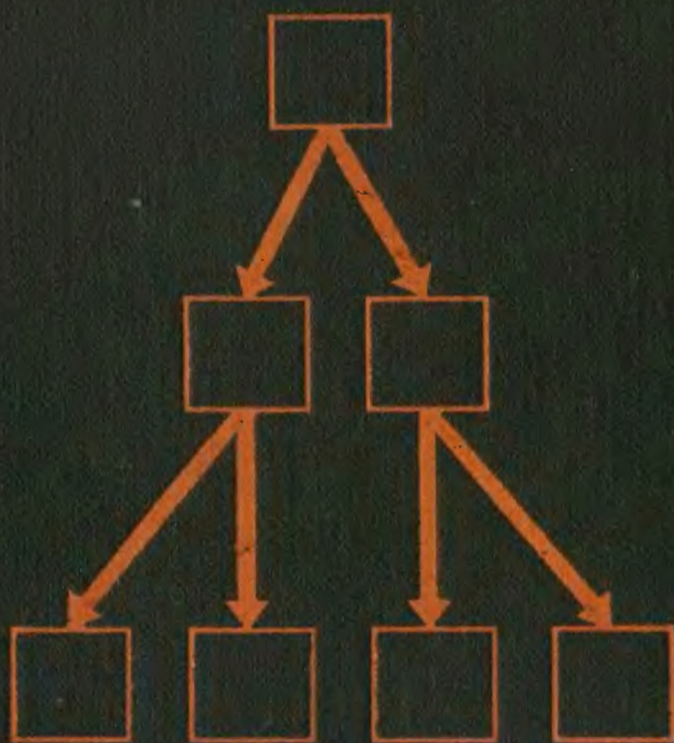


ТЕОРИЯ
ИЕРАРХИЧЕСКИХ
МНОГОУРОВНЕВЫХ
СИСТЕМ



М. Месарович, Д. Мако и И. Такахара

ТЕОРИЯ
ИЕРАРХИЧЕСКИХ
МНОГОУРОВНЕВЫХ
СИСТЕМ

*Перевод
с английского
под редакцией
И. Ф. Шахнова*

*Предисловие
чл.-корр.
АН СССР
Г. С. Поспелова*

ИЗДАТЕЛЬСТВО

«МИР»

МОСКВА 1973

M. D. Mesarovic, D. Macko, and Y. Takahara

THEORY
OF HIERARCHICAL,
MULTILEVEL,
SYSTEMS

*Systems
Research
Center.
Case Western
Reserve
University.
Cleveland,
Ohio*

ACADEMIC
PRESS

NEW YORK
AND LONDON
1970

Интерес к теории «больших систем», с которыми исследователям приходится иметь дело в самых различных областях — в науке об управлении, в экономике, социологии, психологии, физиологии нервной системы, биологии, — в последние годы непрерывно растет. Важным направлением исследования таких «больших» или «сложных» систем является рассмотрение их как многоуровневых систем, или систем с иерархической структурой. Однако стройной математической теории таких иерархических систем пока еще нет.

Книга известного исследователя в области общей теории систем М. Д. Месаровича, написанная им в соавторстве с Д. Мако и И. Такахара, является одной из первых в мировой литературе попыток систематического изложения и математической формализации теории управления в больших системах, построенных по иерархическому принципу. Идеи книги и полученные результаты представляют значительный интерес с точки зрения построения автоматизированных систем управления производством и отраслями народного хозяйства (АСУ).

Книга рассчитана на специалистов — математиков, инженеров и экономистов, работающих в области проблем управления, теории систем и исследования операций, а также на широкий круг научных работников и инженеров, интересующихся кибернетическим моделированием сложных систем любой природы.

Редакция литературы по новой технике

М $\frac{0334-350}{041(01)-73}$ 141-72

ПРЕДИСЛОВИЕ К РУССКОМУ ИЗДАНИЮ

Развитие электронно-вычислительной техники, математических методов кибернетики, системного анализа, исследования операций, экономико-математического моделирования и теории автоматического управления позволили в настоящее время решить ряд серьезных проблем, связанных с автоматическим и автоматизированным управлением сложными техническими объектами и целыми производственными комплексами. Примерами решения подобных задач могут служить автоматизированные системы управления рядом крупных химических производств, энергетических и металлургических комплексов и т. п.

Имеются определенные успехи и в создании теории автоматизированных систем управления предприятиями (АСУП) и отраслями промышленности (ОАСУ). Однако здесь предстоит еще очень много сделать, особенно когда речь идет об ОАСУ и их объединении в системы более высокого уровня. Круг задач, которые предстоит решать с помощью ОАСУ, весьма широк. В первую очередь к этим задачам относятся:

- разработка прогнозов и программ развития отрасли, а также долгосрочное планирование производства по всему комплексу предприятий отрасли;

- среднесрочное и краткосрочное технико-экономическое планирование;

- оперативное управление отраслью, задачи учета, контроля и анализа выполнения планов и результатов производственно-хозяйственной деятельности по предприятиям и организациям отрасли и по отрасли в целом;

- планирование и управление снабжением предприятий отрасли и сбытом готовой продукции;

- планирование и управление капитальным строительством и реконструкцией действующих предприятий;

- планирование и управление исследованиями и разработками (НИР и ОКР), а также всем комплексом мероприятий по научно-техническому прогрессу в отрасли и т. п.

Каждая из таких задач обычно решается силами целых организационных коллективов со сложной иерархической организационной структурой. Поэтому в автоматизированных системах управления такого класса чрезвычайно важную роль играют вопросы

выбора эффективных организационных структур, распределения внутри них и между ними прав и обязанностей, координации и руководства их совместной деятельностью.

При решении проблем управления организационными системами в связи с применением ЭВМ неизбежно приходится рассматривать прежде всего такие вопросы, как совершенствование структуры управления, методы подготовки и принятия решений и соответственно формирования целей и критериев, используемых в организации.

Сложность подобных вопросов усугубляется необходимостью учитывать играющие очень большую роль в организационных системах социальные и психологические факторы. Содержательное представление о функционировании той или иной организационной системы естественно складывается в процессе описания, анализа и моделирования работы этой организации. Умение моделировать работу организационных систем и оценивать их эффективность становится особенно важным, когда мы приступаем к задаче создания новых организационных систем или совершенствования уже сложившихся структур, с чем приходится сталкиваться при создании автоматизированных систем управления предприятиями (АСУП) и отраслями (ОАСУ).

Надо сказать, что некоторый, хотя и небольшой опыт и соответствующие теоретические разработки применительно к определенным аспектам функционирования существующих организационных систем в настоящее время уже имеются ¹⁾. К сожалению, накопленные знания пока еще не позволяют сформировать строгое научное представление об основных закономерностях поведения сложных многоуровневых организационных систем, основных факторах, определяющих это поведение, и выразить их с помощью моделей, сформулированных на строгом математическом языке.

Теория управления подобными сложными многоуровневыми, иерархическими организационными системами еще очень молода (первые публикации в этой области относятся в основном к началу шестидесятых годов) и в настоящее время интенсивно развивается усилиями как наших, так и зарубежных ученых.

Иерархическим системам управления соответствуют многоцелевые и многоуровневые системы принятия решений. В настоящее время хорошо развита — формализована, широко применяется на практике и оснащена мощными математическими методами, — по сути, только теория одноцелевых и одноуровневых решений (все разновидности математического программирования, оптимизационных экономико-математических методов и пр.). В целом слабее развита и значительно меньше применяется теория многоцелевых

¹⁾ См., например, работы [8, 11, 31, 34, 38] в списке дополнительной литературы,

п одноуровневых решений (теория игр). И, как уже указывалось, лишь совсем недавно стала развиваться теория многоцелевых, многоуровневых решений и соответственно теория иерархических систем управления. Имеются и весьма интересные результаты по распространению того же игрового подхода на иерархические системы¹⁾.

Книга М. Д. Месаровича, Д. Мако и И. Такахака является, видимо, первой книгой, в которой более или менее систематически исследуются математические модели иерархических структур управления и анализируются преимущества, которые может дать применение иерархического подхода в различных случаях. Основная ее цель состоит в том, чтобы показать возможности и вскрыть особенности иерархического построения систем управления различными процессами, понимаемыми в широком смысле этого слова (к таким процессам относятся производственно-технологические, экономические процессы, процессы управления множеством движущихся объектов и т. п.).

В предлагаемых авторами математических моделях иерархических систем управления широко используется формальный язык общей теории систем, основанный в свою очередь на теоретико-множественных концепциях. Введенная формализация дает возможность достигнуть (для рассматриваемого класса задач) необходимой точности описания, применять математические методы оптимизации и проводить структурные исследования. Центральное место в книге занимает проблема координации действий (принятия решений) в двухуровневой системе, содержащей n подсистем нижнего уровня, ответственных за управление n подпроцессами некоторого общего процесса и подчиненных единственной высшей системе управления. Такие двухуровневые системы могут использоваться как основные элементы (модули) при синтезе более общих многоуровневых систем. Подробно исследуются стратегии координации при различной степени децентрализации управляющих систем.

Авторы особо останавливаются на характеристиках, важных для систем управления как технологическими процессами, так и процессами в организационных системах, и принципиальных с точки зрения иерархической соподчиненности, в каком бы виде она ни проявлялась. Такими характеристиками являются: вертикальная декомпозиция, сложность принятия решений на разных уровнях, приоритет действий и право вмешательства верхних уровней по отношению к нижним.

Книга состоит из двух частей, разделенных на 8 глав. В первой части дается описание различных иерархических систем, вводят-

¹⁾ См., например, работы [2, 3, 9, 10, 36] в списке дополнительной литературы.

ся основные понятия и определения, дается формальная постановка задачи координации. Вторая часть посвящена математической теории координации двухуровневых систем.

В гл. 1 на трех конкретных примерах (сталелитейная промышленность, нефтехимическое производство, энергетические системы) излагаются основные проблемы многоуровневых иерархических систем. Рассматриваются некоторые причины образования иерархических структур в организационных и экономических системах.

В следующих двух главах излагается содержательный аспект теории. В гл. 2 «Концептуализация» вводятся структурные понятия, которые в дальнейшем становятся объектом математических исследований. На наш взгляд, авторы весьма удачно используют следующие аспекты рассмотрения иерархических структур: уровни описания, уровни последовательных стадий выработки решений и организационные уровни. Для их различения вводятся соответственно термины: страты, слои и эшелоны. Примером стратифицированного описания системы может служить ЭВМ, которая изучается по меньшей мере на двух стратах: электронной и вычислительной. На первой страте система в этом случае описывается с точки зрения законов электричества и основ электроники. На второй страте мы имеем дело с вычислением и программированием, система описывается на языке математических операций и символов. В этой же главе раскрывается взаимосвязь этих трех категорий иерархии: стратифицированности, многослойности и многоэшелонности.

Цель гл. 3 «Формализация» — представить в рамках математической теории систем различные концепции иерархии, введенные в предыдущей главе. Понятия систем, подсистем и их взаимосвязи формулируются на языке теории множеств. Обосновывается важность формализации многоуровневых систем, дающая возможность достигать необходимой точности описания, применять математические методы и проводить структурные исследования.

В заключение части I в гл. 4 авторы переходят к формализации центральной проблемы развиваемой ими теории — проблемы координации элементов иерархической структуры (на примере двухуровневых систем). Вводимая формализация позволяет далее использовать математические средства анализа. С этой целью дается общесистемное описание двухуровневой системы, имеющей n органов управления (блоков принятия решений) нижнего уровня, подчиненных единственному органу управления (блоку принятия решений) верхнего уровня. Вся система управления или принятия решений состоит, таким образом, из элементов, принимающих решения на своем уровне, и средств осуществления принимаемых ими решений. Двухуровневая система, разбираемая авторами, имеет три вида целей: 1) цели органов нижнего уровня; 2) цели вышестоящего органа (координатора) и 3) цель всей

системы. При этом глобальная цель системы и цели координирующего органа (координатора) формулируются, как правило, по-разному.

Важную в методологическом аспекте роль играет вводимый авторами «постулат совместимости» (согласованности) целей, на достижение которых направлена деятельность органов управления вышестоящего и нижестоящего уровней. Выполнение этого постулата эквивалентно, как у нас принято говорить, правильному выбору целей и постановке задач перед всеми органами управления, входящими в систему. Он гарантирует также возможность разумного сочетания централизованного и децентрализованного управления большой системой. В этом случае продвижение к глобальной цели, стоящей перед всей системой, может быть осуществлено за счет соответствующей координации деятельности подсистем, в значительной степени «автономных» с точки зрения выбираемых ими способов действий. При совместимости целей в двухуровневой системе глобальная цель системы и цели координатора не противоречат друг другу и последний формально координирует решения, принимаемые нижестоящими органами управления, не по отношению к глобальной цели, а по отношению к своим целям, что тем не менее не мешает продвижению к глобальной цели, стоящей перед всей системой.

Вторая часть книги (гл. 5—8) посвящена изложению собственно математической теории координации. Основное внимание уделяется трем возможным «принципам координации»: 1) прогнозированию взаимодействий; 2) оценке взаимодействий и 3) «согласованию» взаимодействий.

В гл. 5 исследуется двухуровневая система оптимизации при специальных предположениях. Цель этой главы — привлечь внимание к структурным проблемам и отыскать основные свойства двухуровневой системы, которые не зависят от конкретных технических проблем и более подробных ее описаний. Задачи оптимизации в этой главе ставятся, в частности, без учета ограничений на управления.

Рассматриваемые авторами проблема «разрешения конфликтов» в двухуровневой системе и проблема модификаций целевых функций для элементов нижестоящего уровня, которые допускали бы координацию ранее некоординируемой системы, представляется весьма важной. Авторам удастся в значительной степени обобщить известные ранее способы управления, основанные, например, на «механизме цен» и т. д.

В гл. 6 авторы исследуют задачи координации двухуровневых систем, но с дополнительными предположениями, что объекты системы описываются как подмножества нормированных линейных пространств. Предполагается, что эти подмножества обладают такими свойствами, как выпуклость, компактность, полунепрерыв-

ность и т. п., поэтому результаты этой главы являются более специфическими и конкретными, чем полученные в предыдущей главе. Наконец, для решения задачи координации приводятся некоторые итеративные методы. Класс рассматриваемых здесь систем содержит класс динамических систем со многими переменными, описываемых обыкновенными дифференциальными уравнениями.

В гл. 7 по-прежнему рассматривается двухуровневая система, но учитываются ограничения, налагаемые на управляющие воздействия и на решения, принимаемые элементами нижнего уровня. Изучается влияние этих ограничений на применимость принципов координации. При этом исследуется уже частный случай управления системой, для которой решаемые задачи формулируются в виде задач математического программирования. Анализируется применимость принципов координации при различных предположениях о характере этих задач (задачи выпуклого или линейного программирования, решаемые с помощью процедуры Данцига — Вульфа, и др.).

Глава 8 посвящена двум возможным способам улучшения характеристик работы системы в целом. Поднимаемые автором вопросы в основном сводятся к следующему. 1) Как координатор должен влиять на принятие решений элементами нижележащего уровня с тем, чтобы те, получив от него координирующее указание, выбрали свои локальные управляющие воздействия на ход процесса уже без дальнейшего вмешательства координатора и притом таким образом, чтобы их выбор способствовал улучшению общей характеристики системы? 2) Если задан интервал наблюдения за ходом процесса в двухуровневой системе и существует m моментов времени, в которые координатор может влиять на принятие решений на нижнем уровне, то какова в этом случае должна быть стратегия координатора, чтобы его вмешательство (в каждый разрешенный момент координации) приводило к улучшению характеристики работы всей системы? В этой главе излагаются возможные подходы к решению этих вопросов и показывается применимость ранее развитого математического аппарата к решению таких задач в случае линейных систем.

Как одна из первых попыток строгого формального описания сложных иерархических систем книга, конечно, не свободна от определенных недостатков. В частности, поднимаемые авторами проблемы формулируются на столь высоком уровне общности, что получить для них конструктивное решение пока можно лишь для простейших «линейных» систем. Сама по себе рассматриваемая авторами теория еще далека от завершения, и о возможности непосредственного практического применения достигнутых результатов говорить пока еще рано.

К числу недостатков, на наш взгляд, можно отнести также местами недостаточную «стратификацию» материала самой книги,

затемняющую временами концептуальную направленность рассматриваемых вопросов.

Следует также отметить, что недостаточно четкая формулировка ряда рассматриваемых вопросов и положений, определенная композиционная рыхлость материала книги, а также обилие вновь вводимых терминов и весьма вольная их трактовка повлекли за собой значительные трудности при переводе книги на русский язык. В необходимых случаях пришлось снабдить текст примечаниями, раскрывающими смысл вводимых авторами терминов.

Тем не менее в целом книга М. Месаровича, Д. Мако и И. Такаха «Теория многоуровневых иерархических систем» явится полезным пособием по теоретическим принципам построения управляющих систем с иерархическими структурами. Книга, несомненно, привлечет внимание широкого круга специалистов, работающих в области проблем управления — как в сфере прикладных исследований в связи с задачами конкретного проектирования АСУ, так и в области создания и исследования различных математических моделей систем управления.

Чл.-корр. АН СССР Г. С. Поспелов

Перевод книги выполнен К. К. Фроловым (гл. 1 и 2), О. В. Редькиной (гл. 4, 5 и приложение), Л. В. Ивановой (гл. 6), О. И. Демидовичем, И. Г. Поспеловым и С. А. Федоровым (предисловие авторов, гл. 3, 7 и 8).

ПРЕДИСЛОВИЕ АВТОРА

В данной книге развивается теория больших систем, применимая в случаях, когда изучаемая система может рассматриваться как совокупность иерархически организованных подсистем. Не всякая сложная система обладает подобной структурой; но если она принадлежит к этому классу, ее, несомненно, следует считать сложной.

Тот факт, что система является большой или сложной, еще не означает, что для ее изучения необходима теория больших систем. Прежде всего само понятие сложности системы зависит от точки зрения. То, что психологу представляется сложной или большой системой, может в глазах экономиста оказаться всего лишь элементарным звеном. Иногда мы случайно или по необходимости описываем несомненно сложную систему (скажем, экономику штата) относительно простой моделью (например, дифференциальными уравнениями второго порядка). Изучение такой модели позволяет делать выводы о реальном объекте. Хотя в данном случае система, бесспорно, большая, используемая теория несложна. Мы придерживаемся той точки зрения, что теория, претендующая на изучение сложных систем, должна оперировать моделями, структура которых отражает эту сложность. Модели, в явном виде отражающие тот факт, что система состоит из многих подсистем, взаимодействие которых (в процессе выработки решений) осуществляется иерархическим образом, без сомнения, достаточно сложны. Именно поэтому мы утверждаем, что изложенная в этой книге теория является теорией больших, или сложных систем.

Книга разделена на две части. В первой обсуждаются примеры иерархических систем, взятые из различных областей. Показывается, что при рассмотрении таких систем наиболее важную роль играют три основных аспекта описания их деятельности: принятый уровень абстракции (степень агрегированности моде-

ли); уровень сложности принятия решений и уровни приоритета действий в системах, состоящих из многих подсистем (блоков). Эти три аспекта излагаются затем на формальном языке с использованием терминов, принятых в общей теории систем. Завершает первую часть формальная постановка задачи координации, которая возникает именно при изучении систем с иерархической структурой и имеет для таких систем первостепенную важность.

Вторая часть посвящена изложению математической теории координации. В ней рассматривается двухуровневая система управления с n блоками принятия решений (или «решающими» блоками) на нижнем уровне и одним аналогичным блоком на верхнем¹). Координация формулируется как проблема принятия соответствующих решений блоком верхнего уровня, который должен так воздействовать на блоки нижнего уровня, чтобы была достигнута цель, поставленная перед всей системой. Стратегии координации вырабатываются на основании правил, которые мы называем *принципами координации*. Более подробная теория координации разрабатывается затем для систем, описываемых как самым общим образом, так и с помощью конкретных математических моделей.

Развиваемая математическая теория в основном касается оптимизирующих систем. Это сделано только из соображений удобства. Для того чтобы подчеркнуть главное — связь блоков, находящихся на разных уровнях, — мы сочли удобным исходить из предположения, что проблемы принятия решений для подсистем относительно просты и хорошо определены. В противном случае мы могли бы увязнуть в сложностях индивидуального поведе-

¹ Здесь и далее при использовании термина «решающий блок» или «решающий элемент» (в оригинале «decision unit» — дословно «блок принятия решений») имеется в виду некий орган управления, обладающий, если это необходимо, своей организационной структурой и представляющий собой определенным образом функционирующую человеко-машинную систему. Каждый такой управляющий орган при выполнении порученных ему задач обладает определенной свободой действий и имеет право принимать те или иные решения относительно того, каким способом эти задачи будут выполнены. В этом плане можно считать, что с формальной точки зрения деятельность подобных органов управления направлена на поиск иногда «оптимальных», а чаще — «приемлемых» или «удовлетворительных» решений. Критерии для установления «оптимальности» или «приемлемости» возможных решений при этом считаются фиксированными, известными этим органам, и задаются вышестоящим органом управления. Чтобы отыскать такие решения, нижестоящий орган управления, согласно развиваемой авторами концепции, должен «решить» некоторую «оптимизационную» задачу. Именно в этом смысле и употребляются в данной книге термин «решающий блок» или его синоним — «решающий элемент». — *Прим. ред.*

ния отдельных блоков и потеряли бы возможность изучать вопросы, относящиеся к иерархии в целом. Однако наши основные результаты — координационные принципы — применимы и в более близких к реальности ситуациях, когда принятие решений затруднено наличием неопределенности. Об этом подробнее говорится в гл. 8, а также в гл. 4 при рассмотрении структур «с обратной связью на верхний уровень».

Математическая сторона теории излагается в строгой форме путем формулирования теорем и последующего их доказательства. Такая форма изложения вызвана в первую очередь необходимостью строгого установления справедливости высказываемых утверждений. Как это ни парадоксально, потребность в такой форме изложения больше всего ощущается в абстрактных теориях, где доказательства кратки и по большей части тривиальны, но где ввиду отсутствия подробно разработанной структуры трудно полагаться на интуицию, так что приходится продвигаться вперед осторожно, небольшими шагами. Формулирование теорем сопровождается обсуждением их значения, а иногда и примерами, благодаря чему смысл высказываемых утверждений становится ясным и без ознакомления с доказательствами. При этом строгая схема изложения только помогает чтению, поскольку формальные выкладки, приводящие к доказательству теорем, выделены из основного текста. Поэтому читатель может (скажем, при первом чтении) ограничиться тем, чтобы уяснить себе смысл результатов, не прослеживая дедуктивных построений, но может также специально обратиться к подробному изучению каждого доказательства.

При выборе способа изложения мы преследовали также другую цель, не имеющую прямого отношения к теме данной книги. Мы попытались показать полезность и продемонстрировать способы применения общей теории систем. Эта теория была выдвинута несколько лет назад [4], и ее польза для системотехники вообще и для теории больших систем в частности серьезно оспаривалась. Было установлено, что предлагаемый этой теорией язык для описания очень сложных и плохо определенных систем позволяет ввести ряд строго сформулированных положений относительно свойств таких систем, в то время как раньше, до создания такой теории, подобные системы вообще могли быть описаны лишь на вербальном уровне. Эти положения служат отправной точкой для

описания структуры систем, а затем для развития необходимой математической теории. В этой книге мы постарались впервые показать такой метод в действии.

Наконец, следует указать, что две части книги почти независимы. Для того, кто интересуется лишь иерархическими системами и основами их математической теории, будет вполне достаточно ознакомиться с первой ее частью. Читатель же, для которого представляет интерес только координация, может сразу перейти ко второй. Ему встретятся лишь редкие ссылки на определения и понятия, введенные в первой части (в гл. 3 и 4). Более того, отдельные главы второй части могут рассматриваться как вполне законченные и самостоятельные исследования. Так, например, гл. 8, посвященную «линейной» координации, можно читать независимо от глав 5—7, обращаясь лишь к гл. 4, где вводятся используемые в гл. 8 определения. Аналогично обсуждение задач управления динамическими системами в гл. 6 или 7 можно читать, не изучая предварительно глав 5 и 8.

Часть I

ИЕРАРХИЧЕСКИЕ СИСТЕМЫ

Глава 1

ПРИМЕРЫ

Многоуровневые иерархические системы до сих пор еще не являлись объектом математических теорий или хотя бы подробного теоретического изучения. Это, однако, не означает, что ученые и философы не интересовались «иерархическим порядком» или не признавали его важности; предмет этот достаточно много обсуждался. Тем не менее можно утверждать, что имевшие место обсуждения этих проблем не были направлены на создание теоретических основ, а тем более математического формализма для исследования иерархических явлений и не привели к их созданию. В результате среди исследователей-теоретиков до сих пор отсутствует единое мнение относительно сущности иерархических систем. Можно даже сказать, что важность и широкая распространенность многоуровневых систем еще не достаточно хорошо осознаны.

В данной главе мы постараемся в самой сжатой форме описать некоторые из существующих многоуровневых систем, применительно к которым была разработана излагаемая в этой книге теория.

Основное назначение приводимых ниже примеров — возбудить интерес читателя, показав, насколько широко в действительности распространены многоуровневые системы. Точнее говоря, мы приводим многочисленные примеры из различных областей, отобранные с таким расчетом, чтобы облегчить понимание вводимых далее понятий и используемого на протяжении всей книги математического формализма.

Читатель, не интересующийся описанием существующих систем, но чувствующий необходимость создания теории многоуровневых систем и проявляющий интерес к теории как таковой, может начать чтение непосредственно с гл. 2 или 3.

Для читателя же, интересующегося только проблемой координации, достаточно будет ознакомиться с ч. II, обращаясь, в случае необходимости, к гл. 4 части I.

На этом мы закончим наши вступительные замечания и приступим к качественному рассмотрению различных иерархических систем.

1. ИЕРАРХИЧЕСКИЕ СИСТЕМЫ В КРУПНЫХ АВТОМАТИЗИРОВАННЫХ ПРОМЫШЛЕННЫХ КОМПЛЕКСАХ

Ниже приводятся три примера, взятые из сталелитейной, нефтехимической и энергетической отраслей промышленности. Чтобы избежать трудностей, связанных со специфическими особенностями той или иной отрасли, мы, как правило, будем описывать гипотетические случаи, синтезированные на основании описанных в технической литературе реальных данных.

Сталелитейная промышленность

Уже довольно давно стало очевидно, что для повышения экономичности производства стали необходима более полная автоматизация производственных процессов. Перечислим причины, ускорившие внедрение интегрированных систем управления сталелитейными комплексами на базе ЭВМ.

1) *Размер, сложность, разнообразие.* На большом сталелитейном заводе «может одновременно храниться, готовиться к обработке и перерабатываться 500 000 и более тонн материалов, над которыми производятся сотни операций» [1]. Обработка информации, контроль и управление, осуществляемые в настоящее время руководством предприятий, административными работниками и операторами, требуют поистине колоссального труда. Поэтому использование специального оборудования (например, ЭВМ), облегчающего этот труд и повышающего его производительность за счет более эффективной обработки информации и рационализации процесса принятия решений, представляется весьма привлекательным.

2) *Широкий спектр времени отклика в системе.* Сталелитейный завод, рассматриваемый как динамическая система, подвергается внешним воздействиям с очень широким спектром частот; работа всего завода обычно строго регламентируется недельным планом, составляемым на несколько недель вперед. В то же время прокатный стан работает со скоростью до 1200 м/с, так что «на учете каждая секунда» [1]. Однако в такой системе неизбежно возникают расхождения между заранее составленным планом и его фактическим выполнением. Легко видеть, что отсутствие координации может в такой ситуации привести либо к частичным срывам выполнения заказов (из-за недопустимых отклонений параметров от заданных значений), либо к чрезмерному увеличению складских запасов.

3) *Возрастающие требования к рентабельности.* Фактор издержек производства стали приобретает все большее значение как в связи с увеличивающейся конкуренцией, так и в связи с возрастающими потребностями в продукции высокого качества с минимальными допусками. Доходы сталелитейной промышленности, а так-

же отношение количества готовой стали к количеству исходного сырья неуклонно уменьшаются, как хорошо видно из фиг. 1.1 [2]. Именно для решения этих проблем обычно и используется автоматизация.

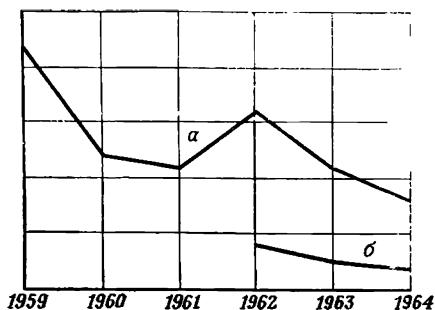
Автоматизация в сталелитейной промышленности основана на принципе объединения всех функций обработки информации и управления в единой системе, охватывающей все этапы от получения заказов до управления скоростью подачи и температурой. В результате возникает неразрывная связь между планированием выпуска продукции и непосредственным управлением производственными операциями в цехах — весьма желательная ситуация как для технологов, так и для руководства.

Обобщенная блок-схема функциональных задач, решаемых в процессе управления сталелитейным заводом, представлена на фиг. 1.2 в виде многоуровневой системы с иерархией организационного типа. Полная задача управления заводом определяется с помощью трех страт, так что вся система представляет собой стратифицированную систему (в соответствии с определением, данным этому термину в гл. 2). Система состоит из большого числа блоков и должна выполнять много задач (на одном подобном предприятии предполагалось использовать 9 ЭВМ [3]). Мы здесь, однако, опишем только ее основные функции.

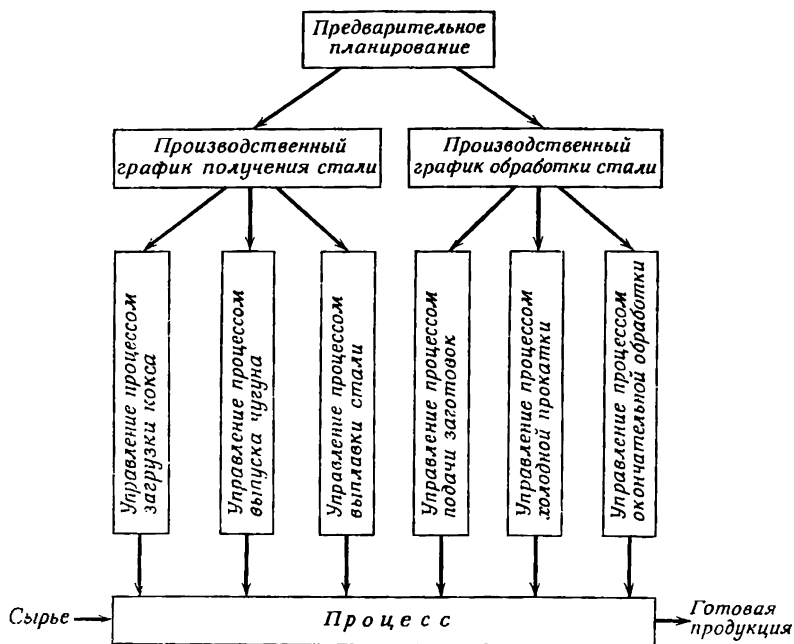
С общесистемной точки зрения рассматриваемая система должна выполнять следующие три основные функции: а) планирование производства, б) составление рабочих заданий и координация работ и в) управление технологическими процессами. Эти функции составляют основу для иерархической организации подсистем.

Блок управления высшего уровня принимает заказы, а затем группирует и распределяет их так, чтобы повысить рентабельность производства в пределах ограничений, накладываемых сроками поставок; на выходе этого уровня получается недельный план. Он составляется на несколько недель вперед и в последнюю минуту корректируется на основании информации, поступающей по каналам обратной связи, о фактическом выполнении производственных планов за истекший период.

Недельный план поступает на вход блоков управления более низкого (среднего) уровня, которые разбивают его на частные



Ф и г. 1.1. а — доход сталелитейной промышленности; б — количество обработанной и отгруженной стали на тонну литья.



Ф и г. 1.2. Блок-схема интегрированной системы управления сталелитейным заводом.

задания по отдельным технологическим процессам. Они сравнивают фактические показатели с плановыми; они получают данные об объеме производства и качестве продукции и могут потребовать полного изменения графика работы всего завода, если это необходимо. Их основная функция — *координирование*. Так как производство непрерывно (а в некоторых подсистемах скорости его весьма велики!), то работа отдельных подсистем должна непрерывно координироваться из единого центра, для того чтобы избежать возникновения узких мест, которые замедляют темпы производства или вызывают непроизводительные расходы. Рассмотрение именно этих процессов, управление которыми требует хорошо развитых методов координации, и привело к созданию концепций и теории, изложенных в гл. 2 этой книги.

Блоки управления нижнего уровня управляют самими технологическими процессами; они осуществляют функции контроля и управления физическим процессом производства продукции отдельными подэлементами, входящими в комплекс. На этом уровне производится оптимизация некоторых подпроцессов (с точки зрения минимизации стоимости продукции); осуществляется текущий контроль за ходом физических процессов, про-

изводится прямое цифровое управление процессами и т. д. К этому уровню также относятся входные и выходные устройства, измерительные приборы и средства индикации.

К этому краткому описанию следует добавить ряд замечаний.

1) Каждый из уровней, показанных на фиг. 1.2, может содержать ряд подуровней; «трехуровневая» структура соответствует первому дополнительному расчленению, или «вертикальной декомпозиции», всей задачи.

2) Система в своих действиях обладает значительной автономией. Сообщалось, например, что подобная система может принимать решения по таким вопросам: когда следует заказать дополнительное количество стали в связи с возрастанием брака; не находится ли в переработке излишнее количество материала и если да, то можно ли использовать его для выполнения других заказов [3]. Весь завод действует как адаптивная система с автоматически меняющимся производственным графиком и может даже отвергнуть некоторые заявки, если их выполнение потребовало бы нежелательной перестройки всего хода производственного процесса.

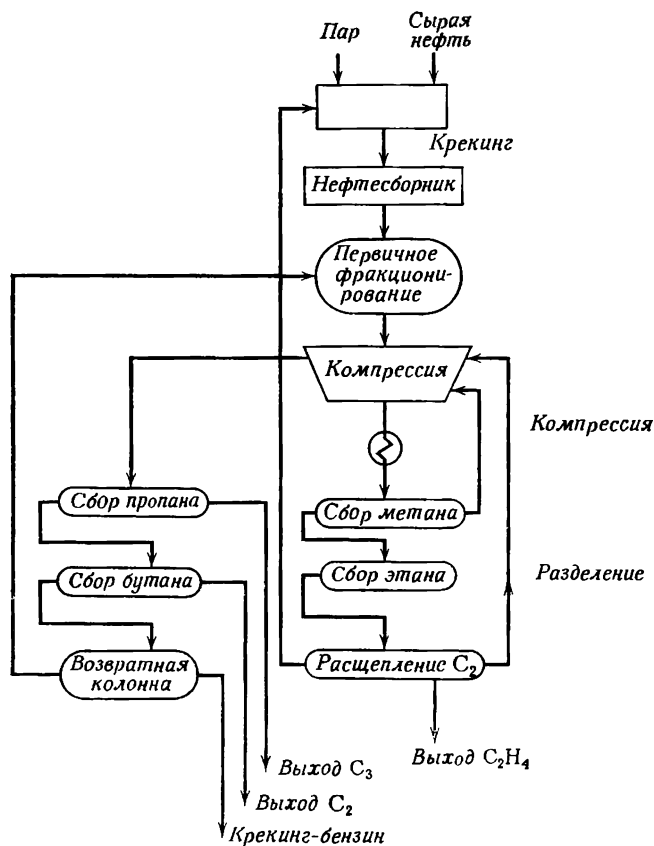
Нефтехимическое производство

Другим хорошим примером многоуровневой системы в больших автоматизированных промышленных комплексах является нефтеочистительный завод, управление которым осуществляется с помощью ЭВМ. Многие аргументы в пользу введения интегрированной системы управления нефтеочистительным заводом аналогичны приведенным выше для управления производством стали: производительность и непрерывность производства являются столь важными факторами, что даже небольшие улучшения в управлении производством приводят к значительной финансовой экономии. Более того, как показывает практика руководства подобными заводами, такие улучшения оказываются наиболее легко реализуемыми с учетом сложности технологической схемы и современных идей в области рационализации рабочих процессов, методов производства и управления. Среди аргументов в пользу иерархического подхода к проблемам управления объектами такого типа можно указать на следующие:

1) *Планирование и практическое управление производством.* Для повышения своей конкурентоспособности многие компании уже прибегают к помощи ЭВМ как инструмента для обработки информации и подготовки решений на уровне компании. В частности, задание по выпуску продукции для нефтеочистительного завода определялось путем проигрывания на ЭВМ различных вариантов, основанных на данных анализа рынка и модели производства, полученной методами линейного программирования. ЭВМ использовались на

заводах также для управления физическими процессами; например, осуществлялось прямое цифровое управление температурой, давлением и т. д. Но наряду с этим мы все еще сталкиваемся между этими двумя уровнями с деятельностью человека-оператора, осуществляющего вручную операции текущего контроля и управления ходом процесса, что, пожалуй, выглядит довольно архаично. Поэтому было бы только естественным попытаться устранить этот разрыв за счет использования современных методов принятия решений и обработки информации.

2) *Сильные взаимодействия и возросшая производительность.* Сильное взаимодействие может существовать как между отдельными звеньями завода в целом, так и между отдельными рабочими процессами. Как одно из следствий этого — возникновение узких мест, ограничивающих производительность завода. Путем улуч-



Ф и г. 1.3. Процесс крекинга нефти с получением этилена,

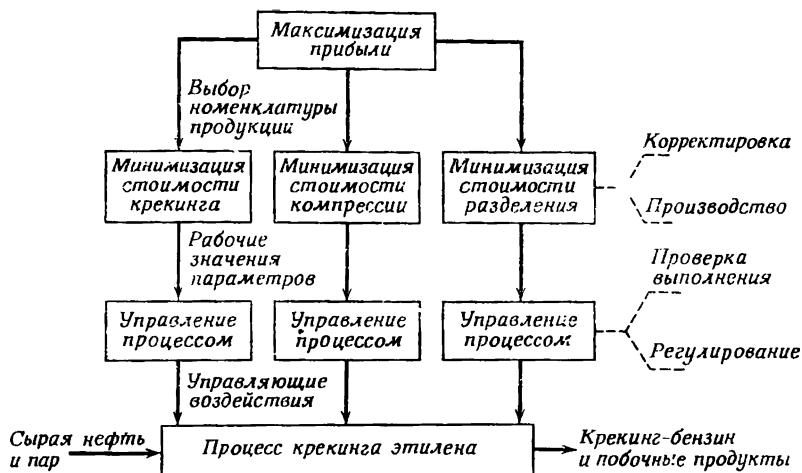
ления планирования можно повысить производительность всего комплекса за счет устранения узких мест на решающих участках, либо за счет как можно более равномерной загрузки всех технологических участков.

3) *Снижение издержек производства.* Параллельные операции, использующие общую подачу топлива и сырья, рециклы и т. д., требуют весьма тщательного составления графиков работы в целях минимизации производственных затрат как для отдельных процессов, так и для всего комплекса в целом.

4) *Приспособление к рынку сбыта.* Большое разнообразие выходных продуктов с различными ценами требует гибкого управления системой. Полная автоматизация позволяет относительно быстро реагировать на изменение рыночных условий.

Рассмотрим в качестве характерного примера производство этилена на нефтеперегонном заводе, имеющем интегрированную систему управления. Обобщенная блок-схема производства, принятая на заводе, показана на фиг. 1.3 [4]. Все производство можно считать состоящим из трех основных подпроцессов: крекинга, компрессии и разделения. Крекинг-бензин от параллельных колонн подается в общий первичный разделитель, сжимается и затем охлаждается для дальнейшего фракционирования, которое происходит в две стадии: низкотемпературная стадия для выделения этиленовых продуктов высокой чистоты и высокотемпературная стадия для выделения бензиновых фракций.

Схема всей системы, состоящей из двух частей — производственной и управляющей, — показана на фиг. 1.4. Очевидно, что это многоуровневая система. Ее построение базируется на двух иерархических понятиях, определенных в гл. 2. А именно, мы имеем здесь многоэшelonную систему организационного типа, так как выделенные элементы системы, ответственные за принятие решений, имеют иерархическое расположение, основанное на «подчиненности» нижестоящих элементов вышестоящим. Более того, это система многослойного типа, поскольку при вертикальной декомпозиции общей задачи произведено выделение нескольких слоев. Имеется три основных слоя, хотя каждый из них может быть представлен несколькими подслоями. Плановое задание по производству вырабатывается на самом верхнем уровне. Основным критерием при формировании планового задания является принцип максимизации прибыли. При этом в первую очередь учитываются внешние (рыночные) условия и значительно меньшее внимание уделяется деталям управления комплексом. На промежуточном слое производство продукции рассматривается с позиций минимальных затрат. В добавление к минимизации по локальной себестоимости на промежуточном слое в рассмотрение вводится ряд адаптивных функций, в частности так называемая корректировочная функция, используемая для корректировки значений



Ф и г. 1. Блок-схема интегрированной системы управления крекингом этилена.

(на основе уточненных исходных данных) коэффициентов, применяемых при оптимизации общей эффективности работы колонны, и позволяющая предсказывать выходные отношения, параметры, связанные с энтальпиями, удельными теплотами и т. д. На первый уровень возлагаются функции контроля и регулирования хода процесса. Управление ходом процесса в данном случае сводится к определению набора контрольных точек, на основе которых производится непосредственное регулирование процесса. В ходе управления процессом используется информация с высших уровней и производится расчет мгновенной скорости оттока и скорости выкипания в дистилляционных колоннах, принимая во внимание скорость подачи и энтальпию входных продуктов.

По поводу такой многослойной иерархии следует сделать несколько замечаний.

1) На первом слое применяется прямое цифровое управление классического типа, в то время как на высших слоях используются более сложные процедуры оптимизации.

2) Реальными физическими переменными управляет лишь первый подслой. Все остальные слои и подслои осуществляют подстройку обобщенных параметров и выбор рабочих точек, т. е. в конечном счете они связаны с перестройкой или адаптацией всей системы управления, в которой имеет место обмен информацией между слоями.

3) Все три слоя работают в реальном масштабе времени, но с точки зрения продолжительности периода принятия решений между ними существует значительное различие. Первый слой

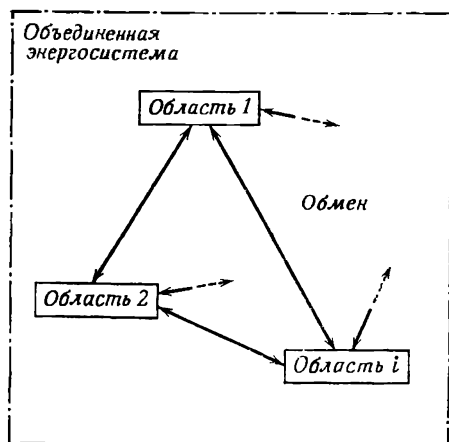
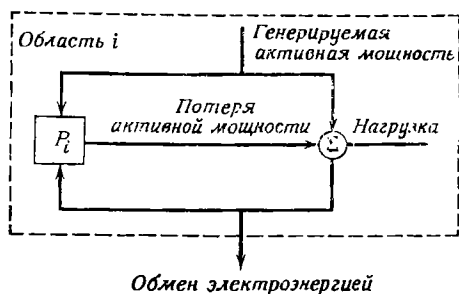
осуществляет практически непрерывное управление. Период принятия решений применительно к задачам оперативного управления для второго слоя составляет от 1 до 5 мин, в то время как для выработки долговременной программы имеется запас времени продолжительностью от 15 до 25 мин. Наконец, период принятия решений для высшего слоя (за исключением аварийных ситуаций) — 1 сутки.

Заслуживает упоминания и еще одна сходная система управления на базе ЭВМ — этиленовый завод с годовым производством этилена в 200 000 т, — относительно которой имеется много данных о характеристиках ее работы, хотя и нет информации о реальной схеме системы управления. Система управления получает информацию об этилене и десяти главных побочных продуктах вместе со статистикой продаж и на основе этих данных вычисляет почасовую прибыль. В случае больших изменений спроса она может за несколько часов довести систему до максимальной производительности на новом уровне производства. Она обеспечивает также (на почасовой основе) полную информацию относительно выхода продукта и других условий на заводе. Надежность и безопасность системы, как утверждается, весьма высоки; в частности, производство продукта в девяти колоннах может быть остановлено простым «нажатием кнопки», тогда как при отсутствии управления с помощью ЭВМ «оператору пришлось бы совершить для этого 15 ручных регулировок» [5].

Энергетические системы

Энергетические системы с их разнообразным оборудованием, огромным числом подсистем и сложными связями между ними — естественный объект для применения многоуровневых подходов. За последние годы энергетические системы значительно выросли размерам, что связано с процессом постепенного объединения существующих систем в более крупные комплексы; на каждом этапе интеграция двух подсистем давала очевидные экономические выгоды, не вызывая в то же время сколько-нибудь существенных технических или управленческих проблем. В результате появились огромные и чрезвычайно сложные энергетические комплексы, объединяющие тесно связанные между собой системы. Управление ими немыслимо без быстрого принятия решений, причем требования к надежности и безопасности функционирования всей системы резко возросли, ибо сбой в ее работе могут иметь серьезные экономические, социальные и политические последствия.

Все это требует нового подхода к структуре больших энергетических систем. В настоящее время все большее внимание привлекает так называемый многослойный подход [6]. Однако мы ограничимся здесь лишь обсуждением более традиционной про-



Ф и г. 1.5. Сеть взаимосвязанных энергетических систем.

блемы диспетчирования активной мощности в энергокомплексе, представляющем собой, пожалуй, наиболее классический пример действующей многоуровневой иерархической системы.

Система разбита на n взаимодействующих подсистем (фиг. 1.5). Обычно границы соответствующих областей выбираются таким образом, чтобы каждая из них представляла одну компанию. В каждой области имеются, конечно, ряд генераторных станций и разнообразных потребителей, но, поскольку нас интересует лишь обмен между областями, мы будем предполагать, что каждая область характеризуется лишь следующими параметрами:

c_i — (полная) нагрузка в области i ;

x_i — (активная) мощность, вырабатываемая элементами в области i ;

y_i — потери (активной) мощности в области i ;

u_i — обмен (точнее «чистый результат» обмена) мощностями через энерголинии, соединяющие область с другими областями.

При заданной нагрузке потери в области зависят как от мощности, вырабатываемой в самой области i , так и от обмена мощностями, т. е. y_i является функцией x_i и u_i ; $y_i = p_i(x_i, u_i)$. Уравнение баланса мощностей можно записать в форме

$$P_i(x_i, u_i) + c_i - u_i + x_i = 0. \quad (1.1)$$

Имеется, конечно, n уравнений такого вида. Кроме того, обмен мощностями между областями также должен быть сбалансирован, т. е.

$$u_1 + \dots + u_n = 0. \quad (1.2)$$

Уравнение (1.1) описывает подпроцессы, уравнение (1.2) — их взаимодействие.

Проблема диспетчирования (активной) мощности в объединенной энергосистеме теперь заключается в том, чтобы по всем линиям найти мощности x_i и объемы обменов u_i , при которых стоимость выработанной электроэнергии будет минимальной. С этой целью введем функцию стоимости $F_i(x_i)$ выработки в каждой из областей мощности x_i ; тогда общая стоимость выработки будет равна

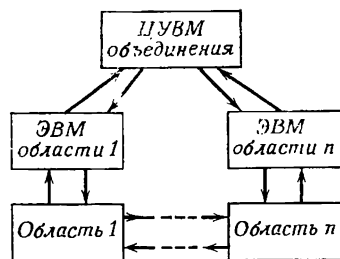
$$F(x_1, \dots, x_n) = F_1(x_1) + \dots + F_n(x_n). \quad (1.3)$$

Проблема оптимального диспетчирования заключается в том, чтобы минимизировать выражение (1.3) при условиях, что переменные $x = (x_1, \dots, x_n)$ и $u = (u_1, \dots, u_n)$ удовлетворяют уравнениям баланса (1.1) и (1.2). Заметим, что вследствие уравнения (1.2) имеется лишь $n - 1$ независимых величин u_i . Проблему можно решить либо посредством одной ЭВМ (на основе так называемой «полной централизации»), либо на основе двухуровневой иерархии ЭВМ.

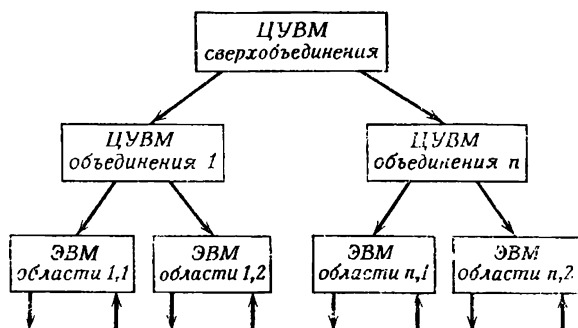
При двухуровневом иерархическом подходе каждая область имеет собственную ЭВМ для решения задачи диспетчирования и, кроме того, есть центральная управляющая вычислительная машина (ЦУВМ). Такая организация показана на фиг. 1.6. Проблема нахождения оптимального распределения производства и обмена мощностями решается с помощью двухуровневой системы, обладающей организационной иерархией.

Задача всей иерархической организации — решение проблемы минимизации. В связи с этим возникает вопрос, как эта задача должна быть распределена между ЭВМ различных областей и ЦУВМ. Этот вопрос весьма подробно изучается в гл. 2. Для довольно близкой проблемы в работе [7] был предложен ряд схем выбора структуры. Однако есть общий метод, основанный на так называемом принципе прогнозирования взаимодействий (см. гл. 4), который довольно естественным образом можно применить и для распределения задач между отдельными ЭВМ. Этот метод заключается в следующем.

ЭВМ для i -й области решает задачу минимизации $F_i(x_i)$ относительно x_i при выполнении условия (1.1). При этом величина u_i^α задается ЦУВМ. Задача ЦУВМ заключается в нахождении опорного уровня обмена, т. е. $u_1^\alpha, \dots, u_n^\alpha$. Если обмен происходит на α -уровне, локальный оптимум



Фиг. 1.6. Двухуровневое диспетчирование в объединенной энергосистеме.



Ф и г. 1.7. Образование сверхобъединения из «двухобластных» объединений при диспетчировании энергосистем.

будет одновременно и глобальным. Когда различия между опорным уровнем обмена (определенным ЦУВМ) и действительным обменом выходят за пределы некоторых предписанных ограничений, новый опорный уровень обмена определяется итерационным процессом (описанным в гл. 4), в котором участвуют ЦУВМ и ЭВМ различных областей.

Мы будем исследовать эту задачу с единственной целью — показать, как теория координации, развитая в гл. 4 и ч. II, может быть применена к задачам такого рода. Хотя мы указали лишь на применение так называемого принципа прогнозирования взаимодействий, принципы согласования и оценки взаимодействий также могут быть использованы. На практике выбор метода зависит от конкретных обстоятельств.

Многоуровневое управление (в данном случае двухуровневая иерархическая система) предпочтительнее централизованного подхода по многим техническим, экономическим и эксплуатационным причинам. Перечислим основные из них.

1) Диспетчирование может осуществляться быстрее (т. е. с затратами меньшего количества машинного времени) и с меньшими требованиями к объему памяти ЭВМ. Это, конечно, зависит как от размеров системы, так и от искусства постановки задачи ЦУВМ, которая действует в этом случае как координатор.

2) Система менее чувствительна к изменениям структуры взаимодействий, — изменениям, которые могут иметь регулярный или случайный характер. Пусть, например, изменения условий в одной области вызывают изменения только в одной из решаемых (локальных) задач. При централизованном, «однообластном», подходе эти изменения не могут быть локализованы и обычно приходится менять программу решения для всей системы (вся система рассматривается как одна область). Кроме того, нетрудно изменить состав общей системы.

3) В центрах областей уже могут иметься ЭВМ, решающие другие технические и эксплуатационные задачи. Тогда ЦУВМ просто добавляется для диспетчерской координации и для других работ на уровне всей системы (например, для подготовки заявок на обмен энергией и т. д.). Следует также напомнить, что одна область обычно относится одновременно к нескольким компаниям; поэтому довольно часто приходится сталкиваться со специфическими административными и юридическими проблемами, в силу которых двухуровневая система управления оказывается предпочтительнее.

Наконец, нужно отметить, что иерархическая структура может быть расширена, в результате чего создается сверхобъединение, как показано на фиг. 1.7.

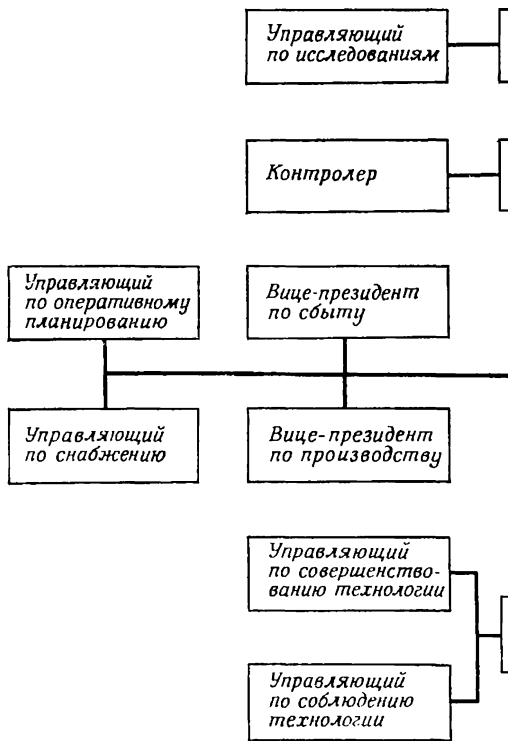
2. ИСПОЛЬЗОВАНИЕ КОНЦЕПЦИЙ МНОГОУРОВНЕВЫХ СИСТЕМ В ТЕОРИИ ОРГАНИЗАЦИЙ

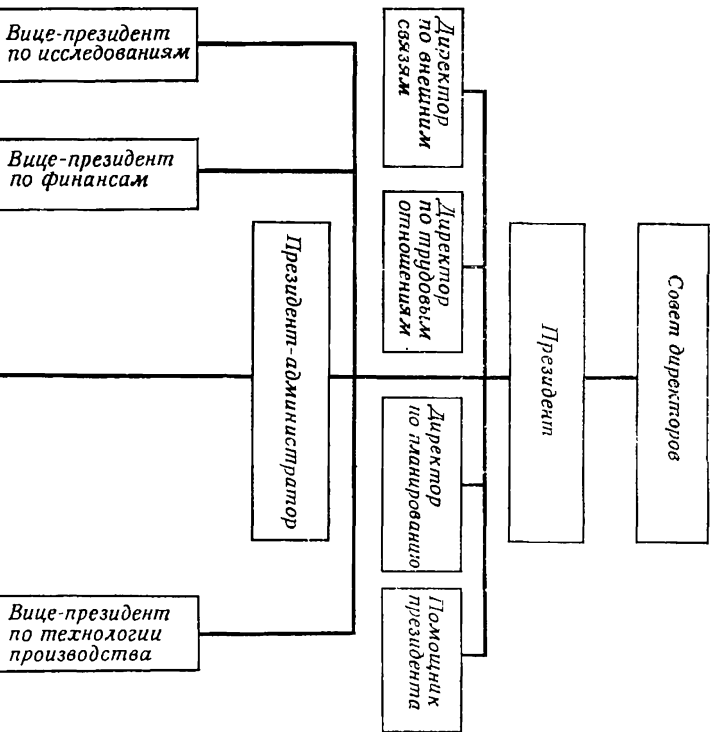
Элементы сходства между организациями людей и многоуровневыми иерархическими системами, рассматриваемыми в данной книге, становятся очевидными уже после быстрого сравнения блок-схемы многоуровневой системы, изображенной на фиг. 1.1, с организационной структурой, представленной на фиг. 1.8. Несмотря на то что имеется ряд довольно отличающихся друг от друга определений организации (зависящих от тех аспектов, на которые делается упор при рассмотрении организации), все они сходятся в том, что организация состоит из семейства взаимодействующих, иерархически расположенных элементов, наделенных правом принимать решения. Отсюда и следует применимость нашей модели.

Важность организации трудно переоценить. Как указал Эрроу, «среди всех творений человека использование организации для осуществления его целей — одно из самых великих и самых ранних его изобретений. Даже при отсутствии других свидетельств было бы ясно, что осуществление таких грандиозных строительных программ, как возведение четко распланированных городов вроде Нара и Киото или таких монументов, как египетские пирамиды, было бы невозможно без создания сложных организаций. В отношении менее материальных целей мы знаем, например, организацию империи инков в Перу, где сложное и раскинувшееся на большой территории государство управлялось весьма четко с использованием всего лишь примитивнейших технических средств: они не знали ни письменности, ни колеса» [8].

Важная роль организации в современном обществе стала очевидной в связи с повсеместным распространением и могуществом крупных организаций в нашу технологическую эру. Гэлбрейт утверждает, что современное государство правильнее всего было

Ф и г. 1.8. Организационная структура компании.





бы рассматривать как индустриальное государство в связи с мощством и преобладающим значением, которые в нем приобрели крупные индустриальные организации [9]. Одна из главных отличительных черт современных организаций — высокая степень их специализации. Касаясь тех, кто руководит деятельностью этих организаций, Гэлбрейт со свойственной ему категоричностью заявляет: «Чтобы управлять организацией, возникшей в результате специализации, нужен специалист в области организации». Совершенно ясно, что в настоящее время руководителю для успешного выполнения его работы необходима всесторонняя помощь. Управление во все большей степени становится коллективной функцией, а не прерогативой ограниченного круга лиц, как это имело место в прошлом.

В свете сказанного выше приходится лишь удивляться тому, как мало мы знаем об организациях и как медленно до сих пор развивалась теория организаций. Это не критика самой области или упрек в адрес работающих в ней ученых. Говоря это, мы хотим лишь обратить внимание на соответствующий круг проблем, подчеркнуть их важность и настоятельность. Без сомнения, незначительный прогресс в этой области обусловлен прежде всего сложностью самого предмета, но он связан также и с недостаточным вниманием к нему, что находит выражение в небольшом числе работающих в этой области исследователей. Можно утверждать, что в настоящее время число исследователей, озабоченных «оптимальным» подбором параметров в системах управления с обратной связью (задачей, которая в лучшем случае может дать лишь узкий коммерческий эффект), значительно выше числа ученых, интересующихся качественными аспектами процессов управления и связи в системах организационного типа.

Исследуя многоуровневые системы, мы стремились более непосредственным образом применить теорию систем к теории организаций. Организацию можно, конечно, представлять как систему, попросту рассматривая внешние воздействия (входы) и связанные с ними отклики (выходы). Если эту зависимость считать динамической (как, например, в так называемой промышленной динамике), то окажется возможным сделать полезные выводы об изменениях во времени, происходящих на протяжении всей истории организации. Однако прогностические возможности модели «вход — выход» (такой подход иногда называют также общесистемным) ограничены в связи со структурными различиями между моделью «вход — выход» и самой организацией. Это замечание сохраняет силу, даже если ввести взаимные обратные связи и представить организацию как одноуровневую, хотя и многопараметрическую, систему управления с обратной связью, как это обычно делается в промышленной динамике. Если бы основная цель теории организаций заключалась в том, чтобы объяснить рост организа-

ции и, может быть, предсказать ее эволюцию в ближайшем будущем, то при наличии ряда ограничений такая модель могла бы оказаться адекватной. Однако одна из первостепенных задач теории организаций — помощь руководителям — лицам, ответственным за преуспевание организации. Руководители — сами члены организации, и им, более чем кому бы то ни было, необходимы знания о том, как воздействовать на организацию изнутри, чтобы улучшить ее функционирование. Для этих целей модель организации, построенная по принципу «вход — выход», окажется непригодной.

Приведенные соображения указывают на необходимость многоуровневой структуры в модели организации, описываемой с позиций теории систем. Структура эта должна отображать самые важные характеристики организации, а именно: 1) что организация состоит из взаимосвязанных подсистем, имеющих право принимать решения; 2) что эти подсистемы образуют иерархию. Поэтому теоретико-системная модель организации — это не что иное, как многошелонная система, описываемая в гл. 3.

Наша цель — показать возможность применения теории многоуровневых систем к решению задач теории организаций. Мы постараемся перекинуть мост, хотя бы временный, между предлагаемым нами формализмом и проблемами, стоящими перед организациями, создав тем самым стимулы для дальнейших теоретических исследований в этой области. Развернутое и обоснованное приложение теоретических результатов к теории организаций — самостоятельная тема, требующая, конечно, значительно больше места и времени, чем мы располагаем в настоящей книге.

Теории организационных систем и многоуровневый системный подход

Ни одна из существующих теорий не может претендовать на то, что единственно она дает правильное описание работы организации. Скорее имеется целый спектр теорий, трактующих проблемы организаций. Да это и естественно. Такая сложная вещь, как организация, имеет множество аспектов, для исследования которых приходится привлекать знания из различных дисциплин. Более того, организации столь многообразны как по размерам, так и по назначению, что говорить просто об «организации» было бы грубым упрощением.

Чтобы определить то новое, что дает теория многоуровневых систем, необходимо сопоставить эту теорию с некоторыми наиболее важными течениями. Мы сделаем это довольно кратко, чтобы не превращать книгу в сводку имеющихся теорий. Наша цель — оценить место многоуровневого системного подхода среди этих теорий.

Теории организаций можно разделить на три категории: *классические* (структурные), *поведенческие* (мотивационные) и *системно-ориентированные*. Мы прежде всего сопоставим эти направления, а затем укажем тот вклад, который дает многоуровневый системный подход в решение следующих вопросов: 1) характерные особенности отдельно взятых членов организации — ее *участников*; 2) способ отображения структуры организации; 3) применимые в этой области инструменты и методы исследования.

Участники

Классическая теория имеет тенденцию рассматривать участника организации просто как «инструмент», выполняющий порученную задачу. При этом молчаливо предполагается, что участника побудили (кнутом или пряником) играть предписанную ему роль; его задача просто-напросто состоит в том, чтобы наиболее эффективно образом выполнить свое задание. Обычно путем фиксации продолжительности выполнения тех или иных функциональных заданий изучаются последовательности событий, из которых складываются действия, необходимые для рассматриваемого вида деятельности. Такой подход, очевидно, применим для исследования действий производственных бригад, обслуживающих автоматическую линию, но он непригоден для описательного или нормативного рассмотрения деятельности таких участников (управляющие и директора), которые находятся на высших уровнях организации.

Теория поведения (мотивации) проявляет особый интерес к участникам организации. Она изучает реакцию участника на различные побудительные стимулы и воздействия как изнутри данной формальной организации, так и со стороны других групп, к которым он одновременно принадлежит. Она в значительной степени рассматривает члена организации безотносительно к тому, какое место в организации он занимает; короче говоря, теория акцентирует свое внимание скорее на его личных качествах, нежели на его роли в организации.

В системно-ориентированных подходах, таких, как промышленная динамика, вообще говоря, отсутствует четкое выделение участников. Функционирование всей системы в целом описывается в динамических терминах (посредством подходящих уравнений или имитационных моделей). При этом ее эволюция исследуется без учета в явном виде влияния человеческих факторов.

Структура

Структура организации в классической теории считается статичной. Организационная структура, показанная на фиг. 1.8, отражает иерархию соподчиненности ее членов. При этом раз-

личным элементам, изображенным на схеме, предписаны конкретные цели. При определении целей и ролей руководствуются некоторыми принципами, сформулированными на основе «опыта хорошего управления». Предполагается, что они учитывают как способности, так и ограниченные возможности участников. Например, принцип «диапазона управления» («сферы полномочий») ограничивает число участников, которые могут контролироваться каждым членом; принцип «делегирования» (или «передачи полномочий») постулирует, что власть должна распределяться сообразно с ответственностью.

При подходе с точки зрения теории поведения акцент делается на участниках как членах неформальной группы (аналогия с «групповой динамикой»). Иерархическая структура рассматривается лишь косвенно, причем молчаливо предполагается, что она играет второстепенную роль.

В системно-ориентированных подходах иерархической структуре также отведено второстепенное место. Основной объект рассмотрения — общая эволюция системы во времени. Система представлена с помощью таких специфических элементов, как «временные задержки», «усиления» и т. д. Иначе говоря, система может иметь любую структуру, лишь бы она обеспечивала нужную зависимость между входом и выходом.

Методология

Классический подход основывается на наблюдениях за действительными процессами и на приобретенном опыте по выполнению определенных задач в реальных организациях. Язык и основные понятия в значительной мере заимствованы из классической механики.

При поведенческом подходе за основу берутся наблюдения за психологическими реакциями и действиями (поведением) участников. По существу многое заимствуется из групповой динамики. Разработанные модели более соответствуют неформальным, нежели формальным организациям. Применимость выводов теории к реальным наблюдениям над формальными организациями — желаемая, но не всегда достижимая цель.

Системные исследования и исследования с помощью ЭВМ в значительной степени сводятся к машинному моделированию, основанному на ряде наблюдений, имеющих отношение к реакции организации в целом. Чтобы можно было успешно изучать динамику, количество переменных должно быть ограничено сравнительно небольшим числом, в связи с чем они неизбежно будут отражать личные интересы и склонности исследователя.

Намеченная выше классификация не предполагает наличия четких границ между различными организационными теориями;

имеются подходы, сочетающие характеристики двух и более классов.

Наиболее плодотворным, на наш взгляд, является подход Саймона [10]. В нем участник организации рассматривается как лицо, имеющее право принимать решения (хотя это понятие никогда не было формализовано). В указанном подходе используются представления современных психологических теорий, касающихся решения задач человеком. Хотя его часто относят к поведенческим, подход этот содержит много системных элементов, как о том свидетельствуют модельные исследования Бонини [11]. Иерархическая структура, в какой бы форме она ни проявлялась в организации и как бы ни изображалась на блок-схемах, играет, как и в поведенческой теории, второстепенную роль. В качестве основных характеристик участников используются имеющиеся у них мотивы, например «побуждения производить» или «побуждения участвовать», вне прямой связи с местом, которое они занимают в организации.

Понятие иерархии встречается при обсуждении глобальной задачи организации. Однако это иерархия многослойного типа, не соответствующая реальной многоэтапной структуре организации. Несмотря на то что применение многослойных иерархий при проектировании или перестройке организаций имеет некоторую нормативную ценность, их использование в описательных исследованиях может привести к путанице, ибо структура реальной организации (имеющей заведомо многоэтапную иерархию) может оказаться такой, что многослойная декомпозиция глобальной задачи не будет иметь смысла. Совершенно очевидно поэтому, что необходимо иметь четкое представление о том, какое именно понятие иерархии имеется в виду.

Другой подход, объединяющий признаки двух направлений, — применение теории игр. С методической точки зрения он прежде всего относится к третьей категории. Однако акцент на процессе нахождения «точек компромисса», приемлемых в пределах рассматриваемой организации, приводит к недооценке иерархической структуры. Аналогичная ситуация имеет место и в теории группового поведения [12].

Место теории многоуровневых систем

Мы утверждаем, что теория многоуровневых систем имеет много общего со всеми тремя перечисленными выше направлениями:

1) она акцентирует внимание на иерархических структурах в смысле организационных блок-схем классической теории; иерархическое расположение элементов, принимающих решение, рассматривается как одна из наиболее важных характеристик организации;

2) она рассматривает участника как систему, принимающую решение (или выбирающую цель) в смысле современных поведенческих или, более определенно, мотивационных подходов. В ней явным образом учитываются уровни удовлетворения и расхождения между действительными и операционными, фактически наблюдаемыми целями;

3) она явным образом учитывает тот факт, что важнейшей особенностью организации неизменно является ее «организующая» роль в налаживании взаимной связи подсистем, принимающих решения.

Можно считать, что предлагаемый подход относится к отдельной категории исследований — *применению математической теории систем в исследовании организаций*. Ссылка на математику может показаться слишком сильным условием. Нужно, однако, заметить, что подход основан на общей математической теории систем и дает основу для формализации слабо структурированных утверждений, выводимых из действительных наблюдений. Такая общая теория может, используя поведенческие структуры, создавать их формальные математические аналоги, не накладывая при этом дополнительных ограничений. Можно даже утверждать, что при такой формализации ничего не теряется, а лишь приобретает повышенная точность в описании реальных явлений.

Потенциальные достоинства теории многоуровневых систем

Среди наиболее непосредственных потенциальных преимуществ, которые сулит применение теории многоуровневых систем к исследованию организации, можно указать на следующие:

1) она создает единую основу для различных подходов, вводя систему понятий и методов, посредством которых различные теории сравниваются, противопоставляются и взаимно дополняют друг друга;

2) она позволяет математически строго сформулировать как основные понятия, так и получаемые результаты;

3) она дает отправные точки для исследования различных аспектов и проблем анализа и проектирования организаций с помощью математических методов и моделирования на ЭВМ.

Сам подход, поскольку он основан на математических методах, прежде всего связан с такими структурными рассмотрениями, как коммуникация, управление, командование, координация и т. д. Нужно, однако, подчеркнуть, что основной его строительный кирпичик — элемент, вырабатывающий и принимающий решения (решающий блок), — стоит ближе к формализации типа добывающегося «удовлетворения» человека (по Саймону), чем к более ортодоксальному типу «человека-оптимизатора». Поэтому это скорее «административный», нежели «экономический» субъект [13].

Формализация в рамках теории многоуровневых систем основных понятий теории организаций

Для иллюстрации утверждений, высказанных в предыдущем разделе относительно потенциальных возможностей применения теории многоуровневых систем к изучению организаций, мы кратко обсудим, каким образом отражаются в предлагаемых формальных моделях основные понятия теории поведения. За основу мы при этом примем классический трактат Марча и Саймона об организациях [10].

Моделирование поведения участника организации

Понятие принимающего решение (или «решающего») элемента, вводимое в гл. 3, включает в себя, как частный случай, саймоновского «добывающегося удовлетворения» администратора. Оно в состоянии поэтому воплотить ряд понятий теории поведения. Например, функция «допустимости» предусматривает наличие «уровня стремлений», тогда как «принятая установка» соответствует множеству альтернатив, оставленных для рассмотрения к моменту принятия решения. Иными словами, принимающие решение элементы, которые являются строительными кирпичиками (блоками) многоуровневых систем, обладают тем, что можно назвать «свободой действий» принимающего решения лица: отклик элемента определяется проблемой удовлетворения, решение которой в общем случае не является однозначным. Соответственно и выбор фактически выполняемого действия зависит от имеющейся у элемента свободы действий. Формализация включает в себя все четыре типа «свободы действий», определенных Саймоном. Повидимому, теория оставляет достаточный простор для представления поведенческих актов участников при принятии решений с позиций современной психологии.

Руководитель и подчиненный

Рассмотрим отношения между руководителем и подчиненным, или, по нашей терминологии, между вышестоящим и нижестоящим элементами. Произведенный Саймоном анализ показывает, что руководитель может влиять на члена организации посредством: а) факторов, связанных с его целями; б) факторов, связанных с ожидаемыми последствиями его решений; в) факторов, связанных с набором альтернативных действий, имеющих в момент принятия решения. Согласно нашей формализации, это отвечает соответственно вмешательству на уровне целей, на уровне представлений (образов, моделей) и на уровне ограничений.

Координация

Саймон утверждает, что организация реагирует на появление конфликтных ситуаций применением либо «аналитических процессов», либо «процессов, связанных с нахождением «точек компромисса»», и что аналитический процесс включает «решение проблем» или «метод побуждения». Для всех этих понятий имеются аналогии в формальной теории координации, развиваемой в ч. II.

Саймон [9] полагает, что при разрешении конфликта по методу *решения проблем* задача сводится к правильной постановке целей для координатора и подчиненных и проблема принятия решения состоит в том, чтобы найти решение, удовлетворяющее установленным критериям. В нашем формализме это соответствует вмешательству на уровне представлений (образов, моделей): функции для оценки деятельности вышестоящих и нижестоящих элементов, если они являются согласованными, остаются неизменными и координация осуществляется путем воздействия на нижестоящие выходные функции или модели. Когда вмешательство на уровне представлений принимает форму прогнозирования взаимодействия, это соответствует введенному Саймоном понятию «сбора информации» (о будущих взаимодействиях). «Выявление новых альтернатив» здесь описывается как вмешательство на уровне ограничений; выходные функции изменяются путем наложения ограничений на множество действий.

Аналогично *метод побуждения* соответствует в нашем формализме вмешательству на уровне целей. Саймон утверждает, что условия для координации по «методу побуждения» возникают тогда, когда «индивидуальные цели могут отличаться внутри организации, однако нет оснований считать цели строго фиксированными». Более того, он утверждает, что «при применении метода побуждения неясным образом предполагается существование общих целей на некотором уровне и что разногласия по поводу подцелей можно устранить, опираясь на эти общие цели». Действительно, в теории координации (ч. II) договоренность относительно целей достигается на вышестоящем уровне и вмешательство в целевую функцию нижестоящего элемента осуществляется на основе общей цели. «Проверка подцелей на совместимость» соответствует выяснению того, обладают ли выбранные цели необходимыми свойствами (межуровневая или внутриорганизационная согласованность). «Различные новые критерии», введенные для координации, соответствуют различным функциям для оценки эффективности, отображающим наше вмешательство в цели.

Наша формализация достаточно широка и включает в себя формализацию идеи Саймона об аналитическом разрешении конфликта либо методом «решения проблем», либо «методом побуждения». Получаемое при этом уточнение основных понятий и аспектов — уже достаточное вознаграждение за формализацию. Отметим все

же мимоходом, что, поскольку формализация является математической, тем самым одновременно закладываются основы для дедукции и сравнения методов побуждения с методами решения проблем. Более того, применяя результаты ч. II, можно создать новые алгоритмы, определяющие, как же в действительности нужно осуществлять процесс координации. Формализация, таким образом, делает определенный шаг в направлении нормативной теории организаций. Любое конкретное нормативное утверждение основывается при этом на более детализованных, специфических структурных особенностях системы и является поэтому менее общим, нежели словесные утверждения, которые в силу объективных причин часто бывают очень туманными. Однако к логически обоснованным нормативным утверждениям можно прийти, лишь прибегая к более конкретным и строгим формулировкам исходных положений.

Можно предположить, что по той же основной схеме может быть разработана и теория координации, базирующаяся на подходе, связанном с нахождением точек компромисса. Этого пока еще никто не пытался сделать, хотя подобные рассуждения представляют важное и интересное направление исследований. Естественно было бы ожидать, что в подходе, связанном с нахождением компромиссных решений, значительную пользу может оказать применение теории игр. Действительно, теория игр уже применялась для изучения ситуаций типа нахождения компромиссных решений, удовлетворительных с точки зрения организаций в экономической и социальной сферах. Следует, однако, отметить, что мы говорим здесь о нахождении точек компромисса как о методе координирования в пределах организации. Это подразумевает, что особая роль отводится «командному» (вышестоящему) элементу, который, возможно, наделяется даже правом изменять правила с тем, чтобы побудить участников прийти к желаемому урегулированию спорных вопросов. Изучение такого подхода проведено в работе [14].

Иерархия

Коснемся, наконец, бегло иерархии в понимании Саймона. В связи с тем, что составляющие систему элементы (блоки принятия решений) обладают ограниченной «решающей» способностью (или ограниченными «интеллектуальными возможностями»), глобальная цель организации, отражающая ее назначение в целом, подразбивается на последовательность подцелей, так что достижение полной цели равноценно достижению совокупности подцелей. Очевидно, что эта концепция соответствует концепции слоя, введенной в гл. 2. Такая иерархия целей может применяться как для проектирования многоуровневой системы, так и для решения сложных задач. Мы еще раз обращаем внимание на важное различие между

иерархией целей и иерархией элементов, принимающих решения; подробнее этот вопрос рассматривается в гл. 2.

Структурные концепции типа линейной и «штатной» подчиненности, сферы полномочий и т. д. в организационных иерархиях очевидным образом могут быть выражены и в нашей формализации. Действительно, организация представляется в теории многоуровневых систем в виде организационной структуры, составленной из (динамических) элементов (блоков), принимающих решения в соответствующих их рангу узловых точках схемы.

Специализация (децентрализация) и координирование

Одной из самых существенных структурных характеристик организации является специализация и неизбежно сопутствующая ей координация. Наша цель — показать, что развитые в гл. 2 методы могут быть использованы при решении проблем координирования, встречающихся при рассмотрении реальных организационных систем. Ряд аргументов в пользу такого утверждения уже был приведен в предыдущем разделе. В дополнение к ним отметим особую применимость в этой области принципов координирования, основанных на взаимодействиях между нижестоящими элементами.

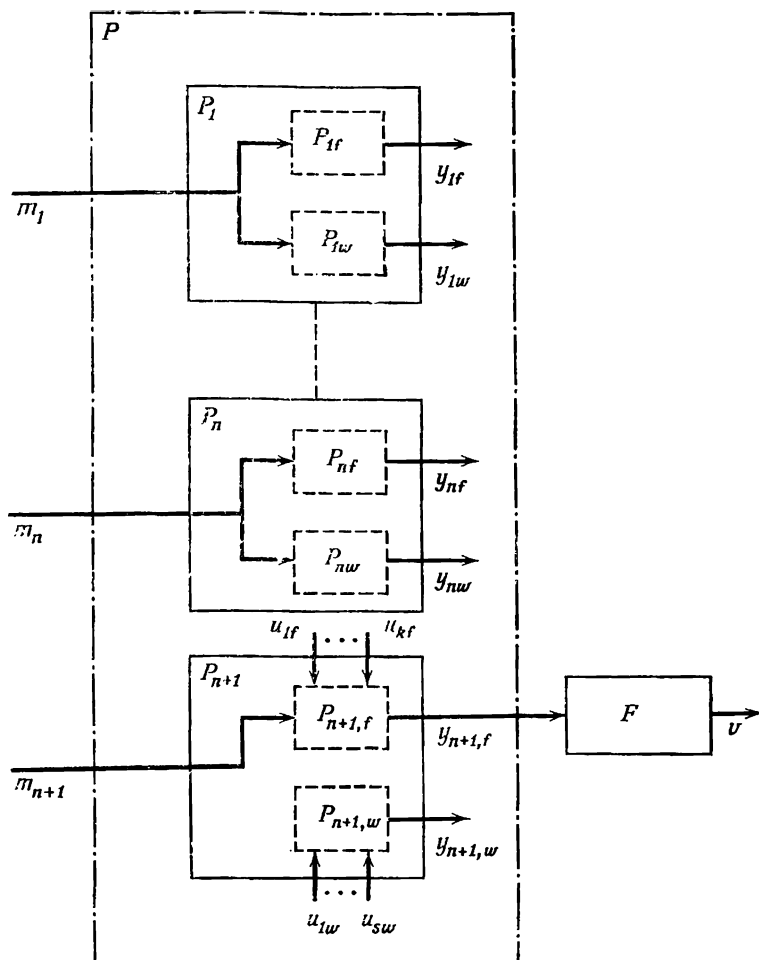
Специализация — одна из основных отличительных особенностей организации. На деле организация всегда возникает в результате выделения отдельных видов работ и передачи их особым специализированным единицам. В широком смысле слова специализация приводит к образованию «целевых» и «функциональных» органов, появлению в системе так называемых «линейных» (line) и «вспомогательных», или «обеспечивающих» (staff) элементов. Задачи, решаемые линейными элементами, отражают назначение организации. Деятельность «вспомогательных» элементов направлена на обеспечение решения линейными элементами своих задач. Линейные элементы обычно определяют, что и когда будет сделано, тогда как вспомогательные элементы вырабатывают, например, рекомендации, как это может быть сделано наилучшим образом.

Специализация любого рода требует, чтобы организация, ее использующая, обладала еще одной характерной особенностью, имеющей решающее значение для ее успешной деятельности. Для полного выполнения задачи *специализированные операции должны быть скоординированы*. Координирование, называемое в теории организаций также управлением, подразделяется естественным образом на две части: установление операционных правил, предписывающих членам организации, как они должны действовать, и практическое обеспечение выполнения этих правил в деятельности организации. Первое называют «управлением в большом», а второе — «управлением в малом» [8].

В нашей формализации «управление в большом» соответствует выбору подходящих функций для оценки эффективности деятельности нижестоящих элементов, или, в более общем смысле, выбору способов координирования. «Управление в малом» соответствует выбору конкретных значений координирующего воздействия. Как правило, мы, говоря о координации, имеем в виду лишь «управление в малом».

В теории организаций [10] учитывается, что один из центральных вопросов, возникающих перед организацией, которая вводит у себя специализированные подразделения, — определение степени самостоятельности элементов организации. Элемент самостоятелен в той степени, в какой условия для его функционирования не зависят от того, что происходит в других элементах системы. В нашей формализации роль переменных, отражающих степень самостоятельности, играют *взаимодействия* между нижестоящими элементами. *Проблема координации, таким образом, связана прежде всего с расчетом взаимодействий нижестоящих элементов.* Решение этой задачи проводится в ч. II с помощью так называемых принципов координации. Последние по существу определяют стратегии, которыми координатор может воспользоваться, чтобы компенсировать то обстоятельство, что отдельные элементы (отделы или функциональные подразделения) действуют так, как если бы они были «самостоятельны». Принципы и связанные с ними методы координирования порождают целое семейство нормативных решений для задачи «управления организацией». Они не только указывают, как координировать специализированные элементы, но, предлагая новые методы координации, позволяют выявлять и новые виды специализации. Еще раз подчеркнем, что предлагаемые решения носят нормативный характер и их применимость для каждого конкретного типа организации должна определяться отдельно.

Принципы координации и методы, развитые в гл. 2, представляют определенный интерес и с точки зрения качественных выводов. Очевидно, что всякая специализация, подобная рассмотренной выше, отражает некоторую степень децентрализации. Классический метод децентрализации состоит в установлении раздельных «центров, обеспечивающих прибыль». Отделы в корпорации рассматриваются как более или менее независимые элементы, наделенные ответственностью действовать наилучшим образом с точки зрения максимизации прибыли при заданных ограничениях, наложенных руководством корпорации. Проблема децентрализации по существу сводится к тому, как следует выбрать налагаемые на отдельные элементы (отделы) ограничения, чтобы обеспечить преуспевание корпорации в целом. Стандартный способ координирования децентрализованных организаций состоит в использовании механизма цен; координация строится по аналогии с прин-



Ф и г. 1.9. Декомпозиция экономической системы P на производственные подсистемы P_i , $1 \leq i \leq n$, и сектор потребления P_{n+1} .

ципами функционирования свободного рынка или основанной на свободной конкуренции экономики. Допускается обмен продуктами между подразделениями, причем для обмениваемых товаров устанавливаются внутренние цены; задача эффективной децентрализации сводится, таким образом, к выбору этих внутренних цен.

Покажем, как механизм цен при децентрализации «экономического типа» может быть образован путем применения одного из координационных принципов, а именно формулируемого ниже принципа согласования взаимодействий. Тем самым мы одно-

временно покажем и ценность теории координирования, развиваемой в ч. II, с точки зрения качественных выводов.

Для простоты рассмотрим вариант обоснования правомерности децентрализации в сфере обслуживания в той форме, как это было изложено Эрроу [15]. Это обоснование, конечно, столь же древнее, как и теория «невидимой руки» Адама Смита, но для наших целей удобнее использовать форму, которую ему придал Эрроу.

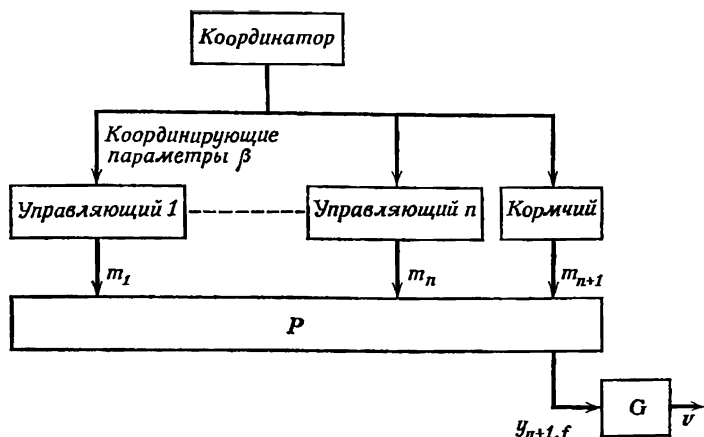
Рассмотрим процесс P , разбитый на $n + 1$ локальных процессов (фиг. 1.9). Подпроцессы P_i , $1 \leq i \leq n$, являются идентичными в том смысле, что каждый процесс P_i имеет один вход m_i и два вида выходов y_{if} и y_{iw} . Подпроцесс P_{n+1} связывает подпроцессы P_i , $1 \leq i \leq n$, как показано на диаграмме, и по-прежнему имеет два вида выходов. Пусть далее взаимодействия между подпроцессами являются входами для P_{n+1} и задаются уравнениями

$$\begin{aligned} u_{if} &= K_{if}(y_{if}, & y_{nf}), & 1 \leq i \leq k, \\ u_{iw} &= K_{iw}(y_{iw}, & y_{nw}), & 1 \leq i \leq s. \end{aligned}$$

Экономическая интерпретация процесса P такова: подпроцессы P_i , $1 \leq i \leq n$, представляют собой производственные процессы или отдельные корпорации в составе экономики, в то время как подпроцесс P_{n+1} есть не что иное, как сектор потребления. Входы подпроцессов P_i , $1 \leq i \leq n$, представляют собой «уровни загрузки» соответствующих подпроцессов в экономике, а выходы — произведенные товары. Вход процесса P_{n+1} представляет собой спрос. Подпроцесс P_{n+1} состоит из части $P_{n+1, f}$, представляющей взаимосвязи между потребителями и произведенными товарами, и части $P_{n+1, w}$, представляющей обмен товаров между отдельными процессами производства. Различие между двумя типами выходов для процессов производства довольно ясно: y_{if} , ..., y_{nf} выпускаются потому, что имеется спрос на эти товары, тогда как y_{iw} , ..., y_{nw} производятся потому, что они необходимы для функционирования системы в целом. Это промежуточные продукты, необходимые для технического функционирования отдельных процессов.

Мы предполагаем, что имеется некоторая функция качества (полезности) G , которая может быть использована для оценки функционирования всей экономики. Обоснование такого предположения можно найти в любом учебнике по математической экономике. Мы предполагаем также, что вся экономика ориентирована на потребителя, так что функция качества зависит только от выходов $y_{n+1, f}$ сектора потребления.

Цель деятельности всей системы — максимизировать функцию полезности (качества) $G(y_{n+1, f})$, варьируя m_1 , ..., m_{n+1} . Такая глобальная задача обычно называется задачей распределения ресурсов в экономике при заданном уровне техники.



Ф и г. 1.10. Двухуровневое иерархическое представление экономической системы.

Эта задача может быть решена на основе полной централизации. Однако возникает следующий вопрос: допустим, что непосредственное управление подпроцессами P_i и сектором потребления P_{n+1} осуществляется с помощью органов локального управления. Пусть такими органами являются «принимающие решения» («решающие») элементы D_i . Принимающими решение элементами D_i для P_i , $1 \leq i \leq n$, являются *управляющие* (производство), тогда как D_{n+1} — принимающий решение элемент для P_{n+1} — это «*кормчий*» [15]. D_{n+1} отражает действия правительства, влияющие на поведение потребителя. Можно ли поставить перед элементами D_i в качестве задач, требующих решения, такие задачи, чтобы, решая их, эти элементы выбирали оптимальные значения m_i по отношению к глобальной функции качества G ? Другими словами, могут ли управляющие и кормчий, преследуя свои собственные интересы, достичь некоего совокупного оптимума по отношению к функции качества G (в данном случае «благо состояния»)?

Рассмотрим двухуровневую систему управления, у которой элементами нижнего уровня (локальными управляющими элементами) являются управляющие и кормчий, а глобальная цель — максимизация благосостояния. Блок-схема системы приведена на фиг. 1.10. Теория координации, развиваемая в ч. II, применима к ситуации подобного рода. Предположим, что применяется координирование, основанное на «развязывании взаимодействий», и что для решения задачи, стоящей перед решающим элементом верхнего уровня (координатором), используется принцип «согласования взаимодействий» (определение этих понятий см. в гл. 4 и 5). Отметим, что функция качества для элементов нижнего уровня до

сих пор не определена. Чтобы предусмотреть координацию, мы используем модифицированный метод получения функций качества для элементов нижнего уровня посредством операторов оценки косвенного эффекта управляющих воздействий, как это делается в ч. II.

Прежде всего введем следующие дополнительные предположения:

1) Выход y_{if} имеет k компонент, $y_{if} = (y_{if}^1, \dots, y_{if}^k)$, тогда как y_{iw} имеет s компонент, $y_{iw} = (y_{iw}^1, \dots, y_{iw}^s)$, где k и s — целые числа, одни и те же для всех подпроцессов.

2) Вход сектора потребления имеет k компонент, $m_{n+1} = (m_{n+1}^1, \dots, m_{n+1}^k)$; кроме того, P_{n+1} и $P_{n+1.w}$ определяются уравнениями

$$y_{n+1, f}^j = u_{jf} + m_{n+1}^j + c_f^j, \quad 1 \leq j \leq k,$$

$$y_{n+1, w}^j = u_{jw} + m_{n+1}^j + c_w^j, \quad 1 \leq j \leq s.$$

3) Взаимодействия описываются уравнениями

$$u_{jf} = \sum_{i=1}^n y_{if}^j, \quad 1 \leq j \leq k,$$

$$u_{jw} = \sum_{i=1}^n y_{iw}^j, \quad 1 \leq j \leq s.$$

Чтобы получить функции качества для элементов нижнего уровня, воздействуя на которые координатор, собственно, и осуществлял бы координирование, мы применим линейные аппроксимации операторов оценки косвенных эффектов управляющих воздействий, как это делается в гл. 6. Функции качества для n управляющих тогда выражаются в виде

$$G_{i\beta}(m_i) = \sum_{j=1}^k \beta_j^i y_{if}^j + \sum_{j=1}^s \beta_w^j y_{iw}^j,$$

тогда как функция качества для потребителя приобретает вид

$$G_{n+1, \beta}(m_{n+1}, n) = G(y_{n+1, f}) - \sum_{j=1}^k \beta_j^i u_{jf} - \sum_{j=1}^s \beta_w^j u_{jw}.$$

Координатор может затем предположить независимость («развязанность») подпроцессов друг от друга и применить принцип согласования взаимодействий для их координирования.

Чтобы получить глобальный оптимум в соответствии с методом координации «развязанных» взаимодействий, элементы нижнего уровня должны максимизировать свои функции качества как по локальным управлениям, так и по взаимодействиям, в то время как координатор должен выбирать координирующие параметры

β_j^i и β_w^i так, чтобы сбалансировать взаимодействия:

$$\sum_{j=1}^n \hat{y}_{jf}^i = \hat{u}_{if}, \quad \sum_{j=1}^n \hat{y}_{jw}^i = \hat{u}_{iw},$$

где $\hat{u}_{if}$ и $\hat{u}_{iw}$ — оптимальные значения, выбранные кормчим, $\hat{y}_{jf}^i$ и $\hat{y}_{jw}^i$ — оптимальные значения, выбранные управляющими.

Этому процессу координирования можно дать соответствующую экономическую интерпретацию. Функция качества управляющих (вычисляемая с использованием проекций и операторов взаимодействия) может рассматриваться как прибыль, получаемая от производства, тогда как координирующими параметрами являются цены продуктов. Следовательно, управляющие максимизируют свои прибыли. Функция качества для кормчего представляет собой разницу между получаемой общественной выгодой и стоимостью производства. Все это снова максимизируется с позиций «сбалансированной» экономики. Координатор выбирает координирующие переменные так, чтобы сбалансировать поставки и спрос на рассматриваемые товары. Таким образом, координатор как бы олицетворяет собой рыночный механизм, а координирующие параметры β_i , $1 \leq i \leq n$, суть не что иное, как цены.

Координируемость системы, таким образом, подразумевает, что существует набор цен, при котором достигается оптимум благосостояния.

Условия координируемости системы такого типа (т. е. действующих в условиях функционирования рыночного механизма) детально исследуются в ч. II. Для простого случая статической системы и выпуклости функции G эти условия вытекают из хорошо известных теорем эконометрии [5]. Следовательно, использование принципа согласования взаимодействий дает, как частный случай, классическое описание оптимального регулирования с помощью рыночного механизма.

Нужно отметить, что правильная экономическая интерпретация также требует, чтобы были выполнены следующие неравенства:

$$m_i \geq 0, \quad 0 \leq i \leq n+1, \quad y_{n+1\ f} \geq 0, \quad y_{n+1\ w} \geq 0.$$

Как показано в гл. 7, это не накладывает ограничений на использование принципа согласования взаимодействий.

Теперь можно оценить как общность теории координации, так и ее потенциальную полезность при разработке новых методов координирования. В связи с этим уместны следующие замечания:

1) В теории децентрализованного функционирования организаций встает вопрос об оптимальности децентрализованного управления и о том, как достигнуть оптимальной координации для (весьма распространенных на практике) случаев, когда взаимосвязь между подразделениями (производственными процессами)

не построена на строго конкурентной основе ввиду внешних воздействий или в связи с дополнительными ограничениями, накладываемыми корпорацией ради достижения ее целей.

На многие из этих вопросов можно ответить, опираясь на общую теорию координации, описанную в ч. II.

2) Основная проблема координирования состоит в том, как скоординировать взаимодействия («ключевые переменные», по терминологии Саймона). Принципы координации, изложенные в ч. II, указывают для этого много различных путей. Очевидно, механизм цен — это частный случай принципа «согласования», используемого при определенных ограничениях. Нет причин, по которым децентрализация в организации не могла бы основываться и на других принципах в условиях, резко отличающихся от тех, которые вытекают из аналогии с рыночным механизмом. Можно предвидеть, что выбор подходящего принципа, а также формы его применения будет зависеть от типа рассматриваемой организации.

3) Следует заметить, что принцип оценки взаимодействий, вводимый в гл. 4, определяется исходя из задачи нахождения *удовлетворительных решений* на уровне нижестоящих решающих элементов и, вероятно, приведет к радикальному отходу от традиционных методов децентрализации в фирмах.

Подчеркнем еще раз, что эти рекомендации имеют в настоящее время только нормативную ценность, указывающую, как следовало бы действовать в специфических условиях. Однако общность условий применимости этих принципов наводит на мысль о возможности их использования и как описательных теорий качественного характера, подобно тому как это имело место в случае действия рыночного механизма.

Заключительные замечания

Для второй половины XX в. характерно появление поистине гигантских организаций: в виде промышленных корпораций и бюрократического государственного аппарата. Это бросает новый вызов исследователям не только в связи с размерами организаций, но и в связи с изменением методов и принципов административного управления ими. Это изменение связано

появлением новой техники, использование которой становится доступным руководителям не только для выполнения тех задач, ради которых была создана организация, но и для улучшения и облегчения управленческого труда. Наилучшим примером может служить применение электронных вычислительных машин в управлении, в частности широкое использование информационных систем. Чтобы полностью использовать преимущества этих новых технических средств, необходим и новый подход к управле-

нию. Поскольку ЭВМ позволяют не только резко улучшить процесс обработки информации, но и расширяют логические возможности принимающего решение лица, руководитель ощущает необходимость в более надежной базе для использования этих новых инструментов. Общая математическая теория многоуровневых систем благодаря своей общности и строгости открывает возможности для создания такой базы. Справедливость такого достаточно смелого заявления можно будет обосновать только после более подробного и тонкого исследования. Если данная книга послужит толчком к такому исследованию, мы будем считать, что наши усилия не были напрасными.

3. ИЕРАРХИЧЕСКИЙ ПОРЯДОК В ПРИРОДНЫХ СТРУКТУРАХ

Наибольших успехов исследователи обычно добиваются, когда выделяют для изучения узкие специализированные области, в которых используют специфические подходы и инструменты. Но существуют объекты исследования, такие, как человек и его естественное или экономическое окружение, которые нельзя выделить и изучать изолированно, как это делается в специализированных областях науки. То, что такое локальное исследование изолированных аспектов способно все же привести к глубоким, содержательным результатам, связано с тем обстоятельством, что в природе существует определенная иерархическая упорядоченность структур, многоуровневое строение естественных явлений. Что представляет собой уровень, каковы главные уровни — на эти вопросы пока нет определенного ответа; на деле ответ на них в значительной степени зависит от подхода, от интересов исследователя и методов анализа. Альберт Сент-Дьердь говорит по этому поводу:

«Если вы попросите химика выяснить, что такое динамо-машина, то первое, что он сделает — растворит ее в соляной кислоте. Молекулярный биохимик, вероятно, разберет ее на части и подробно опишет обнаруженные при этом спиральные обмотки. Если же вы намекнете ему, что то, что приводит в движение машину, возможно, представляет собой некую невидимую субстанцию, называемую «электричеством», он обзовет вас «виталистом» [16].

Некоторые уровни, однако, выявлены достаточно определенно, так что можно говорить об иерархических порядках структур по отношению к этим установленным уровням: это ядерный и атомный уровни в физике, молекулярный и клеточный уровни, уровни тканей и органов в биологии.

Нас интересует здесь математическая теория иерархий, тесно связанная со структурами, являющаяся в то же время формальной и не зависящей от области применения. Поэтому мы не станем обсуждать роль, типы и функции иерархий в природе, а ограничимся рассмотрением их формальных аспектов.

Среди наиболее характерных черт иерархического порядка структур можно указать на следующие:

1) Имеется различие на порядок величины в размерах характеристических элементов различных уровней.

2) Что именно образует элемент данного конкретного уровня — зависит от механизма взаимодействия, существующего на этом уровне. Для успешности анализа необходимо, чтобы рассматриваемое явление было в достаточной степени изолированным: следовательно, явление должно охватывать все сильно взаимодействующие аспекты. Для каждого уровня можно указать подходящее понятие взаимодействия, определяющее элемент, который можно выделить для анализа.

Расплывчатость этих двух критериев («порядок величины» и «интенсивность и тип взаимодействия») отчетливо свидетельствует о трудностях, с которыми приходится встречаться при установлении такой иерархии структур, которая могла бы выдерживать испытание временем даже в течение не очень длинного периода. Вероятно, единственный твердо установленный факт — это существование иерархической упорядоченности по крайней мере в том смысле, что результаты анализа, проведенного над локализованным элементом на некотором заданном уровне, могут быть подтверждены экспериментально.

Существование иерархий в природе неоднократно обсуждалось многими авторами, хотя при этом лишь очень мало добавлялось к тому, что уже было известно, а именно, что иерархические структуры действительно существуют. В ответ на подобные рассуждения ученые-экспериментаторы, ведущие лабораторные исследования, обычно лишь пожимают плечами; как бы ни был важен этот факт, он не может помочь экспериментатору в его работе (и пожалуй, в настоящее время не может также помешать ему!): ученый не может извлечь из него никакой пользы для себя.

Видимо, настала пора изменить такое положение, и к этому нас особенно побуждают следующие факты: 1) возрастает число важных научных проблем, которые просто не могут обойтись без многоуровневого анализа. Пример такого рода проблемы будет приведен ниже; 2) чтобы можно было применить многоуровневый анализ, необходимо достаточно хорошо разработать соответствующие средства и методы анализа на отдельных уровнях. Во многих областях это, по-видимому, уже сделано. Так, например, при исследовании молекулярных свойств можно изучать явления как на клеточном, так и на субклеточном уровнях.

Наиболее убедительные примеры, подтверждающие необходимость многоуровневого анализа, нам дает биология. Хороший отчет о проведенных многоуровневых экспериментах имеется по проблеме дифференциации клеток *in vitro* [17]. Мы кратко изложим его, опуская подробности, не относящиеся к нашей непосред-

ственной цели. Процесс дифференциации клеток невозможно изучать на клеточном уровне путем наблюдения за грубыми морфологическими характеристиками, такими, как форма, конфигурация и т. д. Чувствительность при регистрации изменений значительно увеличивается, если наблюдение проводится на более низком, *субклеточном*, молекулярном уровне. Более того, было обнаружено, что дифференциация клеток происходит наиболее легко, если: а) клеточные агрегаты превосходят по размеру некоторую минимальную величину и б) они находятся в окружении клеток иного рода. Поэтому *надклеточные* аспекты здесь играют важную роль. Экспериментальные наблюдения приводят к такому выводу: «Необходимо признать, что мы имеем дело с многоуровневым явлением и что частью нашей задачи является определение связи между уровнями» [17]. Другими словами, хотя наш интерес и сосредоточен на клеточном уровне, необходимо принимать во внимание и смежные подклеточные и надклеточные уровни. Можно было бы, вероятно, создать такое окружение, которое так взаимодействовало бы с клеткой, что по крайней мере надклеточный уровень можно было бы исключить из рассмотрения. Пока это только теоретическая возможность. Однако, если бы это и оказалось возможным для исследований по дифференциации клеток, в других областях было бы невозможно ограничить наблюдения одним уровнем даже умозрительно, — например, там, где речь идет о динамике развития. Одной из важных проблем, связанных с результатами дробления оплодотворенной яйцеклетки, является проблема межклеточных связей, которая по определению носит многоклеточный характер. Возникновение многоклеточного свойства из исходной одиночной клетки требует одновременного исследования как клеточного, так и надклеточного уровней.

Учитывая очевидное значение проблемы иерархий и отсутствие не только количественной теории, но даже и понятийного аппарата для рассмотрения иерархических систем, имеет смысл обсудить возможные причины такого положения и предложить какие-то меры для его исправления.

Прогрессу или даже просто началу построения теории (концептуальной или математической) иерархических систем препятствовало то обстоятельство, что здесь на неправильно заданные вопросы пытались давать ответы, основанные на неправильных предпосылках; мы имеем в виду проблему иерархической упорядоченности целого, или вопрос о том, что является наилучшим или истинным иерархическим порядком. Более того, проблемы, которые по существу своему носили естественнонаучный характер, пытались формулировать на языке квазифилософских терминов, что заставляло заниматься вопросами о важности и значении иерархического порядка. Мы убеждены в том, что теория иерархических систем (для создания которой время, очевидно, вполне

назрело) должна начинаться с рассмотрения некоторых конкретных, точно сформулированных и в то же время бесспорно многоуровневых проблем, которые должны быть исследованы достаточно подробно.

Одна из таких существенно иерархических проблем — как осуществляется переход с уровня на уровень: как поведение системы на одном уровне влияет на системы, расположенные на соседних уровнях. Заметим, что этот вопрос относится не к иерархии вообще, а лишь к взаимосвязи между двумя смежными уровнями. Аналогичным образом проблема состоит не в том, чтобы выяснить, какие свойства повторяются на каких-то или на всех уровнях, а скорее в том, чтобы выявить межуровневые свойства. Только такими заведомо малыми, но четкими шагами можно продвигаться к ответу на большие вопросы. Одна из таких конкретных иерархических проблем — проблема координации — рассматривается математически в ч. II.

Следует, наконец, сделать еще два замечания, представляющие интерес, в частности, для подробной теории двухуровневых систем принятия решения, изложенной в ч. II.

1) Многоуровневые системы принятия решения дают нам в руки новый тип моделей для изучения физиологических проблем. В качестве иллюстрации можно привести проблему движений глаз в процессе зрения. Хорошо известно, что процесс слежения глазами за зрительными объектами осуществляется при помощи двух различных механизмов: в виде непрерывных и саккадических (скачкообразных) движений глаз. Эти две формы движений могут использоваться как порознь, так и, если это необходимо, совместно. Для объяснения всего диапазона возможных поведений строится двухуровневая модель, в которой второй уровень, представляющий собой координирующий элемент, решает, какой из двух способов управления применить на первом уровне и каким образом. Имеются основания полагать, что введение координационных моделей для моделирования биологических систем будет иметь последствия, сравнимые с последствиями введения моделей управления с обратной связью. Ведутся также аналогичные исследования по изучению сердечно-сосудистой системы.

2) Координация (в частности, принципы координирования) дает в наши руки механизм для достижения интеграции. А именно, в имеющей надлежащую структуру двухуровневой системе усилия, необходимые для достижения глобальной цели, распределяются между элементами различных уровней, так что ни один из этих элементов не «уполномочен» добиваться конечной цели; тем не менее цель каждого индивидуального элемента такова, что глобальная цель будет достигнута, если каждый элемент функционирует надлежащим образом. Рассмотрим двухуровневую систему с n элементами на первом уровне. Каждый элемент имеет

собственную цель, которая зависит от координирующего параметра, получаемого от координатора. Координатор имеет цель, отличную от глобальной цели, и выбирает координирующий параметр так, чтобы обеспечить выполнение своей собственной цели. Если зависимость между этими целями закономерна, система может достигнуть глобальной цели. С позиций внешнего наблюдателя вполне уместно считать, что система преследует некую глобальную цель, хотя попытки найти в системе элемент, задача которого состояла бы в достижении именно этой цели, обречены на провал. Интегральное поведение определяется действиями координатора и его стремлением добиться выполнения собственной цели.

Таким образом, *интеграция достигается посредством координации*. Понимание механизмов интеграции связано с нахождением цели координатора. Однако здесь уместно сделать одно предостережение. Мы говорим о координаторе и процессе координации в функциональном смысле; наличие структурно выделенного координирующего элемента вовсе не необходимо.

В данной книге мы стремимся заложить основы математической теории многоуровневых систем. С этой целью в настоящей главе будут определены основные структурные понятия, которые в дальнейшем будут подвергнуты математическому изучению. Учитывая необъятность предмета иерархических систем и ограниченный объем книги, данная глава преследует лишь следующие конкретные цели:

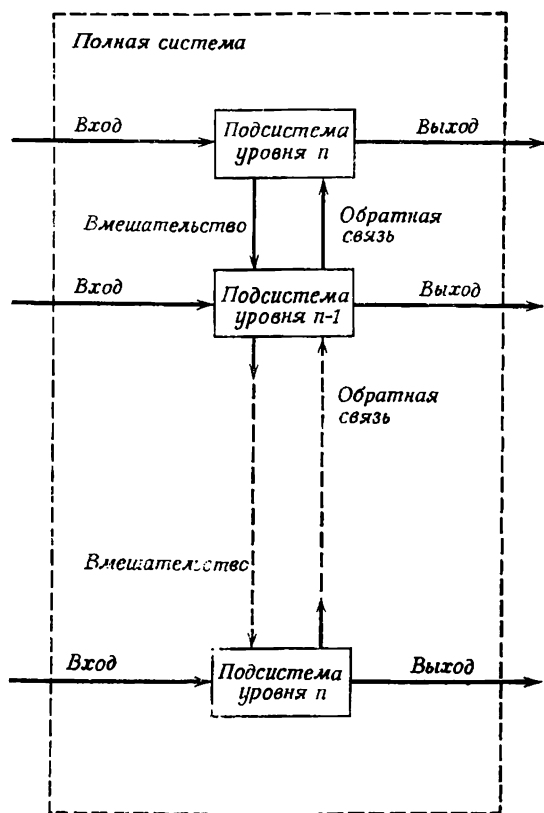
- 1) ввести ряд основных понятий, необходимых для классификации и изучения иерархических систем в целом;
- 2) ввести основные понятия, позволяющие строго сформулировать проблему координирования, которая будет подробно исследована с помощью математических методов в ч. II;
- 3) указать некоторые особенности иерархических систем, которые могли бы быть с успехом использованы в созданных человеком искусственных системах, и попытаться объяснить причину их столь широкого распространения в природе.

1. ЧТО ТАКОЕ МНОГОУРОВНЕВАЯ ИЕРАРХИЧЕСКАЯ СТРУКТУРА?

Понятие многоуровневой иерархической структуры нельзя определить одной краткой и сжатой формулировкой. Уже беглый просмотр приведенных в гл. 1 примеров убеждает нас в том, что исчерпывающее определение потребовало бы перечисления всех возможных альтернатив. Поэтому мы ответим на поставленный вопрос путем указания нескольких существенных характеристик, присущих всем иерархическим системам. К ним относятся: последовательное вертикальное расположение подсистем, составляющих данную систему (вертикальная декомпозиция); приоритет действий или право вмешательства подсистем верхнего уровня; зависимость действий подсистем верхнего уровня от фактического исполнения нижними уровнями своих функций.

Вертикальная соподчиненность

Любая иерархия состоит из вертикально соподчиненных подсистем; это означает, что вся система представляет собой семейство взаимодействующих подсистем, как показано на фиг. 2.1. Под «системой» или «подсистемой» здесь понимается просто осуществление процесса преобразования входных данных в выходные. Это преобразование может либо быть динамическим, протекающим чаще всего в реальном масштабе времени процессом с заранее



Ф и г. 2.1. Вертикальное взаимодействие между уровнями иерархии.

заданным детерминированным алгоритмом и последовательно выполняемыми операциями, либо представлять собой так называемую процедуру «решения проблемы»; в последнем случае декомпозиция носит концептуальный характер: здесь мы имеем совокупность подлежащих выполнению операций, которые могут быть выполнены в разное время и в разной последовательности (системы с недетерминированным алгоритмом). Примеры обеих таких систем будут приведены ниже. Заметим, что как входы, так и выходы могут быть распределены по всем уровням, хотя чаще всего обмен со средой происходит на более низком (или самом низком) уровне. Рассматривая вертикальное расположение, мы будем говорить об элементах верхнего и нижнего уровней с вполне очевидной интерпретацией этих терминов. Укажем также, что взаимодействие между уровнями не обязательно происходит только между каждыми двумя близлежащими уровнями, как для

простоты показано на фиг. 2.1, хотя это в некоторой степени зависит от того, что именно мы рассматриваем в качестве подсистемы на данном уровне.

Право вмешательства

На деятельность подсистемы любого уровня непосредственное и явно выраженное воздействие оказывают вышеразположенные уровни, чаще всего ближайший старший уровень. Это воздействие носит для нижележащих уровней обязывающий характер и в нем находит свое выражение приоритет действий и целей более высоких уровней. В дальнейшем мы это воздействие на более низкие уровни будем называть *вмешательством*. В системах с детерминированным алгоритмом выполнения вмешательство обычно проявляется в виде изменения параметров подсистем нижележащего уровня. В системах же с недетерминированным алгоритмом выполнения приоритет действий задает последовательный порядок получения решений на разных уровнях; обычно проблема (или алгоритм получения решения) на нижележащем уровне не определяется в окончательном виде до тех пор, пока не решена проблема на вышележащем уровне. Чтобы подчеркнуть значение приоритета в установлении порядка действий, мы будем называть элементы верхнего и нижнего уровней соответственно *вышестоящими* (supremal) и *нижестоящими* (infimal).

Взаимозависимость действий

Хотя вмешательство (приоритет действий) направлено сверху вниз, в виде отдачи приказов или команд, успешность действия системы в целом и фактически элементов любого уровня зависит от поведения всех элементов системы. Так как само понятие приоритета подразумевает, что вмешательство *предшествует* действиям более низких уровней, успешность работы верхнего уровня зависит не только от осуществляемых им действий, но и от соответствующих реакций нижних уровней, точнее от их суммарного эффекта. Поэтому можно считать, что качество работы всей системы обеспечивается обратной связью, т. е. реакциями на вмешательство. Информация о которых направляется снизу вверх (фиг. 2.1). Такая взаимозависимость действий особенно очевидна в уже упоминавшемся случае, где обмен с окружающей средой происходит в основном или исключительно на самом нижнем уровне системы.

2. ОСНОВНЫЕ ВИДЫ ИЕРАРХИЙ

Мы введем здесь три типа иерархических систем, которые в некотором смысле отражают классификацию иерархий. На данном этапе нашего рассмотрения это важная задача, поскольку

уточнение высказываемых о системах утверждений и более строгая формулировка задач помогает прояснению сути обсуждаемых проблем.

Как почти всегда бывает в науке, классификацию не следует понимать как строгое разделение; она лишь подчеркивает различия, но отнюдь не исключает возможности существования систем, принадлежащих одновременно к нескольким классам.

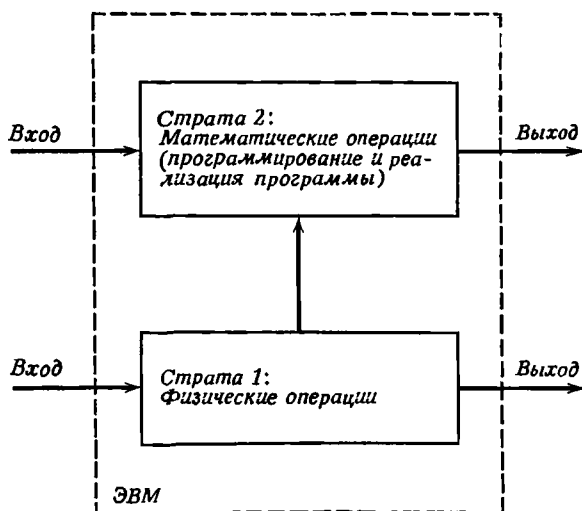
Мы вводим три понятия уровней: а) уровень описания, или абстрагирования; б) уровень сложности принимаемого решения; в) организационный уровень. Для их различения введем следующие термины: «страта», «слой» и «эшелон». Термин «уровень» сохраним как родовой, относящийся к любому из этих понятий, когда нет необходимости в дальнейших уточнениях. Укажем также, что при описании реальных иерархических систем могут одновременно использоваться все три понятия; случай, когда применяется только одно из них, скорее исключение, нежели правило.

Страты. Уровни описания, или абстрагирования

Действительно сложную систему почти невозможно описать полно и детально, что по существу вытекает уже из определения такой системы. Основная дилемма состоит в нахождении компромисса между простотой описания, что является одной из предпосылок понимания, и необходимостью учета многочисленных поведенческих (т. е. типа вход — выход) характеристик сложной системы. Разрешение этой дилеммы ищется в иерархическом описании. Система задается семейством моделей, каждая из которых описывает поведение системы с точки зрения различных уровней абстрагирования. Для каждого уровня существует ряд характерных особенностей и переменных, законов и принципов, с помощью которых и описывается поведение системы. Чтобы такое иерархическое описание было эффективным, необходима как можно большая независимость моделей для различных уровней системы.

Чтобы отличить эту концепцию иерархии от других, мы будем использовать для нее термин *стратифицированная система*, или *стратифицированное описание*. Уровни абстрагирования, включающие стратифицированное описание, будем называть *стратами*.

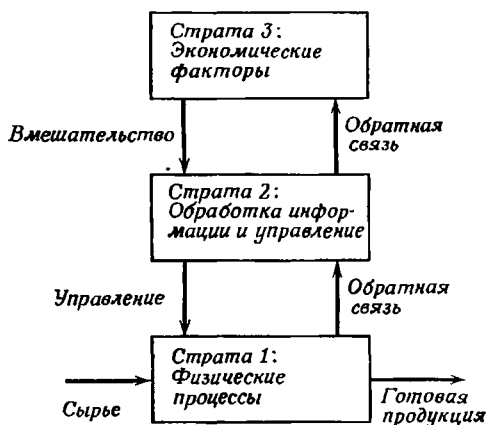
Много примеров стратифицированного описания можно найти в естественных науках. Обсуждавшаяся в гл. 1 иерархическая упорядоченность структур в науках представляет именно такой случай. На каждой страте в иерархии структур имеется свой собственный набор переменных, которые позволяют в значительной степени ограничить изучение только одной стратой. Независимость страт открывает возможность для более глубокого и деталь-



Ф и г. 2.2. Стратифицированное представление ЭВМ с помощью двух страт.

ного изучения поведения системы; однако предположение о полной независимости страт было бы неоправданным, поэтому пренебрежение их взаимной зависимостью может привести лишь к неполному пониманию поведения системы в целом. В самом деле, ограничение, скажем, только биологическим исследованием уже само по себе означает изоляцию, ибо совершенно очевидно, что рассматриваемая система (человек) может быть описана также, с одной стороны, на страте химии или физики, а с другой — на страте экологии или социологии.

Для иллюстрации приведем несколько примеров созданных человеком систем, требующих стратифицированного описания. Рассмотрим модель электронной вычислительной машины. Ее функционирование обычно описывается не менее чем на двух стратах (фиг. 2.2). На первой страте система описывается на языке физических законов, управляющих работой и взаимодействием ее составных частей, в то время как на второй страте мы имеем дело с абстрактными нефизическими понятиями, такими, как двоичные разряды или информационные потоки. На страте физических законов нас интересует правильное функционирование различных электронных компонентов. На страте обработки информации мы имеем дело с проблемами вычисления, программирования и т. д., а стоящие за этим основные физические законы в явном виде не рассматриваются. Разумеется, может представить интерес и описание системы или каких-то ее подсистем на других стратах, помимо этих двух; так, страта атомной физики может представить интерес при



Ф и г. 2.3. Стратифицированное представление автоматизированного промышленного производства.

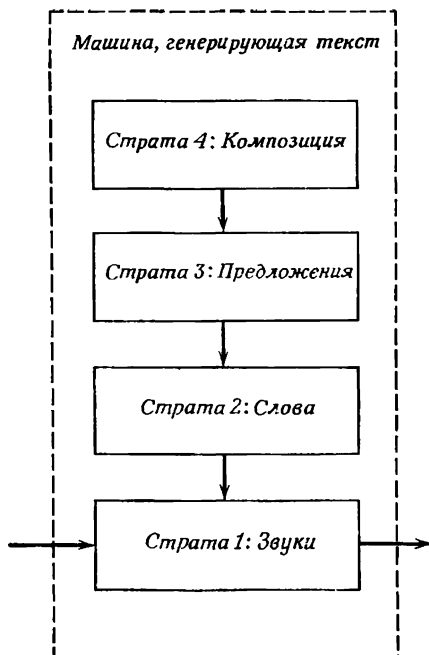
обработки материалов и преобразования энергии, б) управление и обработка информации, в) экономика производства с точки зрения его производительности и прибыльности (фиг. 2.3). Заметим, что на любой из этих трех страт мы имеем дело с тем же самым предметом — основным физическим продуктом. На первой страте он рассматривается как физический объект, который подлежит обработке в соответствии с физическими законами; на второй страте его рассматривают как управляемую переменную; на третьей страте это уже товар как экономическая категория. Для каждого из этих аспектов системы имеется свое описание и своя модель, однако система, конечно, остается одной и той же.

Прекрасным примером использования иерархии моделей одной и той же системы может служить машина, создающая «устные» литературные композиции [18]. Она имеет лишь один выход — реальное физи-

конструировании некоторых электронных компонентов, а так называемая системная страта окажется полезной, например, при разработке систем с разделением времени.

Другой пример стратифицированной системы, созданной человеком, — автоматизированный промышленный комплекс (обсуждавшийся уже в гл. 1). Полностью автоматизированный промышленный комплекс обычно моделируют на трех стратах:

а) физические процессы



Ф и г. 2.4. Представление машины, генерирующей текст, с помощью четырех страт.

ческое «произнесение» литературного текста. Управление же системой может, как видно из фиг. 2.4, рассматриваться минимум на четырех стратах. Первая страта имеет дело с генерацией букв, причем система описывается как машина, производящая звуки. На второй страте осуществляется объединение букв в последовательности, которые воспринимаются как слова в грамматике данного языка: система рассматривается как машина, производящая слова. На третьей страте система рассматривается с точки зрения построения предложений в соответствии с заданными синтаксическими и семантическими правилами. Наконец, на четвертой страте система оценивается в соответствии с определенными литературно-эстетическими стандартами с точки зрения стиля и литературной ценности всей композиции.

Перечень примеров стратифицированных систем можно было бы продолжить. Однако приведенных примеров достаточно, чтобы проиллюстрировать некоторые общие характеристики стратифицированного описания систем.

1. *Выбор страт, в терминах которых описывается данная система, зависит от наблюдателя, его знания и заинтересованности в деятельности системы, хотя для многих систем некоторые страты кажутся естественными, внутренне им присущими.* Это утверждение уже было проиллюстрировано выше. В случае с ЭВМ человек, незнакомый ни с назначением системы как вычислительного устройства, ни со способом ее использования в качестве такового, мог бы ограничиться лишь стратой физических законов; при наличии достаточного времени он мог бы дать весьма подробное и точное описание системы, даже не догадываясь о ее вычислительных свойствах. И наоборот, кое-кто может рассматривать схему обработки информации, не представляя себе, какие физические законы лежат в ее основе. Если нам неизвестен язык, на котором упомянутая выше машина произносит текст, то лучшее, что можно сделать, — это распознать страту машины, производящей звуки; если же мы знаем язык, но не имеем литературных критериев или познаний для оценки смысла и композиционного построения текста, высшая страта будет утрачена.

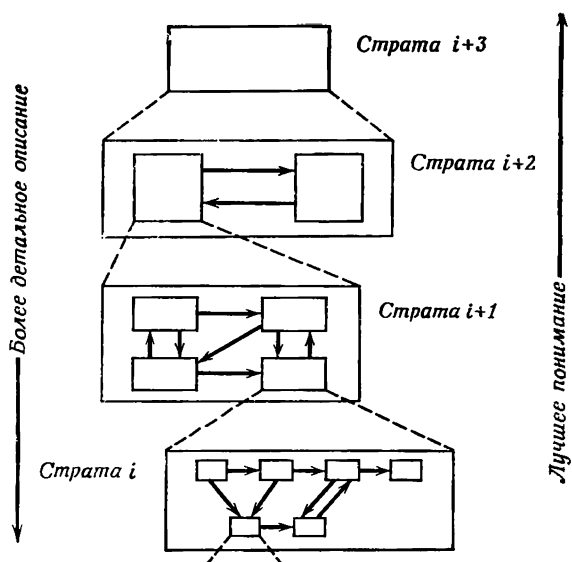
В общем случае стратификация неразрывно связана с интерпретацией производимых системой действий. Контекст, в котором рассматривается и применяется система, определяет, какую страту выбрать как основную и даже, более того, какие страты вообще будут использоваться. Следует, однако, заметить, что почти всегда существуют некоторые страты, хотя и присущие системе, но не представляющие интереса. Возьмем для примера снова машину, произносящую текст: между стратами составления предложений и литературной композиции можно ввести страту стилистических аспектов композиции. Аналогично, рассматривая отдельные серии вычислений, производимых ЭВМ, можно выделить страту спе-

циальных задач, на выполнение которых и запрограммирована машина. ЭВМ целевого назначения могут быть спроектированы для выполнения специальных функций, таких, как управление процессом, для выполнения рутинных операций или обработки деловых данных; здесь ЭВМ вполне может рассматриваться как одностратовая система, так же как цифровой регулятор или аппаратура для коррекции траекторий ракет.

2. *Аспекты описания функционирования системы на различных стратах в общем случае не связаны между собой, поэтому принципы и законы, используемые для характеристики системы на любой страте, в общем случае не могут быть выведены из принципов, используемых на других стратах.* Принципы выполнения расчетов или программирования нельзя вывести из физических законов, лежащих в основе работы ЭВМ на нижней страте, и наоборот. Аналогично экономические принципы и физические законы, лежащие в основе функционирования системы, не связаны между собой. Поэтому стратифицированное описание есть описание одной и той же системы с различных точек зрения.

3. *Существует асимметричная зависимость между условиями функционирования системы на различных стратах.* Требования, предъявляемые к работе системы на любой страте, выступают как условия или ограничения деятельности на нижестоящих стратах. Это находится в соответствии с постулированным приоритетом действия. Например, если ЭВМ используется для определенных расчетов, необходимость выполнения арифметических и других операций накладывает определенные ограничения на используемые для их реализации физические процессы. Генерация конкретного текста зависит от работы системы на страте производства звуков, но выбор последовательности звуков определяется уже требованиями более высоких страт. Ход реального процесса определяется требованиями к поведению системы на верхней страте; для надлежащего функционирования системы на данной страте все нижние страты должны работать правильно. Это означает также наличие в иерархических системах обратной связи с получаемыми результатами.

4. *На каждой страте имеется свой собственный набор терминов, концепций и принципов.* То, что является объектом рассмотрения на данной страте, более подробно раскрывается на ниже-расположенной страте; элемент становится набором; подсистема на данной страте является системой для нижележащей страты. Это отношение между стратами показано на фиг. 2.5. На любой данной страте мы изучаем поведение соответствующих систем с точки зрения их внутреннего механизма и эволюции, в то время как их взаимодействие при образовании новой системы изучается на вышележащей страте. Это замечание весьма существенно, ибо показывает, что изучение на нижней страте не всегда лучше или



Ф и г. 2.5. Взаимосвязь между стратами: система для данной страты является подсистемой для следующей более высокой страты.

основательнее, чем на верхней. На нижней страте мы концентрируем внимание на действиях подсистем, откладывая изучение их взаимодействий для вышестоящих страт. Все было бы иначе, если бы на нижних стратах рассматривалась система в целом, точно так же, как на верхних. Однако это обычно не имеет места, поскольку методологические принципы на одной страте, как правило, не подходят для этого. Следует заметить, что отношение «объект — система» для описаний на различных стратах ведет иерархии соответствующих языков описания. Учитывая, что для каждой страты существует конкретный набор понятий и терминов, используемых при описании системы на этой страте, как правило, имеются и различные языки. Эти языки в свою очередь образуют иерархию с семантическими отношениями между любыми двумя последовательными членами иерархии.

5. *Понимание системы возрастает при последовательном переходе от одной страты к другой: чем ниже мы спускаемся по иерархии, тем более детальным становится раскрытие системы, чем выше мы поднимаемся, тем яснее становится смысл и значение всей системы.* Можно показать, что объяснение назначения системы с помощью элементов той же самой страты по существу есть лишь сжатое описание системы, а для правильного понимания функционирования системы необходимо ее описание с привлечением элементов нижележащих, т. е. более детализированных страт. Так, Брэдли указывает:

«Мы имеем ряд последовательных процессов: биолог объясняет передачу наследственных свойств воспроизведением ДНК; биохимик объясняет это воспроизведение формированием пар основных комплементарных нуклеотидов; химик объясняет парность наличием водородных связей; специалист по молекулярной физике объясняет сами водородные связи межмолекулярными электрическими потенциалами; специалист по квантовой механике объясняет потенциалы на основе волнового уравнения» [19].

Обращаясь к нижним стратам, можно более точно и детально объяснить, каким образом система функционирует, как осуществляется та или иная конкретная операция. С другой стороны, при движении вверх по иерархии описание становится более широким, охватывая большее число подсистем и большие периоды времени. В таком более широком контексте легче понять смысл и назначение подсистем. В примере машины, генерирующей литературный текст, смысл функционирования всей системы становится понятным лишь на высшей страте литературной композиции.

Подводя итоги, можно сказать, что для правильного понимания сложной системы фундаментальную роль играет иерархический подход (стратифицированные модели). Вначале можно ограничиться, скажем, одной стратой, в зависимости от интересующей нас задачи и имеющегося запаса знаний, а затем можем либо детализировать свои знания, двигаясь вниз по иерархии, либо добиться более глубокого понимания системы, двигаясь вверх по иерархии. Выбор исходной страты отчасти определяется также простотой описания на ней.

Слоя. Уровни сложности принимаемого решения

Другое понятие иерархии относится к процессам принятия сложных решений. Почти в любой реальной ситуации принятия решения существуют две предельно простые (настолько простые, что о них слишком часто забывают), но чрезвычайно важные особенности:

1) *Когда приходит время принимать решения, принятие и выполнение решения нельзя откладывать; любая отсрочка просто означает, что не найдено такого нового или изменения старого действия, которое было бы предпочтительнее других альтернатив.*

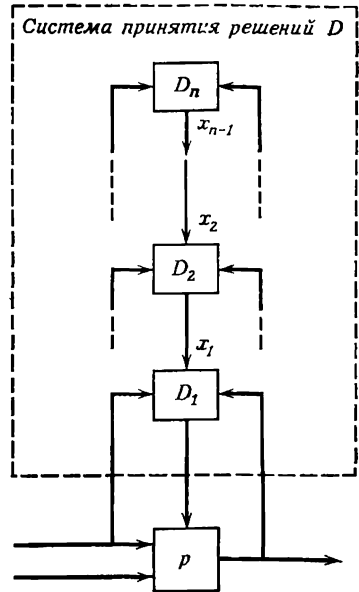
2) *Неясность относительно последствий различных альтернативных действий и отсутствие достаточных знаний о имеющихся связях препятствуют достаточно полному формализованному описанию ситуации, необходимому для рационального выбора действий.*

Эти два фактора приводят к основной дилемме принятия решения: с одной стороны, необходимо действовать немедленно, с другой же — столь же необходимо, прежде чем приступать к дейст-

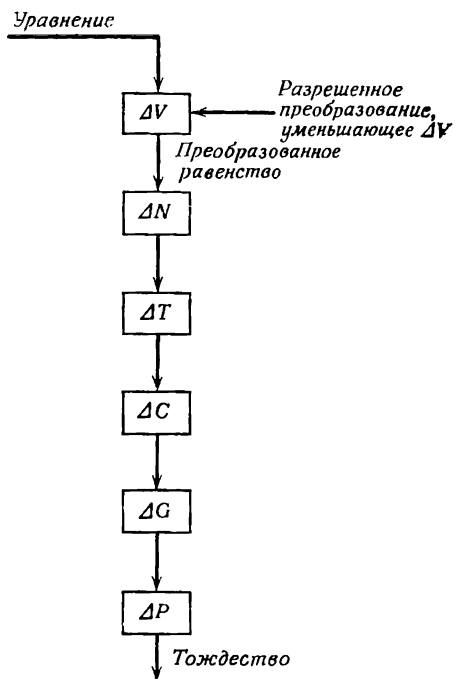
виям, попытаться лучше понять ситуацию. При принятии решения в сложных ситуациях разрешение этой дилеммы ищут в иерархическом подходе. Определяют семейство проблем, которые пытаются разрешить последовательным путем в том смысле, что решение любой проблемы из этой последовательности определяет и фиксирует какие-то параметры в следующей проблеме, так что последняя становится полностью определенной и можно приступить к ее решению. Решение первоначальной проблемы достигнуто, как только решены все подпроблемы. Пример такого разбиения показан на фиг. 2.6. Каждый блок здесь представляет собой принимающий решение элемент. Выход элемента (например, D_2) есть решение или последовательность решений задачи, зависящей от параметра, фиксируемого входом x_2 . Этот вход в свою очередь является выходом принимающего решение элемента более высокого уровня. Таким образом, сложная проблема принятия решения разбивается на семейство последовательно расположенных более простых подпроблем, так что решение всех подпроблем позволяет решить и исходную проблему. Такую иерархию мы будем называть *иерархией слоев принятия решений*, а всю систему принятия решений (обозначенную на фиг. 2.6 через D) — *многослойной системой (принятия решений)*.

Примеры многослойной системы принятия решений легко найти в повседневной жизни. Действительно, личные цели, как правило, весьма расплывчаты и должны еще быть преобразованы в подцели, которые в свою очередь создадут основу для выбора конкретного образа действий. Например, личная цель может заключаться в достижении «счастья» или некоторого уровня удовлетворения, но эту расплывчатую цель еще надо преобразовать в конкретные подцели, ведущие к определенным действиям. Цель надо выбирать так, чтобы ее можно было развернуть в подцели; очень часто лишь после достижения подцелей появляется возможность оценить, приблизились ли мы к первоначальной цели.

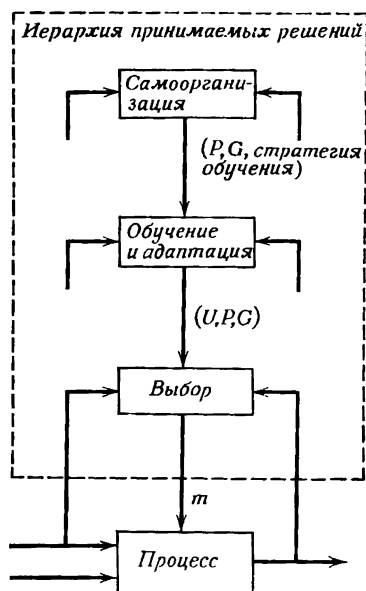
Рассмотрим теперь два примера автоматизированных систем принятия решений, в которых слои иерархии выступают более



Ф и г. 2.6. Многослойная иерархия системы принятия решений.



Ф и г. 2.7. Многослойная стратегия при доказательстве теорем.



Ф и г. 2.8. Функциональная многослойная иерархия решений.

отчетливо. Один пример взят из области «искусственного интеллекта», а другой из области промышленного управления.

В эвристических программах для ЭВМ, предложенных Ньюэллом и др. [20], процесс доказательства теорем в одной из областей математики происходит следующим образом: теорема формулируется как равенство двух математических выражений, к которым могут быть применены преобразования из некоторого допустимого множества. Доказательство теоремы состоит в последовательном преобразовании обеих частей уравнения до получения тождества. Процесс преобразования уравнения в тождество иерархически упорядочен. Этот метод применяется и в области символической логики. Пусть теорема исчисления высказываний задается, скажем, равенством

$$R \wedge (\sim P \rightarrow Q) = (Q \vee P) \wedge R;$$

тогда доказательство теоремы состоит в применении последовательности допустимых преобразований к обеим частям уравнения

вплоть до получения тождества. Общая стратегия доказательства теоремы представляет собой многоуровневую систему (фиг. 2.7). Слои определяются в терминах «различий», которые могут существовать между отдельными выражениями при доказательстве теоремы. Учитываются следующие различия:

- ΔV — указывает на наличие в одном выражении такой переменной, которой нет в другом;
- ΔN — означает, что переменная входит в оба выражения разное число раз;
- ΔT — означает, что различие состоит в отрицании некоторых логических переменных;
- ΔC — означает, что применяются неодинаковые связки;
- ΔG — означает, что переменные по-разному сгруппированы;
- ΔP — означает, что переменные занимают неодинаковые позиции.

Различия затем упорядочиваются в соответствии с введенным приоритетом и используются для выделения различных слоев принятия решений, начиная с высшего, ответственного за наиболее существенное различие. Задача элементов каждого слоя принятия решения — устранить соответствующее различие. Каждый принимающий решение элемент располагает набором преобразований, считающихся полезными для устранения соответствующего различия. Процесс доказательства теоремы начинается с предъявления теоремы высшему элементу, который после устранения соответствующего различия предъявляет преобразованное уравнение следующему элементу. Если каждый из слоев успешно выполняет свою задачу, уравнение в итоге преобразуется в тождество и теорема доказана.

Следует заметить, что на фиг. 2.7 указана лишь принципиальная структурная схема системы доказательства теорем. Полная система значительно сложнее; она предусматривает движение вверх и вниз по иерархии, чтобы избежать тупиков, если в заданный промежуток времени решение на некотором слое не может быть достигнуто; уравнение может быть тогда возвращено на один из предшествующих более высоких слоев, или же оно может быть временно передано следующему, более низкому слою с условием, что в случае необходимости решение будет вновь возвращено на более высокие слои. Но как бы то ни было, даже это упрощенное описание хорошо иллюстрирует необходимость многослойных структур в сложных ситуациях принятия решений.

Второй пример представляет то, что мы называем функциональной иерархией принятия решений или управления. Эта иерархия возникает естественным образом в связи с тремя основными аспектами проблемы принятия решения в условиях полной неопределенности: 1) выбором стратегии, которая должна быть исполь-

зована в процессе решения; 2) уменьшением или устранением неопределенности; 3) поиском предпочтительного или допустимого способа действий, удовлетворяющего заданным ограничениям. Функциональная иерархия, изображенная на фиг. 2.8, состоит из трех слоев:

1. *Слой выбора*: задача этого слоя — выбор способа действий m . Принимающий решение элемент на этом слое получает внешние данные (информацию) и, применяя тот или иной алгоритм (определяемый на верхних слоях), находит нужный способ действий. Алгоритм может быть определен непосредственно как функциональное отображение T , дающее решение для любого набора начальных данных, или косвенно, с помощью процесса поиска. Для примера предположим, что заданы выходная функция P и функция оценки G , а выбор действия, скажем $\hat{m}$, основан на применении функции оценки G к P . Используя теоретико-множественный подход (как это принято в общей теории систем), выходную функцию можно определить как отображение $P: M \times U \rightarrow Y$, где M — множество альтернативных действий; Y — множество возможных результатов на выходе (или «выходов»), а U — множество неопределенностей, адекватно отражающее отсутствие знаний о зависимости между действием m и выходом y . Аналогично функция оценки G есть отображение $G: M \times Y \rightarrow V$, где V — множество величин, которые могут быть связаны с характеристиками качества работы системы. Если множество U состоит из единственного элемента или является пустым, т. е. относительно результата на выходе для данного действия m нет неопределенности, выбор может основываться на оптимизации: найти такое $\hat{m}$ в M , чтобы величина $\hat{v} = G(\hat{m}, P(\hat{m}))$ была меньше, чем $\hat{v} = G(m, P(m))$ для любого другого действия $m \in M$. Если U — более богатое множество, приходится предлагать некоторые другие процедуры для выбора подходящего действия; возможно, при этом придется ввести и некоторые другие отображения помимо P и G . Но в любом случае для того, чтобы определить задачу выбора на первом слое, необходимо уточнить множество неопределенностей U , требуемые отношения P , G и т. д. Это осуществляется на элементах верхних слоев.

2. *Слой обучения, или адаптации*. Задача этого слоя — конкретизация множества неопределенностей U , с которым имеет дело слой выбора. Следует заметить, что множество неопределенностей U рассматривается здесь как множество, включающее в себя все незнание о поведении системы и отражающее все гипотезы о возможных источниках и типах таких неопределенностей. U получают, конечно, с помощью наблюдений и внешних источников информации. Назначение второго слоя — сужение множества неопределенностей U . Если система и окружающая среда стацио-

парны, то множество неопределенностей может быть предельно сужено (до единственного элемента), что соответствует идеальному обучению, как в эксперименте, проводимом в контролируемых условиях. Однако следует подчеркнуть, что U представляет не действительно существующие, а *предполагаемые* системой принятия решения, т. е. учитываемые ею, неопределенности. Второй слой в случае необходимости может полностью изменить U , например расширить его, тем самым как бы допуская, что некоторые базисные гипотезы были несправедливы. Тем не менее основная цель второго слоя (слоя обучения) — насколько возможно сузить множество неопределенностей и таким образом упростить работу слоя выбора.

3. *Слой самоорганизации.* Этот слой должен выбирать структуру, функции и стратегии, используемые на нижележащих слоях, таким образом, чтобы по возможности приблизиться к глобальной цели (обычно определяемой в терминах, которые трудно сделать операционными). Если общая цель не достигается, этот слой может изменить функции P и G на первом слое или стратегию обучения на втором слое в случае неудовлетворительности оценки неопределенности.

Автоматизированные промышленные процессы, описанные в гл. 1, служат хорошими примерами многослойных иерархий. Глобальная цель автоматизации — максимизация прибыли, повышение эффективности и минимизация стоимости производства. Такие грандиозные цели невозможно свести к раз и навсегда выбранным конкретным действиям в обстановке непрерывно меняющихся экономических и технологических условий. Кроме того, необходимо дополнительно учитывать стоимость управления самой автоматизированной системой, а также технологические ограничения, налагаемые характером имеющегося оборудования и сложившейся инженерной практикой; все это приводит к иерархической, многослойной структуре такой сложной автоматизированной системы.

Следует отметить, что функциональная иерархия, изображенная на фиг. 2.8, основана лишь на концептуальном охвате существенных функций в сложной системе принятия решений. Это дает лишь отправную точку для рационального подхода к проблеме выбора функций различных слоев. На практике функция на любом слое выбирается таким образом, чтобы она могла быть реализована с помощью последующей декомпозиции. Так, например, в промышленной автоматике функции слоя выбора обычно осуществляются с помощью прямого управления или регулирования и оптимизации. Задача управления, осуществляющего *регулирование*, — удержать (в условиях неизбежных отклонений) соответствующие переменные около заранее заданных значений, определяемых методами оптимизации и называемых рабочими точками.

Многоэшелонные системы: организационные иерархии

Это понятие **иерархии** подразумевает, что: 1) система состоит из семейства четко выделенных взаимодействующих подсистем; 2) некоторые из подсистем являются принимающими решения (решающими) элементами и 3) принимающие решения элементы¹⁾ располагаются иерархически в том смысле, что некоторые из них находятся под влиянием или управляются другими решающими элементами.

Блок-схема системы такого типа приведена на фиг. 2.9. Уровень в такой системе называется *эшелон*. Эти системы мы будем называть также многоэшелонными, многоуровневыми или многоцелевыми в связи с тем, что различные входящие в систему элементы, обладающие правом принятия решения, имеют обычно «конфликтные» (т. е. противоречащие одна другой) цели. Это противоречие целей является не только побочным результатом эволюции и объединения различных подсистем в одну систему; можно показать, что оно (в некотором смысле и до некоторой степени) даже необходимо для эффективного управления системой в целом.

Наиболее характерный пример систем такого типа — формальные организации людей. Поэтому трудно переоценить важность такого рода иерархий. Много примеров многоуровневых много-



Ф и г. 2.9. Многоуровневая организационная иерархия; многоэшелонная система.

¹⁾ См. примечания на стр. 13 и 117.— Прим. ред.

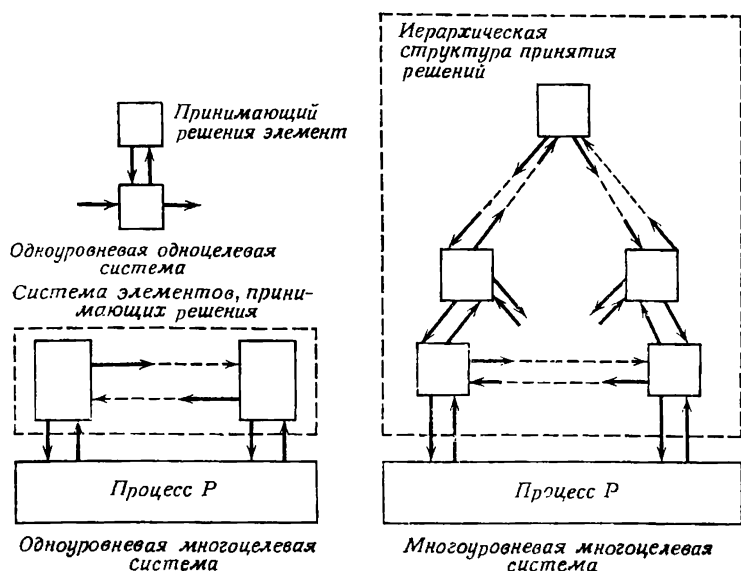
целевых иерархических систем можно найти также в биологии, как, впрочем, и в других областях.

Следует подчеркнуть одну важную характеристику многоуровневых многоцелевых систем, которая отличает их от концептуально более простых (хотя и технически довольно сложных) систем принятия решения с многими переменными. Дело в том, что по самой природе таких многоуровневых многоцелевых систем элементы верхнего уровня в них хотя и обуславливают целенаправленную деятельность элементов нижних уровней, но не полностью управляют ею. Принимающим решения элементам нижних уровней должна быть предоставлена некоторая свобода в выборе их собственных решений; эти решения могут быть, но не обязательно будут, теми решениями, которые выбрал бы верхний уровень. Такая свобода действий — отличительная черта любой социальной или биологической многоуровневой системы. В созданных человеком системах, используемых для принятия решений, затрачиваемые ресурсы могут быть сэкономлены только в том случае, если элементам нижних уровней предоставлена такая свобода. Можно показать, что для эффективного использования многоуровневой структуры существенно, чтобы элементам принятия решения была предоставлена некоторая свобода действий; должно быть проведено рациональное распределение усилий по принятию решений между элементами различных уровней. Только при этом условии будет оправдано само существование иерархии.

Изложенные соображения приводят к концептуально важной классификации систем принятия решений; по характеру иерархического расположения образующих систему элементов (фиг. 2.10) можно указать следующие категории систем принятия решений: а) *одноуровневые одноцелевые системы*, б) *одноуровневые многоцелевые системы*, в) *многоуровневые многоцелевые системы*.

В первом случае цель определяется для всей системы и все переменные выбираются так, чтобы обеспечить достижение этой цели. Технически решение проблемы принятия решения, удовлетворяющее данной цели, может быть очень сложным, так как задача многомерная и может возникнуть необходимость в использовании как методов оптимизации, так и прогнозирования. И все же следует подчеркнуть концептуальную простоту одноуровневых одноцелевых систем, особенно же — отсутствие конфликтов внутри таких систем.

Система, принадлежащая к классу одноуровневых многоцелевых систем, состоит из принимающих решения элементов, имеющих свои собственные цели. Эти цели не обязательно конфликтны; некоторые из элементов, обладающих правом принятия решений, могут образовывать коалиции. Конфликт между принимающими решения элементами может, однако, произойти; тогда



Ф и г. 2.10. Классификация систем принятия решений (управления).

он может быть разрешен только путем вмешательства более высокого уровня.

Наконец, класс многоуровневых многоцелевых систем характеризуется наличием иерархических отношений между принимающими решения элементами этой системы. Существование какого-то высшего командного элемента — принципиальная отличительная особенность таких систем; проблема принятия решений на уровне этого элемента является основной проблемой в теории многоуровневых систем.

Принимая во внимание приведенную выше классификацию, становится ясно, что для изучения многоуровневых систем необходима новая теория. Можно утверждать, что современная теория автоматического управления имеет дело с одноуровневыми одноцелевыми (хотя многопараметрическими и довольно сложными) системами, применительно к которым и рассматривается проблема принятия решений, в то время как для одноуровневых многоцелевых систем мы имеем теорию игр и теорию малых групп. Но ни одна из этих теорий окончательно еще не сформировалась, и необходимо провести еще много исследований, в частности для разработки практических методов (т. е. численных алгоритмов), для управления одноуровневыми одноцелевыми системами или для выяснения природы и влияния конфликтов в одноуровневых многоцелевых системах. Однако в ходе интенсивных исследований, проведенных за последние два или три десятилетия

тия, уже создана основа для развития такой теории, по крайней мере, для определенных классов одноуровневых одноцелевых систем. Очевидно, что для создания теории иерархических многоэшелонных систем необходим новый подход. Разработка основ для развития математической теории таких систем — одна из главных задач настоящей книги.

Связь между различными понятиями уровня

Следует четко определить, какое понятие уровня используется при описании иерархической системы, так как три введенных выше понятия имеют каждое свою область применения, а именно: концепция страт введена для целей моделирования, концепция слоев — для вертикальной декомпозиции решаемой проблемы на подпроблемы, концепция же эшелонов относится к взаимной связи между образующими систему элементами принятия решения. Различие этих трех понятий, пожалуй, лучше всего можно проиллюстрировать, рассмотрев взаимодействие этих концепций при описании многоуровневых систем. Мы рассмотрим три случая.

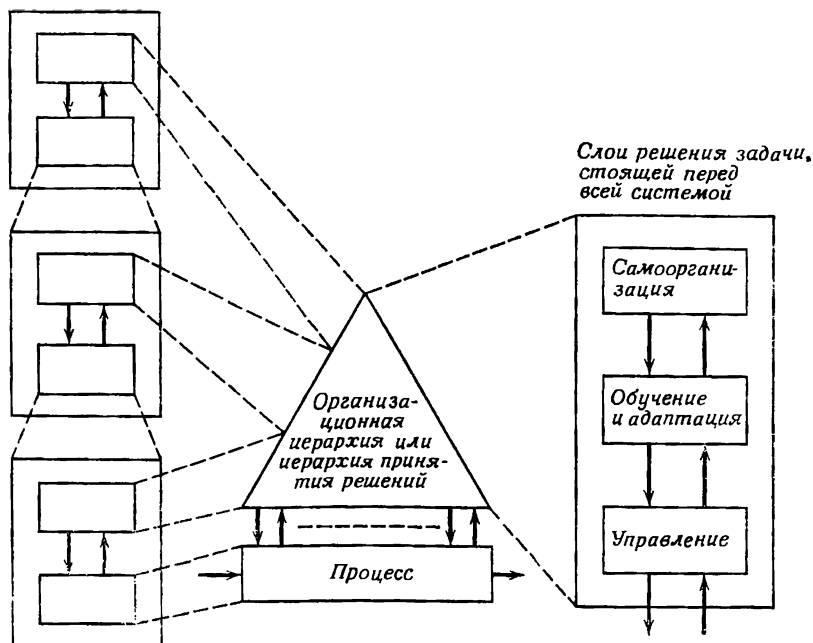
Проектирование многоэшелонной системы

Предположим, что мы строим многоэшелонную (организационного типа) систему. Первая возникающая проблема — распределение задач или ролей, которые должны выполняться различными уровнями или отдельными элементами. Разумная отправная точка обеспечивается при так называемом «системном подходе» к системе в целом и к задаче, которую эта система по предположению должна выполнять. При этом используются иерархические концепции страты и слоя. С одной стороны, происходит стратификация модели всей системы, а с другой — совершается декомпозиция стоящей перед системой задачи на слои. Задания элементам, образующим многоэшелонную систему, в этом случае определяют по отношению к моделям и решаемым проблемам, появляющимся на соответствующей страте или слое (фиг. 2.11). В этой связи следует опять напомнить, что не существует однозначного соотношения между стратами, эшелонами и слоями. Задания для нескольких эшелонов могут быть определены из модели одной и той же страты, в то время как решаемая проблема на данном слое может быть распределена между рядом эшелонов, более того, задание для эшелона может содержать элементы проблем, принадлежащих не одному, а ряду слоев решаемой проблемы.

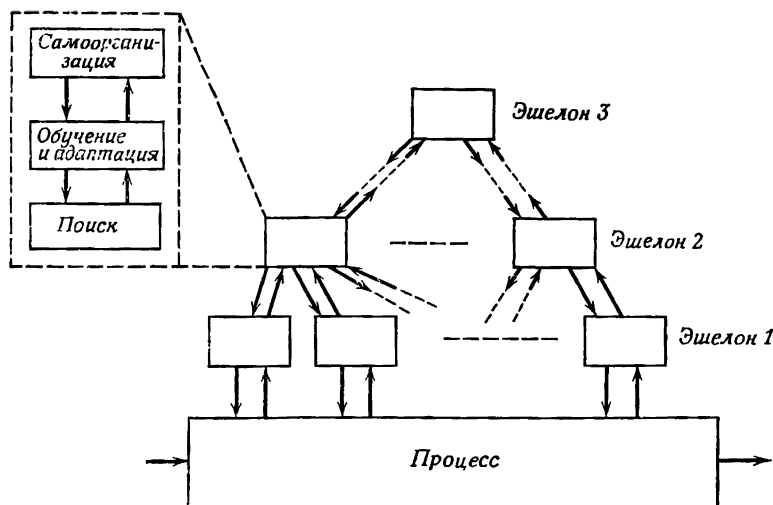
Многослойные элементы в многоэшелонной системе

Концепция многослойной системы была введена применительно к некоторой конкретной решаемой проблеме, а вовсе не обяза-

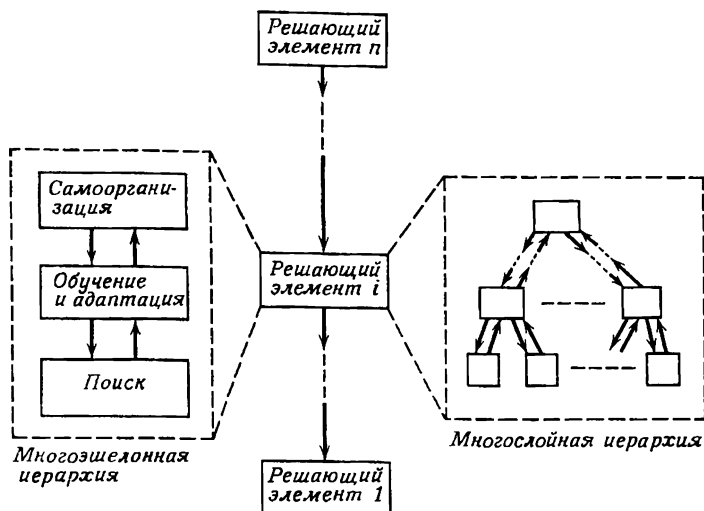
Стратификация
всей системы



Ф и г. 2.11. Вертикальное распределение задач для организационной иерархии.



Ф и г. 2.12. Многослойное представление функционирования решающих элементов многослойной системы.



Фиг. 2.13. Представление решающих элементов, образующих многослойную иерархию, в виде многослойных и многоэтапных иерархий.

тельно для проблемы, стоящей перед всей системой. Следовательно, она может быть и решаемой проблемой, стоящей перед отдельным элементом, являющимся членом большей системы. Так, в примере двухуровневой системы, показанной на фиг. 2.12, каждый принимающий решения элемент использует многослойный подход для решения своих собственных, локальных подпроблем. В этом случае говорят, что многослойная иерархия вложена в многоэтапную систему.

Сложные решающие элементы в многослойной системе

Рассмотрим многослойную систему принятия решений применительно к семейству подпроблем, разрешение которых дает решение исходной проблемы. Каждая из подпроблем может быть достаточно сложной, так что может оказаться целесообразным использовать для ее решения многослойный подход (скажем, функциональную иерархию) или даже сформировать отдельную многоэтапную систему (если, конечно, ресурсы и время позволяют это сделать), которой и будет поручено решение этой конкретной подпроблемы. Фиг. 2.13 иллюстрирует сказанное. Специфические примеры такого рода можно найти в практике работы крупных органов планирования, занимающихся разработкой проблем, стоящих перед корпорациями.

Несмотря на эти различия, существуют и общие для всех трех понятий черты. Некоторые наиболее существенные из них

уже были приведены в разд. 1 этой главы. В основном это были структурные взаимоотношения между подсистемами. В рамках более детального обсуждения иерархии, проводимого в этой главе, укажем еще несколько общих черт, относящихся к задачам и ролям подсистем.

1. *Элемент верхнего уровня имеет дело с более крупными подсистемами или с более широкими аспектами поведения системы в целом.* При многоэшелонной иерархии элемент верхнего уровня является «командным» по отношению к двум или более элементам, и принимаемое им решение координирует их действия в соответствии с целью, определенной для совокупности всех подчиненных ему элементов. Для концепции слоев это следует из ответственности элементов верхнего уровня за поведение системы в течение более длительных промежутков времени. Чтобы собрать информацию, необходимую для сужения множества неопределенностей, слой обучения должен проводить наблюдения в течение ряда периодов принятия решения на первом слое. Чтобы изменить структуру стратегии принятия решений, третий слой (слой самоорганизации) должен наблюдать за действиями нижележащих слоев в течение еще большего периода времени, так как для оценки качества стратегии обучения ее следует испытывать по крайней мере несколько раз. Аналогично и для концепции страт: система на любом уровне образуется из подсистем нижних уровней, и, следовательно, более высокая страта имеет дело с более общим аспектом поведения всей системы.

2. *Период принятия решения для элемента верхнего уровня больше, чем для элементов нижних уровней.* Для концепций слоя и страты это очевидно. Однако это утверждение остается верным и для концепции эшелона. А именно: управляющие воздействия, исходящие от вышестоящего элемента, не могут следовать чаще воздействий, подаваемых нижестоящими элементами, поведение которых он координирует; в противном случае он не сможет оценивать достигаемый эффект (координации).

3. *Элемент верхнего уровня имеет дело с более медленными аспектами поведения всей системы.* Это справедливо для всех трех типов уровней и почти непосредственно вытекает из того, что элемент верхнего уровня имеет дело с более широкими аспектами поведения всей системы и имеет большие периоды принятия решений. Верхние уровни не могут реагировать на такие изменения в окружающей среде или самом процессе, которые происходят быстрее изменений, с которыми имеют дело нижние уровни, так как последние реагируют быстрее и имеют дело с более частными, локальными изменениями.

Эта характеристика особенно интересна для образования многослойной иерархии [24]. Предположим, что все внешние эффекты для системы, рассматриваемой как целое, описаны с помощью

частотных спектров их изменений во времени. Пусть F — диапазон частот для всех возможных возмущений. Множество F можно разбить на подмножества $F_1, \dots, F_n$, образовав многослойную иерархию так, чтобы элемент на любом слое, скажем i -м, имел задачу реагировать на те внешние возмущения, частотный спектр которых лежит в диапазоне F_i . Элементы на i -м слое тогда действуют в предположении, что верхние и нижние слои функционируют надлежащим образом и взяли на себя заботы о всех внешних эффектах, частоты которых лежат за пределами диапазона частот F_i .

4. *Описания и проблемы на верхних уровнях менее структурированы, содержат больше неопределенностей и более трудны для количественной формализации.* Проблема принятия решений на верхних уровнях может рассматриваться как более сложная. Конечно, для решения задачи на верхнем уровне могут использоваться приближенные методы, но тогда точность понижается, и следует соблюдать осторожность при интерпретации результатов. В общем случае для любого уровня существует специфический набор средств для решения соответствующих задач. Например, в многослойной иерархии у каждого слоя существует свой собственный набор методов и алгоритмов: на слое выбора — это управление с обратной связью и численные методы оптимизации; на слое адаптации преобладают статистические методы или методы распознавания образов; на слое самоорганизации используются эвристические методы. Задачи верхнего слоя не удается поставить так, чтобы они имели простое численное решение. Поэтому на практике обычно прибегают к «вмешательству в критических ситуациях», т. е. оценивают общую характеристику и вносят структурные изменения лишь в случае, когда характеристики ухудшаются до такой степени, что изменение становится необходимым. В общем случае, однако, нет априорной гарантии, что изменение приведет к действительному улучшению характеристики.

Сложное переплетение во времени взаимоотношений между уровнями связано также с ограниченной «производительностью» принимающих решения элементов, используемых при построении системы. Предположим, что мы должны построить систему, выполняющую задачу, выходящую за пределы возможностей любого имеющегося в наличии решающего элемента. Тогда мы вынуждены прибегать к многоэшелонной системе, такой, что решающие элементы на верхних эшелонах имеют дело с более общими аспектами задачи и, следовательно, имеют перед собой более сложную проблему принятия решения, чем элементы нижних уровней. Кроме того, ввиду ограниченных возможностей элементов, вырабатывающих решения, элементы более высоких эшелонов вынуждены тратить больше времени на поиск решений своего уровня.

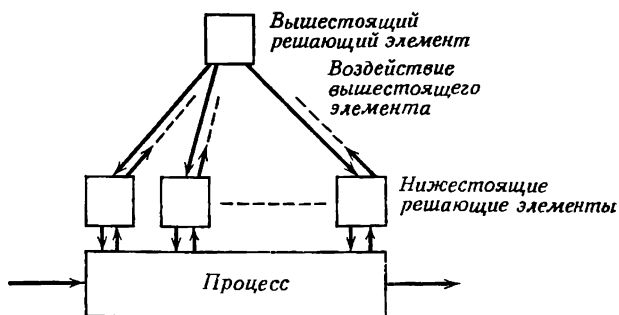
3. ЗАВИСИМОСТЬ МЕЖДУ УРОВНЯМИ И КООРДИНИРУЕМОСТЬ

Как уже указывалось во введении, следует ожидать, что прогресс в развитии теории иерархических систем будет происходить постепенно, отдельными этапами, на каждом из которых будут решаться некоторые специальные задачи, характерные для многоуровневых систем. Одна из таких существенно иерархических задач, которую мы рассмотрим в настоящей книге, касается межуровневых связей в многоуровневой системе, и в частности координации.

Рассмотрим двухуровневую систему принятия решений (фиг. 2.14), имеющую только один вышестоящий (координирующий) элемент и n подчиненных ему (нижестоящих) элементов. Такая система представляет специфический интерес для теории многоуровневых систем по следующим причинам: а) это простейший тип систем, в котором проявляются все наиболее существенные характеристики многоуровневой системы; б) более сложные многоуровневые системы могут быть построены из двухуровневых подсистем, как из модулей. Итак, мы будем исследовать взаимоотношения между уровнями на примере двухуровневой системы.

Ситуация здесь напоминает отношение, существующее между игрой двух лиц с нулевой суммой и более общими играми n лиц; игра двух лиц с нулевой суммой, конечно, представляет собой лишь частный случай общего класса игр, тем не менее она обладает достаточно существенными свойствами, позволившими принять ее за основу при введении ряда важных понятий общей теории игр.

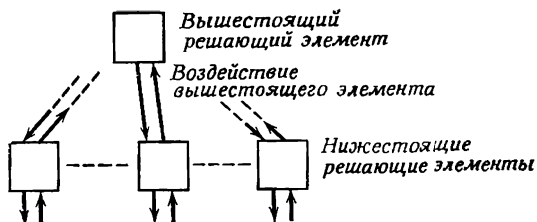
Аналогичным образом мы надеемся развить и теорию двухуровневых систем в качестве первого шага для создания теории многоуровневых систем в общем случае.



Ф и г. 2.14. Двухуровневая организационная иерархия.

Взаимная зависимость уровней

Взаимодействие между вышестоящим элементом и каждым из нижестоящих элементов таково, что действия (и успех) одного из них зависят от действий другого, как это показано на фиг. 2.15.



Ф и г. 2.15. Взаимодействие между вышестоящим и нижестоящими решающими элементами

Так как оба являются элементами, вырабатывающими решения, это означает, что в общем случае проблема, решаемая элементом нижестоящего уровня, зависит от действия вышестоящего элемента, заключающегося в выработке значений определенного параметра; наоборот, проблема, решаемая вышестоящим элементом, зависит от действий (или отклика!) элементов нижестоящего уровня. Очевидно, что получается тупик. Эта дилемма разрешается путем введения приоритета действий вышестоящего элемента. Однако в общем случае следует учитывать динамику: взаимоотношения между вышестоящим и нижестоящими элементами являются динамическими и изменяются во времени.

Времена вмешательства

Вышестоящий элемент может сообщить свое координирующее решение нижестоящим элементам в один из двух моментов времени. Вышестоящий элемент может попытаться скоординировать действия нижестоящих элементов до того, как они примут свои решения. Мы называем такой род действий *вмешательством до принятия решения*. Это основной способ координации, ибо в связи с установленным приоритетом действий задачи, подлежащие решению на уровне нижестоящих элементов, не вполне определены до тех пор, пока не получены конкретизирующие их указания. Вмешательство до принятия решения основано на прогнозировании поведения как самой системы, так и окружающей среды. Более того, вышестоящий элемент в ходе вмешательства до принятия решения определяет функции качества для оценки деятельности элементов нижестоящего уровня и, таким образом, определяет долю участия каждого из них в деятельности ради достижения успеха всей системы.

Через некоторое время после того, как элементы нижестоящего уровня примут и применят свои решения (например, в конце так называемого периода принятия решения), вышестоящий элемент должен снова связаться с нижестоящими. Вышестоящий элемент должен исправить посланные ранее элементам нижестоящего уровня инструкции, если окажется, что допущения, на основе которых эти инструкции были выработаны, стали неверными. Более того, в конце периода принятия решения вышестоящий элемент должен либо подтвердить, либо изменить сообщенные им в ходе вмешательства до принятия решения планы распределения ролей подсистем в обеспечении успеха всей системы. Эти действия координирующего элемента называют *вмешательством после принятия решения, корректирующим* или *поощряющим вмешательством*.

При рассмотрении «вмешательства после принятия решения» следует помнить, что оно вызывается не стремлением упростить работу нижестоящих элементов, а стремлением к достижению целей вышестоящего элемента. Основная цель вмешательства после принятия решения состоит не в уточнении имеющихся у нижестоящих элементов представлений о поведении остальной части системы и не в том, чтобы предоставить им большую роль в достижении цели, как при вмешательстве до принятия решения, а в том, чтобы оказать на нижестоящие элементы такое влияние, которое приведет к улучшению (с точки зрения вышестоящего элемента) поведения системы в целом. Делая это, вышестоящий элемент должен рассматривать поведение всей системы как динамическое и учитывать возможные последствия вмешательства после принятия решения на будущие результаты. Ведь в динамической системе «вмешательство после принятия решения» представляет собой первую фазу «вмешательства до принятия решения» для последующего периода принятия решений. Именно в этом свете следует оценивать различие между вмешательствами до и после принятия решения. Слишком большие расхождения могут отрицательно воздействовать на качество действий нижестоящих элементов.

Дальнейшее изложение (за исключением некоторых замечаний) будет посвящено преимущественно вмешательству до принятия решения.

Основные особенности функционирования вышестоящего элемента

Два вида сигналов связывают вышестоящий и нижестоящие элементы. Сигнал, идущий сверху вниз, конкретизирует задачи, подлежащие решению на уровне нижестоящих элементов, тогда как сигнал, посылаемый вверх, несет вышестоящему элементу информацию о состоянии нижестоящего уровня.

Вмешательство

В связи с приоритетом действий вышестоящий элемент имеет широкие обязанности: во-первых, он указывает нижестоящим элементам, как им следует действовать, а во-вторых, воздействует на них с целью побудить их, если это необходимо, изменить свои действия. Первая обязанность соответствует в теории организации «управлению в большом» и включает в себя выбор алгоритмов и правил поведения в разнообразных предполагаемых обстоятельствах. В нашей формализации ей отвечает задача выбора принципов взаимодействия между вышестоящим и нижестоящими элементами. Мы называем это *выбором способа координации*. Вторая обязанность соответствует в теории координации «управлению в малом» и включает способы и правила «регулирования», целью которых является улучшение качества деятельности. В нашей формализации ей отвечает выбор координационной переменной или переменной реального вмешательства. Мы называем это *координацией* (в широком смысле).

Способ координации определяется тем, как конкретный элемент нижестоящего уровня сообщается с другими элементами своего уровня, а также тем, какие характеристики проблем, решаемых на этом уровне, могут подвергаться изменению в целях улучшения глобального результата. Два перечисленных фактора тесно связаны друг с другом. Они определяются характером задач, для решения которых созданы нижестоящие элементы и сама система. Основываясь на проводимом в ч. II анализе, мы укажем несколько основных категорий.

Взаимосвязь одного элемента с другими элементами того же уровня можно охарактеризовать его действиями, реакцией всей остальной системы и ее воздействием на этот элемент. Это воздействие поступает в виде соответствующего «связующего» сигнала (см. гл. 4) на «внутренний» вход элемента. Центральный вопрос поэтому состоит в том, какое влияние оказывает сигнал, поступивший на внутренний вход, на решения, принимаемые данным нижестоящим элементом. В этом отношении для вышестоящего элемента возможны следующие варианты организации взаимодействия элементов нижестоящего уровня.

1. *Координирование путем прогнозирования взаимодействий.* Вышестоящий элемент посылает нижестоящим элементам значения будущих связующих сигналов. Тогда нижестоящие элементы начинают вырабатывать свои локальные решения в предположении, что связующие сигналы, которые в дальнейшем действительно к ним поступят, окажутся именно такими, какими их предсказал вышестоящий элемент.

2. *Координирование путем оценки взаимодействий.* Вышестоящий элемент задает диапазон значений для связующих сиг-

налов. Нижестоящие элементы рассматривают эти сигналы как возмущения, могущие принимать любое значение в заданном диапазоне.

3. *Координирование путем «развязывания» взаимодействий.* Элементы нижестоящего уровня трактуют связующий сигнал как дополнительную переменную решения. Они решают свои задачи так, как если бы связующие сигналы можно было выбрать произвольно.

4. *Координирование типа «наделения ответственностью».* Элементы нижестоящего уровня знают о наличии других элементов, также принимающих свои решения на том же уровне. Вышестоящий элемент снабжает нижестоящие элементы моделью зависимости между его действиями и откликом системы.

5. *Координирование путем «создания коалиций».* Элементы нижестоящего уровня знают о существовании других решающих элементов на том же уровне. Вышестоящий элемент определяет, какого рода связи разрешены между ними. Это приводит к коалиционным или конкурентным (в теоретико-игровом смысле) отношениям между нижестоящими элементами.

Здесь уместно сделать замечание о пользе и эффективности вышеперечисленных подходов. Последний подход, очевидно, самый уточненный и ближе всего соответствует действительному положению дел в организациях людей. Однако он является и самым сложным и может привести к довольно трудноразрешимым задачам принятия решений для элементов нижестоящего уровня. В предельном случае каждому нижестоящему элементу пришлось бы решать задачи принятия решений и для всех остальных элементов своего уровня. Эффективность такого подхода была бы, следовательно, чрезвычайно низкой.

Многоуровневые структуры позволяют для решения задачи, стоящей перед всей системой в целом, использовать совокупности элементов, каждый из которых по отдельности не в состоянии решить эту задачу. Для этого глобальная задача разбивается на подзадачи и ее решение происходит групповыми усилиями. Искусство синтезирования многоуровневой системы состоит именно в постановке упрощенных задач для вышестоящих и нижестоящих элементов. Первые три подхода имеют очевидные преимущества перед последними в отношении простоты.

В ч. II первые три способа исследуются весьма подробно. Общим для них является то, что внутренние связи элементов нижестоящего уровня рассматривают как каналы, по которым к ним поступают связующие сигналы; остальная часть системы трактуется просто как среда. Между действиями элементов нижестоящего уровня и реакцией системы при этом не предполагается непосредственной причинной связи. Для иллюстративных целей мы кратко отметим ниже ряд аспектов применения двух других способов.

1. Разрешение или запрещение обмена информацией между элементами нижестоящего уровня зависит, как было показано в работе [22], от того, как определяются задачи, стоящие перед этими элементами. Было показано, что в случае, если функции качества (используемые для оценки результатов деятельности нижестоящих элементов) являются линейными, организация коммуникационных связей между элементами одного уровня с общесистемной точки зрения представляется нерациональной, тогда как если указанные функции имеют другой вид (скажем, квадратичный), существование каналов связи может оказаться выгодным. Важно подчеркнуть также семантический аспект коммуникации. В задачи вышестоящего элемента входит не только установление целесообразности тех или иных каналов связи, но и принятие решения по поводу того, какого рода информация будет по ним передаваться. Анализ показывает, что чрезмерно интенсивное общение между элементами одного уровня может иметь такой же эффект, как и отсутствие информационной связи вообще, т. е. ведет к ухудшению глобального результата.

2. Исследование возможности использования на уровне нижестоящих элементов теоретико-игрового подхода показало [14], что вышестоящий элемент может так составить правила поведения, что применение игрового подхода на нижестоящем уровне приводит к глобальному оптимуму. Это важно, например, для проблемы проведения в жизнь оптимального решения в случае социальных систем. Может оказаться, что вышестоящий элемент, зная оптимальное решение, все же не может непосредственно обеспечить его реализацию. Он должен поэтому установить между нижестоящими элементами такие взаимоотношения, чтобы они сами пришли к оптимальному решению. Математический анализ [14] показывает, что такое решение и в самом деле возможно, если на нижестоящем уровне использовать теоретико-игровой подход.

Координация

Координация, сама представляющая собой сложную для решения проблему, имеет два важных аспекта: аспект самоорганизации (изменения структуры) и аспект управления (выбор координирующего вмешательства при фиксированной структуре). Предположим, что способ координации уже определен и что самоорганизация относится лишь к изменениям функций и взаимосвязей, используемых в процессе координации. Эти изменения мы называем *модификациями*. В широком смысле всякая задача принятия решения определяется некоторой целью и образом ситуации, применительно к которой происходит принятие решения. Таким образом, имеются два вида модификаций: *модификация целей* и *модификация образов* (для выбранного способа координации).

Например, может быть принято решение остановиться на способе «развязывания» взаимодействий (решение, касающееся управления в большом), однако при этом может оказаться, что глобальная эффективность работы системы будет все же неудовлетворительной. Тогда вышестоящий элемент может модифицировать функции качества для элементов нижестоящего уровня, например замесить функции вида

$$G_i = m_i^2 + y_i^2 + \beta_i m_i - \beta_i u_i$$

на функции

$$G_i = m_i^2 + y_i^2 + \beta_i m_i^2 - \beta_i u_i^2.$$

Аналогичным образом вышестоящий элемент может модифицировать образ, изменяя структуру моделей, используемых элементами нижестоящего уровня, или ограничения, налагаемые на принимаемые ими решения.

Наконец, после того как выбран способ координации и зафиксирована структура, перед вышестоящим элементом возникает задача управления в малом, связанная с выбором собственно координирующих воздействий. Для простоты мы будем называть эту задачу просто задачей координации, сознательно употребляя этот термин в узком смысле.

Таким образом, работа вышестоящего элемента сводится к следующему: а) выбор способа координации; б) модифицирование функций, определяющих стратегии нижестоящих элементов, если это необходимо, и в) выбор координирующих воздействий после того, как приняты остальные решения.

Проблемы координации, так же как и модификации, подробно исследуются в ч. II для различных способов координирования.

Сбор информации

Требуемая вышестоящему элементу информация о ниже расположенных уровнях зависит фактически от стоящей перед ним и требующей решения задачи, а также от образа (модели), используемого при решении этой задачи. Оба эти фактора, однако, тесно связаны друг с другом, ибо характер проблем принятия решений, которые могут возникать перед вышестоящим элементом, зависит от того, какого рода информацией он располагает. Для этого должен иметься образ (или модель) поведения элементов нижестоящего уровня. Мы упомянем три подхода к проблеме построения таких моделей.

При наиболее простом решении этой проблемы предполагается, что координатор имеет перед собой точное описание поведения элементов нижестоящего уровня. В таком случае задача координации сводится к классической задаче управления: имеются все необходимые данные и задача состоит лишь в том, чтобы принять

наилучшее решение. Прежде чем полностью отказаться от этого подхода, отметим, что в области социальных систем задача координации заключается не только в том, чтобы найти наилучшие условия координации, но и в том, чтобы указать способы их осуществления. Наиболее серьезные проблемы при координации социальных систем связаны главным образом с трудностями *реального осуществления* решения, которое оказалось бы приемлемым с технической, экономической и т. д. точек зрения. Отыскание в принципе приемлемого решения для таких ситуаций — только первый необходимый шаг в процессе координации.

Более содержательный подход к построению модели, используемой вышестоящим элементом, состоит в том, чтобы попытаться упрощенно описать подсистемы нижестоящего уровня. Это, по сути, классический подход, подробно изученный классической теорией управления. В теории больших систем такие приближенные модели можно получить посредством так называемого агрегирования переменных, так что в общем случае цель вышестоящего элемента не зависит от полного набора переменных, используемых на уровне нижестоящих элементов. На деле функция качества для вышестоящего элемента часто может оказаться вполне определенной функцией небольшого числа ключевых переменных, характеризующих эффективность работы элементов нижестоящего уровня. Тогда вышестоящему элементу нет нужды подробно знать поведение и динамику элементов нижестоящего уровня во времени. Успешность такого подхода очевидным образом зависит от того, какие переменные и в каком количестве входят в функцию качества для вышестоящего элемента. Для многоуровневых систем основной недостаток такого подхода заключается в том, что он не учитывает декомпозицию системы на подсистемы на нижнем уровне. Все же, как было показано [23], такой подход оказывается полезным и в применении к многоуровневым системам. Стимулом для некоторых из таких исследований послужили ранние работы области эконометрии. В данной книге подход этот подробно не рассматривается, поскольку, по сути дела, здесь речь идет не о многоуровневой проблеме, а скорее о применении более классических управленческих и программных подходов.

Выбор модели для вышестоящего элемента должен основываться не на прямом упрощении (путем агрегирования или иными средствами) нижестоящего уровня, а скорее на признании того факта, что для вышестоящего элемента управляемый процесс описывается как взаимодействие семейства взаимосвязанных подсистем, каждая из которых преследует собственные цели. Вот почему вышестоящий элемент должен скорее координировать деятельность нижестоящих элементов, нежели управлять ими.

В связи с этим модель для вышестоящего элемента должна основываться на взаимодействиях между нижестоящими элемен-

тами; точнее, на том, каким образом они при выборе своих решений учитывают взаимодействия друг с другом. На деле снова получается агрегирование. В этом случае, однако, каждый нижестоящий элемент агрегирует свои локальные переменные и переходит к новым переменным, существенным с точки зрения вышестоящего элемента. Например, при координации на основе метода согласования функций качества вышестоящий элемент нуждается в информации о локальных характеристиках; локальная функция качества, однако, есть не что иное, как агрегированная переменная. В любом случае именно признание за нижестоящими элементами права принимать решения своего уровня и вытекающая отсюда передача нижестоящим уровням значительной части работы по отысканию и принятию решений и составляют характерную особенность эффективных многоуровневых систем.

Взаимосвязь элементов нижестоящего и вышестоящего уровней

Элементы нижестоящего уровня могут влиять на действия вышестоящего элемента как непосредственно, снабжая его всей запрашиваемой им информацией, так и косвенно, посредством принимаемых ими решений, ибо конечный успех вышестоящего элемента зависит от того, как работают системы нижестоящего уровня. В ходе обмена информацией, имеющего место до принятия решения, вышестоящий элемент имеет превосходство над элементами нижестоящего уровня и может затребовать информацию нужного ему вида. Обычно эта информация касается того, какое решение собираются принять элементы нижестоящего уровня, т. е. она связана с оценкой процесса принятия решения нижестоящими элементами. Эти элементы, в свою очередь, могут использовать посылаемую наверх информацию как дополнительную переменную, определяющую выбор решения на нижестоящем уровне, с целью обеспечения для себя более выгодных условий. Однако, если процесс — динамический и если учитывается вмешательство после принятия решения, то элементы нижестоящего уровня должны принимать во внимание тот факт, что слишком большие расхождения между информацией, посылаемой до принятия решения, и следующим за этим реальным событием могут привести к нежелательной реакции со стороны вышестоящего элемента. Обычно в процессе обмена информацией, происходящего до принятия решения, нижестоящий элемент может сообщить вышестоящему элементу ту информацию, которая максимизирует его потенциальный выигрыш; но он должен при этом соблюдать «чувство меры», ибо от него могут потребовать разумного объяснения любого расхождения, возникающего в период после принятия решения.

4. ПОЧЕМУ ИМЕННО ИЕРАРХИЧЕСКИЕ СТРУКТУРЫ?

При обсуждении иерархических систем постоянно возникает вопрос, почему они так распространены в природе или почему необходимо при проектировании систем наделять их именно такой структурой. Очевидный недостаток многоуровневой системы состоит в сложности ее поведения и управления ею: функционирование такой системы нелегко проанализировать, ею трудно управлять и не всегда легко воздействовать на нее извне. Для проектирования новых систем на основе одноуровневого подхода обычно имеется определенная совокупность знаний и понятий, применимых на самых различных стадиях проектирования. Многоуровневые же системы требуют совершенно новых разработок. Но в таком случае возникает вопрос, дает ли использование многоуровневого подхода какие-то преимущества по сравнению с полностью интегрированным и централизованным подходом. На этот вопрос мы и пытались дать ответ (по крайней мере, в качественной форме) на протяжении данной главы. Имеет смысл выделить и подчеркнуть еще раз ряд доводов в пользу многоуровневого подхода.

Интеграция

Иерархическое упорядочение часто связано с процессом изменения структуры уже существующей системы в целях повышения эффективности ее работы. При создании объединенной (или «интегрированной») системы управления промышленным комплексом редко имеется возможность коренной перестройки и рационализации всего комплекса ввиду наличия ряда экономических, технических и социальных ограничений. По существу приходится исходить из имеющейся уже системы регулирования рабочих процессов и управления на нижнем уровне, добавляя к ней управление более высокого уровня и осуществляя тем самым интеграцию всего управления системой. В этом случае ситуация несколько напоминает раздельное проектирование технологического процесса и системы управления им. В настоящее время уже довольно ясно, что систему нужно проектировать как целое, а не начинать с процесса и затем просто добавлять необходимое управление. Несмотря на то что можно привести примеры, в которых при проектировании технологии процесса учитывается и наличие управляющих подсистем, общесистемный подход, не делающий никаких разделений, все еще не реализован.

Аналогично при проектировании интегрированной управляющей системы начинают с заданного процесса и управления на нижнем уровне с таким расчетом, чтобы координация взаимодействующих подсистем содействовала достижению целей более высокого уровня, охватывающих все более и более обширные части системы.

Стратификация

Описания или модели сложных систем часто могут быть получены лишь на основе стратификации с учетом физических подсистем, управленческих и экономических аспектов и т. п. Кроме того, глобальная задача, для осуществления которой создается система, может быть конкретизирована путем установления иерархии необходимых работ и подзадач.

Ограниченные возможности элементов (модулей) системы

Предположим, что подлежащая выполнению задача такова, что ее нельзя решить применением ни одного из имеющихся в наличии решающих элементов. В системах «линейного» функционирования¹⁾ в таких случаях образуют многоэшелонную иерархию, а в системах с «многофазным» принципом работы используют многослойный подход или декомпозицию многоэшелонного типа, в которой единственный решающий элемент используется для последовательного решения всех подзадач. Очевидным примером является задача оптимизации с помощью ЭВМ, объем памяти которой настолько ограничен, что программированию поддаются только полученные после декомпозиции подзадачи. Следует отметить, что многоуровневый подход к решению сложных задач, как правило, является важным методом в системотехнике. Исходя из сложной глобальной задачи, образуют иерархию подзадач. Затем по очереди решают подзадачи с использованием по возможности единственного решающего элемента. Дрю и др. [24] приводят интересные данные о применении этого подхода при разработке транспортных систем.

Лучшее использование имеющихся ресурсов

Имеющиеся ресурсы используются значительно лучше, если при решении сложных крупномасштабных задач применять многоуровневый подход. Разумеется, это положение не бесспорно, ибо в общем случае применимость такого подхода зависит от умения правильно подобрать многоуровневую структуру. Сравнительно простой анализ, однако, обнаруживает, что это можно сделать для широкого класса систем при условии, что удастся существенно упростить подлежащие решению задачи на верхнем уровне [25]. Анализ основывается на предположении, что оптимизируемая функция (затрата усилий на выработку решений) является выпуклой функцией числа управляющих переменных и линейной функцией числа наблюдаемых переменных. Начиная с некоторого момента (зависящего от конкретной формы выбранной функции для оценки усилий), затрата усилий существенно

¹⁾ См. также стр. 129.— *Прим. ред.*

ается путем декомпозиции задачи и применения двухуровневой структуры. Такой подход оправдан, если удастся упростить задачу координации до такой степени, чтобы она была значительно проще всей решаемой проблемы. В этом случае полная затрата усилий при использовании двухуровневой системы будет меньше, нежели при использовании интегрированной системы, обеспечивающей тот же уровень эффективности.

Адаптивность и надежность

В многоуровневой децентрализованной системе можно локализовать изменения в процедуре выработки решений, вызванные изменениями в протекании подпроцесса, и снизить тем самым затраты времени и средств. В общем случае система при этом быстрее адаптируется. Например, при распределении электроэнергии по многим областям, т. е. при решении задачи диспетчирования, изменения в одной из областей, вырабатывающих электроэнергию, повлекли бы за собой изменение матрицы потерь только для этой области; ничего другого в системе менять не нужно было бы. В то же время при централизованном распределении электроэнергии по всем областям нужно было бы заново рассчитывать всю матрицу потерь. Кроме того, неисправности в работе какой-то части системы при этом не столь быстро распространяются на всю систему. Последнее утверждение, конечно, требует уточнения, ибо на практике это зависит от конкретной системы и от типа возможной неисправности. Однако потенциальные возможности повышения надежности здесь, несомненно, имеются.

Цель данной главы — описать на языке математической теории систем ряд понятий, введенных в предыдущей главе в связи с изучением иерархий. Более строго говоря, мы попытаемся в этой главе на основе теории множеств формализовать различные интуитивные представления, касающиеся систем, подсистем и их взаимосвязей.

Основное содержание данной главы сводится к следующему:

1. Вводится определение функциональной системы как отображения $S: X \rightarrow Y$ абстрактного множества X в абстрактное множество Y , где X и Y представляют соответственно множество входов x и множество выходов y . В более общей формулировке система есть отношение $S \subseteq X \times Y$ над абстрактными множествами X и Y ¹⁾. Выражение системы в виде отношения над абстрактными множествами дает формализованное математическое представление интуитивного понятия системы.

2. Вводится формальное определение системы принятия решений (или «решающей системы»). В такой системе сигнал, приходящий на ее вход, конкретизирует «свободные» параметры решаемой проблемы (бывшие ранее неопределенными), и результатом работы системы является решение поставленной проблемы, получаемое на ее выходе.

3. Эти концепции используются для формального описания различных видов иерархий, обсуждавшихся в предыдущей главе.

Целесообразность предлагаемого нами определения системы как отношения, так же как и целесообразность введения понятия «решающей системы», лучше всего иллюстрируется примерами, подобными тем, которые разобраны в разд. 2 и 3 настоящей главы. Читатель, которому достаточно определений 1 и 2 и который интересуется только теорией координации, может пропустить эту главу и перейти сразу к следующей главе или к гл. 5.

¹⁾ Отношением над множествами X и Y называется произвольное подмножество декартова произведения $X \times Y$ (множества всех упорядоченных пар (x, y) , $x \in X$, $y \in Y$).

Отображение (функция) ставит в соответствие каждому элементу из X единственный элемент из Y . Отображение является, таким образом, частным случаем отношения. Семейство отображений, зависящих от параметра, также является отношением. Автор в дальнейшем пользуется тем, что отношение можно представить в виде семейства отображений, зависящих от параметра, хотя и не обосновывает этого утверждения. Значение параметра автор называет начальным состоянием или начальным условием. — *Прим. перев.*

1. ВВЕДЕНИЕ

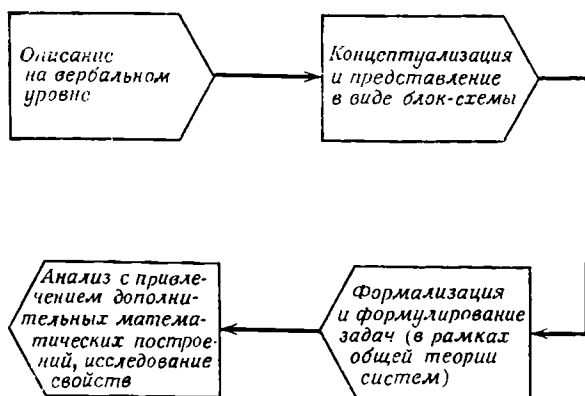
Существуют три основные причины, вызывающие необходимость абстрактной математической формализации многоуровневых систем.

1. *Строгость*. При такой формализации мы отвлекаемся от физической реализации систем. Используемый аппарат опирается лишь на структурные свойства систем и поэтому применим практически без ограничений ко всем возможным объектам, обладающим сходной структурой.

2. *Математическая теория*. Формальные понятия обеспечивают основу для более детального математического изучения иерархических систем, так как позволяют при необходимости вводить и более сложные математические построения.

3. *Исследования структуры*. Наш абстрактный математический подход является тем не менее содержательным, поскольку он позволяет выделить наиболее важные структурные особенности реальных систем. При непосредственном переходе от интуитивных понятий к сложным и детализированным математическим формулировкам (например, к системам, описываемым линейными дифференциальными уравнениями) неизбежно возникают вопросы: насколько те или иные частные выводы зависят от специфики математической модели, на основании которой они были получены? Как можно обобщить результаты? В каком направлении эти результаты следует обобщать? При исследовании той или иной проблемы методами общей теории систем ответ на вопрос, насколько общими являются полученные выводы, однозначно дается математической структурой, используемой для формализации и анализа проблемы.

Выбор «подходящего» уровня абстракции зависит до некоторой степени от планируемых применений, от точки зрения и даже от вкусов исследователя. Однако выбор высокого уровня абстракции имеет существенные преимущества. В качестве иллюстрации укажем на применение принципа координации на основе «согласования взаимодействий», описанного в гл. 6. Легко показать, что этот принцип приложим к двухуровневой системе управления стационарными линейными процессами, когда на нижестоящем уровне функции качества являются квадратичными, в то время как глобальная функция качества есть сумма функций качества для нижестоящего уровня. Однако можно ли применить этот принцип к линейным или нелинейным нестационарным процессам? Должны ли функции качества нижестоящего уровня обязательно быть квадратичными? Какие еще допустимы формы представления глобальной функции качества через функции качества нижестоящего уровня? Короче говоря, насколько общим является этот принцип?



Ф и г. 3.1. Основные этапы формулирования и анализа задач.

На все эти вопросы можно четко ответить, как показано в гл. 5, в изящной и очень простой форме, если подойти к ним с точки зрения общей теории систем. Анализ показывает, что единственным условием применимости этого принципа является требование монотонной зависимости между глобальной функцией качества и функциями качества нижестоящего уровня. Этот факт указывает на большую общность принципа и, следовательно, на возможность его широкого применения. Конечно, для более узких проблем, связанных, например, с координируемостью, рассматриваемой в гл. 5, нужны соответственно и более сложные математические построения, однако общность принципа согласования взаимодействий позволяет пользоваться им и при решении частных вопросов.

Другим примером, демонстрирующим преимущества общего подхода, служит исследование изменений характеристик нижестоящего уровня с помощью операторов оценки эффекта внутриуровневого взаимодействия, которые обычно приводят к нелинейностям.

Роль математической теории абстрактных систем в методологии системотехники схематически показана на фиг. 3.1. Вербальное описание системы позволяет построить ее блок-схему, показывающую взаимодействие подсистем, а также связи между ними. Блок-схему вместе с сопровождающим ее описанием можно затем формализовать и получить модель в виде абстрактной системы. Для такой модели уже легко построить математическое описание и изучать поведение системы аналитически или на ЭВМ. Здесь мы не будем детально рассматривать общую теорию систем, а введем только те понятия, которые нам понадобятся далее в настоящей работе.

ФОРМАЛЬНОЕ ОПРЕДЕЛЕНИЕ АБСТРАКТНОЙ СИСТЕМЫ

Введем прежде всего следующие исходные понятия:

1. *Системой (абстрактной) S* называется отношение над абстрактными множествами X и Y :

$$S \subseteq X \times Y.$$

2. Если S — функция, $S: X \rightarrow Y$ мы будем называть систему *функциональной*.

Для простоты мы будем писать просто «система» без указания на то, является ли она функциональной. Когда это свойство существенно или когда оно вытекает из контекста.

Входящие в определение системы множества X и Y характеризуют входные и выходные *объекты* и называются соответственно входным и выходным множествами, а их элементы — входами и выходами. Таким образом, представление системы в виде отношения есть представление в форме «вход — выход». Входы функциональной системы могут рассматриваться как причины, а выходы как следствия; в этом случае входное и выходное множества мы будем иногда называть множествами причин и следствий (объектами-причинами и объектами-следствиями). Эта терминология относится к моделированию явлений, содержащих причинно-следственные связи. Если система описывается отношением, а не однозначной функцией, причиной служит пара «вход — начальное состояние». В дальнейшем мы не будем, как правило, использовать отношения, а ограничимся рассмотрением функциональных подсистем даже при развитии общей теории. Мы будем употреблять термин «множество входных сигналов», а не «множество причин», подразумевая, что в рассматриваемой системе начальное состояние задано. Это едва ли ограничит общность результатов; в то же время использование функций вместо соответствующих отношений значительно упрощает рассуждения. Избегая длительного обсуждения вопроса о том, почему вводится именно такое понятие системы, мы приведем несколько примеров, показывающих, каким образом это понятие охватывает различные специальные случаи.

Пример 3.1. Система, описываемая разностным уравнением

Рассмотрим разностное уравнение

$$y_k = 2y_{k-1} + x_k, \quad (3.1)$$

описывающее некоторые наблюдения, которые проводятся в дискретные моменты времени $T = \{1, 2, \dots, n\}$. Для заданного начального условия $y_0 = \alpha$ каждому набору из n чисел $x = (x_1, x_2, \dots, x_n) \in R^n$ соответствует единственный набор $y = (y_1, y_2, \dots, y_n) \in R^n$, который удовлетворяет уравнению (3.1) для каждого $k = 1, \dots, n$. Таким образом, определено отобра-

жение $S_\alpha: R^n \rightarrow R^n$, такое, что для всех x из R^n образ $y = S_\alpha(x)$ является единственным решением уравнения (3.1) при заданном начальном условии $y_0 = \alpha$. Если допустимые начальные условия образуют множество $Y_0 \subseteq R$, мы получаем отношение $S \subseteq R^n \times R^n$, причем $S = \bigcup_{\alpha \in Y_0} S_\alpha$. Таким образом, приведенное выше уравнение описывает в общем случае систему $S \subseteq R^n \times R^n$ и, в частности, определяет функциональную систему S_α , когда задано начальное условие $y_0 = \alpha$.

Пример 3.2. Последовательностный автомат

Рассмотрим автомат для продажи кока-колы, в который можно опускать пяти- и десятицентовые монеты, причем стакан кока-колы стоит пятнадцать центов и автомат, когда это требуется, выдает сдачу. Введем следующие множества:

$A = \{5, 10\}$ — множество монет, которые принимает автомат;

$B_1 = \{\emptyset, \text{кока-кола}\}$, где $\emptyset$ означает «кока-колы на выходе нет»;

$B_2 = \{0, 5\}$ — множество монет, которыми автомат выдает сдачу.

Тогда множество выходов представляется декартовым произведением $B = B_1 \times B_2$.

Введем также множество $Q = \{q_0, q_1, q_2\}$ — множество «состояний» автомата. Теперь функцию переходов $f: A \times Q \rightarrow Q$ и функцию выходов $h: A \times Q \rightarrow B$ можно задать следующей таблицей:

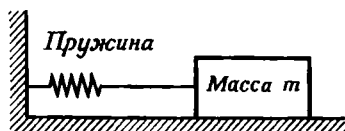
	$a = 5$	$a = 10$		$a = 5$	$a = 10$
$f(a, q_0)$	q_1	q_2	$h(a, q_0)$	$(\emptyset, 0)$	$(\emptyset, 0)$
$f(a, q_1)$	q_2	q_2	$h(a, q_1)$	$(\emptyset, 0)$	$(\text{Кока-кола}, 0)$
$f(a, q_2)$	q_0	q_0	$h(a, q_2)$	$(\text{Кока-кола}, 0)$	$(\text{Кока-кола}, 5)$

Рассмотрим случай, когда в автомат опускают подряд n монет. Пусть A^n и B^n обозначают множества наборов длины n из элементов множества A и B соответственно. Тогда легко видеть, что для заданного начального состояния $q = q_i$ каждому $x \in A^n$ соответствует единственный элемент $y \in B^n$. Другими словами, мы определили отображение $S_q: A^n \rightarrow B^n$, такое, что для всех x из A^n образ $y = S_q(x)$ является однозначно определенным выходом, зависящим от x и начального состояния $q = q_i$. Таким образом, данный автомат представляется системой $S \subseteq A^n \times B^n$, такой, что $S = \bigcup_{q \in Q} S_q$. Иногда мы можем получить стакан кока-колы за пять или десять центов, но если автомат исправен, его

начальным состоянием является q_0 и, следовательно, данный автомат может рассматриваться как функциональная система S_q , где $q = q_0$.

Пример 3.3. Система, описываемая дифференциальным уравнением

Рассмотрим простую динамическую систему:



Обозначим коэффициент упругости невесомой пружины через k , смещение тела массы m из положения равновесия в момент времени t — через $\dot{y}(t)$, а внешнюю силу, действующую на тело в момент времени t , через $x(t)$. Предположим, что трение отсутствует. Тогда связь между $x(t)$ и $y(t)$ задается следующим дифференциальным уравнением:

$$m\ddot{y}(t) = x(t) - ky(t). \quad (3.2)$$

Предположим, что мы наблюдаем $x(t)$ и $y(t)$ в интервале времени $T \in [0, \infty)$. Пусть X — множество всех интегрируемых вещественных функций, определенных на T , а Y — множество всех вещественных функций на T . Тогда для заданных начальных условий $\alpha = (y(0), \dot{y}(0))$ каждому $x \in X$ соответствует единственным образом определенное $y \in Y$ такое, что для каждого $t \in T$

$$y(t) = y(0) \cos \omega t + \dot{y}(0) \sin \omega t + \int_0^t \omega \sin \omega(t - \tau) x(\tau) d\tau, \quad (3.3)$$

где $\omega = (k/m)^{1/2}$. Таким образом, это уравнение описывает однозначное отображение $S_\alpha: X \rightarrow Y$.

Если множеством допустимых начальных условий является $A \subseteq R \times R$, то данная система представляется отношением

$$S = \bigcup_{\alpha \in A} S_\alpha \subseteq X \rightarrow Y.$$

Пример 3.4. Система, описываемая уравнением в частных производных

Рассмотрим одномерную задачу теплопроводности, соответствующую помещенному ниже рисунку:



В общем случае, если в пластине имеются источники тепла, функция распределения температуры $\theta(t, u)$ задается следующим уравнением в частных производных:

$$\frac{\partial \theta}{\partial t} = a \frac{\partial^2 \theta}{\partial u^2} + bz(t, u),$$

где $z(t, u)$ описывает источники тепла внутри пластины, а a и b — константы; ради простоты мы примем, однако, что $z \equiv 0$. Предположим также, что функция $\theta(0, u) = f(u)$ задана на отрезке $[0, \pi]$, а тепловой источник с температурой $x^\circ \text{C}$ начинает действовать у левого конца пластины в момент времени $t = 0$. Пусть Y — множество всех вещественных функций, определенных на $[0, \infty) \times [0, \pi]$. Если для начального распределения температуры f существует разложение в ряд Фурье, то для заданного распределения f каждому $x \in R$ соответствует единственное $\theta \in Y$. Например, когда $f = 0$,

$$\theta(t, u) = x - \frac{xu}{\pi} - \sum_{k=1}^{\infty} \frac{2x}{\pi k} \exp(-kt^2) \sin ku.$$

Эта зависимость представляет собой отображение $S_f: R \rightarrow Y$, где f — заданное начальное распределение. Обозначим через F множество вещественных функций, определенных на $[0, \pi]$, для которых существует разложение Фурье. Тогда отношение

$$S = \bigcup_{f \in F} S_f \subseteq R \times Y$$

описывает данную физическую систему.

3. СИСТЕМА ПРИНЯТИЯ РЕШЕНИЙ

Во всех примерах предыдущего раздела молчаливо предполагалось, что в любой системе существует «механизм», связывающий входы и выходы. Удобным методом описания причинно-следственной связи служат, например, система уравнений. Однако существование причинно-следственной связи, или, иначе говоря, системы, не предполагает в общем случае существования определяю-

щей процедуры или конструктивного описания, такого, какое дается системой уравнений. В действительности в весьма важном случае, когда изучается структурная взаимосвязь подсистем, их конструктивное описание несущественно. Этот случай является объектом рассмотрения общей теории многоуровневых систем, развиваемой ниже в этой работе. Кроме того, конструктивное описание системы может быть задано в неявном виде, т. е. отклик (выход) системы при данных начальных условиях задается как явная функция входа.

Особенно удобно определять систему с помощью задачи «принятия решения», а именно систему $S \subseteq X \times Y$ определяют так, что пара (x, y) принадлежит S тогда и только тогда, когда y является решением определенной задачи, конкретизация которой осуществляется посредством задания x . Более полное определение терминов «решаемая задача» и «проблема (задача) принятия решений» будет дано в этом разделе несколько позднее¹). Мы будем часто предполагать, что определенная таким образом система S является функциональной. В этом случае можно считать, что система содержит алгоритм решения, и мы будем называть ее «решающей» системой, «решателем» или *системой принятия решений*.

Пример 3.5. Оптимизирующая система

Возможно, наиболее удачным примером системы принятия решений («решающей» системы) служит система, которая действует таким образом, чтобы минимизировать значение некоторого заданного критерия качества функционирования G . Пусть X — множество непрерывных вещественных функций на отрезке $[0, 1]$. Пусть для каждого $x \in X$ задана задача оптимизации D_{x1} : минимизировать

$$G(m, g) = \int_0^1 [(y-x)^2 + cm^2] dt$$

при ограничениях

$$y' = ay + bm, \quad y(0) = \alpha,$$

где a, b и $c > 0$ — заданные константы, а m и y — функции, интегрируемые с квадратом на отрезке $[0, 1]$. Решение задачи D_x сводится к решению уравнений

$$\dot{y} = ay - \frac{b^2 p}{2c}, \quad y(0) \quad (3.4)$$

$$\dot{p} = -ap + 2(x - y), \quad p(1) = 0,$$

$$m = -\frac{bp}{2c} \quad (3.5)$$

¹) См. также примечание на стр. 117. — Прим. ред.

Очевидно, для каждого $x \in X$ существует единственный элемент $y \in L_2[0, 1]$, определяемый уравнениями (3.4) и (3.5), который дает решение задачи D_x . Таким образом, имеется система $S \subseteq X \times L_2[0, 1]$, такая, что для любых $x \in X$ и $y \in L_2[0, 1]$ пара (x, y) принадлежит системе S в том и только в том случае, если y представляет собой решение задачи D_x для некоторого начального условия $y(0) = \alpha$.

Пример 3.6. Принцип Гамильтона

Рассмотрим упругую систему, описанную в примере 3.3. Для каждой функции $x(t)$, задающей внешнюю силу на промежутке $T = [0, \infty)$, смещение y в соответствии с принципом Гамильтона определяется решением задачи D_x : найти y , удовлетворяющее условию

$$\delta \int_0^{\infty} \left[\frac{1}{2} m (\dot{y})^2 + \left(xy - \frac{1}{2} ky^2 \right) \right] dt = 0,$$

где δ обозначает первую вариацию. Решение этой задачи дается уравнением Эйлера

$$\frac{d}{dt} \left\{ \frac{\partial}{\partial \dot{y}} \left[\frac{1}{2} m (\dot{y})^2 \right] \right\} - \frac{\partial}{\partial y} \left\{ xy - \frac{1}{2} ky^2 \right\} = 0,$$

из которого следует уравнение движения, т. е. уравнение (3.2). Следовательно, решением задачи D_x является смещение $y(t)$, определяемое выражением (3.3). Иными словами, однозначное отображение $S_\alpha: X \rightarrow Y$ описываемое уравнением (3.3) и начальным условием α ($y(0), \dot{y}(0)$), таково, что для любого $x \in X$ образ $y = S_\alpha(x)$ служит решением задачи D_x . Очевидно, что система S в примере 3.3 может быть описана двумя различными, но эквивалентными способами: один способ обеспечивает конструктивное описание с помощью уравнения движения (3.2), а второй основан на принципе Гамильтона и состоит в том, что задается семейство подлежащих решению задач D_x , $x \in X$.

Чтобы определить решающую систему, постараемся сначала дать читателю более конкретное представление относительно «решаемой» задачи. Ниже рассматриваются два вида таких задач: задача оптимизации и задача нахождения удовлетворительных решений.

Задача оптимизации

Пусть $g: X \rightarrow V$ — функция, отображающая произвольное множество X в множество V , частично или полностью упорядоченное отношением $\leq$. Задача оптимизации состоит в следующем: дано подмножество $X^f \subseteq X$, требуется найти $\hat{x} \in X^f$, такое, что для всех x из X^f

$$g(\hat{x}) \leq g(x). \quad (3.6)$$

Множество X есть множество всех решений, а X^f — множество допустимых (feasible) решений¹⁾. Функция g называется целевой функцией, а множество V — множеством платежей. Задача оптимизации определяется парой (g, X^f) . Элемент x из X^f , удовлетворяющий соотношению (3.6) для всех x из X^f , называется решением задачи (g, X^f) .

Часто функция g задается двумя функциями $P: X \rightarrow Y$ и $G: X \times Y \rightarrow V$:

$$g(x) = G(x, P(x)).$$

В этом случае P называется выходной функцией, или моделью управляемого процесса, а G называется оценочной функцией, или функцией качества. Задача оптимизации может быть тогда определена тройкой (P, G, X^f) или парой (P, G) , если $X^f = X$. Название «модель управляемого процесса» подразумевает тот факт, что задача оптимизации, определенная тройкой (P, G, X^f) , является задачей управления, где принимаемые решения влияют на процесс, выход которого в свою очередь влияет на целевую функцию или платеж. Вопрос о связи модели P с реальным процессом выходит за рамки нашего рассмотрения. В общем случае само существование процесса может быть всего лишь допущением, необходимым для того, чтобы сформулировать задачу оптимизации и, таким образом, иметь возможность определять систему как некий «решатель» (задачи оптимизации) или как некоторую систему принятия решений.

Задача нахождения удовлетворительных решений

Пусть X и Ω — произвольные множества, а g — функция, отображающая $X \times \Omega$ в некоторое множество V , частично или полностью упорядоченное отношением $\leq$, и пусть τ — функция, переводящая Ω в V , т. е. $g: X \times \Omega \rightarrow V$ и $\tau: \Omega \rightarrow V$. Задача нахождения удовлетворительных решений заключается в следующем: дано подмножество $X^f \subseteq X$, требуется найти такое $\hat{x}$ из X^f , что для всех ω из Ω

$$g(\hat{x}, \omega) \leq \tau(\omega). \quad (3.7)$$

Множество Ω называется множеством неопределенностей, τ — функцией допустимости (tolerance), а неравенство (3.7) — критерием удовлетворительности. Остальные элементы задачи имеют тот же смысл, что и в задаче оптимизации. Четверка (g, τ, X^f, Ω) определяет задачу нахождения удовлетворительных решений, а любое $\hat{x}$ из X^f , для которого (3.7) выполняется при всех ω из Ω , является решением этой задачи.

¹⁾ Множество X^f называется также допустимой областью множества возможных решений.

Множество неопределенностей Ω иногда называют множеством возмущений, так как оно представляет собой множество всех факторов, которые могут повлиять на результирующие характеристики системы. Если целевая функция g задана в виде выходной функции $P: X \times \Omega \rightarrow Y$ и оценочной функции $G: X \times \Omega \times Y \rightarrow V$,

$$g(x, \omega) = G(x, \omega, P(x, \omega)),$$

то множество Ω является множеством всех возможных факторов, влияющих на получаемый результат решения $\hat{x}$. Заметим, что благодаря высокому уровню абстракции множество Ω охватывает и так называемые параметрические, и структурные неопределенности. Тогда функция τ определяет «максимально допустимое» значение оценочной функции. Решение $\hat{x}$ считается удовлетворительным, если оно приводит к значению оценочной функции, не превосходящему определенного уровня $\tau(\omega)$ при любых ω из множества неопределенностей Ω .

Пример 3.7.

Пусть функция $P: R \times \Omega \rightarrow R$ задается уравнением

$$y = m^3 + \omega \equiv P(m, \omega),$$

где Ω — множество $\{-1, +1\}$, состоящее только из двух элементов. Пусть функция допустимости $\tau: \Omega \rightarrow R$ — константа, $\tau(\omega) = 1$, для всех ω из Ω , а качество функционирования оценивается абсолютным значением выхода

$$G(m, y) = |y|.$$

Решением задачи нахождения удовлетворительных решений тогда будет управление $\hat{m}$ из R , такое, что

$$|P(\hat{m}, \omega)| \leq 1$$

для всех ω из Ω . Такое управление называется «удовлетворительным».

Очевидно, удовлетворительным управлением является $\hat{m} = 0$. Однако следует отметить, что удовлетворительное управление $\hat{m} = 0$ не будет оптимальным для некоторых возмущений. А именно: если $\omega = -1$, управлением, минимизирующим $|P(m, \omega)|$, будет $\hat{m}(-1) = -1$, в то время как для $\omega = +1$ оптимальное управление $\hat{m}(1) = 1$. Этот пример показывает, что решения задачи нахождения удовлетворительных решений могут быть не оптимальными для некоторых возмущений. Более того, ни одно из оптимальных решений не является удовлетворительным; для $m = 1$ получаем $|P(1, 1)| = 2$, что нарушает допустимый

предел, и аналогично для $m = -1$ имеем $|P(-1, -1)| = 2$, что также не удовлетворяет условию допустимости. На этом примере становится ясным различие между оптимальным и удовлетворительным решениями.

Пример 3.8. Распознавание сообщений при наличии помех в канале связи

Рассмотрим задачу выбора разделяющей (решающей) функции для задачи распознавания при наличии канала с помехами. Предположим, что $A = \{a_1, a_2\}$ — множество сообщений отправителя, а $B = \{b_1, b_2, b_3\}$ — множество сообщений, принимаемых получателем. Они не совпадают из-за наличия шумов в канале передачи. Тогда задача распознавания формулируется следующим образом: пусть принято сообщение b_i ; как определить, используя статистическую информацию о канале, какое сообщение, a_1 или a_2 , было передано с другого конца канала? Другими словами, задача состоит в том, чтобы выбрать функцию $m: B \rightarrow A$, такую, что если получено сообщение b_i , то за отправленное сообщение принимается $a_j = m(b_i)$.

Рассмотрим множество M всех функций, отображающих B в A . Необходимо описать критерии, помогающие выбрать из множества M ту функцию m , которую следует использовать в качестве разделяющей функции. Один из критериев можно сформулировать, если использовать байесовский подход.

Обозначим через $p[a_i]$ априорную вероятность того, что было отправлено сообщение a_i ; предположим, что $p[a_1] = \omega$, и, следовательно, $p[a_2] = 1 - \omega$. Пусть $p[b_i | a_j]$ — условная вероятность получения сообщения b_i при условии, что было отправлено сообщение a_j . Условные вероятности задаются следующей таблицей:

	$b = b_1$	$b = b_2$	$b = b_3$
$p[b a_1]$	2/3	1/6	1/6
$p[b a_2]$	1/4	1/4	1/2

Для удобства обозначим через $\bar{m}$ отрицание результата применения данной разделяющей функции m ; если $m(b_i) = a_1$, то $\bar{m}(b_i) = a_2$, и наоборот.

Вероятность вынесения неправильного решения равна $p[\bar{m}(b_i)]$. Используя теорему Байеса, получим, что условная

вероятность принятия неправильного решения, если принято сообщение b_i , дается формулой

$$p[\bar{m}(b_i) | b_i; \omega] = \frac{p[\bar{m}(b_i)] p[b_i | \bar{m}(b_i)]}{\omega p[b_i | a_1] + (1 - \omega) p[b_i | a_2]}$$

Теперь можно определить функцию потерь

$$g(m, \omega) = \sum_{i=1}^3 p[\bar{m}(b_i) | b_i; \omega] w(\bar{m}(b_i)),$$

где $w: A \rightarrow R$ означает штраф за ошибку. Проблема выбора разделяющей функции может быть тогда сформулирована следующим образом: задана априорная вероятность ω , требуется найти $\hat{m}$ из M , минимизирующую функцию потерь.

В случае когда $\omega = 1/2$ и функция штрафа w такова, что $w(a_1) = 1$ и $w(a_2) = 2$, оптимальная разделяющая функция $\hat{m}$ есть $\hat{m}(b_1) = a_1$, $\hat{m}(b_2) = a_2$, $\hat{m}(b_3) = a_2$. Допустим, что априорные вероятности точно не известны, но $\omega = p(a_1)$ лежит в известном диапазоне, например принадлежит отрезку $\Omega = (2/5, 3/5)$. Тогда, если задать уровень ϵ , где ϵ — постоянное число, можно представить выбор разделяющей функции в виде задачи нахождения удовлетворительных решений, а именно: найти такое m из M , что для всех ω из Ω

$$g(\hat{m}, \omega) \leq \epsilon.$$

Удовлетворительным решением для $\epsilon = 5/3$ тогда будет $\hat{m}(b_1) = a_1$, $\hat{m}(b_2) = a_2$, $\hat{m}(b_3) = a_2$. Заметим, что в данном случае удовлетворительное и оптимальное решения совпадают.

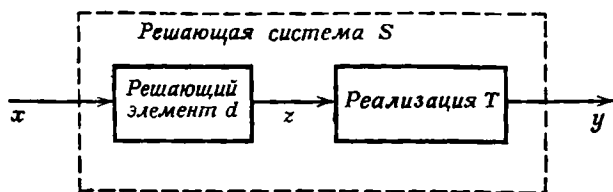
Когда мы рассматриваем задачу нахождения удовлетворительного решения в общем виде, отношение $\leq$ в (3.7) может быть заменено любым отношением вида $R \subseteq V \times V$. В общем случае задача нахождения удовлетворительных решений заключается в отыскании такого $\hat{x}$ из X^f , что для всех ω из Ω

$$g(\hat{x}, \omega) R \tau(\omega),$$

где R — некоторое заданное отношение на множестве платежей V ; иначе говоря, решение $\hat{x}$ из X^f удовлетворяет поставленной задаче, если для любого ω из Ω функция $g(x, \omega)$ связана отношением R с функцией допустимости $\tau(\omega)$.

«Решающая» система

Теперь нам нетрудно определить, что подразумевается под системой принятия решений (или «решающей» системой), блок-схема которой представлена на фиг. 3.2.



Ф и г. 3.2. Система принятия решений.

Система $S \subseteq X \times Y$ называется *решающей системой* (системой принятия решений), если задано семейство задач D_x , $x \in X$, с множеством решений Z и отображение $T: Z \rightarrow Y$. Для любого элемента x из X и любого y из Y пара (x, y) принадлежит системе S в том и только в том случае, если существует элемент $z \in Z$, такой, что он является решением задачи D_x и $T(z) = y$.

В большинстве случаев (хотя и отнюдь не всегда!), которые мы будем рассматривать, выход представляет собой решение поставленной задачи и $Z = Y$ т. е. T — тождественное преобразование.

В заключение сделаем следующие замечания, касающиеся решающей системы:

1) Иногда можно дать конструктивное описание такой системы в виде ряда уравнений. Такое описание особенно удобно, если уравнения имеют аналитическое решение, т. е. если можно указать алгоритм, который для любого входа x из X определяет выход $y = S(x)$. Однако такой алгоритм существует далеко не всегда. В действительности мы требуем только того, чтобы для всякой решающей системы была достаточно точно определена решаемая задача, но не требуем существования какого-либо алгоритма для нахождения решения этой задачи.

2) Всякая система, формализованная посредством моделей «вход — выход», может быть представлена в виде решающей системы, и наоборот. Например, в гл. 4 к двухуровневой системе будут применены оба эти подхода. Аналогично обстоят дело, например, в классической физике, где явление может быть описано и на языке законов движения, и на языке вариационных принципов.

Следует подчеркнуть, что системы, обладающие иерархической структурой, отличаются от всех прочих тем, что функции их подсистем наиболее естественно интерпретируются как поиск и принятие решений. Именно такими системами мы и будем в дальнейшем интересоваться.

3) Задача оптимизации является, очевидно, частным случаем задачи нахождения удовлетворительных решений и получается из последней, если определить $\tau(\omega)$ как минимум $g(x, \omega)$ на $X' \times$

$\times \{\omega\}$. В то же время решение любой задачи отыскания удовлетворительных решений может быть получено как решение задачи оптимизации с соответствующим образом выбранной целевой функцией. Какой формулировкой пользоваться — в значительной степени вопрос вкуса, но мы все-таки будем различать эти два типа задач в связи с некоторыми методологическими соображениями, на которых мы не станем здесь подробно останавливаться. В гл. 6 и 8, где рассматриваются задачи оптимизации и поиска удовлетворительных решений, станет ясным практическое различие между ними.

При обсуждении понятия иерархии в гл. 2 и в следующих разделах мы пользуемся такими понятиями, как цель, целенаправленная деятельность системы, поиск цели и целенаправленные системы. Эти термины, безусловно, тесно связаны с понятиями «поиск», «принятие решений», «решающие системы» и «системы принятия решений». Разграничение между процессом «принятия решения» и «целенаправленным поиском» зависит в основном от принятой точки зрения.

В общем случае мы не будем давать строгого определения понятиям «цель» и «целенаправленная деятельность» и поэтому воздержимся от формулировок, в которых смысл термина «цель» и способы, при помощи которых она может быть достигнута, точно не определены. Однако мы предполагаем, что состояние, в котором цель достигнута, может быть опознано по крайней мере самой системой.

Здесь уместно было бы привести пример из психологии. Для человека целью можно считать достижение «счастья», хотя, что это значит, знает, пожалуй, только он сам, а пути, ведущие к счастью, ему заранее не известны. Преследуя эту цель, он пробует стратегии, которые дают надежду на ее достижение. Он может попытаться получить хорошее образование, постараться разбогатеть, жениться или испробовать любую другую комбинацию стратегий, но ни одна из этих стратегий не гарантирует ему достижения цели. После нескольких попыток он может убедиться в тщетности своего «целенаправленного» поведения. Это не значит, что ситуации, связанные с целенаправленной деятельностью, в принципе не могут быть формализованы. Если такая формализация возможна, она неизбежно приведет к проблеме принятия решений. Более того, целенаправленное поведение и процесс целенаправленного поиска в сущности представляют собой последовательность принимаемых и осуществляемых решений. Переходя далее к «формализованным целям», мы определяем их через «решаемые задачи», которые в свою очередь могут быть сведены к задачам оптимизации или задачам поиска удовлетворительных решений. Поэтому, согласно нашей концепции, цель считается достигнутой, когда найдено решение соответствующей задачи.

4. ОСНОВНЫЕ ПОНЯТИЯ В ТЕОРИИ ИЕРАРХИЧЕСКИХ СИСТЕМ И ИХ ФОРМАЛИЗАЦИЯ

Здесь мы проведем формализацию более специальных понятий, введенных в гл. 1 и развернутых подробнее в гл. 2, а также сформулируем основные принципы описания стратифицированных систем, многослойных и многоэшелонных (организационных) иерархий.

Стратифицированные системы

Стратификация связана с тремя основными свойствами иерархических систем: вертикальной декомпозицией, приоритетом действий и взаимосвязью характеристик качества функционирования системы, описанных в гл. 1. Стратифицированная система изображена на фиг. 2.1.

Отправным пунктом для стратифицированного описания системы $S: X \rightarrow Y$ служит предположение о том, что множество внешних стимулов X и множество откликов Y представимы в виде декартовых произведений; а именно считаются заданными два семейства множеств: X_i , $1 \leq i \leq n$, Y_i , $1 \leq i \leq n$, таких, что

$$X = X_1 \times \dots \times X_n \quad \text{и} \quad Y = Y_1 \times \dots \times Y_n. \quad (3.8)$$

Это предположение означает возможность разбиения откликов и входных стимулов на компоненты.

Если множества X и Y представимы в виде (3.8), то каждая пара (X_i, Y_i) , $1 \leq i \leq n$, приписывается определенной страте. i -я *страта* системы S — это система, представленная как отображение S_i :

- 1) $S_i: X_i \times \mathcal{W}_i \rightarrow Y_i$, если $i = n$,
- 2) $S_i: X_i \times \mathcal{C}_i \times \mathcal{W}_i \rightarrow Y_i$, если $1 < i < n$,
- 3) $S_i: X_i \times \mathcal{C}_i \rightarrow Y_i$, если $i = 1$.

Семейство определенных таким образом систем S_i , $1 \leq i \leq n$, называется *стратификацией* S , если существуют два семейства отображений $h_i: Y_i \rightarrow \mathcal{W}_{i+1}$, $1 \leq i \leq n$, и $c_i: Y_i \rightarrow \mathcal{C}_{i-1}$, $1 \leq i \leq n$, такие, что для каждого элемента x из X и $y = S(x)$

- 1) $y_n = S_n(x_n, h_{n-1}(y_{n-1}))$,
- 2) $y_i = S_i(x_i, c_{i+1}(y_{i+1}), h_{i-1}(y_{i-1}))$,
- 3) $y_1 = S_1(x_1, c_2(y_2))$.

Множество Y_i состоит из откликов i -й страты. $\mathcal{C}_i$ и $\mathcal{W}_i$ представляют собой множества стимулов, исходящих от страт, примыкающих к i -й страте соответственно сверху и снизу. Отображения h_i и c_i называются соответственно *информационной функцией*

и *распределительной функцией* i -й страты; они связывают страты вместе, как, например, в (3.10), образуя систему S .

Рассматривая конкретные свойства отображений h_i и c_i , попытаемся формально определить, насколько «удачны» различные варианты разбиения системы на страты; при этом мы будем различать несколько степеней стратификации.

Система S называется *полностью стратифицированной*, если каждая ее страта S_i , $1 \leq i \leq n$, такова, что для любой пары (γ_i, w_i) из $\mathcal{C}_i \times \mathcal{W}_i$ и любых двух элементов x_i и x'_i из X_i

$$\begin{aligned} h_i(S_i(x_i, \gamma_i, w_i)) &= h_i(S_i(x'_i, \gamma_i, w_i)), \\ c_i(S_i(x_i, \gamma_i, w_i)) &= c_i(S_i(x'_i, \gamma_i, w_i)). \end{aligned} \quad (3.11)$$

Это означает, что для данного воздействия (вмешательства) γ_i и обратной связи w_i отклик подсистемы S_i на произвольное изменение стимула x_i будет таким, что w_{i+1} и γ_{i-1} не изменяются; другими словами, отклик не выходит за пределы i -й страты. Отметим, что полная стратификация зависит не только от преобразований S_i , но также и от отображений h_i и c_i . Иначе говоря, чтобы полностью описать такое разбиение, необходимо задать не только страты, но и взаимные связи между ними.

Требование полной локализации откликов каждой страты есть, несомненно, сильное условие. Более слабым является понятие *устойчивой стратификации*, при которой такая локализация имеет место не для всех, а лишь для некоторых пар «воздействие — обратная связь».

Следует отметить особое положение верхней страты. Она имеет собственное множество внешних стимулов X_n , а ее отклик зависит от всей иерархии, расположенной ниже. В этом случае требования, которые верхний уровень накладывает на нижние, формулируются на основе информации, поступающей по линиям обратной связи w_n , поскольку это единственный канал для сообщений, идущих снизу.

Таким образом, *устойчивую стратификацию* можно характеризовать следующим условием: для некоторых x из X , $y = \bar{S}(x)$ при всяком i , $1 \leq i \leq n$, существует пара (γ_i, w_i) , такая, что

$$\begin{aligned} w_i &= h_{i-1}(y_{i-1}), & 1 \leq i \leq n, \\ \gamma_i &= c_{i+1}(y_{i+1}), & 1 \leq i \leq n, \\ y_i &= S_i(x_i, \gamma_i, w_i), & 1 \leq i \leq n, \end{aligned}$$

и, кроме того, для всех x' из X при любом i , $1 \leq i \leq n$, имеют место равенства (3.11).

Различие между полной и устойчивой стратификациями заключается в том, что в определении последней не требуется, чтобы

страты были независимы для любой пары «воздействие — обратная связь»; необходимым считается только существование некоторых «состояний всей системы», для которых отклики оказываются локализованными в соответствующих стратах.

Разумеется, возникает вопрос — как достичь устойчивого состояния в иерархии, однако для анализа этого вопроса необходимо более подробное знание структуры системы.

Как полная, так и устойчивая стратификация представляют собой идеализированные модели, лишь приближенно отражающие структуру реальных систем. Можно разными способами смягчать условия, получая в результате не полностью стратифицированные системы. Мы не будем заниматься формализацией таких ослабленных условий, а ограничимся лишь несколькими пояснениями.

Равенство (3.11) может выполняться не для всех возмущений из X , а лишь для тех, которые соответствуют «нормальным» условиям работы системы. Для проведения устойчивой или даже полной стратификации при ограничениях, наложенных на стимулы, может возникнуть необходимость объединения нескольких соседних страт в одну. В некоторых случаях вполне может оказаться, что после такого объединения остается единственная страта и, таким образом, уничтожается сама стратификация.

Стратификация подразумевает сокращение объема информации, идущей вверх по иерархии: для вышерасположенных страт многие стимулы, поступающие от нижних страт, несут сходную информацию. Такое «сокращение объема информации» по мере продвижения вверх по иерархической лестнице страт имеет множество интересных следствий, одно из которых указывает на целесообразность введения многоэшелонной иерархии организационного типа.

Рассмотрим случай, когда влияние внешних стимулов имеет место только на нижнем уровне системы. В этом случае чем выше расположен уровень, тем меньший объем информации ему нужно обрабатывать. Отсюда вытекают два важных следствия.

1) Если система построена из блоков, обладающих ограниченной «решательной» способностью, то каждая страта будет состоять из тем меньшего количества таких блоков, чем выше она расположена.

2) Сократить объем информации можно многими способами. Одним из них является агрегирование (объединение). Как говорилось выше, агрегирование приводит к разбиению семейства переменных на такие подсемейства, каждое из которых описывается единственной «агрегированной» переменной. В действительности это означает разбиение нижней страты на подсистемы. Практически информационная обратная связь может быть успешно реализована через переменные, связанные с осуществлением взаимодействия между подсистемами, как, например, в случае координации

с помощью принципов прогнозирования и согласования взаимодействий.

Таким образом, отметив необходимость уменьшения объема информации от уровня к уровню, мы приходим естественным путем к горизонтальной декомпозиции страты на подсистемы. На каждой страте решающие элементы (блоки принятия решений) имеют дело в первую очередь с функционированием самих подсистем, пренебрегая, как правило, взаимодействием между ними. Напротив, решающие элементы более высокой страты в случае, когда подсистемы предшествующего уровня функционируют нормально, обрабатывают только информацию об их взаимосвязях и взаимодействиях. Эти рассуждения приводят нас к понятию многоэшелонной иерархии организационного типа.

Иерархия слоев

Иерархия слоев представляет собой совокупность вертикально расположенных решающих подсистем S_i , как показано на фиг. 3.3. Каждая из таких подсистем может быть, во-первых, описана как отображение $S_i: \mathcal{C}_i \rightarrow \mathcal{C}_{i-1}$ и, во-вторых, представлена в виде решающего элемента. А именно, заданы множество решаемых задач $\mathcal{D}_i(\gamma_i)$, $\gamma_i \in \mathcal{C}_i$, и преобразование T_i , такое, что для любого входа γ_i выход $\gamma_{i-1} = S_i(\gamma_i)$ определяется функцией $\gamma_{i-1} = T_i(x_i)$, где x_i — решение задачи $D_i(\gamma_i)$. Таким образом, входы γ_i выступают в качестве параметров (задаваемых непосредственно вышестоящим элементом), конкретизирующих решаемые задачи в S_i ; соответственно выходы γ_{i-1} , получающиеся после применения преобразования T_i , являются в свою очередь параметрами, задаваемыми непосредственно нижестоящему элементу.

Следует заметить, что многослойная иерархия показана на фиг. 3.3 упрощенно. Весьма важными оказываются следующие два аспекта:

1. Между слоями может существовать обратная связь, как постоянная, так и «временная», т. е. появляющаяся лишь при определенных обстоятельствах. Например, если какой-нибудь слой не решил свою задачу в заданное время, он посылает сигнал обратной связи на вышележащие слои и они определяют ему новую подзадачу. Различные варианты обратной связи весьма многочисленны, и поэтому на данном уровне общности мы не будем пытаться дать их формальное описание.

2. Многие слои могут быть подвержены (и притом одновременно) влиянию внешней среды. Выбор слоев, через которые будет осуществляться взаимодействие с внешней средой, зависит от типа решаемых ими задач и, конечно, от информации о среде, которая может им понадобиться. Это особенно хорошо видно на примере функциональной иерархии решений. Функциональная иерархия

решений, описанная в гл. 1, является весьма важной и поэтому заслуживает более детального рассмотрения. Отправной точкой здесь служит общая проблема принятия решения в условиях неопределенности, сформулированная как проблема нахождения удовлетворительных решений в виде четверки (g, τ, X^f, Ω) . Иначе говоря, требуется найти x из X^f , такое, что для всех ω из Ω

$$g(x, \omega) \leq \tau(\omega),$$

где $\leq$ означает заданное отношение. Как уже говорилось, эта задача приводит к трем функциональным слоям (см. фиг. 2.6), каждый из которых может быть описан как отображение, хотя в более общем случае они представляют собой соответствующие отношения. Представим первый слой, *слой выбора*, отображением

$$S_1: \mathcal{W}_1 \times \mathcal{C}_1 \times \mathcal{U} \rightarrow M,$$

где элементы множества $\mathcal{W}_1$ соответствуют сигналам обратной связи, которые поступают от управляемого объекта (или, может быть, от окружающей среды). Элементами множества $\mathcal{C}_1$ являются сигналы (входы), приходящие с третьего уровня; они определяют структуру слоя S_1 . Сигналы (входы) со второго уровня, образующие множество $\mathcal{U}$, конкретизируют для первого слоя множество неопределенностей. Иными словами, с точки зрения задачи нахождения удовлетворительных решений элемент из $\mathcal{C}_1$ задает первые три элемента этой задачи g, τ, X^f , а элементы из $\mathcal{U}$ определяют последний элемент Ω рассматриваемой задачи, полностью описываемой четверкой (g, τ, X^f, Ω) .

Второй слой, называемый *слоем обучения*, представляется отображением

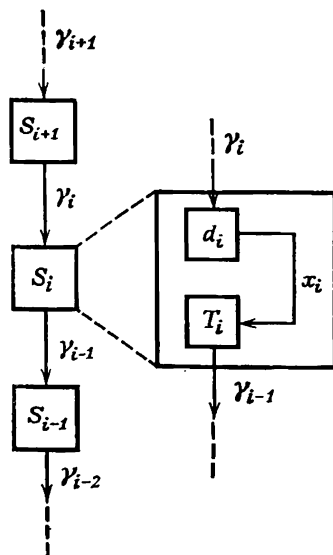
$$S_2: \mathcal{W}_2 \times \mathcal{C}_2 \rightarrow \mathcal{U},$$

где элементы $\mathcal{W}_2$ представляют собой информацию об окружающей среде, $\mathcal{U}$ задает множество неопределенностей для первого слоя, а элементы $\mathcal{C}_2$ являются параметрами, определяющими структуру слоя обучения S_2 , так же как элементы $\mathcal{C}_1$ определяют структуру слоя выбора S_1 .

Третий слой, *слой самоорганизации*, описывается отображением

$$S_3: \mathcal{W}_3 \rightarrow \mathcal{C}_1 \times \mathcal{C}_2,$$

где элементы $\mathcal{W}_3$ представляют собой информацию, поступающую на этот слой через каналы обратной связи.



Ф и г. 3.3. Иерархия слоев.

Многоэшелонная (организационная) иерархия

Особенностью формального описания организационной иерархии является необходимость более точного определения взаимодействия подсистем по вертикали. В иерархии, составленной из страт или слоев, на каждом уровне формально находится один элемент. В эшелонной же иерархии на данном уровне, как правило, располагается несколько элементов. В этом случае становится особенно важным правильное взаимное расположение элементов системы в соответствии с приоритетом действия.

Если $\mathcal{S}$ — (конечное) семейство систем S_i , $i \in I$, где I — конечное множество значений индекса i , а $>$ частично (но строго) упорядочивающее отношение в I , тогда $(\mathcal{S}, >)$ называется *иерархией систем*. Если же $(\mathcal{S}, >)$ — иерархия решающих систем (систем принятия решений), а отношение $>$ таково, что $i > j$, если S_i имеет приоритет действия по отношению к S_j , то $(\mathcal{S}, >)$ называется *иерархической схемой* или просто *иерархией принятия решений*. Выделение эшелонов в иерархии принятия решений легко осуществляется с помощью строгого (частичного) отношения порядка $>$, описывающего приоритет действия следующим образом: первый эшелон составляют минимальные элементы из $\mathcal{S}$; семейство

$$\mathcal{S}^1 = \{S_i: i \in I_1 \subset I\}$$

называется *первым эшелоном*, если $I_1 = \{i: i \text{ — минимальный элемент в } I\}$. Аналогично множество $\mathcal{S}^k = \{S_i: i \in I_k \subset I\}$ называется *k-м эшелоном*, если $I_k = \{i: i \text{ — минимальный элемент множества } I - [I_1 \cup I_2 \cup \dots \cup I_{k-1}]\}$.

Если каждый эшелон содержит не более одного элемента, мы имеем многослойную иерархию при условии, что упорядочение с помощью отношения $>$ определено надлежащим образом.

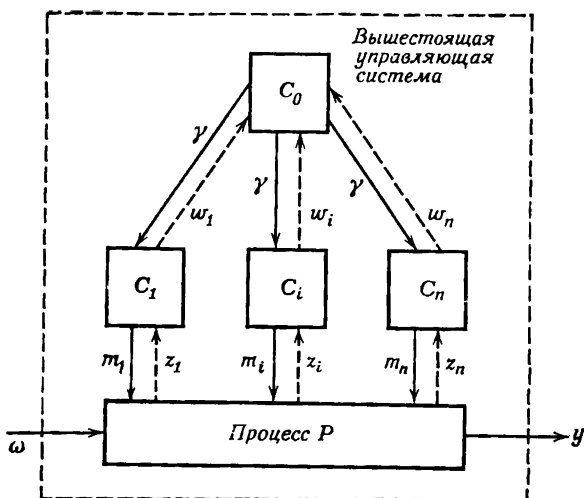
Наконец, мы можем определить многоэшелонные иерархии как подкласс иерархий принятия решений. Иерархия принятия решений $(\mathcal{S}, >)$ является *многоэшелонной иерархией*, если для любых i и j из I существует не более одного $k \in I$, такого, что для любого l из I соотношения $l > i$ и $l > j$ влекут за собой $l > k$. Это условие означает, что для любого члена иерархии $\mathcal{S}$ в эшелоне, расположенном непосредственно над ним, найдется по крайней мере один элемент, обладающий приоритетом действия по отношению к нему. Многоэшелонную иерархию можно интерпретировать следующим, весьма интересным образом. Если отношение $>$ таково, что $i > j$ тогда и только тогда, когда S_j является подсистемой S_i , мы получаем стратифицированную систему в том смысле, что системы нижнего уровня являются подсистемами систем, расположенных на вышележащих уровнях.

Проблему координации в многоуровневой системе с достаточной общностью можно рассмотреть на примере двухуровневой системы. В настоящей главе мы намерены дать такую постановку проблемы координации, которая поддавалась бы анализу математическими средствами. Мы остановили свой выбор именно на двухуровневых системах потому, что, во-первых, такие системы сравнительно просты, а во-вторых, при синтезе более общей многоуровневой системы их можно использовать в качестве основных модулей.

Вначале мы дадим общесистемное описание двухуровневой системы, имеющей n управляющих систем, подчиненных одной вышестоящей системе управления. Мы рассматриваем каждый решающий элемент как систему, выходом которой является если не оптимальное, то удовлетворительное решение стоящей перед данным элементом задачи, тогда как управляющую или «решающую» систему мы рассматриваем как систему, которая содержит в себе решающий элемент и средства осуществления получаемых решений. Применительно к такой структуре системы мы и поставим проблему координации, а в качестве основы для решения проблемы координации введем «постулат совместимости» и принципы координации. Излагаемый ниже материал является в основном описательным, анализ же с помощью математических средств будет дан в последующих главах.

1. ОБЩЕЕ ОПИСАНИЕ ДВУХУРОВНЕВОЙ СИСТЕМЫ

На фиг. 4.1 представлена блок-схема некоторой двухуровневой системы. Отдельные блоки изображают подсистемы, а их взаимное расположение отражает иерархическую структуру всей системы. Система, изображенная на фиг. 4.1, имеет $(n + 2)$ основных подсистем: вышестоящую управляющую систему C_0 , n нижестоящих управляющих систем $C_1, \dots, C_n$ и управляемый процесс P . Отметим два вида вертикального взаимодействия между подсистемами. Один — это передача вниз «командных» сигналов; сигналы от нижестоящих управляющих систем $C_1, \dots, C_n$ к процессу будут называться управляющими воздействиями (входами), тогда как сигналы от вышестоящей к нижестоящим управляющим системам будут называться координирующими сигналами (входами) или вмешательствами. Другой вид вертикального взаимодействия — это передача наверх информационных сигналов, или сигналов обратной связи, различным управляющим системам иерар-



Фиг. 4.1. Двухуровневая система с n нижестоящими управляющими системами и единственной вышестоящей управляющей системой.

хий. Эти передачи сигналов представлены на блок-схеме пунктирными линиями.

Простейший способ описания подсистем двухуровневой системы связан с использованием терминальных переменных: входов и выходов. При этом удобно описывать подсистемы как функциональные в том смысле, что входы однозначно определяют выходы; можно рассматривать эту ситуацию как ситуацию, в которой задано текущее состояние. Поэтому каждый из блоков на фиг. 4.1 представляет собой отображение. Когда мы будем описывать подсистемы, введем соответствующие названия для различных объектов, чтобы подчеркнуть концептуальную роль каждого из них в функционировании двухуровневой системы.

Рассмотрим сначала процесс P , как некую управляемую систему, к которой поступают управляющие воздействия от систем управления нижнего уровня $C_1, \dots, C_n$. К нему приходят входные сигналы двух видов: управляющие сигналы (управляющие входы) $m, m \in M$, где M называется множеством управляющих сигналов, и сигналы (входы) $\omega, \omega \in \Omega$, представляющие собой внешние возмущения, поступающие из окружающей среды. Символом $y, y \in Y$ мы будем обозначать «выход» процесса P и соответственно будем называть множество Y множеством выходов процесса P .

Будем представлять процесс P в виде отображения

$$P: M \times \Omega \rightarrow Y.$$

Поскольку имеется n нижестоящих (локальных) управляющих систем ¹⁾ $C_1, \dots, C_n$, удобно представить множество управляющих сигналов M для процесса P в виде декартова произведения n множеств

$$M = M_1 \times \dots \times M_n,$$

причем i -я локальная управляющая система C_i имеет полномочия выбирать i -ю компоненту m_i управляющего сигнала m , оказывая тем самым соответствующее воздействие на процесс.

Рассмотрим далее i -ю локальную систему управления C_i . Здесь мы будем считать ее просто системой вход — выход. К системе C_i также поступают входные сигналы двух видов: координирующий сигнал γ , $\gamma \in \mathcal{E}$, поступающий от вышестоящей управляющей системы, и информационный сигнал z_i (сигнал обратной связи), поступающий от процесса. Выходом C_i является (локальное) управление m_i , выбираемое из множества M_i . Будем считать, что с помощью рассматриваемой системы реализуется отображение

$$C_i: \mathcal{E} \times \mathcal{Z}_i \rightarrow M_i,$$

где $\mathcal{Z}_i$ — множество информационных сигналов (сигналов обратной связи) Z_i , $z_i \in \mathcal{Z}_i$.

Мы будем обращаться в дальнейшем к множеству $\mathcal{E}$, называя его *множеством координирующих сигналов*, а его элементы γ — соответственно *координирующими сигналами*, так как с помощью этих сигналов управляющая система C_0 воздействует на нижестоящие, локальные управляющие системы $C_1, C_2, \dots, C_n$. Поскольку каждая нижестоящая управляющая система могла бы интерпретировать поступивший координирующий сигнал γ своим особым образом, мы, чтобы избежать этого, будем считать координирующие сигналы γ из $\mathcal{E}$ n -мерными векторами $(\gamma_1, \dots, \gamma_n)$, так что на вход i -й управляющей системы поступает только i -я компонента γ_i .

Управляющую систему C_0 будем также называть *координатором*, так как ее выходные сигналы γ , $\gamma \in \mathcal{E}$ являются координирующими сигналами для систем $C_1, \dots, C_n$. Мы будем рассматривать только один вход для системы C_0 — информацию w , получаемую посредством обратной связи от нижестоящих управляющих систем и используемую для формирования координирующих воздействий (координирующих сигналов) γ . В таком случае мы вправе считать, что управляющая система C_0 по сути дела осуществляет отображение

$$C_0: \mathcal{W} \rightarrow \mathcal{E},$$

¹⁾ Здесь и далее для двухуровневых систем управления часто наряду с термином «*нижестоящие управляющие системы*» используется его синоним «*локальные управляющие системы*». — *Прим. ред.*

где $\mathcal{W}$ представляет собой множество информационных сигналов w , с помощью которых реализуется обратная связь.

Для того чтобы завершить описание двухуровневой системы, мы должны уточнить характер информации, поступающей по каналам обратной связи. Сигналы обратной связи z_i , поступающие на вход локальной управляющей системы C_i , содержат информацию относительно поведения процесса P ; поэтому мы предположим, что они связаны функциональной зависимостью с управляющим сигналом m , внешним возмущением ω и выходом y . Эту зависимость мы будем представлять в виде отображения

$$f_i: M \times \Omega \times Y \rightarrow \mathcal{Z}_i.$$

Аналогично поступающий по каналам обратной связи информационный сигнал w направляется в вышестоящую управляющую систему C_0 и содержит в себе информацию относительно поведения нижестоящих управляющих систем; поэтому он по определению задается отображением

$$f_0: \mathcal{C} \times \mathcal{Z} \times M \rightarrow \mathcal{W},$$

где $\mathcal{Z} = \mathcal{Z}_1 \times \dots \times \mathcal{Z}_n$; $\mathcal{W}$ является функцией координирующего сигнала γ , информационных сигналов обратной связи $z = (z_1, \dots, z_n)$, получаемых нижестоящими управляющими системами, и их управляющих воздействий $m = (m_1, \dots, m_n)$. На фиг. 4.1 информация, поступающая по каналам обратной связи, представлена совокупностью w информационных сигналов $w = (w_1, \dots, w_n)$, где w_i — информационный сигнал обратной связи, поступающий от управляющей системы C_i .

Относительно указанных взаимосвязей между подсистемами следует сделать два замечания.

1. В явном виде не предусматривается прямая коммуникация между нижестоящими управляющими системами. В этом находит отражение тот факт, что мы прежде всего интересуемся только отношениями между смежными уровнями иерархии. Тем не менее некоторые из полученных в последующих главах результатов можно интерпретировать, исходя из коммуникаций между нижестоящими системами через посредство вышестоящей системы. В таких случаях вышестоящая управляющая система действует как передатчик информации между нижестоящими системами.

2. Вышестоящая управляющая система непосредственно не взаимодействует с процессом. Впрочем, это только видимость, так как на самом деле любая нижестоящая управляющая система может передать вышестоящей всю информацию о ходе протекания процесса. В последующих главах мы рассмотрим случай, в котором вышестоящая система получает информацию не только от нижестоящих управляющих систем, но также и непосредственно прямую информацию о ходе протекания процесса.

2. ДЕКОМПОЗИЦИЯ ПОДСИСТЕМ

Для каждой из $(n + 2)$ подсистем двухуровневой системы, показанных на фиг. 4.1, можно произвести дальнейшую декомпозицию. Наиболее важна из них декомпозиция процесса P . Что касается отдельных управляющих систем, то они нуждаются в декомпозиции только в том случае, если их выходными результатами являются уже не сами решения стоящих перед ними задач, а преобразования получаемых решений.

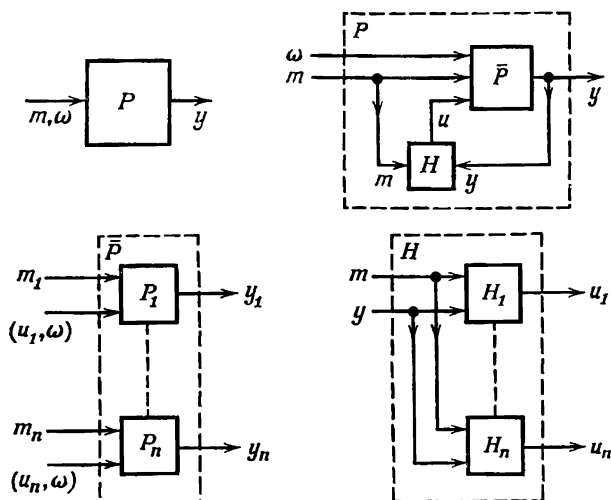
Подпроцессы

Процесс P является первопричиной взаимодействия между нижестоящими управляющими системами C_i , и именно он вызывает необходимость введения координатора, т. е. вышестоящей управляющей системы. Процесс P , показанный на фиг. 4.2, можно рассматривать как состоящий из n подпроцессов, каждый из которых управляется отдельной управляющей системой.

Предположим, что задано n подпроцессов, таких, что каждый i -й подпроцесс есть отображение

$$P_i: M_i \times U_i \times \Omega \rightarrow Y_i,$$

где U_i — множество (входных) сигналов u_i , посредством которых подпроцесс P_i связывается с другими подпроцессами. Формально можно представить себе, что на каждый подпроцесс воздействует одно и то же внешнее возмущение ω из Ω ; однако влияние одного



Фиг. 4.2. Декомпозиция процесса P .

и того же внешнего возмущения может по-разному сказаться на каждом из подпроцессов; в самом деле, внешние возмущения ω из Ω могут быть n -компонентными наборами $(\omega_1, \dots, \omega_n)$, так что на i -й подпроцесс воздействует только i -я компонента ω .

Для каждого i , $1 \leq i \leq n$, мы предполагаем, что задано отображение

$$H_i: M \times Y \rightarrow U_i,$$

которое связывает подпроцессы. Часто (хотя и не всегда) H_i является просто проекционным отображением.

Мы будем называть множества U_i *множествами связующих сигналов*, а их элементы — *связующими сигналами (входами)* (interface inputs). Отображения H_i будем называть *связующими функциями подпроцессов*.

Соотношение между процессом P и его подпроцессом P_i выглядит следующим образом. Положим $U = U_1 \times \dots \times U_n$ и определим функции H на множестве $M \times Y$ и $\bar{P}$ на множестве $M \times U \times \Omega$ в виде

$$H(m, y) = (H_1(m, y), \dots, H_n(m, y)), \quad (4.1)$$

$$\bar{P}(m, u, \omega) = (P_1(m_1, u_1, \omega), \dots, P_n(m_n, u_n, \omega)). \quad (4.2)$$

В этом случае компонентами $\bar{P}$ являются не связанные между собой подпроцессы, в то время как с помощью H осуществляется их соединение. Процесс P состоит из связанных между собой подпроцессов, если условие

$$y = \bar{P}(m, H(m, y), \omega) \iff y = P(m, \omega) \quad (4.3)$$

выполняется для всех (m, y, ω) в $M \times Y \times \Omega$; т. е. существует решение системы уравнений

$$\begin{aligned} y &= \bar{P}(m, u, \omega), \\ u &= H(m, y) \end{aligned} \quad (4.4)$$

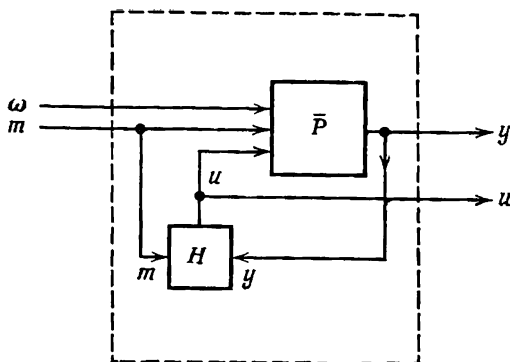
для любого заданного управляющего воздействия m из M и возмущающего воздействия ω из Ω и дает выход $y = P(m, \omega)$. На фиг. 4.3 показано это соотношение между процессом P и подпроцессами P_i , представленными в совокупности блоком $\bar{P}$.

Заметим, что из условия (4.3) следует, что связующие сигналы $u = (u_1, \dots, u_n)$, поступающие на входы подпроцессов, могут быть функционально связаны с управляющими воздействиями m и внешними возмущениями ω . Точнее, u является результатом отображения

$$K: M \times \Omega \rightarrow U,$$

которое в свою очередь определяется уравнением

$$K(m, \omega) = H(m, P(m, \omega)). \quad (4.5)$$



Фиг. 4.3. Взаимосвязь между процессом P и «развязанными» подпроцессами, представленными блоком $\bar{P}$.

Мы будем впредь называть K *функцией взаимодействия подпроцессов*. Таким образом, мы можем считать, что процесс P определяется через подпроцессы, а отображение K — с помощью соотношения

$$P(t, \omega) = \bar{P}(t, K(t, \omega), \omega), \quad (4.6)$$

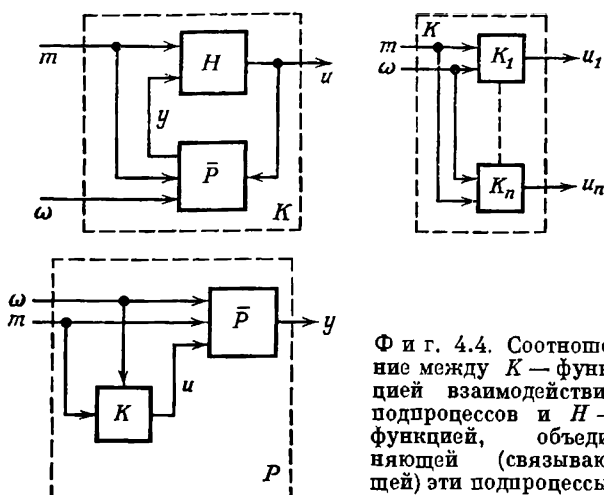
как показано на фиг. 4.4.

Сделаем несколько замечаний относительно процесса и его представления через подпроцессы (или декомпозиции на подпроцессы), важных для последующего рассмотрения понятия координации.

1. Каждая локальная управляющая система $C_1, \dots, C_n$ заинтересована главным образом в каком-нибудь одном аспекте процесса, хотя окончательный результат ее действий зависит от всего процесса. Имея в виду этот «локальный» интерес, мы свяжем каждую i -ю локальную управляющую систему C_i с i -ми компонентами управляющего воздействия t и выхода y ; т. е. i -я локальная управляющая система C_i в первую очередь интересуется связью между управляющим воздействием t_i и выходом y_i , являющимся результатом осуществления i -го подпроцесса P_i .

2. Связующие функции подпроцессов H_i предопределяют характер декомпозиции процесса, и обычно их следует выбирать по возможности более простыми. В большинстве случаев связующие функции H_i будут проекционными отображениями: связующие сигналы u_i будут образованы компонентами терминальных переменных процесса t и y . Например, если $t = (t_1, t_2, t_3)$ и $y = (y_1, y_2, y_3)$, то связующий сигнал u_1 может быть, например, парой (t_2, y_3) .

3. Функция взаимодействия K отражает весь процесс P , так как для любого управляющего сигнала t и возмущающего воздействия ω K определяет (поскольку $K(t, \omega) = u$) связующие сиг-



Ф и г. 4.4. Соотношение между K — функцией взаимодействия подпроцессов и H — функцией, объединяющей (связывающей) эти подпроцессы.

налы, которые поступят на вход подпроцессов P_i , и, кроме того, $u = H(m, P(m, \omega))$. K может также рассматриваться как отображение подпроцесса, который порождает взаимодействия подпроцессов P_i . В нашем рассмотрении несущественно, использовать ли K или сами связующие функции подпроцессов H_i ; но все-таки удобнее использовать функцию взаимодействия K или ее компоненты $K_i: M \times \Omega \rightarrow U_i$, а не функции H_i .

Управляющие подсистемы

Как уже указывалось в гл. 3, мы будем рассматривать управляющую систему как систему, составленную из решающих элементов и реализаторов, связанных каскадно. Функции реализатора в нашем представлении сводятся просто к модификации данных, получаемых на выходе решающего элемента, чтобы сделать их более приемлемыми для использования в другой системе.

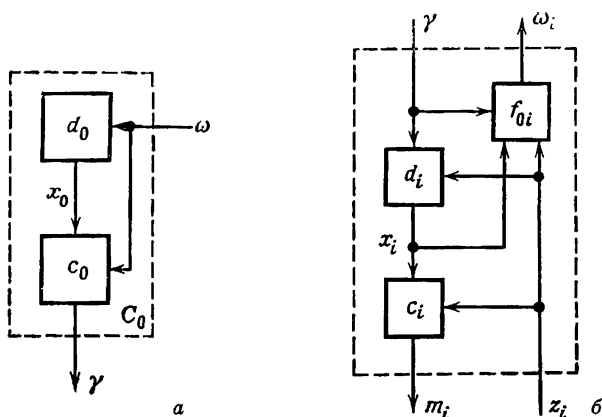
Декомпозиция вышестоящей управляющей системы $C_0: \mathcal{W} \rightarrow \mathcal{C}$ на решающий элемент d_0

$$d_0: \mathcal{W} \rightarrow X_0$$

и реализатор c_0

$$c_0: \mathcal{W} \times X_0 \rightarrow \mathcal{C}$$

показана на фиг. 4.5, а. Предполагается, что существует связанное с управляющей системой C_0 семейство $D_0(w)$, $w \in \mathcal{W}$, решаемых задач и соответствующее множество решений X_0 , таких, что для любого w из $\mathcal{W}$ выход $x_0 = d_0(w)$ есть решение задачи



Ф и г. 4.5. Управляющие системы с двухуровневой иерархией.

$\mathcal{D}_0(w)^1$). Допускается, что выход блока c_0 зависит как от $d_0(w)$, так и от w . Следовательно, при получении по каналу обратной связи любой конкретной информации w из $\mathcal{W}$ координирующий сигнал, исходящий от управляющей системы c_0 , есть

$$\gamma = c_0(w, d_0(w)) = C_0(w).$$

¹⁾ Здесь необходимо сделать одно терминологическое замечание. В оригинале авторы постоянно используют термины «decision unit» и «decision problem», которые в русском тексте переведены соответственно как «решающий элемент» и «решаемая задача». В английском языке имеются два термина: «decision» и «solution», для которых на русском языке обычно используется одно и то же слово — «решение», хотя эти два термина следовало бы различать, ибо они имеют существенно различный смысл. Под «decision» обычно понимается «принятие решений» в смысле выбора (может быть, «волевого» или на основе каких-то субъективных предпочтений) одной из имеющихся альтернатив, тогда как слово «solution» используется для обозначения некоторого конкретного решения (большой частью, единственного) вполне определенной задачи. Таким образом, более близкий к оригиналу перевод термина «decision problem» был бы таков: «задача (проблема) принятия решения». Однако для краткости и учитывая наличие в книге формального определения этой задачи, при переводе, как правило, используется выражение «решаемая задача», причем подразумевается, что ее решения в большинстве случаев неединственны и решающий элемент при ее решении имеет определенную свободу выбора «более предпочтительных» решений (из числа «допустимых»). В этом смысле «решаемая задача» есть одновременно и «задача принятия решения». Надо сказать, что на протяжении всей книги авторы акцентируют свое внимание не на трудностях отыскания решающими элементами «допустимых решений», а именно на вопросах постановки задач перед нижестоящими решающими элементами таким образом, чтобы эти элементы, отыскивая наиболее предпочтительные (с их собственной точки зрения) решения поставленных перед ними задач, естественным образом обеспечивали получение нужных вышестоящему элементу решений. — *Прим. ред.*

Рассмотрим теперь i -ю нижестоящую управляющую систему $C_i: \mathcal{C} \times \mathcal{Z}_i \rightarrow M_i$. Декомпозиция C_i на решающий элемент d_i

$$d_i: \mathcal{C} \times \mathcal{Z}_i \rightarrow X_i,$$

и последовательно соединенный с ним реализатор c_i

$$c_i: \mathcal{Z}_i \times X_i \rightarrow M_i,$$

аналогична декомпозиции вышестоящей управляющей системы и показана на фиг. 4.5, б. Предполагается, что с каждым решающим элементом d_i связано семейство задач $\mathcal{D}_i(\gamma, z_i)$, где $\gamma \in \mathcal{C}$ и $z_i \in \mathcal{Z}_i$, с множеством решений X_i , таких, что для каждой пары (γ, z_i) из $\mathcal{C} \times \mathcal{Z}_i$ выход $x_i = d_i(\gamma, z_i)$ есть решение задачи $\mathcal{D}_i(\gamma, z_i)$. В большей части наших рассмотрений, за исключением нескольких случаев в гл. 8, решаемые задачи $\mathcal{D}_i(\gamma, z_i)$ будут зависеть только от координирующего сигнала γ . Реализатор c_i вырабатывает управляющие воздействия на i -й подпроцесс, функционально зависящие от решения x_i и, возможно, еще от информации по обратной связи z_i ; следовательно, управляющее воздействие, поступающее на i -й подпроцесс, есть

$$m_i = c_i(z_i, d_i(\gamma, z_i)),$$

где γ — заданный координирующий сигнал.

3. КООРДИНИРУЕМОСТЬ

Координирование подсистем означает такое воздействие на подсистемы, которое заставляет их действовать согласованно, подобно тому как обычно координируется деятельность индивидуумов или групп внутри некоторой организации. Чтобы сделать такое представление о координации операциональным, нужно более четко определить, что именно подразумевается под словами «действовать согласованно». В общем случае координация осуществляется в связи с определенной целью или задачей; деятельность частей организации координируется ради общей цели так, чтобы вся организация в целом достигала поставленной цели.

Координация — это сфера деятельности или задача вышестоящей управляющей системы, в ходе которой она пытается добиться, чтобы нижестоящие системы управления функционировали согласованно. Успех вышестоящей управляющей системы в осуществлении надлежащей координации оценивается по отношению к общей глобальной цели, поставленной перед всей (в данном случае двухуровневой) системой. Так как нижестоящие управляющие системы действуют так, чтобы достичь своих собственных индивидуальных целей, то, вообще говоря, между ними возникает конфликт, который приводит к тому, что скорее всего глобальная цель не будет достигнута. Действия координатора направлены как раз на послед-

ствия такого внутриорганизационного конфликта, которые он должен постараться если не полностью исключить, то по крайней мере уменьшить.

Достижение цели, как уже указывалось в гл. 3, может мыслиться и как решение определенной задачи, поэтому классификация и выделение различных блоков управления в двухуровневой системе может производиться исходя из семейства решаемых ими задач. Однако для того чтобы формализовать понятие координации, необходимо ввести еще одну задачу «принятия решения», состоящую в оценке успеха деятельности по координации. Подобная задача определяется по отношению ко всей системе, и в частности ко всему процессу P , поэтому мы будем называть ее *глобальной решаемой задачей*.

Введем два понятия координируемости на примере двухуровневой системы. Первое понятие — это координируемость по отношению к задаче, решаемой вышестоящей управляющей системой, второе — координируемость по отношению к решаемой в настоящий момент глобальной задаче.

Для упрощения представления определим $P(x, \mathcal{D})$ для всех пар $(x, \mathcal{D})$, где $\mathcal{D}$ — произвольная решаемая задача, как предикат

$$P(x, \mathcal{D}) \equiv x \text{ есть решение } \mathcal{D}. \quad (4.7)$$

Следовательно, предикат $P(x, \mathcal{D})$ является истинным тогда и только тогда, когда $\mathcal{D}$ — решаемая задача, а x — одно из ее решений.

Так как мы в настоящее время интересуемся двухуровневой системой только в плане использования такой структуры для системы управления, то мы предполагаем, что информация, идущая по каналам обратной связи между вышестоящей и нижестоящей системами управления, является стандартизованной. Далее мы предполагаем, что решаемые задачи на уровне нижестоящих элементов параметризуются только координирующими сигналами (в этом и заключается координирующее воздействие) и что каждый раз рассматривается только одна задача, решаемая вышестоящей управляющей системой. Итак, пусть $\mathcal{D}_0$ — конкретная задача, решаемая вышестоящей управляющей системой, и каждый координирующий сигнал γ из $\mathcal{C}$ конкретизирует задачу $\mathcal{D}_i(\gamma)$, которую будет решать i -й решающий элемент; пусть далее $\bar{\mathcal{D}}(\gamma) = \{\mathcal{D}_1(\gamma), \dots, \mathcal{D}_n(\gamma)\}$ — совокупность таких задач. Следует заметить, что совокупность $\bar{\mathcal{D}}(\gamma)$ сама по себе также является подлежащей решению задачей, состоящей из n независимых задач $\mathcal{D}_i(\gamma)$; решениями задачи $\bar{\mathcal{D}}(\gamma)$ будут как раз те n -мерные векторы $(x_1, \dots, x_n)$, в которых каждая компонента x_i , $1 \leq i \leq n$, является решением соответствующей задачи $\mathcal{D}_i(\gamma)$.

Ниже мы подробнее остановимся на обоих понятиях координируемости.

Координируемость по отношению к задаче, решаемой вышестоящей управляющей системой

Мы предполагаем, что множество информационных сигналов, проходящих по каналам обратной связи, является фиксированным, поэтому мы можем для простоты положить (без потери общности), что $\mathcal{E} = X_0$, и, следовательно, сигналы вышестоящего решающего элемента непосредственно являются координирующими сигналами, поступающими на вход нижестоящих решающих элементов. Мы будем говорить, что задачи, решаемые нижестоящими элементами, координируемы по отношению к вышестоящей задаче, т. е. задаче, решаемой вышестоящим решающим элементом, тогда и только тогда, когда справедливо следующее предложение:

$$P(\gamma)(\exists x)[P(x, \overline{\mathcal{D}}(\gamma)) \text{ и } P(\gamma, \mathcal{D})]. \quad (4.8)$$

Следовательно, координируемость относительно задачи, решаемой вышестоящим элементом, требует, чтобы эта задача имела решение, и для некоторого координирующего входа γ , решающего данную задачу, множество $\overline{\mathcal{D}}_i(\gamma)$ задач, решаемых нижестоящими элементами, также имело решение. Для дальнейшего анализа удобно представлять условие (4.8) в такой форме, которая давала бы явное (экспlicitное) выражение зависимости решения задачи верхнего уровня $\mathcal{D}_0$ от действий нижестоящих решающих элементов. А именно, вышестоящий решающий элемент воздействует посредством координации на нижестоящие, и ответ на вопрос, будет ли решена задача при выбранном координирующем сигнале, может быть получен путем рассмотрения результатов, появляющихся на выходах нижестоящих решающих элементов.

Зависимость решения задачи $\mathcal{D}_0$ от результатов, получаемых на выходах нижестоящих решающих элементов, выражается формально как

$$P(\gamma, \mathcal{D}_0) \Leftrightarrow (\exists x)[Q_0(\gamma, x)], \quad (4.9)$$

где $Q_0(\gamma, x)$ — заданный предикат, определенный для всех пар (γ, x) из $\mathcal{E} \times X$, а X — декартово произведение множеств решений X_i :

$$X = X_1 \times \dots \times X_n.$$

Условие (4.9) просто утверждает, что данный координирующий сигнал γ решает задачу $\mathcal{D}_0$ тогда и только тогда, когда существует соответствующее решение x , получаемое на выходе нижестоящих элементов, такое, что условие, выраженное предикатом $Q_0(\gamma, x)$, удовлетворяется. Подлежащая решению задача $\mathcal{D}_0$, следовательно, состоит в том, чтобы найти γ из $\mathcal{E}$, такое, что $Q_0(\gamma, x)$ выполняется для решения x , получаемого на выходе нижестоящих решающих элементов. Далее, подставляя (4.9) в (4.8) и считая, что переменная x в (4.9) есть то же самое x , которое фигурирует в (4.8),

мы приходим к предположению

$$(\exists \gamma)(\exists x)[P(x, \overline{\mathcal{D}}(\gamma)) \text{ и } Q_0(\gamma, x)], \quad (4.10)$$

которое выражает координируемость по отношению к задаче, решаемой вышестоящей управляющей системой. Частные виды условия Q_0 будут вводиться в связи с различными формами принципов координации.

Координируемость по отношению к глобальной задаче

Глобальная решаемая задача определяется, как правило, для всего процесса, поэтому ее множество решений можно считать «множеством управлений» M . При фиксированной форме подачи информации через каналы обратной связи управляющие сигналы, имеющие своей целью изменение всего процесса, исходят только от нижестоящих решающих элементов; представим поэтому управляющие сигналы как отображение $\pi_M: X \rightarrow M$.

Тогда мы будем говорить, что задачи, которые будут решаться нижестоящими решающими элементами, координируемы относительно данной глобальной задачи $\mathcal{D}$, если справедливо следующее предложение:

$$(\exists \gamma)(\exists x)[P(x, \overline{\mathcal{D}}(\gamma)) \text{ и } P(\pi_M(x), \mathcal{D})]. \quad (4.11)$$

Координируемость относительно заданной глобальной задачи просто означает, что координатор, т. е. вышестоящая управляющая система, и в самом деле может влиять на нижестоящие решающие элементы так, чтобы их результирующее воздействие на процесс в целом давало решение глобальной задачи. С целью упрощения, а также ввиду важности решения глобальной задачи мы будем говорить, что *двухуровневая система координируема*, если задачи, решаемые на уровне нижестоящих элементов, могут быть скоординированы относительно поставленной глобальной задачи.

Мы будем считать, что нижестоящие решающие элементы координируемы (в определенном смысле), если могут быть скоординированы (в том же смысле) решаемые ими задачи.

4. ПОСТУЛАТ СОВМЕСТИМОСТИ

Для успешной работы двухуровневой системы существенно, чтобы цели (задачи) ее подсистем были согласованы между собой. На необходимость согласованности между глобальной целью иерархической системы и целями ее подсистем уже неоднократно указывалось различными авторами. В этой связи интересно привести высказывание Дж. Гэлбрейта ([9], стр. 204), который подчеркивал необходимость согласованности между рассматриваемыми им общественными целями трех типов: «Отношение между обще-

ством в целом и отдельной организацией должно быть согласовано с отношением этой организации к личности. Должна существовать согласованность целей общества, организации и личности». Он называет это положение принципом согласованности и настаивает на его справедливости в рамках любой социальной системы, и в частности в так называемом «индустриальном обществе».

В двухуровневой системе имеются цели трех типов, формально описываемые тремя типами решаемых задач: глобальными и решаемыми вышестоящими и нижестоящими управляющими системами. Совместимость этих целей или, в рамках нашего рассмотрения, принцип совместимости задач формально вытекает из следующих положений:

1. Только нижестоящие решающие элементы двухуровневой системы являются подсистемами, находящимися в непосредственном контакте со всем процессом. Если должна быть достигнута глобальная цель, то этого можно добиться только через действия нижестоящих решающих элементов; задачи, решаемые на этом уровне, или расположенные на этом уровне решающие элементы должны быть *координируемы* (т. е. обладать свойством координируемости) относительно решаемой глобальной задачи.

2. Вышестоящий решающий элемент, осуществляя координацию, воздействует на нижестоящие элементы, имея в виду свои собственные интересы: координатор выбирает координирующий сигнал так, чтобы продвигаться к осуществлению своей собственной цели. В этом случае задачи, решаемые на уровне нижестоящих элементов, должны быть координируемы по отношению к задачам, решаемым вышестоящим элементом. Следовательно, должно выполняться утверждение (4.10).

3. Глобальная задача, как правило, лежит вне сферы деятельности двухуровневой системы; ни один из решающих элементов внутри иерархии не облечен специально полномочиями решать глобальную задачу и тем самым преследовать общую (глобальную) цель, хотя задача определена в терминах всего процесса.

Для совместимости решаемых задач, а тем самым и целей внутри двухуровневой системы, координация задач, решаемых нижестоящими элементами. относительно задачи вышестоящего решающего элемента должна быть соответствующим образом связана с подлежащей решению глобальной задачей.

Рассматриваемое нами определение совместимости формально дается следующим предложением:

$$\begin{aligned}
 & (\forall \gamma) (\forall x) \{ [P(x, \bar{\mathcal{D}}(\gamma)) \text{ и } Q_0(\gamma, x)] \Rightarrow \\
 & \Rightarrow [P(x, \bar{\mathcal{D}}(\gamma)) \text{ и } P(\pi_M(x), \mathcal{D})] \}, \quad (4.12)
 \end{aligned}$$

которое мы впредь будем называть *постулатом совместимости* для двухуровневой системы. Если предложение (4.12) выполняется для исследуемой двухуровневой системы, мы будем называть

цели этой системы или решаемые этой системой задачи *совместимыми*.

Постулат утверждает, что решаемые на нижнем уровне (локальные) задачи скоординированы относительно решаемой глобальной задачи всякий раз, когда они скоординированы относительно задачи, решаемой на уровне вышестоящего элемента. Если решаемые данной двухуровневой системой задачи совместимы, то глобальная цель достигается тогда, когда вышестоящий решающий элемент координирует нижестоящие элементы по отношению к его собственной цели.

Важно подчеркнуть, что постулат совместимости есть математическое утверждение. Как таковое, оно не зависит от конкретного вида решаемых задач. Оно одинаково применимо в таких областях, как управление системами, оптимизация, искусственный интеллект, и в других применениях, где численные методы до сих пор оказывались малоэффективными.

Чтобы получить условие, налагаемое на задачи вышестоящего уровня и обеспечивающее основу для последующего синтеза задачи, решаемой на уровне вышестоящих элементов, попытаемся скомбинировать в определенном смысле понятия совместимости и координируемости. Указанное условие формулируется в следующем виде: двухуровневая система *координируема при определенным образом выбранной задаче вышестоящего уровня* тогда и только тогда, когда справедливы оба предложения: (4.10) и (4.12), другими словами, если решаемые двухуровневой системой задачи совместимы и задачи нижестоящих решающих элементов координируемы по отношению к задаче вышестоящего элемента. Это выражается предложением

$$(\forall \gamma) (\forall x) \{ [P(x, \bar{\mathcal{D}}(\gamma)) \text{ и } Q_0(\gamma, x)] \Rightarrow P(\pi_M(x), \mathcal{D}) \}, \quad (4.13)$$

что логически эквивалентно постулату совместимости. Поэтому мы будем называть решаемые задачи *совместимыми*, если предложение (4.13) истинно.

5. ПРИНЦИПЫ КООРДИНАЦИИ

Хотя постулат совместимости и требование координируемости указывают, какими свойствами должна обладать задача, решаемая вышестоящим элементом, они отнюдь не исчерпывают решения проблемы синтеза, стоящей перед координатором. Фактически эти условия лишь помогают сформулировать проблему синтеза как структурную проблему; они накладывают ограничения на стратегии, которые координатор может использовать. Однако мы не знаем, какую информацию координатор должен получить и как использовать эту информацию для выбора наилучшего координирующего воздействия, т. е. мы не знаем, какой должна быть фактическая стратегия координатора.

Поучительно взглянуть, как в прошлом решались сходные по своей структуре проблемы. В поисках аналогии обратимся к области автоматического регулирования в период, предшествовавший созданию концепции обратной связи. Ситуация тогда была в основных чертах следующая: дан процесс P , на вход которого поступают как управляющие, так и возмущающие воздействия; спрашивается, как нужно выбрать управление, чтобы противодействовать влиянию возмущений. Для решения этой проблемы был введен *принцип управления с помощью обратной связи*: выход должен сравниваться с желаемым состоянием и наблюдающееся отклонение после соответствующего преобразования должно по цепи обратной связи подаваться на управляющие органы для изменения управления в нужную сторону. Аналогично *принцип оптимальности* Беллмана [26] предполагает, что выбор управления в непосредственно следующий за данным моментом интервал времени должен делаться в предположении, что управление в течение всего остального периода времени будет оптимальным. Если установлен принцип, остается лишь проблема определения вытекающей из него стратегии (например, определение параметров в цепи обратной связи или выбор управления для начального интервала времени) и анализ условий, при которых применима та или иная конкретная стратегия. Заметим, что в истории обоих вопросов первый этап является эвристическим и включает то, что можно было бы назвать «нововведением», тогда как второй этап обычно составляет лишь математическое исследование.

Мы будем действовать аналогичным образом. Прежде всего мы постулируем некоторые принципы координации, которые определяют различные стратегии для координатора (т. е. определяют структуру координации), а затем проанализируем области приемлемости или применимости этих стратегий.

Основная причина возникновения конфликтов в двухуровневой системе связана с взаимодействием подпроцессов и с тем, что каждый из нижестоящих решающих элементов находится в неведении относительно решений, принятых другими решающими элементами того же уровня. Задача координатора, вообще говоря, состоит в оказании на нижестоящие решающие элементы такого влияния, которое приводит к желательным в некотором заранее установленном смысле результирующим взаимодействиям.

Существуют три подхода к рассмотрению такого рода взаимодействий:

1. *Прогнозирование взаимодействий.* Координирующие сигналы могут содержать в себе, помимо всего другого, прогноз связующих входов (сигналов); в этом случае каждый координирующий сигнал u несет с собой прогнозные значения $\alpha^i = (\alpha_1^i, \dots, \alpha_n^i)$ связующих входов, которые будут иметь место в связи с подачей управляющих воздействий.

2. «Развязывание» взаимодействий. Каждый нижестоящий решающий элемент получает право при решении собственной задачи рассматривать связующие входы как дополнительные свободные переменные, которые он волен выбирать по своему усмотрению. Очевидно, что подлежащие решению задачи нижестоящего уровня определяются в этом случае так, как если бы нижестоящие решающие элементы и подпроцессы были полностью «развязанными» (т. е. автономными). Тогда связующий вход u , выбираемый нижестоящими решающими элементами, есть просто часть решения x и будет задаваться отображением $\pi_U: X \rightarrow U$.

3. Оценка взаимодействий. Координатор в этом случае не сообщает точных значений связующих сигналов, а лишь ограничивает области их изменения: каждый координирующий сигнал γ , принадлежащий $\mathcal{E}$, выделяет множество $U^\gamma = U_1^\gamma \times \dots \times U_n^\gamma \subseteq U$; тогда i -й решающий элемент считает U_i^γ установленным диапазоном возмущений.

Ниже мы рассмотрим три принципа координации, основанных на постулате совместности, записанном в виде (4.13), и опишем метод, позволяющий исследовать эффект взаимодействий (роль связующих сигналов).

Принцип прогнозирования взаимодействий

Предположим, что координатор прогнозирует будущие значения связующих сигналов. Довольно естественно предположить, что успех в координации нижестоящих решающих элементов зависит от точности прогнозирования связующих сигналов или, в более широком смысле, — от влияния ошибок прогнозирования.

Самая простая форма принципа прогнозирования взаимодействий дается предложением

$$(\forall \gamma) (\forall x) \{ [P(x, \overline{\mathcal{D}}(\gamma)) \text{ и } K(\pi_M(x)) = \alpha^\gamma] \Rightarrow \\ \Rightarrow P(\pi_M(x), \mathcal{D}) \}. \quad (4.14)$$

Принцип просто утверждает, что подлежащая решению глобальная задача разрешается с помощью управляющего воздействия $m = \pi_M(x)$ всякий раз, когда x является решением задач, поставленных перед нижестоящими элементами, и правильно прогнозируются взаимодействия, т. е. α^γ есть действительно тот самый связующий сигнал, который будет иметь место при управляющем воздействии $m = \pi_M(x)$.

Вместо сравнения прогнозных и фактических значений связующих сигналов можно сравнивать прогнозируемую и фактическую работу подсистем в более обобщенном смысле. В общей форме принцип формулируется в виде предложения

$$(\forall \gamma) (\forall x) \{ [P(x, \overline{\mathcal{D}}(\gamma)) \text{ и } q(\gamma, x) = \\ = \tilde{q}(\gamma)] \Rightarrow P(\pi_M(x), \mathcal{D}) \}, \quad (4.15)$$

где $\tilde{q}$ и q — заданные функции, отображающие $\mathcal{C}$ и $\mathcal{C} \times X$ соответственно на обычное числовое множество (числовую прямую) и используемые для определения точности прогнозирования. Ясно, что (4.14) есть частный случай (4.15). Эта более общая форма принципа будет в дальнейшем называться просто *принципом прогнозирования*.

Принцип согласования взаимодействий

Предположим, что при координации используется тот подход, который мы называем «развязыванием» взаимодействий. Тогда успех в координации нижестоящих решающих элементов можно оценить, исходя из расхождений или рассогласованности между фактическими взаимодействиями и теми, которые были бы желательными с точки зрения нижестоящих решающих элементов.

Принцип согласования взаимодействий дается предложением

$$(\forall \gamma) (\forall x) \{ [P(x, \overline{\mathcal{D}}(\gamma)) \text{ и } K(\pi_M(x)) = \pi_U(x)] \Rightarrow \\ \Rightarrow P(\pi_M(x), \mathcal{D}) \}. \quad (4.16)$$

Этот принцип утверждает, что управляющее воздействие $m = \pi_M(x)$ решает поставленную глобальную задачу всякий раз, когда x является решением задач нижестоящих элементов и желаемые связующие сигналы $u^v = \pi_U(x)$ согласованы (совпадают) с фактическими связующими сигналами $u = K(m)$, имеющими место тогда, когда к процессу приложено управляющее воздействие $m = \pi_M(x)$.

В общей форме принцип, называемый далее просто *принципом согласования*, выражается предложением

$$(\forall \gamma) (\forall x) \{ [P(x, \overline{\mathcal{D}}(\gamma)) \text{ и } q(\gamma, x) = \\ = \tilde{q}(\gamma, x)] \Rightarrow P(\pi_M(x), \mathcal{D}) \}, \quad (4.17)$$

где $\tilde{q}$ и q — заданные функции, отображающие $\mathcal{C} \times X$ на числовую прямую. Этот принцип позволяет устанавливать, достигнуто ли согласование в соответствующем смысле.

Принцип оценки взаимодействий

Предположим теперь, что координатор вместо прогнозирования точных значений самих связующих сигналов определяет области, в пределах которых они могут варьироваться. В соответствии с ранее описанными координационными принципами мы предположим, что успешность координации нижестоящих решающих элементов можно определить на основании точности этих оценок.

Принцип оценки взаимодействий выражается предложением

$$(\forall \gamma) (\forall x) \{ [P(x, \overline{\mathcal{D}}(\gamma)) \text{ и } K(\pi_M(x)) \in U^v] \Rightarrow P(\pi_M(x), \mathcal{D}) \}. \quad (4.18)$$

Этот принцип координации можно рассматривать как обобщение принципа прогнозирования взаимодействий. В самом деле, если оценочные области U^v являются множествами, состоящими из одного элемента, мы имеем принцип прогнозирования взаимодействий. Условие $K(m) \in U^v$ в (4.18) означает, что фактический связующий сигнал, появляющийся при применении управляющего воздействия $m = \pi_m(x)$, попадает внутрь оценочной области U^v .

В общей форме этот принцип, называемый *принципом оценки*, выражается предложением

$$(\forall \gamma)(\forall x) \{ [P(x, \bar{\mathcal{D}}(\gamma)) \text{ и } q(\gamma, x) \in \tilde{q}(\gamma)] \Rightarrow \\ \Rightarrow P(\pi_m(x), \mathcal{D}) \}, \quad (4.19)$$

где $\tilde{q}$ — заданная оценочная функция, определенная на $\mathcal{C}$; функция q считается заданной и определенной на $\mathcal{C} \times X$.

Координируемость и принципы координации

Сравнение постулата совместимости, выражаемого предложением (4.13), с выбранным принципом координации сразу указывает нам предикат $Q_0(\gamma, x)$ в выражении (4.9), а следовательно, и задачу, решаемую вышестоящей системой. Предположим, например, что заданы функции $\tilde{q}$ и q и используется принцип прогнозирования; тогда предикат $Q_0(\gamma, x)$ получается из (4.15) в виде

$$Q_0(\gamma, x) \Leftrightarrow [q(\gamma, x) = \tilde{q}(\gamma)], \quad (4.20)$$

и, следовательно, задача, решаемая на уровне вышестоящего элемента C_0 , состоит в том, чтобы найти γ в $\mathcal{C}$ [такое, что

$$q(\gamma, x) = \tilde{q}(\gamma),$$

где x — результирующее решение задач нижестоящих решающих элементов.

Мы будем ниже называть предикат $Q_0(\gamma, x)$ в принципе совместимости *условием координируемости*, в случае же, если $Q_0(\gamma, x)$ определяется по отношению к выбранному принципу координации, мы будем называть его *условием координируемости для этого принципа*.

Введем следующие понятия:

1. Выбранный принцип координации *применим*, если соответствующее логическое предложение, выражающее этот принцип, истинно. Например, для применимости принципа прогнозирования взаимодействий предложение (4.14) должно быть истинно, тогда как для принципа согласования взаимодействий должно иметь место (4.16).

2. Система *координируема с помощью данного принципа координации*, если принцип применим и существует координирующий

сигнал $\gamma \in \mathcal{E}$, такой, что удовлетворяется соответствующее условие координируемости $Q_0(\gamma, x)$.

Понятие применимости полезно, ибо «применимость» дает гарантию того, что использование принципа не приведет к ошибочным результатам. Но применимость принципа координации еще не гарантирует координируемость с помощью этого принципа. Предположим, например, что для каждого координирующего сигнала γ существует по крайней мере одна задача на уровне нижестоящих решаемых элементов, для которой не удастся получить решение; тогда, даже если и применим какой-то принцип координации, система тем не менее не будет координируема. С другой стороны, предположим, что в данной двухуровневой системе условие координируемости для выбранного принципа координации никогда не удовлетворяется; тогда совершенно ясно, что принцип координации может быть и применим, но нам не удастся скоординировать систему с помощью этого принципа, хотя она, возможно, могла бы быть весьма успешно скоординирована (т. е. является координируемой) каким-то другим путем.

Ранее упоминалось, что, как только принят определенный принцип координации, проблема координации может быть подвергнута математическому рассмотрению. Почти вся вторая часть книги посвящена математическому исследованию задачи координации.

6. РАЗЛИЧНЫЕ АСПЕКТЫ ПРОБЛЕМЫ КООРДИНАЦИИ

Координация, как деятельность вышестоящей управляющей системы, связана с тремя типами решаемых задач: глобальной задачей и задачами, решаемыми на вышестоящем и нижестоящем уровнях. В связи с этим возникают четыре различные проблемы:

1. *Синтез координирующего элемента.* Даны глобальная задача и задачи, решаемые нижестоящими элементами; нужно найти такую задачу $\mathcal{D}_0$, решаемую на уровне координирующего (вышестоящего) элемента C_0 , чтобы система была координируема на основе $\mathcal{D}_0$.

2. *Методы, или процедуры, координации.* Дана двухуровневая система, которая координируема по отношению к задаче $\mathcal{D}_0$; требуется найти эффективный метод (алгоритм) получения координирующего сигнала, который скоординировал бы всю систему.

3. *Проблема модификации.* Дана двухуровневая система, не координируемая по отношению к задаче $\mathcal{D}_0$ (хотя постулат совместности и может выполняться); необходимо найти такую модификацию задач, решаемых на нижестоящем уровне, чтобы эти модифицированные задачи были координируемы относительно задачи $\mathcal{D}_0$.

4. *Декомпозиция.* Поставлена только глобальная задача; нужно найти задачи, подлежащие решению на вышестоящем и нижестоящем уровнях, с тем чтобы двухуровневая система была координируема.

нируемой по отношению к задаче, решаемой на вышестоящем уровне $\mathcal{D}_0$.

Обсудим теперь вкратце связь между этими проблемами.

Проблема синтеза координирующего элемента

Наш подход к проблеме синтеза координирующего элемента основывается на принципах координации, изложенных в предыдущем разделе. Если принцип координации принят, задача $\mathcal{D}_0$ для вышестоящей управляющей системы определена. Тогда немедленно возникают два вопроса: 1) Выполняется ли постулат совместимости? Применим ли избранный принцип координации? 2) Если да, то существует ли координирующий сигнал, который координирует систему и получается в результате решения задачи вышестоящего элемента (координации) $\mathcal{D}_0$, определяемой через условие координируемости для принятого принципа координации? Другими словами, координируема ли система на основе выбранного принципа координации? Во второй части этой книги мы исследуем математическими методами вопросы применимости и координируемости с помощью предлагаемых принципов координации в различных конкретных ситуациях. Многообразие рассматриваемых случаев и общность анализа являются залогом широты применений предлагаемых принципов.

Процедуры координации

Как только задача, решаемая на вышестоящем уровне управления, выбрана, возникает проблема отыскания ее решения. Конечно, для этого существует множество способов, но самые важные из них следующие.

«Линейные» и «многофазные» процедуры итераций

«Улучшить» координирующий сигнал по отношению к найденному решению задачи вышестоящего уровня и добиться удовлетворения условия координируемости при выполнении постулата совместимости можно при определенных условиях за счет объединенных усилий решающих элементов обоих уровней. Для этого может быть использована следующая итеративная процедура: пусть γ^k и m^k — координирующий и управляющий сигналы на k -й стадии итерации; тогда на основании оценки качества работы системы вышестоящий решающий элемент подбирает новый координирующий сигнал γ^{k+1} , который, как он надеется, позволит улучшить характеристики системы по сравнению с γ^k ; используя γ^{k+1} , нижестоящие решающие элементы вырабатывают свои решения и посылают управляющий сигнал (воздействие) m^{k+1} . Эта

итеративная процедура повторяется до тех пор, пока не будет решена задача вышестоящего элемента $\mathcal{D}_0$ или не будет достигнуто желаемое состояние.

В системах с «многофазными» процедурами работы управляющий сигнал не должен поступать на вход управляемого процесса вплоть до того момента, когда будет найден координирующий сигнал, который является решением задачи вышестоящего элемента или сможет скоординировать систему. В таких случаях возникает вопрос, сходятся ли итерации к желаемому состоянию, и если да, то какова скорость этой сходимости.

В системах с «линейной» процедурой работы применение управляющего воздействия нельзя отсрочить и в предельном случае управляющие сигналы (управляющие воздействия) должны подаваться на вход управляемого процесса на каждой стадии итерации, ввиду того что работа системы происходит в одном и том же масштабе времени с процессами принятия решения; поэтому задержка подачи управляющего сигнала может привести к некоторому ухудшению характеристик работы системы. Цель «линейных» процедур координации состоит в том, чтобы улучшать характеристики работы системы на каждой стадии процесса координации.

Системы с обоими типами процедур более подробно рассматриваются в ч. II.

Использование обратной связи

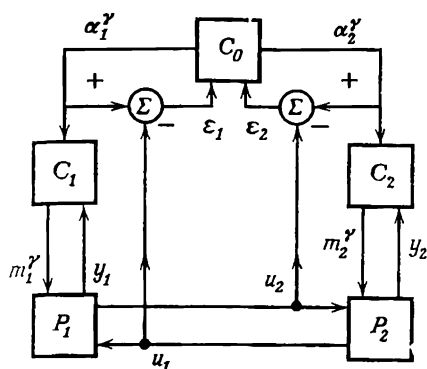
Использование того или иного принципа координации (когда они применимы) имеет то преимущество, что, если условия координируемости не выполняются (для выбранного координирующего сигнала и соответствующих решений нижестоящих элементов), «ошибка» может быть обнаружена и затем использована для улучшения координирующего сигнала.

Предположим, например, что для рассматриваемой двухуровневой системы применим принцип прогнозирования взаимодействий, а множество U представляет собой линейное пространство. Тогда для выбранного координирующего сигнала γ ошибка может быть, например, представлена в виде

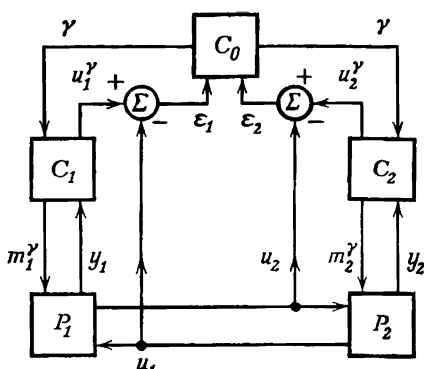
$$\varepsilon = \alpha^\gamma - K(m^\gamma),$$

где α^γ — предсказанный связующий сигнал, $u = K(m^\gamma)$ — фактический связующий сигнал, появляющийся, когда применяется управляющее воздействие $m^\gamma = c(x^\gamma)$, а x^γ — решение, выработанное нижестоящими элементами. Новый координирующий сигнал γ' может быть получен применением подходящего преобразования $\mathcal{T}$ к ошибке ε :

$$\gamma' = \gamma + \mathcal{T}(\varepsilon) = \gamma + \mathcal{T}(\alpha^\gamma - K(m^\gamma)).$$



Ф и г. 4.6. Использование принципа прогнозирования взаимодействий в цепи обратной связи второго уровня.



Ф и г. 4.7. Использование принципа согласования взаимодействий в цепи обратной связи второго уровня.

Блок-схема на фиг. 4.6 иллюстрирует этот подход к использованию принципа прогнозирования взаимодействий. Фиг. 4.7 дает иллюстрацию аналогичного подхода, когда используется принцип согласования взаимодействий. Чтобы определить сигнал ошибки при использовании принципа оценки взаимодействий, необходимо иметь метрику, или норму, на множестве U для нахождения расстояния между точкой и множеством.

После того как выбран принцип координации и установлен сигнал ошибки, поступающий по каналу обратной связи, возникает следующий вопрос: как выбрать преобразование, используемое в цепи обратной связи на второй уровень? ¹⁾ Проблемы такого типа, вообще говоря, решаются путем выделения некоторого класса преобразований и выбора конкретного преобразования из этого класса на основе какого-то заданного критерия. Обычно класс преобразований определяется так, что проблема сводится к выбору параметров в цепи обратной связи на второй уровень. Так как задача цепи обратной связи на второй уровень в общем случае состоит в том, чтобы устранить ошибку, то простейшим подходом было бы включение классического ПИД-регулятора; проблема тогда заключается в выборе коэффициентов усиления и других параметров регулятора. Более утонченный подход состоял бы в выборе более сложной структуры обратной связи, основанной, возможно, на оптимизационном принципе. Например (см. гл. 6), при построении цепей обратной связи может быть использовано аналитическое решение задачи нахождения оптимального координирующего сигнала в линейном и квадратичном случаях. Мы не будем здесь

¹⁾ Здесь и далее авторы под «вторым уровнем» обычно понимают уровень вышестоящего (координирующего) элемента. — Прим. ред.

подробнее останавливаться на этой проблеме, хотя она и отражает плодотворное направление для получения новых интересных и практически важных результатов.

По поводу указанного подхода хотелось бы отметить следующее.

Сигналом для изменения координации служит появление ошибки на выходе самого процесса. Поэтому существует неизбежное запаздывание между осуществлением изменения, откликом системы, информация о котором поступает по каналам обратной связи, и коррекцией ошибки. Короче говоря, любая координация, основанная только на информационной обратной связи на второй уровень и не предусматривающая никакого «упреждения», является «субоптимальной». Естественно поэтому, что положение можно было бы улучшить за счет сочетания информационной обратной связи на второй уровень с упреждающими действиями; на наш взгляд, исследования в этом направлении стоило бы продолжить.

Обратная связь на второй уровень, образуемая при использовании координационных принципов, обладает преимуществом, свойственным любой стандартной системе с отрицательной обратной связью, — уменьшением чувствительности характеристик качества работы системы к возмущениям. Реальный выигрыш, конечно, зависит от системы и рассматриваемых возмущений, но обратная связь на второй уровень дает сравнительно простое решение проблемы координации с использованием «сквозной» процедуры работы. Этот вопрос заслуживает более детального изучения, особенно в плане конкретных применений, когда в полной мере можно использовать конкретные особенности структуры системы.

Модификация

Проблема модификации задач, решаемых нижестоящими элементами, может быть сформулирована следующим образом: предположим, что для рассматриваемой двухуровневой системы, имеющей в качестве решаемых задач нижестоящего уровня (локальные) задачи $\bar{\mathcal{D}}(\gamma)$, $\gamma \in \mathcal{E}$, постулат совместимости справедлив, но система тем не менее не координируема. Проблема модификации тогда заключается в нахождении нового семейства задач, решаемых на нижестоящем уровне, таких, что

(1) $\mathcal{E} \subseteq \mathcal{E}'$ и $\bar{\mathcal{D}}(\gamma) \subseteq \bar{\mathcal{D}}'(\gamma)$ для всех $\gamma \in \mathcal{E}$;

(2) постулат совместимости остается по-прежнему справедливым;

(3) система координируема по отношению к задаче, решаемой на вышестоящем уровне.

Эти условия требуют, чтобы первоначальная задача, решаемая элементами нижестоящего уровня, была погружена в новое семейство задач, чтобы выполнялся постулат совместимости и чтобы

существовал координирующий сигнал, который координировал бы систему на основе решения задачи вышестоящего уровня.

Нередки случаи, когда хотя и удастся конкретизировать свою задачу для каждого из нижестоящих элементов, нет априорной уверенности в том, что при этом удастся добиться координируемости. Несмотря на справедливость постулата совместимости, может оказаться, что необходимо увеличить число возможных задач нижестоящего уровня, как указывает условие (1), в результате чего система станет координируемой на основе решения задачи вышестоящего уровня.

Модификации решаемых задач на нижестоящем уровне могут потребоваться в разных ситуациях. Например, данная двухуровневая система может быть координируемой, но не с помощью ранее выбранной задачи вышестоящего уровня, или же может не удовлетворяться постулат совместимости. Сюда же относится случай, когда задачи, решаемые на нижестоящем уровне, могут оказаться не координируемыми ни по отношению к решаемой задаче вышестоящего уровня $\mathcal{D}_0$, ни по отношению к решаемой глобальной задаче. Модификация задач, решаемых на нижестоящем уровне, есть средство, помогающее избавиться от этих «патологических случаев».

В ч. II проблема модификации будет рассмотрена более подробно. Мы опишем различные методы модификации и проведем их анализ.

Декомпозиция

Под декомпозицией мы понимаем решение следующей проблемы. Дана глобальная задача; найти задачи, которые могли бы быть поставлены перед вышестоящим и нижестоящими решающими элементами так, чтобы выполнялся постулат совместимости.

Мы утверждаем, что проблема декомпозиции сводится к трем предыдущим проблемам, а именно к проблеме синтеза координирующего элемента, к проблеме модификации и к отысканию самой процедуры координации. В самом деле, имеется много способов разложить данную глобальную задачу на подзадачи. Реальная трудность, однако, заключается в том, как их скоординировать; это прежде всего требует выбора принципа координации (т. е. задачи, которая будет поставлена на вышестоящем уровне), затем модификации подзадач и, наконец, разработки метода отыскания координирующего их сигнала.

В любом случае ключевой проблемой является проблема координации; именно она и будет предметом нашего рассмотрения в последующих главах.

Можно считать, что вообще всякий новый метод координации предполагает и свой метод декомпозиции. В этой связи следует

отметить общность развиваемой теории координации. Термин «декомпозиция» часто используется в задачах оптимизации или, точнее, в задачах математического программирования; однако концепции и теория координации, которые мы развиваем здесь, намного шире. Они применимы к гораздо более общему классу систем, например к системам, работающим в реальном масштабе времени (в нашей интерпретации это соответствует использованию «сквозных» процедур координации), для которых обычно отыскивается скорее удовлетворительный, чем оптимальный уровень качества работы.

Часть II

МАТЕМАТИЧЕСКАЯ ТЕОРИЯ КООРДИНАЦИИ

Глава 5

ОБЩАЯ ТЕОРИЯ КООРДИНАЦИИ ДЛЯ СИСТЕМ ОПТИМИЗАЦИИ ПРИ ОТСУТСТВИИ ОГРАНИЧЕНИЙ

Как и в предыдущей главе, мы здесь будем рассматривать двух-уровневую систему. Отправляясь от системы самого общего вида, мы, однако, введем два дополнительных предположения: 1) внешние возмущения отсутствуют, так что все задачи, связанные с принятием решений, можно определить как задачи оптимизации; 2) множества решений не накладывается ограничений, или, точнее, рассматривается случай

$$M' = M_1 \times \quad \times M_n.$$

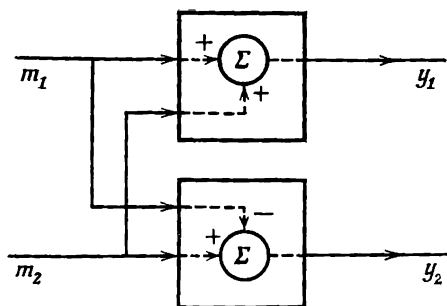
Введя эти два предположения, мы сформулируем глобальную задачу и задачи, решаемые на нижестоящих уровнях, и проанализируем в самом общем виде проблему координации, чтобы показать широкую применимость предлагаемых принципов координации. Наш анализ будет касаться лишь задач оптимизации при отсутствии ограничений; это позволит сосредоточить внимание на главных структурных аспектах получаемых выводов и избежать слишком специальных рассуждений, представляющих скорее технический, нежели принципиальный интерес. Более подробная теория для этого случая будет изложена в гл. 6, а задачи с ограничениями будут рассмотрены в гл. 7. Влияние возмущений на работу системы составит предмет рассмотрения гл. 8.

1. ВВЕДЕНИЕ

Принципы координации являются центральным пунктом в нашем подходе к координации, поскольку на их основе формируются стратегии для выбора координирующих сигналов. В помощь интуиции мы приведем прежде всего несколько простых примеров для иллюстрации содержательного смысла принципов координации.

Рассмотрим простейший тип многоуровневой системы: двух-уровневую систему, имеющую на верхнем уровне один, а на нижележащем уровне — два решающих элемента. Пусть подпроцессы P_1 и P_2 определяются на множестве R^2 уравнениями

$$y_1 = 2m_1 + u_1 \equiv P_1(m_1, u_1), \quad y_2 = 2m_2 - u_2 \equiv P_2(m_2, u_2). \quad (5.1)$$



Ф и г. 5.1.

Предположим, что уравнения

$$u_1 = y_2, u_2 = y_1$$

определяют такое соединение подпроцессов, что процесс в целом (см. фиг. 5.1) есть отображение $P: R^2 \rightarrow R^2$, определяемое системой двух уравнений:

$$y_1 = m_1 + m_2,$$

$$y_2 = -m_1 + m_2.$$

Пусть, далее, глобальная функция качества всей системы G представляет собой отображение множества $R^2 \times R^2$ в область действительных чисел и определяется выражением

$$G(m, y) = m_1^2 + m_2^2 + (y_1 - 1)^2 + (y_2 - 1)^2,$$

и предположим, что глобальной решаемой задачей является задача оптимизации: найти управляющее воздействие $\hat{m}$ из R^2 , такое, что $G(\hat{m}, P(\hat{m})) \leq G(m, P(m))$ для всех других управлений m из R^2 .

Пусть $\beta = (\beta_1, \beta_2)$ — пара действительных чисел, и пусть для каждой такой пары заданные функции качества на нижележащем уровне $G_{1\beta}$ и $G_{2\beta}$ — действительные функции:

$$G_{1\beta}(m_1, u_1, y_1) = m_1^2 + (y_1 - 1)^2 + \beta_1 u_1^2 - \beta_2 y_1^2,$$

$$G_{2\beta}(m_2, u_2, y_2) = m_2^2 + (y_2 - 1)^2 + \beta_2 u_2^2 - \beta_1 y_2^2$$

на $R \times R \times R$. Предположим, что для любой заданной пары β обе задачи, решаемые управляющими элементами этого уровня, представляют собой задачу оптимизации: найти пару $\hat{x}_i$ из R^2 , такую, что $G_{i\beta}(\hat{x}_i, P_i(\hat{x}_i)) \leq G_{i\beta}(x_i, P(x_i))$ для всех пар x_i из R^2 .

Пары β суть координирующие сигналы. Для каждого β решение задач нижестоящего уровня дает локальные управляющие воздействия $m_1(\beta)$ и $m_2(\beta)$. Задача координирующего элемента в этом случае сводится к отысканию такого $\hat{\beta}$, чтобы соответствующие локальные управляющие воздействия $m_1(\hat{\beta})$ и $m_2(\hat{\beta})$ были глобально оптимальными управлениями, т. е. чтобы управляющее воздействие $m(\beta) = (m_1(\beta), m_2(\beta))$ было решением глобальной задачи оптимизации.

Заметим, что при решении задачи оптимизации решающие элементы нижестоящего уровня должны произвести соответствующий выбор не только своих (локальных) управляющих воздействий, но и определенных значений связующих входов u_1 и u_2 . Для данной пары

β пусть $u_1(\beta)$ и $u_2(\beta)$ будут «оптимальными» значениями связующих входов, т. е. их значениями, найденными в процессе оптимизации в предположении, что подпроцессы являются «автономными» и независимыми друг от друга (или, как принято говорить, «развязанными»); связующие входы считаются свободными переменными, которыми можно манипулировать. Однако фактически в силу существования взаимосвязи между подпроцессами, такой свободы нет, поэтому, если управляющие воздействия $m_1(\beta)$ и $m_2(\beta)$ действительно будут приложены к процессу, то вследствие взаимосвязи подпроцессов фактические значения связующих входов (связующих сигналов), реализующихся для процесса P , будут равны

$$u_1 = -m_1(\beta) + m_2(\beta), \quad u_2 = m_1(\beta) + m_2(\beta);$$

и, вообще говоря, не будут совпадать со значениями, выбранными путем оптимизации.

Это расхождение создает предпосылки для выработки стратегии координации. Принцип согласования взаимодействий, сформулированный в гл. 4, в применении к рассматриваемому случаю гласит: управляющее воздействие $m(\beta) = (m_1(\beta), m_2(\beta))$ будет глобально оптимальным при любом β , удовлетворяющем условию согласованности связующих входов:

$$u_1 = u_1(\beta), \quad u_2 = u_2(\beta).$$

Это условие означает, что выбираемые на нижележащем уровне связующие входы должны быть именно теми, которые будут иметь место при приложении к процессу управляющих воздействий $m_1(\beta)$ и $m_2(\beta)$.

Легко проверить, что принцип согласования взаимодействий действительно применим к рассматриваемому случаю. Если β является парой $(0, -1/2)$, то решения оптимизационных задач на нижележащем уровне имеют вид

$$(m_1(\beta), u_1(\beta)) = (0, 2/3), \quad (m_2(\beta), u_2(\beta)) = (2/3, 2/3).$$

При решении глобальной оптимизационной задачи мы находим, что управление $m(\beta) = (0, 2/3)$ есть глобально оптимальное управляющее воздействие; кроме того, если применяется управляющее воздействие $m(\beta)$, то результирующими связующими входами будут $u(\beta) = (2/3, 2/3)$, что согласуется со входами, выбранными в результате оптимизации. Следовательно, достигнуто согласование взаимодействий, и локальные управляющие воздействия дадут оптимальное решение глобальной проблемы.

Применение принципа согласования взаимодействий сводит проблему координации к отысканию такой пары (β_1, β_2) , которая обеспечивает выполнение принципа согласования взаимодействий.

Чтобы проиллюстрировать другой принцип координации, а именно принцип прогнозирования взаимодействий, рассмотрим

опять пример двухуровневой системы, имеющей на нижележащем уровне только два решающих элемента. Пусть, однако, локальные процессы¹⁾ P_1 и P_2 являются действительными отображениями, определенными на множествах $M_1 \times R$ и $M_2 \times R$ соответственно:

$$y_1 = m_1 + (7/3) u_1^3 \equiv P_1(m_1, u_1), \quad y_2 = m_2 - u_2^3 \equiv P_2(m_2, u_2), \quad (5.2)$$

где M_i для $i = 1, 2$ есть множество всех m_i из R , таких, что $|m_i| \leq 1$. Пусть связь подпроцессов между собой такая же, как и ранее: $u_1 = y_2$ и $u_2 = y_1$. Процесс в целом есть отображение $P: M_1 \times M_2 \rightarrow R^2$, полученное исключением связующих входов u_1 и u_2 из уравнений подпроцесса.

Пусть функции качества для нижестоящих (локальных) элементов G_1 и G_2 заданы на $M_1 \times R$ и $M_2 \times R$ соответственно в виде

$$G_1(m_1, y_1) = m_1^2 - (1/3) m_1 + y_1,$$

$$G_2(m_2, y_2) = 4m_2^2 + (1/7) m_2 + y_2,$$

тогда как глобальная функция качества G задана на $M \times R^2$ в виде

$$G(m, y) = m^2 + 4m^2 + y_1 + y_2.$$

Предполагается, что как глобальная задача, так и задачи, решаемые на нижестоящем уровне, представляют собой задачи оптимизации.

Попытаемся скоординировать задачи, подлежащие решению на нижележащем уровне, обязав координатора прогнозировать конкретные значения связующих входов. Координатор будет задавать прогнозные значения α_1 и α_2 связующих входов u_1 и u_2 . Затем решающие элементы нижележащего уровня выберут локальные управляющие воздействия, решая задачи оптимизации для прогнозных значений связующих входов. Обозначим через $m_1(\alpha)$ и $m_2(\alpha)$ решения, полученные для конкретных значений $\alpha = (\alpha_1, \alpha_2)$ связующих входов. Мы утверждаем, что система может быть скоординирована путем прогнозирования соответствующих значений связующих входов.

Пусть $\alpha = (\alpha_1, \alpha_2)$ — набор прогнозных значений связующих входов и предположим, что в результате осуществляется управляющее воздействие $m(\alpha) = (m_1(\alpha), m_2(\alpha))$. Фактические связующие входы тогда имеют вид $(u_2, u_1) = (y_1, y_2) = P(m_1(\alpha), m_2(\alpha))$

¹⁾ Рассматривая двухуровневые системы, авторы часто называют «нижестоящие» решающие элементы также «локальными» управляющими (решающими) элементами, а «вышестоящий» решающий элемент — просто «координатором». Относящиеся к этим уровням задачи соответственно называются «локальными задачами» и «задачами координатора» («задачами координации»). Локальным управляющим элементом обычно поручается управление какими-то «локальными» подпроцессами (в данном случае P_1 и P_2), которые в своей совокупности образуют «процесс в целом» P . — Прим. ред.

и, вообще говоря, не согласуются с прогнозными значениями α_1 и α_2 . Принцип прогнозирования взаимодействий для рассматриваемого примера утверждает: управляющее воздействие $m(\alpha) = (m_1(\alpha), m_2(\alpha))$ является глобально оптимальным всякий раз, когда прогноз $\alpha = (\alpha_1, \alpha_2)$ оправдывается, т. е. когда

$$u_1 = \alpha_1, u_2 = \alpha_2.$$

Легко проверить, что принцип прогнозирования взаимодействий применим для рассматриваемого примера. В самом деле, если $\alpha_1 = -1/7$ и $\alpha_2 = -1/3$, то для результирующих локальных управляющих воздействий имеем

$$m_1(\alpha) = -1/3, m_2(\alpha) = -1/7,$$

причем $m(\alpha) = (-1/3, -1/7)$ является глобально оптимальным управлением, тогда как фактические связующие входы, получающиеся при применении $m(\alpha)$, имеют значения $u_1 = -1/7$ и $u_2 = -1/3$. Кроме того, для любых прогнозных значений α связующих входов результирующее управляющее воздействие $m(\alpha)$ является глобально оптимальным только тогда, когда прогноз α оказывается верным.

Расхождения между прогнозными и фактическими связующими входами могут быть поэтому использованы для того, чтобы судить о правильности выбранного координирующего воздействия; проблема координации сводится, таким образом, к проблеме прогнозирования связующих входов. Кроме того, различия между прогнозными и фактическими связующими входами могут служить для оценки того, насколько далек от оптимума режим работы системы, и могут играть также роль исходных данных для улучшения самих прогнозов.

Приведенные примеры иллюстрируют два основных принципа координации, анализируемые далее в настоящей главе. Сначала мы изучим условия, при которых применение этих принципов приводит к успешной координации, а затем исследуем различные способы поиска координирующих сигналов, обеспечивающих координацию локальных оптимизационных задач. Мы уделим также внимание сравнению указанных принципов с точки зрения их применимости в различных случаях.

2. ДВУХУРОВНЕВАЯ СИСТЕМА ОПТИМИЗАЦИИ

В настоящей главе мы будем заниматься двухуровневой системой, описанной в гл. 4, но добавим следующие более специальные предположения:

1) Влияние внешней среды (возмущения) либо отсутствует, либо полностью известно заранее; т. е. система в целом работает в условиях определенности.

2) Подлежащая решению глобальная задача и задачи, решаемые на нижестоящем уровне, являются оптимизационными задачами.

За исключением этих двух предположений, в остальном рассматриваемая система та же, что и в гл. 4. Но здесь мы стремимся сосредоточиться на структурных проблемах и выявить основополагающие свойства двухуровневой системы, которые не зависят от их конкретных технических особенностей, учитываемых при более детализированных описаниях системы. В силу этого мы считаем, что на решаемые оптимизационные задачи не наложено дополнительных ограничений. Влияние конкретных особенностей системы и наложенных на нее ограничений будет исследовано в последующих главах.

Так как возмущение предполагается известным (фиксированным или отсутствующим), то процесс в целом будет представляться отображением $P: M \rightarrow Y$, а связующие входы и определяются посредством отображения (отношения) $H: M \times Y \rightarrow U$ или $K: M \rightarrow U$, причем последнее отображение определяется через P и H ; при этом подпроцессы описываются отображениями $P_i: M_i \times U_i \rightarrow Y_i$. В связи со вторым предположением наш подход будет строиться на принципе глобальной оптимизации и будет связан с решением проблемы оптимального управления в том виде, как это определено в гл. 3.

Глобальная задача оптимизации

Глобальная задача оптимизации, обозначаемая через $\mathcal{D}$, отражает глобальную цель двухуровневой системы и, как всякая задача оптимизации, определяется, вообще говоря, парой (g, M) , где g — заданная целевая функция. Мы будем предполагать, что g определена с помощью выходной функции процесса в целом P и глобальной функции качества $G: M \times Y \rightarrow V$,

$$g(m) = G(m, P(m)). \quad (5.3)$$

Решением глобальной задачи $\mathcal{D}$ является тогда такое управляющее воздействие $\hat{m} \in M$, что

$$g(\hat{m}) = \min_M g(m). \quad (5.4)$$

Такое управляющее воздействие $\hat{m}$ будет называться *глобально оптимальным* воздействием.

Заметим, что в формальном смысле глобальная оптимизационная задача есть задача оптимизации при отсутствии ограничений. В действительности, однако, M могло бы быть подмножеством более широкого множества, будучи определено как множество таких управляющих воздействий, которые удовлетворяют некоторым предписанным ограничениям; тем не менее предполагается,

что M есть декартово произведение n компонент:

$$M = M_1 \times \dots \times M_n.$$

Локальные оптимизационные задачи

Мы установили в гл. 4, что каждый координирующий сигнал γ , $\gamma \in \mathcal{E}$ определяет подлежащие решению задачи на уровне нижестоящих решающих элементов. Пусть поэтому $\mathcal{D}_i(\gamma)$ означает задачу, решаемую i -м локальным решающим элементом. Решаемые на этом уровне задачи по-прежнему будут оптимизационными задачами; в силу этого предполагается, что i -я локальная задача $\mathcal{D}_i(\gamma)$, $\gamma \in \mathcal{E}$, определяется парой $(g_{i\gamma}, X_{i\gamma})$, где $g_{i\gamma}$ — заданная локальная целевая функция, определяемая на множестве $X_{i\gamma}$, а $X_{i\gamma}$ — заданное подмножество X_i . Мы предполагаем, что $g_{i\gamma}$ определяется посредством выходной функции P_i и локальной функции качества $G_{i\gamma}$:

$$g_{i\gamma}(x_i) = G_{i\gamma}(x_i, P_{i\gamma}(x_i)). \quad (5.5)$$

Решением локальной оптимизационной задачи $\mathcal{D}_i(\gamma)$ является тогда элемент x_i^γ из $X_{i\gamma}$, такой, что

$$g_{i\gamma}(x_i^\gamma) = \min_{X_{i\gamma}} g_{i\gamma}(x_i). \quad (5.6)$$

Такой элемент x_i мы будем называть γ -оптимальным локальным решением или просто оптимальным локальным решением, когда подразумеваемое γ ясно из контекста.

У нас нет никаких оснований для априорного соотнесения величин $P_{i\gamma}$ и $G_{i\gamma}$ с P и G ; тем не менее, принимая во внимание структуру двухуровневой системы, описанную в гл. 4, разумно сделать некоторые предположения на этот счет. Вводимые нами предположения относятся к каждой из локальных оптимизационных задач и сводятся к следующему:

- 1) выходная функция $P_{i\gamma}$ соответствует подпроцессу P_i ; $P_{i\gamma} = P_i$;
- 2) локальная функция качества $G_{i\gamma}$ есть отображение

$$G_{i\gamma}: M_i \times Y_i \times U_i \rightarrow V.$$

Следует заметить, что не предполагается никакой специфической связи между глобальной функцией качества G и локальными функциями качества $G_{i\gamma}$. Локальные оптимизационные задачи $\mathcal{D}_i(\gamma)$ теперь определяются тройкой $(P_i, G_{i\gamma}, X_{i\gamma})$, где $X_{i\gamma}$ есть подмножество $M_i \times U_i$.

Следует отметить, что предположения 1 и 2, сделанные при определении локальных оптимизационных задач, не накладывают особых ограничений и являются довольно естественными.

Прежде всего рассмотрим предположение 1. Использование локальных множеств M_i и Y_i в определении выходной функции

P_{iy} не нуждается в пояснении. Также вполне естественно предположить, что i -й локальный решающий элемент отдает себе отчет в том, что он не может полностью предопределить реализующееся в действительности значение y_i и что имеются некоторые внешние эффекты, которые проще всего объединить в некое дополнительное множество $\mathcal{U}_i$ (принадлежащее области определения выходной функции), что приводит к отображению из $M_i \times \mathcal{U}_i$ в Y_i . Неспособность i -го локального решающего элемента полностью определить реализующееся значение y_i объясняется единственно лишь деятельностью других решающих элементов (оказывающих на него влияние только через взаимодействия). Поэтому естественно предположить, что $\mathcal{U}_i = U_i$. Наконец, мы приходим к наиболее сильно ограничивающему нас предположению, а именно $P_{iy} = P_i$. Реальное значение этого предположения зависит от того, каким образом осуществляется декомпозиция P на подпроцессы. Вообще говоря, если не накладывается никаких ограничений на способ декомпозиции P на подпроцессы, можно сказать, что почти для любого выбора P_{iy} существуют декомпозиция P и соответствующие связующие функции H_i , такие, что $P_{iy} = P_i$; все зависит от выбора связующих функций H_i . Предположение $P_{iy} = P_i$ поэтому в общем случае не является очень сильным ограничением. Тем не менее в последующих рассмотрениях мы часто будем считать H проекционным или, в более широком смысле, линейным отображением. Это предположение будет делаться ради того, чтобы сосредоточиться на центральной проблеме координации, а именно на отношениях между уровнями, а не на точности используемых моделей нижнего уровня, которые нас в настоящей книге меньше интересуют, хотя с технической точки зрения они и являются достаточно важными.

Предположение 2 относительно локальных функций качества G_i не нуждается в обсуждении, так как оно не содержит допущения о наличии какой-либо связи между локальными (G_i) и глобальной (G) функциями качества, о чем уже говорилось выше.

Способы координации и координируемость

Имеются два способа воздействия на локальные задачи оптимизации — через функцию качества G_{iy} и через множество допустимых решений X_{iy} ; следовательно, имеются два способа, которыми координатор может влиять на нижестоящие решающие элементы: один способ, называемый *координацией путем изменения целей*, состоит в изменении локальных функций, а другой способ связан с изменением множества X_{iy} и называется *координацией путем изменения ограничений или «образов»*, так как он подразумевает изменение P_{iy} .

Координация путем изменения целей

Простейший путь описания различных способов координации путем изменения целей — начать с определения n функций $G_{1\beta}$, $G_{n\beta}$,

$$G_{i\beta}: M_i \times Y_i \times U_i \times \mathcal{B} \rightarrow V,$$

где $\mathcal{B}$ — заданное множество. Мы припишем каждому γ из $\mathcal{C}$ свое (единственное) β^γ из $\mathcal{B}$ и получим из $G_{i\beta}$ локальную функцию качества $G_{i\gamma}$

$$G_{i\gamma}(m_i, u_i, y_i) = G_{i\beta}(m_i, y_i, u_i, \beta^\gamma). \quad (5.7)$$

Чтобы подчеркнуть зависимость $G_{i\gamma}$ от β , заменим индекс γ на β , так что $G_{i\beta} = G_{i\gamma}$ и, следовательно, $g_{i\beta} = g_{i\gamma}$, когда $\beta = \beta^\gamma$. Иногда возникает ситуация, когда нет возможностей для координации путем изменения целей. В таких случаях мы будем просто опускать индекс γ (или β) и предполагать, что локальные функции качества суть отображения $G_i: M_i \times U_i \times Y_i \rightarrow V$. Мы будем называть такие локальные функции качества *немодифицированными* в том смысле, что они считаются фиксированными и заданными.

Координация путем изменения ограничений (образов)

Каждое γ из $\mathcal{C}$ определяет для i -го локального решающего элемента множество допустимых решений, которое, вообще говоря, является подмножеством $M_i \times U_i$ согласно предположению 1 относительно выходных функций. Следовательно, множества $X_{i\gamma}$ как собственные подмножества $M_i \times U_i$ представляют собой ограничения, накладываемые на локальные решения. Так как мы не рассматриваем в явном виде задачи оптимизации при наличии ограничений, мы введем третье предположение относительно подлежащих решению локальных задач

$$3) X_{i\gamma} = M_i \times U_i^\gamma,$$

где U_i^γ — заданное подмножество U_i . Тогда координация путем изменения ограничений сводится к выбору соответствующих подмножеств связующих входов; в связи с методами учета взаимодействий, изложенными в гл. 4, мы в этой главе рассмотрим два способа, особенно пригодные для оптимизационных систем.

Способ прогнозирования взаимодействий

Для каждого γ из $\mathcal{C}$ выделяемое множество U^γ есть множество из одного элемента $\{\alpha^\gamma\}$, и, следовательно, i -я локальная задача $\mathcal{Z}_i(\gamma)$ есть задача оптимизации на множестве M_i с заданным (т. е. предсказанным координатором) значением α_i^γ связующего

входа u_i . Для каждого предсказания (прогноза) $\alpha = (\alpha_1, \dots, \alpha_n)$ связующих входов локальные решающие элементы, очевидно, получают новый «образ» остальных подпроцессов; образ для i -го локального решающего элемента есть $P_i^\alpha: M_i \rightarrow Y_i$, где

$$P_i^\alpha(m_i) = P_i(m_i, \alpha_i).$$

Способ «развязывания» взаимодействий

Для каждого γ из $\mathcal{E}$ выпеляемым множеством U^γ является все множество U , и, следовательно, $X_{i\gamma} = X_i \quad M_i \times U_i$ для каждого $i, 1 \leq i \leq n$. Каждая локальная оптимизационная задача полностью определяется независимо от других локальных задач оптимизации и, что более важно, локальные решающие элементы получают инструкцию выбрать оптимальным образом не только локальные управляющие воздействия, но также и локальные связующие входы. Заметим, что несмотря на предположение, что $X_{i\gamma} = M_i \times U_i$, последующий анализ справедлив и для случая, когда $X_{i\gamma}$ есть любое подмножество $M_i \times U_i$, содержащее все пары $(m_i, K_i(m))$, где m принадлежит M .

Координируемость оптимизирующих систем

Для простоты изложения мы на всем протяжении настоящей главы будем пользоваться следующими обозначениями:

1) $\hat{m} \equiv$ глобально оптимальное управляющее воздействие; (5.8)

2) $x^\gamma \equiv$ пара (m^γ, u^γ) , такая, что каждая пара (m_i^γ, u_i^γ) является γ -оптимальной.

Тогда по определению двухуровневая система координируема, если имеет место следующее предложение:

$$(\exists \gamma) (\exists x^\gamma) (\exists \hat{m}) [\pi_M(x^\gamma) = \hat{m}], \quad (5.9)$$

где π_M — проекционное отображение из $M \times U$ на M ; другими словами, система координируема, если существуют γ в $\mathcal{E}$ и γ -оптимальные локальные решения (m_i^γ, u_i^γ) для каждого $i, 1 \leq i \leq n$, такие, что управляющее воздействие $m^\gamma = (m_1^\gamma, \dots, m_n^\gamma)$ является глобально оптимальным, т. е. $m^\gamma = \hat{m}$.

Координирующий сигнал γ , который удовлетворяет предложению (5.9), мы будем называть *оптимальным координирующим сигналом*. Следовательно, двухуровневая система является координируемой тогда и только тогда, когда существует оптимальный координирующий сигнал.

3. ПРИНЦИПЫ КООРДИНАЦИИ И ЗАДАЧА КООРДИНАТОРА

Коль скоро глобальная задача оптимизации считается заданной, а локальные оптимизационные задачи параметризуются в результате подачи координирующих сигналов, то задача координатора состоит фактически в том, чтобы найти оптимальный координирующий сигнал (если таковой существует). Конкретная математическая задача, решаемая координатором, тогда должна быть такой, чтобы ее решение являлось искомым оптимальным координирующим воздействием. Реальная трудность состоит, однако, в конкретизации этой задачи. Можно было бы тривиальным образом выбрать глобальную задачу в качестве задачи, решаемой координирующим элементом, однако при таком подходе нет настоящего разделения труда между решающими элементами различных уровней. Другой подход состоит в использовании постулата совместимости и принципов координации, изложенных в гл. 4; для двухуровневой системы, в которой все решаемые задачи являются оптимизационными, по всей видимости, более всего подходят принципы прогнозирования и согласования, описанные в гл. 4.

Принцип согласования

При использовании принципа согласования способ координации строится на «развязывании» взаимодействий, и поэтому в распоряжении координатора имеется лишь метод координации путем изменения целей¹⁾. Мы проанализируем две частные формы принципа согласования.

Согласование взаимодействий

Эта частная форма принципа согласования уже была рассмотрена в гл. 4; здесь мы можем выразить ее следующим предложением:

$$(\forall \gamma)(\forall x^\gamma)(\exists \tilde{m})\{[(m, u) = x^\gamma \text{ и } K(m) = u] \Rightarrow m = \tilde{m}\}. \quad (5.10)$$

Принцип утверждает, что глобально оптимальное управляющее воздействие обеспечивается оптимальными локальными решениями всякий раз, когда связующие входы согласованы.

Согласование функций качества

При такой форме принципа согласования сравниваются локальные затраты (функции качества), а не сами связующие входы. Пусть для каждого из $\mathcal{C}$ отображение $\bar{g}_\gamma: M \times U \rightarrow V^n$ определяется в виде

$$\bar{g}_\gamma(m, u) = (g_{1\gamma}(m_1, u_1), \dots, g_{n\gamma}(m_n, u_n)). \quad (5.11)$$

¹⁾ Напомним, что «целью» деятельности локальных решающих элементов является отыскание решений, оптимальных по отношению к заданным локальным целевым функциям. Поэтому «изменение целей» в данном случае означает изменение локальных функций качества (целевых функций).—
Прим. ред.

Принцип согласования функций качества выражается тогда предложением

$$(\forall \gamma) (\forall x^\gamma) (\exists m) \{[(m, u) = x^\gamma \text{ и } g_\gamma(m, K(m)) = g_\gamma(m, u)] \Rightarrow m = \bar{m}\}, \quad (5.12)$$

смысл которого сводится к тому, что глобально оптимальное управляющее воздействие складывается из оптимальных локальных решений всякий раз, когда согласованы ожидаемые и фактические локальные затраты.

Принцип прогнозирования

Пусть для данного координирующего сигнала γ из $\mathcal{E}$ m^γ обозначает такое управляющее воздействие из M , что $(m^\gamma, \alpha^\gamma) = x^\gamma$, где α^γ — предсказанное значение связующих входов. Принцип прогнозирования взаимодействий, определяемый соотношением (4.14), может быть тогда выражен предложением

$$(\forall \gamma) (\forall m^\gamma) (\exists \hat{m}) \{[m = m^\gamma \text{ и } K(m) = \alpha^\gamma] \Rightarrow m = \hat{m}\}. \quad (5.13)$$

Эта специальная форма принципа прогнозирования не зависит от того, используется ли координация путем изменения целей: она относится лишь к правильности прогнозирования связующих входов.

Применимость принципа прогнозирования взаимодействий существенно сужается, если не используется координация путем изменений целей: либо шансы правильного прогнозирования малы, либо какие-то правильные предсказания не приводят к глобальному оптимуму. С другой стороны, при координации путем изменения целей точность прогнозирования связующих входов не всегда обеспечивает достижение глобальной оптимальности. Это означает, что необходимо применение более общей формы рассматриваемого принципа.

Для четкого разграничения двух способов координации (прогнозирования взаимодействий и координации с помощью изменения целей) мы будем выражать каждый координирующий сигнал γ как пару (α, β) , где $\alpha = \alpha^\gamma$ — прогнозное значение связующего входа, а $\beta = \beta^\gamma$ — параметр, конкретизирующий локальные функции качества $G_{i\gamma} = G_{i\beta}$. В этом случае множеством возможных координирующих сигналов $\mathcal{E}$ является множество $\mathcal{A} \times \mathcal{B}$, где $\mathcal{A} \subseteq U$, а $\mathcal{B}$ — множество возможных значений β .

Пусть $M \rightarrow \mathcal{B}$ — заданное отображение и

$$q_\eta(m) = (K(m), \eta(m)) \quad (5.14)$$

для всех m из M . Тогда для q_η принцип прогнозирования, задаваемый условием (4.15), выражается предложением

$$(\forall \gamma) (\forall m^\gamma) (\exists \hat{m}) \{[m = m^\gamma \text{ и } q_\eta(m) = \gamma] \Rightarrow m = \hat{m}\}. \quad (5.15)$$

Чтобы указать на применение заданного отображения η , мы будем говорить о принципе прогнозирования (взаимодействий) с использованием отображения η .

Заметим, что если данная двухуровневая система координируема с помощью принципа прогнозирования и применением выбранного отображения η , то область поиска оптимального координирующего сигнала можно ограничить подмножеством $q_\eta(M)$ множества координирующих сигналов $\mathcal{E}$.

4. РАЗРЕШЕНИЕ КОНФЛИКТОВ В ДВУХУРОВНЕВОЙ СИСТЕМЕ

Применим ли тот или иной принцип координации и приводит ли его использование к оптимальному координирующему воздействию, зависит от соотношения между глобальной и локальными целевыми функциями. Вообще говоря, речь фактически идет о зависимости между глобальной и локальными задачами; однако, так как мы ограничиваемся рассмотрением задач оптимизации при отсутствии ограничений, эту зависимость можно выразить через целевые функции. В самом деле, стремление локальных решающих элементов минимизировать свои собственные целевые функции, вообще говоря, не приводит к достижению глобального оптимума. В связи с этим может возникнуть конфликт (несогласованность) между локально принимаемыми решениями. Принципы координации утверждают, что пагубные конфликты разрешаются, когда при принятии локальных решений обеспечивается выполнение некоторых условий «согласования». Однако осуществление таких «согласований» является реально достижимым лишь в том случае, когда система обладает определенными свойствами. В этом разделе мы постараемся описать некоторые основные свойства системы и исследовать их роль в устранении конфликтов. Результаты нашего анализа найдут непосредственное применение в последующих разделах этой главы при изучении условий координируемости.

В общем случае в двухуровневой системе возникают два вида конфликтов: межуровневые и внутриуровневые. *Межуровневый конфликт* есть конфликт между двумя смежными уровнями, или, в общем случае, между двумя различными уровнями в n -уровневой системе. В двухуровневой системе глобальная цель может состоять в достижении минимума глобальных (суммарных) затрат, тогда как локальные цели, возможно, будут сводиться к минимизации локальных затрат. Если глобальная цель и локальные цели не совместимы в том смысле, что достижение минимальных локальных затрат препятствует минимизации суммарных затрат, возникает конфликт между уровнями. *Внутриуровневый конфликт* представляет собой конфликт в пределах отдельного уровня.

Предположим снова, что локальная цель состоит в достижении минимальных локальных затрат. Если локальные цели несовместимы в том смысле, что достижение минимальных локальных затрат одним локальным решающим элементом препятствует другому элементу в достижении минимальных локальных затрат, наличие конфликт внутри данного уровня.

Рассматриваемые в этом разделе специфические свойства систем позволяют устанавливать наличие или отсутствие подобных конфликтов и могут быть использованы не только для характеристики двухуровневых систем и выяснения природы конфликтов, но и в качестве ориентиров для синтеза или видоизменения двухуровневой системы, а также для ее координации. Эти свойства определяются через взаимосвязи между целевыми функциями, отражающими разнообразие целей в двухуровневой системе; поэтому мы будем называть их «целевыми свойствами» двухуровневой системы.

Отправным пунктом при определении целевых свойств являются следующие вспомогательные функции, выраженные через целевые функции данной двухуровневой системы:

1) *Целевая функция глобальных затрат* $g: M \rightarrow V$

2) *Локальные функции затрат* $h_{i\gamma}: M \rightarrow V$.

Для каждого γ из $\mathcal{E}$ и каждого i , $1 \leq i \leq n$, функция $h_{i\gamma}$ задается на M выражением

$$h_{i\gamma}(m) = g_i(m_i, K_i(m)).$$

Для любого управляющего воздействия m из M значение $h_{i\gamma}(m)$ указывает затраты, которые вынужден произвести i -й локальный решающий элемент при подаче локального управляющего воздействия m_i и фактически реализующемся связующем входе $u_i = K_i(m)$.

3) *Межуровневые функции качества* $\psi_\gamma: V^n \rightarrow V$

Для каждого координирующего сигнала γ из $\mathcal{E}$ существует единственное отношение $\Psi \subseteq V^n \times V$

$$\Psi_\gamma = \{((h_{1\gamma}(m), \dots, h_{n\gamma}(m)), g(m)): m \in M\},$$

которое связывает суммарные (глобальные) затраты для любого управляющего воздействия m из M с фактическими локальными затратами $h_{i\gamma}(m)$. Мы называем Ψ_γ межуровневой характеристикой (качества) для фиксированного γ . В общем случае областью определения Ψ_γ является некое подмножество V^n ; это означает, что существуют некоторые n -мерные векторы локальных затрат $(v_1, \dots, v_n)$ в V^n которые не могут иметь места при любом управляющем воздействии. Если Ψ_γ есть функция, то

$$g(m) = \Psi_\gamma(h_{1\gamma}(m), \dots, h_{n\gamma}(m))$$

для всех управляющих воздействий m из M и, следовательно, глобальные (суммарные) затраты являются функцией фактиче-

ских локальных затрат. Поэтому если Ψ_γ есть функция, то мы говорим, что для данного координирующего сигнала γ существует межуровневая функция (качества), обозначаемая через ψ_γ , и мы определяем ее как расширение Ψ_γ на все множество V^n .

4) *Кажущиеся глобальные целевые функции* $g_{\mathcal{C}}: \mathcal{C} \times M \times U \rightarrow V$ Если для каждого координирующего сигнала γ из $\mathcal{C}$ существует межуровневая функция ψ_γ , мы говорим, что существует кажущаяся глобальная целевая функция, и определяем ее как функцию $g_{\mathcal{C}}$ на $\mathcal{C} \times M \times U$,

$$g_{\mathcal{C}}(\gamma, m, u) = \psi_\gamma(g_{1\gamma}(m_1, u_1), g_{n\gamma}(m_n, u_n)). \quad (5.16)$$

Функция $g_{\mathcal{C}}$, если она существует, дает нам суммарные затраты, какими они представляются локальным решающим элементам; она не всегда дает истинные суммарные затраты, потому что учитывает все пары (m, u) из $M \times U$, хотя некоторые из этих пар нарушают условие $u = K(m)$; однако для любого γ из $\mathcal{C}$ и m из M

$$g_{\mathcal{C}}(\gamma, m, K(m)) = g(m).$$

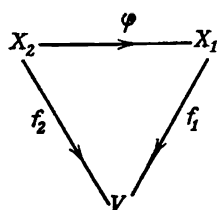
Поэтому $g_{\mathcal{C}}(\gamma, m, u)$ представляет истинные суммарные затраты всякий раз, когда $u = K(m)$.

Свойство монотонности

Существование межуровневых функций и их свойства помогают выявить некоторые существенные характеристики данной двухуровневой системы. Особый интерес как с принципиальной, так и с формальной стороны представляет свойство «монотонности», которое мы сейчас введем и проиллюстрируем на нескольких примерах.

Напомним прежде всего обычное определение монотонной функции. Функция f , отображающая одно частично упорядоченное множество на другое частично упорядоченное множество, является: 1) сохраняющей порядок, или монотонной, если $f(x) \leq f(x')$ при $x \leq x'$ и 2) строго сохраняющей порядок, или строго монотонной, если $f(x) < f(x')$ при $x < x'$. Мы будем использовать более короткий из двух терминов — «монотонность». Это понятие охватывает и тот случай, когда область определения f является подмножеством n -мерного пространства.

Два отображения f_1 и f_2 , имеющие общую область определения и принимающие значения из одного и того же частично упорядоченного множества, определяются как *монотонно связанные*, если значения одного отображения соотнесены монотонным образом с соответствующими значениями другого отображения; точ-



Фиг. 5.2.

нее, f_1 монотонно связана с f_2 , если $f_1(x) \leq f_1(x')$ при $f_2(x) \leq f_2(x')$, и f_1 строго монотонно связана с f_2 , если $f_1(x) < f_1(x')$ при $f_2(x) < f_2(x')$.

Сделаем некоторые важные замечания относительно функций, являющихся монотонно связанными.

1. Отношение между функциями рефлексивно, но не симметрично. Любая функция монотонно связана сама с собой. Тем не менее если данная функция f_1 монотонно связана с другой функцией f_2 , то это вовсе не означает, что f_2 обязательно монотонно связана с f_1 .

2. Функция f_1 монотонно связана с функцией f_2 , если существует такая монотонная функция φ , переводящая $\mathcal{R}(f_2)$ в $\mathcal{R}(f_1)$, что $f_1(x) = \varphi(f_2(x))$ для всех x в соответствующих областях определения, где $\mathcal{R}(f_1)$ и $\mathcal{R}(f_2)$ — области значений функций f_1 и f_2 . Этот результат следует непосредственно из свойств рефлексивности и антисимметричности частичных упорядочений. Кроме того, f_1 строго монотонно связана с f_2 , если функция φ является однозначной.

3. Две заданные функции монотонно связаны друг с другом, если они строго монотонно связаны друг с другом.

Согласно замечанию 2, то обстоятельство, что функция f_1 монотонно связана с функцией f_2 , характеризуется существованием монотонной функции φ , отображающей $\mathcal{R}(f_2)$ на $\mathcal{R}(f_1)$, а диаграмма фиг. 5.2 отражает коммутативный характер этих связей. Если $\mathcal{R}(f_2)$ — собственное подмножество частично упорядоченного множества V , то функцию φ можно продолжить, расширив ее до некоторой функции на V ; продолжение это, однако, обязано быть монотонным только на $\mathcal{R}(f_2)$.

Используем теперь понятие монотонности для того, чтобы определить некоторое характерное свойство двухуровневой системы. Данная двухуровневая система обладает *свойством монотонности*, если ее межуровневые отношения представляют собой монотонные функции. Если данная двухуровневая система обладает *свойством монотонности*, мы будем, учитывая предшествующие замечания, предполагать, что для каждого координирующего сигнала γ из $\mathcal{E}$ существует монотонная функция ψ_γ , переводящая V^n в V , и выступающая в качестве межуровневой функции для этого координирующего сигнала.

В следующей теореме выражен важный аспект двухуровневых систем, обладающих свойством монотонности.

Теорема 5.1

Предположим, что данная двухуровневая система обладает свойством монотонности. Пусть γ — заданный координирующий

сигнал из $\mathcal{E}$. Тогда

$$g(m) = \min g$$

всякий раз, когда управляющее воздействие m из M таково, что

$$h_{i\gamma}(m) = \min h_{i\gamma}, \quad i = 1, \quad n.$$

Доказательство. Так как система обладает свойством монотонности, условие

$$g(m) = \psi_{\gamma}(h_{1\gamma}(m), \quad h_{n\gamma}(m))$$

выполняется для каждого управляющего воздействия m из M , когда ψ_{γ} есть межуровневая функция. Так как ψ_{γ} монотонна, то отсюда непосредственно вытекает содержащееся в теореме утверждение.

Если двухуровневая система обладает свойством монотонности, то глобальная целевая функция монотонно связана с локальными функциями затрат; в этом случае наверняка отсутствует внутриуровневый конфликт: уменьшение каждой из фактических локальных затрат не вызывает увеличения суммарных затрат. В действительности уменьшение каждой из фактических локальных затрат будет в свою очередь вызывать уменьшение суммарных затрат по всей системе каждый раз, когда отношение монотонности является строгим. При стабильном уменьшении фактических локальных затрат суммарные затраты поэтому будут стремиться к минимуму.

Пример 5.1

Пусть $V = R$, т. е. V является множеством действительных чисел, и для каждого γ из $\mathcal{E}$ пусть $\psi_{\gamma} = \psi: R^n \rightarrow R$ является межуровневой функцией для данной двухуровневой системы.

1) Предположим, что $\psi(v) = \sum_{i=1}^n a_i v_i$. Если $a_i \geq 0$ для каждого i , $1 \leq i \leq n$, то ψ монотонна и система обладает свойством монотонности. Если $a_i > 0$ для каждого i , $1 \leq i \leq n$, то ψ строго монотонна и система обладает свойством строгой монотонности.

2) Предположим, что $\psi(v) = |\sum_{i=1}^n v_i|$. Тогда ψ не монотонна; однако если $h_{i\gamma} \geq 0$ для каждого i , $1 \leq i \leq n$, и для всех γ из $\mathcal{E}$, то система обладает свойством монотонности.

3) Предположим, что $\psi(v) = v_1 \times v_2 \times \dots \times v_n$. Тогда система обладает свойством монотонности, если $h_{i\gamma} \geq 0$ для каждого $1 \leq i \leq n$, и для каждого γ из $\mathcal{E}$.

4) В общем случае, если $\psi_i: R^n \rightarrow R$ монотонна, $i = 1, \dots, p$, и $\psi_0: R^p \rightarrow R$ монотонна, то $\psi(v) = \psi_0(\psi_1(v), \dots, \psi_p(v))$ также монотонна.

5) В общем случае если ψ дифференцируема, то она монотонна при условии, что все ее частные производные всюду неотрицательны.

Внутриуровневая и межуровневая согласованность

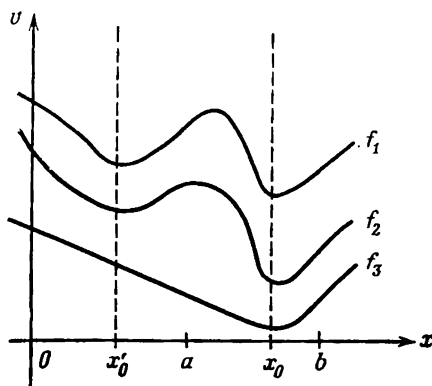
Имеются все основания считать, что решающие элементы двухуровневой системы находятся в состоянии некоторого рода согласованности, если все они могут одновременно достичь каждый своей собственной цели, несмотря на взаимодействия между ними. В случае оптимизирующих систем согласованность можно охарактеризовать как одновременное достижение всеми элементами оптимальных значений целевых функций.

Если данная двухуровневая система обладает свойством монотонности, то естественно обратиться к теореме 5.1 и воспользоваться ее результатами; но при этом у нас нет гарантий существования управляющего воздействия, которое одновременно минимизировало бы каждую из функций локальных затрат. Если такое управляющее воздействие существует, естественно было бы предположить, что такая система и ее локальные функции затрат $h_{i\gamma}$, $1 \leq i \leq n$, обладают внутриуровневой согласованностью. Для того чтобы данная двухуровневая система имела характеристики, выражаемые теоремой 5.1, свойство монотонности является достаточным, но не необходимым. В самом деле, результат теоремы 5.1 справедлив для некоторых двухуровневых систем, которые не обладают свойством монотонности; в связи с этим разумно было бы обобщить этот факт и считать, что в любой двухуровневой системе, наделенной таким свойством, существует согласованность между уровнями.

Введем теперь два определения согласованности, сформулировав их сначала в довольно абстрактной форме, а затем связав эти понятия с согласованностью для двухуровневой системы.

Семейство функций f_i , $1 \leq i \leq n$, обладает свойством *согласованности*, если существует общий элемент, который принадлежит к области определения каждой из функций f_i и минимизирует каждую функцию f_i в ее области определения. Такой элемент мы будем называть *элементом, согласующим функции f_i* , или же просто *согласующим элементом*, если из контекста ясно, о каком семействе функций идет речь. Однако возможны случаи, когда функции f_i не обладают общими элементами: тогда они не могут быть согласованными; с другой стороны, их области определения могут оказаться одним и тем же одноэлементным множеством; в этом случае семейство тривиальным образом обладает свойством согласованности.

Семейство функций f_i , $1 \leq i \leq n$, является *согласованным с функцией f* , если каждый элемент, согласующий функции f_i ,



Ф и г. 5.3.

содержится в области определения f и минимизирует f на ее области определения. Если семейство не обладает свойством согласованности, оно, очевидно, не согласовано ни с какой функцией.

Пример 5.2

Рассмотрим функции f_i , графики которых представлены на фиг. 5.3. Семейство $\{f_1, f_2\}$ обладает согласованностью (согласующим элементом для областей определения рассматриваемых функций будет x_0) и находится в согласованности с f_3 .

Применим теперь эти абстрактные понятия для описания различных видов согласованности, типичных для двухуровневой системы.

Безусловная согласованность

Координирующий сигнал γ из $\mathcal{E}$ приводит к *безусловной локальной согласованности*, если соответствующее семейство локальных функций затрат $h_{i\gamma}$, $1 \leq i \leq n$, обладает свойством согласованности. Следовательно, координирующий сигнал γ влечет за собой безусловную локальную согласованность, если в M существует такое управляющее воздействие, которое минимизирует каждую из локальных функций затрат $h_{i\gamma}$ на M .

Координирующий сигнал γ из $\mathcal{E}$ порождает *безусловную межуровневую согласованность*, если соответствующее семейство локальных функций затрат $h_{i\gamma}$, $1 \leq i \leq n$, оказывается согласованным с глобальной целевой функцией. Мы говорим, что двухуровневая система обладает *безусловной межуровневой согласованностью*, если каждый координирующий сигнал приводит к такой согласованности. Двухуровневая система обладает безусловной межуровневой согласованностью, если определенное управляющее воздействие m из M , минимизирующее суммарные (глобальные) затраты, одновременно минимизирует и локальные функции затрат $h_{i\gamma}$, $1 \leq i \leq n$, при некотором координирующем сигнале γ из $\mathcal{E}$.

Теорема 5.2

Если рассматриваемая двухуровневая система обладает свойством монотонности, она обладает также межуровневой согласованностью.

Доказательство этой теоремы очевидно. Однако обратное утверждение не всегда справедливо. Двухуровневая система может обладать межуровневой согласованностью и при отсутствии свойства монотонности. Покажем это на примере.

Пример 5.3

Предположим, что двухуровневая система такова, что $V \cap R$ и $M = M_1 \times M_2 = R^2$. Предположим также, что

$$g(m) = 2 \sin^2(m_1^2 + m_2^2) + k(m_1^2 + m_2^2)$$

есть заданная глобальная целевая функция, где $k > 0$ и для каждого γ из $\mathcal{E}$ локальные функции затрат $h_{i\gamma}$ имеют вид

$$h_{1\gamma}(m) = m_1^2 + \beta_1^\gamma m_2, \quad h_{2\gamma}(m) = m_2^2 + \beta_2^\gamma m_1.$$

Если координирующий сигнал γ таков, что $\beta_1^\gamma = \beta_2^\gamma = 0$, то γ приводит к локальной согласованности, так как $m = 0$ минимизирует как $h_{1\gamma}$, так и $h_{2\gamma}$. Никакое другое γ не обеспечивает локальной согласованности. Система не обладает свойством монотонности, но она обладает межуровневой согласованностью, так как $g(m)$ принимает минимальное значение в точке $m = 0$, причем $m = 0$ — единственное управляющее воздействие, минимизирующее как $h_{1\gamma}$, так и $h_{2\gamma}$, когда γ таково, что $\beta_1^\gamma = \beta_2^\gamma = 0$.

Ограниченная согласованность

Для существования координирующего сигнала γ , создающего локальную согласованность, требуется наличие управляющего воздействия из множества M , которое одновременно минимизирует каждую из локальных функций затрат $h_{i\gamma}$, $1 \leq i \leq n$, на всем множестве M управляющих воздействий. Если не существует такого управляющего воздействия для какого-либо координирующего входа системы, едва ли много пользы принесет то обстоятельство, что двухуровневая система имеет межуровневую согласованность. Однако мы могли бы в этом случае подходящим образом ограничить область минимизации, чтобы достигалась также и локальная согласованность.

В связи с этим мы рассмотрим несколько способов определения ограниченной согласованности для двухуровневой системы в зависимости от того, какие ограничения налагаются на подмножества, в которых производится поиск локального оптимума. Особое внимание мы уделим ограничениям, порождаемым связующими входами; для заданного связующего входа u_i мы рассмотрим минимизацию функции $h_{i\gamma}(m)$ при условии $K_i(m) = u_i$, так как $h_{i\gamma}(m) = g_{i\gamma}(m_i, u_i)$, когда $K_i(m) = u_i$.

Пусть для каждого i из M и каждого i , $1 \leq i \leq n$, $[m]_{K_i}$ обозначает класс всех m' из M , эквивалентных m в том смысле,

что $K_i(m') = K_i(m)$. Очевидно, что каждое m из M содержится в множествах $[m]_{K_1}, \dots, [m]_{K_n}$, которые оно порождает; кроме того, если K — взаимно однозначное отображение, то это и есть единственное управляющее воздействие m , общее для всех этих множеств. Для любого i , $1 \leq i \leq n$, и m из M пусть $h_{iy}^{(m)}$ обозначает ограничение для h_{iy} на множестве $[m]_{K_i}$. Введем следующие определения ограниченной согласованности в двухуровневой системе.

Координирующий сигнал y из $\mathcal{C}$ влечет за собой *ограниченную локальную согласованность*, если существует такое управляющее воздействие m из M , что семейство ограниченных локальных функций затрат $h_{iy}^{(m)}$, $1 \leq i \leq n$, обладает свойством согласованности. Очевидно, что если для данного m из M семейство $\{h_{iy}^{(m)}, h_{ny}^{(m)}\}$ обладает согласованностью, то в случае существования однозначной функции K данное m есть единственное управляющее воздействие, которое согласует ограниченные локальные функции затрат $h_{iy}^{(m)}$, $1 \leq i \leq n$.

Координирующий сигнал y из $\mathcal{C}$ влечет за собой *ограниченную межуровневую согласованность*, если для всех управляющих воздействий m из M семейство ограниченных локальных функций затрат $h_{iy}^{(m)}$, $1 \leq i \leq n$, оказывается согласованным с глобальной целевой функцией. Мы говорим, что двухуровневая система *обладает ограниченной межуровневой согласованностью*, если каждый координирующий сигнал в системе порождает ограниченную межуровневую согласованность.

Между безусловной и ограниченной согласованностью существует взаимосвязь. Если семейство локальных функций затрат h_{iy} , $1 \leq i \leq n$, обладает свойством согласованности, то по определению существует управляющее воздействие из M , которое приводит к согласованности функций h_{iy} , и для любого такого управляющего воздействия семейство ограниченных локальных функций затрат $h_{iy}^{(m)}$, $1 \leq i \leq n$, является согласованным. Поэтому всякий координирующий сигнал, приводящий к безусловной локальной согласованности, влечет за собой также и ограниченную локальную согласованность. Исходя из этого, мы можем заключить, что двухуровневая система, обладающая ограниченной межуровневой согласованностью, обладает также и безусловной межуровневой согласованностью. Обратное, однако, не всегда справедливо, как показано на следующем примере.

Пример 5.4

Предположим, что имеется двухуровневая система, в которой $V = R$ и $M = M_1 \times M_2 = R^2$. Пусть локальные целевые функции g_1 и g_2 определены на R^2 в виде

$$\begin{aligned} g_1(m_1, u_1) &= m_1^2 + (m_1 + u_1 - 1)^2, \\ g_2(m_2, u_2) &= m_2^2 + (2m_2 + u_2 - 2)^2, \end{aligned}$$

и допустим, что связующие входы в объединенной системе имеют значения $u_1 = 0$ и $u_2 = m_1$. Тогда соответствующие локальные функции затрат будут иметь вид $h_1(m) = g_1(m_1, 0)$ и $h_2(m) = g_2(m_2, m_1)$. Выберем глобальную целевую функцию g в виде

$$g(m) = h_1(m) + h_2(m).$$

Система обладает свойством монотонности и, следовательно, безусловной межуровневой согласованностью. Покажем, что в ней отсутствует ограниченная межуровневая согласованность.

Заметим, что функции взаимодействия K_1 и K_2 для этой системы имеют вид $K_1 \equiv 0$ и $K_2(m) = m_1$. Рассмотрим управляющее воздействие $\tilde{m} = (1/2, 3/5)$. Управляющее воздействие $\tilde{m}$ согласует функции $h_1^{(\tilde{m})}$ и $h_2^{(\tilde{m})}$, где

$$[\tilde{m}]_{K_1} = R^2, \quad [\tilde{m}]_{K_2} = \{1/2\} \times R,$$

и поэтому семейство $\{h_1, h_2\}$ обладает ограниченной локальной согласованностью. Далее, осуществляя минимизацию g , находим, что единственное глобальное оптимальное управление есть $\hat{m} = (7/11, 6/11)$. Согласующее управляющее воздействие $\tilde{m}$ в этом случае не является глобально оптимальным; это означает, что данная система не обладает ограниченной межуровневой согласованностью.

Фиг. 5.3 помогает нам разобраться в рассматриваемой ситуации. Заметим, что функции f_1 и f_2 согласованы с f_3 ; точка x_0 есть единственная точка, в которой достигается минимум как f_1 , так и f_2 на их области определения $[0, b]$; в ней же достигается и минимум f_3 . Однако ограничения, связанные с сужением области определения f_1 и f_2 до сегмента $[0, a]$, разрушают их согласованность с функцией f_3 ; хотя в точке x'_0 имеет место минимум f_1 и f_2 на $[0, a]$, эта точка не дает минимума f_3 .

Использование свойств целевых функций при разрешении конфликта

Изложим теперь некоторые из главных выводов относительно двухуровневых систем, обладающих свойствами монотонности и согласованности (в том или ином смысле), которые мы в дальнейшем будем называть целевыми свойствами. Наша задача состоит в том, чтобы выяснить, каким образом возможность разрешения конфликта между глобальными и локальными целями в данной двухуровневой системе зависит от наличия или отсутствия некоторых свойств целевых функций.

Согласованность и разрешение конфликта

Для каждого координирующего сигнала γ из $\mathcal{C}$ пусть (m^γ, u^γ) будет парой из $M \times U^\gamma$, где $U^\gamma = U_1^\gamma \times \dots \times U_n^\gamma$.

так что для каждого $1 \leq i \leq n$, пара (m_i^γ, u_i^γ) принадлежит $M_i \times U_i^\gamma$ и минимизирует соответствующую i -ю локальную целевую функцию $g_{i\gamma}$ на множестве $M_i \times U_i^\gamma$. Мы будем предполагать, что для каждого координирующего сигнала γ из $\mathcal{C}$ существует (хотя и не обязательно единственная) пара (m^γ, u^γ) . В следующих теоремах формулируются характерные признаки двухуровневых систем, обладающих различными видами безусловной согласованности.

Теорема 5.3

Если двухуровневая система обладает безусловной межуровневой согласованностью и $K(M) \subseteq U^\gamma$ для каждого координирующего сигнала γ из $\mathcal{C}$, то m^γ есть глобально оптимальное управляющее воздействие всякий раз, когда $u^\gamma \in K(m^\gamma)$.

Доказательство. Пусть γ — произвольный элемент из $\mathcal{C}$, и предположим, что $u^\gamma = K(m^\gamma)$; в этом случае для всех $m \in M$ и для каждого i , $1 \leq i \leq n$,

$$h_{i\gamma}(m^\gamma) = g_{i\gamma}(m_i^\gamma, u_i^\gamma) \leq g_{i\gamma}(m_i, K_i(m)) = h_{i\gamma}(m).$$

Тогда, согласно определению неограниченной межуровневой согласованности, любое m из M , одновременно минимизирующее все локальные функции затрат $h_{i\gamma}$, $1 \leq i \leq n$, на M , минимизирует также и глобальную целевую функцию. Этим и доказывается содержащееся в теореме утверждение.

Следствие

Предположим, что для всех координирующих сигналов γ из $\mathcal{C}$ двухуровневая система такова, что $K(M) \subseteq U^\gamma$. Тогда для выбранного координирующего сигнала γ из $\mathcal{C}$ существует такая пара (m^γ, u^γ) , что $u^\gamma \in K(m^\gamma)$ только в том случае, если приводит к ограниченной локальной согласованности.

В следующей теореме делаются более сильные утверждения свойствах неограниченной согласованности.

Теорема 5.4

Предположим, что в данной двухуровневой системе наряду с условием $K(M) \subseteq U^\gamma$ для каждого координирующего сигнала γ из $\mathcal{C}$ существует еще и пара $(m_i^\gamma, u_i^\gamma) = (m_i, K_i(m))$, $1 \leq i \leq n$, для некоторого m из M . Тогда

1) Для того чтобы m^γ было глобально оптимальным всякий раз, когда $u^\gamma \in K(m^\gamma)$, необходимо и достаточно, чтобы имела место безусловная межуровневая согласованность.

2) Существование координирующего сигнала γ из $\mathcal{E}$ обеспечивающего безусловную локальную согласованность, является необходимым и достаточным условием для существования такой пары значений (m^γ, u^γ) , что $u^\gamma = K(m^\gamma)$.

Доказательство. Достаточность условия 1 доказана выше. Чтобы доказать необходимость этого условия, предположим, что γ — произвольный элемент из $\mathcal{E}$, а управление $\hat{m}$ минимизирует каждую из локальных функций затрат $h_{i\gamma}$, $1 \leq i \leq n$, на M . Тогда в силу этого предположения существует такая пара значений (m_i^γ, u_i^γ) для каждого i , $1 \leq i \leq n$, что $(m_i^\gamma, u_i^\gamma) = (\hat{m}_i, K_i(\hat{m}))$. Следовательно, $\hat{m} = m^\gamma$ и $u^\gamma = K(m^\gamma)$. Если теперь m^γ не является глобально оптимальным управлением, то безусловная межуровневая согласованность невозможна, что и доказывает необходимость условия 1. Таким же образом доказывается и достаточность условия 2, что завершает доказательство теоремы.

Обратимся теперь к свойствам ограниченной согласованности.

Теорема 5.5

Если двухуровневая система обладает ограниченной межуровневой согласованностью и $U^\gamma = \{\alpha^\gamma\}$ для каждого координирующего сигнала γ из $\mathcal{E}$, то m^γ будет глобально оптимальным воздействием всякий раз, когда $K(m^\gamma) = \alpha^\gamma$.

Доказательство. Пусть γ — произвольный элемент из $\mathcal{E}$ и предположим, что $\alpha^\gamma = K(m^\gamma)$. Тогда при каждом i , $1 \leq i \leq n$, и m из $[m^\gamma]_{K_i}$ имеем $K_i(m) = \alpha_i^\gamma$,

$$h_{i\gamma}(m^\gamma) = g_{i\gamma}(m_i^\gamma, \alpha_i^\gamma) \leq g_{i\gamma}(m_i, \alpha_i^\gamma) = g_{i\gamma}(m).$$

Поэтому управление m^γ согласует ограниченные локальные функции затрат $h_{i\gamma}(m^\gamma)$ и в соответствии с определением ограниченной межуровневой согласованности оно само является глобально оптимальным управляющим воздействием. Тем самым теорема доказана.

Следствие

Предположим, что при любых координирующих сигналах γ из $\mathcal{E}$ данная двухуровневая система такова, что $U^\gamma = \{\alpha^\gamma\}$. Тогда для выбранного координирующего сигнала γ из $\mathcal{E}$ существует такое управление m^γ , что $K(m^\gamma) = \alpha^\gamma$ только тогда, когда сигнал γ обеспечивает ограниченную локальную согласованность.

В теореме 5.5 и ее следствии мы нигде явно не указывали, использовалась ли координация путем изменения целей, хотя мы явным образом пользовались методом прогнозирования взаимодействий. Если мы убедимся в том, что оба способа могут быть

использованы независимо, мы сможем получить более сильные результаты относительно свойств ограниченной согласованности. Следующая теорема сформулирована для случая, когда координация путем изменения целей не используется; однако ее легко обобщить и на тот случай, когда помимо прогнозирования взаимодействий применяется также координация целей.

Теорема 5.6

Пусть для данной двухуровневой системы координация целей не используется; поэтому локальные целевые функции не будут зависеть от координирующих сигналов. Предположим, что для каждого γ из $\mathcal{C}$ $U^\gamma = \{\alpha^\gamma\}$ и что для каждого $1 \leq i \leq n$, существует такое m_i^γ , что $(m_i^\gamma, \alpha_i^\gamma) = (m_i, K_i(m))$ для некоторого m из M . Кроме того, предположим, что $K(M) \subseteq \{\alpha^\gamma: \gamma \in \mathcal{C}\}$. Тогда

1) Для того чтобы m^γ было глобально оптимальным, необходимо и достаточно, чтобы имела место ограниченная межуровневая согласованность всякий раз, когда $K(m^\gamma) = \alpha^\gamma$.

2) Ограниченная локальная согласованность является необходимым и достаточным условием для существования таких α^γ и m^γ , что $K(m^\gamma) = \alpha^\gamma$.

Доказательство. Достаточность условия 1 вытекает из теоремы 5.5. Чтобы доказать его необходимость, предположим, что m согласует ограниченные локальные функции затрат $h_i^{(m)}$, $1 \leq i \leq n$, которые не зависят от координирующих сигналов. Пусть $\hat{\alpha} = K(\hat{m})$; тогда, поскольку $K(M) \subseteq \{\alpha^\gamma: \gamma \in \mathcal{C}\}$ существует γ в $\mathcal{C}$, такое, что $\alpha^\gamma = \hat{\alpha}$. Согласно предположению, существует $\hat{m}_i^\gamma = \hat{m}_i$ для каждого i , $1 \leq i \leq n$. Следовательно, $m^\gamma = \hat{m}$ и $K(m^\gamma) = \alpha^\gamma$. Если теперь $\hat{m}$ не является глобально оптимальным, ограниченная межуровневая согласованность тем самым исключается. Этим доказывается необходимость условия 1 и достаточность условия 2, что и завершает доказательство теоремы.

Монотонность и разрешение конфликтов

Оставим на время свойства согласованности и займемся более специальным целевым свойством — монотонностью. Теорема 5.2 гласит, что свойство монотонности является более сильным, чем межуровневая согласованность. Монотонность есть глобальное свойство, тогда как межуровневая согласованность есть точечное (локальное) свойство. В силу этого можно ожидать, что мы получим более интересные результаты, если обратимся к анализу двухуровневых систем, обладающих свойством монотонности,

Следующая теорема касается свойств двухуровневой системы, обладающей кажущейся глобальной целевой функцией, но не обладающей свойством монотонности.

Теорема 5.7

Предположим, что двухуровневая система имеет кажущуюся глобальную целевую функцию $g_{\mathcal{E}}$, а координирующий сигнал γ из $\mathcal{E}$ таков, что

1) пара (m^v, u^v) существует и U^v содержит $\hat{u} = K(\hat{m})$ для некоторого глобально оптимального управляющего воздействия $\hat{m}$;

2) $m = m^v$ всякий раз, когда $g_{\mathcal{E}}(\gamma, m, u) = g_{\mathcal{E}}(\gamma, m^v, u^v)$ и (m, u) принадлежит $M \times U^v$

Тогда m^v является глобально оптимальным, если

$$g_{\mathcal{E}}(\gamma, m^v, u^v) = \min_M g(m). \quad (5.17)$$

Доказательство. По определению $g_{\mathcal{E}}(\gamma, m, K(m)) = g(m)$ для любого m из M . В частности, $g_{\mathcal{E}}(\gamma, \hat{m}, K(\hat{m})) = g(\hat{m})$, где $\hat{m}$ является глобально оптимальным, а $\hat{u} = K(\hat{m})$ принадлежит U^v . Следовательно, если $g_{\mathcal{E}}(\gamma, m^v, u^v)$, мы из предположения 2 получаем, что $m^v = \hat{m}$. Этим доказательство завершается.

Равенство (5.17) в теореме 5.7 дает условие, при котором координирующий сигнал γ будет оптимальным, если выполняются некоторые другие условия. Указанные условия требуют, чтобы существовали оптимальные локальные решения, удовлетворялось предположение о единственности в п. 2 и чтобы так называемый глобально оптимальный связующий вход содержался в множестве U^v . Чтобы проверить указанное выше равенство, необходимо знать минимальные глобальные затраты $\hat{v}$. Однако, даже если известно $\hat{v}$, мы все еще не имеем указаний на то, как обращаться с кажущейся глобальной целевой функцией при поиске оптимального координирующего сигнала.

Если двухуровневая система обладает свойством монотонности, она обладает тем самым некоторыми особенностями, которые могут быть использованы при поиске оптимального координирующего сигнала.

Теорема 5.8

Предположим, что двухуровневая система обладает свойством монотонности и, следовательно, кажущейся глобальной целевой

функцией $g_{\mathcal{E}}$. Тогда для любого γ из $\mathcal{E}$ неравенство

$$\inf_{\gamma} \inf_M g_{\mathcal{E}}(\gamma, m, u) \leq \inf_M g(m) \quad (5.18)$$

имеет место всякий раз, когда

$K(M) \subseteq U^v$ или когда существует глобально оптимальное управление $\hat{m}$ и $\hat{u}$

$K(\hat{m})$ принадлежит U^v

Доказательство.

Пусть задано γ из $\mathcal{E}$ и U^v содержит $\hat{u}$

$K(\hat{m})$ для некоторого глобально оптимального управляющего воздействия $\hat{m}$. Так как пара $(\hat{m}, \hat{u})$ принадлежит $M \times U^v$ и $g_{\mathcal{E}}(\gamma, \hat{m}, \hat{u}) = g(\hat{m})$, то отсюда сразу следует неравенство (5.18).

Предположим теперь, что $K(M) \subseteq U^v$. Тогда в соответствии определением кажущейся глобальной целевой функции

$$\inf_{\gamma} \inf_M g_{\mathcal{E}}(\gamma, m, u) \leq \inf_M g_{\mathcal{E}}(\gamma, m, K(m)) = \inf_M g(m).$$

Этим доказательство теоремы завершается.

Смысл неравенства (5.18) в том, что если локальные решающие элементы предполагаются «развязанными», т. е. каждый элемент свободен в выборе связующих входов, содействующих достижению его собственных целей, то кажущиеся глобальные затраты будут ниже фактически реализуемых минимальных глобальных затрат или же равны им. Действительно, как видно из фиг. 5.4, поверхность $v(\gamma) = \inf \{g_{\mathcal{E}}(\gamma, m, u) \mid (m, u) \in M \times U^v\}$ всегда расположена ниже гиперплоскости $v = \hat{v}$.

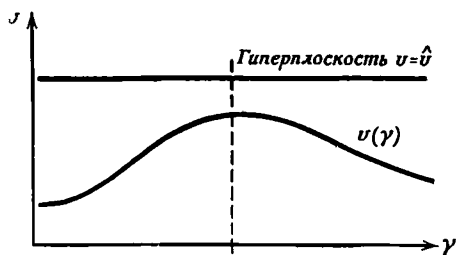
Из этого сразу вытекает важное следствие.

Следствие

Предположим, что двухуровневая система обладает свойством монотонности и, следовательно, имеет кажущуюся глобальную целевую функцию $g_{\mathcal{E}}$. Пусть для каждого γ из $\mathcal{E}$ либо $K(M) \subseteq U^v$, либо U^v содержит $\hat{u} = K(\hat{m})$ для некоторого глобально оптимального управляющего воздействия $\hat{m}$. Тогда

$$\sup_{\mathcal{E}} \inf_{\gamma} \inf_M g_{\mathcal{E}}(\gamma, m, u) \leq \inf_M g(m).$$

Это неравенство иллюстрируется фиг. 5.4. Смысл его в том, что координатор, выбирая γ из $\mathcal{E}$, для того чтобы максимизировать



Фиг. 5.4. Относительное расположение поверхности $v(\gamma) = \inf \{g_{\mathcal{E}}(\gamma, m, u) \mid (m, u) \in M \times U^v\}$ и гиперплоскости $v = \hat{v}$.

$v(\gamma)$, все-таки оставляет минимальные кажущиеся глобальными затраты ниже минимальных фактических затрат.

Теорема 5.7 указывает способ проверки равенства (5.17) и дает условия, при которых оно может быть использовано для установления оптимальности координирующего сигнала. Следствие из теоремы 5.8 указывает, как нужно искать оптимальную координацию. Вместе взятые, они приводят к следующему условию достаточности для координируемости.

Теорема 5.9

Предположим, что двухуровневая система обладает свойством монотонности и, следовательно, имеет кажущуюся глобальную целевую функцию $g_{\mathcal{C}}$. Предположим также, что для каждого $\gamma \in \mathcal{C}$ либо $K(M) \subseteq U^\gamma$ либо U^γ содержит $\hat{u} = K(\hat{m})$ при некотором глобально оптимальном управляющем воздействии $\hat{m}$. Тогда, если межуровневые функции системы строго монотонны, равенство

$$\max_{\mathcal{C}} \min_{U^\gamma} \min_M g_{\mathcal{C}}(\gamma, m, u) = \min_M g(m) \quad (5.19)$$

является достаточным условием координируемости.

Доказательство. Предположим, что выполняется условие (5.19); тогда существует γ из $\mathcal{C}$ такое, что

$$\min_{U^\gamma} \min_M g_{\mathcal{C}}(\gamma, m, u) = g(\hat{m}) = g_{\mathcal{C}}(\gamma, \hat{m}, K(\hat{m})), \quad (5.20)$$

где $\hat{m}$ — глобально оптимальное управляющее воздействие, такое, что $\hat{u} = K(\hat{m})$ принадлежит U^γ . Так как межуровневые функции строго монотонны, то пара (m^γ, u^γ) существует, и для каждого i , $1 \leq i \leq n$, выполняется неравенство $g_{i\gamma}(m_i^\gamma, u_i^\gamma) \leq g_{i\gamma}(\hat{m}_i, K_i(\hat{m}))$. Предположим, что $g_{j\gamma}(m_j^\gamma, u_j^\gamma) < g_{j\gamma}(\hat{m}_j, K_j(\hat{m}))$ для некоторого j , $1 \leq j \leq n$. Тогда в силу строгой монотонности условие (5.20) уже не будет выполняться. Поэтому для каждого i , $1 \leq i \leq n$, пара $(\hat{m}_i, K_i(\hat{m}))$ минимизирует $g_{i\gamma}$ на множестве $M_i \times U_i^\gamma$. Следовательно, система координируема и тем самым теорема доказана.

Наряду с требованием строгой монотонности межуровневых функций существует еще одно затруднение: минимальные глобальные затраты могут быть неизвестны и нельзя будет воспользоваться условием (5.19). Тем не менее если кажущаяся глобальная целевая функция обладает свойством, определяемым ниже-следующей теоремой, мы в качестве достаточного условия координируемости можем использовать условие наличия седловой точки.

Теорема 5.10

Предположим, что двухуровневая система обладает свойством монотонности и ее кажущаяся глобальная целевая функция $g_{\mathcal{C}}$

удовлетворяет неравенству

$$g(m) \leq \sup_{\mathcal{C}} g_{\mathcal{C}}(\gamma, m, u) \quad (5.21)$$

для всех (m, u) из $M \times U$. Тогда имеет место следующее равенство:

$$\inf_M g(m) = \inf_U \inf_M \sup_{\mathcal{C}} g_{\mathcal{C}}(\gamma, m, u).$$

Доказательство. Если (5.21) имеет место для всех (m, u) из $M \times U$, то

$$\inf_M g(m) \leq \inf_U \inf_M \sup_{\mathcal{C}} g_{\mathcal{C}}(\gamma, m, u).$$

Однако для всех m из M

$$g(m) = \sup_{\mathcal{C}} g_{\mathcal{C}}(\gamma, m, K(m)).$$

Следовательно, вместо неравенства должно иметь место равенство, что и завершает доказательство теоремы.

Указанное ниже следствие из теорем 5.9 и 5.10 дает условие типа наличия седловой точки.

Следствие

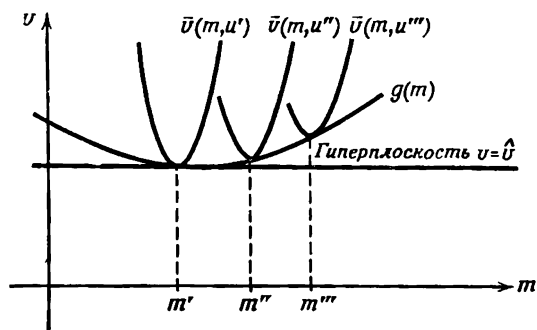
Предположим, что двухуровневая система удовлетворяет предположениям теоремы 5.9. Пусть межуровневые функции системы строго монотонны и условие (5.21) имеет место для всех (m, u) из $M \times U$. Тогда выполнения равенства

$$\max_{\mathcal{C}} \min_{U^V} \min_M g_{\mathcal{C}}(\gamma, m, u) = \min_U \min_M \max_{\mathcal{C}} g_{\mathcal{C}}(\gamma, m, u)$$

достаточно для координируемости.

Условие, выражаемое неравенством в теореме 5.10, иллюстрируется фиг. 5.5. Оно имеет следующую важную интерпретацию. Если координатор ищет максимум на области $\mathcal{C}$ для фиксированной пары (m, u) , то кажущиеся глобальные затраты представляются локальным решающим элементам более высокими, чем минимальные фактические затраты. Если локальные решающие элементы минимизируют кажущиеся глобальные затраты на области $M \times U^V$ для заданного γ , то, как показано на фиг. 5.4, получаемые результаты будут располагаться ниже минимальных фактических затрат. Когда эти противоположные действия будут уравновешены (что и выражено условием седлового типа), система будет скоординирована, если справедливы допущения теоремы 5.9.

В силу предположения $K(M) \subseteq U^V$ полученные выше результаты применимы главным образом к случаю, в котором исполь-



Ф и г. 5.5. Относительное расположение поверхностей $\bar{v}(m, u) = \sup \{g_{\mathcal{C}}(\gamma, m, u) : \gamma \in \mathcal{C}\}$ и гиперплоскости $v = \hat{u}$ для фиксированных значений u . Значения m' , m'' и m''' таковы, что $u' = K(m')$, $u'' = K(m'')$ и $u''' = K(m''')$.

зуется «развязывание» взаимодействий. Другое предположение, согласно которому U^v содержит $\hat{u} = K(\hat{m})$ для некоторого глобально оптимального управляющего воздействия $\hat{m}$, допускает использование прогнозирования взаимодействий при условии, что прогноз дает значения $\alpha^v = \hat{u}$.

Чтобы рассмотреть случай, когда используется способ прогнозирования взаимодействий, возьмем множества $\mathcal{A}$ и $\mathcal{B}$, где $\mathcal{A}$ — такое подмножество U , что $K(M) \subseteq \mathcal{A}$, а каждое β из $\mathcal{B}$ дает семейство локальных целевых функций $g_{i\beta}$, $1 \leq i \leq n$. Предположим, что $\mathcal{C} = \mathcal{A} \times \mathcal{B}$, так что каждый координирующий сигнал γ представляет собой пару (α^v, β^v) , где $\{\alpha^v\} = U^v$, а локальными целевыми функциями являются функции $g_{iv} = g_{i\beta}$, при $\beta = \beta^v$.

Если рассматриваемая система имеет кажущуюся глобальную целевую функцию $g_{\mathcal{C}}$, то функция $g_{\mathcal{B}}$ будет определена на $\mathcal{B} \times M \times U$ так, что

$$g_{\mathcal{C}}(\gamma, m, u) = g_{\mathcal{B}}(\beta^v, m, u).$$

Функция $g_{\mathcal{B}}$ есть кажущаяся глобальная целевая функция для системы, в которой с помощью координирующего сигнала β осуществляется изменение (координация) целей.

В следующей теореме подытоживаются полученные выше результаты применительно к двухуровневой системе при использовании метода прогнозирования взаимодействий.

Теорема 5.11

Предположим, что двухуровневая система обладает свойством монотонности и, следовательно, имеет кажущуюся глобальную

целевую функцию $g_{\mathcal{B}}$, и $\mathcal{C} = \mathcal{A} \times \mathcal{B}$ Тогда

$$\sup_{\mathcal{B}} \inf_{\mathcal{A}} \min_M g_{\mathcal{B}}(\beta, m, \alpha) \leq \inf_M g(m).$$

Кроме того, если межуровневые функции системы строго монотонны, то равенство

$$\max_{\mathcal{B}} \min_{\mathcal{A}} \min_M g_{\mathcal{B}}(\beta, m, \alpha) = \min_M g(m)$$

является достаточным условием для координируемости; в случае же, когда условие (5.24) выполняется для всех (m, u) из $M \times U$, для координируемости достаточно выполнения условия седлового типа

$$\max_{\mathcal{B}} \min_{\mathcal{A}} \min_M g_{\mathcal{B}}(\beta, m, \alpha) = \min_M \min_{\mathcal{A}} \max_{\mathcal{B}} g_{\mathcal{B}}(\beta, m, \alpha).$$

5. ФОРМИРОВАНИЕ ЛОКАЛЬНЫХ ФУНКЦИЙ КАЧЕСТВА И ИХ МОДИФИКАЦИЯ

Обратимся теперь к методам модифицирования локальных оптимизационных задач и к методам получения их из глобальной оптимизационной задачи, имея в виду обеспечение координации. Успех в координировании двухуровневой системы зависит от локальных оптимизационных задач и их взаимосвязей. Стратегия координирования, основанная на выбранном принципе координации, может скоординировать систему только в том случае, если связи между локальными элементами удовлетворяют определенным условиям. По этой причине может возникнуть необходимость в модификации локальных оптимизационных задач.

В настоящей главе мы ограничим наше рассмотрение координацией целей и «развязыванием» или прогнозированием взаимодействий. Поэтому свобода модификации локальных оптимизационных задач состоит главным образом, хотя и не исключительно, в изменении локальных функций качества. В соответствии с этим мы рассмотрим следующие вопросы: 1) как локальные функции качества могут быть образованы или модифицированы простым и естественным путем исходя из глобальной функции качества; 2) как могут быть модифицированы уже заданные локальные функции качества и 3) как вводимые нами специальные операторы оценки эффектов внутреннего взаимодействия могут быть использованы в качестве инструмента для целенаправленной модификации локальных функций качества.

Получение локальных функций качества

Если локальные функции качества не задаются заранее, можно получить их из глобальной функции качества. Это тривиальная

задача, если нам известно глобально оптимальное управляющее воздействие. Например, если $\hat{m}_i$ — i -я компонента глобально оптимального управляющего воздействия $\hat{m}$, то всегда можно определить i -е локальные функции качества как такие действительные функции $f_i(m_i, \hat{m}_i) \geq 0$, что $f_i(m_i, \hat{m}_i) = 0$, если $m_i = \hat{m}_i$. В остальных случаях очевидным способом получения локальных функций качества является выделение тех «частей» глобальной функции качества, которые зависят только от локальных переменных, относящихся к компетенции соответствующих локальных решающих элементов. Для выполнения этой задачи имеются два пути.

Локальные ограничения

Непосредственное получение локальных функций качества из заданной глобальной функции качества состоит в фиксации части входящих в нее переменных.

Пусть G — заданная глобальная функция качества. Тогда i -я локальная функция качества, получаемая из G с помощью заданной пары вход — выход $(\tilde{m}, \tilde{y})$ из области $M \times Y$, есть функция G_i , задаваемая на $M_i \times Y_i$ уравнением

$$G_i(m_i, y_i) = G(\tilde{m}_1, \dots, \tilde{m}_{i-1}, m_i, \tilde{m}_{i+1}, \dots, \tilde{m}_n; \tilde{y}_1, \dots, \tilde{y}_{i-1}, y_i, \tilde{y}_{i+1}, \dots, \tilde{y}_n).$$

Очевидно, что $G_i(m_i, y_i) = G(m, y)$ для всех (m, y) из $M \times Y$, для которых $m_j = \tilde{m}_j$ и $y_j = \tilde{y}_j$ при $j \neq i$. Этим способом можно образовать семейство локальных функций качества G_i , $1 \leq i \leq n$, из заданной глобальной функции качества. Локальные функции качества, полученные таким путем, часто можно упростить: обычно в глобальной функции качества существуют величины (составляющие), которые не оказывают влияния на выбор i -го локального управляющего воздействия и поэтому могут быть исключены из i -й локальной функции качества.

Пример 5.5

Пусть $M = Y = R^2$, и рассмотрим функцию G , заданную на $M \times Y$ в виде

$$G(m, y) = m_1 y_1 + (m_2 + y_2^2) \log m_1.$$

Для данной пары $(\tilde{m}, \tilde{y})$ из $M \times Y$ локальные функции качества G_1 и G_2 , полученные из G путем простых преобразований, таковы:

$$G_1(m_1, y_1) = m_1 y_1 + (\tilde{m}_2 + y_1^2) \log m_1, \\ G_2(m_2, y_2) = m_2 \log \tilde{m}_1$$

Декомпозиция

Другой путь получения локальных функций качества состоит в декомпозиции глобальной функции качества. При этом принимается во внимание соотношение между входами и выходами всего процесса.

Для любой заданной глобальной функции качества G существуют семейство функций G_i , $1 \leq i \leq n$, где $G_i: M_i \times Y_i \times W_i \rightarrow V$, семейство функций θ_i , $1 \leq i \leq n$, где $\theta_i: M \times Y \rightarrow W_i$, и отображение $\psi: V^n \rightarrow V$ такое, что

$$G(m, y) = \psi(G_1(m_1, y_1, \theta(m, y)), \dots, G_n(m_n, y_n, \theta(m, y)))$$

всякий раз, когда $y = P(m)$. Мы будем называть функции G_i , $1 \leq i \leq n$, локальными функциями качества, получаемыми из G декомпозицией, а функции θ_i , $1 \leq i \leq n$, — их функциями взаимодействия. Декомпозиция заданной глобальной функции качества дает не только локальные функции качества, но и определенное отображение, которое в некоторых случаях может служить межуровневой функцией.

Вообще говоря, локальные функции качества, получаемые декомпозицией, включают переменную, отражающую нелокальные эффекты. Если такая переменная может быть исключена или представлена в виде связующих входов подпроцесса, мы будем говорить, что глобальная функция качества «сепарабельна». Точнее, глобальная функция качества G сепарабельна, если существует такое семейство локальных функций качества G_i , $1 \leq i \leq n$, получаемых из G декомпозицией, что связующие подпроцессы функции H_i , $1 \leq i \leq n$, являются их функциями взаимодействия; иначе говоря, $G_i: M_i \times Y_i \times U_i \rightarrow V$, $1 \leq i \leq n$, и

$$G(m, y) = \psi(G_1(m_1, y_1, H_1(m, y)), \dots, G_n(m_n, y_n, H_n(m, y)))$$

при $y = P(m)$. Если глобальная функция качества сепарабельна, она может быть разложена таким образом, что получаемые локальные функции качества зависят только от локальных переменных, и, кроме того, отображение $\psi: V^n \rightarrow V$, получаемое в процессе декомпозиции, есть межуровневая функция для системы.

Пример 5.6

Пусть $M = Y = R^2$; рассмотрим квадратичную форму G

$$G(m, y) = m^T A m + y^T B y$$

на $M \times Y$, где A и B — заданные матрицы размерности 2×2 . Пусть общий процесс P задается уравнениями

$$y_1 = -3m_1 - m_2, \quad y_2 = -4m_1 - m_2.$$

Предположим, что подпроцессы P_1 и P_2 связаны между собой равенствами

$$u_1 = y_2 \equiv H_1(m, y), \quad u_2 = m_1 + y_1 \equiv H_2(m, y).$$

В силу простоты этой системы мы можем выразить нелокальные переменные i -го подпроцесса в виде функций от его локальных переменных. Поэтому G сепарабельна: пусть

$$\begin{aligned} G_1(m_1, y_1, u_1) &= a_{11}m_1^2 + b_{11}y_1^2 - a_{12}m_1(3m_1 + y_1) + b_{12}y_1u_1, \\ G_2(m_2, y_2, u_2) &= a_{22}m_2^2 + b_{22}y_2^2 - \frac{1}{4}a_{21}m_2(m_2 + y_2) + \\ &+ \frac{1}{4}b_{21}y_2(4u_2 + m_2 + y_2). \end{aligned}$$

Тогда

$$G(m, y) = G_1(m_1, y_1, H_1(m, y)) + G_2(m_2, y_2, H_2(m, y))$$

при $y = P(m)$. Если B — диагональная матрица, то G_1 и G_2 не зависят соответственно от u_1 и u_2 ; если A и B обе диагональны, то декомпозиция очевидна.

Пример 5.7

Пусть $M = Y = R^3$; рассмотрим функцию G , заданную на $M \times Y$ в виде

$$G(m, y) = (m_1^2 + y_1^2) |m_2 y_2| + \sin^2 y_3.$$

G — сепарабельна: действительно, пусть $\psi: R^3 \rightarrow R$, так что

$$\psi(v_1, v_2, v_3) = v_1 v_2 + v_3.$$

Тогда

$$\begin{aligned} G_1(m_1, y_1) &= m_1^2 + y_1^2, \quad G_2(m_2, y_2) = |m_2 y_2|, \\ G_3(m_3, y_3) &= \sin^2 y_3. \end{aligned}$$

Модификация локальных функций качества

Модификация локальных функций качества может понадобиться по разным причинам. В некоторых случаях каждому локальному решающему элементу задается заранее только одна функция качества, причем координация этих функций не предусматривается: весь первый уровень задан и проблема координации сводится в основном к учету взаимодействий. В других случаях для каждого локального решающего элемента имеется семейство функций качества, таких, что применим определенный принцип координации, но система не обязательно будет координируема с помощью этого принципа. К проблеме — как сделать систему координируемой с помощью рассматриваемого принципа, можно подойти, пытаясь модифицировать локальные функции качества таким образом, чтобы при всех модификациях сохраня-

лась применимость принципа и по меньшей мере для одной из них система была координируема. В предположении, что существует глобально оптимальное управляющее воздействие, мы изложим несколько общих методов или подходов к модификации локальных функций качества.

Модификации с помощью пар «вход — выход»

Если локальные функции качества получаются путем наложения ограничений (условий) на пары «вход — выход», то представляется естественным использовать эти пары вход — выход как средства для модификации и, следовательно, координации. Пусть G — глобальная функция качества, и пусть семейство локальных функций качества G_i , $1 \leq i \leq n$, получается из G с помощью пар вход — выход. Пусть $\mathcal{R} = M \times Y$; тогда для каждой пары β из $\mathcal{R}$ модифицированная i -я локальная функция качества $G_{i\beta}$ образуется из G с помощью пары β .

Аналогичные средства модификации имеются, когда локальные функции качества получаются декомпозицией, а функции θ_i , $1 \leq i \leq n$, являются соответствующими функциями взаимодействия. Пусть семейство локальных функций качества G_i , $1 \leq i \leq n$, получается из G декомпозицией. Пусть снова $\mathcal{R} = M \times Y$. Тогда для каждой пары β из $\mathcal{R}$ модифицированная i -я локальная функция качества $G_{i\beta}$ задается на $M_i \times Y$ уравнением

$$G_{i\beta}(m_i, y_i) = G_i(m_i, y_i, \theta_i(\beta)).$$

Образование $\psi: V^n \rightarrow V$, получаемое в процессе декомпозиции, в общем случае не является истинной межуровневой функцией для двухуровневой системы.

Согласованные модификации и модификации с нулевой суммой

Очевидно, что если локальные функции качества заданы заранее, причем отсутствуют какие-либо очевидные способы их модификации, то не ясно, как необходимые модификации могли бы быть осуществлены; такая ситуация возникает, когда локальные функции качества получаются путем декомпозиции из сепарабельной глобальной функции качества. Однако существует ли общее правило, которого следовало бы придерживаться, производя такие модификации?

Одно такое общее правило относится к стратегии координации. Предположим, что локальные функции качества таковы, что рассматриваемый принцип координации применим, но система не координируема; так часто случается, ибо соответствующее

условие координируемости редко удовлетворяется без модификации, даже если выбранный принцип применим. Модификация локальных функций качества тогда должна быть такой, чтобы сохранялась применимость принципа координации. Это приводит нас к понятию «согласованных» модификаций.

Пусть $\mathcal{B}$ — множество координирующих цели сигналов в двухуровневой системе; это значит, что для каждого β из $\mathcal{B}$ конкретизируется семейство локальных функций качества $G_{i\beta}$, $1 \leq i \leq n$. Мы будем называть локальные функции качества, определяемые элементами β из $\mathcal{B}$, *модификациями локальных функций качества, определяемыми множеством $\mathcal{B}$* , или просто *модификациями*. Пусть далее $\mathcal{G} = \{G_i: 1 \leq i \leq n\}$ — заданное семейство локальных функций качества, которые могут и не входить в состав семейства, определяемого элементами $\mathcal{B}$.

Для удобства изложения обозначим через g_i и $g_{i\beta}$, где $1 \leq i \leq n$ и $\beta \in \mathcal{B}$, локальные целевые функции, связанные с локальными функциями качества G_i и $G_{i\beta}$.

Модификации локальных функций качества, соответствующие множеству $\mathcal{B}$, являются *ψ -согласованными по отношению к семейству $\mathcal{G}$* , где ψ — заданное отображение $\psi: V^n \rightarrow V$, если уравнение

$$\psi(g_1(m_1, u_1), \dots, g_n(m_n, u_n)) = \psi(g_{1\beta}(m_1, u_1), \dots, g_{n\beta}(m_n, u_n))$$

удовлетворяется для всех (β, m, u) из $\mathcal{B} \times M \times U$ при $u = K(m)$. Если модификации локальных функций качества, определяемые семейством $\mathcal{B}$, являются ψ -согласованными по отношению к этому семейству, то они являются ψ -согласованными и по отношению друг к другу. В самом деле, мы можем использовать локальные функции качества, определяемые каким-либо координирующим цели сигналом, как отправную точку для выяснения того, являются ли модификации локальных функций качества ψ -согласованными или нет.

Если модификации локальных функций качества в двухуровневой системе являются ψ -согласованными, а ψ — межуровневая функция для некоторого, координирующего цели сигнала, то ψ является межуровневой функцией системы для любого координирующего цели сигнала; в этом случае мы будем называть модификации локальных функций качества *согласованными*.

Для данного отображения $\psi: V^n \rightarrow V$ пусть f будет функцией, определяемой на $\mathcal{B} \times M \times U$ равенством

$$f(\beta, m, u) = \psi(g_{1\beta}(m_1, u_1), \dots, g_{n\beta}(m_n, u_n)). \quad (5.22)$$

Тогда модификации локальных функций качества являются ψ -согласованными, если равенство

$$f(\beta, m, K(m)) = f(\beta', m, K(m))$$

имеет место для всех β и β' из $\mathcal{B}$ и t из M . Следовательно, ψ -согласованные модификации локальных функций качества имеют то свойство, что эффект модификации исчезает (в смысле изменения ψ), если связующие входы согласованы. Кроме того, если модификации локальных функций качества являются ψ -согласованными в двухуровневой системе, а ψ — межуровневая функция при некотором координирующем цели сигнале, то функция f , определенная указанным выше способом для системы, представляет собой кажущуюся глобальную целевую функцию системы.

Если задана групповая операция на оцениваемом множестве, мы можем указать некоторые общие правила для получения согласованной модификации локальных функций качества.

Пусть $(V, \cdot)$ — абстрактная группа. Тогда из определения группы непосредственно следует, что для каждого координирующего цели сигнала β из $\mathcal{B}$ и каждого i , $1 \leq i \leq n$, существует такая функция $(\mu_{i\beta})$, определенная на $M_i \times U_i$, что

$$g_{i\beta}(m_i, u_i) = \mu_{i\beta}(m_i, u_i) \cdot g_i(m_i, u_i)$$

для всех (m_i, u_i) из $M_i \times U_i$.

Пусть локальные целевые функции $g_{i\beta}$ выражаются в виде (5.22). Если данное отображение $\psi: V^n \rightarrow V$ удовлетворяет некоторым условиям по отношению к групповой операции в V , то ψ -согласованная модификация локальной функции качества может быть выражена через одни лишь модифицирующие функции.

Лемма

Если заданное отображение $\psi: V^n \rightarrow V$ является гомоморфным, то модификации локальных функций качества будут ψ -согласованными тогда и только тогда, когда для всех $\beta \in \mathcal{B}$ и t из M

$$\mu(\beta, t, K(t)) = e,$$

где μ определяется на $\mathcal{B} \times M \times U$ равенством

$$\mu(\beta, t, u) = \psi(\mu_{1\beta}(m_1, u_1), \dots, \mu_{n\beta}(m_n, u_n)), \quad (5.23)$$

а e — единичный элемент множества V .

Доказательство. Пусть f определена на $\mathcal{B} \times M \times U$ равенством (5.22), Θ — заданный элемент из $\mathcal{B}$, а (β, t, u) из $\mathcal{B} \times M \times U$ произвольны. Так как ψ есть гомоморфизм, то

$$f(\beta, t, u) = \mu(\beta, t, u) \cdot f(\Theta, t, u).$$

Следовательно, модификации локальной функции качества являются ψ -согласованными, если

$$\mu(\beta, t, K(t)) = e.$$

Этим доказательство завершается.

Если для некоторого координирующего сигнала двухуровневая система обладает межуровневой функцией, желательно, чтобы

модификации локальных функций качества были согласованы по отношению к межуровневой функции. Пусть нам известно, что модификации локальной функции качества являются для системы φ -согласованными, в то время как другое отображение, ψ , является межуровневой функцией для некоторого координирующего сигнала. Следующая лемма выражает через ψ и φ достаточное условие того, что модификации локальных функций качества являются ψ -согласованными, когда они выбраны так, чтобы они были φ -согласованными.

Лемма

Пусть ψ и φ — такие гомоморфные отображения из V^n в V , что ядро ψ содержит ядро φ . Тогда φ -согласованная модификация локальной функции качества является также ψ -согласованной.

Доказательство. Пусть μ_ψ и μ_φ определены на $\mathcal{R} \times M \times U$ соответственно с помощью ψ и φ посредством равенства (5.23). Тогда $\mu_\psi(\beta, m, K(m)) = e$ всякий раз, когда $\mu_\varphi(\beta, m, K(m)) = e$, так как ядро ψ содержит ядро отображения φ , что и доказывает лемму.

Заметим, что для данных отображений ψ и φ , удовлетворяющих условиям леммы, существует такой гомоморфизм $\theta: \mathcal{R}(\varphi) \rightarrow \mathcal{R}(\psi)$, что $\psi(v) = \theta(\varphi(v))$ для всех v из V^n , и, кроме того, θ есть изоморфизм, если ψ и φ имеют одно и то же ядро.

Значительный интерес представляет случай, когда групповая операция на множестве оценок V коммутативна и, следовательно, V — аддитивная группа. Если $(V, +)$ — аддитивная группа, то глобальная целевая функция является *аддитивной*, если

$$g(m) = g_1(m_1, K_1(m)) + \dots + g_n(m_n, K_n(m))$$

на M . Следовательно, глобальная функция качества для данной двухуровневой системы является аддитивной, если истинная глобальная стоимость есть «сумма» истинных локальных стоимостей.

Модификации локальных функций качества в случае, когда $(V, +)$ — аддитивная группа, всегда могут быть выражены в виде

$$g_{i\beta}(m_i, u_i) = g_i(m_i, u_i) + \mu_{i\beta}(m_i, u_i),$$

где $\mu_{i\beta}$ — модифицирующая функция. Мы говорим, что модификации локальных функций качества являются модификациями с нулевой суммой, если

$$\sum_{i=1}^n \mu_{i\beta}(m_i, K_i(m)) = 0$$

для всех (β, m) из $\mathcal{R} \times M$. Модификация локальных функций качества, представляющая собой модификацию с нулевой суммой, является согласованной, если глобальная функция качества

аддитивна. В более общем случае такая модификация будет согласованной, если существует межуровневая функция ψ вида

$$\psi(v_1, \dots, v_n) = \theta(v_1 + \dots + v_n),$$

где $\theta: V \rightarrow V$ — гомоморфизм.

Пример 5.8

Пусть $M = Y = R^2$. Обратимся к примеру 5.6 и рассмотрим процесс в целом P и связующие функции H_1 и H_2 подпроцессов P_1 и P_2 . В данном случае подпроцессы таковы:

$$P_1(m_1, u_1) = m_1 + u_1, \quad P_2(m_2, u_2) = m_2 + 2u_2.$$

Пусть $\mathcal{B} = R^2$, и для каждой пары значений $\beta = (\beta_1, \beta_2)$ из $\mathcal{B}$ функции $\mu_{i\beta}$ для $i = 1, 2$ определены на множестве R^2 в виде

$$\mu_{1\beta}(m_1, u_1) = \beta_1 u_1 - \beta_2 (2m_1 + u_1),$$

$$\mu_{2\beta}(m_2, u_2) = \beta_2 u_2 - \beta_1 (m_2 + 2u_2).$$

Тогда эти функции дают нам модификацию с нулевой суммой.

Чтобы убедиться в этом, образуем функции взаимодействия подпроцессов K_1 и K_2 :

$$K_1(m) = H_1(m, P(m)) = -4m_1 - m_2,$$

$$K_2(m) = H_2(m, P(m)) = -2m_1 - m_2.$$

Однако

$$\begin{aligned} \mu_{1\beta}(m_1, u_1) + \mu_{2\beta}(m_2, u_2) &= \beta_1 (u_1 - m_2 - 2u_2) + \\ &+ \beta_2 (u_2 - 2m_1 - u_1). \end{aligned}$$

Следовательно, если $u_1 = K_1(m) = -4m_1 - m_2$ и $u_2 = K_2(m) = -2m_1 - m_2$, то

$$\mu_{1\beta}(m_1, u_1) + \mu_{2\beta}(m_2, u_2) = 0.$$

Если локальные целевые функции g_i для $i = 1, 2$ таковы, что

$$g_{i\beta}(m_i, u_i) = g_i(m_i, u_i) + \mu_{i\beta}(m_i, u_i),$$

где g_i — заданные функции, то локальные модификации являются ψ -согласованными при $\psi(v_1, v_2) = v_1 + v_2$.

Пример 5.9

Рассмотрим ту же систему, что и в примере 5.8, но пусть теперь $M = R^+ \times R^+$, где R^+ — множество положительных действительных чисел. Пусть $\mathcal{B} = R^+$, и для каждого β из $\mathcal{B}$ функции $\mu_{i\beta}$ ($i = 1, 2$) определены на R^2 в следующем виде:

$$\mu_{1\beta}(m_1, u_1) = \beta u_1 / (2m_1 + u_1),$$

$$\mu_{2\beta}(m_2, u_2) = (m_2 + 2u_2) / \beta u_2.$$

Пусть локальные целевые функции $g_{i\beta}$ ($i = 1, 2$) имеют вид

$$g_{i\beta}(m_i, u_i) = g_i(m_i, u_i) \cdot \mu_{i\beta}(m_i, u_i),$$

где функции g_i заданы. Мы утверждаем, что модификации локальных функций качества являются ψ -согласованными при $\psi(v_1, v_2) = v_1 v_2$.

Операторы оценки эффекта внутренних взаимодействий

Выше мы показали, как могут быть модифицированы первоначально заданные локальные функции качества и при каких условиях модификации будут согласованными. Если оцениваемое множество представляет собой группу, то модификации можно выполнить посредством групповой операции. В том случае, когда оцениваемое множество — аддитивная группа, заданные локальные функции качества могут быть модифицированы путем добавления модифицирующего члена.

Если для оцениваемого множества имеется подходящая мера для выражения различия между составляющими это множество элементами, мы можем ввести удобные для нас операторы и использовать их для того, чтобы найти подходящие модифицирующие функции. Для установления меры различия можно использовать либо метрику, либо определенную групповую операцию в случае, когда оцениваемое множество представляет собой группу (если это множество есть аддитивная группа, то меру различия в этом случае можно определить с помощью операции вычитания). Вводимые нами операторы оценки эффекта внутренних взаимодействий предназначены для нахождения вариаций функций качества и «выходов» процесса в зависимости от изменений на «входах».

Мы укажем здесь некоторые из так называемых «операторов оценки эффекта управляющих воздействий», линейные формы которых будут позднее использованы в гл. 6. При этом мы будем пользоваться следующими обозначениями: пусть $\delta: V \times V \rightarrow V$ означает подходящую меру различия объектов оценки (элементов оцениваемого множества). Для выбранных элементов m^0 из M и m_i из M_i обозначим

$$m^0[m_i] \equiv (m_1^0, \dots, m_{i-1}^0, m_i, m_{i+1}^0, \dots, m_n^0).$$

Иначе говоря, $m^0[m_i]$ означает замену m_i^0 на m_i в m^0 . Аналогично определим и элемент $u^0[u_i]$ из U . Воспользуемся также отображением $\bar{P}: M \times U \rightarrow Y$,

$$\bar{P}(m, u) = (P_1(m_1, u_1), \dots, P_n(m_n, u_n))$$

для обозначения «развязанных» подпроцессов. Согласно определению функции взаимодействия K ,

$$P(m) = \bar{P}(m, K(m)),$$

где $\bar{P}$ в явной форме показывает воздействие связующих входов на процесс.

Полные операторы оценки косвенного эффекта управляющих воздействий

Благодаря взаимодействиям между подпроцессами двухуровневой системы изменение в любом локальном управляющем воздействии будет в общем случае передаваться по всей системе через связующие входы подпроцессов. Следовательно, изменение локального управляющего воздействия может повлечь за собой изменение глобальных затрат через *нелокальные* переменные¹⁾. Мера этого изменения глобальных затрат может быть получена с помощью полных операторов «оценки косвенного эффекта управляющих воздействий», которые мы определим следующим образом:

i-й полный оператор оценки косвенного эффекта управляющих воздействий в заданной точке m^0 из M , обозначаемый через $\Gamma_i(m^0)$, есть такое отображение $\Gamma_i(m^0): M_i \rightarrow V$, значение которого в любой точке m_i из M_i , обозначаемое через $\Gamma_i(m^0)m_i$, дается равенством

$$\Gamma_i(m^0)m_i = \delta [G(m^0, \bar{P}(m^0, K(m^0[m_i])), G(m^0, \bar{P}(m^0, K(m^0)))]. \quad (5.24)$$

Величина $G(m^0, \bar{P}(m^0, K(m^0)))$ представляет собой (фактические) глобальные затраты для данного управляющего воздействия m^0 , так как $\bar{P}(m^0, K(m^0)) = P(m^0)$. Если теперь m_i^0 заменить на m_i , то значения связующих входов изменятся, так что вместо $u^0 = K(m^0)$ мы будем иметь $u = K(m^0[m_i])$. Величина $G(m^0, \bar{P}(m^0, K(m^0[m_i])))$ в этом случае представляет собой глобальные затраты, получающиеся при замене m_i^0 на m_i и связанные только с соответствующими изменениями связующих входов. Вообще говоря, это не фактически реализующиеся глобальные затраты, так как равенство $K(m^0) = K(m^0[m_i])$ может и не выполняться.

Частные операторы оценки косвенного эффекта управляющих воздействий

Изменение глобальных затрат в результате изменения локального управляющего воздействия может быть оценено также и по

¹⁾ Т. е. через переменные, функционально связанные с рассматриваемыми *локальными* переменными, но относящиеся к другому подпроцессу и входящие в состав целевой функции другого локального управляющего элемента.— *Прим. ред.*

тому влиянию, которое оно оказывает на определенные связующие входы, а через них — и на глобальные затраты.

Такая мера будет частной по отношению к указанной выше мере изменения, учитывающей изменения во всех связующих входах. Частная мера изменения глобальных затрат может быть найдена с помощью частных операторов «оценки косвенного эффекта управляющих воздействий», которые мы определим следующим образом:

i-й частный оператор оценки косвенного эффекта управляющих воздействий в данной точке m^0 из M есть такое отображение $\Gamma_{ij}(m^0): M_i \rightarrow V$, что

$$\Gamma_{ij}(m^0) m_i = \delta [G(m^0, \bar{P}(m^0, u^0 [K_j(m^0 [m_i])])), \\ G(m^0, \bar{P}(m^0, u^0))] \quad (5.25)$$

для всех m_i из M_i , где $u^0 = K(m^0)$.

Величина $G(m^0, \bar{P}(m^0, u^0))$, как и ранее, выражает (фактические) общие затраты для данного управляющего воздействия m^0 . Изменение в *i*-м локальном управляющем воздействии со значения m_i^0 на m_i вызывает изменение связующего входа для *j*-го подпроцесса со значения $u_j^0 = K_j(m^0)$ на $u_j = K_j(m^0 [m_i])$. Изменяя только связующий вход *j*-го подпроцесса со значения u_j^0 на u_j , мы получим общие затраты

$$G(m^0, \bar{P}(m^0, u^0 [K_j(m^0 [m_i])])).$$

Частные операторы оценки косвенного эффекта управляющих воздействий определяются для всех *i* и *j*, $1 \leq i \leq n$, $1 \leq j \leq n$. Частный оператор оценки косвенного эффекта управляющих воздействий $\Gamma_{ii}(m^0)$ в точке m^0 дает меру изменения глобальной функции качества, вызванного изменением в *i*-м локальном управляющем воздействии в связи с тем, что эффект этого изменения передается по всей системе и возвращается опять к *i*-му подпроцессу.

Операторы оценки эффективности связующих входов

Вместо того чтобы измерять изменения глобальной функции качества, возникающие вследствие изменений в локальном управляющем воздействии по мере того, как их эффекты передаются через связующие входы, мы можем измерить изменения глобальной функции качества, возникающие благодаря изменениям в самих связующих входах. Такая мера задается операторами «оценки эффективности связующих входов», которые мы определим следующим образом:

i-й оператор оценки эффективности связующих входов в данной точке m^0 из M , обозначаемый $\Delta_i(m^0)$, есть такое отображение

$\Delta_i(m^0): U_i \rightarrow V$, что

$$\Delta_i(m^0) u_i = \delta [G(m^0, \bar{P}(m^0, u^0[u_i])), G(m^0, \bar{P}(m^0, u^0))] \quad (5.26)$$

для всех u_i из U_i , где $u^0 = K(m^0)$.

Как видно из сравнения (5.25) и (5.26), эти операторы вполне аналогичны частным операторам оценки косвенного эффекта управляющих воздействий. Ниже мы еще раз вернемся к этому вопросу.

Имеются различные способы введения операторов оценки тех или иных эффектов внутренних взаимодействий. Например, если для рассматриваемой системы существует кажущаяся глобальная целевая функция, то такие операторы могут быть введены через эту функцию, а не через глобальную функцию качества. Если глобальная функция качества аддитивна, то операторы оценки тех или иных эффектов могут быть определены непосредственно через локальные функции качества, а не через глобальную функцию. Операторы оценки влияния внутренних взаимодействий непосредственно на выходные характеристики («выходы») процесса в целом могут быть определены точно так же, если задать на множестве «выходов» процесса соответствующую меру различия.

Соотношения между введенными операторами оценки

Сравнивая равенства (5.25) и (5.26), мы видим, что частные операторы оценки косвенного эффекта управляющих воздействий могут быть выражены через операторы оценки эффективности связующих входов. А именно, для любого фиксированного m^0 из M и любых i и j , $1 \leq i \leq n$, $1 \leq j \leq n$,

$$\Gamma_{ij}(m^0) m_i = \Delta_j(m^0) K_j(m^0[m_i])$$

при всех m_i из M_i .

Укажем теперь на связь между полными и частными операторами оценки косвенных эффектов управляющих воздействий в случае, когда глобальная функция качества G аддитивна в том смысле, что

$$G(m, y) = \sum_{i=1}^n G_i(m_i, y_i)$$

для всех (m, y) из $M \times Y$. Множество V должно быть, следовательно, аддитивной группой. Пусть δ — операция вычитания в группе V . В этом случае для любого m^0 из M и $1 \leq i \leq n$,

$$\begin{aligned} \Gamma_i(m^0) m_i = & \sum_{j=1}^n \{G_j(m_j^0, P_j(m_j^0, K_j(m^0[m_i]))) - \\ & - G_j(m_j^0, P_j(m_j^0, K_j(m^0)))\}, \end{aligned}$$

тогда как при каждом j , $1 \leq j \leq n$,

$$\Gamma_{ij}(m^0) m_i = G_j(m_j^0, P_j(m_j^0, K_j(m^0[m_i]))) - G_j(m_j^0, P_j(m_j^0, K_j(m^0))).$$

Следовательно,

$$\Gamma_i(m^0) m_i = \sum_{j=1}^n \Gamma_{ij}(m^0) m_i.$$

Модификации локальных функций качества

Укажем теперь, как можно использовать операторы оценки косвенного эффекта управляющих воздействий для модификации локальных функций качества. Предположим, что глобальная функция качества аддитивна в указанном выше смысле. Пусть $\mathcal{B} = M$. Тогда для каждого $\beta \in \mathcal{B}$ и i , $1 \leq i \leq n$, модифицирующая функция $\mu_{i\beta}$ определяется на $M_i \times U_i$ в виде

$$\mu_{i\beta}(m_i, u_i) = \Gamma_i(\beta) m_i - \Delta_i(\beta) u_i;$$

если же используется принцип прогнозирования взаимодействий, можно просто положить

$$\mu_{i\beta}(m_i, u_i) = \Gamma_i(\beta) m_i.$$

Эти две формы модифицирующих функций будут рассмотрены в гл. 6, где введенные здесь операторы будут линейаризованы.

6. КООРДИНИРУЕМОСТЬ ПРИ ИСПОЛЬЗОВАНИИ ПРИНЦИПА СОГЛАСОВАНИЯ

В этом разделе мы сосредоточим внимание на координируемости двухуровневой системы при использовании принципа согласования взаимодействий и принципа согласования функций качества в соответствии с (5.10) и (5.12). Полученные результаты относятся к случаю, когда отсутствует конфликтность и рассматриваемая двухуровневая система обладает свойством монотонности и безусловной согласованности.

Прежде всего, напомним некоторые обозначения. Так как будет использоваться координация только путем изменения целей (принципы согласования предназначены главным образом для тех случаев, когда мы имеем дело с «развязанными» взаимодействиями), то мы считаем, что множеством координирующих сигналов является множество $\mathcal{B}$. Тогда для каждого $\beta \in \mathcal{B}$ символ x^β обозначает пару (m^β, u^β) из прямого произведения множеств $M \times U$, такую, что для каждого i , $1 \leq i \leq n$, пара (m_i^β, u_i^β) является β -оптимальной, т. е.

$$g_{i\beta}(m_i^\beta, u_i^\beta) = \min_{M_i \times U_i} g_{i\beta}(m_i, u_i)$$

и, следовательно, (m_i^β, u_i^β) — оптимальное решение локальной задачи оптимизации $\mathcal{D}_i(\beta)$. Элемент $\hat{m}$ из M означает глобально оптимальное управляющее воздействие.

Напомним, что под «координируемостью» с использованием тех или иных принципов согласования мы понимаем следующее.

1) Принцип согласования взаимодействий применим тогда и только тогда, когда справедливо следующее предложение:

$$(\forall \beta) (\forall x^\beta) (\exists \hat{m}) \{[(m, u) = x^\beta \text{ и } u = K(m)] \Rightarrow m = \hat{m}\}. \quad (5.27)$$

Принцип не применим тогда и только тогда, когда существуют такое β и такая пара $(m^\beta, u^\beta) = x^\beta$, что $u^\beta = K(m^\beta)$, но управление m^β не является глобально оптимальным. Двухуровневая система координируема на основе принципа согласования взаимодействий тогда и только тогда, когда принцип применим и существуют такой координирующий сигнал β и такое $x^\beta = (m^\beta, u^\beta)$, что $u^\beta = K(m^\beta)$.

2) Принцип согласования функций качества применим, если справедливо следующее предложение:

$$(\forall \beta) (\forall x^\beta) (\exists \hat{m}) \{[(\hat{m}, u) = x^\beta \text{ и } \bar{g}_\beta(m, K(m)) = \bar{g}_\beta(\hat{m}, u)] \Rightarrow m = \hat{m}\}, \quad (5.28)$$

где

$$\bar{g}_\beta(m, u) = (g_{1\beta}(m_1, u_1), \quad g_{n\beta}(m_n, u_n)).$$

Двухуровневая система координируема на основе принципа согласования функций качества, если принцип применим и существуют такое β в $\mathcal{B}$ и такая пара $(m^\beta, u^\beta) = x^\beta$, что

$$g_\beta(m^\beta, K(m^\beta)) = g_\beta(m^\beta, u^\beta).$$

При анализе координируемости на основе принципов согласования мы будем переходить от более слабых к более сильным требованиям. Имеются два общих свойства — безусловная межуровневая согласованность и свойство монотонности, которые связаны с применимостью принципов, и одно общее свойство — безусловная локальная согласованность¹⁾, — относящееся к существованию оптимального координирующего воздействия. Межуровневая согласованность — точечное свойство в отличие от свойства монотонности и является более слабым из указанных двух свойств; поэтому мы и начнем анализ с рассмотрения систем, наделенных этим свойством²⁾.

Системы с межуровневой согласованностью

Мы покажем здесь, что для применимости обоих принципов согласования достаточно существования межуровневой согласованности, хотя сами по себе эти принципы имеют более широкую область применения. Система не обязательно должна обладать межуровневой согласованностью для того, чтобы были приме-

¹⁾ Т. е. согласованность функций $g_{1\beta}, g_{2\beta}, \dots, g_{n\beta}$.

²⁾ В этом разделе под понятием «согласованность» всюду подразумевается «безусловная согласованность».

нимы принципы согласования; однако при достаточно слабых условиях межуровневая согласованность оказывается необходимой. Мы покажем также, что система, для которой применим определенный принцип согласования, должна иметь локальную согласованность, если предполагается, что система координируема при использовании этого принципа. Однако будет показано, что при сравнительно слабых дополнительных условиях, при которых становится необходимой межуровневая согласованность, локальная согласованность является не только необходимым, но также и достаточным условием для координируемости при использовании этого принципа.

С концептуальной точки зрения чрезвычайно важным представляется прежде всего следующий результат:

Теорема 5.12

Принцип согласования взаимодействий применим для рассматриваемой двухуровневой системы тогда и только тогда, когда для нее применим принцип согласования функций качества.

Доказательство. Пусть β — произвольный элемент из X : положим также, что существует такое $x^\beta = (m^\beta, u^\beta)$, что $\bar{g}_i(m^\beta, u^\beta) = \bar{g}_i(m^\beta, K(m^\beta))$. Тогда пары $(m_i^\beta, K_i(m^\beta))$ будут β -оптимальными, и, кроме того, пара $(m^\beta, K(m^\beta))$ удовлетворяет условию координируемости, отвечающему принципу согласования взаимодействий. Следовательно, существует такое $\hat{m}$, что $m^\beta = \hat{m}$, если применим принцип согласования взаимодействий. С другой стороны, очевидно, что (m^β, u^β) удовлетворяет условию координируемости для принципа согласования функций качества; следовательно, существует такое $\hat{m}$, что $m^\beta = \hat{m}$, если принцип согласования функций качества применим. Это доказывает теорему.

Теорема 5.13

Двухуровневая система координируема на основе принципа согласования взаимодействий тогда и только тогда, когда она координируема на основе принципа согласования функций качества.

Доказательство этой теоремы содержится в доказательстве теоремы 5.12.

Приведенные выше теоремы утверждают, что оба принципа согласования математически эквивалентны; мы можем поэтому ограничить наш анализ рассмотрением принципа согласования взаимодействий. С содержательной же точки зрения эти принципы отличаются один от другого. Использование принципа согласования взаимодействий требует, чтобы фактические связующие входы были либо измеримыми, либо вычислимыми, в противном случае

принцип согласования взаимодействий не может быть использован. Возникающие при этом трудности можно обойти, если использовать принцип согласования функций качества, так как кажущиеся и фактические локальные затраты чаще всего бывают известны.

Перейдем теперь к анализу координируемости двухуровневой системы при использовании принципа согласования взаимодействий. Благодаря приведенным выше теоремам результаты наших последующих рассмотрений будут справедливы и для принципа согласования функций качества.

Теорема 5.14

Принцип согласования взаимодействий применим для всякой двухуровневой системы, обладающей межуровневой согласованностью.

Доказательство. Воспользуемся теоремой 5.3. Предположим, что данная двухуровневая система обладает межуровневой согласованностью. Пусть β — произвольный элемент из $\mathcal{R}$, и предположим, что существует x^β (m^β , u^β). Если связующие входы согласованы, то $u^\beta \in K(m^\beta)$. Тогда из теоремы 5.3 вытекает, что $m^\beta = \hat{m}$. Следовательно, принцип согласования взаимодействий применим и наша теорема доказана.

Предположим теперь, что данная двухуровневая система не обладает межуровневой согласованностью; тогда принципы согласования будут все-таки применимы при условии, что для каждого координирующего воздействия, не приводящего к межуровневой согласованности, либо не выполняется условие координируемости, либо по меньшей мере для одной локальной задачи не достигается оптимальное решение. Это указывает на то, что при определенных (довольно слабых) условиях межуровневая согласованность не только достаточна, но и необходима.

Теорема 5.15

Предположим, что для каждого координирующего сигнала β в рассматриваемой двухуровневой системе и любого i , $1 \leq i \leq n$,

$$\min_{M_i \times U_i} g_{i\beta}(m_i, u_i) = \min_M g_{i\beta}(m_i, K_i(m))$$

имеет место всякий раз, когда существует его правая часть. Тогда наличие в системе межуровневой согласованности как необходимо, так и достаточно для применимости принципа согласования взаимодействий.

Доказательство. Нам нужно доказать только необходимость. Предположим, система не обладает межуровневой согласованностью. Тогда существуют такие β из $\mathcal{R}$ и m из M , что для

каждого i , $1 \leq i \leq n$, величина $g_{i\beta}(m_i, K_i(m))$ представляет собой минимальное значение $g_{i\beta}(m'_i, K_i(m'))$ на M , а m не является глобально оптимальным. Из нашего предположения следует, что существует такая пара $(m^\beta, u^\beta) = x^\beta$, что $(m, K(m)) = (m^\beta, u^\beta)$. Пара (m^β, u^β) удовлетворяет условию координируемости рассматриваемого принципа согласования, но $m^\beta = m$ не является глобально оптимальным. Это исключает применимость принципа. Необходимость, а с ней и вся теорема тем самым доказаны.

Выполнение предпосылок теоремы 5.15 связано с условием: для каждого i , $1 \leq i \leq n$,

$$M_i \times U_i = \{(m_i, K_i(m)): m \in M\}.$$

Это условие удовлетворяется, если для всех i , $1 \leq i \leq n$, связующие входы u_i , получаемые с помощью функций K_i , не зависят от управляющих воздействий m_i , и $U_i = K_i(M)$.

Но даже если данный принцип согласования применим для рассматриваемой двухуровневой системы, эта система может быть некоординируемой при использовании данного принципа. Координируемость на основе того или иного принципа не будет иметь места, если не существует такого координирующего сигнала, для которого выполнялось бы условие координируемости.

Теорема 5.16

Предположим, что принцип согласования взаимодействий в данной двухуровневой системе применим. Тогда для того, чтобы система была координируемой при использовании этого принципа, она должна обладать локальной согласованностью.

Доказательство. Пусть данная двухуровневая система координируема на основе принципа согласования взаимодействий. Тогда существуют β из $\mathcal{B}$ и такая пара $(m^\beta, u^\beta) = x^\beta$, что $u^\beta = K(m^\beta)$. Из следствия к теореме 5.3 вытекает, что система должна обладать локальной согласованностью; следовательно, теорема доказана.

Доказанная теорема утверждает, что для того чтобы условие координируемости при использовании какого-либо из двух принципов согласования удовлетворялось, система должна прежде всего обладать локальной согласованностью. Однако так же, как межуровневая согласованность не является необходимой для применимости принципов согласования, локальная согласованность сама по себе, вообще говоря, еще не гарантирует выполнения условия координируемости.

Теорема 5.17

Предположим, что принцип согласования взаимодействий применим в данной двухуровневой системе и, кроме того, удовлетво-

ряются предпосылки теоремы 5.15. Тогда наличие в системе локальной согласованности будет как необходимым, так и достаточным условием для координируемости на основе принципа согласования взаимодействий.

Доказательство. Нам нужно доказать лишь достаточность; поэтому предположим, что система обладает локальной согласованностью. Тогда существуют такие β из $\mathcal{B}$ и m из M , что для каждого i , $1 \leq i \leq n$, $g_i(m_i, K_i(m))$ есть минимальное значение $g_{i\beta}(m'_i, K_i(m'))$ на множестве M , и $m = \hat{m}$. Следовательно, в силу предположений теоремы 5.15 существует такая пара $(m^\beta, u^\beta) = x^\beta$, что $(m, K(m)) = (m^\beta, u^\beta)$. Таким образом, пара (m^β, u^β) удовлетворяет условию координируемости для принципа согласования взаимодействий. Тем самым достаточность и сама теорема доказаны.

Следствие 5.18

Предположим, что двухуровневая система удовлетворяет условиям теоремы 5.15. Тогда наличие в системе одновременно межуровневой и локальной согласованности необходимо и достаточно для координируемости на основе принципа согласования взаимодействий.

Теперь для углубления нашего анализа необходимо ввести некоторую конкретизацию рассматриваемых систем. Поэтому мы перейдем к рассмотрению системы с более жесткими структурными ограничениями, а именно к двухуровневым системам, обладающим свойством монотонности.

Системы, обладающие свойством монотонности

Если двухуровневая система обладает свойством монотонности, она имеет по определению монотонную межуровневую функцию для каждого координирующего сигнала системы. Поэтому существует определенная зависимость между локальными и общими (глобальными) затратами во всем диапазоне управляющих воздействий; этого может и не быть, если система просто обладает межуровневой согласованностью. Следовательно, мы для системы, обладающей свойством монотонности, должны получить более сильные результаты.

Прежде всего на основе теорем 5.2 и 5.14 получаем следующий результат.

Следствие 5.19

Принцип согласования взаимодействий применим для любой двухуровневой системы, обладающей свойством монотонности.

Теперь мы можем указать принципиальные соображения, по которым нам понадобилось в предыдущем разделе ввести

понятие согласованных модификаций локальных функций качества.

Теорема 5.20

Предположим, что двухуровневая система при некотором координирующем сигнале обладает монотонной межуровневой функцией ψ . Тогда принцип согласования взаимодействий применим, если модификация локальных функций качества является ψ -согласованной.

Доказательство. Если ψ — монотонная межуровневая функция системы при некотором координирующем сигнале и модификации локальных функций качества являются ψ -согласованными, то ψ является межуровневой функцией для всех координирующих сигналов системы. Следовательно, система обладает свойством монотонности. Тогда доказательство теоремы вытекает из следствия 5.19.

Несмотря на свою простоту, а может быть, как раз вследствие ее, предыдущие два результата, дополняя друг друга, дают в наши руки мощное средство для обеспечения координируемости. Следствие 5.19 утверждает, что монотонная зависимость между глобальной и локальными функциями качества создает хорошую основу для синтеза координируемой системы. Взяв за основу систему, обладающую свойством монотонности, можно (не опасаясь нарушить применимость принципа) испробовать различные согласованные модификации локальных функций качества, пытаясь так модифицировать эти локальные функции, чтобы система в результате была координируемой на основе данного принципа.

Цель модификации локальной функции качества — сделать двухуровневую систему координируемой. Имея в виду полученные выше результаты, проанализируем теперь системы, обладающие свойством монотонности, с точки зрения того, при каких условиях эти системы координируемы с помощью принципов согласования. Мы знаем, что для координируемости необходима, а в некоторых случаях и достаточна локальная согласованность. Однако в силу дополнительного свойства монотонности мы можем получить практически более удобные условия. Пусть для данной двухуровневой системы, обладающей свойством монотонности, $g_{\mathcal{A}}$ — кажущаяся глобальная функция затрат, а g пусть обозначает глобальную целевую функцию системы.

Теорема 5.21

Двухуровневая система, обладающая свойством монотонности, координируема на основе принципа согласования взаимодействий, только если выполняется следующее равенство:

$$\max_{\mathcal{B}} \min_{M \times U} g_{\mathcal{A}}(\beta, m, u) = \min_M g(m). \quad (5.29)$$

Доказательство. Предположим, что двухуровневая система обладает свойством монотонности и координируема при использовании принципа согласования взаимодействий. Тогда существуют β из $\mathcal{B}$ и пара $(m^\beta, u^\beta) = x^\beta$, такая, что $u^\beta = K(m^\beta)$. В силу того что принцип применим, $m^\beta = \hat{m}$. Поэтому в силу свойства монотонности

$$\begin{aligned} \min_{M \times U} g_{\mathcal{A}}(\beta, m, u) &= g_{\mathcal{A}}(\beta, m^\beta, u^\beta) = \\ &= g_{\mathcal{A}}(\beta, m^\beta, K(m^\beta)) = g(m^\beta) = g(\hat{m}) = \min_M g(m). \end{aligned}$$

Для завершения доказательства остается применить теорему 5.8.

Вообще говоря, если система обладает свойством монотонности, равенство (5.29) является необходимым, но не достаточным условием для координируемости на основе какого-либо из принципов согласования. Может случиться и так, что ни одна из локальных задач в данной двухуровневой системе, обладающей свойством монотонности, не имеет оптимального решения; тогда система, конечно, не будет координируемой, хотя условие (5.29) и может выполняться.

Теорема 5.22

Предположим, что межуровневые функции двухуровневой системы строго монотонны. Тогда условие (5.29) необходимо и достаточно для координируемости на основе принципа согласования взаимодействий.

Доказательство. Необходимость вытекает из теоремы 5.21, а достаточность — из теоремы 5.9.

Если межуровневые функции двухуровневой системы являются строго монотонными, то при достаточно слабых условиях для координируемости системы необходимо и достаточно существование седловой точки каккущейся глобальной целевой функции $g_{\mathcal{A}}$.

Теорема 5.23

Предположим, что межуровневые функции данной двухуровневой системы строго монотонны и, кроме того, неравенство

$$g(m) \leq \sup_{\mathcal{A}} g_{\mathcal{A}}(\beta, m, u) \quad (5.30)$$

выполняется для всех (m, u) из $M \times U$. Тогда существование седловой точки $g_{\mathcal{A}}$, т. е. выполнение равенства

$$\max_{\mathcal{B}} \min_{M \times U} g_{\mathcal{A}}(\beta, m, u) = \min_{M \times U} \max_{\mathcal{A}} g_{\mathcal{A}}(\beta, m, u),$$

является одновременно необходимым и достаточным условием для координируемости на основе принципа согласования взаимодействий.

Доказательство. В силу неравенства (5.30) получаем из теоремы 5.21 равенство

$$\inf_M g(m) = \inf_{M \times U} \sup_{\mathcal{R}} g(\beta, m, u).$$

Поэтому $g_{\mathcal{R}}$ имеет седловую точку тогда и только тогда, когда выполняется условие (5.29). Для завершения доказательства тогда достаточно применить теорему 5.22.

Двухуровневая система, обладающая свойством монотонности, координируема при использовании определенного принципа согласования (принципа согласования взаимодействий или принципа согласования функций качества) тогда и только тогда, когда существует такой координирующий сигнал, что конкретизированные локальные задачи оптимизации имеют оптимальные решения, которые удовлетворяют условию координируемости для данного принципа координации. Такие координирующие сигналы мы будем называть *компенсирующими*. Полученные выше результаты дают условия существования компенсирующего сигнала в двухуровневой системе, обладающей свойством монотонности. Рассмотрим подробнее свойства компенсирующих сигналов в таких двухуровневых системах.

На основании теоремы 5.21 заключаем, что в двухуровневой системе, обладающей свойством монотонности, данный координирующий сигнал является компенсирующим (относительно какого-либо принципа согласования) только в том случае, если β удовлетворяет неравенству

$$g(\hat{m}) \leq g_{\mathcal{R}}(\beta, m, u) \quad (5.31)$$

для всех (m, u) из $M \times U$. Кажущаяся глобальная целевая функция $g_{\mathcal{R}}$ для компенсирующего координирующего сигнала должна быть ограничена снизу минимальным значением g . Если межуровневые функции двухуровневой системы строго монотонны, то для того, чтобы данный координирующий сигнал был компенсирующим (относительно какого-либо принципа согласования), необходимо и достаточно, чтобы он удовлетворял условию (5.31) для всех (m, u) из $M \times U$.

Если множество оценок V является группой, мы можем выделить из кажущейся глобальной целевой функции $g_{\mathcal{R}}$ модифицирующий член: для данного координирующего сигнала $\odot$ из $\mathcal{R}$ существует функция μ на $\mathcal{R} \times M \times U$, удовлетворяющая условию

$$g_{\mathcal{R}}(\beta, m, u) = \mu(\beta, m, u) \cdot g_{\mathcal{R}}(\odot, m, u) \quad (5.32)$$

на $\mathcal{R} \times M \times U$. Если теперь групповая операция « $\cdot$ » сохраняет

порядок $\leq$ на V , мы можем охарактеризовать компенсирующий координирующий сигнал следующим образом:

Теорема 5.24

Пусть в данной двухуровневой системе, обладающей свойством монотонности, групповая операция на множестве оценок сохраняет порядок. Тогда любой компенсирующий координирующий сигнал β из $\mathcal{B}$ должен удовлетворять неравенству

$$g(\hat{m}) \cdot [g_{\mathcal{B}}(\odot, m, u)]^{-1} \leq \mu(\beta, m, u) \quad (5.33)$$

для всех (m, u) из $M \times U$, где μ задано на $\mathcal{B} \times M \times U$ равенством (5.32).

Следствие

Предположим, что межуровневые функции двухуровневой системы строго монотонны и групповая операция сохраняет порядок на множестве оценок. Тогда компенсирующие координирующие сигналы из $\mathcal{B}$ являются как раз теми координирующими входами из $\mathcal{B}$, которые удовлетворяют (5.33) для всех (m, u) из $M \times U$.

Доказательства обоих утверждений непосредственно вытекают из неравенства (5.31).

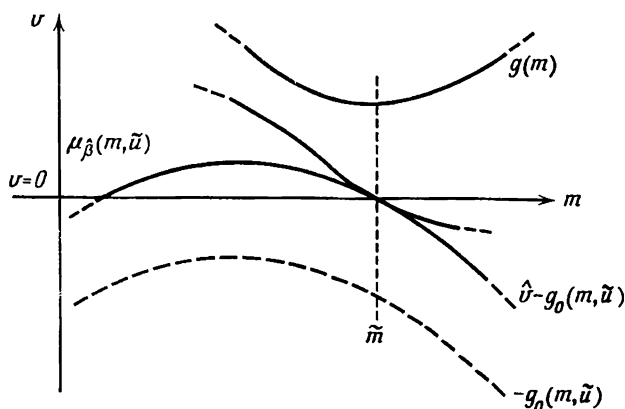
Наконец, рассмотрим специальный случай, когда глобальная функция качества двухуровневой системы является аддитивной и используется модификация локальных функций качества с нулевой суммой. Если двухуровневая система обладает аддитивной глобальной функцией качества, то множество значений этой функции должно быть аддитивной группой; поэтому ее глобальная целевая функция по определению выражается в виде суммы

$$g(m) = \sum_{i=1}^n g_i(m_i, K_i(m))$$

каких-то заданных локальных целевых функций $g_1, \dots, g_n$, и любая из ее локальных целевых функций $g_{i\beta}$ представима в виде суммы g_i и модифицирующей функции $\mu_{i\beta}$, определенной на $M_i \times U_i$. Модифицирующие функции обеспечивают удобное средство для отыскания компенсирующего координирующего сигнала, когда модификации локальной функции качества обладают нулевой суммой.

Теорема 5.25

Допустим, что двухуровневая система обладает аддитивной глобальной функцией качества и используются модификации локальных функций качества с нулевой суммой, и пусть операция сохраняет порядок «добавление аддитивной величины» на множестве оценок качества. Тогда, для того чтобы данный координи-



Фиг. 5.6. Представление модифицирующего члена $\mu_{i\beta}(m, u) = \sum_{i=1}^n \mu_{i\beta}(m_i, u_i)$

при компенсирующем координирующем воздействии $\hat{\beta}$ и $\tilde{\mu} = K(\tilde{m})$, когда все модификации являются модификациями с нулевой суммой и $g_0(m, u) = \sum_{i=1}^n g_i(m_i, u_i)$.

рующий сигнал β был компенсирующим, необходимо и достаточно, чтобы β удовлетворяло неравенству

$$g(\hat{m}) - \sum_{i=1}^n g_i(m_i, u_i) \leq \sum_{i=1}^n \mu_{i\beta}(m_i, u_i) \quad (5.34)$$

для всех (m, u) из $M \times U$

Доказательство. В силу того, что глобальная функция качества аддитивна и используется модификация локальных функций качества с нулевой суммой, аддитивное отображение $\sum V^n \rightarrow V$ представляет собой межуровневую функцию системы для каждого координирующего сигнала. Так как предполагается, что добавление аддитивной величины сохраняет порядок, будет сохраняться и строгое упорядочение; поэтому $\sum$ является строго монотонным отображением. Поскольку равенство

$$\sum_{i=1}^n g_{i\beta}(m_i, u_i) = \sum_{i=1}^n \mu_{i\beta}(m_i, u_i) + \sum_{i=1}^n g_i(m_i, u_i)$$

выполняется на $\mathcal{B} \times M \times U$, доказательство теоремы получается применением следствия теоремы 5.24.

Фиг. 5.6 иллюстрирует равенство (5.34) при $\tilde{u} = K(\tilde{m})$ и компенсирующем координирующем сигнале $\hat{\beta}$. Заметим, что поверх-

ность $\mu_{\beta}(m, \tilde{u}) = \sum_i \mu_{i\beta}(m_i, \tilde{u}_i)$ в пространстве $V \times M \times U$ пересекает гиперплоскость $v = 0$ по кривой $K(\tilde{m})$ $\tilde{u}$, так как используются модификации с нулевой суммой; эта кривая всегда расположена выше поверхности $\hat{v} = g(\tilde{m})$ на расстоянии от нее, равном $-g_0(m, \tilde{u}) = -\sum_i g_i(m_i, \tilde{u}_i)$.

7. КООРДИНИРУЕМОСТЬ ПРИ ИСПОЛЬЗОВАНИИ ПРИНЦИПА ПРОГНОЗИРОВАНИЯ ВЗАИМОДЕЙСТВИЙ

В настоящем разделе мы рассмотрим координируемость двухуровневых систем в случае, когда стратегия координации основана на принципе прогнозирования взаимодействий. Этот принцип можно охарактеризовать следующим образом: глобального оптимума можно достигнуть с помощью локальных действий, если правильно прогнозируются связующие входы. Для широкого класса двухуровневых систем это, к сожалению, не всегда так. Чаще приходится использовать еще и координацию путем изменения целей (модифицировать подходящим образом локальные функции качества). В этом разделе мы введем некоторые общие условия координируемости на основе принципа прогнозирования взаимодействий с использованием и без использования координации путем изменения целей.

Сначала напомним некоторые обозначения, которые использовались нами ранее. Чтобы подчеркнуть применяемый способ координации — прогнозирование взаимодействий и координация путем изменения целей, — будем считать, что множество координирующих сигналов есть множество $\mathcal{C} = \mathcal{A} \times \mathcal{B}$, где $\mathcal{A}$ — такое заданное подмножество U , что $K(M) \subseteq \mathcal{A}$, $\mathcal{B}$ есть множество сигналов, координирующих цели. Для всякого координирующего сигнала $\gamma = (\alpha, \beta)$ символ m^γ означает такой управляющий вход из M , что для каждого i , $1 \leq i \leq n$, локальное управляющее воздействие m_i^γ минимизирует $G_{i\beta}(m_i, \alpha_i)$ на множестве M_i . Если координация путем изменения целей не используется, управляющие воздействия m^γ зависят только от прогнозных значений связующих входов α ; в соответствии с этим мы в таких случаях будем обозначать m^γ через m^α .

Применимость принципа прогнозирования взаимодействий, также координируемость на его основе зависят от того, используется ли, а если да, то как именно, координация путем изменения целей. Мы рассмотрим два случая: 1) координация путем изменения целей не используется (применяется только способ прогнозирования взаимодействий) и 2) координация путем изменения целей осуществляется в соответствии с прогнозом связующих входов. Большинство применений указанного принципа охватывается

этими двумя случаями. Сформулируем еще раз вкратце основные положения, касающиеся координируемости.

1. Принцип прогнозирования взаимодействий применим (без координации путем изменения целей) тогда и только тогда, когда справедливо следующее утверждение:

$$(\forall \alpha) (\forall m^\alpha) (\exists \hat{m}) \{m = m^\alpha \text{ и } K(m) = \alpha \Rightarrow m = \hat{m}\}. \quad (5.35)$$

Двухуровневая система (без координации путем изменения целей) будет координируемой при использовании принципа прогнозирования взаимодействий тогда и только тогда, когда этот принцип применим и существует верный прогноз α ; иначе говоря, для некоторого α из $\mathcal{A}$ существует такое m^α , что $K(m^\alpha) = \alpha$.

Двухуровневая система, для которой используется только способ прогнозирования взаимодействий, вполне может и не быть координируема с помощью принципа прогнозирования взаимодействий (см. пример 5.9). Этот принцип не применим и в том случае, когда имеется правильный прогноз, который, однако, не приводит к глобально оптимальному поведению. Глобальный оптимум достигается при наличии верного прогноза α , только если α оптимально в том смысле, что $\alpha = K(\hat{m})$ для некоторого глобально оптимального управляющего воздействия $\hat{m}$. Даже если принцип применим, координируемость может отсутствовать, если все прогнозы, оптимальные в вышеуказанном смысле, оказываются неверными (т. е. не реализуются на практике). Приведенные причины некоординируемости можно устранить, если выполнить соответствующие модификации локальных функций качества в соответствии с прогнозными значениями связующих входов (пример 5.10).

2) Принцип прогнозирования (взаимодействий) применим при использовании отображения $\eta: M \rightarrow \mathcal{B}$ тогда и только тогда, когда для любого координирующего сигнала (α, β) , для которого $\beta = \eta(\alpha)$, выполняется следующее условие:

$$(\forall \gamma) (\forall m^\gamma) (\exists \hat{m}) \{m = m^\gamma \text{ и } q_\eta(m) = \gamma \Rightarrow m = \hat{m}\}, \quad (5.36)$$

где $q_\eta(m) = (K(m), \eta(m))$. Двухуровневая система координируема при использовании этого принципа при использовании η тогда и только тогда, когда принцип применим при использовании η и, кроме того, существует координирующий сигнал γ , для которого имеется такое m^γ , что $q_\eta(m^\gamma) = \gamma$.

Координируемость при фиксированных целях

Сначала рассмотрим случай, когда координация путем изменения целей не используется. Тогда соответствующие целевые свойства, необходимые для координируемости, представляют

собой свойства ограниченной согласованности. Результаты более или менее сходны с теми, которые относятся к связи свойств безусловной согласованности с координируемостью на основе принципов согласования.

Рассмотрим сначала одно из непосредственных следствий теоремы 5.5.

Теорема 5.26

Принцип прогнозирования взаимодействий применим и без изменения целей в любой двухуровневой системе, обладающей ограниченной межуровневой согласованностью.

Двухуровневая система могла бы и не обладать ограниченной межуровневой согласованностью, и все же принцип прогнозирования взаимодействий мог бы оказаться применимым без модификации целей. Так, мы могли бы рассмотреть случай, когда система не обладает ограниченной межуровневой согласованностью и все предсказания связующих входов являются неверными; однако в этом случае система была бы некоординируемой. Но при соответствующих условиях ограниченная межуровневая согласованность все же необходима в тех случаях, когда не используется координация путем изменения целей.

Теорема 5.27

Пусть двухуровневая система такова, что каждое $\alpha = K(m)$, $m \in M$, удовлетворяет n равенствам

$$\min_{M_i} g_i(m_i, \alpha_i) = \min_{(m \in M: K_i(m) = \alpha_i)} g_i(m_i, K_i(m))$$

всякий раз, когда правые части этих равенств существуют. Тогда существование в системе ограниченной межуровневой согласованности является необходимым и достаточным условием для применимости принципа прогнозирования взаимодействий без изменения целей.

Доказательство. Нам нужно доказать лишь необходимость. Предположим, что система не обладает ограниченной межуровневой согласованностью. Тогда существует такое $m \in M$, что для каждого i , $1 \leq i \leq n$, значение $g_i(m_i, K_i(m))$ является минимальным на множестве значений m' , для которых $K_i(m') = K_i(m)$. Однако управление m не является глобально оптимальным. В силу сделанного предположения существуют такие α из $\mathcal{A}$ и m^α , что $\alpha = K(m)$ и $m = m^\alpha$. Это несовместимо с применимостью рассматриваемого принципа, что и доказывает необходимость, а вместе с тем и всю теорему.

Условия теоремы 5.27 удовлетворяются, если функции взаимодействия подпроцессов K_i , $1 \leq i \leq n$, не зависят от локального управляющего воздействия m_i .

Не предполагая выполнения условий теоремы 5.27, мы все же можем вывести условие, являющееся одновременно необходимым и достаточным для применимости принципа без изменения целей. Это условие связано, к сожалению, не со свойствами ограниченной согласованности, а с глобально оптимальным управляющим воздействием.

Теорема 5.28

Для того чтобы принцип прогнозирования взаимодействий был применим в двухуровневой системе без изменения целей, необходимо и достаточно, чтобы каждое не являющееся глобально оптимальным управляющее воздействие $\tilde{m}$ из M удовлетворяло неравенству

$$\inf_{M_i} g_i(m_i, K_i(\tilde{m})) < g_i(\tilde{m}_i, K_i(\tilde{m})) \quad \text{для некоторого } i, \\ 1 \leq i \leq n.$$

Доказательство. Предположим, что принцип применим без модификации целей. Пусть $\tilde{m}$ — произвольное, но не глобально оптимальное управляющее воздействие из M , и положим $\alpha = K(\tilde{m})$. В этом случае должно быть $\tilde{m} \neq m^\alpha$ для любого m^α , ибо в противном случае применимость принципа была бы невозможна. Следовательно, $\tilde{m}$ удовлетворяет приведенному выше неравенству для некоторого i , $1 \leq i \leq n$. Предположим далее, что принцип не применим без модификации целей; тогда существуют α из $\mathcal{A}$ и не являющееся глобально оптимальным управляющее воздействие m^α , такие, что $\alpha = K(m^\alpha)$. Однако m^α не удовлетворяет неравенству теоремы 5.28 при любом i , $1 \leq i \leq n$. Этим доказательство завершается.

Применимость принципа прогнозирования взаимодействий без изменения целей сама по себе бесполезна, если нельзя получить верный прогноз. Если верный прогноз не может быть получен, то, хотя принцип и применим тривиальным образом без изменения целей, тем не менее система не будет координируема с помощью рассматриваемого принципа. Покажем теперь, что для координируемости с помощью данного принципа без изменения целей необходима ограниченная локальная согласованность.

Теорема 5.29

Если двухуровневая система координируема на основе принципа прогнозирования взаимодействий без использования модификации целей, она должна обладать ограниченной локальной согласованностью.

Доказательство. Если система координируема на основе данного принципа без модификации целей, то по определе-

существуют такие α из $\mathcal{A}$ и m^α , что $\alpha = K(m^\alpha)$. Поэтому оснований следствия из теоремы 5.5 система должна обладать ограниченной локальной согласованностью, что и доказывает теорему. Наличие ограниченной локальной согласованности недостаточно для координируемости. Точно так же ограниченная межуровневая согласованность не является необходимой для применимости принципа.

Найдем условие, при котором ограниченная локальная согласованность является одновременно необходимой и достаточной.

Теорема 5.30

Предположим, что принцип прогнозирования взаимодействий применим без изменения целей в двухуровневой системе, удовлетворяющей предположениям теоремы 5.27. Тогда наличие в системе ограниченной локальной согласованности необходимо и достаточно для координируемости с помощью данного принципа без изменения целей.

Доказательство. Нам нужно доказать лишь достаточность. Пусть система имеет ограниченную локальную согласованность. Тогда существует такое $\tilde{m} \in M$, что для каждого i , $1 \leq i \leq n$, величина $g_i(\tilde{m}_i, K_i(\tilde{m}))$ есть минимум на множестве всех m из M , для которых $K_i(m) = K_i(\tilde{m})$. Поэтому из предположений теоремы 5.27 следует существование таких α из $\mathcal{A}$ и m^α , что $K(m^\alpha) = \alpha$ и $\tilde{m} = m^\alpha$. Так как принцип применим, то $\tilde{m}$ является глобально оптимальным. Этим доказывается координируемость, а с ней и вся теорема.

Следствие

Предположим, что двухуровневая система удовлетворяет условиям теоремы 5.27. Тогда наличие в системе одновременно ограниченной межуровневой согласованности и ограниченной локальной согласованности является необходимым и достаточным условием для координируемости на основе принципа прогнозирования взаимодействий без изменения целей.

В начале этого раздела мы показали, каким образом двухуровневая система может оказаться и не координируемой на основе принципа прогнозирования взаимодействий без изменения целей. Такая некоординируемость может возникнуть даже тогда, когда взаимодействие между локальными решающими элементами осуществляется лишь через связующие входы процесса. Это кажется удивительным, однако подтверждается следующим примером.

Пример 5.9

Пусть $M_i = Y_i \cup U_i \cup R$, где R — множество действительных чисел, а $i = 1, 2$. Рассмотрим двухуровневую систему,

имеющую два подпроцесса P_1 и P_2 , задаваемые уравнениями

$$y_1 = m_1 \equiv P_1(m_1, u_1), \quad y_2 = 3m_2 - u_2 \equiv P_2(m_2, u_2),$$

связанные друг с другом с помощью связующих входов в соответствии с равенствами

$$u_1 = m_2 \equiv K_1(m), \quad u_2 = m_1 \equiv K_2(m).$$

Пусть локальные функции качества G_1 и G_2 задаются на R^2 в виде

$$G_1(m_1, y_1) = m_1^2 + (y_1 - 1)^2, \quad G_2(m_2, y_2) = m_2^2 + y_2^2,$$

и пусть глобальная функция качества является их суммой

$$G(m, y) = m_1^2 + m_2^2 + (y_1 - 1)^2 + y_2^2.$$

Предположим, что прогнозные значения связующих входов суть $\alpha = (3/20, 1/2)$. Тогда, минимизируя локальные затраты и принимая во внимание результаты прогноза, получаем

$$m_1^\alpha = \frac{1}{2}, \quad m_2^\alpha = \frac{3}{20}.$$

Следовательно, α представляет собой верный прогноз, но единственное глобально оптимальное управляющее воздействие есть $\hat{m} = (1/21, 1/7)$. Принцип прогнозирования взаимодействий без изменения целей поэтому не применим в этой системе. Кроме того, эта система обладает тем свойством, что $\alpha = K(\hat{m})$ не является верным прогнозом.

Попробуем объяснить, почему принцип прогнозирования взаимодействий не применим в приведенном выше примере. Хотя единственное взаимодействие между обоими решающими элементами осуществляется через связующие входы процесса, выбор локальных управлений должен делаться с учетом сведений обо всем процессе. Здесь важно, чтобы решающие элементы производили выбор локальных управлений как бы в союзе друг с другом, чтобы можно было учесть их соответственный вклад в глобальную функцию качества. Это в нашем примере не делается. Поскольку связующий вход подпроцесса P_1 не оказывает на него воздействия, выбор локального управления m_1 совершенно не зависит от выбора другого локального управления m_2 .

Координируемость при использовании модификации целей

Двухуровневая система, использующая для координации только способ прогнозирования взаимодействий, без модификации целей, легко может оказаться не координируемой на основе принципа прогнозирования взаимодействий. Мы только что убедились в этом на примере. Но если мы используем модификацию целей в сочетании с прогнозированием взаимодействий, оказы-

вается возможным достигнуть координируемости указанным принципом и в тех случаях, когда иными путями добиться координируемости не удастся.

Пример 5.10

Рассмотрим систему, описанную в примере 5.9, но с модификациями локальных функций качества для каждого β из $\mathcal{B} = R^2$, задаваемыми в виде

$$G_i(m_i, y_i) = G_i(m_i, y_i) + \beta_i m_i.$$

Для каждого прогноза $\alpha = (\alpha_1, \alpha_2)$ связующих входов вектор β задается следующими значениями:

$$\beta_1 = 2(\alpha_2 - 3\alpha_1), \quad \beta_2 = 0.$$

Тогда для данного прогноза α локально оптимальные управляющие воздействия таковы:

$$m_1^\alpha = \frac{1}{2}(1 - \alpha_2 + 3\alpha_1), \quad m_2^\alpha = \frac{3}{10}\alpha_2.$$

Осуществляя теперь объединение подпроцессов с помощью связующих входов $u_1 = m_2$ и $u_2 = m_1$, положим $m_1^\alpha = \alpha_2$ и $m_2^\alpha = \alpha_1$ и далее, имея в виду эти α_1 и α_2 , решим приведенные выше уравнения для m_i^α . Единственным правильным прогнозом будет $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2) = (1/7, 10/21)$.

Система координируема с помощью принципа прогнозирования взаимодействий при использовании отображения $\eta: M \rightarrow \mathcal{B}$, задаваемого в виде

$$\eta_1(m) = 2(m_1 - 3m_2), \quad \eta_2(m) = 0.$$

Отметим, что $-\eta_1(m)$ и $-\eta_2(m)$ соответственно являются линеаризациями полных операторов оценки косвенного эффекта управляющих воздействий $\Gamma_1(m)$ и $\Gamma_2(m)$ в точке m .

Этот пример показывает, что координация с использованием модификации целей может быть применена в сочетании с прогнозированием взаимодействий для того, чтобы скоординировать систему на основе принципа прогнозирования взаимодействий. Вообще говоря, это верно лишь тогда, когда имеется достаточная свобода модификации локальных функций качества. Поэтому вопрос о том, можно ли сделать систему координируемой на основе принципа прогнозирования с использованием модификации целей имеет смысл лишь в том случае, если модификации локальных функций качества задаются заранее или же сводятся к частным классам модификаций.

Пусть $\mathcal{B}$ — некоторое заданное множество возможных координирующих сигналов при использовании также и координации целей; каждое $\beta \in \mathcal{B}$ задает семейство локальных целевых функций

$g_{i\beta}$, $1 \leq i \leq n$. По существу мы считаем, что множество $\mathcal{B}$ определяет модификации локальных функций качества (пример 5.10). Вопрос о том, координируема ли система с помощью принципа прогнозирования при использовании модификации целей, сводится поэтому к вопросу о существовании такого отображения $\eta: M \rightarrow \mathcal{B}$, чтобы имело место (5.36) и, кроме того, существовали координирующий сигнал $\gamma = (\alpha, \beta)$ и управляющее воздействие m , такое, что $g_\eta(m^\gamma) = \gamma$.

Если существует отображение $\eta: M \rightarrow \mathcal{B}$, так что система координируема на основе принципа прогнозирования относительно η , и существует отображение $\theta: \mathcal{A} \rightarrow \mathcal{B}$, такое, что

$$\eta(m) = \theta(K(m)),$$

то подходящие координирующие сигналы β , влекущие за собой изменения целей, могут быть определены непосредственно в виде функции от прогнозных значений связующих входов α ; $\beta = \theta(\alpha)$. Стало быть, в данном случае координатору необходимо лишь правильно прогнозировать связующие входы.

Теорема 5.31

Предположим, что имеется двухуровневая система, для которой функция взаимодействия подпроцессов K является взаимно однозначной. Тогда для существования такого отображения $\eta: M \rightarrow \mathcal{B}$, при использовании которого был бы применим принцип прогнозирования, необходимо и достаточно, чтобы каждое управляющее воздействие $\hat{m}$ из M , которое не является глобально оптимальным, удовлетворяло неравенству

$$\inf_{M_i} g_{i\beta}(m_i, K_i(m)) < g_{i\beta}(\tilde{m}_i, K_i(\tilde{m}))$$

для некоторого $\beta \in \mathcal{B}$ и некоторого $1 \leq i \leq n$.

Доказательство. Предположим, что принцип прогнозирования применим при использовании некоторого отображения $\eta: M \rightarrow \mathcal{B}$. Пусть $\hat{m}$ — произвольное управляющее воздействие из M , не являющееся глобально оптимальным. Пусть $\gamma = (\alpha, \beta)$, где $\alpha = K(\hat{m})$ и $\beta = \eta(\hat{m})$. Но это может быть лишь в том случае, если $\hat{m} \neq m^\gamma$ для любого m^γ ; в противном случае предположение о применимости принципа не выполняется, что и доказывает необходимость.

Предположим теперь, что принцип прогнозирования не применим ни при каком отображении $\eta: M \rightarrow \mathcal{B}$. Тогда существует такое α из $\mathcal{A}$, что для всех $\beta \in \mathcal{B}$ существует m^γ , где $\gamma = (\alpha, \beta)$, которое не является глобально оптимальным, но $K(m^\gamma) = \alpha$. Так как K — взаимно однозначная функция, то $m^\gamma = \tilde{m}$, где $\alpha = K(\tilde{m})$.

Поэтому для всех β из $\mathcal{B}$ и каждого i , $1 \leq i \leq n$, имеем

$$g_{i\beta}(\tilde{m}_i, K_i(\tilde{m})) = \min_{M_i} g_{i\beta}(m_i, K_i(m)),$$

откуда следует достаточность. Тем самым теорема доказана.

Следствие

Предположим, что межуровневые функции двухуровневой системы строго монотонны, а функция взаимодействия подпроцессов K является взаимно однозначной. Тогда для существования такого отображения $\eta: M \rightarrow \mathcal{B}$, при использовании которого принцип прогнозирования будет применим, необходимо и достаточно, чтобы каждое управляющее воздействие $\tilde{m}$ из M , которое не является глобально оптимальным, удовлетворяло неравенству

$$\inf_M g_{\mathcal{B}}(\beta, m, K(\tilde{m})) < g(\tilde{m})$$

для некоторого β из $\mathcal{B}$.

Если функция взаимодействия подпроцессов не является взаимно однозначной, условия, выражаемые сформулированной выше теоремой и следствием из нее, являются только необходимыми условиями; предположение о том, что K является взаимно однозначной функцией, требовалось лишь для доказательства достаточности.

Рассмотрим теперь координируемость двухуровневой системы на основе принципа прогнозирования с использованием отображения $\eta: M \rightarrow \mathcal{B}$. К сожалению, при избранной нами общей постановке задачи мы должны предположить существование координирующего сигнала β из $\mathcal{B}$, модифицирующего цели и координирующего рассматриваемую систему на основе принципа согласования взаимодействий; иначе говоря, для модифицирующего цели сигнала β существует такая пара (m^β, u^β) , что $K(m^\beta) = u^\beta$, и при любых (m^β, u^β) управление m^β является глобально оптимальным всякий раз, когда $K(m^\beta) = u^\beta$. Однако это предположение является более сильным, нежели предположение об ограниченной локальной согласованности.

Теорема 5.32

Предположим, что для двухуровневой системы существует модифицирующий цели координирующий сигнал β из $\mathcal{B}$, который координирует ее на основе принципа согласования взаимодействий. Тогда, если принцип прогнозирования применим при использовании некоторого отображения $\eta: M \rightarrow \mathcal{B}$, то система координируема на основе этого принципа и при использовании другого отображения $\hat{\eta}: M \rightarrow \mathcal{B}$, не обязательно совпадающего с η .

Доказательство. Пусть $\hat{\beta}$ — модифицирующий цели координирующий сигнал из $\mathcal{S}$, который координирует систему по принципу согласования взаимодействий. Тогда существует такая пара значений $(\hat{m}, \hat{u}) = x^{\hat{\beta}}$, что $\hat{u} = K(\hat{m})$ и $\hat{m}$ является глобально оптимальным. Пусть $\hat{\alpha} = \hat{u} = K(\hat{m})$; тогда для каждого i , $1 \leq i \leq n$, управляющее воздействие $\hat{m}_i$ минимизирует $g_i(m_i, \hat{\alpha}_i)$ на M_i . Следовательно, $\hat{m} = m^\gamma$ для некоторого m^γ , где $\gamma = (\hat{\alpha}, \hat{\beta})$ и $K(m^\gamma) = \hat{\alpha}$. Кроме того, любое m^γ является глобально оптимальным, если $K(m^\gamma) = \hat{\alpha}$. Определим теперь $\hat{\eta}: M \rightarrow \mathcal{S}$ так, чтобы выполнялось равенство $\hat{\eta}(\hat{m}) = \hat{\beta}$ и чтобы $\hat{\eta}$ не находилось в противоречии с η для любого $m \neq \hat{m}$. Этим доказательство завершается.

Если принцип согласования взаимодействий применим для рассматриваемой двухуровневой системы, это обстоятельство в ряде случаев может оказаться весьма полезным.

Теорема 5.33

Пусть к данной двухуровневой системе применим принцип согласования взаимодействий. Предположим, что для каждого $\alpha \in \mathcal{A}$ существуют такие $\beta \in \mathcal{S}$ и (m^β, u^β) , что $\alpha = u^\beta$. Тогда существует такое отображение $\eta: M \rightarrow \mathcal{S}$, что применим принцип прогнозирования с использованием η . Кроме того, если система координируема на основе принципа согласования взаимодействий, система координируема также и при помощи принципа прогнозирования с использованием некоторого отображения $\eta: M \rightarrow \mathcal{S}$.

Доказательство. Определим такое $\theta: \mathcal{A} \rightarrow \mathcal{S}$, что для каждого $\alpha \in \mathcal{A}$ $\theta(\alpha) = \beta$ только тогда, когда существует такая пара (m^β, u^β) , что $\alpha = u^\beta$. По предположению такое отображение существует. Определим теперь $\eta: M \rightarrow \mathcal{S}$ так, чтобы для всех m из M выполнялось условие $\eta(m) = \theta(K(m))$. Мы утверждаем, что принцип прогнозирования применим при использовании отображения η . Если система координируема на основе одного из принципов согласования, существуют такое $\beta \in \mathcal{S}$ и такая пара (m^β, u^β) , что $u^\beta = K(m^\beta)$; поэтому, если $\theta(\alpha) = \beta$ для $\alpha = u^\beta$, то система координируема на основе принципа прогнозирования и отображения η . Этим доказательство завершается.

Если принцип согласования взаимодействий применим, а условие $K(M) \subseteq \mathcal{A}$ можно ослабить, мы всегда можем найти подходящее подмножество $\mathcal{A} \subseteq U$, такое, чтобы существовало отображение $\theta: \mathcal{A} \rightarrow \mathcal{S}$, которое было введено при доказательстве предыдущего предложения. Однако в некоторых случаях единственным подходящим подмножеством все же является множество, состоящее из одного элемента $\{\hat{\alpha}\}$, где $\hat{\alpha} = K(\hat{m})$ для некоторого

глобально оптимального управляющего воздействия $\hat{m}$. Это иллюстрируется следующим примером.

Пример 5.11

Пусть $M_i = Y_i = U_i = R$, где R — множество действительных чисел, а $i = 1, 2$. Рассмотрим двухуровневую систему, имеющую два подпроцесса P_i , задаваемых уравнениями

$$y_i = m_i + u_i \equiv P_i(m_i, u_i)$$

со связывающими их соотношениями

$$u_1 = \frac{17}{2} m_2 \equiv H_1(m, y), \quad u_2 = m_1 \equiv H_2(m, y).$$

Пусть глобальная функция качества G задана на $M \times Y$ в виде квадратичной формы

$$G(m, y) = (m_1 - m_2)^2 + y_1^2 + y_2^2.$$

Глобально оптимальное управляющее воздействие есть $\hat{m} = (0, 0)$, и оно является единственным.

Пусть $\mathcal{R} = R^2$, и пусть для каждого $\beta = (\beta_1, \beta_2)$ из $\mathcal{R}$ локальные функции качества $G_{\beta i}$ заданы в следующем виде:

$$G_{\beta i}(m_i, y_i, u_i) = m_i^2 - a_i m_i u_i + y_i^2 + (\beta_i - 1) u_i^2,$$

где $a_1 = 2/17$ и $a_2 = 1$. (Эти модификации локальной функции качества $G_{\beta i}$ не являются модификациями с нулевой суммой, и можно доказать, что система не обладает свойством монотонности.)

Система координируема при использовании принципа согласования взаимодействий. Пусть $\beta = (\beta_1, \beta_2)$ — произвольный, координирующий цели вход. Если $\beta_1 \geq 128/289$ и $\beta_2 \geq 1/8$, то оптимальные решения (m_1^β, u_1^β) и (m_2^β, u_2^β) локальных задач существуют, а именно: $(m_i^\beta, u_i^\beta) = (0, 0)$ при $i = 1, 2$. В противном случае одна из локальных задач не будет иметь оптимального решения. Поэтому принцип согласования взаимодействий применим, и любое β из $\mathcal{R}$ является компенсирующим тогда и только тогда, когда $\beta_1 \geq 128/289$ и $\beta_2 \geq 1/8$.

Отсюда следует, что отображение $\theta: \mathcal{A} \rightarrow \mathcal{R}$, введенное при доказательстве теоремы 5.33, существует лишь тогда, когда $\mathcal{A}$ является множеством из одного элемента $\{(0, 0)\}$. Кроме того, для любого подмножества $\mathcal{A} \subseteq U$, $\mathcal{A} \neq \{(0, 0)\}$, не существует такого отображения $\theta: \mathcal{A} \rightarrow \mathcal{R}$, чтобы принцип прогнозирования был применим при использовании $\theta: M \rightarrow \mathcal{R}$ и $\eta(m) = \theta(K(m))$. Это можно показать следующим образом.

Для каждого координирующего сигнала $\gamma = (\alpha, \beta)$ локальные

задачи имеют единственные оптимальные решения m_1^y и m_2^y :

$$m_1^y = -\frac{8}{17} \alpha_1, \quad m_2^y = \frac{1}{4} \alpha_2.$$

Заметим, что m_1^y не зависит от β ; поэтому, если существует хотя бы одно такое отображение $\theta: \mathcal{A} \rightarrow \mathcal{B}$, достаточно взять любое отображение из $\mathcal{A}$ в $\mathcal{B}$. Выберем $\alpha_1 = 17$ и $\alpha_2 = 8$. Тогда вне зависимости от β $m_1^y = -8$, $m_2^y = 2$; эти управляющие воздействия порождают связующие входы $u_1 = 17/2 m_2^y = 17$ и $u_2 = m_1^y = -8$. Следовательно, α есть правильное предсказание, но управляющее воздействие $m^y = (-8, 2)$ не является глобально оптимальным. В действительности любой прогноз α , такой, что $17\alpha_2 = -8\alpha_1$, является верным прогнозом, но не приводит к глобально оптимальному управляющему воздействию.

Мы утверждаем, что в рассматриваемом примере принцип прогнозирования не применим, каково бы ни было отображение $\eta: M \rightarrow \mathcal{B}$. Путем непосредственных вычислений можно показать, что управляющее воздействие $m = (-8, 2)$ не является глобально оптимальным и не удовлетворяет неравенству теоремы 5.31 для любого $\beta \in \mathcal{B}$ и каждого $i = 1, 2$.

В теореме 5.33 мы сделали вывод, что при подходящих условиях координируемость при использовании какого-либо из принципов согласования есть достаточное условие для координируемости на основе принципа прогнозирования с использованием некоторого отображения $\eta: M \rightarrow \mathcal{B}$. Следующий пример показывает, что координируемость на основе одного из принципов согласования не является необходимой.

Пример 5.12

Рассмотрим систему, описанную в примере 5.10. Мы уже нашли в этом примере отображение $\eta: M \rightarrow \mathcal{B}$, при использовании которого система координируема при помощи любого принципа согласования; однако в данном случае ни один из принципов согласования не применим к исходной системе.

Для всякого фиксированного сигнала $\beta \in \mathcal{B}$, координирующего цели, локально оптимальные решения таковы:

$$m_1^\beta = \frac{1}{2} \beta_1; \quad u_1^\beta \text{ — произвольно;}$$

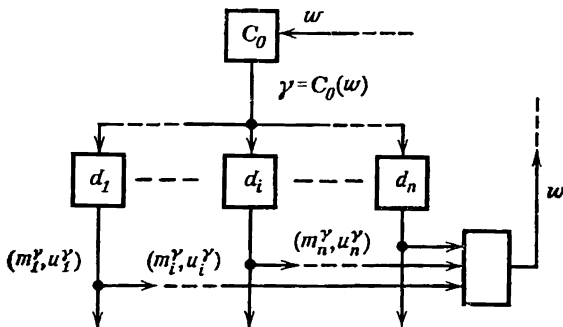
$$m_2^\beta = -\frac{1}{2} \beta_2; \quad u_2^\beta = -\frac{3}{2} \beta_2.$$

Однако $u^\beta = K(m^\beta)$ тогда и только тогда, когда $\beta_1 = -3\beta_2$. Если теперь $\beta_1 = -3\beta_2$, то $m_1^\beta = 3m_2^\beta$. Глобально оптимальное управляющее воздействие есть $\hat{m} = (1/21, 1/7)$, и поэтому оно никогда не достигается, если удовлетворяется условие согласования $u^\beta = K(m^\beta)$.

8. ПРОЦЕСС КООРДИНАЦИИ

Теперь мы можем перейти к проблеме синтеза блока координации двухуровневой системы. В случае оптимизирующих систем координация двухуровневой системы достигается подачей оптимального координирующего воздействия на локальные решающие элементы; следовательно, проблема синтеза состоит в том, чтобы найти преобразование, которое на основе информации, поступающей от нижележащих уровней (локальных решающих элементов и подпроцессов), позволяет найти оптимальное координирующее воздействие. Рассмотрим два подхода к этой проблеме. Первый подход основан на применимости принципа координации и приводит к итеративной процедуре, подразумевающей участие решающих элементов обоих уровней. При втором подходе задача, решаемая на вышестоящем уровне (задача, которую координатор решает для получения оптимального координирующего воздействия), определяется как оптимизационная задача без какой-либо непосредственной связи с принципами координации.

Иерархия управления или принятия решений в двухуровневой системе, рассматриваемая в настоящей главе, еще раз представлена на фиг. 5.7. Координатор C_0 получает информацию w от локальных решающих элементов (локальных блоков принятия решений) d_i , выбирает координирующий сигнал γ , основанный на данной информации, и посылает его каждому из элементов d_i . Получив определенный координирующий сигнал γ , каждый локальный решающий элемент d_i «принимает» то решение $(m_i^\gamma, u_i^\gamma) = d_i(\gamma)$, которое является оптимальным решением конкретной задачи $\mathcal{D}_i(\gamma)$ при условии, что такое решение существует, и посылает определенную информацию об этом обратно к координатору. Можно синтезировать блок координации C_0 , выбрав ту или иную стратегию координации $s_0: \mathcal{C} \times \mathcal{H} \rightarrow \mathcal{C}$, которая будет использоваться для получения оптимального координирующего сигнала. Стратегия координации зависит не только от того, какое отображение



Фиг. 5.7. Двухуровневая иерархия принятия решений.

применено для выработки новых координирующих сигналов, но также и от характера информации, поступающей от локальных элементов.

Применение принципов координации

Если данная двухуровневая система координируема с помощью одного из принципов координации, то можно довольно легко подобрать стратегию координации, которая при подходящих условиях даст оптимальное координирующее воздействие. Предположим, что двухуровневая система координируема на основе выбранного принципа координации; кроме того, предположим, что для любого выбранного координирующего сигнала можно установить, будут ли получающиеся при этом локальные задачи иметь оптимальные решения и удовлетворяется ли условие координируемости. Если условие координируемости не удовлетворяется для выбранного координирующего сигнала, координатор может воспользоваться получившейся в результате несогласованностью для модификации и замены координирующего воздействия.

Пусть для выбранного координирующего сигнала γ в двухуровневой системе символы $m(\gamma)$ и $u(\gamma)$ означают соответственно управляющее воздействие из M и связующий вход из U , такие что для каждого i , $1 \leq i \leq n$, пара $(m_i(\gamma), u_i(\gamma))$ есть оптимальное решение конкретной i -й локальной задачи $\mathcal{D}_i(\gamma)$. Информация, содержащаяся в координирующем сигнале γ , зависит от способов координации, применяемых в системе. Пара преобразований m и u характеризует в агрегированной форме локальные решающие элементы двухуровневой системы. Следует заметить, что $m(\gamma)$ и $u(\gamma)$ могут и не существовать для некоторых координирующих сигналов ¹⁾.

Предположим, что координатор посылает координирующий сигнал γ , но при этом либо $m(\gamma)$ и $u(\gamma)$ не существуют, либо, если они существуют, условие координируемости для данного принципа координации не удовлетворяется. Тогда координатор должен модифицировать или заменить γ . Информация $w(\gamma)$, поступающая от локальных элементов при получении ими координирующего сигнала γ , должна быть, следовательно, такой, чтобы координатору было достаточно знать $w(\gamma)$ и γ , чтобы определить, являются ли локальные решения оптимальными и удовлетворяют ли они условию координируемости. Если стратегия $s_0: \mathcal{E} \times \mathcal{W} \rightarrow \mathcal{E}$, используемая координатором, такова, что $s_0(\gamma, w(\gamma)) = \gamma$ тогда и только тогда, когда $m(\gamma)$ и $u(\gamma)$ суще-

¹⁾ Изменение в обозначениях вызвано желанием подчеркнуть то обстоятельство, что управляющие воздействия и связующие входы, выбираемые локальными решающими элементами, зависят (возможно, функционально) от координирующих сигналов.

ствуют и удовлетворяют условию координируемости для данного принципа координации, мы будем говорить, что стратегия s_0 основана на данном принципе.

Как только определена стратегия координации, сразу же определяется такое преобразование $T: \mathcal{C} \rightarrow \mathcal{C}$, что

$$T(\gamma) = s_0(\gamma, w(\gamma)).$$

Повторное применение преобразования T представляет собой итерационный процесс, подразумевающий участие решающих элементов обоих уровней. После k итераций, начинающихся с координирующего сигнала γ^0 , координатор направляет координирующий сигнал γ^k к локальным решающим элементам. Локальные элементы вырабатывают решения $m(\gamma^k)$ и $u(\gamma^k)$, если стоящие перед ними локальные задачи имеют оптимальные решения, причем эти элементы посылают координатору информацию $w(\gamma^k)$, зависящую от принятых ими решений. Затем координатор отыскивает новый координирующий сигнал $\gamma^{k+1} = s_0(\gamma^k, w(\gamma^k))$. Стратегия координации s_0 должна быть такой, чтобы получаемая последовательность координирующих сигналов сходилась к оптимальному координирующему сигналу (или заканчивалась на нем).

Пример 5.13

Пусть $M_i = Y_i = U_i = R$, где $i = 1, 2$. Рассмотрим двухуровневую систему, имеющую два подпроцесса P_1 и P_2 , описываемые уравнениями

$$y_1 = m_1 - u_1 \equiv P_1(m_1, u_1), \quad y_2 = m_2 - 2u_2 \equiv P_2(m_2, u_2)$$

со связующими входами

$$u_1 = y_2 - \frac{1}{2}m_2 \equiv H_1(m, y), \quad u_2 = y_1 - \frac{1}{2}m_1 \equiv H_2(m, y).$$

Процесс в целом $P: M \rightarrow Y$ и функция взаимодействия подпроцессов $K: M \rightarrow U$ тогда таковы:

$$P(m) = Pm, \quad K(m) = Km,$$

где P и K — матрицы размерности 2×2

$$P = \begin{bmatrix} 0 & \frac{1}{2} \\ 1 & 0 \end{bmatrix}, \quad K = \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Пусть глобальная функция качества G задана на $M \times Y$ в виде суммы $G(m, y) = G_1(m_1, y_1) + G_2(m_2, y_2)$ локальных функций качества G_1 и G_2 , которые являются квадратичными формами

$$G_i(m_i, y_i) = m_i^2 + (y_i - 1)^2,$$

где $i = 1, 2$. Глобально оптимальное управляющее воздействие, найденное путем минимизации $g(m) = G(m, P(m))$ на M , есть $\hat{m} = (1/2, 0)$ и оно единственно.

Пусть задано множество координирующих воздействий $\mathcal{R} = R^2$ и пусть для каждого координирующего сигнала $\beta \in \mathcal{R}$ определены модифицированные локальные функции качества $G_{i\beta}$ в виде

$$G_{1\beta}(m_1, u_1, y_1) = G_1(m_1, y_1) + \beta_1 u_1 + \frac{1}{2}(\beta_2 - 2\beta_1)m_1,$$

$$G_{2\beta}(m_2, u_2, y_2) = G_2(m_2, y_2) + \beta_2 u_2 + \frac{1}{2}(\beta_1 - \beta_2)m_2.$$

Модификации локальных функций качества являются модификациями с нулевой суммой, (т. е. скомпенсированными), и поэтому применим принцип согласования взаимодействий; более того, система координируема на основе данного принципа. Локальные задачи имеют единственные оптимальные решения для каждого координирующего сигнала β

$$\begin{aligned} m_1(\beta) &= -\frac{1}{4}\beta_2, & u_1(\beta) &= -\left(1 + \frac{1}{2}\beta_1 + \frac{1}{4}\beta_2\right), \\ m_2(\beta) &= -\frac{1}{4}\beta_1, & u_2(\beta) &= -\frac{1}{2}\left(1 + \frac{1}{4}\beta_1 + \frac{1}{4}\beta_2\right) \end{aligned} \quad (5.37)$$

Координирующий сигнал $\hat{\beta} = (-8/5, -2)$ координирует систему при использовании этого принципа.

Координатор может прийти к $\hat{\beta}$, если он, выбирая координирующее воздействие, последовательно улучшает его в соответствии со следующим правилом:

$$\beta' = \beta + \lambda [u(\beta) - Km(\beta)], \quad (5.38)$$

где λ — соответствующая диагональная матрица. Подставляя в (5.38) выражения для $m(\beta)$ и $u(\beta)$, получаем β' как функцию только от β :

$$\beta'_1 = \left(1 - \frac{3}{8}\lambda_{11}\right)\beta_1 - \lambda_{11}, \quad \beta'_2 = \left(1 - \frac{3}{4}\lambda_{22}\right)\beta_2 - \frac{1}{2}\lambda_{22}.$$

Эти два выражения определяют преобразование $T: \mathcal{R} \rightarrow \mathcal{R}$. Так как оптимальные решения локальных задач единственны и система координируема на основе принципа согласования взаимодействий, то преобразование T имеет неподвижную точку, которая является оптимальным координирующим сигналом $\hat{\beta}$, т. е. $T(\hat{\beta}) = \hat{\beta}$. Если мы выберем диагональную матрицу λ так, чтобы выполнялись условия $0 < \lambda_{11} < 16/3$ и $0 < \lambda_{22} < 8/3$, то повторное применение T даст последовательность координирующих сигналов, которая сходится к оптимальному координирующему

сигналу β по норме $\|x\| = \max_i |x_i|$. Сходимость не зависит от начального выбора β .

Опираясь на рассмотренный пример, мы можем теперь перейти обсуждению на абстрактном уровне проблемы синтеза блока координации, исходя из стратегий координации, основанных на принципах согласования.

Стратегия координации, основанная на принципе согласования взаимодействий, есть отображение $s_0: \mathcal{R} \times U \times U \rightarrow \mathcal{R}$ где $\mathcal{R}$ — такое множество координирующих цели сигналов, что условие

$$[s_0(\beta, u, u') = \beta] \Rightarrow [u = u'] \quad (5.39)$$

удовлетворяется на всей области определения s_0 , а соответствующее преобразование $T: \mathcal{R} \rightarrow \mathcal{R}$ удовлетворяет соотношению

$$T(\beta) = s_0(\beta, u(\beta), K(m(\beta)))$$

на $\mathcal{R}$. Смысл условия (5.39) очевиден: если для выбранного координирующего сигнала β условие координируемости $u(\beta) =$

$K(m(\beta))$ удовлетворяется, никакой корректировки не требуется. Поэтому, если принцип согласования взаимодействий применим для рассматриваемой системы, любая неподвижная точка преобразования T является оптимальным координирующим сигналом, так как

$$[T(\beta) = \beta] \Rightarrow [u(\beta) = K(m(\beta))].$$

Поэтому T имеет неподвижную точку только в том случае, если система координируема на основе этого принципа согласования. Обратно, если система координируема с помощью принципа согласования, то преобразование T имеет неподвижную точку при условии, что оптимальные решения локальных задач единственны. Если множество координирующих сигналов наделено соответствующей структурой, скажем метрикой, то стратегия координации s_0 должна быть такой, чтобы преобразование T было отображением сжатия.

Стратегия координации, основанная на принципе согласования функций качества, принимает форму, соответствующую условию координируемости для этого принципа. Если данная двухуровневая система обладает свойством строгой монотонности, то примером стратегии координации, основанной на согласовании функций качества, может служить отображение $s_0: \mathcal{R} \times V \times V \rightarrow \mathcal{R}$, удовлетворяющее условию

$$[s_0(\beta, v, v') = \beta] \Rightarrow [v = v']$$

на области своего определения; в этом случае преобразование $T: \mathcal{R} \rightarrow \mathcal{R}$ имеет вид

$$T(\beta) = s_0(\beta, g_{\mathcal{R}}(\beta, m(\beta), u(\beta)), g(m(\beta))),$$

где $g_{\mathcal{R}}$ и g — соответственно кажущаяся глобальная и глобальная целевые функции.

Стратегия координации, основанная на принципе прогнозирования взаимодействий и использующая модификацию целей задается отображением $\eta: M \rightarrow \mathcal{R}$, и она должна учитывать не только точность прогноза связующих сигналов, но также и предполагаемые эффекты координации целей. Следующие примеры показывают, как это может быть осуществлено.

Пример 5.14

Рассмотрим систему, описанную в примере 5.13, с тем отличием, что теперь принцип прогнозирования взаимодействий с использованием координации целей будет применяться как средство для координации системы. Тогда для каждого координирующего сигнала, представленного парой (α, β) , где α — прогнозные значения связующих входов, соответствующие локальные задачи имеют единственные оптимальные решения:

$$m_1(\alpha, \beta) = \frac{1}{8}(4 + 4\alpha_1 + 2\beta_1 - \beta_2),$$

$$m_2(\alpha, \beta) = \frac{1}{8}(4 + 8\alpha_2 - \beta_1 + \beta_2).$$

Принцип прогнозирования взаимодействий применим к системе при использовании отображения $\eta: M \rightarrow \mathcal{R}$, определяемого в виде $\eta(m) = \theta(K(m))$, где $\theta: \mathcal{A} \rightarrow \mathcal{R}$ находится из соотношений

$$\theta_1(\alpha) = 4(-\alpha_1 + 2\alpha_2),$$

$$\theta_2(\alpha) = 4(\alpha_1 - 4\alpha_2 - 1),$$

получающихся из уравнений (5.37) подстановкой $\alpha = u(\beta)$ и последующим их решением относительно β , как функции от α . Легко видеть, что система координируема с помощью принципа прогнозирования при использовании отображения η .

Путем простых алгебраических преобразований можно проверить, что для любого координирующего воздействия в виде пары (α, β) условие: $\beta = \eta(m(\alpha, \beta))$ и $\alpha = Km(\alpha, \beta)$ влечет за собой $m(\alpha, \beta) = m = (1/2, 2/5)$. Более того, пара $(\hat{\alpha}, \hat{\beta})$, где $\hat{\alpha} = (3/10, -1/20)$ и $\hat{\beta} = (-8/5, -2)$, есть та единственная пара координирующих сигналов, для которой $\beta = \eta(m(\hat{\alpha}, \hat{\beta}))$ и $\hat{\alpha} = Km(\hat{\alpha}, \hat{\beta})$.

Пара $(\hat{\alpha}, \hat{\beta})$ может быть найдена путем последовательного улучшения выбираемых координирующих сигналов в соответствии со стратегией координации

$$\alpha' = \alpha + \mu[\beta - \eta(m(\alpha, \beta))], \quad \beta' = \beta + \lambda[\alpha - Km(\alpha, \beta)].$$

где λ и μ — соответствующие диагональные матрицы. Выразив теперь правые части в явном виде через α и β , мы получим преобразование $T_\eta: \mathcal{A} \times \mathcal{B} \rightarrow \mathcal{A} \times \mathcal{B}$ в виде системы уравнений:

$$\alpha'_1 = \alpha_1 + \frac{1}{4} \mu_{11} (16\alpha_1 - 24\alpha_2 + 15\beta_1 - 7\beta_2 + 4),$$

$$\alpha'_2 = \alpha_2 + \frac{1}{4} \mu_{22} (-24\alpha_1 + 40\alpha_2 - 17\beta_1 + 15\beta_2 + 12),$$

$$\beta'_1 = \beta_1 + \frac{1}{16} \lambda_{11} (8\alpha_1 + 8\alpha_2 - 5\beta_1 + 3\beta_2 - 4),$$

$$\beta'_2 = \beta_2 + \frac{1}{16} \lambda_{22} (4\alpha_1 + 8\alpha_2 + 3\beta_1 - 2\beta_2).$$

Если λ и μ выбраны так, что их нормы достаточно малы, то последовательное применение T_η даст нам последовательность координирующих сигналов, которая сходится к оптимальной паре (координирующему сигналу) $(\hat{\alpha}, \hat{\beta})$.

Стратегия координации в этом примере учитывает взаимозависимость между правильностью прогноза и координацией целей. Следует отметить, что в общем случае соотношение $\beta = \theta(\alpha)$ может выполняться не для всех пар (α, β) последовательности координирующих сигналов, порождаемой преобразованием T_η вышеприведенного примера. Если мы потребуем, чтобы все выбираемые пары (α, β) удовлетворяли этому условию, то, подобрав соответствующим образом μ , мы сможем использовать данное правило для внесения поправок в прогнозы.

Пример 5.15

Рассмотрим систему, описанную в примере 5.14, но предположим теперь, что координатор ограничен в своих действиях правом выбора только таких координирующих сигналов, которые удовлетворяют условию $\beta = \theta(\alpha)$. Так как η — композиция θ и K , то координатор должен лишь вносить поправку в прогноз связующих входов согласно правилу

$$\alpha' = \alpha + \mu [\beta - \eta(m(\alpha, \beta))],$$

где $\beta = \theta(\alpha)$; при этом новый сигнал, координирующий цели, находится из соотношения $\beta' = \theta(\alpha')$. Такая «корректировка» прогноза означает применение преобразования $T_0: \mathcal{A} \rightarrow \mathcal{A}$, определяемого соотношениями

$$\alpha'_1 = \alpha_1 + \mu_{11} (-18\alpha_1 + 52\alpha_2 + 8),$$

$$\alpha'_2 = \alpha_2 + \mu_{22} (26\alpha_1 - 84\alpha_2 - 12),$$

которые получаются из нашего правила после подстановки $\beta = \theta(\alpha)$ и последующего исключения β . Если μ_{11} и μ_{22} достаточно близки к нулю, то повторное применение T_0 порождает последовательность прогнозов связующих входов, которая сходится к $\hat{\alpha}$.

Это происходит благодаря однозначности функции θ : если $\beta = \theta(\alpha)$ и $\beta = \eta(m(\alpha, \beta))$, то $\alpha = Km(\alpha, \beta)$. Если бы функция θ не обеспечивала взаимно однозначного отображения, то последовательность сходилась бы к некоторому α , отличному от $\hat{\alpha}$.

В общем случае стратегия координации, основанная на принципе прогнозирования с использованием определенного отображения $\eta: M \rightarrow \mathcal{R}$, есть отображение $s_0: \mathcal{C} \times \mathcal{R} \times U \rightarrow \mathcal{C}$, где $\mathcal{C} = \mathcal{A} \times \mathcal{R}$, такое, что условие

$$[s_0((\alpha, \beta), \beta', u) = (\alpha, \beta)] \Rightarrow [\beta' = \beta \text{ и } u = u^\alpha]$$

удовлетворяется на всей области определения этого отображения; соответственно преобразование, отображающее $\mathcal{C}$ на самое себя, представляется в виде

$$T_\eta(\alpha, \beta) = s((\alpha, \beta), \eta(m(\alpha, \beta)), K(m(\alpha, \beta)))$$

на $\mathcal{C}$. Если координатору необходимо найти только такие координирующие воздействия, которые удовлетворяют условию $\beta = \theta(\alpha)$, где отображение $\theta: \mathcal{A} \rightarrow \mathcal{R}$ определяется так, что $\eta(m) = \theta(K(m))$, то стратегия координации s_0 должна удовлетворять дополнительному условию

$$[s_0(\alpha, \beta), \beta' \text{ и } (\alpha' \beta')] \Rightarrow [\beta' = \theta(\alpha')]$$

на своей области определения.

Если для данной системы принцип прогнозирования применим при использовании некоторого отображения η , то любая неподвижная точка преобразования T_η является оптимальной парой, составляющей координирующий сигнал, и, следовательно, если T_η имеет неподвижную точку, система координируема с помощью этого принципа относительно η . На стратегию координации должны быть наложены дополнительные условия с тем, чтобы T_η было в определенном смысле отображением сжатия.

Описываемые нами здесь итерационные процессы, основанные на принципах координации, будут изучены более подробно в гл. 6.

Решение проблемы координации для данной двухуровневой системы, координируемой на основе выбранного принципа координации, может быть получено как оптимальное решение подходящим образом определенной оптимизационной задачи. Чтобы проиллюстрировать это, предположим, что задано отображение $G_0: U \times U \rightarrow V$ удовлетворяющее условию

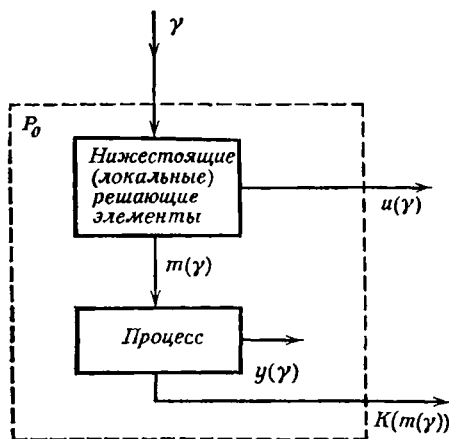
$$[G_0(u, u') = \inf G_0] \Rightarrow [u = u'].$$

Сформулируем теперь оптимизационную задачу для координатора C_0 : найти такое γ из $\mathcal{C}$, где $\mathcal{C}$ — множество координирующих сигналов системы, для которого выполняется условие

$$G_0(u(\hat{\gamma}), K(m(\hat{\gamma}))) = \min_{\mathcal{C}} G_0(u(\gamma), K(m(\gamma)))$$

в предположении, что оптимальные решения локальных задач существуют для каждого координирующего сигнала из $\mathcal{E}$. Тогда для координатора выходная функция является отображением $P_0: \mathcal{E} \rightarrow U \times U$, где $P_0(\gamma) = (u(\gamma), K(m(\gamma)))$; она учитывает деятельность всех локальных решающих элементов, а также все подпроцессы (фиг. 5.8).

Если принцип согласования взаимодействий применим для данной системы, можно положить $\mathcal{E} = \mathcal{R}$. Тогда любое оптимальное решение задачи вышестоящего уровня, сформулированное подобным образом, есть оптимальный координирующий сигнал. Следует, однако, указать, что сложность функции P_0 препятствует практическому применению такого подхода. Оптимизационная задача для координатора может быть решена подходящим итерационным методом, построенным исходя из выбранной стратегии координации, которая, в свою очередь, основана на принципе согласования взаимодействий. Однако эти стратегии предназначены для непосредственного решения задачи координации, и поэтому нет необходимости вводить функцию качества для вышестоящего уровня и, тем более, решать специальную оптимизационную задачу. Можно, конечно, использовать стратегии координации для решения подобной оптимизационной задачи вышестоящего уровня, но, на наш взгляд, это излишне и может привести к недоразумениям.



Ф и г. 5.8.

Оптимизационный подход

Проблема принятия решений на вышестоящем уровне (для двухуровневой системы — на уровне координатора) может быть сформулирована как оптимизационная задача без каких-либо прямых ссылок на принципы координации. Мы опишем два таких подхода.

Задача, подлежащая решению на вышестоящем уровне, как глобальная задача

Самый очевидный способ сформулировать задачу, решаемую на вышестоящем уровне, состоит в представлении ее как глобальной задачи оптимизации для данной двухуровневой системы.

Однако необходимо ввести некоторые уточнения, ибо следует учесть, что решения, принимаемые координатором, связаны с выбором координирующих, а не управляющих воздействий. Функция качества для координатора может поэтому быть глобальной функцией качества G_0 , но тогда выходная функция P_0 должна быть отображением $P_0: \mathcal{C} \rightarrow M \times Y$. Поэтому задача, решаемая на вышестоящем уровне, в этом случае равносильна отысканию такого координирующего сигнала $\hat{\gamma} \in \mathcal{C}$, что

$$G_0(P_0(\hat{\gamma})) = \min_{\mathcal{C}} G_0(P_0(\gamma)).$$

Для того чтобы при таком выборе координирующего сигнала система действительно была скоординирована, выходная функция P_0 должна быть задана в следующем виде:

$$P_0(\gamma) = (m(\gamma), P(m(\gamma))). \quad (5.40)$$

Необходимо заметить, что выходная функция P_0 будет иметь очень сложный вид, так как P_0 включает в себя не только результирующую взаимосвязь между управляющими воздействиями и выходами всего процесса в целом P , но также и действия локальных управляющих элементов, связанные с выбором и подачей локальных управляющих воздействий (именно с их помощью и осуществляется действительное управление процессом). Иными словами, эта функция должна вобрать в себя и весь процесс в целом, и деятельность всех локальных управляющих элементов.

Очевидно, что такой подход к задаче вышестоящего уровня практически не реализуем. Он упомянут здесь лишь для сравнения с другими более реалистичными подходами. А именно, локальные оптимизационные задачи часто бывают предопределены, и задача координатора заключается именно в том, чтобы побудить локальные решающие элементы найти решение глобальной оптимизационной задачи, т. е. отыскать глобально оптимальные управляющие воздействия. Если проблема, решаемая на вышестоящем уровне, представляет собой определенную выше оптимизационную задачу с использованием выходной функции P_0 , задаваемой соотношением (5.40), то задача отыскания оптимального координирующего сигнала, вероятно, окажется даже более трудной, чем решение глобальной задачи.

Трудности, возникающие при таком подходе, вызваны тем, что все взаимодействия локальных элементов при таком рассмотрении оказываются связанными в единый нерасчлененный процесс: подсистемы явным образом нигде не фигурируют. Однако в этом смысле наличие подсистем дает определенные преимущества, ибо позволяет «распределить» между ними труд по отысканию глобального решения; тогда для координатора можно сформулировать более простую задачу. При этом глобальную задачу можно,

как правило, решить путем решения в первую очередь задачи вышестоящего уровня, которая значительно проще, чем сформулированная выше, и тем более проще глобальной задачи. В случае оптимизационных систем глобальный оптимум достигается координатором более экономными средствами, если имеются подходящие условия для применения принципов координации. Часто одна из главных целей построения многоуровневой иерархической системы состоит в уменьшении общих усилий, затрачиваемых на решение задач или управление. Это становится возможным только благодаря использованию какого-нибудь эффективного подхода, например основанного на применении принципов координации.

Но мы хотели бы еще раз подчеркнуть, что экономия усилий — не единственное разумное основание для создания многоуровневой системы. Как уже указывалось в гл. 1, ограничения, налагаемые на размер и сложность подсистем, также играют роль при выборе структуры системы. Что касается социальных систем, то здесь даже в том случае, когда руководителю или координатору известно оптимальное решение, остается еще проблема его реализации, которая представляет обычно колоссальную трудность.

Принципиальный недостаток непосредственного использования глобальной функции качества G при постановке оптимизационной задачи для вышестоящего уровня, помимо сложности такой задачи, состоит в том, что G не зависит явным образом от выбора координирующих сигналов γ , т. е. от переменной, варьируемые которой находятся во власти координатора.

Поэтому на основании самой функции G трудно оценить, является ли тот или иной выбор γ подходящим, а если нет, то как его улучшить. Следовательно, не существует, по-видимому, легкого способа «разделить» между двумя уровнями усилия, затрачиваемые на отыскание глобального оптимума. Если глобальная задача выносится на второй уровень, координатор фактически должен сначала решить эту глобальную задачу, а затем, исходя из полученного решения, отыскать требуемое координирующее воздействие. Естественно, что если деятельность локальных решающих элементов не является необходимой для реализации найденных решений, координатор, который в состоянии найти глобально оптимальное управляющее воздействие, вполне может в обход всей иерархической структуры применить управляющее воздействие непосредственно к процессу P . Это верно также и в том случае, когда в качестве выходной функции используется P_0 ; хотя в этом случае варьируемыми переменными («переменными решения») являются только сами координирующие сигналы, сложность функции P_0 , задаваемой равенством типа (5.40), делает оптимизационную задачу вышестоящего уровня даже более трудной, чем заданная глобальная оптимизационная задача.

Использование кажущейся глобальной целевой функции при формулировании задачи вышестоящего уровня

Некоторые из перечисленных выше трудностей можно частично устранить, если использовать для оптимизационной задачи вышестоящего уровня такую функцию качества, которая явным образом зависит от координирующих сигналов. Кажущаяся глобальная целевая функция $g_{\mathcal{A}}$ двухуровневой системы является именно такой функцией по следующим причинам:

1) $g_{\mathcal{A}}$ определена через все «переменные решения» в системе; значение $g_{\mathcal{A}}$ зависит от выбранных значений координирующего цели сигнала β , управляющего воздействия m и связующего входа u ; следовательно, выбор β , m и u может быть сделан применительно к функции $g_{\mathcal{A}}$.

2) $g_{\mathcal{A}}$ тесно связана по смыслу с глобальной функцией качества: по определению $g_{\mathcal{A}}(\beta, m, u) = G(m, P(m))$ всякий раз, когда $u = K(m)$.

3) Если система обладает свойством монотонности, то при подходящих условиях оптимальное значение глобальной функции качества может быть получено из $g_{\mathcal{A}}$ применением последовательности «операторов экстремизации»: если для некоторого координирующего цели сигнала β^0 и управляющего воздействия m^0 справедливо соотношение

$$g_{\mathcal{A}}(\beta^0, m^0, K(m^0)) = \max_{\mathcal{A}} \min_m \min_u g(\beta, m, u), \quad (5.41)$$

то m^0 будет глобально оптимальным управляющим воздействием.

Пусть $g_{\mathcal{A}}$ — кажущаяся глобальная целевая функция двухуровневой системы, обладающей свойством монотонности. Если равенство (5.41) выполняется для некоторого координирующего цели сигнала β^0 и управляющего воздействия m^0 , то мы имеем два равенства:

$$\max_{\mathcal{A}} g_{\mathcal{A}}(\beta, m(\beta), u(\beta)) = G(\hat{m}, P(\hat{m})), \quad (5.42)$$

если $m(\beta)$ и $u(\beta)$ существуют по крайней мере для $\beta = \beta^0$, и

$$\max_{\alpha} \min_{\mathcal{A}} g_{\mathcal{A}}(\beta, m(\alpha, \beta), \alpha) = G(\hat{m}, P(\hat{m})), \quad (5.43)$$

если $m(\alpha, \beta)$ существует по крайней мере для $\beta = \beta^0$ и $\alpha = K(m^0)$. Процесс решения глобальной задачи можно представить в таком случае в виде экстремизации кажущейся глобальной целевой функции и соответственно возложить решение этой задачи на оба уровня. Равенство (5.42) показывает, что если локаль-

ные решающие элементы минимизируют свои функции качества как по локальным управлениям, так и по связующим входам. координатор должен выбрать β так, чтобы максимизировать $g_{\mathcal{R}}$. Аналогично равенство (5.43) показывает, что если локальные решающие элементы минимизируют свои функции качества только по локальным управляющим воздействиям, то координатор должен выбрать прогнозы α для минимизации, а β — соответственно для максимизации $g_{\mathcal{R}}$.

Сделаем несколько важных замечаний относительно формулировки оптимизационной задачи вышестоящего уровня на основе кажущейся глобальной целевой функции $g_{\mathcal{R}}$ двухуровневой системы, обладающей свойством монотонности.

1. Как легко заметить, локальные решающие элементы могут с полным основанием считать, что $g_{\mathcal{R}}$ дает фактические глобальные затраты: это как раз тот случай, когда выбранные связующие входы согласованы. Так как система обладает свойством монотонности, каждый из локальных решающих элементов минимизирует $g_{\mathcal{R}}$, не выходя из своей области допустимых решений. Координатор же не имеет дела с минимизацией $g_{\mathcal{R}}$ путем выбора соответствующих β ; напротив, он, как указывалось выше, максимизирует $g_{\mathcal{R}}$ на $\mathcal{R}$. В таком случае локальные решающие элементы будут считать, что координатор противодействует их собственным интересам, так что межуровневые связи оказываются такими, как при игре двух лиц: локальные решающие элементы действуют в коалиции против координатора. Поскольку локальные решающие элементы так же, как и координатор, заинтересованы в минимизации фактических глобальных затрат, кажущаяся конкуренция между ними парадоксальна. Этот парадокс можно объяснить тем, что локальные решающие элементы рассматривают связующие входы как свободные управляющие воздействия и выбирают их соответственно своим интересам. Это приводит к тому, что кажущиеся глобальные затраты становятся меньше фактически получаемых затрат (или равными им), когда удовлетворяются ограничения, накладываемые на связующие функции подпроцессов. Тогда координатор должен компенсировать эти действия локальных решающих элементов, ибо тот факт, что кажущиеся глобальные затраты оказываются ниже получаемых в действительности, свидетельствует о том, что в общем случае после удовлетворения условий «связующих ограничений» (возникающих при объединении подпроцессов) фактические глобальные затраты не будут оптимальными. Интересно отметить, что действия координатора кажутся противоположными интересам группы, однако при окончательном анализе они оказываются выгодными всем участникам. Максимизируя $G_{\mathcal{R}}$ на $\mathcal{R}$, координа-

тор признает реальность существующих в процессе взаимодействия и обеспечивает их компенсацию.

2. При подходящих условиях [строго монотонные межуровневые функции и (5.21)] равенство (5.41) справедливо тогда и только тогда, когда кажущаяся глобальная целевая функция $g_{\mathcal{B}}$ имеет седловую точку: в этом случае существует такая тройка значений $(\hat{\beta}, \hat{m}, \hat{u})$, что неравенства

$$g_{\mathcal{B}}(\beta, \hat{m}, \hat{u}) \leq g_{\mathcal{B}}(\hat{\beta}, \hat{m}, \hat{u}) \leq g_{\mathcal{B}}(\hat{\beta}, m, u)$$

удовлетворяются на $\mathcal{B} \times M \times U$. Локальные задачи — минимизация локальных функций качества (и, следовательно, g) как по управляющим, так и по связующим входам, — и стоящая перед координатором задача оптимизации $g_{\mathcal{B}}$ за счет выбора координирующих сигналов β могут считаться «двойственными» задачами. А именно, если дана седловая точка $(\hat{\beta}, \hat{m}, \hat{u})$ функции $g_{\mathcal{B}}$, то минимизация функции $g_{\mathcal{B}}(\hat{\beta}, m, u)$ на $M \times U$ и максимизация функции $g_{\mathcal{B}}(\beta, \hat{m}, \hat{u})$ на $\mathcal{B}$ дают то же самое значение для $g_{\mathcal{B}}$. Кроме того, необходимо отметить, что благодаря соотношению $g(m) = G(m, P(m))$ минимизация $g_{\mathcal{B}}(\hat{\beta}, m, u)$ на множестве

$$\{(m, K(m)) \mid m \in M\}$$

является как раз глобальной оптимизационной задачей.

3. Может показаться, что решающие элементы на каждом уровне имеют свои автономные задачи принятия решений. Хотя это и верно, следует учитывать, что все эти задачи взаимосвязаны. Для того чтобы координатор решил свою задачу максимизации $g_{\mathcal{B}}$ на $\mathcal{B}$, должны быть известны локальные оптимальные решения $m(\beta)$ и $u(\beta)$, но локальные решения не могут быть получены без координирующего воздействия. Чтобы найти оптимальный координирующий сигнал путем максимизации $g_{\mathcal{B}}$, координатор должен знать преобразование, которое переводит координирующее воздействие β в локально оптимальные решения $m(\beta)$ и $u(\beta)$. Такое преобразование может оказаться очень сложным и, следовательно, мы сталкиваемся здесь с теми же трудностями, как и в прямом подходе, использующем глобальную функцию качества. Возникающая на уровне координатора «двойственная» задача может быть решена итерационным процессом, включающим применение таких методов, как метод «наискорейшего спуска» (или «подъема»). Даже если объем вычислений, затрачиваемых на каждом шаге итерации, и невелик, требуемый общий объем вычислений может оказаться значительным. Поэтому такой косвенный подход, использующий кажущиеся глобальные целевые функции

(так называемая «двойственная» задача), имеет, по-видимому, более теоретическую, нежели практическую ценность.

4. Наконец, необходимо сознавать различие между итерационными методами, основанными на принципах координации, и упомянутым выше методом итерации. Последний опирается на изменения величины $g_{\mathcal{B}}$, которые вызываются изменением координирующих сигналов, тогда как первые определяются через условия координируемости вне какой-либо связи с функцией $g_{\mathcal{B}}$. Между обоими методами существует, однако, связь в том смысле, что оба они позволяют достигнуть одного и того же результата: при надлежащих условиях оба сходятся к оптимальному координирующему воздействию (или завершаются им). Поэтому можно считать, что координатор как бы решает «двойственную» задачу, выступающую в данном случае как задача, подлежащая решению на его уровне. Однако, как мы уже упоминали в этом разделе, такая интерпретация может ввести в заблуждение, если в действительности стратегия координатора основана на принципе координации (согласовании взаимодействий), а не на методах отыскания максимума (скажем, методе «наискорейшего подъема») $g_{\mathcal{B}}$. Следует указать, что свойства функции $g_{\mathcal{B}}$ очень сильно сказываются на сходимости итерационных процессов обоих типов. Поэтому сходимость этих итерационных процессов можно изучать путем исследования свойств функций $g_{\mathcal{B}}$ и K ; это и будет сделано в гл. 6 для более детально структурированных систем.

ОПТИМАЛЬНАЯ КООРДИНАЦИЯ
ДИНАМИЧЕСКИХ СИСТЕМ

В этой главе мы будем заниматься вопросами координации двухуровневых систем, описанных в гл. 5, при дополнительном предположении, что свойства таких систем, составляющих их подсистем и их взаимосвязи могут быть описаны на языке соответствующим образом введенных нормированных линейных пространств. Мы сосредоточим наше внимание на использовании принципов координации для получения «скоординированного состояния» двухуровневой системы. При этом мы будем опираться на такие свойства, как выпуклость, компактность, полунепрерывность и т. п., и, следовательно, наши выводы будут более конкретны, чем полученные в предыдущей главе. Разумеется, полученные ранее результаты для формальных абстрактных систем применимы и к системам, которые мы рассматриваем здесь. В этой главе мы не только исследуем существование «оптимальных» координирующих воздействий, но даем также точные их выражения, используя для этого линейные аппроксимации операторов оценки косвенных эффектов управляющих воздействий, введенных в гл. 5 (разд. 5). Наконец, рассматриваются некоторые итеративные методы для решения задачи координатора.

1. ВВЕДЕНИЕ

Класс систем, рассматриваемых в этой главе, включает среди прочих подкласс динамических систем со многими переменными, описываемых обыкновенными векторными дифференциальными уравнениями. Чтобы познакомить читателя с применением принципов координации к задаче «оптимальной» координации динамических систем, мы приведем пример из области линейных динамических систем с квадратичными функциями качества.

Рассмотрим двухуровневую систему, для которой управляемый процесс в целом P описывается на интервале времени $[0, 1]$ векторным дифференциальным уравнением

$$\frac{d}{dt} y = Ay + m, \quad y(0) = 0, \quad (6.1)$$

где A — постоянная матрица $[a_{ij}]$ размерности 2×2 , а m — пара действительных функций, интегрируемых с квадратом на интервале $[0, 1]$. Строго говоря, P есть отображение $P: M \rightarrow Y$, где $M = Y = \mathcal{L}_2[0, 1] \times \mathcal{L}_2[0, 1]$, определяемое уравнением (6.1) для всех m из M .

Предположим, что P состоит из подпроцессов P_i , $i = 1, 2$, где P_i описываются на интервале $[0, 1]$ уравнениями

$$\frac{d}{dt} y_i = a_{ii} y_i + m_i + u_i, \quad y_i(0) = 0 \quad (6.2)$$

и где связями между подпроцессами являются

$$u_1 = a_{12} y_2 \equiv H_1(m, y), \quad u_2 = a_{21} y_1 \equiv H_2(m, y).$$

Следовательно, i -й подпроцесс P_i представляет собой отображение $P_i: M_i \times U_i \rightarrow Y_i$, где $M_i = U_i = Y_i = \mathcal{L}_2[0, 1]$, определяемое уравнением (6.2) для всех (m_i, u_i) , принадлежащих $M_i \times U_i$.

Пусть глобальная функция (функционал) качества задается на $M \times Y$ интегралом

$$G(m, y) = \int_0^1 [m(t)^T m(t) + (y(t) - r)^T Q(y(t) - r)] dt,$$

где Q — симметрическая, положительно полуопределенная матрица $[q_{ij}]$ размерностью 2×2 , а $r = (r_1, r_2)$ — заданный вектор. Глобальная задача оптимизации состоит в минимизации $G(m, P(m))$ на пространстве M .

Прежде чем обратиться к решению задачи координации путем применения соответствующих принципов координации, отметим, что в данном случае глобальная задача оптимизации имеет единственное оптимальное решение. Непосредственное применение принципа максимума дает глобально оптимальное управляющее воздействие (управление) $\hat{m}$

$$\hat{m} = -\frac{1}{2} \lambda,$$

где λ удовлетворяет системе векторных дифференциальных уравнений

$$\begin{aligned} \frac{d}{dt} \lambda &= -A^T \lambda + 2Q(r - y), \quad \lambda(1) = 0, \\ \frac{d}{dt} y &= Ay - \frac{1}{2} \lambda, \quad y(0) = 0, \end{aligned} \quad (6.3)$$

на интервале $[0, 1]$.

Прежде всего мы должны сформулировать локальные (оптимизационные) задачи так, чтобы система могла быть скоординирована путем применения принципа согласования взаимодействий. Для простоты предположим, что Q — диагональная матрица, у которой $q_{11} > a_{21}^2$ и $q_{22} > a_{12}^2$. Для каждой пары $\beta = (\beta_1, \beta_2)$ функций из $\mathcal{L}_2$ определим локальные функции (функционалы)

качества $G_{1\beta}$ и $G_{2\beta}$ следующим образом:

$$G_{1\beta}(m_1, u_1, y_1) = \int_0^1 [m_1^2 + (u_1 - a_{12}r_2)^2 + q'_{11}(y_1 - r_1)^2 + \\ + \beta_1 u_1 - \beta_2 a_{21}y_1] dt, \\ G_{2\beta}(m_2, u_2, y_2) = \int_0^1 [m_2^2 + (u_2 - a_{21}r_1)^2 + q'_{22}(y_2 - r_2)^2 + \\ + \beta_2 u_2 - \beta_1 a_{12}y_2] dt,$$

где $q'_{11} = q_{11} - a_{21}^2$ и $q'_{22} = q_{22} - a_{12}^2$. С помощью простой подстановки легко убедиться, что G обладает свойством аддитивности по отношению ко всем координирующим цели сигналам β :

$$G(m, y) = G_{1\beta}(m_1, a_{12}y_2, y_1) + G_{2\beta}(m_2, a_{21}y_1, y_2). \quad (6.4)$$

При заданном β i -я локальная задача состоит в минимизации $G_{i\beta}(m_i, u_i, P_i(m_i, u_i))$ на пространстве $M_i \times U_i$.

Далее, для каждого координирующего цели сигнала $\beta_1 = (\beta_1, \beta_2)$ обе локальные задачи имеют единственные оптимальные решения. Действительно, для первой локальной задачи мы в соответствии с принципом максимума получаем оптимальное решение в виде пары (m_1^β, u_1^β)

$$m_1^\beta = -\frac{1}{2}\rho_1, \quad u_1^\beta = a_{12}r_2 - \frac{1}{2}\beta_1 - \frac{1}{2}\rho_1, \quad (6.5)$$

где ρ_1 удовлетворяет системе дифференциальных уравнений

$$\frac{d}{dt}\rho_1 = -a_{11}\rho_1 + 2q'_{11}(r_1 - y_1) + \beta_2 a_{21}, \quad \rho_1(1) = 0, \quad (6.6)$$

$$\frac{d}{dt}y_1 = a_{11}y_1 - \rho_1 + a_{12}r_2 - \frac{1}{2}\beta_1, \quad y_1(0) = 0$$

на интервале $[0, 1]$. Оптимальное решение второй локальной задачи получается из выражений (6.5) и (6.6) перестановкой индексов 1 и 2.

Задача координатора заключается в том, чтобы найти такой координирующий цели сигнал β , чтобы управляющее воздействие $m^\beta = (m_1^\beta, m_2^\beta)$ было глобально оптимальным, т. е. обеспечивало оптимальное решение глобальной задачи. Принцип согласования взаимодействий применим к этой системе, так как система, согласно (6.4), обладает свойством монотонности. Применимость может быть показана также и непосредственно путем решения уравнений (6.6) относительно ρ_1 и ρ_2 при условии

$$u_1^\beta = H_1(m^\beta, P(m^\beta)), \quad u_2^\beta = H_2(m^\beta, P(m^\beta)). \quad (6.7)$$

Следовательно, координатор, используя указанный принцип, отыскивает такое β , для которого удовлетворяются условия (6.7).

Такой координирующий сигнал $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)$ существует, а именно:

$$\hat{\beta}_1 = -2\hat{m}_1 + 2a_{12}(r_2 - \hat{y}_2), \quad \hat{\beta}_2 = -2\hat{m}_2 + 2a_{21}(r_1 - \hat{y}_1), \quad (6.8)$$

где $\hat{y} = P(\hat{m})$ и $\hat{m}$ — глобально оптимальное управляющее воздействие. Можно непосредственно проверить, что при $\beta = \hat{\beta}$ пара (m^β, u^β) удовлетворяет уравнению (6.7).

Сформулируем теперь локальные задачи так, чтобы систему можно было скоординировать с помощью принципа прогнозирования взаимодействий. Относительно матрицы Q не будем делать никаких предположений, кроме того, что она диагональна и по крайней мере положительно полуопределенная. Для каждой пары $\beta = (\beta_1, \beta_2)$ функций из $\mathcal{L}_2[0, 1]$ определим локальные функции качества $G_{1\beta}$ и $G_{2\beta}$ в следующем виде:

$$G_{i\beta}(m_i, y_i) = \int_0^1 [m_i^2 + q_{ii}(y_i - r_i)^2 + \beta_i y_i] dt,$$

где $i = 1, 2$. Заметим, что если оптимизация проводится на пространствах $M_i \times U_i$, то задачи локального управления не имеют решения. Система в этом случае не может быть скоординирована с помощью принципа согласования взаимодействий, несмотря на то что принцип применим. Однако координация все же может быть достигнута, если, помимо подачи координирующего сигнала β , координатор сумеет спрогнозировать значения связующих входов и добиться того, чтобы они были приняты в качестве точных значений.

Пусть множеством возможных координирующих сигналов является пространство $\mathcal{A} \times \mathcal{B}$, где $\mathcal{A}$ и $\mathcal{B}$ в свою очередь представляют собой пространство $\mathcal{L}_2[0, 1] \times \mathcal{L}_2[0, 1]$. Элементы $\alpha \in \mathcal{A}$ рассматриваются как прогнозные значения связующих входов, а элементы β из $\mathcal{B}$ определяют локальные функции качества $G_{1\beta}$ и $G_{2\beta}$. В этом случае i -я локальная задача конкретизируется с помощью сигнала $\gamma = (\alpha, \beta)$ и заключается в минимизации $G_{i\beta}(m_i, P_i(m_i, \alpha_i))$ на пространстве M_i . Оптимальным решением является m_i^γ :

$$m_i^\gamma = -\frac{1}{2} \rho_i,$$

где ρ_i удовлетворяет системе дифференциальных уравнений

$$\frac{d}{dt} \rho_i = -a_{ii} \rho_i + 2q_{ii}(r_i - y_i) - \beta_i, \quad \rho_i(1) = 0, \quad (6.9)$$

$$\frac{d}{dt} y_i = a_{ii} y_i - \frac{1}{2} \rho_i + \alpha_i, \quad y_i(0) = 0$$

на интервале $[0, 1]$.

Принцип прогнозирования взаимодействий может быть применен в рассматриваемом примере при использовании отображения $\eta: M \rightarrow \mathcal{R}$, определяемого уравнениями

$$\beta_1 = -2a_{12}m_2 \equiv \eta_1(m_1, m_2), \quad \beta_2 = -2a_{21}m_1 \equiv \eta_2(m_1, m_2).$$

Если $m^\gamma = (m_1^\gamma, m_2^\gamma)$ для заданного $\gamma = (\alpha^\gamma, \beta^\gamma)$, то можно непосредственно проверить, что m^γ является глобально оптимальным управляющим воздействием тогда, когда

$$\alpha_i^\gamma = H_i(m^\gamma, P(m^\gamma)) \quad (6.10)$$

и

$$\beta_i^\gamma = \eta_i(m^\gamma) \quad (6.11)$$

для $i = 1, 2$. Координатору достаточно найти такой координирующий сигнал γ , чтобы удовлетворялись равенства (6.10) и (6.11). Таким координирующим сигналом является $\gamma(\hat{\alpha}, \hat{\beta})$, где

$$\hat{\alpha} = H(\hat{m}, P(\hat{m})), \quad \hat{\beta} = \eta(\hat{m}), \quad (6.12)$$

а $\hat{m}$ — глобально оптимальный управляющий сигнал. Тот факт, что (6.12) дает искомый координирующий сигнал, может быть проверен прямой подстановкой $(\hat{\alpha}, \hat{\beta})$ в (6.9); выполнив такую подстановку, мы убеждаемся, что условия (6.10) и (6.11) удовлетворяются.

2. ПРЕДВАРИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

При анализе координируемости двухуровневой системы мы сосредоточим внимание на принципах координации в применении к двухуровневым системам, описываемым с помощью линейных нормированных пространств или, более конкретно, множествами функций времени. Сперва мы кратко опишем представления нормированного линейного (векторного) пространства как пространства функций времени, а затем представим совокупный процесс и его подпроцессы в форме, подходящей для анализа в линейных нормированных пространствах. Затем будут изложены некоторые результаты, относящиеся к задаче координации.

Модели процесса

В этой главе основным предположением является следующее: системы описываются с помощью нормированных линейных пространств, т. е. процесс в целом и его подпроцессы определены на подмножествах или на нормированных линейных пространствах и обладают, когда необходимо, такими свойствами, как полнота, компактность, выпуклость и т. д. Кроме того, в последующих стандартных процедурах функционального анализа множества

выходов и управляющих воздействий, относящихся к рассматриваемым моделям подпроцессов, определяются посредством операторов проектирования: ограниченных, линейных, идемпотентных преобразований пространства самого в себя. Следовательно, появляющиеся в нашем рассмотрении локальные множества, относящиеся к «локальным» объектам, будут подмножествами «общих множеств», характеризующих совокупность этих локальных объектов в целом.

Чтобы читатель мог следить за дальнейшим изложением, полезно будет напомнить некоторые формы представления нормированного линейного пространства в виде обычных функциональных пространств [27].

Пусть $\mathcal{T}$ — интервал $[a, b]$ действительной прямой R . Пространство $\mathcal{E}[\mathcal{T}]$ всех непрерывных функций $f: \mathcal{T} \rightarrow R$ с однородной нормой

$$\|f\| = \sup \{|f(t)|, t \in \mathcal{T}\}$$

является нормированным линейным пространством. В самом деле, $\mathcal{E}[\mathcal{T}]$ является полным в смысле этой нормы и, следовательно, представляет собой банахово пространство. Аналогично для любого целого $n \geq 1$ пространство $\mathcal{E}[\mathcal{T}]^n$ всех n -мерных непрерывных функционалов на $\mathcal{T}$ с нормой

$$\|f\| = \sup \{|f_i(t)|, t \in \mathcal{T} \quad 1 \leq i \leq n\}$$

является банаховым пространством.

Пусть N — множество целых чисел. Пространство R^N всех функций $f: N \rightarrow R$ с нормой

$$\|f\| = \sup \{|f(n)|, n \in N\}$$

является банаховым пространством, и таковым же является пространство всех функций $f: N \rightarrow R^n$.

Процесс в целом

Мы предполагаем, что процесс в целом P задается оператором с областью определения в виде заданного нормированного линейного пространства $\bar{M}$ и областью значений в виде подмножества заданного нормированного линейного пространства Y . Следовательно, процесс в целом является отображением $P: \bar{M} \rightarrow Y$ (как в гл. 5).

Это определение процесса в целом охватывает системы, описанные в нижеследующих примерах.

Пример 6.1. Динамические системы

Пусть X — топологическое пространство, R — действительная прямая, а B — множество функций на R , замкнутое относительно

по сдвигов во времени: для всех $f \in B$ и $t \in R$ функция f_t , определенная на R уравнением $f_t(\tau) = f(\tau + t)$, принадлежит также B . Определим подверженную внешним воздействиям динамическую систему на $X \times B$ отображением $s: R \times X \times B \rightarrow X$, удовлетворяющим следующим аксиомам:

- (1) $s(0, x, f) = x$ для всех $x \in X$ и $f \in B$;
- (2) s непрерывно в $R \times X$ для всех $f \in B$;
- (3) $s(t + t', x, f) = s(t', s(t, x, f), f_t)$ для всех $t, t' \in R, x \in X$ и $f \in B$.

Тогда для любой постоянной функции f в B система $s_f: R \times X \rightarrow X$, заданная уравнением $s_f(t, x) = s(t, x, f)$, является, согласно Зубову [28], динамической системой на X , т. е.

- (а) $s_f(0, x) = x$, для всех x из X ;
- (б) s_f непрерывно на $R \times X$;
- (в) $s_f(t + t', x) = s_f(t', s_f(t, x))$ для всех $t' \in R$ и $x \in X$.

Приведенное выше описание процесса в целом P охватывает по меньшей мере класс динамических систем (подверженных внешним воздействиям) на $X \times B$, где X и B являются нормированными линейными пространствами. Чтобы показать это, предположим, что X и B — нормированные линейные пространства, а Y — пространство всех непрерывных функций, переводящих R в X с однородной нормой. Определим $P: X \rightarrow Y$ следующим образом:

$$P(x, f)(t) = S(t, x, f) \quad \text{для всех } t \in R.$$

Тогда P будет отображением из нормированного линейного пространства $\bar{M} = X \times B$ в Y . Более того, если s непрерывна на $R \times X \times B$, то P непрерывна в смысле однородной нормы на Y . Интересующим нас множеством $M \subseteq \bar{M}$ может быть множество $\{x\} \times B$; в этом случае траектории выходов y были бы фиксированы при $t = 0$ и изменялись бы в другие моменты времени под действием различных «управлений» из B . С другой стороны, в качестве M можно взять множество $X \times \{f\}$; в этом случае «управление» было бы фиксировано, а траектории выходов y варьировались бы за счет изменения точек, через которые они проходят в момент времени $t = 0$.

Наше определение P охватывает более широкий класс систем, чем класс подверженных внешним воздействиям динамических систем в нормированных линейных пространствах. Действительно, полугрупповое свойство [аксиома (3)] не требуется. Процесс P может быть любым отображением из подмножества нормированного линейного пространства (функций времени) в нормированное линейное пространство (функций времени) без наложения каких-либо ограничений на их изменение во времени.

Данное определение P охватывает системы, заданные обыкновенными дифференциальными уравнениями и дифференциальными

уравнениями в частных производных. Действительно, в работе [28] показано, что понятие динамической системы включает системы, описываемые обыкновенными дифференциальными уравнениями, а понятие динамической системы в функциональном пространстве охватывает системы, описываемые дифференциальными уравнениями в частных производных вида

$$\frac{\partial y}{\partial t} = ay + \sum_{i=1}^n b_i \frac{\partial y}{\partial x_i},$$

где a и b_i при $i = 1, \dots, n$ — заданные константы. Следовательно, наш формализм справедлив и для этих систем.

Помимо рассмотрения процесса в целом P просто как отображения из одного нормированного линейного пространства в другое, мы в следующем примере рассматриваем конкретный случай отображения P , описываемого системой дифференциальных уравнений.

Пример 6.2. Системы, описываемые дифференциальными уравнениями

Пусть $\mathcal{T}$ — интервал $[0, 1]$ на действительной прямой R , а $\bar{M}$ — пространство всех интегрируемых с квадратом функций $m: \mathcal{T} \rightarrow \mathcal{E}_p$, где $\mathcal{E}_p$ — евклидово p -пространство, а Y — пространство всех интегрируемых с квадратом функций $y: \mathcal{T} \rightarrow \mathcal{E}_q$. Предположим, что выход $y = P(m)$ для любого $m \in M$ описывается векторным дифференциальным уравнением

$$\dot{y}(t) = f(t, y(t), m(t)), \quad y(0) = 0,$$

определенным на $\mathcal{T}$. При подходящих условиях, наложенных на f , приведенные выше дифференциальные уравнения определяют отображение $P: \bar{M} \rightarrow Y$. Мы рассматриваем, в частности, случай, в котором f — линейная форма

$$f(t, y(t), m(t)) = Ay(t) + Bm(t),$$

где A и B — матрицы $q \times q$ и $q \times p$, которые могут меняться во времени, но ограничены.

Мы не требуем, чтобы процесс P определялся только системой дифференциальных уравнений; помимо этого, может иметь место и совокупность преобразований, дающих выходы в виде функции решений системы дифференциальных уравнений. Тогда процесс P , заданный как в приведенном выше примере, описывает лишь частичное преобразование состояний системы. Однако более удобные обозначения и другие связанные с этим упрощения оправдывают использование этой более простой формы P ; кроме того, для того типа задач, которым мы будем заниматься, всякое преобразование выходов P будет рассматриваться нами в связи с использованием глобальной функции качества.

Подпроцессы

Предполагается, что множество управляющих воздействий M на процесс в целом является подмножеством линейного нормированного пространства $\bar{M}$. Множества же локальных управляющих воздействий (выходов) считаются такими подмножествами пространств $\bar{M}_i$, $1 \leq i \leq n$, что каждый элемент m в M имеет единственное разложение

$$m = m_1 + \dots + m_n, \quad (6.13)$$

где $m_i \in \bar{M}_i$. Другими словами, предполагается, что $\bar{M}$ — прямая сумма

$$\bar{M} = \bar{M}_1 \oplus \dots \oplus \bar{M}_n. \quad (6.14)$$

Множество локальных управляющих воздействий $\bar{M}_i$, $1 \leq i \leq n$, является тогда подмножеством $\bar{M}_i$, заданным как проекция M на $\bar{M}_i$. В общем случае $M \subseteq M_1 \oplus \dots \oplus M_n$, однако мы далее всюду предполагаем, что

$$M = M_1 \oplus \dots \oplus M_n. \quad (6.15)$$

Случай, когда M — собственное подмножество, рассматривается в гл. 7.

Такая структура множеств локальных управляющих воздействий позволяет нам считать все используемые локальные управляющие воздействия элементами одного и того же множества управляющих воздействий для всего процесса в целом. Это значительно упрощает обозначения. Для удобства мы используем проекционные операторы $\Pi_{iM}: \bar{M} \rightarrow \bar{M}_i$, которые однозначно определяются разложением (6.13) элементов из $\bar{M}$; для каждого $m \in M$ $\Pi_{iM} m = m_i$ представляет проекцию m на $\bar{M}_i$.

Пространство $\bar{M}$ может быть n -мерным декартовым произведением, и, следовательно, множество управляющих воздействий на весь процесс в целом может задаваться в виде $M = M_1 \times$

$\times M_n$ (как в гл. 5), где M_i — соответствующие множества локальных управляющих воздействий. В этом случае декомпозиция m из $\bar{M}$ находится с помощью соотношения (6.13), где для каждого i , $1 \leq i \leq n$,

$$m_i = (0, \dots, 0, m_i, 0, \dots, 0)$$

(m_i — i -я компонента m).

Аналогично множества локальных выходов $\mathcal{Y}_i$, $1 \leq i \leq n$, предполагаются такими, что Y является прямой суммой

$$Y = \mathcal{Y}_1 \oplus \dots \oplus \mathcal{Y}_n.$$

Для каждого i , $1 \leq i \leq n$, Π_{iY} — проекционный оператор, определенный на Y , так что $\Pi_{iY} y = y_i$ — проекция y на $\mathcal{Y}_i$.

Теперь мы можем перейти к декомпозиции процесса в целом. Прежде всего введем множество связующих входов $\bar{U} = \bar{U}_1 \times \dots \times \bar{U}_n$, где множества локальных связующих входов $\bar{U}_i$, $1 \leq i \leq n$, представляют собой нормированные линейные пространства. Следующий шаг заключается в декомпозиции P на n подпроцессов $\mathcal{P}_i$

$$\mathcal{P}_i: \bar{\mathcal{M}}_i \times \bar{U}_i \rightarrow \mathcal{Y}_i$$

во введении n связующих функций H_i

$$H_i: \bar{M} \times Y \rightarrow U_i,$$

таких, что

$$\sum_{i=1}^n \mathcal{P}_i(m_i, H_i(m, P(m))) = P(m)$$

для всех $m \in \bar{M}$.

Покажем теперь на примерах, как упомянутая декомпозиция P может быть реализована.

Пример 6.3

Если использовать оператор $(I - \Pi_{iM})$ и связующую функцию $H_i = (I - \Pi_{iM})$, где I — оператор тождественного преобразования, то для каждого i , $1 \leq i \leq n$, множество связующих входов $\bar{U}_i$ будет образом $\bar{M}$. Связующий вход u_i для i -го подпроцесса в этом случае является суммой нелокальных управляющих воздействий:

$$u_i = m_i + m_{i-1} + m_{i+1} + \dots + m_n,$$

где m — управляющее воздействие на весь процесс в целом. Тогда i -й подпроцесс $\mathcal{P}_i$ определяется на $\bar{\mathcal{M}} \times \bar{U}_i$ уравнением

$$\mathcal{P}_i(m_i, u_i) = \Pi_{iY} P(m_i + u_i). \quad (6.16)$$

Такая декомпозиция P называется канонической относительно входов и всегда возможна.

Пример 6.4

Если в качестве связующей функции H_i использовать оператор $(I - \Pi_{iY})$, то для каждого i , $1 \leq i \leq n$, множество связующих входов H_i будет образом Y . Связующий вход u_i для i -го

подпроцесса состоит в этом случае из нелокальных выходов $u_i = y_1 + y_{i-1} + y_{i+1} + \dots + y_n$. Тогда i -й подпроцесс $\mathcal{P}_i$ определяется на $\mathcal{M}_i \times U_i$ так, чтобы для всех m в M удовлетворялось уравнение

$$\mathcal{P}_i(m_i, (I - \Pi_{iY})P(m)) = \Pi_{iY}P(m).$$

Такая декомпозиция называется *канонической относительно выходов*. Декомпозиция такого рода возникает естественным образом в случае, когда в качестве $y = P(m)$ берется решение дифференциального уравнения $\dot{y} = f(y, m)$, проходящее через заданную точку, как в примере 6.2.

Функция взаимодействия подпроцессов K , как и прежде, является отображением $K: \bar{M} \rightarrow \bar{U}$, определяемым следующим выражением:

$$K(m) = H(m, P(m)),$$

где

$$H(m, y) = (H_1(m, y), \dots, H_n(m, y)).$$

Тогда отображение $\bar{P}: \bar{M} \times \bar{U} \rightarrow Y$ где

$$\bar{P}(m, u) = \sum_{i=1}^n \mathcal{P}_i(m_i, u_i), \quad (6.17)$$

представляет «развязанные» (невзаимодействующие) подпроцессы, поскольку из уравнения (6.16) следует, что

$$P(m) = \sum_{i=1}^n \mathcal{P}_i(m_i, K_i(m)), \quad (6.18)$$

где $K_i: \bar{M} \rightarrow \bar{U}_i$ — i -я компонента отображения K .

Как было показано в гл. 4 и 5, реализующиеся связующие входы подпроцессов возникают в результате взаимодействия подпроцессов и зависят от выбранных управляющих воздействий.

Функции качества

К введенным ранее допущениям, характеризующим рассматриваемые в этой главе двухуровневые системы, надо теперь лишь добавить предположение, касающееся функций качества: все используемые нами функции качества (как глобальная, так и локальные) предполагаются вещественными функциями, определенными на соответствующих множествах. Глобальная функция качества определяется на пространстве $\bar{M} \times Y$, в то время как каждая локальная функция качества определяется на пространстве $\bar{M}_i \times \mathcal{Y}_i \times \bar{U}_i$, за исключением особо оговоренных случаев.

Аддитивное представление глобальных функций качества

Мы будем говорить, что глобальная функция качества $\tilde{G}$ согласована с другой функцией качества G рассматриваемой двухуровневой системы, если глобальные управляющие воздействия, оптимальные с точки зрения $\tilde{G}$, являются в то же время глобально оптимальными по отношению к G . Конкретные значения, которые принимают функции G и $\tilde{G}$, могут быть различными для одного и того же управления процессом в целом — требуется лишь, чтобы выбранное управляющее воздействие было глобально оптимальным относительно G всякий раз, когда оно является глобально оптимальным по отношению к $\tilde{G}$.

Для всякой заданной глобальной функции качества G всегда можно определить локальные функции качества G_i , $1 \leq i \leq n$, $G_i: M_i \times \mathcal{Y}_i \rightarrow R$ так, чтобы функция $\tilde{G}$

$$\tilde{G}(m, y) = \sum_{i=1}^n \frac{1}{n} G_i(m_i, y_i)$$

была согласована с G . Это очевидный и тривиальный факт. Если для G существует глобально оптимальное управляющее воздействие $\hat{m}$, то локальные функции качества G_i можно получить из G введением ограничений или разбиением пары вход — выход $(\hat{m}, \hat{y})$, где $\hat{y} = P(\hat{m})$, на отдельные составляющие. Если же G не имеет глобального оптимума, можно выбрать локальные функции качества G_i так, чтобы $\tilde{G}$ не имела глобального оптимума.

Учитывая это, построим аддитивную глобальную функцию качества $\tilde{G}$, которая была бы согласована (при определенных условиях) с заданной глобальной функцией качества. Предполагая, что семейство локальных функций качества G_i , $1 \leq i \leq n$, задано, постараемся найти действительные числа $a_i \geq 0$, $1 \leq i \leq n$, такие, чтобы глобальная функция качества

$$\tilde{G}(m, y) = \sum_{i=1}^n a_i G_i(m_i, y_i, H_i(m, y)) \quad (6.19)$$

была согласованной с заданной глобальной функцией качества G . Назовем такую функцию G аддитивным представлением G через заданные локальные функции качества G_i , $1 \leq i \leq n$.

Существование аддитивного представления важно по следующим причинам.

Предположим, что двухуровневая система имеет аддитивную глобальную функцию качества, а модификации локальных функций качества являются модификациями с нулевой суммой, т. е. для любого координирующего цели сигнала β , принадлежащего за-

данному множеству $\mathcal{B}$ и $m \in M$, значение глобальной функции качества является суммой реализовавшихся значений локальных функций качества

$$g(m) = \sum_{i=1}^n g_{i\beta}(m_i, K_i(m))$$

для всех $m \in \bar{M}$. Предположим, что локальные целевые функции $g_{i\beta}$ определены через подпроцессы $\mathcal{P}_i$ и заданные локальные функции качества $G_{i\beta}$. Предположим также, что система координируема с помощью выбранного принципа координации. Тогда для любых действительных чисел $a_i \geq 0$, $1 \leq i \leq n$, и глобальной функции качества $\tilde{G}$

$$\tilde{G}(m, P(m)) = \sum_{i=1}^n a_i g_{i\beta}(m_i, K_i(m))$$

система также координируема с помощью этого принципа. Более того, если некая глобальная функция качества G обладает при некотором $\beta \in B$ аддитивным представлением $\tilde{G}$

$$\tilde{G}(m, y) = \sum_{i=1}^n a_i G_{i\beta}(m_i, y_i, H_i(m, y)),$$

то система опять же координируема с помощью этого принципа.

Следовательно, теория координируемости, развитая для систем с аддитивной глобальной функцией качества и модификациями локальных функций качества с нулевой суммой, может быть расширена на все системы, глобальные функции качества которых допускают аддитивное представление через заданные локальные функции качества.

Сформулируем теперь некоторые условия, при которых аддитивное представление существует. Тем самым мы перейдем от более сильных условий к более слабым и представим результаты с помощью глобальной целевой функции g и локальных функций затрат h_i , $1 \leq i \leq n$, которые были введены ранее, определив их на $\bar{M}$ с помощью соотношений

$$g(m) = G(m, P(m)),$$

$$h_i(m) = G_i(m_i, \mathcal{P}_i(m_i, u_i), u_i),$$

где $u_i = H_i(m, P(m))$, G — заданная глобальная функция качества, а G_i , $1 \leq i \leq n$, — заданные локальные функции качества рассматриваемой двухуровневой системы.

Теорема 6.1

Предположим, что M — выпуклое множество и

$$g(m) = \psi(h_1(m), \dots, h_n(m)) \quad (6.20)$$

на $\bar{M}$, где $\psi: R^n \rightarrow R$ — строго монотонная функция. Более того, предположим, что g дифференцируема по Фреше ¹⁾ на $\bar{M}$ и имеет минимум на M и что для каждого i , $1 \leq i \leq n$, функции h_i дифференцируемы по Фреше на $\bar{M}$ и строго выпуклы на M . Тогда существуют такие действительные числа $a_i \geq 0$, $1 \leq i \leq n$, что заданная глобальная функция качества имеет аддитивное представление G в виде выражения (6.19).

Д о к а з а т е л ь с т в о. Пусть $\hat{m}$ — точка в M , минимизирующая функцию g на M . Так как g дифференцируема при $m = \hat{m}$, то

$$g(\hat{m} + \lambda m) - g(\hat{m}) = g'(\hat{m}) \lambda m + \mathcal{O}(\lambda) \geq 0$$

для любого действительного λ , $0 < \lambda < 1$, и таких m , что $\hat{m} +$

$\lambda m \in M$, где $g'(\hat{m})$ — производная Фреше от g при $m = \hat{m}$; следовательно, $g'(\hat{m}) \lambda m \geq 0$. Из выражения (6.20) имеем для любого m в $\bar{M}$

$$g'(m) = \sum_{i=1}^n \psi'_i(h_1(m), \dots, h_n(m)) h'_i(m),$$

где $\psi'_i = \partial \psi / \partial h_i$ в точке $(h_1(m), \dots, h_n(m))$, а $h'_i(m)$ — производная Фреше от h_i в точке m . Так как ψ — монотонная функция, то для каждого i , $1 \leq i \leq n$, функция $\psi'_i(m) \geq 0$. А так как ψ — строго монотонная функция, то $\alpha(m) = \sum_{i=1}^n \psi'_i(m) \neq 0$.

Из того, что $\alpha(m) > 0$ и $g'(\hat{m}) \lambda m' \geq 0$ для любого $\hat{m} + \lambda m'$ в M , получаем для каждого i , $1 \leq i \leq n$, что

$$\lim_{m \rightarrow \hat{m}} \frac{1}{\alpha(m)} [g'(m) \lambda m'] = \sum_{i=1}^n a_i [h'_i(\hat{m})] \lambda m' \geq 0,$$

где

$$a_i = \lim_{m \rightarrow \hat{m}} \frac{1}{\alpha(m)} [\psi'_i(m)]. \quad (6.21)$$

Заметим, что $a_i \geq 0$ для каждого i , $1 \leq i \leq n$, и $\sum_{i=1}^n a_i = 1$. Так как каждая функция h_i строго выпукла на M и каждый коэффициент a_i неотрицательный, то $\sum_{i=1}^n a_i h_i$ также строго выпукла на M . Поэтому $\hat{m}$ является глобальным минимумом на M , и, более того, он единствен. Следовательно, функция $\bar{G}$, заданная выражением (6.19), — аддитивное представление G . Этим доказательство завершается.

¹⁾ Введение производных Фреше необходимо ввиду того, что в общем случае мы имеем дело с функциями, аргументы которых являются элементами функциональных пространств.

При доказательстве теоремы было показано, как может быть построено аддитивное представление заданной глобальной функции качества: для каждого i , $1 \leq i \leq n$, вычисляются коэффициенты a_i , заданные пределом (6.21), и формируются новые локальные функции качества $\tilde{G}_i = a_i G_i$. В случае, если эти коэффициенты существуют и условия теоремы удовлетворяются, аддитивное представление $\tilde{G}$ функции G определяется выражением (6.19).

Следствие

Предположим, что рассматриваемая двухуровневая система такова, что все условия теоремы (6.1) удовлетворяются, за исключением, быть может, дифференцируемости и выпуклости функций h_i . Допустим, однако, что для произвольного i , $1 \leq i \leq n$, коэффициент a_i , найденный с помощью предела (6.21), положителен и h_i — строго выпукла на M . В этом случае функция, определяемая выражением (6.19), является аддитивным представлением рассматриваемой глобальной функции качества.

Доказательство. Пусть коэффициенты a_i для каждого i , $1 \leq i \leq n$, определяются с помощью предела (6.21). Так как функция $a_i h_i$ строго выпукла для произвольного i , $1 \leq i \leq n$, то функция $\sum_{i=1}^n a_i h_i$ также строго выпукла. Следовательно, если точка $\hat{m}$ в M минимизирует g на M , то это единственная точка в M , минимизирующая $\sum_{i=1}^n a_i h_i$ на M . Этим доказательство завершается.

Существование аддитивного представления заданной глобальной функции качества можно доказать для более общего случая, когда требования относительно дифференцируемости ослаблены.

Теорема 6.2

Предположим, что условие (6.20) выполняется, ψ — строго монотонная функция, а M — выпуклое множество. Предположим также, что g имеет минимум на M и для каждого i , $1 \leq i \leq n$, функция h_i строго выпукла и ограничена снизу на M . Тогда данная глобальная функция качества имеет аддитивное представление в виде выражения (6.19).

Доказательство. Пусть $\hat{m}$ — точка в M , минимизирующая функцию g на множестве M . Пусть $\bar{V}$ и V — подмножества пространства R^n , определенные следующим образом:

$$\bar{V} = \{(x_1, \dots, x_n): \text{для некоторого } m \in M, h_i(m) \leq x_i \\ \text{при } i = 1, \dots, n\},$$

$$V = \{(x_1, \dots, x_n): h_i(\hat{m}) > x_i \text{ при } i = 1, \dots, n\}.$$

Тогда, поскольку множество M выпукло и функционалы h_i выпуклы на M , оба множества $\bar{V}$ и $\underline{V}$ выпуклы. Более того, множества $\bar{V}$ и $\underline{V}$ непересекающиеся. Если допустить обратное, то найдется точка $x = (x_1, \dots, x_n)$, общая для множества $\bar{V}$ и множества $\underline{V}$ и, следовательно, должна существовать такая точка m в множестве M , что $h_i(m) \leq x_i < h_i(\hat{m})$ для каждого i , $1 \leq i \leq n$. Но тогда в силу строгой монотонности ψ мы имели бы $\psi(m) < \psi(\hat{m})$, что приводит к противоречию. Кроме того, так как ограничены снизу на M , то $\bar{V}$ и $\underline{V}$ — непустые множества.

Принимая во внимание теорему о разделяющей гиперплоскости, получаем для $i = 1, \dots, n$ такие действительные числа a

и a_i , что $\sum_{i=1}^n |a_i| > 0$ и

$$\sum_{i=1}^n a_i x_i \geq a \text{ для любого } x = (x_1, \dots, x_n) \text{ в } \bar{V},$$

$$\sum_{i=1}^n a_i x_i < a \text{ для любого } x = (x_1, \dots, x_n) \text{ в } \underline{V}.$$

Из определения $\underline{V}$ имеем для каждого i , $1 \leq i \leq n$, что $a_i \geq 0$. Поскольку $x = (h_1(m), \dots, h_n(m))$ находится на границе подмножества $\underline{V}$, то $\sum_{i=1}^n a_i h_i(\hat{m}) \leq a$, а так как $x \in \bar{V}$, то $\sum_{i=1}^n a_i h_i(\hat{m}) = a$. Поскольку $x = (h_1(m), \dots, h_n(m))$ принадлежит $\bar{V}$ для любого m из M , то

$$\sum_{i=1}^n a_i h_i(\hat{m}) = \min_M \sum_{i=1}^n a_i h_i(m).$$

Заметим, что $\sum_{i=1}^n a_i h_i$ строго выпукла на M , так как h_i строго

выпуклы, а $a_i \geq 0$ для $i = 1, \dots, n$ и $\sum_{i=1}^n a_i > 0$. Следовательно,

$\hat{m}$ — единственная точка в M , минимизирующая $\sum_{i=1}^n a_i h_i$ на M .

Тем самым доказательство завершено.

При доказательстве этой теоремы было показано, что вектор коэффициентов $\bar{a} = (a_1, \dots, a_n)$, определяющий разделяющую гиперплоскость, может быть использован при построении аддитивного представления (6.19) заданной глобальной функции качества.

Линейные операторы оценки косвенного эффекта управляющих воздействий

Чтобы сделать систему координируемой, часто приходится видоизменять заданные локальные функции качества. Такие модификации удобно осуществлять с помощью операторов оценки косвенного эффекта управляющих воздействий, определенных в гл. 5. Поэтому ниже мы воспользуемся этими операторами, особенно их линейными аппроксимациями. В связи с этим предположим, что пространства $\bar{M}$ и Y — банаховы пространства, и проведем линеаризацию операторов оценки косвенного эффекта управляющих воздействий.

В гл. 5 такие операторы для двухуровневой системы были определены через глобальную функцию качества, подпроцессы и функции взаимодействия для подпроцессов.

Линеаризованный полный оператор оценки косвенного эффекта управляющих воздействий

В гл. 5 i -м полным оператором оценки косвенного эффекта управляющих воздействий в точке $\tilde{m}$ мы назвали такой оператор $\tilde{\Gamma}_i(m): \tilde{M} \rightarrow R$, что для всех m в M

$$\Gamma_i(\tilde{m}, m) = G(\tilde{m}, \bar{P}(\tilde{m}, K(\tilde{m} + \Pi_{iM}m))) - G(\tilde{m}, P(\tilde{m})). \quad (6.22)$$

Линейным i -м полным оператором оценки косвенного эффекта управляющих воздействий в точке $\tilde{m} \in \bar{M}$ назовем линейный функционал $\Gamma'_i(\tilde{m})$ на $\bar{M}$, определенный для всех m из $\bar{M}$ с помощью предела

$$\Gamma'(\tilde{m})m = \lim_{\lambda \rightarrow 0} \frac{1}{\lambda} [\Gamma_i(\tilde{m}), \lambda(m)], \quad (6.23)$$

при условии что предел существует для всех m в M . Если $\Gamma'_i(m)$ существует, то, подставив выражение (6.22) в (6.23), получим

$$\Gamma'_i(\tilde{m})m = G'_Y(\tilde{m}, \tilde{y}) \bar{P}'_U(\tilde{m}, \tilde{u}) K'(\tilde{m}) \Pi_{iM}m, \quad (6.24)$$

где $\tilde{y} = P(\tilde{m})$ и $\tilde{u} = K(\tilde{m})$. Величины $G'_Y(\tilde{m}, \tilde{y})$, $\bar{P}'_U(\tilde{m}, \tilde{u})$, $K'(\tilde{m})$ являются производными Фреше функций: G по y в точке $(\tilde{m}, \tilde{y})$, $\bar{P}$ по u в точке $(\tilde{m}, \tilde{u})$ и K в точке $\tilde{m}$ соответственно.

Линеаризованные частные операторы оценки косвенного эффекта управляющих воздействий

Назовем j -м частным оператором оценки косвенного эффекта управляющих воздействий в точке $\tilde{m} \in \bar{M}$ такой оператор

$\Gamma_{ij}(\tilde{m}): \bar{M} \rightarrow R$, что для всех m в $\bar{M}$

$$\Gamma_{ij}(\tilde{m}, m) = G(\tilde{m}, \bar{P}(\tilde{m}, \tilde{u} + \mathcal{K}_j(\tilde{m} + \Pi_{iM}m))) - G(\tilde{m}, P(\tilde{m})),$$

где $\tilde{u} = K(\tilde{m})$ и

$$\mathcal{K}_j(m) = (0, \quad 0, K_j(m), \quad 0).$$

Тогда линейным ij -м частным оператором оценки косвенного эффекта управляющих воздействий в точке $\tilde{m} \in \bar{M}$ назовем линейный функционал $\Gamma'_{ij}(\tilde{m})$ на $\bar{M}$, определенный для всех m в M с помощью предела

$$\Gamma'_{ij}(\tilde{m})m = \lim_{\lambda \rightarrow 0} \frac{1}{\lambda} [\Gamma_{ij}(\tilde{m}, \lambda m)]$$

в предположении, что предел этот существует для всех $\tilde{m}$ в $\bar{M}$. Если $\Gamma'_{ij}(\tilde{m})$ существует, то

$$\Gamma'_{ij}(\tilde{m})m = G'_Y(\tilde{m}, \tilde{y}) \mathcal{F}'_{iU}(\tilde{m}_i, \tilde{u}_i) K'_j(\tilde{m}) \Pi_{iM}m, \quad (6.25)$$

где опять $\tilde{y} = P(\tilde{m})$ и $\tilde{u}_i = K_j(\tilde{m})$.

Линеаризованный оператор оценки эффективности связующих входов

Назовем i -м оператором оценки эффективности связующих входов в точке $\tilde{m} \in \bar{M}$ такой оператор $\Delta_i(\tilde{m}): \bar{U}_i \rightarrow R$, что для всех u_i из $\bar{U}_i$

$$\Delta_i(\tilde{m}, u_i) = G(\tilde{m}, \bar{P}(\tilde{m}, \tilde{u} + u_i)) - G(\tilde{m}, P(\tilde{m})),$$

где $\tilde{u} = K(m)$ и $u_i = (0, \quad 0, \quad 0, \quad 0)$.

Линейным i -м оператором оценки эффективности связующих входов в точке $\tilde{m} \in \bar{M}$ будем называть линейный функционал $\Delta'_i(\tilde{m})$ на $\bar{U}_i$, определенный для всех u_i в $\bar{U}_i$ с помощью предела

$$\Delta'_i(\tilde{m})u_i = \lim_{\lambda \rightarrow 0} \frac{1}{\lambda} [\Delta_i(\tilde{m}, \lambda u_i)]$$

в предположении, что предел этот существует для всех u_i в $\bar{U}_i$. Если $\Delta'_i(\tilde{m})$ существует, то

$$\Delta'_i(\tilde{m})u_i = G'_Y(\tilde{m}, \tilde{y}) \mathcal{F}'_{iU}(\tilde{m}_i, \tilde{u}_i)u_i, \quad (6.26)$$

где $\tilde{y} = P(\tilde{m})$ и $\tilde{u}_i = K_i(\tilde{m})$.

Некоторые соотношения для линеаризованных операторов оценки

В этой главе рассматриваются преимущественно системы с аддитивной глобальной функцией качества. А именно, мы будем пола-

гать, что функция G аддитивна в том смысле, что

$$G(m, y) = \sum_{i=1}^n G_i(m_i, y_i, H_i(m, y)) \quad (6.27)$$

Следовательно, когда $\tilde{y} = P(\tilde{m})$ и $\tilde{u} = K(\tilde{m})$, имеем

$$G'_Y(\tilde{m}, \tilde{y}) = \sum_{i=1}^n [G'_{iY}(\tilde{m}_i, \tilde{y}_i, \tilde{u}_i) \Pi_{iY} + G'_{iU}(\tilde{m}_i, \tilde{y}_i, \tilde{u}_i) H'_{iY}(\tilde{m}, \tilde{y})],$$

$$G'_M(\tilde{m}, \tilde{y}) = \sum_{i=1}^n [G'_{iM}(\tilde{m}_i, \tilde{y}_i, \tilde{u}_i) \Pi_{iM} + G'_{iU}(\tilde{m}_i, \tilde{y}_i, \tilde{u}_i) H'_{iM}(\tilde{m}, \tilde{y})].$$

Так как функции H_i являются связующими функциями под-процессов, то они не зависят от y_i и m_i ; следовательно, $H'_{iM}(\tilde{m}, \tilde{y}) \Pi_{iM} \equiv 0$ и $H'_{iY}(\tilde{m}, \tilde{y}) \Pi_{iY} \equiv 0$. Более того, из (6.18) получаем, что

$$P'(\tilde{m}) = \bar{P}'_M(\tilde{m}, \tilde{u}) + \bar{P}'_U(\tilde{m}, \tilde{u}) K'(\tilde{m}) =$$

$$\sum_{i=1}^n [\mathcal{P}'_{iM}(\tilde{m}_i, \tilde{u}_i) \Pi_{iM} + \mathcal{P}'_{iM}(\tilde{m}_i, \tilde{u}_i) K'_i(\tilde{m})].$$

Воспользуемся теперь этими тождествами для получения других форм линеаризованных операторов оценки эффективности, используемых в этой главе.

Предположим, что глобальная целевая функция $g: \bar{M} \rightarrow R$ определяется через P и G в виде

$$g(m) = G(m, P(m))$$

и что для каждого i , $1 \leq i \leq n$, $g_i: \bar{\mathcal{M}}_i \times \bar{U}_i \rightarrow R$ является локальной целевой функцией, соответствующей G_i :

$$g_i(m_i, u_i) = G_i(m_i, \mathcal{P}_i(m_i, u_i), u_i).$$

В этом случае, предполагая, что в точке $\tilde{m}$ существуют соответствующие производные Фреше и что $\tilde{u} = K(\tilde{m})$, получим следующие тождества:

$$(1) \Delta'_i(\tilde{m}) u_i = g'_{iU}(\tilde{m}_i, \tilde{u}_i) u_i, \quad (6.28)$$

$$(2) \Gamma'_{ij}(\tilde{m}) m = g'_{iU}(\tilde{m}_j, \tilde{u}_j) K'_j(\tilde{m}) \Pi_{iM} m = \Delta'_j(\tilde{m}) K'_j(\tilde{m}) \Pi_{iM} m, \quad (6.29)$$

$$(3) \Gamma'_i(\tilde{m}) m = \sum_{j=1}^n g'_{jU}(\tilde{m}_j, \tilde{u}_j) K'_j(\tilde{m}) \Pi_{iM} m = \sum_{j=1}^n \Gamma'_{ij}(\tilde{m}) m =$$

$$= [g'(\tilde{m}) - g'_{iM}(\tilde{m}_i, u_i)] \Pi_{iM} m. \quad (6.30)$$

Эти тождества указывают не только на связь между линеаризованными операторами оценки эффективности, но и проливают

некоторый свет на их смысловое значение в случае, когда глобальная функция качества обладает свойством аддитивности.

3. ПРИМЕНЕНИЕ ПРИНЦИПОВ СОГЛАСОВАНИЯ

Результаты предыдущей главы, относящиеся к применимости принципов согласования и координируемости с помощью этих принципов, могут быть непосредственно использованы и для двухуровневых систем с динамическими подпроцессами. Преимуществом изучения систем со слабыми ограничениями, таких, как в гл. 5, является большая общность получаемых результатов. С другой стороны, большая часть сформулированных в гл. 5 условий, при которых система координируема на основе принципов согласования, трудно поддается проверке. Поэтому мы воспользуемся структурой динамических подпроцессов, чтобы ввести более реалистичные условия. Так как условия применимости принципов согласования взаимодействий и функций качества практически совпадают (теоремы 5.12 и 5.13), нам достаточно рассмотреть только один из них, а именно принцип согласования взаимодействий.

Основные результаты

Прежде чем заняться анализом, напомним некоторые результаты гл. 5.

Пусть для заданной двухуровневой системы: 1) $g: \bar{M} \rightarrow R$ — глобальная целевая функция, соответствующая заданной глобальной функции качества G , $g(m) = G(m, P(m))$; 2) заданное множество $\mathcal{B}$ таково, что каждому элементу β из $\mathcal{B}$ соответствуют локальные функции качества $G_{i\beta}$, $1 \leq i \leq n$; 3) $g_{i\beta}: \bar{\mathcal{M}}_i \times \bar{U}_i \rightarrow R$ — локальные целевые функции, соответствующие $G_{i\beta}$,

$$g_{i\beta}(m_i, u_i) = G_{i\beta}(m_i, \mathcal{P}_i(m_i, u_i), u_i). \quad (6.31)$$

Тогда множество $\mathcal{B}$ является множеством сигналов, координирующих цели, и для каждого β из $\mathcal{B}$ i -я локальная задача заключается в минимизации $g_{i\beta}$ на множестве $\mathcal{M}_i \times U_i$, в то время как глобальная задача состоит в минимизации g на множестве M .

Для нашего анализа наиболее важны следующие два результата гл. 5:

1) Принцип согласования взаимодействий применим к любой двухуровневой системе, обладающей свойством монотонности (следствие 5.19). Напомним, что двухуровневая система обладает свойством монотонности, если для каждого β из $\mathcal{B}$ межуровневая функция системы монотонна и, следовательно, глобальная функция качества монотонно связана с локальными функциями качества. Следовательно, для любого координирующего цели сигнала β из $\mathcal{B}$ всякая пара (m^β, u^β) , такая, что для каждого i , $1 \leq i \leq n$,

пара (m_i^β, u_i^β) является оптимальным решением i -й локальной задачи и связующие входы согласованы, $u^\beta = K(m^\beta)$ представляет собой оптимальное решение, т. е. m^β минимизирует g на M и является, таким образом, глобально оптимальным управляющим воздействием.

2) Если рассматриваемая двухуровневая система обладает свойством монотонности, то для координации с помощью принципа согласования взаимодействий необходимо существование такого координирующего цели сигнала β из $\mathcal{B}$, что

$$\min_{M \times U} g_\beta(\beta, m, u) = \min_M g(m), \quad (6.32)$$

где g_β — кажущаяся глобальная целевая функция системы (теорема 5.21). Более того, если межуровневые функции системы являются строго монотонными, существование координирующего сигнала, для которого справедливо равенство (6.32), необходимо и достаточно для координируемости на основе принципа согласования взаимодействий (теорема 5.22).

Легко привести примеры двухуровневых систем, обладающих свойством монотонности. Некоторые из них уже были указаны в гл. 5. Например, если межуровневая функция заданной двухуровневой системы имеет мультипликативный вид, то система обладает свойством монотонности при условии, что каждая локальная целевая функция неотрицательна, и, кроме того, ее межуровневые функции строго монотонны, если каждая локальная целевая функция строго положительна. В действительности межуровневые функции двухуровневой системы, обладающей свойством монотонности, могут быть любой комбинацией операций сложения и умножения, которая сохраняет порядок предпочтений.

Ограничимся теми двухуровневыми системами, межуровневые функции которых являются аддитивными; считая, например, что мы имеем дело с затратами, получаем, что глобальные затраты равны сумме локальных затрат в том смысле, что для всех m из $\bar{M}$ и β из $\mathcal{B}$

$$g(m) = \sum_{i=1}^n g_{i\beta}(m_i, K_i(m)). \quad (6.33)$$

Такие системы обладают свойством строгой монотонности.

При нашем анализе двухуровневых систем будем считать, что глобальная функция качества нам задана (так же, как и локальные функции качества G_i , $1 \leq i \leq n$) и такова, что

$$g(m) = \sum_{i=1}^n g_i(m_i, K_i(m)), \quad (6.34)$$

где функции g_i , $1 \leq i \leq n$, определяются посредством функций G_i с помощью выражения (6.31); введенная таким образом глобальная

функция качества аддитивна. Если к тому же

$$\min_{M \times U} \sum_{i=1}^n g_i(m_i, u_i) = \min_M g(m), \quad (6.35)$$

то рассматриваемая система координируема с помощью принципа согласования взаимодействий: локальные задачи имеют такие оптимальные решения, что связующие входы согласуются и, следовательно, выбранные таким образом локальные управляющие воздействия являются глобально оптимальными. Если равенство (6.35) не выполняется, то система не координируема на основе принципа согласования взаимодействий, а если все-таки использовать этот принцип, необходимо модифицировать заданные локальные функции качества.

Если глобальная функция качества системы аддитивна, то наиболее целесообразными представляются так называемые аддитивные модификации с нулевой суммой. Модификации локальных функций качества являются аддитивными с нулевой суммой, если для каждого β из $\mathcal{B}$ модифицирующие члены $\mu_{i\beta}$ вводятся согласно равенству

$$G_{i\beta} = G_i + \mu_{i\beta},$$

и обладают следующим свойством для всех m из M :

$$\sum_{i=1}^n \mu_{i\beta}(m_i, K_i(m), \mathcal{F}_i(m_i, K_i(m))) = 0. \quad (6.36)$$

Условие (6.36) означает, что эффект от модификации локальных функций качества сводится к нулю всякий раз, когда связующие входы согласованы, т. е. такие модификации локальных функций качества никак не изменяют сумму фактических значений локальных функций качества. Такие модификации обеспечивают выполнение равенства (6.33). В этом разделе в большинстве случаев используются аддитивные модификации локальных функций качества с нулевой суммой. Поэтому мы начнем наше исследование координируемости на основе принципа согласования взаимодействий со следующего замечания.

Лемма 6.3

Предположим, что глобальная функция качества рассматриваемой двухуровневой системы аддитивна и, кроме того, для нее существует глобально оптимальное (минимизирующее) управляющее воздействие, а используемые модификации локальных функций качества являются аддитивными с нулевой суммой. Тогда для координируемости с помощью принципа согласования взаимодействий необходимо и достаточно существование такого коорди-

нирующего цели сигнала β из $\mathcal{B}$, что

$$\inf_{M \times U} \sum_{i=1}^n g_{i\beta}(m_i, u_i) = \inf_M g(m). \quad (6.37)$$

Доказательство. Функция g_{β} ,

$$g_{\beta}(\beta, m, u) = \sum_{i=1}^n g_{i\beta}(m_i u_i),$$

есть кажущаяся глобальная целевая функция системы. Поскольку функция g имеет минимум на M , то правая часть равенства (6.37) существует и является минимумом g на M . Если равенство (6.37) выполняется, то точка $\hat{m}$, минимизирующая функцию g , порождает пару $(\hat{m}_i, K_i(\hat{m}))$, которая минимизирует $g_{i\beta}$. Следовательно, равенство (6.32) справедливо тогда и только тогда, когда справедливо равенство (6.37). Так как рассматриваемая система обладает свойством строгой монотонности (это следует из исходного предположения), то справедливость леммы вытекает из п. 2 на стр. 236.

Мы будем опираться на лемму 6.3 при выводе условий, гарантирующих координируемость с помощью принципа согласования взаимодействий. Ниже мы введем дополнительные предположения, касающиеся таких свойств, как выпуклость, компактность, линейность и т. д., и определим условия, при которых можно выполнить требования приведенной выше леммы, а следовательно, и обеспечить координирование на основе принципа согласования взаимодействий

Линейная модификация

Один из способов применения аддитивной модификации локальных функций качества с нулевой суммой заключается в выборе для каждого β из $\mathcal{B}$ таких модифицирующих¹ членов $\mu_{i\beta} = G_i - G_{i\beta}$, $1 \leq i \leq n$, что

$$\mu_{i\beta}(m_i, u_i, y_i) = \alpha(m_i) + \beta_i(u_i) + \gamma_i(y_i), \quad (6.38)$$

где α_i , β_i , γ_i — такие заданные и зависящие от β функционалы, что на M

$$\sum_{i=1}^n [\alpha_i \Pi_{iM} + \beta_i K_i + \gamma_i \Pi_{iY} P] = 0. \quad (6.39)$$

Модификации локальных функций качества, определяемые выражением (6.38), являются модификациями с нулевой суммой тогда и только тогда, когда для каждого β из $\mathcal{B}$ на M выполняется равенство (6.39). Это вытекает из равенств (6.36) и (6.38). Следовательно, задача построения аддитивной модификации с нулевой

суммой при наличии ограничения (6.38) сводится к задаче нахождения решений уравнения (6.39). Однако это трудная задача, и в некоторых случаях существует только ее тривиальное решение.

Поэтому мы ограничимся некоторыми частными случаями, которых решения уравнения (6.39) существуют, не тривиальны и легко находятся. Таким является случай, когда связующие функции подпроцессов двухуровневой системы сепарабельны: для всех i , $1 \leq i \leq n$, либо

$$H_i(m, y) = \sum_{j=1}^n H_{ij}(m_j, y_j) \text{ на } \bar{M} \times Y, \quad (6.40a)$$

либо

$$K_i(m) = \sum_{j=1}^n K_{ij}(m_j) \text{ на } \bar{M}. \quad (6.40b)$$

Если подпроцессы двухуровневой системы получаются в результате применения декомпозиции, канонической относительно выходов (пример 6.4), то система обладает свойством (6.40a). В более общем случае система обладает свойством (6.40a) всякий раз, когда функции H_i линейны; однако линейность не необходима. С другой стороны, если подпроцессы двухуровневой системы получаются в результате декомпозиции, канонической относительно входов (пример 6.4), то система обладает свойством (6.40b).

Если заданная двухуровневая система обладает свойством (6.40a) и функционалы β_i линейны для каждого i , $1 \leq i \leq n$, то решения уравнения (6.39) определяются из n уравнений

$$\alpha_i(m_i) + \gamma_i(y_i) = - \sum_{j=1}^n \beta_j H_{ji}(m_i, y_i).$$

Следовательно, аддитивная модификация локальных функций качества с нулевой суммой введена корректным образом, если $\mathcal{R} = U_1^* \times \dots \times U_n^*$, где верхний индекс $*$ обозначает сопряженное пространство и для каждого $\beta = (\beta_1, \dots, \beta_n)$ из $\mathcal{R}$ и каждого i , $1 \leq i \leq n$, модифицирующие члены задаются в виде

$$\mu_{i\beta}(m_i, u_i, y_i) = \beta_i u_i - \sum_{j=1}^n \beta_j H_{ji}(m_i, y_i). \quad (6.41)$$

С другой стороны, если заданная двухуровневая система обладает свойством (6.40b) и функционалы β_i для каждого i , $1 \leq i \leq n$, снова линейны, то решения уравнения (6.39) определяются из $2n$ уравнений:

$$\alpha_i = - \sum_{j=1}^n \beta_j K_{ji}, \quad \gamma_i = 0.$$

Следовательно, аддитивную модификацию с нулевой суммой можно реализовать, если $\mathcal{R} = U_1^* \times \dots \times U_n^*$ и для каждого $\beta =$

$$= (\beta_i, \quad \beta_n) \text{ из } \mathcal{B} \text{ и каждого } 1 \leq i \leq n, \\ \mu_{i\beta}(m_i, u_i, y_i) = \beta_i u_i - \sum_{j=1}^n \beta_j K_{ji}(m_i). \quad (6.42)$$

Две формы модификации, определяемые выражениями (6.41) и (6.42), действительно являются модификациями с нулевой суммой, и, так как координирующие цели входы являются линейными функционалами, мы будем называть эти две формы (6.41) и (6.42) *линейными с нулевой суммой*.

Модифицирующие члены, появляющиеся при использовании линейных модификаций с нулевой суммой для локальных функций качества, с позиций каждого i -го локального решающего элемента можно интерпретировать как линейные функции от связующего входа u_i , который был бы «желателен» для этого элемента («спрос» i -го элемента), и от связующих входов u_{ji} $K_{ji}(m_i)$ или $u_{ji} = H_{ij}(m_i, y_i)$, которые он в свою очередь намерен послать другим решающим элементам (его «предложение»). В этом случае линейные функционалы $\beta_i, \dots, \beta_n$ играют роль цен (как рассматривалось в гл. 5). С точки зрения этой формы модификаций координацию с помощью принципа согласования взаимодействий можно интерпретировать как такое назначение цен, при котором балансируются спрос и предложение, в то время как координацию на основе принципа согласования функций качества можно в свою очередь интерпретировать как назначение цен, но теперь уже такое, что балансируются ожидаемые и действительные затраты.

Предположив, что модификация локальных функций качества является линейной с нулевой суммой, и введя подходящие предположения относительно выпуклости, непрерывности и компактности, можно получить более конкретные, чем в лемме 6.3, условия, гарантирующие координируемость при использовании принципа согласования взаимодействий.

Теорема 6.4

Предположим, что глобальная функция качества рассматриваемой двухуровневой системы аддитивна, что для нее имеется глобально оптимальное управляющее воздействие, а модификации локальных функций качества являются линейными с нулевой суммой. Кроме того, предположим следующее:

1) Для каждого $1 \leq i \leq n$, множество $\mathcal{M}_i \times U_i$ слабо компактно и выпукло, а для каждого β из $\mathcal{B}$ локальные функции качества $g_{i\beta}$ выпуклы и полунепрерывны снизу.

2) Существует элемент $\hat{\beta}$ в $\mathcal{B}$, такой, что

$$\inf_{M \times U} \sum_{i=1}^n g_{i\hat{\beta}}(m_i, u_i) = \sup_{\mathcal{B}} \inf_{M \times U} \sum_{i=1}^n g_{i\beta}(m_i, u_i).$$

Тогда система координируема с помощью принципа согласования взаимодействий.

Доказательство. В силу леммы 6.3 необходимо только найти такое $\beta \in \mathcal{B}$, чтобы выполнялось равенство (6.37). Так как модификации линейны с нулевой суммой, то $\mathcal{B} = U_1^* \times \dots \times U_n^*$. Для удобства определим $\bar{K}_i$, $1 \leq i \leq n$, на $\bar{M} \times \bar{U}$ в виде

$$\bar{K}_i(m, u) = H_i(m, \sum_{j=1}^n \mathcal{F}_j(m_j, u_j)) \text{ или } \bar{K}_i(m, u) = K_i(m) \quad (6.43)$$

в зависимости от того, выражаются ли модификации в соответствии с выражениями (6.41) или (6.42). Следовательно,

$$g_{\mathcal{B}}(\beta, m, u) = \sum_{i=1}^n g_i(m_i, u_i) + \sum_{i=1}^n \beta_i [u_i - \bar{K}_i(m, u)] \quad (6.44)$$

является кажущейся глобальной целевой функцией системы.

Однако

$$\sup_{\mathcal{B}} \sum_i \beta_i (u_i - \bar{K}_i(m, u)) = \infty,$$

если $u_i \neq \bar{K}_i(m, u)$ для некоторого i , $1 \leq i \leq n$.

В силу условия 1 каждая функция $g_{i\beta}$ имеет минимум на множестве $\mathcal{M}_i \times U_i$; поэтому, так как модификация является модификацией с нулевой суммой,

$$\inf_{M \times U} \sup_{\mathcal{B}} g_{\mathcal{B}}(\beta, m, u) = \inf_M \sup_{\mathcal{B}} g_{\mathcal{B}}(\beta, m, K(m)) = \inf_M g(m). \quad (6.45)$$

В силу того же условия 1 функция $g_{\mathcal{B}}$ выпукла и слабо полунепрерывна снизу на $M \times U$ для каждого β в $\mathcal{B}$ и вогнута на β для каждого (m, u) в $M \times U$. Кроме того, множества M и U слабо компактны. Отсюда следует, что

$$\sup_{\mathcal{B}} \inf_{M \times U} g_{\mathcal{B}}(\beta, m, u) = \inf_{M \times U} \sup_{\mathcal{B}} g_{\mathcal{B}}(\beta, m, u). \quad (6.46)$$

Совместное рассмотрение равенств (6.45) и (6.46) показывает, что выбранное в соответствии с предположением 2 значение $\hat{\beta}$ удовлетворяет равенству (6.37). Следовательно, теорема доказана.

Если подпроцессы $\mathcal{F}_i$, $1 \leq i \leq n$, рассматриваемой двухуровневой системы линейны и непрерывны, то предположение 1 предыдущей теоремы о выпуклости и непрерывности можно заменить следующим предположением: для каждого i , $1 \leq i \leq n$, и каждого β в $\mathcal{B}$ модифицированная локальная функция качества $G_{i\beta}$ выпукла и полунепрерывна снизу на $\mathcal{M}_i \times U_i \times \mathcal{Y}_i$. Конечно, предположение 1 не требует выпуклости самих по себе $G_{i\beta}$. Однако, если, кроме линейности и непрерывности подпроцессов, свя-

зующие функции H_i , $1 \leq i \leq n$, также линейны и непрерывны, то, предполагая, что модификация локальной функции качества является линейной с нулевой суммой, получаем, что предположение 1 выполняется, если каждая заданная локальная функция качества G_i , $1 \leq i \leq n$, выпукла и полунепрерывна снизу на своей области определения.

Пожалуй, труднее всего проверить справедливость предположения 2 предыдущей теоремы. Допустим, что для каждого β из $\mathcal{B}$

$$g^0(\beta) = \inf_{M \times U} \sum_{i=1}^n g_{i\beta}(m_i, u_i);$$

предположение 2 требует, чтобы g^0 имело максимум на $\mathcal{B}$. Используя в качестве множеств M , U_i и Y подмножества рефлексивных банаховых пространств, мы можем, усилив предположение 1 и введя дополнительное предположение, проверить справедливость предположения 2, а следовательно, и координируемости с помощью принципа согласования взаимодействий.

Теорема 6.5

Предположим, что глобальная функция качества заданной двухуровневой системы аддитивна, что для нее существует глобально оптимальное управляющее воздействие, а модификации локальных функций качества являются модификациями с нулевой суммой. Более того, считая, что $\bar{M}$ и $\bar{U}_i$ для каждого i , $1 \leq i \leq n$, суть рефлексивные банаховы пространства, предположим следующее:

1) для каждого i , $1 \leq i \leq n$, множества $\mathcal{M}_i$ и U_i замкнуты, ограничены и выпуклы; каждая функция $g_{i\beta}$, $\beta \in \mathcal{B}$, выпукла и полунепрерывна снизу на $\mathcal{M}_i \times U_i$;

2) $g^0(\beta) < g^0(0)$ на сфере $\mathcal{S}_r$ в $\mathcal{B}$ радиуса $r > 0$ с центром в начале координат;

3) функции $\bar{K}_i$ ограничены на $M \times U$.

Тогда система координируема с помощью принципа согласования взаимодействий.

Доказательство. Замкнутое, ограниченное, выпуклое множество в рефлексивном банаховом пространстве слабо компактно, и поэтому множество $\mathcal{M}_i \times U_i$ для каждого i , $1 \leq i \leq n$, слабо компактно и выпукло. Следовательно, необходимо только показать, что g^0 имеет максимум на $\mathcal{B}$; координируемость тогда следует из теоремы 6.4. Чтобы показать это, используем тот факт, что g^0 имеет максимум на $\mathcal{B}$, если функция g^0 вогнута и непрерывна на $\mathcal{B}$, а $g^0(\beta) \rightarrow -\infty$ при $\|\beta\| \rightarrow \infty$.

Чтобы показать вогнутость, выберем β из $\mathcal{B}$ произвольно. В силу линейности модификаций с нулевой суммой, $\mathcal{B} = U_1^* \times \dots \times U_n^*$, а функция $g_{\mathcal{B}}$, определяемая выражением

(6.44), является линейной по β . Тогда для любого действительного числа λ

$$\begin{aligned} g_{\mathcal{B}}(\lambda\beta + (1-\lambda)\beta', m, u) \\ = \lambda g_{\mathcal{B}}(\beta, m, u) + (1-\lambda) g_{\mathcal{B}}(\beta', m, u). \end{aligned}$$

Но для $0 \leq \lambda \leq 1$

$$\begin{aligned} \min_{M \times U} g_{\mathcal{B}}(\lambda\beta + (1-\lambda)\beta', m, u) &\leq \min_{M \times U} \lambda g_{\mathcal{B}}(\beta, m, u) + \\ &+ \min_{M \times U} (1-\lambda) g_{\mathcal{B}}(\beta', m, u), \end{aligned}$$

поэтому

$$g^0(\lambda\beta + (1-\lambda)\beta') \geq \lambda g^0(\beta) + (1-\lambda) g^0(\beta'). \quad (6.47)$$

Следовательно, вогнутость доказана.

Положив теперь $\beta'' = \lambda\beta + (1-\lambda)\beta'$ и $0 < \lambda \leq 1$, получаем из (6.47), что

$$g^0(\beta) \leq \frac{1}{\lambda} [g^0(\beta'') - g^0(\beta')] + g^0(\beta').$$

Пусть $\beta' = 0$, $\beta \neq 0$, а λ такое, что $r = \lambda \|\beta\|$; тогда

$$g^0(\beta) \leq (\|\beta\|/r) [g^0(\beta'') - g^0(0)] + g^0(0).$$

Так как $\beta' = 0$, то $\beta'' = \lambda\beta$ и $\|\beta''\| = r$. В силу предположения 3 $g^0(\beta'') - g^0(0) < 0$, и, следовательно, $g^0(\beta) \rightarrow -\infty$ при $\|\beta\| \rightarrow \infty$. (Заметим, что $1 \leq 1/\lambda = \|\beta\|/r < \infty$.)

Покажем теперь непрерывность функции g^0 на $\mathcal{B}$. Для этого воспользуемся следующей леммой: функционал, выпуклый на топологическом линейном пространстве X , непрерывен на подмножестве X' множества X тогда и только тогда, когда существует такое непустое открытое множество Θ в X' что функция f мажорируема на Θ .

Предположим, что норма в $\mathcal{B}$ задается следующим образом: $\|\beta\| = \sum_i \|\beta_i\|$. Так как множества U_i ограничены, а $\bar{K}_i$ ограничены на $M \times U$, то для любого β из $\mathcal{B}$ существует такое действительное число $a > 0$, что

$$\left| \sum_i \beta_i (u_i - \bar{K}_i(m, u)) \right| \leq \sum_i |\beta_i u_i| + \sum_i |\beta_i \bar{K}_i(m, u)| \leq a \|\beta\|$$

и, следовательно,

$$g^0(\beta) \geq \min_{M \times U} \sum_i g_i(m_i, u_i) - a \|\beta\|.$$

Отсюда вытекает, что $-g^0$ мажорируется на любом ограниченном открытом подмножестве $\mathcal{B}$, а из леммы следует, что g^0 непрерывна на $\mathcal{B}$. Таким образом, доказательство завершено.

Если отображения $\bar{K}_i$, заданные выражениями (6.43), ведут себя подходящим образом в области своего определения, то спра-

ведливость предположения 2 предыдущей теоремы может быть проверена. Заметим, что для выбранного управляющего воздействия m из M связующий вход $u = (u_1, \dots, u_n)$, удовлетворяющий соотношениям $u_i \in \bar{K}_i(m, u)$ для каждого i , $1 \leq i \leq n$, является именно тем связующим входом, который действительно имеет место, когда подпроцессы $\mathcal{P}_i$ корректно «взаимосвязаны» (с точки зрения процесса в целом P). Это следует из соотношений, связывающих подпроцессы со всем процессом в целом, согласно выражению (6.16). Следующая теорема конкретизирует смысл понятия «подходящего поведения» отображения $\bar{K}_i$.

Теорема 6.6

Предположим, что глобальная функция качества рассматриваемой двухуровневой системы аддитивна, что для нее существует глобально оптимальное управляющее воздействие, а модификации локальных функций качества являются линейными модификациями с нулевой суммой. Кроме того, предположим, что для каждого i , $1 \leq i \leq n$, множества $\bar{M}$ и U_i — банаховы рефлексивные пространства и для каждого i , $1 \leq i \leq n$, выполняются следующие условия:

1) множества $\mathcal{M}_i$ и U_i замкнуты, ограничены и выпуклы, все функции $g_{i\beta}$ для β из $\mathcal{B}$ выпуклы и полунепрерывны снизу на $\mathcal{M}_i \times U_i$, а g_i ограничена;

2) для некоторого $\varepsilon > 0$ и для любого e_i из $\bar{U}_i$ с нормой $\|e_i\| = \varepsilon$ существует такая пара (m, u) в $M \times U$ что $e_i = u_i - \bar{K}_i(m, u)$.

Тогда система координируема с помощью принципа согласования взаимодействий.

Доказательство. Достаточно показать, что выполняется предположение 2 теоремы 6.5.

Так как функции g_i ограничены, то существуют такие конечные числа a и b , что для всех пар (m, u) в $M \times U$

$$a \leq \sum_i g_i(m_i, u_i) \leq b. \quad (6.48)$$

Так как модификации локальных функций качества суть модификации с нулевой суммой, то $\mathcal{B} = U_1^* \times \dots \times U_n^*$ и для каждого β из $\mathcal{B}$

$$\inf_{M \times U} g_{\beta}(m, u) \leq b + \inf_{M \times U} \sum_i \beta_i(u_i - \bar{K}_i(m, u)). \quad (6.49)$$

где g_{β} определяется с помощью выражения (6.44). Но в силу предположения 2 существует такое $\varepsilon > 0$, что

$$\inf_{M \times U} \sum_i \beta_i(u_i - \bar{K}_i(m, u)) \leq \inf_{\|e_i\|=\varepsilon} \sum_i \beta_i e_i \leq -\varepsilon \|\beta\|. \quad (6.50)$$

где $\|\beta\| = \sum_i \|\beta_i\|$. Положим $r = b - a + 1$. Тогда для любого β из $\mathcal{B}$, такого, что $\varepsilon \|\beta\| = r$, получаем из выражений (6.48) — (6.50)

$$\inf_{M \times U} g_{\mathcal{B}}(\beta, m, u) \leq b - \varepsilon \|\beta\| \leq a - 1 < a \leq \inf_{M \times U} g_{\mathcal{B}}(0, m, u).$$

Этим доказательство завершается.

Предположение 2 приведенной выше теоремы довольно легко удовлетворить, если линейные модификации с нулевой суммой задаются выражением (6.42), а $K_i(M)$, $1 \leq i \leq n$, ограничены.

Следствие 6.7

Предположим, что глобальная функция качества рассматриваемой двухуровневой системы аддитивна, что для нее существует глобально оптимальное управляющее воздействие, а модификации локальных функций качества введены с помощью выражений (6.42). Предположим, что для каждого i , $1 \leq i \leq n$, множества $\bar{M}$ и $\bar{U}_i$ — рефлексивные банаховы пространства и выполняются следующие условия:

1) каждая функция $g_{i\beta}$ для β из $\mathcal{B}$ выпукла и полунепрерывна снизу на $\mathcal{M}_i \times U_i$, а функция g_i ограничена;

2) множества $\mathcal{M}_i$ и U_i замкнуты, ограничены и выпуклы и U_i содержит множество $K_i(M) + E_i$, где E_i — замкнутая сфера в $\bar{U}_i$ радиусом $\varepsilon > 0$ с центром в начале координат.

Тогда система координируема с помощью принципа согласования взаимодействий.

Возможно, что наиболее серьезным предположением сформулированной теоремы является ограниченность функционалов g_i на $\mathcal{M}_i \times U_i$. Однако ограниченность обычно следует из непрерывности g_i и ограниченности $\mathcal{M}_i \times U_i$.

Применение линейных операторов оценки эффективности

Линейные операторы оценки эффективности, введенные в предыдущем разделе этой главы, можно при подходящих условиях использовать как средство модификации локальных функций качества для того, чтобы достигалась координируемость. Здесь мы покажем, что при соответствующих условиях (теорема 6.8) линейные операторы оценки эффективности позволяют найти оптимальный координирующий сигнал при использовании линейных модификаций с нулевой суммой.

Чтобы использовать линейные операторы оценки эффективности, необходимо предположить их существование в интересующих нас точках M , а для этого нужно предположить дифференцируемость описывающих подпроцессы функций $\mathcal{P}_i$, используемых локальных функций качества G_i и функций взаимодействия подпроцессов K_i .

Лемма

Предположим, что глобальная функция качества рассматриваемой двухуровневой системы аддитивна, что для нее существует глобально оптимальное управляющее воздействие $\hat{m}$ в M и, кроме того, выполняются следующие условия:

1) для каждого i , $1 \leq i \leq n$, функции $\mathcal{P}_i$, G_i и K_i дифференцируемы по Фреше по каждому из своих аргументов;

2) M — выпуклое множество.

Пусть для заданного i , $1 \leq i \leq n$, через $\bar{g}_i$ обозначен функционал

$$\bar{g}_i(m_i, u_i) = g_i(m_i, u_i) - \Delta'_i(\hat{m}) u_i + \Gamma'_i(\hat{m}) m_i, \quad (6.51)$$

где $\Delta'_i(\hat{m})$ и $\Gamma'_i(\hat{m})$ — линейные операторы оценки эффектов внутренних взаимодействий в точке $\hat{m}$, определяемые выражениями (6.26) и (6.24) соответственно. Тогда для всех $\hat{m}_i + m_i$ в $\mathcal{M}_i$ и всех $\hat{u}_i + u_i$ в U_i

$$\bar{g}'_{iM}(\hat{m}_i, \hat{u}_i) m_i \geq 0, \quad (6.52)$$

$$\bar{g}'_{iU}(\hat{m}_i, \hat{u}_i) u_i = 0. \quad (6.53)$$

Доказательство. Так как глобальная функция качества аддитивна, то глобальная целевая функция g определяется выражением (6.34). Для всех $\hat{m} + m$ из M имеем $0 \leq g(\hat{m} + m) - g(\hat{m})$. Но из предположения 1 следует, что функция g дифференцируема в точке $\hat{m}$, а из предположения 2, что

$$0 \leq g'(\hat{m}) m = \lim_{\lambda \rightarrow 0} \frac{1}{\lambda} [g(\hat{m} + \lambda m) - g(\hat{m})] \quad (6.54)$$

для всех $\hat{m} + m$ в M и $\lambda > 0$. Так как обычные правила дифференцирования применимы и к производным Фреше, то

$$\begin{aligned} g'(\hat{m}) m = \sum_{i=1}^n [g'_{iM}(\hat{m}_i, \hat{u}_i) m_i + g'_{iU}(\hat{m}_i, \hat{u}_i) K'_i(\hat{m}) m] = \\ \sum_{i=1}^n [g'_{iM}(\hat{m}_i, \hat{u}_i) + \sum_{j=1}^n g'_{iU}(\hat{m}_j, \hat{u}_j) K'_j(\hat{m}) \Pi_{iM}] m_i. \end{aligned} \quad (6.55)$$

Поскольку $M = \mathcal{M}_1 \oplus \dots \oplus \mathcal{M}_n$, из выражений (6.54) и (6.55) следует, что

$$[g'_{iM}(\hat{m}_i, \hat{u}_i) + \sum_{j=1}^n g'_{jU}(\hat{m}_j, \hat{u}_j) K'_j(\hat{m}) \Pi_{iM}] m_i \geq 0 \quad (6.56)$$

для всех $\hat{m}_i + m_i$ в $\mathcal{M}_i$. Далее, из (6.30) вытекает, что второй член выражения, стоящего в квадратных скобках в (6.56), является линейным оператором оценки косвенного эффекта управляющих

воздействий $\Gamma'_i(m)$ и

$$\bar{g}'_{im}(m_i, u_i) = g'_{im}(m_i, u_i) + \Gamma'_i(m). \quad (6.57)$$

Следовательно, $\bar{g}'_{im}(\hat{m}_i, \hat{u}_i) m_i \geq 0$ для всех $\hat{m}_i + m_i$ в $\mathcal{M}_i$, что доказывает справедливость (6.52). С другой стороны,

$$\bar{g}'_{iv}(\hat{m}_i, \hat{u}_i) u_i = [g'_{iv}(\hat{m}_i, \hat{u}_i) - \Delta'_i(\hat{m})] u_i.$$

Но из выражения (6.28) следует, что $g'_{iv}(\hat{m}_i, \hat{u}_i) = \Delta'_i(\hat{m})$; поэтому $g'_{iv}(\hat{m}_i, \hat{u}_i) \equiv 0$. Этим завершается доказательство равенства (6.53) и леммы.

Заметим, что в предположении 2 приведенной выше леммы m не обязательно является внутренней точкой M , но если m не является внутренней точкой M , необходимо подходящим образом определить дифференцируемость (т. е. одностороннюю производную).

Если $\tilde{m}$ — внутренняя точка, то точка $(\hat{m}_i, \hat{u}_i)$ есть критическая точка $\bar{g}_i$. Заметим также, что для $\bar{g}_i$ на $(\hat{m}_i, \hat{u}_i)$ имеет место локальный минимум, если g_i выпукла, причем этот локальный минимум будет абсолютным, если, кроме того, множество $\mathcal{M}_i \times \bar{U}_i$ выпукло.

Эти рассуждения приводят нас к главному результату, касающемуся использования линейных операторов оценки эффективности, когда для каждого i , $1 \leq i \leq n$, функции, связывающие подпроцессы данной двухуровневой системы K_i , удовлетворяют условию (6.40б), а линейная модификация с нулевой суммой определяется выражением (6.42).

Теорема 6.8

Предположим, что глобальная функция качества рассматриваемой двухуровневой системы аддитивна, что для нее имеется глобально оптимальное управляющее воздействие $\hat{m} \in M$ и выполняется предположение предыдущей леммы о дифференцируемости. Кроме того, предположим следующее:

1) модификации локальных функций качества являются линейными с нулевой суммой; для каждого i , $1 \leq i \leq n$, и β из $\mathcal{R}$

$$g_{i\beta}(m_i, u_i) = g_i(m_i, u_i) + \beta_i u_i - \sum_{j=1}^n \beta_j K_{ji}(m_i);$$

2) для каждого i , $1 \leq i \leq n$, множество $\mathcal{M}_i \times U_i$ выпукло и каждая функция $g_{i\beta}$ выпукла на этом множестве.

Тогда система координируема с помощью принципа согласования взаимодействий, и, более того, если точка β в $U_1^* \times \dots \times U_n^*$ такая, что для каждого i , $1 \leq i \leq n$,

$$\hat{\beta}_i = -\Delta'_i(\hat{m}), \quad (6.58)$$

где $\Delta_i(\hat{m})$ определяется выражением (6.26), то $\hat{\beta}$ является оптимальным координирующим сигналом.

Доказательство. Пусть для любого i , $1 \leq i \leq n$, мы можем доказать, что

$$g'_{i\beta M}(\hat{m}_i, \hat{u}_i) = \bar{g}'_{iM}(\hat{m}_i, \hat{u}_i) \text{ и } \bar{g}'_{i\beta M}(\hat{m}_i, \hat{u}_i) = \bar{g}'_{iU}(\hat{m}_i, u_i),$$

где $\bar{g}_i$ определяется выражением (6.51) и $\hat{u}_i = K_i(\hat{m})$. Тогда, используя предположение 2, можно показать, что $(\hat{m}_i, \hat{u}_i)$ минимизирует $g_{i\beta}$ на $\mathcal{M}_i \times U_i$.

Пусть i , $1 \leq i \leq n$, произвольно; тогда для произвольного m_i из $\bar{\mathcal{M}}_i$ имеем, что в точке $(\hat{m}_i, \hat{u}_i)$

$$g'_{i\beta M}(\hat{m}_i, u_i) m_i = [g'_{iM}(\hat{m}_i, \hat{u}_i) - \sum_{j=1}^n \beta_j K'_{ji}(\hat{m}_i)] m_i.$$

Так как K_j должно удовлетворять равенству (6.406), то

$$K'_{ji}(\hat{m}_i) m_i = K'_j(\hat{m}_i) m_i$$

и, следовательно, из выражения (6.58) получаем

$$g'_{i\beta M}(\hat{m}_i, \hat{u}_i) m_i = [g'_{iM}(\hat{m}_i, \hat{u}_i) + \sum_{j=1}^n \Delta_j(\hat{m}) K'_{ji}(\hat{m}_i)] m_i. \quad (6.59)$$

Но в соответствии с (6.30) второй член в квадратных скобках выражения (6.59) является линейным оператором оценки косвенного эффекта управляющих воздействий $\Gamma'_i(\hat{m})$, поэтому из (6.57) следует, что $g'_{i\beta M}(\hat{m}_i, \hat{u}_i) = \bar{g}'_{iM}(\hat{m}_i, \hat{u}_i)$. С другой стороны, очевидно, что $g'_{i\beta M}(\hat{m}_i, \hat{u}_i) = \bar{g}'_{iU}(\hat{m}_i, \hat{u}_i) \equiv 0$.

Поскольку $M = \mathcal{M}_1 \oplus \dots \oplus \mathcal{M}_n$, то в силу предположения 2 множество M выпукло. Следовательно, $g'_{i\beta M}(\hat{m}_i, \hat{u}_i) m_i \geq 0$ для всех $\hat{m}_i + m_i$ в $\mathcal{M}_i$. Но в силу предположения 2 $(\hat{m}_i, \hat{u}_i)$ минимизирует $g_{i\beta}$ на $\mathcal{M}_i \times U_i$. Из этого вытекает доказательство теоремы.

Теорема 6.8 применима, конечно, и к случаю, в котором подпроцессы $\mathcal{P}_i$ и функции взаимодействия подпроцессов K_i нелинейны. Если функции K_i линейны, необходима только выпуклость функционалов g_i . С другой стороны, если линейны как функции, описывающие подпроцессы $\mathcal{P}_i$, так и функции взаимодействия подпроцессов K_i , необходимо лишь предположить выпуклость рассматриваемых локальных функций качества G_i .

Применение к системам, описываемым линейными дифференциальными уравнениями

Полученные выше результаты непосредственно применимы к различным двухуровневым системам, подпроцессы которых описываются с помощью систем дифференциальных уравнений. В ка-

честве иллюстрации мы рассмотрим ряд линейных систем и попытаемся получить для них в явном виде выражения для оптимальных координирующих сигналов.

Пусть $\bar{M}$ и Y — пространства всех интегрируемых с квадратом функций на интервале $\mathcal{T} = [0, 1]$, определенных в евклидовых пространствах $\mathcal{E}_p$, $p \geq n$, и $\mathcal{E}_q$, $q \geq n$, соответственно.

Предположим, что процесс в целом P задается таким оператором с областью определения $\bar{M}$ и областью значений Y что выход $y = P(m)$ для любого m из $\bar{M}$ является решением векторного дифференциального уравнения

$$\dot{y}(t) = Ay(t) + Bm(t), \quad y(0) = 0 \quad (6.60)$$

на $\mathcal{T}$ где A и B — матрицы $q \times q$ и $q \times p$, зависящие от времени, но ограниченные на $\mathcal{T}$. Следовательно, для любого m из $\bar{M}$ и t из $\mathcal{T}$

$$P(m)(t) = \int_0^t \Phi(t, \sigma) Bm(\sigma) d\sigma,$$

где Φ — фундаментальная матрица уравнений (6.60).

Предполагается, что множество управлений M является подмножеством множества $\bar{M}$, состоящим из всех таких управляющих воздействий m , что $m(t)$ является компактным выпуклым подмножеством Ω множества $\mathcal{E}_p$ почти для всех t в $\mathcal{T}$.

Чтобы выделить подпроцессы $\mathcal{P}_i$, $1 \leq i \leq n$, используем проекционные операторы Π_{iM} и Π_{iY} , определенные ненулевыми диагональными матрицами $p \times p$ и $q \times q$ с элементами 0 и 1, расположенными по диагонали так, что $\sum_i \Pi_{iM}$ и $\sum_i \Pi_{iY}$ являются единичными матрицами. При расширении этих операторов на пространства $\bar{M}$ и Y мы будем понимать их как точечные операторы. Мы будем использовать также матрицы A_{ij} и B_{ij} , заданные следующим образом:

$$A_{ij} = \Pi_{iY} A \Pi_{jM}, \quad B_{ij} = \Pi_{iY} B \Pi_{jM}.$$

Определим теперь для каждого i , $1 \leq i \leq n$, i -й подпроцесс $\mathcal{P}_i$ таким оператором с областью определения $\bar{\mathcal{M}}_i \times \mathcal{Y}_i$ и соответствующим подмножеством $\mathcal{Y}_i$, что выход $y_i = \mathcal{P}_i(m_i, u_i)$ для любого (m_i, u_i) из $\bar{\mathcal{M}}_i \times \mathcal{Y}_i$ является решением дифференциального уравнения

$$\dot{y}_i(t) = A_{ii}y_i(t) + B_{ii}m_i(t) + u_i(t), \quad y_i(0) = 0 \quad (6.61)$$

на $\mathcal{T}$. Следовательно, для любого $(m_i, u_i) \in \bar{\mathcal{M}}_i \times \mathcal{Y}_i$ и $t \in \mathcal{T}$

$$\mathcal{P}_i(m_i, u_i)(t) = \int_0^t \Phi_i(t, \sigma) [B_{ii}(\sigma) m_i(\sigma) + u_i(\sigma)] d\sigma,$$

где $\Phi_i(t, \sigma)$ является фундаментальной матрицей системы (6.61); действительно,

$$\Phi_i(t, \sigma) = \Pi_{iY} \Phi(t, \sigma) \Pi_{iY}.$$

Зависимость $u_i(t)$ от нелокальных входов и выходов (т. е. входящих в состав остальных подпроцессов $\mathcal{P}_j, j \neq i$, и подчиненных другим локальным управляющим подсистемам d_i) описывается выражением

$$u_i(t) = \sum_{j \neq i} [B_{ij} m_j(t) + A_{ij} y_j(t)] = H_i(m, y)(t),$$

которое определяет связующие функции H_i как линейные операторы на $\bar{M} \times Y$. Легко проверить, что уравнение (6.61) удовлетворяется для всех m из $\bar{M}$.

Предположим, что для каждого $i, 1 \leq i \leq n$, локальные функции качества G_i задаются на $\bar{\mathcal{M}}_i \times \mathcal{Y}_i$ интегралами

$$G_i(m_i, u_i, y_i) = \int_{\mathcal{T}} \theta_i(t, m_i(t), u_i(t), y_i(t)) dt, \quad (6.62)$$

а глобальная функция качества G аддитивна:

$$G(m, y) = \sum_{i=1}^n G_i(m_i, H_i(m, y), y_i).$$

Предполагается также, что функции θ_i для каждого $i, 1 \leq i \leq n$, и $t \in \mathcal{T}$ выпуклы и непрерывны.

Множества локальных управлений $\mathcal{M}_i, 1 \leq i \leq n$, являются проекциями M на пространства $\bar{\mathcal{M}}_i$, и, следовательно, (6.14) удовлетворяется. Для того чтобы пространство $\bar{M}$ являлось прямой суммой $\mathcal{M}_1 \oplus \dots \oplus \mathcal{M}_n$, примем, что $\Omega = \Omega_1 \oplus \dots \oplus \Omega_n$, где $\Omega_i = \Pi_{iM}(\Omega), 1 \leq i \leq n$.

Применим теперь ранее полученные в этом разделе результаты и покажем координируемость рассматриваемой двухуровневой системы в том случае, если используются линейные модификации локальных функций качества с нулевой суммой; будем предполагать, что в качестве множества связующих входов $U_i, 1 \leq i \leq n$, выбраны подходящие подмножества пространства $\mathcal{Y}_i$, причем этот выбор производится в соответствии с формой применяемых линейных модификаций с нулевой суммой. Вследствие предположений о линейности можно использовать обе формы. Пространство $\mathcal{R} = Y$ можно использовать в качестве множества координирующих сигналов, так как любой ограниченный линейный функционал на пространстве L_2 будет однозначно определяться соответствующими внутренними произведениями. Далее, мы будем считать, что для каждого β из $\mathcal{R} = Y$ входящие в него β_i являются соответствующими проекциями на $\mathcal{Y}_i$.

Прежде всего рассмотрим случай, когда модификации задаются выражением (6.41). Применяя теорему 6.6, получаем:

Следствие 6.9

Предположим, что для каждого i , $1 \leq i \leq n$, и $\beta \in \mathcal{B}$ локальные функции качества G_i рассматриваемой двухуровневой системы задаются выражением

$$G_{i\beta}(m_i, u_i, y_i) = G_i(m_i, u_i, y_i) + \langle \beta_i, u_i \rangle - \sum_{j=1}^n \langle \beta_j, H_{ji}(m_i, y_i) \rangle. \quad (6.63)$$

Допустим, что для каждого i , $1 \leq i \leq n$, $G_i(m_i, u_i, y_i)$ — ограниченные функции на $\mathcal{M}_i \times U_i$, а U_i — такие замкнутые, ограниченные, выпуклые подмножества $\mathcal{Y}_i$, что для некоторого $\varepsilon > 0$ они содержат все $u_i = H_i(m, y) + e_i$, когда $(m, y) \in M \times Y$ и $e_i \in \mathcal{Y}_i$ с нормой $\|e_i\| \leq \varepsilon$, и при этом на $\mathcal{T}$ имеет место уравнение

$$\dot{y}(t) = Ay(t) + Bm(t) + e_i(t), \quad y(0) = 0. \quad (6.64)$$

В этом случае система координируема с помощью принципа согласования взаимодействий.

Доказательство. Легко видеть, что все предположения теоремы 6.6, за исключением предположения 2, удовлетворяются. Поэтому достаточно показать, что предположение 2 справедливо для некоторого $\varepsilon > 0$.

Пусть для произвольного i , $1 \leq i \leq n$, e_i — какой-либо элемент в $\mathcal{Y}_i$ с нормой $\|e_i\| = \varepsilon$, где $\varepsilon > 0$ и удовлетворяет условиям доказываемой теоремы. Предположим, что точка m в M задана и y из Y удовлетворяет уравнению (6.64) на $\mathcal{T}$. Тогда $u_i = H_i(m, y) + e_i$ принадлежит U_i , и для каждого j , $1 \leq j \leq n$, $j \neq i$ функция $u_j = H_j(m, y)$ принадлежит U_j . Более того, $y_j = P_j(m_j, u_j)$ для каждого j , $1 \leq j \leq n$. Следовательно,

$$e_i = u_i - H_i(m, y) = u_i - H_i\left(m, \sum_{j=1}^n \mathcal{F}_j(m_j, u_j)\right) = u_i - \bar{K}_i(m, u),$$

где $u = (u_1, \dots, u_n)$. Этого достаточно для доказательства предположения 2. Таким образом, следствие 6.10 доказано.

Рассмотрим теперь случай, когда модификации определены согласно (6.42). Непосредственно применяя следствие 6.7, приходим к следующему результату, не требующему доказательства.

Следствие 6.10

Предположим, что для каждого i , $1 \leq i \leq n$, и β из $\mathcal{B}$ локальные функции качества $G_{i\beta}$ рассматриваемой двухуровневой системы имеют вид

$$G_{i\beta}(m_i, u_i, y_i) = G_i(m_i, u_i, y_i) + \langle \beta_i, u_i \rangle - \sum_{j=1}^n \langle \beta_j, H_{ji}(m_i, Pm_i) \rangle. \quad (6.65)$$

Предположим, что для каждого i , $1 \leq i \leq n$, функции G_i ограничены в соответствии со следствием 6.9, а U_i — такое замкнутое, ограниченное, выпуклое подмножество $\mathcal{Y}_i$, что для некоторого $\varepsilon > 0$ оно содержит все $u_i = H_i(m, Pm) + e_i$ для всех m из M и e_i из $\mathcal{Y}_i$ с нормой $\|e_i\| \leq \varepsilon$. Тогда система координируема с помощью принципа согласования взаимодействий.

Рассмотрим теперь квадратичный случай. Для каждого i , $1 \leq i \leq n$, предположим, что подынтегральные выражения в функциях качества (6.62) являются квадратичными формами:

$$\begin{aligned} \theta_i(t, m_i(t), u_i(t), y_i(t)) = & m_i(t)^T Q_{iM} m_i(t) + \\ & + [u_i(t) - a_i(t)]^T Q_{iU} [u_i(t) - a_i(t)] + \\ & + [y_i(t) - b_i(t)]^T Q_{iY} [y_i(t) - b_i(t)], \end{aligned}$$

где матрицы Q_{iM} , Q_{iU} и Q_{iY} — симметрические, по крайней мере положительно полуопределенные и, возможно, зависящие от времени, но ограниченные на $\mathcal{T}$, а a_i и b_i — известные и принадлежащие $\mathcal{Y}_i$ функции. Следовательно, локальные функции качества G_i могут быть выражены в виде

$$\begin{aligned} G_i(m_i, u_i, y_i) = & \langle m_i, Q_{iM} m_i \rangle + \\ & + \langle u_i - a_i, Q_{iU} (u_i - a_i) \rangle + \langle y_i - b_i, Q_{iY} (y_i - b_i) \rangle. \end{aligned} \quad (6.66)$$

Вследствие сделанных предположений о матрицах Q_{iM} , Q_{iU} и Q_{iY} области значений локальных функций качества ограничены снизу значением 0, в то время как $G_i(m_i, u_i, y_i(m_i, u_i))$ ограничены сверху на множестве $M_i \times U_i$, если U_i ограничены. Более того, θ_i выпуклы для каждого t в $\mathcal{T}$, а G_i непрерывны. Следовательно, в этом случае применимы полученные выше результаты

Следствие 6.11

Предположим, что для каждого i , $1 \leq i \leq n$, локальные функции качества задаются выражением (6.66), а модификации определяются выражением (6.63) или (6.65), причем множество связующих входов выбрано так, как указано в следствиях 6.9 или 6.10 соответственно. Тогда рассматриваемая двухуровневая система координируема с помощью принципа согласования взаимодействий.

Линейные операторы взаимодействия могут использоваться для получения оптимального координирующего сигнала, если модификации задаются выражением (6.65), а в некоторых случаях также, если они задаются выражением (6.63). Рассмотрим следующие три случая.

1. Если модификации определяются выражением (6.65), то, применяя теорему 6.8, можно найти оптимальный координирующий сигнал $\hat{\beta}$, используя линейные операторы оценки эффектив-

ности связующих входов Δ'_i в глобально оптимальной точке $\hat{m}$; для каждого i , $1 \leq i \leq n$, и всех u_i из U_i имеем

$$\langle \hat{\beta}_i, u_i \rangle = -\Delta_i(\hat{m}) u_i.$$

Для вычисления $\Delta'_i(\hat{m})$, $1 \leq i \leq n$, найдем производные

$$(1) G'_{iU} = \langle 2Q_{iU}(\hat{u}_i - a_i), u_i \rangle, \text{ где } \hat{u}_i = H_i(\hat{m}, P\hat{m});$$

$$(2) G'_{iY} = \langle 2Q_{iY}(\hat{y}_i - b_i), y_i \rangle, \text{ где } \hat{y}_i = \Pi_{iY} P\hat{m};$$

$$(3) \mathcal{F}'_{iU} = \Phi_i, \text{ где } (\Phi_i x)(t) = \int_0^t \Phi_i(t, \sigma) x(\sigma) d\sigma \text{ для } x \in$$

Найдем, далее, $g'_{iU} = G'_{iU} + G'_{iY} P'_{iU}$ и, подставляя эти величины в выражение (6.28), получим для каждого i , $1 \leq i \leq n$,

$$\Delta'_i(\hat{m}) u_i = 2 \langle Q_{iU}(\hat{u}_i - a_i), u_i \rangle + 2 \langle Q_{iY}(\hat{y}_i - b_i), \Phi_i u_i \rangle.$$

Пусть Φ_i^* — оператор, сопряженный с Φ_i , т. е. для x из $\mathcal{Y}_i$

$$(\Phi_i^* x)(t) = \int_t^1 \Phi_i^T(\sigma, t) x(\sigma) d\sigma,$$

и пусть

$$\hat{\beta}_i = -2 [Q_{iU}(\hat{u}_i - a_i) + \Phi_i^* Q_{iY}(\hat{y}_i - b_i)]. \quad (6.67)$$

Тогда $\hat{\beta} = \sum_i \hat{\beta}_i$ является оптимальным координирующим сигналом.

2. Если модификации локальных функций качества определяются выражением (6.63) и связующие входы зависят в явном виде только от нелокальных управляющих воздействий, то $A_{ij} = 0$ для всех $i \neq j$ и, следовательно,

$$u_i = H_i(m, y) = \sum_{j \neq i} B_{ij} m_j.$$

В этом случае, если $\hat{\beta}_i$ для каждого i , $1 \leq i \leq n$, определяется выражением (6.67), то $\hat{\beta} = \sum_i \hat{\beta}_i$ является оптимальным координирующим входом.

3. Если модификации по-прежнему задаются выражением (6.63), а связующие входы зависят в явном виде только от совокупных нелокальных выходов всей системы, то $B_{ij} = 0$ для всех $i \neq j$ и, следовательно,

$$u_i = H_i(m, y) = \sum_{j \neq i} A_{ij} y_j.$$

Чтобы получить явное выражение для оптимального координирующего сигнала, мы для каждого i , $1 \leq i \leq n$, предполагаем существование такой матрицы, обратной к $(B_{ii}^T B_{ii})$, что

$$Q_i \equiv (B_{ii}^T B_{ii})^{-1}. \quad (6.68)$$

Затем мы рассматриваем локальные выходы y_i в качестве управляющих воздействий, а m_i — в качестве локальных выходов, т. е.

$$m_i = Q_i B_{ii}^T (D - A_{ii}) y_i - Q_i B_{ii}^T u_i, \quad (6.69)$$

где D — оператор дифференцирования. Используя эту инверсию, получим, что оптимальным координирующим сигналом является $\hat{\beta} = \sum_i \hat{\beta}_i$, где

$$\langle \hat{\beta}_i, u_i \rangle = -\bar{\Delta}'_i(\hat{y}) u_i \quad (6.70)$$

для всех u_i из U_i .

Чтобы найти $\bar{\Delta}'_i(\hat{y})$, $1 \leq i \leq n$, определим функционал $\bar{g}_i$ выражением

$$\bar{g}_i(u_i, y_i) = G_i(m_i, u_i, y_i),$$

где m_i определяется из соотношения (6.69). Тогда из (6.28) следует, что

$$\bar{\Delta}'_i(\hat{y}) = \bar{g}'_i u, \quad (6.71)$$

где

$$\bar{g}'_i u = \langle 2Q_{iU}(\hat{u}_i - a_i) \rangle - \langle 2B_{ii}Q_i Q_{iM} \rangle. \quad (6.72)$$

Далее из равенств (6.70) — (6.72) получаем

$$\bar{\beta}_i = 2 [B_{ii} (B_{ii}^T B_{ii})^{-1} Q_{iM} \hat{m} - Q_{iU}(\hat{u}_i - a_i)]. \quad (6.73)$$

Координирующий сигнал $\hat{\beta} = \sum_i \hat{\beta}_i$ и является оптимальным координирующим сигналом.

Нелинейные модификации локальных функций качества

В предыдущих разделах была развита несколько более детализированная теория, относящаяся к координируемости на основе принципа согласования взаимодействий при использовании линейных модификаций локальных функций качества. Общие результаты, полученные ранее, указывают на очень широкую применимость этого принципа, так что следовало бы ожидать, что система может быть скоординирована и при использовании нелинейных модификаций. Тот факт, что система может быть скоординирована с помощью нелинейных модификаций, интересен не только как другой путь координации системы, но и как способ нахождения оптимальной координации с помощью итераций или путем применения двухуровневой обратной связи. Мы не дадим здесь полного рассмотрения этого вопроса, а укажем лишь, что при достаточно общих условиях нелинейные модификации, которые могли бы сделать систему координируемой, действительно возможны. Кроме того, существование линейных модификаций, которые делают систему координируемой, подразумевает при подходящих условиях и существование возможных нелинейных модификаций,

которые играют ту же роль. Это лишний раз подчеркивает важность проведенных в предыдущих разделах исследований.

Пример 6.1

Пусть $M_i = U_i = Y_i = R$, где R — множество действительных чисел, а $i = 1, 2$, и процесс в целом P представляется двумя подпроцессами P_1 и P_2 , описываемыми уравнениями

$$y_1 = 2m_1 + u_1 \equiv P_1(m_1, u_1), \quad y_2 = 2m_2 + u_2 \equiv P_2(m_2, u_2),$$

где $u_1 = m_2$, $u_2 = m_1$ являются связующими входами для этих подпроцессов. Предположим, что глобальная функция качества G определена на $M \times Y$ с помощью выражения

$$G(m, y) = m_1^2 + m_2^2 + (y_1 - 1)^2 + (y_2 - 2)^2.$$

В этом случае управляющее воздействие $\hat{m} = (1/5, 17/10)$ является глобально оптимальным.

Система будет координируемой, если мы воспользуемся принципом согласования взаимодействий и произведем декомпозицию глобальной функции качества относительно локальных переменных, а затем применим линейную модификацию с нулевой суммой. Положим $\mathcal{R} = R^2$, и пусть для каждого β из $\mathcal{R}$ локальными функциями качества $G_{i\beta}$ являются

$$G_{1\beta}(m_1, y_1, u_1) = m_1^2 + (y_1 - 1)^2 + \beta_1 u_1 - \beta_2 m_2,$$

$$G_{2\beta}(m_2, y_2, u_2) = m_2^2 + (y_2 - 2)^2 + \beta_2 u_2 - \beta_1 m_2.$$

Из проведенного в предыдущем разделе анализа следует, что система координируема с помощью принципа согласования взаимодействий; $\hat{\beta} = (-1/5, 4/4)$ — оптимальный координирующий сигнал, а соответствующими локальными оптимальными решениями являются пары $(m_1^\beta, u_1^\beta) = (1/5, 7/10)$ и $(m_2^\beta, u_2^\beta) = (7/10, 1/5)$.

Система может быть сделана координируемой за счет введения нелинейных модификаций самого разнообразного типа, применяемых к одним и тем же исходным (еще не модифицированным) локальным функциям качества ($G_{i\beta}$ для $\beta = 0$). Например, если модификации локальных функций качества выбираются для каждого β из $\mathcal{R}$ в виде

$$\tilde{G}_{1\beta}(m_1, y_1, u_1) = m_1^2 + (y_1 - 1)^2 + \beta_1 \sin \frac{1}{2} u_1 - \beta_2 \sin \frac{1}{8} m_1,$$

$$\tilde{G}_{2\beta}(m_2, y_2, u_2) = m_2^2 + (y_2 - 2)^2 + \beta_2 \sin \frac{1}{2} u_2 - \beta_1 \sin \frac{1}{2} m_2,$$

то система координируема на основе принципа согласования взаимодействий; действительно, оптимальным координирующим сигналом $\hat{\beta}$ является пара

$$\hat{\beta}_1 = -\frac{2}{5 \cos \frac{7}{20}}, \quad \hat{\beta}_2 = \frac{32}{5 \cos \frac{1}{40}}$$

приводящая, конечно, к тем же оптимальным локальным решениям, что и в линейном случае. Система будет координируемой при использовании также таких нелинейных модификаций, как

$$\mu_{1\beta}(m_1, u_1) = \beta_1 u_1^2 - \beta_2 m_1^2 \quad \text{и} \quad \mu_{2\beta}(m_2, u_2) = \beta_2 u_2^2 - \beta_1 m_2^2.$$

Мы не будем подробно останавливаться на этом, так как далее обратимся к более общему случаю.

Предположим, что глобальная функция качества аддитивна и модификации локальных функций качества являются линейными с нулевой суммой; тогда

$$g_{\mathcal{B}}(\beta, m, u) = \sum_{i=1}^n g_i(m_i, u_i) + \sum_{i=1}^n \beta_i (u_i - K_i(m, u)).$$

Предположим, что U_i , $1 \leq i \leq n$, — выпуклое подмножество банахова пространства $\bar{U}_i$, а также заданы отображения $f_i: \bar{U}_i \rightarrow \bar{U}_i$. Функции f_i не предполагаются линейными. Пусть

$$\tilde{g}_{\mathcal{B}}(\beta, m, u) = \sum_{i=1}^n g_i(m_i, u_i) + \sum_{i=1}^n \beta_i (f_i(u_i) - f_i(K_i(m, u))).$$

Тогда имеет место следующая лемма.

Лемма

Пусть для заданного β из $\mathcal{B}$ пара (m^β, u^β) минимизирует $g_{\mathcal{B}}(\beta, m, u)$ на $M \times U$. Предположим, что M и U выпуклы и для каждого i , $1 \leq i \leq n$, функции $g_i, f_i K_i$ дифференцируемы по Фреппе по каждому аргументу при $m_i = m_i^\beta$ и $u_i = u_i^\beta$. Кроме того, предположим, что производная Фреппе $f'_i(u_i^\beta)$ функции f_i^β в точке u_i имеет непрерывное обратное (левое) преобразование T_i , т. е. $T_i f'_i(u_i^\beta)$ является тождественным преобразованием.

Тогда, если $\tilde{g}_\beta$ выпукла на (m, u) , существует такое $\tilde{\beta}$ в $\mathcal{B}$

$$\tilde{\beta} = (\beta_1 T_1, \dots, \beta_n T_n),$$

что пара (m^β, u^β) минимизирует $\tilde{g}_{\mathcal{B}}(\tilde{\beta}, m, u)$ на $M \times U$.

Доказательство. Так как (m^β, u^β) минимизирует $g_{\mathcal{B}}(\beta, m, u)$ на $M \times U$, то для любой точки $(m^\beta + m, u^\beta + u)$ в $M \times U$ справедливы неравенства

$$g'_{\mathcal{B}M}(\beta, m^\beta, u^\beta) m \geq 0, \quad g'_{\mathcal{B}U}(\beta, m^\beta, u^\beta) u \geq 0.$$

Если $\tilde{g}_{\mathcal{B}}$ выпукла, то

$$\begin{aligned} \tilde{g}_{\mathcal{B}}(\tilde{\beta}, m^\beta + m, u^\beta + u) - \tilde{g}_{\mathcal{B}}(\tilde{\beta}, m^\beta, u^\beta) &\geq \\ &\geq \tilde{g}'_{\mathcal{B}M}(\tilde{\beta}, m^\beta, u^\beta) m + \tilde{g}'_{\mathcal{B}U}(\tilde{\beta}, m^\beta, u^\beta) u, \end{aligned}$$

где

$$\tilde{g}'_{\mathcal{B}M}(\tilde{\beta}, m^{\beta}, u^{\beta}) m = \sum_{i=1}^n [g'_{iM}(m_i^{\beta}, u_i^{\beta}) - \tilde{\beta}_i f'_i(u_i^{\beta}) \bar{K}'_{iM}(m^{\beta}, u^{\beta}) m],$$

$$\tilde{g}'_{\mathcal{B}U}(\tilde{\beta}, m^{\beta}, u^{\beta}) u = \sum_{i=1}^n [g'_{iU}(m_i^{\beta}, u_i^{\beta}) + \tilde{\beta}_i f'_i(u_i^{\beta}) \{u_i - \bar{K}'_{iU}(m^{\beta}, u^{\beta}) u\}].$$

Следовательно, если $\tilde{\beta} = (\beta_1 T_1, \dots, \beta_n T_n)$, то

$$\tilde{g}'_{\mathcal{B}M}(\tilde{\beta}, m^{\beta}, u^{\beta}) m = g'_{\mathcal{B}M}(\beta, m^{\beta}, u^{\beta}) m \geq 0,$$

$$\tilde{g}'_{\mathcal{B}U}(\tilde{\beta}, m^{\beta}, u^{\beta}) u = g'_{\mathcal{B}U}(\beta, m^{\beta}, u^{\beta}) u \geq 0.$$

Этим доказательство завершается.

Если $\tilde{g}_{\mathcal{B}}$ невыпукла, то пара (m^{β}, u^{β}) тем не менее удовлетворяет необходимому условию — минимизации $\tilde{g}_{\mathcal{B}}(\tilde{\beta}, m, u)$, когда $\tilde{\beta} = (\beta_1 T_1, \dots, \beta_n T_n)$.

Теперь мы должны показать, что при некоторых условиях функция $f_i(H_i(m, y))$ может быть сепарабельной, так что могут быть определены модификации локальных функций качества. (Заметим, что функция $\tilde{g}_{\mathcal{B}}$ в явной форме не связана с локальными функциями качества.) Чтобы показать это, докажем следующее утверждение.

Лемма

Пусть $\bar{U}_i = \bar{U}_{i1} \times \dots \times \bar{U}_{in}$ и связующие функции подпроцессов $H_i, 1 \leq i \leq n$, удовлетворяют соотношению (6.40а)

$$H_i(m, y) = \sum_{j=1}^n H_{ij}(m_j, y);$$

при этом $H_{ij}(m_j, y_j)$ принадлежат $\mathcal{U}_{ij} = \{u_i \in \bar{U}_i: u_{ik} = 0, k \neq j\}$.

Тогда если функции $f_i: \bar{U}_i \rightarrow \bar{U}_i$ таковы, что $f_i(u_i) = \sum_{j=1}^n f_{ij}(u_{ij})$,

то

$$f_i(H_i(m, y)) = \sum_{j=1}^n f_{ij}(H_{ij}(m_j, y_j)).$$

Доказательство. Справедливость леммы непосредственно вытекает из того факта, что $f_i(u_i) = \sum_{j=1}^n f_{ij}(u_{ij})$ и $H_i(m, y) =$

$$= \sum_{j=1}^n H_{ij}(m_j, y_j).$$

Из приведенных двух лемм немедленно следует

Теорема 6.12

Предположим, что условия этих двух лемм выполняются. Тогда если при использовании линейной модификации с нулевой суммой система координируема с помощью принципа согласования взаимодействий (с оптимальным координирующим сигналом $\hat{\beta}$), то система с нелинейной аддитивной модификацией

$$\mu_{i\beta}(m_i, u_i, y_i) = \beta_i f_i(u_i) - \sum_{j=1}^n \beta_j f_{ji}(H_{ji}(m_i, y_i))$$

также координируема с помощью этого принципа (с оптимальным координирующим сигналом $\hat{\beta} = (\hat{\beta}_1 T_1, \dots, \hat{\beta}_n T_n)$).

Даже если функция $\tilde{g}_{\mathcal{B}}$ не является выпуклой, пара (m^{β}, u^{β}) удовлетворяет необходимому условию — минимизации $\tilde{g}_{\mathcal{B}}$. Следовательно, если мы только докажем, что (m^{β}, u^{β}) является (единственной) парой, которая удовлетворяет необходимому условию минимума, и минимум $\tilde{g}_{\mathcal{B}}$ существует, то сформулированная теорема по-прежнему остается в силе. Таким путем практически можно получить много других случаев нелинейных модификаций.

4. ПРИМЕНЕНИЕ ПРИНЦИПА ПРОГНОЗИРОВАНИЯ

Теперь мы можем вывести более детализированные условия, чем указанные в гл. 5, для применимости принципа прогнозирования взаимодействий и координируемости на его основе. В этом разделе мы рассмотрим два случая: а) когда координация базируется только на прогнозировании связующих входов и б) когда координация основывается на прогнозировании связующих входов и используется координация целей. Наиболее вероятно, что принцип прогнозирования взаимодействий применим во втором случае, когда координация целей осуществляется в сочетании с прогнозированием.

Основные положения

Предположим, что для рассматриваемой двухуровневой системы заданы два множества $\mathcal{A}$ и $\mathcal{B}$, такие, что $K(M) = \mathcal{A}$ и каждое β из $\mathcal{B}$ определяет локальные целевые функции $g_{i\beta}$, $1 \leq i \leq n$, которые мы для удобства первоначально будем считать определенными на множестве $\bar{\mathcal{M}}_i \times \mathcal{A}$, а не на $\bar{\mathcal{M}}_i \times \bar{U}_i$. Для каждого координирующего сигнала, представляющего собой пару $\gamma = (\alpha, \beta)$ из $\mathcal{A} \times \mathcal{B}$, i -я локальная задача, $1 \leq i \leq n$, заключается в минимизации $g_{i\beta}(m_i, \alpha)$ на множестве $\mathcal{M}_i$. Обозначим через m_i^{γ} оптимальное решение этой задачи, а через g — глобальную целевую функцию, определенную на $\mathcal{M}$ с помощью глобальной функции

качества G . Как и прежде, глобальная задача состоит в минимизации g на M .

В фиксированной паре $\gamma = (\alpha, \beta)$ из $\mathcal{A} \times \mathcal{B}$ мы будем прогнозировать α называть *правильным для данного β* , если $\alpha = K(m^\gamma)$, где $m^\gamma = m_1^\gamma \times \dots \times m_n^\gamma$. Далее, если мы собираемся применять принцип прогнозирования взаимодействий, мы должны исключить те пары $\gamma = (\alpha, \beta)$, для которых прогноз α правилен, но управляющее воздействие m^γ не является глобально оптимальным.

В гл. 5 было отмечено, что координация целей в некотором смысле соответствует прогнозированию значений связующих входов. Удобным средством для достижения желаемого соответствия является отображение $\eta: M \rightarrow \mathcal{B}$. Полезность такого отображения следует из того факта, что решение задачи координации существенно упрощается при использовании прогнозирования значений связующих входов, особенно когда η выбирается в соответствии с декомпозицией K , $\eta(m) = \theta(K(m))$ и предполагается также осуществление координации целей.

Существуют два ключевых результата, служащих отправной точкой при анализе координируемости с помощью принципа прогнозирования взаимодействий.

1. Предположим, что функция взаимодействия подпроцессов K в пределах множества M осуществляет взаимно однозначное отображение. Тогда для существования отображения $\eta: M \rightarrow \mathcal{B}$, при использовании которого принцип прогнозирования применим, необходимо и достаточно, чтобы всякое управляющее воздействие $\tilde{m}$ из M , не являющееся глобально оптимальным, удовлетворяло строгому неравенству

$$\inf_{M_i} g_{i\beta}(m_i, K_i(\tilde{m})) < g_{i\beta}(\tilde{m}_i, K_i(\tilde{m}))$$

для каких-то β из $\mathcal{B}$ и каких-то i , $1 \leq i \leq n$ (теорема 5.31).

2. Допустим, что функция взаимодействия подпроцессов K в пределах множества M осуществляет взаимно однозначное отображение и, кроме того, существует управляющее воздействие, глобально оптимальное с точки зрения глобальной функции качества. Тогда существует такое прогнозное значение α из $\mathcal{A}$, что для любого β из $\mathcal{B}$ из $\mathcal{B}$ управляющее воздействие m^γ , если оно существует для $\gamma = (\alpha, \beta)$, является глобально оптимальным, когда $\alpha = K(m^\gamma)$. Таким прогнозным значением является $\hat{\alpha} = K(\hat{m})$, где $\hat{m}$ — глобально оптимальное управляющее воздействие.

Хотя эти два результата, если не принимать во внимание требования взаимной однозначности отображения K в пределах множества M , являются не особенно сильными, они обеспечивают базу для нашего анализа. Чтобы показать применимость принципа прогнозирования, необходимо и достаточно показать, что правильный прогноз может иметь место только тогда, когда пред-

сказанные значения связующих входов соответствуют глобально оптимальному управляющему воздействию. Однако, если глобально оптимальные управляющие воздействия неизвестны, нельзя ответить и на вопрос, соответствует ли данный прогноз такому управляющему воздействию. Выбор координирующих цели сигналов должен производиться так, чтобы предсказываемые значения связующих входов могли оказаться правильными только в том случае, если они соответствуют глобально оптимальному управляющему воздействию. Если теперь функция K не приводит к взаимно однозначному отображению M , то результат 1 дает только необходимое условие существования искомого отображения η , а результат 2 и вовсе перестает быть справедливым; в этом случае имеется возможность того, что прогноз окажется правильным и будет соответствовать глобально оптимальному управляющему воздействию, и тем не менее управляющее воздействие, полученное в результате решения локальных задач, уже не обязательно будет глобально оптимальным.

Допустим теперь, что функции взаимодействия подпроцессов K является взаимно однозначной, а локальные функции качества G_i , $1 \leq i \leq n$, образуются из заданной глобальной функции качества G путем «выделения» в ней локальных переменных, относящихся к i -му подпроцессу. А именно, пусть для каждого i , $1 \leq i \leq n$, функции G_i определены на $\bar{\mathcal{M}}_i \times \mathcal{Y}_i \times \mathcal{A}$ с помощью выражений

$$G_i(m_i, \alpha, y_i) = G(m_i + (I - \Pi_{iM})m^\alpha, y_i + (I - \Pi_{iY})y^\alpha), \quad (6.74)$$

где $m^\alpha = K^{-1}(\alpha)$ и $y^\alpha = P(m^\alpha)$, и пусть g_i суть функционалы

$$g_i(m_i, \alpha) = G_i(m_i, \alpha, \mathcal{F}_i(m_i, \alpha_i)) \quad (6.75)$$

на $\bar{\mathcal{M}} \times \mathcal{A}$. В следующей теореме формулируется первый вывод, касающийся применимости принципа прогнозирования.

Теорема 6.13

Предположим, что локальные функции качества рассматриваемой двухуровневой системы описываются уравнением (6.74), множество M выпукло, а функции g_i , $1 \leq i \leq n$, являются дифференцируемыми и выпуклыми по m_i для каждого α в $\mathcal{A}$. Тогда для любого α из $\mathcal{A}$

$$\max_{\bar{\mathcal{M}}_i^*} \min_{\mathcal{M}_i} [g_i(m_i, \alpha) + \beta_i(m_i^\alpha - m_i)] = g(m^\alpha)$$

и max-min реализуется на $(\beta_i^\alpha, m_i^\alpha)$, где $\beta_i^\alpha = g'_{iM}(m_i^\alpha, \alpha)$ — производная Фреше от g_i по m_i в точке (m_i^α, α) .

Доказательство. Пусть α — произвольный элемент $\mathcal{A}$. Так как $\alpha = K(m^\alpha)$, то из выражений (6.74) и (6.75) получаем,

что $g_i(m_i^\alpha, \alpha) = g(m^\alpha)$. Следовательно, для любого β_i в $\bar{\mathcal{M}}_i^*$

$$\inf_{\mathcal{M}_i} [g_i(m_i, \alpha) + \beta_i(m_i^\alpha - m_i)] \leq g(m^\alpha).$$

Далее, так как функция g_i дифференцируема и выпукла по m_i , то для всех m_i из $\bar{\mathcal{M}}_i$

$$\beta_i^\alpha(m_i - m_i^\alpha) \leq g_i(m_i, \alpha) - g_i(m_i^\alpha, \alpha),$$

где $\beta_i^\alpha = g'_{im}(m_i^\alpha, \alpha)$. Следовательно,

$$\inf_{\mathcal{M}_i} [g_i(m_i, \alpha) + \beta_i^\alpha(m_i^\alpha - m_i)] = g_i(m_i^\alpha, \alpha) = g(m^\alpha),$$

и доказательство завершено.

Предположим, что локальные функции качества рассматриваемой двухуровневой системы определяются выражением (6.74), $\mathcal{B} = \bar{\mathcal{M}}^*$ и для каждого β в $\mathcal{B}$ и i , $1 \leq i \leq n$, $g_{i\beta}$ — функционал на $\bar{\mathcal{M}}_i \times \mathcal{A}$

$$g_{i\beta}(m_i, \alpha) = g_i(m_i, \alpha) + \beta_i(m_i), \quad (6.76)$$

где β_i — проекция β на $\bar{\mathcal{M}}_i$. Тогда теорема 6.13 утверждает, что для любого прогнозного значения α в $\mathcal{A}$, которое не соответствует глобально оптимальному управляющему воздействию, соответствующее β из $\mathcal{B}$ должно быть таким, что

$$\beta_i \neq \beta_i^\alpha = g'_{im}(m_i^\alpha, \alpha) \text{ для некоторого } i, 1 \leq i \leq n; \quad (6.77)$$

в противном случае α было бы правильным прогнозом, но принцип прогнозирования был бы неприменим. Однако условие (6.77) является только необходимым, но не достаточным для применимости принципа.

Далее, i -й линейный полный оператор оценки косвенного эффекта управляющих воздействий Γ'_i , $1 \leq i \leq n$, в точке $\tilde{m} \in \bar{M}$ является ограниченным линейным функционалом на $\bar{M}$, определенным выражением (6.30):

$$\Gamma'_i(\tilde{m}) m_i = g'(\tilde{m}) m_i - g'_{im}(\tilde{m}_i, K(\tilde{m})) m_i. \quad (6.78)$$

Предполагая существование в точке $\tilde{m}$ таких Γ'_i для каждого i , $1 \leq i \leq n$, введем линейный функционал $\Gamma'(\tilde{m})$, ограниченный на $\bar{M}$:

$$\Gamma'(\tilde{m}) m = \sum_{i=1}^n \Gamma'_i(\tilde{m}) m_i.$$

Предположив дифференцируемость целевых функций g и g_i , $1 \leq i \leq n$, можно использовать эти линейные полные операторы оценки косвенного эффекта для выбора подходящих β в $\mathcal{B}$, соответствующих заданному α из $\mathcal{A}$. Прежде всего допустим, что функция взаимодействия подпроцессов K является взаимно однозначной (позже мы ослабим это условие).

Теорема 6.14

Предположим, что локальные функции качества рассматриваемой двухуровневой системы определены выражением (6.74), а модификации осуществляются в соответствии с выражением (6.76). Предположим также, что множество M выпуклое, g — выпуклая и дифференцируемая по Фреше функция, K — взаимно однозначная функция, а функции g_i для каждого i , $1 \leq i \leq n$, и α в $\mathcal{A}$ дифференцируемы по Фреше в точке m_i . Пусть $\eta: M \rightarrow \mathcal{B}$ определяется для всех m из M следующим образом:

$$\eta(m) = \Gamma'(m). \quad (6.79)$$

Тогда принцип прогнозирования взаимодействий применим при использовании отображения η : для всех $\gamma = (\alpha, \beta)$ из $\mathcal{A} \times \mathcal{B}$ управляющее воздействие m^γ будет глобально оптимальным, если $\alpha = K(m^\gamma)$ и $\beta = \eta(m^\gamma)$.

Доказательство. Выберем произвольное $\gamma = (\alpha, \beta)$ в $\mathcal{A} \times \mathcal{B}$. Нам необходимо показать, что m^γ — глобально оптимальное управляющее воздействие, если $\alpha = K(m^\gamma)$ и $\beta = \eta(m^\gamma)$. Предположим, что $\alpha = K(m^0)$ и $\beta = \eta(m^\gamma)$; так как функция K взаимно однозначна, то $m^\gamma = m^\alpha$ и $\beta = \Gamma'(m^\alpha)$. Далее, для всех i , $1 \leq i \leq n$,

$$g_i(m_i^\alpha + m_i, \alpha) + \Gamma'_i(m^\alpha)(m_i^\alpha + m_i) \geq g_i(m_i^\alpha, \alpha) + \Gamma'_i(m^\alpha)m_i^\alpha$$

и, следовательно,

$$g_i(m_i^\alpha + m_i, \alpha) - g_i(m_i^\alpha, \alpha) + \Gamma'_i(m^\alpha)m_i \geq 0,$$

когда $m_i^\alpha + m_i$ принадлежит $\mathcal{M}_i$. С другой стороны, так как g_i дифференцируема, то

$$g_i(m_i^\alpha + \lambda m_i, \alpha) - g_i(m_i^\alpha, \alpha) = g'_{iM}(m_i^\alpha, \alpha) \lambda m_i + o(\lambda)$$

и, следовательно,

$$[\Gamma'_i(m^\alpha) + g'_{iM}(m_i^\alpha, \alpha)] m_i \geq 0, \quad (6.80)$$

когда $m_i^\alpha + m_i$ принадлежит $\mathcal{M}_i$. Поскольку неравенство (6.80) справедливо для каждого i , $1 \leq i \leq n$, M — прямая сумма и имеет место (6.78), то справедливо следующее соотношение:

$$g'(m^\alpha)m = \sum_{i=1}^n [\Gamma'_i(m^\alpha) + g'_{iM}(m_i^\alpha, \alpha)] m_i \geq 0,$$

где $m^\alpha + m \in M$. Следовательно, m^α дает локальный минимум g . Но g — выпуклая функция и M — выпуклое множество, следовательно, m^α — глобально оптимальное управляющее воздействие. Это завершает доказательство.

Эта теорема в противоположность теореме 6.13 показывает, какими должны быть модификации, чтобы принцип прогнозирования взаимодействий был применим в случае взаимной однозначности K . Мы не предполагали существования глобально оптимального управляющего воздействия, так как это не требуется для доказательства применимости, однако существование такого воздействия обуславливается существованием правильного прогноза α из $\mathcal{A}$ при $\beta = \eta(m^\alpha)$.

Далее, можно показать, что если глобально оптимальное управляющее воздействие существует, то при выполнении определенных условий система будет координируема на основе принципа прогнозирования взаимодействий при использовании отображения η .

Теорема 6.15

Предположим, что для рассматриваемой двухуровневой системы выполняются все предположения теоремы 6.14; кроме того, допустим, что существует глобально оптимальное управляющее воздействие, а функции g_i для каждого i , $1 \leq i \leq n$, — выпуклы по m_i для всякого α из $\mathcal{A}$. Тогда система координируема на основе принципа прогнозирования взаимодействий при использовании отображения η , определяемого равенством (6.79): принцип применим при использовании отображения η и существует такое $\gamma = (\alpha, \beta)$ в $\mathcal{A} \times \mathcal{B}$, что $\alpha = K(m^\gamma)$ и $\beta = \eta(m^\gamma)$ для некоторого m^γ .

Доказательство. Применимость при использовании отображения η следует из теоремы 6.14, поэтому необходимо показать только существование такого $\gamma = (\alpha, \beta)$ в $\mathcal{A} \times \mathcal{B}$, что $\alpha = K(m^\gamma)$ и $\beta = \eta(m^\gamma)$ для некоторого m^γ . Пусть α из $\mathcal{A}$ такое, что $K^{-1}(\alpha) = m^\alpha = \hat{m}$, где $\hat{m}$ — глобально оптимальное управляющее воздействие. Так как функция g дифференцируема и множество M выпукло, то $g'(m^\alpha) m = g'(\hat{m}) m \geq 0$ для всех $m^\alpha + m$ в M . Пусть $\beta = \eta(m^\alpha)$, тогда, поскольку функция g_i , $1 \leq i \leq n$, дифференцируема в точке m_i , для каждого $m_i^\alpha + m_i$ из $\mathcal{M}_i$ справедливо следующее неравенство:

$$\begin{aligned} g_{i\beta}^{\alpha}(m_i^{\alpha}, \alpha) m_i &= [g'_{i\beta}(m_i^{\alpha}, \alpha) + \Gamma'_i(m^\alpha)] m_i = \\ &= g'_{i\beta}(m_i^{\alpha}, \alpha) m_i + g'(m^\alpha) m_i - g'_{i\beta}(m_i^{\alpha}, \alpha) m_i = g'(m^\alpha) m_i \geq 0. \end{aligned}$$

Поэтому для каждого i , $1 \leq i \leq n$, m_i^α дает локальный минимум $g_{i\beta}$. Так как функция g_i выпукла по m_i , то и $g_{i\beta}$ выпукла по m_i . Множество $\mathcal{M}_i$ также выпукло. Следовательно, для каждого i , $1 \leq i \leq n$, m_i^α минимизирует $g_{i\beta}$ на $\mathcal{M}_i$. Таким образом,

$$m^\gamma = m^\alpha; \alpha = K(m^\gamma) = K(m^\alpha); \beta = \eta(m^\gamma) = \eta(m^\alpha).$$

Это доказывает теорему.

Ослабим условие относительно взаимной однозначности K . А именно, предположим, что рассматриваемая двухуровневая система обладает аддитивной глобальной функцией качества. Пусть локальные целевые функции g_i , $1 \leq i \leq n$, определены на пространстве $\bar{\mathcal{M}}_i \times \bar{U}_i$:

$$g_i(m_i, u_i) = G_i(m_i, u_i, \mathcal{P}_i(m_i, u_i)), \quad (6.81)$$

где G_i — заданные локальные функции качества. Модификации локальных функций качества, соответствующие выражениям (6.75), будем выбирать такими, чтобы для каждого β в $\bar{M}^*$ и i , $1 \leq i \leq n$, выполнялось условие

$$g_{i\beta}(m_i, u_i) = g_i(m_i, u_i) + \beta_i m_i, \quad (6.82)$$

где β_i — проекция β на $\bar{\mathcal{M}}_i$.

Теорема 6.16

Предположим, что рассматриваемая двухуровневая система обладает аддитивной глобальной функцией качества и используются модификации (6.82). Допустим далее, что M — выпуклое множество, g — выпуклая, дифференцируемая по Фреше функция, а g_i , $1 \leq i \leq n$, — функции, определяемые выражением (6.81). Дифференцируемые по Фреше при каждом u_i по m_i . Тогда принцип прогнозирования взаимодействий применим при использовании отображения η , определяемого выражением (6.79).

Доказательство. Предположим, что существует такое m^γ , что $\alpha = K(m^\gamma)$ и $\beta = \eta(m^\gamma)$ для некоторой пары $\gamma = (\alpha, \beta)$ из $\mathcal{A} \times \mathcal{B}$. Тогда $\beta = \Gamma'(m^\gamma)$. Пусть i , $1 \leq i \leq n$, произвольно и $m_i^\gamma + m_i$ принадлежит $\mathcal{M}_i$; тогда

$$g_i(m_i^\gamma + m_i, \alpha_i) + \Gamma'_i(m_i^\gamma)(m^\gamma + m_i) \geq g_i(m_i^\gamma, \alpha_i) + \Gamma'_i(m^\gamma) m_i^\gamma$$

и, следовательно,

$$g_i(m_i^\gamma + m_i, \alpha_i) - g_i(m_i^\gamma, \alpha_i) + \Gamma'_i(m^\gamma) m_i \geq 0. \quad (6.83)$$

Так как g_i — дифференцируемая функция, то

$$g_i(m_i^\gamma + \lambda m_i, \alpha_i) - g_i(m_i^\gamma, \alpha_i) = g'_{iM}(m_i^\gamma, \alpha_i) \lambda m_i + \mathcal{O}(\lambda).$$

Но из (6.78) или (6.30) получаем, что

$$\Gamma'_i(m^\gamma) m_i = g'(m^\gamma) m_i - g'_{iM}(m_i^\gamma, \alpha_i) m_i. \quad (6.84)$$

Поэтому из (6.83) — (6.84) следует, что

$$g'(m^\gamma) \lambda m_i = [\Gamma'_i(m^\gamma) + g'_{iM}(m_i^\gamma, \alpha_i)] \lambda m_i \geq 0 \quad (6.85)$$

при $\lambda > 0$. Так как неравенство (6.85) справедливо для всех $1 \leq i \leq n$, и $m_i^\gamma + m_i \in \mathcal{M}_i$, множество M выпукло и является

прямой суммой, а функция g выпукла на этом множестве, то m^γ является глобально оптимальным управляющим воздействием. Это завершает доказательство.

Теорема 6.17

Предположим, что для рассматриваемой двухуровневой системы удовлетворяются все предположения теоремы 6.16, и, кроме того, существует глобально оптимальное управляющее воздействие, а локальные целевые функции g_i для всех i , $1 \leq i \leq n$, выпуклы по m_i для каждого u_i . Тогда система координируема с помощью принципа прогнозирования взаимодействий при использовании отображения η , определяемого выражением (6.79).

Доказательство. Применимость доказана в теореме 6.16. Предположим, что пара $\gamma = (\alpha, \beta)$ в $\mathcal{A} \times \mathcal{B}$ такова, что $\alpha = \hat{u} = K(\hat{m})$ и $\beta = \Gamma'(\hat{m})$, где $\hat{m}$ — глобально оптимальное управляющее воздействие. Покажем, что m^γ существует и $m^\gamma = \hat{m}$. Из выпуклости множества M и дифференцируемости функции g следует, что $g'(\hat{m})m \geq 0$ для всех $\hat{m} + m$ в M . Поэтому, поскольку функции g_i , $1 \leq i \leq n$, дифференцируемы, получаем с учетом (6.30), что

$$g'_{i\beta M}(\hat{m}_i, \hat{u}_i)m_i = g'_{iM}(\hat{m}_i, \hat{u}_i)m_i + \Gamma'_i(\hat{m})m_i = g'(\hat{m})m_i \geq 0$$

для всех $\hat{m}_i + m_i$ в $\mathcal{M}_i$. Следовательно, для каждого i , $1 \leq i \leq n$, $\hat{m}_i$ дает локальный минимум $g_{i\beta}(m_i, \hat{u}_i)$, а так как g_i — выпуклые функции и $\mathcal{M}_i$ — выпуклые множества (M выпукло), то этот локальный минимум соответствует абсолютному минимуму на $\mathcal{M}_i$; отсюда следует, что $\hat{m} = m^\gamma$ для некоторого m^γ . Этим доказательство завершается.

Сделаем теперь некоторые замечания относительно применения полученных выше результатов.

1. Модификации локальных функций качества являются линейными, но не с нулевой суммой в отличие от модификаций, рассматривавшихся нами ранее в связи с исследованием применимости принципов согласования (тогда мы, как правило, требовали, чтобы используемые модификации были линейными модификациями с нулевой суммой). Однако можно показать, что теоремы 6.16 и 6.17 остаются справедливыми и в том случае, когда модификации определяются выражением (6.42), а отображение $\eta: M \rightarrow \mathcal{B}$ вводится в виде

$$\eta(m) = (-\Delta'_1(m), \dots, -\Delta'_n(m)), \quad (6.86)$$

где $\Delta'_i(\hat{m})$, $1 \leq i \leq n$, — i -й линейный частичный оператор оценки косвенного эффекта управляющих воздействий, заданный на множестве $\bar{U}_i$ выражением (6.28). Это можно показать следующим

образом. Для каждого β в $\mathcal{B} = \bar{U}_1^* \times \dots \times \bar{U}_n^*$ и i , $1 \leq i \leq n$,

$$g_{i\beta}(m_i, u_i) = g_i(m_i, u_i) + \beta_i u_i - \sum_{j=1}^n \beta_j K_{ji}(m_i).$$

Далее, в точке $\hat{m} \in M$

$$g'_{i\beta M}(\hat{m}_i, \hat{u}_i) m_i = g'_{iM}(\hat{m}_i, \hat{u}_i) m_i - \sum_{j=1}^n \beta_j K'_{ji}(\hat{m}) m_i$$

и, если $\beta_j = -\Delta'_j(\hat{m})$ для всех j , $1 \leq j \leq n$, то из уравнения (6.30) следует, что

$$g'_{i\beta M}(\hat{m}_i, \hat{u}_i) m_i = g'_{iM}(\hat{m}_i, \hat{u}_i) m_i + \Gamma'_i(\hat{m}) m_i = g'(\hat{m}) m_i. \quad (6.87)$$

Следовательно, если модификации определяются выражением (6.42), а $\eta: M \rightarrow \mathcal{B}$ — выражением (6.86), то в силу равенства (6.87) справедливы теоремы 6.61 и 6.17.

2. Из теоремы 6.17 и предыдущего замечания вытекает, что если двухуровневая система с аддитивной глобальной функцией качества и линейными модификациями с нулевой суммой локальных функций качества [определяемыми выражением (6.42)] координируема на основе принципов согласования, то при надлежащих предположениях о дифференцируемости и выпуклости она будет также координируема и на основе принципа прогнозирования взаимодействий при использовании отображения, определяемого выражением (6.86).

Применение к линейным дифференциальным системам

Результаты, полученные в предыдущем разделе, можно использовать при исследовании систем, для которых подпроцессы описываются с помощью дифференциальных уравнений. В этом разделе, так же как и в предыдущем, мы будем рассматривать только линейно-выпуклые задачи, и притом наиболее классический — линейно-квадратичный случай.

Предположим, что процесс в целом P и подпроцессы $\mathcal{P}_i$, $1 \leq i \leq n$, определены на интервале $\mathcal{T} = [0, 1]$ с помощью линейных дифференциальных уравнений

$$\dot{y}(t) = Ay(t) + Bm(t), \quad y(0) = 0,$$

$$\dot{y}_i(t) = A_{ii}y_i(t) + B_{ii}m_i(t) + u_i(t), \quad y_i(0) = 0,$$

где, как и в предыдущем разделе,

$$u_i(t) = \sum_{j \neq i} [B_{ij}m_j(t) + A_{ij}y_j(t)] = H_i(m, y)(t).$$

Пусть локальные и глобальные функции качества задаются так же, как и ранее. Но в этом разделе необходимо сделать некоторые

предположения относительно процесса в целом P , подпроцессов $\mathcal{P}_i$ и функционалов θ_i , а именно:

1) M , а следовательно, и $\mathcal{M}_i$, $1 \leq i \leq n$, — выпуклые и слабо компактные множества;

2) $g(m) = G(m, P(m))$ выпуклая и дифференцируемая по Фреше на $\bar{M}$ функция;

3) $g_i(m_i, \alpha_i) = G_i(m_i, \alpha_i)$, $\mathcal{P}_i(m_i, \alpha_i)$, $1 \leq i \leq n$, и α в $\mathcal{A}$ — выпуклые, слабо полунепрерывные и дифференцируемые по Фреше на $\bar{\mathcal{M}}_i$ функции. Таким образом, условия применения теорем 6.16 и 6.17 выполнены.

Если $\mathcal{R} = \bar{M}$ и для всех β в $\mathcal{R}$ модифицированные локальные функции качества G_i , $1 \leq i \leq n$, определяются выражением

$$G_{i\beta}(m_i, u_i, y_i) = G_i(m_i, u_i, y_i) + \langle \beta_i, m_i \rangle,$$

где β_i — проекция β на $\bar{\mathcal{M}}_i$, то из теоремы 6.17 вытекает следствие 6.18.

Следствие 6.18

Двухуровневая система, описанная выше, координируема с помощью принципа прогнозирования взаимодействий при использовании отображения $\eta: \bar{M} \rightarrow \mathcal{R}$, определяемого уравнениями

$$\langle \eta_i(m), m_i \rangle = \Gamma'_i(\tilde{m}) m_i, \quad (6.88)$$

где $\eta_i(m)$ — проекция $\eta(m)$ на $\bar{\mathcal{M}}_i$, т. е. $\eta(m) = \sum_{i=1}^n \eta_i(m)$.

Естественно, что желательно получить явные выражения для η_i . Мы выведем их при некоторых упрощающих предположениях. Предположим, что подынтегральные выражения θ_i , $1 \leq i \leq n$, являются такими же квадратичными формами, как и в предыдущем разделе, но с нулевой матрицей Q_{iM} . Следовательно, функции G_i , $1 \leq i \leq n$, имеют вид

$$G_i(m_i, y_i) = \langle m_i, Q_{iM} m_i \rangle + \langle y_i - b_i, Q_{iY} (y_i - b_i) \rangle. \quad (6.89)$$

Пусть $Q_{iM} = \Pi_{iM} Q_{iM} = Q_{iM} \Pi_{iM}$ и $Q_{iY} = \Pi_{iY} Q_{iY} = Q_{iY} \Pi_{iY}$. Тогда глобальная функция качества G имеет вид

$$G(m, y) = \langle m, Q_M m \rangle + \langle y - b, Q_Y (y - b) \rangle,$$

где $Q_M = \sum_i Q_{iM}$, $Q_Y = \sum_i Q_{iY}$ и $b = b_1 + \dots + b_n$.

Чтобы найти $\eta_i(\tilde{m})$ для заданного $\tilde{m}$ в $\bar{M}$, необходимо знать производные g и g_i по m_i в точке $\hat{m}$. Представив процесс в целом P и подпроцессы $\mathcal{P}_i$ в виде

$$Pm = \Phi B m, \quad P_i(m_i, u_i) = \Phi_i(B_{i1} m_i + u_i),$$

получим, что

$$g'(\tilde{m}) m_i = 2 \langle Q_M \hat{m}, \Pi_{iM} m \rangle + 2 \langle Q_Y (\hat{y} - b), \Phi B \Pi_{iM} m \rangle$$

и

$$g_{iM}(\hat{m}_i, \hat{u}) m_i = 2 \langle Q_{iM} \hat{m}_i, m_i \rangle + 2 \langle Q_{iY} (\hat{y}_i - b_i), \Phi_i B_{ii} m_i \rangle,$$

где $\hat{y} = P\hat{m}$ и $\hat{u}_i = K\hat{m}$. Из (6.87), (6.30) и тождества

$$\langle Q_M \hat{m}, \Pi_{iM} m \rangle - \langle Q_{iM} \hat{m}_i, \hat{m}_i \rangle = 0$$

следует, что

$$\eta_i(\hat{m}) = 2\Pi_{iM} [B^* \Phi^* Q_Y (\hat{y} - b) - B_{ii}^* \Phi_i^* Q_{iY} (\hat{y}_i - b_i)],$$

где звездочка означает сопряженность. Замечая далее, что

$$Q_{iY} (\hat{y}_i - b_i) = Q_{iY} \Pi_{iY} (\hat{y}_i - b_i) = Q_{iY} (\hat{y} - b)$$

и

$$B_{ii}^* \Phi_i^* Q_{iY} = \Pi_{iM} B^* \Pi_{iY} \Phi^* \Pi_{iY} Q_Y,$$

можно упростить выражение для $\eta_i(\hat{m})$. А именно, имеем

$$\eta_i(\hat{m}) = 2\Pi_{iM} B^* [\Phi^* - \Pi_{iY} \Phi^* \Pi_{iY}] Q_Y (\hat{y} - b) \quad (6.90)$$

или

$$\eta_i(\hat{m})(t) = 2\Pi_{iM} B^T \int_1^1 [\Phi^T(\sigma, t) - \Phi_i^T(\sigma, t)] Q_Y [\hat{y}(\sigma) - b(\sigma)] d\sigma. \quad (6.91)$$

Дальнейший процесс нахождения $\eta_i(\hat{m})$ сводится к интегрированию при условии, что фундаментальная матрица $\Phi(t, \sigma)$ задается уравнением (6.60).

Если связующие входы в явном виде зависят только от нелокальных выходов процесса в целом, т. е. $B_{ij} = 0$ для $i \neq j$, то в качестве множества координирующих входов $\mathcal{R}$ можно выбрать множество Y и модифицировать локальные функции качества следующим образом: для всех β из $\mathcal{R}$ и i , $1 \leq i \leq n$,

$$G_{i\beta}(m_i, y_i) = G_i(m_i, y_i) + \langle \beta_i, \hat{y}_i \rangle, \quad (6.92)$$

где G_i определяется выражением (6.89), а β_i — проекция β на $\mathcal{Y}_i$. Если теперь для каждого i , $1 \leq i \leq n$, существует обратное преобразование (инверсия) $B_{ii}^T B_{ii}$

$$(B_{ii}^T B_{ii})^{-1} \equiv Q_i,$$

то можно не только показать координируемость на основе принципа прогнозирования взаимодействий при использовании некоторого отображения $\bar{\eta}: \bar{M} \rightarrow \mathcal{R}$, но также и получить точное выражение для $\bar{\eta}$.

Примем теперь y_i за управляющее воздействие, а m_i будем рассматривать как локальный выход, определяемый выражением (6.69). Такая инверсия позволяет получить следующий результат.

Следствие 6.19

Если связующие входы зависят в явном виде только от нелокальных выходов процесса в целом, модификации локальных функций качества введены с помощью выражения (6.92), а $Q_i = (B_{ii}^T B_{ii})^{-1}$ существуют для всех i , $1 \leq i \leq n$, то система координируема с помощью принципа прогнозирования взаимодействий при использовании отображения $\bar{\eta}$, определяемого уравнениями

$$(\bar{\eta}_i(\hat{m}), y_i) = \bar{\Gamma}_i(\hat{y}) y_i,$$

где $\hat{y} = P\hat{m}$, $\bar{\Gamma}_i$ — линейный полный оператор оценки косвенного эффекта для преобразованной системы, а $\bar{\eta}_i(\hat{m})$ — проекция $\bar{\eta}(\hat{m})$ на $\mathcal{Y}_i$.

Чтобы найти $\bar{\eta}_i(\hat{m})$ для заданного элемента m из $\bar{M}$, необходимо сначала вычислить величины $\bar{\Gamma}_i(\hat{y})$, $\hat{y} = P(\hat{m})$, определяемые выражениями

$$\bar{\Gamma}_i(\hat{y}) y_i = \sum_{j \neq i} \bar{\Delta}_j(\hat{y}) A_{ij} y_j,$$

где $\bar{\Delta}_j(\hat{y})$ — оператор оценки эффективности связующих входов, уже рассматривавшийся нами в предыдущем разделе [см. (6.71) — (6.73)] и равный

$$\bar{\Delta}_j(\hat{y}) u_j = -2 \langle B_{ij} Q_j Q_{jM} \hat{m}_j, u_j \rangle,$$

а $Q_{jU} = 0$. Поэтому

$$\begin{aligned} \bar{\Gamma}_i(\hat{y}) y_i &= -2 \sum_{j \neq i} \langle B_{ji} Q_j Q_{jM} \hat{m}_j, A_{ji} y_i \rangle = \\ &= -2 \sum_{j \neq i} \langle A_{ji}^* B_{jj} Q_j Q_{jM} \hat{m}_j, y_i \rangle. \end{aligned}$$

Следовательно,

$$\bar{\eta}_i(\hat{m}) = -2 \sum_{j \neq i} A_{ji}^* B_{jj} (B_{jj}^T B_{jj})^{-1} Q_{jM} m_j$$

или

$$\bar{\eta}_i(\hat{m})(t) = -2 \sum_{j \neq i} A_{ji}^T B_{jj} (B_{jj}^T B_{jj})^{-1} Q_{jM} m_j(t). \quad (6.93)$$

Отметим простоту полученного в этом случае выражения для $\bar{\eta}_i(\hat{m})$ по сравнению с выражением (6.90) или (6.91). По этой причине модификации, выраженные в виде (6.92), используются тогда, когда связующие входы зависят в явном виде только от нелокальных выходов процесса в целом и $(B_{ii}^T B_{ii})^{-1}$ существует.

5. ИТЕРАТИВНЫЕ ПРОЦЕССЫ КООРДИНАЦИИ

В гл. 5 были описаны некоторые основные стратегии нахождения оптимального координирующего сигнала с помощью итера-

тивного процесса, в котором участвуют как сам координатор, так и локальные решающие элементы. В ходе такого процесса координатор получает с помощью обратной связи информацию в виде пары $(u^v, K(m^v))$ из $\bar{U} \times \bar{U}$ и на основании предварительно сформированного координирующего сигнала γ и стратегии $s_0: \mathcal{C} \times \bar{U} \times \bar{U} \rightarrow \mathcal{C}$ улучшает координирование системы за счет подачи нового координирующего сигнала γ' . Так как и u^v и $u = K(m^v)$ формируются на основании полученного координирующего сигнала γ , то стратегия s_0 определяет преобразование $T: \mathcal{C} \rightarrow \mathcal{C}$

$$T(\gamma) = s_0(\gamma, u^v, K(m^v)).$$

Тогда итеративность заключается в многократном применении преобразования T , и, следовательно, успешность итеративного процесса зависит от свойств T , которые были описаны в гл. 5. Если итерации сходятся и стратегия s_0 основывается на таком принципе координации, при использовании которого система координируема, то процесс должен сходиться к оптимальному координирующему сигналу. Рассмотрим теперь вопрос о выборе стратегии s_0 , для которой итеративный процесс сходится.

Стратегия, основанная на принципе согласования взаимодействий

Рассмотрим случай, когда двухуровневая система координируема на основе принципа согласования взаимодействий. Координация в этом случае сводится к выбору такого координирующего цели сигнала β из $\mathcal{B}$, чтобы отображения s_0 и T являлись отображениями $s_0: \mathcal{B} \times \bar{U} \times \bar{U} \rightarrow \mathcal{B}$ и $T: \mathcal{B} \rightarrow \mathcal{B}$ соответственно.

Предположим, что стратегия s_0 базируется на принципе согласования взаимодействий, и, в частности, будем считать, что коррекции ранее поданного координирующего сигнала γ пропорциональны (в подходящем смысле) ошибке

$$e^\beta = u^\beta - K(m^\beta)$$

между желаемым связующим входом u^β и действительным связующим входом $u = K(m^\beta)$.

Предположим, что рассматриваемая система обладает аддитивной глобальной функцией качества и модификации локальных функций качества являются аддитивными с нулевой суммой. В частности, будем считать, что пространства $\bar{U}_i$, $1 \leq i \leq n$, — гильбертовы пространства, а множество $\mathcal{B} = \bar{U} = \bar{U}_1 \times \dots \times \bar{U}_n$ таково, что кажущаяся глобальная целевая функция $g_{\mathcal{B}}$ системы имеет вид

$$g_{\mathcal{B}}(\beta, m, u) = \sum_{i=1}^n g_i(m_i, u_i) + \sum_{i=1}^n \langle \beta_i, u_i - K_i(m) \rangle. \quad (6.94)$$

Пусть стратегия s_0 задается n отображениями $s_i: \mathcal{R} \times U_i \times \times U_i \rightarrow \mathcal{R}_i$

$$s_i(\beta, u_i, u_i^b) = \beta_i + \lambda_i(\beta) [u_i - u_i^b],$$

где $\lambda_i(\beta)$ — действительное число, возможно, зависящее от β . Тогда, многократное применение преобразования $T: \mathcal{R} \rightarrow \mathcal{R}_i$, представляющего собой n преобразований $T_i: \mathcal{R} \rightarrow \mathcal{R}_i$, где

$$T_i(\beta) = \beta_i + \lambda_i(\beta) [u_i^b - K_i(m^b)], \quad (6.95)$$

образует итеративный процесс. Если итеративный процесс сходится, то он сходится к точке, в которой связующие входы согласованы; следовательно, если принцип согласования взаимодействий применим, то процесс должен сходиться к оптимальному координирующему сигналу.

Рассмотрим итерационный процесс, определяемый выражением (6.95), и положим его в основу анализа точек накопления. В частности, покажем, что при подходящих условиях существуют такие $\lambda_i(\beta)$, $1 \leq i \leq n$, что последовательность, сформированная с помощью T , удовлетворяет условию согласования всякий раз, когда она имеет точку накопления. Затем покажем, при каких условиях последовательность, образованная с помощью T , имеет такую точку.

Предположим, что отображение $d: \mathcal{R} \rightarrow M \times U$ обобщенно характеризует деятельность локальных решающих элементов в целом; иначе говоря, для всех $\beta \in \mathcal{R}$ отображение d имеет вид: $d(\beta) = (m^b, u^b)$. При этом мы, конечно, требуем существования локально оптимальных решений для каждого β из $\mathcal{R}$. Пусть, далее, g^0 — такая действительная функция на $\mathcal{R}$, что

$$g^0(\beta) = g_{\mathcal{R}}(\beta, m^b, u^b) = g_{\mathcal{R}}(\beta, d(\beta))$$

для всех β из $\mathcal{R}$.

Теорема 6.20

Если функции $g_{\mathcal{R}}$ и d дифференцируемы по Фреше и для всех β в $\mathcal{R}$ пара $(m^b, u^b) = d(\beta)$ является внутренней точкой в $M \times U$, то для каждого β в $\mathcal{R}$ и для каждого i , $1 \leq i \leq n$, существует такое $\lambda_i(\beta) > 0$, что для любой последовательности $\{\beta^h\}$, полученной с помощью преобразования T ,

1) любая точка накопления $\{\beta^h\}$ удовлетворяет условию согласования взаимодействий;

2) последовательность $\{g^0(\beta^h)\}$ монотонно возрастает и сходится к минимуму общих (глобальных) затрат, если $\{g^0(\beta^h)\}$ имеет точку накопления.

Доказательство. Рассматривая приращения функции g^0 , получим, что для всех β в $\mathcal{R}$ и действительных чисел λ

$$g^0(\beta + \lambda h) - g^0(\beta) \geq g_{\mathcal{R}}(\beta, m^b, u^b) \lambda h + o(\lambda),$$

где $g'_{\mathcal{B}}(\beta, m^\beta, u^\beta)$ — производная Фреше от $g_{\mathcal{B}}$ по β в точке $(\beta, m^\beta, u^\beta)$. Из (6.94) следует, что

$$g'_{\mathcal{B}}(\beta, m^\beta, u^\beta) \lambda h = \lambda \sum_{i=1}^n \langle h_i, u_i^\beta - K_i(m^\beta) \rangle.$$

Положим теперь $h = u^\beta - K(m^\beta)$ и для каждого i , $1 \leq i \leq n$, выберем $\lambda_i(\beta) > 0$ так, чтобы выполнялось условие

$$\sum_{i=1}^n \lambda_i(\beta) \|h_i\|^2 = \frac{1}{2} \sup_{\lambda} \sum_{i=1}^n \lambda_i \|h_i\|^2; \quad (6.96)$$

$\sup$ здесь ищется по $\lambda = (\lambda_1, \dots, \lambda_n)$, $\lambda_i > 0$, и при этом

$$g^0(\beta + \lambda h) - g^0(\beta) \geq \sum_{i=1}^n \lambda_i \|h_i\|^2,$$

где $\lambda h = (\lambda_1 h_1, \dots, \lambda_n h_n)$. Тогда для любого β из $\mathcal{B}$ имеем

$$g^0(T(\beta)) - g^0(\beta) \geq \sum_{i=1}^n \lambda_i(\beta) \|u_i^\beta - K_i(m^\beta)\|^2 \geq 0. \quad (6.97)$$

Поэтому последовательность $\{g^0(\beta^k)\}$ монотонно возрастает и сходится.

Предположим, что $\bar{\beta}$ — точка накопления последовательности $\{\beta^k\}$. Функции T , g^0 являются непрерывными на $\mathcal{B}$, поэтому, если β не удовлетворяет условию согласования, т. е. $u^\beta - K(m^\beta) \neq 0$,

$$g^0(T(\bar{\beta})) - g^0(\bar{\beta}) \geq \sum_{i=1}^n \lambda_i(\bar{\beta}) \|u_i^{\bar{\beta}} - K_i(m^{\bar{\beta}})\|^2 > 0.$$

Следовательно, для каких-то β^k , достаточно близких к $\bar{\beta}$, справедливо неравенство

$$g^0(T(\beta^k)) - g^0(\beta^k) > 0,$$

которое противоречит условию монотонного возрастания $\{g^0(\beta^k)\}$. Таким образом, $\bar{\beta}$ удовлетворяет условию согласованности и $\{g^0(\beta^k)\}$ сходится к минимуму глобальных затрат. Этим доказательство завершается.

Теорема 6.21

Допустим, что предположения теоремы 6.20 выполняются и $\lambda_i(\beta)$ выбираются в соответствии с (6.96). Если последовательность $\{\beta^k\}$, полученная с помощью преобразования T , такова, что для каждого i , $1 \leq i \leq n$, последовательность $\{\lambda_i(\beta^k)\}$ ограничена, то $\{\beta^k\}$ имеет точку накопления.

Д о к а з а т е л ь с т в о. Из теоремы 6.20 вытекает, что последовательность $\{g(\beta^k)\}$ сходится к некоторому $\bar{v} < \infty$, а из неравенства (6.97) следует, что

$$\bar{v} \geq g(\beta^k) + \sum_{l=k}^{\infty} \left(\sum_{i=1}^n (\lambda_i(\beta^l) \|u_i^{\beta^l} - K_i(m^{\beta^l})\|^2 \right).$$

Поэтому

$$\sum_{i=1}^n \lambda_i(\beta^l) \|u_i^{\beta^l} - K_i(m^{\beta^l})\|^2 \rightarrow 0 \text{ при } l \rightarrow \infty.$$

Для любого целого $k > 0$

$$\|\beta^{k+1} - \beta^k\| = \sum_{i=1}^n \lambda_i(\beta^k) \|u_i^{\beta^k} - K_i(m^{\beta^k})\|.$$

Тогда, поскольку $\lambda_i(\beta)$ ограничены,

$$\|\beta^{k+1} - \beta^k\| \rightarrow 0 \text{ при } k \rightarrow \infty.$$

Таким образом, последовательность $\{\beta^k\}$ имеет точку накопления, что и завершает доказательство.

Заметим, что практически почти всегда можно с уверенностью допустить существование точки накопления, потому что поиск оптимального координирующего сигнала β проводится на замкнутой ограниченной области $\mathcal{B}$. Если принцип согласования взаимодействий применим, описанная процедура приводит к оптимальной точке.

Стратегия, основанная на прогнозировании взаимодействий

Сходимость итерационного процесса в случае прогнозирования взаимодействий можно доказать совершенно аналогичным путем. Итерационный процесс определяется преобразованиями $T_{i\mathcal{A}}$ и $T_{i\mathcal{B}}$, $1 \leq i \leq n$, которые задаются на множестве $\mathcal{A} \times \mathcal{B}$ формулами

$$T_{i\mathcal{A}}(\gamma) = \alpha_i^\gamma - \mu_i(\gamma) [\eta_i(m^\gamma) - \beta_i^\gamma],$$

$$T_{i\mathcal{B}}(\gamma) = \beta_i^\gamma - \lambda_i(\gamma) [K_i(m^\gamma) - \alpha_i^\gamma],$$

где $\gamma = (\alpha, \beta)$, а $\lambda_i(\gamma)$ и $\mu_i(\gamma)$ — положительные числа. Предполагается, что кажущаяся глобальная целевая функция $g_{\mathcal{B}}$ определяется выражением (6.93). Сходимость доказывается путем рассмотрения приращений первого порядка функции $\hat{g}_{\mathcal{B}}$. Хотя в деталях выкладки несколько отличаются от случая согласования взаимодействий, логика доказательства та же самая, поэтому мы не станем ее здесь воспроизводить.

КООРДИНАЦИЯ ОПТИМИЗИРУЮЩИХ СИСТЕМ ПРИ НАЛИЧИИ ОГРАНИЧЕНИЙ. МАТЕМАТИЧЕСКОЕ ПРОГРАММИРОВАНИЕ

Две предыдущие главы были посвящены главным образом координации оптимизирующих систем без ограничений. Однако мы вовсе не исключали возможности того, что на искомые решения как глобальной, так и локальных задач может быть наложен ряд ограничений. Так, например, мы требовали, чтобы множество допустимых управляющих воздействий (управлений) для процесса в целом обладало некоторым специальным свойством: в гл. 5 предполагалось, что оно является декартовым произведением, а в гл. 6 оно представляло собой прямую сумму. Здесь мы будем рассматривать по существу те же самые двухуровневые системы, что и в предыдущих главах. Различие состоит лишь в том, что теперь мы в явном виде учтем ограничения, наложенные на управляющие воздействия для процесса в целом и на локальные решения. Мы выясним, как влияют ограничения на применимость принципов координации, а затем используем эти принципы при рассмотрении проблемы координирования в задачах (линейного и нелинейного) математического программирования.

1. ВВЕДЕНИЕ

Формулировка глобальной задачи для двухуровневой системы включает в себя конкретизацию как глобальной целевой функции g , определенной на множестве управляющих воздействий (управлений) для всего процесса в целом $M = M_1 \times \dots \times M_n$, так и для заданного множества управляющих воздействий M^j , допустимых для процесса в целом, $M^j \subseteq M$. Задача состоит в том, чтобы минимизировать g на множестве M^j . M^j является подмножеством M ; тем самым подразумевается, что на управляющие воздействия для процесса в целом могут быть наложены ограничения.

Ограничения в задаче оптимизации обычно выражаются условиями, относящимися к входам и выходам. Например, множество входов M и множество выходов Y могут быть пространствами функций, интегрируемых с квадратом на интервале $[0, 1]$. При этом допустимыми (т. е. множеством M^j) являются управляющие воздействия m из M , для которых $\|m(t)\| \leq 1$ почти всюду на отрезке $[0, 1]$, а значение входа $y = P(m)$ в некоторый момент t

из $[0, 1]$ должно лежать в заданной области. Ограничения в задачах математического программирования обычно выражаются совокупностью равенств и неравенств, таких, например, как $m \geq 0$ и $P(m) \geq b$.

Локальные задачи также могут быть оптимизационными задачами с ограничениями. Ограничения выражаются в том, что допустимыми локальными решениями служат не все элементы множеств $M_i \times U_i$, а лишь элементы некоторых их подмножеств. В общем случае мы предполагаем, что заданы такие два множества $\mathcal{A}$ и $\mathcal{B}$, что каждая пара (α, β) из $\mathcal{A} \times \mathcal{B}$ определяет для каждого i , $1 \leq i \leq n$, определенную на $M_i \times U_i$ локальную целевую функцию $g_{i\beta}$ и множество допустимых решений $X_{i\alpha} \subseteq \bar{M}_i \times \bar{U}_i$. При этом i -я локальная задача для выбранной пары (α, β) из $\mathcal{A} \times \mathcal{B}$ состоит в том, чтобы минимизировать $g_{i\beta}$ на множестве допустимых решений $X_{i\alpha}$.

Множества допустимых локальных решений $X_{i\alpha}$ являются, вообще говоря, отношениями, выражающими наложенные ограничения, которые могут быть различными для различных α из $\mathcal{A}$. Конкретизация накладываемых ограничений (а следовательно, и множества допустимых решений) сама по себе является средством координации. Мы, однако, будем интересоваться в основном применением принципов координации; поэтому мы ограничим наш анализ следующими случаями: Пусть для каждого i , $1 \leq i \leq n$, задано множество $X_i \subseteq \bar{M}_i \times U_i$. Тогда 1) при использовании принципов согласования мы будем предполагать, что для каждого локального решающего элемента фиксировано множество допустимых локальных решений $X_{i\alpha} = X_i$; и 2) при применении принципа прогнозирования взаимодействий мы будем предполагать, что $\mathcal{A} = U$ и $X_{i\alpha} = M_{i\alpha} \times \{\alpha_i\}$, где $M_{i\alpha}$ — множество локальных управлений m_i , для которых пара (m_i, α_i) лежит в X_i .

Основной вопрос, с которым мы здесь сталкиваемся, заключается в том, какими должны быть множества допустимых локальных решений, чтобы можно было применять принципы координации.

Прежде чем перейти к ответу на этот вопрос, сделаем следующие замечания.

1) Пусть M_i^f — проекция множества M^f управляющих, воздействующих, допустимых для процесса в целом, на множество M_i :

$$M_i^f = \{m_i: m_i = \pi_i(m), \quad m \in M^f\},$$

где π_i — отображение, проектирующее M на M_i . Тогда в общем случае

$$M^f \subseteq M_1^f \times \dots \times M_n^f$$

в отличие от рассматривавшегося ранее частного случая, когда было $M^f = M_1^f \times \dots \times M_n^f$.

2) Если $M^f = M_1^f \times \dots \times M_n^f$, то множества допустимых локальных решений вводятся сразу же в явном виде, а именно для каждого i , $1 \leq i \leq n$, имеем

$$X_i = M_i^f \times U_i^f,$$

где U_i^f — некоторое подмножество U_i , содержащее все связующие входы для i -го подпроцесса, которые могут иметь место при работе объединенной (связанной) системы и появляются в результате подачи управляющих воздействий, допустимых для процесса в целом: $K_i(M^f) \subseteq U_i^f$. Этот случай, однако, уже рассматривался в двух предыдущих главах. Отметим, что в гл. 6 множество M^f представляло собой прямую сумму, а не декартово произведение.

2. ПРИМЕНИМОСТЬ ПРИНЦИПОВ КООРДИНАЦИИ

Ограничения, наложенные на управляющие воздействия для процесса в целом или на локальные решения, влияют на применимость принципов координации. В самом деле, тот или иной принцип координации, применимый к двухуровневой системе при отсутствии ограничений, может при наличии ограничений оказаться неприменимым вследствие дополнительного требования допустимости полученных решений. Предположим, например, что при отсутствии ограничений система координируема с помощью принципа согласования взаимодействий, причем $\hat{m}$ — единственное глобально оптимальное управляющее воздействие. Тогда принцип согласования взаимодействий не применим, если ограничения наложены таким образом, что $\hat{m}$ не является допустимым управлением. С другой стороны, принцип координации, который нельзя применить к заданной двухуровневой системе без ограничений, может стать применимым, если на локальные решения наложить подходящие ограничения. Например, ограничения, накладываемые на локальные решения, могут оказаться такими, что для некоторого глобально оптимального управляющего воздействия $\hat{m}$ на процесс в целом каждое множество допустимых локальных решений X_i будет состоять из единственной точки $\{(\hat{m}_i, K_i(\hat{m}))\}$.

Проблема ограничений имеет два аспекта. Первый — как влияют ограничения на применимость принципов координации и координируемость, и второй — какими должны быть допустимые локальные решения, чтобы можно было применять принцип координации. Большинство наших выводов, касающихся обоих аспектов, относятся к двухуровневым системам, обладающим по меньшей мере свойством монотонности.

Ограничения, налагаемые на управляющие воздействия для процесса в целом

Предположим, что множества допустимых локальных решений для данной двухуровневой системы таковы, что при каждом i , $1 \leq i \leq n$, $X_i = M_i^f \times U_i^f$, где $K_i(M^f) \subseteq U_i^f$, а M_i^f — проекция M^f на M_i . При применении определенного принципа координации в случае $M^f \neq M_1^f \times \dots \times M_n^f$ может оказаться, что для какого-то координирующего сигнала оптимальные допустимые локальные решения удовлетворяют условию координируемости и тем не менее результирующее управляющее воздействие для процесса в целом не является допустимым. Принцип координации может оказаться неприменимым к данной системе, даже если он применим в случае, когда множество управляющих воздействий, допустимых для процесса в целом, представляет собой декартово произведение $M_1^f \times \dots \times M_n^f$.

Чтобы разрешить вопрос о допустимости решений, возникающий при применении принципов координации, пожалуй, проще всего ввести определенный тест для проверки допустимости. Координатор должен не только проверить, выполняется ли для данных оптимальных локальных решений условие координируемости с помощью выбранного принципа координации, но и определить, является ли результирующее управляющее воздействие для процесса в целом допустимым (глобально допустимым), и если нет, отбросить его.

Приводимые ниже теоремы, касающиеся применимости принципов координации и координируемости систем на их основе при использовании проверки допустимости, просты и не требуют доказательств.

Теорема 7.1

Если выбранный принцип координации применим к двухуровневой системе, для которой множеством глобально допустимых управляющих воздействий является множество $M^f = M_1^f \times$

$\times M_n^f$, то этот принцип, дополненный проверкой допустимости, также применим.

Теорема 7.2

Если двухуровневая система координируема с помощью некоторого принципа координации, когда множеством глобально допустимых управляющих воздействий является множество $M_1^f \times$

$\times M_n^f$, то для координируемости с помощью принципа, включающего проверку допустимости, необходимо выполнение условия

$$\min_{M^f} g(m) = \min_{\dots \times M_n^f} g(m). \quad (7.1)$$

Эти две теоремы справедливы вне зависимости от того, как определены множества допустимых локальных решений.

Хотя такой подход, при котором одновременно с применимостью принципа координации исследуется глобальная допустимость, с принципиальной точки зрения очень прост, при его практическом применении мы сталкиваемся со значительными трудностями:

1. Разумно предположить, что условие (7.1), необходимое для координируемости при таком подходе, выполняется лишь в некоторых частных случаях.

2. Условие (7.1), вообще говоря, не является достаточным для координируемости, и задача координатора может оказаться весьма трудоемкой. Например, если для некоторого координирующего сигнала оптимальные и допустимые локальные решения удовлетворяют условию координируемости, а получающееся в результате управляющее воздействие не является глобально допустимым, то координатор не знает, координируема система или нет. Для ответа на этот вопрос он должен опробовать все координирующие сигналы, для которых оптимальные локальные допустимые решения удовлетворяют условию координируемости, и определить, является ли соответствующее управляющее воздействие глобально допустимым.

3. Даже если система координируема с помощью определенного принципа, включающего проверку допустимости, трудно подобрать такие координирующие сигналы, которые координировали бы систему. Когда проверка допустимости не требуется, корректировку координирующих сигналов можно осуществлять, как показано в разд. 8 гл. 5 и разд. 5 гл. 6, исходя из значений показателей рассогласования системы или погрешностей прогнозных данных о связующих входах. Однако если проверка допустимости необходима, то, за исключением некоторых частных случаев (например, задачи линейного программирования), не ясно, как производить подобную корректировку.

Согласованные ограничения

Описанный подход к задаче координации системы в случае, когда на управляющие воздействия наложены глобальные ограничения, хотя и прост по идее, но не представляет большой практической ценности из-за упомянутых трудностей. Попытаемся поэтому найти такую зависимость между множествами допустимых локальных решений и множеством глобально допустимых (допустимых с точки зрения процесса в целом) управляющих воздействий, которая позволила бы избежать проверки глобальной допустимости.

Уже сами принципы координации содержат указание на характер такой зависимости; а именно локальная допустимость и согласованность подразумевают глобальную допустимость. Пусть X — множество всех пар (m, u) из $M \times U$, таких, что для любого i , $1 \leq i \leq n$, пара (m_i, u_i) лежит в X_i . Условие

$$(\forall m) [(m, K(m)) \in X] \Rightarrow [m \in M^*] \quad (7.2)$$

выражает искомую связь между множествами допустимых локальных решений и множеством глобально допустимых управляющих воздействий.

Для того чтобы без проверки глобальной допустимости можно было использовать либо принцип согласования взаимодействий, либо принцип прогнозирования взаимодействий, достаточно выполнения условия (7.2) при заданных ограничениях. Следует особо подчеркнуть, что оно не гарантирует применимости принципа, а лишь означает, что проверка допустимости не обязательна.

Условие (7.2) не является, однако, достаточным для того, чтобы исключить проверку допустимости при использовании принципа согласования функций качества. Предположим, например, что для некоторого координирующего цели сигнала β существует пара (m^β, u^β) , для которой при каждом i , $1 \leq i \leq n$, пара (m_i^β, u_i^β) является оптимальным допустимым локальным решением и

$$g_{i\beta}(m_i^\beta, u_i^\beta) = g_{i\beta}(m_i^\beta, K_i(m^\beta)).$$

Далее, даже если для рассматриваемой двухуровневой системы выполняется условие (7.2), может оказаться, что некоторые из пар $(m_i^\beta, K_i(m^\beta))$, $1 \leq i \leq n$, не являются допустимыми, и, следовательно, мы не можем утверждать, что m^β — глобально оптимальное допустимое управляющее воздействие.

Если выполняется условие

$$(\forall m) [(m, K(m)) \in X] \Rightarrow [m \in M^*],$$

мы будем говорить, что ограничения данной системы являются согласованными. При согласованных ограничениях для двухуровневой системы, обладающей свойством монотонности, можно сформулировать следующую теорему.

Теорема 7.3

Пусть рассматриваемая двухуровневая система обладает свойством монотонности и наложенные на нее ограничения являются согласованными. Тогда для этой системы применим принцип согласования взаимодействий.

Доказательство. Пусть β — определенный координирующий цели сигнал и существует такая пара (m^β, u^β) , что $u^\beta = K(m^\beta)$. Тогда из условия согласованности ограничений и свой-

ства монотонности вытекает, что m^β является глобально допустимым и

$$g(m^\beta) = g_{\mathcal{B}}(\beta, m^\beta, K(m^\beta)) = \min_{m^f} g_{\mathcal{B}}(\beta, m, K(m)) = \min_{m^f} g(m),$$

где g — глобальная, а $g_{\mathcal{B}}$ — кажущаяся глобальная целевые функции. Этим доказывается применимость принципа, а тем самым и завершается доказательство теоремы.

Для системы, в которой ограничения согласованы, условия координируемости с помощью принципа согласования взаимодействий можно вывести, используя результаты, полученные в двух предыдущих главах. Разумеется, необходимо сделать некоторые предположения относительно множества глобально допустимых управляющих воздействий и множества допустимых локальных решений. Поскольку связанные с этим изменения очевидны, мы не станем переформулировать условия координируемости на случай системы с ограничениями. Стоит, однако, отметить, что вследствие специфического характера результатов гл. 6, касающихся линейных операторов оценки эффективности, эти результаты, вообще говоря, в данном случае неприменимы.

Аналогично условия применимости принципа прогнозирования взаимодействий и координируемости с помощью этого принципа для системы без ограничений, не связанные с использованием линейных операторов оценки эффективности, по-прежнему остаются в силе и их можно применять в том же виде и к системам, для которых ограничения являются согласованными.

Чтобы содержательным образом охарактеризовать класс согласованных ограничений, обратимся сначала к рассмотрению их структуры. Без ограничения общности мы можем предположить, что глобально допустимые управляющие воздействия определяются отображением $\tau: M \rightarrow Q$ и множеством приемлемых решений $Q^0 \subseteq Q$:

$$M^f \quad \{m: \tau(m) \in Q^0\}.$$

Аналогично предположим, что для каждого i , $1 \leq i \leq n$, заданы отображение $\tau_i: M_i \times U_i \rightarrow Q_i$ и множество приемлемых решений $Q_i^0 \subseteq Q_i$, таких, что X_i определяется как

$$X_i = \{(m_i, u_i) \mid \tau_i(m_i, u_i) \in Q_i^0\}.$$

Ограничения в оптимизационных задачах обычно формулируются таким образом. Без потери общности мы в дальнейшем можем считать, что отображения τ и τ_i , $1 \leq i \leq n$, таковы, что области их значений согласованы.

Теорема 7.4

Пусть существует такое отображение $\theta: Q_1 \times \dots \times Q_n \rightarrow Q$, что

1) $\tau(m) = \theta(\tau_1(m_1, K_1(m_1)), \tau_n(m_n, K_n(m_n)))$ для всех m из M^f ;

2) $Q^0 \times \dots \times Q^n$ — прообраз Q^0 при отображении θ .
Тогда ограничения являются согласованными.

Доказательство. Пусть m — произвольный элемент из M . Пара $(m, K(m))$ принадлежит X тогда и только тогда, когда $\tau_i(m_i, K(m_i)) \in Q^0_i$ для всех i , $1 \leq i \leq n$, а в соответствии с условиями 1 и 2 это возможно тогда и только тогда, когда $\tau(m) \in Q^0$. Поскольку $\tau(m) \in Q^0$ тогда и только тогда, когда $m \in M^f$, доказательство завершено.

Эта теорема характеризует класс согласованных ограничений и имеет важное значение, так как указывает такой способ определения множеств допустимых локальных решений, при котором наложенные ограничения согласованны. Существенно, что условие 1 требует, чтобы семейство отображений τ_i было декомпозицией τ относительно отображений K_i , описывающих взаимосвязи и взаимодействия подсистем. Требование 2 дает достаточное (но не необходимое) условие взаимосвязи между множествами приемлемых решений, которое обеспечивает согласованность ограничений.

В качестве примера укажем, что условие 1 удовлетворяется, если τ представляет собой совокупный процесс P , а отображения τ_i — i -е локальные процессы P_i . В этом случае θ будет просто тождественным отображением, а ограничения накладываются на выходы процесса. Тогда, чтобы удовлетворить условию 2, множества приемлемых решений $Q^0 \subseteq Y$ и $Q^0 \in Y$ $1 \leq i \leq n$, должны быть такими, что

$$Q^0 = Q^0_1 \times \dots \times Q^0_n,$$

хотя это условие в общем случае и не является необходимым.

3. КООРДИНАЦИЯ ЗАДАЧ ВЫПУКЛОГО ПРОГРАММИРОВАНИЯ

Изложенные в предыдущем разделе методы, ориентированные на получение допустимых решений, могут быть непосредственно использованы для двухуровневых систем, у которых глобальная и локальные задачи сформулированы в виде задач так называемого математического программирования. Мы рассмотрим, в частности, класс двухуровневых систем, у которых глобальная и локальные задачи суть задачи выпуклого программирования. Заодно мы покажем, как результаты гл. 5 и 6 совместно с результатами, полученными в предыдущем разделе для согласованных ограничений, могут быть использованы для координации в задачах выпуклого программирования. Глобальной задачей рассматриваемой двухуровневой системы, как уже указывалось во введении к этой

главе, является минимизация определенной глобальной целевой функции g на некотором множестве M . Это множество в теории математического программирования принято называть *допустимой областью*. Мы считаем g выпуклой вещественнозначной функцией, определенной на евклидовом пространстве $M = M_1 \times \dots \times M_n$. Множество M_i при любом i , $1 \leq i \leq n$, также может представлять собой многомерное евклидово пространство. Пусть, далее, F_i при $1 \leq i \leq n$ — выпуклые функции, отображающие M в евклидово пространство Y_i , а допустимая область M' определена как множество M всех m , удовлетворяющих ограничениям

$$m \geq 0, \quad F_i(m) \leq 0; \quad i = 1, \dots, n. \quad (7.3)$$

Упорядоченность на Y_i и M понимается в обычном смысле.

Таким образом, глобальная задача есть задача выпуклого программирования: минимизировать $g(m)$ при условиях (7.3).

Локальные задачи также будут задачами выпуклого программирования, но уже в пространствах $M_i \times U_i$, $1 \leq i \leq n$, где U_i — соответствующим образом выбранные пространства. Ограничения, выделяющие допустимые области X_i , $1 \leq i \leq n$, для локальных задач, нужно подобрать так, чтобы они были согласованы с глобальными ограничениями (7.3). Чтобы добиться этого, произведем декомпозицию функций F_i следующим образом: для всех i , $1 \leq i \leq n$, F_i представляется в виде двух отображений $P_i: M \times U_i \rightarrow Y_i$ и $K_i: M \rightarrow U_i$, таких, что

- 1) U_i — евклидово пространство;
- 2) P_i выпукло на $M_i \times U_i$;
- 3) $P_i(m_i, K_i(m)) = F_i(m)$ для всех m из M .

Чтобы привести пример подобной декомпозиции, положим

$$U_i = M_1 \times \dots \times M_{i-1} \times M_{i+1} \times \dots \times M_n$$

и

$$P_i(m_i, u_i) = F_i(u_{i1}, \dots, u_{i,i-1}, m_i, u_i, \dots, u_{in}), \quad (7.5)$$

где

$$u_i = (u_{i1}, \dots, u_{i,i-1}, u_{i,i+1}, \dots, u_{in}).$$

Отображение K_i есть просто проекция M на U_i . Определим для i -й локальной задачи допустимую область X_i как множество всех пар (m_i, u_i) из $M_i \times U_i$, удовлетворяющих ограничениям

$$m_i \geq 0, \quad P_i(m_i, u_i) \leq 0. \quad (7.6)$$

Будем считать, что упорядочение (отношение $\geq$) в M вводится следующим образом: $m \geq m'$ тогда и только тогда, когда $m_i \geq m'_i$ для всех i , $1 \leq i \leq n$. Тогда можно показать, что совокупность локальных ограничений, определяемых для каждого i , $1 \leq i \leq n$, условиями (7.6), согласована с глобальными ограничениями (7.3).

Чтобы сделать это, запишем глобальные ограничения (7.3) в виде

$$\tau(m) \leq 0,$$

где $\tau(m) = (-m, F_1(m), \dots, F_n(m))$. В аналогичном виде представим и локальные ограничения (7.5)

$$\tau_i(m_i, u_i) \leq 0,$$

где $\tau_i(m_i, u_i) = (-m_i, P_i(m_i, u_i))$. Далее, из (7.4) для каждого m из M имеем: $\tau(m) \leq 0$ тогда и только тогда, когда $\tau_i(m_i, K_i(m)) \leq 0$ для всех i , $1 \leq i \leq n$. Таким образом, ограничения являются согласованными. Действительно, теорема 7.4 может быть применена, поскольку отображение θ , для которого условия этой теоремы выполняются, существует.

Предположим теперь, что для всех i , $1 \leq i \leq n$, введены выпуклые вещественнозначные функции g_i , определенные на $M_i \times U_i$ так, что i -я локальная (немодифицированная) задача сводится к минимизации функции g_i в допустимой области, заданной условиями (7.6). Тогда локальные задачи двухуровневой системы будут задачами выпуклого программирования: минимизировать $g_i(m_i, u_i)$ при ограничениях (7.6).

Чтобы для решения задачи координирования рассматриваемой двухуровневой системы можно было применить сформулированные нами принципы координации, допустим, что имеющиеся глобальная (g) и локальные (g_i) целевые функции связаны между собой соответствующим образом; а именно, что g является аддитивной функцией:

$$g(m) = \sum_{i=1}^n g_i(m_i, K_i(m)). \quad (7.7)$$

Если локальные целевые функции получаются с помощью декомпозиции глобальной целевой функции, то декомпозицию глобальной целевой функции и ограничений F_i нужно проводить одновременно, чтобы удовлетворить как условиям (7.4), так и условиям (7.7). Например, в рассматривавшемся выше примере, где мы показали способ декомпозиции функций, посредством которых вводятся ограничения, локальные целевые функции можно ввести следующим образом:

$$g_i(m_i, u_i) = \frac{1}{n} g(u_{i1}, \dots, u_{i, i-1}, m_i, u_{i, i+1}, \dots, u_{in}).$$

В этом случае условия (7.4) и (7.7) будут выполнены.

Поскольку ограничения, определяющие допустимые области для локальных задач, согласованы с ограничениями относительно допустимой области глобальной задачи, принципы согласования или прогнозирования взаимодействий можно попытаться применить для координирования локальных задач относительно глобальной задачи, не проверяя глобальную допустимость. Такое «согласованное распределение» требования глобальной допустимости позволяет непосредственно применять результаты гл. 5 и 6

в задачах координации. Следующая теорема представляет собой прямое следствие теорем 6.4 и 7.4.

Теорема 7.5

Пусть глобальная целевая функция аддитивна, глобальная задача имеет оптимальное допустимое решение и заданные локальные целевые функции модифицируются путем применения линейных модификаций с нулевой суммой, т. е. для каждого i , $1 \leq i \leq n$, и β из $\mathcal{R}$ $U_1 \times \dots \times U_n$

$$g_{i\beta}(m_i, u_i) = g_i(m_i, u_i) + \beta_i^T u_i - \sum_{j=1}^n \beta_j^T K_{ji}(m_i). \quad (7.8)$$

Допустим, далее, что существует координирующий сигнал β из $\mathcal{R}$, такой, что

$$\sum_{i=1}^n \inf_{X_i} g_{i\beta}(m_i, u_i) = \sup_{\mathcal{R}} \sum_{i=1}^n \inf_{X_i} g_{i\beta}(m_i, u_i), \quad (7.9)$$

и для каждого i , $1 \leq i \leq n$, допустимая область X_i компактна и $g_{i\beta}$ для всех β из $\mathcal{R}$ непрерывна на ней. Тогда двухуровневая система координируема с помощью принципа согласования взаимодействий.

Модификации, определяемые уравнениями (7.8), обладают нулевой суммой тогда и только тогда, когда для всех i , $1 \leq i \leq n$, равенство

$$K_i(m) = \sum_{j=1}^n K_{ij}(m_j) \quad (7.10)$$

выполнено при любом m из M . Если бы (7.10) нарушалось при некоторых m из M и j , $i \leq j \leq n$, то нашлось бы такое β из $\mathcal{R}$, что

$$\sum_{i=1}^n g_{i\beta}(m_i, K_i(m)) \neq g(m).$$

Поскольку декомпозиция функций ограничений F_i имеет вид (7.5), существуют функции K_{ij} , $1 \leq j \leq n$, для которых (7.10) удовлетворяется при любом $m \in M$.

Для координируемости требуется существование координирующего сигнала, удовлетворяющего условию (7.9). Однако, если накладываются другие ограничения, нам вовсе не обязательно предполагать существование такого координирующего сигнала в явном виде. Поскольку для рассматриваемой двухуровневой системы ограничения являются согласованными, мы можем, воспользовавшись теоремой 6.6, получить следующий результат.

Теорема 7.6

Пусть глобальная целевая функция аддитивна, глобальная задача имеет оптимальное допустимое решение, заданные локаль-

ные целевые функции преобразуются с помощью линейных модификаций с нулевой суммой и для каждого i , $1 \leq i \leq n$, допустимая область X_i компактна, а $g_{i\beta}$ непрерывна на ней при любом β из $\mathcal{B}$. Допустим еще, что точка начала отсчета является внутренней точкой множества E , состоящего из всех $e = u - K(m)$, где пара (m, u) принадлежит X . Тогда двухуровневая система координируема с помощью принципа согласования взаимодействий.

Компактность допустимой области, требующаяся по условиям теоремы 7.6, является довольно жестким ограничением. Это требование может быть ослаблено, но тогда, чтобы доказать координируемость, мы должны будем предположить, что функции взаимодействия K_i линейны. Это и другие следствия линейности функций взаимодействия K_i устанавливаются в следующей теореме.

Теорема 7.7

Пусть глобальная целевая функция аддитивна, глобальная задача имеет оптимальное допустимое решение и модификации локальных целевых функций являются линейными с нулевой суммой. Если при этом допустимая область M^f имеет внутреннюю точку и для каждого i , $1 \leq i \leq n$, функции K_i линейны, а g_i непрерывны на допустимой области X_i , то двухуровневая система координируема с помощью принципа согласования взаимодействий.

Доказательство. Пусть $U = U_1 \times \dots \times U_n$. Определим функции f и L на $M \times U$ в виде

$$f(m, u) = \sum_{i=1}^n g_i(m_i, u_i), \quad L(m, u) = u - K(m),$$

где $K(m) = (K_1(m), \dots, K_n(m))$. Тогда для всех β из U

$$\sum_{i=1}^n g_{i\beta}(m_i, u_i) = f(m, u) + \beta^T L(m, u). \quad (7.11)$$

Определим теперь подмножества $\bar{V}$ и $\underline{V}$ множества $R \times U$ следующим образом:

$$\begin{aligned} \bar{V} &= \{(r, e): r \in R \text{ и } f(x) \leq r \text{ и } e = L(x) \\ &\quad \text{для фиксированного } x \text{ из } X\}, \\ \underline{V} &= \{(r, 0): r \in R \text{ и } r < f(\hat{x})\}, \end{aligned}$$

где $\hat{x} = (\hat{m}, K(\hat{m}))$, а $\hat{m}$ — оптимальное допустимое решение глобальной задачи. Поскольку такое $\hat{m}$ существует, а K — линейная функция, то множества $\bar{V}$ и $\underline{V}$ отделимы, непусты и выпуклы. Более того, множество $\bar{V}$ имеет внутреннюю точку. Это следует

из предположения о том, что допустимая область имеет внутреннюю точку и для каждого i , $1 \leq i \leq n$, K_i линейна, а g_i непрерывна на X_i . Применяя теперь теорему о разделяющей гиперплоскости, можно получить действительные числа α и r_0 и вектор β из U , такие, что

$$\begin{aligned} \alpha r + \beta^T e &\geq r_0 \text{ для всех } (r, e) \text{ из } \bar{V}, \\ \alpha r + \beta^T e &< r_0 \text{ для всех } (r, e) \text{ из } \underline{V}. \end{aligned} \quad (7.12)$$

Из определения $\underline{V}$ сразу вытекает, что $\alpha \geq 0$. Поскольку для всех $r < f(\hat{x})$ имеем $(r, 0) \in \bar{V}$, а $(f(\hat{x}), 0) \in \underline{V}$, то из (7.12) следует, что для всех $r < f(\hat{x})$ должно выполняться неравенство $\alpha r < r_0 \leq \alpha f(\hat{x})$. Следовательно, $\alpha > 0$. Далее, $r_0 = \alpha f(\hat{x})$, так как если $r_0 > \alpha f(\hat{x})$, то $(f(\hat{x}), 0)$ лежало бы в $\bar{V}$, что невозможно. Тогда из определения множества $\bar{V}$ и (7.12) имеем

$$j(\hat{x}) = \min [f(x) + \hat{\beta}^T L(x)]$$

и, следовательно, из (7.11) и того, что g аддитивна, получаем

$$g(\hat{m}) = \sum_{i=1}^n \min_{X_i} g_{i\beta}(m_i, u_i),$$

что и доказывает теорему.

Пример 7.1

Рассмотрим в качестве глобальной задачи задачу минимизации, когда целевая функция g определена на $\bar{M} = R \times R$ выражением

$$g(m) = \frac{1}{2} (m_1^2 + m_2^2) - (m_1 + 2m_2),$$

а ограничениями служат условия (7.3) при $n = 2$ и

$$F_1(m) = 2m_1 + 3m_2 - 6, \quad F_2(m) = m_1 + 4m_2 - 5.$$

Оптимальное допустимое решение этой задачи есть $\hat{m} = \left(\frac{13}{17}, \frac{18}{17}\right)$

Сформулируем теперь две подзадачи, которые будут параметрически зависеть от $\beta = (\beta_1, \beta_2)$, $\beta \in R \times R$. В первой из них нужно найти минимум целевой функции

$$g_{i\beta}(m_1, u_1) = \frac{1}{2} m_1^2 - (1 + \beta_2) m_1 + \beta_1 u_1$$

при ограничениях $m_1 \geq 0$, $u_1 \geq 0$, $P_1(m_1, u_1) \leq 0$, где

$$P_1(m_1, u_1) = 2m_1 + 2u_1 - 6;$$

во второй — минимум целевой функции

$$g_{2\beta}(m_2, u_2) = \frac{1}{2} m_2^2 - (2 + \beta_1) m_2 + \beta_2 u_2$$

при ограничениях $m_2 \geq 0$, $u_2 \geq 0$, $P_2(m_2, u_2) \leq 0$, где

$$P_2(m_2, u_2) = 4m_2 + u_2 - 5.$$

Если связать эти подзадачи равенствами $m_1 = u_2$ и $m_2 = u_1$, то глобальная целевая функция будет аддитивной, модификации целевых функций для подзадач — линейными с нулевой суммой, а ограничения будут согласованы.

Теперь можно применить принцип согласования взаимодействий для координации двух подзадач относительно решения глобальной задачи. В данном случае подзадачи координируемы с помощью этого принципа. Пусть $\beta = (\beta_1, \beta_2)$, где $\beta_1 = 0$ и $\beta_2 = -4/17$; тогда оптимальные допустимые решения подзадач будут таковы:

$$m_1^\beta = \frac{13}{17}, \quad 0 \leq u_1^\beta \leq \frac{76}{51},$$

$$m_2^\beta = \frac{18}{17}, \quad u_2^\beta = \frac{13}{17}.$$

Если положить $u_1^\beta = 18/17$, станет очевидно, что данное значение параметра β есть оптимальный координирующий сигнал.

Следует подчеркнуть, что значение оптимального координирующего сигнала β в рассмотренном выше примере отличается от того значения, которое получается при использовании линейных операторов оценки эффективности связующих входов и управляющих воздействий в соответствии с теоремой 6.8. Оптимальный координирующий сигнал в общем случае зависит от накладываемых ограничений.

При решении задачи координации для рассматриваемой двухуровневой системы может быть применен принцип прогнозирования (взаимодействий). Если глобальная целевая функция аддитивна и ограничения согласованы, мы модифицируем локальные целевые функции следующим образом: сложим $\mathcal{R} = M$ и пусть для каждого i , $1 \leq i \leq n$, и β из $\mathcal{R}$

$$g_{i\beta}(m_i, u_i) = g_i(m_i, u_i) + \beta_i^T m_i.$$

Мы не можем здесь применить теоремы 6.15 или 6.16 ввиду того, что на локальные решения и управляющие воздействия для процесса в целом наложены ограничения. Однако мы можем опереться на тот факт, что двухуровневая система, координируемая при использовании линейных модификаций (с нулевой суммой) локальных функций качества и применении принципа согласования взаимодействий, координируема и с помощью принципа прогнозирования при использовании некоторого отображения η :

$\bar{M} \rightarrow \mathcal{B}$ Мы определим вид этого отображения в следующем разделе, где будет рассмотрена проблема координации в задачах линейного программирования. Приводимый ниже пример служит иллюстрацией применения принципа прогнозирования взаимодействий к задаче координации системы, описанной в примере 7.1.

Пример 7.2

Рассмотрим ту же глобальную задачу, что и в предыдущем примере. Пусть $A = R^2$ и каждое α в A есть прогнозное значение связующих входов. Сформулируем две подзадачи, зависящие не только от параметра $\beta = (\beta_1, \beta_2)$ из R^2 , но и от прогнозных значений связующих входов. Для фиксированной пары (α, β) i -я подзадача состоит в следующем:

Минимизировать $g_{1\beta}(m_i)$ при ограничениях

$$m_i \geq 0 \text{ и } P_i(m_i, \alpha_i) \leq 0,$$

где P_1 и P_2 — те же, что и в предыдущем примере, а

$$g_{1\beta}(m_1) = \frac{1}{2} m_1^2 - (1 + \beta_1) m_1, \quad g_{2\beta}(m_2) = \frac{1}{2} m_2^2 - (2 + \beta_2) m_2.$$

Эти подзадачи связаны посредством связующих входов, что выражается равенствами $u_1 = m_2$, $u_2 = m_1$.

Теперь можно показать, что существует такое $\hat{\beta}$ в R^2 , что для любого α из A , $\alpha \geq 0$, вектор $m^v = (m_1^v, m_2^v)$, где m_i^v — оптимальное допустимое решение i -й подзадачи для пары (α, β) — будет оптимальным допустимым решением глобальной задачи, если выполнены условия

$$\alpha_1 = m_2^v \text{ и } \alpha_2 = m_1^v. \quad (7.13)$$

Можно показать, что существует $\alpha \geq 0$, удовлетворяющее (7.13).

Пусть $\hat{\beta} = \left(\frac{4}{17}, 0\right)$. Тогда для любого $\alpha \geq 0$, $\alpha \in R^2$, оптимальными допустимыми решениями подзадач будут m_1^v и m_2^v , где

$$m_1^v = \begin{cases} \frac{1}{2}(6 - 3\alpha_1), & \text{если } \frac{1}{2}(6 - 3\alpha_1) \leq \frac{13}{17}, \\ \frac{13}{17}, & \text{если } \frac{1}{2}(6 - 3\alpha_1) > \frac{13}{17}. \end{cases}$$

$$m_2^v = \frac{1}{4}(5 - \alpha_2).$$

Предполагая, что равенства (7.13) выполнены, подставим в них полученные значения m_1^v и m_2^v и разрешим их относительно α_1 и α_2 , требуя при этом, чтобы $\alpha_1 \geq 0$ и $\alpha_2 \geq 0$. Решение единственно:

$$\alpha_1 = \frac{18}{17}, \quad \alpha_2 = \frac{13}{17};$$

следовательно, подзадачи координируемы при помощи принципа прогнозирования взаимодействий при использовании отображения $\eta: M \rightarrow \left\{ \left(\frac{4}{17}, 0 \right) \right\}$

4. Координация задач линейного программирования

Результаты предыдущего раздела могут быть использованы для декомпозиции сложных задач линейного программирования на несколько более простых взаимосвязанных подзадач. Сформулированные таким образом подзадачи будут снова задачами линейного программирования, зависящими от параметра. Параметр можно подобрать так, чтобы скоординировать их относительно исходной задачи в том смысле, что оптимальные допустимые решения подзадач будут непосредственно давать оптимальное допустимое решение исходной задачи.

Рассмотрим глобальную задачу из предыдущего раздела, но теперь в качестве глобальной целевой функции g возьмем линейную форму

$$g(m) = c^T m = \sum_{i=1}^n c_i^T m_i;$$

здесь c_i , $1 \leq i \leq n$, — заданный вектор из M_i , а ограничения F_i , $1 \leq i \leq n$, имеют вид

$$F_i(m) = b_i - A_i m = b_i - (A_{i1}m_1 + \dots + A_{in}m_n), \quad (7.14)$$

где b_i — заданные векторы из Y_i , а A_{ij} , $1 \leq j \leq n$, — матрица соответствующей размерности. Тогда исходная задача становится задачей линейного программирования:

Минимизировать $c^T m$ при ограничениях

$$m \geq 0 \text{ и } A_i m \geq b_i, \quad 1 \leq i \leq n. \quad (7.15)$$

Упорядочивание $\geq$ понимается в том же смысле, как и в предыдущем разделе.

Произведем теперь декомпозицию ограничений (7.15) так, чтобы каждой из n подзадач можно было приписать ограничения, согласованные с ограничениями исходной задачи. Для каждого i , $1 \leq i \leq n$, произведем декомпозицию исходных функций F_i , определяемых с помощью выражений (7.14), на новые функции $P_i: M_i \times U_i \rightarrow Y_i$ и $K_i: M \rightarrow U_i$ следующим образом:

$$P_i(m_i, u_i) = b_i - A_{ii}m_i - B_{ii}u_i, \quad (7.16)$$

$$K_i(m) = K_{i1}m_1 + \dots + K_{in}m_n,$$

где U_i — соответствующее евклидово пространство, а B_{ii} и K_{ij} , $1 \leq j \leq n$, — матрицы соответствующих размерностей, такие, что

для каждого j

$$B_{ii}K_{ij} = \begin{cases} 0, & \text{если } j = i, \\ A_{ij}, & \text{если } j \neq i. \end{cases} \quad (7.17)$$

Если (7.17) выполнено, то все требования условия (7.4) удовлетворяются и мы можем приписать каждой из n подзадач ограничения, согласованные с ограничениями исходной задачи.

Поскольку глобальная целевая функция линейна, естественно выбрать в качестве целевой функции g_i для i -й подзадачи линейную форму

$$g_i(m_i) = c_i^T m_i. \quad (7.18)$$

Предполагая теперь, что матрицы B_i и K_{ij} , $1 \leq i \leq n$, $1 < j < n$, удовлетворяют условиям (7.17), сформулируем i -ю (немодифицированную) подзадачу в виде следующей задачи линейного программирования:

Минимизировать $c_i^T m_i$ при ограничениях

$$m_i \geq 0, \quad A_{ii}m_i + B_{ii}u_i \geq b_i. \quad (7.19)$$

Очевидно, что ограничения для этих подзадач согласованы с ограничениями исходной глобальной задачи и, как следует из выражения (7.18), глобальная целевая функция аддитивна.

Декомпозицию исходной глобальной задачи можно провести многими способами, вопрос только в том, удастся ли скоординировать получающиеся подзадачи.

Мы рассмотрим этот вопрос только для класса декомпозиций (7.16) — (7.17). Таким образом, декомпозиция глобальной целевой функции и ограничений определена так, что координируемость зависит только от способа модификации выбранных целевых функций подзадач.

Применение принципа согласования взаимодействий

Вопрос о применимости принципа согласования взаимодействий к описанной выше двухуровневой системе и о координируемости этой системы с помощью данного принципа уже был решен в предыдущем разделе. Однако, поскольку глобальная задача и подзадачи являются задачами линейного программирования в вещественном евклидовом пространстве, вопрос о координируемости может быть решен при более слабых условиях, а именно: предположение о том, что допустимая область глобальной задачи содержит в себе внутреннюю точку, может быть опущено.

Глобальная целевая функция является аддитивной, поэтому, чтобы скоординировать подзадачи, мы, как и в предыдущем разделе, модифицируем выбранные для них целевые функции с помощью линейного преобразования с нулевой суммой. Пусть

для каждого β из $U = U_1 \times \dots \times U_n$ модифицированная целевая функция i -й подзадачи имеет вид

$$g_{i\beta}(m_i, u_i) = c_i^T m_i + \beta_i^T u_i - \sum_{j=1}^n \beta_j^T K_{ji} m_i. \quad (7.20)$$

Тогда для решения вопроса о координируемости подзадач с модифицированными целевыми функциями можно применить принцип согласования взаимодействий. Более того, подзадачи будут координируемы с помощью этого принципа, если глобальная задача имеет оптимальное допустимое решение.

Теорема 7.8

Если глобальная задача для описанной выше двухуровневой системы имеет оптимальное допустимое решение и для каждого координирующего сигнала β из $\bar{U}$ i -я подзадача состоит в минимизации функции $g_{i\beta}(m_i, u_i)$, определяемой равенством (7.20) при ограничениях (7.19), то система координируема с помощью принципа согласования взаимодействий.

Доказательство. Данное утверждение является прямым следствием теоремы 7.7. Разница только в том, что в общем случае теорема Мазура о разделяющей гиперплоскости требует наличия внутренней точки в допустимой области для глобальной задачи, а в случае линейных ограничений такое условие не обязательно. Эта теорема в принципе решает вопрос о координируемости, но при практическом применении принципа согласования взаимодействий возникают две трудности.

1. Как правило, подзадачи не при любом координирующем сигнале β из U обладают единственными решениями, даже если глобальная задача имеет единственное решение. Таким образом, для того чтобы определить, удовлетворяются ли условия координируемости для того или иного координирующего сигнала, необходимо проанализировать все комбинации всех оптимальных допустимых решений подзадач для данного координирующего сигнала. Эта трудность, очевидно, не возникает, если подзадачи имеют не более одного оптимального допустимого решения. Вероятность получить такое единственное решение возрастает при сужении допустимой области для подзадач. Например, можно ограничить множество связующих входов u_i и считать, что оно содержится в множестве $K_i(M^i)$, где M^i , как обычно, — допустимая область глобальной задачи. Другой подход к устранению указанной трудности может основываться на том факте, что множество всех оптимальных допустимых решений рассматриваемой задачи линейного программирования является выпуклым подмножеством некоторой гиперплоскости и это подмножество обладает конечным числом «угловых точек». Согласование взаимодействий можно проверить, про-

смотрев все возможные сочетания угловых точек, найденных локальными решающими элементами. Можно разработать практические методы такой проверки.

2. Подзадачи могут и не иметь оптимальных допустимых решений при некоторых координирующих сигналах, даже если глобальная задача имеет решение. Это может быть результатом наложенных на подзадачи ограничений. Из исходных ограничений $A_{ij}m_i + B_{ji}u_j \geq b_i$ для i -й подзадачи. Поэтому может оказаться, что последнее неравенство удовлетворяется в неограниченной области, даже если допустимая область глобальной задачи ограничена. Эту трудность можно преодолеть, если наложить дополнительные ограничения на переменные i -й подзадачи. Например, можно добавить к неравенствам (7.19) для i -й подзадачи $n - 1$ неравенств вида

$$A_{ji}m_i + B_{ji}u_i \geq b_j, \text{ где } i \neq j, \quad (7.21)$$

где $j \neq i$ и матрицы B_{ij} подобраны соответствующим образом. Если матрицы B_{ji} таковы, что для любого k , $1 \leq k \leq n$,

$$B_{ji}K_{ik} = \begin{cases} 0, & \text{если } k = i, \\ A_{jk}, & \text{если } k \neq i, \end{cases}$$

то добавление $n - 1$ ограничений, определяемых выражением (7.21), к ограничениям (7.19) даст такие подзадачи, у которых ограничения будут согласованы с глобальными ограничениями. Более того, дополненная таким образом совокупность глобальных ограничений может служить основой для построения ограничений, накладываемых на каждую из подзадач.

Прежде чем закончить этот раздел, посвященный принципу согласования взаимодействий, отметим связь, существующую между оптимальным координирующим сигналом и оптимальным допустимым решением двойственной глобальной задачи.

Задачей, двойственной к исходной глобальной задаче, называется задача:

Максимизировать $b^T y$ при ограничениях

$$y \geq 0 \text{ и } A^T y \leq c,$$

где $b = (b_1, \dots, b_n)$ и A — матрица $[A_{ij}]$.

Ясно, что исходная задача имеет оптимальное допустимое решение в том и только в том случае, когда такое (оптимальное и допустимое) решение имеет двойственная задача. Теперь мы можем переформулировать теорему 7.8: если глобальная задача, двойственная исходной, имеет оптимальное допустимое решение, то подзадачи координируемы на основе принципа согласования взаимодействий. Однако важной отличительной особенностью двойственной задачи является то, что оптимальный координирующий сигнал (при использовании принципа согласования взаимодействий)

вий) в явном виде выражается через оптимальные решения двойственной задачи. Этот факт устанавливается следующей теоремой.

Теорема 7.9

Пусть задача, двойственная к глобальной, имеет оптимальное допустимое решение y , а подзадачи имеют тот же вид, что и в теореме 7.8. Тогда система координируема на основе принципа согласования взаимодействий и координирующий сигнал $\hat{\beta}$ из U , где

$$\hat{\beta}_i = B_{ii}^T \hat{y}_i \text{ для каждого } i, 1 \leq i \leq n, \quad (7.22)$$

является оптимальным.

Доказательство. Пусть $\hat{m}$ и $\hat{y}$ — оптимальные допустимые решения прямой и двойственной глобальных задач соответственно. Тогда $(\hat{m}, \hat{y})$ — седловая точка функционала $\Phi(m, y) = c^T m + y^T (b - Am)$,

$$c^T \hat{m} + y^T (b - A\hat{m}) \leq c^T \hat{m} + \hat{y}^T (b - A\hat{m}) \leq c^T m + y^T (b - Am),$$

для всех $m \geq 0$ и $y \geq 0$. Вследствие использованного нами способа декомпозиции ограничений имеем

$$y^T (b - Am) = \sum_{i=1}^n [y_i^T (b_i - A_{ii}m_i - B_{ii}u_i) + y_i^T B_{ii} (u_i - K_i m)],$$

и, поскольку $\hat{y}^T (b - A\hat{m}) = 0$, имеем

$$c^T \hat{m} \leq c^T m + \sum_{i=1}^n [\hat{y}_i^T (b_i - A_{ii}m_i - B_{ii}u_i) + \hat{y}_i^T B_{ii} (u_i - K_i m)]$$

для всех $m \geq 0$, причем на u_i ограничения не накладываются. Таким образом, если $A_{ii}m_i + B_{ii}u_i \geq b_i$ для каждого $i, 1 \leq i \leq n$, то

$$\begin{aligned} c^T \hat{m} &\leq c^T m + \sum_{i=1}^n \hat{y}_i^T B_{ii} (u_i - K_i m) = \\ &= c^T m + \sum_{i=1}^n [\hat{y}_i^T B_{ii} u_i - \sum_{j=1}^n \hat{y}_j^T B_{jj} K_{ji} m_i] = \\ &= \sum_{i=1}^n g_i \beta (m_i, u_i), \end{aligned}$$

где $\hat{\beta}_i = B_{ii}^T \hat{y}_i, 1 \leq i \leq n$. Следовательно, для каждого $i, 1 \leq i \leq n$, пара $(\hat{m}_i, \hat{u}_i)$, где $\hat{u}_i = K_i \hat{m}$, является оптимальным допустимым решением i -й подзадачи для координирующего сигнала $\hat{\beta}$, определяемого выражением (7.22). Этим доказательство завершается.

Следует отметить, что при координирующем сигнале $\hat{\beta}$, определяемом соотношением (7.22), каждая подзадача имеет оптималь-

ное допустимое решение, если глобальная задача имеет оптимальное допустимое решение. Однако может случиться, что для других координирующих сигналов некоторые из подзадач не имеют оптимальных решений.

Применение принципа прогнозирования

К сожалению, большинство предыдущих результатов, касающихся принципа прогнозирования, нельзя непосредственно использовать в нашем случае. Это легко понять, если вспомнить, что целевые функции g_i вида (7.18) зависят только от локальных переменных m_i , так что линейные операторы оценки косвенного эффекта управляющих воздействий Γ'_i тождественно равны нулю. Тем не менее, чтобы скоординировать систему, приходится несколько модифицировать целевые функции подзадач. В данном случае взаимодействие проявляется в форме ограничений и отражается на функциях качества только через ограничения. Поэтому сразу не ясно, как задать соответствующие операторы оценки эффекта взаимодействия, чтобы нужным образом изменить целевые функции подзадач. Допустим, что целевые функции подзадач имеют вид

$$g_{i\beta}(m_i) = c_i^T m_i + \beta_i^T m_i, \quad (7.23)$$

где $\beta = (\beta_1, \dots, \beta_n)$ — некоторый координирующий сигнал из пространства $\mathcal{R} = M$. Поскольку значения β не определяются линейными операторами оценки косвенного эффекта управляющих воздействий, мы должны определить какой-то иной способ выбора β . Сделаем это, построив отображение $\eta: M \rightarrow \mathcal{R}$.

Без потери общности мы можем теперь записать ограничения в следующей форме:

$$m \geq 0, \quad A_i m = b_i, \quad 1 \leq i \leq n.$$

Очевидно, это видоизменение не нарушает общности, но приводит к существенному упрощению. Для i -й подзадачи ограничения примут вид

$$m \geq 0, \quad A_{ii} m_i + B_{ii} \alpha_i = b_i, \quad (7.24)$$

где α_i — определенное прогнозное значение связующего входа. Для данной пары (α, β) i -я подзадача состоит в минимизации функции (7.23) при условиях (7.24).

Можно привести два соображения, помогающие установить, каковым должно быть отображение η для того, чтобы эти подзадачи были координируемы на основе принципа прогнозирования при использовании отображения η .

1. Известно, что двухуровневая система с линейными модификациями локальных функций качества с нулевой суммой, координируемая на основе принципа согласования взаимодействий,

координируема и с помощью принципа прогнозирования при использовании некоторого отображения. 2. С другой стороны, отображение η можно получить, выразив с помощью теоремы 7.9 оптимальный координирующий сигнал, необходимый при использовании принципа согласования взаимодействий, через оптимальное допустимое решение двойственной глобальной задачи. В частности, для наших подзадач координирующий сигнал выражается через двойственные переменные следующим образом:

$$\beta_i = - \sum_{j \neq i} A_{ji}^T y_j, \quad (7.25)$$

где $y = (y_1, \dots, y_n)$ — двойственная переменная глобальной задачи, или, что эквивалентно, y_i — двойственная переменная i -й подзадачи.

Соотношение (7.25) определяет функцию $\zeta: y \rightarrow \mathcal{B}$, выражаемую через свои компоненты в виде

$$\zeta_i(y) = - \sum_{j \neq i} A_{ji}^T y_j \quad (7.26)$$

для всех $y = (y_1, \dots, y_n)$ из Y . Следующая теорема показывает связь этой функции с задачей координации.

Теорема 7.10

Пусть $\gamma = (\alpha, \beta)$ — определенная пара в $\mathcal{A} \times \mathcal{B}$, и пусть m^v и y^v таковы, что для всех i , $1 \leq i \leq n$, m_i^v — оптимальное допустимое решение i -й подзадачи, а y_i^v — оптимальное допустимое решение соответствующей двойственной задачи. Тогда m^v будет оптимальным допустимым решением глобальной задачи, если выполняются условия

$$\alpha = K(m^v) \quad \text{и} \quad \beta = \zeta(y^v). \quad (7.27)$$

Более того, если глобальная задача имеет оптимальное допустимое решение, то существует пара $\gamma = (\alpha, \beta)$ в $\mathcal{A} \times \mathcal{B}$, такая, что (7.27) выполняется для некоторых m^v и y^v .

Доказательство. Для доказательства первого утверждения допустим, что пара $\gamma = (\alpha, \beta)$ из $\mathcal{A} \times \mathcal{B}$ задана и условия (7.27) выполнены для некоторых m^v и y^v . Тогда для всех i , $1 \leq i \leq n$,

$$c_i + \beta_i - A_{ii}^T y_i^v \geq 0,$$

$$(m_i^v)^T (c_i + \beta_i - A_{ii}^T y_i^v) = 0$$

и, разумеется, выполняются ограничения (7.24). Если $\alpha = K(m_v)$, то, поскольку ограничения являются согласованными, мы вправе заключить, что m^v — допустимое решение глобальной задачи. Предположим теперь, что $\beta = \zeta(y^v)$. Тогда, используя (7.26), получим для всех i , $1 \leq i \leq n$,

$$c_i - \sum_{j=1}^n A_{ji}^T y_j^* \geq 0, \quad (m^*)^T (c_i - \sum_{j=1}^n A_{ji}^T y_j^*) = 0.$$

Следовательно, m^* удовлетворяет необходимым и достаточным условиям для оптимального допустимого решения глобальной задачи.

Для доказательства второй части теоремы возьмем оптимальные допустимые решения прямой и двойственной глобальных задач m^* и y^* , и пусть пара (α, β) такова, что $\alpha = K(\hat{m})$ и $\beta = \zeta(\hat{y})$. Тогда очевидно, что для всех i , $1 \leq i \leq n$, $\hat{m}_i$ и $\hat{y}_i$ будут соответственно оптимальными допустимыми решениями i -й подзадачи, отвечающей паре (α, β) , и двойственной ей подзадачи. Этим доказательство завершается.

Можно считать, что подзадачи как бы состоят из двух задач: данной и двойственной ей. Действительно, если для решения подзадач применяется симплекс-метод, при этом получается также оптимальное допустимое значение двойственной переменной. Более того, переменные можно рассматривать как переменные взаимодействия (связующие входы) двойственных подзадач. Таким образом, выбор пары (α, β) из $\mathcal{A} \times \mathcal{B}$ можно считать предсказанием связующих входов как исходных, так и двойственных им подзадач.

Перейдем теперь к отысканию функции $\eta: M \rightarrow \mathcal{B}$, при использовании которой применим принцип прогнозирования. Пусть θ — заданная функция перехода от $\mathcal{A}$ к множеству двойственных переменных $\mathcal{Y}$ и пусть для всех i , $1 \leq i \leq n$, функции η_i определены на M в виде композиции

$$\eta_i(m) = \zeta_i(\theta(K(m))), \quad (7.28)$$

где ζ_i определяются из (7.26); пусть, далее, η — функция вида

$$\eta(m) = (\eta_1(m), \dots, \eta_n(m)). \quad (7.29)$$

Координатор использует функции $\zeta_i(\alpha) = \zeta_i(\theta(\alpha))$, чтобы для каждого α из $\mathcal{A}$ поставить i -ю подзадачу в следующем виде:

$$\text{Минимизировать } g_{i\alpha}(m_i) = c_i^T m_i + \zeta_i(\alpha)^T m_i \quad (7.30)$$

при ограничениях (7.24). Эти подзадачи имеют такой же вид, как и раньше, с той лишь разницей, что теперь β_i , фигурирующие в целевых функциях, выражены как функции от α . Задача координации, таким образом, свелась к задаче выбора подходящего α из множества $\mathcal{A}$.

Для удобства изложения обозначим через m_i^α и y_i^α соответственно оптимальные допустимые решения прямой и двойственной подзадач с номером i , определяемых условиями (7.30) и (7.24). Пусть также $m^\alpha = (m_1^\alpha, \dots, m_n^\alpha)$; $y^\alpha = (y_1^\alpha, \dots, y_n^\alpha)$.

Если функция θ такова, что $\theta(\alpha) = y^\alpha$ для всех α из $\mathcal{A}$, то принцип прогнозирования взаимодействий может быть применен для координации подзадач при использовании отображения η . Пусть для некоторого α из $\mathcal{A}$ m^α и y^α существуют. Тогда из теоремы (7.10) и условия (7.28) получаем, что m^α — оптимальное допустимое решение глобальной задачи, если $\alpha = K(m^\alpha)$.

Чтобы подзадачи были координируемы с помощью принципа прогнозирования при использовании отображения η , нужно наложить на θ более жесткие ограничения. А именно, из теоремы 7.10 видно, что функция θ должна быть такой, чтобы для некоторых α из $\mathcal{A}$ выполнялись условия

$$\theta(\alpha) = \hat{y} \text{ и } \alpha = K(\hat{m}), \quad (7.31)$$

где $\hat{m}$ и $\hat{y}$ — оптимальные допустимые решения прямой и двойственной глобальных задач.

Подводя итог всему сказанному, можно сформулировать следующую теорему.

Теорема 7.11

Пусть функция η определена соотношениями (7.28) и (7.29), а θ — такая заданная функция, осуществляющая переход от $\mathcal{A}$ к множеству двойственных переменных Y что $\theta(\alpha) = y^\alpha$ для всех α из $\mathcal{A}$. Тогда для координации подзадач применим принцип прогнозирования взаимодействий (при использовании отображения η), т. е. для любого α из $\mathcal{A}$ m^α будет оптимальным допустимым решением глобальной задачи, если выполняется условие

$$\alpha = K(m^\alpha). \quad (7.32)$$

Кроме того, если θ удовлетворяет условиям (7.31) для некоторого α из $\mathcal{A}$, то подзадачи координируемы на основе принципа прогнозирования взаимодействий при использовании отображения η , т. е. существует такое α из $\mathcal{A}$, что выполняется (7.32).

5. ЗАМЕЧАНИЕ ПО ПОВОДУ ДЕКОМПОЗИЦИИ ДАНЦИГА — ВОЛЬФА

В отличие от декомпозиции, описанной в предыдущих разделах, процедура декомпозиции задач линейного программирования Данцига — Вольфа не влечет за собой полной децентрализации или декомпозиции. При декомпозиции задачи методом Данцига — Вольфа допустимое оптимальное решение глобальной задачи получается в виде выпуклой комбинации оптимальных допустимых решений подзадач. Координатор здесь берет на себя роль руководителя, который вырабатывает наилучшее общее глобальное действие, komponуя его из действий, предложенных подчиненными. Этот метод особенно хорошо приспособлен для решения тех задач линейного программирования, ограничения которых имеют блочную структуру.

Процедура декомпозиции Данцига — Вольфа

Рассмотрим задачу линейного программирования в переменных m ($m_1, \dots, m_n$):

$$\text{Минимизировать } c^T m = \sum_{i=1}^n c_i^T m_i \quad (7.33)$$

при ограничениях

$$A m = a, \quad (7.34)$$

$$B_i m_i = b_i, \quad m_i \geq 0, \quad i = 1, \dots, n, \quad (7.35)$$

где m принадлежит евклидову пространству M :

$$M = M_1 \times \dots \times M_n,$$

и каждое из пространств M_i может быть многомерным. Весовые коэффициенты c_i , $1 \leq i \leq n$, суть заданные векторы в пространствах M_i , а постоянные a и b_i — векторы в евклидовых пространствах Y_0 и Y_i соответственно. Предполагается, что матрицы ограничений A_i и B_i , $1 \leq i \leq n$, имеют соответствующую размерность.

Задачу (7.33) — (7.35) мы будем рассматривать как глобальную и будем предполагать, что она имеет допустимое решение.

При декомпозиции глобальной задачи по методу Данцига — Вольфа блочная структура ограничений используется для формулирования подзадач. Учитывая специфику глобальной задачи, мы сформулируем n таких подзадач, причем i -я задача имеет вид:

$$\text{Минимизировать } c_i^T m_i - \pi^T A_i m_i$$

при ограничениях

$$m_i \geq 0, \quad B_i m_i = b_i,$$

где π — координационная переменная, которая определяется решением основной задачи.

Основная задача при декомпозиции Данцига — Вольфа представляет собой задачу линейного программирования в переменных λ_i , $1 \leq i \leq p$:

$$\text{Минимизировать } f_1 \lambda_1 + \dots + f_p \lambda_p \quad (7.36)$$

при ограничениях

$$F_1 \lambda_1 + \dots + F_p \lambda_p = a, \quad (7.37)$$

$$\lambda_1 + \dots + \lambda_p = 1, \quad \lambda_i \geq 0, \quad i = 1, \dots, p, \quad (7.38)$$

где для каждого i , $1 \leq i \leq p$,

$$f_i = c^T \hat{m}_i \quad \text{и} \quad F_i = A \hat{m}_i,$$

а $(\hat{m}^1, \dots, \hat{m}^p)$ — множество угловых точек области M_B , т. е. множество всех m из M , удовлетворяющих ограничениям (7.35).

Определенная таким образом основная задача «эквивалентна» глобальной, если область M_B ограничена. В этом случае M_B

представляет собой выпуклую оболочку множества $(\hat{m}^1, \dots, \hat{m}^p)$. Тогда допустимой областью глобальной задачи служит пересечение M_B с множеством всех тех m в $\bar{M}$, для которых выполняются ограничения (7.34), т. е. $M = M_B \cap \{m: Am = a\}$. Таким образом, если λ — допустимое решение основной задачи, то λ удовлетворяет ограничениям (7.37) и (7.38) и $m = \lambda_1 \hat{m}^1 + \dots + \lambda_p \hat{m}^p$ является допустимым решением глобальной задачи. Более того, некоторый вектор $\hat{m}$ из $\bar{M}$ является оптимальным допустимым решением глобальной задачи тогда и только тогда, когда существует оптимальное допустимое решение $\hat{\lambda}$ основной задачи и $\hat{m} = \hat{\lambda}_1 \hat{m}^1 + \dots + \hat{\lambda}_p \hat{m}^p$.

Если множество M_B не ограничено, нужно добавить к множеству $\{\hat{m}^1, \dots, \hat{m}^p\}$ все лучи, служащие ребрами M_B , и модифицировать условие (7.38) так, чтобы λ_i давали выпуклую комбинацию угловых точек плюс неотрицательную комбинацию этих лучей. Для простоты мы в дальнейшем будем считать M_B ограниченным.

Найдем опорное решение основной задачи. Если ранг матрицы ограничений основной задачи F_i равен k , то допустимое опорное решение включает лишь k угловых точек M_B . Используем теперь симплекс-метод, чтобы улучшить опорное допустимое решение, включив в него еще одну угловую точку M_B . Такая новая угловая точка получается в результате решения подзадач. Если такое добавление не приводит к улучшению, опорное допустимое решение является оптимальным.

Пусть допустимое опорное решение основной задачи задано, причем $\mathcal{F}$ — соответствующий базис (матрица ограничений), а f — вектор весовых коэффициентов f_i . Чтобы узнать, можно ли улучшить данное опорное решение, необходимо лишь проверить, выполняется ли для каждого i , $1 \leq i \leq p$, неравенство

$$0 \leq f_i - \mathcal{F}_i^T \Pi. \quad (7.39)$$

Здесь $\mathcal{F}_i$ — i -й столбец матрицы ограничений основной задачи, а вектор $\Pi = (\mathcal{F}^{-1})^T f$. Если (7.39) выполнено для всех i , $1 \leq i \leq p$, то улучшить данное решение нельзя, т. е. это допустимое опорное решение оптимально. Правая часть неравенства (7.39) называется относительным весовым коэффициентом для λ_i . Компоненты Π называются симплексными множителями, они всегда могут быть найдены с помощью симплекс-метода. Разобьем Π на вектор π и скаляр π_0 , так что $(\pi, \pi_0) = \Pi$. Отметим, что

$$\pi_0 \geq \min_{M_B} (c^T - \pi^T A) m$$

тогда и только тогда, когда для всех i , $1 \leq i \leq p$, выполняется условие (7.39).

При декомпозиции Данцига — Вольфа координатор использует основную задачу, во-первых, чтобы найти выпуклую комбинацию угловых точек M_B , минимизирующую глобальную целевую функцию при ограничениях $At = a$, и, во-вторых, чтобы определить координирующий вектор π , который используется при модифицировании целевых функций для подзадач. Локальные решающие элементы, в обязанности которых входит решение подзадач, в свою очередь решают собственные подзадачи и, таким образом, находят для координатора угловые точки множества M_B . Итерация с участием координатора и локальных решающих элементов протекает следующим образом:

1. Предположим, что имеется какое-то количество угловых точек, например множество $\hat{m}^1, \dots, \hat{m}^k$, достаточное для получения допустимого опорного решения основной задачи. Координатор находит допустимое опорное решение основной задачи, скажем $(\lambda_1, \dots, \lambda_k)$, и вектор симплексных множителей $\Pi = (\pi, \pi_0)$. Вектор π сообщается локальным решающим элементам.

2. Получив вектор π , каждый из n локальных решающих элементов решает свою подзадачу. В результате получается угловая точка $\hat{m}^{k+1} = (\hat{m}_1^{k+1}, \dots, \hat{m}_n^{k+1})$ множества M_B , где $\hat{m}_i^{k+1}$ — оптимальное допустимое решение i -й подзадачи. Эта точка дает новое значение целевой функции v^k :

$$v^{k+1} = (c^T - \pi^T A) \hat{m}^{k+1}.$$

Точка $\hat{m}^{k+1}$ и значение функции v^{k+1} отсылаются к координатору.

3. Координатор сравнивает π_0 и v^{k+1} . Если $\pi_0 \geq v^{k+1}$, то допустимое опорное решение $(\lambda_1, \dots, \lambda_k)$ оптимально и $\hat{m} = \lambda_1 \hat{m}^1 + \dots + \lambda_k \hat{m}^k$ — оптимальное допустимое решение глобальной задачи. Если же $\pi_0 < v^{k+1}$, координатор берет новую угловую точку $\hat{m}^{k+1}$ и улучшает допустимое опорное решение основной задачи.

Сравнение декомпозиции Данцига — Вольфа и декомпозиции, полученной методом согласованных ограничений

Описанная ранее декомпозиция задач линейного программирования, которая приводит к согласованным ограничениям, представляет собой декомпозицию полной матрицы коэффициентов ограничений. В рассмотренном случае такая декомпозиция глобальной задачи дает n ограничений вида

$$A_i m_i + u_i = a_i, \quad B_i m_i = b_i, \quad m_i \geq 0.$$

Декомпозиция по методу Данцига — Вольфа также приводит к n ограничениям вида

$$B_i m_i = b_i, \quad m_i \geq 0.$$

Очевидно, что здесь проявляется «содержательное» различие между декомпозициями, осуществляемыми согласно этим двум методам. Действительно, при декомпозиции Данцига — Вольфа ограничения, связанные с взаимодействиями, не подвергаются декомпозиции.

Кроме того, для применения принципов координации локальные задачи ставятся так, чтобы при оптимальном координирующем сигнале оптимальные локальные решения давали оптимальное решение глобальной задачи. При декомпозиции же Данцига — Вольфа решения подзадач не дают в явном виде допустимого оптимального решения глобальной задачи: допустимым оптимальным решением глобальной задачи является выпуклая комбинация решений подзадач. В методе Данцига — Вольфа не фигурируют такие понятия, как оптимальный координирующий сигнал или параметр; существенно, что необходима некоторая последовательность координирующих сигналов или параметров. Очевидно, что принципы координации, введенные в гл. 5, основываются на предпосылках, отличных от предпосылок, используемых при декомпозиции Данцига — Вольфа.

Глава 8

«ЛИНЕЙНАЯ» КООРДИНАЦИЯ: КООРДИНАЦИЯ В ЦЕЛЯХ УЛУЧШЕНИЯ ХАРАКТЕРИСТИК ФУНКЦИОНИРОВАНИЯ СИСТЕМЫ

В этой главе мы рассмотрим проблему координирования двухуровневой системы в случае, когда ставится цель только улучшить качество работы системы в целом, не стремясь к обязательному достижению «глобального оптимума». Координатор должен так повлиять на выбор локальными управляющими элементами своих решений, чтобы получающиеся при этом управляющие воздействия приводили без дальнейшего вмешательства координатора к улучшению нужных характеристик процесса в целом. «Улучшение» можно оценить, сравнив фактическое значение глобальной функции качества с ее значением при заданном эталонном управляющем воздействии. За «эталон» можно, например, взять управление (набор управляющих воздействий), которое было бы приложено к процессу при отсутствии координирующих воздействий. Когда присутствуют возмущения, улучшение определяется для заданного диапазона возмущений. При описанном подходе локальные задачи формулируются как задачи получения локально улучшенного или удовлетворительного, но не обязательно оптимального управления.

1. ВВЕДЕНИЕ

При рассмотрении многоуровневых систем часто приходится отказываться от требования строгой глобальной оптимальности управляющих воздействий и локальных решений. Дело в том, что в практических ситуациях строгий оптимум по многим причинам оказывается нереализуемым. Чаще всего это связано с недостаточностью информации о факторах, влияющих на результаты выбранных решений или управляющих воздействий. В классических ситуациях управления и принятия решений использование алгоритмов оптимизации оправдывается в первую очередь тем фактом, что они разрешают некоторые проблемы, связанные с имеющимися в данной ситуации неопределенностями. Используя так называемый алгоритм оптимизации, можно выбрать последовательность действий, которая приводит к нужным результатам, если информация и гипотезы, на которых основан алгоритм, достаточно точны. При постановке задачи координации двухуровневой системы мы исходим из предположения о том, что локальные репающие элементы действительно существуют и отвечают

за управление подпроцессами, в то время как прямая обязанность координатора состоит в том, чтобы «согласовывать» деятельность локальных решающих элементов в процессе выработки ими своих решений с целью повышения суммарного эффекта их действий. Оправданием для создания двухуровневой системы является достигаемое при этом улучшение общего поведения системы, поэтому показатель эффективности работы координатора должен основываться на некоторой мере улучшения работы системы в целом, а не на факте достижения «глобальной» оптимальности. Даже в случаях, когда полная информация доступна и глобальной задачей является задача оптимизации, строгий оптимум может оказаться недостижимым из-за ограничений во времени или ограниченных возможностей локальных решающих элементов; поэтому действия координатора должны быть направлены именно на улучшение выбранных характеристик процесса в целом (значений глобальной функции качества).

Понятие координации с целью улучшения характеристик работы системы в целом вполне согласуется с общей концепцией координации, изложенной в гл. 4. Напомним, что сформулированный в гл. 4 постулат «согласованности» требует лишь, чтобы координатор воздействовал на локальные решающие элементы с целью решения задачи, стоящей перед всей системой (глобальной задачей). Но эту задачу можно, конечно, сформулировать в виде требования получить улучшенные характеристики вместо оптимальных. Фактически мы в этой главе рассматриваем «глобальную» задачу именно в такой постановке, а не как задачу оптимизации.

Понятие координации в целях улучшения глобальных характеристик системы отражает реальные ситуации во многих областях человеческой деятельности. Например, вопрос о децентрализации в экономических системах (в том виде, как он поставлен, в частности, в гл. 2) теоретически трактуется как проблема достижения оптимума экономического состояния. Преимущество децентрализации обосновывается тем, что децентрализованными действиями в экономике может быть достигнут глобальный оптимум. Однако, как указывает Саймон [13], ценность децентрализованной модели рынка или экономики состоит в том, что она обеспечивает согласование локальных и глобальной целей. Хотя каждый участник старается улучшить «свое собственное благополучие», одновременно улучшаются характеристики экономики в целом, так что в результате (почти) каждый выигрывает. Именно эта согласованность целей, выраженная в постулате совместимости, служит реальным оправданием децентрализации. Со сходной ситуацией приходится сталкиваться и при автоматизации промышленных процессов крупного масштаба. Первый эшелон управления необходим для обеспечения правильного протекания основных технологических процессов, и в общем случае он существует даже еще

до того, как кто-то начинает думать о создании высших эшелонов управления, деятельность которых связана с улучшением хода процесса с точки зрения его экономических показателей.

В этой главе мы рассмотрим два вопроса, касающихся практических способов осуществления координации для улучшения глобальных характеристик системы.

1. Каким образом координатор должен повлиять на выбор локальными решающими элементами соответствующих локальных управляющих воздействий, чтобы после того, как к ним придет координирующий сигнал, эти элементы вырабатывали такие локальные управляющие воздействия, которые без дальнейшего вмешательства координатора приводили бы к улучшению глобальных характеристик всего процесса в целом?

2. Если задан интервал времени, на протяжении которого ведется наблюдение за поведением функции качества двухуровневой системы, причем интервал этот содержит n моментов времени, в которые координатор может воздействовать на локальные решающие элементы, возникает вопрос: какова должна быть стратегия координатора, чтобы конечным результатом его деятельности явилось улучшение глобальных характеристик или же, при отсутствии противодействующих влияний, чтобы глобальные характеристики в данном интервале времени улучшались монотонно?

Эти два вопроса тесно связаны между собой: решение первого из них дает указания для выбора стратегии при решении второго вопроса. Использование координатором стратегии, основанной на «корректировке» подаваемых координирующих воздействий в определенные (последовательные) моменты времени, будет называться последовательной («секвенциональной») координацией. Если после каждого момента координирования локальные решающие элементы применяют к процессу свои управляющие воздействия без дальнейшего вмешательства координатора, мы будем называть такую координацию просто «линейной» координацией.

Термин «линейная» координация употребляется, чтобы подчеркнуть тот факт, что преимущества, предоставляемые наличием каналов обратной связи от локальных решающих элементов к координатору, никак не используются в промежутки времени от момента получения локальными решающими элементами координирующего сигнала вплоть до момента подачи на процесс выбранных этими элементами управляющих воздействий. В результате координатор в каждый момент координирования сталкивается с проблемой принятия решений в условиях неопределенности. Следует особо подчеркнуть, что координатор, чтобы удовлетворить принципу совместимости, должен рассматривать задачу каждого локального решающего элемента как задачу принятия решений

в условиях неопределенности, даже если неопределенность имеет место лишь в отношении действий остальных локальных решающих элементов.

Прежде чем перейти к изложению теории «линейной» координации, рассмотрим простой пример, иллюстрирующий некоторые из введенных понятий.

Рассмотрим двухуровневую систему, для которой процесс в целом P описывается линейным преобразованием

$$y = At \equiv P(t),$$

где управляющее воздействие t и выход y суть вещественные трехмерные векторы, а A — вещественная матрица $[a_{ij}]$ размерностью 3×3 . Для простоты примем, что внешние возмущения отсутствуют. Предположим, что глобальная функция качества задана и притом определена в виде функции общих (глобальных) затрат

$$G(t, y) = t^T t + (y - \tilde{y})^T (y - \tilde{y}),$$

где $\tilde{y}$ — заданный «эталонный» выход.

Предположим, что рассматриваемый процесс в целом распадается на три подпроцесса, из которых каждый i -й подпроцесс описывается уравнением

$$y_i = a_{ii}t_i + u_i \equiv P_i(t_i, u_i),$$

где управляющее воздействие t_i , связующий вход u_i и выход y_i — действительные числа. Когда эти подпроцессы связаны в один процесс и приложено управляющее воздействие $t = (t_1, t_2, t_3)$, связующий вход u_i для i -го подпроцесса выражается суммой

$$u_i = \sum_{j \neq i} a_{ij}t_j \equiv K_i(t).$$

Задача координации в данном случае заключается в следующем: исходя из некоторого эталонного управления $\tilde{t} = (\tilde{t}_1, \tilde{t}_2, \tilde{t}_3)$, координатор должен воздействовать на локальные решающие элементы, ответственные за управление подпроцессами, таким образом, чтобы они выработали управляющий сигнал $\hat{t} = (\hat{t}_1, \hat{t}_2, \hat{t}_3)$, который, будучи приложен к процессу, давал бы в результате более низкие затраты, чем при управлении $\tilde{t}$, т. е. чтобы для функции затрат выполнялось неравенство

$$G(\hat{t}, P(\hat{t})) < G(\tilde{t}, P(\tilde{t})).$$

Здесь возникают два вопроса: во-первых, каким образом локальные решающие элементы, действуя независимо друг от друга, производят свой выбор управляющих воздействий, и, во-вторых,

какой механизм координатор мог бы использовать для-воздействия на локальные решающие элементы.

Допустим, что локальные решающие элементы на основе имеющейся у них информации выбирают такое управляющее воздействие, которое улучшит их собственную функцию качества. Пусть i -й локальный решающий элемент использует в качестве такой функции функцию затрат $G_{i\mathcal{B}}$:

$$G_{i\mathcal{B}}(m_i, u_i, \beta_i) = 3m_i^2 + \beta_i m_i + (u_i - \tilde{u}_i)^2. \quad (8.1)$$

Сформулируем теперь i -ю локальную задачу следующим образом: для определенного эталонного управления $\tilde{m}_i$, определенного значения параметра $\tilde{\beta}_i$ и заданного («оценочного») диапазона $\tilde{U}_i$, в пределах которого может изменяться связующий вход u_i , найти такое управляющее воздействие $\tilde{m}_i$, чтобы выполнялось условие для всех u_i в оценочном диапазоне $\tilde{U}_i$, где

$$g_{i\mathcal{B}}(\hat{m}_i, u_i, \tilde{\beta}_i) < g_{i\mathcal{B}}(\tilde{m}_i, u_i, \tilde{\beta}_i),$$

$$g_{i\mathcal{B}}(m_i, u_i, \beta_i) = G_{i\mathcal{B}}(m_i, P_i(m_i, u_i), \beta_i).$$

Эти локальные задачи представляют собой так называемые задачи «удовлетворительного» управления, поскольку в них требуется не достижение оптимальных значений функций качества, а лишь получение удовлетворительного уровня значений этих функций при наличии некой неопределенности, характеризуемой в нашем случае диапазонами $\tilde{U}_i$ оценки возможных значений связующих входов. Однако если i -й локальный решающий элемент успешно решит свою задачу и реализующийся связующий вход лежит в оценочном диапазоне $\tilde{U}_i$, то выбранное управляющее воздействие действительно улучшит значение его собственной локальной функции качества.

Принцип оценки взаимодействий (4.18), введенный в гл. 4, гласит, что управляющее воздействие $\hat{m} = (\hat{m}_1, \hat{m}_2, \hat{m}_3)$, выбираемое при этих условиях локальными решающими элементами, приводит к снижению полных затрат, если связующие входы u_1 , u_2 и u_3 , появляющиеся в объединенной «связанной» системе при использовании управления $\hat{m}$, лежат соответственно в оценочных диапазонах $\tilde{U}_1$, $\tilde{U}_2$, $\tilde{U}_3$

$$K_i \hat{m} \in U_i, \quad i = 1, 2, 3.$$

Рассматривая вопрос о постановке локальных задач, можно перейти к более общему случаю, когда значения параметров β_i не фиксированы, а заданы лишь диапазоны $\mathcal{B}_i$, в которых эти параметры могут меняться. В этом случае i -я локальная задача формулируется так: для данного эталонного управления $\tilde{m}_i$ и за-

данных диапазонов $\tilde{\mathcal{X}}_i$ и $\tilde{U}_i$, в которых могут изменяться соответственно параметр β_i и связующий вход u_i , найти такое управление $\tilde{m}$, чтобы для всех β_i из $\tilde{\mathcal{X}}_i$ и всех u_i из $\tilde{U}_i$ выполнялось условие

$$g_i(\hat{m}_i, u_i, \beta_i) < g_i(\tilde{m}_i, u_i, \beta_i).$$

Правильность выбора диапазонов оценок для связующих входов определяется функциями взаимодействия подпроцессов K_i , которые выражают зависимость фактических значений связующих входов в объединенной (связанной) системе от поданных на подпроцессы управляющих воздействий. Можно считать, что параметры β_i в выражении (8.1) описывают непосредственную взаимосвязь локальных функций качества, и ввести соответствующие функции η_i для описания зависимости этих параметров от приложенных управляющих воздействий. При таком предположении из принципа оценки (4.19) вытекает, что управляющее воздействие $\hat{m}_i$ приводит к уменьшению общих затрат, если

$$K_i(\hat{m}) \in \tilde{U}_i \text{ и } \eta_i(\hat{m}) \in \tilde{\mathcal{X}}_i, \quad i = 1, 2, 3, \quad (8.2)$$

т. е. диапазоны оценок $\tilde{U}_i$ и $\tilde{\mathcal{X}}_i$ выбраны правильно для всех i . Задача координатора при выбранном эталонном управлении $\tilde{m}$ состоит в назначении диапазонов $\tilde{U}_i$ и $\tilde{\mathcal{X}}_i$ таким образом, чтобы был применен принцип оценки взаимодействий.

Допустим, что матрица A имеет вид

$$A = \begin{bmatrix} 3 & 2 & 1 \\ 1 & 4 & 0 \\ 1 & 1 & 2 \end{bmatrix}$$

и пусть $\tilde{y} = (5, -5.2)$ — эталонный выход, а $\tilde{m} = (0, 0, 0)$ — эталонное управление. Связующие входы, появляющиеся в объединенной системе при использовании $\tilde{m}$, будут равны $\tilde{u}_1 = \tilde{u}_2 = \tilde{u}_3 = 0$. Нужно отметить, что локальные функции затрат $G_{i\mathcal{X}}$ [см. (8.1)] вводятся через функцию общих затрат, как указано в разд. 3 гл. 8, и поэтому удовлетворяют уравнению (8.20). В этом случае подходящий вид функций η_i дастся формулами (8.22):

$$\begin{aligned} \eta_1(m_1, m_2, m_3) &= 10m_2 + 4m_3 - 6, \\ \eta_2(m_1, m_2, m_3) &= 14m_1 + 8m_3 + 24, \\ \eta_3(m_1, m_2, m_3) &= 6m_1 + 4m_2 + 10. \end{aligned}$$

Оценочные диапазоны $\tilde{U}_i$ и $\tilde{\mathcal{X}}_i$ можно построить двумя способами.

1. Выберем окрестность $N(\tilde{m})$.

$$N(\tilde{m}) = \left\{ \max_i |m_i - \tilde{m}_i| \leq \frac{1}{2}, \quad i = 1, 2, 3 \right\},$$

данного эталонного управления $\tilde{m}$. Используя функции K_i и η_i , построим оценочные диапазоны $\tilde{U}_i$ и $\tilde{\mathcal{R}}_i$:

$$\tilde{U}_i = \{u_i : u_i = K_i(m), \quad m \in N(\tilde{m})\}, \quad (8.3)$$

$$\tilde{\mathcal{R}}_i = \{\beta_i : \beta_i = \eta_i(m), \quad m \in N(m)\} \quad (8.4)$$

для каждого $i = 1, 2, 3$. Отсюда следует, что оценочными диапазонами являются интервалы значений связующих входов u_i и параметров β_i , определяемые неравенствами

$$-\frac{3}{2} \leq u_1 \leq \frac{3}{2}, \quad -\frac{1}{2} \leq u_2 \leq \frac{1}{2}, \quad -1 \leq u_3 \leq 1,$$

$$-13 \leq \beta_1 \leq 1, \quad 13 \leq \beta_2 \leq 33, \quad 5 \leq \beta_3 \leq 15.$$

Каждый локальный решающий элемент решает свою задачу для данного эталонного управления $\tilde{m} = (0, 0, 0)$, учитывая (8.3) и (8.4), результатом чего будет управляющее воздействие $\hat{m} = \left(\frac{1}{3}, -\frac{1}{44}, \frac{1}{4}\right)$. Отметим, что значение $\hat{m}$ не является единственным. Затем координатор проверяет, удовлетворяются ли (8.2). Это имеет место, и поэтому мы можем заключить, что $\hat{m}$ дает меньшее значение затрат, чем $\tilde{m}$. Действительно, $G(\hat{m}, P(\hat{m})) = 42$, в то время как $G(\tilde{m}, P(\tilde{m})) = 54$.

2. Другой способ задания оценочных диапазонов состоит в том, чтобы с помощью определяемой нормы вместо окрестности $\tilde{m}$ определить окрестности $K_i(\tilde{m})$ и $\eta_i(\tilde{m})$. Для каждого $i = 1, 2, 3$ координатор выбирает пару положительных вещественных чисел $(\varepsilon_i, \delta_i)$ и строит оценочные диапазоны

$$\tilde{U}_i = \{u_i : |u_i - K_i(m)| \leq \varepsilon_i\},$$

$$\tilde{\mathcal{R}}_i = \{\beta_i : |\beta_i - \eta_i(m)| \leq \delta_i\}.$$

В нашем примере оказывается, что пары $(\varepsilon_1, \delta_1) = \left(\frac{3}{2}, 7\right)$, $(\varepsilon_2, \delta_2) = \left(-\frac{1}{2}, 11\right)$ и $(\varepsilon_3, \delta_3) = (1, 5)$ порождают те же самые оценочные диапазоны, что и выражения (8.3) и (8.4). Такой подход является эвристическим, поэтому при условии применимости принципа оценки взаимодействий успех координирования системы с целью улучшения значений глобальной функции качества зависит от выбора оценочных диапазонов. Однако нужно отметить, что в случае применимости принципа оценки взаимодействий (как это имело место в предыдущем примере) для того, чтобы сделать вывод об улучшении значений глобальной функции качества, нет необходимости вычислять ее, а достаточно произвести простую проверку с помощью «обратной связи» на второй уровень, т. е. проверить выполнение условия (8.2).

2. КОНЦЕПЦИИ «ЛИНЕЙНОЙ» КООРДИНАЦИИ С ПОЗИЦИЙ ОБЩЕЙ ТЕОРИИ СИСТЕМ

Понятие «линейной» координации, как указано в предыдущем разделе, подразумевает, что координирующие сигналы подаются дискретно в некоторые последовательные моменты времени и что, кроме того, после каждого момента координации локальные решающие элементы осуществляют выбранное ими управление без дальнейшего вмешательства координатора. Цель координатора в каждый момент координации состоит в том, чтобы повлиять на локальные решающие элементы так, чтобы их действия давали в результате улучшение глобальной функции качества. Глобальная задача, стоящая перед двухуровневой системой в каждый момент координации, является, таким образом, задачей улучшения или нахождения удовлетворительного управления, но не задачей оптимизации.

Основные понятия

Пусть двухуровневая система задана следующим образом. Процесс, управляемый двухуровневой иерархической системой, состоящей из координатора и локальных решающих элементов, определяется, как в гл. 4, отображением $P: M \times \Omega \rightarrow Y$ где Ω — множество внешних воздействий или, в более абстрактном виде, — заданное множество неопределенностей. Мы предполагаем, что существует глобальная функция затрат $g: M \times \Omega \rightarrow V$, где V — такое линейно упорядоченное множество, что для данного подмножества управляющих воздействий $M^I \subseteq M$ глобальная задача улучшения может быть сформулирована следующим образом: найти управляющее воздействие $\hat{m}$ в M , такое, что

$$g(\hat{m}, \omega) < g(\tilde{m}, \omega)$$

для всех ω на Ω , где $\tilde{m}$ — определенное эталонное управляющее воздействие, принадлежащее M . Множество M^I называется множеством *допустимых управляющих воздействий*.

Глобальную функцию затрат g можно выразить через процесс в целом P и некоторую функцию оценки затрат G , определенную на $M \times Y$ в виде

$$g(m, \omega) = G(m, P(m, \omega)). \quad (8.5)$$

Такая интерпретация не является необходимой, но может оказаться полезной из-за своей наглядности.

По отношению к глобальной функции затрат g мы теперь можем установить предпочтительность различных управляющих воздействий. Предпочтительность находится в прямой зависимости от значений g на Ω . Более строго говоря, g индуцирует в M упорядочение, выражаемое отношением $>$, такое, что для всех m и m'

из M

$$t > t' \text{ тогда и только тогда, когда } g(t, \omega) < g(t', \omega) \text{ для всех } \omega \text{ из } \Omega. \quad (8.6)$$

Если $t > t'$, мы будем говорить, что t предпочтительнее t' . Следовательно, глобальная задача улучшения формулируется так: известно эталонное управляющее воздействие $\tilde{t}$, найти в M^i управляющее воздействие, которое предпочтительнее $\tilde{t}$.

В общем случае вместо эталонного управляющего воздействия может быть задана специальная так называемая «критериальная» функция $\tau: \Omega \rightarrow V$, которая служит для определения допустимого уровня значения функции качества. В этом случае глобальная задача состоит в том, чтобы найти управляющее воздействие $\hat{t}$, при котором $g(\hat{t}, \omega) < \tau(\omega)$ для всех ω из Ω . Проблему такого типа мы рассматриваем не как задачу улучшения, а как задачу удовлетворительного управления, чтобы подчеркнуть тот факт, что эталонное управление не всегда определяет допустимый уровень функции качества.

Примем, что процесс в целом P состоит из n связанных между собой подпроцессов $P_i: M_i \times U_i \rightarrow Y_i$, как описано в гл. 4, где $M = M_1 \times \dots \times M_n$ и $Y = Y_1 \times \dots \times Y_n$. Предположим также, что связующие входы u_i для каждого i -го подпроцесса, $1 \leq i \leq n$, могут быть выражены в виде функции всех поданных управляющих воздействий и возмущений со стороны среды (или неопределенностей) с помощью отображения $K_i: M \times \Omega \rightarrow U_i$. Иначе говоря, если на процесс P подано управляющее воздействие t и реализуется ω из Ω , то связующий вход, имеющий место для i -го подпроцесса в объединенной («связанной») системе, есть $u_i = K_i(t, \omega)$. Таким образом, связующие входы для i -го подпроцесса объединенной системы в самом деле представляют собой воздействие его окружения, в которое входят не только остальные подпроцессы рассматриваемой двухуровневой системы, но и среда для всей двухуровневой системы.

Мы предполагаем далее, что для каждого i , $1 \leq i \leq n$, определена локальная функция затрат $g_{i\mathcal{B}}: M_i \times U_i \times \mathcal{B}_i \rightarrow V$, где $\mathcal{B}_i$ — некоторое множество значений параметров. i -ю локально решаемую задачу мы конкретизируем несколько позднее, используя определенные локальные функции затрат $g_{i\mathcal{B}}$ и выделенные подмножества множеств $U_i \times \mathcal{B}_i$. Эти подмножества мы будем называть *оценочными диапазонами*, считая, что именно они характеризуют ту неопределенность, которая имеет место при решении локальных задач.

Без ограничения общности можно принять, что локальная функция затрат $g_{i\mathcal{B}}$ для каждого i , $1 \leq i \leq n$, аналогично глобальной функции затрат может быть выражена через подпро-

цесс P_i и заданную функцию оценки затрат $G_{i\mathcal{B}}$, определенную на $M_i \times Y_i \times U_i \times \mathcal{B}_i$, т. е.

$$g_{i\mathcal{B}}(m_i, u_i, \beta_i) = G_{i\mathcal{B}}(m_i, P_i(m_i, u_i), u_i, \beta_i). \quad (8.7)$$

Так же, как и для глобальной функции затрат, эта интерпретация не необходима, но может оказаться весьма полезной.

Очевидно, что координатор может воздействовать на локальные решающие элементы, устанавливая для каждого из них оценочные диапазоны. Поэтому предположим, что у нас есть множество $\mathcal{C}$, состоящее из элементов γ , таких, что при любых γ для каждого i , $1 \leq i \leq n$, определено подмножество $U_i^\gamma \times \mathcal{B}_i^\gamma \subseteq U_i \times \mathcal{B}_i$. Элементы множества $\mathcal{C}$ мы будем называть координирующими сигналами: координатор выбирает γ из $\mathcal{C}$ и устанавливает, таким образом, для каждого i , $1 \leq i \leq n$, оценочный диапазон $U_i^\gamma \times \mathcal{B}_i^\gamma$.

Основное различие между способами координации, применяемыми здесь, и способами, изложенными в предыдущих главах, состоит в том, что оценочные диапазоны зависят от γ , но не являются в общем случае одноэлементными множествами, хотя множества $\mathcal{B}_i^\gamma$ и рассматриваются в частном случае как одноэлементные.

Мы будем анализировать в основном случай, когда глобальная задача поставлена как задача улучшения на основе использования определенного эталонного управления. Поэтому мы сначала введем понятие «линейной» координируемости при управлении на основе ранее выбранного эталонного воздействия, а затем рассмотрим более общие понятия.

Координируемость при использовании эталонного управляющего воздействия

Эталонное управляющее воздействие может быть выбрано самим координатором, или же им может служить управляющее воздействие, подаваемое на процесс в отсутствие координации. В последнем случае цель состоит в том, чтобы улучшить существующее управление.

Каждый координирующий сигнал из $\mathcal{C}$ через оценочные диапазоны и локальные функции затрат индуцирует на множестве M отношение предпочтительности. Для каждого γ из $\mathcal{C}$ введем отношение $(\succ_\gamma)$ на M , такое, что для всех m и m' из M

$$m (\succ_\gamma) m'$$

тогда и только тогда, когда для любого i , $1 \leq i \leq n$, неравенство

$$g_{i\mathcal{B}}(m_i, u_i, \beta_i) \leq g_{i\mathcal{B}}(m'_i, u_i, \beta_i)$$

справедливо в оценочном диапазоне $U_i^\gamma \times \mathcal{R}_i^\gamma$ и в то же время существует j , $1 \leq j \leq n$, для которого в оценочном диапазоне выполняется строгое неравенство

$$g_{j\mathcal{R}}(m_j, u_j, \beta_j) < g_{j\mathcal{R}}(m'_j, u_j, \beta_j).$$

Если $m (>_\gamma) m'$, мы будем говорить, что m γ -предпочтительнее m' .

Для определенного управления $\tilde{m}$ из M в двухуровневой системе координирующий сигнал γ из $\mathcal{E}$ называется $\tilde{m}$ -приемлемым, если существует допустимое управление, которое γ -предпочтительнее $\tilde{m}$.

Изложим теперь концепцию координируемости для управления на основе эталонного воздействия. Пусть $\tilde{m}$ — некоторое эталонное управление из M , а M^I — заданное множество допустимых управлений двухуровневой системы. Двухуровневая система называется *координируемой при $\tilde{m}$* тогда и только тогда, когда существует такой m -приемлемый координирующий сигнал γ из $\mathcal{E}$, что

$$[m (>_\gamma) \tilde{m}] \Rightarrow [m > \tilde{m}]$$

для всех m из M^I

Введенное понятие координируемости означает, что существуют какой-то координирующий сигнал из $\mathcal{E}$ и какое-то глобально допустимое управление, которые улучшают значения по крайней мере одной из локальных функций качества в заданном оценочном диапазоне и не ухудшают значений остальных локальных функций качества (в заданных для них оценочных диапазонах), причем все такие глобально допустимые управляющие воздействия (если их несколько) улучшают значения глобальной функции качества.

Локально решаемые задачи при управлении на основе эталонного воздействия с использованием отношений γ -предпочтительности на M и изложенного понятия координируемости формулируются очевидным образом. А именно, они представляют собой задачи улучшения, причем i -я локальная задача, $1 \leq i \leq n$, ставится так: для данного координирующего сигнала γ из $\mathcal{E}$ и заданного эталонного управляющего воздействия $\tilde{m}$ найти такое управляющее воздействие $\hat{m}_i$ из M_i^I , что

$$g_{i\mathcal{R}}(\hat{m}_i, u_i, \beta_i) \leq g_{i\mathcal{R}}(\tilde{m}_i, u_i, \beta_i)$$

для оценочного диапазона $U_i^\gamma \times \mathcal{R}_i^\gamma$, где M_i^I — подмножество M и представляет собой множество возможных значений i -й компоненты всех глобально допустимых управляющих воздействий. Таким образом сформулированную локальную задачу мы будем называть задачей, *определяемой парой* $(\gamma, \tilde{m})$.

Ясно, что для любой пары $(\gamma, \tilde{m})$ из $\mathcal{E} \times M^f$ управляющее воздействие $\tilde{m}_i$ есть решение i -й локальной задачи, определяемой этой парой, но управление $\tilde{m}$ не является γ -предпочтительным самому себе, так как это отношение неререфлексивно. С другой стороны, если пара $(\gamma, \tilde{m})$ из $\mathcal{E} \times M^f$ такова, что γ является $\tilde{m}$ -приемлемым, то по определению приемлемости каждая локальная задача, определяемая парой $(\gamma, \tilde{m})$, имеет такое решение $\tilde{m}_i$, что управление $\hat{m} = (\hat{m}_1, \dots, \hat{m}_n)$ допустимо и γ -предпочтительнее $\tilde{m}$; поэтому мы в соответствии с введенным выше понятием координируемости будем считать, что для любой такой пары $(\gamma, \tilde{m})$ из $\mathcal{E} \times M^f$ i -й локальный решающий элемент, $1 \leq i \leq n$, в состоянии найти требуемое решение поставленной перед ним задачи.

В основной своей части наш анализ будет сконцентрирован именно вокруг введенного выше понятия координируемости. Однако мы можем его расширить, введя понятие координируемости на заданном подмножестве M ; двухуровневая система называется *координируемой на множестве* $\tilde{M} \subseteq M$ тогда и только тогда, когда она координируема при каждом $\tilde{m}$ из $\tilde{M}$. Такое расширение понятия представляет интерес, поскольку, если система координируема на данном множестве $\tilde{M}$, координатор может выбрать любое управление $\tilde{m}$ из $\tilde{M}$ с уверенностью, что он сможет воздействовать на локальные решающие элементы так, чтобы они выработали глобально допустимое управление, которое улучшит глобальную функцию качества. Очевидно, что это представляет для координатора некоторую разновидность оптимизационной задачи; а именно координатор должен выбрать такое эталонное управляющее воздействие $\tilde{m}$ из $\tilde{M}$, чтобы получить от локальных решающих элементов максимальное улучшение. Однако координатор может быть уверен в некотором улучшении вне зависимости от того, какое управляющее воздействие из $\tilde{M}$ выбрано в качестве эталонного.

При применении изложенной концепции координируемости возникает трудность, состоящая в том, что после того, как к локальным решающим элементам поступят эталонное управление $\tilde{m}$ и координирующий сигнал γ , координатор должен располагать значениями функций затрат $g(\tilde{m}, \omega)$ и $g(\hat{m}, \omega)$, где $\hat{m}$ — управляющее воздействие, выбранное локальными решающими элементами, и иметь возможность сравнивать эти значения для каждого ω из Ω или того ω , которое фактически имеет место, чтобы установить, получено ли действительно улучшение. Хотя это может оказаться не столь сложной задачей, как определение оптимальности в случае решения задачи оптимизации, тем не менее желательно иметь простой тест, который позволял бы опре-

делять, скоординирована ли двухуровневая система или нет. Поэтому, так же как и в ранее изученных случаях оптимизации, мы рассмотрим принцип координации, который основан на взаимодействии, имеющем место между локальными элементами после того, как к процессу приложено выбранное управление.

Взаимодействия между подпроцессами проявляются в связующих входах, поскольку подпроцессы объединены именно посредством связующих входов. Эти взаимодействия описываются отображением $K: M \times \Omega \rightarrow U$, компонентами которого служат отображения K_i . Взаимосвязанность локальных функций качества является результатом взаимодействий между подпроцессами, но она может также проявляться через взаимозависимость значений параметров β_i , фигурирующих в локальных функциях затрат. Мы можем использовать эти взаимосвязи, чтобы установить условие координируемости и, следовательно, соответствующий принцип координации. В качестве принципа координации можно было бы использовать принцип оценки взаимодействий (4.18), который мы здесь представим в следующей форме:

$$(\forall \gamma) (\forall m) (\forall \omega) \{ [m (\succ_\gamma) \tilde{m} \text{ и } K(m, \omega) \in U^\gamma] \Rightarrow m \succ \tilde{m} \}, \quad (8.8)$$

где $U^\gamma = U_1^\gamma \times \dots \times U_n^\gamma$. Мы будем использовать также другую форму принципа оценки взаимодействий, которая включает отображение $\eta: M \times \Omega \rightarrow \mathcal{B}$, где $\mathcal{B}^\gamma = \mathcal{B}_1^\gamma \times \dots \times \mathcal{B}_n^\gamma$. Эта вторая форма принципа получается из (4.19) и может быть записана в следующем виде:

$$(\forall \gamma) (\forall m) (\forall \omega) \{ [m (\succ_\gamma) \tilde{m} \text{ и } q_\eta(m, \omega) \in U^\gamma \times \mathcal{B}^\gamma] \Rightarrow m \succ \tilde{m} \}, \quad (8.9)$$

где $q_\eta(m, \omega) = (K(m, \omega), \eta(m, \omega))$ и $\mathcal{B}^\gamma = \mathcal{B}_1^\gamma \times \dots \times \mathcal{B}_n^\gamma$.

При такой форме принципа оценки взаимодействий мы рассматриваем фигурирующие в локальных функциях затрат параметры β_i как взаимодействия, определяемые отображением η .

Для дальнейшего анализа нам понадобятся некоторые понятия, связанные с координируемостью двухуровневой системы на основе принципа оценки взаимодействий в случае, когда задано ограниченное множество M' глобально допустимых управляющих воздействий и выбрано некое эталонное управление $\tilde{m}$.

Принцип оценки взаимодействий применим при данном $\tilde{m}$ тогда и только тогда, когда предложение (8.8) истинно. Система называется координируемой при данном $\tilde{m}$ на основе этого принципа тогда и только тогда, когда принцип применим для $\tilde{m}$ и существует такая пара (γ, m) в $\mathcal{C} \times M$, которая делает истинным

предложение

$$(\exists \gamma) (\exists m) (\forall \omega) [m (\succ_{\gamma}) \tilde{m} \text{ и } K(m, \omega) \in U\gamma]. \quad (8.10)$$

Понятия применимости и координируемости для второй формы принципа оценки взаимодействий при использовании отображения $\eta: M \times \Omega \rightarrow \mathcal{R}$ определяются совершенно аналогично, за исключением того, что (8.10) заменяется выражением

$$(\exists \gamma) (\exists m) (\forall \omega) [m (\succ_{\gamma}) \tilde{m} \text{ и } q_{\eta}(m, \omega) \in U\gamma \times \mathcal{R}\gamma]. \quad (8.11)$$

Применимость принципа оценки взаимодействий в любой из его формулировок означает просто, что глобальные характеристики улучшаются всякий раз, когда взаимодействия, возникающие в результате приложения более γ -предпочтительного глобально допустимого управления, не выводят нас за пределы оценочных диапазонов. Для координируемости на основе этого принципа требуется выполнение более сильных условий по сравнению с указанными выше условиями просто координируемости, ибо в этом случае требуется, помимо выполнения условия применимости, также существование допустимого координирующего сигнала γ и γ -предпочтительного глобально допустимого управляющего воздействия, при котором получающиеся взаимодействия не выводят нас за пределы оценочных диапазонов. Если принцип применим, можно предложить простой тест, который позволит координатору определить, дадут ли действия локальных решающих элементов улучшение глобальных характеристик, причем такой тест не требует фактического сравнения глобальных затрат.

Принцип оценки взаимодействий обеспечивает основу, на которой можно построить стратегию упомянутой выше последовательной координации для получения оптимального координирующего сигнала в двухуровневой системе, рассмотренной в предыдущих главах. Если двухуровневая система координируема на основе этого принципа в каждый момент координирования, то система может быть скоординирована в каждый из этих моментов так, что улучшения глобальных характеристик будут монотонными. Такого рода итеративный процесс имеет определенное практическое значение при реализации «линейной» координации, поскольку при этом на каждой промежуточной стадии можно быть уверенным в улучшении по крайней мере по отношению к предыдущей стадии.

При практическом применении принципа оценки взаимодействий мы можем столкнуться с определенными трудностями, поскольку процесс решения локальных задач «улучшения» с целью получения γ -предпочтительного глобально допустимого управления может оказаться весьма сложным. Сложность возникает из-за требования улучшить по меньшей мере одну из локальных характеристик во всем оценочном диапазоне, не ухудшив в то же

время остальных. Если оценочные диапазоны сводятся к одноэлементным множествам, $U\gamma \times \mathcal{B}\gamma = \{(u\gamma, \beta\gamma)\}$, обычно удается улучшить значения локальных функций качества, но маловероятно, чтобы при этом было удовлетворено условие координации $K(m, \omega) \in U\gamma$ или $q_\eta(m, \omega) \in U\gamma \times \mathcal{B}\gamma$. С другой стороны, если оценочные диапазоны слишком велики, улучшение локальных функций качества может оказаться невозможным.

Условия координируемости при использовании эталонного управления

Применимость принципа оценки взаимодействий к двухуровневой системе зависит в первую очередь от взаимосвязи локальных функций затрат с глобальной функцией затрат и от выбранного множества глобально допустимых управляющих воздействий. Мы изложим здесь некоторые условия применимости этого принципа в том случае, когда за основу берется некое эталонное управление, локальные функции затрат согласованы с глобальной функцией затрат, а множество глобально допустимых управляющих входов принадлежит к определенному классу связанных подмножеств M ; для этого случая мы выведем условие координируемости системы. При этом подразумевается, что локальные и глобальные функции затрат принимают только вещественные значения. Локальная функция затрат $g_{i\mathcal{B}}$, $1 \leq i \leq n$, считается сбалансированной с глобальной функцией затрат g , если существуют два отображения $\eta_i: M \times \Omega \rightarrow \mathcal{B}_i$ и $h_i: M \times \Omega \rightarrow R$, такие, что

$$G_{i\mathcal{B}}(m_i, K_i(m, \omega), \eta_i(m, \omega)) + h_i(m, \omega) = g(m, \omega) \quad (8.12)$$

для всех m и всех ω из Ω .

Пример 8.1.

Рассмотрим двухуровневую систему с подпроцессами P_1 и P_2 , определенными на $R \times R$ в виде

$$P_1(m_1, u_1) = m_1^2 + u_1, \quad P_2(m_2, u_2) = m_2 u_2,$$

в то время как связующие входы задаются уравнениями

$$u_1 = m_2 + \omega \equiv K_1(m_1, m_2, \omega), \quad u_2 = m_1 + \omega \equiv K_2(m_1, m_2, \omega),$$

где ω — действительное число, характеризующее внешние возмущения. Предположим, что функция для оценки глобальных затрат задана в виде

$$G(m, y) = m^T m + (y_1 - \tilde{y}_1) + (y_2 - \tilde{y}_2).$$

Чтобы получить локальные функции затрат, которые были бы сбалансированы с глобальной функцией затрат g ,

$$g(m, \omega) = G(m, P(m, \omega)) = \\ = m_1^2 + m_2^2 + (m_1^2 + m_2 + \omega - \tilde{y}_1) + (m_2 m_1 + m_2 \omega - \tilde{y}_2).$$

мы определим функции для оценки локальных затрат $G_{1\mathcal{B}}$ и $G_{2\mathcal{B}}$ в виде

$$G_{1\mathcal{B}}(m_1, y_1, u_1, \beta_1) = m_1^2 + (y_1 - \tilde{y}_1) + \beta_1(m_1 + u_1),$$

$$G_{2\mathcal{B}}(m_2, y_2, u_2, \beta_2) = m_2^2 + (y_2 - \tilde{y}_2) + \beta_2 m_2,$$

а функции η_i и h_i выберем такими, чтобы для всех m_1, m_2 и ω выполнялись условия

$$\eta_1(m_1, m_2, \omega) = m_2, \quad h_1(m_1, m_2, \omega) = -\tilde{y}_2; \\ \eta_2(m_1, m_2, \omega) = 1, \quad h_2(m_1, m_2, \omega) = 2m_1^2 - m_1 - \tilde{y}_1.$$

Легко проверить, что для каждого $i = 1, 2$ локальная функция затрат $g_{i\mathcal{B}}$, выраженная через величины, характеризующие подпроцесс P_i и функцию для оценки затрат $G_{i\mathcal{B}}$, действительно сбалансирована с глобальной функцией затрат g .

С частными случаями сбалансированности локальных и глобальной функций затрат мы уже встречались ранее в предыдущих главах. Предположим, например, что глобальная функция затрат аддитивна, а модификации являются сбалансированными, и

$$g(m, \omega) = \sum_{i=1}^n g_{i\mathcal{B}}(m_i, K_i(m, \omega), \beta_i) \quad (8.13)$$

всякий раз, когда все β_i идентичны. В этом случае равенство (8.12) выполняется, если мы определим h_i как функцию вида

$$h_i(m, \omega) = \sum_{j \neq i} g_{j\mathcal{B}}(m_j, K_j(m, \omega), \beta_j),$$

где $\beta_i = \eta_i(m, \omega)$. Ясно, что локальные функции затрат в аддитивном случае (8.13) сбалансированы с глобальной функцией затрат.

Чтобы привести другой пример, определим глобальную и локальные функции затрат g и $g_{i\mathcal{B}}$, $1 \leq i \leq n$, через заданные функции для оценки затрат G и $G_{i\mathcal{B}}$ с помощью выражений (8.5) и (8.6) соответственно. Предположим, что $G_{i\mathcal{B}}$ порождается функцией G в том смысле, что подходящим образом определены функции θ_i , производящие отображение $M \times Y$ в $\mathcal{B}_i$, такие, что

$$G_{i\mathcal{B}}(m_i, y_i, \beta_i) = G(m, y),$$

когда $\beta_i = \theta_i(m, y)$. Тогда локальная функция затрат $g_{i,\beta}$ сбалансирована с g ; искомая функция η_i равна

$$\eta_i(m, \omega) = \theta_i(m, P(m, \omega)),$$

а h_i тождественно равны нулю.

Следует указать, что приведенные ниже условия применимости принципа оценки взаимодействий и координируемости на основе этого принципа требуют, чтобы локальные функции затрат были сбалансированы с глобальной функцией затрат. Более того, необходимо допустить, что отображения K_i , η_i и h_i не зависят от m_i , i -й компоненты управляющего воздействия m . Предположение о независимости отображений K_i от m_i просто означает, что взаимосвязи подпроцессов являются каноническими по отношению ко входам.

Чтобы гарантировать применимость принципа оценки взаимодействий, нам придется ограничить допустимые управления некоторыми специальными подмножествами M .

Введем на $M \times M$ такую целочисленную функцию ρ , что для любых m и m' из M значение $\rho(m, m')$ равно числу компонент, по которым m и m' отличаются друг от друга. Если $\rho(m, m') = 0$, то m и m' не отличаются ни одной из компонент и, следовательно, $m = m'$; $\rho(m, m') = k$ показывает, что m и m' различаются ровно в k компонентах; очевидно, максимальное значение ρ равно n . В действительности функция ρ является метрикой (функцией расстояния) в M . Очевидно также, что ρ неотрицательно, $\rho(m, m') = \rho(m', m)$ и $\rho(m, m') = 0$ тогда и только тогда, когда $m = m'$. Легко проверить, что ρ удовлетворяет неравенству треугольника.

Первым результатом, касающимся применимости принципа оценки взаимодействий при использовании некоего эталонного управления, является следующее предложение.

Теорема 8.1

Предположим, что локальные функции затрат двухуровневой системы сбалансированы с глобальной функцией затрат; для всех i , $1 \leq i \leq n$,

$$g_{i,\beta}(m_i, K_i(m, \omega), \eta_i(m, \omega)) + h_i(m, \omega) = g(m, \omega), \quad (8.14)$$

и отображения K_i , η_i и h_i не зависят от m_i . Пусть $\tilde{m}$ — выбранное эталонное управление, и пусть множество

$$M^f = \{m: \rho(m, \tilde{m}) \leq 1\} \quad (8.15)$$

является множеством допустимых управляющих воздействий. Тогда принцип оценки взаимодействий, выраженный условием (8.9), применим при эталонном управлении $\tilde{m}$.

Д о к а з а т е л ь с т в о. Чтобы доказать применимость принципа, допустим, что существуют такие γ из $\mathcal{E}$ и $\tilde{m}$ из M^f , что

(8.11) справедливо. Нам нужно показать, что $\hat{m} > \tilde{m}$. Учитывая (8.15), предположим, что $\hat{m}$ и $\tilde{m}$ различаются в i -й компоненте. Тогда

$$g_{i\mathcal{B}}(\hat{m}_i, u_i, \beta_i) < g_{i\mathcal{B}}(\tilde{m}_i, u_i, \beta_i)$$

в оценочном диапазоне $U_i^Y \times \mathcal{B}_i^Y$. Для всех $\omega \in \Omega$ из (8.14) получаем, что $K_i(\hat{m}, \omega)$ находится в U_i^Y , а $\eta_i(\hat{m}, \omega) \in \mathcal{B}_i^Y$; поэтому из (8.14) имеем

$$\begin{aligned} g(\hat{m}, \omega) &= g_{i\mathcal{B}}(\hat{m}_i, K_i(\hat{m}, \omega), \eta_i(\hat{m}, \omega)) + h_i(\hat{m}, \omega) < \\ &< g_{i\mathcal{B}}(\tilde{m}_i, K_i(\hat{m}, \omega), \eta_i(\hat{m}, \omega)) + h_i(\hat{m}, \omega) = \\ &= g_{i\mathcal{B}}(\tilde{m}_i, K_i(\tilde{m}, \omega), \eta_i(\tilde{m}, \omega)) + h_i(\tilde{m}, \omega) = g(\tilde{m}, \omega). \end{aligned}$$

Следовательно, $\hat{m} > \tilde{m}$, и теорема доказана.

Для получения следующего более общего результата, касающегося применимости принципа оценки взаимодействий, введем понятие ρ -связности.

Пара $\{m, \tilde{m}\}$ элементов множества M с расстоянием $\rho(m, \tilde{m}) = k > 0$ называется ρ -связанной в подмножестве M' множества M , если существует такая последовательность $\{m^0, \dots, m^k\}$ в M' где $m^0 = m$ и $m^k = \tilde{m}$, что

$$\rho(m^{i-1}, m^i) = 1 \text{ для } i = 1, \dots, k.$$

Такая последовательность называется ρ -связью пары $\{m, \tilde{m}\}$.

Теорема 8.2

Как и в предыдущей теореме, предположим, что локальные функции затрат согласованы с глобальной функцией затрат. Пусть $\tilde{m}$ — заданное эталонное управление и для каждого γ из $\mathcal{C}$ множество допустимых управляющих воздействий $\tilde{M}^\gamma$ имеет вид

$$\tilde{M}^\gamma = \{m : \{m, \tilde{m}\} \text{ являются } \rho\text{-связанными в } M^\gamma\}, \quad (8.16)$$

где

$$M^\gamma = \{m : \{K(m, \omega) : \omega \in \Omega\} \subseteq U^\gamma \text{ и } \{\eta(m, \omega) : \omega \in \Omega\} \subseteq \mathcal{B}^\gamma\}. \quad (8.17)$$

Тогда принцип оценки взаимодействий в формулировке (8.9) применим для $\tilde{m}$.

Доказательство. Предположим, что существуют такие γ из $\mathcal{C}$ и $\tilde{m}$ из $\tilde{M}^\gamma$, что (8.11) справедливо. Нам нужно показать, что $\hat{m} > \tilde{m}$. Пусть $\rho(\tilde{m}, \hat{m}) = k$. Если $k \leq 1$, то нужный нам результат следует из теоремы 8.1. Поэтому положим $k \geq 2$. Пусть $\{m^1, \dots, m^{k-1}\}$ — ρ -связь пары $\{\hat{m}, \tilde{m}\}$ в M^γ . Предположим, что

$\hat{m}$ и $\tilde{m}$ различаются в первых k компонентах; при каждом i , $1 \leq i \leq k$, имеем

$$g_{i\mathcal{B}}(\hat{m}_i, u_i, \beta_i) \leq g_{i\mathcal{B}}(\tilde{m}_i, u_i, \beta_i) \text{ на } U_i^\gamma \times \mathcal{B}_i^\gamma \quad (8.18)$$

со строгим неравенством по меньшей мере для одного i , $1 \leq i \leq k$.

Используя неравенство треугольника, мы можем показать, что $\rho(m^1, \tilde{m}) = k - 1$. Предположим, что $\hat{m}$ и m^1 различаются в i -й компоненте. Тогда $m_i^1 = \tilde{m}_i$; в противном случае, поскольку $m_j^1 = \hat{m}_j$ для $j \neq i$, мы имели бы $\rho(m^1, \tilde{m}) = k$, что противоречит $\rho(m^1, \tilde{m}) = k - 1$. Кроме того, по нашему предположению, $1 \leq i \leq k$. и, следовательно, из $m_i^1 = \tilde{m}_i$ и (8.18) получаем

$$g_{i\mathcal{B}}(\hat{m}_i, u_i, \beta_i) \leq g_{i\mathcal{B}}(m_i^1, u_i, \beta_i) \text{ на } U_i^\gamma \times \mathcal{B}_i^\gamma.$$

Обращаясь к доказательству теоремы 8.1, мы можем теперь заключить, что $\hat{m} \succcurlyeq m^1$, где отношение $\succcurlyeq$ означает $g(\hat{m}, \omega) \leq g(m^1, \omega)$ на Ω .

Заметим, что последовательность $\{m^2, \dots, m^{k-1}\}$ является ρ -связью для пары $\{m^1, \tilde{m}\}$. Поэтому, применяя подобные рассуждения к каждому m^i для $i = 1, \dots, k - 1$, мы получим цепочку

$$\tilde{m} \succcurlyeq m^1 \succcurlyeq \dots \succcurlyeq m^{k-1} \succcurlyeq \tilde{m}.$$

Отношение $\succcurlyeq$, разумеется, транзитивно, и для некоторых (или для некоторого) i , $1 \leq i \leq k$, условие (8.18) выполняется как строгое неравенство. Следовательно, $\hat{m} > \tilde{m}$. Тем самым доказательство завершено.

В отношении (8.16) и (8.17) нужно заметить, что $\tilde{m}$ может и не принадлежать множеству M^γ , поскольку мы не требовали, чтобы M^γ содержало в себе какой-либо из элементов $\tilde{M}^\gamma$.

Теорема 8.3

Предположим, что локальные функции затрат двухуровневой системы сбалансированы с глобальной функцией затрат, как и в теореме 8.1. Пусть $\tilde{m}$ — заданное эталонное управление, и пусть множество M' допустимых управлений имеет вид

$$M' = \{m: \rho(m, \tilde{m}) \leq 2\}.$$

Предположим также, что для каждого координирующего сигнала $\gamma \in \mathcal{E}$ множество M^γ , определяемое условием (8.17), содержит $\tilde{m}$. Тогда для данного $\tilde{m}$ применим принцип оценки взаимодействий.

Доказательство. Допустим, что существуют такие γ из $\mathcal{E}$ и $\hat{m}$ из M' , что (8.11) справедливо. Если $\rho(\tilde{m}, \hat{m}) \leq 1$, то из

теоремы 8.1 следует, что $\hat{m} > \tilde{m}$. Положим поэтому $\rho(\tilde{m}, \hat{m}) = 2$. Далее, известно, что существует такое m в $\bar{M}$, что $\rho(\hat{m}, m) = \rho(m, \tilde{m}) = 1$. Предположим, что $\hat{m}_i \neq m_i$ и $m_j \neq \tilde{m}_j$. Тогда $g_{i\mathcal{B}}(\hat{m}_i, u_i, \beta_i) \leq g_{i\mathcal{B}}(\tilde{m}_i, u_i, \beta_i) = g_{i\mathcal{B}}(m_i, u_i, \beta_i)$ на $U^i \times \mathcal{B}^i$, $g_{j\mathcal{B}}(m_j, u_j, \beta_j) = g_{j\mathcal{B}}(\hat{m}_j, u_j, \beta_j) \leq g_{j\mathcal{B}}(\tilde{m}_j, u_j, \beta_j)$ на $U^j \times \mathcal{B}^j$, причем по крайней мере одно из них является строгим неравенством. Используя предположение о том, что $\tilde{m}$ принадлежит M^v , и теорему 8.1, получаем $\hat{m} \geq m \geq \tilde{m}$. Кроме того, либо $\hat{m} > m$, либо $m > \tilde{m}$. Поэтому $\hat{m} > \tilde{m}$, и доказательство завершено.

Следующая теорема не имеет отношения к принципу оценки взаимодействий как таковому, но дает условие для координируемости на основе эталонного управляющего воздействия.

Теорема 8.4

Предположим, что локальные функции затрат двухуровневой системы сбалансированы с глобальной функцией затрат. Пусть $\tilde{m}$ — заданное эталонное управление, и пусть множество допустимых управляющих воздействий определяется выражением (8.15). Тогда система координируема при $\tilde{m}$, если существует $\tilde{m}$ -приемлемый координирующий сигнал γ из $\mathcal{C}$, такой, что $\tilde{m}$ принадлежит множеству M^v , заданному выражением (8.17).

Доказательство этого утверждения очевидным образом следует из доказательства теоремы 8.1.

Сделаем теперь несколько замечаний, касающихся применения принципа оценки взаимодействий.

При использовании принципа оценки взаимодействий цель координатора, после того как определено эталонное управление, состоит в том, чтобы выбрать такие оценочные диапазоны $U^v \times \mathcal{B}^v$, для которых при всех ω из Ω справедливы отношения: $K(\hat{m}, \omega) \in U^v$, $\eta(\hat{m}, \omega) \in \mathcal{B}^v$, где $\hat{m}$ — управляющее воздействие, выработанное локальными решающими элементами. Сложность этой задачи обусловлена тем фактом, что выбираемые локальными элементами решения координатору заранее не известны. Однако координатору известно эталонное управление, которое он может использовать в качестве основы для построения оценочных диапазонов. Например, пусть $\tilde{m}$ — выбранное эталонное управление и предположим, что на множествах $\bar{U}$ и $\mathcal{B}$ определена норма. Для каждого γ из $\mathcal{C}$ координатор может выбрать положительные числа ε_γ и δ_γ и затем определить оценочный диапазон $U^\gamma \times \mathcal{B}^\gamma$ в виде

$$U^\gamma = \{u \in U: \|u - K(\tilde{m}, \omega)\| < \varepsilon_\gamma \text{ для некоторого } \omega \text{ из } \Omega\},$$

$$\mathcal{B}^\gamma = \{\beta \in \mathcal{B} \mid \|\beta - \eta(\tilde{m}, \omega)\| < \delta_\gamma \text{ для некоторого } \omega \text{ из } \Omega\}.$$

Конкретные значения для ε_γ и δ_γ можно выбрать позднее. Однако нужно отметить, что, хотя при увеличении ε_γ и δ_γ возрастает вероятность того, что с помощью сигнала γ удастся скоординировать систему, при этом в то же время снижается «эффективность» γ в том смысле, что потенциальные возможности улучшения характеристик при использовании γ -предпочтительного управления уменьшаются. Поэтому в практических задачах нужно искать компромисс между потенциальной возможностью улучшения и уверенностью в том, что γ действительно скоординирует систему.

Фигурирующая в теореме 8.2 ρ -связность множества допустимых управляющих воздействий не является серьезным ограничением. Например, любая пара элементов из M^f ρ -связана в M^f , если M^f представляет собой n -кратное декартово произведение $M^f = M_1^f \times \dots \times M_n^f$. Поэтому принцип оценки взаимодействий применим для любого такого множества координирующих сигналов $\mathcal{E}$, что

$$\begin{aligned} \{K(m, \omega) \mid m \in M^f \text{ и } \omega \in \Omega\} &\subseteq U^\gamma \text{ и} \\ \{\eta(m, \omega) \mid m \in M^f \text{ и } \omega \in \Omega\} &\subseteq \mathcal{B}^\gamma. \end{aligned}$$

Однако для выбранного эталонного управления $\tilde{m}$ из M множество M^f может не содержать какого-либо управления $\hat{m}$, которое было бы предпочтительнее $\tilde{m}$ в оценочном диапазоне $U^\gamma \times \mathcal{B}^\gamma$.

Координируемость и «удовлетворительное» управление

«Линейная» координируемость и принцип оценки взаимодействий при использовании эталонного управления рассматривались выше с точки зрения улучшения характеристик процесса по отношению к какому-то выбранному эталонному управлению. Очевидно, что эталонное управление накладывает ограничения на глобальную и локальные функции качества, но эти ограничения могут быть ослаблены за счет того, что границы допустимых значений функций качества будут определены просто через критериальные функции.

Пусть $\tau: \Omega \rightarrow V$ — глобальная критериальная функция, и глобальная задача состоит в том, чтобы найти такое управляющее воздействие $\hat{m}$ из M^f , что $g(\hat{m}, \omega) \leq \tau(\omega)$ для всех ω из Ω .

Аналогично пусть для каждого i , $1 \leq i \leq n$, заданы критериальные функции $\tau_i: \mathcal{B}_i \times U_i \times \mathcal{B}_i \rightarrow V$; тогда i -я локальная задача, определяемая сигналом γ из $\mathcal{E}$, состоит в том, чтобы найти управляющее воздействие $\hat{m}_i$ из M_i , для которого

$$g_i(\mathcal{B}_i)(\hat{m}_i, u_i, \beta_i) \leq \tau_i(\mathcal{B}_i)(u_i, \beta_i)$$

в оценочном диапазоне $U_i^\gamma \times \mathcal{B}_i^\gamma$.

Понятие координируемости и понятия, связанные с принципом оценки взаимодействий при использовании эталонного управле-

ния, непосредственным образом распространяются на случай, когда глобальная и локальные задачи сформулированы в указанном выше виде. Нам нужно лишь подходящим образом переопределить отношение предпочтительности на множестве M — так, чтобы новое определение при переформулировании задачи улучшения и введении критериальных функций было совместимо со старым.

В качестве критериальных функций, особенно когда Ω одноэлементное множество, удобно выбрать константы. Тогда ограничением для функций качества служат просто верхние границы локальных затрат. В этом случае координатор выбирает эти верхние границы затрат $k_1, \dots, k_n$ и, конечно, оценочные диапазоны так, чтобы локальные решающие элементы в результате собственных действий вырабатывали то глобально допустимое управляющее воздействие, при котором общие затраты не превышают некоторой, заранее определенной величины k . Тогда при последовательном уменьшении k и $k_1, \dots, k_n$ значения глобальной функции качества последовательно улучшаются, если только на каждой стадии «корректирования» оценочные диапазоны выбираются так, чтобы система была координируемой.

Конечно, имеются иные возможности выбора критериальных функций. Они могут выбираться и отличными от констант, причем различные координирующие сигналы будут порождать различные множества критериальных функций для локальных решающих элементов. Это имеет место и в случае использования эталонного управления.

3. ПРИЛОЖЕНИЕ К СИСТЕМАМ, ОПИСЫВАЕМЫМ ЛИНЕЙНЫМИ УРАВНЕНИЯМИ

Полученные в предыдущем разделе условия применимости принципа оценки взаимодействий и координируемости могут быть непосредственно применены к двухуровневой системе, когда управляемый процесс в целом описывается линейными уравнениями и используется квадратичная функция качества. Мы здесь покажем, как можно ввести функции для оценки локальных затрат таким образом, чтобы соответствующие локальные функции затрат были согласованы с глобальной функцией затрат. При этом мы в явном виде выпишем нужные для решения этой задачи функции η_i и h_i . Допустим, что множества $M = M_1 \times \dots \times M_n$ и $Y = Y_1 \times \dots \times Y_n$ представляют собой линейные пространства и что их компоненты — пространства M_i и Y_i — являются гильбертовыми, и пусть Ω — некоторое подмножество линейного пространства $\bar{\Omega}$.

Предположим, что процесс в целом P выражается с помощью линейного оператора с областью определения $M \times \bar{\Omega}$ и областью значений Y в следующем виде:

$$P(m, \omega) = Am + B\omega + y^f,$$

где A и B — линейные операторы, определенные на M и $\bar{Q}$ соответственно, а y^f — некоторый элемент из Y , представляющий собой «свободный» выход процесса, реализующийся при отсутствии управляющих воздействий и внешних возмущений, $y^f = P(0, 0)$.

Пусть заданная глобальная функция для оценки затрат G определена на $M \times Y$ в виде квадратичной формы

$$G(m, y) = \sum_{i=1}^n [\langle m_i, Q_{iM} m_i \rangle + \langle y_i - \tilde{y}_i, Q_{iY} (y_i - \tilde{y}_i) \rangle],$$

где Q_{iM} и Q_{iY} , $1 \leq i \leq n$, суть самосопряженные, положительные, ограниченные линейные операторы, а $\tilde{y} = (\tilde{y}_1, \dots, \tilde{y}_n)$ — некоторый фиксированный эталонный выход.

Процесс в целом P может быть представлен своими компонентами, описываемыми уравнениями

$$y_i = A_i m + B_i \omega + y_i^f, \quad (8.19)$$

где $1 \leq i \leq n$; иначе говоря, для данных m из M и ω из $\bar{Q}$ выражение (8.19) определяет i -ю компоненту выхода $y = P(m, \omega)$. Далее, представим процесс в целом P в виде n связанных подпроцессов, определив каждый i -й подпроцесс на $M_i \times U_i$, где $U_i = Y_i$, соотношением

$$P_i(m_i, u_i) = A_{ii} m_i + u_i + y_i^f,$$

где $A_i m = A_{i1} m_1 + \dots + A_{in} m_n$. Связующий вход u_i для i -го подпроцесса будет иметь вид

$$u_i = \sum_{j=1}^n K_{ij} m_j + B_i \omega \equiv K_i(m, \omega),$$

где $K_{ii} = 0$ и $K_{ij} = A_{ij}$ при $i \neq j$. Заметим, что K_i не зависит от m_i ; связующий вход для i -го подпроцесса «связанной» системы не зависит от управляющих воздействий, подаваемых на этот подпроцесс.

Установим теперь вид локальных функций для оценки затрат $G_{i\mathcal{B}}$ так, чтобы соответствующие локальные функции затрат, определяемые выражением (8.7), были сбалансированы с глобальной функцией затрат $g(m, \omega) = G(m, P(m, \omega))$. А именно, для каждого i , $1 \leq i \leq n$, определим локальные функции для оценки затрат $G_{i\mathcal{B}}$ на $M_i \times Y_i \times M_i$ в виде функционала

$$G_{i\mathcal{B}}(m_i, y_i, \beta_i) = \langle m_i, (Q_{iM} + Q_i) m_i \rangle + \langle y_i - \tilde{y}_i, Q_{iY} (y_i - \tilde{y}_i) \rangle + \langle \beta_i, m_i \rangle, \quad (8.20)$$

где

$$Q_i = \sum_{j \neq i} A_{ji}^* Q_{iY} A_{ji},$$

и покажем, что равенство

$$\begin{aligned} G_{i\mathcal{B}}(m_i, P_i(m_i, K_i(m, \omega)), \eta_i(m, \omega)) + h_i(m, \omega) = \\ = G(m, P(m, \omega)) \end{aligned} \quad (8.21)$$

справедливо на множестве $M \times \bar{\Omega}$, если функции η_i и h_i определены на $M \times \bar{\Omega}$ уравнениями

$$\eta_i(m, \omega) = \sum_{j \neq i} 2A_{ji}^* Q_{j\mathcal{Y}} \xi_{ji}(m, \omega), \quad (8.22)$$

$$h_i(m, \omega) = \sum_{j \neq i} [\langle \xi_{ji}(m, \omega), Q_{j\mathcal{Y}} \xi_{ji}(m, \omega) \rangle + \langle m_j, Q_{j\mathcal{M}} m_j \rangle], \quad (8.23)$$

где для каждого j , $1 \leq j \leq n$,

$$\xi_{ji}(m, \omega) = \sum_{k \neq i} [A_{jk} m_k + B_j \omega + y_j^i - \tilde{y}_j]. \quad (8.24)$$

Заметим, что определенные таким образом функции η_i и h_i не зависят от m_i .

Чтобы показать, что (8.21) справедливо на $M \times \bar{\Omega}$ для любого i , $1 \leq i \leq n$, если функции $G_{i\mathcal{B}}$, η_i и h_i определены, как указано выше, выберем из множеств M и $\bar{\Omega}$ произвольные, но фиксированные элементы m и ω соответственно и для каждого i , $1 \leq i \leq n$, положим

$$y_i = A_i m + B_i \omega + y_i^i, \quad z_i(\omega) = B_i \omega + y_i^i - \tilde{y}_i. \quad (8.25)$$

Однако для всех j , $1 \leq j \leq n$, справедливы следующие равенства:

$$\begin{aligned} \langle y_j - \tilde{y}_j, Q_{j\mathcal{Y}}(y_j - \tilde{y}_j) \rangle &= \langle A_j m + z_j(\omega), Q_{j\mathcal{Y}}(A_j m + z_j(\omega)) \rangle = \\ &= \langle A_j m, Q_{j\mathcal{Y}} A_j m \rangle + 2 \langle z_j(\omega), Q_{j\mathcal{Y}} A_j m \rangle + \langle z_j(\omega), Q_{j\mathcal{Y}} z_j(\omega) \rangle = \\ &= \sum_k \sum_l \langle A_{jk} m_k, Q_{j\mathcal{Y}} A_{jl} m_l \rangle + 2 \sum_k \langle z_j(\omega), Q_{j\mathcal{Y}} A_{jk} m_k \rangle + \\ &\quad + \langle z_j(\omega), Q_{j\mathcal{Y}} z_j(\omega) \rangle = \\ &= \langle A_{ji} m_i, Q_{i\mathcal{Y}} A_{ji} m_i \rangle + 2 \langle \sum_{k \neq i} A_{jk} m_k + z_j(\omega), Q_{j\mathcal{Y}} A_{ji} m_i \rangle + \\ &\quad + \langle \sum_{k \neq i} A_{jk} m_k + z_j(\omega), Q_{j\mathcal{Y}} (\sum_{k \neq i} A_{jk} m_k + z_j(\omega)) \rangle. \end{aligned} \quad (8.26)$$

Следовательно, из (8.24) — (8.26) имеем

$$\begin{aligned} \langle y_i - \tilde{y}_i, Q_{i\mathcal{Y}}(y_i - \tilde{y}_i) \rangle &= \langle A_{ji} m_i, Q_{j\mathcal{Y}} A_{ji} m_i \rangle + \\ &+ 2 \langle \xi_{ji}(m, \omega), Q_{j\mathcal{Y}} A_{ji} m_i \rangle + \langle \xi_{ji}(m, \omega), Q_{j\mathcal{Y}} \xi_{ji}(m, \omega) \rangle, \end{aligned} \quad (8.27)$$

а из (8.20) и (8.27) получаем

$$\sum_{j \neq i} \langle y_j - \tilde{y}_j, Q_{jY} (y_j - \tilde{y}_j) \rangle = \langle m_i, Q_i m_i \rangle + \\ + \sum_{j \neq i} \langle 2A_{ji}^* Q_{jY} \xi_{ji}(m, \omega), m_i \rangle + \sum_{j \neq i} \langle \xi_{ji}(m, \omega), Q_{jY} \xi_{ji}(m, \omega) \rangle. \quad (8.28)$$

Наконец, (8.22), (8.23) и (8.28) дают

$$\sum_{j \neq i} [\langle y_j - \tilde{y}_j, Q_{jY} (y_j - \tilde{y}_j) \rangle + \langle m_j, Q_{jM} m_j \rangle] = \\ = \langle m_i, Q_i m_i \rangle + \langle \eta_i(m, \omega), m_i \rangle + h_i(m, \omega). \quad (8.29)$$

Тогда из (8.29) мы можем заключить, что

$$G(m, P(m, \omega)) = \langle m_i, (Q_{iM} + Q_i) m_i \rangle + \\ + \langle P_i(m_i, K_i(m, \omega)) - \tilde{y}_i, Q_{iY} (P_i(m_i, K_i(m, \omega)) - \tilde{y}_i) \rangle + \\ + \langle \eta_i(m, \omega), m_i \rangle + h_i(m, \omega),$$

для всех m из M и ω из $\bar{\Omega}$.

Теперь мы можем использовать результаты предыдущего раздела, касающиеся применимости принципа оценки взаимодействий и координируемости при использовании эталонного управления, и сделать вывод, что система координируема, если локальные функции для оценки затрат даются выражением (8.20). Оценочные диапазоны $U_i^Y \times \mathcal{R}_i^Y$ можно построить, используя заданные подмножества M^Y множества M и отображения K_i и η_i :

$$U_i^Y = \{u_i \mid u_i = K_i(m, \omega) \text{ для некоторых } m \text{ из } M^Y \text{ и некоторых } \omega \text{ из } \Omega\} = K_i(M^Y, \Omega),$$

$$\mathcal{R}_i^Y = \{\beta_i \mid \beta_i = \eta_i(m, \omega) \text{ для некоторых } m \text{ из } M^Y \text{ и некоторых } \omega \text{ из } \Omega\} = \eta_i(M^Y, \Omega).$$

Следует заметить, что в рассматриваемом нами здесь линейно-квадратичном случае функции η_i и h_i можно в явном виде выразить через параметры системы, т. е. A_{jk} , Q_{jM} , Q_{jY} и т. д. Кроме того, между функциями η_i , заданными выражением (8.22), и линейными операторами взаимодействий имеется соотношение, обусловленное видом локальных функций (функционалов) затрат:

$$G_i(m_i, y_i) = \langle m_i, Q_{iM} m_i \rangle + \langle y_i - \tilde{y}_i, Q_{iY} (y_i - \tilde{y}_i) \rangle$$

Полный линейный оператор оценки косвенного эффекта управляющих воздействий Γ_j в точке $(\hat{m}, \omega)$ пространства $M \times \bar{\Omega}$ представляет собой ограниченный линейный оператор, действующий на множестве M_i и вводимый с помощью выражения

$$\Gamma_i(\hat{m}, \omega) m_i = 2 \left\langle \sum_{j \neq i} A_{ji}^* Q_{jY} (A_j \hat{m} + B_j \omega + y_j^f - \tilde{y}_j), m_i \right\rangle.$$

Далее, из (8.24) имеем

$$\Gamma_i^*(\hat{m}, \omega) m_i = 2 \left\langle \sum_{j \neq i} A_{ji}^* Q_{j\gamma} \xi_{ji}(\hat{m}, \omega), m_i \right\rangle + 2 \sum_{j \neq i} \langle A_{ji}^* Q_{j\gamma} A_{ji} \hat{m}_i, m_i \rangle.$$

Следовательно, из (8.21) и (8.22) получаем

$$\Gamma_i^*(\hat{m}, \omega) m_i = \langle \eta_i(\hat{m}, \omega), m_i \rangle + 2 \langle Q_i \hat{m}_i, m_i \rangle. \quad (8.30)$$

Можно считать, что локальные функции для оценки затрат получаются из функционалов G_i добавлением соответствующих модифицирующих членов. Эти модифицирующие члены, в отличие от встречавшихся нам в гл. 6, состоят из двух частей — линейной и квадратичной. Учитывая соотношение (8.30), мы можем утверждать, что квадратичная часть дает «приближение второго порядка» в аппроксимации взаимной зависимости функций качества.

4. ПОСЛЕДОВАТЕЛЬНАЯ КООРДИНАЦИЯ С АДАПТАЦИЕЙ

Подход, изложенный в разд. 2 этой главы, легко использовать для построения последовательной, многоэтапной процедуры координации. На любом этапе процесса координации система координируется так, чтобы было получено улучшение глобальных характеристик. При этом управление, имевшее место на предыдущем этапе, либо непосредственно выступает как эталонное управление для последующего этапа, либо предопределяет его (аналогично дело обстоит и с эталонными значениями функции качества).

Эффективность процесса последовательной координации, разумеется, будет зависеть от выбираемой на каждом этапе постановки локальных задач. Если локальные задачи являются задачами улучшения и на данном этапе применим принцип оценки взаимодействий, то координируемость на этом этапе зависит от того, как задаются оценочные диапазоны. Если для каждого координирующего сигнала γ оценочные диапазоны $U_i^\gamma \times \mathcal{R}_i^\gamma$ определяются парой положительных чисел $(\varepsilon_i^\gamma, \delta_i^\gamma)$, которые устанавливают соответствующие окрестности на множествах U_i^γ и $\mathcal{R}_i^\gamma$ (характеризуемые с помощью заданных норм), то, выбрав приемлемое значение γ так, чтобы ε_i^γ и δ_i^γ были достаточно большими, систему можно сделать координируемой, но в этом случае улучшение глобальных характеристик, вероятно, окажется весьма небольшим. С другой стороны, если локальные задачи поставлены в виде задач нахождения удовлетворительных решений, т. е. таких решений, для которых в оценочных диапазонах локальные затраты не превышают заранее заданной величины k , то желаемую степень улучшения можно гарантировать подходящим выбором k . Но слишком неосторожный выбор k может привести к тому, что система перестанет быть координируемой. В любом случае многократным применением

ем того же самого метода «линейной» координации глобальные характеристики системы можно будет улучшать, хотя и медленно, пока они не станут удовлетворительными.

Процедура последовательной координации может быть использована в качестве итерационного метода, позволяющего повысить характеристики системы до нужного уровня в пределах ее возможностей, и даже достигнуть оптимума, когда внешние возмущения отсутствуют и понятие оптимальности имеет смысл.

Рассмотрим теперь применение процедуры последовательной координации, когда моменты времени, в которые координатор может вмешиваться в действия локальных решающих элементов, образуют последовательность $\{t_0, t_1, \dots, t_j\}$. Через $\tilde{m}^i$ мы будем обозначать управляющие воздействия, выбранные локальными решающими элементами и применяемые ими начиная с момента координирования t_i . В следующий момент координирования t_{i+1} в качестве эталонного управления $\tilde{m}^{i+1}$ на оставшемся интервале времени $[t_{i+1}, t_j]$ берется управление $\hat{m}^i$. Далее, если система координируема при использовании $\tilde{m}^{i+1}$, существует такое $\tilde{m}^{i+1}$ -приемлемое координирующее воздействие γ , что выработанное локальными решающими элементами в момент времени t_{i+1} новое управление $\hat{m}^{i+1}$ будет γ -предпочтительнее, чем $\tilde{m}^{i+1}$.

Если система остается координируемой для всей последовательности моментов координирования, а окружающая среда не меняется, глобальные характеристики системы будут непрерывно улучшаться.

Следует заметить, что координатор фактически не управляет ходом процесса P в интервале между моментами координирования t_i и t_{i+1} . В течение этого времени он может попытаться уменьшить неопределенности и уточнить оценочные диапазоны для связующих входов и внешних возмущений. В том случае, когда неопределенности (внешние возмущения) отсутствуют, координатор может использовать промежуточный период, чтобы улучшить эталонное управление, решив относительно простую задачу, — так называемую *вспомогательную задачу вышестоящего решающего элемента*. Задача состоит в том, чтобы найти управление, более предпочтительное, чем эталонное управление. Для системы, описанной в предыдущем разделе, эта вспомогательная задача может быть сформулирована следующим образом.

Как и выше, обозначим через $\tilde{m}^i$ и $\hat{m}^i$ соответственно эталонное управление и управление, выбранное на i -м этапе координации (в i -й момент координации). Пусть $[\tilde{m}^i, \hat{m}^i]$ — отрезок в M , соединяющий эти две точки:

$$[\tilde{m}^i, \hat{m}^i] = \{m : m = \lambda \tilde{m}^i + (1 - \lambda) \hat{m}^i \text{ для некоторого } \lambda, 0 \leq \lambda \leq 1\}.$$

Тогда на i -м этапе координации вспомогательная задача состоит в следующем: найти действительное число λ из интервала $[0, 1]$ и соответствующее управление $m^i = \lambda \tilde{m}^i + (1 - \lambda) \hat{m}^i$, такие, чтобы выполнялось условие

$$G(m^i, P(m^i)) = \min_{[\tilde{m}^i, \hat{m}^i]} G(m, P(m)). \quad (8.31)$$

Простота задачи, конечно, является следствием того факта, что оптимизация проводится в интервале действительных чисел $[0, 1]$, а не в пространстве функций, зависящих от времени. Можно достичь дальнейшего упрощения, если ослабить условие (8.31) и решать лишь задачу улучшения:

$$G(m^i, P(m^i)) < G(\hat{m}^i, P(\hat{m}^i)). \quad (8.32)$$

Независимо от того, какое из требований — (8.31) или (8.32) — используется, процедура последовательной координации, схематически изображенная на фиг. 8.1, состоит в следующем.

1. В момент координирования t_i координатор направляет локальным решающим элементам эталонное управление $\tilde{m}_i$ и координирующий сигнал γ , который определяет оценочные диапазоны для взаимодействий или неопределенностей (связанных, например, с внешними возмущениями).

2. Каждый локальный решающий элемент ищет управление, которое улучшило бы значения его локальной функции качества по сравнению с эталонным управлением $\tilde{m}^i$; при этом текущие оценочные диапазоны задаются посредством γ . Предположим, что координирующий сигнал γ является $\tilde{m}^i$ -приемлемым; в этом случае такое управление, скажем $\hat{m}^i$, существует.

3. Координатор получает от локальных решающих элементов найденные ими $\hat{m}^i$, решает свою вспомогательную задачу и вырабатывает таким образом управляющее воздействие m_i . Затем координатор принимает в качестве эталонного управления, которое будет использоваться в следующий момент координирования t_{i+1} , управление $\tilde{m}^{i+1} = m^i | [t_{i+1}, t_f]$.

Если в каждый момент координирования t_i система координируема при эталонном управлении $\tilde{m}^i$, имеем

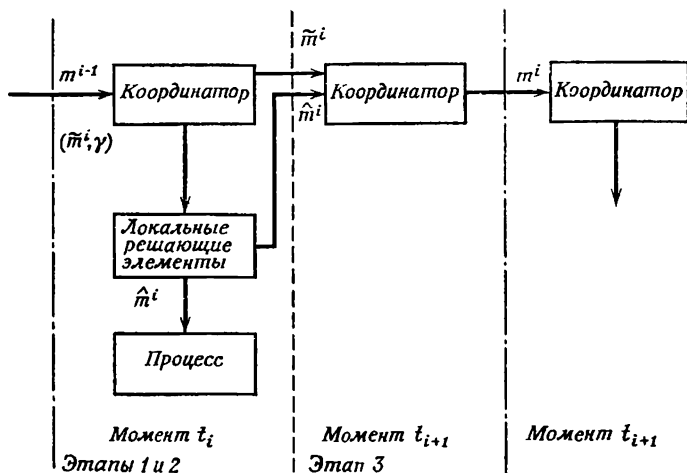
$$m^i \succ \hat{m}^i \succ \tilde{m}^i,$$

поскольку $\tilde{m}^{i+1} = m^i | [t_{i+1}, t_f]$ и характеристики системы улучшаются монотонно.

Пример 8.2

Рассмотрим двухуровневую систему, описанную в примере 8.1, для которой глобальная функция для оценки затрат G имеет вид

$$G(m, y) = m^T m + (y_1 - \tilde{y}_1) + (y_2 - \tilde{y}_2),$$



Ф и г. 8.1. Процесс последовательной координации, основанный на методе «линейной» координации.

а локальные функции для оценки затрат $G_{i\mathcal{B}}$ выражаются в виде

$$G_{i\mathcal{B}}(m_1, u_1, u_1, \beta_1) = m_1^2 + (y_1 - \tilde{y}_1) + \beta_1(m_1 + u_1), \quad (8.33)$$

$$G_{i\mathcal{B}}(m_2, u_2, u_2, \beta_2) = m_2^2 + (y_2 - \tilde{y}_2) + \beta_2 m_2. \quad (8.34)$$

Предположим для простоты, что внешние возмущения отсутствуют, т. е. $\omega = 0$. Покажем, как использовать процедуру последовательной координации для координирования системы с таким расчетом, чтобы ее глобальная характеристика (функция качества) улучшалась монотонно.

Предположим, что мы находимся на k -м этапе процедуры координации и что эталонное управление есть $\tilde{m} = (\tilde{m}_1, \tilde{m}_2)$. Тогда координатор должен установить оценочные диапазоны $U_1^? \times \mathcal{B}_1^?$ и $U_2^? \times \mathcal{B}_2^?$ для локальных решающих элементов с целью улучшения $\tilde{m}$ (если $\tilde{m}$ не является глобально оптимальным). Пусть $\tilde{y}_1 = \tilde{y}_2 = 1$. Локальными функциями затрат являются в этом случае функции

$$g_{2\mathcal{B}}(m_1, u_1, \beta_1) = m_1^2 + (m_1^2 + u_1 - 1) + \beta_1(m_1 + u_1),$$

$$g_{2\mathcal{B}}(m_2, u_2, \beta_2) = m_2^2 + (m_2 u_2 - 1) + \beta_2 m_2.$$

Эти выражения получаются путем исключения y_1 и y_2 соответственно из (8.33) и (8.34).

Далее локальные решающие элементы должны подобрать новые управляющие воздействия так, чтобы уменьшить в пределах оценочного диапазона свои затраты по сравнению с затратами

при эталонном управлении. Для этого они должны найти такие управляющие воздействия m_i , которые удовлетворяли бы на $U_i^\gamma \times \mathcal{R}_i^\gamma$ неравенствам

$$g_{i\mathcal{R}}(m_i, u_i, \beta_i) - g_{i\mathcal{R}}(\tilde{m}_i, u_i, \beta_i) \leq 0. \quad (8.35)$$

Для первого локального решающего элемента неравенство (8.35) имеет вид

$$2(m_1^2 - \tilde{m}_1^2) + \beta_1(m_1 - \tilde{m}_1) \leq 0.$$

Поскольку $\eta_2(m_1, m_2) \equiv 1$, для второго локального решающего элемента это неравенство принимает вид

$$(m_2^2 - \tilde{m}_2^2) + (u_2 + 1)(m_2 - \tilde{m}_2) \leq 0.$$

Заметим, что координатор должен определить области изменения только для β_1 и u_2 , так как $\beta_2 \equiv 1$, а u_1 вообще не фигурирует в задаче первого локального решающего элемента.

Пусть к процессу приложено управление $\tilde{m} = (0, 0)$. Тогда отвечающее этому управлению значение глобальной функции затрат

$$g(m_1, m_2) = m_1^2 + m_2^2 + (m_1^2 + m_2 - 1) + (m_2 m_1 - 1)$$

равно $g(\tilde{m}) = -2$. С помощью процедуры последовательной координации это управление может быть улучшено. Продемонстрируем, как это можно сделать.

Этап 1. Исходное управление $\tilde{m} = (0, 0)$ рассматривается координатором как эталонное для данного этапа. Пытаясь так скоординировать действия локальных решающих элементов, чтобы улучшить $\tilde{m}$, координатор строит оценочные диапазоны

$$\mathcal{R}_1^\gamma = \{\beta_1 : |\eta_1(\tilde{m}) + \beta_1| \leq 1\}$$

и

$$U_2^\gamma = \left\{ u_2 : |K_2(\tilde{m}) + u_2| \leq \frac{1}{2} \right\}.$$

Так как $0 = \eta_1(\tilde{m}) = K_2(\tilde{m})$, то $\mathcal{R}_1^\gamma = [-1, 1]$ и $U_2^\gamma = \left[-\frac{1}{2}, \frac{1}{2}\right]$. Тогда локальные решающие элементы выберут в качестве управляющих воздействий

$$\hat{m}_1 = 0 \quad \text{и} \quad \hat{m}_2 = -\frac{1}{2}.$$

Управление $\hat{m}$ будет γ -предпочтительнее управления $\tilde{m}$ и, поскольку $\eta_1(\hat{m}) = -1/2$ принадлежит $\mathcal{R}_1^\gamma$, а $K_2(\hat{m}) = 0$ принадлежит U_2^γ , то, используя принцип оценки взаимодействий, мы приходим к выводу, что управление $\hat{m}$ приведет к меньшим затратам, чем управление $\tilde{m}$. В самом деле, подавая это управление на

вход процесса, мы получаем $g(\hat{m}) = -9/4$ и

$$g(\hat{m}) = -\frac{9}{4} < -2 = g(\tilde{m}).$$

Э т а п 2. Координатор решает свою вспомогательную задачу (вспомогательную задачу вышестоящего решающего элемента), принимая в качестве эталонного управление $\tilde{m} = (0, -1/2)$, предложенное к процессу на предыдущем этапе. В результате получают следующие оценочные диапазоны:

$$\mathcal{B}_1 = \left\{ \beta_1 : |\eta_1(\tilde{m}) + \beta_1| \leq \frac{1}{4} \right\} = \left[-\frac{3}{4}, -\frac{1}{4} \right],$$

$$U_2 = \left\{ u_2 : |K_2(m) + u_2| \leq \frac{1}{2} \right\} = \left[\frac{1}{2}, \frac{1}{2} \right],$$

поскольку $\eta_1(\tilde{m}) = -1/2$ и $K_2(\tilde{m}) = 0$. Локальные решающие элементы выберут тогда новые значения управляющих воздействий $\hat{m}_1 = 1/8$ и $\hat{m}_2 = -1/2$. Поскольку $\hat{m} = (1/8, -1/2)$ является γ -предпочтительнее $\tilde{m} = (0, -1/2)$ и оценочные диапазоны выбраны правильно, значение глобальной функции качества улучшается. Действительно, при использовании управления $\hat{m} = (1/8, -1/2)$ имеем

$$g(\hat{m}) = -\frac{147}{64} < -\frac{9}{4} < -2.$$

Отметим, что минимум глобальных затрат равен $-16/7$, и полученные только что глобальные затраты отличаются от минимальных на $5/448$.

Мы выбрали сравнительно простую систему для иллюстрации процедуры последовательной координации. Однако процедура остается той же и в случае, когда подпроцессы описываются различными или дифференциальными уравнениями. К сожалению, в этих случаях вычисления становятся весьма громоздкими и трудновыполнимыми.

Приведенный пример наглядно показывает, как процесс последовательной координации может быть использован в качестве итерационного метода координации при решении оптимизационных задач. Преимущество такой процедуры состоит в монотонности улучшения глобальных характеристик, и эта особенность, безусловно, желательна при использовании «линейной» координации. Отметим также, что процедура последовательной координации весьма удобна, когда присутствуют внешние возмущения, хотя в приведенном примере мы ими для простоты пренебрегали. По-видимому, главная трудность при использовании этого метода заключается в трудности установления приемлемых оценочных диапазонов и решении локальных задач.

Приложение

КОММЕНТАРИЙ К СПИСКУ ЛИТЕРАТУРЫ

Настоящая книга является монографией по новому предмету: в ней представлены собственные исследования авторов, причем теория развивается с самых основ как в отношении ее принципиальных положений, так и математических деталей. Предмет этот новый, поэтому нелегко дать полный список литературы или соответствующий исторический обзор. Если бы мы ограничились приведением лишь списка цитированных источников, он оказался бы слишком коротким и едва ли отразил бы важность предмета. Если бы мы, напротив, захотели дать полную библиографию источников, в той или иной степени относящихся к иерархиям и большим системам рассматриваемого здесь типа, он оказался бы непомерно большим. Поэтому мы решили избрать средний путь и упомянем лишь отдельные источники, содержание которых связано с материалом, представленным в различных главах книги.

ОБЩИЕ ОСНОВЫ ТЕОРИИ

Своей книгой мы хотели заложить основы теории многоуровневых систем принятия решений, т. е. систем со структурами организационного типа.

Первый импульс в этом направлении исходил из области автоматки и управления сложными промышленными системами. В этом отношении непосредственным предшественником данной работы, пожалуй, может считаться теория управления многомерными процессами; в частности, настоящая книга является продолжением и завершением работы [29] и отражает результаты многолетних исследований в области управления большими системами.

Второй импульс исходил от исследований организаций. Теория игр фон Неймана и Моргенштерна, относящаяся к конкурентным ситуациям, и более поздняя теория коалиций Маршана — Раднера, касавшаяся кооперативных ситуаций, имели дело с одноуровневыми системами, тогда как нас в основном интересуют многоуровневые системы. Конечно, методологический аспект этих теорий близок и нам: мы также интересуемся математической теорией организационных структур со всеми присущими ей достоинствами и недостатками.

Наконец, плодотворные импульсы постоянно исходили от общих проблем коммуникации, будь то в биологических, социальных или технических системах. В этом отношении вся теория выдержана в духе кибернетики в виновском понимании этого слова и в духе общей теории систем, как она изложена в работах [30] и [31].

ГЛАВА 1

Цитируется только часть статей и книг, имеющаяся литература здесь поистине необъятна, и мы не пытались дать ее обзор.

ГЛАВА 2

Концептуализация основана на общей схеме, изложенной в работе [32] и более ранней публикации [33] (работа эта представляет собой первую публикацию Центра по исследованию систем и явилась источником многих идей, развитых в последующих работах). Более поздние обзоры представлены в работах [34, 35].

Концепции многослойных иерархий основаны на работах [21] и [36]. Многие плодотворные импульсы к исследованию многоуровневых систем

в автоматизации промышленного производства дала основополагающая работа [37]. Иерархическому построению алгоритмов решения сложных задач посвящены также работы [20, 26].

ГЛАВА 3

Формализация осуществляется на основе математического аппарата общей теории систем, применяемого в работах [30, 31, 38]. Формализация многозвенной иерархии почерпнута из работ [39, 40].

ГЛАВЫ 4 и 5

Принципы координации впервые были четко сформулированы в работах [41, 41а], хотя первоначальная идея была высказана еще в работе [32] и развита в исследованиях, описанных в гл. 5 и 6. Понятия, связанные с координируемостью, взяты из наших работ [41а, 42].

ГЛАВА 6

Эта глава в основном написана на материале работ [43, 44]. Однако понятие «развязывания» взаимодействий, впервые введенное в [32], более детально разработано в применении к частному случаю статической оптимизации в работах [46, 47]. Эти исследования продолжены в работах [47, 48], а затем в работах [50, 51] была сделана попытка обобщить их на динамические системы (хотя основные результаты для динамических систем уже сообщались в [40]) и на системы, описываемые дифференциальными уравнениями в частных производных [51]. Однако аппарат функционального анализа и понятие операторов оценки косвенного эффекта управляющих воздействий, введенные в [43], помогают разрешить многие из встречающихся в этой области трудностей; они позволили выявить общий характер принципа согласования взаимодействий и показали ограниченность таких поверхностных аналогий, как тот факт, что линейная модификация с нулевой суммой, порождаемая оператором оценки эффектов внутренних взаимодействий, приводит к кажущейся глобальной функции качества, которая является так называемой лагранжевой формой (рассмотрение случая квадратичной модификации при этом осуществляется столь же просто, как и рассмотрение линейных модификаций!).

ГЛАВА 7

Полученные здесь результаты убедительно свидетельствуют о применимости принципов координации к проблемам программирования. Не делается никакой попытки использовать это обстоятельство для рассмотрения каких-либо других способов декомпозиций, кроме декомпозиции Дандига — Вольфа [52], так как на одной из стадий разработки этой проблемы возникли определенные затруднения. Разумеется, было бы интересно сравнить декомпозиции, вытекающие из принципов координации, с другими, особенно с декомпозициями нелинейного типа. Принципы оценки взаимодействий, по-видимому, стоило бы исследовать также с точки зрения проблем целочисленного программирования.

ГЛАВА 8

Излагаемый здесь материал основан главным образом на работах [53, 54]. Подход с точки зрения отыскания локальными элементами системы удовлетворительных решений почерпнут из работ [36, 55, 56]; его развитию способствовали предшествующие исследования, касающиеся чувствительности оптимального управления [47, 48].

ЛИТЕРАТУРА

1. Miller A., Automation in the steel industry, *Automation*, Nov. (1966).
2. Miller W. E., Systems engineering in the steel industry, *IEEE Trans. Systems Sci. a. Cybern.* SSC-2, 1. Aug. (1966).
3. Hickling B. B., Jones J. T., Information flow and communications in steel-works. *J. Iron. Steel Ind.*, 205, Pt. 5, May (1967).
4. Haalman A., Hoggendorn K., Evers V. M. J., On-line computer control of ethylene production, Proc. IFAC/IFIC Conf. on Computer Contr., Toronto, 1967.
5. Optimum output for computer controlled ethylene plant, *Instrum. Practice*, 22, 2 (1968).
6. Dy Liacco T. E., The adaptive reliability control system, *IEEE Trans. Power Apparatus and Systems*, PAS-86, May (1967).
7. Happ H. H., Multi-computer configurations and diecaptics; dispatch of real power in power pools, Proc. Power Ind. Comuter Appl. Conf., Pittsburgh, 1967.
8. Arrow K. J., Control in large organizations, *Management Sci.*, 10, 3 (1964).
9. Galbraith J. K., The New Industrial State, Houghton-Mifflin, New York, 1967; русский перевод: Гэлбрайт Дж., Новое индустриальное общество, изд-во «Прогресс», М., 1969.
10. March J. G., Simon H. A., Organizations. Wiley, New York, 1958.
11. Bonini C. P., Simulation of Information and Decision Systems in the Firm, Prentice-Hall, New York, 1963.
12. Marschak J., Radner R., The Economic Theory of Teams. Cowles Foundation Monogr., Yale Univ., 1969.
13. Simon H. A., Administrative Behavior, MacMillan, New York, 1947.
14. Chidambaram T., Coordination problems in competitive situation, Systems Res. Center Rep. 101-A-67-43, 1967.
15. Arrow K. J., Hurwicz L., Decentralization and computation in resource allocation, см. Pfouts R. W. (ed.) Essays in Economics and Econometrics, Univ. of North Carolina Press, Chapel Hill, North Car., 1963.
16. Szent-Gyorgi A., Bioelectronics, Academic Press, New York a. London, 1969; Сент-Дьердь А., Биоэлектроника, изд-во «Мир», М., 1971.
17. Goldstein, Levels and ontogeny, *Am. Scientist*, 50, 1 (1962).
18. Polanyi M., Life's irreducible structure, *Science*, 160, 3834 (1969).
19. Bradley D. F., Multilevel systems and biology-view of a submolecular biologist, см. Mesarovic M. D. (ed.). Systems Theory and Biology, Springer, 1968.
20. Newell A., Shaw J. C., Simon H. A., Report on a general problem-solving program, Proc. Intern. Conf. on Inform. Processing, UNESCO, Paris, 1959.
21. Lefkowitz I., Multilevel approach applied to control system design, *Trans. ASME*, 88D, 2 (1966).
22. Mesarovic M. D., Multilevel concept for systems engineering, Proc. Systems Eng. Conf., Chicago, Ill, 1965.
23. Куликовский Р., Оптимальное управление сложными иерархическими системами, см. сб. «Дискретные самонастраивающиеся системы» (Труды III конгресса ИФАК, вып. 3), изд-во «Наука», М., 1971.
24. Drew D. R., Multilevel approach applies to the design of a freeway control system, Presented at 48th Annual Meeting of HRB, 1969.

25. Mesarovic M. D., Self-organizing control systems, IEEE, Trans. Appl. Ind., 83, 74, Sept. (1964).
26. Bellman R., Stratification and control of large systems with applications to chess and checkers, *Inform. Sci.*, 1, 1, Dec. (1968).
27. Taylor A. E., Introduction to Functional Analysis, Wiley, New York, 1963.
28. Зубов В. И., Методы А. М. Ляпунова и их применение. Изд-во Ленингр. ун-та, Л., 1957.
29. Mesarovic M. D., Control of Multivariable System, MIT Press, Cambridge, Mass. a. Wiley, New York, 1960.
30. Mesarovic M. D., Foundations for a general systems theory см. Views on General Systems Theory, Second Systems Symp., Case Inst. of Technol., Cleveland, Ohio, 1964; русский перевод: см. в сб. «Общая теория систем», изд-во «Мир», М., 1966.
31. Mesarovic M. D., Auxiliary functions and constructive specification of general systems, *Math. Systems Th. J.*, 2, 3 (1968).
32. Mesarovic M. D., A conceptual framework for the studies of multi-level multigoal systems, Systems Res. Center Rep. SRC 101-A-66-43, 1966.
33. Mesarovic M. D., A general systems approach to organization theory, Systems Res. Center Rep. SRC 2-A-62-2, 1962.
34. Mesarovic M. D., Macko D., Foundations for a scientific theory of hierarchical systems, см. Whyte, Wilson a. Wilson (eds.), Hierarchical Structures, Elsevier, New York, 1969.
35. Mesarovic M. D., General systems theory and its mathematical foundation, Proc. IEEE Systems Sci. a. Cybern. Conf., Mass., 1967.
36. Mesarovic M. D., Self-organizing control systems, Proc. JACC, New York, 1962.
37. Lefcowitz I., Eckman D. P., Principles of model techniques in optimizing control, доклад на I конгрессе ИФАК, М., 1960.
38. Mesarovic M. D., Mathematical Theory of General Systems, Penns. State Univ. Press, Univ. Park, Penn., 1967.
39. Macko D., General systems theory approach to multilevel systems, Systems Res. Center Rep. SRC 106-A-67-44, 1967.
40. Macko D., Hierarchical and multilevel systems, Proc. IEEE Systems Sci. a. Cybern. Conf., Boston, Mass., 1967.
41. Mesarovic M. D., Macko D., Takahara Y., Structuring of multi-level systems, Proc. IFAC Symp. Multivariable Systems, Dusseldorf, 1968.
- 41a. Mesarovic M. D., Macko D., Takahara Y., Two coordination principles and their application in large-scale systems control, IV IFAC Congr., Warszawa, 1969.
42. Mesarovic M. D., Macko D., Takahara Y., Coordinability of dynamic systems, Proc. JACC, Boulder, Col., 1969.
43. Takahara Y., Multilevel approach to dynamic optimization, Systems Res. Center Rep. SRC-59-A-64-21, 1964.
44. Macko D., A coordination technique for interacting dynamic systems, Proc. JACC, Seattle, Wash., 1966.
45. Lasdon L., A multilevel technique for optimization, Systems Res. Center Rep. SRC 50-C-64-19, 1964.
46. Lasdon L., Schoeffler J. D., A multilevel technique for optimization, Proc. JACC, Troy, N. Y., 1965.
47. Brosilow C. B., Lasdon L., Pearson J. D., Feasible optimization methods for interconnected systems, Proc. JACC, Troy, N. Y., 1965.
48. Brosilow D. B., Lasdon L., A two-level optimization technique for recycle processes, Am. Inst. Chem. Eng. Symp., Ser. № 4, Appl. Math. Models Chem. Eng. Res., Design, Prod., 1965.
49. Pearson J. D., Multilevel control systems, Proc. IFAC Symp. Adaptive Contr., London, 1965; русский перевод: Пирсон Д. Д., Системы

- многоуровневого управления, см. сб. «Теория самонастраивающихся систем управления» (Труды II междунар. симпозиума ИФАК по самонастраивающимся системам, Теддингтон, Великобритания, 1965), изд-во «Наука», М., 1969, стр. 241—247.
50. Varaiya P., A decomposition technique for Nonlinear Programming, Res. Rep. RJ-345, IBM San Jose Res. Lab., San Jose, Cal., 1965.
 51. Wismer D. A., Optimal Control of Distributed Parameter Systems Using Multilevel Techniques, диссертация, UCLA, 1966.
 52. Dantzig G., Linear Programming and Extensions, Princeton Univ. Press, Princeton, N. J., 1963.
 53. Takahara Y., Multilevel systems and uncertainties, Systems Res. Center Rep. SRC 99-A-66-42, 1966.
 54. Mesarovic M. D., Takahara Y., An approach to on-line coordination, Proc. JACC, Boulder, Col., 1969.
 55. Mesarovic M. D., Satisfaction approach to the synthesis and control of systems, III Allerton Conf. Circuit System Th., 1965.
 56. Takahara Y., Mesarovic M. D., On global sensitivity, Proc. IV Allerton Conf. Circuit System Th., 1966.
 57. Macko D., Uncertainty and optimal control, Systems Res. Center Rep. SRC 16-A-62-10, 1962.
 58. Macko D., Mesarovic M. D., Uncertainties and optimal control approach to feedback control problems, *Inform. Control*, 8, 5 (1965).
 59. Straszak A., Optimal and suboptimal multivariable control systems with controller cost constraint, Proc. III IFAC Congr., London (1966).
 60. Miller J. G., Living Systems: basic concepts, structures and process, cross-level hypothesis, *Behav. Sci.*, 10 (1965).

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА ^{1,2)}

1. Аганбегян А. Г., Багриновский К. А., Гранберг А. Г., Система моделей народно-хозяйственного планирования, изд-во «Мысль», 1972.
2. Бурков В. Н., Принципы управления многоуровневыми активными системами, Международный симпозиум по проблемам организационного управления иерархическими системами, Баку, 1971.
3. Ватель И. А., Ерешко Ф. И., Кононенко А. Ф., Игровые модели принятия решений в иерархических системах, VI Симпозиум по кибернетике, Тбилиси, 1972.
4. Власюк Б. А., Моросанов И. С., Иерархия материальных потоков в больших системах, *Изв. АН СССР*, Автоматика и телемеханика, № 7 (1972).
5. Волкович В. Л., Радомский Н. Ф., Системный подход к исследованию иерархических систем управления, Международный симпозиум по проблемам организационного управления и иерархическим системам, Баку, 1971.
6. Волконский В. А., Оптимальное планирование в условиях боль-

¹⁾ Добавлена редактором русского перевода.

²⁾ Представленный перечень дополнительной литературы не претендует на полноту и имеет целью лишь ввести советского читателя в круг разнообразных исследований, ведущихся в области иерархических систем управления и конструирования соответствующих организационных структур. Весьма подробный анализ современного состояния зарубежных исследований в области математической теории организационных систем дан в работе [13]. — *Прим. ред.*

- шой размерности (итеративные методы и принцип декомпозиции), *Экономика и математические методы*, 1, вып. 2 (1965).
- Волконский В. А., Модель оптимального планирования и взаимосвязи экономических показателей, изд-во «Наука», 1967.
8. Гвишиани Д. М., Организация и управление. Социологический анализ буржуазных теорий, изд-во «Наука», 1970.
 9. Гермейер Ю. Б., Об играх двух лиц с фиксированной последовательностью ходов, *ДАН СССР*, 198, № 5 (1971).
 - 9а. Гермейер Ю. Б., Игровые концепции в исследовании систем, *Изв. АН СССР, Техническая кибернетика*, № 2 (1970).
 10. Гермейер Ю. Б., Моисеев Н. Н., О некоторых задачах теории иерархических систем управления, сб. «Проблемы прикладной математики и механики», изд-во «Наука», 1971.
 11. Гурней Б., Введение в науку управления, изд-во «Прогресс», 1969.
 12. Децентрализованные методы управления (материалы семинара), Московский Дом научно-технической пропаганды им. Ф. Э. Дзержинского, об-во «Знание» РСФСР, М., 1972.
 13. Емельянов С. В., Дудин Е. Б., Ларичев О. И., Малевич А. А., Напельбаум Э. Л., Озерной В. М., Подготовка и принятие решений в организующих системах управления, см. *Итоги науки и техники, сер. Автоматика и радиоэлектроника*, вып. «Техническая кибернетика 1969», изд-во ВИНТИ, 1971, стр. 160—184.
 14. Ивановский А. Г., Проблема согласованного планирования в активных системах, Международный симпозиум по проблемам организационного управления и иерархическим системам, Баку, 1971.
 15. Керженцев П. М., Принципы организации, изд-во «Экономика», 1968.
 16. Кузнецов Н. Н., Оптимальное распределение средств в иерархической системе, *Изв. АН СССР, Техническая кибернетика*, № 4 (1970).
 17. Куликовский Р., Агрегация, оптимизация и управление организационной структурой больших систем, *Экономика и математические методы*, 4, вып. 1 (1968).
 18. Майминас Е. З., Процессы планирования в экономике: информационный аспект, изд-во «Экономика», 1971.
 19. Мангейм М. Л., Иерархические структуры. Модель процессов проектирования и планирования, изд-во «Мир», 1970.
 20. Мартынов Г., Пителин А., Экспериментальное исследование аппроксимационной схемы многоступенчатой оптимизации, *Экономика и математические методы*, 5, вып. 4 (1969).
 21. Маршак Дж., Централизация и децентрализация в экономических системах, сб. «Электронное моделирование и машинное управление в экономике», изд-во «Мир», 1966.
 22. Михалевский Б. Н., Система моделей среднесрочного народнохозяйственного планирования, изд-во «Наука», 1972.
 23. Моисеев Н. Н., Информационная теория иерархических систем, 1-я Всесоюзная конференция по исследованию операций, Минск, 1972.
 24. Моррис У. Т., Наука об управлении. Байесовский подход, изд-во «Мир», 1971.
 25. Павловский Ю. Н., К вопросу об агрегировании и построении иерархических управляющих структур для одного класса сложных систем, *Журн. вычисл. мат. и матем. физ.*, 11, № 6 (1971).
 26. Первозванская Т. Н., Первозванский А. А., Децентрализация оптимального планирования в сложной системе, *Автоматика и телемеханика*, № 7 (1968).
 27. Петров В. И., Структура автоматизированных систем управления производством, *Автоматика и вычисл. техника*, № 1 (1970).
 28. Плискин Л. Г., Декомпозиционная динамическая оптимизация

- производства с иерархической структурой управления. Адаптивное управление комплексом из нестационарных подсистем, *Автоматика и телемеханика*, № 4 (1969).
29. Полтерович В. М., О некоторых абстрактных моделях функционирования ячеек иерархии в управляемой экономической системе, *Экономика и математические методы*, 4, вып. 2 (1968).
 30. Проблемы оптимального функционирования социалистической экономики, сб. под ред. акад. Н. П. Федоренко, изд-во «Наука», 1972.
 31. Пронина В. С., Центральные органы управления народным хозяйством, изд-во «Юридическая литература», 1971.
 32. Пугачев В. Ф., Локальный критерий и стимулирование работников в оптимальной экономической системе, *Экономика и математические методы*, 1, вып. 5 (1966).
 33. Раяцкас Р. Л., Интегрированная система планирования народного хозяйства союзной республики, изд-во «Минтис», Вильнюс, 1972.
 34. Управление фирмами в Японии, изд-во «Прогресс», 1968.
 35. Фаткин Ю. М., Оптимальное управление в иерархических структурах, *ДАН СССР*, 202, № 1 (1972).
 36. Флеров Ю. А., Многоуровневые динамические игры, *ДАН СССР*, 187, № 5 (1969).
 37. Черняк Ю. И., Анализ и синтез систем в экономике, изд-во «Экономика», 1970.
 38. Янг С., Системное управление организацией, изд-во «Сов. радио», 1972.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

- Аддитивная глобальная целевая функция (Additive overall objective function) 172, 227—231
 Безусловная локальная согласованность см. Локальная согласованность безусловная
 — межуровневая согласованность см. Межуровневая согласованность безусловная
 Вмешательство (Intervention) 75—79
 — времена (Intervention times) 77, 78
 Выходная функция (Outcome function) 97
 Вышестоящий элемент (Supermal unit) 77, 82—84
 Глобальная решаемая задача (Overall decision problem) 119
 — функция качества (Overall performance function) 97
 — — аддитивное представление 227—231
 — — — декомпозиция 167
 — — — сепарабельная (separable) 167
 — целевая функция (Overall objective function) 97
 — — — кажущаяся см. Кажущаяся глобальная целевая функция
 Глобально оптимальное решение (Overall optimal decision) 144
 — — управляющее воздействие (Overall optimal control input) 140
 Данцига — Вольфа декомпозиция см. Декомпозиция Данцига — Вольфа
 Двойственная глобальная задача (Overall dual problem) 292, 293
 Двухуровневая система (Two-level system) 109—112
 — — конфликты 148
 — — координируемость 121—123
 Декомпозиция 133
 — глобальных функций качества 167
 — Данцига — Вольфа 297—300
 Декомпозиция задач выпуклого программирования 282, 283
 — каноническая относительно входов 225
 — процесса в целом 225
 Допустимости функция см. Критериальная функция
 Достаточные условия координируемости 162—163
 Задача принятия решения. решаемая задача (Decision problem) 96—99
 — — — как задача нахождения удовлетворительных решений (Satisfaction) 97—99
 — — — как задача оптимизации 96—97
 — — — как задача улучшения 309
 — — — совместимость (Consistency) 122, 123
 Задачи, решаемые нижестоящими (локальными) элементами (Infimal decision problems) 80
 — — — — координируемость 120, 121
 — — — — модификации 132, 133
 — — — — обмен информацией между нижестоящими элементами 80, 110—112
 — — — — связь с вышестоящим решающим элементом 77, 84
 Иерархические системы (Hierarchical systems), классификация 54—71
 — — основные характеристики 22, 23, 27—29, 48—52, 54—56, 68—71
 Иерархия принятия решений (Decision-making hierarchy) 108
 Итеративные процедуры координации (Iterative coordination) 129, 269—273
 — — — сходимость 271—273
 Кажущаяся глобальная целевая функция (Apparent overall objective function) 149, 212—215
 Конфликт внутриуровневый (Intra-level conflict) 147
 — межуровневый (Interlevel conflict) 147
 Координатор (Coordinator) 111
 Координация 38—46, 129, 209—215
 — в целях улучшения работы системы 302
 — «линейная» (On-line) 304
 — последовательная (Sequential) 304, 315, 327—329
 — принципы см. Принципы координации
 — способы 79, 80, 142, 143
 — теоретико-игровой подход 38, 81, 82
 Координирование на основе оценки взаимодействий (Interaction estimation coordination) 79
 — — — прогнозирования взаимо-

- действий (Interaction prediction coordination) 79
- — «развязывания» взаимодействий (Interaction decoupling coordination) 80
 - путем «наделения ответственностью» (Load-type coordination) 80
 - — создания коалиций (Coalition-type coordination) 80
- Координирующие сигналы (Coordination inputs) 111, 144, 311
- компенсирующие (Compensating coordination inputs) 186—189
 - множества 111
- «Кормчий» (Helmsman) 44
- Критериальная функция (Tolerance function) 37, 97, 322, 323
- «Линейная» координация (On-line coordination) 304
- Линейные модификации локальных функций качества с нулевой суммой (Linear zero-sum infimal performance modifications) 238—240
- Локальная согласованность (Infimal harmony) 153—155
- — безусловная (Unrestricted) 153
 - — ограниченная (Restricted) 155
- Локально оптимальное решение (Optimal infimal decision) 141
- Локальные функции затрат (Infimal cost functions) 148
- — — сбалансированные (сбалансированные) (Balanced) 316, 323—327
 - — качества (Infimal performance functions), модификации 168, 172, 238—240, 254—258
 - — — нелинейные 254—258
 - — — получение 165—168
- Межуровневая согласованность безусловная 153
- — ограниченная 155
- Межуровневый конфликт (Interlevel conflict) 147
- Многослойные системы (Multilayer systems), понятие 62, 63, 73
- — примеры 21—25
 - — формализация 106, 107
- Многоуровневые системы (см. также Многослойные, Многошелонные, Стратифицированные системы) 84, 85
- — недостатки 84, 85
 - — преимущества 85—87
 - — примеры 18—29
- Многоуровневый системный подход к изучению организаций 35, 36
- Многошелонные системы (Multiecheloned systems), понятие 68, 70
- — примеры 21—28
 - — формализация 108
- Модель управляемого процесса (см. также Выходная функция) 97
- Модификации (Modifications), виды 81, 82
- задач, решаемых нпжестоящими элементами 132, 133
 - локальных функций качества 168, 169, 173, 240, 254
 - — — — нелинейные 254—258
 - — — — с нулевой суммой 172
 - образцов 81, 82
 - целей 81, 82
- Монотонности свойство (Monotonicity property) 150
- Неопределенностей множества (Uncertainty sets) 97
- Ограничения 274—281 (Constraints)
- согласованные (Balanced) 278—281
- Ограниченная локальная согласованность см. Локальная согласованность ограниченная
- межуровневая согласованность см. Межуровневая согласованность ограниченная
- Операторы оценки косвенного эффекта управляющих воздействий (Goal-interaction operators), линеаризованная форма 232—234, 245—254, 266—269
- — — — — полные (Total) 175, 232
- Операторы оценки косвенного эффекта управляющих воздействий, соотношения 177, 178
- — — — — типы 174—176
 - — — — — частные (Partial) 175, 232, 233
 - — эффективности связующих входов 176, 177, 233
- Оптимальный координирующий сигнал (Optimal coordination input) 144, 251—254
- Оценки взаимодействий принцип (Estimation principle) 126, 127
- Оценочные диапазоны (Estimated ranges) 310—323
- Последовательная координация см. Координация последовательная
- — процедуры 327—329
- Постулат совместности (Consistency postulate) 121, 122

- Принцип оценки взаимодействий** (Interaction estimation principle) 126, 127, 313, 314
- — — как основа для последовательной координации 315
 - — — и координируемость 127, 315
 - — — — — при выбранном эталонном управлении 320, 321
 - — — — — применимость 314, 316, 320, 321
 - прогнозирования взаимодействий (Interaction prediction principle) 125)
 - — — с использованием специального отображения 147, 196, 197, 262—269, 287, 297
 - — — при фиксированных целях (Without goal-coordination) 190—194
 - — — стратегия координации 208, 273
 - согласования взаимодействий (Interaction balance principle) 126
 - — — применимость 178—180, 183, 235, 279
 - — — координируемость 236, 237, 240, 247, 251, 252, 284, 285, 290—294
 - — — с использованием нелинейных модификаций 254—258
 - — — стратегия координации 205, 270—273
 - — — экономическая интерпретация 46, 47
- Принцип согласования функций качества** (Performance balance principle) 145
- — — координируемость 178, 179
 - — — — — применимость 178, 179
 - — — — — стратегия координации 205
- Принципы координации, виды** 124, 125, 145, 146
- — общее понятие 40, 41
- Приоритет действий** (Priority of action) 147, 148
- Разрешение конфликтов в двухуровневой системе** 147, 148
- Решаемая проблема (задача) см. Задача принятия решения**
- Связующие сигналы (входы)** (Interface inputs) 114
- функции подпроцессов (Subprocess coupling functions) 114, 115
 - — — сепарабельность 239
- Седлового типа условие для координируемости** (Saddle-condition for coordinability) 163, 165, 185
- Сепарабельная глобальная функция качества см. Глобальная функция качества сепарабельная**
- Сепарабельность связующих функций подпроцессов** 239
- Система, формальное определение** 91
- принятия решений («решающая» система) 94, 95, 100—102
 - — — иерархия 108
 - — — классификация 68, 69
 - — — формальное определение 100, 101
- Слои** (Layers) 23, 24, 66, 67, 71—73
- — функциональной иерархии 66, 67, 72—74
- Согласованность, свойства** (Harmony properties) 152—155
- Стратегия координации** 201—203, 205, 206, 208, 270—273
- — сходимость 271—273
- Стратифицированная система** 19, 56—58
- — понятие 56—58
 - — формализованное определение 103, 104
- Страты** (Strata) 8, 19, 56, 59—62, 82—84, 103—105
- Улучшения задача** (Improvement problem) 309
- Уровни** (Levels) (см. также Страты, Слои, Эшелоны) 19—21, 56
- Условие координируемости** (Coordinating condition) 127
- седлового типа для координируемости 163—165, 185
- Функциональная система** 91
- Функция взаимодействия подпроцессов** (Subprocess interaction function) 115, 116
- допустимости (Tolerance function) см. Критериальная функция
- Целевая функция** (Objective function) 97
- — глобальных затрат (Overall objective cost function) 148
 - — — — — аддитивность 172
 - — — — — сепарабельность 167
- Эшелоны** (Echelons) 68, 71, 108

ОГЛАВЛЕНИЕ

Предисловие к русскому изданию	5
Предисловие автора	12

Часть I. Иерархические системы

Глава 1. Примеры	17
1. Иерархические системы в крупных автоматизированных промышленных комплексах	18
2. Использование концепций многоуровневых систем в теории организаций	29
3. Иерархический порядок в природных структурах	48
Глава 2. Концептуализация	53
1. Что такое многоуровневая иерархическая структура?	53
2. Основные виды иерархий	55
3. Зависимость между уровнями и координируемость	76
4. Почему именно иерархические структуры?	85
Глава 3. Формализация	88
1. Введение	89
2. Формальное определение абстрактной системы	91
3. Система принятия решений	94
4. Основные понятия в теории иерархических систем и их формализация	103
Глава 4. Координация	109
1. Общее описание двухуровневой системы	109
2. Декомпозиция подсистем	113
3. Координируемость	118
4. Постулат совместимости	121
5. Принципы координации	123
6. Различные аспекты проблемы координации	128

Часть II. Математическая теория координации

Глава 5. Общая теория координации для систем оптимизации при отсутствии ограничений	135
1. Введение	135
2. Двухуровневая система оптимизации	139
3. Принципы координации и задача координатора	145
4. Разрешение конфликтов в двухуровневой системе	147
5. Формирование локальных функций качества и их модификация	165
6. Координируемость при использовании принципа согласования	178
7. Координируемость при использовании принципа прогнозирования взаимодействий	189
8. Процесс координации	201
Глава 6. Оптимальная координация динамических систем	216
1. Введение	216
2. Предварительные замечания	220
3. Применение принципов согласования	235
4. Применение принципа прогнозирования	258
5. Итеративные процессы координации	269

Глава 7. Координация оптимизирующих систем при наличии ограничений. Математическое программирование	274
1. Введение	274
2. Применимость принципов координации	276
3. Координация задач выпуклого программирования	281
4. Координация задач линейного программирования	289
5. Замечание по поводу декомпозиции Данцига — Вольфа	297
Глава 8. «Линейная» координация: координация в целях улучшения характеристик функционирования системы	302
1. Введение	302
2. Концепция «линейной» координации с позиций общей теории систем	309
3. Приложение к системам, описываемым линейными уравнениями	323
4. Последовательная координация с адаптацией	327
Приложение. Комментарий к списку литературы	333
Литература	335
Предметный указатель	340

УВАЖАЕМЫЙ ЧИТАТЕЛЬ!

Ваши замечания о содержании книги, ее оформлении, качестве перевода и другие просим присылать по адресу:

129820, Москва, И-110, 1-й Рижский пер., д. 2, издательство «Мир».

М. МЕСАРОВИЧ, Д. МАКО, И. ТАКАХАРА
ТЕОРИЯ ИЕРАРХИЧЕСКИХ МНОГОУРОВНЕВЫХ СИСТЕМ

Редактор В. Я. ФРИДМАН
Художник Е. И. Волков
Художественный редактор Ю. С. Урманчеев
Технический редактор И. К. Держа
Корректор А. Я. Шехтер

Сдано в набор 29/IX 1972 г. Подписано к печати 23/II 1973 г. Бумага № 1,
60×90/16 = 10,75 бум. л. 21,50 печ. л. Уч.-изд. л. 20,99 Изд. № 20/8402

Цена 1 р. 74 к. Зак. 0711

ИЗДАТЕЛЬСТВО «МИР»

Москва, 1-й Рижский пер., 2

Ордена Трудового Красного знамени Московская типография № 7
«Искра революции» Союзполиграфпрома при Государственном комитете
Совета Министров СССР по делам издательств, полиграфии и книжной торговли
Москва, Трехпрудный пер., 9