Władysław Findeisen

STRUKTURY STEROWANIA DLA ZŁOŻONYCH SYSTEMÓW

Warszawa 1997
Oficyna Wydawnicza Politechniki Warszawskiej

S

Władysław Findeisen

STRUKTURY STEROWANIA DLA ZŁOŻONYCH SYSTEMÓW

Warszawa 1997
Oficyna Wydawnicza Politechniki Warszawskiej

Opiniodawcy
Ryszard Gessing
Franciszek Milkiewicz
Antoni Niederliński

Redaktor naukowy
Adam Woźniak

Opracowanie redakcyjne
Anna Fijewska-Makowska

Projekt okładki
Danuta Czudek-Puchalska

ISBN 83–87012-84-X

Oficyna Wydawnicza Politechniki Warszawskiej, ul. Polna 50, 00–644 Warszawa, tel. 25–75–18

Wydanie I. Nakład 200 + 30 egz. Ark. wyd. 14,5. Ark. druk. 10,75. Papier offset. kl. III 80 g.
Oddano do składu w maju 1997 r. Druk ukończono w listopadzie 1997 r. Zam. nr 206/97.

Drukarnia Oficyny Wydawniczej PW, ul. Kopińska 12/16, 02–321 Warszawa, tel. 660–40–26

Spis treści

Przedmowa

W ostatnich paru dziesięcioleciach zakres zainteresowań specjalistów zajmujących się zagadnieniami sterowania zmieniał się i poszerzał. Coraz więcej miejsca w pracach badawczych i w zastosowaniach znajduje to, co można nazwać sterowaniem systemami złożonymi.

Systemy takie wymagają specjalnych metod i sposobów sterowania, polegających głównie na użyciu wielu jednostek decyzyjnych, w stosowny sposób powiązanych ze sobą. Prowadzi to do tworzenia odpowiednich struktur sterowania.

Potrzeba rozpatrywania systemów złożonych jest – być może – jednym ze znaków współczesnego czasu. Występuje ona w bardzo różnych dziedzinach, że wymienimy tu tylko procesy technologiczne i produkcyjne, robotykę, systemy transportowe i energetyczne, sieci telekomunikacyjne, zarządzanie w przedsiębiorstwach, organizacjach i systemach gospodarczych, ochronę środowiska i gospodarkę zasobami naturalnymi. Wyliczenie to nie oznacza, że książka ta pretenduje do bezpośredniej przydatności w dziedzinach tak rozmaitych, różniących się przede wszystkim znacznie stopniem kwantyfikacji występujących w nich zjawisk, zadań i problemów. Być może jednak niektóre koncepcje, zasady i prawidłowości sformułowane na podstawie procesów fizycznych i modeli matematycznych mogą okazać się pomocne także tam, gdzie modeli takich brakuje lub są bardzo niedokładne. Język, sposób opisu i rozumowania oraz większość przykładów niniejszej książki powodują, że adresowana jest ona głównie do osób o wykształceniu technicznym. Powinna ona jednak okazać się czytelna także dla ekonomistów zajmujących się zagadnieniami sterowania lub dla specjalistów zarządzania.

Rozdział 1 tej książki stanowi wprowadzenie w obszar pojęć i problemów sterowania systemami, a rozdział 5 – poszerzenie i podsumowanie, które powinno być zrozumiałe także bez wniknięcia w kwestie szczegółowe, omawiane w rozdziałach 2, 3 i 4.

W rozpoznaniu zakresu tematycznego całego wykładu może być pomocny zamieszczony na końcu skorowidz, który ujmuje ważniejsze hasła i terminy w ich kontekście merytorycznym.

Autor pracował nad tą książką od dawna, z długimi jednak przerwami. Tekst ostateczny zawdzięcza bardzo wiele staranności, wnikliwości i krytycyzmowi redaktora, dr. Adama Woźniaka – w imieniu własnym i przyszłych Czytelników składam Mu za to bardzo serdeczne podziękowanie.

Wyrazy wdzięczności kieruję także pod adresem recenzentów maszynopisu, Profesorów Ryszarda Gessinga, Franciszka Milkiewicza i Antoniego Niederlińskiego, których życzliwe i wnikliwe uwagi były dużą pomocą.

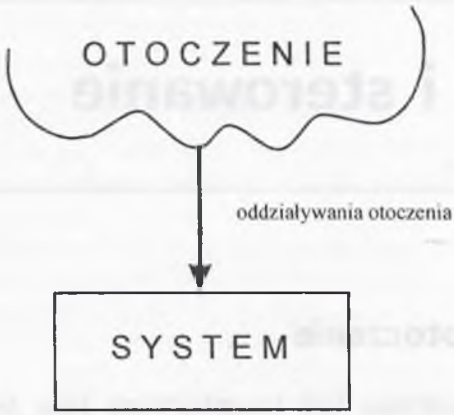
Rozdział 1

Systemy i sterowanie

1.1. System i otoczenie

Granice tego, co powinno być rozpatrywane jako jeden system zależą w pierwszym rzędzie od tego, jaki jest *zasięg kompetencji* podmiotu decyzyjnego, to jest osoby, lub grupy osób, której ten system służy lub która jest za jego funkcjonowanie odpowiedzialna. Jest rzeczą w tym momencie wtórną, czy ta znajdująca się w domyśle osoba będzie podejmować decyzje odnoszące się do działania systemu sama, czy też stanie się autorem reguł postępowania personelu bądź projektantem urządzeń automatycznych. W każdym przypadku możemy mówić o *punkcie widzenia* decydenta. Zależnie od tego spojrzenia systemem będzie ciąg technologiczny w fabryce lub cała fabryka, blok energetyczny albo system elektroenergetyczny kraju, przedsiębiorstwo lub zrzeszenie przedsiębiorstw i tak dalej.

Jest jednak i drugi czynnik, który wpływa na granice tego, co musi być rozpatrywane jako jeden system. Czynnikiem tym jest *możliwość*, a ściślej mówiąc *dopuszczalność wyodrębnienia systemu* spośród reszty świata: wyodrębnienia jako obiektu analizy i jako przedmiotu sterowania czy zarządzania. W większości sytuacji rzeczywistych nie jest to sprawą łatwą i oczywistą, bowiem powiązania fizyczne bądź ekonomiczne procesów zachodzących w jakimkolwiek rozpatrywanym obiekcie sięgają zazwyczaj bardzo daleko. Zarówno analiza, jak i sterowanie systemem wymagają jednak by dokonać pewnego rozgraniczenia, by odróżnić *system* i *otoczenie*. Kwestię tę – na użytek niniejszej książki – rozstrzygniemy w sposób pragmatyczny: otoczenie określimy jako „te aspekty świata zewnętrznego, które mają wpływ na procesy w systemie, ale same nie są pod ich wpływem". Odpowiada temu schemat podany na rysunku 1.1. Zauważmy, że w naszym określeniu „otoczenia" dokonaliśmy rozróżnienia między tym, co ma wpływ na procesy w systemie i tym, co takiego wpływu nie ma. Nie uważamy zatem, że cały zewnętrzny świat jest otoczeniem rozpatrywanego systemu, lecz tylko ta jego część, która ma wpływ na zachodzące w systemie zjawiska i procesy. Najważniejsze jest tu stwierdzenie, że to otoczenie nie znajduje się jednocześnie pod wpływem rozpatrywanego systemu, a mówiąc ściślej – pod wpływem zachodzących w nim procesów i podejmowanych decyzji. Mówiąc dokładniej,



Rys. 1.1. System i otoczenie

uważamy, że interesujące nas procesy przebiegające w systemie nie wpływają na otoczenie w taki sposób, iż wpływ ten powraca poprzez otoczenie z powrotem do systemu. Jeśli bowiem taki powrotny wpływ istnieje, to analizując działanie systemu musielibyśmy to sprzężenie zewnętrzne uwzględnić, tym samym rozpatrywalibyśmy w istocie większy system, a nie „system" i „otoczenie". W większości przypadków rozgraniczenie między systemem a otoczeniem jest umowne, tzn. mówimy o braku sprzężenia zewnętrznego tylko z pewną dokładnością. *Decyzja o rozgraniczeniu systemu i otoczenia jest jednak konieczna*; przy jej podjęciu potrzebna jest duża doza rozsądku, aby właściwie ocenić interakcje mające znaczenie dla danego problemu, czyli zdecydować, które z powiązań ze światem zewnętrznym będą uwzględnione, a które zostaną pominięte. Błędem metodologicznym może być zarówno rozpatrywanie systemu zbyt szerokiego (ze względu na niepotrzebną komplikację analizy), jak też zbyt wąskiego (gdyż pominięte interakcje mogą spowodować, iż wynik analizy będzie niewłaściwy).

Powróćmy na chwilę do wspomnianego na początku „punktu widzenia". Stwierdziliśmy już, że jest to podstawowy czynnik określający granice systemu: dla dyrektora naczelnego linii lotniczych systemem jest całe przedsiębiorstwo, dla kapitana – jego samolot wraz z załogą. Podobnie będzie z dyrektorem fabryki oraz mistrzem zmianowym określonego ciągu technologicznego. To samo występuje przy projektowaniu układu sterowania dla procesu technologicznego: dla projektanta układów regulacji temperatury czy ciśnienia „systemem" będzie pojedynczy reaktor, dla projektanta całości – pełny proces technologiczny, od wejścia surowców do produktów.

Powstaje tu nowa kwestia metodologiczna: czy to, czym steruje (zarządza) jednostka wysoko umieszczona w hierarchii, na przykład dyrektor przedsiębiorstwa, który przecież „nie wchodzi w szczegóły", jest „dużym systemem", to znaczy musi być traktowane jako system opisywany przez znaczną liczbę zmiennych i związków (zależności) między nimi? Dyrektor linii lotniczych operuje przecież stosunkowo niewielką liczbą zmiennych, rozważa na przykład tygodniowy przewóz pasażerów w danej relacji, czy więc nie ma on do czynienia z systemem stosunkowo prostym? Istota odpowiedzi na to pytanie zawiera się w spostrzeżeniu, że związki między zmiennymi rozpatrywanymi przez dyrektora zawierają niejako w swoim wnętrzu wszystko to, co dzieje się czy też działo w obiektach fizycznych systemu, włączając w to ich załogi oraz decydentów niższych szczebli. Nie oznacza to bynajmniej, że trzeba na szczeblu centralnym rozpatrywać wszystkie zmienne szczegółowe, na przykład zachowanie się każdego samolotu i każdego członka załogi; można i trzeba operować wielkościami zbiorczymi, czyli agregowanymi, ale trzeba też brać pod uwagę to, że modele związków między wielkościami agregowanymi są przybliżeniami, które będąc np. całkowicie poprawne w sensie statystycznym mogą okazać się mało miarodajne w odniesieniu do sytuacji jednorazowych. Z tego więc względu przedsiębiorstwo, nawet widziane ze szczebla dyrektora, pozostaje dużym systemem; rozstrzyga o tym, w ostatecznej instancji, wielość zmiennych i zależności opisujących fizyczne czy wykonawcze procesy we wszystkich elementach systemu.

Budowa *agregowanych modeli* zachowania się systemu sterowanego jest sztuką podobną do tej, którą jest rozróżnienie systemu i otoczenia; zdecydować musi rozsądek z uwzględnieniem przeznaczenia modelu. W systemie zasobów wodnych można na przykład pominąć bardzo małe zbiorniki retencyjne jeśli rozpatruje się jeden czy kilka zbiorników dużych, ale co zrobić jeśli zbiorników małych jest dużo oraz suma ich pojemności jest współmierna z pojemnością dużych? Czy można te małe zastąpić jednym zbiornikiem sumarycznym i jak to zależy od rozmieszczenia w terenie? Słuszność modelu przybliżonego zależna jest w dużym stopniu od celu, któremu ma on służyć: jeśli chodzi o bilans zasobów wodnych, można być może zbiorniki połączyć w jeden, jeśli zaś o analizę przepływu fali powodziowej, zrobić tego nie wolno.

1.2. Zadanie sterowania

Rozważania tej książki odnosić się będą najczęściej do sterowania *systemami złożonymi*, to znaczy takimi, które są zbudowane z dających się sensownie wyodrębnić części, czyli *obiektów* lub *podsystemów*.

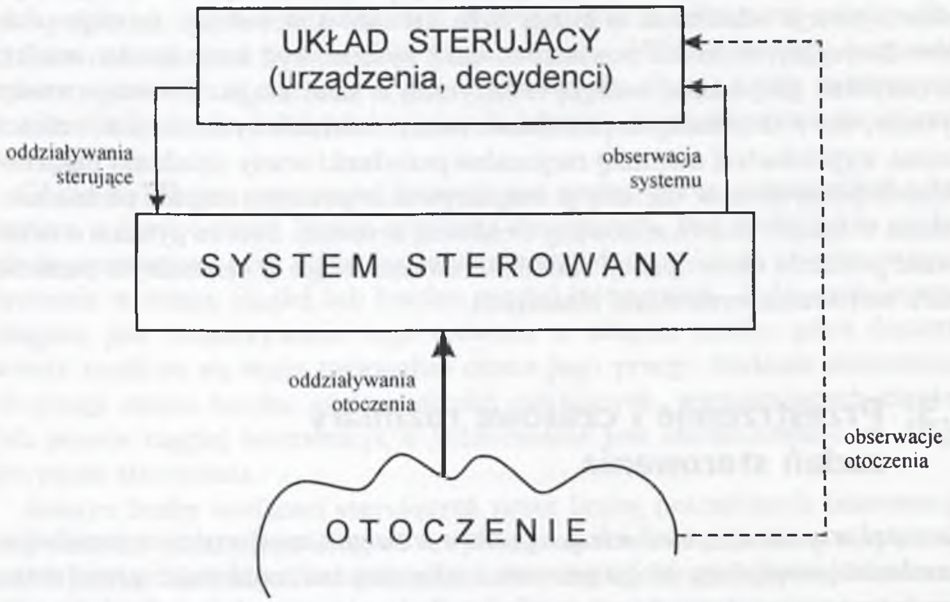
Wyróżnić można trzy czynniki, które *łączą obiekty w system*, to jest skłaniają do rozpatrywania zestawu obiektów jako całości (w odróżnieniu od odrębnego rozpatrywania każdego z obiektów). Pierwszym z nich są *oddzia-*

ływania wzajemne między obiektami, na skutek czego interwencja sterująca wywarta na dany obiekt wywołuje zmiany biegu procesu nie tylko w tym obiekcie, ale także w innych obiektach systemu. Ma to w szczególności miejsce w liniach technologicznych, systemach produkcyjnych i systemach ekonomicznych. Prostym przykładem jest sytuacja, gdy temperatura mieszaniny wypływającej z reaktora w linii technologicznej wpływa na bilans cieplny reaktora następnego. Drugim czynnikiem łączącym grupę obiektów w system może być *wspólność zasobów*. Nie ma wówczas znaczenia czy obiekty oddziałują na siebie wzajemnie, czy też nie; ograniczoność zasobów, z których wszystkie one muszą czerpać powoduje, że trzeba te obiekty rozpatrywać w sposób łączny, a więc jako system. Przykładem takich sytuacji są np. użytkownicy wody, czerpiący z jednego źródła lub zbiornika. Wreszcie trzecim czynnikiem, powodującym potrzebę rozpatrywania systemu jako całości, jest istnienie *wspólnego celu działania* jego obiektów (niezależnie od ewentualnych celów lokalnych), a zatem także istnienie wspólnej miary (ocena skalarna) lub miar (ocena wektorowa) wartościujących bieg procesów w danym zestawie obiektów. Nie ma tu lub może nie być oddziaływań „fizycznych" między obiektami, łączy je jednak w system istnienie wspólnego celu i całościowych ocen. Przykładem takich sytuacji są np. systemy transportu lotniczego czy samochodowego, gdzie każdy pojazd stanowi obiekt fizycznie całkowicie odrębny, ale wszystkie razem służą określonemu zadaniu.

Po określeniu, jaką całość powinniśmy rozpatrywać jako system podlegający sterowaniu, zdefiniować wypada, co rozumieć będziemy przez sterowanie. Przyjmiemy, jako określenie dogodne dla dalszych rozważań, że *sterowanie jest to interwencja* (ze strony układu sterującego) dokonywana *w trakcie trwania procesu sterowanego*, czyli z możliwością wykorzystania bieżących obserwacji. Oznacza to, że skutki decyzji można nie tylko przewidywać (używając sformalizowanego bądź myślowego modelu systemu), lecz także na bieżąco sprawdzać.

W rozważaniach dotyczących zachowania się obiektów bądź systemów pod wpływem oddziaływań sterujących nie można pominąć tego, że każdy obiekt lub system znajduje się równocześnie pod wpływem otoczenia. Otoczenie wpływające na system sterowany może być obserwowane (przez urządzenia sterujące lub przez decydentów), a obserwacje te wykorzystane przy realizacji sterowania (rys. 1.2).

Zadaniem sterowania, mówiąc ogólnie, jest zapewnienie *pożądanego przebiegu procesu sterowanego* w całym systemie, czyli procesów sterowanych we wszystkich jego obiektach. Projektant układu sterowania, tj. struktury i reguł *przekształcania obserwacji zachowania systemu i otoczenia w oddziaływania (interwencje) sterujące*, mieć musi w istocie na widoku ukształtowanie właściwego *reagowania na zmiany w otoczeniu*; mówiąc niezbyt precyzyjnie, układ sterowania ma zapewnić *osłabienie reakcji* procesu sterowanego na jedne oddziaływania bądź cechy otoczenia oraz *pozytywne odpowiadanie* na inne spośród nich.



Rys. 1.2. System sterowany, otoczenie i układ sterujący

Z chwilą, gdy zaczęliśmy mówić o pożądanym przebiegu procesu sterowanego musimy poszerzyć nieco pojęcie otoczenia; muszą to być odtąd nie tylko te aspekty świata zewnętrznego, które mają wpływ na bieg procesu, ale także i te, które mają wpływ na ocenę tego biegu, a więc na cele sterowania i podporządkowane im oddziaływania sterujące. Na przykład decyzje odnoszące się do pracy systemu energetycznego mogą zależeć od czynników takich jak ceny paliw – rynek paliw stanowi zatem otoczenie systemu, chociaż nie ma on wpływu fizycznego na biegnące procesy.

Dla każdego zadania sterowania sprawą podstawową staje się określenie jaki przebieg procesu jest *przebiegiem pożądanym*, tj. tym, który ma być *wywołany* czy *utrzymany przez oddziaływanie sterujące*. Przesłanek do tego określenia można by szukać na wysokim szczeblu, na przykład w wieloletniej strategii rozwoju danego przedsiębiorstwa czy organizacji, ale powiązania takie rzadko są znaczące w odniesieniu do procesów przebiegających w istniejących urządzeniach i instalacjach. Przesłanką do określenia pożądanego biegu procesu będą wówczas prostsze mierniki techniczne i ekonomiczne. Pozostaje jednak nieomal regułą, że przesłanki typu ekonomicznego można w racjonalny i uzasadniony sposób formułować dla systemów stosunkowo dużych i złożonych – zwykle nie da się tego zrobić dla pojedynczego, prostego obiektu.

Mamy więc taką sytuację, że idąc niejako od dołu, to jest starając się rozpatrywać jak najmniejszy system, zatrzymujemy się w punkcie gdzie można już sformułować racjonalne przesłanki oceny procesów w systemie. Może być także sytuacja odmienna, o której była już mowa wcześniej: istnieje podmiot decyzyjny, któremu powierzono dany system, być może bardzo wielki, na przykład gospodarkę energią elektryczną w skali kraju. Powstaje wtedy pytanie, czy z racjonalnych przesłanek oceny działania systemu jako całości można wyprowadzić niemniej racjonalne przesłanki oceny działania poszczególnych podsystemów tak, aby je rozpatrywać w pewnym stopniu oddzielnie, jednak w sposób zharmonizowany z całością systemu. Jest to pytanie o możliwość podziału i koordynacji zadań sterowania leżące w centralnym punkcie teorii sterowania systemami złożonymi.

1.3. Przestrzenne i czasowe rozmiary zadań sterowania

Niewątpliwie mamy przed sobą, z punktu widzenia możliwości racjonalnego określenia pożądanego biegu procesu, zwłaszcza zaś możliwości przetłumaczenia tego żądania na ekonomicznie i technicznie uzasadnioną funkcję celu i inne wymagania, systemy obejmujące wiele, często różnorodnych obiektów, wiele wielkości sterujących (zmiennych decyzyjnych), wiele wielkości wewnętrznych, określających aktualny stan systemu (wiele współrzędnych stanu) oraz wiele wielkości wyjściowych (zmiennych zależnych od sterowań i stanu systemu).

Podobnie, jeśli zastanawiamy się nad skalą czasową, w której rozpatrywać powinniśmy sterowany proces, czyli jeśli myślimy o tzw. *horyzoncie sterowania*, to w wielu wypadkach będzie to czas długi: proces technologiczny, system produkcyjny, system energetyczny lub wodno-gospodarczy, przedsiębiorstwo lotnicze itd. są przykładami systemów, których działanie musi być rozpatrywane w skali wieloletniej, zwłaszcza jeśli uwzględnić wykonywanie i rozłożenie w czasie np. remontów sprzętu i instalacji, ich odnowę, rozbudowę itd. Jednocześnie, w tym samym systemie mamy do czynienia z procesami o szybkiej zmienności, wymagającymi interwencji ciągłej lub prawie ciągłej – start lub lądowanie samolotu, reakcje przebiegające w reaktorach systemu technologicznego, ruchy narzędzi w obrabiarkach w systemie produkcyjnym.

Mówiąc o *rozmiarach przestrzennych* zadania sterowania nie musimy koniecznie mieć na myśli rozmiarów przestrzennych w sensie dosłownym i fizycznym. Istotą rzeczy jest to, że ze względu na racjonalne sformułowanie celu sterowania wypada nam rozpatrywać systemy złożone z wielu obiektów, zawierające bardzo wiele wielkości sterujących.

Podobnie, mówiąc o *długim horyzoncie* sterowania nie musimy mieć na myśli miesięcy czy lat. Horyzont czasowy sterowania całym systemem trzeba porównywać ze zmiennością najszybszych spośród zjawisk w obiektach systemu, czyli z największym dopuszczalnym odstępem między decyzjami sterującymi poszczególnymi procesami. Mówiąc inaczej, horyzont czasowy sterowania całością systemu jest wtedy długi, kiedy zawiera się w nim wielka liczba odstępów między kolejnymi, szczegółowymi decyzjami sterującymi.

Obraz wynikowy opisywanej sytuacji jest następujący: mamy przed sobą system o bardzo dużej liczbie wielkości sterujących, być może także rozległy w przestrzeni, przy czym zapewnienie pożądanego biegu procesu w tym systemie wymaga ciągłej lub bardzo częstej interwencji. Jednocześnie wymagane jest rozpatrywanie tego systemu w długim czasie, gdyż dopiero wtedy możliwa się staje racjonalna ocena jego pracy. Zadanie sterowania obejmuje zatem bardzo wiele wielkości sterujących, wymagających ciągłej lub prawie ciągłej interwencji, a jednocześnie jest sformułowane na długi horyzont sterowania.

Iloczyn liczby wielkości sterujących przez liczbę potrzebnych interwencji w jednostce czasu przez długość horyzontu sterowania mierzoną w tych samych jednostkach daje liczbę decyzji, którą musiałoby podejmować centrum sterujące, gdybyśmy nie sięgnęli do struktur decyzyjnych opierających się na *podziale zadań*, a mianowicie do *zdecentralizowanych* oraz *hierarchicznych* struktur sterowania. Istota rzeczy polega tu na wprowadzeniu wielu jednostek decyzyjnych, zamiast jednej – centralnej, w taki jednak sposób, by *nie stracić całościowego spojrzenia na sterowany system*. Jednostka centralna, ze względu na bardzo wielką liczbę decyzji, byłaby możliwa tylko w wersji całkowicie zmechanizowanej – nawet nadzór przez człowieka czy przegląd decyzji wychodzących z komputera nie byłby możliwy. Regułą większości sytuacji rzeczywistych jest jednak to, że wiele spośród rozstrzygnięć dotyczących dużego systemu i długiego horyzontu czasowego potrzebuje decydenta zdolnego do oceny wartości nie dających się kwantyfikować oraz do oceny i podejmowania ryzyka; decydentem takim musi pozostać człowiek. Zdolności decyzyjne człowieka są ograniczone ilościowo, nie pozostaje zatem nic innego, jak z tego także względu myśleć o strukturach sterowania zawierających więcej niż jedną jednostkę decyzyjną.

1.4. Struktury zdecentralizowane i hierarchiczne

Spróbujmy zdać sobie sprawę z głównych możliwości, a raczej schematów rozmieszczenia *jednostek decyzyjnych* (to znaczy urządzeń sterujących lub decydentów) względem systemu sterowanego.

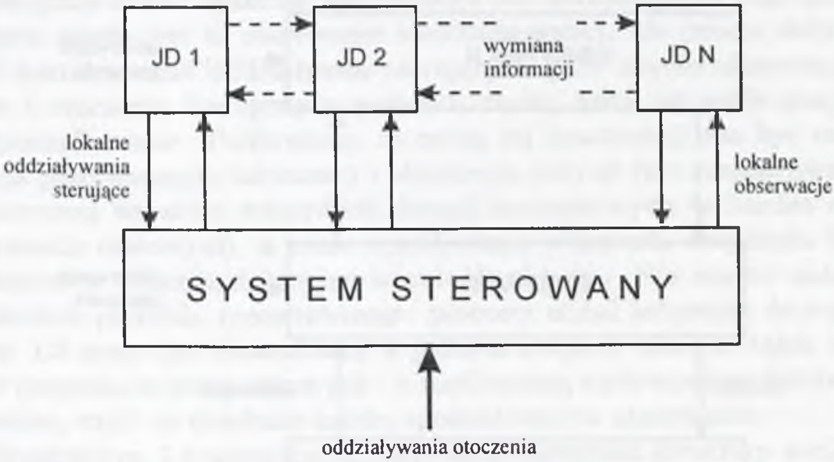
Rysunek 1.3 przedstawia sterowanie za pomocą pojedynczej, centralnej jednostki decyzyjnej, która rozporządza wszystkimi obserwacjami procesu w systemie sterowanym i obserwacjami otoczenia oraz ma wszystkie kompetencje. Mówiąc o pojedynczej jednostce decyzyjnej mamy tu na myśli, że określa ona wszystkie oddziaływania sterujące (czyniąc to poprzez od-

Rys. 1.3. Sterowanie systemem przy użyciu pojedynczej jednostki decyzyjnej

powiednie przekształcanie obserwacji). Wspomnieliśmy na końcu poprzedniego podrozdziału, że jest to struktura mało racjonalna w stosunku do dużych systemów, ze względu na ogromną objętość przetwarzanej informacji, a także na trudność włączenia ocen i intuicji człowieka w tak pomyślany proces decyzyjny.

Na rysunku 1.4 przedstawiona jest druga skrajna możliwość – sterowanie *zdecentralizowane*, polegające na istnieniu wielu, równoległych i równorzędnych, jednostek decyzyjnych. Każdej z nich przydzielono określone kompetencje (wielkości sterujące określonego podsystemu), przy czym nie zachodzą one na siebie; każda z jednostek dysponuje także obserwacjami systemu sterowanego. Będą to zazwyczaj obserwacje odnoszące się do podległego danej jednostce podsystemu, ewentualnie również jakaś część obserwacji pozostałych. Decyzje jednostek zdecentralizowanych działających zgodnie z rys. 1.4 są od siebie niezależne, to znaczy wzajemnie sobie nie podlegają, następować może tylko wymiana informacji – zaznaczono to liniami przerywanymi.

Czego można się spodziewać po sterowaniu systemem przy użyciu struktury z rys. 1.4? Trzeba tu rozróżnić dwa przypadki. Pierwszy z nich to ten, gdy jednostki decyzyjne postrzegają swoje indywidualne interesy, w szczególności *indywidualne cele*. Zauważmy, że ze względu na powiązania w systemie decyzje podejmowane przez każdą z jednostek mogą mieć wpływ na rezultaty osiągane przez inne jednostki, czyli na wyniki ich decyzji; przy istnieniu odrębnych celów jednostki decyzyjne z rys. 1.4 znajdują się w położeniu *wzajemnej konkurencji*.



Rys. 1.4. Sterowanie zdecentralizowane

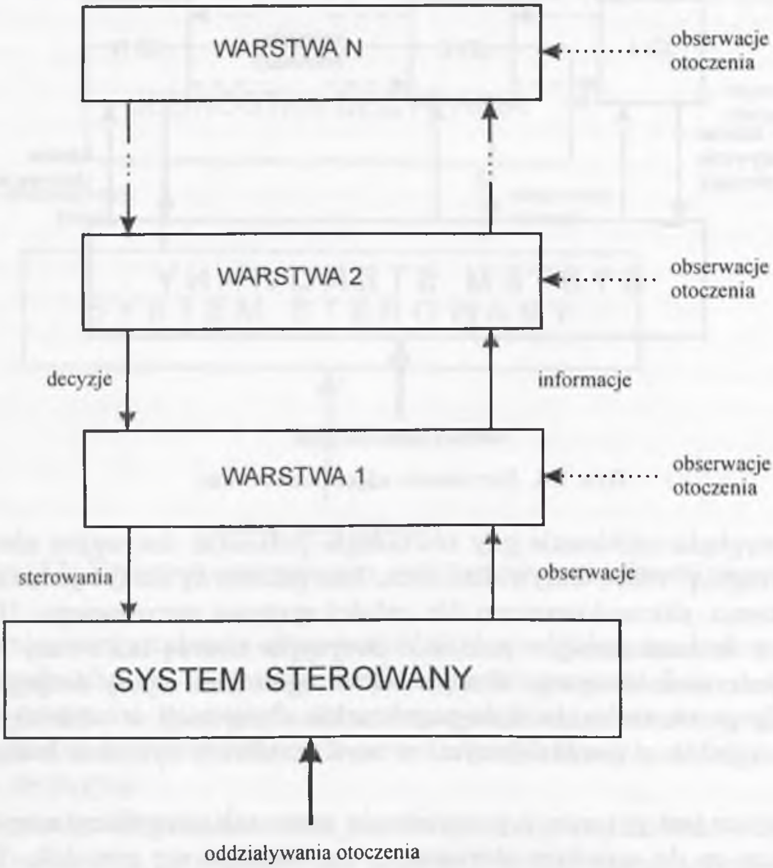
Rzecz wygląda odmiennie gdy równoległe jednostki decyzyjne nie mają (nie postrzegają) celów indywidualnych, lecz gotowe są służyć jednemu celowi ogólnemu, sformułowanemu dla całości systemu sterowanego. Nie ma tu podstaw do konkurencji – jednostki decyzyjne tworzą tak zwany *zespół*. Każda z jednostek akceptuje służące celowi ogólnemu *reguły decyzyjne*, to jest zasady przekształcania dostępnych sobie obserwacji w oddziaływania sterujące, zgodnie z przydzielonymi w myśl struktury systemu kompetencjami.

Interesujące jest pytanie, czy sterowanie przez tak pomyślany zespół jest równoważne co do wyników sterowaniu centralnemu wg rys. 1.3. Odpowiedź teoretyczna jest negatywna, gdyż oddziaływania sterujące poszczególnych jednostek są kształtowane przy użyciu tylko części obserwacji systemu i otoczenia. Wykorzystanie całości informacji do określenia każdego z oddziaływań sterujących, charakterystyczne dla rys. 1.3, może w zasadzie dać wynik lepszy. Na przeszkodzie staje jednak zazwyczaj objętość informacji, czas i niezawodność jej przekazywania oraz inne podobne czynniki, co powoduje, że sterowanie złożonym systemem za pomocą centralnej jednostki decyzyjnej nie może być skuteczne.

Następną ważną możliwością strukturalną jest pionowy układ jednostek decyzyjnych pokazany na rys. 1.5, noszący nazwę *struktury wielowarstwowej*[*]. Istotą rzeczy jest tutaj to, że każda z jednostek decyzyjnych (każda z warstw) podejmuje decyzje odnoszące się do tego samego systemu, ale

[*] Idea sterowania w podziale na warstwy należy do Lefkowitza, patrz I. Lefkowitz (1966). Multilevel approach applied to control system design. Trans. ASME, Series B, J. Bas. Eng. 88 (2) 392–398.

są to decyzje różnego rodzaju. Jedną z możliwości jest rozpatrywanie w kolejnych warstwach coraz to *dłuższego horyzontu czasowego*, w związku



Rys. 1.5. Układ warstwowy

z tym odmiennych aspektów przebiegu procesu oraz prognoz otoczenia systemu*). Dodatkową możliwością jest operowanie w kolejnych warstwach coraz bardziej *agregowanymi spojrzeniami na system sterowany*, na przykład operowanie tygodniowymi czy miesięcznymi wartościami produkcji tak, aby w najwyższej warstwie rozpatrywać proces w długim horyzoncie czasowym, lecz przy użyciu niewielkiej liczby zmiennych. Umożliwia to skorzystanie z intuicji i osądu decydenta, mimo iż sam proces sterowany jest skomplikowany i pełen szczegółów. Jeszcze inną możliwością jest taki rozdział funkcji między jednostki decyzyjne, że każda następna warstwa

*) Tego rodzaju strukturę, zwaną wielohoryzontową, wprowadził i analizował Milkiewicz, patrz: F. Milkiewicz (1968). Problemy optymalizacji systemów produkcyjnych na przykładzie kombinatu rafineryjnego. Zeszyty Naukowe Politechniki Gdańskiej, Elektryka XXII, str. 1–158.

ustala reguły postępowania (*algorytmy działania*) warstwy niższej. Struktura z rys. 1.5 jest w pewnym sensie rozwinięciem struktury z rys. 1.3: obejmujemy całość systemu sterowanego nie korzystając z jego podziału (w tym sensie jest to sterowanie scentralizowane), ale proces decyzyjny, czyli kształtowanie oddziaływań sterujących przy użyciu obserwacji systemu i otoczenia budujemy z pewnych części, które są sobie wzajemnie podporządkowane. Podkreślmy, że cechą tej konstrukcji ma być racjonalizacja przetwarzania informacji i określania decyzji (nie rozpatrujemy np. w pierwszej warstwie wszystkich decyzji szczegółowych w bardzo długim horyzoncie czasowym), a także umożliwienie włączenia decydenta lub decydentów w procesy decyzyjne w całości układu. Nie ma tu natomiast w zasadzie podziału przestrzennego: pionowy układ jednostek decyzyjnych z rys. 1.5 może być umieszczony w jednym miejscu, może to także być zestaw programów komputerowych – z możliwością wpływania na każdy z nich z osobna, czyli na działanie każdej spośród warstw sterowania.

Wreszcie rys. 1.6 przedstawia najprostszy przykład *struktury wielopoziomowej**); jest to sterowanie zdecentralizowane z dodaniem jednostki nadrzędnej, koordynującej działanie jednostek niższego poziomu (zwanych tutaj *lokalnymi jednostkami decyzyjnymi*). Struktura ta ma szczególne znaczenie, gdy jednostki lokalne mają (postrzegają) swoje własne cele, w której to sytuacji – jak już mówiliśmy – mogą konkurować ze sobą. Pierwszym zadaniem jednostki nadrzędnej może tu być zażegnanie konfliktu między jednostkami niższego poziomu, zneutralizowanie go tak dalece, by konflikt interesów lokalnych nie prowadził do wadliwego działania całego systemu. Jednostka nadrzędna posługiwać się będzie *instrumentami koordynacji* (na przykład będzie narzucać ceny na wielkości czy produkty, stanowiące o powiązaniach między podsystemami), myśląc przy tym także o swych własnych celach, to jest starając się optymalizować swoją własną ocenę przebiegu procesu w systemie sterowanym.

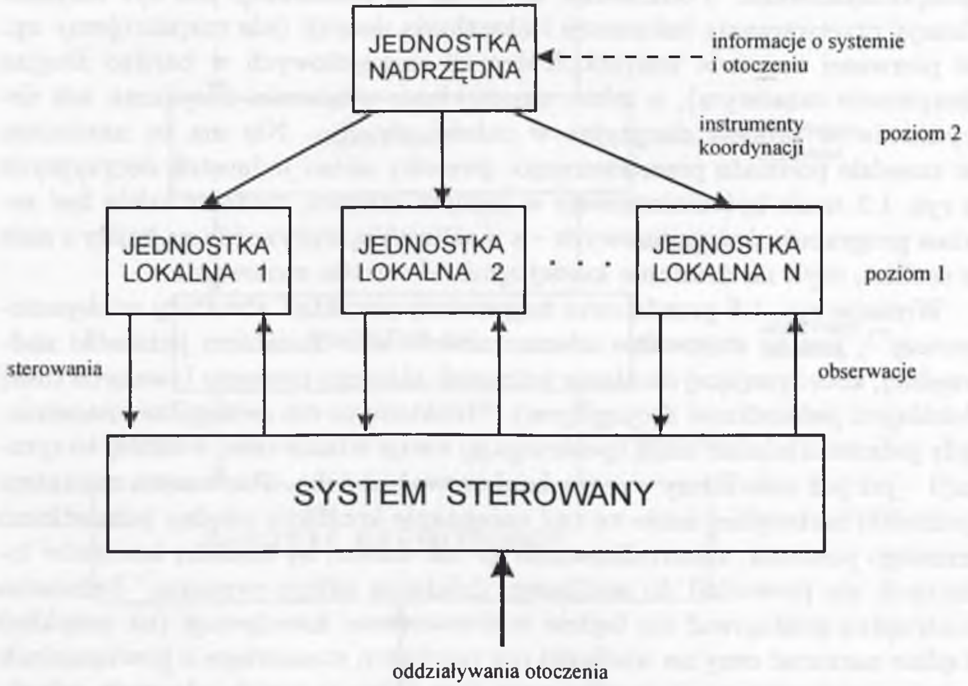
Struktury z rys. 1.5 oraz z rys. 1.6 są strukturami *hierarchicznymi*. Ich cechą wspólną jest istnienie wielu (więcej niż jednej) jednostek decyzyjnych, przy czym nie wszystkie spośród nich mają *dostęp bezpośredni* do systemu sterowanego. Dostęp ten mają tylko jednostki pierwszej warstwy albo pierwszego poziomu.

Powiedzmy jeszcze, że struktura wielowarstwowa (rys. 1.5) i struktura wielopoziomowa (rys. 1.6) mogą być ze sobą połączone, na przykład przez podzielenie zadań jednostki decyzyjnej dolnej warstwy sterowania między kilka jednostek równoległych, co wywoła potrzebę koordynacji. Podobnie, jednostka nadrzędna na rys. 1.6 nie musi operować tym samym horyzontem

*) Idea struktury wielopoziomowej pochodzi z metod dekompozycji i koordynacji w programowaniu matematycznym, patrz np. L.S. Lasdon (1968). Duality and decomposition in mathematical programming. IEEE Trans. Syst. Sci. Cybern. SSC-4 (2): pp. 86–111.

czasowym co jednostki dolnego poziomu, przez co układ wielopoziomowy nabiera zarazem cech układu wielowarstwowego.

W dalszych częściach tej książki termin „sterowanie hierarchiczne" odnosić będziemy głównie do przypadków takich, w których istnieją bądź racjonalne jest utworzenie lokalnych jednostek decyzyjnych, czyli gdzie zachodzi także potrzeba koordynacji (por. rys. 1.6).



Rys. 1.6. Układ wielopoziomowy

W rozdziale 2 omawiamy najważniejsze zjawiska i koncepcje sterowania w tak pomyślanych układach hierarchicznych. Następnie rozwiniemy zagadnienia sterowania wielowarstwowego, wychodząc od pożądanego przebiegu procesów w obiektach systemu (rozdział 3) i omawiając różne aspekty wielowarstwowych struktur sterowania w rozdziale 4.

W sytuacjach rzeczywistych występuje zazwyczaj połączenie koncepcji wielopoziomowej i wielowarstwowej, wraz z wielkim zróżnicowaniem zadań i mechanizmów sterowania; będzie to opisane w rozdziale 5.

1.5. Przykłady złożonych systemów

Hierarchiczne struktury zależności między jednostkami decyzyjnymi, w szczególności gdy są to po prostu podejmujące decyzje osoby, są oczywiście znane od dawna. Występują one we wszelkiego typu organizacjach

– państwowych, militarnych, gospodarczych i tak dalej. Nawet jeśli ograniczyć pole widzenia do sterowania automatycznego, które wymaga pełnego sformalizowania procesu podjęcia decyzj oraz do podejmowania decyzji ze wspomaganiem komputerowym, gdzie co najmniej część procesu decyzyjnego musi być sformalizowana (patrz rozdz. 5.6), dziedziny możliwych zastosowań hierarchicznych struktur struktura sterowania są także bardzo liczne i różnorodne. Rozciągają się one od sterowania procesami technologicznymi oraz systemami produkcyjnymi, poprzez sterowanie systemami użytkowania zasobów, do zarządzania w przedsiębiorstwach. Nie byłaby celowa jakakolwiek próba inwentaryzacji czy klasyfikacji wszelkich istniejących, a zwłaszcza przyszłych – to jest potencjalnych – zastosowań. Leżą one, ogólnie rzecz biorąc, w całej sferze sterowania i zarządzania, a stawać się będą możliwe, realne i racjonalne w miarę tego, jak czynione będą postępy na polu pełnej lub częściowej formalizacji podejmowania decyzji realizujących sterowanie czy zarządzanie. Wchodzi w to m.in. potrzeba budowy dostatecznie wiarygodnych modeli matematycznych.

Na dalszym dopiero planie stawać będą ograniczenia natury technicznej. Możliwości komputerów są już dziś bardzo duże i będą nadal rosnąć; bariery dla realizacji układów sterowania złożonymi systemami można się raczej spodziewać gdzie indziej, a mianowicie w dużym nakładzie wiedzy i wysiłku intelektualnego jaki jest potrzebny dla zaprojektowania i pełnego oprogramowania każdego nowego zastosowania.

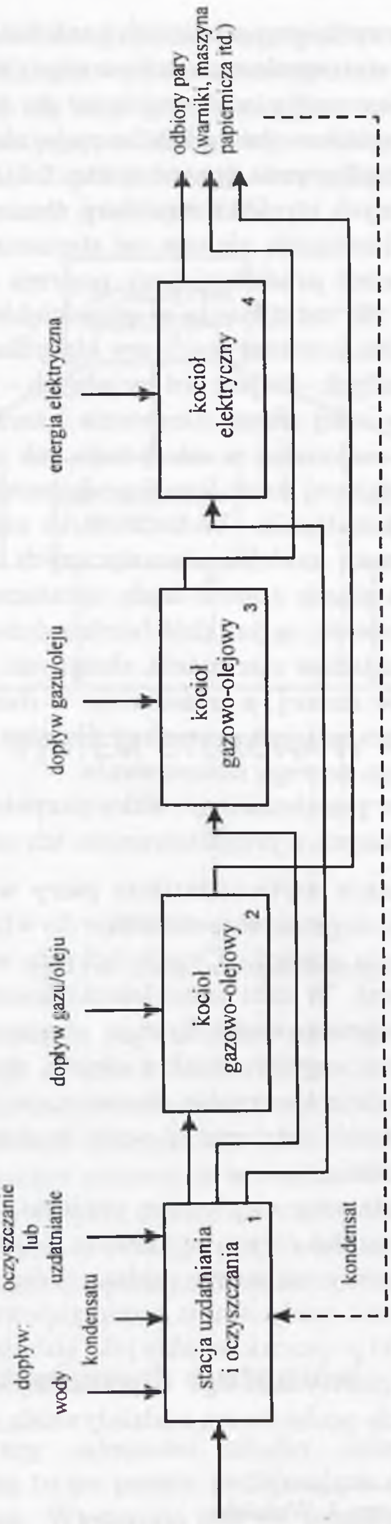
Przejdziemy teraz do przedstawienia kilku przykładów złożonych systemów i zagadnień związanych z projektowaniem ich układów sterowania.

Przykład 1. Sterowanie wytwarzaniem pary w ciepłowni[*)]

Wiele procesów technologicznych potrzebuje do właściwego funkcjonowania ciągłego dostarczania energii. Często odbywa się to za pomocą pary wytwarzanej w ciepłowni. W celu zmniejszenia kosztów i zwiększenia niezawodności ciepłownie przemysłowe bywają obecnie wyposażone, oprócz kotłów opalanych pyłem węglowym albo olejem opałowym, także w kotły elektryczne. Umożliwia to szybkie dostosowywanie ilości wytwarzanej pary do zmiennych potrzeb oraz zmniejszenie kosztów dzięki przełączaniu na różne sposoby zasilania.

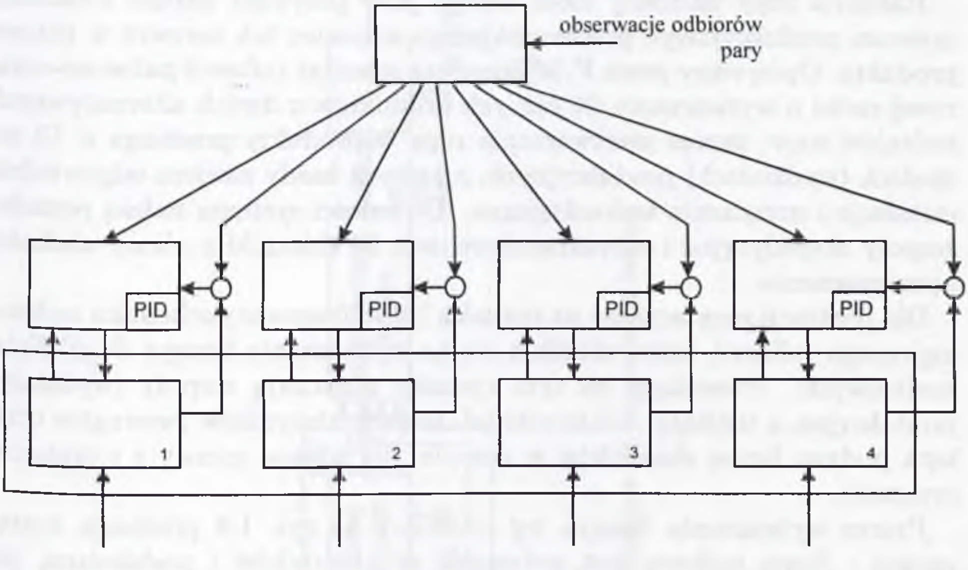
Jako przykład rozpatrzymy ciepłownię papierni, na którą składają się dwa kotły opalane gazem lub olejem opałowym z optymalizatorem procesu spalania, kocioł elektryczny zasilany z podstacji średniego napięcia, stacja uzdatniania i oczyszczania wody, stacja pomp zapewniających jej właściwą cyrkulację oraz jednostki pomocnicze takie jak: zbiorniki oleju, stacja redukcyjna gazu, podstacja elektryczna itp. Uproszczony schemat ciepłowni, na którym zaznaczono także podstawowe oddziaływania sterujące, przedstawia rysunek 1.7.

*) Przykład opracowany przez A. Woźniaka.



Rys. 1.7. Uproszczony schemat ciepłowni

Układ bieżącego sterowania ciepłowni składa się z regulatorów wielofunkcyjnych połączonych z komputerem nadrzędnym. Jego struktura jest więc dwupoziomowa (rys. 1.8). Regulatory funkcjonujące na pierwszym poziomie optymalizują proces spalania, utrzymują zadane: natężenie dopływu gazu



Rys. 1.8. Ciepłownia jako system podlegający sterowaniu

lub oleju opałowego, ciśnienie pary, przepływy wody, stopień oczyszczenia ścieków, zapełnienia różnych zbiorników itd. Ich regułami decyzyjnymi są w większości algorytmy PID. Oprócz tego, regulatory pierwszego poziomu dokonują na polecenie komputera nadrzędnego przełączenia: rodzajów zasilania (z gazu na olej lub odwrotnie), rodzajów kotłów, sposobów pracy stacji uzdatniania wody itd. Komputer nadrzędny (koordynator) dobiera wielkości zadane oraz momenty przełączeń (zmienne koordynacyjne) tak, aby proces wytwarzania pary był prowadzony minimalnym kosztem oraz aby osłabione były reakcje systemu na oddziaływania zewnętrzne (głównie zmiany poboru pary). Możliwe są także oddziaływania bardziej złożone, np. zmiana nastaw w algorytmach PID czy ustalenie nowych parametrów w algorytmach przełączania. Aby koordynator mógł osłabić niekorzystne oddziaływania otoczenia którym dla rozpatrywanego systemu są pozostałe oddziały papierni (warniki, maszyna papiernicza itd.) charakteryzujące się zmiennym poborem pary, dostarczana jest do niego bieżąca informacja o ich stanie, która stanowi podstawę do obliczania nowych wartości zmiennych koordynacyjnych.

Zauważmy też, że wprowadzenie jednostki nadrzędnej może być wykorzystane do podwyższenia niezawodności działania systemu, ponieważ można ją wyposażyć w oprogramowanie umożliwiające bieżącą diagnostykę systemu, a w sytuacji awarii którejkolwiek lokalnej jednostki decyzyjnej – pozwalające przejąć jej funkcje.

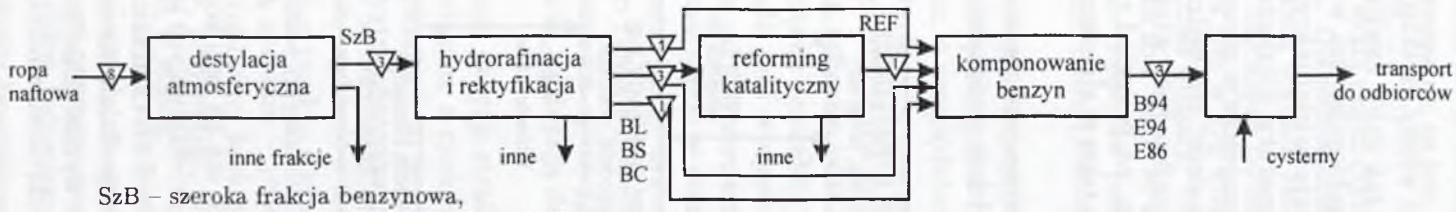
Przykład 2. Rafineria ropy naftowej*)

Rafineria ropy naftowej może służyć jako przykład bardzo złożonego systemu produkcyjnego, przetwarzającego surowiec lub surowce w gotowe produkty. Opisywany przez F. Milkiewicza schemat rafinerii paliwowo-smarowej mówi o wytwarzaniu 24 różnych produktów z dwóch alternatywnych rodzajów ropy; proces przetwarzania ropy w produkty przebiega w 19 zespołach (wydziałach) produkcyjnych, z których każdy zawiera odpowiednie instalacje i urządzenia technologiczne. Do całości systemu należą ponadto zespoły ekspedycyjne i zaopatrzeniowe oraz 94 zbiorniki o różnej wielkości i przeznaczeniu.

Dla ilustracji przytaczamy na rysunku 1.9 te fragmenty schematu technologicznego rafinerii, które składają się na wytwarzanie benzyn do silników spalinowych. Prostokąty na tym rysunku oznaczają zespoły (wydziały) produkcyjne, a trójkąty – zbiorniki lub zespoły zbiorników (wewnątrz trójkąta podano liczbę zbiorników w zespole, dla zdania sprawy z rozmiarów systemu).

Proces wytwarzania benzyn wg schematu na rys. 1.9 przebiega następująco. Ropa naftowa jest pobierana ze zbiorników i poddawana, po uprzednim podgrzaniu, destylacji. Jedną z 5 oddzielanych od siebie frakcji jest tzw. szeroka frakcja benzynowa (SzB), która w następnym zespole produkcyjnym podlega hydrorafinacji (chodzi m.in. o zmniejszenie zawartości związków siarki). W tym samym zespole poddaje się frakcję SzB procesowi rektyfikacji, otrzymując w rezultacie tego procesu benzynę lekką, średnią i ciężką (BL, BS, BC) oraz dwa dalsze produkty – tzw. gaz suchy (głównie metan i etan) oraz gaz płynny (głównie propan, butan). Benzyna średnia trafia do zespołu reformingu katalitycznego, którego zadaniem jest podnoszenie liczby oktanowej doprowadzanego półproduktu. Polega to na zmianie struktury węglowodorów tworzących frakcję BS, a dokonuje się w reaktorach chemicznych z katalizatorem platynowym. Powstaje tzw. reformat o znacznie wyższej liczbie oktanowej oraz pewne produkty uboczne. Bardzo istotne znaczenie ma następny, końcowy w tej sekwencji zespół produkcyjny: komponowanie benzyn. Proces komponowania polega na wymieszaniu ze sobą w odpowiednich proporcjach otrzymanych poprzednio półproduktów (reformat oraz 3 rodzaje benzyny z wyjścia zespołu hydrorafinacji i rektyfikacji) wraz z dodatkami ulepszającymi w taki sposób, by powstał produkt o dokładnie określonych cechach. W zespole, o którym mowa, każdy z rodzajów benzyny (bezołowiowa B94, etylizowana E94 i E86) wytwarzany jest w odrębnym reżimie pracy, czyli nie jednocześnie, a produkt zbiera się

*) Przykład oparty na pracach F. Milkiewicza.

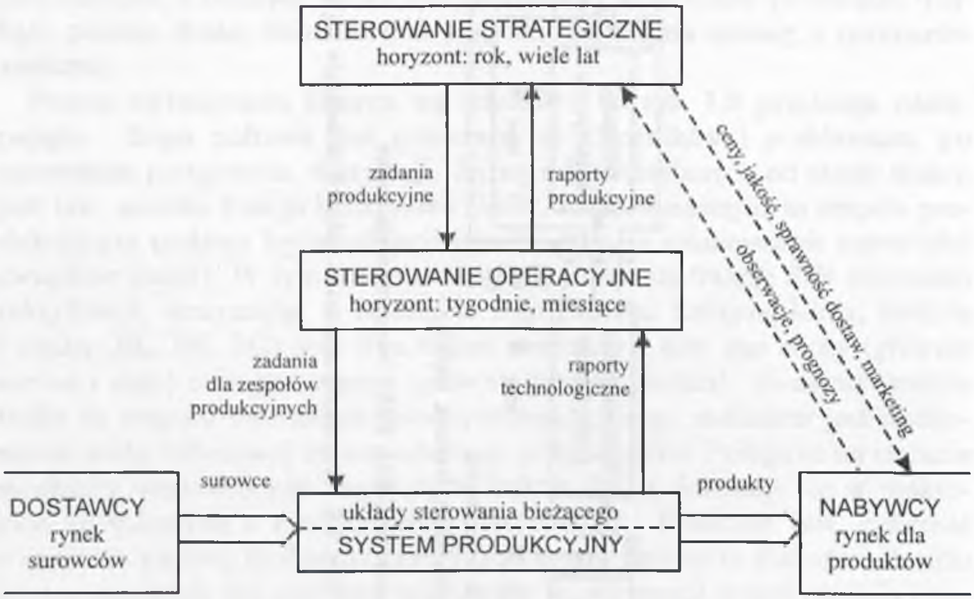


SzB – szeroka frakcja benzynowa,
BL, BS, BC – benzyna lekka, średnia, ciężka,
REF – reformat,
B94 – benzyna bezołowiowa, E94, E86 – benzyny etylizowane.

Rys. 1.9. Fragmenty schematu rafinerii składające się na produkcję benzyn w zbiornikach (mają one – w omawianym przykładzie – pojemność 10000 m^3 każdy).

System produkcyjny, o którym już wiemy, że składa się z instalacji i urządzeń technologicznych, zbiorników i tak dalej, zawiera w sobie także układy bieżącego sterowania procesami przebiegającymi w tych obiektach. Będą to przede wszystkim różnego rodzaju wielofunkcyjne regulatory, służące utrzymaniu zadanej wartości temperatury, ciśnienia, przepływów i innych parametrów, decydujących o biegu procesów w poszczególnych instalacjach. Do układów sterowania bieżącego należeć też będzie określanie jakie wartości parametrów procesowych są najbardziej właściwe dla wykonywania aktualnego polecenia produkcyjnego, na przykład rozdestylowywania ropy na określone frakcje; dla zadań tego typu w sterowaniu bieżącym mogą być użyte stosowne mikrokomputery.

Na rysunku 1.10 sterowanie bieżące jest uważane za integralną część systemu produkcyjnego. Zaznaczyliśmy natomiast tak zwane sterowanie



Rys. 1.10. Rafineria jako system produkcyjny podlegający sterowaniu

operacyjne (bądź operatywne), które ma na celu koordynację działania poszczególnych zespołów (wydziałów) produkcyjnych w taki sposób, by system jako całość realizował postawione przed nim zadanie lub zadania odniesione do produktów finalnych. Koordynacja działania zespołów produkcyjnych (czyli podsystemów) może oznaczać wyznaczanie zadań ilościowych i parametrów jakościowych, w tym także okresowe zmienianie zadań, kiedy jeden i ten sam zespół produkcyjny może realizować różne procesy technologiczne (wytwarzać różne produkty, jak widzieliśmy na przykładzie zespołu komponowania benzyn). Oznacza to, że do sterowania operacyjnego może należeć tak zwane harmonogramowanie, to jest ustalanie kolejności i czasu działania zespołów produkcyjnych.

Zastanówmy się nad skalą czasu czyli nad horyzontem czasowym działań czy decyzji sterujących. Dla sterowania bieżącego czas mierzy się w sekundach, jeśli nie w ułamkach sekund, zaś co najwyżej w minutach – zależnie od dynamiki procesów fizycznych podlegających sterowaniu. W sterowaniu operacyjnym, zwłaszcza gdy w grę wchodzi harmonogramowanie, horyzont czasowy musi być znacznie dłuższy. Milkiewicz*) mówi o kilku tygodniach lub miesiącach – wynika to z tego, że wielkość zbiorników magazynujących pozwala na przełączanie zespołów produkcyjnych z jednego produktu na drugi w dużych odstępach czasu (jest to korzystne jeśli przełączanie pociąga za sobą koszty).

Jeszcze dłuższy horyzont czasowy potrzebny jest dla sterowania strategicznego, którego celem będzie (patrz rys. 1.10) ustalanie zadań produkcyjnych dla całego systemu, z uwzględnieniem co najmniej zmian sezonowych. Oznacza to minimum 1 rok, albo znacznie więcej jeśli uwzględniać będziemy modernizacje i nowe inwestycje dla systemu produkcyjnego.

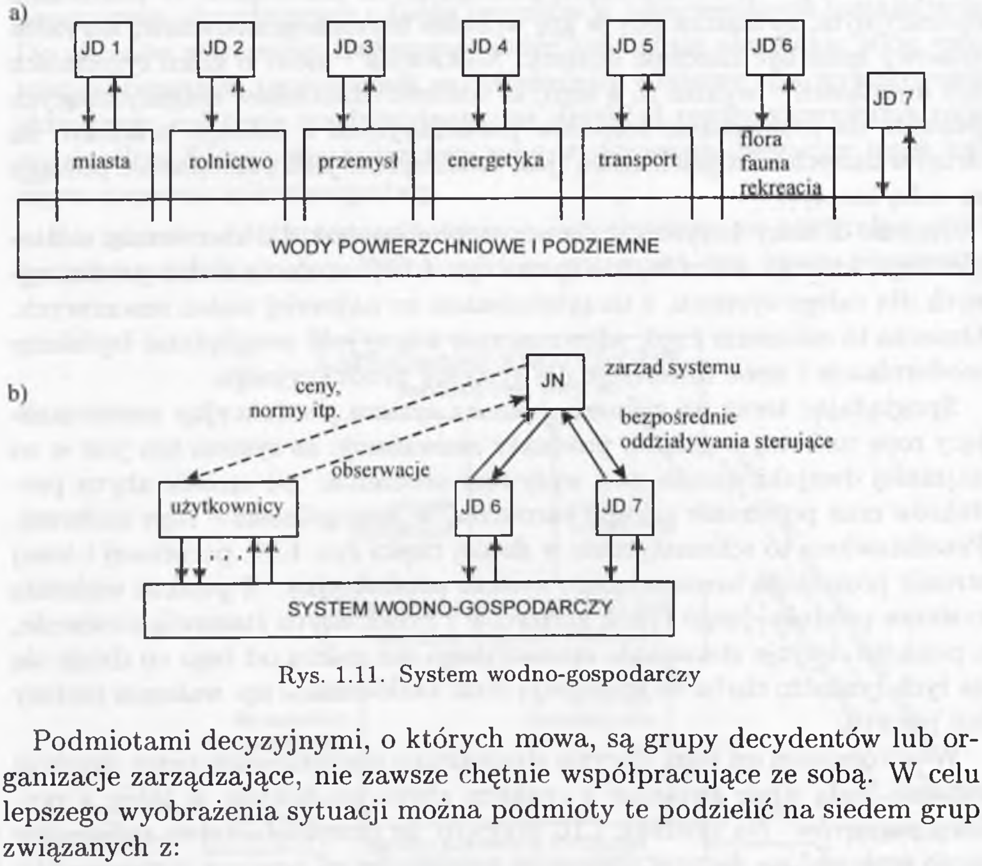
Spoglądając teraz na rafinerię jako na system produkcyjny przetwarzający ropę naftową w gotowe produkty zauważamy, że system ten jest w co najmniej dwojaki sposób pod wpływem otoczenia: po stronie zbytu produktów oraz po stronie zakupu surowców, w szczególności – ropy naftowej. Przedstawiono to schematycznie w dolnej części rys. 1.10, po prawej i lewej stronie prostokąta oznaczającego system produkcyjny. Z punktu widzenia systemu produkcyjnego rynek surowców i rynek zbytu stanowią otoczenie, a ponadto decyzje sterowania operacyjnego nie zależą od tego co dzieje się na tych rynkach, chyba że występują duże zakłócenia – np. wahania podaży lub popytu.

W odróżnieniu od tego, decyzje strategiczne podejmowane przez dyrekcję rafinerii będą silnie związane z rynkiem zbytu produktów, a także z rynkiem surowców. Na rysunku 1.10 przyjęto, że przedsiębiorstwo rafineryjne może wpływać na decyzje nabywców produktów za pomocą instrumentów takich jak: ceny, jakość, szybkość i terminowość dostaw, działania marketingowe. W ten sposób, z punktu widzenia dyrekcji przedsiębiorstwa, rynek zbytu staje się częścią systemu – przestaje być otoczeniem, na które nie mamy wpływu. Nie zrobiliśmy takiego założenia na rys. 1.10 w odniesieniu do rynku surowców, przyjmując, że pojedyncza rafineria wpływu na ten rynek wywierać nie jest w stanie.

---

*) Algorytmy wspomagające w procesie operatywnego sterowania produkcją zakładów z ciągłymi przełączalnymi procesami produkcyjnymi na przykładzie rafinerii nafty. Raport z realizacji projektu badawczego KBN, Katedra Automatyki Politechniki Gdańskiej, 1995.

Przykład 3. System wodno-gospodarczy*)
Kolejnym przykładem systemu złożonego jest system wodno-gospodarczy przedstawiony schematycznie na rysunku 1.11. Granice geograficzne (przestrzenne) takiego systemu najczęściej pokrywają się z granicami zlewni dużej rzeki, a sam system znajduje się w gestii licznych podmiotów decyzyjnych, które wpływają na poziom zasobów wody oraz na stan środowiska naturalnego.



Rys. 1.11. System wodno-gospodarczy

Podmiotami decyzyjnymi, o których mowa, są grupy decydentów lub organizacje zarządzające, nie zawsze chętnie współpracujące ze sobą. W celu lepszego wyobrażenia sytuacji można podmioty te podzielić na siedem grup związanych z:

1) odbiorcami komunalnymi (miastami),
2) rolnictwem,
3) przemysłem bez energetyki,
4) energetyką,
5) transportem,
6) odpowiedzialnością za zachowanie zasobów przyrodniczych i utrzymanie równowagi ekologicznej,
7) odpowiedzialnością za gospodarkę zasobami wody.

*) Przykład opracowany przez A. Woźniaka.

Odpowiada to wydzieleniu w systemie wodno-gospodarczym siedmiu podsystemów powiązanych ze sobą za pośrednictwem pobieranej i zwracanej wody, a także oddziałujących bezpośrednio na siebie (wpływów tych dla większej przejrzystości na rys. 1.11a nie zaznaczono).

Zauważmy, że pięć pierwszych podsystemów to użytkownicy wody, którzy ją zużywają, a ponadto zanieczyszczają. Interesy użytkowników są określone przez naturalne dążenie do zapewnienia pełnego pokrycia swoich potrzeb, przy możliwie minimalnych nakładach na retencję i oczyszczanie wody. Przy ograniczonych zasobach prowadzi to do silnej konkurencji pomiędzy poszczególnymi rodzajami użytkowników, czyli – na rys. 1.11b – jednostkami decyzyjnymi JD 1 do JD 5. Interesy podmiotów decyzyjnych związanych z gospodarką zasobami wodnymi i odpowiedzialnych za zachowanie zasobów przyrodniczych JD 6 i JD 7 są inne – starają się tak gospodarować zasobami, aby pokryć uzasadnione zapotrzebowania wszystkich odbiorców, spełnić wymagania ochrony przeciwpowodziowej oraz maksymalnie ograniczać degradację środowiska naturalnego.

Przedstawione siedem grup podmiotów podejmujących decyzje to decydenci działający stale, o prawnie ustalonym zakresie działania. Obok nich, szczególnie w momencie dyskusji nad projektami rozwoju lub zmian istotnych składników systemu wodno-gospodarczego powstają, często w ramach ciał przedstawicielskich (samorządowych), grupy nacisku reprezentujące interesy zainteresowanych grup ludności albo podkreślające wagę ochrony różnych elementów środowiska naturalnego. W takiej sytuacji konieczność racjonalnego pogodzenia różnych, często sprzecznych interesów staje się jeszcze silniejsza.

Poszczególne podmioty decyzyjne mogą działać niezależnie, czyli w sposób całkowicie zdecentralizowany, wtedy odpowiedni system prawny powinien zmusić je do wzajemnego uzgadniania działań; uzgadnianie to miałoby formę negocjacji. Negocjacje zabierają jednak prawie zawsze dużo czasu, zatem w celu podniesienia efektywności działania systemu sterowania nadaje mu się często strukturę hierarchiczną z wydzieloną jednostką koordynującą.

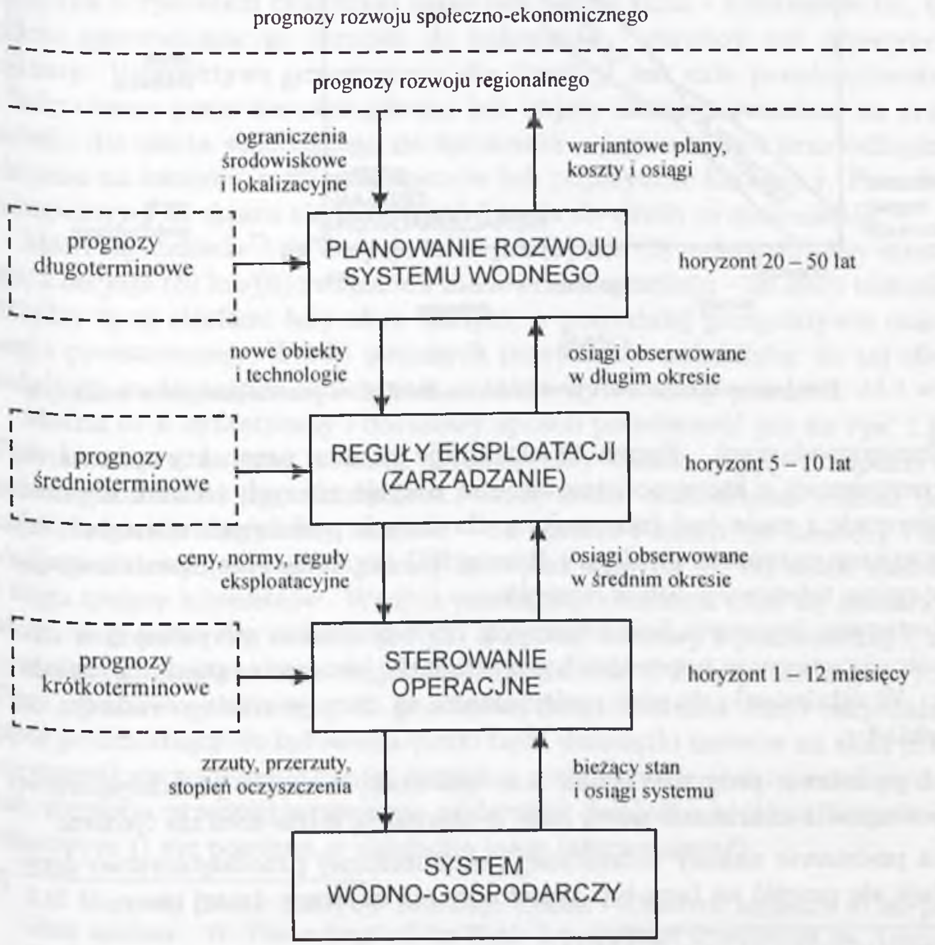
W przypadku systemu wodno-gospodarczego jednostką taką może być tzw. zarząd systemu, któremu bezpośrednio są podporządkowane: bieżące sterowanie gospodarką zasobami wody oraz działania dotyczące zachowania zasobów przyrodniczych i utrzymania równowagi ekologicznej (wszystkie aspekty hydrologiczne i hydrodynamiczne oraz najistotniejsze spośród aspektów biologiczno-ekologicznych). Oznacza to, że dla zarządu system składa się w pierwszym rzędzie z elementów związanych z gospodarowaniem zasobami wód powierzchniowych i podziemnych (rzeki, zbiorniki naturalne i sztuczne, rurociągi przerzutowe, ujęcia wód głębinowych itd.) oraz z wszelkiego rodzaju urządzeń niezbędnych dla utrzymania równowagi ekologicznej (oczyszczalnie, napowietrzacze, osadniki itd.). Na przebieg procesów w tych elementach zarząd może oddziaływać bezpośrednio; na rys. 1.11b przyjęliśmy jednak, że ze względu na pewną odmienność sterowanie gospodarką zasobami i ochroną środowiska powierzono odrębnym jednostkom decyzyjnym, koordynowanym przez jednostkę nadrzędną – zarząd systemu.

Zarząd systemu jest też najczęściej odpowiedzialny za opracowywanie wieloletnich programów rozwoju infrastruktury technicznej, umożliwiającej racjonalną gospodarkę zasobami oraz sprzyjającej ochronie środowiska (budowa zbiorników retencyjnych, oczyszczalni itp.). W ramach swoich uprawnień zarząd systemu może także wpływać pośrednio na decyzje użytkowników, ustanawiając np. ceny na pobieraną wodę oraz normy zanieczyszczeń, czy też wpływając na parametry systemu podatkowego tak, aby skłaniały do działań sprzyjających ochronie środowiska.

Analizując funkcjonowanie układu sterującego dla systemu wodno-gospodarczego wygodnie jest nadać podejmowanym decyzjom strukturę warstwową, przedstawioną na rysunku 1.12. Są tu decyzje dotyczące rozwoju systemu: gdzie, kiedy i jakie mają być zbudowane nowe zbiorniki, rurociągi, oczyszczalnie i inne obiekty. Są decyzje ustalające reguły i przepisy eksploatacyjne, a także określające parametry dla odbiorców, takie jak ceny na wodę i normy zanieczyszczeń. Jednostką odpowiedzialną za wypracowanie obu wymienionych grup decyzji jest zarząd systemu. Są wreszcie decyzje dotyczące bieżącego działania systemu w warunkach normalnych, powodzi lub suszy, określające pobory i zrzuty wody ze zbiorników, przepływy w rurociągach, ograniczenia poboru wody, stopień oczyszczenia itd. Nie należy w tym miejscu zapomnieć o tym, że tylko część decyzji dotyczących bieżącego działania systemu znajduje się w rękach jego zarządu. Pozostałe są podejmowane przez niezależnych użytkowników.

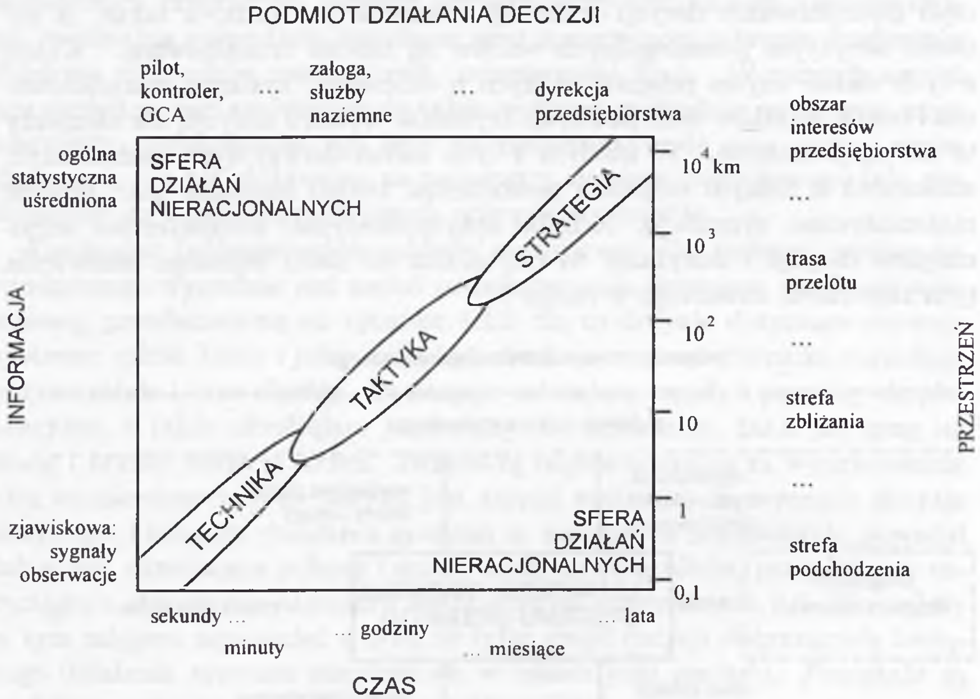
U wierzchołka hierarchii znajduje się programowanie, czyli planowanie rozwoju całego systemu; w działaniu tym rozpatruje się horyzont 20 – 50 lat, dla którego trzeba określić m.in. prognozy zmian zapotrzebowania na wodę, związane np. ze zmianami struktury sieci osadniczej czy przemysłu, a także zmiany jej podaży wynikające np. ze zmian w środowisku naturalnym. Planowanie rozwoju systemu wodno-gospodarczego jest powiązane, w sposób wzajemnie na siebie wpływający, z programami rozwoju danego regionu oraz ogólniejszymi prognozami rozwoju społeczno-gospodarczego, co także zaznaczono na rysunku. W trakcie bieżącego użytkowania systemu mamy do czynienia z oddziaływaniem dwóch rodzajów decyzji: stałych na przestrzeni długich odcinków czasu reguł eksploatacyjnych, cen i norm oraz zmienianych na bieżąco decyzji sterowania operacyjnego – wielkości zrzutu, stopnia oczyszczenia itp. Cechą charakterystyczną reguł i przepisów eksploatacyjnych jest to, że nie są one dostosowywane do danych bieżących, np. prognoz hydrologicznych i bieżącego stanu zasobów, lecz są ustalane przy użyciu uśrednionych i długookresowych, np. wieloletnich, charakterystyk zjawisk hydrologicznych, demograficznych czy procesów gospodarczych. W odróżnieniu od tego, decyzje sterowania operacyjnego są podejmowane „tu i teraz”, z uwzględnieniem aktualnego stanu systemu oraz bieżących prognoz.

Struktura przedstawiona na rys. 1.12 jest przykładem struktury wielowarstwowej, por. rys. 1.5. Widać tu wyraźnie, że częstotliwość interwencji, czyli podejmowania decyzji w każdej z warstw jest inna, a także, iż zadania decyzyjne poszczególnych warstw są mocno zróżnicowane. Każde z tych zadań używa prognoz, danych o otoczeniu, systemu wartościowania i oceny wyników oraz pewnych kryteriów wyboru decyzji, ale elementy te nie są jednakowe. W każdym z tych zadań decyzyjnych można użyć, aczkolwiek z różnym stopniem powodzenia, takich narzędzi, jak: modele matematyczne, symulacja, techniki optymalizacyjne, komputerowe wspomaganie decyzji – odsyłamy tu Czytelnika do nieco bliższego omówienia tych zagadnień, zawartego w rozdz. 5.



Rys. 1.12. Hierarchiczna struktura sterowania systemem wodno-gospodarczym

Przykład 4. Przedsiębiorstwo komunikacji lotniczej*)
Pokażemy w tym przykładzie, w bardzo szkicowy sposób, charakter i powiązania działań i decyzji, które występują w przedsiębiorstwie prowadzącym linie lotnicze. Przykład ten jest przedstawiony na rysunku 1.13 w nieco innej formie aniżeli przykłady dotychczasowe (rafineria, system



Rys. 1.13. Perspektywa czasowa i przestrzenna działań w przedsiębiorstwie lotniczym

wodno-gospodarczy). Forma ta eksponuje różnice perspektywy czasowej i przestrzennej, z którą podejmowane są decyzje różnych szczebli w przedsiębiorstwie i może być interesująca dla innych podobnych rozważań; nie narysujemy natomiast struktury (hierarchii) organizacyjnej, pozostawiając ją w opisie tekstowym lub w domyśle.

Systemem fizycznym linii lotniczych, czyli systemem sterowanym w znaczeniu używanym w poprzednich przykładach, jest posiadana flota samolotów. W odniesieniu do niej podejmowane są decyzje różnego rodzaju, na przykład:

1) na podstawie prognozy rynku oraz zawieranych umów przedsiębiorstwo postanawia uruchomić nową linię, z częstością lotów dwa na tydzień;
2) na podstawie analizy technicznej i ekonomicznej przedsiębiorstwo decyduje się przejść na inny typ samolotów niż używane do tej pory;

*) Przykład oparty na pracach J.M. Morawskiego.

3) analizując różnice w cenach paliwa, pion ekonomiczny przedsiębiorstwa wydaje okresowe zalecenia co do preferowanych portów tankowania;
4) biorąc pod uwagę znane osiągi samolotu oraz komunikaty meteorologiczne wybiera się trasę przelotu;
5) na podstawie danych o widzialności, o stanie pasa oraz o wyposażeniu lotniska, a także znając swój zapas paliwa, kapitan podejmuje decyzję o lądowaniu;
6) na podstawie wskazań przyrządów pokładowych oraz/lub komend GCA (Ground-Controlled Approach) pilot przechyla samolot w celu wycentrowania ścieżki schodzenia.

W przykładach tych można zauważyć istotne różnice nie tylko pod względem podmiotu podejmującego decyzje (dyrekcja przedsiębiorstwa, pion eksploatacyjny, kapitan samolotu, pilot), ale przede wszystkim pod względem perspektywy czasowej i perspektywy przestrzennej: dla dyrekcji przedsiębiorstwa horyzontem czasowym rozważań będzie kilka – kilkanaście lat, dla pilota sprowadzającego samolot do lądowania – sekundy lub pojedyncze minuty. Perspektywą przestrzenną dla dyrekcji jest całe przedsiębiorstwo i cały obszar przez nie obsługiwany lub objęty zainteresowaniem na przyszłość; dla pilota schodzącego do lądowania – jego samolot oraz odległość do pasa na lotnisku, czyli setki metrów lub pojedyncze kilometry. Ponadto, perspektywa ta skraca się w miarę zbliżania do chwili przyziemienia.

Mówi się niekiedy*), że decyzje takie jak (1) lub (2) należą do sfery strategii, a decyzje (5) lub (6) związane z manewrami samolotu – do sfery techniki. Między tymi strefami leży sfera taktyki, o pośredniej perspektywie czasowej i przestrzennej. Wśród podanych przykładów należałyby do tej sfery zalecenia co do portów tankowania, a także wybór trasy przelotu.

Można to w syntetyczny i obrazowy sposób przedstawić jak na rys. 1.13. Pole kwadratu na tym rysunku obrazuje cały obszar objęty perspektywą przedsiębiorstwa i jego działaniem. Skala czasu, odmierzona wzdłuż dolnej krawędzi, to horyzonty czasowe – od sekund i minut do miesięcy i lat, skala przestrzenna, wzdłuż prawej krawędzi zaczyna się od setek metrów i sięga tysięcy kilometrów. Wzdłuż przekątnej kwadratu daje się zaznaczyć sfery: racjonalną dla techniki (krótki horyzont czasowy i przestrzenny), racjonalną dla taktyki oraz racjonalną dla działań i decyzji strategicznych. Poza obszarem zbliżonym do przekątnej leżą działania mało racjonalne: pilot podchodzący do lądowania (setki bądź dziesiątki metrów na skali przestrzennej) nie podejmuje swojej decyzji w perspektywie godzin, miesięcy lub lat, dyrekcja przedsiębiorstwa nie podejmuje działań o krótkim horyzoncie czasowym (i nie powinna w działania takie interweniować).

*) J.M. Morawski (1991). Safety vs. Economy: System – Theoretic approach to the problem analysis. W: Proceedings of the Sixth International Symposium on Aviation Psychology (R.S. Jensen, ed.) Columbus, OH, vol. 2, p. 679–684.

Wzdłuż górnej krawędzi kwadratu, dla lepszej ilustracji problemu rozmieszczono w sposób umowny podmioty wykonujące działania lub podejmujące decyzje – od pilota, związanego z najkrótszą skalą czasu, po dyrekcję przedsiębiorstwa, która operuje skalą najdłuższą.

Jest rzeczą dosyć oczywistą, że podmioty operujące w dłuższym horyzoncie czasowym i przestrzennym korzystać będą z informacji bardziej ogólnej, często statystycznej, podczas gdy w sferze „techniki", to jest decyzji szczegółowych i krótkoterminowych, używać się będzie danych o sytuacji aktualnej „tu i teraz". Zaznaczono to umownie wzdłuż lewej krawędzi kwadratu.

Reguła racjonalności działań podyktowana przez przekątną na rys. 1.13 nazwana została przez J.M. Morawskiego „prawem zgodności skal": w danej strefie (lub na danym poziomie) racjonalne są działania, których perspektywa czasowa, przestrzenna i inne (np. stopień szczegółowości) są ze sobą zharmonizowane.

W kontekście rozpatrywanego przykładu warto jeszcze wspomnieć o dwóch sprawach.

Pierwszą z nich są cele działania („funkcje celu"), które stoją przed podmiotami podejmującymi decyzje. Pilot realizujący lądowanie i inne manewry samolotu kieruje się wyłącznie bezpieczeństwem, nie wolno mu myśleć o ekonomii. Niemal to samo dotyczy kapitana, gdy wybiera trasę przelotu, aczkolwiek tutaj czynnik ekonomiczny (oszczędność paliwa, punktualność przybycia czyli prestiż linii) może już odgrywać pewną rolę. W odróżnieniu od tego, w działaniu dyrekcji linii lotniczych ocena ekonomiczna staje się decydująca. Nie oznacza to postawienia bezpieczeństwa lotów na drugim planie: pewien ustalony poziom bezpieczeństwa, zapewniany przez dobór sprzętu, ścisłe przepisy eksploatacji i pilotażu itd. będzie zachowywany. „Decydujący" charakter oceny ekonomicznej oznacza tyle, że przedsiębiorstwo wybierać będzie tylko warianty swojej strategii przynoszące zyski, odrzucając takie, po których spodziewać się można strat.

Drugą sprawą wartą poruszenia jest zwrócenie uwagi na to, że pilot realizujący manewry samolotu steruje obiektem, w którym jest już wiele „wewnętrznych" układów sterowania (autopilot, serwomechanizmy sterów, układy regulacyjne silników, automatyka ciśnienia w kabinie i in.).

Zauważmy na zakończenie, że decyzje związane z działaniami należącymi do obszaru objętego rys. 1.13 nie mogą być sformalizowane na tyle, by można je było podejmować automatycznie, bez udziału osądu ze strony odpowiednich osób lub grup osób. Nie oznacza to rezygnacji ze wspomagania komputerowego; dotyczy to zarówno załogi samolotu, jak również tzw. naziemnej osłony lotu. Może się w tym m.in. zawierać komputerowa symulacja lotu w przyśpieszonej skali czasu.

* * *

Omówione przykłady, mimo ich różnorodności i różnego stopnia szczegółowości opisu służyć mogą potwierdzeniu tezy, iż rzeczywiste struktury sterowania czy zarządzania złożonymi systemami zawierają wiele jednostek decyzyjnych, chociaż nie zawsze postrzegamy to w taki sposób. Ponadto są one hierarchiczne, a cechą charakterystyczną jest to, że na wyższych szczeblach operuje się dłuższym horyzontem czasowym, lecz mniejszą liczbą szczegółów. Mówiliśmy już o tym wstępnie w podrozdziale 1.1, wspominając o wielkościach i modelach agregowanych. Dzięki takiemu podejściu możliwe jest, na przykład na szczeblu zarządzania całym przedsiębiorstwem, zachowanie spojrzenia na całość systemu. Podkreślmy jednak w sposób wyraźny, że rozpatrując model agregowany, zastępczy czy przybliżony przedsiębiorstwa „widzianego przez dyrektora" nie zawsze wolno nam zapomnieć o tym, że wewnątrz tego obiektu istnieją decydenci różnych niższych szczebli, gdyż mogą oni mieć swoje indywidualne interesy, własne cele. Zmienia to bowiem charakter tego obiektu z „pasywnego" na „aktywny", jeśli użyć terminologii wprowadzonej przez Burkowa[*]. Dopiero gdy układ uzależnień i koordynacji między jednostkami decyzyjnymi jest taki, że konflikty są zneutralizowane, można patrzeć na całość bez użycia modeli uwzględniających konflikt, czyli modeli wziętych z teorii gier. Będzie o tym mowa w rozdziale 2.

*) W.N. Burkow (1977). Osnowy matematiczeskoj teorii aktiwnych sistem, Nauka, Moskwa.

Rozdział 2

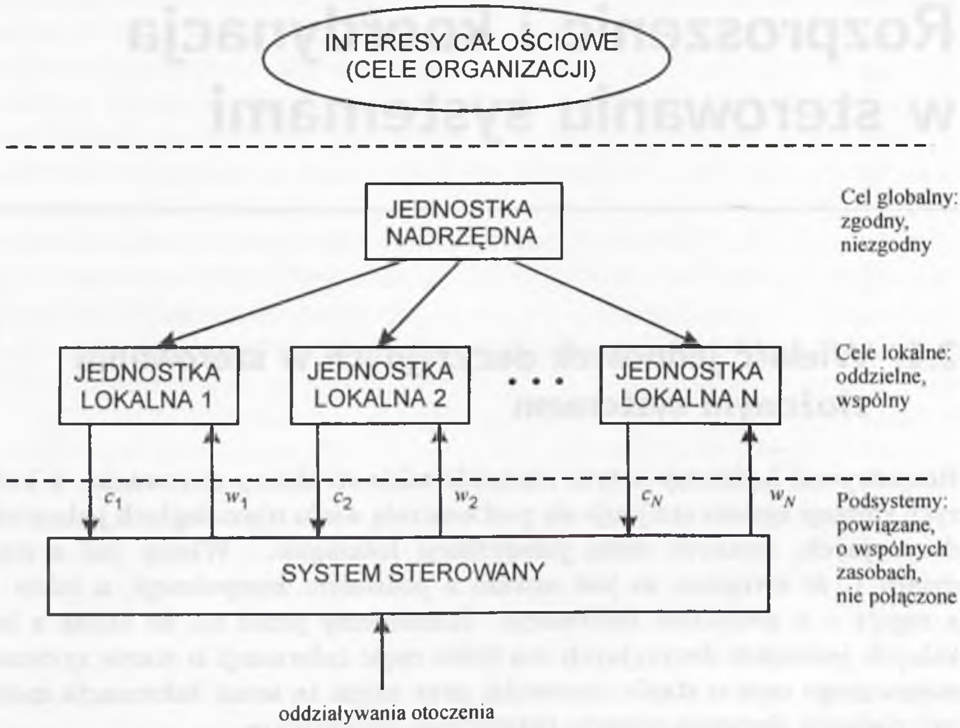
Rozproszenie i koordynacja w sterowaniu systemami

2.1. Wielość jednostek decyzyjnych w sterowaniu złożonym systemem

Rozpatrywać będziemy w tym rozdziale takie struktury sterowania, w których złożony system znajduje się pod kontrolą wielu równoległych jednostek decyzyjnych, zwanych dalej *jednostkami lokalnymi*. Wiemy już z rozdziału 1, że związane to jest zawsze z *podziałem kompetencji*, a także – z reguły – z *podziałem informacji*. Rozumiemy przez to, że każda z lokalnych jednostek decyzyjnych ma tylko część informacji o stanie systemu sterowanego oraz o stanie otoczenia, przy czym ta sama informacja może być niekiedy dostępna różnym jednostkom decyzyjnym.

Zasadniczą strukturę hierarchiczną sterowania złożonym systemem z użyciem jednostek lokalnych oraz odpowiedniej *jednostki nadrzędnej* (zakładamy na razie, że jest ona niezbędna) przedstawia rys. 2.1. Pokazano na nim, że oddziaływania sterujące systemem (częściami systemu) c_1, $c_2, \ldots, c_N$ są określane przez lokalne jednostki decyzyjne, które wykorzystują dostępne im (lokalne) obserwacje $w_1, \ldots, w_N$. Oznacza to, że kompetencje odnośnie sterowania systemem są *rozproszone*. Działanie lokalnych jednostek decyzyjnych koordynowane jest przez jednostkę nadrzędną; całość zaś struktury oraz zasady jej bieżącego działania zbudowano, mając na widoku pewne interesy całościowe.

Do rysunku 2.1 będziemy jeszcze powracali. Na razie przypomnijmy, że każda jednostka decyzyjna dolnego poziomu może mieć swój *cel* lub *cele lokalne*, do których spełnienia będzie starała się dopasowywać podejmowane przez siebie decyzje. Rozróżnić tu trzeba dwie sytuacje. Pierwszą z nich jest ta, gdy cel lokalny może być jednostce narzucony, na przykład w fazie projektowania struktury sterowania. Ma to w szczególności miejsce przy sterowaniu procesami technologicznymi z użyciem rozproszonego sprzętu komputerowego, ale może także być zakładane w tych organizacjach, gdzie osoby podejmujące decyzje lokalne mogą być uważane za nie posiadające własnych celów, czy też powodów do „nierozsądnych" zachowań. Przy swobodzie narzucania łatwo jest osiągnąć zgodność celów lokalnych z ce-



Rys. 2.1. Hierarchiczna (wielopoziomowa) struktura sterowania złożonym systemem

lem jednostki nadrzędnej, czyli tym, którego spełnienie jest jednym z zadań koordynacji.

Drugą sytuacją jest ta, gdzie cele jednostek lokalnych (oraz samo istnienie tych jednostek) pochodzą niejako z natury, są z góry dane, jak to ma na przykład miejsce w opisanym w rozdziale 1.5 systemie wodno-gospodarczym, gdzie każdy z odbiorców wody ma swoje własne cele i rozporządza kompetencjami odnośnie swego działania. Koordynacja jest potrzebna, trudno jednak oczekiwać, że cel jednostki nadrzędnej będzie zgodny z celami lokalnymi.

2.2. Struktura systemu sterowanego i relacje celów

Zgodnie z rys. 2.1 przebieg procesów w systemie podlega sterowaniu za pomocą wielkości $c_1, \ldots, c_N$, przy czym każde c_i leży w gestii oddzielnej (lokalnej) jednostki decyzyjnej. Wielkość c_i może oczywiście być wektorem, to znaczy obejmuje łącznie wszystkie zmienne decyzyjne leżące w kompetencji danej jednostki.

Załóżmy teraz, że lokalne jednostki decyzyjne zdają sobie sprawę z własnych interesów, tak że zbiór zadań sterowania (zadań decyzyjnych) stojących przed tymi jednostkami można zapisać w następującej formie

$$\begin{array}{l} \underset{c_1}{\text{maksymalizować}}\ Q_1^*(c_1, \ldots, c_N) \\ \ldots \\ \underset{c_N}{\text{maksymalizować}}\ Q_N^*(c_1, \ldots, c_N) \end{array} \tag{2.1}$$

W zapisie tym musieliśmy uwzględnić to, że wartość lokalnej funkcji celu Q_i^*, która zostanie osiągnięta w systemie przez jednostkę decyzyjną i przez dobór jej decyzji c_i zależeć może od decyzji wszystkich jednostek pozostałych. W sformułowaniu (2.1) nie zaznaczono, że osiągana wartość funkcji celu może zależeć także od oddziaływań otoczenia oraz że decyzje lokalne mogą podlegać ograniczeniom.

Sformułowanie (2.1) nie jest najbardziej ogólne. Należałoby bowiem zacząć od stwierdzenia, że lokalna jednostka decyzyjna postrzega lokalne interesy, ale że nie muszą one polegać na maksymalizacji czegokolwiek, tym bardziej zaś na maksymalizacji skalarnej funkcji celu Q_i^*. Uznajmy zatem sformułowanie (2.1) za pewien przykład celów lokalnych.

Chcielibyśmy wiedzieć, czy sterowanie systemem za pomocą samych tylko jednostek lokalnych jest dobre czy złe; nie można tego rozstrzygnąć bez wprowadzenia pojęcia wspomnianych już interesów całościowych, związanych z działaniem całości systemu.

Zestaw wielu jednostek decyzyjnych oddziałujących na wspólny dla nich system można nazwać organizacją; nasze pytanie brzmi wówczas: czy ta organizacja ma jakiś *cel* lub *cele globalne* w odróżnieniu od wymienionych już celów lokalnych?

Typowe będą następujące dwa wymagania całościowe:

a) system powinien być stabilny, czyli działania lokalnych jednostek decyzyjnych powinny prowadzić do stabilnego punktu równowagi,

b) osiągnięty stabilny punkt równowagi powinien być taki, by cel lub cele globalne były spełniane w sposób *zadowalający*; można także wymagać aby był to punkt *optymalny*.

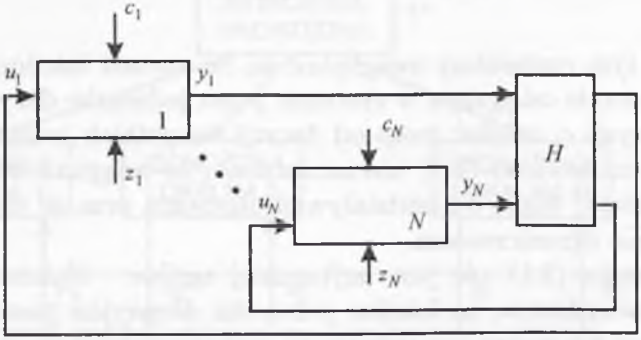
Wybór sposobu spełnienia interesów całościowych w dużym stopniu zależy od dwóch czynników, a mianowicie: od struktury systemu sterowanego oraz od wzajemnego stosunku celu globalnego i celów lokalnych.

Przejdziemy teraz do opisu systemu sterowanego z rys. 2.1. Załóżmy, że jest on zestawem podsystemów opisanych przez związki między wejściami i wyjściami, które można przedstawić w postaci

$$y_i = F_i(c_i, u_i, z_i), \quad i = 1, \ldots, N \tag{2.2}$$

gdzie: y_i jest wyjściem podsystemu (wektorem złożonym z pojedynczych wielkości wyjściowych), c_i jest sterowaniem (wektorem wielkości sterujących), z_i jest wejściem (wektorem wielkości wejściowych) pochodzącym z otoczenia oraz u_i jest wejściem pochodzącym z pozostałych części systemu, które może mieć różny charakter, zależnie od tego, który z trzech dalej opisanych przypadków ma miejsce*).

Pierwszy przypadek przedstawiony jest na rysunku 2.2. Wyjścia jednych podsystemów połączone są z wejściami innych, co daje się opisać za pomocą



Rys. 2.2. Podsystemy powiązane, $u = Hy$

macierzy H, złożonej z zer i jedynek. Część H_i tej macierzy informuje, skąd pochodzi wejście u_i; może ono być dołączone do wyjść jakiegokolwiek spośród podsystemów, tak że

$$u_i = H_i y \tag{2.3}$$

gdzie: $y = \left[y_1^T \ldots y_N^T\right]^T$, czyli jest wektorem złożonym z podwektorów $y_1, \ldots, y_N$, a macierz H_i ma tyle wierszy ile elementów ma wektor u_i oraz tyle kolumn ile wynosi wymiar całego wektora y. Mamy zatem

$$H = \begin{bmatrix} H_1 \\ \vdots \\ H_N \end{bmatrix}$$

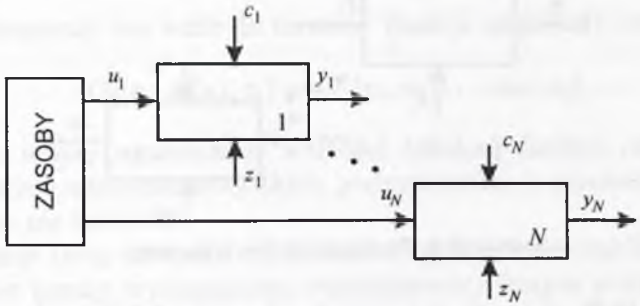
Wektor wyjść y zależeć będzie – przez związki (2.2) oraz macierz połączeń H od wszystkich sterowań i wszystkich oddziaływań otoczenia; zapiszemy zatem

*) Rozpatrujemy tu sytuacje podstawowe. W konkretnym systemie mogą one występować jednocześnie.

$$y = K(c, z), \quad u_i = H_i K(c, z), \quad y_i = K_i(c, z) \tag{2.4}$$

gdzie $K(\cdot)$ jest funkcją dającą się wyznaczyć w postaci jawnej tylko w prostych przypadkach, na przykład gdy zależności (2.2) są liniowe.

Na rysunku 2.3 przedstawiono przypadek drugi. Podsystemy nie są teraz połączone między sobą, wielkości u_i określają zużycie zasobów i podlegają



Rys. 2.3. Podsystemy o wspólnych zasobach, $\sum_{i=1}^{N} u_i \leqslant u_0$

bezpośrednio decyzjom lokalnym, na równi ze sterowaniami c_i. Łączne zużycie zasobów nie może przekroczyć wartości zapasu, stąd też decyzje u_i podlegają globalnemu ograniczeniu nierównościowemu

$$u_1 + \ldots + u_N \leqslant u_0 \tag{2.5}$$

Jak widać, mimo iż podsystemy nie są ze sobą połączone (nie oddziałują na siebie bezpośrednio), nie mogą one być rozpatrywane niezależnie ze względu na wspólność zasobów. Ten właśnie czynnik sprawia, że podsystemy z rys. 2.3 są *systemem*, czyli muszą być rozpatrywane w sposób łączny.

Wreszcie rysunek 2.4 przedstawia przypadek trzeci. Podsystemy nie są połączone, nie występuje też ograniczenie zasobów – wejścia u_i są całkowicie swobodne. Istnieje jednak wspólny cel działania systemu.

Przyjrzyjmy się teraz lokalnym zadaniom sterowania, to jest lokalnym zadaniom decyzyjnym, ich funkcjom celu i uwarunkowaniom.

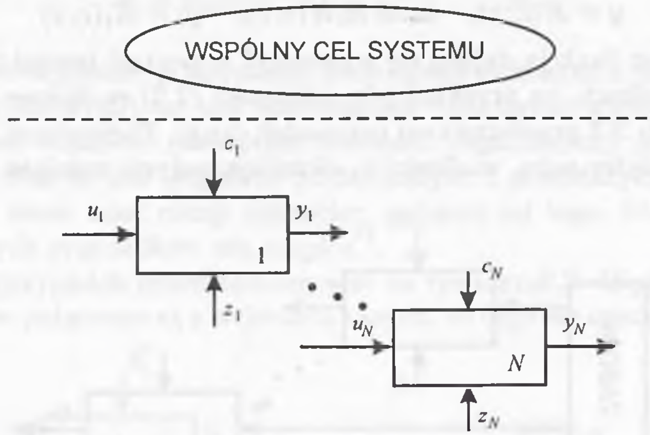
Załóżmy istnienie lokalnej funkcji celu o postaci

$$Q_i^0(c_i, u_i, y_i) \tag{2.6}$$

oraz istnienie pewnych lokalnych ograniczeń, na przykład o postaci

$$(c_i, u_i) \in CU_i(z_i) \tag{2.7}$$

$$y_i \in Y_i \tag{2.8}$$



Rys. 2.4. Podsystemy nie połączone

Zależność (2.7) oznacza wymaganie, że wartości (c_i,u_i) muszą się zawierać w pewnym zbiorze dopuszczalnym, przy czym granice tego zbioru mogą być zależne od wartości z_i, czyli od czynnika zewnętrznego. Wzór (2.8) wskazuje, że wartość wyjścia y_i ma także być ograniczona do pewnego zbioru Y_i.

Postać funkcji celu (2.6) jest dosyć ogólna, podobnie jak zależności ograniczające (2.7) oraz (2.8). Można także napotkać inne sformułowania; trudno jest tylko wyobrazić sobie realne zadania optymalizacji bez jakichkolwiek ograniczeń.

Gdy do funkcji celu (2.6) podstawimy zależność (2.2), to otrzymamy nową funkcję

$$Q_i(c_i, u_i, z_i) = Q_i^0(c_i, u_i, F_i(c_i, u_i, z_i)) \tag{2.9}$$

która ma tę zaletę, że łatwiej jest się nią posługiwać w konkretnych rozważaniach, tę natomiast wadę, że wyraża ona poprawnie wartość pierwotnej funkcji celu tylko wtedy, gdy związek (2.2), to znaczy model $F_i(\cdot)$ podsystemu jest znany z dostateczną dokładnością.

W dalszym ciągu korzystać będziemy z postaci (2.9). Jeśli mamy system połączony jak na rys. 2.2, to podstawiając w miejsce u_i zależność (2.4) otrzymamy

$$Q_i(c_i, H_i K(c, z), z_i) = Q_i^*(c_1, \ldots, c_N, z) \tag{2.10}$$

co wskazuje, że stopień osiągnięcia celu przez lokalną jednostkę decyzyjną zależy nie tylko od jej decyzji c_i oraz od czynników zewnętrznych podsystemu z_i, ale także od całego c i od całego z, zatem od sterowań i wejść zewnętrznych w całości systemu.

W systemie o strukturze według rys. 2.3 będzie nieco inaczej. Lokalne zadania decyzyjne są od siebie niezależne tak długo, jak ograniczenie nierównościowe (2.5) pozostaje bez wpływu, to znaczy dopóki suma poboru zasobów $u_1 + \ldots + u_N$ jest mniejsza od u_0. Z chwilą gdy ograniczenie (2.5) zacznie ingerować, czyli gdy związek (2.5) stanie się równością, mieć będziemy określoną zależność u_i od decyzji pozostałych, co można dogodnie zapisać w postaci wzoru

$$u_i = g_i(u_1, \ldots, u_N) = g_i(u). \tag{2.11}$$

Gdy podstawimy ten wzór do formuły funkcji celu (2.9), otrzymamy

$$Q_i(c_i, g_i(u), z_i) = Q_i^*(c_i, u_1, \ldots, u_N, z_i) \tag{2.12}$$

czyli znowu mamy uzależnienie wartości lokalnej funkcji celu od decyzji podejmowanych odnośnie wszystkich podsystemów, a mianowicie od pobieranych przez nie zasobów.

Rozpatrzmy teraz interesy całościowe. Ograniczenie zasobów (2.5) było już w istocie rzeczy wymaganiem całościowym. Innym przykładem może być postulat by pewna funkcja określona na wyjściach podsystemów

$$J^*(y_1, \ldots, y_N) \tag{2.13}$$

była maksymalizowana lub też by osiągała wartości leżące powyżej pewnego progu, zwanego *poziomem aspiracji*. Bardziej ogólna jest jednak sytuacja, gdy globalna funkcja celu jest określona na wszystkich zmiennych lokalnych

$$J^0(c_1, \ldots, c_N, u_1, \ldots, u_N, y_1, \ldots, y_N) = J^0(c, u, y) \tag{2.14}$$

co należy rozumieć na przykład w ten sposób, że dla podmiotu decyzyjnego odpowiedzialnego za całość systemu nie jest obojętne, jakie będą wartości lokalnych wejść i sterowań; na przykład ilości zużywanych zasobów. Postać (2.14) można przekształcić przy użyciu zależności (2.2), otrzymując

$$J(c_1, \ldots, c_N, u_1, \ldots, u_N, z_1, \ldots, z_N) = J(c, u, z) \tag{2.15}$$

Do postaci (2.15) można odnieść te same komentarze, co do omawianej wcześniej postaci lokalnej funkcji celu (2.9) – zwłaszcza jeśli chodzi o poprawne wyrażanie celu właściwego tj. (2.14).

Zauważmy, że o ile systemy z rys. 2.2 i z rys. 2.3 muszą być rozpatrywane całościowo, czyli właśnie jako systemy ze względu na powiązania fizyczne lub wspólność zasobów, to zbiór podsystemów z rys. 2.4 jest systemem wyłącznie przez to, że istnieje globalna funkcja celu (2.14) bądź (2.13), zatem jest on systemem tylko wtedy, gdy sformułowane zostały interesy całościowe.

Sytuacja naszych rozważań jest obecnie następująca: mamy system o strukturze jak na rys. 2.2, rys. 2.3 lub rys. 2.4, istnieją lokalne jednostki sterujące, istnieje świadomość interesów całościowych – jakie będzie działanie systemu sterowanego?

Problem ten wymaga dalszych rozważań.

2.3. Sterowanie lokalne i potrzeba jednostki nadrzędnej

Weźmy na początek pod uwagę system najprostszy, ten z rys. 2.4, to jest taki, w którym nie ma połączeń między podsystemami i nie ma ograniczenia zasobów. Załóżmy istnienie lokalnych jednostek decyzyjnych $1, \ldots, N$ i porównajmy następujące trzy przypadki lokalnych celów:

1) jednostki lokalne zainteresowane są tylko rezultatami lokalnymi, tak że ich zadania mają postać

$$\underset{(c_i, u_i) \in CU_i}{\text{maksymalizować}} \; Q_i^0(c_i, u_i, y_i), \quad i = 1, \ldots, N \tag{2.16}$$

2) jednostki lokalne zainteresowane są rezultatami, czyli wyjściami nie tylko własnymi, ale także innych podsystemów, tak że zadania lokalne tworzą następujący zbiór

$$\begin{aligned} &\underset{(c_1, u_1) \in CU_1}{\text{maksymalizować}} \; Q_1^0(c_1, u_1, y_1, \ldots, y_N) \\ &\ldots \\ &\underset{(c_N, u_N) \in CU_N}{\text{maksymalizować}} \; Q_N^0(c_N, u_N, y_1, \ldots, y_N) \end{aligned} \tag{2.17}$$

3) funkcje celu wszystkich jednostek lokalnych są identyczne, tak że zadania lokalne mają postać

$$\begin{aligned} &\underset{(c_i, u_i) \in CU_i}{\text{maksymalizować}} \; Q(c_1, \ldots, c_N, u_1, \ldots, u_N, y_1, \ldots, y_N), \\ &i = 1, \ldots, N \end{aligned} \tag{2.18}$$

Przypadek (1) odpowiada prawdziwej niezależności jednostek lokalnych, sterujących – przypomnijmy – niezależnymi podsystemami z rys. 2.4.

W przypadku (2) lokalne jednostki decyzyjne są między sobą uzależnione przez wspólność rezultatów, czyli przez to, że każda z nich przywiązuje jakąś wartość do wyjść podsystemów innych niż swój własny. Jednostki lokalne znajdują się w sytuacji, którą można opisać jako grę N-osobową.

Przypadek (3) jest specyficzny dlatego, że jednostki lokalne nie mają w istocie rzeczy własnych celów, starają się optymalizować jedną i tę samą funkcję celu. Wspominaliśmy o tej sytuacji w rozdziale 1.4, mówiąc, że jednostki lokalne tworzą tu tak zwany *zespół*.

Nie wprowadziliśmy jeszcze wymagań całościowych, celu globalnego, poza – być może – postulatem, by lokalne procesy decyzyjne doprowadzały do stabilnego punktu równowagi.

Rozpatrzmy teraz przypadek (1), niezależnych celów lokalnych (oraz pamiętajmy – niezależnych podsystemów) z dodaniem celu globalnego. Przyjmujemy zatem, że jest ktoś poza jednostkami lokalnymi zainteresowany tym, by optymalizować swoją funkcję celu, na przykład

$$\text{maksymalizować } J^*(y_1, \dots, y_N) \tag{2.19}$$

Gdyby ta nowa jednostka decyzyjna miała w swym ręku część spośród sterowań w podsystemach, czyli mogła podejmować bezpośrednio niektóre decyzje, byłaby ona w sytuacji dającej się opisać jako gra $(N+1)$-osobowa z lokalnymi jednostkami decyzyjnymi. Zauważmy, że byłaby ona w podobnej sytuacji gry z jednostkami lokalnymi z przypadku (2), a także z zespołem z przypadku (3).

Jednostka zainteresowana rezultatami całościowymi może nie interweniować, czyli nie wchodzić w grę z jednostkami lokalnymi, jeśli wyniki działań jednostek lokalnych są takie, iż postulat (2.19) jest spełniony w zadowalającym stopniu, czyli otrzymana wartość J^* czyni zadość aspiracjom całościowym.

Widoczne jest na przykład, że jeśli globalna funkcja celu będzie identyczna z funkcją celu zespołu, o którym mowa jako o przypadku (3), interwencja nie jest potrzebna: jednostki lokalne same ustalą najbardziej sprzyjające celowi ogólnemu wartości c_i, u_i.

Podobnie, jeżeli w przypadku (1) cel globalny miałby postać

$$\text{maksymalizować } \Psi\left(Q_1^0(c_1, u_1, y_1), \dots, Q_N^0(c_N, u_N, y_N)\right) \tag{2.20}$$

gdzie $\Psi(\cdot)$ jest funkcją ściśle monotoniczną*), to decyzje c_i, u_i optymalne z punktu widzenia celów lokalnych służą zarazem najlepiej interesom całościowym i żadna interwencja osiąganych wyników nie poprawia (nie będzie tak jednak w przypadku powiązań między podsystemami wprost bądź poprzez ograniczone zasoby).

Gdy obowiązuje związek (2.20) mówimy, że cel globalny jest *zgodny* (zharmonizowany) z celami lokalnymi, w przypadku przeciwnym mówimy o *niezgodności celów.*

Jeżeli uznamy bierną postawę za niewłaściwą, czyli nie akceptujemy tego co powstaje w wyniku działań lokalnych jednostek decyzyjnych, musimy do układu sterowania wprowadzić nadrzędną jednostkę decyzyjną i przydzielić jej określone kompetencje. Warto tu wrócić do rys. 2.1, gdzie jednostka taka została wprowadzona z zadaniem odpowiedniego działania bieżącego.

*) Funkcja $\Psi(a_1, \dots, a_N)$ jest ściśle monotoniczna, jeśli przy wzroście któregokolwiek argumentu a_i wzrasta także wartość funkcji. Przykładem jest suma $\sum_i^N a_i$, lub iloczyn $\prod_i^N a_i$, z warunkiem, że czynniki w iloczynie są dodatnie.

Istotę *koordynacji*, czyli *interwencji nadrzędnej* w lokalne zadania decyzyjne warto przedstawić na następującym prostym przykładzie.

Mamy zbiór nie powiązanych podsystemów z rys. 2.4 oraz zupełnie niezależne cele lokalne, czyli przypadek (1). Postawmy przed jednostką nadrzędną, czyli przed *koordynatorem* zadanie maksymalizacji funkcji określonej na wyjściach podsystemów, czyli (2.19). Niechaj ma ona prawo wymagać od każdego podsystemu ustalonej przez siebie wartości wyjścia y_i. Traktując rzecz bardzo powierzchownie, cel (2.19) może być spełniony w tym schemacie działania, a jednostki lokalne mogą jeszcze zachować pewną swobodę funkcjonowania i optymalizacji swoich celów, jeśli tylko żądana wartość wyjścia y_i może być otrzymana w podsystemie za pomocą różnych kombinacji wartości c_i, u_i. Realistycznie trzeba tu jednak założyć, że nie każda wartość y_i, która byłaby dogodna dla koordynatora może być osiągnięta w podsystemie, ze względu na istnienie lokalnych ograniczeń, na przykład (2.7). Ograniczenia te muszą być znane jednostce nadrzędnej razem z aktualną wartością czynników zewnętrznych z_i, wraz z ich przeliczeniem na osiągalne wartości y_i; oznacza to, iż jednostka nadrzędna musi mieć w swej dyspozycji bardzo bogatą informację o każdym z podsystemów.

Stanowi to istotną wadę sposobu koordynacji, pomyślanego jako narzucanie wartości określonych zmiennych w podsystemach, szczególnie zaś zmiennych wyjściowych.

Celem opisanego przykładu było wskazanie, że nawet w tym najprostszym przypadku, nie powiązanych podsystemów i niezależnych celów lokalnych, może istnieć potrzeba wprowadzenia jednostki nadrzędnej i działania koordynującego.

W przykładzie użyliśmy tak zwanej *koordynacji bezpośredniej*, polegającej na narzucaniu jednostkom lokalnym zadanych wartości pewnych zmiennych w podsystemach; drugą ważną metodą jest *koordynacja za pomocą cen*; będzie o niej mowa nieco dalej.

2.4. Metody koordynacji

Systemy o wspólnych zasobach

Przyjmijmy teraz, że system z rys. 2.3, o ograniczonych wspólnych zasobach, jest sterowany przez lokalne jednostki decyzyjne. Niech decyzje o wykorzystaniu zasobów u_i będą w ich gestii, tak że każde z zadań lokalnych brzmi tak, jak wskazywał wzór (2.16):

$$\underset{(c_i,u_i)\in CU_i(z_i)}{\text{maksymalizować}} \; Q_i^0(c_i, u_i, y_i)$$

z ograniczeniem (2.5): $u_1 + \ldots + u_N \leqslant u_0$.

Z punktu widzenia lokalnej jednostki decyzyjnej może się okazać, że jej decyzja u_i nie daje się zrealizować, jeśli inne jednostki ubiegły ją w poborze zasobów.

Powstałą sytuację można rozpatrywać na dwa sposoby. Pierwszy z nich polega na rozróżnieniu fazy *przygotowania* i fazy *wykonania decyzji.* W fazie przygotowania jednostki lokalne negocjują między sobą swoje przyszłe decyzje, będąc w sytuacji wzajemnego konfliktu, tak długo aż uzgodnią wartości $u_1, \ldots, u_N$, które nie naruszają ograniczenia zasobów. Następnie decyzje te zostaną wykonane, to znaczy następuje działanie poszczególnych podsystemów.

Drugie podejście polega na wprowadzeniu jednostki nadrzędnej, która mogłaby przydzielać zasoby lub w inny sposób zadbać o nieprzekroczenie wartości u_0 stojącej do dyspozycji systemu. Przydzielenie z góry określonych wartości $u_1, \ldots, u_N$ eliminuje oczywiście sprzeczności między jednostkami lokalnymi, neutralizuje ich konflikt.

Jednostka nadrzędna może wybrać jakikolwiek zbiór wartości $u_1, \ldots, u_N$, byle by spełnione było ograniczenie nierównościowe. Istnieje zatem możliwość takiej alokacji zasobów, aby był to wybór najlepszy z punktu widzenia celów jednostki nadrzędnej. Przyjmijmy, że jednostka nadrzędna chce maksymalizować pewną funkcję określoną na wszystkich zmiennych lokalnych, tak że mamy następujący układ zadań decyzyjnych górnego i dolnego poziomu

jednostka nadrzędna (koordynator)

$$\begin{aligned} &\underset{u_1,\ldots,u_N}{\text{maksymalizować}}\ J^0(c_1, u_1, y_1, \ldots, c_N, u_N, y_N) \\ &\text{z ograniczeniem: } u_1 + \ldots + u_N \leqslant u_0 \end{aligned} \tag{2.21}$$

jednostki lokalne

$$\underset{c_i \in C_i(u_i, z_i)}{\text{maksymalizować}}\ Q_i^0(c_i, u_i, y_i) \tag{2.22}$$

W sformułowaniach (2.21), (2.22) trzeba było przyjąć, że granice zbioru dopuszczalnego dla sterowań c_i zależeć mogą od przydzielonych zasobów, czyli od u_i oraz od wejść zewnętrznych z_i, co zapisaliśmy w postaci warunku $c_i \in C_i(u_i, z_i)$.

Nie ma tu konkurencji między jednostkami lokalnymi, jest natomiast sytuacja, którą można przedstawić jako grę między jednostką nadrzędną a zestawem jednostek lokalnych. Poznajemy to po tym, że rezultat J^0 zależy od decyzji lokalnych $c_1, \ldots, c_N$, a rezultaty lokalne zależą od decyzji nadrzędnych $u_1, \ldots, u_N$, przy tym w podwójny sposób: poprzez występowanie argumentu u_i w funkcjach Q_i^0 oraz przez wpływ na obszary dopuszczalnych sterowań.

Opisywaliśmy dotąd sytuację decyzyjną nie zastanawiając się nad tym w jaki sposób jednostka nadrzędna i jednostki lokalne będą rozwiązywać swoje zadania. W szczególności zauważmy, że będą one posługiwać się modelami, a zwłaszcza modelem (2.2), wiążącym wyjścia podsystemu z wejściami, $y_i = F_i(c_i, u_i, z_i)$.

Dzięki temu w dalszych rozważaniach możemy posługiwać się funkcjami $J(c, u, z)$ oraz $Q_i(c_i, u_i, z_i)$ zamiast J^0 i Q_i^0, por. (2.9), (2.15).

W sytuacji ograniczonych zasobów byłoby rzeczą naturalną dać specjalne uprawnienia jednostce nadrzędnej – będzie ona w tej grze „leaderem", osobą która ma prawo do pierwszego ruchu. Pozostałe jednostki decyzyjne, dolnego poziomu, będą „followerami", czyli tymi którzy nadążają za decyzją leadera. Tworzy to określoną hierarchię, stąd taka gra nosi nazwę *gry hierarchicznej*.

Rozpatrzmy przypadek szczególny i sprzyjający, gdy jednostka nadrzędna ma pełną znajomość zadań decyzyjnych jednostek dolnego poziomu. Jednostka nadrzędna jest wówczas w stanie przewidzieć odpowiedź $\hat{c}_i$ na decyzję u_i, czyli przewidzieć jaką optymalną dla siebie decyzję podejmie jednostka dolnego poziomu, gdy przydzielimy jej zasoby u_i

$$\hat{c}_i(u_i, z_i) = \arg \max_{c_i \in C_i(u_i, z_i)} Q_i(c_i, u_i, z_i) \tag{2.23}$$

We wzorze (2.23) podkreślono, że wartość c_i optymalna dla jednostki lokalnej zależy także od czynnika zewnętrznego z_i.

Po obliczeniu przewidywanych decyzji $\hat{c}_i(\cdot)$ dla wszystkich podsystemów, jednostka nadrzędna może rozwiązać swoje własne zadanie decyzyjne

$$\begin{aligned} &\underset{u_1, \ldots, u_N}{\text{maksymalizować}} \;\; J(\hat{c}(u, z), u, z) \\ &\text{z ograniczeniem: } u_1 + \ldots + u_N \leqslant u_0 \end{aligned} \tag{2.24}$$

gdzie $\hat{c}(u, z) = (\hat{c}_1(u_1, z_1), \ldots, \hat{c}_N(u_N, z_N))$, którego wynikiem jest wartość $\hat{u}(z) = (\hat{u}_1(z), \ldots, \hat{u}_N(z))$, określająca optymalną alokację zasobów z punktu widzenia jednostki nadrzędnej. Zaznaczyliśmy, że alokacja ta zależeć może od wpływu zewnętrznego z. Jest to o tyle ważne, że prowadzi do wniosku, iż jednostka nadrzędna musi mieć nie tylko pełną znajomość zadań decyzyjnych jednostek lokalnych, ale także aktualną informację o podsystemach, mianowicie o wartościach z_i.

Gry hierarchiczne mają bogatą literaturę*) i rozwiniętą teorię. Tutaj wspomnimy tylko, że w przypadku zgodności celu nadrzędnego z celami lokalnymi, to jest w sytuacji gdy zadanie jednostki nadrzędnej można wyrazić w postaci

$$\text{maksymalizować } \Psi(Q_1(c_1, u_1, z_1), \ldots, Q_N(c_N, u_N, z_N)) \tag{2.25}$$

gdzie znaczenie Ψ jest takie samo jak we wzorze (2.20), to znaczy jest to funkcja ściśle monotoniczna, wynik omawianej gry hierarchicznej będzie dla leadera optymalny.

Oznacza to, że rezultat otrzymany przez leadera w postaci wartości funkcji Ψ będzie taki sam, jak gdyby podejmował on wszystkie decyzje $c_1, \ldots, c_N$ wprost, bez pośrednictwa lokalnych jednostek decyzyjnych.

Przydzielając zasoby $u_1, \ldots, u_N$ stosowaliśmy *bezpośrednią metodę koordynacji*, o której była już mowa w rozdz. 2.3. Rozpatrzmy teraz jak można realizować przydział zasobów za pomocą *cen*.

Rozpoczniemy znowu od sytuacji niezgodności celów, to jest gdy funkcja celu jednostki nadrzędnej nie ma postaci szczególnej (2.25).

Załóżmy, że jednostka nadrzędna ustala cenę η za jednostkę zużywanych zasobów, przy czym dzieje się to zgodnie z zasadą gry leader–follower, to znaczy jednostka nadrzędna ma pierwszy ruch.

Zadania decyzyjne dolnego poziomu ulegają modyfikacji, a mianowicie pojawia się w nich koszt zużywanych zasobów $\eta^T u_i$ (jak pamiętamy u_i może być wektorem). Znając w całości zadania decyzyjne dolnego poziomu, jednostka nadrzędna może przewidzieć (obliczyć) ich reakcje

$$(\hat{c}_i(\eta, z_i), \hat{u}_i(\eta, z_i)) = \arg \max_{(c_i, u_i) \in CU_i(z_i)} [Q_i(c_i, u_i, z_i) - \eta^T u_i] \tag{2.26}$$

Jednostka lokalna sama decyduje teraz o poborze zasobów, stąd przewidywać trzeba zarówno jej decyzje $\hat{c}_i(\cdot)$, jak decyzje $\hat{u}_i(\cdot)$.

Wzór (2.26) wskazuje, że przewidywane decyzje lokalne są zależne od ceny η, a także od wpływów zewnętrznych z_i. Jednostce nadrzędnej pozostaje określić cenę dla siebie optymalną. Można to zrobić, rozwiązując zadanie

$$\begin{aligned} &\underset{\eta}{\text{maksymalizować}} \quad J(\hat{c}(\eta, z), \hat{u}(\eta, z), z) \\ &\text{z ograniczeniem: } \hat{u}_1(\eta, z_1) + \ldots + \hat{u}_N(\eta, z_N) \geqslant u_0 \end{aligned} \tag{2.27}$$

*) Patrz np. J.B. Germejer (1976). Igry s neprotiwopołożnymi interesami. Nauka, Moskwa; T. Basar, G.J. Olsder (1982). Dynamic Noncooperative Game Theory. Academic Press, London.

Zadania (2.26), (2.27) są nieco inne niż przy bezpośredniej metodzie koordynacji, to jest przy alokacji zasobów, określanej w jednostkach fizycznych. Warto wspomnieć, że wyliczana przez jednostkę nadrzędną zależność przewidywanego zużycia zasobów od ceny η nosi w ekonomice nazwę „funkcji popytu". Pisząc: $\hat{u}_i(\eta, z_i)$, podkreślaliśmy, iż funkcja popytu może się zmieniać. Zwróćmy też uwagę, że przy sprawowaniu kontroli nad zasobami metodą cen jednostka nadrzędna operuje jedną ceną, niezależnie od liczby podsystemów, podczas gdy przy metodzie alokacji bezpośredniej musi ona ustalić tyle oddzielnych wartości u_i (przydziałów), ile jest podsystemów. Powstać może wrażenie, że metoda cen jest zawsze lepsza; trzeba więc podkreślić, że zachowanie występującego w tym zadaniu globalnego ograniczenia zasobów jest całkowicie pewne tylko przy alokacji bezpośredniej, to jest przy użyciu (2.24). Operując cenami zdajemy się na wiarygodność funkcji popytu $\hat{u}_i(\cdot)$, gdyż same decyzje u_i podejmować będą jednostki dolnego poziomu. Poprawność obliczonych funkcji popytu zależy od znajomości lokalnych zadań decyzyjnych, w szczególności od występujących tam funkcji celu $Q_i(c_i, u_i, z_i)$ oraz aktualnych warunków zewnętrznych, wyrażonych parametrem z_i. Napotykamy zatem na potrzebę uwzględnienia – przy wyborze sposobu sterowania – wiarygodności informacji o sterowanych podsystemach i regułach zachowania się ich lokalnych decydentów.

Dodajmy jeszcze do dyskusji o „alokacji bezpośredniej" i „koordynacji cenowej" ten dodatkowy aspekt, że alokacja bezpośrednia może być niepraktyczna lub nawet niemożliwa gdy podsystemów jest bardzo wiele oraz gdy lokalne jednostki decyzyjne nie są podatne na przestrzeganie przydziałów.

Przy koordynacji zużycia zasobów metodą cen istnieje taka szczególna wartość ceny $\tilde{\eta}$, przy której następuje zużycie całego zapasu, to znaczy ograniczenie (2.5) spełnione jest równościowo. Cena ta nosi nazwę *ceny równowagi* dla zasobów. Ma ona pewną specyficzną własność, mianowicie odpowiadający jej zbiór decyzji lokalnych $\hat{c}_i(\tilde{\eta}), \hat{u}_i(\tilde{\eta})$, $i = 1, \ldots, N$ jest taki, iż następuje ekstremalizacja sumy lokalnych funkcji celu. Jeżeli celem jednostki nadrzędnej nie jest akurat maksymalizacja tej sumy, cena równowagi może nie być optymalna z nadrzędnego punktu widzenia, stąd w sformułowaniu (2.27) mówimy o maksymalizacji względem η. Przy alokacji bezpośredniej jednostka nadrzędna operuje większą liczbą zmiennych decyzyjnych i może wywołać decyzje lokalne c_i, które są dla niej bardziej sprzyjające. Odpowiednikiem tego mogłoby być użycie różnych cen η_i wobec różnych podsystemów; wiadomo, że prowadzi to do zbioru decyzji lokalnych, które maksymalizują sumę ważoną lokalnych funkcji celu Q_i, ze współczynnikami wagi zależnymi od stosunków między cenami. Z uwag tych wynika, że w sytuacji zgodności celów, czyli gdy jednostka nadrzędna ma zadanie o postaci (2.25) użycie cen na zasoby doprowadzi do wyniku całkowicie optymalnego dla jednostki nadrzędnej jeśli $\Psi(\cdot)$ ma postać sumy lub sumy ważonej (trzeba wówczas użyć cen zróżnicowanych). Przy alokacji bezpośredniej wynik optymalny można było uzyskać także dla innych postaci $\Psi(\cdot)$, na przykład dla iloczynu.

Nie należy przeoczyć, że pewność zachowania ograniczenia na zasoby, która jest zaletą metody bezpośredniej, może także obrócić się na niekorzyść. Można tu bowiem dokonać alokacji zasobów nie zważając na funkcje celu jednostek lokalnych, doprowadzając zatem do sytuacji poprawnej tylko z punktu widzenia jednostki nadrzędnej. Przy metodzie cen funkcje celu odgrywają znacznie większą rolę i nie mogą być ignorowane.

Sterowanie systemem powiązanym; koordynacja bezpośrednia

Zacznijmy teraz rozpatrywać system powiązany (rys. 2.2), to jest taki, gdzie podsystemy oddziałują na siebie wzajemnie poprzez wyjścia i wejścia. Załóżmy, że każda z lokalnych jednostek decyzyjnych ma do dyspozycji sterowanie c_i odpowiedniego podsystemu oraz stawia sobie za zadanie maksymalizację lokalnej funkcji celu. Nawet jeśli ta funkcja jest zdefiniowana tylko na zmiennych lokalnych jako $Q_i^0(c_i, u_i, y_i)$ bądź równoważnie $Q_i(c_i, u_i, z_i)$, zdajemy sobie sprawę, że osiągany rezultat lokalny zależy od decyzji w pozostałych podsystemach, stamtąd bowiem pochodzi wejście u_i. Była już o tym mowa poprzednio, a konkluzją staje się stwierdzenie, że jednostki lokalne znajdują się w sytuacji wzajemnego konfliktu.

Konflikt jednostek lokalnych można zneutralizować, wprowadzając jednostkę nadrzędną, która miałaby w swej gestii, w ten lub inny sposób, wartości wejść u_i bądź wyjść y_i w całym systemie.

Najprostszym sposobem wyeliminowania konfliktów, które powstają między lokalnymi jednostkami decyzyjnymi z powodu połączeń między podsystemami jest narzucenie wartości na te połączenia, czyli ich „zamrożenie". Narzucenie wartości na zmienne u_i oraz y_i w całym systemie będzie *koordynacją bezpośrednią*, jej instrumentami są bowiem wprost wartości zmiennych występujących w sterowanym systemie.

Wystarczy oczywiście zadać z góry tylko wartości wyjść y_i albo tylko wejść u_i, ze względu na istniejące połączenia. Jeśli oprócz połączeń istnieje ograniczenie wspólnych zasobów, trzeba koordynacją objąć także zużycie zasobów, podobnie jak to rozważaliśmy wcześniej dla systemu o wspólnych zasobach. System z rys. 2.2 rozpatrywać będziemy, dla uproszczenia, bez ograniczenia na zasoby. Korzystając ze wzorów (2.9), (2.2) oraz (2.7) sformułujemy lokalne zadanie decyzyjne jako

$$\left.\begin{aligned} &\text{maksymalizować}\ Q_i(c_i, u_i, z_i) \\ &\qquad\quad {}_{c_i, u_i} \\ \text{przy warunkach}\quad & \\ &u_i = H_i y_d \\ &F_i(c_i, u_i, z_i) = y_{di} \\ &(c_i, u_i) \in CU_i(z_i) \end{aligned}\right\} \tag{2.28}$$

W zadaniu tym występuje jako parametr wektor y_d, który oznacza narzucone przez jednostkę nadrzędną wartości wyjść dla wszystkich podsystemów $y_d = [y_{d1}^T \dots y_{dN}^T]^T$. Parametr y_d określa także przez macierz H_i wartość wejścia u_i, co należy rozumieć jako *spodziewaną* wartość wejścia u_i – wartość, która wystąpi gdy wszystkie podsystemy spełnią narzucone warunki, to jest dokonają takiego wyboru decyzji c_i, by wyjście danego podsystemu było równe wartości narzuconej y_{di}.

Rozwiązanie zadania decyzyjnego (2.28) w postaci decyzji optymalnej $\hat{c}_i$ oraz maksymalnej wartości funkcji celu $\hat{Q}_i$ będzie zależne od parametru y_d oraz od czynnika zewnętrznego z_i, co można wyrazić zapisem $\hat{c}_i(y_d, z_i)$, $\hat{Q}_i = (y_d, z_i)$.

Przejdźmy teraz do zadania decyzyjnego jednostki nadrzędnej; jeśli będzie to przypadek zgodności celów, por. wzór (2.20), to zadaniem tym będzie

$$\underset{y_d}{\text{maksymalizować}}\ \Psi(\hat{Q}_1(y_d, z_1), \dots, \hat{Q}_N(y_d, z_N)) \tag{2.29}$$

przy czym zastanowić się trzeba nad warunkami ograniczającymi wybór y_d. Oprócz ewentualnego jawnego sformułowania (por. wzór (2.8)) $y_{di} \in Y_i$, $i = 1, \dots, N$, co będziemy zapisywali $y_d \in Y$, trzeba wziąć pod uwagę, że ograniczenia lokalne na (c_i, u_i), zapisane w zadaniu (2.28), także ograniczają osiągalne wartości wyjść y_i – przez właściwości podsystemu $F_i(c_i, u_i, z_i)$. W sumie składa się to na ograniczenie wyboru y_d przez jednostkę nadrzędną do pewnego zbioru

$$y_d \in Y_0(z) \tag{2.30}$$

który jest zbiorem wynikającym ze wszystkich ograniczeń lokalnych oraz z równań podsystemów i może być trudny do wyznaczenia. Granice tego zbioru mogą ponadto zależeć od czynnika zewnętrznego. Jeżeli koordynator ma działać efektywnie, to znaczy wydawać polecenia wykonalne, a zarazem w pełni wykorzystać możliwości podsystemów, potrzebuje dobrej znajomości ich równań i ograniczeń, a także powinien znać aktualne wartości czynników zewnętrznych $z_1, \dots, z_N$.

Zwróćmy się teraz ku przypadkowi niezgodności celów, czyli gdy funkcja celu koordynatora ma postać (por. (2.14))

$$J^0(c, u, y) = J^0(c_1, u_1, y_1, \ldots, c_N, u_N, y_N)$$

co oznacza w istocie rzeczy, że jednostka nadrzędna przypisuje pewne oceny do wejść, wyjść i sterowań lokalnych – inne niż te, które były im przypisywane przez lokalne funkcje celu. Zadanie decyzyjne jednostki nadrzędnej przybiera teraz postać

$$\underset{y_d \in Y_0(z)}{\text{maksymalizować}} \; J^0(\hat{c}_1(y_d, z_1), H_1 y_d, y_{d1}, \ldots, \hat{c}_N(y_d, z_N), H_N y_d, y_{dN}) \tag{2.31}$$

W sformułowaniu tym $\hat{c}_i(\cdot)$ są to decyzje lokalne, będące odpowiedzią na narzuconą wartość y_d. Ponadto przyjęto, iż wejścia i wyjścia podsystemów będą zrealizowane, to znaczy, że jednostki lokalne spełnią narzucone im wymagania.

Różnica w stosunku do sytuacji poprzedniej, to jest do sytuacji zgodności celów (por. (2.29)), jest taka, że jednostka nadrzędna operując wektorem y_d nie osiągnie rezultatu optymalnego, to znaczy tego, który mogłaby osiągnąć mając w ręku wszystkie decyzje, $i = 1, \ldots, N$. Inaczej mówiąc, pośrednictwo lokalnych jednostek decyzyjnych wiąże się ze stratą, jednostka nadrzędna „nie zachowuje pełni władzy" nad systemem sterowanym. W przypadku zgodności celów, (por. (2.29)), jednostka nadrzędna osiąga – mimo istnienia lokalnych jednostek decyzyjnych i prowadzenia przez nie własnych optymalizacji – wynik taki sam, jaki mogłaby osiągnąć, podejmując wszystkie decyzje lokalne bezpośrednio.

Jednostka nadrzędna, przy niezgodności celów, jest oczywiście w sytuacji, którą przedstawiamy jako grę z jednostkami dolnego poziomu. Pisząc zadanie (2.31) założyliśmy w istocie, że jednostka nadrzędna jest leaderem oraz zna spodziewane reakcje followerów w postaci funkcji $\hat{c}_i(y_d, z_i)$.

Zanotujmy jeszcze, że gdyby przy niezgodności celów przedmiotem zainteresowania jednostki nadrzędnej było wyrażenie (2.13)

$$J^*(y_1, \ldots, y_N)$$

to wówczas jednostka nadrzędna nie traciłaby niczego skutkiem istnienia jednostek lokalnych, gdyż sformułowanie to oznacza, iż decyzje lokalne c_i są dla koordynatora obojętne, byle spełnione były wymagania dotyczące wyjść. Jest to jednak przypadek nie mający większego znaczenia praktycznego – zainteresowanie koordynatora samymi tylko wyjściami, bez uwzględnienia sposobu i kosztów ich otrzymania, rzadko będzie racjonalne.

System powiązany; koordynacja przy użyciu cen

Przyjrzyjmy się obecnie sterowaniu systemem powiązanym, to jest o strukturze jak na rysunku 2.2, z użyciem cen jako instrumentu koordynacji działania lokalnych jednostek decyzyjnych. Jednostka nadrzędna otrzymuje prawo ustalania ceny λ na wejścia u, przy czym λ jest wektorem cen złożonym z wektorów częściowych λ_i, które są cenami na wejścia u_i do poszczególnych podsystemów. Ustalając cenę λ na wejścia u określamy zarazem cenę $\mu(\lambda)$ na wyjścia y, gdyż u oraz y są wzajemnie uzależnione przez macierz połączeń H. Konkretnie, cena $\mu_i(\lambda)$ na wyjścia i-tego podsystemu y_i wyrazi się wzorem

$$\mu_i(\lambda) = \sum_{j=1}^{N} H_{ji}^T \lambda_j \tag{2.32}$$

Przy omawianym obecnie sposobie koordynacji musimy nie tylko mieć możność narzucania ceny, ale także spowodować, że lokalna jednostka decyzyjna zmieni swoje zadanie decyzyjne, będzie mianowicie operować zmodyfikowaną funkcją celu

$$Q_{\mathrm{mod}\,i}(c_i, u_i, z_i, \lambda) = Q_i(c_i, u_i, z_i) - \lambda_i^T u_i + \mu_i^T(\lambda) y_i \tag{2.33}$$

Akceptacja formuły (2.33) przez lokalną jednostkę decyzyjną zamiast pierwotnej funkcji celu $Q_i(\cdot)$ nie jest rzeczą oczywistą; robimy tu w szczególności założenie, że $Q_i(\cdot)$, $i = 1, \ldots, N$ wyrażone są w jednostkach współmiernych, co przy koordynacji bezpośredniej wymagane nie było.

Zadanie lokalne z użyciem funkcji celu (2.33) mieć będzie postać

$$\left.\begin{aligned} &\underset{c_i, u_i}{\text{maksymalizować}} \; [Q_i(c_i, u_i, z_i) - \lambda_i^T u_i + \mu_i^T(\lambda) y_i] \\ &\text{przy ograniczeniach} \\ &(c_i, u_i) \in CU_i(z_i) \\ &F_i(c_i, u_i, z_i) \in Y_i \end{aligned}\right\} \tag{2.34}$$

przy czym swobodnymi zmiennymi decyzyjnymi są zarówno c_i jak u_i. Swobodę wyboru u_i rozumieć tu należy w ten sposób, że obliczana jest wartość wejścia pożądana z punktu widzenia decydenta lokalnego, która to wartość nie musi potem wystąpić w samym systemie (wywołując uzasadnione poczucie nieoptymalności z lokalnego punktu widzenia). Zarówno $CU_i(z_i)$ jak Y_i są ograniczeniami lokalnymi (por. wzory (2.7), (2.8)). Rozwiązanie zadania (2.34) będzie zależne tylko od wektorowego parametru λ i czynnika zewnętrznego z_i, to znaczy będziemy mieli $\hat{c}_i(\lambda, z_i)$, $\hat{u}_i(\lambda, z_i)$, $\hat{Q}_{\mathrm{mod}\,i}(\lambda, z_i)$, a także optymalne, z lokalnego punktu widzenia, wyjście $\hat{y}_i(\lambda, z_i)$.

Jeśli koordynator jest przede wszystkim zainteresowany neutralizacją konfliktu między jednostkami lokalnymi, to jego zadanie trzeba sformułować jako postulat *zrównoważenia interakcji*, czyli spowodowanie by optymalne dla poszczególnych jednostek lokalnych wejścia $\hat{u}_i(\cdot)$ oraz wyjścia $\hat{y}_i(\cdot)$ były ze sobą w zgodzie; inaczej mówiąc, by i-ty podsystem otrzymał na wejściu to co uważa za najlepsze ze swego punktu widzenia, a co pochodzi przez macierz H_i z wyjść uważanych za optymalne przez pozostałe podsystemy

$$H_i \hat{y}(\lambda, z) = \hat{u}_i(\lambda, z_i) \tag{2.35}$$

lub w sformułowaniu dla całości systemu

$$\hat{u}(\lambda, z) - H\hat{y}(\lambda, z) = 0 \tag{2.36}$$

Równość (2.36) jednostka nadrzędna ma osiągnąć przez dobór ceny λ. Odpowiednia wartość λ^r nazywa się *ceną równowagi*. Przy tej cenie każda z jednostek lokalnych – w działającym systemie – jest zadowolona ze swej sytuacji (otrzymuje wejście u_i równe temu, co uważa za optymalne), nie ma zatem motywacji do zmiany swych decyzji $\hat{c}_i$ oraz swego wyjścia $\hat{y}_i$. W tym sensie konflikt jest rzeczywiście zneutralizowany. Podkreślmy, że cena równowagi, czyli spełniająca równość (2.36), będzie zależeć od czynnika zewnętrznego z, zatem przy zmianie któregokolwiek z_i stan równowagi systemu może być zachwiany.

Wprowadzenie ceny równowagi neutralizuje konflikt jednostek lokalnych, co uznaliśmy za pierwsze zadanie koordynatora. Co się jednak dzieje z jego własną funkcją celu? Wiadomo, iż cena równowagi ma tę szczególną właściwość, że maksymalizuje sumę lokalnych funkcji celu. Jeśli zatem rzeczywiście $J = \sum\limits_i Q_i$, to koordynator osiąga nie tylko równowagę, ale także swoje własne optimum. Nie jest to przypadek ogólny; wypada raczej założyć, że jednostka nadrzędna ma jako funkcję celu wyrażenie (2.14) lub (2.15). Może być wówczas tak, że przy cenie innej niż cena równowagi wartość funkcji celu jednostki nadrzędnej jest lepsza, przy czym cena optymalna wyniknie z zadania decyzyjnego

$$\underset{\lambda}{\text{maksymalizować}} \; J(\hat{c}(\lambda, z), \hat{u}(\lambda, z), z) \tag{2.37}$$

Rozpatrzmy co się stanie w całości systemu gdy jednostka nadrzędna narzuci cenę wynikającą z (2.37), nie będącą ceną równowagi wynikającą z warunku (2.36). Weźmy najpierw pod uwagę sytuację jednostek dolnego poziomu. Uczyńmy założenie, że ich decyzje c_i są realizowane, to znaczy przykładane do systemu sterowanego, ściśle mówiąc do odpowiednich podsystemów; wówczas powstają określone wartości wejść i wyjść, zależne od wszystkich sterowań łącznie (por. (2.4))

$$u_i = H_i K(c, z), \quad y_i = F_i(c_i, H_i K(c, z), z_i) = K_i(c, z)$$

i w rezultacie zestaw zadań lokalnych ma postać

$$
\left.\begin{array}{l}
\underset{c_1}{\text{maksymalizować}}\ [Q_1(c_1, H_1K(c,z), z_1) - \lambda_1^T H_1 K(c,z) + \\
\qquad\qquad\qquad + \mu_1^T(\lambda) K_1(c,z)] \\
\ldots \\
\underset{c_N}{\text{maksymalizować}}\ [Q_N(c_N, H_NK(c,z), z_N) - \lambda_N^T H_N K(c,z) + \\
\qquad\qquad\qquad + \mu_N^T(\lambda) K_N(c,z)]
\end{array}\right\} \tag{2.38}
$$

Postać (2.38) wskazuje, że jednostki lokalne znajdują się w sytuacji gry pomiędzy sobą, gdyż wartość funkcji celu każdej z nich zależy nie tylko od decyzji własnej czyli od c_i, lecz także od decyzji pozostałych. Załóżmy, że dla każdej wartości parametru λ, czyli dla każdej narzuconej wartości cen $\lambda_i, \mu_i(\lambda)$, gra (2.38) osiąga stabilny punkt równowagi. Przy braku umów (koalicji) między graczami byłby to tzw. punkt równowagi Nasha, tj. taki, że przy danym λ żaden z graczy nie zyskuje jeśli tylko on jeden zmieni swą decyzję.

Nie zmienia to oczywiście faktu, że jednostka lokalna nie jest w pełni zadowolona, dostrzega ona bowiem różnicę między u_i, wynikającym z punktu równowagi gry (2.38) a wartością optymalną, wynikającą z zadania lokalnego (2.34) przy tej samej cenie λ. Położenie punktu równowagi, to jest określające go wartości u^r, y^r oraz c^r zależą od parametru λ; jest tu teraz miejsce na pomyślenie o celu jednostki nadrzędnej, którym zgodnie z (2.15) będzie

$$
\underset{\lambda}{\text{maksymalizować}}\ J(c^r(\lambda,z), HK(c^r(\lambda,z),z)) \tag{2.39}
$$

Postać (2.39) wskazuje wyraźnie, że jednostka nadrzędna znajduje się w sytuacji gry z jednostkami lokalnymi, gdyż osiągana wartość $J(\cdot)$ zależeć będzie zarówno od wartości λ, jak od decyzji c^r wszystkich jednostek dolnego poziomu. Można to rozważać jako grę hierarchiczną jeśli uczynimy założenie, że jednostka nadrzędna ma prawo pierwszego ruchu, to znaczy, iż ona ustala cenę λ, a jednostki dolnego poziomu mają na to tylko odpowiedzieć. Gdyby jednostka nadrzędna znała wszystkie funkcje odpowiedzi jednostek lokalnych $c_i^r(\lambda, z)$ oraz równanie połączonego systemu $u = HK(c,z)$ (por. (2.4)), mogłaby ona obliczyć najlepszą dla siebie wartość ceny $\lambda = \hat{\lambda}$ i taką właśnie wartość narzucać systemowi. Jeżeli jednostka nadrzędna nie może – z braku dostatecznej informacji – obliczyć $\hat{\lambda}$, pozostają dwa warianty postępowania: ustalić λ według najlepszych swoich oszacowań, możliwie bliskie nieznanemu $\hat{\lambda}$ oraz pogodzić się z powstającym wtedy rezultatem działań dolnego poziomu albo też dobrać λ iteracyjnie, czyli obserwować wyniki otrzymywane przy kolejnych wartościach λ w postaci reakcji systemu $c^r(\lambda, z)$, $u^r(\lambda, z)$, dążąc do maksymalizacji swojej funkcji celu. Nietrudno zauważyć, że w tym przypadku jednostka nadrzędna znajduje się w sytuacji ($N + 1$)-osobowej gry z jednostkami lokalnymi, w której nie ma już gracza z prawem pierwszego ruchu. Gra ta może mieć punkt równowagi, w którym jednostka nadrzędna nie zyskuje na zmianie ceny, a żadna z jednostek dolnego poziomu nie zyskuje na zmianie swojej decyzji c_i^r. W teorii gier wykazuje się, że rezultat otrzymywany przez leadera w grze hierarchicznej jest lepszy, a co najmniej równy temu, który osiągnie on w punkcie równowagi Nasha, to jest w grze partnerów równouprawnionych. Dla jednostki nadrzędnej jest zatem lepiej być leaderem, oczywiście zakładając, że może ona obliczyć poprawne $\hat{\lambda}$.

W sformułowaniach (2.37), (2.38), (2.39) zachowaliśmy zmienną z, oznaczającą wpływ otoczenia systemu ażeby podkreślić, że nie wystarczy jeden raz osiągnąć punkt równowagi gry (2.38), czy też gry, w której bierze udział jednostka nadrzędna. Zmiana czynników zewnętrznych może spowodować, że umotywowana będzie zmiana decyzji jednostek lokalnych, w ślad za tym zaistnieje potrzeba zmiany decyzji jednostki nadrzędnej.

Mówiliśmy wcześniej o mechanizmie ustalania przez jednostkę nadrzędną ceny optymalnej w sytuacji, gdy ceną tą nie jest cena równowagi wynikająca ze spełnienia równania (2.36). Przyjrzyjmy się teraz jak można znaleźć cenę λ^r spełniającą (2.36); w terminologii ekonomicznej byłaby to cena równoważąca podaż z popytem.

Pierwszym sposobem jest wyobrażenie sobie, że jednostka nadrzędna zna funkcje $\hat{u}(\lambda, z)$ oraz $\hat{y}(\lambda, z)$ czyli „funkcję popytu" i „funkcję podaży" w całości systemu, może zatem na tej podstawie obliczyć cenę równowagi λ^r. Przekazanie tej ceny jednostkom dolnego poziomu wywoła obliczenie przez nie decyzji $\hat{c}_i$, których wykonanie wywoła w całości powiązanego systemu wartości wejść i wyjść zgodne zarówno z oczekiwaniami jednostki nadrzędnej, jak z przekonaniami jednostek lokalnych co do tego, jakie wejście u_i jest dla każdej z nich najlepsze.

Założenie tak dobrej znajomości systemu przez jednostkę nadrzędną (włączając w to znajomość celów jednostek lokalnych oraz aktualnych wartości wpływów zewnętrznych) jest zapewne dalekie od realizmu. Należy więc pomyśleć o skorzystaniu z informacji posiadanej przez jednostki lokalne lub z obserwacji systemu rzeczywistego, lub z obu tych źródeł.

Najbardziej bezpośrednim podejściem byłaby próba skorzystania z wartości różnicy popytu i podaży, to jest różnicy między $\hat{u}(\lambda, z)$ oraz $H\hat{y}(\lambda, z)$, por. (2.36). Rzecz w tym, że różnicy tej nie można zmierzyć w systemie powiązanym jak na rys. 2.2, gdzie wejścia są na stałe dołączone do wyjść. Można natomiast, jak to już robiliśmy przy zastosowaniu cen do koordynacji zużycia zasobów, wprowadzić fazę *przygotowania decyzji* polegającą na negocjowaniu ceny λ między jednostką nadrzędną a jednostkami lokalnymi. Jednostka nadrzędna zmienia proponowaną cenę λ tak długo, jak długo deklarowane przez jednostki lokalne wartości optymalnych dla nich wyjść $\hat{y}_i(\lambda, z)$ oraz pożądanych przy tym wejść $\hat{u}_i(\lambda, z)$ nie będą takie, że następuje w całości systemu równowaga (2.36). Odpowiadająca temu cena λ^r zostaje podtrzymana przez jednostkę nadrzędną czyli *wprowadzona w życie*, to samo dzieje się z decyzjami lokalnymi $\hat{c}_i(\lambda^r, z)$. Założeniem podstawowym jest tutaj to, że w systemie sterowanym ustalą się dokładnie takie same wartości wejść i wyjść, jakie były obliczone (i deklarowane wobec jednostki nadrzędnej) w fazie przygotowania decyzji. Odpowiada to z kolei założeniu, że jednostki lokalne dysponowały bezbłędnymi modelami swoich podsystemów. Jest to oczywiście mało realne; w teorii sterowania*) mówi się dość dużo o tym, jak można postępować przy istnieniu różnic między modelami a rzeczywistością, wykorzystując przy tym obserwacje systemu w trakcie działania.

W podobny sposób, jak to zrobiliśmy w odniesieniu do znajdowania ceny równowagi, można także wyróżnić fazę przygotowania decyzji w sytuacji poszukiwania ceny optymalnej dla jednostki nadrzędnej według sformułowania (2.37). Proces znajdowania punktu równowagi gry (2.38), a także określenie ceny $\hat{\lambda}$, czyli decyzji jednostki nadrzędnej, może się odbyć w fazie przygotowania, czyli przez negocjacje jednostek decyzyjnych operujących modelami. Dopiero rezultat ostateczny byłby wprowadzony w życie; oszczędza to oczywiście systemowi sterowanemu perturbacji związanych z dochodzeniem do punktu równowagi, wprowadza natomiast nową kwestię, to jest pytanie o skutki różnic między modelem systemu a rzeczywistością.

Przy stosowaniu metody cen do koordynacji działania jednostek lokalnych jest rzeczą potrzebną, by funkcje podaży $\hat{y}_i(\lambda)$ oraz popytu $\hat{u}_i(\lambda)$ były jednoznaczne, to jest by określone λ wywoływało jednoznacznie określony wynik lokalny $\hat{y}_i$ bądź $\hat{u}_i$. Na ogół przyjmuje się, że w sytuacjach praktycznych tak będzie, rozpatrując przy tym najchętniej powiązanie między dwoma podsystemami. W złożonym systemie trzeba jednoznaczność funkcji podaży i funkcji popytu oraz jej skutki rozumieć nieco szerzej.

Rozpatrzmy system jak na rys. 2.5, w którym podsystemy 1 i 2 wytwarzają taki sam produkt, w ilościach y_1 i y_2, a podsystemy 3 i 4 produkt ten zużywają (w ilościach u_3 i u_4).

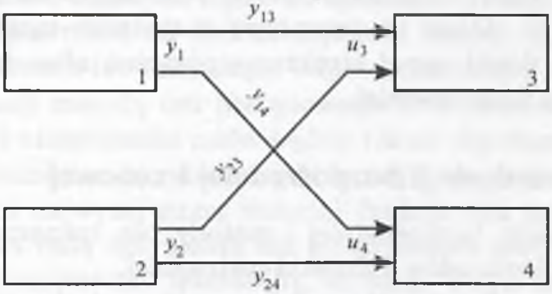
Znane są funkcje odpowiedzi podsystemów produkcyjnych na cenę λ („funkcje podaży"), $\hat{y}_1(\lambda)$ oraz $\hat{y}_2(\lambda)$, a także funkcje odpowiedzi podsystemów zużywających produkt („funkcje popytu") $\hat{u}_3(\lambda)$, $\hat{u}_4(\lambda)$.

Ceną równowagi będzie w tym systemie wartość λ^r, określona przez równanie

$$\hat{y}_1(\lambda) + \hat{y}_2(\lambda) = \hat{u}_3(\lambda) + \hat{u}_4(\lambda)$$

*) Patrz np. W. Findeisen, F.N. Bailey, M. Brdyś, K. Malinowski, P. Tatjewski, A. Woźniak (1980). Control and Coordination in Hierarchical Systems. Wiley, London, rozdz. 3.

a przez cenę równowagi wyznaczone będą konkretne wartości produkcji y_1^r, y_2^r oraz zużycia u_3^r, u_4^r.



Rys. 2.5. Schemat systemu do przykładu

Zwróćmy teraz uwagę, że przy takiej strukturze połączeń jak na rys. 2.5 (każdy z każdym) cena równowagi λ^r określa wartości produkcji i zużycia, lecz nie określa jednoznacznie wartości przepływów y_{13}, y_{14}, y_{23}, y_{24}. Mamy wprawdzie cztery równania

$$\begin{aligned} y_{13} + y_{14} &= y_1^r \\ y_{23} + y_{24} &= y_2^r \\ y_{13} + y_{23} &= u_3^r \\ y_{14} + y_{24} &= u_4^r \end{aligned}$$

ale układ ten nie ma jednoznacznego rozwiązania. Model systemu zachowuje się tak, jak gdyby sieć połączeń między podsystemami była czymś w rodzaju wspólnego kolektora, do którego wprowadza się y_1^r oraz y_2^r i pobiera u_3^r oraz u_4^r, nie wiedząc np. jaka część produkcji podsystemu 1 zużywana jest przez podsystem 3, a jaka przez podsystem 4.

Niejednoznaczność określenia przepływów w systemie jak na rys. 2.5 można usunąć; jeżeli ktokolwiek (jednostka nadrzędna albo jeden z podsystemów) podejmie decyzję odnośnie jednego z przepływów y_{ij}, to pozostałe trzy zostaną także wyznaczone. Decyzja, o której mowa musi przy tym być w zgodzie z potrzebami, np. dla przepływu y_{24}

$$0 \leqslant y_{24} \leqslant u_4^r$$

oraz z możliwościami

$$0 \leqslant y_{24} \leqslant y_2^r$$

Dodajmy jeszcze, że jeśli w strukturze systemu na rys. 2.5 zabraknie jednego z czterech połączeń, to jest to równoważne decyzji, że dany przepływ jest równy zeru. Przepływy w pozostałych trzech połączeniach określone będą jednoznacznie.

Powróćmy na początek rozpatrywanego przykładu: dla każdego podsystemu istniała jednoznaczna funkcja odpowiedzi na cenę, np. $\hat{y}_1(\lambda)$, ale nie istniały tego typu charakterystyki dla poszczególnych przepływów, np. $\hat{y}_{13}(\lambda)$ albo $\hat{y}_{14}(\lambda)$. Mimo to, przepływy w systemie mogły być określone jednoznacznie – dzięki samej strukturze połączeń albo dzięki interwencji typu koordynacji bezpośredniej.

Porównanie koordynacji bezpośredniej i cenowej

Porównanie metody bezpośredniej i metody cen rozpoczniemy od oceny możliwości spełnienia celów jednostki nadrzędnej.

Rozpatrzmy przede wszystkim skuteczność neutralizacji konfliktu między lokalnymi jednostkami decyzyjnymi, który powstaje na skutek połączeń pomiędzy podsystemami bądź też w sytuacji korzystania ze wspólnych zasobów.

W zasadzie obie metody konflikt ten są w stanie zneutralizować. O ile jednak metoda bezpośrednia, która zamraża połączenia bądź ustala sztywny rozdział zasobów, nie jest wrażliwa na nieprzewidziane zmiany celów jednostek lokalnych, to neutralizacja konfliktu metodą cen funkcjonuje poprawnie tylko wtedy, gdy narzucona cena jest istotnie ceną równowagi. Jeśli tak nie jest, lokalne jednostki decyzyjne otrzymują wartości wejść, które nie są dla nich optymalne; wchodzą one zatem we wzajemną grę, która – jeśli jest stabilna – mieć będzie punkt równowagi, nie będący jednak tym samym co punkt odpowiadający cenie równowagi. W szczególności suma osiągniętych wartości lokalnych funkcji celu będzie w zasadzie mniejsza od możliwości maksymalnych systemu.

Drugim dążeniem koordynatora, po neutralizacji konfliktu, jest maksymalizacja jego własnej funkcji celu. Mamy tu do rozróżnienia przypadek zgodności celów i przypadek ich niezgodności.

Przy koordynacji przez narzucanie wartości wyjść bądź wejść postępowanie koordynatora jest w obu przypadkach podobne: wybiera on $y_{di}, i = 1, \ldots, N$ tak, by maksymalizować swoją własną funkcję celu, opierając się na znajomości (przewidywaniu) reakcji jednostek dolnego poziomu, czyli ich decyzji $\hat{c}_i(y_d)$. Różnica między przypadkiem zgodności i niezgodności celów jest tylko taka, że jeśli jednostki lokalne mają funkcje celu nie będące w relacji zgodności z funkcją celu koordynatora, to istnieją decyzje c_i, które byłyby dla koordynatora lepsze niż podejmowane lokalnie $\hat{c}_i(y_d)$. Warto tu spostrzec, że jeśli koordynator nie zna dostatecznie dobrze zadań decyzyjnych dolnego poziomu, lecz opiera się na informacji wynikowej, czyli na zależnościach $\hat{c}_i(y_d)$, to może on nie wiedzieć, że jego funkcja celu znajduje się w niezgodności z celami lokalnymi, czyli że nie zachodzi związek

$$J^0(c, u, y) = \Psi(Q_1^0(c_1, u_1, y_1), \ldots, Q_N^0(c_N, u_N, y_N))$$

W rezultacie koordynator nie zdaje sobie sprawy, że system jako całość mógłby osiągać lepsze wartości $J^0(\cdot)$. Trzeba by jednak w tym celu wpłynąć na zmianę celów lokalnych dążąc do spełnienia relacji zgodności bądź też wpłynąć bezpośrednio na niektóre spośród decyzji c_i, $i = 1, \ldots, N$, podejmując je samemu lub nakładając odpowiednie więzy i ograniczenia.

Przy koordynacji metodą cen postępowanie koordynatora w przypadku zgodności celów i niezgodności celów będzie różne. Zgodność celów ogranicza się przy metodzie cen do sytuacji, gdy funkcja celu jednostki nadrzędnej jest sumą (lub co najwyżej sumą ważoną) funkcji celu jednostek dolnego poziomu; wówczas ceną optymalną dla koordynatora jest cena równowagi (powodująca zrównoważenie interakcji), ta sama która neutralizuje konflikt jednostek lokalnych. Przez optymalność dla koordynatora należy tu rozumieć, że przy cenie równowagi decyzje lokalne są dokładnie takie, jakie podjąłby koordynator bezpośrednio ze swego punktu widzenia. Cena równowagi może być, jak wiemy, znaleziona przez zrównanie podaży i popytu, czyli przez spełnienie pewnej równości, patrz (2.36). Jeżeli funkcja celu jednostki nadrzędnej nie jest sumą lokalnych funkcji celu, cena optymalna dla koordynatora musi być znajdowana na drodze maksymalizacji funkcji celu, patrz (2.39), co jest postępowaniem odmiennym od szukania równowagi. Wynik osiągnięty przez jednostkę nadrzędną, to jest wartość jej funkcji celu, zależy od decyzji lokalnych $\hat{c}_i(\lambda)$. Wynik ten będzie w zasadzie gorszy od tego, co jednostka nadrzędna mogłaby osiągnąć, określając decyzje lokalne bezpośrednio, czyli podporządkowując je wyłącznie swojej własnej funkcji celu – mówimy, że jednostka nadrzędna korzystająca z pośrednictwa jednostek dolnego poziomu „nie zachowała pełni władzy nad systemem". Podobnie było, w przypadku niezgodności celów, przy koordynacji metodą bezpośrednią. Strata na wartości funkcji celu jednostki nadrzędnej w przypadku niezgodności celów występuje zatem przy obu metodach koordynacji; nie można jednak wskazać w sposób ogólny czy strata ta będzie większa przy metodzie bezpośredniej, czy też przy metodzie cen. Nie ma także podstaw do przypuszczenia, że będą one takie same.

Prowadząc porównanie obu metod koordynacji trzeba zwrócić uwagę na zakres informacji potrzebnej na poziomie koordynatora oraz na ewentualne skutki jej braku. Przy metodzie bezpośredniej punktem krytycznym dla koordynatora jest znajomość zbioru decyzji dopuszczalnych Y_0, patrz wzór (2.30). Polecenia koordynatora w postaci zadawanych wartości wyjść y_{di} muszą być wykonalne, w przeciwnym razie trudno jest przewidzieć w jakim stanie znajdzie się cały system, to znaczy jakie ustalą się w nim wartości wszystkich wejść i wyjść, jako wynik gry między lokalnymi jednostkami decyzyjnymi (w nieobecności „zamrożenia" wejść i wyjść sytuacja konfliktowa nie jest zneutralizowana). Informacja o zbiorze dopuszczalnym Y_0, czyli o tym, jakie wyjścia są wykonalne, jest trudna do uzyskania; można ją czerpać od lokalnych jednostek decyzyjnych, wchodzi tu jednak w grę ich interes. Dla jednostki lokalnej opłacalne jest podanie informacji fałszywej, mianowicie takiej, by wynikało z niej narzucenie wartości wyjścia y_i, optymalnego dla jednostki lokalnej, a nie dla koordynatora. Mówimy o tym bliżej w podrozdziale 2.6.

Zakres informacji potrzebnej jednostce nadrzędnej przy koordynacji metodą cen jest inny: nie musi ona znać ograniczeń na osiągalne wyjścia, gdyż przy posługiwaniu się cenami wartości wyjść są ustalane przez jednostki lokalne, które oczywiście swoje ograniczenia znają i uwzględnią. Jeśli koordynator dąży do ustalenia cen równowagi, jedynym jego problemem jest znajomość funkcji odpowiedzi każdego podsystemu $\hat{u}_i(\lambda)$ oraz $\hat{y}_i(\lambda)$. Potrzebne do tego informacje mogą pochodzić bezpośrednio od jednostek lokalnych, które w tym przypadku nie mają wyraźnych bodźców do przekazywania danych niezgodnych z prawdą. Tym niemniej trzeba zwrócić uwagę, że jeśli koordynator oprze swoją decyzję – czyli ustalenie ceny – na danych nie odpowiadających rzeczywistości (na przykład nie dostosowanych do wpływu czynników zewnętrznych z_i), to w systemie ustalą się wartości powiązań inne niż te, które przewidywał koordynator. Będą to jednak nadal wartości wynikające z decyzji lokalnych opieranych na maksymalizacji swoich funkcji celu.

Jeden jeszcze aspekt porównania koordynacji bezpośredniej i cenowej, związany z panowaniem nad całością systemu wart jest zauważenia. Przy koordynacji bezpośredniej jednostka nadrzędna wyznacza wartości wyjść, a zatem i wejść każdego z podsystemów, panuje więc nad każdym „przepływem" z jednego podsystemu do drugiego, czyli nad wartością każdego powiązania. Wielką troską koordynatora – jak już mówiliśmy – staje się przy tym narzucanie takich wartości wyjść i wejść, które okażą się wykonalne, gdyż inaczej jego panowanie nad systemem staje się problematyczne, bowiem w rzeczywistości pojawią się wartości wyjść i wejść przez koordynatora nie zamierzone i nie przewidywane.

Przy koordynacji za pomocą cen panowanie nad wyjściami i wejściami, czyli nad przepływami z podsystemu do podsystemu także ma miejsce, jeśli tylko istnieją jednoznaczne funkcje odpowiedzi podsystemów; istnienie tych funkcji zależy od właściwości zadania decyzyjnego jednostki lokalnej, ogólnie mówiąc – od podsystemu. Jeśli jednoznaczne $\hat{y}_i(\lambda)$, $\hat{u}_i(\lambda)$ istnieją, narzucenie ceny λ wywoła określone wartości y_i, u_i – w tym właśnie sensie jednostka nadrzędna panuje nad wartościami powiązań, może je przewidzieć i wywołać. Może ona na przykład liczyć na to, że swobodne decyzje o poborach zasobów będą takie, iż nie nastąpi przekroczenie ograniczenia, por. dyskusję tego problemu na początku rozdziału 2.4. Trzeba jednak pamiętać, że wartości przewidywanych wyjść podsystemów $\hat{y}_i(\lambda)$ przy metodzie cen nie są z punktu widzenia koordynatora tak pewne, jak w przypadku koordynacji bezpośredniej. Wystarczy zmiana lokalnej funkcji celu lub innego składnika zadania decyzyjnego, by zmieniły się funkcje $\hat{y}_i(\lambda)$, $\hat{u}_i(\lambda)$, a zatem i stan powiązań w systemie przy danym λ. Podkreślmy na koniec jeszcze raz, że przy koordynacji opieranej na cenach równowagi koordynator jest w zasadzie tylko świadkiem zjawisk w systemie; jego zadaniem jest wykrycie ceny równowagi, a następnie narzucenie jej lokalnym jednostkom decyzyjnym, które od tego momentu działają bez konfliktu pomiędzy sobą, a powstające w systemie przepływy wynikają z równowagi podaży i popytu. Przepływy te mogą jednak z różnych względów nie być zadowalające dla jednostki nadrzędnej. Może ona na przykład stwierdzić, iż w systemie z rys. 2.5 wyjście $\hat{y}_1(\lambda^r)$, czyli produkcja podsystemu nr 1 jest przy cenie równowagi za mała, że zdominowana ona została przez bardziej sprawny podsystem nr 2, dający przy cenie λ^r wartość $\hat{y}_2(\lambda^r)$. Jeśli „zbyt mała" wartość $\hat{y}_1(\lambda^r)$ jest niepożądana, mimo iż rachunek ekonomiczny chwili bieżącej mówi inaczej, to jednostka nadrzędna musi pomyśleć o interwencji. Nie musi to być odejście od ogólnej zasady cen równowagi. W systemie z rys. 2.5 istnieć może na przykład możliwość wywarcia wpływu na funkcję podaży $\hat{y}_1 = f_1(\lambda)$ przez dopłatę proporcjonalną do wyjścia $\hat{y}_1$ lub przez dotację w innej formie. Można także uprzywilejować podsystem nr 1 przez wpływ na funkcję podaży podsystemu nr 2, aby zmniejszyć opłacalność, przez odpowiednie opodatkowanie jego wielkości wyjściowej $\hat{y}_2$.

W istocie rzeczy, dopiero przy zastosowaniu tego rodzaju środków co wymienione, jednostka nadrzędna staje się koordynatorem działań podsystemów w zakresie podobnym do tego, który miał miejsce przy koordynacji metodą bezpośrednią. Zachowanie cen jako głównego instrumentu oddziaływania ma tu tę zaletę, że wszystkie lokalne jednostki decyzyjne opierają swoje rozstrzygnięcia na rachunku osiąganych korzyści, bez wymagań dodatkowych dotyczących wartości wyjść czy ograniczających wejścia.

2.5. Droga od struktury systemu do struktury sterowania

Spoglądając syntetycznie na zadanie sterowania systemem złożonym oraz na utworzenie struktury zadanie to realizującej zauważymy, że na początku postawić trzeba dwa pytania podstawowe:

1) jak jest zbudowany system, który ma podlegać sterowaniu?
2) jakie są cele lokalnych jednostek decyzyjnych?

Pierwsze z tych pytań dotyczy struktury systemu, czyli jak dalece poszczególne podsystemy zależą „fizycznie" jeden od drugiego, patrz rys. 2.2, 2.3 oraz 2.4. Drugie pytanie dotyczy indywidualnych celów jednostek decyzyjnych. Jak opisano w rozdziale 2.2, cele te mogą być:

a) określone tylko na rezultatach własnych, wzór (2.16),
b) określone na rezultatach własnych i obcych, czyli zachodzące na siebie, wzór (2.17),
c) identyczne, dając okazję do działania jako zespół, wzór (2.18).

Struktura systemu oraz cele lokalne składają się razem na zachowanie się systemu bez koordynacji, czyli na działanie pod wpływem wyłącznie lokalnych jednostek decyzyjnych. W tym momencie trzeba przywołać ewentualne interesy całościowe: jeśli system sterowany przez jednostki lokalne jest niestabilny albo jeśli jest stabilny, lecz osiągany w systemie punkt równowagi nie jest zadowalający z punktu widzenia celów globalnych, potrzebna jest interwencja. W istocie pojawiają się zatem kolejne dwa pytania:

3) jak zachowuje się system bez koordynacji?
4) jakie są interesy całościowe?

Stosownie do odpowiedzi na te dwa pytania możemy dojść do wniosku, że potrzebne jest wprowadzenie nadrzędnej jednostki decyzyjnej z zadaniem koordynacji działania jednostek lokalnych. Struktura sterowania, czyli struktura podejmowania decyzji staje się hierarchiczna, z jednostkami lokalnymi na dolnym poziomie – por. rozdz. 1.4, a także rys. 2.1.

W ślad za postanowieniem o wprowadzeniu jednostki nadrzędnej musimy określić zasady jej funkcjonowania, na co składają się przede wszystkim kwestie następujące:

5) jaki ma być bieżący cel działania jednostki nadrzędnej?
6) jakie instrumenty (bezpośrednie czy cenowe) mają być użyte do koordynacji?

Liczba wariantów powstających przy różnych odpowiedziach na każde z sześciu wymienionych pytań jest bardzo duża; możemy się spodziewać najróżniejszych kombinacji struktury, celów lokalnych, celów globalnych, metod koordynacji. Istniejąca teoria sterowania bądź podejmowania decyzji w strukturach o wielu jednostkach decyzyjnych może tu być pomocna, trzeba sobie jednak zdawać sprawę, że dla przypadków o stopniu złożoności odpowiadającym praktyce może ona nie być w stanie podać gotowych recept i rozwiązań. Dotyczy to w szczególności reguł podejmowania decyzji gdy pojawia się zjawisko gry między lokalnymi jednostkami decyzyjnymi, a także gry między nimi a jednostką nadrzędną. Tym niemniej istniejąca teoria określa podstawowe prawidłowości przebiegu procesów decyzyjnych i procesów w systemie sterowanym, pomaga przewidzieć jakie zjawiska mogą się pojawić, w pewnym stopniu także jak można im przeciwdziałać.

Projekt konkretnego układu sterowania dla złożonego systemu może znaleźć w teorii swą linię przewodnią, podstawy koncepcyjne, wskazówki wyboru sposobu działania. Szczegółowa struktura, a także algorytmy (reguły) działania jednostek decyzyjnych pozostać muszą dziełem sztuki inżynierskiej, wspólnym rezultatem teorii, doświadczenia, rozsądku i eksperymentu. W dziedzinie sterowania złożonymi systemami będzie to przede wszystkim eksperyment symulacyjny – doświadczenia na „żywych" systemach mogą bowiem być długotrwałe i pod wieloma względami zbyt kosztowne. Powracamy do tego zagadnienia w rozdziale 5.7.

2.6. Podział informacji i procesy decyzyjne w strukturze hierarchicznej

W rozdziale 1 była już mowa o tym, że istotą układu sterowania jakimkolwiek obiektem lub systemem jest przekształcanie obserwacji w oddziaływanie sterujące, przy czym obserwacje dotyczyć mogą zarówno zachowania się systemu sterowanego, jak też zjawisk w jego otoczeniu. W hierarchicznej strukturze sterowania systemem przekształcanie obserwacji w oddziaływanie sterujące jest podzielone między różne jednostki decyzyjne. O każdej z nich można powiedzieć, że ma ona za zadanie przekształcać otrzymywane informacje w podejmowane przez siebie decyzje; u jednostek położonych na najniższym poziomie będą to oddziaływania sterujące w ścisłym znaczeniu tego określenia.

Dotąd zajmowaliśmy się strukturą decyzyjną, czyli kwestią jakie decyzje należeć mają do kompetencji poszczególnych jednostek decyzyjnych. Obecnie zajmiemy się dalszymi aspektami problemu, a mianowicie – na jakiej informacji opierać się będzie dana jednostka decyzyjna oraz jak będzie zorganizowany proces podejmowania decyzji przez całą strukturę, czyli wynikowe przekształcanie dostępnych obserwacji w oddziaływanie na sterowany system. W rozdziale 5.6 omawiamy pokrótce odpowiednie mechanizmy decyzyjne, czyli zajmujemy się bliżej organizacją i przebiegiem procesu przetwarzania informacji wewnątrz każdej jednostki decyzyjnej.

W hierarchicznych strukturach sterowania sprawą bardzo ważną, często decydującą o wyborze sposobu działania, jest zakres informacji o systemie sterowanym i jego otoczeniu oraz o zadaniach sterowania, która może być dostępna poszczególnym jednostkom decyzyjnym. W szczególności, niemożność lub kosztowność posiadania informacji szczegółowej o podsystemach może prowadzić do wyboru metody cen jako sposobu koordynacji działania systemu, gdyż jak to opisano w rozdziale 2.4, jednostka nadrzędna nie potrzebuje wówczas znać lokalnych ograniczeń oraz lokalnej funkcji celu.

Uzyskiwanie informacji przez fazę przygotowania decyzji

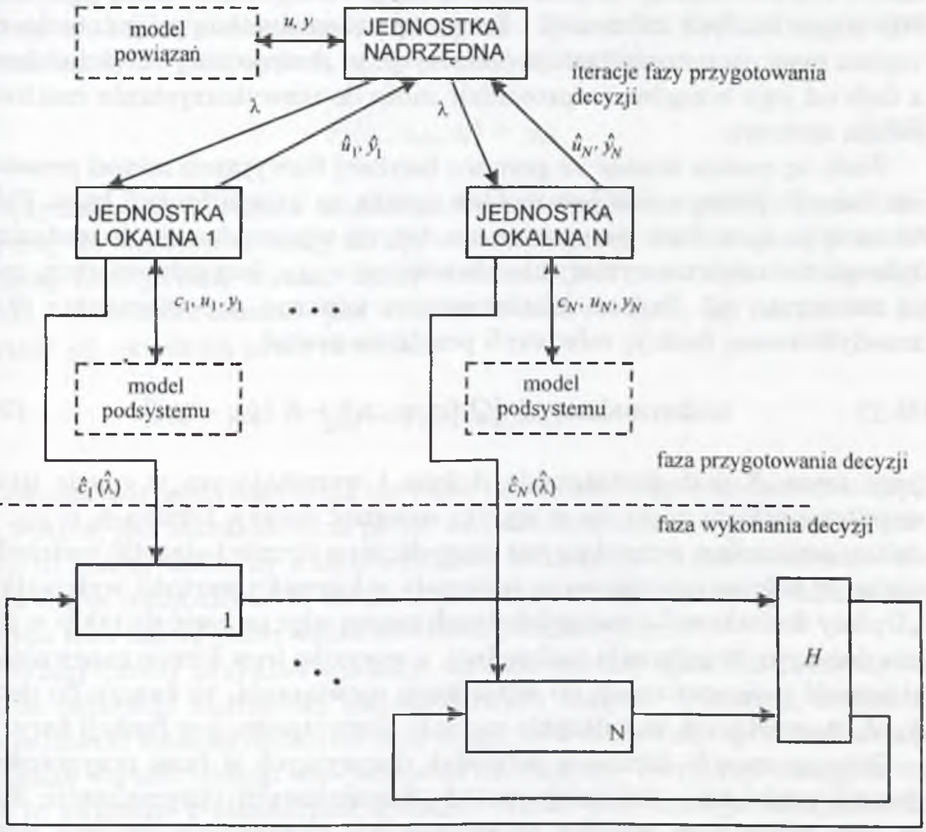
Ważnym sposobem ominięcia trudności związanych z brakiem informacji o podsystemach na szczeblu jednostki nadrzędnej jest podzielenie procesu decyzyjnego na dwie fazy, a mianowicie na fazę *przygotowania* decyzji i fazę jej *wykonania*. Faza przygotowania decyzji może być iteracyjna, to znaczy może polegać na kolejnych propozycjach jednostki nadrzędnej i wysłuchiwaniu odpowiedzi jednostek lokalnych. Na przykład, w kontekście metody cen wyglądałoby to jak następuje.

Jednostka nadrzędna proponuje cenę dotyczącą danej wielkości stanowiącej powiązanie między podsystemami, na przykład cenę jednostkową produktu stanowiącego wyjście podsystemu nr 1 i wejście podsystemu nr 2. Lokalna jednostka decyzyjna nr 1 wykonuje swoje obliczenia (niekoniecznie optymalizacyjne) i informuje o zamierzonej, przy danej cenie, wielkości produkcji, podaje zatem swoją zamierzoną decyzję. Analogicznie, w wyniku podobnego działania, lokalna jednostka decyzyjna nr 2 informuje o zamierzonym zużyciu. Istnienie różnicy między zadeklarowaną produkcją i zadeklarowanym zużyciem świadczy o tym, że zaproponowana cena nie jest ceną równowagi, a ponadto znak różnicy jest wskazówką czy cena ta jest większa czy mniejsza od ceny ostatnio zaproponowanej. Pozwala to jednostce nadrzędnej skorygować cenę we właściwym kierunku, co jest podstawą iteracyjnego procesu dojścia do określenia poszukiwanej decyzji, czyli ceny równowagi. Zwróćmy uwagę, że jednostka nadrzędna nie ma w tym procesie informacji o funkcjach odpowiedzi podsystemów, poznaje tylko ich niektóre wartości – te mianowicie, które odpowiadają kolejno proponowanym cenom.

Omówiony proces dochodzenia do określenia ceny równowagi jest właśnie fazą przygotowania decyzji, zaś decyzją staje się określona przez ten proces cena równowagi. Po jej ustaleniu może nastąpić faza wykonania decyzji: cena równowagi zostaje „wprowadzona w życie" to znaczy podsystemy zostają poinformowane, że będzie ona odtąd obowiązywać w rozliczeniach rzeczywistych. Od tej chwili podsystem nr 1 może kształtować swoje wyjście w sposób swobodny. Podkreślaliśmy w rozdz. 2.4, iż może to oznaczać, że wyjście podsystemu będzie inne niż deklarowane w czasie przygotowania decyzji, gdyż mogą zmieniać się czynniki wpływające na kształt funkcji odpowiedzi, a podsystem kieruje się tylko cenami oraz swoją optymalizacją wewnętrzną.

W fazie przygotowania decyzji zarówno jednostka nadrzędna, jak jednostki lokalne nie mają możliwości konfrontacji swoich przewidywań z rzeczywistością, czyli opierają się na modelach (nie jest rzeczą istotną, czy są to modele sformalizowane czy intuicyjne). W szczególności, jednostka nadrzędna operuje modelem powiązań między podsystemami, a jednostki lokalne – modelami swoich podsystemów. Schematycznie przedstawiono to na rys. 2.6.

Proces podejmowania decyzji w układzie hierarchicznym operującym metodą koordynacji bezpośredniej, natrafić może na jeszcze większe trudności natury informacyjnej niż te, które mogły wystąpić przy metodzie cen. Jeżeli bowiem jednostka nadrzędna chce określić optymalną dla siebie wartość wyjść podsystemów y_d, zgodnie z zadaniem zapisanym jako (2.29) bądź (2.31), to musi ona zarówno umieć przewidzieć decyzje jednostek lokal-



Rys. 2.6. Posługiwanie się modelami w fazie przygotowania decyzji (na przykładzie metody cen)

nych $\hat{c}_i$ bądź ich skutki $\hat{Q}_i$, jak też – co trudniejsze – musi wiedzieć, jakie wartości y_d są realizowalne, to znaczy jakie są w tym zakresie reperkusje wszystkich lokalnych ograniczeń. Posunięte tak daleko scentralizowanie informacji jest zapewne mało realne w praktycznych zastosowaniach; można jednak i tutaj sięgnąć do podziału procesu decyzyjnego na fazę przygotowania i fazę wykonania, uzyskując w fazie przygotowania decyzji odpowiednie dane od podsystemów. Procedura iteracyjna fazy przygotowania decyzji polega w tym przypadku na proponowaniu przez jednostkę nadrzędną kolejnych wartości y_d i wysłuchiwaniu odpowiedzi jednostek lokalnych – na przykład przewidywanych przez nie wartości lokalnych funkcji celu $\hat{Q}_i$, jeśli mamy do czynienia z zadaniem (2.29). Można w ten sposób dojść do wartości $\hat{y}_d$, która maksymalizuje funkcję celu jednostki nadrzędnej, przechodząc potem do fazy wykonania decyzji, czyli do ostatecznego „narzucenia planu". Trudność jednostki nadrzędnej polega m.in. na tym, że kolejne proponowane wartości y_d muszą się mieścić w zbiorze dopuszczalnym Y_0, por. (2.30), muszą być wykonalne. Granice tego zbioru zależą od ograniczeń i równań lokalnych, pośrednio zależą od czynników zewnętrznych z_i. Nie mając ścisłych informacji i licząc się z niepewnością z_i, jednostka nadrzędna musi utrzymywać swoje propozycje w „bezpiecznej" części zbioru Y_0 z dala od jego brzegów, co prowadzić może do niewykorzystania możliwości całego systemu.

Wadę tę można usunąć za pomocą bardziej finezyjnych metod prowadzenia iteracji; jedną z nich jest metoda oparta na użyciu funkcji kary. Polega to na tym, że w fazie przygotowania decyzji wprowadza się do zadania lokalnego nie sztywne wymaganie równości $y_i = y_{di}$, lecz odpowiednią opłatę za różnicę y_i i y_{di}. Zadanie lokalne zawiera wówczas, w porównaniu z (2.28), zmodyfikowaną funkcję celu, czyli przybiera postać

$$\text{maksymalizować } [Q_i(c_i, u_i, z_i) + K\|y_{di} - y_i\|] \tag{2.40}$$

przy czym K jest dostatecznie dużym i wzrastającym w czasie iteracji współczynnikiem po to, by w wyniku osiągnąć decyzję lokalną $\hat{y}_i = y_{di}$. Podobny mechanizm potrzebny jest oczywiście po stronie jednostki nadrzędnej, ażeby jej kolejne propozycje y_d zmierzały w kierunku wartości wykonalnych. „Opłaty dodatkowe" z zadań lokalnych muszą więc pojawić się także w zadaniu decyzyjnym jednostki nadrzędnej, a wszystko to w formie zapewniającej zbieżność procesu iteracji do właściwego rozwiązania, to znaczy do decyzji $\hat{y}_d$, $\hat{c}$, stanowiących rozwiązanie zadania pierwotnego, bez funkcji kary.

Opisany sposób działania jednostek decyzyjnych w fazie przygotowania decyzji pochodzi z dziedziny metod obliczeniowych optymalizacji; znane są tam także inne sposoby zorganizowania efektywnego procesu iteracji w podobnych strukturach*).

Motywacja przekazywania informacji nieprawdziwej

Mówiąc o zakresie informacji dostępnej różnym jednostkom decyzyjnym wspomnieć jeszcze trzeba, że jednostki lokalne mogą być zainteresowane przekazywaniem informacji nieprawdziwej. Łatwo to pokazać na następującym przykładzie.

Niech lokalne zadanie decyzyjne ma postać

$$\left.\begin{array}{l} \text{maksymalizować } Q_i(c_i, u_i, z_i) \\ \text{przy ograniczeniu} \\ (c_i, u_i) \in CU_i(z_i) \end{array}\right\} \tag{2.41}$$

*) Patrz np. W. Findeisen i in., op. cit., rozdz. 2.

Załóżmy, że jednostka nadrzędna operuje koordynacją bezpośrednią. Podsystem dowiaduje się zatem, że wartość wejścia wyniesie u_{di} – nie leży ona w jego gestii. Dla tej wartości wejścia jednostka lokalna może obliczyć decyzję c_i^*, maksymalizującą $Q_i(c_i, u_{di}, z_i)$. Jednostka nadrzędna narzuca jednak także wartość wyjścia y_{di}, postulując równość

$$F_i(c_i, u_{di}, z_i) = y_{di} \tag{2.42}$$

Optymalizacja lokalna z dodaniem tego warunku prowadzić będzie zazwyczaj do wartości $\hat{c}_i$ innej niż c_i^*, w konsekwencji do mniejszej wartości lokalnej funkcji celu niż ta, którą obliczano bez warunku dodatkowego (2.42). Jednostka lokalna jest zainteresowana w tym, by narzucono jej wartość y_{di}^* określoną przez równanie

$$y_{di}^* = F_i(c_i^*, u_{di}, z_i) \tag{2.43}$$

Lokalna jednostka decyzyjna nie może powiedzieć tego jednostce nadrzędnej wprost, jest zatem skłonna podać nieprawdziwe dane o granicach zbioru CU_i, to znaczy takie, by z nich wynikało, że decyzja c_i^* jest możliwa, a decyzja $\hat{c}_i$ jest niemożliwa. W literaturze ekonomicznej ten rodzaj fałszowania danych nosi nazwę „ukrywania zdolności produkcyjnych".

Podany prosty przykład wskazuje na dwa aspekty zjawiska przekazywania do jednostki nadrzędnej nieprawdziwych danych. Pierwszy, że jest to dla jednostki lokalnej opłacalne oraz drugi, że w oczywisty sposób obniża to osiąganą wartość funkcji celu jednostki nadrzędnej. Zwróćmy uwagę, że nie jest to związane z niezgodnością celów. Nawet jeśli funkcja celu jednostki nadrzędnej będzie prostą sumą lokalnych funkcji celu, każda jednostka lokalna może mieć motywację do podania nieprawdziwych danych o swych możliwościach czyli o zbiorze CU_i. Przykład opisany wzorami (2.41)–(2.43) wskazuje, że decyzja c_i^* może być z lokalnego punktu widzenia lepsza niż decyzja $\hat{c}_i$, optymalna z punktu widzenia całości systemu. W rezultacie podania nieprawdziwych („zaniżonych") danych o zbiorze CU_i następuje obniżenie rezultatów osiąganych przez system jako całość.

Decyzje w obliczu niepewności

Przejdziemy teraz do innego aspektu podejmowania decyzji w odniesieniu do systemu istniejącego, działającego w czasie i poddanego wpływom otoczenia, a mianowicie do uwzględnienia niepewności. W sytuacjach niepewności, czyli gdy przyszłe oddziaływania ze strony otoczenia nie mogą być dokładnie przewidziane, ważne jest zdanie sobie sprawy, że można operować decyzjami jednoetapowymi (jednokrotnymi) oraz wieloetapowymi. Istotą *decyzji wieloetapowej* jest to, że podejmując decyzję w chwili t_i zakładamy od razu, że w chwili t_{i+1} podejmiemy decyzję następną (oraz że podobnie będzie dalej). Jest zatem tak, jak gdyby decyzja odnosząca się do sterowania systemem w określonym przedziale czasu nie była dokonywana z góry i od razu, lecz składała się z kolejnych etapów – stąd nazwa „decyzja wieloetapowa". Przeciwieństwem jest *decyzja jednoetapowa*, gdzie w chwili t_i zachowujemy się tak, jak gdyby nie istniała możliwość zmiany decyzji w przyszłości. Jeśli więc ustalamy w chwili t_i oddziaływanie sterujące, na przykład plan zrzutów wody ze zbiornika retencyjnego, robimy to z ostrożnością wybiegającą na całą przyszłość; wszystkie przyszłe dopływy i pobory wody są niepewne, zabezpieczamy się maksymalnie przed ewentualnością przepełnienia lub opróżnienia zbiornika. Przy operowaniu w tej samej sytuacji decyzją wieloetapową również bierzemy w chwili t_i pod uwagę całą przyszłą niepewność dopływów i poborów, ale z mniejszą ostrożnością – wiemy bowiem, że w chwili t_{i+1} będziemy mogli plan zrzutów wody zmienić. Istota różnicy polega zatem na stosunku do niepewności, manifestując się zwłaszcza wtedy, gdy w grę wchodzi obawa przed przekroczeniem nałożonych na sterowany system ograniczeń.

Decyzja jednoetapowa może być powtarzana; jest rzeczą oczywistą że w chwili t_{i+1} można podjąć następną decyzję, anulując poprzednią. Różnica w stosunku do decyzji wieloetapowej polega na tym, że w chwili t_{i+1} podejmowalibyśmy decyzję z taką ostrożnością, jak gdyby następnej decyzji miało już nie być.

Zwróćmy jeszcze uwagę, że charakterystyczna dla decyzji wieloetapowej chwila t_{i+1} nie musi być określona w sposób jawny, może ona być związana z obserwacją systemu i otoczenia; chwila t_{i+1} może na przykład wynikać z zasady, że decyzja jest modyfikowana gdy jakieś zmienne systemu lub zmienne otoczenia przekroczą pewne ustalone wartości progowe.

Pojęcie decyzji jednoetapowej i decyzji wieloetapowej wiąże się ze sposobem podejścia do niepewności, nie pokrywając się z podziałem struktur sterowania na układy otwarte i układy ze sprzężeniem zwrotnym, co widać szczególnie jasno gdy myślimy o ustosunkowaniu się jednostki decyzyjnej do niepewności otoczenia. Tym niemniej warto zauważyć, że w układzie ze sprzężeniem zwrotnym realizowana jest zasada decyzji wieloetapowej, gdyż prawo sterowania przekształcające obserwacje procesu sterowanego w oddziaływanie sterujące zakłada z góry nieustanne modyfikowanie tego oddziaływania na podstawie nowych danych o procesie. Z kolei zaprojektowane na stałe prawo sterowania jest decyzją jednoetapową, obliczoną na wszystkie możliwe stany procesu i wartości jego parametrów, chyba że przewidujemy jakiś mechanizm adaptacji.

Rozdział 3

Przebiegi procesów w obiektach systemu

3.1. Obiekty i procesy

Rozróżnianie wejść i wyjść

W rozdziale 2 opieraliśmy rozważania o systemie na stwierdzeniu, że każdy z wchodzących w dany system podsystemów opisuje się zależnością (2.2), która ma postać

$$y_i = F_i(c_i, u_i, z_i)$$

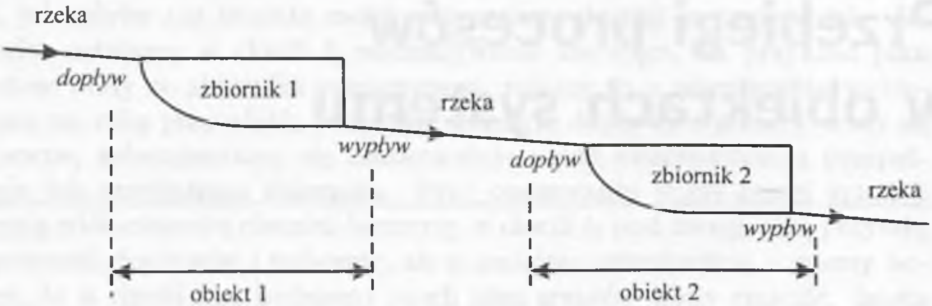
Nie podkreślaliśmy wówczas, że mamy na myśli skutek i przyczynę, że wyrażona tym wzorem – dla każdego z podsystemów – zależność wyjścia y_i od wejścia u_i jest *związkiem przyczynowym*, podobnie jak zależność y_i od sterowania c_i. Stwierdzenie, iż wyjście jest skutkiem wejścia należy rozumieć w ten sposób, że zmiany u_i wywołują zmiany y_i, a nie przeciwnie.

Związek przyczynowy to coś więcej niż korelacja, kiedy dwa zjawiska wykazują – w obserwacjach – pewien związek, ale przyczyna każdego z tych zjawisk leży na zewnątrz. Mówimy o tym w rozdz. 5.3.

W obecnym rozdziale mówić będziemy o „obiektach" jako takich elementach systemu, które – w odróżnieniu od „podsystemu" – nie mogą już podlegać dalszemu racjonalnemu podziałowi. W obiektach przebiegają procesy będące przedmiotem naszej analizy i zainteresowania.

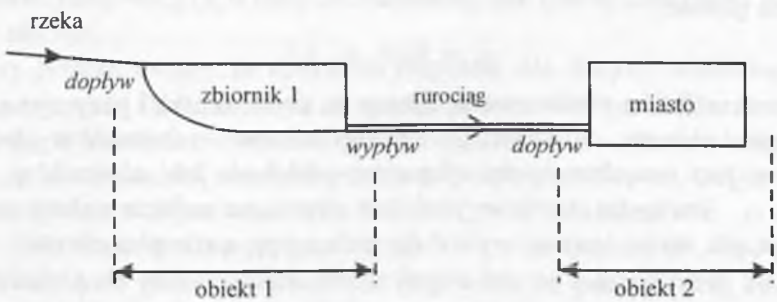
Przyjrzyjmy się najpierw sprawie „wejścia" i „wyjścia" obiektu w systemie na prostym przykładzie.

Na rzece istnieją dwie zapory i powstały dwa zbiorniki wodne, które traktować chcemy jako odrębne obiekty systemu (rys. 3.1). Woda z pierwszego zbiornika jest wypuszczana do rzeki w sposób kontrolowany, przez położenie zasuwy przy zaporze. Powstający przepływ stanowi wyjście pierwszego obiektu oraz wejście obiektu drugiego. Jest tak dlatego, że wypływ z pierwszego zbiornika ma swoje przyczyny w tym właśnie obiekcie (położenie zasuwy będące sterowaniem, dopływ z górnego odcinka rzeki stanowiący wejście). Wypływ z pierwszego zbiornika jest zarazem wejściem do zbiornika drugiego, jest jedną z przyczyn powstających w nim skutków, na przykład ilości nagromadzonej wody.



Rys. 3.1. Dwa zbiorniki wodne

Załóżmy teraz, że wypływ ze zbiornika na rzece jest kierowany przez rurociąg do odbiorców w mieście oraz że o natężeniu przepływu wody decydują pompy i zawory sieci wodociągowej tego miasta (rys. 3.2). Tym razem wypływ wody ze zbiornika jest skutkiem decyzji podejmowanych w obrę-



Rys. 3.2. Zbiornik zasilający miasto

bie innego obiektu; odgrywa on taką samą rolę „przyczyny" jak dopływ z górnego odcinka rzeki.

Klarowne rozróżnienie wejść i wyjść, przyczyn i skutków, nie zawsze jest łatwe. Powstające trudności mogą skłonić lub wręcz zmusić do zaniechania prób wydzielenia poszczególnych obiektów. Na przykład, w systemie z rys. 3.2 przepływ przez rurociąg zależy w sposób fizyczny zarówno od pracy pomp na końcu rurociągu, jak od ciśnienia na wlocie, czyli od poziomu wody w zbiorniku. Przyczyny przepływu leżą w obu obiektach, czy można więc obiekty te rozdzielić i sprowadzić do potrzebnego nam schematu? Sytuacja wyjaśnia się, jeżeli przyjąć, że pompy lub zawory miejskiej sieci wodociągowej są sterowane tak, by wywołać określony (potrzebny miastu) przepływ przez rurociąg zasilający, niezależnie od ciśnienia na jego wlocie;

teraz podział na obiekty jest klarowny, przepływ w rurociągu jest rzeczywiście wejściem do obiektu pierwszego oraz wyjściem obiektu drugiego. Dla zbiornika jest to „pobór wody”, od zbiornika niezależny.

W rzeczywistości, zbiornik z rys. 3.2 mieć będzie także ujście do rzeki, gdyż inaczej nie byłoby możliwości sterowania jego napełnieniem.

Trzeba w tym momencie bardziej precyzyjnie spojrzeć na pojęcie „wejścia” i „wyjścia” obiektu w systemie. Dla rozważań dotyczących sterowania „wejściem” będzie to, czym proces (i sterowanie) obiektu sąsiedniego oddziałują na proces przebiegający u mnie, a „wyjściem” to, co oddziałuje na proces w obiekcie następnym. Mogą także być wejścia (tj. przyczyny takiego, a nie innego przebiegu mojego procesu) ze strony otoczenia oraz wyjścia nie skierowane do innego obiektu, lecz na zewnątrz systemu – zatem oddziałujące na otoczenie.

Należy przeto dokonać ważnego rozróżnienia między tym co można nazwać *schematem technologicznym systemu*, a *schematem opisującym wzajemne oddziaływania*.

Proces odbywający się w obiekcie ma swoje wejścia i wyjścia w sensie fizycznym, strumienie wejściowe i wyjściowe substancji lub energii, dopływy surowców i półfabrykatów, a odpływy produktów. Rysując schemat technologiczny systemu pokazywalibyśmy te właśnie powiązania, traktując strumienie dopływające jako wejścia, a odpływające jako wyjścia. Przypominamy tutaj przykład rafinerii, opisany w rozdziale 1.5, rys. 1.9.

Rzecz w tym, że w schemacie będącym podstawą do rozważania i projektowania struktur sterowania interesują nas nie tyle przepływy substancji, co oddziaływania jednego procesu na drugi. Kierunek oddziaływań może być, jak widzieliśmy, niezgodny z kierunkiem przepływu substancji.

Wspólna dla schematu technologicznego i schematu oddziaływań pozostaje struktura systemu, jego podział na obiekty lub podsystemy.

Co więcej, o ile kierunek przepływu strumieni jest cechą trwałą, wynikającą z koncepcji technologicznej (w szczególnych przypadkach strumień może przepływać naprzemian w obu kierunkach, np. strumień energii w linii przesyłowej), to kierunek oddziaływania jest w istocie rezultatem przyjętej struktury decyzyjnej, ustalenia kto ma o czym decydować. Struktura decyzyjna może zostać zmieniona, kompetencje przesunięte, zasada „tłoczenia” może być zastąpiona „ssaniem”. Kierunek oddziaływania jednego obiektu na drugi (ściśle: oddziaływania procesu przebiegającego w jednym obiekcie na proces przebiegający w drugim) stanie się odwrotny do poprzedniego.

Używając w tej książce, jako poświęconej zagadnieniom sterowania a nie technologii, wzorów w rodzaju (2.2) oraz słów „wejście” i „wyjście” mieć będziemy na myśli kierunki oddziaływania; w ten także sposób trzeba rozumieć tekst rozdziału 2.

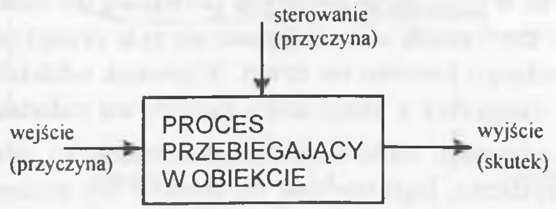
Opis procesu w obiekcie

Przypomnijmy za rozdziałem 1, że zadaniem sterowania jest zapewnienie pożądanego biegu procesu lub procesów w systemie – obecnie rozpatrywać to będziemy bardziej szczegółowo. Zajmiemy się procesami w poszczególnych częściach, bowiem – jak wiemy z rozdziału 2 – proces w systemie składa się z procesów przebiegających w jego częściach, na skutek połączenia wyjść i wejść.

Punktem wyjścia rozważań staje się obiekt, już wydzielony z większej całości, czyli systemu; przedmiotem rozważań jest odbywający się w tym obiekcie proces. Będzie to w istocie rzeczy zagłębienie się we wnętrze obiektu, znalezienie podstaw, na których może być oparty używany dotąd opis (2.2).

Rozważmy jako prosty przykład zbiornik wodny utworzony przez zaporę na rzece, taki jak pokazano na rys. 3.1. Zbiornik służy do zatrzymywania (retencji) wody dopływającej z górnego odcinka rzeki po to, by nie dopuścić do powodzi, a także by zapewnić zasilanie w wodę dolnego odcinka rzeki w okresie małego dopływu z odcinka górnego. Woda wypływa ze zbiornika pod własnym ciężarem, przez otwór przymykany bądź otwierany odpowiednią zasuwą.

Rozpatrywany obiekt ma *wejście* (dopływ wody), *wyjście* (wypływ wody) oraz *sterowanie* (przemieszczanie zasuwy); można to wyrazić prostym schematem, podanym na rys. 3.3.



Rys. 3.3. Schemat obiektu

W rozpatrywanym obiekcie odbywa się *proces*, polegający na gromadzeniu (akumulacji) wody, przy czym gromadzenie to zależy od dopływu i wypływu. Aby proces gromadzenia opisać i rozważać, musimy wprowadzić pojęcie *stanu procesu*. Będzie w nim w naszym przypadku ilość wody w zbiorniku.

W opisie matematycznym procesu odbywającego się w zbiorniku wodnym wystąpi zatem współrzędna stanu (zwana też zmienną stanu) oraz wystąpić muszą zmienne charakteryzujące wejście, wyjście i sterowanie w sposób taki, by można było zmienne te powiązać ze zmienną stanu. W naszym przykładzie:

– współrzędną stanu procesu (zmienną stanu) będzie ilość wody x [m^3],

– wielkością wejściową procesu (zmienną wejściową) będzie natężenie dopływu z [m^3/s],
– wielkością wyjściową procesu (zmienną wyjściową) będzie natężenie wypływu y [m^3/s],
– wielkością sterującą procesu (zmienną sterującą) będzie położenie zasuwy m, w odpowiednich jednostkach.

Związki między zmiennymi sprowadzą się do dwóch podstawowych równań; będą to mianowicie:
– równanie wyjścia (określające wypływ)

$$y(t) = g(x(t), m(t)) \tag{3.1}$$

– równanie określające stan

$$x(t) = \int_{-\infty}^{t} z(\tau)\mathrm{d}\tau - \int_{-\infty}^{t} y(\tau)\mathrm{d}\tau \tag{3.2}$$

W pierwszym z tych równań wyraziliśmy rzecz dosyć oczywistą, że natężenie wypływu w danej chwili t jest funkcją poziomu wody w zbiorniku (czyli jej ilości) oraz położenia zasuwy w tej samej chwili. Drugie równanie określa ilość nagromadzonej wody jako rezultat całego dotychczasowego dopływu i wypływu. Jest to formuła mało praktyczna; dogodnie jest wprowadzić *stan początkowy*, czyli wartość współrzędnej stanu w określonej chwili początkowej t_0 oraz w konsekwencji napisać

$$x(t) = x(t_0) + \int_{t_0}^{t} z(\tau)\mathrm{d}\tau - \int_{t_0}^{t} y(\tau)\mathrm{d}\tau \tag{3.3}$$

Równanie (3.3) określa stan x w chwili bieżącej, czyli $x(t)$ jako funkcję stanu początkowego $x(t_0)$ oraz przebiegów czasowych dopływu i wypływu za okres (t_0, t), czyli od chwili początkowej do obecnej. Równanie (3.3) dobrze oddaje istotę procesu, czyli gromadzenie, ale jest mało poręczne w obróbce matematycznej. Stąd zazwyczaj zapisujemy nie równanie (3.3), lecz równanie powstające przez zróżniczkowanie względem czasu wszystkich jego wyrazów. Otrzymamy wówczas postać

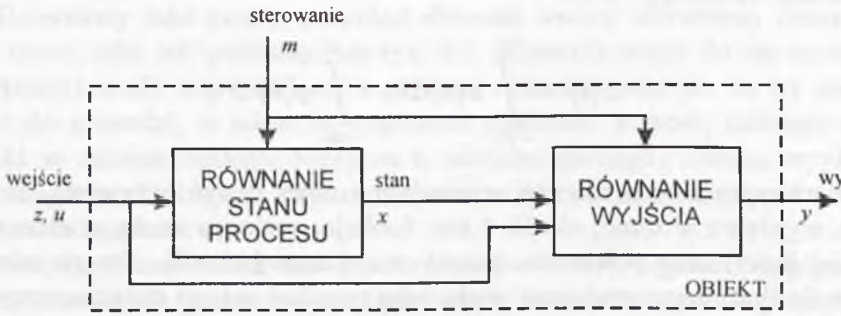
$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = z(t) - y(t)$$

a następnie, biorąc pod uwagę, że $y(t)$ jest funkcją $x(t)$, $m(t)$ (patrz (3.1)), dochodzimy do najczęściej przyjmowanej postaci *równania stanu procesu* w obiekcie

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = f(x(t), m(t), z(t)) \tag{3.4}$$

Warto może podkreślić, że przy operowaniu bardziej zwartym w zapisie równaniem (3.4) także trzeba sięgnąć do wartości stanu początkowego $x(t_0)$ oraz przebiegów wejść sterowań, które miały miejsce po chwili t_0, jeśli zachodzi potrzeba podania wyników ilościowych. Zwykliśmy mówić, że równanie (3.4) wyraża *prędkość akumulacji*, podczas gdy równanie (3.3) określało jej wartość w sposób bezpośredni.

Otrzymany *opis procesu w obiekcie* (mówiąc skrótowo – *opis obiektu*) możemy przedstawić schematycznie jak na rys. 3.4, pokazując związek między równaniem stanu i równaniem wyjścia, w konsekwencji – według jakich praw przekształca się wejściei sterowaniew wyjście.



Rys. 3.4. Równanie stanu i równanie wyjścia

Pozostaje jeszcze odnieść się do formuły (2.2), mającej postać, dla dowolnego obiektu w systemie

$$y_i = F_i(c_i, u_i, z_i)$$

Widzimy, że formuła (2.2) ma postać bardziej ogólną, mówi tylko o tym, co stanowi przyczynę, a co skutek. W szczególności y_i, c_i, u_i, z_i nie oznaczają wartości chwilowych, lecz całe przebiegi czasowe, a $F_i(\cdot)$ oznacza związek między tymi przebiegami – zależny ponadto, jak wiemy z przykładu, od stanu początkowego procesu w obiekcie.

Uważnemu czytelnikowi wypada w tym miejscu wyjaśnić, dlaczego w równaniach (3.1), (3.4) sterowanie oznaczono przez m, a w równaniu (2.2) przez c. Litera m używana jest w tej książce na oznaczenie „wielkości manipulowanej", czyli tej, która stanowi sterowanie w sensie bezpośrednim, fizycznym (położenie zasuwy, zasilanie pompy itp.). Litera c oznacza wielkość sterującą działającą pośrednio, na przykład zadawany pompie (i podtrzymywany przez odpowiedni układ regulacji) przepływ wody. Na poziomie systemu mamy raczej do czynienia ze sterowaniami pośrednimi, stąd użycie oznaczenia c w rozdziale 2. Sprawa ta będzie dokładnie wyjaśniona w rozdziale 4 (przy omawianiu układów dwuwarstwowych).

Wielkość wejściową rozpatrywanego zbiornika oznaczyliśmy przez z dlatego, że traktujemy ją w tym przykładzie jako pochodzącą z zewnątrz systemu. Odpowiada to położeniu lewego zbiornika na rys. 3.1. Dopływ do zbiornika prawego, w dole rzeki, oznaczalibyśmy przez u, jako wejście pochodzące z wyjścia innego obiektu tego samego systemu.

Opisany proces w zbiorniku wodnym jest *procesem dynamicznym*, ponieważ występuje tu zjawisko akumulacji wody. Stan nagromadzenia trzeba było odpowiednio wyrazić, zrobiliśmy to za pomocą współrzędnej stanu x.

Większość procesów, z jakimi mamy do czynienia, jeśli nie wszystkie, są to procesy dynamiczne – tyle, że niekiedy można zjawiska związane z akumulacją pominąć. W rozpatrywanym przykładzie napisaliśmy bez głębszego uzasadnienia równanie (3.1) mówiące, że natężenie chwilowe wypływu wody $y(t)$ zależy bez żadnej inercji od otwarcia zasuwy $m(t)$ – w rzeczywistości będą tu występowały zjawiska hydrodynamiczne, związane z nagromadzeniem energii, tyle że dziejące się w skali czasowej bardzo krótkiej w stosunku do czasowej zmienności zjawisk związanych z akumulacją wody w całym zbiorniku. Intuicyjnie zrobiliśmy opis uproszczony.

Rodzaje procesów i ich przebiegi pożądane

Przebiegi procesów w obiektach systemu, o które dbać będą układy sterowania, związane są z rozwiązaniami technologicznymi tych obiektów.

Mamy więc w szczególności:

- *procesy jednorazowe*, takie jak przepłynięcie statku z portu A do B lub przeprowadzenie reakcji w reaktorze wsadowym,
- *procesy ciągłe*, takie jak proces retencji wody w zbiorniku zbudowanym na rzece, czy też procesy chemiczne w aparatach i instalacjach rafinerii ropy naftowej.

Ciąg powtarzanych procesów jednorazowych tworzy *proces cykliczny*; będzie takim procesem sekwencja rejsów statku, praca samolotu linii lotniczych, produkcja stali w piecu stalowniczym. Zobaczymy za chwilę, że przy spojrzeniu z pewnej odległości – czyli w skali makro – proces cykliczny staje się podobny do procesu ciągłego.

Najbardziej charakterystyczną cechą *procesu ciągłego* jest to, że jego bieg może być rozpatrywany w oderwaniu od *pierwotnego stanu początkowego* (zbiornik na rzece zbudowano wiele lat temu, był wtedy pusty, ale to nie ma znaczenia dla biegu procesu obecnie), a także w oderwaniu od *chwili końcowej* i *stanu końcowego* (zbiornik ma być jeszcze eksploatowany przez dziesiątki lat, kiedyś go całkowicie opróżnimy, ale to nie ma znaczenia dla chwili obecnej).

W odróżnieniu od tego, w przypadku *procesu jednorazowego* istotną rolę odgrywa chwila początkowa i chwila końcowa, a także stan początkowy (statek znajduje się w porcie A, ma zerową prędkość) oraz stan końcowy (port

B, prędkość równa zeru). Ścieżkę przejścia procesu jednorazowego od stanu początkowego do końcowego nazywamy *trajektorią stanu.*

Zwróćmy teraz uwagę, że *proces cykliczny* ma pewną cechę wspólną z procesem ciągłym, a mianowicie jego bieg można często rozpatrywać w oderwaniu od „pierwszej chwili początkowej" (od pierwszego uruchomienia) oraz bez uwzględnienia wpływu „ostatniej chwili końcowej", to jest zaprzestania eksploatacji.

Co więcej, patrząc na proces cykliczny w dłuższej skali czasu oraz bez domagania się szczegółów można go opisać tak jak proces ciągły; na przykład, na szczeblu sterowania produkcją huty piec stalowniczy będzie się prezentował jako obiekt wytwarzający pewną liczbę ton stali na tydzień lub na miesiąc. Fakt, że odbywa się to z zastosowaniem procesu cyklicznego nie odgrywa roli.

Trzeba jeszcze wymienić *procesy złożone z operacji dyskretnych*, polegające na tym, że na przetworzenie produktu wejściowego w produkt wyjściowy składa się szereg operacji technologicznych wykonywanych w pewnej kolejności (na przykład na różnych obrabiarkach), przy czym kolejność ta, czyli uszeregowanie operacji może dopuszczać pewną swobodę, a ponadto mogą w systemie istnieć rozgałęzienia i ścieżki równoległe.

Każda z operacji dyskretnych w takim systemie technologicznym jest procesem jednorazowym w znaczeniu użytym wcześniej (występuje początek, koniec, czas trwania itd.), a procesy jednorazowe powtarzane na tej samej obrabiarce składają się na cykliczny proces pracy tej obrabiarki. Rozpatrywanie całości procesu w systemie jako procesu złożonego z operacji dyskretnych nie jest więc niczym innym jak spojrzeniem w skali makro, które nie wnika w szczegóły każdej z operacji, lecz rozpatruje ich wzajemne uzależnienie i uporządkowanie, na przykład mając na myśli osiąganie możliwie najlepszych rezultatów, z uwzględnieniem wszelkich ograniczeń.

Zwróćmy uwagę, że w istocie rzeczy podobnym spojrzeniem na procesy fizyczne przebiegające w obiektach systemu technologicznego, jest rozpatrywanie go na szczeblu sterowania produkcją w sposób niezależny od tego czy są to procesy ciągłe, czy cykliczne.

Powracając do opisu procesu w obiekcie, który był już w tym rozdziale rozpatrywany przypomnijmy, że istotną rolę w procesie dynamicznym odgrywają zjawiska akumulacji, a zatem przebieg współrzędnych stanu. Myśląc o celu obecnych rozważań, czyli o pożądanym biegu procesów, zobaczymy w dalszym ciągu tego rozdziału, że w *procesach ciągłych* pożądane lub użyteczne mogą się okazać:

– wartości współrzędnych stanu zmienne w dostosowaniu do czynników zewnętrznych (zwłaszcza losowych),
– wartości współrzędnych stanu zmienne z przyczyn wewnętrznych (na przykład dla uzyskania większej wydajności procesu),

– wartości współrzędnych stanu stałe (na przykład z przyczyn technologicznych lub z powodu oddziaływania ograniczeń).

W przebiegu *procesów jednorazowych* i *cyklicznych* wartości współrzędnych stanu są zmienne z samej istoty tych procesów, można jednak rozróżnić dwa przypadki:

– trajektoria stanu, czyli ścieżka przejścia od stanu początkowego do końcowego jest w jakimś stopniu swobodna (można wybierać trasę przelotu samolotu),
– trajektoria stanu jest narzucona i pozostaje tylko jej przestrzeganie (np. ustalony technologicznie przebieg temperatury przy wypalaniu ceramiki).

W następnych podrozdziałach zajmiemy się wszystkimi tymi sytuacjami bardziej szczegółowo, sięgając do przykładów.

3.2. Przebiegi procesów ciągłych

Procesy przebiegające ze zmiennością stanu

Proces retencji wody w zbiorniku na rzece, omawiany w poprzednim podrozdziale, stanowi dobry przykład sytuacji, gdzie pożądany przebieg procesu charakteryzuje się zmiennością stanu w czasie. Łatwo sobie wyobrazić, że przeciwne do tego dążenie do utrzymania stałej zawartości wody w zbiorniku, na przykład zainstalowanie urządzenia stabilizującego poziom wody przez odpowiednie do tego oddziaływanie na zasuwę, byłoby nonsensem; w takiej bowiem sytuacji zbiornik przestałby spełniać swoje zadanie, polegające na przyjmowaniu nadmiaru dopływającej wody w okresach przyboru oraz zwalnianiu posiadanego zapasu w okresie suchym. Przy utrzymywaniu stałego poziomu, czyli przy stabilizacji stanu, zmienność natężenia dopływu przenosiłaby się na wyjście obiektu w postaci identycznej zmienności natężenia odpływu – to samo mielibyśmy w rzece bez zapory i zbiornika.

Podane jakościowe rozumowanie niezbyt łatwo jest uczynić bardziej konkretnym, czyli określić, jaka zmienność stanu w czasie będzie najbardziej właściwa. W dużym uproszczeniu można to przedstawić jak następuje.

Zbiornik zasila w wodę dolny bieg rzeki; można przyjąć, że natężenie wypływu $y(t)$ powinno się zawierać w pewnych zadanych granicach

$$y_{\min} \leqslant y(t) \leqslant y_{\max} \tag{3.5}$$

Granica górna związana jest z ewentualnością wystąpienia rzeki z brzegów i spowodowania strat typu powodziowego, granica dolna może pochodzić z warunków zasilania odbiorców wody, możliwości żeglugi, podtrzymania życia biologicznego w rzece itd.

Zbiornik ma określoną pojemność maksymalną, ale w sytuacji normalnej nie wolno go całkowicie napełnić, gdyż trzeba pozostawić tak zwaną rezerwę powodziową, czyli wolne miejsce na przyjęcie ewentualnego, nagłego przyboru wód. Nie powinno się go także całkowicie opróżnić, aby mieć zapas wody na przypadek szczególnej suszy. Prowadzi to do narzucenia ograniczeń na wartości stanu zbiornika

$$x_{\min} \leqslant x(t) \leqslant x_{\max} \tag{3.6}$$

Zbiornik powinien być tak sterowany, by zachować w trakcie normalnej eksploatacji (poza okresem zagrożenia powodzią lub suszą) obydwa układy ograniczeń, czyli w sumie cztery ograniczenia nierównościowe.

Przebiegi stanu x oraz wyjścia y, które mieszczą się w ograniczeniach, uznać można za jednakowo dobre. W szczególności nie miałoby sensu podtrzymywanie stałej wartości wypływu, na przykład wartości leżącej w środku dopuszczalnego przedziału, gdyby miało temu towarzyszyć niebezpieczeństwo przekroczenia ograniczeń na stan. Istota problemu leży w tym, że przy określaniu decyzji o wypuszczeniu (zrzucie) wody ze zbiornika podejmowanej w chwili t trzeba brać pod uwagę przyszłe stany i dopływy, podczas gdy te przyszłe dopływy są zjawiskiem losowym. W chwili bieżącej znany jest jedynie dopływ obecny, $z(t)$ oraz dopływy w przeszłości, dopływ przyszły można określić tylko w kategoriach probabilistycznych. Zanim to uwzględnimy, przeprowadźmy bardzo proste rozważanie deterministyczne: przebieg dopływu z w przeszłości, na przykład ten, który miał miejsce w roku ubiegłym, nie jest już wielkością losową, jest znaną funkcją czasu $z_{[t_1,t_2]}$. Postawmy pytanie: czy przy tym przebiegu dopływu możliwe było utrzymanie się w zadanych ograniczeniach? Łatwo zauważyć, że możliwe były trzy przypadki:

1) istniało wiele przebiegów zrzutu wody $y_{[t_1,t_2]}$ oraz trajektorii stanu $x_{[t_1,t_2]}$, które przy danym $z_{[t_1,t_2]}$ zachowują zadane ograniczenia,
2) istniał tylko jeden przebieg zrzutu wody $y_{[t_1,t_2]}$ oraz jedna trajektoria stanu $x_{[t_1,t_2]}$, które nie przekraczają ograniczeń przy danym dopływie $z_{[t_1,t_2]}$,
3) zachowanie niektórych lub nawet wszystkich ograniczeń na wyjście oraz stan procesu nie było – przy danym $z_{[t_1,t_2]}$ – możliwe.

Podajmy trywialny przykład wskazujący możliwość zaistnienia przypadku trzeciego: jeśli średnia roczna wartość dopływu $z_{\text{śr}}$ leży poniżej granicy $y_{\min}$, to granicy tej zachować nie można (chyba że w zbiorniku udało się zachować zapas wody z okresu poprzedzającego przedział $[t_1, t_2]$, czyli z lat ubiegłych – wskazuje to m.in. na znaczenie pojemności całkowitej zbiornika odniesionej do wartości $z_{\text{śr}}$).

W pewnym sensie najbardziej wygodny jest przypadek drugi, gdzie zachowane są ograniczenia, a ponadto pożądany przebieg stanu i przebieg wielkości wyjściowej (tym samym także przebieg sterowania) są określone w sposób jednoznaczny, nie wymagając dokonania wyboru.

Przypadek drugi jest oczywiście sytuacją bardzo szczególną, granicą między przypadkiem pierwszym (istnieje wiele rozwiązań) i przypadkiem trzecim (nie istnieje żadne rozwiązanie).

Przejdźmy teraz do sytuacji bardziej rzeczywistej, to jest losowej: przyszły dopływ $z_{[t_1,t_2]}$ nie jest wiadomy, znane są tylko jego cechy probabilistyczne. Określenie pożądanego przebiegu stanu $x_{[t_1,t_2]}$ oraz wyjścia $y_{[t_1,t_2]}$ (bądź sterowania $m_{[t_1,t_2]}$, co jest równoznaczne wobec równania (3.1)) można oprzeć na żądaniu, by minimalne było prawdopodobieństwo, że nastąpi przekroczenie choćby jednego z czterech narzuconych ograniczeń.

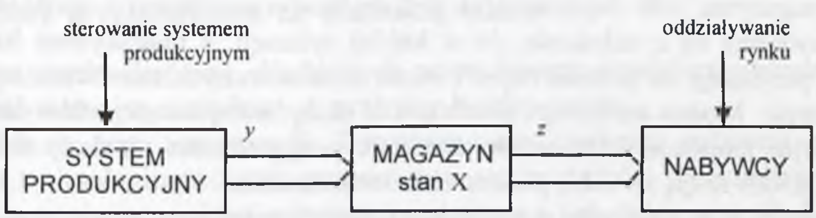
Nie rozwijając dalej wątku rozwiązywania tego zadania, powróćmy do jego sformułowania. Występuje tu bardzo typowa trudność – łatwo jest powiedzieć, że nie należy przekroczyć żadnego ograniczenia, jak jednak ustosunkować się do przekroczeń o prawdopodobieństwach różnych od zera? Czy wszystkie cztery zadane nierówności są tak samo ważne? Czy należy brać pod uwagę czas trwania przekroczenia? Czy dwa przekroczenia po dwie godziny w odstępie kilku dni to jest to samo, co jedno czterogodzinne? Proste te pytania wskazują, jak trudno jest sformułować kryterium, mogące wskazać na najbardziej pożądany przebieg procesu, zwłaszcza przy dominacji czynnika losowego.

Nie ma zatem nic niewłaściwego w tym, że często bardziej racjonalne jest sięgnięcie do intuicyjnej oceny decydenta („ten przebieg jest dobry"), niż matematycznie dokładne rozwiązanie zadania, opartego na mało wiarygodnym lub zbyt uproszczonym wskaźniku jakości.

Tak czy inaczej, w rozpatrywanym przykładzie użyteczny dla nas przebieg stanu procesu jest zmienny w czasie, a na samym procesie ciąży losowość dopływu, czyli losowość zasilania zbiornika.

Można też przedstawić inną charakterystyczną sytuację pracy obiektu o charakterze zbiornika bądź magazynu – taką mianowicie, gdzie chcielibyśmy zachować stałe lub prawie stałe zasilanie, natomiast zmienne, a nawet losowe jest obciążenie, np. ilość produktu zamawianego przez odbiorców.

Sytuację taką przedstawia rys. 3.5. System produkcyjny wprowadza produkt do magazynu, z którego jest on pobierany w zależności od zamówień rynku.



Rys. 3.5. Magazyn pośredniczący między systemem produkcyjnym a nabywcami

Stan magazynu w chwili t wyrazi się wzorem

$$x(t) = x(t_0) + \int_{t_0}^{t} y(\tau)\mathrm{d}\tau - \int_{t_0}^{t} z(\tau)\mathrm{d}\tau$$

w którym y oznacza natężenie produkcji, czyli natężenie dopływu produktu, a z natężenie zbytu.

Podobieństwo między magazynem a zbiornikiem retencyjnym na rzece (por. równanie (3.3)) jest wyraźne. Podobne będzie także zadanie operowania magazynem produktu. W szczególności uważać będziemy za użyteczne by stan napełnienia był zmienny w czasie, będą także ograniczenia nierównościowe na stan, a mianowicie

$$0 \leqslant x(t) \leqslant x_{\max}$$

gdzie $x_{\max}$ oznacza pojemność maksymalną.

Są także istotne różnice. Dopływ wody do zbiornika retencyjnego jest wielkością losową, na którą nie mamy wpływu – jeśli pominiemy działania długofalowe, polegające na zmianach zagospodarowania i ukształtowania zlewni. W sytuacji magazynu zamówienia na produkt są generowane przez nabywców, dla magazynu pozostają czynnikiem zewnętrznym, ale w szerszym kontekście mogą być od nas po części zależne. Przypominamy tutaj rys. 1.10, związany z przykładem rafinerii ropy naftowej.

Dodajmy jeszcze tytułem komentarza, że zadanie dostosowania wielkości produkcji do zamówień może być rozwiązane także bez składowania lub ze składowaniem bardzo ograniczonym. Można mianowicie tworzyć systemy produkcyjne na tyle elastyczne, że są one w stanie zmieniać natężenie swojej produkcji w szerokich granicach, a przez to dostosowywać chwilowe (np. dobowe) y do aktualnego z. Rozwiązanie takie można w pewnym stopniu dostrzec w przykładzie rafinerii: komponowanie benzyn jest elastyczne, można wytwarzać trzy rodzaje benzyny w różnych ilościach, zależnie od potrzeb (na przeszkodzie stają tu jednak różne ograniczenia).

Można także utrzymywać stałą wielkość produkcji, a mimo to obejść się bez magazynu, jeśli dopuszczalna jest realizacja zamówień z opóźnieniami. Wychodzimy tu z założenia, że w każdej sytuacji, z magazynem lub bez, suma produkcji za pewien okres i suma zrealizowanych zamówień będą sobie równe. Można zatem np. produkować stałą liczbę samochodów dziennie, realizując zamówienia klientów zawczasu przygotowane na dany dzień, tj. takie, które mają za sobą pewien czas oczekiwania.

Powróćmy do zbiornika retencyjnego albo do magazynu produktu – w obu tych przykładach zmienność stanu była zjawiskiem użytecznym, leżała u podstaw zaprojektowania procesu. Pożądany przebieg stanu był przez wpływ zewnętrzny niejako podyktowany, jednak nie w sposób trywialny: wskazywaliśmy, jak trudne może być nawet sformułowanie racjonalnych wymagań, nie mówiąc już o wyliczeniu przebiegów najbardziej właściwych.

Można wskazać przypadki nieco odmienne takich sytuacji, gdzie stan obiektu musi być dostosowany do czynnika zewnętrznego – a mianowicie takie, gdzie czynnik zewnętrzny dyktuje pożądany przebieg stanu w sposób bezpośredni. Jedno z klasycznych zastosowań automatyki – śledzenie lotu samolotu przez antenę stacji radiolokacyjnej – jest tu dobrym przykładem. Współrzędne ruchu anteny w kącie kierunku i kącie podniesienia, łącznie cztery współrzędne stanu procesu (kąty i prędkości), muszą być dostosowane do niezależnego od stacji ruchu samolotu. Istota różnicy w stosunku do przykładów poprzednich polega na tym, że pytanie „jak mają one być dostosowane?" ma tutaj odpowiedź bardzo prostą – antena ma być nakierowana na samolot w każdej bieżącej chwili, ma za nim nadążać. Nie ma tutaj zupełnie, charakterystycznego dla magazynów, troski o „przechowanie czegoś na przyszłość".

Zmienność stanu procesu ciągłego może być pożądana także w innych sytuacjach niż pokazane dotąd, a mianowicie z przyczyn wewnętrznych. Istnieją procesy, których sprawność lub wydajność wzrasta gdy są one prowadzone ze zmiennością stanu; może to m.in. zachodzić przy wymianie ciepła (tzw. pulsująca praca wymienników ciepła), a także przy reakcjach chemicznych. Prowadzony z umyślną zmiennością stanu proces ciągły staje się pokrewny procesom cyklicznym; powrócimy do tego w rozdziale 3.3.

Obecnie kontynuować będziemy rozważania dotyczące procesów ciągłych, szukając przypadków, gdy pomimo wpływów zewnętrznych pożądany będzie stan procesu stały w czasie. Sytuacje takie mają duże znaczenie praktyczne, są rozpowszechnione w technologiach przemysłu chemicznego i pokrewnych. Z punktu widzenia techniki sterowania stałość stanu procesu prowadzi do opisu przez *modele statyczne*, nie uwzględniające dynamiki. Znaczne uproszczenie obliczeń powoduje, że daje to możliwość objęcia wspólnym sterowaniem systemów znacznie większych, niż miałoby to miejsce przy potraktowaniu dynamicznym, czyli że można tu osiągać znacznie większą *rozległość sterowania*.

Procesy przebiegające ze stałością stanu

Procesy technologiczne, dla których prawidłowego przebiegu pożądana jest stałość stanu, są spotykane w praktyce bardzo często.

Jakie są tego przyczyny? Pozostają poza dyskusją sytuacje gdy warunki technologiczne biegu procesu wymagają stałości parametrów takich jak: temperatura, skład substancji, ciśnienie i tak dalej, a w ślad za tym potrzebna jest stałość współrzędnych stanu tego procesu. Zastanówmy się jednak czy także poza wymaganiami technologicznymi może zaistnieć sytuacja gdy prowadzenie procesu ze stałością stanu jest prowadzeniem optymalnym.

Przyjmijmy do rozważań proces opisany równaniem stanu

$$\dot{x}(t) = f\left(x(t), m(t), z(t)\right) \tag{3.7}$$

podobnym do równania przykładowego (3.4) z tym, że zmienne x, m, z mogą tu być wektorami.

Uznajmy, że proces ten ma być prowadzony w sposób ciągły, co oznacza, że rozpatrujemy go w pewnym przedziale czasowym $[t_1, t_2]$ z dala od chwili rozruchu t_0 oraz ewentualnej chwili zatrzymania t_k.

Zastanawiamy się nad tym, czy w przedziale $[t_1, t_2]$ pożądany stan procesu jest zmienny czy stały; w przykładzie zbiornika retencyjnego lub magazynu z rozdziału 3.2 stan w przedziale $[t_1, t_2]$ miał być zmienny, przy tym w odpowiedni sposób dostosowany do czynników zewnętrznych. Kiedy mogą wystąpić sytuacje przeciwne?

Załóżmy, że dla procesu opisanego równaniem (3.7) sformułowano zadanie optymalizacji jego biegu w postaci

$$\text{maksymalizować } Q = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} f_0\left(x(t), m(t), z(t)\right) \mathrm{d}t \tag{3.8}$$

co oznacza maksymalizację odpowiednio określonego zysku czy wydajności za okres $[t_1, t_2]$; aby uwolnić się od wpływu jego długości, we wzorze (3.8) zażądano maksymalizacji wartości średniej w tym okresie.

Zadanie opisane przez (3.7), (3.8), po dodaniu ewentualnych ograniczeń na stan $x(t)$ i sterowanie $m(t)$, należałoby rozwiązać, aby się przekonać czy zachodzi przypadek $\hat{x}(t) = \text{const}$, tj. przypadek, gdy stan optymalny procesu (oznaczany przez $\hat{x}$) jest wartością stałą. Przy rozwiązywaniu zadania (3.7), (3.8) dojdzie jeszcze kwestia wartości brzegowych $x(t_1)$, $x(t_2)$. Wartości te trzeba przyjąć za swobodne, aby mogło nastąpić $x(t_1) = \hat{x}(t_1)$ oraz $x(t_2) = \hat{x}(t_2)$, czyli by mogły one leżeć na trajektorii optymalnej.

Nie istnieją reguły ogólne, które na podstawie postaci zadania optymalizacji biegu procesu pozwalałyby stwierdzić, że rozwiązaniem będzie $\hat{x}(t) = \text{const}$. Można tylko powiedzieć, iż ma to miejsce dosyć często, w szczególności zaś w przypadkach gdy wydajność procesu jest determinowana przez jego stan, a równocześnie stan ten podlega nierównościowym ograniczeniom. Zobaczymy to nieco dalej na przykładach, teraz przyjmijmy, że mamy do czynienia z przypadkiem gdy rozwiązanie zadania dynamicznego ma właśnie postać $\hat{x}(t) = \text{const}$, czyli że proces ma być prowadzony ze stałością stanu.

Konkretną wartość $\hat{x}$ można w tej sytuacji wyznaczyć z zadania optymalizacji, które jest łatwiejsze do rozwiązania niż zadanie dynamiczne. Stałość stanu oznacza $\dot{x}(t) = 0$, zatem równanie różniczkowe (3.7) sprowadza się do więzów o charakterze statycznym

$$f(x, m(t), z(t)) = 0 \tag{3.9}$$

Przy takich więzach maksymalizacja wyrażenia całkowego (3.8) sprowadza się do żądania maksymalizacji funkcji podcałkowej

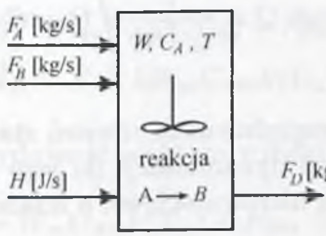
$$\text{maksymalizować } f_0(x, m(t), z(t)) \tag{3.10}$$

co łącznie z (3.9) oraz z ewentualnymi ograniczeniami nierównościowymi stanowi zadanie optymalizacji statycznej (programowania nieliniowego).

W zapisie (3.9), (3.10) zachowaliśmy $m(t)$, $z(t)$, czyli parametryczną zależność zadania od czasu. Nie czynimy bowiem założenia, że rozwiązaniu $\hat{x}(t) = \text{const}$ towarzyszy stałość sterowań i czynników zewnętrznych. Będzie raczej wręcz przeciwnie: utrzymanie $x(t) = \text{const}$ przy zmiennych w czasie wartościach $z(t)$ wymagać będzie odpowiedniego dostosowania $m(t)$ do tej zmienności. Stanie się to zadaniem układów sterowania (układów regulacji), mówić o tym będziemy w rozdziale 4.

Zadanie (3.9), (3.10) pozwala wyznaczyć optymalną wartość zmiennej x, czyli optymalny stan ustalony procesu dynamicznego, bez sięgania do optymalizacji dynamicznej; podkreślmy raz jeszcze, że opiera się to na stwierdzeniu bądź założeniu, że pierwotne dla całego problemu zadanie dynamiczne ma rozwiązanie $\hat{x}(t) = \text{const}$.

Rozpatrzmy teraz przykład sytuacji, o której była dotąd mowa. Na rysunku 3.6 pokazano schemat reaktora chemicznego, będącego naczyniem



Rys. 3.6. Schemat reaktora, w którym prowadzony jest proces ciągły

z mieszadłem, do którego dopływają strumienie dwóch substancji A oraz B. W reaktorze zachodzi przemiana chemiczna substancji A w substancję B; mieszanina – o wzbogaconej w stosunku do dopływu zawartości B – opuszcza reaktor jako ciągły wypływ F_D. Reakcja zachodzi w podwyższonej temperaturze i jest endotermiczna – dopływ ciepła oznaczono przez H.

Proces ma trzy współrzędne stanu, a jego równania – wyprowadzone na podstawie bilansu mas oraz ciepła – mają postać[*]:

$$\frac{\mathrm{d}W(t)}{\mathrm{d}t} = F_A(t) + F_B(t) - F_D(t) \tag{3.11}$$

$$\frac{\mathrm{d}C_A(t)}{\mathrm{d}t} = \frac{1 - C_A(t)}{W(t)} F_A(t) - \frac{C_A(t)}{W(t)} F_B(t) - C_A(t)k(T(t)) \tag{3.12}$$

$$\frac{\mathrm{d}T(t)}{\mathrm{d}t} = \frac{T_A(t) - T(t)}{W(t)} F_A(t) + \frac{T_B(t) - T(t)}{W(t)} F_B(t) + \frac{H(t)}{cW(t)} - \\ - \frac{h}{c} C_A(t)k(T(t)) \tag{3.13}$$

Jako współrzędne stanu wybrano tutaj zapełnienie reaktora W, mierzone w kg, udział wagowy substancji A w mieszaninie C_A, w kg/kg, czyli bez wymiaru, temperaturę mieszaniny T w kelwinach (jest ona miernikiem ilości nagromadzonego ciepła). Istotnym założeniem przy wyprowadzeniu tych równań było tzw. idealne mieszanie, to znaczy jednakowa temperatura i jednakowe stężenie w całej objętości reaktora. Wielkościami wejściowymi są: dopływ substancji A oznaczony F_A [kg/s], dopływ F_B [kg/s], wypływ F_D [kg/s], dopływ ciepła H [J/s], temperatury dopływających substancji T_A oraz T_B. Występują też parametry: stała prędkości reakcji k (załóżmy że rosnąca z temperaturą), ciepło właściwe c, ciepło reakcji h.

Przyjmijmy, że proces w reaktorze ma zapewnić maksimum wypływu substancji B, czyli maksimum wydajności, za rozpatrywany okres pracy ciągłej $[t_1, t_2]$. Żądanie to zapiszemy w postaci maksymalizacji produkcji średniej

$$\text{maksymalizować } Q = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} (1 - C_A(t))F_D(t)\mathrm{d}t \tag{3.14}$$

Wymaganie (3.14) z uwzględnieniem równań stanu (3.11), (3.12), (3.13) jest zadaniem optymalizacji dynamicznej; do jego sformułowania wprowadzimy jeszcze ograniczenia nierównościowe, a mianowicie:

– na zapełnienie reaktora $W(t) \leqslant W_m$ (3.15)

– na temperaturę w reaktorze $T(t) \leqslant T_m$ (3.16)

– na stężenie A w wypływie $C_A(t) \leqslant C_{Am}$ (3.17)

– na dopływ zewnętrzny substancji B $F_B(t) \leqslant F_{Bm}$ (3.18)

[*] W. Findeisen (1974). Wielopoziomowe układy sterowania. PWN, Warszawa, s. 17–18.

Ograniczenie na zapełnienie jest oczywiste, ograniczenie na temperaturę musi być uwzględnione aby nie otrzymać absurdalnego rozwiązania $T \to \infty$, ograniczenie zawartości substancji A w mieszaninie ma charakter wymagania jakościowego, ograniczenia dopływu substancji B zapobiega otrzymaniu odpowiedzi, że reakcja $A \to B$ w reaktorze jest niepotrzebna.

W zadaniu dynamicznym powinny jeszcze wystąpić stany początkowe $W(t_1)$, $C_A(t_1)$, $T(t_1)$ i odpowiednie stany końcowe. Uznamy, że wszystkie te wartości są swobodne.

Teraz należałoby sformułowane zadanie dynamiczne rozwiązać, w szczególności wyznaczyć optymalne przebiegi stanu $\hat{W}(t)$, $\hat{C}_A(t)$, $\hat{T}(t)$ w przedziale $[t_1, t_2]$, aby się przekonać, czy są to wartości stałe, czy zmienne. Przyjmijmy, że optymalne wartości stanu reaktora są stałe; wówczas obok równań stanu (3.11)–(3.13) pojawia się wymaganie

$$W(t) = \text{const}, \quad C_A(t) = \text{const}, \quad T(t) = \text{const} \tag{3.19}$$

równoznaczne wymaganiu zerowych wartości pochodnych. Sprowadza to równania (3.11)–(3.13) do więzów algebraicznych, a nie różniczkowych, a zadanie określenia optymalnych wartości stanu $\hat{W}$, $\hat{C}_A$, $\hat{T}$ staje się zadaniem optymalizacji statycznej.

Zadanie to daje się rozwiązać analitycznie, a jego rozwiązanie ma postać

$$\begin{aligned}
&\hat{W} = W_m \\
&\hat{C}_A = C_{Am} \\
&\hat{T} = T_m \\
&\hat{F}_A = C_{Am}(1 - C_{Am})^{-1} W_m k(T_m) + F_{Bm} \\
&\hat{F}_B = F_{Bm} \\
&\hat{F}_D = (1 - C_{Am})^{-1}\left[W_m C_{Am} k(T_m) + F_{Bm}\right] \\
&c\hat{F}_A\hat{T}_A + cF_{Bm}\hat{T}_B + \hat{H} = hW_m C_{Am} k(T_m) + c\hat{F}_D T_m
\end{aligned} \tag{3.20}$$

przy czym optymalna wydajność reaktora wynosi

$$\hat{Q} = W_m C_{Am} k(T_m) + F_{Bm} \ [\text{kg/s}] \tag{3.21}$$

gdzie $k(T_m)$ jest wartością stałej prędkości reakcji przy $T = T_m$. Pierwszy składnik w sumie (3.21) jest wydajnością reakcji, drugi – przeniesieniem dopływu substancji B na wyjście reaktora.

Otrzymane rozwiązanie ma pewne cechy charakterystyczne, którym warto się przyjrzeć. Optymalne wartości stanu $\hat{W}$, $\hat{C}_A$, $\hat{T}$ leżą na brzegach swoich obszarów dopuszczalnych, to samo odnosi się do dopływu $\hat{F}_B$. Spośród pozostałych wartości wejściowych przepływy optymalne $\hat{F}_A$, $\hat{F}_D$ są określone jednoznacznie, zatem ich ewentualna zmienność w czasie – nawet jeśli nie naruszymy stałości stanu reaktora – oznaczać będzie odejście od sytuacji optymalnej, spadek wydajności. Natomiast pozostałe trzy zmienne wejściowe mają spełnić jeden warunek, wiążący $\hat{H}$, $\hat{T}_A$, $\hat{T}_B$. Oznacza to, że można dopuścić zmienność w czasie tych trzech wielkości, byle zachować przepisane równanie. Fizycznie rzecz biorąc chodzi tu o podtrzymanie bilansu ciepła. Projektując układ sterowania można dopuścić T_A, T_B jako zakłócenia, a ich wpływ kompensować za pomocą sterowania dopływem ciepła H. Rozwiniemy ten wątek przykładu w rozdz. 4.3, mówiąc o strukturach sterowania.

Przedstawiony przykład wydaje się dobrze ilustrować istotę rzeczy, to jest typ sytuacji kiedy pożądany jest bieg procesu w stanie ustalonym. Chodziło mianowicie o maksimum pewnej wydajności, wydajność ta zależała od stanu procesu, stan podlegał ograniczeniom – należało całkowicie wykorzystać istniejące możliwości.

W omówionym przykładzie postępowaliśmy według schematu, który składał się z następujących etapów:

a) *opisanie fizyczne obiektu i procesu:* należało zdać sobie sprawę, co się w danym obiekcie dzieje, co jest – ogólnie biorąc – celem jego działania;
b) *sporządzenie opisu matematycznego procesu:* sformułowanie, choćby przybliżonych, zależności między występującymi w procesie wielkościami (w tym – zdanie sobie sprawy ze współrzędnych określających stan procesu);
c) *określenie pożądanego przebiegu procesu:* sformułowanie warunków, które ma spełnić bieg procesu, aby zadowolić użytkownika; jeśli to możliwe, wyznaczenie *przebiegu optymalnego.*

Po tych trzech etapach powinien nastąpić etap czwarty, a mianowicie

d) *zaprojektowanie układu sterowania procesem* – znalezienie odpowiedzi na pytanie, jak wywołać i jak podtrzymać bieg optymalny procesu, zdefiniowany w etapie c). Powrócimy do tego w rozdziale 4.

3.3. Procesy jednorazowe i cykliczne

Jest wiele sytuacji, gdzie proces podlegający sterowaniu rozpatrywany jest jako proces jednorazowy: przepłynięcie statku z portu A do B, kampania cukrownicza jednego sezonu, proces budowy mostu czy wznoszenia budynku, pojedynczy wytop stali w piecu stalowniczym i tak dalej. Zauważmy, że „jednorazowość” jest tutaj sprawą punktu widzenia, perspektywy czasowej obserwatora czy decydenta. Jeden rejs statku jest fragmentem procesu realizowanego przez przedsiębiorstwo żeglugowe, przedsiębiorstwo budowlane wznosi budynki jeden po drugim (a także równolegle), stalownia produkuje stal przez cały tydzień, miesiąc albo rok.

Proces jednorazowy z reguły jest procesem o stanie zmiennym w czasie; jeśli jego opis ma postać równania różniczkowego

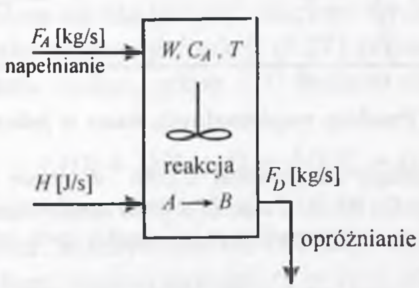
$$\dot{x}(t) = f\left(x(t), m(t), z(t)\right) \tag{3.22}$$

to zazwyczaj charakterystyczne tu będzie, iż $x(t_k) \neq x(t_0)$, przy czym chwila początkowa t_0, chwila końcowa t_k oraz wartość stanu początkowego $x(t_0)$ i końcowego $x(t_k)$ mogą być zarówno dane, jak swobodne; bywają też ograniczone do pewnych zbiorów, na przykład w postaci wymagania $x(t_k) \in X_k$.

Przykładowo, proces płynięcia statku z portu A do B ma zadany stan początkowy (punkt A, zerowa prędkość) i końcowy (punkt B, zerowa prędkość), zadaną chwilę początkową t_0 (nie miał jej żaglowiec, który czekał na wiatr), chwilę końcową t_k swobodną w pewnych granicach. Kampania cukrownicza ma swobodną chwilę końcową (kończy się, gdy przerobimy wszystkie buraki warte kosztu przeróbki) oraz swobodną chwilę początkową (niekiedy warto zacząć przerabiać buraki niezupełnie dojrzałe, gdyż straty cukru w burakach składowanych do końca kampanii mogą być większe).

Proces zwany *cyklicznym* (lub *periodycznym*) powstaje przez powtarzanie procesu jednorazowego, stając się w ten sposób odpowiednikiem procesu ciągłego – zwłaszcza wtedy, gdy nie mówimy o „pierwszym cyklu" i „ostatnim cyklu", czyli o rozruchu i zatrzymaniu ciągu cykli, lecz o jego biegu w przedziale pośrednim (por. rozdz. 3.1).

Rozpatrywanie procesu cyklicznego trzeba zacząć od pojedynczego cyklu; zrobimy to, wykorzystując przykład reaktora z rys. 3.6, zakładając tym razem, że naczynie reaktora jest w pewnej chwili t_1 ładowane substancją A do pełna, to jest do $W = W_m$, w reaktorze następuje przemiana $A \to B$, a w chwili $t_2 - \delta$, gdy skład mieszaniny osiągnie stan $C_A = C_{Am}$ reaktor zostaje opróżniony. Schematycznie przedstawia ten sposób pracy rys. 3.7.



Rys. 3.7. Reaktor z rys. 3.6 pracujący cyklicznie

Przyjmiemy, dla uproszczenia, że temperatura mieszaniny jest utrzymywana z zewnątrz na poziomie T = const = T_m (wiemy z rozdz. 3.2, że jest to temperatura optymalna), oraz że czas ładowania bądź rozładowywania wynosi δ, przy czym w czasie tych operacji reakcja chemiczna nie zachodzi.

Proces w reaktorze będzie przy tych założeniach opisany równaniami (por. wzory (3.11), (3.12)):
w przedziale $t \in [t_1, t_1 + \delta]$

$$\frac{\mathrm{d}W(t)}{\mathrm{d}t} = F_A(t), \quad C_A(t) = 1 \tag{3.23}$$

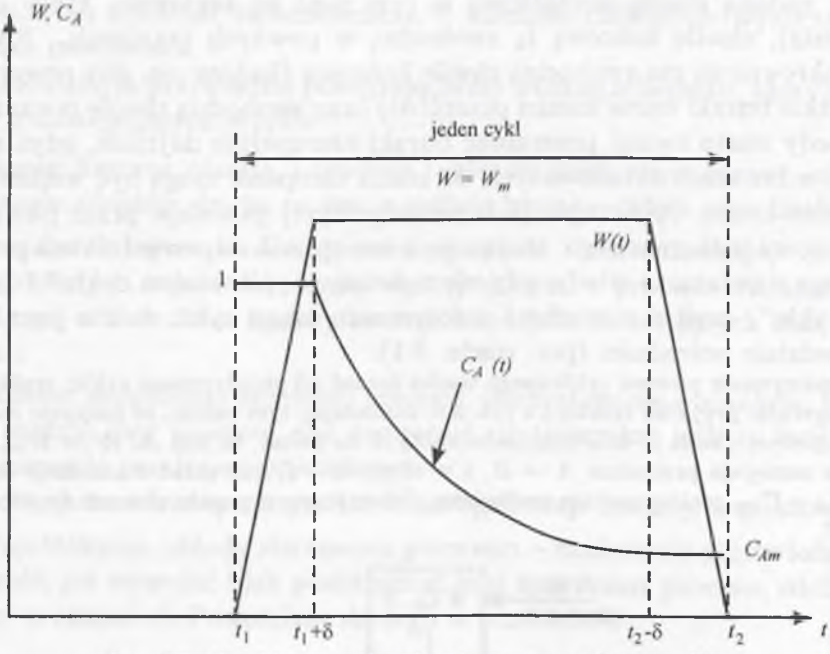
w przedziale $t \in [t_1 + \delta, t_2 - \delta]$

$$\frac{\mathrm{d}C_A(t)}{\mathrm{d}t} = C_A(t)k(T_m), \quad W(t) = W_m \tag{3.24}$$

w przedziale $t \in [t_2 - \delta, t_2]$

$$\frac{\mathrm{d}W(t)}{\mathrm{d}t} = -F_D(t), \quad C_A(t) = C_{Am} \tag{3.25}$$

Przebieg w czasie współrzędnych stanu reaktora $W(t)$, $C_A(t)$ przedstawiono na rys. 3.8; rysując przebieg $W(t)$ założono, że natężenie dopływu w okresie ładowania i natężenie odpływu w okresie rozładowania są stałe i równe W_m/δ.



Rys. 3.8. Przebieg współrzędnych stanu w jednym cyklu

Zwróćmy szczególną uwagę na przebieg $C_A(t)$: w czasie trwania reakcji, tj. dla $t \in [t_1 + \delta, t_2 - \delta]$ mamy $C_A(t) \geqslant C_{Am}$. Produkt opuszczający reaktor ma ten sam skład zadany $C_A = C_{Am}$ co w przypadku pracy ciągłej (por. rozdz. 3.2), a jaka będzie wydajność reaktora?

Wzór (3.20) określił, jako wydajność przemiany przy pracy ciągłej, wartość $W_m C_{Am} k(T_m)$ [kg/s]. W obecnym przypadku, wobec zmiennego w czasie C_A, wydajność wyrażona w sposób podobny, to jest jako wartość średnia za jeden pełny cykl, wyniesie

$$Q = \frac{1}{t_2 - t_1} \int_{t_1+\delta}^{t_2-\delta} W_m k(T_m) C_A(t)\mathrm{d}t = W_m k(T_m) \frac{1}{t_2 - t_1} \int_{t_1+\delta}^{t_2-\delta} C_A(t)\mathrm{d}t$$

We wzorze tym występuje wartość średnia $C_A(t)$; wartość ta, wobec $C_A(t) \geqslant C_{Am}$, jest oczywiście większa niż C_{Am}. Rysunek 3.8 pozwala ocenić, że przewaga $C_{A\acute{s}r}$ nad C_{Am} jest tym większa, im mniejsze jest C_{Am}, gdyż przebieg $C_A(t)$ zaczyna się zawsze od $C_A = 1$. Oznacza to, że przewaga wydajności przy pracy cyklicznej nad wydajnością przy pracy ciągłej jest tym większa, im bardziej „czysty" ma być otrzymywany z reaktora produkt (małe C_{Am}).

Przewaga wydajności pracy cyklicznej nad ciągłą jest zmniejszona przez to, że w czasie ładowania i rozładowania reakcja nie zachodzi: długość cyklu wynosi $t_2 - t_1$, a czas trwania reakcji $t_2 - t_1 - 2\delta$. Ważniejsze bywają inne niedogodności; dopływ musi pochodzić z jakiegoś zbiornika, produkt – otrzymywany w dużych dawkach, a nie w sposób ciągły – musi być magazynowany. Zrozumiała staje się dążność do tego, by procesy produkcyjne kształtować raczej jako procesy ciągłe.

Wspominaliśmy przy końcu rozdziału 3.2 o procesach ciągłych, które zyskują na prowadzeniu ich ze zmiennością stanu pomimo braku wpływów zewnętrznych.

Myśl o przydatności zmian stanu, mówiąc konkretnie – zmian okresowych – nasuwają procesy cykliczne, omawiane w tym rozdziale. Skoro bowiem proces cykliczny prowadzony w reaktorze o pojemności W_m ma większą wydajność niż prowadzony w tym reaktorze proces ciągły, to być może można połączyć zalety jednego i drugiego sposobu prowadzenia procesu: utworzyć proces ciągły o niektórych cechach okresowo zmiennych.

W zapisie formalnym mamy proces dynamiczny ciągły bez czynników zewnętrznych

$$\dot{x}(t) = f\left(x(t), m(t)\right) \tag{3.26}$$

prowadzony w pewnym przedziale czasu $t \in [t_1, t_2]$, z dala od chwili rozruchu t_0 i chwili końcowej t_k, oraz funkcję celu

$$Q = \frac{1}{t_2 - t_1} \int\limits_{t_1}^{t_2} f_0\left(x(t), m(t)\right) \mathrm{d}t \tag{3.27}$$

Poprzednio staraliśmy się dla procesu ciągłego wydzielić sytuacje, gdy rozwiązanie zadania maksymalizacji funkcji (3.27) prowadziło do $\hat{x}(t) = \mathrm{const}$; teraz myślimy o klasie zadań, gdzie $\hat{x}(t)$ będzie okresowe, czyli spełniać będzie zależność

$$\hat{x}\left((k+1)T_c + t\right) = \hat{x}(kT_c + t)$$

gdzie T_c jest okresem (długością cyklu). Może oczywiście istnieć wartość optymalna tego okresu. Korzystając ze swobody wyboru t_1, t_2 w odniesieniu do procesu ciągłego, można przyjąć $t_2 = t_1 + NT_c$, przez co zachodzić będzie równość stanów brzegowych $x(t_2) = x(t_1)$. Optymalna wartość stanu na początku każdego okresu $\hat{x}(t_1) = \hat{x}(t_1 + nTc)$, $n = 1, 2, \ldots$ pozostawałaby do wyznaczenia.

Zadanie wyznaczenia optymalnego przebiegu $x(t)$ i odpowiadającego mu sterowania $m(t)$ sprowadza się zatem do pojedynczego okresu

$$\text{maksymalizować } Q = \frac{1}{T_c} \int_0^{T_c} f_0\left(x(t), m(t)\right) \mathrm{d}t \tag{3.28}$$

przy równaniach procesu (3.26) oraz przy warunku $x(0) = x(T_c)$, przy czym zarówno $x(0)$ jak T_c są swobodne.

Bliższe zdefiniowanie klasy zadań, w których przy braku wpływów zewnętrznych rozwiązanie optymalne $\hat{x}(t)$, $t \in [t_1, t_2]$ jest okresowe, wydaje się trudne.

3.4. Horyzont sterowania procesami

Przebiegi pożądane procesów, które mają podlegać sterowaniu, trzeba rozpatrywać w pewnym horyzoncie czasowym; powstaje ważne dla praktyki pytanie, jaki horyzont jest właściwy.

Przypomnijmy przykład z rozdz. 3.2 – zbiornik retencyjny na rzece. Sformułowaliśmy tam wymagania, że stan zbiornika nie powinien wykraczać poza pewne granice, $x_{\min} \leqslant x(t) \leqslant x_{\max}$, a ponadto wypływ ze zbiornika także ma być ograniczony, $y_{\min} \leqslant y(t) \leqslant y_{\max}$. Widzieliśmy w owym przykładzie, że podjęcie decyzji o zrzucie, czyli wypływie wody w danej chwili t wymaga wzięcia pod uwagę przyszłości procesu, to znaczy tego, czy w przyszłości – w wyniku losowego dopływu wody do zbiornika – nie nastąpi przekroczenie jednego lub drugiego ograniczenia. Powstaje pytanie: jak daleką przyszłość trzeba rozpatrywać w chwili podejmowania decyzji o sterowaniu?

Jest to pytanie o *horyzont sterowania* – przedział czasu, w którym rozpatrujemy przebieg stanu procesu oraz przebieg sterowania, przy czym początkiem tego przedziału jest chwila obecna.

Rozróżnić należy dwie zasadnicze kategorie zadań:

1) zadania sterowania, których horyzont jest narzucony bezpośrednio lub przez stan końcowy,
2) zadania o nie narzuconym horyzoncie.

Pierwsza kategoria obejmuje te zadania, w których chwila końcowa t_k oraz/lub stan (zbiór) końcowy $x(t_k)$ są dane, a ponadto trzeba to brać pod uwagę przy wyznaczaniu sterowania już od pierwszej chwili. Będzie to więc na przykład przepłynięcie statku z portu A do B, przeprowadzenie jednego wytopu stali w piecu stalowniczym, wybudowanie jednego budynku.

Druga kategoria to takie zadania, w których chwila końcowa jest na tyle odległa w czasie, że stan końcowy nie ma wpływu na interesujący nas odcinek pożądanej trajektorii stanu procesu. Należą tu przede wszystkim różnego rodzaju procesy ciągłe, jak wspomniany przed chwilą proces retencji wody, procesy w instalacjach chemicznych, proces wytwarzania energii w systemie elektroenergetycznym. To prawda, że instalacja chemiczna będzie kiedyś wyłączona z ruchu, ale jeśli chwila ta jest bardzo odległa, to sposób przyszłego zatrzymania nie ma wpływu na sterowanie bieżące. W konsekwencji, potrzebny praktycznie horyzont sterowania możemy określić na podstawie innych przesłanek.

Rozpatrzmy to nieco ściślej, interesując się szczególnie możliwością *skrócenia horyzontu sterowania*, czyli skrócenia przedziału czasu, który trzeba rozpatrywać w chwili podejmowania decyzji o sterowaniu.

Załóżmy, że mamy, jak zwykle, proces dynamiczny, opisany równaniem

$$\dot{x}(t) = f\left(x(t), m(t), z(t)\right) \tag{3.29}$$

a zadanie brzmi

$$\text{maksymalizować } Q = \frac{1}{t_k - t_0} \int_{t_0}^{t_k} f_0\left(x(t), m(t), z(t)\right) \mathrm{d}t \tag{3.30}$$

gdzie t_0 jest chwilą obecną, a t_k jest odległą w czasie chwilą końcową (chwilą zatrzymania procesu). Stan obecny procesu jest znany i równy $x(t_0)$.

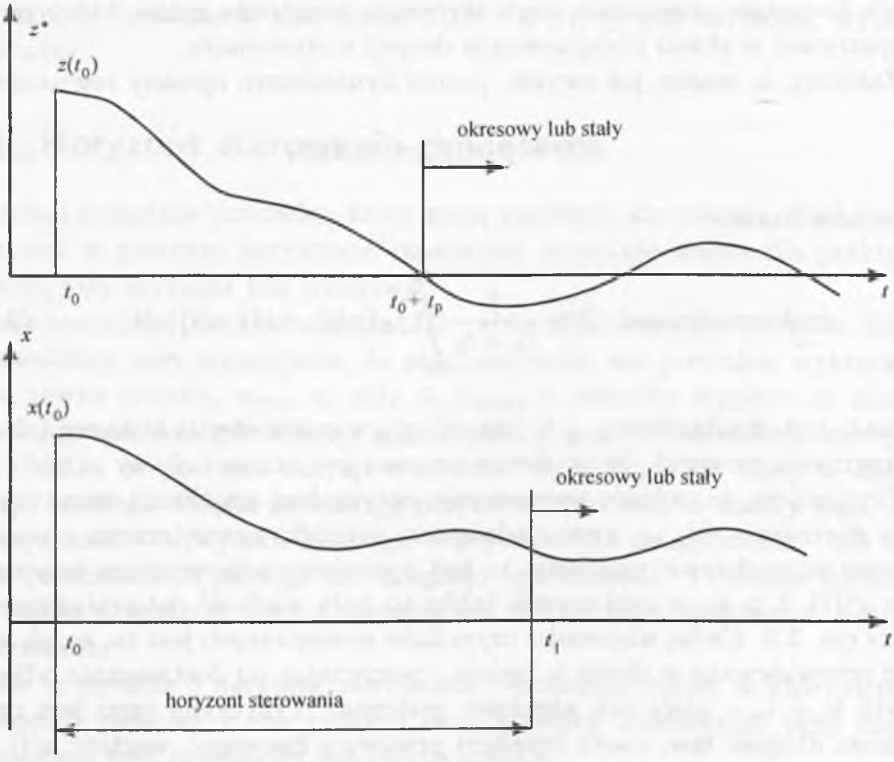
Przyjmijmy, że zadanie wyznaczenia optymalnej trajektorii stanu $\hat{x}_{[t_0,t_k]}$ oraz sterowania $\hat{m}_{[t_0,t_k]}$ z uwzględnieniem czynnika zewnętrznego z rozwiązujemy na podstawie prognozy, to jest operujemy odpowiednim estymatorem $z^*(t)$, $t > t_0$, w taki sposób jakby to była wielkość deterministyczna, patrz rys. 3.9. Cechą większości czynników zewnętrznych jest to, że ich wartość przewidywana w chwili t_0 będzie – poczynając od dostatecznie odległej chwili $t_0 + t_p$ – *stała* lub *okresowo zmienna*. Przyczyną tego jest ograniczona długość tzw. *czasu korelacji* przebiegu losowego: wartość $z(t)$ dla $t \geqslant t_0 + t_p$ przestaje być skorelowana z wartością $z(t_0)$. Znając $z(t_0)$ i ewentualnie także wartości przeszłe nie potrafimy powiedzieć o wartości $z(t_0 + t_p)$ i wartościach późniejszych nic ponad to, co wynika z probabilistycznych właściwości danego przebiegu losowego. Ewentualna zmienność okresowa estymatora pojawia się tutaj jako rezultat istnienia składowej okresowej w przebiegu losowym (przykładem może być obciążenie systemu elektroenergetycznego w obrębie doby albo tygodnia); jeśli składowej okresowej nie ma, estymator $z^*(t)$ dla $t \geqslant t_0 + t_p$ jest wartością stałą.

Przebiegowi $z^*_{[t_0,t_k]}$ traktowanemu jako działająca na proces zewnętrzna wielkość wejściowa odpowiadać będzie wynikający z zadania (3.30) przebieg trajektorii stanu $\hat{x}_{[t_0,t_k]}$ oraz sterowania $\hat{m}_{[t_0,t_k]}$. W trajektorii stanu $\hat{x}_{[t_0,t_k]}$ można będzie wyróżnić trzy odcinki:

- odcinek początkowy, $t \in [t_0, t_1]$, gdzie kształt trajektorii zależy od stanu obecnego $x(t_0)$ oraz obecnej wartości czynnika zewnętrznego $z(t_0)$,

– odcinek środkowy, $t \in [t_1, t_2]$, gdzie kształt trajektorii jest podyktowany przez zadanie sterowania, ale nie zależy ani od stanu obecnego $x(t_0)$ i wartości początkowej $z(t_0)$, ani od stanu w chwili końcowej $x(t_k)$*),

– odcinek końcowy, $t \in [t_2, t_k]$, gdzie kształt trajektorii uzależniony jest od wymagania, że w chwili $t = t_k$ proces ma być zakończony w stanie $x(t_k)$ (stan ten może być swobodny).



Rys. 3.9. Estymator z^* czynnika losowego oraz trajektoria stanu obliczana w chwili t_0

Możemy to zilustrować przykładem samolotu, który odbywa rejs z lotniska A do lotniska B. Przebieg startu i początkowego wznoszenia się można rozpatrywać oddzielnie, to jest z krótkim horyzontem, jeżeli podamy na jaką trajektorię dalszego lotu powinniśmy wejść. Jeśli tej trajektorii (obliczonej oddzielnie) nie podamy, to musimy rozpatrywać od razu zadanie na cały horyzont, zadając punkt docelowy B. W przeciwnym razie przy starcie nie będzie wiadomo, w jakim kierunku i na jaką wysokość mamy się wznieść.

Przypomnijmy, że rozpatrując przebiegi pożądane procesów ciągłych w rozdziale 3.2 analizowaliśmy środkowy odcinek trajektorii. Widzieliśmy wówczas, że istnieją przypadki, gdy mimo zmienności z, optymalny przebieg stanu ma postać $\hat{x}(t) = \text{const}$. W innych przypadkach, gdy z ma wpływ na

*) Ściśle rzecz biorąc, nie ma tak odległej chwili t, dla której nie istniałby już wpływ stanu początkowego $x(t_0)$ na obliczaną trajektorię optymalną – rzecz w tym, że dla $t \gg t_0$ wpływ ten jest znikomo mały.

trajektorię optymalną, przy okresowo zmiennym z musiałoby wystąpić tak samo okresowo zmienne $\hat{x}$; okresowo zmienne $\hat{x}$ może także wystąpić jako rozwiązanie zadania bez czynnika zewnętrznego, czyli przy $z = 0$.

Odnosząc się teraz do rozpatrywanej sytuacji, to jest do obliczenia dokonywanego w chwili t_0 z użyciem estymatora z^*, który dla $t \geqslant t_p$ jest stały lub okresowy, nie potrafimy dla *odległego* przedziału czasu $[t_1, t_2]$ – gdzie $t_1 > t_p$ – obliczyć nic innego jak trajektorię stałą lub okresową. Podkreślmy, że wcale to nie oznacza, że taka ona będzie lub być powinna – po prostu w chwili t_0 nie mamy dostatecznej informacji o wartościach zewnętrznych, jakie w przedziale $[t_1, t_2]$ rzeczywiście wystąpią.

Na rysunku 3.9 chwila t_1 pojawia się później niż chwila t_p. Zawarta jest w tym myśl o wpływie dynamiki samego procesu: obliczany w chwili t_0 stan optymalny nie będzie stały (lub okresowy) już w chwili $t_0 + t_p$, kiedy wielkość zewnętrzna zaczęła być stała (okresowa), lecz nieco później. Odcinek czasu $[t_0 + t_p, t_1]$ jest podobny do tego, który charakteryzowałby czas trwania wpływu stanu początkowego na przebieg trajektorii optymalnej procesu o długim horyzoncie czasowym przy braku czynnika zewnętrznego.

Wniosek z naszych dotychczasowych rozważań jest taki, że rozpatrując w chwili t_0 sterowanie procesem dynamicznym przy użyciu estymatora czynnika zewnętrznego $z^*_{[t_0, t_k]}$ wystarczy brać pod uwagę horyzont czasowy $[t_0, t_1]$, gdzie t_1 jest początkiem przedziału, w którym *obliczana w chwili* t_0 pożądana trajektoria stanu będzie już stała (lub okresowa). Obliczenie trzeba jednak rozpocząć od przedziału $[t_1, t_2]$, a więc od wyznaczenia owej wartości stałej (lub przebiegu okresowego), gdyż wzięta stamtąd wartość optymalna $\hat{x}(t_1)$ będzie zarazem *zadanym stanem końcowym* dla odcinka trajektorii $[t_0, t_1]$.

Mamy więc w istocie dwa zadania:

- wyznaczyć rozwiązanie dla przedziału $[t_1, t_2]$ przy stałej (lub okresowej) wartości $z_{[t_1, t_2]}$ w równaniu (3.29) oraz swobodnych wartościach brzegowych $x(t_1)$, $x(t_2)$,
- rozwiązać zadanie dynamiczne właściwe, z warunkiem początkowym $x(t_0)$, dla horyzontu $[t_0, t_1]$, przy założeniu przebiegu $z_{[t_0, t_1]}$ równego estymatorowi $z^*_{[t_0, t_1]}$.

Przeprowadzone rozważania stanowią w istocie poszukiwania możliwie racjonalnej i umotywowanej odpowiedzi na pytanie: jak daleką przyszłość trzeba brać pod uwagę w chwili obecnej t_0, jeśli mamy do czynienia z procesem, którego ewentualne zatrzymanie jest bardzo odległe? Przenieśmy to na przypadek zbiornika retencyjnego na rzece: czy rozpatrując pożądany przebieg stanu oraz potrzebne do tego sterowanie mamy myśleć o przyszłości odległej o miesiąc, rok, dwadzieścia czy pięćdziesiąt lat? Rozważanie nasze wskazywało, że horyzont czasowy podejmowanej w chwili t_0 decyzji może zależeć od dwóch czynników: *właściwości czynnika zewnętrznego* (to jest od długości czasu korelacji $z(t)$ z wartością $z(t_0)$) oraz *dynamiki obiektu*, czyli – w przykładzie – od korelacji dopływów oraz od stosunku pojemności zbiornika do natężenia przepływów.

Jedna jeszcze uwaga będzie tu na miejscu: mówiliśmy przez cały czas o decyzji podejmowanej w chwili obecnej t_0, uwzględniającej wartości obecnie obserwowane $x(t_0)$, $z(t_0)$. W chwili następnej, $t_0 + \delta$, mieć będziemy do dyspozycji nowe, aktualne wartości stanu x i czynnika zewnętrznego z. Na podstawie $z(t_0 + \delta)$ można ustalić nową, lepszą prognozę $z^*_{[t_0+\delta, t_1+\delta]}$, stan $x(t_0 + \delta)$ można potraktować jako nowy stan początkowy i podstawę do aktualizacji obliczenia pożądanej trajektorii stanu i potrzebnego przebiegu sterowania – teraz już na horyzont, sięgający chwili $t_1 + \delta$.

Opisany schemat postępowania jest podstawą działania tzw. repetycyjnych układów sterowania (w terminologii zarządzania – podstawą planowania kroczącego). Wspomnimy o tym jeszcze w rozdz. 5.6.

Powróćmy do kwestii ustalenia niezbędnego horyzontu sterowania w aspekcie praktycznym – można to mianowicie ująć następująco: jeśli przyjmiemy w obliczeniu jakiś horyzont $[t_0, t_1]$, a potem stwierdzimy (w ponownym obliczeniu) że jego skrócenie nie zmienia początkowego odcinka rozwiązania (a tylko ten jest nam teraz potrzebny), to horyzont skrócony jest wystarczający.

Na ogół nie można odstąpić od uwzględnienia w obliczeniu odpowiedniego stanu końcowego, to jest stanu przewidzianego na koniec horyzontu. Można to objaśnić za pomocą znanego już przykładu zbiornika retencyjnego: jeśli będziemy rozwiązywali zadanie gospodarowania wodą z horyzontem jednego miesiąca, czy nawet roku, ze swobodnym stanem końcowym, to obliczenie wskaże jako optymalny stan końcowy zerowe napełnienie. Potrzebny jest zatem w sformułowaniu zadania jakiś mechanizm, który nakaże zostawić pewien zapas wody na okres następny – mechanizmem tym jest narzucenie (w sposób bezpośredni lub pośredni) stanu końcowego różnego od zera. Dopiero przy bardzo długim horyzoncie (np. 10 lat dla zbiornika pracującego w cyklu rocznym) swobodny stan końcowy nie zniekształci obliczenia początkowego odcinka trajektorii.

Rozdział 4

Sterowanie w podziale na warstwy

4.1. Powstawanie struktur warstwowych

Zapewnienie pożądanego biegu procesów w obiektach systemu, a w ślad za tym – w całym systemie, wymaga ciągłego lub prawie ciągłego oddziaływania na wszystkie wielkości sterujące. Podkreślaliśmy już w rozdziale 1, nawiązując do przestrzennych i czasowych rozmiarów zadań sterowania, że sterowanie bieżące systemem wymaga podejmowania wielkiej liczby decyzji, ale że nie muszą te decyzje wynikać bezpośrednio z zadania postawionego przed całym systemem; prowadziło to do różnych koncepcji podziału funkcji sterowania pomiędzy wiele jednostek decyzyjnych.

W rozdziale 2 przedstawiliśmy podstawowe cechy i zasady działania struktur, których budowa związana była z przestrzennymi rozmiarami zadania sterowania, z jego rozległością w sensie liczby wielkości sterujących oraz przede wszystkim z tym, że w sterowanym systemie można było sensownie wyodrębnić poszczególne podsystemy czy obiekty. Mówiliśmy tam o *rozproszeniu kompetencji* oraz o *koordynacji działania* zdecentralizowanych jednostek decyzyjnych.

Warstwowy podział sterowania, będący – jak mówiliśmy w rozdziale 1.4 – jedną z dwóch podstawowych koncepcji w sterowaniu hierarchicznym, nawiązuje głównie do rozmiarów czasowych, a nie przestrzennych; u podłoża struktury warstwowej leży przekonanie, że podejmowane w sposób ciągły lub prawie ciągły decyzje o bieżącym oddziaływaniu na proces, czyli decyzje sterowania bieżącego, nie muszą być oparte w sposób bezpośredni na zadaniu postawionym przed tym procesem w pełnym horyzoncie czasowym. Można je podporządkować odpowiednio sformułowanym zadaniom prostszym, takim jak: stabilizacja, nadążanie, dojście do określonego punktu, wykonanie określonej operacji, a dopiero te zadania prostsze formułować w nawiązaniu do zadania całościowego.

Powstaje w ten sposób w strukturze sterowania *warstwa sterowania bezpośredniego*, od której oczekuje się oddziaływania na procesy w systemie „na bieżąco", ale którą chciało by się ukształtować na zasadach możliwie prostych. Nie jest to tylko kwestia kosztu lub racjonalności technicznej; sterowanie bieżące wymaga odniesienia się do wszystkich szczegółów dynamiki procesu, uwzględnienia wszystkich zachodzących w nim akumulacji. Powstająca przez to duża wymiarowość, to jest duża liczba współrzędnych stanu, nie pozwala na to, by rozwiązywać w sposób bezpośredni zadania bardziej złożone.

Układy sterowania składające się na warstwę bezpośrednią mieć będą zazwyczaj następujące cechy charakterystyczne:

– sterowanie bezpośrednie korzysta ze *sprzężenia zwrotnego*, to jest z obserwacji procesu sterowanego i wyniku oddziaływania sterującego,
– przekształcanie obserwacji w sterowanie następuje za pomocą możliwie prostego *prawa sterowania*,
– działanie układu sterowania jest *ciągłe* lub *prawie ciągłe*, czyli częste w stosunku do zmienności zachodzących zjawisk – jest to bowiem sterowanie odnoszące się do *dynamiki procesu*,
– istotnym celem sterowania bieżącego jest *eliminacja wpływu* nie mierzonych i trudno prognozowalnych *czynników zewnętrznych* oraz *niedokładności modelu procesu sterowanego*,
– sterowanie bezpośrednie jest zazwyczaj *zdecentralizowane*, składa się z wielu równoległych układów, z których każdy steruje procesem o małej złożoności,
– *implementacja* układów sterowania bezpośredniego w zastosowaniu do procesów technologicznych ma zazwyczaj postać *sterowania automatycznego*.

Przez *prawo sterowania* rozumiemy takie postępowanie przekształcające obserwacje w sterowanie, które daje się sformułować w postaci określonej reguły, to jest funkcji, tabeli lub algorytmu, według którego przetwarzane będą obecne i przeszłe obserwacje procesu oraz/lub jego otoczenia w bieżące oddziaływanie sterujące. Przeciwieństwem jest wyznaczanie sterowania przy użyciu procedury z jawnie wprowadzoną funkcją celu, ograniczeniami, modelem procesu, prognozą oddziaływania zewnętrznego itd., co jest właściwe raczej dla interwencji podejmowanych w większych odstępach czasu (mówimy o tym nieco szerzej w rozdziale 5.6). Istnieje wiele dróg i metod projektowania prawa sterowania, nie koniecznie prowadzących do tych samych wyników. Dzieje się tak dlatego, że projektant musi sobie zazwyczaj sam sformułować pewną liczbę dodatkowych założeń, aby zadanie wyrażone na przykład jako „utrzymanie stałości wielkości regulowanej" doprowadzić do jednoznacznego i technicznie sensownego rozwiązania.

Układy sterowania bezpośredniego zapanują nad dynamiką procesu, wyeliminują oddziaływanie otoczenia, spełnią postawione przed nimi zadania (np. stabilizacji lub nadążania); rzecz teraz w tym, że zadania te muszą być podporządkowane celom działania systemu sterowanego w taki sposób, by ich realizacja dawała w wyniku pożądany bieg procesu. Nie ma już bowiem innego dostępu do procesów w systemie jak dostęp pośredni, przez układy stanowiące pierwszą warstwę sterowania. Mówimy więc teraz o *warstwie wyznaczania zadań.*

W istocie rzeczy, w całym rozdziale 2 przyjęliśmy milcząco, że istnieje już co najmniej warstwa sterowania bezpośredniego, nie oznaczaliśmy tam bowiem wielkości sterujących przez m, lecz przez c. Litera ta jest w całej książce odnoszona do wielkości sterujących pośrednich. Struktury wielopoziomowe omawiane w rozdziale 2 były zatem strukturami, w których wyznaczano zadania dla niższej warstwy sterowania, a nie sterowania bezpośrednie dla procesu.

Problem wyznaczania zadań dla sterowania bezpośredniego najłatwiej objaśnić na przykładzie, gdy przebieg procesu sterowanego można ocenić przy użyciu skalarnej funkcji celu (oznaczonej przez Q): pożądany jest przebieg, przy którym wartość tej funkcji osiąga maksimum. Do dyspozycji są wielkości sterujące procesu (oznaczone przez m); one zatem są zmiennymi decyzyjnymi w podstawowym, pierwotnym zadaniu optymalizacji. Można postawić pytanie, czy nie da się przeformułować zadania optymalizacji przebiegu procesu w taki sposób, by jego rozwiązanie określało nie wielkości sterujące, lecz zadania dla układów sterowania bezpośredniego (wyrażone poprzez zmienne c), a ponadto – co najważniejsze – by realizacja zadań sterowania bezpośredniego powodowała powstawanie tych właśnie wartości wielkości sterujących, które są rozwiązaniem pierwotnego zadania optymalizacji procesu.

Mówiąc skrótowo, zadanie pierwotne polegające na optymalizacji względem m

$$\underset{m}{\text{maksymalizować}}\ Q \tag{4.1}$$

z ograniczeniami wynikającymi z równań procesu chcemy zastąpić odpowiednim nowym zadaniem, względem c

$$\underset{c}{\text{maksymalizować}}\ Q \tag{4.2}$$

z ograniczeniami wynikającymi z równań procesu oraz zastosowanych układów sterowania bezpośredniego. Narzucenie tym układom wartości zmiennych $c = \hat{c}$ wynikających z rozwiązania zadania (4.2) powinno wywołać zaistnienie sterowań $m = \hat{m}$ rozwiązujących zadanie (4.1), to znaczy sterowań potrzebnych do wywołania pożądanego biegu procesu. Ten zwłaszcza warunek jest najważniejszy – nie jest natomiast potrzebne, by postać funkcji celu w zadaniu (4.2) była taka sama, jaka występuje w zadaniu pierwotnym (4.1).

Przeformułowanie, o którym mowa, mieć będzie sens praktyczny wówczas, gdy rozwiązywanie zadania wyznaczania zadań dla układów sterowania bezpośredniego będzie mniej kosztowne lub mniej wrażliwe na niedokładność modelu procesu w realizacji bieżącej, czyli przy wykonywaniu w trakcie biegu procesu, od bieżącego rozwiązywania zadania pierwotnego, tj. optymalizacji biegu procesu względem wielkości sterujących.

Możliwość przeformułowań mających postulowane cechy wynika ze spostrzeżenia, że oprócz wielkości sterujących są w każdym procesie wielkości od nich zależne, takie jak współrzędne stanu i wielkości wyjściowe. Może się zdarzyć, że w myśl zadania optymalizacji procesu, wielkości sterujące trzeba dostosowywać do czynników zewnętrznych, natomiast wartości optymalne współrzędnych stanu są od tych czynników niezależne. Jeśli w takim przypadku zadanie wyznaczania optymalnych sterowań przekształcimy w zadanie utrzymania trajektorii stanu, czyli zadanie nadążania za trajektorią, to wystarczy odpowiednią trajektorię optymalną wyznaczyć tylko jeden raz, w dodatku bez potrzeby znajomości wspomnianych czynników zewnętrznych.

Myśl tego rodzaju mógł już Czytelnikowi nasunąć przykład optymalizacji biegu procesu w reaktorze chemicznym, opisany w rozdz. 3.2, począwszy od równań stanu (3.11), (3.12), (3.13). Doszliśmy tam do wzorów (3.20) – podających rozwiązania zadania optymalizacji procesu. Wzory te określają m.in., że wartość optymalnego dopływu ciepła $\dot{H}$ trzeba uzależnić od bieżącej wartości zakłóceń T_A, T_B, zatem trzeba by powtarzać rozwiązanie zadania optymalizacji na bieżąco, dla każdych nowych wartości T_A, T_B. Jeśli jednak wprowadzimy układ sterowania bezpośredniego (układ regulacji) z zadaniem podtrzymywania zadanej temperatury T, to wartość optymalną tej zadanej temperatury wystarczy obliczyć raz jeden, gdyż wynosi ona T_m i jest od wpływów bieżących niezależna. Z kolei, układ regulacji temperatury oddziałujący na dopływ ciepła H i podtrzymujący $T = T_m$ będzie reagował na zakłócenia bieżące T_A, T_B dokładnie w taki sposób, że spełnione będzie odpowiednie równanie z zestawu (3.20). Przykład ten rozwiniemy jeszcze w rozdz. 4.2.

Ujmując rzecz bardziej ogólnie stwierdzimy, że *zadania dla układów sterowania bezpośredniego* powinny być sformułowane tak, aby:

a) realizacja zadań sterowania bezpośredniego zapewniała pożądany (optymalny) bieg procesu,
b) modyfikacje bieżące zadań sterowania bezpośredniego były możliwie mało kosztowne, a więc interwencje rzadkie, dotyczące niewielkiej liczby zmiennych, określane za pomocą niezbyt skomplikowanych procedur decyzyjnych.

W dziedzinie sterowania ciągłymi procesami technologicznymi, jak widzieliśmy we wspomnianym przykładzie sterowania dopływem ciepła do reaktora, jest to najczęściej problem odpowiedniego wyboru wielkości regulowanych – takich, których dotrzymanie (powierzone układom sterowania bezpośredniego) dałoby w wyniku optymalny bieg procesu. Można przy tym próbować wybrać jako wielkości regulowane takie zmienne procesu lub ich proste funkcje, by zmiana wartości zadanych była zbędna lub tylko rzadko potrzebna, oraz by dotyczyła tylko części spośród wszystkich wielkości regulowanych. Zagadnienie to będzie jeszcze omawiane w rozdz. 4.4; jest ono także dobrze znane praktykom.

Podkreślmy, że układy sterowania stabilizujące wielkości regulowane wykonują swoje zadanie przez ciągłe lub prawie ciągłe oddziaływanie na proces sterowany, w sposób wynikający z wpływów zewnętrznych i właściwości samego procesu; wyznaczanie zadań ma odbywać się rzadziej (w skrajnym przypadku – tylko jeden raz).

Można na zagadnienie wyboru wielkości regulowanych spojrzeć także z innego punktu widzenia, a mianowicie od strony ekonomicznej: powinny to być zmienne procesu najbardziej cenne (w sensie ich wagi w ocenie jakości biegu procesu), czyli te, których odchylenia od wartości optymalnych przyniosą największe straty.

Spełnienie zatem warunku a) oraz dobre zbliżenie się do warunku b) może być trudne do jednoczesnej realizacji. Powstaje potrzeba kompromisu: bliskie do optymalnego zachowanie się procesu za cenę rzadszego lub łatwiejszego wyznaczania nowych zadań dla układów sterowania bezpośredniego.

Łatwiejsze wyznaczanie zadań to m.in. sprawa możliwie małej liczby zmiennych oraz prostoty modeli. Zauważmy tutaj, że istnienie układów sterowania bezpośredniego zmienia obraz procesu: proces zaopatrzony w układy sterowania bezpośredniego zachowuje się inaczej niż proces bez tych układów. W szczególności, układy sterowania bezpośredniego – jako oparte na ujemnym sprzężeniu zwrotnym – „przyspieszają" dynamikę (reakcja na zmianę wartości zadanej regulatora jest szybsza niż byłaby reakcja samego procesu na zmianę wielkości sterującej), a ponadto – co bardzo ważne – czynią tę dynamikę mniej zależną od wahań parametrów własnych procesu.

Obie te właściwości są korzystne dla rozpatrywanych w tej chwili „zadań wyznaczania zadań". Radykalnym, tym niemniej w wielu przypadkach zupełnie uzasadnionym, postępowaniem praktycznym jest np. wyznaczanie wartości zadanych wielkości regulowanych procesu ciągłego przy użyciu jego modelu statycznego, z całkowitym pominięciem dynamiki. Wprowadzając kolejne, nowe wartości zadane wiemy o tym, że po każdej zmianie istnieć będzie pewien przebieg przejściowy, dochodzenie do nowej wartości ustalonej; jeśli jednak trwa on krótko w stosunku do odstępu między kolejnymi zmianami, można go pominąć (objaśniamy to bliżej w rozdz. 4.5). Następuje pewna strata na optymalności, ale duży zysk na prostocie modelu – mamy model statyczny zamiast dynamicznego. Otwiera to m.in. możliwość zwiększenia rozległości sterowania przez jednostkę decyzyjną, położoną w warstwie wyznaczania zadań, to znaczy objęcia przez nią systemu czy podsystemu znacznie większego niż można by to zrobić bez wprowadzenia tak daleko idącego uproszczenia modelu.

Duże znaczenie ma także cecha druga – większa niezależność zachowania się procesu zaopatrzonego w układy sterowania bezpośredniego od wahań jego własnych parametrów, (a zatem także od stopnia ich znajomości przy opisie procesu) oraz towarzysząca temu większa niezależność procesu od czynników zewnętrznych. Pozwala to dość często realizować wyznaczanie zadań (np. nowych wartości zadanych dla wielkości regulowanych) bez sprzężenia zwrotnego, korzystając tylko z obserwacji otoczenia procesu. Na przykład, obserwując zmianę składu surowca dochodzącego do reaktora chemicznego żądamy przestawienia go na nowe wartości temperatury, ciśnienia i innych parametrów technologicznych, ale nie sprawdzamy już za pomocą sprzężenia zwrotnego czy nowy bieg procesu jest optymalny – decyzje o zmianie zadań dla regulatorów podjęliśmy na podstawie modelu procesu (uwzględniającego istnienie tych regulatorów). Postępowanie takie, będące sterowaniem w układzie otwartym, jest uzasadnione jeżeli użyty przy nim model procesu z regulatorami (zazwyczaj model statyczny) jest odpowiednio wiarygodny. Nie należy przy tym przeoczyć, że alternatywa w postaci sprzężenia zwrotnego opartego na obserwacji wartości funkcji celu (na przykład wydajności reaktora) może być mało skuteczna ze względu na pomiar tej wartości, który może wymagać uśrednienia w dłuższym okresie czasu. Pomiar taki może jednak np. być użyty do poprawienia (adaptacji) modelu, co można przypisać wyższej warstwie układu sterowania.

Podział sterowania na warstwę bezpośrednią i warstwę wyznaczania zadań przedstawiliśmy w odniesieniu do sterowania ciągłym procesem technologicznym. Zastosowaną tu metodę postępowania można przedstawić słownie w szerszym – być może – kontekście: tworzymy strukturę decyzyjną, w której warstwa niższa podejmuje wszystkie decyzje bieżące, a warstwa wyższa wyznacza zadania, którym decyzje te są podporządkowane; w sprzyjającym przypadku decyzje bieżące są podejmowane przy wykorzystaniu prostych reguł, a decyzje warstwy wyższej – stosunkowo rzadko.

Rozważane zmiany zadań dla układów sterowania bezpośredniego można uważać za pierwszy przykład interwencji, które mogą być dokonywane rzadziej aniżeli potrzebne dla utrzymania pożądanego biegu procesu zmiany wielkości sterujących.

Ponieważ można pomyśleć np. o dokonywaniu co pewien czas zmiany zasad, na których opiera się zmienianie zadań, można sobie wyobrazić coraz to wyższe warstwy sterowania, o coraz to malejącej częstotliwości interwencji. W literaturze, datującej się od cytowanej już pracy Lefkowitza, przyjęto nazywać warstwę najniższą warstwą *stabilizacji*, warstwę następną – warstwą *optymalizacji*, jeszcze następną – warstwą *adaptacji*. Obrazowe te określenia trzeba traktować nieco umownie; w szczególności warstwa najniższa, będąca warstwą sterowania bezpośredniego, może mieć cele bardzo odmienne od stabilizacji, a wyznaczane jej zadania nie muszą wynikać z optymalizacji procesu.

W istocie rzeczy, cechą niemal każdej struktury warstwowej jest nie tylko coraz mniejsza częstotliwość interwencji kolejnych warstw, lecz także operowanie – w wyższych warstwach – coraz dłuższym horyzontem czasowym. Niekiedy widoczne to jest bardzo wyraźnie, jak w przykładzie sterowania systemem wodno-gospodarczym, opisanym w rozdziale 1.5.

Malejąca, w górę warstw, częstotliwość interwencji jest cechą podstawową każdej struktury wielowarstwowej, cechą stanowiącą o jej istocie i rozróżnieniu między warstwami. Towarzyszą jej cechy w pewnym stopniu wynikowe, takie jak:

– rosnąca, w górę warstw, *złożoność* rozpatrywanych zadań decyzyjnych,
– możliwość operowania, w kolejnych warstwach, coraz to *mniej szczegółowymi modelami* procesu sterowanego,
– operowanie, w poszczególnych warstwach, *różnymi przesłankami* dla podejmowanych decyzji; od tych, które dotyczą nadążania za wyznaczonym zadaniem, do funkcji celu sformułowanej dla pełnego horyzontu sterowania,
– różnego charakteru *instrumenty oddziaływania* warstwy wyższej na niższą (zadania ilościowe, reguły postępowania, ograniczenia, przepisy itd.),
– różny rodzaj i postać *informacji* przekazywanej od warstwy niższej do wyższej (wartości zmiennych procesu, oceny statystyczne, ekonomiczne itd.),
– malejąca potrzeba i zasadność, często także *możliwość sterowania automatycznego*, tj. podejmowania decyzji bez udziału człowieka; automatyzacja nie ma dużego znaczenia przy decyzjach podejmowanych np. raz na tydzień, miesiąc czy rok, a ponadto nie jest możliwa przy mało wiarygodnej kwantyfikacji ocen wartościujących oraz przy intuicyjnej ocenie ryzyka.

Podane cechy były mniej lub więcej widoczne w przykładach omówionych w rozdz. 1.5. W rozdziale niniejszym przedstawiliśmy jak dotąd wielowarstwową koncepcję sterowania procesami w sposób opisowy, nie sięgając do wzorów i rozważań sformalizowanych. W dalszych częściach rozdziału przedstawiamy niektóre z zagadnień bardziej szczegółowo.

Trzeba jednak już teraz powiedzieć, że koncepcja warstwowa nie poddaje się łatwo analizie formalnej oraz że najczęściej nie można tej analizy doprowadzić do całkowicie miarodajnego rezultatu końcowego. O ile bowiem w przypadku struktur wielopoziomowych, takich jak omawiane w rozdziale 2, opartych na koordynacji równoległych zadań częściowych, można wykazać czy i kiedy są one całkowicie równoważne sterowaniu scentralizowanemu, czyli zadaniu niepodzielonemu, to w przypadku struktury warstwowej mamy prawie zawsze do czynienia z pewną stratą na jakości przebiegu procesu sterowanego. Strata ta jest równoważona prostotą i mniejszym kosztem, niekiedy wprost możliwością techniczną realizacji sterowania.

Sterowanie w podziale na warstwy jest szeroko stosowane, ale trudno jest jego racjonalność ująć ilościowo w sposób ogólny, czyli inaczej niż przez analizę kosztów i korzyści konkretnej realizacji.

4.2. Droga od procesu do układu sterowania

W rozdziale 3 zajmowaliśmy się przebiegiem pożądanym procesów, mających miejsce w obiektach systemu, nie zastanawiając się nad tym, w jaki sposób można go będzie zrealizować – czyli wywołać i podtrzymać – za pomocą układów sterowania. Schemat postępowania był wówczas następujący:

— *opis fizyczny procesu*: co się w danym obiekcie dzieje, co jest celem jego działania,

— *opis matematyczny procesu*: sformułowanie zależności między występującymi w procesie wielkościami,

— *określenie pożądanego przebiegu procesu*: sformułowanie warunków, które ma spełnić bieg procesu; jeśli to możliwe, wyznaczenie *przebiegu optymalnego*.

Mając za sobą te trzy etapy postępowania, a zatem mając do dyspozycji rezultaty przeprowadzonej analizy, myśleć można o etapie następnym, mającym na celu *zaprojektowanie układu sterowania* procesem.

Chodzi tu, ogólnie rzecz biorąc, o ustalenie wielkości mierzonych (obserwowanych) oraz zaproponowanie sposobu przekształcania obserwacji w oddziaływanie sterujące, przy czym przekształcanie to może – w przypadku sterowania złożonym systemem – odbywać się w strukturach różnego rodzaju.

Pierwszym przedmiotem zainteresowania będą układy sterowania bezpośredniego, czyli te, które mają styczność bezpośrednią z fizycznymi wielkościami sterującymi procesu; są one zawsze potrzebną pierwszą warstwą sterowania. Nie należy przy tym pomijać pytania: czy struktura sterowania procesem nie może poprzestać na układach sterowania bezpośredniego, czyli czy nie można tych układów ukształtować w taki sposób, by wyższe warstwy czy poziomy sterowania nie były potrzebne?

Przyjmijmy, dla ustalenia uwagi, że mamy proces opisany równaniem stanu

$$\dot{x} = f\left(x(t), m(t), z(t)\right) \tag{4.3}$$

oraz że przebieg pożądany procesu ma spełniać zadanie

$$\text{maksymalizować } Q = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} f_0\left(x(t), m(t), z(t)\right) \mathrm{d}t \tag{4.4}$$

gdzie $[t_1, t_2]$ jest rozpatrywanym przedziałem czasowym, a stany brzegowe $x(t_1)$, $x(t_2)$ mogą być swobodne lub zadane.

Zmienna $z(t)$ reprezentuje w zapisie (4.3), (4.4) wpływ zewnętrzny na proces, czyli oddziaływanie otoczenia. Będzie to z reguły wielkość losowa. W związku z tym sformułowanie oceny Q użyte we wzorze (4.4) można uważać za poprawne tylko *ex post*; mówiąc inaczej, dopiero w chwili t_2 będziemy wiedzieli, jaki był przebieg $z_{[t_1,t_2]}$ oraz ile wyniosła osiągnięta wartość Q. Do celów sterowania musimy rozporządzać oceną wartości Q dokonywaną w chwili t_1, czyli oceną probabilistyczną. Najczęściej odpowiednia tu będzie wartość oczekiwana, tak że zadanie (4.4) przekształci się do postaci

$$\text{maksymalizować } Q = E\left[\frac{1}{t_2 - t_1}\int_{t_1}^{t_2} f_0\left(x(t), m(t), z(t)\right) \mathrm{d}t\right] \tag{4.5}$$

Rezultatu zadania (4.5) możemy poszukiwać w różnych formach. Na przykład, dla sterowania w układzie otwartym potrzebne jest wyznaczenie przebiegu sterowania $m_{[t_1,t_2]}$, a dla sterowania opartego na sprzężeniu zwrotnym – wyznaczenie prawa sterowania $d(\cdot)$, aby go użyć do wyznaczenia sterowania bieżącego wg wzoru $m(t) = d(w(t))$, gdzie $w(t)$ oznacza bieżącą obserwację.

Przystępując do zagadnienia bardziej ogólnie zwróćmy najpierw uwagę, że analiza zadania (4.4) lub (4.5) pozwala stwierdzić, czy wszystkie wielkości wejściowe procesu brane pod uwagę jako *możliwe do użycia* wielkości sterujące m są *istotne*. Jeśli się okaże, że ten sam rezultat możemy otrzymać traktując nie wszystkie wejścia m jako wielkości sterujące, tzn. jeżeli

$$\max_{m^1} Q = \max_{m^1, m^2} Q \tag{4.6}$$

gdzie m^1, m^2 są rozłącznymi podzbiorami m, to wielkości wejściowe m^2 można nazwać *nieistotnymi* dla danego zadania sterowania. Użycie ich jako wielkości sterujących nie poprawi wyniku, to jest osiąganej wartości funkcji celu. Implikacją praktyczną jest to, że można sterowaniem objąć mniejszą liczbę wielkości wejściowych, pozostawiając m^2 jako wielkości swobodne. Jest to cecha pozytywna, pamiętać jednak należy, że trzeba tu uwzględnić dopuszczalne lub możliwe obszary zmienności m^1, m^2. Oznaczając te obszary przez M_1, M_2 napiszemy warunek (4.6) nieistotności zmiennych m^2 bardziej poprawnie jako

$$\max_{m^1 \in M_1} Q = \max_{\substack{m^1 \in M_1 \\ m^2 \in M_2}} Q \tag{4.7}$$

przy czym nie wzięliśmy jeszcze pod uwagę tego, czy sterowanie za pomocą samego tylko m^1 nie będzie bardziej kosztowne niż sterowanie za pomocą zarówno m^1 jak m^2. Jeśli przyjmiemy, że w sformułowanie Q włączono także koszty sterowania, wnioski z (4.7) będą w pełni poprawne.

Przypomnijmy, że dosyć podobne rozważania prowadziliśmy w rozdz. 3.4 poświęconym horyzontowi sterowania, zmierzając tam do ustalenia, jaki co najmniej horyzont trzeba brać pod uwagę w chwili t_0, aby poprawnie ustalić sterowanie.

Wyliczony z zadania (4.4), czyli *ex post*, przebieg optymalny sterowania $\hat{m}_{[t_1,t_2]}$ będzie w ogólnym przypadku zależny od przebiegu czynnika zewnętrznego $z_{[t_1,t_2]}$ oraz od parametrów zawartych w modelu procesu (4.3), które z kolei pochodzą od parametrów obiektu, w którym analizowany proces się odbywa. Mówić trzeba oddzielnie o wpływie czynnika zewnętrznego, który jest zmienny w czasie oraz o wpływie *niepewności modelu*, która jest czynnikiem stałym. Na niepewność tę składają się błędy co do wartości parametrów procesu oraz wątpliwości co do wykorzystanych praw fizycznych i innych, użytych przy układaniu równań procesu sterowanego. Dojść do tego może zamierzona niedokładność modelu, wynikająca z celowo wprowadzonych uproszczeń, na przykład z pominięcia niektórych zjawisk, agregacji zmiennych, przyjęcia zależności liniowych w miejsce rzeczywistych nieliniowych i tak dalej.

Niemożność wyznaczenia z góry dostatecznie wiarygodnego sterowania $\hat{m}_{[t_1,t_2]}$ jest przyczyną, dla której niemal każda spotykana w praktyce struktura sterowania procesami (nie tylko technicznymi) korzysta z informacji o przebiegu procesu rzeczywistego, czyli korzysta ze sprzężenia zwrotnego. W zasadzie odróżnia to sterowanie od *planowania*, aczkolwiek i tam dokonuje się korekty planu w miarę obserwacji jego wykonywania, jako tzw. *planowanie kroczące* – wspomnieliśmy o tym w rozdz. 3.4.

Dochodzimy teraz do kwestii, co chcemy w procesie obserwować albo lepiej – na jakiej informacji zwrotnej chcielibyśmy oprzeć sterowanie, poczynając od bieżących decyzji sterowania bezpośredniego, a kończąc na decyzjach sterowania długofalowego całością systemu na najwyższym szczeblu.

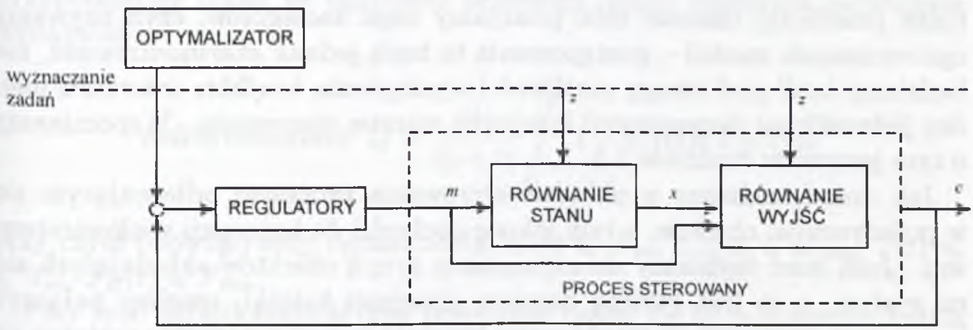
Zatrzymajmy się przy informacji zwrotnej potrzebnej sterowaniu bezpośredniemu: wydaje się rzeczą naturalną zacząć od pytania, co będzie stawiane temu sterowaniu jako zadanie do spełnienia ze strony warstwy następnej (warstwy wyznaczania zadań), a następnie dostosować informację zwrotną o biegu procesu do tego zadania. Z kolei, jak mówiliśmy w rozdziale 4.1, warstwa wyznaczania zadań powinna mieć możliwie mało powodów do interwencji, do zmieniania swoich postulatów. Wiąże się to z ewentualnym przeformułowaniem zadania stawianego przed procesem tak, by optymalizować nie względem pierwotnych i naturalnych wielkości sterujących m, lecz względem zmiennych wybranych inaczej, ewentualnie wprowadzonych do zadania w tym właśnie celu – zmiennych c. W rozdz. 4.1

objaśniliśmy to na przykładzie sterowania dopływem ciepła do reaktora, zrealizowanego jako regulacja temperatury.

Z punktu widzenia struktury sterowania mamy tu do czynienia z układem dwuwarstwowym, w którym warstwa sterowania bezpośredniego ma postać układów regulacji (rys. 4.1). Wprowadza się wielkości regulowane, związane ze zmiennymi procesu równaniami wyjść

$$c(t) = h\left(x(t), m(t), z(t)\right) \tag{4.8}$$

przy czym warstwa wyznaczania zadań ma określić wartości zadane dla c w postaci wektora c_d. Układy regulacji mają zapewnić równość $c(t) = c_d(t)$, czyli mają zapewnić nadążanie, natomiast wybór wielkości regulowanych



Rys. 4.1. Dwuwarstwowy układ sterowania

powinien być tak dokonany, ażeby optymalność c zapewniała zaistnienie optymalnych sterowań

$$c \text{ dążące do } \hat{c} \text{ ma wywołać } m \text{ dążące do } \hat{m} \tag{4.9}$$

Wybór wielkości regulowanych to w istocie ustalenie struktury związków (4.8), spełnienie zaś wymagania (4.9) zależy od odpowiedniego dostosowania (4.8) do równań procesu. Ważny ten problem omawiamy szczegółowo w rozdz. 4.4. Jest i druga kwestia: tak wybrać wielkości c, czyli strukturę związków (4.8), aby wartości optymalne c mało zależały od czynników zewnętrznych oraz niepewności i niedokładności modelu; to także omawiamy w rozdz. 4.4, ale najpierw pokazujemy na przykładzie w rozdz. 4.3.

Ustalenie (wybór) wielkości regulowanych to pierwszy krok w projektowaniu struktury sterowania procesem w danym obiekcie. Jeśli się okaże, iż wartości zadane wielkości regulowanych nie powinny być stałe, lecz wymagają – dla podtrzymania pożądanego biegu procesu – zmian w trakcie jego trwania, czyli wewnątrz przedziału $[t_1, t_2]$, trzeba do warstwy sterowania bezpośredniego dodać warstwę zmiany zadań.

Realizacja zmiany zadań może się opierać na obserwacji czynników zewnętrznych (działanie w układzie otwartym), może też korzystać z obserwacji procesu (działanie oparte na sprzężeniu zwrotnym). Istotą rzeczy pozostaje tutaj to, że w warstwie zmiany zadań dla sterowania bezpośredniego opierać się będziemy, w ten czy inny sposób, na modelu procesu sterowanego oraz operować będziemy określonym horyzontem sterowania. Stąd też, w warstwie następnej możemy zatroszczyć się o dłuższy horyzont bądź o poprawianie modelu procesu, bądź też o jedno i drugie jednocześnie. Ważne będzie to, że w warstwie wyznaczania zadań i w warstwach wyższych mamy na celu zapewnienie pożądanego biegu ciągle tego samego procesu „fizycznego”, mamy więc w zasadzie na widoku podstawową, pierwotną funkcję celu Q. Nie przeszkadza to temu, że w różnych warstwach możemy operować inaczej zbudowanymi sformułowaniami zadania, na przykład rozpatrujemy różne przedziały czasowe albo pomijamy część szczegółów, czyli używamy uproszczonych modeli – postępowania te będą jednak *zharmonizowane*, nie będziemy brali pod uwagę możliwości wystąpienia *konfliktu interesów* między jednostkami decyzyjnymi kolejnych warstw sterowania. Wspominamy o tym jeszcze w rozdziale 5.5.

Jak dotąd, mówimy o układzie sterowania procesem odbywającym się w pojedynczym obiekcie, o tym jak się dochodzi do koncepcji wielowarstwowej. Jeśli mieć będziemy do czynienia z grupą obiektów składających się na system, a to jest główną domeną niniejszej książki, musimy połączyć koncepcję warstwową ze strukturami korzystającymi z podziału systemu na części, omawianymi w rozdz. 2. Synteza ta pojawić się może już na szczeblu warstwy drugiej, to jest warstwy wyznaczania zadań dla sterowania bezpośredniego. Wspominaliśmy już o tym, że w rozdziale 2 braliśmy to pod uwagę przez oznaczenie zmiennych decyzyjnych literą c. Dla wyższych warstw i poziomów trudno o reguły, trzeba się raczej odwoływać do przykładów, czyli sięgać po rozwiązania indywidualne. Jedyną regułą ogólną pozostaje to, że na szczycie struktury ma się znaleźć jednostka *obejmująca cały system* i operująca *pełnym horyzontem czasowym.* Szczegółowe omówienie tego problemu można znaleźć w rozdziale 5.

4.3. Przykład układu dwuwarstwowego

W niniejszym rozdziale przedstawimy drogę prowadzącą od określenia pożądanego przebiegu procesu sterowanego do struktury i podstawowych cech układu sterowania na konkretnym przykładzie, gdyż to pozwoli na pokazanie jak bardzo droga ta jest naturalna i logicznie zrozumiała.

Omawiać będziemy kolejno:

– punkt wyjścia, czyli pożądany przebieg procesu sterowanego,
– wybór wielkości sterujących,

– wybór wielkości regulowanych,

– ukształtowanie warstwy sterowania bezpośredniego,

– warstwę optymalizacji,

a przykład zakończymy analizą cech powstałego układu.

Obiektem, który pozwoli na konkretyzację rozważań będzie reaktor chemiczny omawiany w rozdziale 3.2 i przedstawiony schematycznie na rys. 3.6. Przebiega w nim przemiana $A \to B$; reaktor pracuje w trybie ciągłym (przepływowym). Stan procesu w reaktorze można określić za pomocą trzech współrzędnych stanu W, C_A, T; przy założeniu idealnego zmieszania zawartości reaktora równania stanu mają postać równań różniczkowych zwyczajnych, podanych w rozdz. 3.2 jako wzory (3.11)–(3.13). Objaśniono tam także wszystkie zmienne i parametry związane z rozpatrywanym procesem. Przypomnijmy także, że żądaliśmy przebiegu, który zapewnia największą wydajność

$$\text{maksymalizować } Q = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} (1 - C_A(t))\, F_D(t) \mathrm{d}t$$

przy czym obowiązywały ograniczenia $W(t) \leqslant W_m$, $C_A(t) \leqslant C_{Am}$, $T(t) \leqslant T_m$, $F_B(t) \leqslant F_{Bm}$.

Przy tych ograniczeniach oraz przedziale czasu $[t_1, t_2]$ położonym z dala od chwili rozruchu i od chwili zatrzymania procesu, rozwiązanie zadania optymalizacji miało postać stanu ustalonego. Optymalne wartości zmiennych procesu w stanie ustalonym, czyli przy $W(t) = \text{const}$, $C_A(t) = \text{const}$, $T(t) = \text{const}$ można było wyznaczyć z zadania optymalizacji niedynamicznej, którego pełna postać jest następująca

$$\text{maksymalizować } Q = (1 - C_A) F_D(t) \tag{4.10}$$

przy ograniczeniach nierównościowych

$$W \leqslant W_m, \quad C_A \leqslant C_{Am}, \quad T \leqslant T_m, \quad F_B \leqslant F_{Bm} \tag{4.11}$$

oraz ograniczeniach równościowych, powstałych z równań stanu

$$F_A(t) + F_B(t) - F_D(t) = 0 \tag{4.12}$$

$$\frac{1 - C_A}{W} F_A(t) - \frac{C_A}{W} F_B(t) - C_A k(T) = 0 \tag{4.13}$$

$$\frac{T_A(t) - T}{W} F_A(t) + \frac{T_B(t) - T}{W} F_B(t) + \frac{H(t)}{cW} - \frac{h}{c} C_A k(T) = 0 \tag{4.14}$$

W zapisie zadania optymalizacji stanu ustalonego zachowaliśmy ewentualną zmienność w czasie wielkości wejściowych, tak jak to wskazywały wzory ogólne (3.9), (3.10). Stałość współrzędnych stanu nie implikuje stałości wielkości wejściowych, lecz przeciwnie – zachowanie między nimi, w każdej chwili czasu, określonych relacji. Będzie to szczególnie ważne gdy dojdziemy do rozpatrywania wpływu wielkości zewnętrznych – utrzymanie stanu ustalonego procesu powinno być możliwe także przy czynnikach zewnętrznych zmiennych w czasie.

Zadanie optymalizacji sformułowane przez (4.10)–(4.14) jest zadaniem programowania nieliniowego (optymalizacji statycznej). Przy jego rozwiązywaniu parametr t może być pominięty, a konkretny wynik ma postać wzorów podanych w rozdz. 3.2 jako (3.20). Wzory te będziemy obecnie analizować dokładniej, warto więc je przytoczyć ponownie – tym razem przypominając istnienie parametru t

$$
\begin{aligned}
&\hat{W} = W_m \\
&\hat{C} = C_{Am} \\
&\hat{T} = T_m \\
&\hat{F}_A = C_{Am}(1 - C_{Am})^{-1}[W_m k(T_m) + F_{Bm}] \\
&\hat{F}_B = F_{Bm} \\
&\hat{F}_D = (1 - C_{Am})^{-1}\left[W_{Am} C_{Am} k(T_m) + F_{Bm}\right] \\
&c\hat{F}_A \hat{T}_A(t) + cF_{Bm}\hat{T}_B(t) + \hat{H}(t) = hW_m C_{Am} k(T_m) + cF_D T_m
\end{aligned}
\tag{4.15}
$$

Wzory (4.15) kończą etap rozważań, mający na celu określenie pożądanego przebiegu procesu. Zawierają one m.in. informację, że potrzebne dla tego przebiegu wartości wejść F_A, F_B oraz F_D są określone jednoznacznie przez czynniki stałe zadania wyjściowego, nie dopuszcza się ich zmienności w czasie bez odejścia od optymalności. Zauważmy jednak, że gdyby granica obszaru dopuszczalnego F_{Bm} była zmienna w czasie jako $F_{Bm}(t)$, to odpowiednio zmienne w czasie byłyby także wartości optymalne przepływów F_A, F_B i F_D. Napisalibyśmy wówczas we wzorach (4.15) symbole $\hat{F}_A(t)$, $\hat{F}_B(t)$, $\hat{F}_D(t)$. Czytelnik może sprawdzić, że zmienne te będą przy tym zachowywały warunek bilansu przepływów określony przez równanie (4.12), wynikający z postulatu stałości zapełnienia $W(t) = \text{const}$.

Przejdźmy teraz do następnego etapu, to jest do *zaprojektowania układu sterowania.* Punktem wyjścia będą wzory (4.15) – determinują one warunki i relacje, które muszą być spełnione, aby przebieg procesu był przebiegiem pożądanym (optymalnym). Postępowanie przeprowadzimy według pewnego schematu.

Pierwszą kwestią związaną z przyszłą strukturą sterowania jest *wybór wielkości sterujących.*

Które z wielkości wejściowych procesu należy przyjąć za wielkości kontrolowane, czyli sterujące, a które można pozostawić jako wejścia swobodne, czyli jako wpływy zewnętrzne?

Zauważmy, że w równaniach stanu (3.11)–(3.13) znajdujemy trzy współrzędne stanu W, C_A oraz T i sześć wielkości wejściowych procesu F_A, F_B, F_D, H, T_A oraz T_B. Wszystkie te zmienne występują jako zmienne decyzyjne w zadaniu statycznym (4.10)–(4.14). Jest ich dziewięć.

Rozwiązanie (4.15) nie determinuje dziewięciu zmiennych, daje ono tylko siedem warunków. Wspominaliśmy już w rozdz. 3.2, że ostatni z warunków (3.20), czy też (4.15) postuluje spełnienie określonej relacji między trzema spośród zmiennych decyzyjnych – przy czym są to trzy spośród wielkości wejściowych procesu.

Patrząc na wzory (4.15) pod kątem wielkości wejściowych procesu, widzimy, że dają one dla sześciu wielkości cztery warunki do spełnienia – dla podtrzymania optymalnego stanu ustalonego potrzebne są zatem cztery niezależne od siebie wzajemnie wielkości sterujące. Muszą wśród nich być F_A, F_B oraz F_D, gdyż rozwiązanie (4.15) nie zostawia im swobody oraz jedna spośród pozostałych trzech wielkości wejściowych: H lub T_A lub T_B. Wspominaliśmy już w rozdz. 3.2, iż wybór padnie tu zapewne na wielkość H; zwróćmy bowiem uwagę, że realizacja sterowania dopływem H może być prosta (zawór sterujący), podczas gdy realizacja sterowania temperaturą T_A lub T_B wymagałaby urządzenia bardziej skomplikowanego, np. wymiennika ciepła. Przepływy F_A, F_B, F_D oraz H są istotnymi (dla danego zadania sterowania) wielkościami wejściowymi procesu (przypominamy wzór (4.7)).

Wybór wielkości sterujących został dokonany, można przejść do określenia następnego elementu struktury sterowania – do *wyboru wielkości regulowanych* – a następnie do określenia praw bądź algorytmów sterowania, tj. przekształcania obserwacji tych wielkości w oddziaływania sterujące.

Przestrzec tu należy przed utworzeniem struktury, która by na przykład – za pomocą układów regulacji przepływów – stabilizowała wielkości wejściowe F_A, F_B oraz F_D. Myśl taka nasunąć się może w sposób dość naturalny, byłoby to dosłowne wykorzystanie faktu jednoznacznego określenia tych wielkości przez wzory (4.15). Zauważmy jednak, że stabilizowane odpowiednimi regulatorami przepływy muszą spełniać warunek

$$F_A + F_B - F_D = 0$$

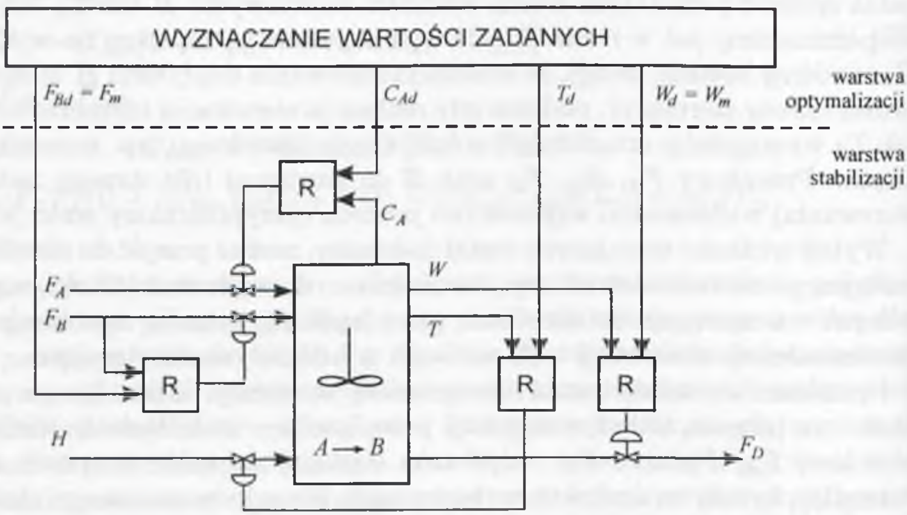
a naruszenie tego warunku wywołałoby powolne przepełnianie się lub opróżnianie reaktora. Z kolei, nie można gwarantować – choćby ze względu na błędy pomiaru przepływów – że warunek ich ściśle zerowego bilansu będzie mógł być zachowany w praktyce.

Punktu wyjścia do projektu struktury układu sterowania trzeba w naszym przypadku poszukać wcześniej, przed wzorami (4.15), a mianowicie w pierwotnym, w stosunku do całego zadania (4.10)–(4.14) wymaganiu, by zapewnić w reaktorze stan ustalony. Oznacza to, iż myślimy w istocie rzeczy o dwóch kolejnych kwestiach:

1) należy zapewnić stałość współrzędnych stanu, tj. W = const, C_A = = const oraz T = const,
2) ustalone wartości W, C_A, T oraz wartości wielkości sterujących powinny być optymalne.

Punkt 1 sugeruje, żeby wielkości W, C_A i T przyjąć za wielkości regulowane i w ten sposób zapewnić stabilizację procesu w reaktorze. Punkt 2 odnosi się natomiast do optymalizacji procesu. Realizacja tych zadań, przez wzajemne nałożenie, stanowić będzie istotę dwuwarstwowego układu sterowania.

Stałość współrzędnej stanu W skutecznie zapewnić można w układzie sterowania ze sprzężeniem zwrotnym (układzie regulacji) spełniającym wprost wymaganie stabilizacji W = const. Pokazano to na rys. 4.2, podając zarazem układy stabilizacji pozostałych dwóch współrzędnych stanu, tj. C_A



Rys. 4.2. Układ sterowania reaktora z rys. 3.6

oraz T. Trzeba tu było dokonać wyboru, którą z wielkości wejściowych sterujących powiązać z regulatorem danej współrzędnej stanu. Przyjęliśmy na rys. 4.2, że regulator zapełnienia W posługuje się odpływem F_D, regulator stężenia C_A ma do dyspozycji dopływ F_A, regulator temperatury T – dopływ ciepła H. Nie była to jedyna możliwość wyboru powiązań; właściwe ukształtowanie układów stabilizacji wymaga m.in. uwzględnienia dynamiki i rozwijać tego teraz nie będziemy. Zwróćmy tylko uwagę, że dla zapewnienia stabilizacji stanu procesu potrzeba nam było tyle wielkości sterujących, ile jest współrzędnych stanu.

Przechodząc do zagadnienia optymalizacji zauważmy, że po wprowadzeniu trzech regulatorów stabilizujących mamy następujące zadanie

$$\text{maksymalizować } Q = (1 - C_A)F_D$$

to znaczy to samo zadanie co poprzednio, lecz z tą różnicą że swobodnymi zmiennymi decyzyjnymi są teraz W_d, C_{Ad}, T_d, F_B, T_A, T_B. Pisząc to przyjmujemy oczywiście, że zadania stabilizacji będą wykonywane dokładnie, tj. będzie $W = W_d$, $C_A = C_{Ad}$, $T = T_d$. Przeformułowane zadanie trzeba rozwiązać z uwzględnieniem tych samych ograniczeń co poprzednio, tj. (4.11), a rozwiązanie mieć będzie postać

$$\begin{aligned} \hat{W}_d &= W_m \\ \hat{C}_{Ad} &= C_{Am} \\ \hat{T}_d &= T_m \\ \hat{F}_B &= F_{Bm} \end{aligned} \tag{4.16}$$

T_A, T_B – dowolne.

Rozwiązanie to można otrzymać bez znajomości stałej prędkości reakcji $k(T)$, a tylko przy założeniu, że stała ta rośnie wraz z temperaturą T.

Pierwsze trzy z otrzymanych liczb określają wartości zadane, które należy ustawić w regulatorach stabilizujących współrzędne stanu. Czwarta liczba podaje optymalną wartość dopływu F_B. Wartości T_A, T_B są dowolne, gdyż $T = \text{const}$ gwarantuje stałość bilansu cieplnego, a $T = \text{const} = T_m$ zapewnia optymalność tego stanu ustalonego. Wszystkie te wartości były oczywiście dane przez rozwiązanie zadania optymalizacji, wzory (4.15). Tu chcemy podkreślić, że w sterowaniu dwuwarstwowym wykorzystujemy odpowiednio wybrane elementy tego rozwiązania.

Zauważmy, że budując układy stabilizacji stanu wykorzystaliśmy trzy spośród czterech wielkości sterujących. Wzory (4.16), podobnie jak (4.15), wskazują, że czwarta wielkość sterująca czyli dopływ F_B, nie może pozostać poza kontrolą – mamy dla niej ściśle określoną wartość optymalną. Stąd też na rys. 4.2 umieściliśmy regulator stabilizujący przepływ F_B.

Może jeszcze nasunąć się wątpliwość, czy po zastosowaniu układów sterowania bezpośredniego stabilizujących W, C_A, T na wartościach, odpowiednio: W_m, C_{Am}, T_m, za pomocą sterowań F_D, F_A, H ustalą się w procesie te same optymalne wartości wejść F_D, F_A, H, które były określone przez wzory (4.15). Gdyby tak nie było, układ dwuwarstwowy nie zapewniałby optymalności procesu. W istocie trzeba spełnić warunek (4.9), w którym chodziło o to, by narzucenie procesowi określonej wartości wektora wielkości regulowanych c, czyli W, C_A, T, wywołało zaistnienie określonej i optymalnej wartości wektora wielkości sterujących m – w tym przypadku zmiennych F_D, F_A, H. Zależy to od struktury równań procesu, które muszą spełniać omawiany w rozdz. 4.4 *warunek jednoznaczności*. W omawianym przykładzie, równania (3.11)–(3.13) warunek ten spełniają, zatem – przy działaniu układów stabilizacji stanu – ustalą się jednoznaczne i optymalne wartości wielkości sterujących F_D, F_A, H. Będą to w zasadzie wartości określone przez wzory (4.15); w zasadzie, bowiem wartości rzeczywiste F_D, F_A, H wynikną z „równań rzeczywistych", tj. rzeczywistych właściwości procesu, podczas gdy wartości (4.15) są wyznaczone na podstawie modelu. Wartości rzeczywiste powstałe w układzie z rys. 4.2 będą bardziej poprawne niż oparte na modelu, a mianowicie takie, że będą przy nich dokładnie spełnione decydujące o optymalności ograniczenia $W \leqslant W_m$, $C_A \leqslant C_{Am}$, $T \leqslant T_m$. Narzucenie procesowi np. wartości $\hat{F}_A$ wynikającej ze wzorów (4.15), w przypadku błędu w określeniu stałej prędkości reakcji $k(T_m)$, dałoby $C_A \neq C_{Am}$ – zatem albo przekroczenie ograniczenia, albo wydajność mniejszą od optimum. Widać tu jeden jeszcze element wyższości wyboru wielkości regulowanych W, C_A, T nad ewentualną stabilizacją wejść F_D, F_A, H.

Stwierdźmy na zakończenie, że *układ dwuwarstwowy* z rys. 4.2 ma w naszym konkretnym przykładzie kilka zalet, a mianowicie:

– stabilizacja stanu reaktora jest osiągnięta niezależnie od dokładnej znajomości jego równań,
– stan optymalny wyznaczany jest niezależnie od stabilizacji, zatem ewentualne błędy optymalizatora nie wywołują wytrącenia reaktora z równowagi,
– do wyznaczenia optymalnych wartości zadanych można nie znać wartości stałej prędkości reakcji $k(T)$ – wystarczyła informacja, że rośnie ona z temperaturą.

W przykładzie nie można było pokazać, czy są potrzebne – oraz jak często – interwencje optymalizatora w trakcie trwania procesu. Było raczej przeciwnie: układy sterowania bezpośredniego przy prawidłowo określonych wartościach zadanych zapewniały optymalny bieg procesu – także przy zakłóceniach takich jak zmiany właściwości procesu, np. zmiany stałej prędkości reakcji. Optymalizacja bieżąca była zbędna, ale mogłaby się okazać potrzebna, np. przy zmianie sformułowania funkcji celu Q. Warto jeszcze zwrócić uwagę, że do określenia praw działania poszczególnych warstw potrzebne są różne modele tego samego procesu – *model statyczny* dla optymalizacji, a *model dynamiczny* dla konkretnego projektowania układów regulacji.

Po określeniu struktury układów sterowania bezpośredniego, to jest po ustaleniu wielkości sterujących i wielkości regulowanych trzeba ustalić *prawa* *sterowania*; na rys. 4.2 przyjęto bez głębszej analizy, że użyte będą cztery oddzielne regulatory, bez tzw. powiązań skrośnych, ale i tak trzeba jeszcze określić czy mają to na przykład być regulatory PI czy PID oraz o jakich nastawach. Nie będziemy tego rozwijać, zauważmy tylko że dla zaprojektowania regulatorów wystarczy model procesu sterowanego obowiązujący dla małych odchyleń, zatem model powstały w wyniku linearyzacji równań dynamicznych procesu.

Inny model procesu był lub będzie potrzebny w warstwie wyznaczania zadań – mamy tu do czynienia z optymalizacją stanu ustalonego, model ma wyrażać zależność statyczną funkcji celu Q od zmiennych decyzyjnych tej warstwy sterowania, czyli od wektora c_d, w razie potrzeby także zależność wyjścia procesu sterowanego od tych zmiennych.

Widać bardzo wyraźnie, że model używany w warstwie drugiej uwzględniać ma istnienie i działanie układów sterowania bezpośredniego. Według oznaczeń stosowanych w rozdziale 2 jest to model typu (2.2), (2.6)

$$\begin{aligned} y &= F(c, u, z) \\ Q &= Q^0(c, u, y) \end{aligned}$$

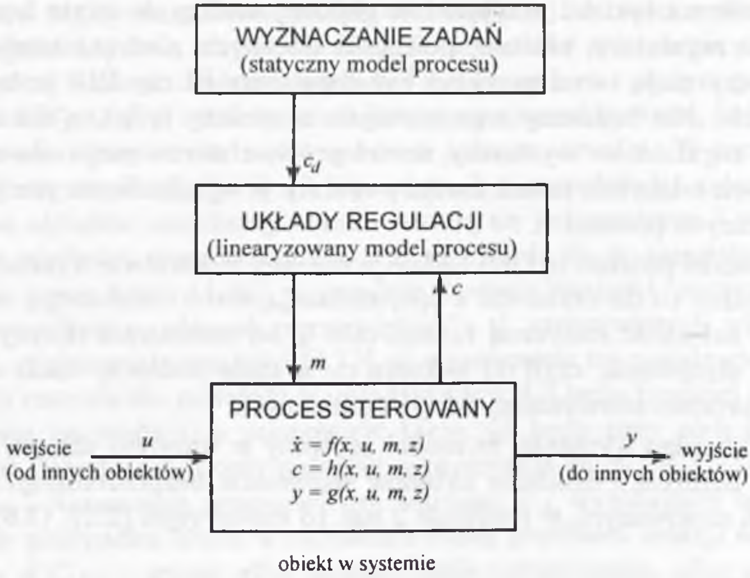
Przypomnijmy, że w tych wzorach zmienna u oznacza wejście procesu pochodzące z innych części tego samego systemu – w naszym przykładzie nie istniejące*). Dla rozpatrywanego reaktora modele potrzebne drugiej warstwie są statyczne (w ogólnym przypadku być nie muszą) i mają postać

$$y = \begin{bmatrix} (1 - C_A)^{-1}(WC_Ak(T) + F_B) \\ C_A \\ T \end{bmatrix}$$

$$Q = WC_Ak(T) + F_B$$

Rysunek 4.3 przedstawia strukturę dwuwarstwową z naszego przykładu w sposób bardziej ogólny niż rys. 4.2, za to z podkreśleniem zróżnicowania modeli procesu, które są potrzebne poszczególnym warstwom sterowania: do określenia praw sterowania w warstwie pierwszej bądź do sformułowania zadania decyzyjnego w warstwie drugiej.

*) Reaktor rozpatrywaliśmy tak, jak gdyby pracował on samodzielnie, nie był ogniwem procesu ciągłego w linii technologicznej. Stąd także mogliśmy uważać za swobodne wszystkie przepływy, tj. F_A, F_B, F_D oraz H.



Rys. 4.3. Rozszerzenie przykładu: reaktor może być obiektem w systemie

4.4. Zasady wyboru wielkości regulowanych

W poprzednim podrozdziale przedstawiliśmy przykład, w którym odpowiednio dobrane układy regulacji całkowicie zastępowały optymalizację bieżącą – zapewniały pożądany bieg procesu i maksymalną wydajność, mimo wpływu czynników zewnętrznych, a także niepełnej znajomości parametrów występujących w równaniach procesu sterowanego. Istotną przyczyną tego powodzenia był odpowiedni wybór wielkości regulowanych. Teraz postaramy się znaleźć reguły ogólne, a także wskazać, kiedy optymalizacja bieżąca stanie się niezbędna.

Rozpatrujemy układ (patrz np. rys. 4.1), w którym zamiast wyznaczać sterowanie m wprost z zadania optymalizacji

$$\underset{m}{\text{maksymalizować}}\ Q$$

wprowadzamy układy regulacji (sterowania bezpośredniego) dla wielkości regulowanych c, a następnie wyznaczamy optymalne wartości zadane c_d, maksymalizujące tę samą funkcję celu. Główny sens tej idei polega na tym, że dzięki spełnianiu zadań nadążania c za c_d upraszcza się często, ułatwia lub nawet eliminuje zadanie bieżącej optymalizacji procesu. Innymi słowy, właściwy wybór wielkości regulowanych pozwolić może na uniknięcie stosowania optymalizatora.

Rozpatrzyć należy w istocie dwa zagadnienia. Po pierwsze, wybór wielkości regulowanych c musi być taki, aby nadanie im określonych wartości rozstrzygało jednoznacznie zadanie optymalizacji

$$c = \hat{c} \;\text{ powoduje }\; Q = \hat{Q} \tag{4.17}$$

Nazwiemy to *warunkiem jednoznaczności.*

Po drugie, możemy się starać o taki wybór c, aby wartości optymalne wszystkich lub dużej części elementów c nie zależały od czynników zewnętrznych

$$\hat{c}_j \neq \hat{c}_j(z), \quad j = 1, \ldots, k \tag{4.18}$$

Jeśli (4.18) będzie spełnione dla wszystkich elementów c_j, to zadanie sterowania procesem sprowadzi się do stabilizacji c (bądź też do regulacji programowej c, jeśli $\hat{c} = \hat{c}(t)$, tzn. dla $\hat{c}$ jawnie zależnego od czasu).

Będzie to zatem *warunek inwariantności* $\hat{c}$ (optymalnej wartości wektora wielkości regulowanych) względem czynników zewnętrznych, które w tym kontekście nazywać można *zakłóceniami.*

Spełnianie warunku jednoznaczności

Rozpatrywać będziemy najpierw warunek (4.17), który jest warunkiem koniecznym poprawności wyboru wektora wielkości regulowanych c.

Przyjmijmy, że zadanie optymalizacji brzmi

$$\text{maksymalizować } Q, \quad \text{przy} \quad \dot{x} = f(x, m, z) \tag{4.19}$$

oraz że zarówno x, jak i m są zmiennymi istotnymi – to znaczy, że $Q = \hat{Q}$ otrzymujemy tylko wówczas, gdy $x = \hat{x}$ oraz $m = \hat{m}$, gdzie $\hat{x}$, $\hat{m}$ są to określone wartości optymalne. W tej sytuacji warunek (4.17) zastąpić można równoważnie warunkiem określoności rozwiązań przy c jako zmiennej niezależnej

$$c = \hat{c} \quad \text{powoduje} \quad m = \hat{m} \quad \text{oraz} \quad x = \hat{x} \tag{4.20}$$

Zauważmy, że najprostszym wyborem wydaje się $c = m$, bowiem wówczas wartości współrzędnych stanu, jako zmiennych zależnych, będą określone równaniami procesu. Ponieważ jednak równania procesu są równaniami różniczkowymi, x zależeć będzie nie tylko od $c = m$, ale także od warunków początkowych. Zmodyfikujemy więc warunek (4.20) tak, aby żądać narzucenia trajektorii $m(t)$, $x(t)$ tylko dla $t \geqslant t_1 > 0$

$$c = \hat{c}(t) \quad \text{powoduje} \quad m = \hat{m}(t) \quad \text{oraz} \quad x = \hat{x}(t) \quad t \geqslant t_1 > 0 \tag{4.21}$$

niezależnie od warunków początkowych $x(0)$. Pozostanie oczywiście wymaganie sterowalności procesu, w sensie możliwości osiągnięcia stanu $\hat{x}(t_1)$

w chwili t_1, przy stanie początkowym $x(0)$; wymaganiem tym nie będziemy się tu zajmować.

Wymaganie (4.21) wyrażające możliwość narzucenia dowolnej trajektorii $x(t)$ jest równoważne – dla procesów opisywanych przez równania liniowe – wymaganiu stabilizacji stanu x wbrew działaniu dowolnych oddziaływań zewnętrznych $z(t)$. Powróćmy jednak do rozpatrywania stanu ustalonego. Załóżmy, że obieramy wektor c jako pewną funkcję wektorowych argumentów x, m

$$c = h(x, m) \tag{4.22}$$

Mamy równanie procesu w stanie ustalonym

$$f(x, m, z) = 0 \tag{4.23}$$

oraz nakładamy warunek (4.20) wyboru c, co sprowadza się do żądania, aby x, m były jednoznacznymi funkcjami c

$$x = \varphi^1(c), \quad m = \varphi^2(c) \tag{4.24}$$

Równania (4.22) i (4.23) są łącznie układem równań ze zmiennymi c, x, m. Warunek wystarczający istnienia i jednoznaczności funkcji (4.24) sprowadza się do żądania, aby wszystkie funkcje h_i, f_i były ciągłe wraz z pierwszymi pochodnymi oraz do warunku na rząd macierzy zbudowanej z pochodnych funkcji f_i i h_i względem ich argumentów

$$\text{rząd} \begin{bmatrix} \dfrac{\partial f_i}{\partial x_j} & \dfrac{\partial f_i}{\partial m_j} \\ \dfrac{\partial h_i}{\partial x_j} & \dfrac{\partial h_i}{\partial m_j} \end{bmatrix} = \dim x + \dim m \tag{4.25}$$

dla każdego x i m.

Liczba kolumn w macierzy (4.25) jest równa $\dim x + \dim m$, natomiast liczba wierszy wynosi $\dim f + \dim h$, przy czym $\dim f = \dim x$. Jeśli założymy, że wiersze $\begin{bmatrix} \dfrac{\partial f_i}{\partial x_j} & \dfrac{\partial f_i}{\partial m_j} \end{bmatrix}$ nie są liniowo zależne*), to wynika stąd, że liczba funkcji h_i, czyli liczba elementów wektora c musi wynosić

$$\dim c = \dim m \tag{4.26}$$

a ponadto spełniony ma być warunek (4.25). Warunek ten musi być spełniony dla wszystkich spodziewanych wartości parametrów z, opisujących możliwe oddziaływania zewnętrzne, czyli zakłócenia.

*) Można wykazać, przynajmniej dla procesów liniowych, że każdy proces sterowalny spełnia ten warunek.

Podajmy prosty przykład – niech równanie procesu ma postać

$$x - mz = 0$$

a jako wielkość regulowaną wybierzmy $c = x - m$. Stąd macierz (4.25) ma postać

$$\begin{bmatrix} 1 & -z \\ 1 & -1 \end{bmatrix}$$

a rząd jej wynosi jeden zamiast dwa, jeśli $z = 1$. Dla tej wartości z narzucenie liczby c nie określi wartości x, m. Dobór c był niewłaściwy.

Wprowadzone wymagania (4.25) ulegają uproszczeniu, jeśli elementami wektora c będą wprost niektóre elementy wektora x oraz niektóre elementy wektora m. Jeśli elementy pozostałe x, m, które nie weszły do c, oznaczymy przez d, to warunek jednoznaczności (4.20) sprowadza się do żądania, aby same tylko równania procesu określały d jako jednoznaczne funkcje c

$$d = \varphi(c)$$

Jeśli zapiszemy równania procesu z dokonaną zamianą zmiennych

$$f_i(c, d, z) = 0, \quad i = 1, \ldots, \dim x \tag{4.27}$$

to warunek jednoznaczności określenia d pociąga za sobą żądanie

$$\text{rząd} \left[\frac{\partial f_i}{\partial d_j} \right] = \dim d = \dim x \tag{4.28}$$

które jest odpowiednikiem (4.25) z przypadku ogólniejszego.

Zauważmy, że przy spełnionym warunku (4.28) możemy zadanie optymalizacji stanu ustalonego zastąpić zadaniem ekstremalizacji bezwarunkowej względem c; mianowicie zadanie

$$\max_{c,d} f_0(c, d), \quad \text{przy} \quad f(c, d) = 0$$

przechodzi w zadanie

$$\max_{c} f_0(c, \varphi(c))$$

Oznacza to właśnie, że wielkości c zostały wybrane właściwie: można je narzucić z zewnątrz układami stabilizacji i jednocześnie przez $c = \hat{c}$ zapewnić spełnienie zadania optymalizacji.

Macierz (4.28) powstaje z wybranych kolumn macierzy zbudowanej na podstawie równań stanu procesu

$$\left[\frac{\partial f_i}{\partial x_j}\frac{\partial f_i}{\partial m_j}\right]=\begin{bmatrix}\frac{\partial f_1}{\partial x_1}\cdots\frac{\partial f_1}{\partial x_n} & \frac{\partial f_1}{\partial m_1}\cdots\frac{\partial f_1}{\partial m_r}\\ \cdots\ \cdots\ \cdots & \cdots\ \cdots\ \cdots\\ \frac{\partial f_n}{\partial x_1}\cdots\frac{\partial f_n}{\partial x_n} & \frac{\partial f_n}{\partial m_1}\cdots\frac{\partial f_n}{\partial m_r}\end{bmatrix} \tag{4.29}$$

Praktycznie należy najpierw zestawić macierz (4.29), a następnie przez jej odpowiedni przegląd wydzielić podmacierz $(\partial f_i/\partial d_j)$, to znaczy wybrać zmienne c tak, aby zmienne pozostałe spełniały wymaganie (4.20).

Przedstawmy wybór wielkości wyjściowych procesu na przykładzie reaktora z rys. 3.6, przy założeniu, że mamy do dyspozycji cztery wielkości sterujące F_A, F_B, F_D, H. Równania stanu dla procesu w tym reaktorze mają postać podaną wzorami (3.11)-(3.13); pozwalają one określić macierz (4.29), która będzie miała kształt następujący

$$\left[\frac{\partial f_i}{\partial x_j}\frac{\partial f_i}{\partial m_j}\right]=\begin{bmatrix}0 & 0 & 0 & 1 & 1 & -1 & 0\\ a_{21} & a_{22} & a_{23} & b_{21} & b_{22} & 0 & 0\\ a_{31} & a_{32} & a_{33} & b_{31} & b_{32} & 0 & b_{34}\end{bmatrix}$$

przy czym jako wektor x przyjęto $x = [W\, C_A\, T\,]^T$, a jako wektor sterowań $m = [\,F_A\, F_B\, F_D\, H\,]^T$, por. rys. 3.6.

Przegląd kolumn macierzy $\left[\frac{\partial f_i}{\partial x_j}\frac{\partial f_i}{\partial m_j}\right]$ wskazuje, że ze względu na warunek (4.28) dopuszczalne są następujące zestawy:

zmienne swobodne (c)	zmienne zależne (d)
W, C_A, T, F_D	F_A, F_B, H
W, C_A, T, F_B	F_A, F_D, H
W, C_A, T, F_A	F_B, F_D, H
W, F_B, F_D, H	C_A, T, F_A
...	...

Zwróćmy raz jeszcze uwagę na interpretację fizyczną; jeśli narzucimy wartości zmiennych swobodnych, np. W, C_A, T, F_B, to zmienne zależne czyli pozostałe, wynikną jednoznacznie poprzez równania obiektu.

Rysunek 4.2 podawał strukturę układu regulacji reaktora, wykorzystującą jako wielkości regulowane W, C_A, T, F_B.

Podkreślmy, że w rozpatrywanym przykładzie nie jest np. zestawem dopuszczalnym $c = [\,F_A\, F_B\, F_D\, H\,]^T$, bowiem wówczas macierz

$$\frac{\partial f_i}{\partial d_j}=\begin{bmatrix}0 & 0 & 0\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{bmatrix}$$

nie ma rzędu $\dim d$. Oznacza to, że narzucenie F_A, F_B, F_D, H nie określa jednoznacznie stanu ustalonego procesu w reaktorze.

Interesujący dla wielu sytuacji rzeczywistych przypadek powstaje gdy $\dim m < \dim x$. Przyjmijmy, że dla reaktora z rys. 3.6 mamy tylko dwie wielkości sterujące: F_A oraz H. Mamy natomiast trzy współrzędne stanu, W, C_A oraz T. Pozostałe wielkości wejściowe F_B, F_D, T_A, T_B będą zakłóceniami.

Poszukajmy odpowiedzi na pytanie: jak można wybrać wielkości regulowane c? Zestawmy macierz

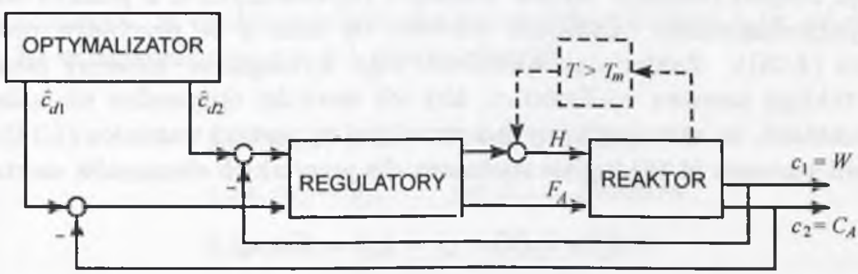
$$\left[\frac{\partial f_i}{\partial x_j}\ \frac{\partial f_i}{\partial m_j}\right] = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ a_{21} & a_{22} & a_{23} & b_{21} & 0 \\ a_{31} & a_{32} & a_{33} & b_{31} & b_{34} \end{bmatrix}$$

Przegląd tej macierzy wskazuje, że można wybrać następujące zestawy zmiennych, spełniające warunki (4.26) oraz (4.28)

zmienne swobodne (c)	zmienne zależne(d)
W, C_A	T, F_A, H
W, T	C_A, F_A, H
C_A, T	W, F_A, H
W, H	C_A, T, F_A
C_A, H	W, T, F_A
T, H	W, C_A, F_A

Możemy więc np. przyjąć: $c = [\,W\,C_A\,]^T$.

Przyjrzyjmy się jeszcze skutkom, jakie $\dim m < \dim x$ przyniesie dla zadania optymalizacji i dla realizacji sterowania. Załóżmy, że na podstawie analizy macierzy $\left[\frac{\partial f_i}{\partial x_j}\ \frac{\partial f_i}{\partial m_j}\right]$ przyjęliśmy $c = [\,W\,C_A\,]^T$; powstał układ dwuwarstwowy jak na rys. 4.4.



Rys. 4.4. Reaktor o dwóch wielkościach sterujących i jego układ sterowania

Zauważmy, że mając do dyspozycji dwie wielkości sterujące F_A oraz H, zamiast wszystkich czterech istotnych wielkości sterujących F_A, F_B, F_D, H będziemy osiągać wartości Q mniejsze od poprzedniego $\hat{Q}$, czyli wydajność reaktora mniejszą od maksymalnie możliwej przy danych jego rozmiarach. Zauważmy ponadto, że mając $\dim m < \dim x$ nie jesteśmy w stanie stabilizować stanu reaktora, tj. utrzymać stałości wszystkich współrzędnych stanu W, C_A, T wbrew dowolnym zakłóceniom. Obrane rozwiązanie (rys. 4.4) stabilizuje W (zawartość masy) oraz C_A (stężenie), natomiast temperatura mieszaniny T będzie musiała podlegać wahaniom. Jeśli dla odbiorcy strumienia F_D ważniejsza będzie stałość jego temperatury niż stałość stężenia, należy za wielkości regulowane obrać W oraz T. Jeśli ważna jest zarówno temperatura, jak i stężenie, obierzemy za wielkości regulowane C_A, T. Wówczas zawartość reaktora W będzie musiała wahać się wraz z zakłóceniami, co oczywiście wpłynie na spadek wydajności: średnia wartość $W(t)$ będzie niższa niż W_m, co odpowiadałoby reaktorowi o mniejszych rozmiarach. Widzimy na tym przykładzie, jak ważne dla zadania optymalizacji oraz dla stabilizacji procesu jest posiadanie odpowiednich wielkości sterujących.

Wartości zadane optymalne $c_d = \hat{c}_d$ musielibyśmy wziąć z rozwiązania zadania optymalizacji; można wyliczyć, że będzie to

$$\hat{c}_{d1} = W_m, \qquad \hat{c}_{d2} = \frac{F_D(t) - F_B(t)}{W_m k(T_m) + F_D(t)}$$

Jeśli rozwiązania zadania optymalizacji procesu są poprawne, to tak długo jak utrzymamy $c = \hat{c}_d$ nie nastąpi naruszenie żadnego z ograniczeń nierównościowych, w szczególności także ograniczenia $T \leqslant T_m$. W tym celu trzeba utrzymać $\hat{c}_d$, czyli bieżąco zmieniać c_d zależnie od zakłóceń F_B, F_D. Nie zapominajmy, że $\hat{c}_d$ wylicza się na podstawie równań procesu, zatem może ono być obarczone błędem; w szczególności może nastąpić przekroczenie ograniczenia na temperaturę. Jeśli chcemy się przed tym zabezpieczyć, możemy do układu regulacji na rys. 4.4 dodać elementy zabezpieczenia przed przekroczeniem niekontrolowanego dotąd bezpośrednio ograniczenia. Należałoby w rozpatrywanym przykładzie mierzyć temperaturę w reaktorze oraz oddziaływać na sterowanie H lub na wyznaczoną przez optymalizator wartość zadaną c_{d2} w kierunku zmniejszenia dopływu ciepła w razie przekroczenia temperatury. Zaznaczono to na rys. 4.4, wprowadzając dodatkowy regulator, którego decyzja ma zmajoryzować decyzję regulatora sterującego wielkością H w stanie normalnej pracy. Zwróćmy uwagę, że w układzie na rys. 4.4 mamy trzy wielkości mierzone (regulowane), a tylko dwie wielkości sterujące – zgodnie z założeniem na początku tego przykładu, że do dyspozycji mamy tylko wielkości F_A oraz H.

Niezależność od zakłóceń

Dotąd rozpatrywaliśmy wybór wielkości regulowanych c z punktu widzenia jednoznaczności określenia wartości m oraz x w przebiegu procesu (patrz (4.25)). Zakładając spełnienie tego wymagania, możemy poszukiwać takiego zestawu wielkości c_j, aby ich wartości optymalne nie zależały od zakłóceń, co zaznaczyliśmy już wcześniej w postaci warunku (4.18).

Jeśli warunek (4.18) będzie spełniony dla wszystkich elementów wektora c

$$\hat{c}_j \neq \hat{c}_j(z), \quad j = 1, \ldots, \dim c,$$

to układy regulacji (stabilizacji w przypadku optymalizacji stanu ustalonego) zapewnią optymalność sterowania i nie będzie potrzebna optymalizacja bieżąca.

Sytuacja taka wystąpiła w przykładzie omówionym w rozdz. 4.3. Dla reaktora z rys. 3.6 wartości optymalne wybranych wielkości regulowanych $\hat{W}$, $\hat{C}_A$, $\hat{T}$, $\hat{F}_B$ nie zależały od czynników zewnętrznych (T_A, T_B) i parametrów procesu (stała prędkości reakcji k i in.), natomiast pozostałe zmienne zadania cechy tej nie miały; na przykład optymalny dopływ ciepła $\hat{H}$ zależał od T_A, T_B i innych wielkości. Reaktor zaopatrzony we wskazane cztery regulatory będzie zachowywał stan i sterowanie optymalne, $x = \hat{x}$ oraz $m = \hat{m}$ bez udziału optymalizatora. Stało się tak dlatego, że rozwiązania zadania optymalizacji leżą na brzegu obszaru ograniczeń, przy tym dla wszystkich wartości z w tym samym punkcie tego brzegu.

Rozpatrzmy optymalizację stanu ustalonego dla procesów opisywanych równaniami różniczkowymi zwyczajnymi

$$\begin{aligned} &\text{maksymalizować } Q(x,m,z) = f_0(x,m,z), \\ &\text{przy } f_i(x,m,z) = 0, \quad i = 1,\dots,\dim x \end{aligned}$$

przy ograniczeniach nierównościowych niezależnych od czynników zewnętrznych

$$g_j(x,m) \geqslant b_j, \quad j = 1,\dots$$

Przyjmijmy, że w rozwiązaniu $\hat{x}(z)$, $\hat{m}(z)$ aktywnych jest k ograniczeń nierównościowych, to znaczy że $\hat{x}(z)$, $\hat{m}(z)$ spełniają oprócz równań procesu równania

$$g_j(\hat{x}(z),\hat{m}(z)) = b_j, \quad j = 1,\dots,k, \tag{4.30}$$

przy czym zachodzi to dla wszystkich wartości z, tzn. aktywne są stale te same ograniczenia.

Wybór wielkości regulowanych polega na ustaleniu funkcji

$$c_j = h_j(x,m), \quad j = 1,\dots,\dim m. \tag{4.31}$$

Spełnienie przez x, m równań procesu oraz równań (4.31) będzie narzucone; jeśli ma być $x = \hat{x}$ oraz $m = \hat{m}$, to spełnione mają być ponadto (4.30). Mamy więcej równań niż zmiennych, co wymaga aby część z nich była liniowo zależna. W ogólnym przypadku mamy zatem funkcje $h_j(x,m)$ ustalić tak, by w układzie $\dim x + k + \dim m$ równań

$$\begin{aligned} f_i(x,m,z) &= 0, && i = 1,\dots,\dim x \\ g_j(x,m) &= b_j, && j = 1,\dots,k \dim m \\ h_j(x,m) &= c_j, && j = 1,\dots,\dim m \end{aligned}$$

było dokładnie $\dim x + \dim m$ równań niezależnych. Jeśli f_i, g_j są układami niezależnych od siebie funkcji x, m, to możemy tylko wprowadzić h_j zależne od g_j. Najprościej przyjąć identyczność $h_j(x,m) \equiv g_j(x,m)$, $j = 1,\dots,k$, a w takim przypadku tylko pozostała część funkcji h_j jest nowa i nasz układ przybiera postać

$$\begin{aligned} f_j(x,m,z) &= 0, && i = 1,\dots,\dim x \\ g_j(x,m) &= b_j = c_j, && j = 1,\dots,k \\ h_j(x,m) &= c_j, && j = k+1,\dots,\dim m \end{aligned} \tag{4.32}$$

Układ ten mieć będzie jednoznaczne rozwiązanie dla x, m, jeśli spełniony będzie warunek (4.25). Stabilizacja wielkości regulowanych c zastąpi optymalizację, jeśli rozwiązaniem tym będzie właśnie $x = \hat{x}(z)$, $m = \hat{m}(z)$. Pozostaje nam wskazać, jak powinny być zbudowane funkcje $h_j(x, m)$, $j = k+1, \ldots, \dim m$, występujące w układzie (4.32). Funkcje te muszą być takie, aby $\hat{x}$, $\hat{m}$, spełniały układ równań

$$h_j(\hat{x}(z), \hat{m}(z)) = c_j, \quad j = k+1, \ldots, \dim m$$

z tym, że układ ten ma być niezależny od z w punkcie $\hat{x}(z)$, $\hat{m}(z)$ dla każdej wartości z

$$\frac{\mathrm{d}h_j(\hat{x}(z), \hat{m}(z))}{\mathrm{d}z} = 0, \quad j = k+1, \ldots, \dim m \tag{4.33}$$

Rozważania te wskazują, iż istnienie zestawu wielkości regulowanych przy którym wartość optymalna c będzie niezależna od z wymaga spełnienia następujących warunków:

a) równania procesu są zależne od z, w przeciwnym przypadku układ (4.32) nie mógłby mieć rozwiązań x, m zależnych od z,

b) równania spełnionych ograniczeń nierównościowych g_j, $j = 1, \ldots, k$, są niezależne od z oraz te same dla wszystkich z,

c) dobierane funkcje h_j, $j = k+1, \ldots, \dim m$, spełniają warunek (4.33),

d) funkcje f_i, g_j, h_j spełniają warunek (4.25).

Przykład opisany w rozdz. 4.3, odnoszący się do reaktora z rys. 3.6 był taki, że spełnione były cztery ograniczenia nierównościowe, dając równości typu (4.30), przy czym były one niezależne od zakłóceń

$$\hat{W} = W_m, \quad \hat{C}_A = C_{Am}, \quad \hat{T} = T_m, \quad \hat{F}_B = F_{Bm}$$

Na tej podstawie wybrano wielkości regulowane c_j według reguły, aby układ równań (4.31) był identyczny z układem spełnionych ograniczeń nierównościowych: $c_1 = W$, $c_2 = C_A$ itd. Układ spełnionych ograniczeń (4.30) dostarczył przy tym od razu optymalnych wartości wielkości regulowanych.

Znalezienie takiego zestawu wielkości c, aby wartość $\hat{c}$ nie zależała od z nie zawsze jest możliwe. Wystarczy, aby równania procesu nie zawierały z albo przy różnych z spełnione były inne ograniczenia – a już nie da się spełnić warunku inwariantności $\hat{c}$.

W przypadkach takich mamy dwie możliwości. Pierwsza z nich polega na wybraniu takiego c, aby możliwie wiele jego elementów miało wartości optymalne niezależne od z, $\hat{c}_j \neq \hat{c}_j(z)$. Wariant ten jest dobry w sytuacji, gdy mamy zamiar zastosować optymalizację bieżącą, bowiem optymalizacja ta będzie zadaniem o mniejszej wymiarowości (dotyczyć będzie tylko tych elementów c_j, gdzie $\hat{c}_j$ zależy od z). Druga możliwość ma sens przy braku optymalizacji bieżącej – wtedy warto znaleźć takie c, aby $c = \hat{c} = \mathrm{const}$ dawało możliwie najlepszą wartość funkcji celu.

Rozpatrzmy najpierw pierwsze rozwiązanie. Zastosowaliśmy je w istocie w przypadku maksymalizacji wydajności reaktora przy użyciu tylko dwóch wielkości sterujących (rys. 4.4). Optymalna wartość zapełnienia $\hat{W} = W_m$ jest stała i optymalizator nie musiałby jej wyliczać, przedmiotem optymalizacji byłaby więc jedna zmienna zamiast dwóch.

Biorąc ogólnie, mamy taką sytuację, że liczba aktywnych ograniczeń nierównościowych niezależnych od zakłóceń jest mniejsza od $\dim m$

$$g_j(x, m) = b_j, \quad j = 1, \ldots, k < \dim m \tag{4.34}$$

i z tych ograniczeń może wynikać wybór k elementów wektora c, $c_j = c_j(x, m)$, jak to opisywaliśmy wcześniej. Tak jak przedtem, założymy, że ograniczenia powyższe są aktywne przy wszystkich wartościach z. Dzięki temu k elementów wektora $\hat{c}$ może być stałych. Pozostałe natomiast elementy $\hat{c}$ będą zależne od czynników zewnętrznych z, czyli będą wymagać nastawiania w trakcie biegu procesu. Możliwe jest przy tym, że będą aktywne dalsze ograniczenia nierównościowe, ale zależne od z

$$g_j(x, m, z) = b_j, \quad j = k + 1, \ldots, \dim m \tag{4.35}$$

co dałoby w zasadzie dość jawne zależności $\hat{c}_j$ od z, ale możliwe jest też, że dla różnych z aktywne będą różne zestawy ograniczeń albo też, że nie będą aktywne żadne z ograniczeń. Wymaga to bardziej skomplikowanych algorytmów działania optymalizującego.

Interesujący jest przypadek, gdy na przykład wiadomo z badania rozwiązania zadania optymalizacji, że spośród wszystkich ograniczeń nierównościowych zawsze aktywnych jest $\dim m$ ograniczeń, ale w różnych kombinacjach dla różnych wartości z. Niech na przykład będą dwa ograniczenia

$$g_1(x, m) \geqslant b_1$$
$$g_2(x, m) \geqslant b_2$$

przy jednej wielkości sterującej m i wiadomo, że w punkcie $\hat{x}(z)$, $\hat{m}(z)$ aktywne jest zawsze albo jedno, albo drugie z tych ograniczeń. Prowadzi to do koncepcji dwóch układów stabilizacji, zbudowanych dla wielkości

$$c_1 = g_1(x, m)$$
$$c_2 = g_2(x, m)$$

przy czym to, który z nich jest włączony, zależy od wartości czynnika zewnętrznego.

Przejdźmy teraz do drugiej sytuacji – zamierzamy poprzestać na układach stabilizacji, $c(t) = \text{const}$; interesuje nas zatem taki wybór strukturalny, to jest wybór funkcji $h(x,m)$, ażeby osiągnąć najlepszą możliwą w tych warunkach wartość funkcji celu. Trzeba przy tym uwzględnić losowy charakter czynnika zewnętrznego z, stąd też – po wyborze $h(x,m)$ – zadanie optymalizacyjne mieć będzie postać

$$\underset{c}{\text{maksymalizować}} \ \underset{z}{E}\ Q(x,m,z) \tag{4.36}$$

przy warunkach $f(x,m,z) = 0$ oraz $c = h(x,m)$.

Funkcje $h(x,m)$ wolno nam wybrać, mając na widoku najlepszy rezultat zadania (4.36), ale nie ma tu pełnej dowolności. Spełniony być musi warunek (4.25), stanowiący o jednoznacznym określaniu, w działającym układzie, wartości x oraz m przez narzucanie c. Mieć zatem będziemy $x = \varphi^1(c,z)$ oraz $m = \varphi^2(c,z)$. Zadanie (4.36) otrzymuje wówczas postać

$$\underset{c}{\text{maksymalizować}} \ \underset{z}{E}\ Q(\varphi^1(c,z), \varphi^2(c,z), z) \tag{4.37}$$

Warunek (4.25) można wbudować w mechanizm optymalizacyjny, na przykład w postaci wymagania, by wyznacznik macierzy (4.25) spełniał pewne kryterium ilościowe

$$\begin{vmatrix} \dfrac{\partial f_i}{\partial x_j} & \dfrac{\partial f_i}{\partial m_j} \\ \dfrac{\partial h_i}{\partial x_j} & \dfrac{\partial h_i}{\partial m_j} \end{vmatrix}^2 \geqslant \delta^2 \tag{4.38}$$

Tytułem przykładu załóżmy, że funkcja $h(x,m)$ zbudowana jest z pewnych funkcji standardowych

$$c = \alpha_1 h^1(x,m) + \alpha_2 h^2(x,m) + \ldots \tag{4.39}$$

gdzie współczynniki $\alpha_1, \alpha_2, \ldots$ można dobrać. Oznaczając wartość zadaną dla c przez c_d możemy teraz sformułować zadanie (4.36) jako

$$\underset{c_d,\alpha_1,\alpha_2,\ldots}{\text{maksymalizować}} \ \underset{z}{E}\ Q(\varphi^1(c_d,z,\alpha_1,\alpha_2,\ldots), \varphi^2(c_d,z,\alpha_1,\alpha_2,\ldots), z) \tag{4.40}$$

z uwzględnieniem nierówności (4.38). W wyniku otrzymamy zarówno postać funkcji $h(x,m)$ jak optymalną wartość – podlegających stabilizacji – wielkości regulowanych.

Podsumujemy krótko omówione w tym rozdziale aspekty wyboru wielkości regulowanych:

1) konieczny jest taki wybór c, aby spełnić warunek jednoznaczności określania x, m (4.20), przedstawiony jako kryterium (4.25);

2) pośród c spełniających (4.20) można szukać wyboru optymalnego:
   a) jeśli przewiduje się bieżącą optymalizację, to poszukiwanie idzie w kierunku c spełniających warunek (4.18) dla możliwie największej liczby elementów (redukcja wymiaru zadania optymalizacji),
   b) jeśli nie przewiduje się bieżącej optymalizacji, wybór c należy prowadzić w kierunku wymagania (4.37) (optymalna stabilizacja).

Przedstawimy teraz te same stwierdzenia w sposób bardziej jakościowy. Mówimy o „wyborze wielkości regulowanych", mamy więc na myśli budowę układów regulacji – czyli nałożenie na proces sterowany pewnych reguł postępowania, podporządkowanych zadaniom typu „utrzymać stałość wielkości c", czy też „niech wielkość c nadąża za narzucanym c_d". Rzecz w tym, by spełnienie zadania postawionego dla c wywołało wewnątrz obiektu sterowanego pożądane przez nas przebiegi procesu. W związku z tym:

1) wielkości c trzeba tak wybrać, by ich dotrzymanie wywoływało bieg procesu jednoznacznie określony, nie pozostawiało swobody w odniesieniu do tych zmiennych, które są istotne,
2) można ponadto poszukiwać takich c, aby zredukować potrzebę interwencji (tj. potrzebę zmiany wartości zadanych) bądź też – rezygnując z interwencji – ponieść jak najmniejsze straty w przypadku działających na proces czynników zewnętrznych.

4.5. Częstotliwość interwencji i modele statyczne

Niektóre właściwości struktury dwuwarstwowej, w której sterowanie procesem dynamicznym podzielono między warstwę sterowania bezpośredniego o prostym zadaniu stabilizacji bądź nadążania oraz warstwę wyznaczania zadań były już widoczne poprzednio, zwłaszcza w przykładzie opisanym w rozdz. 4.3.

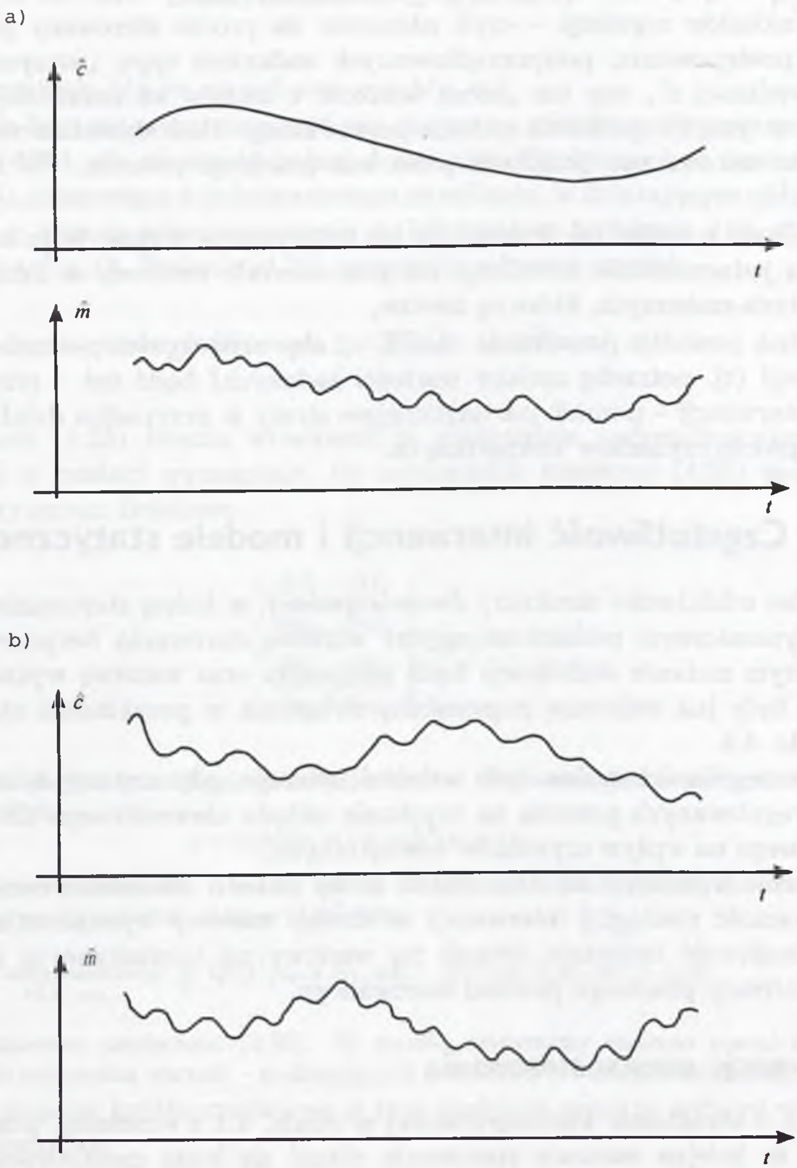
W szczególności można było wskazać sytuacje, gdy zręczny dobór wielkości regulowanych pozwala na uzyskanie układu niewrażliwego albo mniej wrażliwego na wpływ czynników zewnętrznych.

Obecnie wskażemy na dwie dalsze cechy układu dwuwarstwowego: dopuszczalność nieciągłej interwencji ze strony warstwy wyznaczania zadań oraz możliwość opierania decyzji tej warstwy na uproszczonym zadaniu optymalizacji przebiegu procesu sterowanego.

Interwencje warstw sterowania

Mówiąc o strukturze wielowarstwowej w rozdz. 4.1 i wcześniej, podkreślaliśmy, że kolejne warstwy sterowania różnić się będą częstotliwością interwencji. Rozpatrzymy teraz bliżej tę kwestię w odniesieniu do układu dwuwarstwowego.

Załóżmy, że dla procesu sterowanego przez układ dwuwarstwowy obliczyliśmy przebieg czasowy optymalnego sterowania $\hat{m}$ i optymalnej wartości $\hat{c}$, które odpowiadają pewnym konkretnym przebiegom czynników zewnętrznych, działających na ten proces albo występujących w zadaniu optymalizacji, rys. 4.5.



Rys. 4.5. Hipotetyczne przebiegi czasowe wartości optymalnych m oraz c

W przypadku jak na rys. 4.5a przebieg $\hat{c}$ jest bardziej powolny aniżeli przebieg $\hat{m}$, co oznacza, że optymalizacja skierowana na ustalanie $\hat{c}$ mogłaby być dokonywana rzadziej niż miałoby to miejsce w przypadku optymalizacji bezpośredniej sterowania m.

Sytuacja jak na rys. 4.5a zaistnieje wówczas, gdy wartość optymalna m zależy od wszystkich czynników zewnętrznych, a wartość optymalna c tylko od niektórych spośród nich, przy czym są to wielkości zmieniające się powoli. Nie będzie to przypadek rzadko spotykany – widzieliśmy już wcześniej, że wartość optymalna c może w ogóle nie zależeć od czynników zewnętrznych, możliwe jest zatem że będzie zależeć tylko od niektórych.

W odróżnieniu od tego, w przypadku jak na rys. 4.5b, wyznaczanie $\hat{c}$ musiałoby być równie częste jak wyznaczanie $\hat{m}$ i układ dwuwarstwowy może być niewłaściwy; optymalizator musiałby interweniować równie często jak układy sterowania bezpośredniego. Będzie on jednak nadal racjonalny, jeśli $\hat{c}$ będzie łatwiejsze do wyznaczania niż $\hat{m}$.

Uproszczenie zadania optymalizacji

Rozpatrywaliśmy korzyści, związane z częstotliwością interwencji optymalizatora – być może niższą, niż to byłoby potrzebne w przypadku optymalizacji bezpośredniej. Spojrzymy teraz na samo zadanie optymalizacji, zmierzając do wykrycia jego możliwych uproszczeń w przypadku, gdy zadanie to dotyczyć będzie wielkości c zamiast m.

Nie będziemy zajmować się sytuacją, znaną z rozdz. 3.2, gdy wartości optymalne współrzędnych stanu procesu są stałe i wystarczy je wyznaczyć korzystając z zadania optymalizacji statycznej, tylko jeden raz. Rozważymy przypadek bardziej ogólny, proces dynamiczny o nie narzuconym horyzoncie sterowania, zatem taki, dla którego horyzont ten wynika z rozważań przeprowadzonych w rozdz. 3.4. Mamy więc zadanie typu

$$\text{maksymalizować } Q = \int_{t_0}^{t_1} f_0\left(x(t), m(t), z^*(t)\right) \mathrm{d}t \tag{4.41}$$

z równaniem procesu

$$\dot{x}(t) = f\left(x(t), m(t), z^*(t)\right) \tag{4.42}$$

i warunkami brzegowymi: $x(t_0)$ dane, $x(t_1)$ dane i równe $\hat{x}(t_1)$.

Przypomnijmy, że z^* oznacza predykcję czynnika zewnętrznego, a $\hat{x}(t_1)$ zadany optymalny stan na krańcu horyzontu sterowania $[t_0, t_1]$, wyznaczony z odpowiedniego zadania sformułowanego dla przedziału $[t_1, t_2]$ i mającego rozwiązanie $\hat{x}_{[t_1,t_2]}$ stałe bądź okresowe. Wspomnieliśmy na końcu rozdz. 3.4 o sposobie sterowania, polegającym na przesuwaniu horyzontu: w chwili $t_0 + \delta$ powtarzamy zadanie (4.41), (4.42), sięgając do chwili $t_1 + \delta$, a wychodząc ze znanych, nowych wartości $x(t_0 + \delta)$, $z^*_{[t_0+\delta, t_1+\delta]}$.

Rozpatrzmy teraz rozwiązanie przybliżone dla optymalizacji z przesuwanym horyzontem. Przyjmijmy mianowicie, że trajektoria optymalna jest realizowana w układzie dwuwarstwowym, w którym stan $x(t)$ ma nadążać za zadanym $x_d(t)$. Załóżmy, dla uproszczenia, że układy sterowania bezpośredniego nie są dynamiczne, tj. realizują statyczne prawo sterowania

$$m(t) = r\left(x_d(t), x(t)\right)$$

dzięki czemu zadanie dynamiczne (4.41), (4.42) można przeformułować na zadanie o tym samym charakterze, nie zawierające zmiennej m

$$\text{maksymalizować } Q = \int_{t_0}^{t_1} f_0\left(x(t), x_d(t), z^*(t)\right) \mathrm{d}t \tag{4.43}$$

z równaniem procesu

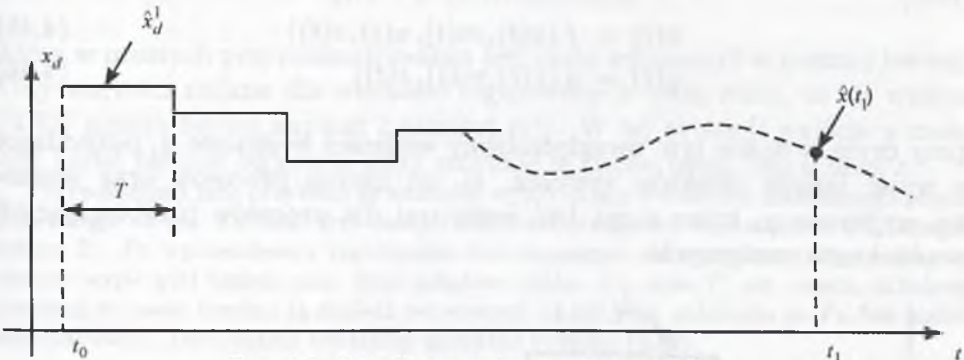
$$\dot{x}(t) = f\left(x(t), x_d(t), z^*(t)\right) \tag{4.44}$$

i warunkami brzegowymi: $x(t_0)$ dane, $x(t_1) = \hat{x}(t_1)$ dane. Charakter zadania (4.43), (4.44) jest taki sam jak (4.41), (4.42), jedynie funkcje $f_0(\cdot)$, $f(\cdot)$ będą nieco inne. Rolę sterowania m przejęło x_d.

Weźmy pod uwagę przypadek, gdy zamierzamy zmieniać x_d tylko w pewnych odstępach czasu T, utrzymując x_d stałe wewnątrz tych przedziałów. Ponieważ x_d odgrywa rolę sterowania w obecnym zadaniu dynamicznym (4.43), (4.44), sytuacja nasza odpowiada nałożeniu ograniczenia „m odcinkami stałe" na wybór sterowania $m_{[t_0,t_1]}$ w pierwotnym zapisie, tj. w zadaniu (4.41), (4.42).

Zauważmy, że $\hat{x}^1_d$, to znaczy wartość $\hat{x}_d$ dla pierwszego przedziału o długości T (rys. 4.6) będąca liczbą (lub wektorem liczb), będzie musiała być wyznaczona za pomocą procedury optymalizacji dynamicznej dyskretnej, z horyzontem $[t_0, t_1]$ i czasem dyskretyzacji T. Rozwiązanie określi całą trajektorię $\hat{x}_d$, zawierać więc będzie także zależność wartości $\hat{x}^1_d$ od stanu początkowego $x(t_0)$, stanu końcowego zadanego $\hat{x}(t_1)$, wartości $z(t_0)$ i tych cech probabilistycznych zakłócenia, które łącznie z obserwacją $z(t_0)$ określają predykcję $z^*_{[t_0,t_1]}$.

Będzie to zależność $\hat{x}^1_d$ od wektorów $x(t_0)$, $\hat{x}(t_1)$, $z(t_0)$. Można próbować ją wyrazić modelem statycznym, np. drogą aproksymacji wyników rozwiązania zadania dynamicznego dla różnych $x(t_0)$, $\hat{x}(t_1)$, $z(t_0)$, a następnie używać tej aproksymacji do sterowania bieżącego. Będzie to model „prawa sterowania", reguła decyzyjna (statyczna), a nie model statyczny samego procesu; nie odstępujemy tu od zadania dynamicznego.



Rys. 4.6. Schodkowo zmienne „sterowanie" x_d

Warto zauważyć, że przy dużej wymiarowości x oraz z liczba obliczeń potrzebnych do sporządzenia tabeli wyników, a następnie statycznej reguły decyzyjnej może być tak duża, że wykonywanie zadania dynamicznego na bieżąco, zatem tylko dla tych wartości $x(t_0)$, $\hat{x}(t_1)$, $z(t_0)$, które istotnie się zdarzą, może okazać się bardziej racjonalne.

Znacznie ważniejsza dla praktyki jest sytuacja, gdy sam proces dynamiczny można – do celów sterowania – opisać *modelem statycznym*, a w ślad za tym – optymalizacja biegu procesu jest optymalizacją statyczną. Sytuacja taka powstaje, kiedy:

– pożądana trajektoria stanu procesu x_d zadawana jest jako wielkość odcinkami stała,
– w obrębie jednego takiego odcinka stan rzeczywisty procesu jest także, przez prawie cały czas, stały (dzięki sprawnemu nadążaniu za wartością zadaną).

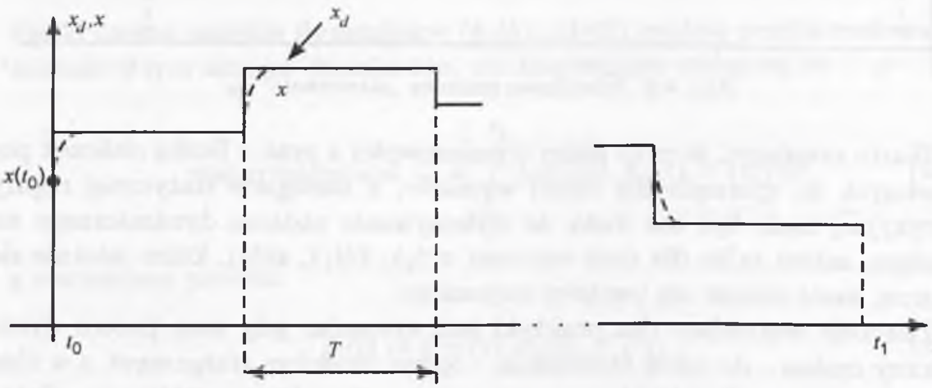
Przypadek taki ilustruje rys. 4.7. Powstać on może wówczas, gdy zmienność trajektorii pożądanej nie jest podyktowana dynamiką procesu, lecz wynika głównie ze zmian czynnika zewnętrznego, przy czym zmiany te są – w stosunku do dynamiki procesu – powolne. Inaczej mówiąc, horyzont decyzji podejmowanej w chwili t_0 i dotyczącej ciągu wartości x_d sięga chwili t_1 nie dlatego, że tak daleko sięga wpływ wartości stanu początkowego na trajektorię optymalną, lecz dlatego, że tak daleko sięga korelacja przewidywanego przebiegu z, tj. czynnika zewnętrznego z wartością początkową $z(t_0)$; wtedy właśnie, dzieląc $[t_0, t_1]$ na pewną liczbę (niezbyt krótkich) odcinków T otrzymamy obraz działania układu taki jak na rys. 4.7. Można wówczas przyjąć, że proces będzie się znajdował w kolejnych, następujących po sobie stanach ustalonych. Rozpatrzmy tę sytuację nieco bliżej celem wykazania, że optymalizację można przeprowadzać dla każdego przedziału T oddzielnie.

Mamy do czynienia z procesem dynamicznym, opisanym przez równanie stanu i równanie wyjścia

$$\dot{x}(t) = f\left(x(t), m(t), u(t), z(t)\right) \tag{4.45}$$

$$y(t) = g\left(x(t), m(t), z(t)\right) \tag{4.46}$$

przy czym w opisie tym uwzględniliśmy wielkości wejściowe u, pochodzące z wyjść innych obiektów systemu, tj. od innych procesów oraz wielkości wyjściowe y, które mogą być wejściami dla procesów przebiegających w obiektach następnych.



Rys. 4.7. Szybko reagujący proces nadąża za odcinkami stałym x_d

Zakładając stan ustalony, tj. przyjmując $\dot{x} = 0$, otrzymamy model *statyczny procesu* (4.45) (4.46)

$$f\left(\bar{x}, m(t), u(t), z(t)\right) = 0 \tag{4.47}$$

$$y(t) = g\left(\bar{x}, m(t), u(t), z(t)\right) \tag{4.48}$$

gdzie $\bar{x}$ oznacza ustaloną wartość stanu (w danym przedziale czasu). Model ten jest opisem samego procesu, bez układów sterowania bezpośredniego. Załóżmy teraz istnienie regulatorów oraz odpowiedni wybór wielkości regulowanych (por. rozdz. 4.4)

$$c = h(x, m) \tag{4.49}$$

dzięki czemu związki (4.47)–(4.49) określają jednoznaczne zależności $m = \varphi^2(c, z)$ oraz $\bar{x} = \varphi^1(c, z)$. Teraz model statyczny (4.47), (4.48) samego procesu przechodzi w model statyczny procesu wraz z regulatorami, czyli w model procesu widzianego przez wyższą warstwę sterowania

$$f\left(\varphi^1(c(t), z(t)), \varphi^2(c(t), z(t)), u(t), z(t)\right) = 0 \tag{4.50}$$

$$y(t) = g\left(\varphi^1(c(t), z(t)), \varphi^2(c(t), z(t)), u(t), z(t)\right) \tag{4.51}$$

Model (4.50), (4.51) opisuje w istocie zależność

$$y(t) = F\left(c(t), u(t), z(t)\right) \tag{4.52}$$

którą w prostych przypadkach można być może wyznaczyć w postaci jawnej. Gdy wartości zadane dla wielkości regulowanych będą stałe, to we wzorze (4.52) można będzie napisać $\bar{c}$ zamiast $c(t)$. W tej sytuacji wyjście y może być nadal zależne od czasu, gdyż zmienne w czasie mogą być u oraz z.

Przypomnijmy jako przykład, że strumień wypływający z reaktora chemicznego przedstawionego na rys. 3.6 miał trzy cechy: natężenie przepływu F_D, stężenie C_A, temperaturę T. Po wprowadzeniu regulatorów utrzymujących stan ustalony trójwymiarowy wektor wyjść $y(t)$ będzie miał dwie składowe stałe: $\bar{C}_A$ oraz $\bar{T}$, ale trzecią składową zmienną w czasie (można ją znaleźć we wzorach (4.15) przy założeniu że F_B nie będzie stabilizowane). Otrzymamy konkretny przykład związku (4.52)

$$y(t) = \begin{bmatrix} \bar{W} \\ \bar{C}_A \\ (1 - \bar{C}_A)^{-1}[\bar{W}\bar{C}_A k(\bar{T}) + F_B(t)] \end{bmatrix}$$

Przejdźmy teraz do funkcjonowania warstwy wyznaczającej stałe wartości zadane dla regulatorów, czyli określającej optymalne wartości $c_d(t) = \bar{c}$. Niech podstawą będzie zadanie optymalizacji procesu w przedziale czasu o długości T

$$\text{maksymalizować } Q = \frac{1}{T}\int_0^T f_0(\bar{c}, u(t), y(t))\mathrm{d}t \tag{4.53}$$

Sformułowanie (4.53) nie jest najbardziej ogólne, przyjęliśmy w nim mianowicie, że koszty bądź zyski przypisywane są tylko wielkościom wejściowym i wyjściowym oraz decyzjom sterującym $\bar{c}$. Koszt przypisywany decyzji $\bar{c}$ ma zatem reprezentować także koszt uzyskania stanu $\bar{x}$ i wytworzenia sterowań samego procesu m.

Zadaniu maksymalizacji (4.53) towarzyszą ograniczenia równościowe w postaci modelu (4.52) – wyrażonego jawnie lub jako układ równań (4.50), (4.51). Zadanie to nie jest zadaniem dynamicznym (nie ma więzów różniczkowych ani też troski o stan początkowy i końcowy), ale nie jest to także zwykłe zadanie optymalizacji statycznej: we wzorach (4.52) i (4.53) występuje zależność od czasu.

Zauważmy, że chodzi o wyznaczenie – w chwili zerowej – optymalnego $\bar{c}$ na przyszłość, to jest na przedział o długości T. Występujące w sformułowaniu (4.52), (4.53) funkcje czasu stanowią tu istotną trudność. Mamy w zasadzie trzy możliwości:

– użyć predykcji $u^*_{[0,T]}$ oraz $z^*_{[0,T]}$ i wykonać obliczenie korzystając z tej funkcji czasu,

– znając cechy probabilistyczne u oraz z przeformułować zadanie (4.53) na maksymalizację wartości oczekiwanej,
– zdecydować się na użycie estymatorów $\bar{u}$, $\bar{z}$ jako wartości stałych w czasie, otrzymując w ten sposób zadanie optymalizacji statycznej o zwykłej postaci, tj. bez zależności od czasu.

Trzecia możliwość wydaje się posunięciem bardzo radykalnym. Zwróćmy jednak uwagę, że możliwa jest sytuacja, gdy wartość optymalna $\bar{c}$, czyli rozwiązanie zadania (4.52), (4.53), w ogóle nie zależy od wartości u, z. Rozwiązanie to możemy wówczas wyznaczyć, kładąc $u(t) = \text{const}$ i $z(t) = \text{const}$. Nie przeszkadza to użyciu następnie wzoru (4.52) w pełnej postaci, ażeby obliczyć jaka wartość wyjściowa $y(t)$ pojawi się w rzeczywistości.

Przytoczmy tu raz jeszcze przykład reaktora. Rozwiązanie zadania optymalizacji ma postać $W = W_m$, $C_A = C_{Am}$, $T = T_m$ i jest niezależne od ewentualnych zmian wejścia F_B oraz czynników zewnętrznych T_A, T_B. Uwzględnienie ich zmienności w czasie w trakcie obliczania optymalnego stanu ustalonego byłoby niepotrzebne. Model procesu uwzględniający zależność od czasu pozostaje jednak użyteczny wskazując np. że natężenie F_D będzie zmienne (w takt zmian F_B).

Rozważania te wiążą się z jednym z istotnych problemów w projektowaniu układów dwuwarstwowych. Jeśli projektujemy układ, w którym zamierzamy narzucać stan procesu w odstępach T jako wartość stałą w każdym z tych przedziałów, to może – ale nie musi – okazać się możliwe wyznaczanie wartości optymalnej tego stanu z zadania statycznego. Możliwość taka wystąpi w sytuacji, gdy czynniki zewnętrzne wpływające na optymalną wartość stanu są powolne, a proces – dzięki układom sterowania bezpośredniego – szybko reaguje na zmianę wartości zadanych. Długość przedziału $[t_0, t_1]$ wynika w takim przypadku z czasu korelacji czynników zewnętrznych, a nie z dynamiki procesu. Jednocześnie przedział T, będący częścią $[t_0, t_1]$, jest dostatecznie długi na to, by uznać, że $x(t) = \text{const}$ przez prawie cały czas.

Na zakończenie rozważań zwróćmy uwagę, że posługiwaliśmy się stale modelami procesu, w różnych formach do różnych celów. Modele te zapisywaliśmy w postaci mówiącej o związku między zmiennymi (patrz np. wzór (4.52)), nie dotykając kwestii kształtu tego związku. Przy przechodzeniu do sytuacji konkretnych trzeba by ustalać strukturę modelu oraz wartości występujących w nim parametrów. Zarówno parametry, jak i struktura modelu mogą być co pewien czas zmieniane, aby model lepiej odpowiadał zmieniającej się rzeczywistości. Czynność taka byłaby *adaptacją* modelu, można ją powierzyć następnej warstwie sterowania.

4.6. Zróżnicowanie horyzontów sterowania

Zarówno w rozdziale 1, jak w rozdziale 4.1 wspominaliśmy o tym, że kolejne warstwy układu sterowania mogą operować różnym horyzontem sterowania – coraz to dłuższym w warstwach wyższych. Było to m.in. jedną z ważnych cech układu sterowania systemem wodno-gospodarczym, przytoczonym jako jeden z przykładów w rozdz. 1.5. Zróżnicowaniu horyzontu sterowania towarzyszą z reguły dwie dalsze cechy: wyższe warstwy mają coraz *mniejszą częstotliwość* interwencji oraz używane w nich modele procesu sterowanego są coraz to *mniej szczegółowe*.

Zanim przedstawimy problem bardziej formalnie, przyjrzyjmy się prostemu przykładowi, a mianowicie zadaniu sterowania zasobami wody. Ze względu na cykl pór roku oraz rozmiary głównych zbiorników retencyjnych horyzont sterowania musiałby wynosić nie mniej niż rok (kwestia wyboru horyzontu sterowania jest omawiana w rozdz. 3.4). W takim rocznym planowaniu można z pewnością pominąć średnie i małe zbiorniki albo też zsumować ich zawartości i traktować to jako jedną zmienną. Byłoby to właśnie uproszczenie modelu dla najwyższej warstwy. Schodząc w dół hierarchii, dojdziemy do warstwy najniższej, gdzie oczywiście trzeba rozpatrywać każdy, nawet najmniejszy zbiornik – gdyż trzeba określić jaka ma być decyzja sterująca tym zbiornikiem, to znaczy z jaką intensywnością mamy go opróżniać lub napełniać. Korzystamy tu jednak z możliwości rozpatrywania krótszego horyzontu.

Jako podobny przykład może służyć sterowanie zakładem metalurgicznym: na najwyższym poziomie potrzebne jest planowanie długofalowe, np. kilkuletnie, ale nie potrzeba tam określać – na rok z góry – co będzie robione z każdym odrębnym wlewkiem czy kawałkiem stali. Na najniższym poziomie musimy wydać konkretne dyspozycje operatorom, nie możemy zatem pomijać szczegółów.

Przejdźmy do opisu formalnego omawianej struktury warstwowej, przedstawionej schematycznie na rys. 4.8. Niech będzie dane zadanie sterowania optymalnego

$$\text{maksymalizować} \int_{t_0}^{t_k} f_0^1(x^1, m^1, z^1)\mathrm{d}t \tag{4.54}$$

dla procesu o równaniu stanu

$$\dot{x}^1 = f^1(x^1, m^1, z^1)$$

przy $x^1(t_0)$ danym oraz $x^1(t_k)$ swobodnym.

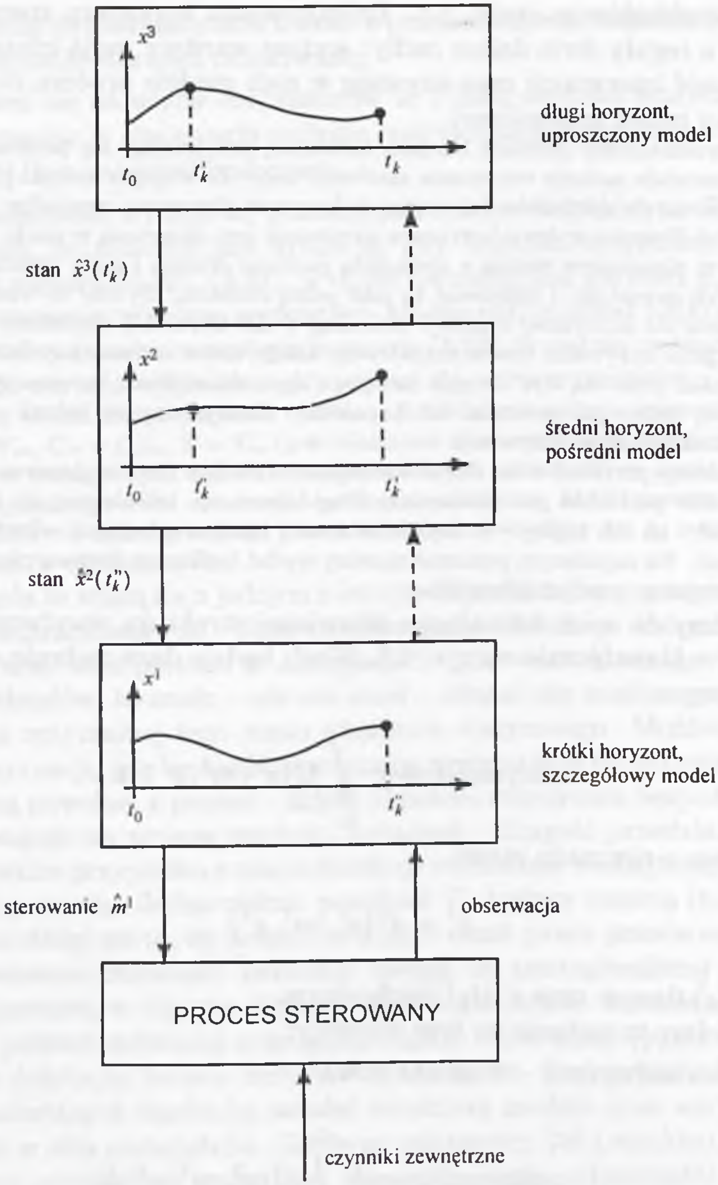
Rozdzielmy to zadanie na trzy warstwy:

A. Warstwa najwyższa – długookresowa

$$\text{maksymalizować} \int_{t_0}^{t_k} f_0^3(x^3, m^3, z^3)\mathrm{d}t \tag{4.55}$$

$\dot{x}^3 = f^3(x^3, m^3, z^3)$, $x^3(t_0)$ dane, $x^3(t_k)$ swobodne, przy czym x^3 jest uproszczonym wektorem stanu, $\dim x^3 < \dim x^1$, m^3 jest uproszczonym wektorem sterowania, z^3 uproszczonym (zastępczym) oddziaływaniem zewnętrznym.

Rozwiązanie zadania długookresowego wyznacza m.in. $\hat{x}^3(t_k')$, to znaczy stan, jaki ma być osiągnięty w chwili $t_k' < t_k$. Stan ten określa warunek końcowy dla zadania następnej, niższej warstwy.



Rys. 4.8. Zróżnicowanie horyzontów w warstwach sterowania

B. Warstwa druga – średniookresowa

$$\text{maksymalizować} \int_{t_0}^{t'_k} f_0^2(x^2, m^2, z^2)\mathrm{d}t \tag{4.56}$$

$$\dot{x}^2 = f^2(x^2, m^2, z^2),\ x^2(t_0) \text{ dane},\ x^2(t_k) \text{ zadane przez wartość } \hat{x}^3(t'_k).$$

Zadanie wartości końcowej $x^2(t_k)$ przez $\hat{x}^3(t'_k)$ nie może być bezpośrednie, bowiem wektor x^2 ma inny wymiar niż x^3, $\dim x^2 < \dim x^3$, zgodnie z zasadą zwiększania szczegółowości modelu, idąc w dół hierarchii. Musimy wprowadzić pewną funkcję γ^2 i żądać

$$\gamma^2\left(x^2(t'_k)\right) = \hat{x}^3(t'_k)$$

Funkcja γ^2 jest związana z upraszczaniem modeli („agregacją" wektora stanu, idąc od x^1 do x^2 i do x^3) i trzeba ją ustalić razem z wprowadzonymi uproszczeniami.

Zadanie drugiej warstwy wyznacza m.in. wartość $\hat{x}^2(t''_k)$, to jest stan do osiągnięcia w chwili t''_k. Będzie to wykorzystane w warstwie pierwszej.

C. Warstwa pierwsza – krótkookresowa

$$\text{maksymalizować} \int_{t_0}^{t''_k} f_0^1(x^1, m^1, z^1)\mathrm{d}t \tag{4.57}$$

$\dot{x}^1 = f^1(x^1, m^1, z^1)$, $x^1(t_0)$ dane, $x^1(t''_k)$ zadane przez warunek

$$\gamma^1\left(x^1(t''_k)\right) = \hat{x}^2(t''_k)$$

Szczegóły tego zadania są podobne do poprzednich. Zauważmy tylko, że funkcje opisujące proces f^1 i wyrażające cel sterowania f_0 są tu dokładnie te same co w zadaniu wyjściowym (oznacza to „pełny" model), natomiast horyzont jest znacznie krótszy, $t''_k < t'_k < t_k$. Zadanie najniższej warstwy wyznacza ostateczne decyzje wykonawcze, m^1.

Zauważmy, że bez upraszczania modeli obecnie omawiany podział na warstwy nie miałby sensu. Gdyby bowiem używać w najwyższej warstwie pełnego modelu, to zarówno przebieg $\hat{x}^1$, jak decyzje $\hat{m}^1$ byłyby tam wyznaczone od razu i to nie tylko na czas od t_0 do t''_k, ale aż do t_k. Niższe warstwy powtarzałyby tylko te same obliczenia.

Wprowadźmy teraz sprzężenia zwrotne, to znaczy wykorzystajmy obserwacje aktualnego działania systemu do poprawy sterowania wyznaczonego początkowo z użyciem modeli. Jedną z możliwości jest ta, by rzeczywiście osiągnięta wartość $\gamma^1(x^1(t''_k))$ stanowiła nowy stan początkowy $x^2(t''_k)$ dla warstwy drugiej. Oznacza to, że w chwili t''_k ponowilibyśmy zadanie warstwy drugiej, biorąc za punkt wyjścia nie teoretyczny, lecz rzeczywisty stan $x^1(t''_k)$. Podobnie w chwili $2t''_k$, $3t''_k$ i tak dalej. Analogiczne sprzężenie zwrotne można wprowadzić między warstwę drugą a trzecią, biorąc za stan początkowy $x^3(t'_k) = \gamma^2(x^2(t'_k))$, $x^3(2t'_k) = \gamma^2(x^2(2t'_k))$ i tak dalej.

Dla wykorzystania informacji zwrotnej z systemu rzeczywistego żądamy zatem ponowienia obliczeń optymalizacyjnych. Ponawianie to jest jednak i tak potrzebne, gdyż obliczenia dokonywane w chwili początkowej t_0 oparte będą na dostępnej wówczas informacji o czynnikach zewnętrznych i ze względu na to muszą być aktualizowane. Nie można np. w chwili początkowej wyznaczyć poprawnie $x^1(t_k'')$ dla wszystkich chwil kt_k'' w przedziale (t_0, t_k), na przykład na wszystkie dni roku.

Stosowanie sprzężenia zwrotnego, tj. wprowadzanie rzeczywiście osiągniętego stanu do aktualizacji obliczeń w każdej z warstw stanowi o różnicy między „otwartym" i „zamkniętym" układem sterowania.

Zróżnicowanie horyzontu sterowania w poszczególnych warstwach układu nie zawsze będzie miało taką postać, jak to przed chwilą opisaliśmy. Odmienna jest np. sytuacja, gdy proces dynamiczny poddany jest jakiemuś wpływowi okresowemu, a ponadto trzeba jego przebieg – z innych przyczyn – rozpatrywać w długim horyzoncie czasowym. Możemy wówczas nałożyć na siebie dwa zadania, a mianowicie:

– w niższej warstwie sterowanie ciągłe w kolejnym okresie, uzależnione od stanu $x(k)$ na początku okresu i stanu zadawanego na jego koniec $x(k+1)$,
– w wyższej warstwie sterowanie o długim horyzoncie wyznaczające pożądane stany $x(1), \ldots, x(N)$.

Przykład tego rodzaju jest opisany w książce wydanej wcześniej*) w odniesieniu do systemu przedstawionego na rys. 5.2 w rozdziale 5.

*) W. Findeisen: op. cit., s. 115–118.

Rozdział 5

Złożoność struktur hierarchicznych

5.1. Powstawanie struktur złożonych

W rozdziale 1, stanowiącym wprowadzenie w zadania i struktury sterowania, wspomnieliśmy bardzo ogólnie o *układach warstwowych* i *układach wielopoziomowych*, patrz rys. 1.5 i 1.6.

W rozdziale 2 zajęliśmy się bardziej szczegółowo *układami wielopoziomowymi*, jako takimi, które powstają w sytuacji gdy z różnych przyczyn sterowanie systemem powierzone jest równolegle działającym, „lokalnym" jednostkom decyzyjnym. Mówiliśmy o decentralizacji, czyli o rozproszeniu kompetencji pomiędzy jednostki lokalne oraz o przyczynach, dla których może być potrzebne wprowadzenie jednostki nadrzędnej, mającej za zadanie koordynację działania jednostek lokalnych. Układ tego rodzaju nazwaliśmy wielopoziomowym chociaż rozpatrywaliśmy tylko dwa poziomy. Nietrudno jest sobie wyobrazić, jak może powstać układ o trzech poziomach, gdzie jednostka nadrzędna znajdująca się na drugim poziomie będzie działała równolegle do jednostek nadrzędnych obejmujących rozległością swego sterowania inne grupy obiektów lub podsystemów, a nad nimi wszystkimi znajduje się jednostka koordynująca, obejmująca cały system. Tworzone w ten sposób układy o trzech lub więcej poziomach nie były omawiane dlatego, że nie wnoszą one nowych elementów lub zagadnień jakościowo odmiennych od tych, które występują w strukturze dwupoziomowej; ich budowa polega na powieleniu podstawowego schematu rozproszenia i koordynacji.

Wielopoziomowym układom sterowania, ogólnie rzecz biorąc, można przypisać następujące cechy charakterystyczne:

– jednostki niższych poziomów obejmują zasięgiem swoich kompetencji mniejsze części systemu, niż jednostki poziomów wyższych (zróżnicowanie rozległości sterowania),
– jednostka decyzyjna znajdująca się na szczycie obejmuje swym zasięgiem cały system,

– zadania (interesy) jednostek równoległych, tj. znajdujących się na tym samym poziomie, są we wzajemnym konflikcie,
– jednostka nadrzędna może zneutralizować konflikt jednostek lokalnych jeżeli rozporządza odpowiednimi instrumentami koordynacji,
– reprezentowane przez jednostkę nadrzędną zadanie całościowe może być w konflikcie z interesami jednostek lokalnych, nawet gdy konflikt niższego poziomu został zneutralizowany,
– jednostki lokalne i jednostka nadrzędna mogą dysponować różnym zakresem informacji o systemie sterowanym, może istnieć motywacja do przekazywania informacji nieprawdziwej,
– w każdym przypadku jednostki decyzyjne muszą się liczyć z niedokładnością posiadanych modeli, stąd potrzeba sięgania do obserwacji biegu procesu w sterowanym systemie.

Większość tych cech była w dostateczny sposób widoczna i objaśniona w rozdziale 2, jedynie kwestia sięgania do obserwacji biegu procesu sterowanego była tam zaledwie wspomniana, nie rozwinięta. Zagadnienie to jest skomplikowane, a ponadto wymaga odrębnego potraktowania w odniesieniu do zadań sterowania stanem ustalonym i w odniesieniu do zadań dynamicznych*). W rozważaniach rozdziału 2 rozróżnienie takie potrzebne nie było.

W rozdziale 4 omawialiśmy układy warstwowe, to jest układy hierarchiczne polegające na pionowym usytuowaniu jednostek decyzyjnych o różnych funkcjach bądź zadaniach, ale odnoszących się do tego samego obiektu lub systemu. Nie ma tu zróżnicowania rozległości sterowania, typowe jest natomiast zróżnicowanie częstotliwości interwencji i horyzontu czasowego. Miejsce szczególne w tej strukturze zajmuje warstwa *sterowania bezpośredniego*, mająca styczność ze sterowanym procesem fizycznym.

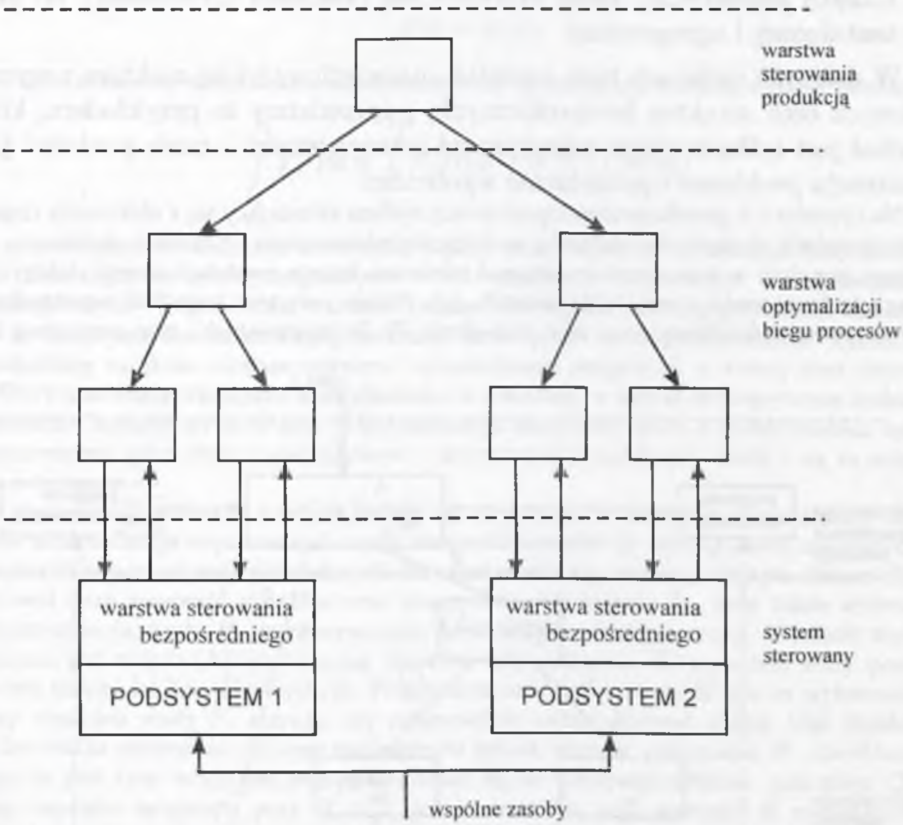
W odróżnieniu od koncepcji rozproszenia i koordynacji, gdzie schemat dwupoziomowy pełni rolę podstawową, a struktura o większej liczbie poziomów polega na jego powtórzeniu, realizacja funkcji sterowania w podziale na warstwy opiera się na zróżnicowaniu tych warstw.

Sterowanie systemami o dużej liczbie obiektów i w długim horyzoncie czasowym, zwłaszcza gdy można w nich rozsądnie wyodrębnić luźniej ze sobą powiązane podsystemy, opierać się musi w praktyce na *strukturach złożonych*, wykorzystujących zasady sterowania wielopoziomowego i wielowarstwowego równocześnie.

Struktury te powstają gdy w jednej lub kilku warstwach układu wielowarstwowego wprowadza się realizację zadania danej warstwy na sposób

*) W. Findeisen (1982). Decentralized and hierarchical control under consistency or disagreement of interests. Automatica, vol. 18, p. 647–664; W. Findeisen (1986). Interests, structures and decisions in hierarchical control. Control-Theory and Advanced Technology, vol. 2, No. 2, p. 77–103.

wielopoziomowy, tj. z podziałem równoległym i koordynacją. Przeprowadzony podział jest kontynuowany we wszystkich warstwach niższych, przy czym niektóre z nich mogą być ponownie rozdzielone. Przedstawia to schematycznie rys. 5.1. Przyjęto tam, dla ustalenia uwagi, sterowanie systemem złożonym z dwóch podsystemów, połączonych przez wspólne zasoby. Każdy z tych podsystemów jest wewnątrz podzielony tak, że uzasadnia to strukturę dwupoziomową w warstwie drugiej. Nazwy warstw użyte na rys. 5.1 wiążą się z przykładami sterowania procesami produkcyjnymi.



Rys. 5.1. Przykład złożonej struktury hierarchicznej

Struktury hierarchiczne wykorzystujące obie koncepcje podstawowe mogą mieć cechy charakterystyczne pochodzące od nich obu. Cechy układów wielopoziomowych były wymienione w tym rozdziale, cechy sterowania z podziałem na warstwy sformułowaliśmy przy końcu rozdz. 4.1. Łącząc jedno i drugie możemy się spodziewać, że w hierarchicznych strukturach sterowania rozumianych w sensie ogólnym, mieć będziemy, na kolejnych – licząc od dołu – warstwach lub poziomach, następujące cechy:

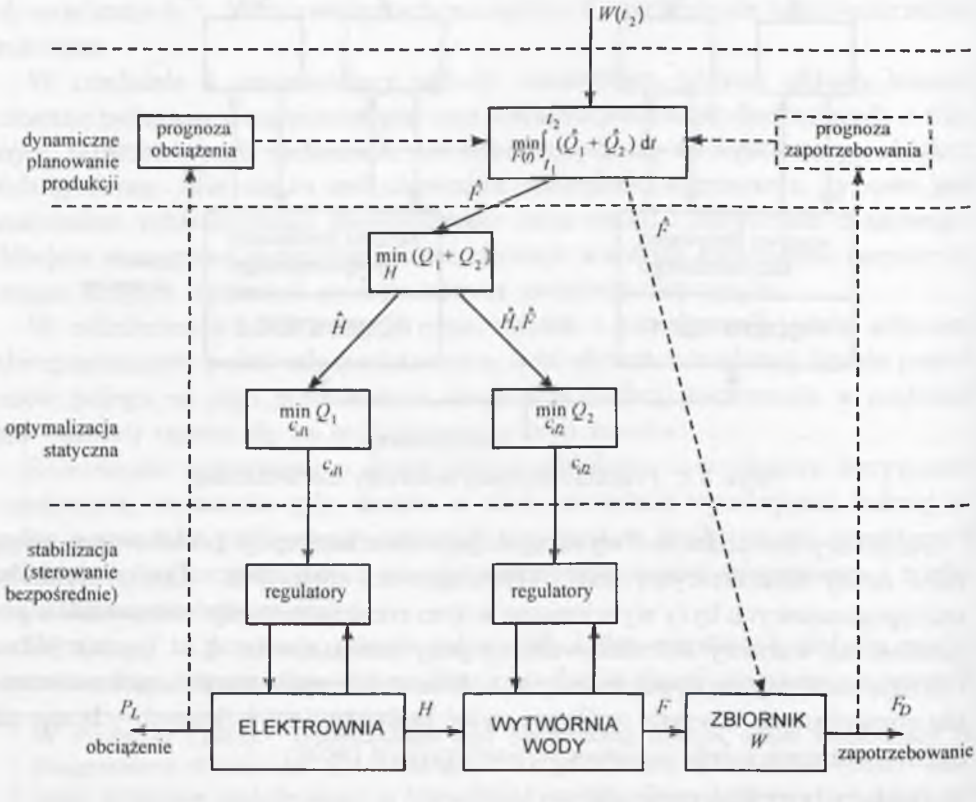
– rosnący horyzont czasowy,

– zwiększający się zasięg przestrzenny,

– malejącą częstotliwość interwencji,
– malejącą szczegółowość modeli,
– odmienne zadania, od wycinkowych do coraz to bardziej zbliżonych do „celów działania całości systemu”,
– odmienne środki (instrumenty) działania, od sterowań bezpośrednich do wyznaczania różnego rodzaju celów, ograniczeń, reguł itd.,
– odmienną informację, od fenomenologicznej (wartości chwilowe obserwowanych zmiennych), przez statystyczną (wartości uśrednione), do przekształconej i agregowanej.

W dalszych częściach tego rozdziału naświetlimy bliżej niektóre z wymienionych cech struktur hierarchicznych; poprzedzimy to przykładem, który – choć jest tylko szkicem odległym od rzeczywistości – może posłużyć jako ilustracja problemu i pobudzenie wyobraźni.

Na rysunku 5.2 przedstawiono hipotetyczny system składający się z elektrowni cieplnej oraz instalacji służącej do odsalania wody przez odparowanie i ponowne skroplenie. Załóżmy, że należy w tym systemie osiągnąć minimum kosztu produkcji energii elektrycznej oraz odsolonej wody w przedziale czasu $[t_1, t_2]$. Należy przy tym zaspokoić zapotrzebowanie na energię określoną przez moc obciążenia P_L (w megawatach) oraz potrzebną ilość



Rys. 5.2. Przykład struktury złożonej

wody F_D (w m^3 na godzinę). Jeśli natężenie strumienia kosztów oznaczymy przez Q (składają się na to koszty paliwa, energii elektrycznej, potrzeb własnych itd.), przebieg optymalny procesu w całym systemie będzie określony przez zadanie

$$\text{minimalizować } J = \int_{t_1}^{t_2} Q(t)\mathrm{d}t$$

przy ograniczeniach:

– moc produkowana równa obciążeniu

$$P(t) = P_L(t)$$

– ilość odsolonej wody równa zapotrzebowaniu minus dopuszczalna zmiana zasobu

$$\int_{t_1}^{t_2} F(t)\mathrm{d}t = \int_{t_1}^{t_2} F_D(t)\mathrm{d}t - W(t_1) + W(t_2)$$

Przypomnijmy, że struktura sterowania będzie miała warstwę sterowania bezpośredniego, działającą na zawory regulacyjne i inne elementy wykonawcze w instalacjach technologicznych. Odpowiednie wartości zadane dla regulatorów będą wówczas zmiennymi dycyzyjnymi dla zadania optymalizacji, które przed chwilą napisaliśmy. Zadanie to podzielimy na dwie odrębne warstwy: *optymalizacji statycznej*, w której stan zbiornika $W(t)$ rozważany nie będzie oraz *planowania produkcji*, w której rozwiązywane będzie odpowiednie zadanie dynamiczne. Optymalizacja statyczna może z powodzeniem być dwupoziomowa, gdyż dwie części systemu – elektrownia i wytwórnia wody – są ze sobą słabo powiązane.

Na rysunku 5.2 pokazano schemat całości tej struktury. W warstwie optymalizacji następuje minimalizacja odpowiednich części natężenia kosztów Q_1 oraz Q_2, przy zmiennych decyzyjnych w postaci wartości zadanych dla regulatorów c_{d1} oraz c_{d2}. Optymalizator dla elektrowni musi zapewnić nie tylko moc równą mocy obciążenia P_L, lecz także wytwarzanie strumienia ciepła H, przekazywanego do instalacji odsalania wody. Wielkość tego strumienia jest zmienną koordynacyjną. Sąsiedni optymalizator, dla wytwórni wody, podlega dwu zmiennym koordynacyjnym. Przy narzuconej ilości ciepła H ma on wytwarzać zadany strumień wody F, starając się jednocześnie minimalizować koszty tego działania. Jednostka nadrzędna dla optymalizatorów ustala wartość optymalną $\hat{H}$. Zwróćmy uwagę, że pod tym względem jednostki lokalne są ze sobą w konflikcie: minimum Q_1 byłoby zapewne osiągnięte przy $H = 0$, a minimum Q_2 przy wartości H stosunkowo dużej. Jednostka nadrzędna mogłaby też operować ceną na przekazywane ciepło; pozostawiamy Czytelnikowi refleksję, czy ilość ciepła przekazywana wówczas „dobrowolnie" przez elektrownię odpowiadałaby potrzebom wynikającym z zapotrzebowania na wodę oraz od czego by to zależało (patrz porównanie metod koordynacji, rozdz. 2.4.)

Warstwa najwyższa mieć będzie do rozwiązania zadanie typu planowania produkcji

$$\text{minimalizować } J = \int_{t_1}^{t_2} \left(\hat{Q}_1(t) + \hat{Q}_2(t)\right) \mathrm{d}t$$

przy czym zmienną decyzyjną jest natężenie strumienia produkowanej wody $F(t)$. Symbole $\hat{Q}_1$, $\hat{Q}_2$ oznaczają wartości natężenia kosztów, uzyskane po wykonaniu optymalizacji statycznej biegu procesu. Dla zadania planowania produkcji są to wielkości zmienne w czasie, ale w sposób powolny, „z godziny na godzinę" albo jeszcze wolniej.

Zadaniu dynamicznemu towarzyszy równanie procesu magazynowania wody, czyli równanie procesu sterowanego rozpatrywanego w tej warstwie

$$\frac{\mathrm{d}W(t)}{\mathrm{d}t} = F(t) - F_D(t)$$

przy czym zadane są stany brzegowe $W(t_1)$, $W(t_2)$.

Rozwiązanie zadania wymaga znajomości $F_D(t)$ w przedziale $[t_1, t_2]$, a także obciążenia elektrycznego $P_L(t)$ na ten przedział, gdyż od wartości tej zależy koszt $\hat{Q}_1$. Na rysunku 5.2 zaznaczono, że potrzebne są odpowiednie prognozy. Wyliczona z zadania optymalna wartość wytwarzanej wody $\hat{F}(t)$ jest zadaniem dla położonej niżej warstwy optymalizacji statycznej.

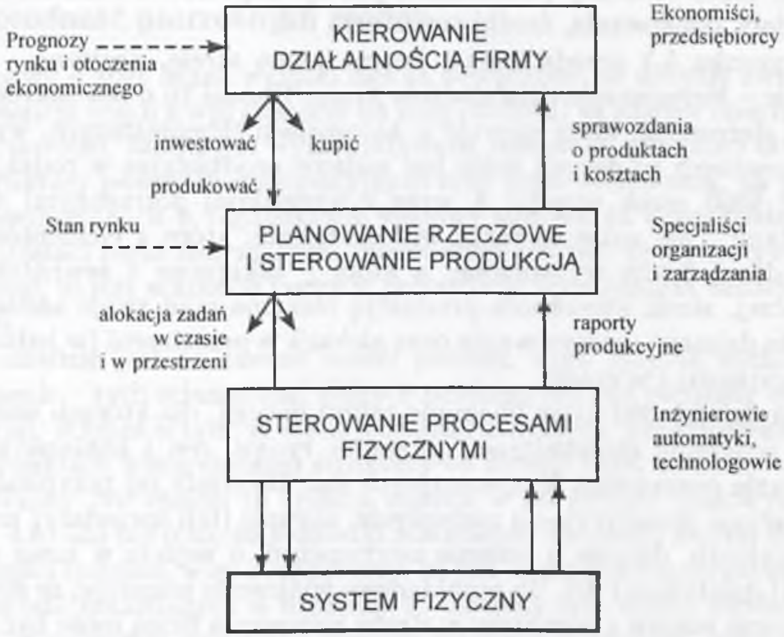
Zauważmy, że zadanie dynamiczne przypisane najwyższej warstwie może podlegać podziałowi w czasie, jak to omawialiśmy w rozdz. 4.6. Można zmieniać stan docelowy zasobu $W(t_2)$, zatem także różnicę $W(t_2) - W(t_1)$, w kolejnych przedziałach, według pewnej taktyki długofalowej. Gdyby $[t_1, t_2]$ wynosiło 24 godziny, można by stosować ujemne $W(t_2) - W(t_1)$ od poniedziałku do piątku, a następnie uzupełnić zapas wody wzmożoną produkcją w sobotę i niedzielę, gdy zapotrzebowanie na energię elektryczną jest mniejsze. Wymagałoby to oczywiście by zbiornik na wodę miał odpowiednią pojemność.

5.2. Zróżnicowanie zadań sterowania

Schemat przedstawiony na rys. 5.1, a także przykłady zamieszczone w rozdz. 1.5 pokazują zróżnicowanie zadań, które stać mogą przed jednostkami różnych szczebli (różnych warstw i poziomów) hierarchicznego układu sterowania. Zależnie od tego, jak daleko jesteśmy od procesów fizycznych i wykonawczych, zmieniać się może nie tylko horyzont sterowania, ale także charakter rozpatrywanych wielkości, w pewnym sensie – charakter branych pod uwagę zjawisk.

Na rys. 5.3 pokazano schematycznie strukturę hierarchiczną sterowania, przedstawioną z punktu widzenia zróżnicowania charakteru zadań rozpatrywanych przez jednostki decyzyjne.

W warstwie najniższej, lub grupie warstw najniższych, mamy do czynienia z zadaniem *sterowania procesami fizycznymi*. Będzie zazwyczaj istnieć warstwa sterowania bezpośredniego i warstwa optymalizacji biegu procesu, tak jak to opisywaliśmy w rozdziale 4. Zadania sterowania bezpośredniego będą stosunkowo proste, ich wykonanie – zdecentralizowane. Każdy reaktor, obrabiarka czy pojazd będzie w tym zakresie autonomiczny. Warstwa optymalizacji biegu procesu, mówiąc ogólniej – warstwa wyznaczania zadań dla sterowania bezpośredniego, może stać przed zadaniami o dość różnym charakterze. Może to być optymalizacja stanu ustalonego procesu ciągłego, wyznaczanie trajektorii procesu jednorazowego (przebieg procesu w reaktorze po jego napełnieniu, trajektoria przelotu statku powietrznego), harmonogramowanie w odniesieniu do procesu złożonego z operacji dyskretnych. Wyznaczanie zadań dla układów sterowania bezpośredniego może się odbywać w schemacie wielopoziomowym, jak pokazano na rys. 5.1, może też być wydzielona warstwa adaptacji modeli; w tej części układu sterowania mamy jednak nadal bliski związek z rzeczywistością fizyczną.



Rys. 5.3. Zróżnicowanie charakteru zadań w strukturze hierarchicznej

Sterowanie procesami fizycznymi i wykonawczymi, realizowane w całej ewentualnej złożoności swoich warstw i poziomów, może być podporządkowane zadaniom o charakterze bardziej ogólnym, ale ciągle jeszcze mającym charakter rzeczowy. Będą to zadania produkcyjne w rodzaju „przerobić w ciągu kampanii cukrowniczej 200 tys. ton buraków", czy też „wyprodukować 9500 pojazdów w ciągu miesiąca". Wchodzimy tu jakby w inną strefę, strefę *sterowania produkcją* czy też *planowania rzeczowego*. Warstwy i poziomy tej części układu sterowania muszą przetworzyć rzeczowe zadanie ogólne na sformułowania, które można by przekazać jako zadania dla procesów fizycznych. Dla cukrowni może to być wyznaczenie zadanego dobowego przerobu buraków, w przypadku technologii dyskretnych może to być zadanie odniesione do jednej zmiany na jednej linii produkcyjnej, np. „wykonać 50 sztuk wyrobu A i 120 sztuk wyrobu B".

Wydaje się rzeczą oczywistą, że strefa planowania rzeczowego i sterowania produkcją może być bardzo różnie ukształtowana zależnie od tego, czy mamy do czynienia z jednym czy wieloma zakładami, czy produkcja jest jednorodna (np. cukier) czy zróżnicowana (farmaceutyki), czy wyrób finalny jest jednorodny czy też składa się z dużej liczby oddzielnych elementów i tak dalej. Zróżnicowane będą, używane w tej strefie, narzędzia pomocne w podejmowaniu decyzji: programowanie matematyczne ciągłe i dyskretne (optymalizacja), harmonogramowanie (produkcji, a nie poszczególnych operacji), podejścia wielokryterialne. W strefie planowania rzeczowego ustalane są także różne elementy pomocnicze, jak np. zapotrzebowanie na surowce i materiały, opakowania, środki transportu itd.

Na rysunku 5.3 przedstawiono jeszcze jedną strefę, nazwaną – trochę umownie – *kierowaniem działalnością firmy.* Chodzi tu o ten zakres planowania i sterowania, który operuje w kategoriach ekonomicznych; wynikiem podejmowanych tu decyzji może być zadanie produkcyjne w rodzaju „wytworzyć 9500 sztuk wyrobu A wraz z wszystkimi potrzebnymi do tego elementami", ale może to także być określenie, które z elementów mają być w danej firmie wytworzone, a które – zakupione z zewnątrz. Tak czy inaczej, strefa sterowania produkcją otrzyma stąd swoje zadanie rzeczowe do dalszego sprecyzowania oraz alokacji w przestrzeni (w jednostkach produkcyjnych) i w czasie.

Strefa działalności firmy obejmuje zakres decyzji, dla których ważne jest szersze otoczenie ekonomiczne: prognozy rynku, cen i koniunktury, nawiązywanie porozumień kooperacyjnych itp. Do strefy tej przypisalibyśmy także decyzje inwestycyjne o rozbudowie, zakupie (lub sprzedaży) urządzeń produkcyjnych, decyzje o zmianie asortymentu, o wejściu w nowe rodzaje (branże) działalności itd. To przykładowe wyliczenie pokazuje, że struktura wewnętrzna warstw i poziomów w strefie kierowania firmą może być bardzo złożona, a także, iż bardzo różne mogą być sposoby i mechanizmy podejmowania decyzji – zwłaszcza długofalowych i liczących się z niepewnością otoczenia. Podkreślaliśmy to m.in. omawiając przykład przedsiębiorstwa lotniczego w rozdz. 1.5.

Na rysunku 5.3 wyodrębniliśmy trzy strefy sięgające od procesów fizycznych po „firmę", to jest pewną samodzielną całość ekonomiczną. Można to zapewne zrobić inaczej, a także bardziej szczegółowo, zwłaszcza gdyby schemat był dostosowany do sytuacji konkretnej. W tej chwili istotne wydaje się zwrócenie uwagi, że zaznaczone na rys. 5.3 strefy odpowiadają zainteresowaniom różnych specjalistów. Sterowanie procesami fizycznymi jest obszarem działania inżynierów automatyków, strefa planowania rzeczowego i sterowania produkcją – inżynierów specjalistów organizacji i zarządzania, strefa kierowania firmą – ekonomistów i przedsiębiorców.

Stosownie do tego, często – a może nawet z reguły – spotyka się rozważania o sterowaniu hierarchicznym, odniesione do jednej tylko strefy, na przykład do sterowania procesem fizycznym albo do strefy planowania i zarządzania ekonomicznego. Nie ma w tym żadnej niewłaściwości, pamiętać tylko należy, że w różnych strefach mówi się w zasadzie o tym samym systemie, ale o różnych przejawach jego funkcjonowania, stąd różnym językiem, oraz – co najważniejsze – co innego uważając za „proces w systemie" lub „model systemu". Zróżnicowanie to opisujemy nieco bliżej w następnym podrozdziale.

5.3. Modele systemu sterowanego

Różnorodność potrzebnych modeli

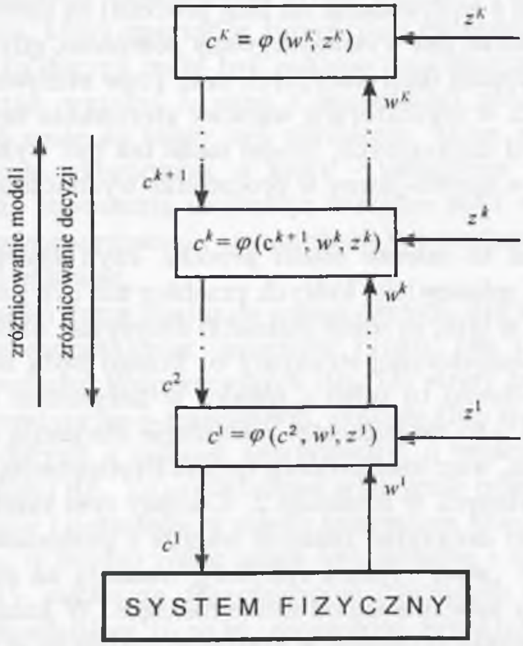
Bez obawy o zbyt liczne wyjątki można powiedzieć, że decyzje związane ze sterowaniem (czyli z wpływaniem na bieg procesu) są zawsze oparte na modelach procesu. Model jest wykorzystywany *pośrednio*, gdy służy do określenia struktury powiązań informacyjnych oraz praw sterowania, na przykład implementowanych w regulatorach warstwy sterowania bezpośredniego czy też w postaci reguł decyzyjnych; model może też być wykorzystywany *bezpośrednio*, to jest w sposób jawny w procedurze wyznaczania każdej kolejnej decyzji.

W zasadzie jest to zawsze *model procesu*, czyli zjawisk zachodzących w systemie – tych mianowicie, których przebieg ma być poddany wpływowi i kontroli. Kłopot w tym, że różne jednostki decyzyjne, umiejscowione w różnych punktach hierarchicznej struktury co innego będą uważały za proces w systemie. Nie chodzi tu tylko o różnice w horyzoncie czasowym (patrz rozdz. 4.6) ani też o to, że jednostki równoległe obejmują swymi decyzjami inne części systemu, więc inne procesy, co jest istotną cechą struktur dwupoziomowych, omawianych w rozdziale 2. Chcemy tym razem zwrócić uwagę na to, że jednostki decyzyjne różnych warstw i poziomów, zwłaszcza gdy należą do różnych „stref" (patrz rys. 5.3), uważają za proces w systemie sterowanym rzeczy znacznie się od siebie różniące. W każdej ze stref mówić się będzie o sterowaniu procesem w systemie; będzie to w zasadzie ten sam system, ale zupełnie inne przejawy rzeczywistości będą uważane za „proces sterowany".

Problem ten można uporządkować w sposób następujący: w każdej ze stref uważa się za system sterowany wszystko to, co znajduje się poniżej, to znaczy system fizyczny wraz z wszystkimi jednostkami decyzyjnymi niższych szczebli. Dla strefy planowania rzeczowego i sterowania produkcją wszystko to, co znajduje się poniżej, stanowi „system produkcyjny"; dla strefy kierowania firmą przedmiotem znajdującym się poniżej są zakłady produkcyjne bądź ich grupy, wraz z całym wewnętrznym zarządzaniem. Z tego właśnie powodu model systemu, potrzebny i rozpatrywany w danej strefie, może być odmienny od pozostałych, może opisywać inne bądź inaczej ujęte procesy, może także eksponować inne wzajemne oddziaływania między podsystemami.

Wszystko to jest ze sobą, w układzie pionowym, wzajemnie powiązane, gdyż decyzje oddziałują z góry w dół, a obserwacje – odpowiednio dostosowanego rodzaju – przekazywane są w górę i służą przy podejmowaniu tych decyzji. Formalnie można to przedstawić jak na rys. 5.4, który w istocie rzeczy ilustruje każdą strukturę wielowarstwową. Decyzja warstwy wyższej c^{k+1} jest parametrem w zadaniu decyzyjnym warstwy niższej; wchodzi tu ponadto obserwacja otoczenia z^k oraz informacja o procesie w^k, przekazywana przez jednostkę decyzyjną warstwy położonej poniżej. Można to zapisać wzorem

$$c^k = \varphi(c^{k+1}, w^k, z^k) \tag{5.1}$$



Rys. 5.4. Decyzje i obserwacje w pionowej strukturze hierarchicznej

Zapis ten nie ma oznaczać, że sposobem podejmowania decyzji c^k jest reguła decyzyjna; może to być jakikolwiek mechanizm podejmowania decyzji. Przypomnijmy, na podstawie sytuacji przedstawionej w rozdz. 2.6, że jednostka niższego szczebla może mieć motywację do przekazywania informacji nieprawdziwej.

Każda z warstw operuje, w sposób pośredni lub bezpośredni, modelem opisującym swój „proces sterowany"; zawiera się w tym spodziewana reakcja wszystkich warstw niższych na decyzję c^k, przy czym wyrazem odpowiedzi rzeczywistego systemu jest wielkość w^k.

Wspominaliśmy już, że modele w różnych warstwach mogą opisywać różne przejawy zjawisk w systemie. Zwróćmy teraz uwagę, że mogą to być modele o różnym charakterze – modele procesów ciągłych, procesów jednorazowych, procesów złożonych z operacji dyskretnych i in. Modele takie mogą się przeplatać, występować na przemian, na przykład: model dynamiczny do celów sterowania bezpośredniego, model statyczny dla optymalizacji stanu ustalonego, ponownie model dynamiczny dla sterowania w długim horyzoncie itp. Proces fizyczny podstawowy może być ciągły, a w jednej z górnych warstw hierarchii pojawi się model procesu złożonego z operacji dyskretnych, jak na przykład w warstwie planowania rozwoju systemu wodno-gospodarczego, rys. 1.12.

Charakter decyzji c^k, przekazywanych z warstwy do warstwy, może być różny, tak jak różne są rozpatrywane procesy i ich modele. Wielkość c^k może być zadaną wartością zmiennych fizycznych (temperatura, przepływ), zmiennych rzeczowych (wielkość produkcji), zmiennych ekonomicznych lub innych, włączając w to także agregaty zmiennych tego samego rodzaju (sumaryczna wielkość produkcji kilku zakładów). Mogą to także być wartości ograniczeń dla zmiennych występujących w zadaniu decyzyjnym, wartości zasobów itd. Wreszcie, mogą to być wartości parametrów dla reguł decyzyjnych niższej warstwy, czy też innego rodzaju zmiany przepisów i regulaminów postępowania. Trzeba tu jeszcze zwrócić uwagę, że przy całym zróżnicowaniu zagadnień rozpatrywanych w różnych warstwach decyzja c^k, język i sposób wyrażenia polecenia warstwy wyższej dla niższej musi być zrozumiały dla nich obu. Jest to bowiem zmienna, która ma swoje miejsce zarówno w jednym, jak i w drugim modelu systemu.

W tym miejscu warto jeszcze wspomnieć, że sprawa modelu systemu sterowanego, widzianego przez daną warstwę sterowania, komplikuje się gdy pomyślimy o możliwej sprzeczności interesów jednostki danej warstwy i jednostek warstw niższych. Powracamy do tego w rozdz. 5.5.

Mówiliśmy dotąd o potrzebie posiadania modeli procesów w systemie, czyli modeli opisujących zjawiska zachodzące w samym systemie; dodać do tego należy *modele otoczenia*, czyli modele obrazujące zachowanie się otoczenia systemu oraz – jeśli zachodzi taka potrzeba – prognozujące jego stan w przyszłości.

Dalej, dla wygody Czytelnika nie w pełni zorientowanego w budowie i stosowaniu modeli, podajemy objaśnienie ważniejszych pojęć i terminów z tej dziedziny.

Cechy i rodzaje modeli

Model jest to narzędzie, za pomocą którego można opisać system (ściślej – proces w systemie) i przewidzieć jego zachowanie się pod wpływem warunków zewnętrznych. Nie mając modelu systemu musielibyśmy wykonywać eksperymenty na systemie rzeczywistym. Wynika stąd m.in., że szczególne znaczenie mają te modele, za pomocą których badamy zachowanie się systemów jeszcze nie istniejących lub systemów które mają działać w warunkach znacznie odbiegających od stanu obecnego.

W potocznym odczuciu słowo „model" zawiera w podtekście niepełną zgodność z rzeczywistością. Tak też zwykle jest: model może być umyślnie uproszczoną reprezentacją rzeczywistości do określonych celów. Model może też być zniekształconym poglądem na rzeczywistość i wówczas prowadzić będzie do fałszywych konkluzji.

a) *Modele sformalizowane i intuicyjne*

Modele, które opisują zależności i zjawiska rzeczywiste przy użyciu matematycznych równań lub wyrażeń logicznych są modelami *sformalizowanymi*. Modele takie są, w swym zachowaniu, całkowicie zdeterminowane, to znaczy nie podlegają one zakłóceniom zewnętrznym (chyba że są one przewidziane w samym modelu): przy wielokrotnym powtarzaniu pytania model sformalizowany daje tę samą odpowiedź.

Modele *intuicyjne*, jak np. model zawarty w umyśle eksperta, opierają się na myślowych dedukcjach i ocenach, które zawierają zawsze sporą dozę niejednoznaczności. Zadając ekspertowi to samo pytanie w różnym czasie możemy otrzymać nieco inną odpowiedź.

Nie należy jednak twierdzić, że modele sformalizowane są zawsze lepsze od intuicyjnych. Model sformalizowany jest w istocie często zależny od przekonań i osądów eksperta, który model ten opracowywał, zatem model sformalizowany też niełatwo może osiągnąć poziom „pełnej obiektywności"

b) *Modele przyczynowe i korelacyjne*

Model *przyczynowy* jest potrzebny wszędzie tam, gdzie chcemy ustalić, że: „jeśli moje działanie będzie m, to skutek będzie y"; jest to z reguły potrzebne w rozważaniach dotyczących sterowania.

Aby zbudować poprawny model przyczynowy trzeba albo znać *wewnętrzne mechanizmy* procesu zachodzącego w systemie, to znaczy prawa według których jedne zjawiska wywołują inne zjawiska, albo też trzeba przeprowadzić tak zwany *czynny eksperyment identyfikacyjny*.

W odróżnieniu od tego istnieją sytuacje, gdzie nie mamy zamiaru podejmować interwencji, chcemy tylko przewidzieć bieg zjawisk – na przykład zjawisk w otoczeniu systemu. Wystarczy wówczas model, wskazujący na istnienie *korelacji*; model taki można zbudować korzystając z biernych obserwacji i ich obróbki statystycznej, nie dochodząc co jest przyczyną, a co skutkiem. Można tu np. wykryć korelację między dwoma skutkami tej samej przyczyny, obserwatorowi nie znanej i nie obserwowanej, znajdującej się jakby „za ścianą".

c) *Modele statyczne i dynamiczne*

Modele *statyczne* pozwalają opisać stany ustalone (stany równowagi statycznej) procesów dynamicznych (mówiliśmy o takich modelach w rozdz. 3.2 oraz 4.5).

Modele *dynamiczne* pozwalają opisać – oprócz stanu ustalonego – przebiegi przejściowe, na przykład pozwalają przewidzieć w jaki sposób proces w systemie przejdzie z jednego stanu równowagi do drugiego. Modele dynamiczne były potrzebne w rozważaniach rozdz. 3.2, 3.3, 3.4, a także 4.6. Istotą modelu dynamicznego jest branie pod uwagę możliwej zmienności *stanu procesu*.

Nie należy utożsamiać pojęcia modelu statycznego z „systemem statycznym". System może być i jest dynamiczny, to tylko my interesujemy się pewnym wybranym aspektem zachodzących w nim zjawisk. Aspektem takim może być stan ustalony procesu dynamicznego, ale może też być np. wartość średnia przebiegów czasowych na wejściu i wyjściu systemu. Przykładowo, wartość średnia $y_{\text{śr}}$ dopływu naturalnego do odcinka rzeki za długi okres czasu (np. miesiąc) może się okazać w przybliżeniu proporcjonalna do średniej opadów $z_{\text{śr}}$, byłby to zatem model statyczny $y_{\text{śr}} = kz_{\text{śr}}$. Jednocześnie, model $y(t) = kz(t)$, gdzie $y(t)$, $z(t)$ byłyby wartościami chwilowymi jest oczywiście dla tej sytuacji fizycznej nieprawdziwy – nie można by go np. użyć do opisu sytuacji powodziowej.

Istnieją jednak sytuacje, gdy można użyć modeli statycznych nawet dla opisu związków między wartościami chwilowymi wejść i wyjść systemu; ma to miejsce wówczas, gdy zmienność wejść jest dostatecznie powolna, a ponadto proces ma właściwość zapominania – co oznacza, że dostatecznie dawne wartości wejść nie mają wpływu na stan obecny (właściwości tej nie ma np. proces magazynowania). Bliższe omówienie takiej sytuacji można znaleźć w jednym z rozdziałów podręcznika z 1985 r[*)].

Przypominamy tu także rozważania dotyczące dopuszczalności użycia modelu statycznego dla procesu rozpatrywanego łącznie z regulatorami bezpośrednimi, przeprowadzone w rozdz. 4.5.

Trzeba mocno podkreślić, że przy rozpatrywaniu systemów o dużej złożoności jest rzeczą bardzo ważną móc zdecydować, iż dla niektórych procesów w systemie można przyjąć modele statyczne. Oznacza to bowiem uproszczenie modelu dynamicznego dla całości systemu (obniżenie rzędu modelu), a w ślad za tym ułatwienie bądź wręcz umożliwienie rozpatrywania systemu jako całości.

d) *Modele deterministyczne i probabilistyczne*

Modele są właściwie zawsze deterministyczne w tym sensie, że przecież wiemy jaki model zbudowaliśmy. Jeżeli jednak chcemy za pomocą modelu wyrazić działanie systemu w warunkach niepewności, na przykład analizować proces opisany przez równania

$$\dot{x}(t) = f\left(x(t), u(t), z(t)\right) \tag{5.2}$$

$$y(t) = g\left(x(t), u(t), z(t)\right) \tag{5.3}$$

gdzie – w procesie rzeczywistym – $z(t)$ jest nie znanym i nie obserwowanym wpływem zewnętrznym, to możemy na przykład najpierw zbudować model zgodny z równaniami (5.2), (5.3), a następnie przyłożyć na jego wejście symulowany przebieg losowy o takich samych cechach probabilistycznych, jakie ma przebieg z, działający na proces rzeczywisty. Na tym

*) W. Findeisen, red. Analiza systemowa – podstawy i metodologia. PWN, Warszawa 1985.

„tle" możemy badać, w modelu, różne cechy procesu rzeczywistego – na przykład jak zależy trajektoria wyjścia $y_{[t_0,t_k]}$ od wejścia $u_{[t_0,t_k]}$ i stanu początkowego $x(t_0)$. Będzie to oczywiście zależność o charakterze losowym, inna w każdym kolejnym eksperymencie o długości $[t_0, t_k]$ ze względu na to, że inna będzie za każdym razem wartość, czyli tak zwana realizacja zakłócenia $z_{[t_0,t_k]}$.

Wyniki takiej *symulacji stochastycznej* można poddać obróbce statystycznej, obliczając np. wartość średnią i inne cechy statystyczne wyjścia $y_{[t_0,t_k]}$.

Jeżeli, w tym samym zadaniu, odnoszącym się do zbadania zachowania się systemu w warunkach niepewności, uda nam się wyznaczyć w modelu *cechy probabilistyczne* wyjścia $y_{[t_0,t_k]}$, to jest wyrazić *analitycznie* jego wartość oczekiwaną, wariancję, funkcję autokorelacji i tak dalej, w zależności od odpowiednich cech stanu $x(t_0)$ i wejścia $z_{[t_0,t_k]}$, to otrzymujemy *model probabilistyczny*: udziela on potrzebnej odpowiedzi bez konieczności poddania go eksperymentom polegającym na użyciu symulowanych przebiegów losowych.

Modele probabilistyczne są trudne do uzyskania, a symulacja stochastyczna jest pracochłonna i nie zawsze możliwa. Nader często analizujemy sytuację opisaną przez równania (5.2), (5.3) w ten sposób, że zaniedbując losowy charakter $z_{[t_0,t_k]}$ przyjmujemy jego wartość oczekiwaną, lub wartość najbardziej prawdopodobną (prognozę) $z^*_{[t_0,t_k]}$ za deterministyczną wielkość wejściową systemu. Korzystaliśmy z tego podejścia w rozdz. 3.4.

e) *Modele dyskretne procesów ciągłych*

Wartości zmiennych u, x, y, z w procesie dynamicznym ciągłym są określone dla każdej chwili t; czas jest ciągły. Można dla procesów ciągłych budować *modele dyskretne*: takie, które będą operować wartościami zmiennych procesu rzeczywistego tylko dla chwil $t = kT$, $k = 0, 1, \ldots$, czas modelu jest ziarnisty.

Spróbujmy interesować się wartościami stanu $x(t)$ procesu ciągłego w chwilach kT. Napiszmy równanie stanu zbiornika lub magazynu w postaci eksponującej te właśnie wartości

$$x((k+1)T) = x(kT) + \int_{kT}^{(k+1)T} u(t)\mathrm{d}t - \int_{kT}^{(k+1)T} y(t)\mathrm{d}t \tag{5.4}$$

gdzie $u(t)$ oznacza chwilową wartość zasilania, a $y(t)$ wartość odpływu ze zbiornika. Uznajmy, że chwilowy odpływ zależy od chwilowej wartości stanu; otrzymamy

$$x((k+1)T) = x(kT) + \int_{kT}^{(k+1)T} u(t)\mathrm{d}t - \int_{kT}^{(k+1)T} g(x(t))\,\mathrm{d}t \tag{5.5}$$

Równanie to jest ścisłe. Aby go użyć, trzeba by jednak znać wartości stanu $x(t)$ także wewnątrz przedziału $[kT, (k+1)T]$, co przeczy założeniu budowy modelu dyskretnego. Trzeba zatem zdecydować się na przybliżenie, przyjmując np. że $x(t) = x(kT)$ w całym przedziale. Czyniąc podobnie w stosunku do przebiegu $u(t)$, dojdziemy do *dyskretnego równania stanu* o postaci ogólnej

$$x((k+1)T) = f^*(x(kT), u(kT)) \tag{5.6}$$

a normując oś czasu tak, by było $T = 1$ oraz zapisując również równanie wyjścia, otrzymamy standardową postać modelu dyskretnego

$$x(k+1) = f^*(x(k), u(k)) \tag{5.7}$$
$$y(k) = g(x(k), u(k)) \tag{5.8}$$

Funkcja $g(\cdot)$ występująca w (5.8) jest identyczna z takąż funkcją występującą w opisie ciągłym, natomiast postać funkcji $f^*(\cdot)$, różni się od funkcji $f(\cdot)$ występującej w ciągłym równaniu stanu; zależy to od sposobu, w jaki zastąpimy ciągłe w czasie wartości zmiennych po prawej stronie równości (5.5) przez wartości dyskretne. Można, w szczególności, zamiast przyjmować w odpowiedniej całce, że $x(t) = x(kT)$, przyjąć, że $x(t)$ zmienia się liniowo od $x(kT)$ do $x((k+1)T)$.

Model dyskretny informuje tylko o wartościach stanu procesu ciągłego w chwilach kT (leży to w jego założeniach), gorzej, że w ogólnym przypadku informuje o tych wartościach z błędem. Znamy już pochodzenie tego błędu: przybliżone obliczenie prawej strony równania (5.5). W niesprzyjających przypadkach błąd określania wartości stanu $x(kT)$ przez model dyskretny może narastać w miarę oddalania się od chwili początkowej, kiedy stan był znany i wynosił $x(0)$. Błędy wartości stanu $x(kT)$ z reguły maleją przy zmniejszaniu odstępu dyskretyzacji T: musi on być mały w stosunku do zmienności przebiegów w systemie.

Istnieją sytuacje gdy model dyskretny jest dokładny, niezależnie od odstępu dyskretyzacji. Jeżeli na przykład rozpatrujemy zbiornik retencyjny o dopływie z oraz zrzucie wody u, to napiszemy dla stanu $x((k+1)T)$ równanie

$$x((k+1)T) = x(kT) + \int_{kT}^{(k+1)T} z(t)\mathrm{d}t - \int_{kT}^{(k+1)T} u(t)\mathrm{d}t \tag{5.9}$$

Wprowadzając pojęcia „dopływu całkowego" $z_{\text{int}}(kT)$ i „zrzutu całkowego" $u_{\text{int}}(kT)$ (będą one wyrażone w jednostkach objętości, a nie natężenia przepływu) napiszemy

$$x((k+1)T) = x(kT) + z_{\text{int}}(kT) - u_{\text{int}}(kT) \tag{5.10}$$

Gdy będziemy określać dopływ jako z_{int} oraz narzucać zrzut wody w postaci u_{int}, to model (5.10) – mimo że jest dyskretny – jest modelem dokładnym: wyznacza ściśle wartości stanu w chwilach dyskretnych $t = kT$, $k = 1, 2, \ldots$, przy czym dokładność ta nie zależy od odstępu dyskretyzacji T. Ma to istotne znaczenie np. dla obliczeń związanych z polityką retencji czy innego rodzaju magazynowania. Obliczenia takie mogą być robione dla odstępów dziennych, tygodniowych czy miesięcznych bez obawy, że będzie narastał „błąd dyskretyzacji". Przyczyną nieobecności tego błędu jest to, że równanie (5.10) opisuje proces o właściwościach całkujących, tj. taki, dla którego ciągłe równanie stanu ma postać

$$\dot{x}(t) = f\left(u(t), t\right) \tag{5.11}$$

czyli po prawej stronie nie występuje zmienna x.

f) *Modele zjawisk ziarnistych i procesów dyskretnych*

Warto odróżnić, na zakończenie, modele dyskretne procesów ciągłych od modelowania zjawisk, które są *ziarniste z natury.* Na przykład zjawiskiem takim jest długość kolejki statków oczekujących na wejście do portu albo zadań oczekujących na wykonanie wewnątrz komputera. Jest pewnego rodzaju ciekawostką, że dla tego rodzaju zjawisk ziarnistych stosuje się często modele ciągłe i właśnie modele ciągłe są teraz modelami przybliżonymi (występują w nich np. ułamkowe liczby oczekujących statków czy zadań).

Odrębnym zagadnieniem są modele procesów dyskretnych, rozumianych jako procesy złożone z wyodrębnionych operacji, następujących po sobie bądź odbywających się równolegle, przy tym w różny sposób wzajemnie uwarunkowanych bądź uzależnionych – używa się tu czasem nazwy *kompleks operacji.* Tego rodzaju obraz zjawisk w systemie może się pojawić w różnych warstwach sterowania – np. tam gdzie dokonuje się alokacji zadania produkcyjnego i możliwe są różne tego warianty przestrzenne i czasowe albo w warstwie rozbudowy systemu, gdzie rozpatruje się zadania inwestycyjne oraz harmonogram ich realizacji.

Modele procesów dyskretnych, będących kompleksami operacji, operują odmiennym językiem opisu i specyficznymi metodami analizy. Tematu tego rozwijać nie będziemy.

5.4. Zróżnicowanie informacji

Sterowanie – jak to określiliśmy w rozdz. 1 – jest to interwencja dokonywana w trakcie trwania procesu sterowanego; daje to możliwość korzystania z obserwacji zarówno przebiegu tego procesu, jak też zjawisk zachodzących w otoczeniu danego systemu. Dane takie zaliczymy do *informacji bieżącej.* Decyzje jednostek sterujących czy też reguły ich podejmowania oparte są ponadto na innego rodzaju danych niż obserwacje procesów i zjawisk, a mianowicie na wiedzy o właściwościach systemu i otoczenia. Będzie to tzw. *informacja a priori.*

Informacja *a priori* w odniesieniu do systemu sterowanego to przede wszystkim opis struktury oraz modele przebiegających w systemie procesów, a także znajomość zasobów i różnego rodzaju ograniczeń. Jeśli system sterowany zawiera jednostki decyzyjne, jak to ma miejsce gdy rozpatrujemy go z wyższego szczebla sterowania, to ważna jest znajomość celów (interesów) jednostek decyzyjnych, a także sposobu (racjonalności) ich postępowania – na przykład czy istotnie chcą maksymalizować swoją funkcję celu, czy nie stosują innych kryteriów decyzyjnych, jak ustosunkowują się do ryzyka, tj. do niepewności (patrz np. rozdz. 2.6). Całość tej informacji *a priori* składa się na *model systemu*; podkreślaliśmy różnorodność tych modeli z punktu widzenia różnych warstw i poziomów sterowania, w poprzednim podrozdziale.

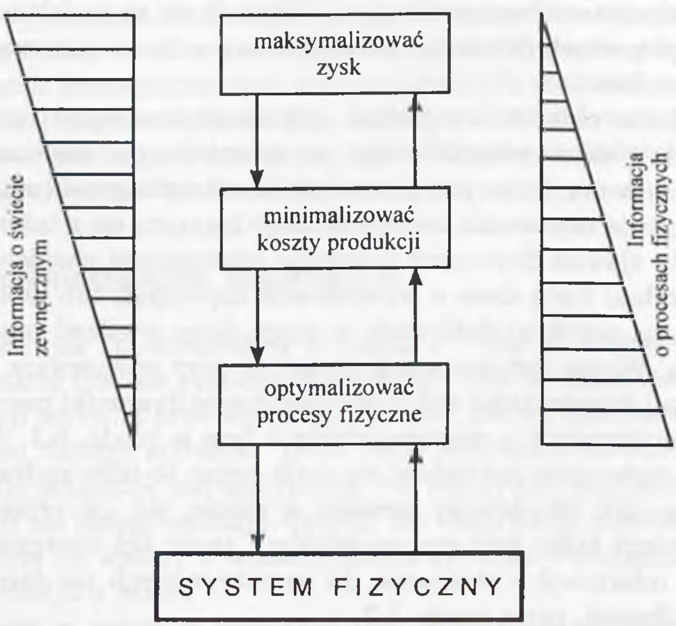
Z kolei, dane *a priori* w odniesieniu do otoczenia systemu to znajomość obowiązujących tam praw i odbywających się procesów, jeśli mają one wpływ na zjawiska zachodzące w systemie sterowanym (patrz określenie „otoczenia" w rozdz. 1.1). Dla systemu wodno-gospodarczego, np. takiego jak na rys. 1.11, będzie to przede wszystkim znajomość prawidłowości hydrologicznych, a także odpowiednich dla danego horyzontu sterowania prognoz (z punktu widzenia decyzji sterujących prognoza jest informacją *a priori*). Dla najwyższych szczebli decyzyjnych ważne będą inne prognozy, np. dotyczące rozwoju sieci osiedleńczej, rodzajów przemysłu itd.

Informacja bieżąca także daje się podzielić na informację o systemie i informację o otoczeniu. W odniesieniu do systemu mamy tu dostarczać jednostce decyzyjnej informacji o tym, jaki jest rzeczywisty bieg procesu; bieg ten jest wynikiem zarówno sterowań, jak czynników zewnętrznych. Jednostka decyzyjna jest zainteresowana przede wszystkim tym, jak proces zareagował na jej dotychczasowe decyzje sterujące. Pojawia się tu problem wydzielenia informacji potrzebnej (filtracji), rozwiązywany w teorii sterowania na wiele różnych sposobów.

W kontekście układów sterowania systemami złożonymi rzeczą najważniejszą wydaje się zwrócenie uwagi na zróżnicowanie informacji bieżącej o działaniu systemu, która jest potrzebna na różnych szczeblach. Na samym dole, w warstwie sterowania bezpośredniego korzysta się z informacji odnoszącej się do zjawisk fizycznych i operuje wartościami chwilowymi. Nieco wyżej potrzebne będą dane o wielkościach fizycznych lub rzeczowych, ale uśrednione, na przykład ilość wody w ciągu doby, wielkość produkcji tygodniowej itp. Ważne jest zwrócenie uwagi, że przy operowaniu wartościami uśrednionymi dopuszczalne być może pominięcie dynamiki procesu, tj. użycie modeli statycznych – wspomnieliśmy o tym w rozdz. 5.3. W wyższych warstwach sterowania potrzebny się staje coraz to inny rodzaj informacji bieżącej, na ogół uśrednionej zarówno w czasie, jak „w przestrzeni" (np. suma produkcji kilku linii czy wydziałów), może też następować zmiana charakteru informacji – stosownie do rozpatrywanych na danym szczeblu procesów i decyzji, patrz rozdz. 5.2.

Informacją bieżącą o otoczeniu systemu będą aktualne dane o oddziaływaniach zewnętrznych: obecny dopływ wody do zbiornika retencyjnego, obecne zapotrzebowanie na produkt, aktualne ceny surowców itd. Informacją bardzo istotną dla wszystkich zadań decyzyjnych liczących się z dynamiką procesów, czyli wybiegających naprzód w czasie, będą jednak także aktualne prognozy oddziaływania otoczenia; prognoza pozostaje informacją *a priori* w odniesieniu do pojedynczej decyzji, staje się jednak działaniem bieżącym w układzie sterowania. To samo można powiedzieć o aktualizacji innych elementów wiedzy o systemie, na przykład modeli procesów.

W strukturach hierarchicznych wydaje się rzeczą typową i charakterystyczną, że zakres informacji o systemie fizycznym, odbywającym się w nim procesie, maleje w miarę przechodzenia do wyższych warstw i poziomów sterowania; wygląda to tak, że z dużej liczby danych „konkretnych", o zjawiskach i ich przebiegu, wydobywa się stopniowo obraz coraz bardziej syntetyczny, dostosowany do potrzeb jednostki decyzyjnej (patrz rozdz. 5.2). Jest w tych danych coraz mniej informacji o zjawiskach, ale być może więcej o ich uwarunkowaniach, o tym dlaczego zjawiska – w sensie średnim – są takie, jakie są. Jednocześnie zwiększa się zakres informacji o otoczeniu systemu. Układy sterowania bezpośredniego działają zazwyczaj bez jakiejkolwiek obserwacji oddziaływań zewnętrznych, kompensują ich wpływ na zasadzie samego tylko sprzężenia zwrotnego. Optymalizacja biegu procesu często musi brać pod uwagę zmiany w otoczeniu; przejście do dalszych stref układu hierarchicznego (patrz rys. 5.2) powoduje potrzebę posiadania coraz to większej liczby danych o świecie zewnętrznym. Można to przedstawić obrazowo jak na rys. 5.5.



Rys. 5.5. Zmiany w zakresie informacji o systemie i otoczeniu, potrzebnej różnym warstwom i poziomom sterowania

5.5. Zgodność i konflikty interesów

Przypomnijmy, że rozpatrując rozproszenie kompetencji między równolegle działające jednostki decyzyjne, w rozdziale 2 wyróżniliśmy trzy przypadki (patrz wzory (2.16)–(2.18)):

1) jednostki lokalne są zainteresowane tylko własnymi rezultatami

$$Q_i = Q_i(c_i, u_i, y_i)$$

2) jednostki lokalne są zainteresowane także wyjściami innych podsystemów

$$Q_i = Q_i(c_i, u_i, y_1, \ldots, y_N)$$

3) jednostki lokalne mają wspólną funkcję celu

$$Q_i = Q(c_1, \ldots, c_N, u_1, \ldots, u_N, y_1, \ldots, y_N)$$

Podkreśliliśmy wówczas, że tylko w trzecim z tych przypadków nie ma wątpliwości, że jednostki decyzyjne nie będą ze sobą w konflikcie – tworzą one tak zwany zespół.

Przypadek drugi jest sytuacją niezgodności interesów, gdyż indywidualne funkcje celu są określone, po części, na wspólnych zmiennych w postaci wyjść podsystemów. Na przykład podsystem i w obliczu dużych kosztów, pociąganych za sobą przez decyzje sterujące c_i, może ustalić małą wartość swego wyjścia y_i. Będzie to dla niego optymalne, ale nie zadowoli innych podsystemów.

Przypadek pierwszy, mimo że funkcje celu są określone tylko na zmiennych lokalnych, jest pozbawiony wzajemnych konfliktów tylko wtedy, gdy podsystemy nie są ze sobą powiązane, tzn. nie korzystają ze wspólnych zasobów, nie mówiąc o powiązaniach przez wejścia i wyjścia. Przypominamy tu trzy podstawowe warianty struktury systemu, rys. 2.2, 2.3 oraz 2.4, a także wzór (2.10), ilustrujący faktyczną zależność osiąganej wartości każdego Q_i od wszystkich sterowań.

Sytuacją najbardziej typową dla równoległych jednostek decyzyjnych, które działają ze *świadomością swoich celów* jest niezgodność interesów. W rezultacie, jednostki równoległe znajdują się w sytuacji gry, która może prowadzić do niepożądanego zachowania się systemu zaopatrzonego wyłącznie w zdecentralizowane sterowanie. Jest to jedna z przyczyn, dla których potrzebna może być koordynacja (patrz rys. 1.6 oraz 2.1). Pisaliśmy już w rozdz. 2.2, że interesy jednostki koordynującej sprowadzają się do dwóch głównych punktów:

– neutralizacji konfliktu między jednostkami lokalnymi,

– spowodowania, by punkt równowagi systemu sterowanego przez jednostki lokalne był zadowalający lub optymalny z punktu widzenia celu globalnego.

Jeśli jednostka koordynująca ma na celu maksymalizację globalnej funkcji celu Q, to staje się rzeczą istotną jaka jest jej relacja do funkcji $Q_1, Q_2, \ldots, Q_N$, których optymalizację mają na celu jednostki lokalne. W rozdziale 2 mówiliśmy o tym, że jeżeli

$$Q = \Psi(Q_1, Q_2, \ldots, Q_N)$$

gdzie $\Psi(\cdot)$ jest funkcją ściśle monotoniczną (por. rozdz. 2.3, wzór (2.20)), to cele jednostki nadrzędnej i jednostek lokalnych są zgodne. Rozumieć to trzeba w ten sposób, że poprawa celów lokalnych sprzyja celowi globalnemu, nie jest mu przeciwna albo obojętna. Tym niemniej indywidualna optymalizacja celu lokalnego Q_i pozostaje w konflikcie z pozostałymi celami lokalnymi Q_j, $j \neq i$, tym samym – może być w konflikcie z celem Q jednostki nadrzędnej. Stąd właśnie pochodzi motywacja jednostki lokalnej do przekazywania nieprawdziwej informacji o podsystemie, mimo „zgodności celów", jak to opisywaliśmy w rozdz. 2.4.

Przypomnijmy, że najprostszym przykładem funkcji ściśle monotonicznej jest suma

$$Q = Q_1 + Q_2 + \ldots + Q_N$$

a także, iż relacja taka zapewne nie jest powszechna w praktyce sterowania systemami.

Jeśli związek Q z celami lokalnymi $Q_1, \ldots, Q_N$ nie daje się wyrazić przez funkcję typu $\Psi(\cdot)$, mamy niezgodność interesów jednostki nadrzędnej ze zbiorem interesów jednostek lokalnych. W rozdziale 2.4 określiliśmy to jako sytuację gry $(N+1)$-osobowej, sugerując, że powinna to być gra hierarchiczna, z prawem pierwszego ruchu dla jednostki nadrzędnej.

Dla jednostki nadrzędnej sytuacja zgodności lub niezgodności interesów ma określone konsekwencje. W sytuacji zgodności jednostka nadrzędna „zachowuje pełnię władzy": decyzje lokalne w strukturze dwupoziomowej z koordynacją (neutralizującą konflikt na dolnym poziomie) są takie, jakie podjąłby sam koordynator przy maksymalizacji swego Q względem decyzji lokalnych $c_1, c_2, \ldots, c_N$, tj. z pominięciem lokalnych jednostek decyzyjnych. Można powiedzieć, że z punktu widzenia koordynatora jednostki lokalne działają jak zgadzający się z nim zespół.

W sytuacji niezgodności interesów, czyli wtedy gdy powstaje gra $(N+1)$-osobowa jednostki nadrzędnej i jednostek lokalnych, jednostka nadrzędna, czyli koordynator, nie zachowuje w pełni władzy nad systemem, pominięcie jednostek lokalnych i podjęcie samemu decyzji $c_i, \ldots, c_N$ dałoby mu wynik lepszy. Prawo pierwszego ruchu problemu tego nie rozwiązuje, daje tylko to, że rezultat koordynatora może być lepszy (a z pewnością nie gorszy, o czym wspomnieliśmy w rozdz. 2.4) od rezultatu, który byłby do osiągnięcia w $(N+1)$-osobowej grze graczy równoprawnych.

Przejdźmy teraz do struktury o pionowym układzie jednostek decyzyjnych, czyli do struktury wielowarstwowej. Tutaj także może wystąpić konflikt interesów – w istocie rzeczy taki właśnie konflikt opisywały wzory (2.21), (2.22), odnoszące się równie dobrze do relacji jednostki nadrzędnej z jedną jednostką dolnego poziomu.

Używając oznaczeń jak na rys. 5.4, tj. przypisując dwu sąsiadującym warstwom decyzje c^k, c^{k+1}, wprowadźmy ich indywidualne interesy w postaci maksymalizacji funkcji celu:
w warstwie $k+1$

$$\underset{c^{k+1}\in C^{k+1}}{\text{maksymalizować}}\ Q^{k+1}\left(c^{k+1}, w^{k+1}(c^k, c^{k-1}, \ldots, c^1)\right) \tag{5.12}$$

w warstwie k

$$\underset{c^{k}\in C^{k}}{\text{maksymalizować}}\ Q^{k}\left(c^{k}, w^{k}(c^{k-1}, c^{k-2}, \ldots, c^1)\right) \tag{5.13}$$

W zapisie (5.12) przyjęliśmy, że w^{k+1} oznacza te cechy przebiegu procesu w systemie sterowanym, które mają znaczenie dla oceny tego przebiegu przez jednostkę decyzyjną warstwy $k+1$. Cechy te zależą od decyzji podejmowanych przez wszystkie jednostki decyzyjne znajdujące się w warstwach poniżej $k+1$, w szczególności – od decyzji c^k warstwy podporządkowanej bezpośrednio. Wzory (5.12), (5.13) wskazują na istnienie sytuacji typu gry nie tylko między dwiema warstwami sąsiednimi, ale między wszystkimi warstwami pionowej struktury. Trzeba bowiem zauważyć, że decyzja c^k podjęta w warstwie k zależeć będzie od decyzji wyższej c^{k+1}, na przykład przez to, iż decyzja c^{k+1} będzie określać zbiór dopuszczalny dla decyzji c^k, oznaczony we wzorze (5.13) przez C^k.

Byłoby rzeczą naturalną w parze jednostek decyzyjnych $k+1$, k przyznać prawo pierwszego ruchu jednostce wykonującej zadanie (5.12), tworząc układ *leader-follower*. Relacje między jednostkami w pionowej strukturze nie sprowadzają się jednak do jednej pary – mielibyśmy więc do czynienia z „wielopiętrową” grą hierarchiczną.

Przy omawianiu struktur warstwowych w rozdziale 4 nie wspominaliśmy o tym, że może w takiej strukturze wystąpić sytuacja typu gry. Rozpatrywaliśmy wyłącznie sytuacje bez konfliktu interesów. Przede wszystkim trzeba zwrócić uwagę, że konfliktu takiego nie ma w strukturze dwuwarstwowej, gdzie tylko warstwa wyznaczania zadań (optymalizacji) operuje funkcją celu, związaną z biegiem procesu sterowanego. Warstwa sterowania bezpośredniego ma tylko nadążać za poleceniami. Omawiając układy o większej liczbie warstw, w szczególności w rozdz. 4.6, odnoszącym się do zróżnicowania horyzontu sterowania operowaliśmy – w każdej warstwie – tą samą, funkcją celu Q, mieliśmy zatem sytuację typu zespołu: wszystkie jednostki decyzyjne przyczyniały się do tego samego celu.

Nie można osądzić, na ile występowanie konfliktu między warstwami jest istotne dla sytuacji praktycznych. Można przypuszczać, że jeżeli struktura warstwowa jest rezultatem określonego zamysłu organizacyjnego, będzie w niej raczej panowała sytuacja zgodności stawianych celów. Mamy tu do czynienia, mówiąc w skrócie, z *podporządkowaniem kompetencji*, podczas gdy w strukturze z jednostkami równoległymi występuje ich *rozproszenie*.

Podkreślmy, że rozproszenie kompetencji wymaga koordynacji – także wtedy, gdy występuje opisana wcześniej „zgodność celów". Podporządkowanie kompetencji dodatkowej koordynacji nie potrzebuje, zawiera ją jakby w swoim wnętrzu.

5.6. Mechanizmy działania jednostek decyzyjnych

Podstawowe rodzaje mechanizmów decyzyjnych

Przekształcanie obserwacji w sterowanie, czyli informacji w decyzję, może być dokonywane w różny sposób, to jest przy użyciu różnych mechanizmów. Najprostszym z nich jest wspominane już parokrotnie prawo sterowania, czyli *reguła decyzyjna*: funkcja, która przyporządkowuje każdej wartości obserwacji systemu i otoczenia odpowiadającą im wartość oddziaływania sterującego, czyli określoną decyzję

$$c = d(w, z) \tag{5.14}$$

przy czym stanowiąca regułę decyzyjną funkcja $d(\cdot)$ jest dana w postaci wzoru, wykresu lub tabeli.

Istotną cechą reguły decyzyjnej jest to, że znajomość właściwego zadania sterowania, a więc funkcji celu, ograniczeń itp. elementów, a także modelu procesu jest potrzebna tylko przy projektowaniu tej reguły. W trakcie stosowania reguły decyzyjnej, czyli przy określaniu konkretnych oddziaływań (konkretnych decyzji), elementy zadania decyzyjnego nie są już potrzebne. Operowanie regułą decyzyjną przez regulator lub przez program komputerowy nie budzi zastrzeżeń, natomiast w przypadku zalecenia takiego postępowania operatorowi ludzkiemu, trzeba uczynić założenie, że nie ma on własnych celów sterowania lub innych motywacji, by swoją decyzję uczynić inną niż to wynika z narzuconej reguły. Przypominaliśmy o takich sytuacjach w poprzednim podrozdziale.

Drugi sposób przekształcania informacji w sterowanie polega na tym, że dla każdej nowej wartości obserwacji w oraz z przeprowadza się obliczenie (bądź rozumowanie), biorące pod uwagę pierwotne zadanie sterowania, to znaczy funkcję celu, ograniczenia, właściwości obiektu, prognozę oddziaływania zewnętrznego itd., otrzymując w wyniku odpowiadającą tym elementom wartość decyzji c. W przypadku realizacji komputerowej jest to równoważne wzorowi (5.14), z tą różnicą, że zamiast posługiwać się ustaloną *a priori* funkcją bądź tabelą przewidującą wszelkie możliwe wartości obserwacji w, z, obliczenie przeprowadzane jest na bieżąco, zatem tylko dla tych kolejnych wartości w, z, które rzeczywiście wystąpiły. Główną zaletą takiego mechanizmu decyzyjnego jest to, że łatwo można modyfikować elementy zadania decyzyjnego, na przykład zmieniać funkcję celu albo wartość ograniczeń – z różnych przyczyn, m.in. w myśl wskazań nadchodzących od nadrzędnej jednostki decyzyjnej.

Mówimy o *jawnym mechanizmie decyzyjnym*, w odróżnieniu od reguły decyzyjnej, gdzie te same elementy zadania były wzięte pod uwagę przy jej ustalaniu, ale nie są już widoczne.

Wyznaczenie decyzji może być wykonane przez program komputerowy tylko przy wiarygodnie określonej funkcji celu, ograniczeniach, modelu systemu sterowanego oraz pełnym sformalizowaniu zarówno niepewności jak stosunku do ryzyka. Nie musi to oznaczać „sterowania automatycznego"; decyzja z komputera może być wprowadzona w życie z udziałem operatora. Nie oczekuje się jednak od niego kwestionowania słuszności decyzji wyznaczonej przez program w komputerze.

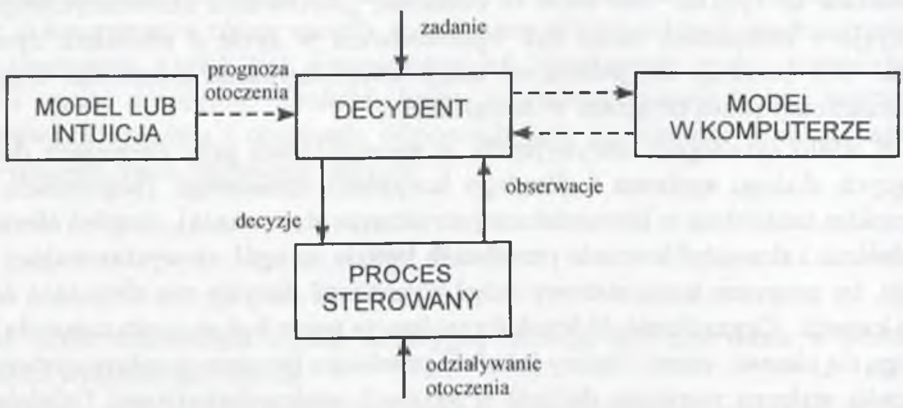
W wielu sytuacjach decyzyjnych, w szczególności przy decyzjach dotyczących dużego systemu i długiego horyzontu czasowego (odpowiada to wysokim szczeblom w hierarchicznej strukturze sterowania), stopień sformalizowania i skwantyfikowania przesłanek będzie na ogół niewystarczający do tego, by program komputerowy mógł wyznaczyć decyzję nie ulegającą żadnej kwestii. Czynnikami, których formalizacja może być za mało miarodajna mogą się okazać: oceny (miary jakości) przebiegu procesu w całym systemie, sposób wyboru wariantu decyzji w sytuacji wielowskaźnikowej (wielokryterialnej), prognoza wpływu otoczenia, a zwłaszcza to, jaki scenariusz, czy scenariusze oddziaływań otoczenia przyjąć za miarodajne dla rozpatrywanej decyzji, ograniczenia i ich podział na „miękkie" (przekraczalne) i „twarde" (nieprzekraczalne). Niepewny może być także przyczynowo-skutkowy model procesu sterowanego, to znaczy model, który służy do określenia przewidywanej reakcji na sterowanie.

Jeśli choć jeden element spośród użytych w komputerowej procedurze wyznaczania decyzji jest nie w pełni wiarygodny, nie w pełni wiarygodna jest także wyznaczona decyzja. Trzeba w tej sytuacji sięgnąć do osądu decydenta; osąd ten może być wykorzystany:

a) do *weryfikacji* decyzji „komputerowej", na podstawie jej przewidywanych skutków; w sytuacji odrzucenia decyzji decydent zmienia niektóre założenia lub parametry w sformułowaniu problemu decyzyjnego, na przykład sposób analizowania sytuacji wielokryterialnej lub współczynniki w funkcji celu, a następnie oczekuje na ponowne wykonanie całości procedury wyznaczenia decyzji – jest to postępowanie iteracyjne;

b) do *współdziałania interakcyjnego* z procedurą wyznaczania decyzji przez komputer w tych miejscach, gdzie przesłanki sformalizowane są niepewne; decydent wprowadza swoje decyzje (dokonuje wyborów) w takich elementach jak: uznanie określonej prognozy za najbardziej miarodajną, wskazanie priorytetu jednych ograniczeń wobec innych, określenie wzajemnych wag i znaczenia różnych ocen itd. Decyzja wyznaczona w wyniku działania interakcyjnego jest dla decydenta bardziej wiarygodna, gdyż jest ona podsumowaniem jego własnych decyzji cząstkowych; postępowanie iteracyjne w postaci powrotu na sam początek procedury może już być niepotrzebne.

Przykładową strukturę mechanizmu decyzyjnego ze wspomaganiem komputerowym przedstawia rys. 5.6. Można sobie wyobrazić, że mechanizmy takie znajdują się na różnych szczeblach struktury sterowania, np. tej z rys. 5.3.



Rys. 5.6. Struktura mechanizmu decyzyjnego ze wspomaganiem komputerowym

Jako bardzo prosty przykład wyobraźmy sobie wyznaczanie decyzji o zrzuceniu wody ze zbiornika retencyjnego w sytuacji powodziowej. Decydent może decyzję sugerowaną przez model komputerowy odrzucić, uważając, że fala powodziowa wyliczona przez komputer jako skutek zrzutu wody jest za duża (decydent uznaje model procesu zawarty w komputerze za poprawny, ale nie zgadza się z decyzją) albo że rzeczywista fala powodziowa przy proponowanym zrzucie może okazać się większa i zbyt niebezpieczna (decydent kwestionuje decyzję, ale nie dowierza także modelowi systemu łącznie z przyjętą prognozą opadów). Postępowanie *iteracyjne* polegałoby na powtórzeniu procedury ze zmienioną funkcją celu lub zmienionymi ograniczeniami; w postępowaniu *interakcyjnym* można by przedstawić decydentowi warianty prognoz opadów, prosząc o wybór prognozy najbardziej miarodajnej dla rozpatrywanej sytuacji decyzyjnej (zapewne nie będą to opady o największym prawdopodobieństwie, lecz wariant bardziej krytyczny).

W sumie, w układach sterowania złożonymi systemami mogą być używane następujące mechanizmy decyzyjne:

1) prawa sterowania lub reguły decyzyjne, w których nie występują w sposób jawny takie elementy, jak funkcja celu i model procesu,
2) komputerowe procedury wyznaczania decyzji, do których jawnie wprowadza się przesłanki, takie jak funkcje celu, model procesu, ograniczenia,
3) procedury komputerowego wspomagania decyzji, wykorzystujące osąd decydenta ludzkiego, w sposób iteracyjny lub interakcyjny.

Spośród wymienionych, tylko mechanizmy pierwszej grupy są z góry przygotowane do użycia sprzężenia zwrotnego, to jest obserwacji stanu samego systemu; w gruncie rzeczy ich istotę stanowi zaprojektowanie sprzężenia zwrotnego, realizującego zadanie sterowania przebiegającym procesem. Procedury wyznaczania decyzji z jawnie wprowadzanymi przesłankami, takimi jak funkcja celu i inne elementy, pochodzą w zasadzie z dziedziny decyzji jednokrotnych, w szczególności z dziedziny programowania matematycznego. Realizacja czynności sterowania wymaga wielokrotnego użycia mechanizmu decyzyjnego, gdyż ma to być interwencja podejmowana w trakcie biegu procesu sterowanego. Odnosząc to do mechanizmów typu (2) oraz (3), rozróżnić trzeba dwa przypadki, a mianowicie sterowanie w układzie otwartym oraz ze sprzężeniem zwrotnym.

Układy otwarte i ze sprzężeniem zwrotnym

Przy sterowaniu w układzie otwartym powtarzanie czynności podejmowania decyzji związane jest tylko ze zmianami obserwowanymi w otoczeniu systemu. Podejmując decyzję w chwili t_i mieć będziemy na uwadze nie tylko obecny stan otoczenia $z(t_i)$, ale i jego stany przyszłe, wraz z wchodzącą tu w grę niepewnością. Jeśli w grę wchodzi np. naruszenie ograniczeń, to decyzja w chwili t_i musi być odpowiednio ostrożna: musi być bardzo ostrożna, jeśli postanowimy być zabezpieczeni na całą przyszłość $t > t_i$, może być mniej ostrożna jeśli podejmując decyzję w chwili t_i uwzględnimy, że w niezbyt odległej chwili t_j będziemy podejmować decyzję następną; mówiliśmy o tym już w rozdziale 2.6, klasyfikując decyzje w obliczu niepewności na jednoetapowe i wieloetapowe.

Sterowanie w układzie ze sprzężeniem zwrotnym ma wykorzystać obserwacje procesu sterowanego; wystąpią tu istotne różnice w zależności od tego, czy z punktu widzenia danej jednostki decyzyjnej mamy do czynienia z zadaniem statycznym czy dynamicznym. Dopuszczalność „statycznego" potraktowania procesu dynamicznego można uzasadnić tylko na drodze ilościowej, tj. przez odpowiedź na pytanie, na ile gorszy będzie wynik sterowania; rozważać ją zatem trzeba w każdym konkretnym przypadku. Tym niemniej wiele jest takich sytuacji, gdy decyzje wyższej warstwy sterowania można realizować jako optymalizację statyczną, tj. jako wyznaczanie kolejnych pożądanych stanów ustalonych procesu sterowanego; mówiliśmy o tym dość obszernie w rozdziale 4.5. Istotą rzeczy jest tu uznanie, że w procesie sterowanym nie następuje jakakolwiek akumulacja. Decyzja podejmowana w chwili t_i nie musi więc np. brać pod uwagę, że wyczerpuje się jakiś zasób, którego zabraknie na chwilę późniejszą; nie potrzeba myśleć o przyszłości, nie potrzeba brać pod uwagę horyzontu sterowania poza tym, co wynika z możliwych zmian otoczenia systemu – ale i to tylko w okresie do następnej chwili decyzyjnej t_j. Postępujemy więc analogicznie jak przy sterowaniu w układzie otwartym. Użycie sprzężenia zwrotnego przy podejmowaniu decyzji sterowania statycznego w chwili t_i wiąże się tylko z brakiem pewności co do modelu procesu, z chęcią sprawdzenia czy określone zmienne w systemie rzeczywistym przyjmują wartości zgodne z przewidywanymi przez model. Wykorzystanie takich sprzężeń do ulepszenia decyzji sterowania statycznego, a zwłaszcza do algorytmów koordynacji, może mieć istotne znaczenie. Omawiane sprzężenie zwrotne jest m.in. jedynym sposobem, dzięki któremu przy koordynacji metodą cen w systemie działającym, tj. bez przerwanych połączeń między „produkcją" i „konsumpcją" można korygować ceny w dostosowaniu do różnic między modelem a rzeczywistością, (por. rozdz. 2.4 oraz obszerną analizę w monografii z 1980 r.)*).

Przy sterowaniu procesem dynamicznym decyzja podejmowana w chwili t_i musi uwzględniać przyszłość wynikającą ze zjawiska akumulacji w systemie, tj. musi być podejmowana z odpowiednim horyzontem czasowym. Stosownie do tego, sprzężenie zwrotne musi nieść informację o stanie zasobów, w terminologii bardziej ścisłej – o stanie procesu. W mechanizmach decyzyjnych typu „prawo sterowania" dzieje się to przez bieżące mierzenie współrzędnych stanu lub ich mniej lub więcej pośrednią, ale również bieżącą obserwację. Opowiednikiem tego dla mechanizmów wyznaczających decyzje z jawnym użyciem funkcji celu, ograniczeń i innych elementów zadania decyzyjnego jest tak zwana struktura repetycyjna. W myśl tej struktury, zwanej także regulacją predykcyjną**), w danej chwili t_i, wyznacza się sterowanie korzystając ze zmierzonego stanu procesu $x(t_i)$ w taki sposób, jakby to było sterowanie w układzie otwartym (stan początkowy $x(t_i)$, wyliczenie sterowania oraz trajektorii stanu do końca horyzontu sterowania). W chwili

*) W. Findeisen i in., op. cit., rozdz. 3.

**) K. Nowosad (1995). Strukturalne cechy regulacji predykcyjnej. Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa.

$t_j = t_i + h$, gdzie h jest odstępem czasu krótszym od horyzontu sterowania (zwykle znacznie krótszym), nowy zmierzony stan procesu $x(t_j)$ staje się „stanem początkowym" dla ponownie przeprowadzonego, analogicznego jak w chwili t_i, wyliczenia sterowania i trajektorii stanu. Z pierwszego wyliczenia realizowane jest tylko sterowanie w przedziale $[t_i, t_i + h]$. Z drugiego wyliczenia – sterowanie w przedziale $[t_j, t_j + h]$, i tak dalej. Struktura repetycyjna daje się m.in. dobrze wykorzystać przy podejmowaniu decyzji z udziałem człowieka; w istocie rzeczy tak zapewne działają decydenci kierujący procesami dynamicznymi bez pomocy modeli matematycznych i komputerów.

Jednostki decyzyjne różnych warstw i poziomów sterowania

Wielokrotnie już mówiliśmy, poczynając od rozdziału 1, o zróżnicowaniu rozległości oraz horyzontu czasowego w kolejnych warstwach i poziomach hierarchicznej struktury sterowania (większa rozległość i dłuższy horyzont na wyższych szczeblach). W wielu przypadkach dochodzi do tego zróżnicowanie środków (instrumentów) oddziaływania. Równocześnie ze zróżnicowaniem środków występuje zróżnicowanie zadań – cele do osiągnięcia przez jednostkę decyzyjną wysokiego szczebla stają się coraz bliższe co do swego charakteru celom stawianym przed całym systemem, podczas gdy niżej formułowane są cele lokalne.

W ślad za zróżnicowaniem zasięgu przestrzennego i czasowego, instrumentów działania, celów do osiągnięcia, a także – co bardzo ważne – częstotliwości interwencji, idzie zróżnicowanie potrzebnych, czy też możliwych do zastosowania mechanizmów decyzyjnych oraz środków technicznych, którymi będą się one posługiwać.

Nie jest oczywiście możliwe ustalenie absolutnych zasad, np. takich, że sterowanie dotyczące bezpośrednio zjawisk dynamicznych w systemie (pierwsza warstwa sterowania) powinno być realizowane za pomocą reguł decyzyjnych (praw sterowania), a ponadto powinno być automatyczne. Można by raczej powiedzieć, że sterowanie bezpośrednie na ogół *może być* realizowane za pomocą reguł decyzyjnych, zatem jest stosunkowo łatwe do zautomatyzowania.

Dla wyższych warstw i poziomów sterowania nie da się podać nawet takich stwierdzeń, poza konstatacją ogólną, że im wyższy poziom, większa rozległość sterowania i dłuższy horyzont czasowy, tym trudniej o obiektywizację wyznaczania decyzji w stopniu pozwalającym całkowicie wyręczyć decydenta. Będziemy więc tu raczej mieli komputerowe wspomaganie decyzji, a nie sterowanie automatyczne; ponadto częstotliwość interwencji jednostek decyzyjnych wyższych warstw i poziomów sterowania jest mała i decydent może być w stanie decyzje te podejmować.

Zastanawiając się nad tym, jak ukształtować mechanizm jednostki decyzyjnej przewidzianej w określonym miejscu struktury sterowania, wziąć trzeba pod uwagę następujące elementy:

1) jakie są instrumenty oddziaływania danej jednostki decyzyjnej na jednostki niższego poziomu (wartości zadane określonych zmiennych, parametry funkcji celu, wartości ograniczeń, przepisy (regulaminy) funkcjonowania itd.);
2) jakie jest zadanie stawiane przez jednostkę nadrzędną (osiągnięcie określonego punktu bądź stanu, realizacja przebiegu czasowego, nieprzekroczenie zasobu, optymalizacja narzuconej funkcji celu, uwzględnienie czynników niekwantyfikowalnych, stosunek do ryzyka itd.);
3) jaka jest relacja do innych jednostek decyzyjnych tego samego poziomu, w szczególności czy ma miejsce interferencja, tj. wzajemna zależność skutków decyzji między jednostkami równoległymi;
4) jaki jest charakter i jakość przesłanek do decyzji (model obiektu sterowania sformalizowany czy intuicyjny, na ile wiarygodny, jaki stopień rozpoznania otoczenia, na ile sprecyzowane cele do osiągnięcia, jaka informacja bieżąca o systemie sterowanym jest lub może być dostępna, co jest wiadome o decydentach „równoległych" itd.);
5) jaki jest horyzont czasowy i jak często dana jednostka decyzyjna ma podejmować swoje decyzje (ma to wpływ zarówno na dopuszczalny stopień skomplikowania procesu decyzyjnego, jak i na potrzebę, racjonalność oraz stopień automatyzacji).

Można takich elementów wypisać więcej, można je też inaczej ująć; w istocie chodzi o wyłuskanie odpowiedzi na następujące pytania, związane już dosyć blisko z wyborem mechanizmu działania jednostki decyzyjnej oraz ukształtowaniem samego procesu (algorytmu) ustalania decyzji sterującej:

a) czy można z dostatecznym stopniem wiarygodności zobiektywizować oraz sformalizować proces wyznaczania decyzji?
b) czy sytuacja decyzyjna w grupie jednostek decyzyjnych (na różnych lub na tym samym poziomie) jest sytuacją, w której dochodzi do głosu niezgodność interesów, czy też układ celów i wybranych instrumentów koordynacji stworzył sytuację działania bezkonfliktowego?

Pozytywna odpowiedź na pytanie (a) składa się z wielu oddzielnych elementów (cele, model, ograniczenia, otoczenie, prognoza, stosunek do ryzyka itd.). Jeśli formalizacja pełna jest możliwa, można proces decyzyjny zmechanizować całkowicie, tj. można ułożyć program dla komputera. Jeśli odpowiedź na pytanie (a) jest częściowo pozytywna, sięgnąć trzeba do komputera wspomaganego intuicją, rozumowaniem i doświadczeniem decydenta ludzkiego, najlepiej w formie interakcyjnej. W szczególnych przypadkach może to także być grupa decydentów, reprezentujących różne poglądy na te same sprawy, z programem komputerowym typu wspomagania decyzji grupowej.

Jeśli odpowiedź na pytanie (a) jest całkowicie negatywna, tj. żaden z elementów używanych w procesie podjęcia decyzji nie może być sformalizowany, mechanizm wspomagania komputerowego decyzji indywidualnej lub grupowej może być nadal przydatny. Można, na przykład, używać nieformalnych, tj. intuicyjnych modeli obiektu sterowania w taki sposób, w jaki używamy modeli sformalizowanych, z symulacją stochastyczną włącznie.

Odpowiedź na pytanie (b) o rodzaj sytuacji decyzyjnej prowadzi z punktu widzenia poszukiwania algorytmów i procedur wyznaczania decyzji, w jeden z dwóch obszarów: w obszar objęty teorią hierarchicznych układów sterowania o silnej (tj. neutralizującej konflikt) koordynacji, w zasadzie już dość dobrze znany, bądź też w obszar należący do teorii gier (w tym gier hierarchicznych i teorii przetargów), w którym mało jest dotąd rozwiązań konstruktywnych, tj. pozwalających budować algorytmy postępowania dla konkretnych zadań.

Decyzje w warunkach niepewności i ograniczeń

Wspomnieliśmy w tym rozdziale, a także w rozdziale 2.6, o decyzjach jedno- i wieloetapowych, w kontekście losowości czynników zewnętrznych oraz obowiązujących dla przebiegu procesu sterowanego ograniczeń.

Sytuacji takiej warto poświęcić nieco więcej uwagi.

Przypomnijmy, że sterowaniu podlega proces dynamiczny, opisany równaniem stanu

$$\dot{x}(t) = f\left(x(t), u(t), z(t)\right) \tag{5.15}$$

gdzie przebieg czynnika zewnętrznego w przedziale $[t_0, t_k]$ jest w chwili t_0 nie znany, ale wiemy że zawiera się w znanym nam „zbiorze niepewności"

$$z_{[t_0,t_1]} \in Z \tag{5.16}$$

Załóżmy, że obowiązują ograniczenia na stan procesu i na sterowanie (mogą one być zadane jako funkcje czasu)

$$x(t) \in X(t) \tag{5.17}$$

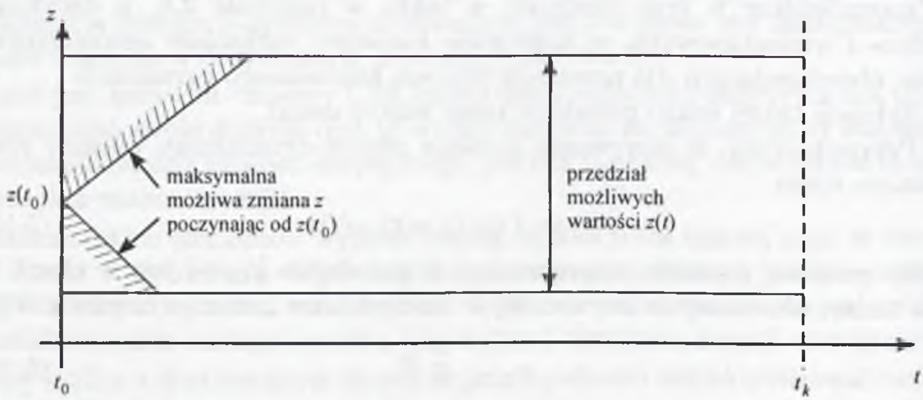
$$u(t) \in U(t) \tag{5.18}$$

oraz że znamy stan początkowy $x(t_0)$.

Pierwszym narzucającym się pytaniem jest: czy dla *każdego* przebiegu $z_{[t_0,t_k]}$ istnieje *sterowanie dopuszczalne* $u^d_{[t_0,t_k]}$, tj. takie, przy którym nie następuje przekroczenie ograniczeń? Odpowiedź twierdząca oznacza potencjalną możliwość sterowania tym procesem bez przekroczenia ograniczeń; nie oznacza to jeszcze, że potrafimy to zrealizować, bowiem w chwili t_0, kiedy obliczamy sterowanie albo ustalamy regułę decyzyjną znamy co najwyżej wartość $z(t_0)$, a nie przebieg przyszły $z_{[t_0,t_k]}$.

Na drugim jakby krańcu jest pytanie: czy istnieje *sterowanie bezpieczne* $u^b_{[t_0,t_k]}$, tj. takie, przy którym nie nastąpi przekroczenie ograniczeń dla *dowolnego* spośród przebiegów, mieszczących się w znanym nam zbiorze Z? Gdyby sterowanie takie istniało i dało się wyznaczyć, można by je wprost zastosować w układzie otwartym. Jeżeli takich sterowań jest wiele, można dokonać wyboru porównując osiągany wskaźnik jakości.

Istnienie sterowania bezpiecznego na cały przedział $[t_o, t_k]$ jest mało prawdopodobne w krytycznych sytuacjach decyzyjnych („szeroki" zbiór Z, „wąskie" ograniczenie na stan), ale można liczyć na istnienie sterowania bezpiecznego, gwarantującego nieprzekroczenie ograniczeń w początkowym odcinku przedziału sterowania $[t_0, t_i]$, gdzie $t_i < t_k$. Sprzyja temu fakt, że znamy $x(t_0)$ – podczas gdy późniejsze $x(t)$ będą losowe – oraz znamy być może wartość $z(t_0)$; jeśli z nie jest białym szumem, lecz przebiegiem, mającym pewną autokorelację lub ograniczoną prędkość zmian, zbiór niepewności Z będzie dla rozważań odnoszących się do początkowego odcinka czasu „zawężony", patrz rys. 5.7.



Rys. 5.7. Zbiór niepewności „zawężony" dzięki znajomości wartości początkowej $z(t_0)$

W tym właśnie zawiera się istota decyzji wieloetapowych: w chwili t_0 dbamy tylko o bezpieczeństwo w czasie $[t_0, t_i]$, w chwili t_i zaczynamy od nowa – znając nowe $x(t_i)$ oraz ewentualne nowe „zawężenie" zbioru niepewności Z.

Wspomnieliśmy, że w przypadku sterowania procesem dynamicznym sprawa nieco się komplikuje; jeżeli bowiem przy sterowaniu bezpiecznym na odcinku $[t_0, t_i]$ stan procesu w chwili t_i znajdzie się na brzegu obszaru dopuszczalnego („dotknie" ograniczeń), to sterowanie na początku następnego odcinka – będąc ograniczone – może nie być w stanie zapobiec przekroczeniu tej granicy (zależy to od postaci równań procesu, nie mówiąc już o wpływie ewentualnych opóźnień). Układ sterowania powinien to odpowiednio uwzględnić.

Sterowanie procesem (5.15), poddanym czynnikowi zewnętrznemu (5.16) może być realizowane za pomocą reguły decyzyjnej, czyli prawa sterowania, być może zmiennego w czasie

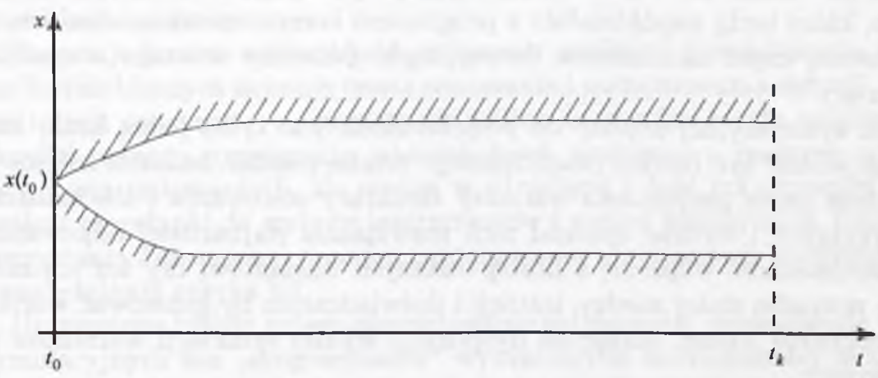
$$u(t) = d\,(x(t), t) \tag{5.19}$$

Bezpieczne prawo sterowania będzie to taka postać funkcji $d(\cdot)$, przy której generowane będzie sterowanie dopuszczalne $u^d_{[t_0,t_k]}$ dla dowolnego spośród przebiegów $z_{[t_0,t_k]} \in Z$. Określenie takiego prawa w chwili t_0, w obliczu całej przyszłej niepewności, może się nie udać. Można jednak także tutaj zastosować podejście wieloetapowe: w chwili t_0 ustalić prawo sterowania bezpieczne na odcinek $[t_0, t_i]$, w chwili t_i operację tę powtórzyć, itd. Byłoby to zadanie dla drugiej warstwy sterowania.

Omawiając sterowanie w podziale na warstwy w rozdz. 4.3–4.5, mówiliśmy o sytuacjach, gdy warstwa wyższa narzuca trajektorię stanu procesu, a warstwa sterowania bezpośredniego ma zapewnić nadążanie za tym zadaniem. W kontekście niepewności i ograniczeń oznacza to potrzebę zastanowienia nad tym, jaką trajektorię wolno nam narzucić.

Potrzebne jest pojęcie *trajektorii realizowalnej*, to jest takiej, którą da się utrzymać przy *dowolnym* spośród $z_{[t_0,t_k]} \in Z$, bez przekroczenia ograniczeń na sterowanie.

Trajektoria realizowalna może nie istnieć, na przykład może nie być możliwe ścisłe utrzymanie ciśnienia czy temperatury w reaktorze, przy dowolnym spośród zakłóceń, może jednak być możliwe utrzymanie współrzędnych stanu w jakimś przedziale. Można wtedy mówić o *realizowalnym paśmie*, jako o sytuacji mniej wymagającej (rys. 5.8).



Rys. 5.8. Realizowalne pasmo – przedział, w którym można utrzymać trajektorię $x_{[t_0,t_k]}$ przy dowolnym $z_{[t_0,t_k]} \in Z$

Narzucenie realizowalnej trajektorii czy realizowalnego pasma w sposób dla siebie dogodny należałoby do wyższej warstwy sterowania; podobnie jak przy narzucaniu reguły decyzyjnej, może to być powtarzane w chwilach t_i.

5.7. Narzędzia analizy i projektowania

Przypomnijmy, że w zadaniu sterowania chodzi o zapewnienie pożądanego przebiegu procesu (procesów) w sterowanym systemie; zadanie to trzeba rozumieć bardzo wszechstronnie (por. np. rys. 5.3), a jego realizacja – w rozpatrywanych sytuacjach – będzie powierzona układowi o złożonej strukturze, zawierającemu wiele jednostek decyzyjnych. Będą wśród nich „urządzenia sterujące", działające automatycznie, ale będą też miejsca gdzie decyzje są podejmowane z udziałem człowieka lub wyłącznie przez jakąś osobę albo grupę osób. Całość składa się na strukturę o wielu jednostkach decyzyjnych, a w tej chwili zastanawiamy się nad narzędziami analizy i projektowania takich struktur.

Zadanie *analizy* struktury już istniejącej bądź też kolejnego wariantu struktury, podlegającej projektowaniu, polega na określeniu jej przewidywanego zachowania w różnych warunkach, które mogą zaistnieć w rzeczywistości. Chodzi tu o działanie systemu sterowanego wraz z układami sterowania w przyszłości, tj. w przyszłych warunkach stworzonych przez otoczenie.

Nie potrzeba uzasadniać, że jedynym narzędziem które można tu skutecznie zastosować jest *symulacja*, przy której używa się modeli systemu, modeli otoczenia i modeli jednostek decyzyjnych.

Mówimy z premedytacją o „modelach" jednostek decyzyjnych. Jeśli bowiem mieć będziemy, w pewnych punktach struktury, mechanizmy decyzyjne korzystające z ocen bądź wyborów dokonywanych przez człowieka, to w analizie symulacyjnej musimy to odpowiednio modelować. Może to m.in. prowadzić do sensowności korzystania z odpowiednio przygotowanych osób, które będą współdziałały z programem komputerowym, stanowiącym pozostałą część mechanizmu decyzyjnego. Mówimy wówczas o symulacji typu gry*).

Od symulacyjnej analizy do *projektowania* jest tylko jeden krok, który może jednak być bardzo pracochłonny: trzeba poddać badaniu kolejne pomyślane przez projektanta warianty struktury sterowania i mechanizmów decyzyjnych i wybrać spośród nich rozwiązanie najbardziej odpowiednie. Pracochłonność wiąże się z liczbą badanych wariantów, czy też ich mutacji – potrzeba dużej wiedzy, intuicji i doświadczenia by generować warianty potencjalnie lepsze, mając do dyspozycji wyniki symulacji wariantów poprzednich.

---

*) H.J. Miser, E.S. Quade, eds. (1988). Handbook of Systems Analysis – Craft Issues and Procedural Choices. North-Holland.

Specyfika projektowania układu o wielu jednostkach decyzyjnych polega m.in. na tym, że trzeba rozstrzygnąć następujące kwestie:

– kto, to znaczy która jednostka decyzyjna podejmuje określony rodzaj decyzji – *alokacja kompetencji*,
– jakie jest zadanie stawiane przed daną jednostką decyzyjną (uniezależnienie procesu od wpływów otoczenia lub przeciwnie, dostosowanie do potrzeb zewnętrznych, maksymalizacja przydzielonej funkcji celu itd.) – *alokacja zadań*,
– na jakiej podstawie będą opierane decyzje tej jednostki – *alokacja informacji*,
– jakie mają być użyte *mechanizmy decyzyjne* w poszczególnych miejscach w strukturze sterowania.

Odpowiedzi na te pytania składają się na *projekt struktury*, którą można następnie poddawać analizie, jak to już wspomnieliśmy wcześniej.

Trzeba może podkreślić, w ślad za rozdz. 5.2, że niekoniecznie trzeba się zajmować całością układu; konkretna potrzeba może dotyczyć jednej tylko strefy z przykładowego rys. 5.3, a więc np. tylko sterowania procesami fizycznymi albo tylko strefy planowania rzeczowego i sterowania produkcją itp.

W takim podejściu nie ma nic niewłaściwego, zwłaszcza gdy zauważymy, że działanie istniejących w rzeczywistości innych, tj. wyższych bądź niższych stref i szczebli sterowania może nie być dostatecznie rozpoznane, tym bardziej zaś – sformalizowane. Błędem może być natomiast ich ignorowanie, w szczególności zaś przeoczenie, że mogą się tam znajdować jednostki decyzyjne o własnych celach. Ważną tę sprawę sygnalizowaliśmy już w rozdz. 1.

Istotne jest także pamiętanie o tym, że zajmujący się różnymi strefami specjaliści będą nazywali „systemem", a zwłaszcza „procesem w systemie", rzeczy dosyć różne.

Próbowaliśmy wyjaśnić w rozdz. 5.2, a także w 5.3, na ile te różnice są pozorne, jaki jest związek między „procesami" rozpatrywanymi na różnych szczeblach struktury.

Pozostaje jeszcze wskazać jakie miejsce w analizie i projektowaniu struktur hierarchicznych zajmuje teoria sterowania i podejmowania decyzji. Miejsce to jest znaczące – wprawdzie nie można na drodze czysto teoretycznej określić pełnego rozwiązania jakiegokolwiek problemu o realnych rozmiarach i uwarunkowaniach, ale można w istniejącej i dość już okrzepłej teorii znaleźć przesłanki do wyboru instrumentów i metod koordynacji, sposobów korzystania z obserwacji procesu, postępowania w sytuacjach konfliktowych, uwzględniania ryzyka itd.

Uprawnione będzie zatem sformułowanie ostrzeżenia: prowadzenie badań symulacyjnych bez „drogowskazów" wynikających ze znajomości teorii sterowania i podejmowania decyzji może być nie tylko bardziej pracochłonne, ale grozi rozpatrywaniem wariantów dalekich od wykorzystania możliwości systemu, niedostrzeżeniem rozwiązań najbardziej skutecznych.

Monografie i podręczniki

Zestawienie literatury przedmiotu, a raczej obszaru, którego dotyka ta książka byłoby zadaniem znacznie przekraczającym możliwości autora – w grę wchodzi zapewne parę tysięcy pozycji bibliograficznych. Zestawienie pełnej literatury nie było też konieczne, gdyż książka nie pretenduje do roli przewodnika wskazującego sposoby rozwiązywania zagadnień szczegółowych.

Poszukiwaliśmy raczej pewnej syntezy albo lepiej – filozofii patrzenia na zadania i struktury sterowania, konkretne problemy i przykłady traktując jedynie jako ilustracje. Stąd też przywoływane bezpośrednio pozycje literaturowe są podane w odnośnikach na odpowiednich stronach tekstu, a teraz – na zakończenie – zamieszczamy już tylko listę związanych z naszym tematem monografii i podręczników. Spis ten stanowi zarazem *sui generis* kronikę rozwoju myśli w zakresie sterowania złożonymi systemami. Z tego właśnie powodu ułożony on został nie w porządku alfabetycznym, lecz chronologicznie.

M.D. Mesarović, D. Macko, Y. Takahara (1970). Theory of Hierarchical, Multilevel Systems. Academic Press, New York.

R. Kulikowski (1970). Sterowanie w wielkich systemach. WNT, Warszawa

L.S. Lasdon (1970). Optimization Theory for Large Systems. MacMillan, London.

D.A. Wismer, ed. (1971). Optimization Methods for Large Scale Problems. McGraw – Hill, New York.

J. Marschak, R. Radner (1972). Economic Theory of Teams. Yale University Press, New Haven, Connecticut (tłumaczenie polskie: Ekonomiczna teoria zespołów. PWE. Warszawa 1977).

D.M. Himmelblau, ed. (1973). Decomposition of Large-Scale Problems. North-Holland, Amsterdam.

W. Findeisen (1974). Wielopoziomowe układy sterowania. PWN, Warszawa.

P. Bernhard (1976). Commande optimale, decentralisation et jeux dynamiques. Dunod, Paris.

J.B. Germejer (1976). Igry s neprotiwopołożnymi interesami. Nauka, Moskwa.

W.N. Burkow (1977). Osnowy matematiczeskoj tieorii aktiwnych sistem. Nauka, Moskwa.

M.G. Singh (1977). Dynamical Hierarchical Control. North-Holland, Amsterdam.

M.G. Singh, A. Titli (1978). Systems: Decomposition and Control. Pergamon Press, Oxford.

D.D. Siljak (1978). Large-Scale Dynamic Systems. North-Holland, New York.

I. Dirickx, L.P. Jennergren (1979). Systems Analysis by Multilevel Methods: With Applications to Economics and Management. Wiley, London.

W. Findeisen, F.N. Bailey, M. Brdyś, K. Malinowski, P. Tatjewski, A. Woźniak (1980). Control and Coordination in Hierarchical Systems. Wiley, London.

W. Findeisen, red. (1985). Analiza systemowa – podstawy i metodologia. PWN, Warszawa.

H. Tamura, T. Yoshikawa, eds. (1990). Large-Scale Systems Control and Decision Making. M.Dekker, New York-Basel.

M.A. Brdyś, K. Malinowski, eds. (1994). Computer-Aided Control System Design. World Scientific, Singapore-New Jersey-London-HongKong.

K. Malinowski, red. (1995). Sterowanie w systemach wodnych. Monografie Komitetu Gospodarki Wodnej PAN, z. 7. Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa.

Skorowidz

A

B

C

D

E

F

G

H

M

N

P

Q

R

T

Y

Z

9,

Potrzeba rozpatrywania systemów złożonych jest – być może – jednym ze znaków współczesnego czasu. Występuje ona w bardzo różnych dziedzinach, że wymienimy procesy technologiczne i produkcyjne, systemy transportowe i energetyczne, sieci telekomunikacyjne, zarządzanie w przedsiębiorstwach, ochronę środowiska, gospodarkę zasobami naturalnymi... Książka nie pretenduje do bezpośredniej przydatności w dziedzinach tak rozmaitych... Być może jednak niektóre koncepcje i prawidłowości sformułowane na podstawie procesów fizycznych i modeli matematycznych mogą okazać się pomocne także tam, gdzie modeli takich brakuje lub są bardzo niedokładne... W rozpoznaniu zakresu tematycznego może być pomocny skorowidz, który ujmuje ważniejsze hasła i terminy w ich kontekście merytorycznym...

Z przedmowy autora

Książkę przeczytałem z ową rzadką przyjemnością, jaką daje kontakt z myślą klarowną i uporządkowaną, będącą owocem bogatego i różnorodnego doświadczenia...

ISBN 83-87012-84-X