AKADEMIA EKONOMICZNA W KRAKOWIE

CZESŁAW KULIK, RYSZARD TADEUSIEWICZ

ELEMENTY CYBERNETYKI EKONOMICZNEJ

KRAKÓW 1974

AKADEMIA EKONOMICZNA W KRAKOWIE

CZESŁAW KULIK RYSZARD TADEUSIEWICZ

ELEMENTY CYBERNETYKI EKONOMICZNEJ

KRAKÓW 1974

OPRACOWANIE

Czesław Kulik: rozdz. IV - punkty 7.1, 7.2, 8 - 13, V oraz Wstęp

Ryszard Tadeusiewicz: rozdz. I, II, III, IV - punkty 1 - 5, 7.3, 14, 15, VI oraz Zakończenie

RECENZENT

Stefan Mynarski

Wydano za zgodą Rektora
Akademii Ekonomicznej w Krakowie
pismo z dnia 20 II 74 r., nr N 33a/39/74

Nakład: 1500 egz.
Objętość: 16,0 ark. wyd.
Cena: 20,- zł
Rze. 1762/75 S-62-102 1500 egz

SPIS TREŚCI

Rozdział IV

AUTOMATY SKOŃCZONE I ALGORYTMY

Rozdział V

ELEMENTY LOGIKI MATEMATYCZNEJ

Rozdział VI

ELEMENTY TEORII INFORMACJI

W s t ę p

Dokonujący się gwałtowny rozwój automatyzacji wielu procesów oraz informatyki, wytwarza zapotrzebowanie na specjalistów znających nowe działy matematyki, tj. teorię zbiorów, logikę matematyczną, a przede wszystkim cybernetykę.

Wiele zagadnień bada się obecnie z różnych stron przez matematyków, statystyków, elektroników i innych specjalistów, którym często jest trudno porozumiewać się między sobą. Te same pojęcia w każdej grupie specjalistów noszą zupełnie inne nazwy. Wiele uciążliwych badań powtarza się bez potrzeby wielokrotnie, bez wykorzystania wyników otrzymanych w sąsiednich dziedzinach.

Celowe jest podjęcie jednolitych wysiłków przez specjalistów z różnych dziedzin, aby przełamać bariery terminologiczne, metodologiczne i podchodzić do rozwiązywania zadań z ogólnego punktu widzenia. Wysiłki w tym kierunku zostały zapoczątkowane w czterdziestych latach przez grupę wybitnych specjalistów z różnych gałęzi nauki, zainteresowanych jednym problemem - problemem sterowania.

Inżynierowie opracowali i wykonali aparaturę elektroniczną do sterowania. Matematycy badali własności sygnałów w różnych systemach i podali analityczny ich opis. Inni specjaliści rozwijali teorię informacji. Statystycy rozpatrywali potok informacji w organiźmie jako podstawę regulowania jego funkcji.

W wyniku tej współpracy wyodrębniła się nowa dziedzina nauki "o ogólnych zasadach sterowania i sprzężeń w żywych organizmach", którą nazwano cybernetyką.

Aktualnie śledzimy intensywny rozwój cybernetyki, która obejmuje kompleks nauk, takich jak:

- cybernetyka teoretyczna,
- cybernetyka biologiczna,

- cybernetyka techniczna,
- cybernetyka przemysłowa,
- cybernetyka ekonomiczna i inne.

Podstawowe pojęcia w cybernetyce to: system, struktura, informacja, sygnał, sterowanie, sprzężenie zwrotne.

Jednym z centralnych problemów w rozwoju nauki i rozwoju świata stał się problem tworzenia wielkich systemów i sterowania nimi. Przedsiębiorstwa przemysłowe i handlowe, organa planowania, rynek, gospodarka narodowa lub jej gałęzie, mogą być rozpatrywane jako wielkie systemy. Przy badaniu wielkich systemów konieczne jest analizowanie ogromnej liczby powiązań między elementami i zjawiskami, uwzględnianie współdziałania i udziału części w całości, uwzględnianie nieokreśloności w prowadzeniu systemu oraz jego połączeń z otoczeniem i współdziałania z nim.

W przemyśle, w badaniach naukowych i technicznych, dzięki zastosowaniu automatycznych systemów, stało się możliwe rozwiązanie wielu złożonych zadań sterowania. W miarę wzrostu złożoności struktury sterowanych obiektów, zwiększenia objętości informacji o zachodzących w nich procesach, człowiek często nie był w stanie w najlepszy sposób realizować funkcji zarządzania. Tłumaczy się to brakiem czasu, w ciągu którego powinna być podjęta optymalna decyzja, trudnością zmobilizowania w krótkim czasie dużych obszarów pamięci, właściwością zapominania informacji i szeregiem innych czynników.

Złożone systemy automatycznie sterowane cechują się dużą prędkością pracy i dużą pojemnością urządzeń pamięciowych. Powinny one realizować takie funkcje, jak: porównanie różnych wariantów rozwiązania zadań, wybór najlepszego wariantu przy uwzględnieniu określonych kryteriów i ograniczeń, uwzględnienie zmian czynników zewnętrznych i dokonywanie związanych z tym zmian rozwiązania i kryteriów. Konieczne jest przy tym uczenie się przewidywania.

Wprowadzenie elektronicznych maszyn cyfrowych rozszerzyło możliwości rozwiązania złożonych zadań przetwarzania informacji i automatyzacji sterowania, a ponadto przyczyniło się do rozwoju cybernetyki, teorii gier, teorii informacji, teorii automatów i wielu innych dziedzin wiedzy.

Specjaliści związani z badaniem, budową i prowadzeniem złożonych systemów przetwarzania informacji i sterowania, powinni znać te dziedziny wiedzy i opanować wspólny język porozumiewania się, bowiem pracować będą w zespołach obejmujących różnych specjalności inżynierów, ekonomistów, matematyków i innych.

Celem niniejszej pracy jest zapoznanie ekonomistów i innych specjalistów w możliwie przystępny sposób z cybernetyką współczesną, której aparat pojęciowy i metodologiczny ułatwi im pracę w zespołach, a także ukształtowanie postępowego światopoglądu.

Praca składa się z 6 rozdziałów.

W pierwszym rozdziale omówiono cele, zadania i podstawowe pojęcia cybernetyki oraz zasady modelowania obiektów, między innymi, ekonomicznych.

Drugi rozdział jest poświęcony przedstawieniu sprzężeń układów m.in. ekonomicznych oraz głównych zagadnień teorii regulacji.

Trzeci obszerny i ważny rozdział został przeznaczony na omówienie teorii wielkich systemów i optymalnego sterowania systemami gospodarczymi. Omówiono w nim metody sterowania wielkimi systemami z uwzględnieniem metod programowania matematycznego, teorii gier, analizy sieci zależności i wielu działów teorii podejmowania decyzji.

W rozdziale czwartym zostały podane podstawy teorii automatów oraz teorii algorytmów i niektórych jej zastosowań, ułatwiające zrozumienie problemów związanych z programowaniem, budową i działaniem EMC.

Następne dwa rozdziały zostały poświęcone logice matematycznej i teorii informacji, w zarysie umożliwiającym przyswojenie głównych pojęć, terminów i metod, przez czytelników którzy mogą mieć trudności w studiowaniu odrębnych pozycji bogatej literatury z tego zakresu, głównie w językach obcych.

W literaturze krajowej nieliczne pozycje są poświęcone cybernetyce, a zwłaszcza cybernetyce ekonomicznej. Niniejsza praca przyczyni się do wypełnienia tej luki, co stało się pilne i ważne w związku z uruchomieniem w uczelniach ekonomicznych kierunku studiów: cybernetyka ekonomiczna i informatyka.

Objaśnienia do posługiwania się książką

Podręcznik ten zredagowany został z myślą o tzw. nauczaniu programowanym. Innymi słowy, można w zależności od indywidualnych zdolności i przygotowania zapoznać się z przedstawianym w nim materiałem szybciej lub wolniej. Każda partia nowego materiału zakończona jest testem, który pozwoli ocenić stopień przyswojenia nowych treści, W zależności od wyników testu czytelnik zostanie skierowany do odpowiednich partii materiału, które zawierać będą niezbędne wyjaśnienia przy negatywnym wyniku, lub stanowic będą kontynuację. Przy niestarannym czytaniu możliwe jest kierowanie przez test powtórnie do niestarannie przerobionego materiału.

Czytając książkę należy przysłaniac tekst odpowiedzi, wypisanych poniżej znaku xxx do czasu udzielenia samodzielnych odpowiedzi na pytania testowe.

Naturalnie można nie stosować się do powyższych wskazówek, można "ściągać" odpowiedzi lub nie powtarzać partii materiału, pomimo negatywnego wyniku testu, ale wówczas utraci się możliwość krytycznej oceny swoich wiadomości.

Rozdział I

POJĘCIE UKŁADU. ZASADY MODELOWANIA CYBERNETYCZNEGO OBIEKTÓW EKONOMICZNYCH

1. Cele, zadania i podstawowe pojęcia cybernetyki

1. Istnieje bardzo wiele definicji określających sens słowa C Y B E R N E T Y K A, daje się nawet zauważyć pewna tendencja do wyszukiwania oryginalnych sformułowań, my jednak zamiast tworzyć jeszcze jedną definicję, posłużymy się pewnym cytatem:
"/.../ naukowcy uświadomili sobie, że problemy komunikacji, sterowania i mechaniki statystycznej stanowią zasadnicze jedną całość, niezależnie od tego czy dotyczą maszyny, czy istoty żywej. Z drugiej jednak strony poważną przeszkodą był brak takiej jednosci w literaturze tematu, brak wspólnej terminologii, wreszcie nawet brak oddzielnej nazwy dla tej dziedziny./.../. Zostaliśmy zmuszeni, jak to często bywa, do ukucia przynajmniej jednego sztucznego, neogreckiego wyrażenia, które wypełniłoby lukę. Zdecydowaliśmy się nadać całej dziedzinie teorii sterowania i komunikacji w maszynach i zwierzętach nazwę cybernetyki, którą utworzyliśmy od greckiego κυβερνητης czyli "sternik". Zacytowany fragment pochodzi z książki Norberta Wienera [1/], a jego wartość polega na tym, że właśnie wówczas po raz pierwszy padło słowo cybernetyka.

2. Zapamiętajmy więc, że cybernetyka jest nauką o sterowaniu i komunikacji. Przedmiotem jej są procesy przetwarzania i przesyłania INFORMACJI, niezależnie od tego, czy zachodzą one

---

1/Norbert Wiener: Cybernetics or control and communication in the animal and the machine, The MIT-Press, Cambridge 1948.

w urządzeniu technicznym, systemie ekonomicznym, społeczeństwie czy w żywym organiźmie.

3. Termin "informacja" użyty w poprzednim punkcie, jest w cybernetyce rozumiany szerzej niż potocznie. Właśnie informacja jest podstawową treścią procesów, którymi zajmuje się cybernetyka, jej przetwarzanie jest podstawowym celem wszystkich systemów cybernetycznych, natomiast zagadnienia materialnej formy systemów cybernetycznych i procesów energetycznych w nich zachodzących są z punktu widzenia cybernetyki nieistotne.

4. Naturalnie nie można całkowicie oderwać pojęcia informacji od pojęć materialnych, gdyż system cybernetyczny musi być zbudowany z określonego tworzywa, zaś przekazywaniu i przetwarzaniu informacji towarzyszyć muszą procesy energetyczne, jednakże uwaga cybernetyki koncentruje się na poziomie wyabstrahowanego pojęcia informacji. Wyjątkiem są przypadki, kiedy struktura materialna lub proces energetyczny niesie informację, na przykład kiedy funkcje informacyjne są uwarunkowane tworzywem, tak jak w przypadku kabla w izolacji - do przekazywania informacji służą metalowe przewody, natomiast otoczka nie spełnia funkcji informacyjnych, lub kiedy działanie źródła energii zmienia się w czasie, przekazując odbiorcom energii informację.

5. Będziemy przyjmowali, że informacje są treścią S Y G N A Ł Ó W. Sygnał posiadać może określone znaczenie własne i niekiedy inne znaczenie symboliczne. Przykładem znaczenia własnego sygnału jest treść napisu "ciotka przyjechała". Napis ten może jednak mieć dla wtajemniczonych pewne ukryte znaczenie symboliczne, na przykład może powiadamiać o ruchach wojsk. Mówimy wówczas, że sygnał jest K O D E M innego sygnału, proces zmiany sygnału na sygnał symboliczny nazwiemy K O D O W A N I E M; nazwa D E K O D O W A N I E, jako określenie operacji odwrotnej narzuca się sama.

TEST 1

P. Jakimi procesami zajmuje się cybernetyka ?

xxx

O. Procesami sterowania i komunikacji, polegającymi na przetwarzaniu i przesyłaniu informacji.
Jeśli odpowiedziałeś nieprawidłowo przeczytaj ponownie punkty 3 i 4.

P. Które z wymienionych pojęć można utożsamić z pojęciem informacji?
a/ list,
b/ seria impulsów napięcia, biegnąca w kablu telefonicznym,
c/ treść ogłoszenia prasowego,
d/ dźwięk budzika.

xxx

O. c/ i d/
Jeśli odpowiedziałeś nieprawidłowo, przeczytaj punkty 3 i 4 oraz zapoznaj się z przykładem 1.

P. Czy usprawiedliwione jest twierdzenie, że cybernetyka zajmuje się techniką elektronicznych maszyn liczących?

xxx

O. Nie, ponieważ cybernetykę interesują jedynie procesy informacyjne, a te w maszynie liczącej są dużo mniej ciekawe niż np. w rozwijającym się zarodku.

P. Czy stale palące się światło może być sygnałem?

xxx

O. Nie, ponieważ fakt stałego świecenia nie pozwala na żadne zmiany, a tym samym nie jest przekazywana żadna informacja. Aby sygnał niósł informację musi posiadać najmniej dwie wartości "Tak - Nie".
Jeśli odpowiedziałeś poprawnie na oba pytania możesz pominąć czytanie następujących dalej przykładów i przystąpić od razu do czytania punktu 6.

Przykład 1

Ważnym jest stwierdzenie punktu 3 głoszące, że informacja jest czymś odrębnym, oderwanym od materialnego nośnika informacji oraz od procesu energetycznego towarzyszącego przekazywaniu informacji. Nieco upraszczając zagadnienie można powiedzieć, że informacją jest T R E Ś Ć sygnału, a nie jego forma materialna czy energetyczna. Dlatego też list nie jest informacją, jest on sumą informacji i materialnego nośnika. Podobnie czasowe zmiany napięcia lub prądu w kablu telefonicznym stanowią zjawisko energetyczne, które niesie informację, lecz nie jest z nią tożsame, gdyż ktoś nierozsądny mógłby użyć tego prądu na przykład do ogrzewania pokoju. Natomiast treść ogłoszenia jest informacją, podobnie jak dźwięk budzika [2/], gdyż treści te stanowią nie zjawiska, a wiadomości. Oczywiście treść listu również będzie informacją, w przeciwieństwie do całego listu.

TEST 2

P. Czym jest desygnat /materialny odpowiednik/ pojęcia "Pan Tadeusz? Czy jest to:

a/ książka,

b/ zbiór wszystkich książek pod tym tytułem,

c/ uporządkowany zbiór myśli Adama Mickiewicza?

xxx

O. Ani a/, ani b/, ani c/. Jest to właśnie informacja. Jeśli miałeś wątpliwości przestudiuj przykład 2.

Przykład 2

Dzieła literackiego nie można utożsamiać z książką, ponieważ jak należałoby wówczas traktować inną książkę o tym samym tytule i treści - jako jeszcze jeden poemat? Ten sam czy inny? Jeśli ten sam, to każde wznowienie powodowałoby dalsze "rozmycie" odpowiednika pojęcia "dzieło literackie"

[2/] Dźwięk rozumiemy tu jako subiektywne doznanie psychiczne, a nie jako falowe zjawisko akustyczne polegające na rozchodzeniu się obszarów zwiększonego ciśnienia.

na dalsze dziesiątki tomów. Zbiór wszystkich książek nas również nie zadawala, bo czym jest wówczas inscenizacja czy recytacja? Z myślami poety utożsamić dzieła również niepodobna, ponieważ, niezależnie tego czy poeta żyje, czy zmarł dawno, myśl już nie istnieje, podczas gdy o istnieniu dzieła jesteśmy przekonani. Jest to dobra ilustracja niezależności pojęcia informacji od materialnej formy jej nośnika, która, być może podbuduje zaufanie Czytelnika do twierdzenia, że informacja, podstawowy przedmiot cybernetyki, posiada byt obiektywny, a jednocześnie odmienny od energii i materii. Dla nieprzekonanych przytaczamy przykład 3, nie mający wątpliwości mogą go pominąć.

Przykład 3

Rozpatrzmy prosty mechanizm: turbinkę napędzaną wodą. /Rys. 1/. Aby turbinka wykonywała pracę, niezbędna jest obecność energii płynącej wody, gdyż przy jej braku /Rys. 1b/ turbinka jest nieruchoma, oraz materii /Rys.1c/, są to jednak warunki niepełne. Dla kompletu potrzeba jeszcze czegoś, czego brak jest na rysunku 1d, a mianowicie struktury /odwrotnie założony wirnik nie będzie się poruszał, gdyż woda opływa jego łopatki/. Określona struktura, organizacja materii, jest przejawem działania informacji, jest z nią tożsama. Podkreślamy: struktura, a nie materialny twór o określonej strukturze.

TEST 3

P. Co zawiera więcej informacji - zegarek czy szyna stalowa?

xxx

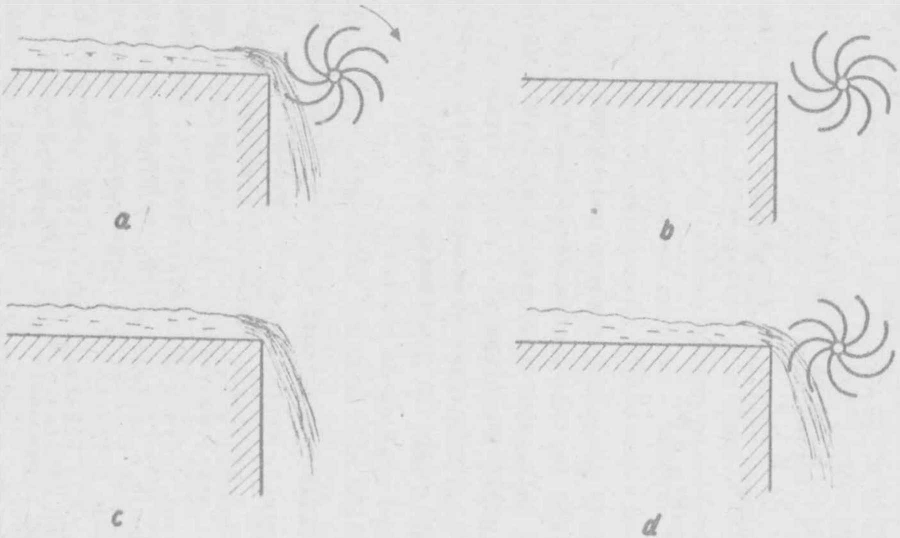
O. Oczywiście zegarek, posiadający daleko bardziej złożoną strukturę. Jeśli miałeś wątpliwości przeczytaj uważnie jeszcze raz przykład 3.

P. Jaki organ żywego człowieka jest dla cybernetyki najbardziej interesujący:

a/ układ oddechowy,

b/ trawienny,

c/ nerwowy,



Rys. 1. Ilustracja znaczenia struktury dla działania układu

d/ mięśniowo-szkieletowy?

xxx

O. Oczywiście układ nerowy, spełniający funkcje przyjmowania informacji od otoczenia, przesyłania informacji pomiędzy mózgiem a poszczególnymi organami i przeróbki informacji /podejmowanie decyzji, uczenie/.
Jeśli nie odpowiedziałeś na to pytanie przeczytaj raz jeszcze punkty 1. 2. 3 i 4.

P. Który z dwu podanych zapisów jest kodem?
a/ Ala ma Asa,
b/ B - 234 - xy/W = : 9.

xxx

O. Zarówno a/, jak i b/ może być kodem pod warunkiem istnienia odpowiedniego systemu kodowania i dekodowania.
Brak poprawnej odpowiedzi powinien zachęcić Cię do powtórnego zapoznania się z punktem 5.
Czy chciałbyś pogłębić swoje wiadomości na temat pojęcia informacji? Jeśli tak, to przeczytaj przykład 4, w przeciwnym przypadku przejdź do punktu 6.

Przykład 4

Istnieje możliwość przesyłania sygnałów zarówno w przestrzeń, jak i w czasie. Przesyłanie przestrzenne jest bardziej oczywiste i znane z praktyki, natomiast przepływ informacji poprzez czas jest mniej oczywisty. Wyjaśnijmy na wstępie, że komunikacja ta jest jedynie jednokierunkowa, można mianowicie przesyłać informacje w przyszłość /np. robiąc notatki/ lub odbierać informacje z przeszłości /np. wykopaliska archeologiczne/. Można na tej podstawie naświetlić dodatkowo problem pamięci. Ponieważ wspomnienie jest informacją przesyłaną w czasie, więc może być w cybernetyce traktowane na równi i w identyczny sposób, jak każda informacja przesłana dowolnym kanałem.

6. Kończąc rozdział dotyczący podstawowych pojęć cybernetyki, niepodobna nie wskazać kilku zasadniczych cech charak-

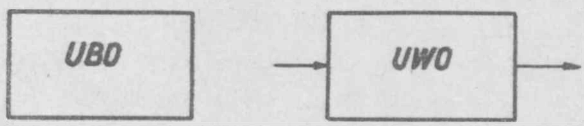
teryzujących tę naukę na tle innych dziedzin badań naukowych. Ze względu na przedmiot zainteresowań cybernetyki, jakim jest informacja, której podstawową cechą jest zmienność, niemożliwe jest stosowanie w badaniach cybernetycznych rozważań stanów stacjonarnych, tak chętnie i szeroko stosowane w innych naukach. Podstawowymi własnościami systemów badanych przez cybernetykę jest ich zmienność, a badanie dynamiki tej zmienności pod wpływem oddziaływań sterujących, stanowi jeden z zasadniczych problemów cybernetyki. Systemy, którymi cybernetyka się zajmuje nie są odosobnione, lecz mogą tworzyć zbiór, przy czym jednym z kierunków badań aktualnie prowadzonych jest śledzenie wpływu właściwości systemu - składnika na właściwości systemu - zbioru. Wyodrębnienie obiektu badań cybernetycznych z otoczenia ma charakter informacyjny /czyli system cybernetyczny nie wymienia z otoczenia informacji niekontrolowanej/, natomiast nie musi być to równoznaczne z wyodrębnieniem fizycznym, stąd duża abstrakcyjność stosowanych pojęć i trudność szybkiego stosowania osiągnięć cybernetyki, oraz z drugiej strony niebezpieczeństwo pominięcia w pobieżnej analizie niekiedy istotnych czynników.

2. Układ. Układ względnie odosobniony. Wejście, wyjście i stan układu. Transmitancja i funkcja przejścia.

1. Zakończenie poprzedniego paragrafu wprowadziło pojęcie systemu wyodrębnionego pod względem informacyjnym i wskazało na pewien dualizm, pewną relatywność pojęć, polegającą na tym, że ten sam obiekt może być raz traktowany jako system, a innym razem jako element systemu, w skład którego wchodzi poza nim jeszcze szereg innych podsystemów.
Aby nie komplikować rozważań wprowadzimy następującą terminologię: obiekt badań cybernetycznych, którego strukturą wewnętrzną zajmować się nie będziemy i który może ewentualnie stanowić "cegiełkę" do budowy większej struktury, nazwiemy UKŁADEM, natomiast nazwę SYSTEM zarezerwujemy dla wię-

kszej struktury złożonej z pewnej ilości połączonych we wiadomy lub niewiadomy sposób układów. Naturalnie nie przeszkodzi nam to, jeśli zajdzie potrzeba, w traktowaniu SYSTEMU jako całości, bez wnikania w jego strukturę, a więc tak, jakby był on UKŁADEM.

2. Obiekty cybernetyczne rozpatrywane są wyłącznie w aspekcie informacyjnym i tylko w tym sensie muszą być wyodrębnione z otoczenia. Rozważmy układ /rys. 2/, który nie pobiera



Rys. 2. Układ bezwzględnie i względnie odosobniony

z zewnątrz informacji, ani informacji na zewnątrz nie wysyła. Układ taki jest więc w sensie cybernetyki doskonale odizolowany od otoczenia, chociażby przez układ ten np. przepływał bez przerwy strumień jakiegoś medium. Układ ten jest jednak nieciekawy - nie mamy możliwości wpływać na niego /sterować/, ani obserwować jego działania. Układ taki nazwiemy Układem Bezwzględnie Odosobnionym. /UBO/

3. Układ bezwzględnie odosobniony jest idealizacją zbyt daleką, ograniczmy więc nasze rozważania do układu, do którego informacja z zewnątrz dopływa ściśle określonymi kanałami, które nazwiemy W E J Ś C I A M I układu, a wydawana jest /po przetworzeniu w układzie/ również wyłącznie pewnymi kanałami, które nazwiemy W Y J Ś C I A M I. Układem takim można sterować oddziaływując na jego wejścia i obserować jego zachowanie poprzez wyjścia. Procesy zachodzące wewnątrz są nie-

dostępne i można na ich temat mówić jedynie w kontekście wyjście - wejście. Układ taki nazwiemy U K Ł A D E M W Z G L Ę D N I E O D O S O B N I O N Y M /UWO/.

TEST 1

P. Jak nazywamy kanał przekazywania informacji do układu bezwzględnie odosobnionego?

xxx

O. Oczywiście nie ma takiego kanału, więc i nie ma nazwy. Jeżeli nie dostrzegłeś tego, to znak, że czytasz nieuważnie i dobrze Ci zrobi ponowne przeczytanie punktów 1. 2. i 3.

P. Jeżeli samochód potraktujemy jako układ względnie odosobniony, to co można nazwać wejściami do tego układu, a co jego wyjściami?

xxx

O. Wejściami są kierownica i pedał przyspiesznika /"gazu"/ oraz hamulca, natomiast wyjściami prędkość i kierunek ruchu.
Jeżeli nie uporałeś się z tym zadaniem możesz zrehabilitować się odpowiadając na następne, łatwe pytanie:

P. Dlaczego nie traktuje się wejścia do UWO, jakim jest samochód, wlewu benzyny, a jako wyjścia koloru?

xxx

O. Wlew benzyny pozwala na wprowadzenie do UWO pewnej ilości materii /lub energii, jeśli uwzględnić, że spalanie benzyny jest źródłem energii układu/, natomiast nie wprowadza żadnej informacji. Natomiast kolor samochodu, pozostając niezmiennym, również /aczkolwiek z innych powodów/ nie niesie żadnej informacji o układzie ani o sterowaniach. Oba czynniki mają więc inny, niż informacyjny charakter, nie są więc wejściem i wyjściem.
Gdybyś przypadkiem nie uporał się z tym pytaniem należycie, powinieneś przeczytać punkty 2, 3 i 4 oraz odpowiedzi

do pytań z rozdziału poprzedniego, jak również punkt 3 niniejszego rozdziału.

P. Co jest wyjściem UWO, jakim jest zegarek? Czy układ ten ma wejście?

xxx

O. Położenie wskazówek. Nie ma, nakręcanie jest czynnością energetyczną, a nie informacyjną!
Brak odpowiedzi na to pytanie powinien zachęcić Cię do zapoznania się z przykładem i poprawna odpowiedź upoważnia do pominięcia przykładu i przystąpienia do dalszych pytań /TEST 2/.

Przykład 1

Człowiek jest z punktu widzenia cybernetyki układem względnie odosobnionym. Posiada on bowiem określony zespół kanałów informacyjnych wejściowych /narządów zmysłów/ i wyjściowych /efektorów/. Poza tymi kanałami wymiana informacji nie zachodzi ani do, ani z układu. Na zasadzie wąskiego skanalizowania informacji dopływającej z otaczającego świata opartych było szereg systemów filozoficznych, rozważających wiarogodność napływającej tym kanałem informacji. Jako ciekawostkę warto przytoczyć pogląd solipsystów, którzy uważali całą informację zmysłową za mylną i wnioskowali, że poza danym konkretnym człowiekiem nic nie istnieje, gdyż wszystkie doznania są mirażem.

Nie czas tu ani miejsce poddawać dyskusji te poglądy, warto jednak znać ich cybernetyczny aspekt.

TEST 2

P. Czy moglibyśmy obserwować układ bezwzględnie odosobniony?

O. Nie, ponieważ obserwacja byłaby równoważna z otrzymywaniem informacji od układu, a to jest z definicji niemożliwe.

P. Określ, które z wymienionych przedmiotów można traktować jako UWO oraz wskaż ich wejścia i wyjścia:
a/ młotek,

b/ telefon,
c/ lampa,
d/ mięsień,
e/ kolonia bakterii,
f/ dynamika kształtowania się ceny rynkowej,
g/ księgowość przedsiębiorstwa handlowego,
h/ sprężyna,
i/ równanie matematyczne.

xxx

0. Wszystkie są układami względnie odosobnionymi. Jeżeli miałeś wątpliwości co do trzech lub większej ilości przedmiotów powinieneś raz jeszcze przeczytać punkt 3, a następnie możesz zrehabilitować się podając samodzielnie informacje wejściowe i wyjściowe wszystkich układów:

xxx

1/ informacją wejściową jest wielkość i kierunek działającej na trzonek siły, a wyjściową szybkość i kierunek ruchu,
2/ informacją wejściową jest położenie słuchawki /podniesiona lub nie/, ruch tarczy numerowej oraz głos docierający do mikrofonu, a wyjściową sposób zmian napięcia w przewodzie łączącym telefon z siecią abonencką i głos płynący ze słuchawki,
3/ położenie wyłącznika - obecność lub brak światła,
4/ impulsy mioneuronów znajdujących się w rdzeniu kręgowym - stopień skrócenia i siła rozwijana przez mięsień,
5/ ilość pożywki i chemizm środowiska - szybkość rozprzestrzeniania się kolonii i ilość osobników,
6/ podaż i popyt - wysokość ceny rynkowej,
7/ faktury dostaw i raporty sklepowe - zestawienia i plany,
8/ siła nacisku - stopień skrócenia sprężyny,
9/ parametry, współczynniki i dane wyjściowe - rozwiązanie.

Jeżeli przy udzielaniu odpowiedzi napotkałeś na duże trudności powinieneś przed dalszym czytaniem zapoznać się z książką H. Greniewski "Cybernetyka niematematyczna" [3] lub książką J.Tiepłow "O cybernetyce" [4], gdzie w początkowych rozdziałach omawiana jest w popularnej formie rozważana tu problematyka.

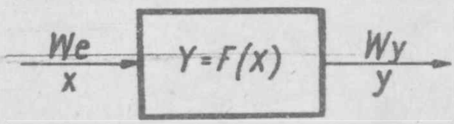
P. Czym różni się system od układu?

xxx

O. Stopniem złożoności struktury; system może być złożeniem dużej ilości układów. Rozróżnienie to w dużej mierze wynika z punktu widzenia badającego konkretny obiekt - jeśli ma możliwość wniknąć w strukturę obiektu mówi o systemie, jego częściach /układach/ i organizacji wewnętrznej, przy braku tej informacji mówi o całości jako o układzie.

Brak odpowiedzi odsyła do ponownego przeczytania punktu 1.

4. Pomiędzy informacją wejściową a wyjściową zachodzą w układzie względnie odosobnionym określone relacje. Aby je wyjaśnić skupimy się chwilowo na przypadku UWO o jednym wejściu i jednym wyjściu /rys. 3/.



Rys. 3 Funkcja przejścia

Układ taki nazwiemy J E D N O W Y M I A R O W Y M układem względnie odosobnionym. W układzie takim wielkość wyjś-

[3] H. Greniewski: Cybernetyka niematematyczna,PWN Warszawa 1969
[4] L. Tiepłow: O cybernetyce, WNT, Warszawa 1967.

ciowa Y /liczbowa miara wielkości sygnału pojawiającego się na wyjściu/ jest pewną funkcją wielkości wejściowej X.

$$Y = F/X/ \qquad /1/$$

Postać funkcji F zależy oczywiście od konkretnych właściwości obiektu. Będziemy nazywali ją F U N K J Ą P R Z E J Ś C I A. Wielkość wyjściowa zależna jest, jak wynika z równania /1/ zarówno od właściwości obiektu /funkcja F/, jak i od sygnału wejściowego /argumentu X/. Chcąc w wygodnej postaci przedstawić same właściwości obiektu zapiszemy równanie /1/ w postaci:

$$Y = G . X \qquad /2/$$

Łatwo zauważyć, że człon G zastąpił funkcję F /./, a zamiast obliczania funkcji z sygnału X dokonujemy teraz mnożenia sygnału przez człon G. Nie wdając się w dość złożone dociekania matematyczne, można stwierdzić, że człon G będzie pewnym operatorem działającym na sygnał X i przekształcającym go na sygnał Y. Ponieważ sygnały X i Y są w ogólnym przypadku pewnymi funkcjami czasu /są przebiegami zachodzącymi w określonym czasie/, a układ może sygnały te całkować, przesuwać w czasie i dokonywać z nimi bardziej złożonych operacji, więc operator G może mieć niekiedy postać bardzo złożoną. Niezależnie jednak od stopnia komplikacji będzie on zawsze zależał wyłącznie od właściwości obiektu, można go więc przypisać obiektowi jako matematyczną formę zapisu jego właściwości przekształcania informacji /sygnałów/. Taki operator G, zależny wyłącznie od obiektu nazwiemy T R A N S M I T A N C J Ą obiektu.

5. Układ względnie odosobniony przedstawiać więc będziemy na rysunkach jako prostokąt, do którego dochodzi jedna lub więcej linii zakończonych strzałką i od którego odchodzi pewna ilość linii.

Linie dochodzące symbolizują wejścia układu, będziemy przy nich pisali symbole lub funkcje czasu, reprezentujące sygnały wejściowe /na przykład X lub X/t/ /.

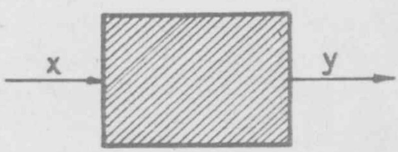
Linie wychodzące, reprezentujące wyjścia zaopatrzymy w podobne oznaczenia, dotyczące sygnałów wyjściowych. W samym prostokącie wpisujemy transmitancję układu. Prawidłowo narysowany UWO przedstawia rys. 4.

Rys.4 Transmitancja

Ponieważ układ przy tym sposobie notacji zastąpiony został rysunkowo blokiem, więc schematy takie jak rys. 4 nazywamy SCHEMATAMI BLOKOWYMI. Znając strukturę systemu i transmitancje jego poszczególnych układów można narysować schemat blokowy systemu.

6. Stosunkowo często zdarza się, że znamy sygnały wejściowe i sygnały wyjściowe, natomiast ich związek i wzajemna zależność jest nieznana. Jak na przykład określić transmitancję wiążącą spostrzeżenia wzrokowe z ruchami ręki u człowieka?

W przypadku niemożności określenia transmitancji rysujemy schemat blokowy podobny jak poprzednio /rys. 5/, ale zamiast



Rys. 5 "Czarna skrzynka"

wpisania transmitancji zakreskowujemy wnętrze bloku. Symbolizuje to naszą niewiedzę o wewnętrznych zależnościach w układzie, który zwykło się w tym przypadku nazywać CZARNĄ SKRZYNKĄ.

TEST 3

P. Układem jest zwykła waga uchylna /ze wskazówką/. Przyjmując, że wielkością wejściową jest ciężar ciała ważonego /Q/, a wyjściową wychylenie wskazówki /L/ określić postać funkcji /1/.

xxx

O. Ponieważ w dobrej wadze wychylenie jest proporcjonalne do ciężaru, więc

$$L = k\,Q,$$

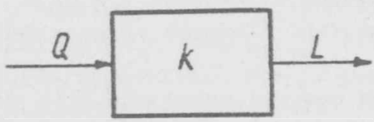
gdzie k jest współczynnikiem proporcjonalności, wyrażającym ilość działek na jednostkę ciężaru.

Złe rozwiązanie tego zadania kosztuje ponowne przeczytanie punktów 4. 5. i 6.

P. Dla układu z poprzedniego zadania określić transmitancję i schemat blokowy.

xxx

O. Zależność proporcjonalna ma tę właściwość, że współczynnik proporcjonalności jest czynnikiem, przez który mnożony jest sygnał wejściowy dla otrzymania sygnału wyjściowego, spełnia więc rolę transmitancji. Schemat blokowy przedstawia rysunek 6. W przypadku trudności z zadaniem



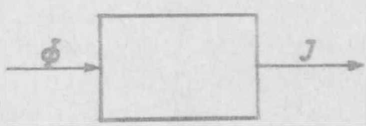
Rys. 6 Schemat blokowy

należy przeczytać jeszcze raz punkt 5.

P. Niech naszym UWO będzie fotokomórka elektryczna. Określić funkcję przejścia układu, wiedząc, że przy braku oświetlenia płynie przez fotokomórkę pewien mały prąd I_0 /prąd ciemny/, a przy oświetleniu fotokomórki strumieniem światła Φ_1 otrzymano z fotokomórki prąd I_1. Wejściem układu niech będzie strumień padającego światła Φ, a wyjściem prąd I.

xxx

O. Schemat układu podany jest na rysunku 7. Funkcja przej-



Rys. 7 Schemat blokowy fotokomórki

ścia ma postać:

$$I = \frac{I_1 - I_0}{\Phi_1}\Phi + I_0$$

P. Czy możliwe jest w poprzednim przykładzie określenie transmitancji? Dlaczego?

xxx

O. Niemożliwe, ponieważ zależności $I = F/\Phi/$ nie da się przedstawić w postaci iloczynu pewnego mnożnika i zmiennej Φ.

P. Jakie należałoby uczynić założenie, aby określenie transmitancji stało się możliwe? Jaką miałaby ona wówczas postać?

xxx

O. Trzeba założyć, że prąd ciemny jest pomijalnie mały w porównaniu do całego prądu z fotokomórki:

$$I_0 \ll I$$

Transmitancja ma wówczas postać:

$$G = \frac{I_1}{\Phi_1} \qquad \text{ponieważ} \qquad I = \frac{I_1}{\Phi_1}\, \Phi$$

Założenie to jest technicznie uzasadnione, ponieważ w rzeczywistości prąd ciemny dobrze wykonanej fotokomórki jest niewielki. Zastosowany sposób określenia transmitancji nosi nazwę linearyzacji, ponieważ doprowadziliśmy do operatora G mającego postać mnożenia przez stałą. Niebawem poznamy inny sposób znajdowania transmitancji, zwany metodą przyrostową. Zanim to jednak nastąpi trzeba podsumować zadanie o fotokomórce:

Jeśli rozwiązałeś je bezbłędnie lub pomyłki były nieistotne, możesz nie czytać następnych pytań i przystąpić od razu do zadania o zużyciu energii elektrycznej. W przeciwnym razie musisz dalej studiować systematycznie.

P. Niech naszym układem względnie osodobnionym będzie wydział produkcyjny przedsiębiorstwa. Wielkością wejściową będzie ilość dostarczanego surowca S, zaś wyjściową wartość towarów zgromadzonych w magazynie po zakończeniu produkcji W. Zanim rozpoczęto produkcję w magazynie zgromadzone były narzędzia na sumę W_0 zł. Po przerobieniu ilości surowca S_1 sumaryczna wartość towarów w magazynie wynosiła W_1. Zakładamy, że środki produkcji nie uległy amortyzacji i ich wartość nie uległa zmianie. Analogicznie, jak w przykładzie poprzednim, określić funkcję, ocenić możliwości wprowadzenia pojęcia transmitancji i po poczynieniu odpowiednich założeń określić transmitancję.

xxx

xxx

O. Funkcja przejścia ma postać:

$$W = \frac{W_1 - W_o}{S_1} S + W_o$$

Transmitancję można wprowadzić, zakładając /co jest nieco mniej uzasadnione, niż w poprzednim przypadku/, że W_o stanowi pomijalnie małą część całkowitej wartości W. Wówczas

$$W = \frac{W_1}{S_1} S \quad \text{oraz} \quad G = \frac{W_1}{S_1}$$

Przykład jest celowo tak dobrany, aby był w pełni analogiczny do poprzednio omówionego. Jeśli jednak nie odniosłeś sukcesu, proponujemy zapoznanie się z książką: A.Lerner "Zarys cybernetyki".

P. Niech układem względnie odosobnionym będzie odbiornik radiowy. Jako wejście rozpatrywać będziemy gałkę regulatora siły głosu, a jako wyjście głośnik /wielkością wejściową będzie kąt położenia gałki, a wyjściową siła głosu/. Jakie założenia trzeba przyjąć, aby wprowadzić pojęcie transmitancji? Jaką ma ona postać?

xxx

O. Trzeba założyć, że w rozpatrywanym przedziale czasu nie zmienia się intensywność sygnału docierającego do anteny. Założenie to uzależnia nasze rozważania od czynników pozaukładowych, takich jak siła propagacji,radiostacji i stan atmosfery. Czynniki takie zwykle nazywamy Z A K Ł Ó C E N I A M I. Dodatkowo potrzebne jest założenie, że zakres zmian rozpatrywanych wielkości jest daleki od ograniczeń, to znaczy, że gałka nie jest dokręcona "do oporu" i że głośnik ma jeszcze rezerwę mocy. Przy tych ograniczeniach możemy przyjąć, że siła głosu jest proporcjonalna do położenia gałki i napisać:

$$I = q\,A,$$

gdzie q jest poszukiwaną transmitancją, I - intensywnością dźwięku, zaś A kątem położenia gałki. Dla ciekawych warto dodać, że nasze subiektywne odczucie głośności nie pokrywa się z natężeniem dźwięku wyrażanym mocą głośnika; pomiędzy wielkościami tymi istnieje zależność logarytmiczna, powodująca, że dźwięk 10 razy silniejszy słyszymy jako zaledwie o stopień głośniejszy. Mechanizm ten, stanowiący obronę organizmu przed zbyt silnymi dźwiękami jest uwzględniany przy konstrukcji odbiorników przez wprowadzenie tzw. potencjometrów logarytmicznych do regulacji siły głosu.

P. Niech naszym UWO będzie system ekonomiczny, w którym wartość inwestycji I będzie proporcjonalna do wielkości dochodu narodowego brutto D. Ponadto istnieć będą pewne inwestycje stałe w wielkości I_0, pomijalnie małe przy zgrubnych obliczeniach.
Podać postać funkcji przejścia i transmitancji.

xxx

O. Funkcja przejścia $I = k\,D + I_0$, transmitancja wynosi k.
Poprawna odpowiedź w dwu ostatnich pytaniach upoważnia do odpowiedzi na dwa następne pytania, pomyłki kosztują ponowne czytanie punktów 4 i 5.

P. Celem wykrycia zależności pomiędzy zamożnością pojedynczej rodziny Q, a ilością zużywanej przez nią energii elektrycznej p wykonano badania statystyczne, których opracowane wyniki zestawiono poniżej:

zamożność Q	0	1	2	3	4	5	6	7	8
zużycie energii elektrycznej p	0	1	4	9	16	25	36	49	64

Traktujemy rodzinę jako czarną skrzynkę, /rys. 8a/, której wejściem jest zamożność Q /średni dochód na jednego członka rodziny/, a wyjściem całkowite zużycie energii elektrycznej p. Należy określić funkcję przejścia. Czy można podać sposób obliczenia transmitancji?

Rys. 8 Transmitancja zamożności rodziny

xxx

O. Jedyną zależnością matematyczną spełniającą wyniki pomiarów jest:

$$p = Q^2$$

Zależności tej nie da się w prosty sposób przedstawić w postaci transmitancji. O układach, których funkcje przejścia są niesprowadzalne do postaci transmitancji mówimy, że są N I E L I N I O W E.

P. Rozważmy następujący, dość teoretyczny problem: Weźmy pod uwagę rodzinę, której zamożność w danej chwili wynosi Q_o. Zamożność tej bez wątpienia odpowiada pewna wielkość zużycia energii elektrycznej p_o, przy czym

$$p_o = Q_o^2$$

Przypuśćmy, że w wyniku różnych czynników zamożność rodziny wykazuje niewielkie wahania wokół wymienionej wyżej wartości ustalonej tak, że pojawiają się m a ł e odchylenia

$$\Delta Q = Q_{aktualne} - Q_o$$

Odchylenia te powodują również m a ł e wahania w ilości zużywanej energii elektrycznej.

$$\Delta p = p_{aktualne} - p_o$$

Rozpatrzmy teraz UWO, którego wielkościami wejściowymi będą odchyłki zamożności Q, a wielkościami wyjściowymi odchyłki zużycia energii elektrycznej p. Jaka będzie funkcja przejścia takiego układu? Czy można wprowadzić transmitancję po przyjęciu rozsądnych założeń? Ile ona wynosi?

xxx

O. Z zależności

$$p_{aktualne} = Q^2_{aktualne}$$

otrzymujemy po podstawieniu i dokonaniu odpowiednich przeliczeń

$$\Delta p = Q_o^2 - p_o + 2Q_o \Delta Q + /\Delta Q/^2$$

czyli, po uwzględnieniu, że $p_o = Q_o^2$ mamy funkcję przejścia

$$\Delta p = 2Q_o \Delta Q + /\Delta Q/^2$$

Ponieważ wielkość ΔQ jest mała, więc $/\Delta Q/^2$ jest dużo mniejsze od $2Q_o \Delta Q$ i może być pominięta, dając transmitancję /rys.8b./

$$\Delta p = 2Q_o \Delta Q$$

Jeżeli nie rozwiązałeś zadowalająco tego zadania zapoznaj się z następnym zadaniem, jeśli dałeś sobie radę, przeczytaj od razu punkt 7.

P. Powiedzmy, że mierzymy głębokość studni wrzucając do niej kamień i mierząc czas upływający do chwili usłyszenia plusku. Jak wiadomo z fizyki spadanie kamienia opisuje wzór:

$$x = \frac{g}{2} t^2$$

w którym g oznacza przyspieszenie /przyrost prędkości na jednostkę czasu/, z jakim spadają wszystkie ciała w polu przyciągania ziemi, /wartość ta jest stała i wynosi około 10 $m.s^{-2}$/, x jest drogą, jaką przebył kamień, a t czasem spadania. Pomijając opór powietrza i czas rozchodzenia się dźwięku upadku można więc przyjąć, że głębokość studni wyraża się wzorem:

$$H = \frac{g}{2} T^2$$

gdzie H jest głębokością, a T czasem od chwili upuszczenia kamienia do chwili usłyszenia plusku. Załóżmy, że znamy głębokość studni w normalnych warunkach H_o i odpowiadający tej głębokości czas T_o. /Oczywiście $H_o = \frac{g}{2} T_o^2$ /. Głębokość studni ulega małym zmianom na skutek deszczu lub długotrwałej suszy oΔH, co powoduje zmiany wielkości czasu spadania o Δ T. Obliczyć transmitancję wiążącą przyrosty czasu spadania /wielkość wejściowa/ z przyrostami głębokości studni. /wielkość wyjściowa/. Ocenić, kiedy pomiar będzie dokładniejszy: przy małej czy dużej głębokości studni H_o?

xxx

O. Wychodząc z zależności:

$$H_o + \Delta H = \frac{g}{2} /T_o + \Delta T/^2$$

i uwzględniając, jak poprzednio, że $T_o \gg \Delta T$ otrzymamy

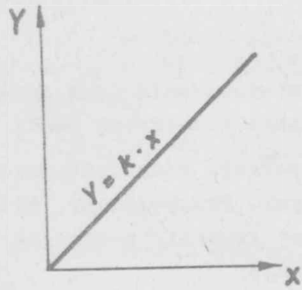
$$\Delta H = \frac{gH_o}{2} 2 \Delta T$$

czyli transmitancja wynosi gH_o. Pomiar jest tym dokładniejszy, im mniejsze przyrosty Δ H odpowiadają tym samym przyrostom ΔT, ponieważ przy pomiarze czasu możemy popełnić drobny błąd, który powinien możliwie mało wpływać na określenie głębokości. Wynika z tego, że dla minimalizacji błędu wartość transmitancji powinna być mała, co oznacza w efekcie, że pomiar będzie dokładniejszy w płytszej studni /mniejsze H_o/.

7. Podczas rozwiązywania zadań testowych pojawiło się kilka nowych pojęć, które można obecnie sformułować w sposób ścisły. Układy, dla których potrafimy określić transmitancję, nazywaliśmy L I N I O W Y M I, ponieważ funkcja przejścia dla tych układów miała ogólną postać:

wielkość wyjściowa = współczynnik . wielkość wejściowa

Jak wiemy z elementarnej matematyki, zależność tego typu daje się przedstawić w prostokątnym układzie współrzędnych /rys. 9 /, na którego osi pionowej odkładamy wielkości wyjściowe, a na poziomej wejściowe, w postaci linii prostej.



Rys. 9 Zależność liniowa

Nieco bardziej ogólnie można stwierdzić, że liniowymi będą wszystkie te obiekty, /układy względnie odosobnione/, dla których istnieje operator G spełniający dwa warunki: warunek a d d y t y w n o ś c i i warunek j e d n o r o d n o ś c i.

Addytywność oznacza, że działanie operatora na sumę sygnałów daje wynik identyczny z zesumowaniem wyników działania operatora na każdy z sygnałów z osobna:

$$G. /X + Y/ = GX + GY$$

Jednorodność oznacza, że działanie operatora na sygnał pomnożony przez dowolną l i c z b ę q jest identyczne z pomnożeniem wyniku działania operatora na sam sygnał przez tę liczbę q:

$$G /qX/ = q GX$$

Łatwo sprawdzić, że operator polegający na mnożeniu przez stałą jest operatorem liniowym:

$$k /X + Y/ = kX + kY \quad i \quad k/qX/ = kq X$$

TEST 4

P. Sprawdź, czy zależność:

$$Y = AX + B \qquad /B \neq 0/$$

gdzie A i B są stałymi, jest operatorem liniowym.

xxx

O. Nie jest

$$Y_1 = AX_1 + B \quad Y_2 = AX_2 + B \quad Y = A/X_1 + X_2/+B \neq Y_1 + Y_2$$

oraz

$$Y = AqX + B \neq q\left(AX + B\right)$$

W przypadku niewykonania tego zadania musisz koniecznie jeszcze raz uważnie przeczytać punkt 7.

P. Udowodnij, że operacje różniczkowania i całkowania są operatorami liniowymi. Przypominamy, że różniczkowanie /znajdowanie pochodnej funkcji/ polega na znajdowaniu granicy ilorazu różnicowego

$$\frac{dY/x/}{dx} = \lim_{h \to o} \frac{Y/x + h/ - Y/x/}{h}$$

a całkowanie jest w stosunku do różniczkowania operacją odwrotną.

xxx

O. Wprowadzamy dwie funkcje zmiennej x:$Y_1/x/$ i $Y_2/x/$. Mamy wtedy:

$$\frac{d\ Y_1/x/+/Y_2/x/}{dx} = \lim_{h \to o} \frac{Y_1/x+h/+Y_2/x+h/-Y_1/x/-Y_2/x/}{h} =$$

$$= \lim_{h \to o} \frac{Y_1/x+h/-Y_1/x/}{h} + \lim_{h \to o} \frac{Y_2/x+h/-Y_2/x/}{h} = \frac{dY_1/x/}{dx} + \frac{dY_2/x/}{dx}$$

oraz

$$\frac{d\ /k\ Y/x//}{dx} = \lim_{h \to o} \frac{k\ Y/x+h/-kY/x/}{h} = k \lim_{h \to o} \frac{Y/x+h/-Yx/}{h} =$$

$$= k\ \frac{dY/x/}{dx}$$

Liniowość operatora różniczkowania została wykazana.

Liniowość operatora całkowania dowiedziemy nie wprost. /Można oczywiście dowodzić inaczej, nie jest to błędem, my jednak chcemy udowodnić więcej, niż wymaga pytanie testu, a mianowicie, że operator odwrotny do operatora liniowego też jest liniowy/.

Niech $\int y/x/\ dx = Y/x/$ /pomijamy dla prostoty stałe całkowania/. Wówczas $y/x/ = \frac{dY/x/}{dx}$ /z definicji całki/. Sprawdźmy więc, czy

$$\int\left(y_1/x/ + y_2/x/\right) dx \neq \int y_1/x/\ dx + \int y_2/x/\ dx = Y_1/x/ + Y_2/x/.$$

ale z liniowości różniczkowania wynika, że

$$\frac{d\left(Y_1/x/ + Y_2/x/\right)}{dx} = y_1/x/ + y_2/x/$$

co daje sprzeczność dowodzącą addytywności całki. Jednorodności dowiedziemy natychmiast

$\int Ky/x/\ dx \neq K\int y/x/\ dx$ implikuje, że $K\,\frac{dY/x/}{dx} \neq \frac{d\ KY/x/}{dx}$ co jest sprzeczne.

Jeżeli nie zdołałeś przeprowadzić dowodu powinieneś przeczytać ponownie punkt 7. Jeżeli natomiast ponadto podczas czytania odpowiedzi miałeś wątpliwości, dobrze zrobisz, jeśli rozwiejesz je przy pomocy jakiegokolwiek podręcznika rachunku różniczkowego np. F.Leja "Rachunek różniczkowy i całkowy" [5].

Czy operacja

$$Y = X^2$$

jest liniowa ?

xxx

Nie jest :

$X_1 = X_1^2 \quad Y_2 = X_2^2 \quad Y = /X_1 + X_2/^2 \neq X_1^2 + X_2^2$

---

[5] F.Leja: Rachunek różniczkowy i całkowy ze wstępem do równań różniczkowych, PWN, Warszawa 1969.

i podobnie

$$Y = /qX/^2 = q^2 X^2 \neq q X^2$$

Nieprawidłowa odpowiedź dowodzi, że albo nieuważnie czytasz, albo jesteś zmęczony. Zrób krótką przerwę, a potem przeczytaj wszystko, poczynając od punktu 7 jeszcze raz.

8. W pytaniach testu 3, poprzedzających punkt 7 wskazano, że niekiedy układ N I E L I N I O W Y może zostać zlinearyzowany przez przyjęcie odpowiednich założeń. Poznaliśmy dwie metody LINEARYZACJI:pomijanie nieistotnych składników i linearyzację wokół pewnego ustalonego punktu, przeprowadzaną dla małych przyrostów. Obie te metody są chętnie i szeroko stosowane w cybernetyce przy określaniu funkcji przejścia, czy transmitancji układów. Proces określania funkcji przejścia dla układu, który uprzednio był "czarną skrzynką", nazywamy I D E N T Y F I K A C J Ą . Podczas indentyfikacji niemal zawsze konieczne jest pomijanie wpływu całego szeregu czynników, które są albo niekontrolowane, albo mają trudny do uchwycenia wpływ na pracę układu. Przykład tego typu procedury mieliśmy w teście dotyczącym funkcji przejścia odbiornika radiowego. Umówimy się, podobnie jak w teście, nazywać w s z y s t k i e n i e k o n t r o l o w a n e w i e l k o ś c i w e j ś c i o w e d o u k ł a d u Z A K Ł Ó C E N I A M I. Jest to bardzo ważne pojęcie, wielokrotnie używane w dalszych rozważaniach.

TEST 5

P. Jeżeli naszym układem względnie odosobnionym będzie piec, wielkością wejściową ilość dostarczanego paliwa W, a wyjściową temperatura pieca T, to jakie potrafisz wymienić zakłócenia?

xxx

O. Zakłóceniami będą wszystkie czynniki wpływające na wielkość wyjściową, a nie będące wielkościami wejściowymi, a zatem:

1/ wartość opałowa paliwa,

2/ ilość powietrza docierająca do pieca,

3/ wilgotność paliwa,

4/ szczelność pieca /np. otwarcie drzwiczek jest zakłóceniem/,

5/ ilość ciepła odbierana od ścian pieca /np. przez polanie go wodą ilość tę można wydatnie zwiększyć/.

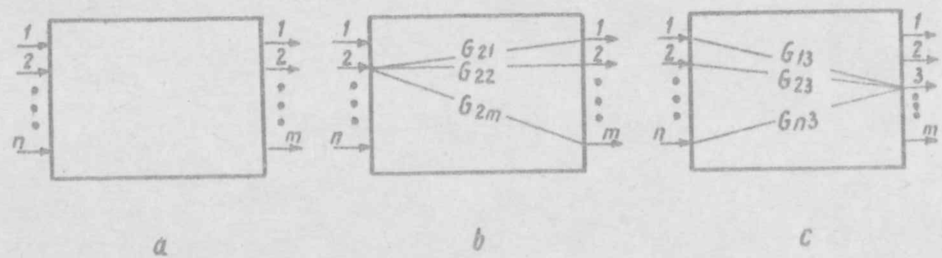
Jeśli wymieniłeś co najmniej 2 zakłócenia możesz już czytać punkt 9, w przeciwnym przypadku zrób następne zadanie.

P. Co stanowi zakłócenia w UWO jakim jest samochód /wejście - kierownica, wyjście - kierunek jazdy/?; Szybkością jazdy nie interesujemy się.

xxx

O. Nierówności drogi, boczny wiatr, niewłaściwe ustawienie lub wyważenie kół.

9. Wszystkie dotychczas rozważania prowadzone były dla układów o jednym wejściu i jednym wyjściu, czyli jednowymiarowym. Analiza układów o wielu wejściach i wielu wyjściach jest dużo bardziej skomplikowana. Rozważając układ o n wejściach i m wyjściach /rys. 10a / musimy brać pod uwagę nm transmitancji /lub funkcji przejścia/, a mianowicie wszystkie połączenia pierwszego wejścia z każdym z m wyjść /razem m funkcji/, wszystkie połączenia drugiego wejścia z każdym z m wyjść /razem z pierwszymi 2m funkcji/, i tak dalej, aż do n-tego wejścia, co daje w sumie właśnie tych nm połączeń. Na rysunku 10b przedstawiono wszystkie połączenia biorące początek w drugim wejściu, a na rysunku 10c wszystkie połączenia dochodzące do trzeciego wyjścia. T r a n s m i t a n c j ę łączącą k-te wejście z j-tym wyjściem oznaczamy G_{kj}. W związku z tym wszystkie połączenia na rysunku 10b mają pierwszy indeks 2, a wszystkie połączenia na rysunku 10c mają drugi indeks 3.



Rys. 10 Transmitancja układu wielowymiarowego

TEST 6

P. Niech UWO będzie zakładem gastronomicznym, wielkościami wejściowymi niech będą - ilość sprzedanych potraw X_1 i ilość sprzedanego alkoholu X_2, natomiast wielkością wyjściową - całkowity zysk Y_1.
Czym będą transmitancje G_{11} i G_{21}?

O. G_{11} jest marżą za potrawy, a G_{21} marżą za alkohol. Zysk jest w przypadku równoczesnej sprzedaży potraw i alkoholu sumą:

$$Y_1 = G_{11} X_1 + G_{21} X_2 .$$

Układ jest liniowy.

P. Niech UWO będzie zakładem produkcyjnym. Wielkością wejściową X_1 będzie plan produkcji wyrażony w zł., a wielkościami wyjściowymi ilości wyprodukowanego towaru pierwszego Y_1, towaru drugiego Y_2 i planowane koszty remontów Y_3.
Określić transmitancje G_{11}, G_{12}, oraz G_{13}.

xxx

O. Transmitancją G_{11} jest udział towaru pierwszego w planowanej produkcji, G_{12} udział towaru drugiego, a G_{13} udział remontów. Oczywiście transmitancja G_{13} jest ujemna. Wielkości wyjściowe są określone równaniami:

$$Y_1 = G_{11} X_1$$
$$Y_2 = G_{12} X_1$$
$$Y_3 = G_{13} X_1$$

Jeżeli podczas odpowiadania miałeś wątpliwości, przeczytaj jeszcze raz punkt 9 zanim przystąpisz do czytania następnego pytania.

P. Niech UWO będzie pewien obszar uprawny. Wielkościami wyjściowymi będą przyrosty plonów żyta /Y_1/ i pszenicy /Y_2/, uzyskiwane dzięki stosowaniu nawozów sztucznych: azotowych /X_1/, potasowych /X_2/ i fosforowych /X_3/.

Doświadczalnie stwierdzono, że przyrost plonów żyta wynosi "a" procent na kilogram nawozu azotowego, "b" procent na kilogram nawozu potasowego i "c" procent na kilogram nawozu fosforowego.

Idetyczne dane dla pszenicy wynoszą: "d" dla nawozów azotowych, "e" dla potasowych i "f" dla fosforowych. Określić wszystkie możliwe transmitancje.

xxx

Transmitancje wynoszą:

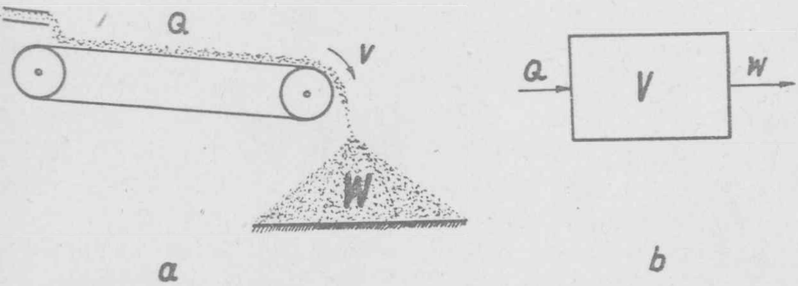
$G_{11} = a, G_{21} = b, G_{31} = c, G_{12} = d, G_{22} = e, G_{32} = f.$

Zrobiono tu silne założenia, że przyrosty plonów wywołane stosowaniem poszczególnych grup nawozów, sumują się, oraz że przyrost plonów jest proporcjonalny do dawki nawozu. Ich poprawność nie będzie tu dyskutowana, gdyż nie chodziło o analizę plonów, lecz o przykład stosowania metod cybernetycznych.

10. Analiza układów wielowymiarowych jest mocno skomplikowana nawet w przypadku, kiedy daje się określić wszystkie transmitancje i kiedy wypadkowa wielkość wyjściowa układu jest sumą wielkości wyjściowych pochodzących od poszczególnych wejść układu. Trudności jednak rosną niezwykle szybko przy dopuszczeniu nieliniowych funkcji przejść. Poza możliwościami nieliniowych zależności, wiążących odpowiednie wejście z rozpatrywanym wyjściem, należy uwzględniać wówczas nieliniowe zależności pomiędzy wejściami w postaci iloczynów lub bardziej złożonych funkcji wielu zmiennych. Rozpatrzmy następujący przykład:

TEST 7

P. Niech układem względnie odosobnionym będzie transporter taśmowy przenoszący materiał sypki /rys. 11/. Niech wejściowymi wielkościami będą: ilość materiału /w kilogramach na metr/, przypadająca na jednostkę długości Q i prędkość biegu taśmy /w metrach na minutę/ V, zaś wielkością wyjściową niech będzie szybkość gromadzenia się materiału sypkiego na końcu przenośnika W /w kilogramach na minutę/.



Rys. 11 Transmitancja przenośnika taśmowego

Wyznaczyć funkcję przejścia układu. Czy, i w jakich okolicznościach można by tu mówić o transmitancji?

xxx

0. Ilość materiału gromadzona na końcu przenośnika w jednostce czasu jest związana z szybkością przenośnika i jego zapełnieniem zależnością nieliniową:

$$W = Q \cdot V$$

Wpływu jednej wielkości wejściowej nie da się rozpatrywać w oderwaniu od drugiej, nie można zatem mówić o transmitancjach G_{11} i G_{21}. Można jednak przyjąć założenie, że na przykład prędkość przenośnika jest stałą, natomiast zmienny stopień zapełnienia przenośnika Q wywiera wpływ na wyjście układu /możliwe jest również założenie odwrotne/. Jednakże wówczas mamy już do czynienia z układem jednowymiarowym, gdyż drugie wejście /prędkość taśmy/ przestało przenosić informację przy raz na stałe ustalonej prędkości. Prędkość ta stanowi wówczas liczbową wartość transmitancji układu /rys. 11b/.

Niech UWO będzie zakład produkcyjny, którego wejściami są liczba zatrudnionych robotników N i średnia wydajność pracy W, natomiast wyjściem jest wartość uzyskanej produkcjiP.

Jakie transmitancje można tu wyróżnić?
Przyjmujemy, że P = N . W

xxx

O. Można wyróżnić transmitancję WP i transmitancję NP .[6/]
Transmitancję NP otrzymujemy, gdy założymy, że wydajność pracy W jest stała. Ta stała W pełni wówczas rolę transmitancji. Transmitancję WP otrzymujemy, gdy ilość zatrudnionych jest stała, natomiast zmienia się wydajność. Transmitancją jest wówczas stała liczba N.

11. Dla usystematyzowania zapisu transmitancji układów wielowymiarowych wprowadzimy notację wektorową dla oznaczeń wielkości wejściowych i wyjściowych i macierzową dla oznaczeń transmitancji.

Najpierw krótkie przypomnienie z matematyki:

Wektorem w przestrzeni n-wymiarowej możemy nazywać uporządkowany zestaw n liczb, z których pierwszą możemy utożsamiać z pierwszą współrzędną końca wektora, /z wielkością odcinka odłożonego na pierwszej osi prostokątnego układu współrzędnych w tej przestrzeni/, drugą - z drugą współrzędną i tak dalej, aż do n-tej współrzędnej. Początek wektora umieszczamy w początku układu współrzędnych.

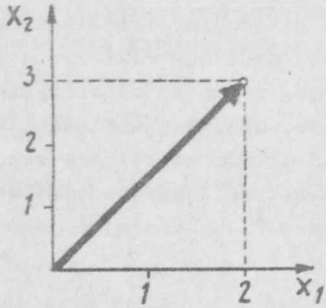
TEST 8

P. Narysuj w układzie współrzędnych na płaszczyźnie wektor [2,3]. Początek wektora połącz z końcem odcinkiem ze strzałką przy końcu wektora.

xxx

O. Rozwiązanie przedstawia rysunek 12. Jeżeli popełniłeś błąd przeczytaj koniecznie jeszcze raz punkt 11, a następnie odpowiedz na następne pytanie, w przeciwnym wypadku czytaj od razu punkt 12.

6/ Transmitancja NP oznacza, że jako wielkość /jedyną/ przyjęto N, a jako wyjściową P. Pozostałe oznaczenia są analogiczne.

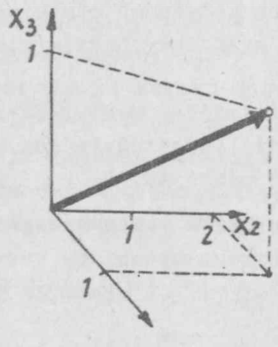


Rys. 12 Wektor na płaszczyźnie

P. Narysuj /stosując perspektywę/ wektor $[1, 2, 1]$.

xxx

O. Rozwiązanie podaje rysunek 13. W przypadku trudności na-



Rys. 13 Wektor w przestrzeni

leży zapoznać się z odpowiednim podręcznikiem algebry liniowej lub geometrii analitycznej np. W.Smirnow "Matematyka wyższa", Tom III cz. pierwsza.[7]

12. W postaci uporządkowanego zbioru liczb /wektora/ podawać będziemy aktualne wartości wielkości wejściowych

7/ W.Smirnow: Matematyka wyższa, PWN, Warszawa 1965.

odpowiednio najpierw na pierwszym wejściu /pierwsza współrzędna/, następnie na drugim, trzecim i na końcu na n-tym. Wektor tak utworzony nazwiemy wektorem wejściowym i oznaczymy go X, uważając, że ma on współrzędne /X_1, X_2,...,X_n/. Zupełnie identycznie, uważając wielkości występujące na kolejnych wyjściach układu za kolejne współrzędne wektora w m-wymiarowej przestrzeni, wprowadzimy wektor wyjściowy, który oznaczymy przez Y, wyróżniając jego kolejne współrzędne /Y_1, Y_2, ..., Y_m/.

TEST 9

P. Jeżeli układ wielowymiarowy ma osiem wejść i trzy wyjścia, to ile współrzędnych będzie miał wektor wyjściowy X, a ile wektor wyjściowy Y?

xxx

O. Wektor wejściowy będzie miał osiem współrzędnych, ponieważ na kolejnych współrzędnych podawane będą wartości wielkości wejściowych kolejnych ośmiu wejść, z tych samych powodów wektor wyjściowy będzie miał trzy współrzędne.

Jeżeli nie odpowiedziałeś musisz jeszcze raz czytać punkt 12., a następnie odpowiedzieć na następne pytanie, inaczej zaraz czytaj punkt 13.

P. Określ ilość współrzędnych wektora wejściowego i wyjściowego we wszystkich trzech pytaniach testu 6, następujących po punkcie 9.

xxx

1/ wektor wejściowy 2 współrzędne, wektor wyjściowy 1 współrzędna

2/ wektor wejściowy 1 współrzędna, wektor wyjściowy 3 współrzędne

3/ wektor wejściowy 3 współrzędne, wektor wyjściowy 2 współrzędne

13. Przy spełnionym warunku liniowości układu kolejne

składowe wektora wyjściowego można otrzymać przez zsumowanie wszystkich składowych wektora wejściowego, pomnożonych przez odpowiednie transmitancje G_{jk}. Można to zapisać w postaci układu równań:

$$
\begin{aligned}
Y_1 &= G_{11} X_1 + G_{21} X_2 + G_{31} X_3 + \dots + G_{n1} H_n, \\
Y_2 &= G_{12} X_1 + G_{22} X_2 + G_{32} X_3 + \dots + G_{n2} X_n, \qquad /3/ \\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
Y_m &= G_{1m} X_1 + G_{2n} X_2 + G_{3m} X_3 + \dots + G_{nm} X_n.
\end{aligned}
$$

Współczynniki tych równań G_{kl} stanowią odpowiednio oznaczone /zgodnie z oznaczeniami przyjętymi w punkcie 9./ transmitancje łączące odpowiednie wyjście i wejście. Utworzymy teraz macierz współczynników, będącą prostokątną tablicą liczb o wymiarach n kolumn i m wierszy. Pomnożenie prawostronne tej macierzy przez wektor wejściowy X daje prawe strony równań /3 /. Przypominamy, że prawostronne mnożenie macierzy przez wektor oznacza mnożenie macierzy przez wektor następujący po macierzy. Działanie to jest wykonalne tylko w przypadku, kiedy ilość współrzędnych wektora jest równa ilości kolumn macierzy. Wynikiem jest kolumna sum, której k-ta współrzędna powstaje przez zesumowanie iloczynu pierwszego elementu k-tego wiersza macierzy przez pierwszą współrzędną wektora, plus drugi element k-tego wiersza razy druga współrzędna wektora i tak aż do n-tych składowych.
Wynika z tego, że równanie /3/ można przedstawić w postaci macierzowej:

$$
\begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ Y_m \end{bmatrix} = \begin{bmatrix} G_{11} & G_{21} & G_{31} & \dots & G_{n1} \\ G_{12} & G_{22} & G_{32} & \dots & G_{n2} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ G_{1m} & G_{2m} & G_{3m} & \dots & G_{nm} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ X_n \end{bmatrix} \qquad /3a/
$$

Zestawienie wzorów /3 i 3a/ powinno rozproszyć resztę wątpliwości co do sposobu mnożenia macierzy przez wektory i istoty przyjętej notacji. Ewentualne niejasności warto usunąć w oparciu o wskazaną wyżej literaturę.

Reasumując, po wprowadzeniu n-wymiarowego wektora wejściowego X, m-wymiarowego wektora wyjściowego Y i liczącej n kolumn i m wierszy macierzy transmitancji G możemy pełny opis liniowego układu wielowymiarowego przedstawić w postaci równania /3a/, które krócej zapiszemy relacją macierzową:

$$Y = [G]\, X \qquad /4/$$

Widać, że macierz G w relacji tej pełni rolę transmitancji, rozsądnym więc będzie nazwanie jej transmitancją macierzową.

TEST

P. Pomnóż prawostronnie macierz A przez wektor B:

$$A = \begin{bmatrix} 2 & 3 & 5 \\ 1 & 0 & 9 \end{bmatrix} \qquad B = \begin{bmatrix} 1 \\ 6 \\ 2 \end{bmatrix}$$

xxx

O. W wyniku otrzymujemy wektor $[A]\, B = \begin{bmatrix} 30 \\ 19 \end{bmatrix}$

Prawidłowa odpowiedź zwalnia od konieczności robienia następnego zadania.

P. Pomnóż prawostronnie macierz G przez wektor X, jeśli

$$G = \begin{bmatrix} g_{11} & g_{21} \\ g_{12} & g_{22} \end{bmatrix} \qquad X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

xxx

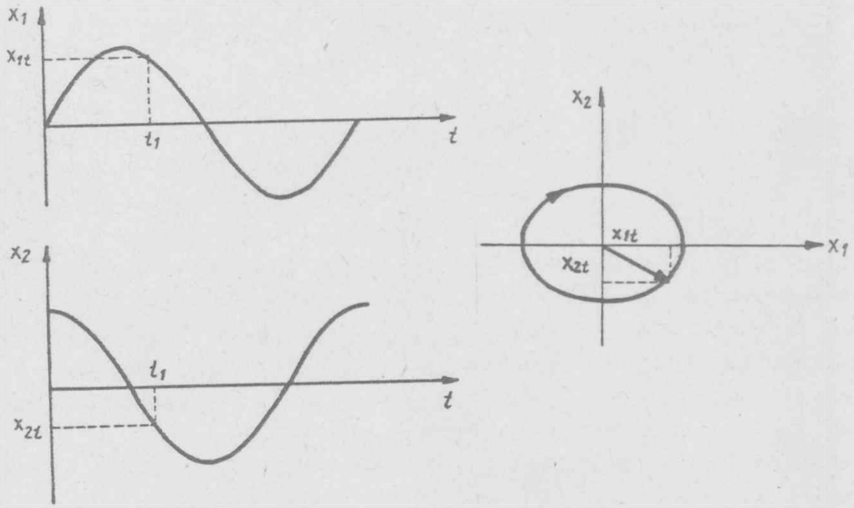
O. Wynik wynosi $[G]\, X = \begin{bmatrix} g_{11}X_1 + g_{21}X_2 \\ g_{12}X_1 + g_{22}X_2 \end{bmatrix}$.

W razie złej odpowiedzi należy jeszcze raz przeczytać punkt 13.

3. Przestrzeń stanów. Trajektorie.

1. Stwierdziliśmy uprzednio dwa fakty: po pierwsze, zbiór wielkości wyjściowych występujących odpowiednio na wszystkich wyjściach w danej chwili czasowej można przedstawić w postaci wektora wyjściowego, albo, innymi słowy, punktu w pewnej przestrzeni. Przestrzeń ta ma tyle współrzędnych /jest tylo-wymiarowa/, ile jest wyjść układu; po drugie, ponieważ wielkości wyjściowe są zwykle zmienne w czasie, zmieniać się będą w czasie również i współrzędnè punktu w rozpatrywanej przestrzeni, co obserwować będziemy jako ruch punktu. Przyjmując dość rozsądne założenie, że wielkości wyjściowe zawierają pełną informację o aktualnym stanie układu /właśnie w ten sposób należy je wybrać, aby zawierały pełną informację/, zgodzimy się nazwać przestrzeń, o której mowa, P R Z E S T R Z E N I Ą S T A N Ò W. Zmiany wielkości wyjściowych powodują, że położenie punktu, obrazującego stan aktualny będzie się zmieniało. Punkt ten będzie wędrował w przestrzeni stanów po pewnym torze, który nazwiemy T R A J E K T O R I Ą.

2. Opis zachowania się układu w oparciu o przestrzeń stanu i trajektorie wbrew pozorom jest dużo prostszy, niż opis zachowania się układu poprzez podanie przebiegów czasowych na wszystkich wyjściach układu. Zanalizujmy rysunek 14a. Przedstawia on czasowe wykresy dwu wyjść układu układu względnie odosobnionego. Przyznajmy, że stosunkowo niewiele potrafimy powiedzieć o związkach pomiędzy tymi przebiegami. Przenieśmy się teraz do przestrzeni stanu, której współrzędnymi są rozpatrywane wyjścia. Narysujemy trajektorię nanosząc punkt po punkcie wartości obydwu współrzędnych dla kolejnych chwil czasu /na rysunku pokazano sposób naniesienia punktu odpowiadającego chwili t_1/. Teraz, na podstawie trajektorii przedstawionej na rysunku 14b potrafimy powiedzieć następujące rzeczy: Oba przebie-



Rys. 14 Konstrukcja trajektorii

gi są periodyczne i posiadają ten sam okres, ponieważ trajektoria jest linią zamkniętą /po upływie pewnego czasu powracamy do tych samych punktów na trajektorii, czyli wartości zarówno x_1, jak i x_2 okresowo powtarzają się/. Amplituda przebiegu x_1 jest większa niż amplituda przebiegu x_2, ponadto amplitudy są przybierane na przemian: wielkość x_1 jest maksymalna dokładnie wtedy, kiedy x_2 jest równa zero i na odwrót. W miarę upływu czasu punkt porusza się po trajektorii zgodnie z ruchem wskazówek zegara.

TEST

P. Narysuj trajektorię dla UWO, którego dwa wyjścia generują przebiegi opisane równaniami:

$x_1 = A \sin t$ A B - stałe, dodatnie.

$x_2 = B \cos t$

WSKAZÓWKA: Nie rysować przebiegów x_1 i x_2 w funkcji czasu, tylko wstawiając kolejno wartości czasu t w niedużych odstępach obliczać wartości x_1 i x_2 i wyrysować kolejne punkty na płaszczyźnie stanów.

xxx

O. Poprawne rozwiązanie stanowi elipsa o dłuższej osi wynoszącej 2A i skierowanej poziomo, krótszej osi wynoszącej 2B i skierowanej pionowo oraz o środku w początku układu.

P. Niech podobnie jak w zadaniu poprzednim UWO ma dwa wyjścia których stany w kolejnych momentach czasu opisują równania:

$$x_1 = \sin t$$

$$x_2 = \sin 2t$$

Jakie wnioski można wyciągnąć z trajektorii?

xxx

O. Trajektoria przypomina poziomo położoną ósemkę. Widać z niej, że oba przebiegi wyjściowe zerują się równocześnie /trajektoria dwukrotnie przecina początek układu/, że są oba periodyczne /trajekria jest linią zamkniętą/ i że przebieg x_2 przyjmuje wartości maksymalne dwa razy częściej niż przebieg x_1.

W przypadku napotkania trudności należy zapoznać się z książką: A.Lerner "Zarys cybernetyki".

3. Ponieważ na przebieg czasowy wyjściowych wielkości układu mają wpływ /poprzez odpowiednie funkcje przejścia/ wielkości wejściowe, więc ruch w przestrzeni stanu jest sterowany przez wielkości wejściowe układu. Sterowanie to może polegać na wyborze jednej z kilku możliwych trajektorii, wyborze kierunku i sposobu poruszania się po zadanej trajektorii lub może pozwalać na zupełnie dowolne ruchy w przestrzeni stanów, czyli na swobodny wybór kształtu trajektorii. Z punktu widzenia najszerszej teorii ten ostatni przypadek jest najbardziej interesujący, jednakże

będziemy starali się w niniejszej książce ograniczać do przypadków, kiedy trajektoria jest ustalona, gdyż upraszcza to rozważania.

TEST

Niech UWO ma jedno wejście i dwa wyjścia. Wielkości wyjściowe na wyjściu pierwszym /Y_1/ mogą przyjmować wyłącznie wartości 0, +1, i +2. Wielkości wyjściowe na drugim wyjściu /Y_2/ mogą przyjmować wyłącznie wartości +1, +3 i +5. Na jednym wejściu /X/ mogą się pojawiać sygnały o wartościach +1, -1, -2, przy czym sygnał +1 powoduje zwiększenie wielkości wyjściowej Y_1 do następnej możliwej wartości /o ile ruch jest niemożliwy pojawienie się wielkości wejściowej nie wywołuje skutku/, zaś wielkość wyjściowa Y_2 pozostaje niezmieniona. Sygnał -1 na wejściu powoduje zmniejszenie /o ile to możliwe/ wielkości wyjściowej Y_1 do następnej możliwej wartości, bez zmiany wartości wielkości wyjściowej Y_2. Sygnały +2 i -2 powodują analogiczne zmiany wielkości wyjściowej Y_2, pozostawiając bez zmian wielkość wyjściową Y_1, a mianowicie sygnał +2 powoduje zwiększenie wielkości Y_2 do następnej możliwej / o ile nie jest już osiągnięte maksimum/, zaś sygnał -2 powoduje zmniejszenie Y_2.

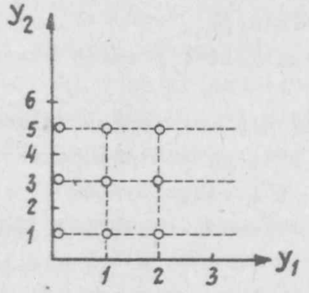
Narysuj przestrzeń stanu opisanego układu, sytuując oś Y_1 poziomo, a oś Y_2 pionowo. Zaznacz wszystkie punkty, w których układ może się znajdować.

xxx

O. Porównaj rysunek 15, w przypadku wystąpienia różnic koniecznie jeszcze raz przeanalizuj zadanie i ewentualnie przeczytaj punkty 1, 2 i 3.

P. Jaki kształt będzie miała w poprzednim zadaniu trajektoria gdy układ początkowo znajdował się w punkcie 0,1 i podano sygnały:

+1, +1, +1, +1, -1, -1, -1, -1?



Rys.15 Przestrzeń stanu

xxx

O. Trajektoria będzie linią poziomą położoną na wysokości 1 nad osią Y_1. Układ porusza się po tej trajektorii najpierw na prawo, a potem na lewo i wraca do punktu wyjścia.

Jeśli odpowiedź twoja była prawidłowa i pełna, pomiń następne pytanie, w przeciwnym przypadku odpowiedz na nie.

P. Jak wyglądać będzie trajektoria tego układu, jeśli w chwili początkowej układ znajduje się w stanie 1,3 , czyli na wyjściu Y_1 jest wartość 1 a na wyjściu Y_2 wartość 3, a następnie otrzymuje serię sygnałów na wejściu: -1, +1, +1, -1, +1, +1?

xxx

O. Trajektoria jest linią poziomą leżącą na wysokości 3 nad osią Y_1. Układ wykonał ruch w lewo, w prawo, znowu w lewo i jeszcze raz w prawo kończąc ruch w punkcie 2,3 .

P. Jaki kształt mają trajektorie powstające przy podawaniu sygnałów złożonych z ciągów +2 i -2?

xxx

O. Są liniami pionowymi, ich położenie na płaszczyźnie stanu zależy od położenia punktu, w którym był układ w chwili początkowej, zaś sposób ruchu zależy od sposobu podawania sygnałów +2 i -2 na wejściu.

Jeżeli nie odpowiedziałeś, przerób następujący przykład:

Narysować trajektorię dla omawianego układu, powstającą w wyniku podania do układu, znajdującego się w chwili początkowej w punkcie 1 , ciągu sygnałów: +2, +2, -2, -2, +2, -2, -2. Wykonaj to samo dla układu znajdującego się w chwili początkowej w stanie 2,5 .

P. W jakim stanie końcowym znajdzie się układ, jeśli na początku znajdował się w stanie 1,3 , a następnie poddano sekwencję sygnałów wejściowych: -1, -2, -1,+2, +2, -1, +2, -2, +1, +2, +1, +1, -2?

xxx

O. Punktem końcowym jest 2,3 .
Jeśli otrzymałeś inny wynik prześledź jeszcze raz zachowanie układu, aż do uzyskania prawidłowej odpowiedzi.

4. Oryginał i model. Izomorfizm a homeomorfizm. Modele matematyczne.

1. Dwa układy względnie odosobnione mogą być pod różnymi względami podobne. Podobieństwo to może być czysto zewnętrzne, np. polegające na podobieństwie wielkości, koloru czy kształtu. Jak łatwo się domyślić, ten rodzaj podobieństwa będzie dla nas najmniej interesujący. Istnieją jednak układy podobne pod względem funkcjonalnym, podobne pod względem prawideł przetwarzania informacji w nich zachodzących i ten właśnie rodzaj podobieństwa weźmiemy teraz pod uwagę.

2. Jeżeli dwa układy charakteryzują się podobieństwem prawideł przekazywania lub przetwarzania informacji to o jednym z nich możemy powiedzieć, że jest M O D E L E M cybernetycznym drugiego układu. Naturalnie w sensie tej definicji każdy układ jest sam własnym modelem.

Pojęcie modelu jest bardzo cenne i chętnie stosowane, ponieważ zdarza się, że trudno jest badać jeden z dwu podobnych układów i wówczas można posłużyć się badaniami modelu. Procedura taka jest uzasadniona, jeśli tylko w pełni zdajemy sobie sprawę, jak daleko sięga podobieństwo pomiędzy modelem a jego oryginałem.

Przykład 1

Podczas prac projektowych nad nowymi typami samolotów zachodzi zwykle konieczność określania ich właściwości areodynamicznych, czyli ich zachowania podczas szybkiego ruchu w powietrzu. Dokonuje się tego, ze względu na złożone obliczenia, drogą eksperymentu. Jednakże niecelowe byłoby budowanie prototypu samolotu w celu zbadania wspomnianych jego właściwości /badania takie byłyby nader kosztowne i niebezpieczne/, wykonuje się więc makietę, zachowującą jedynie podobny kształt samolotu i bada się ją w silnym strumieniu powietrza w tzw. tunelu aerodynamicznym. Na podstawie tych badań można wprawdzie wypowiedzieć się, jak będą opływały samolot strugi powietrza /wniosek o oryginale, wyciągnięty na podstawie modelu/, jednak absurdalne byłoby np. spekulacje dotyczące ciężaru samolotu, snute na podstawie zważenia jego np. gipsowej makiety użytej do prób.

3. Rozważając relację model - oryginał z punktu widzenia cybernetyki możemy obydwa te układy względnie odosobnione wyobrażać sobie w postaci czarnych skrzynek. Jeżeli wejścia i wyjścia modelu i oryginału są identyczne, oraz jeśli identyczne sygnały wejściowe wchodzące do modelu i oryginału powodują powstanie identycznych sygnałów wyjściowych w modelu i oryginale, to będziemy mówili, że model jest IZOMORFICZNY z oryginałem. Z naciskiem podkreślamy, że chodzi o izomorfizm funkcji, a nie struktury. Wiedząc, że jeden układ jest w stosunku do drugiego izomorficzny możemy, badając zachowanie się jednego układu, dokładnie określić zachowanie się drugiego, natomiast nie możemy wypowiadać się na podstawie budowy jednego układu o budowie drugiego. Jeśli więc po-

trafimy do danego UWO, będącego czarną skrzynką, dobrać układ o znanej strukturze i funkcji przejścia, który będzie jego izomorficznym modelem, to nie możemy uważać, że zidentyfikowaliśmy układ; zrobiliśmy o wiele mniej - zrozumieliśmy działanie układu. Dlatego naiwnym wydaje się mniemanie, że rozwój elektronicznych maszyn liczących i programów tzw. sztucznej inteligencji przybliża nas do poznania struktury naszego mózgu, możemy mieć co najwyżej nadzieję, że kiedyś zostanie zbudowany jego izomorficzny model.

4. Wymaganie pełnej odpowiedniości wejść i wyjść w oryginale i w modelu jest założeniem bardzo silnym. Niezmiernie rzadko uda się do danego układu dobrać izomorficzny model, bardziej będą nas wobec tego interesowały modele odwzorowujące jedynie niektóre, szczególnie dla nas istotne właściwości oryginału. Modele takie, różniąc się we wszystkich nieuwzględnianych szczegółach, w ramach swego podobieństwa mogą być wykorzystywane jak modele w pełni izomorficzne. Jest to właściwie jedynie realnie istniejąca klasa modeli, ponieważ pomiędzy każdymi dwoma nieidentycznymi układami zawsze znajdzie się jakąś drobną i nieistotną różnicę przeczącą definicji izomorfizmu. Takie "z grubsza podobne" modele nazwiemy modelami H O M E O M O R F I C Z N Y M I.
Homeomorfizm jest więc ograniczeniem izomorfizmu do cech wyłącznie istotnych. Nie ulega wątpliwości, że każdy model izomorficzny jest równocześnie /tym bardziej!/ modelem homeomorficznym.

TEST 1

P. Czy można podać izomorficzny model danego człowieka?

xxx

O. Nie, ponieważ wszyscy ludzie różnią się pewnymi szczegółami między sobą i nawet bliźnięta jednojajowe mają te różnice na tyle głębokie, że o izomorfiźmie nie może być mowy.

xxx

P. Czy zegar ścienny i zegarek na rękę są izomorficzne?

xxx

O. Mogą być, ponieważ ich wyjścia i wejścia informacyjne są identyczne. Warunkowy tryb wypowiedzi wynika z faktu, że zegar ścienny może posiadać wyjścia nie mające odpowiednika w zegarku ręcznym /np. wydzwanianie godzin/, lub odwrotnie /np. sekundnik/.
Jeśli źle odpowiedziałeś, przeczytaj jeszcze raz punkt 3.

P. Czy mając dwa układy podobne można określić, który jest oryginałem, a który modelem?

xxx

O. Nie, ponieważ podział na oryginał i jego model ma charakter umowny. Podział ten wynika z faktu, że z reguły jeden z dwu podobnych układów jest łatwiejszy do zbadania i dlatego uważamy go za model układu mniej dostępnego.
Jeśli odpowiedziałeś źle, przeczytaj punkty 1 i 2.

P. Czy jeden oryginał może mieć wiele modeli?

xxx

O. Naturalnie, że tak, ilość układów wykazujących cechę podobieństwa jest nieograniczona, a każdy z nich może być modelem wszystkich pozostałych.

P. Czy dwa homeomorficzne modele danego układu są względem siebie izomorficzne?

xxx

O. Ależ nie, każdy z nich ma istotne cechy takie jak oryginał, natomiast cechy nieistotne w obu modelach mogą być dowolne i zazwyczaj będą w jednym modelu inne niż w drugim.

Jeżeli odpowiedziałeś bez wahania i poprawnie, możesz pominąć przykład 2.

Przykład 2

Jeżeli zamierzamy wykonać jakiś przedmiot metalowy, to z reguły wykonujemy uprzednio jego rysunki. Rysunki te można traktować jak homeomorficzny model oryginału /przedmiotu metalowego/, jeśli za jedynie interesujące cechy uznać kształty przedmiotu, a nie na przykład jego ciężar czy połysk. Jeśli przedmiot jest wystarczająco prosty, możemy rysunek zastąpić opisem słownym, który również stanowić będzie homeomorficzny model przedmiotu, jeśli jednak jest bardzo złożony, wykonujemy najpierw makietę z materiału zastępczego /np. drewna/, która również jest homeomorficznym modelem. Łatwo zauważyć, że poszczególne modele homeomorficzne /rysunek, opis makieta/ na tyle różniły się między sobą, że mówienie o ich izomorfiźmie jest bezcelowe.

P. Czy dany układ może równocześnie być i nie być modelem homeomorficznym innego układu?

xxx

O. Może, ponieważ fakt bycia modelem uzależniony jest od ilości i charakteru wielkości wejściowych i wyjściowych uznanych za "istotne". Generalnie można stwierdzić, że im więcej wielkości uznać za nieistotne, to tym obszerniejsza staje się klasa modeli /układów podobnych/, w szczególności jeśli za jedyną istotną cechę uznać fakt materialnego bytowania, to wszystkie znane nam przedmioty są nawzajem swoimi modelami, i na odwrót - jeśli postawić wymaganie pełnej identyczności /wszystkie cechy istotne - pełny izomorfizm/, to wówczas jedynym modelem danego układu byłby tylko ten sam układ.

P. Jak można by matematycznie zapisać fakt, że układ A o n wejściach i m wyjściach jest izomorficzny z układem B? /Naturalnie układ B posiada taką samą ilość wejść i wyjść, jak układ A, gdyż inaczej mowy być nie może o izomorfiźmie/.

xxx

O. Oznaczamy przez $X_{1A}/t/$ czasowy przebieg zmienności i-tej wielkości wejściowej układu A, odpowiedni przebieg dla układu B oznaczamy $X_{1B}/t/$. Podobne oznaczenia z wykorzystaniem symbolu Y reprezentują przebiegi czasowe wielkości wyjściowych w obu układach. Izomorfizm układów zapiszemy wówczas implikacją:

$$\{X_{1A}/t/ = X_{1B}/t/,\ X_{2A}/t/ = X_{2B}/t/,\ \ldots,\ X_{nA}/t/ = X_{nB}/t/\} \text{ implikuje}$$

$$\{Y_{1A}/t/ = Y_{1B}/t/,\ Y_{2A}/t/ = Y_{2B}/t/,\ \ldots,\ Y_{mA}/t/ = Y_{mB}/t/\}$$

Implikacja oznacza zdanie warunkowe: $[\{A\}$ implikuje $\{B\}]$ jest równoważne zdaniu: ["Jeżeli $\{A\}$ to $\{B\}$ "]. Praktyczny przykład:

j e ż e l i nie wykonałeś powyższego zadania i w dodatku niezupełnie rozumiesz wywód podany w rozwiązaniu, t o zapoznaj się z książką K.Szczerbińska "Elementy logiki matematycznej i algebry wyższej" [8/].

P. Korzystając z rozwiązania poprzedniego zadania spróbuj zapisać matematycznie fakt homeomorfizmu dwu układów, z których pierwszy ma n wielkości wejściowych /tylko r pierwszych istotne/, i m wielkości wyjściowych /tylko k pierwszych istotne/, a drugi ma s wielkości wejściowych i v wielkości wyjściowych /ilość oraz numery wejść i wyjść istotnych - jak w pierwszym układzie/.
Oczywiście $n > r$ i $s > r$ oraz $m > k$ i $v > k$.

xxx

$$\{X_{1A}/t/ = X_{1B}/t/,\ X_{2A}/t/ = X_{2B}/t/,\ \ldots,\ X_{rA}/t/ = X_{rB}/t/,$$

[8/] K. Szczerbińska, O. Berejniewicz-Rajca, J.Walichniewicz: Elementy logiki matematycznej i algebry wyższej, Gliwice 1970.

$X_{r+1A}, \ldots, X_{nA}, X_{r+1B}, \ldots X_{sB}$ - dowolne} implikuje

$\{Y_{1A}/t/ = Y_{1B}/t/, Y_{2A}/t/ = Y_{2B}/t/, \ldots, Y_{kA}/t/ = Y_{kB}/t/,$

$Y_{k+1A}, \ldots, Y_{mA}, Y_{k+1B}, \ldots, Y_{vB}$ - dowolne}

5. Szczególnie intereujące są w cybernetyce modele matematyczne. Modelem matematycznym danego układu względnie odosobnionego nazywać będziemy formułę matematyczną, lub zbiór formuł, wiążącą wartości wielkości wejściowych i wielkości wyjściowych. Model matematyczny jest oczywiście zawsze jedynie modelem homeomorficznym, a więc w określonym stopniu niedokładnym, jednakże cechuje się on wyjątkową operatywnością i łatwością badania. Eksperymenty dokonywane na modelu matematycznym mogą być niezwykle dokładnie kontrolowane i są zupełnie niekosztowne/ o ile model nie jest zbyt skomplikowany/.

6. Modele matematyczne poznaliśmy już wcześniej, zanim pojawił się ten termin, mianowicie funkcje przejścia i transmitancja są właśnie matematycznymi odpowiednikami danego układu, czyli jego modelami. Powtarzanie tego faktu raz w terminologii funkcji przejścia, a raz w terminologii modeli ma na celu podkreślenie jednorodności teorii cybernetycznej, dla której modelem z równym powodzeniem może być zwierzę, maszyna, układ równań czy społeczeństwo. Mając do dyspozycji model matematyczny danego układu możemy bez trudu określić, jak zachowa się on po zaistnieniu określonych czasowych zmian jego wielkości wejściowych, aczkolwiek niekiedy droga prowadząca do odpowiedzi może być trudna matematycznie. W wielu przypadkach model matematyczny /którego złożoność musi być odpowiednia do złożoności układu będącego oryginałem/, może być tak skomplikowany, że jedynie szybkie komputery są w stanie dostarczyć potrzebnych rozwiązań w rozsądnie krótkim czasie. Mówi się wówczas często o modelowaniu procesu /układu/ na maszynie matematycznej. Nie wydaje się to zbyt szczęśliwym,gdyż istotne jest nie zastosowanie maszyny do uzyskania potrzeb-

nych odpowiedzi, a przejście do abstrakcyjnego modelu, wyrażającego się w odpowiednich formułach matematycznych. Będziemy zatem i w tych przypadkach poprzestawali na terminie - model matematyczny.

TEST 2

P. Co jest modelem matematycznym wagi /patrz test 3, s. 25/

xxx

O. Równanie $L = k \cdot Q$

P. Zapisz model matematyczny poprawy plonów w wyniku stosowania nawozów sztucznych. /test 6, s. 39/

xxx

O. Układ równań:

$$Y_1 = aX_1 + bX_2 + cX_3,$$

$$Y_2 = dX_1 + eX_2 + fX_3.$$

Jeżeli miałeś wątpliwości przy udzielaniu odpowiedzi przeczytaj punkty 5 i 6 raz jeszcze.

7. Nasze dotychczasowe rozważania na temat modeli matematycznych układów względnie odosobnionych ograniczone były w znacznym stopniu przez nieuwzględnianie zależności czasowych. Sygnały na wejściu i na wyjściu traktowaliśmy jak liczby, podczas gdy są one najczęściej f u n k c j a m i c z a s u. Naturalnie może się zdarzyć, że sygnały wejściowe i wyjściowe są stałe w czasie lub zmieniają się w sposób pozwalający na traktowanie tych zmian jako kolejnego podawania stałych wartości. Może się również zdarzyć, że układ nie zmienia dynamiki przebiegów, czyli mówiąc prościej, że czasowe zmiany wielkości wyjściowych równe są czasowym zmianom wielkości wejściowych, co najwyżej pomnożonym przez określoną stałą lub na przykład zesumowanym.

Przykład 3

Niech naszym UWO będzie tucznik, którego waga zmienia się w czasie /rośnie i ewentualnie okresami maleje/, zgodnie z pewną funkcją czasu X/t/. Przebieg czasowy X/t/ będzie wejściem naszego układu. Wyjściem będzie przebieg czasowy wartości rynkowej tegoż tucznika Y/t/. Pomiędzy wielkościami tymi istnieje oczywisty i jednoznaczny związek:

$$Y/t/ = k \cdot X/t/.$$

Jeśli cena za kilogram żywca pozostaje w omawianym okresie czasu niezmieniona /k = const./, to transmitancja takiego układu jest po prostu stałym mnożnikiem.

Przykład 4

Waga jest układem, w którym przebieg czasowy wielkości wejściowej /ciężaru ważonych przedmiotów/ może być uważany za kolejne podawanie stałych wartości /na ogół po położeniu na szalkę, ważony przedmiot już wagi nie zmienia/. Wielkości wejściowe ulegają wprawdzie pewnym zafalowaniom /po położeniu przedmiotu waga się "kiwa" i wskazanie jej wskazówki jest zmienne w czasie/, ale proces ten trwa krótko i można go nie brać pod uwagę, tak że również wielkości wyjściowe są sekwencją stałych wartości. Dla każdej wartości X_i odpowiadająca jej wartość Y_i spełnia równanie:

$$Y_i = k X_i$$

gdzie i jest numerem okresu czasu, czyli kolejnym numerem ważonego przedmiotu.

W obydwu powyższych przykładach udało się ominąć konieczność uwzględniania zależności czasowych przy określaniu funkcji przejścia. Gdybyśmy jednak wzięli pod rozwagę przykład wahadła jako układu o wejściu w postaci siły wytrąoającej wahadło z położenia równowagi i wyjściu w postaci kąta wychylenia, to pominięcie zależności czasowych byłoby zbyt grubym przybliżeniem. Wahadło wytrącone z równowagi, nawet pojedynczym impulsem siły, drga przez długi czas, a jego stan ustalony jest zupełnie niereprezentatywny, gdyż niezależnie od działających sił po ich zniknięciu ustala się ostatecznie w swym najniższym położeniu, co nie niesie żadnej informacji o zaistniałych wielkościach wejściowych.

U w z g l ę d n i a n i e   z a l e ż n o ś c i   c z a s o w y c h   j e s t   w i ę c   n i e k i e d y   k o n i e c z n e.

8. Najprostszy sposób wprowadzenia dynamiki rozpatrywanego układu polega na przyjęciu założenia, że szybkość zmian wielkości wyjściowej jest pewną funkcją wielkości wejściowej w danej chwili, wielkości wyjściowej w danej chwili oraz w ogólnym przypadku danej chwili:

$$\frac{dY/t/}{dt} = \dot{F}\left(Y/t,\ X/t/,\ t\right) \qquad /5\ /$$

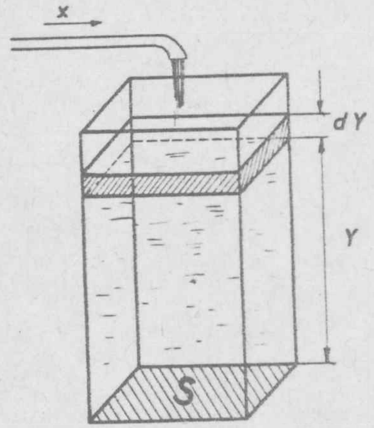
Równanie różniczkowe /5/ można obecnie uważać za nową funkcję przejścia układu, jednakże wyrażoną w postaci niejawnej. Jak wiadomo /a jeśli nie wiadomo, to wypadałoby przypomnieć sobie np. z książki F.Leja "Rachunek różniczkowy i całkowy"/, aby równanie /5/ było rozwiązalne, czyli aby stanowiło tzw. problem Cauchy´ego, potrzebne jest zadanie warunków początkowych:

$$Y/0/ = Y_0. \qquad /6\ /$$

Warunki początkowe określają wartość wielkości wyjściowej, w chwili kiedy zaczęliśmy obserwować układ i stanowią skondensowaną formę informacji o przeszłym zachowaniu układu. O ile niezbyt pewnie czujesz się na gruncie równań różniczkowych, przeczytaj przykłady 5, 6 i 7, natomiast nie musisz czytać punktu 9, który jest przeznaczony dla zaawansowanych i nie jest konieczny dla rozumienia całości książki.

<u>Przykład 5</u>

Niech naszym UWO będzie zbiornik o stałym polu powierzchni przekroju poziomego S. /rys. 16/. Wielkością wejściową tego układu X jest natężenie przepływu wody w rurociągu zasilają-



Rys.16 Napełnianie zbiornika

cym zbiornik /wyrażone w m^3 na sekundę/. Wielkością wyjściową jest aktualny poziom wody w zbiorniku Y. Ponieważ ilość wody, jaka została pomieszczona w warstewce wody o grubości dY /patrz rys. 16/, dlatego spełnione jest równanie:

$$X/t/\ dt = S\ /dY/t/$$

czyli po przekształceniu

$$\frac{dY/t/}{dt} = \frac{1}{S} X/t/. \qquad /7/$$

Jak widać ten prosty układ poddaje się opisowi w postaci uproszczonej wersji równania /5/. Dla ilustracji zastanówmy się, jak będzie się ten układ zachowywał przy różnych przebiegach wielkości wejściowej X/t/. Dla uproszczenia przyjmiemy $S = 1m^2$

a/ Niech X/t/ będzie stałe w czasie:

$$X/t/ = k = \text{const.}$$

Wówczas rozwiązanie /przez proste całkowanie/ równania /7/ prowadzi do przebiegu Y/t/ narastającego w czasie ze stałą szybkością:

$$Y/t/ = k\ t + Y_o$$

Y_o oznacza początkowy poziom wody w zbiorniku w chwili, w której rozpoczęliśmy obserwację. Wynik ten jest logiczny, gdyż z doświadczenia wiemy, że poziom wody w zbiorniku napełnianym z kranu o stałej wydajności równomiernie narasta.

b/ Niech X/t/ narasta w czasie poczynając od pewnej wartości X_o

$$X/t/ = X_o + r\ t.$$

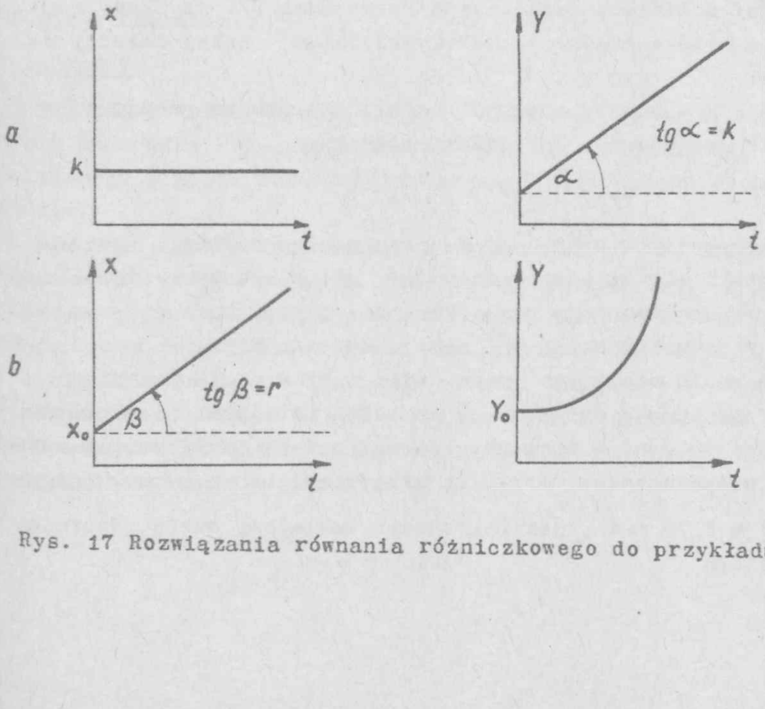
Dokonamy całkowania równania /7/:

$$\int dY = \int /X_o + r\ t/\ dt,$$

$$Y/t/ = \frac{1}{2} r\ t^2 + X_o\ t + Y_o.$$

Poziom wody narasta teraz szybciej niż proporcjonalnie do czasu, co jest również uzasadnione, ponieważ dopływ wody jest coraz intensywniejszy.

Omówione przebiegi narysowane są na rysunku 17. Wprowadzimy teraz po prawej stronie równania zależność od Y/t/.



Rys. 17 Rozwiązania równania różniczkowego do przykładu 5

Przykład 6

Niech w dnie omawianego poprzednio zbiornika będzie wywiercony otwór, przez który woda wycieka pod działaniem własnego ciężaru, a więc natężenie wypływu wody będzie proporcjonalne /jest to założenie!/ do wysokości słupa wody nad dnem zbiornika, ze współczynnikiem proporcjonalności a. Łatwo zauważyć, że wówczas sumaryczne natężenie dopływu wody do zbiornika będzie różnicą X/t/ - a Y/t/, a równanie dynamiki układu da się zapisać w postaci:

$$\frac{dY/t/}{dt} = \frac{1}{S}\left[X/t/ - a\,Y/t/\right] \qquad /8/$$

Równanie /8/ jest już trudniejsze do rozwiązania, gdyż rozwiązanie go wymaga rozdzielania zmiennych, jednakże jest to typowe równanie różniczkowe liniowe pierwszego rzędu, którego rozwiązanie szczegółowo omówione jest w każdym podręczniku matematyki.

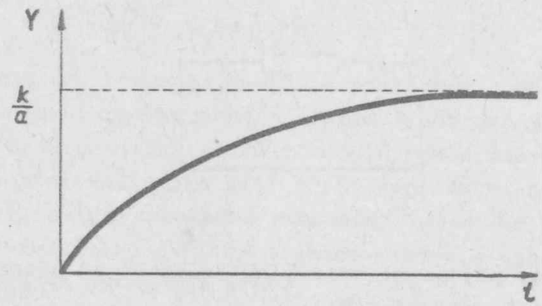
Zakładając znowu $S = 1\ m^2$ i przyjmując X/t/ jako stałe /przypadek /a/ i liniowo narastające /przypadek b/, uzyskamy rozwiązania, które okażą się nader przydatne w całej dalszej analizie.

a/ X/t/ = k = const, a ponadto zakładamy, że na początku zbiornik jest pusty. Uzyskujemy wówczas

$$Y/t/ = \frac{k}{a}\ /1 - e^{-at}/.$$

Rozwiązanie to wskazuje, że przy stałym dopływie poziomu wody ustali się na pewnej wartości, która będzie wynikała z równowagi dopływu wody przez rurociąg i odpływu wody przez otwór. Ponieważ funkcja e^{-t} maleje od zera dopiero po upływie nieskończenie wielkiego czasu, więc teoretycznie przebieg zmian Y/t/ będzie narastał do wartości ustalonej nieskończenie długo, niemniej w rzeczywistości zmiany bardzo szybko staną się niezauważalne. Przebieg przedstawiono na rysunku 18.

b/ $X/t/ = X_0 + r\,t$, jak poprzednio zakładamy pusty zbiornik na początku .



Rys. 18 Rozwiązanie równania różniczkowego do przykładu 6

Wówczas

$$Y/t/ = \frac{r}{a}t + \left(\frac{r}{a^2} - \frac{X_0}{a}\right)e^{-at} + \left(\frac{X_0}{a} - \frac{r}{a^2}\right)$$

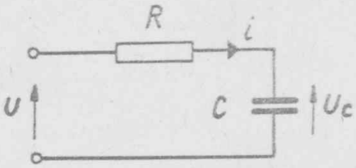
Przykład poprzedni był o tyle interesujący, że równania różniczkowe postaci /8/ opisują bardzo wiele zjawisk w przyrodzie i technice.

Przykład 7

Ten sam model matematyczny może służyć do opisu np. zjawiska ładowania kondensatora elektrycznego, nagrzewania bloku metalowego w piecu lub analizy uproszczonego układu ekonomicznego.

Wyobraźmy sobie kondensator o pojemności C /w Faradach/ ładowany ze źródła napięcia /np. baterii/ o napięciu U /woltów/ przez opornik o oporności R /omów/. Niech wielkością wyjściową z tego układu będzie napięcie na kondensatorze U_c, a wielkością wejściową napięcie baterii ładującej U /rys.19/. Wówczas jak wiadomo z fizyki, szybkość wzrostu napięcia na kondensatorze będzie proporcjonalna do wartości prądu i /w amperach/ ładującego kondensator:

$$\frac{dU_c}{dt} = \frac{1}{C} \; i/t/$$



Rys. 19 Ładowanie kondensatora

Ale wartość płynącego prądu jest zależna od różnicy napięć i wartości oporności R:

$$i/t/ = \frac{1}{R}\left[U/t/ - U_c/t/\right]$$

Podstawiając otrzymamy równanie w znacznym stopniu podobne do równania /8/, a mianowicie

$$\frac{dU_c}{dt} = \frac{1}{RC}\left[U/t/ - U_c/t/\right] \qquad /9/$$

Przez odpowiednie rozbudowanie układu moglibyśmy, wprowadzając wielkość modelującą a, bez trudu upodobnić jeszcze bardziej wzór /9/ do wzoru /8/.

Rozważmy teraz proces nagrzewania. Szybkość wzrostu temperatury grzanego bloku metalu jest proporcjonalna do ilości ciepła dostarczanego do bloku w jednostce czasu /w kaloriach/:

$$\frac{dT/t/}{dt} = b\,Q/t/$$

gdzie T/t/ jest temperaturą grzanego bloku, Q/t/ tzw. strumieniem ciepła /ilością ciepła w jednostce czasu dostarczaną do bloku/, natomiast b współczynnikiem proporcjonalności. Ale intensywność wymiany cieplnej jest tym większa, im większa jest różnica temperatur pomiędzy bryłą a temperaturą pieca $T_p/t/$ /temperatura pieca może być zmienna w czasie!/

Zatem można napisać:

$$Q/t/ = c\left[T_p/t/ - T/t/\right],$$

gdzie c jest współczynnikiem proporcjonalności. Podstawiając mamy:

$$\frac{dT/t/}{dt} = bc\left[T_p/t/ - T/t/\right] \qquad /10/$$

Otrzymaliśmy równanie zupełnie analogiczne do równań /9/ i /8/. Podamy jeszcze jeden przykład z dziedziny ekonomii.

Niech Z/t/ oznacza ilość zatrudnionych w danym przedsiębiorstwie pracowników, a W/t/ ilość wyprodukowanych wyrobów. Wówczas w idealnym przypadku szybkość przyrostu produktu byłaby proporcjonalna do ilości pracowników ze współczynnikiem k wyrażającym wydajność pracy:

$$\frac{dW/t/}{dt} = k\,Z/t/.$$

Jednakże w miarę wzrostu ilości wyprodukowanych wyrobów coraz więcej pracowników musi przechodzić ze stanowisk produkcyjnych do pracy przy transporcie wyrobów, gospodarce magazynowej, dystrybucji i administracji. Jest to oczywiście naiwny model, niezbyt dokładnie oddający rzeczywiste zależności, ma on jednak jedynie stanowić ilustrację, zgódźmy się więc na jego naiwność.

Zatem liczba ludzi pracujących rzeczywiście produkcyjnie wynosi:

$$Z/t/ = Z/t/ - l\,W/t/$$

gdzie l jest współczynnikiem. Wobec tego

$$\frac{dW/t/}{dt} = k\left[Z/t/ - l\,W/t/\right] \qquad /11/$$

I znowu równanie jest analogiczne jak poprzednie.

Możemy więc powiedzieć, że zjawiska napełniania zbiornika, ładowania kondensatora, nagrzewania bloku i produkcji przemysłowej są h o m e o m o r f i c z n e z d o k ł a d n o ś c i ą d o p r z y j ę t y c h z a ł o ż e ń. A zatem każde z nich może być przyjęte do badań jako model pozostałych. Naturalnie najwartościowszym modelem ze względu na łatwość badań jest model kondensatora i dlatego chcąc się czegoś dowiedzieć na temat zachowania bloku stali w piecu o zmiennej temperaturze mierzymy napięcie na kondensatorze poddanym ładowaniu ze źródła o zmiennym napięciu. Jest to daleko wygodniejsza droga niż rozwiązywanie nawet tak prostych równań, jak: /8/,/9/,/10/ czy /11/. Metoda ta jest jeszcze bardziej owocna i daje bez porównania większe korzyści przy równaniach bardziej złożonych, i dlatego buduje się nawet specjalne maszyny matematyczne, złożone poza opornikami i kondensatorami z całego szeregu innych elementów pomocniczych, które m o d e l u j ą rozmaite procesy na podobnej zasadzie do podanej tutaj, a że działają poprzez analogię z rzeczywistym procesem, nazywa się je m a s z y n a m i a n a l o g o w y m i. Są one wielokrotnie szybsze w pracy niż najszybsze komputery cyfrowe.

9. Opis dynamiki procesów zachodzących w układach względnie odosobnionych przy pomocy równań /5/ z warunkiem /6/ jest naiwnie prosty w zestawieniu z rzeczywistą złożonością tej dynamiki. Dlatego trzeba przynajmniej wspomnieć, jakimi metodami usiłuje cybernetyka sprostać rosnącej złożoności opisywanych układów.

Po pierwsze zazwyczaj zachodzi potrzeba angażowania do opisu równań różniczkowych wyższych rzędów:

$$\frac{d^nY/t/}{dt^n} = F\left(\frac{d^{n-1}Y/t/}{dt^{n-1}}, \frac{d^{n-2}Y/t/}{dt^{n-2}}, \ldots, \frac{dY/t}{dt}, Y/t/, X/t/, t\right) \qquad /12/$$

z zespołem warunków początkowych:

$$\begin{aligned} Y/0/ &= Y_0 \\ \frac{dY/0/}{dt} &= Y_1 \\ \frac{d^2Y/0/}{dt^2} &= Y_2 \qquad 12.a \\ &\ldots\ldots\ldots\ldots \\ \frac{d^{n-1}Y/0/}{dt^{n-1}} &= Y_{n-1} \end{aligned}$$

Takie postawienie zagadnienia pozwala już wprawdzie odwzorować dowolnie złożoną dynamikę, ale w dalszym ciągu /poprzez zespół warunków początkowych/ uzależnia całe rozwiązanie od chwili początkowej. Inaczej można to ograniczenie interpretować jako założenie, że stan układu w pełni zdeterminowany przez stan w chwili obecnej i przyszłe wielkości wejściowe. W wielu przypadkach to wystarcza, jednakże w skrajnie złożonych układach, takich jak biologiczne czy ekonomiczne, w układach posiadających pamięć, założenie to jest nie do przyjęcia.

Musimy więc wprowadzić uzależnienie przyszłych stanów układu od jego historii, czego dokonujemy wprowadzając tzw. równania różniczkowo-różnicowe, albo równania z odchylonym argumentem:

$$\frac{dY/t/}{dt} = F\left(Y/t - ł/, \quad X/t - ł/, t\right) \qquad /13/$$

gdzie ł jest czasem opóźnienia, czyli czasem, w którym przeszłe oddziaływania mogą uwidaczniać się w zachowaniu układu. Dla rozwiązania takiego równania nie wystarczają warunki początkowe, potrzebny jest pełny przebieg funkcji Y/t/ w czasie poprzedzającym chwilę początkową aż do ł. Jeżeli weźmiemy pod uwagę fakt, że we wszystkich omawianych przypadkach rozpatrywaliśmy jedynie układy o jednym wejściu i jednym wyjściu, a w rzeczywistości występują zwykłe układy wielowymiarowe, to zrozumiemy, że większość naprawdę wartościowych osiągnięć cybernetyki może być przekazana jedynie językiem zaawansowanej matematyki. Będziemy jednak starali się unikać skomplikowanych zapisów matematycznych, na czym zyska popularność książki, niestety, ze stratą dla ścisłości opisu.

Rozdział II

SPRZĘŻENIE ZWROTNE I JEGO ROLA W UKŁADACH EKONOMICZNYCH

1. Sprzężenie układów. Pojęcie systemu. Sprzężenie szeregowe i równoległe

1. Zajmiemy się obecnie łączeniem układów względnie odosobnionych pomiędzy sobą przy pomocy S P R Z Ę Ż E Ń. Sprzężeniem nazywać będziemy kanał przekazywania informacji, który zaczyna się na wyjściu układu i kończy na wejściu tego samego lub innego układu. Sprzężenia będziemy rysunkowo oznaczali jako kreski, zaczynające się na wyjściu a kończące na wejściu układu, zaopatrzone w strzałki wskazujące kierunek przepływu informacji.

2. Zbiór układów względnie odosobnionych powiązanych sprzężeniami nazywać będziemy S Y S T E M E M. /Patrz punkt 1 rozdziału I podr.2. Naturalnie, jeśli niektóre układy wchodzące w skład systemu będą miały wejścia nie połączone z innymi układami systemu, a przyjmujące z zewnątrz, to wejścia te nazwiemy wejściami systemu. Podobnie wyjścia.

Każdy system można rozpatrywać jako układ, ponieważ można nie interesować się jego wewnętrzną strukturą, a jedynie zwracać uwagę na wejścia i wyjścia. I na odwrót, każdy z rozpatrywanych poprzednio układów można nazwać systemem, ponieważ wniknąwszy w jego strukturę potrafilibyśmy wyróżnić w niej pewną ilość części, które same są układami, a po połączeniu dają system, który dotychczas nazywaliśmy układem.

Przykład 1

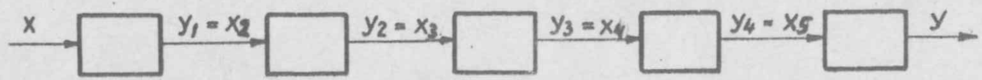
Samolot możemy rozpatrywać jako układ, którego wejściami są stery, a wyjściami położenie i szybkość w powietrzu. Jednakże ten układ można rozbić na szereg powiązanych układów, takich jak stery, silnik, układ wskaźników, itp. jest więc on systemem. /Podział ten można by kontynuować, wyróżniając z kolei części składowe silnika itd./. Równocześnie samolot, jako układ, jest częścią systemu samolot-pilot, który to system jako układ, jest elementem systemu dywizjon lotniczy itd.

3. Rozpatrzymy najprostszy sposób powiązania układów w system: S P R Z Ę Ż E N I E S Z E R E G O W E. Jeżeli rozważymy układy o jednym wejściu i jednym wyjściu oraz połączymy je sprzężeniami, tak aby wyjście pierwszego układu było sprzężone z wejściem drugiego, wyjście drugiego z wejściem trzeciego itd. /rys. 20/, to uzyskamy sprzężenie szeregowe. Wejściem do takiego systemu jest wejście pierwszego układu, wyjściem wyjście ostatniego układu. Systemy takie spotyka się nader często we wszystkich dziedzinach leżących w zakresie zainteresowań cybernetyki. W podanym przykładzie 1, i we wszystkich dalej w tym rozdziale podawanych przykładach, uwzględnione będą następujące dziedziny: ekonomia /E/, technika /T/, biologia /B/ i matematyka /M/, zatem Czytelnik może wybrać sobie przykład z dziedziny, którą zna najlepiej i nie interesować się, o ile nie chce, przykładami z innych dziedzin.

Przykład 1

E. Przedsiębiorstwo przemysłowe jest systemem składającym się z wielu układów, powiązanych najczęściej szeregowo, gdyż produkt wyprodukowany na jednym stanowisku pracy przechodzi następnie jako półfabrykat na inne stanowisko do dalszej obróbki. W systemie takim istnieją układy /wydziały fabryki, stanowiska pracy/, przyjmujące surowce z zewnątrz /wejścia systemu/ i układy wydające na zewnątrz gotowy produkt.

B. Często przytaczany przykład wzajemnego uzależniania biosfery jest przykładem szeregowego połączenia układów /poszczególnych gatunków roślin i zwierząt/ w system. Wejściem tego systemu jest energia promieniowania słonecznego zamieniania przez rośliny w energię chemiczną. Energia ta następnie staje się



Rys. 20 System szeregowy

wejściową dla układu zwierząt mięsożernych. O wyjściu tego systemu trudno mówić, chyba że umówimy się iż wyjściem tym jest ilość zwierząt będących pokarmem!

M. Dowód matematyczny jest systemem, w którym poszczególne jego układy /etapy dowodu/ są połączone szeregowo - wniosek poprzedni stanowi punkt wyjścia dla dalszych rozważań. Wejściem tego systemu jest ustalony zbiór aksjomatów, a wyjściem teza dowodzonego twierdzenia.

4. Rozważymy funkcje przejścia i transmitancję systemu szeregowego. Wielkość wyjściowa pierwszego układu Y_1 jest związana z wielkością wejściową systemu X funkcją przejścia pierwszego układu:

$$Y_1 = F_1/X/$$

Wielkość ta /jak zaznaczono na rysunku /20/ jest wielkością wejściową do następnego układu, więc

$$Y_2 = F_2/X_2/ = F_2\left(F_1/X_1/\right)$$

Kontynuując te rozważania dochodzimy do wniosku, że wielkość wyjściową systemu podanego na rysunku 20 można określić ze wzoru:

$$Y = F_4\left(F_3\left(F_2\left(F_1/X/\right)\right)\right) \qquad /14/$$

Wzór /14/ ma dość złożoną postać. Będzie ona prostsza, jeśli założymy, że układy wchodzące w skład systemu są liniowe, czyli, że można opisać je przy pomocy transmitancji. Mamy wówczas

$$\begin{aligned} Y_1 &= G_1\, X, \\ Y_2 &= G_2\, X_2 = G_1 G_2\, X, \qquad /15/ \\ Y &= G_1 G_2 G_3 G_4\, X. \end{aligned}$$

Jak widać, dla systemu złożonego z szeregowo połączonych układów można wprowadzić T R A N S M I T A N C J Ę Z A S T Ę P C Z Ą, będącą i l o c z y n e m transmitancji wszystkich układów składowych:

$$Y = G\,X \qquad G = G_1 G_2 G_3 G_4 \qquad /16/$$

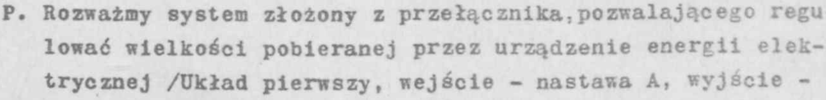
Naturalnie zasada ta stosuje się do wszystkich liniowych systemów szeregowych o dowolnej ilości układów składowych.

TEST 1

P. Rozważmy system złożony z przełącznika,pozwalającego regulować wielkości pobieranej przez urządzenie energii elektrycznej /Układ pierwszy, wejście - nastawa A, wyjście - pobór energii P, zależność proporcjonalna/, z licznika energii /Układ drugi, wejście - zużycie energii P, wyjście - wskazanie N/, kontrolera elektrowni /Układ trzeci, wejście - wskazanie licznika, N, wyjście - koszt rachunku wystawionego K/ i elektrowni /Układ czwarty, wejście - wpływ z rachunku K, wyjście - czysty zysk Z, będący stałym procentem K/. Narysować system, podać transmitancje poszczególnych układów i określić transmitancję zastępczą.

xxx

O. System jest narysowany na rysunku /21/. Poszczególne ukła-



Rys. 21 Pobór energii jako system szeregowy

dy opisują równania:

$$P = kA$$
$$N = l\,P$$
$$K = m\,N$$
$$Z = n\,K$$
$$Z = G\,A \qquad G = klmn$$

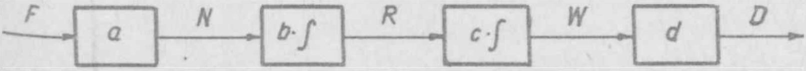
gdzie k, l, m, n, oznaczają odpowiednie współczynniki proporcjonalności.

Brak poprawnej odpowiedzi zmusza do ponownego zapoznania się z punktem 4.

P. Rozważymy system złożony z czterech układów, z których pierwszy jest szkołą zawodową, i wejściem jego jest ilość funduszy F przeznaczonych na szkolenie, a wyjściem-liczba zatrudnionych nauczycieli N /zakładamy, że są to wielkości pozostające w zależności prostej proporcjonalności/. Układ drugi ma na wejściu liczbę nauczycieli N, a na wyjściu liczbę wykwalifikowanych robotników R, opuszczających szkołę. Zależność ta jest tu podobna, jak w przykładzie 5 /z rozdziału I podr.4/. Układ trzeci uzależnia ilość wyrobów /wyjście/ W, od liczby robotników R. Układ ostatni określa zależność /proporcjonalną/ pomiędzy ilością wyrobów W, a dochodem D. Narysować system, określić transmitancję zastępczą i scharakteryzować związek pomiędzy F a D.

xxx

O. Schemat systemu przedstawiono na/rys.22/. Transmitancje



Rys. 22 Szkoła przyzakładowa jako system szeregowy

wynoszą:

$$N = a\,F$$

$$\frac{dR}{dt} = b\,N \qquad R = b\int N\,dt$$

$$\frac{dW}{dt} = c\,R \qquad W = c\int R\,dt$$

$$D = d\,W$$

$$D = d\;c\int b\left(\int a\,F\,dt\right)dt = abcd\iint F\,/t/\,dt\;dt$$

Zależność jest tego rodzaju, że stopa wzrostu dochodu jest proporcjonalna do stopy wzrostu nakładów na oświatę.

Jeśli i tym razem miałeś trudności z rozwiązaniem, radzimy zapoznać się z książką O.Kyn i P.Pelikan "Cybernetyka a ekonomia".

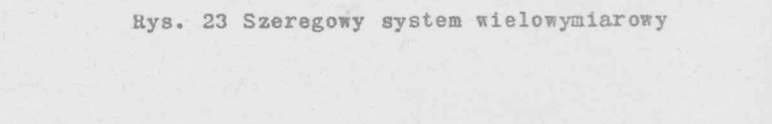
5. W przypadku układów wielowymiarowych możliwe jest również łączenie ich w systemy, w szczególności możliwe jest również łączenie ich w systemy, w szczególności możliwe jest szeregowe połączenie pewnej ilości układów wielowymiarowych, polegające na połączeniu wszystkich wyjść układu poprzedzającego z wszystkimi wejściami układu następnego. Jeśli to zachodzi, to transmitancja macierzowa utworzonego w ten sposób systemu jest iloczynem transmitancji macierzowych układów wchodzących w jego skład, wzór /16/ pozostaje w mocy, a jedynie należy występujące w nim wielkości rozumieć jako odpowiednie macierze i wektory.

Przypominamy /patrz punkt 13 rozdziału I podr.2/, że macierz G posiada tyle wierszy, ile dany układ posiada wyjść, i tyle kolumn, ile ma wejść. Wynika z tego, że postulowane przez wzór /16/ mnożenie transmitancji macierzowych będzie wykonalne, ponieważ odpowiednie ilości kolumn i wierszy /ilości wejść i odpowiednich wyjść/ będą jednakowe. Trzeba tylko zmienić kolejność mnożenia, gdyż mnożenie macierzy, w przeciwieństwie do mnożenia liczb jest nieprzemienne:

$$Y = [G_4]\,[G_3]\,[G_2]\,[G_1]\,X \qquad /16a/$$

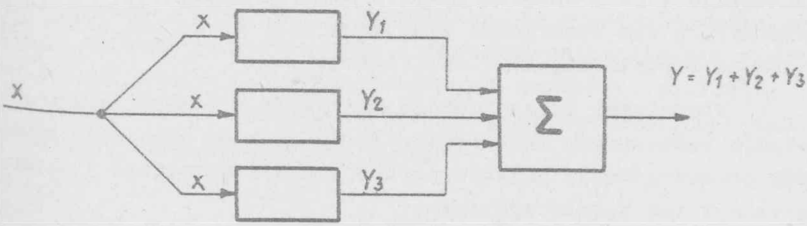
Wzór /16a/ można uważać za modyfikację dla potrzeb rachunku macierzowego /wzoru 16/.

Szeregowe połączenie układów wielowymiarowych ilustruje rysunek /23/.



Rys. 23 Szeregowy system wielowymiarowy

6. Poza połączeniem szeregowym równie popularnym jest SPRZĘŻENIE R Ó W N O L E G Ł E układów, charakteryzujące się tym, że pewna ilość układów posiada to samo wejście, a ich wyjścia są sumowane /rys. 24/.



Rys. 24 System równoległy

Podamy jak poprzednio kilka przykładów z różnych dziedzin, ilustrujących powszechność systemów z równoległymi sprzężeniami.

Przykład 2

E. Grupa klientów sklepu spożywczego stanowi przykład zbioru układów, których wejścia pobudzane są przez te same wielkości wejściowe /widok towarów będących na zbyciu/, a wyjścia, z reguły różne będące ilością pieniędzy pozostawionych w wyniku dokonania zakupów, są sumowane i określają wyjście systemu, będące sumarycznie utargiem sklepu.

T. Zespół pomp odwadniających, sterowanych jednym wyłącznikiem, jest przykładem systemu równoległego. Wejściem /wspólnym dla wszystkich pomp/ jest położenie wyłącznika /włączone - wyłączone/, wyjściami poszczególnych układów są wydajności minutowe poszczególnych pomp /mogą być one np. zmienne w czasie/, a wyjściem systemu-sumaryczna wydajność pomp, mierzona ilością wody usuwanej w ciągu minuty.

B. Mięśnie kręgowców stanowią przykład systemu równoległego, którego składowymi układami są włókna mięśniowe /miofibryle/, posiadające zdolność do kurczenia się pod działaniem impulsów nerwowych docierających do tzw. płytek ruchowych. Impulsy te są zwykle generowane przez pojedynczy neuron i obsługują dużą ilość włókien /wejście/. W wyniku pobudzenia następuje skurcz włókien /zwykle niejednoczesny i nierównomierny/, a ich sumaryczna siła jest siłą wywieraną przez mięsień na odpowiednią kość szkieletową.

7. Rozpatrzmy teraz funkcję przejścia i transmitancję układów połączonych równolegle. Ze względu na sumowanie wyników poszczególnych układów postać funkcji przejścia i transmitancji też będzie addytywna:

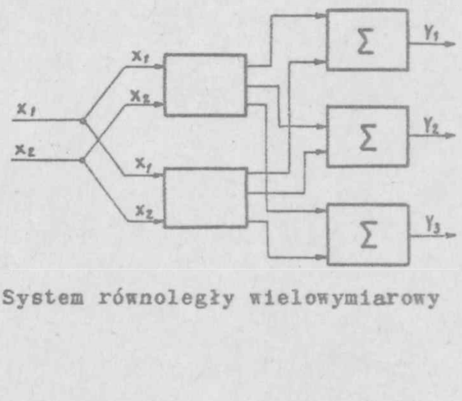
$$Y = Y_1 + Y_2 + Y_3 = F_1/X/ + F_2/X/ + F_3/X/ \qquad /17/$$

W przypadku układu liniowego zesumują się transmitancje:

$$Y = G_1 X + G_2 X + G_3 X = /G_1+G_2+G_3/ X \qquad /18/$$

Dla dowolnej, większej lub mniejszej ilości układów połączonych równolegle postać wzorów /17/ i /18/ będzie podobna, zmienią się jedynie ilości składników sumy.

8. Uogólnienie na przypadek układów wielowymiarowych jest proste, ale wymaga przyjęcia dość mocnych założeń: wszystkie układy wchodzące w skład systemu muszą mieć taką samą ilość wejść /bo wejścia te są wspólne/, oraz taką samą ilość wyjść, przy czym sumowane są odpowiednio pierwsze wyjścia wszystkich układów, dając pierwsze wyjście systemu, analogicznie drugie itd. /rys. 25/.



Rys. 25 System równoległy wielowymiarowy

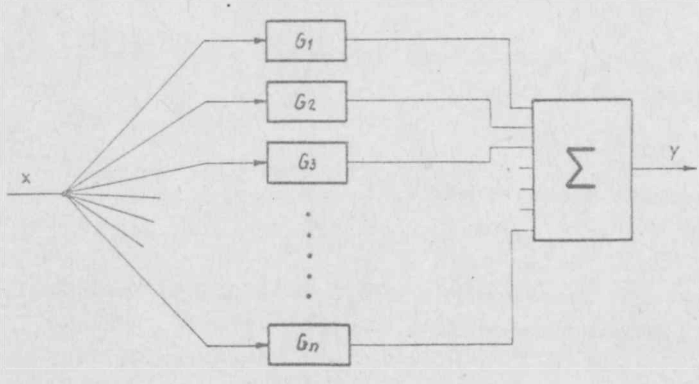
Założenia te są dość sztuczne, a systemy równoległe wielowymiarowe stanowią rzadkość. Niemniej należy wspomnieć, że transmitancja macierzowa takiego systemu jest sumą transmitancji macierzowych poszczególnych układów wchodzących w skład systemu /wzór 18/. Przypominamy, że sumowanie macierzy polega na sumowaniu elementów i jest wykonalne jedynie wtedy, gdy liczby kolumn i wierszy sumowanych macierzy, są identyczne. Warunek ten jest tu spełniony, ponieważ wszystkie układy mają tę samą ilość wejść i wyjść.

TEST 2

P. Narysować strukturę i określić transmitancję systemu opisanego w przykładzie 2 E, jeżeli jako zmienną X przyjmiemy atrakcyjność towaru /w skali od 1 do 100/, przez G_1 - podatność pierwszego klienta na atrakcyjność towaru /kwota zakupów w zł na jednostkę atrakcyjności towaru/, przez G_2, G_3, ..., G_n analogiczne wielkości dla drugiego, trzeciego itd. klienta, przez Y_1, ... Y_2 kwotę zakupów poszczególnych klientów, a przez Y utarg sklepu.

xxx

O. System przedstawiony jest na rysunku 26. Jego transmitan-



Rys. 26 Sklep jako system równoległy

cję obliczymy ze wzoru /18/ :

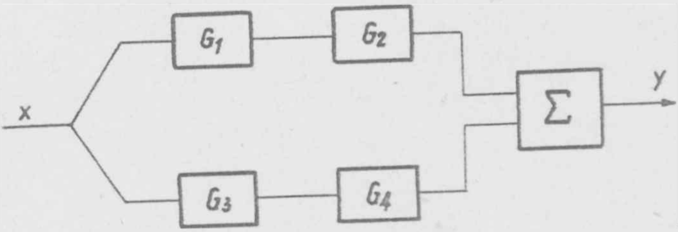
$$G = \sum_{i=1}^{n} G_i ,$$

a sumaryczne wyjście ze wzoru

$$Y = G\ X.$$

Jeśli nie odpowiedziałeś, koniecznie przeczytaj jeszcze raz punkt 7.

P. Znaleźć transmitancje zastępczą dla systemu przedstawionego na rysunku 27.



Rys. 27 System mieszany /do testu 2 /

xxx

P. Układy G_1 i G_2 są połączone szeregowo, ich transmitancja zastępcza wynosi więc

$$G_{12} = G_1\ G_2$$

Analogicznie dla systemów G_3 i G_4:

$$G_{34} = G_3\ G_4$$

"Podsystemy" szeregowe G_{12} i G_{34} są połączone równolegle, ich transmitancja wypadkowa wynosi

$$G = G_{12} + G_{34} = G_1\ G_2 + G_3\ G_4$$

9. Ostatnie zadanie testu 2 wprowadziło nas w sferę S Y S T E M Ó W M I E S Z A N Y C H. Są to systemy złożone z pewnej ilości układów, z których jedne są połączone szeregowo, inne równolegle, a całe ich grupy mogą być też łączone szeregowo bądź równolegle itd. Zjandowanie transmitancji takich systemów polega na znajdowaniu transmitancji wszystkich układów składowych /identyfikacja składników systemu/, rozwiązaniu zagadnień podsystemów połączonych "czysto" szeregowo czy równolegle, a następnie rozwiązaniu systemu przy traktowaniu podsystemów jako układów składowych.

TEST 3

P. Znaleźć transmitancje systemu podanego na rysunku 28.

xxx

O. Jeżeli odpowiedziałeś samodzielnie, prześledź uważnie poniższy wywód i porównaj go z uwagami zawartymi w punkcie 9.

$$G_{45} = G_4\ G_5$$

$$G_{345} = G_3 + G_{45} = G_3 + G_4\ G_5$$

$$G_{2345} = G_2\ \ G_3 + G_4\ G_5$$

$$G_{23456} = G_2\ G_{345}\ \ + G_6 = G_2\ \ G_3 + G_4\ G_5\ \ + G_6$$

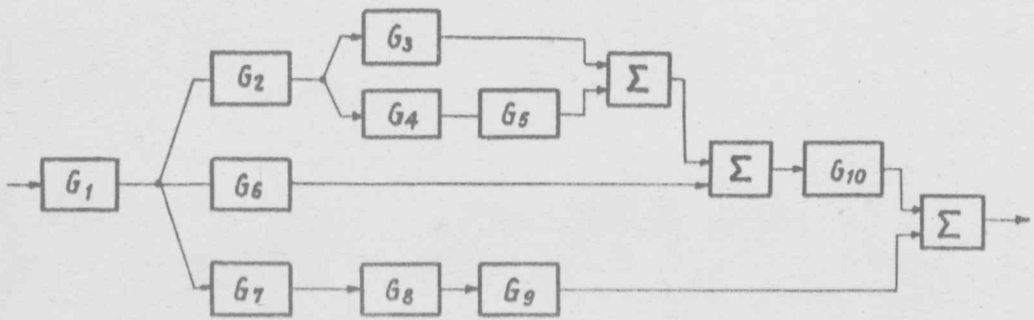
$$G_{2345610} = G_{10}\ G_{23456}\ \ = G_{10}\ G_2\ \ G_3 + G_4\ G_5\ \ + G_6$$

$$G_{789} = G_7\ G_8\ G_9$$

$$G_{2345678910} = G_{10}\ G_2\ G_3 + G_4 G_5\ \ + G_6 + G_7 G_8 G_9$$

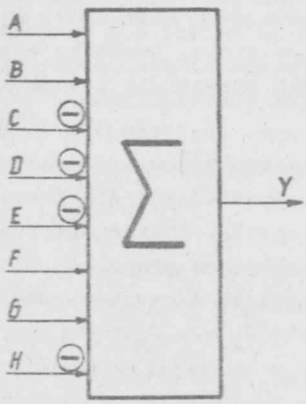
$$G = G_{12345678910} = G_1\ G_{10}\ G_2\ G_3 + G_4 G_5\ \ + G_6 + G_7 G_8 G_9$$

10. Przy poznawaniu systemów równoległych posługiwaliśmy się układem, który sumował wejściowe sygnały, dając na wyjściu sygnał sumaryczny. Taki wielowejściowy i jednowyjściowy



Rys. 28 System mieszany /do testu 3/

układ nazywać będziemy dalej sumatorem. Trzeba jednak wspomnieć o pewnej niewykorzystanej jeszcze możliwości sumatora: może on dodawać sygnały nadając niektórym z nich przeciwny znak, niż miały w rzeczywistości, czyli mówiąc obrazowo układ ten jest "dodawczo-odejmowaczem". Wejścia, na które podawane są sygnały do odejmowania, zaznaczać będziemy symbolem "-" pisanym koło wejścia. Przy tej umowie układ pokazany na rysunku 29 będzie wykonywał działanie:



Rys.29 Sumator

$$Y = A + B - C - D - E + F + G - H$$

2. Wpływ rodzaju sprzężenia na niezawodność systemu

1. Dotychczasowe rozważania prowadzone były dla systemów, których funkcje przejścia lub transmitancje były raz na zawsze ustalone i nie zależały od czasu. Tymczasem realne układy, będące częściami składowymi systemów, ulegają wpływom czynników niekontrolowanych /zakłóceń/, pod działaniem których zmieniają swoje właściwości w sposób mniej lub bardziej przypadkowy. Zmiany te w krańcowym przypadku mogą być tak dalekie i nieodwracalne, że mówimy o uszkodzeniu systemu. I tu dochodzimy do ważnego i z punktu widzenia praktyki ciekawego zagadnienia: jak budować systemy z zawodnych układów, tak aby system był niezawodny?

2. Podstawimy i udowodnimy na wstępie tezę, że szeregowe sprzężenie powoduje zwiększenie jego wrażliwości na uszkodzenia. Istotnie, niech system złożony będzie z n jednakowych układów połączonych szeregowo. Jasnym jest, że przy takim połączeniu uszkodzenie któregokolwiek elementu powoduje "zerwanie łańcucha", czyli przerwanie funkcjonowania systemu. Porównania z łańcuchem użyliśmy celowo, ponieważ większość dotychczasowych koncepcji zwiększania niezawodności systemów bazowania na wzmacnianiu najsłabszego ogniwa, czyli zwiększaniu niezawodności najbardziej zawodnych elementów. Wykażemy, że koncepcje te nie mogą dać zbyt wiele, ponieważ w samej strukturze szeregowego sprzężenia mieści się źródło niesprawności.

3. Załóżmy, że prawdopodobieństwo poprawnej pracy pojedynczego elementu /układu/ wynosi odpowiednio: dla pierwszego P_1, drugiego P_2, itd. aż do n-tego P_n. Prawdopodobieństwo poprawnej pracy całego systemu jest równe prawdopodobieństwu poprawnej pracy w s z y s t k i c h elementów, czyli równe jest iloczynowi prawdopodobieństw $P_1 \cdot \ldots \cdot P_n$.

$$P = P_1 \, P_2 \cdot \ldots \cdot P_n \qquad /19/$$

Jeśli najbardziej niesprawnym elementem będzie element "i", to prawdopodobieństwo poprawnej pracy całego systemu będzie nie większe niż P_i, gdyż wzór /19/ można przepisać w postaci:

$$P = P_i \cdot /P_1 . P_2 . \ldots . P_{i-1} \cdot P_{i+1} \ldots P_n/,$$

a wszystkie prawdopodobieństwa wchodzące w skład iloczynu w nawiasie są mniejsze lub równe jedności. Gdyby jednak udało się nam polepszyć pracę i-tego układu do tego stopnia, że byłby on zupełnie niezawodny /$P_i = 1$/, to niezawodność całego układu nie wzrosłaby przy tym ponad

$$P' = P_1 \, P_2 \cdot \ldots \cdot P_{i-1} \, P_{i+1} \cdot \ldots \cdot P_n$$

i byłaby dalej niska. Poprawianie najsłabszego elementu nie wnosi więc istotnej poprawy, chyba, że element ten miał rzeczywiście bardzo niską jakość w stosunku do pozostałych.

4. Rozważmy teraz system złożony z n układów o jednakowym prawdopodobieństwie poprawnej pracy P_1. Zgodnie ze wzorem /19/ prawdopodobieństwo poprawnej pracy całości będzie wynosiło

$$P = P_1^n \qquad /20/$$

Łatwo się przekonać, że dla $P_1 < 1$, czyli dla zawodnych układów składowych, prawdopodobieństwo poprawnej pracy systemu złożonego z n takich układów staje się dowolnie bliskie zeru dla dostatecznie dużych n, niezależnie od wartości P_1. Warto sobie równocześnie zdawać sprawę, że już dla P = 0,5 zawodność systemu jest tego rodzaju, że równie prawdopodobne jest działanie, jak i niedziałanie systemu!

TEST 1

P. System szeregowy złożony z 4 układów, których prawdopodobieństwa pracy są odpowiednio równe: P_1 = 0,9 P_2 = 0,8 P_3 = 0,9 i P_4 = 0,7

Wskazać najsłabsze ogniwo, obliczyć prawdopodobieństwo poprawnej pracy systemu ze "słabym ogniwem" i po zwiększeniu niezawodności najsłabszego układu najpierw do 0,9, a następnie do 1,0.

xxx

O. Niezawodność, zgodnie ze wzorem /19/ wynosi:

$$P = 0{,}9\ 0{,}8\ 0{,}9\ 0{,}7 = 0{,}4536$$

Najsłabszy jest układ 4. Jeśli Będziemy go wzmacniać otrzymamy kolejno:

$$P = 0{,}9\ 0{,}8\ 0{,}9\ 0{,}9 = 0{,}5812$$

$$P = 0{,}9\ 0{,}8\ 0{,}9\ 1{,}0 = 0{,}648$$

Wyciągnijmy dwa wnioski: Już nawet tak prosty system o stosunkowo niedużej zawodności układów, wykazuje bardzo dużą niepewność działania. Coż można więc mówić o systemach złożonych z setek czy tysięcy układów?! Po wtóre, zwiększenie niezawodności elementów niewiele daje. Aby się o tym ostatecznie upewnić zrobimy następne zadanie, przedtem jednak, jeśli miałeś trudności z tym zadaniem, koniecznie jeszcze raz przeczytaj punkty 2, 3, i 4.

P. Obliczyć niezawodność systemu złożonego kolejno z trzech, czterech, pięciu, sześciu, dziewięciu i dwudziestu siedmiu układów o prawdopodobieństwie poprawnego działania równym 0,99.

xxx

O. Mamy kolejno:

$0,99^3 = 0,97$ $0,99^4 = 0,96$ $0,99^5 = 0,95$ $0,99^6 = 0,94$

$0,99^9 = 0,91$ $0,99^{27} = 0,75$

Jak widać, nawet przy bardzo dużej niezawodności układów system staje się zawodny po przekroczeniu pewnej wielkości.

P. Przeprowadzić takie same obliczenia jak w poprzednim zadaniu, ale dla systemu złożonego z układów o niezawodności 0.95.

xxx

O. $0,95^3 = 0,86$ $0,95^4 = 0,82$ $0,95^5 = 0,78$ $0,95^6 = 0,74$

$0,95^9 = 0,64$ $0,95^{27} = 0,26$

W tym przypadku wpływ rosnących wymiarów systemu zaznaczył się jeszcze wyraźniej.

5. Problem niezawodności jest obecnie jednym z najtrudniejszych problemów cybernetyki, a jednocześnie, w związku z koniecznością tworzenia coraz większych systemów /technicznych, społecznych czy ekonomicznych/, jest on bardzo aktualny. Wytworzyła się więc na gruncie cybernetyki cała teoria niezawodności, której celem pośrednim jest określenie, jaką strukturę powinny mieć systemy, aby były niezawodnymi. Rozwiązalności tak postawionego zadania dowodzą dwa fakty: stru-

ktura /szeregowa/ potrafi pogorszyć niezawodność systemu poniżej niezawodności układów składowych, można więc oczekiwać również istnienia struktury polepszającej niezawodność. Ponadto mamy stale kontakt z pewnym bardzo wielkim systemem, który zbudowany z bardzo zawodnych elementów, działa niezawodnie. Jest to mózg człowieka. System ten zawiera niewyobrażalnie wielką liczbę 10^{10} komórek nerwowych, które są układami bardzo zawodnymi - codziennie obumiera ich w mózgu kilka tysięcy /bezpowrotnie, gdyż tkanka nerwowa nie regeneruje się/. Jednak ten ogromny system dzięki swej strukturze działa niezawodnie wiele lat. Nie wiemy jeszcze, jak on to robi, ale fakt, że to robi, stanowi dla nas doping i potwierdza słuszność założenia, że struktura pozwala zwiększyć niezawodność.

6. Rozważmy teraz system o strukturze równoległej. Jeśli postawimy pytanie, kiedy wielkość wyjściowa takiego systemu nie ulegnie w ogóle zmianie w wyniku nieprzewidzianych awarii układów, to dojdziemy do takiej samej konkluzji, jak w przypadku systemu szeregowego: system będzie niezawodny, jeśli wszystkie jego układy będą pracowały niezawodnie, a to zachodzi z prawdopodobieństwem określonym wzorem /19/.

Zauważmy jednak pewną istotną różnicę: O ile awaria któregokolwiek układu systemu szeregowego powodowała przerwanie pracy systemu i całkowitą zmianę wielkości wyjściowej, o tyle awaria jednego z równolegle sprzężonych układów jedynie nieznacznie odbija się na wartości wielkości wyjściowej i wpływ ten jest tym bardziej nieznaczny, im więcej układów zepniemy równolegle.

7. Rozważmy wobec tego zagadnienie, jak wielkie będzie prawdopodobieństwo tego, że sygnał na wyjściu systemu równoległego nie zaniknie zupełnie, pomimo że prawdopodobieństwa niezniknięcia sygnału /bezawaryjnej pracy/ poszczególnych układów systemu wynoszą tylko P_1, P_2 ..., P_n.

Sygnał zaniknie całkowicie, jeśli zniszczeniu ulegną w s z y s t k i e układy. Prawdopodobieństwo zniszczenia pierwszego układu wynosi $/1-P_1/$, drugiego $/1-P_2/$, itd. aż do n-tego $/1-P_n/$. Prawdopodobieństwo awarii wszystkich układów wynosi więc

$$P_a = /1-P_1/\ /1-P_2/\ \ldots\ /1-P_n/, \qquad /21/$$

a prawdopodobieństwo niezaniknięcia sygnału na wyjściu wynosi:

$$P = 1-P_a = 1 - /1-P_1/\ /1-P_2/\ \ldots\ /1-P_n/ \qquad /22/$$

Dla przypadku jednakowych prawdopodobieństw $P_1 = P_2 = \ldots = P_n$ wzór /22/ uprości się do postaci

$$P = 1 - /1-P_1/^n \qquad /23/$$

Z prawdopodobieństwem określonym wzorem /23/ możemy więc oczekiwać, że system pomimo awarii będzie działał, aczkolwiek jego wielkość wyjściowa może być różna od wielkości wyjściowej w stanie poprawnej pracy.

8. Rozważmy przypadek, kiedy każdy z układów spiętych w równoległy system sam może dawać wielkość wyjściową równą pożądanej, a sumator działa "z nasyceniem" : jeśli suma wielkości wejściowych do niego przekracza wielkość pożądaną, to zamiast sumy daje on wielkość pożądaną. Wówczas prawdopodobieństwo określone wzorem /5/ określa niezawodność systemu jako całości. Z takim właśnie prawdopodobieństwem otrzymamy z systemu pożądany sygnał, ale prawdopodobieństwo określone wzorem /23/ może być, przy dostatecznie dużym n, o wiele w i ę k s z e niż prawdopodobieństwo P_1, czyli system jest bardziej niezawodny od swoich elementów.

Przykład 1

Ośrodek zarządzania ma być powiadomiony o jakimś ważnym wydarzeniu, przy czym zależy nam na bardzo dużej niezawodności dotarcia sygnału. Wobec tego instalujemy w ośrodku nie jeden telefon /którego prawdopodobieństwo bezawaryjnej pracy w chwili przekazywania wiadomości P_1 nas nie zadawala/, lecz kilka, np. dwa telefony, i przekazujemy wiadomość przez obydwa. Telefonista, niezależnie od tego z ilu telefonów otrzymał wiadomość, przekaże ją tylko raz /"sumator z nasyceniem"/, lecz bezawaryjność pracy takiego systemu powiadamiania jest większa. Jest to powszechnie znany i stosowany sposób zwiększania niezawodności przez tworzenie rezerw, doszliśmy jednak do tego sposobu nie intuicyjnie, a poprzez ścisłe rozumowanie. Pozwala to między innymi uściślić niejasne przeświadczenie osób stosujących podobne systemy, że "jeśli dam dwa telefony zamiast jednego, to mam dwa razy większą pewność". Otóż jak łatwo zauważyć ze wzoru /23/ "pewność" zwiększy się mniej, niż dwa razy. Jeszcze dalsze od prawdy są podobnie naiwne rozumowania przy jeszcze większych n.

9. Systemy opisane w punkcie 8 i przykładzie 1 nazywać będziemy systemami Z E Z W I E L O K R O T N I E N I E M. Zwielokrótnienie jest istotnie skutecznym sposobem zmniejszania zawodności systemu w stosunku do jego układów składowych, jest jednak metodą kosztowną. Istotnie, tworzenie rezerw, których jedynym zadaniem jest zwiększanie niezawodności, jest trudne do przyjęcia na szerszą skalę, jest jednakże na obecnym etapie rozwoju teorii i praktyki cybernetycznej drogą jedyną. Tym bardziej drażniący jest przykład mózgu, dla którego wykazano, że niezawodność jego działania oparta jest przypuszczalnie na innych zasadach niż zwielokrotnieniowe.

TEST 2

P. Mamy system, złożony ze spiętych równolegle trzech układów, o prawdopodobieństwach działania $P_1 = 0,9$ $P_2 = 0,8$ i $P_3 = 0,9$. Obliczyć prawdopodobieństwo, że na wyjściu nie zaniknie sygnał.

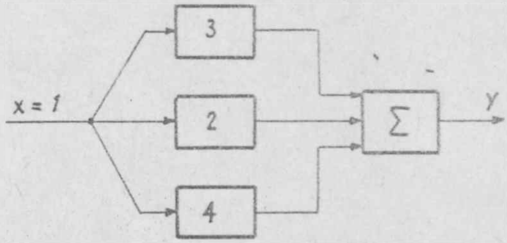
xxx

O. $P = 1 - /1-0,9/\ /1-0,8/\ /1-0,9/ = 1 - 0,1\ 0,2\ 0,1 = 1-0,002$

$$P = 0,998$$

Jak widać niezawodność jest znaczna.

P. Jeśli system, o którym mówiono w poprzednim zadaniu, przedstawiony jest na rysunku /30//podano tam wartość wielkości wejściowej i wszystkie transmitancje/, to jakie wartoś-



Rys. 30 System zawodny

ci będzie przyjmowała wielkość wyjściowa systemu Y i z jakimi prawdopodobieństwami?

xxx

O. Jeśli nie zrobiłeś zadania, to przed przystąpieniem do wnikliwej analizy rozwiązania przeczytaj przykład 2.

Wielkość wyjściowa	Prawdopodobieństwo wystąpienia
3+2+4 = 9	$P_1P_2P_3 = 0,648$
3+2 = 5	$P_1P_2/1-P_3/ = 0,072$
3 + 4 = 7	$P_1/1-P_2/p_3 = 0,162$
2 + 4 = 6	$/1-P_1/P_2\ P_3 = 0,072$
3	$P_1/1-P_2/\ /1-P_3/ = 0,018$
2	$/1-P_1/\ /1-P_2/P_3 = 0,018$
4	$/1-P_1/P_2/1-P_3/ = 0,008$

Prawdopodobieństwo wystąpienia jakiejkolwiek niezerowej wartości na wyjściu jest równe

$$P = 0,648 + 0,072 + 0,162 + 0,072 + 0,018 + 0,018 + 0,008 = 0,998 .$$

zgodnie z wynikiem poprzedniego zadania.

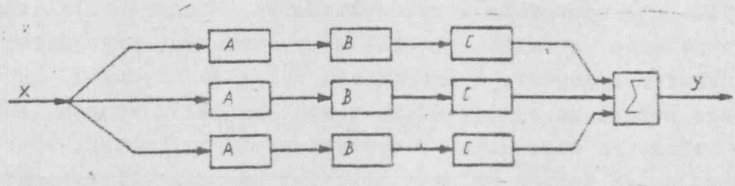
Przykład 2

Jeśli chcemy określić prawdopodobieństwo wystąpienia określonej kombinacji wielkości wyjściowych układów systemu, musimy obliczyć prawdopodobieństwo zdarzenia, polegające na równoczesnym wystąpieniu wielkości wyjściowych pewnych układów, i niewystąpieniu innych. Zatem jeśli chcemy wiedzieć jakie jest prawdopodobieństwo wystąpienia sygnałów wyjściowych na drugim i trzecim wyjściu a niewystąpienia na pierwszym, musimy obliczyć iloczyn odpowiednich prawdopodobieństw, pamiaętając, że niewystąpienie sygnału na pierwszym wyjściu zachodzi z prawdopodobieństwem $/1-P_1/$. Mamy więc:

$$P_{23} = /1-P_1/P_2\ P_3 = 0,1\quad 0,8\quad 0,9 = 0,072$$

TEST 3

P. Budulcem dla naszego systemu będą układy posiadające niezawodność $P_1 = 0,9$. Zmuszeni jesteśmy zbudować z nich system zawierający trzy takie układy sprzężone szeregowo. Rys. 31. Ilu układów musielibyśmy użyć, aby zbudować sy-



Rys. 31 Zwielokrotniony system szeregowy

stem ze zwielokrotnieniem, posiadający niezawodność nie mniejszą niż 0,95?

xxx

O. Niezawodność podsystemu sprzężonego szeregowo bez zwielokrotnienia wynosi:

$$P_s = 0{,}9^3 = 0{,}729$$

Aby spełnić warunki zadania należy rozwiązać nierówność

$$1 - /1 - 0{,}729/^x \geqslant 0{,}95$$

$$0{,}05 \geqslant /0{,}271/^x$$

$$\log 0{,}05 \geqslant x \log 0{,}271$$

$$x \geqslant \frac{\log 0{,}05}{\log 0{,}271} \quad 2{,}28$$

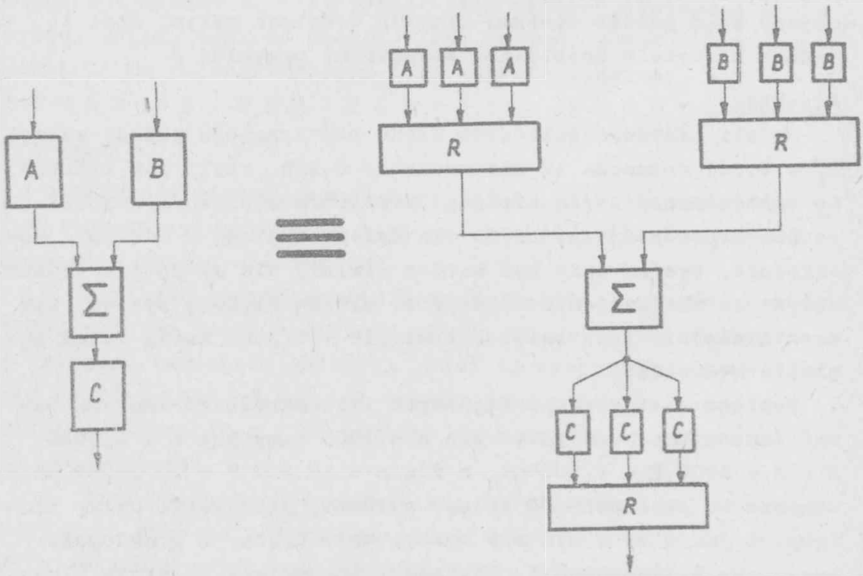
Ponieważ liczba równoległych gałęzi musi być całkowita, przyjmujemy trzy gałęzie. Dokonujemy sprawdzenia niezawodności:

$$P = 1 - 0{,}271^3 = 1 - 0{,}020 = 0{,}980$$

Wynika z tego, że do budowy systemu potrzebne nam będzie 9 układów: trzy równoległe gałęzie po trzy układy szeregowo sprzężone w każdej, oraz jeden "sumator z nasyceniem".

10. Dotychczas rozważaliśmy zawodność układów polegającą na ich zupełnej niesprawności: zepsuty układ nie przekazywał żadnych sygnałów. Jest to sytuacja wyidealizowana i łagodniejsza niż spotykane w rzeczywistości. Mianowicie niesprawności układów wchodzących w skład systemu mogą przejawiać się nie brakiem sygnałów, a obecnością sygnałów błędnych. Ma to miejsce zwłaszcza w systemach, w których skład wchodzą ludzie jako układy, a więc m.in. w systemach ekonomicznych. Teorię niezawodności takich układów stworzył amerykański matematyk J. von Neumann. Brak tu miejsca na przytoczenie nawet skrótu tej teorii, warto jednak przytoczyć jej podstawową myśl: każdy z układów systemu zastępuje się przez n układów połączonych równolegle, ale zamiast sumatora czy "sumatora z nasyceniem" stosuje się układ nazwany przez von Neumanna "urządzeniem rejestrującym". Układ ten dokonuje na wielkościach wejściowych z układów równolegle sprzężonych, swoistego głosowania i czyni swoje wejście /a tym samym wyjście z "podsystemu"

zastępującego pojedynczy układ/ równe wyjściu reprezentowanemu przez większość układów. Jest to rozsądne, ponieważ jeśli układy się mylą lub kłamią, to mało prawdopodobne jest, aby większość z nich myliła się tak samo. Jeśli każdy układ daje inny wynik, "urządzenie rejestrujące" sygnalizuje awarię. Rys. 32 przedstawia zwielokrotnienie systemu przy n=3 .



Rys. 32. System zwielokrotniony

Pomysł tu przytoczony wydaje się być dość banalnym bo tak właśnie postępujemy i my w warunkach niepewności. Jednak, podobnie jak przy naszych rozważaniach na temat systemów ze zwielokrotnieniem, mamy teraz dowód, że postępowanie takie jest w danej grupie rozwiązań optymalne. Z teorii von Neumanna wynika również wzór, pozwalający oszacować prawdopodobieństwo błędu dla całego systemu, jeśli zamiast każdego jego zawodnego układu zastosowano n takich samych układów połączonych z "urządzeniem restaurującym". Wzór ten można zapisać krótko w postaci:

$$P = a\, n^{-1/2}\, 10^{-bn} \qquad /24/$$

gdzie n oznacza wielkość zwielokrotnienia /ilość równolegle wpiętych układów/, a stałe a i b zależą od prawdopodobieństwa błędu pojedynczego układu P_1. Zatem dając dostatecznie duże n możemy błąd całego systemu uczynić dowolnie małym. Jest to jednak niebywale kosztowne. Rozpatrzmy przykład 3.

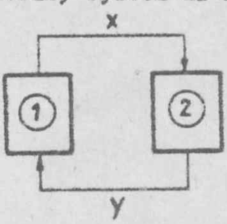
Przykład 3

Jeżeli prawdopodobieństwo błędu pojedynczego układu wynosi $P_1 = 0{,}005$ /oznacza to niezawodność 0,995, czyli już wysoką/, to zastosowanie tysiąckrotnego zwielokrotnienia /n = 1000/ daje prawdopodobieństwo błędu dla całego systemu P = 0,027. Naturalnie, system może być bardzo wielki, nie ma to już jednak wpływu na wielkość niezawodności, ale im większy system, tym kosztowniejsze jego zwielokrotnienie - trzeba każdy układ powielić n-krotnie!

Dopiero zastosowanie większych zwielokrotnień zaczyna dawać imponujące rezultaty: dla n = 2000 mamy już P = 0,0026, dla n = 5000 P = 0,000004, a dla n = 10 000 P = 0, 0000000016. Oznacza to niezawodność całego systemu, jakkolwiek byłby złożony, i jakie by w nim nie występowały błędy, z prawdopodobieństwem 0,9999999984. Doprawdy, imponujący to wynik, lecz osiągnięty za cenę budowy systemu dziesięć tysięcy razy większego, niż jest nam potrzebny.

3. Sprzężenie zwrotne

1. Rozpatrzmy system na rysunku 33. Są tam dwa układy: 1 i 2.



Rys. 33 Sprzężenie zwrotne

Zacznijmy rozpatrywanie tego systemu od układu 1. Wielkością wejściową jest y, wyjściową x. Wielkość wyjściowa podana jest na wejście następnego układu 2, dając wielkość ... wejściową y. A zatem wartość wielkości wejściowej y zależy poprzez funkcje przejścia obydwu układów od ... siebie samej. Identyczny wniosek uzyskamy zaczynając rozpatrywanie systemu od wielkości x i układu 2. W systemie takim wartość wielkości wyjściowej zależy sama od siebie, jest bowiem z w r o t n i e podawana na wejście. System taki nazywać będziemy S Y S T E M E M Z E S P R Z Ę Ż E N I E M Z W R O T N Y M.

Przykład 1

E. Wielkość dochodu narodowego zależy od wielkości inwestycji produkcyjnych, a ta jest zależna od wielkości dochodu. Mamy więc sprzężenie zwrotne.

T. Wielkość oporów mechanicznych /tarcia i oporów powietrza/ podczas ruchu samochodu zależy od szybkości samochodu, a szybkość samochodu zależy z kolei od wielkości oporów tarcia /a właściwie od różnicy oporów tarcia i siły napędowej silnika/. Mamy więc i tu zwrotne uzależnienie szybkości od od samej siebie.

B. Żywy organizm komunikuje się z otoczeniem poprzez efektory /np. ruchowe/. W wyniku oddziaływań na otoczenie organizm dokonuje zmian w otoczeniu, a tym samym w oddziaływaniu otoczenia na niego i dostrzegając to przy pomocy receptorów /wzroku/ modyfikuje swoje działanie itd. Zachodzi zatem stałe zwrotne sprzężenie organizmu i środowiska.

M. W procesie obliczeniowym wielkości wyników pośrednich decydują o kierunku dalszych obliczeń, np. przy rozwiązywaniu równania kwadratowego o przebiegu dalszych obliczeń decyduje wartość i znak wyróżnika . A więc wynik oddziaływuje na dalsze obliczenia, które decydują o nowych wynikach itd.

2. Sprzężenie zwrotne może prowadzić do wzmocnienia zaistniałych w systemie zmian lub im przeciwdziałać. Jeżeli rozwa-

żymy system, w którym dwa układy są sprzężone zwrotnie w ten sposób, że wzrost sygnału Y na wejście pierwszego układu powoduje wzrost sygnału na wyjściu tego układu X, zaś wzrost sygnału na wejściu X drugiego układu powoduje wzrost sygnału na jego wyjściu Y, to jakiekolwiek odchylenie od zera którejkolwiek z wielkości X i Y będzie się w miarę upływu czasu pogłębiało. Sprzężenie zwrotne, przy którym oba sprzężone układy reagują wzrostem wielkości wyjściowej na wzrost wielkości wejściowej, nazywamy D O D A T N I M S P R Z Ę Ż E N I E M Z W R O T N Y M.

Przykład 2

Systemem z dodatnim sprzężeniem zwrotnym jest system złożony z nienormalnego palacza i kotła parowego. Jeśli palacz stwierdzi, że ciśnienie pary w kotle wzrosło, to zacznie ze zdwojoną energią dorzucać węgla na palenisko, co spowoduje dalszy wzrost ciśnienia pary, dopingujący palacza do zwiększenia wysiłku, itd. Przykład ten sugeruje, że sprzężenie zwrotne, gdy jest dodatnie, kryje w sobie pewne groźne możliwości.

3. Efekty sprzężenia zwrotnego dodatniego polegające na wzmacnianiu raz zainicjowanych zmian mogą być zarówno szkodliwe, jak i pożyteczne. Przytoczymy teraz po dwa przykłady z różnych dziedzin, kiedy sprzężenie dodatnie raz prowadzi do zjawisk pożądanych, a raz do katastrofalnych.

Przykład 3

E. Po złożeniu pieniędzy w PKO otrzymujemy odsetki, które są tym większe, im większą kwotę złożymy. Z kolei odsetki te powiększają złożony kapitał, powodując dalsze zwiększenie odsetek itd. Ten system ze sprzężeniem zwrotnym dodatnim w znany sposób powoduje coraz szybszy wzrost zgromadzonego kapitału; że jest to zjawisko korzystne, nie trzeba chyba nikogo specjalnie przekonywać.

T. Przykładem niekorzystnego wpływu dodatniego sprzężenia zwrotnego w technice jest zjawisko powstawania uszkodzeń części np. mechanicznych. Jeśli dwie współpracujące części ulegną częściowemu zużyciu, to powstałe przy tym luzy wpływać będą na wzrost zużycia, który powiększy luz itd.

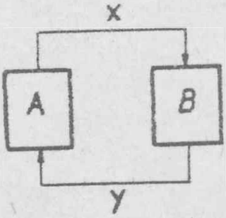
B. Korzystnym zjawiskiem w biologii dla określonego gatunku zwierząt jest fakt lawinowego przyrostu populacji: ilość potomstwa jest tym większa, im więcej jest osobników dorosłych, a tych w następnym pokoleniu jest tym więcej, im więcej było potomków; w efekcie następuje szybki rozwój gatunku, łatwy do zaobserwowania, szczególnie u gryzoni.

E. Niekorzystnym objawem działania dodatniego sprzężenia zwrotnego jest inflacja: wzrost cen zmusza do wzrostu poborów ludności, a ten z kolei, przy brakach na rynku, wymusza wzrost cen itd.

T. Korzystnie wpływa dodatnie sprzężenie przy detonowaniu materiałów wybuchowych - temperatura spalanego materiału jest tym wyższa, im szybsze jest spalanie, a spalanie jest tym szybsze, im wyższa jest temperatura. W efekcie następuje detonacja.

B. Pewne choroby psychiczne są wynikiem zaistnienia dodatniego sprzężenia zwrotnego: jeśli chory uroi sobie, że ktoś na niego "dybie", to wpłynie to na wzrost podenerwowania i zmniejszy krytycyzm obserwacji, co z kolei powoduje pogłębienie urojeń, poprzez dostrzeganie fikcyjnych dowodów, co pogarsza stan nerwowy itd.

4. Przytoczone przykłady mogły sugerować, że wynikiem działania dodatniego sprzężenia zwrotnego musi być nieograniczony wzrost obu wielkości wyjściowych w układach podsystemu. Otóż tak jest najczęściej, ale tak być wcale nie musi. Aby się przekonać, od czego to zależy, rozpatrzmy przykład, w którym /rys. 34/ układy mają transmitancję odpowiednio równą A i B dla pierwszego i drugiego układu. Niech w chwili początkowej wartość X wynosi x_0. Wówczas w kolejnych chwilach czasowych wielkości X i Y będą przyjmowały wartości określone tabelą:



Rys. 34 System ze sprzężeniem zwrotnym dodatnim

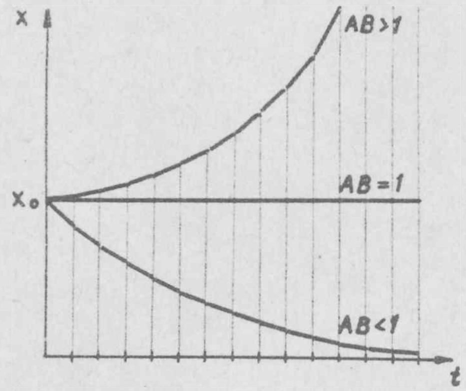
chwila czasu	wartość X	wartość Y
1	x_0	$B\ x_0$
2	$AB\ x_0$	$AB^2\ x_0$
3	$A^2B^2\ x_0$	$A^2B^3\ x_0$
4	$A^3B^3\ x_0$.	$A^3B^4\ x_0$

co można zapisać ogólnie w postaci wzoru:

$$X/n/ = A^{n-1}\ B^{n-1}\ x_0$$

czyli wartości zmiennej X w kolejnych chwilach czasowych są równe wartości początkowej, mnożonej przez iloczyn transmitancji układów w coraz wyższych potęgach. Przebieg tej zależności czasowej warunkowany jest wartością iloczynu AB. Jeśli AB 1, to kolejne wartości X szybko rosną i efekt jest właśnie taki, jakiego oczekiwaliśmy od systemu z dodatnim sprzężeniem. Jeśli AB = 1, to przebieg nie zmienia wartości początkowej, natomiast przy AB 1 przebieg maleje do zera /rys. 35/.

5. Wprowadzimy nowe pojęcie: mianowicie system ze sprzężeniem zwrotnym /lub inny system/, w którym przebiegi samorzutnie rosną nieograniczenie, nazwiemy systemem N I E - S T A B I L N Y M. System, w którym przebiegi wprawdzie nie rosną, ale i nie chcą maleć, nazwiemy systemem na granicy stabilności, natomiast system, w którym przebiegi /bez in-



Rys. 35 Zachowanie systemu ze sprzężeniem dodatnim

gerencji z zewnątrz systemu/ maleją, nazwiemy S T A B I L - N Y M. Zatem opisany system z dodatnim sprzężeniem zwrotnym jest stabilny dla AB 1, a niestabliny dla AB 1. Warunek, przy spełnieniu którego system jest stabilny jest nazywany W A R U N K I E M S T A B I L N O Ś C I. A zatem warunkiem stabilności dla omawianego systemu jest AB 1.

TEST 1

P. Czy podczas czytania tej książki funkcjonuje jakaś forma sprzężenia zwrotnego?

xxx

O. Tak, ponieważ sposób czytania tekstów wyjaśniających wpływa na ilość poprawnych odpowiedzi w testach, a te z kolei decydują o sposobie czytania tekstów, gdyż wskazują, że pewne partie można pominąć, a inne trzeba przeczytać jeszcze raz. Na przykład zła odpowiedź na postawione pytanie wskazuje na potrzebę ponownego przeczytania przykładu 1.

P. Czy pomiędzy nauczycielem a uczniem w procesie nauczania występuje sprzężenie zwrotne? Czy jest ono dodatnie? Czy jest ono stabilne?

xxx

O. Zwiększenie ilości wiadomości przekazanych przez nauczyciela powoduje zwiększenie ilości wiedzy przyswojonej przez ucznia. Gdy fakt przyswojenia zostanie stwierdzony przez nauczyciela, zwiększa on dawkę wiedzy przekazywaną uczniowi, co zmusza tego ostatniego do zwiększenia zasobu swoich wiadomości itd., czyli zachodzi sprzężenie zwrotne dodatnie. Jest ono niestabilne, ponieważ prowadzi do stałego zwiększania zasobu wiadomości u ucznia i ilości wiadomości przekazanych przez nauczyciela. Niestabilność ta jest nader korzystna, gdyż warunkuje postęp.

P. Czy możliwe jest, aby system niestabilny stał się stabilnym lub odwrotnie?

xxx

O. Tak, ponieważ warunek stabilności systemu wyrażony jest poprzez parametry układów składowych /transmitancje/, a te mogą pod działaniem zakłóceń ulegać zmianom.

P. Czy możesz wyjaśnić, dlaczego układ na granicy stabilności należy traktować jako potencjalnie niestabilny?

xxx

O. Nawet nieznaczna zmiana parametrów któregoś z układów systemu może powodować, że warunek granicy stabilności zmieni się na warunek niestabilności /w omawianym w punktach 4 i 5 systemie wystarczy, aby nieco tylko zwiększyła się jedna z transmitancji/.

6. Jeżeli jeden z dwu układów przedstawionego na rysunku 33 systemu reaguje w ten sposób, że zwiększenie jego wielkości wejściowej powoduje zmniejszenie wielkości wyjściowej, to nazwiemy go systemem ze S P R Z Ę Ż E N I E M Z W R O T N Y M U J E M N Y M.

Przykład 4

Powróćmy do znanego nam z przykładu 2 palacza, ale teraz palacz, widząc, że ciśnienie pary rośnie, z m n i e j s z y ilość dosypywanego węgla, co spowoduje obniżenie ciśnienia, a to zapewni zwiększenie intensywności palenia itd. Zachowanie takie będzie typowe dla ujemnego sprzężenia zwrotnego, ponieważ teraz wzrost ciśnienia powoduje zmniejszenie ilości podawanego paliwa. /Kocioł dalej reaguje wzrostem ciśnienia na wzrost ilości paliwa/.

7. Ujemne sprzężenie zwrotne jest jednym z podstawowych mechanizmów spotykanych w przyrodzie, technice i życiu gospodarczym. Zwykle /dokładniej przedyskutujemy to w punkcie 8/ prowadzi ono do stabilizacji działania systemu, co jest zjawiskiem ze wszech miar korzystnym.

Przykład 5

E. Mechanizm ujemnego sprzężenia zwrotnego stabilizuje cenę w warunkach gospodarki wolnorynkowej. Mianowicie wzrost popytu powoduje wzrost ceny, co wpływa na zmniejszenie popytu, to z kolei powoduje obniżenie ceny itd. Analogiczne, ale w przeciwnym kierunku, są wpływy ceny na podaż i podaży na cenę. Mechanizm ten może doprowadzić albo do stabilizacji ceny na poziomie ceny równowagi ostatecznej, albo drgania /oscylacje cen/ będą się potęgowały, prowadząc raz do kryzysów nadprodukcji, a raz do spekulacji i załamań rynku.

T. Grzejnik z termostatem jest technicznym układem ze sprzężeniem zwrotnym ujemnym. Wzrost temperatury ponad normę powoduje oddziaływanie termostatu na grzejnik, zmierzające do zmniejszenia mocy grzejnej /zakręcenie kaloryfera/, a to powoduje spadek temperatury, który jeśli jest zbyt silny, powoduje oddziaływanie termostatu zmierzające do zwiększenia mocy grzejnej, co daje wzrost temperatury itd.

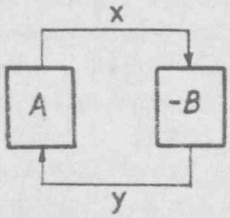
B. Ujemne sprzężenie zwrotne chyba po raz pierwszy było badane i opisywane właśnie na gruncie biologii pod nazwą homeostazy. Przez homeostazę rozumiano zdolność organizmów żywych do utrzymywania stałych parametrów /temperatura, skład chemiczny/ wewnętrznych. przy zmieniających się warunkach środowiska. Mechanizmy takie odpowiedzialne są za utrzymanie stałej ciepłoty ciała, za zwiększanie lub zmniejszanie wentylacji płuc w zależności od wysiłku /a dokładniej w zależności od zawartości dwutlenku węgla we krwi/, za odnawianie zapasu wody itd. Rozpatrzmy jednak inny przykład, a mianowicie oddziaływanie wzajemne osobnika i środowiska. Bodziec pochodzący od środowiska zwiększa się, wobec czego osobnik zwiększa intensywność swych działań obronnych, co powoduje zmniejszenie bodźca /np. ucieczka od intensywnego dźwięku/ itd.

9. Już podczas lektury przykładu 5 nasuwały się chyba uważnemu czytelnikowi wątpliwości, czy istotnie ujemne sprzężenie zwrotne zawsze działa stabilizująco. Odpowiedź ostateczna jest negatywna. W wyniku działania sprzężenia układ może się ustabilizować, ale mogą też pojawić się niegasnące oscylacje lub, co gorsza, oscylacje narastające. Od czego to zależy? Zanim damy odpowiedź matematyczną w postaci znowu warunku stabilności, spróbujemy ją przewidzieć dedukcyjnie.

Przykład 6

Łódka na wodzie przechyliła się, wobec czego wioślarz przeważył ją ciałem w przeciwnym kierunku /gdyby przeważał ją w tym samym kierunku mielibyśmy sprzężenie zwrotne dodatnie/. W wyniku tego łódka powróciła do położenia początkowego lub nawet lekko przeważyła w przeciwną stronę. W tym ostatnim przypadku wioślarz ponownie wychyleniem ciała doprowadził ją do pionu i tak drogą bezpośredniego wyprostowania lub kilku kolejnych coraz słabszych wychyleń łódka ustabilizowała się. Jeśli jednak wioślarz zadziała zbyt energicznie, to wychylenie łódki w przeciwną stronę będzie wcale nie mniejsze niż początkowe, co spowoduje nową interwencję i nowe wychylenie w przeciwną stronę itd. Oscylacje mogą nie zanikać, ale wręcz narastać. Ten sam efekt, co zbyt nerwowy wioślarz, może spowodować zbyt czuła na balansowanie wioślarza łódka, krótko mówiąc; jeśli zbyt duże są występujące w systemie transmitancje układów /transmitancja wioślarza to stosunek gwałtowności jego reakcji do wielkości wychylenia łódki, a transmitancja łódki to stosunek wielkości wyważania przez wioślarza do wielkości wychylenia/, to układ ze sprzężeniem zwrotnym ujemnym destabilizuje się.

Nadajmy teraz naszym rozważaniom formę matematyczną. W tym celu rozpatrzmy system przedstawiony na rysunku 36. Zwracamy



Rys. 36 System ze sprzężeniem zwrotnym ujemnym

uwagę, że transmitancja wiążąca X na wejściu i Y na wyjściu /drugi układ/ wynosi -B, a nie B jak na rysunku 34. Układ jako całość odpowiada więc wzrostem sygnału wyjściowego na zmniejszenie sygnału wejściowego, zapewniając ujemne sprzężenie w systemie.

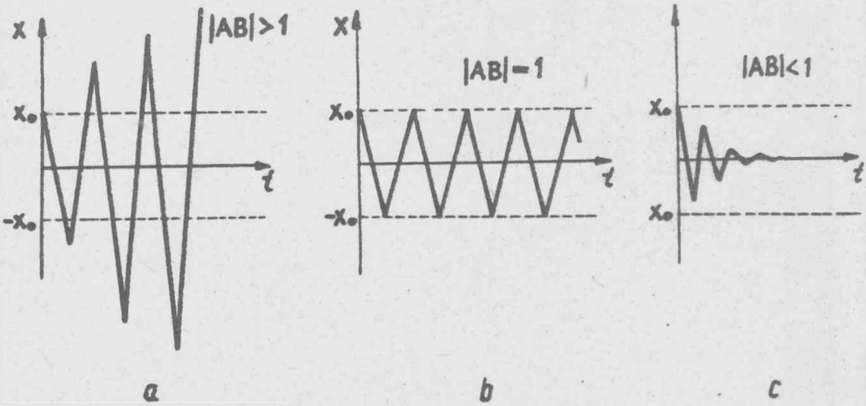
Podobnie, jak w punkcie 4, rozpatrzmy zachowanie systemu wychodząc od stanu początkowego $X = x_o$.

chwila czasu	wartość X	wartość Y
1	x_o	$-Bx_o$
2	$-AB\ x_o$	$+A\ B^2\ x_o$
3	$+\ A^2B^2\ x_o$	$-A^2B^3\ x_o$
4	$-A^3B^3\ x_o$	$+A^3\ B^4\ x_o$

co można zapisać ogólnym wzorem

$$X/n/ = / - AB/^{n-1}\ x_o \qquad /25/$$

Zatem wartości zmiennej X w kolejnych chwilach czasowych będą na przemian dodatnie i ujemne, a ponadto będą rosły co do wartości bezwzględnej przy AB 1, pozostawały na stałym poziomie przy AB = 1 i malały do zera przy AB 1. /rys.37/.



Rys.37 Zachowanie systemu ze sprzężeniem zwrotnym ujemnym

Opierając się na warunku stabilności wprowadzonym w punkcie 5 możemy stwierdzić, że system jest stabilny jeśli AB 1, jest na granicy stabilności przy AB = 1 i jest niestabilny przy AB 1.

10. Widzimy więc, że zakorzenione wśród fachowców z dziedziny automatyki przeświadczenie o niestabilności systemów z dodatnim sprzężeniem zwrotnym i stabilności systemów z ujemnym sprzężeniem zwrotnym nie jest w pełni uzasadnione. Najczęściej jednak spotyka się stabilne systemy z ujemnym sprzężeniem i niestabilne z dodatnim.

TEST 2

P. Jeśli obserwujemy system, o którym wiemy, że jest systemem z ujemnym sprzężeniem zwrotnym i stwierdzimy w tym systemie obecność drgań o ustalonej amplitudzie, to czy jest to powód do obaw?

xxx

O. Owszem, ponieważ dowodzi to przebywania systemu na granicy stabilności, a to, jak stwierdziliśmy w teście 1, jest niebezpieczne. Przy złej odpowiedzi przeczytaj jeszcze raz punkt 9.

P. Czy na stabilność systemu /z dodatnim lub ujemnym sprzężeniem zwrotnym/ można wpłynąć przez wprowadzenie odpowiednich sygnałów zewnętrznych?

xxx

O. Nie. Warunek stabilności systemu zależy od transmitancji wchodzących w jego skład układów, a więc od struktury systemu, a nie od sygnałów docierających do systemu. Dokładniej rozpatrzymy to zagadnienie dalej, podczas rozważania systemów ze sprzężeniami zwrotnymi, zaopatrzonych w wejścia i wyjścia.

P. Opisać dynamikę kształtowania się ceny rynkowej w terminach systemów ze sprzężeniem zwrotnym, przyjmując, że popyt podporządkowany jest prawu:

$$P_p/n/ = A \cdot C/n/ + a,$$

gdzie $P_p/n/$ oznacza popyt w danym okresie czasu, $C/n/$ - cenę w danym okresie, natomiast A i a są stałymi współczynnikami, charakteryzującymi rynek. Podobnie dla podaży mamy:

$$P_d/n/ = B \cdot C/n-1/ + b,$$

gdzie $P_d/n/$ jest podażą w danym okresie, $C/n-1/$ ceną w okresie poprzednim /producent przy podjęciu produkcji kieruje się ceną, jaka jest w chwili, gdy podejmuje produkcję, a nie ceną, jaka będzie, gdy wyrób będzie gotowy/. Określić funkcje przejścia, a następnie metodą przyrostową zlinearyzować je, wprowadzając transmitancje. Określić, jakie przebiegi będą miały miejsce w systemie, w jakich warunkach, i jaki jest warunek stabilności systemu.

xxx

O. Do budowy systemu użyjemy trzech układów. Jeden z nich na podstawie ceny z poprzedniej chwili czasowej /n-1/ ustala podaż towarów,

$$P_d/n/ = B\ C/n-1/ + b$$

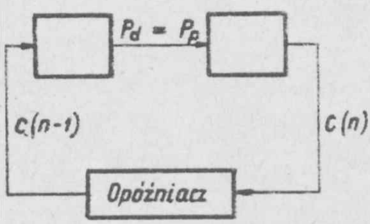
drugi na podstawie równowagi podaży i popytu /warunek okresowej równowagi rynku/ ustala cenę w danym okresie:

$$C/n/ = \frac{P_p/n/ - a}{A} = \frac{P_d/n/ - a}{A}$$

ponieważ $P_p/n/ = P_d/n/$

Trzeci układ działa jak biurokrata - nic nowego nie wnosi, a jedynie opóźnia przepływającą informację - mając na wejściu cenę aktualną, podaje na wyjściu cenę z poprzedniego okresu.
Nazwiemy go opóźniaczem.

Naturalnie trudno mówić o transmitancjach poszczególnych bloków, gdyż występujące stałe czynniki a i b komplikują funkcję przejścia. Omawiany system przedstawiony jest na rysunku 38a , z którego łatwo wyciągniemy wniosek, że



Rys. 38a Model dynamiki ceny rynkowej

a

jest to system ze sprzężeniem zwrotnym - cena zależy od samej siebie, z poprzedniego okresu czasu. Sprzężenie zwrotne będzie ujemne, jeśli stałe A i B będą różnych znaków. Tak też jest najczęściej w rzeczywistości: stała A jest ujemna /popyt maleje ze wzrostem ceny/, a B dodatnia. Jeżeli jednak /np. ze względu na snobizm nabywców/ stała A byłaby dodatnia, albo przy subwencjonowanej produkcji tanich produktów stała B była ujemna, wówczas mielibyśmy sprzężenie zwrotne dodatnie ze wszystkimi jego konsekwencjami.

Dalsza analiza ilościowa, a w szczególności badanie stabilności niemożliwe jest bez wprowadzenia transmitancji. Posłużymy się tu znaną z testu 3 rozdziału I podr.2 metodą przyrostową. Ustalimy więc najpierw stan spoczynkowy, czyli tzw. cenę równowagi ostatecznej, ze wzoru wynikającego z warunku równowagi w stanie stałej ceny:

$$A\,C_r + a = B\,C_r + b, \qquad /A \neq B$$

gdzie C_r jest ceną równowagi. Obliczając otrzymamy:

$$C_r = \frac{b - a}{A - B}$$

Wprowadzimy teraz zamiast ceny C/n/ wielkość odchylenia ceny od ceny równowagi:

$$\Delta C/n/ = C/n/ - C_r$$

z czego otrzymamy następujące równania dla podaży i popytu:

$$P_d/n/ = B\,C_r + B\,\Delta C/n-1/ + b$$

$$C_r + \Delta C/n/ = \frac{P_d/n/ - a}{A}$$

po podstawieniu pierwszego równania do drugiego mamy:

$$C_r + \Delta C/n/ = \frac{B\,C_r + b - a + B\;C/n-1/}{A}$$

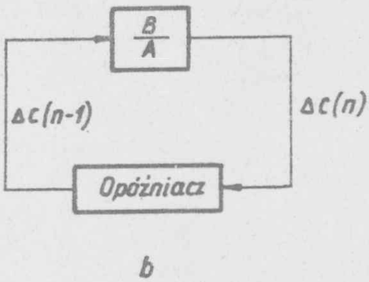
ale

$$\frac{B\,C_r + b - a}{A} = C_r$$

zatem

$$\Delta C/n/ = \frac{B}{A}\;\Delta C/n-1/$$

Nasz system zlinearyzowany metodą przyrostową zawiera więc dwa bloki: układ o transmitancji $\frac{B}{A}$ i opóźniacz, spięte sprzężeniem zwrotnym /rys. 38b/. Zbadamy stabilność



Rys.38b Model dynamiki ceny rynkowej

tego systemu. W tym celu założymy, że z jakichś powodów nastąpiło odchylenie od ceny równowagi o $\Delta C/1/$.

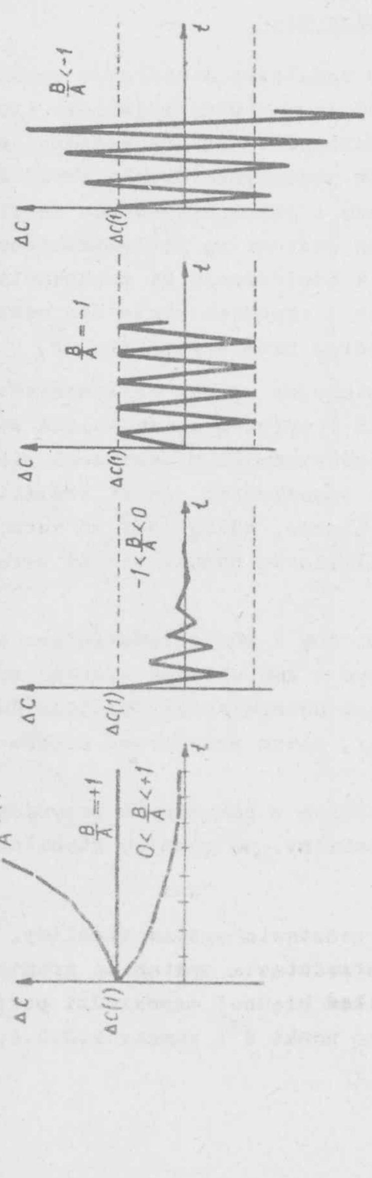
Ciąg kolejnych odchyleń ma postać:

$$\Delta C/2/ = \frac{B}{A} \Delta C/1/, \quad \Delta C/3/ = \left(\frac{B}{A}\right)^2 \Delta C/1/, \quad \Delta C/4/ = \left(\frac{B}{A}\right)^3 \Delta C/1/, \ldots$$

Aby system był stabilny potrzeba, aby kolejne odchylenia $\Delta C/n/$ malały do zera, a do tego warunkiem koniecznym jest, aby transmitancja $\frac{B}{A}$ była na moduł mniejsza od jedności.

$$\left|\frac{B}{A}\right| < 1$$

Jeśli transmitancja ta jest ujemna, czyli A lub B są ujemne, to mamy sprzężenie ujemne i zanikanie odchylenia ma charakter oscylacyjny; jeśli sprzężenie jest dodatnie, zanikanie ma charakter aperiodyczny, czyli "gładki". Jeśli warunek stabilności nie jest spełniony, to odchylenia nie zanikają /na granicy stabilności/, lub potęgują się /niestabilność/. Omówione przebiegi ilustruje rysunek 39. Jeżeli przykład ten sprawił ci wiele trudności, musisz koniecznie zapoznać się raz jeszcze gruntownie z punktami 1 do 10 niniejszego rozdziału, a także, jeśli uznasz za potrzebne, z rozwiązaniami zadań testu 3 z rozdziału I podr.2. Dalszych wyjaśnień możesz szukać w książce O.Lange



Rys. 39 Przebiegi zmian ceny rynkowej

"Wstęp do cybernetyki ekonomicznej".

4. Stabilnośc w przestrzeni stanu. Warunek stabilności systemu wielowymiarowego

1. Dotychczas badaliśmy stabilność systemów ze sprzężeniem zwrotnym, rozważając wartości wyjściowe tych systemów w kolejnych chwilach czasowych. Jest to wygodne, o ile mamy system złożony z układów jednowymiarowych, czyli mających jedną wielkość wyjściową i jedną wejściową. Jeśli jednak układy wchodzące w skład systemu są wielowymiarowe, to stabilność trzeba określić w odniesieniu do zachowania wszystkich wielkości wyjściowych i wygodnym staje się odwołanie do poznanej w rozdziale I podr.3 przestrzeni stanów.

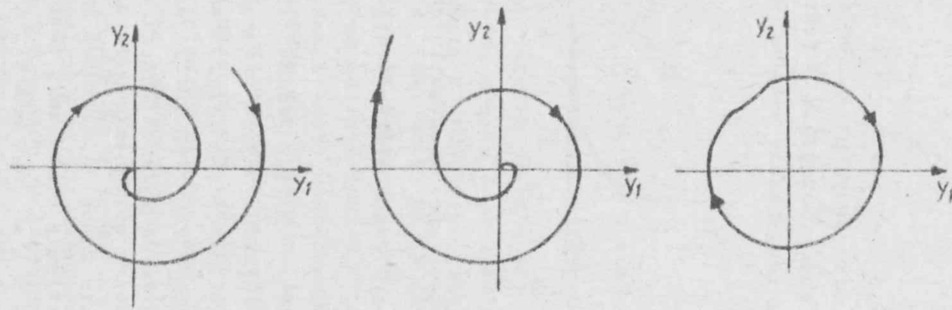
2. Powiemy zatem, że system wielowymiarowy jest stabilny, jeśli trajektoria z upływem czasu zbliża się do zera /do początku układu współrzędnych przestrzeni stanów/, jest natomiast na granicy stabilności, jeśli trajektoria pozostaje w stałej, średnio biorąc, odległości od zera, oraz jest niestabilny, jeśli trajektoria oddala się od zera.

TEST 1

P. Na rysunku 40, 40b i 40c przedstawiono trajektorie układu, będącego jednym z dwu układów systemu ze sprzężeniem zwrotnym. Układ jest dwuwymiarowy /posiada dwie wielkości wyjściowe X_1 i Y_2/, zatem przestrzeń stanów tego układu jest płaszczyzną.
Określić, w którym z pokazanych przypadków /a, b i c/ system jest: stabilny, na granicy stabilności i niestabilny?

xxx

O. Rysunek 40aprzedstawia system stabilny, /40b/ niestabilny, a /40c/ przedstawia system na granicy stabilności.
Jeśli udzieliłeś błędnej odpowiedzi powinieneś jeszcze raz przeczytać punkt 2 i punkty 1,2,3,4,5 i 6 rozdziału I, podr.3.

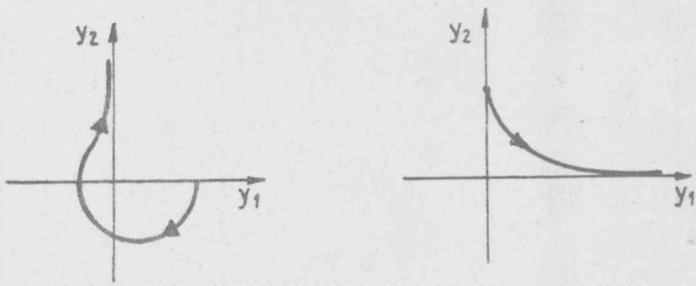


Rys. 40 Stabilność systemu przestrzeni stanu

P. Czy do tego, aby system wielowymiarowy był stabilby, wystarcza, aby jedna lub kilka z jego wielkości wyjściowych malało z czasem do zera?

xxx

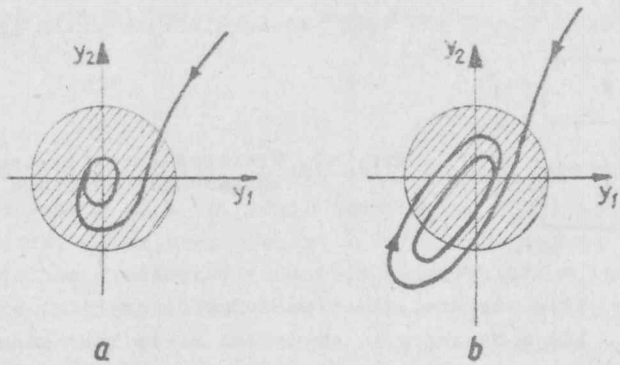
O. Nie, niezbędne jest zmniejszanie się w s z y s t k i c h wielkości wyjściowych, gdyż w przeciwnym przypadku system oddala się od początku układu współrzędnych w przestrzeni stanów. Rysunek 41 przedstawia trajektorie systemu dwuwy-



Rys. 41. Trajektoria systemów niestabilnych

miarowego niestabilnego, którego jedna wielkość wyjściowa rośnie, a druga maleje do zera.

3. Wymagania stawiane systemowi, o którym - w myśl naszej definicji powiemy, że jest stabilny, są bardzo ostre. W praktyce zadawala nas, jeśli system po odchyleniu od zera wróci w jego pobliże i już się od niego nie oddali, natomiast samo osiągnięcie zera może nie nastąpić nigdy, ponieważ system jest stale pod działaniem zakłóceń, które utrudniają osiągnięcie trwałej równowagi. Wymaganie takie można zilustrować w przestrzeni stanów przy pomocy zakreślenia wokół zera pewnego obszaru, który nazwiemy O B S Z A R E M P R A K T Y C Z N E J S T A B I L N O S C I, a który reprezentować będzie nasze wymagania odnośnie odległości pozostawania trajektorii. Rysunek 42a będzie więc przedstawiał system stabilny /praktycznie/, a 42b niestabilny. Na obydwu rysunkach obszar prak-



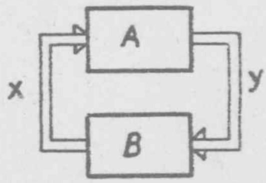
Rys. 42. Stabilność techniczna

tycznej stabilności zakreskowano.

4. Z określenia stabilności systemu w przestrzeni stanu wynika pewien sposób sprawdzania stabilności systemu. Tworzymy mianowicie w przestrzeni stanów pewną powierzchnię zamkniętą, np. taką kulę, aby w jej wnętrzu mieścił się zarówno początek układu współrzędnych, jak i aktualnie przebywany przez system odcinek trajektorii. Jeśli w miarę upływu czasu możemy budować takie powierzchnie o coraz mniejszych rozmiarach /kule "kurczą się"/, to oczywiście trajektoria zbliża się do początku układu, czyli system jest stabilny. Metoda ta nosi nazwę drugiej metody Lapunowa /od nazwiska rosyjskiego matematyka, który jest jej twórcą/ i jest szeroko stosowana do badania stabilności złożonych systemów.

5. My jednak ograniczymy swoje rozważania jedynie do liniowych systemów wielowymiarowych i sformułujemy kryterium stabliności w sposób analogiczny do sposobu określonego w poprzednim rozdziale dla systemów jednowymiarowych.

Jeżeli system nasz składa się z układu pierwszego o n wejściach /wektor n wymiarowy X/ i m wyjściach /wektor m wymiarowy Y/ i posiada transmitancję macierzową $[A]$, zaś układ drugi ma odpowiednio m wejść /Y/ i n wyjść /X/ oraz transmitancję macierzową B /rys. 43/, to zmiany w kolejnych momentach czasu wielkości X i Y będą zdeterminowane przez iloczyn



Rys. 43. Wielowymiarowy system ze sprzężeniem zwrotnym

transmitancji macierzowych B A, będący oczywiście macierzą o wymiarach: tyle wierszy, ile w macierzy B, czyli n, oraz tyle kolumn, ile w macierzy A, czyli też n. Opiszmy zachowanie naszego systemu znaną nam już tabelką, pamiętając jednak, że wielkości w niej występujące są odpowiednio wektorami i macierzami.

chwila czasowa	wektor X	wektor Y
0	X_o	AX_o
1	BAX_o	$ABAX_o$
2	$BABAX_o$	$ABABAX_o$
3	$BABABAX_o$	$ABABABAX_o$

Ze względu na nieprzemienność mnożenia macierzy, nie można zależności tych zapisać jednym ogólnym wzorem. Jeżeli jednak macierze A i B są przemienne, czyli AB = BA, do czego koniecznym warunkiem jest jednakowa ilość wierszy i kolumn w obu macierzach /n = m/, wówczas otrzymujemy znany już wzór:

$$X/n/ = A^n B^n X_o \qquad /26/$$

Interpretacja tego wzoru jest jednak daleko trudniejsza niż poprzednich. W szczególności orzeczenie, czy system jest z dodatnim, czy ujemnym sprzężeniem nie jest zupełnie banalne - pewne wejścia mogą wpływać na pewne wyjścia w kierunku dodatnim, a na inne w ujemnym, i to zarówno w macierzy A, jak i w macierzy B. W ten sposób prześledzenie, które wejście jest zwrotnie sprzężone dodatnio, a które ujemnie, jest zadaniem niemal niewykonalnym.

Przykład 1

Niech transmitancje A i B będą dane następującymi wzorami:

$$A = \begin{bmatrix} -1 & 2 \\ 1 & 0 \\ 3 & -1 \end{bmatrix} \qquad B = \begin{bmatrix} 4 & -2 & 5 \\ 1 & 0 & -1 \end{bmatrix}$$

Widać, że wejście X_1 wpływa słabo ujemnie na wyjście Y_1, słabo dodatnio na Y_2 i silnie dodatnio na Y_3. Z kolei możemy stwierdzić, że poprzez macierz B wejście Y_1 wpływa silnie dodatnio na X_1, wejście Y_2 silnie ujemnie i wejście Y_3 bardzo silnie dodatnio. Niełatwo jest ostatecznie odpowiedzieć na pytanie, czy wielkość wejściowa X_1 jest sprzężona zwrotnie dodatnio, czy ujemnie. Ponieważ przykład rozpatrywany jest prosty /niewielkie wymiary macierzy A i B/ można przeprowadzić dokładną analizę:

Niech $X_1 = 1$. Wówczas wartości wyjść bloku A wynoszą:
$Y_1 = -1$, $Y_2 = 1$, $Y_3 = 3$. Nowa wartość X_1 wyniesie więc
$X_1 = 4/-1/ + /-2/\ 1 + 5\ 3 = +9$, czyli wynikałoby, że X_1 jest dodatnie. Analiza nasza jest jednak niepełna, nie uwzględniamy mianowicie faktu, że wyjścia Y_1, Y_2 i Y_3 zależą również od wielkości wejściowej X_2. Nawet jeśli na początku ta wielkość wejściowa była zerowa, to już po jednokrotnym obejściu pętli sprzężenia zwrotnego będzie miała wartości niezerowe, a mianowicie:
$X_2 = 1/-1/ + 0\ 1 + /-1/\ 3 = -4$. Przy następnym obchodzeniu pętli, celem ustalenia następnej wartości wielkości wejściowej X_1 wpływ wielkości X_2 ujawni się i może nawet zmienić charakter oddziaływania zwrotnego z dodatniego na ujemny. Analizy tej jednak dalej prowadzić nie będziemy, ponieważ celem przykładu było nie zanalizowanie konkretnego przypadku, a wskazanie na trudności takiej analizy.

Rozpatrywany system należał do prostszych, wektory były zaledwie dwu i trzywymiarowe. Można sobie więc wyobrazić, na jakie trudności napotyka próba analizy wysoko wielowymiarowych systemów, w dodatku np. nieliniowych lub opisywanych równaniami różniczkowymi.

6. Zakończenie punktu 5 przyniosło wzór /26/, pozwalający dyskutować zagadnienie stabilności wielowymiarowych systemów w podobny sposób, jak systemów jednowymiarowych. Otóż na to, aby systęm był stabilny koniecznym jest, aby kolejne wartości X/n/ malały ze wzrostem n, a dokładniej, aby malały wszystkie składowe wektora X/n/. Jednak mnożniki, przez które mnożony jest w kolejnych chwilach wektor początkowy X_o, są macierzami i nie można kryterium stabilności określić w sposób tak prosty i bezpośredni, jak w przypadku układów jednowymiarowych.

7. Aby trudność tę obejść odwołamy się do rozwiązania tzw. zagadnienia własnego macierzy. Aby nie zaciemniać wykładu trudnościami czysto rachunkowej natury, uprościmy jeszcze nieco wzór /1/, przyjmując, że układ B opisywany jest macierzą jednostkową I, która posiada wyłącznie jedynki na głównej przekątnej i zera we wszystkich innych miejscach:

$$I = \begin{bmatrix} 1 & 0 \ldots & 0 \\ 0 & 1 \ldots & 0 \\ \ldots & \ldots & \ldots \\ 0 & 0 \ldots & 1 \end{bmatrix}$$

Macierz I spełnia w rachunku macierzowym tę samą rolę liczba 1 w konwencjonalnej arytmetyce. W szczególności mnożenie dowolnej macierzy przez macierz I lub jej dowolną potęgę nie zmienia tej macierzy:

$$A\ I = I\ A = I^n\ A = A\ I^m = A \qquad /27/$$

Na podstawie /27/ można zależność /26/ przepisać w postaci:

$$X/n/ = A^n\ I^n X_o = A^n X_o \qquad /28/$$

8. Przypomnimy teraz kilka faktów z teorii rachunku macierzowego. Dla każdej macierzy A istnieje taki wektor X i taka wartość λ, że zamiast mnożenia macierzy przez wektor wystarczy pomnożyć wektor przez liczbę :

$$A\ X = \lambda\ X \qquad /29/$$

Znajdowanie liczb spełniających warunek /29/ nazywane jest rozwiązywaniem zagadnienia własnego macierzy A, a znalezione wartości nazywane są wartościami własnymi macierzy. Aby rozwiązać zagadnienie własne macierzy, przekształcimy równanie /29/ do postaci :

$$/A - \lambda\ I/\ X = 0 \qquad /30/$$

Zależność /30/ przedstawia układ równań liniowych jednorodnych ze względu na zmienną X. Układ ten ma nietrywialne /niezerowe/ rozwiązania tylko wtedy, jeśli wyznacznik macierzy $/A - \lambda\ I/$ jest zerowy. Z tego właśnie warunku obliczymy wartości własne macierzy.

$$\det\ /A - \lambda\ I/ = 0 \qquad /31/$$

Zależność /31/ jest równanien algebraicznym stopnia równego stopniowi macierzy A /pojęcie stopnia macierzy można przypomnieć sobie np. z książki W.Smirnow "Matematyka wyższa" tom 3. część 1./

Równanie to ma więc dokładnie tyle rozwiązań, ile wynosi stopień macierzy A. Na podstawie równań /28/ i /29/ możemy zapisać:

$$X/n/ = A^n\ X_o = \lambda^n\ X_o \qquad /32/$$

Wartości X/n/ obliczone z równania /32/ będą malały, jeśli wartości własne λ będą mniejsze od jedności. Jest to właśnie poszukiwane przez nas kryterium stabilności:
system wielowymiarowy liniowy jest stabilny, jeśli wszystkie wartości własne jego transmitancji są mniejsze od jedności.

TEST 2

P. Niech w systemie jak na rysunku 43 transmitancje A i B wynoszą:

$$A = \begin{bmatrix} -1 & 2 \\ 0 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 4 \\ -2 & 0 \end{bmatrix}$$

Stwierdzić, czy do określania przebiegów wielkości

X i Y można stosować wzór /26/, oraz obliczyć te przebiegi dla $X_{01} = 1$ i $X_{02} = 1$.

xxx

0. Sprawdzamy przemienność mnożenia macierzy A i B :

$$A\,B = \begin{bmatrix} -5 & -4 \\ -2 & 0 \end{bmatrix}, \qquad B\,A = \begin{bmatrix} -1 & 6 \\ 2 & -4 \end{bmatrix}$$

Ponieważ AB $\neq$ B A, więc macierze są nieprzemienne i wzoru /26/ stosować nie wolno. Jeśli twój wniosek był inny, przed czytaniem dlaszego ciągu rozwiązania przeczytaj jeszcze raz punkt 5.

Celem obliczenia przebiegów zbudujemy podobną jak w punkcie 5. tabelkę wartości w poszczególnych momentach czasu:

chwila czasowa	wektor X	wektor Y
0	$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$
1	$\begin{bmatrix} 5 \\ -2 \end{bmatrix}$	$\begin{bmatrix} -9 \\ -2 \end{bmatrix}$
2	$\begin{bmatrix} -17 \\ 18 \end{bmatrix}$	$\begin{bmatrix} 53 \\ 18 \end{bmatrix}$
3	$\begin{bmatrix} 125 \\ -106 \end{bmatrix}$	$\begin{bmatrix} -337 \\ -106 \end{bmatrix}$
	$\begin{bmatrix} -761 \\ 674 \end{bmatrix}$	$\begin{bmatrix} 2109 \\ 674 \end{bmatrix}$

Jeżeli sam nie wykonałeś tego zadania, to przynajmniej sprawdź na podstawie tabeli z punktu 5 wszystkie przeliczenia, które doprowadziły do podanych w rozwiązaniu wyników.

P. Czy system jest stabilny?

xxx

O. Wartości obydwu współrzędnych wektora X narastają z upływem czasu, zatem układ jest niestabilny.

P. Narysować zachowanie systemu w przestrzeni stanów X_1, X_2

i w przestrzeni Y_1, Y_2.

xxx

O. Poprawne trajektorie przedstawiają rysunki 44a i 44b.

P. Czy system jest z dodatnim, czy z ujemnym sprzężeniem zwrotnym?

xxx

O. O systemach wielowymiarowych z reguły nie wypowiada się zdań w rodzaju "system z dodatnim sprzężeniem" lub "system z ujemnym sprzężeniem", ale zachowanie naszego systemu /obecność narastających, znakozmiennych oscylacji/ upodabnia go do niestabilnego systemu z ujemnym sprzężeniem zwrotnym.

P. Czy omawiany system jest stabilny praktycznie?

xxx

O. Nie, ponieważ nie potrafimy wskazać żadnego obszaru w pobliżu początku układu przestrzeni stanu, w którym trajektoria pozostawałaby dostatecznie długo /w razie wątpliwości sprawdź określenie w punkcie 3/.

P. Czy gdyby macierz B zastąpiono macierzą I system stałby się stabilnym?

xxx

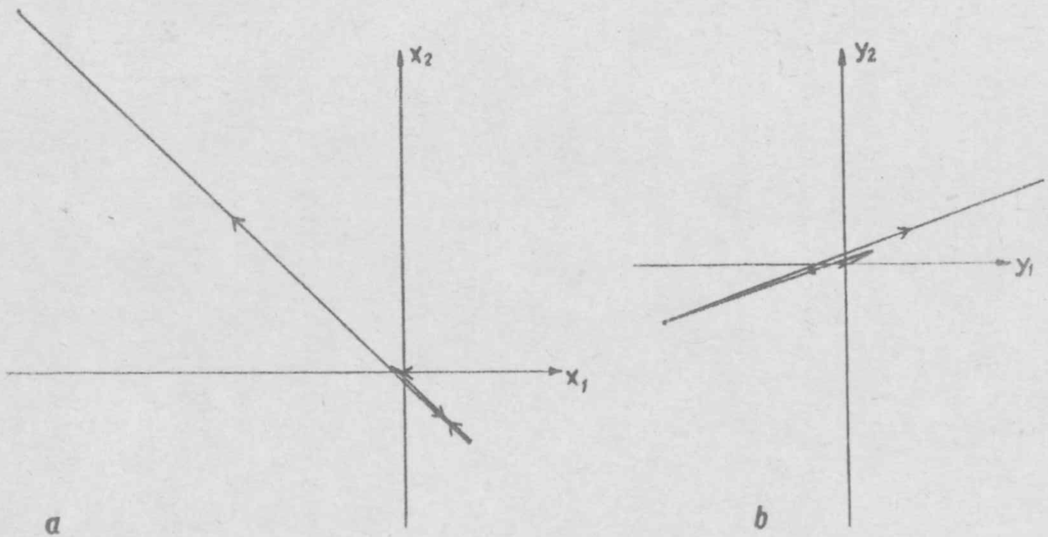
O. Aby to rozstrzygnąć, obliczamy ze wzoru /31/ wartości własne macierzy A:

$$\begin{bmatrix} -1-\lambda & 2 \\ 0 & 1-\lambda \end{bmatrix} = -1 /1+\lambda/ /1-\lambda/ = 0$$

Wartości własne wynoszą +1 i -1, układ jest więc na granicy stabilności.

P. Czy gdyby macierz A zastąpiono macierzą I system stałby się stabilnym?

xxx



Rys. 44. Zachowanie systemu /do testu 2/

0. Identycznie jak poprzednio obliczamy wartości własne:

$$\begin{bmatrix} 1 & -\lambda & 4 \\ -2 & & -\lambda \end{bmatrix} = \lambda^2 - \lambda + 8 = 0$$

Równanie to ma dwa pierwiastki zespolone sprzężone

$$\lambda_1 = \frac{1}{2} + i\,\frac{\sqrt{31}}{2} \qquad \lambda_2 = \frac{1}{2} - i\,\frac{\sqrt{31}}{2}$$

W celu orzeczenia o stabilności musimy zbadać moduł tych wartości własnych:

$$|\lambda_1| = |\lambda_2| = \frac{1}{2}\sqrt{1 + 31} = \frac{1}{2}\sqrt{32} = 2\sqrt{2} > 1$$

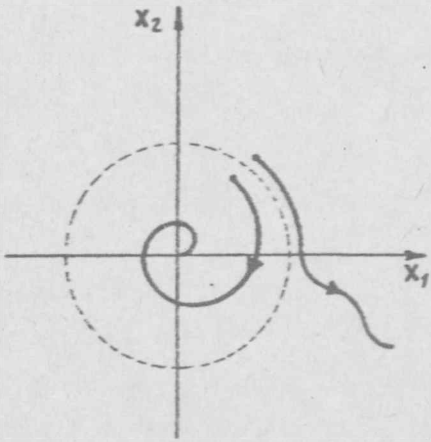
Jak widać system będzie niestabilny.

W przypadku, gdyby dwa ostatnie zadania sprawiły ci trudności, powinieneś jeszcze raz przeczytać punkt 8 i ewentualnie skorzystać ze wskazanej tam literatury.

9. Wszystkie dotychczasowe rozważania prowadziliśmy dla systemów liniowych, to znaczy takich, których układy dawały się opisywać za pomocą transmitancji. W systemach nieliniowych pojawia się pewne istotne novum w stosunku do systemów liniowych: stabilność systemu nieliniowego może zależeć od wielkości sygnałów w nim występujących.

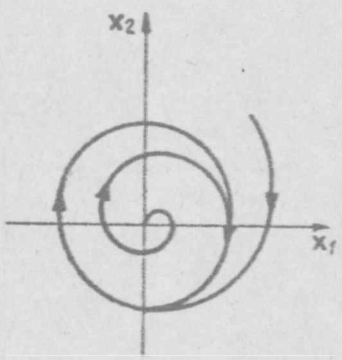
Jest to zrozumiałe, jeśli przypomnimy sobie, w jaki sposób metodą przyrostową wprowadziliśmy dla układów nieliniowych pojęcie transmitancji. Wartość uzyskiwanych transmitancji uzależniona była zwykle od wartości odpowiadającej punktowi spoczynkowemu /patrz s. /, gdzie transmitancja wynosiła $2\,Q_0$/. Ponieważ równocześnie warunek stabilności uzależniony jest od wartości transmitancji, często może się zdarzać, że układ stabilny dla małych sygnałów staje się niestabilny, gdy sygnały przekroczą pewien poziom, lub odwrotnie - wzrost amplitud może układ stabilizować.

10. Omawiane właściwości systemów nieliniowych również wygodnie jest zilustrować za pomocą trajektorii w przestrzeni stanów. Rysunek 43a przedstawia typową sytuację - system



Rys. 43a . System warunkowo stabilny

jest stabilny, o ile nie oddali się od początku układu poza przerywaną linię /trajektoria 1 powraca do początku układu/. Przy zbyt wielkich odchyleniach system staje się niestabilny i zupełnie traci równowagę /trajektoria 2/. Możliwe są i bardziej złożone kombinacje - np. system jest niestabilny w pobliżu początku układu /wzbudza się/, ma obszar stabilnych oscylacji i jest stabilny poza obszarem oscylacji /rys.45/.

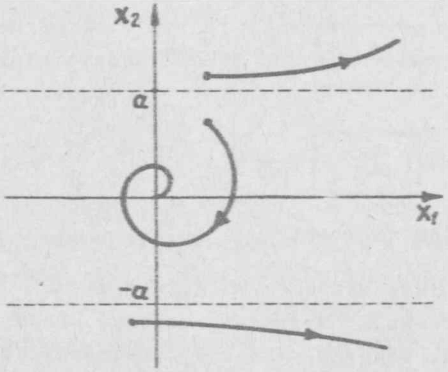


Rys.45. System generujący drgania

System taki będzie generował drgania o parametrach określonych zamkniętą trajektorią na rysunku.

TEST 3

P. Określić zachowanie i stabilność systemu, którego przestrzeń stanu przedstawia rysunek 46.



Rys. 46. System warunkowo stabilny

xxx

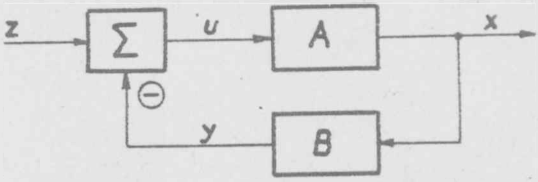
O. Dla X_2 nie przekraczającego na moduł wartości a system jest stabilny i po wytrąceniu z równowagi powraca do zera kilkoma malejącymi oscylacjami. Natomiast jeśli X_2 przkroczy a, system staje się niestabilny.

Jeśli nie odpowiedziałeś, to niezawodny znak, że czas na powtórkę!

5. Teoria regulacji

1. Rozpatrzmy teraz system ze sprzężeniem zwrotnym, na który będziemy mogliz zewnątrz oddziaływać, czyli system z wejściem. Zbudujemy go w sposób następujący:
sygnał z zewnątrz systemu Z, który nazywać będziemy WYMUSZENIEM, jest sumowany z sygnałem Y, przy czym sygnał Y może do sygnału Z być dodawany lub odejmowany, co oznaczać będziemy zgodnie z konwencją przyjętą na rysunku 29. Sumę sygnałów

nazwiemy odchyleniem i oznaczymy przez U. Odchylenie U jest sygnałem wejściowym bloku o transmitancji A, którego wyjście, będące równocześnie wyjściem systemu, oznaczymy przez X i nazwiemy W I E L K O Ś C I Ą R E G U L O W A N Ą. Sygnał X podawany jest również na układ o transmitancji B, wytwarzający sygnał Y. Strukturę systemu przedstawia rysunek 47.



Rys. 47. System regulacji

2. Rozpatrywany system jest systemem ze sprzężeniem zwrotnym. Aby się o tym przekonać wystarczy założyć, że wymuszenie Z jest równe zeru. Wówczas sygnał U jest równy +Y lub -Y i mamy klasyczny, znany z poprzednich rozdziałów, system ze sprzężeniem zwrotnym dodatnim, o ile sumator dodaje Z do Y celem uzyskania U, i ujemnym, o ile sumator odejmuje Y od Z.

3. Ponieważ system ten posiada określoną wielkość wyjściową /X/ i określoną wielkość wejściową Z, sensowne jest mówienie o jego transmitancji. Wyprowadzimy wzór na transmitancję takiego systemu:

Z określenia sumatora mamy

$$U = Z \pm Y,$$

przy czym + odnosi się do systemu z dodatnim, a - do systemu z ujemnym sprzężeniem zwrotnym.

Z kolei

$$X = A\,U \quad \text{i} \quad Y = B\,X \qquad /AB \neq 1/ \quad /AB \neq -1/$$

Dokonując odpowiednich podstawień mamy:

$$X = A\,/Z \pm Y/ = A\,/Z \pm B\,X/$$

$$X\,/1 \mp AB/ = AZ$$

i ostatecznie

$$X = \frac{A}{1 \mp AB}\,Z \qquad /33/$$

Należy zwrócić uwagę na zmianę kolejności znaków + i - . Obecnie - w mianowniku odpowiada systemowi z dodatnim sprzężeniem zwrotnym a, + systemowi z ujemnym sprzężeniem. Ostatecznie transmitancja naszego systemu ze sprzężeniem zwrotnym ma postać:

$$G = \frac{A}{1 \mp AB} \qquad /34/$$

4. Wzory /33/ i /34/ są bardzo ważne ze względu na szerokie zastosowanie, jakie mają systemy typu przedstawionego na rysunku 47. Systemy takie nazwiemy S Y S T E M A M I R E G U L A C J I [9/] ze względu na to, że najobszerniejsze zastosowanie i najszersze opracowanie naukowe systemy te znalazły w dziale automatyki zwanym teorią regulacji.
Rozpatrzmy działanie systemu regulacji w przypadku kiedy B = 1. Wówczas w systemie z ujemnym sprzężeniem zwrotnym U jest równe różnicy pomiędzy wielkością wymuszenia Z a wielkością wyjściową X.
Jeżeli więc transmitancja A będzie dodatnia, system taki będzie dążył do uzgodnienia swej wielkości wyjściowej X z wymuszeniem Z, gdyż X mniejszym od Z zaistnieje dodatnie U, które przez transmitancję A wpłynie na wzrost X, w przeciwnym przypadku ujemne U obniży X.

5. System tak zorganizowany ma możliwości utrzymywać wielkość X równą każdorazowo wielkości Z pomimo istnienia dowolnych zakłóceń. Zagadnienie panowania nad systemem, przy obecności zakłóceń, jest znane w biologii pod nazwą homeostazy,

---

[9/] Systemy regulacji są szczególnym przypadkiem systemów sterowania. Obejmują one układy zamknięte, najczęściej z ujemnym sprzężeniem zwrotnym.

a w naukach technicznych i ekonomii pod nazwą sterowalności /kierowalności/. Jest to problem ważny i trudny do rozwiązania na drodze innej, niż wskazany system regulacji. Rozważmy bowiem inne możliwości:

Można usiłować ograniczyć dostęp zakłóceń do systemu. Przykładem tego typu rozwiązania jest skorupa żółwia. Jest to jednak rozwiązanie nieeleganckie, kosztowne, i co gorsza, najczęściej nieskuteczne - zakłócenia często są niezbędną częścią funkcjonowania systemu.

Można usiłować mierzyć lub w inny sposób oceniać aktualną wielkość zakłóceń i starać się im na bieżąco przeciwdziałać. Taka metoda "kompensacyjna" ma wielu zwolenników, ale również nie jest wolna od wad: aby zakłócenia mierzyć, trzeba je znać lub przynajmniej domyślać się jakie są, tymczasem większości zakłóceń w rzeczywistych systemach zwykle nie znamy lub nie znamy ich wpływu na system.

Pozostaje ostatnia możliwość - likwidować nie przyczyny, a skutki. Nie analizując co było powodem odchylenia, zwalczamy je, stosując oddziaływanie proporcjonalne do wielkości odchylenia. Metoda ta, to nic innego, jak tylko opis działania naszego systemu regulacji.

6. Aby być w pełni obiektywnym należy wskazać również i na słabą stronę regulacyjnej koncepcji zwalczania zakłóceń - otóż aby system regulacji zaczął interweniować, musi zaistnieć odchylenie od wymaganego przebiegu, ponieważ właśnie to odchylenie użyte jest do sterowania systemem.

Dla złagodzenia tego mankamentu stosowany jest układ o transmitancji B. Jeśli B jest duże, to nawet niewielka odchyłka spowoduje silne oddziaływanie i szybki powrót do warunków poprawnej pracy. Układ B może być zresztą bardziej złożony, może on np. obliczać pochodną sygnału wyjściowego, czyli jeśli tylko sygnał zacznie wzrastać /dodatnia pochodna/, to już pojawi się oddziaływanie itp. Przyjęło się w technicznych zastosowaniach nazywać układ A obiektem, a układ B regulatorem.

7. Transmitancji regulatora nie można jednak zwiększać nieograniczenie, gdyż w systemie naszym obowiązuje ten sam warunek stabilności, co w systemie bez wejść:

$$|A\ B| > 1 \qquad /35/$$

Przykład 1

E. Systemem regulacji byłby w ekonomii fundusz wyrównawczy, stabilizujący dochody rolników niezależnie od warunków losowych typu: nieurodzaj, klęski żywiołowe itp. Wielkość Z byłaby wówczas przyjętą średnią rocznych dochodów, a wielkość X rzeczywistymi dochodami. Gdyby dochody realne w roku szczególnie urodzajnym przekroczyły założoną średnią, nadwyżka ich /U/ byłaby akumulowana, natomiast w latach nieurodzaju ten fundusz nadwyżek mógłby być podstawą subwencjonowania. System taki stabilizowałby dochody niezależnie od przyczyn wpływających na ich wahanie. Na podobnej zasadzie funkcjonują wszelkie instytucje ubezpieczeniowe.

T. Podczas długotrwałego lotu funkcję pilota w dużych samolotach komunikacyjnych przejmuje automat, który sprawuje kontrolę nad samolotem, porównując kierunek i wysokość lotu - podane mu przez pilota /Z/ - z aktualnym położeniem i kierunkiem lotu /X/. W razie odchylenia /U/ stanowi ona podstawę do interwencji automatu, który organami sterowania dąży do zmniejszenia tej odchyłki do zera. Automat taki zdolny jest zlikwidować zaburzenia lotu niezależnie od powodów ich zaistnienia.

B. Jeśli wyciągam rękę, aby pochwycić jakiś przedmiot, to ruch ten jest pod kontrolą wzroku. Wielkość oddziaływania układu nerwowego na odpowiednie mięśnie /U/ jest proporcjonalna do obserwowanej różnicy pomiędzy położeniem przemiotu do którego dążę, /Z/, a aktualnym położeniem ręki /X/, zaobserwowanym przez oczy /Y/.

8. Wzór /1/ pozwala w systemie regulacji określić ze znanej wielkości wejściowej /Z/ wielkość wyjściową /X/. Pamię-

tajmy przy tym, że zarówno Z, jak i X są najczęściej funkcjami czasu.

Przekształcimy teraz ten wzór tak, aby uzyskać zależność odwrotną wiążącą znaną wartość X z nieznaną Z. Wzór taki określi nam, jakie wymuszenie musimy stosować wobec systemu, jeśli chcemy osiągnąć określony przebieg wielkości wyjściowej.

$$Z = /\frac{1}{A} \pm B/\ X \qquad /36/$$

Możemy również postawić pytanie, jaki powinien być regulator /transmitancja B/, aby przy określonym A i określonym Z zapewnić wymagany przebieg X:

$$B = \mp\ /\frac{Z}{X} - \frac{1}{A} \qquad /37/$$

Wzory /35/, /36/ i /37/ będą używane przy rozważaniu konkretnych systemów ze sprzężeniem zwrotnym, które rozpatrywać będziemy w teście.

TEST 1

P. Wykazać, że w systemie regulacji wielkość wyjściowa jest uzależniona od siebie samej /jest sprzężona zwrotnie/.

xxx

O. Od wielkości X zależy wielkość Y, ta zaś z kolei ma wpływ na wielkość U, decydującą o wielkości X.

Przy braku pozytywnej odpowiedzi powinieneś jeszcze raz przeczytać punkty od 1 do 4.

P. Czy system regulacji wykorzystuje przy usuwaniu zakłóceń informację o ich charakterze i wielkości?

xxx

O. Nie. Jedyną wykorzystywaną informacją jest stwierdzenie faktu zaistnienia odchylenia, natomiast jego przyczyny są obojętne.

Przy nieprawidłowej odpowiedzi należy przypomnieć sobie wiadomości z punktów 5,6 i 7.

P. Co należy wybrać, znając zakłócenia wpływające na obiekt: system regulacji czy pomiar i kompensację zakłóceń?

xxx

O. Pomiar i kompensację, ponieważ system regulacji z zasady swego działania musi dopuszczać do pojawienia się odchyłek /gdyż kompensuje je dopiero do ich zaistnieniu/, natomiast system z pomiarem zakłóceń, o ile jest prawidłowo wykonany, będzie interweniował już po zaobserwowaniu przyczyny zła, nie czekając na jej efekty.

W razie innej odpowiedzi koniecznie przemyśl jeszcze raz punkty 5 i 6.

P. Wykazać, że model gospodarki narodowej - proponowany przez Keynesa - jest systemem ze sprzężeniem zwrotnym, i to systemem regulacji. Przyjmiemy, że model ten zakłada, iż dochód narodowy netto X jest sumą wydatków przeznaczonych na inwestycje Z i wydatków przeznaczonych na konsumpcję Y, przy czym zakłada się, że

$$Y = B\,X \qquad 0 \quad B \quad 1$$

gdzie B jest stopą konsumpcji.
Określić transmitancję tego systemu i ocenić jego stabilność.

xxx

O. Z warunków modelu łatwo ustalimy, że:

$$X = Z + Y = Z + B\,X$$

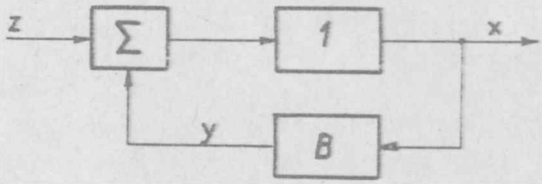
czyli

$$X = \frac{1}{1 - B}\,Z$$

Porównując ten wzór ze wzorem /35/ dochodzimy do wniosku, że mamy do czynienia z systemem regulacji o dodatnim sprzężeniu zwrotnym i transmitancji obiektu A = 1. Transmitancja systemu wynosi

$$G = \frac{1}{1 - B}$$

System jest stabilny, jeśli $A\ B > 1$, co w naszym przypadku zachodzi. System przedstawia rysunek 48.



Rys. 48. Model Keynesa

Jeśli napotkałeś przy rozważaniu tego zadania na trudności, powinieneś zapoznać się z książką O.Lange "Wstęp do cybernetyki ekonomicznej".

P. Postawimy teraz systemowi regulacji opisującemu model Keynesa zadanie określenia potrzebnej wielkości nakładów inwestycyjnych Z, zapewniających zadaną wielkość produkcji czystej P.

xxx

O. Musi oczywiście być spełniony warunek stabilności rynku:

$$X = P$$

/Suma pieniędzy wydatkowanych musi być równa wartości produkcji/.

Wykorzystując wzór /42/ mamy wówczas

$$Z = /1 - B/\ X = /1 - B/\ P$$

P. Powiedzmy, że w modelu Keynesa zadana jest wielkość nakładów inwestycyjnych Z i wartość docelowej produkcji czystej. Czy system pozwala na wyznaczenie stopy konsumpcji?

xxx

O. Stopa konsumpcji odgrywa w naszym modelu rolę regulatora, możemy więc wyznaczyć ją ze wzoru /43/ :

$$B = \frac{Z}{P} - 1 = \frac{Z - P}{P}$$

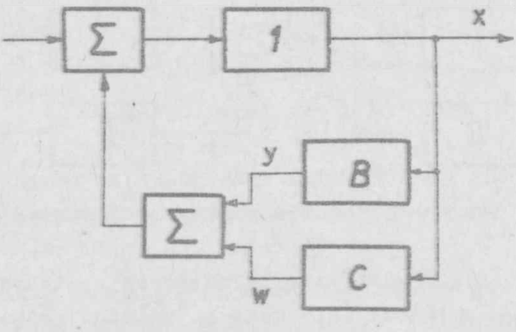
P. Zajmiemy się teraz rozbudowywaniem modeli Keynesa i związanych z nimi naszych systemów regulacji. Najpierw zrobimy założenia, że dochód narodowy netto X jest sumą inwestycji Z, konsumpcji Y i inwestycji wtórnych W, proporcjonalnych do wielkości dochodu narodowego X:

$$W = C\,X$$

Narysować schemat i podać jego transmitancję. Jaki jest obecnie warunek stabilności?

xxx

O. Strukturę systemu przedstawia rysunek 49. Na podstawie



Rys. 49. Model Keynesa z uwzględnieniem inwestycji i konsumpcji

wzoru /33/ i znanych z rozdziału II, podr.1 wzorów dla systemów równoległych możemy napisać:

$$X = \frac{1}{1 - /B + C/}\,Z$$

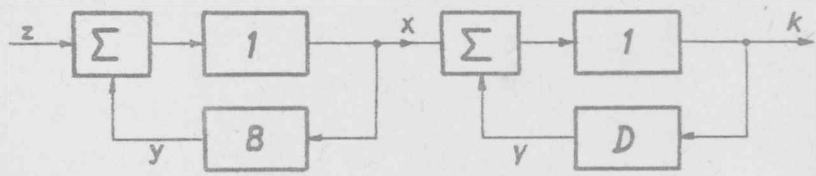
Warunek stabilności jest obecnie rozbudowany:

$$|B + C| < 1$$

P. Obecnie rozbudujemy schemat Keynesa w innym kierunku: niech, jak w pierwotnym modelu, na dochód narodowy składają się jedynie inwestycje Z i konsumpcja Y. Jednakże niech gospodarka, którą modelujemy, partycypuje w handlu międzynarodowym i uzyskuje na rynkach światowych kredyty proporcjonalne do całkowitego posiadanego kapitału K, ze współczynnikiem proporcjonalności D, oznaczającym stopę nakładów pracy żywej i będącym sumą dochodu narodowego X i kredytów V, przy czym V = D K. Narysować schemat systemu regulacji odpowiadającego temu modelowi i określić jego transmitancję.

xxx

O. Schemat systemu przedstawia rysunek 50. Jak widać system



Rys. 50. Model handlu międzynarodowego

jest złożony z dwu, szeregowo połączonych systemów regulacji. Systemy składowe opisywane są zależnościami:

$$X = \frac{1}{1 - B} Z, \qquad K = \frac{1}{1 - D} X,$$

a cały system opisuje zależność:

$$K = \frac{1}{1 - B} \, \frac{1}{1 - D} \, Z$$

P. System regulacji stanowi również model dla schematów reprodukcji Marksa. Najpierw rozważmy gospodarkę narodową jako całość. Produkt globalny netto X jest sumą nakładów pracy żywej /V + M/ i nakładów środków produkcji /pracy uprzedmiotowionej/ C.

Między ilością pracy uprzedmiotowionej, a produktem globalnym zachodzi zależność:

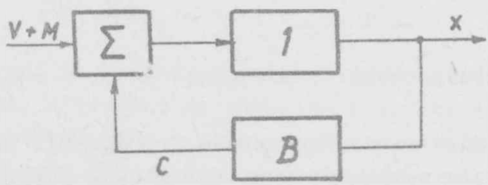
$$C = B X$$

B jest współczynnikiem nakładów środków produkcji.

Narysować schemat systemu regulacji odpowiadający temu modelowi.

xxx

O. Schemat przedstawia rysunek 51.



Rys. 51. Model reprodukcji prostej Marksa

Jeżeli nie wykonałeś tego zadania samodzielnie, jeszcze raz przeanalizuj rozwiązania wszystkich poprzednich zadań tego testu.

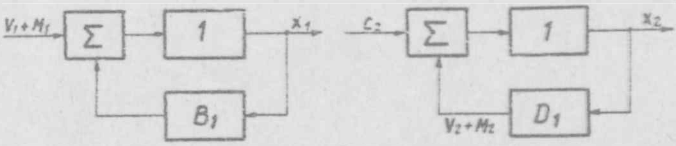
P. Obecnie rozważymy schemat obejmujący dwa działy gospodarki: dział produkcji środków produkcji i dział produkcji środków konsumpcji. W dziale produkcji środków produkcji należy przyjąć, że wielkość produktu globalnego X_1 jest, jak poprzednio, sumą nakładów pracy żywej $/V_1 + M_1/$ i nakładów środków produkcji $C_1 = B_1X_1$, natomiast w dziale drugim przyjmiemy, że X_2 jest sumą C_2 i $/V_2+M_2/ = D_2X_2$, gdzie D_2 oznacza współczynnik nakładów pracy żywej w dziale drugim.

Narysować schemat obu działów, a następnie połączyć je w schemat całości gospodarki, wykorzystując znany warunek reprodukcji prostej:

$$C_2 = V_1 + M_1$$

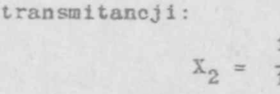
xxx

0. Schematy obu działów, zbudowane analogicznie jak w poprzednim zadaniu, przedstawia rysunek 52. Aby dokonać ich połą-



Rys. 52. Model reprodukcji prostej w dwu działach

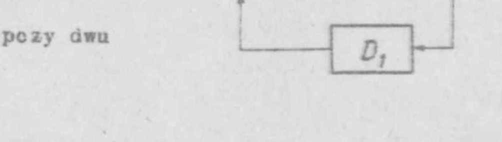
czenia posługując się warunkiem reprodukcji prostej, musimy jeden ze schematów odwrócić metodą wskazaną wzorem /36/. Jeśli zrobimy to ze schematem pierwszego działu, otrzymamy zeń system o transmitancji /$1 - B_1$/, wejściu X_1 i wyjściu $V_1 + M_1$ /rys. 53/



Rys. 53. Model zastępczy działu 1

Ponieważ z warunku reprodukcji wyjście tego wystemu jest równe wejściu drugiego, nie zmienionego systemu, możemy je więc połączyć szeregowo /rys.54/, uzyskując system o transmitancji:

$$X_2 = \frac{1 - B_1}{1 - D_1} X_1$$



Rys.54. Model zastępczy dwu działów

P. W identyczny sposób zbudujemy system regulacji odpowiadający modelowi gospodarki narodowej w warunkach reprodukcji rozszerzonej. W tym celu wartość produktu dodatkowego M w obu działach rozbijemy na sumę produktu przeznaczonego na zwiększenie zasobów środków produkcji M_c, produktu przeznaczonego na zwiększenie zatrudnienia M_v i części produktu konsumowanego nieprodukcyjnie M_o:

$$M = M_c + M_v + M_o$$

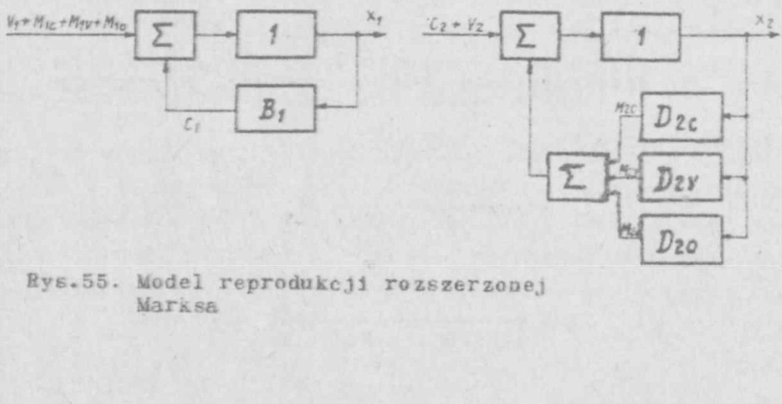
Warunek równowagi będzie obecnie bardziej rozbudowany, a mianowicie:

$$C_2 + M_{2c} = V_1 + M_{1v} + M_{1o}$$

Oznacza on, że zapotrzebowanie drugiego działu na środki produkcji, potrzebne na aktualne potrzeby i na rozbudowę, musi być równoważone przez zapotrzebowanie pierwszego działu na środki konsumpcji potrzebne na wypłaty robotnikom aktualnie pracującym, nowo zatrudnionym i na konsumpcję nieprodukcyjną. Stosując postępowanie identyczne jak w poprzednim zadaniu, należy określić transmitancje obu działów, a następnie transmitancję całości gospodarki.

xxx

O. Transmitancje obu systemów przedstawiających działy wynikają z rysunku 55.



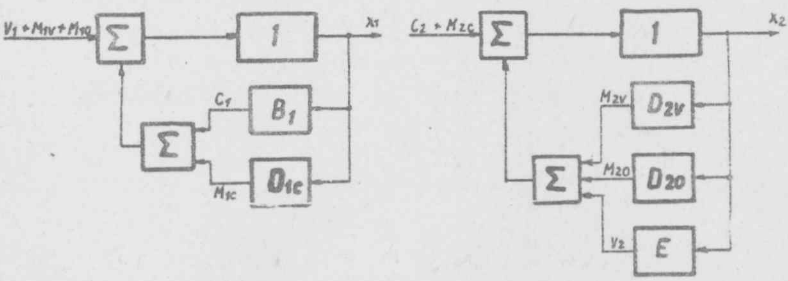
Rys.55. Model reprodukcji rozszerzonej Marksa

$$X_1 = \frac{1}{1 - B_1} /V_1 + M_{1c} + M_{1v} + M_{1o}$$

$$X_2 = \frac{1}{1 - /D_{2c} + D_{2v} + D_{2o}/} /C_2 + V_2/$$

Widać, że współczynnik D_2 rozbity został na składniki D_{2c}, D_{2v} i D_{2o}, których znaczenie wynika z rysunku 55.

W celu połączenia systemów musimy jednak dokonać pewnych przegrupowań, tak aby wejściem do pierwszego systemu było tylko $V_1 + M_{1v} + M_{1o}$, a do drugiego $C_2 + M_{2c}$. Strukturę systemów po przegrupowaniu przedstawia rysunek 56.



Rys. 56. Przekształcony model reprodukcji rozszerzonej

Transmitancje wynoszą teraz

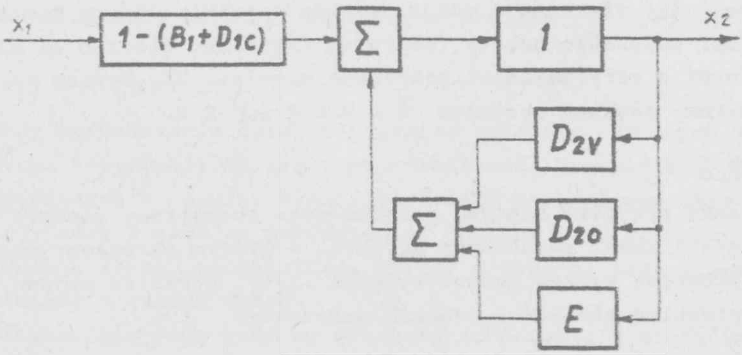
$$X_1 = \frac{1}{1 - /B_1 + D_{1c}/} /V_1 + M_{1v} + M_{1o}/,$$

$$X_2 = \frac{1}{1 - /D_{2v} + D_{2o} + E} /C_2 + M_{2c}/$$

E jest tu współczynnikiem wiążącym V_2 i X_2 /współczynnikiem akumulacji kapitału zmiennego/.

Stosując znowu wzór /36/ i wzór na transmitancję systemu szeregowego otrzymamy strukturę /rys. 57/ i transmitancję

$$X_2 = \frac{1 - /B_1 + D_{1c}/}{1 - /D_{2v} + D_{2o} + E/} X_1$$



Rys. 57. Model zastępczy reprodukcji rozszerzonej

9. Wzory /34/ i /36/ oraz /37/ mogą być również stosowane w przypadkach dynamicznych, kiedy zmienne Z, X i Y są funkcjami czasu, a bloki A i B zawierają nie tylko liczbowe transmitancje, ale i np. operatory całkowania, różniczkowania i ewentualnie bardziej złożone modele matematyczne dynamiki układów. Przypadki takie są jednak nieporównanie trudniejsze w analizie i nawet orzekanie o stabilności takich dynamicznych systemów może się okazać trudne.

10. Z właściwości systemu regulacji omawianych w punkcie 5 wiemy, że system ten posiada zdolność do utrzymywania wielkości wyjściowej X w pobliżu wymuszenia Z. Jeżeli więc wartość Z będzie niezmienna w czasie, to pomimo działania zakłóceń, wielkość X będzie również pozostawała praktycznie nie-

zmienną, system taki możemy więc nazwać S Y S T E M E M S T A B I L I Z A C J I. Wielkość Z możemy również zmieniać według pewnego zadanego programu. System regulacji będzie dążył do utrzymania zmian wielkości X w pobliżu zmian wielkości Z, zatem X będzie również zmieniało się zgodnie z założonym programem. System możemy wówczas nazwać S Y S T E M E M P R O G R A M O W Y M. Ostatni przypadek może polegać na tym, że wielkość Z jest od nas niezależna, a chcielibyśmy, aby wielkość X odwzorowywała wszelkie zmiany wielkości Z /np. położenie anteny radarowej X powinno śledzić za nieznanymi z góry zmianami położenia samolotu Z/. System taki nazwiemy wówczas systemem Ś L E D Z Ą C Y M.

TEST 2

P. Jako przykład systemu programowego rozpatrzmy znany z testu 1 model gospodarki Keynesa, w którym określony mamy przebieg wypłat inwestycyjnych /Z/t/. Określić czasowy przebieg wielkości dochodu narodowego.

xxx

O. T poprzednio wyprowadzonego wzoru mamy

$$X/t/ = \frac{1}{1 - B} Z/t/$$

P. Posługując się jak poprzednio modelem Keynesa, ustalić należy przebieg czasowy potrzebnych inwestycji, jeśli z prognoz demograficznych wynika, że ilość zatrudnionych będzie wzrastała zgodnie z funkcją N/t/, a wielkość produkcji czystej jest proporcjonalna do ilości zatrudnionych:

$$P/t/ = A\ N/t/$$

xxx

O. W teście pierwszym wyprowadzono wzór na wielkość nakładów inwestycyjnych niezbędną do uzyskania danej wartości produkcji czystej:

$$Z = /1 - B/\ P$$

Podstawiając i wprowadzając zależności czasowe mamy:

$$Z/t/ = /1 - B/\ P/t/ = /1-B/\ A\ N/t/$$

P. Na jakich założeniach oparty jest uzyskany powyżej wzór? Co należałoby zmienić, aby był on bliższy rzeczywistości?

xxx

O. Przy wprowadzaniu ostatniego wzoru założono, że zarówno stopa konsumpcji B, jak i pracochłonność produkcji A są niezmienne w czasie. Tymczasem dzięki naturalnemu dążeniu ludzi B jest /a przynajmniej powinno być/ rosnącą funkcją czasu, natomiast A, dzięki postępowi technicznemu, powinno z czasem maleć.
Ostatecznie więc wzór na przebieg inwestycji w stosunku do przewidywanego zatrudnienia powinien mieć postać

$$Z/t/ = /1 - B/t/\ /\ A/t/\ N/t/$$

Rozdział III

TEORIA WIELKICH SYSTEMÓW I OPTYMALNEGO STEROWANIA SYSTEMAMI GOSPODARCZYMI

1. Wielkie systemy

1. Systemy dotychczas rozpatrywane składały się z kilku zaledwie układów stosunkowo prosto ze sobą połączonych. Dzięki temu analiza tych systemów nie nastręczała większych trudności, niż analiza samych układów. Sytuacja zmienia się w istotny sposób, gdy zamiast kilku układów na system składa się ich kilkaset, oraz kiedy system jest wielowymiarowy. Niemożliwe staje się wówczas określenie przebiegów wszystkich wielkości w systemie, i to zarówno "ręcznie", jak i przy użyciu obecnie stosowanych maszyn matematycznych.

2. Pojęcie systemu "wielkiego" jest względne, i jest wyrazem nie tyle wielkości samego systemu, lecz raczej trudności analizy. Podobnie, jak dla pewnych prymitywnych plemion każda ilość przekraczająca trzy nazywa się "dużo", tak i dla współczesnej cybernetyki, każdy system, który ze względu na swą wielkość nie poddaje się analizie, nazywany jest wielkim.

3. Nie możemy więc ściśle określić, które systemy są "wielkie", mamy jednak pewność, że wszystkie systemy naprawdę adekwatne do realnych systemów gospodarczych i nie zawierające drastycznych uproszczeń będą bardzo wielkie, gdyż zarówno mnogość i wielorakość sprzężeń, jak i wielowymiarowość tych systemów jest nieporównanie większa, niż jakichkolwiek innych systemów, z wyjątkiem może jedynie biologicznych.

4. W systemie wielkim, ze względu na niezbędną ilość

obliczeń, nie można dokonać pełnej identyfikacji wszystkich układów, stanowiących części składowe systemu. Niemożliwa jest również pełna identyfikacja systemu jako całości, gdyż wymagałoby to zebrania materiałów o wpływie każdego wejścia na każde wyjście, co wymaga kolosalnej pracy. Identyfikację /znalezienie formuł opisu systemu/ prowadzimy więc najczęściej jedynie wyrywkowo, interesując się konkretnie wpływem danego wejścia /lub grupy wejść/ na dane wyjście /lub grupę wyjść/, traktując z konieczności wszystkie inne powiązania jako zakłócenia.

5. Inną metodą identyfikacji wielkiego systemu jest badanie go metodami statystyki matematycznej. Obserwując więc system przez długi okres czasu gromadzimy kolejne obserwacje wartości wejściowych na odpowiednich wejściach, oraz równocześnie bądź z uwzględnieniem opóźnień występujących wartości wyjściowych na odpowiednich wyjściach.

Wielkości w ten sposób uzyskane traktujemy jako realizacje pewnych procesów przypadkowych i dokonujemy obliczeń mających na celu określenie współczynników regresji liniowej pomiędzy określonymi wielkościami wejściowymi i wyjściowymi. Współczynniki te mogą służyć jako pierwsze przybliżenie wielkości transmitancji wiążących odpowiednie wejścia i wyjścia.

Metoda ta posiada wszystkie wady metod statystycznych, a mianowicie: wymaga do uzyskania wiarogodnych wyników stosowania wielkiej ilości obserwacji, zależności uzyskane muszą być poddane dodatkowo testowi na statystyczną istotność, aby mogły być uznane za miarodajne /lub nie!/ i daje wyniki silnie przybliżone, szczególnie w przypadku, jeśli rzeczywiste zależności są nieliniowe.

Zaletą tej metody jest fakt, że stosowanie jej nie wymaga żadnych apriorycznych danych o strukturze rozpatrywanego systemu, oraz że metoda ta niweluje wpływ zakłóceń na obserwowane w czasie pomiarów wartości, jeśli tylko zakłócenia mają charakter czysto przypadkowy.

TEST 1

P. Rozpatrzmy dla prostoty wyłącznie jedno /wybrane/ wejście do systemu i jedno wybrane wyjście z systemu. Będziemy się starali określić przybliżoną wartość transmitancji wiążącej dane wejściowe z danym wyjściem, mając do dyspozycji następujące obserwacje:

X	1	2	3	4	5
Y	2	3	7	7	11

xxx

O. Będziemy się starali dobrać współczynniki A i B /współczynniki regresji liniowej/ tak, aby podane wartości możliwie najlepiej odpowiadały wartościom obliczonym z równania

$$Y = A + B X \qquad /38/$$

Dopasowanie uznamy za najlepsze, jeśli suma kwadratów różnicy pomiędzy wartościami rzeczywistymi a wartościami obliczonymi ze wzoru /38/ będzie minimalna /jest to tzw. dopasowanie metodą najmniejszych kwadratów/. Współczynniki A i B wyrażają się wzorami /patrz Z.Hellwig "Elementy rachunku prawdopodobieństwa i statystyki matematycznej"/[10]:

$$B = \frac{E/XY/ - E/X/\,E/Y/}{E/X^2/ - E^2/X/}, \qquad /39/$$

$$A = E/Y/ - B\,E/X/\,. \qquad /40/$$

Zapis E/X/ oznacza średnią wartość zmiennej X, podobnie wszystkie inne zapisy. W związku z tym mamy:

$E/X/ = 3 \quad E/XY/ = 22{,}4 \quad E/Y/ = 6 \quad E/X^2/ = 11$

Podstawiając te wartości do wzorów /39/ i /40/ otrzymamy

$A = -0{,}6 \qquad B = 2{,}2$

10/ Z.Hellwig: Elementy rachunku prawdopodobieństwa i statystyki matematycznej, PWN, Warszawa 1972.

Przybliżona funkcja przejścia wynosi więc

$$Y = -0,6 - 2,2\ X \qquad /41/$$

Ewentualne wątpliwości, jakie nasunęły się przy rozwiązywaniu /lub czytaniu rozwiązania/ tego zadania, powinieneś usunąć przy pomocy książki Z.Hellwiga lub innego podręcznika statystyki matematycznej.

P. Czy obliczona w poprzednim zadaniu funkcja przejścia dokładnie opisuje zachowanie systemu? Jaka jest tego przyczyna?

xxx

O. Opis jest przybliżony. Upewnimy się o tym obliczając ze wzoru /41/ wartości Y w tych samych punktach X, w których mamy dane eksperymentalne:

X	1	2	3	4	5
Y	1,6	3,8	6,0	8,2	10,4

Jak widać wartości te nie zgadzają się z eksperymentalnymi. Przyczyn tego stanu rzeczy jest wiele: dane eksperymentalne mogły być obarczone błędem pomiaru, zależność rzeczywista mogła być inna niż założona we wzorze /38/, np. nieliniowa, nie uwzględnialiśmy wpływu innych wejść itd. Niemniej otrzymana aproksymacja jest najlepszą z możliwych przy przyjętych założeniach.

6. **W wielkich systemach** zadanie nie polega na stabilizacji czy programowych zmianach pewnych wielkości, jak to miało miejsce w prostych systemach regulacji. System posiada zwykle jakiś cel, który powinien wypełniać w sposób w określonym sensie optymalny, należy więc w kontekście tych systemów podać wiadomości, dotyczące S T E R O W A N I A O P T Y M A L N E G O.

2. Sterowanie optymalne

1. Każdą działalność człowiek stara się wykonywać w określonym sensie optymalnie. Jeśli jednak chcemy, aby optymalne było również zachowanie tworu cybernetycznego, jakim jest system, musimy słowu optymalnie nadać sens ścisły i matematyczny. W tym celu wprowadza się pojęcie K R Y T E R I U M   J A K O Ś C I. Kryterium to jest pewną funkcją wielkości wyjściowych systemu, zależną od celów, jakie stawiamy systemowi. Ponieważ wielkości, od których zależy wartość kryterium, są często funkcjami czasu /przebiegi wielkości wyjściowych/, więc kryterium jakości może być również funkcjonałem.

2. Zagadnienie optymalizacji systemu sformułować teraz można jako zagadnienie sformułowania kryterium jakości i jego maksymalizacji lub minimalizacji. Jest to z matematycznego punktu widzenia już zadanie dobrze postawione: należy znaleźć ekstremum funkcjonału przy ograniczeniach wynikających z matematycznego opisu systemu. Zagadnienie takie leży zasadniczo w obrębie działu matematyki zwanego rachunkiem wariacyjnym, jednakże ze względu na jego specyfikę i stopień trudności wypracowano w cybernetyce szereg metod specjalnych, służących do rozwiązywania zadań optymalizacji systemów.

Na uwagę zasługują szczególnie dwie metody: oparte na pracach radzieckiego matematyka Pontriagina metody zwane zasadą maksimum [11/] i zaproponowana amerykańskiego badacza R. Bellmana koncepcja programowania dynamicznego [12/].

3. Obie wymienione wyżej metody matematyczne wymagają dla pełnego zrozumienia bardzo zaawansowanego przygotowania ma-

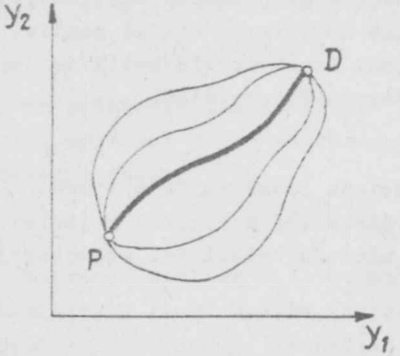
---

11/ Pontriagin: Matematyczna teoria procesów optymalnych, MON, Warszawa 1968.

12/ R.Bellman, S.Dryfus: Programowanie dynamiczne. Zastosowania, PWE, Warszawa 1967.

tematycznego, postaramy się więc jedynie zreferować istotę i przedstawić przykłady, natomiast szczegóły wywodów matematycznych zostaną całkowicie pominięte.

4. Przyjmijmy na początek, że kryterium jakości systemu jest czas przejścia układu z jednego stanu w drugi, ściśle określony. Zadanie to można przedstawić w przestrzeni stanów przez oznaczenie punktu początkowego P i docelowego D /rys.58/.



Rys. 58. Trajektoria optymalna

W zależności od sterowania system może przebyć drogę od stanu P do stanu D po różnych trajektoriach, z których tylko jedna /pogrubiona na rysunku/ zapewnia minimalny czas przejścia. Trajektorię tę nazwiemy optymalną, a sterowanie, które zmusza system do poruszania się po trajektorii optymalnej, nazwiemy sterowaniem optymalnym.

5. Charakter sterowania optymalnego zależny jest nie tylko od rodzaju postawionego kryterium optymalności i od charakteru równań opisujących system, ale w głównej mierze od ograniczeń nałożonych na możliwe sterowania.

Gdyby ograniczeń nie było, dla uzyskania maksymalnej szybkości przebiegów trzeba było operować nieskończenie wielkimi mocami. Jest to rozwiązanie tyleż nierealne /nie dysponujemy możliwością rozwinięcia mocy większej niż określona/, co i nieinteresujące z matematycznego punktu widzenia: takie rozwiązanie można wymyślić zwyczajnie zastanowiwszy się nad istotą ruchu w przestrzeni stanów, a bez żadnych obliczeń.

6. Jeżeli sterowanie dokonywane jest z uwzględnieniem ograniczeń, to mają one zwykle właśnie postać ograniczenia mocy sterowania lub ograniczenia całkowitej ilości energii /nakładów/ zużytej na sterowanie.

Sterowanie optymalne polega więc na rozdziale posiadanego zasobu środków w najlepszych porcjach. Z zasady maksimum wynika, że należy w tych warunkach dostarczać środki możliwe w dużych "porcjach", gdyż wówczas uzyskuje się możliwie największą szybkość przebiegów. Rozważymy przykład.

Przykład 1

Niech system jako całość posiada jedno wejście i dwa wyjścia. Niech prędkość zmian pierwszej wielkości wyjściowej Y_1 będzie równa aktualnej wartości wielkości wejściowej:

$$\frac{dY_1}{dt} = X \qquad /42/$$

natomiast przyspieszenie zmian /druga pochodna/ drugiej wielkości wyjściowej Y_2 niech będzie też równe wielkości wejściowej:

$$\frac{d^2Y_2}{dt^2} = X \qquad /43/$$

Nietrudno zauważyć, że przy tak postawionym zadaniu Y_1 jest niczym innym, jak szybkością zmian wielkości Y_2.

$$Y_1 = \frac{dY_2}{dt} \qquad /44/$$

Przyjmijmy, że wielkość sygnału sterującego X nie może przewyższać wartości +1 lub -1.

Z zasady maksimum wynika, że aby nasze sterowanie było optymalne w sensie minimalnego czasu przeprowadzenia systemu z jednego stanu do innego, powinniśmy operować maksymalnymi sygnałami, czyli $X = +1$ lub $X = -1$. Musimy teraz wyłącznie określić, kiedy jakim sygnałem trzeba działać.

Postawmy konkretne pytanie: Należy w najkrótszym czasie sprawdzić naszy system z określonego punktu A w przestrzeni stanu do początku układu współrzędnych tej przestrzeni. Określimy najpierw, jakie trajektorie są w przestrzeni stanu możliwe przy dostępnych nam sterowaniach. Jeżeli działać będziemy sygnałem $X = + 1$, to czasowy przebieg zmiennej Y_1 wyznaczymy całkując wyrażenie /42/:

$$Y_1 = +t + C_1 \qquad /45/$$

Na podstawie równania /44/ czasowy przebieg zmian wielkości Y_2 uzyskamy przez scałkowanie wyrażenia /45/ jeszcze raz względem czasu:

$$Y_2 = \frac{1}{2} t^2 + C_1 t + C_2 \qquad /46/$$

Wartości stałych C_1 i C_2 określimy z warunku, że w chwili początkowej /t=0/ system znajdował się w punkcie A o współrzędnych Y_{1A} i Y_{2A}. Wstawiając te wartości do równań /45/ i /46/ otrzymamy, że

$$C_1 = Y_{1A}, \qquad C_2 = Y_{2A} .$$

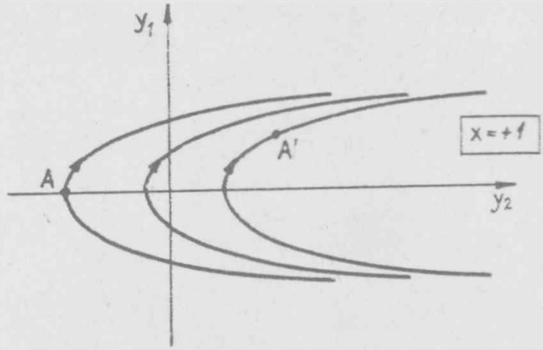
Ruch systemu opisują więc dwie zależności czasowe:

$$\begin{aligned} Y_1 &= t + Y_{1A}, \\ Y_2 &= \frac{1}{2} t^2 + Y_{1A} t + Y_{2A} \end{aligned} \qquad /47/$$

Z równań /47/ można wyeliminować czas /np. obliczając go z pierwszego równania i wstawiając do drugiego/. Otrzymamy wówczas równanie trajektorii we współrzędnych Y_1, Y_2 .

$$Y_2 = \frac{1}{2} Y_1^2 + /Y_{2A} - \frac{1}{2} Y_{1A}^2 / \qquad /48/$$

Trajektorie opisane równaniem /48/ przedstawiają parabole w przestrzeni stanów /rys. 59/. Położenie paraboli zależy od położenia punktu A.

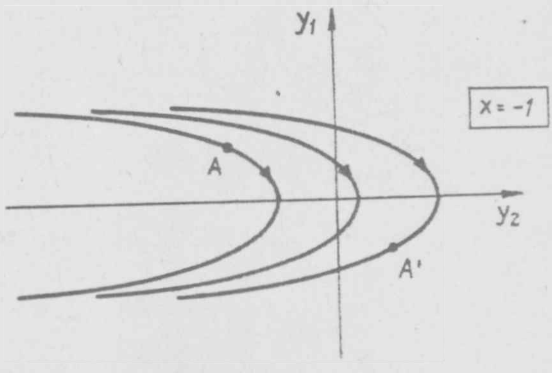


Rys. 59. Trajektoria przy sterowaniu x = + 1

Dla X = - 1 możemy przeprowadzić analogiczne rozumowanie, otrzymując w rezultacie równanie trajektorii

$$Y_2 = -\frac{1}{2} Y_1^{\,2} + /Y_{2A} + \frac{1}{2} Y_{1A}^2 / \qquad /49/$$

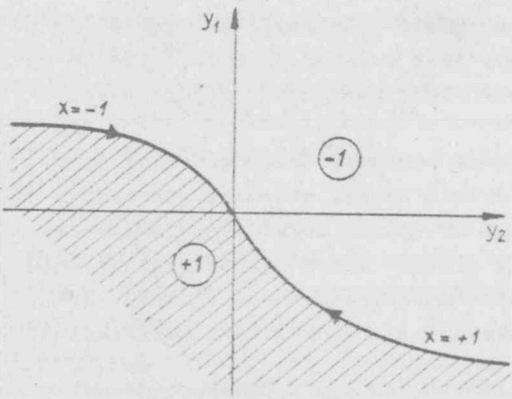
Rodzinę tych trajektorii dla różnych położeń punktu A przedstawia rysunek 60.



Rys. 60. Trajektoria przy sterowaniu x = - 1

Z wyjątkiem przypadkowych położeń punktu A względem początku układu współrzędnych nie ma trajektorii, która by bezpośrednio doprowadzała system do początku układu współrzędnych. Sterowanie optymalne będzie więc polegało na wybraniu takiej drogi po kawałkach trajektorii odpowiadających +1 i -1, aby doprowadziła ona z zasadnego punktu do początku układu. Minimalny czas takiego "przejazdu"zagwarantowany jest przez przyjęcie, wynikającej z zasady maksimum, "marszruty" wyłącznie po częściach trajektorii odpowiadających maksymalnym możliwym wartościom sygnałów sterujących.

Określimy teraz algorytm wyboru trajektorii, a więc i sygnałów sterujących. Z rysunku 59 wynika, że jest tylko jedna trajektoria przy X = +1 pozwalająca dojść do punktu docelowego /początku układu/. Przebiega ona w IV ćwiartce układu współrzędnych /dalszy ciąg tej trajektorii nie jest uwzględniany, ponieważ prowadzi od początku układu/. Jedna jest też taka trajektoria na rysunku 60. Obie te trajektorie dają linię dojścia, pozwalającą na dotarcie do celu. Linia ta dzieli cały obszar przestrzeni stanu na dwa obszary /rys. 61/.



Rys. 61. Sterowanie optymalne

W obszarze pod linią /zakreskowanym/ należy najpierw zastosować sterowanie pełnym dodatnim sygnałem /X = +1/, a po dotarciu do linii dojścia włączyć sygnał X = - 1. Dokładnie przeciwna procedura obowiązuje przy sterowaniu z punktów leżących powyżej krzywej.

Algorytm sterowania jest więc następujący:
Dla linii granicznej z równań /48/ i /49/ otrzymujemy zależności:

$$Y_2 + \frac{1}{2} Y_1^2 \operatorname{sign} Y_1 = 0 \qquad /50/$$

Funkcja sign występująca we wzorze /50/ oznacza znak i przyjmuje tylko wartości +1 i -1.

$$\operatorname{sign} Y_1 = \begin{cases} +1 & \text{dla } Y_1 > 0 \\ -1 & \text{dla } Y_1 < 0 \end{cases} \qquad /51/$$

Łatwo sprawdzić, że powyżej linii dojścia równanie /50/ zmienia się w nierówność:

$$Y_2 + \frac{1}{2} Y_1^2 \operatorname{sign} Y_1 > 0, \qquad /52/$$

a poniżej, w nierówność

$$Y_2 + \frac{1}{2} Y_1^2 \operatorname{sign} Y_1 < 0 \qquad /53/$$

Równocześnie w obszarze, gdzie obowiązuje zależność /53/, początkowy sygnał sterujący jest dodatni, a w obszarze, w którym obowiązuje zależność /52/, początkowy sygnał sterujący jest ujemny.

Reasumując stwierdzimy, że tam gdzie wartość lewej strony wyrażenia /50/ jest dodatnia należy stosować sterowanie ujemne, a tam gdzie ujemna, dodatnie. Regułę tę nazywamy regułą optymalnego /w sensie minimum czasu potrzebnego na zmianę położenia w przestrzeni stanów/ sterowania systemu. Regułę tę znaleźliśmy w oparciu o zasadę maksimum Pontriagina.

7. Z obszernie przeliczonego przykładu wynikają dwa wnioski: Nawet w tak prostym systemie, jak rozpatrywany przez nas, i przy tak prostym kryterium jak minimalny czas działania systemu, ilości obliczeń potrzebne do wyznaczenia sterowania optymalnego są dość znaczne. Sytuacja komplikuje się ogromnie w przypadku systemów o większej ilości wymiarów i

bardziej ciekawych kryteriach /wskaźnikach/ jakości.

Po drugie, przy optymalizacji systemu decydujący wpływ na wybór określonego sterowania mają ograniczenia, jakim podlegają nasze możliwości generacji sygnałów sterujących. Wniosek ten ugruntujemy w przykładzie 2, który zilustruje zastosowanie metody programowania dynamicznego.

Przykład 2

Niech naszym systemem będzie ferma hodowlana. Wskaźnikiem jakości niech będzie dochód ze sprzedaży bydła w ciągu N lat: Jeżeli cenę bydła /za sztukę/ oznaczymy przez C, a ilość sprzedanego bydła w poszczególnych latach przez Y_i /w i-tym roku/, to wskaźnik ten ma postać:

$$Q = \sum_{i=1}^{N} C\,Y_i, \qquad /54/$$

czyli jest po prostu sumą dochodów pochodzących ze sprzedaży w poszczególnych latach. W przeciwieństwie do poprzedniego przykładu, w tym przypadku chodzi nam o maksymalizację wskaźnika. Przyjmiemy najpierw, że nie uwzględniamy żadnych ograniczeń, że na początku posiadamy X_0 sztuk bydła, i że roczny wzrost pogłowia jest k - krotny. /tj. z X_0 sztuk po roku, jeśli nic nie sprzedamy, będziemy mieli $X_1 = k\,X_0$ sztuk/.

Zadanie, które aktualnie sobie stawiamy, to znaczy określenie najwłaściwszych ilości bydła sprzedawanych w kolejnych latach, wygodnie jest rozwiązywać od końca, to jest poczynając od momentu, kiedy w N-tym roku likwidujemy hodowlę i sprzedajemy całą pozostałą część bydła. Koncepcja ta stanowi istotny moment programowania dynamicznego Bellmana: należy najpierw ustalić decyzję ostatniego etapu, następnie przedostatniego itd. aż do stanu początkowego.

Niech ilość bydła, którą sprzedamy w N-tym /ostatnim/ roku będzie Y_N. Otrzymamy za nie $C\,Y_N$ złotych. Ilość Y_N jest, zgodnie z założeniem o likwidacji, równa ilości posiadanego w N-tym roku bydła.

$$Y_N = X_N = k\, X_{N-1} \qquad /55/$$

Ustalmy teraz, jaką ilość bydła możemy sprzedać w roku N-1.

Jeżeli sprzedamy Y_{N-1} sztuk, to o tyle mniejsza będzie wartość X_{N-1}. Za tych Y_{N-1} sprzedanych sztuk otrzymamy $C\, Y_{N-1}$ złotych, ale kwota otrzymana w N-tym roku zmaleje o $k\, C\, Y_{N-1}$ złotych, zatem sprzedanie w roku poprzedzającym likwidację chociażby jednej sztuki naraża nas na straty /k jest większe od jedności!/. Rozumowanie to można przenieść na N-2, N-3 i dalsze lata, aż do pierwszego roku, dochodząc generalnie do wniosku, że maksimum wskaźnika /54/ /zysku/ uzyskamy sprzedając całe bydło w ostatnim roku, nie dokonując żadnych sprzedaży w pozostałych latach. Uzyskamy wówczas dochód wynoszący

$$Q_{max} = k^N\, C\, X_o \qquad /56/$$

Wprowadźmy teraz do naszych rozważań ograniczenie w postaci warunku, że nie możemy równocześnie hodować więcej niż X_{max} sztuk bydła.

W pierwszym roku musimy więc sprzedać

$$Y_1 = k\, X_o - X_{max}$$

w drugim i wszystkich dalszych, aż do N-1

$$Y_2 = Y_3 = \ldots = Y_{N-1} = X_{max}\, /k - 1/$$

oraz na koniec w roku N-tym

$$Y_N = X_{max} \cdot k$$

Możemy teraz, korzystając ze wzoru /54/, określić uzyskany dochód:

$$Q = C \left(X_o\, k + /N - 1/\, /k - 1/\, X_{max}\right) \qquad /57/$$

Porównując wzory /56/ i /57/ widzimy, że uwzględnianie ograniczeń doprowadziło do strategii optymalnej, dającej mniejszą wartość kryterium. Możemy stąd wyciągnąć jeszcze jeden wniosek ogólny: sterowanie optymalne zależy silnie od ograniczeń, i to nie tylko w sensie zależności strategii optymalnej, ale i w sensie uzyskiwanych efektów. Uwzględnianie ograniczeń zwykle prowadzi do komplikacji procesu obliczeniowego przy poszukiwaniu strategii optymalnej, sprawia, że strategia ta jest dużo bardziej skomplikowana niż bez uwzględniania ograniczeń oraz wpływa na drastyczne pogorszenie efektów, jakie możemy uzyskać przez zastosowanie systemu optymalnego.

TEST 1

P. Jaki charakter ma sterowanie w systemach, w których wielkość sterująca jest ograniczona co do maksymalnej wartości, a kryterium jest minimalny czas przejścia od jednego do drugiego stanu?
Z czego to wynika?

xxx

O. Musimy wówczas operować maksymalnymi dopuszczalnymi wartościami sygnału sterującego, przełączając ewentualnie z sygnału "cała naprzód" na sygnał "cała wstecz" lub odwrotnie. Niejednokrotnie /zwłaszcza w systemach wielowymiarowych, tutaj nie omawianych/ potrzebna jest spora liczba takich przełączeń. Od stosowanej metody sterowania /"uderzeniowo", pełnymi wielkościami/ systemy takie i strategia sterowania w nich stosowana noszą nazwę "systemów Bang-Bang". Metoda takiego sterowania oparta jest na wywodach teorii Pontriagina, a w szczególności na sformułowanej przez niego tzw. zasadzie maksimum. Ze względu na skomplikowaną postać matematyczną teorii tej przytaczać nie będziemy.

P. Jakie skutki pociąga za sobą nieuwzględnienie ograniczeń przy opracowywaniu strategii sterowania optymalnego?

xxx

O. Opracowana strategia jest na ogół nierealna lub nie prowadzi do oczekiwanych rezultatów, a oczekiwania w stosunku do osiąganych wskaźników jakości są zbyt optymistyczne. Niewykluczone, że wiele niepowodzeń w procesie planowania i rozczarowań realnymi osiągnięciami wynika z nieuwzględniania istniejących ograniczeń.

P. Jeżeli mielibyśmy optymalizować proces wieloetapowy, to w jaki sposób należałoby do tego zagadnienia przystąpić? Jak się nazywa taka metodyka postępowania?

xxx

O. Należałoby zacząć od ostatniego etapu, gdzie stan docelowy nakłada najwięcej ograniczeń. Metoda ta, zwana programowaniem dynamicznym, szczególnie dobrze nadaje się do programowania przy pomocy elektronicznych maszyn liczących. Zainteresowanych / i odpowiednio przygotowanych matematycznie!/ czytelników odsyłamy do pracy R.Bellman "Dynamic programing" lub tego autora "Adaptacyjne procesy sterowania".

8. Do zagadnień optymalizacji i sterowania optymalnego w wielkich systemach należy również problematyka tzw. programowania matematycznego. Jest to dział matematyki stosowanej, który powstał i rozwinął się jeszcze przez powstaniem cybernetyki, a obecnie został przez nią zaanektowany i znajduje szerokie zastosowanie.

9. Najprostszym z problemów programowania matematycznego jest problem tzw. programowania liniowego. Zagadnienie to występuje wówczas, gdy kryterium jakości systemu jest funkcją liniową współrzędnych stanu lub wielkości wyjściowych systemu /np. tak jak we wzorze /54/, zaś ograniczenia i transmitancje systemu są również funkcjami liniowymi. Na gruncie programowania liniowego wyrosło obecnie szereg inte-

resujących i bardzo pożytecznych metod, których szczegółowy wykład nie jest celem tej książki, ograniczymy się więc jedynie do krótkiej wzmianki.
Rozpoczniemy przykładem:

Przykład 3

Przyjmijmy, że systemem nazwiemy zakład produkcyjny wytwarzający n towarów w ilościach $X_1, X_2, \ldots, X_N$. Jeżeli jednostkowy zysk zakładu n i-tym wyrobie oznaczymy C_i /i = 1,2 ..., N/, to oczywistym kryterium jakości przyjętej polityki produkcyjnej będzie forma liniowa:

$$Q = \sum_{i=1}^{N} C_i X_i \qquad /58/$$

Zależeć nam będzie na maksymalizacji tej formy. Musimy jednak liczyć się z obecnością ograniczeń. Do produkcji jednostki i-tego towaru potrzebny jest mianowicie j-ty surowiec, w ilości równej A_{ij}. Tymczasem podaż j-tego surowca jest ograniczona i nie przewyższa B_j. Musimy więc planując produkcję zapewnić spełnienie warunku nieprzekraczania przez ł ą c z n e zapotrzebowanie na j-ty surowiec podaży tego surowca:

$$\sum_{i=1}^{N} A_{ji} X \leqslant B_j \qquad j = 1, 2, \ldots, M \quad /59/$$

Nierówności typu /22/ mamy tyle, ile używamy surowców, czyli dla j = 1,2, ..., M. Zagadnienie będzie w pełni określone, jeśli w dodatku postawimy wymaganie, aby produkcja była nieujemna:

$$X_i \geqslant 0 \qquad i = 1,2, \ldots, N \quad /60/$$

Programowaniem liniowym nazwiemy metodę znajdowania maksimum /lub minimum/ funkcji liniowej postaci /58/ przy obecności systemu ograniczeń /59/ i /60/.

Wskażemy teraz metody rozwiązania tego problemu, bardziej dociekliwych odsyłając do książki D.Gale "Teoria liniowych modeli ekonomicznych".

Zauważmy najpierw, że ograniczenia /60/ zawężają obszar, w którym moglibyśmy poszukiwać rozwiązania, do tej części przestrzeni stanu, w której wszystkie współrzędne są dodatnie. Na płaszczyźnie byłaby to pierwsza ćwiartka układu współrzędnych. Następnie ograniczenia /59/ przedstawiają, po zastąpieniu nierówności znakiem =, równania hiperpłaszczyzn w przestrzeni N - wymiarowej /na płaszczyźnie są to równania prostych/. Zespół tych ograniczeń wyznacza obszar, w którym możemy poszukiwać optymalnego zespołu zmiennych $X_1, \ldots, X_N$. W przestrzeni wielowymiarowej jest to wielościan, a na płaszczyźnie wielokąt. Otóż jedno z podstawowych twierdzeń teorii programowania liniowego głosi, że rozwiązanie optymalne, jeśli istnieje, odpowiada jednemu z naroży tego wielościanu /jednemu z wierzchołków wielokąta/. Rozpatrzmy to na bardziej szczegółowym przykładzie 4.

<u>Przykład 4</u>

Niech w omawianym powyżej systemie będzie N=2, M=2, ceny niech wynoszą odpowiednio $C_1 = 2$ i $C_2 = 1$, a ilości surowców zużywanych do produkcji $A_{11} = \frac{1}{2}$, $A_{21} = 3$, $A_{12} = 1$, $A_{22} = 1$.

Ujemna wartość współczynnika A_{21} oznacza, że przy produkcji pierwszego wyrobu drugi surowiec powstaje jako produkt uboczny, co można oznaczyć, tak jak to zrobiono, jako ujemne zużycie. Ilości surowców będące na rynku wynoszą $B_1 = 2$, $B_2 = 9$.

Kryterium /59/ ma więc postać

$$Q = 2X_1 + X_2 \qquad /61/$$

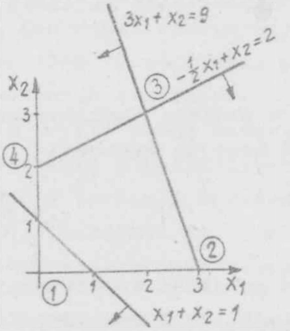
i należy je maksymalizować przy ograniczeniach /59/:

$$\begin{aligned} \tfrac{1}{2} X_1 + X_2 &\leqslant 2 \\ 3 X_1 + X_2 &\leqslant 9 \end{aligned} \qquad /62/$$

oraz ograniczeniach /60/

$$X_1 \geqslant 0 \qquad X_2 \geqslant 0 \qquad /63/$$

Wymienione ograniczenia narysowane zostały w układzie X_1, X_2 w postaci prostych /rysunek 62/ ograniczających pewien obszar. Strzałki wskazują, po której stronie linii granicznej leży obszar dopuszczalny /obszar, w którym nierówności są spełnione/.



Rys. 62. Programowanie liniowe

Zgodnie z twierdzeniem, na które powołano się w zakończeniu poprzedniego przykładu, kryterium powinno przybierać wartość maksymalną w którymś z czterech, ponumerowanych na rysunku wierzchołków obszaru dopuszczalnego. Należy wobec tego sprawdzić wartości kryterium w wierzchołkach. Dla wierzchołka 1 Q = 0 + 0 = 0, dla wierzchołka 2 Q = 2 . 3 + 0 = 6, dla wierzchołka 3 Q = 2 . 2 + 3 = 7 i dla wierzchołka 4 Q = = 0 + 2 = 2.

Jak widać kryterium przybiera maksymalną wartość w narożu 3, a więc optymalny program produkcji powinien przewidywać produkcję dwu jednostek pierwszego produktu i trzech jednostek drugiego produktu. Zastosowana metoda "przeglądania" naroży nosi w literaturze nazwę metody Simplex. Metodę tę trudno jest stosować "ręcznie" do systemów wielowymiarowych, których obszary dopuszczalne zawierają nieraz setki naroży, i dlatego metoda Simplex jest opracowana w wersji pozwalającej na zastosowanie maszyny cyfrowej. Program rozwiązywania zadania programowania liniowego metodą Simplex jest standardowym programem wszystkich maszyn matematycznych ogólnego zastosowania.

TEST 2

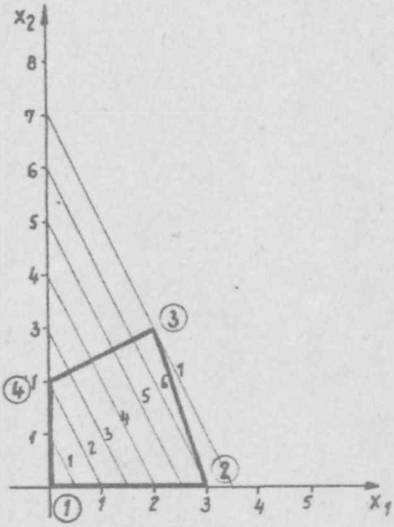
P. W przestrzeni stanu z przykładu 4 wykreślić powierzchnie /linie/, na których jednakowa jest wartość kryterium jakości Q i wykazać, że optymalne naroże leży na linii o największej wartości kryterium.

xxx

O. Linie stałej wartości otrzymamy w postaci analitycznej, wstawiając do równania /61/ za Q kolejne wartości. Np. dla Q = 1 linia ma równanie

$$1 = 2 X_1 + X_2$$

Zbiór takich linii naniesiono na układ współrzędnych X_1, X_2 na rysunku 63. Widać z niego, że linię odpowiada-



Rys. 63. Linie stałej wartości kryterium przy programowaniu liniowym

dającą najwyższej wartości Q układ osiąga w narożu 3, natomiast wszystkie pozostałe punkty obszaru dopuszczalnego odpowiadają mniejszym wartościom kryterium.

10. Poza programowaniem liniowym w zakres programowania matematycznego jako metody optymalizacji wchodzą programowania nieliniowe. Stanowią one w zasadzie uogólnienie programowania liniowego na przypadki, kiedy kryterium lub ograniczenia są funkcjami nieliniowymi. Niestety, uogólnienie to nie pozwala również uogólniać metody Simplex - w zagadnieniach programowania nieliniowego do rozwiązania optymalnego dochodzi się zwykle drogą niezwykle mozolnych obliczeń, wymagających szybkiej i posiadającej obszerną pamięć maszyny cyfrowej. Metod tych nie będziemy omawiali, zainteresowanych odsyłamy do literatury np. C.W.Meriam "Teoria optymalizacji i projektowania układów sterowania automatycznego".

11. Optymalizację systemu można prowadzić w różnych kierunkach. Można między innymi poszukiwać optymalnego sterowania kompleksem operacji reprezentującym systemem, w sensie minimalnego czasu trwania kompleksu. Metodą matematyczną służącą do tego jest analiza sieciowa, znana również jako kompleks metod Pert.

Nie wdając się w szczegóły, które czytelnik łatwo znajdzie np. w monografii A.Idzikiewicz "Pert - metody analizy sieciowej", przedstawimy istotę tej metody.

12. Zanim zastosuje się metody analizy sieciowej, trzeba system, jako kompleks czynności, przedstawić w postaci rysunku siatki, której gałęzie reprezentują poszczególne czynności. Gałęzie te biorą początek w węzłach i kończą się w węzłach; uważa się, że czynność reprezentowana gałęzią wychodzącą z węzła nie może być rozpoczęta, zanim nie zostaną ukończone wszystkie czynności dochodzące do węzła.

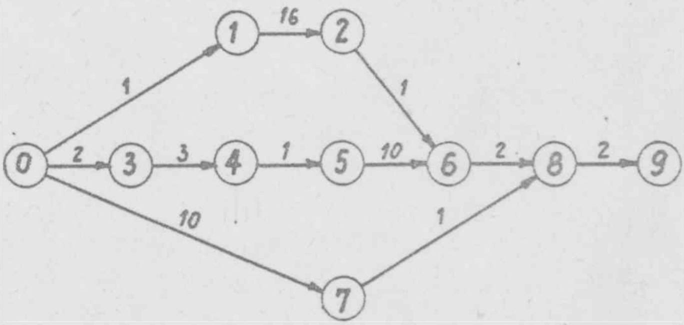
Rozpatrzmy ponownie prosty przykład.

Przykład 5

Niech kompleksem operacji, którego czas chcemy zoptymalizować, będzie zamontowanie silnika elektrycznego dużej mocy. W skład tego kompleksu wejdą następujące operacje:

Zamówienie płyty fundamentowej /0-1/
Wyprodukowanie płyty fundamentowej /1-2/
Przywóz płyty fundamentowej /2-6/
Przygotowanie podłoża pod fundament /0-3/
Przygotowanie deskowania /3-4/
Betonowanie /4-5/
Twardnienie betonu /5-6/
Montaż płyty fundamentowej /6-8/
Zamówienie i odbiór silnika /0-7/
Przewóz silnika /7-8/
Montaż silnika /8-9/

Wymienione operacje są wzajemnie powiązane, niektórych z nich nie można wykonać przed zakończeniem innych, co można przedstawić w postaci sieci, w której wymienione wyżej czynności figurować będą jako gałęzie łączące węzły wymienione w nawiasach w spisie operacji. Schemat ten przedstawiony jest na rysunku 64.



Rys. 64. Siatka czynności

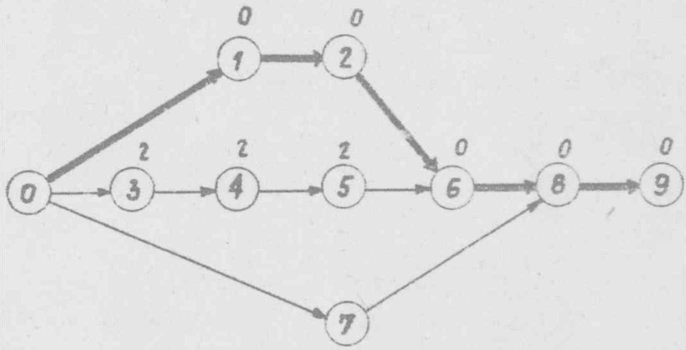
Widać, że kompleks operacji zaczyna się w węźle "0", kiedy to rozpoczynają się operacje zamówienia płyty fundamentowej, zamówienia silnika oraz przygotowania podłoża, a kończy w węźle 9, kiedy to silnik jest już zamontowany. W węźle 6 następuje połączenie zakończenia twardnienia betonu i przywozu płyty, co pozwala na rozpoczęcie montażu płyty, podobnie w węźle 8 zakończenie montażu płyty i przywóz silnika pozwalają na rozpoczęcie montażu silnika.

Każda z wymienionych operacji ma określony czas trwania czasy te zaznaczono na rysunku 64 w postaci liczb przy gałęziach reprezentujących poszczególne czynności. Posługując się tym rysunkiem i spisem z łatwością zauważymy na przykład, że czas twardnienia betonu wynosi 10 dni.

Ponieważ żadna operacja biorąca początek w węźle i-tym nie może się rozpocząć przed zakończeniem wszystkich operacji dochodzących do tego węzła, więc operacja biorąca początek w danym węźle zacznie się w czasie odpowiadającym maksymalnemu czasowi dotarcia od węzła początkowego do danego węzła. Czas trwania całego kompleksu operacji wyznaczany jest więc przez czas odpowiadający sumie czasów na drodze o maksymalnym czasie przejścia.

W naszym schemacie są trzy drogi: 0-1-2-6-8-9, 0-3-4-5-6-8-9, i 0-7-8-9. Obliczmy czasy przejścia dla tych dróg. Dla pierwszej drogi mamy: 1+16+1+2+2 = 22, dla drugiej 2+3+1+10+2+2 = 20 a dla trzeciej 10+1+2 = 13. Widać więc, że najdłuższa, w sensie czasowym, jest droga pierwsza, a czas jej przejścia jest jednocześnie minimalnym czasem kompleksu operacji.
Drogę taką nazywa się drogą krytyczną.

Mając określony czas ukończenia kompleksu operacji, wynikający z drogi krytycznej, możemy dla poszczególnych węzłów sieci określić rezerwy czasowe. Będą to różnice pomiędzy rzeczywistymi chwilami dotarcia do danych węzłów, a czasami wynikającymi z czasu przebycia drogi krytycznej. Na rysunku 65 oznaczono pogrubioną linią drogę krytyczną, a przy węzłach zapisano wielkości rezerw czasowych. Oczywiście wszystkie węzły wzdłuż drogi krytycznej nie mają rezerw czasowych, ponieważ każde opóźnienie na tej drodze prowadzi nieuchronnie do opóźnienia ukończenia kompleksu. Natomiast w węźle 7 mamy rezerwę czasową wynoszącą 9 dni, gdyż zamówienie i odbiór silnika trwa 10 dni, przewóz 1, a operacji biorącej



Rys. 65. Analiza drogi krytycznej

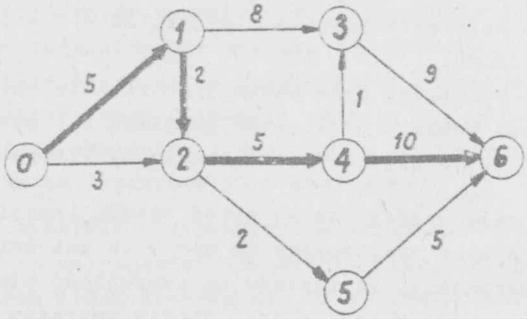
początek w węźle 8 nie będzie można rozpocząć przed upływem 20 dni, gdyż tyle potrzeba na dotarcie do tego węzła na drodze krytycznej.

Informacje uzyskane dzięki analizie drogi krytycznej pozwalają określić minimalny czas przejścia, drogę krytyczną i miejsca występowania rezerw. Sterowanie optymalne polega teraz na skupieniu środków nad skróceniem drogi krytycznej. Można to osiągnąć, bez uciekania się do środków z zewnątrz, poprzez przerzucenie części środków z dróg posiadających znaczne rezerwy. Stan optymalny nastąpi wtedy, gdy wszystkie drogi staną się krytyczne, gdyż wówczas środki będą rozdysponowane stosownie do potrzeb.

Poszukiwanie drogi krytycznej i określenie rezerw w kompleksach obejmujących tysiące czynności jest również zadaniem rozwiązywalnym jedynie przy pomocy maszyn matematycznych, ale też większość maszyn posiada w swych bibliotekach programy obliczeń sieciowych metodami Pert.

TEST 3

P. System stanowi kompleks operacji powiązanych zależnościami, jak na rysunku 66.



Rys. 66. Droga krytyczna /do testu 3 /

Rysunek ten zawiera również oznaczone na gałęziach czasy trwania poszczególnych operacji. Określić czasy przejścia wszystkich możliwych dróg w tym systemie i określić drogę krytyczną.

xxx

O. Wyniki zestawimy w postaci tabelki:

Droga	czas
0-1-3-6	22
0-1-2-4-6	22
0-1-2-4-3-6	22
0-2-4-6	18
0-2-4-3-6	18
0-2-5-6	10
0-1-2-5-6	14

Krytyczna jest więc droga 0-1-2-4-6.

P. Określić rezerwę czasu w punkcie 3 i 5.

xxx

O. W punkcie 3 operacja musi rozpocząć się nie później niż w czasie 22-9 = 13 dni od rozpoczęcia, a może się rozpocząć w chwili upływu 13 dni od rozpoczęcia /czas dojścia po drodze 0-1-2-4-3/. Rezerwa wynosi więc 0 dni. Podobnie obliczymy, że rezerwa w punkcie 5 wynosi 22 - 5 = 17, 17 - 2 - 3 = 12 dni.

Jeżeli nie wykonałeś tych zadań samodzielnie, powinieneś jeszcze raz uważnie przeczytać przykład 5 i ewentualnie uzupełnić swoje wiadomości ze wskazanej literatury.

13. Istnieje cały szereg innych metod optymalizacji w wielkich systemach. Wykład przytoczony tu nie jest ani ścisły, ani w pełni wyczerpujący, ze względu na szczupłość miejsca i konieczność operowania wyłącznie dość ubogim aparatem matematycznym. Może on jednak stanowić pewną wskazówkę, pozwalającą ocenić istniejące metody i określić ich użyteczność w konkretnych problemach.

3. Sterowanie hierarchiczne, adaptacyjne i dualne

1. Problemy wynikające przy optymalizacji wielkich systemów stanowią niezwykle złożony kompleks zagadnień obliczeniowych. Ilości obliczeń, niezbędne w celu uzyskania sterowania optymalnego, przekraczają możliwości najszybszych maszyn matematycznych, jeśli będziemy starać się określać wielkości sterujące dla wszystkich układów systemu i brać pod uwagę ogromne ilości wielkości wyjściowych z systemu.

2. Problem ten jednak można obejść wprowadzając odpowiednią strukturę sterowania systemu. Mianowicie, w żadnym z istniejących wielkich systemów nie podejmuje się wszystkich decyzji o rodzaju przyjętego sterowania w jednym miejscu, centralnie. Wprost przeciwnie, w organiźmie, który jest wielkim systemem biologicznym, większość procesów sterowania ustalanych jest przez oddzielne podsystemy zarządzania,a dopiero decyzje o znaczeniu zasadniczym, a jednocześnie dotyczące większej ilości podsystemów, podejmowane są centralnie, przez układ nerwowy.

Procesy przemiany materii w tkankach znajdują się pod kontrolą i są sterowane przez jądra komórkowe, zespoły tkanek znajdują się pod sterowaniem układów odruchów nieświadomych, a dopiero duże grupy narządów są sterowane przez celowe i świadome oddziaływanie systemu ośrodkowego.

Podobnie organizacja państwowa ma budowę hierarchiczną: losami konkretnych obywateli zajmują się układy decyzyjne najniższego szczebla, grup ludzi - wyższego szczebla, a całego państwa dopiero organy centralne.

3. W systemie hierarchicznym większość decyzji podejmowana jest na najniższych szczeblach, tam też opracowywana jest największa ilość informacji. Są to jednak decyzje mające małe znaczenie i mało istotne informacje. Do szczebla wyższego dociera informacja już wstępnie zagregowana i uśredniona, a więc w mniejszej ilości, lecz o większej wadze. Decyzje na tym szczeblu podejmowane mają też bardziej ogólny charakter, ale równocześnie ich wpływ na wartość kryterium, osiąganą przez system jako całość, jest dużo większy.

Wreszcie organ centralny wydaje decyzje o zasadniczym znaczeniu, w oparciu o niewielką ilość informacji syntetycznej.

W ten sposób możemy realizować sterowanie optymalne nie dławiąc się nadmiarem szczegółowej informacji, ani nie rozdrabniając się na szczegółowe instrukcje dla poszczególnych układów.

4. W praktyce zasadę tę stosuje się w wielkich systemach w ten sposób, że cały system dekomponuje się na możliwie mało nawzajem powiązane podsystemy /układy tych podsystemów mają większość sprzężeń zamykających się w ramach podsystemów i nader nieliczne powiązania międzysystemowie/. Poszczególne podsystemy rozwiązują swoje własne, mniejsze, a więc łatwiejsze do wykonania, zagadnienia optymalizacji, czyli poszukują sterowania, które zapewniłoby im optimum wskaźnika jakości. Zeby jednak jeden podsystem optymalizując swoje działanie nie psuł optymalności innych podsystemów, ustanawia się nad nimi wyższy hierarchicznie system koordynacji,

który tak zmienia podsystemom kryteria optymalności, aby te, dążąc do ich optymalizacji, nie szkodziły sobie wzajemnie.

Piramidę taką można ciągnąć dalej, gdyż to, co obecnie nazywaliśmy systemem wielkim /np.przedsiębiorstwo/, samo jest z kolei podsystemem jeszcze większego systemu /np.zjednoczenie/, a to jeszcze większego /np.gospodarka narodowa/ itd. Pamiętać tylko należy, że do optymalizacji całości systemu nie wystarczy sama optymalizacja podsystemów, niezbędna jest jeszcze na wszystkich etapach koordynacja.

TEST 1

P. Z czego wynika konieczność dekomponowania wielkiego systemu i optymalizowania go w strukturze hierarchicznej?

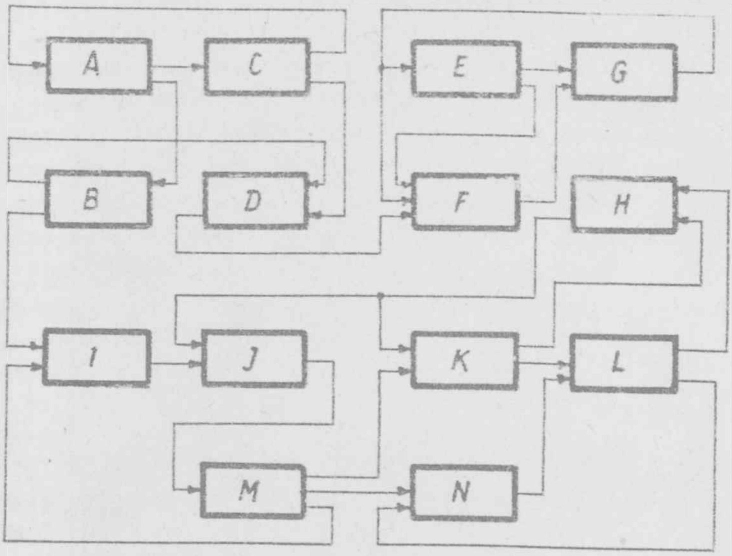
xxx

O. Z ogromnej ilości obliczeń, koniecznej przy centralnym ustalaniu decyzji sterujących dla wszystkich układów wchodzących w skład systemu. Struktura hierarchiczna zmniejsza niezbędną ilość obliczeń i pozwala wykonywać je w kilku ośrodkach zarządzania różnych szczebli równocześnie, w związku z tym możliwe się staje ustalenie decyzji o sterowaniu na bieżąco.
Przy nieprawidłowej odpowiedzi musisz jeszcze raz przeczytać punkty 1 i 2.

P. Jeżeli system ma strukturę przedstawioną na rysunku 67 to na jakie należałoby go rozbić podsystemy w razie konieczności zastosowania dekompozycji?

xxx

O. Podsystemy należy zgodnie ze wskazaniami punktu 4 dobierać tak, aby większość sprzężeń brała początek i kończyła się wewnątrz podsystemu, natomiast by minimalna ich ilość docierała do układów, zaliczonych przez nas do innego podsystemu. W tym sensie należy wybrać jako składniki pierwszego podsystemu układy A, B, C i D, drugiego pod-



Rys. 67. Dekompozycja systemu

systemu układy E, F, G i trzeciego podsystemu układy H, I, J, K, L, M i N. Podsystem pierwszy łączy wówczas z drugim i trzecim po jednym sprzężeniu, a układy systemów drugiego i trzeciego w ogóle nie kontaktują się.

Gdybyś miał wątpliwości, jeszcze raz przeczytaj punkty 3 i 4.

5. W rozdziale III, podr.1 wskazano na podstawową w teorii wielkich systemów trudność, polegającą na kłopotliwości określenia właściwości systemu w postaci modelu matematycznego. Uzyskanie potrzebnych transmitancji wiąże się zwykle z długotrwałym i kosztownym badaniem systemu, polegającym bądź to na obserwacji obiektu i po zgromadzeniu dostatecznie obfitego materiału statystycznego obliczaniu zależności regresyjnych /patrz test 1 w rozdziale III, podr.1/, albo, co jest jeszcze kosztowniejsze, na dokonywaniu celowych eksperymentów z syste-

mem.

. Ta ostatnia metoda poza kłopotliwością dokonywania eksperymentów na wielkim systemie ma jeszcze jedną poważną wadę: przeprowadzający eksperymenty z reguły nie wie, jaki charakter mają równania opisujące system. Natomiast z praktyki wynika, że zwykle są to zależności nieliniowe. Grozi nam więc niebezpieczeństwo, że system, który w warunkach swej normalnej pracy był stabilny, w wyniku eksperymentów może zostać wprowadzony w niestabilny obszar przestrzeni stanu, co zakończy się katastrofą /przypominamy, że w systemach nieliniowych, i tylko w takich zresztą, stabilność zależy od przesyłanych do systemu sygnałów/.

6. Natomiast wszystkie metody obliczeń sterowania optymalnego bezwzględnie wymagają dokładnej znajomości równań opisujących optymalizowany system. Jest to bardzo ważne i dlatego powtarzamy jeszcze raz: żadna metoda obliczania optymalnego sterowania nie jest wolna od pewnego niedostatku. Jeżeli wartości wstawiane do równań tej metody będą niepewne, to równie niepewne będą wyniki uzyskane z metody. Stosowanie algorytmów sterowania optymalnego do systemów, których opis matematyczny jest niepewny, jest wyjątkowo kosztownym i skomplikowanym sposobem uzyskiwania błędnych wyników.

Przykład 1

Powiedzmy, że chcemy optymalizować czas wykonania pewnego kompleksu operacji metodą Pert. Powinniśmy zatem ustalić sieć zależności kolejności wykonywania czynności. Jest to zadanie stosunkowo łatwe, jeśli jednak popełnimy tu błąd, cała dalsza analiza będzie bezwartościowa.

Powiedzmy jednak, że sieć została skonstruowana prawidłowo. Należy teraz określić czasy poszczególnych operacji. W tym celu powinniśmy przez długi czas obserwować działanie systemu, na podstawie zaobserwowanych wartości czasów określić średnie i dopiero rozpocząć właściwe obliczenia. Jest to kosztowne, gdyż po pierwsze angażuje pewną ilość personelu do obserwacji, a ponadto przez czas zbierania danych proces przebiega nieoptymalnie, na co my godzimy się, pomimo, że moglibyśmy go polepszyć. Cóż więc robimy? Zwykle pytamy ekspertów i na podstawie ich opinii ustalamy arbitranie czasy, stanowiące parametry w modelu sieciowym. O ile to jest dobre?

O tyle, że mamy wynik natychmiast. A o ile złe? O tyle, że jeśli ekspert się myli, to może nas to drogo kosztować, tym drożej, im bardziej się pomylił i im dłużej będziemy takie błędne "optymalne" sterowanie stosowali.

7. Jedyną radą na wskazane niedostatki jest koncepcja sterowania dualnego, sformułowana przez Feldbauma [13/]. Polega ona na tym, że równocześnie ze sterowaniem dokonujemy identyfikacji systemu. Mianowicie na początku, kiedy stajemy wobec konieczności sterowania systemu, o którym niewiele wiemy, wypracowujemy na podstawie nielicznych posiadanych danych, model systemu, o którym wiemy, że jest niedokładny, ale też nie traktujemy go zbyt serio. Dla tego "modelu pierwszego przybliżenia" opracowujemy sterowanie optymalne i zaczynamy je ostrożnie stosować. Sterowanie nasze wywołuje w systemie określony skutek, na ogół różna od tego, jakiego oczekiwaliśmy. Ta różnica pozwala nam poprawić nasz model ponieważ wiemy gdzie i o ile model różni się od rzeczywistości. W tym udoskonalonym modelu znowu obliczamy sterowanie optymalne, stosujemy je, stwierdzamy niezgodności itd.

8. Koncepcja ta jest chyba najlepszą, jaką można by zaproponować. Nie tracimy czasu na eksperymenty, ponieważ nasze sterowanie jest już eksperymentem. Równocześnie cały czas sterujemy nejlepiej, jak w danej chwili potrafimy i cały czas powiększamy swoją wiedzę o obiekcie.
Nazwa metody sterowanie dualne pochodzi od dwojakiego celu,

[13/] A. Feldbaum: Podstawy teorii optymalnych układów sterowania automatycznego, PWN, Warszawa 1967.

jaki przyświeca naszym oddziaływaniom na system optymalizacji i identyfikacji.

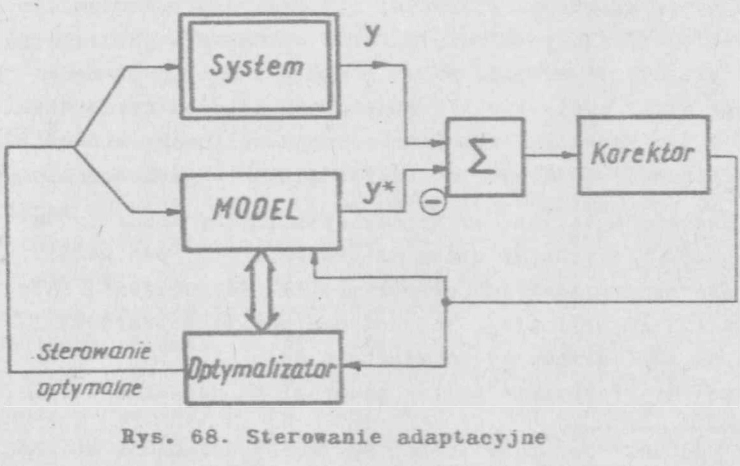
9. Od omówionej metody sterowania dualnego tylko krok dzieli nas od najbardziej zaawansowanej teorii cybernetycznej dotyczącej sterowania systemów-sterowania adaptacyjnego. System, którym sterujemy, nie pozostaje nigdy przez długi czas niezmiennym. Elementami systemu wielkiego są:

1/ urządzenia techniczne, które się zużywają lub zostają zastąpione przez nowe, sprawniejsze i bardziej wydajne,

2/ ludzie, którzy zdobywają kwalifikacje, lub przeciwnie, zaczynają mniej wydajnie pracować,

3/ zasoby surowców lub sytuacja rynkowa itd. Zatem nawet najdokładniej określony model matematyczny i najdoskonalsze sterowanie optymalne po upływie pewnego czasu dewaluuje się. Czy jest na to rada? Naturalnie, i to taka sama jak przy sterowaniu dualnym: model, który stanowi podstawę do obliczeń optymalizacyjnych musi ulegać zmianom, aby "gonić" zmienną rzeczywistość.

10. Koncepcja takiego zmiennego modelu przedstawiona jest na rysunku 68. Wypracowane sterowanie optymalne stosowane



Rys. 68. Sterowanie adaptacyjne

jest równocześnie w systemie, który reaguje na nie, przebiegiem wielkości wyjściowych Y, i w modelu, który daje przewidywany przebieg teoretyczny Y. Oba sygnały: ten prawdziwy, z systemu, oraz ten teoretyczny z modelu są odejmowane w sumatorze i stanowią podstawę pracy korektora, który w przypadku zaistnienia zbyt dużych zmian pomiędzy zachowaniem procesu i zachowaniem modelu zmienia charakterystyki modelu tak, aby zbliżyć je do charakterystyk obiektu.

Zapewne dostrzegłeś już Czytelniku, że jest to zwykły system ze sprzężeniem zwrotnym, lecz elementami jego są nie układy a systemy, i to nawet wielkie systemy.

11. Taki zmieniany na bieżąco model niejako adaptuje się do zmiennej rzeczywistości, stąd nazwa. Model jest w tym systemie stale dość bliski rzeczywistemu systemowi, w przeciwieństwie do sterowania dualnego, kiedy początkowo model jest bardzo odległy od rzeczywistości.

TEST 2

P. Jaka jest różnica powodów stosowania sterowania adaptacyjnego i dualnego?

xxx

O. Sterowanie dualne stosuje się w przypadku braku apriorycznej informacji o modelu matematycznym optymalizowanego systemu, natomiast sterowanie adaptacyjne znajduje zastosowanie w przypadkach, gdy dysponujemy już stosunkowo niezłym modele, a liczymy się tylko z możliwościami jego zmian pod wpływem trudnych do przewidzenia czynników.

P. Kiedy człowiek stosuje praktycznie zasadę sterowania dualnego?

xxx

O. Stosunkowo często, kiedy staje wobec nieznanej sytuacji, a musi wobec niej zachować aktywną postawę. Najjaskrawszy jest przykład pilota oblatywacza, który sterując optymalnie, boe inaczej nie wróci!/ nowym typem samolotu musi jednocześnie identyfikować jego właściwości lotne.

4. Teoria gier i podejmowania decyzji w sytuacjach konfliktowych

1. Dotychczasowe rozważania dotyczące optymalnego sterowania systemami wielkimi prowadziliśmy w oparciu o niesformułowane explicite założenie, że system poddaje się biernie naszym oddziaływaniom. Tymczasem, szczególnie w procesach ekonomicznych, systemy przejawiają często aktywność skierowaną przeciwko wysiłkom sterującego. Tak jest na przykład, jeśli za system uznamy gospodarkę wolnokonkurencyjną, a za cel sterowania sprzedanie maksymalnej ilości towarów na wolnym rynku. Wiadomo wówczas, że każde nasze sterowanie napotyka na czynne przeciwdziałanie konkurencji. Podobna sytuacja istnieje w przypadku sterowania zespołami wojsk w warunkach konfliktu zbrojnego.

2. Najprostszym przypadkiem sytuacji, w której sterowanie napotyka na aktywne i inteligentne przeciwdziałanie, jest gra. Rozpatrywać będziemy jedynie gry dwuosobowe /jako osoby można tu uważać dwa konkurujące zespoły/. Gry wieloosobowe, w których możliwe jest współdziałanie graczy i tworzenie koalicji nie będą rozpatrywane, są one po prostu dużo bardziej skomplikowane w analizie.

3. Podstawowe założenie jakie musimy zrobić, aby mieć możliwość przewidywania posunięć przeciwnika, jest następujące: przeciwnik jest inteligentny i życzy nam wszystkiego najgorszego, można więc być pewnym, że w każdej sytuacji wykona ruch, który postawi nas w najbardziej niekorzystnym położeniu. Będziemy się starali w tych najbardziej niekorzystnych warunkach uzyskać maksymalną wygraną. Strategia nasza będzie więc najlepsza w najgorszych warunkach, ale uwaga:

gdyby przeciwnik "patałaszył", można by wygrać więcej, stosując inną strategię, a więc w walce z nieinteligentnym przeciwnikiem nasza strategia nie jest optymalna.

4. Strategię opartą na powyższym założeniu, zapewniającą maksymalne minimum wygranej /minimum, bo przeciwnik nie da wygrać więcej/, nazywamy strategią MINI - MAX. Strategia ta nosi również nazwę strategii punktu siodłowego, a to z powodów, które staną się jasne po jej pełnym przedstawieniu. Zanim jednak przedstawimy metody wyboru strategii mini-max musimy określić kilka pojęć, aby ich potoczne znaczenia nie utrudniały Czytelnikowi śledzenia wykładu.

5. Strategią będziemy nyzywali ciąg ruchów, jakie mogą być wykonane przez danego gracza. Większość interesujących gier jest wielochodowa, tak więc strategia jest ciągiem ruchów, czyli w efekcie ciągiem decyzji podejmowanych na poszczególnych etapach gry. Jeżeli jednak zdecydujemy się ponumerować wszystkie możliwe strategie i rozpatrywać je wyłącznie jako całości, od pierwszego do ostatniego ruchu, to wprawdzie otrzymamy bardzo wiele takich strategii czystych, ale operować będziemy już wyłącznie numerami, a nie ciągami decyzji.

Rozpatrując grę jako zestawienie dwu strategii czystych /strategii pierwszego gracza i strategii drugiego gracza/, można każdej parze strategii jednoznacznie przypisać wynik gry, gdyż znając zasady gry i ciągi ruchów, jakie wykonują obaj gracze, można określić sytuację końcową gry, a wartość wygranej określonego gracza w oczywisty sposób zależy od sytuacji końcowej. Jeżeli przyjmiemy, że wartość wygranej określać będziemy dla pierwszego gracza, oraz że wygraną drugiego gracza przedstawimy w postaci ujemnej wygranej pierwszego gracza, to każdej parze numerów strategii odpowiada określona wartość tej wygranej. Nazywać ją będziemy wypłatą. Wypłata jest z matematycznego punktu widzenia funkcją dwu zmiennych: numeru strategii gracza I i numeru strategii gracza I i numeru strategii gracza II.

6. Dla tak postawionego zadania możemy zbudować tablicę której wierszami będą strategie czyste gracza pierwszego, a kolumnami strategie czyste gracza drugiego. Tablica taka ma więc n wierszy i m kolumn, jeśli gracz I ma n strategii, a gracz II m strategii.

Na przecięciu odpowiedniego wiersza i odpowiedniej kolumny wpiszemy w tabeli wartość /z uwzględnieniem znaku/ wypłaty odpowiadającej danej kombinacji strategii. Niech wartość ta na przecięciu i-tego wiersza i j-tej kolumny wynosi a_{ij}.

7. Strategię optymalną gracza pierwszego ustalimy w następujący sposób: gracz przewiduje, że jego przeciwnik zagra tak, żeby wygraną gracza pierwszego uczynić minimalną, będzie więc dobierał takie j /taką strategię/, aby uzyskać

$$\min_{j} a_{ij} \qquad /64/$$

Zapis wzoru /64/ należy czytać jako "minimum po j /po kolumnach/ z wartości wypłat a_{ij}".

Gracz I powinien wybrać swoją strategię /wiersz tabeli/ tak, aby dostać maksymalną z tych minimalnych wypłat:

$$\max_{i} \min_{j} a_{ij} \qquad /65/$$

Jeżeli analogiczne rozważania prowadzić będziemy z punktu widzenia drugiego gracza, to dojdziemy do reguły symetrycznej, a mianowicie:

$$\min_{j} \max_{i} a_{ij} \qquad /66/$$

8. Dochodzimy teraz do posiadającego wielkie znaczenie twierdzenia o punkcie siodłowym, sformułowanego przez jednego z pionierów cybernetyki, wspomnianego już w tej książce, J.von Neumanna: jeżeli dla danej gry istnieje rozwiązanie optymalne, będące strategię czystą dla gracza I, ustaloną na podstawie zasady /65/, to wartość wypłaty odpowiadającej temu rozwiązaniu, a wynikająca z wyboru /65/, jest równa war-

tości wypłaty przy dokonywaniu wyboru strategii czystej drugiego gracza na podstawie reguły /66/.

$$\min_j \max_i a_{ij} = \max_i \min_j a_{ij} \qquad /67/$$

Z twierdzenia tego wynika, że jeśli istnieją strategie czyste dla obydwu graczy, to wartość, jaką otrzymają gracze w formie wygranej po zakończeniu partii, będzie identyczna, niezależnie od tego, z punktu widzenia którego z graczy ustalana strategie. Warunek /67/ jest równocześnie warunkiem stosowalności strategii czystych. Jeśli on nie zachodzi, optymalnymi okazują się być strategie mieszane, czyli stosowanie kilku strategii z różnymi prawdopodobieństwami.

Wartość wypłaty, wynikającą z doboru strategii według zasady /65/ lub /66/, nazywamy wartością gry.

<u>Przykład 1</u>

Niech nasza gra będzie tego rodzaju, że pierwszy gracz, wybierając pierwszą strategię czystą, wygrywa 2 zł w przypadku, gdy drugi gracz wybierze pierwszą strategię, a w przypadku, gdy drugi gracz wybierze drugą strategię wygrywa 4 zł. Natomiast wybierając drugą strategię gracz pierwszy w przypadku wybrania przez gracza drugiego pierwszej strategii przegrywa 1 zł, a w przypadku wybrania przez gracza drugiego drugiej strategii wygrywa 8 zł. Tablica, obrazująca wartości wypłat ma postać:

$$\begin{bmatrix} 2 & 4 \\ -1 & 8 \end{bmatrix}$$

Obliczymy minimum po j /optymalne zagrywki gracza II/. Są to wartości 2 i -1. Maksimum tych minimów /wzór 65/ jest 2 i odpowiada pierwszej strategii gracza I i pierwszej strategii gracza II. Podobnie obliczając maksima po i, mamy 2 i 8. Minimalne z tych maksimów /wzór 66/ wynosi 2, co odpowiada strategiom jak poprzednio. Widać więc, że gra posiada punkt siodłowy i istnieją optymalne strategie czyste dla obu graczy.

TEST 1

P. Znaleźć wartość gry i określić strategie optymalne w grze opisanej tablicą:

$$\begin{bmatrix} 2 & 6 & 1 & -1 \\ 3 & 5 & 8 & 4 \end{bmatrix}$$

xxx

O. Minima po j wynoszą -1 i 3, ich maksimum zaś 3, co odpowiada drugiej strategii gracza pierwszego i pierwszej strategii gracza drugiego. Maksima po i wynoszą 3,6,8 i 4. Ich minimum jest znowu 3, a więc określony zestaw strategii jest optymalny, gdyż gra ma punkt siodłowy.

O ile zrobiłeś to zadanie, możesz pominąć następne i od razu przystąpić do studiowania punktu 9. Natomiast jeśli nie zrobiłeś go, musisz jeszcze raz przeczytać punkty 5,6,7 i 8, a także przykład 1, i musisz potem zrobić zadanie następne.

P. Określić strategie optymalne dla gry określonej macierzą:

$$\begin{bmatrix} 5 & 11 & 5 & 5 \\ 6 & 3 & 4 & 4 \\ 1 & 2 & 3 & 8 \end{bmatrix}$$

xxx

O. Minima po j; 5,3,1, ich maksimum 5, odpowiada strategii pierwszej dla pierwszego i pierwszej, trzeciej lub czwartej dla drugiego. Maksima po i: 6,11,5 i 8, ich minimum 5, pokrywa się z poprzednio określonym minimum dla pierwszego wiersza i trzeciej kolumny, zatem wskazane strategie są optymalne /65/ i /66/.

9. Strategie optymalne wyznaczone z zależności /65/ i /66/ przy warunku /67/ mają tę właściwość, że przy idealnie grającym partnerze każde odchylenie od tej strategii przynosi szkody. Na przykład obranie przez pierwszego gracza w przykładzie 1 strategii drugiej nieuchronnie prowadzi do przegranej - 1.

Natomiast partnerowi trzymającemu się sztywno strategii optymalnej żadne zmiany strategii przeciwnika nie mogą odebrać wygranej w wysokości wartości gry /a pewne zmiany, wynikłe z błędów partnera, mogą mu przynieść spory zysk/. Jednakże w przypadku gry z nieinteligentnym partnerem sztywne trzymanie się strategii optymalnej nie zapewnia korzyści, jakie można by odnieść wykorzystując sprzyjające okoliczności. Na przykład, gdy w pierwszym zadaniu testu gracz drugi nie zastosuje strategii pierwszej tylko drugą, to gracz pierwszy stosując strategię optymalną ma zapewnioną wygraną wysokości 5, czyli nawet więcej od wartości gry /wynik błędu partnera/, ale mógłby, odstąpiwszy od strategii optymalnej, wygrać 6.

10. Podsumujmy: strategia mini-max daje najlepsze wyniki w najgorszej sytuacji i gwarantuje we wszystkich innych sytuacjach wygrane przynajmniej równe wartości gry. Jednakże w przypadku pomyłek partnera strategia ta nie jest optymalna. I jeszcze jedna drobna uwaga: jak widać było z przykładów, gra tak zdefiniowana jak tutaj, nie musi być sprawiedliwa. We wszystkich rozpatrywanych przypadkach wartość gry była dodatnia, co oznacza, że nawet broniąc się optymalnie drugi gracz musi co najmniej tyle przegrać, ile wynosi wartość gry. Gra, która spełniałaby w naszym przekonaniu warunek sprawiedliwości, powinna być zatem z zerową wartością gry. Gry takie nie różnią się strategicznie od pozostałych gier, zostały one jednak nazwane grami o sumie zerowej.

TEST 2

P. Rozważmy konfliktową sytuację gospodarczą, polegającą na konieczności dokonania zakupu surowca do produkcji towaru, którego zbyt jest chwilowo nieznany. W przypadku, gdyby zbyt był mały /strategia pierwsza gracza II/, wystarczy 10 t surowca, gdyby zbyt był średni potrzeba 15 t, a gdyby był duży, aż 20 t. Jednocześnie jednak zmienią się ceny surowca: jeśli okaże się, że zbyt jest mały, pozostaną na obecnym poziomie 200 zł za tonę, jeśli będzie średni wzrosną do 300 zł za tonę, a przy dużym zbycie aż do 400 zł za tonę.

Gracz pierwszy /zaopatrzeniowiec/ ma trzy możliwe strategie: kupić 10 t surowca, kupić 15 t lub kupić 20 t. Gracz drugi ma też trzy możliwe strategie: utrzymać niski zbyt, ustanowić średni zbyt lub ustanowić duży zbyt. Określić tablicę gry i optymalną strategię pierwszego gracza.

xxx

O. Określimy elementy pierwszego wiersza tablicy, co odpowiada skutkom decyzji o zakupie 10 t surowca przy różnych warunkach koniunkturalnych. Jeśli zbyt będzie niski, wówczas koszt poniesiony przez zakład wyniesie a_{11} = 2 tys. zł /10 x 200/. Jeśli zbyt będzie średni zajdzie konieczność dokupienia 5 ton surowca w cenie 300 zł, co podniesie koszty do 3,5 tys. zł. I na koniec przy zbycie wysokim trzeba będzie zakupić 10 t po 400 zł, co da łączny koszt 6 tys. zł.

Prowadząc identyczne rozumowanie dalej określamy pozostałe elementy tablicy, która w całości wygląda:

$$\begin{bmatrix} -2 & -3{,}5 & -6 \\ -3 & -3 & -5 \\ -4 & -4 & -4 \end{bmatrix}$$

Wszystkie wartości są ujemne, ponieważ oznaczają "przegrane" /koszty/ gracza I. Łatwo stwierdzić, że tablica ta ma punkt siodłowy odpowiadający trzeciej strategii pierwszego gracza i trzeciej strategii drugiego gracza. Wartość gry wynosi wówczas -4 tys. zł.

P. Określić strategię optymalną dla gry opisanej tablicą:

$$\begin{bmatrix} 1 & 8 & 4 & 6 \\ 2 & 6 & 17 & 81 \\ 9 & 15 & 0 & 5 \\ -2 & 1 & 7 & 3 \end{bmatrix}$$

xxx

O. Gra nie posiada punktu siodłowego, nie ma więc strategii optymalnej w sensie strategii czystych. Istotnie minimum po j wynoszą odpowiednio 1,2,0 i -2, ich maksimum wynosi 2, natomiast maksima po i wynoszą 9,15,17,31, ich minimum wynosi zaś 9. Warunek /67/ nie jest spełniony, nie ma więc czystej strategii optymalnej.

P. Rozważmy sytuację konfliktową, w jakiej aktualnie się znajdujesz, Czytelniku. Masz do wyboru dwie strategie: czytać tę książkę dalej lub odłożyć ją i zająć się przyjemniejszymi rzeczami, np. pójść na spacer. W wyniku podjęcia decyzji o czytaniu masz zapewnioną dobrą ocenę na egzaminie, jeśli pytanie egzaminatora będzie dotyczyło partii materiału zawartych w dalszej części, oraz jedynie trudne do określenia korzyści intelektualne, jeśli pytanie będzie dotyczyło materiału już przerobionego. Powiedzmy równocześnie, że przyjemność spaceru jest trzy razy większa, niż przyjemność otrzymania dobrej oceny na egzaminie.
Co powinieneś wybrać?

xxx

O. Zadanie należy do typu zadań, w których nie ma podanych wartości "wypłat" w poszczególnych przypadkach, a jedynie ich względne wartości /spacer trzy razy przyjemniejszy niż dobra ocena/. Tablicę gry możemy jednak napisać, przyjmując, że przyjemność spaceru wynosi a, przyjemność dobrej oceny wynosi b /3b = a/, a korzyści odniesione z rozszerzenia wiedzy bez wpływu na wynik egzaminu wynoszą c. Tablica ma wówczas postać:

$$\begin{bmatrix} b & c \\ a-b & a \end{bmatrix}$$

Pierwszym graczem jesteś Ty, a drugim egzaminator, który ma dwie możliwe strategie:

1/ zadać pytanie z dalszych rozdziałów i

2/ nie zadawać pytania z dalszych rozdziałów.

Gra ma punkt siodłowy niezależnie od wartości c, jeśli tylko c /a - b/. Odpowiada on drugiej strategii czytelnika i pierwszej egzaminatora, wartość wypłaty jest wówczas a - b = 2b.

Zanim jednak pójdziesz na spacer, zwrócić Czytelniku uwagę na dwa fakty: uzyskany wynik jest rezultatem nie tyle teorii gier, co przyjętego sposobu wartościowania /spacer trzy razy milszy niż piątka/, teoria gier nie ponosi za to żadnej odpowiedzialności!

11. W pytaniach testu zetknęliśmy się z sytuacją, kiedy gra nie posiadała punktu siodłowego, mówiliśmy wówczas o konieczności wyboru strategii mieszanej. Strategia mieszana polega na stosowaniu /w tajemnicy przed przeciwnikiem/ na przemian kilku strategii prostych z określonymi prawdopodobieństwami. Wybór strategii musi być naprawdę przypadkowy, gdyż inaczej przeciwnik przejrzy naszą strategię i powiększy swoją wygraną stosując odpowiednie kontrstrategie.

Przykład 2

Rozważmy sytuację konfliktową, polegającą na tym, że dowódca armii musi zdecydować się na zaatakowanie jednego z dwu możliwych obiektów /2 możliwe strategie/, z których pierwszy jest 3 razy ważniejszy. Przeciwnik musi obiektów bronić, ale ma ilość sił pozwalającą na skuteczną obronę tylko jednego z nich.

Tablica gry ma więc postać :

$$\begin{bmatrix} 0 & 3 \\ 1 & 0 \end{bmatrix}$$

Gra ta nie ma punktu siodłowego:

$$\max_{i} \min_{j} a_{ij} = 0$$

$$\min_{j} \max_{i} a_{ij} = 1.$$

Naturalnie przyjęcie którejkolwiek strategii czystej jest niebezpieczne, gdyż o ile atakujący wybierze strategię pierwszą /atak na ważniejszą pozycję/ przeciwnik zgromadzi tam siły i wynik gry będzie zerowy. I odwrotnie, skupienie obrony na ważniejszym obiekcie sprawi, że przeciwnik zdobędzie łatwiejszy obiekt bez walki. Jedynym wyjściem jest stosowanie przez obie strony każdej ze strategii z określonym prawdopodobieństwem. Dla przypadku dwu możliwych strategii z teorii gier [14/] wynika, że należy stosować odpowiednie strategie z prawdopodobieństwami:

$$P_{1I} = \frac{a_{22} - a_{21}}{/a_{11} + a_{22}/-/a_{12} + a_{21}/} \qquad /68/$$

$$P_{1II} = \frac{a_{22} - a_{12}}{/a_{11} + a_{22}/-/a_{12} + a_{21}/} \qquad /69/$$

$$P_{2I} = \frac{a_{11} - a_{12}}{/a_{11} + a_{22}/-/a_{12} + a_{21}/} \qquad /70/$$

$$P_{2II} = \frac{a_{11} - a_{21}}{/a_{11} + a_{22}/-/a_{12} + a_{21}/} \qquad /71/$$

Na podstawie wzorów /68/ -/71/ z łatwością określimy, że pierwszy gracz powinien stosować strategię pierwszą z prawdopodobieństwem $P_{1I} = 1/4$, a strategię drugą z prawdopodobieństwem $P_{2I} = 3/4$, zaś gracz drugi strategię pierwszą z prawdopodobieństwem $P_{1II} = 3/4$, a strategię drugą $P_{2II} = 1/4$. Przy tak wybranych strategiach średni zysk gracza I będzie wynosił 3/4.

[14/] Patrz A.Pospiełow: "Gry i automaty",

12. W przypadku stosowania przez graczy strategii mieszanych można również mówić o wartości gry: jest nią wówczas średnia wartość wypłaty. Wartość ta jest już jednakowa z punktu widzenia obydwu graczy, na przykład w przykładzie 2 średnia strata gracza II wynosi też 3/4. Strategie mieszane pozwalają więc więc ekstrapolować pojęcia teorii gier z punktem siodłowym na gry bez tego punktu, jednakże w przypadku strategii mieszanych możliwa do przewidzenia jest tylko średnia wartość wypłaty, gdyż w każdej konkretnej rozgrywce może zajść inna kombinacja strategii obydwu graczy i wartości wygranych mogą się znacznie różnić.

13. Rozpatrzymy jeszcze przypadek tzw. dominowania. Zachodzi on wówczas, jeśli przy rozpatrywaniu gry o znacznej liczbie strategii pewne strategie od razu eliminujemy z rozważań, jako szczególnie niekorzystne.

Przykład 3

Niech gra będzie opisana tablicą:

$$\begin{bmatrix} 2 & -1 & 2 \\ 3 & 4 & 0 \\ 1 & 2 & -2 \end{bmatrix}$$

Łatwo zauważyć, że - niezależnie od strategii przyjętej przez gracza drugiego - trzecia strategia jest dla pierwszego gracza gorsza od drugiej, można więc w ogóle nie brać jej pod uwagę, gdyż nie ma ona żadnych szans, aby być wybraną jako optymalna. Podobnie strategia pierwsza jest dla drugiego gracza gorsza niż druga, nie będzie więc nigdy wybrana. Można więc, nie zmieniając wyniku gry, w ogóle nie brać wymienionych strategii pod uwagę i zapisać tablicę gry jako:

$$\begin{bmatrix} -1 & 2 \\ 4 & 0 \end{bmatrix}$$

TEST 3

0. Określ strategię optymalną dla macierzy gry z przykładu 3.

xxx

O. Gra nie posiada punktu siodłowego, stosować więc musimy strategie mieszane z prawdopodobieństwami określonymi wzorami /68/ - /71/.
Mamy więc:

$$P_{1I} = \frac{4}{7}, \quad P_{2I} = \frac{3}{7}, \quad P_{2II} = \frac{2}{7}, \quad P_{3II} = \frac{5}{7}$$

oraz oczywiście, ze względu na wyjątkowo niekorzystny wynik

$$P_{3I} = P_{1II} = 0$$

Jeżeli nie zrobiłeś tego zadania, musisz jeszcze raz przeczytać przykłady 2 i 3.

P. Określ strategię optymalną dla gry:

$$\begin{bmatrix} 3 & 1 & 3 \\ 2 & 5 & 5 \\ 2{,}5 & 2{,}5 & 5 \end{bmatrix}.$$

xxx

O. Strategia trzecia drugiego gracza może być usunięta jako wyjątkowo niekorzystna. Po jej usunięciu strategia trzecia gracza pierwszego jest zdominowana przez kombinację strategii pierwszej i drugiej. Pozostaje więc tablica

$$\begin{bmatrix} 3 & 1 \\ 2 & 5 \end{bmatrix}$$

Gra ta nie ma punktu siodłowego. Obliczając odpowiednie prawdopodobieństwa mamy:

$$P_{1I} = 3/5, \quad P_{2I} = 2/5, \quad P_{3I} = 0,$$

$$P_{1II} = 4/5, \quad P_{2II} = 1/5 \quad P_{3II} = 0.$$

Jeżeli pomyliłeś się, lub nie zrobiłeś tego zadania, powinieneś zapoznać się z rozdziałem 12 książki A.Lerner

"Zarys cybernetyki" [15/]

5. Podejmowanie decyzji na podstawie kryterium minimalnego ryzyka

1. Metody podejmowania decyzji, sformułowane w poprzednim rozdziale, opierały się bardzo silnie na założeniu o inteligentnym przeciwniku działającym na naszą szkodę. Założenie to nie zawsze można przyjąć. W szczególności jeśli naszym "przeciwnikiem" jest przyrodna, czyli jeśli mamy optymalizować np. proces wydobywania jakichś kopalin lub politykę zasiewów, to zakładanie istnienia tendencji do przeciwdziałania istniejących w przyrodzie jest nieuzasadnione. W większości takich przypadków zachowanie naszego przeciwnika wykazuje pewną regularność, możemy mianowicie określić z dużym poziomem ufności prawdopodobieństwa wystąpienia określonych reakcji.

2. Większość metod podejmowania decyzji w sytuacji, kiedy znane są rozkłady prawdopodobieństwa zachowania systemu w różnych warunkach, oparta jest na pochodzącej od T.Bayesa [16/] koncepcji minimalizacji średniego ryzyka błędu. Rozpatrzymy tu przypadek szczególny tego bayesowskiego podejścia, polegający na minimalizacji średnich kosztów podjętej decyzji, w przypadku kiedy ustalone są straty powstające w wyniku popełnienia określonych błędów oraz koszty poprawnego postępowania po podjęciu prawidłowej decyzji.

3. Niech system nasz scharakteryzowany będzie wartością jednej ze swych wielkości wyjściowych, a mianowicie wielkością X. System ten może znajdować się w jednym z dwu możliwych stanów - w stanie pracy lub w stanie awaryjnym. W stanie pracy system nie wymaga interwencji, natomiast w stanie awaryjnym niezbędne jest podjęcie środków zaradczych. Stan pracy

[15/] A. Lerner: Zarys cybernetyki, WNT, Warszawa 1971.

[16/] Wg Z. Hellwig: Elementy rachunku prawdopodobieństwa i statystyki matematycznej, PWN, Warszawa 1972.

charakteryzuje wartość wielkości wyjściowej X, leżąca w pobliżu pewnej wartości średniej $\bar{X}_1$. Ściślej można powiedzieć, że w stanie pracy systemu wartość X podlega rozkładowi normalnemu o wartości oczekiwanej $\bar{X}_1$ i dyspercji D /jeżeli niezbyt dokładnie pamiętasz, co to jest rozkład normalny, wartość oczekiwana oraz dyspersja, to koniecznie przed dalszą lekturą przypomnij to sobie np. z książki Z.Hellwig "Elementy rachunku prawdopodbieństwa i statystyki matematycznej"/[17]/

Jak wiadomo, gęstość prawdopodobieństwa wielkości X dla sprawnego systemu będzie można opisać wzorem:

$$f_1/X/ = \frac{1}{\sqrt{2\pi}\,D} \exp -\left(\frac{/X - \bar{X}_1/^2}{2\,D^2} \right) \qquad /72/$$

Fakt przypadkowego "rozmycia" wartości wielkości wyjściowej wynika z bezustannego oddziaływania na system zakłóceń i sygnałów sterujących, które również poza stanem systemu /sprawny - niesprawny/ mają wpływ na aktualnie obserwowaną wartość wielkości wyjściowej.

W przypadku, gdy system jest niesprawny, wartości wielkości wyjściowej X leżą w pobliżu innej wartości oczekiwanej $\bar{X}_2$, ale podlegają takiemu samemu rozkładowi normalnemu, z tą samą dyspersją D /przyczyny "rozmycia" wartości są w systemie uszkodzonym takie same, jak w nieuszkodzonym, więc charakter rozkładu musi być taki sam/.

Wartości wielkości wyjściowej podlegają więc w niesprawnym systemie rozkładowi o gęstości:

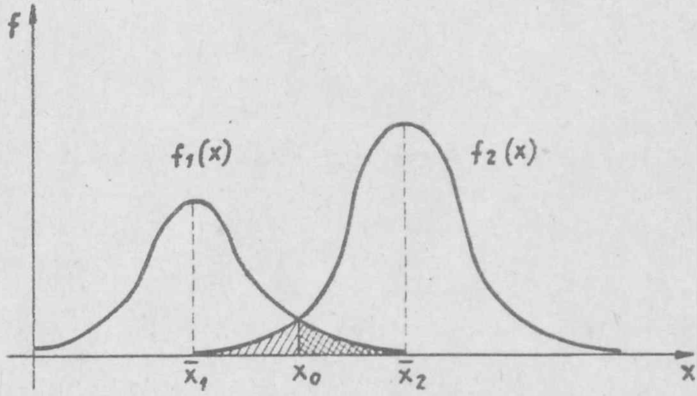
$$f_2/X/ = \frac{1}{\sqrt{2\pi}\,D} \exp -\left(\frac{/X - \bar{X}_2/^2}{2\,D^2} \right) \qquad /73/$$

4. Oba rozkłady narysowane są na rysunku 69. Widać, że istnieje określone, niezerowe prawdopodobieństwo, że stan

---

[17]/ Z. Hellwig: Elementy rachunku prawdopodobieństwa i statystyki matematycznej, PWN, Warszawa 1972 .

/wielkość X/ systemu sprawnego będzie leżał w pobliżu wartości charakterystycznych dla systemu niesprawnego lub odwrotnie. Jeżeli więc umówimy się podejmować decyzję, że system jest sprawny w przypadku, gdy wartość obserwowana X jest mniejsza od wartości X_0, a decyzję niesprawności w przeciwnym przypadku, to zakreskowane na rysunku /69/ pola obrazują prawdopodobieństwa popełniania błędów, polegających na zaliczeniu systemu wadliwego do systemów dobrych /kreskowanie pojedyncze/, albo systemu dobrego do złych /kreskowanie podwójne/.



Rys. 69. Podejmowanie decyzji w sytuacjach konfliktowych

Odpowiednie prawdopodobieństwa wynoszą więc:

prawdopodobieństwo przyjęcia systemu dobrego za dobry

$$P_{11} = \int_{-\infty}^{x_0} f_1/X/\, dX \qquad /74/$$

przyjęcia dobrego za złły

$$P_{12} = \int_{x_0}^{\infty} f_1/X/\, dX \qquad /75/$$

złego za zły

$$p_{22} = \int_{x_0}^{\infty} f_2/X/ \; dX \qquad /76/$$

i złego za dobry

$$p_{21} = \int_{-\infty}^{x_0} f_2/X/ \; dX \qquad /77/$$

5. Z każdym z tych czterech możliwych zdarzeń wiążemy określoną cenę: przyjęcie dobrego systemu za doby wiąże się jedynie z niewielkim kosztem C_{11}, odpowiadającym kosztom kontroli. Pomyłka polegająca na przyjęciu dobrego za zły wiąże się z dużymi kosztami niepotrzebnego zatrzymania dobrego agregatu i remontu C_{12}. Stwierdzenie, że system wadliwy jest zły pociąga za sobą koszty remontu C_{22}. Natomiast błąd niezauważenia awarii wiąże się z wielkimi kosztami katastrofy, jaką spowoduje nie wyłączony w porę system C_{21}.
Oczywiście powinno być:

$$C_{11} < C_{22} < C_{12} < C_{21} \qquad /78/$$

6. Zadanie optymalizacji polega teraz na znalezieniu takiej wartości progowej X_0, która gwarantowałaby wartość średniego kosztu. Przyjmijmy, że prawdopodobieństwo wystąpienia awarii urządzenia wynosi P. Oczywiście prawdopodobieństwo pracy bezawaryjnej wynosi /1-P/. Wówczas możemy zapisać, że średnia /oczekiwana/ wartość kosztów wynosi:

$$C_{śr} = /1-P/ \; /C_{11} \; p_{11} + C_{12} \; p_{12}/ + P/C_{22} \; p_{22} + C_{21} p_{21}/ \qquad /79/$$

7. Węzłowa koncepcja Bayesowskiej metody podejmowania decyzji polega na znalezieniu wartości progowej X_0, która minimalizowałaby wyrażenie /79/. Jak wiemy warunkiem na to, aby funkcja /zależność 79 jest funkcją granicy X_0, która występuje w całkach 74 - 77/ miała ekstremum, jest zerowanie się

pierwszej pochodnej.
Obliczamy więc

$$\frac{dC_{śr}}{dX_o} = /1-P/ \ /C_{11} \ f_1/X_o/ - C_{12} \ f_1/X_o/ \ / + P/C_{21} \ f_2/X_o/ -$$

$$- C_{22} \ f_2/X_o/ = 0 \qquad /80/$$

Przekształcając zależność otrzymamy:

$$\frac{f_2/X_o/}{f_1/X_o/} = \frac{/1-P/ \ C_{12} - C_{11}/}{P \ /C_{21} - C_{22}/} \qquad /81/$$

Należałoby jeszcze sprawdzić, czy warunek /80/ istotnie odpowiada minimum funkcji, ponieważ średnia cena rośnie zarówno przy zbytnim przesunięciu wartości X_o w lewo /wzrasta ilość błędów zaliczenia dobrego systemu jako wadliwego/, jak i w prawo /zwiększa się ilość błędów zaliczenia złego jako dobrego/, więc intuicyjnie jest jasne, że jedynym ekstremum tej funkcji jest minimum.

8. Ponieważ analizowanie złożonego wzoru /81/ jest zadaniem dość niewdzięcznym, uprościmy sobie zadanie, przyjmując, że koszty poprawnego stwierdzenia wadliwej pracy $/C_{11}/$ i koszty prawidłowo przeprowadzonego remontu $/C_{22}/$ są znikome w porównaniu z kosztami błędów. Mamy wówczas, po wstawieniu zależności /72/ i /73/ :

$$\frac{\exp -\left(\frac{/X_o - \bar{X}_2/^2}{2 \ D^2}\right)}{\exp -\left(\frac{/X_o - \bar{X}_1/^2}{2 \ D^2}\right)} = \frac{/1 - P/ \ C_{12}}{P \ C_{21}} \qquad /\bar{X}_1 = \bar{X}_2/ \qquad /82/$$

Po przekształceniu otrzymujemy ze wzoru /82/ zależność:

$$\frac{\bar{X}_1 + \bar{X}_2}{2} - \frac{D^2}{\bar{X}_1 - \bar{X}_2} \ \ln \frac{/1-P/ \ C_{12}}{P \ C_{21}} \qquad /83/$$

Wartość progowa X_1 obliczona ze wzoru /83/ charakteryzuje się tym, że jeśli tylko założenia co do cen są poprawne, to nasze orzekanie o stanie systemu przez porównywanie obserwowanych wartości X z progiem X_0 zapewni minimalny koszt pracy systemu.

Rozdział IV

AUTOMATY SKOŃCZONE I ALGORYTMY

1. Automat skończony. Alfabety wejściowy i wyjściowy

1. Wrócimy znowu do omawiania pojedynczych układów. Podejście nasze będzie o tyle nowe, że zarówno o sygnałach wejściowych, jak i o sygnałach wyjściowych z układu założymy, że są ciągami symboli, a nie płynnymi zmianami.

Niech na i-tym wejściu do układu może pojawiać się wyłącznie jeden z symboli tworzących zbiór $\{X_i\}$. Na innym, j-tym wejściu, może pojawiać się wyłącznie jeden z symboli tworzących zbiór $\{X_j\}$. Zbiory $\{X_i\}$ i $\{X_j\}$ możemy nazwać odpowiednio repertuarami wejść i-tego i j-tego.

Możemy teraz rozpatrzyć zbiór wszystkich symboli, które mogą pojawić się na którymkolwiek wejściu $\{X\}$. Oczywiście zbiór $\{X\}$ jest sumą teoriomnogościową zbiorów $\{X_1\}$, $\{X_2\}$..., $\{X_n\}$:

$$\{X\} = \{X_1\} \cup \{X_2\} \cup \dots \cup \{X_n\}. \qquad /84/$$

Zbiory $\{X_1\}$, $\{X_2\}$, ..., $\{X_n\}$ oznaczają tu zbiory symboli wejściowych, występujących odpowiednio na wejściach 1, 2, ..., n. Tak zdefiniowany zbiór $\{X\}$ nazwiemy A L F A B E T E M W E J Ś C I O W Y M naszego układu.

2. W zupełnie analogiczny sposób można określić repertuary kolejnych wyjść $\{Y_1\}$, $\{Y_2\}$, ..., $\{Y_m\}$. ALFABETEM WYJŚCIOWYM nazwiemy obecnie zbiór $\{Y\}$ wszystkich możliwych symboli, jakie mogą się pojawić na którymkolwiek wyjściu. Dla układu o m wyjściach alfabet wyjściowy jest również sumą teoriomnogościową:

$$\{Y\} = \{Y_1\} \cup \{Y_2\} \cup \{Y_3\} \cup \ldots \cup \{Y_m\} \qquad /85/$$

3. W dalszym ciągu rozpatrywać będziemy układy, dla których zarówno alfabet wejściowy, jak i alfabet wyjściowy są zbiorami zawierającymi pewne s k o ń c z o n e ilości symboli. Naturalnie ilość symboli może być bardzo duża i inna w alfabecie wejściowym niż w wyjściowym, ważne jest, aby była skończona.

Układy operujące skończonym alfabetem wejściowym i skończonym alfabetem wyjściowym nazwiemy A U T O M A T A M I. S K O Ń C Z O N Y M I.

Automat skończony cechuje się wyjątkową łatwością analizy. Zamiast zapisu zależności sygnału wyjściowego od sygnału wejściowego w postaci transmitancji lub funkcji przejścia można podać po prostu tabelkę sygnałów wyjściowych, odpowiadających poszczególnym sygnałom wejściowym. Ilość wierszy takiej tabelki byłaby równa ilości różnych możliwych sygnałów wejściowych.

4. Ilość możliwych sygnałów wejściowych i wyjściowych w automacie skończonym nietrudno określić. Mianowicie, jeśli alfabet wejściowy $\{X\}$ zawiera l symboli, a ilość wejść jest n, to liczba wszystkich możliwych różniących się między sobą symboli jest l^n.

Przykład 1

Niech alfabet $\{X\}$ składa się z trzech symboli: A, B, C oraz niech automat posiada dwa wejścia. Ilość różnych kombinacji sygnałów wejściowych jest więc $3^2 = 9$. Wypiszemy je:

AA, AB, AC, BA, BB, BC, CA, CB, CC.

5. Naturalnie w przypadku, kiedy alfabet wyjściowy zawiera k symboli i układ ma m wyjść liczebność zbioru rozróżnialnych sygnałów wyjściowych jest k^m.

Automat, o którym dotychczas mówimy, przyporządkowuje wielkościom wejściowym /które są zespołami symboli występujących równocześnie na różnych wejściach/ wielkości wyjściowe, będące kombinacjami symboli występujących równocześnie na wyjściach. Dla uproszczenia nazewnictwa zespół symboli występujących w danej chwili na wejściu nazwiemy SŁOWEM WEJŚCIOWYM, a odpowiedni zespół na wyjściu S Ł O W E M W Y J Ś C I O W Y M.

Automat nasz po otrzymaniu określonego słowa wejściowego odpowiada wygenerowaniem odpowiedniego słowa wyjściowego.

6. Automat tak określony jest bardzo ubogi w możliwości. Jego stopień organizacji odpowiada organizacji automatów sprzedających po wrzuceniu pieniędzy określone przedmioty. Słowami wejściowymi takiego automatu są zespoły przedmiotów wrzucanych do wrzutni, a słowami wyjściowymi zespoły przedmiotów wydane na wylocie. Automaty takie są zbyt prymitywne, by się nimi zajmować na gruncie cybernetyki. Możemy je jednak uczynić nader interesującymi przez bardzo prosty zabieg: wprowadzimy pojęcie stanu wewnętrznego automatu.

7. Niech automat, o którym mówimy, ma p możliwych, rozróżnialnych stanów wewnętrznych. Teraz słowo wyjściowe, jakie wygeneruje automat otrzymawszy dane słowo wejściowe, może być różne, w zależności od tego, w jakim automat jest stanie. Konkretnieje można to sformułować mówiąc, że słowo wyjściowe jest funkcją /jest zależne/ od słowa wejściowego i aktualnego stanu. Ilość wierszy tabelki opisującej zachowanie automatu wzrosła teraz do $p.l^n$, ponieważ rozróżniane są nie tylko słowa wejściowe, ale i stany wewnętrzne, przy których zostały te słowa odebrane, a ponieważ każde słowo może być odebrane w każdym z p stanów, więc ilość rozróżnialnych zespołów słowo - stan wewnętrzny jest p razy większa niż ilość słów.

8. Funkcję pozwalającą na podstawie znanej wartości wejścia /znanego słowa wejściowego/ i znanego stanu określić nową wielkość wyjściową /słowo wyjściowe/ nazwiemy funkcją σ. Automat, w którym funkcja σ zależy tylko od słów wejściowych /rozpatrywany na początku/, nazwiemy automatem trywialnym

albo po prostu przekształtnikiem kodów. Nie będziemy się nimi chwilowo zajmowali.

Automat, w którym funkcja δ zależy wyłącznie od stanu wewnętrznego S, nazwiemy /od nazwiska jego badacza/ automatem Moore'a. Natomiast najpełniejszy automat, w którym δ jest funkcją zarówno stanu wewnętrznego, jak i słowa wejściowego, nazwiemy automatem Mealy'ego.

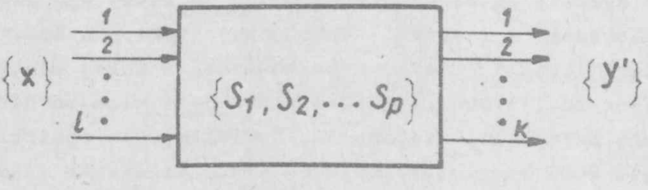
9. Automat powinien mieć możliwość zmiany swego stanu wewnętrznego, gdyż inaczej fakt, że posiada szereg stanów wewnętrznych, byłby tylko nieistotną ciekawostką, skoro i tak przebywałby stale w jednym stanie.

Zmiana stanu wewnętrznego powinna zachodzić na skutek pojawienia się określonego słowa wejściowego w czasie, gdy automat znajduje się w określonym stanie. Funkcję przyporządkowującą słowu wejściowemu i stanowi nowy stan, nazywa się funkcją λ.

10. Zatem bardziej abstrakcyjny opis automatu skończonego można podać definiując automat jako piątkę uporządkowaną :

$$A = \left\{ \{X'\}, \{Y'\}, \{S\}, \delta, \lambda \right\} \qquad /86/$$

gdzie $\{X'\}$ jest zbiorem słów wejściowych /utworzonych z symboli alfabetu $\{X\}$ /, $\{Y'\}$ zbiorem słów wyjściowych, $\{S\}$ zbiorem stanów wewnętrznych, a funkcje δ i λ zostały wyżej określone /rys. 70/.



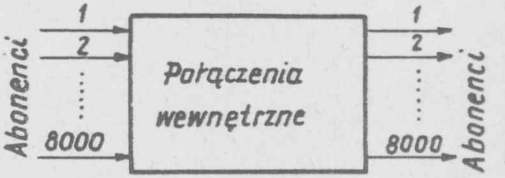
Rys. 70. Automat skończony

TEST 1

P. Czy centrala telefoniczna jest automatem skończonym?

xxx

O. Tak, jest to klasyczny przykład automatu skończonego. Teoria automatów skończonych pojawiła się zresztą właśnie w powiązaniu z konstruowaniem dużych sieci przełączających, a do takich należy w szczególności automatyczna łącznica telefoniczna /rys. 71/



Rys. 71. Łącznica telefoniczna

Aby dowieść jej skończoności należy kolejno rozpatrzyć zbiór słów wejściowych, wyjściowych i stanów wewnętrznych i o każdym z nich wykazać, że złożony jest ze skończonego zbioru elementów.

Zbiór słów wejściowych.

Centrala posiada tyle wejść, ilu abunentów obsługuje. Jest to oczywiście ilość skończona. Na każdym z wejść może pojawić się tylko jeden z możliwych symboli. Jeżeli na przykład centrala jest czterocyfrowa, to ilość symboli, jakie mogą pojawić się na dowolnym wejściu, wynosi 10 001 /10 000 symboli są to numery od 00-00 do 99-99 wykręcane przez abonenta, a 1 symbol reprezentuje fakt nie korzystania w danej chwili przez danego abonenta z danej centrali/. Ilość możliwych słów wejściowych jest więc bardzo duża, ale naturalnie skończona. /Powiedzmy, że centrala obsługuje 8000 abonentów. Wówczas ilość możliwych słów wejściowych jest zawrotna: 10001^{8000} 10^{32000}, nie ulega jednak wątpliwości, że jest to liczba skończona/. Ilość stanów wewnętrznych tego automatu jest również ogromna:

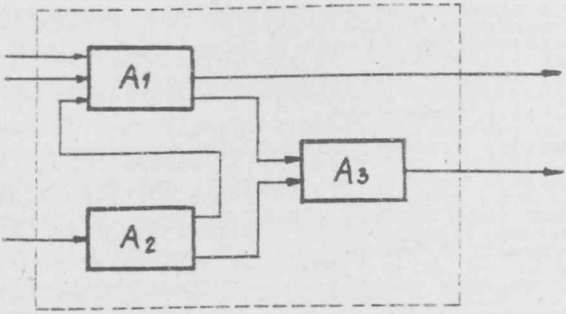
stanem wewnętrznym będzie stan połączeń aktualnie realizowanych przez centralę. Ilość takich różnych połączeń równa jest w naszym przykładzie ilości kombinacji dwuelementowych w zbiorze 8000 elementowym, czyli

$$C^{2}_{8000} = \frac{8000.7999}{2} \quad 32\ 000\ 000.$$

Za to ilość słów wyjściowych jest względnie niewielka. Na każdym wyjściu, których jest tyle, ilu abonentów, może pojawiać się jeden z trzech sygnałów: sygnał zajęcia, sygnał wywołania i brak sygnału. Ilość słów wyjściowych jest więc równa 3^{8000} 10^{3820}.

Układ rozpatrywany jest więc automatem skończonym, gdyż zarówno zbiór słów wejściowych, jak i zbiór stanów wewnętrznych i zbiór słów wyjściowych są skończone. Funkcja decyduje o tym, czy przy danym stanie połączeń abonent po wykręceniu numeru usłyszy sygnał wywołania, czy sygnał zajęcia, zaś funkcja decyduje o połączeniu, jakiego dokona centrala po wykręceniu określonego numeru przy określonym stanie dotychczasowych połączeń.

?. Czy system zbudowany przez połączenie automatów skończonych będzie automatem skończonym? /rys. 72/



Rys. 72. System automatów

xxx

O. Owszem, ponieważ wejściem do systemu będą wejścia pewnych układów wchodzących w skład systemu, a te naturalnie spełniają warunki stawiane automatom skończonym, gdyż należą do takich automatów, stany wewnętrzne systemu będą wypadkową stanów wewnętrznych poszczególnych układów, a więc ilość ich będzie skończona, chociaż z reguły bardzo duża /ilość stanów wewnętrznych systemu jest iloczynem ilości stanów jego części składowych/. Wielkości wyjściowe systemu są oczywiście wielkościami wyjściowymi automatów składowych, są więc skończone /składają się ze skończonej ilości słów/.

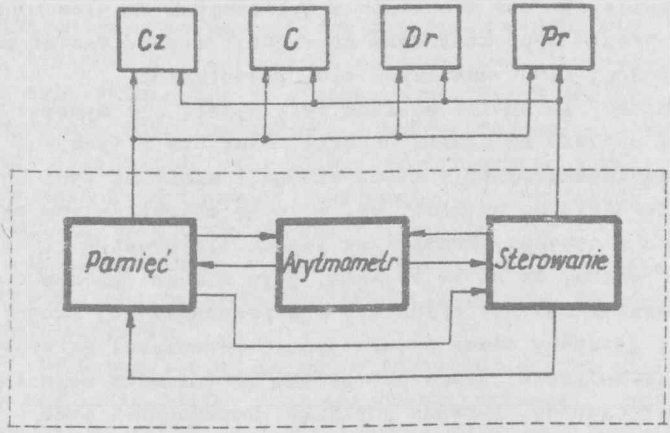
Jeżeli zasadniczy sens twoich odpowiedzi na pytania postawione w poprzednim i niniejszym zadaniu był zgodny z odpowiedzią podaną /nie musiałeś przeprowadzać samodzielnie całego wywodu, wystarczyło przekonanie/, to możesz czytać dalej. Inaczej przeczytaj jeszcze raz punkty 1 - 10.

P. Wykaż, że elektroniczna maszyna cyfrowa jest automatem skończonym.

xxx

O. Maszyna cyfrowa składa się z trzech zasadniczych układów: /rys. 73/ pamięci, arytmometru i sterowania. Wykażemy o każdym z tych układów, że jest automatem skończonym, a na podstawie poprzedniego zadania będziemy upoważnieni do wnioskowania, że cały system, jako złożony z automatów skończonych, jest skończony.

Pamięć maszyny cyfrowej posiada dwa wejścia. Jedno z nich przekazuje jedynie dwa sygnały: zapisać - odczytać. Repertuar tego wejścia składa się więc z trzech możliwych symboli: Z /zapamiętaj/, W /wydaj/ i O /nic nie rób/. Drugie wejście określa nadawcę wiadomości do zapamiętania lub odbiorcę wiadomości wydawanej; możliwe są więc na tym wejściu symbole: Cz /czytnik taśmy perforowanej lub kart jako nadawca wiadomości dostarczanej z zewnątrz/, C /konsola maszyny - dalekopis lub elektryczna maszyna do pisania, która może zarówno dostarczać danych do



Rys. 73. Maszyna cyfrowa

pamięci, jak i być ich odbiorcą/, A /arytmometr, do którego dostarczamy liczby z pamięci, lub z którego zapamiętujemy wyniki częściowe/, S /blok sterowania, któremu dostarczamy z pamięci kolejne rozkazy/, Dr /drukarka, na której wypisuje się wyniki/ lub Pr /perforator perforujący wyniki na taśmie/.

Czytelnikowi lepiej zorientowanemu w problematyce maszyn matematycznych należy się tu wyjaśnienie: wiadomo, że komunikacja maszyny z urządzeniami zewnętrznymi /czytniki, drukarki, perforatory/.odbywa się zwykle za pośrednictwem elementu układu arytmometru tzw. akumulatora. Jednak okoliczność ta w niczym nie zmienia istoty rzeczy, polegającej na wprowadzeniu i wyprowadzeniu danych do p a m i ę c i, a jedynie komplikuje opis. Do celów, które tutaj sobie stawiamy, w zupełności więc wystarczy uproszczony opis obejmujący fikcyjną, bezpośrednią komunikację pamięci z urządzeniami wejścia - wyjścia.

Z podanego opisu wynika, że pamięć jest układem, na którego wejściu może się pojawić co najwyżej 3.6 = 18 słów wejściowych /nie wszystkie z wyliczonych tu słów mają sens np. W Cz - wyprowadź dane na czytnik itp./. Pamięć ma więc skończoną ilość możliwych słów wejściowych.

Przyjmiemy, że pamięć posiada tyle wyjść, ile symboli mogło się pojawić na drugim wejściu. Niektóre z tych "wyjść/ są inwertowane: w rzeczywistości kanałami tymi dopływa informacja do pamięci, np. wyjście odpowiadające sygnałowi Cz - czytnik. Przyjmiemy jednak dla prostoty i zwartości opisu, że są to wyjścia, gdyż nie chodzi nam o opis działania maszyny cyfrowej, a o wykazanie jej skończoności; gdybyśmy nasze pseudowyjścia przenieśli na wejście, to skomplikowalibyśmy obliczenia ilości słów wejściowych i wyjściowych, podczas gdy fakt skończoności tych zbiorów słów takie przegrupowanie nie miałoby wypływu.
Każda informacja jest w pamięci przechowywana i przesyłana do /od/ innych bloków w postaci wyrażeń złożonych ze skończonej ilości jedynej i zer /liczby binarne/. W każdej pojedynczej informacji, niezależnie od tego, czy jest ona liczbą, rozkazem czy np. napisem, może być ściśle określona ilość miejsc, na których mogą być jedynki lub zera. Na przykład w maszynie cyfrowej "ODRA 1013" taka długość informacji wynosi 39 pozycji. Ponieważ na każdym miejscu może być jedynka lub zero, więc ilość różnych zestawień /ilość różnych symboli na rozważanym wyjściu/ wynosi $2^{39} \approx 10^{12}$. Przy ilości wyjść równej 6 daje to mniej niż 10^{13} różnych możliwych słów wyjściowych.
Jest to ilość znaczna, lecz naturalnie skończona.
Ilość stanów wewnętrznych pamięci określimy równie łatwo: pamięć zawiera określoną, zwykle dość dużą ilość miejsc, tzw. komórek, w których mieścić się mogą pojedyncze informacje. Niech nasza pamięć zawiera 10 000 komórek. Stanem pamięci będzie określona zawartość wszystkich komórek. Ponieważ w każdej komórce może znajdować się jedna z 10^{12}

możliwych informacji /patrz wyżej/, więc stan pamięci może być podany na 10 000 . 10^{12} = 10^{16} sposobów. Tyle jest więc możliwych stanów.
Wykazaliśmy, że pamięć jest automatem skończonym.

Arytmometr jest układem maszyny, w którym wykonywane są wszystkie operacje arytmetyczne. Układ ten posiada dwa wejścia: jedno /od pamięci/, którym dostarczany jest argument /liczba/ do wykonania operacji arytmetycznej i drugie wejście /od układu sterowania/, którym nadchodzi instrukcja, co należy wykonać.

Ilość możliwych symboli docierających pierwszym wejściem równa jest ilości możliwych kombinacji jedynek i zer, czyli około 10^{12}.

Ilość symboli na drugim wejściu równa jest ilości operacji, jakie zdolny jest dokonywać arytmometr. Nie należy sądzić, że są to wyłącznie cztery działania, repertuar możliwości arytmometru jest dużo bogatszy, w skład jego możliwości wchodzą również przesuwania liczby o określoną ilość pozycji /np. z liczby 123 robi się przesunięciem w lewo 123000, a przesunięciem w prawo 12/, badania parzystości, sprawdzanie tożsamości liczb itp. Ilość takich działań rzadko jednak przewyższa 20, taką więc wartość możemy założyć w obliczeniach. W efekcie mamy nie więcej niż 10^{13} słów wejściowych, a więc ilość skończoną.

Ponieważ wynikiem działania arytmometru jest rezultat, czyli liczba, więc istnieje tylko jedno wyjście arytmometru, na którym może pojawić się jeden z 10^{12} symboli, tyleż jest więc możliwych słów wyjściowych.

Stan wewnętrzny arytmometru utożsamić można z liczbą aktualnie się w nim znajdującą /arytmometr przy wykonywaniu operacji, np. dodawania, pobiera z pamięci tylko jedną liczbę, druga jest w nim uprzednio zmagazynowana/.
Ilość możliwych stanów wewnętrznych jest więc znowu równa 10^{12}.

Zajmiemy się z kolei blokiem sterowania. Jedynym jego wejściem jest połączenie z pamięcią, którym docierają rozkazy uprzednio zapamiętanego programu /pomijamy instrukcje przekazywane z pulpitu/.

Każda maszyna cyfrowa posiada określoną, skończoną, i na ogół względnie krótką listę rozkazów. Ilość tych rozkazów nie przewyższa zwykle kilkudziesięciu do kilkuset. Tyle jest więc symboli wejściowych dla układu sterowania. Ilość wyjść układu sterowania związana jest z rozbudową zaplecza maszyny, ogólnie można jednak liczyć, że sterowanie posiada wyjście do kontaktowania się z pamięcią /wyjście to określa numer komprki pamięci z której ma być pobrana informacja, lub w której informację należy zapisać/, wyjście do kontaktowania się z arytmometrem, komunikujące charakter wykonywanej operacji arytmetycznej, wyjście do urządzeń wejściowych /czytnika kart, taśmy,konsoli maszyny/, na którym pojawiają się polecenia wprowadzenia danych, oraz wyjście połączone z urządzeniami wyprowadzania danych /drukarka/, na którym pojawiają się rozkazy wydruku odpowiednich informacji z pamięci. Jeśli więc przyjmiemy, że rozważana maszyny posiada tylko jeden czytnik i tylko jedną drukarkę, to ilość wyjść bloku sterowania będzie wynosiła 4, a ilość słów wyjściowych nie przekroczy $10\ 000 \cdot 20 \cdot 2 \cdot 2 = 8 \cdot 10^5$, a więc ilość ta jest skończona.

Skrajnie ubogi jest też asortyment stanów wewnętrznych bloku sterowania. Można wyróżnić jedynie trzy interesujące stany:

1/ aktualnie wykonywana operacja została ukończona,

2/ aktualnie wykonywana operacja jeszcze trwa,

3/ maszyna jest zatrzymana.

Reasumując, możemy stwierdzić, że wszystkie układy wchodzące w skład systemu będącego maszyną cyfrową są automatami skończonymi, a więc i utworzony z nich system jest automatem skończonym. Dowód ukończono: maszyna cyfrowa jest automatem skończonym. Niezwykle ważne jest przy tym, że maszyna cyfrowa dysponuje jedynie skończoną /choć w miarę postępu techniki coraz pojemniejszą/pamięcią.

Ważność tego stwierdzenia docenimy dopiero wówczas, gdy poznamy inny system cybernetyczny, który będzie w zasadzie również automatem skończonym, ale dysponującym pamięcią nieskończenie wielką.

Twór ten nazwiemy maszyną Turinga i przytoczymy niezwykle ekscytujące twierdzenie, że maszyna ta potrafi wykonać każdą operację, którą da się ściśle określić zbiorem reguł postępowania. Ale to później, chwilowo jednak - kubełek zimnej wody; maszyna cyfrowa jest automatem skończonym, bardzo złożonym i ciekawym, ale podlegającym wszystkim ograniczeniom typowym dla automatów skończonych, w szczególności nie jest maszyną Turinga i nie ma dowodu, że zdolna jest wykonać dowolną, ściśle opisaną operację. Może ten zdecydowanie niepełny wywód będzie jakąś odpowiedzią na pytania w rodzaju: czy maszyny cyfrowe mogą przejawiać, po odpowiedniej rozbudowie i oprogramowaniu, złożone formy działania inteligentnego?

Zadziwiające jest, jak wiele optymistycznych opinii w postawionej kwestii wypowiadanych jest nawet przez ludzi posiadających znaczną wiedzę i praktykę w dziedzinie teorii i zastosowań maszyn cyfrowych. A tymczasem wystarczy sobie uświadomić, że nie wszystkie formy umysłowej działalności poddają się formalizacji i "ścisłemu opisowi", a nawet z tych, których opis jest możliwy, nie wszystkie leżą w zakresie możliwości maszyny cyfrowej /w przeciwieństwie do wspomnianej już maszyny Turinga/.

Wywód powyższy był bardzo obszerny i dosyć szczegółowy. Nie mamy prawa wymagać, aby Czytelnik przeprowadził go w pełni samodzielnie. Jeśli jednak znasz dziedzinę maszyn cyfrowych w sposób, który wystarczał do przeanalizowania postawionego w zadaniu pytania i udzielenia uzasadnionej odpowiedzi, a nie podjąłeś tego trudu, to zastanów się, czy przypadkiem nie traktujesz nauki zbyt lekko i nie starasz się iść drogą najmniejszego oporu. Uświadom sobie, że jest psychologicznie uzasadnioną prawidłowością zasada łatwego zapominania rzeczy, których przyswojenie nie wiązało się z pewnym minimalnym wysiłkiem.

Jeśli zastanowienie nad ostatnimi uwagami skłoniło Cię do przyznania im słuszności, postaraj się raz jeszcze przeanalizować prowadzony wyżej wywód, starając się wyszukać jego słabe strony i uzupełnić jego luki. Na przykład: jak sądzisz, czym byłyby wspomniane w punktach 8 i 9 funkcje δ i λ w poszczególnych układach maszyny cyfrowej? ...

P. Załóżmy, że rozpatrujemy automat o jednym wejściu i jednym wyjściu. Niech alfabet wejściowy /a tym samym i zbiór słów wejściowych/ składa się jedynie z dwu symboli: A i B. Niech automat nasz posiada trzy stany wewnętrzne: S1, S2 i S3. Alfabet wyjściowy składa się z pięciu symboli: 1, 6, 4, !, Q. Funkcje δ i λ zadane będą tabelką. Ile wierszy musi liczyć ta tabelka?

xxx

O. Ilość wejść n równa się 1, ilość symboli l = 2, a ilość stanów wewnętrznych p = 3. Ilość wierszy tabelki wynosi więc

$$p \cdot l^n = 3 \cdot 2^1 = 6$$

Jeżeli miałeś wątpliwości, to rozsądnie uczynisz, rozproszywszy je definitywnie przy pomocy powtórnej lektury punktów od 4 do 7.

P. Niech funkcje δ i λ określone będą następującą tabelą:

słowo wejściowe	stan	słowo wyjściowe	nowy stan
A	S1	1	S2
A	S2	0	S1
A	S3	Q	S3
B	S1	1	S3
B	S2	1	S1
B	S3	4	S2

Które kolumny tablicy odpowiadają funkcji δ, a które funkcji λ ?

xxx

O. Kolumny: pierwsza wraz z drugą oraz trzecia opisują funkcję δ, zaś kolumny: pierwsza wraz z drugą, oraz czwarta opisują funkcję λ. Nadmieniamy, że do poprawnej odpowiedzi potrzebna była znajomość materiału zawartego w punktach 8 i 9.

P. Dla omawianego w poprzednich dwu zadaniach automatu należy określić przebieg wielkości wyjściowej /sekwencję słów wyjściowych/, jeżeli w chwili, gdy automat był w stanie S3, podano sygnał A a następnie sekwencję:

A,A,A,B,A,B,B,B,A,B,A,B.

xxx

O. Dla przykładu przytoczymy całość rozumowania:
W chwili początkowej nadszedł sygnał A przy stanie S3, co spowodowało wystąpienie na wyjściu sygnału "Q" i utrzymanie stanu S3 /patrz - trzeci wiersz tabelki/. Trzy kolejne pojawiające się sygnały "A" spowodują /ze względu na nie zamieniający się stan/ te same reakcje, a zatem wygenerowany będzie ciąg słów wyjściowych "Q,Q,Q". Następnie, przy stanie S3 przychodzi sygnał B. Powoduje to wygenerowanie sygnału "4" i przejście do stanu S2.

W tym stanie sygnał A powoduje wygenerowanie "0" i przejście do S1, z kolei B powoduje na wyjściu "!" i stan S3, dalej B - "4" i S2, znów B - "1" i S1, teraz A, więc "1" i S2, następnie B, więc "1" i S1, A więc 1 S2, i na koniec B, dające "1" i S1.

Reasumując, automat wygenerował sekwencję słów wyjściowych: Q,Q,Q,Q,4,0,!,4,1,1,1,1.
i po zakończeniu działania znalazł się w stanie S1.

Dla większej przejrzystości będziemy sekwencje słów wejściowych i wyjściowych zapisywać razem, tak aby słowa występujące równocześnie występowały pionowo, jedne pod drugimi:

A A A A B A B B B A B A B
Q Q Q Q 4 0 ! 4 1 1 1 1 1

Przykład ten potraktuj raczej treningowo i jeśli nie poradziłeś sobie z nim samodzielnie, wystarczy jeśli dokładnie prześledziłeś rozwiązanie. Jeśli jednak rozwiązanie uzyskałeś niezależnie i samodzielnie, to możesz nie czytając następnego zadania od razu przystąpić do studiowania rozdziału IV podr.2.

P. Przyjmijmy, że mamy automat o dwu wejściach, trzech wyjściach i dwu stanach wewnętrznych. Stany, jak poprzednio oznaczymy S1 i S2. Alfabet wejściowy niech składa się z symboli "x" i "o", alfabet wyjściowy z symboli "T" i "M". Określić ilość możliwych słów wejściowych, ilość możliwych słów wyjściowych oraz ilość wierszy dla funkcji i .

xxx

O. Ilość słów wejściowych - 4, ilość słów wyjściowych - 27, ilość wierszy tabelki - 8.

P. Dla opisanego wyżej automatu funkcje δ i λ opisuje tabela:

słowo wejściowe	stan	słowo wyjściowe	nowy stan
xx	S1	TTT	S2
xx	S2	TRT	S1
ox	S1	RRR	S1
ox	S2	RTR	S1
xo	S1	MRR	S2
xo	S2	RRM	S2
oo	S1	MMM	S2
oo	S2	TRM	S1

Określić sekwencję słów wyjściowych opisywanego automatu w przypadku, gdy automat początkowo znajduje się w stanie S2 i otrzymuje sekwencję sygnałów:

oo ox ox xx xo oo xo xo xx xx

xxx

0. Wygeneruje on wówczas sekwencję słów wyjściowych:
TRM RRR RRR TTT RRM TRM MRR RRM TRT TTT,
a końcowym stanem automatu będzie S2.

Jeżeli ostatnie zdanie sprawiło Ci trudności, powinieneś w zasadzie jeszcze raz przeczytać cały rozdział i wnikliwie prześledzić rozwiązanie poprzedniego zadania.

2. Przedstawienie automatu w postaci grafu

1. Posługiwanie się przy badaniu automatów tabelką jest przy większych rozmiarach automatu niezmiernie niewygodne. Istnieje jednak pewna forma graficznego przedstawienia l o g i k i działania automatu /podkreślić należy, że nie ma to nic wspólnego z graficznym przedstawieniem s t r u k t u r y automatu/.

2. Posłużymy się w tym celu grafem, którego wierzchołkami będą stany automatu /na rysunku oznaczane kółkiem z wpisanym

wewnątrz numerem lub nazwą stanu/, a których gałęziami będą "ruchy" automatu pod działaniem określonych sygnałów wejściowych. Innymi słowy, gałęzie grafu ilustrują działanie σ i λ.

3. Z powyższego opisu wynika, że z każdego wierzchołka powinno w y c h o d z i ć tyle gałęzi, ile jest możliwych słów wejściowych /l^n/, a ze wszystkich wierzchołków /bo tyle jest stanów wewnętrznych/ wychodzi łącznie p.l^n gałęzi. Liczba ta jest nam już znana - tyle było wierszy w tabelce. Istotnie, tak musi być, gdyż każda gałąź opisuje "ruch" automatu, który znajdując się w określonym stanie otrzymał określone słowo wejściowe.

4. Każda gałąź grafu kończy się również w wierzchołku. Jest to mianowicie ten wierzchołek, który odpowiada stanowi, do którego przechodzi automat, jeśli - będąc w stanie odpowiadającym wierzchołkowi, z którego gałąź "wyrasta" - otrzymał sygnał odpowiadający danej gałęzi.

5. Podobnie jak wierzchołki gałęzie są również opisane i to podwójnie. Na każdej gałęzi musi być napisane, w jakich warunkach /przy jakim słowie wejściowym/ opuszczamy wierzchołek po tej właśnie gałęzi oraz jaki przy tym zostaje wygenerowany sygnał /słowo wyjściowe/.
Opis ten będziemy wykonywali w postaci ułamka:

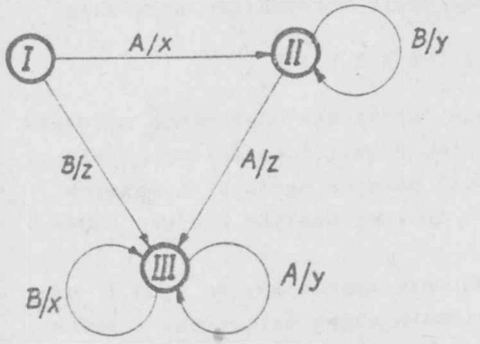
$$\frac{\text{słowo wejściowe}}{\text{słowo wyjściowe}}$$

Przykład 1

Rozważmy bardzo prosty automat o trzech stanach wewnętrznych /nazwiemy je I, II i III/, jednym wejściu z alfabetem A i B oraz jednym wyjściu z alfabetem X, Y, Z. Funkcje σ i λ podaje tabela:

słwo wejściowe	stan	słowo wyjściowe	stan
A	I	X	II
B	I	Z	III
A	II	Z	III
B	II	Y	II
A	III	Y	I·II
B	III	X	III

Konstruowanie grafu tego automatu rozpoczynamy od narysowania trzech kółek /rys.74/, w których wpiszemy numery stanów. Będą to wierzchołki grafu. Następnie, rozpatrując kolejno wiersze tabelki rysujemy gałęzie. Na przykład pierwszy



Rys.74. Graf automatu

wiersz tabelki odpowiada gałęzi wychodzącej z wierzchołka I i idącej do wierzchołka II, ponieważ wierzchołek II odpowiada stanowi po "ruchu" automatu. Na gałęzi tej wpiszemy "ułamek", w którego liczniku wpiszemy symbol wejściowy, pod działaniem którego dany "ruch" zachododzi, a w mianowniku wygenerowane przy tym słowo wyjściowe. W naszym przypadku będzie to A/X. Postępując analogicznie z pozostałymi wierszami tabelki wypełnimy cały graf.

6. Opis automatu w postaci grafu jest dużo wygodniejszy do analizy. Mając podaną sekwencję sygnałów wejściowych i stan wewnętrzny, z którego startujemy, możemy teraz bez

trudu określić sekwencję sygnałów wejściowych, "wędrując" po wierzchołkach grafu i zapisując kolejno "mianowniki".

Przykład 2

W automacie z poprzedniego przykładu aokreślimy sekwencję sygnałów wyjściowych, jeśli stan początkowy był I, a automat otrzymał słowa wejściowe: A, B, B, B. A, B, A, A, A, B, A. Startujemy z wierzchołka I i po gałęzi odpowiadającej słowu A wędrujemy do wierzchołka II. Powstaje przy tym sygnał X. Następnie zataczamy pętlę powracając do stanu II /pod działaniem słowa B/ i otrzymujemy słowo Y. Pętlę tę zataczamy jeszcze dwukrotnie otrzymując sygnały Y, Y. Następnie pod działaniem sygnału przechodzimy do stanu III /wygenerowane zostaje Z/ i już stanu tego nie opuszczamy poruszając się wyłącznie po pętlach. W rezultacie otrzymujemy sekwencję słów wyjściowych

X Y Y Y Z X Y Y Y X Y

7. Graf nie tylko ułatwia określenie zachowania automatu w odpowiedzi na konkretny ciąg sygnałów wejściowych, ale również pozwala na określenie pewnych ogólnych i ważnych właściwości automatu, które inaczej mogłyby być nie zauważone.

Popatrzmy na rysunek 74. Od razu zauważymy, że stan I nie może być w czasie pracy automatu nigdy osiągnięty - żadna gałąź do niego nie dochodzi, natomiast wszystkie odchodzą. Stan taki może być jedynie punktem startowym, i to natychmiast i na zawsze przez automat opuszczonym. Stan taki nazywać będziemy P A G Ó R K I E M G R A F U.

Z kolei stan III jest swoistą "pułapką" - automat po dotarciu do niego pozostaje w nim na zawsze, gdyż ze stanu tego nie ma gałęzi wychodzących. Nazwiemy go S T U D N I Ą G R A F U.

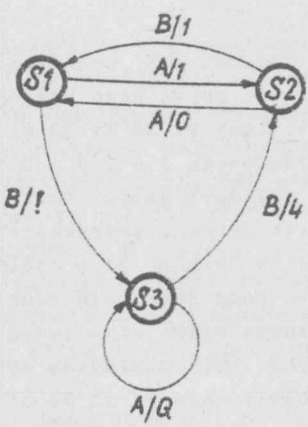
Stwierdzenie istnienia w grafie studni i pagórków pozwala sporo powiedzieć o zachowaniu automatu nawet bez znajomości sekwencji sygnałów wejściowych, które zostaną podane do automatu.

TEST 1

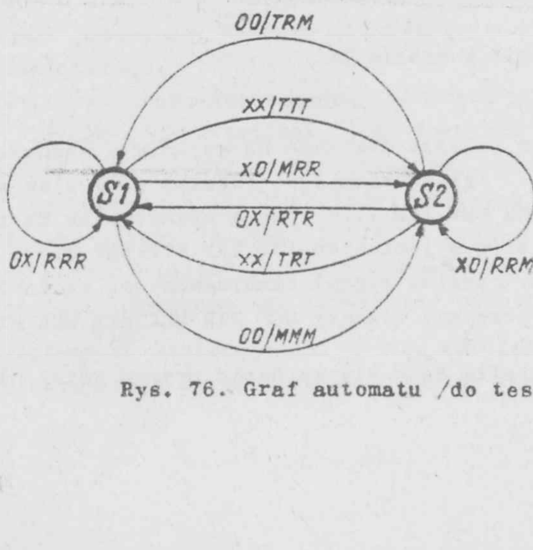
P. Narysować graf automatu opisanego w zadaniu testu 1 poprzedniego rozdziału na str. oraz automatu opisanego w zadaniu na str.

xxx

O. Graf dla automatu ze str. przedstawia rysunek /75/, a graf dla automatu ze str. rysunek /76/.



Rys. 75. Graf automatu /do testu 1/



Rys. 76. Graf automatu /do testu 1/

P. Czy grafy omawianych automatów zawierają studnie lub pagórki?

xxx

O. Nie, zawierają za to tzw. CYKLE, o których zaraz szerzej napiszemy. Zanim jednak przystąpisz do studiowania punktu 8 powinieneś zastanowić się, czy nie przeczytać jeszcze raz punktów od 1 do 7. W rozmyślaniach tych powinno pomóc Ci pytanie pomocnicze: czy zrobiłeś samodzielnie i poprawnie oba zadania testu?

8. Łatwo zauważyć, że automat z rysunku /75/ pod działaniem sekwencji AAAAAAAAA będzie /jeśli początkowo znajdował się w stanie S§/ kolejno przybierał stany S1 S2 S1 S2 S1 S2 ..., generując ciąg sygnałów wyjściowych 1 0 1 0 1 0 1 0 ... . Stan taki odpowiada w grafie układowi gałęzi tworzących zamknięty kontur, pozwalający na okrężną wędrówkę stale przez te same wierzchołki. Nazywamy to CYKLEM, gdyż cyklicznie powtarza się ten sam stan i ta sama sekwencja słów wyjściowych pod działaniem powtarzanego słowa wejściowego.

Graf 75 zawiera jeszcze jeden cykl: pod działaniem sygnału BBBBB ... powtarza się sekwencja stanów S1 S3 S2 S1 S3 ..., dając w efekcie sygnał wyjściowy ! 4 1 ! 4 1 ! 4 1 ...

TEST 2

P. Wskazać cykle w grafie 76.

xxx

O. Cykl tworzą gałęzie związane na wejściu z sygnałem oo oo oo oo oo ... /generowany jest wówczas ciąg słów wyjściowych MMM TRM MMM TRM .../ oraz z sygnałem xx xx xx xx .../ ciąg słów wyjściowych TTT TRT TTT TRT TTT .../. Cykl tworzy również sygnał kombinowany xo ox xo ox xo ox ..., generujący sygnały MRR RTR MRR RTR MRR RTR ... .

9. Gdy w grafie dają się wyróżnić przynajmniej dwa cykle

nie w pełni rozłączone /posiadające przynajmniej jeden wierzchołek wspólny/, to możliwe są rozmaite cykle kombinowane. Na przykład w grafie /75/ można również wskazać cykl związany z sygnałem BBABBABBA ... i dający sygnał wyjściowy ! 4 0 ! 4 0 ! ..., cykl AAABBAABBAA ..., cykl BBBAABBBAA ... i wiele innych. Nie będziemy się tym jednak bardziej wnikliwie zajmować.

10. Reasumując, reprezentacja automatu w postaci grafu daje w zasadzie najobszerniejszą i najbardziej poglądową informację o logice działania automatu. Ze względu na prostotę i względne ubóstwo możliwości automatu skończonego, wykonalne jest zbudowanie dla automatu w pełni izomorficznego modelu. Takim modelem jest właśnie graf automatu; znając graf potrafimy odtworzyć tabelę /funkcje i /, a znając tabelę możemy narysować graf. Jak się niebawem przekonamy, dla pewnych automatów możliwe jest również zbudowanie innego izomorficznego modelu w postaci pewnych funkcji logiki formalnej. Model ten daje się jednak zbudować jedynie dla automatów operujących dwuliterowym alfabetem wejściowym i wyjściowym. Musimy więc najpierw poświęcić nieco uwagi automatom binarnym.

3. Automat binarny

1. A U T O M A T E M B I N A R N Y M nazywać będziemy taki automat skończony, którego alfabet wejściowy i wyjściowy jest dwuliterowy.
Zgodnie z tradycją dwa symbole tego alfabetu nazywać będziemy 0 i 1. Ze względu na związek, jaki istnieje pomiędzy teorią automatów binarnych a logiką formalną, wygodnie będzie symbolom tym przypisać pewne wartości logiczne.

Umówimy się zatem, że tam, gdzie to będzie potrzebne, 0 oznaczać będzie "fałsz", a 1 = "prawda". Interpretacja ta jednak nie jest niezbędna dla rozważanej teorii.

2. Zaczniemy od stwierdzenia, że dla każdego automatu skończonego istnieje automat binarny zdolny realizować

wszystkie funkcje, które realizował automat pierwotny. Automat binarny będzie miał jednak z reguły większą liczbę wejść i wyjść niż automat wzorcowy.

3. Jak wiemy, ilość różnych słów wejściowych, jakie może odebrać automat o n wejściach i alfabecie wejściowym liczącym l symboli, wynosi l^n. Automat binarny, aby mógł robić to samo co omawiany automat, musi dysponować nie mniejszą ilością słów wejściowych. Alfabet tego automatu jest jednak wyłącznie dwuliterowy, tak więc jeśli przyjmiemy, że automat binarny posiada x wejść, to ilość jego słów wejściowych wynosi 2^x. Ze względu na konieczność spełniania tych samych funkcji co w automacie wzorcowym, musi być spełniona nierówność:

$$2^x \geqslant l^n \qquad /87/$$

którą łatwo przekształcić do postaci :

$$x \geqslant n \log_2 l \qquad /88/$$

Ponieważ z oczywistych względów x musi być liczbą całkowitą, więc należy wybrać najmniejszą liczbę całkowitą spełniającą zależność /88/.

<u>Przykład 1</u>

Niech automat, który chcemy zamienić na binarny, posiada 3 wejścia, każde o trzech możliwych symbolach /n = 3 i l =3/. Ilość możliwych słów wejściowych wynosi wówczas 27.

Zanim skorzystamy ze wzoru /88/, w celu określenia ilości niezbędnych wejść automatu binarnego, przystosujemy ten wzór do używania popularniejszych logarytmów dziesiętnych.

$$x \geqslant n \cdot \frac{1}{\log 2} \cdot \log l = n \cdot 3{,}32 \cdot \log l \qquad /89/$$

W naszym przypadku

$$x \geqslant 3 \cdot 3{,}32 \cdot \log 3 = 9{,}96 \cdot 0{,}4771 = 4{,}75 \approx 5$$

Nasz automat binarny będzie więc posiadał 5 wejść. Pozwala to wprawdzie na utworzenie aż 2^5 = 32 słów wejściowych, ale tylko niektóre z nich będą miały dla automatu znaczenie.

4. Po określeniu ilości wejść moeżmy określić zupełnie identycznym sposobem niezbędną ilość wyjść równoważonego automatu binarnego, wstawiając we wzorach /87/, /88/ i /89/ zamiast ilości wejść n, ilość wyjść m, a zamiast liczebności alfabetu wejściowego l, liczebność alfabetu wyjściowego k. Otrzymamy w ten sposób wzór:

$$y \geqslant m \,.\, 3{,}32 \,.\, \log k \qquad /90/$$

gdzie y jest poszukiwaną ilością wyjść automatu binarnego. Ilość stanów ewnętrznych w automacie binarnym jest taka sama jak w automacie wyjściowym.

5. Mając ustaloną ilość wejść i wyjść automatu binarnego możemy sporządzić "słownik" pozwalający przyporządkować słowa automatu wzorcowego słowom automatu binarnego i w ten sposób w pełni utożsamić ze sobą te dwa automaty.

Przykład 2

Napiszemy "słownik" pozwalający tłumaczyć słowa wejściowe automatu opisanego w poprzednim przykładzie na ich wersję binarną.

W tym celu przyjmiemy najpierw, że wspomniane w poprzednim przykładzie trzy symbole wejściowe są to x, y i z.

xxx	00000	xxz	00001	xxy	00010
xyx	00011	xzx	00100	yxx	00101
zxx	00110	xzz	00111	xzy	01000
xyz	01001	xyy	01010	zzx	01011
zzz	01100	zzy	01101	zyx	01110
zyz	01111	zyy	10000	zxz	10001
zxy	10010	yxz	10011	yxy	10100
yzx	10101	yzz	10110	yzy	10111
yyx	11000	yyz	11001	yyy	11010

Pozostało nie wykorzystanych 32 - 27 = 5 słów języka binarnego: 11011, 11100, 11101, 11110, 11111.
Słowa te dla naszego automatu binarnego są "bez znaczenia".

TEST 1

P. Określić liczbę wejść i wyjść automatu binarnego, który mógłby zastąpić automat z rysunku 75.

xxx

O. Niezbędna ilość wejść może być obliczona ze wzoru /89/, ale prościej jest zauważyć, że automat, o którym mowa, był od strony wejściowej binarny, gdyż rozporządzał dwuelementowym alfabetem wejściowym A i B. Wystarczy więc sporządzić "słownik" postaci:

A - 0 B - 1

i sprawa wejść jest załatwiona.

Dla obliczenia ilości wyjść musimy posłużyć się wzorem /90/:

$$y \geqslant 1 \;.\; 3{,}32 \;.\; \log 5 = 3{,}32 \;.\; 0{,}7 = 2{,}324 \approx 3$$

Sporządzimy prosty słownik wyjściowy:

1 000 0 001 ! 010 4 011 Q 100

Pozostają nie wykorzystane trzy słowa binarne /2^3 - 5 = = 3/: 101, 110, 111.

Jeżeli nie odpowiedziałeś prawidłowo, powinieneś jeszcze raz przeczytać punkty 3, 4 i 5.

P. Wykonać to samo dla automatu z rysunku 76.

xxx

O. Podobnie jak poprzednio, nie zachodzi potrzeba przeliczania ilości wejść, gdyż wejścia rozpatrywanego automatu były binarne. Wystarczy tylko zrobić przyporządkowanie x = 1, o = 0, aby można było od razu wypisać słownik:

oo 00 ox 01 xo 10 xx 11

Gdybyśmy zwracali uwagę na ilość wyjść i ilość symboli alfabetu wyjściowego, to musielibyśmy zastosować w automacie binarnym 5 wyjść /patrz przykład 2/. Jednakże efektywnie wykorzystywanych jest jedynie 8 słów wyjściowych /a nie 27/, więc wystarcza zastosowanie trzech wyjść automa-

tu binarnego, gdyż $2^3 = 8$.
Możemy więc napisać słownik:

RRR	000	TRM	001	TTT	010	MRR	011
RTR	100	TRT	101	MMM	110	RRM	111

Oba zadania przytoczone w teście wskazują, że zamiast mechanicznie stosować wzory, warto przemyśleć zagadnienie, gdyż można uzyskać tożsamość automatu binarnego i oryginalnego niejednokrotnie "tańszym kosztem", niż by to wynikało ze wzorów.

6. W dalszym ciągu rozpatrywać będziemy jedynie automaty binarne o jednym wyjściu. Wydaje się, że są to twory bardzo proste, zwłaszcza jeśli dodamy, że automaty te będą miały wyłącznie jeden stan wewnętrzny. Okazuje się jednak, że ilość możliwych różnych automatów tego rodzaju rośnie bardzo szybko ze zwiększeniem ilości wejść.

Ponieważ automat o n wejściach może rozróżniać 2^n słów wejściowych, a różne będą automaty, które przynajmniej jednemu z tych 2^n słów w inny sposób przyporządkowują 0 lub 1 na wyjściu, więc ilość różnych automatów będzie równa ilości różnych rozstawień dwu symboli w 2^n wierszach tabelki.
Ilość ta wynosi

$$A/n/ = 2^{2^n} \qquad /91/$$

Funkcja A/n/ bardzo szybko rośnie ze wzrostem n:

n	A/n/
1	4
2	16
3	256
4	65 536
5	4 294 967 296
6	18 446 744 073 709 551 616

Dla ilustracji wypiszemy wszystkie możliwe automaty dla n = 1 i dla n = 2. Automaty zadawać będziemy w postaci wierszy tabelki, określających sygnał wyjściowy automatu dla kombinacji sygnałów wejściowych wskazanej u góry kolumny

n = 1

symbol na wejściu	1	0
A_1	0	0
A_2	0	1
A_3	1	0
A_4	1	1

n = 2

symbol na wejściu 1	1	0	1	0
symbol na wejściu 2	1	1	0	0
A_1	0	0	0	0
A_2	0	0	0	1
A_3	0	0	1	0
A_4	0	0	1	1
A_5	0	1	0	0
A_6	0	1	0	1
A_7	0	1	1	1
A_8	1	0	0	0
A_9	1	0	0	1
A_{10}	1	0	1	0
A_{11}	1	0	1	1
A_{12}	1	1	0	0
A_{13}	1	1	0	1
A_{14}	1	1	1	0
A_{15}	1	1	1	1
A_{16}	0	1	1	0

Okaże się, że automaty takie, to znaczy binarne o jednym wyjściu i jednym stanie wewnętrznym można utożsamiać z pewnymi wyrażeniami algebry logiki. Zanim to jednak nastąpi musimy poznać kilka faktów czysto matematycznych.

4. Automaty logiczne i ich synteza

1. Logika matematyczna jest działem matematyki, operującym nie na liczbach, a na zdaniach. Zdanie może być prawdziwe lub fałszywe tylko takimi zdaniami logika operuje/. Fakt prawdziwości oznaczymy przez "1", a fałszu przez "0".
Podobnie jak liczby w klasycznej algebrze zastępuje się symbolami literowymi, tak i my zastąpimy zdania literami. Fakt, że jakieś zdanie, które oznaczymy przez A, jest prawdziwwe,

$A = 1$

i odpowiednio stwierdzenie, że zdanie to jest fałszywe przez

$A = 0$

2. Mając dane zdanie możemy łatwo utworzyć jego NEGACJĘ przez dodanie słowa "nie", na przykład ze zdania "Pada deszcz" utworzymy zdanie "Nie pada deszcz". Oczywiście, zdanie zanegowane jest prawdziwe wtedy i tylko wtedy, kiedy zdanie proste jest prawdziwe i na odwrót.

Zdanie zanegowane oznaczać będziemy tym samym symbolem co zdanie proste, ale z kreseczką u góry: "nie A" = $\bar{A}$.

Właściwości negacji, jako funkcji można przedstawić tabelką:

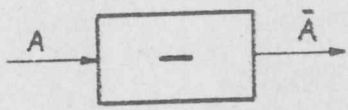
A	0	1
$\bar{A}$	1	0

Tabelka ta, zwana w literaturze "tabelą prawdy", pokazuje, dla jakich wartości zmiennej A /w górnym wierszu/ jej funkcja $\bar{A}$ przybiera wartości odpowiednio 0 i 1.

3. Funkcję negacji mógłby realizować prościtki automat binarny, który na wejściu miałby sygnał odpowiadający aktualnej wartości logicznej zdania A /prawda lub fałsz/, a na wyjściu dawałby sygnał będący wartością logiczną zdania zanego-

wanego. Tabelką takiego automatu może być tabela prawdy dla negacji czytana z góry na dół /kolumnami/.

Automat taki nazwiemy negatorem i oznaczymy graficznie jak na rysunku 77 .



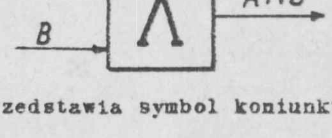
Rys. 77. Negator

4. Jeśli mamy dwa zdania, to możemy tworzyć ich kombinacje /będące również zdaniami/ używając różnych spójników. My rozważymy tylko dwa spójniki: "i" oraz "lub".

Jeśli dwa zdania połączymy spójnikiem "i", to zdanie złożone, które powstanie, będzie prawdziwe wtedy i tylko wtedy, kiedy obydwa łączone zdania będą prawdziwe. Uwidocznimy to w postaci tabeli prawdy:

A	0	1	0	1
B	0	0	1	1
$A \land B$	0	0	0	1

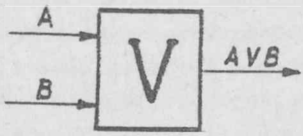
W najniższym wierszu tabeli użyto oznaczenia, jakie przyjmiemy dla operacji łączenia zdań spójnikiem "i", mianowicie będziemy zapisywać zamiast "A i B" $A \land B$. Operację tę nazwiemy KONIUNKCJĄ, a prosty automat, który posiada dwa wejścia i jedno wyjście, na którym pojawia się 1 lub zero zgodnie z przytoczoną tabelką, nazwiemy koniunktorem. Rysunek 78



Rys. 78. Koniunktor

przedstawia symbol koniunktora.

5. Operację łączenia zdań spójnikiem "lub" nazwiemy ALTERNATYWĄ. Zdanie złożone powstałe przy pomocy alternatywy jest prawdziwe, gdy chociaż jedno ze zdań składowych jest prawdziwe. Używając dla alternatywy oznaczenia /"A lub B" = $A \vee B$/, zapiszemy więc tabelę prawdy, która będzie równocześnie tabelką dla automatu zwanego alternatorem /rys. 79/.



Rys. 79. Alternator

A	0	1	0	1
B	0	0	1	1
$A \vee B$	0	1	1	1

TEST 1

P. Oznaczymy literami następujące zdania:

A = "Lato następuje po wiośnie"

B = "Wisła jest rzeką"

C = "Trawa jest czerwona"

D = "Lipiec ma 72 dni"

E = "Jakie to piękne"

Przypisać poszczególnym literom odpowiednie wartości logiczne.

xxx

O. A = 1, B = 1, C = 0, D = 0.

O zdaniu E nie można orzec, czy jest ono prawdziwe, czy fałszywe, zatem zdanie to nie może być zmienną logiczną, jest w pewnym sensie "symbolem nielegalnym", tak jak w zwykłej arytmetyce np.log/-3/.

Jeżeli nie odpowiedziałeś, musisz wrócić do początku tego rozdziału i przeczytać wszystko jeszcze raz.

P. Przyjmując, że A, B, C i D są zdaniami znanymi nam z poprzedniego zadania określić wartość logiczną formuł:

$\bar{A}$, $C \wedge D$, $A \wedge C$, $A \wedge \bar{B}$, $A \wedge B \wedge C$, $A \wedge B \wedge \bar{D}$, $A \vee C$, $\bar{A} \vee B$, $A \vee B \vee C$, $\bar{A} \vee \bar{B} \vee C \vee D$, $/A \vee B/ \wedge C$.

xxx

O. $\bar{A} = 0$, $C \wedge D = 0$, $A \wedge C = 0$, $A \wedge \bar{B} = 0$, $A \wedge B \wedge C = 0$
Do tego przypadku konieczne jest dodatkowe wyjaśnienie: jeśli mamy alternatywę lub koniunkcję większej ilości wyrazów /zdań/, to operacje realizujemy kolejno. W ostatnim przypadku najpierw rozstrzygnęliśmy, że $A \wedge B = 1$, a następnie utworzyliśmy koniunkcję tego zdania ze zdaniem C: $/A \wedge B/ \wedge C = 0$, ponieważ $C = 0$.
$A \wedge B \wedge \bar{D} = 1$, $A \vee C = 1$, $\bar{A} \vee B = 1$, $A \vee B \vee C = 1$,
$\bar{A} \vee \bar{B} \vee C \vee D = 0$, $/A \vee B/ \wedge C = 0$.

6. Zadanie testu wprowadziło nas w obręb złożonych zdań, będących skomplikowaną kombinacją zdań prostych, ich negacji oraz alternatyw i koniunkcji. O prawidłowości lub fałszu takich zdań najprościej orzec kolejno wstawiając wartości logiczne poszczególnych zdań i posługując się tabelami prawdy. Jeśli nie ma nawiasów, działania te wykonuje się w następującej kolejności: najpierw wszędzie tam, gdzie występują zmienne negowane, wstawia się /zgodnie z tabelką dla negacji/ zamiast zer jedynki i odwrotnie. Następnie wykonuje się koniunkcje, a dopiero na końcu alternatywy.

<u>TEST 2</u>

P. Sprawdzić prawdziwość zdań dla A = 0, B = 1, C = 1 i D= 0.
a/ $A \wedge \bar{C} \vee A \wedge \bar{B} \wedge C \vee D \wedge A$
b/ $A \vee (\overline{B \vee C}) \wedge C \vee (\overline{D \wedge C})$
c/ $A \vee B \vee D \wedge C \wedge (A \vee B \vee C)$

xxx

O. a - fałsz, b - prawda, c - prawda.

Jeżeli miałeś nieprzezwyciężone kłopoty, względnie jeśli nie rozwiązałeś poprawnie tego zadania, to dobrze by było, gdybyś najpierw zapoznał się z rozdziałem V.

7. Nie mamy powodów zadręczać Czytelnika mnożeniem coraz bardziej skomplikowanych przykładów zdań złożonych do sprawdzania, gdyż podręcznik ten nie ma na celu nauki logiki matematycznej. Chyba jednak każdy przyzna, że takie sprawdzanie jest uciążliwe i nieciekawe, zwłaszcza, że po zakończonej analizie może okazać się, że błędnie określiliśmy wartość logiczną któregoś ze zdań prostych i całą zabawę trzeba zaczynać od początku.

Tymczasem mamy przecież nasze automaty z rysunków 77 ; 78 i 79. Możemy, traktując te automaty jako układy, zbudować z nich system, który będzie automatem binarnym o wejściach odpowiadających zdaniom prostym i wyjściu odpowiadającemu wartości logicznej zdania złożonego.

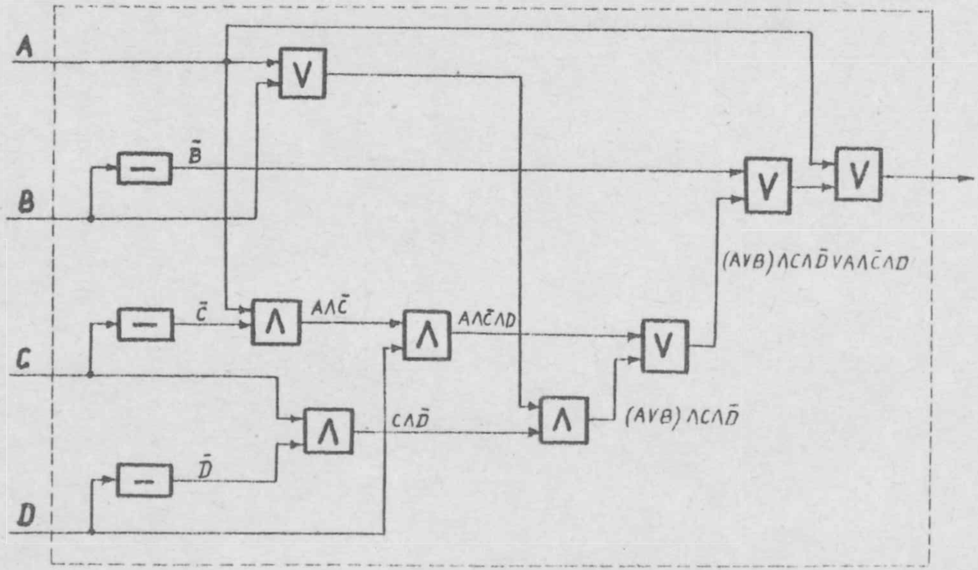
Przykład 1

Zbudujemy system, który będzie rozstrzygał o prawidłowości zdania złożonego:

$$A \vee \bar{\bar{B}} \vee C \wedge \bar{D} \wedge (A \vee B) \vee A \wedge \bar{C} \wedge D$$

Konstrukcję systemu rozpoczniemy od narysowania odpowiedniej liczby wejść /rysunek 80 lewa strona/. Przy wejściach oznaczamy odpowiadające im sygnały/ /wartości logiczne zdań A,B,C,D/. Następnie, zgodnie z kolejnością wykonywania działań, konstruujemy najpierw negacje, następnie koniunkcje i ewentualnie występujące w nawiasach alternatywy, a na końcu alternatywy bez nawiasów.

Dokładne prześledzenie rysunku 80 pomoże Ci zrozumieć zasady konstruowania takich systemów. Jesli przykład ten okazał się za trudny, możesz przeczytać łatwiejszy przykład 2, a następnie, po zrozumieniu metody konstruowania omawianych systemów, jeszcze raz powrócić do rysunku 80. Jeśli jednak w pełni zrozumiałeś ten przykład, przystąp bezpośrednio do zadań testu 3.



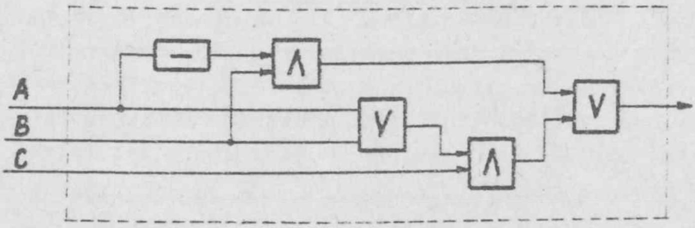
Rys. 80. Automat logiczny /przykład 1/

Przykład 2

Skonstruować system rozstrzygający o prawdziwości zdania

$$\bar{A} \wedge B \vee C \wedge (A \vee B)$$

Konstrukcję rozpoczynamy od realizacji występujących negacji. W rozważanym przykładzie zanegowana jest jedynie zmienna A, więc wystarczy jeden negator połączony z wejściem A /rys. 81/.



Rys.81. Automat logiczny /przykład 2/

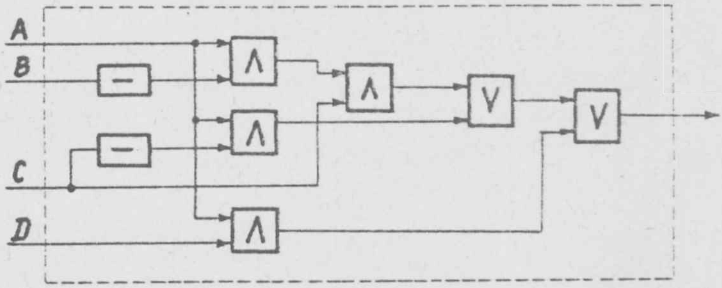
W drugiej kolejności realizujemy koniunkcję $\bar{A}$ B oraz występującą w nawiasie alternatywę A B. Wprowadzając wynik działania alternatora oraz zmienną C na wejście koniunktora trzeciej warstwy, otrzymujemy drugą występującą w zdaniu koniunkcję. Wyniki obu koniunkcji wprowadzamy na wejście alternatora, uzyskując wartość odpowiadającą wartości logicznej całego zdania złożonego.

TEST 3

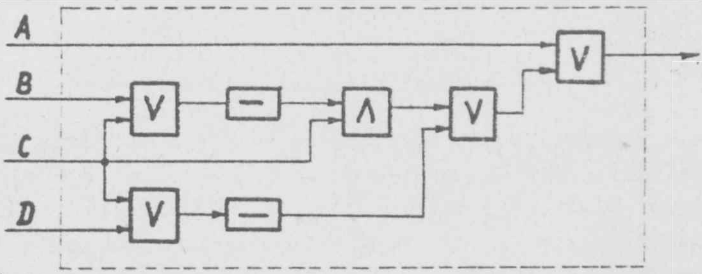
P. Narysować systemy pozwalające oceniać prawdziwość zdań a/, b/ i c/ testu 2/ str.

xxx

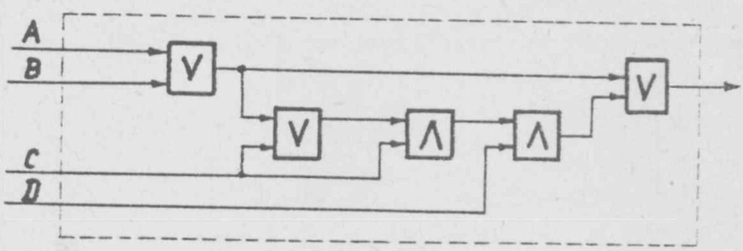
O. System dla zdania
a/ przedstawia rysunek 82, dla zdania
b/ rysunek 83, a dla zdania
c/ rysunek 84 .



Rys. 82. Automat logiczny /test 3/



Rys. 83. Automat logiczny /test 3/



Rys. 84. Automat logiczny /test 3/

Jeżeli nie tylko nie zrobiłeś samodzielnie tego zadania, ale w dodatku dalej nie rozumiesz metody budowania tego typu systemów, to masz zasadniczo trzy możliwości: albo spokojnie jeszcze raz przeczytasz oba przykłady i postarasz się prześledzić sposób konstruowania rysunków 82, 83 i 84.
Jeśli Ci się to uda - możesz czytać dalej.

Możesz również poszukać odpowiedzi w literaturze. Godna polecenia jest tu zwłaszcza książka J.Bromirski "Teoria automatów". Możesz jednak po prostu przejść nad swoim niepowodzeniem do porządku i czytać dalej, pomijając do końca rozdziału niejasne fragmenty, ponieważ rozdział ten można traktować nieco "nadobowiązkowo".

8. Dla automatów logicznych opis przy pomocy funkcji zdaniowej w sposób jednoznaczny określa strukturę i działanie automatu, możemy więc mówić, że funkcja zdaniowa stanowi izomorficzny model automatu, podobnie jak dający się łatwo sporządzić graf automatu. Jest to,dodajmy, najelegantszy i najoszczędniejszy model.

9. Funkcje zdaniowe /zdania złożone/ dają się przekształcać podobnie jak zwykłe formuły algebraiczne, choć według nieco innych reguł. Ponieważ każdej formie zdania złożonego będzie odpowiadał pewien automat, będący systemem złożonym z negatorów, koniunktorów i alternatorów, zachodzi istotne pytanie, czy ilość tych elementów, zużyta do budowy automatu według jednej formy funkcji zdaniowej, jest taka sama jak ilość elementów zużyta do budowy automatu według innej /przekształconej/ formy funkcji zdaniowej.

Okazuje się, że ilości te są różne, co więcej, odpowiednio przekształcając funkcję zdaniową można wielokrotnie zmniejszyć niezbędną ilość automatów - układów.

Problematyka wyboru funkcji zdaniowej, wymagającej do realizacji w postaci systemu minimalnej ilości elementów, jest rozległa i nawet przytaczanie tylko gotowych rezultatów zajęłoby zbyt wiele miejsca. Zainteresowanych odsyłamy więc do literatury, na przykład do cytowanej już wyżej książki J.Bromirskiego. Tutaj, jedynie informacyjnie, przytoczymy kilka nazw metod poszukiwania takiej minimalnej formy:

Całokształt teorii metod przekształceń formuł logicznych /w tym i interesujących nas funkcji zdaniowych/ nosi nazwę algebry Boole'a. Stosując metody tej algebry, możemy poszukiwać sposobu przekształcenia złożonej i wymagającej wielu elementów funkcji zdaniowej do postaci prostszej i tańszej w realizacji. Metoda ta jest jednak żmudna i mało efektywna. Do znajdowania bezpośrednio minimalnej formy funkcji zdaniowej służą diagramy Veitha i matryce Karnaugha. Metody te są wygodne, o ile zdanie złożone nie składa się z więcej niż 6 różnych zdań prostych /automat ma nie więcej niż 6 wejść/. Dla bardziej rozbudowanych form istnieją metody związane z nazwiskami Quine'a i McCluskey'a. Przegląd ten nie jest bynajmniej pełny, lecz sądzimy, że zainteresowany /i odpowiednio przygotowany/ Czytelnik w razie potrzeby dotrze i do pozostałych metod.

10. Następnym problemem, jaki nasuwa się przy studiowaniu zagadnień automatów logicznych, jest problem ilości elementów /automatów - układów/ niezbędnych do skonstruowania każdej, dowolnej funkcji zdaniowej. Dotychczas używaliśmy trzech: negatora, koniunktora i alternatora. Okazuje się, że można swobodnie z jednego z tych elementów zrezygnować, gdyż w algebrze Boole'a istnieją tzw. prawa de Morgana:

$$\overline{A \vee B} = \bar{A} \wedge \bar{B} \qquad /92/$$

$$\overline{A \wedge B} = \bar{A} \vee \bar{B} \qquad /93/$$

Prawa te można odczytać następująco: zanegowana alternatywa równa jest koniunkcji zanegowanych zdań, /prawo 92/, oraz zanegowana koniunkcja równa jest alternatywie zanegowanych zdań /prawa 93/.

Praw tych nie będziemy wyprowadzali, tylko je po prostu sprawdzimy:

Możliwe są cztery kombinacje wartości zmiennych A i B: 00, 01, 10, 11. Jeśli we wszystkich powyższych kombinacjach wartość logiczna lewej strony np. wyrażenia /92/ będzie równa wartości logicznej prawej strony, to wyrażenie to będzie zawsze prawdziwe /w logice wyrażenie takie nazywa się tautologią/.

Sprawdzamy: dla 00 lewa strona ma wartość 1 i prawa 1, dla 01 lewa strona ma wartość 0 i prawa 0, dla 10 lewa 0 i prawa 0, oraz dla 11 lewa 0 i prawa 0. Analogicznie można sprawdzić wyrażenie /93/. Jak z tego wynika możliwe jest zastąpienie koniunkcji przez alternatywę i negację lub alternatywy przez koniunkcję i negację, zatem do zbudowania dowolnej funkcji wystarcza koniunkcja i negacja lub alternatywa i negacja. Warto jednak dodać, że zmniejszenie asortymentu elementów prowadzi zwykle do zwiększenia ich ilości ponad tę, jaka byłaby konieczna do budowy systemu z wszystkich trzech układów.

11. Skoro można zbudować funkcję z dwu typów elementów, to warto zastanowić się, czy istnieją elementy tylko jednego typu /pełna inifikacja/, z których można by zbudować funkcje. Rozważania prowadzące do odpowiedzi na to pytanie są dość złożone, ale odpowiedź jest twierdząca: owszem, istnieją automaty, z których można zbudować dowolną funkcję zdaniową używając wyłącznie elementów jednego typu. Automaty te już poznaliśmy uprzednio, lecz wówczas nie znaliśmy ich znaczenia. Są to mianowicie automaty A_8 i A_{14} z tabeli na str. Automat A_8 nosi w literaturze miano modułu Peirce´a, bywa również nazywany NOR od angielskich słów NOt OR, gdyż automat ten odpowiada zanegowanej alternatywie, zaś automat A_{14} nazywany bywa modułem Sheffera lub NAND od Not AND.

Każdą funkcję zdaniową można więc zbudować z samych N O-R Ó W lub z samych N A N D Ó W, aczkolwiek wymaga to pewnej zręczności. Można by ponownie postawić zagadnienie minimalizacji. czyli poszukiwać postaci funkcji zdaniowej, pozwalającej na realizację systemu przy użyciu najmniejszej ilości

N O R Ò W lub N A N D Ò W. Nie będziemy się jednak zagadnieniem tym zajmować.

12. Dotychczas poznaliśmy metody realizacji w postaci systemu negatorów, koniunktorów i alternatorów dowolnej w zasadzie funkcji zdaniowej, która, jak wiemy, może być interpretowana jako izomorficzny model automatu.

W większości przypadków, kiedy w praktyce mamy do czynienia z koniecznością budowy automatu, wymagania co do jego działania zadane są nie w postaci funkcji, a w postaci opisu słownego. Na przykład chcemy, aby automat o trzech wejściach odpowiadał sygnałem "1" na wyjściu wtedy i tylko wtedy, gdy na wszystkich trzech wejściach pojawią się jednakowe sygnały. /Same "0" lub same "1"/.

13. Przedstawimy /bez podawania wyprowadzeń teoretycznych/ metodę budowania funkcji zdaniowej dla tak postawionego zadania. Najpierw sporządzimy tabelkę, w której uwidocznimy wszystkie możliwe słowa wejściowe automatu /będzie ich oczywiście przy n wejściach 2^n, a w naszym przypadku 8/, i dla każdego słowa wejściowego na podstawie słownego opisu ustalimy słowo wyjściowe. Dla rozważanego wyżej przykładu tabelka ta będzie miała postać:

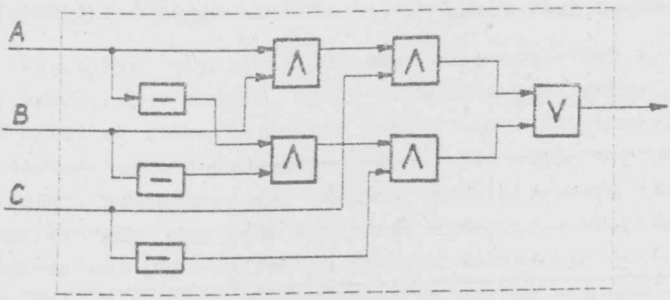
A	1	0	1	0	1	0	1	0
B	1	1	0	0	1	1	0	0
C	1	1	1	1	0	0	0	0
wyjście	1	0	0	0	0	0	0	1

Na podstawie takiej tabelki możemy przystąpić do konstruowania funkcji zdaniowej. Wybieramy z tabelki kolumny reprezentujące słowa wejściowe, dla których wyjście automatu przybiera wartość "1".

Dla rozważanego przykładu są to kolumny pierwsza i ostatnia. Następnie tworzymy koniunkcję wszystkich wejść /dla przykładu jest koniunkcja $A\wedge B\wedge C$ tylokrotnie, ile kolumn tabeli wyróżniliśmy /dla przykładu dwie/. W koniunkcjach tych zdania proste /wejścia/ występować będą w postaci prostej, o ile w rozważanej kolumnie dany sygnał wejściowy miał wartość "1", lub w postaci zanegowanej, gdy w kolumnie sygnał ten miał wartość "0". Dla przykładu z pierwszej kolumny otrzymamy więc koniunkcję $A\wedge B\wedge C$, a z ostatniej kolumny koniunkcję $\bar{A}\wedge\bar{B}\wedge\bar{C}$. Uzyskane w ten sposób koniunkcje /przypominamy, że jest ich tyle, dla ilu kolumn tabeli sygnał wyjściowy miał wartość "1" / łączymy znakiem alternatywy, uzyskując poszukiwaną funkcję zdaniową:

$$A\wedge B\wedge C\vee\bar{A}\wedge\bar{B}\wedge\bar{C}$$

Łatwo sprawdzić, że funkcja ta ma wartość "1" wtedy i tylko wtedy, kiedy wszystkie sygnały wejściowe są identyczne, a więc spełnia warunki postawione automatowi w punkcie 12. Funkcję tę z łatwością zrealizujemy w postaci systemu /rys. 85/, który będzie naszym poszukiwanym automatem.



Rys. 85. Synteza automatu

14. Konstrukcję funkcji zdaniowej można również oprzeć na kolumnach tabelki odpowiadających zerom na wyjściu, nie podajemy jednak tej metody wychodząc z założenia, że korzystniej będzie, gdy Czytelnik przyswoi sobie jedną metodę, ale dobrze, niż wiele słabo.

TEST 4

P. Sprawdzić metodą podstawiania wszystkich 8 możliwych kombinacji trzech sygnałów binarnych, że występujące formuły funkcji zdaniowych są równoważone:

a/ $A \wedge B \wedge C \vee \bar{A} \wedge \bar{B} \wedge \bar{C} \stackrel{?}{=} (A \vee \bar{B} \vee \bar{C}) \wedge (A \vee B \vee \bar{C}) \wedge (\bar{A} \vee \bar{B} \vee C) \wedge$

$\wedge (\bar{A} \vee \bar{B} \vee C) \wedge (A \vee \bar{B} \vee C) \wedge (\bar{A} \vee B \vee C)$

b/ $A \vee B \vee C \stackrel{?}{=} A \wedge B \wedge C \vee A \wedge B \wedge \bar{C} \vee A \wedge \bar{B} \wedge C \vee A \wedge \bar{B} \wedge \bar{C} \vee$

$\vee \bar{A} \wedge B \wedge C$

xxx

O. Obie formuły są równoważne, co dowodzi postawionej w punkcie 9 tezy, że odpowiednie przekształcenie funkcji zdaniowej może prowadzić do uzyskania jej postaci minimalnej, oraz, że proces ten jest wysoce opłacalny.

P. Zbudować automat, który dawałby sygnał "1" na wyjściu wtedy i tylko wtedy, gdy większość z trzech sygnałów wejściowych jest równa "1".

xxx

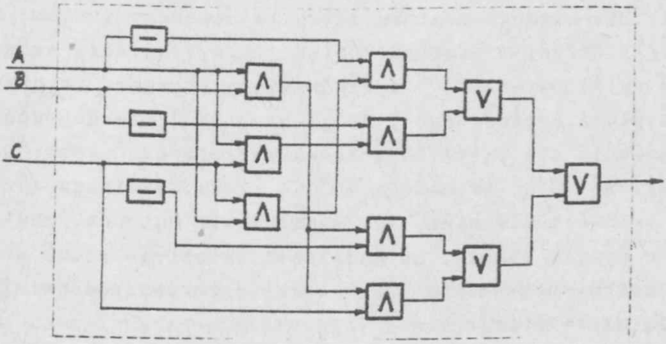
O. Tabelka ma postać:

A	1	0	1	0	1	0	1	0
B	1	1	0	0	1	1	0	0
C	1	1	1	1	0	0	0	0
wyjście	1	1	1	0	1	0	0	0

Odpowiadająca jej funkcja wynosi:

$$A \wedge B \wedge C \vee \bar{A} \wedge B \wedge C \vee A \wedge \bar{B} \wedge C \vee A \wedge B \wedge \bar{C}$$

a schemat automatu przedstawia rysunek 86.



Rys. 86. Synteza automatu /test 4/

4. Automaty probabilistyczne

1. Wszystkie dotychczas omawiane automaty miały wspólną cechę każdorazowo zespół: słowo wejściowe + aktualny stan w pełni określał słowo wyjściowe i nowy stan. Zależności te były zdeterminowane funkcjami i . Nie mogły się one zmieniać, można powiedzieć, że automat nie miał wyboru i musiał działać sztywno. Nie jest to jednak właściwość wspólna wszystkim automatom, gdyż automaty, których zachowania nawet przy znajomości sekwencji słów wejściowych nie da się przewidzieć.

2. Gdyby zachowanie takiego automatu nie podlegało zupełnie żadnym ograniczeniom, byłby to jednak po prostu generator sygnałów przypadkowych, a więc twór nieinteresujący.

Zajmiemy się więc przypadkiem pośrednim: automatem, który daje się opisać przy pomocy pewnych prawdopodobieństw zależnych od słów wejściowych i stanów wewnętrznych.

Dokładniej każdemu słowu wejściowemu i stanowi odpowiada określony rozkład prawdopodobieństwa słów wyjściowych.

Przykład 1

Niech automat posiada 3 słowa wejsciowe, 2 stany i 4 słowa wyjściowe. Dla każdego zestawu słowa wejściowego i stanu będziemy mogli określić prawdopodobieństwa wystąpienia wszystkich słów wyjściowych. Na przykład możemy przyjąć, że dla trzeciego słowa wejściowego i drugiego stanu, prawdopodobieństwo pojawienia się pierwszego słowa wyjściowego wynosi 0,5, słowa drugiego -0,3, trzeciego -0,1 i słowa czwartego też 0,1.

W tej sytuacji nie wiemy jak zachowa sie automat, znajdujący się w drugim stanie, pd nadejściu trzeciego słowa wejściowego, możemy stwierdzić jedynie, że najprawdopodobniej pojawi się słowo pierwsze.

3. Automaty probabilistyczne również opisuje się tabelką, jest ona jednak dużo większa niż w rozpatrywanych uprzednio automatach zdeterminowanych. Musi zawierać tyle samo wierszy co tabelka automatu zdeterminowanego, to znaczy przy n wejściach, l literach alfabetu wejściowego oraz p stanach $p \cdot l^n$. Zamiast jednak dwu kolumn /jednej dla nowego stanu i jednej dla słowa wyjściowego/ musi ona zawierać tyle kolumn, ile jest słów wyjściowych i stanów wewnętrznych łącznie.

Przykład 2

Rozważmy automat z poprzedniego przykładu, zakładając, że ma on słowa wejściowe A, B i C, stany wewnętrzne S1 i S2 i słowa wyjściowe oo, ox, xo i xx.

Narysujemy tabelkę opisaną w punkcie 3, wpisując na przecięciu wierszy i kolumn wartości prawdopodobieństw wystąpienia odpowiednich słów lub stanów /zgodnie z napisem u szczytu kolumny/ przy wystąpieniu słowa wejściowego i stanu wewnętrznego odpowiadającego wierszowi tabelki

słowo wejściowe	stan wewnętrzny	słowo wyjściowe				stan	
		00	0X	X0	XX	S1	S2
A	S1	0,1	0,1	0,2	0,6	0,2	0,8
B	S1	0,4	0,1	0,4	0,1	0,1	0,9
C	S1	0,0	0,0	0,5	0,5	0,7	0,3
A	S2	0,9	0,0	0,1	0,0	0,5	0,5
B	S2	0,5	0,1	0,1	0,3	0,2	0,8
C	S2	0,5	0,3	0,1	0,1	0,4	0,6

Łatwo sprawdzimy, że sytuacja opisana w pierwszym przykładzie odpowiada ostatniemu wierszowi tej tabelki.

4. Uważny Czytelnik zauważył już zapewne szczególną właściwość liczb wypełniających tabelkę: ich suma w każdym odcinku wiersza równa jest jedności. Musi tak być, ponieważ na przykład cztery pierwsze liczby pierwszego wiersza odpowiadają prawdopodobieństwom pojawienia się odpowiednich słów wyjściowych po pojawieniu się sygnału A przy stanie wewnętrznym S1, a przecież j a k i e ś słowo musi się pojawić, więc suma prawdopodobieństw musi równać się jedności. Podobnie suma prawdopodobieństw wystąpienia pierwszego i drugiego stanu /0,2 i 0,8 w pierwszym wierszu/ musi być równa 1.

5. Automaty, którymi dotychczas zajmowaliśmy się można również przedstawić w postaci podanej wyżej tabelki, w której wierszach zawsze tylko jeden element dla słów wyjściowych i jeden dla stanów byłby równy jedności, a wszystkie pozostałe byłyby zerami.

6. Algorytmy

1. W każdym systemie sterowania zbiera się, przetwarza i nadaje informacje. Przetwarzanie informacji polega na wykonaniu pewnych operacji w określonej kolejności. W wyniku wykonania ciągu operacji otrzymuje się rezultat lub rozwiązanie.

Ciągi czynności przedstawia się w tzw. algorytmach. W matematyce przyjmuje się, że algorytm jest ścisłym przepisem określającym proces obliczeniowy i prowadzącym do wykazujących zmienność danych wejściowych do poszukiwanego wyniku. Ścisła definicja algorytmu jest trudna do sformułowania.

Operacje występujące przy przetwarzaniu informacji w sytemach zarządzania i sterowania można podzielić na dwie klasy:

operacje arytmetyczne,
operacje logiczne.

Przy przetwarzaniu informacji operacje arytmetyczne i logiczne układa się w ściśle określonej kolejności.

2. W cybernetyce algorytm nazywa się zbiór reguł, instrukcji, opisów kolejnych czynności lub ograniczeń, które określają porządek wykonywania poszczególnych operacji w celu otrzymania pewnego wyniku.

Inaczej mówiąc algorytm jest zbiorem opisów prostych kroków i ścisłych reguł ich wykonywania. Jeśli w rezultacie zastosowania danego algorytmu do jakiegokolwiek obiektu, na podstawie zbioru danych wejściowych otrzymuje się rozwiązanie, to uważa się, że taki algorytm jest odpowiedni dla tego obiektu.

Definicja powyższa jest bardzo ogólna, ale dzięki temu można traktować jako algorytm zarówno wzór matematyczny, receptę produkcyjną lub instrukcję obsługi jakiegoś sprzętu, jak i na przykład program rozwoju organizmu zawarty w kodzie genetycznym.

TEST

P. Jak określa się pojęcie algorytmu w cybernetyce?

xxx

O. Algorytm jest to zbiór reguł, instrukcji, opisów kolejnych czynności lub ograniczeń, które określają porządek wykonywania poszczególnych operacji w celu otrzymania pewnego wyniku.

7. Algorytmy arytmetyczne i logiczne

7.1. A l g o r y t m   E u k l i d e s a

Jest to algorytm dla znajdowania największego wspólnego dzielnika dwóch dodatnich liczb całkowitych a i b. Algorytm ten zapiszemy w postaci kolejnych instrukcji :

1. Pobierz dane liczby a i b.
2. Porównaj obserwowane liczby /a = b; a > b; a < b/. Jeśli obserwowane liczby są równe, to każda z nich daje poszukiwany wynik. Jeśli nie, to przejdź do następnej instrukcji.
3. Jeśli pierwsza z obserwowanych liczb jest mniejsza od drugiej, to zamień je miejscami.
4. Odejmij drugą liczbę od pierwszej i porównaj dwie liczby: odjemnik i różnicę. Przejdź do instrukcji 2.

TEST

P. Znaleźć największy wspólny dzielnik algorytmu Euklidesa dla liczb 12 i 15

xxx

O. 3.

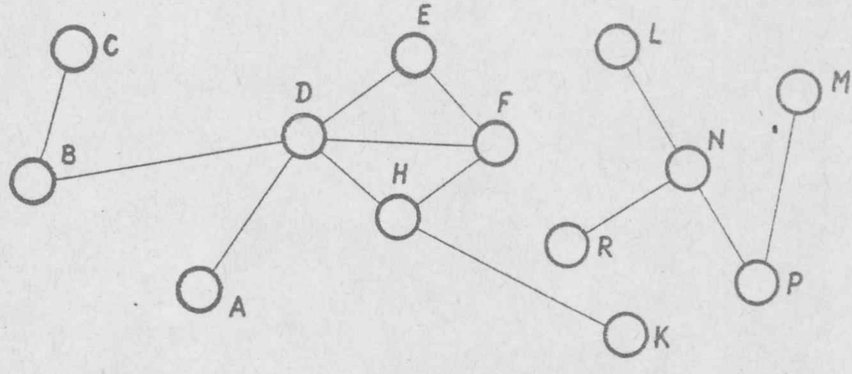
Jeśli rozwiązanie jakiegokolwiek zadania sprowadza się do działań arytmetycznych, to odpowiedni algorytm nazywa się arytmetyczny lub obliczeniowy. Taki właśnie jest algorytm Euklidesa.

Prócz arytmetycznych istnieją algorytmy logiczne.

7.2. A l g o r y t m   p o s z u k i w a n i a   d r o g i   w s k o ń c z o n y m   l a b i r y n c i e

1. Labirynt składa się ze skonczonej liczby pól, z których rozchodzą się korytarze. Każdy korytarz łączy dwa pola zwane sąsiadującymi. Pola, z których wychodzi tylko jeden korytarz

nazywa się ślepymi zaułkami. Labirynt geometryczny ma postać systemu kół A,B,C,D,..., wyobrażających pola, połączonych odcinkami linii prostych, przedstawiających korytarze /rys.87/.



Rys. 87. Labirynt skończony

Pole Y można osiągnąć z pola X, jeśli istnieje droga wiodąca przez pośredniczące korytarze i pola z X do Y. Przy tym X i Y mogą być albo sąsiadującymi polami, albo szeregiem sąsiadujących pól; X, X_1, X_2, ..., X_n, Y.

Jeśli pole Y w ogóle jest osiągalne z pola X, to jest ono osiągalne prostą drogą /bez pętli/. Na prostej drodze każde pole występuje tylko raz. Np. jedną z prostych dróg z punktu B do K jest BDHK.

Pole N jest nieosiągalne z punktu B.

TEST

P. Podaj przykłady arytmetycznych i logicznych algorytmów.

xxx

O. - Algorytm Euklidesa

- Algorytm poszukiwania drogi w skończonym labiryncie.

7.3. A L G O R Y T M   N A D   A L F A B E T E M

Alfabetem nazwiemy zbiór symboli, a algorytm będzie zbiorem reguł przekształcania zespołów takich symboli, które w skrócie nazywać będziemy słowami.

Przykład 1

Niech alfabet, nad którym określimy algorytm, składa się z symboli A,B,C,D. Regułą postępowania niech będzie: "jeśli występuje parzysta liczba jednakowych symboli, to należy je ze słowa usunąć, jeśli natomiast występuje nieparzysta ilość symboli, to należy za każdy symbol tej nieparzystej grupy podstawić symbol występujący w alfabecie wcześniej, np. za DDD podstaw CCC itp". Rozważymy efekt działania tego algorytmu na słowa: DDCBADABBBCD i AABCCCCDBCBBA.

Zaczynając od lewej strony najpierw usuwamy z pierwszego słowa grupę DD, następnie zamieniamy C na B, usuwamy grupę BB itd. otrzymując kolejno:

CBADABBBCD BBADABBBCD ADABBBCD ACABBBCD ABABBBCD AAABBBCD AAAAAACD CD BD AD AC AB AA i na koniec słowo całkowicie znikło. Postępując podobnie z drugim słowem otrzymamy:

BCCCCDBCBBA ACCCCDBCBA ADBCBBA ACBCBBA ABBCBBA ACBBA ABBBA AAAAA.

Tym razem słowo nie znikło, ale postępowanie nasze jest zakończone, ponieważ słowo,które powstało, nie daje się już zgodnie z przyjętymi regułami przekształcić.

3. W przykładzie 1 zastosowaliśmy pewien sposób stosowania reguł algorytmu nad alfabetem: mając słowo poszukujemy, poczynając od lewej strony, pierwszego zespołu symboli, jaki może być przekształcony przy pomocy reguł algorytmu. Taki pierwszy zespół nazwiemy pierwszym wchodzeniem algorytmu. Dokonujemy przekształcenia pierwszego wchodzenia i otrzymujemy nowe słowo. W tym słowie ponownie poszukujemy pierwszego wchodzenia itp.

4. W algorytmach, nie tylko w algorytmach nad alfabetem ,

mamy do czynienia z pewnymi kategoriami. Na przykład w algorytmie z przykładu 1 kategoriami były "parzysta ilość jednakowych elementów" i "nieparzysta ilość jednakowych elementów". Ogólnie K A T E G O R I Ą nazwiemy zbiór wszystkich zespołów symboli alfabetu, które podlegają przekształceniom zgodnie z jedną regułą algorytmu.

Przykład 2

Gramatyka dowolnego języka jest pewnym algorytmem, pozwalającym na tworzenie ze zbioru symboli alfabetu /symbolami tymi są słowa/, zespołów symboli /fraz, zdań, wypowiedzi/. Kategorią dla tego algorytmu są np. rzeczowniki, przymiotniki itp. Algorytm ten jest bardzo skomplikowany i zawiera wiele zakazów szczegółowych powodujących, że nie każde wyrażenie typu "przymiotnik + rzeczownik + czasownik" jest poprawną wypowiedzią. /Porównaj: "Mały kotek idzie" z "Czerwony dźwięk odpoczywa"/. Podobną strukturę mają gramatyki sztucznych języków, służących do programowania maszyn cyfrowych, takich jak ALGOL, są to jednak twory prostsze i bardziej spójne logicznie, a więc łatwiejsze do zinterpretowania jako algorytmy.

5. W przykładzie 1 tak się szczęśliwie złożyło, że proces przekształcania słów skończył się po skończonej ilości zastosowań algorytmu. Tak być wcale nie musi, gdyż wystarczy wziąć pod uwagę prosty algorytm nakazujący podstawić za A B, a za B A, by to "chodzenie w kółko" nie miało końca.

Aby tego uniknąć, poprawnie zbudowany algorytm powinien zawierać operacje końcowe, po których wykonaniu przerwany zostaje tok działania algorytmu.

Przykład 3

Niech algorytmem naszym będzie następujący zbiór prawideł: literami alfabetu niech będą A B i C. Jeśli w słowie występuje grupa liter jednakowych, należy je zastąpić pojedynczą literą /taką samą/, gdy natomiast litery są różne, należy dokonywać zamiany cyklicznej w grupie: A B C,B C A, C A B ... Przekształcić słowo: AAABCCCABC

Kolejne przekształcenia mają postać:

ABCCCABC BCACCABC CABCCABC ...

Widać, że nie docieramy nawet do "głębiej" położonych informacji, tylko obracamy się w koło. Jeżeli jednak uznamy, że przekształcenie cykliczne, w którym zespół liter niejednakowych pojawi się w ustawieniu identycznym jak przed przekształceniami, jest przekształceniem kończącym, to dokonamy jedynie czterech przekształceń.

8. Empiryczne własności algorytmów

1. Własność zdeterminowana. Metoda obliczeń /działań logicznych/ powinna być ścisła i w pełni zrozumiała. Żaden dowolny wybór nie jest dopuszczalny. Istotę metody można uogólnić /w dowolnej postaci/ w postaci skończonej liczby wskazań. Czynności przedsiębrane w związku z tymi wskazaniami stanowią zdeterminowany proces i nie zależą od woli działającej osoby. Ten proces może być w dowolnym czasie powtórzony przez drugą osobę.

2. Własność masowości. Algorytm daje możliwość rozwiązywania całej klasy zadań. Wskazania przedstawiające istotę algorytmu dotyczą początkowych danych, które mogą się zmieniać. Algorytm Euklidesa np. jest odpowiedni dla dowolnej pary liczb całkowitych $a > 0$ i $b > 0$;

Wzory do rozwiązania /systemu/ układu równań dają rozwiązanie przy dowolnych współczynnikach systemu; metoda poszukiwania jest odpowiednia dla dowolnego, dowolnie złożonego skończonego labiryntu itd.

Zadania z pewnej klasy uważa się za rozwiązywalne, jeśli dla tej klasy znaleziony jest algorytm rozwiązania.

Jeśli jednak dla rozwiązania wszystkich zadań danej klasy brak jest algorytmu, to /w szczególnych przypadkach/ trzeba szukać metod rozwiązania, które chociaż dają rozwiązanie w konkretnym przypadku, to jednak okazują się nieodpowiednie dla innych przypadków. Np. nie ma algorytmu pozwalającego dla dowolnego $n = 1,2,3, \ldots$ określić, czy równanie $X^n + Y^n = r^n$ posiada rozwiązanie będące liczbami naturalnymi. Dla konkretnych wartości n to zadanie może być rozwiązane. Przy $n = 2$ można dobrać trzy liczby /$x = 3$, $y = 4$, $z = 5$/ spełniające to równanie.

Dla $n = 3$ zostało dowiedzione, że to równanie nie ma rozwiązań w liczbach naturalnych. Dowód ten jednak nie jest odpowiedni dla innych n. Opis kolejnych czynności potrzebnych dla rozwiązania konkretnego zadania nie jest nazywany algorytmem.

3. Własność rezultatywności. Algorytm, zastosowany do dowolnego zadania określonej klasy, po skończonej liczbie kroków powinien doprowadzić do zatrzymania operacji. Po zatrzymaniu otrzymujemy wynik. Tę własność nazywa się niekiedy ukierunkowaniem algorytmu. Np. jeśli zastosuje się algorytm Euklidesa dla dowolnych dwóch liczb $a \geqslant 1$, $b \geqslant 1$, to wcześniej lub później nastąpi zatrzymanie pracy algorytmu i zostanie znaleziony największy wspólny dzielnik.

Algorytm poszukiwania w dowolnym i dowolnie złożonym, skończonym labiryncie musi doprowadzić do zatrzymania. Według tego na jakim polu nastąpiło zatrzymanie, można przeprowadzić dowód, czy poszukiwane pole jest osiągalne z pola startowego. Formalnie można stosować algorytm Euklidesa dla dowolnych liczb całkowitych $a \geqslant 0$, $b \geqslant 0$, a także dla liczb ujemnych. Może się jednak wtedy zdarzyć, że nigdy nie nastąpi zatrzymanie procedury algorytmicznej. Np. przy $a = 0$, $b = 6$ /największy wspólny dzielnik równy 6/, stosując instrukcje 1-5, otrzymamy pary: 0, 6; 0, 6; 0, 6; ... i tak do nieskończoności.

Z własności rezultatywności algorytmu wpływa pojęcie obszaru odpowiedniości algorytmu. Jest to największy obszar /zbiór/ początkowych danych, dla których algorytm jest rezultatywny. Jeśli warunki zostały wzięte z obszaru odpowiedniości, to algorytm przetwarza warunki w rozwiązanie zadania i następuje zatrzymanie z wydaniem wyników. Jeśli jednak początkowe warunki mieszczą się w obszarze odpowiedniości, to albo procedura algorytmiczna ciągnie się do nieskończoności, albo następuje zatrzymanie, lecz nie otrzymuje się wyniku.

Dla rozpatrzonych przykładów obszarami odpowiedniości są: zbiór liczb całkowitych dodatnich / 1,2,3, ... /dla algorytmu Euklidesa/, zbiór wszystkich skończonych labiryntów. Liczba operacji niezbędna dla tej lub innej procedury algorytmicznej wcześniej nie jest znana. Jest ona określana przez wybór początkowych danych. Wykonalność algorytmu w ogólnym pojęciu należy rozumieć jako proces potencjalnie możliwy. Rzeczywiście, dla pewnych konkretnych zadań z obszaru odpowiedniości algorytmu, w praktyce może nie starczyć miejsca, czasu arytmometru, lub pamięci przy realizacji algorytmu na EMC.

Omówione trzy ogólne własności algorytmów - własności zdeterminowania, masowości i rezultatywności /ukierunkowania/ - są własnościami empirycznymi. Zostały one sformułowane na podstawie doświadczeń, analizy wszystkich istniejących aktualnie algorytmów.

Naturalnie tych własności nie należy traktować w ścisłym, matematycznym znaczniu pojęcia "algorytm".

TEST

P. Jakie są empiryczne własności algorytmów.

xxx

O.-Własność zdeterminowania
-własność masowości /uniwersalności/,
-własność rezultatywności.

8. Elementy teorii algorytmów

1. Operatory alfabetyczne i algorytmy.

Alfabetem abstrakcyjnym przyjęto nazywać skończony zbiór obiektów nazywanych literami danego alfabetu. Obiekty te mogą być dowolnego rodzaju. Jako litery alfabetów abstrakcyjnych można rozpatrywać osobne znaczenia wielu wartości zmiennej

/parametru/ procesów technologicznych, określone stany obiektów sterowania, konkretne sytuacje przemysłowe, litery alfabetu jakiegokolwiek języka, cfyry, rysunki itp. Można rozpatrywać alfabet abstrakcyjny, którego literami są całe słowa tego lub innego języka. Podstawowe ograniczenie, nakładane przy określaniu, polega na tym, żeby alfabet był skończony, tj. żeby składał się ze skończonej liczby liter.

Słowem w alfabecie abstrakcyjnym jest dowolny skończony, uporządkowany szereg liter. Np. w alfabecie A = A/x,y/, składającym się z dwóch liter x i y, dowolne szeregi x, xy, yyx, xxxyy, ..., są słowami. Długość słowa zależy od liczby liter.

Prócz słów określonej długości /nie mniej niż jedna litera/ dla uogólnienia celowe jest rozpatrywać tzw. słowo puste, nie zawierające ani jednej litery. Dla oznaczenia pustego słowa można użyć literę 0 lub na odpowiadającym mu miejscu nie zapisywać żadnej litery.

Alfabet rozszerza się przez włączanie w jego skład nowych liter. Przy tym pojęcie słowa może się istotnie różnić. Np. wyrażenie 32 + 45 przedstawia dwa słowa /32 i 45/ w alfabecie A, złożone z 10 cyfr połączonych znakiem +. To wyrażenie można jednak rozpatrywać jako jedno słowo w rozszerzonym alfabecie A, które otrzymało się przez przyłączenie do niego nowej litery + .

2. Alfabetycznym operatorem lub odwzorowaniem nazywa się każdy sposób ustalania odpowiedniości pomiędzy porównywanym słowem w określonym alfabecie,a słowa w tym samym lub jakims innym alfabecie. Pierwszy alfabet nazywa się przy tym wejściowym a drugi wyjściowym alfabetem danego operatora. Jeśli alfabety wejściowy i wyjściowy pokrywają się /zgadzają się/, to operator alfabetyczny jest zadany w zgodnym /odpowiednim/ alfabecie. Będziemy rozpatrywać jednoznaczne operatory alfabetyczne tj. takie które przyporządkowują każdemu słowu w wejściowym alfabecie /słowu wejściowemu/ nie więcej niż jedno słowo w wyjściowym alfabecie /słowo wyjściowe/. Jeśli

alfabetyczny operator nie przypisuje danemu wejściowemu słowu R żadnego wyjściowego słowa /w tym również pustego/, to mówi się, że nie jest on określony dla tego słowa.

3. Polem określoności alfabetycznego operatora nazywa się zbiór wszystkich słów, dla których jest on określony. W ten sposób pod pojęciem operator alfabetyczny będziemy odtąd rozumieć jednoznaczne, ściśle określone odwzorowanie zbioru słów wejściowego alfabetu operatora w zbiór słów e jego alfabecie wyjściowym. Operatory alfabetyczne nie mogą być stosowane dla wszystkich słów, dlatego zawsze można uważać, że alfabet wejściowy i wyjściowy operatora odpowiadają sobie. W tym celu wystarczy ujednolicić wejściowy i wyjściowy alfabet danego operatora w ogólny alfabet A.

Jednymi z prostszych operatorów alfabetycznych są takie, które realizują odwzorowania literowe. Każda litera x wejściowego słowa R zoszaje zamieniona odpowiednią literą y wyjściowego alfabetu. Takie odwzorowanie nie zależy od struktury wejściowego słowa R i w pełni określa się zadaniem odpowiedniości między literami wejściowego i wyjściowego alfabetu.

4. Ważną postacią odwzorowania jest tzw. odwzorowanie kodujące. Przy prostym kodowaniu słowa w alfabecie A koduje się słowami w drugim alfabecie B w następujący sposób: każda litera a_i alfabetu A wiąże się z pewnym skończonym szeregiem liter $b_{i_1}, b_{i_2}, \ldots, b_{i_k}$ w alfabecie B. Ten szereg nazywa się kodem odpowiedniej litery. Różnym literom alfabetu A powinny odpowiadać różne kody.

Odwzorowanie kodujące słowa R realizuje się przy zamianie wszystkich jego liter odpowiednimi kodami. Przy tym otrzymamy pewne słowo w alfabecie B, nazywane kodem wejściowym słowa R. Jeden z podstawowych warunków polega na tym, że odwzorowanie kodujące powinno być odwracalne. Inaczej mówiąc powinien być spełniony warunek wzajemnej jednoznaczności odwzorowania kodującego.

Jasne jest, ze spełnienie warunku, żeby różnym literom odpowiadały różne kody, jeszcze nie zapewnia odwracalności odwzorowania. Przyjmijmy, że np. literze a_1 odpowiada kod b, a literze a_2 - kod bb. Wówczas kod bbb może w równym stopniu odpowiadać zarówno słowu a_2a_1, jak i słowom $a_1a_1a_1$ oraz a_1a_1. Po to aby kodowanie było odwracalne niezbędne jest spełnienie dwóch następujących warunków:

1/ kody różnych liter wejściowego alfabetu A powinny być różne,

2/ kod dowolnej litery alfabetu A nie powinien się pokrywać z żadnym z odcinków kodów innych liter tego alfabetu.

Początkowym odcinkiem słowa q = pr nazywa się słowo p, przy czym r - dowolne słowo /w tej liczbie również puste/. Przyjmiemy, że słowo $q = b_{i_1} b_{i_2} \dots b_{i_n}$ jest kodem pewnego słowa $p = a_{j_1} a_{j_2} \dots a_{j_m}$ w alfabecie A. Jeśli są spełnione warunki 1/ oraz 2/, to można wykazać, że według kodu q słowo p odwzorowuje się jednoznacznie.

Rzeczywiście z 2/ wynika, że tylko jeden odcinek początkowy słowa q może się pokrywać z kodem dowolnej litery alfabetu A. Jest jasne, że takim odcinkiem jest kod litery a_{j_1}. Odrzucając ten odcinek, otrzymujemy kod q_1 słowa $p = a_{j_2} a_{j_3} \dots a_{j_m}$.

Wykorzystując takie samo rozważanie, jednoznacznie odwzorowujemy /odtwarzamy/ literę a_{j_2} słowa p, itd.

5. Jeśli kody wszystkich liter wyjściowego alfabetu A mają taką samą długość, kodowanie nazywa się normalnym. Stosując kodowanie, można sprowadzić badanie dowolnych alfabetycznych odwzorowań do odwzorowań alfabetycznych w wybranym w określony sposób alfabecie standardowym.

Najczęściej takim standardowym alfabetem jest alfabet dwójkowy, składający się z dwóch liter np. 0 oraz 1.

Niech A, to dowolny, a B standardowy alfabet, składający się z więcej niż jednej litery. Jeśli w alfabecie A liczba liter wynosi n, a w alfabecie B - m, to zawsze można wybrać liczbę k tak, żeby spełniona była nierówność

$$m^k \geqslant n \qquad /94/$$

Liczba różnych słów o długości k w m-literowym alfabecie wynosi m^k. Dlatego też ze wzoru /94/ wynika, że wszystkie litery a alfabecie A można zakodować słowami o długości k w alfabecie B tak, żeby kody różnych liter były odmienne.

Pojęcie operatora alfabetycznego jest wystarczająco ogólne. Łączy ono w rzeczywistości dowolne procesy przekształcania informacji.

TEST

P. Co to jest alfabet abstrakcyjny.

xxx

O. Jest to skończony zbiór obiektów nazywanych literami danego alfabetu.

P. Co to jest słowo i odwzorowanie w alfabecie abstrakcyjnym.

xxx

O. Słowem w alfabecie abstrakcyjnym jest dowolny, skończony, uporządkowany szereg liter. Alfabetyczne odwzorowanie /operator/ jest to każdy sposób ustalania odpowiedniości danego słowa w określonym alfabecie, słowa w tym samym lub innym alfabecie.

10. Słowa w rachunku łącznym /asocjatywnym/

1. Opisany poprzednio przykład poszukiwania był rozpatrzony dla dowolnego, skończonego labiryntu. Problem słów w abstrakcyjnym alfabecie, w określonym pojęciu, można rozpatrywać jako uogólnienie zadania poszukiwania w nieskończonym labiryncie. Rozpatrzmy dwa słowa L i M w pewnym abstrakcyjnym alfabecie A. Jeśli L jest częścią M to mówi się, że sło-

wo L wchodzi w słowo M, inaczej mówiąc występuje zachodzenie słowa L na M.

Np. słowa ab i c d a b a.

Zachodzenie w ogólnym przypadku może być wielokrotne:

dad b d b d a d a d b .

2. Przekształcenie słów w abstrakcyjnym alfabecie opisuje się pewną procedurą, która daje możność otrzymywania ze słów zadanych, słów nowych.

TEST

P. Czy w słowach tor i kantor występuje zachodzenie?

xxx

O. Tak.

Niech w pewnym alfabecie będzie zadany skończony układ dopuszczalnych podstawień:

L - M; S - R; ... U - V;

gdzie L, M, R, S, ... U, V - słowa w tym alfabecie.

Dowolne z zadanych podstawień można przypisać pewnemu słowu P tego alfabetu. Jeśli np. słowo P raz lub kilkakrotnie wchodzi słowo S, to dowolne z tych zachodzeń można zamienić na słowo R i na odwrót, jeśli występuje zachodzenie słowa R, to można je zamienić na słowo S. Dla słowa a b c b c b a b podstawienie ab - bcb można zrealizować czterema sposobami. Zamiana każdego z dwóch zachodzeń b c b daje słowa a a b c b a b a b c a b a b, a zamiana każdego z dwóch zachodzeń ab daje słowa b c b c b c b a b, a b c b c b b c b.

Dla słowa b a c b podstawienia ab - dcb nie można przypisać, ponieważ do niego nie wchodzi ani ab, ani dcb. Jeśli do otrzymanych nowych słów przypisze się inne zadane podstawienia, to otrzyma się inne nowe słowa itd.

3. Rachunkiem łącznym nazywa się zbiór wszystkich słów w danym alfabecie razem z układem dopuszczalnych podstawień. Określić rachunek łączny - znaczy określić alfabet i układ podstawień.

4. Przyległymi nazywa się dwa słowa P_1 i P_2 w określonym rachunku łącznym, jeśli jedno z nich może być przekształcone na drugie, jednokrotnym przypisaniem pewnego dopuszczalnego podstawienia.

5. Dedukcyjnym łańcuchem prowadzącym od słowa P do słowa Q nazywa się szereg słów P, P_1, P_2, ..., Q, jeśli każde z dwóch obok siebie stojących słów tego łańcucha są przyległe.

6. Równoważnymi nazywa się dwa słowa P i Q, jeśli istnieje łańcuch dedukcyjny wiodący od słowa P do słowa Q.

Do oznaczenia stosunku równoważności stosuje się taki sam znak jak w rachunku zdań: $P \sim Q$. Ponieważ dopuszczalne podstawienia można dokonywać w obie strony, słuszne jest, że $/P \sim Q/$, $/Q \sim P/$.

Przykład

Zadany rachunek łączny:

/a, b, c, d, e/ - alfabet;

ae - ca
ad - da
bc - cb
bd - db
abac - dba
eca- ae
edb - be

Słowa abcde i acbde przylegają w tym rachunku, ponieważ słowo abcde przekształca się na acbde jednym podstawieniem bc - cb. Słowo aaabb nie ma słów przyległych - do niego nie można przypisać ani jednego podstawienia.

Wskutek występowania dedukcyjnego łańcucha abcde, acbde, cabde, cadbe, cadedb, słowa abcde i cadedb są równoważne. Przy budownie łańcucha dedukcyjnego kolejno zastosowano 3, 1, 4 i 5 podstawienie.

7. Rachunek łączny można porównać z pewnym nieskończonym labiryntem. Każdemu słowu pewnego alfabetu można przyporząd-

kować określone pole labiryntu. Labirynt będzie nieskończony, ponieważ z liter danego alfabetu abstrakcyjnego można zestawić nieskończoną liczbę słów. Dowolne dwa pola tego labiryntu odpowiadające przylegającym słowom połączone są korytarzem.

Równoważność słów P i Q oznacza, że w zbudowanym w taki sposób labiryncie pole odpowiadające słowu Q jest osiągalne z pola odpowiadającego słowu P.

TEST

P. Co to jest rachunek łączny?

xxx

O. Jest to zbiór wszystkich słów w danym alfabecie razem z układem dopuszczalnych podstawień.

8. W pewnych przypadkach celowe jest rozpatrywanie specjalnego rodzaju rachunku łącznego, który określa się alfabetem i układem ukierunkowanych podstawień postaci R $\rightarrow$ S. Strzałka oznacza, że podstawienie jest możliwe tylko z lewej w prawą stronę, tj. można zamienić R na S, lecz nie i na odwrót. Podobny rachunek łączny odpowiada nieskończonemu labiryntowi, w którym każdy korytarz można przechodzić tylko jeden raz, w jednym kierunku.
Oczywiście w takim rachunku łącznym z równoważności $P \sim Q$ nie wynika, że $Q \sim P$.

9. Problem słów dla dowolnego rachunku łącznego określa się następująco: dla każdych dwóch słów w danym rachunku należy ustalić czy są one równoważne, czy nie. W takim ujęciu problem jest równoważny z problemem osiągalności w przypadku labiryntu. Jednak teraz mamy do czynienia z nieskończonym labiryntem, dlatego wcześniej opisana metoda nie ma tu zastosowania. Zbadanie nieskończonego labiryntu w skończonym czasie jest niemożliwe.

10. Problem równoważności w dowolnym rachunku łącznym - to nieskończona liczba jednego typu zadań. Rozwiązanie tego problemu powinno przedstawiać algorytm dla ustanowienia równo-

ważności lub nierównoważności dowolnej pary słów.

11. Tak więc logiczne zadanie poszukiwania drogi w labiryncie można sformułować w terminach rachunku łącznego. W języku rachunku łącznego można opisywać również procesy logiczne. Np. dowolny wzór logiki matematycznej można rozpatrywać jako zapis słowa w pewnym alfabecie, którego literami są symbole logiczne $-, \vee, \wedge, \rightarrow$ itp., zmienne logiczne i funkcje logiczne /predykaty/.

Równoważności stosowane w logice matematycznej do przekształcania różnych wyrażeń rozpatruje się jako określone podstawienia, np. $\overline{X \vee Y}$ można zamienić na $\bar{X} \wedge \bar{Y}$ i na odwrót.

TEST

P. Czy w rachunku łącznym, zawierającym wyłącznie podstawienia jednokierunkowe, prawdziwe jest twierdzenie: "Jeśli $R \sim Q$ oraz $Q \sim R$ to $R \sim P$"?

xxx

O. Oczywiście, na ogół nie. Jeśli popełniłeś omyłkę, musisz jeszcze raz przestudiować punkt 6.

11. Algorytmy równoważne. Normalny algorytm Markowa

1. Dwa algorytmy a_1 i a_2 w pewnym alfabecie nazywa się równoważnymi, jeśli obszary ich stosowalności pokrywają się i wyniki przekształcenia nimi dowolnego słowa z ich ogólnego obszaru stosowalności także pokrywają się. Jeśli algorytm a_1 jest odpowiedni dla pewnego słowa P, to a_2 także powinien być odpowiedni dla tego słowa i odwrotnie. Przy tym oba algorytmy powinny przekształcać słowo P w takie samo słowo Q.

2. Markow uściślił pojęcie algorytmu w sensie określenia matematycznego. Zbudował on normalny algorytm w terminach systemu podstawień w następujący sposób. Przyjmuje się alfabet A i określa system podstawień. Wychodząc od dowolnego słowa P w alfabecie A należy przejrzeć formuły podstawień

w tym porządku, w jakim są one podane w schemacie. Przy tym odszukuje się formuły z lewej strony wchodzącej do P.

Jeśli takiej formuły brakuje, procedura algorytmiczna zatrzymuje się. Jeśli formuła jest, dokonuje się kolejno sprawdzeń i podstawień jej prawej strony w miejsce pierwszego zachodzenia jej lewej strony na P. Otrzymuje się nowe słowo P_1 a alfabecie A. Na tym kończy się pierwszy krok. Drugi krok jest taki sam, tylko teraz wejściowym słowem jest P_1 zamiast P. Analogiczne kroki wykonuje się do tej pory, dopóki nie nastąpi zatrzymanie.

To może nastąpić w dwóch przypadkach:

1/ przy otrzymaniu takiego słowa P_n, do którego nie wchodzi ani jedna z lewych stron formuł podstawień;

2/ gdy przy otrzymaniu słowa P_n zastosowano ostatnie podstawienie.

W jednym i w drugim przypadku mówi się, że dany algorytm przekształca słowo P w słowo P_n .

3. Różne normalne algorytmy różnią się tylko alfabetami i układami dopuszczalnych podstawień. Określić normalny algorytm - znaczy określić alfabet i układ podstawień.

Przykład

Określimy alfabet i układ podstawień w następujący sposób:

$$A = \{1, +\} \qquad \begin{aligned} 1+ &\to +1 \\ +1 &\to 1 \\ 1 &\to 1 \end{aligned}$$

Strzałkami przyjęto oznaczać układ podstawień w normalnym algorytmie Markowa /w odróżnieniu od dowolnego rachunku łącznego/.

Wejściowe słowo ma postać 1111 + 11 + 111.

Stosując normalny algorytm Markowa, kolejno otrzymujemy słowa:

1111 + 11 + 111
111 + 111 + 111
11 + 1111 + 111
1 + 11111 + 111
+ 111111 + 111
+ 1111 + 11111
+ 111 + 111111
+ 11 + 1111111
+ 1 + 11111111
+ + 11111111
+ 11111111
11111111
11111111

Proces algorytmiczny zatrzymuje się po zastosowaniu ostatniego podstawienia $1 \rightarrow 1$, które przekształca słowo 11111111 w siebie. Uściślenie pojęcia algorytmu, dokonane przez Markowa, polega na założeniu, że dowolny algorytm w alfabecie A jest równoważony z pewnym normalnym algorytmem w tym samym alfabecie.

TEST

P. Jakie algorytmy nazywa się równoważnymi.

xxx

O. Równoważnymi nazywa się takie algorytmy, których obszary stosowalności pokrywają się i wyniki przekształcenia nimi dowolnego słowa z ich ogólnego obszaru stosowalności także pokrywają się.

12. Problemy algorytmicznie niedopuszczalne

1. Normalny algorytm Markowa można rozpatrywać jako wygodną "formę standardową" dla określenia dowolnego algorytmu. Inaczej mówiąc można założyć, że ogólnie dowolny algo-

rytm może być określony w postaci normalnego algorytmu Markowa. To założenie jest hipotetyczne, lecz gdy przyjmie się taką hipotezę, można określić w jaki sposób ściśle dowieść niedopuszczalności algorytmicznej określonego kręgu problemów. Np. dowodzenie niedopuszczalności algorytmicznej problemu równoważności słów sprowadza się do dowodzenia tego, że w odpowiednim rachunku łącznym nie istnieje normalny algorytm rozpoznający równoważność słów P i Q.
To, że algorytm dla rozwiązania dowolnej klasy zdań nie istnieje, nie oznacza niedopuszczalności w ogóle. Rozpatrywana klasa zadań może być na tyle szeroka, że nie istnieje jedyna efektywna metoda ich rozwiązania. Dla konkretnych zadań jednak nie wyklucza się możliwości znajdowania szczególnych warunków rozwiązania.

Do uściślenia pojęcia "algorytm" w matematyce były rozpowszechnione dwa punkty widzenia.

1. Wszystkie problemy są algorytmicznie dopuszczalne. Po prostu dla rozwiązania niektórych z nich nie został znaleziony algorytm. Nie wystarcza środków we współczesnej matematyce do jego skonstruowania.

Stronnicy tego punktu widzenia uważali, że do rozwiązania problemów określanych jako algorytmicznie nierozwiązywalne, po prostu nie starcza środków współczesnej matematyki, a zbudowanie poszukiwanych algorytmów będzie dziełem przyszłości.

2. Istnieje klasa zadań dla rozwiązywania których w ogóle nie istnieją algorytmy, tj. niektórych problemów nie można rozwiązywać mechanicznie za pomocą formalnych rozważań i obliczeń. Te problemy wymagają twórczego myślenia.

Jest to bardzo silne przeświadczenie, rozciągające się na wiele epok i środków. Częściowym dowodem jest tu twierdzenie Gödla. Dzięki istnieniu hipotez o występowaniu "standardowych form" /np. normalny algorytm/, w którym mogą być wyrażone dowolne algorytmy, stało się możliwe sformułowanie pojęcia "algorytm" i "problem algorytmicznie niedopuszczalny" w ścisłych terminach.

13. Sprowadzanie dowolnego algorytmu do algorytmu arytmetycznego. Metoda Gődla

1. Jak to było widoczne w rozpatrywanych przykładach, określenie dla artymetycznych i logicznych algorytmów jest takie samo. W tym i drugim przypadku algorytmem nazwano zbiór reguł dla rozwiązania pewnej klasy zadań. W pierwszym i drugim przypadku wskazany zbiór reguł charakteryzuje się własnościami zdeterminowania, masowości i rezultatywności /ukierunkowania/.

2. Bardziej ścisłe jest pojęcie algorytmu przy wprowadzeniu do rozważań normalnego algorytmu Markowa. Okazuje się, że dowolny algorytm logiczny można stosunkowo prostymi metodami sprowadzić do algorytmu arytmetycznego. Z tego powodu teorię algorytmów arytmetycznych można uważać za uniwersalny sparat do badania wszystkich problemów algorytmicznych.

3. Problemy algorytmiczne można sprowadzić do obliczania wartości pewnej funkcji /w liczbach całkowitych/ przy wartościach argumentów wyrażonych w liczbach całkowitych.
Oznaczmy wszystkie warunki zadania, opracowywanego danym algorytmem , w postaci szeregów z całkowitymi nieujemnymi indeksami - numerami:

$$A_0, A_1, A_2, \ldots, A_n, \ldots$$

Rozwiązania można przedstawić jako szereg ponumerowanych wartości

$$B_0, B_1, B_2, \ldots, B_m, \ldots$$

Po wprowadzeniu numeracji będziemy dokonywać operacji nie na wartościach warunków i rozwiązań, lecz na ich numerach. Teraz można przedstawić algorytm, który przerabia numer zapisu warunku na numer zapisu rozwiązania. Ten algorytm realizuje obliczanie wartości funkcji

$$m = \varphi /n/$$

będąc algorytmem arytmetycznym.

4. Jeśli istnieje algorytm rozwiązujący wejściowe zadanie, to istnieje algorytm obliczający wartości odpowiedniej funkcji. Rzeczywiście, w celu znalezienia wartości $\varphi /n/$ przy $n = n^*$ można wybrać zapis warunku dla n^*, następnie za pomocą posiadanego algorytmu znaleźć zapis rozwiązania i według tego określić odpowiedni numer m^*. Tak więc:

$$\varphi /n^*/ = m^*$$

5. Rozpatrzmy następnie szeroko stosowaną do numeracji metodę Gödl'a.

Przedstawmy pewną liczbę n w postaci:

$$n = 2^{a_1} \cdot 3^{a_2} \cdot 5^{a_3} \cdot 7^{a_4} \dots p_{m-1}^{a_m},$$

gdzie $p_0 = 2$; $p_1 = 3$; $p_2 = 5$ itp, tj. p_m = liczba pierwsza.

Przyjmujemy, że każdej liczbie jednoznacznie odpowiada zbiór $a_1, a_2, \dots, a_m$ i na odwrót, każdemu zbiorowi $a_1, a_2, \dots, a_m$, jednoznacznie odpowiada liczba n.

Np. jeśli n = 60, to

$$60 = 2^2\ 3^1\ 5^1, \text{ tj. } a_1 = 2, a_2 = 1, a_3 = 1.$$

W ten sposób można numerować dowolne, uporządkowane szeregi m liczb.

Przykłady

1. Każdej parze liczb a_1 i a_2, dla której największy wspólny dzielnik zapisujemy jako q, można przypisać odpowiednio gödlowski numer tej pary

$$n = 2^{a_1}\ 3^{a_2}$$

Wówczas algorytm Euklidesa sprowadza się do obliczania funkcji

$$q = \varphi /n/$$

2. Przypiszemy numer n równaniu n-tego stopnia, zapisanemu w ogólnej postaci:

$$x^n + b_1 x^{n-1} + b_2 x^{n-2} + \dots + b_n = 0$$

Według n można łatwo ustalić zapis równania. Przy n = 2 mamy:

$$x^2 + b_1 x + b_2 = 0$$

Rozwiązanie uzyskuje się przez obliczenie wartości według wzoru

$$x = \frac{-b_1}{2} \pm \sqrt{\left(\frac{b_1}{2}\right)^2 - b_2}$$

Można to zapisać w postaci

$$x = , b_1 : 2 + - \sqrt{/b_1 . b_2 : 4 - b_2/}$$

Załóżmy, że trzeba znaleźć pierwiastki równania n-tego stopnia. Przy dowolnej postaci tego rozwiązania może ono składać się tylko z symboli:

$$+, -, ., :, (,), 1, b_1, b_2, \dots, b_n, \sqrt{}, \sqrt[3]{}, \dots, \sqrt[r]{}.$$

Oznaczmy te symbole pewnymi liczbami :

$\sqrt[r]{}$	odpowiada	liczba	2r	/	odpowiada	liczba	13
+	"	"	3	1	"	"	17
-	"	"	5	b_1	"	"	19
.	"	"	7	b_2	"	"	23
:	"	"	9	....			
	"	"	11	b_n	"	"	2n+15 .

Dowolnemu wyrażeniu zestawionemu z wymienionych symboli, odpowiada określony zbiór liczb. Np. dla wyrażenia

$$\sqrt[3]{/b_1 + b_2 /}$$

otrzymujemy zbiór liczb

6, 11, 19, 3, 23, 13.

Zbiorowi liczb można odpowiednio przypisać jego gödlowski numer

$$2^6 \cdot 3^{11} \cdot 5^{19} \cdot 7^3 \cdot 11^{23} \cdot 13^{13}$$

Z drugiej strony, według danego gődlowskiego numeru można zestawić zbiór liczb, a następnie dla każdej z nich określić odpowiedni symbol i utworzyć zapis dowolnego wzoru. W ten sposób można numerować dowolne wyrażenia zestawione zarówno z cyfr, jak i innych symboli - liter, znaków operacji itp.

3. Za pomocą metody Gődla można ponumerować wszystkie słowa w pewnym alfabecie A. Każdej literze przypisuje się odpowiednio pewną liczbę. Wówczas dowolnemu słowu w alfabecie A odpowiada szereg liczb, od którego można łatwo przejść do gődlowskiego numeru, zależnego od wybranego systemu odpowiedniości liter i liczb. Następnie można ponumerować wszystkie szeregi słów /np. wszystkie łańcuchy dedukcyjne/. Nie tylko arytmetyczne algorytmy sprowadzają się do obliczenia wartości funkcji w liczbach całkowitych. Dowolny, normalny algorytm Markowa przy zastosowaniu metody Gődla także można sprowadzić do obliczania wartości funkcji w liczbach całkowitych. Tak więc algorytm obliczania wartości funkcji w liczbach całkowitych można traktować jako uniwersalną formę algorytmu.

TEST

P. Na czym polega metoda numeracji Gődla?

xxx

O. Polega ona na przedstawianiu liczby n w postaci

$$n = 2^{a_1} \cdot 3^{a_2} \cdot 5^{a_3} \cdot 7^{a_4} \dots p_{m-1}^{a_m}$$

gdzie p_m - liczba pierwsza

Jeśli miałeś trudności z podaniem odpowiedzi na pytanie testowe, przestudiuj jeszcze raz rozdział IV, podr.7.

4. Algorytmy posiadają ścisły związek z teorią automatów skończonych, gdyż w wielu przypadkach można łatwo zbudować automat, który realizowałby przekształcenia zbiorów symboli

zgodnie z wymogami algorytmu. Automat taki nazwiemy realizatorem algorytmu. Niestety, nie wszystkie algorytmy posiadają swoje realizatory w dziedzinie automatów skończonych. Klasycznym przykładem algorytmu, który nie może być realizowany przez żaden automat, jest algorytm wydzielania z ciągu liczb, pełnych kwadratów.
Sformułujmy dokładniej: mamy zbiór jednakowych symboli, na przykład pałeczek. Należy rozpatrywać zbiór takich pałeczek pod kątem ich ilości i wydzielać takie zbiory, które zawierają ilości pałeczek będące pełnymi kwadratami. Należy więc rozróżnić

IIII

od

I II III IIIII IIIIII

lecz utożsamić z

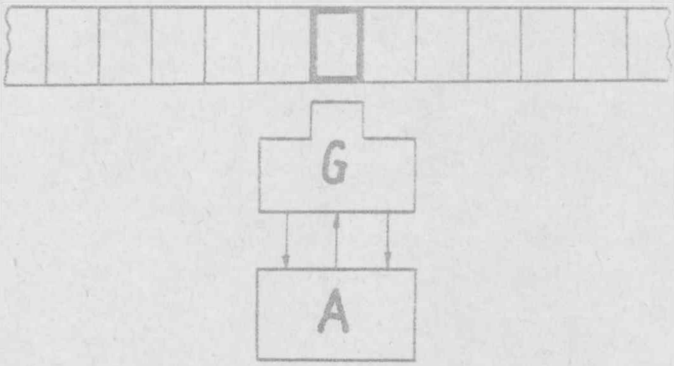
IIIIIIIII IIIIIIIIIIIIIIII

Udowodniono /szczegóły patrz M.Arbib; "Mózg, maszyna, matematyka"/, że nie potrafi tego dokonać żaden automat skończony.
Do realizacji d o w o l n e g o algorytmu potrzebujemy więc innego układu. Zaraz go poznamy.

14. Maszyna Turinga

1. Wyobraźmy sobie automat skończony połączony z nieskończenie długą taśmą podzieloną na kwadraty.

Elementem bezpośrednio komunikującym się z taśmą jest głowica G. I to wszysto, mamy już całą maszynę Turinga, złożoną z automatu A /patrz rys.88/, taśmy T i głowicy G. Najistotniejsza jest tu obecność taśmy, gdyż dzięki temu automat A może pokonać swoją "skończoność", którą mu tak wciąż wypominamy /nawet w nazwie/ i wykonywać czynności przewidziane dowolnym algorytmem.



Rys. 88. Maszyna Turinga

2. Podkreślmy mocniej tezę zawartą w ostatnim zdaniu poprzedniego punktu: maszyna Turinga zdolna jest wykonać k a ż d y algorytm. Znowu, jak i przy problemie normalizacji algorytmu, pojawia się trudność w dowodzie: jak w sposób formalny i matematycznie ścisły ująć pojęcie wszystkich algorytmów?

Naturalnie zrobić tego chwilowo nie potrafimy i dlatego stwierdzenie o uniwersalności maszyny Turinga jest jedynie hipotezą, ale nic nie wskazuje na to, by hipoteza ta miała upaść.

3. Na poparcie przytoczonej wyżej hipotezy przytoczymy kilka udowodnionych na gruncie cybernetyki twierdzeń, dotyczących maszyny Turinga. Naturalnie nie podajemy tu dowodów, które są dość skomplikowane i wymagają biegłego operowania złożonym aparatem matematycznym i pojęciowym, a przytaczamy gotowe wyniki, i to w nieco "zbeletryzowanej" formie.

4. Dowodzi się, że można zbudować maszynę Turinga, zdolną realizować dowolną procedurę rekurencyjną. Przez procedurę rekurencyjną rozumiemy czynność polegającą na kolejnym powtarzaniu grupy takich samych operacji skończoną lub nieskończo-

ną ilość razy. Na przykład,liczenie dowolnych przedmiotów jest jest procedurą rekurencyjną, gdyż polega na powtarzaniu dwu czynności:

1/ sprawdzenię, czy jest jeszcze jakiś przedmiot; w razie odpowiedzi "tak" -

2/ powiększenie pamiętanej liczby o jeden.

5. Dowodzi się ponadto, że ilość wszystkich możliwych maszyn Turinga jest przeliczalna. Określenie "jest przeliczalna", oznacza, że można by te maszyny ponumerować /aczkolwiek byłoby tych numerów nieskończenie wiele/. Jest to ważna właściwość, gdyż nie wszystkie zbiory zawierające nieskończoną ilość elementów dają się ponumerować. Nie szukając daleko - zbiór wszystkich liczb rzeczywistych zawartych pomiędzy zero a jeden jest nieprzeliczalny.

6. Ponieważ poszukiwanie określonego numeru jest czynnością rekurencyjną, więc na mocy wniosku podanego w punkcie 4 istnieje maszyna Turinga zdolna poszukiwać maszyny Turinga, mogącej wykonać dowolną operację rekurencyjną. Dowiedliśmy, że i s t n i e j e u n i w e r s a l n a m a s z y n a T u r i n g a. Inna sprawa, że nikt jej nie zbudował - brakuje /bagatela/ nieskończenie długiej taśmy.

7. Warto zapoznać się teraz nieco bliżej z konstrukcją i działaniem maszyny Turinga.

Na taśmie mogą znajdować się symbole z pewnego skończonego alfabetu, który będzie równocześnie alfabetem wejściowym automatu A. W każdym kwadracie może znajdować się co najwyżej jeden symbol. Ponieważ kwadratów jest nieskończenie wiele, a na taśmie znajduje się równocześnie jedynie skończona ilość symboli, więc niemal wszystkie kwadraty są niezapełnione. Głowica G kontaktuje się w danym momencie czasu ze ściśle określonym i tylko jednym kwadratem taśmy. Odczytuje ona wówczas zawartość tego kwadratu i przesyła ją jako słowo wejściowe do automatu A. Automat A reaguje na to pojawieniem się symboli na swoich dwu wyjściach. Na jednym pojawia się symbol z tego samego alfabetu jak wejściowy, który to symbol głowica zapisuje w "oglądanym" aktualnie kwadracie /usuwając jego poprzednią zawartość/. Dla prostoty przyjmiemy, że fakt braku symbolu w kwadracie jest również symbolem omawianego alfabetu oznaczymy Ø . Pojawienie się więc na wyjściu automatu A symbolu Ø jest więc poleceniem dla głowicy, aby "wymazała" zawartość kwadratu.

Na drugim wyjściu automatu A pojawia się jeden z trzech symboli: L, N, P. Symbol L oznacza, że głowica ma się przesunąć na kwadrat z lewej strony sąsiadujący z obecnie "oglądanym", P, że należy przejść do kwadratu po prawej, a N oznacza spoczynek głowicy. Dzięki drugiemu wyjściu maszyna może w sposób celowy poruszać się wzdłuż taśmy.

Automat A posiada odpowiedni do zadań, jakie ma spełniać, zbiór stanów wewnętrznych plus jeden stan szczególny, a mianowicie "STOP". Po p-przejściu do tego stanu maszyna zaprzestaje działalności, która w przeciwnym wypadku trwałaby w nieskończoność.

Przykład 1

Rozważymy maszynę Turinga, dla której symbolami pisanymi na taśmie będą pałeczki I, a która będzie operowała na ciągach takich pałeczek i będzie w efekcie usuwała dwie najbardziej na prawo położone pałeczki.

Możliwe są więc dwa słowa wejściowe "I" i "Ø". Przyjmiemy że automat A, wchodzący w skład maszyny, ma trzy stany /poza "STOP"/ S1, S2, S3. Wypiszemy tabelkę dla tego automatu, przy czym wypełnimy ją symbolami jedynie w tych miejscach, które mają wpływ na pracę maszyny Turinga, pozostałe pozycje tabelki oznaczymy "-", co oznacza, że mogą tam być wielkości dowolne, gdyż nie ma to znaczenia.

Maszyna ta jest bardzo prosta, można jednak podać opis maszyny przeznaczonej do bardziej złożonego celu, na przykład do liczenia ilości symboli Q i wpisywania cyfrą ich ilości. Rozpatrzenie działania takiej maszyny odłożymy jednak do testu, a tymczasem sprawdzimy, jak zachowuje się nasza maszyna.

słowo z taśmy	stan	słowo do zapisania	stan
I	S1	Ø P	S2
Ø	S1	Ø P	S1
I	S2	Ø P	S3
Ø	S2	Ø P	S2
I	S3	STOP	-
Ø	S3	STOP	-

Stanem startowym jest zawsze stan S1. Będąc w tym stanie maszyna sprawdza, czy w kwadracie aktualnie oglądanym znajduje się pałeczka. Jeśli tak, to kasuje ją, przechodzi do stanu S2 i szuka dalszych pałeczek wykonując ruch w prawo. Jeśli na początku pałeczki nie znaleziono, maszyna nie zmienia stanu, tylko porusza się dalej w prawo aż do znalezienia pałeczki. Po znalezieniu /i skasowaniu/ pierwszej pałeczki maszyna czyni podobne poszukiwania, utrzymując jednak stan S2. Znalezienie drugiej pałeczki powoduje jej skasowanie i przejście maszyny do stanu S3, w którym to stanie napotkanie dowolnego symbolu powoduje przejście do stanu STOP.

TEST 1

P. Prześledzić zachowanie się maszyny Turinga, opisanej podaną niżej tabelą. Automat A posiada aż 12 symboli wejściowych /wliczając w to symbol pusty/ i 4 stany wewnętrzne, tabela musiałaby więc zawierać 48 wierszy, co jest niewygodne. Przedstawimy więc ten automat w postaci prostokątnej tablicy, której wiersze odpowiadać będą symbolom wejściowym, a kolumny stanom wewnętrznym. W każdym tak utworzonym polu wpiszemy trzy symbole: pierwszy odpowiada symbolowi wyjściowemu drukowanemu na taśmie, drugi poleceniu ruchu głowicy, a trzeci nowemu stanowi.

	S1	S2	S3	S4
0	0 N S4	1 P S3	0 P S3	0 P S4
1	1 N S4	2 P S3	1 P S3	1 P S4
2	2 N S4	3 P S3	2 P S3	2 P S4
3	3 N S4	4 P S3	3 P S3	3 P S4
4	4 N S4	5 P S3	4 P S3	4 P S4
5	5 N S4	6 P S3	5 P S3	5 P S4
6	6 N S4	7 P S3	6 P S3	6 P S4
7	7 N S4	8 P S3	7 P S3	7 P S4
8	8 N S4	9 P S3	8 P S3	8 P S4
9	9 N S4	0 L S2	9 P S3	9 P S4
Q	Ø L S2	Q L S2	Q P S3	Q P S4
Ø	Ø L S1	1 P S3	Q L S1	Ø N STOP

Ze względu na nowość zapisu przed przystąpieniem do rozwiązywania zadań przećwiczmy posługiwanie się tabelą. Jaka będzie reakcja maszyny. Jeśli będąc w stanie S2 odczyta z taśmy symbol 9?

xxx

O. Wpisze na taśmę symbol "0", wykona ruch w lewo i pozostanie w stanie S2.

Jeśli odpowiedziałeś prawidłowo,jesteś gotów do rozwiązywania następnego zadania. Jeśli nie, przemyśl jeszcze raz strukturę powyższej tabeli.

P. Jak zachowa się maszyna, jeśli ją umieścić po prawej stronie ciągu symboli: ... ØØØØØØØQQQQQQØØØØ...? Stanem startowym jest S1. Co by było, gdyby symboli Q było więcej, w szczególności więcej niż 10?

xxx

O. Maszyna będzie się poruszała, nie zmieniając stanu, w lewo aż napotka pierwszy symbol Q /ostatni wiersz, pierwsza kolumna tabeli/. Po napotkaniu symbolu Q maszyna skasuje go, przejdzie do stanu S2 i zacznie poruszać się po ciągu symboli Q w lewo, aż napotka cyfrę lub puste miejsce. Za pierwszym razem oczywiście natrafi na puste miejsce. Po natrafieniu na cyfrę lub puste miejsce wpisuje cyfrę o jeden większą /lub jedynkę w puste miejsce/ i zaczyna powrót poprzez ciąg symboli Q w prawo. Zwiększenie liczby o 1 nastąpi również po natrafieniu na liczbę 9, lecz wykonane będzie w dwu taktach.

Powrót w prawo kończy się na "wyrwie" - pustym miejscu, z którego maszyna wymazała Q. Maszyna wpisuje tam Q z powrotem, przechodzi do stanu S1 i wyrywa następny z lewej symbol Q, wędruje w lewo aż do napotkania cyfry itd. Trwa to do chwili, aż maszyna dopisze cyfrę odpowiadającą ostatniemu Q, wróci o krok napotka pustkę, zapełni ją przez Q i będąc w stanie S1 uda się w lewo. Napotka tam od razu którąś z liczb, co spowoduje przejście do stanu S4. W stanie tym przebędzie w prawo cały ciąg symboli Q i wykona stop na końcu tego ciągu.

W efekcie dla postawionego zadania na taśmie pojawi się ciąg symboli: .. ØØØØ6QQQQQQØØØ...

Jeśli zadania tego nie rozwiązałeś sam w całości, musisz, po zapoznaniu się z powyższym opisem, określić, jaki ciąg symboli pojawi się na taśmie w wyniku podobnej jak wyżej obróbki ciągu: ...ØØØØØØQQQQQQQQQQQQQQQØØØ ... /wiadomo, że będzie to ciąg

... ØØØØ15QQQQQQQQQQQQQQQØØØØ ..,

chodzi jednak o to, by prześledzić krok po kroku jego powstawanie, zwłaszcza dla liczb większych od 9.

8. W punkcie 7 wykazaliśmy, że maszyna Turinga jest w zasadzie automatem skończonym, zaopatrzonym jedynie w pewne "przystawki". Złożoność maszyny Turinga można oceniać na podstawie bogactwa alfabetu wejściowego tego automatu i ilości jego stanów wewnętrznych. Właściwym miernikiem jej złożoności jest iloczyn tych wielkości. Za bardziej skomplikowaną uznamy

maszynę, dla której iloczyn ilości symboli alfabetu i ilości stanów wewnętrznych będzie mniejszy.

9. W punkcie 6 dowiedzieliśmy się, że istnieje uniwersalna maszyna Turinga. Powstaje więc problem: podać koncepcję minimalnej uniwersalnej maszyny Turinga, czyli maszyny uniwersalnej o minimalnej wartości wspomnianego iloczynu.

Zagadnienie to nie zostało definitywnie rozwiązane. Najlepszym z dotychczasowych było rozwiązanie A.Trittera o iloczynie 24 /4 symbole i 6 stanów/. Czy jest to optimum? Nie wiadomo.

10. W rozdziale tym poznaliśmy jeden z najwdzięczniejszych tworów cybernetyki, o którym wiele daje się udowodnić, ale z którego nie ma bezpośredniego pożytku. Poznaliśmy również cudowne dziecko cybernetyki - maszynę cyfrową, z której pożytek jest ogromny, ale o której właściwie nic nie daje się udowodnić. W miarę postępu techniki maszyn matematycznych pojemność ich pamięci będzie ustawicznie rosła i dlatego coraz szerszy będzie wachlarz zagadnień, dla których maszyna cyfrowa właściwie nie będzie różniła się od maszyny Turinga. Rysuje to świetne perspektywy przed dyscypliną obliczeń numerycznych. O ile się one sprawdzą - pokaże przyszłość. Skoro jednak doszliśmy do zagadnień możliwości układów cybernetycznych, należy przytoczyć i przedyskutować jeden z najpoważniejszych argumentów, przytaczanych przez zwolenników tezy, że możliwości dowolnych tworów cybernetycznych nigdy nie osiągną poziomu ludzkich zdolności intelektualnych. W tym celu zrobimy wycieczkę w dziedzinę najbardziej abstrakcyjnej matematyki i poznamy zarys twierdzenia Gödla - twierdzenia, które wstrząsnęło podstawami matematyki, a którego konsekwencje filozoficzne niepodobna przecenić.

Czytelnicy nie zainteresowani tematyką skrajnych możliwości maszyn i mózgu ludzkiego mogą ten rozdział opuścić, jednak zachęcamy do jego przeczytania, ponieważ daje wiele do myślenia, a kwestie w nim poruszane nie należą bynajmniej do ostatecznie rozstrzygniętych.

15. Twierdzenie Gödla

1. Na początku konieczne jest kilka słów komentarza z dziedziny historii matematyki, aby naświetlić klimat, w jakim doszło do sformułowania twierdzenia Gödla.

Od czasów starożytnych rozwijano matematykę, a w szczególności geometrię, w sposób aksjomatyczno-dedukcyjny. Sposób ten polega na przyjęciu pewnej ilości twierdzeń pierwotnych, tzw. aksjomatów, bez dowodu, opierając się na "oczywistości", a następnie wywodzeniu z nich kolejnych twierdzeń metodami logicznego i ścisłego wnioskowania. System taki ma wiele zalet, z których wymienić można klarowną i logiczną strukturę, łatwość prowadzenia dowodów itp. Wada jest jedna, ale zasadnicza: cały system jest tylko o tyle prawdziwy, o ile prawdziwe są aksjomaty leżące u jego podstaw.

Aksjomaty leżące u podstaw geometrii euklidesowej przez wieki nie budziły zastrzeżeń, jako że opierały się na najelementarniejszych spostrzeżeniach w rodzaju: "Przez dwa nie pokrywające się punkty przechodzi tylko jedna linia prosta".

W miarę jak ludzkość rozwijała zdolności abstrakcyjnego myślenia, owe prawdy "oczywiste" zaczęto podawać w wątpliwość. Szczególnie ostre ataki przypuszczone na aksjomat, mówiący że przez punkt nie leżący na danej prostej można przeprowadzić *tylko jedną prostą równoległą do danej*.

Prace poszukujące odpowiedzi na pytanie, czy aksjomat ten jest istotnie do przyjęcia, doprowadziły do zbudowania /przez Riemanna i innych/ geometrii negującej ten aksjomat, a w całej reszcie równie spójnej i matematycznie poprawnej jak euklidesowa. Geometria Riemanna twierdzi, że przez punkt nie leżący na danej prostej można przeprowadzić aż nieskończenie wiele prostych równoległych do danej. System ten, aczkolwiek wydaje się być sprzeczny z obserwacją codzienną, lepiej opisuje otaczający nas Wszechświat niż "naiwna" geometria euklidesowa, co niedwuznacznie wynika z teorii względności Einšteina. Aspekt ten nie będzie przez nas bliżej omawiany, chodzi bowiem jedynie o zasygnalizowanie faktu, że system aksjomatyczny można w matematyce budować w sposób dość dowolny, zmieniając aksjomaty.

Wyłania się jednak poważny problem: system aksjomatów /i cały wzniesiony na nim system matematyczny/ może być wewnętrznie sprzeczny.

Sprzeczność ta ujawniałaby się, gdybyśmy, posługując się aksjomatami i poprawnymi regułami wnioskowania, zdołali udowodnić w sposób ścisły dane twierdzenie i jego zaprzeczenie. Pytanie o niesprzeczność geometrii poruszyło najtęższe umysły matematyczne. Problem miał zresztą dużo szerszy zasięg: posługując się geometrią analityczną można udowodnić wszystkie twierdzenia geometrii na gruncie algebry. Zatem geometria jest niesprzeczna, o ile niesprzeczna jest algebra. Ale algebra wywodzi się z arytmetyki, a raczej z elementarnej teorii liczb operującej twierdzeniami w rodzaju "1 + 1 = 2" /przysłowiowe 2x2 jest już w tej teorii bardzo skomplikowanym twierdzeniem/. Stanęło więc na koniec przed matematykami pytanie: czy arytmetyka /a wraz z nią cała współczesna matematyka/ nie jest czasem wewnętrznie sprzeczna? Fakt, że nie odkryto jeszcze dowodu twierdzenia i jego zaprzeczenia na gruncie znanej matematyki nie stanowi gwarancji, gdyż w każdej chwili dowód taki mógłby się pojawić.

Program poszukiwania ścisłego dowodu niesprzeczności matematyki nakreślił Hilbert, a zrealizował go częściowo Russel w swym monumentalnym dziele "Pricipia Mathematica".

2. W tej sytuacji wystąpił w 1931 roku ze swym twierdzeniem Gödel. W obszernym artykule zatytułowanym "O formalnie nierozstrzygalnych problemach Principia Mathematica i systemów pokrewnych"[18/] Gödel udowodnił, że niemożliwe jest udowod-

[18/] K. Gödel: Über formal unentscheidbare Sätre der Priuzipia Mathematica und verwandter Systeme, 1931.

nienie niesprzeczności logiki arytmetycznej za pomocą metod dających się odwzorować w tej logice. Problem niesprzeczności jest więc drogą formalną n i e r o z s t r z y g a l n y. Zwróćmy uwagę na fakt, że niemożliwe jest rozstrzygnięcie metodami f o r m a l n y m i, czyli metodami dającymi się zalgorytmizować. Istotnie, podano później dowód niesprzeczności arytmetyki, lecz Gentzen, który to zrobił, zastosował metodę rozumowania zwaną indukcją ε_0, polegającą na nieskończonym rozszerzaniu metody indukcji matematycznej. Jest to metoda nieformalna i nie dająca się ująć w algorytm. Nawiasem mówiąc, niesprzeczność tej metody jest równie wątpliwa, jak niesprzeczność arytmetyki, której miała dowodzić.

3. Wydobądźmy jeszcze raz silnie interesujący nas aspekt twierdzenia Gödla: istnieje nader ważki problem w łonie nauki o podstawach matematyki, a mianowicie problem niesprzeczności systemów matematycznych, który, co ściśle dowiedziono, n i e m o ż e b y ć rozstrzygnięty metodami dającymi się zalgorytmizować.

4. Gödel udowodnił jeszcze więcej. Wykazał mianowicie, że każdy system formalny jest niezupełny, to znaczy istnieją takie p r a w d z i w e twierdzenia w tym systemie, których nie da się udowodnić w ramach tego systemu.

5. Postawmy kropkę nad "i".
Przecież dowiedzieliśmy się w poprzednim rozdziale /i wprawiło nas to w zachwyt/, że maszyna Turinga może podołać każdemu sformalizowanemu zadaniu. Okazuje się jednak, że metody formalne posiadają ograniczenia, ba, że ograniczenia te stanowią ich część integralną, niezależnie od formy algorytmizacji.
Konkludujemy: istnieją prawdy, których nie może osiągnąć żaden twór zbudowany w sposób formalny i niesprzeczny. Wszystkie znane twory cybernetyczne, z maszyną Turinga na czele są niesprzeczne. Zatem prawdy te są dla maszyn raz na zawsze niedostępne. Są one jednak dostępne dla człowieka, tak jak

chociażby dowód niesprzeczności arytmetyki Gentzena. Maszyna postępuje ściśle, pewnie, szybko, logicznie. Człowiek postępuje intuicyjnie, rozpaczliwie wolno i heurystycznie. Jednak w ostatecznej konfrontacji, konfrontacji, której często się obawiamy w dobie komputerów, człowiek ma matematycznie zagwarantowaną przewagę, właśnie dlatego, że jest wolny od chłodnej logiki automatów.

Rozdział V

ELEMENTY LOGIKI MATEMATYCZNEJ

W algebrze działania na liczbach zamienia się na działania na literach. Słowa "dodać", "odjąć", "pomnożyć", "podzielić itp. zamienia się w algebrze na znaki - symbole: +, -, ·, : itp. Posługując się oznaczeniami algebraicznymi równanie sześcienne można krótko zapisać: $x^3 + ax = b$. Do tak krótkiego sposobu zapisu uczeni doszli dopiero w XVIII wieku.

Wprowadzenie algebraicznych oznaczeń do badań na polu logiki rozpoczęło się w drugiej połowie XVII wieku, lecz wyodrębnienie się logiki matematycznej w samodzielną gałąź nauki nastąpiło w połowie XIX wieku.

W ostatnich latach obserwujemy proces zastosowań logiki matematycznej w różnych dziedzinach takich, jak: automatyka, matematyka, informatyka i inne.

1. Zadania. Prawdziwość i fałszywość zdań

1. Jednym z działów logiki matematycznej jest rachunek zdań. Przedmiotem badań rachunku zdań są zdania, np. dwa razy dwa równa się siedem, liczba 4 jest nieparzysta, śnieg jest biały, lipiec to letni miesiąc. Możemy postawić pytanie, które z tych zdań są prawdziwe, a które fałszywe.

Spośród przytoczonych zdań 1 i 2 są fałszywe, a 3 prawdziwe. Czwarte zdanie jest prawdziwe dla półkuli północnej, a fałszywe - dla południowej. Istnieją ponadto takie zdania, o których trudno powiedzieć, czy są prawdziwe czy fałszywe.

Np. 2 to liczba nieparzysta. Wprawdzie wiemy co znaczy określenie "liczba całkowita nieparzysta", lecz co oznacza "liczba niewymierna nieparzysta" - tego nie wiemy. W dalszych rozważaniach będziemy używać tylko takich zdań, o których wiadomo, czy są one prawdziwe czy fałszywe.

2. Istnieją zdania złożone z pewnych zdań prostych, powiązanych między sobą spójnikami: i, lub, jeśli - to itp. Tę samą myśl można wyrazić za pomocą zdań prostych lub zdań złożonych. Np. zdanie proste "zarobek Kowalskiego jest nie większy niż tys.zł", można zastąpić zdaniem złożonym "zarobek Kowalskiego jest mniejszy niż 1 tys.zł lub równy 1 tys.zł ". Zagadnienie dotyczące prawdziwości lub fałszywości zdań jest podstawowym problemem logiki matematycznej.

3. Badając zdania, abstrahuje się od ich treści, występowania i innych ważnych właściwości, a koncentruje się uwagę na dwóch zagadnieniach:

1/ jak zależy prawdziwość tego lub innego zdania złożonego od prawdziwości wchodzących w jego skład bardziej prostych zdań?

2/ jak zależy prawdziwość niektórych /lub wszystkich/ prostych zdań wchodzących w skład złożonego, od prawdziwości tego zdania złożonego i pozostałych prostych zdań, wchodzących w jego skład?

4. Zdania będziemy krótko oznaczać dużymi literami łacińskiego alfabetu:

$$X, Y, Z, U, V, \ldots$$

Przy tym różne litery odpowiadają różnych zdaniom, a te same litery - jednakowym zdaniom. Zdania złożone będziemy oznaczać albo jedną z liter, albo literami, odpowiadającymi składowym zdaniom, połączonymi odrębnymi znakami.

2. Spójniki logiczne

1. Do opisu połączeń logicznych wprowadzimy pięć następujących znaków:

1/ $\bar{X}$ /czytą się "nie X"/ oznacza zaprzeczenie X. Inaczej mówiąc $\bar{X}$ oznacza zdanie, które jest prawdziwe, gdy X jest fałszywe, lub zdanie fałszywe, jeśli X jest prawdziwe.

2/ $X \wedge Y$ /czyta się "X i Y"/ oznacza zdanie, które jest prawdziwe tylko w tym przypadku; gdy X i Y są prawdziwe.

3/ $X \vee Y$ /czyta się X lub Y/ oznacza zdanie, które jest prawdziwe tylko w tym przypadku, gdy co najmniej jedno z dwóch zdań X lub Y jest prawdziwe.

4/ $X \rightarrow Y$ /czyta się "jeśli X, to Y"/ oznacza zdanie, które jest fałszywe tylko w tym przypadku, gdy X jest prawdziwe, a Y fałszywe.

5/ $X \sim Y$ /czyta się "X jest równoważone z Y"/ oznacza zdanie, które jest prawdziwe wtedy i tylko wtedy, gdy X i Y są prawdziwe, lub gdy X i Y są fałszywe. Tak więc $X \sim Y$ oznacza, że X i Y posiada to samo znaczenie prawdziwości lub fałszywości.

2. Odnośnie trzeciego określenia należy zauważyć, że zdanie złożone typu "X lub Y" może przyjmować dwa różne znaczenia. Np. "obiekt r wchodzi do zbioru A lub do zbioru B" można rozumieć następująco:

1/ obiekt r wchodzi do jednego i tylko jednego z dwóch zbiorów,

2/ obiekt r wchodzi albo do zbioru A albo do zbioru B, tj. że do obydwu tych zbiorów obiekt r nie może wchodzić; ale można to twierdzenie rozumieć i tak, że obiekt r wchodzi przynajmniej do jednego z dwóch zbiorów, tj. nie wyklucza się możliwości wchodzenia obiektu r do obydwu zbiorów.

My będziemy rozumieli spójnik "lub" i stosunek 3 tylko w tym drugim rozumieniu.

3. Wykluczając "albo - albo" /w pierwszym znaczeniu/ można wyrazić za pomocą pewnych kombinacji podstawowych znaków.

"Albo X albo Y" jest zaprzeczeniem w stosunku do zdania $X \sim Y$ i można je zapisać $\overline{X \sim Y}$.

4. Określenia 4 /"jeśli X, to Y"/ nie należy rozumieć jako wyrażenie stosunku założenia i tezy. Odwrotnie, zdanie X Y jest prawdziwe nawet w tym przypadku, gdy X jest fałszywe a Y prawdziwe. Np. zdaniami prawdziwymi są: jeśli "3 x 3 = 9", to "miedź jest metalem", jeśli 3 x 3 = 7", to "miedź jest metalem", jeśli "3 x 3 = 7", to "miedź jest niemetalem", natomiast fałszywym byłoby zdanie: jeśli "3 x 3= 9", to "miedź jest niemetalem". Stosunek $X \to Y$ posiada w tym sensie wspólne znaczenie ze stosunkiem założenia i tezy, że w przypadku prawdziwości $X \to Y$ z prawdziwości Y można wnioskować o prawdziwości Y.

5. Stosunek 5 /$X \sim Y$/ zachodzi między dowolnymi dwoma prawdziwymi, lub dowolnymi dwoma fałszywymi zdaniami. Np. zdania "3 x 3 = 9" $\sim$ "miedź jest metalem", "3 x 3 = 7" $\sim$ "miedź jest niemetalem" są prawdziwe.

Odrębnie należy podkreślić, że zgodnie z naszym określeniem podstawowych spójników logicznych-prawdziwość lub fałszywość zdania złożonego zależy tylko od prawdziwości i fałszywości składowych zdań, a nie od ich treści.

6. Wprowadzimy jeszcze dwa oznaczenia. Zdania prawdziwe będziemy oznaczać literą P, a fałszywe - literą F. Tak więc np. połączenie "$\to$" charakteryzuje się tym, że zdania $P \to P$, $F \to P$ i $F \to F$ są prawdziwe, a zdanie $P \to F$ jest fałszywe. Dla spójnika "$\wedge$" zdanie $P \wedge P$ jest prawdziwe, a wszystkie pozostałe: $P \wedge F$, $F \wedge P$, $F \wedge F$ są fałszywe. Następnie $P \vee P$, $P \vee F$, $F \vee P$ - są prawdziwe, a $F \vee F$ jest fałszywe. Spójnik "$\sim$" charakteryzuje się tym, że $P \sim P$, $F \sim F$ są prawdziwe, a $P \sim F$ i $F \sim P$ - fałszywe. W końcu $\bar{P}$ jest fałszywe, $\bar{F}$ jest prawdziwe.

7. W ten sposób będziemy rozpatrywać podstawowe spójniki jako funkcje prawdziwości, tj. jako określenie funkcji dla

której w charakterze argumentów i wartości funkcji rozpatruje się tylko P lub F.

TEST

P. Nazwij i wyjaśnij wszystkie postacie spójników logicznych. /podaj przykłady/.

xxx

0. Nie. np. $\bar{X}$ /nie X/.
I. np. $X \wedge Y$ /X i Y/.
Lub. np. $X \vee Y$ /X lub Y/.
Jeśli, to. np. $X \rightarrow Y$ /jeśli X, to Y/.
Równoważne. np. $X \sim Y$ /X jest równoważne z Y/.

Jeśli wyjaśnienia sprawiały Ci trudności, to wróć do 6.2. a jeśli nie - to przejdź do rozdz.VI podr.3.

3. Równoważność. Zamienność podstawowych połączeń logicznych

1. Każde zdanie złożone, tak jak i proste połączenia zdań, przedstawia określoną funkcję prawdziwości.
Rozpatrzmy

$$\left[(X \rightarrow Y) \wedge (Y \rightarrow Z)\right] \wedge (X \vee Z)$$

Dla X,Y,Z możliwe jest osiem trójek wartości:
P,P,P; P,P,F; P,F,P; P,F,F; F,P,P; F,P,F; F,F,P; F,F,F,
Przy podstawianiu do podstawowego wzoru dowolnej trójki, otrzymuejmy wartość P lub F. Np. kombinacji F,P,F odpowiada wartość F.

$$\left[(F \rightarrow P) \wedge (P \rightarrow F)\right] \wedge (F \vee F)$$

$$P \wedge F \wedge F$$

$$F \wedge F$$

$$F$$

2.Zauważmy, że niektóre inne sposród tych połączeń są równoznaczne tj. wyrażają tę samą funkcję prawdziwości. Tak więc, $\bar{\bar{X}}$ jest równoznaczne X. Podobne równoznaczne połączenia będziemy nazywać ekwiwalentnymi:

$$X \text{ ekw } X \qquad /95/$$

$$X \wedge Y \text{ ekw } Y \wedge X \qquad /96/$$

$$X \wedge (Y \wedge Z) \text{ ekw } (X \wedge Y) \wedge Z \qquad /97/$$

$$X \vee Y \text{ ekw } Y \vee X \qquad /98/$$

$$X \vee (Y \vee Z) \text{ ekw } (X \vee Y) \vee Z \qquad /99/$$

$$X \vee (Y \wedge Z) \text{ ekw } (X \vee Y) \wedge (X \vee Z) \qquad /100/$$

Z ekwiwalentności /96/ - /100/ wynikają prawa: przemienności, łączności i rozdzielności.

3. Z przytoczonych praw widać, że w logicznych wyrażeniach można, jak w algebrze, mnożyć lub wynosić poza nawias wspólny czynnik. Według analogii z algebrą $X \vee Y$ nazywa się sumą logiczną, a $X \wedge Y$ iloczynem logicznym. W odróżnieniu jednak od algebry w rachunku zdań działa inne prawo rozdzielności:

$$X \wedge (Y \vee Z) \text{ ekw } (X \wedge Y) \vee (X \wedge Z) \qquad /101/$$

W ten sposób z powodzeniem moglibyśmy nazwać $X \vee Y$ iloczynem logicznym, a $X \wedge Y$ logiczną sumą. Ponieważ jednak w logice odnośnie używania słów "suma" i "iloczyn" istnieje nieokreśloność będziemy w miarę możliwości unikać tych wyrażeń.

$X \wedge Y$ będziemy nazywać koniunkcją,

$X \vee Y$ " " alternatywą,

$X \rightarrow Y$ " " implikacją.

4. Zgodnie z prawem przemienności i łączności wieloczłonowe koniunkcje, alternatywy można zapisywać bez nawiasów. Prócz tego, w celu zmniejszenia liczby nawiasów, przyjmujemy, że $\wedge$ łączy silniej niż $\vee$, a $\vee$ z kolei silniej niż $\sim$ i $\rightarrow$. Znaku $\wedge$ można nie stawiać, podobnie jak w lagebrze nie musi się stawiać znaku mnożenia.

Rozpatrzmy jeszcze raz ekwiwalentności:

$$X \wedge X \text{ ekw } X \qquad /102/$$

$$X \vee X \text{ ekw } X \qquad /103/$$

$$X \wedge P \text{ ekw } X \qquad /104/$$

$$X \wedge F \text{ ekw } F \qquad /105/$$

$$X \vee P \text{ ekw } P \qquad /106/$$

$$X \vee F \text{ ekw } X \qquad /107/$$

$$P \rightarrow X \text{ ekw } X \qquad /108/$$

$$F \rightarrow X \text{ ekw } P \qquad /109/$$

$$X \sim P \text{ ekw } X \qquad /110/$$

$$X \sim F \text{ ekw } \bar{X} \qquad /111/$$

i niektóre bardziej złożone ekwiwalentności:

$$\overline{X \wedge Y} \text{ ekw } \bar{X} \vee \bar{Y} \qquad /112/$$

$$\overline{X \vee Y} \text{ ekw } \bar{X} \wedge \bar{Y} \qquad /113/$$

$$X \rightarrow Y \text{ ekw } \overline{X \wedge Y} \qquad /114/$$

Wykorzystując /112/, można /114/ zapisać w postaci

$$X \rightarrow Y \text{ ekw } \bar{X} \vee \bar{Y},$$

a następnie zgodnie z /95/

$$X \rightarrow Y \text{ ekw } \bar{X} \vee Y \qquad /115/$$

$$X \vee Y \text{ ekw } \bar{X} \rightarrow Y \qquad /116/$$

$$X \rightarrow Y \text{ ekw } \bar{Y} \rightarrow \bar{X} \qquad /117/$$

$$X \sim Y \text{ ekw } (X \rightarrow Y) \wedge (Y \rightarrow X) \qquad /118/$$

z określenia relacji $\sim$ bezpośrednio otrzymujemy, że

$$X \sim Y \text{ ekw } Y \sim X \qquad /119/$$

$$X \sim Y \text{ ekw } \bar{X} \sim \bar{Y} \qquad /120/$$

Z /112/ i /113/ otrzymujemy:

$$X \wedge Y \text{ ekw } \overline{\bar{X} \vee \bar{Y}} \qquad /121/$$

$$X \vee Y \text{ ekw } \overline{\bar{X} \wedge \bar{Y}} \qquad /122/$$

5. Oczywiście jest, że niektóre z podstawowych spójników logicznych są zbyteczne. Ze wzoru /118/ widać, że można się obejść bez znaku $\sim$. Następnie z /114/ i /121/ wynika, że znaki $\rightarrow$ i $\vee$ również są zamienne i można się obejść tylko znakami $\wedge$ i $-$. Ze wzorów /115/ i /122/ wynika, że można się ograniczyć do znaków $\vee$ i $-$.
Rozpatrzmy jeszcze równoważności ważne przy przedstawianiu stosunków równoznaczności

$$X \sim Y \text{ ekw } (\bar{X} \vee Y) \wedge (\bar{Y} \vee X) \qquad /123/$$

$$X \sim Y \text{ ekw } (X \wedge Y) \vee (\bar{X} \wedge \bar{Y}) \qquad /124/$$

6. Najbardziej słuszne wydaje się stosowanie trzech znaków $\wedge$, $\vee$, $-$, ponieważ na podstawie równoważności /96/ -/100/, otrzymuje się przy tym szczególnie prosty zapis zdań logicznych oraz procedury obliczeniowej.

TEST

P. Podaj podstawowe postacie przekształceń równoważnych dla połączeń koniunkturalnych i alternatywnych.

xxx

O. $X \wedge Y \text{ ekw } Y \wedge X$

$X \wedge (Y \wedge Z) \text{ ekw } (X \wedge Y) \wedge Z$

$X \vee Y \text{ ekw } Y \vee X$

$X \vee (Y \vee Z) \text{ ekw } (X \vee Y) \vee Z$

$X \vee (Y \wedge Z) \text{ ekw } (X \vee Y) \wedge (X \vee Z)$

$X \wedge (Y \vee Z) \text{ ekw } (X \wedge Y) \vee (X \wedge Z)$

P. Podaj przekształcenia równoważne dla przedstawienia implikacji i równoznaczności.

xxx

0. $X \rightarrow Y$ ekw $\bar{X} \vee \bar{\bar{Y}}$ ekw $\bar{X} \vee Y$

$X \sim Y$ ekw $(X \rightarrow Y) \wedge (Y \rightarrow X)$

4. <u>Normalna postać wyrażeń logicznych</u>

1. Każde zdanie złożone można sprowadzić do postaci normalnej drogą równoważnego przekształcania. Ta postać normalna składa się z pewnej koniunkcji alternatyw, w której każdy człon alternatywny jest albo zdaniem prostym, albo jego zaprzeczeniem.

Reguły przekształcania.

1/ Znakami $\wedge$ i $\vee$ można posługiwać się tak jak w algebrze, stosując prawa przemienności, łączności i rozdzielności.

2/ $\bar{\bar{X}}$ można zamieniać na X.

3/ $\overline{X \wedge Y}$ można zamieniać na wyrażenie $\bar{X} \vee \bar{Y}$, a $\overline{X \vee Y}$ - na $\bar{X} \wedge \bar{Y}$.

4/ $X \rightarrow Y$ można zamieniać na wyrażenie $\bar{X} \vee Y$, a $X \sim Y$ na $(\bar{X} \vee Y) \wedge (\bar{Y} \vee X)$

Przykład: znaleźć koniunktywną postać normalną wyrażenia

$$\overline{(XY \wedge \bar{Y}) \vee (Z \wedge Y)}$$

Kolejno otrzymujemy:

$$\overline{(XY \wedge \bar{Y})} \wedge \overline{(Z \wedge Y)}$$

$$(\overline{XY} \vee \bar{\bar{Y}}) \wedge (\bar{Z} \vee \bar{Y})$$

$$(\bar{X} \wedge \bar{Y}) \bar{\bar{Y}} \wedge \bar{Z}\bar{Y}$$

Teraz zastosujemy prawo rozdzielności:

$$\bar{X} \bar{\bar{Y}} \wedge \bar{Y} \bar{\bar{Y}} \wedge \bar{Z} \bar{Y}$$

$$\bar{X} Y \wedge \bar{Y} Y \wedge \bar{Z} \bar{Y}$$

Otrzymaliśmy koniunktywną postać normalną. Istnieje również druga postać normalna - alternatywna. Przedstawia ona pewną alternatywę koniunkcji, w której każdy człon koniunktywny jest zaprzeczonym lub niezaprzeczonym zdaniem prostym.

Przykład

Znaleźć alternatywną postać normalną zdania złożonego

$X \wedge (X \rightarrow Y)$

Posługując się równoważnością /115/ otrzymamy:

$X \wedge (\bar{X} \vee Y)$

Otwierając nawiasy dostajemy:

$(X \wedge \bar{X}) \vee (X \wedge Y)$

TEST

P. Sformułuj reguły sprowadzania zdań do normalnej postaci.

xxx

O. 1. Znakami $\wedge$ i $\vee$ można posługiwać się tak jak w algebrze, stosując prawa przemienności, łączności i rozdzielności.

2. $\bar{\bar{X}}$ można zamienić na X.

3. $\overline{X \wedge Y}$ można zamienić na $\bar{X} \vee \bar{Y}$, a wyrażenie $\overline{X \vee Y}$ na wyrażenie $\bar{X} \wedge \bar{Y}$.

4. $X \rightarrow Y$ można zamienić na $\bar{X} \vee Y$, a $X \sim Y$ na $(\bar{X} \vee Y) \wedge (\bar{Y} \vee X)$

5. Zdania zawsze prawdziwe i zawsze fałszywe /tautologia/

1.Głównym zadaniem logiki jest znajdowanie takich połączeń, które są zawsze prawdziwe, niezależnie od tego, czy przedstawiają one proste zdania prawdziwe, czy fałszywe. Ponieważ każdemu wyrażeniu logicznemu możemy przypisać odpowiednio równoważne mu wyrażenie w postaci normalnej, należy ustalić, kiedy wyrażenie w postaci normalnej jest zdaniem zawsze prawdziwym.

2. Zdaniami zawsze prawdziwymi okazują się zdania, które w koniunktywnej postaci normalnej charakteryzują się tym, że w każdej alternatywie co najmniej jedno zdanie proste spotyka się jednocześnie z jego zaprzeczeniem.

Przykład :

$X \wedge Y \rightarrow X$

$\overline{X \wedge Y} \vee X$, /zgodnie z regułą 4/

$$\bar{X} \vee \bar{Y} \vee X \text{ , /zgodnie z regułą 3/}$$

Ostatnia alternatywa zawiera X i $\bar{X}$, a więc jest ona prawdziwa.

3. Za pomocą alternatywnej postaci normalnej można ustalić, czy istnieją zdania zawsze fałszywe.
Zdarza się to w tym i tylko w tym przypadku, gdy każdy człon alternatywny jednocześnie ze zdaniem prostym zawiera również zdanie mu przeciwstawne.

Przykład:

$$\bar{X}\,Y \wedge \bar{Y}\,Z \wedge X \wedge \bar{Z} \qquad /125/$$

Zastosujemy drugie prawo rozdzielności:

$$(\bar{X} \wedge \bar{Y} \wedge X \wedge \bar{Z}) \vee (\bar{X} \wedge Z \wedge X \wedge \bar{Z}) \vee (Y \wedge \bar{Y} \wedge X \wedge \bar{Z}) \vee (Y \wedge Z \wedge X \wedge Z)$$

Jak widać, każdy alternatywny człon zawiera pewne zdanie proste z jego zaprzeczeniem. Wynika z tego, że zdanie /125/ jest zawsze fałszywe.

TEST

P. W jakim przypadku zdanie złożone jest zawsze fałszywe.

xxx

O. Tylko w tym przypadku, gdy każdy człon alternatywny jednocześnie ze zdaniem prostym zawiera również zdanie mu przeciwstawne.

Jeśli miałeś trudności w sformułowaniu odpowiedzi na ostatnie pytania testowe, przerób ponownie materiał zawarty w 97, 98, 99.

6. Predykaty/orzeczniki/

1. Dalszym rozwinięciem rachunku zdań jest rachunek predykatów. Obejmuje on cały rachunek zdań, tj. proste zdania, przyjmując dwie wartości P i F oraz wszystkie operacje rachunku zdań. Prócz tego jednak w rachunku predykatów rozpatruje się zdania odnoszące się do przedmiotów. Tu już zdania rozczłonkowuje się na podmiot i predykat /orzecznik/.

2. Rozpatrzmy pewien zbiór przedmiotów M, i niech a,b,c, d - będą określonymi przedmiotami tego zbioru. Zdania o tych przedmiotach oznaczymy następująco:
P/a/, Q/a,b/, S/b,c,d/.

Przyjmijmy np. że zbiór M jest ciąg liczb naturalnych, a litery a,b,c,d odpowiednio liczbami 2,6,7,9. Wówczas P/b/ może być zdaniem "6 to liczba pierwsza", Q/a,b/ - "2 jest mniejsze od 6", S/b,c,d/ - "6,7,9, są liczbami parzystymi". Podobne zdania mogą być zarówno prawidłowe, jak i fałszywe. Będziemy je rozpatrywać, tak samo jak w rachunku zdań, tylko z tego punktu widzenia, czy są one prawdziwe czy fałszywe /P lub F/. Teraz jednak będziemy uważać, że wartości P i F przypisuje się określonym przedmiotom lub grupom przedmiotów. W powyższych przykładach zdanie P/b/ jest fałszywe w odniesieniu do 6, a Q/a,b/ jest prawdziwe w odniesieniu do pary 2 i 6.

3. Niech M będzie dowolnym zbiorem niepustym, x - dowolnym elementem tego zbioru.

Wówczas wyrażenie F/x/ oznacza zdanie, które staje się określone, gdy x zastąpi się konkretnym elementem z M. F/a/, F/b/ ... są już w pełni określonymi zdaniami.

Np. jeśli M jest ciągiem liczb naturalnych to zdanie F/x/ może oznaczać: "x jest liczbą pierwszą". Jest to zdanie nieokreślone. Stanie się ono określone, jeśli zastąpimy x pewną liczbą: "3 jest liczbą pierwszą", "6 jest liczbą pierwszą" itd. Niech S/x,y/ oznacza: "x jest mniejsze od y". To zdanie stanie się określone, gdy x i y zastąpi się konkretnymi liczbami: "2 jest mniejsze od 6", "9 jest większe od 2" itd.

4. Ponieważ każde zdanie przedstawia P lub F, F/x/ oznacza, że każdemu elementowi ze zbioru M jest przypisany odpowiednio jeden z dwóch symboli P lub F. Inaczej mówiąc F/x/ jest funkcją określoną na zbiorze M i przyjmującą tylko dwie wartości - P lub F. Tak samo nieokreślone zdania o dwóch i więcej przedmiotach S/x,y/, G/x,y,z/ itp. są funkcjami dwóch, trzech

i więcej zmiennych. Zmienne x,y,z ... przebiegają zbiór M, a funkcja przyjmuje wartości tylko P i F.

Takie nieokreślone zdania, funkcje jednej lub wielu zmiennych, nazywają się funkcjami logicznymi lub predykatami. Wykorzystując uprzednio wprowadzoną terminologię przy zestawianiu funkcji i stosunkach między elementami zbioru, odpowiednie logiczne funkcje często nazywa się jednomiejscowymi, dwumiejscowymi predykatami itd.

5. Jednomiejscowy predykat może wyrażać cechę przedmiotu, np. "x jest liczbą pierwszą", "y jest sygnałem wyjściowym" itd. Predykaty wielu zmiennych pozwalają wyrażać różne stosunki między przedmiotami. Niech np. M będzie zbiorem materiałów. Wówczas predykatami można wyrazić ich charakterystyki porównawcze: "x jest twardszy od y", "x i y to metale" itp.

6. Wszystkie wprowadzone pojęcia zawsze odnoszą się do pewnego dowolnego zbioru M, nazywanego polem. Elementy pola oznacza się małymi literami /niekiedy z indeksami/.

Nieokreślone elementy pola oznacza się literami z końca alfabetu:

$$x, y, z, u, v, x_1, x_2, \ldots$$

Nazywa się je zmiennymi swobodnymi /przedmiotowymi/.
Literami z początku alfabetu

$$a, b, c, d_1, d_2, \ldots$$

oznacza się określone elementy pola.

Nazywa się je przedmiotami indywidualnymi lub stałymi przedmiotowymi.
Podobnie jak w rachunku zdań, dużymi literami

$$A, B \ldots, X, A_1, A_2, \ldots$$

oznacza się zmienne, przyjmujące wartości P lub F.
Nazywa się je zdaniami zmiennymi.

7. Będziemy nazywać elementarnymi formułami zdania wyrażone dużymi literami, zarówno zmienne jak i stałe, a także wyrażenia

$$P/a/,\ S/a,b/,\ \ldots$$

gdzie P i S to predykaty, a i b - przedmioty indywidualne. Nazwę tę wprowadza się po to, żeby odróżnić takie formuły od złożonych, zestawionych z elementarnych.

8. Ponieważ formuły elementarne /oraz zdania i predykaty/ zawsze przyjmują tylko wartości P lub F, można je łączyć spójnikami,

$$\wedge,\ \vee,\ \rightarrow,\ -,\ \sim,$$

przy czym te operacje określa się tak samo jak w rachunku zdań. Otrzymane w ten sposób złożone formuły mogą określać zdania lub predykaty. Np.

$$A \vee F/x/;$$

$$A/x,y/ \rightarrow B \wedge \bar{A}/x_1,\ x_2/\ ;$$

$$G/x,y/ \rightarrow G/x_1,x_2/;$$

$$L/x/ \sim L/y/ \text{ itp.}$$

Pierwszy wzór przy stałych A i F/x/ określa pewien predykat. Czwarty wzór przy każdym L jest predykatem dwóch zmiennych x i y. Ten predykat przyjmuje wartość P przy x = y.

TEST

P. Co to są funkcje logiczne /predykaty/.

xxx

O. Nieokreślone zdania, funkcje jednej lub wielu zmiennych, przyjmujące wartości tylko P i F.

P. Podaj przykłady zapisów złożonych predykatów logicznych.

O. $A/x,y/ \rightarrow (B \wedge \bar{A}/x_1,x_2/)$;

$G/x,y/ \rightarrow G/x_1,x_2/$.

7. Operacje kwantyfikacji

1. Prócz rozpatrzonych operacji algebry logiki, w rachunku predykatów wprowadza się dwie specyficzne operacje - kwantyfikacje.

Niech Q/x/ będzie określonym predykatem, przyjmującym wartość P lub F dla każdego elementu $x \in M$.
Wówczas wyrażenie

$$\forall x\, Q/x/$$

będziemy nazywać prawdziwym, jeśli Q/x/ jest prawdziwe dla każdego elementu $x \in M$. Symbol nazywa się kwantyfikatorem ogólnym /dużym/ [19]. Wyrażenie $\forall x$ to kwantyfikator ogólny po zmiennej x. Przejście od predykatu Q/x/ do predykatu $\forall x$ Q/x/ nazywa się kwantyfikacją na predykat Q/x/ kwantyfikatora ogólnego po zmiennej x.

Przykład

Niech M będzie zbiorem procesów. Predykat Q/x/ dla $x \in M$ ma interpretację: "proces x przebiega w czasie". W rezultacie kwantyfikacji kwantyfikatora ogólnego mamy wzór $\forall$ xQ/x/, który interpretuje się następująco: "wszystkie procesy x przebiegają w czasie", co oczywiście jest prawdziwe.

2. Niech znowu Q/x/ będzie predykatem jednej zmiennej x. Połączymy z nim wzór

$$\exists x\, Q/x/,$$

przyjmujący wartość P, jeśli istnieje element pola M, dla którego Q/x/ jest prawdziwe, oraz wartość F, jeśli takie elementy nie istnieją.

Symbol $\exists$ nazywa się kwantyfikatorem /szczególnym/ małym, a wyrażenie $\exists$ x - kwantyfikatorem małym po zmiennej x [20].

[19] Znak $\forall$ został przyjęty jako odwróoona litera A z angielskiego określenia "for All".

[20] Znak $\exists$ został przyjęty jako odwrócona litera E z angielskiego określenia "Exist".

Przejście od Q/x/ do $\exists$ x Q/x/ nazywa się kwantyfikacją na predykat Q/x/ kwantyfikatora małego po zmiennej x.

Przykład

Niech M będzie zbiorem przyrządów pomiarowych. Predykat Q/x/ dla $x \in M$ można interpretować następująco:"x posiada indykację cyfrową". Wówczas przy kwantyfikacji kwantyfikatora $\exists$ otrzymamy wzór x Q/x/, który będzie się interpretować następująco: "istnieje x posiadający cyfrową indykację", co oczywiście jest prawdziwe.

8. Teoriomnogościowe pojęcie kwantyfikatorów

1. Przyjmijmy, że M jest zbiorem, dla którego określa się predykaty. Umówuliśmy się nazywać ten zbiór polem. Każdemu jednomiejscowemu predykatowi F/x/ można przypisać zbiór tych elementów $a \in M$, dla których F/a/ jest prawdziwe. Oznaczmy ten zbiór jako E_p. Z drugiej strony każdemu zbiorowi E, zawierającemu się w M, można przypisać predykat P/x/ przedstawiający zdanie $x \in E$. Predykat przyjmuje wartość P w E oraz F poza E. Tak więc E jest E_R.

2. Między podzbiorami M i predykatami jednej zmiennej, określonymi na M, istnieje stosunek wzajemnie jednoznaczny. Zbiór M przyjmuje się jako niepusty. Z teorii zbiorów wiadomo że sumą zbiorów

$$E_1 \cup E_2 \cup \ldots$$

nazywa się zbiór składający się z wszystkich elementów zbiorów składowych.

Iloczynem albo przekrojem zbiorów

$$E_1 \cap E_2 \cap \ldots$$

nazywa się zbiór wszystkich elementów należących jednocześnie do wszystkich zbiorów - czynników.

3. Ten fakt, że wielkości logiczne, zdania i predykaty zawsze przyjmują jednakowe wartości P lub F, oznaczymy jako $\equiv$. Równe zbiory będziemy łączyć znakiem $=$.
Niech

$$P/x/ \equiv P_1/x/ \vee P_2/x/$$

wówczas

$$E_R = E_{R_1} \cup E_{R_2}$$

Rzeczywiście, jeśli $x \in E_R$, to $P/x/$ jest prawdziwe; to znaczy, że $P_1/x/$ lub $P_2x/$ jest prawdziwe. W pierwszym przypadku $x \in E_{R_1}$, w drugim $x \in E_{R_2}$, stąd

$$x \in E_{R_1} \cup E_{R_2}$$

Odwrotnie, niech $x \in E_1 \cup E_{R_2}$. Wówczas $x \in E_{R_1}$ lub $x \in E_{R_2}$, tj. $P_1/x/$ jest prawdziwe lub $P_2/x/$ prawdziwe. Stąd $P/x/$ jest prawdziwe i $x \in E_r$.

Tak samo można wykazać, że jeśli

$$P/x/ \equiv P_1/x/ \wedge P_2/x/$$

to

$$E_R = E_{R_1} \cap E_{R_2}$$

4. Zbiór odpowiadający predykatowi $\bar{P}/x/$ jest dopełnieniem do zbioru odpowiadającego predykatowi $P/x/$. W symbolach stosowanych w teorii zbiorów można zapisać

$$E_{\bar{R}} = C E_R$$

gdzie CE_R - zbiór elementów pola M, nie należących do E_R, lub inaczej dopełnienie zbioru E_R .

5. Formuły rachunku zdań oznaczone dużymi literami można traktować jako predykaty zachowujące dla wszystkich przedmiotów tę samą wartość P lub F. Zgodnie z określonymi wyżej warunkami powinniśmy przypisać takim predykatom w pierwszym

przypadku całe pole M, a w drugim przypadku zbiór pusty. Ponieważ między funkcjami logicznymi a zbiorami ustanowiona została odpowiedniość, to prawom rachunku predykatów odpowiadają znane prawa dla operacji z teorii zbiorów.

Np. pierwszemu prawu rozdzielności

$$F/x/ \vee \left[G/x/ \wedge H/x/\right] = \left[F/x/ \vee G/x/\right] \wedge \left[F/x/ \vee H/x/\right]$$

odpowiada prawo z teorii zbiorów

$$P \cap (Q \cup S) = (P \cap Q) \cup (P \cap S)$$

gdzie P,Q,S są dowolnymi zbiorami.

Drugiemu prawu rozdzielności

$$F/x/ \wedge \left[G/x/ \vee H/x/\right] \equiv \left|F/x/ \wedge G/x/\right| \vee \left|F/x/ \wedge H/x/\right|$$

odpowiada prawo z teorii zbiorów

$$P \cup (Q \cap S) = (P \cup Q) \cap (P \cup S).$$

6. W podobny sposób można ustalić związek między zbiorami a predykatami wielu zmiennych. Jako przykład rozpatrzmy przypadek dwumiejscowego predykatu. Niech $M^{/2/}$ będzie zbiorem wszystkich par $/x,y/ \in M$. Przy tym pary różnią się nie tylko składem, lecz i porządkiem elementów.

Predykatowi $P/x,y/$ przypiszemy zbiór par $/x,y/ \in M^{/2/}$, dla których $P/x,y/$ jest prawdziwe. Oznaczmy ten zbiór jako E_R.

7. Rozpatrzmy pojęcie kwantyfikatorów w teorii zbiorów. Niech

$$F/x/ \equiv \exists\, xP/x,y/$$

Zbiór E_p odpowiadający predykatowi P składa się tylko z tych elementów pola M, dla których $f/x/$, tj. $\exists\, yP/x,y/$ jest prawdziwe. Ostatnie wyrażenie jest prawdziwe dla danego x_0, jeśli istnieje takie y, że $P/x_0,y/$ jest prawdziwe. Funkcji $P/x,y/$ odpowiada część E_R zbioru $M^{/2/}$. Tak więc E_p składa się z wszystkich elementów $x \in M$ i dla każdego z nich można znaleźć parę $/x,y/$ należącą do E_R.

Będziemy nazywać x_0 rzutem dowolnej pary $/x_0, y/$, a rzutem zbioru - zbiór projekcji należących do niego par. Jasne jest, że E_p jest rzutem E_R.

Niech M będzie zbiorem liczb rzeczywistych. Zbiór $M^{/2/}$ można interpretować jako płaszczynę, której punkty posiadają współrzędne x,y; a M jako oś OX tej płaszczyzny. Wówczas punkt x jest projekcją /rzutem/ punktu /x,y/ bezpośrednio w znaczeniu geometrycznym. Dlatego E_p odpowiadające predykatowi F/x/, równemu $\exists$ y P/x,y/ pokrywa się z rzutem ortogonalnym zbioru E_R na osi OX. Rzut zbioru H na M oznaczymy jako pr_xH. Wówczas

$$E_p = pr_x E_R$$

W celu ustalenia pojęcia /z teorii zbiorów/ kwantyfikatora ogólnego posłużyły się operację zaprzeczenia.

Niech

$$F/x/ \equiv \forall\ yP/x,y/$$

Wówczas

$$\forall\ y\ P/x,y/ = \overline{\exists\ y\bar{P}/x,y/}$$

8. Logiczna operacja zaprzeczenia odpowiada operacji dopełnienia w teorii zbiorów. Możemy zapisać

$$E_p = C\ pr_x\ CE_R$$

tj. zbiór odpowiadający funkcji $\forall$ y P/x,y/ jest dopełnieniem do rzutu na M, dopełnienia do E_R.

Prawdziwe jest również twierdzenie odwrotne. Każdy zbiór R będący rzutem na M zbioru U należącego do $M^{/2/}$

$$R = pr_x\ U$$

można przedstawić jako E_p, gdzie F/x/ jest $\exists$ yP/x,y/, przy czym $U = E_R$. Rzeczywiście zbiorowi R odpowiada predykat F/x/, określony na M, a zbiorowi U - predykat P/x,y/ określony na $M^{/2/}$, oraz oczywiście

$$F/x/ \equiv \exists\ yP/x,y/$$

Jeśli jednak R′będzie dopełnieniem do rzutu U, to R′odpowiada predykatowi $\forall$ y$\bar{P}$/x,y/. Rzeczywiście, predykatowi

$\forall y \bar{P}/x,y/$ odpowiada zbiór $Cpr_x\ CE_{\bar{R}} = E_P = U$, stąd

$Cpr_x\ CE_{\bar{R}}$, lecz $CE_R = E_R = U$,

Tak więc kwantyfikatory są związane z geometryczną operacją rzutowania.

TEST

P. Podaj przykłady zapisu predykatów z zastosowaniem operacji kwantyfikacji.

xxx

O. $\forall x\ Q/x/$

$\exists x\ Q/x/$.

Rozdział VI

ELEMENTY TEORII INFORMACJI

1. Nieokreśloność i entropia

1. Z całości tej książki, a w szczególności z jej początkowych rozdziałów wynikało, że podstawowym przedmiotem badań cybernetyki są procesy przekazywania, przetwarzania i obróbki informacji. Układy i systemy cybernetyczne są wyodrębnione pod względem informacyjnym, ich wejścia i wyjścia są kanałami przekazywania informacji, a ich rola polega na przekształcaniu informacji wejściowej w informację wyjściową według odpowiednich reguł.

2. Tymczasem do tej pory nie powiedzieliśmy ani słowa o podstawowym pojęciu - o samej informacji. Co więcej - nawet w tej chwili nie podamy definicji pojęcia informacji, po prostu dlatego, że pojęcie to nie zostało dotychczas w sposób zadowalający zdefiniowane. Fakt ten może wywołać zrozumiałe wzburzenie, gdyż wynika z niego, że cały czas zajmowaliśmy się - nie wiadomo dokładnie czym. Nie jest to jednak prawdą, gdyż pojęcie informacji, dotychczas nie zdefiniowane, jest jednak na tyle dokładnie określone i intuicyjnie zrozumiałe, że można o nim mówić bez obawy nieporozumień.

3. Trudności w zdefiniowaniu pojęcia informacji wynikają z faktu, że informacja jest wieloaspektowa i trudno zdecydować się, który aspekt wziąć za podstawę przy formułowaniu definicji. Informacja ma mianowicie określoną objętość, wyrażającą się niezbędną ilością symboli alfabetu lub sygnałów w urządzeniach technicznych, konieczną do jej przekazania. Ma ona określone znaczenie, wyrażające się zwykle określoną

zmianą działania odbiorcy. Ma ona również określoną wartość polegającą na tym, że pewne informacje mają dla odbiorcy lub nadawcy doniosłe znaczenie, a inne są pospolite lub nieważne.

4. Nie będziemy więc wdawali się w zawiłe spory terminologiczne, które toczą latami naukowcy, dążąc do sprecyzowania odpowiedzi na najprostsze pytanie: "Co to jest informacja?" Zajmiemy się za to kilkoma aspektami praktycznymi, związanymi z pojęciem informacji, a mianowicie: podamy miarę pozwalającą określać ilość informacji, podejmiemy dyskusję nad problemem najekonomiczniejszego przedstawienia informacji przy pomocy symboli oraz rozważymy, co się dzieje z informacją przesłaną kanałem, w którym istnieją zakłócenia.

Nie zajmiemy się zagadnieniem skuteczności informacji, wyrażającej się jej wpływem na działalność odbiorcy, gdyż trudno jest określić chwilowo miarę "zmiany działania odbiorcy". Podobnie abstrahować będziemy od problemów adekwatności przesyłanych symboli do określonych treści /semantyka/ i wartości informacji dla odbiorcy, gdyż zarówno w jednym, jak i w drugim przypadku, pojęcia leżące u podstaw są zbyt subiektywne i matematycznie niewymierne.

5. Zaczniemy od próby dania miary ilości informacji. Zaczniemy od dość trywialnie wyglądającego stwierdzenia, że ilość informacji równa jest wielkości niepewności, jaka została usunięta w wyniku otrzymania informacji. Jeśli się na takie postawienie sprawy zgodzimy, to przerzucamy problem ilości informacji w rejony łatwiejsze do operowania matematyką, a mianowicie w rejony mierzenia stopnia niepewności, chaosu, bezładu.

6. Aparatem matematycznym najbardziej odpowiednim do operowania zdarzeniami nieokreślonymi jest rachunek prawdopodobieństwa. Nim też właśnie będziemy się posługiwali. O konieczności minimalnego przygotowania matematycznego ze strony Czytelnika nie wspominamy, ponieważ już trzeci raz pojawia się rachunek prawdopodobieństwa w naszej książce.

7. Sformułujemy wymagania, jakim powinna odpowiadać poszukiwana przez nas obecnie miara nieokreśloności. Po pierwsze, jeśli jakieś wyrażenie występuje z prawdopodobieństwem p /czyli mamy "szansę" p, że wydarzenie to wystąpi/, to nasza niepewność co do wyniku jest tym większa, im zdarzenie to jest mniej prawdopodobne, czyli im mniejsze jest p. Zapiszemy to w postaci warunku:

$$H/p_1/ \quad H/p_2/ \quad \text{jeśli} \quad p_1 \quad p_2 \qquad /126/$$

$H/p/$ jest oznaczeniem niepewności co do wyniku doświadczenia o prawdopodobieństwie sukcesu p.

Przykład 1

Jeśli mamy n jednakowo prawdopodobnych zdarzeń, to prawdopodobieństwo zajścia jednego z nich wynosi $p = 1/n$ i jest tym mniejsze, im "konkurentów" jest więcej. W tej sytuacji niepewność co do wyniku jest tym większa, im więcej mamy zdarzeń "do wyboru", a więc im mniejsze jest prawdopodobieństwo p, zgodnie z relacją /126/.

8. Drugie wymaganie będzie bardziej skomplikowane, choć równie oczywiste. Jeśli mamy mianowicie dwa zdarzenia, z których jedno zachodzi z prawdopodobieństwem p_1 a drugie z prawdopodobieństwem p_2, to chcielibyśmy, aby nieokreśloność zdarzenia łącznego, polegającego na zajściu zarówno zdarzenia 1 jak i 2, była równa sumie nieokreśloności zdarzeń 1 i 2. Prawdopodobieństwo zdarzenia łącznego wynosi /zgodnie z zasadami rachunku prawdopodobieństwa/ p_1p_2, zatem wymaganie nasze zapiszemy w postaci:

$$H/p_1p_2/ = H/p_1/ + H/p_2/ \qquad /127/$$

Jest to tzw. postulat przeliczalnej addytywności miary nieokreśloności.

Przykład 2

Przypuśćmy, że jedno wydarzenie polega na wyborze jednego przedmiotu z n możliwych, a drugie jednego przedmiotu z m możliwych. Nieokreśloności tych wydarzeń są tym większe, im większe są ilości n i m, zgodnie z wnioskiem z przykładu 1. Ilość możliwych kombinacji dwuelementowych, złożonych z jednego przedmiotu z pierwszego zbioru i jednego z drugiego zbioru wynosi m, zatem prawdopodobieństwo wybrania danego zestawu dwu elementów wynosi 1/nm, a nieokreśloność tego wydarzenia jest tym większa, im większy jest ten iloczyn. Postulat przeliczalnej addytywności wymaga, aby nieokreśloność ta była równa sumie nieokreśloności pojedynczych wyborów ze zbioru pierwszego i drugiego.

9. Ostatnie wymaganie jest już zupełnie proste. Chcielibyśmy, aby w przypadku, gdy zdarzenie jest pewne /p=1/, nieokreśloność wynosiła zero:

$$H/1/ = 0 \qquad /128/$$

Przykład 3

Jeżeli mamy tylko jedno wydarzenie możliwe, to mamy pewność, że ono zajdzie i nieokreśloność tego wydarzenia słusznie możemy uważać za zerową.

10. Można udowodnić, że jedyną zależnością matematyczną spełniającą postulaty /127/ i /128/ jest zależność logarytmiczna. Można się było tego spodziewać po postulacie /127/, ponieważ funkcja H zamieniała iloczyn na sumę, a to jest właściwość funkcji logarytmicznych. Ponadto funkcja logarytmiczna dla jedynki przybiera wartość zero, czyli spełnia postulat /128/. Gorzej z postulatem /126/, gdyż stwierdza on, że funkcja H powinna maleć, a wiadomo, że logarytm jest funkcją rosnącą. Zatem nieokreśloność określimy jako m i n u s logarytm prawdopodobieństwa:

$$H/p/ = - \log p \qquad /129/$$

Przykład 4

Dla rozważanych w przykładach przypadków zbiorów jednakowo prawdopodobnych wydarzeń wzór /129/ można przekształcić:

$$H/p/ = - \log \frac{1}{n} = \log n = H/n/ \qquad /130/$$

W zależności /130/wynika, że nieokreśloność wyboru jednego z n elementów jest równa logarytmowi ilości elementów. Ta postać funkcji nieokreśloności była odkryta wcześniej niż postać /129/ przez Hartleya i przez długi czas uznawano ją za jedyną. Potwierdziły jej słuszność badania psychologów, którzy eksperymentalnie stwierdzili, że stopień trudności decyzji typu "wybór z n jednakowych elementów" jest proporcjonalny do logarytmu ilości elementów.

11. Wzór /129/ jest słuszny niezależnie od podstawy logarytmu w nim występującego. W zależności od podstawy będziemy jedynie mierzyli wielkość nieokreśloności w różnych jednostkach.
Ze względów teoretycznych najodpowiedniejszymi byłyby logarytmy naturalne /o podstawie tzw. liczby e/, lecz taka miara nieokreśloności nie przyjęła się/. Jednostką tej miary był tzw. "nit"/.
Ze względów praktycznych /łatwo dostępne tablice/ dogodne byłyby logarytmy dziesiętne, lecz i te się nie przyjęły /"dziesiętna" jednostka nieokreśloności nazywa się hartley/. Przyjęto powszechnie logarytmy o podstawie 2. Zadecydowała ła - twość interpretacji: jednostkową jest nieokreśloność wyboru pomiędzy dwoma jednakowo prawdopodobnymi sytuacjami. Jednostką nieokreśloności przy dwójkowej podstawie logarytmu jest BIT /od angielskiego binary digit/. W dalszym ciągu będziemy posługiwali się wyłącznie bitami.

12. Wzór /129/ określa miarę nieokreśloności dla zdarzeń pojedynczych, które mogą zachodzić z prawdopodobieństwem p. Tymczasem zwykle spotykamy się z sytuacją, kiedy mamy kilka możliwych zdarzeń, zachodzących odpowiednio z prawdopodobieństwami $p_1, p_2, \ldots, p_n$. W tak postawionym zadaniu możemy mówić o nieokreśloności każdego zdarzenia osobno, na ogół jednak interesuje nas nieokreśloność całej tej sytuacji.

Przykład 5

Jeżeli mamy jakieś urządzenie np. automat skończony, które może wysyłać określone symbole wyjściowe, przy czym znamy jedynie prawdopodobieństwa wysyłania poszczególnych symboli, to możemy interesować się nieokreślonością tego, co pojawi się na wyjściu. Jeśli ustalimy miarę takiej złożonej nieokreśloności, to możliwa stanie się odpowiedź na pytanie, który z dwu rozpatrywanych automatów probabilistycznych jest bardziej nieokreślony w swoich reakcjach.

13. Wprowadzimy więc pojęcie średniej nieokreśloności zbioru symboli, z których każdy może pojawić się z prawdopodobieństwem $p_1, p_2, \ldots, p_n$. Skorzystamy tu z ogólnie znanego wzoru na wartość średnią.
Otrzymamy wówczas:

$$H/A/ = - \sum_{i=1}^{n} p_i \log_2 p_i \qquad /131/$$

H/A/ oznacza średnią nieokreśloność zdarzenia A /pojawienie się jednego z n symboli o prawdopodobieństwach $p_1, \ldots, p_n$/.

14. Średnia nieokreśloność wyrażona wzorem /131/ wprowadzona została do cybernetyki przez C.Shannona w 1948 roku i fakt ten leży u podstaw narodzin niezwykle interesującej technicznie i filozoficznie dziedziny, zwanej w całości teorią informacji. Wielkość H/A/ wyraża się w jednostkach bit/symbol /ponieważ wyraża średnią ilość nieokreśloności w bitach, przypadającą na jeden symbol/. Ze względu na pewne dość głębokie analogie z pojęciem ENTROPII wprowadzonym w termodynamice, wielkość H/A/ nazywa się również zwykle entropią.

15. Entropia jest miarą bezładu i stopnia przypadkowości. Można się przekonać, aczkolwiek ścisły dowód jest dość długi, że entropia przybiera w danym, n-elementowym zbiorze zdarzeń /symboli/ wartość maksymalną w przypadku, kiedy wszystkie zdarzenia są jednakowo prawdopodobne. Wówczas $p_1 = p_2 = \ldots$ $\ldots = p_n = 1/n$ i po wstawieniu do wzoru /131/ otrzymujemy

$$H_{max}/A/ = - \sum_{i=1}^{n} \frac{1}{n} \log_2 \frac{1}{n} = \log_2 n \qquad /132/$$

Z porównania wzorów /132/ i /130/ wynika, że miara nieoznaczoności wprowadzona przez Hartleya [21] jest maksymalną wartością entropii w przypadku, kiedy wszystkie symbole są jednakowo prawdopodobne. Jest to logiczne, gdyż jakiekolwiek zróżnicowanie prawdopodobieństw wprowadza uporządkowanie, które musi zmniejszać entropię. Przy skrajnie dużym uporządkowaniu, jedno z wydarzeń jest pewne a inne niemożliwe, entropia jest równa zeru./Człony w sumie, odpowiadające $p_i = 0$ zerują się dzięki pierwszemu czynnikowi, a człony dla $p_j = 1$ zerują się dzięki drugiemu czynnikowi, w rezultacie wzór /131/ zamienia się w sumę samych zer/.

16. Druga zasada termodynamiki mówi, że w układzie fizycznie odosobnionym entropia może tylko wzrastać. Po przetłumaczeniu na język cybernetyki oznacza to, że każdy układ odosobniony będzie dążył do coraz większego bezładu. Świat nasz jest niewątpliwie układem odosobnionym; jak daleko sięga obserwacja w głąb kosmosu, wszędzie zachodzą procesy wzrostu entropii. Zatem świat dąży do powszechnego chaosu. Przejawem tej tendencji są zarówno gigantyczne erupcje energii gwiazd, jak i mieszanie się dwu gazów zawartych w jednym naczyniu lub rozpadanie się budowlni wzniesionych przez Człowieka. Na skutek tego powszechnego prawa każda struktura, każda organizacja poddawana jest wpływom otoczenia dążącym do jej dezorganizacji. Są jednak we Wszechświecie enklawy względnie odosobnione, które idą pod prąd i dążą do zwiększenia swej organizacji wewnętrznej. Są to organizmy żywe, a wśród nich my sami.

Proces powstawania nowego nowego organizmu, kiedy to z niekształtnego zarodka tworzy się wielokroć od niego wyżej zorganizowany osobnik dorosły jest wydarzeniem, którego powszechność w naszym otoczeniu maskuje nieco jego unikalność w skali kosmosu. Proces ten jednak odbywa się kosztem wielekroć większych wzrostów entropii w otoczeniu, z którym organizm kontaktuje się, tak że ogólny bilans wskazuje jedyny możliwy w naszym świecie kierunek - ku kompletnej anarchii.

21/ R. Hartlej: Transmission of information. Bell System Techn., f. vol. 7, nr 3, 1928.

Zresztą proces dezorganizacji nie oszczędza i organizmów żywych, dlatego starzejemy się i umieramy. Jednak przeżywa i odradza się wciąż na nowo gatunek, jedyny twór, znany cybernetyce, nie podlegający prawu wzrostu entropii. Prawu, którego konsekwencje są chyba jeszcze bardziej pesymistyczne niż twierdzenia Gödla.

2. Informacja

1. Skoro potrafimy określić ilościową miarę niewiedzy /nieokreśloności/, możemy też określić ilość informacji, przyrównując ją do ilości nieokreśloności, jaką informacja ta usunęła. Mówiąc prościej, jeśli początkowo pewna sytuacja charakteryzowała się pewną nieokreślonością, a następnie otrzymaliśmy wiadomość, która usun;ła nieokreśloność /H/A/ po otrzymaniu wiadomości wynosi 0/, to powiemy, że ilość informacji otrzymanej z tą wiadomością /w bitach, nitach lub hartleyach/ jest równa ilościowej mierze owej początkowej nieokreśloności.

TEST 1

P. Jaką ilość informacji / w bitach/ otrzymamy uzyskując odpowiedź na pytanie, posiadające jedynie dwie możliwe /i jednakowo a priori prawdopodobne/ odpowiedzi?

xxx

O. Jeden bit.

Jeśli nie odpowiedziałeś, musisz niestety wracać do poprzedniego rozdziału i przeczytać jeszcze raz punkty 5, 7, 8, 9, 10 i 11.

P. Ile pytań trzeba postawić w grze, polegającej na odpowiadaniu jedynie "tak" lub "nie", aby uzyskać od przeciwnika 3 bity informacji?

xxx

O. Trzy.

Jeśli masz wątpliwości, przypomnij sobie postulat zawarty w punkcie 8 poprzedniego rozdziału.

P. Ile informacji niosłaby jedna litera alfabetu, gdyby przyjąć, że wszystkie litery występują z jednakowym prawdopodobieństwem i że jest ich 32 /odstęp między literami jest również literą/?

xxx

O. Pięć bitów /$\log_2 32 = \log_2 2^5 = 5$/ Zastosowaliśmy wzór 7 ze względu na postulat jednakowego prawdopodobieństwa liter.

P. Ile informacji wnosiłby przy jednakowo prawdopodobnych literach tekst 30 liter? Czy ilość informacji rzeczywiście wnoszona przez ten tekst odpowiada obliczonej? Dlaczego?

xxx

O. Ze względu na postulat z punktu 8 poprzedniego rozdziału entropie, a więc i ilości informacji wnoszone przez pojedyncze symbole można sumować, czyli omawiany tekst wnosi $5 \cdot 30 = 150$ bitów informacji. W rzeczywistości wnosi jednak daleko mniej, gdyż litery nie występują w tekście zupełnie przypadkowo, z jednakowymi prawdopodobieństwami, a wprost przeciwnie, wystąpienie na przykład litery "a" jest w polskim tekście dużo bardziej prawdopodobne, niż wystąpienie litery "x", a więc istnieje pewna aprioryczna organizacja, zmniejszająca nieokreśloność, a tym samym zmniejszająca ilość informacji wnoszoną przez tekst.

2. Tak więc możemy ustalić ilość informacji wnoszoną przez komunikat, jeśli w wyniku odebrania tego komunikatu usunięta została nieokreśloność odpowiedniej wartości. Zapiszemy to prostym wzorem:

$$I/A/ = H/A/ \qquad /133/$$

Wzór ten czytamy: informacja o A jest równa początkowej niepewności co do A.

Możliwa jest jednak sytuacja, w której otrzymany komunikat nie znosi niepewności, lecz ją zmienia, na przykład wówczas, gdy komunikat taki zmienia odpowiednie prawdopodobieństwa symboli. Jaką ilość informacji wnosi taki komunikat?

Aby odpowiedzieć na to pytanie wprowadzimy pojęcie entropii warunkowej.

3. Omówimy najpierw pojęcie entropii zdarzenia złożonego z dwu zdarzeń. Jeżeli dwa rozpatrywane zdarzenia są niezależne, entropia zdarzenia złożonego, na podstawie postulatu punktu 8 poprzedniego rozdziału, jest równa sumie entropii obu zdarzeń:

$$H/AB/ = H/A/ + H/B/ \qquad /134/$$

Dzieje się tak w przypadku, gdy prawdopodobieństwo zdarzenia złożonego AB jest równe iloczynowi prawdopodobieństw tych zdarzeń /jeśli zdarzenia są niezależne/.

Przykład 1

Gdy tworzymy dwuliterową nazwę wybierając litery w sposób zupełnie przypadkowy i z jednakowymi prawdopodobieństwami, to prawdopodobieństwo wybrania określonej litery wynosi $\frac{1}{32}$, zarówno dla pierwszej, jak i dla drugiej litery. Natomiast prawdopodobieństwo danego napisu dwuliterowego jest równe iloczynowi tych prawdopodobieństw, jako że litery są wybierane zupełnie przypadkowo, i wynosi $\frac{1}{32} \cdot \frac{1}{32}$

Entropia w takim wypadku wynosi $H/A/ + H/B/ = 5 + 5 = 10$ bitów.

4. Jeśli jednak omawiane dwa zdarzenia będą uzależnione od siebie nawzajem, to znaczy jeśli fakt, że znamy pierwszy symbol, będzie zmieniał rozkład prawdopodobieństw, z jakimi będziemy wybierali drugi, to możemy mówić, że prawdopodobieństwo zdarzenia łącznego, polegającego na pojawieniu się zespołu tych dwu symboli; będzie równe iloczynowi prawdopodobieństwa pierwszego symbolu przez prawdopodobieństwo warunkowe symbolu, pod warunkiem, że wiemy, jaki jest pierwszy symbol /jest to jeden z podstawowych faktów ustalowych w klasycznym rachunku prawdopodobieństwa, którego obszerne wyjaśnienie

znajdzie Czytelnik w każdej monografii z tej dziedziny/.

Taka zmiana rozkładu prawdopodobieństwa drugiego zdarzenia na skutek określonej realizacji pierwszego zdarzenia nie może nie odbić się na łącznej nieokreśloności zdarzenia łącznego. Zbadajmy to bliżej.

5. Niech pierwsze z dwu zdarzeń, które tradycyjnie nazywamy A, ma n realizacji możliwych: $a_1, a_2 \ldots, a_n$ przyjmowanych z prawdopodobieństwami $p_{a1}, p_{a2}, \ldots, p_{an}$, zaś drugie niech ma m możliwych realizacji: $b_1, b_2, \ldots, b_m$, występujących z prawdopodobieństwami $p_{b1}, p_{b2}, \ldots, p_{bm}$, /gdy występują samodzielnie/, oraz z prawdopodobieństwami $p_{b1/ai}, p_{b2/ai}, \ldots, p_{bm/ai}$, pod warunkiem, że poprzednio nastąpiła realizacja numer "i" zdarzenia A.

Zdefiniujemy teraz tzw. entropię warunkową zdarzenia B, pod warunkiem, że zdarzenie A przyjęło realizację a_i. Będzie to oczywiście suma:

$$H/B/a_i/ = \sum_{i=1}^{m} p_{bj}/a_i \cdot \log_2 p_{bj}/a_i \qquad /135/$$

Suma ta naturalnie wynika ze wzoru /131/ poprzedniego rozdziału, jeśli zamiast odpowiednich prawdopodobieństw bezwzględnych podstawić warunkowe.

Można teraz określić pełną entropię warunkową zdarzenia B, pod warunkiem, że zaszło A, biorąc średnią /wartość oczekiwaną/ wszystkich wyrażeń typu /135/ przy wszystkich możliwych realizacjach zdarzenia A:

$$H/B/A/ = \sum_{i=1}^{n} p_{a_i} \cdot H/B/a_i/ \qquad /136/$$

lub, po dokonaniu odpowiednich podstawień:

$$H/B/A/ = \sum_{i=1}^{n} p_{a_i} \sum_{j=1}^{m} p_{bj/a_i} \cdot \log_2 p_{bj/a_i} \qquad /137/$$

6. Wykorzystując wzór /137/ możemy teraz zapisać entropię zdarzenia złożonego z dwu zdarzeń, które mogą być od siebie uzależnione.

Nie będziemy wzoru tego wyprowadzali, gdyż wyprowadzenie to jest żmudne i długie, a jednocześnie nieskomplikowane, bo polegające jedynie na przekształceniach i grupowaniu symboli. Posłużymy się zamiast tego pewną analogią.

W przypadku dwu zdarzeń niezależnych prawdopodobieństwo zdarzenia łącznego jest równe iloczynowi prawdopodobieństw, a entropia zdarzenia łącznego jest równa sumie entropii:

$$p/AB/ = p/A/ \cdot p/B/ \qquad H/AB/ = H/A/ + H/B/ \qquad /138/$$

W przypadku dwu zdarzeń zależnych prawdopodobieństwo zdarzenia łącznego jest równe iloczynowi prawdopodobieństwa zdarzenia pierwszego przez prawdopodobieństwo zdarzenia drugiego, pod warunkiem, że pierwsze zaszło. Entropia zdarzenia łącznego będzie w tym wypadku równa sumie entropii zdarzenia pierwszego i entropii warunkowej zdarzenia drugiego, pod warunkiem, że pierwsze zaszło:

$$p/AB/ = p/A/\ p/B|A/ \qquad H/AB/ = H/A/ + H/B|A/ \qquad /139/$$

Wzór powyższy łącznie ze wzorem /137/ pozwoli nam w następnym rozdziale sformułować twierdzenie o granicznej informacji wzajemnej.

7. Zajmiemy się teraz zagadnieniem, które interesowało nas na początku tego rozdziału: jaką ilość informacji otrzymujemy, jeśli w wyniku komunikatu zmieniają się prawdopodobieństwa poszczególnych realizacji zdarzenia A?

Początkowy rozkład prawdopodobieństw zdarzenia A powodował, że miało ono pewną entropię H/A/, którą łatwo obliczyć ze wzoru /138/ poprzedniego rozdziału. Nastąpiło wydarzenie B, w wyniku czego poszczególne prawdopodobieństwa realizacji zdarzenia A zamienione zostały przez odpowiednie prawdopodobieństwa warunkowe /p/A | B/. Nowa entropia zdarzenia A wynosi teraz H/A | B/. W wyniku otrzymania komunikatu B, nasza niepewność co do zdarzenia A zmieniła się z H/A na H/A | B/, słusznym więc będzie przypisanie B ilości informacji wynoszącej

$$I/B|A/ = H/A/ - H/A|B/ \qquad /140/$$

Wzór ten odczytamy: ilość informacji zawarta w zdarzeniu B o zdarzeniu A równa jest różnicy pomiędzy początkową nieokreślonością zdarzenia A, a nieokreślonością zdarzenia A, jaka pozostaje po ustaleniu wyniku zdarzenia B.

8. Istnieją dwie ważne konsekwencje praktyczne wzoru /140/. Entropia warunkowa zdarzenia A pod warunkiem B nie może być wyższa, niż antropia bezwarunkowa zdarzenia A

$$0 \leqslant H/A \,|\, B/ \leqslant H/A/ \qquad /141/$$

z czego wynika, że nie istnieje ujemna wartość ilości informacji określonej wzorem /140/, można co najwyżej nie uzyskać informacji w ogóle, co następuje przy wypełnieniu warunku /141/ z równością.

Entropia jest zawsze nieujemna, co najwyżej zerowa, więc ilość informacji, jaką możemy otrzymać o zdarzeniu A nie może być większa od początkowej nieokreśloności zdarzenia A. Może być jednak mniejsza przy nieodpowiednim doborze wiadomości B, o czym przekonamy się w następnym rozdziale.

3. Twierdzenie o granicznej informacji wzajemnej

1. Zaczniemy ponownie od stwierdzenia, którego nie będziemy dowodzili, lecz odwołamy się do pewnej analogii z elementarnego rachunku prawdopodobieństwa.

Prawdopodobieństwo zdarzenia złożonego równe jest iloczynowi prawdopodobieństwa pierwszego zdarzenia i prawdopodobieństwa warunkowego drugiego zdarzenia, pod warunkiem, że pierwsze zaszło, lub iloczynowi prawdopodobieństwa drugiego zdarzenia razy prawdopodobieństwo pierwszego, pod warunkiem, że drugie zaszło. Zapiszemy to wzorem:

$$p/AB/ = p/A/\; p/B/A/ = p/B/\; p/A/B/ \qquad /142/$$

Wzór ten wskazuje na istnienie symetrii zdarzeń we wzorze na prawdopodobieństwo zdarzenia złożonego, wzór ten jest prawdziwy niezależnie od tego, które ze zdarzeń uznamy za pierwsze, a które za drugie /uzależnione od poprzedniego/. Na tej zasadzie możemy rozszerzyć wzór /132/ z poprzedniego paragrafu:

$$H/A \;.\; B/ = H/A/ + H/B/A/ = H/B + H/A/B \qquad /143/$$

Wynika z tego, że nieokreśloność zdarzenia złożonego możemy otrzymać przez zesumowania entropii jednego wydarzenia z warunkową entropią drugiego wydarzenia, przy czym obojętne jest, które ze zdarzeń uznamy za pierwsze.

2. Tymczasem we wzorze na ilość informacji /140/ z poprzedniego paragrafu zdarzenia A i B nie są dowolne, lecz mają określone znaczenie: A jest sytuacją, o której chcemy się czegoś dowiedzieć, a B jest komunikatem, który otrzymujemy. Spróbujmy wzór ten poprzekształcać

$$I/B \mid A/ = H/A/ - H/A \mid B/ = H/A/ - H/AB/ + H/B/ =$$

$$= H/A/ - H\,A - H/B \mid A/ + H/B/ = H/B/ - H/B \mid A/$$

W przekształceniach tych dwukrotnie korzystaliśmy ze wzoru /143/. Ostateczny rezultat tych przekształceń ma znajomą postać. Oczywiście, jest to wyrażenie na ilość informacji o zdarzeniu B, zawartej w A:

$$H/B/ - H/B \mid A/ = I/A \mid B/ \qquad /144/$$

Cóż więc wynika ze wzoru /140/ poprzedniego paragrafu, powyższego wzoru /144/ i przeprowadzonych wyżej przekształceń?

$$I/B \mid A/ = H/A/ - H/A \mid B/ = H/B/ - H/B \mid A/ = I/A \mid B/ \qquad /145/$$

Wynik jest tyleż oczywisty, co zaskakujący. Okazuje się, że ilość informacji, zawarta w komunikacie i dotycząca zdarzenia, jest dokładnie równa ilości informacji zawartej w zdarzeniu o komunikacie. Istotnie. jeśli komunikat zawiera jakąś prawdę o zdarzeniu /a tylko wtedy wnosi niezerową informację/, to musi być od tego zdarzenia uzależniony. Zatem znając zdarzenie możemy z równym powodzeniem wypowiadać się o komunikacie, jak znając komunikat, wypowiadać się o procesie.

3. Wzór /145/ wyraża zasadniczy sens twierdzenia o granicznej informacji wzajemnej, które nieco popularniej można streścić stwierdzeniem, że dla danej pary: komunikat - zdarzenie, ilość informacji jest stała, niezależnie od punktu widzenia. Twierdzenie to ma doniosłą wagę teoretyczną, ale i praktycznie można z niego wyciągnąć ciekawe wnioski. Przypomnijmy sobie mianowicie, że na zakończenie poprzedniego rozdziału stwierdziliśmy, że ilość informacji, jaką możemy uzyskać o zdarzeniu A nie może przewyższać jego entropii /wzór 134/ poprzedniego paragrafu/, jednak pod pewnym tylko warunkiem może tę wartość osiągnąć. Warunek ten wynika z drugiej połowy wzoru /145/. Łatwo zauważyć, że ilość informacji I/A B/ nie może przekroczyć wartości H/B/, a osiąga ją jedynie w przypadku, gdy H/B A/ = 0, co w praktyce oznacza, że wynik doświadczenia A w pełni określa charakter komunikatu /zdarzenia B/. Wynika z tego /po uwzględnieniu i lewej połowy wyrażenia /145/, że ilość informacji zawarta w komunikacie nie może przekraczać jego własnej entropii. Jest to niezmiernie istotne stwierdzenie. Nie możemy więc przesłać określonej ilości informacji w komunikacie, którego entropia jest mniejsza od ilości przesłanej informacji. Stwierdzenie to bywa podawane, w oderwaniu od twierdzenia o granicznej informacji wzajemnej, jako tak zwane prawo dostatecznej różnorodności /entropia jest miarą różnorodności, to tylko nowa nazwa/. Prawem tym szafują szczególnie specjaliści amerykańscy, chyba nieco przeceniając jego znaczenie.

4. Przedyskutujemy jeszcze jeden aspekt twierdzenia /145/, który pozwoli nam zrozumieć, dlaczego jest w nim mowa o g r a n i c z n e j informacji. Na ilość informacji o zdarzeniu w komunikacie lub odwrotnie mają wpływ obie strony równości /145/ na zasadzie swoistego prawa "weta". Jeśli zdarzenie odznacza się dużą entropią, to możliwa jest /z jego strony/ duża ilość przekazanej informacji, lecz na przeszkodzie temu może stanąć mała entropia komunikatu /trudno w

trzech literach streścić wyniki konferencji/ lub duża niejednoznaczność odwzorowania zdarzenia w komunikat, powodująca, że pozostaje spora entropia H/B | A/. Entropia H/B | A/ jest miarą niepewności co do tego, jaki komunikat zostanie wysłany, jeśli nawet wiemy, jakie zdarzenie zaszło. Człon ten uwzględnia niedoskonałości systemów kodowania i przekazywania informacji, ale o tym dalej.

W wyniku dysproporcji pomiędzy dużą różnorodnością zdarzeń H/A/, a małą ilością informacji I/A | B/ pozostaje duża entropia H/A | B/, będąca miarą naszej niepewności - nawet po otrzymaniu komunikatu - dotyczącej tego, jakie zdarzenie zaszło.

Możliwa jest również sytuacja odwrotna. Entropia komunikatu jest duża /przeznaczono dużo liter na napisanie wiadomości/, ale albo zdarzenie ma małą różnorodność /małe H/A//, albo niedokładnie znamy jego przebieg w chwili wysyłania komunikatu /duże H/A / /. W wyniku tego ilość informacji będzie znowu mała, a duże będzie H/B | A/, które można w tym wypadku interpretować jako stopień dowolności komunikatu, o ile znamy zdarzenie A. W tym H/B | A/ mieści się zresztą tzw. kwiecistość stylu.

5. Sytuacja opisana jest właściwie regułą we wszystkich komunikatach. Już sam język, lub jego graficzna postać - pismo, odznacza się daleko większą entropią, niż niezbędna do przekazywania tej ilości informacji, jaką pismo to niesie. Jak wspomniano wyżej, jedna litera wykazuje entropię rzędu 5 bitów i taką ilość informacji może przenosić. Tymczasem dokładne badania wykazały, że informacja przenoszona przez pojedynczą literę w tekście nie przewyższa 1 bitu, a często bywa zdecydowanie niższa. Zjawisko to występuje we wszystkich systemach przekazywania i przetwarzania informacji. Zawsze realnie przenoszona informacja stanowi zaledwie kilka procent entropii sygnału, a w przypadkach szczególnych może wynosić ułamki procent. Ten nadmiar informacyjnego potencjału zawartego w sygnale nosi nazwę R E D U N D A N C J I.

6. Redundancja jest zjawiskiem powszechnie spotykanym w procesie przekazywania informacji, co jednakże nie powinno nasuwać podejrzeń, że wszystkie istniejące systemy przekazywania i przetwarzania informacji są wadliwe i wymagają udoskonalenia. Wprawdzie w wielu technicznych wypadkach dąży się do uekonomicznienia sposobu przekazywania informacji i wysiłki te zmierzają właśnie do wyeliminowania redundancji, jednak pewna nadmierność jest znakomitym zabezpieczeniem przed nieuchronnie następującymi w trakcie transmisji zniekształceniami tekstu.

7. Ze względu na tendencję do zwiększania entropii, będącą ogólnym prawem w przypadkach przetwarzania informacji, żaden układ przekazujący lub przetwarzający informację nie może zwiększyć jej ilości, a wprost przeciwnie, każda operacja dokonana nad informacją częściowo ją bezpowrotnie deformuje. Przykładem może być znana dziecinna zabawa w "głuchy telefon". Jeśli cała zawarta w symbolach komunikatu informacja jest istotna, to zamiana /na skutek nieprawidłowego zadziałania układu przekazującego informację/ jednego symbolu na inny jest niewykrywalna, bo powstały wskutek pomyłki nowy zespół symboli jest również sensowny, aczkolwiek najzupełniej błędny.

Przykład 1

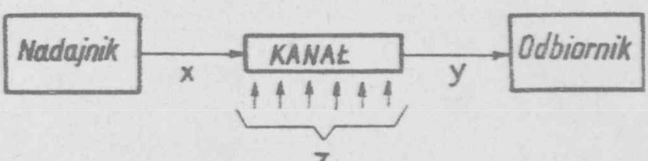
Czytając słowo " B A Ł K A " łatwo zauważymy, że zawiera ono błąd i w potocznym tekście na pewno przeczytamy je jako "B U Ł K A", ponieważ zbiór liter B A Ł K A jest bezsensowny. Wykrycie podobnych błędów zawdzięczamy faktowi, że entropia pojedynczej litery w tekście jest dużo mniejsza, niż entropia tejże litery "solo" lub w bezsensownej grupie. Gdyby jednak ktoś przekazał nam numer rejestracyjny samochodu z takim samym błędem /np. zamiast KU - 1234 podałby KA - 1234/, to faktu tego nie bylibyśmy w stanie rozpoznać, bo oba zestawy symboli są sensowne. Wynika to z faktu, że w zestawach typu numer rejestracyjny fakt kontekstu innych symboli nie zmienia w najmniejszym stopniu entropii pojedynczych symboli i H/B|A/ = 0.

Widzimy więc, że istnienie redundancji pomaga wykryć i skorygować zniekształcenia komunikatu i dlatego jest ona niezbędnym składnikiem każdego naturalnego sposobu przekazywania i przetwarzania informacji. Nie zdziwi nas zatem wiadomość, że we wszystkich językach stosunek entropii przypadającej na jeden znak pisarski, do entropii przypadającej na ten sam znak w kontekście jest orientacyjnie stały i wynosi 0,2. Znaczy to, że każdy język ukształtował się w trakcie kontaktów międzyludzkich tak, aby poprzez 80% liczący nadmiar zmniejszyć wrażliwość na zakłócenia /np. hałasy/ w czasie przekazywania informacji. W niektórych przypadkach, kiedy zależy nam na szczególnej precyzji przekazywania informacji, stosujemy języki o jeszcze większej redundancji, np. języki służące do porozumiewania się z załogami samolotów i statków kosmicznych są złożone ze starannie dobranych słów, szczególnie dobrze rozróżnialnych.

W językach tych redundancja dochodzi do 99%.

4. Przekazywanie informacji przez kanał

1. Ponieważ istota zagadnień związanych z utratą ilości informacji, zachodzą podczas obróbki tej informacji jest podobna w wielu zagadnieniach, rozpatrywać ją będziemy na najprostszym możliwym przykładzie przekazywania informacji. Układ służący do przekazywania informacji nazwiemy K A N A Ł E M, a jego wejście i wyjście połączone będzie z układami, które nazwiemy odpowiednio N A D A J N I K I E M i O D B I O R N I K I E M /rys.89/



Rys. 89. Przekazywanie informacji przez kanał

2. Sygnał na wyjściu kanału Y nie musi być identyczny z sygnałem na wejściu kanału X. Spowodowane jest to szeregiem przyczyn wewnątrz i na zewnątrz kanału, takich jak niedoskonałość kanału /kanał zniekształca/, dostawanie się innych sygnałów do kanału, co powoduje, że sygnał wyjściowy Y jest funkcją nie tylko sygnału wejściowego X, ale i zakłóceć Z itp. W języku przyjętym w rozdziale pierwszym niniejszej książki moglibyśmy powiedzieć, że kanał powinien posiadać tylko jedno wejście i mieć transmitancję równą 1, a tymczasem ma najmniej dwa wejścia i transmitancję różną od 1, zresztą najczęściej jest on w większym lub mniejszym stopniu nieliniowy.

3. Wszystkie te czynniki, które powodują zniekształcenie nadawanej wiadomości nazwiemy w tym rozdziale krótko S Z U-M A M I i nie będziemy dociekali ich źródeł ani podawali środków zaradczych, a jedynie rozpatrzymy ich wpływ na jakość transmisji informacji.

Istnienie szumów jest logiczną i konieczną konsekwencją prawa wzrostu entropii, nie łudźmy się więc, że zdołamy je całkowicie wyeliminować, postaramy się jednak zabezpieczyć przesyłaną informację przed ich wpływami przez odpowiednie K O D O W A N I E.

4. Kodowaniem nazwiemy przedstawienie treści informacji przy pomocy symboli pewnego alfabetu. Naturalnie jedną i tę samą informację można rozmaicie zakodować, używając różnych alfabetów lub tego samego alfabetu z innymi znaczeniami przypisanymi poszczególnym symbolom. Przykładami takiego kodowania są wiadomości podawane alfabetem Morse'go lub zakodowane przy pomocy "klucza" np. w celu utajnienia.
Nie ulega chyba wątpliwości, że sam proces kodowania nie może zmieniać ilości informacji zawartej w komunikacie. Ściślej należałoby powiedzieć, że nie może z w i ę k s z y ć ilości informacji, gdyż nieumiejętne kodowanie może prowadzić do utraty informacji. Może jednak zmieniać wielkość redundancji i chociażby dlatego może zmieniać podatność komunikatu na dzia-

łanie szumów.

5. Rozpatrzymy to dokładniej, najpierw jednak zdefiniujemy jeszcze kilka pojęć oraz przyjmiemy określone założenia. Załóżymy mianowicie, że źródło wiadomości można scharakteryzować pewną entropią, którą nazwiemy entropią nadajnika H/X/. Entropia ta obliczana jest według wzoru /131/ przy założeniu, że możliwy jest pewien zbiór komunikatów $x_1, x_2, \ldots, x_n$ występujących z prawdopodobieństwami $p_1, p_2 \ldots, p_n$.
Możemy podobnie zdefiniować entropię odbiornika /lub sygnału odbieranego H/Y/. Ponieważ sygnał jest zniekształcony w kanale, więc można również określić niezerową entropię H/Y/X/, wyrażającą stopień niepewności nadawcy /znającego nadane X/ co do tego, jaki sygnał zostanie odebrany u odbiorcy Y. Podobnie niezerowa będzie entropia H/X/Y/, wyrażająca stopień niepewności odbiorcy /po odebraniu sygnału Y/ co do tego, jakie X zostało nadane. Ta ostatnia wielkość ma szczególne znaczenie w teorii przekazywania informacji, ponieważ stanowi przeciętną /na symbol/ wieloznaczność symbolu odebranego. Nazwano ją E N T R O P I A E K W I W O K A C Y J N Ą.

6. Zakładając, że źródło nadaje swe symbole jeden po drugim, a kanał przenosi je i zniekształca n i e z a l e ż n i e jeden od drugiego, możemy zdefiniować teraz pojęcie poziomu transmisji. Będziemy więc rozumieli przez poziom transmisji R różnicę pomiędzy entropią źródła, a entropią ekwiwokacyjną:

$$R = H/X/ - H/X \mid Y/ \qquad /146/$$

Łatwo zauważyć, że poziom transmisji jest to po prostu ilość informacji, jaką uzyskamy o źródle wiadomości po otrzymaniu sygnału na wyjściu kanału.

7. Zbliżamy się do sformułowania podstawowego twierdzenia Shannona, dotyczącego kodowania optymalnego. Zanim je jednak podamy, jeszcze jedna definicja.

Dla każdego kanału z szumami możemy przez zastosowanie odpowiedniego kodowania uzyskiwać /przy tym samym źródle/ różne wartości entropii ekwiwokacyjnej. Ten sam kanał zapewnia więc różne wartości poziomu transmisji przy różnych sposobach kodowania. Oznaczmy maksymalny poziom transmisji, dostępny dla danego kanału przez C:

$$C = \max R = \max \quad H/X/ - H/X \mid Y/ \qquad /147/$$

Ten maksymalny poziom transmisji nazwiemy P O J E M N O Ś C I Ą K A N A Ł U. Daje się on teoretycznie obliczyć dla określonego kanału z jego danych technicznych.

8. Otóż twierdzenie Shannona mówi, że dla każdego kanału z szumami istnieje taki sposób kodowania informacji, że możliwe jest przesłanie jej przy poziomie transmisji dowolnie bliskim pojemności kanału. Innymi słowy, po odpowiednim zakodowaniu można przesyłać informacje praktyczne bez strat /przy zerowej ekwiwokacji/, pomimo,że w kanale są szumy i odbierany sygnał może znacznie różnić się od nadawanego. Niemożliwa jest natomiast transmisja przy poziomie wyższym od pojemności kanału. Jeśli zastosujemy nadajnik informacji o większej entropii niż pojemność kanału /o dużej zdolności produkowania wiadomości/, to wzrośnie ekwiwokacja do tego stopnia, że odbierając sygnał będziemy mieli znaczną wieloznaczność jego interpretacji.

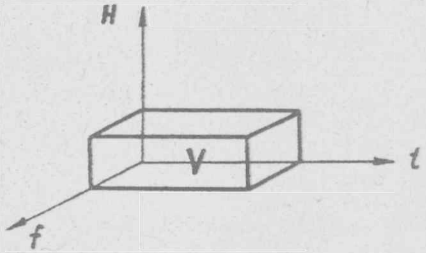
Przykład 1

Jeśli entropia źródła wynosi H/X/ = 1 bit/symbol i mamy przekazać n symboli liczący tekst kanałem o pojemności niższej niż 1 bit, to niezbędne staje się zastosowanie kodowania. Możemy mianowicie wiadomość tę zakodować z użyciem k symboli /k $>$ n/, dzięki czemu entropia tak zakodowanej wiadomości zmaleje poniżej 1 bit/symbol /Więcej symboli a tyle samo bitów!/ i możliwe będzie przekazanie informacji przez kanał bez zniekształceń.

Proszę zwrócić uwagę, że zastosowano tu do zwiększenia odporności sygnału na szumy nie co innego jak właśnie redundancję.

9. Cały komunikat złożony z pewnej ilości symboli posiadających określoną entropię ma również pewną entropię, będącą iloczynem ilości informacji, przypadającej najeden symbol, przez ilość symboli. Wartość tej całościowej informacji nazwiemy O B J Ę T O Ś C I Ą I N F O R M A C J I. Objętość tę można przedstawić jako prostopadłościan w pewnym układzie współrzędnych. Na osiach tego układu będziemy odkładać entropię pojedynczego symbolu /średnią/ H, częstotliwość pojawienia się symboli /ilość symboli przesyłanych w ciągu jednostki czasu/ f i czas przesyłania sygnału t.

Iloczyn tych wielkości jest właśnie objętością informacji



Rys.90. Objętość informacyjna sygnału

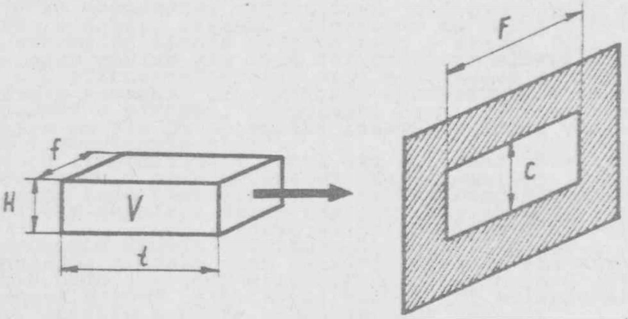
$$V = H f t \qquad /148/$$

10. Objętość informacji jest wielkością stałą, gdyż określoną ilość bitów przesłać musimy, natomiast możemy robić to naróżne sposoby. Możemy na przykład zastosować system kodowania przypisujący każdemu symbolowi dużą entropię, dzięki czemu będziemy mogli wysłać niewiele symboli, czyli pracować z małą ilością symboli przesyłanych w ciągu sekundy f, oraz uporamy się z transmisją w krótkim czasie t. Możemy też przesyłać symbole niosące niewiele informacji, ale wysyłać ich dużo, na przykład nadając długi czas itp.

Sposób przesyłania b y ł b y dowolny, gdyby nie ograniczenia narzucane przez kanał.

Otóż kanał posiada określoną pojemność C, która ogranicza możliwości zwiększenia pionowego wymiaru objętości informacji H, jako że niecelowe jest przesyłanie kanałem informacji o entropii przewyższającej C, bo i tak szumy cały ten nadmiar zniszczą. Szybkość nadawania nie może również być zbyt wysoka, gdyż kanał "nie nadąży" informacji przekazywać i zacznie ją zniekształcać. Ograniczenie to istnieje dla każdego realnego kanału i przedstawiane bywa w postaci określenia największej możliwości szybkości przekazywania wiadomości F.

11. Można więc sobie wyobrazić kanał jako "okienko" o wymiarach C i F /rys. 91/, przez które trzeba "przepchnąć"



Rys. 91. Zdolność przepustowa kanału

określoną objętość sygnału V. Na szczęście sygnał jest, jak wskazaliśmy wyżej, plastyczny. Można mianowicie zmniejszyć jego wymiar pionowy /średnią entropię symboli/ przez zastosowanie odpowiedniego redundantnego kodowania, jak w przykładzie 1, można też zmniejszyć częstotliwość sygnałów przez wydłużenie czasu nadawania. Objętość informacji przekazywanej musi jednak pozostać bez zmian, a za każde ograniczenia, jakie kanał wprowadzi, zapłacimy wydłużonym czasem transmicji.

12. Widzimy więc, że teoria informacji pokazuje, jak można kanałem o dużej zawartości szumów i dopuszczającym jedynie powolne nadawanie przekazać dużą ilość informacji. Rozwiewa jednak równocześnie złudzenia, że uda się to zrobić za darmo - każde ustępstwo od wymagań wobec kanału powoduje spowolnienie transmisji. Wymagajmy więc zawsze dobrych kanałów przekazywania informacji!

ZAKOŃCZENIE

Przedstawione w tej pracy teorie nie wyczerpują bynajmniej wszystkich treści złożonej i bardzo wielokierunkowej nauki, jaką jest cybernetyka. Nauka ta ma w chwili obecnej dopiero 25 lat, znajduje się w stanie ciągłego rozwoju. Pojawiają się wciąż nowe jej dziedziny, a pewne działy, uprzednio rozwijane, kostnieją i tracą na znaczeniu. Dlatego trudno w chwili obecnej definitywnie wypowiedzieć się, czy działy cybernetyki, zaprezentowane w poprzednich rozdziałach, stanowią w pełni reprezentatywny podzbiór treści składających się na całość cybernetyki. Wybór takiej a nie innej reprezentacji podyktowany był obecnym stanem wiedzy cybernetycznej oraz zainteresowaniami potencjalnych odbiorców pracy.

Cybernetyka nie okrzepła jeszcze jako nauka w ostatecznej formie. Poszczególne jej działy takie, jak: Teoria Regulacji Automatycznej, Teoria Wielkich Systemów, Teoria Gier, Teoria Automatów Skończonych, Teoria Algorytmów, Teoria Informacji, nie stanowią dotychczas fragmentów monolitycznej całości, lecz zbiór odrębnych dziedzin, z których każda rozwija się w znacznym stopniu niezależnie od pozostałych, a związek pomiędzy nimi jest raczej formalny. Przypuszczalnie dalsze badania zapełnią istniejące obecnie luki i cybernetyka stanie się nauką równie zwartą jak fizyka. Stan obecny jest jednak od tego ideału dość odległy.

Zasadniczym walorem cybernetyki, który powoduje, że studiowanie jej nawet na obecnym etapie jej rozwoju jest celowe i korzystne - jest uniwersalność i ścisłość cybernetycznego opisu zjawisk. Cybernetyka jest nauką interdyscyplinarną, leżącą na pograniczu matematyki, techniki, fizyki, biologii, ekonomii i socjologii.

Jednak cybernetyka nie zajmuje się wyłącznie kompilowaniem osiągnięć tych dyscyplin szczegółowych i przenoszeniem osiągnięć jednej dziedziny do innych dziedzin.

Główna rola cybernetyki polega na tym, by przez wprowadzenie jednolitego aparatu pojęciowego i jednolitej terminologii ułatwić porozumienie pomiędzy specjalistami różnych dziedzin, oraz na tym, by przez nadanie obiektom nauk szczegółowych nowego, informacyjnego wymiaru przyczyniać się do głębszego rozumienia prawideł ich działania i lepszego czynnego formowania ich właściwości.

W sferze poznawczej cybernetyka jest czynnikiem inspirującym, w sferze działania wytycza cele i środki optymalnych decyzji, a w sferze wytwórczej daje możliwości i metody tworzenia systemów o ogromnej złożoności oraz efektywne metody ich badania i sterowania nimi.

Cybernetyka w przeciwieństwie do matematyki, podaje nie tylko metody ścisłego opisu zjawisk, ale określa również metody uzyskania danych doświadczalnych, niezbędnych w tym opisie, oraz podaje zasady celowego oddziaływania w celu formowania tych zjawisk zgodnie z nadrzędnymi celami.

Zależność cybernetyki od nauk szczegółowych jest dwukierunkowa. Istnieje tu typowe sprzężenie zwrotne, gdyż cybernetyka pobiera informacje od nauk szczegółowych, aby analizować je swoimi metodami, a nauki szczegółowe mogą użytkować wyniki tych analiz w celach praktycznych. W szczególności nauki ekonomiczne mogą wiele zyskać dzięki cybernetycznemu podejściu. Metody rachunku ekonomicznego można by znacznie wzbogacić wprowadzając cybernetyczną teorię dynamiki systemów, teorię struktur niezawodnych, optymalnego sterowania czy optymalnych strategii, które mogą znaleźć natychmiastowe zastosowanie, zaś teoria wielkich systemów umożliwia kompleksową analizę zjawisk ekonomicznych, które wymykają się, ze względu na ogromną złożoność, jakimkolwiek mniej subtelnym metodom analizy. Teoria systemów dostarcza wygodnego aparatu matematycznego, przydatnego przy formalizacji opisu struktur ekonomicznych, a teoria informacji pozwala optymalizować przepływ i gromadzenie informacji.

Wprowadzenie metod cybernetycznych w ekonomii pozwoli na lepszą orientację w całokształcie procesów ekonomicznych, a tym samym na optymalizację działań sterujących i informacyjnych w tych procesach. Z tego względu nie jest wskazane oczekiwanie na ostateczne okrzepnięcie cybernetyki, lecz należy zacząć ją stosować w obecnej postaci, gdyż potrzeby praktyki mogą i powinny wskazać nowe drogi rozwoju teorii cybernetycznych.

Aby stosować cybernetykę, należy ją możliwie dobrze poznać. Praca ta powinna, w zamierzeniach autorów, zapoznanie to ułatwić.

L I T E R A T U R A

1. Arbib M.A.; Mózg, maszyna, matematyka, PWN, Warszawa 1968.
2. Ashby W.R.: Wstęp do cybernetyki, PWN,Warszawa 1963.
3. Beer S.: Cybernetyka a zarządzanie, PWN, Warszawa 1966.
4. Bellman R.: Adaptacyjne procesy sterowania, PWN, Warszawa 1968.
5. Belmann R.; Dreyfus S.: Programowanie dynamiczne./Zastosowania/, PWE, Warszawa 1967.
6. Bereśniewicz-Rajca O., Szczerbińska K., Walichniewicz J.: Elementy logiki matematycznej i algebry wyższej, Gliwice 1970.
7. Borkowski L., Suszko R.: Wstęp do teorii mnogości i logiki matematycznej, Warszawa 1966.
8. Brillouin L.: Nauka a teoria informacji, PWN, Warszawa 1969.
9. Bromirski J.: Teoria automatów, WNT Warszawa 1969.
10. Campbell D.P.: Dynamika procesów, PWN, Warszawa 1962.
11. Charkiewicz A.: Zarys ogólnej teorii łączności, MON, Warszawa 1967.
12. Feldbaum A.: Podstawy teorii optymalnych układów sterowania automatycznego, PWN, Warszawa 1967.
13. Findeisen W.: Technika regulacji automatycznej, PWN, Warszawa 1965.
14. Gale D.: Teoria liniowych modeli ekonomicznych, PWN, Warszawa 1969.
15. Głuszkow W.M.: Synteza automatów cyfrpwych, WNT, Warszawa 1968.
16. Greniewski H.: Cybernetyka niematematyczna, PWN, Warszawa 1969.

17. Hartley R.: Transmission of information, Bell System Techn. J. vol.7, nr 3, 1928.

18. Hellwig Z.: Elementy rachunku prawdopodobieństwa i statystyki matematycznej, PWN, Warszawa 1972.

19. Jagłom A., Jagłom I.: Prawdopodobieństwo i informacja, KiW, Warszawa 1963.

20. Krassowski A., Pospiełow G.: Podstawy automatyki i cybernetyki technicznej, WNT, Warszawa 1965.

21. Kulik Cz.: Metody automatycznego sterowania i kontroli procesów wytwórczych, WSE, Kraków 1972.

22. Kulikowski R.: Procesy optymalne i adaptacyjne w układach regulacji automatycznej, PWN, Warszawa 1965.

23. Kyn O., Pelikan P.: Cybernetyka a ekonomia, PWE, Warszawa 1967.

24. Lange O.: Wstęp do cybernetyki ekonomicznej, PWN, Warszawa 1965.

25. Leja F.: Rachunek różniczkowy i całkowy ze wstępem do równań różniczkowych, PWN, Warszawa 1969.

26. Lerner A.J.: Zarys cybernetyki, WNT, Warszawa 1970.

27. Lyndon R.: O logice matematycznej, PWN, Warszawa 1968.

28. Merriam C.W.: Teoria optymalizacji i projektowania układów sterowania automatycznego, WNT, Warszawa 1967.

29. Mostowski A., Pawlak Z.: Logika dla inżynierów, PWN, Warszawa 1970.

30. Nilsson N.J.: Maszyny uczące się, PWN, Warszawa 1968.

31. Nowacki P., Szklaski L., Górecki H.: Podstawy teorii układów regulacji automatycznej, tom 1 i 2, PWN, Warszawa 1958 i 1962.

32. Osiowski J.: Zarys rachunku operatorowego, WNT, Warszawa 1965.

33. Poletajew I.: Zagadnienia cybernetyki, WNT, Warszawa 1961.

34. Pontriagin L.: Matematyczna teoria procesów optymalnych, MON, Warszawa 1968.

35. Rasiowa H.: Wstęp do logiki matematycznej i teorii mnogości, PWN, Warszawa 1971.

36. Seidler J.: Teoria kodów, PWN, Warszawa 1965.

37. Słupecki J., Borkowski L.: Elementy logiki matematycznej i teorii mnogości, PWN, Warszawa 1963.

38. Smirnow W.: Matematyka wyższa, PWN, Warszawa 1965.

39. Tiepłow L.:O cybernetyce, WNT, Warszawa 1967.

40. Tou T.J.: Nowoczesna teoria sterowania, WNT, Warszawa 1967.

41. Trybuś E.Wstęp do logiki matematycznej, WSE, Wrocław 1971.

42. Wentcel J.: Elementy programowania dynamicznego, PWE, Warszawa 1966.

43. Wentcel J.: Elementy teorii gier, PWN, Warszawa 1961.

44. Węgrzyn S.: Podstawy automatyki, PWN, Warszawa 1963.

45. Wiener N.: Cybernetyka, PWN, Warszawa 1970.

46. Wiener N.: Cybernetyka a społeczeństwo, KiW, Warszawa 1960.

47. Williams J.D.: Strateg doskonały, Wprowadzenie do teorii gier, PWN, Warszawa 1965.