorów opierających się na kryterium J_2 wykazują oscylacyjny charakter. Oscylacje można zmniejszyć, jeśli zastosuje się kryterium wrażliwe nie tylko na błąd przejściowy, ale też na jego pochodną. Przebieg będzie mniej oscylacyjny i tym bardziej będzie dążył do przebiegu aperiodycznego, im więcej pochodnych wyższych rzędów zostanie uwzględnione.

Uogólnione kryterium kwadratowo-całkowe może mieć postać

$$J_{2n} = \int_0^\infty [y^2(t) + \beta_1 \dot{y}^2(t) + \beta_2 \ddot{y}^2(t) + \dots + \beta_n (y^{(n)})^2(t)] dt \quad (9.4)$$

gdzie dowolne liczbowe współczynniki $\beta_i > 0$ reprezentują wagę poszczególnych pochodnych.

Powyższą całkę można przedstawić jako sumę całek. Każdą z nich można przedstawić jako funkcję parametrów, obliczając transformatę pochodnej $y^{(l)}(t)$ i stosując wzory (9.3) lub (9.4).

Przykład 9.5

Dany jest system jak w przykładzie 9.2 (rys. 9.8) z transmitancją dla $K = 1$

$$G(s) = \frac{1}{s^2 + 2s\xi + 1}.$$

Jaka wartość parametru ξ daje najmniejszy sumaryczny wskaźnik jakości

$$J_{21} = \int_0^\infty [\varepsilon_p^2(t) + \beta \dot{\varepsilon}_p^2(t)] dt = J_2 + J_{20}.$$

Z przykładu 9.2 wynika, że całka pierwsza J_2 równa się $J_2 = \frac{4\xi^2 + 1}{4\xi}$. Całkę drugą wylicza się z warunku

$$\mathcal{L}[\dot{y}(t)] = \frac{s}{s^2 + 2s\xi + 1} \cdot \frac{1}{s}.$$

Współczynnik: $b_0 = 0$, $b_1 = 1$, $a_2 = 1$, $a_1 = 2\xi$, $a_0 = 1$.

Stosując wzory (9.3), otrzymuje się:

$$J_{20} = \beta \int_0^\infty \dot{\varepsilon}_p^2(t) dt = \frac{\beta}{2a_0^2 \cdot \Delta} [B_1 \Delta_1], \quad \Delta = \begin{vmatrix} a_0 & -a_2 \\ 0 & a_1 \end{vmatrix} = a_0 a_1 = 2\xi,$$

$$B_1 = 1, \quad \Delta_1 = \begin{vmatrix} a_0 & a_1 \\ 0 & a_2 \end{vmatrix} = 1, \quad J_{20} = \beta \frac{1}{4\xi},$$

$$J_{21} = \frac{4\xi^2 + 1 + \beta}{4\xi}, \quad \frac{\partial J_{21}}{\partial \xi} = \frac{4\xi^2 - (1 + \beta)}{4\xi^2} = 0, \quad \xi^o = \frac{\sqrt{1 + \beta}}{2}, \quad J^o = \sqrt{1 + \beta}.$$

Dla $\beta = 1$ $\xi^o = \frac{1}{\sqrt{2}}$ i optymalna transmitancja ma postać:

$$G^o(s) = \frac{1}{s^2 + \sqrt{2}s + 1}, \quad J_{21}^o = \sqrt{2}.$$

Przykład 9.6

Wyliczyć optymalne nastawy regulatora PI o transmitancji $G_R = [K + 1/(Ts)]$ dla likwidacji wpływu zakłócenia skokowego z_1 jak na rysunku 6.2, przychodzącego między obiekt i regulator. Obiektem sterowania jest model III rzędu (całkujący z inercją II rzędu)

$$G(s) = \frac{1}{s(s+1)(s+2)}.$$

Nastawy optymalne K i T mają minimalizować kryterium

$$J_{21} = \int_0^\infty [\varepsilon_p^2(t) + \dot{\varepsilon}_p^2(t)] dt.$$

Transmitancja zamknięta i wskaźnik jakości wyrażają się wzorem

$$G_z(s) = \frac{sT}{Ts^4 + 3Ts^3 + 2Ts^2 + KTs + 1} = \frac{Y(s)}{Z(s)}.$$

$$\Delta = KT^3(6-K) - 9T^2, \quad \Delta_1 = 6T^2 - KT^2, \quad \Delta_2 = 3T, \quad B_1 = T^2, \quad B_2 = T^2,$$

$$J_{21} = \frac{B_1 \Delta_1 + B_2 \Delta_2}{2\Delta} = \frac{(6-K)T^2 + 3T}{2[KT(6-K) - 9]},$$

$$\frac{\delta J_{21}}{\delta K} = 0 \Rightarrow T(12K - K^2 - 36) + 6K - 9 = 0,$$

$$\frac{\delta J_{21}}{\delta T} = 0 \Rightarrow T^2 K (36 - 12K + K^2) + T(18K - 108) - 27 = 0.$$

Z pierwszego równania oblicza się T . Po wstawieniu do drugiego równania otrzymuje się równanie na optymalne K^o :

$$T = \frac{6K - 9}{K^2 - 12K + 36}, \quad 4K^2 - 3K - 45 = 0 \Rightarrow K^o = 30/8 \Rightarrow T^o = 8/3, J^o = 8/9.$$

Inne nastawy tego regulatora zapewniają w tym układzie realizację najlepszego nadążania za sygnałem referencyjnym $r = 1(t)$, przy nieobecności zakłóceń. Minimalizacja kryterium J_{21} daje wyniki:

$$K^o = 1,283 \Rightarrow 1/T^o = 0, \quad J^o = (3K^2 + 7K + 12)/(12K - 2K^2) = 2,14.$$

Z postacią uogólnioną całki (9.5) wiąże się ciekawe spostrzeżenie. Z punktu widzenia teorii optymalizacji dotyczące zadania optymalizacji parametrycznej polegały na minimalizacji kryterium w obecności ograniczeń (optymalizacja z ograniczeniami). Tymi ograniczeniami jest równanie procesu (9.1). Poszukiwane są bowiem przebiegi najlepsze $y(t)$, ale tylko z klasy funkcji spełniających równanie (9.1). Można rozwiązać zadanie znalezienia kształtu $y(t)$, który daje minimum uogólnionego kryterium całkowego (9.5), ale bez ograniczeń. Jest to oczywiście inne zadanie optymalizacji. Aby go rozwiązać, wystarczy całkując przez części wzór (9.5), wykazać, że zachodzi zależność (przy zerowych warunkach początkowych na pochodne)

$$\begin{aligned} J_n &= \int_0^\infty [y^2(t) + \beta_1 \dot{y}^2(t) + \beta_2 \ddot{y}^2(t) + \dots + \beta_n (y^{(n)})^2(t)] dt = \\ &= \int_0^\infty [y(t) + b_1 \dot{y}(t) + b_2 \ddot{y}(t) + \dots + b_n y^{(n)}(t)]^2 dt + b_1 y^2(0) \end{aligned} \quad (9.6)$$

Całka z sumy kwadratów pochodnych została zastąpiona całką z kwadratu sumy pochodnych.

Między współczynnikami β_i i b_i zachodzą wzory:

$$\begin{aligned} \beta_0 &= 1, & \beta_k &= b_k^2 - 2b_{k-1}b_{k+1} + \dots + 2(-1)^k b_{2k} \\ \beta_1 &= b_1^2 - 2b_2, & \vdots & \\ \beta_2 &= b_2^2 - 2b_1b_3 + 2b_4, & \beta_n &= b_n^2 \\ &\vdots & & \end{aligned} \quad (9.7)$$

Warto zauważyć, że zależności powyższe są identyczne jak w problemie danym przez (9.3), przy założeniu $b_0 = 1$. Ponieważ wartość $b_1 y^2(0)$ w (9.6) jest stała, to minimum globalne (9.6) jest osiągane w przypadku zerowania się całki, czyli zachodzi to wtedy, jeśli przebieg $y(t)$ jest rozwiązaniem równania różniczkowego

$$b_n y^{(n)}(t) + \dots + b_2 \ddot{y}(t) + b_1 \dot{y}(t) + y(t) = 0 \quad (9.8)$$

z warunkami początkowymi:

$$y(0) = y_0, \quad \dot{y}(0) = \ddot{y}(0) = \dots = y^{(n)}(0) = 0 \quad (9.9)$$

Aby znaleźć b_i , oraz całe równanie (9.8) i jego rozwiązanie (przy zadanych wagach β_i), trzeba najpierw rozwiązać układ równań (9.7).

Można również sformułować zadanie odwrotne: dane jest równanie różniczkowe (9.8) ze współczynnikami b_i i warunkami początkowymi (9.9) – należy znaleźć jego rozwiązanie. Po wyliczeniu z (9.7) współczynników β_i można numerycznie minimalizować jedną lub drugą całkę z (9.6) po różnych funkcjach $y(t)$. Funkcja zerująca całkę będzie rozwiązaniem (9.8).

9.5. Całkowe kryterium z modułu błędu

Regulatory optymalne zaprojektowane przy wykorzystaniu minimalizacji całkowego kryterium J_1 , odpowiadającego normie błędu regulacji w $L^1[0, \infty]$, dają mniej oscylacyjne przebiegi niż regulatory otrzymane w wyniku minimalizacji kryterium J_2 . Ponieważ do analitycznego wyliczenia pola powierzchni pod modelem błędu $\varepsilon_p(t)$ wymagana jest znajomość chwil przejść przez zero tego sygnału, to otrzymanie analitycznych wzorów możliwe jest tylko dla układów II rzędu (ewentualnie III). Poniżej wyprowadzimy te formuły dla układu II rzędu.

Przykład 9.7

Dany jest obiekt podwójnie całkujący, pracujący z regulatorem PD: $G_R(s) = K + 2\xi\sqrt{K}s$. Transmitancja układu zamkniętego w torze zakłócenia ma postać jak w przykładzie 9.2

$$G(s) = \frac{1}{s^2 + 2\xi\sqrt{K}s + K}.$$

Przebieg błędu $\varepsilon_p(t)$ na skok jednostkowy zakłócenia dla oscylacyjnego przypadku $0 \leq \xi < 1$ według (4.18) ma postać

$$\varepsilon_p(t) = \frac{1}{K} \frac{1}{\sqrt{1-\xi^2}} e^{-\xi\sqrt{K}t} \sin(\sqrt{K}\sqrt{1-\xi^2}t + \Phi), \text{ gdzie } \Phi = \arccos \xi.$$

Momenty przejścia przez zero tej funkcji są dane wzorem

$$t_i = \frac{i\pi - \Phi}{\sqrt{K}\sqrt{1-\xi^2}} : i = 1, 2, \dots$$

Wskaźnik jakości J_1 ma postać:

$$J_1 = \int_0^\infty |\varepsilon_p(t)| dt = \int_0^{t_1} \varepsilon_p(t) dt - \int_{t_1}^{t_2} \varepsilon_p(t) dt + \int_{t_2}^{t_3} \varepsilon_p(t) dt - \dots = J_{11} - J_{12} + J_{13} - \dots$$

Aby go policzyć, przypomnijmy podstawowy wzór:

$$\int e^{at} \sin(bt+c) dt = \frac{e^{at}}{a^2+b^2} [a \sin(bt+c) - b \cos(bt+c)],$$

stąd

$$\begin{aligned} &\int e^{-\xi\sqrt{K}t} \sin(\sqrt{K}\sqrt{1-\xi^2}t + \Phi) dt = \\ &= -\frac{e^{-\xi\sqrt{K}t}}{\sqrt{K}} \left[\xi \sin(\sqrt{K}\sqrt{1-\xi^2}t + \Phi) + \sqrt{1-\xi^2} \cos(\sqrt{K}\sqrt{1-\xi^2}t + \Phi) \right]. \end{aligned}$$

Składowe J_{1i} mają postać:

$$J_{11} = \frac{1}{\sqrt{K} K \sqrt{1-\xi^2}} \left[e^{-\xi \frac{\pi-\Phi}{\sqrt{1-\xi^2}}} \cdot \sqrt{1-\xi^2} + \xi \sin \Phi + \sqrt{1-\xi^2} \cdot \cos \Phi \right],$$

$$J_{12} = \frac{1}{\sqrt{K} K \sqrt{1-\xi^2}} \left[-e^{-\xi \frac{2\pi-\Phi}{\sqrt{1-\xi^2}}} \sqrt{1-\xi^2} - e^{-\xi \frac{\pi-\Phi}{\sqrt{1-\xi^2}}} \left(\sqrt{1-\xi^2} \right) \right],$$

$$J_{13} = \frac{1}{\sqrt{K} K \sqrt{1-\xi^2}} \left[e^{-\xi \frac{3\pi-\Phi}{\sqrt{1-\xi^2}} t} \sqrt{1-\xi^2} + e^{-\xi \frac{2\pi-\Phi}{\sqrt{1-\xi^2}} t} \sqrt{1-\xi^2} \right].$$

Ostatecznie

$$J_1 = \frac{1}{\sqrt{K} K} \left[\frac{\xi \sin \Phi}{\sqrt{1-\xi^2}} + \cos \Phi + 2 \left(e^{-\xi \frac{\pi-\Phi}{\sqrt{1-\xi^2}}} + e^{-\xi \frac{2\pi-\Phi}{\sqrt{1-\xi^2}}} + e^{-\xi \frac{3\pi-\Phi}{\sqrt{1-\xi^2}}} \dots \right) \right].$$

Uwzględniając granicę szeregu geometrycznego, otrzymamy,

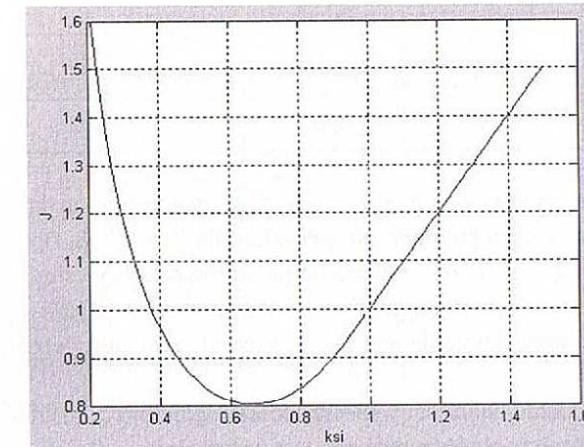
$$J_1 = \frac{1}{\sqrt{K} K} \left[\frac{\xi \sin \Phi}{\sqrt{1-\xi^2}} + \cos \Phi + 2 \cdot e^{\frac{\xi \Phi}{\sqrt{1-\xi^2}}} \cdot \frac{-\frac{\xi \pi}{\sqrt{1-\xi^2}}}{1 - e^{-\frac{\xi \pi}{\sqrt{1-\xi^2}}}} \right].$$

Ponieważ $\Phi = \arccos \xi$, a $\sin \Phi = \sqrt{1 - \cos^2 \Phi}$, to

$$\frac{\xi \sin \Phi}{\sqrt{1-\xi^2}} + \cos \Phi = \frac{\xi \cdot \sqrt{1-\xi^2}}{\sqrt{1-\xi^2}} + \xi = 2\xi.$$

$$\text{Ostatecznie } J_1 = \frac{2}{\sqrt{K} K} \left[\xi + \frac{e^{-\xi \frac{(\pi-\Phi)}{\sqrt{1-\xi^2}}}}{1 - e^{-\xi \frac{\pi}{\sqrt{1-\xi^2}}}} \right], \quad \text{gdzie } \Phi = \arccos \xi.$$

Przebieg funkcji J_1 w funkcji ξ dla $K = \sqrt[3]{4}$, pokazano na rysunku 9.10. Wynika z niego, że w zakresie oscylacyjności $0 \leq \xi < 1$ istnieje optymalny współczynnik tłumienia niezależny od K , $\xi^o = 0,663$, który minimalizuje kryterium $J_1 = 0,8026$.



Rys. 9.10. Przebieg wskaźnika J_1 w funkcji współczynnika tłumienia ξ

Dla przypadku $\xi \geq 1$ nie występują oscylacje i uzyskuje się prosty rezultat na J_1

$$J_1 = \int_0^\infty \varepsilon_p(t) dt.$$

Według (4.20) dla $\xi \geq 1$

$$\varepsilon_p(t) = \frac{e^{-\xi \sqrt{K} t}}{K} \left[\frac{\xi}{\sqrt{\xi^2 - 1}} \sinh \left(\sqrt{K} \sqrt{\xi^2 - 1} t \right) + \cosh \left(\sqrt{K} \sqrt{\xi^2 - 1} t \right) \right]$$

lub

$$\varepsilon_p(t) = \frac{1}{K} \left[\frac{\xi + \sqrt{\xi^2 - 1}}{2\sqrt{\xi^2 - 1}} \cdot e^{-\sqrt{K}(\xi - \sqrt{\xi^2 - 1})t} - \frac{\xi - \sqrt{\xi^2 - 1}}{2\sqrt{\xi^2 - 1}} \cdot e^{-\sqrt{K}(\xi + \sqrt{\xi^2 - 1})t} \right].$$

Po wyliczeniu całki z powyższego wzoru otrzymuje się prosty wynik na J_1 ,

$$J_1 = \frac{2\xi}{\sqrt{K} \cdot K}.$$

Jego graficzna reprezentacja stanowi przedłużenie wykresu z rysunku 9.10 linią prostą $J_1 = \xi$, (dla $\xi \geq 1$ i $K = \sqrt[3]{4}$). Wynik ten można otrzymać również z ogólnego wzoru dla układów przetłumionych, podanego w następnym podrozdziale.

Z przykładów 9.2, 9.5 i 9.6 widać, że dla $K = 1$ i obiektu $G(s) = \frac{1}{s^2 + 2\xi s + 1}$ można zestawić optymalne wartości ξ (tab. 9.2) minimalizujące różne kryteria jakości – również maksymalny moduł z transmitancji $\left| G(\omega) \right|_{\max} = \frac{1}{2\xi \sqrt{1-\xi^2}}$.

Tabela 9.2

J	Tłumienie do 1/4	$J_2 = \int \varepsilon^2 dt$	$J_1 = \int \varepsilon dt$	$J_{21} = \int [\varepsilon^2 + \beta \dot{\varepsilon}^2] dt$	$ G(\omega) _{\max}$
ξ^o	0,215	0,5	0,663	$(\sqrt{1+\beta})/2$	$1/\sqrt{2} \approx 0,707$

Widać, że zmieniając współczynnik wagi β w kwadratowym kryterium J_{21} , można otrzymać różne optymalne regulatory. Na przykład dla $\beta = 0,758$, optymalny współczynnik tłumienia równa się $\xi^o = 0,663$ i minimalizuje on również kryterium modułowe J_1 , a dla $\beta = 1$, $\xi^o = 1/\sqrt{2}$ i jest on optymalnym również dla kryterium $|G(\omega)|_{\max}$. Dla $\beta = 3$ optymalnym współczynnikiem tłumienia jest $\xi^o = 1$, a układ zamknięty jest wtedy aperiodyczny $G(s) = 1/(s+1)^2$.

W zasadzie, dla każdej wartości współczynnika regulatora ξ , można dobrać jakieś kryterium, względem którego współczynnik ten jest optymalny.

Problem doboru właściwego kryterium jest problemem kluczowym przy strojeniu regulatorów i powinien być rozstrzygnięty z udziałem technologów procesu.

9.6. Liniowe kryterium jakości

Jak już wspomniano, dla układów wyższych rzędów nie da się analitycznie policzyć wartości kryterium odpowiadającego normie w $L^1[0, \infty)$. Jeśli jednak jest pewność, że przebiegi przejściowe będą się odbywać bez przeregulowania (nie będzie zmian znaku pola $\varepsilon_p(t)$), to można zastosować kryterium całkowe liniowe [9NOP].

$$J_1 = \int_0^\infty \varepsilon_p(t) dt \quad (9.10)$$

Przy skokowym wymuszeniu i liniowym obiekcie n -tego rzędu odpowiedź ma postać

$$Y(s) = \frac{b_m s^m + b_{m-1} s^{m-1} + \dots + b_1 s + b_0}{a_n s^n + a_{n-1} s^{n-1} + \dots + a_1 s + a_0} \cdot \frac{1}{s}$$

Wykorzystując twierdzenia graniczne, można zapisać dla wskaźnika jakości (9.10) zależności

$$\begin{aligned} J_1 &= \int_0^\infty (y_\infty - y(t)) dt = \lim_{t \rightarrow \infty} \int_0^t (y_\infty - y(t)) dt = \\ &= \lim_{s \rightarrow 0} s \left[\frac{Y_\infty}{s} - Y(s) \right] - \lim_{s \rightarrow \infty} s \left[\frac{Y_\infty}{s} - Y(s) \right], \end{aligned}$$

$$J_1 = \lim_{s \rightarrow 0} \left[\frac{b_0}{a_0 s} - \frac{b_m s^m + \dots + b_0}{s(a_n s^n + \dots + a_0)} \right] - \lim_{s \rightarrow \infty} \left[\frac{b_0}{a_0 s} - \frac{b_m s^m + \dots + b_0}{s(a_n s^n + \dots + a_0)} \right] = \frac{a_1 b_0 - a_0 b_1}{a_0^2} \quad (9.11)$$

Można aproksymować odpowiedź układu inercyjnego n -tego rzędu przez układ I rzędu w sensie kryterium (9.10).

Odpowiedź zastępczego układu inercyjnego I rzędu ze stałą czasową T , na sterowanie skokowe

$$T \dot{\varepsilon}(t) + \varepsilon(t) = b_0 / a_0 \cdot 1(t)$$

$$\text{postać } \varepsilon(t) = b_0 / a_0 \cdot \left(1 - e^{-\frac{t}{T}} \right).$$

$$\text{Błąd przejściowy wynosi } \varepsilon_p(t) = b_0 / a_0 - b_0 / a_0 \cdot \left(1 - e^{-\frac{t}{T}} \right) = b_0 / a_0 \cdot e^{-\frac{t}{T}}.$$

Pole powierzchni pod krzywą $\varepsilon_p(t)$ ma wartość

$$\int_0^\infty \left[\frac{b_0}{a_0} e^{-\frac{t}{T}} \right] dt = \frac{b_0}{a_0} T \quad (9.12)$$

Przyrównując (9.11) i (9.12), można otrzymać stałą czasową zastępczego układu I rzędu

$$T_z = \frac{a_1}{a_0} - \frac{b_1}{b_0}$$

aproksymującego w sensie kryterium (9.10) aperiodyczny układ n -tego rzędu.

9.7. Inżynierskie metody strojenia regulatorów

Jak widać z powyższych rozważań, optymalne nastawy regulatora wyrażają się w funkcji parametrów obiektu regulowanego. Trzeba więc znać model tego obiektu. Im jest on wyższego rzędu, tym więcej parametrów należy zidentyfikować (możliwie dokładniej). W zastosowaniach inżynierskich, zwłaszcza w zadaniach wstępnego strojenia, funkcjonują uproszczone reguły wyboru parametrów regulatorów na podstawie dwóch założeń co do modelu obiektu regulowanego:

- 1) przyjmuje się obiekt całkujący z opóźnieniem lub obiekt rzędu pierwszego z opóźnieniem (obiektem tym można aproksymować modele wyższych rzędów),
- 2) trzy parametry tego obiektu K , T , τ odczytywane są z jego odpowiedzi na skok jednostkowy.

Proponowane reguły strojenia wynikają z obszernych badań eksperymentalnych, jakie przeprowadzono w latach 40. dla powyższego typu obiektu, dobierając najlepsze zestawy parametrów regulatora w funkcji parametrów obiektu i dokonując aproksymacji tych zależności.

ności w postaci prostych funkcji matematycznych. Otrzymane wyniki nie wynikają z procedur optymalizacyjnych, gwarantują natomiast stabilność układu zamkniętego i spełnienie kryterium „tłumienia do 1/4” omawianego w podrozdziale 4.6.1.

9.7.1. Pierwsza metoda Zieglera–Nicholsa dla układów ciągły

Pierwsza metoda Zieglera–Nicholsa (1942 r.) (tab. 9.3) doboru parametrów regulatorów PID dla pracy w układzie zamkniętym oparta jest na modelu odpowiedzi obiektu I rzędu z opóźnieniem na skok jednostkowy (w układzie otwartym) [09TAY].

Tabela 9.3

Proponowane nastawy ZN1

Regulator	K_R	T_I	T_D
P	$T/(K\tau)$	–	–
PI	$0,9T/(K\tau)$	$3,3\tau$	–
PID	$1,2T/(K\tau)$	2τ	$0,5\tau$

Transmitancja obiektu dana jest wzorem

$$G(s) = \frac{K e^{-\tau s}}{T s + 1} \quad (9.13)$$

Praktyczna uniwersalność metody została przebadana też dla wielu aperiodycznych obiektów wyższego rzędu bez opóźnienia, których odpowiedź może być aproksymowana metodą Küpfmüllera przez odpowiedź obiektu (9.13), z praktycznym wskazaniem, że liczba R wyrażająca stosunek $K/T = R$ powinna być odczytana z odpowiedzi jednostkowej obiektu wysokiego rzędu, jako maksymalne nachylenie stycznej do tej odpowiedzi (czyli w punkcie przegięcia). Styczna ta odcina na osi czasu zastępco opóźnienie τ i stałą czasową T .

Poniżej podane zostaną przykłady stosowania metody Zieglera–Nicholsa.

Przykład 9.8

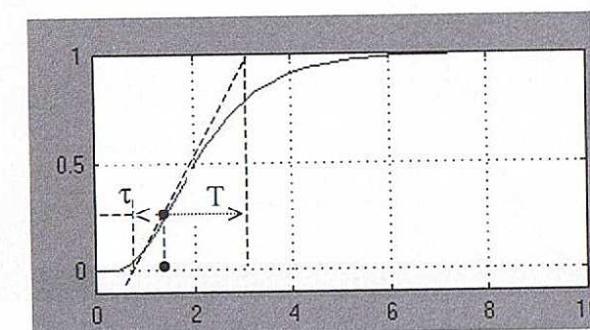
Odpowiedź układu V rzędu przedstawiona jest na rysunku 9.11.

$$G(s) = \frac{120}{(s+1)(s+2)(s+3)(s+4)(s+5)} \quad (9.14)$$

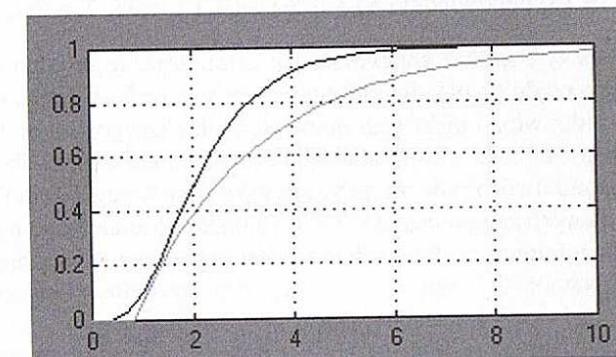
Numeryczna analiza odpowiedzi (9.14) w celu policzenia maksymalnej wartości pierwszej pochodnej $y'(t)$ wskazała na punkt przegięcia $t_p = 1,62$, $y(t_p) = 0,33$, $y'(t_p) = 0,41$, stąd $R = 1/T = 0,41$. Stała czasowa $T = 2,44$, a opóźnienie $\tau = 0,82$.

Z takiej graficznej analizy wynika, że obiekt (9.14) jest aproksymowany transmitancją

$$G_1(s) = \frac{e^{-0,82s}}{2,44s + 1} \quad (9.15)$$

Rys. 9.11. Odpowiedź na $1(t)$

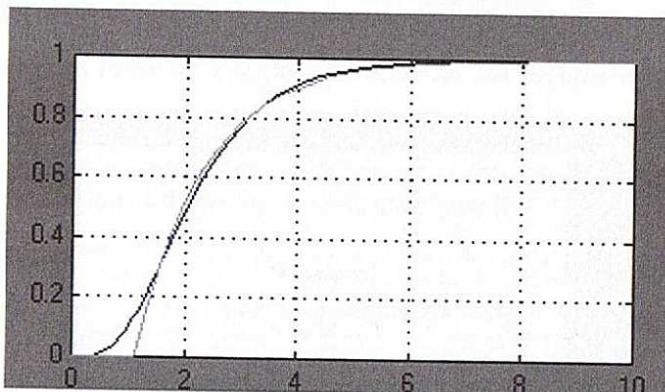
Porównanie odpowiedzi na skok jednostkowy obiektów (9.14) i (9.15) wykazuje jednak ich różnicę (rys. 9.12). Całka z modułu różnicy wynosi $J_1 = 0,92$.

Rys. 9.12. Odpowiedzi na $1(t)$

Inny wynik daje aproksymacja dokładna przeprowadzona poprzez numeryczną procedurę minimalizacji całki z modułu różnicy pomiędzy pełnymi odpowiedziami obiektu (9.14) i obiektu typu (9.13) względem T i τ . Najlepsze parametry obiektu (9.13) gwarantujące $\min J_1 = 0,127$ mają wartości $T = 1,24$ i $\tau = 1,1$. Stąd transmitancja $G_2(s)$ dana jest wzorem

$$G_2(s) = \frac{e^{-1,1s}}{1,24s + 1} \quad (9.16)$$

Odpowiedzi wyrysowane na rysunku 9.13. Widać lepszą zgodność niż na rysunku 9.12.

Rys. 9.13. Odpowiedzi na $1(t)$

Łatwo policzyć z tabeli 9.3, że zalecany regulator PID powinien mieć:

- dla obiektu G_1 (9.15) parametry $K_R = 1,2 \cdot 2,44 / 0,82 = 3,57$, $T_i = 1,64$, $T_D = 0,41$;
- dla obiektu G_2 (9.16) odpowiednio $K_R = 1,2 \cdot 1,24 / 1,1 = 1,35$, $T_i = 2,2$, $T_D = 0,55$.

Są to różne nastawy i można sprawdzić, jak ostatecznie te regulatory współpracują z obiektem wysokiego rzędu (9.14), dla którego przeprowadzano te różne aproksymacje.

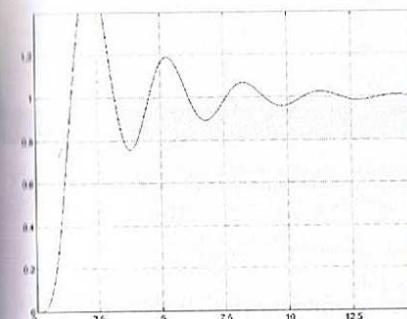
W tabeli 9.4 przedstawiono efekt tych numerycznych eksperymentów w układzie zamkniętym, pokazując wartości całki z kwadratu błędu i całki z modułu błędu dla obiektu V rzędu (9.14) przy zastosowaniu dwóch powyższych zestawów nastaw regulatora PID według Zieglera–Nicholsa obliczonych za pomocą tabeli 9.3. Dodatkowo znaleziono nastawy optymalne dające rzeczywiste minimum tych dwóch kryteriów regulatora PID i obiektu (9.14).

Tabela 9.4
Zebrane nastawy optymalne

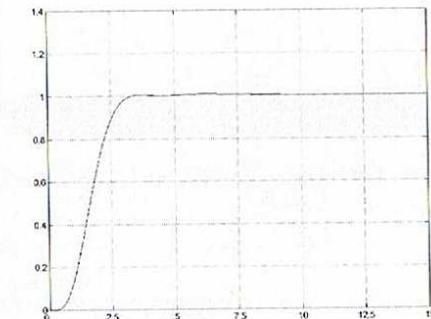
Nr	Kryterium	K_R	T_i	T_D	$\int e^2 dt$	$\int e dt$
1	ZN wg (9.15)	3,57	1,64	0,41	1,34	2,59
2	ZN wg (9.16)	1,35	2,2	0,55	1,31	1,70
3	$\min \int e^2 dt$ (9.14)	3,5	3,24	0,71	0,978	1,68
4	$\min \int e dt$ (9.14)	2,5	3,24	0,65	1,056	1,45

Na rysunkach 9.14–9.17 przedstawiono przebieg czterech odpowiedzi $y(t)$ układu V rzędu dla czterech różnych nastaw regulatorów PID w kolejności z tabeli 9.4.

Widać, że nastawy PID według klasycznej metody Zieglera–Nicholsa z aproksymacją Klopfmullera, czyli według modelu (9.15), gwarantują „tłumienie do $1/4$ ” (rys. 9.14). Nastawy oparte na lepszej modelowej aproksymacji obiektu (9.16), dają przebiegi $y(t)$ tłumione, z krótszym czasem regulacji T_{95} , o mniejszej wartości całki z modułu błędu (rys. 9.15).

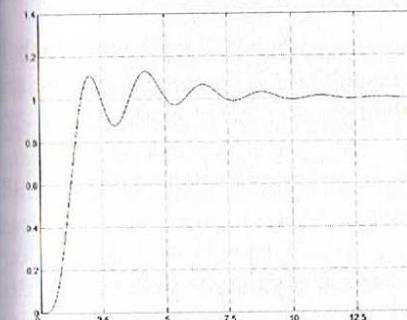
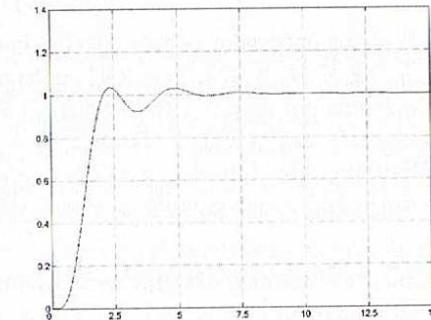


Rys. 9.14. Nastawy ZN według (9.15)



Rys. 9.15. Nastawy ZN według (9.16)

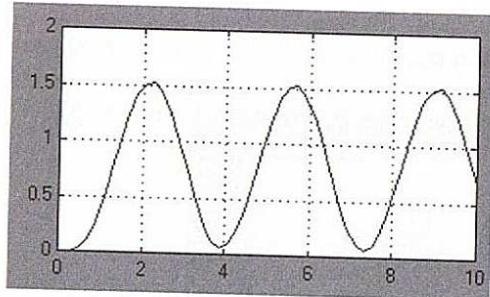
Nastawy PID otrzymane z minimalizacji całki z kwadratu błędu, dają przebiegi bardziej oscylacyjne niż według minimalizacji całki z modułu błędu (rys. 9.16 i 9.17).

Rys. 9.16. Nastawy według $\min \int e^2 dt$ Rys. 9.17. Nastawy według $\min \int |e| dt$

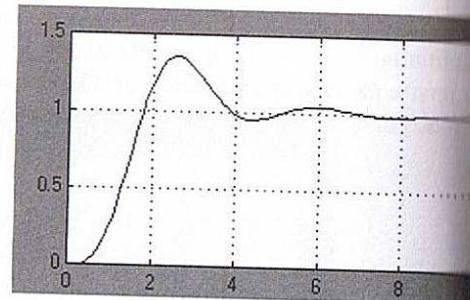
9.7.2. Druga metoda Zieglera–Nicholsa dla układów ciągłych

W przypadku gdy otrzymanie i zarejestrowanie odpowiedzi jednostkowej układu otwartego ze względów technologicznych nie jest możliwe, Ziegler i Nichols zaproponowali inną inżynierską metodę wyliczenia nastaw regulatora PID. Eksperyment polega na przyjęciu parametru stałej całkowania $T_I = \infty$ i różniczkowania $T_D = 0$ (eliminacja bloku całkowania i różniczkowania), a następnie na zwiększaniu wzmocnienia regulatora w układzie zamkniętym aż do uzyskania drgań ustalonych (granica stabilności). Po odnotowaniu krytycznej wartości wzmocnienia regulatora K_{kr} i otrzymanego okresu drgań P odpowiedzi $y(t)$ nastawy określone są z tabeli 9.5.

W podanym przykładzie obiektu (9.14) parametry krytyczne K_{kr} i P można otrzymać metodą prób w pakiecie Simulink lub dokładnie z granicy stabilności wynikającej z kryterium Hurwicza. Wynoszą one $K_{kr} = 3,884$ i $P = 3,5$. Przebieg krytyczny widać na rysunku 9.18.



Rys. 9.18. Krytyczny współczynnik wzmacnienia



Rys. 9.19. Nastawy PID według tabeli 9.5

Tabela 9.5
Proponowane nastawy ZN2

Regulator	K	T_I	T_D
P	$0,50 K_{kr}$	–	–
PI	$0,45 K_{kr}$	$P/1,2$	–
PID	$0,60 K_{kr}$	$P/2$	$P/8$

Przebieg optymalny odpowiedzi na skok jednostkowy dla nastaw PID uzyskanych z tabeli 9.5 ($K_{kr} = 2,33$, $T_I = 1,75$, $T_D = 0,43$) pokazany jest na rysunku 9.19. Przebieg daje mniejsze powierzchnie pól $\int \epsilon^2 dt = 1,164$ i $\int |\epsilon| dt = 1,791$ i wizualnie sprawia lepsze wrażenie niż uzyskany z pierwszej metody ZN (tab. 9.4).

Inne kryteria strojenia, takie jak optymalnego modułu i kryteria całkowe dla układów z opóźnieniem, przedstawione są w pracy [9GOH].

9.7.3. Metody Zieglera–Nicholsa dla układów dyskretnych

Poniżej podajemy za literaturą [9ISR] odpowiedniki dwóch metod wzorów Zieglera–Nicholsa dla układów dyskretnych, do stosowania we wzorze na równanie różnicowe regulatora PID (6.3).

Metoda I oparta jest na odpowiedzi na skok jednostkowy. Czas dyskretyzacji oznaczamy Δ (tab. 9.6).

Tabela 9.6
Proponowane nastawy ZN1

Regulator	K_R	Δ/T_I	T_D/Δ
P	$T / (\tau + \Delta)$	–	–
PI	$\frac{0,9T}{\tau + \Delta/2} - \frac{0,135T\Delta}{(\tau + \Delta/2)^2}$	$\frac{0,27T\Delta}{K(\tau + \Delta/2)^2}$	–
PID	$\frac{1,2T}{\tau + \Delta} - \frac{0,3T\Delta}{(\tau + \Delta/2)^2}$	$\frac{0,6T\Delta}{K(\tau + \Delta/2)^2}$	$\frac{0,5T}{K\Delta}$

Metoda II oparta jest na krytycznej wartości wzmacnienia K_{kr} i okresie drgań P odpowiedzi (tab. 9.7).

Tabela 9.7

Proponowane nastawy ZN2

Regulator	K_R	Δ/T_I	T_D/Δ
P	$K_{kr}/2$	–	–
PI	$0,45 K_{kr} (1 - 0,6\Delta/P)$	$0,54 \frac{K_{kr}}{K} \frac{\Delta}{P}$	–
PID	$0,6 K_{kr} (1 - \Delta/P)$	$1,2 \frac{K_{kr}}{K} \frac{\Delta}{P}$	$\frac{3}{40} \frac{K_{kr}}{K} \frac{P}{\Delta}$

W pracy [9ISR] można również znaleźć omówienie wielu innych typów regulatorów dyskretnych.

9.8. Strojenie regulatorów w zadaniach z niepewnością

9.8.1. Zakłócenia najbardziej niebezpieczne

Projektant układu regulacji powinien zdecydować, czy rozwiązuje zadanie najlepszej likwidacji wpływu zakłócenia na zmienne procesowe (zadanie stabilizacji), czy zadanie najlepszego śledzenia przez zmienne procesowe zmieniającego się punktu pracy, czyli zadanej wartości referencyjnej (zadanie nadążania). Ze względu na powszechność występowania zakłóceń, optymalizacja pierwszego zadania wydaje się istotniejsza. Wyniki optymalizacji parametrycznej dla tego zadania będą zależały od miejsca lokalizacji zakłócenia oraz jego kształtu. Jeśli charakterystyki sygnału referencyjnego i zakłócającego są podobne, to można sformułować spostrzeżenie, że najbardziej niebezpiecznym punktem przyłożenia zakłócenia jest wejście do układu ze sprzężeniem zwrotnym, a najmniej niebezpiecznym – wyjście układu. Stąd tłumienie zakłóceń będzie tym silniejsze, im dalej od wejścia znajdować się będzie element, na który ono oddziaływa. Z reguły, jeśli zakłócenie nie jest mierzone, to przyjmuje się typowe jego kształty, takie jak $\delta(t)$, $1/t$, $\sin(\omega t)$, biały szum. Są jednak takie procesy, w których projektanci powinni rozważyć nietypowe kształty zakłóceń. W obiektach, w których względy bezpieczeństwa odgrywają największą rolę, na wszelki wypadek powinno się rozwiązywać zagadnienia stabilizacji w obecności zakłóceń najbardziej niebezpiecznych. Do takich obiektów należą np. elektrownie jądrowe, samoloty pasażerskie, rakiety.

Określenie „zakłócenia najbardziej niebezpieczne” (*most dangerous disturbances*) lub „zakłócenia najgorsze” (*worst disturbances*) implikuje zarówno potrzebę określenia klasy sygnałów (funkcji), w której takie zakłócenie będzie poszukiwane, jak i określenie kryterium, według którego będzie oceniany wpływ tych zakłóceń. Tego typu pesymistyczne podejście ma dwojakie uzasadnienie. Po pierwsze będzie stosowane w przypadku zadań, które noszą cechy gry z inteligentnym przeciwnikiem. Przeciwnik zna nasz obiekt (obserwuje jego wyjścia) i stara się poprzez zakłócenie wyrządzić największą szkodę. Przestrajanie

regulatora ma zminimalizować efekt jego działania. Dla każdych nastaw regulatora istnieje inne zakłócenie najgorsze. Regulator optymalny wynika z rozwiązania zadania typu min–max – będzie najlepszy w najgorszym przypadku. Drugie uzasadnienie rozważania przypadku najgorszego może wynikać z faktu braku innych informacji o zakłóceniu, poza ogólniejszą cechą – ograniczonoscią jego amplitud lub energii.

W zagadnieniach „cywilnych” – gdy zakłócenia generowane są przez otaczające środowisko, zakłócenia najgorsze mogą się nie pojawić, ale przyjęcie pesymistycznej strategii strojenia pozwala ocenić granice możliwości układu regulacji, jeśli taki nietypowy przypadek zakłóceń będzie jednak miał miejsce. Jako przykładowe procesy można wymienić:

- nakładanie się odchyleń w analogowych przyrządach nawigacyjnych (żyrokompasach) przy szybko manewrującym statku, gdy częstotliwość kołyseń zbiega się z częstotliwością rezonansową przyrządu; pierwsze prace na ten temat publikowano już w 1939 r. [9BUB];
- drgania skrzydeł samolotu spowodowane wibracjami podwieszonych silników, przy ich rezonansowej szybkości obrotowej;
- lawinowe narastanie obciążenia połączonych sieci energetycznych i ich odłączenie w przypadku niekorzystnego zbiegu okoliczności kilku awarii – przykład awaria w USA w 2003 r. (*blackout*);
- „kołyseń” zespołu turbogeneratora synchronicznego; projektowanie układów regulacji dla najgorszych sygnałów sinusoidalnych wiąże się z minimalizacją normy systemu w przestrzeni H^∞ (maksimum modułu transmitancji widmowej).

Problematyce projektowania regulatorów w warunkach niepewności poświęcone były prace autora [9BYW1, 9BYW2, 9BYW3]. Tam też zamieszczona jest obszerna literatura na ten temat. W takich zadaniach decydującym warunkiem nakładanym na system jest stabilność w sensie BIBO (podrozdz. 7.1.2). Czułość rozwiązań na małe zmiany prawej strony równania różniczkowego, przy danych warunkach początkowych i przy założeniu, że zaburzenia powodują sygnały z klasy funkcji o ograniczonej amplitudzie, w pięćdziesiąt lat później pracach Lapunowa, badali B. Bułagakow [9BUB], I.G. Malkin [9MAI], K.P. Persidzki [9PEK], N.N. Krasowski [9KRN]. Odchyłki najgorsze dla klasy sygnałów o ograniczonej wartości średnio kwadratowej badał E.A. Barbaszin [9BAE]. Jak wspomniano, zadania tego typu prowadzą się do rozwiązywania problemów minimalno-normowych. Wartość funkcjonalu opisującego wyjście układu, określonego na kuli sygnałów zakłócających (ograniczonych), będzie można oszacować poprzez twierdzenia analizy funkcjonalnej o osiąganiu kresu górnego funkcjonalu w przestrzeni Banacha (rozdz. 20). Poniżej przytoczymy pewne wyniki

9.8.2. Oszacowania maksymalnych odchyłek wyjścia

Przeprowadzona zostanie analiza odchyłek maksymalnych dla przestrzeni funkcji $C[0, \infty)$. Zakładamy, że zakłócenia $z(t)$ należą do klasy funkcji ciągkich i ograniczonych.

Dany jest obiekt liniowy opisany równaniem

$$\sum_{i=0}^n a_i y^{(i)}(t) = \sum_{i=0}^m b_i z^{(i)}(t) \quad (9.17)$$

Przy zerowych warunkach początkowych i dla $n = m$, odpowiedź skokowa może mieć jedynie skok w chwili $t = 0$, równy $p_0 = b_n/a_n$.

Dla zakłócenia $z \in C[0, t_k]$, o ograniczonej amplitudzie $\max |z(t)| \leq 1$, zachodzi oszacowanie

$$\|y\|_c \leq \|F\| \|z\|_c,$$

gdzie operator F związany z obiektem (9.17) dany jest całką Stieltjesa, a jego norma jest wyrażana funkcji $h(t)$

$$[F z](t) = \int_0^t z(\tau) dh(\tau).$$

Stąd

$$y(t) = p_0 z(t) + \int_0^t g(t-\tau) z(\tau) d\tau \quad (9.18)$$

Wartość maksymalną $y(t)$ otrzymuje się dla sygnału $z(t)$ przedziałami stałego

$$z(t-\tau) = \operatorname{sgn}[g(\tau)] \quad (9.19)$$

Jeżeli $m < n$, to $h(t)$ redukuje się do funkcji absolutnie ciągłej, dla której z twierdzenia Lebesgue'a wynika, że

$$\int_0^t g(\tau) d\tau = h(t) - h(0),$$

$$\max_z y(t_k) = \int_0^{t_k} |g(\tau)| d\tau.$$

Rozwiązyując równanie $g(\tau) = 0$, znajduje się chwile t_1, t_2, \dots , w których $g(\tau)$ zmienia znak i które określają przedziały monotoniczności funkcji $h(t)$, a jednocześnie chwile przejęć sygnału najbardziej niebezpiecznego.

$$\begin{aligned} y_{\max}(t_k) &= \int_0^{t_1} g(\tau) d\tau - \int_{t_1}^{t_2} g(\tau) d\tau + \dots + \int_{t_{k-1}}^{t_k} g(\tau) d\tau = \\ &= h(t_1) - h(0) - [h(t_2) - h(t_1)] + \dots + [h(t_k) - h(t_{k-1})]. \end{aligned}$$

Jeżeli $h(0) = 0$, to

$$y_{\max}(t_k) = 2 \left[h(t_1) - h(t_2) + h(t_3) - h(t_4) + \dots + (-1)^{k+1} h(t_k) \right].$$

Powyższy zbiór określa wahanie funkcji $h(t)$

$$y_{\max}(t_k) = V_o^{t_k}[h] = \|F\|.$$

Z własnością wahania wynika, że jego maksimum zostanie osiągnięte dla $t_k \rightarrow \infty$.

Jak wiadomo (rozdz. 20, aneks 2), przestrzeń funkcji $C[0, 1]$ nie jest refleksywna, stąd nie istnieje w niej element z , który realizuje normę funkcjonalu. Sygnał najbardziej niebezpieczny jest funkcją nieciągłą (9.19). Jednak prawdziwe jest spostrzeżenie, że dla każdej funkcji h o wahaniu ograniczonym, określonej na przedziale $[0, 1]$ i dla

$$\forall \varepsilon > 0 : \exists z \in C[0, 1], \text{ takie że } |z| = 1$$

oraz

$$\int_0^1 z(\tau) dh(\tau) + \varepsilon \geq V_o^1[h].$$

Czyli normę dowolnego funkcjonału liniowego, ciągłego w $C[0, 1]$ można zrealizować z dowolną dokładnością. Fakt ten wynika również z możliwości rozszerzenia omawianego funkcjonału z zachowaniem normy, na przestrzeń funkcji istotnie ograniczonych $L^\infty[0, 1]$ (twierdzenie Hahna–Banacha).

Wyprowadzenie wzoru na normę funkcjonału dla systemu II rzędu

Dana jest transmitancja

$$G(s) = \frac{b_1 s + b_o}{s^2 + a_1 s + a_o} = \frac{b_1}{a_o} \frac{\omega^2 (s + 1/\alpha)}{s^2 + 2\xi\omega s + \omega^2},$$

Współczynniki $0 \leq \alpha = b_1/b_o < \infty$, $\omega = \sqrt{a_o}$ – częstotliwość drgań własnych. Wartość $h_\infty = b_1/a_o$ oznacza stan ustalony dla $u = 1(t)$. Współczynnik tłumienia $\xi = 0,5 a_1 / \sqrt{a_o}$.

Układ jest oscylacyjny dla $0 < \xi < 1$, a przetłumiony dla $\xi \geq 1$. Ze wzorów podanych w podrozdziale 4.6.1 wynikają czasy przełączenia t_n najgorszego sygnału wejściowego, dla $\xi < 1$, dającego y_{\max} :

$$t_n = \frac{(n-1)\pi + \Theta}{\sqrt{1-\xi^2}} : \quad \operatorname{tg} \Theta = \frac{\alpha \omega \sqrt{1-\xi^2}}{\alpha \xi \omega - 1} : \quad A = h_\infty \sqrt{\frac{\alpha^2 \omega^2 - 2\alpha \omega \xi + 1}{1-\xi^2}}.$$

Oznaczając przez γ dekrement tłumienia, $\gamma = e^{-d\pi} < 1$; $d = \frac{\xi}{\sqrt{1-\xi^2}}$, otrzymuje się:

$$\begin{aligned} y_{\max} &= V_o^\infty[h] = \left(h_\infty + A e^{-d\Theta} \sin \psi \right) - \\ &- \left(h_\infty + A e^{-d\Theta} \cdot \gamma \cdot \sin(\pi + \psi) - h_\infty - A e^{-d\Theta} \sin \psi \right) + \\ &+ \left(h_\infty + A e^{-d\Theta} \cdot \gamma^2 \sin(2\pi + \psi) - h_\infty - A e^{-d\Theta} \gamma \sin(\pi + \psi) \right) - \dots = \\ &= h_\infty + 2A e^{-d\Theta} \sin \psi \cdot \frac{1}{1-\gamma} = h_\infty \left[1 + 2\sqrt{\alpha^2 \omega^2 - 2\alpha \omega \xi + 1} \cdot \frac{e^{-d\Theta}}{1-\gamma} \right]. \end{aligned} \quad (9.20)$$

Jeżeli w liczniku transmitancji brak operatora różniczkowania, tzn. $\alpha = 0$, to $\Theta = \pi$,

$$V_o^\infty[h] = h_\infty \left[1 + \frac{2\gamma}{1-\gamma} \right] = h_\infty \left[\frac{1+\gamma}{1-\gamma} \right] = h_\infty \operatorname{ctgh} \left(\frac{d\pi}{2} \right) \quad (9.21)$$

Dla przypadku przetłumionego, $\xi \geq 1$ z $\alpha \neq 0$ (jedno przeregulowanie), po przekształceniach otrzymuje się wzór

$$\begin{aligned} V_o^\infty[h] &= h(t_1) - [h_\infty - h(t_1)] = 2h(t_1) - h_\infty = \\ &= h_\infty \left[1 + 2\sqrt{\alpha^2 \omega^2 - 2\xi\omega\alpha + 1} \cdot e^{-\bar{d}\Theta} \right] \end{aligned} \quad (9.22)$$

$$\text{gdzie } \bar{d} = \frac{\xi}{\sqrt{\xi^2 - 1}} : \quad \operatorname{tgh} \Theta = \frac{\alpha \omega \sqrt{\xi^2 - 1}}{\alpha \xi \omega - 1}.$$

Jeśli $\xi \geq 1$ i $\alpha = 0$ (brak przeregulowania, $t_1 = 0$), wtedy najgorszym sygnałem jest skok jednostkowy, a norma jest równa h_∞ .

$$V_o^\infty[h] = h_\infty \quad (9.23)$$

Przykład 9.9

Dla testowego systemu podwójnie całkującego z przykładu 9.2 i regulatora PD

$$G_R = K + 2\xi\sqrt{K} s,$$

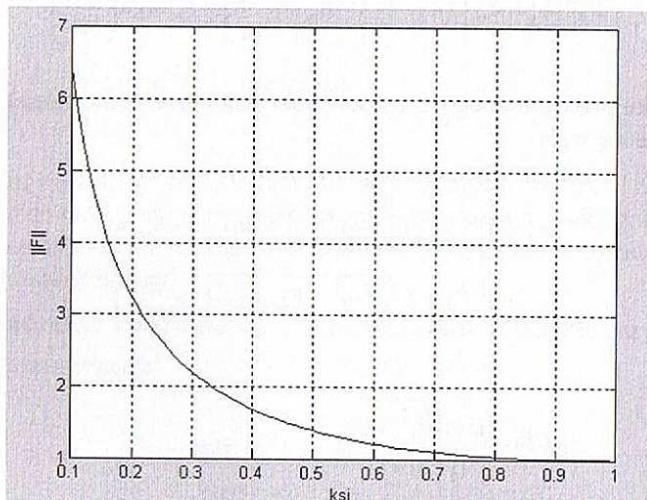
transmitancja systemu zamkniętego w torze zakłócenia wynosi

$$G_Z(s) = \frac{1}{s^2 + 2\xi\sqrt{K}s + K}.$$

Zgodnie ze wzorem (9.21) wahanie funkcji $h(t)$, która jest odpowiedzią skokową tego systemu, określa maksymalną wartość odpowiedzi systemu na zakłócenie najgorsze typu signum (9.19) i wyraża się poniższym wzorem

$$y_{\max} = \frac{1}{K} \frac{\frac{1}{1+e^{-\frac{\xi}{\sqrt{1-\xi^2}}\pi}} - \frac{\xi}{\sqrt{1-\xi^2}\pi}}{\frac{\xi}{\sqrt{1-\xi^2}\pi} - \frac{1-e^{-\frac{\xi}{\sqrt{1-\xi^2}}\pi}}{1+e^{-\frac{\xi}{\sqrt{1-\xi^2}}\pi}}} = \|F\| \quad (9.24)$$

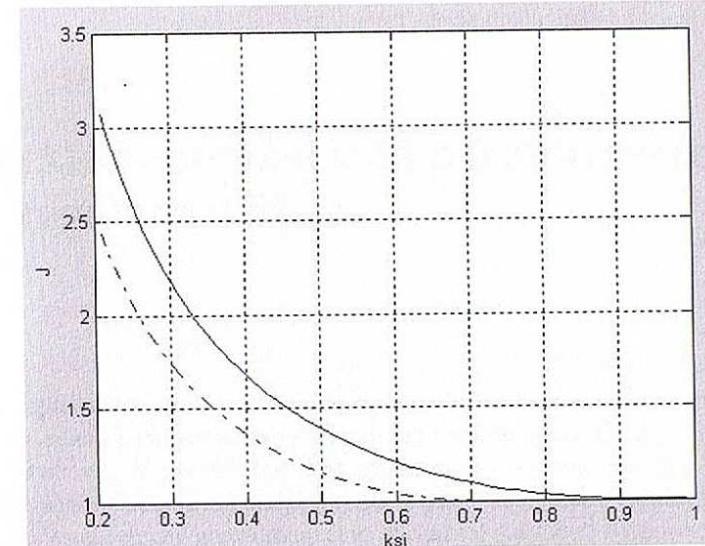
Dla współczynnika $\xi > 1$ norma z F , pozostaje na poziomie 1. Kształt funkcji y_{\max} dla $K = 1$ w funkcji ξ przedstawiony jest na rysunku 9.20.



Rys. 9.20. Przebieg normy systemu w funkcji ξ

Maksymalna odchyłka y_{\max} (9.24) spowodowana sygnałem najbardziej niebezpiecznym typu signum (*bang-bang*) o amplitudzie $|u| = 1$ jest większa niż spowodowana sygnałem najbardziej niebezpiecznym z klasy sygnałów sinusoidalnych (o częstotliwości rezonansowej ω_r) i o takiej samej jednostkowej amplitudzie, a określona przez $|G(\omega_r)| = 1/(2\xi\sqrt{1-\xi^2})$ (wg rys. 4.53). Dla systemu II rzędu przedstawiono to na rysunku 9.21, ($|G(\omega_r)|$ – linia przerywana).

Wzmocnienie sygnałów sinusoidalnych obejmuje tylko $0 < \xi < 0,707$.

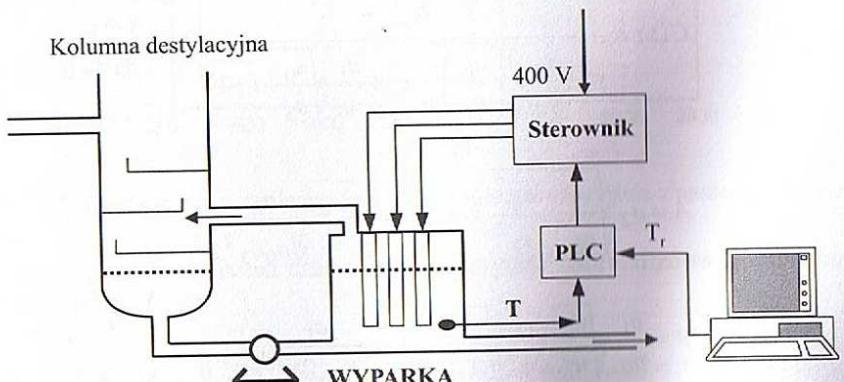


Rys. 9.21. Porównanie maksymalnych odchylek dla dwóch klas funkcji sterujących
Objaśnienia w tekście

Kombinacja wskaźnika (9.24) i dowolnego wskaźnika rosnącego w zakresie $0 < \xi < 1$ (np. czasu regulacji T_{95}) da sumarycznie funkcję wypukłą posiadającą minimum, dla której można określić optymalną wartość ξ^* .

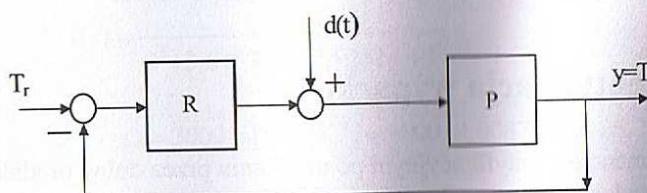
10. Przykład optymalizacji parametrycznej dla regulatora PI

Jako kompleksowy przykład zadania optymalizacji parametrycznej rozważymy rozwiązany przez autora problem syntezy układu regulacji temperatury (rys. 10.1) w wyparce olejowo-elektrycznej. W przykładzie tym zastosowany zostanie model obiektu III rzędu i regulator PI. Celem układu regulacji jest stabilizacja temperatury dolnego produktu kolumny destylacyjnej (cieczy wyczerpanej) w obecności zakłóceń skokowych mocy pochodzących od takich efektów jak awaria grzałki czy wyłączenie jednej z faz trójfazowego układu zasilania.



Rys. 10.1. Schemat techniczny układu regulacji

Wartość zadana temperatury T_r jest przesyłana z komputera nadzorującego. Sygnał wyjściowy z regulatora wchodzi na tyrystorowy człon wykonawczy (sterownik), który dawkuje prąd do dziewięciu grzałek (3 grzałki /fazę). Schemat blokowy przedstawiono na rysunku 10.2.



Rys. 10.2. Schemat blokowy

10.1. Transmitancja wyparki

Dla rozwiązania zadania syntezy potrzebna jest transmitancja układu zamkniętego pośród wyjściem i sygnałem zakłócającym. Na podstawie eksperymentu wybrano model III rzędu dla aproksymacji dynamiki aparatu wypartego; układ III rzędu omówiono szczegółowo w podrozdziale 4.7.

Ogólna postać transmitancji

$$G(\hat{s}) = \frac{K_o}{a_3 \hat{s}^3 + a_2 \hat{s}^2 + a_1 \hat{s} + a_0} = \frac{\frac{K_o}{a_o}}{\frac{a_3}{a_o} \hat{s}^3 + \frac{a_2}{a_o} \hat{s}^2 + \frac{a_1}{a_o} \hat{s} + 1},$$

po wprowadzeniu nowego przeskalowanego operatora Laplace'a s

$$\sqrt[3]{\frac{a_3}{a_0}} \cdot \hat{s} = s, \text{ czyli } \hat{s}^2 = \sqrt[3]{\left(\frac{a_0}{a_3}\right)^2} \cdot s^2,$$

przyjmuje postać znormalizowaną układu tylko z trzema parametrami K_p , A i B .

$$G(s) = \frac{K_p}{s^3 + \frac{a_2}{a_0} \sqrt[3]{\left(\frac{a_0}{a_3}\right)^2} \cdot s^2 + \frac{a_1}{a_0} \sqrt[3]{\frac{a_0}{a_3}} \cdot s + 1} = \frac{K_p}{s^3 + As^2 + Bs + 1},$$

$$A = a_2 \sqrt[3]{\frac{1}{a_0^3} \cdot \frac{a_0^2}{a_3^2}} = a_2 \sqrt[3]{\frac{1}{a_0 a_3^2}} = \frac{a_2}{\sqrt[3]{a_0 a_3^2}},$$

$$B = \frac{a_1}{a_0} \sqrt[3]{\frac{a_0}{a_3}} = a_1 \sqrt[3]{\frac{1}{a_0^3} \cdot \frac{a_0}{a_3}} = \frac{a_1}{\sqrt[3]{a_0^2 a_3}},$$

$$\Omega = \sqrt[3]{\frac{a_0}{a_3}}, \quad \hat{s} = \sqrt[3]{\frac{a_0}{a_3}} \cdot s = \Omega s.$$

To podstawienie w dziedzinie częstotliwości ma swój odpowiednik w przeskalowaniu osi czasu. Można wprowadzić bezwymiarową zmienną $t = \hat{t} \cdot \Omega$.

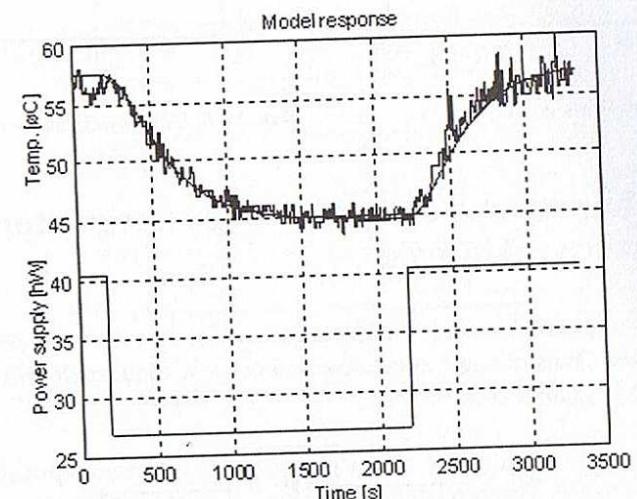
10.2. Identyfikacja wyparki

W eksperymencie identyfikacyjnym po uzyskaniu przez dolny produkt ustalonej temperatury 58°C wyłączonego trzy z dziewięciu grzałek, co interpretowane było jako ujemny skok maksymalnej mocy zasilania ($14 \text{ h}\cdot\text{W}$) (rys. 10.3). Po zebraniu danych eksperymental-

nych użyty został przybornik Identification z pakietu Matlab Toolbox do otrzymania współczynników transmitancji układu otwartego

$$\hat{G}(\hat{s}) = \frac{0,9}{1500\,000 \hat{s}^3 + 50\,000 \hat{s}^2 + 450 \hat{s} + 1} \quad (10.1)$$

Biegunki transmitancji mają wartość: $\hat{s}_1 = -0,02$, $\hat{s}_2 = -0,021$, $\hat{s}_3 = -0,0033$. Odpowiedź tego modelu (10.1) prezentuje rysunek 10.3.



Rys. 10.3. Przebieg mocy zasilania, temperatura mierzona i wyjście z modelu (linia ciągła)

Dokonując skalowania osi czasu, można otrzymać znormalizowaną transmitancję ze współczynnikiem $a_3 = 1$.

$$\Omega = \sqrt[3]{\frac{a_0}{a_3}} = \frac{1}{\sqrt[3]{a_3}} = 0,0087358, \quad \frac{1}{\Omega} = 114,471424 = \sqrt[3]{a_3}, \quad \hat{s} = \sqrt[3]{\frac{a_0}{a_3}} \cdot s = \Omega s,$$

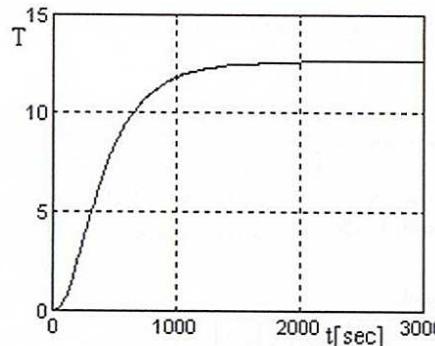
$$A = \frac{a_2}{\sqrt[3]{a_3^2}} = 3,815714142 = \frac{a_2}{a_3} \sqrt[3]{a_3}, \quad B = \frac{a_1}{\sqrt[3]{a_3}} = 3,9311121 = \frac{a_1}{a_3} \sqrt[3]{a_3^2}.$$

Postać znormalizowana ma trzy biegunki: $s_1 = -2,2894$, $s_2 = -1,1447$, $s_3 = -0,3775$:

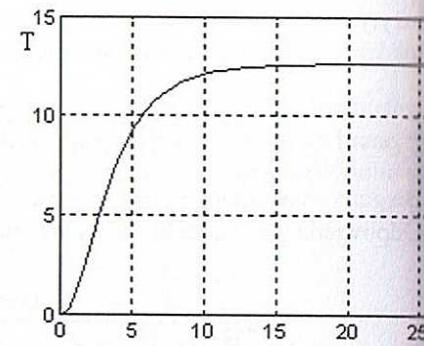
$$G(s) = \frac{0,9}{s^3 + 3,8157 s^2 + 3,9311 s + 1} \quad (10.2)$$

$$t = \hat{t} \cdot \Omega = 3000 [\text{s}] \cdot 0,0087358 = 26,2.$$

Odpowiedź transmitancji (10.2) na przeskalowanej osi czasu w przedziale $[0, 26,2]$ (rys. 10.5) ma ten sam kształt jak na rysunku 10.4.



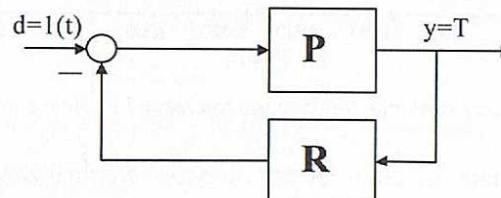
Rys. 10.4. Wyjście układu (10.1)



Rys. 10.5. Odpowiedź układu otwartego (10.1)

10.3. Stabilność układu zamkniętego z regulatorem PI dla toru zakłócenia

Zakłada się, że w punkcie pracy temperatura wynosi $T = T_r$. W chwili $t = 0$ wystąpi zakłocenie skokowe. Transmitancję układu zamkniętego, w torze zakłócenia, można otrzymać ze schematu na rysunku 10.6.



Rys. 10.6. Schemat blokowy układu zamkniętego

Transmitancja układu zamkniętego G_z z regulatorem PI dana jest wzorem:

$$G_R = K_R + \frac{\alpha_R}{s}, \quad G_Z = \frac{G_P}{1+G_P \cdot G_R} = \frac{Y(s)}{d(s)},$$

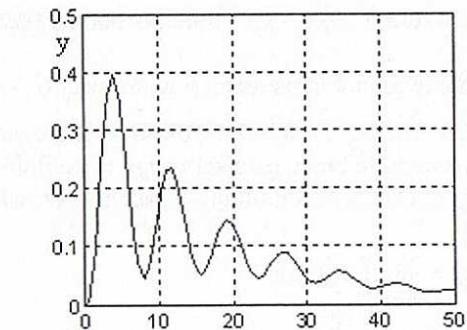
$$G_Z(s) = \frac{\frac{K_P}{s^3 + As^2 + Bs + 1}}{1 + \frac{K_P}{s^3 + As^2 + Bs + 1} \left(K_R + \frac{\alpha_R}{s} \right)} = \frac{K_P}{s^3 + As^2 + Bs + 1 + K_P \left(K_R + \frac{\alpha_R}{s} \right)},$$

$$G_Z(s) = \frac{K_P s}{s^4 + As^3 + Bs^2 + s + K_p K_R s + K_P \alpha_R},$$

Oznaczając $K_P \cdot K_R = K$, $K_P \cdot \alpha_R = \alpha$, otrzymujemy

$$G_Z(s) = \frac{K_P s}{s^4 + As^3 + Bs^2 + (1+K)s + \alpha}.$$

Przebiegi dynamiczne błędu dla przykładowego regulatora: $K_p = 1$, $K = 3$, $\alpha = 0,2$, $A = 6$, $B = 2$ są pokazane na rysunku 10.7.



Rys. 10.7. Przykładowy przebieg błędu

Dla układu zamkniętego można podać warunki konieczne i wystarczające stabilności układu w funkcji parametrów A , B , K i α . Z kryterium Hurwitza otrzymuje się wyznacznik:

$$\Delta = \begin{vmatrix} A & 1 & 0 & 0 \\ K+1 & B & A & 1 \\ 0 & \alpha & K+1 & B \\ 0 & 0 & 0 & \alpha \end{vmatrix} > 0,$$

$$A > 0, \alpha > 0, AB - (K+1) > 0 \Rightarrow K < AB - 1,$$

$$(K+1)^2 - (K+1)AB + \alpha A^2 < 0 \Rightarrow \hat{K} = K + 1,$$

$$\hat{K}_1 = \frac{AB + A\sqrt{B^2 - 4\alpha}}{2}, \quad \hat{K}_2 = \frac{AB - A\sqrt{B^2 - 4\alpha}}{2} \quad (10.3)$$

Układ zamknięty będzie stabilny, gdy dla danej wartości α wzmacnienie regulatora PI będzie z przedziału $\hat{K}_1 < \hat{K} < \hat{K}_2$. Z zależności $B^2 - 4\alpha > 0$ wynika warunek na przedział stabilności na α : $0 < \alpha < \frac{B^2}{4}$. Dla każdego α z tego przedziału ze wzoru (10.3) można otrzymać K_1 i K_2 .

10.4. Sformułowanie zadania optymalizacji parametrycznej

Regulator optymalny układu zamkniętego powinien gwarantować najlepszą jakość tłumienia błędu przejściowego spowodowanego zakłóceniem skokowym. Parametry optymalne regulatora będą wyliczane z analitycznych formuł i procedur numerycznych minizujących wybrane wskaźniki jakości.

Rozważane będą cztery wskaźniki:

- 1) całka z kwadratu błędu,
- 2) całka z ważonej sumy kwadratu błędu i kwadratu pochodnej błędu,
- 3) całka z modułu błędu,
- 4) kryterium H_∞ (maksymalny moduł transmitancji widmowej $G_Z(j\omega)$).

Z twierdzenia Parsevala (rozdz. 9) wynika postać analityczna dwóch pierwszych wskaźników jakości: całki z kwadratu błędu przejściowego i uogólnionego kryterium jakości ze współczynnikiem wagi β . Dla nich formułuje się zadania i rozwiązuje je analitycznie

$$\min_{K,\alpha} J_{20} = \min_{K,\alpha} \int_0^\infty \varepsilon_p^2(t) dt \quad (10.4)$$

$$\min_{K,\alpha} J_{21} = \min_{K,\alpha} \int_0^\infty (\varepsilon_p^2(t) + \beta \dot{\varepsilon}_p^2(t)) dt \quad (10.5)$$

$$\min_{K,\alpha} J_1 = \min_{K,\alpha} \int_0^\infty |\varepsilon_p(t)| dt \quad (10.6)$$

$$\min_{K,\alpha} J_{H\infty} = \min_{K,\alpha} \max_{\omega} |G_Z(\omega)| \quad (10.7)$$

Minimalizacja kryterium trzeciego i czwartego będzie przeprowadzana numerycznie.

10.5. Analityczne formuły dla kwadratowych wskaźników jakości

10.5.1. Optymalny regulator PI według całkowego wskaźnika kwadratu błędu

Poniżej wyprowadzone zostaną wzory dla rozważanego modelu (10.1).

Odpowiedź skokowa przybiera postać

$$L[y(t)] = Y(s) = \frac{K_p s}{s^4 + As^3 + Bs^2 + (k+1)s + \alpha} \cdot \frac{1}{s} \quad (10.8)$$

gdzie: $K_p \cdot K_R = K$, $K_p \cdot \alpha_R = \alpha$.

Na podstawie wzorów z podrozdziału 9.3 mamy: $n = 4$, $m = 1$,

$$\begin{aligned} b_0 &= 0, & a_0 &= \alpha, & B_0 &= 0, \\ b_1 &= K_p, & a_1 &= 1+K, & B_1 &= K_p^2, \\ b_2 &= 0, & a_2 &= B, & & \\ b_3 &= 0, & a_3 &= A, & a_4 &= 1, \end{aligned}$$

wyznaczniki przyjmują postać:

$$\Delta = \begin{vmatrix} a_0 & -a_2 & a_4 & 0 \\ 0 & a_1 & -a_3 & 0 \\ 0 & -a_0 & a_2 & -a_4 \\ 0 & 0 & -a_1 & a_3 \end{vmatrix}, \quad \Delta_1 = \begin{vmatrix} a_0 & a_1 & a_4 & 0 \\ 0 & a_0 & -a_3 & 0 \\ 0 & 0 & a_2 & -a_4 \\ 0 & 0 & -a_1 & a_3 \end{vmatrix},$$

$$\begin{aligned} \Delta &= \begin{vmatrix} \alpha & -B & 1 & 0 \\ 0 & (1+K) & -A & 0 \\ 0 & -\alpha & B & -1 \\ 0 & 0 & -(1+K) & A \end{vmatrix} = \alpha \begin{vmatrix} (1+K) & -A & 0 \\ -\alpha & B & -1 \\ 0 & -(1+K) & A \end{vmatrix} = \\ &= \alpha \cdot (1+K) \cdot \begin{vmatrix} B & -1 \\ -(1+K) & A \end{vmatrix} - \alpha \cdot (-A) \cdot \begin{vmatrix} -\alpha & -1 \\ 0 & A \end{vmatrix} = \\ &= \alpha \cdot (1+K) \cdot [AB - (1+K)] + \alpha \cdot A(-\alpha \cdot A), \end{aligned}$$

$$\Delta = \alpha \cdot (1+K) \cdot AB - \alpha \cdot (1+K)^2 - \alpha^2 \cdot A^2.$$

Oznaczmy: $1+K = \bar{K}$, wtedy

$$\Delta = \alpha \cdot \bar{K} \cdot AB - \alpha \cdot \bar{K}^2 - \alpha^2 \cdot A^2,$$

$$\bar{K} = 1+K$$

$$\Delta_1 = \begin{vmatrix} \alpha & (1+K) & 1 & 0 \\ 0 & \alpha & -A & 0 \\ 0 & 0 & B & -1 \\ 0 & 0 & -(1+K) & A \end{vmatrix} = \alpha \begin{vmatrix} \alpha & -A & 0 \\ 0 & B & -1 \\ 0 & -(1+K) & A \end{vmatrix} = \alpha^2 (AB - \bar{K}),$$

$$\begin{aligned}
 J &= \frac{1}{2a_0^2\Delta} [B_0\Delta_0 + B_1\Delta_1 + \dots + B_m\Delta_m - 2b_0b_1\Delta] = \\
 &= \frac{K_p^2}{2\alpha^2(\alpha\bar{K}AB - \alpha\bar{K}^2 - \alpha^2A^2)} [\alpha^2(AB - \bar{K})] = \\
 &= \frac{K_p^2\alpha^2(AB - \bar{K})}{2\alpha^2(\alpha KAB - \alpha\bar{K}^2 - \alpha^2A^2)}, \\
 J_1 &= \frac{K_p^2(AB - \bar{K})}{2\alpha\bar{K}AB - 2\alpha\bar{K}^2 - 2\alpha^2A^2} \tag{10.9}
 \end{aligned}$$

Obliczamy optymalne \bar{K} i α z warunków koniecznych optymalności:

$$\frac{\delta J}{\delta \bar{K}} = 0, \quad \frac{\delta J}{\delta \alpha} = 0.$$

Oznaczając licznik i mianownik wskaźnika jakości przez LI i MI , otrzymujemy $J = \frac{LI}{MI}$. Po zastosowaniu wzoru na pochodną otrzymuje się:

$$\begin{aligned}
 \frac{\delta J}{\delta \bar{K}} &= \frac{-1(MI) - (AB - \bar{K})(2\alpha AB - 4\alpha\bar{K})}{MI^2} = 0, \\
 -MI - (AB - \bar{K})(2\alpha AB - 4\alpha\bar{K}) &= 0, \\
 -2\alpha\bar{K}AB + 2\alpha\bar{K}^2 + 2\alpha^2A^2 - (AB - \bar{K})(2\alpha AB - 4\alpha\bar{K}) &= 0, \\
 -\bar{K}AB + \bar{K}^2 + \alpha A^2 - (AB - \bar{K})(AB - 2\bar{K}) &= 0 \tag{10.10}
 \end{aligned}$$

$$\begin{aligned}
 \frac{\delta J}{\delta \alpha} &= \frac{-(AB - \bar{K})[2\bar{K}AB - 2\bar{K}^2 - 4\alpha A^2]}{MI^2} = 0, \\
 (AB - \bar{K})[2\bar{K}AB - 2\bar{K}^2 - 4\alpha A^2] &= 2(AB - \bar{K})[\bar{K}AB - \bar{K}^2 - 2\alpha A^2] = 0,
 \end{aligned}$$

$$\bar{K}(AB - \bar{K})^2 - (AB - \bar{K})2\alpha A^2 = 0,$$

$$\alpha = \frac{\bar{K}(AB - \bar{K})^2}{2A(AB - \bar{K})}, \quad \text{stąd dla } \bar{K} \neq AB, \quad \alpha = \frac{\bar{K}(AB - \bar{K})}{2A^2}.$$

Podstawiając wartość α do równania (10.10), otrzymujemy:

$$\begin{aligned}
 -\bar{K}AB + \bar{K}^2 + \frac{\bar{K}(AB - \bar{K})}{2A^2} A^2 - (AB - \bar{K})(AB - 2\bar{K}) &= 0, \\
 -2\bar{K}AB + 2\bar{K}^2 + \bar{K}AB - \bar{K}^2 - 2A^2B^2 + 6AB\bar{K} - 4\bar{K}^2 &= 0, \\
 -3\bar{K}^2 + 5AB\bar{K} - 2A^2B^2 &= 0,
 \end{aligned}$$

$$\Delta = 25A^2B^2 - 24A^2B^2 = A^2B^2,$$

$$\bar{K}_1^o = \frac{-5AB + \sqrt{A^2B^2}}{2(-3)}, \quad \bar{K}_1^o = \frac{2}{3}AB \Rightarrow K_1^o = \frac{2}{3}AB - 1,$$

$$\bar{K}_2^o = \frac{-5AB - \sqrt{A^2B^2}}{2(-3)}, \quad \bar{K}_2^o = AB \Rightarrow K_2^o = AB - 1,$$

$$\alpha = \frac{\bar{K}(AB - \bar{K})}{2A^2}, \quad \bar{K}_1^o = \frac{2}{3}AB, \quad \alpha_1^o = \frac{\frac{2}{3}AB \left(AB - \frac{2}{3}AB \right)}{2A^2} = \frac{\frac{2}{3}AB \left(\frac{1}{3}AB \right)}{2A^2} = \frac{B^2}{9},$$

$$\bar{K}_2^o = AB, \quad \alpha_2^o = 0 \quad (\text{przypadek układu na granicy stabilności}).$$

Ostatecznie optymalne parametry regulatora PI są dane formułami:

$$K^o = K_1^o = \frac{2}{3}AB - 1, \quad \alpha^o = \alpha_1 = \frac{B^2}{9}, \quad J^o = \frac{27}{2AB^3} \tag{10.11}$$

$$K_R^o = \frac{\frac{2}{3}AB - 1}{K_p}, \quad \alpha_R^o = \frac{B^2}{9K_p} \tag{10.12}$$

Jeśli amplituda skoku $u(t) = M 1(t)$, to wartość wskaźnika kwadratowego wynosi

$$J_1 = \frac{M^2 K_p^2 (AB - \bar{K})}{2\alpha\bar{K}AB - 2\alpha\bar{K}^2 - 2\alpha^2A^2}.$$

Dla obiektu nieznormalizowanego odpowiednie wzory mają postać

$$\begin{aligned} Y(\hat{s}) &= \frac{K_p \hat{s}}{a_3 \hat{s}^4 + a_2 \hat{s}^3 + a_1 \hat{s}^2 + (\bar{K} + 1) \hat{s} + \bar{\alpha}} \cdot \frac{1}{\hat{s}} = \\ &= \frac{\frac{MK_p}{a_3} \hat{s}}{\hat{s}^4 + \frac{a_2}{a_3} \hat{s}^3 + \frac{a_1}{a_3} \hat{s}^2 + \frac{(\bar{K} + 1)}{a_3} \hat{s} + \frac{\bar{\alpha}}{a_3}} \cdot \frac{1}{\hat{s}}. \end{aligned}$$

Wtedy:

$$\begin{aligned} \frac{\hat{K}^o + 1}{a_3} &= \frac{2}{3} \frac{a_2 a_1}{a_3 a_3} - 1 \Rightarrow \hat{K}_R^o = \frac{\frac{2}{3} \frac{a_2 a_1}{a_3} - 1}{K_p}, \\ \hat{\alpha}_R^o &= \left(\frac{a_1}{a_3} \right)^2 \frac{1}{9} \Rightarrow \hat{\alpha}_R^o = \frac{a_1^2}{9 a_3 K_p}. \end{aligned}$$

Zależności pomiędzy wzorami normalizowanymi i nienormalizowanymi, które należy użyć do eksperymentu, są następujące:

$$\begin{aligned} \hat{K}_{R \text{ exp}}^o &= K_R^o, \quad \hat{\alpha}_{R \text{ exp}}^o = \alpha_R^o \Omega \\ J_{\text{exp}} &= \frac{J_1 \cdot M^2 K_p^2}{\Omega} \end{aligned} \tag{10.13}$$

10.5.2. Optymalny regulator P według całkowego wskaźnika kwadratu błędu

Dla szczególnego przypadku regulatora proporcjonalnego P ($\alpha = 0$) zachodzą wzory

$$Y(s) = \frac{K_p}{s^3 + A s^2 + B s + 1 + K} \cdot \frac{1}{s} = \frac{K_p}{s^3 + A s^2 + B s + \bar{K}} \cdot \frac{1}{s}.$$

Ze wzorów z podrozdziału 9.3 mamy: $n = 3$, $m = 1$,

$$\begin{aligned} a_0 &= 1 + K, \quad b_0 = K_p, \quad B_0 = K_p^2, \\ a_1 &= B, \quad a_2 = A, \quad a_3 = 1, \quad \bar{K} = 1 + K, \end{aligned}$$

Wyznaczniki przyjmują postać:

$$\Delta = \begin{vmatrix} 1+K & -A & 0 \\ 0 & B & -1 \\ 0 & -(1+K) & A \end{vmatrix} = (1+K)[AB - (1+K)],$$

$$\Delta_0 = \begin{vmatrix} B & -A & 0 \\ 1+K & B & -1 \\ 0 & -(1+K) & A \end{vmatrix} = AB^2 - B(1+K) + A^2(1+K),$$

$$\begin{aligned} J &= \frac{1}{2a_0^2 \Delta} [B_0 \Delta_0 + B_1 \Delta_1 + \dots + B_m \Delta_m - 2b_0 b_1 \Delta] = \\ &= \frac{K_p^2 [AB^2 + (A^2 - B)\bar{K}]}{2\bar{K}^3 (AB - \bar{K})}. \end{aligned}$$

Z warunków koniecznych optymalności, $\frac{\delta J}{\delta \bar{K}} = 0$, otrzymuje się równanie

$$3(A^2 - B)\bar{K}^2 + \bar{K}[6AB^2 - 2A^3B] - 3A^2B^3 = 0,$$

z niego rozwiążanie optymalne

$$\bar{K}^o = \frac{AB}{3} \cdot \frac{A^2 - 3B + A\sqrt{A^2 + 3B}}{A^2 - B}.$$

10.5.3. Optymalny regulator PI według wskaźnika z kwadratu pochodnej błędu

W podrozdziale tym wyprowadzone zostaną wzory analityczne dla wskaźnika kwadratowego opartego na pochodnej błędu dla obiektu III rzędu i regulatora PI, gdy $L[y'(t)] = sY(s)$

$$J = \int_0^\infty \dot{\varepsilon}_p^2(t) dt = \int_0^\infty \dot{y}^2(t) dt = \int_0^\infty \bar{y}^2(t) dt.$$

Ponieważ

$$Y(s) = \frac{s}{s^4 + A s^3 + B s^2 + (K + 1) s + \alpha} \cdot \frac{1}{s},$$

to

$$\bar{Y}(s) = \frac{s^2}{s^4 + A s^3 + B s^2 + (K + 1) s + \alpha} \cdot \frac{1}{s},$$

i ze względu na ogólną formułę

$$Y(s) = \frac{b_m s^m + b_{m-1} s^{m-1} + \dots + b_1 s + b_0}{a_n s^n + a_{n-1} s^{n-1} + \dots + a_1 s + a_0} \cdot \frac{1}{s},$$

można wyprowadzić wartość całki.

W tym przypadku: $n = 4, m = 2$:

$$b_0 = 0, \quad a_1 = K+1, \quad B_0 = 0,$$

$$b_1 = 0, \quad a_2 = B, \quad B_1 = 0,$$

$$b_2 = 1, \quad a_3 = A, \quad B_2 = 1,$$

$$a_0 = \alpha, \quad a_4 = 1,$$

$$\Delta = \begin{vmatrix} a_0 & -a_2 & a_4 & 0 \\ 0 & a_1 & -a_3 & 0 \\ 0 & -a_0 & a_2 & -a_4 \\ 0 & 0 & -a_1 & a_3 \end{vmatrix} = \begin{vmatrix} \alpha & -B & 1 & 0 \\ 0 & (1+K) & -A & 0 \\ 0 & -\alpha & B & -1 \\ 0 & 0 & -(1+K) & A \end{vmatrix} =$$

$$= \alpha \begin{vmatrix} (1+K) & -A & 0 \\ -\alpha & B & -1 \\ 0 & -(1+K) & A \end{vmatrix} =$$

$$= \alpha(1+K) \begin{vmatrix} B & -1 \\ -(1+K) & A \end{vmatrix} - \alpha(-A) \begin{vmatrix} -\alpha & -1 \\ 0 & A \end{vmatrix} = \alpha(1+K)[AB - (1+K)] + \alpha A(-\alpha A),$$

$$\Delta = \alpha(1+K)AB - \alpha(1+K)^2 - \alpha^2 A^2.$$

Oznaczając: $1+K = \bar{K}$, otrzymujemy

$$\Delta = \alpha \bar{K}AB - \alpha \bar{K}^2 - \alpha^2 A^2.$$

Zastępując kolumnę numer 3 w wyznaczniku Δ wektorem kolumnowym $[a_1 \ a_0 \ 0 \ 0 \ 0 \dots 0]^T$, otrzymujemy Δ_2

$$\Delta_2 = \begin{vmatrix} \alpha & -B & (K+1) & 0 \\ 0 & (K+1) & \alpha & 0 \\ 0 & -\alpha & 0 & -1 \\ 0 & 0 & 0 & A \end{vmatrix} = \alpha \begin{vmatrix} (K+1) & \alpha & 0 \\ -\alpha & 0 & -1 \\ 0 & 0 & A \end{vmatrix} =$$

$$= \alpha(K+1)(0) - \alpha^2(\alpha A - 0) = \alpha^3 A,$$

$$J = \frac{1}{2a_0^2 \Delta} [B_0 \Delta_0 + B_1 \Delta_1 + \dots + B_m \Delta_m - 2b_0 b_1 \Delta].$$

Stąd

$$J = \frac{1}{2a_0^2 \Delta} [B_2 \Delta_2],$$

$$J = \frac{1}{2\alpha^2 (\alpha \bar{K}AB - \alpha \bar{K}^2 - \alpha^2 A^2)} [1\alpha^3 A] = \frac{\alpha A}{2(\alpha \bar{K}AB - \alpha \bar{K}^2 - \alpha^2 A^2)}.$$

Wskaźnik ogólny ma postać:

$$J_{21} = \int_0^\infty \varepsilon_p^2(t) dt + \beta \int_0^\infty \dot{\varepsilon}_p^2(t) dt,$$

$$J_{21} = \frac{AB - \bar{K} + \beta \alpha A}{2\alpha \bar{K}AB - 2\alpha \bar{K}^2 - 2\alpha^2 A^2} \quad (10.14)$$

$$\frac{\delta J_2}{\delta \bar{K}} = 0, \quad \frac{\delta J_2}{\delta \alpha} = 0,$$

$$\begin{aligned} \frac{\delta J_2}{\delta \alpha} &= \beta A \left[2\alpha \bar{K}AB - 2\alpha \bar{K}^2 - 2\alpha^2 A^2 \right] - \\ &\quad - [AB - \bar{K} + \beta \alpha A] \left[2\bar{K}AB - 2\bar{K}^2 - 4\alpha A^2 \right] = 0, \\ &\quad \left[2\beta A^2 B \bar{K} \alpha - 2\beta A \bar{K}^2 \alpha - 2\beta A^3 \alpha^2 \right] - \\ &\quad - \left[2A^2 B^2 \bar{K} - 4A B \bar{K}^2 - 4A^3 B \alpha + 2\bar{K}^3 + \right. \\ &\quad \left. + 4A^2 \bar{K} \alpha + 2\beta A^2 B \bar{K} \alpha - 2\beta A \bar{K}^2 \alpha - 4\beta A^3 \alpha^2 \right] = 0, \end{aligned}$$

$$\begin{aligned} &2\beta A^2 B \bar{K} \alpha - 2\beta A \bar{K}^2 \alpha - 2\beta A^3 \alpha^2 - 2A^2 B^2 \bar{K} + \\ &+ 4A B \bar{K}^2 + 4A^3 B \alpha - 2\bar{K}^3 - 4A^2 \bar{K} \alpha - \\ &- 2\beta A^2 B \bar{K} \alpha + 2\beta A \bar{K}^2 \alpha + 4\beta A^3 \alpha^2 = 0, \end{aligned}$$

$$\begin{aligned} &\left[4\beta A^3 - 2\beta \hat{A}^3 \right] \alpha^2 + \left[2\beta A^2 B \bar{K} - 2\beta A \bar{K}^2 + 4A^3 B - 4A^2 \bar{K} - 2\beta A^2 B \bar{K} + 2\beta A \bar{K}^2 \right] \alpha + \\ &+ \left[-2\bar{K}^3 + 4A B \bar{K}^2 - 2A^2 B^2 \bar{K} \right] = 0, \\ &(2A^3 \beta) \alpha^2 + (4A^3 B - 4A^2 \bar{K}) \alpha - 2\bar{K}(\bar{K}^2 - 2A B \bar{K} + A^2 B^2) = 0, \\ &(2A^3 \beta) \alpha^2 + 4A^2 (AB - \bar{K}) \alpha - 2(\bar{K} - AB)^2 \bar{K} = 0 \quad (10.15) \end{aligned}$$

$$\frac{\delta J_2}{\delta \bar{K}} = 0,$$

$$-1 \left[2\alpha \bar{K}AB - 2\alpha \bar{K}^2 - 2\alpha^2 A^2 \right] - [AB - \bar{K} + \beta \alpha A] \left[2\alpha AB - 4\alpha \bar{K} \right] = 0,$$

$$\begin{aligned}
& -2\alpha AB\bar{K} + 2\alpha\bar{K}^2 + 2\alpha^2 A^2 - 2\alpha A^2 B^2 + 2\alpha AB\bar{K} - 2\alpha^2 A^2 B\beta + 4AB\alpha\bar{K} - 4\alpha\bar{K}^2 + 4\alpha^2 \beta\bar{K} = 0, \\
& 2\alpha\bar{K}^2 + 2\alpha^2 A^2 - 2\alpha A^2 B^2 - 2\alpha^2 A^2 B\beta + 4AB\alpha\bar{K} - 4\alpha\bar{K}^2 + 4\alpha^2 A\beta\bar{K} = 0, \\
& -2\alpha\bar{K}^2 + 2\alpha^2 A^2 - 2\alpha A^2 B^2 - 2\alpha^2 A^2 B\beta + 4AB\alpha\bar{K} + 4\alpha^2 A\beta\bar{K} = 0, \\
& 2\alpha\bar{K}^2 - 2\alpha^2 A^2 + 2\alpha A^2 B^2 + 2\alpha^2 A^2 B\beta - 4AB\alpha\bar{K} - 4\alpha^2 A\beta\bar{K} = 0, \\
& \bar{K}^2 - \alpha A^2 + A^2 B^2 + \alpha A^2 B\beta - 2AB\bar{K} - 2\alpha A\beta\bar{K} = 0, \\
& \bar{K}^2 - 2AB\bar{K} - 2\alpha A\beta\bar{K} + A^2 B^2 + A^2 B\beta\alpha - \alpha A^2 = 0, \\
& \bar{K}^2 - 2A[B + \alpha\beta]\bar{K} + A^2 [B^2 + B\beta\alpha - \alpha] = 0
\end{aligned} \tag{10.16}$$

Z równania (10.16) mamy:

$$\begin{aligned}
& \bar{K}^2 - 2AB\bar{K} - 2\alpha A\beta\bar{K} + A^2 B^2 + A^2 B\beta\alpha - \alpha A^2 = 0, \\
& \alpha(-2A\beta\bar{K} + A^2 B\beta - A^2) = 2AB\bar{K} - \bar{K}^2 - A^2 B^2, \\
& \alpha = \frac{\bar{K}^2 - 2AB\bar{K} + A^2 B^2}{-A^2(\beta B - 1) + 2A\beta\bar{K}}
\end{aligned} \tag{10.17}$$

Po podstawieniu (10.17) do równania (10.15), otrzymujemy końcową zależność w postaci wielomianu trzeciego stopnia na optymalny współczynnik K

$$\bar{K}^3 - \frac{A(4\beta B - 7)}{4\beta}\bar{K}^2 + \frac{A^2(\beta^2 B^2 - 6\beta B + 3)}{4\beta^2}\bar{K} + \frac{A^3 B(\beta B - 2)}{4\beta^2} = 0 \tag{10.18}$$

Dla pewnych wartości wagi β , równanie (10.18) ma proste rozwiązania, np. po podstawieniu $\beta = 2/B$ otrzymuje się:

$$\begin{aligned}
& \bar{K}^2 - \frac{AB}{8}(8 - 7)\bar{K} + \frac{A^2 B^2}{16}(4 - 6,2 + 3) = 0, \\
& \bar{K}^2 - \frac{AB}{8}\bar{K} - \frac{5}{16}A^2 B^2 = 0, \\
& 16\bar{K}^2 - 2AB\bar{K} - 5A^2 B^2 = 0,
\end{aligned}$$

$$\begin{aligned}
\Delta &= 4A^2 B^2 - 4(16)(-5A^2 B^2) = 4A^2 B^2 + 320A^2 B^2 = 324A^2 B^2, \quad \sqrt{\Delta} = 18AB, \\
\bar{K} &= \frac{2AB \pm 18AB}{32} = \frac{20AB}{32} \quad \text{lub} \quad \frac{-16AB}{32}.
\end{aligned}$$

Ostatecznie optymalne K

$$K_1^o = \frac{5}{8}AB - 1, \quad K_2^o = -\frac{AB}{2} - 1 \quad (K_2 \text{ przypadek układu niestabilnego}).$$

Przez podstawienie K_1^o do równania (10.17) otrzymuje się

$$\alpha^o = \frac{\left(\frac{5}{8}AB\right)^2 - 2AB\frac{5}{8}AB + A^2 B^2}{-A^2(\beta B - 1) + 2A\beta\frac{5}{8}AB} = \frac{3}{32}B^2.$$

$$\text{Ostatecznie dla } \beta = 2/B: K^o = \frac{5}{8}AB - 1, \quad \alpha^o = \frac{3}{32}B^2, \quad J_2^o = \frac{64}{3AB^3}.$$

Dla $\beta = 0$, otrzymuje się inną optymalną wartość, wyrażoną wzorem (10.11):

$$K^o = \frac{2}{3}AB - 1, \quad \alpha^o = \frac{1}{9}B^2, \quad J_1^o = \frac{27}{2AB^3}.$$

Wagę $\beta = \infty$ można interpretować jako szczególny przypadek, gdy J_{21} z (10.5) i (10.4) redukuje się tylko do drugiej całki, tzn. z kwadratu pochodnej błędu.

$$J_{20} = \int_0^\infty [\dot{\epsilon}(t)]^2 dt.$$

Równanie (10.18) redukuje się wtedy do $\bar{K}^2 - AB\bar{K} + A^2 B^2 / 4 = 0$,

$$\Rightarrow (\bar{K} - AB/2)^2 = 0, \quad \alpha = 0 \Rightarrow K^o = AB/2 - 1, \quad \alpha^o = 0, \quad J_{20} = 2/(AB^2).$$

Oznacza to, że całka jest najmniejsza, gdy w regulatorze nie ma akcji całkującej. Taki regulator zgodnie z (10.8) pozostawia jednak błąd w stanie ustalonym

$$\varepsilon_{ust} = 1 - K_p / (K + 1).$$

Otrzymywanie poziomów uogólnionego wskaźnika jakości

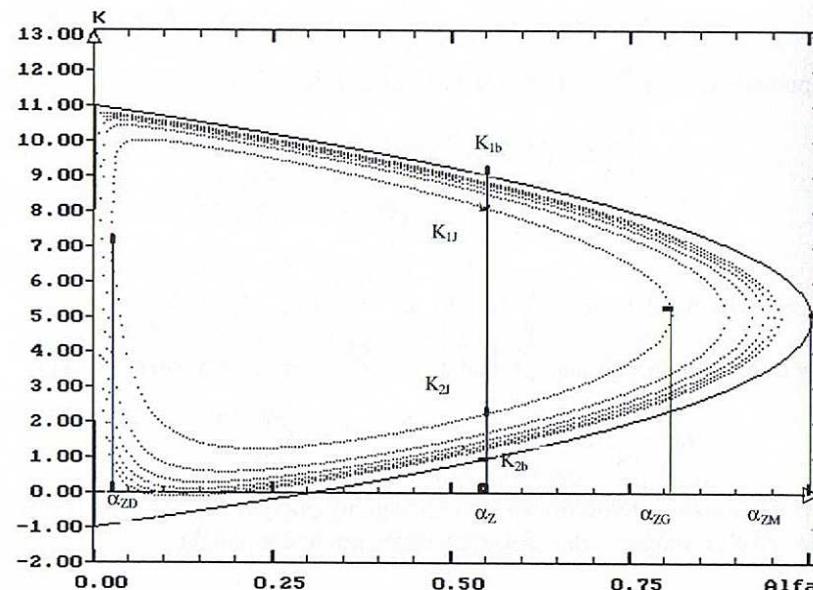
Z równania (10.14) można otrzymać równanie parametryczne poziomów dla wskaźnika jakości J_{21} . Dla zadanej wartości poziomicy J_{2Z} i danej α_Z ze wzoru (10.14), można otrzymać równanie kwadratowe na dwa współczynniki wzmacnienia K_{1J} i K_{2J} , należące do tej poziomicy.

$$\hat{K}^2 - \hat{K} \left(AB + \frac{1}{2\alpha_Z J_{2Z}} \right) + \alpha_Z A^2 + \frac{\alpha_Z \beta A + AB}{2\alpha_Z J_{2Z}} = 0 \tag{10.19}$$

$$K_{1J,2J} = \frac{AB + \frac{1}{2\alpha_Z J_{2Z}} \pm \sqrt{\Delta}}{2} - 1,$$

gdzie Δ jest liczone ze wzoru (10.19).

Rysunek 10.8 pokazuje przykład zbioru poziomie.



Rys. 10.8. Obszar stabilności i poziomice J_2 dla parametrów $A = 6,0$, $B = 2,0$ i $\beta = 5$

Dla każdej wybranej poziomicy współczynnik α_z może się zmieniać w przedziale $\alpha_z \in [\alpha_{ZD}, \alpha_{ZG}]$, gdzie α_{ZD} (dolny) i α_{ZG} (górny) są określone z warunku $\Delta > 0$. Z zależnością tych można zbudować procedurę wykreślania zbioru poziomic na płaszczyźnie parametrów regulatora K, α . Poziomice kreśli się wewnątrz obszaru stabilności układu zamkniętego.

Brzegiem obszaru stabilności jest parabola wyliczona z (10.19) dla $J_{2Z} = \infty$, tzn.

$$\bar{K}^2 - AB\bar{K} + \alpha_z A^2 = 0$$

i wzór na wartości brzegowe K_b jest dany wzorem jak w (10.19)

$$K_{1b,2b} = \frac{AB \pm A\sqrt{B^2 - 4\alpha_z}}{2}$$

z warunkiem $B^2 - 4\alpha_z > 0$. Stąd α_z powinno spełniać nierówność $0 < \alpha_z < B^2/4 = \alpha_{ZM}$.

10.6. Numeryczne wyniki symulacji układu optymalnego dla kryterium kwadratowego

Na podstawie wyników otrzymanych w poprzednim rozdziale przedstawiona zostanie graficzna prezentacja lokalizacji optymalnych nastaw regulatora w obszarze stabilności układu zamkniętego (rys. 10.9).

Dla współczynnika wagi $\beta = 0$, rozwiązanie pierwszego zadania optymalizacji parametrycznej (całka z kwadratu błędu) jest następujące:

$$K^o = \frac{2}{3}AB - 1, \quad \alpha^o = \frac{1}{9}B^2, \quad J_1^o = \frac{27}{2AB^3}, \quad J_{exp} = \frac{J_1^o M^2 K_p^2}{\Omega}.$$

Dla $\beta \neq 0$ rozwiązanie drugiego zadania optymalizacji parametrycznej (całka z ważonej sumy kwadratu błędu i kwadratu pochodnej błędu) sprowadza się do rozwiązania równania trzeciego stopnia na optymalne K , a następnie znalezienia optymalnego α .

$$\bar{K}^3 - \frac{A(4\beta B - 7)}{4\beta} \bar{K}^2 + \frac{A^2(\beta^2 B^2 - 6\beta B + 3)}{4\beta^2} \bar{K} + \frac{A^3 B(\beta B - 2)}{4\beta^2} = 0,$$

$$\alpha = \frac{\bar{K}^2 - 2AB\bar{K} + A^2 B^2}{-A^2(\beta B - 1) + 2A\beta\bar{K}}.$$

Dla specjalnie dobranej wartości wagi $\beta = 2/B$ rozwiązanie optymalne ma postać:

$$K^o = \frac{5}{8}AB - 1, \quad \alpha^o = \frac{3}{32}B^2, \quad J_2^o = \frac{64}{3AB^3},$$

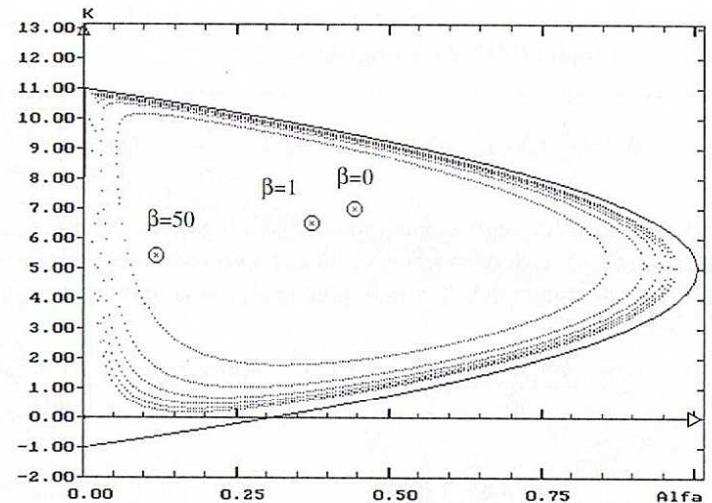
$$\hat{K}_{Rexp}^o = K_R^o, \quad \hat{\alpha}_{Rexp}^o = \alpha\Omega.$$

Numeryczne wyniki dla dwóch różnych testowych obiektów III rzędu i trzech różnych wag β prezentuje tabela 10.1. Pierwszy obiekt jest oscylacyjny z biegunami $[-5,678, -0,16 \pm 0,387j]$, a drugi inercyjny z biegunami $[-2,29, -1,14, -0,37]$. Nastawy podane są również dla obiektu nieznormalizowanego. Nastawy naniesiono też na rysunkach 10.9 i 10.16.

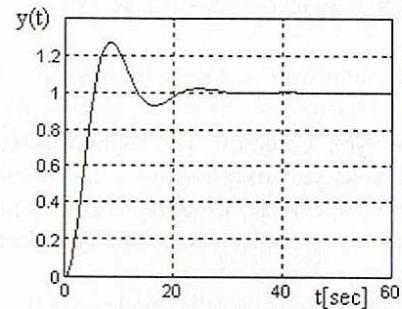
Na tych rysunkach przykładowy przebieg poziomic odpowiada tylko wadze $\beta = 0$.

Tabela 10.1
Optymalne parametry regulatora PI dla obiektów testowych

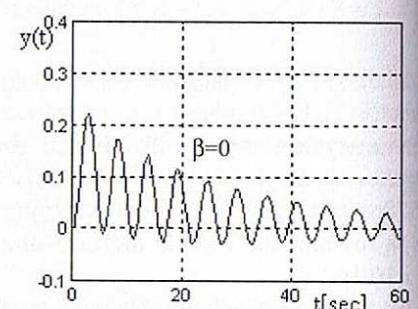
Obiekt	A	B	K_p	β	K^o	α^o	J^o	K_R^o	α_R^o	$\hat{\alpha}_R^o$	J_{exp}
I	6	2	1	0	7,00	0,444	0,281	7,00	0,444	0,444	0,281
	6	2	1	1	6,50	0,375	0,444	6,50	0,375	0,375	0,444
	6	2	1	50	5,38	0,120	5,500	5,38	0,120	0,120	5,500
II	3,8157	3,9311	0,9	0	9,00	1,717	0,058	10,0	1,908	0,0166	185185
	3,8157	3,9311	0,9	1	8,10	1,298	0,120	9,00	1,442	0,0125	381560
	3,8157	3,9311	0,9	50	6,85	0,340	2,070	7,61	0,377	0,0033	6582000

Rys. 10.9. Lokalizacja parametrów optymalnych dla obiektu $A = 6, B = 2$, i $\beta = 0, 1, 50$

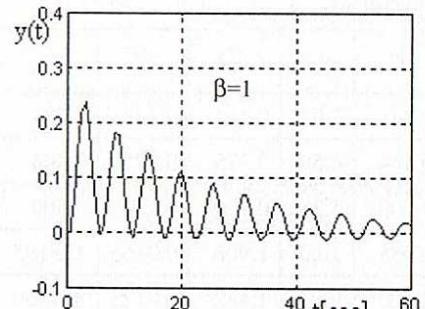
Na rysunkach 10.10 i 10.14 przedstawione są odpowiedzi układu otwartego, a na rysunkach 10.11–10.13 i 10.15 – odpowiedzi układu zamkniętego na skoki jednostkowe.



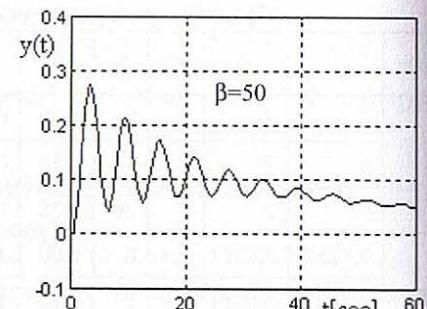
Rys. 10.10. Odpowiedź obiektu otwartego



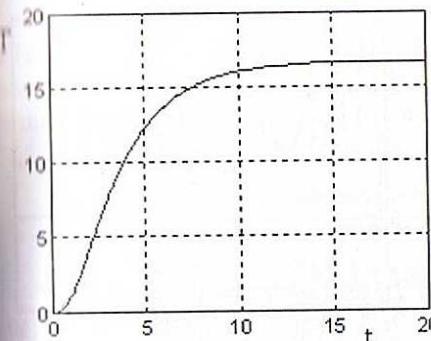
Rys. 10.11. Odpowiedź układu zamkniętego



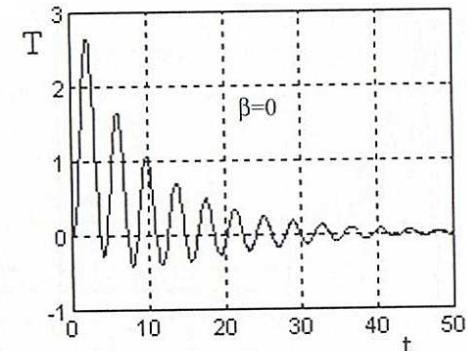
Rys. 10.12. Odpowiedź układu zamkniętego



Rys. 10.13. Odpowiedź układu zamkniętego

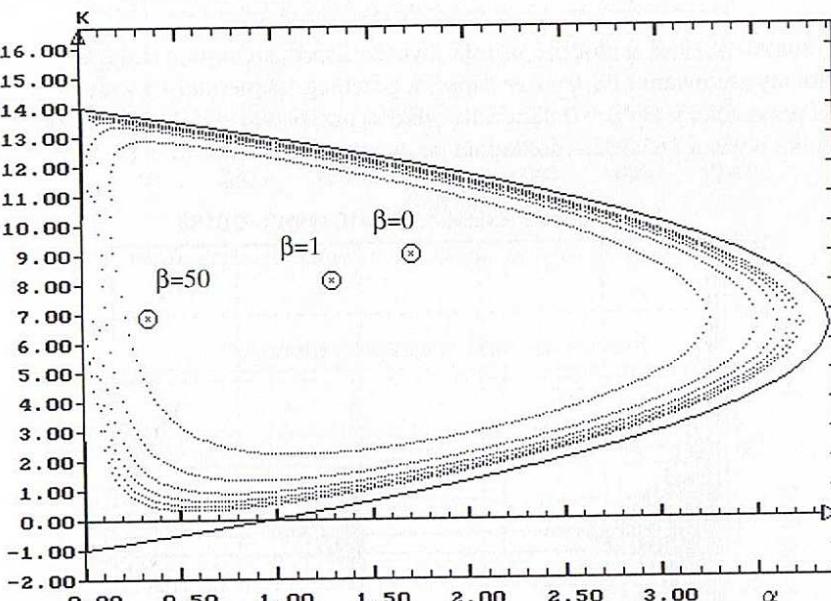


Rys. 10.14. Odpowiedź obiektu otwartego



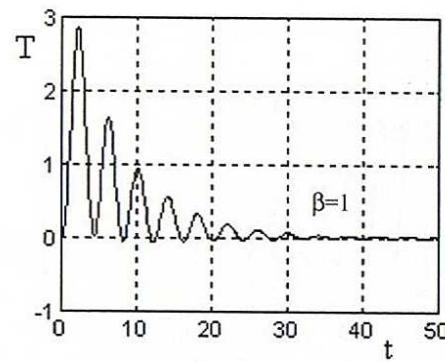
Rys. 10.15. Odpowiedź układu zamkniętego

Przykładowy przebieg poziomu (rys. 10.16) odpowiada tylko wadze $\beta = 0$.

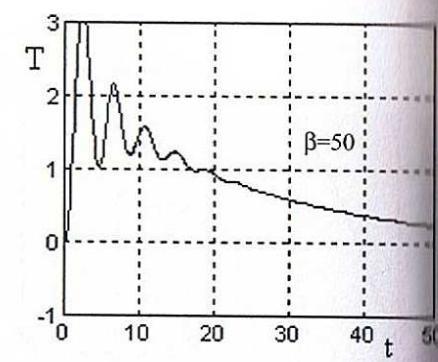
Rys. 10.16. Lokalizacja parametrów optymalnych dla obiektu $A = 3,815, B = 3,931, \beta = 0, 1, 50$

Wykorzystując optymalny regulator PI ($\hat{K}_R = 10, \hat{\alpha}_R = 0,0166$), można wykonać symulację zamkniętego układu rzeczywistego dla dwóch przypadków (modele przyrostowe):

- 1) dla nadążania za skokiem jednostkowym wartości zadanej z poziomu 0°C do 50°C (wartość mocy zasilania zmieniła się z $0 \text{ h}\cdot\text{W}$ do $55,55 \text{ h}\cdot\text{W}$) w chwili $t = 0 \text{ s}$,
- 2) dla likwidacji skoku jednostkowego zakłócenia (spadek mocy grzewczej w zasilaniu do $55,55/3 \text{ h}\cdot\text{W} = 18,5 \text{ h}\cdot\text{W}$; stąd $M = 18,5 \text{ h}\cdot\text{W}$ – gdy trzy grzałki z dziewięciu zostały wyłączone) w chwili $t = 15\,000 \text{ s}$.



Rys. 10.17. Odpowiedź układu zamkniętego



Rys. 10.18. Odpowiedź układu zamkniętego

Rysunki 10.17 i 10.18 przedstawiają odpowiedzi układu zamkniętego na skok jednostkowy.

Na rysunku 10.19 są widoczne wyniki dwóch eksperymentów – nadążania i stabilizacji, tzn. zmiany sterowania PS (power supply), przebieg temperatury i wartość wskaźnika jakości dla przypadku wagi $\beta = 0$ naliczaną tylko w przypadku stabilizacji. Wartość końcowa wskaźnika wynosi 185 185 – dokładnie jak w tabeli 10.1.

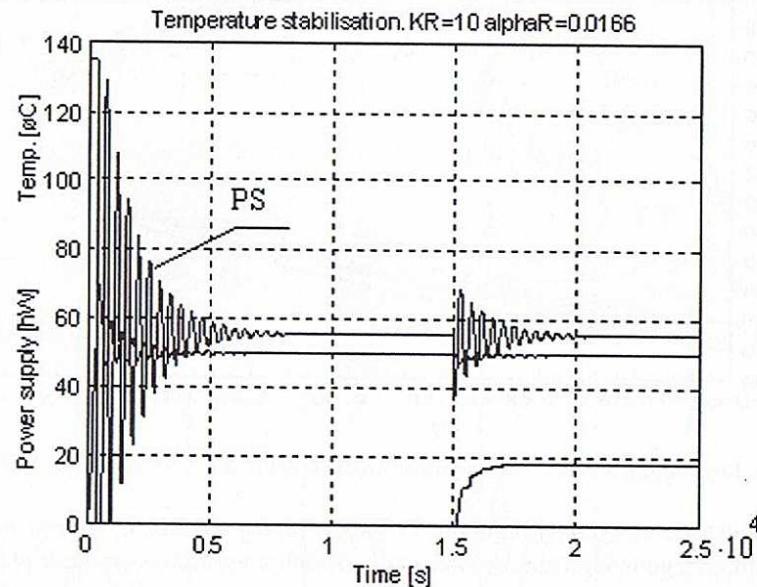
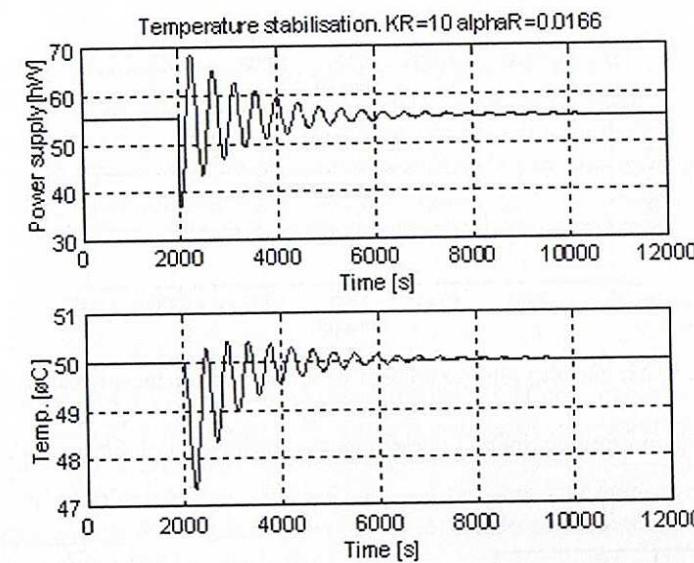


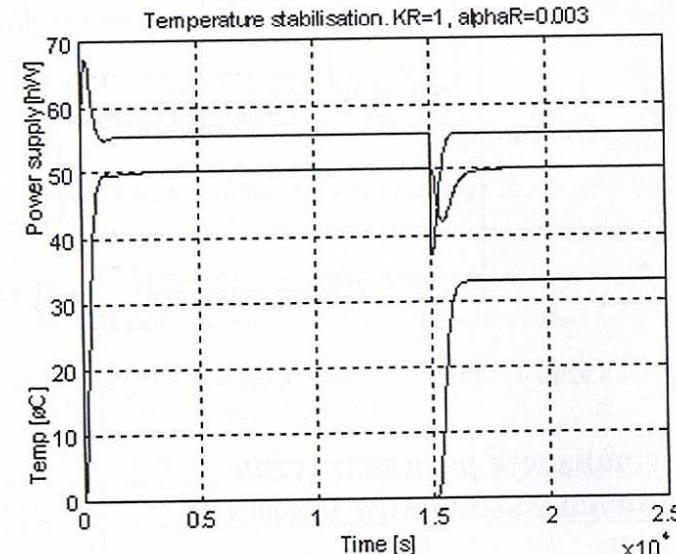
Fig. 10.19. Przebiegi sterowania i wyjścia systemu zamkniętego

Dokładniejsze wykresy z drugiego eksperymentu (stabilizacji) pokazano na rysunku 10.20. Maksymalna amplituda błędu temperatury wynosiła $2,5^{\circ}\text{C}$.

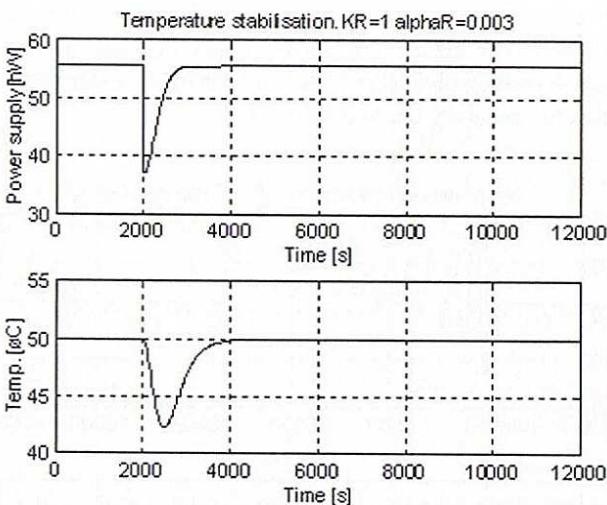
Eksperyment ten powtórzono dla nieoptymalnych (dla $\beta = 0$) parametrów regulatora ($\hat{K}_R = 1$, $\hat{\alpha}_R = 0,003$) (rys. 10.21 i 10.22). Wartość wskaźnika jakości w tym przypadku jest równa 3 304 830. Przebieg nie jest oscylacyjny zarówno w sterowaniu, jak i w wyjściu, maksymalny błąd temperatury dochodzi do 7°C .



Rys. 10.20. Przebiegi sterowania i zmiany temperatury przy stabilizacji

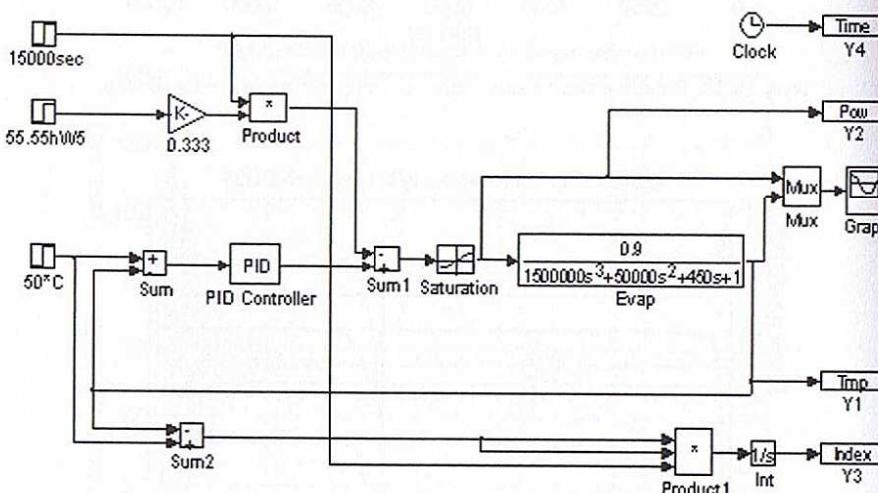


Rys. 10.21. Sterowanie i odpowiedzi zamkniętego nieoptymalnego układu sterowania



Rys. 10.22. Przebieg samego procesu stabilizacji układu nieoptymalnego

Schemat układu symulacyjnego wykonanego w pakiecie Matlab/Simulink przedstawiony na rysunku 10.23.



Rys. 10.23. Schemat blokowy wykonany w pakiecie Matlab/Simulink

10.7. Optymalizacja parametryczna dla innych wskaźników jakości

Jako trzeci wskaźnik dla testowego znormalizowanego obiektu III rzędu $A = 6, B = 2$ rozważana będzie całka z modułu błędu, wyliczana numerycznie przez symulację prze-

biegu $\varepsilon_p(t)$ na skończonym (ale zmiennym) przedziale $[0, t_k]$, gdzie $|\varepsilon_p(t_k)| < \varepsilon_z$ – zadany błąd.

$$J_3 = \int_0^{t_k} |\varepsilon_p(t)| dt \quad (10.20)$$

Jako czwarte, ostatnie zadanie optymalizacyjne, przyjęto zadanie mini-max w dziedzinie częstotliwości (H_∞). Minimalizowanym kryterium jest maksymalny moduł transmitancji widmowej układu zamkniętego. Układ optymalny gwarantuje najlepszą odporność dla przypadku najgorszego zakłócenia sinusoidalnego, tzn. o częstotliwości pokrywającej się z częstotliwością maksymalnego modułu. Zakłócenia o innych częstotliwościach dadzą błąd mniejszy (tzn. mniejszą amplitudę odchyłki błędu w stanie ustalonym):

$$J_4 = \max_{\omega} |G_z(\omega)| \quad \text{dla } \omega \in [0, \infty), \quad \min J_4 = \min_{\alpha} \max_{\omega} |G_z(\omega)|.$$

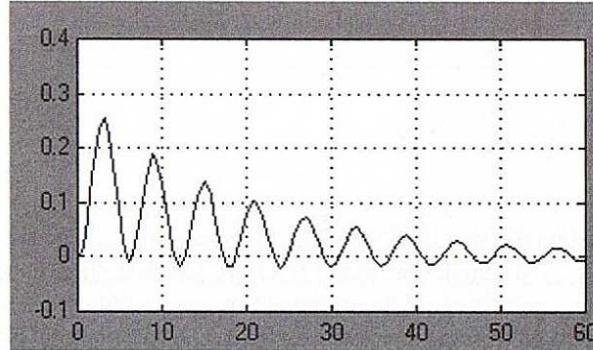
Analiza wykazuje że nastawy optymalne regulatora PI dającego minimum J_4 leżą na ograniczeniu $\alpha^o = 0, K^o \neq 0$ (regulator P). Takie rozwiązanie nie gwarantuje błędu w stanie ustalonym $\varepsilon_u(\infty) = 0$. Warunkowi $\varepsilon_u(\infty) = 0$ odpowiada w dziedzinie częstotliwości warunek, aby dla $\omega = 0$ $|G_z(0)| = 0$. Dlatego zostanie przyjęte zmodyfikowane kryterium J_4 uwzględniające z wagą W karę za $|G_z(0)| \neq 0$ i jednocześnie za wartość pochodnej w $t = 0$ (aby nie była za duża). Stąd J_4 ma końcową postać:

$$\begin{aligned} J_4 &= \max_{\omega} |G_z(\omega)| + W \left(\frac{|G_z(0)|}{\Delta\omega} + \frac{|G_z(\Delta\omega)| - |G_z(0)|}{\Delta\omega} \right) = \\ &= \max_{\omega} |G_z(\omega)| + W \frac{|G_z(\Delta\omega)|}{\Delta\omega} \end{aligned} \quad (10.21)$$

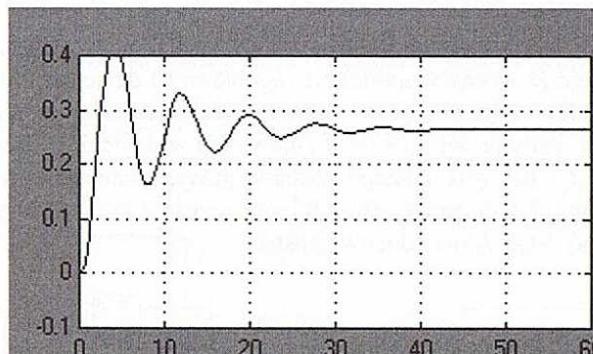
Obliczenia numeryczne dały dla modelu testowego $A = 6, B = 2$ najlepsze nastawy dla J_3 i J_4 (tab. 10.2).

Tabela 10.2
Optymalne parametry regulatora PI dla obiektu testowego

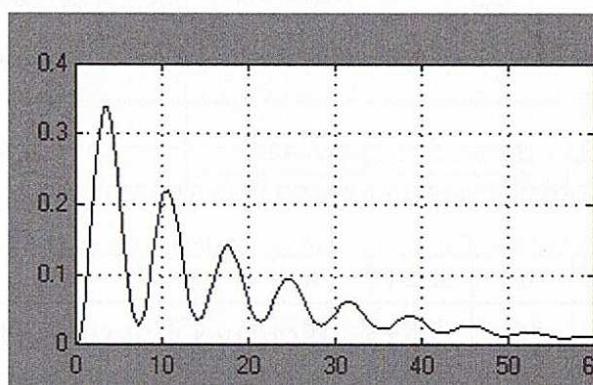
Nr	K^o	α^o	$\min J_3$	J_3		J_4		
				W	K^o	α^o	$\min J_4$	
1	5,63	0,41	2,688	0	2,77	0	0,92	
2	–	–	–	0,1	3,77	0,24	1,65	
3	–	–	–	0,5	4,05	0,42	2,84	
4	–	–	–	1	4,15	0,5	3,93	



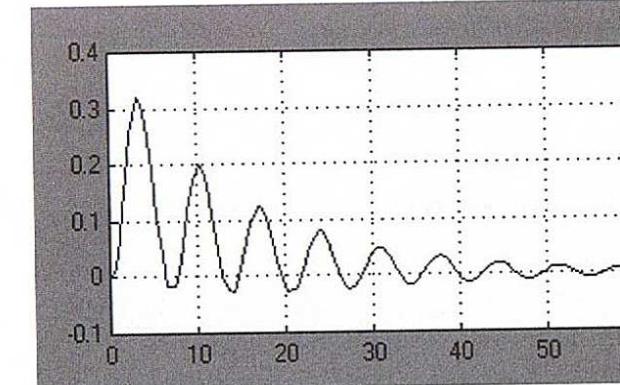
Rys. 10.24. Przebieg odpowiedzi na skok jednostkowy w oparciu o parametry PI nr 1 i kryterium J_3



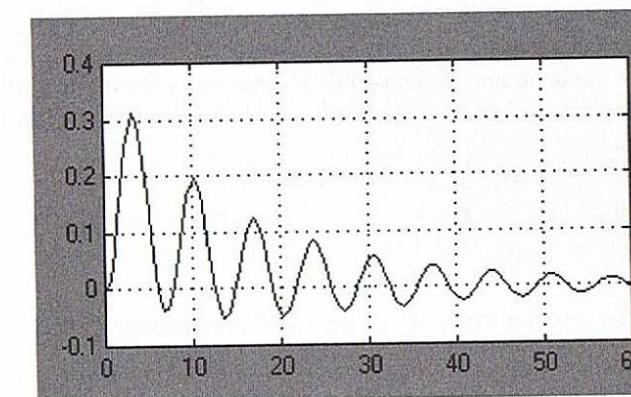
Rys. 10.25. Przebieg odpowiedzi na skok jednostkowy w oparciu o parametry PI nr 1 i kryterium J_4



Rys. 10.26. Przebieg odpowiedzi na skok jednostkowy w oparciu o parametry PI nr 2 i kryterium J_4



Rys. 10.27. Przebieg odpowiedzi na skok jednostkowy w oparciu o parametry PI nr 3 i kryterium J_3



Rys. 10.28. Przebieg odpowiedzi na skok jednostkowy w oparciu o parametry PI nr 4 i kryterium J_4

Z rozważań przeprowadzonych w tym rozdziale i uzyskanych rysunków 10.14–10.18 oraz rysunków 10.24–10.28 widać, że dla różnych kryteriów jakości parametry optymalne regulatora lokują się w różnych punktach obszaru stabilności układu zamkniętego (rys. 10.16), gwarantując różnorodność kształtów przebiegu błędu regulacji. Inżynier automatyk przed i w czasie rozwiązywania zadania syntezy musi współpracować z technologią danego procesu, aby z jego pomocą mógł wybrać odpowiedni kształt i dopuszczalny przebieg błędu zmiennej procesowej oraz kryterium jakości.

II. Metody projektowania regulatorów stanu

11.1. Metoda lokowania zer i biegunów

Dla liniowych wielowymiarowych układów stabilizacji stanu $x \in R^n$ podstawowe założenie formułuje się w postaci kanonicznej, przyjmując, że sprowadzanie odchyłki stanu do punktu zerowego z niezerowych warunków początkowych będzie realizowane za pomocą jednowymiarowego, wielowymiarowego proporcjonalnego regulatora od stanu z prostokątną macierzą współczynników wzmocnień K , $[r \times n]$, dającego dla każdego $t > 0$ sterowanie $u \in R^r$.

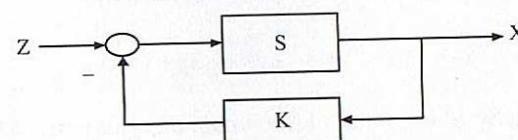
$$u(t) = -Kx(t) = -\begin{bmatrix} K_{11} & \cdots & K_{1n} \\ \vdots & \ddots & \vdots \\ K_{r1} & \cdots & K_{rn} \end{bmatrix} \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}.$$

Dla układów jednowymiarowych sprowadza się to do wektora współczynników K

$$u(t) = -K x(t) = -[K_1 \quad \cdots \quad K_n] \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix} \quad (11.1)$$

Sterowalny względem stanu układ S (rys. 11.1), objęty sprzężeniem zwrotnym, opisują wzory:

$$\begin{aligned} \dot{x}(t) &= A x(t) + B u(t), \\ u(t) &= z(t) - K x(t), \\ \dot{x}(t) &= [A - BK] x(t) + B z(t) \end{aligned} \quad (11.2)$$



Rys. 11.1. Schemat blokowy

Jedną z metod doboru współczynników regulatora K jest lokowanie wartości własnych macierzy stanu (biegunów transmitancji). Temat ten jest omówiony w źródłowej pozycji [11WOW] i np. w pracach [11KAT1, 11NIA].

Poniżej omówimy dwie metody lokowania zer i biegunów.

I METODA LOKOWANIA ZER I BIEGUNÓW

Tworząc wielomian charakterystyczny macierzy stanu układu zamkniętego $|sI - A + BK| = 0$ i porównując jego współczynniki ze współczynnikami α_i wielomianu pożądanego, który ma założone z góry bieguny \bar{s}_i

$$\begin{aligned} |sI - A + BK| &= (s - \bar{s}_1)(s - \bar{s}_2) \cdots (s - \bar{s}_n) = \\ &= s^n + \alpha_{n-1}s^{n-1} + \dots + \alpha_1s + \alpha_0 = 0, \end{aligned}$$

otrzymuje się n równań na współczynniki regulatora K , realizujące ten wielomian. Układ równań będzie miał rozwiązanie dla układu sterowalnego. Stąd sterowalność (A, B) jest warunkiem koniecznym i wystarczającym istnienia takiego regulatora.

Przykład 11.1

Dla obiektu III rzędu $G_o(s)$, znaleźć wektor parametrów regulatora $K \in R^3$ jak w (11.1) tak, aby macierz stanu układu zamkniętego $A_Z = [A - BK]$ (11.2) miała potrójną wartość własną $s = -1$:

$$G_o(s) = \frac{1}{s(s+2)(s+3)}, \quad G_Z(s) = \frac{1}{(s+1)^3}.$$

Równanie stanu układu zamkniętego i równanie wyjścia:

$$\begin{aligned} \dot{x}(t) &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -K_1 & -2 - K_2 & -3 - K_3 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} z(t) = A_Z x(t) + B z(t), \\ y(t) &= x_1(t). \end{aligned}$$

Równanie charakterystyczne powyższej macierzy $[sI - A_Z]$ ma postać

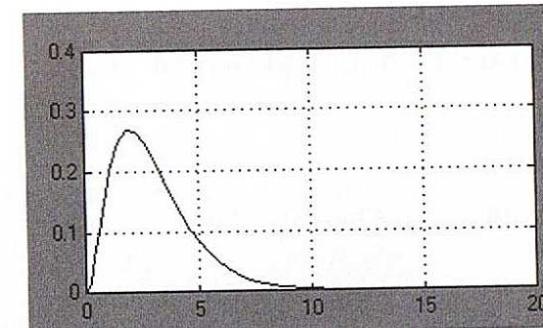
$$s^3 + s^2(3 + K_3) + s(2 + K_2) + K_1 = 0,$$

a pożądany wielomian charakterystyczny układu zamkniętego

$$(s+1)(s+1)(s+1) = s^3 + 3s^2 + 3s + 1 = 0.$$

Stąd, przyrównując współczynniki obu równań, otrzymujemy: $K_1 = 1, K_2 = 1, K_3 = 0$.

Podanie impulsu Diraca na układ zamknięty z rysunku 11.1 daje odpowiedź $y(t)$ tego układu jak na rysunku 11.2.



Rys. 11.2. Odpowiedź układu na $\delta(t)$

Wartość całki z kwadratu zmiennej stanu x_1 (kwadrat błędu) wynosi $J = 0,1875$, a równanie stanu układu zamkniętego:

$$\begin{aligned} \dot{x}(t) &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -3 & -3 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} z(t), \quad G_Z(s) = \frac{1}{(s+1)^3}. \\ y(t) &= [1 \ 0 \ 0] x(t) \end{aligned}$$

II METODA LOKOWANIA ZER I BIEGUNÓW (METODA ACKERMANNA)

Przez α_Z oznacza się pożądaną postać wielomianu charakterystycznego układu zamkniętego

$$\alpha_Z(s) = |sI - A + BK| = s^n + \alpha_{n-1}s^{n-1} + \dots + \alpha_1s + \alpha_0 \quad (11.3)$$

odpowiadającą mu postać wielomianu macierzowego

$$\alpha_Z(A) = A^n + \alpha_{n-1}A^{n-1} + \dots + \alpha_1A + \alpha_0I \quad (11.4)$$

następnie, wykorzystując macierz sterowalności Q_c (8.2), można zastosować uniwersalny wzór, podany przez I. Ackermann [11FRG], na współczynniki macierzy regulatora realizującego (11.3).

$$K = [0 \ 0 \ 0 \ \dots \ 1] [B \ AB \ A^2B \ \dots \ A^{n-1}B]^{-1} \alpha_Z(A) \quad (11.5)$$

Przykład 11.2

Dane z przykładu 11.1:

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -2 & -3 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad Q_c = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & -3 \\ 1 & -3 & 7 \end{bmatrix}, \quad Q_c^{-1} = \begin{bmatrix} 2 & 3 & 1 \\ 3 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Wielomian pożądany $(s+1)(s+1)(s+1) = s^3 + 3s^2 + 3s + 1 = 0$.

$$\alpha_Z(A) = A^3 + \alpha_2 A^2 + \alpha_1 A + \alpha_0 = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & -2 & -2 \end{bmatrix}.$$

$$K = [0 \ 0 \dots 1] Q_c^{-1} \alpha_Z(A) = [0 \ 0 \dots 1] \begin{bmatrix} 2 & 3 & 1 \\ 3 & 4 & 1 \\ 1 & 1 & 0 \end{bmatrix} = [1 \ 1 \ 0].$$

Otrzymane współczynniki K_i są równe współczynnikom otrzymanym w metodzie I.

Należy zauważyć, że dla układów wielowymiarowych MIMO będzie istniało nieskończenie wiele różnych macierzy regulatora K realizujących zadany zestaw wartości własnych macierzy stanu układu zamkniętego $A - BK$.

Dla układu dyskretnego, odpowiadającego układowi ciągłemu (11.2), cała powyższa metodyka wyliczania regulatorów, łącznie ze wzorami Ackermanna pozostaje prawdziwa.

Dyskretny regulator **Dead-beat** od stanu

Poniżej przytoczymy dla systemu dyskretnego wersję czasooptymalną regulatora stabilizującego stan, w której wykorzystuje się metodologię lokowania biegunów. Regulator ma zdolność sprowadzania do zera dowolnego stanu początkowego, w co najwyżej n krokach i zwany jest regułatorem *Dead-beat*.

Dla układu:

$$x_{k+1} = A_D x_k + B_D u_k, \quad x_0 \neq 0,$$

należy zaprojektować regulator K , który pracując w sprzężeniu zwrotnym $u_k = -K x_k$, w co najwyżej n krokach sprowadzi stan układu do zera $x_N = 0$.

Równanie układu zamkniętego

$$x_{k+1} = (A_D - B_D K) x_k = A_{DZ} x_k \quad (11.6)$$

Rozwiązań równania (11.6) wynikające ze wzoru (5.13) ma postać

$$x_k = A_{DZ}^k x_0.$$

Aby stan w n -tym kroku był zero, musi zachodzić zależność

$$x_n = 0 \Leftrightarrow (A_D - B_D K)^n = 0.$$

Rezultat ten można uzyskać pod warunkiem, że wszystkie wartości własne macierzy stanu układu zamkniętego A_{DZ} będą zerowe. Wynika to z przekształcenia macierzy stanu do jej postaci Jordana za pomocą macierzy modalnej P : $x_k = P \xi_k$:

$$P \xi_{k+1} = (A_D - B_D K) P \xi_k, \quad \xi_{k+1} = P^{-1} (A_D - B_D K) P \xi_k,$$

$$\xi_n = \left[P^{-1} (A_D - B_D K) P \right]^n \xi_0 = P^{-1} (A_D - B_D K)^n P \xi_0 = J^n \xi_0.$$

Wtedy przy zerowych wartościach własnych widać efekt zerowania się kolejnych potęg J :

$$J = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad J^2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad J^3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Dla układu sterowanego zawsze można dobrą regulator K tak, by wartości własne układu zamkniętego A_{DZ} były wielokrotne i zerowe $\lambda_i = 0$. W n -tym kroku wystąpi $J^n = 0$ w konsekwencji $x_n = 0$.

Regulator **Dead-beat** dla układu II rzędu

Poniżej wyprowadzono wzory ogólne na regulator K lokujący w zerze wszystkie wartości własne zamkniętego dyskretnego układu drugiego rzędu.

$$A_D = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}, \quad B_D = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad K = [K_1 \ K_2].$$

Z warunku $|sI - A_D + B_D K| = 0$ można otrzymać równanie kwadratowe na s , w którym, przyrównując do zera współczynniki przy pierwszej i zerowej potędze s , otrzymuje się dwa równania na K_1 i K_2 .

Rozwiązań ma postać:

$$K_1 = \frac{(a_1 + a_4)(a_3 b_1 - a_1 b_2) - (a_2 a_3 - a_1 a_4)b_2}{a_3 b_1^2 - a_2 b_2^2 - b_1 b_2(a_1 - a_4)},$$

$$K_2 = \frac{(a_1 + a_4)(a_4 b_1 - a_2 b_2) + (a_2 a_3 - a_1 a_4)b_1}{a_3 b_1^2 - a_2 b_2^2 - b_1 b_2(a_1 - a_4)}.$$

Dla często spotykanego przypadku szczególnego, $a_1 = 1$, $a_3 = 0$, równania upraszczają się do postaci:

$$K_1 = \frac{1}{b_1 - b_1 a_4 + b_2 a_2}, \quad K_2 = K_1 \frac{a_2 b_2 + a_2 b_2 a_4 - a_4^2 b_1}{b_2} \quad (11.7)$$

Przykład 11.3

Rozważmy przykład z podrozdziału 5.3.1, dyskretyzując człon całkująco-inercyjny:

$$G(s) = \frac{1}{s(s+1)} \Rightarrow \ddot{y}(t) + \dot{y}(t) = u(t),$$

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \dot{x} = \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u, \quad y = [1 \ 0] x,$$

$$A_D = e^{AT} = \begin{bmatrix} 1 & 1-e^{-T} \\ 0 & e^{-T} \end{bmatrix}; \quad B_D = \int_0^T e^{A\tau} B d\tau = \begin{bmatrix} T+e^{-T}-1 \\ 1-e^{-T} \end{bmatrix},$$

$$A_{DZ} = A_D - B_D \cdot K.$$

Przyjmując okres próbkowania $T = 0,1$ i stosując wzory (11.7) i (11.6), otrzymujemy

$$A_D = \begin{bmatrix} 1 & 0.0952 \\ 0 & 0.9048 \end{bmatrix}, \quad B_D = \begin{bmatrix} 0.0048 \\ 0.0952 \end{bmatrix}, \quad K = [105, 083 \quad 14, 675],$$

$$A_{DZ} = \begin{bmatrix} -0,4917 & 0,0242 \\ -10,00 & 0,4917 \end{bmatrix}, \quad A_{DZ}^2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Dla takiego regulatora, w drugim kroku obie zmienne stanu osiągają zero z dowolnych warunków początkowych.

11.2. Minimalnonormowe sterowania programowe

Powyżej przedstawiony regulator czasoptymalny typu *Dead-beat*, pracujący w sprzężeniu zwrotnym w oparciu o ostatnią próbkę pomiarową x_k , realizuje ostatecznie taką samą trajektorię stanu jak, wyprowadzony w (8.3) n -krokowy sterownik programowy od stanu początkowego x_0 , wyrażony wzorem

$$u = -Q_c^{-1} \cdot A_D^n x_0.$$

Wzór ten dla układu SISO i stanu końcowego $x_N = 0$, można zmodyfikować dla odwrotnej kolejności próbek sterowania

$$-A_D^N x_0 = [A_D^{N-1} B_D \ \dots \ B_D] \begin{bmatrix} u_0 \\ \vdots \\ u_{N-1} \end{bmatrix} = P \cdot u$$

Jeżeli $N = n$ i układ jest sterowalny, to macierz P jest kwadratowa i otrzymujemy przepis na czasooptymalne sterowanie w układzie otwartym, od zadanego warunku początkowego

$$u^o = -P^{-1} A_D^n x_0$$

Regulator *Dead-beat* i n -krokowy sterownik programowy mogą generować dużą amplitudę impulsów sterujących, niemożliwą do realizacji poprzez praktycznie stosowane układy nastawników. Dalej sformułowane zostanie zadanie zastępcze, jakie może być zrealizowane przez **sterownik programowy**. Zadanie to da rozwiązanie bardziej dostosowane do możliwości technicznych układu, należy jednak założyć z góry większą liczbę kroków $N > n$, doprowadzających stan końcowy do zera, $x_N = 0$. Takich sterowań jest wtedy nie-

okonczona liczba. Można z pośród nich wybrać sterowanie **minimalnonormowe**, tzn. takie, że zachodzi minimum normy sterowania J .

$$J = \|u\|^2 = u^T u = \sum_{i=0}^{N-1} u_i^2.$$

Jeżeli $N > n$, to poziomo-prostokątna macierz P ma wymiar $[n \times N]$ i można zastosować jedynie jej **minimalnonormową prawą pseudoodwrotność** (podrozdz. 19.8). Daje ona jednocześnie minimum J .

$$u^o \equiv -P^T [P P^T]^{-1} A_D^N x_0 \quad (11.8)$$

Jeśli istnieje macierz odwrotna do A_D , to można otrzymać trochę inny, zmodyfikowany, ale równoważny wzór, mnożąc wzór wyjściowy przez A_D^{-N} .

$$x_0 = - \begin{bmatrix} A_D^{-1}B_D & \dots & A_D^{-N}B_D \end{bmatrix} \begin{bmatrix} u_0 \\ \vdots \\ u_{N-1} \end{bmatrix} = \hat{P} \cdot u \Rightarrow u^o = -\hat{P}^T [\hat{P} \hat{P}^T]^{-1} x_0.$$

Inny przypadek sterownika programowego występuje, jeśli narzucony jest skrócony horyzont czasowy sterowania (mała liczba N dopuszczalnych impulsów sterujących). Dla przypadku $N < n$, pionowo-prostokątna macierz P ma wymiary $[n \times N]$ i może nie istnieć sterowanie dyskretnie, które doprowadziłoby układ do stanu końcowego $x_N = 0$. Można wtedy tylko szukać sterowania **minimalizującego normę odchyłki** stanu od zera.

$$\begin{aligned} J &= \|x_N\|^2 = x_N^T x_N, \\ x_N &= A_D^N x_0 + [A_D^{N-1} B_D, \dots, B_D] \cdot u = A_D^N x_0 + P \cdot u, \\ \|x_N\|^2 &= x_N^T x_N = x_0^T (A_D^N)^T A_D^N x_0 + 2x_0^T (A_D^N)^T P u + u^T P^T P u, \\ \frac{\partial J}{\partial u} &= 2P^T A_D^N x_0 + 2P^T P u = 0, \\ u^o &= -[P^T P]^{-1} P^T A_D^N x_0. \end{aligned}$$

W powyższym wzorze wykorzystuje się lewą pseudoodwrotność macierzy P (podrozdz. 19.8) do policzenia optymalnego sterowania gwarantującego minimum odchyłki J .

Powyżej policzone przypadki sterowania dają dwa rodzaje sterowania od stanu – z minimalną energią i z minimalnym błędem. Są to sterowania w **układzie otwartym** (programowe). W podrozdziale 11.3 zostanie przedstawiony przypadek optymalnego sterowania w **sprzężeniu zwrotnym**, dającego minimum mieszanego wskaźnika jakości LQR (*Linear Quadratic Regulator*).

11.3. Sterowanie modalne

Dla pewnych szczególnych przypadków struktury układu można łatwo pokazać możliwość separacji poszczególnych modów układu i bezpośredniego sterowania poszczególnych zmiennych stanu.

Jeżeli istnieją macierze odwrotne do kwadratowych macierzy B i C , tzn. o wymiarze $r = n$ i $m = n$, gdzie $u \in R^r$ i $y \in R^m$, to sterowanie od wyjścia staje się równorzędne sterowaniu od stanu:

$$u(t) = -Ky(t) = -KCx(t), \quad \dot{x}(t) = (A - BKC)x(t) = A_zx(t).$$

Aby zrealizować z góry zadaną macierz stanu układu zamkniętego A_z , wystarczy dobrą kwadratową macierz regulatora

$$K = B^{-1}(A - A_z)C^{-1}.$$

Jeśli zachodzi jeszcze bardziej szczegółowy przypadek, że macierz B równa się macierzy transformacji T dla macierzy A do jej postaci Jordana, $B = T$, a macierz $C = T^{-1}$, wtedy

$$\begin{aligned} A &= TJT^{-1}, \quad u = -Ky = -KT^{-1}x, \\ \dot{x}(t) &= (TJT^{-1} - BKT^{-1})x(t) = T[J - K]T^{-1}x(t), \\ \dot{y}(t) &= [J - K]y(t). \end{aligned}$$

Współczynniki diagonalnej macierzy regulatora K wpływają bezpośrednio na wartości własne macierzy A , ustawiając dynamikę każdego modu z osobna i rozseparowując każdą zmienną wyjścia y_i .

W ogólnym przypadku, $m \leq r < n$, macierz regulatora jest macierzą prostokątną ($r \times n$) wymiarową. Za pomocą $r \cdot n$ elementów można osiągnąć tylko częściową korekcję n^2 elementów macierzy A . Wybiera się wtedy $q < n$ modów dominujących (powolnych, dla których wartości własne są ulokowane blisko osi urojonej) i steruje się tylko q elementami skróconego wektora stanu.

Zaleca się, żeby $q = r = m$ (liczba sterowań = liczbie wyjść):

$$u = -Kx, \quad x = Tz,$$

gdzie z – cały modalny wektor stanu.

Nowe zmienne stanu $z(t)$ trzeba tak dobrać, żeby zachodziło równanie

$$\dot{z}(t) = \left[T^{-1}AT - T^{-1}BKT \right] z(t) = \left[J - \begin{bmatrix} H_q \\ 0 \end{bmatrix} K \begin{bmatrix} F_q & 0 \end{bmatrix} \right] z(t).$$

Skrócony wektor modalny z_q obejmuje tylko mody dominujące i wpływa na dominujące wartości własne J_q : $z_q(t) = [z_1, \dots, z_q]^T$.

$$\dot{z}_q(t) = \left[J_q - H_q K_q F_q \right] z_q(t) = J_{qZ} \cdot z_q(t).$$

Regulator K ma postać macierzy $[q \times n]$ -wymiarowej

$$K = \begin{bmatrix} K_q & 0 \end{bmatrix},$$

dzie $K_q = H_q^{-1} [J_q - J_{qZ}] F_q^{-1}$ – jest macierz kwadratowa $[q \times q]$.

11.4. Optymalny regulator liniowo-kwadratowy LQR

W rozdziale 9 omawiano wskaźniki jakości sterowania oparte na kształcie błędu regulacji $\epsilon_p(t)$ i jego pochodnych. Innymi typowymi wskaźnikami jakości są całkowe wskaźniki wracające również uwagę na koszt sterowania na zadanym odcinku czasu.

Należą do nich:

- koszt energii sterowania $J = \int_{t_0}^{t_1} u^2(t) dt$,
- wielkość wydatku (zużycia) paliwa $J = \int_{t_0}^{t_1} |u(t)| dt$,
- kompromis między jakością stabilizacji i kosztami sterowania ustalany macierzami wagowymi Q i R $J = \int_{t_0}^{t_1} [x^T(t)Qx(t) + u^T(t)Ru(t)] dt$.

Poniżej przedstawiony zostanie podstawowy problem optymalizacji i jego rozwiązanie dla regulatora liniowego od stanu z kwadratowym wskaźnikiem jakości LQR. Problematyka regulacji liniowo-kwadratowej jest treścią wielu opracowań i monografii, np. [11KWH; 11GOH1; 11LEF; 11ANB; 11ATM]. W pozycjach tych można znaleźć odnośniki do źródłowych publikacji z początku lat 60. XX w., autorstwa R. Kalmana. Nowe wyniki, np. dla układów z opóźnieniem, można znaleźć w pracy [11KOA].

Zakładamy, że dany jest system liniowy stacjonarny:

$$\dot{x}(t) = Ax(t) + Bu(t),$$

$$y(t) = Cx(t), \quad x(0) \neq 0.$$

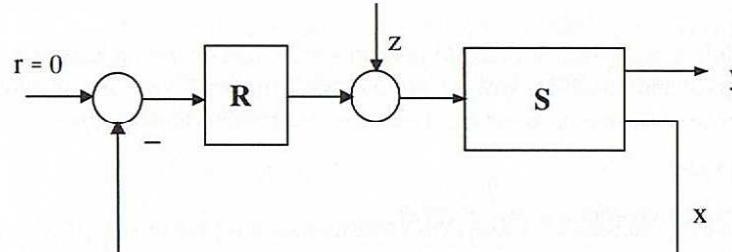
Mogą przyjąć, że: $x(t) \in R^n$, $u(t) \in R^r$, $y(t) \in R^m$ dla $\forall t > 0$, a funkcje sterujące $u_i(\cdot) \in L^2[0, T]$ nie są więc ograniczone co do amplitudy.

Wskaźnik jakości obejmuje odcinek $[0, T]$ i ma postać

$$J = 0,5x^T(T)Fx(T) + 0,5 \int_0^T [x^T(t)Qx(t) + u^T(t)Ru(t)] dt \quad (11.9)$$

gdzie macierze wagowe mają odpowiednie wymiary i są symetryczne, F i Q są dodatnio półokreślone, a R jest dodatnio określona (rys. 11.3).

Zadaniem jest znalezienie optymalnego sterowania $u(t)$, które sprowadzi układ z wejściu początkowego stanu $x(0)$ do zera $x(T) = 0$, w skończonym czasie T . Wskaźnik (11.9) ujmuję punktowy koszt nieosiągnięcia po czasie T stanu zerowego z wagą F i rozłożony (całka) kompromisowy koszt błędu stabilizacji i energii sterowania. Takie klasyczne sformułowanie zadania LQR przedstawione przez Kalmana dotyczy najlepszej stabilizacji układu w stanie zerowym. Do niezerowej wartości $x_0(0)$ w umownej chwili $t = 0$, stan zeroowy może być przesunięty, np. przez zakłócenie impulsowe $z(t)$ w postaci delty Diraca i „mocy” o wartości x_0 (rys. 11.3).



Rys. 11.3. Schemat blokowy

Rozwiążanie zadania można otrzymać np. z rachunku wariacyjnego lub zasady maksimum Pontriagina. W tym celu wprowadza się tzw. funkcję Hamiltona (hamiltonian)

$$H = 0,5 (x^T Q x + u^T R u) + \psi^T (Ax + Bu).$$

Wektor $\psi(t)$ reprezentuje wektor zmiennych sprzężonych do stanu $x(t)$ pełniący rolę mnożników (funkcji) Lagrange'a dla ograniczenia $(Ax + Bu)$ i spełniający równanie

$$\dot{\psi}(t) = -\frac{\partial H}{\partial x} = -A^T \psi(t) - Qx(t).$$

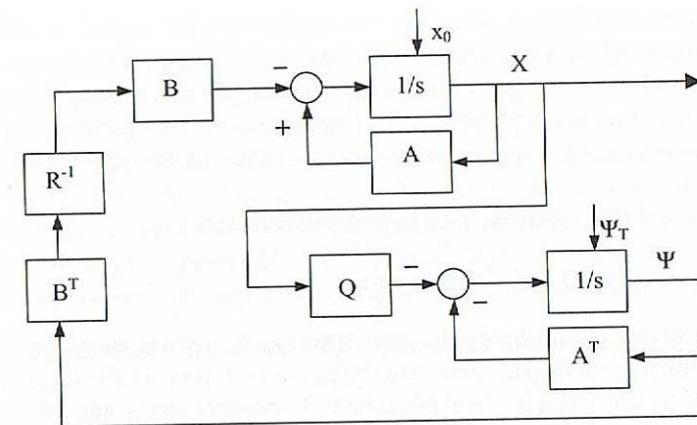
Z warunków koniecznych optymalności zachodzi $\frac{\partial H}{\partial u} = u^T R + \psi^T B = 0$.

Występuje więc liniowa zależność sterowania od zmiennych sprzężonych

$$u^*(t) = -R^{-1}B^T \psi(t).$$

Macierzowy układ dwóch równań różniczkowych ma wymiar $2n \times 2n$ (rys. 11.4)

$$\begin{cases} \dot{x}(t) = Ax(t) - BR^{-1}B^T \psi(t), & x(0), \\ \dot{\psi}(t) = -A^T \psi(t) - Qx(t), & \psi(T) = Fx(T). \end{cases}$$



Rys. 11.4. Schemat blokowy układu kanonicznego

Z powodu dwugranicznosci warunków (część dana na początku $x(0)$, część na końcu $\psi(T)$) nie jest łatwo go rozwiązać. Wprowadzając rozszerzony wektor stanu $\bar{x}(t) = [x, \psi]^T$, otrzymuje się zapis

$$\dot{\bar{x}}(t) = \begin{bmatrix} A & -BR^{-1}B^T \\ -Q & -A^T \end{bmatrix} \bar{x}(t) \quad (11.10)$$

Ważnym wynikiem obliczeń teoretycznych jest fakt, że stan sprzężony jest związany liniowo ze stanem systemu poprzez kwadratową, symetryczną macierz funkcyjną $K(t)$ $[n \times n]$

$$\psi(t) = K(t) x(t) \quad (11.11)$$

Ponieważ $\psi(T) = Fx(T) \Rightarrow K(T) = F$.

Sterowanie optymalne wyraża się więc wzorem

$$u^*(t) = -R^{-1}B^T \psi(t) = -R^{-1}B^T K(t) x(t) = -K(t)x(t).$$

Sterowanie jest liniową kombinacją zmiennych stanu poprzez prostokątną macierz $K(t)$ $[r \times n]$.

Kwadratowa macierz $K(t)$ jest funkcją czasu, stąd regulator optymalny $K(t)$ policzony dla skońzonego przedziału $[0, T]$ jest niestacjonarny.

Macierz $K(T)$ spełnia nieliniowe macierzowe równanie różniczkowe Riccatiego uzyskiwane poprzez zróżniczkowanie (11.11) i wykorzystanie (11.10)

$$\dot{K}(t) = -K(t)A - A^T K(t) + K(t)B R^{-1}B^T K(t) - Q \quad (11.12)$$

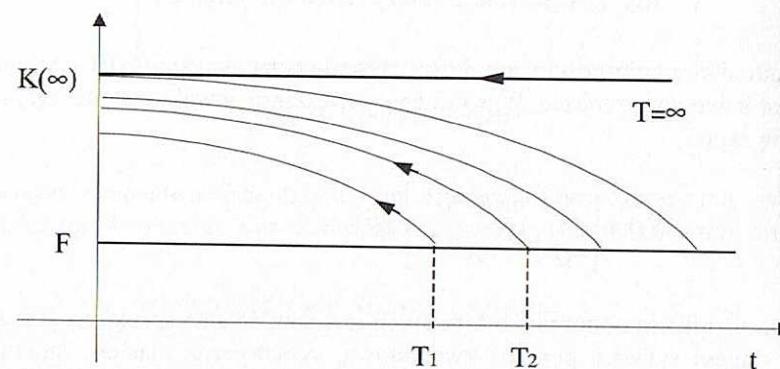
z warunkiem końcowym $K(T) = F$.

Ponieważ rozwiązywanie tego równania nie zależy od $x(0)$, a tylko od macierzy F , można go rozwiązać od tego warunku końcowego do $t = 0$ za pomocą komputera off-line i zapamiętać jego przebieg w rejestrach. Następnie funkcję macierzową $K(t)$ wykorzystuje się w każdej chwili czasu z przedziału $[0, T]$ do mnożenia ze zmierzonym stanem układu $x(t)$, generując sterowanie optymalne. W sposób naturalny tworzy się zamknięty układ sterowania.

Optymalna wartość wskaźnika jakości jest dana wzorem

$$J^o = 0,5 x^T(0) K(0) x(0).$$

Można zauważyc, że wydłużając czas sterowania T i rozwiązyując „od tyłu” (11.12), macierz $K(t)$ stabilizuje się na poziomie ustalonym dla $t = 0$ (rys. 11.5). Tak więc rozwiązania optymalne $K^o(t)$ dla dużego, ale skończonego horyzontu sterowania, na początku jest prawie stałe.



Rys. 11.5. Rozwiązania równania Riccatiego

Stąd specjalny przypadek zachodzi dla przejścia granicznego $T \rightarrow \infty$ i postaci wskaźnika jakości z macierzą $F = 0$.

$$J = 0,5 \int_0^\infty [x^T(t) Q x(t) + u^T(t) R u(t)] dt \quad (11.13)$$

Ponieważ nieskończony horyzont sterowania jest nierealizowalny, to tak uzyskane sterowanie jest suboptimalne, ale ma cenną właściwość – macierz K jest stała. Stała macierz regulatora K minimalizująca powyższy wskaźnik J jest równa wartości ustalonej rozwiązanemu równaniu różniczkowemu Riccatiego $K(\infty)$. Jest też jednocześnie rozwiązaniem równania algebraicznego ARE (Algebraic Riccati Equation).

$$0 = KA + A^T K - KBR^{-1}B^T K + Q \quad (11.14)$$

Tak rozwiązana macierz stała K^o daje stacjonarny regulator dla nieskończonego horyzontu sterowania. Istnieje bogata literatura dotycząca numerycznych metod rozwiązywania, zarówno równania różniczkowego jak i algebraicznego Riccatiego, np. [11BIS] lub [11KWH].

Uwagi:

- Jeżeli para (A, B) jest sterowalna (wystarczy stabilizowalna), to istnieje ograniczone, symetryczne rozwiązanie ARE w postaci macierzy K^o .
- Jeśli dodatkowo para (A, C) jest obserwowały, to K^o jest jedyna i dodatnio określona.
- Najważniejszym stwierdzeniem jest, że system zamknięty z regulatorem stacjonarnym LQR jest asymptotycznie stabilny. Nie może bowiem żadna ze składowych zmiennej stanu x narastać do nieskończoności, gdy wskaźnik jakości ma skończoną wartość stałą.
- Przyjęcie macierzy wagowej $Q = C^T C$ daje $x^T(t) Q x(t) = x^T(t) C^T C x(t) = y^T(t) y(t)$, co można interpretować jako najlepszą stabilizację względem wyjścia $y(t)$.

Wersja dyskretna regulatora LQR dla czasu skończonego T

Dla systemu:

$$x_{i+1} = A_D x_i + B_D u_i, \quad x_0 \neq 0,$$

$$J = 0,5 x_N^T F x_N + 0,5 \sum_{i=1}^{N-1} (x_i^T Q x_i + u_i^T R u_i),$$

$$u_i = -(R + B_D^T K_{i+1} B_D)^{-1} B_D^T K_{i+1} A_D x_i,$$

$$K_i = A_D^T K_{i+1} A_D - A_D^T K_{i+1} B_D (R + B_D^T K_{i+1} B_D)^{-1} B_D^T K_{i+1} A_D + Q,$$

rozwiązanie startuje dla wartości końcowej czasu dyskretnego $T = N$ i wartości $K_N = F$.

Minimalna wartość wskaźnika $J^o = 0,5 x_0^T K^o x_0$.

W każdym kroku iteracji macierzy K_i jest możliwe jednoczesne próbowanie stanu x_i i wyliczanie u_i – tzw. „regulator jeden krok/próbkę”. Inne algorytmy proponują w każdej próbce kontynuację iteracji K_j aż do uzyskania nowego stanu ustalonego K (co ma miejsce zwłaszcza przy zmiennych parametrach systemu i systemach adaptacyjnych) i dopiero wtedy wyliczanie u_i – tzw. „regulator ustalony/próbkę”.

Z powyższych wzorów wynika, że synteza stabilizującego optymalnego regulatora stanu typu LQR wymaga znajomości parametrów procesu i możliwości pomiaru pełnego wektora stanu. W przypadku nieznanego procesu należy dodatkowo zastosować on-line identyfikator dyskretny – dla obiektu dyskretnego, zaś w wersji ciągłej, dla obiektu ciągłego. W przypadku braku dostępu do stanu należy dodatkowo zastosować obserwator stanu – asymptotyczny lub dokładny.

Przykład 11.4

Dany jest system liniowy podwójnie całkujący $G(s) = 1/s^2$ lub $\ddot{y}(t) = u(t)$.

Odpowiada mu równanie stanu ze sterowalnymi i obserwowanymi parami macierzy (A, B) i (A, C) :

$$x_1(t) = x_2(t)$$

$$x_2(t) = u(t) \Rightarrow A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C = [1, \ 0].$$

$$y(t) = x_1(t)$$

Znaleźć liniowy statyczny regulator LQR minimalizujący wskaźnik kwadratowy z wagą r

$$J = \int_0^\infty [y^2(t) + r \cdot u^2(t)] dt.$$

Sterowanie optymalne dane jest wzorem $u^o(t) = -r^{-1}B^T K x(t) = -\mathbf{K} x(t)$. Stała symetryczna macierz $K[2 \times 2]$ spełnia równanie algebraiczne Riccatiego

$$KA + A^T K - KB r^{-1} B^T K + Q = 0.$$

$$J = \int_0^\infty [y^2(t) + r \cdot u^2(t)] dt = \int_0^\infty [x^T(t) \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} x(t) + r \cdot u^2(t)] dt, \text{ stąd waga } Q = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

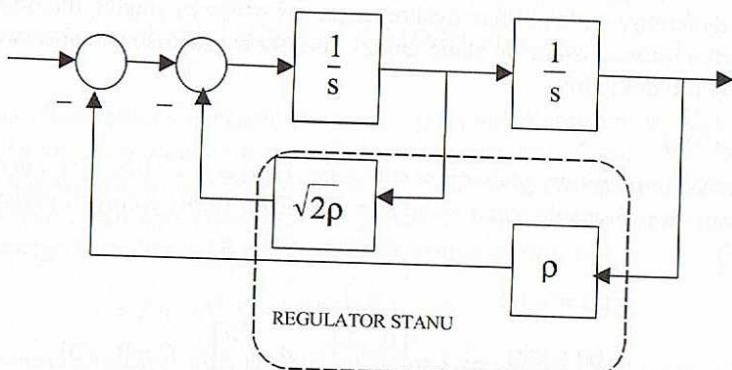
$$\begin{aligned} & \begin{bmatrix} k_1 & k_2 \\ k_2 & k_3 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} k_1 & k_2 \\ k_2 & k_3 \end{bmatrix} - \\ & - \begin{bmatrix} k_1 & k_2 \\ k_2 & k_3 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} k_1 & k_2 \\ k_2 & k_3 \end{bmatrix} \cdot \frac{1}{r} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

Po przekształceniach otrzymuje się trzy równania:

$$\left. \begin{array}{l} k_2^2 - r = 0 \\ rk_1 - k_2 k_3 = 0 \\ 2rk_2 - k_3^2 = 0 \end{array} \right\}, \text{ których rozwiązanie ma postać: } k_2 = \sqrt{r}, \quad k_1 = \sqrt{2} \cdot \sqrt[4]{r}, \quad k_3 = \sqrt{2} \cdot \sqrt[4]{r^3},$$

$$u^o(t) = -r^{-1}B^T K x(t) = -\mathbf{K} x(t) = -[\rho, \sqrt{2\rho}] \cdot \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}, \text{ gdzie } \rho = \frac{1}{\sqrt{r}}.$$

Schemat blokowy układu zamkniętego przedstawiono na rysunku 11.6.



Rys. 11.6. Schemat blokowy układu blokowego

$$\text{Macierz stanu układu zamkniętego } A_Z = [A - BK] = \begin{bmatrix} 0 & 1 \\ -\rho & -\sqrt{2\rho} \end{bmatrix}.$$

$$\text{Transmitancja układu zamkniętego } G(s) = \frac{1}{s^2 + s\sqrt{2\rho} + \rho} = \frac{1}{s^2 + s\sqrt{2} + 1}.$$

Warto zauważyć, że jest to taka sama transmitancja układu zamkniętego jak w przykładzie 9.5 w podrozdziale 9.4 z kryterium kwadratowym od błędu i jego pochodnej. Widoczna wyraźnie funkcję zastosowanego tam regulatora PD, również jako obserwatora stanu.

Przykład 11.5

Dany jest system liniowy podwójnie całkujący: $G(s) = \frac{1}{s^2}$, lub $\ddot{y}(t) = u(t)$. Odpowiada mu równanie stanu:

$$\begin{aligned} x_1(t) &= x_2(t) \\ x_2(t) = u(t) \Rightarrow A &= \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C = [1, 0]. \\ y(t) &= x_1(t) \end{aligned}$$

Znaleźć liniowy statyczny regulator LQR minimalizujący wskaźnik kwadratowy z wagą $r = 1$

$$\begin{aligned} J &= \int_0^\infty [x_1^2(t) + x_2^2(t) + u^2(t)] dt = \int_0^\infty (x_1^2 + 2x_1 x_2 + x_2^2 + u^2) dt = \\ &= \int_0^\infty [x^T(t) Q x(t) + u^2(t)] dt. \end{aligned}$$

$$\text{Macierz wagowa } Q = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Po obliczeniach uzyskujemy współczynniki macierzy regulatora:

$$k_1 = \sqrt{3} - 1, \quad k_2 = 1, \quad k_3 = \sqrt{3}.$$

$$\text{Sterowanie optymalne dane jest wzorem } u^o(t) = -[1, \sqrt{3}]x(t).$$

$$\text{Macierz stanu układu zamkniętego } A_Z = [A - BK] = \begin{bmatrix} 0 & 1 \\ -1 & -\sqrt{3} \end{bmatrix}.$$

$$\text{Transmitancja układu zamkniętego } G(s) = \frac{1}{s^2 + s\sqrt{3} + 1}.$$

Przykłady 11.4 i 11.5 potwierdzają fakt, że przyjęcie różnych macierzy wagowej i R generuje różne rozwiązania.

Przykład 11.6

Dany jest obiekt I rzędu.

Poniżej wyprowadzono ogólne wzory na regulator LQR dla skończonego odcinka czasu T :

$$\dot{x}(t) = ax(t) + bu(t), \quad J = \int_0^T [qx^2(t) + ru^2(t)] dt,$$

$$u^o(t) = -\frac{b}{r}\psi(t) = -\frac{b}{r}K(t)x(t); \quad K(T) = f,$$

$$\dot{K}(t) = \frac{b^2}{r}K^2(t) - 2aK(t) - q.$$

Pierwiastki prawej strony dane są wzorami:

$$K_{1,2} = \alpha \pm \beta, \quad \alpha = \frac{ra}{b^2}, \quad \beta = \frac{\sqrt{a^2r^2 + b^2qr}}{b^2}.$$

Po scałkowaniu równania różniczkowego na $K(t)$ otrzymuje się zależność

$$\frac{r}{b^2} \frac{1}{2\beta} \left[\ln \frac{[f-K_1][K(t)-K_2]}{[f-K_2][K(t)-K_1]} \right] = T - t.$$

Rozwiązanie na kwadratową macierz wzmacnień regulatora $K(t)$ dane jest wzorem:

$$K(t) = \frac{(f-K_2)K_1 - (f-K_1)K_2 e^{-\xi(T-t)}}{f-K_2 - (f-K_1)e^{-\xi(T-t)}} = K_1 + \frac{K_1 - K_2}{\gamma e^{\xi(T-t)} - 1},$$

$$\gamma = \frac{f-K_2}{f-K_1}, \quad \xi = \frac{2b^2\beta}{r}.$$

Jeżeli współczynnik wagi $f=0$, to

$$K(t) = K_1 \left[1 + \frac{K_1 - K_2}{K_2 e^{\xi(T-t)} - K_1} \right].$$

Dla czasu $T \rightarrow \infty$, regulator K staje się regulatorem stacjonarnym:

$$K = K_1 = \frac{r}{b^2} \left[a + \sqrt{a^2 + \frac{b^2}{r}q} \right], \quad u^o(t) = -\left[\frac{a}{b} + \sqrt{\frac{a^2}{b^2} + \frac{q}{r}} \right] x(t).$$

Dla układu stabilnego $a < 0$, przyjęcie $q=0$ we wskaźniku J oznacza pominięcie błędu zmiennej stanu i poszukiwanie sterowania stabilizującego, ale najtańszego (z minimalną

formą). Z ostatniego wzoru (dla $T \rightarrow \infty$), otrzymuje się wtedy rozwiązanie trywialne, $K=0$, co oznacza, że najlepiej systemem nie sterować, gdyż i tak sam powróci do stanu zerowego.

Ten charakterystyczny przypadek, dla skończonego czasu sterowania T daje już rozwiązanie nietrywialne. Dla problemu LQR ze skończonym czasem sterowania T i założonymi zerowymi macierzami wagowymi $F=0$ i $Q=0$, wskaźnik (11.9) reprezentuje tylko energię (koszty) sterowania

$$J = 0,5 \int_0^T [u^T(t)Ru(t)] dt,$$

sterowanie optymalne, sprowadzające sterowalny system z x_0 do $x(T)$ w zadany czasie T , ma tzw. postać minimalno-energetyczną

$$u(t) = R^{-1}B^T e^{-A^T t} G^{-1} \cdot \left[e^{-AT} x(T) - x_0 \right],$$

gdzie nieosobliwa macierz Grama G ma postać (8.1)

$$G = \int_0^T e^{-A\tau} BB^T e^{-A^T \tau} d\tau.$$

Szczegóły wyprowadzenia można znaleźć np. w [11LEF]. Postać ta reprezentuje uogólnienie minimalnonormowego zadania dyskretnego (11.8) na układy ciągłe, dla $x(T) \neq 0$ i niejednostkowej macierzy wagowej R .

Innym ciekawym wynikiem teorii sterowania łączącym tematykę podrozdziałów 11.1 i 11.3 jest spostrzeżenie, że dla układów wielowymiarowych MIMO istnieje nieskończenie wiele regulatorów stanu $u = -Kx$, realizujących zadany zestaw wartości własnych macierzy stanu układu zamkniętego $A - BK$. Można tę nieokreśloność zmniejszyć, łącząc zadanie lokowania zer i biegunków z zadaniem regulatora LQR, ze wskaźnikiem (11.13) o nieskończonym horyzoncie sterowania. Warunkiem rozwiązania jest swobodna macierz wagowa Q , której wartości nie narzuca się z góry, ale wylicza się ją iteracyjnie za pomocą równania ARE (11.14). Tak więc dla zadanej wartości macierzy wagowej R można wyliczyć macierz wagową błędu stabilizacji Q i macierz regulatora K , które minimalizują wskaźnik jakości (11.13) i gwarantują zadane wartości własne macierzy układu zamkniętego $(A - BK)$, [11SOO; 11KAT].

11.5. Nieliniowy czasooptymalny regulator od stanu

Wśród wielu zadań nadążania i stabilizacji, czas osiągnięcia pożądanego punktu w przestrzeni stanu lub czas powrotu do punktu równowagi może być najbardziej pożądaną cechą, bez względu na koszty sterowania (np. szybkość pozycjonowania głowicy dysku HD decyduje o jego czasie dostępu, czas jazdy suwnicy o wypracowanym zysku, a czas przechwytcia rakiety o powodzeniu akcji). Intuicja mówi, że funkcja sterująca realizująca takie zadanie powinna przyjmować maksymalną dopuszczalną wartość, a więc powinna osiągać

swoje ograniczenie. Formalnym, minimalizowanym wskaźnikiem jakości, będzie czas przejścia od jednego stanu do innego stanu. Kryterium to można więc przedstawić jako całkowy wskaźnik jakości

$$J = \int_0^T d\tau = T,$$

przy założonym ograniczeniu na amplitudę sterowania $|u(t)| \leq M$ (np. w przestrzeni funkcji ciągły i ograniczonych odpowiada to normie $\|u\| = \sup |u(t)| < M$).

Poniżej podamy ważne twierdzenie przedstawione jest po raz pierwszy przez R. Bellmana w 1956 r. [11ATM].

Twierdzenie 11.1

Dany sterowalny, stacjonarny układ liniowy, posiadający macierz stanu A z rzeczywistymi wartościami własnymi:

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bu(t), \\ x_1(0) &= x_1 \neq 0, \quad x_2(T_{\min}) = 0,\end{aligned}$$

gdzie $x(t) \in R^n$, $u(t) \in R^r$, dla $\forall t > 0$. Na to, aby przeprowadzić ten układ sterowaniem ograniczonym $|u_i(t)| \leq M_i$ z dowolnego punktu x_1 do $x_2 = 0$ w minimalnym czasie T_{\min} należy zastosować maksymalne dopuszczalne sterowanie przedziałami stałe $\pm M_i$, a liczba koniecznych przełączeń dla każdej składowej u_i będzie co najwyżej $n-1$.

Powyższe twierdzenie jest prawdziwe **tylko przy założeniu istnienia sterowania czasooptymalnego**. Dla układów stabilnych zawsze takie sterowanie istnieje. Dla układów niestabilnych dopuszczalna amplituda M może być za mała w stosunku do odlegiego startowego warunku początkowego i układ nie będzie sterowalny praktycznie. Wtedy sterowanie optymalne nie istnieje. Zostało to już omówione w podrozdziale 8.1 przy okazji rozważań o sterowalności praktycznej.

Na podstawie twierdzenia 11.1 rozwiązanie problemu będzie więc polegało na znalezieniu kolejnych czasów przełączeń i znaku sterowania w pierwszym przedziale. Również w tym zadaniu wprowadza się jak w (11.10) hamiltonian

$$H = 1 + \psi^T(t)[Ax(t) + Bu(t)] \quad (11.15)$$

Wektor $\psi(t)$ zmiennych sprzężonych do stanu $x(t)$ spełnia równanie

$$\dot{\psi}(t) = -\frac{\partial H}{\partial x} = -A^T \psi(t).$$

Ze względu na liniowość hamiltonianu względem u , jego minimum leży na ograniczeniu, jest więc osiągane przez sterowanie przyjmujące wartości graniczne (funkcja przedziałami stała)

$$u(t) = -M \cdot \text{sign}(\psi^T(t)B).$$

Wiele szczegółów i dowodów twierdzeń związanych ze sterowaniem czasooptymalnym przedstawiono w monografiach [11GOH2, 11POL, 11BOW].

Klasyczna synteza regulatora czasooptymalnego przeprowadzona zostanie na przykładzie obiektu II rzędu przy wykorzystaniu analizy trajektorii stanu na płaszczyźnie fazowej.

Przykład 11.7

Dany jest układ II rzędu (podwójnie całkujący) $\ddot{y}(t) = u(t)$. Odpowiadające mu równanie stanu:

$$\begin{aligned}\dot{x}(t) &= \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix}u(t), \quad e^{At} = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}, \\ y(t) &= \begin{bmatrix} 1 & 0 \end{bmatrix}x(t),\end{aligned}$$

Rozwiążanie $x(t)$ dla warunków początkowych i sterowania skokiem jednostkowym $u(t) = M \cdot 1(t)$ ma postać

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} x_{10} + t x_{20} \\ x_{20} \end{bmatrix} + M \begin{bmatrix} t^2/2 \\ t \end{bmatrix} \quad (11.16)$$

Łatwo sprawdzić, że po wyrugowaniu czasu zależność pomiędzy $x_1(t)$ i $x_2(t)$ jest funkcją paraboliczną. Można to również wyprowadzić z równań różniczkowych stanu

$$\begin{aligned}\frac{dx_1}{dt} &= x_2 \\ \frac{dx_2}{dt} &= u \Rightarrow \frac{dx_1}{dx_2} = \frac{x_2}{u} \Rightarrow \int_{x_{10}}^{x_1} u \, dx_1 &= \int_{x_{20}}^{x_2} x_2 \, dx_2 \Rightarrow x_1 = \frac{x_2^2}{2M} + x_{10} - \frac{x_{20}^2}{2M}.\end{aligned}$$

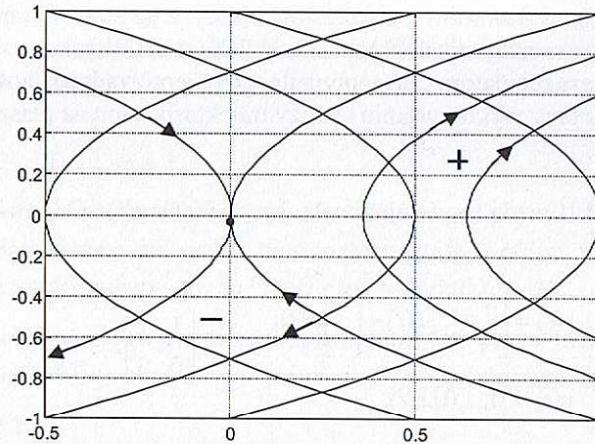
Przy sterowaniu sygnałem stałym $+M$ dla różnych warunków początkowych powstaje rodzina trajektorii stanu z wartościami dającymi do $+\infty$. Przy sterowaniu sygnałem stałym $-M$ dla różnych warunków początkowych powstaje rodzina trajektorii stanu z wartościami x_1 i x_2 dającymi do $-\infty$ (rys. 11.7).

Dla obu sterowań istnieją trajektorie przechodzące dokładnie przez początek układu, opisane równaniem

$$x_1 = \frac{x_2^2}{2M}.$$

Z dolnej połówki trajektorii dla sterowania dodatniego i z górnej połówki trajektorii dla sterowania ujemnego można zbudować jedną linię przełączającą dzielącą płaszczyznę fazową na dwie części. Równanie tej linii to

$$x_1 = -\text{sgn}(x_2) \cdot \frac{x_2^2}{2|M|} = L(x_2).$$

Rys. 11.7. Trajektorie fazowe przy sterowaniu $+M \cdot 1(t)$ i $-M \cdot 1(t)$ dla $M = 0,5$

Sterowanie czasooptymalne, sprowadzające układ z dowolnych warunków początkowych do zera, powinno być funkcją przedziałami stałą $\pm M$, z odpowiednimi momentami przełączeń:

- dla warunków początkowych leżących na prawo od linii $L(x_2)$ sterowanie powinno najpierw przyjmować wartość ujemną, a po osiągnięciu linii $L(x_2)$ przełączyć znak na dodatni (sekwencja $-M, +M$);
- dla warunków początkowych leżących na lewo od linii $L(x_2)$ sterowanie powinno najpierw przyjmować wartość dodatnią, a po osiągnięciu linii $L(x_2)$ przełączyć znak na ujemny (sekwencja $+M, -M$);
- dla warunków początkowych leżących na linii $L(x_2)$ dla $x_1 < 0$ sterowanie powinno przyjąć wartość ujemną $-M$;
- dla warunków początkowych leżących na linii $L(x_2)$ dla $x_1 > 0$ sterowanie powinno przyjąć wartość dodatnią $+M$.

Łatwo więc policzyć i zaprojektować układ sterowania optymalnego w otwartej pętli, przełączając programowo sterowanie $u = \pm M$.

Optymalne czasy sterowania wyrażają się wzorami wynikającymi z (11.16) i z rysunku 11.8.

Dla warunków początkowych x_{10} i x_{20} , leżących powyżej krzywej przełączeń, czasy te wynoszą:

$$t_1 = \frac{x_{20}}{M} + \sqrt{\frac{x_{20}^2}{2M^2} + \frac{x_{10}}{M}} - \text{długość przedziału sterowania ze znakiem } -M,$$

$$t_2 = \sqrt{\frac{x_{20}^2}{2M^2} + \frac{x_{10}}{M}} - \text{długość przedziału sterowania ze znakiem } +M.$$

$$T_{\min} = t_1 + t_2.$$

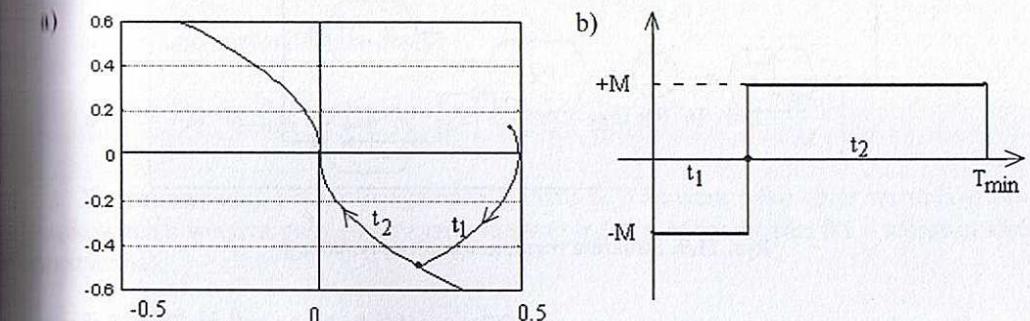
Dla warunków początkowych leżących poniżej krzywej przełączeń symetryczne wzajemną postać:

$$t_1 = -\frac{x_{20}}{M} + \sqrt{\frac{x_{20}^2}{2M^2} - \frac{x_{10}}{M}} - \text{długość przedziału sterowania ze znakiem } +M,$$

$$t_2 = \sqrt{\frac{x_{20}^2}{2M^2} - \frac{x_{10}}{M}} - \text{długość przedziału sterowania ze znakiem } -M.$$

$$T_{\min} = t_1 + t_2.$$

Czasy te poglądowo przedstawiono na rysunku 11.8.

Rys. 11.8. Minimalny czas przejścia T_{\min} na trajektoriach fazowych (a) i na wykresie przebiegu sterowania u (b)

Jeśli jednak w czasie ruchu układu wystąpią jakieś zakłócenia zmiennych stanu $x_1(t)$ i $x_2(t)$, to programowo zadane przełączenie może okazać się za wcześnie lub za późne w stosunku do linii przełączającej i obiekt po wyliczonym z góry czasie T_{\min} nie osiągnie zera układu. Niezawodne staje się dopiero zastosowanie bieżącego pomiaru położenia i prędkości oraz specjalnego regulatora w pętli sprzężenia zwrotnego. W celu syntezy równania regulatora konieczne jest lokalizowanie położenia obiektu na płaszczyźnie stanu względem linii przełączającej L . Odchyłkę ε położenia obiektu od linii przełączającej względem zmiennej x_1 można wyznaczyć następująco:

$$\varepsilon = x_1 - L(x_2) > 0 \Rightarrow u(t) = -M,$$

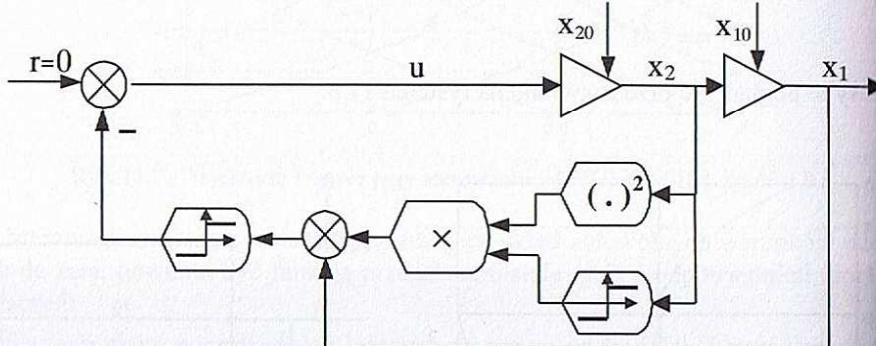
$$\varepsilon = x_1 - L(x_2) < 0 \Rightarrow u(t) = +M,$$

$$\varepsilon = x_1 - L(x_2) = 0 \Rightarrow \begin{cases} x_2 < 0 \Rightarrow u(t) = +M, \\ x_2 > 0 \Rightarrow u(t) = -M. \end{cases}$$

Stąd wynika ogólny zapis równania dla regulatora czasooptymalnego

$$u^o(t) = -M \operatorname{sgn} \left[x_1 + \operatorname{sgn}(x_2) \cdot \frac{x_2^2}{2|M|} \right].$$

Jest to regulator nieliniowy, o strukturze jak na rysunku 11.9, w którym występuje podwójna operacja określania znaku (przekaźniki) oraz operacja mnożenia i potęgowania zmiennej stanu.



Rys. 11.9. Struktura regulatora czasooptymalnego

W porównaniu z reguatorem liniowym LQR z przykładu 11.4 regulator czasooptymalny ma więc bardziej skomplikowaną strukturę. Można go uprościć, wprowadzając strukturę częściowo liniową. Powstaje ona z linearyzacji linii przełączającej i zamiany parabol na linię prostą o zadanym nachyleniu K . Główny przekaźnik decydujący o przełączeniu musi jednak pozostać. Równanie takiego quasi-optymalnego regulatora ma więc postać

$$u^o(t) = -M \operatorname{sgn}[x_1 + K x_2].$$

Dla takiego regulatora nie obowiązuje już twierdzenie 11.1 o $(n-1)$ liczbie przełączeń. Przełączenia może być wiele w zależności od współczynnika wzmacnienia K i warunku początkowego. Dla takiego regulatora występuje też charakterystyczne zjawisko ruchu ślizgowego po końcowym odcinku prostej przełączającej (w pobliżu zera), gdy z uwagi na nie skończoną liczbę przełączeń idealnego przekaźnika ruch punktu do zera odbywa się bezprzewodowo wzdłuż linii przełączającej.

Przykład 11.8

Zastosowanie metodologii z przykładu 11.7 do syntezy regulatora czasooptymalnego dla stabilnego układu całkującąco-inercyjnego:

$$\ddot{x}(t) + \dot{x}(t) = u(t), \quad |u(t)| \leq M$$

równaniu stanu

$$\begin{aligned} \dot{x}(t) &= \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t), \\ y(t) &= [1, 0] x(t). \end{aligned}$$

dalej wzór na formułę linii przełączającej $L(x_2) = L$ i regulatora optymalnego

$$u^o(t) = -M \cdot \operatorname{sgn}[x_1 - L] = -M \cdot \operatorname{sgn} \left[x_1 + x_2 - \operatorname{sgn}(x_2) \cdot M \cdot \ln \left(1 + \frac{|x_2|}{M} \right) \right].$$

Inną formułę otrzymuje się dla niestabilnego układu całkującąco-inercyjnego:

$$\ddot{x}(t) - \dot{x}(t) = u(t), \quad |u(t)| \leq M.$$

Sterowanie optymalne ma postać

$$u^o(t) = -M \cdot \operatorname{sgn}[x_1 - L] = -M \cdot \operatorname{sgn} \left[x_1 - x_2 - \operatorname{sgn}(x_2) \cdot M \cdot \ln \left(1 - \frac{|x_2|}{M} \right) \right].$$

W tym wypadku obszar osiągalności punktu zero zawiera tylko zbiór warunków początkowych z wnętrza pasa płaszczyzny fazowej $|x_2(t)| < M$ (podrozdz. 8.1 – sterowalność praktyczna).

Przykład 11.9

Dany jest układ podwójnie inercyjny, asymptotycznie stabilny, z wartościami własnymi macierzy stanu $\lambda_2 < \lambda_1 < 0$, $\gamma = \lambda_2/\lambda_1 > 1$. Dla tego układu równanie wyjścia i transmitancja są następujące:

$$\ddot{y}(t) - (\lambda_1 + \lambda_2) \dot{y}(t) + \lambda_1 \lambda_2 y(t) = K u(t), \quad G(s) = \frac{K}{(s - \lambda_1)(s - \lambda_2)} \quad (11.17)$$

Przyjmując jego wewnętrzną równoległą strukturę, można otrzymać równanie stanu w postaci,

$$\begin{aligned} \dot{x}(t) &= \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} x(t) + \begin{bmatrix} -\lambda_1 \\ -\lambda_2 \end{bmatrix} u(t), \\ y(t) &= \frac{K}{\lambda_1 - \lambda_2} \begin{bmatrix} -1 & 1 \end{bmatrix} x(t). \end{aligned}$$

Takie zmienne stanu x związane są z naturalnymi zmiennymi stanu (y, \dot{y}) zależnością:

$$\begin{aligned} x_1(t) &= (1/K)[\lambda_1 \lambda_2 y(t) - \lambda_1 \dot{y}(t)] \\ x_2(t) &= (1/K)[\lambda_1 \lambda_2 y(t) - \lambda_2 \dot{y}(t)] \end{aligned} \quad (11.18)$$

Moduł sterowania jest ograniczony, $|u(t)| \leq M$.

Zastosowanie analizy rozwiązań na płaszczyźnie stanu prowadzi do wzoru na linię przełączającą

$$x_2 = L(x_1) = -M \cdot \text{sgn}(x_1) \cdot \left[1 - \left(1 + \frac{|x_1|}{M} \right)^\gamma \right] \quad (11.19)$$

Sterowanie czasooptymalne określone jest przez formułę

$$u^o(t) = \text{sgn} \left[x_2 + M \cdot \text{sgn}(x_1) \cdot \left[1 - \left(1 + \frac{|x_1|}{M} \right)^\gamma \right] \right].$$

Ze wzorów (11.18) wynika, że nie jest możliwym (dla dowolnego γ) wyprowadzenie równania linii przełączającej dla naturalnych zmiennych stanu (y, \dot{y}) obiektu (11.17). Podstawiając zmienne x z (11.18) do równania (11.19) nie da się bowiem rozwiązać tego wzoru ze względu na (y, \dot{y}) i stąd otrzyma się równie proste formuły na regulator czasooptymalny w przestrzeni (y, \dot{y}) . Jakkolwiek nietrudno udowodnić, że sterowanie czasooptymalne znalezione w jednej przestrzeni stanu X pozostanie identyczne dla innej przestrzeni Z uzyskanej przez nieosobliwą liniową transformację zmiennych stanu $x = Pz$. Minimalny czas dojścia do $x(T) = z(T) = 0$ z $x(0)$ będzie taki sam jak z $z(0) = P^{-1}x(0)$.

Stąd tak ważnym zagadnieniem jest umiejemny dobór zmiennych stanu dla celów syntez efektywnych algorytmów stabilizacji i konstrukcji regulatora.

Jak wynika z powyższych rozważań, regulator czasooptymalny od stanu wymaga tak jak i regulator LQR dostępu do wszystkich zmiennych stanu. W rozwiązaniach technicznych nie zawsze można zagwarantować spełnienie tego warunku. Pomocne wtedy staje się zastosowanie obserwatorów stanu.

12. Klasyczne metody obserwacji stanu

12.1. Obserwatory asymptotyczne

12.1.1. Rola obserwatorów

Jak wspomniano w podrozdziale 8.2 i rozdziale 11 obserwatory stanu znalazły zastosowanie głównie w odtwarzaniu niemierzalnego wektora stanu systemu dla celów sterowania. Struktura obserwatorów pozwala też na wykonywanie innych zadań takich jak filtracja szumów czy identyfikacja parametrów. Można je również wykorzystywać do wykrywania usterek w obiekcie sterowanym. Różnice w mierzonym poziomie i w jego estymowanej wartości mogą bowiem świadczyć o nieszczelności zbiornika. Fakt algorytmicznego odtwarzania zmiennych niemierzalnych uzasadnia stosowanie dla nich nazwy *softsensory*. Stąd pojawiają się też ich zastosowania w diagnostyce torów pomiarowo-kontrolnych [12KOJ, 12KOZ]. Łatwiejsze może być sprawdzenie wiarygodności czujnika sprzętowego mierzącego ważną zmienną stanu przez wyliczenie z modelu jej hipotetycznej wartości niż instalowanie dodatkowego czujnika.

Poniżej omówiona zostanie podstawowa funkcja obserwatorów jako rekonstruktore stanu w systemach sterowania. Pokazana jest ona na rysunku 12.1 (trzeci schemat), gdy braku możliwości dostępu do stanu regulator korzysta z wyliczanej estymaty stanu.

Jak przedstawiono w podrozdziale 8.6, dla obserwowlanego systemu liniowego o znanym modelu:

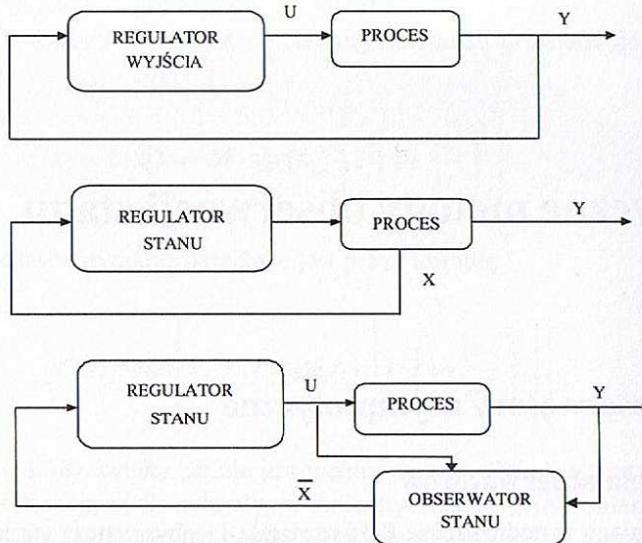
$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), & x(t) \in R^n, y(t) \in R^m, m < n \\ y(t) &= Cx(t), \end{aligned} \quad (12.1)$$

ale nieznanym warunku początkowym $x(0)$, asymptotyczny obserwator stanu ma też postać równania różniczkowego. Jego rozwiązanie $\bar{x}(t)$ startując z arbitralnie przyjętego stanu początkowego $\bar{x}(0)$, nadąża za stanem rzeczywistym $x(t)$. Nadążanie jest asymptotycznie, bieżące wtedy, gdy macierz F będzie asymptotycznie stabilna i równa $F = A - GC$.

$$\dot{\bar{x}}(t) = F \bar{x}(t) + Bu(t) + G y(t), \quad \bar{x}(0) = \bar{x}_0,$$

$$\lim_{t \rightarrow \infty} \varepsilon(t) \rightarrow 0 \Leftrightarrow F \text{ będzie asymptotycznie stabilna},$$

$$\dot{\bar{x}}(t) = (A - GC) \bar{x}(t) + Bu(t) + G y(t), \quad \bar{x}(0) = \bar{x}_0 \quad (12.2)$$



Rys. 12.1. Trzy schematy regulacji

Jeśli w równaniu wyjścia obiektu (12.1) występuje człon $Du(t)$, to równanie obserwatora przyjmuje ogólniejszą postać:

$$\dot{\bar{x}}(t) = (A - G C) \bar{x}(t) + (B - G D) u(t) + G y(t), \quad \bar{x}(0) = \bar{x}_0.$$

Dobór nastaw współczynników wzmocnień obserwatora G wykonuje się wybraną metodą w zależności od typu przyjętego obserwatora. Jeżeli pomiary sterowania i wyjścia nie są zaszumione (zakłócenia pomijalne) i podstawowym problemem jest nieznajomość stanu początkowego, to stosowana deterministyczna odmiana (12.2) nosi nazwę obserwatora Luenbergera (1962), [12ORJ]. Współczynniki wzmocnień obserwatora G dobiera się metodą lokowania wartości własnych macierzy F . Jeżeli system jest obserwowlany, to dla dowolnie wybranych wartości własnych macierzy F można zawsze znaleźć odpowiednie współczynniki macierzy G . Ogólne wzory analityczne dla systemu MISO daje formuła Ackermannia (12.4) [11FRG].

Jeżeli odtwarzanie stanu odbywa się w obecności dużych szumów typu „gausowski szum biały”, to obserwator oprócz asymptotycznej zbieżności powinien przede wszystkim cechować się dobrymi własnościami filtracyjnymi, tzn. estymata stanu powinna gwarantować najmniejszą wariancję błędu. Oznacza to, że nawet przy zgodnych warunkach początkowych pojawiający się (ze względu na szумy pomiarowe) stochastyczny błąd odtwarzania powinien być najmniejszy w sensie wybranego kryterium. Takimi własnościami cechuje się obserwator (12.9) ze specjalnie dobraną macierzą wzmocnień G . Teorię tego obserwatora w przestrzeni stanu opracował R. Kalman najpierw dla systemu dyskretnego (1960, [12KAR1]), a następnie wspólnie z R. Bucy dla systemu ciągłego (1961) [12KAR2; 12CHC]. Od tej pory ta odmiana optymalnego obserwatora stanu nosi nazwę filtra Kalmana lub filtra Kalmana-Bucy. Aby znaleźć taką optymalną macierz G , trzeba rozwiązać dodatkowe równania macierzowe (12.12).

12.1.2. Obserwatory asymptotyczne w układzie ze sprzężeniem zwrotnym

Lącząc w układzie ze sprzężeniem zwrotnym obiekt (12.1), obserwator (12.2) i proporcjonalny regulator stabilizujący stan K dający sterowanie

$$u(t) = -K \bar{x}(t) + r(t),$$

można otrzymać złożony system opisany $2n$ -wymiarowym układem równań

$$\begin{bmatrix} \dot{\bar{x}}(t) \\ \dot{\hat{x}}(t) \end{bmatrix} = \begin{bmatrix} A & -BK \\ GC & A - GC - BK \end{bmatrix} \begin{bmatrix} \bar{x}(t) \\ \hat{x}(t) \end{bmatrix} + \begin{bmatrix} B \\ B \end{bmatrix} r(t).$$

Przyjmując oznaczenie dla błędu estymacji jako $\varepsilon(t) = x(t) - \bar{x}(t)$ i wprowadzając nowe zmienne przez specjalną transformację $T = T^{-1}$, otrzymuje się układ:

$$\begin{bmatrix} x(t) \\ \varepsilon(t) \end{bmatrix} = \begin{bmatrix} I & 0 \\ I & -I \end{bmatrix} \begin{bmatrix} \bar{x}(t) \\ \hat{x}(t) \end{bmatrix},$$

$$\begin{bmatrix} \dot{\bar{x}}(t) \\ \dot{\hat{x}}(t) \end{bmatrix} = \begin{bmatrix} A - BK & BK \\ 0 & A - GC \end{bmatrix} \begin{bmatrix} \bar{x}(t) \\ \hat{x}(t) \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} r(t) \quad (12.3)$$

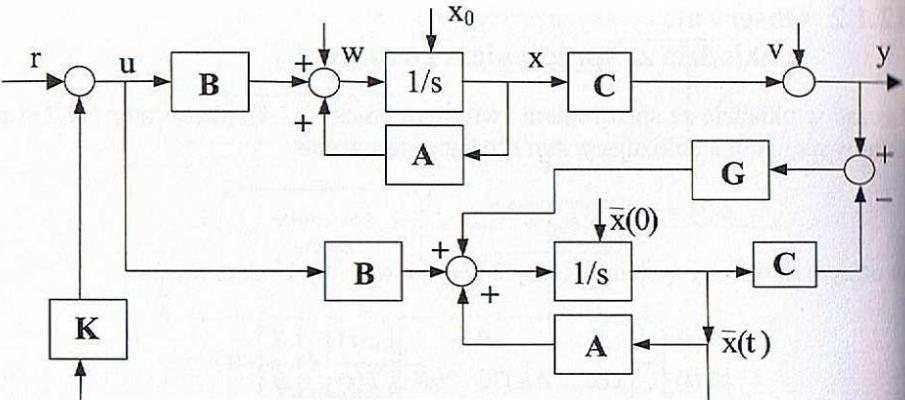
Z tej postaci widać niesterownalność błędu estymacji od sygnału referencyjnego $r(t)$. Zbiór wartości własnych układu zamkniętego składa się z wartości własnych macierzy $A - BK$ i macierzy $A - GC$. Wynika z tego ważna własność układu zamkniętego, tzw. własność separacji, która mówi, że na wartość jednej połowy biegunów układu zamkniętego wpływa macierz regulatora K , a na drugą połowę biegunów macierz obserwatora G . Cały układ ma rzad $2n$, a jego struktura jest przedstawiona na rysunku 12.2.

Traktując regulator i obserwator jak jeden dynamiczny układ korekcyjny w sprzężeniu zwrotnym, często stosuje się do niego nazwę kompensatora.

Transmitancja układu zamkniętego ma postać:

$$G_z(s) = \frac{Y(s)}{r(s)} = [C \ 0] \begin{bmatrix} Is - (A - BK) & -BK \\ 0 & Is - (A - GC) \end{bmatrix}^{-1} \begin{bmatrix} B \\ 0 \end{bmatrix} = C [Is - (A - BK)]^{-1} B.$$

Transmitancja układu ze sprzężeniem zwrotnym i obserwatorem nie zależy od parametrów obserwatora, a tylko od parametrów regulatora, tak jak gdyby stan układu był dostępny do pomiaru. Jest to zrozumiałe, gdy weźmie się pod uwagę fakt, że transmitancję definiuje się dla warunków początkowych równych zero (równość warunków początkowych stanu i estymaty), a wtedy obserwator odtwarza stan dokładnie i jest niewidoczny w torze sterowania. Jednak przy wszelkich niezgodnościach stanu i estymaty (wynikających np. z niedokładnością przyjętego modelu (12.1)) ujawnia się podwójna wymiarowość $[2n \times 2n]$ systemu zamkniętego.



Rys. 12.2. Schemat blokowy z obserwatorem

Jeśli stosowany obserwator jest typu Kalmana (optymalny filtr LQF – *Linear Quadratic Filter*), a regulator jest optymalnym regulatorem (LQR), to całościowy kompensator nosi nazwę LQG (*Linear Quadratic Gaussian*). Generalnie zastosowanie obserwatorów asymptotycznych-różniczkowych w układzie ze sprzężeniem zwrotnym pogarsza jego dynamiczność, co związane jest z podnoszeniem rzędu układu.

12.2. Obserwator Luenbergera

Metoda Ackermanna dla obserwatora Luenbergera pełnego rzędu

Aby wyliczyć kolumnową macierz współczynników wzmacniania G [$n \times 1$] obserwatora Luenbergera (12.2) pełnego rzędu, dla układu obserwowlanego (A, C) i wyjścia skalarnego $y \in R$, można zastosować formuły Ackermanna [11FRG].

Gdy oznaczy się pożądaną postać wielomianu charakterystycznego macierzy obserwatora F ze współczynnikami α_i i pierwiastkami s_i przez $\alpha(s)$

$$\begin{aligned}\alpha(s) &= |sI - A + GC| = (s - s_1)(s - s_2) \cdots (s - s_n) = \\ &= s^n + \alpha_{n-1}s^{n-1} + \dots + \alpha_1s + \alpha_0,\end{aligned}$$

odpowiadająca mu postać wielomianu macierzowego ma postać

$$\alpha(A) = A^n + \alpha_{n-1}A^{n-1} + \dots + \alpha_1A + \alpha_0I.$$

Można zastosować gotowy wzór na współczynniki macierzy obserwatora pełnego rzędu G [$n \times 1$], wykorzystując macierz obserwowlności Q_o (8.11)

$$G = \alpha(A) \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \quad (12.4)$$

Obserwator Luenbergera o zredukowanym rzędzie [12ORJ]

W 1962 roku D.G. Luenberger zauważył, że ponieważ do obserwatora stanu $x \in R^n$ dociera informacja z wyjścia układu $y \in R^m$, gdzie $m < n$, to możliwa jest wersja obserwatora stanu o niższym rzędzie równania różniczkowego niż n . Zaproponował postać obserwatora zredukowanego, który wymaga mniejszej liczby obliczeń związanych z rozwiązywaniem równania obserwatora. Równanie różniczkowe wewnętrznego stanu obserwatora $z(t)$ ma rzad $n - m$, a estymata stanu jest otrzymywana jako równanie algebraiczne wyjścia obserwatora.

$$\begin{aligned}\dot{z}(t) &= Fz(t) + Hu(t) + Gy(t) \\ \bar{x}(t) &= Nz(t) + My(t) \\ z \in R^{n-m}, \quad x \in R^n, \quad \bar{x} \in R^n, \quad u \in R^r, \quad y \in R^m.\end{aligned}\quad (12.5)$$

Zmodyfikowana została również definicja błędu estymacji przez wprowadzenie dodatkowej macierzy transformacji P

$$\varepsilon(t) = z(t) - Px(t).$$

Wtedy otrzymujemy:

$$\begin{aligned}\dot{\varepsilon}(t) &= \dot{z}(t) - P\dot{x}(t) = Fz(t) + Hu(t) + Gy(t) - PAx(t) - PBu(t) = \\ &= F[\varepsilon(t) + Px(t)] + Hu(t) + GCx(t) - PAx(t) - PBu(t) = \\ &= F\varepsilon(t) + (FP + GC - PA)x(t) + (H - PB)u(t).\end{aligned}$$

Aby otrzymać warunek:

$$\dot{\varepsilon}(t) = F\varepsilon(t),$$

$$\varepsilon(t) = e^{Ft}\varepsilon(0),$$

musi zachodzić ogólniejsza niż w (12.2) zależność na macierz F :

$$PA - GC = FP \text{ lub } PA = FP + GC,$$

lub inaczej

$$PA = [F, \quad G] \cdot \begin{bmatrix} P \\ C \end{bmatrix} \quad (12.6)$$

oraz $H = PB$.

Algorytm wyliczania poszczególnych macierzy jest następujący:

1. Prostokątną macierz $P[(n-m) \times n]$ należy wybrać tak, aby dobudowana do macierzy C dala nieosobliwą kwadratową macierz $[n \times n]$

$$\begin{bmatrix} P \\ C \end{bmatrix}, \text{ taką że } \begin{bmatrix} P \\ C \end{bmatrix} \bar{x}(t) = \begin{bmatrix} z(t) \\ y(t) \end{bmatrix}.$$

2. Po wyliczeniu macierzy odwrotnej definiuje się dwie podmacierze $N[n \times (n-m)]$ i $M[n \times m]$.

$$\bar{x}(t) = \begin{bmatrix} P \\ C \end{bmatrix}^{-1} \begin{bmatrix} z(t) \\ y(t) \end{bmatrix} = [N \quad M] \begin{bmatrix} z(t) \\ y(t) \end{bmatrix}.$$

Otrzymuje się założone w (12.5) równanie wyjścia obserwatora

$$\begin{aligned} \bar{x}(t) &= Nz(t) + My(t) = N[\varepsilon(t) + Px(t)] + MCx(t) = \\ &= N\varepsilon(t) + (NP + MC)x(t). \end{aligned}$$

Ponieważ jednak zachodzi równanie ograniczeń

$$[N \quad M] \begin{bmatrix} P \\ C \end{bmatrix} = NP + MC = I,$$

to ostatecznie estymata nadają za stanem $x(t)$ z pewnym błędem

$$\begin{aligned} \bar{x}(t) &= N\varepsilon(t) + (NP + MC)x(t) = x(t) + N\varepsilon(t), \\ \bar{x}(t) - x(t) &= N\varepsilon(t) = Ne^{Ft}\varepsilon(0). \end{aligned}$$

Asymptotyczna stabilność F gwarantuje asymptotyczną zbieżność estymaty do stanu $x(t)$.

3. Ponieważ

$$PA = [F \quad G] \cdot \begin{bmatrix} P \\ C \end{bmatrix} \Rightarrow [F \quad G] = PA[N \quad M],$$

stąd macierze

$$\left. \begin{array}{l} F = P A N \\ G = P A M \\ H = P B \end{array} \right\} \quad (12.7)$$

Naturalna kolejność wyliczania macierzy to:

- przyjęcie P ,
- wyliczenie z macierzy odwrotnej N i M ,
- wyliczenie F , G , oraz H .

Kolejność ta nie gwarantuje jednak otrzymania macierzy stanu obserwatora F o zadanach wartościach własnych. Zadanie odwrotne, tzn. przyjęcie macierzy F i wyliczanie odpowiedniej macierzy P , wymaga procedur iteracyjnych.

Końcowe analityczne i nieiteracyjne wzory dla obserwatora zredukowanego o zadanach z góry wartościach własnych można podać tylko dla specjalnych postaci obiektu obserwowanego. Dotyczy to np. przypadku, gdy dla dowolnej macierzy stanu A jedna zmiana na stanu jest bezpośrednio mierzona. Zastosować można wówczas odmianę wzoru Ackera (12.4) prezentowaną dalej.

Dany jest system ze stanem $x = [x_1, x_2]^T \in R^n$, wyjściem skalarnym $y(t) = x_1(t) \in R$, sterowaniem $u(t) \in R^r$.

$$\begin{aligned} \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} &= \begin{bmatrix} a_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u(t), \\ y(t) &= [1 \quad 0 \quad \dots \quad 0] \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}. \end{aligned}$$

Obserwator zredukowany rzędu $(n-1)$ z kolumnową macierzą wzmacnień $G_r[(n-1) \times 1]$, oddziałującą niemierzalne współrzędne $x_2(t)$ wektora stanu $x(t)$, ma postać:

$$\dot{\bar{x}}_2(t) = [A_{22} - G_r A_{12}] \bar{x}_2(t) + [A_{21} - G_r a_{11}] y(t) + [B_2 - G_r B_1] u(t).$$

Oznaczając przez $\alpha_r(s)$ postać wielomianu charakterystycznego macierzy obserwatora zredukowanego $F_r = (A_{22} - G_r A_{12})$ ze współczynnikami α_i i pożądanymi pierwiastkami s_i ,

$$\begin{aligned} \alpha_r(s) &= |sI - A_{22} + G_r A_{12}| = (s - s_1)(s - s_2) \cdots (s - s_{n-1}) = \\ &= s^{n-1} + \dots + \alpha_1 s + \alpha_0, \end{aligned}$$

oraz odpowiadającą mu postać wielomianu macierzowego przez α_r

$$\alpha_r(A_{22}) = A_{22}^{n-1} + \dots + \alpha_1 A_{22} + \alpha_0 I,$$

otrzymuje się gotowy wzór na współczynniki wzmacnień macierzy obserwatora G_r ,

$$G_r = \alpha_r(A_{22}) \begin{bmatrix} A_{12} \\ A_{12}A_{22} \\ \vdots \\ A_{12}A_{22}^{n-2} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

Dla układu obserwowalnego jednowierszowa macierz A_{12} i kwadratowa $(n-1) \times (n-1)$ macierz A_{22} gwarantują odwracalność macierzy w powyższym wzorze.

Inny wynik można otrzymać również dla obserwatora zredukowanego o rzędzie $(n-1)$, w przypadku obserwowalnego systemu MISO o postaci Frobeniusa dla macierzy stanu A .

Dla zadanych wartości własnych obserwatora s_i oraz mierzonej bezpośrednio jednej współrzędnej stanu $x_n(t)$, otrzymuje się wyniki [12ORJ], przedstawione poniżej:

- dany obserwowalny system MISO typu IIc (wzory (3.22)) z macierzami A , B , C :

$$A = \begin{bmatrix} 0 & 0 & 0 & -a_0 \\ 1 & 0 & 0 & -a_1 \\ .. & 1 & .. & .. \\ .. & .. & 1 & -a_{n-1} \end{bmatrix}_{n \times n}, \quad B = [b_1 \quad \dots \quad b_r]_{n \times r}, \quad b_i \in R^n,$$

$$C = [0 \quad 0 \quad \dots \quad 1]_{1 \times n}, \quad D = 0,$$

- pożądana macierz stanu F , obserwatora zredukowanego o rzędzie $(n-1)$ i macierz P

$$F = \begin{bmatrix} 0 & 0 & 0 & -\alpha_0 \\ 1 & 0 & 0 & -\alpha_1 \\ .. & 1 & .. & .. \\ .. & .. & 1 & -\alpha_{n-2} \end{bmatrix}_{(n-1) \times (n-1)}, \quad P = \begin{bmatrix} 1 & 0 & 0 & -\alpha_0 \\ 0 & 1 & 0 & -\alpha_1 \\ .. & 0 & .. & .. \\ .. & .. & 1 & -\alpha_{n-2} \end{bmatrix}_{(n-1) \times n}$$

powinny również (tak jak A) mieć postać Frobeniusa

Współczynniki α_i można otrzymać z wielomianu charakterystycznego macierzy stanu obserwatora F dla zadanych wartości własnych obserwatora s_i :

$$\begin{aligned} \det |sI - F| &= (s - s_1)(s - s_2) \cdots (s - s_{n-1}) = \\ &= s^{n-1} + \alpha_{n-2}s^{n-2} + \dots + \alpha_1s + \alpha_0. \end{aligned}$$

Wtedy macierze N i M (12.7) mają postacie:

$$N = \begin{bmatrix} I_{n-1} \\ 0 \end{bmatrix}_{n \times (n-1)}, \quad M = [-\alpha_0 \quad -\alpha_1 \quad \cdots \quad -\alpha_{n-2} \quad 1]_{n \times 1}^T,$$

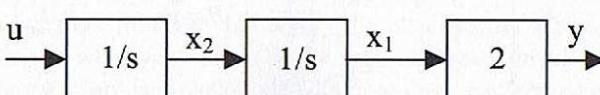
a jednokolumnowa macierz wzmocnień obserwatora G_r wyraża się wzorem

$$G_r = P \cdot A \cdot M = \begin{bmatrix} a_0 & -\alpha_0(a_{n-1} + \alpha_{n-2}) \\ a_1 & +\alpha_0 - \alpha_1(a_{n-1} + \alpha_{n-2}) \\ a_2 & +\alpha_1 - \alpha_2(a_{n-1} + \alpha_{n-2}) \\ \vdots & \\ a_{n-2} & +\alpha_{n-3} - \alpha_{n-2}(a_{n-1} + \alpha_{n-2}) \end{bmatrix}_{(n-1) \times 1} \quad (12.8)$$

Przykład 12.1

Dla układu $G(s) = 2/s^2$ podać obserwatory Luenbergera pełnego i niepełnego rzędu (rys. 12.3).

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C = [2, \quad 0].$$



Rys. 12.3. Obiekt obserwowany

- Macierze wzmocnień obserwatora pełnego rzędu G i stanu F są postaci:

$$G = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}, \quad F = A - GC = \begin{bmatrix} -2g_1 & 1 \\ -2g_2 & 0 \end{bmatrix}, \quad \det(sI - F) = s^2 + 2g_1s + 2g_2 = 0,$$

$$s_{12} = -g_1 \pm \sqrt{g_1^2 - 2g_2}$$

Dla założonych $s_1 = -2$ i $s_2 = -4$ z powyższych wzorów otrzymuje się $g_1 = 3$ i $g_2 = 4$. Te same wartości otrzymuje się ze wzoru (12.4). Obserwator (12.2) ma postać:

$$\dot{\bar{x}}(t) = \begin{bmatrix} -6 & 1 \\ -8 & 0 \end{bmatrix} \bar{x}(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t) + \begin{bmatrix} 3 \\ 4 \end{bmatrix} y(t).$$

- Obserwator niepełnego rzędu jest określony przez przykładową macierz P ze współczynnikami p ze wzoru (12.6):

$$P = [p \quad 1], \quad \begin{bmatrix} P \\ C \end{bmatrix} = \begin{bmatrix} p & 1 \\ 2 & 0 \end{bmatrix}, \quad \begin{bmatrix} P \\ C \end{bmatrix}^{-1} = [N, \quad M] = \begin{bmatrix} 0 & 0,5 \\ 1 & -p/2 \end{bmatrix},$$

$$F = PAN = p, \quad G = PAM = -p^2/2, \quad H = PB = 1$$

Dla wybranego parametru $p = -2$, równania obserwatora są dane wzorami:

$$\dot{z}(t) = -2z(t) + u(t) - 2y(t),$$

$$\bar{x}(t) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} z(t) + \begin{bmatrix} 0,5 \\ 1 \end{bmatrix} y(t).$$

12.3. Optymalny filtr Kalmana

12.3.1. Ciągły filtr Kalmana

Dany jest stacjonarny ciągły układ pracujący w obecności szumów sterowania (lub stanu) i wyjścia:

$$\dot{x}(t) = Ax(t) + Bu(t) + w(t),$$

$$y(t) = Cx(t) + Du(t) + v(t).$$

gdzie $w(t)$ i $v(t)$ są białymi szumami Gaussowskimi o wartości średniej $E\{w(t)\} = 0$ i $E\{v(t)\} = 0$ oraz macierzy kowariancji $Q \geq 0$, $S \geq 0$, $R > 0$ ($R > 0$ dla uniknięcia osobliwości zadania)

$$E \left\{ \begin{bmatrix} w(t_1) \\ v(t_1) \end{bmatrix} \begin{bmatrix} w^T(t_2) & v^T(t_2) \end{bmatrix} \right\} = \begin{bmatrix} Q(t_1) & S(t_1) \\ S(t_1) & R(t_1) \end{bmatrix} \delta(t_1 - t_2).$$

Znaleźć najlepszy układ filtru (obserwatora) pełnego rzędu, który dla $\forall t$ zagwarantuje najlepsze odfiltrowanie szumów odtwarzanej zmiennej stanu $\bar{x}(t)$.

$$\dot{\bar{x}}(t) = (A - GC)\bar{x}(t) + (B - GD)u(t) + Gy(t) \quad (12.9)$$

Przyjmuje się założenia o warunkach początkowych $x(0)$ – zmienna losowa niezależna od w i v o wartości średniej $E\{x(0)\} = \bar{x}(0)$ i wariancji $E\{[x(0) - \bar{x}(0)][x(0) - \bar{x}(0)]^T\} = Q_0$. Chwilowy błąd estymacji stanu dany jest równaniem

$$\varepsilon(t) = x(t) - \bar{x}(t).$$

Wartość średnia błędu i macierz wariancji błędu:

$$E\{\varepsilon(t)\} = \tilde{\varepsilon}(t), \quad E\{[\varepsilon(t_1) - \tilde{\varepsilon}(t_1)][\varepsilon(t_1) - \tilde{\varepsilon}(t_1)]^T\} = P(t_1).$$

Macierz momentów drugiego rzędu

$$E\{\varepsilon(t) \varepsilon^T(t)\} = \tilde{\varepsilon}(t) \tilde{\varepsilon}^T(t) + P(t).$$

Zadaniem jest znalezienie formuły na macierz wzmocnień filtru G , która zagwarantuje najmniejszy błąd średniokwadratowy, tzn. minimum wskaźnika jakości danego wzorem

$$\min_{x_0, G} J = \min_{x_0, G} E\{\varepsilon^T(t) \cdot \varepsilon(t)\},$$

$$E\{\varepsilon^T(t) \cdot \varepsilon(t)\} = \tilde{\varepsilon}^T(t) \tilde{\varepsilon}(t) + \text{trace}[P(t)].$$

Pierwszy człon osiąga minimum dla $\tilde{\varepsilon}(t) = 0$. Można to uzyskać, przyjmując $\tilde{\varepsilon}(0) = 0$, tzn. zakładając, że $\bar{x}(0) = \tilde{x}(0)$ (np. rzeczywisty warunek początkowy jest znany).

Wtedy macierz wariancji błędu równa się

$$P(t) = E\{\varepsilon(t) \cdot \varepsilon^T(t)\}$$

a formuła na błąd średniokwadratowy ma postać

$$J = E\{\varepsilon^T(t) \cdot \varepsilon(t)\} = \text{trace}[P(t)] \quad (12.10)$$

Wskaźnik J osiąga minimum dla optymalnej macierzy współczynników wzmocnień, która jest niestacjonarna $G(t)$ [12KWH] i wyraża się wzorem

$$G(t) = [P(t) \cdot C^T + S] \cdot R^{-1} \quad (12.11)$$

Macierz kwadratowa $P(t)$ spełnia nieliniowe macierzowe równanie różniczkowe Riccatiego

$$\dot{P}(t) = AP(t) + P(t)A^T - G(t)RG^T(t) + Q \quad (12.12)$$

dla $S = 0$

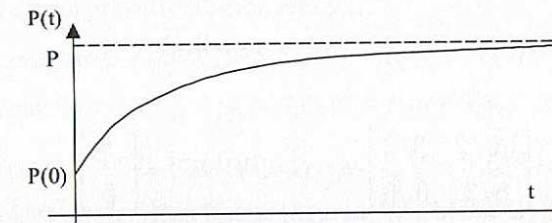
$$\dot{P}(t) = AP(t) + P(t)A^T - P(t)C^T R^{-1} C P^T(t) + Q$$

warunkiem początkowym $P(0) = P_0$.

Rozwiązyaniem równania różniczkowego Riccatiego (12.12) jest macierz $P(t)$, która zmienia się z czasem, dając niestacjonarny współczynnik wzmocnienia filtru $G(t)[n \times m]$ (12.11).

Równania (12.11) i (12.12) można rozwiązać off-line i po wyliczeniu stosować macierze wzmocnień $G(t)$ w systemie czasu rzeczywistego on line.

Przy założeniu dodatniej określoności (półokreśloności) $P(0)$ i obserwonalności pary macierzowej (A, C) rozwiązanie $P(t)$ zmierza asymptotycznie do jedynej dodatnio określonej (półokreślonej) stałej macierzy $P[n \times n]$, gdy czas $t \rightarrow \infty$ (rys. 12.4).



Rys. 12.4. Rozwiązanie równania Riccatiego

Macierz stała P jest również rozwiązaniem algebraicznego równania Riccatiego ARE.

$$0 = AP + PA^T - PC^T R^{-1} CP^T + Q \quad (12.13)$$

Stosując stałą macierz P z (12.13), otrzymuje się według (12.11) stałą macierz wzmocnień filtru G (filtr stacjonarny), który dla początkowych czasów nie jest jeszcze optymalny.

Dla $S = 0$, $D = 0$ i $A - GC = F$ rozwiązanie równania stacjonarnego filtru Kalmana (12.9) ma postać dwóch całek splotowych

$$\bar{x}(t) = e^{Ft} \bar{x}_0 + \int_0^t e^{F(t-\tau)} Gy(\tau) d\tau + \int_0^t e^{F(t-\tau)} Bu(\tau) d\tau.$$

Przykład 12.2

Dla obiektu z przykładu 12.1a, znaleźć ciągłą stacjonarną wersję filtru Kalmana:

$$\begin{aligned} \dot{x}(t) &= \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t) + \begin{bmatrix} 0 \\ \gamma \end{bmatrix} w(t), \\ y(t) &= \begin{bmatrix} 2 & 0 \end{bmatrix} x(t) + v(t). \end{aligned}$$

Kowarianca szumu $w(t)$ oznaczana jest przez Q , kowarianca szumu $v(t)$ przez $R = r$, a wariancja błędu P , przy czym:

$$Q = \begin{bmatrix} 0 \\ \gamma \end{bmatrix} \begin{bmatrix} 0 & \gamma \end{bmatrix} w_1 = \begin{bmatrix} 0 & 0 \\ 0 & \gamma^2 w_1 \end{bmatrix}, \quad P = \begin{bmatrix} p_1 & p_2 \\ p_2 & p_4 \end{bmatrix} = P^T.$$

Równanie ARE ma postać:

$$0 = AP + PA^T + Q - PC^T R^{-1} CP,$$

$$0 = \begin{bmatrix} p_2 & p_4 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} p_2 & 0 \\ p_4 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \gamma^2 w_1 \end{bmatrix} - \frac{4}{r} \begin{bmatrix} p_1^2 & p_1 p_2 \\ p_1 p_2 & p_2^2 \end{bmatrix},$$

$$\left. \begin{array}{l} 2p_2 - \frac{4}{r} p_1^2 = 0 \\ p_4 - \frac{4}{r} p_1 p_2 = 0 \\ \gamma^2 w_1 - \frac{4}{r} p_2^2 = 0 \end{array} \right\} \text{rozwiązanie ma postać } p_1 = \frac{1}{2} \sqrt{\beta} r, \text{ gdzie } \beta = \gamma \sqrt{\frac{w_1}{r}},$$

$$p_2 = \frac{\beta}{2} r, \quad p_4 = \beta \sqrt{\beta} r,$$

$$P = \begin{bmatrix} \sqrt{\beta}/2 & \beta/2 \\ \beta/2 & \beta\sqrt{\beta} \end{bmatrix} \cdot r, \quad G = PC^T R^{-1} = \begin{bmatrix} \sqrt{\beta} \\ \beta \end{bmatrix}.$$

Macierz stanu F obserwatora ma postać

$$F = A - GC = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} 2\sqrt{\beta} & 0 \\ 2\beta & 0 \end{bmatrix} = \begin{bmatrix} -2\sqrt{\beta} & 1 \\ -2\beta & 0 \end{bmatrix}.$$

Wielomian charakterystyczny i pierwiastki macierzy F :

$$s^2 + 2\sqrt{\beta}s + 2\beta = 0, \quad s_{1,2} = -\sqrt{\beta} \pm j\sqrt{\beta},$$

$$\dot{\bar{x}}(t) = (A - GC)\bar{x}(t) + Bu(t) + Gy(t),$$

$$\dot{\bar{x}}(t) = \begin{bmatrix} -2\sqrt{\beta} & 1 \\ -2\beta & 0 \end{bmatrix}x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix}u(t) + \begin{bmatrix} \sqrt{\beta} \\ \beta \end{bmatrix}y(t).$$

Dla $\beta = 9$ otrzymuje się pierwiastki i optymalną postać obserwatora (por. przykład 12.1a):

$$s_{1,2} = -3 \pm j3, \quad \dot{\bar{x}}(t) = \begin{bmatrix} -6 & 1 \\ -18 & 0 \end{bmatrix}\bar{x}(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix}u(t) + \begin{bmatrix} 3 \\ 9 \end{bmatrix}y(t).$$

12.3.2. Dyskretny filtr Kalmana

W zastosowaniach komputerowych systemów sterowania często obiekt sterowania jest modelowany liniowymi równaniami różnicowymi. Dalej przedstawione są dwie wersje

filtru Kalmana dla przypadku obiektu dyskretnego (5.12) niestacjonarnego z dodatkową macierzą wzmocnienia szumu W i macierzą kowariancji $S = 0$:

$$x_{i+1} = A_{Di}x_i + B_{Di}u_i + W_i w_i,$$

$$y_i = C_{Di}x_i + v_i.$$

W dalszych wzorach dla zwiększenia czytelności indeks D macierzy dyskretnych będzie opuszczany, $A_{Di} = A_i$.

Poniżej przedstawione zostaną dwie wersje jednokrokkowego predyktora Kalmana.

1. Wersja jednokrokkowego predyktora Kalmana pracującego przy założeniu, że predykowana estymata stanu $\bar{x}_{i+1} = \bar{x}(i+1| i)$ zależy od bieżącego pomiaru wyjścia y_i i nie będzie korygowana przyszłym pomiarem y_{i+1} .

Macierz wzmocnienia: $G_i = A_i P_i C_i^T [C_i P_i C_i^T + R_i]^{-1}, \quad P(0) = P_0$

Predykcja stanu: $\bar{x}_{i+1} = [A_i - G_i C_i]\bar{x}_i + B_i u_i + G_i y_i, \quad \bar{x}(0) = \bar{x}_0 \quad (12.14)$

Macierz wariancji: $P_{i+1} = A_i P_i A_i^T - G_i C_i P_i C_i^T + W_i Q_i W_i^T$

2. Wersja jednokrokkowego filtra Kalmana (predyktor-korektor) pracującego przy założeniu, że predykowana estymata stanu $\hat{x}_{i+1} = \hat{x}(i+1| i+1)$ zależy od bieżącego pomiaru wyjścia y_i i będzie dodatkowo korygowana przyszłym pomiarem y_{i+1} . Wartość predykowana stanu tylko na podstawie bieżącego pomiaru wyjścia (bez korekcji), tak jak i w wersji 1, będzie oznaczona $\bar{x}_{i+1} = \bar{x}(i+1| i)$.

Oznacza się predykcję stanu: $\bar{x}_{i+1} = \bar{x}(i+1| i)$

i skorygowaną predykcję stanu: $\hat{x}_{i+1} = \hat{x}(i+1| i+1)$

Predykcja stanu: $\bar{x}_{i+1} = A_i \hat{x}_i + B_i u_i, \quad \hat{x}(0) = x_0$

$$\bar{P}_{i+1} = A_i \hat{P}_i A_i^T + W_i Q_i W_i^T$$

Innowacja (residuum): $e_{i+1} = y_{i+1} - \bar{y}_{i+1} = y_{i+1} - C_{i+1} \bar{x}_{i+1}$

Wariancja innowacji: $L_{i+1} = C_{i+1} \bar{P}_{i+1} C_{i+1}^T + R_{i+1}$

$$\begin{aligned} \text{Wzmocnienie filtra:} \quad G_{i+1} &= \bar{P}_{i+1} C_{i+1}^T L_{i+1}^{-1} = \\ &= \bar{P}_{i+1} C_{i+1}^T [C_{i+1} \bar{P}_{i+1} C_{i+1}^T + R_{i+1}]^{-1} \end{aligned}$$

Korekcja wariancji: $\hat{P}_{i+1} = [I - G_{i+1} C_{i+1}] \bar{P}_{i+1}$

$$\begin{aligned} \text{Korekcja predykcji stanu} \quad \hat{x}_{i+1} &= \bar{x}_{i+1} + G_{i+1} e_{i+1} = \\ -\text{czyli równanie filtra:} \quad &= [I - G_{i+1} C_{i+1}] \bar{x}_{i+1} + G_{i+1} y_{i+1} = \\ &= A_i \hat{x}_i + B_i u_i + G_{i+1} e_{i+1} \end{aligned}$$

Warto zwrócić uwagę, że w zastosowaniach do sterowania on-line, w których regulator w bieżącym kroku wypracowuje sterowanie dla następnego okresu próbkowania na podstawie bieżącego pomiaru wyjścia i jednokrokowej predykcji stanu, zgodnie ze wzorem $u_{i+1} = -K\bar{x}_{i+1}$, formalnie używany może być tylko predyktor Kalmana, gdyż dopiero predykowany stan \bar{x}_{i+1} i zależne od niego sterowanie u_{i+1} wygenerują przyszły prawdziwy stan x_{i+1} i prawdziwe wyjście $y_{i+1} = f(u_{i+1})$, którym już nie można korygować użytego \bar{x}_{i+1} . Wykorzystanie predykowanego stanu \hat{x}_{i+1} skorygowanego przez y_{i+1} do generacji sterowania w tej samej próbce jest w sprzężeniu zwrotnym nierealizowne. Dyskretny filtr Kalmana (w wersji predyktor-korektora) najczęściej służy więc do najlepszej filtracji na bieżąco sygnałów pomiarowych (np. wizji czy fonii), bez możliwości wykorzystania tego sygnału w sprzężeniu zwrotnym. Dla malejących czasów próbkowania obie omówione wersje dążą do jednej wspólnej wersji, która jest równoważna wersji ciągłej.

12.3.3. Rozszerzony dyskretny filtr Kalmana

Zastosowanie filtru Kalmana w wersji dyskretnej predyktora-korektora zostało również z powodzeniem zastosowane do filtracji sygnałów z modeli nieliniowych [12MA, 12CHC]. Zmodyfikowana procedura korzysta w każdym kroku dyskretyzacji z modelu nieliniowego, a w algorytmie filtru Kalmana wykorzystuje macierze modelu zlinearyzowanego. Ta wersja zwana jest w literaturze rozszerzonym filtrem Kalmana EKF (*Extended Kalman Filter*).

Dany jest dyskretny układ nieliniowy, opisany równaniem:

$$x_{i+1} = f_i(x_i) + W_i w_i,$$

$$y_i = g_i(x_i) + v_i,$$

przy czym $f_i : R^n \rightarrow R^n$, $g_i : R^n \rightarrow R^m$, gdzie n – wymiar zmiennych stanu, m – liczba wyjść. Założone $\{w_i\}$ i $\{v_i\}$ są szumami białymi, gaussowskimi o zerowej wartości średniej i znanych macierzach kowariancji odpowiednio Q_i i R_i , tzn:

$$E(w_i w_i^T) = Q_i \delta_{i,i}, \quad E(v_i v_i^T) = R_i \delta_{i,i}, \quad E(w_i v_i^T) = 0, \quad E(w_i x_0^T) = 0, \quad E(v_i x_0^T) = 0.$$

Warunek początkowy jest określony jako zmienna losowa o następujących momentach:

$$\bar{x}_0 = E(x_0), \quad \bar{x}_{0|0} = f(\hat{x}_{0|0}), \quad E(x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^T = P_{0,0}.$$

Należy określić $\hat{x}_{i|i}$ dla $i = 1, 2, \dots$ przy równaniu predykcji danym następującym wyrażeniem

$$\bar{x}_{i|i-1} = f_i(\bar{x}_{i-1}).$$

Po rozwinięciu w szereg Taylora funkcji $f_i(x_i)$ w punkcie $\hat{x}_{i|i}$ oraz $g_i(x_i)$ w punkcie $\bar{x}_{i|i-1}$ otrzymujemy:

$$f_i(x_i) \equiv f_i(\hat{x}_{i|i}) + A_i(x_i - \hat{x}_{i|i}),$$

$$g_i(x_i) \equiv g_i(\bar{x}_{i|i-1}) + C_i(x_i - \bar{x}_{i|i-1}),$$

gdzie macierze jacobianowe dane są wzorami:

$$A_i = \left[\frac{\partial f_i(x_i)}{\partial x_i} \right]_{\hat{x}_{i|i}}, \quad C_i = \left[\frac{\partial g_i(x_i)}{\partial x_i} \right]_{\bar{x}_{i|i-1}}.$$

Pozostałe zmienne są zdefiniowane następująco:

$$u_i = f_i(\bar{x}_i) - A_i x_i,$$

$$z_i = y_i - g_i(\bar{x}_{i|i-1}) + C_i \bar{x}_{i|i-1}.$$

Lincaryzacja jest możliwa, jeśli estymata $\hat{x}_{i|i}$ jest określona. W ten sposób model nieliniowy jest aproksymowany modelem liniowym:

$$x_{i+1} = A_i x_i + u_i + W_i w_i,$$

$$z_i = C_i x_i + v_i,$$

dla którego sposób określenia poszczególnych macierzy dla obliczeń w czasie rzeczywistym jest pokazany poniżej.

Pełny algorytm rozszerzonego filtru Kalmana

$$P_{0|0} = \text{Var}(x_0), \quad \bar{x}_0 = E(x_0)$$

$$A_{i-1} = \left[\frac{\partial f_{i-1}(\hat{x}_{i-1|i-1})}{\partial x_{i-1}} \right], \quad C_i = \left[\frac{\partial g_i(\bar{x}_{i|i-1})}{\partial x_i} \right], \quad \text{dla } i = 1, 2, 3, \dots$$

$$\bar{P}_{i|i-1} = A_{i-1} \cdot \hat{P}_{i-1|i-1} A_{i-1}^T + W_{i-1} Q_{i-1} W_{i-1}^T, \quad \hat{P}_0 = P_0$$

$$\bar{x}_{i|i-1} = f_i(\hat{x}_{i-1|i-1}), \quad \hat{x}_0 = x_0 \tag{12.16}$$

$$G_i = \bar{P}_{i|i-1} C_i^T \cdot \left[C_i \bar{P}_{i|i-1} C_i^T + R_i \right]^{-1}$$

$$\hat{P}_{i|i} = [I - G_i C_i] \cdot \bar{P}_{i|i-1}$$

$$\hat{x}_{i|i} = \bar{x}_{i|i-1} + G_i (y_i - g_i(\bar{x}_{i|i-1}))$$

Rozszerzony filtr Kalmana zastosowany dla modelu nieliniowego, nie jest filtrem liniowym i optymalnym, a zmienne $\hat{x}_{i|i}$ i $\hat{P}_{i|i}$ oznaczają aproksymację wartości średniej wa-

runkowej oraz macierzy kowariancji. Wyrażenie służące do obliczania macierzy wzmacnianie G_i jest powiązane z równaniem modelu, ponieważ zależy od macierzy A_i i W_i , co oznacza, że nie można dokonać obliczeń tej macierzy off-line i należy ją obliczać w każdym kroku.

13. Całkowe metody dokładnej obserwacji stanu

13.1. Obserwatory dokładne w przestrzeni $L^2[0, T]$

Do estymacji wektora stanu powszechnie używa się filtru Kalmana. Filtr ten charakteryzuje właściwości wymienione poniżej:

- sformułowanie teorii filtracji oparte jest na procesach stochastycznych;
- estymata stanu jest optymalna w sensie minimum średniokwadratowego błędu odtwarzania;
- estymata stanu startuje z przyjętej początkowej wartości (która jest również wektorem losowym) i dąży asymptotycznie do stanu rzeczywistego;
- obliczenia opierają się na bieżącym pomiarze wartości sterowania i wyjścia i nie wymagają użycia bufora pamięciowego dla próbek z przeszłości.

Cechy te wynikają z zasadniczego założenia: struktura filtru jest oparta na liniowych równaniach różniczkowych (lub różnicowych), podobnych do równania obiektu.

Z tego założenia wynika, że [13FIR]:

- algorytm obserwatora wymaga informacji o stanie początkowym;
- estymata stanu różni się od stanu początkowego i dąży do niego asymptotycznie;
- algorytm potrzebuje informacji o macierzach kowariancji szumów;
- nigdy nie jest znana wartość błędu estymacji, czyli rzeczywista odległość od stanu;
- mogą pojawić się problemy ze zbieżnością algorytmu dla długich obliczeń.

Nowoczesna teoria sterowania i możliwości komputerów umożliwiają zastosowanie innych metod obserwacji. W tych metodach struktura obserwatora związana jest bezpośrednio z definicją obserwonalności, a do opisu modelu i obserwatora wykorzystuje się operatory całkowe. Operacje całkowe na skończonym przedziale mają własność uśredniania wpływu zakłóceń. W przeciwieństwie do nich operacje różniczkowania są bardzo czułe na występowanie zakłóceń w danych pomiarowych. Dodatkową pozytywną cechą operacji całkowych na skończonym odcinku czasu jest pomijanie wpływu dawnych zakłóceń, które występowały w przeszłości, a które już nie zakłócają przebieg procesu. W równaniu różniczkowym taki okresowy wpływ skutkuje teoretycznie aż do nieskończoności.

Poniżej podano najważniejsze cechy obserwatora całkowego (ciągłe) określone na skończonym odcinku czasu [13BYW3, 13FUS]:

- 1) Deterministyczne sformułowanie zadania obserwacji i obserwacji optymalnej.
- 2) Całkowa struktura opisu obserwatora.
- 3) Wykorzystanie ciągłych lub dyskretnych pomiarów wejścia i wyjścia obiektu.
- 4) Jeśli pomiary są idealne (niezaszumione), to po ustalonym czasie obserwacji (znany) obserwator odtworzy dokładną wartość wektora stanu, bez potrzeby wstępного ustalania warunków początkowych.
- 5) W przypadku „zaszumionego” wejścia u i wyjścia y systemu, obserwator posiada własności filtrujące.
- 6) Istnieją ogólne i uniwersalne formuły na optymalną minimalnonormową postać obserwatora dającego minimalną normę błędu odtwarzania stanu.

Najważniejszą własnością omawianych obserwatorów przy idealnych pomiarach jest fakt, że po zadanym (wybranym) czasie T **obserwatory całkowe odtwarzają stan dokładnie**.

Ma to ważne znaczenie w komputerowych systemach sterowania czasu rzeczywistego, gdzie pożądany jest wszelki determinizm obliczeniowy dla wykonania zadań, podlegających „sztywnym” ograniczeniom czasowym.

13.1.1. Przykład obserwatora całkowego odtwarzającego dokładnie stan końcowy

Dany jest obserwowlany system:

$$\begin{aligned}\dot{x}(t) &= A x(t) + B u(t) \\ y(t) &= C x(t)\end{aligned}\quad (13.1)$$

gdzie $x(t) \in R^n$, $u(t) \in R^r$ i $y(t) \in R^m$, $m < n$ dla $\forall t \geq 0$. Macierze A , B , C mają odpowiednie wymiary i stan początkowy jest nieznany.

Załóżmy, że mierzymy ciągłe sterowanie u i wyjście y na przedziale $[0, T]$, gdzie T jest ustalonym okresem obserwacji. Funkcje y i u należą do klasy funkcji $L^2[0, T]$. Przestrzeń wyjścia i sterowań $Y = (L^2[0, T])^m$, $U = (L^2[0, T])^r$, $y \in Y$, $u \in U$. Celem jest określenie stanu $x(T) \in R^n$. Jest to bardziej przydatne w zadaniach sterowania on-line niż odtwarzanie $x(0)$.

Wyjście systemu dla $t \in [0, T]$ ma postać

$$y(t) = Ce^{-A(T-t)}x(T) - C \int_t^T e^{A(t-s)}Bu(s)ds \quad (13.2)$$

W dalszym tekście transpozycje macierzy oznaczone będą znakiem „'”. Podobnie jak miało to miejsce w podrozdziale 8.3, po pomnożeniu wzoru (13.2) przez macierz transponowaną $[Ce^{-A(T-t)}]'$, otrzymuje się kwadratową macierz

$$e^{-A'(T-t)}C'C e^{-A(T-t)} \quad (13.3)$$

która jest osobliwa dla każdego t .

Całkując obustronnie, otrzymuje się

$$\begin{aligned}\int_0^T e^{-A'(T-t)}C'y(t)dt &= \\ &= \int_0^T e^{-A'(T-t)}C'C e^{-A(T-t)} dt \cdot x(T) - \int_0^T \left[e^{-A'(T-t)}C'C \int_t^T e^{A(t-s)}Bu(s)ds \right] dt.\end{aligned}$$

Dla systemu obserwowlanego symetryczna macierz Grama M_T jest nieosobienna dla $\forall T > 0$.

$$M_T = \int_0^T e^{-A'(T-t)}C'C e^{-A(T-t)} dt \quad (13.4)$$

Wtedy $x(T)$ wyraża się wzorem, w którym po zmianie granic całkowania, można wyłączyć sterowanie spod całki wewnętrznej.

$$\begin{aligned}x(T) &= M_T^{-1} \int_0^T e^{-A'(T-t)}C'y(t)dt + M_T^{-1} \int_0^T \left[e^{-A'(T-t)}C'C \int_t^T e^{A(t-s)}Bu(s)ds \right] dt = \\ &= \int_0^T M_T^{-1} e^{-A'(T-t)}C'y(t)dt + \int_0^T \left[M_T^{-1} \int_0^s e^{-A'(T-t)}C'C e^{A(t-s)}dt \right] Bu(s)ds\end{aligned}\quad (13.5)$$

Wzór (13.5) można zapisać skrótnie za pomocą dwóch operatorów macierzowych G_1 i G_2 .

$$x(T) = \int_0^T G_1(t) y(t) dt + \int_0^T G_2(s) u(s) ds \quad (13.6)$$

Daje to ogólną postać obserwatora całkowego w przestrzeniach $L^2[0, T]$.

Jak to zostało pokazane, istnieje nieskończoność wiele zestawów macierzy G_1 i G_2 , które odtwarzają dokładnie $x(T)$ według formuły (13.6). Przykładowe macierze G_1 i G_2 uzyskane przekształceniami (13.3)–(13.5) są więc jedne z wielu możliwych i są uogólnieniem przypadku z podrozdziału 8.3, gdzie wykorzystywano macierz M_0 (8.5).

$$\begin{aligned}G_1(t) &= M_T^{-1} e^{-A'(T-t)}C' = e^{AT} M_0^{-1} e^{A't} C' \\ G_2(t) &= \int_0^t G_1(s) C e^{A(s-t)} ds \cdot B = M_T^{-1} \left(\int_0^t e^{-A'(T-s)} C'C e^{A(s-t)} ds \right) B = \\ &= e^{AT} M_0^{-1} \cdot \left(\int_0^t e^{A's} C'C e^{As} ds \right) e^{-At} B\end{aligned}\quad (13.7)$$

13.1.2. Ogólne warunki istnienia obserwatora całkowego

Odpowiedź systemu obserwowlanego (13.1), przy wymaganym warunku końcowym $x(T)$, jest określona wzorem (13.2)

$$y(t) = Ce^{-A(T-t)}x(T) - C \int_t^T e^{A(t-s)}Bu(s)ds.$$

Dla tego systemu należy określić ogólne warunki dla macierzy G_1 i G_2 tak, aby mogły być używane w ogólnej formule (13.6) na dokładny obserwator stanu $x(T)$. Ma on postać

$$x(T) = \int_0^T G_1^o(T, \tau)y(\tau)d\tau + \int_0^T G_2^o(T, \tau)u(\tau)d\tau \quad (13.8)$$

gdzie macierze funkcyjne $G_1^o(T, \tau)$ i $G_2^o(T, \tau)$ mają wymiary $(n \times m)$ i $(n \times r)$.

Jak wynika z przykładowej postaci tych macierzy (13.5), są one funkcjami dwóch parametrów: zadanego czasu obserwacji T i zmiennej całkowania $\tau \in [0, T]$. Dla przejrzystości w zapisie będziemy omijać pierwszy parametr T . Aby znaleźć warunki istnienia obserwatora, wzór (13.2) zostanie podstawiony do (13.8). Otrzyma się ogólną formułę identyczną

$$\begin{aligned} x(T) &= \int_0^T G_1^o(\tau)Ce^{-A(T-\tau)}d\tau \cdot x(T) - \\ &- \int_0^T G_1^o(\tau)C \left[\int_{\tau}^T e^{A(\tau-s)}Bu(s)ds \right] d\tau + \int_0^T G_2^o(\tau)u(\tau)d\tau \end{aligned}$$

lub po zmianie kolejności całkowania

$$\begin{aligned} x(T) &= \int_0^T G_1^o(\tau)Ce^{-A(T-\tau)}d\tau \cdot x(T) - \\ &- \int_0^T \left[\int_0^s G_1^o(\tau)Ce^{A(\tau-s)}Bd\tau \right] u(s)ds + \int_0^T G_2^o(\tau)u(\tau)d\tau. \end{aligned}$$

Aby prawa strona równała się lewej, musi zachodzić podstawowa zależność

$$\int_0^T G_1(\tau)Ce^{-A(T-\tau)}d\tau = \mathbf{I} \quad (13.9)$$

gdzie \mathbf{I} jest $(n \times n)$ macierzą jednostkową. Formułę (13.9), należy traktować jako ograniczenie na wszystkie możliwe macierze $G_1(\tau)$, określające dopuszczalną klasę macierzy obserwacji,

Drugie ograniczenie dotyczy macierzy $G_2(\tau)$ i wynika z warunku koniecznego na zerowanie się dalszej części wzoru

$$G_2(\tau) = \int_0^{\tau} G_1(s)Ce^{-A(\tau-s)}Bds \quad (13.10)$$

Gdy wybierzemy macierz $G_1(\tau)$ spełniającą (13.9) to $G_2(\tau)$, musi być wyliczona z (13.10).

Jak wspomniano, jest nieskończoność wiele zestawów macierzy G_1 i G_2 liczących ze wzorów (13.9) i (13.10). Pozwalają one na dokładne wyliczenie wektora stanu $x(T)$, jeśli jest spełnione założenie, że model i pomiary $y(t)$ i $u(t)$ są dokładne, tzn. pozbawione zakłóceń pomiarowych. Jednym z zestawów (G_1 i G_2) jest ten przedstawiony w formułach (13.7).

13.1.3. Optymalna postać obserwatora całkowego z minimalną normą

Inaczej wygląda problem, gdy założymy, że w pomiarach y i u wystąpią zakłócenia dodatnie z_1 i z_2 . Stan $x(T)$ jest wtedy odtwarzany z błędem $\epsilon(t) = x(t) - \bar{x}(t)$

$$\begin{aligned} \bar{x}(T) &= \int_0^T G_1^o(\tau) [y(\tau) + z_1(\tau)] d\tau + \int_0^T G_2^o(\tau) [u(\tau) + z_2(\tau)] d\tau = x(T) + \epsilon(T), \\ \epsilon(T) &= \int_0^T G_1^o(\tau) z_1(\tau) d\tau + \int_0^T G_2^o(\tau) z_2(\tau) d\tau. \end{aligned}$$

Sensowne staje się postawienie zadania znalezienia z klasy macierzy dopuszczalnych takich G_1 i G_2 , aby dawały one najmniejszy błąd odtwarzania. Wymaga się również sprecyzowania założień o przychodzących zakłócenach. Sygnały u i y oraz zakłócenia należą do tych samych przestrzeni. Ponieważ $y \in [L^2[0, T]]^m$, a $u \in [L^2[0, T]]^r$, to $z_1 \in [L^2[0, T]]^m$, a $z_2 \in [L^2[0, T]]^r$.

Z uwagi na deterministyczną wersję obserwatora nie będą przyjmowane żadne założenia o stochastycznych własnościach tych zakłóceń. Przyjmie się tylko najbardziej niezbędnie założenie (sensowne w warunkach rzeczywistych), że zakłócenia mają ograniczoną normę $\|z_1\| \leq 1$ i $\|z_2\| \leq 1$. Przyjęcie normalizacji do kuli jednostkowej w $L^2[0, T]$ nie zmniejsza ogólności rozważań. Z ciągłości i liniowości odwzorowań wynika, że pary macierzy funkcyjnych (G_1 i G_2) mogą być potraktowane jako realizacje poszczególnych obserwatorów, będące elementami w przestrzeni $Y \times U = [L^2[0, T]]^{m \times n} \times [L^2[0, T]]^{r \times n}$. Pary te tworzą domkniętą liniową rozmaitość S . Dla elementów tej rozmaitości można wprowadzić półnormę. Jej kwadrat wyraża się wzorem

$$\|(G_1, G_2)\|^2 = \int_0^T \left[\sum_{i=1}^n \sum_{j=1}^m (g_1^{ij}(\tau))^2 + \sum_{i=1}^n \beta_i \sum_{j=1}^r (g_2^{ij}(\tau))^2 \right] d\tau = J \quad (13.11)$$

Wagi $\beta_i \geq 0$ dotyczą sum kwadratów elementów i -tego wiersza macierzy G_2 .

Wielkość tej normy może być traktowana jako wskaźnik J jakości odtwarzania stanu. Widać to na podanym niżej oszacowaniu.

$$\varepsilon(T) = \int_0^T G_1^o(\tau) z_1(\tau) d\tau + \int_0^T G_2^o(\tau) z_2(\tau) d\tau.$$

W przestrzeni L^2 wartość normy błędu $\|\varepsilon\|$ może być zrealizowana przez specjalnie dobrane sygnały zakłócające (najgorsze z kuli jednostkowych) o kształtach równych elementom macierzy $g_1^{ij}(t)$ i $g_2^{ij}(t)$. Dla $\beta = 1$ zachodzi więc oszacowanie

$$\begin{aligned} \max_{(z_1, z_2)} \|\varepsilon\|_{R^n}^2 &= \max_{(z_1, z_2)} \| \langle G_1 | z_1 \rangle + \langle G_2 | z_2 \rangle \|^2 \leq 2 \left[\|G_1\|_Y^2 + \|G_2\|_U^2 \right] = \\ &= 2 \|(G_1, G_2)\|_{Y \times U}^2 = 2J \end{aligned} \quad (13.12)$$

Norma J szacuje więc od góry (w sposób najbardziej pesymistyczny) możliwy błąd estymacji stanu dla najgorszych zakłóceń o normie jeden. Minimalizacja normy obserwatora (sumaryczne pole powierzchni pod kwadratami funkcji $g_1^{ij}(t)$ i $g_2^{ij}(t)$) zagwarantuje najlepsze jego własności w sensie błędu średniokwadratowego). Waga β może być interpretowana jako norma zakłóceń z_2 .

Zadanie optymalizacji ma postać

$$J^o = \min_{(G_1, G_2)} J.$$

Mogą utworzyć funkcję Lagrange'a z funkcji (13.11) i z ograniczenia na G_1 (13.9) w postaci

$$L = J + 2 \sum_{i=1}^n \left[\left[e_i^T - \left[\langle g_1^i, h_1^1 \rangle, \dots, \langle g_1^i, h_1^n \rangle \right] \right] \cdot \lambda_i \right],$$

gdzie e_i to wektor bazowy z R^n z jedynką na i -tej pozycji, g_1^i to i -ty wiersz macierzy G_1 , h_1^i to i -ta kolumna macierzy $C e^{-A(T-t)}$, nawiasy trójkątne $\langle \cdot, \cdot \rangle$ oznaczają iloczyny skalarnego w przestrzeni funkcyjnej, a $\lambda_i \in R^n$ wektory mnożników Lagrange'a.

Biorąc pod uwagę, że w funkcji J obowiązuje zależność pomiędzy macierzami G_1 i G_2 (13.10), to cały funkcjonal Langrange'a można przedstawić jako zależny tylko od nieznanego elementów g_1^i , czyli wierszy macierzy G_1 .

Warunek optymalności $\frac{\delta L}{\delta g_1^i} = 0$ dla każdego wiersza macierzy G_1 doprowadzi do zapisu macierzowego na równanie opisujące optymalne macierze G_1 i G_2 („'” oznacza transpozycję)

$$G_1'(t) = - \int_t^T C e^{A(t-\tau)} B G_2'(\tau) d\tau \beta + C e^{-A(T-t)} \lambda \quad (13.13)$$

β to diagonalna macierz wag β_i , a λ to macierz mnożników Lagrange'a złożona z kolumn λ_i . Równanie (13.13) wraz z (13.10)

$$G_2(\tau) = \int_0^\tau G_1(s) C e^{-A(\tau-s)} B ds$$

daje układ dwóch równań, które można rozwiązać. W tym celu zauważmy, że obowiązują pewne zależności:

$$G_1'(T) = C \lambda, \quad G_2(0) = 0, \quad G_2(T) = B.$$

Wprowadźmy kwadratowe macierze $[n \times n]$, P_1 i P_2 :

$$G_1'(t) = C P_1'(t), \quad G_2'(t) = B' P_2'(t).$$

Równania (13.13) i (13.10) przyjmą wtedy postać:

$$\begin{aligned} P_1'(t) &= e^{-A(T-t)} \lambda - \int_t^T e^{A(t-\tau)} B B' P_2'(\tau) d\tau \cdot \beta \\ P_2'(t) &= \int_0^t e^{-A'(t-\tau)} C' C P_1'(\tau) d\tau \end{aligned} \quad (13.14)$$

z warunkami brzegowymi: $P_1'(T) = \lambda$, $P_2(0) = 0$, $P_2'(T) = I$.

Widać wyraźnie, że równania (13.14) mają charakterystyczną postać rozwiązań dwóch równań różniczkowych liniowych:

$$\begin{aligned} P_1'(t) &= A P_1'(t) + B B' P_2'(t) \beta \\ P_2'(t) &= C' C P_1'(t) - A' P_2'(t) \end{aligned} \quad (13.15)$$

z mieszanymi warunkami brzegowymi: $P_1'(T) = \lambda$, $P_2(0) = 0$, $P_2'(T) = I$.

Oznaczmy przez $p_1^i(t), p_2^i(t)$ kolumny macierzy $P_1'(t), P_2'(t)$, oraz wprowadźmy macierze fundamentalne $\Phi_i(t)$ z indeksem i , dla $i = 1, \dots, n$.

$$\Phi_i(t) = e^{W_i t} = \begin{bmatrix} \Phi_{11}^i(t) & \Phi_{12}^i(t) \\ \Phi_{21}^i(t) & \Phi_{22}^i(t) \end{bmatrix} \quad (13.16)$$

gdzie

$$W_i = \begin{bmatrix} A & \beta_i BB' \\ C'C & -A' \end{bmatrix} \quad (13.17)$$

Dla $2n$ -wymiarowego złożenia wektorów $p_1^i(t), p_2^i(t)$ (przy warunku $p_2^i(0) = 0$) otrzymuje się rozwiązanie

$$\begin{aligned} \begin{bmatrix} p_1^i(t) \\ p_2^i(t) \end{bmatrix} &= \begin{bmatrix} \Phi_{11}^i(t) & \Phi_{12}^i(t) \\ \Phi_{21}^i(t) & \Phi_{22}^i(t) \end{bmatrix} \begin{bmatrix} p_1^i(0) \\ p_2^i(0) \end{bmatrix} = \\ &= \begin{bmatrix} \Phi_{11}^i(t) & \Phi_{12}^i(t) \\ \Phi_{21}^i(t) & \Phi_{22}^i(t) \end{bmatrix} \begin{bmatrix} p_1^i(0) \\ 0 \end{bmatrix} = \begin{bmatrix} \Phi_{11}^i(t) p_1^i(0) \\ \Phi_{21}^i(t) p_1^i(0) \end{bmatrix} \end{aligned} \quad (13.18)$$

Aby ostatecznie podać postać rozwiązania na $p_1^i(t), p_2^i(t)$, trzeba znaleźć warunki początkowe dla wektorów $p_1^i(0)$, przez podstawienie $p_1^i(t)$ z (13.18) do ograniczenia (13.9)

$$\int_0^T e^{-A'(T-\tau)} C' C \left[\Phi_{11}^1(\tau) p_1^1(0), \dots, \Phi_{11}^n(\tau) p_1^n(0) \right] d\tau = I.$$

Stąd oznaczając kolejne liczbowe kwadratowe transponowane macierze M'_i jako

$$M'_i = \int_0^T e^{-A'(T-\tau)} C' C \Phi_{11}^i(\tau) d\tau,$$

można znaleźć kolejne warunki początkowe jako

$$M'_1 p_1^1(0) = [1 \ 0 \ \dots \ 0]', \quad M'_2 p_1^2(0) = [0 \ 1 \ \dots \ 0]', \quad M'_n p_1^n(0) = [0 \ 0 \ \dots \ 1].$$

Ostatecznie rozwiązanie na kolejne kolumny macierzy $P_1'(t), P_2'(t)$ ma postać:

$$\begin{aligned} p_1^i(t) &= \Phi_{11}^i(t) [M'_i]^{-1} e_i \\ p_2^i(t) &= \Phi_{21}^i(t) [M'_i]^{-1} e_i \end{aligned} \quad (13.19)$$

gdzie e_i to wektory bazowe w R^n z jedynką na i -tej pozycji. Wektory $p_1^i(t), p_2^i(t)$ po transpozycji tworzą wiersze optymalnych macierzy $G_1^o(t)$ i $G_2^o(t)$ według wzorów:

$$G_1^o(t) = P_1(t) C', \quad G_2^o(t) = P_2(t) B.$$

Wzory (13.19) dają najogólniejszą postać rozwiązania optymalnego na obserwator całkowy z minimalną normą, dla różnych wag β .

13.1.4. Optymalna postać obserwatora dla wybranych wag

Dla różnych wartości współczynników wagowych β_i we wskaźniku jakości (normie (13.11)) można otrzymać odpowiednie wzory na obserwatory optymalne.

Rozważymy dwa charakterystyczne przypadki.

Przypadek 1

Współczynniki $\beta_i = 0$. Wskaźnik jakości ma postać

$$J = \int_0^T \left[\sum_{i=1}^n \sum_{j=1}^m (g_1^{ij}(\tau))^2 \right] d\tau \quad (13.20)$$

Oznacza to, że nie jest uwzględniana wielkość funkcji, z których zbudowana jest macierz $G_2(t)$. Średniokwadratowej minimalizacji podlega tylko macierz $G_1(t)$, a odpowiednia macierz $G_2(t)$ jest budowana na podstawie wzoru (13.10). Pominięcie macierzy G_2 w zadaniu minimalizacji normy obserwatora może być interpretowane jako faktyczny brak wpływu wielkości G_2 na błąd obserwacji stanu, co może mieć miejsce w szczególnym przypadku – braku zakłóceń w pomiarach sterowania. Założenie takie nie zawsze ma uzasadnienie.

Układ (13.15) przyjmuje postać dwóch równań:

$$\begin{aligned} P_1'(t) &= AP_1'(t), \\ P_2'(t) &= C'C P_1'(t) - A' P_2'(t), \end{aligned} \quad \begin{aligned} \dot{P}'(t) &= \begin{bmatrix} A & 0 \\ C'C & -A' \end{bmatrix} \begin{bmatrix} P_1'(t) \\ P_2'(t) \end{bmatrix} \end{aligned}$$

z warunkami brzegowymi: $P_1'(T) = \Lambda$, $P_2(0) = 0$, $P_2'(T) = I$.

Rozwiązanie dla macierzy $P_1'(t), P_2'(t)$, przyjmuje zwartą postać

$$\begin{bmatrix} P_1'(t) \\ P_2'(t) \end{bmatrix} = \begin{bmatrix} e^{At} & 0 \\ \Phi_{21}(t) & e^{-A't} \end{bmatrix} \begin{bmatrix} P_1'(0) \\ P_2'(0) \end{bmatrix} = \begin{bmatrix} e^{At} & P_1'(0) \\ \Phi_{21}(t) & P_1'(0) \end{bmatrix} \Rightarrow P_1'(t) = e^{At} P_1'(0).$$

Aby ostatecznie podać postać rozwiązań na $P'_1(t), P'_2(t)$, trzeba znaleźć warunki początkowe dla macierzy $P_1(0)$, przez podstawienie $P'_1(t)$ do ograniczenia (13.9)

$$\int_0^T e^{-A'(T-\tau)} C' C e^{A\tau} d\tau \cdot P'_1(0) = I.$$

Oznaczmy liczbową macierz M' jako

$$M' = \int_0^T e^{-A'(T-\tau)} C' C e^{A\tau} d\tau = e^{-A'T} \int_0^T e^{A'\tau} C' C e^{A\tau} d\tau.$$

Wykorzystując $M_T z$ (13.4), można przedstawić macierz M' również jako:

$$M' = M_T e^{AT} \quad \text{lub} \quad M = e^{A'T} M_T.$$

Z powyższych wzorów wynika, że warunki początkowe i końcowe mają postać:

$$P'_1(0) = (M')^{-1} = e^{-AT} [M_T]^{-1}, \quad P'_1(T) = \Lambda = M_T^{-1}.$$

Rozwiązanie na macierze P jest postaci:

$$\begin{aligned} P'_1(t) &= e^{At} \cdot P'_1(0) = e^{At} \cdot (M')^{-1} = e^{-A(T-t)} [M_T]^{-1}, \\ P'_2(t) &= \int_0^t e^{-A'(t-\tau)} C' C P'_1(\tau) d\tau = \int_0^t e^{-A'(t-\tau)} C' C e^{-A(T-\tau)} d\tau \cdot M_T^{-1}. \end{aligned}$$

Stąd można otrzymać wzór na $\phi_{21}(t)$:

$$P'_2(t) = \phi_{21}(t) P'_1(0) = \phi_{21}(t) e^{-AT} M_T^{-1} \Rightarrow \phi_{21}(t) = \int_0^t e^{-A'(t-\tau)} C' C e^{A\tau} d\tau.$$

Ostatecznie rozwiązanie na optymalne macierze obserwatora ma postać:

$$\begin{aligned} G_1^o(t) &= P_1(t) \cdot C' = M^{-1} e^{A't} C' = M_T^{-1} e^{-A'(T-t)} C' \\ G_2^o(t) &= M^{-1} \int_0^t e^{A'\tau} C' C e^{A\tau} d\tau \cdot e^{-At} B = M_T^{-1} \int_0^t e^{-A'(T-\tau)} C' C e^{-A(t-\tau)} d\tau \cdot B \end{aligned} \quad (13.21)$$

Jest to ta sama postać macierzy, jaką otrzymano we wzorach (13.7), przy wyprowadzaniu równań obserwatora techniką mnożenia przez macierz transponowaną. Przyjmując jak wyżej współczynniki $\beta_i = 0$, należy pamiętać, że tak otrzymane macierze obserwacji nie gwarantują najmniejszego błędu obserwacji, gdy zakłócenia pomiarowe pojawią się zarówno w pomiarach wyjścia y i wejścia u (w rzeczywistości jest to często występujący przypadek).

Przypadek 2

Współczynniki $\beta_i = 1$. Wskaźnik jakości ma postać uwzględniającą w równym stopniu wpływ wielkości funkcji G_1 i G_2 na normę obserwatora, czyli uwzględnienie obecności zakłóceń zarówno w sygnale wyjścia, jak i wejścia.

$$J = \int_0^T \left[\sum_{i=1}^n \sum_{j=1}^m (g_1^{ij}(\tau))^2 + \sum_{i=1}^n \sum_{j=1}^r (g_2^{ij}(\tau))^2 \right] d\tau \quad (13.22)$$

Ze wzorów (13.15)–(13.19) wynika, że zachodzi równość wszystkich macierzy $M_i = M$ ostatecznie

$$M = \left[\int_0^T \Phi'_{11}(\tau) C' C e^{A\tau} d\tau \right] e^{-AT},$$

$$\text{macierz odwrotna } M^{-1} = e^{AT} \left[\int_0^T \Phi'_{11}(\tau) C' C e^{A\tau} d\tau \right]^{-1}.$$

Macierz W i macierz fundamentalna mają postaci:

$$W = \begin{bmatrix} A & BB' \\ C' C & -A' \end{bmatrix}, \quad \Phi(\tau) = e^{W\tau} = \begin{bmatrix} \Phi_{11}(\tau) & \Phi_{12}(\tau) \\ \Phi_{21}(\tau) & \Phi_{22}(\tau) \end{bmatrix} \quad (13.23)$$

Optymalne macierze obserwacji wyrażają się wzorami:

$$\begin{aligned} G_1^o(T, \tau) &= P_1(\tau) \cdot C' = M^{-1} \Phi'_{11}(\tau) C' \\ G_2^o(T, \tau) &= P_2(\tau) \cdot B = M^{-1} \Phi'_{21}(\tau) B \end{aligned} \quad (13.24)$$

Przykład 13.1

Dany jest system I rzędu:

$$\begin{aligned} \dot{x}(t) &= -ax(t) + u(t), \\ y(t) &= x(t), \quad t \in [0, T]. \end{aligned}$$

Należy podać zależności na całkowy obserwator optymalny stanu $x(T)$ ze współczynikiem wagowym we wskaźniku jakości $\beta \neq 0$.

Pozorne trywialne zadanie dla systemu skalarnego może mieć sensowne uzasadnienie w przypadku zakłóceń pomiarowych, gdyż zastosowany obserwator będzie spełniał rolę całkowego filtra uśredniającego.

Przyjmując zastępczy współczynnik wagowy $\sqrt{a^2 + \beta} = \mu$ oraz wykorzystując wzory (13.15)–(13.17) dla powyższego przypadku, otrzymuje się:

$$W = \begin{bmatrix} -a & \beta \\ 1 & a \end{bmatrix}, \quad \Phi = e^{Wt} = \frac{1}{\mu} \begin{bmatrix} \mu \cosh(\mu t) - a \sinh(\mu t) & \beta \sinh(\mu t) \\ \sinh(\mu t) & \mu \cosh(\mu t) + a \sinh(\mu t) \end{bmatrix}.$$

Stąd:

$$G_1^o(t) = \frac{\mu \cosh(\mu t) - a \sinh(\mu t)}{\sinh(\mu T)}, \quad G_2^o(t) = \frac{\sinh(\mu t)}{\sinh(\mu T)} \quad (13.25)$$

Po dłuższych obliczeniach otrzymuje się normę obserwatora liczoną według (13.11). Jest ona malejącą funkcją czasu obserwacji T

$$\|(G_1^o, G_2^o)\| = \sqrt{\int_0^T [G_1^o(t)]^2 + \beta [G_2^o(t)]^2 dt} = \sqrt{\mu \operatorname{ctgh}(\mu T) - a} = f(T),$$

dającą do asymptoty $\lim_{T \rightarrow \infty} f(T) = \sqrt{\mu - a} = f(\infty)$.

Wynika z tego ważny wniosek, że im dłuższe jest okno obserwacyjne, tym obserwator gwarantuje mniejszy błąd odtwarzania. Wydłużenie okna nawet do nieskończoności nie prowadzi jednak do wyzerowania błędu, chociaż uzyskuje się jego najmniejszą wartość $f(\infty)$.

Wykorzystując wcześniejsze przedstawione oszacowania najgorszego błędu średniokwadratowego przez normę obserwatora (13.12), $|\varepsilon_{\max}|/\sqrt{2} \leq \|(G_1, G_2)\|$, można policzyć minimalny czasu obserwacji (szerokość okna T) gwarantujący błąd nie większy niż zadane ε .

Po krótkim wyprowadzeniu otrzymamy

$$T_{\min} = \ln((\varepsilon^2 + 2a + 2\mu)/(\varepsilon^2 + 2a - 2\mu))/(2\mu).$$

Przyjęcie współczynnika $\beta = 0$ w tym zadaniu oznacza pominięcie we wskaźniku jakości kary za wielkość elementów macierzy G_2 , co może być interpretowane jako brak zakłóceń pomiarowych sterowania. W takim przypadku $\mu = a$ i optymalne macierze (13.25) mają postać:

$$G_1^o(t) = \frac{ae^{-at}}{\sinh(at)}, \quad G_2^o(t) = \frac{\sinh(at)}{\sinh(at)} \quad (13.26)$$

Norma takiego obserwatora ma postać

$$\|(G_1^o, G_2^o)\| = \sqrt{\int_0^T [G_1^o(t)]^2 dt} = \sqrt{a[\operatorname{ctgh}(at) - 1]},$$

której wartość dla $T \rightarrow \infty$ dąży do zera. Wzory (13.26) można uzyskać bezpośrednio z formuły przypadku 1 (13.21).

Analiza dokładności odtwarzania stanu

Warto przeprowadzić analizę dokładności odtwarzania stanu przez obserwator optymalny, policzony dla wybranej wartości wagi β , w przypadku pojawienia się zakłóceń z_2 sterowaniu, o normie $\hat{\beta}$, innej niż spodziewana β . Według (13.12), dla optymalnych macierzy obserwatora (wynikających z przyjętej wagi β) zachodzi oszacowanie błędu

$$\max_{(z_1, z_2)} \|\varepsilon\| \leq \sqrt{2(\mu \operatorname{ctgh}(\mu T) - a)} \quad (13.27)$$

Jeśli w trakcie stosowania tego obserwatora (optymalnego dla β) pojawią się zakłócenia o innej normie, $\hat{\beta} \neq \beta$, wtedy dla określenia błędu tego obserwatora trzeba zastosować formułę (13.12)

$$\begin{aligned} \max_{(z_1, z_2)} \|\varepsilon\| &\leq \sqrt{2 \int_0^T \{ [G_1^o(t)]^2 + \hat{\beta} [G_2^o(t)]^2 \} dt} = \\ &= \sqrt{[2a^2 + (\beta + \hat{\beta})] \operatorname{ctgh}(\mu T) / \mu - 2a + T(\beta - \hat{\beta}) / \operatorname{sh}^2(\mu T)}. \end{aligned}$$

W tabeli 13.1 przedstawiono wartości błędu policzone powyższą formułą dla różnych zestawów $\hat{\beta}$ i β .

Tabela 13.1

Wartości błędów

$\hat{\beta}$	0	1	5	10	20
β	0,052	0,707	1,578	2,231	3,155
0	0,052	0,707	1,578	2,231	3,155
1	0,164	0,688	1,503	2,119	2,992
5	0,578	0,817	1,414	1,915	2,646
10	0,900	1,038	1,465	1,866	2,481
20	1,310	1,385	1,654	1,938	2,408

Zwłaszcza pierwszy wiersz tabeli jest ważny, gdyż pokazuje wartości błędów dla przypadku optymalnego obserwatora policzonego dla $\beta = 0$, pracującego jednak w obecności zakłóceń o normach $0 < \hat{\beta} = 1, 5, 10, 20$. Stosowanie takiego obserwatora, danego wzorami (13.7) z macierzami obserwacyjnymi, które dla $0 < \hat{\beta} = 1, 5, 10, 20$ nie są już optymalne, powoduje powstanie dodatkowych błędów. Widać, że minimalne wartości błędów występują w kolumnach (dla zadanego poziomu zakłóceń $\hat{\beta}$) tylko dla przypadku obserwatora policzonego z zadanym $\beta = \hat{\beta}$ (na przekątnej), czyli danych optymalnym wzorem (13.27).

Kontynuując przykład 13.1, można odczytać również rozwiązanie zadania obserwacji dla systemu całkującego, tzn. gdy $a = 0$. Wtedy $\sqrt{\beta} = \mu$. W tym wypadku przyjęcie dodat-

kowo wagi $\beta = 0$ (jak w (13.26)), prowadzi do prostych wzorów:

$$G_1^o(t) = \frac{1}{T}, \quad G_2^o(t) = \frac{t}{T}, \quad \|(G_1, G_2)\| = \sqrt{\int_0^T G_1^2(t) dt} = \frac{1}{\sqrt{T}}, \quad T_{\min} = \frac{2}{\varepsilon^2}.$$

Najlepszy obserwator stanu $x(T)$ (o minimalnej normie) odpowiadający przypadkowi spodziewanych zakłóceń tylko na zmienną wyjściową y ma postać filtra uśredniającego

$$x(T) = \frac{1}{T} \int_0^T y(t) dt + \frac{1}{T} \int_0^T t u(t) dt.$$

Przykład 13.2

Dla układu II rzędu należy podać zależności na całkowy obserwator optymalny stanu końcowego $x(T)$, dla współczynnika wagi we wskaźniku jakości $\beta = 1$.

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t) \quad (13.28)$$

$$y(t) = [2 \ 0] x(t)$$

Macierze $\Phi_{11}(t)$ i $\Phi_{21}(t)$ obliczone na podstawie zależności (13.23) mają postać:

$$\Phi_{11}(t) = \begin{bmatrix} \cosh t \cos t & \frac{1}{2}(\sinh t \cos t + \cosh t \sin t) \\ \sinh t \cos t - \cosh t \sin t & \cosh t \cos t \end{bmatrix},$$

$$\Phi_{21}(t) = \begin{bmatrix} 2(\sinh t \cos t + \cosh t \sin t) & 2 \sinh t \sin t \\ -2 \sinh t \sin t & \sinh t \cos t - \cosh t \sin t \end{bmatrix}.$$

Macierze M i M^{-1} mają postać:

$$M = \begin{bmatrix} 2 \sinh T \cos T + 2 \cosh T \sin T & -2 \sinh T \sin T \\ 2 \sinh T \sin T & \sinh T \cos T - \cosh T \sin T \end{bmatrix},$$

$$M^{-1} = \frac{1}{2 \sinh^2 T - 2 \sin^2 T} \begin{bmatrix} \sinh T \cos T - \cosh T \sin T & 2 \sinh T \sin T \\ -2 \sinh T \sin T & 2(\sinh T \cos T + \cosh T \sin T) \end{bmatrix}$$

Ostatecznie, dla optymalnej pary obserwatorów $G_1^o(t)$ i $G_2^o(t)$, obliczonej na podstawie (13.24) i odpowiadającej przypadkowi spodziewanych zakłóceń na y i u , można podać poniższe wzory:

$$\begin{aligned} G_1^o(t) &= M^{-1} \begin{bmatrix} 2 \cos t \cosh t \\ \cos t \sinh t + \sin t \cosh t \end{bmatrix} \\ G_2^o(t) &= M^{-1} \begin{bmatrix} -2 \sin t \sinh t \\ \cos t \sinh t - \sin t \cosh t \end{bmatrix} \end{aligned} \quad (13.29)$$

Kształt tych przebiegów dla $T = 0,5$ i $T = 2$ przedstawiono na rysunku 13.1.

Optymalny obserwator dokładny dla powyższego układu II rzędu ma postać jak w (13.8)

$$\begin{bmatrix} x_1(T) \\ x_2(T) \end{bmatrix} = \int_0^T \begin{bmatrix} G_{11}^o(\tau) \\ G_{12}^o(\tau) \end{bmatrix} y(\tau) d\tau + \int_0^T \begin{bmatrix} G_{21}^o(\tau) \\ G_{22}^o(\tau) \end{bmatrix} u(\tau) d\tau.$$

Norma obserwatora $\|(G_1, G_2)\|(T)$ jest funkcją czasu obserwacji T

$$\|(G_1, G_2)\|(T) = \sqrt{\frac{3 \sinh 2T + \sin 2T}{4(\sinh^2 T - \sin^2 T)}}.$$

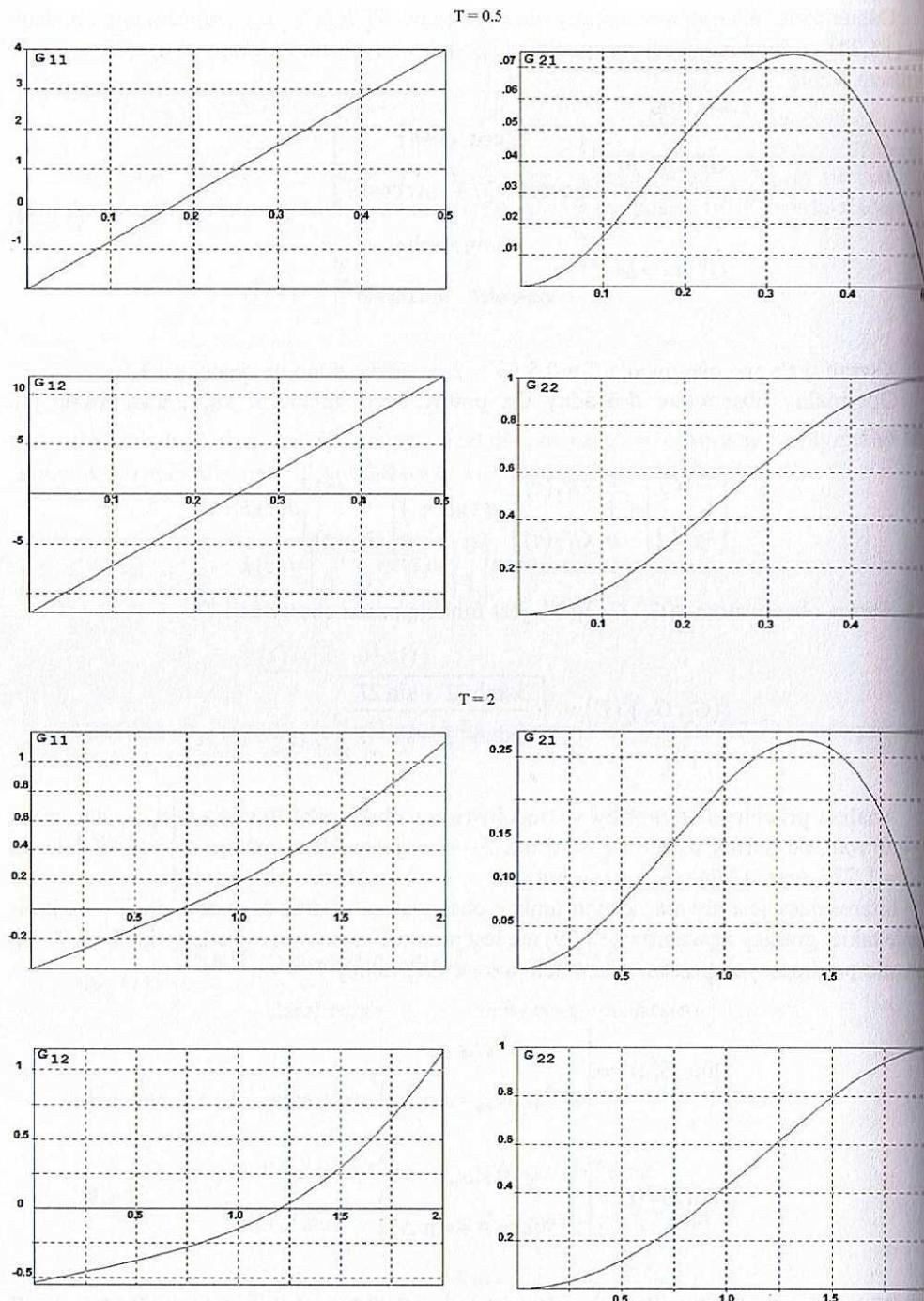
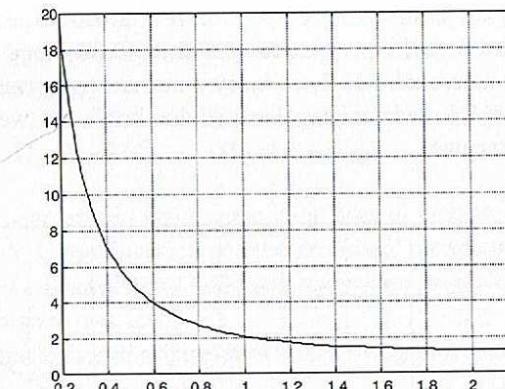
Analiza przebiegu tej normy w funkcji czasu obserwacji T wykazuje, że dla czasu $T \rightarrow 0$ wartość normy rośnie do ∞ , a dla $T \rightarrow \infty$, jej wartość maleje monotonicznie do $\sqrt{1.5} = 1.225$ (rys. 13.2).

Interesujący jest również kształt funkcji obserwatora wzdłuż osi czasu, dla $T = \infty$. Policzanie takiej granicy ze wzorów (13.29) nie jest proste. Zastosowanie podstawienia $t = T - \Delta$, pozwala po dłuższych przekształceniach przejść do granicy $T \rightarrow \infty$:

$$\lim_{T \rightarrow \infty} G_1^o(t) = \begin{bmatrix} e^{-\Delta} \cos \Delta \\ e^{-\Delta} (\cos \Delta - \sin \Delta) \end{bmatrix},$$

$$\lim_{T \rightarrow \infty} G_2^o(t) = \begin{bmatrix} e^{-\Delta} \sin \Delta \\ e^{-\Delta} (\cos \Delta + \sin \Delta) \end{bmatrix}.$$

Widac, że dla bardzo długiego okna obserwacji T i początkowych chwil obserwacji (tzn. $\Delta \rightarrow \infty$) funkcje G_1 i G_2 są prawie równe zero, a dla końcowych chwil obserwacji (tzn. $\Delta \rightarrow 0$), funkcje przyjmują wartości $G_1(T = \infty, \Delta = 0) = [1, 1]^T$, $G_2(T = \infty, \Delta = 0) = [0, 1]^T$.

Rys. 13.1. Kształt $G_{11}(t)$, $G_{12}(t)$ i $G_{21}(t)$, $G_{22}(t)$ dla okien $T_1 = 0,5$ i $T_2 = 2$ 

Rys. 13.2. Przebieg normy obserwatora dla układu II rzędu

Podobnie jak w przykładzie 13.1, przyjęcie wagi $\beta = 0$ daje prostsze wzory na M^{-1} i na optymalną parę obserwatorów $G_1^o(t)$ i $G_2^o(t)$:

$$\Phi_{11}(t) = \begin{bmatrix} 1, & t \\ 0, & t \end{bmatrix}, \quad \Phi_{21}(t) = \begin{bmatrix} 4t, & 2t^2 \\ -2t^2, & -2t^3/3 \end{bmatrix}$$

$$M^{-1} = \frac{1}{2T^3} \begin{bmatrix} -T^2, & 3T \\ -3T, & 6 \end{bmatrix} \quad (13.30)$$

$$G_1^o(t) = \frac{1}{T^3} \begin{bmatrix} 3T & -T^2 \\ 6t-3T & \end{bmatrix}, \quad G_2^o(t) = \frac{1}{T^3} \begin{bmatrix} T^2 & t^2-T & t^3 \\ 3T & t^2-2 & t^3 \end{bmatrix}$$

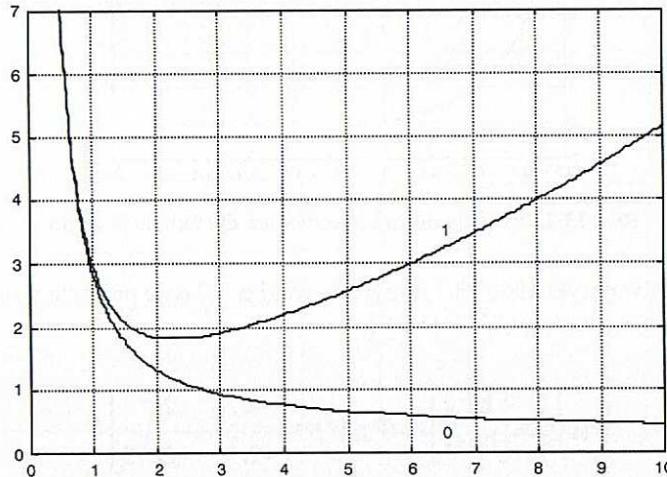
Największy błąd obserwacji z powodu najgorszych zakłóceń tylko na wyjściu, jest oszacowany przez normę obserwatora $\|(G_1, G_2)\|(T)$ (tylko z macierzy G_1 , bo $\beta = 0$) i jest on malejącą funkcją czasu obserwacji T

$$\|\varepsilon\| \leq \sqrt{2} \cdot \|(G_1, G_2)\| = \sqrt{2} \cdot \sqrt{\int_0^T [G_{11}(t)]^2 + [G_{12}(t)]^2 dt} = \sqrt{2} \cdot \sqrt{\frac{1}{T} + \frac{3}{T^3}} \quad (13.31)$$

Jeśli jednak dla tych macierzy obserwacji wystąpią dodatkowo niebezpieczne zakłócenia na sterowaniu $\hat{\beta} = 1$, to maksymalny błąd obserwacji należy oszacować według wzoru

$$\begin{aligned} \|\varepsilon\| &\leq \sqrt{2} \cdot \|(G_1, G_2)\| = \\ &= \sqrt{2} \cdot \sqrt{\int_0^T [G_{11}(t)]^2 + [G_{12}(t)]^2 + [G_{21}(t)]^2 + [G_{22}(t)]^2 dt} = \\ &= \sqrt{2} \cdot \sqrt{(T^6 + 39T^4 + 105T^2 + 315)/(105T^3)} \end{aligned} \quad (13.32)$$

Funkcja (13.32) przyjmuje minimum dla $T_m = 2,286$. Stosowanie dłuższego okna obserwacji, z nieoptymalnymi w takim przypadku wzorami (13.30), daje wzrost błędu obserwacji. Widać to wyraźnie na rysunku 13.3, na którym zaznaczono przebiegi według wzoru (13.31) przy założonej $\beta = 0$ i rzeczywistej $\hat{\beta} = 0$ (dolna krzywa) i według wzoru (13.32) przy $\beta = 0$, $\hat{\beta} = 1$ (górsza krzywa).



Rys. 13.3. Przebieg funkcji szacujących błąd obserwacji (13.31) i (13.32)
Objaśnienia w tekście

13.2. Ogólna teoria obserwacji w przestrzeniach Hilberta

Przedstawioną w poprzednich rozdziałach teorię obserwacji dokładnej stanu w przestrzeniach funkcyjnych $L^2[0, T]$ można uogólnić na inne przestrzenie, w których zdefiniowany jest iloczyn skalarny (liniowe przestrzenie unormowane, zupełne – zwane przestrzeniami Hilberta) [13BYW1]. Pozwoli to na sformułowanie relacji ogólnych umożliwiających rozwijanie zadań odtwarzania skończonego wektora stanu, opisywanych równaniami cząstkowymi lub z opóźnieniem oraz w wersji dyskretnej.

13.2.1. Warunki istnienia obserwatora

Rozważane będzie zadanie dokładnego odtworzenia nieznanego, skończonego wektora parametrów x na podstawie pomiarów wejścia u i wyjścia y na skończonym odcinku czasu $[0, T]$. Przyjmuje się ogólną klasę systemów liniowych, których wyjście daje się opisać operatorowo jako suma składowych pochodzących od nieznanego skończonego wymiarowego parametru x i od funkcji sterującej u (patrz np. wzór 13.7).

$$y = \mathbf{H}_1 x + \mathbf{H}_2 u \quad (13.33)$$

gdzie dwa liniowe, ciągłe odwzorowania \mathbf{H}_1 i \mathbf{H}_2 są określone w funkcyjnych przestrzeniach Hilberta U i Y (przestrzenie liniowe, unormowane, zupełne (Banacha) ze zdefiniowanym iloczynem skalarnym),

$$\mathbf{H}_1 : X \rightarrow Y, \quad \mathbf{H}_2 : U \rightarrow Y,$$

nieznany parametr x należy do skończonego wymiarowej przestrzeni $X = R^n$.

Przestrzenie X, Y, U są odpowiednio przestrzeniami parametrów, wyjścia i sterowania. Element $y \in Y$ dany zależnością (13.33) jest nazywany wyjściem związanym z parametrem $x \in X$ i sterowaniem $u \in U$.

Operator \mathbf{H}_1 ma więc strukturę wektora wierszowego, $H_1^T = [h_1^1 \dots h_1^n]$, z elementami $h_i^j \in Y$:

$$\mathbf{H}_1 x = H_1^T x = \sum_{i=1}^n h_1^i x_i,$$

$$\mathbf{H}_1 = \begin{bmatrix} h_1^1 \\ \vdots \\ h_1^n \end{bmatrix} \in Y^n, \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in R^n.$$

Można podać interpretację powyższych oznaczeń dla przestrzeni L^2 (wzór (13.7)). Odwzorowaniem elementów h_1^i są kolumny macierzy $C e^{-A(T-t)}$. Operator \mathbf{H}_2 to odpowiedni operator całkowy, a nieznany parametr x – to liczbowy wektor stanu końcowego $x(T)$.

Dla uproszczenia zapisu wprowadza się symbol $\langle . | . \rangle$ oznaczający macierzowy iloczyn skalarny w przestrzeni Hilberta, tzn. dający w wyniku strukturę macierzy lub wektora, oraz symbol $\langle . , . \rangle$ oznaczający standardowy iloczyn skalarny – dający w wyniku skalar.

Liniowy obserwator dla systemu (13.33) powinien, w oparciu o znajomość odcinków funkcji y i u , odtwarzać stan $x \in X = R^n$, dlatego też powinien on mieć postać dwóch liniowych ciągłych macierzowych funkcjonalów na przestrzeniach Hilberta Y i U .

Z twierdzenia Frécheta–Riesza o reprezentacji (aneks 2), które mówi, że każdy liniowy ciągły (lub liniowy ograniczony) funkcjonał w przestrzeni Hilberta może być wyrażony jako iloczyn skalarny, wynika dodatkowo, że obserwator dokładny musi mieć postać dwóch macierzowych iloczynów skalarnych.

$$x = \mathbf{G}_1 y + \mathbf{G}_2 u = \langle G_1 | y \rangle_Y + \langle G_2 | u \rangle_U \quad (13.34)$$

Obserwator będzie więc miał postać dwóch liniowych odwzorowań \mathbf{G}_1 i \mathbf{G}_2 z przestrzeni Hilberta Y i U na przestrzeń R^n :

$$\mathbf{G}_1 : Y \rightarrow X, \quad \mathbf{G}_2 : U \rightarrow X$$

takich, że (13.34) zachodzi dla każdego x, y, u spełniającego (13.33).

Ponieważ x jest wektorem kolumnowym, to dla jednowymiarowego (skalarnego) y operatory G_1 i G_2 też muszą mieć strukturę wektorów kolumnowych z elementami g_1^1, \dots, g_1^n . Dla wielowymiarowych y i u operatory mają strukturę macierzy, w której kolejne g_1^i oznaczają całe wiersze macierzy G_1 , a g_2^i – odpowiednio macierzy G_2 :

$$G_1 = \begin{bmatrix} g_1^1 \\ \vdots \\ g_1^n \end{bmatrix}, \quad G_2 = \begin{bmatrix} g_2^1 \\ \vdots \\ g_2^n \end{bmatrix}.$$

Wtedy:

$$\mathbf{G}_1 y = \langle G_1 | y \rangle_Y = \left\langle \begin{bmatrix} g_1^1 \\ \vdots \\ g_1^n \end{bmatrix} \mid y \right\rangle = \begin{bmatrix} \langle g_1^1, y \rangle \\ \vdots \\ \langle g_1^n, y \rangle \end{bmatrix}, \quad \text{gdzie } g_1^i \in Y, G_1 \in Y^n,$$

$$\mathbf{G}_2 u = \langle G_2 | u \rangle_U = \left\langle \begin{bmatrix} g_2^1 \\ \vdots \\ g_2^n \end{bmatrix} \mid u \right\rangle = \begin{bmatrix} \langle g_2^1, u \rangle \\ \vdots \\ \langle g_2^n, u \rangle \end{bmatrix}, \quad \text{gdzie } g_2^i \in U, G_2 \in U^n.$$

Wstawiając równanie (13.33) do (13.34) i wykorzystując operator sprzężony do \mathbf{H}_2 otrzymuje się

$$\begin{aligned} x &= \langle G_1 | H_1^T x \rangle_Y + \langle G_1 | \mathbf{H}_2 u \rangle_Y + \langle G_2 | u \rangle_U = \\ &= \langle G_1 | H_1^T x \rangle_Y + \left\langle \begin{bmatrix} \mathbf{H}_2^* g_1^1 \\ \vdots \\ \mathbf{H}_2^* g_1^n \end{bmatrix} \mid u \right\rangle_U + \langle G_2 | u \rangle_U. \end{aligned}$$

Z przedstawionego równania wynika, że lewa strona równa się prawej wtedy i tylko wtedy, gdy:

- zachodzi podstawowy warunek obserwonalności systemu (13.33) – jądro operatora \mathbf{H}_1 musi być zerowe, tzn.: $\ker \mathbf{H}_1 = 0$, czyli elementy h_1^i muszą być liniowo niezależne;

2) dla $\forall u \in U$ pomiędzy operatorami zachodzą relacje:

$$\langle G_1 | H_1^T \rangle = \mathbf{I}, \quad \text{gdzie } \mathbf{I} \text{ – macierz jednostkowa o wymiarach } n \times n, \text{ lub inaczej:} \\ \left\langle G_1 \mid H_1^T \right\rangle = \left\langle \begin{bmatrix} g_1^1 \\ \vdots \\ g_1^n \end{bmatrix} \mid \begin{bmatrix} h_1^1 & \dots & h_1^n \end{bmatrix} \right\rangle = \begin{bmatrix} \langle g_1^1, h_1^1 \rangle & \dots & \langle g_1^1, h_1^n \rangle \\ \vdots & \ddots & \vdots \\ \langle g_1^n, h_1^1 \rangle & \dots & \langle g_1^n, h_1^n \rangle \end{bmatrix} = \mathbf{I} \quad (13.35)$$

3)

$$\begin{bmatrix} g_2^1 \\ \vdots \\ g_2^n \end{bmatrix} = - \begin{bmatrix} \mathbf{H}_2^* g_1^1 \\ \vdots \\ \mathbf{H}_2^* g_1^n \end{bmatrix}, \quad \text{gdzie } \mathbf{H}_2^* \text{ jest operatorem sprzężonym do } \mathbf{H}_2$$

lub w zapisie po transpozycji

$$G_2^T = -\mathbf{H}_2^* G_1^T \quad (13.36)$$

Warunek 2) to warunek wprowadzający ograniczenie na dowolność doboru operatora G_1 , a warunek 3) wprowadza formułę na wyliczenie G_2 (po wyborze G_1).

Istnieje nieskończona liczba par obserwatorów (G_1 i G_2) dla systemu (13.33), które spełniają zależność obserwatora (13.34), ograniczenie (13.35) i zależność (13.36).

13.2.2. Optymalny obserwator skończonej liczby parametrów układu liniowego

Dany jest obserwowlany, liniowy, stacjonarny system opisany równaniem (13.33)

$$y = \mathbf{H}_1 x + \mathbf{H}_2 u$$

oraz obserwator nieznanego wektora parametrów $x \in R^n$ opisany równaniem (13.34)

$$x = \langle G_1 | y \rangle_Y + \langle G_2 | u \rangle_U.$$

Oznaczając przez S zbiór wszystkich obserwatorów dla tego systemu (par (G_1, G_2)) w przestrzeni $Y^n \times U^n$, można zdefiniować dla jej elementów normę (półnormę)

$$\| (G_1, G_2) \| = \sqrt{\sum_{i=1}^n \langle g_1^i, g_1^i \rangle_Y + \sum_{i=1}^n \beta_i \langle g_2^i, g_2^i \rangle_U} \quad (13.37)$$

$$\beta_i \geq 0, \quad i = 1, \dots, n.$$

Obserwator optymalny to obserwator z minimalną normą, czyli taka para $(G_1^o, G_2^o) \in S$, że

$$\min_{(G_1, G_2)} \| (G_1, G_2) \|^2 = \| (G_1^o, G_2^o) \|^2.$$

Norma obserwatora może być użyta jako wskaźnik jakości obserwacji.

Obserwatory minimalnonormowe znajdują uzasadnienie w zadaniach min-max, dając najlepsze efekty odtwarzania wektora parametrów w przypadku występowania w pomiarach sterowania i wyjścia addytywnych zakłóceń o ograniczonej normie i najbardziej niebezpiecznym kształcie, odpowiadającym kształtu funkcji filtrujących $G_1(t)$ i $G_2(t)$. Minimalizacja przyjętej normy obserwatora odpowiada, więc zadaniu minimalizacji błędu odtwarzania stanu dla zakłóceń najgorszych, przyjmowanych z klasy zakłóceń należących do kuli jednostkowej w odpowiedniej przestrzeni funkcyjnej.

Poniżej oszacujemy wpływ normy obserwatora na błąd obserwacji, np. dla wagi $\beta = 1$.

Niech z_1 i z_2 będą znormalizowanymi, ograniczonymi zakłóceniami pomiarowymi wejścia i wyjścia należącymi do kul jednostkowych w odpowiednich przestrzeniach $z_1 \in Y$, $z_2 \in U$ oraz $\|z_1\|_Y \leq 1$, $\|z_2\|_U \leq 1$.

Wtedy w chwili $t = T$:

- stan rzeczywisty: $x = \langle G_1 | y \rangle + \langle G_2 | u \rangle,$
- estymata stanu rzeczywistego: $\bar{x} = \langle G_1 | y + z_1 \rangle + \langle G_2 | u + z_2 \rangle,$
- błąd obserwacji: $\varepsilon = x - \bar{x}.$

Zadanie optymalizacji dla zakłóceń najbardziej niebezpiecznych z kuli można zapisać jako:

$$\begin{aligned} \min_{(G_1, G_2)} \max_{(z_1, z_2)} \|\varepsilon\|_{R^n}^2 &= \min_{(G_1, G_2)} \max_{(z_1, z_2)} \|\langle G_1 | z_1 \rangle + \langle G_2 | z_2 \rangle\|_{R^n}^2 \leq \\ &\leq 2 \min_{(G_1, G_2)} \left[\max_{\|z_1\|=1} \|\langle G_1 | z_1 \rangle\|_{R^n}^2 + \max_{\|z_2\|=1} \|\langle G_2 | z_2 \rangle\|_{R^n}^2 \right] \stackrel{\text{def}}{=} 2 \min_{(G_1, G_2)} \|(G_1, G_2)\|_{Y \times U}^2. \end{aligned}$$

Norma obserwatora określa maksymalny błąd odtwarzania stanu i może być traktowana jako wskaźnik jakości obserwacji

$$J_1 = \|(G_1, G_2)\|_{Y^n \times U^n}^2.$$

13.2.3. Minimalnonormowa postać obserwatora

Optymalizacja wskaźnika J_1 (13.37) zostanie przeprowadzona z wykorzystaniem techniki mnożników Lagrange'a. Funkcjonal Lagrange'a wykorzystujący ograniczenie (13.35) ma postać

$$L = J_1 + 2 \sum_{i=1}^n \left[e_i^T - \left(\langle g_1^i, h_1^i \rangle, \dots, \langle g_1^i, h_1^i \rangle \right) \right] \lambda_i \quad (13.38)$$

gdzie wektory $\lambda_i \in R^n$ są mnożnikami Lagrange'a, a e_i^T są wektorami bazowymi w R^n .

Wykorzystując zależność (13.36), otrzymuje się

$$\begin{aligned} J_1 &= \sum_{i=1}^n \langle g_1^i, g_1^i \rangle_Y + \sum_{i=1}^n \beta_i \langle H_2^* g_1^i, H_2^* g_1^i \rangle_U = \\ &= \sum_{i=1}^n \langle g_1^i, g_1^i \rangle_Y + \sum_{i=1}^n \beta_i \langle g_1^i, H_2 H_2^* g_1^i \rangle_U. \end{aligned}$$

Warunki konieczne optymalności dla każdego elementu $g_1^i : \frac{\partial L}{\partial g_1^i} = 0$ mają postać:

$$g_1^i + \beta_i H_2 H_2^* g_1^i - \sum_{j=1}^n h_1^j \lambda_i^j = 0 \quad \text{dla } i = 1, \dots, n,$$

gdzie λ_i^j jest j -tą składową wektora λ_i .

Przyjmując, że kolumny λ_i tworzą macierz Λ , a β_i macierz diagonalną β

$$\left[g_1^1, \dots, g_1^n \right] + H_2 H_2^* \left[g_1^1, \dots, g_1^n \right] \beta = H_1^T \Lambda,$$

otrzymuje się ogólną postać operatorową warunków optymalności obserwatora stanu

$$G_1^T + H_2 H_2^* G_1^T \beta = H_1^T \Lambda \quad (13.39)$$

Po uwzględnieniu zależności $G_2^T = -H_2^* G_1^T$ inna postać (13.39) to

$$G_1^T - H_2 G_2^T \beta = H_1^T \Lambda.$$

Poniżej przedstawione zostaną dwa szczególne przypadki wag.

1) Dla przypadku szczególnego wag $\beta = \mathbf{0}$ otrzymuje się z (13.39)

$$G_1^T = H_1^T \Lambda.$$

Podstawiając tę zależność do ograniczenia (13.35), $\langle H_1 | G_1^T \rangle = I$, otrzymuje się

$$\langle H_1 | H_1^T \rangle \Lambda = I \Rightarrow \Lambda = \langle H_1 | H_1^T \rangle^{-1} \Rightarrow G_1^T = H_1^T \langle H_1 | H_1^T \rangle^{-1}.$$

Jest to minimalnonormowa postać obserwatora G_1 dla równania liniowego (13.35).

2) Dla przypadku szczególnego macierzy wag $\beta = \beta \mathbf{I}, \beta \in R$ otrzymuje się z (13.39)

$$(I + \mathbf{H}_2 \mathbf{H}_2^* \beta) G_1^T = H_1^T \lambda \Rightarrow G_1^T = \mathbf{F}^{-1} H_1^T \lambda,$$

gdzie operator $\mathbf{F} = (I + \mathbf{H}_2 \mathbf{H}_2^* \beta)$, a operator odwrotny \mathbf{F}^{-1} jest dodatnio określony, symetryczny i działa na elementach H_1^T . Jego pierwiastek $\mathbf{F}^{-0.5}$ jest również dodatnio określony i symetryczny.

Uzgadniając ograniczenie (13.35), $\langle H_1 | G_1^T \rangle = I$, otrzymuje się

$$\langle H_1 | \mathbf{F}^{-1} H_1^T \rangle \lambda = I \Rightarrow \lambda = \langle H_1 | \mathbf{F}^{-1} H_1^T \rangle^{-1} = \langle \mathbf{F}^{-0.5} H_1 | \mathbf{F}^{-0.5} H_1^T \rangle^{-1}.$$

Macierz λ jest dodatnio określona, rzeczywistą macierzą symetryczną. Stąd ostatecznie obserwator optymalny ma zwartą postać minimalnonormową

$$\begin{aligned} G_1^{T \circ} &= \mathbf{F}^{-1} H_1^T \langle H_1 | \mathbf{F}^{-1} H_1^T \rangle^{-1} \\ G_2^{T \circ} &= -\mathbf{H}_2^* G_1^{T \circ} \end{aligned} \quad (13.40)$$

W podrozdziale 13.1 przedstawiono już zastosowanie formuł (13.39) dla optymalnej postaci obserwatora dokładnego dla obiektu opisywanego równaniem różniczkowym zwyczajnym w przestrzeni $L^2[0, T]$. W następnych podrozdziałach przedstawimy przykłady zastosowania wyprowadzonych tu ogólnych formuł obserwatora dla układu o parametrach rozłożonych (przestrzeń $L^2[0, T], (0, 1)$) i dla układu dyskretnego (przestrzeń Hilberta R^N).

13.3. Przykład obserwacji całkowej w układzie o parametrach rozłożonych

13.3.1. Problemy obserwacji w układach o parametrach rozłożonych

Dla właściwego modelowania pewnych fizycznych zjawisk nieodzowne jest zastosowanie modeli o parametrach rozłożonych, opisywanych przez równania różniczkowe cząstkowe. Obserwonalność takich systemów oznacza, że stan początkowy (lub końcowy) może być jednoznacznie określony z idealnych pomiarów wejścia i wyjścia. Ogólnie dla układów o parametrach rozłożonych przestrzeń stanów początkowych jest nieskończoność wymiarowa. Stąd powstają pewne pytania odnośnie do warunków obserwonalności, np. jak wiele czujników powinno być użyte do zagwarantowania jednoznaczności obserwowanego stanu, gdzie te czujniki powinny być ulokowane i czy stan początkowy zależy w sposób ciągły od danych pomiarowych. Ostatnie pytanie jest ważne w przypadkach, gdy dane pomiarowe nie są idealne i małe błędy pomiarowe mogą generować duży błąd w rekonstrukcji stanu początkowego. Wypełnienie warunków obserwonalności dla systemu PDE nie zapewnia ciągłej zależności odtwarzanego stanu od pomiarów wyjścia, stąd w pewnych

przypadkach problem jest źle postawiony. Te zagadnienia były badane w wielu pracach [13ROS; 13DOS1; DOS2; 13GOR; 13JAA; 13KUA; 13KOT; 3SAY]. Standardowo w tych badaniach były rozważane systemy bez sterowania. Autorzy wychodzili z założenia, że sterowanie z reguły jest generowane przez operatora i jest znane. Stąd można jego efekt na wyjściu policzyć i odjąć od obserwowanego wyjścia, odtwarzając stan początkowy z tak otrzymanej różnicy. Takie podejście w zastosowaniach do przemysłowych systemów sterowania nie zawsze jest właściwe, z powodu zakłóceń zmiennej sterującej, a więc sygnał sterowania wchodzący na obiekt powinien być mierzony. W tym podrozdziale wykorzystamy optymalne formuły na obserwator dokładny dla stanu skończenie wymiarowego wyprowadzone w podrozdziale 13.2.

Dla ominienia omawianych trudnych problemów ciągłości dla modeli o parametrach rozłożonych rekonstrukcja stanu zostanie przetransformowana do rekonstrukcji skończenie wymiarowej nieznanego parametru $x \in R^n$ (gdzie x_i oznacza nieznane amplitudy dla n sinusoid, których suma tworzy stan początkowy) [13BYW2].

13.3.2. Model procesu cieploprzewodnictwa

Jednowymiarowe równanie cieploprzewodnictwa (zwane niekiedy równaniem dyfuzji) opisuje przepływ ciepła przez jednorodny, odizolowany, cienki pręt o długości L . $T(t, z)$ oznacza temperaturę w chwili t i w punkcie z .

$$k^2 \cdot \frac{\partial^2 T(t, z)}{\partial z^2} = \frac{\partial T(t, z)}{\partial t} \quad (13.41)$$

Stała $k^2 = K/c\rho$ jest nazywana stałą dyfuzji, gdzie K – przewodność cieplna, c – ciepło właściwe, ρ – masa jednostkowa, a z oznacza zmienną przestrzenną.

Dla jednoznaczności rozwiązania dane są:

- warunek początkowy: $T(0, z) = \varphi(z)$ dla $0 < z < L$,
- brzegowe warunki Dirichleta: $T(t, 0) = \Psi_1(t), T(t, L) = \Psi_2(t)$.

Obserwacje będą generowane przez dwa typy czujników ulokowanych w punkcie α , $0 < \alpha < L$, dających:

- 1) pomiar idealny punktowy,
- 2) pomiar rzeczywisty przestrzennie rozłożony (uśredniony).

Daje to różne równania dla wyjścia systemu:

- 1) $y(t, \alpha) = T(t, \alpha), \quad 0 < \alpha < L$,
- 2) $y(t, \alpha) = \frac{1}{2\epsilon} \int_{\alpha-\epsilon}^{\alpha+\epsilon} T(t, \xi) d\xi$, gdzie 2ϵ jest średnicą termopary ulokowanej w punkcie α .

Wynika stąd różna forma operatorów $\mathbf{H}_1, \mathbf{H}_2$ w ogólnym równaniu wyjścia (13.33).

Przyjmuje się, że wektor $x \in R^n$ ma skończony wymiar i reprezentuje nieznany rozkład początkowy temperatury $T(0, z) = \phi(z)$.

$$T(0, z) = \phi(z) = \sum_{j=1}^n x_j \cdot \sin\left(\frac{j \cdot \pi}{L} z\right) \quad (13.42)$$

Dla warunków brzegowych Dirichleta: $u(t, 0) = \psi_1(t)$, $u(t, L) = \psi_2(t)$, rozwiązanie jest następujące [13KON]

$$T(t, z) = \sum_{i=1}^{\infty} T_i(t) \cdot \sin\left(\frac{i \cdot \pi}{L} z\right) \quad (13.43)$$

gdzie

$$T_i(t) = e^{-\left(\frac{i \pi k}{L}\right)^2 t} \left[C_i + \frac{2i \pi k^2}{L^2} \int_0^t e^{\left(\frac{i \pi k}{L}\right)^2 \tau} [\psi_1(\tau) - (-1)^i \psi_2(\tau)] d\tau \right] \quad (13.44)$$

i C_i reprezentuje wpływ warunków początkowych

$$C_i = T_i(0) = \frac{2}{L} \int_0^L \phi(\xi) \sin\left(\frac{i \pi \xi}{L}\right) d\xi \quad (13.45)$$

Końcowa formuła będzie wyprowadzona dla warunku brzegowego z lewej strony preta $u(t) = T(t, 0) = \Psi_1(t)$, podczas gdy izolowana strona prawa $T(t, L) = \Psi_2(t) = 0$, dla $t > 0$.

Wtedy rozwiązanie wyraża się wzorem

$$T(t, z) = \sum_{i=1}^{\infty} e^{-\left(\frac{i \pi k}{L}\right)^2 t} \cdot T_i(0) \cdot \sin\left(\frac{i \cdot \pi}{L} z\right) + \sum_{i=1}^{\infty} \frac{2i \pi k^2}{L^2} \int_0^t e^{-\left(\frac{i \pi k}{L}\right)^2 (t-\tau)} \psi_1(\tau) d\tau \cdot \sin\left(\frac{i \cdot \pi}{L} z\right)$$

Równanie to można napisać w postaci

$$T(t, z) = T_0(t, z) + T_u(t, z) \quad (13.46)$$

gdzie T_0 zależy od warunków początkowych, a T_u od brzegowego sterowania.

Podstawiając warunki początkowe (13.42) do (13.46) i uwzględniając, że

$$\int_0^L \sin\left(j \pi \frac{\xi}{L}\right) \sin\left(i \pi \frac{\xi}{L}\right) d\xi = 0 \quad \text{dla } \forall i \neq j \quad \text{i} \quad \int_0^L \sin\left(j \pi \frac{\xi}{L}\right) \sin\left(j \pi \frac{\xi}{L}\right) d\xi = \frac{L}{2},$$

w wyniku otrzymuje się sumę tylko skończonej liczby elementów w $T_0(t, z)$

$$T_0(t, z) = \sum_{j=1}^n x_j \cdot e^{-\left(\frac{j \pi k}{L}\right)^2 t} \cdot \sin\left(\frac{j \cdot \pi}{L} z\right) \quad (13.47)$$

nieskończoną sumę elementów w $T_u(t, z)$

$$T_u(t, z) = \sum_{i=1}^{\infty} \left[\frac{2i \pi k^2}{L^2} \int_0^t e^{-\left(\frac{i \pi k}{L}\right)^2 (t-\tau)} \psi_1(\tau) d\tau \cdot \sin\left(\frac{i \cdot \pi}{L} z\right) \right] \quad (13.48)$$

Uwagi do metody obliczeniowej całki we wzorze (13.48)

Wzór (13.48) może służyć do symulacji zachowania się modelu o parametrach rozłożonych. Jednak dla dużych czasów przy liczeniu tej całki z użyciem różnych numerycznych metod całkowania otrzymywano niewystarczającą dokładność obliczeń.

Można zauważyć, że współczynniki $T_{ui}(t)$ w (13.48)

$$T_u(t, z) = \sum_{i=1}^{\infty} T_{ui}(t) \cdot \sin\left(\frac{i \cdot \pi}{L} z\right)$$

mają postać

$$T_{ui}(t) = \int_0^t e^{-\left(\frac{i \pi k}{L}\right)^2 (t-\tau)} \frac{2i \pi k^2}{L^2} \psi_1(\tau) d\tau$$

i można je potraktować jako rozwiązanie poniższego równania różniczkowego (aproksymującego (13.41)) z zerowymi warunkami początkowymi

$$\frac{dT_{ui}(t)}{dt} + \left(\frac{i \pi k}{L}\right)^2 T_{ui}(t) = \frac{2i \pi k^2}{L^2} [\psi_1(t)]$$

lub w zapisie macierzowym

$$\begin{aligned} \dot{\tilde{T}}_u(t) &= A \tilde{T}_u(t) + Bu(t) \\ T_u(t, z) &= C \tilde{T}_u(t) \end{aligned} \quad (13.49)$$

gdzie $\tilde{T}_u(t) = [T_{u1}(t), T_{u2}(t), \dots, T_{uN}(t)]^T$, a A jest macierzą diagonalną

$$A = \text{diag} \left[-\frac{(\pi k)^2}{L^2}, -\frac{(2\pi k)^2}{L^2}, \dots, -\frac{(N\pi k)^2}{L^2} \right]_{N \times N},$$

$$B = \left[\frac{2\pi k^2}{L^2}, \frac{4\pi k^2}{L^2}, \dots, \frac{2N\pi k^2}{L^2} \right]_{N \times 1}^T,$$

$$C = \left[\sin\left(\pi \frac{z}{L}\right), \sin\left(2\pi \frac{z}{L}\right), \dots, \sin\left(N\pi \frac{z}{L}\right) \right]_{1 \times N}.$$

Liczba N oznacza skończoną liczbę wyrazów szeregu (13.48).

Do rozwiązania (13.49) można użyć procedury **lsim** w pakiecie Matlab.

13.3.3. Obserwacja punktowa i uśredniona wyjścia

Problem obserwacji punktowej idealnej

Idealny punktowy czujnik ulokowany w punkcie α , $0 < \alpha < L$, pozwala ze wzorów (13.33), (13.42), (13.47) i (13.48) otrzymać operatory \mathbf{H}_1 i \mathbf{H}_2 , gdzie przestrzeń sterowa $U = L^2[0, T]$, a przestrzeń wyjść $Y = L^2([0, T] \times [0, L])$

$$y_1(t, \alpha) = T_0(t, \alpha) = \mathbf{H}_1 x = \sum_{j=1}^N h_1^j x_j = \begin{bmatrix} h_1^1 & h_1^2 & \dots & h_1^N \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix},$$

$$\text{gdzie: } h_1^j = e^{-\left(\frac{j\pi k}{L}\right)^2 t} \sin\left(\frac{j\pi}{L} \alpha\right),$$

$$y_2(t, \alpha) = T_u(t, \alpha) = \mathbf{H}_2 u = \int_0^t \left[\frac{2\pi k^2}{L^2} \sum_{i=1}^{\infty} i e^{-\left(\frac{i\pi k}{L}\right)^2 (t-\tau)} \sin\left(\frac{i\pi}{L} \alpha\right) \right] \psi_1(\tau) d\tau \quad (13.50)$$

Z ostatnich wzorów widać, że warunek obserwowalności $\ker \mathbf{H}_1 = 0$, nie jest spełniony dla niektórych α , $0 < \alpha < L$.

Problem obserwacji uśrednionej

Dla obserwacji przestrzennie uśrednionej wyjście z czujnika o średnicy 2ϵ ulokowanego w punkcie α jest dane wzorem

$$y(t, \alpha) = \frac{1}{2\epsilon} \int_{\alpha-\epsilon}^{\alpha+\epsilon} T(t, \xi) d\xi = \frac{1}{2\epsilon} \int_{\alpha-\epsilon}^{\alpha+\epsilon} T_0(t, \xi) d\xi + \frac{1}{2\epsilon} \int_{\alpha-\epsilon}^{\alpha+\epsilon} T_u(t, \xi) d\xi \quad (13.51)$$

Uwzględniając (13.47), rozwinięcie pierwszej całki daje

$$\begin{aligned} y_1(t, \alpha) &= \frac{1}{2\epsilon} \int_{\alpha-\epsilon}^{\alpha+\epsilon} T_0(t, \xi) d\xi = \frac{1}{2\epsilon} \int_{\alpha-\epsilon}^{\alpha+\epsilon} \sum_{j=1}^n x_j \cdot e^{-\left(\frac{j\pi k}{L}\right)^2 t} \sin\left(\frac{j\pi}{L} \xi\right) d\xi = \\ &= \frac{1}{2\epsilon} \sum_{j=1}^n x_j \cdot e^{-\left(\frac{j\pi k}{L}\right)^2 t} \int_{\alpha-\epsilon}^{\alpha+\epsilon} \sin\left(\frac{j\pi}{L} \xi\right) d\xi \end{aligned}$$

Ostatecznie, na $y_1(t, \alpha)$ otrzymuje się wzór

$$y_1(t, \alpha) = \frac{L}{\pi\epsilon} \sum_{j=1}^n \frac{1}{j} x_j \sin\left(\frac{j\pi}{L} \alpha\right) \sin\left(\frac{j\pi}{L} \epsilon\right) \cdot e^{-\left(\frac{n\pi k}{L}\right)^2 t} \quad (13.52)$$

Formułę tę można zapisać jako iloczyn skalarny wektora funkcyjnego \mathbf{H}_1 i wektora parametrów $x = [x_1, x_2, \dots, x_n]^T$

$$y_1(t, \alpha) = \sum_{j=1}^n h_1^j x_j = \begin{bmatrix} h_1^1 & \dots & h_1^n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{H}_1 x,$$

$$\text{gdzie } h_1^j = \frac{L}{j\pi\epsilon} e^{-\left(\frac{j\pi k}{L}\right)^2 t} \sin\left(\frac{j\pi}{L} \alpha\right) \sin\left(\frac{j\pi}{L} \epsilon\right).$$

Również w tym przypadku warunek obserwowalności $\ker \mathbf{H}_1 = 0$ nie jest spełniony dla pewnych α .

Druga całka w równaniu (13.51) jest dana jako

$$\begin{aligned} y_2(t, \alpha) &= \frac{1}{2\epsilon} \int_{\alpha-\epsilon}^{\alpha+\epsilon} T_u(t, \xi) d\xi = \\ &= \frac{1}{2\epsilon} \int_{\alpha-\epsilon}^{\alpha+\epsilon} \sum_{i=1}^{\infty} \frac{2i\pi k^2}{L^2} \int_0^t e^{-\left(\frac{i\pi k}{L}\right)^2 (t-\tau)} \psi_1(\tau) d\tau \cdot \sin\left(\frac{i\pi}{L} \xi\right) d\xi, \end{aligned}$$

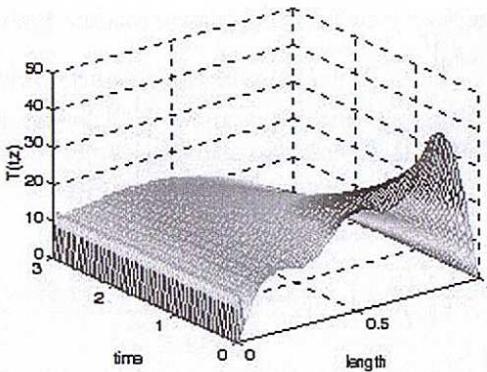
Po transformacji można otrzymać również końcową postać operatora \mathbf{H}_2

$$\begin{aligned} y_2(t, \alpha) &= \frac{\pi k^2}{\epsilon L^2} \sum_{i=1}^{\infty} i \int_0^t e^{-\left(\frac{i\pi k}{L}\right)^2 (t-\tau)} \psi_1(\tau) d\tau \int_{\alpha-\epsilon}^{\alpha+\epsilon} \sin\left(\frac{i\pi}{L} \xi\right) d\xi = \\ &= \frac{2k^2}{\epsilon L} \int_0^t \left[\sum_{i=1}^{\infty} e^{-\left(\frac{i\pi k}{L}\right)^2 (t-\tau)} \sin\left(\frac{i\pi}{L} \alpha\right) \sin\left(\frac{i\pi}{L} \epsilon\right) \right] \psi_1(\tau) d\tau = \mathbf{H}_2 u \end{aligned} \quad (13.53)$$

Wzory (13.50), (13.52) i (13.53) są użyte do symulacji dwóch typów wyjścia. Na podstawie pomiaru temperatury w punkcie α obserwator dokładny z przesuwany oknem czasowym (*Moving Window Observer*) będzie użyty dla rekonstrukcji temperatury w pewnym punkcie β niedostępnym do pomiaru.

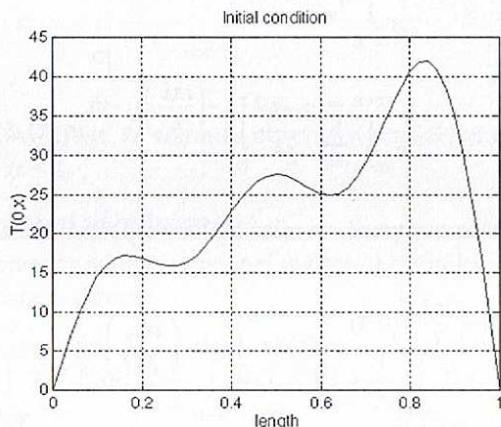
13.3.4. Obserwator dokładny. Przykład numeryczny

Na rysunku 13.4 przedstawiono wyniki symulacji temperatury dla układu z parametrami $k = 0,2$, $L = 1,0$, $n = 5$, $N = 50$, $\psi_1(t) = 10 \cdot 1(t)$, krok numeryczny dyskretyzacji $\Delta t = 0,001$, okno czasowe $T = 0,1$ (100 próbek), $\Delta z = 0,01$, promień czujnika $\epsilon = 3\Delta z$



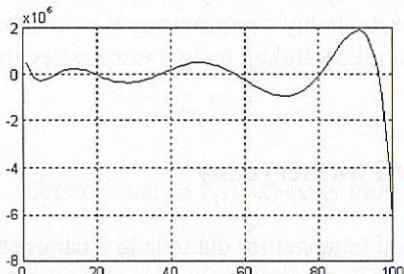
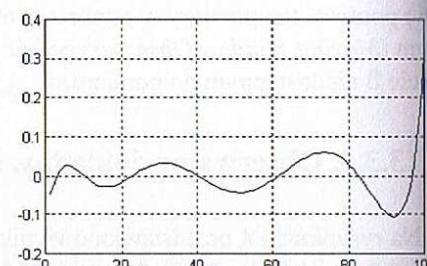
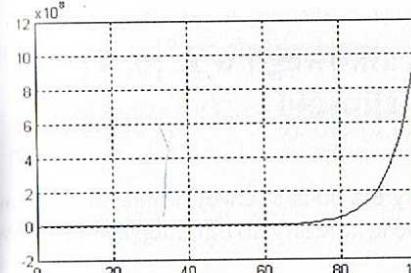
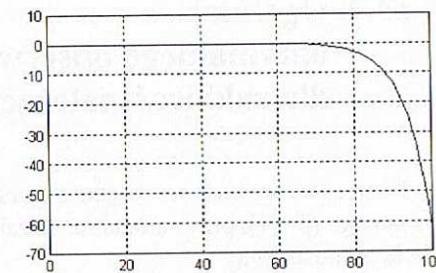
Rys. 13.4. Przebieg temperatury

Wygenerowany przebieg warunku początkowego złożony z ośmiu sinusoid jest widoczny na rysunku 13.5.

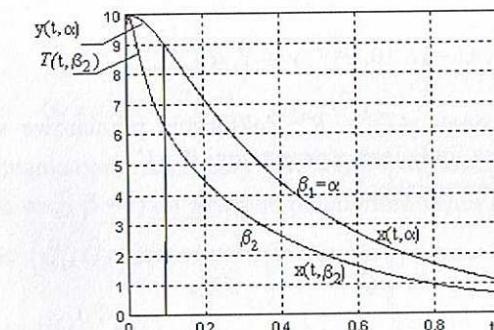


Rys. 13.5. Warunek początkowy

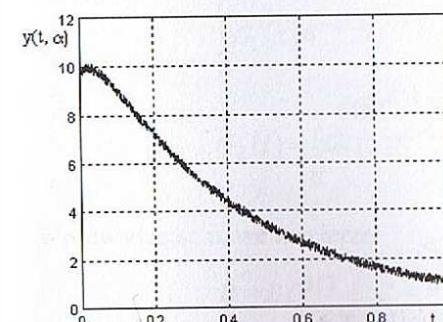
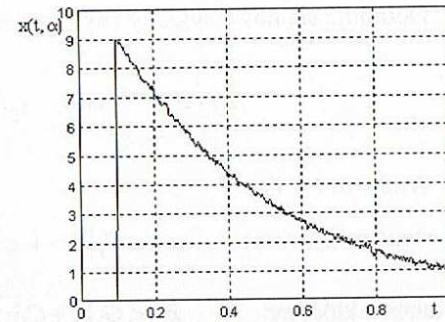
Na rysunkach 13.6–13.9 przedstawiono kształty przykładowych elementów macierzy G_1 i G_2 .

Rys. 13.6. Kształt $G_1(2)$ Rys. 13.7. Kształt $G_1(8)$ Rys. 13.8. Kształt $G_2(2)$ Rys. 13.9. Kształt $G_2(8)$

Na rysunku 13.10 widoczne są efekty rekonstrukcji stanu $x(t, \beta)$ w punkcie $\beta_1 = 0,2$, $\beta_2 = 0,4891542$, podczas gdy czujnik ulokowany jest w punkcie $\alpha = 0,4891542$ (tu β nie oznacza wag wskaźnika jakości). Widoczne jest dokładne odtworzenie stanu po czasie $t = 0,1$, zarówno dla punktu pomiaru α , jak i dla punktu β_2 – przez ruchome okno czasowe $T = 0,1$.

Rys. 13.10. Rzeczywiste i odtworzone temperatury dla czasu $T \leq t \leq 1$ w dwóch punktach pręta $\beta_1 = 0,2$ i $\beta_2 = 0,4891542$, $\alpha = 0,4891542$

Na rysunkach 13.11 i 13.12 przedstawiono efekty filtracji i odtwarzania stanu przez obserwator z ruchomym oknem czasowym $T = 0,1$ przy zaszumionych pomiarach. Wydłużenie okna polepsza efekt filtracji.

Rys. 13.11. Zaszumiony pomiar w punkcie $\alpha = 0,4891542$ Rys. 13.12. Odtworzony i odfiltrowany stan w punkcie α dla $t > T = 0,1$

13.4. Ogólniejsza wersja optymalnego obserwatora całkowego w $L^2[0, T]$ dla zakłóceń należących do elipsoid

Poniżej wyprowadzone zostaną szersze formuły dla podstawowej wersji obserwatora optymalnego ($\beta = 1$), przy założeniu, że zakłócenia będą należały do ogólniejszych zbiorów niż kula jednostkowa.

Dany jest obserwowlany system:

$$\begin{aligned}\dot{x}(t) &= A x(t) + B u(t) + B w(t) \\ y(t) &= C x(t) + C_1 v(t)\end{aligned}\tag{13.54}$$

przestrzenie: $Y = (L^2[0, T])^m$, $U = (L^2[0, T])^r$, $y \in Y$, $u \in U$.

Celem jest określenie stanu $x(T) \in R^n$. Zakłócenia pomiarowe wejścia i wyjścia $w(t) \in R^r$, $v(t) \in R^m$, $\forall t \geq 0$, są funkcjami z przestrzeni W i V .

Zbiory te można zdefiniować jako:

$$W = \left\{ w(\cdot) \in U : \|Q^{-1/2} w\|_U \leq 1 \right\},$$

$$V = \left\{ v(\cdot) \in Y : \|R^{-1/2} v\|_Y \leq 1 \right\},$$

a Q i R są symetrycznymi, dodatnio określonymi macierzami $[r \times r]$ i $[m \times m]$ wymiarowymi. C_1 jest nieosobliwe.

Definiuje się nowe sygnały zakłócające:

$$z_1(t) = R^{-1/2} v(t), \quad z_2(t) = Q^{-1/2} w(t).$$

W chwili $t = T$:

- stan rzeczywisty: $x = \langle G_1 | y \rangle + \langle G_2 | u \rangle,$
- stan zakłócony: $\bar{x} = \langle G_1 | y + C_1 v \rangle + \langle G_2 | u + w \rangle,$
- błąd obserwacji: $\varepsilon = \bar{x} - x.$

Błąd obserwacji można oszacować, wprowadzając zmodyfikowane macierze obserwacji i definiując normę obserwatora.

$$\begin{aligned}\min_{(G_1, G_2)} \max_{(w, v)} \|\varepsilon\|_{R^n}^2 &= \min_{(G_1, G_2)} \max_{(w, v)} \left\| \langle G_1 | C_1 v \rangle + \langle G_2 | w \rangle \right\|_{R^n}^2 \leq \\ &\leq \min_{(G_1, G_2)} 2 \left[\max_{v \in V} \left\| \langle G_1 | C_1 v \rangle \right\|_{R^n}^2 + \max_{w \in W} \left\| \langle G_2 | w \rangle \right\|_{R^n}^2 \right] = \\ &= \min_{(G_1, G_2)} 2 \left[\max_{\|z_1\|=1} \left\| \langle G_1 C_1 R^{1/2} | z_1 \rangle \right\|_{R^n}^2 + \max_{\|z_2\|=1} \left\| \langle G_2 Q^{1/2} | z_2 \rangle \right\|_{R^n}^2 \right] = \\ &= \min_{(\hat{G}_1, \hat{G}_2)} 2 \left[\max_{\|z_1\|=1} \left\| \langle \hat{G}_1 | z_1 \rangle \right\|_{R^n}^2 + \max_{\|z_2\|=1} \left\| \langle \hat{G}_2 | z_2 \rangle \right\|_{R^n}^2 \right] \stackrel{df}{=} \min_{(\hat{G}_1, \hat{G}_2)} 2 \left\| (\hat{G}_1, \hat{G}_2) \right\|_{Y \times U}^2.\end{aligned}$$

Nowe macierze obserwacji:

$$\begin{aligned}\hat{G}_1 &= G_1 C_1 R^{1/2}, \quad \hat{G}_2 = G_2 Q^{1/2} \\ G_1 &= \hat{G}_1 R^{-1/2} C_1^{-1}, \quad G_2 = \hat{G}_2 Q^{-1/2}\end{aligned}\tag{13.55}$$

W celu otrzymania optymalnej pary $(\hat{G}_1^{op}, \hat{G}_2^{op})$ wykorzystane zostaną wyniki z podrozdziału 13.1 (dla wagi $\beta = 1$) po pewnym przeddefiniowaniu problemu.

Zauważmy, że (\hat{G}_1, \hat{G}_2) jest obserwatorem dla systemu:

$$\begin{aligned}\dot{\hat{x}}(t) &= A \hat{x}(t) + B Q^{1/2} u(t) + \hat{B} z_2(t) = A \hat{x}(t) + \hat{B} u(t) + \hat{B} z_2(t) \\ \hat{y}(t) &= R^{-1/2} C_1^{-1} C \hat{x}(t) + z_1(t) = \hat{C} \hat{x}(t) + z_1(t)\end{aligned}\tag{13.56}$$

gdzie zakłócenia $z(\cdot)$ są już brane z kul jednostkowych.

Ze wzorów (13.9), (13.10) i (13.55) mamy:

$$\int_0^T \hat{G}_1(\tau) R^{-1/2} C_1^{-1} C e^{-A(T-\tau)} d\tau = I,$$

$$\hat{G}_2(t) = \int_0^t \hat{G}_1(\tau) R^{-1/2} C_1^{-1} C e^{-A(t-\tau)} B Q^{1/2} d\tau.$$

Wprowadzając nowe macierze:

$$\hat{B} = B Q^{1/2} \text{ i } \hat{C} = R^{-1/2} C_1^{-1} C,$$

można bezpośrednio wykorzystać wyniki z podrozdziału 13.1 dla optymalnego obserwatora systemu (13.56).

Z zależności (13.23) wynikają wzory na optymalne macierze obserwacji:

$$\begin{aligned}\hat{G}_1^{op}(t) &= \hat{M}^{-1} \hat{\Phi}'_{11}(t) \hat{C}' = \hat{P}_1(t) \hat{C}' \\ \hat{G}_2^{op}(t) &= \hat{M}^{-1} \hat{\Phi}'_{21}(t) \hat{B}' = \hat{P}_2(t) \hat{B}'\end{aligned}\quad (13.57)$$

gdzie rzeczywista macierz \hat{M} i macierz fundamentalna $\hat{\Phi}$ dane są wzorami:

$$\hat{M} = \int_0^T \hat{\Phi}'_{11}(\tau) \hat{C}' \hat{C} e^{-A(T-\tau)} d\tau, \quad \hat{\Phi}(t) = e^{\hat{W}t} = \begin{bmatrix} \hat{\Phi}_{11} & \hat{\Phi}_{12} \\ \hat{\Phi}_{21} & \hat{\Phi}_{22} \end{bmatrix} \quad (13.58)$$

$$\text{macierz } \hat{W} = \begin{bmatrix} A & BQB' \\ C'C_1^{-1}R^{-1}C_1^{-1}C & -A' \end{bmatrix}.$$

Macierze $G_1^{op}(t), G_2^{op}(t)$ dla systemu (13.54) otrzymuje się ze wzorów (13.55) i (13.57)

$$\begin{aligned}G_1^{op}(t) &= \hat{G}_1 R^{-1/2} C_1^{-1} = \\ &= e^{AT} \left[\int_0^T \hat{\Phi}'_{11}(\tau) C'C_1^{-1} R^{-1} C_1^{-1} C e^{A\tau} d\tau \right]^{-1} \hat{\Phi}'_{11}(t) C C_1^{-1} R^{-1} C_1^{-1} \\ G_2^{op}(t) &= \hat{G}_2 Q^{-1/2} = \\ &= e^{AT} \left[\int_0^T \hat{\Phi}'_{11}(\tau) C'C_1^{-1} R^{-1} C_1^{-1} C e^{A\tau} d\tau \right]^{-1} \hat{\Phi}'_{21}(t) B\end{aligned}\quad (13.59)$$

Przez analogię nietrudno podać również formuły dla niejednostkowej macierzy wagowej β .

14. Obserwatory całkowe w układzie on-line

14.1. Obserwatory całkowe z przesuwany i rozszerzanym oknem obserwacji

Przedstawiony w poprzednich rozdziałach całkowy opis obserwatorów na skończonym oknie obserwacji według formuły

$$x(T) = \int_0^T G_1(T, \tau) y(\tau) d\tau + \int_0^T G_2(T, \tau) u(\tau) d\tau \quad (14.1)$$

może być wykorzystany w systemach obserwacji i sterowania on-line. W tym celu można zaproponować strukturę MWO (*Moving Window Observer*) przesuwanego okna o stałej szerokości T . Formuła (14.1) jest prawdziwa w układach stacjonarnych dla dowolnego odcinka zarejestrowanych funkcji y i u , a więc również dla funkcji przesuniętych.

Dla ustalonego T , policzonych $G_1(T, \tau), G_2(T, \tau)$ i $t \geq T$ obowiązuje więc formuła MWO przedstawiana różnymi możliwymi postaciami całek:

$$x(t) = \int_0^T G_1(\tau) y(t-T+\tau) d\tau + \int_0^T G_2(\tau) u(t-T+\tau) d\tau \quad (14.2)$$

$$x(t) = \int_{-T}^0 G_1(T+\tau) y(t+\tau) d\tau + \int_{-T}^0 G_2(T+\tau) u(t+\tau) d\tau,$$

$$x(t) = \int_{t-T}^t G_1(T-t+\tau) y(\tau) d\tau + \int_{t-T}^t G_2(T-t+\tau) u(\tau) d\tau \quad (14.3)$$

Z ostatniej postaci wynika, że taki obserwator reprezentuje okno o stałej szerokości T związane z macierzami obserwacji G , przesuwane na tle pomiarów wzdłuż osi czasu i po obliczeniu całek dające stan $x(t)$ dla $t \geq T$. W praktycznych zastosowaniach wygodną postacią jest (14.2), z której widać, że macierze G_1, G_2 nie zależą od bieżącego czasu t i muszą być policzone w przedziale $[0, T]$ tylko raz (off-line), a następnie zapamiętane wystarczą-

co gęsto, aby zagwarantować dokładność numeryczną liczenia całek. Użyty sprzęt komputerowy musi być wystarczająco szybki, aby w jednym przedziale dyskretyzacji zdążyły się policzyć w przedziale T iloczyny funkcji G_1y i G_2u oraz ostateczne wartości dwóch całek (np. procedurą Simpsona). Są to jednak wymagania techniczno-numeryczne niezwiązane z koniecznością dyskretyzacji algorytmu modelu i obserwatora.

Teoretycznie, przy pomijalnie małych zakłócenach, długość okna T może być wybierana dowolnie, gdyż nie wpływa ona na wynik obserwacji. Ponieważ wybrana wielkość przedziału obserwacji T implikuje ostatecznie liczbę obserwacji pomiarowych koniecznych do zapamiętania w komputerze oraz odpowiadającą im liczbę punktów, w których zapamiętane będą przebiegi G_1 i G_2 , to z punktu widzenia szybkości procedur numerycznych zalecana jest mała szerokość okna. Trzeba jednak wziąć pod uwagę fakt, że dla zmniejszającego się T , norma obserwatora rośnie i dla $T \rightarrow 0$, dąży do nieskończoności. Duża norma obserwatora nie psuje dokładności wyniku, jednak w wypadku wystąpienia dużych zakłóceń pomiarowych taki obserwator będzie je również wzmacniał i może spowodować powstawanie błędów obserwacji.

Dla zwiększającego się przedziału T norma obserwatora maleje i teoretycznie obserwator staje się mniej czuły na zakłócenia. W praktyce jednak, przy zwiększaniu czasu obserwacji mogą wystąpić efekty niekorzystne w zastosowaniach on-line, związane ze wspomnianą liczbą obliczeń, a przede wszystkim z faktem, że obserwatory całkowe nie mogą odtworzyć prawidłowo stanu w pierwszym okresie obserwacji $0 \leq t \leq T$ z powodu braku danych pomiarowych w rejestrach pierwszego okna. Stąd tendencja do minimalizacji T . Można więc sformułować kryterium i określić optymalny okres obserwacji (przykład 13.1).

Warto w tym miejscu naświetlić inną ideę budowy obserwatora całkowego odtwarzającego stan dokładnie. Będzie to obserwator z nieruchomym oknem obserwacji o rozszerzonej szerokości EWO (*Expanding Window Observer*) [14BYW1].

Na bieżąco odtwarzany stan dla $t > 0$ jest dany przez relację podobną do (14.1)

$$x(t) = \int_0^t G_1(t, \tau) y(\tau) d\tau + \int_0^t G_2(t, \tau) u(\tau) d\tau \quad (14.4)$$

Można to interpretować jako rozszerzające się okno z ustalonym lewym końcem ($t = 0$, początek obserwacji) i z bieżąco przesuwany prawym końcem ($t = T$, koniec obserwacji). Ponieważ macierze $G_1(t, \tau), G_2(t, \tau)$ są funkcjami bieżącego czasu t , to dla policzenia całek w przedziale $\tau \in [0, t]$ konieczne jest też pamiętanie pomiarów y, u w całym przedziale $[0, t]$. Stąd liczba pomiarów do zapamiętania ciągle rośnie, a co ważniejsze i trudniejsze do policzenia – optymalne macierze $G(T, \tau)$, składające się dla $\forall T = t$ z coraz to innych przebiegów funkcji w przedziale $[0, t]$, muszą być liczone od nowa dla każdej kolejnej chwili $T = t$ w przedziale $\tau \in [0, t]$. Struktura taka w sposób płynny dostosowywałaby się do rozszerzającego się przedziału obserwacji. Z powodu narastającej liczby obliczeń, jest ona realizowalna tylko dla ograniczonych zakresów czasu. Idea ta sugeruje jednak wersję pośrednią (jaka zostanie przedstawiona w podrozdziale 17.2) polegającą na przygotowaniu off-line kilku różnych obserwatorów o różnych oknach T_i , które można stosować równolegle, zaczynając od najkrótszego okna i przełączając się z biegiem czasu sekwencyjnie na coraz dłuższe okna.

Idea rozszerzanego okna obserwacyjnego EWO została tu przedstawiona jednak z innego, ważniejszego powodu. W rozdziale 16 zostaną omówione dwie różniczkowe wersje obserwatorów odtwarzających również stan dokładnie, a odpowiadające obserwatorom całkowym MWO i EWO. Jak zostanie to wykazane, wersja różniczkowa obserwatora EWO doprowadzi do analogicznych równań, jakie występują przy filtracji Kalmana. Procedury dojścia do tych zależności można więc będzie traktować jako nową, deterministyczną wersję filtracji optymalnej w przestrzeni funkcyjnej $L^2[0, T]$, bez uciekania się do rozbudowanego aparatu procesów stochastycznych, jakie obowiązują w zrozumieniu i wyprowadzeniu teorii filtra Kalmana.

14.2. Analiza stabilności obserwatorów całkowych w układzie ze sprzężeniem zwrotnym

W podrozdziale 12.1.2 przedstawiono opis struktury układu stabilizacji stanu z regulatorem proporcjonalnym (np. LQR) i obserwatorem asymptotycznym Luenbergera (filtrem Kalmana). Podstawowym stwierdzeniem był ważny fakt gwarancji stabilności tak zbudowanego układu ze sprzężeniem zwrotnym LQG. Istotną negatywną cechą okazał się jednak fakt dwukrotnego podniesienia rzędu dynamiki całego układu (do $2n$).

Poniżej zostanie przedstawiony dowód, prezentowany po raz pierwszy w pracy [14BYW2], że zastosowanie obserwatora całkowego MWO, współpracującego w sprzężeniu zwrotnym on-line z regulatorem stabilizującym K , gwarantuje stabilność układu zamkniętego i dodatkowo nie zmienia rzędu układu zamkniętego, który pozostaje równy rzędowi obiektu n .

Przyjmijmy, że dla standardowego, liniowego, obserwowlanego systemu (bez zakłóceń)

$$\dot{x}(t) = Ax(t) + Bu(t)$$

wymagane jest sprzężenie zwrotne od stanu w postaci

$$u(t) = -K x(t),$$

gdzie K jest stałą macierzą regulatora o odpowiednich rozmiarach.

Otrzymuje się następujący układ równań opisujących zachowanie systemu zamkniętego z regulatorem i obserwatorem stanu:

$$\dot{x}(t) = A x(t) + B u(t) \quad (14.5)$$

$$u(t) = -K \int_{t-T}^t G_1(T-t+\tau) C x(\tau) d\tau - K \int_{t-T}^t G_2(T-t+\tau) u(\tau) d\tau$$

Przed przystąpieniem do analizy własności układu zamkniętego, udowodnimy istnienie i jednoznaczność rozwiązania $u(t)$ równania (14.5). Równanie to zawiera opóźnienie

rozłożone. Bez straty ogólności przyjmijmy, że równanie (14.5) rozważać będziemy dla $t \geq 0$ i warunków początkowych:

$$x(t) = \varphi_1(t); \quad u(t) = \varphi_2(t) \quad \text{dla } t \in [-T, 0].$$

Lemat 14.1 [14BYW2]

Dla każdego $T_1 > 0$, $x \in L^2(-T, T_1; R^n)$, $\varphi_2 \in L^2(-T, 0; R^r)$ istnieje jedynie absolutne ciągłe rozwiązanie u równania (14.5) dla $t \geq 0$ i przy założeniu, że $u(t) = \varphi_2(t)$, $t \in [-T, 0]$, a elementy macierzy $G_1(t)$, $G_2(t)$ należą do $L^2(0, T)$.

Dowód

Z własności $x(t)$ i $G_1(t)$ wynika, że pierwsza całka w (14.5) jest funkcją absolutnie ciągłą. Oznaczmy ją przez $\psi(t)$, a całkę drugą rozbijmy na dwie części

$$u(t) = \psi(t) - K \int_{-T}^0 G_2(T-t+\tau) \varphi_2(\tau) d\tau - K \int_0^t G_2(T-t+\tau) u(\tau) d\tau \quad (14.6)$$

Równanie to jest całkowym równaniem typu Volterra, stąd posiada ono jedynie rozwiązanie w przedziale $[0, T]$ [14KOA]. Po rozszerzeniu analizy na przedziały $[(i-1)T, iT]$, gdzie $i \geq 2$, teza zostaje udowodniona.

Lemat 14.2

Każde rozwiązanie u równania (14.5) spełnia również poniższe równanie różniczkowe w przedziale $[0, \infty]$ dla prawie każdego t ,

$$\begin{aligned} \dot{u}(t) = & K \int_{-T}^t \dot{G}_1(T-t+\tau) Cx(\tau) d\tau + K \int_{-T}^t \dot{G}_2(T-t+\tau) u(\tau) d\tau - \\ & - KG_1(T)Cx(t) - KG_2(T)u(t) + KG_1(0)Cx(t-T) + KG_2(0)u(t-T) \end{aligned} \quad (14.7)$$

dla $u(t) = \varphi_2(t)$, $t \in [-T, 0]$.

Dowód

Różniczkowanie obustronne po zmiennej t wzoru (14.5) z wykorzystaniem reguły

$$\frac{d}{dt} \int_{\alpha(t)}^{\beta(t)} f(t, \tau) d\tau = \int_{\alpha(t)}^{\beta(t)} \frac{\delta f(t, \tau)}{\delta t} d\tau + \frac{d\beta(t)}{dt} f[t, \beta(t)] - \frac{d\alpha(t)}{dt} f[t, \alpha(t)],$$

prowadzi bezpośrednio do równania różniczkowego (14.7).

Rozwiązanie równania (14.7) jest również jedynie.

Równanie (14.7) razem z równaniem stanu obiektu daje pełny opis dynamiki układu zamkniętego.

$$\begin{aligned} \begin{bmatrix} \dot{x}(t) \\ \dot{u}(t) \end{bmatrix} = & \begin{bmatrix} A & B \\ -KG_1(T)C & -KG_2(T) \end{bmatrix} \begin{bmatrix} x(t) \\ u(t) \end{bmatrix} + \\ & + \begin{bmatrix} 0 & 0 \\ KG_1(0)C & KG_2(0) \end{bmatrix} \begin{bmatrix} x(t-T) \\ u(t-T) \end{bmatrix} + \begin{bmatrix} 0 \\ K\xi(t) \end{bmatrix} \end{aligned} \quad (14.8)$$

dla $t \geq 0$ i dla $\xi(t) = \int_{-T}^t \dot{G}_1(T-t+\tau) Cx(\tau) d\tau + \int_{-T}^t \dot{G}_2(T-t+\tau) u(\tau) d\tau$,

z warunkami początkowymi: $x(t) = \varphi_1(t)$, $u(t) = \varphi_2(t)$ dla $t \in [-T, 0]$.

Poniżej przedstawiony zostanie dowód stabilności systemu ze sprzężeniem zwrotnym.

Twierdzenie 14.3 [14BYW2]

Równanie (14.8) opisujące system z całkowym obserwatorem z przesuwanym oknem MWO oraz regulatorem statycznym K w sprzężeniu zwrotnym zamkniętego układu sterowania posiada te same wartości własne co system tylko z samym regulatorem (bez obserwatora).

$$\dot{x}(t) = Ax(t) - BKx(t) \quad (14.9)$$

System (14.8) z obserwatorem jest więc asymptotycznie stabilny wtedy i tylko wtedy, gdy system (14.9) bez obserwatora jest asymptotycznie stabilny.

Dowód

Dla wykazania tezy wyliczone zostaną wartości własne λ układu (14.8). W tym celu należy znaleźć wszystkie wykładnicze rozwiązania tego układu.

Podstawiając wykładnicze funkcje:

$$\begin{bmatrix} x(t) \\ u(t) \end{bmatrix} = \begin{bmatrix} x_0 \\ u_0 \end{bmatrix} e^{\lambda t}, \quad x_0 \in R^n, \quad u_0 \in R^r, \quad t \in [-T, 0], \quad \lambda - \text{skalar},$$

do wzoru (14.8), otrzymuje się dwa równania:

$$\begin{aligned} \lambda x_0 e^{\lambda t} = & Ax_0 e^{\lambda t} + Bu_0 e^{\lambda t} \\ \lambda u_0 e^{\lambda t} = & K \int_{-T}^t \dot{G}_1(T-t+\tau) Cx_0 e^{\lambda \tau} d\tau + K \int_{-T}^t \dot{G}_2(T-t+\tau) u_0 e^{\lambda \tau} d\tau - \\ & - KG_1(T)Cx_0 e^{\lambda t} - KG_2(T)u_0 e^{\lambda t} + KG_1(0)Cx_0 e^{\lambda(t-T)} + KG_2(0)u_0 e^{\lambda(t-T)} \end{aligned} \quad (14.10)$$

Stosując po prawej stronie równania reguły całkowania przez części, otrzymuje się równości

$$\begin{aligned} K \int_{t-T}^t \dot{G}_1(T-t+\tau) C x_0 e^{\lambda \tau} d\tau &= K G_1(T-t+\tau) C x_0 e^{\lambda \tau} \Big|_{t-T}^t - \\ - K \int_{t-T}^t G_1(T-t+\tau) C x_0 \lambda e^{\lambda \tau} d\tau &= \\ = K G_1(T) C x_0 e^{\lambda t} - K G_1(0) C x_0 e^{\lambda(t-T)} - K \int_{t-T}^t G_1(T-t+\tau) C x_0 \lambda e^{\lambda \tau} d\tau \end{aligned}$$

oraz dla drugiego składnika

$$\begin{aligned} K \int_{t-T}^t \dot{G}_2(T-t+\tau) u_0 e^{\lambda \tau} d\tau &= K G_2(T-t+\tau) u_0 e^{\lambda \tau} \Big|_{t-T}^t - K \int_{t-T}^t G_2(T-t+\tau) u_0 \lambda e^{\lambda \tau} d\tau = \\ = K G_2(T) u_0 e^{\lambda t} - K G_2(0) u_0 e^{\lambda(t-T)} - K \int_{t-T}^t G_2(T-t+\tau) u_0 \lambda e^{\lambda \tau} d\tau. \end{aligned}$$

Stąd całe równanie ma postać (po odpowiednich uproszczeniach)

$$\lambda u_0 e^{\lambda t} = -K \int_{t-T}^t G_1(T-t+\tau) C x_0 \lambda e^{\lambda \tau} d\tau - K \int_{t-T}^t G_2(T-t+\tau) u_0 \lambda e^{\lambda \tau} d\tau.$$

Gdy pomnoży się przez $e^{-\lambda t}$ równanie ostatnie i równanie (14.10), otrzymuje się dwa równania:

$$\lambda x_0 = A x_0 + B u_0 \quad (14.11)$$

$$u_0 = -K \left[\int_{t-T}^t G_1(T-t+\tau) C x_0 e^{\lambda \tau} d\tau + \int_{t-T}^t G_2(T-t+\tau) u_0 e^{\lambda \tau} d\tau \right] e^{-\lambda t} \quad (14.12)$$

Stąd zgodnie z definicją obserwatora (14.3)

$$x(t) = \int_{t-T}^t G_1(T-t+\tau) y(\tau) d\tau + \int_{t-T}^t G_2(T-t+\tau) u(\tau) d\tau$$

i założonymi wzorami $x(t) = x_0 e^{\lambda t}$ i $u(t) = u_0 e^{\lambda t}$, wzór (14.12) można zapisać w postaci

$$u_0 = -K x(t) e^{-\lambda t} = -K x_0.$$

Łącząc ten wzór oraz (14.11), otrzymuje się końcową zależność dla systemu (14.8)

$$(A - BK)x_0 = \lambda x_0.$$

Wynika z niej, że system (14.8) ma te same wartości własne jak system bez obserwatora (14.9)

$$\dot{x}(t) = (A - BK)x(t).$$

Jest to interesujące i ważne stwierdzenie, że pomimo funkcyjno-różniczkowego równania (FDE) z opóźnieniem opisującego układ ze sprzężeniem zwrotnym, po początkowym okresie obserwacji $[0, T]$, dynamika systemu jest taka sama jak dynamika układu ze statycznym regulatorem K bez obserwatora. Jeśli tylko regulator K jest regulatorem stabilizującym, to cały układ będzie asymptotycznie stabilny. Po czasie $t > T$ dynamika obserwatora jest niewidoczna bez względu na nieznajomość początkowych warunków stanu. Przy pojawienniu się zakłóceń impulsowych w chwili t_z , dynamika ta ujawnia się tylko w okresie $[t_z, t_z+T]$.

Jest to sytuacja bardziej korzystna niż w przypadku używania różniczkowych postaci obserwatorów Luenbergera czy Kalmana, które powodują podwyższenie wymiarowości i rzędu całego systemu, nawet dwukrotnie, a których dynamika w praktycznych zastosowaniach musi się ujawnić od początku, w związku z nierównością rzeczywistego stanu początkowego i założonej początkowej estymaty stanu, oraz każdorazowo przy pojawienniu się zakłóceń. Efekty tej dynamiki trwają teoretycznie nieskończonie długo.

15. Dyskretne wersje obserwatorów dokładnych

15.1. Prosta wersja obserwatora przy dyskretnych pomiarach wyjścia

W metodach ciągłej obserwacji stanu wcześniej omówionych, wymagana jest znajomość przebiegu sygnałów pomiarowych wyjścia i sterowania dla każdej chwili t w całym przedziale obserwacji. Jeśli wyjście rejestrowane jest tylko w wybranych dyskretnych momentach skończonego przedziału obserwacji, a przebieg sterowania znany jest dla każdego $t \in [0, T]$, to można zaproponować dyskretno-ciągły sposób dokładnej obserwacji stanu systemu. Twierdzenie 15.1 gwarantuje możliwość doboru momentów obserwacji dla systemu obserwowalnego, a stąd możliwość budowy takiego algorytmu.

Dany jest system:

$$\begin{aligned}\dot{x}(t) &= Ax(t) \\ y(t) &= Ce^{A(t-t_0)}x(t_0), \quad x \in R^n, \quad y \in R^m \quad \text{dla } \forall t \in [0, T]\end{aligned}\tag{15.1}$$

Twierdzenie 15.1

System (15.1) jest obserwowywalny w przedziale $[t_0, t_1]$ jeśli istnieje liczba q i chwile czasu $\tau_1, \tau_2, \dots, \tau_q \in [t_0, t_1]$ takie, że macierz $V(\tau_1, \tau_2, \dots, \tau_q, t_0)$ o wymiarze $[m \times q, n]$ ma rząd równy n .

$$V(\tau_1, \tau_2, \dots, \tau_q, t_0) = \begin{bmatrix} Ce^{A(\tau_1-t_0)} \\ Ce^{A(\tau_2-t_0)} \\ \vdots \\ Ce^{A(\tau_q-t_0)} \end{bmatrix}, \quad \text{rząd } V = n.$$

Dowód

Należy oszacować określoność macierzy $V'V$. Pozwoli to wyznaczyć nieosobliwość macierzy Grama M . Symetryczna macierz $V'V$ o wymiarze $[n \times n]$ jest dodatnio określona $V'V > 0$, gdy V ma rząd n . Macierz $V'V$ można zapisać jako

$$V'V = \begin{bmatrix} Ce^{A(\tau_1-t_0)} \\ Ce^{A(\tau_2-t_0)} \\ \vdots \\ Ce^{A(\tau_q-t_0)} \end{bmatrix} \begin{bmatrix} Ce^{A(\tau_1-t_0)} \\ Ce^{A(\tau_2-t_0)} \\ \vdots \\ Ce^{A(\tau_q-t_0)} \end{bmatrix} = \sum_{i=1}^q e^{A'(\tau_i-t_0)} C'Ce^{A(\tau_i-t_0)} \quad (15.1)$$

Niech $0 < \Delta < 0,5 \min |\tau_i - \tau_j|$, $i \neq j$, stąd $(\tau_i + \Delta) \in [t_0, t_1]$ dla każdego τ_i .

$$\begin{aligned} M(t_1, t_0) &= \int_{t_0}^{t_1} e^{A'(\tau-t_0)} C'Ce^{A(\tau-t_0)} d\tau \geq \sum_{i=1}^q \int_{\tau_i}^{\tau_i + \Delta} e^{A'(\tau-t_0)} C'Ce^{A(\tau-t_0)} d\tau \approx \\ &\approx \Delta \cdot V'V + v(\Delta), \quad \text{jeśli } \Delta \rightarrow 0. \end{aligned}$$

Macierz $v(\Delta)$ odpowiada macierzy reszt

$$\lim_{\Delta \rightarrow 0} \frac{\|v(\Delta)\|}{\Delta} = 0.$$

Jeżeli rząd $V = n$, $\rightarrow \Delta \cdot V'V + v(\Delta) > 0$, $\rightarrow M(t_1, t_0) > 0$ i system (15.1) jest obserwowywalny.

Wersja tego twierdzenia jest również prawdziwa dla układów niestacjonarnych [15RUD].

Poniżej wyprowadzony zostanie wzór na odtworzenie stanu $x(T)$ przy dyskretnych pomiarach wyjścia, przykładowo dla systemu SISO.

Przyjmuje się, że jest do dyspozycji n pomiar zmiennej wyjściowej y w chwilach $[0 = \tau_1 < \tau_2 < \dots < \tau_n = T]$. Z pomiarów tych można sformować wektor $y_D \in R^n$, $y_D(\tau_i) = [y(0), y(\tau_1), \dots, y(\tau_n)]^T$. Dla każdego τ_i zachodzi wzór

$$Ce^{-A(T-\tau_i)}x(T) = y(\tau_i) + C \int_{\tau_i}^T e^{A(\tau_i-s)}Bu(s)ds.$$

Wykonując n pomiarów w chwilach τ_i i wprowadzając oznaczenie M na odpowiednią nieosobliwą macierz $[n \times n]$, można podać postać wzoru dla obserwatora dokładnego.

$$M = \begin{bmatrix} Ce^{-A(T)} \\ Ce^{-A(T-\tau_2)} \\ \vdots \\ C \end{bmatrix}, \quad x(T) = M^{-1}y_D + M^{-1}C \begin{bmatrix} \int_{\tau_2}^T e^{-A(s)}Bu(s)ds \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (15.3)$$

Wykonanie tylko n pomiarów zmiennej wyjściowej wystarczy do dokładnego odtworzenia stanu $x(T)$ przy braku jakichkolwiek zakłóceń. W obecności zakłóceń wersja (15.3) nie wykazuje jednak żadnych własności optymalnych. Taka wersja obserwatora, sugerowana po raz pierwszy w 1966 r. [15GIJ], może dawać duże błędy odtwarzania.

W przypadku obecności zakłóceń powinno się wykonać dużą liczbę pomiarów (większą niż n) i zastosować wersję minimalnonormową obserwatora. Taką wersję omówimy poniżej.

15.2. Uproszczona wersja minimalnonormowa obserwatora przy dyskretnych pomiarach wyjścia

Dla standardowego systemu:

$$\dot{x}(t) = A x(t) + B u(t),$$

$$y(t) = C x(t),$$

zakłada się, że wyjście $y \in R^m$ jest mierzone w przedziale $[0, T]$ w skończonej liczbie dyskretnych chwil czasu $t_i = T - h_i$, $i = 0, \dots, k$. Natomiast wejście $u \in [L^2[0, T]]^r$ jest mierzone w sposób ciągły w przedziale $[0, T]$. Celem jest określenie stanu końcowego $x(T)$. Równanie wyjścia systemu dla $\forall t_i \in [0, T]$ ma postać:

$$y(t_i) = Ce^{-A(T-t_i)}x(T) - C \int_{t_i}^T e^{A(t_i-s)}Bu(s)ds \quad (15.4)$$

$$y(T-h_i) = Ce^{-Ah_i}x(T) - C \int_{T-h_i}^T e^{A(T-h_i-s)}Bu(s)ds \quad (15.5)$$

W celu obliczenia $x(T)$ należy pomnożyć obie strony równania przez $e^{-A'h_i}C'$

$$\begin{aligned} e^{-A'h_i}C'y(T-h_i) &= e^{-A'h_i}C'Ce^{-Ah_i}x(T) - \\ &- e^{-A'h_i}C'C \int_{T-h_i}^T e^{A(T-h_i-s)}Bu(s)ds. \end{aligned}$$

Kwadratowa macierz $[n \times n]$ $e^{-A'h_i}C'Ce^{-Ah_i}$ jest osobliwa dla każdego h_i . Dlatego należy zsumować k różnych równań w punktach pomiarowych $T - h_i$. Jeżeli system jest obserwowywalny, wtedy istnieje takie k , dla którego rzeczywista macierz Grama M_k nie jest osobliwa, i wtedy $x(T)$ dane jest formułą:

$$M_k = \sum_{i=0}^k e^{-A'h_i}C'Ce^{-Ah_i} \quad (15.6)$$

$$\begin{aligned} x(T) = & M_k^{-1} \sum_{i=0}^k e^{-A'h_i} C'y(T-h_i) + \\ & + M_k^{-1} \sum_{i=0}^k e^{-A'h_i} C'C \int_{T-h_i}^T e^{A(T-h_i-s)} Bu(s) ds \end{aligned} \quad (15.7)$$

Przy równomiernym próbkowaniu można wprowadzić oznaczenie $T - h_i = t_i = i\Delta$ oraz $T = k\Delta$. Po zmianie kolejności całkowania i sumowania oraz zmianie granic całkowania równanie (15.7) przyjmie ostateczną postać:

$$\begin{aligned} x(T) = & M_k^{-1} \sum_{i=0}^k \left[Ce^{-A(k-i)\Delta} \right] y(i\Delta) + \\ & + M_k^{-1} \int_0^T e^{-A'k\Delta} \sum_{i=0}^{\text{Fix}(s/\Delta)} \left[e^{A'i\Delta} C'C e^{Ai\Delta} \right] e^{-As} Bu(s) ds \end{aligned} \quad (15.8)$$

gdzie przez $\text{Fix}(s/\Delta)$ oznaczono część całkowitą z dzielenia s/Δ .

Zaprezentowany powyżej obserwator o skończonej pamięci był nazywany *dead-beat* i był przedmiotem badań w 1992 r. w pracy [15MEA]. Jak można zauważyć na podstawie wyników z rozdziału 13, wersja ta jest analogiem uproszczonej wersji ciągiej (13.5) i nie jest pełną dyskretną wersją minimalnonormową. Jest optymalna tylko wtedy, gdy w pomiarach sterowania nie ma zakłóceń (współczynnik wagi w normie obserwatora $\beta = 0$).

15.3. Pełna wersja minimalnonormowa obserwatora przy dyskretnych pomiarach wyjścia i wejścia

Ponieważ przestrzeń skończenie wymiarowa R^n jest jednym z przykładów przestrzeni Hilberta, to dla podania pełnej wersji obserwatora dyskretnego można wykorzystać uzyskane w podrozdziale 13.2 ogólne wzory. Dla ich przejrzystości skupimy się na wersji SISO.

Odpowiednimi przestrzeniami Hilberta będą $Y = R^{N+1}$, $U = R^{N+1}$, $X = R^n$. Rozważona zostanie podstawowa wersja obserwatora dyskretnego dla współczynnika $\beta = I$ [15PEM].

Wektory y i u składają się z $N+1$ próbek pomiarowych skalarnego pomiaru $y(t)$ i $u(t)$ na przedziale $[0, T]$. Końcowy stan $x(N)$ jest nieznany.

Obiekt dyskretny, którego stan należy odtworzyć, opisany jest równaniem

$$\begin{bmatrix} y_0 \\ \vdots \\ y_N \end{bmatrix} = \mathbf{H}_1 \begin{bmatrix} x_1(N) \\ \vdots \\ x_n(N) \end{bmatrix} + \mathbf{H}_2 \begin{bmatrix} u_o \\ \vdots \\ u_N \end{bmatrix} \quad (15.9)$$

W i -tym kroku dyskretyzacji $i\Delta$ wzór (15.9) ma postać

$$y(i\Delta) = Ce^{-A(N-i)\Delta} x(N\Delta) - Ce^{Ai\Delta} \int_{i\Delta}^{N\Delta} e^{-A\tau} Bu(\tau) d\tau.$$

Operatory \mathbf{H}_1 i \mathbf{H}_2 wynikają z poniższych wzorów

$$\begin{bmatrix} y_0 \\ \vdots \\ y_{i\Delta} \\ \vdots \\ y_{N\Delta} \end{bmatrix} = \begin{bmatrix} Ce^{-AN\Delta} \\ Ce^{-A(N-i)\Delta} \\ \vdots \\ C \end{bmatrix} x(N\Delta) + \mathbf{H}_2 \begin{bmatrix} u_0 \\ \vdots \\ u_N \end{bmatrix} \quad (15.10)$$

Macierz H_2 ma postać, która upraszcza się po wprowadzeniu nowej zmiennej $s = \tau - j\Delta$

$$\begin{aligned} H_2(i\Delta)u = & -Ce^{Ai\Delta} \left(\sum_{j=i}^{N-1} \int_{j\Delta}^{(j+1)\Delta} e^{-A\tau} Bu(\tau) d\tau \right) = \\ = & -Ce^{Ai\Delta} \left(\sum_{j=i}^{N-1} e^{-Aj\Delta} \int_0^\Delta e^{-As} Bu(s+j\Delta) ds \right). \end{aligned}$$

Uwzględniając obecność członu podtrzymującego zerowego rzędu, dla $u(j\Delta) = \text{const}$ otrzymujemy się:

$$H_2 = \begin{bmatrix} -CB_D & -Ce^{-A\Delta}B_D & \dots & -Ce^{-A(N-1)\Delta}B_D \\ 0 & -CB_D & \dots & -Ce^{-A(N-2)\Delta}B_D \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \vdots & \vdots & -CB_D \end{bmatrix}_{N \times N}, \quad B_D = \int_0^\Delta e^{-As} B ds \quad (15.11)$$

Dyskretny obserwator dany macierzami $G_1[n \times (N+1)]$, $G_2[n \times (N+1)]$ ma postać

$$x(N\Delta) = G_1 y + G_2 u \quad (15.12)$$

Norma obserwatora

$$\| (G_1^T, G_2^T) \| = \frac{1}{\Delta} \sqrt{ \left[\sum_{i=1}^n \langle g_1^i, g_1^i \rangle_Y + \sum_{i=1}^n \langle g_2^i, g_2^i \rangle_U \right] }$$

przyjmuje minimum dla optymalnych macierzy otrzymywanych według (13.40)

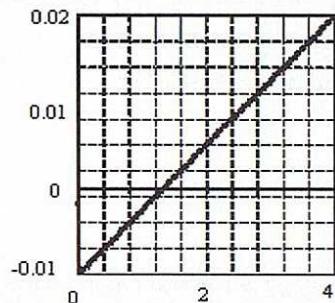
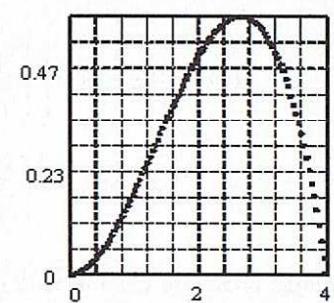
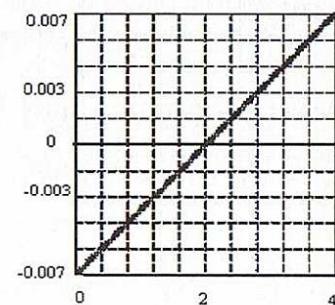
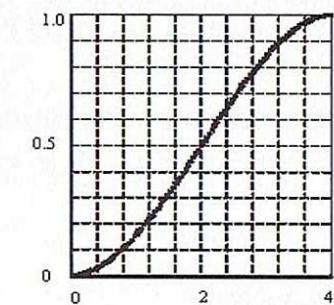
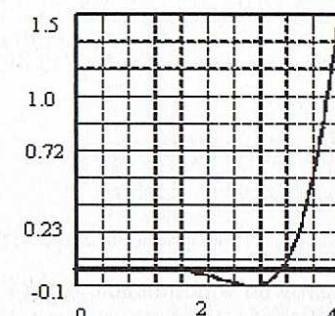
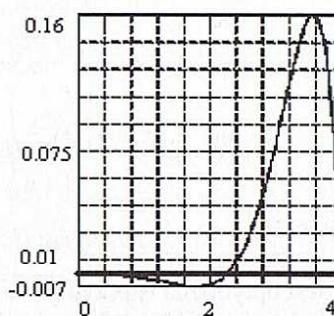
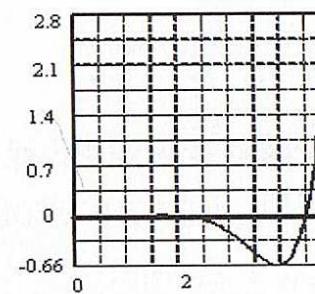
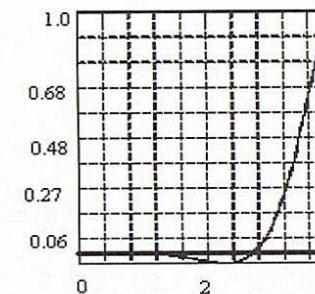
$$G_1^o = (H_1^T F^{-1} H_1)^{-1} H_1^T F^{-1}; \quad G_2^o = -G_1^o H_2 \quad (15.13)$$

gdzie $F = (I + H_2 H_2^T)$.

Wersja ta jest optymalna w przypadku, gdy zarówno w pomiarach y_i , jak i w u_i wystąpią zakłócenia najgorsze o ograniczonej i jednakowej normie. Można też podać analog dyskretny do ogólniejszej wersji ciągiej (13.16)–(13.19) dla $0 < \beta \neq 1$.

Przykład 15.1

Dany jest system II rzędu jak w przykładzie 13.2. Na rysunkach 15.1–15.4 przedstawiono zmiennosć elementów macierzy G_1 i G_2 w obserwatorze *dead-beat* danym wzorem (15.8), dla $T = 4$, $\Delta = 0,04$, $\beta = 0$. Na rysunkach 15.5–15.8 przedstawiono zmiennosć elementów macierzy G_1 i G_2 w obserwatorze minimalnonormowym danym wzorem (15.12), dla $T = 4$ i $\Delta = 0,04$, $\beta = 10$. Przyjęcie współczynnika wagi $\beta \neq 0$ pokazuje duże różnice w optymalnych kształtach macierzy obserwatora.

Rys. 15.1. Kształt G_{11} , $\beta = 0$ Rys. 15.2. Kształt G_{12} , $\beta = 0$ Rys. 15.3. Kształt G_{21} , $\beta = 0$ Rys. 15.4. Kształt G_{22} , $\beta = 0$ Rys. 15.5. Kształt G_{11} , $\beta = 10$ Rys. 15.6. Kształt G_{12} , $\beta = 10$ Rys. 15.7. Kształt G_{21} , $\beta = 10$ Rys. 15.8. Kształt G_{22} , $\beta = 10$

16. Różniczkowe wersje obserwatorów dokładnych

Ze względu na prostotę algorytmu i moce obliczeniowe maszyn cyfrowych najczęściej używaną formą opisu dla obserwatorów pracujących on-line były równania różniczkowe zwyczajne modelujące strukturę asymptotycznych estymatorów stanu. W związku ze wzrostającymi możliwościami komputerów sterujących, realne stało się oprogramowanie algorytmów dla dokładnych, całkowych obserwatorów stanu. Jak powiedziano poprzednio, taki opis zadania odtwarzania stanu jest opisem źródłowym. Obserwatory są bowiem ciągłymi, liniowymi funkcjonalami określonymi na przestrzeni funkcji obserwacji (pomiarów) i mają naturalną reprezentację jako iloczyny skalarne (całki lub sumy). Opis różniczkowy posiada pewne znane, niedogodne własności, np. problem akumulacji błędów w procedurach numerycznych, zwłaszcza w systemach niestabilnych. Może to prowadzić do rozbieżności procedur filtru Kalmana przy długotrwałych obliczeniach. Dla równania różniczkowego występuje konieczność podania warunków początkowych, co przy nieznajomości rzeczywistego stanu początkowego powoduje asymptotyczność zanikania błędu.

Oba poprzednio omówione obserwatory całkowe z rozszerzonym i przesuwanym oknem obserwacji mają swoje odpowiedniki w postaci równań różniczkowych [16BYW]. Interesujące jest wyprowadzenie tych postaci różniczkowych i przeprowadzenie dyskusji o zakresie ich stosowania. Poniżej zostaną one wyprowadzone dla podstawowej wersji obserwatora (13.24) ze współczynnikiem wagi $\beta=1$ w normie obserwatora.

16.1. Różniczkowa wersja obserwatora z rozszerzonym oknem obserwacji EWO

Dla wyprowadzenia odpowiedniego równania różniczkowego obserwatora EWO powtórzony zostanie wzór na jego podstawową postać całkową

$$\begin{aligned} x(T) &= \int_0^T G_1(\tau) y(\tau) d\tau + \int_0^T G_1(\tau) y(\tau) d\tau = \\ &= M(T)^{-1} \cdot \left[\int_0^T \Phi'_{11}(\tau) C' y(\tau) d\tau + \int_0^T \Phi'_{21}(\tau) B u(\tau) d\tau \right] \end{aligned} \quad (16.1)$$

oraz wzory (13.24):

$$\begin{aligned} G_1^{op}(T, t) &= M^{-1}(T) \cdot \Phi'_{11}(t) C' = P_1(t) C' \\ G_2^{op}(T, t) &= M^{-1}(T) \cdot \Phi'_{21}(t) B = P_2(t) B \end{aligned} \quad (16.2)$$

i

$$M(T) = \left[\int_0^T \Phi'_{11}(\tau) C' C e^{A\tau} d\tau \right] e^{-AT}, \quad M^{-1}(T) = e^{AT} \left[\int_0^T \Phi'_{11}(\tau) C' C e^{A\tau} d\tau \right]^{-1} \quad (16.3)$$

Przy założeniu, że końcowy czas obserwacji $T=t$ jest czasem bieżącym, z powyższych równań wynika, że macierz $M(t)$ będzie zależna od t . Obustronne różniczkowanie po T równania (16.1) prowadzi do wzoru

$$\begin{aligned} \dot{x}(T) &= \dot{M}^{-1}(T) \left[\int_0^T \Phi'_{11}(\tau) C' y(\tau) d\tau + \int_0^T \Phi'_{21}(\tau) B u(\tau) d\tau \right] + \\ &+ M^{-1}(T) [\Phi'_{11}(T) C' y(T) + \Phi'_{21}(T) B u(T)]. \end{aligned}$$

Pochodna $dM(T)/dT$ wyraża się wzorem

$$\dot{M}(T) = [\Phi'_{11}(T) C' C e^{AT}] e^{-AT} - M(T) A.$$

Pochodną $\dot{M}^{-1}(T)$ liczymy zgodnie ze wzorem na pochodną macierzy odwrotnej

$$\dot{M}^{-1}(T) = \frac{d}{dT} M^{-1}(T) = -M^{-1}(T) \dot{M}(T) M^{-1}(T).$$

Ostatecznie

$$\begin{aligned} \dot{M}^{-1}(T) &= -M^{-1}(T) [\Phi'_{11}(T) C' C - M(T) A] M^{-1}(T) = \\ &= -M^{-1}(T) \Phi'_{11}(T) C' C M^{-1}(T) + A M^{-1}(T) \end{aligned} \quad (16.4)$$

stąd

$$\begin{aligned} \dot{x}(T) &= [A M^{-1}(T) - M^{-1}(T) \Phi'_{11}(T) C' C M^{-1}(T)] \cdot \\ &\cdot \left[\int_0^T \Phi'_{11}(\tau) C' y(\tau) d\tau + \int_0^T \Phi'_{21}(\tau) B u(\tau) d\tau \right] + \\ &+ M^{-1}(T) [\Phi'_{11}(T) C' y(T) + \Phi'_{21}(T) B u(T)]. \end{aligned}$$

Przy jednostkowym ograniczeniu na macierz G_1 (13.9) i podstawieniu $t=T$, otrzymuje się równanie

$$G_2(T, T) = \int_0^T G_1(T, \tau) C e^{-A(T-\tau)} B d\tau = I \cdot B.$$

Ponieważ jednocześnie $G_2^{op}(T, T) = M^{-1}(T) \Phi'_{21}(T) B$, stąd $M^{-1}(T) \Phi'_{21}(T) = I$.

Ostateczny wzór na pochodną odwzorowanego wektora stanu ma postać równania różniczkowego, gdzie T pełni rolę czasu bieżącego

$$\dot{x}(T) = [A - G_1(T) C] x(T) + G_1(T) y(T) + B u(T) \quad (16.5)$$

Jest to znana postać (12.2) zaproponowana przez Kalmana jako optymalny estymator stanu.

Ze wzorów (16.2) wynika postać symetrycznej macierzy $P_1(T)$

$$P_1(T) = M^{-1}(T) \Phi'_{11}(T) \quad (16.6)$$

Jej bezpośrednie różniczkowanie ze względu na T daje wzór

$$\dot{P}_1(T) = \dot{M}^{-1}(T) \Phi'_{11}(T) + M^{-1}(T) \dot{\Phi}'_{11}(T) \quad (16.7)$$

Pochodną $\dot{\Phi}'_{11}(T)$ można policzyć z (13.23):

$$\begin{aligned} e^{Wt} &= \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{bmatrix}, \quad W = \begin{bmatrix} A & BB' \\ C'C & -A' \end{bmatrix} \\ \dot{\Phi}(t) &= \begin{bmatrix} \dot{\Phi}_{11} & \dot{\Phi}_{12} \\ \dot{\Phi}_{21} & \dot{\Phi}_{22} \end{bmatrix} = \begin{bmatrix} A & BB' \\ C'C & -A' \end{bmatrix} \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{bmatrix} \\ \dot{\Phi}'(t) &= \begin{bmatrix} \dot{\Phi}'_{11} & \dot{\Phi}'_{21} \\ \dot{\Phi}'_{12} & \dot{\Phi}'_{22} \end{bmatrix} = \begin{bmatrix} A' & C'C \\ BB' & -A \end{bmatrix} \begin{bmatrix} \Phi'_{11} & \Phi'_{21} \\ \Phi'_{12} & \Phi'_{22} \end{bmatrix} = \begin{bmatrix} \Phi'_{11} & \Phi'_{21} \\ \Phi'_{12} & \Phi'_{22} \end{bmatrix} \begin{bmatrix} A' & C'C \\ BB' & -A \end{bmatrix} \end{aligned} \quad (16.8)$$

Stąd $\dot{\Phi}'_{11}(T) = \Phi'_{11}(T) A' + \Phi'_{21}(T) BB' = A' \Phi'_{11} + C'C \Phi'_{12}$.

Zgodnie ze wzorem (16.4)

$$\dot{M}^{-1}(T) = -M^{-1}(T) \Phi'_{11}(T) C' C M^{-1}(T) + A M^{-1}(T)$$

z zależności (16.7) otrzymuje się równanie różniczkowe na macierz $P(T)$

$$\begin{aligned}\dot{P}_1(T) = & \left[-M^{-1}(T) \Phi'_{11}(T) C' C M^{-1}(T) + A M^{-1}(T) \right] \Phi'_{11}(T) + \\ & + M^{-1}(T) \left[\Phi'_{11}(T) A' + \Phi'_{21}(T) B B' \right]\end{aligned}$$

Ponieważ $M^{-1}(T) \Phi'_{21}(T) = I$, stąd macierz $P_1(T)$ spełnia równanie Riccatiego

$$\dot{P}_1(T) = A P_1(T) - P_1(T) C' C P_1(T) + P_1(T) A' + B B' \quad (16.9)$$

z warunkami początkowymi wynikającymi z (16.6) dla dowolnego $t_0 = T$. Wzór (16.6) musi być więc liczony tylko raz dla $T = t_0$. Następnie należy rozwiązać równanie (16.9) i wyliczyć $G_1(T)$

$$G_1(T) = P_1(T) C' \quad \text{dla } T > t_0,$$

a ze (16.5) wyznaczyć $x(t)$.

Można zuważyć, że taka różniczkowa postać obserwatora daje dokładny stan rzeczywisty dla $t > T = t_0$, tylko dzięki wcześniejszemu wyliczeniu rzeczywistego stanu początkowego $x(t_0)$. Jeśli jednak w czasie estymacji wystąpią nawet chwilowe zakłócenia, ujawnią się asymptotycznie własności takiego obserwatora, tzn. powstały błąd estymacji będzie dążył do zera dopiero w nieskończoności.

16.2. Różniczkowa wersja obserwatora EWO w ogólniejszym problemie z zakłóceniami niejednostkowymi

W niniejszym podrozdziale przedstawione jest rozwiązywanie dla optymalnego problemu obserwacji przy ogólniejszych założeniach dotyczących zakłóceń. W tym przypadku różniczkową postać obserwatora można otrzymać, łącząc wyniki podrozdziałów 13.4 i 16.1. Postać całkowa obserwatora dla systemu (13.54) jest dana wzorem

$$\begin{aligned}x(T) = & \int_0^T \hat{G}_1(T, \tau) R^{-1/2} C_1^{-1} y(\tau) d\tau + \int_0^T \hat{G}_2(T, \tau) Q^{-1/2} u(\tau) d\tau = \\ = & e^{AT} \left[\int_0^T \hat{\Phi}'_{11}(\tau) \hat{C}' \hat{C} e^{A\tau} d\tau \right]^{-1} \left[\int_0^T \hat{\Phi}'_{11}(\tau) \hat{C}' R^{-1/2} C_1^{-1} y(\tau) d\tau + \int_0^T \hat{\Phi}'_{21}(\tau) B u(\tau) d\tau \right].\end{aligned}$$

Po obustronnym zróżniczkowaniu tego równania ze względu na bieżący czas T i skorzystaniu ze wzorów (13.9), (13.10) i (13.59) otrzymuje się tę samą różniczkową postać obserwatora jak w (16.5)

$$\dot{x}(T) = Ax(T) - \hat{G}_1(T) C x(T) + \hat{G}_1(T) y(T) + Bu(T) \quad (16.10)$$

Z (13.59) i z zależności: $\hat{C}' = C' C_1^{-1} R^{-1/2}$ otrzymuje się

$$G_1(T) = \hat{G}_1(T) R^{-1/2} C_1^{-1} = \hat{P}_1(T) \hat{C}' R^{-1/2} C_1^{-1} = \hat{P}_1(T) C' C_1^{-1} R^{-1} C_1^{-1}.$$

Macierz $\hat{P}_1(T)$ formalnie spełnia to samo równanie różniczkowe Riccatiego jak (16.9), ale dla systemu (13.56).

$$\dot{\hat{P}}_1(T) = A \hat{P}_1(T) - \hat{P}_1(T) \hat{C}' C \hat{P}_1(T) + \hat{P}_1(T) A' + \hat{B} \hat{B}' \quad (16.11)$$

W podrozdziale 13.4 przyjęto oznaczenia:

$$\hat{B} = B Q^{1/2} \quad \text{i} \quad \hat{C} = R^{-1/2} C_1^{-1} C.$$

skąd wynika pełniejsza postać równania Riccatiego, znana z teorii filtracji Kalmana

$$\dot{\hat{P}}_1(T) = A \hat{P}_1(T) - \hat{P}_1(T) C' C_1^{-1} R^{-1} C_1^{-1} C \hat{P}_1(T) + \hat{P}_1(T) A' + B Q B' \quad (16.12)$$

gdzie warunki początkowe dla dowolnego $T = t_0$ są dane przez

$$\begin{aligned}P_1(T) = & e^{AT} \left[\int_0^T \Phi'_{11}(\tau) C' C e^{A\tau} d\tau \right]^{-1} \Phi'_{11}(T) = \\ = & e^{AT} \left[\int_0^T \Phi'_{11}(\tau) C' C_1^{-1} R^{-1} C_1^{-1} C e^{A\tau} d\tau \right]^{-1} \Phi'_{11}(T)\end{aligned}$$

i muszą być obliczone w ten sposób tylko raz.

Postać $\Phi'_{11}(\tau)$ wynika z (13.58).

Wyniki otrzymane w tym podrozdziale i podrozdziale 13.4 dają deterministyczny opis całkowych i różniczkowych wersji obserwatorów z rozszerzonym oknem obserwacji w najogólniejszej sytuacji, gdy zakłócenia dla systemu (13.56) mają niestandardową postać. Otrzymane rezultaty mają identyczną formę jak otrzymywane w stochastycznej teorii filtra Kalmana.

Ze względu na identyczną różniczkową postać opisu dynamiki obserwatora EWO i filtra Kalmana można się doszukiwać analogii pomiędzy stochastyczną (Gaussowską) i deterministyczną ($w L^2$) interpretacją problemów kwadratowych. Autor nie znalazł w literaturze przedmiotu deterministycznej wersji optymalnego filtra Kalmana z interpretacją macierzy Q i R (reprezentujących macierze kowariancji szumów). W niniejszej teorii macierze Q i R są macierzami transformacji zakłóceń z elipsoid do zakłóceń z kul jednostkowych, a optymalność obserwatora wynika z jego minimalnonormowości.

16.3. Różniczkowe wersje obserwatora z przesuwanyem oknem obserwacji MWO

16.3.1. Różniczkowa wersja obserwatora z opóźnieniem

W systemie obserwacji stanu on-line obserwator z przesuwanyem oknem obserwacji jest dany przez (14.3)

$$\dot{x}(t) = \int_{t-T}^t G_1(T-t+\tau) y(\tau) d\tau + \int_{t-T}^t G_2(T-t+\tau) u(\tau) d\tau.$$

Dla jego wersji podstawowej z $\beta = 1$ macierz M dana przez (16.3) jest teraz macierzą stałą, gdyż okres obserwacji T jest ustalony, a obserwator przyjmuje postać

$$x(t) = M^{-1} \cdot \left[\int_{t-T}^t \Phi'_{11}(T-t+\tau) C'y(\tau) d\tau + \int_{t-T}^t \Phi'_{21}(T-t+\tau) Bu(\tau) d\tau \right] \quad (16.13)$$

Po obustronnym zróżniczkowaniu równania (16.13) ze względu na bieżący czas $t \geq T$ otrzymuje się

$$\begin{aligned} \dot{x}(t) &= M^{-1} \left[- \int_{t-T}^t \dot{\Phi}'_{11}(T-t+\tau) C'y(\tau) d\tau + \Phi'_{11}(T) C'y(t) - \Phi'_{11}(0) C'y(t-T) \right] + \\ &+ M^{-1} \left[- \int_{t-T}^t \dot{\Phi}'_{21}(T-t+\tau) Bu(\tau) d\tau + \Phi'_{21}(T) Bu(t) - \Phi'_{21}(0) Bu(t-T) \right]. \end{aligned}$$

Z (16.8) wynika, że pochodne transponowanych podmacierzy fundamentalnych mają postać:

$$\begin{aligned} \dot{\Phi}'_{11}(\cdot) &= A'\Phi'_{11}(\cdot) + C'C\Phi'_{12}(\cdot) \\ \dot{\Phi}'_{21}(\cdot) &= A'\Phi'_{21}(\cdot) + C'C\Phi'_{22}(\cdot) \end{aligned} \quad (16.14)$$

Ze wzorów (16.14) otrzymuje się po przekształceniach różniczkową postać obserwatora

$$\begin{aligned} \dot{x}(t) &= -M^{-1}A' \left[\int_{t-T}^t \Phi'_{11}(T-t+\tau) C'y(\tau) d\tau + \int_{t-T}^t \Phi'_{21}(T-t+\tau) Bu(\tau) d\tau \right] - \\ &- M^{-1}C'C \left[\int_{t-T}^t \Phi'_{12}(T-t+\tau) C'y(\tau) d\tau + \int_{t-T}^t \Phi'_{22}(T-t+\tau) Bu(\tau) d\tau \right] + \\ &+ M^{-1}\Phi'_{11}(T) C'y(t) - M^{-1}\Phi'_{11}(0) C'y(t-T) + \\ &+ M^{-1}\Phi'_{21}(T) Bu(t) - M^{-1}\Phi'_{21}(0) Bu(t-T). \end{aligned}$$

Po uwzględnieniu dodatkowo zależności (16.2) i (16.13) oraz wartości $\Phi(0)$

$$\Phi(0) = \begin{bmatrix} \Phi_{11}(0) & \Phi_{12}(0) \\ \Phi_{21}(0) & \Phi_{22}(0) \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix},$$

jak również wykazanej w podrozdziale 16.1 zależności $M^{-1}\Phi'_{21}(T) = I$, otrzymuje się końcową formułę na obserwator różniczkowy MWO

$$\dot{x}(t) = -M^{-1}A'M x(t) - M^{-1}C'C.$$

$$\begin{aligned} &\cdot \left[\int_{t-T}^t \Phi'_{12}(T-t+\tau) C'y(\tau) d\tau + \int_{t-T}^t \Phi'_{22}(T-t+\tau) Bu(\tau) d\tau \right] + \\ &+ G_1(T) y(t) - M^{-1}C'y(t-T) + Bu(t) \end{aligned} \quad (16.15)$$

Jest to równanie różniczkowe stanu $x(t)$ ze skupionym i rozłożonym opóźnieniem wyjścia $y(t)$ i rozłożonym opóźnieniem sterowania u . Równanie nie jest wygodne jako wzór aplikacyjny. Z powodu ustalonej szerokości okna obserwacji i stałości macierzy M nie występuje tu równanie Riccatiego. Rozwiązywanie startuje z warunku początkowego $x(T)$.

16.3.2. Różniczkowa wersja obserwatora różnicowego bez opóźnienia

Można wyprowadzić prostszą różniczkową postać obserwatora z przesuwanyem oknem. W tym celu należy zauważyc, że w każdej chwili czasu obserwator z przesuwanyem oknem może być formalnie rozważany jako różnica dwóch obserwatorów z rozszerzonym oknem o początkach w chwilach $t_0 = 0$ i końcach w t oraz $t - T$, ale ze stałą macierzą M (16.3) policzoną dla T .

Zgodnie z równaniem (16.13) dla $t \geq T$ zachodzi

$$x(t) = M^{-1} \cdot \left[\int_{t-T}^t [\Phi'_{11}(T-t+\tau) C'y(\tau) + \Phi'_{21}(T-t+\tau) Bu(\tau)] d\tau \right].$$

Jest to równoważne ze wzorem

$$\begin{aligned} x(t) &= M^{-1} \int_0^t [\Phi'_{11}(T-t+\tau) C'y(\tau) + \Phi'_{21}(T-t+\tau) Bu(\tau)] d\tau - \\ &- M^{-1} \int_0^{t-T} [\Phi'_{11}(T-t+\tau) C'y(\tau) + \Phi'_{21}(T-t+\tau) Bu(\tau)] d\tau \end{aligned} \quad (16.16)$$

Po wprowadzeniu dwóch nowych n -wymiarowych zmiennych wektorowych $\eta_1(t)$ i $\eta_2(t)$, które spełniają macierzowe równanie różniczkowe pierwszego rzędu o wymiarze $(2n \times 2n)$, z wybranymi zerowymi warunkami początkowymi $\eta_1(0) = 0$ i $\eta_2(0) = 0$

$$\begin{bmatrix} \dot{\eta}_1(t) \\ \dot{\eta}_2(t) \end{bmatrix} = \begin{bmatrix} -A' & -C'C \\ -BB' & A \end{bmatrix} \begin{bmatrix} \eta_1(t) \\ \eta_2(t) \end{bmatrix} + \begin{bmatrix} C'y(t) \\ Bu(t) \end{bmatrix} = -W'\eta(t) + \begin{bmatrix} C'y(t) \\ Bu(t) \end{bmatrix} \quad (16.17)$$

Rozwiązanie (16.17) ma postać:

a) dla czasu t

$$\eta(t) = \int_0^t e^{-W'(t-\tau)} \begin{bmatrix} C'y(\tau) \\ Bu(\tau) \end{bmatrix} d\tau \quad (16.18)$$

b) dla czasu $t-T$

$$\eta(t-T) = \int_0^{t-T} e^{-W'(t-T-\tau)} \begin{bmatrix} C'y(\tau) \\ Bu(\tau) \end{bmatrix} d\tau \quad (16.19)$$

Z (16.8) wiadomo, że zachodzi

$$\begin{aligned} \Phi'(T-t+\tau) &= \begin{bmatrix} \Phi'_{11}(T-t+\tau), \Phi'_{21}(T-t+\tau) \\ \Phi'_{12}(T-t+\tau), \Phi'_{22}(T-t+\tau) \end{bmatrix} = \\ &= e^{W'(T-t+\tau)} = e^{W'T} e^{-W'(t-\tau)}. \end{aligned}$$

Stąd łącząc a) i b), można utworzyć różnicowy wektor $X(t) = e^{W'T} \eta(t) - \eta(t-T)$

$$\begin{aligned} X(t) &= e^{W'T} \cdot \int_0^t e^{-W'(t-\tau)} \begin{bmatrix} C'y(\tau) \\ Bu(\tau) \end{bmatrix} d\tau - \\ &- \int_0^{t-T} e^{-W'(t-T-\tau)} \begin{bmatrix} C'y(\tau) \\ Bu(\tau) \end{bmatrix} d\tau = e^{W'T} \eta(t) - \eta(t-T). \end{aligned}$$

Pierwsza składowa (n -wymiarowa) wektora $X(t)$ pomnożona przez macierz M^{-1} realizuje odtwarzany stan $x(t)$ według (16.16) w formie różnicowego równania „wyjścia” takiego obserwatora różniczkowego.

$$x(t) = [M^{-1}, 0] X(t) = [M^{-1}, 0] e^{W'T} \cdot \eta(t) - [M^{-1}, 0] \eta(t-T) \text{ dla } t \geq T \quad (16.20)$$

Stąd ostatecznie otrzymuje się dwa równania do opisu obserwatora:

$$\dot{\eta}(t) = -W'\eta(t) + \begin{bmatrix} C'y(t) \\ Bu(t) \end{bmatrix}, \quad \eta(0) = 0 \quad (16.21)$$

$$x(t) = [M^{-1}, 0] e^{W'T} \cdot \eta(t) - [M^{-1}, 0] \eta(t-T) \quad (16.22)$$

Równania (16.21) i (16.22) dają równoważną z równaniem (16.16) różniczkową postać obserwatora z przesuwany oknem. Różniczkowe równanie stanu obserwatora (16.21) jest bez opóźnienia. Opóźnienie występuje tylko w algebraicznym równaniu wyjścia obserwatora (16.22). Dla każdego t równanie różniczkowe musi być rozwiązane, a wartości $\eta(t)$ w przedziale $[t-T, t]$ muszą być pamiętane i użyte do obliczenia stanu dokładnego z równania wyjścia obserwatora. Z powodu ustalonej szerokości okna obserwacji i stałości macierzy M , tak jak i w podrozdziale 16.3.1, tu również nie występuje równanie Riccatiego, stąd złożoność obliczeń nie jest duża.

Należy zauważyć, że równanie różniczkowe (16.21) jest niestabilne, gdyż wartości własne macierzy W są dodatnie. Stąd składniki równania (16.22) są też niestabilne, jednakże ich różnica daje dobrą estymatę stanu. Z powodu tej niestabilności, dla dużych czasów T , różniczkowa postać obserwatora z przesuwany oknem tworzonym jako różnica okien rozszerzanych, nie jest polecana do obliczeń numerycznych. W postaci całkowej z macierzami (13.24) niestabilność macierzy W jest bez znaczenia, z powodu ustalonej i skończonej szerokości okna T .

Wykorzystując wyniki z podrozdziału 13.4, można otrzymać wersje różniczkowe obserwatora dla ogólniejszego problemu optymalizacji i zakłóceń nienależących do kul jednostkowych.

Podsumowując, można stwierdzić, że dla realizacji wersji rozszerzanego okna obserwacyjnego bardziej naturalna jest postać różniczkowa obserwatora w formie filtra Kalmana, a nie postać całkowa EWO. Wersja różniczkowa nie realizuje jednak podstawowego postulatu – dokładnego odtwarzania stanu.

Do realizacji przesuwanego okna obserwacyjnego o skończonej długości bardziej naturalna i stabilna jest całkowa postać obserwatora MWO.

17. Adaptacyjne algorytmy obserwatorów całkowych

Do zalet obserwatora z rozszerzonym oknem obserwacyjnym w wersji EWO można zaliczyć możliwość jego funkcjonowania od chwili $t > 0$ oraz fakt, że z biegiem czasu norma obserwatora maleje, a więc następuje naturalne uodparnianie się obserwatora na ewentualne pojawienie się zakłóceń pomiarowych w y i u .

Obserwator o stałej szerokości okna MWO może podjąć pracę po czasie $t > T$, a jego norma jest stała. Im wybrany czas obserwacji T jest dłuższy, tym norma obserwatora jest mniejsza i obserwator jest mniej wrażliwy na zakłócenia, przy zachowaniu podstawowej cechy – dokładnego odtwarzania stanu.

Filtr Kalmana jest optymalnym filtrem dla zakłóceń typu szum biały, realizującym również asymptotyczne odtwarzanie stanu.

Zalety tych trzech typów obserwatorów zostały wykorzystane w dwóch wersjach adaptacyjnych obserwatorów przedstawionych poniżej. Obserwator adaptacyjny pierwszego typu oparty jest na strukturze przełączanej KF/MWO/KF, w której te dwa typy obserwatorów współpracują ze sobą sekwencyjnie. Obserwator adaptacyjny drugiego typu wykorzystuje strukturę przełączaną MWO/MWO, w której w różnych chwilach czasu pracę podejmują obserwatory o różnych szerokościach okien w zależności od normy pojawiającego się zakłócenia.

17.1. Sekwencyjna współpraca obserwatora całkowego MWO i filtru Kalmana

Dobrze znana różniczkowa formula filtru Kalmana KF (podrozdz.12.3.1)

$$\dot{\tilde{x}}(t) = (A - G(t)C)\tilde{x}(t) + Bu(t) + G(t)y(t) \quad (17.1)$$

wymaga podania wartości startowej estymaty warunków początkowych $\tilde{x}(0)$.

Jeśli w pomiarach nie występują zakłócenia i $\tilde{x}(0) = x(0)$, to filtr Kalmana odtwarza stan dokładnie. Jeśli $\tilde{x}(0) \neq x(0)$, to występuje tylko asymptotyczne nadążanie za stanem rzeczywistym.

Gdy szумy białe występują i znane są dla nich macierze kowariancji $Q \geq 0$ i $R > 0$, to optymalną niestacjonarną macierz wzmacnień filtru $G(t)$ można otrzymać jako

$$G(t) = P(t)C^T R^{-1} \quad (17.2)$$

gdzie kwadratowa macierz wariancji błędu $P(t)$ spełnia macierzowe, nieliniowe różniczkowe równanie Riccatiego:

$$\dot{P}(t) = AP(t) + P(t)A^T - G(t)RG(t)^T + Q, \quad P(0) = P_0 \quad (17.3)$$

W praktyce poprzestaje się na stacjonarnej macierzy P , spełniającej algebraiczne równanie Riccatiego ARE (*Algebraic Riccati Equation*)

$$0 = AP + PA^T - PC^T R^{-1} CP^T + Q \quad (17.4)$$

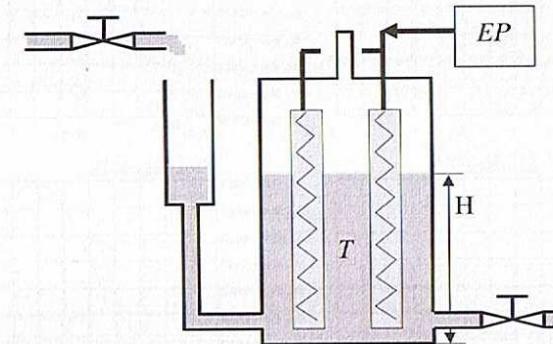
Taki obserwator jest więc optymalny ze względu na własności filtracyjne i daje minimum wartości oczekiwanej błędu średniokwadratowego ze względu na wspomniane szумy pomiarowe. Nie daje on jednak informacji na temat bieżącej wartości błędu rekonstrukcji stanu w przypadku, gdy $\tilde{x}(0) \neq x(0)$. Filtr Kalmana nie jest więc optymalny pod względem nadążania za skokowymi zmianami stanu. Gwarantuje tylko asymptotyczne zanikanie takiego błędu do zera, dając w granicy $\epsilon(\infty) = 0$. Jak już wspomiano, w przypadku braku zakłóceń i równości $\tilde{x}(0) = x(0)$ filtr Kalmana odtwarza stan dokładnie i to przy mniejszym nakładzie obliczeniowym niż obserwator MWO. Niestety tę cenną własność KF traci przy pojawieniu się impulsowych zakłóceń na stanie rzeczywistym (skokowe zmiany stanu). Wtedy ponownie ujawniają się jego własności asymptotyczne. Najlepsze własności będzie więc posiadać obserwator o zmiennej strukturze, w której przełączane są obserwatory KF/MWO/KF w zależności od detekcji momentu wystąpienia impulsowego zakłócenia stanu.

Przetestowany algorytm [17BYW] przedstawiony jest poniżej.

1. W chwili startu procesu obserwacji ($t = 0$) oba obserwatory KF i MWO rozpoczęają przetwarzanie sygnałów, ale tylko wyjście z filtra KF jest od razu użyte do stabilizacji, jako robocza i bezpieczna estymata stanu.
2. Po czasie T , gdy obserwator dokładny MWO zakończy proces obliczeń stanu dokładnego $x(T)$, stan ten zostaje przesłany do KF jako nowy stan początkowy. Od tej chwili filtr Kalmana również odtwarza bieżący stan $x(t)$ dokładnie. Obserwator MWO przerwa pracę.
3. Jeśli na jednej lub kilku zmiennych stanu w chwili $t_k > T$ wystąpi skokowe odchylenie stanu rzeczywistego od estymaty, spowodowane impulsowym zakłóceniem (a nie sterowaniem), to obserwator MWO zostanie ponownie uaktywniony w tle na okres $[t_k, t_k+T]$, w celu odtworzenia wartości dokładnej stanu $x(t_k+T)$. W chwili t_k+T ponownie dokonywana jest korekcja estymaty początkowej dla filtra Kalmana, jak w punkcie 1. Fakt wystąpienia zakłócenia powinien być raportowany przez specjalny (ekspercki) układ detekcji zakłóceń pracujący na podstawie oceny jakości układu sterowania (realizowane w komputerowym systemie czasu rzeczywistego jako przerwanie programowe). Przy braku możliwości wychwycenia momentu zakłócenia obserwator MWO powinien być włączany „na wszelki wypadek” asynchronicznie, jako jedno z zadań o przyznanych priorytetach, lub cyklicznie, jako przerwanie zegarowe. Gdy moce obliczeniowe pozwalają na pracę równoległą z KF w tle dwóch obserwatorów MWO₁ i MWO₂ o różnych szerokościach okien: $[t_k-T_1, t_k]$ i $[t_k-T_2, t_k]$, to ich wskazania w chwili t_k mogą służyć jako detektor wystąpienia zakłócenia. Przy braku zakłóceń wyniki ich pracy są bowiem identyczne, a przy pojawieniu się zakłócenia wyniki będą

różne z powodu różnych norm obserwatorów. Pomimo występującej w takiej sytuacji możliwości stosowania bezpośrednio tylko obserwatora MWO, użycie rozwiązania ze strukturą przełączaną będzie bardziej korzystne, w przypadku gdy problemy optymalnej filtracji szumów są ważniejsze niż dokładne nadążanie za stanem.

Powyższy algorytm został przetestowany na obiekcie III rzędu; badany układ to model wyparki olejowej (rys. 17.1). Taki aparat służy w inżynierii chemicznej do odparowywania lekkich frakcji mieszanin wieloskładnikowych. Do celów badawczych użyto znormalizowanego liniowego modelu III rzędu.



Rys. 17.1. Schemat techniczny wyparki

Transmitancja obiektu (stosunek transformat temperatury cieczy odparowywanej do mocy elektrycznej) dana jest wzorem [17GOH]

$$\frac{T(s)}{EP(s)} = \frac{0,9}{s^3 + 3,8157s^2 + 3,9311s + 1} \quad (17.5)$$

Dla systemu (17.5) przyjęto standardową postać przestrzeni stanu z odpowiednią macierzą Frobeniusa stanu A . Za pomocą programu Matlab wyliczono optymalną statyczną macierz wzmocnień regulatora LQR dla przyjętych macierzy wagowych Q i R we wskaźniku jakości.

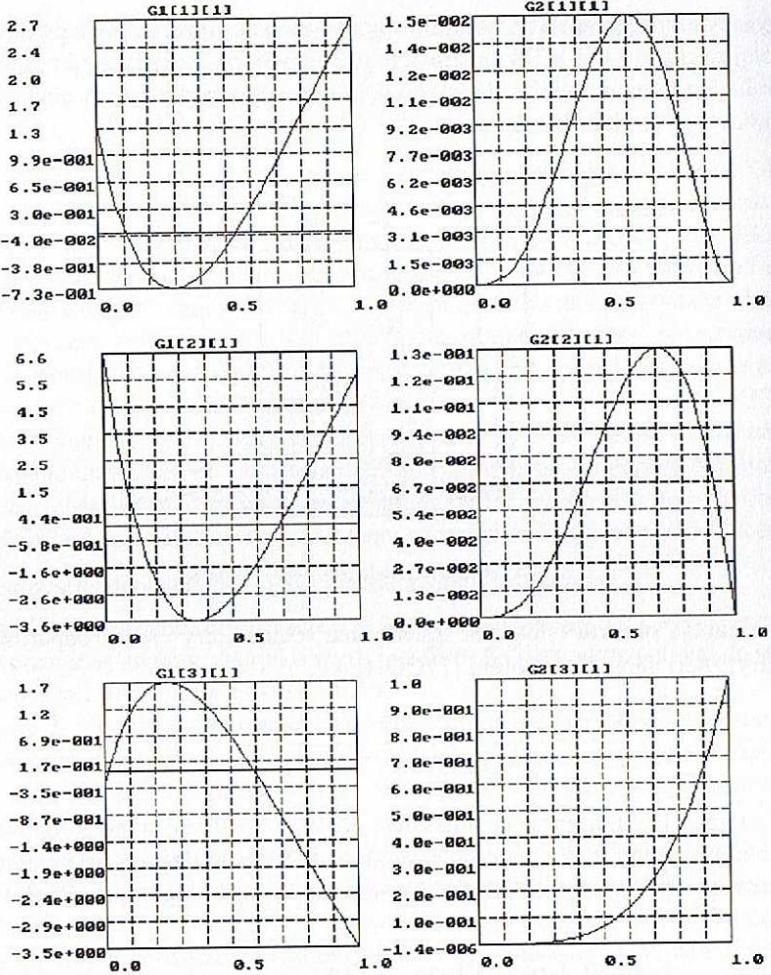
$$K = [1,4495 \quad 2,1938 \quad 1,0530],$$

$$Q = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix}, \quad R = 1.$$

Wyliczono również optymalną macierz wzmocnień G filtra KF, przy macierzach kowariancji białych szumów w pomiarach wejścia i wyjścia.

$$G = \begin{bmatrix} 0,1035 \\ 0,0048 \\ -0,0303 \end{bmatrix}, \quad Q = 0,1, \quad R = 0,1.$$

Dla obserwatora MWO przyjęto szerokość okna $T = 1$ i trzy współczynniki wagi w normie obserwatora $\beta_i = 0,1$. Dla tych parametrów i systemu (17.5) wyliczono wcześniej elementy macierzy funkcyjnych G_1 i G_2 [3×1], których przebieg przedstawiono na rysunku 17.2.



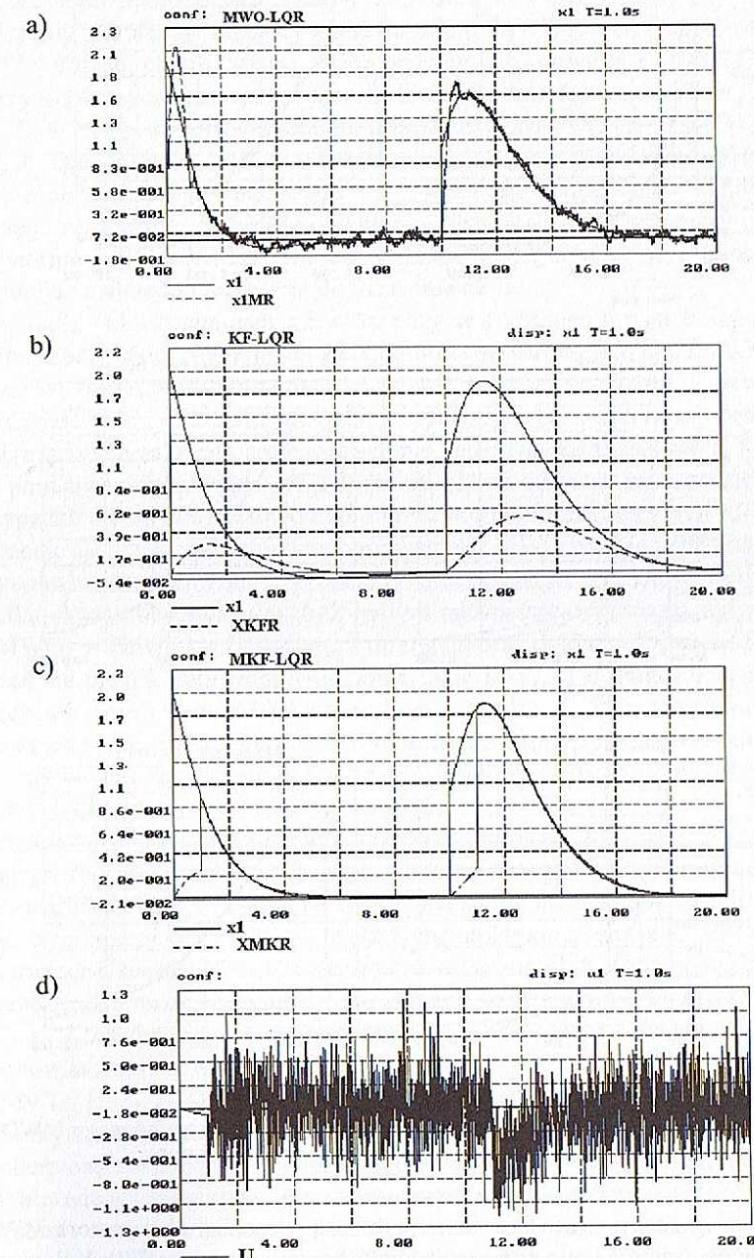
Rys. 17.2. Elementy macierzy obserwacji G_1 i G_2

Wszystkie eksperymenty wykonano przy założeniu obecności białych szumów i skokowego zakłócenia zjawiającego się na wszystkich zmiennych stanu w chwili $t = 0$ i $t = 10$ s. Warunki początkowe stanu przyjęto jako $x(0) = [2, 0, -1]^T$.

Rekonstruowany był stan dla układu otwartego i dla układu zamkniętego w wersji online z regulatorem LQR. W obu eksperymentach zanotowano odpowiedzi układu dla każdego obserwatora pracującego odrębnie i dla obserwatorów współpracujących w wersji przełączanej struktury KF/MWO. Użyto własnego oprogramowania napisanego w języku C [17PEM].

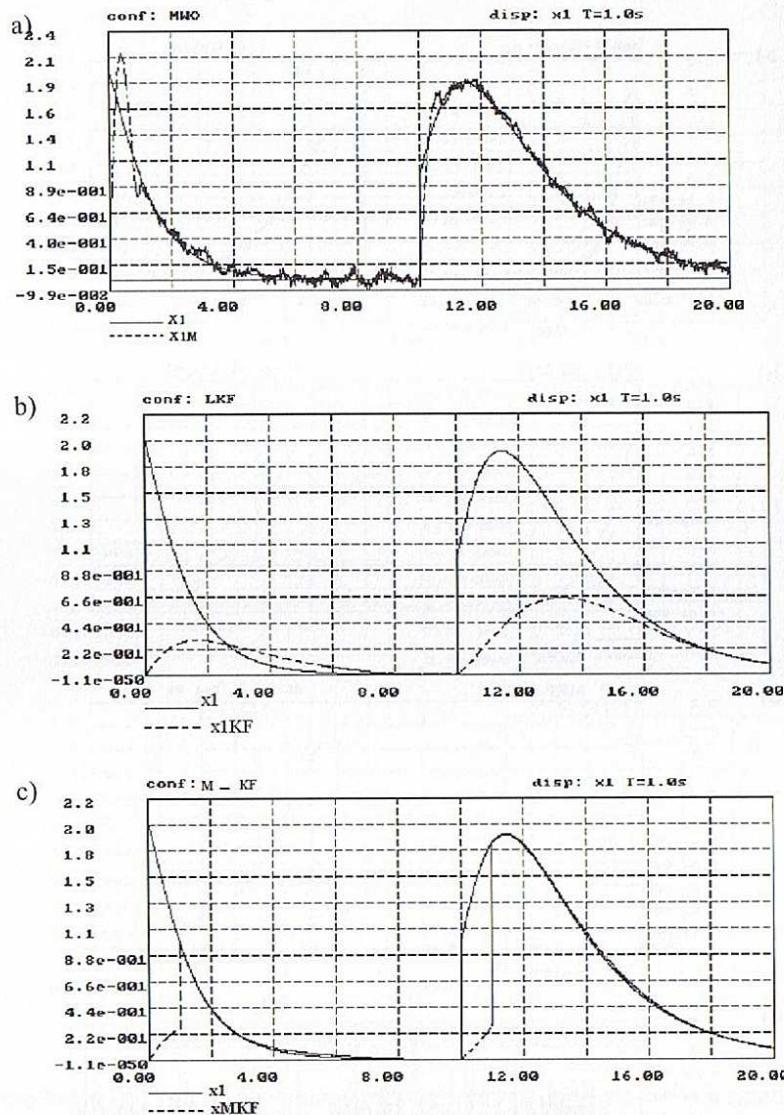
Na rysunkach 17.3a, b i c widoczne są efekty stabilizacji w układzie zamkniętym pierwszej zmiennej stanu (wyjścia) z wykorzystaniem LQR współpracującego kolejno

z MWO, KF i z obserwatorami w strukturze przełączanej MWO/KF, na tle stanu rzeczywistego otrzymywanego w bezpośredniej stabilizacji LQR. Na rysunku 17.3d widoczny jest przykładowy poziom szumów w pomiarze zmiennej sterującej.



Rys. 17.3. Stabilizacja stanu x_1 w układzie zamkniętym: a) obserwator MWO (X1M); b) filtr KF (X1KF); c) struktura MWO/KF (XMKF); d) poziom szumów w zmiennej wejściowej

Na rysunku 17.4 przedstawiono przebiegi odtwarzania stanu w układzie otwartym z użyciem MWO, KF i struktury przełączanej KF/MWO, na tle stanu rzeczywistego.



Rys. 17.4. Odtwarzanie stanu dla zmiennej x_1 w układzie otwartym: a) obserwator MWO (X1M); b) filtr KF (X1KF); c) struktura KF/MWO (XMKF)

Z wykresów można wywnioskować, że lepsze są właściwości obserwatora MWO w odtwarzaniu stanu po czasie T_1 i po wystąpieniu skokowej zmiany stanu. Filtr Kalmana ma lepsze właściwości filtracji białego szumu. Najlepszy efekt daje struktura przełączana KF/MWO.

W eksperymencie potwierdzone zostały lepsze właściwości struktury przełączanej.

17.2. Równoległa współpraca obserwatorów całkowych o różnych długościach okna

W niniejszym podrozdziale zostanie przedstawiona idea współpracy i adaptacyjnego przełączania pomiędzy kilkoma obserwatorami MWO o różnych oknach obserwacyjnych. W wybranej chwili czasu t_k pracę podejmuje obserwator i MWO(T_i) wybrany ze zbioru przygotowanych zawczasu obserwatorów MWO(T_1), ..., MWO(T_n). Ponieważ $0 < T_1 < T_2 < \dots < T_n$, to normy spełniają nierówność $\|MWO(T_1)\| > \|MWO(T_2)\| > \dots > \|MWO(T_n)\|$. Dla szybkiego startu w układzie zamkniętym, pracę rozpoczyna obserwator MWO(T_1) o największej normie. Dla $t > T_1$, w zależności od normy pojawiącego się ewentualnego zakłócenia, na bieżąco wybierany jest obserwator najlepszy (optymalny) o mniejszej normie. Struktura MWO/MWO dostosowuje się adaptacyjnie do poziomu zakłóceń, utrzymując nadziraną tendencję do skracania okna.

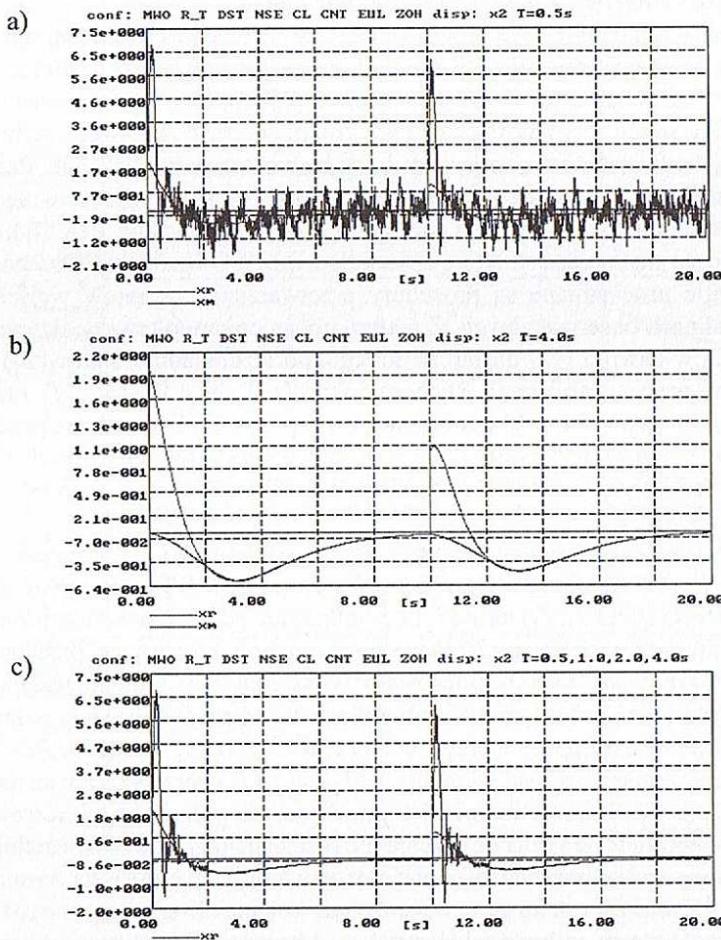
Stabilizacja wolnozmieniennych procesów i użycie szybkiego sprzętu komputerowego pozwala na realizację takiej równoległej wersji multiobserwatora MWO o krótkim czasie startu, poprzez wykorzystanie pierwszego, możliwe najkrótszego okna obserwacyjnego T_1 . Jakkolwiek obserwator optymalny dla takiego krótkiego czasu obserwacji będzie miał dużą normę, to przy braku zakłóceń pomiarowych nie będzie to miało znaczenia. Przy obecności zakłóceń pomiarowych estymata $x(T_1)$ może się różnić od stanu rzeczywistego, w najgorzszym przypadku o błąd szacowany góry przez normę obserwatora $\|(G_1^0(T_1), G_2^0(T_1))\|$.

Proponuje się przełączaną strukturę obserwatorów MWO według poniższego algorytmu.

1. Równolegle uruchamiane są procedury przetwarzania sygnałów wejścia i wyjścia w kilku oknach obserwacyjnych. Z małym opóźnieniem, pierwszy aktywny obserwator MWO₁ w chwili $t = T_1$ dostarcza do regulatora „dokładny” stan $x_1(T_1)$ i może rozpoczęć się proces stabilizacji. Ruchome okno $[t-T_1, t]$ o długości T_1 realizuje dalej proces obserwacji stanu $x_1(t)$ dla bieżącego $t: T_1 < t < T_2$. Równolegle przetwarzane są w dalszym ciągu sygnały pomiarowe w oknach o większych długościach T_i , na podstawie indywidualnych kształtów optymalnych funkcji obserwacyjnych G_{1i}, G_{2i} przygotowanych off-line.
2. W chwili $t = T_2$, po osiągnięciu przez pierwsze okno $[t-T_1, t]$ prawego końca okna drugiego T_2 i po zakończeniu procesu obliczeń przez MWO₂, następuje zamiana bieżącej wartości stanu $x_1(T_2)$ na $x_2(T_2)$ i przełączenie procesu obserwacji na obserwator drugi o mniejszej normie z równoczesnym uruchomieniem jego przesuwu $[t-T_2, t]$. Obserwator pierwszy MWO₁ może zostać wyłączony lub – gdy pozwala to na to moce obliczeniowe – może kontynuować równolegle w tle swoje obliczenia w krótkim oknie $[t-T_1, t]$. Z chwilą osiągnięcia przez ruchome okno T_2 brzegu okna T_3 , $T_3 > T_2$, następuje ponowna zamiana wartości estymaty $x_2(T_3)$ na $x_3(T_3)$ i przełączenie na następną wersję MWO₃ $[t-T_3, t]$. Obserwator MWO₂ może zostać wyłączony lub również równolegle kontynuować w tle swoje obliczenia. Po osiągnięciu brzegu maksymalnego okna T_n uruchomiony zostaje przesuw ostatniego obserwatora o najmniejszej normie.
3. Jeśli bieżące wartości stanu generowane przez kolejne obserwatory nie różnią się między sobą – oznacza to brak zakłóceń, a stąd możliwość realizacji obserwacji przez obserwator o najkrótszym oknie. Równoległa praca co najmniej dwóch obserwatorów MWO pełni więc dodatkowo ważną funkcję. Jest przy okazji konstrukcją detektora wyłapującego moment wystąpienia zakłóceń impulsowych.

4. Jeśli na jednej lub kilku zmiennych stanu w chwili $t_k > T_1$ wystąpi odchylenie stanu rzeczywistego od aktualnej estymaty, to rekonstrukcja nastąpi dokładnie po czasie $t_k + T_i$, gdzie T_i jest bieżącą realizowaną szerokością okna. Czas odtworzenia można skrócić, do $t_k + T_1$, dokonując w t_k uaktywnienia obserwatora o najkrótszym czasie T_1 . W czasie $[t_k, t_k + T_1]$ może nastąpić wyliczenie i przełączenie z bieżącej estymaty na wartość estymaty dostarczaną przez szybki obserwator MWO₁. Kolejne przełączenia korygujące stan nastąpią zgodnie z regułami punktu 2.

Opisany obserwator MWO zmieniający w sposób dynamiczny swoje własności filtracyjne gwarantuje w najlepszy sposób połączenie cech szybkiego reagowania na zmiany stanu i płynnego polepszania dokładności odtwarzania stanu. Łatwo zauważać, że przedstawiona koncepcja przełączania obserwatorów nawiązuje wyraźnie do „dyskretnej” realizacji całkowego obserwatora typu EWO.



Rys. 17.5. Jakość stabilizacji drugiej zmiennej stanu modelu wyparki: a) z obserwatorem MWO $T = 0.5$; b) z obserwatorem MWO $T = 4$; c) z dynamicznie przełączanymi obserwatorami $T_1 = 0.5$, $T_2 = 1$, $T_3 = 2$ i $T_4 = 4$.

Na rysunku 17.5 przedstawiono fragmenty testów badań dotyczących stabilizacji stanu układu wyparki za pomocą czterech przełączanych obserwatorów MWO z narastającymi oknami $T_1 = 0.5$, $T_2 = 1$, $T_3 = 2$ i $T_4 = 4$. Widać wyraźnie z jednej strony szybkość odtwarzania, a z drugiej – wrażliwość na szумy pomiarowe obserwatora z krótkim oknem $T_1 = 0.5$ (rys. 17.5a). Zaobserwować można również lepsze własności filtracyjne i dłuższą reakcję obserwatora o długim oknie $T_4 = 4$ (rys. 17.5b). Najlepsze efekty otrzymuje się przy strukturze przełączanej wszystkich obserwatorów [17PEM].

18. Modelowanie, obserwacja i sterowanie kolumną destylacyjną

18.1. Kolumna jako obiekt sterowania

Jednym z ważniejszych obiektów przemysłowych, który jest stałym przedmiotem badań i jednocześnie służy jako obiekt testowy dla nowoczesnych algorytmów sterowania, jest proces destylacji wielokrotniej, ciąglej, przeprowadzany w wielopółkowych kolumnach rektyfikacyjnych [18NAI; 18LIL]. Zarówno dynamika, jak i statyka tego procesu jest opisywana równaniami nieliniowymi. Szeregowo połączenie wielu przestrzeni międzypółkowych w kolumnie destylacyjnej sugeruje duży rzad problemu, a liczba zmiennych sterujących i liczba wyjść wskazuje na wielowymiarowość modelu. Sam proces chemiczny wymiany masy i energii w przeciwwprądzie faz parowej i ciekłej, którą tworzą surowiec i strumień powrotu (refluks, orosienie), otwiera duże możliwości zadaniom modelowania. Skomplikowane zjawiska zachodzące na półkach, wieloskładnikowość surowca, możliwość odbiorów bocznych i odzysku energii powodują, że kolumna destylacyjna jest też jednym z trudniejszych obiektów do sterowania. Charakterystycznymi cechami dynamiki modelu kolumny są nieminimalnofazowość i złe uwarunkowanie. Pierwsza cecha związana jest z dynamiką fazy parowej i chwilowym zwiększeniem stężenia destylatu przy zwiększeniu fazy parowej i chwilowym wzrostem stężenia destylatu przy zwiększeniu mocy wyparki, a następnie z jego spadkiem, druga – z dużą czułością stężeń na sterowanie przepływem zewnętrznym, np. odbiorem dolnym B, i mniejszą czułością na sterowania przepływami wewnętrznymi, np. refluksu R, czyli ogólnie mówiąc – wrażliwością na kierunek sterowania [18SKS1; 18SKS2; 18COA; 18JAE]. Modele kolumn przeznaczone do zadań projektowania konstrukcji (inżynieria chemiczna) mogą być nieprzydatne do zadań sterowania, stąd dla celów automatyki przyjmuje się zestaw założeń upraszczających. Literatura dotycząca zagadnień modelowania i sterowania kolumn jest bogata [18LUW; 18SHF]. W ostatnich latach rozszerzona została ona o tematykę sterowania adaptacyjnego [18SAS1; 18SAS2; 18MOA; 18GEJ; 18MAS; 18SHF; 18YAD; 18UNH; 18THM; 18MOM]. Przez wiele lat autor niniejszej monografii zajmował się powyższą problematyką z racji badań naukowych i prac prowadzonych przez zespół Instytutu Automatyki AGH przy wdrażaniu komputerowego systemu sterowania nadzawanego dla przemysłowej instalacji destylacji styrenu w ZCH Oświęcim. Wyniki badań były publikowane w pracach [18BYW1; 18BYW2; 18BYW3; 18BYW4; 18BYW5; 18BYW6; 18BYW7; 18BYW8; 18GOH].

Tradycyjne systemy sterowania kolumną w warstwie bezpośredniej posługują się metodami liniowymi i są dwuwymiarowe. Głównym celem jest stabilizacja dwóch wyjść, którymi są skład destylatu i cieczy wyczerpanej, poprzez zmianę dwóch sterowań: przepływ cieczy wyczerpanej i refluksu. W tym celu należy zidentyfikować cztery transmitancje (2×2)-wymiarowej macierzy transmitancji, która reprezentuje zredukowaną dynamikę stężeń i która będzie służyła do projektowania pojedynczych pętli regulacji PID. W tym podejściu uproszczone modele (z reguły inercyjne pierwszego rzędu z opóźnieniem) nie dają wglądu w dynamikę stężeń na poszczególnych półkach. Testy dokładności sterowania takich układów są prezentowane, np. w [19MOM]. W kolumnach z odbiorami bocznymi (petrochemia) stabilizacja całego profilu stężeń w kolumnie (na wszystkich półkach) byłaby najbardziej pożądanym zadaniem sterowania. Stąd modelowanie dynamiki całej kolumny w przestrzeni stanu jest lepszym, choć bardziej skomplikowanym rozwiązaniem. Stężenie w cieczy na poszczególnej półce jest zmienną stanu, stąd rząd modelu zależy od liczby półek. Tylko stężenia na ostatniej półce (destylat) i w wyparce (ciecz wyczerpana) są dostępne do pomiarów chromatograficznych. Stąd zastosowanie wielowymiarowego regulatora, np. LQR, wymusza potrzebę użycia obserwatora stanu.

W warstwie sterowania nadzorującego kolumną ważną rolę odgrywa warstwa optymalizacji statycznej wykorzystująca procedury identyfikacji modeli statycznych procesu, których na podstawie określony jest punkt pracy ustalonej dla kolumny i wyliczana jest w tym punkcie wartość wskaźnika jakości, czyli zysk. Dla najprostszego modelu kolumny dwuskładnikowej punkt pracy określany jest przez przepływ zasilania surowcem F , jego skład x_F i np. pożądany skład cieczy wyczerpanej x_{B_2} . Uzguglając ograniczenia techniczne (np. wydajność pomp), technologiczne (np. wydajność wyparki) i marketingowe (np. pożądany skład destylatu x_D), można wyliczyć dla procedury optymalizacji sterowania optymalne u^* w stanie ustalonym, tzn. wartości przepływów cieczy wyczerpanej B i orosienia R ([18BYW3; 18BYW4; 18DUJ]).

Po wyborze nowego punktu pracy, układ tranzycji przeprowadza instalację do tego punktu, wykonując przyrostowe korekty wartości zadanych Δu dla lokalnych pętli regulacji ΔB (ΔV), ΔR , a warstwa sterowania dynamicznego odpowiedzialna za stabilizację w punkcie pracy, na podstawie pomiarów składów destylatu x_D i cieczy wyczerpanej x_B przeprowadza dynamiczne korekty tych wartości, δB i δR , w celu likwidacji wpływu zakłóceń. W warstwie tej mogą być wykorzystane procedury obserwacji wektora stanu (stężeń na półkach) i przeprowadzana jest identyfikacja dynamiczna nieznanych parametrów i zakłóceń, np. niemierzalnych zmian stężenia surowca $\Delta \bar{x}_F$ (rys. 18.2). Dla modelowania dynamiki odchyłek wektora stanu kolumny od punktu pracy wykorzystywane są układy równań różniczkowych opisujących zachowanie się składu mieszaniny na każdej półce. Podstawowym problemem jest wyznaczenie ustalonego punktu pracy, w którym spełnione są bilanse masowe i energetyczne. Dla odchyłkowego układu równań różniczkowych będzie on reprezentował punkt równowagi z zerowymi warunkami początkowymi.

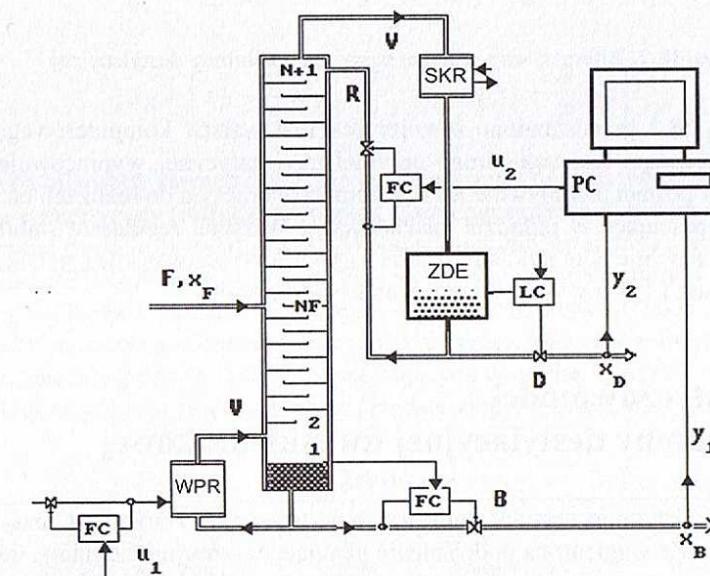
Ustalone wartości przepływów i poziomów, wyliczane w nadzorującym systemie sterowania, przesyłane są do lokalnych regulatorów stabilizujących, jako ich wartości zadane. Sterowaniemi mogą być wartości zadawane przepływu R i V (V – przepływ oparów

proporcjonalny do mocy grzewczej), a wyjściami – stężenia cieczy wyczerpanej x_B i destylatu x_D .

Stabilizacja poziomu zbiornika destylatu jest wykonywana poprzez zmianę przepływu D , a stabilizacja poziomu wyparki (*reboiler*) jest realizowana poprzez odsyłanie B . Jak wspomniano, wyjściem układu jest stężenie cieczy wyczerpanej i destylatu. Przykładowa konfiguracja dla pojedynczej kolumny jest przedstawiona na rysunku 18.1. Dla celów symulacji w następnych rozdziałach będzie w dalszym ciągu przyjmowana właśnie ta struktura.

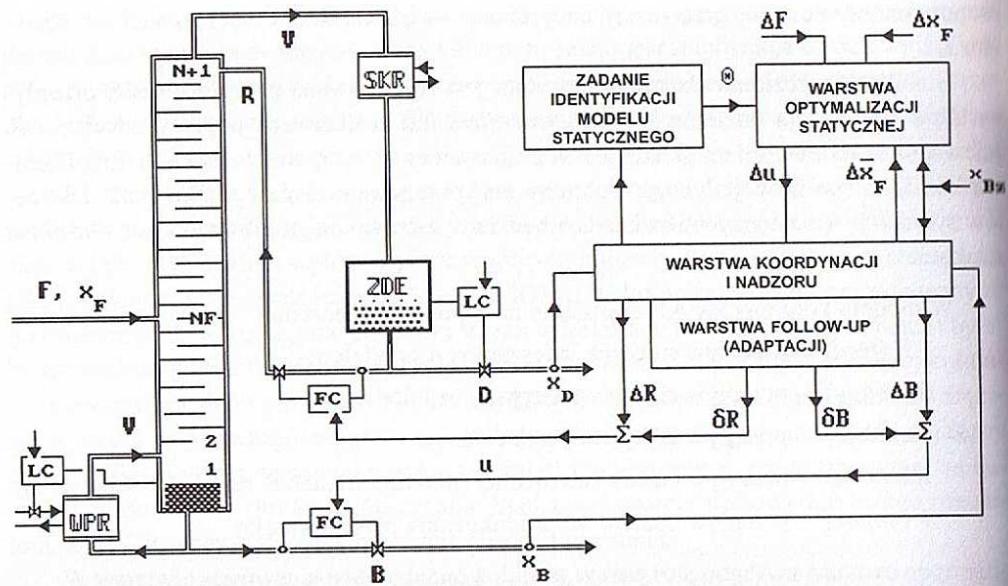
W modelu kolumny zwykle stosujemy następujące oznaczenia:

- F (feed) – przepływ surowca, mieszanina o składzie x_F ,
- B (bottom) – przepływ cieczy wyczerpanej o składzie x_B ,
- D (distillate) – przepływ destylatu o składzie x_D ,
- R (reflux) – przepływ cieczy zwróconej (powrót, oroszenie, refluks) o składzie x_D ,
- V (vapor) – przepływ oparów wyprodukowany przez wyparkę.



Rys. 18.1. Schemat sterowania kolumny destylacyjnej
WPR – wyparka, SKR – skrapacz, ZDE – zbiornik destylacyjny

Innym układem sterowania kolumną może być układ, dla którego składowymi sterowania są refluks R i odsyłanie cieczy wyczerpanej B . Stabilizacja poziomu zbiornika destylatu jest wykonywana poprzez zmianę przepływu D , a stabilizacja poziomu wyparki jest realizowana poprzez zmianę mocy grzewczej, czyli w konsekwencji ilości oparów V . Układ ten przedstawiono na rysunku 18.2.



Rys. 18.2. Alternatywny schemat sterowania kolumny destylacyjnej

Na rysunku 18.2 przedstawiono również schemat warstw komputerowego systemu sterowania nadzawanego. Warstwa góra – optymalizacji statycznej, wypracowuje optymalny punkt pracy, w postaci przepływów R i B . Wartości te przesyła do realizacji na klasyczne regulatory PID, pracujące w układzie kaskadowym. Warstwa regulatora stabilizującego (follow-up) wypracowuje poprawki sterowania na podstawie zaawansowanych algorytmów sterowania od stanu LQR z wykorzystaniem obserwatora stanu.

18.2. Statyczny model kolumny destylacyjnej dwuskładnikowej

W półkowej kolumnie destylacyjnej następuje mieszanie i wymiana masy i energii w przeciwpładzie. Ze względu na podciśnienie panujące w szczycie kolumny, do góry kolumny przesuwają się opary mieszaniny będące w przeciwpładzie ze spadającą przelewowo z półek wrzącą cieczą. Opary przedostają się również przez ciecz zgromadzoną na każdej półce, poprzez zaworki grzybkowe, umieszczone na całej powierzchni półki. Opary przesuwając się, wzbogacają swój skład, zabierając z cieczy lżejszy składnik (lotniejszy) i tracąc cięższy. Skład mieszaniny parowej nad każdą półką (zawartość składnika lżejszego) jest wyższy niż w cieczy na półce. Ostatecznie stężenie mieszaniny rośnie wraz z numerem półki. Opary z półki najwyższej są kierowane do skraplacza. Po skroplieniu otrzymana ciecz tworzy destylat, ale tylko jego pewna część jest odbierana jako produkt końcowy. Większa jego część w formie cieczy jest ponownie kierowana na szczyt kolumny (tzw. refluks R), skąd spływa ponownie w dół kolumny i bierze czynny udział w podnoszeniu stężenia destylatu.

Tylko dzięki temu wewnętrznemu powrotowi cieczy jest wielokrotnie poddawana destylacji, co umożliwia osiąganie wysokich stężeń, np. 0,9999. Takie wielokrotne przetwarzanie tego samego strumienia wymaga doprowadzenia dużych ilości energii do wyparki, w której następuje proces wrzenia i wytwarzania strumienia oparów V , wprowadzanych pod dolną półkę. Ilość energii doprowadzana do wyparki nie może być ani za duża (aby nie „wysuszyć” dołu kolumny), ani za mała (aby nie zalać wyparki cieczą wyczerpaną).

Bilans zewnętrzny stanu ustalonego kolumny dwuskładnikowej (binarnej) obejmuje bilans masy całkowitej i masy składnika lżejszego

$$F = B + D \quad (18.1)$$

$$F x_F = B x_B + D x_D$$

Stąd wypływały wszystkie inne powiązania pomiędzy składami i przepływanymi:

$$\frac{B}{F} = \frac{x_D - x_F}{x_D - x_B}, \quad \frac{D}{F} = 1 - \frac{B}{F} = \frac{x_F - x_B}{x_D - x_B}, \quad \frac{B}{D} = \frac{x_D - x_F}{x_F - x_B} \quad (18.2)$$

$$x_B = \left(x_F - \frac{D}{F} x_D \right) / \frac{B}{F}, \quad x_D = \left(x_F - \frac{B}{F} x_B \right) / \frac{D}{F} \quad (18.3)$$

W tych bilansach zewnętrznych nie jest widoczny i nie jest uwzględniony bilans wewnętrznego energetycznego (refluksów i wyparki). Ustalenie stężenia x_B cieczy wyczerpanej daje przewidywane stężenie destylatu x_D (18.3) pod warunkiem, że znane są wymagane przepływy względne B/F i D/F i skład surowca. Przepływy te oraz wartość odpowiedniego refluksu R/F należy wyliczyć z modelu statycznego, w którym narzuca się jakiś dodatkowy warunek, np. żeby stężenie na półce zasilającej było równe x_F . Wielkość refluksu R jest też związana jednoznacznie z ilością oparów opuszczających wyparkę V , gdyż $V = R + D$.

Przykładowy bilans zewnętrzny jest przedstawiony w tabeli 18.1.

Tabela 18.1

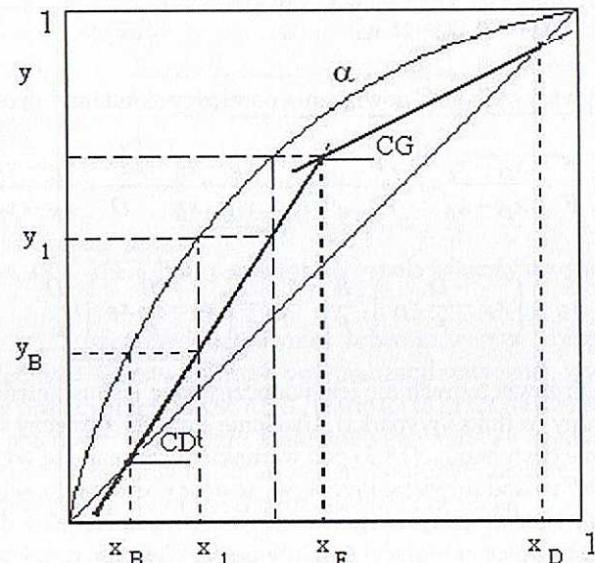
F kmol/s	B kmol/s	D kmol/s	R kmol/s	V kmol/s	x_F	x_B	x_D
14,5	14,0	0,5	7,0	7,5	0,065	0,0022	0,999

Dla mieszanki złożonej z dwóch składników o różnych lotnościach, w warunkach równowagi pomiędzy cieczą i zebraną nad nią parą nasyconą (zbiornik zamknięty) zachodzi znana w chemii zależność pomiędzy stężeniem składnika lekkiego w cieczy x i jego stężeniem w oparach y . Im większe stężenie w cieczy, tym większe w oparach. Ogólnie, zależność ta jest nieliniowa, ale z pewnym przybliżeniem można przyjąć, że stosunek względnego stężenia składnika lżejszego w oparach do wzajemnego stężenia składnika lżejszego

szego w cieczy jest dla niektórych zestawów prawie stały. Jest on nazywany współczynnikiem lotności względnej α i jego wartość dla wielu zestawów dwuskładnikowych można znaleźć w tablicach chemicznych (np. benzen-toluen $\alpha \in [2.3 \div 2.6]$) [18ZIZ; 18SEM].

$$\alpha = \frac{y/(1-y)}{x/(1-x)} \Rightarrow y = \frac{\alpha x}{1+(\alpha-1)x} \quad (18.4)$$

Zależność $y(x)$ daje linię równowagi w postaci łuku i może być naniesiona na standaryzowany wykres stężeń $[y, x]$ w osiach jednostkowych (rys. 18.3).



Rys. 18.3. Wykres równowagowy i proste operacyjne CD i CG

Główym celem modelu statycznego jest określenie składu na każdej półce: w cieczy x_i i w oparach y_i , na podstawie założonego składu cieczy wyczerpanej x_B . Znając również przepływy, skład destylatu x_D można policzyć z (18.3), ale rozkład stężeń wzdłuż kolumny nie jest łatwo wyznaczyć. Jest on rozwiązaniem zbioru równań liniowych i nieliniowych – po dwa na każdą półkę. Algorytm tego rozwiązania wynika z analizy graficznej wykresu równowagowego. W tym celu, w modelu kolumny wprowadza się pojęcie tzw. prostych operacyjnych: dolnej o współczynniku nachylenia CD (tangens kąta) i górnej o współczynniku nachylenia CG . Surowiec podawany jest na półkę zasilającą, która formalnie dzieli kolumnę na dwie części: część górną wzmacniającą i dolną odprowadową. W części górnej odciek (*liquid*) cieczy L_G jest równy tylko R , w dolnej części dochodzi jeszcze przepływ surowca F , czyli $L_D = F+R$. W obu częściach kolumny ilość oparów idąca do góry jest jednakowa, stała i równa V . Po skroplieniu przepływ V rozdziela się na destylat D i refluks R , czyli $V = D+R$.

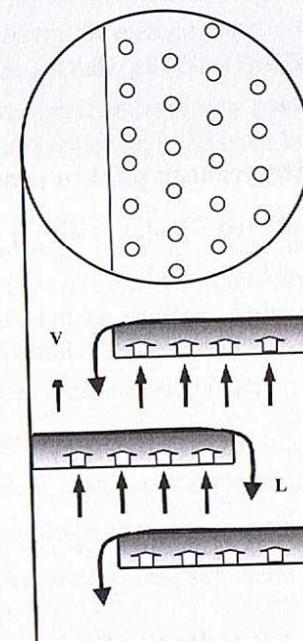
Nachylenia prostych operacyjnych ujmują właśnie stosunek odcieku L do oparów V , stosunek ten w dolnej części kolumny jest inny niż w górnej jej części.

$$CD = \frac{L_D}{V} = \frac{R+F}{R+F-B}, \quad CG = \frac{L_G}{V} = \frac{R}{R+D} = \frac{R}{R+F-B} \quad (18.5)$$

Tylko przy specjalnym nachyleniu prostej operacyjnej dolnej CD , startując z zadanego x_B , można po N_F krokach dojść do półki zasilającej tak, by $x_{NF} = x_F$. Również przy jednym nachyleniu CG po $N_T - N_F$ krokach można dojść do półki ostatniej tak, by $y_{NT} = x_{NT+1} = x_D$. Nachylenia znajduje się iteracyjnie. Znając ostateczne nachylenia prostych operacyjnych, można jednoznacznie z (18.5) określić konieczne przepływy.

$$\frac{R}{F} = \frac{CG}{CD-CG}, \quad \frac{D}{F} = \frac{1-CG}{CD-CG}, \quad \frac{B}{F} = \frac{CD-1}{CD-CG}, \quad \frac{B}{D} = \frac{CD-1}{1-CG} \quad (18.6)$$

Równania modelu statycznego kolumny wynikają z modelu półki teoretycznej, do której dopływa ciecz z półki wyższej i opary z półki niższej, a odpływa w dół ta sama ilość cieczy i w górę – oparów (rys. 18.4).



Rys. 18.4. Przekrój przez kolumnę półkową

Półki numeruje się od dołu kolumny; numer półki zasilającej N_F , numer półki najwyższej N_T .

W dolnej części kolumny $n \in [1, N_F - 1]$ zachodzą następujące zależności dla n -tej półki:

$$Lx_{n+1} + Vy_{n-1} = Lx_n + Vy_n \Rightarrow \text{równania dolnej linii operacyjnej}$$

$$y_{n-1} = CDx_n - (CDx_{n+1} - y_n) \quad \text{lub} \quad x_{n+1} = x_n + \frac{y_n - y_{n-1}}{CD} \quad (18.7)$$

oraz:

$$y_{n-1} = CDx_n - (CD - 1)x_B \quad \text{lub} \quad x_n = x_B + \frac{y_{n-1} - x_B}{CD} \quad (18.8)$$

W górnej części kolumny $n \in [N_F + 1, N_T]$ zachodzą następujące zależności dla półek i równań linii operacyjnej górnej:

$$y_{n-1} = CGx_n - (CGx_{n+1} - y_n) \quad \text{lub} \quad x_{n+1} = x_n + \frac{y_n - y_{n-1}}{CG} \quad (18.9)$$

oraz:

$$y_{n-1} = CGx_n + (1 - CG)x_D \quad \text{lub} \quad x_n = \frac{y_{n-1}}{CG} - \frac{1 - CG}{CG}x_D \quad (18.10)$$

Algorytm modelowania statycznego punktu pracy kolumny

1. Przyjmuje się: $x_B \Rightarrow y_B = \alpha x_B / (1 + (\alpha - 1)x_B)$, $x_0 = x_B$, $y_{-1} = x_B$, $y_0 = y_B$.
2. Przyjmuje się: $x_F \Rightarrow y_F = \alpha x_F / (1 + (\alpha - 1)x_F)$.
3. Zakłada się wstępnie maksymalne możliwe nabylenie $CD_m = (y_F - x_B) / (x_F - x_B) > 1$ i odkłada się na wykresie liczbe N_F schodków-półek według kolejności $x_i \rightarrow y_i \rightarrow x_{i+1} \rightarrow y_{i+1} \dots$
4. $n = 0$, $x_1 = x_B + \frac{y_B - x_B}{CD} \Rightarrow y_1 = \alpha x_1 / (1 + (\alpha - 1)x_1)$.
5. $n = 1$, $x_2 = x_1 + \frac{y_1 - y_B}{CD} \Rightarrow y_2 = \alpha x_2 / (1 + (\alpha - 1)x_2)$.
6. $n = 2$, $x_3 = x_2 + \frac{y_2 - y_1}{CD} \Rightarrow y_3 = \alpha x_3 / (1 + (\alpha - 1)x_3)$.
⋮
- $n = NF - 1$, $x_{NF} = x_{NF-1} + \frac{y_{NF-1} - y_{NF-2}}{CD} \Rightarrow$ jeśli $x_{NF} < x_F \Rightarrow CD = CD - \Delta$ i powrót do kroku 4.

Pochylając stopniowo prostą operacyjną dolną, powtarza się od początku wyliczanie rozkładu stężeń w cieczy i nad półką aż do momentu, gdy na półce o numerze NF zajdzie równość $x_{NF} = x_F$. Takie CD należy zapamiętać. W punkcie $[x_F, y_{NF-1}]$ należy zaczepić górną linię operacyjną. Jej maksymalne wstępne nabylenie $CG_m = (1 - y_{NF-1}) / (1 - x_F)$. Następnie w kolejnych iteracjach pochyla się prosta, przyjmując $CG_{i+1} = CG_i - \Delta$.

Dla kolejnych CG_i wyznacza się składy x_i , zaczynając od $x_{NF} = x_F$ i kończąc na warunku $y_{NT} = x_D$, gdzie $x_D = (y_{NT} - y_{NF-1}) / CG_i + x_F$, z zachowaniem liczby półek w górnej części kolumny. Jeśli warunek nie jest spełniony, procedura pochylenia górnej linii jest powtarzana. Nabylenie górnej prostej operacyjnej CG , dla którego zachodzi $y_{NT} = x_{NT+1} = x_D$ należy zapamiętać. Takie dwa nabylenia CD i CG stosuje się do wyliczenia przepływów zgodnie z (18.6) i rysunkiem 18.3. Z procedury otrzymuje się dodatkowo pełny rozkład ustalony stężeń na półkach.

18.3. Dynamiczny nieliniowy model kolumny destylacyjnej dwuskładnikowej

W celu przedstawienia nieliniowego dynamicznego modelu binarnej kolumny destylacyjnej wokół wyliczonego modelem statycznym punktu równowagi, wprowadzone zostaną odpowiednie założenia [18SHF; 18DUG; 18EDJ; 18SCP; 18SKS2]:

- dwuskładnikowość surowca przychodzącego w postaci wrzącej cieczy,
- stały współczynnik lotności względnej α dla całej kolumny,
- stały profil ciśnienia,
- pomijalna dynamika fazy parowej wzdłuż kolumny $V(t) = V$,
- dokładne mieszanie na półkach i istnienie punktu równowagi ciecz-paro,
- dokładna lokalna regulacja poziomów (pomijalna dynamika przepływów).

Stosowane będą również następujące oznaczenia:

M_i – zatrzymanie cieczy na i -tej półce;

N_T – liczba półek + dół kolumny;

L_i – przelew (ilość cieczy opuszczającej półkę i);

F, B, R, D – przepływ surowca, cieczy wyczerpanej, orosienia i destylatu;

x_F, x_D, x_B – stężenie molowe lekkiego składnika w surowcu, destylacie i cieczy wyczerpanej;

x_i, y_i – stężenie składnika lekkiego na i -tej półce w cieczy i w oparach;

V – natężenie przepływu oparów wewnątrz kolumny;

NF – nr półki zasilającej;

α – współczynnik lotności względnej.

Z bilansu masowego na każdej półce wynikają nieliniowe równania różniczkowe I rzędu: Skraplacz (zbiornik destylatu) $y_{NT} = x_D$:

$$\frac{dM_D(t)}{dt} = V(t) - R(t) - D(t) \quad (\text{równanie ciągłości ogólne});$$

$$\frac{d(M_D x_D)}{dt} = Vy_{NT} - (R + D)x_D \quad (\text{równane ciągłości składnika lekkiego}).$$

Półka ostatnia ($n = NT$)

$$\frac{dM_{NT}}{dt} = R - L_{NT}, \quad \frac{d(M_{NT} x_{NT})}{dt} = Rx_D - L_{NT}x_{NT} + Vy_{NT-1} - Vy_{NT}.$$

Półka przedostatnia ($n = NT-1$):

$$\frac{dM_{NT-1}}{dt} = L_{NT} - L_{NT-1}, \quad \frac{d(M_{NT-1} x_{NT-1})}{dt} = L_{NT}x_{NT} - L_{NT-1}x_{NT-1} + Vy_{NT-R} - Vy_{NT-1}.$$

⋮

Półka n -ta (powyżej zasilania):

$$\frac{dM_n}{dt} = L_{n+1} - L_n, \quad \frac{d(M_n x_n)}{dt} = L_{n+1}x_{n+1} - L_nx_n + Vy_{n-1} - Vy_n.$$

Półka zasilająca ($n = NF$):

$$\frac{dM_{NF}}{dt} = L_{NF+1} - L_{NF} + F, \quad \frac{d(M_{NF} x_{NF})}{dt} = L_{NF+1}x_{NF+1} - L_{NF}x_{NF} + Vy_{NF-1} - Vy_{NF} + Fx_F.$$

Półka najniższa ($n = 1$):

$$\frac{dM_1}{dt} = L_2 - L_1, \quad \frac{d(M_1 x_1)}{dt} = L_2x_2 - L_1x_1 + Vy_B - Vy_1.$$

Wyparka i dół kolumny:

$$\frac{dM_B}{dt} = L_1 - V - B, \quad \frac{d(M_B x_B)}{dt} = L_1x_1 - Vy_B - Bx_B.$$

$D = V - R$ – stały poziom w zbiorniku destylatu ZDE, gwarantowany przez szybki regulator poziomu PID, $D = f_1(M_D)$, jak na rysunku 18.1,

$B = R + F - V$ – stały poziom w wyparce W i w kubie kolumny, gwarantowany przez szybki regulator poziomu PID, $B = f_2(M_B)$, jak na rysunku 18.1.

Zależność pomiędzy stężeniem w cieczy i w oparach dla każdej półki wynosi

$$y_i = \frac{\alpha x_i}{1 + (\alpha - 1)x_i}.$$

Zależność pomiędzy zatrzymaniem i przelewem M_i i L_i wynosi

$$L_i = L_{i0} + (M_i - M_{i0}) / \beta,$$

gdzie L_{i0} , M_{i0} stałe określające punkt pracy, β współczynnik.

Liczba zmiennych: $4NT(M_i, L_i, x_i, y_i) + 3(M_B, x_B, y_B) + 3(M_D, x_D, y_D) + 2(F, x_F) + 2(R, V)$.

Liczba równań: $4NT + 6$

Przepływ F i skład surowca x_F są zadawanymi parametrami (zakłóceniami).

Dwie zmienne R i V są sterowaniami i są zadawane z systemu sterowania nadziedzennego, którego struktura jest przedstawiona na alternatywnym schemacie destylacji (rys. 18.2).

Założenie upraszczające: przyjmując, że zatrzymanie na półkach M jest jednakowe i stałe, uzyskuje się redukcję jednego równania różniczkowego dla każdej półki i uproszczenie obliczeń $L_C = R$, $L_D = R + F$.

Z bilansu masowego na każdej półce wynikają nieliniowe równania różniczkowe I rzędu: Skraplacz (zbiornik destylatu): $y_{NT} = x_D$, $V(t) = R(t) + D(t)$ (równanie ciągłości ogólne)

$$M_D \frac{d x_D}{dt} = V y_{NT} - (R + D)x_D \quad (18.11)$$

Półka ostatnia ($n = NT$)

$$M_{NT} \frac{d x_{NT}}{dt} = R x_D - R x_{NT} + V y_{NT-1} - V y_{NT}.$$

Półka przedostatnia ($n = NT-1$)

$$M_{NT-1} \frac{d x_{NT-1}}{dt} = R x_{NT} - R x_{NT-1} + V y_{NT-R} - V y_{NT-1}.$$

⋮

Półka n -ta (powyżej zasilania)

$$M_n \frac{d x_n}{dt} = R x_{n+1} - R x_n + V y_{n-1} - V y_n.$$

⋮

Półka zasilająca ($n = NF$)

$$M_{NF} \frac{d x_{NF}}{dt} = R x_{NF+1} - (R + F)x_{NF} + V y_{NF-1} - V y_{NF} + Fx_F.$$

⋮

Półka najniższa ($n = 1$)

$$M_1 \frac{d x_1}{dt} = (R + F)x_2 - (R + F)x_1 + V y_B - V y_1.$$

⋮

Wyparka i dół kolumny

$$M_B \frac{d x_B}{dt} = (R + F)x_1 - V y_B - Bx_B.$$

W niniejszym rozdziale przedstawione zostaną przeprowadzone przez autora badania symulacyjne jakości stabilizacji punktu pracy kolumny binarnej, z wykorzystaniem wcześniej prezentowanych algorytmów obserwacji i regulatora LQR. Punkt pracy kolumny wyliczany był na podstawie opracowanego przez autora w 1985 r. dokładnego modelu statycznego kolumny liczonej z półki-na-półkę (*tray-to-tray*) i procedury numerycznej modelującej dowolną liczbę półek. Model ten był wykorzystany później w zadaniu konstrukcji uproszczonego modelu wejście-wyjście aproksymowanego funkcją logarytmiczną, który był wykorzystany w rzeczywistym przemysłowym systemie sterowania [18BYW1; 18BYW8].

18.4. Zlinearyzowany model kolumny destylacyjnej

Po linearyzacji równania równowagi (18.4) otrzymuje się

$$K_i = \frac{dy_i}{dx_i} = \frac{\alpha}{(1 + (\alpha - 1)x_i)^2} \quad (18.12)$$

Linearyzacja równań modelu uproszczonego (18.11) wokół punktu równowagi daje model

$$\frac{d\tilde{x}(t)}{dt} = F(\tilde{x}, \tilde{u}).$$

Macierz stanu systemu zlinearyzowanego w wybranym punkcie pracy \tilde{x}_0, \tilde{u}_0

$$A = \left[\frac{\partial F(\tilde{x}, \tilde{u})}{\partial \tilde{x}} \right]_{\tilde{x}_0, \tilde{u}_0},$$

a cały model ma postać równań z macierzami A, B, C, E :

$$\begin{aligned} \dot{x}(t) &= A x(t) + B u(t) + E \begin{bmatrix} x_F \\ F \end{bmatrix} \\ y(t) &= C x(t) \end{aligned} \quad (18.13)$$

Stan $x(t) = [x_0, \dots, x_N]^T$ tworzą odchyłki stężeń $x_i(t)$ od punktu pracy na każdej półce, gdzie $x_0 = x_B, x_N = x_D$, a N oznacza liczbę półek. Sterowaniem jest $u(t) = [V, R]^T$, a wyjściem są mierzone stężenia cieczy wyczerpanej i destylatu $y(t) = [x_B, x_D]^T$, tak jak na rysunku 18.1.

Macierz A jest trójdziagonalna $(N+1 \times N+1)$ wymiarowa, o współczynnikach:

$$\begin{aligned} i = N: \quad a_{i,i-1} &= K_{i-1} V/M, \quad a_{i,i} = -(R + K_i V)/M; \\ NF < i < N: \quad a_{i,i-1} &= K_{i-1} V/M, \quad a_{i,i} = -(R + K_i V)/M, \quad a_{i,i+1} = R/M; \\ 1 \leq i \leq NF: \quad a_{i,i-1} &= K_{i-1} V/M, \quad a_{i,i} = -(R + K_i V)/M, \quad a_{i,i+1} = (R+F)/M; \\ i = 0: \quad a_{i,i-1} &= -(R+F+K_i V)/M, \quad a_{i,i} = (R+F)/M. \end{aligned}$$

Macierz sterowań B o wymiarze $((N+1) \times 2)$:

$$\begin{aligned} i = N: \quad b_{N,1} &= 0, \quad b_{N,2}; \\ 1 \leq i < N: \quad b_{i,1} &= (x_{i+1} - x_i)/M, \quad b_{i,2} = -(y_i - y_{i-1})/M; \\ i = 0: \quad b_{0,1} &= (x_1 - x_0)/M, \quad b_{0,2} = -(y_0 - x_0)/M. \end{aligned}$$

Macierz wyjścia C o wymiarze $(2 \times (N+1))$

$$C = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}.$$

Dla przykładowej kolumny z $N = 10$ [18BYW9], współczynnika lotności względnej $\alpha = 2,2$, $NF = 6$ oraz założonych stężeń $x_F = x_{NF} = 0,45$ i $x_B = 0,1$, zatrzymań na półkach $M_i = 25,0$ moli i w wyparce $M_B = 200$ moli otrzymano z modelu statycznego konieczne dwa sterowania V, R oraz stężenie x_D :

- surowiec: $F = 10,00$ mol/min, stężenie $x_F = 0,45$;
- opary: $V = 14,025$ mol/min, stężenie $x_B = 0,10$;
- orosienie: $R = 9,71$ mol/min, stężenie $x_D = 0,9112$.

Punkt pracy (stan ustalony) kolumny:

$$x(\infty) = [x_0, \dots, x_N] = [0,1; 0,168; 0,248; 0,328; 0,397; 0,45; 0,525; 0,616; 0,721; 0,823; 0,9112].$$

Linearyzacja modelu nieliniowego kolumny (18.11) wokół tego punktu pracy daje trójdziagonalną macierz liczbową A . Poniżej pierwsza kolumna oznacza dolną diagonalę A , druga kolumna główną diagonalę A , trzecia kolumna górną diagonalę A , czwarta i piąta kolumna oznaczają macierz B , szósta i siódma – macierz E . Macierz C – jak powyżej.

Diag.dol.	Diag.gl.	Diag.gór.	B1	B2	E1	E2
0,9838	-0,1514	0,0985	-0,00048	0,00034	0,00	0,00034
0,8537	-1,6421	0,7884	-0,00448	0,00319	0,00	0,00319
0,7326	-1,5210	0,7884	-0,00449	0,00320	0,00	0,00320
0,6352	-1,4236	0,7884	-0,00389	0,00277	0,00	0,00277
0,5659	-1,3543	0,7884	-0,00296	0,00210	0,00	0,00210
0,5204	-1,3088	0,3884	-0,00204	0,00294	0,40	0,00000
0,4656	-0,8540	0,3884	-0,00258	0,00373	0,00	0,00000
0,4077	-0,7961	0,3884	-0,00289	0,00418	0,00	0,00000
0,3548	-0,7432	0,3884	-0,00283	0,00409	0,00	0,00000
0,0781	-0,7007	0,3884	-0,00243	0,00351	0,00	0,00000
	-0,1403		0,00000	0,00000	0,00,	0,00000

Kolumna jako obiekt dynamiczny jest asymptotycznie stabilna i obserwowały.

Dla symulacji stanów przejściowych w układzie bez sprzężenia zwrotnego zadawano warunek początkowy dla stężeń na półkach x_{pi0} – różny od stanu ustalonego:

$$x_{pB0} = x_1(0), x_{p20} = x_2(0), \dots, x_{pN0} = x_N(0), x_{D0} = x_{N+1}(0),$$

$$x_{pi0} = [0,06; 0,14; 0,2; 0,29; 0,42; 0,41; 0,56; 0,65; 0,68; 0,79; 0,94].$$

Dla symulacji pracy filtra Kalmana zadawano początkową estymatę stanu

$$\bar{x}_{pi0} = [0,3; 0,20; 0,3; 0,4; 0,40; 0,50; 0,5; 0,6; 0,7; 0,70; 0,600].$$

$$\text{Wariancja początkowa stanu: } P[i, i] = 0,5, \quad P[i, j] = 0,$$

$$\text{Kowariancja zakłóceń stanu: } \bar{Q}[i, i] = 1e-5, \quad \bar{Q}[i, j] = 0,$$

$$\text{Koariancja szumów wyjścia: } \bar{R}[i, i] = 1e-5, \quad \bar{R}[i, j] = 0.$$

Używając najprostszej metody dyskretyzacji Eulera, można otrzymać równanie różnicowe modelu (18.13)

$$\tilde{x}_{k+1} = \tilde{x}_k + h \cdot F(\tilde{x}_k, \tilde{u}_k) = f(\tilde{x}_k, \tilde{u}_k),$$

gdzie h jest krokiem dyskretyzacji:

$$x_{k+1} = A_d x_k + B_d u_k,$$

$$y_k = C x_k,$$

przy czym

$$A_d = \left[\frac{\partial f(\tilde{x}_k, \tilde{u}_k)}{\partial \tilde{x}_k} \right]_{\tilde{x}_k, \tilde{u}_k} = I + h \cdot A.$$

Zmienne stanu reprezentują odchyłki od stężeń w punkcie ustalonym. Sterowania są odchyłkami od ustalonych przepływów V i R , a wyjścia odchyłkami od x_B i x_D .

Dla symulacji działania rozszerzonej wersji filtru Kalmana EKF w każdym kroku znajdowana jest inna wartość A_{dk} .

Dla symulacji pracy kolumny w zamkniętej pętli sprzężenia zwrotnego, wyliczono dla powyższego modelu za pomocą pakietu Matlab regulator LQR dla diagonalnej macierzy wag $Q = [10^3, 1, 1, 1, \dots, 1, 10^3]$ i $R = I$. Otrzymano macierz wzmacnienia K (2×11)

$$K = \begin{bmatrix} -143,5; -11,38; -7,3; -4,72; -3,04; -1,94; -1,29; -1,65; -3,18; -6,56; -49,81 \\ 5,72; 0,074; -0,19; -0,21; -0,01; 0,46; 1,35; 2,35; 3,65; 5,5; 3,188 \end{bmatrix} \quad (18.14)$$

Z uwagi na postać macierzy A i C model liniowy kolumny jest obserwowałny.

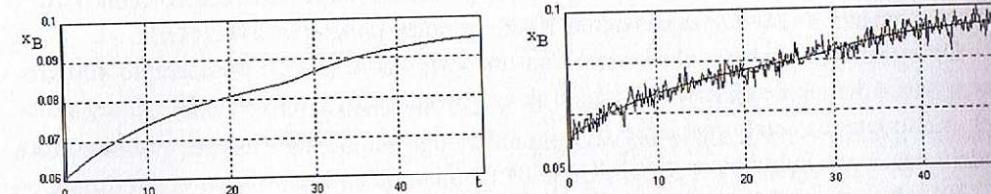
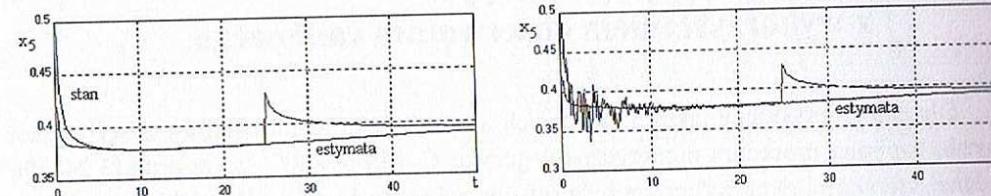
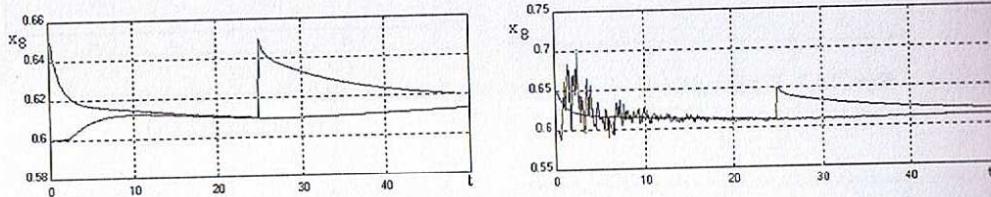
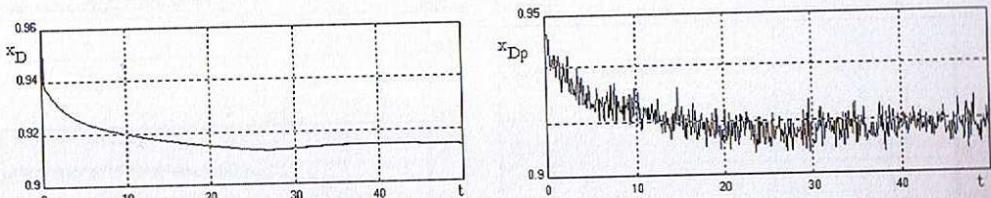
18.5. Eksperymenty symulacyjne z wykorzystaniem rozszerzonego filtra Kalmana

W podrozdziale niniejszym przedstawione są wyniki odtwarzania całego wektora stanu w modelu 10-półkowej kolumny destylacyjnej opisywanej układem nieliniowym 11 rzędów za pomocą rozszerzonego filtru Kalmana [18BYW9; 18BYW10].

Testy przeprowadzono własnym programem w języku C, symulującym nieliniowy model dynamiczny kolumny destylacyjnej.

Na rysunku 18.5 przedstawiono jakość estymacji stężeń na półkach x_D , x_8 , x_5 , x_B (ze skokowymi zmianami x_8 , x_5 – bez szumów pomiarowych).

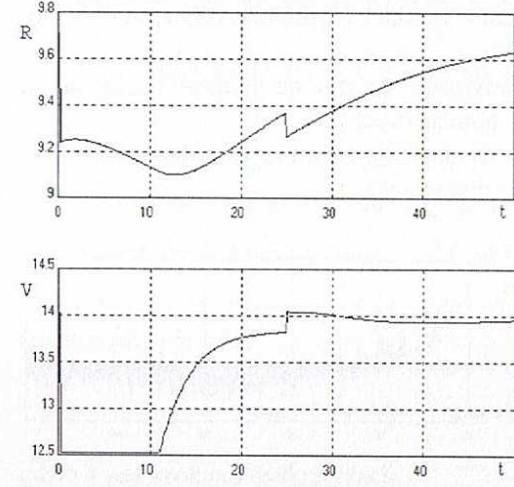
Na rysunku 18.6 przedstawiono jakość estymacji stężeń na półkach x_D , x_8 , x_5 , x_B (przy skokowych zmianach x_8 , x_5 i szumach pomiarowych).



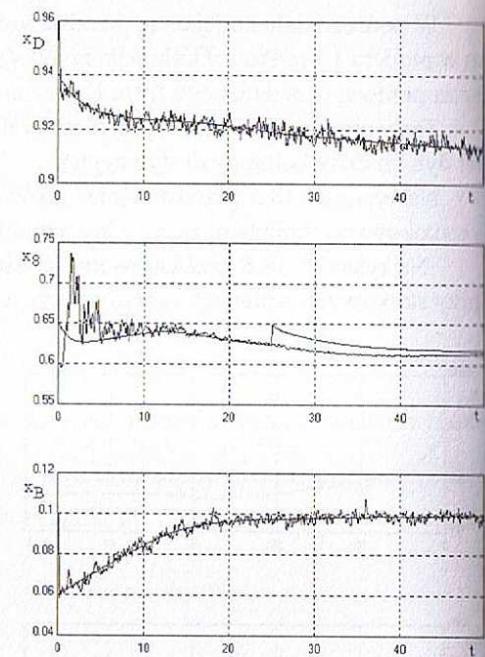
Rys. 18.5. Estymacja stanu przez EKF

Rys. 18.6. Estymacja stanu przez EKF

Na rysunku 18.7 przedstawiono przebieg sterowań R i V , a na rysunku 18.8 jakość stabilizacji kolumny w systemie LQG z EKF i regulatorem LQR na przykładowych półkach x_D , x_8 , x_5 , x_B (przy szumach pomiarowych i skokowych zmianach x_8 , x_5).



Rys. 18.7. Przebiegi R i V
z regulatorem LQG



Rys. 18.8. Stabilizacja kolumny
z regulatorem LQG

18.6. Eksperymenty symulacyjne z wykorzystaniem obserwatora całkowego

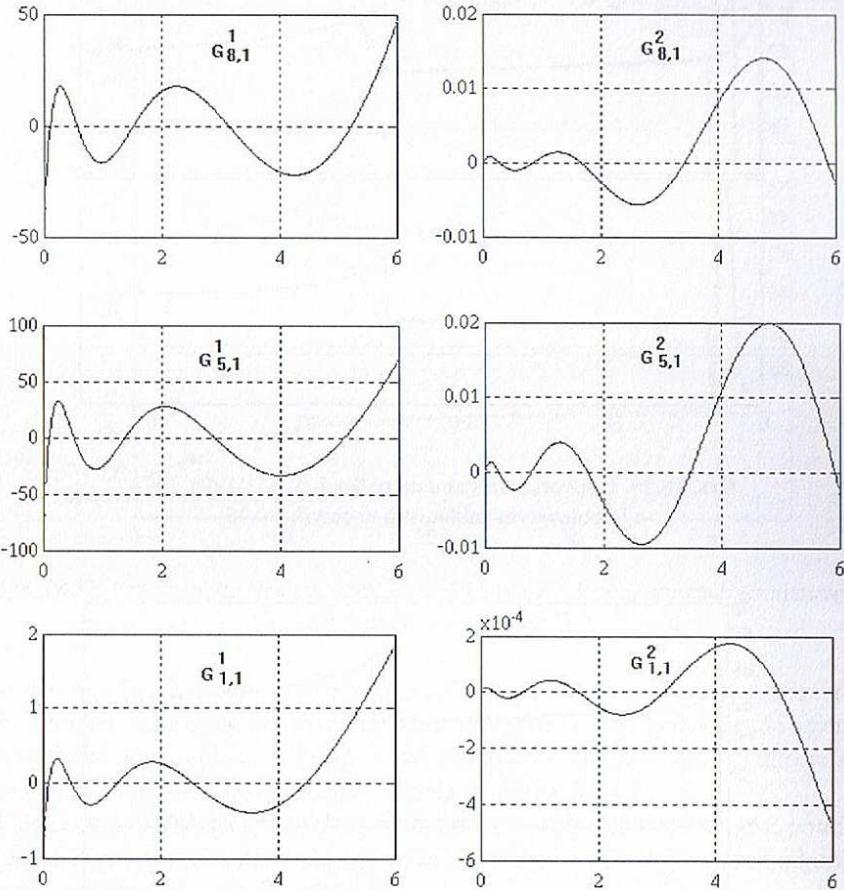
Dla celów testowania jakości całkowych obserwatorów stanu kolumny destylacyjnej została napisana procedura numeryczna w języku C. Macierz M^{-1} ze wzoru (13.24) dla zadanej szerokości okna wyliczana była off-line pakietem Matlaba. Dla dokładnego wyliczenia tej macierzy np. dla $T = 6$, wyniknęła konieczność doboru kroku całkowania $5 \cdot 10^{-5}$, co przy macierzy $\Phi(22 \times 22)$ zajęło około 10 minut pracy komputera P4/2,5GHz.

Do symulacji działania obserwatora on-line (wg wzoru (14.2)) wystarczyło 400 kroków w bieżącym oknie obserwacji, tzn. krok całkowania 0,015 dla $T = 6$. W tylu też próbkach zapamiętane zostały macierze G_1 i G_2 o wymiarach 11×2 dla wyżej wymienionego modelu liniowego kolumny: $x \in R^{11}$, $y \in R^2$, $u \in R^2$.

Kształt funkcji obserwatora dla modelu liczbowego z podrozdziału 18.4 i czasu $T = 6$ jest przedstawiony na rysunku 18.9. Funkcje z lewej przedstawiają wybrane elementy

pierwszej kolumny macierzy G_1 (mnożone następnie przez $y_1(t) = x_B(t)$, a funkcje z prawej – elementy pierwszej kolumny macierzy G_2 , (mnożone następnie przez $u_1(t) = V(t)$). Elementy dotyczą półek nr 8, 5 i 1.

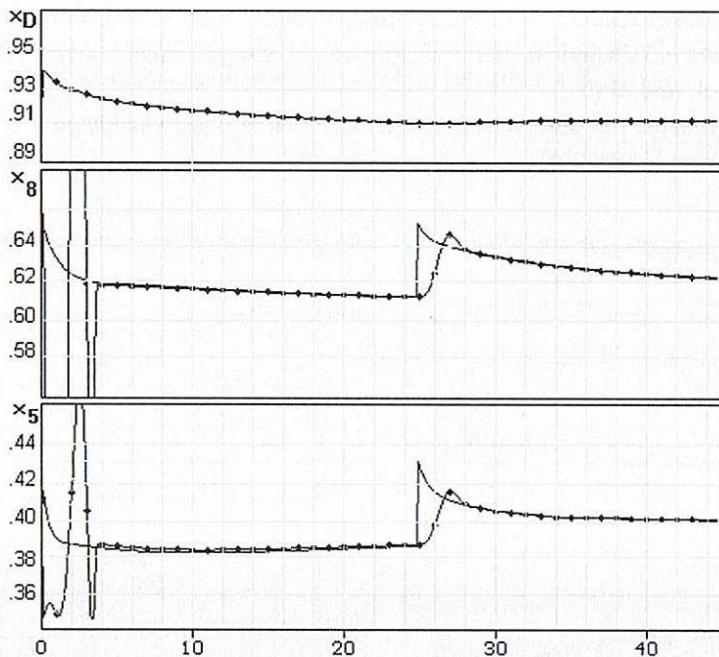
Te same funkcje dla obserwatora o szerokości okna obserwacyjnego $T = 4$ są o rząd wielkości większe [18BYW9].



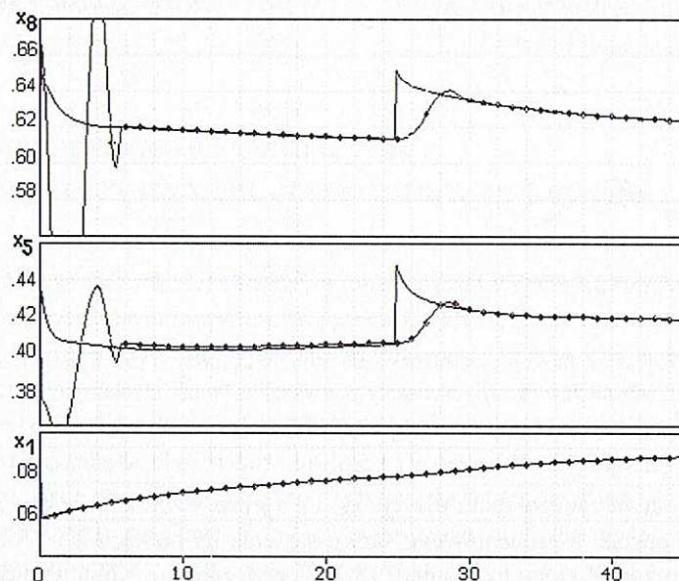
Rys. 18.9. Kształt funkcji obserwatora dla modelu kolumny i czasu obserwacji $T = 6$.
Funkcje z lewej – elementy pierwszej kolumny macierzy G_1 .
Funkcje z prawej – elementy pierwszej kolumny macierzy G_2

Przebieg odtwarzanego stanu dla x_D , x_8 , x_5 , x_B przy szerokości okna $T = 4$, zakłócenach stanu w postaci 3-procentowego skoku stężenia na półkach 8 i 5 i braku zakłóceń pomiarowych przedstawiono na rysunku 18.10, a przy oknie $T = 6$ na rysunku 18.11.

Na rysunkach widać idealną dokładność odtwarzania dla obserwatora z przesuwany oknem (MWO) po czasie obserwacji równym dokładnie szerokości okna T .

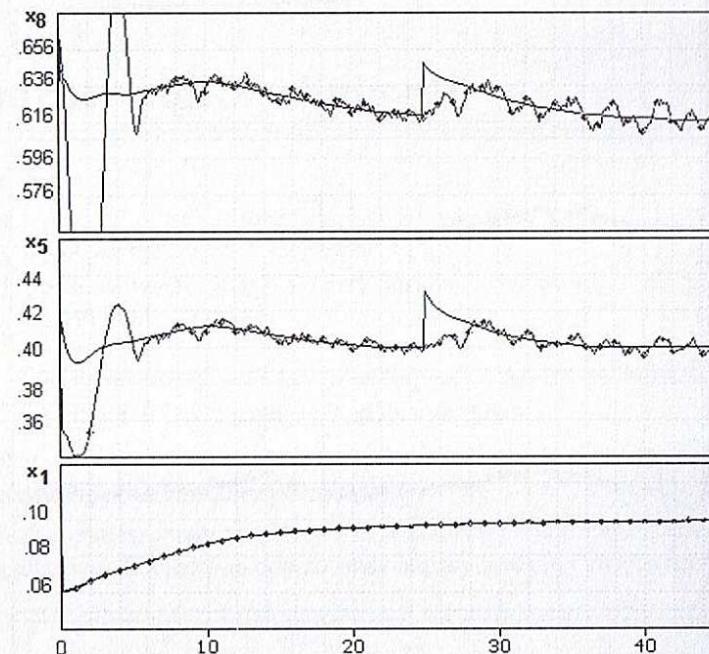


Rys. 18.10. Odtworzenie stanu na półkach 5, 8 i 11 dla $T = 4$
i impulsowym zakłóceniu w chwili $t = 25$



Rys. 18.11. Odtworzenie stanu na półkach 1, 5 i 8 dla $T = 6$
i impulsowym zakłóceniu w chwili $t = 25$

Ten sam test dla obserwatora $T = 6$ w obecności szumów pomiarowych daje wyniki pokazane na rysunku 18.12.



Rys. 18.12. Odtworzenie stanu na półkach 1, 5 i 8 dla $T = 6$ przy szumach pomiarowych
i zakłóceniu w chwili $t = 25$

Na rysunku 18.13 pokazano rezultaty stabilizacji kolumny z wykorzystaniem regulatora LQR (18.14) i całkowego obserwatora stanu MWO dla $T = 6$. Symulowano szum pomiarowy o wartości wariancji $\text{var} = 0,003$ oraz dodatkowo w chwili $t = 25$ symulacje zakłóceń stanu w postaci 3-procentowego skoku stężenia na półkach 8 i 5.

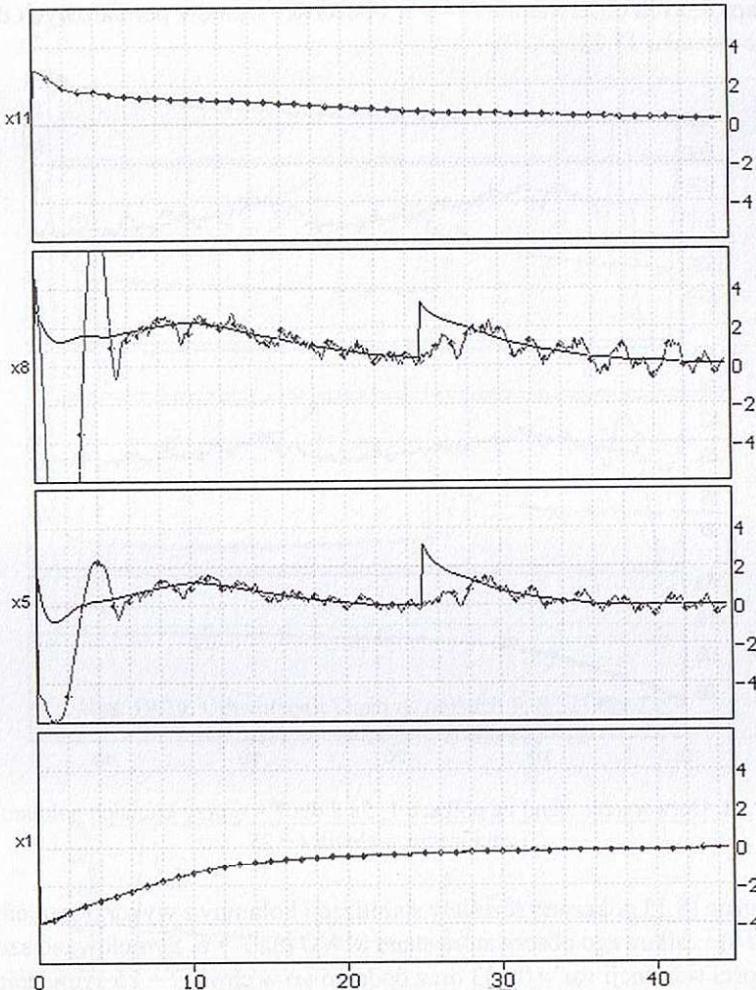
Widac dobrą stabilizację wyjść $x_D = x_{11}$, $x_B = x_1$ w obecności szumów, pomimo gorszych estymat półek wewnętrznych, co wiąże się z większymi wartościami odpowiednich elementów macierzy G_1 i G_2 (rys. 18.9).

Wykonane obliczenia normy obserwatora MWO dla powyższego modelu kolumny dały wyniki:

$$\|G_1, G_2\|_{T=4} = 1,7049 \cdot 10^6,$$

$$\|G_1, G_2\|_{T=6} = 4,4131 \cdot 10^4,$$

$$\|G_1, G_2\|_{T=8} = 4,1878 \cdot 10^3.$$



Rys. 18.13. Stabilizacja kolumny z wykorzystaniem regulatora LQR i całkowego obserwatora stanu MWO dla $T = 6$

19. Aneks 1. Elementy algebry liniowej

Przegląd podstawowych pojęć algebry liniowej opracowano na podstawie prac [19GAF; 19KAT; 19KAT1; 19TUA].

19.1. Macierz i jej zagadnienie własne

Definicja formalna macierzy liczbowej

Układ liczb rozmieszczony w tablicy o m wierszach i n kolumnach nazywamy macierzą $[m \times n]$ -wymiarową. Stanowi on odwzorowanie pary liczb (i, j) w element a_{ij} .

$$A = \begin{bmatrix} a_{11} & a_{1n} \\ \dots & \ddots & \dots \\ a_{m1} & a_{mn} \end{bmatrix}.$$

Definicja operatorowa macierzy

Każdy liniowy operator przekształcający skończenie wymiarową przestrzeń $X^n \rightarrow X^m$, (przy zadanych wektorach bazowych) ma postać macierzy $[m \times n]$ -wymiarowej.

Macierz kwadratowa przekształca elementy (wektory) przestrzeni X^n w inne elementy tej samej przestrzeni.

$$Ax = y, \quad x, y \in X^n.$$

Dla każdej macierzy kwadratowej $[n \times n]$ istnieją w przestrzeni X^n charakterystyczne wektory, które po przekształceniu przez tę macierz pozostają kolinearne, tzn. zmieniają tylko swoją długość z pewnym współczynnikiem λ . Wektory te nazywają się wektorami własnymi macierzy A , a współczynniki proporcjonalności nazywane są wartościami własnymi tej macierzy.

$$Aw = \lambda w \Rightarrow (A - \lambda I)w = 0 \quad (19.1)$$

Aby powyższy układ posiadał niezerowe rozwiązanie dla wektora w , rzęd macierzy $(A - \lambda I)$ musi być mniejszy od n . Stąd z warunku $\det[A - \lambda I] = 0$ otrzymuje się wielomian

charakterystyczny, który posiada n pierwiastków λ_i . Dla danego $\lambda = \lambda_i$ wyznacza się z układu równań (19.1) wektor w_i , ale tylko z dokładnością do współczynnika c , tzn. wektor $c \cdot w_i$ (c – dowolna stała) również spełnia ten układ. Z układu (19.1) wyznacza się, więc nie wektory, a raczej kierunki własne, które nie ulegają zmianie przy przekształceniu przez macierz A . Stąd wektory własne z reguły mogą być znormalizowane do jedności.

Jeśli wartości własne są pojedyncze, to można wyznaczyć n wektorów własnych liniowo niezależnych. Macierz W , której kolumny są utworzone z wektorów własnych macierzy A , nazywa się macierzą modalną macierzy A .

19.2. Przekształcenie podobieństwa i jego związek z bazą przestrzeni

W skończenie wymiarowej przestrzeni wektorowej bazę stanowi uogólniony układ współrzędnych n wektorów liniowo niezależnych. Przy przekształceniach liniowych przestrzeni w siebie wygodnie jest dobrać odpowiednią bazę. Dany wektor x w bazie E ma inne współrzędne \bar{x} w bazie W . Wynikają one ze związku

$$x = P\bar{x}$$

gdzie P – nieosobliwa macierz transformacji z bazy E do W .

Gdy dane jest przekształcenie przestrzeni w siebie $Ax = y; x, y \in R^n$ w pewnej bazie, to w innej bazie otrzymamy:

$$AP\bar{x} = P\bar{y}, \quad \text{gdzie } x = P\bar{x}, \quad y = P\bar{y},$$

$$\bar{y} = P^{-1}AP\bar{x} = B\bar{x}.$$

Takie przekształcenie macierzy A w B nazywane jest przekształceniem podobieństwa.

W nowym układzie odniesienia macierz B gwarantuje te same własności geometryczne, co A w poprzedniej bazie. Jakkolwiek kierunki własne ulegają zmianie, to wartości własne i wielomiany charakterystyczne pozostają te same.

Ogólnie, dwie macierze A i B tych samych wymiarów, nazywamy podobnymi, jeśli istnieje macierz P nieosobliwa i taka, że:

$$A = P^{-1}BP, \quad |P| \neq 0.$$

Jeśli istnieje również nieosobliwa macierz Q :

$$B = Q^{-1}CQ, \quad |Q| \neq 0,$$

wtedy:

$$A = P^{-1}Q^{-1}CQP = (QP)^{-1}CQP, \quad |QP| \neq 0.$$

Niech macierz kwadratowa $A[n \times n]$ ma różne wartości własne i związane z nimi wektory własne w_i . Macierz modalna W to macierz, w której kolumny tworzone są z wektorów własnych w_i macierzy A . Korzystając z tego, zapisujemy równanie (19.1) jako:

$$AW = W \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} = WJ.$$

Stąd można otrzymać też zapis

$$J = W^{-1}AW,$$

a macierz diagonalna J nazywa się macierzą Jordana podobną do macierzy A .

Nowa baza jest utworzona przez wektory własne. Jeżeli macierz A ma n różnych wartości własne, to przekształcenie podobieństwa za pomocą macierzy modalnej $T = W$ daje macierz diagonalną Jordana macierzy A . W jest tak zwaną prawostronną macierzą modalną, zbudowaną z prawostronnych wektorów własnych.

Można też tworzyć wektory własne lewostronne v_i

$$v_i^T A = \lambda_i v_i^T$$

i budować lewostronną macierz modalną V , której wiersze tworzą transformowane wektory własne lewostronne. W zapisie macierzowym zależność ma postać

$$VA = JV.$$

Łatwo pokazać, że $V = W^{-1}$.

$$AW = WJ \Rightarrow W^{-1}AW = W^{-1}WJ \Rightarrow W^{-1}AWW^{-1} = JW^{-1} \Rightarrow$$

$$\Rightarrow W^{-1}A = JW^{-1} \Rightarrow V = W^{-1} \Rightarrow V \cdot W = I.$$

Przykład 19.1

Dana jest macierz typu Frobeniusa $A = \begin{bmatrix} 0 & 1 \\ -2 & -2 \end{bmatrix}$;

$$|sI - A| = 0 \Rightarrow s_1 = -1 + j, \quad s_2 = -1 - j.$$

W przestrzeni rzeczywistej R^2 nie istnieją kierunki własne dla powyższej macierzy A . Prawostronna macierz modalna T dla postaci Frobeniusa ma postać macierzy Vandermonda. Stąd można ją zapisać bezpośrednio:

$$T = \begin{bmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1+j & -1-j \end{bmatrix}, \quad J = T^{-1}AT = \begin{bmatrix} -1+j & 0 \\ 0 & -1-j \end{bmatrix}.$$

Można to też wykazać poprzez inną postać podobną.

Stosując transformację $P = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$, można znaleźć macierz podobną B o tych samych wartościach własnych:

$$P^{-1}AP = B = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ -1 & -1 \end{bmatrix}, \quad \lambda = \alpha \pm j\beta = -1 \pm j.$$

Dla macierzy B jej macierz modalna ma postać $W = \begin{bmatrix} 1 & 1 \\ +j & -j \end{bmatrix}$.

Stąd diagonalna postać Jordana

$$J = W^{-1}BW = (PW)^{-1}A(PW) = \begin{bmatrix} -1+j & 0 \\ 0 & -1-j \end{bmatrix}.$$

Widać, że modalna macierz transformacji wektorów własnych dla macierzy A

$$T = PW = \begin{bmatrix} 1 & 1 \\ -1+j & -1-j \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{bmatrix}.$$

W celu rozszerzenia powyższych rozważań na przypadki wartości własnych wielokrotnych i podania reguł uzyskiwania ogólnych blokowo-diagonalnych postaci macierzowych Jordana podane zostaną pewne uzupełniające własności dla macierzy wielomianowych.

19.3. Macierze wielomianowe

Każda macierz kwadratowa $W(\lambda)$ o wymiarach $[n \times n]$, rzędu r , której elementami są wielomiany zmiennej λ , można przez przekształcenia elementarne sprowadzić do tzw. równoważnej postaci normalnej (diagonalnej) macierzy Smitha $S(\lambda)$ [19TU]

$$W(\lambda) = \begin{bmatrix} w_{11}(\lambda) & \cdots & w_{1n}(\lambda) \\ \cdots & \ddots & \cdots \\ w_{n1}(\lambda) & \cdots & w_{nn}(\lambda) \end{bmatrix}.$$

Oznacza to, że istnieją macierze wielomianowe $P(\lambda)$ i $Q(\lambda)$ o stałych (liczbowych) wyznacznikach różnych od 0 takie, że dla każdego λ ,

$$S(\lambda) = P(\lambda)W(\lambda)Q(\lambda),$$

gdzie

$$S(\lambda) = \begin{bmatrix} p_1(\lambda) & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ \cdots & & p_r(\lambda) & 0 \\ \cdots & 0 & 0 & 0 \end{bmatrix}_{n \times n}.$$

W szczególnie interesującym przypadku, gdy wielomiany w macierzy $W(\lambda)$ są pierwszego stopnia, macierze przekształcenia P i Q są stałe (twierdzenie Weierstrassa).

Wielomiany $p_i(\lambda)$ są zwane **czynnikami niezmienniczymi** i posiadają tę własność, że $p_i(\lambda)$ jest dzielnikiem wielomianu $p_{i+1}(\lambda)$ dla $i = 1, 2, \dots, r-1$, a współczynniki przy ich najwyższych potęgach są równe 1.

Każdy czynnik niezmienniczy jest iloczynem pewnej ilości tzw. **dzielników elementarnych**. Dzielniki elementarne są już wielomianami nierozikładalnymi w danym ciele liczbowym (tzn. wielomianami I stopnia lub II stopnia z $\Delta < 0$ w ciele liczb rzeczywistych i wielomianami I stopnia w ciele liczb zespolonych) podniesionymi do odpowiedniej potęgi. Stąd znajomość dzielników elementarnych i rzędu macierzy $W(\lambda)$ jest równoważna znajomości wszystkich czynników niezmienniczych.

Na przykład dana jest macierz wielomianowa $W(\lambda)$ o wymiarze $n = 12$ i rzędzie $r = 6$. Macierz Smitha może mieć postać

$$S(\lambda) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda-2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda-2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & (\lambda-2)^2(\lambda+3)^2 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & (\lambda-2)^2(\lambda+3)^3(\lambda+4) & \cdots & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{12 \times 12}$$

Występujące tu dzielniki elementarne są wielomianami pierwszego stopnia podniesionymi do pewnych potęg.

19.4. Macierz charakterystyczna i jej własności

Właściwości macierzy charakterystycznej przedstawiono poniżej.

- Dla liczbowej macierzy kwadratowej A , macierz wielomianowa $W(\lambda) = (\lambda I - A)$ nazywana jest macierzą charakterystyczną macierzy A , a jej wyznacznik $|W(\lambda)|$ wielomianem charakterystycznym.
- Pierwiastkami wielomianu charakterystycznego są wartości własne λ_i macierzy A . Wielomian charakterystyczny równa się też iloczynowi wszystkich czynników niezmienniczych występujących w równoważnej do $W(\lambda)$ postaci normalnej Smitha $S(\lambda)$.
- Macierz A spełnia swoje własne równanie charakterystyczne $|W(A)| = 0$. Ostatni czynnik niezmienniczy w postaci normalnej Smitha $S(\lambda)$ nazywa się wielomianem minimalnym $M(\lambda)$, gdyż jest to wielomian najniższego stopnia spełniający równanie $M(A) = 0$.
- Wstawiając wartość własną λ_i do macierzy $W(\lambda) = (\lambda I - A)$, otrzymuje się macierz liczbową o rzędzie $r_i < n$. Różnicę $d_i = n - r_i$ nazywa się spadkiem rzędu macierzy charakterystycznej $W(\lambda)$ dla danej wartości λ_i .

- Liczba wektorów własnych związań z daną λ_i równa się spadkowi rzędu d_i .
- Przy jednokrotnej wartości własnej λ_i spadek rzędu jest też jednokrotny, $d_i = n - r_i = 1$, istnieje więc jeden kierunek własny.
- Przy wielokrotnej wartości własnej λ_i o krotności m_i , spadek rzędu $d_i = n - r_i$, i spełniona jest nierówność $1 \leq d_i \leq m_i$.
- Liczba kierunków własnych dla wielokrotnej wartości własnej zawiera się od 1 do m_i . Gdy dla jakiejś λ_i występuje pełny spadek rzędu macierzy $(\lambda_i I - A)$, tzn. $d_i = m_i$, można wyznaczyć dla tego $\lambda_i m_i$ wektorów własnych. Przy jednokrotnym spadku $d_i = 1$, mimo m_i krotności λ_i występuje tylko jeden wektor własny.
- Liczba wektorów własnych dla danej λ_i , czyli spadek rzędu d_i , można odczytać również z postaci Smitha. Równa się on liczbie czynników niezmienniczych, w których dana λ_i występuje (np. 6-krotna wartość własna $\lambda_i = 2$ z powyższego przykładu występuje w czterech czynnikach niezmienniczych, czyli ma tylko cztery dzielniki elementarne i istnieją związane z nią cztery kierunki własne). Oznacza to, że przy pełnym spadku rzędu $d_i = m_i$, w postaci Smitha występuje m_i czynników niezmienniczych z dzielikami elementarnymi I stopnia, a przy jednokrotnym spadku rzędu, $d_i = 1$, musi wystąpić jeden czynnik niezmienniczy, zawierający jeden dzielnik elementarny w postęce m_i .

19.5. Sprowadzanie macierzy A do postaci Jordana

W przypadku gdy w macierzy $(\lambda_i I - A)$ dla każdej λ_i o krotności $m_i (i = 1, n)$ występuje pełny spadek rzędu $d_i = m_i$ (tzn. istnieją pojedyncze wartości własne lub wielokrotne z liniowymi dzielikami elementarnymi), to można znaleźć n liniowo niezależnych wektorów własnych. Zbudowana z nich macierz modalna T transformuje macierz A do diagonalnej postaci Jordana z wartościami własnymi na głównej przekątnej.

$$J = T^{-1}AT = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}.$$

W przypadku gdy w macierzy $(\lambda_i I - A)$ dla pewnej λ_i o krotności m_i występuje niepełny spadek rzędu, $d_i < m_i$, to można znaleźć tylko d_i liniowo niezależnych wektorów własnych związań z tą wartością λ_i . Wektory własne należy wtedy uzupełnić brakującymi wektorami głównymi w liczbie $m_i - d_i$. Zbudowana z wektorów własnych i głównych macierz modalna T transformuje A do postaci pseudodiagonalnej Jordana (blokowo diagonalnej).

Dla każdej wielokrotnej wartości własnej powstaje blok Jordana, o wymiarze $m_i \times m_i$, składający się z klatek Jordana. Liczba klatek jest równa spadkowi rzędu d_i dla danej λ_i (liczba dzielików elementarnych), a wymiar poszczególnych klatek jest równy poszczególnym potęgom dzielików.

Stąd dla przykładowej macierzy Smitha $[12 \times 12]$ (podrozdz. 19.3) postać Jordana będzie następująca

$$J = \begin{bmatrix} 2 & & & & & & & & & & & \\ 2 & 2 & & & & & & & & & & \\ & 2 & 1 & & & & & & & & & 0 \\ & 0 & 2 & 1 & & & & & & & & 0 \\ 0 & & 2 & 1 & 2 & & & & & & & \\ & & & 0 & 2 & & & & & & & \\ & & & & & 3 & 1 & 0 & & & & \\ & & & & & 0 & -3 & & & & & \\ & & & & & & -3 & 1 & 0 & & & 0 \\ 0 & & & & & & 0 & -3 & 1 & & & \\ & & & & & & 0 & 0 & -3 & & & \\ & & & & & & 0 & 0 & 0 & -3 & & \\ & & & & & & & & & & & -4 \end{bmatrix}_{12 \times 12}$$

Klatki Jordana mogą mieć różną postać w zależności od sposobu konstrukcji wektorów głównych. Standardowo przy prawostronnych wektorach własnych i głównych otrzymuje się się postać klatek z jedynkami nad główną przekątną.

$$J_i = \begin{bmatrix} \lambda_i & 1 & 0 & 0 \\ 0 & \lambda_i & 1 & 0 \\ 0 & 0 & \ddots & 1 \\ 0 & 0 & 0 & \lambda_i \end{bmatrix}.$$

Taka postać wynika z poniższego algorytmu liczenia wektorów własnych i głównych:

$$\begin{aligned} (A - \lambda_i I) w_{g1} &= 0, \\ (A - \lambda_i I) w_{g2} &= w_{g1}, \\ &\dots \\ (A - \lambda_i I) w_{gm} &= w_{g(m-1)}. \end{aligned}$$

Wektor główny w_{g1} jest równy wektorowi własnemu $w_{g1} = w_w$.

Przy wektorach lewostronnych jedynki wystąpią pod przekątną.

Poniżej przedstawimy trzy przypadki dla potrójnej wartości własnej.

Przykład 19.2

Dana jest macierz $[3 \times 3]$:

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad \det[A - \lambda I] = (2 - \lambda)^3 = 0 \Rightarrow \lambda_{1,2,3} = 2,$$

$\text{rząd } [A - 2I] = 0$, spadek rzędu $d = 3 \Rightarrow 3$ klatki Jordana.

Postać Smitha

$$S(\lambda) = \begin{bmatrix} \lambda - 2 & 0 & 0 \\ 0 & \lambda - 2 & 0 \\ 0 & 0 & \lambda - 2 \end{bmatrix}.$$

Występują tylko liniowe dzielniki elementarne, czyli trzy wektory własne

$(A - 2I)w = 0 \Rightarrow 0 \cdot w_1 + 0 \cdot w_2 + 0 \cdot w_3 = 0 \Rightarrow w_1, w_2, w_3$ – dowolne, np.:

$$T = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad T^{-1} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad T^{-1}AT = J = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Przykład 19.3

Dana jest macierz A [3×3]:

$$A = \begin{bmatrix} 0 & 1 & 0 \\ -4 & 4 & 0 \\ -2 & 1 & 2 \end{bmatrix}, \quad \det[A - \lambda I] = \lambda^3 - 6\lambda^2 + 12\lambda - 8 = (\lambda - 2)^3 = 0 \Rightarrow \lambda_{1,2,3} = 2,$$

rzad $[A - 2I] = 1$, spadek rzędu $d = 2 \Rightarrow 2$ klatki Jordana.

Postać Smitha

$$S(\lambda) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda - 2 & 0 \\ 0 & 0 & (\lambda - 2)^2 \end{bmatrix}.$$

Występuje jeden liniowy dzielnik elementarny i związany z nim wektor własny, oraz dzielnik elementarny w potędze 2 i związany z nim wektor własny i główny:

$$(A - 2I)w = 0 \Rightarrow \begin{cases} -2w_{11} + w_{12} = 0 \\ -4w_{11} + 2w_{12} = 0 \\ -2w_{11} + w_{12} = 0 \end{cases} \quad \begin{cases} -2w_{21} + w_{22} = 0 \\ -4w_{21} + 2w_{22} = 0 \\ -2w_{21} + w_{22} = 0 \end{cases} \quad \begin{cases} -2w_{g1} + w_{g2} = \gamma \\ -4w_{g1} + 2w_{g2} = 2\gamma \\ -2w_{g1} + w_{g2} = \gamma \end{cases}$$

Pierwszy wektor własny: $w_{13} = \alpha$, $w_{12} = \beta$, $w_{11} = \beta/2$.

Drugi wektor własny: $w_{23} = \gamma$, $w_{22} = k$, $w_{21} = k/2$.

Wektor główny do drugiego wektora własnego: $\gamma = k/2$, $w_{g3} = \delta$, $w_{g2} = \eta$, $w_{g1} = (\eta - \gamma)/2$.

Stąd macierz T ma np. postać:

$$T = \begin{bmatrix} \beta/2 & \gamma & (\eta - \gamma)/2 \\ \beta & 2\gamma & \eta \\ \alpha & \gamma & \delta \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 2 & 1 \\ 1 & 1 & 0 \end{bmatrix},$$

$$T^{-1} = \begin{bmatrix} -1 & 0 & 1 \\ 1 & 0 & 0 \\ -2 & 1 & 0 \end{bmatrix}, \quad T^{-1}AT = J = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}.$$

Przykład 19.4

Dana jest macierz A [3×3]:

$$A = \begin{bmatrix} 3 & -1 & -2 \\ -1 & 3 & 2 \\ 0 & -2 & 0 \end{bmatrix}, \quad \det[A - \lambda I] = \lambda^3 - 6\lambda^2 + 12\lambda - 8 = (\lambda - 2)^3 = 0 \Rightarrow \lambda_{1,2,3} = 2,$$

rzad $[A - 2I] = 2$, spadek rzędu $d = 1 \Rightarrow 1$ klatka Jordana.

Postać Smitha

$$S(\lambda) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & (\lambda - 2)^3 \end{bmatrix}.$$

Występuje jeden liniowy dzielnik elementarny w potędze 3 i związany z nim jeden wektor własny, oraz związane z nim dwa wektory główne:

$$(A - 2I)w = 0 \Rightarrow \begin{cases} w_1 - w_2 - 2w_3 = 0 \\ -w_1 + w_2 + 2w_3 = 0 \\ -2w_2 - 2w_3 = 0 \end{cases} \quad \left. \begin{array}{l} w_{g11} - w_{g12} - 2w_{g13} = \alpha \\ -2w_{g12} - 2w_{g13} = \alpha \end{array} \right\} \quad \left. \begin{array}{l} w_{g21} - w_{g22} - 2w_{g23} = \beta + \alpha/2 \\ -2w_{g22} - 2w_{g23} = \beta \end{array} \right\}$$

Wektor własny: $w_3 = \alpha$, $w_2 = -\alpha$, $w_1 = \alpha$,

Pierwszy wektor główny: $w_{g13} = \beta$, $w_{g12} = -\beta - \alpha/2$, $w_{g11} = \beta + \alpha/2$,

Drugi wektor główny: $w_{g23} = \delta$, $w_{g22} = -\delta - \beta/2$, $w_{g21} = \delta + (\beta + \alpha)/2$.

Stąd macierz T ma np. postać:

$$T = \begin{bmatrix} 2 & 1 & 1 \\ -2 & -1 & 0 \\ 2 & 0 & 0 \end{bmatrix}, \quad T^{-1} = \begin{bmatrix} 0 & 0 & 0,5 \\ 0 & -1 & -1 \\ 1 & 1 & 0 \end{bmatrix}, \quad T^{-1}AT = J = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}.$$

19.6. Własności wektorów własnych prawo- i lewostronnych

Znając wektory własne macierzy A , prawostronne w_i (kolumny macierzy W) i lewostronne v_i (wiersze macierzy $V = W^{-1}$), można za ich pomocą wyliczyć wrażliwość poszczególnych wartości własnych λ_i na małą zmianę dowolnego elementu a_{kj} macierzy A , tzn określić pochodną λ_i względem a_{kj} . Zależność taka może być przydatna w zadaniach syntez i analizy układów sterowania, gdy elementy macierzy stanu są niepewne lub ulegają zmianie. Założymy, że macierz $A [n \times n]$ ma n różnych wartości własnych, wtedy zachodzi:

$$Aw_i = \lambda_i w_i \quad (19.2)$$

$$\frac{\partial A}{\partial a_{kj}} w_i + A \frac{\partial w_i}{\partial a_{kj}} = \frac{\partial \lambda_i}{\partial a_{kj}} w_i + \lambda_i \frac{\partial w_i}{\partial a_{kj}}$$

gdzie:

$\frac{\partial A}{\partial a_{kj}}$ – macierz zerowa z jedynką w k -tym wierszu i j -tej kolumnie,

$\frac{\partial w_i}{\partial a_{kj}}$ – wektor kolumnowy,

$\frac{\partial \lambda_i}{\partial a_{kj}}$ – skalar.

Mnożąc równanie (19.2) obustronnie przez v_i^T otrzymuje się

$$v_i^T \frac{\partial A}{\partial a_{kj}} w_i + v_i^T A \frac{\partial w_i}{\partial a_{kj}} = v_i^T \frac{\partial \lambda_i}{\partial a_{kj}} w_i + v_i^T \lambda_i \frac{\partial w_i}{\partial a_{kj}} \quad (19.3)$$

Uwzględniając, że $v_i^T w_i = 1$ oraz $v_i^T (A - \lambda_i I) = 0$, z (19.3) otrzymuje się

$$v_i^T \frac{\partial A}{\partial a_{kj}} w_i = \frac{\partial \lambda_i}{\partial a_{kj}} \quad (19.4)$$

Ponieważ zachodzi

$$v_i^T \frac{\partial A}{\partial a_{kj}} w_i = v_{ik} w_{ij} \quad (19.5)$$

(v_{ik} – k -ty element wektora v_i , w_{ij} – j -ty element wektora w_i), to ostatecznie

$$\frac{\partial \lambda_i}{\partial a_{kj}} = v_{ik} w_{ij} \quad (19.6)$$

Wrażliwość (pochodna cząstkowa) i -tej wartości własnej λ_i na zmianę elementu a_{kj} macierzy A jest wyliczana jako iloczyn dwóch liczb – k -tej składowej i -tego wektora lewostronnego i j -tej składowej i -tego wektora prawostronnego. Zapis macierzowy dla wszystkich elementów a_{kj} z wykorzystaniem iloczynu diadycznego całych i -tych wektorów własnych ma postać

$$\frac{\partial \lambda_i}{\partial A} = v_i^T w_i^T \quad (19.7)$$

Na przykład dla przypadku dwuwymiarowego, oznaczając pierwszym wskaźnikiem numer wartościowej, otrzymujemy następujące macierze modalne:

$$W = \begin{bmatrix} w_{11} \\ w_{12} \end{bmatrix} \begin{bmatrix} w_{21} \\ w_{22} \end{bmatrix}, \quad V = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix},$$

to dla λ_1

$$v_1^T w_1^T = \begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} \begin{bmatrix} w_{11} & w_{12} \end{bmatrix} = \begin{bmatrix} \frac{\partial \lambda_1}{\partial a_{11}} & \frac{\partial \lambda_1}{\partial a_{12}} \\ \frac{\partial \lambda_1}{\partial a_{21}} & \frac{\partial \lambda_1}{\partial a_{22}} \end{bmatrix} = \frac{\partial \lambda_1}{\partial A},$$

a dla λ_2

$$v_2^T w_2^T = \begin{bmatrix} v_{21} \\ v_{22} \end{bmatrix} \begin{bmatrix} w_{21} & w_{22} \end{bmatrix} = \begin{bmatrix} \frac{\partial \lambda_2}{\partial a_{11}} & \frac{\partial \lambda_2}{\partial a_{12}} \\ \frac{\partial \lambda_2}{\partial a_{21}} & \frac{\partial \lambda_2}{\partial a_{22}} \end{bmatrix} = \frac{\partial \lambda_2}{\partial A}.$$

Sumaryczne zmiany λ_i przy zmianach wszystkich wyrazów a_{kj} o Δ_{kj} wynikną z sumy wszystkich iloczynów Δ_{kj} i elementów macierzy $w_1 v_1^T$.

Ponieważ $V \cdot W = I$, to iloczyny skalarne v_i i w_i są jednostkowe i dla każdej wartościowej własnej λ_i zachodzi

$$v_i^T w_i = [v_{i1} \ v_{i2} \ \dots \ v_{in}] \begin{bmatrix} w_{11} \\ w_{12} \\ \vdots \\ w_{in} \end{bmatrix} = \sum_{k=1}^n v_{ik} w_{ik} = \sum_{k=1}^n \frac{\partial \lambda_i}{\partial a_{kk}} = 1 \text{ dla każdej wartości } i \in [1, n].$$

Czyli ślad macierzy (19.7), $[\partial \lambda_i / \partial A]$, jest jedynką.

Ponieważ również $W \cdot V = I$, to dla każdego elementu a_{kk} zachodzi

$$[v_{1k} \ v_{2k} \ \dots \ v_{nk}] \begin{bmatrix} w_{1k} \\ w_{2k} \\ \vdots \\ w_{nk} \end{bmatrix} = \sum_{i=1}^n v_{ik} w_{ik} = \sum_{i=1}^n \frac{\partial \lambda_i}{\partial a_{kk}} = 1 \text{ dla każdej wartości } k \in [1, n].$$

19.7. Ogólne własności macierzy

Poniżej zebrane podstawowe własności macierzy.

- 1) Iloczyn wartości własnych macierzy A jest równy jej wyznacznikowi

$$|A| = \det[A] = \lambda_1 \cdot \lambda_2 \cdot \dots \cdot \lambda_n = \prod_{i=1}^n \lambda_i.$$

Macierz jest osobliwa, jeżeli przynajmniej jedna $\lambda_i = 0$.

- 2) Zachodzi nierówność $\det[A] \leq \sum_{i=1}^n a_{ii} = a_{11} \cdot a_{22} \cdots \cdot a_{nn}$.

- 3) Suma elementów a_{ii} na głównej przekątnej A nazywana jest śladem macierzy (*trace*) i równa się sumie wartości własnych:

$$\text{trace}[A] = \sum_{i=1}^n a_{ii} = \lambda_1 + \lambda_2 + \dots + \lambda_n.$$

- 4) Ślad macierzy ma poniższe własności:

$$\text{trace}[A] = \text{trace}[A^T], \quad \text{trace}[A + B] = \text{trace}[A] + \text{trace}[B],$$

$$\text{trace}(AB) = \text{trace}(BA) = \text{trace}(A^T B^T) = \text{trace}(B^T A^T),$$

$$\text{trace}[xy^T] = \text{trace}[x^T y] = x^T y \text{ (skalar).}$$

- 5) Nierówność Sylvestera: $\text{rzad } A + \text{rzad } B - n \leq \text{rzad } (AB) \leq \min(\text{rzad } A, \text{rzad } B)$.
 6) Wartości własne kwadratowej macierzy symetrycznej $A = A^T$ są liczbami rzeczywistymi, a jej wektory własne w_i są ortonormalne i zawsze istnieje dla A diagonalna postać Jordana.

$$w_i^T w_i = 1, \quad w_i^T w_j = 0.$$

- 7) Dla dowolnej macierzy A , macierz AA^T jest kwadratowa i symetryczna.
 8) Jeśli $A^T = -A$, to A jest macierzą skośno symetryczną z zerami na przekątnej.
 9) Każdą macierz kwadratową można rozłożyć na część symetryczną $(A+A^T)/2$ i skośno symetryczną $(A-A^T)/2$.
 10) Jeśli A ma wartości własne λ_i to A^n będzie miała wartości własne λ_i^n i te same wektory własne. Ogólnie A^2 ma inne wartości własne niż $A^T A$.
 11) Jeśli nieosobliwa macierz A ma wartość własną λ , to A^{-1} ma wartość własną λ^{-1} i te same wektory własne

$$Aw = \lambda w \Rightarrow A^{-1}Aw = A^{-1}\lambda w \Rightarrow \lambda^{-1}w = A^{-1}w.$$

- 12) Dla macierzy A i B zachodzą związki:

$$(AB)^T = B^T A^T, \quad (A+B)^T = A^T + B^T, \quad (A^{-1})^{-1} = A,$$

$$(A^{-1})^n = (A^n)^{-1}, \quad ((A^{-1})^T = (A^T)^{-1}, \quad \det(A^{-1}) = (\det A)^{-1},$$

$$\det(A^T) = \det(A), \quad \det(AB) = \det(A)\det(B), \quad (A^T)^n = (A^n)^T.$$

Jeśli A i B są kwadratowe i nieosobliwe, to: $(AB)^{-1} = B^{-1}A^{-1}$.

- 13) Jeżeli A i B są symetryczne, to $A+B$ jest symetryczna, ale AB niekoniecznie.

- 14) Jeżeli $A^T = A^{-1}$, to macierz A nazywa się unitarna i jest ortonormalna $A^T A = I$ oraz $|A| = 1$.

- 15) $\log(AB) = \log A + \log B \Leftrightarrow AB = BA$.

- 16) $\exp(A+B) = \exp(A) \cdot \exp(B) \Leftrightarrow AB = BA$.

- 17) Pochodna (całka) macierzy jest macierzą pochodnych (całek) każdego elementu.

$$\frac{d}{dt} A(t) = \left[\frac{da_{i,j}}{dt} \right].$$

- 18) Pochodna macierzy odwrotnej wyraża się wzorem

$$\frac{d}{dt} A^{-1}(t) = -A^{-1}(t) \frac{dA(t)}{dt} A^{-1}(t).$$

Wynika to z prostej zależności

$$\frac{d}{dt} [A^{-1}(t)A(t)] = 0 = \frac{d}{dt} [A^{-1}]A + A^{-1} \frac{dA}{dt}.$$

- 19) Twierdzenie Cayleya–Hamiltona:

Każda macierz kwadratowa A spełnia swoje własne równanie charakterystyczne

$$\det(sI - A) = a_n s^n + a_{n-1} s^{n-1} + \dots + a_1 s + a_0.$$

To znaczy, że dla $a_n = 1$ zachodzi

$$A^n + a_{n-1} A^{n-1} + a_{n-2} A^{n-2} + \dots + a_1 A + a_0 I = 0.$$

Twierdzenie to można wykorzystać do liczenia macierzy odwrotnej

$$a_0 I = - \left[A^n + a_{n-1} A^{n-1} + \dots + a_1 A \right],$$

$$A^{-1} = -\frac{1}{a_0} \left[A^{n-1} + a_{n-2} A^{n-2} + \dots + a_1 I \right].$$

Jeśli nie są znane współczynniki a_i wielomianu charakterystycznego, to można je też wyliczyć z formuł:

$$\begin{aligned} a_{n-1} &= -T_1, \\ a_{n-2} &= -(a_{n-1}T_1 + T_2)/2, \\ a_{n-3} &= -(a_{n-2}T_1 + a_{n-1}T_2 + T_3)/3, \\ &\dots \\ a_0 &= -(a_1T_1 + a_2T_2 + \dots + a_{n-1}T_{n-1} + T_n)/n, \\ T_i &= \text{trace}(A^i). \end{aligned}$$

20) Macierz A jest dodatnio określona $\Leftrightarrow x^T Ax > 0$ dla \forall niezerowego x i wszystkie $\lambda_i > 0$.

Macierz A jest dodatnio półokreślona $\Leftrightarrow x^T Ax \geq 0$ dla \forall niezerowego x i wszystkie $\lambda_i \geq 0$.

Macierz A jest ujemnie określona $\Leftrightarrow x^T Ax < 0$ dla \forall niezerowego x i wszystkie $\lambda_i < 0$.

Macierz A jest ujemnie półokreślona $\Leftrightarrow x^T Ax \leq 0$ dla \forall niezerowego x i wszystkie $\lambda_i \leq 0$.

21) Zachodzą zależności:

$$\begin{aligned} \frac{\partial}{\partial x} [y^T x] &= \frac{\partial}{\partial x} [x^T y] = y, \quad \frac{\partial}{\partial x} [y^T Ax] = A^T y, \quad \frac{\partial}{\partial x} [Ax] = A, \\ \frac{\partial}{\partial x} [x^T Ax] &= Ax + A^T x = (A^T + A)x = 2Ax \quad (\text{gdy } A = A^T), \\ \frac{\partial}{\partial x^2} [x^T Ax] &= A + A^T = 2A \quad (\text{gdy } A = A^T). \end{aligned}$$

Obszerny zestaw reguł działania na macierzach blokowych można znaleźć w [19KAT].

19.8. Rozwiązywanie algebraicznego układu liniowego

Dany jest macierzowy układ równań liniowych z zadanym wektorem y i macierzą C oraz nieznanym wektorem x .

$$y = Cx \tag{19.8}$$

Można rozważyć trzy przypadki:

1) Jeżeli $y \in R^n, x \in R^n, C [n \times n]$ – macierz nieosobliwa, to

$$x = C^{-1}y.$$

2) Jeżeli $y \in R^m, x \in R^n, m < n$, to $C [m \times n]$ jest macierzą poziomo-prostokątną, rząd $[C] = k \leq m$.

Jeżeli układ jest niesprzeczny (z twierdzenia Kroneckera–Capelliego \Rightarrow rząd $[Cy] = \text{rząd } [C]$), to ze względu na to, że jest za mało równań, a za dużo niewiadomych, istnieje nieskończona liczba rozwiązań x_i spełniających (19.8), czyli $n-k$ zmiennych można przyjąć dowolnie (np. przyjąć, że równe są zero – rozwiązanie bazowe).

Często w zastosowaniach automatyki interesujące może być rozwiązanie będące najkrótszym wektorem x , prostopadłym do hiperpowierzchni (19.8), czyli rozwiązanie z minimalną normą.

Rozwiązanie to musi wtedy spełniać równanie (19.8) z macierzą $P [n \times m]$

$$x = Py \tag{19.9}$$

Z (19.8) musi zachodzić $Py = PCx \Rightarrow PC = I [n \times n]$.

Aby znaleźć odpowiednią macierz P i rozwiązanie x , można zastosować metodę mnożników Lagrange'a. Pomóżmy (19.9) lewostronnie przez C

$$Cx = CPy = y \Rightarrow CP = I [m \times m] \tag{19.10}$$

Wprowadza się hamiltonian w postaci kwadratu normy rozwiązania i kary za odchyłkę od zależności (19.10). Hamiltonian ten należy zminimalizować

$$J = x^T x + [y - CPy]^T \lambda.$$

Wektor mnożników Lagrange'a $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_n]^T$ jest nieznany.

$$J = [Py]^T Py - [Py]^T C^T \lambda + y^T \lambda.$$

Zadanie $\min \|x\|$ przy spełnieniu warunku (19.8) znajdujemy z warunku koniecznego minimum

$$\frac{\partial J}{\partial x} = \frac{\partial J}{\partial [Py]} = 2Py - C^T \lambda = 0 \tag{19.11}$$

Równanie mnożymy obustronnie przez C i wykorzystujemy zależność $CP = I$.

$$2y = CC^T \lambda \Rightarrow \lambda = 2[CC^T]^{-1}y.$$

Stąd z (19.11) otrzymuje się wzór $2Py - C^T \lambda = 2Py - 2C^T[CC^T]^{-1}y = 0$ i ostatecznie otrzymuje się poszukiwaną postać macierzy P

$$P = C^T[CC^T]^{-1}.$$

Daje ona wg (19.9) minimalnonormowe rozwiązanie na x

$$x = Py = C^T[CC^T]^{-1}y.$$

Macierz P nazywana jest prawą pseudoodwrotnością macierzy C .

Przykład 19.5

Dane jest równanie:

$$8 = x_1 + 4x_2 = [1 \ 4] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = Cx,$$

$$x_{\min} = \begin{bmatrix} 1 \\ 4 \end{bmatrix} \left[\begin{bmatrix} 1 & 4 \\ 4 & 1 \end{bmatrix} \right]^{-1} \cdot 8 = \begin{bmatrix} 1 \\ 4 \end{bmatrix} \cdot \frac{1}{17} \cdot 8 = \begin{bmatrix} 0,4705 \\ 1,8823 \end{bmatrix},$$

$$P = [1/17 \ 4/17]^T, \quad \lambda = 16/17, \quad J = (8^2 + 32^2)/17^2 = 3,7647,$$

$$\|x_{\min}\| = \sqrt{x^T x} = 1,9403.$$

Znalezione rozwiązanie x_{\min} jest prostopadłe do prostej ograniczającej

$$x_2 = -0,25x_1 + 2.$$

- 3) Jeżeli $y \in R^m$, $x \in R^n$, $m > n$, to macierz $C [m \times n]$ jest macierzą pionowo-prostokątną. Ze względu na to, że jest za dużo równań w stosunku do liczby niewiadomych, z reguły układ (19.8) jest sprzeczny. W przeciwnym przypadku $m=n$ równań liniowo zależnych można odrzucić, przyjąć kwadratową, nieosobliwą podmacierz $[n \times n]$ i zastosować przypadek A.
Jeśli układ jest sprzeczny, można zastosować technikę minimalnokwadratową. Rozwiązywanie minimalizuje sumę kwadratów błędów obu stron równania.

$$\varepsilon = y - Cx \Rightarrow J = \varepsilon^T \varepsilon = (y - Cx)^T (y - Cx) = y^T y - 2y^T Cx + x^T C^T Cx.$$

Zadanie: \min_J rozwiązuje się z warunku koniecznego:

$$\frac{\partial J}{\partial x} = -2C^T y + 2C^T Cx = 0,$$

$$x^o = [C^T C]^{-1} C^T y = Ly.$$

Jest to dobrze znany wzór na estymator minimalnokwadratowy w metodzie najmniejszych kwadratów (MNK), wyprowadzony przez Gaussa w 1809 r. i stosowany przez niego do obliczeń astronomicznych. Macierz $L = [C^T C]^{-1} C^T$ nazywana jest **lewą pseudoodwrotnością** macierzy C , a kwadratowa symetryczna $[n \times n]$ macierz $[C^T C]$ nazywa się macierzą regresji liniowej. Powinna oczywiście być nieosobliwa. Zarówno macierz pseudoodwrotna P , jak i L mają następującą własność:

$$CPC = C \quad \text{i} \quad CLC = C.$$

Przykład 19.6

Dane jest równanie macierzowe:

$$y = \begin{bmatrix} 6 \\ 2 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 & 6 \\ 1 & 1 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \Rightarrow C^T C = \begin{bmatrix} 3 & 11 \\ 11 & 53 \end{bmatrix}, \quad [C^T C]^{-1} = \frac{1}{38} \begin{bmatrix} 53 & -11 \\ -11 & 3 \end{bmatrix},$$

$$L = \frac{1}{38} \begin{bmatrix} -13 & 42 & 9 \\ 7 & -8 & 1 \end{bmatrix},$$

$$x^o = \begin{bmatrix} \frac{30}{19} \\ \frac{19}{16} \\ \frac{16}{19} \end{bmatrix}, \quad Cx^o = \left[6 \frac{12}{19}, \quad 2 \frac{8}{19}, \quad 4 \frac{18}{19} \right]^T \neq y, \quad \min J = \sum_{i=1}^3 \varepsilon_i^2 = 1 \frac{3}{19}.$$

19.9. Własności formy kwadratowej

Iloczyn skalarny wektora $x \in R^n$ i wektora $Ax \in R^m$, gdzie A jest rzeczywistą macierzą $[n \times n]$, nazywa się formą kwadratową i jest oznaczany jako

$$\langle x, Ax \rangle = x^T Ax = \sum_{i=1}^n a_{ii} x_i x_j.$$

Tworzy on wielomian drugiego stopnia n zmiennych x_i . Macierz A jest z reguły symetryczna.

Iloraz Rayleigha to stosunek dwóch liczb

$$\frac{x^T Ax}{x^T x} = R(x).$$

Wartość ilorazu Rayleigha nie zależy od długości wektora x i jest stała na danym kierunku.

Dla wektora kolinearnego $y = ax$, gdzie $a \in R$

$$R(y) = \frac{a^2 x^T Ax}{a^2 x^T x} = R(x).$$

Stąd własności $R(x)$ rozważa się na sferze jednostkowej $x^T x = 1$.

Ekstremalizacja formy kwadratowej

Niech $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ będą rzeczywistymi wartościami własnymi A , a w_1, w_2, \dots, w_n odpowiednimi znormalizowanymi wektorami własnymi A .

Stąd zachodzą związki:

$$Aw_i = \lambda_i w_i \Rightarrow w_i^T Aw_i = \lambda_i w_i^T w_i$$

$$\lambda_i = w_i^T Aw_i \text{ dla sfery } \|w_i\| = 1 \quad (19.12)$$

Twierdzenie 19.1

Forma kwadratowa osiąga swoje ekstremum dla kierunku wektora własnego macierzy A .

Dowód

Warunek konieczny ekstremum formy kwadratowej na sferze jednostkowej:

$$\frac{\partial}{\partial x} \left[x^T Ax + (1 - x^T x) \lambda \right] = 0, \quad 2Ax - 2\lambda x = 0 \Rightarrow Ax = \lambda x.$$

Stąd ekstremalnym kierunkiem jest kierunek wektora własnego, a mnożnik Lagrange'a jest wartością własną. \square

Ponieważ $\lambda_1 = \max_i \{\lambda_i\}$, a $\lambda_n = \min_i \{\lambda_i\}$, to zgodnie z (19.12):

$$\lambda_1 = w_1^T Aw_1 = \max_{\|x\|=1} R(x), \quad \lambda_n = w_n^T Aw_n = \min_{\|x\|=1} R(x).$$

Stąd ważne oszacowanie dla wartości formy

$$\lambda_1 \geq \langle x, Ax \rangle \geq \lambda_n \text{ dla } \|x\| = 1.$$

Powyższa nierówność pozwala na szybką ocenę rozrzutu krańcowych wartości własnych λ_1, λ_n przez wykonanie kilku numerycznych eksperymentów obliczenia wartości $x^T Ax$, dla kilku różnych x ze sfery jednostkowej.

Definiując zbiór S_2 jako podzbiór sfery $x^T x = 1$, składający się z wektorów prostopadłych do wektora własnego w_1 związanego z maksymalną wartością własną λ_1 (koło wielkiej sfery, do którego płaszczyzny w_1 jest prostopadły), $S_2 = \{x: x^T x = 1, w_1^T x = 0\}$, można wykazać, że dla zbioru S_2 wektorem x maksymalizującym $R(x)$ jest drugi w kolejności wektor własny $w_2 \leq w_1$.

$$\lambda_2 = \max_{x \in S_2} (x^T Ax) = w_2^T Aw_2.$$

Budując kolejno zawężone podzbiory sfery jednostkowej, otrzymujemy następujące zależności:

$$S_k = \{x: x^T x = 1, w_j^T x = 0; j = 1, 2, \dots, k-1\},$$

$$\lambda_k = \max_{x \in S_k} (x^T Ax) = w_k^T Aw_k.$$

Właściwości formy kwadratowej dla niesymetrycznej kwadratowej macierzy rzeczywistej A

$$Aw_i = \lambda_i w_i \Rightarrow w_i^T A^T = \lambda_i w_i^T.$$

Dla symetryzacji macierzy mnożymy obie strony równania przez Aw_i

$$w_i^T A^T Aw_i = \lambda_i w_i^T Aw_i = \lambda_i^2 w_i^T w_i \Rightarrow \frac{w_i^T A^T Aw_i}{w_i^T w_i} = \lambda_i^2.$$

Macierz $A^T A$ jest symetryczna o wartościach własnych rzeczywistych dodatnich μ_i .

$$\mu_1 \geq \mu_2 \geq \dots \geq \mu_n \geq 0.$$

Z poprzednich rozważań wynika, że:

$$\mu_1 \geq \frac{w_i^T A^T Aw_i}{w_i^T w_i} \geq \mu_n,$$

$$\mu_1 \geq \lambda_i^2 \geq \mu_n \Rightarrow \sqrt{\mu_1} \geq |\lambda_i| \geq \sqrt{\mu_n} \text{ dla } i = 1, 2, \dots, n.$$

Wszystkie wartości własne λ_i macierzy A (która nie musi być symetryczna), spełniają powyższą nierówność.

Iloraz Rayleigha służy do zdefiniowania **normy spektralnej** prostokątnej macierzy A .

$$\|A\|_S = \max_{\|x\| \neq 0} \sqrt{\frac{x^T A^T Ax}{x^T x}} = \max_{i=1..n} \sqrt{\mu_i} = \sqrt{\mu_1},$$

gdzie μ_i są wartościami własnymi kwadratowej macierzy $A^T A$.

Uwaga

Promień spektralny kwadratowej macierzy A definiuje wzór

$$r(A) = \max_{i=1..n} |\lambda_i| = |\lambda_1|.$$

Dana jest macierz prostokątna A o wymiarach $[m \times n]$ i rzędzie r . Wartości

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0 = \sigma_{r+1} = \dots = \sigma_n,$$

będące dodatnimi pierwiastkami kwadratowymi z wartości własnych macierzy kwadratowej symetrycznej $A^T A [n \times n]$

$$\sigma_i = \sqrt{\mu_i},$$

nazywają się wartościami szczególnymi lub singularnymi macierzy A (*singular values*).

Dla każdej macierzy prostokątnej $A \in \mathbb{R}^{m \times n}$ istnieją kwadratowe macierze unitarne $U \in \mathbb{R}^{m \times m}$ i $V \in \mathbb{R}^{n \times n}$ ($U^T = U^{-1}$, $V^T = V^{-1}$) takie, że:

$$A = U \Sigma V^T,$$

$$\sum_{m \times n} = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \sigma_m & 0 \end{bmatrix}, \quad m < n \quad \text{lub} \quad \sum_{n \times m} = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma_n \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad m > n.$$

Jeśli A jest macierzą kwadratową i symetryczną $A = A^T$, to $\sigma_i = |\lambda_i|$ dla $i = 1, 2, \dots, r$.

Przykład 19.7

Macierz A ma podwójną wartość własną $\lambda_{1,2} = -1$, niezależną od parametru ε :

$$A = \begin{bmatrix} -1 & 0 \\ 1/\varepsilon & -1 \end{bmatrix}.$$

$$\text{Macierz } A^T A = \begin{bmatrix} 1+1/\varepsilon^2 & -1/\varepsilon \\ -1/\varepsilon & 1 \end{bmatrix},$$

ma wielomian charakterystyczny $\lambda^2 - \lambda(2+1/\varepsilon^2) + 1 = 0$ i dwa pierwiastki:

$$\mu_1 = 1 + \frac{1}{2\varepsilon^2} + \frac{1}{2\varepsilon^2}\sqrt{1+4\varepsilon^2}, \quad \mu_2 = 1 + \frac{1}{2\varepsilon^2} - \frac{1}{2\varepsilon^2}\sqrt{1+4\varepsilon^2}.$$

Jeśli ε dąży do zera, $\varepsilon \rightarrow 0$, to wartości własne $A^T A$: $\mu_1 \rightarrow \infty$, $\mu_2 \rightarrow 0$.

Wartości singularne macierzy A :

$$\sigma_1 = \sqrt{\mu_1}, \quad \sigma_2 = \sqrt{\mu_2},$$

pokazują złe uwarunkowanie macierzy A , jeśli $\varepsilon \rightarrow 0$. Wtedy $\lim \sigma_1 \rightarrow \infty$, $\lim \sigma_2 \rightarrow 0$. Nie widać tego po wartościach własnych A , $\lambda_{1,2} = -1$.

20. Aneks 2.

Podstawowe pojęcia analizy funkcjonalnej

Przegląd podstawowych pojęć analizy funkcjonalnej opracowano na podstawie prac [20LUD; 20KUJ; 20PEA; 20KOA; 20KUS].

- 1) Przestrzeń metryczna (X, ρ) to zbiór elementów X wraz ze zdefiniowaną metryką $\rho(x, y)$, czyli miarą odległości między dwoma dowolnymi elementami x, y :

$$\rho(x, y) = 0 \Leftrightarrow x = y,$$

$$\rho(x, y) = \rho(y, x),$$

$$\rho(x, z) \leq \rho(x, y) + \rho(y, z).$$

- 2) Ciąg elementów przestrzeni metrycznej nazywa się ciągiem Cauchy'go, jeśli zachodzi dla niego warunek $\lim_{k,l \rightarrow \infty} \rho(x_k, x_l) = 0$.
- 3) Przestrzeń metryczna (X, ρ) nazywa się zupełną, jeśli każdy ciąg Cauchy'go $\{x_k\} \subset X$ posiada granicę $x \in X$.
- 4) W przestrzeni metrycznej zupełnej X , zbiór $A \subset X$ nazywa się zwartym, jeżeli każdy ciąg jego elementów $x_n \in A$ zawiera w sobie podciąg zbieżny do granicy $\hat{x} \in A$.
- a) Każdy zbiór zwarty jest ograniczony i domknięty, tzn.

$$\exists M > 0 : \forall x, y \in A : \rho(x, y) \leq M,$$

np. odcinek liczb rzeczywistych $[0, 1]$ jest zbiorem zwartym.

- b) Nie każdy zbiór ograniczony i domknięty jest zwarty.

- 5) Zbiór X nazywa się przestrzenią liniową, jeżeli suma każdych dwóch jego elementów x, y też należy do X : $x + y \in X$ i iloczyn $\alpha x \in X$ dla $\forall \alpha \in R$. Działania dodawania i mnożenia muszą spełniać osiem aksjomatów – patrz [20KOA]. Przestrzeń musi zawierać element zerowy.

- a) Elementy x_1, \dots, x_n przestrzeni liniowej X są liniowo niezależne \Leftrightarrow gdy zachodzi implikacja

$$\sum_{i=1}^n \lambda_i x_i = 0 \Rightarrow \forall \lambda_i = 0.$$

- b) Jeżeli każdy zbiór elementów liniowo niezależnych zawiera co najwyżej n elementów, to przestrzeń jest n -wymiarowa.
- c) Dla n liniowo niezależnych elementów x_1, \dots, x_n , zbiór ich wszystkich kombinacji liniowych tworzy podprzestrzeń liniową n -wymiarową

$$X_o = \left\{ y : y = \sum_{i=1}^n \lambda_i x_i : \lambda_i \in R \right\}.$$

d) Przykłady przestrzeni liniowych:

- przestrzeń euklidesowa R^n (skończenie wymiarowa);
 - przestrzeń funkcji ciągłych na odcinku $[0, 1]$: $C[0, 1]$ (nieskończona wymiarowa);
 - zbiór wszystkich funkcji $\sin(nt)$ i $\cos(nt)$, $n \in N$ w przestrzeni $C[0, 1]$ jest liniowo niezależny.
- 6) Przestrzeń unormowana $(X, \| \cdot \|)$ to przestrzeń liniowa, w której każdemu elementowi x przypisano jego długość, czyli normę $\|x\|$. Norma to funkcjonał spełniający trzy aksjomaty:

$$\|x\| = 0 \Leftrightarrow x = 0,$$

$$\|\alpha x\| = |\alpha| \cdot \|x\| : \forall \alpha \in R, x \in X,$$

$$\|x + y\| \leq \|x\| + \|y\| : \forall x, y \in X.$$

W każdej przestrzeni unormowanej można za pomocą normy zdefiniować metrykę i stworzyć przestrzeń metryczną $\rho(x_1, x_2) = \|x_1 - x_2\|$.

7) Unormowana przestrzeń metryczna i zupełna nazywa się przestrzenią Banacha.

- a) Jeśli w przestrzeni R^n wprowadzi się normę $\|x\| = \sqrt{\sum_{i=1}^n x_i^2}$, przestrzeń ta będzie przestrzenią Banacha.
- b) Jeśli w liniowej przestrzeni funkcji ciągłych ograniczonych $C[0, 1]$ wprowadzimy jako normę ich amplitudę $\|x\|_c = \max_{t \in [0, 1]} |x(t)|$, to przestrzeń ta będzie przestrzenią Banacha.
Z inną normą $C[0, 1]$ może już nie być przestrzenią zupełną.
- c) Jeśli w liniowej przestrzeni funkcji całkowalnych w sensie Lebesgue'a z p -tą potegą $L^p(0, 1)$ wprowadzamy normę

$$\|x\|_{L^p} = \left[\int_0^1 |x(t)|^p dt \right]^{\frac{1}{p}} : 1 \leq p < \infty,$$

to przestrzeń ta będzie przestrzenią Banacha. Pojedynczym elementem jest tu klasa funkcji różniących się tylko na zbiorach miary zero, czyli w przeliczalnej ilości punktów, bo dla nich wartość całki jest ta sama.

W przestrzeni Banacha zachodzi nierówność Minkowskiego

$$\|x + y\|_p \leq \|x\|_p + \|y\|_p.$$

Zachodzi również zawieranie

$$1 \leq p \leq q \leq \infty; C[0, 1] \subset L^\infty(0, 1) \subset L^q(0, 1) \subset L^p(0, 1) \subset L^1(0, 1).$$

$C[0, 1]$ jest domkniętą podprzestrzenią $L^\infty(0, 1)$.

d) Liniowa przestrzeń funkcji mieralnych w sensie Lebesgue'a i istotnie ograniczonych $L^\infty(0, 1)$ z normą równą kresowi dolnemu supremów wszystkich funkcji równoważnych, składających się na dany element X (istotne supremum), jest przestrzenią Banacha.

$$\|x\| = \text{ess sup}_{t \in [0, 1]} |x(t)| = \inf_{E \subset [0, 1]; t \in [0, 1] \setminus E} \sup_{t \in E} |x(t)|, \text{ gdzie miara } E = 0.$$

e) Liniowa przestrzeń funkcji posiadających $(m-1)$ pochodną absolutnie ciągłą, a m -tą pochodną należącą do $L^p(0, 1)$, oznaczana jako przestrzeń Sobolewa $W_m^p(0, 1)$ z normą

$$\|x\| = \left[\int_0^1 |x(t)|^p dt + \int_0^1 |\dot{x}(t)|^p dt + \dots + \int_0^1 |x^{(m)}(t)|^p dt \right]^{\frac{1}{p}},$$

jest przestrzenią Banacha.

f) Przestrzeń funkcji x o wahaniu ograniczonym, lewostronne ciągłych w $[0, 1]$ i $x(0) = 0$ oznaczana $NBV[0, 1]$, z normą równą wahaniu x , $\|x\| = V_0^1[x]$, jest przestrzenią Banacha.

Funkcja x zadana na odcinku $[0, 1]$ nazywa się funkcją o wahaniu ograniczonym wtedy i tylko wtedy, gdy $\exists \alpha > 0$ takie, że dla każdego podziału $[0, 1]$: $0 = t_0 < t_1 < t_2 \dots < t_n = 1$ zachodzi

$$\sum_{i=1}^n |x(t_i) - x(t_{i-1})| \leq \alpha.$$

Wahaniem funkcji x nazywamy kres górnego sum, po wszystkich możliwych podziałach. Jeżeli x składa się z funkcji skokowej i funkcji absolutnie ciągły, to

$$V_0^1[x] = \sum_{i=0}^{\infty} |p_i| + \int_0^1 |g(\tau)| dt.$$

8) Iloczyn skalarny to odwzorowanie, które parze elementów x, y przyporządkowuje liczbę $(x, y) : X \times X \rightarrow R$ i dla $\forall x, y, z \in X$ spełnia cztery aksjomaty:

- 1) $(x, y) = (y, x)$;
- 2) $(\alpha x, y) = \alpha(x, y), \forall \alpha \in R$;
- 3) $(x + z, y) = (x, y) + (z, y)$;
- 4) $(x, x) = 0$ dla $x = 0$ i $(x, x) > 0$ dla $x \neq 0$.

a) Przykłady iloczynów skalarnych:

- $X = \mathbb{R}^n; (x, y) = x^T y;$
- $X = L^2(0, 1); (x, y) = \int_0^1 x(t)y(t)dt;$
- $X = W_m^2(0, 1);$

$$(x, y) = \int_0^1 x(t)y(t)dt + \int_0^1 \dot{x}(t)\dot{y}(t)dt + \dots + \int_0^1 x^{(m)}(t)y^{(m)}(t)dt;$$

- $|(x, y)| \leq \sqrt{(x, x)} \cdot \sqrt{(y, y)}$ – nierówność Schwartza.

b) Elementy $x, y \in X$ nazywamy ortogonalnymi (prostopadłymi), jeżeli $(x, y) = 0$.

Na przykład:

- $x = [1, 4]^T, y = [-4, 1]^T, (x, y) = x^T y = 0 \Rightarrow x \perp y,$
- funkcje $x_i(t) = \sin(i\omega t), y_j(t) = \cos(j\omega t)$, dla $t \in [0, 2\pi/\omega]$, dla $i \neq j$, są prostopadłe,

$$(x_i, x_j) = \int_0^{2\pi/\omega} x_i(t)x_j(t)dt = 0 \Rightarrow x_i \perp x_j,$$

$$(y_i, y_j) = \int_0^{2\pi/\omega} y_i(t)y_j(t)dt = 0 \Rightarrow y_i \perp y_j,$$

$$(x_i, y_j) = \int_0^{2\pi/\omega} x_i(t)y_j(t)dt = 0 \Rightarrow x_i \perp y_j,$$

$$(x_i, x_i) = (y_i, y_i) = \int_0^{2\pi/\omega} x_i^2(t)dt = \pi/\omega = \|x_i\|^2.$$

- 9) Przestrzeń unitarna X , to przestrzeń liniowa ze zdefiniowanym iloczynem skalarnym.
- 10) Przestrzeń unitarna, zupełna X nazywana jest przestrzenią Hilberta.
- 11) Przestrzeń Hilberta to podzbiór przestrzeni Banacha. Normę dla X można zawsze zdefiniować, jako pierwiastek z ilocznego skalarnego elementu $x \in X$.

$$\|x\| = \sqrt{(x, x)}.$$

- $X = \mathbb{R}^n : \|x\| = \sqrt{x^T x}$.

$$\bullet \quad X = L^2(0, 1) : \|x\| = \sqrt{\int_0^1 x^2(t)dt}.$$

$$\bullet \quad X = W_m^2(0, 1) : \|x\| = \sqrt{\int_0^1 x^2(t)dt + \int_0^1 \dot{x}^2(t)dt + \dots + \int_0^1 x^{(m)}(t)^2 dt}.$$

- Przestrzeń $C[0, 1], L^1(0, 1)$ i $L^\infty(0, 1)$ ze standardowymi normami, nie należą do przestrzeni Hilberta.
- W przestrzeni Hilberta iloczyn skalarny jest ciągły.
- Jeżeli Y jest domkniętą liniową podprzestrzenią przestrzeni Hilberta X , to dla każdego $x \in X$, istnieje jedyny element $y_m \in Y$ najbliższy x , nazywany rzutem ortogonalnym x na Y

$$\|x - y_m\| = \min_{y \in Y} \|x - y\|.$$

12) Definicje przekształceń na dwóch dowolnych niepustych zbiorach X, Y .

a) Przekształcenie $F: X \rightarrow Y$ nazywa się injekcją wtedy i tylko wtedy, gdy

$$x_1 \neq x_2 \Rightarrow y_1 \neq y_2 : \quad y = F(x).$$

b) Przekształcenie $F: X \rightarrow Y$ nazywa się surjekcją wtedy i tylko wtedy, gdy

$$\forall y \in Y, \exists x \in X : \quad y = F(x).$$

c) Przekształcenie $F: X \rightarrow Y$ nazywa się bijekcją \Leftrightarrow gdy jest injekcją i surjekcją.

d) Dla bijekcji $F: X \rightarrow Y$ istnieje przekształcenie odwrotne $G: Y \rightarrow X$ ($G = F^{-1}$), które spełnia warunki:

$$G(f(x)) = x : \quad \forall x \in X; \quad F(G(y)) = y : \quad \forall y \in Y.$$

e) Dla przestrzeni metrycznych X, Y przekształcenie $F: X \rightarrow Y$ jest ciągłe w punkcie

$$x \in X \Leftrightarrow \forall \varepsilon > 0, \quad \exists \delta > 0 : \quad \forall x_i \in K(x, \delta), \quad F(x_i) \in K(F(x), \varepsilon),$$

gdzie kula $K(x, \delta) = \{x_i \in X, \rho(x, x_i) < \delta\}$.

f) Przekształcenie F nazywa się homeomorfizmem, jeżeli jest ciągłe i posiada ciągłą odwrotność F^{-1} .

g) Przekształcenie $f: X \rightarrow R$ nazywa się funkcją.

13) Twierdzenie Banacha o odwzorowaniu zwężającym.

Dla przestrzeni (X, ρ) , która jest zupełna, dane jest odwzorowanie $F: X \rightarrow X$, które jest zwężające, tzn.

$$\rho(F(x), F(y)) \leq \alpha \rho(x, y) : \quad \forall x, y \in X, \quad 0 < \alpha < 1.$$

Wtedy równanie $x = F(x)$ posiada dokładnie jedno rozwiązanie $\hat{x} \in X$.

- 14) Przekształcenie $A: X \rightarrow Y$ na przestrzeniach liniowych X i Y nazywa się operatorem liniowym, gdy jest addytywne i jednorodne

$$A(\alpha x_1 + \beta x_2) = \alpha A(x_1) + \beta A(x_2): \quad \forall x_1, x_2 \in X, \quad \forall \alpha, \beta \in R.$$

Jądro operatora oznaczamy $\ker A$ (*kernel*).

Jądrem operatora liniowego $A: X \rightarrow Y$ nazywamy

$$\ker A = \{x \in X : A(x) = 0\}.$$

- 15) Dla operatora liniowego ciągłego $A: X \rightarrow Y$ na przestrzeniach unormowanych X i Y można wprowadzić pojęcie normy operatora $\|A\|$

$$\|A\| = \sup_{x \in X} \frac{\|Ax\|}{\|x\|} = \sup_{\|x\|=1} \|Ax\|.$$

- 16) Dany jest operator A na przestrzeniach liniowych X, Y

$$A: X \rightarrow Y: Ax = y.$$

Przy ustalonym $y \in Y$:

- a) może istnieć jedno rozwiązanie $x \in X$,
- b) może nie istnieć rozwiązanie,
- c) może być więcej niż jedno rozwiązanie.

Jeżeli X i Y są przestrzeniami Banacha, to operator odwrotny (jeżeli istnieje) jest liniowy i ciągły. Operator A^{-1} istnieje, gdy A jest wzajemnie jednoznaczne i obejmuje całą Y .

$$A^{-1}y = x.$$

- 17) Operator sprzężony $A^*: X^* \rightarrow Y^*$ jest określony równaniem $(Ax, y) = (x, A^*y)$ lub $(x, A^*y^*) = (Ax, y^*)$ i jest on liniowy i ograniczony oraz $\|A^*\| = \|A\|$.

- a) Jeżeli $X = R^n = Y$, to A jest macierzą $n \times n$, a $A^* = A^T$.

- b) Jeżeli $X = L^2[0, 1] = Y$ i operator $Ax = \int_0^1 G(t, \tau)x(\tau) dt: t \in [0, 1]$, to wykonując zamianę zmiennych i zmieniając kolejność całkowania, otrzymuje się

$$A^*y = \int_0^1 G(\tau, t)y(\tau)d\tau.$$

- 18) Jeżeli dla równania operatorowego $Ax = y$ rozwiązanie nie istnieje, to może istnieje przybliżone rozwiązanie z twierdzenia o rzucie ortogonalnym.

Jeśli X, Y są przestrzeniami Hilberta, to dla ustalonego $y \in Y$, wektor $x \in X$ minimalizuje normę $\|y - Ax\|$ wtedy i tylko wtedy, gdy $A^*Ax = A^*y$: $x = (A^*A)^{-1}A^*y$.
Zakłada się, że istnieje operator odwrotny.

- 19) Jeśli równanie $Ax = y$ posiada więcej niż jedno rozwiązanie, to można wybrać rozwiązanie minimalnonormowe $\bar{x}, \bar{x} = A^*(AA^*)^{-1}y$.
Zakłada się, że istnieje operator odwrotny.

- 20) Funkcjonały liniowe:

a) Każdy funkcjonał liniowy określony na R^n jest ciągły.

b) Funkcjonał liniowy określony na przestrzeni nieskończoności wymiarowej, nie musi być ciągły.

c) Funkcjonał liniowy określony na przestrzeni Banacha jest ciągły \Leftrightarrow gdy jest ograniczony.

d) Zbiór wszystkich funkcjonałów liniowych i ciągłych (czyli ograniczonych) f określonych na przestrzeni unormowanej X tworzy przestrzeń liniową X^* zwaną przestrzenią sprzężoną do X ($f \in X^*$). Można w niej zdefiniować normę elementu.

$$\|f\| = \sup_{x \neq 0} \frac{|f(x)|}{\|x\|} = \sup_{\|x\|=1} |f(x)|.$$

X jest przestrzenią zupełną Banacha, bo R jest przestrzenią zupełną.

- 21) Przestrzenie sprzężone.

a) Przestrzeń sprzężona do R^n jest też przestrzenią R^n .

b) Przestrzeń sprzężona do $C[0, 1]$ to przestrzeń funkcji o wahaniu ograniczonym lewostronne ciągłych w $[0, 1]$ i zerujących się w zerze $NBV[0, 1]$.

c) Przestrzeń sprzężona do $L^p(0, 1)$ to $L^q(0, 1)$ $[L^p(0, 1)]^* = L^q(0, 1)$ dla $1/p + 1/q = 1$.

d) Przestrzeń sprzężona do $W_m^p(0, 1)$ to $W_m^q(0, 1)$; $1/p + 1/q = 1$.

e) Zachodzi zawieranie: $[L^1(0, 1)]^* \subset [L^p(0, 1)]^* \subset [L^q(0, 1)]^* \subset [C[0, 1]]^*$.

f) Dla przestrzeni sprzężonej X^* unormowanej, można określić ponownie przestrzeń funkcjonałów liniowych i ciągłych, czyli przestrzeń do niej sprzężoną X^{**} – również unormowaną.

g) Przestrzeń X nazywamy refleksywną, gdy $X = X^{**}$.

Każda przestrzeń refleksywna jest przestrzenią Banacha.

Każda przestrzeń Hilberta jest refleksywna.

Na przykład przestrzenie $X = R^n, X = L^p(0, 1); X = W_m^p(0, 1)$: $1 < p < \infty$, są refleksywne.

Nie są refleksywne przestrzenie $C[0, 1], L^1(0, 1), L^\infty(0, 1)$.

- 22) Wnioski z twierdzenia Hahna–Banacha o rozszerzaniu funkcjonalu.

a) Dla $\forall x \in X$ (Banacha) istnieje funkcjonał liniowy, ciągły $F \in X^*$ taki, że $F(x) = \|F\| \cdot \|x\|$. Czyli, jeśli x będzie wybierany z kuli jednostkowej, to istnieje dla niego funkcjonal, który zrealizuje swoją normę $F(x) = \|F\|$.

b) Dla $\forall x \in X$, istnieje funkcjonał liniowy, ciągły $F \in X^*$ o normie jednostkowej $\|F\| = 1$ taki, że $F(x) = \|x\|$.

Dla x_o funkcjonał ten wyznacza hiperpłaszczyznę $\{x : F(x) = \|x_o\|\}$, która podpiera kulę $\{x : \|x\| \leq \|x_o\|\}$ w punkcie x_o , który jest wspólnym dla kuli i hiperpłaszczyzny.

- c) Twierdzenie odwrotne do b) jest słusne tylko dla przestrzeni refleksywnych X , tzn.: dla każdego funkcjonału liniowego ciągłego $F \in X^*$, $\exists x \in X$, że $\|x\| = 1$ i $F(x) = \|F\|$. Czyli dla każdego funkcjonału liniowego, ciągłego istnieje element z kuli jednostkowej pozwalający zrealizować normę tego funkcjonału. Lub ogólnie – na przestrzeni refleksywnej (np. Hilberta) każdy funkcjonał liniowy, ciągły realizuje swoją normę.

- 23) Twierdzenie Frecheta–Riesza o reprezentacji.

Każdy funkcjonał liniowy, ciągły (a więc ograniczony) $F(x)$ określony na przestrzeni Hilberta X , można przedstawić w postaci iloczynu skalarnego elementów x i elementu y .

$$F(x) = (x, y), \quad \forall x \in X, y \in X,$$

tzn. dla wszystkich x , istnieje dokładnie jeden taki element (wektor) $y \in X$. Zachodzi również własność, że dla każdego $y \in X$, istnieje jeden funkcjonał

$$F \in X^* : \|F\| = \|y\|.$$

Element y nazywa się reprezentacją funkcjonału.

Na przykład jeżeli $X = L^2[0, 1]$, to każdy funkcjonał liniowy, ciągły $F(x)$ musi mieć postać całki

$$F(x) = \int_0^1 x(t) y(t) dt \quad \text{i} \quad \|F\| = \sqrt{\int_0^1 y^2(t) dt}.$$

Na pewno też istnieje x , $\|x\| = 1$, takie, że funkcjonał realizuje swoją normę.

- 24) Twierdzenie Riesza o reprezentacji.

Każdy funkcjonał liniowy, ciągły określony na przestrzeni $C[0, 1]$ ma postać całki Stieltjesa

$$F(x) = \int_0^1 x(t) dh(t),$$

gdzie $h(t)$ jest funkcją o wahaniu ograniczonym, $h(t) \in NBV[0, 1]$, oraz $\|F\| = \|h\| = V_o[h]$.

- 25) Zawieranie się przestrzeni funkcyjnych:

UNITARNE \subset UNORMOWANE \subset LINIOWE \subset METRYCZNE,
 HILBERTA \subset REFLEKSYJNE \subset BANACHA \subset ZUPEŁNE \subset METRYCZNE,
 BANACHA \subset UNORMOWANE,
 HILBERTA \subset UNITARNE.

21. Zakończenie

W niniejszej monografii przedstawiono metodologię konstrukcji nowych algorytmów obserwacji skończenie wymiarowego wektora stanu, na tle innych zagadnień automatyki. Zaproponowane całkowe algorytmy zostały przetestowane dla obiektów o niskich i wysokich rzędach i wykazały się dokładnością odtwarzania stanu w skończonym i z góry zadanym czasie T , co ma decydujące znaczenie dla zadań prowadzonych przez komputery pod kontrolą systemów operacyjnych czasu rzeczywistego, gdy każde zadanie powinno mieć ścisłe określony czas swojego wykonania.

Ze względu na fakt, że całkowe obliczenia przetwarzające sygnały są prowadzone niezależnie dla wejścia i wyjścia, można zaproponować jednocześnie użycie dwóch komputerów lub jednej maszyny dwuprocesorowej (dwurdzeniowej). Pozwoli to skrócić czas obliczeń prowadzonych jednocześnie on-line w dwóch niezależnych oknach, przesuwanych na przedziałie $[t-T, t]$.

Ogólnie należy stwierdzić, że do niedawna kształt opracowywanych algorytmów identyfikacji, obserwacji i sterowania był dostosowywany do możliwości sprzętu obliczeniowego. Stąd na przykład od lat 60. XX w., jako jedyne stosowane były algorytmy obserwacji, oparte na równaniu różniczkowym (filtr Kalmana, obserwator Luenbergera). Nie implementowano również innych algorytmów, np. znanego wzoru na identyfikację parametrów Θ systemu liniowego, według wersji podstawowej metody najmniejszych kwadratów $\Theta = [C^T C]^{-1} C^T y$, [1SOT], który działa na skończonym oknie obserwacji i daje wyniki reprezentatywne tylko jako wartości średnie dla tego okna. Stosuje się raczej szybszą, rekurencyjną odmianę tej metody RLS (Recursive Least Squares, [1MAK]), niewymagającą odwracania macierzy w systemie sterowania on-line.

Algorytmy rekurencyjne związane są metodologicznie z równaniami różniczkowymi i różnicowymi. Stąd można otrzymywać rozwiązania z próbki pomiarowej na próbce, ale do wyniku ostatecznego rozwiązanie zbliża się asymptotycznie, a wynik w każdej chwili reprezentuje średnią wartość na podstawie całej informacji nagromadzonej od początku eksperymentu w takim „rozszerzanym oknie”. Algorytmy oparte na oknach o skończonej długości (które reprezentują grupę danych zwaną batchem) w wersji on-line tworzą przesuwane okna, stąd wynik odpowiada tylko informacji bieżącej zgromadzonej w ostatnim oknie. Im okno krótsze, tym wynik aktualniejszy, ale narażony na większe błędy w przypadku szumów pomiarowych. W związku z ciągłą potrzebą usuwania najstarszych danych, algorytmy te (w przeciwieństwie do algorytmów rekurencyjnych) muszą posiadać pamięć danych w ilości równej długości okna. Dla takich algorytmów, jak to wykazano w podroz-

Skorowidz

- algorytm pozycyjny 199
- algorytm przyrostowy 199
- aproksymacja
 - opóźnienia 111
 - Padégo 112
 - pochodnych 150
 - transformat 169
 - transmitancji dyskretnych 169
- bijekcja 489
- charakterystyka logarytmiczna 121
- charakterystyki Bodego 118
 - częstotliwościowe 118
 - ciąg Cauchy'go 485
 - czas regulacji 175
 - częstotliwość drgań swobodnych 131
 - drgań własnych 130
 - rezonansowa 135
 - człon całkujący 54
 - podtrzymujący 155
 - inercyjny 50
 - różniczkujący z inercją 51, 108
- delta Diraca 48
- dyskretny obserwator dokładny 417
- dyskretny regulator PID 198
- estymatory różniczkowe 273
- filtr FIR 173
 - IIR 172
 - Kalmana 367, 435
 - Kalmana rozszerzony 372, 373, 459
- filtracja 29
- forma kwadratowa 481
- funkcja o wahaniu ograniczonym 487
- funkcjonał liniowy 491
- identyfikacja 30
- iloczyn skalarny 487
- impulsator 155
- impulsowa funkcja przejścia 82, 96, 166
- iniekcja 489
- kolumna destylacyjna 445
- kryteria jakości 277, 280
 - — całkowe 280, 283, 288
 - — stabilności 214
- kryterium Gerszgorina 222
 - Hurwita 214
 - Michajłowa 218
 - Nyquista 219
- linearyzacja 238, 265
- linia przełączająca 353
- lokowanie zer i biegunów 335
- macierz 465
 - charakterystyczna 469
 - Frobeniusa 75, 79
 - fundamentalna 69, 71
 - obserwacji 71
 - Smitha 468
 - stanu 71
 - sterowania 71
 - Toeplitza 78, 80
 - transformacji 74
 - Vandermonda 75
 - wyjścia 71
- metoda funkcji opisującej 245
 - Lapunowa 237
 - Zieglera–Nicholsa 296, 299, 300
- metoda Popowa 241

metody obserwacji 359
 model ciepłoprzewodnictwa 399
 — kolumny destylacyjnej, statyczny 448
 — kolumny destylacyjnej, dynamiczny 453
 — wewnętrzny 188
 modele statyczne 40, 42
 modelowanie 28, 37
 norma operatora 490
 — spektralna 483
 obiekt inercyjny 55
 obserwacja stanu 31
 — w przestrzeniach Hilberta 392
 obserwator całkowy 375, 376, 378, 460
 — adaptacyjny 435
 — — z minimalną normą 379
 — dokładny 375, 417, 425
 — — różniczkowy 425
 — EWO 410
 — Luenbergera 362
 — MWO 409
 obserwowałość 266, 268
 odchyłka maksymalna 302
 odwrotna transformata Laplace'a 91
 okres dyskretyzacji 155, 174
 — próbkowania 155
 operator liniowy 490
 — sprężony 490
 opóźnienie transportowe 111
 optymalizacja parametryczna regulatora PI 309
 optymalne nastawy regulatorów 277

parametry Markowa 160
 płaszczyzna fazowa 137
 postacie obserwowań 77
 — sterowalne 76
 przebiegi przejęciowe 42
 przekążnik 45, 206, 247
 — z histerezą 46
 przekształcenie w 232
 przestrzeń Banacha 486
 — Hilberta 488
 — metryczna 485
 — refleksywna 491
 — sprężona 491
 — unitarna 488
 — unormowana 486
 — zupełna 485

przesuwnik fazowy 108
 pseudoodwrotność lewa 480
 — prawa 479
 punkt pracy 40
 — równowagi 236

realizowalność fizyczna transmitancji 96
 regulator czasooptymalny 351
 — Dead-beat 338
 — liniowo-kwadratowy 343, 347
 — PID 193
 — samonastrajalny 184
 — stanu 335
 reguły całkowania 151
 rozwiązywanie algebraicznego układu równań 478
 równania nieliniowe 61, 63
 — różniczkowe 65
 równanie różnicowe 149
 — różniczkowe stanu 69
 — — Lotki–Volterry 63
 — — Lienarda 63
 — — Riccatiego 62
 — — Van der Pola 63

schematy blokowe 99
 silnik prądu stałego 53
 słowa kluczowe 33
 splot 82
 sprężenie zwrotne 27, 110, 180, 186
 — do przodu 186
 stabilizowałość 262
 stabilność 208
 — względem warunku początkowego 235
 — BIBO 211
 — układów dyskretnych 229
 — układów nieliniowych 234
 stan układu 66
 — ustalony 225, 227
 stany przejęciowe 42
 sterowalność względem stanu 253, 260
 — praktyczna 261
 sterowanie adaptacyjne 32
 — — z modelem odniesienia 185
 — minimalnonormowe 340, 341
 — modalne 342
 — odporne 32
 — optymalne 344
 — predykcyjne 33
 sterowniki logiczne PLC 201
 struktury systemów sterowania 186

suma splotowa 158
 surjekcja 489
 sygnały wykładnicze 116
 synteza regulatora 31
 szereg Fouriera 123
 — Taylora 56, 59

trajektoria fazowa 207
 transformacja odwrotna 91
 transformacje stanu 74
 transformata Fouriera 125
 — Laplace'a 87
 — Z 162
 transmitancja operatorowa 95
 — wielowymiarowa 104
 — widmowa 116
 — wyparki 310
 twierdzenie Banacha 489
 — Frecheta–Riesza 492
 — Hahna–Banacha 491
 — Paleya–Wienera 129
 — Riesza 492

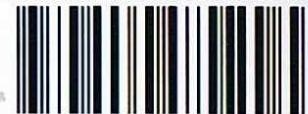
układ adaptacyjny 183
 — całkująco-inercyjny 356

— impulsowy 168
 — inercyjny 107
 — kaskadowy 182
 — podwójnie całkujący 353
 — przesuwający fazę 108
 — regulacji 180
 — II rzędu 52, 63
 — III rzędu 139
 — z modelem wewnętrznym 188
 układy automatyki procesowej 177
 — automatyki zabezpieczeniowej 177
 — nieminimalnofazowe 122

warstwy sterowania 192
 wektory własne 471
 współczynnik tłumienia 130
 wykres Wyszniewgradskiego 141

zakłócenia 180
 — najbardziej niebezpieczne 301
 zasady ciągłości 49
 zbiór zwarty 485
 zjawisko *windup* 197
 zmienne odchyłkowe 98
 — stanu 101

BIBLIOTEKA GŁÓWNA AGH



1000214331

