

Парадигма развития науки

Методологическое обеспечение

А.Е. Кононюк

КОНСАЛТОЛОГИЯ

**ОБЩАЯ ТЕОРИЯ
КОНСАЛТИНГА**

Книга 4

**Киев
Освита України
2011**

УДК 51 (075.8)

ББК В161.я7

К65.

Рецензент: *Н.К. Печурин* - д-р техн. наук, проф.
(Национальный авиационный университет).

Кононюк А.Е.

К65 Консалтология. Общая теория консалтинга
К.4: "Освита Украины", 2011. К.4- 508 с.

ISBN 978-966-7599-50-8

Настоящая работа посвящена вопросам развития общей теории консалтинга с использованием научных методов формирования рекомендаций для решения задач консультируемых проблем. Сформулированы основные положения построения автоматизированных консультационных процессов. Рассмотрены принципы построения систем автоматизированного консультирования (САК). С позиций пользователя (лица, формирующего рекомендации) изложены основные положения, связанные с разработкой, исследованием и реализацией сформированных рекомендаций по решению задач консультируемых проблем различных проблемных областей.

Книга предназначена для научных работников, магистров, аспирантов, докторантов соответствующих специальностей.

ББК В161.я7

ISBN 978-966-7599-50-8

©А.Е. Кононюк, 2011

Оглавление

8. Методы анализа математических моделей консультируемых проблем	5
8.1. Требования к методам анализа.....	5
8.2. Анализ статических и динамических режимов.....	9
8.3. Метод анализа переходных процессов.....	28
8.4. Выбор метода численного интегрирования.....	35
8.5. Методы анализа повышенной эффективности.....	59
8.6. Методы анализа логических и функциональных схем консультируемых проблем	66
8.7. Методы многовариантного анализа	71
8.8. Анализ чувствительности	75
8.9. Анализ стационарных режимов.....	82
8.10. Задачи анализа структур консультируемых проблем.....	90
8.11. Введение в анализ функциональных и системных структур КП	93
8.12. Анализ структур КП.....	104
8.13. Анализ КП на начальных стадиях консультирования	109
8.14. Анализ структурно-топологических характеристик	113
8.15. Анализ количественных характеристик структур консультируемых проблем	122
8.16. Декомпозиция структуры консультируемой проблемы.....	138
9. Синтез формируемых рекомендаций по решению задач консультируемых проблем в САК.....	146
9.1. Постановка задачи синтеза формирования рекомендаций....	146
9.2. Структурный синтез и параметрическая оптимизация формируемых рекомендаций	152
9.3. Разновидности консультационных задач оптимизации.....	162
9.4. Показатели эффективности сформированных рекомендаций и выбор методов поиска экстремума.....	166
9.5. Задачи формирования рекомендаций по оптимизации допусков и консультационных требований.....	176
9.6. Методы формирования рекомендаций по оптимизации различных процессов	182
9.7. Формирование рекомендаций при решении консультационных задач функционального и структурного синтеза КП	189
9.8. Постановка задачи синтеза КП	198
9.9. Критерии синтеза	202
9.10. Особенности решения задач структурного синтеза	205
9.11. Методы выбора структуры	208

9.12. Методы и алгоритмы оптимизации структур	225
9.13. Последовательные методы в задачах консультирования сложных проблем	237
9.14. Алгоритм распределения консультационных функций по модулям САК	246
9.15. Синтез комплекса технических средств САК.....	252
10. Формирование и функционирование консультационных модулей	264
10.1. Формирование консультационных модулей	266
10.1.1. Базовая форма представления математических моделей консультируемых проблем в САК	266
10.1.2. Формирование консультационных модулей для их автономного функционирования	294
10.1.3. Формирование среды обмена информацией между КМ (при вертикальном взаимодействии)	338
10.1.4. Формирование КМ на базе управляющего модуля САК	375
10.2. Функционирование консультационных модулей	409
10.2.1. Содержание процедур функционирования консультационных модулей	409
10.2.2. Основные понятия метода учета неопределенности при формировании рекомендаций (метод ПРИНН)	419
10.2.3. Процедура формирования рекомендации при функционировании консультационного модуля	432
10.2.4. Формальное (математическое) обоснование описания многоцелевой системы консультирования.....	435
10.2.5. Расчет n-обобщенных потерь	463
10.2.6. Алгоритмы выбора формируемых рекомендаций при неопределенности внешних условий и задач функционирования консультируемой проблемы	473
Литература.....	503

8. Методы анализа математических моделей консультируемых проблем

8.1. Требования к методам анализа

Многообразие применяемых методов. Анализ консультируемых проблем в САК основан на их математическом моделировании, т. е. на исследовании консультируемых проблем путем оперирования их математическими моделями.

В разделе 6 показано, что функциональными моделями консультируемых проблем на макроуровне являются системы ОДУ, которые могут быть представлены в общем виде (6.75), либо предварительно приведены линеаризацией к виду (6.76), либо алгебраизацией и линеаризацией к виду системы линейных алгебраических уравнений (6.77). К таким же формам уравнений с помощью методов конечных разностей или конечных элементов приводятся ММ консультируемых проблем на микроуровне.

Таким образом, математические модели консультируемых проблем на микро- и макроуровнях сводятся к системам обыкновенных дифференциальных и конечных уравнений (под конечными уравнениями понимаются алгебраические и трансцендентные уравнения). ***Оперирование такими моделями в процедурах одновариантного анализа означает решение соответствующих уравнений.*** Поэтому ***методы одновариантного анализа на этих уровнях суть численные методы решения систем дифференциальных и конечных уравнений.***

Многовариантный анализ заключается в многократном повторении решения систем названных уравнений при варьировании внутренними и (или) внешними параметрами. Типовыми процедурами ***многовариантного анализа***, реализуемыми в САК, являются ***процедуры анализа чувствительности и статистического анализа.***

Особенностью некоторых консультируемых проблем является использование на функционально-логическом уровне консультирования в качестве ***моделей систем логических уравнений***, а в качестве ***методов анализа — методов решения этих систем.***

В вычислительной математике известно большое количество методов численного решения систем уравнений. Однако применение большинства из них в САК оказывается неэффективным, что объясняется особенностями ММ консультируемых проблем. Поэтому при создании математического обеспечения САК усилия направлены

не только на разработку математических моделей, но и в не меньшей мере на развитие **численных методов и алгоритмов анализа**. Поскольку эффективность метода зависит от особенностей решаемой задачи, целесообразна реализация в САК более чем одного метода для каждого класса решаемых уравнений. Выбор метода в большинстве случаев возлагается на пользователя (консультанта), что требует от него соответствующей подготовки в области численных методов анализа.

При неудачном выборе моделей или методов анализа пользователь САК может столкнуться с рядом трудностей: чрезмерной продолжительностью вычислений, несходимостью или неустойчивостью вычислительного процесса, малой точностью получаемых результатов.

В САК целесообразно применять методы, исключаящие возможность возникновения подобных ситуаций, т. е. методы, обладающие свойствами высокой экономичности, надежности и точности. Однако эти требования противоречивы и не всегда удается их одновременное удовлетворение в должной мере, поэтому важно уметь распознавать неблагоприятные ситуации и знать факторы, изменение которых может привести к исправлению положения.

Экономичность. Экономичность метода характеризуется затратами вычислительных ресурсов (машинного времени T_m и машинной памяти P_m) на его применение в некоторых заранее оговоренных условиях (например, в тестовых задачах, в среднем по группе задач определенного класса и т.п.). На показатели T_m и P_m обычно оказывают влияние многие факторы и в первую очередь **размерность решаемой задачи** N . В качестве N принимают порядок решаемой системы уравнений, число элементов, из которых состоит моделируемая проблема, и т. п.

При сравнении методов по экономичности часто не интересуются абсолютными показателями T_m и P_m в конкретной ситуации, а исследуют характер зависимости T_m и P_m от N . Наиболее эффективные методы имеют линейную или близкую к линейной зависимость показателей экономичности от сложности задачи. Для многих численных методов характерна полиномиальная зависимость T_m от N :

$$T_m \approx cN^\alpha$$

где c — коэффициент пропорциональности. При $\alpha \geq 2$ метод имеет заметное ограничение по сложности решаемых задач.

Надежность. Надежность метода оценивается как вероятность получения правильных результатов при использовании метода для решения задач заданного класса. Обычно условия применимости

метода связаны с такими характеристиками ММ анализируемых проблем, которые пользователь (консультант) не может оценить заранее имеющимися в его распоряжении средствами, поэтому возможны ситуации, когда вычислительный процесс оказывается неустойчивым или отсутствует сходимость, что может выражаться в заикливании или останове ЭВМ из-за переполнения разрядной сетки. В САК стараются применять надежные методы. Однако высоконадежные методы часто характеризуются недостаточной экономичностью. В этом случае целесообразно комбинирование методов с переходом к трудоемким, но надежным методам только в результате автоматического распознавания ситуаций несходимости или неустойчивости вычислений.

Точность. Погрешности решения задачи определяются особенностями используемых моделей, численных методов, ограниченностью разрядной сетки ЭВМ. Каждый источник погрешности должен контролироваться, с тем чтобы погрешности не превысили предельно допустимые. Обычно точность результатов, получаемых с помощью численного метода, зависит от некоторых параметров, выбираемых «по умолчанию» или задаваемых среди исходных данных. С помощью этих параметров можно управлять погрешностями решения, но необходимо помнить, что снижение погрешностей возможно лишь до некоторого отличного от нуля предела и, как правило, сопровождается увеличением затрат машинного времени. Целесообразно в математическом обеспечении САК иметь не один, а несколько методов одинакового целевого назначения, но с различными возможностями компромиссного удовлетворения противоречивых требований точности и экономичности.

Пользователь САК должен также знать, что явления заикливания вычислений или переполнения разрядной сетки могут происходить не только из-за недостатков выбранного численного метода, но и из-за ошибок в задании исходных данных. Некоторые ошибки, связанные с нарушением формальных правил грамматики входного языка, распознаются автоматически. Однако ряд ошибок не может быть выявлен формальными средствами без участия пользователя. Примерами таких ошибок могут быть ошибки в задании численных значений параметров или в задании соединений в анализируемой схеме. Если эти ошибки приводят к получению модели самовозбуждающейся схемы, то возможны явления заикливания и переполнения разрядной сетки.

Направления повышения эффективности методов анализа. Высокие размерности задач консультирования, необходимость

выполнения многих вариантов решения систем уравнений при формировании рекомендаций для решения задач сложных консультируемых проблем обуславливают большие затраты вычислительных ресурсов. Поэтому повышение экономичности методов анализа при соблюдении требований точности является актуальной задачей создания и совершенствования математического обеспечения САК. Эта задача решается на основе идей и методов, группируемых в несколько направлений.

Декомпозиция—деление модели консультируемой проблемы на части и раздельный анализ получающихся частей. Если $T_m = cN^\alpha$, то после деления модели на m равных частей затраты машинного времени приближенно оцениваются величиной $cm(N/m)^\alpha$, т. е. уменьшаются в $q \approx m^{\alpha-1}$ раз. Однако раздельный анализ происходит в условиях принятия упрощающих предположений о взаимном влиянии частей, т. е. сопровождается увеличением погрешностей расчетов. Декомпозиция составляет основу блочно-иерархического подхода к консультированию. Это направление предлагается широко использовать как в автоматизированных, так и в неавтоматизированных методах консультирования.

Диакоптика — направление исследования сложных консультируемых проблем по частям, отличающееся от декомпозиции тем, что раздельный анализ осуществляется без упрощающих предположений о влиянии частей друг на друга. Экономичность диакоптических методов соизмерима с экономичностью обычных композиционных методов, а точность выше.

Учет разреженности матриц — направление экономичной организации операций над разреженными матрицами. Матрицу называют разреженной, если в ней преобладают нулевые элементы. Отказ от хранения нулевых элементов и реализация алгоритмов, в которых игнорируются арифметические действия над нулевыми элементами, могут дать значительную экономию T_m и P_m .

Учет событийности — направление, называемое также учетом временной разреженности моделей и основанное на исключении из вычислительного процесса действий над неактивными переменными. *Неактивной* на интервале $[t, t+\Delta t]$ переменной называют величину, изменения которой на этом интервале не превышают достаточно малого заранее заданного значения. В моделях сложных консультируемых проблем в каждый момент модельного времени большинство переменных неактивно. Моделирование, основанное на учете событийности, принято называть *событийным моделированием*. В алгоритмах событийного моделирования необходимо реализовать

критерии своевременного включения переменных и соответствующих им частей моделей в группу неактивных (латентных) и своевременного их исключения из этой группы.

Комбинирование моделей и методов — одновременное использование при решении конкретной задачи нескольких разнотипных моделей или методов анализа одинакового целевого назначения. Комбинирование может быть пространственным, если разнотипные модели или методы применяют в разных частях общей модели, или временным, если их применяют на разных этапах вычислительного процесса. Пространственное комбинирование является частным случаем диакоптического подхода, так как подразумевает разделение модели на части (фрагменты). Повышение эффективности при комбинировании моделей и методов основано на использовании наиболее подходящих моделей и методов для данного фрагмента и данного этапа вычислений. Пространственное комбинирование моделей, относящихся к разным иерархическим уровням, называют *многоуровневым* (или *смешанным*) *моделированием*.

8.2. Анализ статических и динамических режимов

В общем случае *математическая модель консультируемой проблемы* представляет собой совместную *систему нелинейных дифференциально-алгебраических уравнений вида*

$$F(x', x, t) = 0, \quad (8.1)$$

где x — вектор переменных модели; F — вектор-функция.

Методы решения системы (8.1), обеспечивающие получение динамических и статических характеристик консультируемой проблемы, состоят из *процедур численного интегрирования систем обыкновенных дифференциальных уравнений и процедур решения систем нелинейных алгебраических уравнений*. Естественно, что выбор конкретных методов решения для каждой из этих процедур обуславливает однозначно вид формул алгебраизации и линеаризации компонентных соотношений.

Осуществляя алгебраизацию компонентных уравнений, содержащих производные по времени, с помощью выражений, аналогичных уравнениям (6.44)—(6.45), преобразуем исходную систему уравнений (8.1) в систему нелинейных алгебраических уравнений

$$F(x_{n+l}, t_{n+l}) = 0. \quad (8.2)$$

Следует отметить, что уравнения (8.2) используются для описания статических режимов консультируемой проблемы в целом

$$F(x, t) = 0. \quad (8.3)$$

Однако в выражении (8.2) вид функционала F зависит от интервала времени решения, в результате чего это выражение решается на каждом временном шаге. Следовательно, задача определения динамического режима консультируемой проблемы сводится к многократной процедуре нахождения ее квазистатических состояний для каждого интервала времени, на которых проводится алгебраизация инерционных компонентных соотношений. Полный процесс совместного решения системы нелинейных дифференциально-алгебраических уравнений состоит из следующих операций.

1. Выбор временного интервала ($t_n = t_{n-1} + h$).
2. Алгебраизация реактивных компонентных уравнений.
3. Формирование системы нелинейных алгебраических уравнений.
4. Решение системы нелинейных алгебраических уравнений.
5. Проверка погрешности решения.
6. Получение результата.

В свою очередь, нелинейное алгебраическое уравнение типа (8.3) или (8.2) решается итерационно, например, с помощью метода Ньютона — Рафсона, при этом

$$x_{n+1}^{(m+1)} = x_{n+1}^{(m)} - \frac{F(x_{n+1}^{(m)}, t_{n+1})}{F'(x_{n+1}^{(m)}, t_{n+1})}, \quad (8.4)$$

где

$$F'(x_{n+1}^{(m)}, t_{n+1}) = \frac{\partial F}{\partial x_{n+1}^{(m)}}$$

— матрица Якоби; m — номер текущей итерации ($m = 0$ соответствует некоторое начальное значение x_{n+1}^0). Матрица Якоби, определяемая при линейаризации компонентных зависимостей членами с первыми производными ряда Тейлора, аналогична выражению (6.41):

$$F'(x_{n+1}^{(m)}, t_{n+1}) = A(x_{n+1}^{(m)}), \quad (8.5)$$

и совпадает по форме с матрицами математических моделей консультируемой проблемы, рассмотренных в разделе 6.

Подставляя уравнение (8.5) в выражение (8.4), находим

$$A(x_{n+1}^{(m)}) x_{n+1}^{(m+1)} = b(x_{n+1}^{(m)}), \quad (8.6)$$

где

$$b(x_{n+1}^{(m)}) = A(x_{n+1}^{(m)}) x_{n+1}^{(m)} - F(x_{n+1}^{(m)}, t_{n+1}).$$

Таким образом, система линейных уравнений (8.6) итеративно решается до тех пор, пока максимальная погрешность составляющих вектора переменных x становится меньше заданной, т. е.

$$E_{\max} = \max_i \left| \frac{x_{i(n+1)} - x_{i(n+1)}^{(m)}}{x_{\max i}} \right| < \varepsilon, \quad i = 1, 2, \dots, n. \quad (8.7)$$

Полный процесс решения нелинейных алгебраических уравнений состоит из следующих операций.

1. Выбор начального значения x_{n+1}^i .
2. Линеаризация нелинейных компонентов.
3. Формирование системы линейных алгебраических уравнений.
4. Решение системы линейных алгебраических уравнений.
5. Сравнение решения и начального значения.
6. Получение результата.

Следовательно, анализ уравнений модели консультируемой проблемы, по существу, сводится к многократному (сотни, а иногда тысячи раз) решению системы линейных алгебраических уравнений (8.6), в котором выделяются два цикла:

1) внешний цикл временных итераций по индексу n , когда корректируются аппроксимации для производных по времени от переменных модели;

2) внутренний цикл итераций по индексу m , когда на каждом выбранном временном интервале с помощью метода Ньютона уточняется линейная аппроксимация нелинейных переменных модели. Внутренний цикл решения по индексу m имеет самостоятельное значение при анализе статических режимов работы.

Модификации метода Ньютона. Итерации Ньютона могут не сходиться, если начальное приближение вектора неизвестных x_{n+x}^0 , контролируемое консультантом консультируемой проблемы, окажется далеким от решения и, следовательно, определенное матрицей Якоби направление сходимости будет значительно отличаться от истинного.

На практике применяют различные модификации метода Ньютона. Так, для демпфирования колебаний составляющих вектора x в процессе итерационного решения систем (8.2) и (8.3) искусственно ограничивают величину шага приращения колеблющихся переменных:

$$\tilde{x}_i^{(m+1)} = x_i^{(m)} + \lambda (x_i^{(m+1)} - x_i^{(m)}), \quad (8.8)$$

где $x_i^{(m+1)}$ — полученное из выражения (8.4) значений i -й компоненты вектора x ; λ — демпфирующий параметр, выбираемый при сопоставлении на m -й и $(m+1)$ -й итерации знаков погрешности определения той составляющей вектора x , которая обусловила максимум относительной погрешности на m -й итерации. При этом

$\lambda = \lambda \lambda_0$, если знак погрешности изменился, в противном случае $\lambda = \lambda \lambda_0$. Если при оценке окажется, что $\lambda > 1$, то выбирается значение $\lambda = 1$. Начальное значение λ также равно 1, а λ_0 выбирается эмпирически ($\lambda_0 \approx 0,75$).

При статическом анализе консультируемых проблем наряду с вопросами обеспечения сходимости метода Ньютона возникает опасность переполнения разрядной сетки ЭВМ при оценке трансцендентных компонентных зависимостей, используемых, например, для описания формируемых рекомендаций по размещению торговых точек. В этом случае исходная зависимость

$$i = i_0(e^{\theta x} - 1)$$

заменяется парой уравнений:

$$\begin{aligned} i &= i_0(e^{\theta x} - 1) \quad \text{при } x < x_{\max}; \\ i &= i_M [1 + \theta(x - x_{\max})] - i_0 \quad \text{при } x \geq x_{\max}, \end{aligned}$$

где

$$i_M = i_0 e^{\theta x_{\max}}.$$

Кроме того, при оценке приращений переменных математической модели консультируемой проблемы применяется модификация формулы (8.8), заключающаяся в выборе очередного приближения согласно следующему выражению:

$$\bar{x}_i^{(m+1)} = x_i^{(m)} + \text{sign}(\Delta x_i^{(m+1)}) \frac{r}{k} \ln [1 + k |\Delta x_i^{(m+1)}|], \quad (8.9)$$

где

$$\Delta x_i^{(m+1)} = x_i^{(m+1)} - x_i^{(m)}; \text{sign}(\Delta x_i^{(m+1)})$$

— знак приращения; r и k — эмпирические константы.

Для повышения сходимости и точности применяются также итерационные методы Ньютона, в соответствии с которыми

$$\bar{x}_i^{(m+1)} = x_i^{(m)} + \Delta x_1^{(m)} + \Delta x_2^{(m)} + \Delta x_3^{(m)} + \dots, \quad (8.10)$$

где

$$\begin{aligned} \Delta x_1^{(m)} &= -\frac{F(x_1^{(m)})}{F'(x_1^{(m)})}, \quad x_1^{(m)} = x_i^{(m)}; \\ \Delta x_2^{(m)} &= -\frac{F(x_2^{(m)})}{F'(x_2^{(m)})}, \quad x_2^{(m)} = x_i^{(m)} + \Delta x_1^{(m)}; \\ \Delta x_3^{(m)} &= -\frac{F(x_3^{(m)})}{F'(x_3^{(m)})}, \quad x_3^{(m)} = x_i^{(m)} + \Delta x_1^{(m)} + \Delta x_2^{(m)}. \end{aligned}$$

Если при анализе статического режима метод Ньютона не сходится к решению за заданное максимальное число итераций, то целесообразно перейти к одной из *разновидностей метода продол-*

жения решения по параметру. Например, можно осуществить постепенный перевод консультируемой проблемы из априорно известного начального состояния в искомый статический режим введением специального параметра $\tau \leq 1$ ($0 \leq \tau \leq 1$). При $\tau=0$ решение уравнений модели проблемы известно, а значение $\tau=1$ соответствует рассчитываемому режиму. Для каждого из фиксированных значений параметра τ решается частная задача статики. Метод Ньютона обладает **локальной сходимостью**, поэтому можно для достаточно малых значений приращений $\Delta\tau$ гарантировать сходимость задачи нахождения статического режима, если система уравнений исследуемого проблемы непрерывно дифференцируема на всем интервале изменения τ . Выбор приращений $\Delta\tau$ целесообразно осуществлять по процедуре, принятой для оценки локальной погрешности метода численного интегрирования.

Решение линейных систем. Для решения базовой линейной системы уравнений вида (8.6)

$$Ax = b \quad (8.11)$$

часто используется **метод последовательного исключения Гаусса** или одна из его модификаций. Если **правая часть** b в уравнении (8.11) меняется **многократно**, то рекомендуется применять метод

LU-преобразования, в соответствии с которым матрица A решаемой системы уравнений представляется произведением нижней треугольной матрицы с единичной диагональю L и верхней треугольной матрицы U :

$$A = LU. \quad (8.12)$$

При этом элементы матриц L и U определяются с помощью следующей рекуррентной процедуры:

$$\left. \begin{aligned} u_{sj} &= a_{sj} - \sum_{k=1}^{s-1} l_{sk}u_{kj}, \quad j = s, s+1, \dots, n; \\ l_{is} &= (a_{is} - \sum_{k=1}^{s-1} l_{ik}u_{ks})/u_{ss}, \quad i = s+1, \dots, n, \end{aligned} \right\} s = 1, 2, \dots, n,$$

где n — размерность решаемой системы уравнений (8.11), а символами a , l , u обозначены элементы матриц A , L и U соответственно. Например, для $s=1, 2, 3$ получаем соответственно:

$$\left. \begin{array}{l}
 u_{11} = a_{11}; \\
 u_{12} = a_{12}; \\
 \vdots \\
 u_{1n} = a_{1n}; \\
 l_{21} = a_{21}/a_{11}; \\
 l_{31} = a_{31}/a_{11}; \\
 \vdots \\
 l_{n1} = a_{n1}/a_{11};
 \end{array} \right\} s = 1$$

$$\left. \begin{array}{l}
 u_{22} = a_{22} - l_{21}u_{12}; \\
 u_{23} = a_{23} - l_{21}u_{13}; \\
 \vdots \\
 u_{2n} = a_{2n} - l_{21}u_{1n}; \\
 l_{32} = (a_{32} - l_{31}u_{12})/u_{22}; \\
 l_{42} = (a_{42} - l_{41}u_{12})/u_{22}; \\
 \vdots \\
 l_{n2} = a_{n2} - l_{n1}u_{12}/u_{22};
 \end{array} \right\} s = 2$$

$$\left. \begin{array}{l}
 u_{33} = a_{33} - l_{31}u_{13} - l_{32}u_{23}; \\
 u_{34} = a_{34} - l_{31}u_{14} - l_{32}u_{24}; \\
 \vdots \\
 u_{3n} = a_{3n} - l_{31}u_{1n} - l_{32}u_{2n}; \\
 l_{43} = (a_{43} - l_{41}u_{13} - l_{42}u_{23})/u_{33}; \\
 l_{53} = (a_{53} - l_{51}u_{13} - l_{52}u_{23})/u_{33}; \\
 \vdots \\
 l_{n3} = (a_{n3} - l_{n1}u_{13} - l_{n2}u_{23})/u_{33};
 \end{array} \right\} s = 3$$

(8.14)

После LU -преобразования матрицы A решение системы (8.11) заменяется последовательным решением двух систем линейных уравнений с треугольными матрицами:

$$\left. \begin{array}{l}
 Ly = b; \\
 Ux = y,
 \end{array} \right\} \quad (8.15)$$

которые решаются простой обратной подстановкой

$$y_i = b_i - \sum_{s=1}^{i-1} l_{is}y_s, \quad i = 1, 2, \dots, n; \quad (8.16)$$

$$x_t = (y_t - \sum_{s=t+1}^n u_{ts}x_s)/u_{tt}, \quad t = n, n-1, \dots, 1. \quad (8.17)$$

Пример 1. Рассмотрим процедуру LU -преобразования на примере системы уравнений

$$\begin{bmatrix} 2 & -7 & 4 \\ 1 & 9 & -6 \\ -3 & 8 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 9 \\ 1 \\ 6 \end{bmatrix}.$$

Используя выражения (8.12) — (8.14), определяем

$$\begin{bmatrix} 2 & -1 & 4 \\ 1 & 9 & -6 \\ -3 & 8 & 5 \end{bmatrix} = \begin{bmatrix} 1 & & \\ & 1/2 & 1 \\ -3/2 & -1/5 & 1 \end{bmatrix} \begin{bmatrix} 2 & -7 & 4 \\ 25/2 & -8 \\ 47/5 \end{bmatrix}.$$

На основании формулы (8.15) и (8.16) строим частную систему уравнений для прямого хода решения

$$\begin{bmatrix} 1 & & \\ & 1/2 & 1 \\ -3/2 & -1/5 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 9 \\ 1 \\ 6 \end{bmatrix}$$

и получаем из нее $y = [9, -7/2, 188/10]^t$.

Далее, пользуясь выражениями (8.15) и (8.17), формируем частную систему уравнений для обратного хода решения

$$\begin{bmatrix} 2 & -7 & 4 \\ 25/2 & -8 \\ 47/5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 9 \\ -7/2 \\ 188/10 \end{bmatrix},$$

из которой находим искомый вектор переменных

$$x = [4, 1, 2]^t.$$

Во многих случаях решаемые системы линейных уравнений являются разреженными, поэтому при их решении основное внимание уделяется вопросам целесообразного хранения (кодирования) информации о решаемой системе уравнений, поддержанию наибольшей степени разреженности на всех этапах вычислительного процесса и разработке специальных методов оптимальной организации вычислений.

Дополнительно к способам кодирования, рассмотренным в разделе 6, применительно к процедуре LU -преобразования используется кодировка по способу «строк и столбцов». В этом случае массив IC содержит индексы строк НЭ матрицы L и индексы столбцов НЭ матрицы U , а в массиве VA последовательно размещаются НЭ столбцов матрицы L и строк матрицы U . Границы участков массивов IC и VA , соответствующие НЭ отдельных столбцов матрицы L и НЭ отдельных строк матрицы U , фиксируются в массиве IR , длина которого

увеличивается вдвое по сравнению с ранее рассмотренными. Дополнительно формируется массив IVA , в котором указываются номера ячеек массива VA , содержимое которых изменяется в процессе LU -преобразования матрицы A . Благодаря измененной форме «упаковки» матрицы, согласованной с последовательностью процесса LU -преобразования, отсутствуют логические операции и процедуры поиска обрабатываемых НЭ на каждом шаге вычислений. При этом формируемые элементы матрицы L и U записываются на позиции «обработанных» элементов матрицы A , что существенно экономит массив используемой памяти.

Необходимость принятия специальных мер для сохранения разреженности уравнений модели консультируемой проблемы в процессе их решения объясняется тем, что при LU -преобразовании матрицы A возможно появление новых ненулевых элементов (ННЭ). Количество ННЭ существенно зависит от того, какие из элементов матрицы A и в какой последовательности будут выбраны в качестве главных (диагональных).

При многократном решении уравнений (8.11) целесообразно процесс LU -преобразования построить так, чтобы по возможности уменьшить количество ННЭ и изменений в уже имеющихся ненулевых элементах (ИНЭ). При этом сокращается число арифметических операций при решении системы (8.11), повышается точность решения и уменьшаются затраты памяти, необходимые для хранения информации о системе уравнений (8.11) и (8.6). Оптимизация процесса LU -преобразования достигается предварительным упорядочением решаемой системы уравнений.

Существует большое количество стратегий упорядочения, представляющих собой n -шаговую процедуру, каждый шаг которой включает:

- выбор среди элементов матрицы A текущего главного элемента, удовлетворяющего критерию упорядочения и условия вычислительной устойчивости;
- перестановку строк и столбцов матрицы A так, чтобы уже выбранные главные элементы лежали на главной диагонали;
- анализ полученной ненулевой структуры с целью определения позиций ННЭ и проведение соответствующей коррекции информационных массивов.

Перестановка строк и столбцов может выполняться и однократно после определения всех главных элементов, при этом на каждом шаге упорядочения новые номера строк и столбцов запоминаются в специальных векторах перестановок.

Одним из наиболее распространенных критериев упорядочения разреженных систем уравнений является критерий Марковица. В соответствии с этим критерием главный элемент выбирается на пересечении тех строк и столбцов матрицы, среди которых главный элемент еще не выбирался и которым соответствует минимальный вес

$$\omega_{ij} = \min_{ij} [(c_i - 1)(r_j - 1)], \quad (8.18)$$

где r_j , c_i — число НЭ соответственно в j -й строке и i -м столбце, расположенных в непреобразованной части упорядоченной матрицы A . Вес ω_{ij} соответствует максимальному числу ННЭ, которые могут появиться, если в качестве главного будет выбран элемент a_{ji} .

При упорядочении уравнений модели консультируемой проблемы обычно трудно обеспечить выполнение условий вычислительной устойчивости, связанных с выбором для каждого шага LU -преобразования в качестве главного элемента наибольшего по абсолютному значению элемента некоторого столбца (строки) из непреобразованной еще части матрицы A . Это объясняется, с одной стороны, противоречивостью требований сохранения разреженности матрицы A и условий вычислительной устойчивости и, с другой стороны, сложностью предварительной оценки величин, изменяющихся в процессе вычислений элементов матрицы. Поэтому на практике ограничиваются контролем величины главных элементов упорядоченной системы уравнений, стремясь обеспечить, чтобы все выбранные в качестве главных элементы были больше некоторого минимально допустимого для них значения. При этом на каждом шаге упорядочения выделяемый согласно уравнению (8.18) элемент проверяется на малость по величине. Если величина элемента больше допустимого минимального значения, то он выбирается в качестве главного, в противном случае поиск продолжается. При упорядочении уравнений модели консультируемой проблемы учитываются особенности уравнений. Например, уравнения модели OM_{20} по методу узловых величин, частично GM_{10} (таблично-узловой модели) и других характеризуются тем, что элементы главной диагонали матриц их коэффициентов доминируют по величине в своих строках. Поэтому в процессе упорядочения при выборе главных элементов даже при равных весах ω_{ij} предпочтение целесообразно отдать диагональным элементам этих матриц.

Способы повышения эффективности решения. При многократном решении уравнений (8.11) наряду с использованием их разреженности важно учесть также различия в частоте изменения элементов матриц решаемых уравнений. Коэффициенты уравнений

консультируемой проблемы образуются из вкладов нелинейных, реактивных и постоянных его компонентов, а также величин ± 1 . Величины вкладов в матрицу A системы (8.11) всех нелинейных компонентов изменяются при решении уравнений консультируемой проблемы чаще (они переоцениваются на каждой итерации Ньютона), чем вклады линейных реактивностей (на каждом временном шаге), а вклады постоянных компонентов остаются неизменными до завершения анализа данного варианта решения. Учет характера компонентов консультируемой проблемы, участвующих в образовании элементов матрицы A , дает возможность прогнозировать частоту их изменения и учитывать эту информацию на всех этапах обработки уравнений модели. При этом каждому ненулевому элементу матрицы A соответствует некоторое целое число, характеризующее его тип: 1, 2 — для элементов ± 1 соответственно; 3 — для постоянного элемента; 4 — для элемента, величина которого изменяется во времени; 5 — для нелинейного элемента, зависящего от составляющих вектора переменных модели объекта.

Учитывая распределение НЭ системы уравнений (8.11) по типам, можно представить процесс ее решения с помощью соотношений (8.12)—(8.17) в виде последовательности вложенных циклов так, что элементы высших типов обрабатываются на внутренних циклах. При этом LU -преобразование матрицы уравнений объекта разделяется на несколько этапов в соответствии с представлением выражения (8.11) в виде:

$$\left(A_{cg} + A_T \alpha M + A_x \frac{\partial f}{\partial x} \right) x_{n+1}^{m+1} = b_c + b_T + b_x, \quad (8.19)$$

где g , αM , $\partial f/\partial x$ — векторы мгновенных вкладов в матрицу схемы соответственно постоянных, реактивных и нелинейных компонентов; A_c , A_T , A_x — матрицы связи векторов мгновенных вкладов с элементами матрицы уравнений консультируемой проблемы; b_c , b_T , b_x — составляющие правой части уравнения, соответствующие постоянным, реактивным и нелинейным компонентам.

Вначале выполняется однократное для данного варианта решения исключение постоянных элементов матрицы A ($LU = \text{const}$), затем многократно повторяемое исключение изменяющихся элементов ($LU = \text{var}$), которое включает два подэтапа $LU-T$ и $LU-X$, соответствующих исключению элементов типов 4 и 5. Этап $LU-T$ выполняется на каждом временном шаге только один раз, а $LU-X$ повторяется после каждой переоценки нелинейных компонентов консультируемой проблемы. По аналогии прямой ход решения системы уравнений (8.16) разбивается с учетом типов элементов

вектора b на этапы, соответствующие учету постоянных ($ПХ — С$), зависящих от времени ($ПХ — Т$) и нелинейных ($ПХ — X$) элементов вектора b , а обратный ход (8.17) выполняется, как правило, полностью на каждой итерации решения системы (8.6) или (8.11). Учет частоты изменения элементов математических моделей исследуемых консультируемых проблем оказывает влияние на используемую стратегию упорядочения, так как становится выгодным минимизировать при упорядочении количество НЭ и ИНЭ высших типов. Поэтому в процессе упорядочения строкам и столбцам матрицы уравнений консультируемой проблемы, среди которых ищется главный элемент, также присваиваются типы (они определяются максимальными типами элементов, находящихся в этих строках и столбцах). Тогда среди нескольких элементов матрицы, удовлетворяющих критерию (8.18) и проверке на малость по абсолютной величине, в качестве главного выбирается тот элемент, строка и столбец которого имеют минимальный тип.

Повысить скорость процедуры решения разреженных систем уравнений (8.11), кроме вышеописанного разбиения на отдельные этапы в соответствии с частотой изменения НЭ, можно применением метода кодирования формул (МКФ) и метода генерации программ решения (ГПР). Сущность МКФ заключается в представлении процесса LU -преобразования матрицы уравнений консультируемой проблемы и решения системы $LUx=b$ в виде последовательности вычислений, выполняемых по введенным заранее типам формул (определенным аналитическим выражениям). Тип каждой из формул и координаты НЭ из «позиционного» списка NR располагаются последовательно в специальном целочисленном массиве ICF . Тогда для выполнения какого-то этапа решения системы (8.11) необходимо просмотреть соответствующий участок массива ICF и выполнить вычисления по отмеченным в нем типам формул. Поскольку массив ICF всегда обрабатывается последовательно, то хранить его можно во внешней памяти ЭВМ и загружать по частям, непосредственно перед обработкой.

При использовании метода ГПР последовательность операций решения системы (8.11) набирается с помощью специальной программы-кодировщика, не содержащей циклов и ветвлений, непосредственно на языке высокого уровня или в машинных кодах и оформляется как подпрограмма. После трансляции и включения в систему сгенерированная таким образом программа решения может быть многократно использована в процессе анализа консультируемой проблемы.

Однако при использовании МКФ и ГПР увеличиваются затраты основной памяти ЭВМ для хранения массива ICF или программы решения и практически невозможно обеспечить эффективный контроль численной устойчивости процесса решения. Последнее обстоятельство связано с тем, что при анализе характеристик консультируемых проблем значения нелинейных параметров их компонентов изменяются в очень широких пределах. Это может привести к появлению на главной диагонали используемой фиксированной ненулевой структуры уравнений модели близких к нулю элементов, прерыванию процесса решения, проведению частичного переупорядочения уравнений модели и повторению этапа настройки процедуры решения (набор массива ICF или ГПР).

Среди эффективных мер, принимаемых для обеспечения численной устойчивости процесса решения разреженных систем (8.11), следует отметить метод диагональной модификации (МДМ) и выделение в нижнем правом углу матрицы A заполненного субблока, при LU -разложении которого главные элементы выбираются по всему его полю.

Метод диагональной модификации не нарушает однородности процесса решения системы вида (8.11) и не требует проведения перестановок их строк и столбцов. В соответствии с МДМ при появлении на главной диагонали решаемой системы уравнений малого по величине элемента (например, a_{kk}) его модифицируют добавлением некоторой константы g_k и продолжают вычисления с модифицированной матрицей:

$$A' = A + e_k g_k e_k^t \quad (8.20)$$

где e_k — единичный вектор, элементы которого равны нулю, кроме k -го, равного единице.

Для получения истинного решения модифицированная система уравнений должна быть решена столько раз с различными векторами правых частей, сколько диагональных модификаций было выполнено в исходной системе. Например, при однократном использовании МДМ истинное решение определяется выражением

$$x = x' - \left(\frac{x_k}{z_k - \frac{1}{g_k}} \right) z, \quad (8.21)$$

где x' , z — решения систем уравнений $A'x' = b$ и $A'z = e_k$ соответственно, а x'_k , z_k — их k -е составляющие.

Дополнительные затраты на получение истинного решения зависят от того, в каких строках выполнялись диагональные модификации.

Точность решения линейной системы (8.11) зависит от степени обусловленности матрицы A уравнений модели консультируемой проблемы, под которой понимают величину

$$\text{cond}(A) = \|A\| \|A^{-1}\| = \frac{|\lambda_{\max}|}{|\lambda_{\min}|}, \quad (8.22)$$

где λ_{\max}^2 и λ_{\min}^2 — соответственно максимальное и минимальное по величине собственные значения эрмитовой матрицы $A^t A$.

Например, если вектор правой части системы (8.11) задан с погрешностью δb , то погрешность при нахождении вектора переменных x определяется выражением

$$\frac{\|\delta x\|}{\|x\|} \leq \text{cond}(A) \cdot \frac{\|\delta b\|}{\|b\|}, \quad (8.23)$$

где $\|\cdot\|$ означает норму вектора, в частности, евклидову

$$\|x\| = \sqrt{x^t x} = \sqrt{|x_1|^2 + |x_2|^2 + \dots + |x_n|^2}.$$

Если с погрешностью задана матрица уравнений $(A+\delta A)$, то вызываемая погрешность при определении вектора переменных - оценивается в соответствии с выражением

$$\frac{\|\delta x\|}{\|x + \delta x\|} \leq \text{cond}(A) \frac{\|\delta A\|}{\|A\|}. \quad (8.24)$$

При решении систем уравнений, характеризующихся большим разбросом собственных значений λ_i , могут возникать существенные вычислительные затруднения.

Пример 2. Рассмотрим характер вычислительных затруднений на простой системе линейных уравнений второго порядка

$$\begin{bmatrix} 1 & 2 \\ 1 & 1,999 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 4,998 \end{bmatrix},$$

для которой

$$|\lambda_{\max}| = \sqrt{9,996} \text{ и } |\lambda_{\min}| = \sqrt{10^{-7}}$$

и степень обусловленности $(A) = 10^4$.

Если вектор правой части изменить на величину $\delta b = [0,0,001]^t$, т. е. выбрать $\|\delta b\|/\|b\| = 1,4 \cdot 10^{-4}$, то в соответствии с выражением (8.23) можно оценить относительную погрешность

$$\frac{\|\delta x\|}{\|x\|} \leq \text{cond}(A) \frac{\|\delta b\|}{\|b\|} = 1,4.$$

В реалистичности этой оценки можно убедиться, если решить непосредственно модифицированную систему уравнений

$$\begin{bmatrix} 1 & 2 \\ 1 & 1,999 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 4,999 \end{bmatrix},$$

для которой точное решение $x=[3, 2]^t$ в отличие от точного решения исходной системы $x_0 = [1, 2]^t$.

Таким образом, в действительности $\delta x = [2, -1]^t$, $\|\delta x\|/\|x\| = 1$. Определитель решаемой системы уравнений мал (10^{-3}), но это не означает, что степень обусловленности матрицы всегда плохая. Можно привести пример матрицы 10-го порядка, для которой $|\lambda_1|=|\lambda_2|=...=|\lambda_9|=0,1$ и $\lambda_{10}=1$, а определитель равен $\pm 10^{-9}$, но $\text{cond}(A) = 10$.

Консультанту, моделирующему консультируемую проблему на САК, следует учитывать изложенное выше и стремиться по возможности к снижению величины $\text{cond}(A)$ то ли правильным выбором модели из имеющихся вариантов, существенно отличающихся этой величиной, то ли использованием процедуры, аналогичной методу диагональной модификации, то ли согласованным выбором параметров компонентов и округлением их величин.

При частотном анализе линейной консультируемой проблемы уравнение (8.11) с учетом операторных компонентных выражений преобразуется к виду

$$(\text{Re}A + j\text{Im}A)(\text{Re}x + j\text{Im}x) = \text{Re}b + j\text{Im}b,$$

в результате чего решается система уравнений удвоенной размерности

$$\begin{bmatrix} \text{Re} A & -\text{Im} A \\ \text{Im} A & \text{Re} A \end{bmatrix} \begin{bmatrix} \text{Re} x \\ \text{Im} x \end{bmatrix} = \begin{bmatrix} \text{Re} b \\ \text{Im} b \end{bmatrix}. \quad (8.25)$$

При этом, если на входе консультируемой проблемы синусоидальный источник единичной амплитуды и нулевой фазы, то значения действительной и мнимой составляющих передаточной характеристики T на частоте ω_k получаем как соответственно действительную и мнимую части реакции схемы x_k :

$$T(j\omega_k) = \text{Re} x_k + j \text{Im} x_k,$$

откуда легко определяются значения амплитудно-частотной (АЧХ) и фазочастотной (ФЧХ) характеристик консультируемой проблемы соответственно:

$$T(\omega) = |T(j\omega)| = \sqrt{(\text{Re} x_k)^2 + (\text{Im} x_k)^2}; \quad (8.26)$$

$$\varphi(\omega) = \text{arctg} \frac{\text{Im} x_k}{\text{Re} x_k}. \quad (8.27)$$

Таким образом, решив уравнения консультируемой проблемы (8.25) на частоте $\omega=\omega_k$, определим значение на этой частоте всех интересующих консультанта передаточных функций от одного входа к произвольному числу выходов. Если необходимо найти передаточные функции от разных входов, то для каждого из входов необходимо

повторить прямой и обратный ход решения уравнений консультируемой проблемы (8.26) с соответствующей правой частью.

Кусочно-линейные методы. Количество последовательных решений линейной системы уравнений (8.6), необходимое для решения нелинейной системы (8.2) или (8.3), можно снизить, если применить кусочно-линейные методы и изменить характер используемой линеаризации компонентных уравнений (6.41).

В результате кусочно-линейной аппроксимации компонентных уравнений с сравнительно небольшим числом участков аппроксимации уравнения нелинейной модели консультируемой проблемы $f(x) = y$, где y — вектор задающих сил, преобразуются в совокупность линеаризованных алгебраических уравнений вида

$$y = f(x) = J^{(m)}x + B^{(m)}, \quad m = 1, 2, \dots, l, \quad (8.28)$$

где l — число выпуклых областей разбиения пространства переменных модели консультируемой проблемы, границы которых в виде гиперплоскостей определяются точками излома аппроксимированных характеристик отдельных нелинейных компонентов.

Вследствие непрерывности оператора преобразования непрерывного пространства в линейно-секционированное предполагается, что матрицы $J^{(m)}$ и $J^{(m-1)}$ отличаются только одним элементом, поэтому

$$J^{(m)} - J^{(m-1)} = cr^t, \quad (8.29)$$

где $c = x(J^{(m)} - J^{(m-1)})$; r — вектор, нормальный к границе областей.

Применяя соотношение (8.29), используют метод приспособывания для решения линейных систем уравнений двух смежных областей, в соответствии с которым треугольные матрицы $L^{(m)}$ и $U^{(m)}$, определяемые матрицей $J^{(m)} = L^{(m)}U^{(m)}$, находят непосредственно по известным матрицам $L^{(m-1)}$ и $U^{(m-1)}$, соответствующим предыдущей области решения с матрицей $J^{(m-1)} = L^{(m-1)}U^{(m-1)}$.

Задача нахождения x^* , удовлетворяющего уравнению (8.28) при заданном y^* , при использовании кусочно-линейных методов формулируется как задача проведения траектории движения изображающей точки в секционированном пространстве, начиная с $x^{(0)}$. Причем в области значений функции f (подпространстве y) траектория $L(y)$ имеет вид прямой линии, соединяющей точки $y^{(0)}$ и y^* , а в области значений аргументов (подпространство x) эта траектория имеет форму кусочно-линейной кривой $L(x)$, точки излома которой находятся на граничных гиперплоскостях. При выборе начального значения $x^{(0)}$ исходят из условий обеспечения сходимости решения, для чего в

качестве начальной выбирают область, в которой принципиально может находиться только одно решение, и вычисляют

$$y^{(0)} = f(x^{(0)}) = J^{(0)} x^{(0)} + B^{(0)}.$$

Далее строят часть кривой решения $L(x)$, расположенной в начальной области R_0

$$x^{(0)}(\lambda) = x^{(0)} + \lambda [J^{(0)}]^{-1} (y^* - y^{(0)}) \quad (8.30)$$

по двум точкам

$$x^{(0)} \text{ и } x^{(0)}(1) = x^{(0)} + [J^{(0)}]^{-1} (y^* - y^{(0)}),$$

при этом кривая решения $L(y)$ имеет вид

$$y^{(0)}(\lambda) = \lambda y^* - (1 - \lambda) y^{(0)}. \quad (8.31)$$

В выражениях (8.27) и (8.28) коэффициент λ изменяется в диапазоне $0 \leq \lambda \leq \lambda^{(0)}$, где значение $\lambda^{(0)} < 1$ определяется из условия, что $x^{(0)}[\lambda^{(0)}]$ лежит на граничной гиперплоскости между областями $R^{(0)}$ и $R^{(1)}$ и не попадает в точку пересечения нескольких областей, т. е. из условий

$$\left. \begin{aligned} x^{(1)} &= x^{(0)}[\lambda^{(0)}]; \\ y^{(1)} &= J^{(1)} x^{(1)} + B^{(1)}. \end{aligned} \right\}$$

Далее аналогичным образом строятся части «кривых решения» $L(x)$ и $L(y)$ в области $R^{(1)}$:

$$\begin{aligned} x^{(1)}(\lambda) &= x^{(1)} + \lambda [J^{(1)}]^{-1} (y^* - y^{(1)}); \\ y^{(1)}(\lambda) &= \lambda y^* - (1 - \lambda) y^{(1)}, \end{aligned} \quad (8.32)$$

и также оценивается величина λ , для этой области и т. д.

Процесс вычислений (перебора областей) продолжается до тех пор, пока λ в одной из них, например в s -й области, не достигнет значений $\lambda = 1$. Тогда

$$\begin{aligned} x^* &= x^{(s)}(1) = x^{(s)} + [J^{(s)}]^{-1} (y^* - y^{(s)}); \\ y^{(s)}(1) &= y^*. \end{aligned}$$

Если матрицы $J^{(m)}$ и $J^{(m-1)}$ для соседних областей имеют одинаковые знаки, то при прохождении границы областей коэффициент $\lambda > 0$, что быстро ведет к сходимости вычислений. Если указанные матрицы имеют различные знаки, то граница областей пересекается при $\lambda < 0$ и решение $y^{(m)}$ уходит от y^* , что в принципе может привести к образованию циклов в траектории движения $L(x)$.

Следует подчеркнуть, что ускорение процедуры решения нелинейной алгебраической системы уравнений применением кусочно-линейных методов достигается за счет точности самого решения.

Выбор метода решения системы алгебраических уравнений.

Решение систем алгебраических уравнений (АУ) имеет место во

многих консультационных процедурах и прежде всего в процедурах функционального консультирования. Эффективность решения этих консультационных задач вносит существенный вклад в общую эффективность выполнения консультационных процедур, поэтому необходимо правильно выбрать метод решения системы АУ. Такой выбор приходится осуществлять разработчику пакета прикладных программ (ППП) для подсистем функционального консультирования. Если же пакет выполнен открытым по отношению к численным методам решения систем АУ и, следовательно, содержит ряд модулей, реализующих альтернативные методы, то выбор метода возлагается на консультанта.

На эффективность применения метода оказывают влияние не только особенности самого метода, но и в не меньшей мере особенности решаемой задачи и используемой ЭВМ. Среди наиболее существенных особенностей задач, называемых ниже факторами, отметим размерность n (порядок системы уравнений), число обусловленности C и разреженность S матрицы Якоби, а среди особенностей ЭВМ — быстродействие B , определенное для класса консультационных задач, емкость оперативной памяти и разрядность машинного слова. Разработчик ППП должен ориентироваться на некоторые диапазоны значений этих факторов, характерные для моделей консультируемых проблем в соответствующей предметной области. Эти диапазоны должны быть либо указаны в техническом задании на разработку ППП, либо спрогнозированы самим разработчиком на основе исследования статистических данных об имеющихся или предполагаемых моделях.

Собственно выбор метода основан на установлении взаимосвязи между факторами и характеристиками используемых ЭВМ, с одной стороны, и такими показателями эффективности применения метода, как время решения, вероятность получения правильного результата и его точность, с другой стороны.

Рассмотрим, как устанавливаются подобные взаимосвязи для основных методов решения АУ. Для этих методов

$$T_m = n\gamma I/B,$$

где T_m — затраты машинного времени; n — порядок решаемой системы АУ, принимаемый за оценку сложности задачи; γ — среднее число арифметических операций, приходящихся на единицу сложности задачи, на одной итерации; I — среднее число итераций.

Для решения системы линейных алгебраических уравнений (ЛАУ) вида $\mathbf{AV}=\mathbf{B}$ выбирают *либо метод Гаусса, либо итерационные методы.*

Для метода Гаусса $I=1$, и если не учитывать разреженность матрицы коэффициентов A , то $\gamma \approx 2(n^2/3 + 2n)$. Неучет разреженности ограничивает целесообразность применения метода Гаусса решением задач только невысокой размерности. При $n > 50$ учет разреженности становится необходимым. Для метода Гаусса при учете разреженности и оптимальном упорядочении строк и столбцов матрицы A в задачах переключаемых электронных схем имеем $\gamma = \text{const}$. Так, для моделей переключаемых электронных схем $\gamma \approx 25$, а для распределенных моделей с трехдиагональной матрицей коэффициентов при применении метода прогонки $\gamma \approx 8$.

Для решения систем ЛАУ итерационными методами с учетом разреженности матрицы коэффициентов имеем $I > 1$, а $\gamma = 2Qn$, где

$Q = 1 - S$ — насыщенность матрицы. Так как $Q = K/n$, где K — среднее арифметическое для числа ненулевых элементов в одной строке матрицы A , то $\gamma = 2K$. Так, для моделей переключаемых электронных схем получаем по результатам статистических исследований $\gamma \approx 7,8$, т. е. одна итерация выполняется быстрее, чем по методу Гаусса. Однако из-за того, что $I > 1$, итерационные методы по показателю T_m практически всегда проигрывают методу Гаусса.

Решение систем нелинейных АУ выполняется итерационными методами, при этом на требуемое число итераций I в методе Ньютона решающее влияние оказывает выбор начального приближения, а в остальных итерационных методах — число обусловленности C матрицы Якоби решаемой системы уравнений.

В методе Ньютона, применяемом в рамках методов установления или продолжения решения по параметру, обычно I не превышает трех. В случаях, если I превышает некоторый порог $I_{пр}$ (например, $I_{пр} = 7$), лучше уменьшать значения коэффициентов, управляющих процессом установления, чем продолжать итерации при $I > I_{пр}$. Следует отметить, что при решении нелинейных АУ величина γ растет, так как при ее подсчете должны быть учтены затраты на вычисление элементов матрицы Якоби.

В методе простых итераций I может достигать неприемлемо больших значений, поэтому целесообразно ввести на I ограничение $I_{гп}$ сверху. Если принять $I_{гп} = 1,5 \cdot 10^4$, то из соотношения $I_{гп} = -0,5 C \lg \varepsilon$ при $\varepsilon = 10^{-3}$ получаем, что метод простых итераций можно применять только к решению системы уравнений, у которых матрица Якоби имеет $C < 10^4$. Методы Зейделя, Якоби, последовательной верхней релаксации (ПВР) имеют аналогичный характер зависимости I от C , хотя скорость сходимости у них часто оказывается несколько выше, чем в методе простых итераций.

Экономичность метода решения систем АУ определяется также затратами оперативной памяти. При неучете разреженности только на хранение матрицы Якоби нужно n^2 ячеек памяти. Поэтому если для хранения одного слова используется 8 байт, то при $n=100$ для хранения требуется 80 кбайт, а при $n=500$ — уже 2 Мбайт. Итак, подтверждается вывод о необходимости учета разреженности при решении задач с $n > n_{np}$, где n_{np} зависит от характеристик используемой ЭВМ и, как правило, составляет несколько десятков. В задачах анализа распределенных моделей, в которых n может превышать 10^4 , экономичность метода по затратам машинной памяти становится одной из важнейших характеристик. В таких случаях применяют либо релаксационные методы, либо метод Ньютона с использованием на каждой итерации метода Гаусса, но в рамках рассматриваемого ниже диакоптического подхода.

На точность решения задачи оказывают влияние задаваемые консультантом в исходных данных значения допустимых погрешностей ε_1 или ε_2 , а также обусловленность модели. Однако задаваемые значения ε_1 или ε_2 могут вообще оказаться недостижимыми или из-за несходимости, или из-за слишком медленной сходимости вычислительного процесса. Поэтому если создаваемый ППП ориентирован на решение систем уравнений с широким диапазоном значений C , то нужно принимать специальные меры по обеспечению точности решения. При реализации метода Гаусса нужно перейти к представлению чисел в ЭВМ с повышенной разрядностью (например, с удвоенной разрядностью), в случае метода простых итераций — к уменьшению шага h .

Надежность применения метода определяется не только фактом принципиальной сходимости к корню, но и тем, каковы затраты времени T_m на получение решения с требуемой точностью. Ненадежность итерационных методов проявляется либо при неудачном выборе начального приближения к корню (метод Ньютона), либо при плохой обусловленности задачи (методы релаксационные и простых итераций), либо при повышенных требованиях к точности решения (метод простых итераций), либо при высокой размерности задач (метод Гаусса при неучете разреженности). Поэтому при создании узкоспециализированных программ необходимы предварительный анализ особенностей ММ заданного класса задач (значений n , C , допустимых погрешностей) и соответствующий выбор конкретного метода. При создании ППП с широким спектром решаемых задач необходима реализация средств автоматической адаптации метода решения к конкретным условиям. Такая адаптация в современных

ППП чаще всего применяется в рамках методов установления или продолжения решения по параметру.

8.3. Метод анализа переходных процессов

Классификация методов численного интегрирования систем обыкновенных дифференциальных уравнений (ОДУ).

Методы численного интегрирования ОДУ являются методами преобразования дифференциальных уравнений в алгебраические. После дискретизации независимой переменной t система ОДУ

$$\Psi(\dot{\mathbf{V}}, \mathbf{V}, t) = 0$$

в каждой точке t_k этой переменной представляется в виде системы n алгебраических уравнений

$$\Psi(\dot{\mathbf{V}}_k, \mathbf{V}_k, t_k) = 0 \tag{8.33}$$

с $2n$ неизвестными $\dot{\mathbf{V}}_k = \dot{\mathbf{V}}(t_k)$ и $\mathbf{V}_k = \mathbf{V}(t_k)$. Система (8.33) доопределяется уравнениями

$$\dot{\mathbf{V}}_k = \mathbf{f}(\mathbf{V}_k), \tag{8.34}$$

задаваемыми выбранным методом численного интегрирования. Система алгебраических уравнений (8.33), (8.34) решается в каждой точке t_k , $k=1, 2, \dots, III$, где III — число точек дискретизации (шагов интегрирования).

Формулу численного интегрирования (8.34), в которой в качестве неизвестных величин фигурируют $\dot{\mathbf{V}}_k$ и \mathbf{V}_k , и соответствующие этой формуле методы интегрирования называют *неявными*. В неявных формулах кроме $\dot{\mathbf{V}}_k$ и \mathbf{V}_k могут присутствовать значения переменных $\dot{\mathbf{V}}$ и (или) \mathbf{V} в p предыдущих точках дискретизации t_{k-i} , $i=1, 2, \dots, p$. При $p \geq 2$ метод интегрирования называют *многошаговым*. Следует отметить, что к моменту решения системы (8.34), (8.35) значения $\dot{\mathbf{V}}_{k-i}$ и \mathbf{V}_{k-i} для $i \geq 1$, фигурирующие в (8.35), уже вычислены на предыдущих шагах. Название метода «многошаговый» происходит из-за использования в формуле интегрирования результатов нескольких предыдущих шагов. Величину p при этом называют *порядком многошагового метода*. Вместо $\dot{\mathbf{V}}_{k-i}$ или \mathbf{V}_{k-i} , $i \geq 2$, в формуле интегрирования могут присутствовать производные \mathbf{V} по t порядка выше первого или заменяющие их результаты некоторых дополнительных вычислений на данном шаге. В этом случае метод называется *одношаговым*, а порядок одношагового метода совпадает с порядком старшей из используемых производных.

Систему алгебраических уравнений, решаемых на каждом шаге численного интегрирования, можно записать также в следующем виде:

$$\Psi(\dot{\mathbf{V}}_{k-1}, \mathbf{V}_{k-1}, t_{k-1}) = 0, \quad (8.35)$$

$$\dot{\mathbf{V}}_{k-1} = \mathbf{f}(\mathbf{V}_k), \quad (8.36)$$

где $\dot{\mathbf{V}}_{k-1}$ и \mathbf{V}_k — неизвестные величины; \mathbf{V}_{k-1} вычислены на предыдущем шаге.

Формулу численного интегрирования (8.36) и соответствующие ей методы интегрирования называют *явными*. Явные методы по аналогии с неявными могут быть одно- и многошаговыми, аналогично определяются порядки явных методов.

Очевидно, что необязательно на каждом шаге интегрирования численно решать систему из $2n$ конечных уравнений. В большинстве случаев выполняют предварительное исключение неизвестного вектора $\dot{\mathbf{V}}_k$ из (8.33) или $\dot{\mathbf{V}}_{k-1}$ из (8.35) с помощью формул интегрирования (8.34) или (8.36) в общем виде и на каждом шаге численно решают систему n уравнений с неизвестным вектором \mathbf{V}_k .

Методы численного интегрирования ОДУ, применяемые в САК.

В практике машинных вычислений наиболее распространены для решения ОДУ *методы Гира, Адамса и Рунге — Кутты*.

Общий вид формул интегрирования в неявных методах Гира

$$\dot{\mathbf{V}}_k = \left(\sum_{i=0}^p a_i \mathbf{V}_{k-i} \right) / h_k,$$

где $h_k = t_k - t_{k-1}$ — величина k -го шага интегрирования; a_i — коэффициенты, значения которых зависят от порядка p метода и величин порядков последних шагов. Формулы Гира называют также *формулами дифференцирования назад* (ФДН) по той причине, что в них аппроксимация производных в точке t_k производится с помощью значений функций, относящихся к данному и предыдущим моментам времени. Формула Гира при $p=1$ совпадает с *неявной формулой Эйлера*

$$\dot{\mathbf{V}}_k = (\mathbf{V}_k - \mathbf{V}_{k-1}) / h_k.$$

Часто применяют формулу Гира второго порядка, называемую также *формулой Шихмана*, которая при $h = \text{const}$ имеет вид

$$\dot{\mathbf{V}}_k = \left(\frac{3}{2} \mathbf{V}_k - 2\mathbf{V}_{k-1} + \frac{1}{2} \mathbf{V}_{k-2} \right) / h.$$

Общий вид формул интегрирования в *явных методах Адамса* при $p \geq 2$

$$\dot{\mathbf{V}}_{k-1} = a_1 (\mathbf{V}_k - \mathbf{V}_{k-1})/h_k + \sum_{i=2}^p a_i \dot{\mathbf{V}}_{k-i};$$

в неявных методах Адамса при $p \geq 2$

$$\dot{\mathbf{V}}_k = a_0 (\mathbf{V}_k - \mathbf{V}_{k-1})/h_k + \sum_{i=1}^{p-1} a_i \dot{\mathbf{V}}_{k-i}.$$

Явная формула Адамса при $p=1$ называется также *явной формулой Эйлера*:

$$\dot{\mathbf{V}}_{k-1} = (\mathbf{V}_k - \mathbf{V}_{k-1})/h_k;$$

явная формула Адамса при $p = 2$

$$\dot{\mathbf{V}}_{k-1} = \frac{2}{3} (\mathbf{V}_k - \mathbf{V}_{k-1})/h_k + \dot{\mathbf{V}}_{k-2}/3.$$

Неявный метод Адамса второго порядка точности называют также *методом трапеций*, ему соответствует формула интегрирования

$$\dot{\mathbf{V}}_k = 2 (\mathbf{V}_k - \mathbf{V}_{k-1})/h_k - \dot{\mathbf{V}}_{k-1}. \quad (8.37)$$

Рассмотренные методы при $p \geq 2$ являются многошаговыми. К одношаговым методам относится метод Рунге — Кутты.

В САК в основном используются неявные методы трапеций и Гира, а в отдельных случаях применяют явный метод Эйлера.

Использование методов возможно, если порождаемый ими вычислительный процесс является устойчивым. Неустойчивость вычислений может возникнуть в связи с катастрофическим ростом погрешностей. Различают локальную погрешность интегрирования, допущенную на данном шаге интегрирования, и погрешность, накопленную к моменту t_k за все предыдущие шаги. В неустойчивых методах погрешность решения увеличивается от шага к шагу, что приводит к полному искажению результатов и, возможно, к переполнению разрядной сетки.

Среди рассмотренных методов интегрирования имеются А-устойчивые и ограниченно устойчивые методы. А-устойчивым называют метод, при применении которого к интегрированию системы линейных ОДУ

$$\dot{\mathbf{V}} = \mathbf{A}\mathbf{V} \quad (8.38)$$

с начальными условиями $\mathbf{V}_0 \neq 0$ погрешность решения стремится к нулю при любом значении постоянного шага $h > 0$ и при $t_k \rightarrow \infty$. Здесь \mathbf{A} — постоянная матрица, собственные значения которой имеют отрицательные действительные части. К А-устойчивым относятся неявные методы Гира и Адамса первого и второго порядков точности.

Ограниченно устойчивыми являются остальные из рассмотренных методов, для них характерно сохранение устойчивости вычислений только при выполнении ограничений, накладываемых на значение шага интегрирования. Так, для явного метода Эйлера при $h = \text{const}$ в задаче (8.38) условие устойчивости имеет вид неравенства

$$|1 + h\lambda_j| < 1, \quad (8.39)$$

которое должно выполняться для всех собственных значений λ_j матрицы \mathbf{A} . Для матрицы \mathbf{A} с отрицательными вещественными собственными значениями λ_j условие (8.39) можно представить в виде

$$0 < h < -2/\lambda_j.$$

Если при этом система уравнений (8.33) есть модель динамической консультируемой проблемы (например, электронной схемы), то величины $-1/\lambda_j$ принято называть постоянными времени τ_j . Тогда условие устойчивости явного метода Эйлера приводится к виду

$$0 < h < 2\tau_{\min}, \quad (8.40)$$

где τ_{\min} — минимальная постоянная среди постоянных времени моделируемой консультируемой проблемы.

Условия (8.39) или (8.40) устойчивости методов интегрирования в применении к нелинейным системам ОДУ можно рассматривать как приближенные, при этом под λ_j понимают собственные значения матрицы Якоби $\mathbf{Y} = \partial \dot{\mathbf{V}} / \partial \mathbf{V}$.

Так как в нелинейных задачах элементы матрицы Якоби непостоянны, то непостоянны и ее собственные значения. Поэтому априорный выбор значения постоянного шага h , удовлетворяющего условиям устойчивости на всем интервале интегрирования $[0, T_{\text{кон}}]$, оказывается практически невозможным (случай гарантированного выполнения условий устойчивости за счет выбора $h = \tau_{\min}$ неприемлем, так как приводит к чрезмерным затратам машинного времени).

Интегрирование с постоянным шагом нецелесообразно и в А-устойчивых методах, так как h влияет на точность и время решения. Влияние h на точность решения по-разному проявляется на различных участках моделируемого переходного процесса. Поэтому минимизация затрат машинного времени при соблюдении точностных ограничений возможна только в условиях интегрирования с переменным шагом.

Большинство алгоритмов автоматического выбора шага основано на контроле локальных погрешностей интегрирования. Локальные погрешности включают в себя погрешности методические, обуславливаемые приближенностью формул интегрирования, и

округления, обуславливаемые представлением чисел с помощью ограниченного количества разрядов. Локальная методическая погрешность многошагового метода порядка p , допущенная на k -м шаге интегрирования, зависит от значения шага h_k и оценивается по формуле

$$\varepsilon_k = c \| \mathbf{V}^{(p+1)}(\tau) \| h_k^{p+1}, \quad (8.41)$$

где c — коэффициент, зависящий от характера и порядка метода; $\| \mathbf{V}^{(p+1)}(\tau) \|$ — норма $(p+1)$ -х производных \mathbf{V} по t , вычисленных в точке $\tau \in [t_{k-p}, t_k]$. В алгоритмах интегрирования задаются значением допустимой локальной погрешности δ , а значение шага h_k ограничивают сверху

$$\varepsilon_k < \delta, \quad (8.42)$$

при этом вектор $(p+1)$ -х производных приближенно оценивают с помощью вектора $(p+1)$ -х конечных разностей, вычисленных по уже найденным значениям \mathbf{V}_{k-i} , $i=1, 2, \dots, (p+2)$. Однако выполнение условия (8.42) с большим запасом приводит к заниженным значениям h_k и увеличивает T_m . Поэтому наряду с (8.42) вводят также ограничение на допущенную погрешность ε_k снизу и значение шага выбирают автоматически, например по следующему правилу:

$$h_{k+1} = \begin{cases} m_1 h_k & \text{при } \varepsilon_k > \delta_2; \\ h_k & \text{при } \delta_2 > \varepsilon_k > \delta_1; \\ m_2 h_k & \text{при } \varepsilon_k < \delta_1, \end{cases}$$

где $m_1 < 1$; $m_2 > 1$; $\delta_2/\delta_1 = 2$. (Значения m_1 , m_2 , δ_2 выбирают, исходя из имеющегося опыта вычислений по данному алгоритму.)

Для того чтобы сделать несущественным влияние погрешностей округления, предусматривают в соответствующих частях алгоритма вычисления с удвоенным числом разрядов, а также не допускают снижения значения шага ниже некоторого минимального уровня h_{\min} .

Сравнение методов и обоснование их выбора для конкретных задач автоматизированного консультирования.

Эффективность метода численного интегрирования оценивается его влиянием на экономичность и точность вычислений.

Качественно характер влияния значения шага на погрешность интегрирования для различных методов можно представить с помощью графиков, изображенных на рис. 8.1.

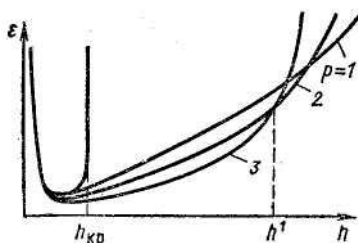


Рис. 8.1. Зависимость погрешности интегрирования от величины шага и порядка метода

В области малых $h < h_{\min}$ значительны погрешности округления. В явных методах наблюдается резкий рост погрешностей при $h > h_{кр}$, где $h_{кр}$ — максимально допустимая по условиям устойчивости величина шага интегрирования. С ростом порядка метода снижаются погрешности интегрирования, но только при умеренных значениях $h < h'$. Поскольку величина h' зависит не только от порядков сравниваемых методов, но и от особенностей переходного процесса, что подтверждается формулой (8.41), то заранее определить, метод какого порядка точности даст оптимальный компромисс между показателями точности и экономичности, затруднительно. В связи с этим в программах анализа ряда консультируемых проблем (например, электронных схем) распространен метод ФДН, основанный на автоматическом выборе не только значения шага, но и порядка p неявной формулы Гира. Если на предшествующем $(k-1)$ -м шаге использовалась формула порядка p_{k-1} , то на следующем шаге выбор производился между формулами порядков $p_{k-1}-1$, p_{k-1} или $p_{k-1}+1$. Выбирается та формула, использование которой при заданной погрешности обеспечивает наибольшую величину шага h_k . При этом порядок p_k формул Гира не должен выходить за пределы диапазона [1,6].

Среди неявных методов интегрирования при $p = \text{const}$ применяют методы Эйлера, трапеций, Шихмана. Их положительными особенностями являются А-устойчивость и сравнительно малый объем памяти, требующийся для хранения результатов интегрирования, полученных на предыдущих шагах. Однако метод Эйлера не обеспечивает необходимой точности при анализе переходных процессов в слабодемпфированных системах. Метод трапеций в его первоначальном виде (8.37) имеет недостаток, заключающийся в появлении в численном решении ложной колебательной составляющей уже при сравнительно умеренных значениях шагов, поэтому метод

трапеций удобен только при принятии мер, устраняющих ложные колебания. Значительное уменьшение ложных колебаний, но при несколько больших погрешностях, дает формула Шихмана.

В целом затраты машинного времени на анализ переходных процессов неявными методами существенно зависят от экономичности алгоритмов численного решения конечных уравнений, применяемых на каждом шаге интегрирования. Обычно для решения конечных уравнений используют метод Ньютона, тогда

$$T_m \approx bN^\alpha ИШ, \quad (8.43)$$

где b — коэффициент пропорциональности, зависящий в основном от быстродействия используемой ЭВМ; N — показатель сложности анализируемого объекта; $И, Ш$ — число ньютоновских итераций на одном шаге и шагов интегрирования; α [1, 3] и зависит от свойств выбранного алгоритма решения системы линейных алгебраических уравнений (ЛАУ) на каждой ньютоновской итерации. Если разреженность матрицы Якоби не учитывается, то $\alpha=3$ и возможности применения неявных методов ограничиваются задачами сравнительно малой размерности. Поэтому в САК сложных консультируемых проблем (таких, как БИС) необходим учет разреженности матриц. При этом α в (8.43) оказывается в интервале [1, 2] и существенно повышает эффективность неявных методов. Величины $И$ и $Ш$ зависят от особенностей ММ консультируемой проблемы и характера анализируемых процессов.

Явные методы интегрирования целесообразно применять к решению систем ОДУ, представленных в нормальной форме Коши

$$\dot{\mathbf{V}} = \Phi(\mathbf{V}, t). \quad (8.44)$$

Тогда отпадает необходимость решения систем конечных уравнений на каждом шаге. Например, подставляя формулу Эйлера

$$\dot{\mathbf{V}}_{k-1} = (\mathbf{V}_k - \mathbf{V}_{k-1})/h_k \text{ в } \dot{\mathbf{V}}_{k-1} = \Phi(\mathbf{V}_{k-1}, t_{k-1}),$$

получаем явное относительно искомого вектора \mathbf{V}_k выражение

$$\mathbf{V}_k = \mathbf{V}_{k-1} + h_k \Phi(\mathbf{V}_{k-1}, t_{k-1}) \quad (8.45)$$

(явность этого выражения и послужила причиной названия явных методов интегрирования). Вычисления по (8.45) фактически сводятся к расчету правых частей системы ОДУ (8.44) и соответствуют значению α , близкому к единице, в формуле (8.43). Малый объем вычислений на одном шаге и малый объем требующейся оперативной памяти — положительная черта явных методов. Однако в них могут стать недопустимо большими значения $Ш$. В самом деле, $Ш = T_{\text{кон}}/h_{\text{ср}}$, где $T_{\text{кон}}$ — отрезок интегрирования, соизмерим с максимально допустимой

постоянной времени τ_{\max} ; h_{cp} — средняя величина шага, из-за соображений устойчивости соизмерима с τ_{\min} . Теперь (8.43) принимает вид

$$T_M \approx bN\tau_{\max}/\tau_{\min}.$$

Очевидно, что явные методы применяются к анализу переходных процессов только в консультируемых проблемах с умеренными значениями $Ц = \tau_{\max}/\tau_{\min}$ (обычно при $Ц$ не более 10^4 — 10^5). В то же время в реальных задачах анализа фрагментов сложных консультируемых проблемах значения $Ц$ могут оказаться значительно больше.

Таким образом, сравнение явных и неявных методов интегрирования ОДУ свидетельствует о большей универсальности последних. Поэтому неявные методы являются основными методами анализа переходных процессов в подпроблемах сложных консультируемых проблем. Явные методы могут давать лучшие результаты только в отдельных случаях анализа консультируемых проблем с хорошо обусловленными ММ.

8.4. Выбор метода численного интегрирования

Методы численного интегрирования при анализе динамических свойств консультируемых проблем могут применяться, с одной стороны, непосредственно к уравнениям переменных состояния, сформированным для всей консультируемой проблемы, т. е. к модели консультируемой проблемы, а с другой стороны — для алгебраизации реактивных компонентных уравнений при формировании моделей консультируемой проблемы по уравнениям (8.1) и (8.2).

Существует множество различных методов численного интегрирования систем дифференциальных уравнений первого порядка с начальными условиями вида

$$\frac{dx}{dt} = \varphi(x, t), \quad x_0 = x(0), \quad (8.45)$$

которые различаются между собой формулами вычисления следующего приближения, способами оценки локальной погрешности, стратегией выбора шага (порядка), правилами принятия решения о преэмптентности решения. При моделировании сложных консультируемых проблем в САК (особенно, когда математические модели имеют сильно разнесенные собственные значения) необходимы мощные методы интегрирования или даже различные комбинации отдельных методов, обеспечивающие автоматический выбор и смену методов (например, явных или неявных) в процессе решения исследуемой системы дифференциальных или алгебро-дифференциальных уравнений.

Формулы вычисления следующего приближения. Методы интегрирования можно подразделить на одношаговые (например, Эйлера, Рунге—Кутты и др.) и многошаговые (прогноза — коррекции, Адамса, Гира и др.). Это подразделение определяется количеством информации о предыдущих значениях интегрируемой функции $x_{n-k}, x_{n-k+1}, \dots, x_{n-1}$, используемой для нахождения приближения x_n к точному значению $x(t_n)$ с помощью соответствующих разностных формул.

В зависимости от того, используется ли в соответствующих разностных формулах приближения для x_n производная $x'_{n-1} = \varphi(x_{n-1}, t_{n-1})$ или $x'_n = \varphi(x_n, t_n)$, методы численного интегрирования подразделяются на **явные и неявные**.

Разбивая интервал интегрирования (t_0, T) на отдельные временные участки $h = t_k - t_{k-1}$, $k = 1, 2, \dots, N$ и интегрируя выражение (8.45), получаем для некоторого интервала времени t_{k-q}, t_{k+p}

$$x_{k+p} = x_{k-q} + \int_{t_{k-q}}^{t_{k+p}} \varphi[x(\tau), \tau] d\tau. \quad (8.46)$$

Если заменить $\varphi[x(\tau), \tau]$ некоторым интерполяционным полиномом, принимающим значения $x_n = \varphi(x_n, t_n)$ на множестве точек t_n , то можно получить из соотношения (8.46) различные формулы численного интегрирования.

Применим, например, интерполяционный полином в форме Ньютона, использующий обратные разности высоких порядков,

$$P(t) = \sum_{j=0}^r (-1)^j \binom{-s}{j} \nabla^j \varphi_k, \quad (8.47)$$

где

$$\begin{aligned} \nabla^{(0)} \varphi_k &= \varphi_k; \\ \nabla^{(p)} \varphi_k &= \nabla^{(p-1)} \varphi_k - \nabla^{(p-1)} \varphi_{k-1}; \\ s &= \frac{t - t_k}{h}; \\ \binom{t}{j} &= \frac{t(t-1) \dots (t-j+1)}{j!}. \end{aligned}$$

Подставляя полином (8.47) в выражение (8.46), получаем выражение

$$x_{k+p} = x_{k-q} + h \sum_{j=0}^r \alpha_j \nabla^{(j)} \varphi_k, \quad (8.48)$$

в котором

$$\alpha_i = (-1)^i \int_{-q}^p \binom{-s}{i} ds.$$

Эту же формулу, раскрывая разности высоких порядков, представляем в виде

$$x_{k+p} = x_{k-q} + h \sum_{j=0}^r \beta_j \varphi_{k-j}, \quad (8.49)$$

где

$$\beta_j = (-1)^j \left[\binom{j}{j} \alpha_j + \binom{j+1}{j} \alpha_{j+1} + \dots + \binom{r}{j} \alpha_r \right].$$

Различные конкретные формулы численного интегрирования можно получить на основании выражения (8.49), если выбрать значения для p и q ($p \geq 0$, $q \geq 0$). Например, при $p = 1$ и $q = 0$ реализуется известный экстраполяционный явный метод численного интегрирования Адамса — Бэшфорта:

$$x_{k+1} = x_k + h [\beta_0 \varphi_k + \beta_1 \varphi_{k-1} + \dots + \beta_r \varphi_{k-r}], \quad (8.50)$$

а при $p = 0$ и $q = 1$ реализуется интерполяционный неявный метод Адамса—Мултона в виде

$$x_k = x_{k-1} + h [\beta_0^* \varphi_k(x_k) + \beta_1^* \varphi_{k-1} + \dots + \beta_r^* \varphi_{k-r}], \quad (8.51)$$

который удобнее для последующего сравнения с выражением (8.50) переписать следующим образом:

$$x_{k+1} = x_k + h [\beta_0^* \varphi_{k+1} + \beta_1^* \varphi_k + \dots + \beta_r^* \varphi_{k+1-r}]. \quad (8.52)$$

Нетрудно видеть, что при выборе $r=0$ в формулах (8.50) и (8.52) реализуется соответственно явный и неявный методы Эйлера:

$$\begin{aligned} x_{k+1} &= x_k + h \varphi_k = x_k + h \dot{x}_k \\ x_{k+1} &= x_k + h \varphi_{k+1} = x_k + h \dot{x}_{k+1}. \end{aligned} \quad (8.53)$$

Может показаться, что явный многошаговый метод является самым простым методом с точки зрения вычислений. Однако на практике явные методы используются редко из-за низкой устойчивости вычислений, характеризующей накопление погрешности от шага к шагу. Для заданного числа r неявный метод (8.52) является в принципе более точным, чем явный метод (8.50). Кроме того, в явных методах значение шага h должно быть гораздо меньше, чем в неявных методах (вследствие абсолютной устойчивости последних).

Рассмотрим теперь практический подход к реализации неявных методов. На основании выражений (8.52) и (8.45) можно получить

нелинейное алгебраическое уравнение, которое необходимо решить для нахождения x_{k+1} :

$$\varepsilon = x_{k+1} - h\beta_0^* \Phi(x_{k+1}, t_{k+1}) - \sum_{i=1}^r \beta_i^* \Phi_{k+1-i} = 0. \quad (8.54)$$

Если воспользоваться методом Ньютона, то получим

$$x_{k+1}^{(m+1)} = x_{k+1}^{(m)} - (P_{k+1}^{-1}) \varepsilon(x_{k+1}^{(m)}), \quad (8.55)$$

где

$$P_{k+1} = \left(\frac{\partial \varepsilon}{\partial x} \right)_{x_{k+1}^{(m)}} = 1 - h\beta_0^* \left(\frac{\partial \Phi}{\partial x} \right)_{x_{k+1}^{(m)}} = 1 - h\beta_0^* J_{k+1}^m, \quad (8.56)$$

где J_{k+1}^m — матрица Якоби для решаемой системы уравнений (8.45). В качестве начального приближения $x_{k+1}^{(0)}$ удобно выбирать значение вектора переменных, предсказанное с помощью формулы явного метода (8.50). Тогда весь комбинированный процесс использования формул (8.50) и (8.52) называют методом **предсказания коррекции**, в котором явный метод вычислений по формуле (8.50) называют **предсказывающим**, а неявный метод вычислений по формуле (8.52) — **корректирующим**. Очевидно, что решение нелинейного алгебраического уравнения (8.53) можно заменить многократной последовательной оценкой x_{k+1} по формуле (8.52) с учетом начального значения, определенного по уравнению (8.50), до тех пор, пока не будет достигнута сходимость, т. е. последовательные итерации не станут достаточно близкими друг к другу, при этом

$$x_{k+1}^{(m+1)} = x_k + h[\beta_0^* \Phi(x_{k+1}^{(m)}, t_{k+1}) + \beta_1^* \Phi(x_k, t_k) + \dots]. \quad (8.57)$$

Но и по точности, и по вычислительным затратам в таких случаях, как правило, ощущаются потери.

Следует отметить, что формулы, аналогичные выражениям (8.52) и (8.50), можно вывести и для общего случая, когда происходит изменение шага h и порядка r формулы, если интерполяционный полином в уравнении (8.46) выразить через разделенные разности.

С учетом выражения (8.48), выбирая значения для p , q и r , получаем много других формул численного интегрирования. В частности, для линейных многошаговых методов в общем случае справедливо выражение

$$x_{n+1} = \sum_{i=0}^q a_i x_{n-i} + h \sum_{i=-1}^q b_i \Phi(x_{n-i}, t_{n-i}), \quad (8.58)$$

которое соответствует как явным ($b_{-1}=0$), так и неявным ($b_{-1} \neq 0$) методам.

На базе уравнения (8.58) получена известная неявная формула Гира для жестких дифференциальных уравнений:

$$x_{n+1} = hb_{-1}\varphi(x_{n+1}, t_{n+1}) + \sum_{i=1}^k a_i x_{n-i}, \quad k \leq 6, \quad (8.59)$$

из которой легко определяется аппроксимация для производной:

$$\dot{x}_{n+1} = -\frac{1}{h} \sum_{i=0}^k \alpha_i^* x_{n+1-i}, \quad (8.60)$$

коэффициенты которой в предположении, что вычисления ведутся с переменным шагом,

$$\alpha_i^* = \left(\frac{t_n - t_{n-1}}{t_n - t_{n-i}} \right) \prod_{\substack{j=1 \\ j \neq i}}^k \left(\frac{t_n - t_{n-j}}{t_{n-i} - t_{n-j}} \right), \quad i = 1, 2, \dots, k.$$

Ей сопутствует формула прогноза, аналогичная выражению (8.50) при использовании раздельных разностей,

$$x_{n+1}^{(0)} = \sum_{i=1}^{k+1} \gamma_i x_{n+1-i}, \quad (8.61)$$

где

$$\gamma_i = \prod_{\substack{j=1 \\ j \neq i}}^{k+1} \left(\frac{t_n - t_{n-j}}{t_{n-i} - t_{n-j}} \right), \quad i = 1, \dots, k+1.$$

Неявный метод интегрирования с обратными разностями первого порядка Брайтона, или формула дифференцирования назад (ФДН) отличается от выражения (8.60) и (8.61) введением в соответствующие выражения обратной разности

$$\Delta x_{n-i} = x_{n+1-i} - x_{n-i}, \quad i = 1, 2, \dots, k.$$

При этом получаем:

$$\dot{x}_{n+1} = -\frac{1}{h} \sum_{i=0}^{k-1} \hat{\alpha}_i \Delta x_{n-i}, \quad (8.62)$$

$$x_{n+1}^{(0)} = x_n + \sum_{i=1}^k \hat{\gamma}_i \Delta x_{n-i}, \quad (8.63)$$

где

$$\hat{\alpha}_i = \sum_{j=0}^i \alpha_j^*, \quad \hat{\gamma}_i = -1 + \sum_{j=1}^i \gamma_j.$$

Неявный метод численного интегрирования жестких систем дифференциальных уравнений, основанный на применении разностей высших порядков, предусматривает прогнозируемое значение вектора неизвестных математической модели консультируемой проблемы x_{n+1} определять с помощью интерполяционного полинома Ньютона для неравных промежутков:

$$x_{n+1}^{(0)} = \sum_{i=0}^k \rho_i \nabla_n^i x_n^t, \tag{8.64}$$

где

$$\begin{aligned} \nabla_n^{(0)} &= 1, \rho_0 = 1; \\ \rho_l &= \rho_{l-1} \frac{t_{n+1} - t_{n+1-l}}{t_n - t_{n-l}}, \quad l = 1, 2, \dots, k; \\ \nabla_{n+1}^i &= \rho_i \nabla_n^i. \end{aligned}$$

При этом для алгебраизации производной по времени используется формула

$$h\dot{x}_{n+1} = c_0 x_{n+1} - \sum_{i=0}^{k-1} c_i \nabla_n^i x_n, \tag{8.65}$$

где

$$c_i = \rho_i \sum_{j=1}^{k-1} \frac{h}{t_{n+1} - t_{n-j}}; \tag{8.66}$$

$$\nabla_n^i x_n = \nabla_n^{i-1} x_n - \rho_{i-1} \nabla_{n-1}^{i-1} x_{n-1}, \quad i = 1, \dots, k+2, \tag{8.67}$$

— разделенные разности, вычисленные в предыдущий момент времени t_n откуда

$$\begin{aligned} \nabla_n x_n &= x_n - x_{n-1}; \\ \nabla_n^2 x_n &= x_n - x_{n-1} - \rho_1 \nabla_{n-1} x_{n-1}; \\ &\dots \dots \dots \dots \dots \dots \dots \\ \Delta_n^{k+1} x_n &= x_n - x_{n-1} - \rho_1 \nabla_{n-1} x_{n-1} - \dots - \rho_k \nabla_{n-1}^k x_{n-1}. \end{aligned}$$

Формулы (8.60), (8.62) и (8.65) могут использоваться для сведения дифференциально-алгебраической системы уравнений (8.1) к системе нелинейных алгебраических уравнений (8.2), соответствующей временной точке t_{n+1} . При оценке начального значения искомого вектора неизвестных $x_{n+1}^{(0)}$ система нелинейных алгебраических уравнений решается методом Ньютона с использованием формул (8.61), (8.63) и (8.64).

Формулы (8.60), (8.62) и (8.65) эквивалентны друг другу, если их вывод проводить при одинаковых исходных предположениях

(постоянном или переменном шаге h), и между их коэффициентами существует однозначная связь. Например, формула второго порядка ($k = 2$) для каждого из перечисленных методов приводится к виду, аналогичному формуле Шихмана:

$$x_{n+1} = \frac{(t_{n+1} - t_{n-1})(t_{n+1} - t_n)}{(2t_{n+1} - t_n - t_{n-1})} \dot{x}_{n+1} + \frac{(t_{n+1} - t_{n-1})^2}{(2t_{n+1} - t_n - t_{n-1})} x_n - \frac{(t_{n+1} - t_n)^2}{(2t_{n+1} - t_n - t_{n-1})} x_{n-1}. \quad (8.68)$$

Об эквивалентности методов (8.60), (8.62) и (8.65) свидетельствует и их одинаковая численная устойчивость при $k \leq 6$ (рис. 8.2, а и б).

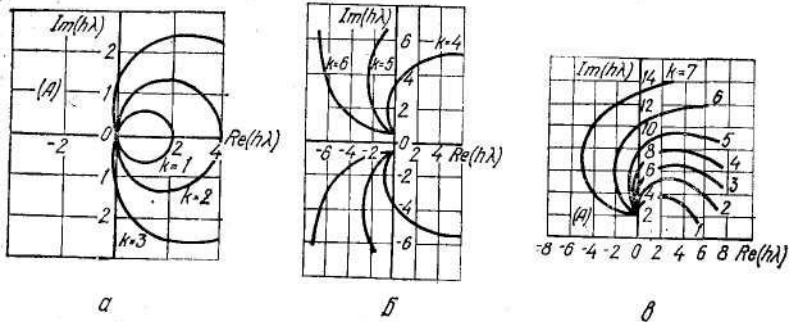


Рис. 8.2. Годографы устойчивости неявных многошаговых методов численного интегрирования Гира—Брайтона (а, б) и Энрайта (в); (А) — зона устойчивости

Известно, что при $k \leq 2$ эти методы обладают А-устойчивостью, т. е. область их абсолютной устойчивости включает всю левую полуплоскость $\text{Re}(h\lambda) < 0$, где λ — собственные значения матрицы Якоби для правой части уравнения (8.45). Для $k = 3, \dots, 6$ они, с одной стороны, А(α)-устойчивы, т. е. область их абсолютной устойчивости включает бесконечный клин $|\arg(-h\lambda)| < \alpha$, а с другой, — жестко устойчивы, так как область их абсолютной устойчивости ограничена частью левой полуплоскости, при этом в окрестности начала координат обеспечивается не только устойчивость, но и точность. В табл. 8.1 приведены значения α для формул различных порядков $k \leq 6$ при относительном изменении шага вычислений $q = h/h_{r.1}$.

Таблица 8.1

$A(\alpha)$ -устойчивость многошаговых неявных формул численного интегрирования

q	α^0 , при k						
	1	2	3	4	5	6	7
1	90	90	86	74	52	18	0
1,01	90	90	85	72	48	10	0
1,02	90	90	85	71	45	1	0
1,05	90	90	83	65	32	0	0
1,1	90	89	81	55	1	0	0
1,2	90	87	70	25	0	0	0
1,5	90	75	25	0	0	0	0
2,0	90	50	0	0	0	0	0

Среди часто используемых на практике методов вычисления приближения x_{n+1} , кроме рассмотренных выше, следует отметить следующие.

1. Одношаговый A -устойчивый метод трапеций, для которого

$$\begin{aligned} x_{n+1} &= x_n + \frac{h}{2} (\dot{x}_{n+1} + \dot{x}_n) = \\ &= x_n + \frac{h}{2} \{ \varphi(x_{n+1}, t_{n+1}) + \varphi(x_n, t_n) \}. \end{aligned} \quad (8.69)$$

Для повышения точности вычислений рекомендуется этот метод использовать совместно со схемой экстраполяции

$$\tilde{x}_{n+1} = x_{2n+2}(h/2) + \frac{[x_{2n+2}(h/2) - x_{n+1}(h)]}{3}, \quad (8.70)$$

где $x_{n+1}(h)$ и $x_{2n+2}(h/2)$ — решения, полученные по формуле (8.69) соответственно с шагом h и $h/2$ для каждой точки t_{n+1} .

2. Многошаговый неявный метод Энрайта, в котором используется вторая производная,

$$x_{n+1} = x_n - h \sum_{j=-1}^k \beta_j \dot{x}_{n-j} - h^2 \ddot{x}_{n+1} = 0, \quad (8.71)$$

годографы устойчивости которого показаны на рис. 8.2, ϵ для $k \leq 7$. При решении нелинейных задач для вычисления второй производной

$$\ddot{x}_{n+1} = \frac{\partial \varphi}{\partial x} (x_{n+1}^{(m)}) \dot{x}_{n+1}^{(m)} \quad (8.72)$$

необходимо вычислять матрицы Якоби на каждой итерации и на каждом шаге, поэтому данный метод слишком трудоемкий. Он обладает большей точностью и устойчивостью, чем методы, выраженные уравнением (8.60), (8.62) и (8.65).

3. Композиционный многошаговый неявный метод Скила, соответствующий по смыслу методу Энрайта и представляющий собой комбинацию метода Гира — Брайтона, или ФДН, и метода Адамса — Мулттона (8.51):

$$\{AM^{(k+1)}\} + \gamma^{(k)}hJ \{ФДН^{(k)}\}, \quad (8.73)$$

где $J = \frac{\partial \varphi}{\partial x}$ — матрица Якоби для уравнения (8.45).

Выражение (8.73) имеет порядок $k+1$ для каждого шага вычислений k . Параметр $\gamma^{(k)}$ в уравнении (8.73) выбирают исходя из улучшения численной устойчивости метода, которая в данном случае охватывает методы до 12-го порядка включительно.

Данные о значениях $\gamma^{(k)}$ и параметре α для характеристики $A(\alpha)$ устойчивости сведены в табл. 8.2.

Таблица 8.2

$A(\alpha)$ -устойчивость композиционных методов неявного численного интегрирования

Шаг k	Порядок $k+1$	$\gamma^{(k)}$	$\alpha, ^\circ$	Шаг k	Порядок $k+1$	$\gamma^{(k)}$	$\alpha, ^\circ$
1	2	0,5	90	7	8	0,08754865	77,4
2	3	0,166667	90	8	9	0,08105623	70,2
3	4	0,125	90	9	10	0,07599874	60,7
4	5	0,1284997	89,4	10	11	0,07192936	47,6
5	6	0,1087264	87	11	12	0,06857227	28,7
6	7	0,09625961	82,9				

4. Блочные или циклические методы, основанные на идее одновременного вычисления «блока» приближения $x_{n+1}, x_{n+2}, \dots, x_{n+l}$. Блок просчитывается с постоянным шагом на выбранном интервале $t_n = t_0 + nh$, где $n = ml + j$ ($j = 1, 2, \dots, l$ и $m = 0, 1, 2, \dots$), при этом сразу используется несколько независимых многошаговых методов:

$$\sum_{j=-k+1}^l [a_{ij}x_{ml+j} - hb_{ij}\dot{x}_{ml+j}] = 0, \quad i = 1, 2, \dots, l. \quad (8.74)$$

Например, при выборе исходных многошаговых методов третьего порядка ($k = 3$) получаем:

$$\begin{aligned} -2x_{ml-2} + 9x_{ml-1} - 18x_{ml} + 11x_{ml+1} - hx_{ml+1} &= 0; \\ -2x_{ml-1} + 9x_{ml} - 18x_{ml+1} + 11x_{ml+2} - h6\dot{x}_{ml+2} &= 0; \\ 9x_{ml+1} - 12x_{ml+2} + 3x_{ml+3} - h(-4\dot{x}_{ml+1} - 4\dot{x}_{ml+2} + 2\dot{x}_{ml+3}) &= 0. \end{aligned}$$

Следовательно, если блок приближений $[x_{ml-2}, x_{ml-1}, x_{ml}]^t$ известен, то по приведенным формулам определяется следующий блок $[x_{ml+1}, x_{ml+2}, x_{ml+3}]^t$, а затем процесс приближений повторяется.

Неявные циклические методы характеризуются большей точностью и устойчивостью, чем методы, выраженные уравнениями (4.48), (4.50) и (4.53), что следует из табл. 8.3.

Таблица 8.3

A(α)-устойчивость циклических неявных методов численного интегрирования

k	$\alpha, \text{ }^\circ$	
	Метод Гира	Циклический метод
3	86	89,4
4	74	81
5	52	78
6	18	63
7	0	34

Их целесообразно применять для моделирования консультируемых проблем, комплексные собственные значения моделей которых расположены вблизи мнимой оси плоскости $h\lambda$. Установлено, что эти методы жестко устойчивы при порядке $k \leq 25$.

Оценка локальной погрешности. Каждый из методов численного интегрирования характеризуется локальной погрешностью, или погрешностью на данном шаге вычислений. Для k -шагового метода, в котором приближение x_{n+1} к точным значениям $x(t_{n+1})$ вычисляется через ранее вычисленные приближенные значения $x_{n-k}, x_{n-k+1}, \dots, x_{n-1}$, можно записать обобщенную формулу вычислений в виде

$$\sum_{i=0}^k \alpha_i x_{n-i} |h - \varphi(x_{n+1}, \dots, x_{n-k}) = 0. \quad (8.75)$$

Раскладывая функцию в уравнении (8.75) в ряд Тейлора в окрестности точки x_{n+1} , получаем оценку для локальной погрешности дискретизации как оценку отсечения. Если метод имеет k -й порядок, то приближенно

$$e_T \approx [C_{k+1} x^{(k+1)}(\tau) h^{k+1} + 0(h^{k+2})], \quad (8.76)$$

где $x^{(k+1)}(\tau)$ — $(k+1)$ -я производная от решения $x(t)$ в промежуточной точке $t_n < \tau < t_{n+1}$; $0(h^{k+2})$ — члены отсечения, содержащие h в степени $k+2$ и выше и которыми обычно пренебрегают; C_{k+1} — константа, характеризующая численный метод.

Например, для формулы Адамса—Мултона (8.51)

$$C_{k+1} = \frac{1}{(k+1)!} \left[(k-1)^{k+1} - (k-2)^{k+2} - (k+1) \sum_{i=-1}^{k-3} \beta_i^* (k-2-i)^k \right],$$

для формулы Гира (8.60) при постоянном шаге

$$C_{k+1} = \frac{1}{(k+1)!} \left[k^{k+1} - \sum_{i=0}^{k-2} a_i (k-1-i)^{k+1} - b_{-1} (k+1) k^k \right],$$

для метода трапеций с учетом $k=1$ $C_2 = 1/12$ и др.

Следует отметить, что выражение (8.76) мало удобно для практических применений, поэтому для оценки локальной погрешности ϵ_T применяют различные косвенные методы.

Прежде всего, можно производные высоких порядков выразить приближенно через разности высоких порядков, используемые, например, в формуле (8.65):

$$\epsilon_T \approx \frac{h}{t_{n+1} - t_{n-k}} \nabla_{n+1}^{k+1} x_{n+1} = \frac{1}{t_{n+1} - t_{n-k}} \left(x_{n+1} - \sum_{i=0}^k \theta_i \nabla_n^i x_n \right). \quad (8.77)$$

Далее можно применить экстраполяцию значений переменных $x_{n+1}(h)$ и $x_{n+1}(h/2)$, полученных интегрированием от точки t_n до точки t_{n+1} один раз с шагом h и другой раз с шагом $h/2$, при этом

$$\epsilon_T = \frac{x_{n+1}(h) - x_{n+1}(h/2)}{h(1-2^{-k})}, \quad (8.78)$$

где k — порядок используемого метода интегрирования.

Наконец, можно использовать способ вложенных методов, когда на каждом шаге интегрирование выполняется дважды: методом порядка k и методом порядка $k+1$. С помощью разности полученных значений

$$x_{n+1}^{(k)} \text{ и } x_{n+1}^{(k+1)}$$

определим оценку локальной погрешности

$$\epsilon_T = \frac{x_{n+1}^{(k+1)} - x_{n+1}^{(k)}}{h}. \quad (8.79)$$

Здесь, как и при экстраполяции, имеется выбор: продолжать вычисления по значению, полученному методом порядка k или методом порядка $k+1$.

Если учесть, что формулы прогноза (8.61), (8.63) и (8.64) имеют порядок на единицу больше, чем основные расчетные уравнения (8.60), (8.62) и (8.65), то к способу вложенных методов можно отнести следующую оценку локальной погрешности:

$$\epsilon_T = \frac{h}{t_{n+1} - t_{n-k}} (x_{n+1} - x_{n+1}^{(0)}). \quad (8.80)$$

Следует отметить, что в выражение (8.80) входит то же значение $x_{n+1}^{(0)}$, которое используется в итерациях Ньютона при решении нелинейного уравнения (8.2). Справедливость уравнения (8.80)

непосредственно следует из выражения (8.77), если учесть формулу (8.64).

Рассмотрим взаимосвязь локальной погрешности ε_t с общей (глобальной) погрешностью ε_{\max} , определяемой консультантом при моделировании различных консультируемых проблем и равной разности вычисляемых x_{n+1} и точных x^*_{n+1} значений вектора переменных уравнений модели.

Чаще всего исходными при моделировании являются интервал интегрирования T (t_0, t_0+T) и максимальная допустимая глобальная погрешность E_{\max} на этом интервале. Тогда допустимая на единичном шаге вычислений локальная погрешность определяется по формуле

$$\varepsilon_{T, n+1} \leq \frac{E_{\max}}{T} h_{n+1}. \quad (8.81)$$

Экспериментальные исследования показывают, что при использовании выражения (8.81) решение получается часто более точным, чем требуется, особенно в начале интервала интегрирования, в результате чего выполняется завышенное число шагов вычислений. Определенную избыточность решения можно уменьшить, если вместо соотношения (8.81) ввести другое, учитывающее действительный характер изменения переменных при решении уравнения модели объекта.

Например, линеаризация уравнения (8.1)

$$F(\dot{x}, x, t) = 0$$

с учетом возможных аппроксимаций для производной вида (8.60)

$$\dot{x}_{n+1} = \beta_0 x_{n+1} - \frac{1}{h} \sum_{i=1}^k \alpha_i x_{n+1-i}$$

приводит к выражению:

$$\left[\beta_0 \frac{\partial F}{\partial \dot{x}} + \frac{\partial F}{\partial x} \right]_{x_{n+1}} x_{n+1} - \left[\frac{\partial F}{\partial \dot{x}} \right]_{x_{n+1}} \sum_{i=1}^k \frac{\alpha_i}{h} x_{n+1-i} = 0. \quad (8.82)$$

Теперь предположим, что общая погрешность невелика, так что производные, вычисленные при значениях x_{n+1} и x^*_{n+1} можно считать равными, и учтем, что для точных значений вектора переменных уравнения модели консультируемой проблемы справедливо уравнение

$$x^*_{n+1} = \beta_0 x^*_{n+1} - \frac{1}{h} \sum_{i=1}^k \alpha_i x^*_{n+1-i} - \varepsilon_{n+1},$$

где

$$\varepsilon_{n+1} = \frac{\varepsilon_{T, n+1}}{h_{n+1}}$$

— локальная погрешность вычислений в единицу времени, определяемая с учетом уравнений (8.80) или (8.76). Тогда по аналогии с предыдущим записываем

$$\left[\beta_0 \frac{\partial F}{\partial x} + \frac{\partial F}{\partial x} \right]_{x_{n+1}} x_{n+1}^* - \left[\frac{\partial F}{\partial x} \right]_{x_{n+1}} \left[\sum_{i=1}^k \frac{\alpha_i}{h} x_{n+1-i}^* - \varepsilon_{n+1} \right] = 0 \quad (8.83)$$

и, вычитая линеаризованные уравнения (8.82) и (8.83), находим искомое уравнение для общей (глобальной) погрешности

$$A E_{n+1} = B \left[\sum_{i=1}^k \frac{\alpha_i}{h} E_{n+1-i} + \varepsilon_{n+1} \right], \quad (8.84)$$

где

$$A = \left[\beta_0 \frac{\partial F}{\partial x} + \frac{\partial F}{\partial x} \right]_{x_{n+1}}; \quad B = \left[\frac{\partial F}{\partial x} \right]_{x_{n+1}}.$$

Следует отметить, что матрица A в полученном уравнении (8.84) совпадает с матрицей линеаризованного уравнения модели консультируемой проблемы, поэтому вычисления глобальной погрешности при известных матрицах L и U выполняются только процедурой прямой и обратной подстановок согласно выражениям (8.16) и (8.17). В результате оценка величины E_{n+1} по выбранной ε_{n+1} (и наоборот) не требует ни коренной переделки программы решения, ни чрезвычайно большего времени работы процессора.

Найденное по уравнению (8.84) значение общей погрешности E_{n+1} используется наряду с максимально допустимой ее величиной E_{\max} для управляемого выбора допустимого значения локальной погрешности

$$\varepsilon_{Tj} = \varepsilon_{T, j-1} \left[\frac{\delta E_{\max}}{E_{j-1}} \right]^{\frac{k+1}{k}}, \quad (8.85)$$

где $\delta = 0,5$ ($0 < \delta < 1$) — коэффициент надежности, при этом выбирают $\varepsilon_{T, 0} = 0,1 E_{\max}$ и ограничивают $\varepsilon_{T, \min} \neq 0$.

Выражение (8.85) можно получить, если учесть из уравнения (8.84) пропорциональность E_{n+1} и ε_{n+1} , а из выражения (8.76) зависимость h и ε , т. е. записать

$$E_{n+1} \sim \frac{\varepsilon_{T, n+1}}{h_{n+1}} \sim \frac{\varepsilon_{T, n+1}}{\frac{k+1}{\sqrt{\varepsilon_{T, n+1}}}} = \varepsilon_{T, n+1}^{\frac{k+1}{k}}. \quad (8.86)$$

Стратегии выбора шага (порядка). Подставляя формулу (8.76) в выражение (8.81), находим взаимосвязь максимально допустимой погрешности на единичном шаге вычислений ε с максимальным размером шага h :

$$\varepsilon = [C_{k+1} x^{(k+1)}(\tau)] h^k, \quad (8.87)$$

которая изображена на рис. 8.3 для $k = 1, 2, 3$, и 4.

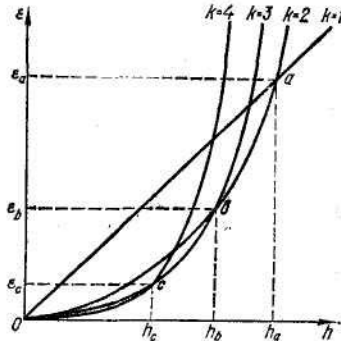


Рис. 8.3. Кривые зависимости допустимой погрешности на единичном шаге вычислений от изменений порядка k алгоритма

Из рисунка следует, что при $\varepsilon > \varepsilon_a$ алгоритм первого порядка (неявный метод Эйлера) обеспечивает наибольший допустимый шаг вычислений, при $\varepsilon_b < \varepsilon < \varepsilon_a$ — алгоритм второго порядка (метод Шихмана), при $\varepsilon_c < \varepsilon < \varepsilon_b$ — алгоритм третьего порядка и при $\varepsilon < \varepsilon_c$ — алгоритм четвертого и более высоких порядков. Так как положение кривых на рис. 8.3 зависит от текущих значений C_{k+1} и $x^{(k+1)}(t_{n+1})$, то даже при выбранном постоянном ε оптимальные порядок k алгоритма и шаг вычислений h могут изменяться от одного временного шага к другому. Как следует из рис. 8.3, любое увеличение допустимой погрешности ε не приведет к увеличению порядка формулы для максимизации шага вычислений и любое уменьшение ε не может привести к уменьшению оптимального порядка k используемого метода. Для достижения максимальной вычислительной эффективности целесообразно в процессе интегрирования одновременно изменять порядок алгоритмов и шаг с помощью многошаговых алгоритмов.

С точки зрения программирования изменение порядка алгоритма эквивалентно смене набора коэффициентов, определяющих ту или иную многошаговую формулу. При увеличении порядка k алгоритма

увеличивается число коэффициентов в наборе. Это сопровождается ростом загрузки памяти ЭВМ для хранения «прошлых» значений вектора переменных, используемых для оценки приближения формулами высоких порядков. Как правило, указанные дополнения не требуют существенных затрат. Наоборот, изменение шага связано со значительными вычислительными затратами, так как часто хранимые «прошлые» значения вектора переменных, соответствующие шагу h_{j-1} необходимо интерполировать для получения преобразованных «прошлых» значений, соответствующих новому шагу вычислений h_j . Процедура интерполяции упрощается, если «прошлые» значения хранить в виде вектора обратных разностей

$$z_n = [x_n, x_{n-1}, \dots, x_{n-k}, h\dot{x}_n, \nabla^1(h\dot{x}_n), \nabla^2(h\dot{x}_n), \dots, \nabla^n(h\dot{x}_n)]^t$$

или в виде вектора Нордсика

$$y_n = \left[x_n, h\dot{x}_n, \frac{h^2 x_n''}{2}, \frac{h^3 x_n'''}{3!}, \dots, \frac{h^{(2k-1)} x_n^{(2k+1)}}{(2k+1)!} \right]^t,$$

старые и новые значения которых связаны матрицей, НЭ которой зависят только от коэффициента $q = h_j/h_{j-1}$.

Успешно используются разные стратегии изменения порядка и шага методов интегрирования. В соответствии с одной из них, предложенной Гиром, выбор шага осуществляется по формуле

$$h = \left[\frac{E_{\max}}{TC_{k+1} x^{(k+1)}(t_n)} \right]^{1/k}, \tag{8.88}$$

получаемой на основании выражений (8.87) и (8.81).

На основании одного из выражений (8.76) или (8.80) оценивается погрешность вычисления каждой составляющей вектора x сразу для методов трех ближайших порядков

$$\varepsilon_{k-1}, \varepsilon_k, \varepsilon_{k+1}, \quad 1 \leq k \leq 6. \tag{8.89}$$

Далее выбираются погрешности, соответствующие наименее точным составляющим вектора переменных x :

$$e_1 = \max \varepsilon_{k-1}, \quad e_2 = \max \varepsilon_k, \quad e_3 = \max \varepsilon_{k+1}.$$

Порядок метода после каждых k шагов вычислений выбирается равным порядку, при котором величина e_i минимальна ($i = 1, 2, 3$). Новое значение шага h определяется в соответствии с выражением (8.88). При этом шаг вычислений при неизменном порядке метода корректируется не сразу, а только при достижении определенных соотношений между действительной и заданной погрешностями вычислений между используемым и разрешенным значениями шага интегрирования. Напомним, что в соответствии с табл. 8.1 частые и большие изменения шага не разрешаются, при этом

$$q_{\text{доп}} = h_j/h_{j-1} = \sqrt[k+1]{4}.$$

Изменение шага и порядка не разрешается только для $(k + 1)$ -го шага после последнего изменения шага.

Для начала интегрирования используется метод с $k = 1$ и начало интегрирования не представляет трудностей. Порядок постепенно достигает оптимальной величины по мере увеличения длины шага до максимально возможной. Если порядок ошибочно выбран слишком большим, то разности, фигурирующие в процессе выбора порядка, становятся бессмысленными, при этом в одной и той же точке программа отвергает несколько шагов с последовательно уменьшающейся длиной, а затем продолжает счет.

Суть стратегии переменного шага и переменного порядка не столько в том, что эти величины могут изменяться, сколько в том, что их можно правильно выбирать автоматически. Для большей части вычислений они не должны меняться сильно. Программа является гибкой и адаптируется в зависимости от характера решаемой задачи.

Другая стратегия изменения h и k также обеспечивает максимизацию временного шага, но достигает это следующим образом.

На основании выражения (8.77) применительно к алгоритму (8.65) оценивается погрешность каждой составляющей вектора переменных x для методов 1,2..... $(k + 1)$ -го порядков:

$$\epsilon^1, \epsilon^2, \dots, \epsilon^k, \quad 1 \leq k \leq 6. \quad (8.90)$$

Среди вычисленных погрешностей выбираются погрешности, соответствующие наименее точным компонентам вектора x ,

$$e_i = \max(\epsilon^l), \quad i = 1, 2, \dots, k + 1,$$

и порядок метода выбирается равным порядку, при котором величина e_i ($i = 1, 2, \dots, k + 1$) минимальна.

Новый временной шаг определяется в соответствии с выражением

$$h_j = \sqrt[k]{\frac{e_i}{E_{ik}}} h_{j-i}, \quad (8.91)$$

где $e_i = |\epsilon x_i|$;

$$E_{ik} = \frac{h}{t_{n+1} - t_{n-k}} (x_{(n+1)i} - x_{(n+1)i}^{(0)});$$

ϵ — допустимая относительная погрешность; i — номер компоненты вектора x .

При этом величина изменения шага ограничивается, как и раньше, коэффициентом

$$q = h_j/h_{j-1} \leq \sqrt[k+1]{4}.$$

Описанная схема изменения порядка метода интегрирования и временного шага может выполняться всякий раз при изменении временного интервала. В ней благодаря использованию разностей высших порядков уменьшаются вычислительные ошибки и сокращаются затраты при определении коэффициентов формул прогноза (8.64) и алгебраизации (8.65).

Кроме того, достигается высокая гибкость программ при изменении порядка алгоритма интегрирования. Благодаря легкости оценки локальных погрешностей алгоритмов интегрирования различных порядков рассмотренная стратегия позволяет легко изменять порядок в пределах от 1-го до $(k + 1)$ -го на каждом временном шаге (а не через $(k + 1)$ шагов в пределах от $(k - 1)$ -го до $(k + 1)$ -го), что обеспечивает ее высокую экономичность по вычислительным затратам.

Сравнение методов интегрирования. Кратко рассмотрим некоторые известные модельные задачи и их решение, полученное с помощью описанных выше методов. Из набора 25 тестовых задач выберем только пять как наиболее сложные для отдельных методов.

В5. Система линейных дифференциальных уравнений с комплексными собственными значениями ($\lambda_i = -0,1; -0,5; -1; -4; -10; \pm j100$):

$$\begin{cases} \dot{x}_1 = -10x_1 + 100x_2, & x_1(0) = 1; \\ \dot{x}_2 = -100x_1 - 10x_2, & x_2(0) = 1; \\ \dot{x}_3 = -4x_3, & x_3(0) = 1; \\ \dot{x}_4 = -x_4, & x_4(0) = 1; \\ \dot{x}_5 = -0,5x_5, & x_5(0) = 1; \\ \dot{x}_6 = -0,1x_6, & x_6(0) = 1. \end{cases}$$

Интегрирование ведется в интервале $T = [0, 20]^t$, минимальный начальный шаг $h_0 = 10^{-2}$.

С5. Система нелинейных дифференциальных уравнений, в решении которой быстро осциллирующие компоненты нелинейно зависят от плавных компонент:

$$\begin{cases} \dot{x}_1 = -x_1 + 2, & x_1(0) = 1; \\ \dot{x}_2 = -10x_1 + 20x_1^2, & x_2(0) = 1; \\ \dot{x}_3 = -40x_3 + 80(x_1^2 + x_2^2), & x_3(0) = 1; \\ \dot{x}_4 = -100x_4 + 200(x_1^2 + x_2^2 + x_3^2), & x_4(0) = 1, \end{cases}$$

интегрирование ведется в интервале $T = [0, 20]^t$, минимальный начальный шаг $h = 10^{-2}$.

Д6. Система нелинейных дифференциальных уравнений с действительными собственными значениями

$$(\lambda_i = 0; -1,0 \rightarrow -8,6; -3 \cdot 10^7 \rightarrow 4 \cdot 10^7,$$

где знак \rightarrow показывает изменение λ_i от начальных до конечных значений):

$$\begin{aligned} \dot{x}_1 &= -x_1 + 10^8 x_3 (1 - x_1), & x_1(0) &= 1; \\ \dot{x}_2 &= -10x_2 + 3 \cdot 10^7 x_3 (1 - x_2), & x_2(0) &= 0; \\ \dot{x} &= -\dot{x} - \dot{x}_2, & x_3(0) &= 0. \end{aligned}$$

Интегрирование ведется до $T_{\max} = 1$, минимальный начальный шаг $h_0 = 3,3 \cdot 10^{-8}$.

Е4. Система нелинейных дифференциальных уравнений с комплексными собственными значениями:

$$\dot{x} = -Ut \begin{bmatrix} -10 & -10 & 0 & 0 \\ 10 & -10 & 0 & 0 \\ 0 & 0 & 10^3 & 0 \\ 0 & 0 & 0 & 10^{-2} \end{bmatrix} Ux + G(x); \quad x(0) = \begin{bmatrix} 0 \\ -2 \\ -1 \\ -1 \end{bmatrix};$$

$$U = \frac{1}{2} \begin{bmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix}; \quad G(x) = U \begin{bmatrix} (z_1^2 - z_2^2)/2 \\ z_1 z_2 \\ z_3^2 \\ z_4 \end{bmatrix}; \quad \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} = Ux.$$

Интегрирование ведется до $T_{\max} = 10^3$ с начальным шагом $h_0 = 10^{-3}$. Собственные значения решаемой системы изменяются в процессе решения, при этом

$$\begin{aligned} \lambda_1 &= : -2,0 \rightarrow 1,0 \cdot 10^{-2}; \\ \lambda_2, \lambda_3 &= : 8,0 \pm j10 \rightarrow 7,1 \pm j7,9 \rightarrow 9,0 \pm j0,03 \rightarrow 12 \pm j13 \rightarrow 0,19 \pm j29 \rightarrow \\ &\rightarrow -17 \pm j18 \rightarrow -10 \pm j10; \\ \lambda_4 &= : -10^3. \end{aligned}$$

Е5. Система нелинейных дифференциальных уравнений с комплексными собственными значениями:

$$\begin{aligned} \dot{x}_1 &= -7,89 \cdot 10^{-10} x_1 - 1,1 \cdot 10^7 x_1 x_3; \\ \dot{x}_2 &= 7,89 \cdot 10^{-10} x_1 - 1,13 \cdot 10^9 x_2 x_3; \\ \dot{x}_3 &= 7,89 \cdot 10^{-10} x_1 - 1,1 \cdot 10^7 x_1 x_3 + 1,13 \cdot 10^9 x_4 - 1,13 \cdot 10^9 x_2 x_3; \\ \dot{x}_4 &= 1,1 \cdot 10^7 x_1 x_3 - 1,13 \cdot 10^9 x_4; \\ x(0) &= [1,76 \cdot 10^{-3}; 0; 0; 0]t. \end{aligned}$$

Интегрирование ведется до $T_{\max} = 10^3$ с начальным шагом $h_0 = 5 \cdot 10^{-5}$. Собственные значения решаемой задачи изменяются в процессе решения, при этом

$$\lambda_1 = : 0;$$

$$\lambda_2, \lambda_3 = : -7,5 \cdot 10^{-10} \pm j9,2 \cdot 10^{-4} \rightarrow 2,9 \cdot 10^{-4} \pm j8,7 \cdot 10^{-4} \begin{cases} \rightarrow -9,4 \cdot 10^{-8}; \\ \rightarrow -0,019; \end{cases}$$

$$\lambda_4 = : -2,0 \cdot 10^4.$$

Основные характеристики решений для задач, полученных для заданной точности $E = 10^{-4}$, приведены в табл. 8.4. Примеры решались четырьмя основными методами, указанными в табл. 8.4:

Таблица 8.4

Задачи при погрешности $E = 10^{-4}$

Номер задач	Метод решения характеристики	Метод			
		1	2	3	4
B5	Число шагов	2752	471	351	277
	Количество вычислений матрицы Якоби	24	471	11	21
	Количество обращений матрицы (<i>LU</i> -преобразований)	24	471	34	21
	Число вычислений функции	6738	3297	545	922
C5	Число шагов	312	89	123	103
	Количество вычислений матрицы Якоби	31	203	13	17
	Количество обращений матрицы (<i>LU</i> -преобразования)	31	376	22	17
	Число вычислений функции	881	1188	282	344
D6	Число шагов	53	33	42	29
	Количество вычислений матрицы Якоби	9	36	8	7
	Количество обращений матрицы (<i>LU</i> -преобразований)	9	62	12	7
	Число вычислений функции	125	170	106	99
E4	Число шагов	1060	679	185	170
	Количество вычислений матрицы Якоби	26	818	18	28
	Количество обращений матрицы (<i>LU</i> -преобразований)	26	1541	32	28
	Число вычислений функции	2963	4813	355	558

1 — методом Гира в соответствии с выражениями (8.59) и (8.60), реализованными в программе *GEAR*;

2 — методом Брайтона (ФДН) в соответствии с выражениями (8.62), (8.63) и формулами (8.80), (8.88), (8.89);

3 — циклическим методом в соответствии с выражением (8.74) и формулами (8.80), (8.88) и (8.89), реализованными в программе *STINT*;

4 — основным методом Скила в соответствии с выражением (8.73) совместно с формулами (8.88) и (8.89), реализованными в программе *DIFS UB*, при этом для оценки локальной погрешности используется вместо уравнения (8.79) следующее выражение:

$$r_T = \frac{h_n^k}{k+1} (x_n^{(k)} - x_{n-1}^{(k)}),$$

так как на соседних шагах порядок метода отличается на единицу.

Данные, соответствующие решению приведенных задач методом Рунге — Кутты (явным и неявным) и трапецией, в табл. 8.4 не введены, так как они значительно уступают другим использованным в комплексном эксперименте методам.

По столь небольшой выборке, как выделенные пять задач, трудно делать категорические выводы и рекомендации. Однако (с учетом и других данных) можно отметить, что новые многошаговые неявные методы типа Скила и циклические методы значительно эффективней, чем широко используемые на практике методы Брайтона и Гира, особенно в случаях, когда матрица Якоби $\partial f/\partial x$ имеет собственные значения с большими мнимыми числами.

Задача E5 по химической кинетике с заданными начальными условиями, по существу, не решена ни одним из методов при заданной погрешности вычислений $E=10^{-4}$. Объясняется это недостаточно выбранной точностью при том разбросе значений коэффициентов, который характерен для рассматриваемой задачи. Если погрешность вычислений снизить хотя бы до 10^{-7} , то задача E5 успешно решается методами 2, 3 и 4 при экономном числе шагов.

В табл. 8.5 сравниваются неявный метод Эйлера ($k = 1$), метод Шихмана второго порядка ($k = 2$) и ФДН ($k \leq 6$) на примере решения задачи E5. Из табл. 8.5 следует, что многошаговые методы высоких порядков, в свою очередь, в несколько раз превосходят по быстродействию неявные методы интегрирования первого и второго порядков.

Таблица 8.5

Результаты решения задачи E5 неявными методами разных порядков при погрешности 10^{-7}

Число оценок функции	Число оценок Якоби	Число решений	Число переменных шагов	Число отброшенных шагов	Максимальный порядок метода	Время, с
3056	556	382	556	0	1	43
1059	207	231	207	1	2	19
372	66	108	51	0	6	6

По-видимому, каждой предметной области консультирования, ее конкретным задачам соответствуют свои оптимальные (один или несколько) методы численного интегрирования, которые должны быть реализованы в соответствующей подсистеме консультационного

анализа. Если решаемые задачи «жесткие» и нелинейные, что справедливо для многих консультируемых проблем, то желательно выбрать методов выполнить среди рассмотренных в этом пункте.

Знание особенностей и возможностей используемых при моделировании численных методов необходимы консультанту для оптимального выбора констант настройки исполнительных программ на решаемую задачу. Несмотря на то, что обычно в моделирующем консультационном комплексе осуществляют настройку управляющих констант на наихудший случай, при котором рассчитаны практически все тестовые задачи, при исследовании консультируемых проблем со сложной реакцией и с нетипичными сочетаниями переменных возможны случаи замедления процесса определения результатов, а в некоторых случаях и получения неверных результатов. Опыт эксплуатации моделирующих программ показал, что наиболее «важными» константами настройки являются следующие.

Значение минимального шага интегрирования *MINSTEP*, с помощью которого отсекаются те высокие частоты изменения переменных, которые не интересуют разработчика. Величина *MINSTEP* выбирается произвольно, однако ее сильное уменьшение может привести к замедлению расчетов («топтанию на месте»), а увеличение — к снижению точности.

Порядок метода интегрирования *ORDER*, который может фиксироваться или изменяться до $k \leq 6$, что существенно сказывается при ожесточении требований к погрешности метода интегрирования и усложнении реакции консультируемой проблемы.

Параметр *REJC*, используемый при оценке необходимости отбрасывания текущего шага при анализе консультируемой проблемы во временной области. Число отброшенных шагов зависит от сложности исследуемых функций, но обычно не превышает 10—20% общего числа шагов.

Относительная погрешность *CONVTA* итераций Ньютона при расчете динамики, влияющая на суммарное число итераций и, следовательно, на общее время расчетов. Величина *XMAX*, или исходное значение максимальной величины составляющих вектора неизвестных, которая является пороговым значением абсолютного изменения переменных. Выбор ее значения зависит от используемой системы единиц. Во всех случаях снижение значения *XMAX* приводит к повышению точности результатов расчета и к увеличению общего времени вычислений. Относительная локальная погрешность метода интегрирования *TRUN*, определяющая условия выбора нового временного шага и влияющая на общее число временных шагов.

Чрезмерное ее уменьшение приводит к большим временным затратам на проведение расчетов. Выбор отмеченных параметров позволяет эффективно управлять процессами интегрирования и уменьшать суммарное время вычислений в несколько раз.

Кроме указанных возможностей управления характеристиками, влияющими на качество работы исполнительных процедур, существуют управляющие параметры, изменяющие режим вычислений, например параметр *DAMP* для изменения режима демпфирования колебаний переменных на итерациях по Ньютону, параметр *IVZ* для изменения частоты переоценки матрицы Якоби, параметр *TAYDC* для изменения режима расчета статики и перехода к методу продолжения решения по параметру и др.

Таким образом, существующие возможности управления характеристиками основных исполнительных процедур обеспечивают достаточно полный комплекс мер, повышающий надежность подсистемы консультационного анализа в целом. Как показывает опыт, наличие подобных средств влияния консультанта на исполнительные подпрограммы увеличивает степень их универсальности, что существенно сказывается на эффективности процедуры моделирования при работе с широким ассортиментом нелинейных консультируемых проблем. В то же время, поскольку выбор отдельных управляющих параметров лежит полностью или частично на консультанте-пользователе САК, то степень его заинтересованности в решении поставленной задачи, личный опыт и знание возможностей моделирующего консультационного комплекса имеют существенное значение.

Интегрирование уравнений распределенных консультируемых проблем. Рассмотренные выше методы численного интегрирования могут использоваться и при нахождении характеристик некоторых консультируемых проблем, отдельные компоненты которых описываются уравнениями в частных производных.

Например, если консультируемая проблема предстает длинной активной линией передачи (рис. 8.4, а), справедливы следующие пространственно-временные соотношения:

$$\frac{\partial u}{\partial x} = -Ri - L \frac{\partial i}{\partial t} + e; \quad (8.92)$$

$$\frac{\partial i}{\partial x} = -Gu - C \frac{\partial u}{\partial t} + j, \quad (8.93)$$

где R, G, L, C, e, j — параметры выделенной секции такой линии;
 $x_0 < x < x_n$.

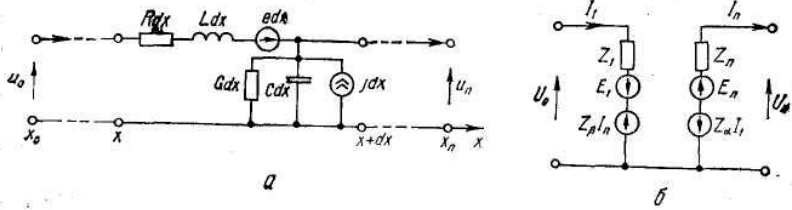


Рис. 8.4. Схема замещения длинной линии: *a* — многосекционная, *б* — эквивалентным четырехполюсником

Используя формулу неявного метода интегрирования (8.60) для алгебраизации частных производных по времени, получаем

$$\frac{\partial i}{\partial t} \Big|_{t=t_k} = -\frac{\alpha_0^*}{h} i_{k+1} - \frac{1}{h} \sum_{j=1}^p \alpha_j^* i_{k+1-j}; \quad (8.94)$$

$$\frac{\partial u}{\partial t} \Big|_{t=t_k} = -\frac{\alpha_0^*}{h} u_{k+1} - \frac{1}{h} \sum_{i=1}^p \alpha_i^* u_{k+1-i}. \quad (8.95)$$

На основании исходных выражений (8.92) и (8.93) с учетом соотношений (8.94) и (8.95) находим преобразованные уравнения длинной линии, справедливые для каждого данного момента времени t_{k+1} :

$$\frac{d}{dx} \begin{bmatrix} U \\ I \end{bmatrix} = - \begin{bmatrix} 0 & Z \\ Y & 0 \end{bmatrix} \begin{bmatrix} U \\ I \end{bmatrix} + \begin{bmatrix} E \\ J \end{bmatrix}, \quad (8.96)$$

где

$$Z = R_{k+1} + \lambda L_{k+1};$$

$$Y = G_{k+1} + \lambda C_{k+1};$$

$$E = e_{k+1} + \lambda L_{k+1} \sum_{j=1}^p \left(\frac{\alpha_j^*}{\alpha_0^*} \right) i_{k+1-j};$$

$$J = j_{k+1} + \lambda C_{k+1} \sum_{i=0}^p (\alpha_i^* / \alpha_0^*) u_{k+1-i};$$

$$\lambda = -\frac{\alpha_0^*}{h}; \quad U = u_{k+1}, \quad I = i_{k+1},$$

Параметры Z , Y , E , J , вошедшие в уравнение (8.96), примут другие значения, если вместо неявного метода (8.60) использовать метод (8.73) или (8.74).

Полученные уравнения (8.96) затем интегрируются по пространственной переменной x . Если линия линейная и однородная, то

параметры Z и Y не зависят от индекса временного шага k и общее решение можно получить в аналитическом виде. Если линия нелинейна и неоднородна, то для решения уравнения (8.96) используются численные методы. Например, применяя аппроксимирующие разности первого порядка, находим

$$\frac{\partial U}{\partial x} \Big|_{x=x_i} = \frac{U_i - U_{i-1}}{\Delta x_i}, \quad i = 1, \dots, n; \tag{8.97}$$

$$\frac{\partial I}{\partial x} \Big|_{x=x_i} = \frac{I_{i+1} - I_i}{\Delta x_i}, \quad i = 1, \dots, n-1, \tag{8.98}$$

где n — общее число дискретных точек (секций), выделяемых в линии.

На основании выражений (8.96)—(8.98) можно записать следующую систему $2n - 1$ преобразованных уравнений длинной линии:

$$\begin{aligned} U_i - U_{i-1} &= -Z_i I_i + E_i, \quad i = 1, \dots, n; \\ I_{i+1} - I_i &= -Y_i U_i + J_i, \quad i = 1, \dots, n-1. \end{aligned} \tag{8.99}$$

Структура уравнений (8.99) такова, что внутренние переменные длинной линии легко исключаются из них, в результате чего они сводятся к уравнениям эквивалентного четырехполюсника (рис. 8.4, б). Параметры Z_α и Z_β такого четырехполюсника находятся в процессе гауссова исключения и приведения системы уравнений вида

Y_1	1				-1			U_1	J_1
-1	Z_2	1						I_2	E_2
	-1	Y_2						U_2	J_2
	\dots							\dots	\dots
			Y_{n-2}	1				U_{n-2}	J_{n-2}
			-1	Z_{n-1}	1			I_{n-1}	E_{n-1}
				-1	Y_{n-1}		1	U_{n-1}	J_{n-1}
1					Z_1	-1		I_1	E_1
					-1	Z_n	1	I_n	E_n
							1	U_0	U_0
							1	U_n	U_n

(8.100)

к системе с верхней и нулевой нижней левыми субматрицами

1	A ₁				B ₁					U ₁	D ₁
	1	A ₂			B ₂ ^{''}					I ₂	D ₂ ^{''}
		1			B ₂ [']					U ₂	D ₂ [']
			
			1	A _{n-2}	B _{n-2} [']	C _{n-2} [']				U _{n-2}	D _{n-2} [']
			1	A _{n-1}	B _{n-2} ^{''}	C _{n-2} ^{''}				I _{n-1}	D _{n-2} ^{''}
				1	B _{n-1} [']	C _{n-1} [']				U _{n-1}	D _{n-1} [']
0					Z ₁	Z _β	-1			I ₁	E ₁
					Z _α	Z _n		1		I _n	E _n
							1			U ₀	U ₀
								1		U _n	U _n

(8.101)

В случае нелинейной и неоднородной длинной линии параметры $Z_\beta, Z_\alpha, Z_1, Z_n$ оцениваются для каждого временного шага.

Для консультируемых проблем, распределенных в двух или трех направлениях, вместо уравнений (8.97) и (8.98) используются более сложные разностные аппроксимации, учитывающие близлежащие выделяемые точки консультируемой проблемы, лежащие на регулярной или нерегулярной сетке. Для решения уравнений с частными производными широко применяются также методы коллаций, Галеркина, Ритца и конечного элемента.

8.5. Методы анализа повышенной эффективности

Диакоптические методы. Диакоптические методы основаны на фрагментации модели сложной консультируемой проблемы, организации отдельных вычислений по фрагментам с периодическим согласованием результатов, получаемых в отдельных фрагментах. Диакоптические методы применяют для решения систем различных уравнений совместно с традиционными численными методами.

Для решения систем линейных алгебраических уравнений (ЛАУ) $AV=B$ применяют *диакоптический вариант метода Гаусса*, основанный на приведении матрицы коэффициентов к блочно-диагональному виду с окаймлением (БДО). При анализе схем этот вариант называют *методом подсхем*. В методе подсхем исходную схему разбивают на фрагменты (подсхемы). Фазовые переменные

(например, узловые потенциалы) делят на внутренние переменные фрагментов и граничные переменные. Вектор фазовых переменных

$$\mathbf{V} = (\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_l, \mathbf{V}_L),$$

где \mathbf{V}_i — подвектор внутренних переменных i -го фрагмента; l — число фрагментов; \mathbf{V}_L — подвектор граничных переменных ($L = l+1$). Матрица коэффициентов имеет структуру БДО

\mathbf{A}_{11}	0	0	.	0	\mathbf{A}_{1L}
0	\mathbf{A}_{22}	0	.	0	\mathbf{A}_{2L}
0	0	\mathbf{A}_{33}	.	0	\mathbf{A}_{3L}
0	0	0	.	\mathbf{A}_{ll}	\mathbf{A}_{lL}
\mathbf{A}_{L1}	\mathbf{A}_{L2}	\mathbf{A}_{L3}	.	\mathbf{A}_{Ll}	\mathbf{A}_{LL}

Прямой ход метода Гаусса состоит из L этапов. На i -м этапе исключаются переменные \mathbf{V}_i , при этом пересчет коэффициентов по формуле Гаусса производится только в подсистеме уравнений

$$\begin{array}{|c|c|} \hline \mathbf{A}_{ii} & \mathbf{A}_{iL} \\ \hline \mathbf{A}_{Li} & \mathbf{A}_{LL} \\ \hline \end{array} \begin{array}{|c|} \hline \mathbf{V}_i \\ \hline \mathbf{V}_L \\ \hline \end{array} = \begin{array}{|c|} \hline \mathbf{B}_i \\ \hline \mathbf{B}_L \\ \hline \end{array},$$

где \mathbf{V}_i и \mathbf{V}_L — подвекторы правых частей соответствующих уравнений.

Следовательно, решение одной сложной задачи заменяется решением L задач меньшей размерности, что при $\alpha > 1$ в формуле (8.43) дает снижение T_m . Кроме того, при нехватке оперативной памяти в методе подсchem можно минимизировать число обменов информацией между оперативной и внешней памятью.

Для решения систем нелинейных конечных уравнений используют *диакоптический вариант метода Ньютона* с контролем сходимости итерационного процесса отдельно по выделенным фрагментам. Выполнение условий сходимости в i -м фрагменте является основанием для прекращения вычислений по уравнениям этого фрагмента. Очевидно, что раздельное интегрирование означает и раздельное решение подсистем ЛАУ, относящихся к отдельным фрагментам.

Раздельное по фрагментам интегрирование дифференциальных уравнений довольно просто организуется лишь при использовании явных методов. Покажем это на примере решения методом Эйлера системы ОДУ, представленной в нормальной форме Коши и разделенной на две подсистемы:

$$\dot{\mathbf{V}} = \varphi_1(\mathbf{V}, \mathbf{Z}, t); \quad \dot{\mathbf{Z}} = \varphi_2(\mathbf{V}, \mathbf{Z}, t),$$

где \mathbf{V} и \mathbf{Z} — подвекторы фазовых переменных, относящиеся к разным фрагментам (граничные переменные предполагаются включенными в

один из подвекторов \mathbf{V} или \mathbf{Z}). Обозначая величины шагов в подсистемах h и H и применяя формулу Эйлера для момента времени t_k

$$\dot{\mathbf{V}}(t_k) = \frac{\mathbf{V}(t_k + h) - \mathbf{V}(t_k)}{h}; \quad \dot{\mathbf{Z}}(t_k) = \frac{\mathbf{Z}(t_k + H) - \mathbf{Z}(t_k)}{H},$$

имеем

$$\begin{aligned} \mathbf{V}(t_k + h) &= \mathbf{V}(t_k) + h\Phi_1(\mathbf{V}(t_k), \mathbf{Z}(t_k), t_k), \\ \mathbf{Z}(t_k + H) &= \mathbf{Z}(t_k) + H\Phi_2(\mathbf{V}(t_k), \mathbf{Z}(t_k), t_k), \end{aligned}$$

откуда вычисляем $\mathbf{V}(t_k+h)$ и $\mathbf{Z}(t_k+H)$. Пусть $h < H$. Для выполнения очередного шага по переменным \mathbf{V} нужно вычислить $\mathbf{Z}(t_k+h)$, что осуществляют с помощью интерполяционного полинома степени p , построенного по $p+1$ последним рассчитанным точкам кривой $\mathbf{Z}(t)$. Здесь p — порядок метода интегрирования. При применении метода Эйлера $p=1$, и значение $\mathbf{Z}(t_k+h)$ вычисляют на основе линейной интерполяции по известным точкам $\mathbf{Z}(t_k)$ и $\mathbf{Z}(t_k+h)$. После осуществления шага по переменным \mathbf{V} шаги повторяют — очередной шаг делают по переменным \mathbf{V} или \mathbf{Z} в зависимости от того, в каком фрагменте моделируемое время имеет меньшее значение.

Интегрирование подсистем ОДУ с оптимальным для каждого фрагмента значением шага может привести к существенной экономии затрат машинного времени, особенно при применении неявных методов интегрирования. Однако организация неявного пофрагментного интегрирования оказывается более сложной, чем явного. Примеры методов пофрагментного неявного интегрирования — *методы однонаправленных моделей и релаксации формы сигнала (РФС)*.

Методы однонаправленных моделей и релаксации формы сигнала. Модели многих сложных элементов консультируемых проблем являются однонаправленными. В них могут быть выделены входные и выходные фазовые переменные, причем выходные не влияют на входные. Примерами однонаправленных моделей служат большинство моделей консультируемых проблем, элементы которых представлены логическими элементами.

При представлении макромоделей логического элемента в схемной форме ее типичная структура имеет вид, показанный на рис. 8.5.

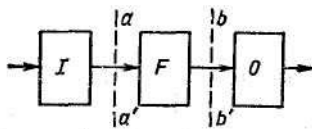


Рис. 8.5. Структура модели логического элемента

Входной блок I отражает входные характеристики схемы, выходной блок O — выходное сопротивление, блок F — функциональные преобразования сигналов и задержку их распространения. Передача сигналов через линии $a—a'$ и $b—b'$ происходит только в направлении слева направо.

Метод однонаправленных моделей применяется при выделении фрагментов в схеме в соответствии со следующими правилами:

- 1) границы фрагментов проходят через схемы однонаправленных моделей по линиям $a—a'$ или $b—b'$;
- 2) контур любой обратной связи должен полностью находиться в пределах одного фрагмента.

После выделения фрагментов необходимо их ранжирование — упорядочение в соответствии с последовательностью прохождения сигналов. Ранг 1 получают фрагменты, на входы которых поступают только внешние воздействия. Ранг 2 получают фрагменты, на входы которых поступают сигналы с выходов фрагментов ранга 1 и, возможно, также внешние воздействия, и т. д. Другими словами, ранг r присваивают фрагменту, на входы которого поступают сигналы с выходов только ранжированных фрагментов (внешние воздействия рассматриваются как сигналы с выходов фрагментов ранга 0), причем старший из рангов предшествующих фрагментов равен $r—1$.

После фрагментации и ранжирования выполняют раздельное численное интегрирование подсистем дифференциальных уравнений, относящихся к различным фрагментам в порядке увеличения их рангов. Интегрирование выполняют на всем заданном отрезке интегрирования $T_{\text{кон}}$. При интегрировании уравнений фрагмента с рангом r в качестве входных воздействий используют результаты интегрирования уравнений фрагментов с более низкими рангами.

Раздельное интегрирование позволяет организовать вычисления в каждом фрагменте с оптимальным для фрагмента значением шага, что может привести к значительной экономии вычислительных затрат. Однако метод однонаправленных моделей имеет ограниченное применение из-за необходимости соблюдения указанных правил фрагментации. Эти ограничения устраняются в методе РФС.

Метод РФС является итерационным методом раздельного интегрирования дифференциальных уравнений. Условие однонаправленности моделей снимается благодаря введению фрагментации схем с перекрытием, поясняемой рис. 8.6.

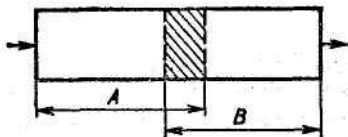


Рис. 8.6. Фрагментация схемы с перекрытием

Заштрихованный участок соответствует подсхеме, включаемой при раздельном интегрировании и в фрагмент A , и в фрагмент B . Чем шире зона перекрытия, тем точнее учитывается нагрузка для фрагмента A и точнее рассчитываются входные сигналы для фрагмента B . Если в схеме нет межфрагментных обратных связей, то достаточно ранжирования фрагментов и выполнения одной итерации пофрагментного интегрирования на интервале $[0, T_{\text{кон}}]$. При наличии межфрагментных обратных связей требуется выполнение нескольких итераций, с помощью которых форма сигналов на выходах фрагментов приближается к истинной (происходит релаксация формы сигналов).

Очевидно, что выполнение пофрагментного интегрирования на всем заданном интервале $[0, T_{\text{кон}}]$ может привести к излишне большому числу итераций и к колебаниям фазовых переменных в процессе релаксации с излишне большой амплитудой, что потребует значительных затрат машинного времени. Во избежание этих неприятностей можно использовать *метод прогнозируемой реакции*. В отличие от метода РФС здесь пофрагментное интегрирование производится на отрезках $[t_k, t_k + H_j]$, где H_j — шаг прогноза, соизмеримый с длительностью задержки прохождения сигнала по j -му контуру обратной связи. Значение H_j определяют автоматически аналогично тому, как рассчитываются значения шагов интегрирования.

Комбинированные методы и алгоритмы анализа. При решении консультационных задач анализа в САК получает достаточно широкое распространение временное комбинирование численных методов. Наиболее известны рассмотренные выше *алгоритмы ФНД* для численного интегрирования ОДУ, являющиеся алгоритмами комбинирования формул Гира. Другим примером временного комбинирования методов служат *циклические алгоритмы неявно-явного интегрирования ОДУ*. В этих алгоритмах циклически меняется формула интегрирования—следом за шагом неявного интегрирования следует шаг явного интегрирования. В бивовом алгоритме неявно-явного интегрирования используют формулы первого порядка точности — формулы Эйлера. Такой комбинированный алгоритм оказывается реализацией А-устойчивого метода второго порядка точности, повышение точности объясняется взаимной компенсацией

локальных методических погрешностей, допущенных на последовательных неявном и явном шагах. Следует отметить, что в качестве результатов интегрирования принимаются только результаты неявных шагов, поэтому в алгоритме комбинированного неявно-явного интегрирования устраняются ложные колебания, присущие наиболее известному методу второго порядка точности— методу трапеций.

Комбинирование неявных и явных формул интегрирования применяют для повышения эффективности решения нестационарных двумерных задач на микроуровне в рамках метода, называемого **методом переменных направлений**. Если на очередном временном шаге использовать неявные формулы аппроксимации производных по обеим пространственным координатам x_1 и x_2 , то потребуется решать систему конечных уравнений порядка pq , где p и q — уменьшенное на единицу число участков дискретизации по осям x_1 и x_2 . Однако можно циклически менять характер аппроксимации, например на i -м временном шаге для производных по x_1 использовать неявную, для производных по x_2 — явную формулы, а на $(i+1)$ -м шаге, наоборот, для производных по x_1 — явную, для производных по x_2 — неявную формулы. Тогда на каждом шаге по времени нужно решать q систем уравнений порядка p или p систем уравнений порядка q .

Учет латентности фрагментов. Локальные погрешности интегрирования зависят от значения шага интегрирования h и от характера переходных процессом. Если фазовые переменные претерпевают быстрые изменения, то погрешность не выше заданной обеспечивается при малых h . Если же фазовые переменные меняются медленно, то значения h при тех же погрешностях могут быть существенно больше. В сложных схемах консультируемых проблем, как правило, большинство фрагментов в любой момент времени относится к **неактивным (латентным)**, т. е. к таким, в **которых не происходит изменений фазовых переменных**, причем отрезки латентности $T_{\text{лат}}$ могут быть довольно продолжительными. В латентных фрагментах допустимо увеличивать шаг интегрирования вплоть до значения $T_{\text{лат}}$, что эквивалентно исключению уравнений фрагментов из процесса интегрирования на период их латентности. Такое исключение выполняется в алгоритмах учета латентности, относящихся к алгоритмам **событийного моделирования**. Основу этих алгоритмов составляет проверка условий латентности. Примером таких условий может служить

$$\|\Delta V\| < \varepsilon, \quad (8.102)$$

где ΔV — вектор изменений фазовых переменных; ε — малая положительная константа.

При проверке (8.102) у нелатентных фрагментов в вектор ΔV входят изменения как внешних, так и внутренних переменных фрагмента, и если условие (8.102) выполняется на протяжении нескольких шагов интегрирования подряд, то фрагмент включается в группу латентных. При проверке (8.102) у латентных фрагментов вектор ΔV состоит из изменений только внешних фазовых переменных, происшедших за время с начала латентности, и фрагмент исключается из числа латентных, если условие (8.102) нарушается.

Следует отметить, что в таких методах, как РФС, латентность учитывается естественным образом, в них не требуется специальных проверок статуса латентности фрагментов.

Адаптивное моделирование. Адаптивное моделирование — метод автоматического выбора подходящих моделей для фрагментов в процессе анализа сложной консультируемой проблемы. Метод адаптивного моделирования составляют способы решения следующих основных вопросов:

1. Фрагментация.
2. Критерий смены моделей.
3. Определение начальных значений переменных во включаемых моделях.

Рассмотрим один из возможных вариантов метода адаптивного моделирования. В нем используются модели трех уровней сложности: M_1 — простейшая макро модель фрагмента; M_2 — промежуточная по сложности макро модель, в которой блоки I и O (см. рис. 8.5) адекватно отражают входные и выходные цепи схемы, M_3 — полная модель фрагмента. Модели M_1 и M_2 имеют блочную структуру (см. рис. 8.5). Фрагментацию используют двоякую: во-первых, функциональную фрагментацию, выделяющую фрагменты, которым соответствуют сменяемые модели; во-вторых, диакоптическую фрагментацию, выделяющую раздельно интегрируемые фрагменты. Диакоптическая фрагментация выполняется с перекрытием, зону перекрытия составляют блоки I макромоделей. При функциональной фрагментации нужно стремиться к получению максимально возможного числа типовых фрагментов, под которыми понимаются фрагменты, для которых заранее разработаны и включены в библиотеку макро модели типов M_1 и M_2 вместе с рассчитанными для макромоделей типа M_1 областями адекватности (ОА). Диакоптический фрагмент есть функциональный фрагмент вместе с зоной перекрытия, которой соответствует блок I из модели схемы-нагрузки.

Предполагается, что анализ начинается при представлении всех фрагментов моделями M_1 . При дальнейшем анализе возможны замены

моделей М1 моделями М3. Такие замены выполняются для тех фрагментов, в которых происходит нарушение условий адекватности, обнаруживаемое при выходе вектора внешних переменных фрагмента за пределы ОА. Одновременно с включением полной модели М3 для некоторого фрагмента А происходит замена блоков О в моделях схем, являющихся источниками сигналов для фрагмента А, и замена блоков I в моделях схем, являющихся нагрузками для фрагмента А. Эта замена — включение блоков из моделей М2 вместо блоков из моделей М1,

Включение новой модели сопровождается появлением новых фазовых переменных. Очевидно, что для продолжения анализа должны быть определены значения этих переменных в момент $t_{\text{вкл}}$ включения модели. Расчет этих значений целесообразно выполнять повторным интегрированием подсистемы дифференциальных уравнений включаемой модели на интервале $[0, t_{\text{вкл}}]$.

8.6. Методы анализа логических и функциональных схем консультируемых проблем

Методы решения логических уравнений. Анализ переходных процессов в логических схемах консультируемых проблем выполняют с помощью асинхронных моделей (6.93), т. е. на основе *асинхронного моделирования*. К началу очередного такта t_i известны значения векторов внутренних $\mathbf{V}_i = (v_{1i}, v_{2i}, \dots, v_{ni})$ и входных \mathbf{U}_i переменных. Подставляя \mathbf{V}_i и \mathbf{U}_i в правую часть выражений (6.94), получаем новые значения v'_{ki} , которые примут внутренние переменные в моменты времени $t_i + \tau_k$, где τ_k — внутренняя задержка распространения сигнала v_k в соответствующем элементе схемы. Далее переходим к следующему такту, в котором вычисления по (6.94) повторяются со значениями векторов \mathbf{V} и \mathbf{U} , соответствующими новому моменту времени $t_{i+1} = t_i + 1$ (напомним, что время измеряется в количестве тактов). Асинхронное моделирование называют **потактовым**.

Если длительность такта превышает задержку в некоторых элементах, т. е. для некоторых k имеем $\tau_k \approx 0$, то в модели (6.94) последовательностной схемы появляются отдельные неявные относительно v_k выражения, а это приводит к необходимости решать подсистемы логических уравнений в пределах каждого такта. Если задержки не учитывать во всех элементах, то имеем синхронную модель (6.93), с помощью которой анализируются установившиеся состояния в схеме и могут определяться статические и динамические риски сбоя. Синхронная модель — это система логических уравнений

$$\mathbf{V} = \mathbf{F}(\mathbf{V}, \mathbf{U}). \quad (8.103)$$

Для *синхронного моделирования* (решения систем логических уравнений) используются итерационные методы *простой итерации и Зейделя*.

Алгоритм *метода простой итерации* при решении (8.103) совпадает с алгоритмом асинхронного моделирования при $\tau_k = 1$. На первой итерации (такте) выбирается начальное приближение V_0 и подставляется в правую часть (8.103), при этом определяется новое приближение V_1 . На второй итерации рассчитывается V_2 при подстановке V_1 в правую часть (8.103) и т. д. Условием прекращения итераций является или совпадение результатов двух последних итераций $V_i = V_{i-1}$, где i — номер итерации, или выполнение равенства $i = i_{\max}$, где i_{\max} — предельно допустимое число итераций. В последнем случае делается заключение о несходимости вычислительного процесса из-за генерирования колебаний в моделируемой схеме.

Итерации по *методу Зейделя* отличаются от простых итераций тем, что новое значение k -го элемента вектора V_i сразу же после его вычисления на i -й итерации заменяет старое значение и используется для вычисления новых значений следующих элементов вектора V_i на той же i -й итерации, т. е.

$$v_{k,i} = F(v_{1,i}, v_{2,i}, \dots, v_{k-1,i}, v_{k,i-1}, \dots, v_{n,i-1}).$$

В правильно составленной комбинационной схеме каждому входному вектору U соответствует одно определенное решение V^* системы логических уравнений (8.103). Корень V^* может быть найден при произвольном начальном приближении как методом простой итерации, так и методом Зейделя. Однако в последовательностной схеме система уравнений (8.103) имеет несколько решений, соответствующих различным устойчивым состояниям схемы. Обычно требуется найти то состояние, в которое переходит схема из заданного исходного состояния при заданном входном воздействии. Пусть заданному исходному состоянию соответствует значение $V_{\text{исх}}$ вектора V . Тогда для решения сформулированной задачи рекомендуется выбрать начальное приближение $V_0 = V_{\text{исх}}$ и, кроме того, в методе Зейделя необходимо привести в соответствие последовательности вычислений по уравнениям модели и прохождения сигналов в схеме. Последнее условие выполняется с помощью алгоритмов ранжирования, что способствует определению искомого корня за малое число итераций. Например, для комбинационной схемы решение получается за одну итерацию.

Ранжирование логических схем и уравнений. Алгоритмы ранжирования комбинационных схем аналогичны алгоритмам ранжирования электронных схем в методе однонаправленных моделей.

Ранг 0 присваивается входным цепям, по которым поступают сигналы на схему извне. Ранг r присваивается элементам, все входы которых ранжированы и старший из рангов равен $r - 1$. Ранг r присваивается также выходам элементов ранга r . После этого логические уравнения упорядочиваются по значениям рангов соответствующих элементов схемы.

В последовательных схемах ранжирование может быть выполнено таким же образом после преобразования схемы в комбинационную с помощью разрывов обратных связей. Ранжирование последовательностей схемы начинается так же, как и комбинационной, но проранжировать таким образом удастся только часть элементов. Далее нужно выполнить разрывы контуров обратной связи. Контур обратной связи ищутся среди элементов, которые проранжировать не удалось. Выбирается любой из этих элементов, например A , к нему добавляется элемент B , выход которого связан с не имеющим ранга входом элемента A . Очевидно, что такой элемент B имеется, иначе элемент A был бы проранжирован. Далее таким же образом присоединяется элемент C к элементу B и т. д., пока в цепочке из элементов $A, B, C...$ не встретится повторно какой-либо из элементов, уже включенных в цепочку. Эта ситуация означает обнаружение контура обратной связи. Далее одна из связей в контуре обратной связи обрывается. Входы элементов от разорванных связей, называемые псевдовходами, получают ранг 0. После этого ранжирование возобновляется. Операции выделения контуров и ранжирования повторяются до тех пор, пока в схеме не останется непроранжированных элементов.

Повышение эффективности моделирования логических и функциональных схем. Для повышения эффективности решения уравнений методом Зейделя целесообразно использовать *диакоптический подход*, в рамках которого итерации выполняются отдельно по фрагментам логической схемы. Введем следующие понятия: *составной элемент* — множество контуров обратной связи, имеющих попарно общие связи; *фрагмент логической схемы* — составной элемент или комбинационная схема, состоящая из взаимосвязанных логических элементов, не вошедших в составные элементы.

Очевидно, что после обнаружения контуров обратной связи нетрудно найти составные элементы и, следовательно, представить моделируемую схему, состоящую из фрагментов. В такой схеме нет межфрагментных обратных связей и, следовательно, система логических уравнений может быть решена путем однократного

обращения к проранжированным подсистемам уравнений. При этом решение подсистем комбинационных фрагментов будет выполнено за одну итерацию Зейделя, а решение подсистем составных элементов — за число итераций, определяемое особенностями данного составного элемента и способом разрыва в нем обратных связей. Это число итераций можно уменьшить, если осуществить разрывы всех контуров обратной связи в минимальном количестве межэлементных связей. Пофрагментное выполнение итераций существенно сокращает общие затраты машинного времени.

Другим способом повышения эффективности является *параллельное моделирование*, основанное на том, что для представления логической переменной достаточно k разрядов, где $k=1$ в двузначном алфавите и $k=2$ в трехзначном. Тогда моделирование одной и той же схемы можно выполнять одновременно для $m=s/k$ различных наборов входных сигналов, где s — количество разрядов в разрядной сетке ЭВМ. Подобное параллельное моделирование эффективно используется при синтезе тестов для проверки логических схем, где требуется определить реакцию схемы на большое количество входных тестовых наборов.

Наиболее общим направлением повышения эффективности математического обеспечения как синхронного, так и асинхронного моделирования является *учет событийности*.

При *анализе логических и функциональных схем событием называют изменение состояния* любого элемента или, что то же самое, *изменение значения* любой переменной состояния.

В процессе событийного моделирования вычисления производят только по уравнениям активных элементов, т. е. таких элементов, на входах которых на данном такте или итерации произошли события.

Рассмотрим алгоритм асинхронного событийного моделирования. В алгоритме используются *списки текущих и будущих событий*. Все события привязаны к *моментам дискретного модельного времени*. Ссылки на события, происходящие в текущий момент, находятся в *списке текущих событий*, а ссылки на те события, наступление которых можно предвидеть, помещаются в *список будущих событий*. Моделирование текущих событий означает обращение к моделям соответствующих элементов, после чего становятся известными изменяющиеся выходные переменные и задержки в их изменении по отношению к текущему моменту модельного времени. *Эти изменения суть события, и сведения о них заносятся в список будущих событий*. После того как список текущих событий полностью

отработан, модельное время увеличивается и соответствующая часть списка будущих событий становится списком текущих событий. Так как в сложных схемах на каждом такте активными являются не более нескольких процентов элементов, то учет событийности позволяет существенно уменьшить затраты машинного времени.

В маршрутах консультирования сложных консультируемых проблем к числу основных консультационных процедур относятся ***верификация логических и функциональных схем, синтез и анализ тестов***. В этих процедурах требуется многократное выполнение моделирования логических схем. Однако высокая размерность задач логического моделирования (сложные консультируемые проблемы насчитывают десятки—сотни тысяч элементов) существенно ограничивает возможности многовариантного анализа. Преодоление затруднений, обусловливаемых чрезмерной трудоемкостью вычислений, происходит в двух направлениях. Первое из них основано на использовании общих положений блочно-иерархического подхода и выражается в переходе к представлениям подуровня регистровых передач, рассмотренным в п.6.11. Второе направление основано на применении специализированных вычислительных средств логического моделирования, называемых спецпроцессорами или машинами логического моделирования (МЛМ).

Логико-электрическое моделирование. Логико-электрическое моделирование представляет собой разновидность многоуровневого моделирования, при котором в анализируемой схеме одновременно используются модели и методы, относящиеся как к схемотехническому, так и к функционально-логическому иерархическому уровню.

В анализируемой схеме выделяются подсхемы, подлежащие анализу с помощью логических и электрических моделей. ***Сопряжение моделей подсхем осуществляется с помощью специальных переходных моделей элементов и алгоритмов синхронизации событий в логической и электрической частях. Переходные модели служат для отображения процессов в элементах с преобразованием аналоговых переменных в логические и наоборот.***

Логико-электрическое моделирование является незаменимым способом анализа сложных цифроаналоговых схем, поскольку позволяет резко сократить размерность задач, благодаря использованию логических моделей для цифровой части при сохранении необходимой точности анализа, благодаря использованию электрических моделей для аналоговой части схемы.

8.7. Методы многовариантного анализа

Анализ чувствительности. Анализ чувствительности входит составной частью в алгоритмы решения многих консультационных задач, в частности в алгоритмы оптимизации параметров консультируемых проблем градиентными методами. Для анализа чувствительности задаются ММ консультируемой проблемы и вектор тех внутренних и внешних параметров \mathbf{X} , влияние которых на вектор выходных параметров \mathbf{Y} требуется определить.

В большинстве случаев анализ чувствительности выполняют на основе численного дифференцирования, при котором поочередно задают приращения Δx_i элементам вектора \mathbf{X} и определяют получающиеся при этом изменения Δy_j выходных параметров. Тогда **абсолютный коэффициент влияния** (**коэффициент чувствительности**) i -го элемента x_i вектора \mathbf{X} на j -й выходной параметр y_j определяется по формуле

$$a_{ji} = dy_j/dx_i \approx \Delta y_j / \Delta x_i;$$

относительный коэффициент влияния

$$b_{ji} = a_{ji} x_{i\text{ном}} / y_{j\text{ном}}$$

где $x_{i\text{ном}}$ и $y_{j\text{ном}}$ — номинальные значения параметров x_i и y_j соответственно.

Такой метод анализа чувствительности называют **методом приращений**. Если n — размерность вектора \mathbf{X} , то в методе приращений требуется $n+1$ раз выполнить одновариантный анализ. Сравнительно большие трудоемкость и погрешности вычислений, присущие численному дифференцированию, относятся к недостаткам этого метода, а универсальность метода — к его преимуществам.

Повысить точность вычислений можно, если, определяя a_{ji} , производить одновариантный анализ при значениях параметра x_j , равных $x_{j\text{ном}} + \Delta x_j$ и $x_{j\text{ном}} - \Delta x_j$ (остальные внутренние параметры при этом сохраняют номинальные значения). Тогда a_{ji} есть отношение разности значений y_j , полученных в этих двух вариантах, к $2\Delta x_j$. Однако в этом методе увеличивается трудоемкость — нужно $2n$ раз выполнить одновариантный анализ.

Применяют также методы, имеющие менее универсальный характер, например методы анализа чувствительности функционалов зависимостей $\mathbf{V}(t)$, получающихся при интегрировании систем ОДУ. В этих методах или сокращается число вариантов интегрирования уравнений, или упрощается система интегрируемых уравнений.

Статистический анализ. Статистический анализ имеет целью получение информации о распределении вектора выходных

параметров Y при заданном законе распределения случайного вектора X внутренних параметров консультируемой проблемы.

Основным методом статического анализа в САК является метод *статистических испытаний (метод Монте-Карло)*. Каждое k -е статистическое испытание заключается в присвоении элементам x_i вектора X случайных значений x_{ik} и расчете вектора выходных параметров Y_k , с помощью одновариантного анализа. После выполнения запланированного числа N статистических испытаний их результаты Y_k обрабатываются с целью оценки числовых характеристик распределений выходных параметров.

В основе алгоритма задания случайных значений параметрам лежит формула

$$\xi = \int_{-\infty}^{x_{ik}} p(x_i) dx_i,$$

с помощью которой случайное значение ξ величины, равномерно распределенной в интервале $[0, 1]$, преобразуется в случайное значение x_{ik} величины x_i , имеющей плотность распределения $p(x_i)$. Для выработки значений ξ используют стандартные подпрограммы, имеющиеся в программном обеспечении любой универсальной ЭВМ. Выработанное значение ξ интерпретируется как значение $F(x_{ik})$ функции распределения величины x_i , так как

$$F(x_{ik}) = \int_{-\infty}^{x_{ik}} p(x_i) dx_i.$$

Функция распределения $F(x_i)$ монотонная, обычно представлена в памяти ЭВМ в табличной форме, и, следовательно, преобразование ξ в x_{ik} сводится к поиску нужного интервала таблицы и определению результата с помощью интерполяции.

Вместо значений ξ равномерно распределенной случайной величины можно использовать значения u_k нормированного нормально распределенного вектора $U = (u_1, u_2, \dots, u_n)$ и в алгоритме осуществлять преобразование u_{ik} в x_{ik} . Но элементы вектора U некоррелированы, поэтому если требуется выработка значений коррелированных случайных величин x_i , то вводятся промежуточный вектор коррелированных нормальных величин Z и матрица A преобразования U в Z . Тогда алгоритм задания случайных значений параметрам x_i сначала вырабатывает n случайных значений u_{ik} , затем преобразует их в вектор $Z_k = AU_k$ и далее значения этого вектора — в искомые значения x_{ik} коррелированных параметров элементов.

Матрица преобразования \mathbf{A} , так же как и таблица преобразования \mathbf{Z} (или \mathbf{U}) в \mathbf{X} , определяется на основе обработки результатов предварительно выполненных измерений параметров на партии тестовых образцов либо проблем данного или аналогичного типа.

Результаты статистических испытаний \mathbf{Y}_k используются для *построения гистограмм, подсчета математических ожиданий и дисперсий выходных параметров*. Можно рассчитать также *коэффициенты корреляции* между выходными y_j и внутренними x_i параметрами, которые используются для определения *коэффициентов регрессии* y_j на x_i . Поскольку *относительные коэффициенты регрессии являются аналогами коэффициентов влияния* x_i на y_j , регрессионный анализ, совмещаемый со статистическим анализом, следует рассматривать как возможный *подход к анализу чувствительности*.

Точность и трудоемкость статистических испытаний зависят от их числа N . Для получения большинства интересующих консультанта результатов статистического анализа с приемлемой погрешностью требуется выбирать $N=50-200$. Однако получение некоторых результатов (таких, как вероятность выхода годных изделий при значениях этой вероятности, близких к единице или нулю) с приемлемой точностью требует значительно большего числа испытаний. Отсюда следует вывод о значительной трудоемкости статистического анализа. Именно по этой причине статистический анализ проводят лишь на заключительных итерациях процесса консультирования.

Разработка тестов консультируемых проблем. Консультирование проблем включает в себя разработку тестов, с помощью которых проверяется эффективность (достоверность) сформированных рекомендаций по решению задач консультируемой проблемы.

Назовем тестируемый объект *блоком*. *Входное воздействие* обозначим вектором $\mathbf{X}=(x_1, x_2, \dots, x_n)$, а *выходную реакцию* — вектором $\mathbf{Y}=(y_1, y_2, \dots, y_m)$, где x_i — булева переменная на i -м входе; y_j — то же на j -м выходе. Определенному значению X_k вектора \mathbf{X} в блоке эффективных рекомендаций соответствует значение Y_k вектора \mathbf{Y} . Пару $\{X_k, Y_k\}$ будем называть *элементарной проверкой*.

С помощью одной элементарной проверки нельзя выявить все неэффективных рекомендации даже ограниченного класса и тем более нельзя локализовать неэффективную рекомендацию. *Поэтому тест представляет собой множество элементарных проверок*. Число элементарных проверок в тесте называют его *длиной*. Тест, предназначенный только для установления факта наличия блока

неэффективных рекомендаций, называют **контролирующим (проверяющим)**, а тест, с помощью которого дополнительно устанавливается неэффективная рекомендация в блоке, — **диагностическим**. Тест, который выявляет все неэффективные рекомендации заданного класса, будем называть **полным**, а тест, из которого нельзя исключить ни одну элементарную рекомендацию без изменения его полноты, — **неизбыточным**.

Очевидно, что тест, в который включены всевозможные комбинации значений входных сигналов, является полным. Однако громадное число таких комбинаций, равное 2^n в комбинационном блоке и 2^{n+q} в последовательностном, делает нереальным применение подобных полных тестов (здесь q — число элементов памяти в блоке). Поэтому возникают **задачи синтеза и анализа тестов. При синтезе генерируются входные наборы X_k , а при анализе устанавливаются характеристики теста.**

Обычно тесты рассчитывают на выявление только **одиночных константных неэффективных рекомендаций**. Это класс устойчивых неэффективных рекомендаций, возникающих поодиночке в единственном элементе блока. К ним относятся неэффективные рекомендации «константный нуль», «константная единица», связанные с наличием постоянного низкого или высокого уровня напряжения, и «инверсная неисправность», связанная с появлением непредусмотренного инвертирования сигнала.

Для синтеза тестов применяют **вероятностные и детерминированные методы**. В вероятностных методах наборы генерируются с помощью датчиков случайных чисел. Основные затраты машинного времени приходятся при этом на анализ проверяющих возможностей генерируемых наборов. Анализ каждого набора состоит в расчете реакции на воздействие X_k как эффективной рекомендации, так и всех ее возможных разновидностей. Если блок состоит из N элементов, то имеем $3N$ таких разновидностей и общее число вариантов моделирования блока окажется пропорциональным произведению sN , где s — число рекомендаций входных наборов. Практика показывает, что при заданной полноте теста s зависит от N и в результате затраты машинного времени T_m оказываются пропорциональными N_α , где $\alpha=2\div 3$.

Большие значения T_m обуславливают применение для анализа тестов наиболее экономичных методов моделирования логических и функциональных схем. Обычно используют **параллельное синхронное трехзначное моделирование**. Трехзначный алфавит целесообразен для отбраковки входных векторов X_k , приводящих к состязаниям сигналов

в блоке, из-за которых результаты применения теста могут стать неопределенными.

В **детерминированных методах синтеза тестов** для каждой неэффективной рекомендации из заданного списка подбирают свой входной набор. Одним из наиболее распространенных алгоритмов, реализующих детерминированный подход к синтезу тестов, является **алгоритм Рота**. В соответствии с этим алгоритмом для очередной неэффективной рекомендации, связанной с рекомендацией Р тестируемой схемы, подбираются входные для Р воздействия, такие что выходы эффективной рекомендации и неэффективной рекомендации Р будут иметь неодинаковые значения. После этого ищется путь передачи информации о наличии неэффективной рекомендации к выходам схемы, которые доступны для наблюдения. Поиск пути заключается в установлении цепочки рекомендаций между рекомендацией Р и выходом схемы и в определении таких значений входов рекомендаций цепочки, при которых неэффективная рекомендация транспортируется к выходу. Очевидно, что не любая цепочка может служить путем транспортировки, поскольку требуемые для транспортировки значения входов могут оказаться противоречивыми. Поэтому в алгоритме Рота осуществляется последовательное наращивание цепочки с проверкой непротиворечивости условий транспортировки для ранее и вновь включаемых рекомендаций. После установления пути требуемые значения входов у рекомендаций цепочки пересчитываются в значения тех входов тестируемой схемы, которые доступны для управления. Эти значения входов и представляют собой очередной искомый набор теста.

В программах синтеза тестов обычно используют комбинирование вероятностных и детерминированных методов. С помощью первых генерируется и анализируется некоторое заранее заданное число наборов. Если полнота теста после этого недостаточна, то выбирают неэффективные рекомендации, не выявляемые уже полученными наборами, и для этих неэффективных рекомендаций ищут входные наборы детерминированным методом.

8.8. Анализ чувствительности

Остановимся более детально на анализе чувствительности. С помощью аппарата теории чувствительности можно оценить направление и скорость изменения характеристик консультируемой проблемы при условии, что вектор переменных состояния

консультируемой проблемы смещен относительно своего номинального значения. Это смещение может вызываться, например, влиянием внешних факторов на параметры компонентов консультируемой проблемы. Теория чувствительности находит широкое применение в процедурах параметрической оптимизации (при определении составляющих градиента и гессовой матрицы целевой функции), анализа допусков, расчета характеристик устройств, соответствующих режиму «худшего случая» и т. д.

Формальное определение чувствительности обычно соответствует выражению

$$S_{x_i}^{\gamma} = \frac{\partial (\ln \gamma)}{\partial (\ln x_i)}, \text{ или } S_{x_i}^{\gamma} = \frac{x_i}{\gamma} \frac{\partial \gamma}{\partial x_i}, \quad (8.104)$$

где γ — выходной параметр консультируемой проблемы (переменная, параметр качества и др.); x_i ($i = 1, 2, \dots, n$) — варьируемые параметры компонентов консультируемой проблемы.

Частную производную $\partial \gamma / \partial x_i$ называют **функцией чувствительности**, или **коэффициентом влияния** параметра x_i на выходную характеристику схемы γ .

Для учета влияния на выходной параметр консультируемой проблемы одновременных изменений нескольких переменных вводят понятие чувствительности по многим переменным, соответствующее градиенту:

$$\nabla \gamma = \left[\frac{\partial \gamma}{\partial x_1}, \frac{\partial \gamma}{\partial x_2}, \dots, \frac{\partial \gamma}{\partial x_n} \right]^t. \quad (8.105)$$

Тогда относительное изменение выходного параметра консультируемой проблемы, обусловленное малыми и независимыми отклонениями параметров его компонентов, в первом приближении можно определить как

$$\delta \gamma = \frac{\Delta \gamma}{\gamma} \approx (\nabla \gamma)^t \delta x = \sum_{i=1}^n \frac{\partial \gamma}{\partial x_i} \frac{\Delta x_i}{x_i}, \quad (8.106)$$

где δx — вектор относительных изменений варьируемых параметров, i -я составляющая которого $\delta x_i = \Delta x_i / x_i$.

Исследование чувствительности выполняется, как правило, в сочетании с каким-то основным видом анализа (статического, динамического, частотного), который оказывает существенное влияние на выбор конкретной методики расчета функций чувствительности.

Статические характеристики консультируемой проблемы определяются для заданного режима функционирования,

динамические — при анализе скорости изменения характеристик консультируемой проблемы во временной области, а *частотные характеристики* определяют при расчете зависимости требуемых передаточных функций консультируемой проблемы от частоты. Для описания и расчета каждого из отмеченных видов анализа применяется специфический математический аппарат, учет особенностей которого необходим при анализе чувствительности.

При решении всех рассмотренных выше типов задач моделирования чувствительности применяют различные модификации методов присоединенных консультируемых проблем, моделей чувствительности и вариационного метода. В общем случае консультируемая проблема описывается совокупностью уравнений независимых сечений или контуров, в которые частично или полностью введены компонентные соотношения:

$$\begin{cases} \eta - \varphi_1(\xi, x) = 0; \\ \xi - \varphi_2(\eta, x) = 0, \end{cases} \quad (8.107)$$

где η, ξ — вектор соответственно последовательных и параллельных переменных модели консультируемой проблемы; x — вектор варьируемых параметров.

Модель чувствительности для консультируемой проблемы получают в результате дифференцирования выражения (8.107) по x

$$\left. \begin{aligned} \frac{\partial \eta}{\partial x} - \frac{\partial \varphi_1}{\partial \xi} \frac{\partial \xi}{\partial x} &= \frac{\partial \varphi_1}{\partial x}; \\ \frac{\partial \xi}{\partial x} - \frac{\partial \varphi_2}{\partial \eta} \frac{\partial \eta}{\partial x} &= \frac{\partial \varphi_2}{\partial x}. \end{aligned} \right\} \quad (8.108)$$

Решение для уравнения (8.108) эквивалентно решению m систем линейных уравнений типа

$$[A] \begin{bmatrix} \eta_j^i \\ \xi_j^i \end{bmatrix} = b_j, \quad j = 1, 2, \dots, m, \quad (8.109)$$

где

$$\eta_j^i = \frac{\partial \eta}{\partial x_j}, \quad \xi_j^i = \frac{\partial \xi}{\partial x_j}, \quad b_j = \left[\frac{\partial \varphi_1}{\partial x_j}, \frac{\partial \varphi_2}{\partial x_j} \right]^t.$$

Исходная модель консультируемой проблемы согласно выражению (8.107) при решении также подвергается линеаризации и алгебраизации, в результате чего получаем математическую модель:

$$\left. \begin{aligned} \eta - \frac{\partial \varphi_1}{\partial \xi} \xi &= W_1; \\ \xi - \frac{\partial \varphi_2}{\partial \eta} \eta &= W_2, \end{aligned} \right\} \quad (8.110)$$

где W_1 и W_2 — векторы задающих источников, учитывающие вклады процедур линеаризации и алгебраизации, или

$$[A] \begin{bmatrix} \eta \\ \xi \end{bmatrix} = W. \quad (8.111)$$

Сравнив выражения (8.111) и (8.109), приходим к заключению, что модель чувствительности для консультируемой проблемы характеризует другую консультируемую проблему, топологически эквивалентную исходной, переменные которой (последовательные и параллельные) соответствуют производным переменных исходной консультируемой проблеме по варьируемому параметру x_j , при этом устраняются все задающие источники, воздействующие на исходную модель консультируемой проблемы; нелинейные компоненты исходной консультируемой проблемы заменяются линейными эквивалентами; компоненты консультируемой проблемы, характеристики которых зависят от x_j , дополняются параллельно включенным источникам $\partial\varphi_1/\partial x_j$ (если это Y-компонент) или последовательно включенным источникам $\partial\varphi_2/\partial x_j$ (если это Z-компонент).

Математические модели исходной консультируемой проблемы согласно выражению (8.111) и модели чувствительности согласно выражению (8.109) представляют собой системы линейных алгебраических уравнений (для каждой временной точки), в которых различны только значения векторов правых частей.

В соответствии с процедурой моделей чувствительности для определения скорости изменения n выходных характеристик консультируемой проблемы (типа последовательных или параллельных переменных) по m варьируемым параметрам на всем рассматриваемом интервале времени $[t_0, t_k]$ необходимо решить на этом же участке времени $(m+1)$ систему уравнений: нелинейную систему дифференциально-алгебраических уравнений исходной консультируемой проблемы и m линейных дифференциально-алгебраических систем уравнений соответствующих моделей чувствительности. По существу, анализируется одна модель чувствительности, запрашиваемая различными входными источниками

$$\eta_k = \frac{\partial\varphi_{1k}}{\partial x_j} \quad \text{и} \quad \xi_s = \frac{\partial\varphi_{2s}}{\partial x_j},$$

значения которых определяются на основании анализа исходной консультируемой проблемы.

Если консультируемая проблема описывается некоторой интегральной характеристикой

$$\gamma(x) = \int_0^T e(\eta, \xi, x, t) dt, \quad (8.112)$$

то отдельные составляющие градиента такой характеристики представляются в виде

$$\frac{\partial \gamma(x)}{\partial x_j} = \int_0^T \left(\frac{\partial e}{\partial \eta} \frac{\partial \eta}{\partial x_j} + \frac{\partial e}{\partial \xi} \frac{\partial \xi}{\partial x_j} + \frac{\partial e}{\partial x_j} \right) dt, \quad (8.113)$$

где составляющие $\partial \eta / \partial x_j$ и $\partial \xi / \partial x_j$ определяются как реакции соответствующей модели чувствительности.

В результате однократного анализа модели чувствительности формируется соответствующая строка матрицы первых производных

$$J = \left[\begin{array}{c} \dots \\ \dots \frac{\partial \eta_1}{\partial x_j} \dots \\ \vdots \\ \dots \frac{\partial \xi_1}{\partial x_j} \dots \\ \dots \end{array} \right] n. \quad (8.114)$$

$\underbrace{\hspace{10em}}_m$

Если необходимо одновременно выделить все функции чувствительности одной переменной консультируемой проблемы по всем варьируемым параметрам, например выделить строку матрицы J согласно выражению (8.114)

$$\left[\frac{\partial \eta_k}{\partial x_1}, \frac{\partial \eta_k}{\partial x_2}, \dots, \frac{\partial \eta_k}{\partial x_m} \right],$$

то применяют метод присоединенных консультируемых проблем.

Из уравнения (8.109) можно найти полную матрицу, определяемую выражением (8.114):

$$J = \begin{bmatrix} \eta' \\ \xi' \end{bmatrix} = [A]^{-1} B, \quad (8.115)$$

где

$$B = [b_1, b_2, \dots, b_m]^t.$$

Умножая выражение (8.115) на квадратную единичную матрицу 1_k с k -м столбцом, содержащим единичные элементы, получаем

$$y'_k = 1_k^t J = 1_k^t A^{-1} B = [(A^t)^{-1} 1_k]^t B = \hat{y}_k^t B = \tilde{y}_k^t y_k, \quad (8.116)$$

где y'_k соответствует η'_k или ξ'_k .

Вектор

$$\hat{y}_k^t = [\eta_k \hat{\xi}_k]$$

является решением уравнений присоединенной консультируемой проблемы типа

$$[A^t] \begin{bmatrix} \hat{\eta}_k \\ \hat{\xi}_k \end{bmatrix} = l_k, \quad (8.117)$$

а

$$y_k = \begin{bmatrix} \eta_k \\ \xi_k \end{bmatrix}$$

— вектор решения уравнения (8.111), так как вектор B из выражения (8.109) соответствует совокупности значений переменных, полученных при анализе уравнений исходной консультируемой проблемы. В соответствии с процедурой присоединенных консультируемых проблем для определения функций чувствительности n выходных характеристик консультируемой проблемы по m варьируемым параметрам в некоторый момент времени $t=t_i$ необходимо, кроме расчета реакций исходной консультируемой проблемы на участке $[t_0, t_i]$, где t_0 — нижняя граница рассматриваемого интервала времени, проинтегрировать в обратном масштабе времени в диапазоне $[t_i, t_0]$ уравнения присоединенных консультируемых проблем. При этом уравнения каждой из присоединенных консультируемых проблем линейны (их отдельные компоненты определяются значениями составляющих вектора переменных исходной консультируемой проблемы в соответствующий момент времени) и отличаются лишь векторами правых частей. По существу, анализируется одна присоединенная консультируемая проблема, питающаяся от различных источников сигнала единичной величины.

Уравнения присоединенной консультируемой проблемы (8.115) строят непосредственно по уравнениям исходной консультируемой проблемы (8.111).

Если в модели исходной консультируемой проблемы используется матрица

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{bmatrix}, \quad (8.118)$$

то матрица модели присоединенной консультируемой проблемы

$$A^t = \begin{bmatrix} A_{11}^t & -A_{21}^t \\ -A_{12}^t & A_{22}^t \end{bmatrix} = \begin{bmatrix} U_{11}^t & 0 \\ -U_{12}^t & U_{22}^t \end{bmatrix} \begin{bmatrix} L_{11}^t & -L_{21}^t \\ 0 & L_{22}^t \end{bmatrix}. \quad (8.119)$$

Для однородных моделей консультируемых проблем это преобразование существенно упрощается, при этом

$$A = A_{11} \text{ и } A^t = A_{11}^t.$$

Метод присоединенной консультируемой проблемы является разновидностью вариационного метода, в котором используются сопряженные уравнения, решаемые в обратном масштабе времени. Сам вариационный метод удобно применять, когда определяются составляющие градиента некоторой интегральной характеристики, сложно зависящей от переменных консультируемых проблем. При решении задач параметрической чувствительности наиболее эффективным оказывается метод моделей чувствительности.

Остановимся несколько подробнее на вопросах одновременного интегрирования математических моделей исследуемой консультируемой проблемы и всех необходимых моделей чувствительности. Наиболее очевидный способ решения поставленной задачи заключается в раздельном интегрировании каждой из $(m + 1)$ -й систем дифференциально-алгебраических уравнений. При этом уравнения анализируемой консультируемой проблемы должны интегрироваться с опережением по времени хода решения остальных систем уравнений, поскольку компоненты моделей чувствительности зависят от текущего значения вектора переменных исходной консультируемой проблемы. Такой подход характеризуется значительными суммарными затратами, так как при выполнении вычислений совершенно не учитывается родственность исследуемых моделей.

В качестве альтернативы рассматривается методика, в соответствии с которой все исследуемые консультируемые проблемы интегрируются с общим шагом. При интегрировании математических моделей исходной консультируемой проблемы (желательно, дискретного принципа действия) и всех моделей чувствительности с общим временным шагом коэффициенты алгебраизации типа приведенных в уравнениях (8.60) и (8.65) будут одинаковыми для всех моделей и, следовательно, уравнения отдельных моделей вида (8.109) будут отличаться только значениями векторов правых частей. Тогда после получения решения уравнений всех исследуемых моделей в некоторый момент времени t_n необходимо для каждой из них оценить локальные погрешности ϵ_t для различных методов интегрирования и определить желательные для данной модели значения нового временного шага h_i и порядка метода интегрирования. В качестве общего нового шага интегрирования $(m+1)$ -й систем выбирается наименьший из желательных шагов

$$h = \min |h_i|, \quad i = 1, 2, \dots, m + 1.$$

Он же определит и используемый порядок метода интегрирования. Последовательность решения уравнений отдельных моделей чувствительности целесообразно организовать так, чтобы сначала решались уравнения тех моделей, которые имели меньшую величину желательного шага интегрирования.

Для консультируемых проблем непрерывного принципа действия временные характеристики, как правило, отличаются по форме от своих функций чувствительности. При интегрировании уравнений консультируемой проблемы и всех ее моделей с общим временным шагом характеристики всех исследуемых консультируемых проблем просчитываются с дискретностью, необходимой для точного моделирования той системы уравнений, которая описывает самые активные процессы на данном временном участке. Нерациональность такого положения очевидна.

Рассмотрим еще один из возможных подходов, названный **методом динамического группирования моделей**, при котором несколько моделей чувствительности объединяются в отдельные группы без существенной коррекции желательных для них значений текущего временного шага, что позволяет не проводить для каждой из них переоценку текущих значений нелинейных компонентов консультируемой проблемы.

В настоящее время созданы высокоэффективные машинные процедуры расчета параметрической чувствительности, которые гармонично сочетаются с современными методами моделирования консультируемых проблем на ЭВМ.

Пользователь САК может управлять процессом анализа чувствительности при изменении значения констант настройки, к которым относятся следующие.

Величина *XMAXSM* (аналог *XMAX*), или начальное значение максимальной величины составляющих вектора переменных каждой модели чувствительности.

Порядок *ORDERSM* (аналог *ORDER*) метода интегрирования, применяемого при решении уравнений моделей чувствительности.

Допустимое значение *TRUNCSM* (аналог *TRUNC*) локальной погрешности метода интегрирования при решении уравнений моделей чувствительности.

Параметр *REJCSM* (аналог *REJC*), используемый при оценке необходимости отбрасывания текущего временного шага для отдельных моделей чувствительности.

8.9. Анализ стационарных режимов

Анализ установившихся периодических процессов в нелинейных консультируемых проблемах является одной из важнейших задач консультирования из-за широкого применения в технике устройств периодического принципа действия. Вследствие значительного времени переходной реакции таких консультируемых проблемах при непосредственном применении эффективных методов численного интегрирования, рассмотренных в п. 8.4, вычислительные затраты могут быть большими.

В подсистемах консультационного анализа САК для расчета установившихся периодических процессов рекомендуется применять специальные методы, среди которых выделим **квазиньютоновские, градиентный и экстраполяционный методы**. Каждый из указанных подходов ориентирован на определение вектора начальных значений $x(0)$, соответствующего установившемуся режиму работы консультируемой проблемы, для которой

$$\delta[x(0)] = x(0) - x(T) = 0, \quad (8.120)$$

где T — период рассчитываемого процесса; $x(0)$, $x(T)$ — значения вектора переменных x в моменты времени $t=0$, $t=T$. При этом математическая модель анализируемой консультируемой проблемы представлена системой уравнений вида (8.1), т. е.

$$F(\dot{x}, x, t) = 0.$$

Квазиньютоновские методы. При решении уравнения (8.120) методом Ньютона (8.4) получим

$$x^{i+1}(0) = x^i(0) - [I - W]^{-1} [x^i(0) - x(T)], \quad (8.121)$$

где

$$W [x^i(0), T] = \partial x [x^i(0), T] / \partial x^i(0)$$

— матрица перехода состояний, или матрица чувствительностей переменных модели к своим начальным значениям.

Для сокращения времени вычислений матрица $W[x^i(0), T]$ определяется одновременно с $x^i(t)$ в процессе интегрирования системы уравнений в вариациях, получающейся на основе выражения (8.1),

$$\frac{\partial F}{\partial \dot{x}} \dot{z} + \frac{\partial F}{\partial x} z = 0, \quad (8.122)$$

где

$$z = \partial x^i(t) / \partial x^i(0)$$

— матрица частных производных.

Если для аппроксимации производной воспользоваться, например, выражением (8.60)

$$z_{n+1} = -\frac{1}{h} \sum_{i=0}^k \alpha_i^* z_{n+1-i},$$

то уравнение (8.122) преобразуется к виду

$$\left(\frac{\partial F}{\partial x} - \frac{\alpha_0^*}{h} \frac{\partial F}{\partial \dot{x}} \right) z_{n+1} = \frac{1}{h} \frac{\partial F}{\partial \dot{x}} \sum_{i=1}^k \alpha_i^* z_{n+1-i}, \quad (8.123)$$

где

$$\left(\frac{\partial F}{\partial x} - \frac{\alpha_0^*}{h} \frac{\partial F}{\partial \dot{x}} \right)$$

— матрица Якоби.

Таким образом, на каждой итерации при решении уравнения (8.121) на интервале $[0, T]$ интегрируется система уравнений (8.1) с одновременным вычислением $\hat{\partial}x^i(t_{n+1})/\hat{\partial}x^i(0)$, в соответствии с выражением (8.123). При этом вектор $x^i(0)$ начальных значений, соответствующий установившемуся режиму, определяется при одновременном выполнении условий сходимости итерационного процесса согласно выражению (8.121)

$$\max_k \left| \frac{\Delta x^{i+1}(0)}{x_{\max}^i(0)} \right| < \epsilon_{\text{доп}}$$

и условия периодичности (8.120)

$$\left| \frac{x^i[x^i(0), T] - x^i(0)}{x^i(0)} \right| < \epsilon_0$$

соответственно.

Перепишем выражение (8.121) в виде, известном как метод Эйприла — Трика:

$$x^{i+1}(0) = x^i(0) - \lambda \Delta x^{i+1}(0),$$

где

$$\Delta x^{i+1}(0) = [1 - W]^{-1} \{x^i(0) - x^i[x^i(0), T]\},$$

а демпфирующий параметр $\lambda \leq 1$ выбирается по аналогии с выражениями (8.8) и (8.9).

При анализе автоколебательных режимов консультируемых проблем период T неизвестен и его требуется определить. В этом случае вектор переменных модели консультируемой проблемы видоизменяется путем введения дополнительной составляющей, соответствующей периоду T^i , вместо одной из переменных выражения

$$x(0) = [x_1(0), \dots, x_{j-1}(0), T^i, x_{j+1}(0), \dots, x_n(0)]^t.$$

Корректируются также j -й столбец матрицы перехода состояний W^i , определяемый выражением

$$\frac{\partial x^i(T^i)}{\partial T^i} = -\frac{1}{h} \sum_{\ell=0}^k \alpha_{\ell}^* x_{n+1-\ell}, \quad (8.124)$$

и j -й диагональный элемент единичной матрицы 1, приобретающий нулевое значение. Позиция j выбирается на каждой итерации, например, как номер компоненты $\partial x^i(T^i)/\partial T^i$, соответствующего значению максимальной производной из выражения (8.124).

Квазиньютоновские методы достаточно эффективны при анализе периодических режимов моделируемых консультируемых проблем. Скорость их сходимости повышается, если в качестве начальных на первой итерации используется значение вектора x , полученное путем предварительного интегрирования системы уравнений (8.1) обычными методами в течение нескольких периодов рассчитываемого процесса.

Градиентный метод. Задача анализа установившихся периодических режимов в нелинейных консультируемых проблемах может быть сформулирована как задача поиска минимума целевой функции

$$\Phi[x(0), T] = \delta^t[x(0), T] \delta[x(0), T], \quad (8.125)$$

где $\delta[x(0), T]$ определяется из выражения (8.120).

Градиент такой целевой функции соответствует

$$\frac{\partial \Phi[x(0), T]}{\partial x(0)} = g[x(0)] = 2 \left\{ \frac{\partial \delta[x(0), T]}{\partial x(0)} \right\}^t \delta[x(0), T]$$

или с учетом выражения (8.120)

$$\begin{aligned} g[x(0), T] &= 2 \left\{ I - \frac{\partial x[x(0), T]}{\partial x(0)} \right\}^t \delta[x(0), T] = \\ &= 2 \left\{ \delta[x(0), T] - \frac{\partial x[x(0), T]}{\partial x(0)} \delta[x(0), T] \right\}. \end{aligned} \quad (8.126)$$

Матрица $W = \partial x[x(0), T]/\partial x(0)$ может быть получена по аналогии с предыдущим. Из выражения (8.122) следует, что

$$\dot{z} = - \left(\frac{\partial F}{\partial \dot{x}} \right)^{-1} \frac{\partial F}{\partial x} z = Cz. \quad (8.127)$$

Полученному уравнению (8.127) можно поставить в соответствие сопряженное уравнение

$$\dot{y} = C_{x[t, x(0)]}^t y = \left[\left(\frac{\partial F}{\partial \dot{x}} \right)^{-1} \frac{\partial F}{\partial x} \right]_{x[t, x(0)]}^t y, \quad (8.128)$$

при этом

$$z^t(t) y(t) = z^t(0) y(0) = \text{const}. \quad (8.129)$$

В справедливости соотношения (8.129) легко убедиться, если продифференцировать его по времени и принять во внимание выражения (8.127) и (8.128):

$$\begin{aligned} \frac{d}{dt} [z^t(t) y(t)] &= \dot{z}^t(t) y(t) + z^t(t) \dot{y}(t) = \\ &= -z^t C^t y(t) + z^t C^t y(t) = 0 = \text{const.} \end{aligned}$$

Если учесть, что $z = \partial x[t, x(0)] / \partial x(0)$, то для $t = T$ получаем

$$\frac{\partial x[x(0), T]}{\partial x(0)} y(T) = \frac{\partial x[x(0), 0]}{\partial x(0)} y(0) = y(0). \quad (8.130)$$

Следовательно, если положить $y(T) = \delta[x(0), T]$ и определить $y(0)$ путем интегрирования сопряженных уравнений на интервале $[T, 0]$ в обратном времени, начиная с $y(T)$, то в соответствии с выражениями (8.130) и (8.126) найдем

$$g[x(0)] = 2 \{ \delta[x(0), T] - y(0) \delta[x(0), T] \}. \quad (8.131)$$

В результате основными этапами вычисления градиента целевой функции являются: интегрирование исходной модели консультируемой проблемы по выражению (8.1) на интервале $[0, T]$, начиная с $x(0)$, и запоминание решений $x[t, x(0)]$; итерационное вычисление вектора невязки согласно уравнению (8.120); решение сопряженных уравнений (8.128) в обратном времени на интервале $[T, 0]$ вдоль траектории $x[x(0), t]$, начиная с $\delta[x(0), T]$; определение градиента в соответствии с выражением (8.131).

Для случая, когда период процесса T неизвестен, в градиенте целевой функции появляется дополнительная составляющая

$$\frac{\partial \Phi[x(0), T^t]}{\partial T^t} = -2\delta^t[x(0), T^t] \frac{\partial x[x(0), T^t]}{\partial T^t}$$

или с учетом выражения (8.124)

$$\frac{\partial \Phi[x(0), T^t]}{\partial T^t} = 2\delta^t[x(0), T^t] \frac{1}{h} \sum_{i=0}^k \alpha_i \dot{x}_{n+1-t}. \quad (8.132)$$

После этого, используя одну из процедур оптимизации, определяем значения $\Delta x^{i+1}[x^i(0), T^{i+1}]$ с учетом значений $\Phi[x^i(0), T]$ и $g[x^i(0)]$ и корректируем величину $x^i(0)$ вектора начальных значений. Критерий сходимости используется в виде

$$\left| \frac{\Phi[x^t(0), T^t]}{[x^t(0)]^t x^t(0)} \right| \leq \epsilon_0.$$

Экстраполяционный метод базируется на ϵ -алгоритме, который предназначен для нахождения числовой последовательности с экс-

понижающимися членами и может быть представлен следующей экстраполяционной схемой:

$$\begin{array}{ccccccc}
 E_{-1}^{(0)} & & & & & & \\
 & E_0^{(0)} & \dots & & & & \\
 E_{-1}^{(1)} & E_0^{(1)} & & & E_{2m}^{(0)}, & q = 0, 1, \dots & \\
 & \vdots & & & & & \\
 & E_0^{(q-1)} & \dots & & & & \\
 E_{-1}^{(q)} & & & & & &
 \end{array}$$

При этом последующие значения $E_{k+1}^{(q)}$ вычисляются согласно выражению

$$E_{k+1}^{(q)} = E_{k-1}^{(q-1)} + [E_k^{(q+1)} - E_k^{(q)}]^{-1}, \quad (8.133)$$

где

$$E_{-1}^{(q)} = 0; E_0^{(q)} = \{S_q\}, \quad k, q = 0, 1, \dots, \{S_q\}$$

- последовательность скалярных либо векторных величин.

Оказывается справедливым следующее соотношение:

$$\sum_{v=0}^m C_v S_{c+v} = \left(\sum_{v=0}^m C_v \right) S, \quad q = 0, 1, \dots, \quad (8.134)$$

где

$$S = \lim_{q \rightarrow \infty} \{S_q\},$$

а C_v — некоторые действительные числа. Тогда применение ε — алгоритма к последовательности $\{S_q\}$ приводит к

$$E_{2m}^{(q)} = S, \quad q = 0, 1, \dots, \text{ если } \sum_{v=0}^m C_v \neq 0 \text{ и } E_k^{(q+1)} \neq E_k^{(q)}.$$

Пусть последовательность $\{x_q\}$ соответствует значениям вектора переменных из итерационной формулы

$$x_{q+1}^i = \psi(x_q^i), \quad q = 0, 1, \dots, 2m,$$

начиная с $x_0^i = y^i$, $i = 0, 1, \dots$, и пусть y — вектор начальных значений переменных консультируемых проблем, соответствующий установившемуся режиму консультируемой проблемы, т. е. решению уравнения вида $y = \psi(y)$.

На каждом шаге вычисления последовательности $\{x_q\}$ получаем

$$x_{q+1}^i = y + \psi'(y)(x_q^i - y) + O(\|x_q^i - y\|^2). \quad (8.135)$$

Если

$$\sum_{v=0}^m C_v \lambda^v$$

соответствует минимальному многочлену $\psi'(y)$ степени m по отношению к $(x'_0 - y)$, то согласно выражению (8.135) находим

$$\sum_{v=0}^m C_v x_{q+v}^i = \left\{ \sum_{v=0}^m C_v [\psi'(y)]^v (x'_0 - y) + \sum_{v=0}^m C_v [y + 0(\|x'_0 - y\|^2)] \right\}. \quad (8.136)$$

Так как первая сумма в правой части уравнения (8.136) в силу свойств минимального многочлена равна нулю, то

$$\sum_{v=0}^m C_v x_{q+v}^i = \left(\sum_{v=0}^m C_v \right) [y + 0(\|x'_0 - y\|^2)]. \quad (8.137)$$

Сравнивая выражения (8.137) и (8.134), находим, что $E_{2m}^{(q)} = y + 0(\|y' - y\|^2)$, что справедливо, в частности, и для $E_{2m}^{(0)}$.

В результате

$$y^{t+1} = y + 0(\|y^t - y\|^2), \quad (8.138)$$

или решение стремится к y по крайней мере по квадратичному закону.

При расчете установившихся периодических режимов последовательность $\{x_q\}$, $q = 1, \dots, 2m$ формируется путем интегрирования уравнений исходной модели консультируемой проблемы (8.1) на интервале $[0, 2mT]$, где T — период моделируемого процесса. При интегрировании выбираются некоторые начальные значения $x_{i=0}^j = y^i$, начиная с $i = 0$.

Применяя ε -алгоритм к сформированной последовательности, находят

$$E_{2m}^{(0)} = y^{i+1}.$$

Далее по аналогии с выражением (8.7) вычисляют и проверяют максимальную относительную погрешность

$$E_s = \max_s \left| \frac{y^{i+1} - y^i}{x_{\max}} \right| < \varepsilon, \quad s = 1, 2, \dots, n,$$

где x_{\max} — вектор максимальных значений $x_{i=0}^j$.

Если погрешность превышает допустимую, то изменяют вектор максимальных значений

$$x_{\max} = \max_s [\|y^{i+1}\|, x_{\max}], \quad s = 1, \dots, n$$

и, полагая $x_{i=0}^{j+1} = y^{i+1}$, повторяют цикл вычислений.

Отсутствие в рассмотренном методе трудоемких этапов оптимизации, вычисления вспомогательных матриц (например, W) и довольно простое согласование алгоритма с программами анализа динамических процессов выгодно отличает рассмотренный подход от упомянутых

выше и позволяет без особых усилий расширить функциональные возможности комплексов моделирующих программ.

Пример 3. Сформируем рекомендации по определению различными методами установившегося периодического режима схемы, показанной на рис. 8.7.

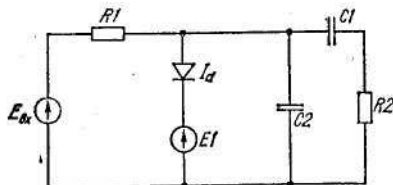


Рис. 8.7. RC-схема

Ее математическая модель, как модель консультируемой проблемы:

$$\begin{aligned} 0,1x_1 + 0,01(x_1 - x_2) &= 0, \\ 0,005x_2 + I_d + 0,01(x_2 - E_{вх}) - 0,01(x_1 - x_2) &= 0, \\ x_3 &= x_2 - E_1, \end{aligned}$$

где

$$\begin{aligned} x_1 &= u_{C1}; \quad x_2 = u_{C2}; \quad x_3 = u_d; \quad I_d = 10^{-8}(e^{40u_d} - 1); \\ E_1 &= 5В; \quad E_{вх} = 2 \sin(2\pi 500t). \end{aligned}$$

Вычисленные характеристики установившегося периодического режима этой схемы приведены в табл. 8.6.

Таблица 8.6

К определению установившегося периодического режима

Алгоритм расчета	Число итераций	Число пройденных периодов	Погрешность вычислений
Квазиньютоновский	4	4	$0,415 \cdot 10^{-8}$
Экстраполяционный	2	8	$0,5168 \cdot 10^{-7}$
Обычное интегрирование	100	100	$0,2809 \cdot 10^{-6}$

Как и ранее, пользователь САК может эффективно влиять на процесс моделирования установившихся периодических режимов, если выберет минимальный шаг интегрирования, относительную погрешность решения нелинейных систем методом Ньютона, относительную погрешность метода интегрирования, допустимую относительную погрешность решаемой задачи и др.

8.10. Задачи анализа структур консультируемых проблем

Консультируемые проблемы характеризуются структурами, определяющими множество их качественных свойств, обуславливаемых множеством элементов и связей между ними. При анализе структур консультируемых проблем (КП) и их подпроблем в целях выявления их общих свойств конкретные (физические) реализации отдельных элементов и связей не имеют существенного значения и обычно не рассматриваются. Подобное абстрагирование позволяет производить анализ КП в наиболее общем виде — как схему «чистых» отношений ее элементов и связей. Основным свойством структур является однозначное соответствие их некоторому подмножеству K_s качественных свойств КП, причем каждой структуре соответствует определенный вид этого подмножества.

Однако обратного однозначного соответствия подмножества K_s и структуры КП не существует, т. е. одно и то же подмножество качественных свойств может принадлежать нескольким (в общем случае — множеству) различным структурам КП, причем, чем меньше мощность подмножества K_s , тем больше мощность множества соответствующих структур s .

Указанное обстоятельство имеет существенное значение в процессе консультирования КП, так как им обуславливается необходимость выбора (формирования) рекомендаций оценки качества структур или соответствующих критериев, определяющих метод выбора предпочтительных или оптимальных структур из множества s .

Одним из *свойств структур* является также *их устойчивость к влиянию внешних воздействий*, что может выражаться в нарушении тех или иных связей, вследствие чего уменьшается число их качественных свойств, определяемое мощностью подмножества K_s или наоборот, исключают или препятствуют использованию конкретной КП по ее прямому назначению.

Одной из *главных задач структурного анализа* является *построение наглядной формальной модели, отображающей существующую систему отношений элементов как между собой, так и с внешней средой*. Структурная модель КП является *многоуровневой*, причем конкретизация структуры дается на стольких уровнях, сколько их требуется для создания полного представления об основных свойствах КП (рис. 8.8).

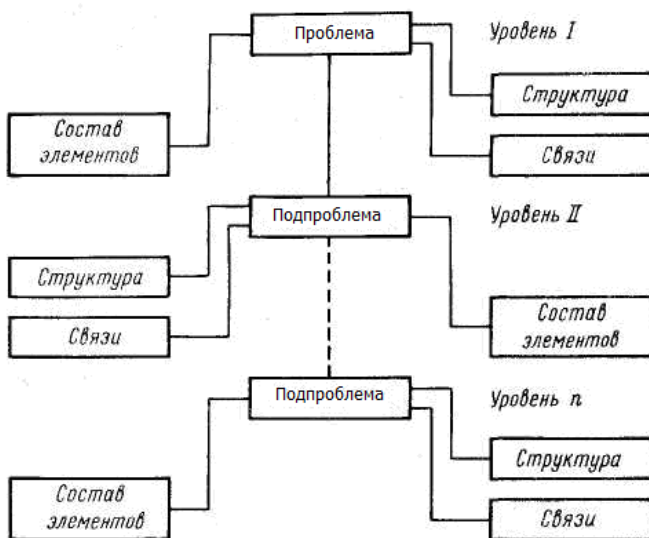


Рис. 8.8. Структурная модель КП

Рассмотрим структурную модель КП на нескольких крупных уровнях: организационном, функциональном, техническом. При анализе структур КП решаются следующие задачи:

организационная:

- 1) описание состава КП и построение ее структурной схемы;
- 2) определение функции отдельных подпроблем КП, раскрытие их структурной схемы;
- 3) описание материальных, вещественных, информационных и других видов связи;
- 4) построение обобщенной структурной модели КП;

функциональная:

- 1) изучаются вопросы функционирования каждой подпроблемы КП;
- 2) выбирается состав функций;
- 3) определяются их взаимосвязи;
- 4) составляется обобщенная структура задач функционирования КП;

техническая:

- 1) определяются основные элементы, участвующие в реализации функциональных задач КП;

2) составляется формальная структурная модель системы технических средств с учетом топологии расположения элементов КП и взаимодействия их как между собой, так и с внешней средой.

Независимо от уровня рассмотрения **общая задача структурного анализа КП состоит в том, чтобы, исходя из заданного описания элементов КП и непосредственных связей между ними, получить заключение о структурных свойствах КП в целом и ее подпроблем.** При решении практических задач структурного анализа КП будем принимать **три уровня описания связей между элементами:**

- 1) наличие связи;
- 2) направление связи;
- 3) вид и направление воздействий, определяющих взаимодействие элементов.

На первом уровне, когда исходят из наличия или отсутствия связей между элементами, структуре КП может соответствовать неориентированный граф, вершинами которого являются элементы КП, а ребрами — связи между элементами. Основные задачи структурного анализа на этом уровне сводятся:

- 1) к определению связности (целостности) КП; если КП не является связной, то ставят задачу выделения изолированных несвязных подпроблем со списками входящих в них элементов;
- 2) к выделению циклов;
- 3) к определению минимальных и максимальных последовательностей элементов (цепей), разделяющих элементы друг от друга.

На втором уровне, когда задано направление связи, КП соответствует ориентированный граф, направления дуг которого совпадают с направлениями связей. На этом уровне результаты структурного анализа оказываются более содержательными. К задачам структурного анализа в этом случае относят:

- 1) определение связности КП;
- 2) топологическую декомпозицию с выделением сильно связанных подпроблем;
- 3) выделение уровней в структуре и определение их взаимосвязи;
- 4) определение максимальных и минимальных длин путей;
- 5) определение характеристик топологической значимости элементов;
- 6) получение информации о слабых местах структуры и др.

На третьем уровне описания связей между элементами КП учитывается не только направленность связи, но и раскрываются

состав и характер функций взаимодействия элементов (входные, входные и управляющие воздействия).

Далее будут рассмотрены задачи структурного анализа при минимальном объеме априорной информации о структуре КП, когда учитывается только факт наличия связи и ее направленность.

8.11. Введение в анализ функциональных и системных структур КП

Разработка любых сложных систем, в том числе и систем автоматизированного консультирования, обычно начинается с анализа требований к системе и идентификации основных функций и отношений между функциями. Этот процесс может быть представлен диаграммой, изображенной на рис. 8.9.

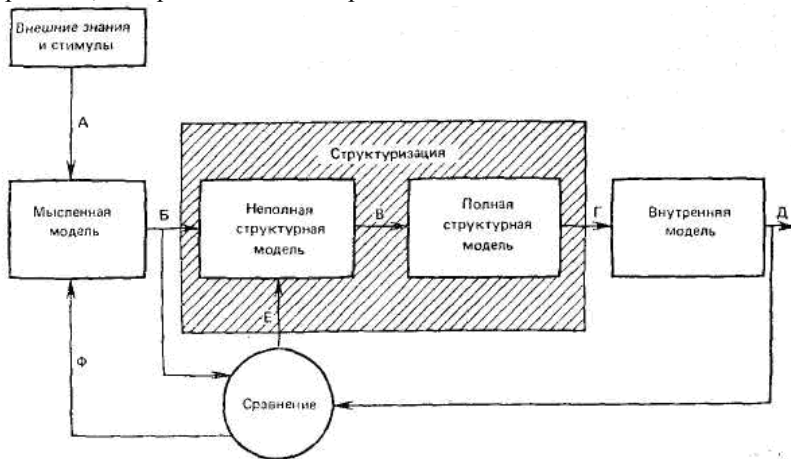


Рис.8.9. Диаграмма процесса структурного моделирования КП

Под воздействием внешних стимулов консультант, используя знания, полученные в результате анализа (А), создает мысленную модель КП. Используя ее, консультант генерирует неполную структурную модель КП (Б), на базе которой строит полную структурную модель КП путем разбиения и группирования (В). Для получения полной структурной модели используются методы адекватного структурирования частей в целостную проблему. Полученная полная модель КП, представленная в виде внутренней модели (Д), сравнивается с мысленной моделью КП. При этом обычно требуется уточнение мысленной модели и внесение

изменений и исправлений в ее структуру. Кроме того, в ходе этого этапа обычно достигается более глубокое понимание структуры проблемы, интерпретируется и корректируется полная структурная модель КП. Надо также отметить большую важность процесса визуализации и документирования структурной модели КП.

Описанный процесс будем называть структурным моделированием (анализом) КП и он позволяет:

- идентифицировать ключевые элементы КП и их отношения;
- показать существование или не существование требуемых элементов и связей;
- определить уровни в проблемной иерархии;
- специфицировать подпроблемы внутри уровней;
- определить критические и сильносвязанные элементы.

Процесс структурного моделирования КП включает в себя эмпирические и интуитивные шаги, поэтому его строгое определение затруднительно.

Пусть A и T — множество и мощность множества A , t — подмножество T , тогда $t \times t$ — декартово произведение этого множества на себя. Определим отношение в $t \times t$ как подмножество пар в $t \times t$, а свойство в отношении — как присвоение информации (направлений, знаков, весов и других описателей) каждой паре в отношении. Если рассматривать множество отношений

$R = \{r_1, \dots, r_j, \dots, r_n\}$, $r_j \in t \times t$ и множество соответствующих свойств $P = \{p_1, \dots, p_i, \dots, p_n\}$, то $S = (A, R)$ и $U = (A, R, P)$ обозначают соответственно структуру и нагруженную структуру.

Рассмотрим две нагруженные структуры $U_1 = (A_1, R_1, P_1)$ и $U_2 = (A_2, R_2, P_2)$. Если между U_1 и U_2 имеются следующие отношения: $A_2 = A_1 \cup \Delta A$, ΔA — приращение элементов, $R_2 = R_1 \cup \Delta R$, ΔR — приращение отношений, $P_2 = P_1 \cup \Delta P$, ΔP — приращение свойств, то структура U_2 является *более расширенной*, чем U_1 , или структура U_1 является *более элементарной*, чем U_2 .

Процесс получения более расширенной нагруженной структуры из элементарной будем называть структурированием.

Когда структура реализуется для реальной КП, то ее будем называть *структурной моделью консультируемой проблемы*. Может быть получен широкий спектр структурных моделей, формируемых путем структуризации, сгенерированной из первоначальной неполной структурной модели КП. Конечная модель КП должна объединять в себе различные точки зрения на структуру КП.

Процесс структурного моделирования можно формально представить рис. 8.10.

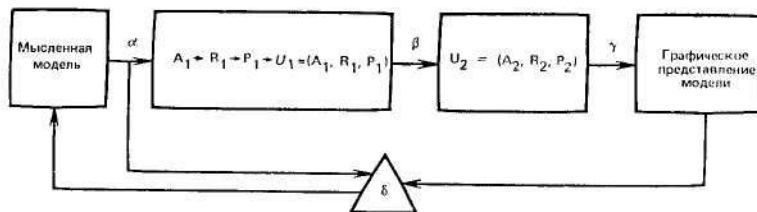


Рис. 8.10. Формальное представление процесса структурного моделирования

Среди показанных подпроцессов (α — генерация, β — структуризация, γ — визуализация, δ — модификация) α и β являются эмпирическими или интуитивными, а β и γ — формализованными. На данном рисунке не показаны процессы документирования и анализа внешних знаний и стимулов, приведенные на рис. 8.9.

Для *представления структурной модели КП* будем использовать *таблицы, матрицы и графы* вместо отношений. При этом обозначения удобно заменить графическим представлением, что облегчает работу со структурной информацией. Графическое представление определяется как двумерная диаграмма структурной модели КП или ее части. На рис. 8.11 показаны графы, которые рекомендуется широко применять в структурных моделях КП, где a — ненаправленный, b — направленный, e — знаковый, z — нагруженный, d, e — гиперграфы, $ж$ — композиционный граф.

Применение метода структурного моделирования рассмотрим на примере функциональной структуры. Исходным при этом является список элементов (переменных) и текстовое описание проблемы. Поэтому для начальной идентификации элементов консультационной процедуры удобен способ табличного описания, позволяющий иерархически описывать консультационную процедуру путем последовательного заполнения четырех граф: *условие, событие, действие и результат*. Таким образом, на каждом уровне иерархии описания в графах «условие» и «результат» идентифицируются элементы-данные, а в графах «событие» и «действие» — элементы-функции. Последовательность описаний внутри каждого уровня и между ними задает структуру управления для данной консультационной процедуры.

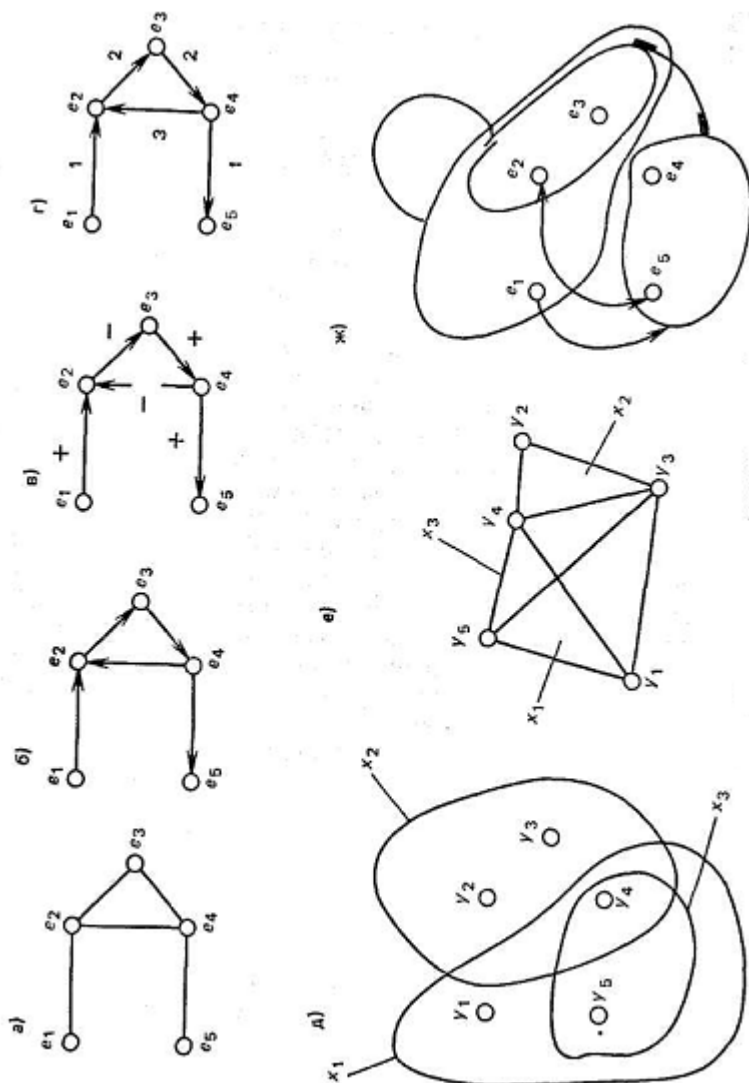


Рис. 8.11. Графы структурных моделей КП

На нижнем уровне иерархии процесс консультирования, представляющий собой композицию консультационных процедур, описывается в терминах **элементарных функций и данных** с точки зрения конкретной предметной области. Рассмотрим пример:

пусть описание функциональной структуры состоит из множества функций

$$F = \{f_1, f_2, \dots, f_n\}.$$

Для любых f при $f_i \neq f_j \Rightarrow i \neq j$, тогда можно построить матрицу смежности функций

$$A = \begin{matrix} & \begin{matrix} f_1 & f_2 & \dots & f_n \end{matrix} \\ \begin{matrix} f_1 \\ \cdot \\ \cdot \\ \cdot \\ f_n \end{matrix} & \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1n} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ z_{n1} & z_{n2} & \dots & z_{nn} \end{bmatrix} \end{matrix},$$

где $z_{ij} = 0$ при $f_i \bar{R} f_j$ и $z_{ij} = 1$ при $f_i R f_j$.

Символ R представляет собой бинарное отношение между функциональными элементами системы.

Зная матрицу A , можно определить матрицу достижимости функций путем сложения матрицы смежности и единичной матрицы I и умножения полученной матрицы саму на себя $(n - 1)$ раз:

$$D = (A + I)^n = (A + I)^{n-1} \neq (A + I)^{n-2}.$$

Эта матрица показывает, что длина пути, по которому данные из функционального элемента k достигают функциональный элемент m , находится в диапазоне от 0 до $n - 1$.

В соответствии с методом структурного анализа далее определяются разбиения множества F , уточняющие функциональную структуру путем выполнения следующих процедур:

— вычисления R^* — рефлексивно-транзитивного замыкания отношения R ;

— выделения уровней иерархии L_i на множестве F :

$$\Pi_1[F] = [L_0, L_1, \dots, L_m],$$

где $\Pi_1[F]$ — разбиение F на уровни иерархии в соответствии с алгоритмом $L_0 = 0$;

$$L_j = \{f_i \in \bar{L}_j | Q_j(f_i) = Q_j(f_i) \cap P(f_i)\},$$

где

$$\bar{L}_j = F - L_0 - L_1 - \dots - L_{(j-1)}; \quad Q_j(f_i) = \{f_k \in L_j | f_i R^* f_k\} -$$

множество элементов \bar{L}_j , достижимых из f_i ;

$P_j(f_i) = \{f_k \in \bar{L}_j | f_k R^* f_i\}$ — множество элементов \bar{L}_j , из которых достигим f_i ;

— анализа на достижимость, выполняемого на основе выделения всех компонент связности графа, задающего анализируемую структуру, и установления недостающих связей в графе;

— анализа на согласованность, выполняемого путем определения циклов внутри каждого уровня иерархии: $\Pi_2[L_i] = \{S_i, C_i\}$, где S_i — множество не связанных внутри уровня L_i элементов, C_i — множество элементов, образующих циклы в L_i , $S_i = \{f_j \in L_i \mid Q_i(f_j) = \{f_j\}\}$, где $Q_i(f_k) = \{f_k \in L_i \mid R^* f_k\}$ — множество элементов уровня L_i , достижимых из f_k ;

— предметной интерпретации выделенных циклов и сравнения полученной интерпретации каждого цикла с предметными интерпретациями других элементов, принадлежащих тому же уровню, что и рассматриваемый цикл, а также выделения транзитивных связей между элементами несмежных уровней;

— декомпозиции на подпроблемы $\Pi_3[F] = [U_1, U_2, \dots, U_k]$, где

U_i — подпроблема, которая определяется как последовательность сильно связанных функций, достижимых и предшествующих друг другу, с минимизацией связей с другими подпроблемами.

Программная реализация метода интерпретирующего структурного моделирования обеспечивает: **создание и редактирование структурной модели сложной КП, анализ имеющейся структуры, протоколирование хода работ консультанта с интерактивным комплексом программ интерпретирующего структурного моделирования (ИКП ИСМ), создание выходных документов, отражающих текущую структуру моделируемой КП.**

Преимущества ИКП ИСМ заключаются в обеспечении диалогового режима работы и в использовании внешней памяти на магнитных носителях хранения данных о структуре исследуемой КП. Использование системы меню, а также контроль ответов консультанта позволяют удобно и оперативно «проигрывать» различные варианты структуры моделируемой проблемы с получением документации на каждом месте.

Метод структурного анализа обеспечивает выполнение следующих функций:

- 1) создание, модификацию, просмотр матриц взаимодействия элементов проблемы;
- 2) вычисление матриц достижимости;
- 3) вычисление уровней иерархии по рассчитанной ранее матрице достижимости;
- 4) распечатку всех интересующих консультанта матриц и уровней иерархии в удобной для чтения форме;

- 5) определение циклов (сильносвязанных множеств) проблемы внутри каждого уровня иерархии;
- 6) прекращение сеанса работы с ИКП ИСМ по требованию консультанта с сохранением результатов до следующего сеанса;
- 7) ведение диалога с неподготовленным консультантом, имеющим общее представление о ИКП ИСМ;
- 8) выделение транзитивных (избыточных) связей между элементами исследуемой проблемы.

Входными данными для работы ИКП ИСМ является множество ключевых терминов (словарь), каждый из которых описывает простой функционально законченный на некотором уровне рассмотрения элемент исследуемой проблемы. Это множество выделяется в процессе анализа предметной области консультантом соответствующей предметной области.

На первом этапе работы консультант предметной области с помощью ИКП ИСМ в диалоговом режиме специфицирует структуру на заданном множестве элементов проблемы, определяя связи между элементами. В результате выполнения первого этапа формируется матрица связности исследуемой проблемы.

На последующих этапах работы с ИКП ИСМ анализируется заданная структура с целью выявления из перечня необходимых свойств, поддерживаемых проблемой.

Конечным результатом работы с ИКП ИСМ является набор документов, отражающих результаты анализа исследуемой структуры КП в терминах данной предметной области.

Для обеспечения мобильности все машинно-зависимые функции, использованные в ИКП ИСМ, выделяют в отдельные модули с их полной спецификацией.

В качестве примера рассмотрим построение функциональной структуры подсистемы технологического проектирования процессов и оснащения сборки в машиностроении (рис. 8.12).

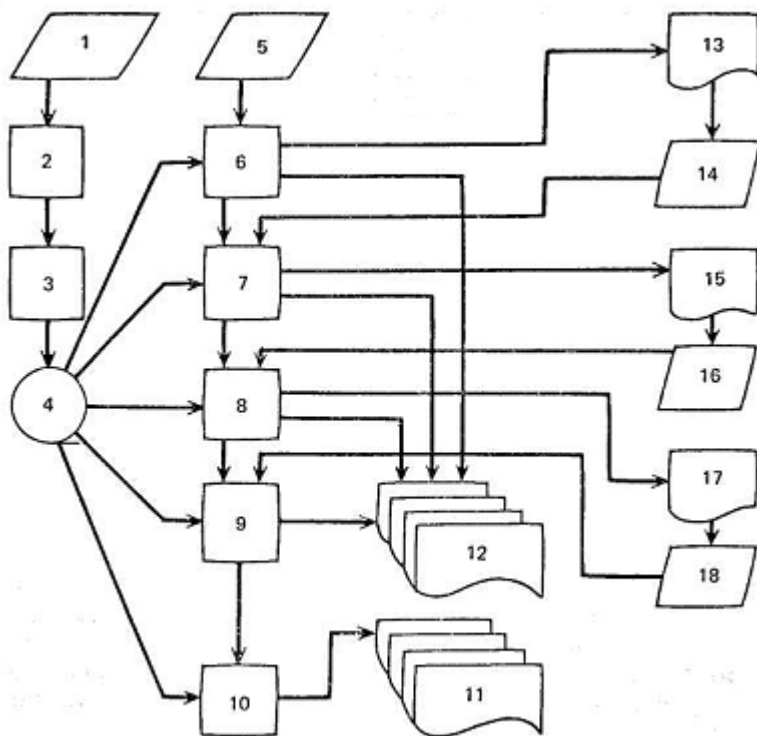


Рис. 8.12. Функциональная структура подсистемы технологического проектирования

В подсистеме осуществляется автоматизированное проектирование технологических процессов и оснащения сборки, а также оценка технологичности сборочных единиц путем реализации последовательности функциональных преобразований, где первоначальная функциональная декомпозиция рассматриваемой подсистемы состоит из следующих этапов:

1. Разработка модели производственной системы (МПС).
2. Отладка МПС.
3. Печать МПС.
4. Запись МПС на носитель информации.
5. Подготовка исходных данных.
6. Выбор схемы базирования.

7. Определение конструктивной схемы сборочного приспособления.
8. Проектирование технологического процесса.
9. Формирование технологической документации.
10. Определение последовательности установки элементов сборочной единицы.
11. Печать документации.
12. Печать промежуточных результатов проектирования и диагностирования МПС.
13. Выбор вариантов схем базирования.
15. Выбор вариантов конструктивных схем сборочных приспособлений.
17. Выбор вариантов последовательности установки.
- 14, 16, 18. Дополнение и изменение исходных данных.

Применение метода интерпретирующего структурного анализа позволяет сформировать рекомендации по построению упорядоченной с точки зрения агрегации данных функциональную структуру, из которой следует несогласованность по данным между элементом 4 и элементами 6, 7, 8, 9, 10, а также между элементом 12 и элементами 6, 7, 8, 9 (рис. 8.13).

Предметная интерпретация полученных результатов показывает, что основными функциональными преобразованиями в рассматриваемом процессе проектирования являются преобразования, выполняемые элементами 6, 7, 8 и 9. Анализ текстуального описания элементов 6, 7, 8 и 9 показывает, что они представляются в виде композиции следующих функциональных элементов:

Для элемента 6

- 6.1. Формирование множества возможных схем.
- 6.2. Проверка схемы на отсутствие нерациональных сочетаний сборочных баз.
- 6.3. Предварительная проверка схемы по точности.
- 6.4. Проверка схемы на существование возможной последовательности установки.
- 6.5. Выбор из всех возможных схем базирования схемы, оптимальной по заданному критерию, в качестве которого принимается технологическая себестоимость.

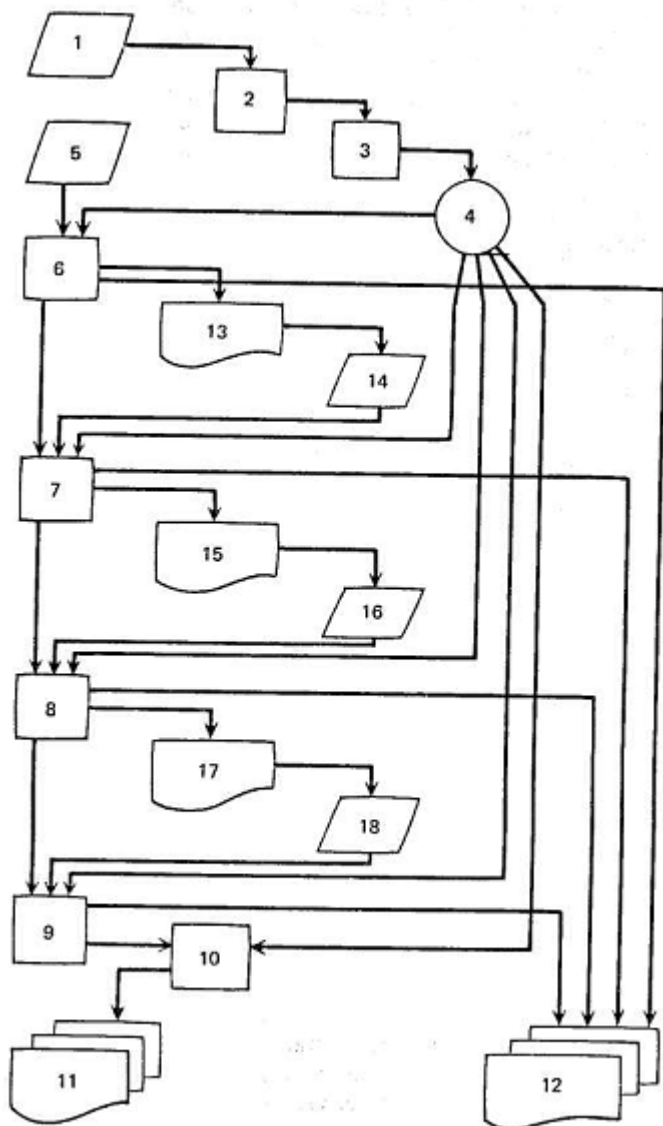


Рис. 8.13. Иерархически упорядоченная функциональная структура подсистемы технологического проектирования

Для элемента 7

- 7.1. Уточнение схемы базирования.
- 7.2. Определение состава базовых элементов приспособления.
- 7.3. Укрупненный расчет размерных характеристик базовых элементов.
- 7.4. Выбор типовой конструкции каркаса сборочного приспособления.
- 7.5. Выбор конкретных конструктивных элементов сборочного приспособления.
- 7.6. Определение конкретных конструктивных элементов сборочного приспособления.
- 7.7. Определение размерных характеристик конкретных конструктивных элементов сборочного приспособления.
- 7.8. Прочностные расчеты конкретных конструктивных элементов сборочного приспособления.
- 7.9. Выбор из числа возможных конструктивных схем сборочного приспособления схемы, оптимальной по заданному критерию оптимизации.

Для элемента 8

- 8.1. Формирование массива перестановок, возможных по условиям базирования и доступа.
- 8.2. Сокращения пространства возможных перестановок.
- 8.3. Поиск последовательности устоновки, оптимальной по заданному критерию.

Для элемента 9

- 9.1. Определение состава и последовательности этапов сборки.
- 9.2. Предварительная оценка рекомендаций по укрупненным показателям.
- 9.3. Детализация этапов сборки до уровня операций.
- 9.4. Выбор вида оборудования, инструмента и оснастки.
- 9.5. Предварительная оценка рекомендаций по укрупненным показателям.
- 9.6. Нормирование.
- 9.7. Расчет технико-экономических показателей.
- 9.8. Выбор оптимального варианта технологического процесса по заданному критерию.

Повторное применение метода интерпретирующего структурного анализа к полученному множеству элементов формирует уточненную функциональную структуру анализируемой подсистемы (рис. 8.14).

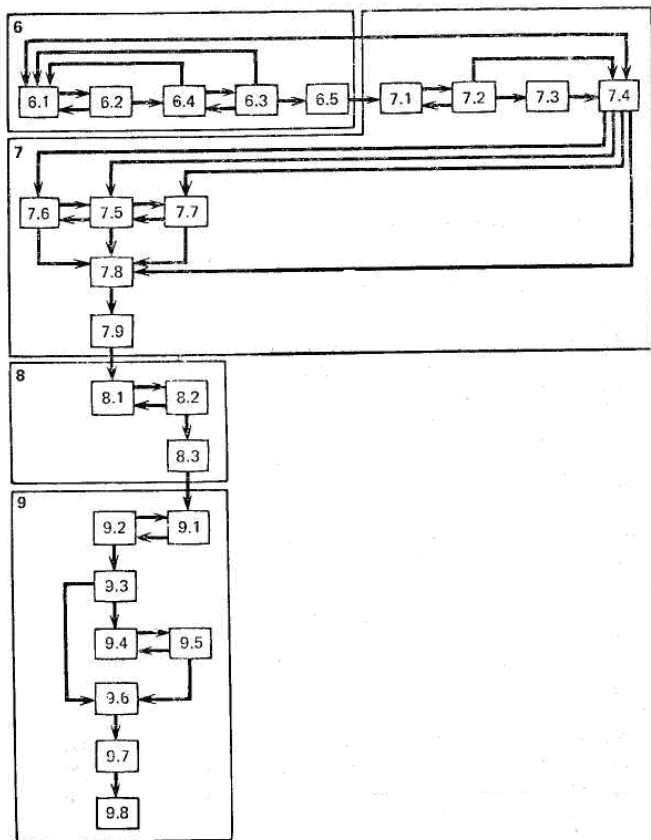


Рис. 8.14. Уточненная функциональная структура

Анализ структуры на достижимость и сравнение с результатами анализа, полученными на предыдущем этапе, показывают отсутствие связи элемента 7 с элементом 6, что свидетельствует о противоречивости исходной структурной декомпозиции и необходимости ее устранения, например, путем применения метода синтеза функциональной структуры.

8.12. Анализ структур КП

Классификация всего многообразия *структур КП* может быть произведена по различным признакам, наиболее употребительными из

которых являются: *этап создания, природа связей, отношение между частями, структурные признаки.*

При использовании *первого классификационного признака* все структуры КП будем разделять на подмножества исходных S_n , разрабатываемых S_p и производных S_{np} структур, причем

$$S = S_n \cup S_p \cup S_{np}.$$

Подмножество S_n составляет структуры, задаваемые или однозначно определяемые множеством требований Z .

Элементами подмножества S_n являются структурные, функциональные, электрические, кинематические, пневматические, гидравлические и другие схемы КП, причем, как правило, ряд свойств этих элементов не известен и подлежит определению в процессе консультирования, а следовательно,

$$S_n \cap S_p \neq \emptyset.$$

Подмножество S_p составляет как структуры, разрабатываемые в процессе консультирования, так и некоторые элементы подмножества, исходная информация о которых является недостаточной для их материального воспроизведения.

Элементами подмножества S_p являются: схема связей между элементами структуры, отдельные модули КП, а также геометрические, механические, информационные и другие структуры.

Подмножество S_{np} составляют те структуры, которые не разрабатываются в процессе консультирования КП, а являются производными — вытекающими из подмножества исходных и разрабатываемых структур. Очевидно, что для всего подмножества производных структур имеет место соотношение:

$$(S_n \cup S_p) \cap S_{np} = \emptyset.$$

При использовании *второго классификационного признака* все структуры КП будем разделять на подмножества S_k ($k = 1, 2, \dots, n$) по *природе связей структуры*, т. е.

$$S = \bigcup_{k=1}^n S_k.$$

Элементами S_k (подмножествами) множества S являются электрические, кинематические, информационные, функциональные, геометрические и другие структуры.

При использовании *третьего классификационного признака* структуры КП будем разделять на подмножества *сложных структур* S_c (элементы которых имеют собственную структуру) и подмножество

элементарных структур S_3 (элементы которых не имеют собственной структуры), причем

$$S_c \cup S_a = S; S_c \cap S_a = \emptyset.$$

Следует отметить, что деление структур на сложные и элементарные является условным и определяется задачами и целями консультирования КП.

При использовании *четвертого классификационного признака* все структуры КП по *признаку направления действия связей* будем разделять на подмножества *ориентированных* S_o , *неориентированных* S_{no} и *частично ориентированных* $S_{чо}$ структур, структур с обратными связями S_{oc} и т. д., причем

$$S_o \cup S_{no} \cup S_{чо} = S;$$
$$S_{oc} \subset S.$$

КП как каждая большая система в развитии проходит последовательно несколько этапов — от идеи создания до демонтажа.

Первый этап — появление идеи формирования рекомендаций по решению задач КП, когда определяется ее назначение. На этом этапе, как правило, уже известно, к какому классу будет принадлежать консультируемая проблема и после формулировки основных функций определяется и семейство, к которому она относится. Если заранее известен и комплекс условий, то неопределенность в выборе вида системы уменьшается.

Второй этап — выработка КЗ на формирование рекомендаций по решению задач КП. Основная задача этапа состоит в том, чтобы определить различные показатели КП. Определение всего множества характеристик может вызвать определенные трудности, поскольку научно-обоснованных методик для этой процедуры еще практически не разработано. При назначении характеристик обычно руководствуются следующим: основные характеристики КП должны быть лучше, а остальные не хуже, чем у ее прототипа, что является явно недостаточным аргументом и руководящим началом.

Третий этап — предварительное формирование рекомендаций. Его основная задача — формирование рекомендаций по определению структуры КП и алгоритма (процесса) ее функционирования. Формализация процесса консультирования пока решена слабо, и в основе деятельности консультанта лежит, как правило, интуиция, практический опыт и поиск аналогий.

Четвертый этап — формирование рекомендаций, когда отыскивается рекомендация, определяемая структурой КП и множеством элементов.

Пятый этап — оценка эффективности сформированных рекомендаций по решению задач КП и ее подпроблем.

Шестой этап — оценка соответствия сформированных рекомендаций заданным.

Седьмой, заключительный, этап — функционирование. В процессе функционирования КП можно правильно оценить или рассчитать по полученным статистическим данным все эксплуатационные характеристики КП, а после некоторого периода функционирования КП дать правильные выводы о действительной эффективности использованных рекомендаций.

Вынашивая идею формирования рекомендаций по решению задач КП, необходимо прежде всего принять решение относительно того, формировать рекомендации или не формировать. Выдавая консультационное задание на формирование рекомендаций, необходимо принять решение относительно значений различных характеристик КП. Приступая к консультированию, следует принять решение относительно структуры будущей КП и т. д.

Всем процессам консультирования присущи две основные черты:

1) выбор из множества всевозможных рекомендаций одной, вполне определенной;

2) рекомендация формируется ради определенной цели.

Из сказанного следует, что **каждый процесс формирования рекомендаций можно описать функцией, аргументами которой являются допустимые варианты рекомендаций, а значениями — числа, которые описывают меру достижения поставленной цели.** Эту функцию принято называть **целевой**.

Задача формирования рекомендации сводится тем самым к нахождению максимального (минимального) значения целевой функции, а также к нахождению того конкретного аргумента, при котором это значение достигается.

В результате в каждом процессе формирования рекомендаций возникают две проблемы:

1) описать множество допустимых рекомендаций и целевую функцию;

2) найти максимум целевой функции и допустимую рекомендацию, реализация которой позволит осуществить этот максимум.

Первая проблема является задачей математического описания условий, в которых протекает процесс формирования рекомендаций, а также цели, ради которой он проводится.

Различные варианты второй проблемы называют экстремальными задачами.

Следует заметить, что составление математических описаний, математических моделей КП требует прежде всего основательных знаний в области функционирования КП, а также опыта математических рассуждений. Наиболее перспективными в этом направлении исследований являются подходы к разработке модели КП с помощью специальных языков моделирования. Однако, использование языков моделирования не исключает создания чисто математических, аналитических моделей, в которых действие различных факторов представляется в некоторой абстрактной форме и учитываются только наиболее существенные черты исследуемой КП.

В таком же положении и вторая проблема оптимизации: для одной части задач существуют хорошо разработанные методы решения, для другой — их вовсе нет.

Наиболее хорошо разработаны методы оптимизации для тех математических моделей консультируемых проблем, которые ***функционируют в евклидовом или целочисленном пространстве***. В этом случае для решения задач оптимизации пригодны: методы вариационного исчисления, принципы оптимальности Понтрягина и Бэллмана и всевозможные их разновидности; классический анализ, линейное, нелинейное и динамическое программирование; целочисленные задачи линейного программирования, некоторые нелинейные задачи и некоторый класс задач на графах.

При анализе КП, используют два типа методик:

- 1) быстрого анализа КП, цель которого — установить целесообразность формирования рекомендаций по решению задач консультируемой проблемы, формы ее структуры и др.;
- 2) подробного обследования КП.

Следует отметить, что при консультировании проблемы не всегда проводится предварительное ее исследование, поэтому часто имеют случаи, когда неэкономичность и малая эффективность рекомендаций по решению задач КП выявляются уже после израсходования значительных средств и времени на формирование рекомендаций. Методика формирования рекомендаций становится методикой лишь в том случае, если метод, положенный в ее основу, содержит в себе алгоритм решения консультационной задачи, обеспечен исходными данными и дает возможность решить консультационную задачу за конечное (реальное) время. Для начальных этапов консультирования КП в настоящее время еще не разработаны эффективные методики автоматизированного консультирования КП. Такое положение можно объяснить отсутствием четко сформулированного основного направления, вокруг которого должны сгруппироваться все остальные

направления исследования КП на ранних этапах их консультирования. Таким стержневым направлением должен явиться **структурный анализ** и более высокая его ступень — **синтез оптимальных структур КП**.

Малое количество исходной информации на начальных этапах консультирования КП требует поиска таких моделей КП, которые были бы обеспечены исходными данными и «работали» бы при минимуме исходной информации. **Такой моделью является структура КП совместно с совокупностью отношений на ней.**

Проведение структурного анализа позволяет:

- получить информацию о степени «нагруженности» и значимости элементов КП;
- сравнить КП с различными структурами;
- получить информацию о «слабых местах» КП, что дает возможность своевременно произвести доработку структуры КП;
- скорректировать программу обеспечения требуемых характеристик и качества КП.

Таким образом, можно сделать вывод, что **методика исследования КП, включающая элементы структурного анализа, относится к классу методик предварительного быстрого изучения проблемы, цель которых — установить целесообразность принятой рекомендации по решению задачи консультируемой проблемы.**

8.13. Анализ КП на начальных стадиях консультирования

Выбирая модели КП на начальных стадиях консультирования необходимо руководствоваться практическими соображениями, а именно: модель КП должна быть проще самой КП во всех аспектах, за исключением тех, которые определяют выполнение выбранного отношения эквивалентности.

Выбор модели КП во многом предопределяет методику исследования и получение конечных результатов – сформированных рекомендаций по решению задач консультируемой проблемы. При формировании рекомендаций по решению задач консультируемой проблемы на начальных стадиях возникает вопрос, какую абстрактную систему выбрать в качестве ее модели. Этот выбор будет определяться и обуславливаться следующими обстоятельствами: объемом исходной информации о КП; требуемой степенью детализации конечных результатов исследования. Все это в целом определит уровень абстракции при проведении исследования КП. Необходимо иметь в виду, что на каждом этапе познания имеется определенная степень

абстракции, а каждый этап познания, имея объектом исследования присущую ему модель, дает вполне определенные, соответствующие именно этому этапу, результаты.

Любая консультируемая проблема, которая представима системой, в процессе эволюции, прежде чем стать системой проходит ряд этапов. Каждый из этапов по-своему важен и по-своему специфичен. Однако представляется, что наибольшее внимание должно быть обращено на ранний этап создания, поскольку здесь формируются основные контуры будущей системы. Кроме того, на этом этапе должна быть сформированы рекомендации о целесообразности построения системы, о путях реализации желаемых свойств на этапе конструктивно-технологической разработки системы. Следовательно, начальный этап определит, что будет заложено в будущую систему. От правильно сформированных рекомендаций во многом зависит стоимость системы, так как все последующие изменения, доработки и доделки обходятся очень дорого, не принося подчас должного эффекта.

Чтобы сформировать эффективные рекомендации, необходимо рассмотреть и исследовать ряд вариантов, поскольку процесс окончательного создания определенной системы с желаемыми свойствами включает в себя выбор. Следовательно, модель, являющаяся эквивалентом изучаемой проблемы, должна быть относительно простой.

Характерной особенностью начального этапа консультирования является ограниченность информации о КП. Исходные данные, которые можно использовать в процессе исследования, обычно содержат: общие требования к характеристикам КП; структурно-функциональной схемы КП с весьма общим описанием принципа ее действия.

В таких условиях стремятся получить максимум возможного из этих минимальных сведений, что заставляет в первую очередь обратиться к структурно-функциональной схеме и содержащейся в ней информации.

Основным и общим во всех работах, связанных со структурными исследованиями, является стремление очистить объект исследования от всего вторичного, рассмотреть лишь его наиболее характерные признаки. Так, при структурных исследованиях КП в них не остается иного содержания, кроме связей, их числа, дифференциального порядка, знака и конфигурации. При исследовании проблем на ранних этапах их консультирования объект исследования оказывается «очищенным» естественным образом из-за недостатка информации о его свойствах. В подобных условиях исследуются элементы проблемы и существующие между ними связи, т. е. структурная схема КП.

Анализ КП требует использования основ математической теории отношений, которая оперирует о произвольными по своей природе объектами.

Теория отношений является составной частью теории множеств. Рассмотрим элементы теории отношений, необходимые для более глубокого понимания существа выбранной модели КП и правил оперирования о нею.

Отношением в множестве M называют закон, который выделяет некоторые P элементов (P -парные отношения) из M с указанием, какой из них является первой компонентой, какой второй и т. д.

В дальнейшем будут использоваться лишь бинарные отношения. Под бинарным понимают закон, который выделяет некоторые пары элементов (X, Y) из M с указанием, какой из них считается первой компонентой пары и какой второй компонентой.

Совокупность отношений представим матрицами отношений. В свою очередь матрицы отношений могут быть истолкованы как аналитический образ графа, или, наоборот, граф может быть истолкован как геометрический образ матрицы отношений, что позволяет совокупности отношений в системе поставить в соответствие граф отношений и изучать в дальнейшем только его свойства. Теория графов позволяет разработать общие формальные приемы исследования конкретной КП, не зависящие от ее сложности.

Информацию, содержащуюся в некотором графе, можно представить в алгебраическом виде матрицей отношений. Эта связь графа и матрицы отношений имеет важное значение, так как позволяет перевести структурные особенности КП на язык чисел, фигурирующих в математических уравнениях, описывающих структуры КП. При изучении структур КП с помощью графов необходимо помнить, что два графа обладают одной и той же структурой, если в одном графе столько же вершин и ребер, как и в другом, и если вершины и ребра одного графа соединены друг с другом так же, как и в другом графе. Равноструктурные графы называются эквивалентными.

Рассмотрим ряд правил оперирования с совокупностью отношений как математическим объектом. Доказано, что бинарные отношения на фиксированном множестве образуют булеву алгебру. В связи с этим операции конъюнкции, дизъюнкции и отрицания определяются обычным способом и все тождества булевой алгебры применимы к алгебре отношений.

Выводы для практики исследования КП, следующие отсюда, позволяют отметить: используя тождественные формулы булевой алгебры, можно отыскивать новые конфигурации структурных

схем КП, эквивалентные исходной, но более удобные для анализа. Кроме того, поиск эквивалентных структур позволяет найти и указать общие свойства различных типов и видов КП, на первый взгляд, различных по своей структуре.

Известно также, что матрицы отношений можно складывать и умножать как обычные матрицы. Отношения представимы также алгебрами нумероидов.

Нумероидом называют множество A с определенными на нем сложением $x+y$ и умножением $x \cdot y$, если:

1) сложение и умножение суть однозначные операции, применимые к любым элементам из A ;

2) сложение и умножение ассоциативны и коммутативны:

$$x+(y+z) = (x+y)+z; x(yz) = (xy)z; x+y = y+x; xy = yx.$$

3) сложение имеет хотя бы один нуль, такой, что $0+x = x+0 = x$ для всех x из A , а умножение — хотя бы одну единицу, такую, что $1x = 1x = x$ для всех x из A ;

4) умножение дистрибутивно относительно сложения:

$$x(y+z) = xy + xz.$$

Нумероид называют абсорбентом, если сложение в нем абсорбтивно по отношению к умножению, т. е., если для x и y из нумероидов справедлив закон абсорбции (поглощения) $x + xy = x$.

Вывод для практики, следующий отсюда, сводится к тому, что, если рассматривается вопрос только лишь о наличии либо отсутствии отношений между элементами КП, то все арифметические операции в алгебре отношений должны подчиняться правилам проведения операций в абсорбентах, т. е.

$$1 + x = 1; 0 \cdot x = x \cdot 0 = 0; xx = x.$$

Это справедливо также и потому, что единичный поднумероид (т. е. содержащий только нули и единицы) есть двухэлементная булева алгебра.

Когда в совокупности отношений устанавливается логическое понятие « r влечет S » (или «имеет следствием»), определенное следующим образом: $r \subset S$, означающее, что из $\alpha r \beta$ следует $\alpha S \beta$, то могут потребоваться другие виды исчислений.

Действительно, если некоторые элементы находятся в отношении r , $\alpha r \subset S$, то эти элементы тем более находятся в отношении S .

В этом случае для исследования совокупности отношений можно привлечь одно из экстремальных исчислений, а именно, максимумпликативное. Это исчисление получается, если на множестве неотрицательных действительных чисел положить

$$x \hat{+} y = \max(x, y); x \hat{\cdot} y = xy; \hat{0} = 0; \hat{1} = 1.$$

Здесь знак « \wedge » (крышка) означает, что арифметические операции выполняются в одном из экстремальных исчислений.

Другим экстремальным исчислением, находящим применение, является миниаддитивное. Оно получается, если множество действительных чисел, включая и бесконечность, положить:

$$\begin{aligned}x \dot{+} y &= \min(x, y); \quad x \dot{-} y = x \dot{+} y; \quad \min(\infty, x) = \min(x, \infty) = x; \\ \infty \dot{+} x &= x \dot{+} \infty = \infty.\end{aligned}$$

Поскольку все числа, которыми в последующем мы будем оперировать, являются положительными, то введенное в рассмотрение миниаддитивное исчисление будет сильным и абсорбентным.

8.14. Анализ структурно-топологических характеристик

Консультирование проблемы начинается с составления структурных схем, которые характеризуются множеством качественных свойств, обуславливаемых множеством элементов КП и связей между ними. Как уже отмечалось, структурная схема КП — это условное графическое изображение элементов КП и связей между ними, которое удобно анализировать с помощью графов. При таком подходе в ряде случаев можно найти адекватные задачи в теории графов, что позволяет воспользоваться для решения задач консультирования известными математическими методами.

Граф дает наглядное представление порядка и вида отношений между элементами двух множеств, что еще раз указывает на целесообразность его использования для изображения структурных схем КП. При решении одних задач элементам КП можно поставить в соответствие ребра графа, а связям — его вершины; при решении других задач элементам КП ставят в соответствие вершины графа, а связям между ними — ребра. Граф, полученный в первом случае, будем называть реберным, а во втором случае — вершинным. Эти два подхода не исключают друг друга. Даже при исследовании одной и той же проблемы возникает необходимость преобразовывать один граф в другой.

Рассмотрим консультируемую проблему, которую представим структурной схемой, например, производственной системы (рис. 8.15). Наиболее общую модель производственной системы можно получить, если расчленить ее на функциональные подсистемы. В нашем примере производственная система расчленена на само производство, систему управления производством и обеспечение ресурсами производства.

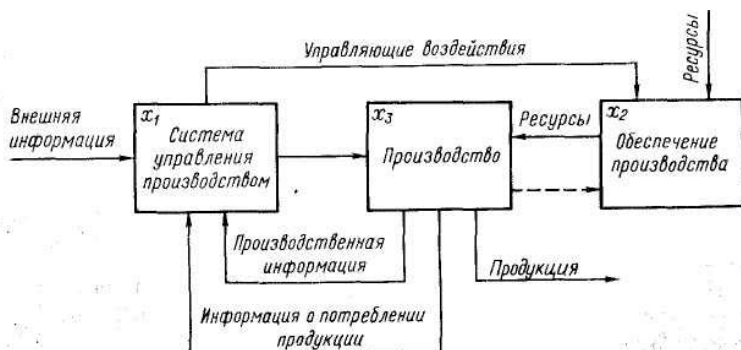


Рис. 8.15. Структурная схема производственной системы

Представим рассматриваемую структурную схему производственной системы в форме графа: подсистемы этой системы — вершинами графа, а потоки информации, ресурсов и продукции — ребрами графа. На рис. 8.16 изображена графовая модель структурной схемы рассматриваемой производственной системы.

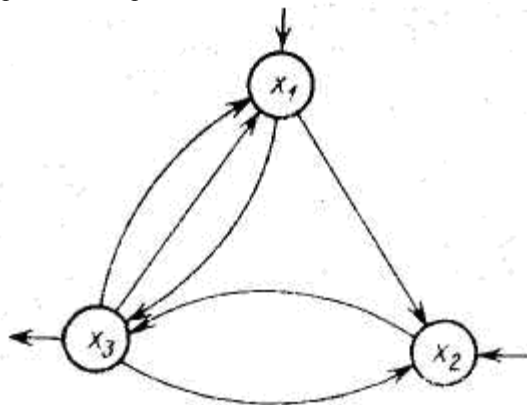


Рис. 8.16. Граф структурной схемы производственной системы

Данный пример иллюстрирует методику построения вершинного графа.

Модели сложных систем, к которым относятся и КП, представляемые в виде вершинных графов, имеют следующий недостаток: как физическое содержание отдельных элементов, так и логические условия их осуществления объединены в одних и тех же элементах —

в вершинах графа. Это обстоятельство чрезвычайно затрудняет анализ, делая его индивидуальным для каждой структуры, представленной вершинным графом.

Переход от вершинных графов к реберным дает возможность придать все физические свойства элементов дугам графа, а все логические условия сосредоточить в вершинах. Это существенно упрощает формирование логической структуры КП и позволяет разработать полностью формализованные методы построения структурных схем КП с вершинами, в которых могут реализоваться любые логические функции.

Как известно, в теории графов существуют две классические задачи: задача Эйлера о кенингсбергских мостах и задача Гамильтона. Задача Эйлера состоит в отыскании путей, проходящих по ребрам графа, а задача Гамильтона состоит в отыскании путей, проходящих по вершинам графа.

Таким образом, задачу Эйлера можно рассматривать как задачу упорядочения дуг реберного графа, а задачу Гамильтона — как задачу упорядочения вершин вершинного графа.

Доказанные Эйлером теоремы позволяют однозначно по виду графа решать вопросы о существовании эйлеровых путей. Для гамильтоновых путей такого критерия не существует. Для большинства таких графов не разработаны даже удовлетворительные алгоритмы, позволяющие установить наличие гамильтоновых путей.

Таким образом, возникает противоречие: анализ сложной проблемы позволяет установить перечень элементов проблемы и определить систему бинарных отношений на множестве этих элементов, т. е. построить матрицу непосредственных путей. Матрица непосредственных путей всегда позволяет построить ориентированный вершинный граф, анализ которого (т. е. поиск целесообразных путей, проходящих через его вершины) связан с очень большими, иногда непреодолимыми трудностями.

В то же время матрица непосредственных путей в большинстве случаев не позволяет непосредственно построить ориентированный реберный граф, анализ которого в принципе всегда возможен.

Построение реберного графа, эквивалентного вершинному, подчинено более сложным закономерностям. Ниже будут рассмотрены правила и порядок их использования при построении реберного графа. При этом будем рассматривать только ориентированные графы. Если исходный вершинный граф неориентированный, то его необходимо преобразовать в ориентированный с помощью операции удвоения,

когда любой неориентированный путь представляется парой противоположно ориентированных путей (рис. 8.17).

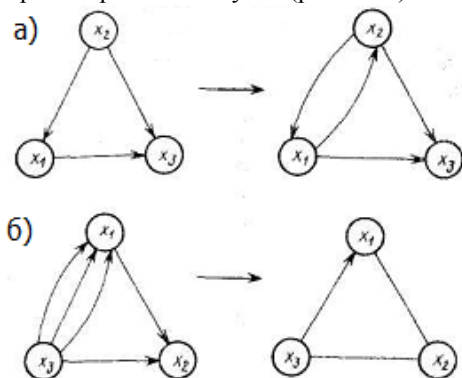


Рис. 8.17. Преобразование графов

Далее имеющиеся в исходном вершинном графе параллельные пути объединяются в один путь на основании правил операций в адсорбентах.

Построим реберный граф, эквивалентный вершинному, воспользовавшись графом структуры производственной системы (рис. 8.16).

Объединяя параллельные пути, идущие из третьей вершины в первую, получим преобразованный граф следующего вида (рис. 8.18).

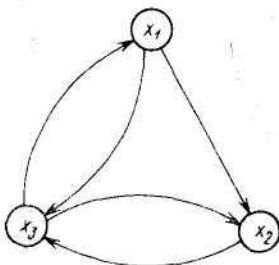


Рис. 8.18. Преобразованный граф производственной системы

Запишем для него матрицу отношений:

$$\|a_{ij}\| = \begin{array}{c} \begin{array}{ccc} x_1 & x_2 & x_3 \\ \hline & 1 & 1^* \\ \hline & & 1 \\ \hline 1 & 1^* & \\ \hline \end{array} \begin{array}{l} x_1 (2) \\ x_2 (1) \\ x_3 (2) \end{array} \end{array} \quad (8.139)$$

В пустых клетках здесь и далее будут подразумеваться нули. По матрице (8.139) построим еще две матрицы $\|b_{ij}\|$ и $\|c_{ij}\|$. Элементы первой из них определим по формуле

$$b_{ij} = a_{ij} \left(\sum_{i=1}^n a_{ij} + \sum_{j=1}^n a_{ij} \right), \quad (8.140)$$

где

$$\sum_{i=1}^n a_{ij}, \quad \sum_{j=1}^n a_{ij}$$

— суммы единиц по строкам и столбцам матрицы $\|a_{ij}\|$; их значения проставлены в круглых скобках в (8.139).

Производя вычисление по (8.140), получим матрицу $\|b_{ij}\|$ следующего вида:

$$\|b_{ij}\| = \begin{array}{c} \begin{array}{ccc} x_1 & x_2 & x_3 \\ \hline & 4 & 4 \\ \hline & & 3 \\ \hline 3 & 4 & \\ \hline \end{array} \begin{array}{l} x_1 \\ x_2 \\ x_3 \end{array} \end{array} \quad (8.141)$$

Элементы матрицы c_{ij} вычислим по формуле:

$$c_{ij} = a_{ij} (\delta_{ij} b_{ij} + \delta_{i/j} b_{ij}), \quad (8.142)$$

где $\delta_{ij} b_{ij} = (b_{ij} - \min b_{ij})i$ — превышение суммы полустепеней исхода и захода данного элемента $a_{ij}=1$ в i -й строке по сравнению с элементом, имеющим в этой строке минимальное значение $b_{ij} \neq 0$; $\min b_{ij}$ — минимальное по j в i -й строке значение суммы полустепеней захода и исхода для элементов $a_{ij} = 1$; $\delta_{i/j} b_{ij} = (b_{ij} - \min b_{ij})j$ — превышение суммы полустепеней исхода и захода данного элемента $a_{ij} = 1$ в j -м столбце по сравнению с элементом, имеющим в этой строке минимальное значение $b_{ij} \neq 0$; $\min b_{ij}$ — минимальное по i в j -столбце значение суммы полустепеней захода и исхода для элементов $a_{ij}=1$.

Подсчитаем величины элементов матрицы $\|c_{ij}\|$.

$$\begin{aligned}
 c_{12} &= a_{12} (\delta_{2/1} b_{12} + \delta_{1/2} b_{12}); \\
 \delta_{2/1} b_{12} &= (b_{12} - \min_2 b_{12})_1 = 4 - 4 = 0; \\
 \delta_{1/2} b_{12} &= (b_{12} - \min_1 b_{12})_2 = 4 - 4 = 0; \\
 c_{12} &= 0;
 \end{aligned}$$

$$\begin{aligned}
 c_{13} &= a_{13} (\delta_{3/1} b_{13} - c_{13} b_{13}); \\
 \delta_{3/1} b_{13} &= (b_{13} - \min_3 b_{13})_1 = 4 - 3 = 1;
 \end{aligned}$$

$$\delta_{1/3} b_{13} = (b_{13} - \min_1 b_{13})_3 = 4 - 4 = 0; \quad c_{13} = 1.$$

Продолжая вычисления, получим, что $c_{23} = 0$, $c_{32} = 1$, $c_{31} = 0$. Таким образом, окончательно имеем

$$\|c_{ij}\|_1^3 = \begin{array}{|c|c|c|} \hline x_1 & x_2 & x_3 \\ \hline & 0 & 1 \\ \hline & & 0 \\ \hline 0 & 1 & \\ \hline \end{array} \begin{array}{l} x_1 \\ x_2 \\ x_3 \end{array} \quad (8.143)$$

Пометим в матрице $\|a_{ij}\|_1^3$ звездочками те элементы, для которых c_{ij} в матрице (8.143) не равны нулю. Такими элементами будут a_{32} и a_{13} . Расширим матрицу $\|a_{ij}\|_1^3$, дополнив ее двумя строками и столбцами, а элементы a_{13} и a_{32} заменим парой новых, как это показано штриховой линией в матрице (8.144).

$$\|a_{ij}\|_i^5 = \begin{array}{c|ccccc|c} & x_1 & x_2 & x_3 & x_4 & x_5 & \\ \hline & & 2 & \bullet \dashrightarrow 1 & & & x_1 \\ \hline & & & \vdots & & & x_2 \\ \hline 1 & \bullet \dashrightarrow 1 & \dashrightarrow 1 & \dashrightarrow 1 & \dashrightarrow 1 & & x_3 \\ \hline & & \vdots & \vdots & & & x_4 \\ \hline & & \vdots & 1 & & & x_5 \\ \hline & & \vdots & & & & \\ \hline & & 1 & & & & \\ \hline \end{array} \quad (8.144)$$

По матрице $\|a_{ij}\|_i^5$ вновь построим матрицы $\|b_{ij}\|_i^5$ и $\|c_{ij}\|_i^5$.

Они будут иметь вид:

$$\|b_{ij}\|_i^5 = \begin{array}{c|ccccc|c} & x_1 & x_2 & x_3 & x_4 & x_5 & \\ \hline & & 4 & & & 3 & x_1 \\ \hline & & & 3 & & & x_2 \\ \hline 3 & & & & & 3 & x_3 \\ \hline & & & 3 & & & x_4 \\ \hline & & 3 & & & & x_5 \\ \hline \end{array} \quad (8.145)$$

$$\|c_{ij}\|_i^5 = \begin{array}{c|ccccc|c} & x_1 & x_2 & x_3 & x_4 & x_5 & \\ \hline & & 2 & & & & x_1 \\ \hline & & & & & & x_2 \\ \hline & & & & & & x_3 \\ \hline & & & & & & x_4 \\ \hline & & & & & & x_5 \\ \hline \end{array} \quad (8.146)$$

У матрицы (8.146) только один элемент $c_{ij} = c_{12}$, соответствующий $a_{ij} = a_{12} = 1$, не равен нулю.

Дополним матрицу (8.144) еще одной строкой и одним столбцом, а элемент a_{12} заменим парой новых

$$\left\| a_{ij} \right\| = \begin{array}{c|cccccc} & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ \hline x_1 & & \bullet & \text{---} & \text{---} 1 \text{---} & \text{---} & \text{---} 1 \text{---} \\ \hline x_2 & & \vdots & 1 & & & \\ \hline x_3 & 1 & \vdots & & & 1 & \\ \hline x_4 & \bullet & \vdots & 1 & & & \\ \hline x_5 & & \vdots & & & & \\ \hline x_6 & & \vdots & & & & \\ \hline \end{array} \quad (8.147)$$

Процесс преобразования исходной матрицы (8.140) заканчивается на (8.147), так как построенная для нее матрица $\|c\|_1^6$ будет иметь вес $c_{ij} = 0$ для соответствующих им $a_{ij} = 1$.

Таким образом, преобразование и расширение исходной матрицы бинарных отношений следует производить до тех пор, пока в соответствующей ей матрице $\|c_{ij}\|$ все элементы не станут равными нулю.

Дальнейший этап преобразования сводится к определению нумерации вершин нового реберного графа и распределению между ними ребер, эквивалентных исходному вершинному графу. Рассмотрим алгоритм определения нумерации вершин нового реберного графа.

1. Если в матрице $\|a_{ij}\|_1^n$ имеется больше, чем один пустой столбец, то к матрице слева и сверху приписываются еще один столбец и строка, причем новый столбец остается пустым, а в новой строке на местах, соответствующих пустым столбцам, проставляются единицы.
2. Если имеется более, чем одна пустая строка, то к матрице приписываются слева и снизу столбец и строка, причем приписанная строка остается пустой, а в новом столбце на местах, соответствующих пустым строкам, проставляются единицы.

3. Пустому столбцу присваивается первый номер. Этот же номер присваивается строке, имеющей тот же индекс, который проставляется справа от матрицы в колонке под начальным номером $N_{\text{нач}}$. Если в исходной матрице $\|a_{ij}\|_I^n$ нет пустых столбцов, то первый номер может быть присвоен произвольному столбцу.

4. Отметим первый столбец и опустимся по нему вниз до встречи с элементом, отличным от нуля. Вычеркиваем его. Затем, перемещаясь по строке, где был встречен первый элемент, отличный от нуля, вычеркиваем все элементы $a_{ij} \neq 0$. Если в очередном столбце нет более элементов, отличных от нуля, то ему присваивается тот же номер (индекс). Эти же номера (индексы) присваиваются тем же строкам. Они проставляются в колонке под рубрикой $N_{\text{нач}}$. Операция повторяется до тех пор, пока в отмеченном столбце не останется ни одного не вычеркнутого элемента. Затем переходят к следующему столбцу матрицы, присваивая ему очередной номер (индекс). Операция повторяется, пока в матрице $\|a_{ij}\|_I^n$ не останется ни одного невычеркнутого элемента.

5. Полученные номера (индексы) столбцов выписывают в виде дополнительной строки $N_{\text{нач}}$ под матрицей $\|a_{ij}\|_I^n$.

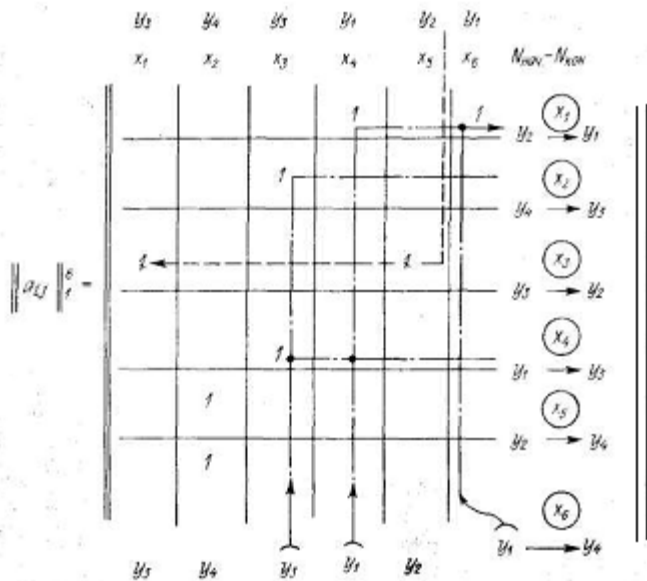
6. Составляют второй дополнительный столбец $N_{\text{кон}}$ (номер конечный) путем проектирования $N_{\text{нач}}$ на матрицу $\|a_{ij}\|_I^n$. Проектирование осуществляется следующим образом. Номер (индекс) из строки $N_{\text{нач}}$ перемещается по столбцу до встречи с $a_{ij} \neq 0$. Затем по этой строке номер (индекс) из строки $N_{\text{нач}}$ перемещается в столбец $N_{\text{кон}}$.

7. На этом процедура преобразования заканчивается. Номера (индексы) столбцов $N_{\text{нач}}$ и $N_{\text{кон}}$ указывают на вершины, которые связаны ребром с номером (индексом), соответствующим номеру (индексу) строки, бывшем ранее вершиной в исходном вершинном графе.

8. По полученной матрице вместе со столбцами $N_{\text{нач}}$ и $N_{\text{кон}}$ строится эквивалентный граф.

Осуществим с помощью описанного алгоритма преобразование матрицы (8.147). В ней нет строк и столбцов, содержащих только нули, поэтому процедуры, определенные первыми двумя пунктами алгоритма, отпадают. Выбор первого столбца осуществляется произвольно. Пусть им будет столбец 6. Присвоим ему индекс y_1 . Строке 6 также присваивается этот же индекс и проставляется в столбце $N_{\text{нач}}$. Элемент $a_{16} = 1$. Вычеркиваем его. Двигаясь по первой строке, вычеркиваем элемент $a_{14} = 1$. В столбце 4 нет больше отличных от нуля элементов. Ему также присваивается индекс y_1 и соответственно строке 4 индекс проставляется в столбце $N_{\text{нач}}$.

Переходим к следующему столбцу. Пусть им будет столбец 5. Присваиваем ему индекс y_2 (в строке 5). В столбце 5 отличен от нуля элемент a_{35} . Вычеркиваем его. Двигаемся по строке 3 до встречи с отличным от нуля элементом a_{13} . Вычеркиваем его. В столбце 1 нет больше ненулевых элементов. Ему присваивается индекс y_2 (в первой строке). Переходим к следующему столбцу. Пусть им будет столбец 3. Присвоим ему (и строке 3) индекс y_3 и т. д. Получим новую нумерацию в виде столбца $N_{нач}$. Запишем новую строку $N_{кон}$ и спроектируем ее на матрицу, что даст столбец $N_{кон}$.



По матрице (8.148) видно, что эквивалентный граф имеет четыре вершины y_1, y_2, y_3, y_4 , соединенные шестью ребрами.

Полученный граф эквивалентен исходному (см. рис. 8.12), поскольку в нем сохраняется та же система бинарных отношений.

Изложенная методика построения графа достаточно сложна для ее ручной реализации. Поэтому существует программа, ориентированная на реализацию ее на ЭВМ.

8.15. Анализ количественных характеристик структур консультируемых проблем

Проводя структурный анализ КП и их подпроблем, необходимо располагать методикой, позволяющей определять некоторые структурные характеристики КП и давать им количественную оценку.

Качественная характеристика оценки структур КП и их подпроблем необходима для определения на начальной стадии консультирования проблем степени пригодности данной структурной схемы для выполнения поставленных перед проблемой задач.

Весьма важными показателями при оценке структурных схем КП являются установление роли и значения отдельных элементов КП, определение наиболее уязвимых мест и слабых схемных решений. Кроме того, при разработке множества вариантов структурных схем КП и их подпроблем, возникает необходимость в качественном сравнении этих схем.

Представляя структурную схему КП или ее подпроблемы математической моделью в виде графа, еще ничего нельзя сказать относительно качественных показателей самой КП. Для определения качественных показателей структур КП в количественном выражении требуется разработать методику определения этих количественных выражений. Используя указанные методики, можно разработать алгоритмы и программы реализации этих методик в подсистеме автоматизированного анализа структурных схем КП.

Параметры структурной схемы КП и ее подпроблем, определяющие качество схемы, должны быть простыми и не требовать для своего определения данных больше, чем их содержится в самой структурной схеме. Для определения *качественных характеристик структур КП* и их подсистем можно ввести *следующие критерии*.

1. **Связность структуры.** Этот критерий позволяет определить отсутствие требуемых связей между элементами КП, «висящие» вершины и др.

Связность элементов структуры, если она представлена в форма ориентированного графа, можно определить с помощью матрицы связности $S = ||s_{ij}||$. Элементы этой матрицы вычисляются с использованием матрицы смежности этого графа:

$$A_{\Sigma} = \sum_{k=1}^n A^k.$$

В этом случае элемент

$$s_{ij} = \begin{cases} 1, & \text{если } a_{ij}^{\Sigma} \geq 1; \\ 0, & \text{если } a_{ij}^{\Sigma} = 0. \end{cases} \quad (8.149)$$

В случае представления структуры КП в форме неориентированного графа связность его элементов определяется с помощью следующего выражения:

$$\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_{ij} \geq n - 1; \quad i \neq j. \quad (8.150)$$

В этом выражении правая часть неравенства определяет необходимое минимальное число связей в структуре неориентированного графа, содержащего n вершин.

2. Структурная избыточность. Под структурной избыточностью будем понимать такой структурный параметр, который отражает превышение общего числа связей над минимально необходимым.

В соответствии с (8.150) структурную избыточность N будем определять из выражения

$$N = \frac{1}{2} \left[\sum_{i=1}^n \sum_{j=1}^n a_{ij} \right] \frac{1}{n-1} - 1.$$

Эту характеристику структуры КП будем использовать для косвенной оценки экономичности и эффективности КП. КП, представляемые моделью «полный граф», имеют максимальную избыточность. У таких КП величина параметра структурной избыточности больше нуля ($N > 0$). У КП с минимальной избыточностью этот параметр равен нулю ($N = 0$). Несвязные структуры характеризуются отрицательной величиной N ($N < 0$).

Отсюда следует, что наибольшей эффективностью обладают КП, у которых наибольшее численное значение параметра структурной избыточности.

В ряде задач анализа структурной эффективности КП будем использовать такой **структурный параметр**, который учитывает **неравномерность распределения связей** μ^2 .

Как известно, в неориентированном графе, имеющем m ребер и n вершин, равномерное распределение связей можно охарактеризовать средней степенью вершины $\bar{\rho} = 2m/n$. Введем понятие отклонения

$\rho_i - \bar{\rho}$, где $\bar{\rho}$ — действительная степень i -й вершины заданного графа.

Определим квадратичное отклонение заданного распределения степеней вершин от равномерного:

$$\begin{aligned} \mu^2 &= \sum_{i=1}^n (\rho_i - \bar{\rho})^2 = \sum_{i=1}^n \rho_i^2 - 2\bar{\rho} \sum_{i=1}^n \rho_i + 4m^2/n = \\ &= \sum_{i=1}^n \rho_i^2 - 2 \cdot 2m/n \cdot 2m + 4m^2/n = \sum_{i=1}^n \rho_i^2 - 4m^2/n. \end{aligned}$$

Структурный параметр μ^2 характеризует уровень «недогрузки» заданной структуры, имеющей m ребер и n вершин, в достижении максимальной связности. Этот показатель в относительном выражении можно использовать для сравнения различных вариантов структур КП.

3. Структурная компактность. Под структурной компактностью будем понимать величину D , отражающую общую структурную близость элементов между собой в системе. Близость двух элементов i и j между собой будем определять через минимальную длину пути для ориентированного графа и цепи для неориентированного. Численное значение величины D определим из выражения

$$D = \sum_{i=1}^n \sum_{j=1}^n d_{ij}, \quad (i \neq j).$$

Можно использовать не абсолютное значение параметра D , а относительное

$$D_{\text{отн}} = (D/D_{\min}) - 1,$$

где $D_{\min} = n(n-1)$ — минимальное значение компактности для структуры проблемы, промоделированной полным графом.

Структурную компактность можно характеризовать и другой характеристикой — диаметром структуры:

$$d = \max_{ij} d_{ij}.$$

Параметры $D_{\text{отн}}$, D и d дают интегральную оценку инерционности различных процессов в проблеме, а в случае равных значений μ^2 и N их возрастание отражает увеличение числа разделяющих связей в структуре, характеризуя тем самым снижение общей эффективности системы.

4. Степень централизации в структурах. Этот параметр характеризует степень распределения связей в структуре проблемы. Численное значение этого параметра будем определять из выражения:

$$\alpha = (n-1) (2z_{\max} - n) \frac{1}{z_{\max} (n-2)},$$

где z_{\max} — максимальное значение величины

$$z_i = \frac{D}{2} \left(\sum_{l=1}^n d_{il} \right)^{-1}, \quad i = \overline{1, n}, \quad i \neq j;$$

значение этого параметра колеблется в пределах от 0 до 1.

На рис. 8.19 изображены основные модели структур КП.

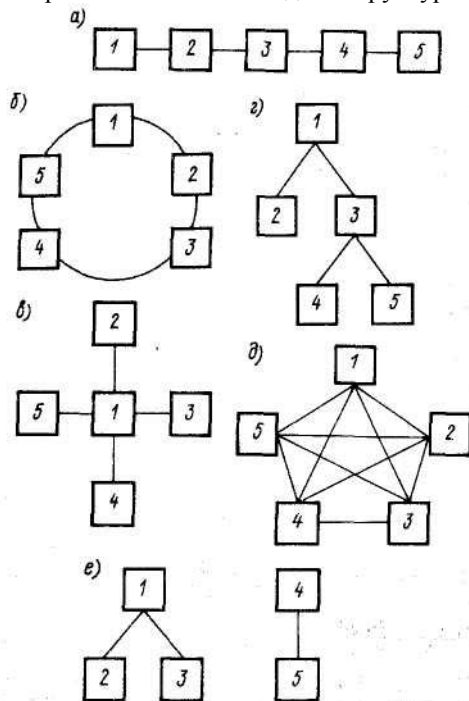


Рис. 8.19. Модели типовых структур КП

Структура, изображенная на рис. 8.19, в, имеет максимальную степень централизации — $\alpha=1$. При равномерном распределении связей величина степени централизации равна нулю (рис. 8.19, б и д)

5. Ранг элемента. Под рангом элемента будем понимать отношение рассматриваемого элемента графа структурной схемы КП с другими элементами этого графа.

Данный критерий позволяет распределить элементы структурной схемы КП в порядке их значимости. Значимость элемента здесь определяется только качеством связей данного элемента с другими.

Величина ранга еще не дает нам полной характеристики главенства того или иного элемента, поскольку не учитывает ряд других

характеристик элементов — точностных, скоростных и др. Однако характеризуя значимость элемента рангом — структурным критерием качества, можно высказать предположение, что чем выше ранг элемента, тем сильнее он связан с другими элементами схемы — тем тяжелее будут условия работы КП и ее подпроблем в случае выхода из строя элемента, обладающего наивысшей величиной ранга. Подробная методика определения ранга элемента приводится далее.

6. Множество сочленений. Под множеством сочленения структуры КП будем понимать такое множество ее элементов, удаление которых делает структуру несвязной. При подобных исследованиях может возникнуть вопрос, удаление каких ребер нарушит связность графа, что равносильно выходу из строя элементной структурной схемы КП.

Значимость элемента в этом случае можно определять и оценивать его принадлежностью к множеству сочленения, т. е. такому множеству, удаления которого из исходного графа делает его несвязным.

В дальнейшем будем говорить не вообще о множестве сочленения, поскольку граф в целом уже есть множество сочленения, а о минимальном множестве сочленения, так как консультанта в первую очередь интересует тот минимум элементов, удаление которых разрушит схему. Подробная методика определения множества сочленения будет приведена далее.

Приведенные критерии оценки качества структурных схем КП и их подпроблем позволяют количественно оценить их качество. Это позволяет оценить топологические свойства структур проблем. С точки зрения топологии внутренних связей можно выделить следующие основные типы структур (рис. 8.13): *a* — последовательная; *б* — кольцевая; *в* — радиальная; *г* — древовидная; *д* — типа «полный граф»; *е* — несвязная.

Рассмотрим пример применения количественных характеристик к анализу топологических свойств этих структур. Результаты анализа представлены в табл. 8.7.

Таблица 8.7

Вид структуры	<i>У</i>	<i>п²</i>	<i>D</i>	<i>α</i>	<i>R</i>
<i>a</i>	0	1,2	1,0	4	0,7
<i>б</i>	0,25	0	0,5	2	0
<i>в</i>	0	7,2	0,6	2	1,0
<i>г</i>	0	3,2	0,7	3	0,7
<i>д</i>	1,50	0	0	1	0
<i>е</i>	-0,25	—	—	—	—

Анализируя данные табл. 8.7, можно сделать следующие выводы.

1. Для несвязных структур $N < 0$, для структур без избыточности (последовательная, радиальная, древовидная) $N = 0$, для структур с избыточностью по связям (кольцевая, типа «полный граф») — $N > 0$.

2. Структуры (последовательная, радиальная, древовидная) с $N = 0$ различаются по показателю μ^2 . Наибольшую неравномерность связей имеет радиальная структура.

3. Наибольшую близость элементов (показатель D) имеет структура типа «полный граф», наименьшую — последовательная; радиальная и кольцевая структуры, неразличимые по показателю α , имеют различные значения D .

4. Радиальная и древовидная структуры, имеющие одинаковые или близкие значения N , D , α , значительно отличаются по показателю μ^2 и R , что соответствует физическому смыслу, ибо отход от полной централизации в структуре ведет к большой равномерности распределения связей по элементам.

Из приведенных критериев оценки структурных схем КП и их подпроблем особый интерес представляет структурный параметр — ранг элемента.

Ниже будет изложена методика определения ранга и полного ранга элемента структурной схемы КП и алгоритм реализации этой методики.

Рассмотренные выше структурные характеристики КП получают на основе информации о составе элементов и их связях. Дальнейшее развитие методологии построения структурных параметров для решения задач структурного анализа может быть основано на неструктурной информации за счет введения числовых функций на графах. Это позволяет наряду с составом элементов и направленностью их взаимодействия учитывать при решении задач структурного анализа другие интересующие стороны взаимодействия (временные, надежностные, стоимостные и др.).

Рассмотрим метод определения величины ранга элемента структуры. При решении задач структурного анализа КП и их подпроблем возникает вопрос, какой элемент структуры наиболее значимый. Иначе говоря, необходимо знать, какое влияние оказывает на общую надежность структуры КП выход из строя или иного ее элемента.

На начальных этапах консультирования достаточно определить значимость элемента некоторыми структурными параметрами, и, наоборот, анализировать и оценивать качество структурной схемы

этим структурными параметрами значимости. При дальнейшем исследовании КП и их подпроблем необходимо уточнить полученные на ранней стадии формирования схем показатели значимости, дополнив их параметрами, характеризующими функционирование принятых схем КП и определяющими эффективность их функционирования. На начальном этапе для такой оценки обычно данные отсутствуют.

Количественная оценка значимости сформулирована при анализе структуры отношения доминирования в теории математических отношений, которая оперирует с произвольными по своей природе объектами. В связи с этим некоторые ее положения могут быть перенесены и на теорию анализа структур КП.

Одним из критериев теории отношения является ранг элемента структуры объекта, которым определяется количество отношений этого элемента структуры с другими ее элементами, определяющее значимость элемента в структуре. Более высокие ранги соответствуют элементам, имеющим большее число связей с другими элементами структуры. Численно ранг элемента равен числу всех одночленных и двучленных отношений доминирования, которые данный элемент может осуществлять, т. е. сумме чисел всех ориентированных, непосредственных и двузвенных транзитных путей, соединяющих данный элемент с остальными элементами структуры. Ранги элемента будем обозначать буквой R с индексом, соответствующим номеру элемента на структурной схеме.

Из теории матричного исчисления известно, что матрица M , определяющая ранги элементов, может быть найдена из матрицы непосредственных путей A из выражения $M = A + A^2$.

Число непосредственных путей R_{in} , связывающих элемент с остальными элементами структуры, определяются по формуле

$$R_{in} = \sum_{i=1}^n a_{ij} \quad (8.151)$$

или в матричной форме

$$R_n = \left\| \begin{array}{c} \sum_{j=1}^n a_{1j} \\ \sum_{j=1}^n a_{2j} \\ \vdots \\ \sum_{j=1}^n a_{nj} \end{array} \right\|,$$

где n — ранг матрицы A .

Число двузвенных транзитных путей R_{iT_2} , связывающих элемент i с остальными элементами структуры, определяется по формуле:

$$R_{iT_2} = \sum_{j=1}^n [a_{ij}]^2 \tag{8.152}$$

или в матричной форме

$$R_{T_2} = \left\| \begin{array}{c} \sum_{j=1}^n [a_{1j}]^2 \\ \sum_{j=1}^n [a_{2j}]^2 \\ \vdots \\ \sum_{j=1}^n [a_{nj}]^2 \end{array} \right\|,$$

где $[a_{ij}]^2$ элемент, принадлежащий i -й строке и j -му столбцу матрицы A^2 .

На основании выражений (8.151) и (8.152) имеем:

$$R_i = R_{in} + R_{iT_2} = \sum_{j=1}^n (a_{ij} + [a_{ij}]^2)$$

или в матричной форме

$$R = R_n + R_{T_2} = \left\| \begin{array}{c} \sum_{j=1}^n (a_{1j} + [a_{1j}]^2) \\ \sum_{j=1}^n (a_{2j} + [a_{2j}]^2) \\ \vdots \\ \sum_{j=1}^n (a_{nj} + [a_{nj}]^2) \end{array} \right\|.$$

В тех случаях, когда различные пути между вершинами графа, отображающего рассматриваемую структуру, неравноценны, соответствующим звеньям приписываются некоторые коэффициенты (веса), учитывающие их характеристики по выбранному показателю. Ранг элементов в этом случае определяется аналогично рассмотренным выше приемам, но, в отличие от них, ранги отдельных (или всех) элементов структуры не выражаются целыми числами.

Приведенное выше определение понятия ранга элемента в ряде случаев оказывается недостаточным, так как оно не оценивает существования транзитных путей, содержащих более, чем два звена. В подобных случаях, наиболее часто встречающихся при оценке качества сложных структур, вместо этого понятия будем применять понятие — **полный ранг элемента**, определяемый как сумма всех ориентированных непосредственных и транзитных путей, соединяющий данный элемент с остальными элементами структуры.

Вычисление полного ранга элемента производится аналогично описанному выше с использованием машинных методов. Для определения полного ранга элементов последовательно находятся матрицы A , A^2 и т. д. до тех пор, пока все элементы одной из этих матриц не станут равными 0. Полный ранг элемента определим по формуле:

$$R_{ni} = \sum_{k=1}^m \sum_{j=1}^n [a_{ij}]^k$$

или в матричной форме

$$R_n = \begin{pmatrix} \sum_{k=1}^m \sum_{j=1}^n [a_{1j}]^k \\ \sum_{k=1}^m \sum_{j=1}^n [a_{2j}]^k \\ \dots \\ \sum_{k=1}^m \sum_{j=1}^n [a_{nj}]^k \end{pmatrix},$$

где $[a_{ij}]^k$ — элемент, принадлежащий i -й строке и j -му столбцу матрицы $A^k = ([a_{ij}]^k = a_{ij})$.

При оценке качества структуры по рассматриваемому показателю следует иметь в виду, что наличие одного или нескольких элементов с высокими рангами заставляет предъявлять к этим элементам повышенные требования по безотказности, поэтому при разработке

структур надо стремиться к тому, чтобы все элементы структуры имели близкие друг к другу и относительно невысокие ранги.

Оценивая соотношения элементов в КП рангом, можно определить не только их структурную значимость, но и функциональную. Для этого необходимо знать описание характера последствий изменения параметров того или иного элемента КП. Здесь достаточно иметь лишь качественное описание, примерно в такой форме: «изменение параметра первого элемента вызывает изменение параметров второго и третьего элементов ...» и т. д.

На основании словесного описания можно построить граф (рис. 8.20), отражающий взаимное влияние элементов в КП, и матрицу отношений.

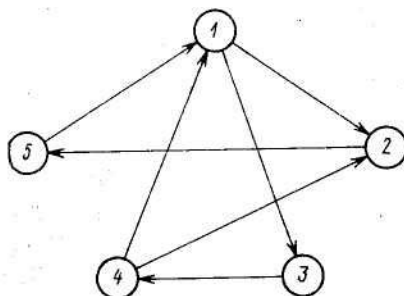


Рис. 8.20. Граф взаимного влияния элементов КП

В случае, если степень влияния точно количественно неизвестна, то на месте i -го элемента матрицы отношений ставится единица. Это означает лишь только то, что i -й элемент влияет на j -й. В том случае, когда степень влияния можно характеризовать некоторым числом или набором чисел, то в соответствующей клетке матрицы доминирования ставится этот весовой коэффициент. Нуль, как и прежде, означает отсутствие влияния элементов друг на друга. Порядок вычисления ранга, определяющего функциональную значимость элемента, остается тем же, что и при вычислении ранга, определяющего структурную значимость элемента.

При вложении конкретного содержания в понятие «отношение» может образоваться несколько групп рангов, определяющих структурную, функциональную и иные значимости элементов. Для определения уровня ранжировки, на котором требуется остановиться, воспользуемся **информационным подходом к решению поставленной задачи.**

Каждое ранжирование можно рассматривать как опыт с некоторым исходом (векторным) R . Компоненты этого вектора — ранги элементов системы, сумма которых равна единице.

Известно, что множество положительных дробей, в сумме составляющих единицу, можно рассматривать как соответствующие некоторому множеству элементы, которые обнаруживают разнообразие. В этом случае *энтропия H может служить средней мерой разнообразия на одну составляющую*. Условимся, что до ранжирования разнообразие по рангам элементов в КП оценивалось H_0 , а после ранжирования — H_1 . Тогда *величиной*

$$J_n = H_0 - H_1, \quad (8.153)$$

называемой *количеством информации, можно характеризовать эффект, даваемый ранжированием как действием, ограничивающим разнообразие (устраняющим неопределенность задачи)*.

Изменение неопределенности задачи под влиянием пришедшего сигнала можно интерпретировать как *процесс запасаения полезной информации*, если за нулевой уровень принять запас информации при равномерной значимости элементов (что соответствует равновероятностному исходу опыта). Тогда

$$J_n = \log n - H_1, \quad (8.154)$$

где n — число элементов КП, а

$$H_1 = - \sum_{i=1}^n R_i \log R_i. \quad (8.155)$$

Поскольку количество информации J достигает максимума при $R=1/n$, то любая первая ранжировка даст приращение объема полезной информации $J_n > 0$. Пусть имеются другие ранжировки. Для выявления целесообразности использования их результатов необходимо выполнить следующие действия. По формуле (8.155) подсчитать для всех k ранжировок значения H_k . Затем расположить H_k в порядке их убывания и подсчитать по формуле

$$\Delta J_n = H_k - (H_k - 1) \quad (8.156)$$

прирост полезной информации, даваемый каждой последующей ранжировкой относительно предыдущей. Задаваясь определенным уровнем прироста полезной информации ΔJ_n ранжировки, которые не дают прироста информации (возможен случай «дезинформации» при $\Delta J_n < 0$), из дальнейшего рассмотрения следует исключить.

Сравнивать эффект ранжировок следует по группе сравнимых, эквивалентных факторов, определяющих качество функционирования

КП. В подобной ситуации возможно получать конечные формулы, оценивающие прирост полезной информации.

Допустим, что в k -й ранжировке ранги неполно отражают значимость элементов, отличаясь от «истинных» на величину $\pm \Delta R_i$, а в $(k + 1)$ -й ранжировке определены истинные значения рангов. Тогда для i -х составляющих будем иметь в k -й ранжировке

$$R_i \log_2 R_i = \frac{R_{0i}}{\ln 2} (1 \pm \delta_i) [\ln R_{0i} + \ln (1 + \delta_i)] \approx \\ \approx \frac{R_{0i} \ln R_{0i}}{\ln 2} \pm \frac{R_{0i}}{\ln 2} \delta_i,$$

где $\delta_i = \Delta R_i / R_{0i}$;

в $(k + 1)$ -й ранжировке —

$$R_{0i} \log_2 R_{0i} = (1/\ln 2) R_{0i} \ln R_{0i}.$$

Отсюда, согласно (8.156), будем иметь

$$\Delta J_n \approx 1,5 \sum_{i=1}^n (\pm \Delta R_i). \tag{8.157}$$

Из (8.157) следует, что последующее ранжирование будет целесообразным в том случае, когда оно усиливает найденные значимости элементов, если при этом происходит более резкое разграничение элементов по их значимости.

Рассмотрим метод определения множества сочленения. Как отмечалось ранее, под множеством сочленения графа понимают такое множество его элементов, удаление которых делает граф несвязным.

Принадлежностью к множеству сочленения можно оценивать значимость как вершин, так и ребер графа. Однако большая сложность и трудоемкость определения принадлежности элемента к множеству сочленения делает предпочтительным ранжирование при определении значимости вершин графа — элементов КП. Для оценки значимости ребер графа, что соответствует оценке значимости связей в КП, не просматривается ничего иного, кроме множества сочленения. Поскольку оценить значимость ребер не представляется, возможно, ничем иным, кроме множества сочленения, то приходится мириться с неизбежными осложнениями.

Метод определения принадлежности элементов графа к множеству сочленения сводится к следующему. Пусть имеется граф, вершины которого соответствуют элементам КП, а ребра — отношениям (связям) между ними. Строится вспомогательный граф отображением множества вершин во множество ребер.

Построение вспомогательного графа осуществляется следующим образом: вершины вспомогательного графа, число которых равно

числу ребер исходного графа, размечаются индексами (буквами), соответствующими разметке ребер исходного графа (рис. 8.21); если в исходном графе ребра были смежными, т. е. имели общую вершину, то во вспомогательном графе соответствующие вершины соединяются ребром. Например, ребра y_1, y_2, y_3 исходного графа смежны (рис.8.21, а).

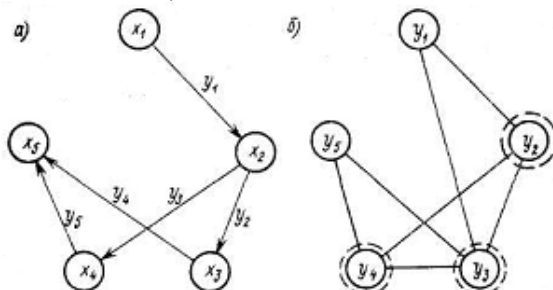


Рис. 8.21. Граф: а — основной; б — вспомогательный

Следовательно, во вспомогательном графе вершины y_1, y_2, y_3 следует соединить ребрами (рис. 8.21, б). Для определения принадлежности элемента к множеству сочленения следует придерживаться следующего порядка.

1. Определить центры и периферийные точки графа.

Рассмотрим некоторые определения. В конечном или бесконечном графе отклонением $\delta(x, y)$ его вершины x от вершины y называется длина кратчайшего пути из x в y . При этом $\delta(x, y)$ удовлетворяет условиям:

- а) $\delta(x, x) = 0$;
- б) $\delta(x, y) + \delta(x, z) > \delta(x, x)$;
- в) для симметричного графа $\delta(x, y) = \delta(y, x)$.

Отклоненностью вершины x называют число:

$$l(x) = \max_{y \in x} \delta(x, y).$$

Если наименьшая из отклоненностей является конечным числом, то вершина, в которой этот минимум достигается, называется **центром графа**, вершина с наибольшей отклоненностью — **периферийной точкой графа**. Исключаем из рассмотрения периферийные точки, так как они не могут быть точками сочленения.

2. Строится полная матрица связей. Последовательным перебором оставшихся вершин определяем, удаление каких именно из них нарушит связность графа.

Как видно, операция нахождения множества сочленения весьма трудоемка. Однако, если рассматривать реальные КП, которые имеют

ограниченное число входов и выходов, то размерность задачи значительно уменьшится. Поэтому, определяя связность графа и условия ее нарушения, не имеет смысла рассматривать возможность соединения любых двух вершин, а следует ограничиться лишь входами и выходами. Так, в приведенном примере (рис. 8.21) строго к множеству сочленения относятся элементы y_2, y_3, y_4 . Если же входом и выходом являются элементы x_1 и x_5 соответственно, то ко множеству сочленения достаточно отнести элементы y_3 и y_4 .

На рис. 8.22 представлена структура более сложной КП. Здесь две подпроблемы A и B связаны системой различных связей. Необходимо определить, удаление каких связей разрушит структуру КП.

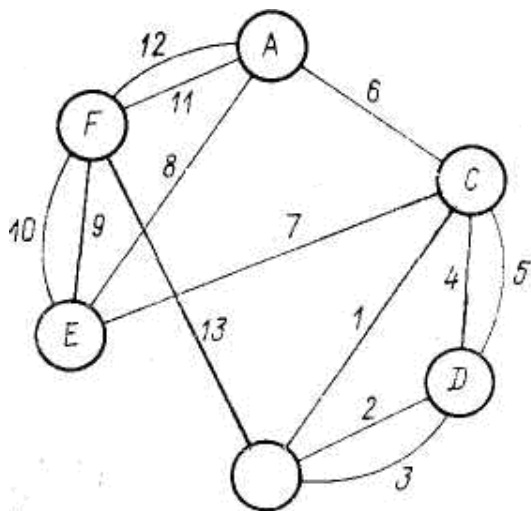


Рис. 8.22. Граф структуры сложной КП

Придерживаясь описанного порядка и проделав все необходимые операции, определим, что к множеству сочленения во вспомогательном графе (рис. 8.23) относятся вершины 6, 7, 13. Следовательно, разрушение связей (ребер основного графа) 8, 7, 13 разорвет и нарушит цепь функциональных связей между подпроблемами A и B .

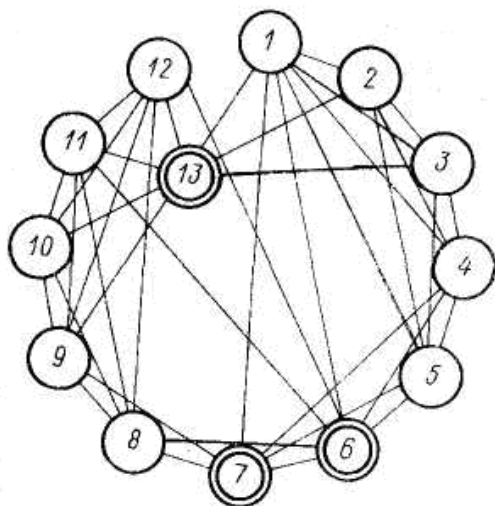


Рис. 8.23. Вспомогательный граф сложной КП

В заключение следует отметить, что множество сочленения, как структурный параметр, не вводит упорядочение в проблему отношений. Принадлежность к нему отвечает лишь на вопрос о возможности разрушения этой проблемы отношений.

К чему же необходимо стремиться при использовании критерия качества «множество сочленения»? При чисто структурных исследованиях, когда изучаются проблемы функциональных связей широкого плана, необходимо стремиться к тому, чтобы к множеству сочленения принадлежало как можно большее число элементов проблемы (ребер графа). Это достигается введением в структуру проблемы дополнительных связей, установления обходных путей, обычного резервирования. Чем большим будет в этом случае множество сочленения, тем более живуча консультируемая проблема, тем труднее нарушить ее связность.

Когда проблема отношений отражает взаимные влияния элементов друг на друга, то необходимо стремиться, чтобы к множеству сочленения принадлежало как можно меньшее число элементов. Достаточно изъять их, и вся система взаимных влияний нарушится.

Рассмотрение критериев, оценивающих качество структурной схемы КП, убедительно показывает, что применять их следует последовательно, идя от наиболее простого к более сложному, требующему большего количества данных о функционировании КП.

Окончательная оценка качества структурной схемы КП, оценка значимости элементов схемы в большинстве случаев является следствием компромиссного решения применительно к конкретной ситуации.

8.16. Декомпозиция структуры консультируемой проблемы

Произвести декомпозицию КП — это значит выделить в ней отдельные сильно связанные подпроблемы, т. е. такие подпроблемы, все остальные части которых благодаря обратным связям взаимно достижимы. Граф такой системы бисвязен. При декомпозиции выделяются также слабо связанные подпроблемы, все составные части которых связаны неориентированным путем. Граф такой подпроблемы связан.

Процесс декомпозиции КП можно формализовать сокращением ориентированного графа. Тогда для КП можно рекомендовать следующий алгоритм декомпозиции.

1. Составить матрицу смежности A графа $G(X,U)$.

2. Вычислить матрицу $R_1 = A + E$, где E — единичная матрица; «+» — знак логического сложения; R_1 — матрица первой достижимости, i -я строка которой представляет все ориентированные пути по графу из i -й вершины до всех остальных, если длина пути равна одному ребру.

3. Определить $R_2 = R_1^{*2}$, где знак «*» означает, что при вычислении $R_1 \times R_2$ применяется логическое умножение и суммирование элементов матриц.

Аналогично определяются все матрицы вплоть до $R = R_n = R_1^{*n}$, где R — матрица достижимости графа $G(X,U)$, i -я строка которой представляет все ориентированные пути по графу длиной от одного до n ребер, из i -й вершины ко всем остальным. Матрицы A и R имеют размерность $n \times n$.

При вычислении R не обязательно R_1 возводить в n -ю степень. Если $R_1^{*k} = R_1^{*(k-1)}$, то $R = R_1^{*k}$, где $k < n$.

4. Проанализировать матрицу R . Если $R = Q$, где $Q = [q_{ij}]$, такая универсальная матрица, что для всех i и j $q_{ij} = 1$, то граф бисвязен и декомпозиция системы невозможна, т. е. система состоит из одной сильно связанной подпроблемы. Если $R \neq Q$, то необходимо перейти к следующему шагу. В тех случаях, когда известно, что граф связан, следует перейти к шагу 8.

5. Определить матрицу достижимости неориентированного графа $G^0(X,U)$, соответствующего ориентированному графу проблемы

$G(X,U)$. Матрица $R^0=(A+A^T+D)^{*0}$, где знак T означает транспонирование.

6. Определить связные подграфы ориентированного графа $G(X,U)$. Известно, что множество вершин связного подграфа, содержащего вершину i , определено единицами в i -й строке матрицы R^0 . Если $R^0=Q$, то граф $G(X,U)$ состоит из одного связного подграфа. В этом случае следует переходить к шагу 8. Если же $R^0 \neq Q$, то надо перейти к следующему шагу.

7. Упорядочить вершины графа $G(X,U)$ (матрицы A) по связным подграфам.

8. Образовать матрицу связности $C=R+R^T$. Здесь сложение обычное, арифметическое.

9. Выделить из матрицы C бисвязные подграфы. Бисвязный подграф, содержащий вершину i , определен двойками в i -й строке матрицы C .

10. Упорядочить A так, чтобы бисвязные подграфы образовали квадратные подматрицы $E_\varphi \subset A$, $\varphi = 1, 2, \dots, p$.

11. Образовать матрицу $R_+=(A^0_++A^T_++E_+)^{*0}$, где A_+ —матрица смежности подграфа с множеством вершин

$V=W/\bigcup_{\varphi=1}^n B_\varphi$ и B_φ — подмножество составных частей φ -й сильно связанной подпроблемы, и выполнить п. 6 и 7.

12. По упорядоченной матрице A' определить связи (ребра) между подпроблемами, которые должны быть разорваны в результате декомпозиции.

Рассмотрим пример реализации описанного алгоритма. Пусть имеется КП, структура которой представлена графом (рис. 8.24).

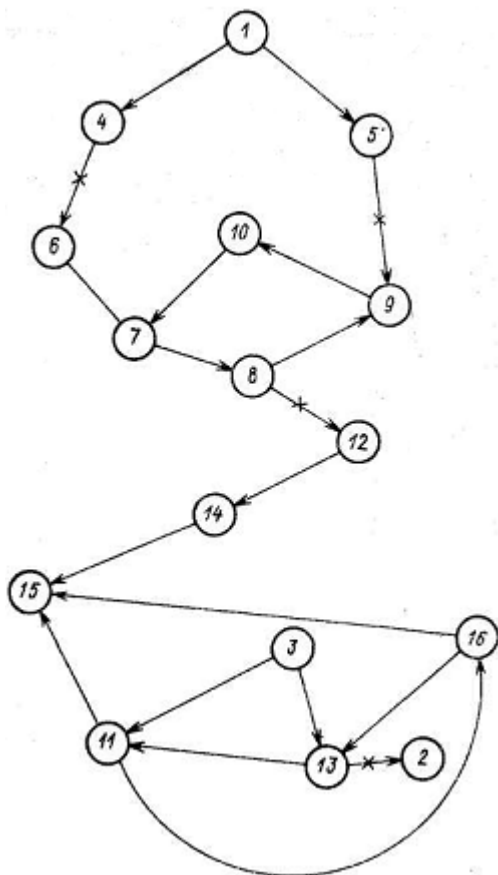


Рис. 8.24. Граф структурной схемы КП

Составим матрицу смежности A этого графа:

Следовательно, имеются две сильно связанные подпроблемы $B_1=6, 7, 8, 9, 10$ и $B_2=3, 11, 13, 15, 16$.

По матрице

$$R_+^0 = \begin{matrix} & \begin{matrix} 1 & 2 & 4 & 5 & 12 & 14 \end{matrix} \\ \begin{matrix} 1 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \\ 13 \\ 14 \\ 15 \\ 16 \end{matrix} & \begin{matrix} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ 1 & & & 1 & & \\ 1 & & & & 1 & \\ & & & & & 1 & \\ & & & & & & 1 & \\ & & & & & & & 1 & \\ & & & & & & & & 1 & \\ & & & & & & & & & 1 & \\ & & & & & & & & & & 1 & \\ & & & & & & & & & & & 1 & \\ & & & & & & & & & & & & 1 & \\ & & & & & & & & & & & & & 1 & \end{matrix} \end{matrix} \quad (8.162)$$

получаем еще три слабо связанные подпроблемы $S_1=\{1, 4, 5\}$, $S_2=\{12, 14\}$ и $S_3=\{2\}$.

Отсюда следует, что в упорядоченной матрице смежности графа КП необходимо разорвать связи, обозначенные на матрице звездочкой (на рисунке крестиками)

$$A^0 = \begin{matrix} & \begin{matrix} 1 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 12 & 14 & 3 & 11 & 13 & 15 & 16 & 2 \end{matrix} \\ \begin{matrix} 1 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 12 \\ 14 \\ 3 \\ 11 \\ 13 \\ 15 \\ 16 \\ 2 \end{matrix} & \begin{matrix} 1 & & & & & & & & & & & & & & & \\ & 1 & & & & & & & & & & & & & & \\ & & 1^* & & & & & & & & & & & & & \\ & & & 1^* & & & & & & & & & & & & \\ & & & & 1 & & & & & & & & & & & \\ & & & & & 1 & & & & & & & & & & \\ & & & & & & 1 & & & & & & & & & \\ & & & & & & & 1 & & & & & & & & \\ & & & & & & & & 1 & & & & & & & \\ & & & & & & & & & 1 & & & & & & \\ & & & & & & & & & & 1 & & & & & \\ & & & & & & & & & & & 1 & & & & \\ & & & & & & & & & & & & 1 & & & \\ & & & & & & & & & & & & & 1 & & \\ & & & & & & & & & & & & & & 1 & \\ & & & & & & & & & & & & & & & 1 \end{matrix} \end{matrix}$$

$\underbrace{\hspace{1.5cm}}_{S_1} \quad \underbrace{\hspace{2.5cm}}_{B_1} \quad \underbrace{\hspace{1.5cm}}_{S_2} \quad \underbrace{\hspace{2.5cm}}_{B_2} \quad \underbrace{\hspace{1.5cm}}_{S_3}$

После декомпозиции матрицу A' можно разрушить, сохранив лишь подматрицы смежности D_s и матрицу смежности сокращенного графа M , которые естественным образом получаются из A' .

Замкнутые контуры в подпроблемах B_1 и B_2 устраняются с использованием матриц смежности

$$\begin{aligned}
 D_{B_1} &= \begin{vmatrix} & 6 & 7 & 8 & 9 & 10 \\ & & 1 & & & \\ & & & 1 & & \\ & & & & 1 & \\ & & & & & 1 \\ 1 & & & & & \end{vmatrix} \quad \text{и } D_{B_2} = \\
 &= \begin{vmatrix} & 3 & 11 & 13 & 15 & 16 \\ & & & 1 & & \\ 1 & & & & 1 & 1 \\ & & 1 & & & \\ & & & & & 1 \\ & & & 1 & & \end{vmatrix}
 \end{aligned}$$

После декомпозиции и устранения замкнутых контуров можно представить КП сокращенным ориентированным графом $\Gamma(\Delta, \Omega)$, где Δ — множество вершин (подпроблем); Ω — множество ориентированных ребер (связей между подпроблемами). Сокращенный граф является удобной моделью для решения **динамических, информационных и диагностических проблем** при решении задач консультирования. Он обладает всеми основными топологическими свойствами исходной КП, поскольку КП преобразовывалась таким образом, что топологическое пространство, представленное ориентированным графом $G(X, U)$, непрерывно отображалось в топологическое пространство, представленное графом $\Gamma(\Delta, \Omega)$.

Так, для рассмотренного примера матрица смежности сокращенного графа имеет вид:

$$M = \begin{matrix} & B_1 & B_2 & S_1 & S_2 & S_3 \\ \begin{matrix} B_1 \\ B_2 \\ S_1 \\ S_2 \\ S_3 \end{matrix} & \left\| \begin{array}{cccccc} & & & & 1 & \\ & & & & & 1 \\ 1 & & & & & \\ & & 1 & & & \\ & & & & & \\ & & & & & \end{array} \right\| & , \end{matrix} \quad (8.162)$$

а сам граф $\Gamma(\Delta, \Omega)$ представлен на рис. 8.18.

Во всех матрицах рассмотренного примера, за исключением (8.160), в свободных клетках подразумеваются нули; в матрице (8.160) в свободных клетках единицы.

Изложенный метод декомпозиции основан только на анализе топологических свойств модели КП, т. е. на анализе свойств графа КП. Функциональная сторона анализа в данном случае не рассматривается, но предполагается, что она должна явиться логическим продолжением топологического анализа.

Следует отметить, что прием декомпозиции структурной схемы КП дает возможность осуществить не только преобразование графа КП к виду, удобному для решения последующих задач (появляется возможность существенно уменьшить размерность графа), но и одновременно появляется возможность до некоторой степени косвенно оценить качество структурной схемы. Так, рассмотренный алгоритм дает возможность указать те связи, разрыв которых приводит к распаду КП как единого целого на отдельные подпроблемы. Тогда, если смотреть на решение задачи с использованием указанного алгоритма с позиций **надежности и живучести исследуемой КП**, то становится очевидным, что они позволяют (в первом приближении, поскольку учитываются только топологические свойства) определить наиболее жизненно важные элементы связи КП.

В заключение следует отметить, что структурный анализ является необходимым этапом общего исследования КП. Требуя для своего проведения минимум исходной информации — структурную схему — его результаты позволяют в принципе сформулировать рекомендации, обеспечивающие функциональную разгрузку отдельных элементов, усиление слабых звеньев КП введением структурной и функциональной избыточности и позволяющие, наконец, целесообразно распределить усилия для обеспечения и достижения заданного качества функционирования КП. Рассмотр-

ренные методы и алгоритмы служат основой для подсистемы САК КП «Структурный анализ».

9. Синтез формируемых рекомендаций по решению задач консультируемых проблем в САК

9.1. Постановка задачи синтеза формирования рекомендаций

Понятие «синтез» формируемых рекомендаций в широком смысле слова близко по содержанию к понятию «консультирование». Задача синтеза формирования рекомендаций состоит в том, чтобы по заданному функциональному назначению консультируемой проблемы или по закону ее функционирования получить сформированные рекомендации по решению задач консультируемой проблемы в виде некоторого описания.

Синтез формирования рекомендаций нацелен на создание новых вариантов, а анализ используется для оценки этих вариантов, т. е. синтез и анализ выступают в процессе консультирования в диалектическом единстве.

При синтезе заданы допустимый набор используемых (функционирующих) элементов (элементов при синтезе рекомендаций по созданию электронных схем, набор балок и блочных конструкций при синтезе рекомендаций по проектированию и созданию строительных сооружений и т. д.), возможные правила их соединения между собой и способы определения по синтезированной структуре КП функции, которую она реализует.

Напомним, что под *структурой консультируемой проблемы* понимается набор составляющих ее элементов и связей между ними. Структура определяет, как устроена КП, из каких физических частей она состоит и как эти части связаны друг с другом.

К проблемам консультирования относятся структура и физически реализованная КП, причем под *физически реализованной КП* будем понимать материализованную (и не только) совокупность соединенных между собой элементов, выполняющих заданные функции.

На структуру и физическую реализацию любой КП накладывается множество различных ограничений. При этом одна группа ограничений относится к методу решения консультационной задачи и охватывает такие вопросы, как наличие знаний, сроки и имеющиеся в

распоряжении технические средства консультирования. Другая группа ограничений связана с требованиями КЗ на параметры КП, с требованиями стандартов и консультационных процессов формирования рекомендаций по решению задач КП и их подпроблем. Третья группа ограничений формируется физическими принципами реализации закона функционирования КП и получения ее предельно желаемых характеристик. Дополнительные ограничения накладываются способами и формами взаимодействия КП с внешней средой, а также методами организации взаимодействия консультанта с КП в процессе функционирования.

При консультировании проблемы выбирается или задается в КЗ критерий оптимальности. В зависимости от характера и назначения КП в качестве критерия оптимальности может быть принята ее стоимость, КПД, потребляемая мощность или другой, более сложный критерий.

При консультировании проблем важное значение имеет определение оптимальных вариантов структур и функций КП и подпроблем, параметров схем, режимов функционирования КП и т. д. Под *оптимальным* будем понимать такой вариант структуры или функции КП, параметры которой удовлетворяют всем системным, конструктивным, технологическим, электрическим и экономическим требованиям КЗ, а критерий оптимальности, описывающий качество структуры или функции, принимает наилучшее (минимальное или максимальное) значение.

Формирование оптимальных рекомендаций по решению задач консультируемых проблем становится возможным и достижимым как по срокам консультирования, так и по стоимости реализации консультационных процедур только при автоматизированном консультировании, когда появляется возможность синтезировать и исследовать множество вариантов структур и функций КП, а также проводить количественное изучение консультируемых проблем, которые в прошлом изучались лишь качественно.

Любому варианту КП соответствуют свои структура и функции. При автоматизированном консультировании для порождения множества альтернативных структур КП, эквивалентных по функциональному назначению, но различных по тактико-техническим характеристикам, необходима разработка математической модели КП, представляющей собой формальное описание КП на принятом уровне детализации.

При оптимальном консультировании необходимо обосновать критерий оптимальности КП и определить множество показателей $\theta = (v_1, \dots, v_n)$, на которые наложено множество ограничений

$$V=(V_1, \dots, V_n).$$

Для решения задачи синтеза рекомендаций по решению задач консультируемой проблемы будем выделять некоторую совокупность независимых переменных $\mathbf{X}=(x_1, \dots, x_m)$, фиксация значений которых определяет один из вариантов сформированной рекомендации и ее количественные характеристики, в том числе, значение критерия оптимальности, а также показателей, принятых в качестве ограничений.

Переменные x_1, \dots, x_m будем называть *переменными консультирования* и, в зависимости от физической природы КП, они имеют различную интерпретацию, в частности могут характеризовать количество элементов (подпроблем) каждого типа в КП, указывать на включение или на невключение той или иной подпроблемы в структуру КП, представлять геометрические размеры КП и т. д.

Критерий оптимальности $F(\mathbf{X})=F(x_1, \dots, x_m)$ и показали $\theta_i(\mathbf{X})=\theta_i(x_1, \dots, x_m)$, на значения которых наложены ограничения, являются функциями независимых переменных x_1, \dots, x_m .

В реальных задачах консультирования функция критерия оптимальности $F(\mathbf{X})$, функции ограничений $\theta_i(\mathbf{X})$, как правило, нелинейно зависят от значений множества переменных \mathbf{X} . Кроме того, во многих задачах консультирования на значения переменных накладывают условия целочисленности.

В формализованном виде задача синтеза рекомендаций заключается в определении значений независимых переменных x_1, \dots, x_m , при которых критерий оптимальности рекомендации

$$F(\mathbf{X}) = F(x_1, x_2, \dots, x_m) \quad (9.1)$$

принимает экстремальное (максимальное или минимальное) значение при условиях

$$\theta_i(x_1, \dots, x_m) \leq 0, \quad i = 1, 2, \dots, m; \quad (9.2)$$

$$a_j \leq x_j \leq b_j, \quad j = 1, 2, \dots, n. \quad (9.3)$$

Ограничения вида

$$\theta_i(x_1, \dots, x_m) \geq 0 \quad (9.4)$$

путем умножения левой части неравенства на минус единицу могут быть приведены к выражению (9.2).

Ограничения вида (9.2) и (9.4) в своей исходной постановке могут быть заданными в виде уравнений; в случае задания системы неравенств всегда можно перейти от неравенств к уравнениям

$$\theta_i(x_1, \dots, x_m, x_{m+i}) = 0 \quad (9.5)$$

путем введения дополнительных переменных x_{m+i} , причем для неравенств вида (9.2) $x_{m+i} > 0$, а для неравенств вида (9.4) $x_{m+i} < 0$. На ряд переменных может быть наложено условие целочисленности, т. е.

x_p — целые числа, $p=1,2,\dots,q$, $q \leq m$.

В задаче оптимизации, в которой ограничения имеют вид уравнений, количество ограничений n не может быть больше числа переменных m , т. е. $n \leq m$. Разность $m-n$ определяет число степеней свободы в данной задаче. Только $m-n$ переменных берутся произвольными, значения же остальных переменных определяются из системы ограничений. Если $m=n$, то число степеней свободы равно нулю и задача в этом случае является алгебраической. Оптимизация целевой функции $F(X)$ при этом не требуется.

Задача оптимального консультирования (формирования рекомендаций) в виде (9.1) — (9.5) представляет собой **задачу математического программирования**. При этом если **целевая функция (9.1) и все ограничения линейны**, то задачу оптимизации называют **задачей линейного программирования**, если же **целевая функция или хотя бы одно ограничение нелинейны**, то задача оптимизации является **задачей нелинейного программирования**.

Во многих задачах математического программирования некоторые переменные могут принимать лишь определенные дискретные значения (например, диаметр обмоточного провода, выбираемый из определенного сортамента, номиналы конденсаторов и т. д.) либо только целочисленные значения (например, число выпускаемых станков, самолетов и т. д.). **В этом случае консультационная задача может быть сформулирована в терминах дискретного программирования.**

При консультировании проблем с использованием моделей и методов математического программирования оказывается удобной геометрическая иллюстрация процесса формирования оптимальных рекомендаций. Рассмотрим геометрическую интерпретацию задачи математического программирования с линейной целевой функцией и с системой ограничений, образующих выпуклую оболочку области существования задачи оптимизации, т. е. пусть имеется система уравнений

$$\theta_i(x_1, \dots, x_m) \leq 0, \quad i = 1, 2, \dots, m; \quad (9.6)$$

$$a_j \leq x_j \leq b_j, \quad j = 1, 2, \dots, n, \quad (9.7)$$

и целевая функция вида

$$F(X) = \sum_{j=1}^n c_j x_j. \quad (9.8)$$

Требуется найти значения переменных, удовлетворяющих ограничениям (9.6) и (9.7) и обращающих в минимум целевую функцию (9.8).

В данной задаче $m-n$ переменных могут быть выбраны произвольно, в частности взяты равными нулю. Это дает возможность проиллюстрировать задачу геометрически в $m-n$ -мерном пространстве E^{m-n} . В этом пространстве каждой точке $\mathbf{X}^{(r)}$ соответствует совокупность чисел x_1^r, \dots, x_{m-n}^r равных проекции вектора, проведенного из начала координат в точку $\mathbf{X}^{(r)}$, на координатные оси пространства E^{m-n} .

Функция $F(x_1, \dots, x_n)$ в каждой точке пространства имеет определенное значение, следовательно, пространство E^{m-n} является скалярным полем критерия оптимальности $F(\mathbf{X})$ и функций ограничений $\theta_i(\mathbf{X})$. Функциям ограничений (9.6) соответствуют граничные гиперповерхности (в частном случае — гиперплоскости). Ограничениям (9.7) соответствуют гиперплоскости, выделяющие в пространстве определенную пространственную область. Если ограничения (9.6) и (9.7) представляют собой выпуклую область, то решения задачи оптимизации будут соответствовать такой точке этой области со скалярным полем критерия $F(\mathbf{X})$, в которой критерий $F(\mathbf{X})$ принимает минимальное значение. Поясним вышесказанное на примере.

Пример 9.1. Имеется следующая задача нелинейного программирования: минимизировать целевую функцию

$$F(\mathbf{X}) = x_1 - x_2, \tag{9.9}$$

при ограничениях

$$x_1^2 + x_2^2 \leq 4; \quad -2x_1 + x_2 \leq 1; \quad x_1 - 2x_2 \leq 1. \tag{9.10}$$

Введя дополнительные переменные x_3, x_4 и x_5 , (9.10) превратим в систему уравнений, которую разрешим относительно этих введенных переменных. При этом получим

$$\begin{aligned} x_3 &= 4 - x_1^2 - x_2^2; \quad x_3 \geq 0; \\ x_4 &= 1 + 2x_1 - x_2; \quad x_4 \geq 0; \\ x_5 &= 1 - x_1 + 2x_2; \quad x_5 \geq 0. \end{aligned} \tag{9.11}$$

В данной задаче число переменных $m = 5$, число уравнений $n = 3$, тогда $m-n=2$, что дает возможность дать геометрическую интерпретацию задачи в пространстве E^2 , т. е. на плоскости.

Поскольку все переменные должны быть положительными:

$$x_j \geq 0, \quad j=1,2,3,4,5 \tag{9.12}$$

каждое из неравенств (9.12) определяет некоторую допустимую область в пространстве E^2 . Так, неравенство $x_3 \geq 0$ определяет верхнюю полуплоскость, неравенство $x_4 \geq 0$ — полуплоскость, лежащую по одну сторону от прямой $2x_1 - x_2 + 1 = 0$, а именно ту, которая содержит начало координат. Область, соответствующая $x_4 < 0$, является запрещенной, и ее удобно отметить (на рис. 9.1, а она отмечена штриховкой).

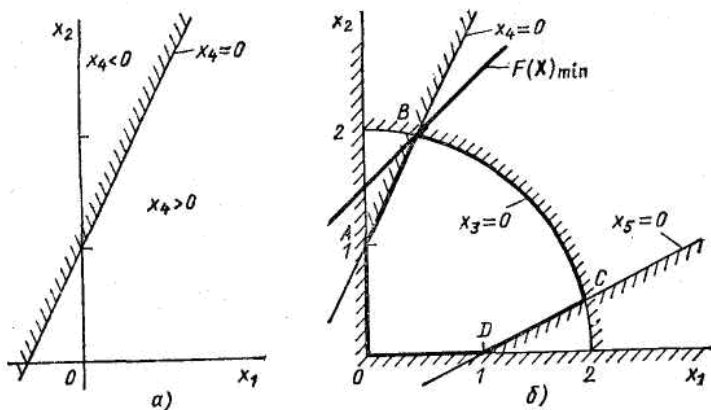


Рис. 9.1. Допустимая и запрещенная полуплоскости (а) и область существования задачи нелинейного программирования (б)

Подобные построения для всех x_i приведены на рис. 9.1, б. Областью существования задачи является область $OABCD$. В рассматриваемом примере функция (9.9) принимает минимальное значение на границе допустимой области в точке B .

Отметим ряд особенностей задачи нелинейного программирования с нелинейной целевой функцией. В общем случае заранее нельзя сказать о расположении точки, в которой функция $F(\mathbf{X})$ принимает максимальное или минимальное значение. Эта точка может находиться как на границе допустимой области, так и внутри нее. Функция $F(\mathbf{X})$ может достигнуть экстремального значения как в одной точке, так и на некотором множестве (гиперлинии или гиперповерхности).

Другая особенность задачи математического программирования состоит в том, что в общем случае нелинейная целевая функция $F(\mathbf{X})$ может иметь несколько локальных экстремумов в допустимой области, включая ее границу. Поэтому при оптимальном консультировании возникает проблема определения глобального экстремума задачи оптимизации, что в практических задачах консультирования нередко вызывает значительные трудности при вычислениях.

В общем виде задача нелинейного программирования пока не имеет строгого математического решения. Однако в связи с тем что данный класс задач довольно часто встречается в практических задачах консультирования, разработано большое число методов и эвристических алгоритмов решения конкретных задач нелинейного программирования.

9.2. Структурный синтез и параметрическая оптимизация формируемых рекомендаций

Задача синтеза формируемых рекомендаций включает в себя создание структуры КП и расчет ее параметров. Эти две части синтеза будем называть *структурным и параметрическим синтезом формируемых рекомендаций*.

Структурный синтез. Задача структурного синтеза заключается в поиске оптимальной или рациональной структуры (схемы) КП для реализации заданных функций в рамках выбранного принципа действия. Результаты структурного синтеза могут быть представлены в виде перечня элементов вместе с таблицей соединений (связей), схемы расположения элементов в линейном евклидовом пространстве с указаниями их типов, схемы алгоритмов и т. п.

Задачи структурного синтеза относят к задачам консультирования, наиболее сложным с точки зрения возможностей формализации. Сложность формализованного синтеза структур заключается прежде всего в наличии большого числа факторов, влияющих на разновидности, свойства и параметры синтезируемых структур, а также в трудностях решения задач оптимизации большой размерности при высокой степени детализации описания синтезируемых рекомендаций.

Так, при синтезе рекомендаций по созданию структур специализированных ЭВМ и микропроцессорных систем возникают трудности решения ряда специфичных проблем, связанных с согласованием структур и алгоритмов функционирования ЭВМ и систем с характеристиками решаемых задач. Результатами структурного синтеза ЭВМ являются рекомендации по номенклатуре входящих в состав ЭВМ блоков, число блоков каждого типа, топология информационных и управляющих взаимосвязей между блоками, а также расписание функционирования каждого блока при организации вычислительного процесса.

Большая размерность консультационных задач КП делает целесообразным блочно-иерархический подход, при котором консультационный процесс разбивается на взаимосвязанные иерархические уровни. Формирование рекомендаций по структурному синтезу составляет существенную часть консультационного процесса и также организуется по блочно-иерархическому принципу. Это означает, что синтезируется не вся сложная рекомендация целиком, а на каждом уровне в соответствии с выбранным способом декомпозиции синтезируются определенные функциональные блоки

рекомендации с соответствующим уровнем детализации. Можно предложить различные способы классификации задач формирования рекомендаций на стадии структурного синтеза. Так, в частности, в зависимости от стадии консультирования будем различать следующие процедуры структурного синтеза: выбор основных принципов функционирования КП, формирование рекомендации в рамках заданных принципов функционирования, выпуск технической документации. В зависимости от типа синтезируемых структур будем различать *задачи одномерного, схемного и геометрического синтеза*. В зависимости от возможностей формализации будем различать *задачи, в которых возможен полный перебор сформированных рекомендаций, задачи, которые не могут быть решены путем полного перебора за приемлемое время, задачи поиска вариантов рекомендаций по синтезу структур в счетном множестве допустимых вариантов и задачи синтеза, решение которых является проблематичным*.

В самом общем случае для решения задачи необходим перебор вариантов рекомендаций, и сокращение перебора является актуальной проблемой. Иерархический подход уменьшает число вариантов рекомендаций на каждом уровне и делает решение задачи определения оптимальной структуры КП реальным.

Отметим существенное различие между консультационными задачами синтеза оптимальных структур и задачами анализа качества структур КП. В анализе необходимо убедиться, что решение существует, а численные методы анализа устойчивы. При структурном синтезе не гарантировано даже существование номинальной структуры, удовлетворяющей всем требованиям КЗ на КП. Существующие и разрабатываемые ММ синтезируемых КП, как правило, оказываются довольно чувствительными к начальным условиям, к размерности задачи оптимизации, к виду целевых функций и ограничений. Поэтому необходимым условием для решения задач синтеза оптимальных рекомендаций по синтезу структур КП различной природы является использование методов и средств автоматизированного консультирования. Естественно, что формализованные модели и методы для САК, с одной стороны, должны характеризоваться высокой степенью общности и достоверности, а с другой стороны, должны быть разрешимыми с вычислительной точки зрения.

Существо получения ММ КП для решения задач структурного синтеза поясним на примерах компоновки, размещения и трассировки, довольно часто встречающихся в задачах конструирования ЭВА,

распределения оборудования по производственным цехам, размещения цехов по территории завода, при проектировании линий электропередачи транспортных средств и т. п.

Пример 9.2. Задача компоновки. Под задачами компоновки понимают задачи разбиения множества $\mathbf{D} = \{d_1, \dots, d_n\}$ из n элементов на ряд непересекающихся подмножеств \mathbf{D}_k , $k=1, 2, \dots, N$, чтобы при этом выполнялись заданные ограничения и достигался экстремум некоторой функции качества $F(\mathbf{X})$.

При заданном числе N подмножеств разбиения задача компоновки формулируется следующим образом:

$$F(\mathbf{X}) \rightarrow \min;$$

$$\forall k, l \in \{1, 2, \dots, N\} (\mathbf{D}_k \cap \mathbf{D}_l = \emptyset), \bigcup_{k=1}^N \mathbf{D}_k = \mathbf{D}, \quad (9.13)$$

где $\mathbf{D}_k \subset \mathbf{D}$ — множество элементов, принадлежащих k -му подмножеству разбиения.

Общее число разбиений

$$\prod_{k=1}^N C_{n - \sum_{q=1}^{k-1} n_q}^{n_k} = \frac{n!}{n_1! n_2! \dots n_N!}$$

при условии, что мощность каждого подмножества \mathbf{D}_k является заданной, т. е.

$$|\mathbf{D}_k| = n_k, \quad \sum_{k=1}^N n_k = n.$$

Поэтому при реальном консультировании (при $n > 100$) получить решение задачи компоновки путем перебора всех вариантов разбиения даже с использованием современных ЭВМ практически невозможно. Для уменьшения перебора задачу компоновки можно сформулировать в терминах целочисленного программирования.

Пусть требуется распределить n компонентов электронной схемы между N блоками таким образом, чтобы суммарное число связей между блоками было минимально. Введем вектор \mathbf{X} переменных консультирования, компоненты x_{ik} ($i=1, 2, \dots, n$, $k=1, 2, \dots, N$) которого указывают на включение или невключение элемента $d_i \in \mathbf{D}$ в подмножество \mathbf{D}_k , т. е.

$$x_{ik} = \begin{cases} 1, & \text{если компонент } d_i \text{ включается в подмножество } \mathbf{D}_k; \\ 0 & \text{в противном случае.} \end{cases}$$

Содержательно функция качества $F(\mathbf{X})$ может характеризовать число связей между подмножествами \mathbf{D}_i (при заданном числе N подмножеств), число подмножеств N , число типов подмножеств \mathbf{D}_i , опре-

деляемых данным разбиением. Очевидно, все перечисленные функции качества $F(\mathbf{X})$ следует минимизировать.

Пусть $F(\mathbf{X})$ характеризует общее число связей между подмножествами \mathbf{D}_k , $k=1,2,\dots,N$. Тогда задача компоновки формулируется следующим образом: минимизировать целевую функцию

$$F(\mathbf{X}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{k=1}^{N-1} \sum_{l=k+1}^N x_{ik} x_{jl} \pi_{ij} \quad (9.14)$$

при ограничениях

$$\sum_{k=1}^N x_{ik} = 1, i = \overline{1, n}; \sum_{i=1}^n x_{ik} = n_k, k = \overline{1, N}; \quad (9.15)$$

$$\sum_{i=1}^n \theta_i^{(S)} x_{ik} \leq V_S^{(k)}. \quad (9.16)$$

Здесь π_{ij} — число связей между элементами d_i и d_j ; $v_i^{(S)}$ — значение параметра S для элемента d_i ; $V_S^{(k)}$ — ограничение по параметру S , накладываемое на подмножество \mathbf{D}_k , причем под параметром $v_i^{(S)}$ элемента d_i может подразумеваться любой показатель, подчиняющийся свойству аддитивности (объем, масса, стоимость, энергоёмкость и т.д.).

Целевая функция (9.14) является квадратичной, поэтому задача компоновки, сформулированная в виде задачи (9.14) — (9.16), является квадратичной задачей дискретного программирования.

Пример 9.3. Задача размещения. После того как решена задача компоновки, требуется определенным образом расположить компоненты, входящие в один блок. От того, как будут размещены микросхемы на определенной печатной плате (или любые элементы на плоскости), зависит длина соединительных проводников (связей), от которой в свою очередь зависят уровень помех и время распространения сигналов. Подобные задачи получили название задач размещения. В общем случае требуется найти такое размещение компонентов d_1, d_2, \dots, d_n на множестве q_1, q_2, \dots, q_m ($m > n$) позиций монтажного пространства, при котором суммарная длина электрических соединений между компонентами была бы минимальной. Введем псевдобулевы переменные

$$x_{ik} = \begin{cases} 1, & \text{если компонент } d_i \text{ назначается на позицию } q_k; \\ 0 & \text{в противном случае.} \end{cases}$$

Тогда задача размещения может быть сформулирована в следующем виде: минимизировать целевую функцию

$$F(\mathbf{X}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{k=1}^m \sum_{s=1}^m x_{ik} x_{js} l_{ks} m_{ij} \quad (9.17)$$

при ограничениях

$$\sum_{k=1}^m x_{ik} = 1, \quad i = \overline{1, n}; \quad \sum_{i=1}^n x_{ik} \leq 1, \quad k = \overline{1, m}. \quad (9.18)$$

Здесь l_{ks} — расстояние между позициями q_k и q_s ; m_{ij} — число связей между компонентами d_i и d_j .

Первое ограничение гарантирует, что каждая компонента разместится только на одной позиции; второе ограничение гарантирует, что на каждую позицию будет назначено не более одной компоненты.

Следует отметить, что среди известных критериев размещения наибольшее распространение получили минимум суммарной длины соединительных проводников, минимум наибольшей длины из всех длин соединительных проводников, минимум числа пересечений проводников и др.

Пример 9.4. Задача трассировки. Задача заключается в определении трасс соединений между компонентами схемы с учетом консультационных ограничений, причем *трассой* называют множество связанных отрезков, соединяющих точки цепи. Консультационные задачи трассировки встречаются при конструировании печатных плат, при разработке систем водоснабжения, канализации, электроснабжения и т. д.

Критериями оптимальности в консультационных задачах трассировки могут выбираться минимум суммарной длины трасс, минимум числа соединений трасс длины больше заданной, минимум числа переходов между слоями в многослойных структурах и др.

Рассмотрим формальную постановку одной из разновидностей консультационной задачи трассировки, а именно — задачи построения связывающих сетей минимальной длины для цепей α_k . Соединяемые по цепи α_k точки образуют множество U_k мощностью $|U_k| = n_k$, в котором каждому элементу $u_k \in U_k$ в пространстве соответствует одна точка. Введем псевдобулевы переменные

$$x_{ij} = \begin{cases} 1, & \text{если ребро } (i, j) \text{ длиной } l_{ij} \text{ включается в связывающую сеть;} \\ 0 & \text{в противном случае.} \end{cases}$$

Задача построения минимальной связывающей сети имеет вид: минимизировать целевую функцию

$$L(\mathbf{x}) = \sum_{i=1}^{n_k-1} \sum_{j=i+1}^{n_k} l_{ij} x_{ij} \quad (9.19)$$

при ограничениях

$$\sum_{i=1}^{n_k} x_{ij} \leq K_0, \quad i = \overline{1, n_k}; \quad (9.20)$$

$$y_{ij}^{(s)} = \sum_{r=1}^{n_k} y_{ir}^{(s-1)} x_{rj}; \quad y_{ij}^{(1)} = x_{ij}; \quad (9.21)$$

$$\sum_{s=1}^{n_k-1} y_{ij}^{(s)} \geq 1, \quad i = \overline{1, n_k}, \quad j = \overline{2, n_k}, \quad s = \overline{2, n_k-1}. \quad (9.22)$$

Здесь K_0 — максимально допустимое число соединений, инцидентных одной точке цепи; $y_{ij}^{(s)}$ — вспомогательные переменные.

Условия (9.21), (9.22) гарантируют связность определяемой сети.

Приведенные примеры показывают, что во многих случаях консультационные задачи структурного синтеза являются экстремальными комбинаторными задачами, которые могут быть сведены к задачам дискретного программирования. Оценка трудоемкости получения точных решений задач этого класса позволяет сделать вывод, что при реальном консультировании получение точных решений либо невозможно, либо требует больших затрат машинного времени. Поэтому для структурного синтеза каждого класса КП необходима разработка специальных приближенных методов, позволяющих получать эффективные рекомендации, близкие к оптимальным, а точные методы при этом служат для оценки качества сформированных рекомендаций.

Параметрический синтез. Консультационная задача параметрического синтеза заключается в формировании рекомендаций по определению наилучших значений параметров для выбранной структуры КП с учетом всех требований КЗ на консультируемую проблему.

Функционирование любой консультируемой проблемы подчиняется определенным физическим законам. Закон функционирования КП описывается аналитическими соотношениями между входными, внутренними и выходными переменными КП. Эти переменные связаны определенными соотношениями с переменными консультирования (формирования рекомендаций) X , под которыми понимаются внутренние переменные, допускающие варьирование. В консультационном процессе параметрического синтеза варьирование переменных консультирования (формирования рекомендаций) X ведет к изменению выходных параметров Y КП.

Для пояснения сущности задач параметрического синтеза будем использовать геометрическую интерпретацию, связанную с введением m -мерного пространства E^m пространства параметров консультирования (управляемых параметров) и k -мерного пространства E^k выходных параметров. Каждой точке пространства E^m и E^k соответствуют векторы \mathbf{X} и \mathbf{Y} значений переменных консультирования и выходных параметров соответствующего варианта КП.

Для постановки и решения консультационной задачи параметрического синтеза необходимо формирование целевой функции $F(\mathbf{X})$, отражающей качество функционирования КП. Векторный характер критериев оптимальности (многокритериальность) в задачах консультирования обуславливает сложность проблемы постановки задач оптимизации.

Формально задачу параметрического синтеза можно представить как задачу нахождения вектора $\mathbf{X} \in E^m$, которым минимизирует целевую функцию

$$F(\mathbf{X}) \rightarrow \min \quad (9.23)$$

при ограничениях

$$g(\mathbf{X})=0 \text{ и } h(\mathbf{X}) \leq 0, \quad (9.24)$$

где $g(\mathbf{X})$ и $h(\mathbf{X})$ — векторные функции от \mathbf{X} , описывающие систему ограничений на параметры консультирования \mathbf{X} .

В качестве целевой функции целесообразно выбирать один из параметров, наиболее полно характеризующих свойства КП. Если частный критерий при консультировании выбрать затруднительно, то будем прибегать к формированию обобщенных критериев. В зависимости от целей консультирования и типов математических моделей КП целевые функции могут задаваться по-разному, для чего в САК должна быть предусмотрена **библиотека целевых функций**. Примерами целевых функций могут служить целевая функция максимального модуля отклонения характеристик КП от заданных

$$F(\mathbf{X}) = \max_i |y_j^{(0)} - y_j^{(p)}(\mathbf{X})|, \quad j = \overline{1, m}, \quad (9.25)$$

и среднеквадратичная целевая функция

$$F(\mathbf{X}) = \sqrt{\sum_{j=1}^m (y_j^{(0)} - y_j^{(p)}(\mathbf{X}))^2}, \quad (9.26)$$

где $y_j^{(0)}$ — заданное значение параметра y_j ; $y_j^{(p)}$ — реальное значение этого параметра.

Очевидно, что функции вида (9.25) и (9.26) в задачах оптимизации следует минимизировать. Оптимизация может быть осуществлена

различными методами, включающими весьма сложные аналитические и численные математические процедуры.

Если в задачах оптимального консультирования все консультационные переменные и состояний являются непрерывными, то для формирования рекомендаций по решению консультационных задач параметрического синтеза могут быть использованы методы решения задач нелинейного программирования, основанные на хорошо разработанных процедурах поиска экстремума функций. Однако не всегда все элементы в консультируемых проблемах могут принимать любые значения в пределах некоторой допустимой области. Это связано прежде всего со стандартизацией и унификацией объектов в различных проблемных областях. Так, в радиотехнике параметры резисторов и конденсаторов могут принимать только определенные значения из разрешенной шкалы номиналов, в строительстве плиты перекрытия, балки и другие комплектующие изделия имеют ряд определенных стандартных размеров. Кроме того, на параметры разрабатываемых объектов также накладывается ряд ограничений, учитывающих условия стандартизации и унификации. Так, в электротехнике и радиоэлектронике разрешается использовать только определенные значения питающих напряжений, в вычислительной технике существуют стандартные градации емкости устройств памяти и коммутационного оборудования.

Поэтому для решения консультационных задач оптимизации при консультировании проблем с дискретными значениями параметров методы оптимизации непрерывных объектов непосредственно неприменимы. Эти задачи относятся к задачам дискретного программирования. Если при оптимизации часть параметров дискретна, а часть имеет непрерывный характер, то задача должна решаться методами частично дискретного программирования. Из-за недифференцируемости выходных параметров в задачах дискретного программирования довольно часто возникают трудности при вычислениях.

Рассмотрим пример задачи параметрического синтеза.

Пример 6.5. Формирование рекомендаций на проектирование трехстержневой фермы. Цель проектирования — выбор конструкции трехстержневой фермы (рис. 9.2) минимальной массы.

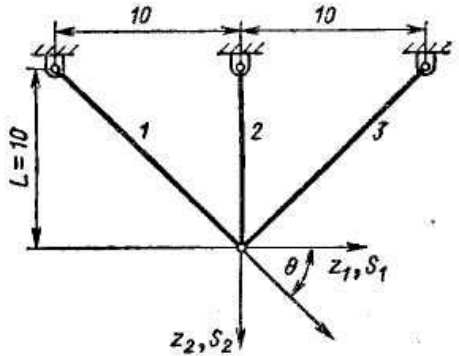


Рис. 9.2. Трехстержневая ферма

Проектирование сводится к выбору площадей поперечных сечений отдельных стержней x_1, x_2 и x_3 (переменные проектирования), так чтобы ферма была по возможности легкой и удовлетворялись ограничения на напряжение, устойчивость при продольном изгибе, смещение и размеры стержней.

Целевая функция $F(\mathbf{X})$ представляет собой массу конструкции (при $\theta = 45^\circ$):

$$F(\mathbf{X}) = \rho g (10\sqrt{2}x_1 + 10x_2 + 10\sqrt{2}x_3), \quad (9.27)$$

где ρg — плотность материала, из которого изготовлена ферма.

Горизонтальное и вертикальное смещения z_1 и z_2 общего узла получены из линейных соотношений теории упругости. Для данной конструкции эти уравнения имеют вид

$$\mathbf{K}(x)\mathbf{Z} - \mathbf{S} = 0, \quad (9.28)$$

где $\mathbf{Z} = [z_1, z_2]^T$;

$$\mathbf{K}(x) = \frac{\sqrt{2}E}{40} \begin{bmatrix} (x_1 + x_3) & (x_1 - x_3) \\ (x_1 - x_3) & (x_1 + x_3 + 2\sqrt{2}x_2) \end{bmatrix}, \mathbf{S} = \begin{bmatrix} P \cos \theta \\ P \sin \theta \end{bmatrix} \quad (9.29)$$

— положительно определенная матрица жесткости и вектор нагрузки соответственно; E — модуль Юнга; θ — угол приложения нагрузки, отсчитываемый от горизонтали.

Вычисляя деформации ε в зависимости от смещения узла и применяя закон Гука, найдем напряжение в каждом стержне:

$$\sigma_1 = E(z_1 + z_2)/20, \quad \sigma_2 = Ez_2/10, \quad \sigma_3 = E(z_2 - z_1)/20. \quad (9.30)$$

С помощью допустимых напряжений

$$\sigma_i^{(0)}, \quad i = \overline{1, 3},$$

определенных для каждого стержня, ограничения на напряжения примут вид

$$\frac{E}{20} |z_1 + z_2| - \sigma_1^{(0)} \leq 0; \quad (9.31)$$

$$\frac{E}{10} |z_2| - \sigma_2^{(0)} \leq 0; \quad (9.32)$$

$$\frac{E}{20} |z_2 - z_1| - \sigma_3^{(0)} \leq 0. \quad (9.33)$$

Чтобы наложить ограничения по устойчивости, необходимо задать вид зависимости момента инерции от площади поперечного сечения для каждого стержня. Общей при инженерных расчетах является зависимость вида $I = \beta x^2$, где β — безразмерная постоянная. Подобная зависимость получается, если зафиксировать форму поперечного сечения и все его размеры менять в одинаковой пропорции. Осевые усилия имеют вид $\Phi_i = \sigma_i x_i$, $i = 1, 2, 3$, растяжения стержней считаются положительными. Ограничения по устойчивости имеют вид

$$-\Phi_i \leq \pi^2 EI_i / L_i^2, \quad i = \overline{1, 3}. \quad (9.34)$$

Используя равенства (9.28), можно записать:

$$\begin{aligned} -E(z_1 + z_2)x_1/20 - \pi^2 E \beta x_1^2 / 200 &\leq 0; \\ -E z_2 x_2 / 10 - \pi^2 E \beta x_2^2 / 100 &\leq 0; \\ -E(z_2 - z_1)x_3/20 - \pi^2 E \beta x_3^2 / 200 &\leq 0. \end{aligned} \quad (9.35)$$

Ограничения на смещения могут быть сформулированы в виде

$$|z_1| - z_1^{(0)} \leq 0, \quad |z_2| - z_2^{(0)} \leq 0, \quad (9.36)$$

где $z_1^{(0)}$ и $z_2^{(0)}$ — заданные верхние границы для z_1 и z_2 соответственно. Наконец, требуется, чтобы площади поперечных сечений были неотрицательны, так что следует наложить ограничения

$$-x_i \leq 0, \quad -x_2 \leq 0, \quad -x_3 \leq 0. \quad (9.37)$$

Теперь задачу проектирования можно рассматривать как задачу выбора таких переменных проектирования x_1 , x_2 и x_3 , которые минимизируют $F(\mathbf{X})$ и удовлетворяют ограничениям (9.28), (9.31) — (6.37).

Заметим, что смещения z_1 и z_2 играют в данной формулировке важную роль. Их определяют из уравнения для конструкции (9.28), коль скоро заданы переменные проектирования. Эти переменные представляют собой отклик системы на приложенную нагрузку, их называют *переменными состояниями*, а уравнения (9.28) — *уравнениями состояния*.

9.3. Разновидности консультационных задач оптимизации

В задачах оптимального формирования рекомендаций по решению задач консультируемых проблем вектор переменных консультирования $\mathbf{X} = \{x_1, \dots, x_n\}$ выбирают в результате определения экстремума целевой функции $F(\mathbf{X})$ в допустимой области, заданной системой ограничений на параметры консультируемой проблемы. В самом общем виде целевая функция и ограничения являются нелинейными функциями переменных консультирования \mathbf{X} .

Задачи, в которых экстремум ищут в пределах неограниченного пространства переменных консультирования, относятся к задачам *безусловной оптимизации*. Найденные при этом *экстремумы* называют *безусловными*. Наличие ограничений любого вида приводит к задачам *условной оптимизации*, решение которых дает *условный экстремум*.

При решении задач оптимизации первоначально проверяют условия, которым должен удовлетворять вектор переменных консультирования \mathbf{X} , минимизирующий (максимизирующий) критерий качества $F(\mathbf{X})$. Эти условия проверяют для отыскания стационарных точек, среди которых находится искомый вектор \mathbf{X} .

Функция $F(\mathbf{X})$, определенная в E^m , имеет абсолютный минимум в $\mathbf{X}^* \in E^m$, если

$$F(\mathbf{X}^*) \leq F(\mathbf{X}) \quad (9.38)$$

для всех $\mathbf{X} \in E^m$. Минимум является строгим, если в (9.38) стоит знак строгого неравенства для $\mathbf{X} \neq \mathbf{X}^*$.

Функция $F(\mathbf{X})$ имеет в \mathbf{X}^* относительный минимум, если существует $\varepsilon > 0$ такое, что для любой точки $\mathbf{X} \in S_\varepsilon(\mathbf{X}^*)$ выполняется неравенство

$$F(\mathbf{X}) - F(\mathbf{X}^*) > 0, \quad (9.39)$$

где $S_\varepsilon(\mathbf{X}^*)$ — окрестность точки \mathbf{X}^* . При определении максимума $F(\mathbf{X})$ (9.39) должно быть заменено на неравенство

$$F(\mathbf{X}) - F(\mathbf{X}^*) < 0.$$

Если $F(\mathbf{X})$ имеет в точке \mathbf{X}^* абсолютный минимум, то эта функция имеет там также и относительный минимум. Обратное не обязательно верно. Часто относительный минимум называют *локальным*, а абсолютный минимум — *глобальным*.

Рассмотрим необходимые и достаточные условия экстремума. Классические методы оптимизации используют тогда, когда известно аналитическое выражение функции $F(\mathbf{X})$ и известно, что она по крайней мере дважды дифференцируема по переменным консультирования. Тогда для определения экстремума будем

использовать необходимые и достаточные условия безусловного экстремума. Эти условия легко получить с помощью разложения $F(\mathbf{X})$ в окрестностях экстремальной точки \mathbf{X}^* в ряд Тейлора:

$$F(\mathbf{X}) - F(\mathbf{X}^*) = \sum_{i=1}^m \frac{\partial F(\mathbf{X}^*)}{\partial x_i} (x_i - x_i^*) + \frac{1}{2} \sum_{i=1}^m \sum_{k=1}^m \frac{\partial^2 F(\mathbf{X}^*)}{\partial x_i \partial x_k} (x_i - x_i^*) (x_k - x_k^*). \quad (9.40)$$

Если \mathbf{X}^* — точка максимума, то линейные члены в (9.40) равны нулю, тогда равны нулю составляющие вектора—градиента функции $F(\mathbf{X})$. Следовательно, необходимым условием экстремума является условие

$$\text{grad } F(\mathbf{X}) = 0,$$

где

$$\text{grad } F(\mathbf{X}) = \frac{\partial F(\mathbf{X})}{\partial \mathbf{X}} = \left(\frac{\partial F(\mathbf{X})}{\partial x_1}, \frac{\partial F(\mathbf{X})}{\partial x_2}, \dots, \frac{\partial F(\mathbf{X})}{\partial x_m} \right).$$

Все точки, где выполняется это условие, называются *стационарными*.

Достаточное условие максимума на основании (9.39) и (9.40) заключается в том, что сумма членов со вторыми частными производными должна быть отрицательной при любых $\mathbf{X} \in S_\varepsilon(\mathbf{X}^*)$, т. е.

$$(\Delta \mathbf{X})^T \Gamma (\Delta \mathbf{X}) < 0, \quad (9.41)$$

где $\Delta \mathbf{X} = \mathbf{X} - \mathbf{X}^*$ — вектор-столбец; $((\Delta \mathbf{X})^T = (\mathbf{X} - \mathbf{X}^*)^T$ — вектор-строка);

$$\Gamma = \begin{bmatrix} \frac{\partial^2 F(\mathbf{X})}{\partial x_1^2} & \frac{\partial^2 F(\mathbf{X})}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 F(\mathbf{X})}{\partial x_1 \partial x_m} \\ \frac{\partial^2 F(\mathbf{X})}{\partial x_1 \partial x_2} & \frac{\partial^2 F(\mathbf{X})}{\partial x_2^2} & \dots & \frac{\partial^2 F(\mathbf{X})}{\partial x_2 \partial x_m} \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 F(\mathbf{X})}{\partial x_1 \partial x_m} & \frac{\partial^2 F(\mathbf{X})}{\partial x_2 \partial x_m} & \dots & \frac{\partial^2 F(\mathbf{X})}{\partial x_m^2} \end{bmatrix}$$

— матрица Гессе; Γ — знак транспонирования матрицы.

Условие (9.41) есть достаточное условие максимума. Матрицу Γ , удовлетворяющую условию (9.41) при любых $\Delta \mathbf{X}$, называют *отрицательно определенной*, а в случае $(\Delta \mathbf{X})^T \Gamma (\Delta \mathbf{X}) > 0$ для любых $\Delta \mathbf{X}$ — *положительно определенной*. Поэтому достаточные условия

экстремума можно представить как требование отрицательной определенности матрицы Гессе для максимума или положительной определенности для минимума в экстремальной точке.

Если достаточные условия не выполняются, то имеем не экстремальную, а седловую точку.

При наличии ограничений на переменные консультирования X решение задачи оптимизации ищут в некоторой допустимой области S согласно следующему направляющему принципу. Если $F(x_1, \dots, x_m)$ является функцией нескольких переменных, определенных на допустимой области S , то максимальное значение $F(X)$, если оно существует, достигается в одной или более точках, которые могут принадлежать следующим множествам:

$K_1 = \{ (x_1, \dots, x_m) \mid (x_1, \dots, x_m) \text{ — точка в области } S, \text{ в которой функция } F \text{ стационарна} \};$

$K_2 = \{ (x_1, \dots, x_m) \mid (x_1, \dots, x_m) \text{ — точка границы области } S \};$

$K_3 = \{ (x_1, \dots, x_m) \mid \text{— функция } F(X) \text{ не дифференцируема в точке } (x_1, \dots, x_m) \}.$

На рис. 9.3 приведен пример геометрической интерпретации многоэкстремальной задачи оптимального консультирования.

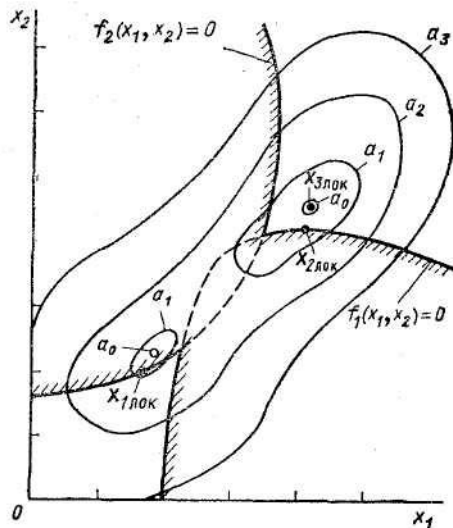


Рис. 9.3. Многоэкстремальная задача оптимального консультирования

На рисунке показаны линии равного уровня целевой функции $F(\mathbf{X})(a_3 > a_2 > a_1 > a_0)$ и видны три локальных оптимума, которые находятся в областях, определяемых общим направляющим принципом (точки $\mathbf{X}_{1\text{лок}}$, $\mathbf{X}_{2\text{лок}}$, $\mathbf{X}_{3\text{лок}}$ являются точками локальных оптимумов, причем точка $\mathbf{X}_{3\text{лок}}$ совпадает с глобальным оптимумом).

К сожалению, отсутствуют формальные признаки многоэкстремальных ситуаций. Исключением являются задачи, где целевая функция выпуклая (в задачах минимизации) или вогнутая (при максимизации).

Напомним, что функцию $F(\mathbf{X})$ с числовыми значениями, определенными на выпуклом множестве S , называют *вогнутой*, если для любой пары точек $\mathbf{X}_1, \mathbf{X}_2 \in S$ и для всех чисел λ ($0 \leq \lambda \leq 1$) выполняется неравенство $F(\lambda\mathbf{X}_1 + (1-\lambda)\mathbf{X}_2) \geq \lambda F(\mathbf{X}_1) + (1-\lambda)F(\mathbf{X}_2)$. Если $F(\lambda\mathbf{X}_1 + (1-\lambda)\mathbf{X}_2) \leq \lambda F(\mathbf{X}_1) + (1-\lambda)F(\mathbf{X}_2)$, то функцию $f(\mathbf{X})$ называют *выпуклой*. Если имеют место строгие неравенства, то говорят, что функция *строго вогнута* или *строго выпукла*. Для дважды дифференцируемой функции критерий вогнутости или выпуклости формируется следующим образом. Дифференцируемая функция $F(\mathbf{X})$ строго вогнута в некоторой окрестности точки $\mathbf{X}^{(0)} = (x_1^{(0)}, \dots, x_m^{(0)})$, если выполняются условия

$$F_{11}(\mathbf{X}^{(0)}) < 0, \left| \begin{array}{cc} F_{11}(\mathbf{X}^{(0)}) & F_{12}(\mathbf{X}^{(0)}) \\ F_{21}(\mathbf{X}^{(0)}) & F_{22}(\mathbf{X}^{(0)}) \end{array} \right| > 0, \\ \left| \begin{array}{ccc} F_{11}(\mathbf{X}^{(0)}) & F_{12}(\mathbf{X}^{(0)}) & F_{13}(\mathbf{X}^{(0)}) \\ F_{21}(\mathbf{X}^{(0)}) & F_{22}(\mathbf{X}^{(0)}) & F_{23}(\mathbf{X}^{(0)}) \\ F_{31}(\mathbf{X}^{(0)}) & F_{32}(\mathbf{X}^{(0)}) & F_{33}(\mathbf{X}^{(0)}) \end{array} \right| < 0, \dots$$

т. е. если знаки этих определителей чередуются указанным образом. Здесь $F_{ij}(\mathbf{X}^{(0)})$ — частная производная второго порядка, вычисленная в точке $\mathbf{X}^{(0)}$.

Функция $F(\mathbf{X})$ строго выпукла в малой окрестности точки $\mathbf{X}^{(0)}$, если все определители, указанные выше, положительны.

Достаточные условия для определения максимума или минимума формулируются следующим образом: для того, чтобы в точке $\mathbf{X}^{(0)}$ достигался внутренний локальный максимум, достаточно равенства нулю всех частных производных и строгой вогнутости функции в некоторой окрестности этой точки; для того чтобы в точке $\mathbf{X}^{(0)}$ достигался внутренний локальный минимум, достаточно, чтобы все частные производные обращались в нуль и чтобы в малой окрестности этой точки функция была строго выпуклой.

В большинстве задач консультирования при отсутствии аналитического задания целевых функций проверка $F(\mathbf{X})$ на

выпуклость или вогнутость, как правило, невозможна, поэтому для решения задач оптимального консультирования следует использовать методы поисковой оптимизации, основанные на исследовании малой окрестности оптимальной точки в допустимой области. Основные требования, предъявляемые к методу поиска,— высокая алгоритмическая надежность, приемлемые затраты машинного времени и требуемой памяти.

Методы поиска экстремума классифицируются по следующим признакам: в зависимости от характера экстремума существуют методы условной и безусловной, локальной и глобальной оптимизации; по числу переменных консультирования различают методы одномерного и многомерного поиска, а по характеру информации о виде целевой функции — методы нулевого, первого и второго порядков, причем в методах первого порядка используют градиент целевой функции, поэтому эти методы называются градиентными, в методах второго порядка применяют вторые производные, а в методах нулевого порядка производные не используют.

9.4. Показатели эффективности сформированных рекомендаций и выбор методов поиска экстремума

В задачах оптимального консультирования в самой общей постановке целевая функция и ограничения являются нелинейными функциями переменных консультирования.

В общем случае эти задачи многоэкстремальны, поэтому для их решения следует применять методы определения поиска глобального экстремума, которые включают в себя один из способов генерации начальных точек, а также поиска локальных оптимумов.

Поиск локального оптимума состоит из следующих этапов определения: направления движения к оптимуму, длины шага поиска, окончания поиска.

Алгоритмы поиска локального оптимума X^* являются, как правило, итеративными, т. е. порождают последовательность векторов $\{X^{(k)}\} = X_1, X_2, \dots, X_k$, сходящуюся к вектору X^* .

Будем говорить, что вектор X^* является пределом сходящейся последовательности $\{X^{(k)}\}$, если для любого $\varepsilon > 0$ найдется такой номер N , что при $k > N$ выполняется неравенство $|X_k - X^*| < \varepsilon$. Отсюда следует, что допустимая область S должна вместе с любой сходящейся последовательностью содержать и ее предел. Такую **область** будем

называть **замкнутой**. Примером замкнутой области может служить множество всех точек, удовлетворяющих ограничениям (9.2) и (9.3).

Множество всех точек пространства E^n , которые не содержатся в замкнутой области $S(E^n \setminus S)$, будем называть **открытым**. В замкнутой области S , если она не совпадает со всем пространством E^n , всегда можно найти точки, в ε -окрестности которых имеются точки из $E^n \setminus S$. Такие точки области будем называть **граничными**. Множество всех граничных точек образует **границу области** S . В частности, если область S определяется условиями (9.2) и (9.3), его границу составляют те точки, в которых хотя бы одно из ограничений выполняется как **строгое равенство**.

Эффективность методов поиска локального оптимума определяется **скоростью их сходимости** к X^* , а критериями оценки качества выбора направления являются:

- улучшение значения критерия оптимальности во вновь выбранной точке по сравнению с его величиной в данной точке;
- наиболее быстрое убывание (возрастание) критерия в окрестности данной точки;
- наиболее вероятное расположение экстремума с учетом кривизны гиперповерхности, представляющей критерий оптимальности.

Использование каждого из трех критериев выбора направления движения к оптимуму требует различного числа обращений к модели КП. Необоснованное усиление критерия выбора направления поиска может привести к резкому возрастанию числа обращений к модели КП, а ослабление — к беспорядочному блужданию в окрестности оптимума. В обоих случаях возрастают затраты машинного времени на консультирование.

Проведем краткий анализ методов поиска экстремума. Особенности методов будем иллюстрировать примерами их применения к поиску экстремума функции $F(X)$ в двумерном пространстве переменных консультирования.

Методы безусловной оптимизации. Для решения задачи безусловной оптимизации будем использовать итерационные процессы вида

$$X_k = X_{k-1} + \alpha_k \Delta X_k \quad (9.42)$$

где ΔX_k — вектор, определяющий направление движения из точек X_{k-1} , α_k — числовой множитель, значение которого определяет длину шага в направлении ΔX_k . Для большинства методов

$$\Delta X_k = P_k / \|P_k\|. \quad (9.43)$$

где P — вектор, указывающий направление поиска; $\|P_k\|$ — норма вектора P ; k — индекс, обозначающий номер шага поиска.

Процесс (9.42) будет определен, если указаны способы построения вектора $\Delta \mathbf{X}_k$ и вычисления величины α_k на каждой итерации. От того, каким образом строится вектор $\Delta \mathbf{X}_k$ и определяется множитель α_k , непосредственно зависят свойства процесса: поведение функции $F(\mathbf{X})$ на элементах последовательности $\{\mathbf{X}^{(k)}\}$, сходимость последовательности к решению, скорость сходимости и др. В то же время различные способы построения вектора $\Delta \mathbf{X}_k$ и множителя α_k требуют различных затрат машинного времени и различной емкости оперативной памяти ЭВМ.

Чтобы приблизиться к точке \mathbf{X}^* , естественно двигаться от точки \mathbf{X}_{k-1} в одном из направлений убывания функции $F(\mathbf{X})$ (в направлении спуска). Если точка \mathbf{X}_{k-1} не является точкой минимума или стационарной точкой, то существует бесконечно много векторов $\Delta \mathbf{X}_k$, определяющих направления спуска из точки \mathbf{X}_{k-1} , причем каждый из них определяется условием

$$(F'(\mathbf{X}_{k-1}), \Delta \mathbf{X}_k) < 0,$$

вытекающим из (9.40).

Поскольку местонахождение точки \mathbf{X}^* неизвестно, процесс поиска экстремума может быть прекращен в точке \mathbf{X}_k , при этом число шагов r , разделяющих точки \mathbf{X}_k и \mathbf{X}_{k-1} , определяют из условия $|\mathbf{X}_k - \mathbf{X}_{k-r}| < \varepsilon$. Следовательно, поиск прекращается, если расстояние, на которое продвинулась отображающая точка в пространстве переменных консультирования за последние r шагов, оказывается меньше заданного числа ε .

Прямые методы оценки направлений. Наиболее простым является *метод покоординатного спуска* (метод Гаусса — Зейделя). Направление поиска выбирают поочередно вдоль всех координатных осей, т. е. вектор \mathbf{P}_k в (9.43) состоит из нулевых элементов за исключением одного, равного единице.

Пусть $\mathbf{X}_0 = (x_1^{(0)}, \dots, x_n^{(0)})$ — начальная точка поиска (рис. 9.4, а).

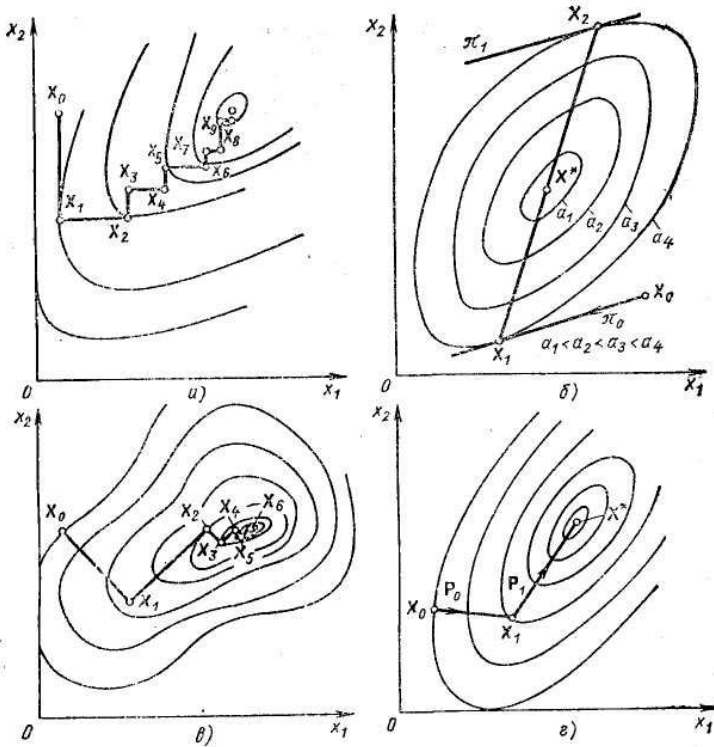


Рис. 9.4. Поиск экстремума квадратичной функции методом покоординатного спуска (а), методом параллельных касательных (б), методом наискорейшего спуска (в) и методом сопряженных градиентов (г)

Результат первой итерации X_1 получается из X_0 при выполнении однопараметрической оптимизации по параметру x_1 , т. е.

$$X_1 = (x_1^{(1)}, x_2^{(0)}, \dots, x_n^{(0)}),$$

причем

$$F(x_1^{(1)}, x_2^{(0)}, \dots, x_n^{(0)}) = \min_{x_1} F(x_1^{(1)}, x_2^{(0)}, \dots, x_n^{(0)}).$$

Результат второй итерации X_2 получается из X_1 путем оптимизации $F(X)$ по параметру x_2 и т. д. Далее процесс продолжается из точки X_n снова путем варьирования параметра x_1 и т. д. Величину шага α_k выбирают по способу оптимального шага.

Величина r за один цикл поиска равна размерности пространства, т. е. $r = n$.

Модификацией алгоритма покоординатного спуска является *метод ортогональных направлений (метод Розенброка)*, который основан на вращении системы координат в соответствии с изменением скорости убывания критерия оптимальности. При этом направление одной оси соответствует наиболее вероятному направлению скорейшего убывания на данной итерации критерия оптимальности, а остальные находятся из условия ортогональности.

Если в задаче оптимального консультирования поверхность отклика ограничена концентрическими эллипсоидами, то точное местоположение оптимума не более чем за $(2n-1)$ одномерных итераций позволяет получить *метод параллельных касательных*. Идея этого метода для $n=2$ иллюстрируется на рис. 9.4, б.

Метод заключается в поиске центра системы концентрических эллипсов. Первоначально определяют направление касательной π_0 из точки X_0 к линии уровня (точка X_1), затем строят произвольную прямую π_1 , параллельную π_0 , и на ней точку X_2 , соответствующую точке касания π_1 и линии уровня. Местоположение центра X^* определяется путем поиска вдоль линии X_1X_2 .

Метод параллельных касательных пригоден для поиска оптимума на любой унимодальной функции, требует небольшого количества информации, но сопряжен с громоздкими вычислениями.

Градиентные методы. Методы поиска, в которых направление движения от итерации X_{k-1} к X_k определяется градиентом (антиградиентом), вычисленным в точке X_{k-1} , носят название *градиентных методов*.

Наиболее применяемым из методов минимизации является *метод наискорейшего спуска*. В большой степени широкому распространению метода способствуют его сравнительная простота и возможность применения для минимизации весьма широкого класса функций. При определении направления поиска выбирают наибо́льшее убывание целевой функции $F(X)$, т. е.

$$P_k = -\text{grad}F(X_{k-1}).$$

Работа метода заключается в следующем. После определения градиента критерия оптимальности в точке X_k движутся вдоль направления антиградиента до точки, в которой достигается минимальное значение функции. Затем в этой точке снова определяют градиент и движутся по прямой согласно направлению нового антиградиента и т. д., пока не достигнут точки, имеющей наименьшее значение функции $F(X)$. На рис. 9.4, в приведен пример движения при

поиске методом наискорейшего спуска оптимума для критерия оптимальности, зависящего от двух переменных. Направление $\text{grad } F(\mathbf{X}_{k-1})$ является касательным к поверхности уровня в точке \mathbf{X}_k , и, следовательно, $\text{grad } F(\mathbf{X}_k)$ в точке \mathbf{X}_{k-1} ортогонален $\text{grad } F(\mathbf{X}_{k-1})$.

Простейшим среди градиентных методов является метод, при котором движение из точки \mathbf{X}_{k-1} осуществляется *по антиградиенту с постоянным шагом* $\alpha_k = \text{const}$.

Если окажется, что $F(\mathbf{X}_k) > F(\mathbf{X}_{k-1})$, то для определения экстремума используют один из возможных алгоритмов *поиска с возвратом*. Так, в частности, поиск может производиться путем возврата в точку \mathbf{X}_{k-1} и дальнейшего перемещения из этой точки вдоль антиградиента с шагом $\alpha/2$. Процедура дробления шага производится до тех пор, пока его значение не станет меньше некоторого малого положительного числа ε .

Описанные варианты реализации градиентного метода отличаются друг от друга способом выбора длины шага. Скорость сходимости этих методов примерно одинакова, а трудоемкость каждой итерации вариантов процесса (9.42) различна только в способах определения параметра α_k . Как правило, вычисления градиента в меньшем числе точек требует метод наискорейшего спуска.

Градиентные методы эффективны для решения задач минимизации гладких и выпуклых функций. В практике автоматизированного консультирования для ускорения сходимости решения задач оптимизации градиентные методы следует использовать в комбинации с другими, более эффективными на начальной стадии решения задачи.

Метод сопряженных градиентов. В градиентных методах для поиска экстремума использовались свойства ортогональности векторов. В методе сопряженных градиентов оптимум целевой функции ищется на основе свойств ортогональности приращений вектора градиентов. Для этой цели наряду с градиентом используют матрицу Гессе Γ критерия оптимальности. С помощью матрицы Γ удастся выбрать направление поиска, наиболее полно учитывающее особенности критерия оптимальности. Напомним, что векторы \mathbf{A} и \mathbf{B} называются сопряженными относительно симметричной и положительно определенной матрицы Γ , если скалярное произведение векторов \mathbf{A} и $\Gamma \mathbf{B}$ равно нулю, т. е. $\langle \mathbf{A}, \Gamma \mathbf{B} \rangle = 0$. Направление поиска \mathbf{P}_{k+1} на $k+1$ -м шаге определяется как

$$\mathbf{P}_{k+1} = -\text{grad } F(\mathbf{X}_k) + \beta_k \mathbf{P}_k,$$

где

$$\beta_k = \frac{\langle \text{grad } F(\mathbf{X}_k), \text{grad } F(\mathbf{X}_k) \rangle}{\langle \text{grad } F(\mathbf{X}_{k-1}), \text{grad } F(\mathbf{X}_{k-1}) \rangle}.$$

Если критерий оптимальности представлен квадратичной функцией, то минимум функции достигается ровно за n шагов (рис. 9.4, z). В случае критерия оптимальности произвольного вида метод позволяет для заданной погрешности получить приближенное решение быстрее, чем это позволяют сделать методы наискорейшего спуска и параллельных касательных.

Методы Ньютона и переменной метрики. Ускорение поиска экстремума связано с улучшением выбора сопряженных направлений. Довольно эффективным является поиск сопряженных направлений с одновременным накоплением информации о матрице Гессе критерия оптимальности. Используют соотношение

$$\mathbf{X}_{k+1} = \mathbf{X}_k - \alpha_k \Gamma_k^{-1} \text{grad } F(\mathbf{X}_k), \alpha_k > 0, k = 0, 1, \dots,$$

которое является формулой *метода Ньютона с регулировкой шага*. Обычный метод Ньютона соответствует случаю $\alpha_k=1$.

Метод Ньютона с регулировкой шага сходится к решению независимо от выбора начальной точки \mathbf{X}_0 и обладает либо линейной, либо квадратичной скоростью сходимости в зависимости от вида функции $F(\mathbf{X})$. В обычном методе Ньютона сходимость гарантируется лишь при наличии достаточно хорошего начального приближения.

При решении задач минимизации выпуклых функций метод Ньютона обеспечивает более высокую скорость сходимости последовательных приближений к решению по сравнению с градиентными методами, однако количество вычислений на итерации метода Ньютона высоко за счет необходимости вычисления и обращения матрицы вторых производных. Минимизация квадратичных функций происходит за один шаг.

На основе метода Ньютона разработан эффективный метод, получивший название *метода переменной метрики*. Идея метода заключается в использовании информации о градиенте критерия оптимальности для приближенного вычисления матрицы Гессе. Этот метод — итерационный. Поиск в нем ведется по формуле

$$\mathbf{X}_{k+1} = \mathbf{X}_k - h_k \mathbf{H}_k \text{grad } F(\mathbf{X}_k),$$

где \mathbf{H}_k — приближенно вычисляемая обратная матрица Гессе; h_k — величина шага, определяемая одномерной минимизацией целевой функции на луче — $\mathbf{H}_k \text{grad } F(\mathbf{X}_k)$.

Главное преимущество метода переменной метрики перед методом Ньютона — отказ от вычислений матрицы Гессе на каждой итерации. Положительно определенная матрица

$$\mathbf{H}_k = \mathbf{H}_{k-1} + \frac{\Delta \mathbf{X}_{k-1} (\Delta \mathbf{X}_{k-1})^T}{\langle \Delta \mathbf{X}_{k-1}, \mathbf{R}_{k-1} \rangle} + \frac{\mathbf{H}_{k-1} \mathbf{R}_{k-1} (\mathbf{R}_{k-1})^T \mathbf{H}_{k-1}}{(\mathbf{R}_{k-1})^T \mathbf{H}_{k-1} \mathbf{R}_{k-1}},$$

где $\Delta \mathbf{X}_{k-1}$ — приращение вектора переменных консультирования на предыдущем $(k-1)$ -м шаге, имеет форму вектор-столбца; $(\Delta \mathbf{X}_{k-1})^T$ — транспонированный вектор $\Delta \mathbf{X}_{k-1}$, т. е. вектор-строка;

$$\mathbf{R}_{k-1} = \text{grad } F(\mathbf{X}_k) - \text{grad } F(\mathbf{X}_{k-1})$$

— вектор-строка.

В начале вычислений нужно задаться произвольной положительно определенной матрицей \mathbf{H}_0 , в частности \mathbf{H}_0 может быть единичной матрицей. Шаг h_k выбирают по методу одномерной оптимизации.

Ввиду того что в методе переменной метрики достаточно полно учитывается локальная информация, его целесообразно применять в окрестности оптимального решения.

Методы одномерной оптимизации. Эти методы позволяют найти оптимум для функций одной переменной. ***Они не требуют условия дифференцируемости критерия оптимальности, а требуют только его непрерывности.*** Среди этих методов наиболее распространены ***методы чисел Фибоначчи, золотого сечения и полиномиальной аппроксимации.*** Для их применения необходимо знать интервал $[a, b]$ переменной x , на котором функция $F(x)$ имеет единственный минимум.

При применении ***метода чисел Фибоначчи*** должно быть зафиксировано число точек N , в которых производится вычисление критерия оптимальности.

Пусть на k -й итерации интервал изменения переменной x сократился до $[a_k, b_k]$, $a \leq a_k \leq b_k \leq b$. Тогда для вычисления следующего интервала $[a_{k+1}, b_{k+1}]$ выбирают точки x_k и x'_k (рис. 9.5) по формулам

$$x_k = \frac{\Phi_{N-1-k}}{\Phi_{N+1-k}} (b_k - a_k) + a_k; \quad x'_k = \frac{\Phi_{N-k}}{\Phi_{N+1-k}} (b_k - a_k) + a_k,$$

где Φ_k — числа Фибоначчи, определяемые с помощью рекуррентных соотношений

$$\Phi_k = \Phi_{k-1} + \Phi_{k-2}; \quad \Phi_0 = \Phi_1 = 1.$$

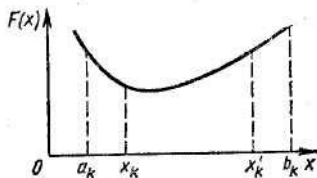


Рис. 9.5. Расположение интервалов поиска экстремума методом чисел Фибоначчи

Если $F(x_k) < F(x'_k)$, то в качестве следующего интервала выбирают $[a_{k+1}, b_{k+1}] = [a_k, x'_k]$, если $F(x_k) > F(x'_k)$, то выбирают $[a_{k+1}, b_{k+1}] = [x_k, b_k]$, если $F(x_k) = F(x'_k)$, то может быть выбран любой из интервалов.

Последние точки задаются формулами

$$x'_{N-1} = (0,5 + \varepsilon) (b_{N-1} - a_{N-1}) + a_{N-1};$$

$$x_{N-1} = 0,5 (b_{N-1} - a_{N-1}) + a_{N-1},$$

где ε — произвольно малое число, вводимое на последней итерации. Следует отметить, что длина последнего интервала неопределенности определяется как

$$b_N - a_N = (b_0 - a_0) / (2\Phi_N). \quad (9.44)$$

Выражение (9.44) позволяет определить количество вычислений критерия оптимальности исходя из требуемой точности поиска.

В *методе золотого сечения* сохраняется постоянным отношение длин двух последовательных интервалов неопределенности:

$$\tau = L_k / L_{k+1} = 1,61803\dots$$

По результатам двух экспериментов устанавливают, какую область неопределенности оставить для дальнейших исследований. Процесс поиска оптимума можно продолжать сколь угодно долго. После N испытаний длина интервала неопределенности составляет $L_N = 1/\tau^{N-1}$.

Отметим, что метод золотого сечения требует сравнительно небольшого объема памяти ЭВМ и прост в реализации.

Метод полиномиальной аппроксимации заключается в определении полинома, аппроксимирующего функцию $F(\mathbf{X})$ (чаще всего — квадратичного полинома), и поиске его минимума.

В конкретных задачах оптимального консультирования довольно часто зависимость критерия оптимальности F от параметров консультирования \mathbf{X} получается слишком сложной. В этих случаях вместо вышеизложенных регулярных методов оптимизации рекомендуется использовать *методы случайного поиска*. В этих методах направление поиска \mathbf{P}_k выбирают случайно, например, равномерно в пределах гиперсферы с центром в точке \mathbf{X}_{k-1} .

Существует огромное число алгоритмов случайного поиска. Следует отметить, что регулярные алгоритмы поиска являются частным (а точнее, вырожденным) случаем стохастических алгоритмов.

Методы условной оптимизации. Задачи условной оптимизации, заключающиеся в минимизации некоторого критерия оптимальности с ограничениями на область существования переменных консультирования, относятся к классу задач математического программирования.

Одним из наиболее простых и широко известных методов решения задачи математического программирования является **метод штрафных функций**. Основная идея метода состоит в приближенном сведении задачи минимизации функции $F(\mathbf{X})$ при ограничениях $Q_i(\mathbf{X}) \leq 0, i=1, 2, \dots, n$, к задаче минимизации функции

$$\Phi(\mathbf{X}, t) = F(\mathbf{X}) + tR(\mathbf{X}) \quad (9.45)$$

без ограничений. При этом вспомогательную функцию $\Phi(\mathbf{X}, t)$ подбирают так, чтобы она совпадала с функцией $F(\mathbf{X})$ внутри допустимой области S и быстро возрастала вне ее.

В выражении (9.45) $R(\mathbf{X})$ — дифференцируемая функция штрафа, удовлетворяет следующим условиям: $R(\mathbf{X}) = 0$, если $\mathbf{X} \in S$, и $R(\mathbf{X}) > 0$, если хотя бы для одного k будет $Q_k(\mathbf{X}) > 0, (k = 1, 2, \dots, n)$, t — некоторое положительное число — коэффициент штрафа.

Примерами функций $R(\mathbf{X})$ могут служить выражения

$$R_1(\mathbf{X}) = \sum_{i=1}^n \max R_i \{(\mathbf{X}), 0\}^\alpha, \quad \alpha \geq 1;$$

или

$$R_2(\mathbf{X}) = \sum_{i=1}^n \left[\frac{R_i(\mathbf{X}) + |R_i(\mathbf{X})|}{2} \right]^2.$$

Рис. 9.6 иллюстрируется метод штрафных функций в одномерном случае.

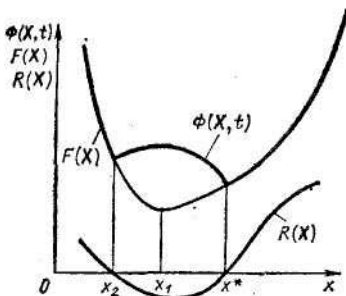


Рис. 9.6. Метод штрафных функций

Допустимая область S определяется ограничением $R(\mathbf{X}) \geq 0$, в этой области $F(\mathbf{X})$ и $\Phi(X, t)$ совпадают. В области, где $R(\mathbf{X}) < 0$, функция $\Phi(\mathbf{X}, t)$ резко возрастает. На рисунке \mathbf{X}_1 и \mathbf{X}^* — точки безусловного и условного минимумов.

В задачах условной оптимизации, в которых ограничения заданы только в виде неравенств, возможно построение обобщенного критерия оптимальности с помощью *барьерных функций*. Значения, принимаемые барьерной функцией, неограниченно возрастают при приближении к границе допустимой области.

Примером барьерной функции является

$$R(\mathbf{X}) = -\sum_{i=1}^n \ln [-R_i(\mathbf{X})].$$

Эта функция существует только внутри допустимой области S . Вне области S и на ее границе функция $R(\mathbf{X})$ не определена, а при приближении к границе области S она неограниченно возрастает.

Исходя из организации поиска условного оптимума иногда метод штрафных функций называют *методом внешней точки*, а метод барьерных функций — *методом внутренней точки*.

Таким образом, задачу нелинейного программирования удастся свести к задаче или последовательности задач безусловной минимизации.

9.5. Задачи формирования рекомендаций по оптимизации допусков и консультационных требований

Качество формируемых рекомендаций по решению задач консультируемой проблемы в значительной мере определяется характером постановки консультационной задачи параметрического синтеза, реализуемой при консультировании, т. е. тем, насколько сформулированные целевая функция и ограничения отражают объективно существующие требования к свойствам КП. При формализации КЗ такие требования выражаются в виде *условий функционирования* КП. Условие функционирования КП — это *требуемое соотношение между выходным параметром u_r значения которого зависят от принимаемых сформированных рекомендаций, и предельно допустимым значением — нормой $y_j^{(0)}$* . Величину $y_j^{(0)}$ будем называть также *консультационным требованием на параметр u_r* . Условия функционирования КП могут иметь одну из следующих форм:

$$y_j < y_j^{(0)}, \quad (9.46a)$$

$$y_j > y_j^{(0)}, \quad (9.46б)$$

$$y_{j\min}^{(0)} < y_j < y_{j\max}^{(0)}. \quad (9.46в)$$

Формы (9.46б) и (9.46в) могут быть сведены к форме (9.46а), поэтому в дальнейшем будем считать, что все условия функционирования КП в КЗ имеют вид (9.46а).

Область в пространстве **ХП** управляемых параметров, в которой все условия функционирования КП, а заданные прямые ограничения на управляемые параметры x_i вида

$$x_{i\min} \leq x_i \leq x_{i\max} \quad (9.46г)$$

удовлетворяются, будем называть **областью функционирования КП** Z_0 .

Задача формирования рекомендаций по оптимизации **допусков функционирования КП** сводится к определению размеров допусковой области Z_d и ее расположения в пространстве **ХП**. **Цель оптимизации допусков** — максимизация размеров области Z_0 при выполнении ограничений на степень несовпадения областей Z_0 и Z_0 .

Решение задачи оптимизации допусков выполняется в два этапа: на первом ищут какую-либо точку $X_0 \in Z_0$, называемую **опорной**; на втором определяют **оптимальную допусковую область**.

На первом этапе для определения опорной точки целесообразно использовать постановку задачи оптимизации параметров, известную под названием **максиминной постановки**. Последняя приводит к получению опорной точки внутри области Z_0 на достаточном удалении от границ, что удобно для реализации алгоритмов второго этапа.

При максиминной постановке вводится количественная оценка s_j степени выполнения j -го условия функционирования КП. Каждая из оценок s_j может носить детерминированный или статистический характер. При детерминированном подходе используют формулу

$$s_j = (y_j^{(0)} - y_j) / y_j^{(0)},$$

при статистическом — формулу

$$s_j = (y_j^{(0)} - M_j) / \delta_j,$$

где M_j — оценка математического ожидания; δ_j — оценка рассеяния параметра y_j .

В частности, величина δ_j может рассматриваться как некоторый весовой коэффициент, указание его физического смысла упрощает

правильное задание его численного значения без трудоемких статистических расчетов.

Так как вероятность надежного функционирования КП определяется главным образом наименьшей из вероятностей выполнения отдельных условий функционирования КП, то в первую очередь нужно **увеличивать наименьший из запасов** s_j . Поэтому в качестве целевой функции $F(\mathbf{X})$ следует выбрать наименьший из запасов, и задача оптимизации параметров КП формулируется как максиминная задача нелинейного программирования:

$$\max_{\mathbf{X} \in \mathbf{Z}_0} \min_i s_j(\mathbf{X}).$$

Ограничения задачи при этом будут прямые ограничения (9.46г).

Максиминный критерий запаса функционирования КП применим при наличии у консультируемой проблемы параметров с условиями функционирования КП любого вида. Этот критерий в зависимости от конкретной ситуации может рассматриваться либо как детерминированный, либо как статистический.

На втором этапе формулируется **задача вписывания гиперпараллелепипеда допусков** \mathbf{Z}_d в область функционирования \mathbf{Z}_0 . В этой задаче к исходным данным относятся:

1. Область функционирования КП \mathbf{Z}_0 , задаваемая прямыми ограничениями (9.46г) и условиями функционирования КП, приведенными к виду $y_j(\mathbf{X}) \leq y_j^{(0)}$.

2. Точка $\mathbf{X}_3 \in \mathbf{Z}_0$.

3. Соотношения α_i , между допусками g_i , равноценными с позиций затрат на их получение при производстве.

При решении задачи сначала нормируют управляемые параметры в соответствии с заданными соотношениями между допусками

$$u_i = \alpha_i (x_i - x_{ai}) / x_{ai} + 1,$$

где u_i — нормированный параметр x_i ; $\alpha_i = g_1 / g_i$.

При таком нормировании оптимальной формой допусковой области является **гиперкуб с ребрами, параллельными координатным осям**. Областям \mathbf{Z}_0 и \mathbf{Z}_d в нормированном пространстве управляемых параметров соответствуют области \mathbf{U}_0 и \mathbf{U}_1 .

Задача вписывания формулируется как **задача определения центра гиперкуба** $\mathbf{U}^* \in \mathbf{U}_0$, имеющего максимальную длину ребра при условии, что оценка рассогласования положений областей работоспособности КП \mathbf{U}_0 и допусковой области \mathbf{U}_1 не превышает заданной величины.

Рассмотрим один из алгоритмов вписывания, основанный на линеаризации зависимостей $y_j(\mathbf{U})$ и ориентированный на случай, когда

$U_1 \subset U_0$. Этот случай часто называют оптимизацией параметров и допусков в условиях 100%-ного выхода годных. Принимается также допущение о постоянстве знаков коэффициентов влияния $a_{ji} = \partial y_j / \partial x_i$ т.

$$\text{sgn}(a_{ji}) = \text{const}, \quad (9.47)$$

в пределах U_0 , $i=1, 2, \dots, n$; n — количество управляемых параметров.

Для j -го участка границы области U_0 , выражаемого уравнением

$$y_j^{(0)} - y_j(\mathbf{U}) = 0, \quad (9.48)$$

известно положение нормали, проходящей через центр \mathbf{U}^* гиперкуба.

Система уравнений этой нормали в параметрическом виде

$$u_i - u_i^* = a_{ji} h_j, \quad (9.49)$$

где h_j — параметр; a_{ji} определяется в точке пересечения нормали с поверхностью (9.48).

Запишем систему уравнений для диагонали искомого гиперкуба, имеющей ту же ориентацию, что и j -я нормаль (9.49):

$$u_i - u_i^* = \text{sgn}(a_{ji}) \gamma_j, \quad (9.50)$$

где γ_j — параметр.

В соответствии с (9.47) вычисление a_{ji} может быть выполнено в точке U_3 . Проведем линеаризацию границ (9.48) области U_0 , используя разложение $y_j(\mathbf{U})$ в ряд Тейлора в окрестности опорной точки U_3 :

$$y_j^{(0)} - y_j(\mathbf{U}_3) - \sum_{i=1}^n a_{ji} (u_i - u_i^*) - \sum_{i=1}^n a_{ji} (u_i^* - u_{3i}) = 0. \quad (9.51)$$

Обозначим

$$c_j(\mathbf{U}^*) = y_j^{(0)} - y_j(\mathbf{U}_3) - \sum_{i=1}^n a_{ji} (u_i^* - u_{3i}).$$

Тогда (9.51) перепишем в виде

$$c_j(\mathbf{U}^*) - \sum_{i=1}^n a_{ji} (u_i - u_i^*) = 0. \quad (9.52)$$

Точка пересечения диагонали (9.50) с границей (9.52) определяется из совместного решения уравнений (9.50) и (9.52):

$$\begin{aligned} c_j(\mathbf{U}^*) - \gamma_j \sum_{i=1}^n \text{sgn}(a_{ji}) a_{ji} &= 0, \\ \gamma_j &= c_j(\mathbf{U}^*) / \sum_{i=1}^n |a_{ji}|. \end{aligned} \quad (9.53)$$

Здесь γ_j имеет смысл половины длины ребра гиперкуба, имеющего центр в точке \mathbf{U}^* и вершину на гиперповерхности (9.52). Всего имеем

m условий функционирования КП и, следовательно, не более чем m вершин на границах U_0 .

Чтобы выполнить условие $U_1 \subset U_0$, нужно контролировать принадлежность области U_0 вершины, которой соответствует минимальная величина γ_q среди величин γ_j . Чтобы область U_1 имела максимально возможные размеры, нужно, максимизировать γ_q , варьируя положение центра U^* .

Таким образом, задача вписывания допусковой области в область функционирования КП формулируется как задача линейного программирования

$$\max_{U^* \in U_0} \gamma_q(U^*)$$

при ограничениях

$$\gamma_p(U^*) - \gamma_q(U^*) \geq 0; \quad u_i^* \geq 0; \quad \|U_s - U^*\| < \gamma_q.$$

Последнее ограничение ставится в том случае, если точка U_s должна принадлежать допусковой области.

Отметим, что основные затраты машинного времени на реализацию алгоритма связаны с анализом чувствительности. Анализ чувствительности методом приращений требует $n+1$ раз обращаться к математической модели КП. Первое обращение производится при значении вектора управляемых параметров U_s и позволяет вычислить $y_j(U_s)$, фигурирующие в (9.51). Каждое последующее обращение позволяет вычислить очередную строку матрицы чувствительности и в итоге дает значения a_{ij} . Теперь полностью определена линеаризованная модель КП (9.53). Манипулирование ею при решении задач линейного программирования не требует заметных затрат машинного времени.

При большом числе n управляемых параметров (несколько сотен и более) применение алгоритма, включающего анализ чувствительности методом приращений, становится нерациональным. С ростом n более предпочтительными оказываются алгоритмы, основанные на **статистических испытаниях**.

Одним из таких алгоритмов является **алгоритм центрирования по методу статистического градиента**. На каждом шаге алгоритма выполняются N статистических испытаний с выбором случайных точек в пределах некоторой области $U_{ст}$. По результатам испытаний выделяются те точки $U_p \in U_{ст}$, которые оказались в области функционирования КП U_0 . Для следующего шага в качестве координат u_i^* центра U^* допусковой области $U_{ст}$ принимаются средние арифметические значения координат u_{ip} выделенных точек U_p .

Если в процессе испытаний область U_{CT} выбирается в соответствии с некоторыми условиями (например, $U_0 \in U_{CT}$ и U_{CT} имеет форму гиперсферы) и если выполнено достаточное количество шагов, то точка U^* может быть принята в качестве центра области U_0 , а окончательная допусковая область U_d устанавливается в соответствии с характером распределения точек U_p на последнем шаге центрирования. Задача оптимизации допусков обычно решается на том иерархическом уровне консультирования, на котором в качестве управляемых параметров фигурируют **параметры базовых элементов**.

Сформированные рекомендации по расчету допусков используются для выбора унифицированных деталей и узлов по справочникам и каталогам либо служат непосредственными исходными данными для последующего технологического проектирования.

На промежуточных иерархических уровнях нисходящего функционального или конструкторского проектирования также возникают задачи, подобные задаче оптимизации допусков. Предположим, что на k -м иерархическом уровне управляемыми параметрами КП являются параметры y_j . На следующем, $(k+1)$ -м иерархическом уровне эти же параметры рассматриваются уже как выходные параметры подпроблем, а управляемыми параметрами здесь оказываются другие параметры x_i . Для выполнения консультирования на $(k+1)$ -м иерархическом уровне на выходные параметры y_j нужно задать условия функционирования КП. Очевидно, что эти условия должны быть результатом сформированных рекомендаций на k -м уровне, т. е. должны быть определены не только некоторая оптимальная точка Y^* в пространстве параметров y_j , но и консультационные требования $y_j^{(0)}$ на эти параметры.

Задача оптимального расчета консультационных требований $y_j^{(0)}$ по своей постановке и методам решения близка к рассмотренной выше задаче оптимизации допусков. Исходными данными при ее решении являются условия функционирования КП, задаваемые на параметры КП, а результатом должны быть условия функционирования КП для подпроблем. При оптимизации консультационных требований на последующем $(k+1)$ -м уровне исходными данными будут условия функционирования КП на параметры подпроблем, а результатом — условия функционирования КП для элементов подпроблем и т. д.

Таким образом, блочно-иерархический подход приводит к формулировке основных оптимизационных задач нисходящего консультирования как задач оптимального преобразования консультационного задания на КП k -го иерархического уровня в

консультационные задания на КП $(k+1)$ -го иерархического уровня. *Эти задачи решаются с помощью статистических или детерминированных алгоритмов вписывания гиперфигур в заданную область n -мерного пространства параметров.*

9.6. Методы формирования рекомендаций по оптимизации различных процессов

Комплексные САК охватывают все этапы консультирования проблем. Автоматизация формирования рекомендаций по решению консультационных задач различных процессов включает в себя разработку принципиальных схем процессов, маршрутной технологии, операционной технологии и получение управляющей информации на машинных носителях для программно-управляемого процессорного оборудования.

Любой процесс независимо от его физической природы всегда можно представить в виде некоторой КП, а следовательно, для его организации следует применять системный подход, сущность которого заключается в комплексном, едином рассмотрении всех частей технологических систем и в гармоническом их сочетании.

Постановку задачи оптимизации рекомендаций по решению задач проектирования заданного процесса - как консультируемой проблемы, можно представить следующим образом (рис. 9.7).

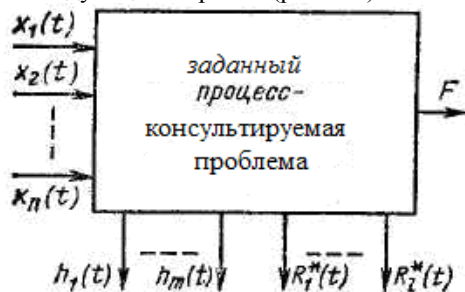


Рис. 9.7. Постановка задачи оптимизации рекомендаций по решению задач проектирования заданного процесса - как консультируемой проблемы

Заданный процесс рассматривается как КП, на вход которого поступает вектор $\mathbf{X}(t)=(x_1(t), \dots, x_n(t))$ входных переменных, а скалярный выход F характеризует качество заданного процесса. Требуется сформировать рекомендации по такому вектору управляющих воздействий

$$\mathbf{R}^*(t) = (R_1^*(t), \dots, R_l^{(*)}(t)),$$

который минимизировал бы значение показателя F , т. е.

$$F(\mathbf{X}(t), \mathbf{R}^*(t)) = \min_{\mathbf{R}(t)} \{F_k(\mathbf{X}(t), \mathbf{R}(t))\}.$$

Выходы $h_1(t), \dots, h_m(t)$ характеризуют состояние заданного процесса и индицируют нежелательные режимы работы процессных средств или выход контролируемых параметров заданного процесса за установленные пределы:

$$h_j(t) \leq b_j, \quad j = \overline{1, m},$$

где b_j — требуемые или допустимые значения соответствующих параметров.

Таким образом, оптимизация рекомендаций по решению задач проектирования заданного процесса рассматривается как задача определения оптимального вектора управления $\mathbf{R}_{\text{ит}}^*(t)$, минимизирующего целевую функцию $F(\mathbf{X}(t), \mathbf{R}(t))$ при условии выполнения заданных ограничений.

Подобные задачи оптимизации решают в два этапа. На первом этапе определяют идеальный вектор управления $\mathbf{R}_{\text{ит}}^*(t)$, обеспечивающий оптимизацию заданного процесса. Практически реализовать это не представляется возможным, и вектор $\mathbf{R}_{\text{ит}}^*(t)$ является эталоном, к которому надо стремиться. Зная $\mathbf{R}_{\text{ит}}^*(t)$, на втором этапе выбирают реализуемый квазиоптимальный вектор управления, с помощью которого стараются получить рекомендацию, наименее отличающуюся от идеальной и в то же время реализуемую наиболее просто.

Следует отметить, что в большинстве практических случаев оптимизированные рекомендации по решению задач проектирования заданных процессов дополнительно подвергаются наладке и корректировке, поскольку при построении математических моделей процессов невозможно учесть все влияющие на процесс факторы.

На любой заданный процесс оказывает влияние множество случайных факторов. Так, например, на технологический процесс изготовления механических изделий влияет неточность технических средств, неточности режущего инструмента и приспособлений, внутренние напряжения обрабатываемой детали, разброс параметров у электрорадиоэлементов при монтаже электронных схем и т. д. Поэтому параметры изготавливаемых изделий являются случайными величинами, причем влияние действующих случайных факторов на изменение параметров изделий можно, как правило, определить исходя из статистического анализа. В подобных случаях очень важно учитывать характер взаимосвязи между случайными величинами. Для

количественного выражения этой взаимосвязи служат регрессия и корреляция. Остановимся более подробно на этих понятиях.

Пусть x и y — случайные величины, характеризующие параметры некоторой КП, причем упорядоченная пара (x, y) характеризует параметры одного варианта рекомендации КП и может быть изображена точкой на плоскости. Полная совокупность вариантов рекомендаций изображается множеством точек, показанных на рис. 9.8. Математические ожидания случайных величин x и y равны соответственно $M_{(x)}$ и $M_{(y)}$, и среднеквадратичные отклонения σ_x и σ_y характеризуют рассеивание величин x и y относительно их математических ожиданий.

Рассмотрим зависимость $\bar{y}(x)$, являющуюся условным математическим ожиданием $M_{(y|x)}$. Используя выражение для условного математического ожидания и обозначая через $p(x, y)$ совместную вероятность данных значений x и y , находим

$$\bar{y}(x) = M_{(y|x)} = \frac{\sum_y y p(x, y)}{\sum_y p(x, y)}. \quad (9.54)$$

Определяя $\bar{y}(x)$ при различных x , можно построить линию, графически выражающую эту зависимость и называемую линией регрессии y по x (рис. 9.8).

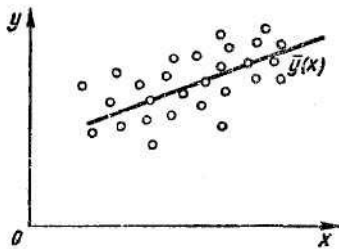


Рис. 9.8. Линия регрессии

Аналогично может быть получена зависимость $\bar{x}(y)$, называемая регрессией x по y .

На практике наиболее часто встречается случай линейной регрессии, уравнение которой записывается в виде

$$\bar{y}(x) = a + b(x - M_{(x)}). \quad (9.55)$$

Коэффициенты a и b выбирают такими, чтобы получить наибольшую концентрацию точек (x, y) вблизи прямой $\bar{y}(x)$, что выражается условием

$$\varphi(a, b) = M \{ [y - \bar{y}(x)]^2 \} = \min. \quad (9.56)$$

Выражение (9.56) с учетом (9.55) дает следующую систему уравнений для определения коэффициентов a и b :

$$M_{(y)} - a = 0; \quad M [y(x - M_{(x)})] - b\sigma_x^2 = 0. \quad (6.57)$$

Величину $\mu_{xy} = M[y(x - M_{(x)})]$ называют ковариацией между x и y . Она служит мерой взаимной связи между случайными величинами a и b .

Из (9.57) находят значения

$$a = M_{(y)} \text{ и } b = \mu_{xy} / \sigma_x^2,$$

определяющие линию регрессии.

Ковариация μ_{xy} зависит от дисперсий самих случайных величин, поэтому для оценки взаимосвязи между случайными величинами более удобен коэффициент корреляции $r_{xy} = \mu_{xy} / (\sigma_x \sigma_y)$, который может меняться от нуля для независимых случайных величин до единицы, если случайные величины связаны линейной функциональной зависимостью. При технологическом проектировании в качестве критериев оптимальности рекомендаций могут рассматриваться такие показатели эффективности, как себестоимость производства изделий, производительность технологических процессов, основное технологическое время и т. д.

Методы оптимизации формирования рекомендаций по решению задач проектирования заданных процессов рассмотрим на примерах управления технологическим процессом производства магнитоуправляемых контактов и управления технологической установкой.

Пример 9.6. Формирование рекомендаций по управлению технологическим процессом на основе текущего регрессионного анализа. Рассмотрим технологический процесс производства магнитоуправляемых контактов (МК). Основная задача производства МК — получение изделий с заданными величинами напряженности магнитного поля Θ и зазора S между контактами.

Одной из задач автоматизации формирования рекомендаций по проектированию технологического процесса производства МК является определение функциональной связи между величинами Θ и S с последующей реализацией математической модели процесса управления заварки лепестков МК на управляющей ЭВМ.

Схема управления заваркой лепестков МК приведена на рис. 9.9.

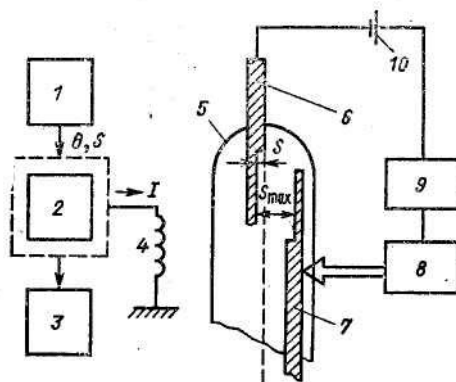


Рис. 9.9. Схема управления заваркой лепестков МК:

1 — пульты статистического контроля Θ и S ; 2 — управляющая ЭВМ; 3 — блок выдачи рекомендаций технологу; 4 — катушка МК; 5 — стеклянный баллон; 6, 7 — верхний и нижний пружинные контакты; 8 — исполнительный механизм; 9 — регистрирующее устройство; 10 — источник питания

Технологический процесс установки зазора по заданной напряженности магнитного поля состоит в следующем: заваривается верхний пружинный контакт 6; нижний пружинный контакт 7 отводится исполнительным механизмом 8 на максимальное расстояние; в катушку 4 подается ток, значение которого обеспечивает заданную величину напряженности магнитного поля для срабатывания МК; начинается движение контакта 7 к контакту 6.

Движение контактов происходит до момента их замыкания под действием поля катушки 4. Значение поля пропорционально силе тока I , протекающего в катушке. При замыкании регистрирующее устройство 9 срабатывает и останавливает исполнительный механизм 8; нижний пружинный контакт 7 заваривается, зазор становится равным величине S . Исследования показывают, что величины Θ и S после заварки не связаны жесткой функциональной зависимостью. Характер геометрического места точек $\Theta(S)$ зависит от многих факторов — качества исходного материала, режимов операции заварки и т. д.

С пультов статистического контроля I данные экспериментов поступают в управляющую ЭВМ 2, в которой непрерывно строятся регрессионные модели и вырабатываются рекомендации по реализации управляющих воздействий на технологический процесс

(корректируется сила тока в катушке, выдаются различные рекомендации).

Математически процесс управления заваркой МК можно свести к следующей модели. Пусть Θ и S — требуемые значения напряженности магнитного поля и зазора между контактами. Для решения задачи оптимизации сформулируем целевую функцию вида

$$F(\Theta, S) = \left(\frac{S_T - S}{S_T}\right)^2 + \left(\frac{\Theta_T - \Theta}{\Theta_T}\right)^2,$$

которую следует минимизировать с учетом ограничения $\Theta = kS + b$, определяемого уравнением регрессии.

Сформулированная задача является **задачей квадратичного программирования, которую можно решить с использованием неопределенных множителей Лагранжа.**

Функция Лагранжа имеет вид

$$F(\Theta, S, \lambda) = \left(\frac{S_T - S}{S_T}\right)^2 + \left(\frac{\Theta_T - \Theta}{\Theta_T}\right)^2 + \lambda(\Theta - kS - b).$$

Беря частные производные, получаем

$$\begin{aligned} \frac{\partial F}{\partial S} &= -2 \left(\frac{S_T - S}{S_T}\right) - k\lambda = 0; \quad \frac{\partial F}{\partial \Theta} = \\ &= -2 \left(\frac{\Theta_T - \Theta}{\Theta_T}\right) + \lambda = 0; \quad \frac{\partial F}{\partial \lambda} = \Theta - kS - b = 0. \end{aligned} \quad (9.58)$$

Решением системы будут значения S^* и Θ^* , при этом Θ^* косвенно определяет силу тока I в катушке.

Пусть $\Theta_T = 40 \pm 2$ А/м; $S_T = 10 \pm 2$ мкм; $\Theta = 3S + 6$; число витков в катушке $\omega = 50$.

Решая систему уравнений (9.58), получаем $\Theta^* = 40,44$ А/м, $S^* = 10,48$ мкм.

Оптимальные значения Θ и S находятся в поле допуска. Необходимый ток в катушке $I = 0,8$ А. После установки $I = 0,8$ А строят новую регрессионную зависимость и рассчитывают новое значение I . В результате многократной коррекции силы тока I находят такое значение, при котором процент выхода МК, отвечающих требованиям консультационных условий, будет максимальным.

Пример 9.7. Формирование рекомендаций по оптимизации управления технологической установкой.

При конструкторском проектировании и при технологической подготовке производства узлов ЭВА для формирования графической информации используют приборы последовательного действия (ППД),

в частности координатографы, сверлильные станки с числовым программным управлением и др.

Управление ППД производится подачей на их входы совокупности команд A , которые формируют изображение M . Требуется сформировать рекомендации для получения такой совокупности команд, которая обеспечивает функционирование ППД (получение изображения M) с минимальным числом переходов от фрагмента к фрагменту.

Допустим, выработана последовательность команд $A_s = (\alpha_1, \dots, \alpha_n)$ с индексом s . Тогда время работы ППД при формировании полного изображения M определяется соотношением

$$T^{(s)} = T_n^{(s)} + T_M = \sum_{i=1}^{n-1} t_{\alpha_i, \alpha_{i+1}} + \sum_{i=1}^n \Delta t_i,$$

где $T_n^{(s)}$ — общее время переходов между фрагментами изображения; T_M — время выполнения команд ($T_M = \text{const}$); $t_{\alpha_i}, t_{\alpha_{i+1}}$ — время перехода от команды α_i к команде α_{i+1} ; Δt_i — время формирования i -го фрагмента изображения M .

Стоимость формирования полного изображения M последовательностью команд A , с учетом затрат на оптимизацию последовательности будет

$$C^{(s)} = C_{\text{опт}}^{(s)} + C_{\text{ППД}}^{(s)} = C_{\text{ЭВМ}} T_{\text{ЭВМ}}^{(s)} + C_{\text{ППД}} T^{(s)} q,$$

где $C_{\text{ЭВМ}}, C_{\text{ППД}}$ — соответственно удельная стоимость работы ЭВМ, производящей оптимизацию, и ППД; $T_{\text{ЭВМ}}^{(s)}$ и $T_{\text{ППД}}^{(s)}$ — соответственно времени работы ЭВМ и ППД; q — необходимое число копий изображения M .

Для оптимизации последовательностей A обычно выбирают такие методы, для которых

$$C_{\text{опт}}^{(s)} \ll C_{\text{ППД}}^{(s)}.$$

Таким образом, оптимизация последовательных команд A заключается в минимизации переходов T_n .

Для получения решения сформируем матрицу $\|t_{ij}\|_{n \times n}$, в которой элемент t_{ij} равен времени перехода от выполнения команды α_i к команде α_j .

Введем псевдодобулевы переменные

$$x_{ij} = \begin{cases} 1, & \text{если } i = \alpha_h, j = \alpha_{h+1}; \\ 0, & \text{если } i = \alpha_h, j \neq \alpha_{h+1} \end{cases}$$

и предположим, что последовательность команд образует замкнутый контур (начинается и заканчивается командой a_i). Тогда наша задача совпадает с классической задачей о коммивояжере

$$F(x) = \sum_{1 \leq i+j \leq n} t_{ij} x_{ij} \rightarrow \min$$

при ограничениях

$$\sum_{\substack{i=1 \\ i+j}}^n x_{ij} = 1, \quad j = \overline{1, n}; \quad \sum_{\substack{j=1 \\ j+i}}^n x_{ij} = 1, \quad i = \overline{1, n}, \quad (9.59)$$

$$u_i - u_j + nx_{ij} \leq n - 1, \quad i, j = \overline{1, n}, \quad (9.60)$$

где u_i, u_j — произвольные действительные числа.

Условия (9.59) гарантируют непрерывность маршрута, а условие (9.60) — получение последовательности команд, не распадающейся на ряд замкнутых циклов.

Задача может быть решена методом ветвей и границ.

9.7. Формирование рекомендаций при решении консультационных задач функционального и структурного синтеза КП

При консультировании проблемы консультант, как правило, определяет совокупность функциональных компонент, из которых состоит КП, вводит отношения между определенными компонентами и принимает это как начальное представление о КП.

Далее перед консультантом возникает задача полноты определенного множества функциональных структур, их избыточности и согласованности по параметрам. В терминах общей теории консалтинга это сводится к *решению проблемы синтеза рекомендаций по созданию или консультированию структуры КП*, которая представляется следующим образом: *задана система данных, включающая в себя множество компонент проблемы и отношения между ними; необходимо идентифицировать наилучшее представление этой системы с помощью структуры проблемы, элементы которой связаны с подмножествами переменных, описывающих компоненты.*

Процедура решения состоит из генерации наиболее приемлемых с точки зрения начального поведения проблемы структурных предположений рассматриваемой проблемы, анализа каждого из них на основе задающей системы данных и начальной структуры и

сравнения результатов анализа с исходными данными с использованием различных критериев оценки полученного структурного предположения.

Например, при проектировании САК можно выделить четыре рекомендуемые компоненты, обеспечивающие *управление, диалог, поиск, вычисления*. Для каждой конкретной САК, которую будем рассматривать как консультируемую проблему, эти компоненты могут быть различными способами реализованы.

Всего для этих компонент можно получить 2^{2n} структур САК. Естественно, не все структуры этого множества приемлемы для более детального анализа (уточнение параметров выделенных компонент системы и определение требований к связям между этими компонентами).

Задачей синтеза рекомендаций в данном случае является формирование рекомендаций по выбору структуры из множества возможных структур, которая обеспечивала бы устойчивое и эффективное функционирование КП. Для проведения синтеза структур КП необходимо сформировать некоторую начальную структуру, т. е. выделить компоненты КП и определить отношения между ними. Далее, используя метод, изложенный ниже, можно синтезировать рекомендации для получения новых структур проблемы. Однако проводить синтез структур, не зная, как оценивать синтезируемые структуры, бессмысленно. Следовательно, необходимо иметь рекомендованный метод, который позволял бы оценивать структуры и выбирать наиболее приемлемые для последующего анализа и синтеза. Чтобы этот метод использовался при решении задачи синтеза структур КП, он должен обладать инвариантными свойствами и обеспечивать формальное описание КП.

Для структурных предположений введем частичный порядок. В этом случае множество структурных предположений вместе с частичным порядком образуют решетку, которая позволяет проводить процесс поиска «наилучшего» структурного предположения не по всему множеству возможных структурных предположений, а только по ограниченной части его.

Определим структуру как семейство подмножеств множества переменных V . Множеством всех возможных структур по отношению к множеству V будет множество

$$S = \{S_i | S_i \in P(V)\},$$

где $P(V)$ — мощность множества V .

Проведем классификацию введенного множества S . Определим отображение

$$r: S \rightarrow R,$$

где R — множество всех симметричных бинарных отношений, определенных на множестве V , и $r(S_i)$ — бинарное отношение, в котором переменные v_i и v_j связаны тогда и только тогда, когда они обе принадлежат, по крайней мере, одному из подмножеств множества V , через которое определена S_i ;

$$r(S_i) = \{(v_i, v_j) | v_i, v_j \in V, \exists (E_a \in S_i)(v_j \in E_a, v_i \in E_a)\},$$

где E_a — подсистема структуры S_i .

Отображение r не взаимнооднозначное, так как $S_i \equiv S_j$, если $r(S_i) = r(S_j)$. Будем использовать символ S/r для обозначения эквивалентных классов в S посредством r .

Некоторые структуры из множества S могут не сохранять первоначального понятия об объекте исследования, следовательно, необходимо, вводя некоторые ограничения на множество S , сузить множество всех возможных структур до множества значимых структур, дающих полное представление о проблеме. Введем эти ограничения:

— множество связанных переменных структуры должно быть полным;

— каждый элемент структуры должен быть определен непустым множеством переменных;

— не должно существовать элемента (подпроблемы), который состоял бы только из переменных, включенных полностью в некоторый другой элемент этой же структуры. В результате получается множество S_G , называемое множеством приемлемых структур:

$$S_G = \left\{ S_G^i \mid S_G^i \subset P(V), \bigcup_{E_a \in S_G^i} E_a = V, \right. \\ \left. E_a \not\subset E_b, \forall E_a, E_b \in S_G^i (a \neq b) \right\}.$$

Пусть требуется разработать комплекс программ, который позволял бы генерировать рекомендации по созданию всего множества приемлемых структур и оценивать их для заданного множества переменных. Как отмечалось ранее, множество приемлемых структур вместе с определенным на этом множестве частичным порядком образуют решетку. Рассмотрим подробно понятие «частичный порядок» для множества приемлемых структур.

Предположим, что S_G^i и S_G^j принадлежат множеству S_G , где $S_G^i \leq S_G^j$ тогда и только тогда, когда для каждого множества $E_a \in S_G^i$ существует множество $E_b \in S_G^j$ такое, что $E_a \subseteq E_b$. Если $S_G^i \leq S_G^j$, то будем называть S_G^i «уточнением» структуры S_G^j и, наоборот, S_G^j

«агрегатом» S_G^i . Кроме того, S_G^i — непосредственное «уточнение» S_G^j тогда и только тогда, когда $S_G^i \leq S_G^j$ и не существует такой S_G^k , что $S_G^i \leq S_G^k$ и $S_G^k < S_G^j$. Аналогично определяется непосредственный «агрегат».

Определим отображение как

$$r_G: S_G \rightarrow R^*,$$

где R^* — множество всех симметричных и рефлексивных бинарных отношений, определенных на множестве V . Это множество можно интерпретировать как множество ненаправленных графов с петлями. Символом S_G/r_G обозначим множество классов эквивалентности, определенных на множестве приемлемых структур посредством r_G . Каждому классу эквивалентности соответствует единственная каноническая структура, содержащая минимальное число элементов. Между множеством канонических структур S_C и множеством ненаправленных графов, определенных на данном множестве переменных, существует взаимно однозначное соответствие. Каноническая структура всегда будет агрегатом для структур, принадлежащих классу эквивалентности, определяемому этой структурой. Введем также структуры, которые будут являться уточнением для любой структуры из определенного класса эквивалентности. Они построены так же, как пары переменных, которые связаны в графе, представляющем класс эквивалентности, и как единичные переменные, которые изолированы в графе. Будем называть введенные структуры соответственно C - и P -структуры. Очевидно, что $C \geq P$ и все остальные структуры (из определяемого этими структурами класса эквивалентности) будут всегда находиться между C и P .

Из этого следует, что структурные предположения, не входящие в указанное множество S_G , могут быть игнорированы без каких-либо потерь для множества анализируемых синтезированных структурных предположений.

Процесс синтеза структурных предположений КП можно проводить различными способами в зависимости от ограничений, накладываемых консультантом на начальную структуру КП.

Во-первых, если консультант считает, что все связи между функциональными компонентами, заданными в начальной структуре, «жесткие», т. е. удаление какой-либо связи повлечет за собой коренное изменение в функциональной принадлежности начально определенных агрегатов компонент проблемы, то процесс синтеза новых структурных предположений необходимо проводить в рамках выделенного класса эквивалентности, соответствующего начальной структуре. В этом

случае механизм синтеза структурных предположений заключается в агрегировании либо в декомпозиции элементов начальной структуры КП.

Во-вторых, если консультант считает, что удаление связей возможно и даже необходимо, то процесс синтеза основывается на переходе из одного класса эквивалентности структурных предположений в другой, причем связи можно и удалять и добавлять.

Для правильного понимания физического смысла таких понятий, как «удаление связей», «добавление связей», «агрегирование и декомпозиция элементов» и т. д., необходимо уяснить одно из центральных понятий структурного синтеза — поведение КП, которое описывается следующим образом:

— определяется множество переменных, через которые наблюдается КП — множество V (например, множество функциональных компонент, необходимых для функционирования КП);

— каждой переменной задается определенное состояние из множества состояний данной переменной. Совокупность состояний множества переменных — множество X (например, конкретные реализации функциональных модулей);

— на множестве V определяется функция f , ставящая в соответствие каждой переменной (функциональной компоненте) определенное состояние (конкретную реализацию);

— определяется множество агрегатов состояний A (в данном случае — множество возможных конфигураций консультируемой проблемы);

— вводится отображение b , ставящее в соответствие каждому агрегату состояний некоторое число их $[0, 1]$, в данном случае — назначение каждой конфигурации консультируемой проблемы определенной вероятности (если сложно выделить приоритеты различных конфигураций КП, то возможные конфигурации принимаются равновероятными).

Описав таким образом поведение КП, нетрудно проследить качественные изменения, происходящие в процессе синтеза новых структурных предположений.

В качестве исходных данных для процедур синтеза имеется начальная структура КП, а также сформулированное описанным способом поведение КП. Используя процедуры генерации, можно получить множество новых структурных предположений в рассматриваемой проблеме и оценить эти сгенерированные структурные предположения.

Для проведения оценки необходимо выяснить, как соотносятся эти сгенерированные структурные предположения с начальным поведением проблемы. Предлагается использовать структурное предположение (совокупность подмножеств множества переменных или функциональных модулей КП) в качестве некоторой маски, накладываемой на начальное поведение КП (множество реализаций рекомендаций) для получения нового поведения, соответствующего сгенерированному структурному предположению. Далее эти оба поведения (новое и начальное) можно сравнить по множеству агрегатов состояний A и значению вероятностей появления данного конкретного агрегата в поведении КП, определенном структурным предположением. Естественно, новое поведение, соответствующее сгенерированному структурному предположению, может иметь агрегаты состояний (возможные конфигурации проблемы), которые не были описаны в начальном поведении. В этом случае к рассмотрению этих агрегатов состояний необходимо подойти с особой внимательностью, и если полученные агрегаты состояний представляют собой допустимые конфигурации, то целесообразно включить эти агрегаты состояний в начальное поведение и последующие итерации проводить уже с модифицированным начальным поведением. Таким образом, структура и начальное поведение определяют поведение КП, имеющей эту структуру. Следовательно, одним из критериев выбора структуры КП может являться факт сохранения начального поведения КП для сгенерированного структурного предположения. Сохранение означает совпадение у начального и нового поведения агрегатов состояний и значений вероятностей появления этих агрегатов.

В этом случае структурные предположения, полученные в результате первой итерации и сохранившие начальное поведение, выбираются в качестве начальных структур для последующей итерации и т. д.

В конкретных реализациях сформированных рекомендаций, естественно, кроме критерия выбора структур для последующих итераций используется и ряд других критериев, например минимизация количества подпроблем в структуре и т. п.

Учитывая тот факт, что совокупность структурных предположений образует решетку, можно определить направление синтеза новых структурных предположений.

С чисто практической точки зрения наиболее интересным является синтез структурных предположений в рамках определенного класса эквивалентности. Как правило, консультант считает свою модель КП

(начальное поведение и структуру) «жесткой», т. е. установленные им связи (отношения между состояниями переменных) неизменными. На самом деле, начальная структура в большинстве случаев избыточна и содержит различного рода противоречия. Для того, чтобы корректно задать начальное поведение и структуру КП, необходимо на первом шаге промоделировать исходное структурное предположение. Используя метод интерпретирующего структурного моделирования, можно выделить компоненты связанности в проблеме, проверить структуру КП на достижимость переменных, определить уровни иерархии КП и т. д. Только после проведенного таким образом анализа начальной структуры можно приступать к синтезу новых структурных предположений КП.

Например, формируя рекомендации для проектирования системы документирования, на вход которой поступает информация от других систем, и необходимо сформировать на выходе этой системы описания свойств, спецификацию компонент в виде таблицы, а также построить структурную схему. Представим, что состояние информации в системе описывается четырьмя переменными $V = (v_1, v_2, v_3, v_4)$:

$$v_1 = \begin{cases} 1 & \text{— наличие входной информации;} \\ 0 & \text{— отсутствие входной информации;} \end{cases}$$

$$v_2 = \begin{cases} 0 & \text{— отсутствие описания;} \\ 1 & \text{— описание свойств;} \\ 2 & \text{— описание свойств и элементов;} \\ 3 & \text{— описание свойств и внутренних связей;} \\ 4 & \text{— описание свойств, элементов и внутренних связей;} \end{cases}$$

$$v_3 = \begin{cases} 0 & \text{— отсутствие спецификации компонентов;} \\ 1 & \text{— наличие таблицы спецификации;} \end{cases}$$

$$v_4 = \begin{cases} 0 & \text{— отсутствие структурной схемы;} \\ 1 & \text{— наличие структурной схемы.} \end{cases}$$

Пусть в результате предварительного анализа было получено требуемое поведение системы, представленное табл. 9.1.

Таблица 9.1

v_1	v_2	v_3	v_4
0	0	0	0
1	1	0	0
1	2	1	0
1	3	0	1
1	4	1	1

Появление каждого агрегата в этом поведении принимается равновероятным с вероятностью, равной 0,2.

Пусть в качестве начальной структуры выбрана структура вида

$$S_0 = (v_1, v_2, v_3, v_4).$$

Это означает, что S_0 состоит из одного элемента и все переменные попарно связаны. Далее, применяя процедуры декомпозиции данной структуры на подсистемы, получаем новые структурные предположения.

В соответствии с начальным поведением для каждой синтезированной структуры вычисляются поведение, а также степень отличия вычисленного поведения от начального. Например, для структур

$$S_1 = (v_1, v_2, v_3)(v_2, v_3, v_4);$$

$$S_2 = (v_1, v_2, v_4)(v_2, v_3, v_4);$$

$$S_3 = (v_1, v_2, v_3)(v_1, v_2, v_4)$$

вычисленное поведение совпадает с начальным и степень отличия равна нулю. Взяв эти структуры в качестве исходных для последующих итераций синтеза и применив процедуру декомпозиции, получаем множество структур, в которых изменились связи между переменными по сравнению с исходными структурами. Для полученных структур также вычисляется поведение и сравнивается с исходным.

Структуры, выделенные из множества синтезированных структур, согласно критерию сохранения поведения систем показаны на рис. 9.10.

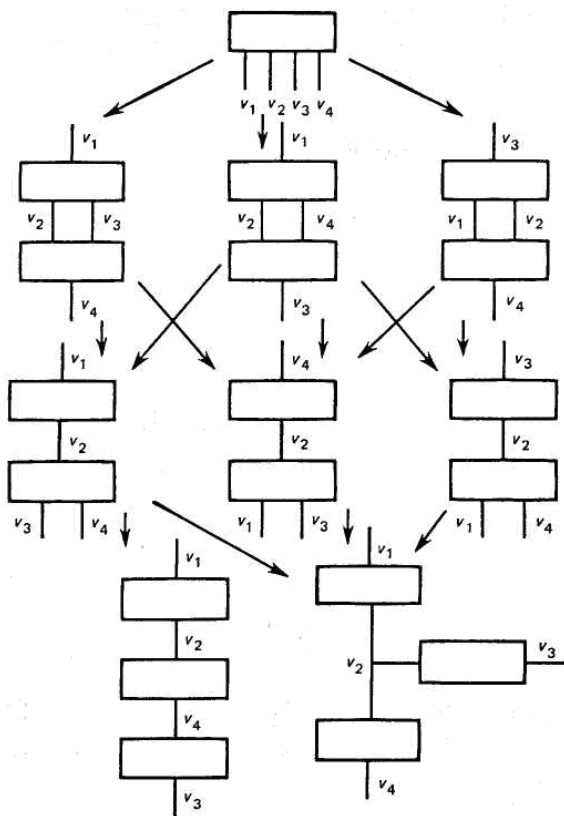


Рис. 9.10. Возможные структуры системы документирования

В качестве структуры проектируемой системы можно принять любую из этих структур, но необходимо также учитывать некоторые другие критерии, по которым можно сузить это множество, например критерий минимума подсистем в системе или минимума связей между подсистемами. Проведенный анализ синтезированных структур показал, что начальному поведению системы удовлетворяют и наиболее приемлемы структуры, изображенные на рис. 9.11.

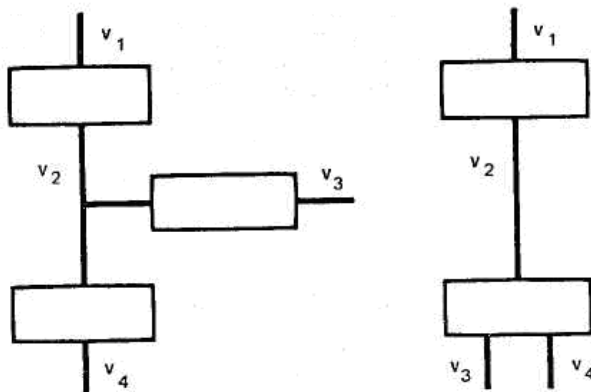


Рис. 9.11. Структуры системы, удовлетворяющие начальному поведению

9.8. Постановка задачи синтеза КП

Чтобы определить функции и структуры КП, необходимо:

- 1) составить задачу функционирования КП;
- 2) выбрать алгоритм их решения;
- 3) сформировать общую структуру КП и распределить задачи ее функционирования по уровням КП;
- 4) скомпоновать комплекс технических средств КП и ее подпроблем.

Перечисленные выше этапы синтеза КП взаимно связаны, и задачи каждого из них решаются с учетом ресурсов, выделяемых на консультирование проблемы. Синтезированная КП считается оптимальной, если достигается максимум (минимум) выбранного показателя эффективности, отражающего основные свойства КП с точки зрения выполнения поставленных задач.

На начальном этапе синтеза КП функции КП представляются в виде совокупности взаимосвязанных задач, которые, в свою очередь, могут быть разбиты на совокупности операций. По известным характеристикам операций и их взаимосвязей могут быть получены соответствующие характеристики задач.

При формализации взаимосвязей между функциями обычно учитываются порядок следований операций и их длительности (временные связи). Формализация взаимосвязей обычно производится на основе теории графов. Постановки задачи синтеза КП будем осуществлять, исходя из следующих условий:

- 1) задано или подлежит выбору множество реализуемых функций;
- 2) заданы или подлежат выбору взаимосвязи между функциями, в том числе и временные;
- 3) заданы или подлежат выбору элементы комплекса технических средств (КТС);
- 4) учитывается или не учитывается расположение элементов на плоскости (в пространстве);
- 5) заданы или подлежат выбору связи между элементами КП;
- 6) учитывается или не учитывается возможность выполнения задачи несколькими элементами.

Одновременно с этими условиями задачи синтеза КП могут различаться:

- а) видом показателей эффективности;
- б) типом учитываемых характеристик элементов;
- в) видом ограничений, накладываемых на учитываемые ресурсы (временные, сырьевые, технико-экономические и т. д.).

При решении задач синтеза КП находят применение различные модели и методы. Широкое распространение, как мы знаем, получили модели математического программирования. Рассмотрим постановку задачи синтеза КП на этапе синтеза структуры КП по изготовлению объектов.

Считаем, что предполагаемая топология размещения модулей КП известна; функции, реализуемые модулями, перечислены в виде последовательности задач, которые необходимо распределить между модулями КП. При построении структуры КП необходимо распределить функции (задачи) ($i = 1, 2, \dots, I$) между модулями КП ($j=1, 2, \dots, J$), выбрать алгоритмы реализации функции ($k = 1, 2, \dots, K$), типы технических средств модуля ($l = 1, 2, \dots, L$) и вариант организации функциональной связи между модулями. При этом должны учитываться такие технико-экономические характеристики КП, как затраты на создание A и эксплуатацию B , длительность выполнения цикла реализации функции (задачи) T , надежность КП P , масса W , энергопотребление E и др. Для формализации постановки задачи введем переменные: $x_{ikjl} = 1$, если i -я функция (операция) реализуется k -м вариантом в j -м модуле при помощи l -го технического средства; $x_{jl} = 1$, если j -й модуль имеет в своем составе l -е техническое средство; $x_{ji}=1$, если необходимо создать функциональную связь между модулями j и j' ; $x_{ikjl} = x_{jl} = x_{ji} = 0$ — в противном случае. Тот факт, что каждый вариант построения структуры должен включать в себя лишь один способ распределения функций (операций) по модулям проблемы и один способ выполнения каждой задачи, учитывает ограничение.

$$\sum_{k, l, i} x_{ijkil} = N, \quad i = 1, 2, \dots, I.$$

Переменные x_{jl} , x_{jj}' зависят от x_{ikjl} и используются для удобства записи аналитических выражений качества различных характеристик вариантов структуры:

$$x_{jl} = \begin{cases} 1, & \text{если } \sum_{i, k} x_{ijkil} \geq 1; \\ 0, & \text{если } \sum_{i, k} x_{ijkil} = 0; \end{cases}$$

$$x_{ll'} = \begin{cases} 1, & \text{если } \sum_{ikl'k'll'} x_{ijkil} x_{l'k'l'l'} \geq 1; \\ 0, & \text{если } \sum_{ikl'k'll'} x_{ijkil} x_{l'k'l'l'} = 0. \end{cases}$$

Для определения характеристик вариантов структуры введем следующие обозначения: A_l — стоимость технического средства либо затраты на разработку и изготовление перспективных средств; A_{jj}' — стоимость создания функциональной связи; A_{ikjl} — затраты на разработку необходимых видов обеспечения решения (реализации) l -й функции (операции) в k -м варианте в j -м модуле при наличии l -го технического средства; $B_{ij(i+1)j'}$ — затраты, понесенные в результате перехода от i -й функции (операции), реализуемой в j -м модуле, к $(i + 1)$ -й функции (операции), реализуемой j' -м модуле; t_{ikjl} — время реализации i -й функции k -м способом в j -м модуле при наличии l -го технического средства; $t_{ij(i+1)j'}$ — время, затрачиваемое на переход от i -й функции, реализуемой в j -м модуле, к $(i + 1)$ -й функции, реализуемой в j' -м модуле; P_{ikil} — надежность реализации функции (технического средства), определяемая либо вероятностью безотказной работы, либо ее функцией, $P_{ij(i+1)j'}$ — надежность функциональной связи, определяемая как и P_{ikil} ; W_e — масса технического средства; E_e — энергопотребление технического средства.

Используя введенные обозначения, запишем математические выражения для определения характеристик различных вариантов структуры КП.

Капитальные затраты A включают в себя стоимость технических средств в модуле, стоимость создания функциональных связей между модулями КП, затраты на разработку алгоритмов выполнения функций (операций):

$$A = \sum_{l, l'} A_l x_{ll} + \sum_{l, l'} A_{ll'} x_{ll'} + \sum_{i, k, j, l} A_{ijkil} x_{ijkil}.$$

Эксплуатационные затраты B включают в себя затраты на реализацию функционирования КП и затраты на транспортирование полуфабрикатов между модулями:

$$B = \sum_{i, k, j, l} B_{ikjl} x_{ikjl} + \sum_{k', l, k, j, l, l', l''} B_{il(i+1)l'} x_{ikjl} x_{(i+1)k'j'l''}.$$

Время выполнения цикла функционирования КП вычисляется аналогично:

$$T = \sum_{i, k, j, l} t_{ikjl} x_{ikjl} + \sum_{k', l, k, j, l, l', l''} t_{il(i+1)l'} x_{ikjl} x_{(i+1)k'j'l''}$$

При определении надежности КП будем считать КП невосстанавливаемой, в результате чего ее надежность, характеризуемая вероятностью безотказной работы элементов, равна произведению вероятностей безотказной работы элементов. Поэтому

$$\bar{P} = \sum_{i, j, k, l} \bar{p}_{ikjl} x_{ikjl} + \sum_{k', l, j, k, l, l', l''} \bar{p}_{il(i+1)l'} x_{ikjl} x_{(i+1)k'j'l''}, \quad (9.61)$$

где

$$\bar{p}, \bar{p}_{ikjl}, \bar{p}_{il(i+1)l'}$$

— логарифмы соответствующих величин. Масса и энергопотребление технических средств в модулях КП:

$$W_j = \sum_{i, k, l} W_i x_{ikjl};$$

$$E_j = \sum_{i, k, l} E_i x_{ikjl}, \quad j = 1, 2, \dots, J. \quad (9.62)$$

При решении задачи синтеза структуры любая из указанных характеристик может быть выбрана в качестве показателя эффективности (в зависимости от цели и назначения КП), который оптимизируется, а другие учитываются в ограничениях. Поэтому в общем виде задача оптимизации структуры КП выглядит следующим образом:

$$\min (\max) f(x_{ikjl}, x_{jl}, x_{jj'}). \quad (9.63)$$

При ограничениях:

$$\begin{aligned} f_s(x_{ikjl}, x_{jl}, x_{jj'}) &\leq D_s, \quad s = 1, 2, \dots, S; \\ \sum_{k, j, l} x_{ikjl} &= N, \quad i = 1, 2, \dots, I; \\ x_{ikjl} &= (0; 1); \quad x_{jl} = (0; 1); \quad x_{jj'} = (0; 1). \end{aligned} \quad (9.64)$$

Из приведенной постановки задачи следует, что она является задачей дискретного программирования и носит комбинаторный характер. При решении таких задач возникают трудности принципиального характера, а именно, необходимо исключить явный перебор всех допустимых решений и стремиться к эффективному частичному перебору сравнительно малого числа допустимых вариантов решения соответствующей задачи и неявному перебору остальных.

9.9. Критерии синтеза

Выбор путей консультирования проблемы будем производить в два этапа. На первом этапе решим вопрос о целесообразности проведения консультационных работ по результатам оценки экономической эффективности проблемы, на втором, если принято решение о целесообразности выполнения консультационных работ, произведем выбор рациональной структуры и оптимального варианта КП с учетом определенного числа факторов и условий.

Рассмотрим основное содержание первого этапа. Прежде всего следует выполнить прикидочную оценку целесообразности проведения консультационных работ, установить их рентабельность. Разработка и внедрение консультационных рекомендаций представляет практический интерес, если суммарный экономический выигрыш за будет превышать суммарные расходы на их формирование и реализацию. Это означает, что в начальный период времени реализованные рекомендации окупают те затраты, которые в них вложены, а в дальнейшем до конца жизненного цикла проблемы дают экономию (прибыль).

Основной критерий экономической эффективности (абсолютная прибыль от реализации рекомендаций) выразится соотношением

$$F(x) = \sum_{i=1}^n F_i x_i, \quad (9.65)$$

где F_i — абсолютная прибыль от внедрения i -го варианта рекомендации.

Значение F_i определим по формуле:

$$F_i = u_i T - C_i(T). \quad (9.66)$$

Здесь u_i — среднегодовой экономический выигрыш при внедрении i -го варианта рекомендации; C_i — суммарная стоимость

$$C_i(T) = C_{0i} + s_i T, \quad (9.67)$$

где C_{0i} — стоимость капиталовложений при реализации i -го варианта рекомендации; s_i — годовые эксплуатационные расходы от реализации i -го варианта рекомендации.

При $F_i \leq 0$ реализация i -го варианта рекомендации является нецелесообразной, так как такой вариант рекомендации является нерентабельным.

Оценка целесообразности выполнения консультационных работ производится по наилучшему варианту, удовлетворяющему соотношению

$$F^* = F(X^*) = \max \sum_{i=1}^n F_i x_i. \quad (9.68)$$

Можно оценку целесообразности производить по совокупности критериев (9.65), (9.68), пользуясь *теорией аддитивной последовательности*. Тогда, обозначив:

$\bar{F}_1(X)$ — относительное значение абсолютного дохода от реализации рекомендации;

$\bar{F}_2(X)$ — относительное время, в течение которого реализованная рекомендация дает прибыль, окупив полную стоимость ее реализации.

Запишем

$$\bar{F} = \bar{F}_1(X)g_1 + \bar{F}_2(X)g_2. \quad (9.69)$$

где g_1 и g_2 — вес каждого из критериев (в частности, можно положить, что $g_1 = g_2 = 1$).

Наилучший вариант рекомендации, используемый для оценки целесообразности ее разработки, определим на основании соотношения:

$$\bar{F}^* = \bar{F}(X^*) = \max \sum_{i=1}^n (\bar{F}_{i1}g_1 + \bar{F}_{i2}g_2) x_i, \quad (9.70)$$

где

$$F_{i1} = u_i/u_0 - c_i(T)/u_0T; \quad F_{i2} = 1 - c_i(T)/u_iT; \quad u_0$$

— предельное значение среднегодового экономического дохода (при внедрении некоторой идеальной рекомендации).

Максимальное значение критерия $F(X^*)$ или $\bar{F}(X^*)$ по формуле (9.68) или (9.70) находят при выполнении ряда условий или ограничений на переменные x_i .

Ограничение на численные значения переменных X отражает тот факт, что $x_i=1$, если 1-й вариант рекомендации выбирается для оценки, и $x_i = 0$, если не выбирается, а также то, что для оценки выбирается

наилучший вариант:

$$\sum_{i=1}^n x_i = 1. \quad (9.71)$$

При нахождении решения (9.68) и при использовании в качестве критерия абсолютного экономического дохода должно выполняться ограничение:

$$F_i x_i \geq F^{(u)}. \quad (9.72)$$

Наложим ограничение на критерий времени

$$F_i x_i \geq F^{(l)}, \quad (9.73)$$

где $F^{(l)}$ — минимально допустимое значение времени, в течение которого реализованная рекомендация дает прибыль.

При нахождении (9.70) должны выполняться условия:

$$\bar{F}_i^{(1)} x_i \geq \bar{F}^{(u)}; \quad (9.74)$$

$$\bar{F}_i^{(2)} x_i \geq \bar{F}^{(l)}. \quad (9.75)$$

Ограничение по максимально допустимой стоимости на формирование и реализацию рекомендации можно записать в следующем виде:

$$\sum_{i=1}^n c_i x_i \leq c_m.$$

Может быть наложено ограничение по максимально допустимой стоимости на формирование и реализацию рекомендации, по условию допустимой стоимости

$$c'_m < c_m.$$

Следовательно, в общем случае, допустимые затраты на формирование и реализацию рекомендации определяются из соотношения $c = \min(c_m, c'_m)$, и соответствующее ограничение запишутся в виде

$$\sum_{i=1}^n c_i x_i \leq c. \quad (9.76)$$

Кроме ограничения сверху на суммарную стоимость реализации рекомендации бывает необходимым введение ограничения по минимально возможной стоимости реализации рекомендации. Это ограничение отражает то обстоятельство, что для КП данного назначения существует минимальное значение стоимости, уменьшение которой приводит к принципиальной невозможности реализации рекомендации. Иными словами, существует тот минимум средств,

который позволяет решить задачу наиболее простой КП данного назначения. Указанное ограничение запишем в виде

$$\sum_{i=1}^n c_i x_i \leq \underline{c}, \quad (9.77)$$

где \underline{c} — минимально возможная стоимость реализации рекомендации

Объединяя неравенства (9.76) и (9.77), получим ограничение по суммарной стоимости реализации рекомендации

$$\sum_{i=1}^n c_i x_i \leq \bar{c}. \quad (9.78)$$

Таким образом, оценка целесообразности формирования и реализации рекомендаций по решению задач консультируемой проблемы производится:

- при выборе в качестве критерия абсолютного дохода от реализации рекомендаций — с помощью решения задач (9.68), (9.71), (9.72), (9.78);

- при выборе в качестве критерия времени, в течение которого реализованная рекомендация дает прибыль — путем решения задач (9.68), (9.71), (9.73), (9.78);

- при использовании двухкритериальной оценки — решением задач (9.70), (9.71), (9.74), (9.75), (9.78).

В этом случае, если решение существует, то формирование рекомендаций целесообразно; если же указанного решения не существует, то формирование рекомендаций нецелесообразно.

9.10. Особенности решения задач структурного синтеза

Задача структурного синтеза заключается в выборе принципа действия КП и в определении оптимальной структуры КП для реализации заданных функций. Структура КП определяется природой входящих в нее элементов и физической реализацией связей между ними в составе КП. Структуры наглядно изображают, как устроена КП — из каких частей она состоит и как эти части связаны друг с другом. Одной из форм математического описания структуры, как мы знаем, является граф. Напомним, что вершины графа отождествляются с элементами КП (на принятом уровне детализации), а дуги и ребра графа — со связями между соответствующими элементами. Форма отображения структуры — схема. Схема и граф тождественны по своему содержанию, но различаются по форме. В схемах для

обозначения элементов используют различные геометрические фигуры, разнообразие форм которых облегчает чтение схем.

Принцип, по которому объединение элементов приводит к появлению новых свойств, отличных от свойств элементов, называют *принципом организации*. **Организация** — понятие более высокого ранга, чем функция и структура. Различные принципы организации могут приводить к построению КП, различающихся своими структурами, но тождественным по своему функциональному назначению.

Сложность — свойство КП, заключающееся в том, что функция, реализуемая КП, не может быть представлена в виде композиции функций, реализуемых элементами КП. Например, при структурном синтезе ЭВМ рассматривается как система, состоящая из взаимосвязанных функциональных блоков и узлов, организованных таким образом, чтобы их функционирование приводило к реализации заданных функций — вычислениям на основе алгоритмов. Одни и те же функции могут быть реализованы различными структурами, обеспечивающими производительность решения задач при различных затратах оборудования. Закон функционирования ЭВМ невозможно рассмотреть только с точки зрения электрических процессов, происходящих в цепях ЭВМ. Функции ЭВМ выявляются лишь при рассмотрении процессов в ЭВМ в информационном и алгоритмическом аспектах. **Это объясняется эффектом организации, порождающим в совокупностях элементов новые свойства.**

Процесс формирования структуры сложной КП представляет собой многоэтапную процедуру, осуществляемую по блочно-иерархическому принципу:

Этап 1. Построение допустимого множества (каталога) принципиально возможных типов, входящих в КП элементов. Примерами могут служить: каталоги типовых рекомендаций в САК больших КП; каталог допустимых консультационных модулей в САК; каталог типовых консультационных проектов в САК и др. Каталоги входят в состав информационного обеспечения САК.

Этап 2. Выбор множества допустимых элементов, используемых при структурном синтезе. Поскольку различные элементы, входящие в каталог применяемых элементов, могут быть реализованы на основе разнообразных физических принципов и с помощью различных решений, не все элементы могут стыковаться друг с другом. Поэтому основной задачей данного этапа является определение требований к

параметрам и принципам функционирования отдельных элементов, входящих в состав КП.

Этап 3. Построение вариантов структур КП с учетом их функционирования. Здесь должны быть известны состав элементов, правила их соединения между собой и способ определения по структуре синтезируемого объекта функцию, которую он реализует.

При структурном синтезе небольшой сложности возможно построение полного множества допустимых структур (для реализации полного перебора вариантов). При этом в ЭВМ должны быть заложены правила генерации всех вариантов структур КП.

При синтезе сложных структур прямой перебор уже невозможен и необходима разработка процедур и алгоритмов направленного поиска оптимальной структуры КП. Эти процедуры обычно базируются на использовании методов математического программирования (в основном — дискретного программирования), последовательных и итерационных алгоритмов синтеза, сетевых и графовых моделей консультирования, а также методов теории эвристических решений и методов решений консультационных задач.

Этап 4. Оценка вариантов и выбор компромиссной структуры КП. Как правило, оценка варианта структуры требует формирования и анализа математической модели синтезированной структуры и выполнения параметрической оптимизации, так как для объективной оценки сравнивать варианты структуры имеет смысл при оптимальных значениях параметров. Эти процедуры сложны и громоздки, в связи с чем полный перебор вариантов при таком подходе практически неосуществим.

Для уменьшения сложности этого этапа целесообразно либо использовать косвенные критерии предпочтения вариантов, либо искать оценки варианта структуры без исследования громоздких математических моделей. При таком подходе вводят параметр, характеризующий качество КП. Это может быть число элементов в КП, ее стоимость, занимаемый объем, максимальное число элементов, находящихся в активном состоянии (мощность), вероятность нерешения проблемы, максимальная длина путей (в задачах размещения и трассировки) и т.д.

Другим путем уменьшения сложности решения задачи структурного синтеза является организация диалогового режима консультанта с ЭВМ на 3-м и 4-м этапах синтеза. При этом консультант сам решает, какие программы анализа и оптимизации будет использовать для оценки вариантов. Сокращение времени на формирование рекомендации в диалоговом режиме происходит за счет эвристических

способностей человека, за счет возможности прерывания построения заведомо бесперспективного варианта структуры и за счет поиска не оптимального, а допустимого варианта синтезируемой структуры.

Этап 5. Коррекции консультационного задания. Под коррекцией КЗ понимают изменения заданных ограничений в тех случаях, когда не существует вариантов рекомендаций, обеспечивающих требуемое ограничение. Если коррекция произведена, то соответствующие этапы консультирования повторяют при новых значениях ограничений.

Объем решаемых на каждом этапе задач настолько велик, что проведение исследований в полном объеме невозможно без средств автоматизации консультирования. При этом на различных этапах к консультациям привлекаются различные специализированные коллективы научных работников и инженеров. Следует отметить, что при использовании отдельных элементов, устройств и подпроблем, входящих в состав сложной КП, следует проводить их анализ и синтез на различных уровнях. В процессе консультирования приходится учитывать существующие многократные перекрестные связи между элементами, что существенно усложняет задачу структурного синтеза.

Исходя из вышесказанного следует, что особенность задач структурного синтеза заключается в том, что для получения оптимального варианта структуры КП необходимо наличие ее математической модели, представляющей собой формальное описание множества структур КП на принятом уровне детализации. В этом случае задача структурного синтеза сводится к выбору компромиссного варианта в счетном множестве.

9.11. Методы выбора структуры

Методы выбора структуры КП рассмотрим на примере системы автоматизированного консультирования (САК)

Структура САК характеризуется прежде всего числом уровней функционирования, наличием в составе САК различных модулей САК, числом и типами комплексов технических средств различного функционального назначения, схемой организации потоков консультационных требований (запросов) между элементами и модулями САК и т. д. Смысл критериев выбора (9.65), (9.70) в том, что возможные их значения определяют возможные варианты САК с присущей им структурой. При таком представлении критериев выбора будем говорить о вариантном методе выбора структуры САК. Если же задан вектор параметров выбора

$$X = (x_1, \dots, x_k), \quad (9.79)$$

где компоненты x_k ($k = 1, \dots, \bar{k}$) означают число комплектов технических средств каждого k -го типа, входящих в САК, то возможные значения вектора X характеризуют различные структуры САК.

Изменение вектора (9.79) в общем случае приводит к изменению как номенклатуры, так и количества средств САК.

Метод решения задачи выбора структуры САК в терминах переменных вектора (9.79) назовем **методом вариации состава средств САК**. Задачи выбора структуры САК в методологическом плане аналогичны задачам оценки целесообразности ее разработки по полной совокупности вариантов САК. Различие состоит лишь в том, что на этапе выбора структуры САК учитывается существенно больше ограничений, более детально и глубже отражающих возможность реализации той или иной структуры, того или иного варианта САК. Ниже изложены два метода выбора рациональных структур САК: **вариантный и метод вариации состава средств САК**.

Вариантный метод выбора структур САК изложим для случая использования в качестве критерия выбора абсолютного дохода от внедрения САК за весь срок ее функционирования. Показатели экономической эффективности (9.68), (9.70) запишем:

для стохастического принципа

$$F(X) = T \sum_{i=1}^n \sum_{y \in G} u_i(y) P(y) - \sum_{i=1}^n c_i x_i; \quad (9.80)$$

для детерминированного принципа

$$F(X) = T \sum_{i=1}^n u_i x_i - \sum_{i=1}^n c_i x_i, \quad (9.81)$$

где $u_i(y)$ и u_i — среднегодовой экономической выигрыш за счет внедрения i -го варианта САК соответственно при случайном векторе состояний внешней среды и при осредненном его значении; $P(y)$ — плотность вероятности; c_i — суммарная стоимость разработки и внедрения i -го варианта САК.

Переменные x_i представленных выше показателей экономической эффективности должны быть подчинены ряду условий (ограничений), учитываемых при решении задачи. Правильная формализация и учет достаточно большого числа противоречивых ограничений является важной основой решения задач выбора рационального варианта структуры САК. Следует учитывать, что система ограничений зависит также от выбора показателя экономической эффективности. В качестве показателя экономической эффективности может быть принят один из

приведенных в предыдущем пункте показателей, тогда некоторые из оставшихся функций могут быть переведены в разряд ограничений. Например, можно, взяв за показатель среднегодовой экономической выигрыш, стремиться его максимизировать при заданных ограничениях сверху на различные виды стоимости или, наоборот, выбрав стоимость, стремиться минимизировать затраты на САК при ограничениях снизу на среднегодовой экономической выигрыш от внедрения САК. Выше было определено, что за показатель экономической эффективности принимается абсолютный экономический доход от внедрения САК.

Сформулируем основные ограничения на переменные показатели (9.80) или (9.81) экономической эффективности внедрения САК применительно к вариантному методу выбора:

- по численным значениям переменных

$$\sum_{i=1}^n x_i = 1; \quad x_i = 0 \quad \text{или} \quad 1; \quad (9.82)$$

- по стоимости капиталовложений на создание и внедрение САК

$$\sum_{i=1}^n C_{0i} x_i \leq C_0, \quad (9.83)$$

где C_{0i} — капиталовложения на создание i -го варианта САК; C_0 — допустимые капиталовложения при внедрении САК.

Ограничения по различным составляющим капиталовложений:

- по стоимости технологической части САК

$$\sum_{i=1}^n \bar{C}_{0i} x_i \leq \bar{C}_0 \quad (9.84)$$

(\bar{C}_{0i} — стоимость технологической части i -го варианта САК; \bar{C}_0 — допустимые затраты на технологическую часть САК);

- по стоимости капитального строительства

$$\sum_{i=1}^n \tilde{C}_{0i} x_i \leq \tilde{C}_0 \quad (9.85)$$

(\tilde{C}_{0i} — стоимость строительства сооружений i -го варианта САК;

\tilde{C}_0 — допустимые затраты на капитальное строительство).

Ограничение по стоимости эксплуатационных расходов при внедрении САК можно определить из следующего выражения:

$$\sum_{i=1}^n C_{э_i} x_i \leq C_э, \quad (9.86)$$

где $C_{э_i}$ — эксплуатационные затраты при внедрении i -го варианта САК; $C_э$ — допустимые эксплуатационные затраты.

Ограничение по суммарной стоимости внедрения САК определяется как

$$\sum_{i=1}^n C_i x_i \leq C, \quad (9.87)$$

где C — допустимые суммарные затраты на разработку и внедрение САК.

Ограничение по абсолютной окупаемости САК. Отношение суммарной стоимости системы к среднегодовому доходу за счет внедрения САК представляет собой период (срок) абсолютной окупаемости внедряемых САК. Очевидно, для того, чтобы внедрение САК было экономически целесообразно, должны выполняться условия:

$$\sum_{i=1}^n [C_i x_i / \sum_{y \in G} u_i(y) p(y)] \leq T^{(A)} \quad (9.88)$$

или

$$\sum_{i=1}^n (C_i x_i / u_i) \leq T^{(A)}, \quad (9.89)$$

где $T^{(A)}$ — допустимый срок абсолютной окупаемости САК.

Ограничение по окупаемости САК. Часто наряду с ограничением по абсолютной окупаемости используют ограничение по окупаемости. Указанное ограничение выражается в виде:

$$\sum_{i=1}^n [C_{0_i} x_i / \sum_{y \in G} u_i(y) p(y)] \leq T^{(0)} \quad (9.90)$$

или

$$\sum_{i=1}^n (C_{0_i} x_i / u_i) \leq T^{(0)}, \quad (9.91)$$

где $T^{(0)}$ — допустимый срок окупаемости САК.

Ограничение общей численности личного состава, занятого эксплуатацией САК. Это условие записывается в форме следующего неравенства:

$$\sum_{i=1}^n N_i x_i \leq N, \quad (9.92)$$

где N_i — численность личного состава, занятого обслуживанием i -го варианта САК; N — допустимая общая численность личного состава на эксплуатацию САК.

Ограничения по числу специалистов того или иного профиля, занятых эксплуатацией САК. Пусть $\bar{\mu}$ — общее число профилей специалистов, необходимых для привлечения к эксплуатации САК; N_{μ_i} — число специалистов μ -го профиля ($\mu = 1, \dots, \bar{\mu}$), эксплуатирующих i -й вариант САК; N_{μ} — допустимое число специалистов μ -го профиля, тогда

$$\sum_{i=1}^n N_{\mu_i} x_i \leq N_{\mu}, \quad \mu = 1, \dots, \bar{\mu}. \quad (9.93)$$

Ограничения производственных возможностей промышленных предприятий по поставке серийных средств автоматизации. Пусть для внедрения i -го варианта САК необходимо осуществить в определенные сроки поставки m_{ω_i} серийных средств автоматизации (комплектующих изделий) ω -го типа, тогда

$$\sum_{i=1}^n m_{\omega_i} x_i \leq m_{\omega}, \quad \omega = 1, \dots, \bar{\omega}, \quad (9.94)$$

где m_{ω} — число серийных средств автоматизации ω -го типа, которое может выпустить промышленность за требуемый период поставки этих средств для внедрения САК с учетом уже имеющихся заказов на поставку аналогичных изделий по другим работам; величина m_{ω} характеризует свободные производственные мощности промышленных предприятий по выпуску изделий данного вида.

Ограничения по возможности разрабатывающих организаций.

Если для внедрения i -го варианта САК требуется разработка некоторых средств (изделий) автоматизации заново, то необходимо учитывать возможности таких разработок. По аналогии с предыдущим указанное ограничение запишется

$$\sum_{i=1}^n \alpha_{\psi_i} x_i \leq \alpha_{\psi}, \quad \psi = 1, \dots, \bar{\psi}, \quad (9.95)$$

где α_{ψ_i} — число специалистов (или производственных мощностей) ψ -го вида, которых необходимо привлечь для разработки новых изделий в определенные сроки при создании i -го варианта САК; α_{ψ} — число свободных специалистов (свободные производственные мощности) разрабатывающих организаций.

Ограничения по срокам внедрения САК определяются как

$$\sum_{i=1}^n \bar{t}_i x_i \leq T_0, \quad (9.96)$$

где \bar{t}_i — срок создания и внедрения i -го варианта САК; T_0 — максимально допустимый срок внедрения и ввода в эксплуатацию САК.

Изложенные задачи выбора оптимального (исходя из принятого показателя экономической эффективности) варианта САК заключаются в нахождении такого решения системы линейных неравенств (9.82)—(9.88), (9.90)—(9.96), которое обращало бы в максимум линейную функцию (9.80), либо в нахождении решения системы линейных неравенств (9.82)—(9.87), (9.89), (9.91)—(9.96), обращаемого в максимум линейную функцию (9.81).

Рассмотрим кратко **метод решения** сформулированной задачи для более общего случая, когда требуется определить максимум функции (9.80).

Рассмотрим произвольный i -й шаг. К началу этого шага известны векторы

$$\|x_i^{(t-1)}\| = 0 \quad \text{и} \quad \|r_i^{(t-1)}\|.$$

Для первого шага

$$\|r_i^{(0)}\| = \|T \sum_{y \in G} u_i(y) p(y) - C_i\|;$$

находим

$$r_i^{(r-1)} = \max_i r_i^{(r-1)}$$

Затем проверяем неравенства:

$$\begin{aligned}
 C_{0,t} \leq C_0, \quad \bar{C}_{0,t} \leq \bar{C}_0; \quad \hat{C}_{0,t} \leq \hat{C}_0, \quad C_{\mu,t} \leq C_{\mu}, \quad C_{t,t} \leq C; \\
 \frac{C_{t,t}}{\sum_{y \in G} u_{t,t}(y) \rho(y)} \leq T^{(A)}, \quad \frac{C_{0,t,t}}{\sum_{y \in G} u_{t,t}(y) \rho(y)} \leq T^{(0)}; \\
 N_{t,t} \leq N; \quad N_{\mu,t} \leq N_{\mu}, \quad \mu = 1, \dots, \bar{\mu}; \\
 m_{\omega,t} \leq m_{\omega}; \quad \omega = 1, \dots, \bar{\omega}; \quad \alpha_{\psi,t} \leq \alpha_{\psi}; \quad \psi = 1, \dots, \bar{\psi}; \quad \bar{t}_t \leq T_0.
 \end{aligned}$$

Если все неравенства удовлетворяются, то полагаем $x_{i_t} = 1$, $r_{i_t}^{(t)} = r_{i_t}^{(t-1)}$ и процесс нахождения максимума функции (9.80) заканчивается. Если хотя бы одно из приведенных выше неравенств нарушено, то полагаем x_{i_t} и $r_{i_t}^{(t)} = 0$, после чего переходим к следующему шагу, повторяя изложенную процедуру. Процесс поиска заканчивается через число шагов $t \leq m$.

Рассмотрим принцип выбора структурных схем САК *методом вариации состава средств* САК. С учетом смыслового содержания введенного выше вектора (9.79) среднегодовой экономической доход от внедрения САК приближенно может быть представлен в виде

$$u(X) = u_0 [1 - \exp(-\sum_k \beta_k x_k)], \quad (9.97)$$

где u_0 — предельный среднегодовой экономический доход при внедрении наиболее совершенной (идеальной) САК; β_k — коэффициент, учитывающий прирост среднегодового экономического дохода при введении в состав САК одного КТС i -го типа.

Из приведенной зависимости видно, что искомая величина u зависит от переменных x_k , значения которых и подлежат определению. Выражение (9.97) можно записать в виде

$$u = u_0 \{1 - \exp[-\sum_{\delta \neq k} \rho_{\delta} x_{\delta}] \exp(\beta_k x_k)\}$$

или

$$u = u_0 [-A(x) \exp(-\beta_k x_k)],$$

где

$$A(X) = \exp(-\sum_{\delta \neq k} \beta_{\delta} x_{\delta}).$$

При $x_k = 1$ имеем

$$\bar{u} = u_0 [1 - A(X) \exp(-\beta_k x_k)],$$

откуда

$$\beta_k = -\ln [(u_0 - u)/A(X) u_0].$$

Полученное выражение показывает, что, строго говоря, коэффициент $\beta_k = \beta_k(X)$. Однако с достаточной для практики точностью можно полагать β_k некоторым средним значением по множеству X или рассчитывать β_k для одного из вариантов структурной схемы САК. Вычисления показывают, что зависимость β_k от X не является резко выраженной, поэтому изложенный выше прием оправдан.

Выражение для β_k можно преобразовать следующим образом. При

$$x_k = 0 \text{ имеем } u = u_0 [1 - A(X)], \text{ откуда } A(X) = (u_0 - u)/u_0.$$

Подставляя значение $A(X)$ в формулу для β_k , получаем окончательное выражение для β_k :

$$\beta_k = -\ln [(u_0 - \bar{u})/(u_0 - \bar{u})], \quad (9.98)$$

где \bar{u} и \bar{u}' рассчитываются для некоторого фиксированного варианта структурной схемы САК.

Точность приближения β_k тем выше, чем ближе этот фиксированный вариант САК к оптимальному в смысле принятого показателя экономической эффективности.

Применительно к рассматриваемому методу выбора структуры САК показатели экономической эффективности (9.80) и (9.81) записываются в форме

$$F(x) = T [1 - \exp(-\sum_k \beta_k x_k)] \sum_{y \in G} u_0(y) p(y) - \sum_k C_k x_k \quad (9.99)$$

и

$$F(X) = u_0 T \left[1 - \exp\left(-\sum_k \beta_k x_k\right) \right] - \sum_{k=1}^{\bar{k}} C_k x_k. \quad (9.100)$$

Рассмотрим ограничения на переменные показатели экономической эффективности (9.99), (9.100).

Ограничения по численным значениям переменных. Пусть I — множество КТС ξ -го функционального назначения ($\xi = 1, \dots, \bar{\xi}$):

$$\sum_{k \in J_\xi} x_k \leq M_\xi, \quad \xi = 1, \dots, \bar{\xi}, \quad (9.101)$$

где M_ξ — максимальное число КТС ξ -функционального назначения, которые могут входить в САК; $x_k = 0, 1, 2, \dots$ — целочисленные значения.

Ограничения по стоимости капитальных вложений на создание САК

$$\sum_{k=1}^{\bar{k}} C_{0k} x_k \leq C_0. \quad (9.102)$$

Ограничения по различным составляющим капиталовложений:

- по стоимости технологической части САК

$$\sum_{k=1}^{\bar{k}} \bar{C}_{0k} x_k \leq \bar{C}_0 \quad (9.103)$$

(\bar{C}_{0k} — стоимость k -го элемента САК; \bar{C}_0 — допустимые затраты на технологическую часть САК);

- по стоимости строительства

$$\sum_{k=1}^{\bar{k}} \hat{C}_{0k} x_k \leq \hat{C}_0 \quad (9.104)$$

(\hat{C}_{0k} — стоимость строительства сооружений для k -го элемента САК;

\hat{C}_0 — допустимые затраты на строительство).

Ограничения по стоимости эксплуатационных расходов

$$\sum_{k=1}^{\bar{k}} C_{эk} x_k \leq C_{э} \quad (9.105)$$

где $C_{эk}$ — стоимость эксплуатационных расходов k -го элемента САК;

$C_{э}$ — допустимая стоимость эксплуатационных расходов при внедрении САК.

Ограничение по суммарной стоимости САК

$$\sum_{k=1}^{\bar{k}} C_k x_k \leq C, \quad (9.106)$$

где C — допускаемая суммарная стоимость САК.

Ограничения по абсолютной окупаемости САК

$$\frac{\sum_{k=1}^{\bar{k}} C_k x_k}{[1 - \exp(-\sum_k \beta_k x_k)] \sum_{y \in G} u_0(y) p(y)} \leq T^{(A)} \quad (9.107)$$

или

$$\frac{\sum_{k=1}^{\bar{k}} C_k x_k}{u_0 [1 - \exp(-\sum_k \beta_k x_k)]} \leq T^{(A)}. \quad (9.108)$$

Ограничения по окупаемости САК

$$\frac{\sum_k^{\bar{k}} C_{0k} x_k}{[1 - \exp(-\sum_k \beta_k x_k)] \sum_{y \in G} u_0(y) p(y)} \leq T^{(0)} \quad (9.109)$$

или

$$\frac{\sum_{k=1}^{\bar{k}} C_{0k} x_k}{u_0 [1 - \exp(-\sum_k \beta_k x_k)]} \leq T^{(0)}. \quad (9.110)$$

Ограничения по общей численности эксплуатационного состава САК

$$\sum_{k=1}^{\bar{k}} N_k x_k \leq N, \quad (9.111)$$

где N_k — число людей, занятых эксплуатацией k -го элемента САК; N — допустимая общая численность эксплуатационного состава САК.

Ограничение по числу специалистов различного профиля, занятых эксплуатацией САК:

$$\sum_{k=1}^{\bar{k}} N_{\mu k} x_k \leq N_{\mu}, \quad \mu = 1, \dots, \bar{\mu}, \quad (9.112)$$

где $N_{\mu k}$ — число специалистов μ -го профиля, занятых эксплуатацией k -го элемента САК; N_{μ} — допустимое число специалистов μ -го профиля.

Ограничения по производственным возможностям промышленных предприятий по поставке серийных технических средств:

$$\sum_{k=1}^{\bar{k}} m_{\omega k} x_k \leq m_{\omega}, \quad \omega = 1, \dots, \bar{\omega}, \quad (9.113)$$

где $m_{\omega k}$ — число серийных технических средств (комплектующих изделий) ω -го типа для k -го элемента САК; m_{ω} — число серийных технических средств ω -го типа, которое может выпустить промышленность за требуемый период поставки этих средств для внедрения САК с учетом уже имеющихся заказов на поставку аналогичных изделий по другим работам.

Ограничения по трудовым ресурсам разрабатывающих организаций.

$$\sum_{p=1}^{\bar{p}} \alpha_{\psi p} l_k \left(\sum_{k \in K_p} \min \{x_k, 1\} \right) + \sum_{k \in K_0} \alpha_{\psi k} \min \{x_k, 1\} \leq \alpha_{\psi}, \quad \psi = 1, \dots, \bar{\psi}, \quad (9.114)$$

где K_p ($p = 1, \dots, \bar{p}$) — подмножества типов КТС, на которые распространяются требования по унификации; K_0 — подмножество типов КТС, на которые не распространяется требование унификации; $l_k(x_k)$ — коэффициент, характеризующий уменьшение трудоемкостей при унификации разработки КТС и представляющей собой слабо возрастающую функцию от суммы

$$\sum_{k \in K_p} \min \{x_k, 1\}$$

в области

$$\sum_{k \in K_p} \min \{x_k, 1\} > 1$$

и $l_k(0), l_k(1) = 1$; $\alpha_{\psi k}$ — число специалистов ψ -го профиля, которых необходимо привлечь для разработки или доработки k -го элемента САК,

$$\alpha_{\psi p} = \max_{k \in K_p} \{\alpha_{\psi k}\};$$

α_{ψ} — число свободных специалистов ψ -го профиля в разрабатывающих организациях.

Ограничение по срокам разработки и внедрения САК. Для рассматриваемого метода выбора структуры САК это ограничение зависит от способов организации внедрения, принятого плана работ по внедрению и в общем случае выражается нелинейными функциями от переменных. Запишем это ограничение для самого простого случая, когда внедрение САК осуществляется последовательно:

$$\sum_{k=1}^k \bar{t}_k x_k \leq T_0, \quad (9.115)$$

где \bar{t}_k — время, необходимое для внедрения комплекса (устройства) k -го вида.

Ограничения, связанные с выполнением требований по унификации КТС. При выборе структур комбинированным методом требования унификации учитываются на уровне формирования вариантов САК. В излагаемом здесь методе выбора структур рассматриваются множества

I_{ξ_i} ($\xi = 1, \dots, \bar{\xi}$) КТС, которые различаются как типом выполнения, так и функциональным назначением, так что в ходе решения задачи в структуру САК могут быть включены КТС различных типов одного функционального назначения. В ряде случаев это может быть неприемлемым, и поэтому требуется ввести специальное условие, исключающее применение разнотипных средств.

Пусть K_p ($p = 1, \dots, \bar{p}$) — подмножества типов КТС, на которые распространяются требования по унификации. На каждое подмножество K_p типов КТС распространяется унификация p -го вида. Унификация p -го вида означает, что в K_p -группе КТС к реализации может быть принят только один тип КТС. Указанные ограничения записываются в виде равенств нулю попарных произведений x' и x'' для любых пар $(k', k'') \in K_p$, т. е.

$$x_{k'} \cdot x_{k''} = 0; \quad k' < k''; \quad (k', k'') \in K_p; \quad p = 1, \dots, \bar{p}. \quad (9.116)$$

Таким образом, сформулированная задача выбора варианта структурной схемы САК состоит в нахождении максимального значения функционала (9.99) при ограничениях (9.101)—(9.107), (9.109), (9.111), (9.115), (9.116) на переменные задачи. Решение такой экстремальной задачи обычно проводится известными численными методами. При усреднении случайного вектора y , т. е. при детерминированном подходе задача сводится к нахождению максимума функционала (9.100) при ограничениях (9.101)—(9.106), (9.110)—(9.116). Функционалы (9.99), (9.100) являются выпуклыми от переменных, а ограничения (9.101)—(9.115), за исключением ограничений (9.107)—(9.111), (9.116), представляют собой линейные неравенства.

Сформулированная задача приближенно может быть решена так называемым *методом максимального элемента*. Изложим суть этого метода применительно к максимизации функционала (9.100).

Отыскание максимума производится за ряд последовательных шагов. Рассмотрим произвольный t -й шаг. К началу t -го шага известны векторы

$$\|x_k^{(t-1)}\|, \quad \|r_k^{(t-1)}\| = \left\| \frac{\partial F(X)^{(t-1)}}{\partial X_k} \right\|, \quad \|M_{\xi}^{(t-1)}\|, \\ \|N_{\mu}^{(t-1)}\|, \quad \|m_{\omega}^{(t-1)}\|, \quad \|\alpha_{\psi}^{(t-1)}\|$$

и величины

$$C_0^{(t-1)}, \quad \bar{C}^{(t-1)}, \quad \widehat{C}_0^{(t-1)}, \quad C_9^{(t-1)}, \quad C^{(t-1)}, \\ T_0^{(A, t-1)}, \quad T^{(0, t-1)}, \quad N^{(t-1)}, \quad T_0^{(t-1)}.$$

Для 1-го шага

$$\begin{aligned} x_k^{(0)} &= 0; \quad r_k^{(0)} = u_0 T \beta_k - C_k; \quad M_{\xi}^{(0)} = M_{\xi}; \\ C_0^{(0)} &= C_0; \quad \bar{C}_0^{(0)} = \bar{C}_0; \quad \widehat{C}_0^{(0)} = \widehat{C}_0; \quad C_9^{(0)} = C_9; \quad C^{(0)} = C; \\ T^{(A, 0)} &= T^{(A)}; \quad T^{(0, 0)} = T^{(0)}; \quad N^{(0)} = N; \quad N_{\mu}^{(0)} = N_{\mu}; \\ m_{\omega}^{(0)} &= m_{\omega}; \quad \alpha_{\varphi}^{(0)} = \alpha_{\varphi}; \quad T_0^{(0)} = T_0. \end{aligned}$$

Находим

$$r_k^t = \max_k r_k^{(t-1)}.$$

Определяем новые значения переменных

$$x_k^{(t)} = \begin{cases} X_k^{(t-1)} + 1 & \text{при } k = k_t \notin K_p, \text{ или } k = k_t \in K_p, \\ & K_p \cap \{k_1, \dots, k_{t-1}\} \neq 0, k_t \in \{k_1, \dots, k_{t-1}\}, \text{ или} \\ k = k_t \in K_p, & K_p \cap \{k_1, \dots, k_{t-1}\} = 0; \\ X_k^{(t-1)} & \text{при } k \neq k_t \text{ или } k = k_t \in K_p, \\ & K_p \cap \{k_1, \dots, k_{t-1}\} \neq 0, k_t \in \{k_1, \dots, k_{t-1}\}. \end{cases}$$

Если

$$k = k_t \in K_p, \quad K_p \cap \{k_1, \dots, k_{t-1}\} \neq 0,$$

$k_t \notin \{k_1, \dots, k_{t-1}\}$ то полагаем $r_k^{(t-1)}$ -й элемент вектора $\|r_k^{(t-1)}\|$ равным нулю, получая таким образом вектор $\|r_k^{(t)}\|$, и переходим к следующему шагу. Затем определяем

$$r_k^{(t)} = \begin{cases} r_k^{(t-1)} e^{\beta_k} - C_k (1 - e^{-\beta_k}) & \text{при } k = k_t; \\ r_k^{(t-1)} & \text{при } k \neq k_t; \end{cases}$$

$$M_{\xi}^{(t)} = \begin{cases} M_{\xi}^{(t-1)} - 1 & \text{при } k_t \in J_{\xi}; \\ M_{\xi}^{(t-1)} & \text{при } k_t \notin J_{\xi}; \end{cases}$$

$$C_0^{(t)} = C_0^{(t-1)} - C_{0k_t},$$

$$\bar{C}_0^{(t)} = \bar{C}_0^{(t-1)} - \bar{C}_{0k_t},$$

$$\widehat{C}_0^{(t)} = \widehat{C}_0^{(t-1)} - \widehat{C}_{0k_t}, \quad C_9^{(t)} = C_9^{(t-1)} - C_{9k_t},$$

$$C^{(t)} = C^{(t-1)} - C_{k_t}, \quad T_0^{(t)} = T_0^{(t-1)} - t_{k_t};$$

$T^{(A,t)} = T^{(A)}$ при выполнении соотношения (9.104) и $T^{(A,t)} = 0$ в противном случае; $T^{(0,t)} = T^{(0)}$ при выполнении соотношения (9.106) и $T^{(0,t)} = 0$ в противном случае.

Процесс поиска максимального значения $F(x)$ заканчивается, если вектор $\|r_k^{(t)}\| = 0$. Найденные значения вектора X^* представляют собой оптимальный с точки зрения принятого критерия выбора варианта структурной схемы САК.

Рассмотрим пример решения задачи выбора структуры **методом вариации состава технических средств САК**. Предположим, что нужно определить рациональную структуру подсистемы САК при следующих исходных данных.

В составе подсистемы САК рассматриваются КТС трех функциональных назначений: КТС-1, $\xi = 1$; КТС-2, $\xi = 2$; КТС-3, $\xi = 3$. Выбор рациональной структуры осуществляется на базе восьми типов КТС ($\bar{k} = 8$), при этом подмножество КТС-1 включает $J_1 = \{k = 1, 2, 3\}$; КТС-2 $J_2 = \{k = 4, 5, 6\}$; КТС-3 $J_3 = \{k = 7, 8\}$. Стоимость КТС характеризуется вектором $C = \|C_k\| = \|5, 8, 15, 5, 8, 10, 2, 1,5\|$ (C_k измеряется в миллионах рублей). Ограничение по суммарной стоимости $C = 60$ млн. руб. Коэффициенты β_k заданы вектором $\beta = \|\beta_k\| = \|0,1, 0,2, 0,3, 0,2, 0,5, 0,1, 0,05\|$. Численность состава обслуживающего персонала КТС каждого типа $\|N_k\| = \|20, 22, 25, 20, 22, 22, 2, 2\|$, а ограничение по общей численности обслуживающего персонала $N = 115$ чел. Число специалистов высшей квалификации для эксплуатации КТС каждого типа ($\mu = 1$) $\|N_{1k}\| = \|7, 10, 10, 7, 8, 8, 1, 1\|$, а ограничение по общему числу специалистов высшей квалификации $N_1 = 43$ чел. Число специалистов средней квалификации для обслуживания КТС каждого типа ($\mu = 2$) $\|N_{2k}\| = \|13, 12, 15, 13, 14, 14, 1, 1\|$, а ограничение по общему числу специалистов средней квалификации $N_2 = 77$ чел. Предельный среднегодовой экономический доход от внедрения подсистемы САК $u_0 = 15$ млн. руб., срок службы $T = 10$ лет, допустимый период абсолютной окупаемости $T^A = 4,1$ года.

Максимальное число КТС первого функционального назначения $M_1 = 1$, второго — $M_2 = 3$, третьего — $M_3 = 9$. Унификация распространяется на два подмножества КТС: $k_1 = \{k = 4, 5, 6\}$, $k_2 = \{k = 7, 8\}$.

Математически сформулированная задача запишется в следующем виде.

Критерии выбора:

$$F(X) = 150[1 - \exp\{- (0,1x_1 + 0,2x_2 + 0,3x_3 + 0,2x_4 + 0,5x_5 + 0,5x_6 + 0,1x_7 + 0,05x_8)\}] - (5x_1 + 8x_2 + 15x_3 + 5x_4 + 8x_5 + 10x_6 + 2x_7 + 1,5x_8).$$

Ограничения:

$$\sum_{k \in J_1} X_k = x_1 + x_2 + x_3 \leq 1 \quad (\xi = 1, J_1 = \{1, 2, 3\});$$

$$\sum_{k \in J_2} X_k = x_4 + x_5 + x_6 \leq 3 \quad (\xi = 2, J_2 = \{4, 5, 6\});$$

$$\sum_{k \in J_3} X_k = x_7 + x_8 \leq 9 \quad (\xi = 3, J_3 = \{7, 8\});$$

$$x_4x_5 = 0, \quad x_4x_6 = 0,$$

$$x_5x_6 = 0 \quad (p = 1, k_1 = \{4, 5, 6\});$$

$$x_7x_8 = 0 \quad (p = 2, k_2 = \{7, 8\});$$

$$\sum_{k=1}^8 C_k X_k = 5x_1 + 8x_2 + 15x_3 + 5x_4 + 8x_5 + 10x_6 + 2x_7 + 1,5x_8 \leq 60;$$

$$[5x_1 + 8x_2 + 15x_3 + 5x_4 + 8x_5 + 10x_6 + 2x_7 + 1,5x_8] \cdot 15^{-1} \times \\ \times [1 - \exp\{- (0,1x_1 + 0,2x_2 + 0,3x_3 + 0,2x_4 + 0,3x_5 + 0,5x_6 + 0,1x_7 + 0,05x_8)\}]^{-1} \leq 41;$$

$$\sum_{k=1}^8 N_k X_k = 20x_1 + 22x_2 + 25x_3 + 20x_4 + 22x_5 + 22x_6 + \\ + 2x_7 + 2x_8 \leq 115;$$

$$\sum_{k=1}^8 N_{1k} X_k = 7x_1 + 10x_2 + 10x_3 + 7x_4 + 8x_5 + 8x_6 + \\ + x_7 + x_8 \leq 43 \quad (\mu = 1);$$

$$\sum_{k=1}^8 N_{2k} X_k = 13x_1 + 12x_2 + 15x_3 + 13x_4 + 14x_5 + \\ + 14x_6 + x_7 + x_8 \leq 77.$$

Первый шаг:

$$\begin{aligned} & \|r_k^{(0)}\| = \|10; 22; 30; 25; 37; 65; 13; 6\|; \\ M_1^{(0)} &= 1; M_2^{(0)} = 3; M_3^{(0)} = 9; C^{(0)} = 60; T^{(A, 0)} = 0; N^{(0)} = 115; \\ N_1^{(0)} &= 43; N_2^{(0)} = 77; r_{k_1}^{(0)} = r_6^{(0)} = \max_k r_k^{(0)}; k_1 = 6; \\ x_k^{(1)} &= \begin{cases} 0 + 1 & \text{для } k = k_1 = 6 \in k_1, k_1 \cap \{0\} = 0; \\ 0 & \text{для } k \neq 6; \end{cases} \\ X^{(1)} &= \|0; 0; 0; 0; 0; 1; 0; 0\|; \\ \tilde{r}_k^{(1)} &= \begin{cases} 67e^{-0,5} - 4 = 35 & \text{при } k = 6; \\ r_k^{(0)} & \text{при } k \neq 6; \end{cases} \\ \|\tilde{r}_k^{(1)}\| &= \|10; 22; 30; 25; 37; 35; 13; 6\|; \\ M_\xi^{(1)} &= \begin{cases} M_\xi^{(0)} - 1 = 3 - 1 = 2 & \text{при } \xi = 1; \\ M_\xi^{(0)} & \text{при } \xi \neq 1; \end{cases} \\ \|M_\xi^{(1)}\| &= \|1; 2; 9\|; C^{(1)} = C^{(0)} - C_{k_1} = 60 - 10 = 50; \\ \frac{\sum_k C_k X_k}{\mu_0 \left[1 - \exp\left(-\sum_k \beta_k x_k\right) \right]} &= \frac{10}{15(1-0,5)} = 1,67 \leq 4,1; \\ N^{(1)} &= N^{(0)} - N_{k_1} = N^{(0)} - N_6 = 115 - 22 = 93; \\ N_1^{(1)} &= N^{(0)} - N_{1k_1} = 43 - 8 = 35; \\ N_2^{(1)} &= N_2^{(0)} - N_{2k_1} = 77 - 14 = 63. \end{aligned}$$

Второй шаг:

$$\begin{aligned} r_{k_2}^{(1)} &= r_5^{(1)} = \max_k r_k^{(1)}, k = 5; \\ x_k^{(2)} &= X_k^{(1)}, \text{ так как } k_2 \in k_1, k_1 \cap \{k_1\} \neq 0, k_2 \notin \{k_1\}; \\ \|r_k^{(2)}\| &= \|10; 22; 30; 25; 0; 35; 13; 6\|. \end{aligned}$$

Третий шаг:

$$\begin{aligned}
 r_{k_3}^{(2)} &= r_6^{(2)} = \max r_k^{(2)}, \quad k_3 = 6; \\
 x_k^{(3)} &= \begin{cases} x_k^{(2)} + 1 = 2 \text{ где } k = k_3 = 6, \quad k_3 \in k_1, \\ k_1 \cap \{k_1, k_2\} \neq, \quad k_3 \in \{k_1, k_2\}; \\ x_k^{(2)} \text{ для } k \neq 6; \end{cases} \\
 x^{(3)} &= \|0; 0; 0; 0; 0; 0; 2; 0; 0\|; \\
 \tilde{r}_k^{(3)} &= \begin{cases} 35e^{-0,5} - 15 = 2 - 1 - 4 \text{ для } k = k_3 = 6; \\ r_k^{(2)} \text{ для } k \neq 6; \end{cases} \\
 \|\tilde{r}^{(3)}\| &= \|10; 22; 30; 25; 0; 17; 13; 6\|; \\
 M_\xi^{(3)} &= \begin{cases} M_\xi^{(2)} - 1 = 2 - 1 = 1 \text{ при } \xi = 1; \\ M_\xi^{(2)} \text{ при } \xi \neq 1; \end{cases} \\
 \|M_\xi^{(3)}\| &= \|1; 1; 9\|; \quad C^{(3)} = C^{(2)} - C_{k_3} = 50 - 10 = 40; \\
 20/15 (1 - 0,45) &= 2,42 \leq 4,1; \quad T^{(A, 3)} = 4,1; \\
 N^{(3)} &= N^{(2)} - N_{k_3} = 93 - 22 = 71; \quad N_1^{(3)} = N_1^{(2)} - N_{1k_3} = 35 - 8 = \\
 &= 27; \\
 N_2^{(3)} &= N_2^{(2)*} - N_{2k_3} = 63 - 14 = 49; \\
 \|r_k^{(3)}\| &= \|10; 22; 3; 25; 0; 17; 13; 6\|.
 \end{aligned}$$

На 12-м шаге процесс заканчивается, так как наступает ограничение по абсолютной окупаемости подсистемы САК. Получаем решение $x^{(12)} = x^* = \|0, 0, 1, 0, 0, 3, 6, 0\|$, из которого следует, что структура подсистемы САК включает в себя: один КТС, используемый в САК-1 третьего типа ($k = 6$); шесть КТС, используемых в САК-3 седьмого типа ($k = 7$).

В заключение отметим, что, в отличие от комбинированного метода, где все особенности структуры органически содержатся в каждом оцениваемом варианте, в рассмотренном методе особенности структуры формируются не только в ходе решения задачи выбора, но и на этапе формирования исходных данных для задачи, а также после получения решения, когда устанавливается схема функциональной взаимосвязи, исходя из совокупности КТС данной САК.

9.12. Методы и алгоритмы оптимизации структур

Из множества формализуемых задач структурного синтеза значительная их часть может быть сведена к определению экстремального значения целевой функции

$$F(\mathbf{X}) = \sum_{j=1}^m c_j x_j \rightarrow \max (\min) \quad (9.117)$$

при ограничениях

$$\sum_{j=1}^m a_{ij} x_j \leq b_i \left(\sum_{j=1}^m a_{ij} x_j \geq b_i \right), \quad i = \overline{1, n}, \quad (9.118)$$

$$x_j \geq 0, \quad j = \overline{1, m}; \quad (9.119)$$

$$x_j \text{ — целые числа, } j = 1, 2, \dots, p \ (p \leq m). \quad (9.120)$$

Если в сформулированной задаче ограничения (9.120) отсутствуют, то имеет место классическая задача линейного программирования, если ограничения (9.120) имеются и $p=m$, то данная задача является полностью целочисленной, при $p < m$ задача является частично целочисленной.

Задача линейного программирования. В настоящее время теория линейного программирования хорошо разработана и имеется целый арсенал методов решения задач линейного программирования — это, например симплекс-метод, реализующий последовательную процедуру направленного поиска оптимального значения целевой функции (9.117) при существующих ограничениях вида (9.118) и (9.119).

Симплекс-метод. Первоначально система неравенств (9.118) путем введения дополнительных переменных $x_{m+i} \geq 0$ преобразуется в систему уравнений таким образом, чтобы имело место одно из двух выражений

$$a_{i1} x_1 + a_{i2} x_2 + \dots + a_{im} x_m + x_{m+i} = b_i;$$

$$a_{i1} x_1 + a_{i2} x_2 + \dots + a_{im} x_m - x_{m+i} = b_i, \quad i = \overline{1, n}.$$

Такое изменение приводит просто к увеличению числа переменных, что не меняет существа задачи. Введем ряд понятий, широко применяемых в задачах линейного программирования.

Базисом называют любой набор из n таких переменных, что определитель, составленный из коэффициентов при этих переменных, не равен нулю. Отметим, что остальные $m-n$ переменных называют **свободными**.

Не уменьшая общности, рассмотрим сущность симплекс-метода на примере задачи максимизации целевой функции (9.117) при наличии ограничений (9.118) и (9.119). Для определения первоначального

базисного решения какие-либо $m-n$ переменные принимают за свободные, т.е. приравнивают нулю, при этом все базисные переменные выражают через свободные, после чего решают систему полученных уравнений. Если некоторые из базисных переменных окажутся отрицательными, то полученное базисное решение является *недопустимым* и производится переход к новому базису путем выбора новой совокупности свободных переменных. Базисное решение, в котором отсутствуют отрицательные переменные, называют *допустимым*.

После того как найдено допустимое базисное решение, проверяют, не достигнут ли максимум целевой функции $F(\mathbf{X})$. Если нет, то ищут новое допустимое базисное решение, но не любое, а такое, которое увеличивает значение целевой функции $F(\mathbf{X})$. Затем процедуру повторяют. Данный метод довольно быстро приводит к цели, так как позволяет исключить из рассмотрения большое число базисных решений, заведомо не обращающих в максимум целевую функцию $F(\mathbf{X})$.

Проверку того, не достигнут ли при найденном решении максимум целевой функции, можно сделать путем поиска нового базисного решения, при котором значение целевой функции $F(\mathbf{X})$ будет больше предыдущего. Для прихода к новому допустимому базисному решению одну из свободных переменных следует сделать базисной, при этом она будет отличной от нуля, т. е. возрастет. Следовательно, если какая-либо из свободных переменных входит в выражение для целевой функции со знаком « + », а значит, при ее увеличении целевая функция увеличивается, то максимум целевой функции не достигнут и данную свободную переменную следует перевести в базисную.

Однако при возрастании свободной переменной некоторые из базисных переменных начнут уменьшаться. Так как отрицательные значения переменных недопустимы, в качестве новой свободной переменной следует взять ту из базисных, которая раньше других обратится в нуль.

Задачи целочисленного программирования. В общем случае условие целочисленности накладывает дополнительные ограничения, вследствие которых максимальное значение целевой функции (в задачах максимизации) оказывается, как правило, меньше максимального значения целевой функции соответствующей задачи линейного программирования; в последней отсутствуют условия целочисленности переменных.

Комбинаторные методы. Методы решения задач целочисленного линейного программирования в классе комбинаторных методов

базируются на максимальном учете характера задач и конечности множества вариантов их решения. В их вычислительных схемах используется идея частичного перебора вариантов решения задач. Это достигается путем отбрасывания некоторых подмножеств вариантов, которые согласно известным свойствам оптимального решения заведомо могут считаться неперспективными.

Отличительной особенностью комбинаторных алгоритмов является и то, что во многих из них вообще не используется решение задач линейного программирования, соответствующей исходной целочисленной задаче линейного программирования.

Таким образом, для большинства комбинаторных методов не требуется специальных доказательств их конечности, что и способствовало широкому применению методов этого типа на практике. Большая группа комбинаторных методов базируется на достаточно общей схеме методов, которые объединены под названием — методы «ветвей и границ». В настоящее время схема метода «ветвей и границ» широко используется на практике.

Приближенные методы. Одной из основных причин бурного развития этой группы методов следует назвать сравнительную сложность реализации точных методов, а зачастую и просто невозможность применения последних для задач большой размерности и для задач, время решения которых ограничено. Другим фактором, способствующим на практике развитию приближенных методов, является то, что для многих практических задач в значительной мере точные решения не обеспечиваются из-за малой достоверности исходных данных.

Среди приближенных методов решения целочисленных задач линейного программирования следует выделить применение различного рода эвристических алгоритмов, основанных на:

- а) методах, построенных на использовании случайного поиска;
- б) методах, сочетающих случайный поиск с идеей локальной оптимизации;
- в) методах, вычислительные схемы которых строятся на максимальном учете специфики конкретных типов задач и др.

Рассмотрим два основных подхода к отысканию точного оптимального решения задач целочисленного программирования, базирующихся на методах отсекающих плоскостей и методах возврата.

Методы отсекающих плоскостей (методы отсечения). Исходным моментом решения задачи целочисленного программирования является оптимальное решение соответствующей задачи линейного программирования, полученной после отбрасывания условий целочис-

ленности. На каждой итерации добавляется линейное ограничение, удовлетворяющее целочисленному решению исходной задачи, но исключающее текущее нецелочисленное решение. Вычислительный процесс прекращается, как только будет достигнуто любое целочисленное решение. Сходимость обеспечивается за конечное, но иногда очень большое число итераций.

Сущность алгоритмов, основанных на методе отсечения, легко уяснить, обратившись к геометрическим представлениям в пространстве решений (см. п. 9.1). Определим выпуклую оболочку множества допустимых целочисленных точек (решений) как минимальное выпуклое множество, содержащее все эти точки. Допустимыми решениями будет не вся область допустимых решений, находящаяся внутри и на границе выпуклой оболочки, а лишь отдельные дискретные точки этой области, имеющие все целочисленные координаты. Целевая функция достигает оптимального значения в одной из вершин этой выпуклой оболочки, которая представляет собой одно из допустимых целочисленных решений.

Выпуклую оболочку можно представить конечным множеством линейных ограничений (9.118), как изображено на рис. 9.12.

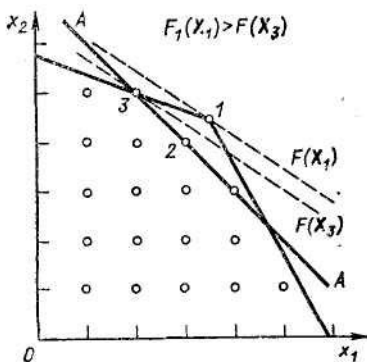


Рис. 9.12. Геометрическая иллюстрация принципа отсечения

Можно, не считаясь с условиями целочисленности, найти решение, определяемое точкой 1, а затем, округлив это решение до ближайших целых чисел, получить целочисленное решение в точке 2. Однако при этом может получиться решение, далекое от оптимального. Оптимальным целочисленным решением будет точка 3.

Для определения оптимального решения в алгоритмах отсечения вначале рассмотрим выпуклую оболочку, определенную линейными

ограничениями (9.118) и условиями неотрицательности переменных исходной задачи, и отыщем экстремальную точку этой оболочки (точка I на рис. 9.12). Если такое решение оказывается нецелочисленным, то добавляем ограничение, отсекающее текущую экстремальную точку и уменьшающее «объем» выпуклой оболочки (прямая $A-A$). Однако новое ограничение не отсекает ни одной экстремальной точки выпуклой оболочки, принадлежащей допустимым целочисленным решениям. В конечном итоге вводится такое число дополнительных ограничений, что экстремальная точка усеченной выпуклой оболочки представляет собой целочисленное решение исходной задачи.

Покажем, как формально построить дополнительные линейные ограничения, которым должно удовлетворить любое решение задачи (9.117) — (9.120). Рассмотрим ограничение

$$\sum_{j=1}^m a_{ij} x_j = b_i. \tag{9.121}$$

Возможно, одна или несколько величин a_{ij} и b_i являются дробными. Обозначим символом $[a_{ij}]$ целую часть a_{ij} , т. е. наибольшее целое число, меньшее действительного числа a_{ij} или равное ему. Поскольку на величины x_j наложено ограничение (9.119), любые значения x_j , соответствующие условию (9.121), должны удовлетворять более слабому ограничению

$$\sum_{j=1}^m [a_{ij}] x_j \leq b_i. \tag{9.122}$$

Поскольку сумма в левой части неравенства (9.122) должна быть целочисленной, (9.122) можно усилить следующим образом:

$$\sum_{j=1}^m [a_{ij}] x_j \leq [b_i]. \tag{9.123}$$

Наконец, неравенство (9.123) можно преобразовать в равенство, добавив дополнительную переменную x_{m+i} :

$$\sum_{j=1}^m [a_{ij}] x_j + x_{m+i} = [b_i], \tag{9.124}$$

где x_{m+i} — целое, положительное.

Таким образом, если в исходную задачу добавить линейное ограничение (9.124), задача все равно останется полностью целочисленной.

Уравнение отсекающей плоскости получается в результате вычитания из (9.124) уравнения (9.121):

$$\sum_{j=1}^m -f(a_{ij})x_j + x_{m+i} = -f(b_i), \quad (9.125)$$

где $f(a_{ij}), f(b_i)$ — дробная часть чисел a_{ij} и b_i .

Заметим, что текущее решение задачи линейного программирования не удовлетворяет ограничению (9.125), поскольку значение $x_{m+i} = -f(b_i)$ строго отрицательно.

Алгоритм отсечения (алгоритм Гомори) состоит из следующих шагов.

1. Найти оптимальное решение задачи линейного программирования (9.117) — (9.119) без условия целочисленности (9.120).

2. Прекратить вычисления, если текущее решение задачи является целочисленным. В противном случае выбрать какую-либо дробную базисную переменную. Составить ограничение (9.125) из уравнения, содержащего эту базисную переменную в текущем оптимальном решении задачи линейного программирования.

3. Добавить к исходной задаче линейного программирования новое ограничение (9.125), найти оптимальное решение задачи с дополнительным ограничением и вернуться к шагу 2.

Введение на шаге 3 отсекающего ограничения (9.125) наряду с условием $x_{m+i} > 0$ делает текущее решение задачи линейного программирования недопустимым. Отсечение текущего оптимального решения означает, что на шаге 3 значение целевой функции в задаче максимизации не может возрасти, а может лишь уменьшиться. Отметим, что алгоритмы отсечения не гарантируют получения допустимого целочисленного решения до самой последней итерации. Вследствие этого при преждевременном прекращении вычислений можно допустимое решение и не получить.

Методы возврата. В этой группе методов имеются различные модификации. Наиболее распространенным среди них является **метод ветвей и границ**, который предназначен для решения частично целочисленных задач. Как и в методе отсечения, решение задачи начинается с отыскания оптимального решения задачи линейного программирования без учета условия целочисленности. Затем формируется семейство связанных, но различных задач линейного программирования. Термин «возврат» определяет специфический способ формирования и решения последовательности задач.

Рассмотрим задачу максимизации (9.117) — (9.120). Допустим, что для каждой целочисленной переменной можно задать верхнюю и нижнюю границы, в пределах которых безусловно содержатся ее оптимальные значения

$$L_j \leq x_j \leq U_j, \quad j = \overline{1, p}. \quad (9.126)$$

Идея метода ветвей и границ основана на следующем элементарном факте. Рассмотрим любую переменную x_j и примем, что R есть некоторое целое число, где $L_j \leq R \leq U_{j-1}$. Тогда оптимальное решение задачи будет удовлетворять одному из ограничений

$$x_j \geq R + 1 \text{ или } x_j \leq R. \quad (9.127)$$

К примеру, если при решении задачи линейного программирования получено $x_i = 2,6$, то можно поставить и решить две задачи линейного программирования, причем в одну из них вводится согласно (9.127) условие $3 \leq x_i \leq U_i$, а к другую — условие $L_i \leq x_i \leq 2$. Предположим, что каждая из этих задач имеет оптимальное решение, удовлетворяющее условию целочисленности (9.120). Тогда решение, доставляющее большее значение целевой функции, является оптимальным решением исходной целочисленной задачи.

Метод ветвей и границ основан на решении некоторого множества задач линейного программирования. Границы (9.126) на каждую переменную x_j служат для конкретизации диапазона изменений переменных в задаче целочисленного программирования, что в свою очередь определяет мощность множества задач линейного программирования, используемых в процедуре реализации метода ветвей к границ. Поэтому трудоемкость вычислений определяется числом целочисленных переменных, содержащихся в задаче.

Алгоритм решения задачи целочисленного программирования методом ветвей и границ заключается в следующем. На каждой итерации (обозначим номер итерации через t) имеются нижняя оценка $F_t^*(\mathbf{X})$ оптимального значения целевой функции и список задач линейного программирования, подлежащих решению. Процедура решения состоит в последовательном улучшении оценки $F_t^*(\mathbf{X})$ и приближении ее к оптимальному значению $F_{\text{opt}}(\mathbf{X})$.

На итерации 1 список задач содержит одну задачу (9.117) — (9.119). На итерации t из списка выбирают и решают задачу линейного программирования. Если она не имеет допустимого решения или если полученное оптимальное значение целевой функции $F_{t \text{ опт}}(\mathbf{X}) \leq F_t^*(\mathbf{X})$, то нижняя оценка остается прежней и из списка выбирают очередную задачу для решения. Если полученное решение удовлетворяет условию целочисленности (9.120) и $F_{t \text{ опт}}(\mathbf{X}) > F_t^*(\mathbf{X})$, то полученное оптимальное решение $F_{t \text{ опт}}(\mathbf{X})$ на итерации t принимают в качестве

нижней оценки для последующих итераций. Если полученное оптимальное решение задачи линейного программирования не удовлетворяет условиям целочисленности (9.120), то выбирают нецелочисленную переменную x_i и решаемую задачу разбивают на две новые задачи линейного программирования путем введения в каждую из них по одному ограничению (9.127).

При остановке алгоритма в случае, если допустимому решению соответствует значение целевой функции $F_{i \text{ opt}}(\mathbf{X})=F_i^*(\mathbf{X})$, полученное решение оптимально, в противном случае допустимого решения не существует. Метод ветвей и границ особенно эффективен для решения комбинаторных задач, в частности задачи коммивояжера.

В общем виде процедура реализации этого метода заключается в следующем (рассмотрим ее применительно к задаче минимизации). Предположим, что имеется возможность получить нижнюю оценку качества решения $F_{i \text{ opt}}(\mathbf{X})$ как для всего множества V возможных решений, так и для его различных подмножеств. Разобьем множество на два непересекающихся подмножества A и B с точными нижними границами критерия качества $F_{A \text{ opt}}(\mathbf{X})$ и $F_{B \text{ opt}}(\mathbf{X})$, связанных с $F^*(\mathbf{X})$ очевидным соотношением

$$F_{\text{opt}}(\mathbf{X}) = \min \{ F_{A \text{ opt}}(\mathbf{X}), F_{B \text{ opt}}(\mathbf{X}) \}.$$

Поскольку множества A и B имеют меньшее число элементов, чем множество V , т.е. возможность получить нижние оценки качества $F_{A \text{ opt}}^*(\mathbf{X})$ и $F_{B \text{ opt}}^*(\mathbf{X})$ более близкие к $F_{A \text{ opt}}(\mathbf{X})$ и $F_{B \text{ opt}}(\mathbf{X})$, чем в первоначальном множестве V , это означает, что оценка $\min \{ F_{A \text{ opt}}^*(\mathbf{X}), F_{B \text{ opt}}^*(\mathbf{X}) \}$ будет более близка к $F_{\text{opt}}(\mathbf{X})$, чем оценка $F_V^*(\mathbf{X})$.

Можно предположить, что оптимальное решение будет с большой вероятностью принадлежать тому из подмножеств A и B , которое имеет меньшую нижнюю оценку. Если оказалось, что $F_{B \text{ opt}}^*(\mathbf{X}) < F_{A \text{ opt}}^*(\mathbf{X})$, то есть все основания для детального исследования подмножества B . Последнее разбивают на два подмножества C и D с соответствующими нижними оценками $F_{C \text{ opt}}^*(\mathbf{X})$ и $F_{D \text{ opt}}^*(\mathbf{X})$, и если оказалось, что $F_{D \text{ opt}}^*(\mathbf{X}) < F_{C \text{ opt}}^*(\mathbf{X})$, то подобному разбиению подвергается подмножество D . Эту процедуру продолжают до тех пор, пока не придут к подмножеству, состоящему из одного элемента с оценкой $F^*(\mathbf{X})=F_0(\mathbf{X})$. Процедура разбиения обычно представляется в виде дерева решений (рис. 9.13).

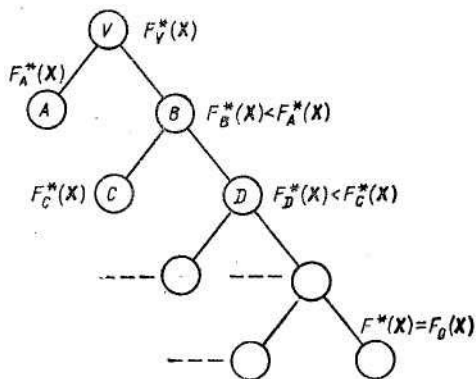


Рис. 9.13. Геометрическая иллюстрация метода ветвей и границ

Найденное решение необходимо проверить на оптимальность. Для этого проверяют, нет ли среди нерассмотренных множеств A , C и т. д. элемента со значением нижней оценки $F^*(X) < F_0(X)$. При этом множества, у которых $F^*(X) \geq F_0(X)$, дальше не рассматриваются.

Разбиение перспективных множеств либо приведет к решению с оценкой $F^*(X) < F_0(X)$, которое должно быть принято за оптимальное, либо позволит убедиться в оптимальности полученного ранее решения с оценкой $F_0(X)$. Из процедуры реализации метода ветвей и границ становится ясным, почему этот метод получил название метода возврата.

Метод ветвей и границ наряду с методами отсечения обладает существенными достоинствами с вычислительной точки зрения. Алгоритмы, построенные на этих методах, сравнительно легко программируются на ЭВМ и реализуются на любой итерации без вмешательства человека, однако их эффективность резко снижается при увеличении размерности решаемой задачи.

Применение моделей и методов математического программирования при консультировании проблем было рассмотрено в примерах п. 9.2. Ниже приводятся примеры постановки типовых задач структурного синтеза в терминах математического программирования.

Пример 9.8. Формирование рекомендаций по решению задачи оптимизация структуры сети электросвязи. Процесс консультирования по решению задачи проектирования региональных сетей электросвязи состоит из ряда взаимосвязанных этапов:

- построения структуры первичной сети связи, обеспечивающей связь между заданными пунктами региона по критерию минимальной стоимости;

- корректировки полученной структуры сети путем введения дополнительных связей между пунктами с целью получения живучей сети электросвязи;

- построения вторичной сети связи, заключающейся в определении конфигурации сетки пучков каналов электросвязи для обеспечения необходимой пропускной способности каналов между различными пунктами региона.

Формально консультационную задачу по формированию рекомендаций синтеза структуры первичной сети связи можно представить в виде следующей задачи математического программирования. Задана матрица расстояний $D=|d_{ij}|$ размерности $n \times n$ между всеми n пунктами данного региона. Необходимо определить такую структуру сети, которая обеспечивала бы связь между всеми пунктами региона по критерию минимальной стоимости. При этом будем считать, что стоимость канала связи между пунктами i и j пропорциональна расстоянию d_{ij} между ними.

Введем псевдодобулевы переменные:

$$x_{ij} = \begin{cases} 1, & \text{если имеется прямой канал связи между пунктами } i \text{ и } j; \\ 0 & \text{в противном случае.} \end{cases}$$

Задача синтеза оптимальной структуры сети связей по критерию минимальной стоимости заключается в определении таких переменных $x_{ij} \in \{0, 1\}$, которые обращали бы в минимум целевую функцию

$$F(\mathbf{X}) = \sum_{i=1}^n \sum_{j=1}^n d_{ij} x_{ij}$$

при ограничениях

$$\sum_{i=1}^n x_{ij} \geq 1, \quad \sum_{j=1}^n x_{ij} \geq 1, \quad \sum_{i=1}^n \sum_{j=1}^n x_{ij} \geq n-1.$$

Решение сформулированной задачи математического программирования удобно определять методом ветвей и границ.

Пример 9.9. *Консультационная задача формирования рекомендаций осуществления синтеза структуры памяти специализированной ЭВМ.* Требуется синтезировать структуру внутренней памяти (СОЗУ—ОЗУ—ПЗУ) специализированной ЭВМ, работающей в режиме реального времени, совместно с определением необходимого пакета объектных программ и их размещением в различных блоках системы памяти.

Пусть ЭВМ ориентирована на решение множества задач (заявок)

$C = \{c_1, \dots, c_n\}$, причем каждая задача $c_i \in C$ встречается с относительной частотой g_i , и пусть на множестве C задано его разбиение на подмножества C_l , $l=1, 2, \dots, u$, ($u \leq n$), так что в подмножество $C_l \subset C$ входят только те задачи $c \in C_l$, которые имеют l -й приоритет, причем чем меньше l , тем выше приоритет.

Для решения поступающих в ЭВМ задач имеется множество алгоритмов $\mathcal{L}' = \{L_1, \dots, L_p\}$, $p \leq n$, по отношению к которым множество C разбито на подмножества $C(L_i) \subseteq C$ таким образом, что если $c \in C(L_i)$.

то задача может быть решена с помощью алгоритма $L_i \in \mathcal{L}'$. Каждый из алгоритмов $L_i \in \mathcal{L}'$ может быть реализован m_i методами, т. е. для каждой c_i можно записать множество упорядоченных пар $L_i = \{(i, j)\}$, $i=1, 2, \dots, p$, $j=1, 2, \dots, m_i$, где любая пара (i, j) описывает метод решения задачи c_i . Каждый из методов $(i, j) \in L_i$ характеризуется: способом хранения программы (в ОЗУ или ПЗУ); числом требующихся ячеек a_{ij} ОЗУ или ПЗУ для записи программ; числом b_{ij} ячеек СОЗУ для хранения часто встречающихся данных и промежуточных результатов и числом h_{ij} ячеек ОЗУ для хранения исходных данных при обработке заявки c_i в реальном времени; количеством $r_{ij}^{(k)}$ ресурса k , где под ресурсами будем понимать габариты, массу, стоимость, число используемых ячеек или панелей, потребляемую мощность и т. д. Поскольку программы могут храниться как в ПЗУ, так и в ОЗУ, это влияет на время реализации программы, надежность и оперативность в смене программ. Часть параметров, например, время решения задачи, стоимость, габариты, потребляемая мощность и др., подчиняется условиям аддитивности, в то время как такие параметры, как надежность, диагностическая возможность, ремонтпригодность и др., являются неаддитивными.

Задача заключается в формировании рекомендаций по определению структуры внутренней памяти и такого набора программ Π решения поступающих задач, чтобы удовлетворялись все ограничения на параметры ЭВМ, а выбранный критерий оптимальности достигал своего экстремального значения.

В частности, в качестве критерия оптимальности можно выбрать суммарный объем ОЗУ и ПЗУ, который требуется для хранения программ и данных и который следует минимизировать.

Дадим математическое описание задачи. Введем псевдобулевы переменные:

объем СОЗУ и ОЗУ соответственно; T_l — максимально допустимое время реализации задач, имеющих l -й приоритет.

Задача (9.128) — (9.131) также является задачей дискретного программирования с псевдодобулевыми переменными. Подставляя в целевую функцию задачи оптимизации другие параметры, в частности стоимость, время решения задач, энергетические и другие параметры системы памяти, можно оптимизировать структуру системы памяти ЭВМ по соответствующим критериям.

Рассмотренная задача может быть использована для синтеза технических и программных средств САК.

9.13. Последовательные методы в задачах консультирования сложных проблем

Выбор оптимального варианта структуры консультируемой проблемы методами, базирующимися на полном переборе вариантов, является дорогостоящей, трудоемкой и, как правило, неосуществимой процедурой. Использование методов математического программирования для формирования рекомендаций по решению задач структурного синтеза КП требует большой предварительной подготовки для исследования пространства рекомендаций и не всегда оправдано из-за больших трудностей учета многочисленных факторов, влияющих на корректность постановки задачи оптимального консультирования, и из-за существенных вычислительных трудностей решения задач математического программирования большой размерности.

При консультировании сложных проблем довольно эффективными оказываются последовательные методы анализа и синтеза.

Последовательные методы анализа основаны на направленной генерации множества вариантов формируемых рекомендаций и осуществлении процедуры анализа вариантов с целью выбора наилучшего путем последовательного отсеивания неперспективных вариантов.

При решении консультационных задач можно рекомендовать метод последовательного анализа вариантов, основанный на обобщении идей теории последовательных статистических решений А. Вальда.

Последовательные алгоритмы синтеза основаны на наращивании структуры путем добавления по определенным правилам элементов к некоторому начальному элементу.

Метод последовательного формирования рекомендаций, анализа и отсеивания вариантов. В основе этого метода лежит идея

процесса формирования рекомендаций в виде многоступенчатой структуры. Каждая ступень связана с проверкой наличия определенных свойств у подмножества вариантов и либо ведет к непосредственному сокращению исходного множества вариантов, либо подготавливает возможность такого сокращения в будущем. Для решения задачи необходимо определить отличительные свойства, которыми должен обладать искомый вариант. Первоначально из множества признаков выбирают наиболее легко проверяемые и присущие одновременно возможно большему числу вариантов. После этого выбор численной схемы решения состоит в выборе рационального порядка проверки признаков, позволяющего провести отсев неконкурентоспособных вариантов и найти оптимальный.

Алгоритмы последовательного анализа вариантов основаны на принципе оптимальности, который представляет собой естественное обобщение принципа оптимальности динамического программирования для решения многошаговых задач оптимизации.

Напомним, что в методе динамического программирования выбор рекомендации (управления) на отдельном шаге производится не с точки зрения интересов данного шага, выражающихся в минимизации потерь на данном шаге, а с точки зрения всего многошагового процесса формирования рекомендаций в целом, выражающихся в минимизации суммарных потерь на всех последующих шагах. Отсюда следует основное свойство оптимального процесса формирования рекомендаций, заключающееся в том, что каковы бы ни были начальное состояние и начальная рекомендация, последующие рекомендации на каждом шаге должны быть оптимальными относительно состояния, являющегося результатом применения первой рекомендации. Из этого свойства следует, что ***оптимизация выбора рекомендации для многошагового процесса формирования рекомендаций заключается в выборе рекомендаций только на последующих шагах процесса.***

Основное правило отсева бесперспективных вариантов — монотонная рекурсивность, идейно родственная критерию оптимальности динамического программирования.

Пусть имеются множество $W = \{w\}$ вариантов рекомендаций и множество опытов $\Pi = \{\pi_a\}$. Каждый вариант $w \in W$ описывается некоторым множеством признаков. Задача состоит в определении подмножества $W^* \subseteq W$, инвариантного относительно любого π_a и содержащего оптимальную рекомендацию $w_0 \in W^*$. Для определения подмножества W^* необходимо поставить опыты по анализу и оценке свойств элементов $w \in W$. Исходы опытов позволяют отбросить неперспективные варианты w , которые не имеют общих частей с

элементами подмножества W^* , и сделать заключение о целесообразности постановки последующих опытов с целью определения элементов, входящих в подмножество W^* .

Приведем описание правила отсева. Пусть задано некоторое базовое множество X . Обозначим множество конечных последовательностей вида

$$\mathbf{p} = (x_1, \dots, x_l, \dots, x_{k_p}), \quad x_l \in X, \quad 1 \leq l \leq k_p,$$

через $P(X)$. В этом множестве выделено некоторое подмножество допустимых последовательностей $W(X) \subseteq P(X)$, в свою очередь во множестве $W(X)$ выделено подмножество $W_0(X) \subseteq W(X)$ полных допустимых последовательностей. Пусть задана последовательность \mathbf{p} . Последовательность

$$\mathbf{p}_l = (x_1, \dots, x_l), \quad 1 \leq l \leq k_p,$$

называют l -м начальным отрезком последовательности \mathbf{p} , а последовательность

$$\mathbf{p}^{(q)} = (x_q, x_{q+1}, \dots, x_{k_p}), \quad 1 \leq q \leq k_p,$$

называют q -м конечным отрезком. Если $q=l+1$, то соответствующие части последовательности \mathbf{p} называют сопряженными. Рассмотрим две допустимые последовательности \mathbf{p}_1 и \mathbf{p}_2 . В \mathbf{p}_1 выделены l_1 -й начальный отрезок \mathbf{p}_{1l_1} и $(l_1 + 1)$ -й конечный отрезок \mathbf{p}_{1l_1+1} , в \mathbf{p}_2 выделены l_2 -й начальный отрезок \mathbf{p}_{2l_2} и (l_2+1) -й конечный отрезок \mathbf{p}_{2l_2+1} . Если функционал Φ , определенный на множестве $W(X)$, обладает тем свойством, что из

$$\mathbf{p}_{1l_1} \in W(X); \quad \mathbf{p}_{2l_2} \in W(X); \quad \mathbf{p}_{1l_1+1} \equiv \mathbf{p}_{2l_2+1}; \\ \Phi(\mathbf{p}_{1l_1}) < \Phi(\mathbf{p}_{2l_2})$$

следует $\Phi(\mathbf{p}_1) < \Phi(\mathbf{p}_2)$, то его называют **монотонно-рекурсивным**.

Вариант $w_0 \in W^*$, характеризующийся максимальным значением функционала Φ , определяется согласно обобщенному принципу оптимальности: если заданы монотонно-рекурсивный функционал Φ и две допустимые последовательности \mathbf{p}_1 и \mathbf{p}_2 , причем

$$\Phi(\mathbf{p}_1) < \Phi(\mathbf{p}_2), \quad P(\mathbf{p}_1) \subseteq P(\mathbf{p}_2),$$

то последовательности, у которых начальным является отрезок \mathbf{p}_1 будут неперспективными и не входят в подмножество W^* . Здесь $P(\mathbf{p}_i)$ — множество конечных отрезков, сопряженных с последовательностью \mathbf{p}_i .

Таким образом, схема последовательного выбора оптимального варианта сводится к следующим повторяющимся процедурам:

1. Множество вариантов рекомендаций W разбивают на несколько подмножеств, каждое из которых обладает специфическими свойствами.

2. На основании обобщенного принципа оптимальности отсеивают те варианты $w \in W$, значение функционала Φ для которых заведомо не может быть максимальным.

3. Последовательно формируются и анализируются не отсеянные ранее варианты рекомендаций.

4. Для множества, состоящего из вновь образованных в п. 3 допустимых последовательностей, и последовательностей не исключенных и не продолженных ранее, выполняют операции п. 2.

Далее операции п. 2, 3 и 4 циклически повторяются. Если на каком-то этапе процесса формирования рекомендаций не останется ни одной последовательности, требующей своего развития до получения полной последовательности, то процесс завершен и в качестве рекомендации берется одна из рассмотренных полных допустимых последовательностей с наибольшим значением функционала $\Phi(\mathbf{p})$.

Основным достоинством метода последовательного анализа вариантов является значительная экономия в вычислительной процедуре за счет отсеивания бесперспективных начальных частей вариантов до их полного построения. Экономия является тем существенней, чем больше определяющих свойств задачи использовано для построения процедур анализа и отсева.

Последовательные алгоритмы структурного синтеза.

Последовательные алгоритмы синтеза относятся к классу

эвристических. Их достоинством является сравнительно высокая экономичность по затратам машинного времени и требуемому объему оперативной памяти за счет отсутствия процедуры многократного анализа вариантов структуры. Однако, как правило, последовательные алгоритмы дают не оптимальные, а близкие к оптимальным рекомендации.

Развитие диалоговых средств общения консультанта с ЭВМ способствует широкому применению последовательных методов и алгоритмов в структурном синтезе КП из разнообразных областей. В качестве иллюстрации рассмотрим идею формирования последовательных алгоритмов для решения консультационных задач формирования рекомендаций по осуществлению конструкторского проектирования ЭВА — задач компоновки, размещения и трассировки.

При иерархической организации конструкции ЭВА под *компоновкой* понимают определение состава типовых конструкций каждого уровня. Задача компоновки обычно решается «снизу — вверх», т.е. известные схемы $i-1$ -го уровня необходимо распределить по конструкциям i -го уровня. Так, например, на самом низшем уровне элементами могут выступать корпуса элементов, а конструкциями (блоками) — типовые элементы замены, связанные друг с другом путем разъемных соединений.

В качестве критериев оптимальности при решении задач компоновки наиболее часто используют критерии либо минимума суммарного числа N_i типов модулей

$$N_i = \sum_j x_{ij},$$

где x_{ij} — число модулей j -го типа i -го уровня схемы, либо минимума межблочных соединений

$$R_i = 0,5 \sum_{k=1}^{N_i} R_{ik},$$

где R_{ik} — число внешних связей каждого модуля i -го уровня.

Первый критерий непосредственно связан с конструктивными характеристиками аппаратуры и показателем технологичности стоимости, второй критерий ведет к повышению надежности конструктивной реализации схемы за счет сокращения числа разъемных соединений, уменьшению помех и задержек сигналов благодаря снижению числа межблочных соединений.

Для иллюстрации рассмотрим принцип построения последовательных алгоритмов компоновки по критерию минимума межблочной связности. Этот критерий широко используется при компоновке оборудования в различных технических приложениях.

Идея алгоритмов заключается в следующем. Первоначально выбирают исходный элемент (модуль) схемы. Выбор начального элемента основывается на схемотехнических соображениях.

В первый компоновываемый узел включены все элементы, смежные с начальным, и сам начальный элемент. Если полученное число элементов равно максимально допустимому числу элементов в первом узле, то компоновка этого узла заканчивается. Если это число больше или меньше максимально допустимого, то выполняются соответствующие операции по устранению лишних или добавлению недостающих элементов, причем из нескомпонованных элементов выбирают такой, который имеет наибольшее число связей с элементами, уже вошедшими в состав компоновываемого узла. Далее сформированный узел

удаляют из схемы и компонуют новые узлы. Процесс повторяется до тех пор, пока схема не будет разбита на требуемое число частей или не будет выяснена невозможность этого.

Сформулируем описанный алгоритм в терминах теории графов. Пусть задан граф схемы $G=(X, U)$, который необходимо разбить на l частей G_1, G_2, \dots, G_l с числом вершин в каждом соответственно

$$n_1, n_2, \dots, n_l \quad \left(\sum_{i=1}^l n_i = n, \quad |X| = n \right).$$

Первоначально в графе G определяют вершину $x_i \in X$ с наибольшей локальной степенью $\rho(x_i)$ [напомним, что локальной степенью $\rho(x_i)$ вершины $x_i \in X$ называют число ребер, инцидентных этой вершине графа]. Если таких вершин несколько, то предпочтение отдается той, которая имеет большее число кратных ребер. Вершина x_i и все смежные с ней вершины включаются в граф G_l . Обозначим это множество вершин через Γx_i . Если $|\Gamma x_i| = n_l$ то G_l образован, если же $|\Gamma x_i| > n_l$, то из графа G_l удаляют вершины, связанные с остающимися вершинами графа G меньшим числом ребер. Когда $|\Gamma x_i| < n_l$, то выбирают вершину $x_j \in \Gamma x_i$, удовлетворяющую условию

$$\sigma(x_j) = \max_{x_k \in \Gamma x_i} \{ \sigma(x_k) \} = \max_{x_k \in \Gamma x_i} \{ \rho(x_k) - a_k \},$$

где a_k — число ребер, соединяющих вершину x_k со всеми невыбранными вершинами графа G .

Строят множество вершин Γx_j , смежных x_j , и процесс выборки вершин G_l повторяют. Образованный подграф G_l исключают из исходного и получают граф $G^*=(X^*, U^*)$, где $X^* = X \setminus X_l$, $U^* = U \setminus U_l$. Далее в графе G^* выбирают вершину с наибольшей локальной степенью, включают ее в G_2 и процесс повторяют до тех пор, пока граф G не будет разрезан на l частей.

Первоначальную компоновку можно улучшить с помощью итерационных алгоритмов, основанных на реализации методов парных или групповых перестановок элементов из одной части схемы в другую таким образом, чтобы улучшилось значение целевой функции с учетом заданных ограничений.

Задача *размещения* заключается в определении оптимального (с точки зрения выбранного критерия оптимальности) положения элементов и связей между ними в монтажном пространстве типовой конструкции с учетом заданных конструктивно-технологических ограничений. Исходными данными в задаче решения являются принципиальная электрическая схема узла или устройства, метрические параметры и топологические свойства монтажного пространства.

Главная цель размещения — создание наилучших условий для трассировки с учетом обеспечения тепловых режимов и электромагнитной совместимости электрорадиоэлементов. Несмотря на обилие существующих критериев размещения (минимума пересечений, минимума суммарной длины соединений и т.д.) истинной целью размещения компонентов является максимальное упрощение процесса трассировки соединений, т. е. достижение минимального числа непроведенных трасс. При размещении n электрорадиоэлементов в регулярном монтажном пространстве с числом позиций m общее число размещений $N(n, m)$ определяется как

$$N(n, m) = n! C_m^n = m! I(m - n)!$$

В связи с этим поиск оптимального размещения с помощью перебора нецелесообразен уже при $n > 15$.

Имеется много разновидностей последовательных алгоритмов размещения. Основной идеей этих алгоритмов является идея упорядочения электрорадиоэлементов по определенным признакам. Сначала устанавливают очередность электрорадиоэлементов, а затем для каждого из них определяют наилучшую позицию по выбранному критерию, например по суммарной длине связей с уже размещенными компонентами. Затем процесс повторяют для оставшихся компонентов и свободных позиций. Связность размещаемых элементов задается матрицей смежности R графа $G=(X, U)$. Для выбора размещаемого элемента используют различные оценки степени связности.

Пусть на k -м шаге алгоритма размещено $I_k \subset I$ элементов, тогда $I'_k = I \setminus I_k$ — множество еще не размещенных элементов. Основными правилами для выбора элемента на $(k+1)$ -м шаге алгоритма являются:

а) максимум суммарной связности h_i со всеми размещенными элементами

$$h_i = \max_{i \in I'_k} \left\{ \sum_{j \in I_k} r_{ij} \right\}, \quad i \neq j;$$

б) максимум разности связей f_i между размещенными и неразмещенными элементами

$$f_i = \max_{i \in I'_k} \left\{ \sum_{j \in I_k} r_{ij} - \sum_{j \in I'_k} r_{ij} \right\}, \quad i \neq j.$$

Выбранный для размещения элемент устанавливают в такую позицию среди оставшихся незаполненных, при которой будет иметь наименьшее значение некоторая целевая функция. Для многих задач размещения в качестве такой функции может быть выбрана суммарная длина связей с уже размещенными элементами.

Последовательные алгоритмы размещения требуют небольших затрат машинного времени, относят их к классу полиномиальных алгоритмов со сложностью $O(n)$, приводящих к неоптимальным решениям. Улучшить решение можно путем применения итерационных алгоритмов компоновки, основанных на изменении позиций одиночных элементов или групп элементов. Итерационные алгоритмы также относятся к классу полиномиальных со сложностью порядка $O(n^2) — O(n^4)$.

Задача *трассировки* заключается в определении конкретной геометрии печатного или проводного монтажа, реализующего соединения между элементами схемы. Исходными данными для трассировки являются список цепей, метрические параметры и топологические свойства типовой конструкции и ее элементов, а также результаты решения задачи размещения, по которым находят координаты выводов элементов.

При решении задачи трассировки строят множество трасс, соединяющих выводы элементов соответствующих цепей схемы. Разработка отдельной трассы представляет собой построение на фиксированных вершинах минимального покрывающего или связывающего дерева, а разработка множества трасс сводится к построению леса непересекающихся минимально покрывающих или связывающих деревьев. Известно, что на n вершинах можно построить n^{n-2} различных деревьев, поэтому точное решение задачи трассировки методом полного перебора практически нереализуемо.

В последовательных алгоритмах трассировки трассы цепей проводятся в определенном порядке одна за другой, при этом каждая проложенная трасса становится препятствием для всех последующих цепей. В последовательных алгоритмах производят локальную оптимизацию качества трассировки каждой отдельной трассы без учета влияния размещения данной трассы на возможность проведения последующих. Это приводит к тому, что некоторые участки платы могут оказаться заблокированными.

Большинство известных алгоритмов трассировки основывается на волновом алгоритме (алгоритм Ли). Основные принципы волнового алгоритма Ли заключаются в следующем. Плоскость трассировки разбивают на прямоугольные площадки — дискреты заданного размера. Размер дискретной площадки определяется допустимыми размерами проводников и расстояниями между ними. Задача проведения трасс сводится к получению последовательности дискретов, соединяющих элементы a и b , соответствующие началу и концу проводимой трассы.

Вводим целевую функцию $F=F(f_1, \dots, f_n)$ как критерий качества пути. Начиная с элемента a дискретам, соседним с ранее просмотренным, присваивают определенное значение целевой функции $F_{ij}=m(i, j)$. Этот этап проводится итерационно до элемента b , которому присваивают некоторое значение веса $m(i_b, j_b)$. Затем, начиная от элемента b , перемещаются к элементу a по пройденным дискретам таким образом, чтобы значения целевых функций дискретов монотонно убывали. В результате получается трасса, соединяющая элементы a и b .

Обычно работа алгоритма Ли реализуется следующим образом. На трассируемой плоскости из источника a моделируется распространение волны до тех пор, пока не будет достигнута точка b или пока на некотором шаге фронт волны не сможет включить ни одного незанятого дискрета. Эту часть алгоритма называют распространением волны. После этого проводят трассу, начиная от конечной точки b , по дискретам с последовательно уменьшающимися весами. На рис. 9.14 цифры в квадратах соответствуют весам дискретов, занятые дискреты заштрихованы, а построенная трасса показана штриховой линией.

8	9	10	11	10	11	b	13
7		9	8	9			12
6		8	7	8	9	10	11
5		5	6	7			
4	3	4	5	6	7		
3	2	3			6		
2	1				5	6	7
1	a	1	2	3	4	5	6

Рис. 9.14. Пример соединения элементов a и b с помощью волнового алгоритма

Существует несколько вариантов проведения пути, из которых конструктор (или ЭВМ) выбирает один, наиболее удовлетворяющий заданным требованиям. Имеется многообразие модификаций волновых алгоритмов, направленных на повышение быстродействия трассировки, уменьшение объема требуемой оперативной памяти ЭВМ и т. д. Волновые алгоритмы широко применяют в различных

областях техники, в частности при разработке сетей связи и сетей ЭВМ.

9.14. Алгоритм распределения консультационных функций по модулям САК

Пусть задано множество консультационных функций, реализуемых в САК ($i = 1, 2, \dots$), задано множество функциональных модулей САК ($j = 1, 2, \dots$). Необходимо так распределить функции по модулям системы, чтобы достигнуть максимального значения эффективности, не выходя из области допустимых ограничений. Пусть C_{ij} — затраты на реализацию i -й функции (консультационных операций) в j -м модуле, t_{ij} — время реализации функции в j -м модуле.

Вводим дополнительную переменную

$$x_{ij} = \begin{cases} 1, & \text{если } i \text{ — } i\text{-я функция реализуется в } j\text{-м модуле,} \\ 0 & \text{— в противном случае.} \end{cases}$$

Оптимизация распределения консультационных функций по модулям может производиться по одной из следующих целевых функций:

$$\min_{x_{ij}} \sum_{i=1}^I \sum_{j=1}^J C_{ij} x_{ij}; \tag{9.132}$$

$$\min_{x_{ij}} \sum_{i=1}^I \sum_{j=1}^J t_{ij} x_{ij}; \tag{9.133}$$

$$\min_{x_{ij}} \left[\max_j \left(\sum_{i=1}^J t_{ij} x_{ij} \right) \right]. \tag{9.134}$$

Функция (9.132) соответствует минимизации затрат, функция (9.133) — минимизации общего времени реализации консультационных функций, а функций (9.134) — минимизации максимального времени, реализации консультационной функции в модуле. При этом могут учитываться следующие ограничения:

а) связи между функциями (обычно задаются графом $G(J)$, где J — множество функций);

б) связи между модулями, в состав которых входят технические средства (элементы) системы (обычно задаются графом $G(J)$, где J — множество модулей);

в) ограничение на общее время реализации консультационных функций

$$\sum_{i=1}^I \sum_{j=1}^J t_{ij} x_{ij} \leq T_{\text{в}},$$

если минимизируются затраты, либо ограничение на общие затраты по реализации функций в САК

$$\sum_{i=1}^I \sum_{j=1}^J C_{ij} x_{ij} \leq C_{\text{в}},$$

если минимизируется время реализации всех функций;

г) ограничение на загрузку консультационными процедурами каждого модуля

$$\sum_{i=1}^I \lambda_i t_{ij} x_{ij} \leq \rho_{\text{в}j}, \quad j = 1, 2, \dots, J,$$

где λ_i — интенсивность поступления i -й консультационной задачи на формирование рекомендаций по ее (задачи) решению; $\rho_{\text{в}j}$ — допустимая загрузка j -го модуля;

д) если каждая функция реализуется только в одном модуле системы, то

$$\sum_{j=1}^J x_{ij} = 1, \quad i = 1, 2, \dots, I.$$

В зависимости от того, как выбрана целевая функция и какие ограничения учитываются, возникает ряд частных постановок задачи оптимального распределения консультационных функций по модулям.

Минимизация общих затрат (общего времени) при ограничениях на загрузку каждого из модулей и при условии, что каждая функция реализуется только в одном модуле системы, т. е.

$$\min \sum_{i=1}^I \sum_{j=1}^J C_{ij} x_{ij} \quad \text{либо} \quad \min \sum_{i=1}^I \sum_{j=1}^J t_{ij} x_{ij} \quad \text{при} \quad \sum_{i=1}^I \lambda_i t_{ij} x_{ij} \leq \rho_{\text{в}j};$$

$$\sum_{j=1}^J x_{ij} = 1; \quad j = 1, 2, \dots, J; \quad i = 1, 2, \dots, I.$$

Минимизация общих затрат (общего времени) при ограничениях на общее время (общие затраты) и при условии, что каждая функция реализуется только в одном модуле системы, т. е.

$$\min \sum_{i=1}^I \sum_{j=1}^J C_{ij}x_{ij};$$

$$\sum_{i=1}^I \sum_{j=1}^J t_{ij}x_{ij} \leq T_8;$$

$$\sum_{j=1}^J x_{ij} = 1; i = 1, 2, \dots, J; x_{ij} = 0; 1. \quad (9.135)$$

В дальнейшем будем рассматривать только задачу (9.135). Дерево ветвления будем строить следующим образом. Подмножество первого уровня разбиения формируем, фиксируя возможность реализации первой функции различным модулям $(x_1, x_2, \dots, x_{j_1}, \dots, x_j)$. Множество x_{j_1} включает в себя варианты, где первая функция реализуется в модуле j_1 , а распределение остальных функций по модулям произвольное. Аналогично, множеством второго уровня формируем, фиксируя соответствие второй функции различным модулям. Множество $x_{j_1 j_2}$ включает в себя все варианты реализаций, где первая функция реализуется в модуле j_1 , вторая функция — в модуле j_2 , а остальные функции имеют произвольное распределение по модулям и т. д. Для каждого из подмножеств (вершин дерева) необходимо построить оценки целевой функции и ограничения. Общее выражение оценки целевой функции для множества вариантов $x_{j_1, j_2, \dots, j_i, \dots, j_e}$ в данной задаче может быть построено следующим образом:

$$V_c(x_{j_1, j_2, \dots, j_i, \dots, j_e}) = \sum_{i \leq e} C_{ij_i} + \sum_{i > e} \min_l C_{ij_i}, \quad (9.136)$$

а общее выражение для оценки ограничения может быть построено аналогично:

$$V_t(x_{j_1, j_2, \dots, j_i, \dots, j_e}) = \sum_{i \leq e} t_{ij_i} + \sum_{i > e} \min_l t_{ij_i}, \quad (9.137)$$

где $j_1, j_2, \dots, j_i, \dots, j_e$ — множество модулей системы, закрепленных за соответствующими функциями $1, 2, \dots, i, \dots, e$.

Оценка для функции ограничения (9.137) необходима в данном случае для исключения из процесса ветвления множества заведомо неподходящих вариантов с учетом принятого ограничения на общее время реализации функций. Для ускорения процесса поиска используются всякие дополнительные приемы, учитывающие специфику задачи. Рассмотрим их на численном примере. Пусть заданы: матрицы стоимости

$$C = \begin{pmatrix} 3 & 7 & 2 & 2 \\ 4 & 8 & 1 & 3 \\ 5 & 9 & 6 & 2 \\ 6 & 10 & 7 & 1 \\ 7 & 5 & 3 & 1 \end{pmatrix};$$

временных затрат

$$T = \begin{pmatrix} 1,5 & 3 & 2 & 9 \\ 2 & 6 & 5 & 10 \\ 3 & 7 & 6 & 11 \\ 4 & 8 & 7 & 12 \\ 4 & 9 & 8 & 5 \end{pmatrix};$$

значение $T_8 = 20$.

Так как каждая функция может быть реализована только одним модулем системы, то

$$\min_{x_{ij}} \sum_{i=1}^I \sum_{j=1}^J C_{ij}x_{ij} = \sum_{i=1}^I \min_{x_{ij}} \sum_{j=1}^J C_{ij}x_{ij}.$$

Используя это условие и принимая во внимание ограничение по времени, можно исключить из матриц C и T элементы, которые не влияют на выбор оптимального решения. Такими элементами в каждой строке являются:

$$C_{ij} > \min C_{ij} = C_{ip}; \quad t_{ij} > t_{ip}.$$

После проведения такого преобразования матрицы $C^{(0)}$ и $T^{(0)}$ будут выглядеть следующим образом:

$$C^{(0)} = \begin{pmatrix} 3 & - & 2 & - \\ 4 & - & 1 & - \\ 5 & 9 & 6 & 2 \\ 6 & 10 & 7 & 1 \\ 4 & - & - & - \end{pmatrix};$$

$$T^{(0)} = \begin{vmatrix} 1,5 & - & 2 & - \\ 2 & - & 5 & - \\ 3 & 7 & 6 & 11 \\ 4 & 8 & 7 & 12 \\ 4 & - & - & 5 \end{vmatrix}.$$

Принимая во внимание ограничение по времени $T_3 = 20$, можно из матриц $C^{(0)}$ и $T^{(0)}$ также вычеркнуть элементы, при которых всегда будет нарушено ограничение по времени. Распишем для выделения таких элементов данное ограничение в следующем виде:

$$\sum_{i=1}^{r-1} \min_j t_{ij} + t_{rj} + \sum_{i=r+1}^J \min_i t_{ij} \leq T_3. \quad (9.138)$$

Изменяя индекс строки r ($r = 1, 2, \dots, i, \dots, J$) и просматривая все элементы этой строки в соответствии с условием (9.138), можно еще раз преобразовать матрицы $C^{(0)}$ и $T^{(0)}$ и вычеркнуть такие элементы, если они имеются. Проведем такое преобразование матриц $C^{(0)}$ и $T^{(0)}$. В результате получим

$$C^{(1)} = \begin{vmatrix} 3 & - & 2 & - \\ 4 & - & 1 & - \\ 5 & 9 & 6 & - \\ 6 & 10 & 7 & - \\ 7 & - & - & 1 \end{vmatrix};$$

$$T^{(1)} = \begin{vmatrix} 1,5 & - & 2 & - \\ 2 & - & 5 & - \\ 3 & 7 & 6 & - \\ 4 & 8 & 7 & - \\ 4 & - & - & 5 \end{vmatrix}.$$

Из матриц $C^{(1)}$ и $T^{(1)}$ видно, что исходные матрица значительно упростилась после проведения дополнительных преобразований. Перейдем теперь непосредственно к решению задачи на основе алгоритма ветвей и границ. Дерево решения с учетом построенных оценок (9.136) и (9.137) представлено на рис. 9.15.

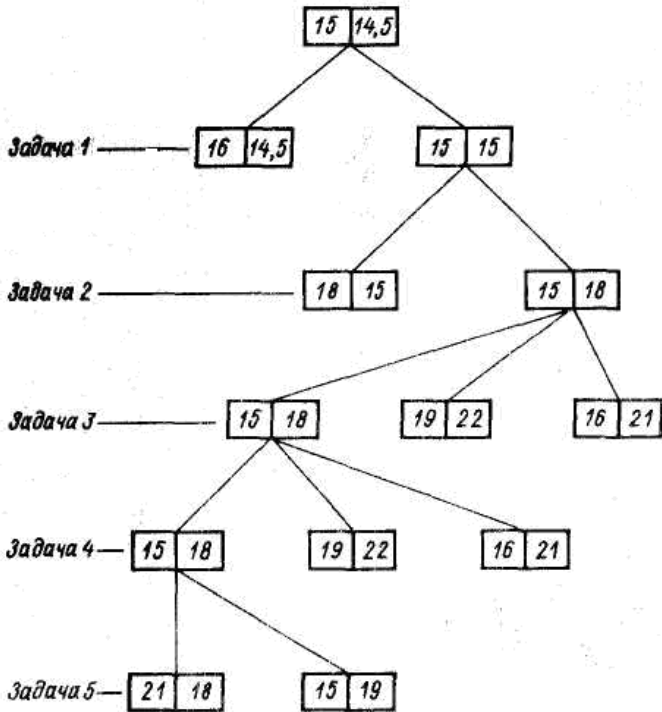


Рис. 9.15. Дерево решений с оценками (9.136) и (9.137)

В узлах дерева слева представлена оценка целевой функции, справа — оценка ограничения; зачеркнутые узлы, не удовлетворяющие ограничению. В результате решения получили следующее оптимальное распределение консультационных функций по модулям: 1—3, 2—3, 3—1, 4—1, 5—4 (функция — модуль). Жирной чертой на рис. 9.13 представлен путь поиска оптимального решения на дереве вариантов.

Из примера видно, что учет специфики задачи значительно повышает эффективность вычислительной процедуры метода «ветвей и границ».

В заключение следует отметить, что в ряде случаев структурный синтез хотя и может быть формализован в виде оптимальных задач дискретного программирования, однако он трудно разрешим в силу большой размерности задач, наличия большого числа случайных факторов, из-за векторного характера показателя эффективности. В таких ситуациях структурный синтез обычно понимают как выбор

структуры элементов и принципов их взаимодействия из некоторого числа рекомендуемых вариантов, перепективвших для дальнейшего использования. Предварительный выбор таких вариантов по совокупности свойств осуществляется группой квалифицированных консультантов, имеющих опыт консультирования подобных проблем, а детальное исследование количественных показателей основных свойств этих вариантов проводится для каждого из них в отдельности. После этого проводится сравнительный анализ и окончательный выбор основного варианта рекомендации по построению структуры системы. В этом случае будем говорить, что структурный синтез ведется на основе методов анализа.

9.15. Синтез комплекса технических средств САК

Задача выбора элементов КТС САК является наиболее важной при проектировании САК и их подсистем. В составе элементов КТС САК будем различать функционально законченные устройства, комплексы и подсистемы.

Задачу выбора элементов КТС САК сформулируем следующим образом. Пусть имеется совокупность $\{i\} = \{1, 2, \dots, n\}$ объектов одинакового функционального назначения, которые можно использовать для создания САК и ее подсистем. Каждый i -й объект имеет множество $\{k\} = \{1, 2, \dots, k\}$ характеристик. Известны требования к каждой k -й характеристике объекта. Необходимо из совокупности объектов выбрать такой, который бы наиболее полно удовлетворял предъявленным к нему требованиям.

Все характеристики того или иного объекта можно разделить на три вида.

Характеристики *первого вида* могут задаваться и определяться различными численными размерными или безразмерными физическими величинами, и чем больше численное значение характеристики объекта, тем он лучше при прочих равных условиях

Характеристики *второго вида* также измеряются численными физическими величинами, но чем меньше значение этой характеристики, тем лучше объект при прочих равных условиях (например, величиной точности позиционирования объекта, стоимость, потребляемая мощность, время формирования рекомендации, период окупаемости изделия и т. д.).

Требования к характеристикам первого и второго видов задаются в виде численных значений физических величин.

Характеристики *третьего вида* не измеряются численными значениями физических величин, а выражаются интегральной

возможностью реализации определенных материальных свойств и качеств данного объекта (например, возможность перевозки объекта в неработающем состоянии на тех или иных видах транспортных средств, демонтажа и повторного монтажа оборудования, функционирования изделия в подвижном состоянии, возможность удовлетворения той или иной конструктивной группе). Характеристики этого вида могут формулироваться с использованием ряда частных характеристик первого и второго видов, однако интегральная оценка общей характеристики выражается утверждением о наличии или отсутствии у рассматриваемого объекта определенного свойства или качества. Требования к характеристикам третьего вида задаются в виде словесной формулировки о необходимости реализации объектом того или иного свойства или качества. Оценка факта реализации требования объектом производится в виде логического утверждения: «Требование выполнено» или «Требование не выполнено».

Обозначим: K_1 , K_2 и K_3 — соответственно подмножества характеристик первого, второго и третьего видов. Численные значения k -й характеристики первого или второго вида для i -го объекта задаются величинами y_{ik} для всех $k \in K_1 \cup K_2$, а численные значения соответствующих требований — величиной Y_k .

Для характеристик третьего вида i -го изделия условно будем полагать, что если k -е требование выполняется объектом, то $y_{ik} = 1$, если не выполняется, то $y_{ik} = 0$ для всех $k \in K_3$. Само же k -е требование при формально-логическом описании задается в виде значения $y_k = 1$ для всех $k \in K$.

Величины y_{ik} и y_k , $k = 1, 2, \dots, k$ характеризуют условия выбора. Вектор условий выбора запишем в виде

$$Y = \| \| y_{11}, \dots, y_{nk}, y_1, \dots, y_k \| \| . \quad (9.139)$$

Разность $y_{ik} - y_k$ характеризует степень (полноту) выполнения (реализации) k -го требования i -м объектом. Отрицательные значения разностей говорят о том, что соответствующие характеристики i -го объекта меньше заданного требования на полученные разности, положительные — характеризуют превышение этого требования. Казалось бы

$$\sum_k (y_{ik} - y_k)$$

должна характеризовать суммарную полноту реализации i -м объектом заданных требований по всей их совокупности. Но это справедливо лишь в том случае, когда все требования имеют одинаковые размерности отражающих их физических величин, что на практике

практически не встречается. Обычно требования задаются физическими величинами различных измерений. Таким образом, указанная сумма не имеет физического смысла. Для того, чтобы характеризовать определенным образом выполнение совокупности требований i -м объектом, нужно эту сумму представить в безразмерных величинах или с одинаковой размерностью.

Пусть a_k — коэффициент приведения разностей $y_{ik} - y_k$ к одинаковой размерности. Тогда

$$\sum_k \alpha_k (y_{ik} - y_k)$$

в определенной мере характеризуют суммарную полноту реализации заданных требований, которое дает i -й объект. Однако и эта сумма не является исчерпывающей характеристикой, так как вычисляя ее, можно получить значение, равное нулю или близкое к нему, что характеризует факт полной реализации заданных требований, т. е. создается впечатление весьма благоприятной картины при выборе i -го объекта. На самом деле может оказаться, что ряд требований не выполняется, т. е. разности $y_{ik} - y_k$ имеют отрицательные для всех $k \in K_1$ и положительные для $k \in K_2$ значения, а ряд требований перевыполняется, и указанные разности имеют положительные для $k \in K_1$ и отрицательные для $k \in K_2$ значения. Во избежание указанного положения на величины y_{ik} для $k = K_1 \cup K_2$ должны быть наложены соответствующие ограничения.

Рассмотрим особенности задания ограничения к характеристикам первого и второго видов. Подмножество характеристик первого вида ($k \in K_1$) в свою очередь может содержать четыре подмножества.

Первое подмножество K_1^1 характерно тем, что требования к характеристикам этого подмножества задаются в виде ограничения снизу y_k .

Второе подмножество K_1^2 характеризуется тем, что требования к характеристикам этого подмножества задаются в виде ограничения снизу y_k и ограничения сверху A_k . Ограничение сверху вызвано тем, чтобы в ряде случаев при задании требований к наиболее важным характеристикам не допускать большой избыточности в реализации этих характеристик, что может повлечь за собой удорожание оборудования, увеличение габаритов, потребляемой мощности, усложнение и удорожание эксплуатации и т. д. Можно потребовать ограничить сверху производительность станка или требуемый диаметр обрабатываемой детали.

Третье подмножество характеристик K_1^3 , требования к которым задаются в виде ограничений снизу y_k , но при этом допускается его

снижение до величины B_k . Так, требуемая пропускная способность тракта накопителя деталей может быть задана величиной, например, $y_k \geq 200$ шт./ч, но допускается снижение ее до $B_k = 160$ шт./ч.

Четвертое подмножество характеристик K_1^4 , требования к которым задаются в виде ограничений снизу y_k и ограничений сверху, но при этом допускается снижение ограничения снизу до величины B_k . Например, мощность, расходуемая станком, задается величиной $y_k = 10$ кВт, но эта величина не должна превосходить значения $A_k = 15$ кВт. Однако допускается снижение мощности до значения $B_k = 8$ кВт.

Характеристики второго вида, как и первого, могут также разбиваться на четыре подмножества.

Первое подмножество характеристик K_1^2 , для которых требования задаются в виде ограничения сверху y_k . Например, вероятность брака обработанных деталей не должна превышать определенной наперед заданной величины.

Второе подмножество характеристик K_2^2 , требования к которым задаются в виде ограничений сверху y_k и ограничений B_k . Так же как и для характеристик первого вида, это связано с необходимостью избежать громоздких и дорогостоящих рекомендаций при существенно лучших характеристиках, чем требуемые. Например, требуемая вероятность потерь от брака не должна превышать величины $y_k = 10^{-2}$, но в то же время нет необходимости применения станков, обеспечивающих вероятность получения брака меньше $B_k = 10^{-3}$.

Третье подмножество характеристик K_2^3 , для которых требования задаются в виде ограничений сверху y_k , но при этом допускается его увеличение до A_k . К примеру, среднее время восстановления изделия ограничено величиной $y_k = 0,5$ г, однако может быть увеличено до значения $A_k = 0,6$ г.

Четвертое подмножество характеристик K_2^4 , требования к которым задаются в виде ограничений сверху y_k и ограничения снизу B_k , но при этом допускается его увеличение до значения A_k .

Если в примере для подмножества характеристик K_2^2 допустить возможность увеличения требуемой вероятности брака до значения $A_k = 10^{-1}$, то получим пример для подмножества характеристик K_2^4 . Второй пример для этого подмножества состоит в том, что требуемая величина среднего времени восстановления вышедшего из строя изделия ограничивается значением $y_k = 0,8$ ч, но при этом, исходя из функционального назначения рассматриваемого изделия, нет надобности стремиться к снижению этого времени ниже $B = 0,3$ ч, а при необходимости значение y_k можно увеличить до $A_k = 1$ ч. Величины A_k и B_k должны содержаться в ТЗ на разработку САК

С учетом указанных ограничений, представленная сумма с достаточной для лица, формирующего рекомендации степенью строгости характеризует суммарную полноту реализации заданных требований i -м модулем САК. Однако, для принятия более обоснованной рекомендации и выработки более аргументированного суждения о качествах i -й рекомендации целесообразно учитывать важность или вес g_k каждого k -го требования.

Тогда сумма

$$\sum g_k L_k (y_{ik} - y_k) \tag{9.140}$$

с учетом введенных ограничений наиболее полно характеризует суммарную полноту реализации заданных требований i -м модулем.

Если попытаться определить смысл введенных коэффициентов g_k и L_k , то обратимся к понятию качества функционирования данного комплекса, подсистемы или САК в целом, куда входят рассматриваемые модули, т. е. к качеству функционирования объекта оснащения САК.

Существует функциональная зависимость качества функционирования объекта оснащения САК, включающего объекты рассматриваемого функционального назначения, от величин y_{ik} :

$$M = M(y_{i1}, \dots, y_{i\bar{k}})$$

Откуда

$$M = (y_{i1}, \dots, y_{i\bar{k}}) - M(y_1, \dots, y_k) \approx \sum_{k=1, \bar{k}} \frac{\partial M(y_k)}{\partial y_k} (y_{jk} - y_k), \tag{9.141}$$

где $\partial M(y_k)/\partial y_k$ — частная производная $\partial M(y_{ik})/\partial y_{ik}$ в точке $y_{ik} = y_k$. Для всех $k = 1, 2, \dots, m$.

Сравнивая выражения (9.140) с (9.141), получаем

$$g_k = \frac{\partial M(y_k)}{\partial y_k} y_k; \quad L_k = 1/y_k. \tag{9.142}$$

Таким образом, все требования g_k представляют собой изменение качества функционирования комплекса, включающего рассматриваемый объект при отклонении k -й характеристики объекта на единицу, а коэффициент nk приводит разность $y_{ik} - y_k$ к относительным безразмерным величинам. Полученные выводы следуют и из чисто физических соображений.

Вектор параметров выбора представим в виде

$$X = \|x_1, \dots, x_n\|, \tag{9.143}$$

где его компоненты $x_i=1$, если из совокупности n объектов выбирается i -й объект, и $x_i=0$ — в противном случае. Тогда критерий

эффективности выбора с учетом приведенных выше рассуждений запишется в виде

$$F(X, Y) = \sum_{i=1, \bar{n}} \sum_{k=1, k} g_k \frac{y_{ik} - y_k}{y_k} x_i, \quad (9.144)$$

где g_k определяется по формуле (9.142). Эквивалентной записью критерия (9.144) является выражение

$$F(X, Y) = \sum_{i=1, \bar{n}} \sum_{k=1, k} g_k \frac{y_{ik}}{y_k} x_i. \quad (9.145)$$

Отношения y_{ik}/y_k , входящие в формулу (9.145), выражают также полноту реализации k -го требования i -м объектом, а

$$\sum_k g_k \frac{y_{ik}}{y_k}$$

характеризует суммарную реализацию всей совокупности требований i -м объектом с учетом веса каждого требования. Примем следующие ограничения на переменные x_i которые должны выполняться при использовании критериев (9.144) и (9.145). Ограничения на численные значения переменных:

$$\sum_{i=1, \bar{n}} x_i = 1. \quad (9.146)$$

Ограничения на пределы изменения величин y_{ik} для подмножеств характеристик следующие:

для K^1_1

$$\sum_{i=1, \bar{n}} y_{ik} x_i \geq y_k, \quad k \in K^1_1; \quad (9.147)$$

для K^2_1

$$\sum_{i=1, \bar{n}} y_{ik} x_i \geq y_k, \quad \sum_{i=1, \bar{n}} y_{ik} x_i \leq A_k, \quad k \in K^2_1; \quad (9.148)$$

для K^3_1

$$\sum_{i=1, \bar{n}} y_{ik} x_i \geq B_k, \quad k \in K^3_1; \quad (9.149)$$

для K^4_1

$$\sum_{i=1, \bar{n}} y_{ik} x_i \geq B_k, \quad \sum_{i=1, \bar{n}} y_{ik} x_i \leq A_k, \quad k \in K^4_1; \quad (9.150)$$

для K^1_2

$$\sum_{i=1, \bar{n}} y_{ik} x_i \leq y_k, \quad k \in K^1_2; \quad (9.151)$$

для K^2_2

$$\sum_{i=1, \bar{n}} y_{ik} x_i \leq y_k, \quad \sum_{i=1, \bar{n}} y_{ik} x_i \geq B_k, \quad k \in K^2_2; \quad (9.152)$$

для K^3_2

$$\sum_{i=1, \bar{n}} y_{ik} x_i \leq A_k, \quad k \in K^3_2; \quad (9.153)$$

для K^4_2

$$\sum_{i=1, \bar{n}} y_{ik} x_i \leq A_k, \quad \sum_{i=1, \bar{n}} y_{ik} x_i \geq B_k, \quad k \in K^4_2; \quad (9.154)$$

для подмножества характеристик $k \in K_3$ должно выполняться условие вида

$$\sum_{i=1, \bar{n}} y_{ik} x_i = 1. \quad (9.155)$$

Кроме того, требования к некоторым характеристикам могут быть заданы только в виде ограничений (9.147)—(9.155), а не входить непосредственно в выражение для критерия эффективности выбора. Это целесообразно делать в том случае если не представляется возможность установить количественную связь между качеством функционирования технического средства и характеристикой объекта, т. е. определить вес или важность данной характеристики. Такими характеристиками объекта могут быть, например, мощность потребляемой электроэнергии, занимаемая объектом площадь, его объем, весовые характеристики, стоимость, некоторые характеристики третьего вида ($k \in K_3$) и т. д. В этом случае в формулах (9.144) и (9.145) сумма по k распространяется только на те характеристики, которые вводятся в выражение для критерия эффективности выбора. Если обозначить через K_0 подмножество характеристик, которые входят в выражение для критерия эффективности выбора, формулы (9.144) и (9.145) для этого критерия соответственно принимает вид

$$F(X, Y) = \sum_{i=1, \bar{n}} \sum_{k \in K_0} g_k \frac{y_{ik} - y_k}{y_k} x_i; \quad (9.156)$$

$$F(X, Y) = \sum_{i=1, \bar{n}} \sum_{k \in K_0} g_k \frac{y_{ik}}{y_k} x_i. \quad (9.157)$$

Эта запись критериев является более общей.

Задача выбора объекта того или иного функционального назначения из совокупности n объектов сводится к отысканию такого решения системы (9.146)—(9.155) которое обращало бы в максимум критерии

эффективности выбора (9.156) либо (9.157), т. е. суммарную полноту выполнения требований с учетом веса каждого требования. Объекты, соответствующие решению этой задачи, наиболее полно удовлетворяют поставленным требованиям.

Эквивалентной формой записи ограничений на компоненты вектора условий выбора (9.139) являются ограничения на величины отклонения характеристик y_{ik} от задаваемых требований y_k . Для рассматриваемого случая эти ограничения (вместо 9.147)— (9.155) записываются в виде:

$$\sum_{i=1, \bar{n}} (y_{ik} - y_k) x_i \geq 0, \quad k \in K_1; \quad (9.158)$$

$$\sum_{i=1, \bar{n}} (y_{ik} - y_k) x_i \geq 0, \quad \sum_{i=1, \bar{n}} (y_{ik} - A_k) x_i \leq 0, \quad k \in K_1^2; \quad (9.159)$$

$$\sum_{i=1, \bar{n}} (y_{ik} - B_k) x_i \geq 0, \quad k \in K_1^3; \quad (9.160)$$

$$\sum_{i=1, \bar{n}} (y_{ik} - B_k) x_i \geq 0, \quad \sum_{i=1, \bar{n}} (y_{ik} - A_k) x_i \leq 0, \quad k \in K_1^4; \quad (9.161)$$

$$\sum_{i=1, \bar{n}} (y_{ik} - y_k) x_i \leq 0, \quad k \in K_2^1; \quad (9.162)$$

$$\sum_{i=1, \bar{n}} (y_{ik} - y_k) x_i \leq 0, \quad \sum_{i=1, \bar{n}} (y_{ik} - B_k) x_i \geq 0, \quad k \in K_2^2; \quad (9.163)$$

$$\sum_{i=1, \bar{n}} (y_{ik} - A_k) x_i \leq 0, \quad k \in K_2^3; \quad (9.164)$$

$$\sum_{i=1, \bar{n}} (y_{ik} - A_k) x_i \leq 0, \quad \sum_{i=1, \bar{n}} (y_{ik} - B_k) x_i \geq 0, \quad k \in K_2^4; \quad (9.165)$$

$$\sum_{i=1, \bar{n}} (y_{ik} - y_k) x_i = 0, \quad k \in K_3. \quad (9.166)$$

Второй эквивалентной формой записи ограничений на y_{ik} и y_k являются ограничения на относительные отклонения характеристик от задаваемых требований.

При этом в рассматриваемом случае получим ограничения:

$$\sum_{i=1, \bar{n}} \frac{y_{ik}}{y_k} x_i \geq 1, \quad k \in K_1; \quad (9.167)$$

$$\sum_{i=1, \bar{n}} \frac{y_{ik}}{y_k} x_i \geq 1, \quad \sum_{i=1, \bar{n}} \frac{y_{ik}}{y_k} x_i \leq \bar{A}_k, \quad k \in K_1^2, \quad (9.168)$$

где $A_k = A_k / y_k$;

$$\sum_{i=1, \bar{n}} \frac{y_{ik}}{y_k} x_i \geq \bar{B}_k, \quad k \in K_1^2, \quad (9.169)$$

где $B_k = B_k / y_k$;

$$\sum_{i=1, \bar{n}} \frac{y_{ik}}{y_k} x_i \geq \bar{B}_k, \quad \sum_{i=1, \bar{n}} \frac{y_{ik}}{y_k} x_i \leq \bar{A}_k, \quad k \in K_1^4; \quad (9.170)$$

$$\sum_{i=1, \bar{n}} \frac{y_{ik}}{y_k} x_i \leq 1, \quad k \in K_2^1; \quad (9.171)$$

$$\sum_{i=1, \bar{n}} \frac{y_{ik}}{y_k} x_i \leq 1, \quad \sum_{i=1, \bar{n}} \frac{y_{ik}}{y_k} x_i \geq \bar{B}_k, \quad k \in K_2^2; \quad (9.172)$$

$$\sum_{i=1, \bar{n}} \frac{y_{ik}}{y_k} x_i \leq \bar{A}_k, \quad k \in K_2^3; \quad (9.173)$$

$$\sum_{i=1, \bar{n}} \frac{y_{ik}}{y_k} x_i \leq \bar{A}_k,$$

$$\sum_{i=1, \bar{n}} \frac{y_{ik}}{y_k} x_i \geq \bar{B}_k, \quad k \in K_2^4; \quad (9.174)$$

$$\sum_{i=1, \bar{n}} \frac{y_{ik}}{y_i} x_i = 1, \quad k \in K_8. \quad (9.175)$$

Сформулированные выше задачи записаны для детерминированного случая, когда вектор условий выбора (9.139) не является случайной величиной. Это наиболее простой случай.

Рассмотрим еще три случая, когда величины y_{ik} и y_k в отдельности или одновременно могут быть случайными. При этом случайные величины y_{ik} и y_k заменяются их математическими ожиданиями a_{ik} и b_k .

Необходимо отметить, что случайные величины y_{ik} и y_k в полученных моделях могут быть как независимыми, так и зависимыми. В большинстве случаев численные значения как характеристик y_{ik} , так и требований y_k , являются взаимно независимыми для различных номеров характеристик и требований (различных значений k). Если же ряд требований и характеристик является функционально связанными (зависимыми) величинами, то из последовательности этих зависимых величин выделяется последовательность независимых величин, которые и вводятся в условия выбора. Например, коэффициент

готовности, средняя наработка на отказ и среднее время восстановления объекта есть последовательность зависимых величин, а любая пара из них является независимыми величинами. Следовательно, в условиях выбора должны вводиться только две из перечисленных выше характеристик и требований к ним. Однако, даже в том случае когда все же случайные величины y_{ik} и y_k по тем или иным причинам являются зависимыми, то целесообразно прибегать к допущению об их независимости, так как в противном случае получение выражений для критериев и некоторых видов ограничений может вызвать существенные трудности.

Приведенная выше совокупность моделей выбора содержит модели двух классов. **Модели первого класса**, в которых компоненты y_{ik} и y_k вектора условий выбора Y являются детерминированными величинами, соответствуют детерминированному принципу выбора в условиях определенности. **Модели второго класса**, в которых компоненты y_{ik} или y_k , либо обе вместе являются случайными величинами с известными распределениями вероятностей, соответствуют вероятностному принципу выбора в условиях риска.

В обоих случаях весьма существенным является нахождение подходящей формы критериев в эффективности выбора.

Форма записи критериев эффективности выбора (9.156)—(9.157), является универсальной и наиболее общей. В некоторых случаях можно предложить более удобную запись. Например, как следует из выражений (9.140)—(9.141), более естественной записью критерия эффективности (9.156) является

$$F(X, Y) = \sum_{i=1, \bar{n}} \sum_{k \in K_0} g_k (y_{ik} - y_k) x_i,$$

где $g_k = \partial M(y_{11}, \dots, y_{nk0}) / \partial y_{ik}$ в точках $y_{ik} = y_k$ для всех $k \in K_0$, а $\alpha_k = 1$.

В этом случае размерность критерия $F(X, Y)$ совпадает с размерностью функции $M(y_{11}, \dots, y_{nk0})$. Однако такая запись была бы оправдана, если бы всегда представлялась возможность получить функцию $M(y_{11}, \dots, y_{nk0})$. Но в ряде случаев функциональную зависимость качества функционирования объекта консультирования от его характеристик получить не представляется возможным. Тогда будем прибегать к некоторым способам определения весов в виде безразмерных величин и нормированию значений y_{ik} или $y_{ik} - y_k$, т. е. представление в виде y_{ik} / y_k или $(y_{ik} - y_k) / y_k$ является необходимым. Отсюда следует, что форма записи критериев эффективности выбора, приведенная в предыдущем пункте, является наиболее общей и позволяет унифицировать вычислительные процедуры при любых способах определения весов характеристик g_k .

Рассмотрим некоторые приемы определения весов g_k при отсутствии зависимости между характеристиками объекта и качеством его функционирования.

Одним из распространенных приемов является определение весов характеристик с помощью экспертов или, как принято говорить, с помощью *экспертных оценок*.

Приведенные выше рассуждения позволяют сформулировать четыре характерных случая вынесения экспертных оценок.

1. Экспертные оценки при разнородном составе экспертов, т. е., когда эксперты различны по своей квалификации и степени объективности. В этом случае веса характеристик определяются из соотношения

$$g_k = \frac{\sum_{\beta=1, \bar{\beta}} W_{k\beta} q_{k\beta}}{\sum_{\beta=1, \bar{\beta}} W_{k\beta} q_{k\beta}}, \quad (9.176)$$

где $\bar{\beta}$ — число экспертов, участвующих в оценке изделий; $W_{k\beta}$ — коэффициент, выражающий количественно степень квалификации β -го эксперта ($\beta = 1, \bar{\beta}$) при оценке k -й характеристики, $0 \leq W_{k\beta} \leq 1$ (чем выше квалификация эксперта, тем больше значение коэффициента $W_{k\beta}$); $q_{k\beta}$ — коэффициент, отражающий степень объективности β -го эксперта, при оценке k -й характеристики; $q_{k\beta}$ — вес k -й характеристики, назначаемой β -м экспертом.

Веса характеристик целесообразно назначать в интервале от нуля до некоторого максимального значения, например в интервалах (0; 1), (0; 10), (0; 100), хотя верхняя граница интервала не обязательно должна быть единицей или кратной десяти. В принципе она может быть любым положительным числом. Однако, удобнее всего работать в указанных выше интервалах и предпочтительнее в интервале (0; 1). Что касается оценки степени объективности экспертов, то вопрос определения численного значения коэффициента $q_{k\beta}$ решается более сложно, чем простое приписывание тому или иному эксперту определенного значения этого коэффициента. Эксперт при проведении оценки характеристик изделия по тем или иным причинам может обнаружить тенденции к завышению весов одних характеристик и к занижению весов других характеристик. Поэтому с учетом упомянутых тенденций можно предложить следующую формулу для расчета величины $q_{k\beta}$:

$$q_{k\beta} = \begin{cases} \bar{q}_{k\beta} & \text{при тенденции к завышению;} \\ \min \left\{ \frac{1}{\bar{q}_{k\beta}}, \frac{q_{\max}}{q_{k\beta}} \right\} & \text{при тенденции к занижению,} \end{cases} \quad (9.177)$$

где $q_{k\beta}$ — некоторый коэффициент ($0 \leq q_{k\beta} \leq 1$), характеризующий степень объективности β -го эксперта (чем больше значение $q_{k\beta}$, тем больше объективность β -го эксперта при оценке k -й характеристики); q_{\max} — верхняя граница интервала, на котором оцениваются веса характеристик объекта.

Например, пусть веса характеристик оцениваются в интервале $(0; 10)$, т. е. $q_{\max} = 10$ и есть основания полагать, что β_1 -й эксперт попытался зависить k -ю характеристику, но не очень сильно, и ему приписывается значение $\bar{q}_{k\beta_1} = 0,8$, а β_2 -й эксперт обнаруживает усиленную тенденцию к занижению этой характеристики и ему приписывается значение $\bar{q}_{k\beta_2} = 0,4$. Первый эксперт дает $q_{k\beta_1} = 9$, а второй полагает $q_{k\beta_2} = 3$. Тогда по формуле (9.176) находим:

$$\bar{q}_{k\beta_1} = \bar{q}_{k\beta_2} = 0,8; \bar{q}_{k\beta_2} = \min \{1/0,4; 0/3\} = \min \{2,5; 3,33\} = 2,5$$

и с учетом степени объективности экспертов получим

$$q_{k\beta_1} g_{k\beta_1} = 0,8 \cdot 9 = 7,2; \quad q_{k\beta_2} g_{k\beta_2} = 2,5 \cdot 3 = 7,5.$$

2. Экспертные оценки при разнородном составе экспертов. Разнородность характеризуется тем, что эксперты различны по своей квалификации, но одинаковы по степени объективности.

Тогда вес k -й характеристики определим по формуле:

$$g_k = \frac{\sum_{\beta=1, \beta} W_{k\beta} g_{k\beta}}{\sum_{\beta=1, \beta} W_{k\beta}}. \tag{9.178}$$

Этот случай может также иметь место тогда, когда эксперты хотя и различны по своей объективности, но степень объективности каждого установить не представляется возможным.

3. Экспертные оценки при разнородном составе экспертов. Разнородность характеризуется тем, что эксперты различны по степени объективности, но одинаковы по своей квалификации. В этом случае:

$$g_k = \frac{\sum_{\beta=1, \beta} g_{k\beta} q_{k\beta}}{\sum_{\beta=1, \beta} q_{k\beta}}. \tag{9.179}$$

По этой формуле может производиться расчет весов характеристик и тогда, когда эксперты хотя и различны по своей квалификации, но коэффициенты $W_{k\beta}$ определить не представляется возможным.

4. Экспертные оценки при однородном составе экспертов, т. е., когда эксперты одинаковы по своей квалификации и степени объективности. Тогда веса характеристик g_k определяют по формуле:

$$g_k = \left(\sum_{\beta=1, \bar{\beta}} g_{k\beta} \right) / \bar{\beta}. \quad (9.180)$$

Аналогичным образом можно поступать и тогда, когда состав экспертов различен по квалификации и объективности, но приписать экспертам значения $W_{k\beta}$ и $q_{k\beta}$ не представляется возможным; состав экспертов различен по квалификации и одинаков по объективности, но определение $W_{k\beta}$ затруднительно или невозможно по объективности, определение $q_{k\beta}$ затруднительно или невозможно; состав экспертов одинаков по квалификации и различен по объективности, но определить $q_{k\beta}$ не представляется возможным.

10. Формирование и функционирование консультационных модулей

Важным фактором ускорения научно-технического прогресса является создание и широкое внедрение в сферу производственной и социальной деятельности автоматизированных систем, в том числе систем автоматизированного консультирования (САК).

Развитие САК требует решения комплекса сложных научно-технических проблем, связанных как собственно с разработкой систем, предназначенных для консультирования конкретных проблем, так и с применением этих систем для решения практических задач в рамках целостной технологии автоматизированного проектирования. По мнению автора (и это подтверждается отечественным и мировым опытом), необходимым условием широкого и эффективного внедрения методов и средств автоматизации в практику консультирования является простота и доступность их использования консультантами. Исходя из этой предпосылки в настоящей работе предлагается и обосновывается новый подход к организации автоматизированного консультирования сложных проблем. В основе этого подхода лежит рассмотрение САК как организационно-технической системы (ОТС), представляющей собой структурированное объединение всех видов ресурсов, используемых при автоматизированном консультировании. Процесс автоматизированного консультирования при этом состоит из многократно повторяющихся процедур формирования и функционирования консультационных модулей (КМ).

Понятие «консультационный модуль», впервые введенное автором, означает автоматизированную систему, предназначенную для выполнения отдельной консультационных операции, и отражает объективно существующую необходимость декомпозиции процесса

консультирования. Каждый модуль гибко формируется на базе ресурсов САК под конкретную консультационную операцию, а их интеграция с учетом информационной и организационной связности обеспечивает реализацию процесса целостного консультирования сложных проблем.

Реализация такого подхода потребовала разработки достаточно сложных в теоретическом отношении методов и алгоритмов. Однако они построены так, что все трудоемкие операции, связанные с формированием и функционированием консультационных модулей, скрыты от консультанта, на долю которого остаются функции, лежащие в сфере его профессиональных интересов.

10.1. Формирование консультационных модулей

10.1.1. Базовая форма представления математических моделей консультируемых проблем в САК

10.1.1.1. Содержание процедур формирования консультационных модулей

Основная особенность формирования КМ состоит в том, что данная процедура проводится на базе предварительно сформированной САК и состоит в целенаправленном выборе соответствующих элементов из множеств членения компонент этой системы. Выбор элементов производится из условия их соответствия конкретной консультационной операции, выполнение которой возлагается на формируемый КМ. Выбранные элементы «замещают» соответствующие компоненты ОТС, образуя функционально завершённую автоматизированную систему, ориентированную на выполнение заданной консультационной операции.

Предполагается, что в составе исходной САК достаточно средств для того, чтобы на их основе была выполнена каждая консультационная операция. В случае недостаточности имеющихся средств в процессе формирования КМ должна быть получена информация о необходимых дополнениях базовой системы. Эта информация передается в другие, функционально-ориентированные на этот вид работ, модули разрабатываемой ОТС. В роли таких модулей могут выступать, в частности, АСНИ, АСО и т. п., на базе которых должно быть проведено развитие или модификация используемой САК.

Формирование КМ представляется состоящим из процедур двух видов. Процедуры первого вида непосредственно связаны с построением функционально завершенной автоматизированной системы, способной выполнить заданную консультационную операцию.

Присутствие процедур второго вида вызвано тем, что результаты выполняемых в процессе консультирования операций могут находиться во взаимозависимости, предопределяющей взаимосвязи между соответствующими автономно функционирующими КМ. В дальнейшем будем говорить, что КМ, результаты автономного функционирования которых влияют друг на друга, находятся в информационной взаимосвязи.

Сушественно важным при определении информационных взаимосвязей между КМ является тот факт, что такого рода связи в процессе консультирования не являются статическими, а могут появляться, исчезать или менять свое содержание в процессе консультирования. Такие вариации связаны, в частности, с изменениями в процессе функционирования КМ физических принципов действия агрегатов консультируемой проблемы, определение параметров которых является задачей, возлагаемой на рассматриваемые модули.

Наличие информационной связи между КМ порождает, как правило, необходимость согласования результатов их автономного функционирования (согласованию не подлежит информация, передаваемая из КМ, в состав которого входит ЛФР, занимающее главенствующее место в организационной структуре по сравнению с ЛФР,ом, входящим в состав ПМ, принимающем информацию). Методика и организационные формы процедур согласования определяются положением, занимаемым ЛФР,ами, входящими в состав КМ, результаты автономного выполнения которых должны быть согласованы.

В итоге процедуры второго вида, выполняемые при формировании КМ, состоят в «отслеживании» информационных связей этого модуля с другими КМ, а также определении методического и организационного обеспечений, необходимых для реализации сходящихся процедур согласования. Эти виды обеспечений определяются на основе сопоставления информационных и организационных связей между КМ и составляют содержание соответствующих компонентов для взаимодействующих консультационных модулей.

При формировании КМ предполагается, что консультационная задача, решение которой возлагается на рассматриваемый модуль, а

также ЛФР, несущее ответственность за результаты ее решения, известны. В частности, задача формулируется вышестоящим ЛФР,ом и является для формируемого модуля директивной информацией. ЛФР определяется, как правило, одновременно с формулировкой задачи. В результате можно считать, что формированию в каждом конкретном случае подлежат компоненты «информационная среда» и «технические средства».

Одним из основных условий, налагаемых на формирование КМ, является то, что оно должно производиться гибко без ограничений как на исходную постановку сформулированной консультационной задачи, так и на методы и методики ее решения. Данное условие может быть выполнено лишь в случае формализации процесса формирования КМ и создания инвариантных средств управления этим процессом.

В качестве инвариантной составляющей САК, управляющей процессом формирования КМ, рассматривается управляющий модуль (УМ) системы. Его структуру (рис. 10.1) и назначение определим следующим образом.

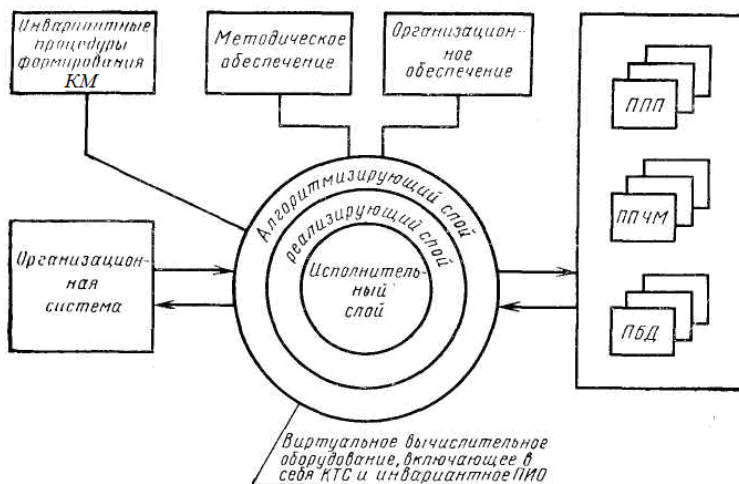


Рис. 10.1. Структура управляющего модуля САК

Как уже отмечалось, в основе автоматизированного консультирования лежит использование фундаментальных знаний и опыта, накопленного в области консультируемой проблемы, которые имеют «машинную» форму представления. Основными формами такого их представления являются:

- математические модели класса проблем, к которым относится консультируемая проблема. Наиболее распространенной формой их представления в САК являются пакеты прикладных программ (ППП);

- конкретные данные, полученные в результате проведения различного рода исследований и экспериментов. В САК они представляются в виде прикладных баз данных (ПБД);

- методы, методики и алгоритмы решения задач различного рода, которые представляются в виде пакетов программ, называемых далее пакетами программ численных методов (ППЧМ).

Вычислительную систему (совокупность технических средств, программного и информационного обеспечений), на базе которой функционирует САК, разделим на две составляющие. К первой составляющей, называемой проблемно-ориентированной, отнесем ППП, ППЧМ, ПБД, а ко второй — оставшуюся часть и представим ее в виде виртуальных вычислительных средств.

Процесс решения консультационных задач на базе представленных виртуальных вычислительных средств, охватывающих как формирование, так и функционирование КМ, в общем случае может быть разделен на три этапа.

На первом из них для каждой рассматриваемой задачи формируется с учетом данных ПБД соответствующая логика обработки программных модулей из ППП и ППЧМ. Этот этап адекватен проектированию соответствующего рассматриваемой операции консультационного модуля.

На втором этапе, адекватном «изготовлению» спроектированного КМ, формируются соответствующие вычислительные схемы по обработке программных модулей из ППП и ППЧМ по ранее определенной логике.

На третьем этапе сформированные вычислительные схемы реализуются, что соответствует отдельному такту функционирования сформированного КМ.

Основываясь на вышеизложенном, выделим в составе базового вычислительного средства САК три слоя: алгоритмизирующий, реализующий и исполнительный. При этом алгоритмизирующий и реализующий слои определяют собственное программное и информационное обеспечения УМ.

Исполнительный слой также входит в состав УМ и отражает компоненты «технические средства» и «стандартное программное обеспечение», на базе которых функционирует САК.

Данное представление вычислительной системы САК позволяет при определении УМ не конкретизировать используемые технические

средства, а также соответствующее стандартное программное обеспечение вне зависимости от их реализации (на базе одной ЭВМ, распределенной системы или сети ЭВМ). Поэтому в дальнейшем при определении УМ ограничимся рассмотрением программного и информационного обеспечений, соответствующих алгоритмизирующему и реализующему слоям.

Основную информацию, на базе которой функционирует УМ, формируя требуемые КМ, составляют: математическая модель, описывающая консультируемую проблему; методы и методики решения различного рода консультационных задач на базе математических моделей; данные различного характера, полученные к текущему моменту времени и отражающие директивную и нормативно-справочную информацию. При этом математическая модель играет центральную роль и используется: во-первых, для обеспечения непротиворечивости рекомендаций, формируемых в процессе консультирования, а во-вторых, совместно с методами и методиками — для проведения расчетных работ в процессе анализа сформированных рекомендаций и выбора из них наиболее обоснованного варианта.

Рассмотрим сказанное подробнее.

Так как в рамках предлагаемого подхода предполагается, что математическая модель, описывающая консультируемую проблему, считается известной, то все действия, выполняемые в процессе консультирования, должны быть согласованными с имеющейся моделью. В частности, если в составе модели имеются связи $a=f_1(b, c)$, $l=f_2(c)$, то это означает, что взаимонезависимое задание параметров консультируемой проблемы, описываемых переменными a, b, c, l , недопустимо. Обратное, если для консультируемой проблемы известен лишь параметр, описываемый переменной c , а возможность определения значений переменной b отсутствует, то значения переменной a неопределимы и, соответственно, консультационная операция, для выполнения которой существенно знание этого значения, невыполнима.

Рассмотренные два факта могут быть сопоставлены с реальными консультационными ситуациями. Вначале предположим, что в задании на выполнение консультационной операции в качестве директивных данных присутствуют значения полученных независимо параметров, описываемых переменными a, b, c, l . Если эти параметры носят непрерывный характер, то вероятность того, что их значения непротиворечивы, равна нулю. Консультационная операция в таких условиях, очевидно, невыполнима. Если же значения тех же

параметров не фиксированы, а ограничены, то требуется провести сопоставление диапазонов их возможных значений, и если они непротиворечивы, то наиболее «слабые» ограничения могут быть выведены из рассмотрения, «освободив» тем самым соответствующие параметры.

Теперь предположим, что в процессе функционирования одного КМ были определены параметры, описываемые переменными a и b , а другого КМ — параметр описываемый переменной c . Данное обстоятельство с учетом имеющейся связи ($a = f(b, c)$) свидетельствует о том, что результаты функционирования рассматриваемых КМ являются взаимозависимыми. Это является признаком наличия информационной связи между этими КМ и необходимости согласования результатов их автономного функционирования.

Далее, на основе такого же рода анализа взаимозависимости различных параметров, описывающих консультационную операцию (критериев, фиксированных и определяемых параметров), можно определить требуемый способ ее выполнения. Так, если в качестве директивных данных фигурируют параметры, соответствующие переменным b и c , а требующие определения параметры соответствуют переменным a , l , то основным содержанием КМ в таком случае является выполнение расчетов на базе математической модели.

Если же значения требуемых параметров на основе математической модели не могут быть определены расчетным путем, то их определение возможно лишь на основе оптимизационных процедур. При этом для такого их определения необходимо, чтобы используемый критерий мог быть вычислен на основе имеющейся модели по значениям выбираемых параметров и известным к текущему моменту времени значениям фиксированных (директивных) параметров.

Приведенные простейшие примеры показывают, что на основе анализа возможности определения значений одних переменных математической модели через другие возможно выполнение таких возлагаемых на управляющий модуль САК функций, как выявление информационных связей между различными КМ, а также анализ корректности задания на формирование требуемого КМ.

Непосредственное формирование КМ в рассматриваемом подходе состоит в гибком и оперативном формировании требуемых программно и информационно обеспечений. Это трактуется как выбор из ППП (программной реализации математической модели), ППЧМ (программной реализации методов и методик решения консультационных задач) и ПБД (различного рода данных, связанных с консультируемой проблемой) их определенных элементов,

составляющих целостную информационную среду, адекватную консультационной задаче, возлагаемой на формируемый КМ. При этом полагается, что выбор методов и методик осуществляется ЛФР,ом, а формирование требуемых моделей как «цепочек» из модулей ППП может осуществляться УМ на основе формальных алгоритмов или ЛФРом.

Формирование КМ возлагается на определенный выше алгоритмирующий слой УМ. В основу структуризации данного слоя положен принцип, согласно которому каждый элемент, далее называемый алгоритмирующим, взаимнооднозначно соответствует определенному классу задач, решаемых в САК. Такого рода структуризация базируется на последовательной декомпозиции процесса консультирования на иерархическую совокупность инвариантных действий (задач) и постановке каждому из этих действий в соответствие того или иного алгоритмирующего элемента, на базе которого оно должно выполняться.

В общем случае все проблемы, решаемые в процессе консультирования, могут быть разделены на задачи, имеющие:

- отличия в их математической постановке;
- общую математическую постановку, но отличающиеся методиками (методами) их решения;
- общую математическую постановку, при решении которых используется единая методика, но отличающиеся в своей предметной постановке, определяемой математической моделью конкретно рассматриваемой подпроблемы консультируемой проблемы.

Тогда, обозначив некоторую конкретную консультационную задачу через Z_α , ее можно представить в виде:

$$Z_\alpha = \langle H_\alpha, A_\alpha, M_\alpha, D_\alpha \rangle,$$

где H_α — математическая постановка рассматриваемой задачи; A_α — методика или метод ее решения; M_α — используемая математическая модель; D_α — исходные данные и результаты.

Методика решения той или иной задачи представляет собой логику выполнения некоторых действий, которые, в свою очередь, могут рассматриваться как задачи следующего уровня иерархии. Обозначим множество индексов этих задач, порожденных методикой A_α , через n_α . Тогда можно записать

$$A_\alpha \Rightarrow \{Z_\beta\}_{\beta \in n_\alpha}.$$

При этом сама методика должна соответствовать математической постановке решаемой задачи:

$$H_\alpha \Rightarrow A_\alpha.$$

Для решения задач в одной и той же постановке (H_α) можно использовать различные методики, совокупность которых обозначим

$$\mathfrak{M}_\alpha = \mathfrak{M}_\alpha (H_\alpha):$$

$$\mathfrak{M}_\alpha = \{A_\alpha : H_\alpha \Rightarrow A_\alpha\}.$$

Выбранная из \mathfrak{M}_α конкретная методика A_α определяет как задачи следующего уровня, требующие решения при ее реализации:

$$n_\alpha = n_\alpha(A_\alpha),$$

так и необходимые математические модели:

$$M_\alpha = M_\alpha(A_\alpha).$$

Требуемой модели в составе имеющихся может не оказаться. В этом случае ее необходимо агрегировать из элементарных моделей (модулей ППП). Задание на агрегирование требуемой для выбранной методики математической модели будем представлять парой векторов — входных и выходных переменных (I_α, T_α):

$$A_\alpha \Rightarrow \{I_\alpha, T_\alpha\}.$$

В итоге, процесс формирования КМ по предложенной схеме сводится к следующим операциям (рис. 10.2):

- определению, к какому классу формальных постановок относится решаемая задача (H_α);

- определению множества допустимых методик ($\mathfrak{M}(H_\alpha)$): выбору из множества допустимых методик некоторой конкретной методики ($A_\alpha \in \mathfrak{M}_\alpha$);

- формированию математической модели, требуемой для реализации выбранной методики ($M_\alpha(I_\alpha, T_\alpha)$);

- определению взаимосвязанной совокупности задач $\{Z_\beta\}_{\beta \in n_\alpha}$, формулирование каждой из которых сводится к повторению указанной процедуры.

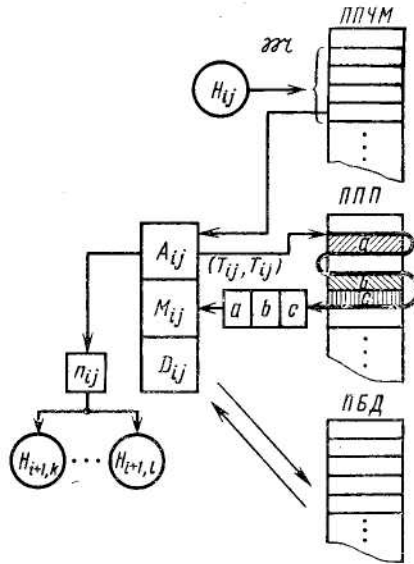


Рис. 10.2. Схема формирования алгоритмирующих элементов

Описанная процедура предполагает, что каждая из рассматриваемых консультационных задач представляется в виде некоторой иерархической совокупности задач, состав которых на каждом уровне определяется выбираемыми методиками решения задач предшествующих уровней. Обратим внимание на то, что сказанное предполагает отсутствие циклических связей в получаемой структуре задач. Данное условие может быть выполнено всегда, в частности, путем представления совокупности задач, образующих циклическую структуру, как некоторой единой задачи, на следующем уровне за которой располагаются задачи из рассматриваемой совокупности. При этом целью методики решения такого рода задач является, как правило, организация некоторого сходящегося итерационного процесса.

В итоге структура алгоритмирующего слоя представляет собой ациклический альтернативный граф, где вершины типа И могут рассматриваться как некоторые шаблоны с ячейками для: модулей ППЧМ; моделей, агрегированных из модулей ППП; информационного обеспечения. При этом процесс алгоритмизации решения той или иной консультационной задачи, т. е. формирования алгоритмирующих элементов сводится к заполнению указанных шаблонов,

производимому ЛФР,ами, в части, выбора методик и формальными операторами при агрегировании требуемой математической модели.

Представленная структура УМ, где каждому из элементов ставится в соответствие определенный класс консультационных процедур в их инвариантной постановке, позволяет определить полноту САК классом задач, допускающих формирование рекомендаций на их основе. Пополнение этого класса задач не ограничивается и связано с введением в состав УМ дополнительных алгоритмизирующих элементов. Их введение в существующую структуру означает, что связанные с ними элементы:

- более высоких уровней получают возможность решения возлагаемых на них задач в более широком спектре методов;
- более низких уровней определяют тот класс методов на базе которых могут решаться задачи, возлагаемые на вновь вводимые элементы.

Таким образом, представленный подход к формированию КМ обеспечивает возможность свободного расширения класса консультационных задач, решаемых САК, без ограничений на используемые при этом методы и методики, выбираемые в каждом конкретном случае ЛФР,ами.

Итак, формирование КМ основывается на наличии в составе САК прикладного программного и информационного обеспечений, представленных в виде ППП, ППЧМ и ПБД. Обязательным условием такого формирования, возлагаемого на УМ, является то, что оно должно осуществляться ЛФР,ом, работающим вначале в составе УМ, а после того, как КМ будет сформирован, — в его составе при обосновании сформированных рекомендаций.

В таком контексте работа ЛФР,а в составе как консультационного, так и управляющего модулей должна носить естественный характер и основываться на оперировании консультационными понятиями, связанными с консультируемой проблемой. Это возможно лишь при условии формализации основных, требующих выполнения в процессе формирования и функционирования КМ, процедур, их алгоритмизации и возложения на средства автоматизации, оставляя за ЛФР,ом функции задания и изменения исходных для алгоритмов данных, которые должны быть связаны с привычными для него понятиями.

Другими словами, в составе САК должны присутствовать средства, обеспечивающие гибкое и оперативное формирование как КМ, так и наиболее эффективных методик выбора (формирования) непротиворечивых и обоснованных рекомендаций. Данные средства должны играть роль «посредника» между ЛФР,ами и вычислительной

системой, в составе которой находятся математические модели (ППП), методы и методики (ППЧМ) и различного рода данные (ПБД), рассматриваемые как накопленный опыт и знания в области консультирования. Основным принцип функционирования этих средств должен основываться на выполнении следующей последовательности действий:

- сведения процедур формирования и функционирования КМ в соответствии с заданием, полученным от ЛФР,а, к совокупности формальных задач, решаемых на основе имеющихся ППП; ППЧМ и ПБД;

- решении этих формальных задач;

- интерпретации результатов решения формальных задач, как консультационной информации, сообщаемой ЛФР,ам.

Реализация названного принципа делает необходимым определение основных процедур, требующих выполнения в процессе формирования и функционирования КМ. Это определение базируется на формальном анализе процесса консультирования, который был рассмотрен в разделе 4 (см. книгу 1 настоящей работы).

10.1.1.2. Основные компоненты математических моделей

Согласно предлагаемому подходу математическая модель консультируемой проблемы является основной исходной информацией, на базе которой производится формирование консультационных модулей. При этом, учитывая инвариантный характер выявленных процедур формирования, для разработки методов и алгоритмов их реализации математическая модель должна быть представлена в формализованном виде, отражающем тот факт, что каждая конкретная модель представляется в САК в виде ППП модульной структуры.

Под математической моделью понимают совокупность переменных, описывающих консультируемую проблему и связей между этими переменными .

Под переменными математической модели понимают числа, векторы, функции, с помощью которых определяется моделируемая консультируемая проблема. Набор переменных модели будем обозначать через $P = \{p^1, p^2, \dots, p^{N_p}\}$, а каждую конкретную реализацию модели — задавать N_p -мерным вектором $\langle P \rangle = \{\langle p^1 \rangle, \dots, \langle p^{N_p} \rangle\}$ (где N_p — общее число переменных; $\langle a \rangle$ — значения компонент вектора a).

Под связями математической модели понимают совокупности отношений, связывающих значения отдельных групп переменных

модели. Основными видами связей, используемых в математических моделях, являются *отображение и отношение порядка*.

Отображения описывают некоторую совокупность свойств моделируемой консультируемой проблемы и ее взаимодействия с внешней средой. Они реализуются оператором, обозначаемым в дальнейшем через Φ :

$$\Phi: \chi \Rightarrow \lambda(\chi \mathbf{U} \lambda = P). \quad (10.1)$$

При этом переменные χ называют аргументами или входами модели, а λ — ее значениями или выходами.

Отношения порядка, обозначаемые в дальнейшем Ψ , описывают, как правило, условия существования (допустимого) моделируемой консультируемой проблемы. Эти условия практически всегда могут быть приведены к виду

$$\Psi: \langle p^i \rangle \leq (\geq) \langle p^j \rangle; \quad p^i, p^j \in P \quad (10.2)$$

определяющему отношения порядка между отдельными компонентами вектора переменных. В частности, в тех случаях, когда ограничения на переменные модели заданы в виде изопараметрических неравенств, переход к отношениям (10.2) возможен путем определения левой части ограничений как отображения со значениями, присутствующими в отношениях порядка.

В связях, определенных отношениями порядка, присутствуют обычно переменные, которые имеют смысл допустимых верхних или нижних пределов изменения значений переменных, описывающих консультируемую проблему. Эти переменные, в большинстве своем, носят характер директивных или нормативных данных. Простейшими примерами такого рода отношений могут служить $\sigma \leq [\sigma]$ или $N \leq N^{TP}$, определяющие допустимость некоторых решений (рекомендаций) из условий устойчивости и надежности, соответственно (где $[\sigma]$ — допустимый коэффициент устойчивости; N^{TP} — требуемый уровень надежности). С учетом сказанного, при представлении математической модели консультируемой проблемы можно ограничиться рассмотрением только связей-отображений, полагая при этом, что их переменные в процессе функционирования КМ могут быть как фиксированы, так и ограничены.

Неотъемлемым атрибутом модели являются ее области определения X и значений Λ , пара которых задает область применения модели \wp , т. е.

$$\wp = \{X, \Lambda\}.$$

В итоге математическая модель консультируемой проблемы в целом (будем называть ее исходной моделью — M) может быть представлена в виде

$$M = \langle \Phi, \chi, \lambda, \wp \rangle. \quad (10.3)$$

Известно, что модель M , имеющая в качестве выхода вектор λ , может быть представлена в виде совокупности из N_λ моделей-отображений со скалярными выходами $\lambda^k \in \lambda$, $k = 1, 2, N_\lambda$, где N_λ — размерность вектора λ . Такие модели будем называть скалярными.

При программной реализации математической модели в качестве самостоятельной единицы (прикладного программного модуля) используются подмодели, являющиеся блоками различной размерности, составленные из скалярных моделей. В дальнейшем такого рода блоки будем называть элементарными моделями и обозначать буквой m .

Каждая j -я элементарная модель может быть представлена в виде, аналогичном выражению (10.3);

$$m_j = \langle f_j, \chi_j, \lambda_j, \wp_j \rangle, \quad (10.4)$$

а компоненты исходной модели могут быть записаны через компоненты элементарных моделей следующим образом:

$$\begin{aligned} \Phi &= \{f_i\}_{i \in [1, N_m]}; \\ \lambda &= \bigcup_{i=1}^{N_m} \lambda_i; \quad \chi = \left(\bigcup_{i=1}^{N_m} \chi_i \right) | \lambda; \quad P = \bigcup_{i=1}^{N_m} P_i; \\ \wp_i &= \text{Pr}_{p_i}(\wp); \quad i = 1, 2, \dots, N_m, \end{aligned}$$

где $\text{Pr}_{p_i}(\wp)$ — проекция множества \wp на гиперплоскость, координатами которой являются компоненты вектора P_i ; N_m — число элементарных моделей в исходной модели.

Будем называть модель M связной, если для каждой совокупности элементарных моделей $\{m_{i'}\}_{i' \in I}$ ($I \subset [1, N_m]; [1, N_m] \setminus I \neq \emptyset$) найдется модель $m' \notin \{m_{i'}\}_{i' \in I}$, такая, что векторы P' и $\{P_{i'}\}_{i' \in I}$ пересекутся, т. е.

$$\forall \{m_{i'}\}_{i' \in I} \subset M, \quad \exists j \Rightarrow \{P_{i'}\}_{i' \in I} \cap P_j \neq \emptyset; \quad j \notin I.$$

При невыполнении этого условия модель M называется несвязной. Она может быть представлена в виде ряда связных подмоделей, каждая из которых допускает автономное рассмотрение. Поэтому в дальнейшем основное внимание будем уделять связным моделям.

Особо отметим такую характерную особенность математических моделей, как *разрывность их связей, под которой понимается наличие*

в них конечного числа точек разрыва, где возможны изменения как оператора, задающего связь, так и состава содержащихся в этих связях переменных.

Отметим, также, что часть аргументов связей модели может не входить непосредственно в левую часть отношений, а присутствовать лишь в задании области определения. Поэтому представляется целесообразным различать явные и неявные аргументы связей модели. *Под явными аргументами понимают аргументы, непосредственно входящие в левую часть отношения, а под неявными — аргументы, фигурирующие только в задании области определения.*

Переменные модели, носящие дискретный характер, как правило, не являются выходными и не присутствуют в явном виде в числе входных и, в основном, задают область определения связей модели. Например, если рассматривается модель консультируемой проблемы представлена как летательный аппарат (ЛА), то такие параметры, как тип системы подачи топлива, аэродинамическая схема ЛА и т. п., не являются явными аргументами ни в одной из связей модели. Однако при определении модели двигательной установки первая из названных переменных определяет ту или иную методику расчета, а вторая существенна при выборе модели для определения аэродинамических характеристик ЛА. При этом в каждой расчетной методике в качестве исходных данных могут присутствовать различные переменные. Так, при расчете массы двигательной установки ЛА с турбонасосной подачей топлива используют такие данные, как число оборотов турбины и т. п., которые, очевидно, не требуются при определении массы двигательной установки с вытеснительной системой подачи топлива.

Можно привести аналогичные примеры с непрерывными неявными аргументами. В частности, при формировании рекомендаций по расчету ряда аэродинамических характеристик ЛА в используемых моделях отсутствует в числе явных аргументов переменная «удлинение крыла», в то время как значения этих характеристик определяются различными формулами для крыльев малого, среднего и большого удлинений.

Путем введения в состав неявных аргументов таких характеристик моделей, как трудоемкость, обеспечиваемая точность вычислений и т. п., можно «развязать» модели, имеющие общие выходные переменные, т. е. считать, что в составе исходной модели отсутствуют элементарные модели с пересекающимися областями применения и выходными переменными. Далее предполагается, что такая возможность реализована и модель M обладает свойством:

$$\lambda_i \cap \lambda_j = \emptyset \text{ при } \wp_i \cap \wp_j \neq \emptyset (i \neq j) \quad i, j = 1, 2, \dots, N_m).$$

В таком случае каждая разрывная модель может быть представлена в виде совокупности непрерывных моделей с непересекающимися областями применения.

Заметим, что обычно задание областей определения связей модели дискретными переменными производится фиксацией их конкретных значений, а задание этих же областей непрерывными переменными — путем определения диапазонов изменения таких переменных. Несложно проследить связь между введенным ранее понятием «концепция» (4.1) и областями применения математических моделей. Концепцию теперь можно определить как совокупность сформированных рекомендаций, математические модели которых имеют непрерывную область \wp :

$$\wp \Leftrightarrow \Omega. \quad (10.5)$$

В п. 4.2.1 было проведено разделение процесса формирования рекомендаций на две последовательные стадии. На первой из них производится формирование концепций, в результате чего возможно выделение из общей модели, описывающей множество концептуально отличающихся рекомендаций, непрерывной модели с областью \wp , соответствующей рассматриваемой в текущий момент концепции Ω . На второй, достаточно хорошо формализуемой стадии, выбирают рациональный вариант в рамках каждой концепции на основе уже непрерывной модели. Учитывая, что процедура определения концепции является первой и при ее выполнении математическое моделирование не используется, в представлении математических моделей возможно не выделять области \wp в явном виде. Тогда для дальнейшей разработки вопросов формирования КМ является достаточным представлением математической модели консультируемой проблемы в виде совокупности элементарных моделей:

$$M = \{m_i\}_{i=1, 2, \dots, N_m}$$

каждая из которых может быть определена следующими компонентами:

$$m_j = \langle f_j, \chi_j, \lambda_j \rangle. \quad (10.6)$$

Между приведенными компонентами и компонентами процесса формирования рекомендаций имеет место определенное соответствие, заключающееся в том, что вектор переменных $P = \{\chi, \lambda\}$ отражает параметры консультируемой проблемы — $\{x, y, z, K\}$:

$$P \Leftrightarrow \{x, y, z, K\},$$

а связи $\Phi = \{f_j\}_{j=1,2,\dots, N_m}$ — условия допустимости (V) и правило вычислений значений критериев оценки вариантов проекта (F):

$$\Phi \Leftrightarrow \{V, F\}$$

Приведенное определение математической модели соответствует ее представлению в виде ППП модульной структуры, где каждый модуль — элементарная модель — выступает в роли «черного ящика» с обозначенными лишь входами-выходами. В таком контексте основная информация, которая может быть использована в процессе формирования КМ, содержится в структуре исходной модели, отражающей информационные связи между входящими в ее состав элементарными моделями.

10.1.1.3. Сетевая структура математических моделей

Наиболее употребляемыми являются представления структуры математических моделей в виде графа. При этом возможно использование ориентированных графов, в которых вершины соответствуют операторам элементарных моделей $\{f_i\}_{i=1,2,\dots, N_m}$, а дуги — переменным, причем входящие дуги каждой вершины f_i определяют векторы χ_i , а исходящие — λ_i . Такое представление структуры модели применяют, обычно, при решении задач планирования вычислений на ППП модульной структуры или, другими словами, при гибком построении расчетных моделей, что является частью одной из процедур формирования КМ. В основе планирования в данном случае лежит тот факт, что множество всевозможных путей на графе исходной модели консультируемой проблемы в целом представляет собой множество подмоделей, которые можно сформировать из элементарных моделей, порождающих этот граф. Каждый такой j -й путь представляется естественно упорядоченным множеством чисел $T_j = \{t_1, t_2, \dots, t_{N_{T_j}}\}$, элементами которого являются номера вершин, лежащих на этом пути. При этом предполагается, что граф модели является предварительно упорядоченным по условию: $m^* \sqsubset m^{**}$, если на информационном графе существует путь из m^* в m^{**} , а его связи пронумерованы, соответственно, числами натурального ряда:

$$m^* = m_i, \quad m^{**} = m_j, \quad i < j.$$

В общем случае очевидно, что между некоторыми вершинами путь может не существовать. Относительная нумерация таких вершин может быть произвольной. Случаи, когда ряд вершин образуют на графе цикл, алгоритмами планирования на базе использования

ориентированных графов исходной модели принципиально не охватываются. Таким образом, на базе исходной модели M , описывающей консультируемую проблему в целом, путем объединения элементарных моделей в различных комбинациях можно сформировать множество моделей $M_{Tj} = \{m_i\}_{i \in T_j}$, соответствующих

отдельным агрегатам и системам этой консультируемой проблемы и различным аспектам ее функционирования. Выше такие модели были названы **агрегированными**.

В качестве примера рассмотрим некоторую совокупность элементарных математических моделей, описывающих геометрию и аэродинамику корпуса ЛА :

$$\begin{aligned} m_1 : \{D_\phi, R_{\text{зат}}, \theta_{\text{кон}}\} &\Rightarrow Cx_{\text{нос}}, \\ m_2 : \{Cx_{\text{тр}}, Cx_{\text{дн}}, Cx_{\text{корм}}, Cx_{\text{нос}}\} &\Rightarrow Cx_{\text{оф}}, \\ m_3 : \{M, D_\phi, R_{\text{зат}}, \theta_{\text{кон}}, \lambda_{\text{ц}}, C^{\alpha}y_{\text{оф}}\} &\Rightarrow C^{\alpha}y_{\text{ф}}, \\ m_4 : \{C^{\alpha}y_{\text{ф}}, \alpha\} &\Rightarrow Cy_{\text{ф}}, \\ m_5 : \{Cx_{\text{оф}}, \lambda_{\text{нос}}, M, Cy_{\text{ф}}, \alpha\} &\Rightarrow Cx_{\text{иф}}, \end{aligned}$$

где D_ϕ — диаметр корпуса; $R_{\text{зат}}$ — радиус затупления носка корпуса; $\theta_{\text{кон}}$ — полуугол раствора носового конуса; M — число Маха набегающего потока; $Cx_{\text{нос}}$ — коэффициент сопротивления носовой части корпуса (затупленный конус); $Cx_{\text{оф}}$ — коэффициент лобового сопротивления корпуса при $\alpha = 0$; $Cx_{\text{тр}}$ — коэффициент сопротивления трения корпуса; $Cx_{\text{дн}}$ — коэффициент донного сопротивления корпуса; $Cx_{\text{корм}}$ — коэффициент сопротивления кормовой части корпуса.

Здесь m_1 определяется следующими скалярными моделями:

$$\begin{aligned} \bar{r} &= 2R_{\text{зат}}/D_\phi; \\ \lambda'_{\text{нос}} &= 1/2 \operatorname{tg} \theta_{\text{кон}}, \\ C'x_{\text{нос}} &= f_1(M, \lambda'_{\text{нос}}); \\ Cx_{\text{зат}} &= f_2(M); \end{aligned}$$

$Cx_{\text{нос}} = C'x_{\text{нос}} (1 - \bar{r}^2 \operatorname{COS}^2 \theta_{\text{кон}}) + Cx_{\text{зат}}$,
 где \bar{r} — относительный радиус затупления носка корпуса; $\lambda'_{\text{нос}}$ — удлинение носовой части корпуса без затупления; $C'x_{\text{нос}}$ — коэффициент сопротивления носовой части корпуса (заостренный конус); $Cx_{\text{зат}}$ — коэффициент сопротивления затупленной носовой части; $f_1(M, \lambda'_{\text{нос}})$ — таблица значений функции $C'x_{\text{нос}} = f_1(M, \lambda'_{\text{нос}})$; $f_2(M)$ — таблица значения функции $Cx_{\text{зат}} = f_2(M)$.

На рис. 10.3 показан граф, описывающий структуру данной модели, а на рис. 10.4, $a-g$ — примеры принципиально возможных агрегированных моделей, которые могут быть составлены на базе исходной модели.

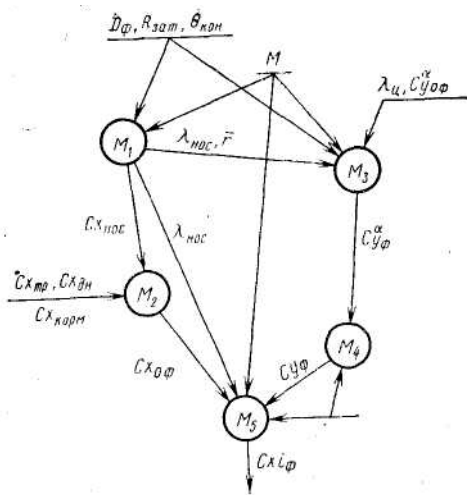


Рис. 10.3. Пример ориентированного информационного графа модели.

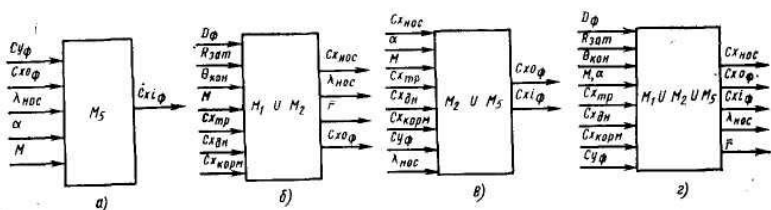


Рис. 10.4. Примеры агрегированных моделей.

Гибкость исходной модели будем оценивать числом агрегированных моделей, которые могут быть сформированы на ее основе. Уровень такого рода гибкости определяется двумя факторами. Первым из них является степень разбиения исходной модели на элементарные модели, представляющие собой **программные модули** — **неделимые составные части ППП**. Предельным в данном случае является отождествление скалярных и элементарных моделей. Однако при этом надо учитывать, что элементарная модель может включать в себя ряд связей, выделение каждой из которых в автономную программную единицу не имеет смысла (см. модель m_1), так как

вычисляемые ими значения не представляют самостоятельного интереса. Разделение исходной модели на элементарные модели производится при ее программной реализации, что исключает влияние способов реализации процедур формирования КМ на рассматриваемый фактор.

Вторым фактором, определяющим гибкость исходной модели, является применяемый способ формирования агрегированных моделей из ее элементарных составляющих.

В частности, при использовании способов, базирующихся на представлении структуры исходной модели в виде ориентированного графа, допускается формирование лишь таких агрегированных моделей, входы и выходы которых содержатся во входах и выходах, соответственно, исходной модели. Такого рода формирование является наиболее рациональным и исследованным.

В то же время из требования гибкости, предъявляемого к процедурам формирования КМ, следует, что на одной и той же исходной математической модели должно допускаться решение задач в различных постановках. При этом переменные, которые в одних задачах являются входными, в других могут быть выходными и наоборот. Такого рода гибкость допустима лишь при описании элементарных моделей наборами входящих в них переменных без деления их на входные и выходные. Такое рассмотрение модели практически исключает ее представление в виде ориентированного графа, но обеспечивает расширенные возможности как при формировании расчетных моделей для выполнения консультационных операций, так и при выполнении других процедур формирования КМ. Покажем это.

Будем считать, что задание на формирование расчетных моделей всегда может быть определено парой (I, T) , где I — вектор исходных переменных; T — требуемые результаты моделирования. В частности, если выбор формируемых рекомендаций при выполнении некоторой консультационной операции S_{ij} оптимизационного статуса основан на использовании свертки критериев, то для этой операции набор переменных, определяющих требуемую расчетную модель, может быть представлен в виде

$$I = \{x_{ij}, z_{ij}, y_{ij}\}; \quad (10.7)$$

$$T = \{k^1_{ij}, k^2_{ij}, \dots, k^{N_{k_{ij}}}_{ij}\}. \quad (10.8)$$

Для выполнения операций расчетного статуса требуемая расчетная модель может быть представлена парой:

$$I = \{x_{ij}, z_{ij}\}; \quad (10.9)$$

$$T = y_{ij}. \quad (10.10)$$

Рассмотрим некоторую скалярную модель $\lambda_j = f_j(\chi_j)$, предполагая f_j явной функцией действительных переменных. Она соответствует заданию (I, T) , для которого $T = \lambda_j$, $I \subseteq \chi_j$. Однако нетрудно показать, что на базе этой же модели можно решить и ряд других задач, для которых

$$I \in p_j, T \text{—скаляр}; \quad (10.11)$$

$$(p_j \setminus T) \subseteq I, \text{ где } p_j = \lambda_j \cup \chi_j.$$

Данные условия следуют из возможности решения уравнения $\lambda_j - f_j(\chi_j) = 0$ относительно любой из компонент вектора p_j , а не только λ . При этом разрешимость приведенного уравнения относительно некоторой переменной из $p_j \setminus \lambda_j$ равносильна переориентации графа модели.

В случае элементарной модели m_j , для которой $|\lambda_j| > 1$, в дальнейшем будем считать, что условие $|T| = |\lambda_j|$, являющееся аналогом условия (10.11), обеспечивает разрешимость m_j относительно любых $|\lambda_j|$ компонент вектора p_j . Реализация переориентации m_j в этом случае может быть произведена, в частности, на базе оптимизационных методов решения уравнений. Значения $|\lambda_j|$ далее будем называть рангом модели m_j .

Исходя из сказанного, для последующего изложения исходную математическую модель консультируемой проблемы целесообразно представлять, как и раньше, в виде совокупности элементарных моделей. Но при этом не разделять их переменные на входные и выходные, а учитывать их ранг, т. е.

$$m_j = \langle f_j, p_j, |\lambda_j| \rangle. \quad (10.12)$$

Отметим, что если в качестве элементарной модели рассматривается модель, вектор выходных переменных которой является многомерным, то дуги, исходящие из определенной вершины, соответствуют различным переменным. Для того чтобы обеспечить полноту описания, приходится делать на дугах пометки (как в приведенном выше примере). Введение их определяет граф исходной модели как граф с помеченными дугами, исследование которого затруднительно. Все это делает целесообразным переход к ***представлению структуры модели в виде двудольного графа.***

В дальнейшем будем представлять структуру исходной математической модели консультируемой проблемы в виде ***сети, интерпретируемой неориентированным двудольным графом***, состоящим из двух групп вершин и соединяющих их ребер: в первую группу входят вершины, образованные операторами элементарных математических моделей, а во вторую — переменные, входящие в эти модели; ребра определяют факт присутствия переменных в соот-

ветствующих моделях. Далее такого рода графом будем представлять информационный граф модели. На рис. 10.5 в качестве примера показан информационный граф, соответствующий фрагменту математической модели ЛА, приведенному на рис. 10.3.

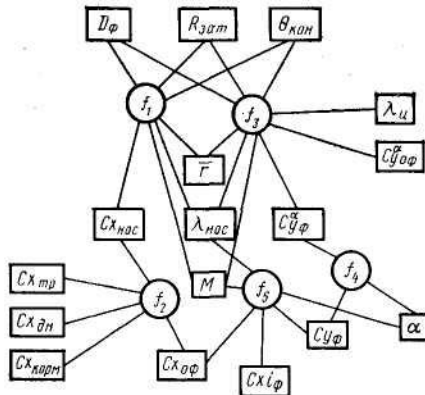


Рис. 10.5. Пример неориентированного двудольного информационного графа модели (□ — вершины-переменные, ○ — вершины-связи)

В ряде работ рассматривались решения задач планирования вычислений на базе ППП с представлением их модулей в виде (10.12). Однако это касалось лишь решения сугубо расчетных задач и основным препятствием при этом являлась проблема разрешимости систем уравнений, представленных программными модулями пакета. **В данной работе использование аппарата планирования распространяется на гибкое формирование расчетных моделей для решения задач оптимизации параметров сформированных рекомендаций.** При этом проблема разрешимости указанных систем уравнений естественно сводится к регулярно решаемым в настоящее время задачам условной оптимизации.

Кроме того, предлагаемое использование сетевого характера математических моделей консультируемых проблем распространяется помимо процедур формирования расчетных моделей и на процедуры определения статуса консультационных модулей, а также определения информационных связей между КМ.

10.1.1.4. Определение тандемной модели – базовой формы представления многоуровневых математических моделей

Многоуровневый характер математических моделей консультируемых проблем является следствием реализации формирования рекомендаций, как процесса последовательной детализации рекомендаций и порождается следующими характерными чертами этого процесса:

1) ограниченностью размерности задач, решаемых на каждом уровне детализации;

2) более полным описанием рекомендаций на каждом последующем уровне детализации по сравнению с предыдущими.

Ограничения на размерность задачи, решаемой на каждом q -м уровне детализации, непосредственно отражаются на размерности вектора переменных P_q в соответствующей консультационной модели M_q , на базе которой эта задача решается.

Необходимость ограничения размерности вектора переменных модели требует ранжирования его компонент по степени их влияния на вычисляемую величину. Наиболее значимые из них (d) непосредственно входят в модель, а влияние других учитывается в обобщенном виде с помощью специально вводимых в состав модели переменных (v). В роли данных переменных могут выступать как параметры консультируемой проблемы, имеющие физическую интерпретацию (например, коэффициенты весового совершенства конструкций), так и статистические (эмпирические) коэффициенты. В итоге состав вектора переменных P_q в общем случае может быть представлен двумя группами компонент — d_q и v_q , т. е.

$$P_q = \{d_q, v_q\}. \quad (10.13)$$

Несложно заметить, что вектор переменных v как по смыслу, так и по назначению соответствует реакциям z_q , а вектор d_q — всем другим компонентам, описывающим q -й уровень детализации (S_q). Учитывая это, установим следующие соответствия:

$$d_q \Leftrightarrow (x_q, y_q, K_q); \quad (10.14)$$

$$v_q \Leftrightarrow z_q. \quad (10.15)$$

Ограничения на размерность вектора переменных модели P_q распространяются в первую очередь, на размерность d_q . Причем, так как значения векторов x_q всегда являются заданными, а критерии, как правило, — производные от x_q, y_q , то действительные ограничения на размерность d_q определяются размерностью y_q . При рассмотрении более поздних уровней детализации увеличивается число сформированных рекомендаций. Этому соответствует увеличение

размерности вектора x и, соответственно, уменьшение ограничений на размерность вектора P .

Итак, вычисление той или иной характеристики консультируемой проблемы, в частности критериев оценки формируемых рекомендаций, может быть произведено с использованием различных моделей. Вначале могут использоваться модели, учитывающие лишь основные факторы, т. е. влияние которых на вычисляемые характеристики наиболее сильное, а затем, по мере детализации формируемых рекомендаций и расширения вектора x , возможно применение более полных моделей. Эту возможность будем использовать при формировании рекомендаций для сложных консультируемых проблем. Так, например, при формировании рекомендаций для проектирования ЛА прочностные и аэродинамические расчеты на начальных этапах формирования рекомендаций осуществляют по приближенным формулам, а далее — на основе методов типа конечных элементов, панельного и т. п., для использования которых необходимо задание значительно большего числа переменных, описывающих проектируемый ЛА. Очевидно, что применение этих методов возможно лишь после того, как сформированные рекомендации достаточно детализированы и основной массив исходных данных известен (в виде рекомендаций, сформированных на предыдущих уровнях).

Проиллюстрируем сказанное выше на примере моделей массовых характеристик ЛА, используемых при формировании рекомендаций для его проектирования. Здесь на ранних операциях детализации рекомендаций широко будем использовать приведенные уравнения, полученные на основе статистических данных. Переменными v в них являются: относительная масса топливных баков — $a_{то}$; относительная масса двигательной установки — $\gamma_{дв}$ и др. При этом сами связи задаются, как правило, в виде линейных функций. Так, в частности, масса топливного отсека $m_{то}$ задается как линейная функция от $a_{то}$ и массы топлива m_T , т. е.

$$m_{то} = a_{то}m_T. \quad (10.16)$$

В данном случае известно сильное влияние переменной $a_{то}$ на результаты расчетов. Последнее объясняется тем, что она представляет такие существенные факторы, влияющие на массу топливного отсека, как данные об отсеках различных форм, нагрузки на них, плотности компонентов и т. д. При этом формула (10.16) не учитывает перечисленные выше факторы, а следовательно, имеет малую достоверность.

После конкретизации компоновки аппарата используют зависимости, имеющие также чаще всего статистический характер, но

учитывающие тип и форму бака (Φ_6), давление наддува $P_{\text{над}}$, максимальные перегрузки H_{max} , плотность компонентов ρ_k , а также массу топлива (m_T). В этом случае формула, описывающая связь приведенных выше характеристик, уже будет нелинейной, а состав ее аргументов существенно расширен по сравнению с формулой (10.16):

$$m_{\text{то}} = m_{\text{то}}(\Phi_6, m_T, \rho_k, H_{\text{max}}, P_{\text{над}}, v_1). \quad (10.17)$$

Здесь переменные v_1 уже не имеют столь четко выраженного смысла, как $a_{\text{то}}$ в формуле (10.16).

По мере разработки рекомендаций на проектирование на последующих уровнях детализации формируют рекомендации по определению конструктивно-силовой схемы топливного отсека, действующие на него нагрузки, состав и массу устанавливаемой в нем арматуры. Далее, опираясь на результаты сформированных рекомендаций по прочностным расчетам, определяют массу топливного отсека как сумму масс входящих в него элементов: силовой конструкции ($m_{\text{ск}}$), арматуры ($m_{\text{арм}}$) и других элементов ($m_{\text{др}}$), в том числе и элементов автоматики:

$$m_{\text{то}} = m_{\text{ск}} + m_{\text{арм}} + m_{\text{др}}. \quad (10.18)$$

При этом массы каждого из представленных элементов топливного отсека в зависимости от параметров этих элементов, а также условий их работы ($\rho_k, H_{\text{max}}, P_{\text{над}}, m_T$ и др.) определяются своими моделями, которые могут иметь многоуровневый характер. Агрегируя эти модели и модель (10.18), выражение, определяющее $m_{\text{то}}$, можно представить в виде

$$m_{\text{то}} = m_{\text{то}}(m_T, \rho_k, H_{\text{max}}, P_{\text{над}}, m_T, P_{\text{над}}, P_{\text{ск}}, P_{\text{арм}}, P_{\text{др}}, v_2) \quad (10.19)$$

где $P_{\text{ск}}$ — параметры силовой конструкции; $P_{\text{арм}}$ — параметры арматуры; $P_{\text{др}}$ — параметры других элементов топливного отсека.

Приведенные примеры подтверждают, что по мере накопления данных по текущим рекомендациям для вычисления одних и тех же характеристик консультируемой проблемы возможно использование все более полных (в смысле учитываемых факторов) математических моделей.

Присутствие вектора v в составе переменных модели, как отмечалось выше, призвано компенсировать огрубление этих моделей, вызванное ограничениями на число входящих в них переменных. Значения компонент этого вектора выбирают, как правило, из условия сведения к минимуму погрешности, возникающей при огрублении модели.

Пусть известны некоторая модель $M_q = \langle d_q, v_q, f_q \rangle$:

$$f_q: \{ d'_q, v_q \} \Rightarrow \lambda_q,$$

где $d'_q = d_q \setminus \lambda_q$, а также полный набор факторов, влияющих на значения λ , представляемый совокупностью переменных, образующих вектор a :

$$a \supset d_q. \quad (10.20)$$

Предположим также, что известны истинные значения λ в N_q отдельных точках, т.е. $\{\langle a_i \rangle\}_{i=1,2,\dots, N_q}$. В этих условиях задача определения значений v может быть сформулирована как минимизация тем или иным образом выраженной погрешности I (где $I = I(\{\langle a_i \rangle\}_{i=1,2,\dots, N_q}, f_q, d_q, v_q)$) вычислений переменных λ на модели M_q :

$$v_q = \arg \min I(\{\langle a_i \rangle\}_{i=1,2,\dots, N_q}, f_q, d_q, v_q).$$

Подобного рода задачу принято называть **задачей идентификации математических моделей**.

Как правило, оператор f_q в задачах идентификации является известным и может рассматриваться как некоторая константа. Тогда

$$v_q = v_q(\{\langle a_i \rangle\}_{i=1,2,\dots, N_q}, d_q).$$

При рассмотрении многоуровневых моделей в роли точек $\{\langle a_i \rangle\}_{i=1,2,\dots, N_q}$ могут выступать реализации вычислений переменных λ_q на базе моделей, более полно учитывающих факторы, влияющие на λ , чем d_q . Тогда v_q можно представить в виде

$$v_q = v_q(d_q, d_b, v_i)$$

при условии

$$d_q \subset d_b. \quad (10.21)$$

где d_i — вектор переменных модели, по результатам вычислений на которой проводится идентификация.

Условие (10.21) является необходимым для того, чтобы по результатам численных экспериментов на базе модели $M_i = (d_i, v_b, f_i)$ можно было идентифицировать M_q . Его нарушение равносильно сопоставлению результатов выполнения двух (численных) экспериментов, отличающихся условиями их проведения.

Представим решение задачи идентификации как реализацию некоторого оператора Q :

$$Q: \{d_q, d_b, v_i\} \Rightarrow v_q.$$

Компоненты v_q отражают влияние на вычисляемую величину переменных d_i/d_q , поэтому результаты решения задачи идентификации можно представить в виде

$$v_q = Q(d_i/d_q, v_i). \quad (10.22)$$

При выполнении консультационных операций на базе математического моделирования последнее выражение с учетом соотношений (10.14) и (10.15) тождественно формуле (4.7).

Обобщая сказанное, введем формальное понятие **танDEMной модели** — канонического представления многоуровневых моделей,

положенного в основу формирования консультационных модулей. Модели $M_{q_1}, M_{q_2}, \dots, M_{q_k}$ будем называть одноименными с признаком одноименности \mathfrak{K} , если все они в составе своих выходных переменных имеют переменные \mathfrak{K} , т. е. если

$$\bigcap_{i=1}^k \mathfrak{K}_{q_i} = \mathfrak{K}$$

В частности, модели (10.16)—(10.18) являются одноименными с признаком одноименности $m_{\text{то}}$.

Тандемной моделью с признаком одноименности \mathfrak{K} будем называть совокупность одноименных моделей с признаком \mathfrak{K} , допускающую линейное упорядочение по условию:

$$i > j \Leftrightarrow M_i \text{ f } M_j \Leftrightarrow d_i \supset d_j.$$

В качестве примера такой совокупности могут служить модели (10.16), (10.17), (10.19). **Тандемные модели** будем обозначать через ${}^t M(\mathfrak{K})$.

Важным свойством тандемных моделей является то, что переменные v , присутствующие в моделях каждого уровня, могут быть определены через переменные моделей более низких уровней решением задачи идентификации. В результате, к совокупности связей исходной модели $\{f_i\}_{i=1,2,\dots, N_m}$ добавляются связи (10.22), имеющие место на каждой паре уровней тандемных моделей.

В общем случае на базе исходной модели может быть сформировано множество тандемных моделей, отличающихся признаками одноименности \mathfrak{K} . Уровни этих моделей являются, как правило, агрегированными из модулей ППП — элементарных моделей. Потребность в агрегировании определяется необходимостью обеспечения условий (10.21). Так, например, модель (10.16)—(10.18) не является тандемной. Однако агрегирование формулы (10.18) с моделями, вычисляющими $m_{\text{ск}}$, $m_{\text{арм}}$, $m_{\text{др}}$, порождает модель (10.19), которая в совокупности с соотношениями (10.16), (10.17) уже представляет тандемную модель.

Таким образом, каждый уровень тандемной модели ${}^t M(\mathfrak{K})$, обозначаемый далее ${}^t M_i(\mathfrak{K})$, представляет собой в общем случае совокупность элементарных моделей $\{m\}_i$. В состав модели ${}^t M_i(\mathfrak{K})$ включаются те элементарные модели, которые в своей совокупности обеспечивают выполнение условия (10.21) и удаление любой из них нарушает это основное условие тандемности. При этом возможно, что отдельные элементарные модели могут входить в различные тандемные модели.

Будем считать, что в тандемной модели выполняется условие (там, где очевидно, что речь идет о тандемной модели, индекс \mathfrak{x} будем опускать):

$$\lambda_{M_i} \subseteq \lambda_{M_j} \text{ при } i < j, \quad (10.23)$$

где λ_{M_k} — вектор выходных переменных модели M_k . Данное предположение можно выполнить всегда. В частности, если оно не выполняется, то в M_i следует включить то подмножество элементарных моделей из M_i , выходами из которых являются переменные $\lambda_{M_i} \setminus \lambda_{M_i}$. Тогда из условия (10.23) следует, что

$$\lambda_{M_i} = \mathfrak{x}, \quad (10.24)$$

т. е. вектор выходных переменных модели первого уровня является признаком одноименности тандемной модели.

Аналогично тому, как было введено понятие элементарной модели, введем определение *элементарной тандемной модели*, под которой будем понимать тандемную модель со скалярным признаком одноименности. Тандемные модели, имеющие в качестве признака одноименности вектор переменных, будем называть *агрегированными*. При этом очевидно, что такие модели представляют собой объединение элементарных тандемных моделей с признаками, являющимися компонентами указанного вектора, т. е.

$$M(\mathfrak{x}) = \bigcup_{i=1}^{N_{\mathfrak{x}}} M(\mathfrak{x}_i),$$

где $N_{\mathfrak{x}}$ — размерность вектора \mathfrak{x}

Из условия (10.24) следует, что моделью первого уровня элементарной тандемной модели является элементарная модель. Модели второго и последующего уровней элементарной тандемной модели могут уже быть агрегированными, имея в своем составе несколько элементарных моделей, в том числе и одноименную с моделью верхнего уровня.

Пусть имеется элементарная тандемная модель $M(\mathfrak{x}) = \{M_1, M_2, \dots, M_N\}$ с вектором выходных переменных на каждом уровне λ_i ($\mathfrak{x} \in \lambda_i$, $i=1, 2, \dots, N$). Любой компоненте $\lambda_i^j \in \lambda_i$ можно поставить в соответствие элементарную модель, имеющую в качестве выходной переменной эту компоненту. Тогда, учитывая условие (10.24), можно заключить, что каждой элементарной модели i -го уровня соответствует по одноименной ей модели на всех нижележащих уровнях.

Если рассмотреть некоторую модель M_i то входящие в ее состав элементарные модели можно разделить на две группы:

- модели, имеющие на выходе переменные из λ_{i-1} , т. е. модели, одноименные которым содержатся в M_{i-1} ;

- модели, имеющие одноименные только на нижележащих уровнях, т.е. модели, имеющие на выходе переменные из $\lambda_i \setminus \lambda_{i-1}$.

Каждую из элементарных моделей $m_j \in M_i$, относящуюся ко второй из названных групп, можно рассматривать как первый уровень элементарной тандемной модели с признаком одноименности — выходной переменной из m_j . Поэтому каждую элементарную модель $M(\lambda_i)$ можно рассматривать как объединение элементарных тандемных моделей, признаками одноименности которых являются переменные из $\lambda_i \setminus \lambda_{i-1}$ ($i=2, 3, \dots, N-1$). Так, например, элементарные модели, вычисляющие $m_{ск}$, $m_{арм}$, $m_{др}$ в виде, подобном соотношению (10.17), и входящие как составные части в агрегированную модель (10.19), могут рассматриваться как первый уровень тандемных моделей соответственно с признаками одноименности $m_{ск}$, $m_{арм}$, $m_{др}$. Их последующими уровнями являются модели, аналогичные моделям (10.18), (10.19).

В итоге структуру тандемной модели в общем случае можно представить в виде, показанном на рис. 10.6.

Структура элементарных моделей на каждом уровне была приведена ранее на рис. 10.5.

На базе элементарных моделей путем их агрегирования могут быть сформированы различные тандемные модели. При этом, основываясь на представленной структуризации тандемных моделей, их формирование в целом может быть сведено к построению двухуровневых скалярных моделей, первым уровнем которой является некоторая элементарная модель $m^* = {}^t M_1(\lambda^*)$, а вторым (${}^t M_2(\lambda^*)$) — как правило, агрегированная модель. При этом, если m^* , в свою очередь, входит в состав агрегированной модели, являющейся вторым уровнем в другой, ранее сформированной тандемной модели, то путем подстановки ${}^t M_2(\lambda^*)$, вместо m^* можно сформировать третий уровень этой модели и т. д.

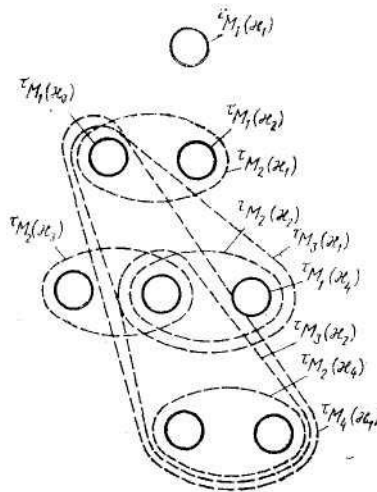


Рис. 10.6. Структура тандемной модели

Таким образом, базовая форма представления математических моделей консультируемых проблем характеризуется:

- наборами переменных;
- связями между этими переменными, непосредственно присутствующими в моделях, и связями идентификации;
- структурой (сетевой и многоуровневой).

Использование данной базовой формы позволяет свести реализацию основных процедур формирования КМ к решению формальных задач, формулируемых как задачи анализа структурных свойств исходной модели.

10.1.2. Формирование консультационных модулей для их автономного функционирования

10.1.2.1. Обобщение процедур формирования консультационных модулей

Использование предложенного представления математических моделей консультируемых проблем позволяет обобщить процедуры, связанные с формированием КМ, и представить их как формальные задачи, решаемые на базе этих моделей.

Вначале рассмотрим процедуры: определения статуса КМ, гибкого формирования расчетных моделей, выявления горизонтальных информационных связей между отдельными КМ.

Все названные процедуры так или иначе связаны с анализом взаимозависимости отдельных параметров консультируемой проблемы, что в условиях выполнения данных процедур на базе математического моделирования может трактоваться как выявление возможности выражения одних переменных модели через другие с помощью связей исходной модели консультируемой проблемы. Так, во-первых, задача гибкого формирования расчетной модели может быть сведена к выявлению связей модели, с помощью которых вычисляемые переменные могут быть выражены через переменные с известными их значениями. Эта задача в условиях представления исходной математической модели в виде неориентированного графа равносильна выделению из связей этой модели упорядоченной совокупности минимально-замкнутых систем.

Во-вторых, статус консультационных модулей, в основном, определяется возможностью однозначного выражения переменных, соответствующих искомым формируемым рекомендациям y , через переменные, соответствующие исходным данным x , z . Если такая возможность существует, то статус модуля определяется как расчетный. В противном случае он будет либо оптимизационным (если возможность указанного выражения отсутствует), либо невыполнимым (если компоненты x , z взаимозависимы, а заданные их значения не соответствуют друг другу). В этом случае, очевидно, ни одно значение y не может соответствовать исходным данным x , z , т. е. $Y = \emptyset$.

В-третьих, наличие информационных связей между отдельными консультационными модулями непосредственно основывается на факте взаимозависимости формируемых рекомендаций, выбираемых при функционировании различных КМ.

Покажем, что выполнение указанных трех процедур может быть обобщено и сведено, в основном, к решению единой для всех них формальной задачи. Постановка последней базируется на использовании сетевого характера структуры исходной математической модели консультируемой проблемы, т. е. на представлении этой структуры в виде сети на базе двудольного неориентированного графа.

Далее для упрощения изложения будем полагать, что исходная математическая модель состоит из скалярных моделей. Такое допущение не нарушает общности изложения, так как при этом предполагается, что проведена эквивалентная замена каждой

элементарной модели m_i совокупностью из $|\lambda_j|$ скалярных моделей с одинаковыми для всех них векторами входов χ_j и операторами f_j , но с отличающимися выходными переменными, в роли которых последовательно выступают отдельные компоненты λ_j .

Пусть имеется некоторый информационный граф модели, вершинами-связями которого являются операторы скалярных моделей в неявном виде:

$$\Phi_j(P_j) = 0, j=1, 2, \dots, N_m,$$

где

$$\Phi_j(P_j) = \lambda_j - f_j(\chi_j), j=1, 2, \dots, N_m.$$

Будем каждый раз исключать из состава переменных модели те переменные, значения которых в текущий момент зафиксированы, например, в результате выполнения операций более высокого уровня детализации. В результате этого рассматриваемый информационный граф, отражая текущее состояние рекомендаций, постоянно изменяется.

Пронумеруем вершины-связи графа исходной модели произвольным образом. Некоторое подмножество индексов вершин-связей обозначим через L , а множество вершин-переменных, входящих в связи с индексами из L , — через $P(L)$:

$$P(L) = \bigsqcup_{j \in L} P_j.$$

Пусть заданы некоторое множество L и вектор переменных $J \subseteq P(L)$, значения которых полагаем известными. Обозначим через $\sigma(J, L)$ — вектор переменных из $P(L)$, значения которых полагаем неизвестными:

$$\sigma(J, L) = P(L) \setminus J.$$

Тогда на каждом из $L \subseteq [1, N_m]$ возможно одно из следующих трех условий:

$$|L| = |\sigma(J, L)|; \tag{10.25}$$

$$|L| < |\sigma(J, L)|; \tag{10.26}$$

$$|L| > |\sigma(J, L)|; \tag{10.27}$$

Будем считать, что связи, входящие в исходную математическую модель, являются независимыми и что каждая из них может быть разрешена относительно любой из входящих в нее переменных. Тогда множество L , для которого:

-имеет место условие (10.27), задает переопределенную систему уравнений $\{\Phi_i(P_i)=0\}_{i \in L}$. Вектор J в данном случае является противоречивым, поскольку одна часть его компонент может быть выражена через другую, и их независимое задание является недопустимым;

- имеет место условие (10.25) и ни на каком его подмножестве не выполняется условие (10.27), определяет замкнутую систему уравнений. При этом каждая переменная из $\sigma(J, L)$ может быть однозначно выражена через компоненты J с помощью совокупности связей $\{\Phi_i(P)\}_{i \in L}$;

- на каждом подмножестве выполняется условие (10.26), задает недоопределенную систему уравнений, что свидетельствует о невозможности выражения ни одной из переменных $\sigma(J, L)$ через J .

В дальнейшем будем говорить:

- вектор I является непротиворечивым, если ни при каком $L \subseteq [1, N_m]$ не выполняется условие (10.27), и противоречивым, если это условие соблюдается хотя бы при одном L ;

- вектор J является корректным, если он непротиворечив, и существует $\tilde{L} \subseteq [1, N_m]$ такое, что $|\tilde{L}| = |\sigma(J, \tilde{L})|$;

- вектор J является τ -полным, если он корректен и

$$\sigma(J, \tilde{L}) \not\subset \tau; \tag{10.28}$$

- вектор J является частично τ -полным, если он корректен и

$$\sigma(J, \tilde{L}) \cap \tau \neq \emptyset; \tau \not\subset \sigma(J, \tilde{L}); \tag{10.29}$$

- вектор J является неполным, если при каждом $L \subseteq [1, N_m]$ выполняется условие (10.27).

Теперь рассмотрим возможность выполнения названных в начале данного пункта трех процедур формирования КМ на основе анализа на полноту, непротиворечивость и корректность некоторого составляемого каждый раз специальным образом вектора J .

Вначале рассмотрим процедуру, связанную с гибким формированием расчетных моделей. Как и раньше, исходные данные на такое формирование будем задавать парой (I, T) , где I определяет набор входных переменных формируемой расчетной модели, а T — набор выходных переменных этой модели.

В общем случае задание на формирование расчетной модели может не соответствовать имеющемуся ППП, программно отражающему исходную математическую модель, оперируя модулями которого и необходимо сформировать требуемую модель. В частности, вектор исходных переменных I может быть как недостаточен, так и избыточен для проведения вычислений переменных T на основе имеющегося пакета. Поэтому рассматриваемая процедура сводится к выполнению двух групп действий. Первая из них связана с анализом задания (I, T) на корректность с последующей (при необходимости) корректировкой

задания. Вторая группа действий связана непосредственно с выделением из $\{\Phi_i(P_i)\}_{i=1,2,\dots,N_m}$ при условии корректного задания (I, T) совокупности минимальных замкнутых систем и определением порядка их разрешения.

Задачи формирования агрегированных моделей из элементарных моделей в условиях представления их структуры неориентированным информационным графом рассматривались ранее в различных работах, но без учета возможных некорректностей заданий (I, T) . При этом априори предполагалось существование замкнутых систем, а некорректность задания (I, T) , приводящая к их отсутствию, рассматривалась как следствие ненахождения таких систем. Источники некорректностей в задании не анализировались и, тем более, не обеспечивалось их целенаправленное устранение.

В настоящей работе основное внимание уделяется выполнению действий первой группы — связанных с анализом и корректировкой заданий (I, T) , так как разработанные методы, выделяющие из $\{\Phi_i(P_i)\}_{i \in [1, N_m]}$ упорядоченные совокупности минимальных замкнутых систем при корректных (I, T) , на наш взгляд, исчерпывают данную проблему.

Отождествляя J с I , а τ с T , можно утверждать, что задача формирования расчетной модели может быть сведена к анализу вектора J на τ -полноту.

Рассмотрим теперь процедуру, связанную с определением статуса консультационных модулей. В этом случае положим:

$$J = |x, z|; \quad \tau = y.$$

Считая компоненты вектора y независимыми, на основе анализа вектора J на полноту и непротиворечивость можно сделать следующие выводы:

1) если вектор J неполон, то это означает отсутствие возможности отображения x, z в y на базе имеющейся модели, что соответствует оптимизационному статусу консультационного модуля;

2) если вектор J не является даже частично τ -полным, то в этом случае можно сделать вывод, аналогичный п. 1;

3) если вектор J τ -полон, то возможно выражение формируемых рекомендаций y через исходные данные x, z , что соответствует расчетному статусу консультационного модуля. Однако, если полученное расчетным путем значение y не содержится в множестве исходных альтернатив, то модуль переходит в разряд невыполнимых;

4) если вектор J частично τ -полон, то часть сформированных рекомендаций — содержащихся в $\sigma(\tilde{L}, J)$, может быть определена расчетным путем, а другая их часть, которая не может быть выражена

через x, z , — путем оптимизации. В этом случае будем говорить, что модуль имеет смешанный статус;

5) если вектор J противоречив, то это означает взаимозависимость компонент исходных векторов x, z . При этом можно сделать вывод, что в рассматриваемом случае отсутствует возможность нахождения значений y , удовлетворяющих одновременно всем исходным данным, т. е. соответствующий модуль является невыполнимым.

Последний вывод распространяется, в первую очередь, на тот случай, когда значения компонент исходных векторов зафиксированы, а не ограничены. В противном случае выводы о невыполнимом статусе консультационных модулей преждевременны. В частности, если взаимозависимые компоненты вектора J ограничены и ограничения на одну из таких компонент, пересчитанные в ограничения на другие, позволяют выделить общие области в рамках каждого ограничения, то соответствующие модули имеют, как правило, оптимизационный статус. В дальнейшем такого рода взаимозависимые и непротиворечивые компоненты представляются в векторе J одной из них с возможным изменением ранее наложенных на данную компоненту ограничений. Сказанное основывается на следующих соображениях.

Пусть определены две группы взаимозависимых компонент вектора J , обозначаемые в дальнейшем J_1 и J_2 ($J_1, J_2 \subset J$). Это означает, что на базе элементарных составляющих исходной модели, образующих переопределенную систему, возможно построение расчетной модели, реализующей некоторую связь $\varphi: J_1 \Rightarrow J_2$.

Будем считать, что компоненты каждого из векторов J_1 и J_2 независимы, а на их значения наложены ограничения $J_i^j \in [J_i^{j \min}, J_i^{j \max}]$, где $i = 1, 2, \dots, N_j, N_j$ — размерность J_j , т. е.

$$J_j \in \prod_{i=1}^{N_j} [J_i^{j \min}, J_i^{j \max}] = J_j^*, j=1,2.$$

Определим значения:

$$J_2^{j \min} = \min_{J_1 \in J_1^*} \varphi(J_1), \quad i=1, 2, \dots, N_2;$$

$$J_2^{j \max} = \max_{J_1 \in J_1^*} \varphi(J_1), \quad i=1, 2, \dots, N_2;$$

Тогда, если $J_2^* \cap J_2^{**} = \emptyset$ (где $J_2^{**} = \prod_{i=1}^{N_2} [J_2^{i \min}, J_2^{i \max}]$), то

консультационный модуль имеет статус невыполнимого, в противном случае он, как правило, выполним и имеет оптимизационный статус.

В частных случаях, когда $J_2^* \supset J_2^{**}$, ограничения на значения компонент вектора J_2 «слабее» ограничений, порождаемых J_1^* , т. е. ограничениями на J_1 и могут быть опущены из рассмотрения. Если $J_2^* \subset J_2^{**}$, то в рамках сделанных выше допущений из рассмотрения могут быть выпущены ограничения на J_1 .

Представленный анализ непротиворечивости областей возможных значений взаимозависимых компонент вектора J носит численный характер. Его реализация не представляет серьезных затруднений, если, во-первых, выявлены взаимосвязанные компоненты вектора J и, во-вторых, может быть сформирована агрегатированная модель для взаимного пересчета значений этих компонент. Ответам на данные два вопроса ниже уделяется основное внимание.

Рассмотрим процедуру формирования КМ, направленную на выявление горизонтальных информационных связей между консультационными операциями. Согласно определению эти связи существуют между каждой парой КМ, результаты выполнения которых являются взаимозависимыми, или, другими словами, между модулями, сформированные рекомендации которых могут быть выражены друг через друга с помощью связей модели консультируемой проблемы. При этом подчеркнем, что в данном случае рассматривается выявление взаимозависимостей, проявляющихся через связи модели, которые не позволяют получить обобщенные прогнозные параметры z , т. е. не содержат соотношений типа (10.22). Такого рода взаимозависимости определяют горизонтальные информационные связи.

Рассмотрим некоторую пару консультационных операций S_{ij} и S_{ik} . Отметим, что если та или иная операция $S_{\alpha\beta} \in [S_{ij}, S_{ik}]$ имеет оптимизационный статус, то ее вектор формируемых рекомендаций содержится в составе соответствующего набора входных переменных используемой расчетной модели [см., например, выражения (10.7), (10.9)], т. е. $u_{\alpha\beta} \subset I_{\alpha\beta}$. Если же сформированная рекомендация $S_{\alpha\beta}$ имеет расчетный статус, свидетельствующий о взаимном соответствии значений векторов $u_{\alpha\beta}$ и входных переменных используемой расчетной модели $I_{\alpha\beta}$, то отдельные компоненты $I_{\alpha\beta}$ могут быть выражены через $u_{\alpha\beta}$. Таким образом возможность взаимовыражения компонент векторов y_{ij} и y_{ik} всегда приводит к возможности взаимовыражения компонент I_{ij} и I_{ik} . При этом вектор, представляющий их объединение, в рассматриваемом случае всегда противоречив, так как каждый из векторов I_{ij} , I_{ik} в отдельности является корректным. В результате

можно сделать вывод, что взаимосвязь векторов формируемых рекомендаций y_{ij} , y_{ik} тождественна противоречивости вектора I_{ij} и I_{ik} . Тогда, приняв

$$J = I_{ij} \cup I_{ik}$$

можно утверждать, что если этот вектор противоречив, то соответствующие консультационные модули находятся в информационном взаимодействии. Итак, показана возможность обобщения трех рассмотренных процедур формирования КМ. При этом принцип их обобщения состоит в том, что все эти процедуры могут быть сведены к анализу взаимозависимости и полноты компонент некоторого вектора J , формируемого при рассмотрении каждой процедуры специальным образом (табл. 10.1).

Таблица 10.1.

Наименование процедур формирования КМ	Формирование векторов J и τ	Вектор J			
		τ -полон	частично τ -полон	неполон	противоречив
Определение статуса консультационного модуля	$J = \{x_{ij}, z_{ij}\}$ $\tau = y_{ij}$	Расчетная	Операции Смешанная		Невыполнимая
Формирование расчетной модели	$J = I_{ij}$ $\tau = T_{ij}$	Задание (I, T) корректно	Задание (I, T) должно быть скорректировано		
Выявление информационных связей между консультационными операциями S_{ij} и S_{ik}	$J = I_{ij} \cup I_{ik}$	Информационная связь отсутствует			Информационная связь имеет место

Теперь рассмотрим процедуру выявления вертикальных информационных связей между КМ. Реализация этой процедуры базируется на использовании многоуровневого характера математических моделей консультируемых проблем, а именно, на использовании введенного выше понятия тандемной модели.

Вертикальные информационные связи согласно изложенному в п. 4.2 представлению процесса формирования рекомендаций соединяют каждую консультационную операцию S_{ij} с теми операциями детализирующих уровней, формируемые рекомендации которых влияют на результаты S_{ij} посредством реакций z_{ij} . Поэтому рассматриваемая функция состоит в выделении для каждой S_{ij} группы

детализирующих операций, результаты выполнения которых позволяют определить значения реакций z_{ij} .

При выполнении консультационных операций на базе математического моделирования в роли реакций z выступают переменные v . Значения этих переменных определяются из решения задачи идентификации (10.22). Необходимым условием решения данной задачи является условие (10.20), а в случае идентификации модели по результатам вычислительных экспериментов, проводимых на базе более полных моделей, условие (10.21). Последнее определяет: каждая модель может идентифицироваться по результатам вычислений только на тех моделях, совместно с которыми она представляет тандемную модель, являясь в ней моделью более высокого уровня.

Поэтому, если операция S_{ij} выполняется на основе расчетной модели M_{ij} , определяющей, в частности, значения критерия κ_{ij} , то v_{ij} зависит от тех переменных, которые являются дополнительными по отношению к d_{ij} аргументами в модели: следующим за M_{ij} уровнем в некоторой

${}^{\tau}M(\mathbf{x}=\kappa_{ij}) \ni M_{ij}$. Соответственно, консультационные операции, в процессе выполнения которых принимаются решения, описываемые этими дополнительными переменными, являются вертикально связанными с S_{ij} .

Итак, если известна модель M_{ij} , то информационные связи к S_{ij} от операций детализирующих уровней могут быть определены как результат формирования тандемной модели ${}^{\tau}M(\mathbf{x}=\kappa_{ij})$, $M_{ij} = {}^{\tau}M_1(\mathbf{x}=\kappa_{ij})$. При этом связь к S_{ij} от S_{kl} ($\kappa > i$) имеет место, если

$$\tilde{P}_{kl} \cap ({}^{\tau}d_m / {}^{\tau}d_l) \neq \emptyset \quad (m=2,3,\dots,N_{\kappa_{ij}})$$

где \tilde{P}_{kl} — вектор переменных, описывающих результаты выполнения S_{kl} ; ${}^{\tau}d_r$ — вектор переменных r -го уровня тандемной модели

${}^{\tau}M(\kappa_{ij})$ (${}^{\tau}d_l = d_{ij}$); $N_{\kappa_{ij}}$ — число уровней в ${}^{\tau}M(\kappa_{ij})$.

Тандемные модели в общем случае не являются заранее сформированными, а получаются путем агрегирования из элементарных моделей — модулей ППП. При этом имеется в виду, что если сформирована некоторая тандемная модель ${}^{\tau}M(\mathbf{x}) = \{ {}^{\tau}M_1(\mathbf{x}), {}^{\tau}M_2(\mathbf{x}), \dots, {}^{\tau}M_n(\mathbf{x}) \}$, то момент, когда в процессе формирования рекомендаций становятся известными данные, достаточные для вычисления \mathbf{x} по некоторой совокупности (α) элементарных моделей, удовлетворяющих условию $\bigcup_{j \in \alpha} d_j \supseteq {}^{\tau}d_r$, определяет появление $(r + 1)$ -го уровня в

рассматриваемой тандемной модели. Другими словами, появление

$(r + 1)$ -го уровня ${}^rM(x)$ означает достаточность известных данных по консультируемой проблеме для вычисления переменной x по более полной, чем ${}^rM_r(x)$, модели. При этом имеется в виду, что такая более полная модель принципиально может быть сформирована из модулей имеющегося ППП.

Учитывая описанную выше структуру тандемных моделей, процедуру формирования ${}^rM_{r+1}(x)$ можно рассматривать как формирование ${}^rM(\lambda^*)$, где λ^* — выходная переменная одной из элементарных моделей, составляющих ${}^rM_r(x)$.

Резюмируя сказанное, можно сделать вывод, что процедура по выявлению вертикальных информационных связей сводится к постоянному «отслеживанию» появления в процессе формирования рекомендаций совокупностей элементарных моделей, которые, во-первых, могут быть вычислены (сформированы) на основе полученных к данному моменту времени результатов и, во-вторых, составляют второй уровень тандемных моделей, где первым уровнем является элементарная модель, содержащаяся в расчетной модели той или иной уже выполненной операции.

10.1.2.2. Оператор КОН как основа выполнения процедур формирования консультационных модулей

Согласно предложенному выше обобщению процедур формирования КМ в основе их выполнения лежит определение взаимозависимости компонент некоторого вектора J , специальным образом составляемого при выполнении каждой процедуры. При этом основная информация заключается в характеристике рассматриваемого вектора: является ли он *корректным, противоречивым или неполным*. Эта информация позволяет установить содержание требуемых действий в процессе формирования КМ (например, предусмотреть взаимодействие формируемого КМ с соседними модулями при выявлении информационной связи между ними). Однако она еще не позволяет организовать целенаправленного выполнения этих действия (в частности, в приведенном примере для этого необходимо установить конкретные консультационные параметры, значения которых должны быть согласованы).

Определим характеристики каждого состояния вектора J (противоречивого, корректного, неполного), их консультационную интерпретацию и использование в процессе формирования КМ. Для этого введем в рассмотрение некоторую переменную $x(J)$, значениями которой могут быть $J0$, J^+ или J^- :

$$\kappa(J) = \begin{cases} J^0 \\ J^+ \\ J^- \end{cases}$$

где J^0 , J^+ , J^- — соответственно определяют факт корректного, противоречивого и неполного вектора.

Если вектор J противоречив (будем обозначать такой вектор J^+), то представляется необходимым определить те его компоненты, которые являются взаимозависимыми.

Взаимозависимые компоненты вектора J^+ будем обозначать через $\Phi St(J^+)$. Определение данных компонент позволит:

1) при выполнении процедуры по формированию расчетных моделей обеспечить целенаправленную корректировку вектора исходных данных I , удаляя из него зависимые компоненты;

2) при выполнении процедуры по определению статуса консультационных модулей выявить взаимозависимые компоненты исходных ограничений, задающих область допустимых рекомендаций. Если эти компоненты являются не фиксированными, а ограниченными, то можно найти их «пересечения» и определить, являются ли они взаимоисключающими. Если же взаимозависимые компоненты исходных данных являются взаимоисключающими, то их выявление позволит, в случае необходимости, обеспечить целенаправленную их корректировку;

3) при выполнении процедуры по выявлению информационных связей конкретно указать взаимозависимые формируемые рекомендации различных операций и организовать процесс их согласования.

Очевидно, что каждый вектор J^+ может быть преобразован в корректный путем удаления из него отдельных взаимозависимых компонент. Обозначим через $St(J^+)$ число компонент из J^+ , при удалении которых из данного вектора он становится корректным.

Для достижения корректности из J^+ должны удаляться переменные в определенных сочетаниях. Например, рассмотрим исходную модель, описываемую следующими уравнениями:

$$\left. \begin{aligned} f_1(x_1, x_2, x_3, x_4) &= 0; \\ f_2(x_1, x_2, x_3) &= 0; \\ f_3(x_1, x_2, x_4) &= 0; \\ f_4(x_4, x_5, x_5) &= 0; \end{aligned} \right\} \quad (10.30)$$

При $J = \{x_1, x_2, x_5, x_6\}$ получим $\sigma = \{x_2, x_4\}$. При $L = \{1, 2, 3, 4\}$, т. е. при рассмотрении всех четырех уравнений имеет место условие (10.27), определяющее названный вектор J как противоречивый. Следовательно, $\varkappa(J) = J^+$. При этом $St(J^+) = 2$, а $\Phi St(J^+) = \{x_1, x_3, x_5, x_6\}$. Удалим из J элементы x_5, x_6 . В этом случае $J = \{x_1, x_3\}$ противоречив, так как при $L = \{1, 2, 3\}$ имеет место условие (10.27). В то же время, если из J удалить x_5, x_1 , то полученный вектор $J = \{x_3, x_6\}$ является корректным. Также корректными являются векторы $J = \{x_1, x_5\}$; $J = \{x_1, x_6\}$ и др.

В дальнейшем каждую i -ю совокупность компонент противоречивого вектора J^+ , путем удаления которых данный вектор становится корректным, будем обозначать $USt_i(J^+) \subseteq \Phi St(J^+)$:

$$|\{USt_i(J^+)\}| = St(J^+); J^+ \setminus USt_i(J^+) = J^0.$$

Таким образом, для противоречивого вектора J важными являются характеристики $\Phi St(J^+)$, $St(J^+)$, $\{USt_i(J^+)\}_{i=1,2,\dots,N_{J^+}}$ (где N_{J^+} — число возможных комбинаций взаимозависимых компонент вектора J^+ , путем удаления которых из J^+ этот вектор может быть преобразован в корректный). При этом весьма существенным является то, что способы сведения противоречивого вектора к корректному могут быть неоднозначны, т. е. в общем случае $N_{J^+} > 1$.

Последнее обстоятельство делает необходимым участие ЛФР,а в выборе конкретного варианта $USt_i(J^+)$, смысл которого определяется выполняемой процедурой формирования КМ. Так, при формировании расчетной модели выбор ЛФР,ом того или иного варианта $USt_i(J^+)$ приводит к конкретизации первоначально заданного состава исходных данных, оставляя часть из них в том же качестве и переводя другую их часть в состав вычисляемых величин.

При выделении информационных связей между консультационными модулями выбор $USt_i(J^+)$ равносильен определению векторов r_{ij} (см. п. 10.2.2), т. е. конкретных параметров формируемых рекомендаций, по которым будет происходить непосредственно согласование результатов этих модулей. Например, пусть при выполнении некоторого консультационного модуля используются данные по нескольким параметрам, определяемым в процессе

функционирования другого модуля, которые в свою очередь являются взаимозависимыми. Тогда все эти данные будут содержаться в $\Phi St(J^+)$, а определение $USI(J^+)$ равносильно выделению из этих параметров, тех, значения которых должны стать предметом согласования.

При определении невыполнимого статуса консультационных модулей множество $\Phi St(J^+)$ выделяет из всей совокупности исходных для выполнения рассматриваемого модуля данных те, которые противоречивы. Их определенные сочетания, задаваемые в виде векторов $USI_i(J^+)$ ($i=1, 2, \dots, N_{j^+}$), представляют ЛФР,у информацию о возможных путях снятия этих противоречий.

Рассмотрим теперь характеристики вектора J в случае его неопределенности, обозначая этот вектор через J . Ими являются наборы переменных модели, дополняя которыми тот или иной вектор J , можно получить корректный вектор J . Такого рода дополнения будем обозначать $\Delta(J)$:

$$JU\Delta(J) = J^0.$$

Несложно показать, что для каждого конкретного J его дополнение до корректного неединственно. Так, например, пусть исходная модель представлена в виде уравнений (10.30). Тогда при $J = x_i$ получим $\sigma = \{x_2, x_3, x_4, x_5, x_6\}$. Здесь ни при каком L , т. е. ни при каких сочетаниях приведенных связей, условия (10.25), (10.27) не выполняются. Следовательно, $\aleph(J)=J$. Рассмотрим следующие сочетания переменных модели $\{x_2\}$, $\{x_3\}$, $\{x_5, x_6\}$ и др. Каждое из них может быть определено как $\Delta(J)$.

Из приведенного примера следует, что размерность вектора $\Delta(J)$ может изменяться, так же как и разрешимые при каждом $\Delta(J)$ системы связей модели. Практически наибольший интерес представляют случаи, когда вектор $\Delta(J)$ выбирается из условия минимизации его размерности, а в ряде случаев и из условия доведения J до τ -полного, т. е. обеспечивающего вычисление некоторых заранее определенных переменных модели. В дальнейшем вектор $\Delta(J)$ минимальной размерности будем обозначать $\Delta(J^-)$.

Использование характеристики $\Delta(J^-)$ предусматривается, в первую очередь, при формировании расчетных моделей, когда задаваемых исходных данных для вычисления требуемой характеристики консультируемой проблемы недостаточно. В данной ситуации ЛФР,у должны быть представлены альтернативные варианты дополнения исходных данных, а именно — $\Delta(J^-)$.

И, наконец, рассмотрим характеристики вектора J в случае его корректности (обозначаемого J^0). Такими его характеристиками являются:

1) множество разрешимых при J^0 связей исходной модели, обозначаемое далее L^0 . Это множество является исходной информацией для определения порядка разрешимости входящих в него связей;

2) множество переменных модели, которые могут быть выражены через J^0 с помощью связей L^0 . Обозначим эти переменные через $\sigma(J^0, L^0)$. На их основе может быть сделан вывод о τ -полноте и частичной τ -полноте вектора J^0 , используя соответственно условия (10.28) и (10.29).

Здесь же заметим, что вне зависимости от способа доведения J^+ до J^0 , т. е. от выбранного варианта $USt(J^+)$, получающиеся множества L^0 и $\sigma(J^0, L^0)$ не отличаются. Это позволяет отнести характеристики корректного вектора и на тот случай, когда данный вектор противоречив.

Итак, рассмотрены основные характеристики вектора переменных модели J , определяемого в зависимости от рассматриваемой процедуры формирования КМ специальным образом. Эти характеристики существенным образом зависят от состава и структуры исходной модели консультируемой проблемы, которые в процессе формирования рекомендаций постоянно изменяются. В частности, переменные, описывающие уже принятые сформированные рекомендации и являющиеся для рассматриваемых рекомендаций известными константами, из рассмотрения исключаются. Это непосредственно приводит к изменениям структуры модели, выражающимся, например, в удалении из информационного графа соответствующих вершин и связанных с ними ребер.

Представим теперь получение характеристик вектора J при текущем состоянии исходной модели консультируемой проблемы как реализацию некоторого оператора KON:

$$KON : \{ \{ \Phi_i(P_i) \}_{i \in [1, N]}, J \} \Rightarrow \{ \varkappa, \eta, \xi, \beta \} \quad (10.31)$$

где

$$\varkappa = J^0 \vee J^+ \vee J;$$

$$\eta(\kappa) = \begin{cases} St(J^+), \Phi St(J^+) \text{ при } \kappa = J^+; \\ \{USt_i(J^+)\}_{i=1, 2, \dots, N_{J^+}} \\ \overline{\Delta(J^-)} \text{ при } \kappa = J^-; \\ \emptyset \text{ при } \kappa = J^0; \end{cases}$$

$$\xi(\kappa) = \begin{cases} L^0 \text{ при } \kappa = J^0 \vee J^+; \\ \emptyset \text{ при } \kappa = J^-, \end{cases}$$

$$\beta(\kappa) = \begin{cases} \sigma(J^0, L^0) \text{ при } \kappa = J^+ \vee J^0; \\ \emptyset \text{ при } \kappa = J^-. \end{cases}$$

На основании изложенного выше можно сделать вывод, что на базе оператора KON могут быть выполнены такие процедуры в процессе формирования консультационных модулей, как определение статуса КМ, гибкого формирования расчетных моделей, выявления горизонтальных информационных связей между КМ. Использование этого же оператора может быть положено также в основу выявления вертикальных информационных связей между КМ. Покажем это. Пусть известна некоторая совокупность (T) операций — $\{S_{kl}\}_{kl \in T}$, выполнение которых предшествует рассмотрению текущей — операции S_{ij} ($ij \notin T$).

Определение вертикальных информационных связей между консультационными модулями выливается непосредственно в формирование тандемных моделей. При этом операция S_{ij} вертикально связана с совокупностью детализирующих операций, объединение расчетных моделей которых образует с M_{ij} — расчетной моделью, на базе которой выполняется S_{ij} , — нижележащие уровни тандемной модели.

Сформируем вектор J , определяя его компонентами результаты выполнения операций с индексами из T . Обозначим этот вектор через J^T и без нарушения общности рассуждений будем считать его корректным. При данном векторе J^T , как результаты выполнения оператора KON, могут быть определены $L^T = L^0(J^T)$ и $\sigma(J^T, L^T)$, т. е. набор разрешимых элементарных моделей и получаемых при их разрешении переменных. Эти наборы фактически отражают ту информацию, которая получена при формировании рекомендаций в период, предшествующий выполнению рассматриваемой операции S_{ij} . Дополнительная информация такого же плана, получаемая в результате выполнения S_{ij} , может быть представлена наборами:

$$L^{\Delta_{ij}} = L^{ij} \setminus L^T; \quad \sigma^{\Delta_{ij}} = P(L^{\Delta_{ij}}) \setminus J^T,$$

где

$$P(L^{\Delta_i}) = P(L^j) \setminus P(L^T); \quad L^{ij} = L^0(J^{ij});$$

J^{ij} — вектор, компонентами которого наряду с J^+ являются результаты выполнения S^{ij} .

Если $\sigma^{\Delta_{ij}} \cap \sigma^T \neq \emptyset$, то это означает, что в составе элементарных моделей $\{m_l\}_{l \in L^{\Delta_{ij}}}$ и $\{m_l\}_{l \in L^T}$ содержатся одноименные модели.

Обозначим их некоторую пару через m_α и m_β , соответственно: $\lambda_\alpha = \lambda_\beta$.

В том случае, когда $d_\alpha \supset d_\beta$, или, в более общем случае, если найдется такая совокупность $\{m_l\}_{l \in R}$, (где $\alpha \in R, (R/\alpha) \subset T$,

$$m_\beta \notin \{m_l\}_{l \in R}), \text{ что } \bigcap_{l \in R} d_l \supset d_\beta \text{ и } \bigcup_{l \in R} d_l \not\supset d_\beta \text{ при } \forall R' \subset R, \text{ то}$$

модели m_β и $\{m_l\}_{l \in R}$ образуют тандемную модель с признаком одноименности $\varkappa = \lambda_\beta$, т. е. ${}^tM_1(\lambda_\beta) = m_\beta, {}^tM_2(\lambda_\beta) = \{m_l\}_{l \in R}$.

В результате определения тандемной модели $\{m_\beta\}, \{m_l\}$ могут быть выделены вертикальные информационные связи к консультационным модулям, в составе расчетных моделей которых присутствует m_β (в частности к КМ, выполняющему операцию S_{ij}). Эти связи к данным модулям проходят от модулей, результаты выполнения которых пересекаются с $(\bigcup_{l \in R} d_l \setminus d_\beta)$ или, что то же самое, от модулей,

расчетные модели которых содержат элементарные модели из

$$\{m_l\}_{l \in R}.$$

Заметим, что выявление тандемных моделей в процессе формирования рекомендаций может быть совмещено с выполнением других процедур. Так, наборы L^T и $\sigma(L^T, J^T)$ могут формироваться путем накопления в базе данных САК перечня элементарных моделей, использованных в расчетных моделях ранее выполненных операций, а также вычисленных на их базе переменных. Наборы L^{ij} и $\sigma(L^{ij}, J^{ij})$ могут быть определены попутно с формированием расчетных моделей для выполнения текущей операции S_{ij} .

Дадим укрупненное описание процесса формирования консультационного модуля, предназначенного для выполнения некоторой операции S_{ij} на базе оператора KON (рис. 10.7).

Пусть определена концепция выбираемых в процессе выполнения рассматриваемой операции формирования рекомендаций, и, как результат, из исходной модели консультируемой проблемы выделено

ее непрерывное подмножество $\{\Phi_i(P_i)\}_{i \in [1, N]}$, соответствующее данной концепции.

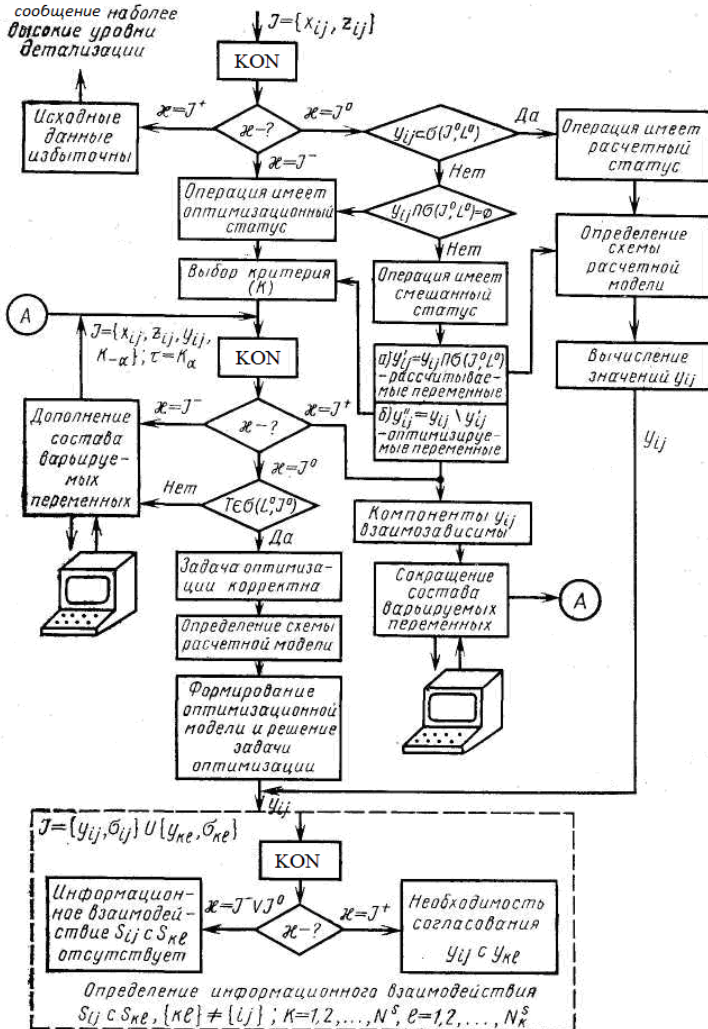


Рис. 10.7. Структура формирования консультационного модуля на базе использования оператора KON.

Первой задачей, которую требуется решить в процессе формирования КМ, является определение статуса выполняемой операции. Данная задача может быть решена путем выполнения оператора KON, для чего необходимо положить $J = \{x_{ij}, z_{ij}\}$. При этом,

- если $\kappa(J)=J$, то Y_{ij} множественно, а S_{ij} является оптимизационной;
- если $\kappa(J)=J^0$ и $y_{ij} \cap \sigma(L^0, J^0) = \emptyset$, то S_{ij} является также оптимизационной;

- если $\kappa(J) = J^0$, $y_{ij} \not\subset \sigma(L^0, J^0)$ и $J_{ij} \cap \sigma(L^0, J^0) = \emptyset$, то операция S_{ij} является смешанной, т. е. часть вектора y_{ij} , а именно $y'_{ij} = y_{ij} \cap \sigma(L^0, J^0)$ является рассчитываемой, а другая ее часть ($y_{ij} = y_{ij} \setminus y'_{ij}$) может быть определена лишь при решении оптимизационной задачи;

- если $\kappa(J) = J^0$ и $y_{ij} \subset \sigma(L^0, J^0)$, то $|Y_{ij}| = 1$, а сама операция S_{ij} является расчетной;

- если $\kappa(J)=J^+$, то это свидетельствует о взаимозависимости переменных, определяющих исходные данные для выполнения S_{ij} .

Анализ согласованности этих данных приводит либо к исключению части из них из рассмотрения, либо к определению статуса S_{ij} как невыполнимого. Каждый из рассмотренных случаев определяет соответствующую структуру дальнейших действий по формированию КМ для выполнения S_{ij} :

1) если $Y_{ij}=\emptyset$, то по характеристикам $St(J^+)$ и $\Phi St(J^+)$ можно определить число и состав противоречивых исходных данных и далее найти консультационные операции, в результате выполнения которых были получены эти данные.

Устранение имеющихся противоречий не относится к компетенции рассматриваемой операции. Ее функцией в данном случае является их выявление и сообщение об этом на более высокие уровни детализации рекомендаций, породившие выявленные противоречия;

2) если операция определена как расчетная, что, полагая $I = \{x_{ij}, z_{ij}\}$, $T=y_{ij}$, необходимо сформировать из элементарных моделей соответствующую расчетную модель и провести по ней расчет y_{ij} (для операций смешанного статуса вместо y_{ij} следует принять y_{ij}'');

3) если операция определена как оптимизационная, то требуется определить критерий (или критерии) оценки альтернативных сформированных рекомендаций и (полагая, например, при скалярном критерии $I = \{x_{ij}, z_{ij}, y_{ij}\}$, $T = \kappa_{ij}$, а при векторном критерии —

$I = \{x_{ij}, z_{ij}, y_{ij}, K_{-a}\}$, $T = \kappa_a, \kappa_a$, $\bigcup K_{-a} = K_{ij}$ (см. рис. 10.8) (для операций смешанного статуса вместо y_{ij} следует принять y_{ij}') сформировать расчетную модель, а затем модель оптимизации и выбора рекомендаций.

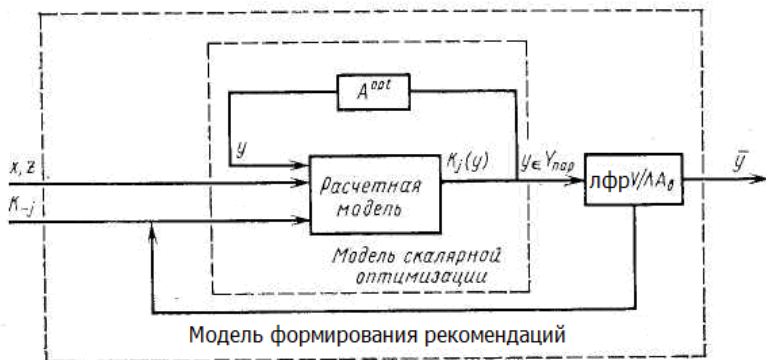


Рис. 10.8. Структура модели формирования информации для формирования рекомендаций на базе минимизации (максимизации) одной из компонент вектора критериев

При формировании расчетных моделей по заданным (I, T) также может использоваться оператор KON, в частности, для анализа корректности формируемой задачи. При оптимизационном статусе выполняемой операции возможны следующие случаи:

1) если $\varkappa = J^0$, $T \in \sigma(L^0, J^0)$, то это свидетельствует о корректности сформулированной задачи оптимизации, когда выбранные критерии и заданные ограничения согласуются с составом варьируемых переменных, в роли которых выступают компоненты вектора формируемых рекомендаций y_{ij} ;

2) если $\varkappa = J$, или $\varkappa = J^0$, и $T \notin \sigma(L^0, J^0)$, то это свидетельствует о несогласованности состава вектора y_{ij} с выбранным критерием оптимизации. В данном случае необходимо расширить число варьируемых переменных. Данное расширение $\Delta(J)$, в силу его неоднозначности, должно производиться под управлением ЛФР,ов. В то же время, исходя из условия минимизации размерности формируемой задачи оптимизации, наиболее целесообразным является определение $\Delta(J^-)$;

3) если $\varkappa = J^+$, то это свидетельствует о взаимозависимости компонент x_{ij} , z_{ij} , y_{ij} . Поскольку статус операции был определен ранее как оптимизационный, то взаимозависимость x_{ij} , z_{ij} , исключается, что свидетельствует о взаимозависимости компонент y_{ij} . В этом случае необходимо сокращение числа варьируемых переменных, что можно сделать, используя такую генерируемую оператором KON характеристику, как ΦSt , представляющую собой перечень

взаимозависимых переменных. Выделяемый ЛФР,ом из ΦSt перечень переменных — USt определяет выводимые из числа варьируемых компоненты y_{ij} , значения которых далее должны определяться расчетным путем.

Теперь рассмотрим решение задачи по определению информационных связей между консультационными модулями на базе использования оператора KON . Для определения связи между консультационными модулями, выполняющими операции S_{ij} и S_{kl} , необходимо определить вектор J в виде $J=I_{ij} \cup I_{kl}$, и если $\kappa(J)=J \vee J^0$, то информационная связь между рассматриваемыми операциями отсутствует. В противном случае данные операции находятся в информационном взаимодействии и результаты их автономного выполнения требуют согласования. Число $St(J^+)$ определяет степень их информационной связности, а $\Phi St(J^+)$ — состав «общих» переменных, используемых при выполнении этих операций.

При этом каждый из вариантов $USt(J^+)$, генерируемых оператором KON , определяет параметры из $\Phi St(J^+)$, которыми должны обмениваться взаимодействующие операции. Конкретный из этих вариантов должен выбрать ЛФР.

Итак, выполнение рассмотренных процедур формирования консультационных модулей может быть обеспечено путем реализации введенного оператора KON . Выходные переменные этого оператора, интерпретированные каждый раз в зависимости от выполняемой функции соответствующим образом, представляют ЛФР,ам информацию, обеспечивающую выполнение ими работ по формированию консультационных модулей для автоматизированного выполнения возлагаемых на них операций.

10.1.2.3. Постановка основной задачи, решаемой при формировании консультационных модулей для их автономного функционирования

Сформулируем задачу, решение которой на базе математической модели сетевой структуры, интерпретируемой двудольным неориентированным графом, обеспечивает получение выходных переменных оператора KON .

Введем в рассмотрение граф $G(\Phi, \sigma)$, получающийся из информационного графа исходной модели путем удаления из него вершин — переменных, соответствующих компонентам вектора J . Подмножества этого графа, определяемые вершинами-отношениями с

индексами из L и связанными с ними вершинами-переменными из σ , будем обозначать через $G(L)$, $L \subseteq [1, N]$.

Введем также в рассмотрение дефицит множества вершин $\Phi(L) = \{\Phi_i\}_{i \in L}$ двудольного графа $G(L)$, обозначая его через $d(G(L))$ и определяя в виде:

$$d(G(L)) = |\Phi(L)| - |\sigma(L, J)|,$$

или с учетом взаимоднозначного соответствия между $\Phi(L)$ и L :

$$d(G(L)) = |L| - |\sigma(L, J)|,$$

Покажем, что реализация оператора KON в целом может быть сведена к нахождению $\bar{L} \subseteq [1, N]$, на котором значение дефицита достигает максимума:

$$\bar{L} = \arg \max_{L \subseteq [1, N]} d(G(L)). \quad (10.32)$$

Действительно:

1) если $d(G(\bar{L})) = 0$, то это свидетельствует о том, что условие (10.27) не имеет места при $\forall L \subseteq [1, N]$, а при $L = \bar{L}$ соблюдается условие (10.25). В данном случае можно сделать вывод о корректности J , т. е.

$$\kappa(J) = J^0; L^0 = \bar{L}; \sigma(L^0, J^0) = \sigma(\bar{L}, J);$$

2) если $d(G(\bar{L})) < 0$, то это свидетельствует о недоопределенности J , т. е. $\kappa(J) = J$. При этом полученная система $\{\Phi_i(P_i)\}_{i \in \bar{L}}$ — является минимально недоопределенной и можно утверждать, что минимальное дополнение J до J^0 содержит — $d(G(\bar{L}))$ элементов, т. е.

$$|\Delta(J^-)| = -d(G(\bar{L})).$$

Конкретное определение элементов $\Delta(J^-)$ в силу их неоднозначности должно производиться ЛФР,ом. Причем их правильное задание должно приводить к увеличению значения $d(G(\bar{L}))$, что и может быть положено в основу оценки целенаправленности действий ЛФР,ов и формирования им «подсказки»;

3) если $d(G(\bar{L})) > 0$, то это свидетельствует о переопределенности вектора J , т. е. $\kappa(J) = J^+$:

$$St(J^+) = d(G(\bar{L}));$$

$$\Phi St(J^+) = J \cap P(\bar{L}).$$

При этом в состав $USt(J^+)$ вводятся элементы, уменьшающие значение $d(G(\bar{L}))$. Ввиду неоднозначности элементов множеств $\{USt_i(J^+)\}_{i=1,2,\dots,NJ^+}$, конкретное их определение должно производиться ЛФР,ами. Правильность их действий, а также формирование в случае необходимости для них «подсказки» определяются по изменениям величины $St(J^+)$.

Таким образом, путем решения задачи (10.32) возможно определение характеристик вектора J , заданного тем или иным образом. Соответственно, в основе оператора KON, на который возлагаются функции по генерированию этих характеристик, лежит решение этой задачи.

Данная задача может рассматриваться как задача нахождения максимальных паросочетаний на двудольном графе.

В основе предлагаемого метода лежит условие, согласно которому множество одноименных вершин двудольного графа может быть разделено на два непересекающихся подмножества, на одном из которых (критическом множестве, возможно, пустом) функция дефицита достигает неотрицательного максимума, а на другом (множестве без дефицита) — функция дефицита всегда отрицательна.

Данное условие применительно к рассматриваемой задаче свидетельствует о том, что после выделения из множества вершин отношений информационного графа модели подмножества с индексами из \bar{L} , на котором функция дефицита достигает неотрицательного максимума, на оставшейся части вершин функция дефицита всегда отрицательна. Тогда, обозначив через L_N множество индексов вершин-отношений исходного графа ($L_N = \{1, 2, \dots, N_m\}$), а через L^* множество без дефицита — дополнение \bar{L} до L_N , полагая

$$d(G(\bar{L})) \geq 0, \tag{10.33}$$

можно записать

$$d(G(L^*)) < 0, \tag{10.34}$$

где

$$L^* = L_N \setminus \bar{L}.$$

В дальнейшем при рассмотрении \bar{L} будем учитывать как ранее введенное его определение в виде (10.32), так и условие (10.33). В случае, если $\bar{L} = \emptyset$, то это свидетельствует о том, что значение $d(G(\bar{L}))$ в условии (10.32) отрицательно, а вектор J недоопределен.

Итак, в основу предлагаемого метода положено членение исходного множества L_N на два непересекающихся подмножества \bar{L} и L^* :

$$\begin{aligned}\bar{L} \cap L^* &= \emptyset \\ \bar{L} \cup L^* &= L_N,\end{aligned}$$

удовлетворяющих соответственно условиям (10.32) и (10.33), (10.34).

Отличительной чертой данного метода является целенаправленность формирования множеств \bar{L} и L^* , удовлетворяющих приведенным выше условиям. Эта целенаправленность заключается в том, что исходный граф $G(L_N)$ представляется в виде совокупности подмножеств определенных свойств (классов). Подмножества отдельных классов при этом однозначно определяются как входящие в $G(\bar{L})$. Подмножества других классов анализируются и либо относятся к $G(L^*)$, либо также безусловно переходят в $G(\bar{L})$. Такая логика действий, основанная на последовательном отделении от исходного множества L_N подмножеств, безусловно принадлежащих \bar{L} , обеспечивает алгоритмам, основанным на использовании предлагаемого метода, линейную вычислительную трудоемкость их реализации в зависимости от размерности исходного графа.

В основе формируемого метода лежат:

- 1) условия, которые позволяют на каждой паре подмножеств (L_1 и L_2 , $L_1 \cup L_2 = L_N$, $L_1 \cap L_2 = \emptyset$) определить их эквивалентность \bar{L} и L^* ;
- 2) определение классов подмножеств исходного графа со специальными свойствами;
- 3) операции анализа подмножеств различных классов с целью безусловного отнесения этих подмножеств или их частей к \bar{L} .

Рассмотрим условия, позволяющие на каждой паре подмножеств L_1 и L_2 проверить их соответствие условиям (10.32)—(10.34), т. е. определить эквивалентность этих подмножеств критическому множеству и множеству без дефицита.

Введем в рассмотрение остов графа модели $G(L_N) = \{\Phi(L_N), \sigma(L_N)\}$ и будем его обозначать $g(L_N) = \{\Phi(L_N), \sigma_{L_N}\}$. Данный остов представляет собой древовидный двудольный граф с вершинами-связями $\Phi(L_N)$ и вершинами-переменными σ_{L_N} . Здесь каждое множество $\sigma_L (L \leq L_N)$ представляет собой совокупность вершин-переменных, соединенных с вершинами-связями на древовидном графе $g(L)$.

Определим на множестве вершин-отношений остовного графа функцию дефицита:

$$d(g(L))=|L|-\sigma_L \text{ при } \forall L \subseteq L_N. \quad (10.35)$$

Поскольку из определений для $G(L)$ и $g(L)$ следует, что

$$\sigma(L) \supseteq \sigma_L \text{ при } \forall L \subseteq L_N \quad (10.36)$$

то

$$d(g(L)) \geq d(G(L)) \text{ при } \forall L \subseteq L_N. \quad (10.37)$$

Приступим теперь непосредственно к формированию названных условий. Они следуют из следующего утверждения.

Если на исходном графе $G(L_N)=\{\Phi(L_N), \sigma(L_N)\}$ построена пара основных деревьев $g(L_1)=\{\Phi(L_1), \sigma_{L_1}\}$ и $g(L_2)=\{\Phi(L_2), \sigma_{L_2}\}$:

$$L_1 \cup L_2 = L_N; \quad (10.38)$$

$$L_1 \cap L_2 \neq \emptyset; \quad (10.39)$$

$$\sigma_{L_1} \cup \sigma_{L_2} = \sigma_{L_N} = \sigma(L_N); \quad (10.40)$$

$$\sigma_{L_1} \cap \sigma_{L_2} = \emptyset, \quad (10.41)$$

таких, что

$$d(g(L')) < 0 \text{ при } \forall L' \subseteq L_1 \quad (10.42)$$

$$d(g(L_2)) \geq 0; \quad (10.43)$$

$$\sigma_{L_1} \cap \sigma(L_2) = \emptyset, \quad (10.44)$$

то

$$L_1 = L^*; \quad (10.45)$$

$$L_2 = \bar{L}. \quad (10.46)$$

Покажем справедливость данного утверждения.

Предположим, что указанная пара остовов построена. Тогда из неравенств (10.37) и (10.43) непосредственно следует, что $d(G(L')) < 0$ при $\forall L' \subseteq L_1$, т. е. L_1 — есть множество без дефицита L^* .

Из выражений (10.40), (10.41), (10.44) и очевидного условия

$$\sigma(L_1) \cup \sigma(L_2) = \sigma(L_N),$$

следующего, в частности, из соотношений (10.36) и (10.40), можно получить

$$\sigma_{L_2} = \sigma(L_2)$$

Тогда $d(g(L_2))=d(G(L_2))$ и согласно неравенству (10.37) $d(G(L_2)) \geq 0$, т. е. L_2 является критическим множеством \bar{L} .

Теперь покажем, что если условия (10.45), (10.46) выполняются, то построенные на L_1 и L_2 остовы всегда удовлетворяют условиям (10.42)—(10.43).

По определению, при условиях (10.44), (10.45) $d(G(L')) < 0$ при $\forall L' \subseteq L_1$ и $d(G(L_2)) \geq 0$.

Из неравенств (10.37) и $d(G(L_2)) \geq 0$ непосредственно следует, что $d(g(L_2)) \geq 0$.

На L_1 будем строить остов, исходя из условия максимизации его дефицита, т. е. вершины $\sigma(L_1) \cap \sigma(L_2)$ будем считать отнесенными в $g(L_2)$:

$$\sigma(L_1) \cap \sigma(L_2) \subset \sigma_{L_2}; \quad (10.47)$$

$$\sigma_{L_1} = \sigma(L_1) \setminus \sigma(L_2). \quad (10.48)$$

Предположим, что в данных условиях найдется подмножество $L' \subseteq L$, на котором $d(g(L')) \geq 0$. Данное предположение равносильно, учитывая (10.47), тому, что $d(G(L_2 \cup L')) \geq d(G(L_2))$. Это противоречит исходному условию (10.45), что делает данное предположение неверным и доказывает истинность неравенства (10.42).

Таким образом, задача разделения исходного множества L_N на множество вершин без дефицита — L^* и критическое множество — \bar{L} , т. е. на множества, удовлетворяющие условиям (10.32)—(10.34), может быть сведена к построению на исходном графе пары остовов, удовлетворяющих условиям (10.38)—(10.44).

Само по себе построение пары остовов на исходном графе модели, удовлетворяющих условиям (10.38)—(10.43), не представляет проблем. Однако в общем случае здесь не будет выполняться условие (10.44) (рис. 10.9). На достижение выполнения этого условия и направлен рассматриваемый метод.

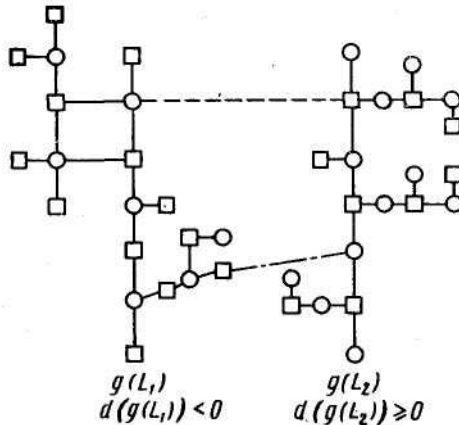


Рис. 10.9. Членение остова информационного графа на критическое множество и множество без дефицита (— — — — ребра, удовлетворяющие условию (10.44); — . — . — — ребра, не удовлетворяющие условию (10.44); □ — вершины-переменные; ○ — вершины-связи)

В основу предлагаемого формирования $g(L_1)$ и $g(L_2)$ положим их представление как совокупностей остовов определенных, рассматриваемых ниже классов.

10.1.2.4. Основные операторы решения задачи

Представим остов $g(L_N)$ исходного графа модели в виде некоторой совокупности остовов $\{g_q(\Phi_q, \sigma_q)\}_{q \in Q}$:

$$\left. \begin{aligned} \bigcup_{q \in Q} \Phi_q &= \Phi(L_N); \\ \Phi_j \cap \Phi_i &= \emptyset \quad (i \neq j; i, j \in Q); \\ \bigcup_{q \in Q} \sigma_q &= \sigma_{L_N}; \\ \sigma_i \cap \sigma_j &= \emptyset \quad (i \neq j; i, j \in Q). \end{aligned} \right\} \quad (10.49)$$

При этом ребра между этими остовами при их построении вначале рассматривать не будем.

Обозначим терминальные вершины некоторого остова $g_q(\Phi_q, \sigma_q)$ через $T_q = \{t_{qi}\}_{i=1,2,\dots,N_{qt}}$ (где N_{qt} — множество терминальных вершин на g_q), а вершины, являющиеся точками ветвления на этом дереве, — через $V_q = \{v_{qi}\}_{i=1,2,\dots,N_{qv}}$ (где N_{qv} — множество точек ветвления на g_q).

Введем в рассмотрение совокупности остовов следующих четырех классов:

1) совокупность остовов класса A , обозначаемую далее

$$g_A(\Phi_A, \sigma_A) = \{g_{ai}\}_{i \in N_A}, \text{ у которых:}$$

- все терминальные вершины соответствуют переменным модели ($T_A \subseteq \sigma_A$);
- ветвление происходит только в вершинах, соответствующих связям модели ($V_A \subseteq \Phi_A$);

2) совокупность остовов класса B , обозначаемую далее

$$g_B(\Phi_B, \sigma_B) = \{g_{bi}\}_{i \in N_B}, \text{ у которых:}$$

- все терминальные вершины соответствуют переменным модели ($T_B \subseteq \sigma_B$);
- ветвление происходит только в вершинах, соответствующих переменным модели ($V_B \subseteq \Phi_B$);

3) совокупность остовов класса C , обозначаемую далее,

$$g_C(\Phi_C, \sigma_C) = \{g_{ci}\}_{i \in N_C}, \text{ у которых:}$$

- все терминальные вершины, кроме одной, соответствуют переменным модели ($|T_C \cap \Phi_C| = 1$);
- ветвление происходит только в вершинах, соответствующих переменным модели ($V_C \subseteq \sigma_C$);

4) совокупность остовов класса D , обозначаемую далее

$$g_D(\Phi_D, \sigma_D) = \{g_{di}\}_{i \in N_D}, \text{ у которых:}$$

- функция дефицита всегда нулевая ($d(g_D) = 0$);
- переменные, входящие в состав связей Φ_D , не содержатся в g_A и g_B :

$$\sigma(L_D) \cap (\sigma_A \cup \sigma_B) = \emptyset. \quad (10.50)$$

Определим значения функции дефицита на остовах различных классов. Как следует из построения g_{a_i} , значение функции дефицита на остовах этого класса всегда отрицательно и, в частности,

$$d(g_{a_i}) = 1 - N_{a_i t} \quad (i = 1, 2, \dots, N_A).$$

При этом дефицит на g_A , представляющем объединение g_{a_i} ($i = 1, 2, \dots, N_A$), определяется в виде

$$d(g_A) = \sum_{i=1}^{N_A} d(g_{a_i}) = N_A - \sum_{i=1}^{N_A} N_{a_i t} \quad (10.51)$$

Из определения остовов классов B, C, D следует, что

$$d(g_{b_i}) = -1 \quad (i=1, 2, \dots, N_B); \quad (10.52)$$

$$d(g_B) = \sum_{i=1}^{N_B} d(g_{b_i}) = -N_B; \quad (10.53)$$

$$d(g_{c_i}) = 0 \quad (i=1, 2, \dots, N_C); \quad (10.54)$$

$$d(g_C) = \sum_{i=1}^{N_C} d(g_{c_i}) = 0; \quad (10.55)$$

$$d(g_{d_i}) = 0; \quad (10.56)$$

$$d(g_D) = \sum_{i=1}^{N_D} d(g_{d_i}) = 0. \quad (10.57)$$

Итак, значения функции дефицита на остовах классов A и B всегда отрицательны, а на остовах классов C и D — нулевые. Данные свойства позволяют рассматривать объединение σ_A и g_B как подмножества множества $g(L^*)$:

$$g_A, g_B \subseteq g(L^*),$$

а остовы класса D , учитывая эквивалентность условий (10.44) и (10.50), как подмножества $g(\bar{L})$:

$$g_D \subseteq g(\bar{L}).$$

Остовы класса C далее рассматриваются как переходные. В частности, если на них выполняется условие

$$\sigma(L_C) \cap (\sigma_A \cup \sigma_B) = \emptyset, \quad (10.58)$$

то они переходят в состав g_D . Если же это условие не выполняется, то они трансформируются в подмножества g_B, g_A и может быть g_D .

Представление графа исходной модели совокупностью остовов названных классов позволяет предложить следующую направленность действий в процессе формирования критического множества \bar{L} . На графе исходной модели несложно построить совокупность остовов классов A и B таких, что все вершины-переменные исходного графа войдут в их число. Эту совокупность остовов примем за начальное (нулевое) приближение:

$$\{\sigma^{(0)}_A, \sigma^{(0)}_B\} = \sigma(L_N)$$

При этом, в общем случае, найдутся вершины-отношения исходного графа, которые не войдут в $g^{(0)}_A, g^{(0)}_B$. Обозначим множества таких вершин через $\Phi^{(0)}$:

$$\Phi^{(0)} = \Phi(L_N) / \{\Phi^{(0)}_A, \Phi^{(0)}_B\}.$$

Основные операции метода в таком случае могут быть сведены к последовательному объединению вершин из $\Phi^{(0)}$ с одним из имеющихся остовов. Причем выполнение этих операций должно быть определено таким образом, чтобы в результате получались остовы лишь названных классов. По мере получения остовов класса C на них проверяется условие (10.58), в результате чего они переходят в классы A, B или D . Таким образом, в начале выполнения каждого очередного шага исходный граф представляется остовами трех классов — A, B, D .

Тогда, как результат выполнения конечного числа ($|\Phi^{(0)}|$) таких операций определяются g_A и g_B , задающие множество без дефицита L^* , а также g_D , определяющее критическое множество \bar{L} . Здесь же отметим, что при выполнении каждого очередного шага сформированное множество g_D может лишь пополняться за счет перехода в него отдельных частей из g_A, g_B , т. е. вершины, попавшие в состав остовов класса D , в g_A и g_B перейти не могут. Данное утверждение следует непосредственно из определения остовов класса D и условия (10.49).

Определим процедуры объединения вершин из Φ^H с остовами различных классов, составляющие основные операции рассматриваемого метода.

Некоторая вершина $\Phi_i \in \Phi^H$ может быть объединена с остовом g_{ai} если $\sigma_j \cap \sigma_{q_i} \neq \emptyset$. В зависимости от того, с какими классами остовных деревьев может быть объединена та или иная вершина, введем в рассмотрение следующие подмножества Φ^H :

$$\Phi^A = \{\Phi^a : \Phi^a \in \Phi^H; |\sigma^a \cap \sigma_A| > 1\}; \quad (10.59)$$

$$\Phi^{AB} = \{\Phi^{ab} : \Phi^{ab} \in \Phi^H; \sigma^{ab} \cap \sigma_A \neq \emptyset; \sigma^{ab} \cap \sigma_B \neq \emptyset\}; \quad (10.60)$$

$$\Phi^B = \{\Phi^b : \Phi^b \in \Phi^H; |\sigma^b \cap \sigma_B| > 1\}; \quad (10.61)$$

$$\Phi^D = \{\Phi^d : \Phi^d \in \Phi^H; \sigma^d \subseteq \sigma_D\}; \quad (10.62)$$

$$\Phi^E = \{\Phi^e : \Phi^e \in \Phi^H; |\sigma^e \setminus \sigma_D| = 1\}, \quad (10.63)$$

где σ^q — вектор переменных, входящих в связь Φ^q ($q = a, b, ab, d, e$).

Учитывая, что каждая связь модели имеет в своем составе не менее двух переменных, а также промежуточный характер остовных деревьев класса C , нетрудно отметить, что Φ^H всегда может быть представлено его приведенными подмножествами, причем, как правило, неоднозначно.

Учитывая приведенную классификацию элементов Φ^H и изложенный выше метод, определим набор операций по выделению из $g(L_N)$ его подмножеств, удовлетворяющих условиям (10.38)—(10.46). При этом рассмотрим все операции, представляющие практический интерес, имея в виду, что при реализации конкретных алгоритмов можно использовать не все из этих операций.

Первой рассмотрим операцию объединения некоторой вершины Φ^a с g_A . Эта операция может выполняться различными способами в зависимости от положения элементов:

$$n^a = \sigma^a \cap \sigma_A$$

в структуре g_A .

В частности, особого вида операций требуют элементы Φ^a , переменные которых пересекаются больше, чем с одной терминальной вершиной из g_A . Обозначим такого рода элементы Φ^{At} :

$$\Phi^{At} = \{\Phi^{at} : |n^a \cap T_A| > 1\}. \quad (10.64)$$

В данном случае на паре (g_A, Φ^{at}) всегда может быть построен остов класса A . В основе этого построения лежит:

1) объединение с помощью Φ^{at} двух остовов g_{ai} и g_{aj} , если $n^a \cap T_{ai} \neq \emptyset$ и $n^a \cap T_{aj} \neq \emptyset$ (рис. 10.10, а).

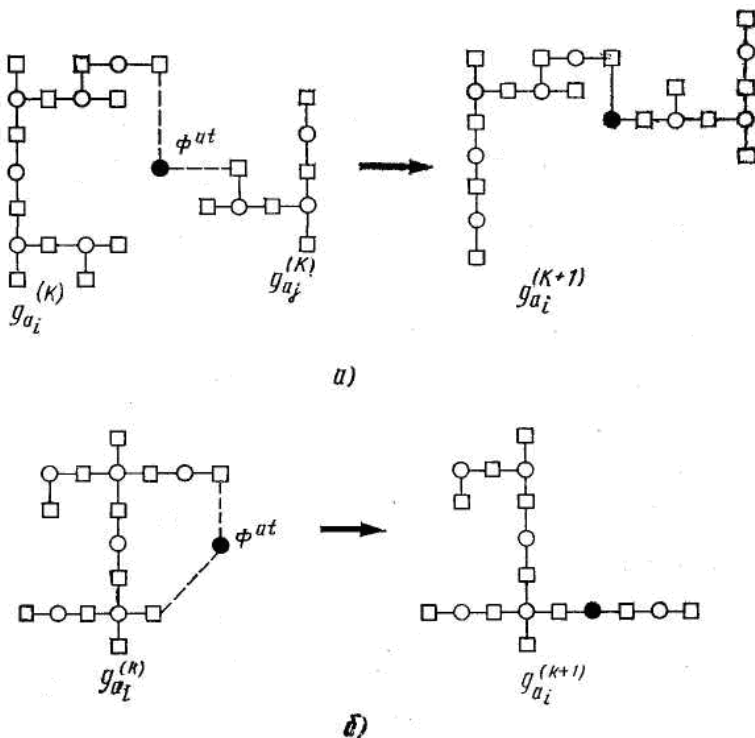


Рис. 10.10. Выполнение оператора
 $KOT(a) \ n^a \not\subseteq T_{a_i}; \ \sigma \cdot n^a \subseteq T_{a_i}$

При этом число остовов в g_A уменьшается на единицу, а число терминальных вершин — на две вершины. В итоге значение функции дефицита на g_A , как следует из формулы (10.51), увеличивается на единицу;

2) трансформация g_{a_i} если $n^a \subseteq T_{a_i}$, которая заключается в присоединении к одной из терминальных вершин через Φ^{at} части g_{a_i} начиная с его другой терминальной вершины до ближайшей точки ветвления (рис. 10.10, б). При этом число остовов в g_A остается без изменения, а число терминальных вершин в g_{a_i} и, соответственно, в g_A уменьшается на единицу. В итоге значение функций дефицита на g_A увеличивается на единицу.

Таким образом, операция объединения g_A и Φ^{al} приводит лишь к трансформации g_A и может быть представлена (см. рис. 10.10) в виде некоторого оператора:

$$\text{КОТ: } \{g^{(k)}_A, \Phi^{al}\} \Rightarrow g^{(k+1)}_A, \quad (10.65)$$

где k — номер приближения.

Операция объединения элементов из Φ^A , переменные которых не пересекаются с двумя и более терминальными вершинами из g_A , требуют разделения этих элементов на две группы. К первой из них, обозначаемой Φ^{Al} , будем относить элементы, все переменные которых пересекаются с g_A на одном из линейных участков некоторого g_{ai} . Обозначим множество индексов вершин-отношений, лежащих на таких линейных участках, через

$$l^j_{ai} \subset g_{ai} \ (l^j_{ai} \cap V_{ai} = \emptyset, j=1, 2, \dots, N_{l_{ai}}),$$

где $N_{l_{ai}}$ — число линейных участков в g_{ai} . Тогда Φ^{Al} можно представить в виде:

$$\Phi^{Al} = \{ \Phi^{Al} : \sigma^{al} \subseteq l^j_{ai}; g_{aj} \in g_A; j \in [1, N_{l_{ai}}] \} \quad (10.66)$$

Ко второй группе, обозначаемой далее Φ^{A0} , будем относить все оставшиеся связи из Φ^A :

$$\Phi^{A0} = \Phi^A \setminus \Phi^{Al} \setminus \Phi^{Al} \quad (10.67)$$

Операция объединения g_A с Φ^{Al} приводит к расчленению остова g_{ai} на два, каждый из которых представляет собой подмножество g_{ai} , лежащее по одну или другую сторону от l^j_{ai} , начиная от ближайшей к l^j_{ai} точки ветвления. Другими словами, в данном случае происходит удаление из g_{ai} линейного участка l^j_{ai} . При этом множество l^j_{ai} , объединенное с Φ^{al} , представляет собой остов класса C (рис. 10.11, *а*). В частном случае, когда линейный участок содержит терминальную вершину, т.е. $l^j_{ai} \cap T_{ai} \neq \emptyset$, из g_{ai} также удаляется линейный участок, который совместно с Φ^{al} образует остов класса C (рис. 10.11, *б*). В итоге в остове g_{ai} уменьшается число терминальных вершин, в результате чего значение $d(g_A)$ увеличивается на единицу.

В общем случае операция по объединению g_A с Φ^{al} может быть представлена в виде некоторого оператора:

$$\text{КОЛ: } \{g_A^{(k)}, \Phi^{al}\} \Rightarrow \{g_A^{(k+1)}, g_C\}. \quad (10.69)$$

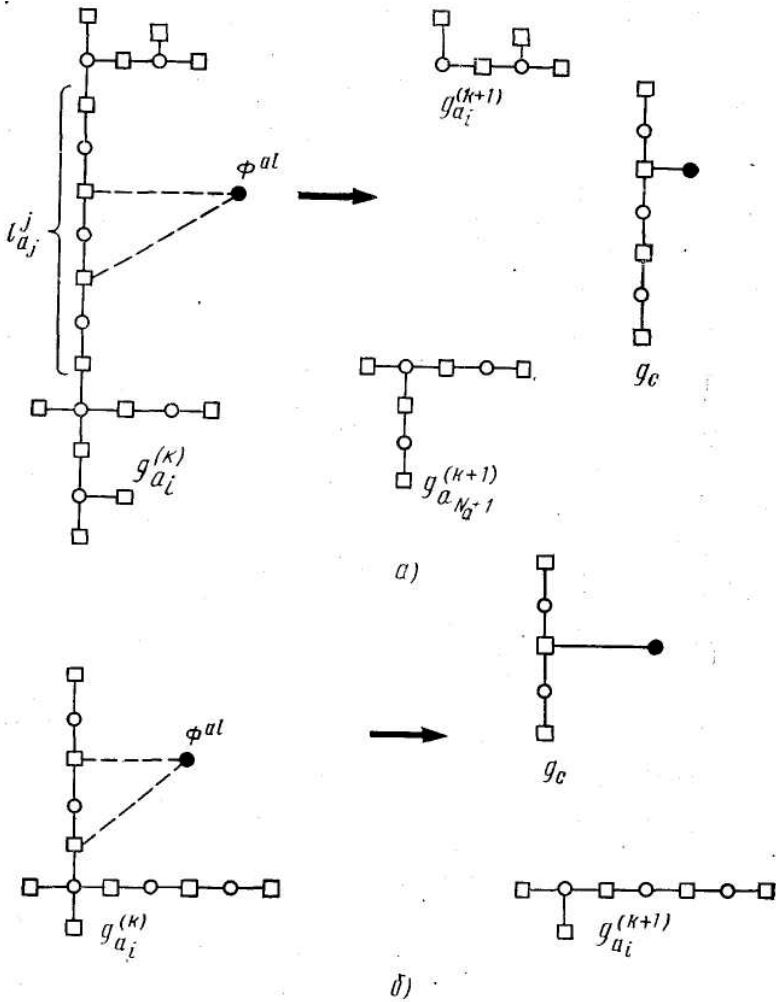


Рис. 10.11. Выполнение оператора
 KOL (а) $t_{a_i}^j \cap \tau_{a_i} = \emptyset$; б) $t_{a_i}^j \cap \tau_{a_i} \neq \emptyset$)

При объединении g_A с Φ^{al} происходит как трансформация исходного остова g_A , так и порождение дополнительного подмножества в g_B . Данная трансформация заключается в выделении из g_A пары линейных

участков, содержащих переменные из σ^A , которые объединяются с помощью Φ^{a0} и образуют остов класса B (рис. 10.12).

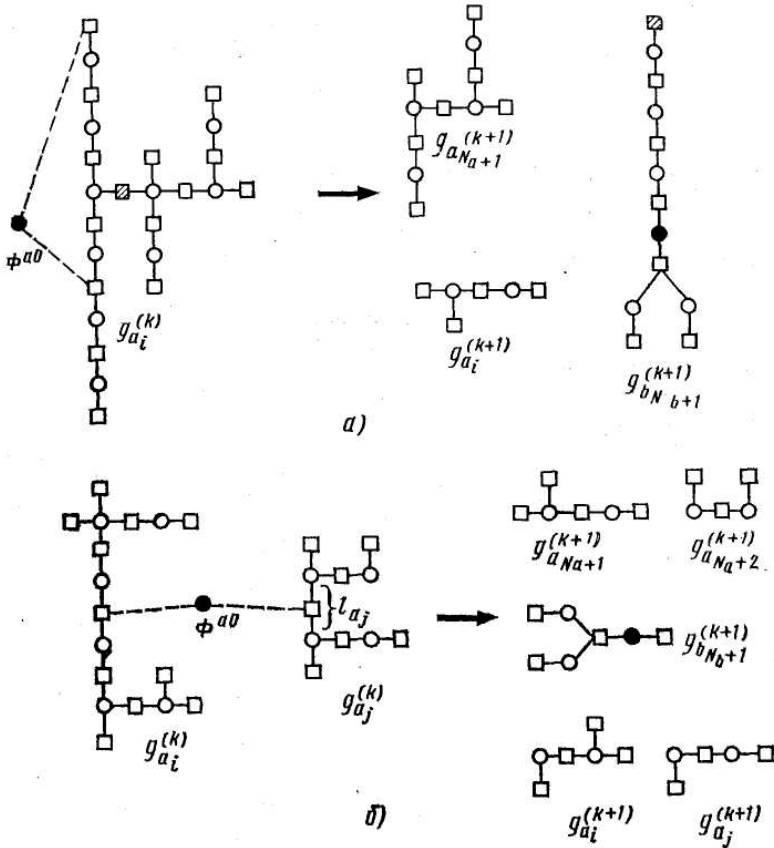


Рис. 10.12. Выполнение оператора
 $KOO(a) \sigma^{a0} \subseteq g_{a_i}; \sigma^{a0} \subseteq g_{a_j}$

Как уже отмечалось, удаление из g_A линейного участка приводит к увеличению значения $d(g_A)$ на единицу. В рассматриваемом случае из $d(g_A)$ выделяются два таких участка и порождается остов класса B , имеющий, по определению, дефицит, равный -1 , т. е. в итоге

суммарное значение функции дефицита на имеющихся остовах увеличивается на единицу.

Рассмотренную операцию по объединению g_A с Φ^{a0} можно представить как реализацию некоторого оператора:

$$\text{КОО} : \{g^{(k)}_A, \Phi^{a0}\} \Rightarrow \{g^{(k+1)}_A, g^{(k+1)}_B\}. \quad (10.69)$$

Таким образом, рассмотренные операции (10.65), (10.68), (10.69) полностью определяют действия по объединению элементов из Φ^A с имеющимися остовами g_A .

Теперь перейдем к рассмотрению операции объединения элементов Φ^{AB} с остовами классов A и B . Объединение Φ^{AB} с g_A и g_B приводит к следующей трансформации: из g_{ai} выделяется линейный участок l_{ab} содержащий $\tilde{\sigma} = \sigma_A \cap \sigma^{AB}$, и он с помощью Φ^{ab} присоединяется к g_B в вершине, соответствующей одной из переменных σ_{ab} . При этом дефицит g_A уменьшается на единицу и, если $l_{ai} \cap T_{ai} \neq \emptyset$, g_{ai} расщепляется на два подмножества (рис. 10.13).

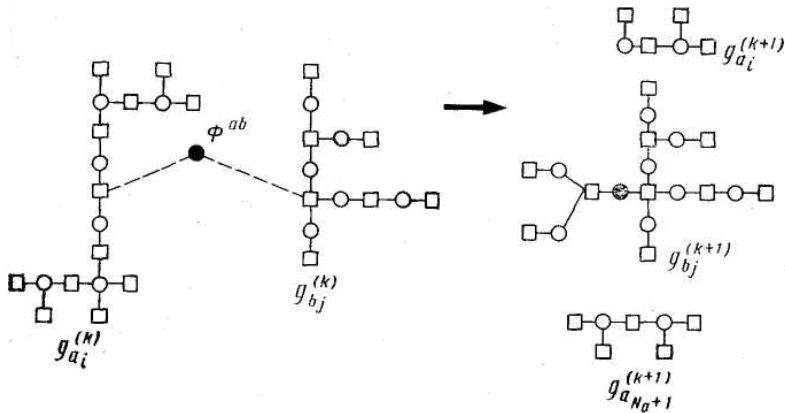


Рис. 10.13. Выполнение оператора КОВ

Данная операция может быть представлена как реализация некоторого оператора:

$$\text{КОВ} : \{g^{(k)}_A, g^{(k)}_B, \Phi^{ab}\} \Rightarrow \{g^{(k+1)}_A, g^{(k+1)}_B\}. \quad (10.70)$$

При проведении объединения Φ^b с g_B возможны два случая. Первому из них соответствует условие

$$(\sigma^b \cap \sigma_B) \subseteq g_{b_i} \quad (10.71)$$

согласно которому все содержащиеся в некотором отношении $\Phi^b \in \Phi^B$ переменные находятся в одном из остовов класса B . Далее множество такого рода вершин из Φ^B будем обозначать Φ^{B1}

Во втором случае рассматриваются все оставшиеся в Φ^B элементы, совокупность которых обозначим:

$$\Phi^{B0} = \Phi^B \setminus \Phi^{B1}. \quad (10.71)$$

Операция по объединению Φ^{B1} с g_B приводит к выделению из g_B одного остова, удовлетворяющего условию (10.71), который в совокупности с Φ^{B1} образует остов класса C (рис. 10.14).

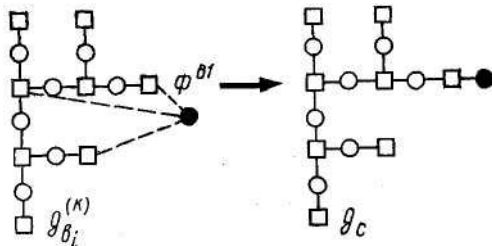


Рис. 10.14. Выполнение оператора KB1

При этом число остовов класса B уменьшается на единицу и, соответственно, на единицу увеличивается суммарный дефицит имеющихся остовов, так как дефицит вновь появившегося g_C по определению, равен нулю. Данная операция может быть представлена как реализация оператора:

$$KB1 : \{g^{(\kappa)}_B, \Phi^{B1}\} \Rightarrow \{g^{(\kappa+1)}_B, g_C\}. \quad (10.73)$$

В случае объединения Φ^{B0} с g_B , а именно: некоторых g_{bj} и g_{bi} , в состав которых входят переменные из $\sigma^{b0} \cap \sigma_B$, — трансформация этих остовов заключается в их объединении с помощью Φ^{b0} (рис. 10.15).

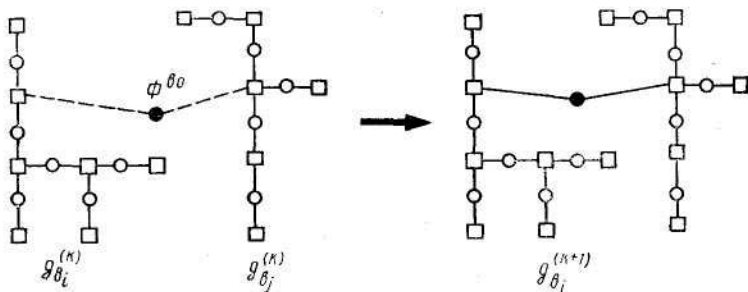


Рис. 10.15. Выполнение оператора KB0

При этом число остовов класса B уменьшается на единицу. Выполнение данной операции представим как реализацию некоторого оператора:

$$\text{КВО} : \{g^{(k)}_B, \Phi^{b0}\} \Rightarrow g_B^{(k+1)} \quad (10.74)$$

Приступим к определению операции объединения Φ^E с имеющимися остовами. По определению Φ^E и g_D , выражающийся в том, что остовы g_C по мере их появления, сразу «расформировываются» на деревья других классов, данного рода связи могут быть объединены лишь с g_A и g_B . При этом, по определению, Φ^e может иметь лишь одну переменную в составе g_A или g_B . Соответственно, трансформации в данном случае подвергается один из остовов $g_{ai} \in g_A$ или $g_{bj} \in g_B$, определение которого производится по условию:

$$g_{a_j} \wedge g_{b_i} \cap \sigma^l \neq \emptyset. \quad (10.75)$$

В зависимости от получаемых при этом результатов элементы из Φ^E будем обозначать Φ^{eA} или Φ^{eB} .

Трансформация g_{ai} заключается в выделении из g_{ai} линейного участка, содержащего переменную из σ^e , который в совокупности с Φ^{ea} образует остов класса C (рис. 10.16).

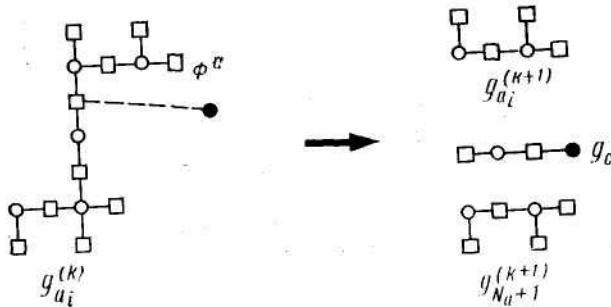


Рис. 10.16. Выполнение оператора КЕА

При этом значение $d(g_A)$ увеличивается на единицу. Данную трансформацию далее будем рассматривать как реализацию оператора:

$$\text{КЕА} : \{g^{(k)}_A, \Phi^{ea}\} \Rightarrow \{g_A^{(k+1)}, g_C\} \quad (10.76).$$

Объединение g_{bj} с Φ^{eb} представляет в результате остовное дерево класса C (рис. 10.17) и может рассматриваться как реализация оператора:

$$\text{КЕВ} : \{g^{(k)}_B, \Phi^{eb}\} \Rightarrow \{g_B^{(k+1)}, g_C\} \quad (10.77).$$

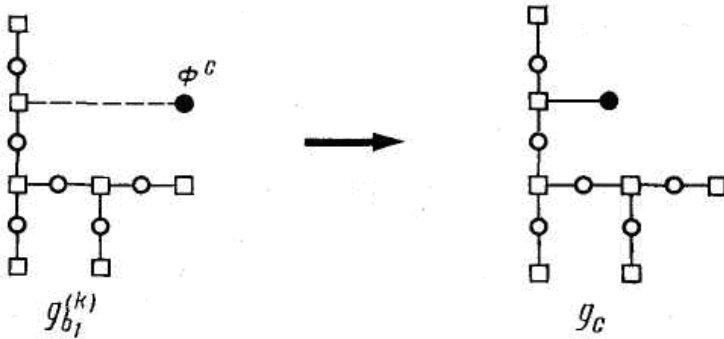


Рис. 10.17. Выполнение оператора КЕВ

Заметим, что остовные деревья класса D в рассмотренных случаях никак не фигурируют. Эти деревья получаются как результат анализа остовов класса C на выполнение условия (10.41), когда в роли $g(L_2)$ выступает g_c , а в роли $g(L_1)$ — объединение g_A и g_B . В результате этого анализа g_c расщеляется на подмножества класса B и D (рис. 10.18).

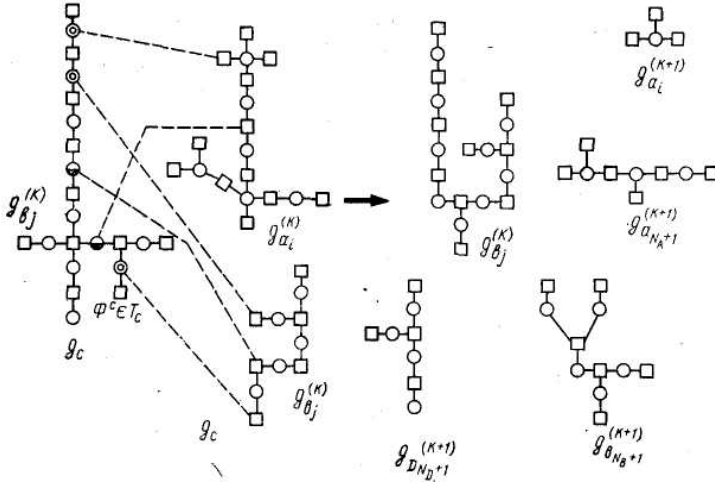


Рис. 10.18. Выполнение оператора КС (●- элементы Φ^6 ;

○- элементы Φ^6)

Данное членение будем производить следующим образом:

1) вначале определим на g_c вершины-связи Φ^6 такие, что содержащиеся в них переменные пересекаются с g_A или g_B , т. е.

$$\tilde{\Phi}^C = \{\tilde{\Phi}^c : \tilde{\Phi}^c \in g_C; \tilde{\sigma}^C \cap (g_A \cup g_B) \neq \emptyset\};$$

2) выделим из $\tilde{\Phi}$ вершины, каждая из которых является ближайшей к единственной на g_C терминальной вершине-связи; обозначим и $\mathcal{B}^{\mathcal{A}}$

$$\mathcal{B}^{\mathcal{A}} \subseteq \mathcal{B}^{\mathcal{B}};$$

3) каждая ветвь из g_C , начинающаяся с вершины-связи из $\mathcal{B}^{\mathcal{A}}$ и лежащая по другую сторону от терминальной вершины-связи, отделяется от g_C , и если переменные, содержащиеся в $\mathcal{B}^{\mathcal{A}}$, пересекаются с g_B , то «пристыковывается» к g_B в соответствующей вершине; если переменные, содержащиеся в $\mathcal{B}^{\mathcal{B}}$, пересекаются с g_A , то образует совместно с линейным участком из $g_{ai} \in g_A$, содержащим соответствующую вершину, остов класса B ;

4) далее, учитывая, что при выполнении действий 1)–3) множество переменных, входящих в остовы класса B , пополняется, процедура повторяется, начиная с первого действия.

В результате выполнения приведенной итерационной процедуры оставшееся от g_C подмножество, т. е. не перешедшее в состав g_B , удовлетворяет условию (10.51) и пополняет остовы класса D . Рассмотренная операция может быть представлена как реализация следующего оператора:

$$KC : \{g_C, g_A^{(k)}, g_B^{(k)}\} \Rightarrow \{g_B^{(k+1)}, g_D^{(k+1)}, g_A^{(k+1)}\}. \quad (10.78)$$

Приведенные операции (10.65), (10.68)–(10.70), (10.73), (10.74), (10.76)–(10.78) обеспечивают выполнение объединения всех, кроме Φ_D , элементов множества Φ^H с исходным графом модели, представленным остовными деревьями рассмотренных классов. Вершины Φ^D согласно их определению могут быть объединены исключительно с остовами класса D . Соответствующие операции будут рассмотрены в п. 10.1.2.5.

10.1.2.5. Структура оператора KON

Структура оператора KON может быть представлена тремя процедурами. Первая из них связана с максимизацией функции дефицита или, что то же самое, с выделением из информационного графа модели критического множества. В результате выполнения этой процедуры определяется \aleph — характеристика, свидетельствующая о корректности, недоопределенности или избыточности анализируемого вектора J . Вторая и третья процедуры связаны с преобразованиями переопределенного и недоопределенного, соответственно, вектора J в корректный или, что то же самое, с определением таких характеристик,

как ΦSt , USt и $\overline{\Delta J}$. Остальные генерируемые оператором KON данные: $\sigma(L^0, J^0)$, L^0 — перечень переменных, которые могут быть определены через компоненты вектора J , и связи модели, через которые это определение должно быть произведено, получаются как результат формирования критического множества с нулевым дефицитом в результате выполнения указанных процедур.

Вначале приведем процедуру, связанную с выделением из исходного графа модели критического множества \overline{L} (рис. 10.19). В основе этой процедуры лежат рассмотренные выше операторы по объединению вершин связей из Φ^h с классифицированными остовами, представляющими информационный граф исходной модели.

На начальном этапе (на первом шаге) этот граф представим остовами класса A , что, очевидно, можно сделать всегда, т. е.

$$\begin{aligned} \sigma_A^{(0)} &= \sigma(L_N); \\ g_B^{(0)} &= g_D^{(0)} = g_C^{(0)} = \emptyset \\ \Phi^{h(0)} &= \Phi(L_N) \setminus \Phi_A^{(0)}. \end{aligned}$$

Задачу построения остова $g_A^{(0)}$ возложим на некоторый оператор:

$$KOST : G(L_N) \Rightarrow \{g_A^{(0)}, \Phi^{h(0)}\},$$

выполнение которого является началом исполнения рассматриваемого алгоритма выделения критического множества \overline{L} .

На каждом последующем (k -м) шаге, как правило, считаются известными $g_A^{(k)}$, $g_B^{(k)}$, $g_D^{(k)}$. Выполнение этого шага начинается с выбора элемента из $\Phi^{h(k)}$, обозначаемого $\Phi^{*(k)}$:

$$\Phi^{*(k)} = \Phi^{h(0)} \setminus \bigcap_{i=1}^{k-1} \Phi^{*(i)}.$$

Выбранный элемент Φ^* далее классифицируется по условиям (10.59)—(10.63), (10.64), (10.66), (10.67), (10.71), (10.72), (10.75) и далее в зависимости от полученных результатов выполняется соответствующий из рассмотренных выше операторов.

Согласно приведенным выше условиям классификации элементов $\Phi^* \in \Phi^h$ эти элементы в ряде случаев могут быть отнесены к различным классам. Например, условия (10.59) и (10.70) не исключают отнесение одного и того же элемента как к Φ^A , так и к Φ^B .

Отнесение $\Phi^* \in \Phi^h$ к тому или иному классу определяет последующую операцию его объединения с соответствующими остовами. При этом в результате выполнения одних операций остовы класса C порождаются, а других — нет. Порождение g_C требует выполнения оператора KC , т. е. проведения дополнительных операций, целью которых является выделение из g_C подмножеств g_D .

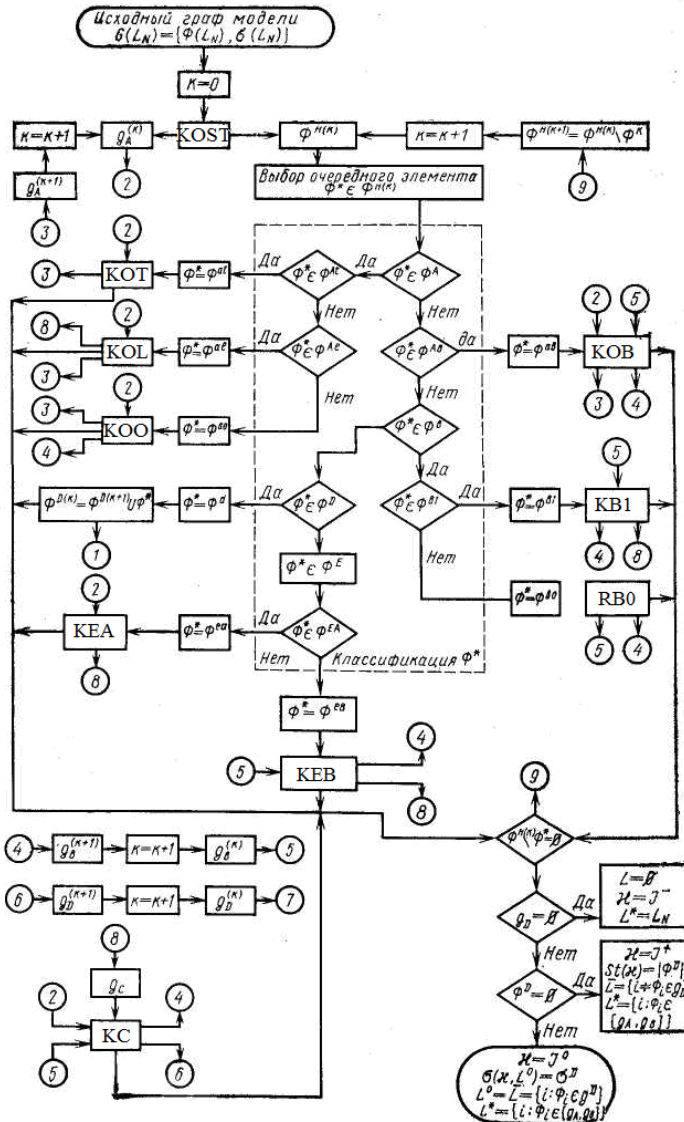


Рис. 10.19. Структура процедуры выделения критического множества из исходного графа модели

Если некоторый элемент Φ^* может быть отнесен к двум классам и в одном случае соответствующая операция его объединения с тем или иным остовом порождает g_C , а в другом нет, то из порожденного g_C выделение подмножества g_D невозможно. Данное утверждение основывается:

1) на показанной в п. 10.1.2.3 единственности результатов решения рассматриваемой задачи, т. е. единственности \bar{L} ;

2) на безусловности включения остовов в класс D (подмножества g_D в рассматриваемом методе перейти в состав остовов других классов не могут)

3) на условии $g_D \subseteq g(\bar{L})$.

Учитывая сказанное, можно дать рекомендации о наиболее целесообразной последовательности проведения анализа $\Phi^* \in \Phi^H$ на принадлежность тому или иному классу. Эта последовательность должна быть таковой, чтобы отнесение Φ^* к классам (Φ^{B1} , Φ^{eB} , Φ^{eA} , Φ^{ab}), приводящим к порождению элементов g_C , производилось в последнюю очередь.

Отнесение рассматриваемого элемента $\Phi^* \in \Phi^H$ к тому или иному классу определяет конкретную операцию его объединения с тем или иным остовом. В результате ее выполнения происходит изменение отдельных остовов, в том числе возможно появление остовов класса C . В случае появления остова класса C на нем проверяется возможность выделения из него элементов g_D , что сводится к реализации оператора KC .

Таким образом, на каждом шаге производится объединение одного элемента из Φ^B с известными остовами, в результате чего эти остовы трансформируются соответствующим образом, оставаясь в рамках названных классов A, B, D .

Условием окончания рассматриваемой процедуры является условие $\Phi^H^{(k+1)} = \Phi^D$, свидетельствующее о «присоединении» всех вершин из Φ^H к остовам g_A и g_B . Итогом при этом являются совокупности остовов g_A, g_B, g_D , а также вершины Φ^D , связанные исключительно с g_D . При этом

$$\begin{aligned} \bar{L} &= \{i: \Phi_i \in g_D \cup \Phi^D\}; \\ L^* &= \{i: \Phi_i \in g_A \cup g_B\}; \\ d(G(\bar{L})) &= |\Phi^D| \\ \sigma(L^0, J^0) &= \{\sigma_i: \sigma_i \in g_D, \sigma^d\}. \end{aligned}$$

Верхняя оценка вычислительной трудоемкости выполнения алгоритмов, базирующихся на предложенном методе, в общем случае может быть определена в виде

$$T = |\Phi^{H(0)}| S |P^{\max}|,$$

где T — верхняя оценка трудоемкости; $|\Phi^{H(0)}|$ — число шагов в алгоритме; S — число переменных в $g_A^{(0)}$; $|P^{\max}|$ — максимальное число переменных, являющихся аргументами в $\Phi_i \in \Phi^H$.

Отметим, что $|P^{\max}|$ является величиной, не зависящей от размерности исходной модели, под которой принято понимать число вершин в информационном графе. Также практически не зависит от размерности модели и величина $|\Phi^{H(0)}|$, которая может трактоваться как максимальная переопределенность вектора J . В итоге можно считать, что $T = KS$, где $K = |\Phi^{H(0)}| |P^{\max}|$ — коэффициент, не зависящий от размерности исходной модели.

Теперь рассмотрим процедуры, которые обеспечивают получение таких выходных переменных оператора KON, как ΦSt и USt при $\varkappa = J^+$, а также ΔJ^- при $\varkappa = J$. Основным назначением этих переменных, как уже указывалось, является обеспечение целенаправленной корректировки вектора J в случае, если он некорректен или недоопределен. Каждому из данных случаев соответствует либо отрицательное, либо положительное максимальное значение функции дефицита на двудольном информационном графе модели. Определение $USt \subset J^+$ и $\Delta J \subset \sigma(L_N)$ при этом связано с уменьшением или увеличением, соответственно, числа компонент в исходном векторе J таким образом, чтобы вновь получаемое критическое множество \bar{L} было, во-первых, не пусто, а во-вторых, имело нулевой дефицит.

Вектор переменных ΦSt в качестве своих компонент содержит взаимозависимые переменные из J . Он может быть определен в виде

$$\Phi St (J^+) = J^+ \cap P(\bar{L}).$$

Рассмотрим основные операции, связанные с формированием USt . Эти операции требуют выполнения при $d(G(\bar{L})) > 0$, т. е. когда $\Phi^D \neq \emptyset$ и $g_D \neq \emptyset$, где Φ^D и g_D определяются на рассмотренном выше этапе выполнения оператора KON.

Как было указано в п. 10.1.2.2, множество $USt(J^+)$ может быть определено различными комбинациями из $St(J^+)$ компонент вектора ΦSt . В то же время не любые $St(J^+)$ компонент определяют USt . Учитывая неоднозначность определения компонент USt , ниже полагается, что инициатива выбора той или иной из них принадлежит ЛФР,у, который последовательно выводит из состава вектора J^+ отдельные его компоненты. Задачей в данном случае является установление факта истинности принадлежности очередной такой компоненты допустимому USt с учетом ранее выведенных компонент или — при организации работ по принципу «меню» — сообщение

ЛФР, у списка компонент вектора J^+ , допускающих их выведение из состава этого вектора, который изменяется по мере выбора ЛФР, от той или иной компоненты.

При известных g_D и Φ^D компоненты вектора J^+ могут быть разделены на две группы (здесь и далее для упрощения записей будем считать, что $\Phi St = J^+$). К первой из этих групп, обозначаемой $J^\Phi \subseteq J^+$, будем относить компоненты, содержащиеся в составе переменных, входящих в отношения из Φ^D . Обозначим эти переменные через P^D , а остальные компоненты вектора J^Φ — через J^* , тогда

$$J^\Phi = \{J^* \cdot J^* \cap P^D \neq \emptyset\}.$$

Ко второй группе, обозначаемой J^D , отнесем все оставшиеся компоненты вектора J^* :

$$J^D = J^+ \setminus J^\Phi$$

Выведение из состава J^+ компонент J^D и J^Φ имеет принципиальное отличие. Так, если некоторая компонента вектора J^+ принадлежит J^Φ , то она всегда может быть включена в состав USt . Ее выведение из состава J^+ приводит к образованию дополнительного остова класса D , состоящего из этой переменной и связи — элемента множества Φ^D , — в состав переменных которой она входит. Получающийся таким образом остов пополняет g_D . Определим такого рода операцию как выполнение некоторого оператора $KJ\Phi$:

$$KJ\Phi : \{\Phi^D, J^{*\Phi}\} \Rightarrow g_D^{(k+1)}.$$

Переменные из J^D в ряде случаев не могут быть выведены из состава J^+ . Возможность их выведения определяется следующими условиями.

Выведение из вектора J^+ компонент J^D и пополнение ими вектора σ приводит к тому, что один остов g_D переходит в класс B . Как следствие, оставшиеся остовы класса D переходят в класс C . Проверка выполнения на вновь полученных множествах g_B и g_C условия (10.51) и реализация операции по трансформации этих множеств таким образом, чтобы данное условие выполнялось, приводит к построению остовов классов B и D . При этом, если имеется возможность объединения полученного остова g_B хотя бы с одним элементом из Φ^D , то рассматриваемая компонента вектора J^D может быть включена в состав USt , а в противном случае она не включается.

Структура алгоритма определения множеств USt может быть представлена следующими действиями, связанными с определением принадлежности некоторой компоненты J^* вектора J^+ множеству USt :

1) проверкой условия $J^* \in J^\Phi$, и если оно выполняется, то реализуется оператор $KJ\Phi$;

2) если $J^* \in J^D$, то выделение из g_D остова класса B как реализация оператора KC , где в роли g_C выступает g_D при $g_B = J^*$; $g_A = \emptyset$;

3) проверкой условия (10.51) — наличия пересечения переменных из вновь получаемого остова g_B с переменными, содержащимися в $P^D \setminus J^*$. Если такое пересечение пусто, то $J^* \notin USt$, в противном случае $J^* \in USt$.

После введения очередной компоненты J^* в USt происходит переход к рассмотрению следующей компоненты из J^+ и т. д. Реализацию процедуры по формированию множества USt будем рассматривать как выполнение оператора:

$$KSt: \{g_D, \Phi^D, J^+, Pr\} \Rightarrow USt,$$

где Pr — лицо, формирующее рекомендации.

Теперь рассмотрим процедуру формирования множества $\overline{\Delta J}$. Оно представляет собой минимальную совокупность переменных модели, дополнение которыми некоторого недоопределенного J позволяет получить корректный вектор:

$$J^0 = J \cup \overline{\Delta J^-}.$$

Причем здесь заранее предполагается, что выбор $\overline{\Delta J}$ производится из условия τ -полноты вектора J^0 :

$$\sigma(J^0, L^0) \ni \tau,$$

т. е. из условия обеспечения вычисления некоторой заранее определенной переменной τ , в частности, критерия оценки сформированных рекомендаций.

Принципиально, вычисление некоторой переменной τ может быть произведено на каждой из элементарных моделей, содержащих ее в своем составе. Обозначим индексы такого рода моделей через I_τ :

$$I_\tau = \{i: P_i \cap \tau \neq \emptyset\}.$$

Рассмотрим векторы:

$$J_i^c = J \cup (P_i \setminus \tau); \quad i \in I_\tau,$$

каждый из которых, если его рассматривать в качестве исходных данных, очевидно, является достаточным для вычисления τ . В то же время некоторые из этих векторов могут быть переопределены. Проведя их анализ на базе использования оператора KL , можно получить значения $\{St(J_i^c)\}_{i \in I_\tau}$. Величина $|P_i| - St(J_i^c)$ задает

размерность $\overline{\Delta J^-}$ при $\forall i \in I_\tau$ позволяет определить индекс элементарной модели, по которой возможно вычисление τ при минимальной величине $|\Delta J^-|$. Обозначим этот индекс i^* .

Следующим шагом определения вектора $\overline{\Delta J}$ является назначение его компонент, содержащихся в P_{i^*} . Если $St(J_i^c) = 0$, то очевидно

$\overline{\Delta J} = P_{i^*}$. В противном случае $\overline{\Delta J}$ может быть определено на основе рассмотренного выше оператора KSt при $J^+ = J_{i^*}$.

В целом процедура определения минимального дополнения недоопределенного ΔJ до корректорного может быть представлена как выполнение оператора:

$$KJ : \{J, \tau, Pr\} \Rightarrow \overline{\Delta J}.$$

Таким образом, выполнение оператора KON может быть представлено как последовательное выполнение составляющих его операторов KL, KSt, KJ .

В заключение остановимся на способах формирования последовательности разрешения связей $\{\Phi_i\}_{i \in L^0}$, определяющих в совокупности замкнутую корректную систему. Данная задача заключается в разложении $\{\Phi_i\}_{i \in L^0}$ на упорядоченную совокупность замкнутых систем минимальной размерности. Так, например, на рис. 10.20 показана замкнутая система из шести связей с шестью неизвестными переменными.

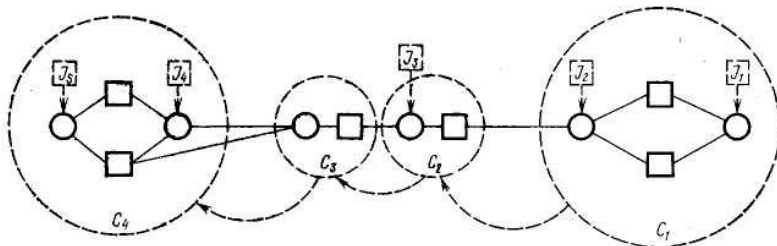


Рис. 10.20. Разложение корректной системы связей модели на упорядоченную совокупность замкнутых систем минимальной размерности.

Ее разрешение может быть сведено к разрешению вначале системы C_1 (из двух связей с двумя переменными), далее систем C_2 и C_3 (каждая из которых состоит из одной связи и одной переменной) и, наконец, системы C_4 (из двух связей с двумя переменными).

В настоящее время методы и алгоритмы выделения из $\{\Phi_i\}_{i \in L^0}$ упорядоченной совокупности замкнутых систем (называемых также минимально-замкнутыми системами) являются достаточно разработанными. Наиболее эффективным из них, на наш взгляд, является алгоритм, в основе которого лежит выделение на графе

модели сильносвязанных компонент. Верхняя оценка трудоемкости реализации данного алгоритма является линейной функцией от размерности исходного графа. Выполнение этого алгоритма, которому должно предшествовать выполнение оператора KON, далее будем определять оператором KFOR.

Итак, в рассмотренном разделе показано, что для разработки методов и алгоритмов формирования консультационных модулей является целесообразным представление структуры используемых математических моделей в виде сети, интерпретируемой двудольным неориентированным графом. В таком случае ряд процедур формирования КМ может быть формально обобщен таким образом, что выполнение каждой из этих процедур сводится к решению единой для всех них задачи. В основе решения этой задачи лежит анализ на полноту и взаимозависимость компонент некоторого вектора J , формируемого при рассмотрении той или иной процедуры специальным образом. Введенный в рассмотрение оператор KON позволяет упорядочить информацию, получаемую в процессе такого анализа, и дать ей соответствующую практическую интерпретацию в зависимости от того, какая из процедур формирования выполняется на базе этого оператора в текущий момент времени.

10.1.3. Формирование среды обмена информацией между КМ (при вертикальном взаимодействии)

10.1.3.1. Постановка задачи вертикального согласования

Формирование среды взаимодействия КМ производится после того как выявлены КМ, между которыми имеются информационные связи и, соответственно, результаты автономного функционирования которых должны быть согласованы. При этом подразумевается, что выявлены взаимосвязанные параметры, описывающие консультационные операции, выполняемые этими КМ. Данным параметрам при определении горизонтальных связей, в частности, соответствуют множества переменных ΦSt (см. п. 10.1.2.2). Из них ЛФР_{ми} выделяются подмножества $USt \subseteq \Phi St$, описывающие те параметры, по которым непосредственно предполагается проводить согласование результатов автономного функционирования взаимодействующих КМ. При определении вертикальных связей требующими согласования параметрами являются: для КМ более высокого уровня параметры z , которым соответствуют переменные v , содержащиеся в используемой этим модулем модели, а для КМ более

низких уровней параметры, по которым формируются рекомендации в процессе автономного функционирования этих модулей.

Среда взаимодействия КМ включает в себя две определяющие составляющие: **модель**, трансформирующую результаты автономного функционирования одних КМ в исходные данные для других и наоборот; **методику** обмена информацией между взаимодействующими КМ.

Определение первой из этих составляющих было описано в разделах 4 и 10.1.1. В частности, модель, позволяющая «пересчитать» результаты автономного функционирования одного КМ в исходные данные для другого, при их горизонтальном взаимодействии может быть определена в виде $L^0(J) \setminus (L^* \cup L^{**})$ (где $J = I^* \cup I^{**}$; I^* , I^{**} — входные переменные для расчетных моделей рассматриваемой пары КМ, а L^* , L^{**} — индексы элементарных моделей, составляющих эти расчетные модели; $L^0(J)$ — индексы всех вычисляемых отношений исходной модели консультируемой проблемы). При вертикальном взаимодействии в роли такого рода модели выступает **модель идентификации**. Таким образом, при формулировании среды взаимодействия КМ остается решить вопрос **о методике обмена информацией между взаимодействующими КМ**. Выбор такой методики, как и вообще методик при формировании КМ, является в общем случае прерогативой ЛФР. Данные методики отражают различные правила поиска компромисса. **Одна из таких методик может быть проиллюстрирована рис. 10.21.**

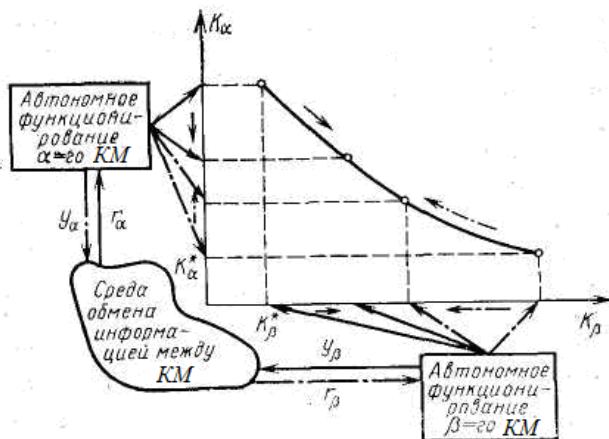


Рис. 10.21. Пример процедуры согласования результатов автономного функционирования КМ

Здесь в координатах K_α, K_β (где K_α, K_β — критерии оценки рекомендаций, формируемых некоторыми α -м и β -м консультационными модулями) в процессе поиска компромисса строится множество парето-оптимальных рекомендаций методом уступок. Причем каждая рекомендация из этого множества представляет собой результат автономного функционирования одного из КМ при фиксированных значениях реакций от другого модуля. Эти значения определяются данным модулем из условий минимизации собственных вынужденных потерь, т. е. отклонений от тех значений критерия (K^*), которые могли бы быть им получены, если бы отсутствовало взаимодействие с другим КМ. В процессе поиска компромисса КМ функционируют и порождают парето-оптимальные точки параллельно, обмениваясь информацией о возможных уступках в значениях выбираемых параметров и оценивая каждый раз, во что выливаются эти уступки. Наиболее целесообразными, на наш взгляд, являются **методики обмена информацией между КМ, когда эта информация представляет собой не «точечные» значения согласуемых параметров, а множества возможных значений этих параметров.** В этом случае функционирование каждого КМ происходит в условиях неопределенности. К работе в таких условиях аппарат выбора сформированных рекомендаций должен быть специально приспособлен.

В ряде случаев итерационный обмен «точечными» значениями согласуемых параметров не дает требуемый результат. Таким случаем, в частности, является обмен информацией между вертикально взаимодействующими консультационными модулями в процессе согласования результатов их автономного функционирования. **Рассмотрим эту задачу и предложим метод обмена информацией в процессе согласования результатов автономного функционирования вертикально связанных КМ, основанный на том, что информация от КМ нижнего уровня отражает множественность возможных формируемых рекомендаций этих модулей.**

Необходимость проведения вертикального согласования порождается наличием вертикальных информационных связей между консультационными операциями. Такого рода связи согласно п. 10.1.2 могут быть выделены путем формирования тандемных моделей. При этом согласованию подлежат результаты выполнения, с одной стороны, некоторой операции S_{ij} , а с другой — совокупностей операций, объединения расчетных моделей которых образуют с M_{ij} тандемную модель. Каждая такая пара представляет собой уровни детализации консультируемой проблемы по параметру консультируемой

проблемы, соответствующему переменной — признаку одноименности рассматриваемой тандемной модели. Причем эта же модель, учитывая приведенную в п. 10.1.2.3 структуру тандемных моделей, может входить в состав более общей модели, но уже с другим признаком одноименности, где M_{ij} является частью одного из ее уровней и т. д. Таким образом, если определена некоторая N -уровневая модель

$${}^tM(\mathbf{x}) = \{ {}^tM_1(\mathbf{x}), {}^tM_2(\mathbf{x}), \dots, {}^tM_N(\mathbf{x}) \},$$

то каждому ее уровню может быть поставлена в соответствие совокупность консультационных операций, образующих уровень детализации консультируемой проблемы по параметру \mathbf{x} :

$${}^tM_i(\mathbf{x}) \Leftrightarrow S_i \quad (i = 1, 2, \dots, N). \quad (10.79)$$

При этом состав S_i определяется консультационными операциями, вертикально связанными с операциями из S_{i-1} , и т. д.

Допустим, что операция верхнего уровня ($S_i \equiv S_{i,1}$) носит оптимизационный характер и сводится к решению задачи скалярной оптимизации по некоторому критерию K . Тогда признак одноименности рассматриваемой модели определяется этим критерием:

$$\mathbf{x} \Leftrightarrow K. \quad (10.80)$$

Переменные каждого уровня тандемной модели, определяемые разностью $d_i \setminus d_{i-1}$ соответствуют результатам выполнения операций 1-го уровня:

$$d_i \setminus d_{i-1} \Leftrightarrow y_i. \quad (10.81)$$

Директивные данные x_i для рассматриваемого уровня соответствуют вектору d_{i-1} , содержащему все данные о результатах выполнения операций предшествующих уровней, необходимые для выполнения S_i :

$$d_i \setminus d_{i-1} \Leftrightarrow x_i \quad (i = 1, 2, \dots, N, d_0 \Leftrightarrow x_0 = x_1). \quad (10.82)$$

Реакции z , как уже неоднократно отмечалось, соответствуют переменным v :

$$v_i \Leftrightarrow z_i \quad (i = 1, 2, \dots, N). \quad (10.83)$$

Определение этих переменных возможно на базе решения задачи идентификации, представляемой как реализация некоторого оператора Q :

$$v_i = Q(d_{i+1} \setminus d_i, v_{i+1}) \quad (i = 1, 2, \dots, N-1; v_N \Leftrightarrow z_N = z_0). \quad (10.84)$$

Структура взаимодействия уровней тандемной модели может быть представлена в виде, показанном на рис. 10.22.

Обозначим оператор модели ${}^tM_i(\mathbf{x})$ через F_i и представим его в виде:

$$F_i : \{ d_{i-1}, d_i \setminus d_{i-1}, v_i \} \Rightarrow \mathbf{x} \quad (i = 1, 2, \dots, N), \quad (10.85)$$

а для модели нижнего уровня:

$$F_N : \{ d_0, d_N \setminus d_0, v_N \} \Rightarrow \mathbf{x}, \quad (10.86)$$

где

$$d_N \setminus d_0 = \prod_{i=1}^N (d_i \setminus d_{i-1}) \Leftrightarrow \bigcup_{i=1}^N y_i = y_0. \quad (10.87)$$

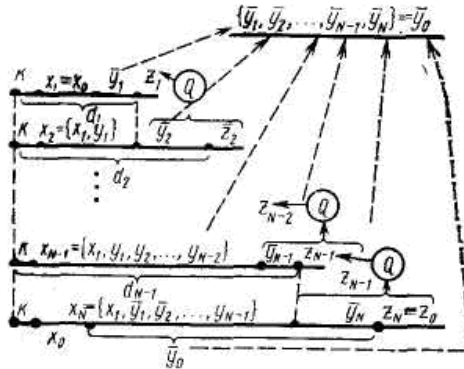


Рис. 10.22. Структура взаимодействия уровней тандемной модели

Тогда решение задачи

$$\overline{(d_i \setminus d_{i-1})} = \arg \min_{d_i \setminus d_{i-1}} F_i(d_{i-1}, d_i \setminus d_{i-1}, v_i) \quad (10.88)$$

является тождественным выполнению i -го уровня детализации, а задача

$$\overline{(d_N \setminus d_0)} = \arg \min_{d_N \setminus d_0} F_N(d_0, d_N \setminus d_0, v_N) \quad (10.89)$$

эквивалентна исходной задаче формирования рекомендаций.

Таким образом, решение задачи формирования рекомендаций по схеме последовательной детализации может быть представлено как последовательность автономного решения задач (10.88) на базе различных уровней тандемной модели с учетом условий (10.84), порождающих необходимость согласования получающихся при этом результатов. Исходной задаче формирования рекомендаций в таком случае соответствует решение задачи на модели нижнего уровня тандемной модели. Условием того, что формирование рекомендаций по схеме последовательной детализации (включая и проведение согласования) корректно является следующее:

$$\overline{(d_N \setminus d_0)} = \bigcup_{i=1}^N \overline{(d_i \setminus d_{i-1})}. \quad (10.90)$$

Основной проблемой при реализации процесса формирования рекомендаций по схеме последовательной детализации является организация процедур вертикального согласования. Их основным назначением является обеспечение такого выбора реакций z_i (или v_i) при выполнении каждого i -го уровня детализации, чтобы в итоге было выполнено условие (10.90). Как правило, при организации процедур вертикального согласования по традиционной итерационной схеме данное условие из рассмотрения выпускается, заменяясь условиями:

$$F_i(x_i, \bar{y}_i, z_i) = F_j(x_j, \bar{y}_j, z_j) \quad (i \neq j; i, j=1, 2, \dots, N). \quad (10.91)$$

Эти условия означают совпадение результатов вычислений на базе моделей различных уровней в точках, соответствующих автономно выбираемым рациональным формируемым рекомендациям.

Здесь и далее, учитывая соответствия (10.79)—(10.83), модель i -го уровня (10.85) можно представить:

$$F_i : \{x_i, y_i, z_i\} \Rightarrow K$$

а вектора x_i :

$$x_i = \{x_{i-n}, y_{i-n+1}, y_{i-n+2}, \dots, y_{i-1}\} \quad (i = 1, 2, \dots, N, n=1, 2, \dots, i-1).$$

Покажем, что условия (10.91) не являются достаточными для обеспечения выполнения условия (10.90) или тождественного ему:

$$\prod_{i=1}^N \bar{y}_i = \bar{y}_0, \quad (10.92)$$

где \bar{y}_i определяют из решения задачи на модели i -го уровня; \bar{y}_0 находят полностью из решения задачи на модели нижнего уровня.

Рассмотрим двухуровневую задачу, решаемую на базе пары моделей $M_1 : K = F_1(x_1, y_1, z_1)$ и $M_2 : K = F_2(x_1, y_1, y_2, z_2)$ при заданных значениях x_1 и z_2 .

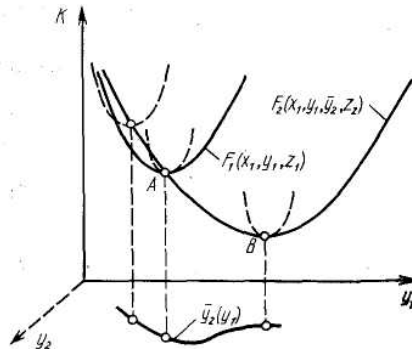


Рис. 10.23. Взаимное расположение графиков функций, соответствующих различным уровням тандемной модели

Из рис. 10.23, иллюстрирующего взаимное расположение графиков этих моделей, видно, что в точке A происходит совпадение результатов вычислений значений K на моделях M_1 и M_2 при $\{x_1, \bar{y}_1\}$ и

$\{x_1, \bar{y}_1, \bar{y}_2\}$, где

$$\bar{y}_1 = \arg \min_{y_1} F_1(x_1, y_1, z_1); \quad \bar{y}_2 = \arg \min_{y_2} F_2(x_1, \bar{y}_1, y_2, z_2).$$

В точке A соблюдаются все условия получения результатов решения задач на моделях различных уровней и достигнуто совпадение значений критерия, полученных с использованием этих моделей. Однако условие (10.92), согласно которому точки A и B должны быть совмещены, не выполняется. Таким образом, рассмотренное условие (10.91) является недостаточным, чтобы судить о корректности решения задачи формирования рекомендаций путем последовательной детализации объекта.

Обеспечение корректности решения задачи формирования рекомендаций путем последовательной детализации связано, в первую очередь, с организацией процедур формирования информации для одних уровней детализации по результатам выполнения других из этих уровней. В дальнейшем, определяя задачу вертикального согласования, будем иметь в виду, что ее целью является обеспечение такого выбора формируемых рекомендаций $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_N$ на базе соответствующих моделей ${}^tM_1(K), {}^tM_2(K), \dots, {}^tM_N(K)$, чтобы выполнялось условие (10.92).

Резюмируя сказанное, дадим постановку формальной задачи, решение которой лежит в основе вертикального взаимодействия КМ на базе многоуровневых моделей. При этом известными считаются:

- тандемная модель $\{{}^tM(K)\}_{i=1,2,\dots,N}$:

$$K = F_i(x_i, y_i, z_i) \quad (i=1, 2, \dots, N);$$
- значения $x_1 (x_1 \equiv x_0)$ и $z_N (z_N \equiv z_0)$;
- множество рассматриваемых сформированных рекомендаций

$$Y_i \quad (i = 1, 2, \dots, N).$$

Необходимо определить значения $\bar{y}_i \quad (i = 1, 2, \dots, N)$ из решения задач:

$$\bar{y}_i = \arg \min_{y_i \in Y_i} F_i(x_i, y_i, z_i) \quad (i = 1, 2, \dots, N) \quad (10.93)$$

при связях:

$$z_i = Q(\bar{y}_{i+1}, z_{i+1}) \quad (i=1, 2, \dots, N-1); \quad (10.94)$$

$$x_i = \{x_{i-1}, \bar{y}_i\} \quad (i = 2, 3, \dots, N) \quad (10.95)$$

таким образом, чтобы выполнялось условие:

$$\prod_{i=1}^N \bar{y}_i = \bar{y}_0. \quad (10.96)$$

где

$$\bar{y}_0 = \arg \min_{y \in Y} F_N(x, y, z_N),$$

$$Y = \prod_{i=1}^N Y_i$$

Предложим подход к решению поставленной задачи.

10.1.3.2. Формирование подхода к решению задачи

Проанализируем принципиально возможные способы организации процедур вертикального согласования формируемых рекомендаций различных уровней. Источник рассогласования рекомендаций в данном случае состоит в том, что при проведении детализации на каждом уровне используется прогнозная информация о результатах выполнения операций последующих уровней, которая при непосредственном выполнении этих операций не подтверждается. Эта информация для операций уровня i задается в виде вектора z_i , влияющего на формируемые рекомендации как i -го, так и последующих уровней. Рассмотрим это влияние на примере изолированной пары уровней детализации — S_i и S_{i+1} , считая реакции x_i и z_{i+1} известными константами.

Влияние z_i на выполнение операций соответствующего (i -го) уровня непосредственное, так как этот вектор присутствует в явном виде в условиях, задающих множество допустимых рекомендаций Y_i и значения критериев на его элементах.

На результаты выполнения операций последующих уровней z_i влияет косвенно и это влияние проявляется двояко. Во-первых, оно проявляется с помощью связей $\bar{y}_i = y_i(x_i, z_i)$ и $x_{i+1} = x_{i+1}(x_b, \bar{y}_i)$, после использования которых допустимые и рациональные на $(i + 1)$ -м уровне рекомендации можно представить в виде

$$Y_{i+1} = Y_{i+1}(x_i, z_i, z_{i+1}); \quad \bar{y}_{i+1} = \bar{y}_{i+1}(x_b, z_b, z_{i+1}).$$

Во-вторых, его проявление связано с тем, что данные Y_{i+1} и \bar{y}_{i+1} соответствуют рекомендациям предыдущего (i -го) уровня, полученным при некотором фиксированном z_i . Однако не каждый элемент $y_{i+1} \in Y_{i+1}$ является соответствующим этому z_i . Данное соответствие наблюдается лишь на некотором подмножестве из Y_{i+1} , являющемся

пересечением Y_{i+1} с множеством, обозначаемым далее Y_{i+1, z_i} на котором ранее сделанный прогноз по z_i выполняется. При этом Y_{i+1, z_i} является множеством решений уравнения $Q(\varphi''_i(y_{i+1}, z_{i+1})) = z_i$:

$$Y_{i+1, z_i} = \{y_{i+1} : Q(\varphi''_i(y_{i+1}, z_{i+1})) = z_i\}.$$

Итак, вторым проявлением влияния z_i на результаты выполнения операций следующего уровня является дополнительное ограничение на множество допустимых рекомендаций, которое теперь может быть представлено в виде

$$Y'_{i+1} = Y_{i+1} \cap Y_{i+1, z_i}.$$

При этом Y_{i+1} определяет допустимые рекомендации, соответствующие \bar{y}_i , при формировании которого использовался z_i , а Y_{i+1, z_i} — рекомендации подтверждающие этот z_i (рис. 10.24).

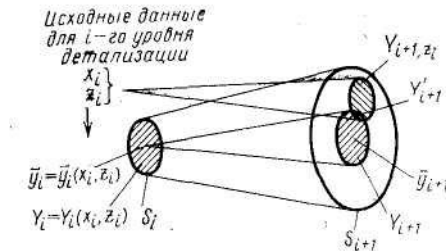


Рис. 10.24. Определение совместимости формируемых рекомендаций различных уровней детализации (Y_{i+1} — формируемые рекомендации $(i+1)$ -го уровня, соответствующие рекомендациям i -го уровня при некотором фиксированном z_i ; Y_{i+1, z_i} — формируемые рекомендации $(i+1)$ -го уровня, соответствующие прогнозу (z_i) о них, принятому при выполнении операций i -го уровня; Y'_{i+1} — формируемые рекомендации $(i+1)$ -го уровня, совместимые с рекомендациями, принятыми при выполнении операций i -го уровня, т. е. соответствующие \bar{y}_i и z_i , при которых были определены эти \bar{y}_i)

Распространяя полученный результат, т. е. не ограничиваясь рассмотрением двух уровней, получим

$$Y'_i = Y_i \cap Y_{i+1, z^+_{i+1}} \quad (i = 2, 3, \dots, N^s, Y'_1 = Y_1),$$

где $Y_{i-1, z^+_{i+1}} = Y_{i, z_1, z_2, \dots, z_{i-1}}$ определяется как множество решений системы уравнений

$$Q(\varphi''_j(y_{j+1}, y_{j+2}, \dots, y_{N^s}, z_0)) = z \quad (j=1, 2, \dots, i-1)$$

относительно y_k , $k = j+1, j+2, \dots, N^s$, т. е.

$$Y_{i,z^+_{i-1}} = \{y_i : Q(\varphi''_j(y_{j+1}, y_{j+2}, \dots, y_{N_s}, z_0)) = z_j; j = 1, 2, \dots, i-1\}$$

$$(i = 2, 3, \dots, N^s).$$

Таким образом, область допустимых на очередном уровне рекомендаций, совместимых с рекомендациями предшествующих уровней, можно представить в виде пересечения двух множеств, первое из которых — Y_i определяется непосредственно сформированными рекомендациями, принятыми на предшествующих уровнях детализации объекта, а второе — $Y_{i,z^+_{i-1}}$ состоит из рекомендаций, подтверждающих прогноз о них, который был использован при выполнении предшествующих операций. Организация выполнения операций на каждом i -м уровне из условия «попадания» выбираемых здесь рекомендаций одновременно и в Y_i и в $Y_{i,z^+_{i-1}}$ обеспечивает линейную последовательность прохождения различных уровней детализации, минуя необходимость проведения итераций. Достигаемая при этом совместимость формируемых рекомендаций различных уровней обеспечивается при некотором образом фиксированных значениях z_b , определение которых представляет следующую самостоятельную проблему.

Множество $Y_{i,z^+_{i-1}}$ может быть также представлено как множество решений системы уравнений:

$$\bar{y}_i = \bar{y}_i(\bar{y}_1, \bar{y}_2, \dots, \bar{y}_{i-1}, \dots, \bar{y}_{N^s}) \quad (i=1, 2, \dots, N^s), \quad (10.97)$$

содержащей N^s уравнений с $\sum_{i=1}^{N^s} |y_i|$ неизвестными. Здесь, как

правило, $|y_i| > 1$, т. е. система (10.97) формально незамкнута и допускает множество решений. Эта система в общем случае порождается связями (10.93) — (10.95) — аналогами (10.91). Ее незамкнутость может рассматриваться как иллюстрация к выводу о недостаточности условий (10.91) для формирования рациональных рекомендаций и необходимости привлечения для этого условий (10.96).

Итак, многоэлементность $Y_{i,z^+_{i-1}}$ порождает множественность z_b , обеспечивающих соответствие формируемых рекомендаций различных уровней детализации. Тогда на определение z_i ($i = 1, 2, \dots, N^s - 1$) требуется наложить условие обеспечения рациональности формируемых рекомендаций, а определение z_i на каждом уровне производить наравне с y_i . В таком случае схема взаимодействия

уровней детализации формируемых рекомендаций может быть представлена в виде линейной последовательности действий, связанных с выбором рациональных y_i и z_i ($i = 1, 2, \dots, N^s$). При этом функционал и условия определения области допустимых рекомендаций Y_i при выполнении операций каждого уровня не изменяются, только теперь операции максимизации (минимизации) должны выполняться не только по y_i , но и по z_i . Область допустимых значений z_i обозначаемая далее \tilde{Z} , должна формироваться из условий соответствия множества допустимых на последующих уровнях рекомендаций и принципиально может быть определена следующим образом:

$$\tilde{Z}_i = \bigcap_{y_k \in Y_k, k=i+1, i+2, \dots, N^s} Q(\varphi''_i(y_{i+1}, y_{i+2}, \dots, y_{N^s}, z_0)).$$

Реальное получение множеств \tilde{Z}_i невозможно, в основном, вследствие наличия корреляционных связей между компонентами каждого из векторов z_i ($i = 1, 2, \dots, N^s$). Определение этих связей невозможно в силу специфики реакций z_i , получаемых с помощью операторов Q , реализация которых связана, как было определено, с обработкой экспериментальных данных. При этом возможно лишь определение границы допустимых значений по каждой компоненте z_i , прямое произведение которых, обозначаемое далее Z , «вписывает» в себя искомое \tilde{Z} .

Очевидно, что Z содержит и элементы, соответствующие нереализуемым (несовместимым) сформированным рекомендациям, т. е. не удовлетворяющим системе (10.97). Поэтому возможно, что выбранный рациональным в процессе оптимизации $\tilde{z}_i \in Z_i$ окажется нереализуемым, т. е. окажется, что на последующих за i -м уровнях $Y'_j = \emptyset$ ($j > 0$). Это тем более вероятно, что большинство практически используемых функционалов являются монотонными по компонентам z_i , вследствие чего \tilde{z}_i будут, в основном, определяться «угловыми» элементами Z , не попадающими в \tilde{Z} (рис. 10.25).

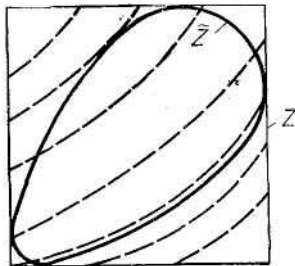


Рис. 10.25. К определению Z и \tilde{Z} (----- линии равных значений минимизируемого критерия)

Заметим, что даже в случае получения конкретных связей между компонентами z подход к организации выполнения операций, требующий формирования рекомендаций как по y , так и по z , отличается методической неверностью. Во-первых, на каждом уровне здесь должны формироваться рекомендации (по z), которые не относятся к компетенции данного уровня, а во-вторых, — назначения z , носящих характер обобщенных параметров, приводят к неявному ограничению путей дальнейшего развития z вариантов формирования рекомендаций.

Если при формировании рекомендаций на каждом уровне не пытаться рационализировать z , то очевидно, что при выполнении операций последующих уровней целесообразно ставить целью формирование рекомендаций, соответствующих ранее зафиксированному z . Но если при формировании рекомендаций на i -м уровне не учитывать условий «попадания» в $Y_{i,z^{+}i-1}$, а рассматривать в качестве допустимого все множество Y_i , то далее необходимо проверить «попадание» выбранной сформированной рекомендации в $Y_{i,z^{+}i-1}$ и в случае неудовлетворения ему произвести коррекцию вектора z_{i-1} так, чтобы выбранная рекомендация лежала в скорректированном множестве $Y_{i,z^{+}i-1}$. Затем, очевидно, необходимо произвести изменения Y_i и \bar{Y}_i на тех уровнях, где изменились z , т. е. выполнить итерацию и т. д. Данная процедура описывает обеспечение соответствия результатов выполнения различных уровней путем непосредственного решения системы уравнений (10.97) методом итераций.

Проблема нахождения корней систем нелинейных уравнений является достаточно непростой, поскольку в данном случае в роли

уравнений выступают совокупности операций детализации сформированных рекомендаций, т. е. вычисления правой части каждого такого уравнения определяются временем выполнения операций соответствующего уровня детализации рекомендаций. Кроме того, множественность решений системы (10.97), как правило, не позволяет получить сформированные рекомендации рациональными в целом (см. рис. 10.23).

Организации процедур вертикального согласования, сводящейся к решению системы уравнений (10.97), присущи и другие недостатки. В частности, как при автономном выполнении операций, так и при решении соответствующих систем уравнений для согласования результатов выполнения этих операций никак не ограничивается область нахождения рекомендаций. Ею на каждой итерации является множество возможных рекомендаций соответствующего уровня — Y_i ($i = 1, 2, \dots, N^s$). При этом информация, накопленная к текущему моменту, никак не используется в интересах ограничения множеств рекомендуемых альтернатив, требующих анализа. Ввиду этого очевидна перспективность предлагаемой далее организации итераций (с последовательным ограничением области поиска рациональных рекомендаций) с точки зрения сокращения трудоемкости решения задач согласования.

Итак, из сказанного следует, что невозможность однозначного определения значений реализуемых реакций \bar{z}_i ($i = 1, 2, \dots, N^s$), обеспечивающих рациональность рекомендаций в целом, порождает необходимость организации итерационных процедур согласования. Формальной причиной этому являются принципиальные сложности построения множества допустимых реакций (\bar{Z}), соответствующих рекомендациям следующих уровней. Однако возможно построение «описывающего» его множества, называемого далее **расширенным множеством допустимых реакций (Z)**.

Ниже предлагается при выполнении операций очередного уровня рассматривать два значения z_i , первое из которых (z_i^a) является рациональным на расширенном множестве Z_i и может не принадлежать строгому множеству допустимых значений \tilde{Z}_i , т. е. являться нереализуемым; второе значение (z_i^b) — заведомо реализуемое значение z_i , на рациональность которого не обращается внимание. Рациональное реализуемое значение \bar{z}_i лежащее где-то между указанными точками, предполагается искать, сближая эти точки (рис. 10.26).

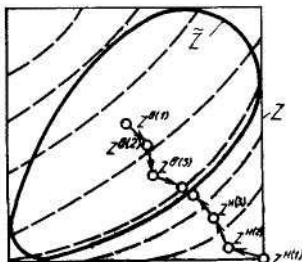


Рис. 10.26. Определение значений \bar{z} путем «сближения» значений z^b и z^h .

В качестве указанной пары значений z_i в общем случае, могут быть приняты «угловые точки» множества Z_i . Их определение требует решения самостоятельной задачи. Предложение об учете указанных точек фактически выливается в предложение не фиксировать некоторые конкретные значения z при выполнении операций каждого уровня, а учитывать их некоторое минимальное множество, гарантированно содержащее элемент, соответствующий искомым рациональным рекомендациям следующих за рассматриваемым уровнем. Здесь под минимальностью такого рода множества будем понимать отсутствие в нем элементов, соответствующих заведомо нерациональным рекомендациям.

Основной направленностью формируемой с учетом данного предположения методики вертикального согласования является минимизация (сжатие) множества учитываемых реакций Z вокруг элемента, соответствующего рациональным рекомендациям последующих уровней. Такая минимизация производится за счет «отбраковки» заведомо нерациональных и нереализуемых рекомендаций в процессе их детальной проработки. Причем результаты проработки одной из рекомендаций предполагается в определенной мере переносить на соседние с ней рекомендации, что позволяет эффективно использовать накапливаемую в процессе формирования рекомендаций информацию и минимизировать трудоемкость процедур согласования.

10.1.3.3. Определение множеств конкурентоспособных рекомендаций и путей их сужения

Рассмотрим особенности и принципы оценки и выбора формируемых рекомендаций на каждом уровне при непосредственном учете множества возможных рекомендаций последующих уровней. Этот учет требует представления реакций последующих уровней при выполнении текущих операций в виде некоторого множества Z .

Очевидно, что непосредственное использование в качестве реакций нижних уровней некоторого неконкретизированного множества Z приводит к неоднозначности оценок рассматриваемых альтернативных рекомендаций, так как, в частности, «размывает» значения функционала $F(x, y, z)$, являющегося критерием оценки анализируемых рекомендаций, в некоторое множество. Это множество для каждого y при некотором фиксировании x будем задавать его верхней и нижней границами — $F(x, y, z^H)$ и $F(x, y, z^B)$, которые можно определить следующим образом:

$$F(x, y, z^H) = \min_{z \in Z} F(x, y, z);$$

$$F(x, y, z^B) = \max_{z \in Z} F(x, y, z).$$

Получаемые таким образом z^B и z^H являются упоминавшимися выше «угловыми» элементами множества Z .

Истинная оценка рекомендации y при каждом конкретном $z \in Z$ принадлежит диапазону $\delta(x, y, Z)$:

$$\delta(x, y, Z) = [F(x, y, z^H), F(x, y, z^B)].$$

В дальнейшем $F(x, y, z^H)$ и $F(x, y, z^B)$ будем называть функциями нижних и верхних границ, соответственно.

Неоднозначность реакций влечет за собой необходимость включения в множество допустимых формируемых рекомендаций каждой рекомендации, допустимой хотя бы при одном $z \in Z$, т. е. хотя бы при одном из возможных путей дальнейшей детализации рекомендаций. Обозначим определенное таким образом множество допустимых рекомендаций через $Y(x, Z)$:

$$Y(x, Z) = \bigcap_{z \in Z} \{y: V(x, y, z) \leq 0\}.$$

Выделим из множества $Y(x, Z)$ ту рекомендацию, которой соответствует лучшая оценка, и обозначим ее y^* . Эту рекомендацию можно определить следующим образом:

в задачах максимизации критерия

$$y^* = \arg \max_{y \in Y(x, Z)} F(x, y, z^B);$$

в задачах минимизации критерия

$$y^* = \arg \min_{y \in Y(x, Z)} F(x, y, z^H); \quad (10.98)$$

В данном случае несложно заметить, что рекомендации y^* определяются при самом оптимистическом прогнозе о рекомендациях последующих уровней или, другими словами, при значениях $z \in Z$, доставляющих максимум (минимум) используемому функционалу $F(x, y, z)$. Как было показано выше, такие значения z могут быть недопустимы (нереализуемы). Поэтому значения критерия $F(x, y^*, z^H)$ или $F(x, y^*, z^B)$ могут рассматриваться лишь как предельная оптимистическая оценка рекомендаций из $Y(x, Z)$, которая может и не подтвердиться.

Выделим теперь из множества $Y(x, Z)$ ту рекомендацию, которой соответствует лучшая оценка при пессимистическом прогнозе рекомендаций последующих уровней. Эта рекомендация, обозначаемая далее y^{**} , может быть определена следующим образом:

- в задачах максимизации критерия

$$y^{**} = \arg \max_{y \in Y(x, Z)} F(x, y, z^B);$$

- в задачах минимизации критерия

$$y^{**} = \arg \min_{y \in Y(x, Z)} F(x, y, z^H); \quad (10.99)$$

Здесь значения критерия $F(x, y^{**}, z^H)$ и $F(x, y^{**}, z^B)$ являются соответственно нижней и верхней границами значений используемого критерия, определяемыми худшим вариантом дальнейшего расширения вариантов рекомендаций. При этом можно гарантировать достижимость данных значений.

Обозначим через Δ диапазон изменения значений критерия, в котором гарантированно находится значение, соответствующее искомой рациональной рекомендации. Ее границы определяются рациональными y при наилучших и наихудших $z \in Z$. Этот диапазон может определяться соотношениями:

- в задачах максимизации критерия

$$\Delta = [F(x, y^{**}, z^H); F(x, y^*, z^B)] = [\Gamma^H, \Gamma^B];$$

- в задачах минимизации критерия

$$\Delta = [F(x, y^*, z^H); F(x, y^{**}, z^B)] = [\Gamma^H, \Gamma^B].$$

В дальнейшем Δ будем называть гарантированным диапазоном с границами Γ^H и Γ^B .

Критерий с таким гарантированным диапазоном значений имеет следующую область определения:

$$Y^{\text{кон}} = \{y: \delta(x, y, Z) \cap \Delta \neq \emptyset\}. \quad (10.100)$$

Ниже формируемые рекомендации из множества $Y^{\text{кон}}$ будем называть конкурентоспособными, а само это множество — множеством конкурентоспособных рекомендаций (рис. 10.27).

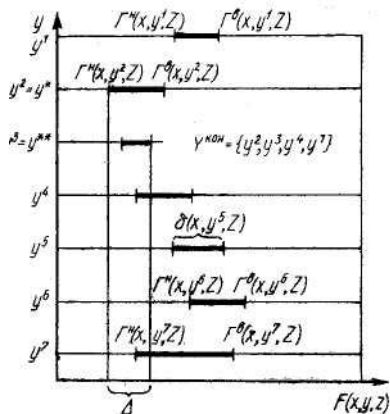


Рис. 10.27. Определение множества конкурентоспособных рекомендаций (для задачи минимизации критерия; x, Z — фиксированы)

Получаемые конкурентоспособные рекомендации являются неразличимыми в отношении их рациональности и, располагая имеющейся информацией о Z , не представляется возможным отдать предпочтение ни одной из этих рекомендаций. Все остальные рекомендации, не вошедшие в состав конкурентоспособных, теперь заведомо являются нерациональными, так как каждая из них даже при «лучшем» $z \in Z$ имеет значение критерия, худшее по сравнению с гарантированной оценкой.

Теперь рассмотрим схему процесса вертикального согласования формируемых рекомендаций, обеспечивающую последовательное сужение множеств конкурентоспособных рекомендаций. В основе рассматриваемой схемы лежит анализ формируемых рекомендаций каждого уровня с учетом множественности возможных способов их детализации на следующих уровнях, который позволяет разделить все исходные альтернативы Y на две группы. В первой из них содержатся рекомендации «подозрительные на рациональность» — конкурентоспособные, а в другой — рекомендации заведомо нерациональные, которые из дальнейшего рассмотрения могут быть безусловно исключены. При этом, если выделение рациональной рекомендации из множества конкурентоспособных проводить путем

рассмотрения каждой из рекомендаций с учетом конкурентоспособных вариантов их детализации, то для этого необходимо многократное выполнение операций последующих уровней при различных исходных (директивных) данных. В этом случае требуется фактически совместить выполнение операций различных уровней, т. е. решать задачу размерности, являющейся суммой размерностей задач совмещаемых уровней, что лишает смысла членение исходной задачи формирования рекомендаций на уровни детализации. Ниже предлагается схема взаимодействия операций различных уровней, основывающаяся на разделении процессов решения задач различных уровней.

Рассмотрим пути, обеспечивающие сужение множеств конкурентоспособных рекомендаций в процессе взаимодействия операций различных уровней. При этом ограничимся рассмотрением взаимодействия изолированной пары уровней — S_i и S_{i+1} , т.е. будем считать x_i и Z_{i+1} известными константами и предполагать, что решения y_{i+1} полностью детализируют y_i . Это позволяет положить:

$$x_{i+1} = \{x_b, y_{ij}\}.$$

Для достижения поставленной цели возможны два пути:

- сужение гарантированного диапазона — Δ ;
- сужение диапазонов оценок каждой из анализируемых рекомендаций — $\delta(x_b, y_b, Z_i)$ при $\forall y_i \in Y_i^{кон}$.

Соотношение (10.100), на основе которого осуществляется сужение (для определенности будем рассматривать задачу минимизации критерия), может быть представлено в виде

$$Y_i^{кон} = \{y_i: F_{ii}(x_b, y_b, z_i^u) \leq F(x_b, y_i^{**}, z_i^b)\},$$

или

$$Y_i^{кон} = \{y_i: F_{ii}(x_b, y_b, z_i^u) \leq \Gamma^b\}. \quad (10.101)$$

Верхняя граница гарантированного диапазона Γ^b дает заведомо достижимое значение критерия. Поскольку принято, что y_{i+1} полностью детализирует y_b , то каждое значение функции $F_{i+1}(x_b, y_i, y_{i+1}, z_{i+1})$ является достижимым и за границу Γ^b может быть принято значение критерия на произвольной паре (y_i, y_{i+1}) . При этом следует иметь в виду, что значение $\Gamma^b = F_i(x_b, y_i^{**}, z_i^b)$ получается в процессе выполнения операций i -го уровня, а $F_{i+1}(x_b, y_b, y_{i+1}, z_{i+1})$ — при детализации рекомендаций y_b , т. е. на $(i+1)$ -м уровне. Тогда, если в результате детализации некоторого фиксированной рекомендации \tilde{y}_i на $(i+1)$ -м уровне определяются соответствующие \bar{y}_{i+1} и $F_{i+1}(x_b, \tilde{y}_i, \tilde{y}_{i+1}, z_{i+1})$:

$$F_{i+1}(x_b, \tilde{y}_i, \bar{y}_{i+1}, z_{i+1}) = \min_{y_{i+1} \in Y_{i+1}} F_{i+1}(x_b, \tilde{y}_i, \bar{y}_{i+1}, z_{i+1}), \quad (10.102)$$

такие, что

$$F_{i+1}(x_b, \tilde{y}_i, \bar{y}_{i+1}, z_{i+1}) < F(x_b, y_i^{**}, z^B),$$

то в качестве верхней границы гарантированного диапазона можно принять $F_{i+1}(x_b, \tilde{y}_i, \bar{y}_{i+1}, z_{i+1})$. Если же условие (10.102) не выполняется, то рекомендация \tilde{y}_i должна быть исключена из состава неконкурентоспособных:

$$\Gamma^B = \begin{cases} F_i(x_i, y_i^{**}, z_i^B), & \text{если } F_i(x_i, y_i^{**}, z_i^B) < \\ & < F_{i+1}(x_i, \tilde{y}_i, \bar{y}_{i+1}, z_{i+1}); \\ F_{i+1}(x_i, \tilde{y}_i, \bar{y}_{i+1}, z_{i+1}), & \text{если } F_i(x_i, y_i^{**}, z_i^B) > \\ & > F_{i+1}(x_i, \tilde{y}_i, \bar{y}_{i+1}, z_{i+1}). \end{cases} \quad (10.103)$$

Таким образом, в результате детальной проработки на $(i+1)$ -м уровне отдельных рекомендаций y_i из состава конкурентоспособных возможно уменьшение (понижение) верхней границы гарантированного диапазона значений критерия и, как следствие, сужение исходного множества конкурентоспособных рекомендаций на i -м уровне (рис. 10.28).

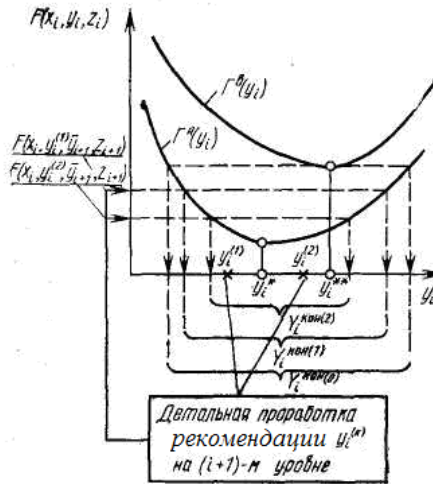


Рис. 10.28. Сужение множества конкурентоспособных рекомендаций при уменьшении верхней границы гарантированного диапазона

Вторым путем сужения множества конкурентоспособных рекомендаций является изменение (повышение) функции нижних границ $F_i(x_b, y_b, z_i^H)$ при $\forall y_i \in Y^{кон}_i$, определяемой наиболее оптимистическими прогнозами из Z_b , которые, как подчеркивалось, могут быть нереализуемы.

Уточнение нижних оценок для каждого $\forall y_i \in Y^{кон}_i$ может быть произведено лишь в результате выполнения операций, где непосредственно рассматриваются рекомендации следующего уровня, а это связано с совмещением решения задач i -го и $(i + 1)$ -го уровней. В таком аспекте уточнение функции нижних границ выливается в непосредственное решение задачи совместного выбора рациональных \bar{y}_i и \bar{y}_{i+1} .

Как будет показано ниже, можно «развязать» решения задач различных уровней при построении функции нижних оценок при $\forall y_i \in Y^{кон}_i$. Для этого необходимо осуществить детальную проработку лишь некоторых из определенного ряда рекомендаций $\forall y_i \in Y^{кон}_i$. Совокупность этих рекомендаций далее будем называть программой детализации и обозначать $\pi_i \subseteq Y^{кон}_i$. Ее строят в зависимости от анализируемого множества рекомендаций $Y^{кон}_i$ и переопределяют при его изменениях. На определение этой программы существенное влияние оказывают уже проработанные при выполнении предшествующих действий рекомендации из $Y^{кон}_i$. В дальнейшем такие рекомендации будем обозначать \hat{Y}_i . В итоге получим

$$\pi_i = \pi_i(Y^{кон}_i, \hat{Y}_i),$$

где $\hat{Y}_i \subset Y^{кон}_i$.

По мере проведения детализации конечного числа рекомендаций $y_i \subset \pi_i$ оказывается возможным провести коррекцию множества прогнозов о рекомендациях следующего уровня (т. е. Z_i) таким образом, что формируемая на его основе функция нижних границ является более близкой к реальным оценкам рекомендаций $\forall y_i \in Y^{кон}_i$. Образно говоря, в данном случае за счет проведенных детальных проработок происходит падение неоправданного оптимизма, результатом чего является увеличение нижних гарантированных оценок критерия при каждом $\forall y_i \in Y^{кон}_i$ (рис. 10.29).

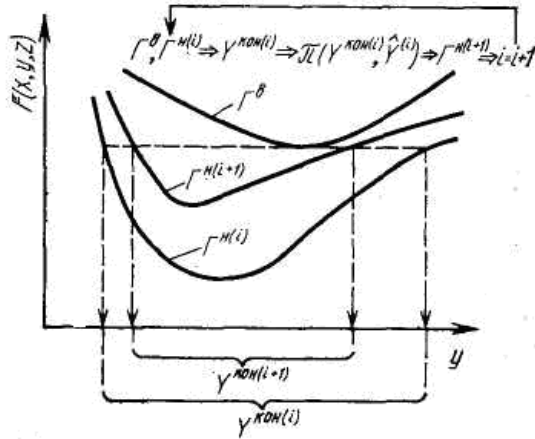


Рис. 10.29. Сужение множества конкурентоспособных рекомендаций при уточнении нижних гарантированных оценок

Рассмотренные два пути сужения множества конкурентоспособных рекомендаций взаимно дополняют друга друга. В частности, при детализации рекомендаций $y_i \in \pi_i$, проводимых с целью уточнения значений z_i^H , возможно получение результатов, позволяющих снизить значения Γ^B , а проработка рекомендаций y_i , которая предположительно наибольшим образом позволяет снизить Γ^B , дает информацию для уточнения z_i^H .

Опишем теперь процесс взаимодействия операций различных уровней, основанный на выделении и последовательном сужении множеств конкурентоспособных рекомендаций. Выполнение каждого приближения в этом процессе заключается в выполнении следующей последовательности действий:

- определение множества реакций Z_i соответствующего прогнозируемому рекомендациям, в данном случае $(i + 1)$ -го уровня, т. е. множеству \mathcal{Y}_{i+1} .

Множество Z_i при этом достаточно представить его «угловыми» элементами — z_i^H и z_i^B , определяющими диапазоны оценок каждой сформированной рекомендации $y_i \in \mathcal{Y}_i$;

- определение множества конкурентоспособных рекомендаций $Y_i^{кон}$ на основе полученного Z_i ;

- определение программы $\pi_i (Y_i^{кон}, \hat{Y}_i) \subseteq Y_i^{кон}$, представляющей собой конечную совокупность рекомендаций i -го уровня, детализация

которых на $(i + 1)$ -м уровне позволяет уточнить прогнозируемые значения z_i^H ;

- определение рационального варианта детализации рекомендации \tilde{Y}_i , что связано с выполнением операций следующего уровня; в результате чего находятся \tilde{Y}_{i+1} (\tilde{Y}_i) и соответствующее значение

$F_{i+1}(x_{i+1}, \tilde{Y}_i, \tilde{Y}_{i+1}, z_{i+1})$, которые являются информацией, передаваемой на i -й уровень;

- анализ полученной информации о результатах детализации рекомендаций, производимый на i -м уровне; при этом:

а) если выполняется условие $\Gamma^B > F_{i+1}(x_b, \tilde{Y}_i, \bar{Y}_{i+1}, z_{i+1})$, то за верхнюю гарантированную оценку принимают значение

$$F_{i+1}(x_b, \tilde{Y}_i, \bar{Y}_{i+1}, z_{i+1}),$$

б) если детализированы все рекомендации из π_i , т. е. $\pi_i \setminus \hat{Y}_i = \emptyset$, то это является условием, позволяющим уточнить z_i^H ;

в) если ни одно из указанных условий не выполняется, то необходимо выбрать для детализации следующую рекомендацию из известной программы и передать ее на $(i+1)$ -й уровень для детальной проработки.

Рассматриваемый процесс выделения рациональной рекомендации проводится до тех пор, пока полученный разброс значений минимизируемого критерия, обозначенный ранее Δ и определяемый как разность между верхней и нижней границами гарантированного диапазона:

$$\Delta = \Gamma_i^B - F_i(x_i, y_i^*, z_i^H) \tag{10.104}$$

не достигнет наперед заданных значений.

Распространим полученные результаты на общий случай. Основным отличием рассмотренного взаимодействия от общего случая является то, что при выполнении операций $(i + 1)$ -го уровня в общем случае должны рассматриваться не «точечные» значения z_{i+1} , как предполагалось выше, а некоторое множество Z_{i+1} . В этом случае информацией о результатах детализации некоторой рекомендации \tilde{Y}_i на $(i + 1)$ -м уровне будут уже не значения $\bar{Y}_{i+1}(\tilde{Y}_i)$ и $F_{i+1}(x_b, \tilde{Y}_i, \bar{Y}_{i+1}, z_{i+1})$, а множество Y_{i+1}^{KOH} и границы его гарантированного диапазона: $\Gamma_{i+1}^B, F_{i+1}(x_b, \tilde{Y}_i, y_{i+1}^*, z_{i+1}^H)$. Принципиально это ничего не меняет. В частности, если выше предполагалось значение $F_{i+1}(x_b, \tilde{Y}_i, \bar{Y}_{i+1}, z_{i+1})$

использовать для коррекции и $\Gamma_{i^b}^b$ и $z_{i^b}^h$, то теперь значение Γ_{i+1}^b должно использоваться для коррекции $\Gamma_{i^b}^b$, а значение $F_{i+1}(x_b, \tilde{y}_i, y_{i+1}^*, z_{i+1}^h)$ — для коррекции $z_{i^b}^h$.

В общем случае представляется возможной организация трех схем вертикального согласования на базе процедур взаимодействия соседних пар уровней детализации:

- с максимальной детализацией;
- с минимальной детализацией;
- с детализацией по условию.

В первой из указанных схем предполагается, что после определения конкурентоспособных рекомендаций на каждом уровне выбирается одна из них и передается для детализации на следующий уровень. При этом информация на предшествующий уровень передается лишь после выбора на всех последующих уровнях единой рациональной рекомендации, соответствующей полученному заданию. Это означает, что получаемая на каждом уровне информация «снизу» полностью детализирует те рекомендации, которые были заданы для детальной проработки.

Вторая схема противоположна первой. Здесь на основе задания в виде некоторой рекомендации, например \tilde{y}_i , на $(i + 1)$ -м уровне определяется соответствующее $Y_{i+1}^{\text{кон}}$, и информация сразу передается на i -й уровень.

Названные схемы взаимодействия характеризуются такими двумя показателями, как «время исполнения задания по детализации одной рекомендации» и «конкретность детализации одной рекомендации», определяемая величиной Δ . Имеет место прямая пропорциональность между этими показателями и предпочтительность уменьшения первого из них и увеличения второго. Очевидно, что реализация взаимодействия по первой схеме обеспечивает максимум значений обоих показателей, а по другой схеме — минимум значений обоих показателей.

Практически наибольший интерес, видимо, представляет организация вертикального взаимодействия соседних пар операций по третьей схеме. Здесь предполагается, что при формулировании задания на детализацию каждой очередной рекомендации одновременно задаются и условия, по которым требуется сообщать информацию о результатах детализации. При этом очевидно, что организация взаимодействия по первой и второй схемам является частным случаем организации взаимодействия по третьей схеме.

Итак, предложенная схема вертикального согласования осно-

ываается на учете множественности реакций z при выполнении каждого уровня детализации рекомендаций. Она носит итеративный характер, где на каждой k -й итерации из некоторого множества альтернатив ($\mathcal{Y}^{(k)}$) выделяются конкурентоспособные рекомендации ($Y^{\text{кон}}^{(k)}$). При этом требуется обеспечить условия:

$$Y^{\text{кон}}^{(k)} \subset \mathcal{Y}^{(k)}_i, \quad (10.105)$$

где $\mathcal{Y}^{(1)}_i = \mathcal{Y}_{0i}$ ($i = 1, 2, \dots, N$);

$$\mathcal{Y}^{(k)}_i = Y^{\text{кон}}^{(k-1)}_i, \quad k > 1 \quad (i=1, 2, \dots, N)$$

и

$$\bar{Y}_0 \in \prod_{i=1}^N Y_i^{\text{кон}}^{(k)} \quad \text{при } \forall k. \quad (10.106)$$

Тогда согласно принципу сжатых отображений рассмотренный метод можно считать сходящимся к рациональным рекомендациям в целом.

Определим основные операторы метода и покажем, что при их реализации выполняются условия (10.105) и (10.106).

10.1.3.4. Основные операторы метода

Согласно предложенной схеме основными процедурами метода являются:

- определение множества реакций $Z_i = Z_i(\mathcal{Y}_{i+1}, \mathcal{Y}_{i+2, \dots}, \mathcal{Y}_N, Z_N)$, соответствующих прогнозируемому множеству рекомендаций, следующих за рассматриваемым уровнем;

- определение множества конкурентоспособных рекомендаций $Y^{\text{кон}}_i = Y^{\text{кон}}_i(x_i, Z_i)$;

- определение программы $\pi_i(Y^{\text{кон}}_i, \hat{Y}_i) \subset Y^{\text{кон}}_i$ детализации рекомендаций из $Y^{\text{кон}}_i$ с целью уточнения функции нижних границ.

Поставим в соответствие каждому из названных действий некоторый оператор и сформулируем процедуры выполнения этих операторов.

Рассматриваемый метод вертикального согласования основывается на том, что при выполнении операций каждого текущего уровня непосредственно учитывается множество реакций последующих уровней — Z_i ($i = 1, 2, \dots, N - 1$), которое должно соответствовать множеству рассматриваемых на этих уровнях рекомендаций \mathcal{Y}_j ($j=i+1, i+2, \dots, N$). Требуемые соответствия формально представляются в виде

$$Z_i \Leftrightarrow Q(\mathcal{Y}_{i+1}, \mathcal{Y}_{i+2, \dots}, \mathcal{Y}_{i+k} z_{i+k}),$$

$$k = 1, 2, \dots, N-i; \quad i = 1, 2, \dots, N-1 \quad (10.107)$$

и означают, что для каждого $y_{i+k} \in Y_{i+k}$, $z_{i+k} \in Z_{i+k}$ ($k = 1, 2, \dots, N-1$) найдется такое значение $z'_i \in Z_i$ при котором

$$F_{i+k}(x_b, y_b, y_{i+1}, \dots, y_{i+k}, z_{i+k}) = F_i(x_b, y_b, z'_i), \quad (10.108)$$

$$k = 1, 2, \dots, N-i; \quad i = 1, 2, \dots, N-1.$$

Заметим, что условия (10.108) будут выполняться, если Z_i ($i = 1, 2, \dots, N-1$) таковы, что:

1) функция $F_i(x_b, y_b, z^B_i)$ является мажорантой семейств функций

$F_{i+k}(x_i, y_i, y_{i+l}, \dots, y_{i+k}, z_{i+k})$; $y_{i+l} \in \mathcal{Y}_{i+l}$; $l = 1, 2, \dots, k$, $z_{i+k} \in Z_{i+k}$ каждое из которых определяется при различных $k \in [1, N-i]$, т. е.

$$F_i(x_b, y_b, z^B_i) \geq F_{i+k}^{Di}(x_b, y_b, y_{i+1}, \dots, y_{i+k}, z_{i+k}) \quad (10.109)$$

при $\forall y_{i+l} \in \mathcal{Y}_{i+l}$, $l=1, 2, \dots, k$; $\forall z_{i+k} \in Z_{i+k}$ где F_{i+k}^{Di} — проекция на плоскость с координатами — компонентами вектора $\{y_b, F_{i+k}\}$ семейства функций $F_{i+k}(x_b, y_b, y_{i+1}, \dots, y_{i+k}, z_{i+k})$ при $y_{i+l} \in \mathcal{Y}_{i+l}$ ($l = 1, 2, \dots, k$);

2) функция $F_i(x_b, y_b, z^H_i)$ является минорантой этих же семейств, т. е.

$$F_i(x_b, y_b, z^H_i) \leq F_{i+k}^{Di}(x_b, y_b, y_{i+1}, \dots, y_{i+k}, z_{i+k}) \quad (10.110)$$

при $\forall l = 1, 2, \dots, k$; $\forall z_{i+k} \in Z_{i+k}$.

Выполнение формальных условий (10.109), (10.10) обеспечивает наличие в Z_i значений, соответствующих каждому элементу из $\prod_{k=1}^N \mathcal{Y}_{i+k}$

т. е. выполнение требуемого соответствия (10.107). Эти же условия могут быть записаны в следующем виде:

$$F_i(x_b, y_b, z^B_i) \geq F_{i+k}^{Di}(x_{i+k}, y_{i+k}, z^B_{i+k}), \quad k = 1, 2, \dots, N-1 \quad (10.111)$$

$$F_i(x_b, y_b, z^H_i) \leq F_{i+k}^{Di}(x_{i+k}, y_{i+k}, z^H_{i+k}), \quad k = 1, 2, \dots, N-1 \quad (10.112)$$

Как уже отмечалось, не столь важным является знание самих множеств Z_i ($i = 1, 2, \dots, N-1$), сколько функций $F_i(x_i, y_b, z^H_i)$ и $F_i(x_b, y_b, z^B_i)$, определяемых значениями z^H_i и z^B_i . Причем $F_i(x_b, y_b, z^B_i)$ используются лишь при определении первого приближения и даже здесь может быть заменено любым реально достижимым значением критерия. Поэтому определение Z_i в общем случае может быть заменено нахождением значений z^H_i , задающих функцию нижних границ $F_i(x_i, y_b, z^H_i)$.

Функции $F_i(x_i, y_b, z^H_i)$ при $y_i \in \mathcal{Y}_i$ определяются, с одной стороны, условиями минорирования $F_{i+k}^{Di}(x_b, y_b, y_{i+1}, \dots, y_{i+k}, z^H_{i+k})$, $k=1, 2, \dots, N-i$ при $y_{i+k} \in \mathcal{Y}_{i+k}$ ($k=0, 1, \dots, i-1$), а с другой, — условиями максимального приближения к минорируемым функциям. В соответствии с этими условиями z^H_i могут быть определены из решения следующей задачи:

$$\bar{z}^H_i = \arg \min_{z_i \in Z^*_i} V(A^*, F_i(x_b, y_b, z_i), \mathcal{Y}_i), \quad i=1, 2, \dots, N-1, \quad (10.113)$$

здесь

$$V(A^*, F_i(x_b, y_b, z_i), \mathcal{Y}_i) = \int_{y_i \in \mathcal{Y}_i} (A^* - F_i(x_b, y_b, z_i)) dy_i; \quad (10.114)$$

$$Z^*_i = \{z_i: F_i(x_b, y_b, z_i) \leq F^{Di}_{i+k}(x_i, y_b, y_{i+1}, \dots, y_{i+k}, z_{i+k}); \quad (10.115)$$

$$y_{i+l} \in \mathcal{Y}_{i+l}; \quad l = 0, 1, \dots, k; \quad k = 1, 2, \dots, N-i\},$$

$$i = 1, 2, \dots, N-1,$$

где A^* — произвольная константа, удовлетворяющая условию:

$$A^* \geq \max_{y_i \in \mathcal{Y}_i} F_i(x_b, y_i, z^H_i).$$

Учитывая выражение (10.112), условия (10.115) могут быть записаны в виде:

$$Z^*_i = \{z^H_i: F_i(x_b, y_b, z^H_i) \leq F^{Di}_{i+1}(x_i, y_b, y_{i+1}, z^H_{i+1}); \quad (10.116)$$

$$y_j \in \mathcal{Y}_j; \quad j = i, i+1\}, \quad i = 1, 2, \dots, N-1.$$

Введем в рассмотрение функцию

$$\tilde{F}_i(x_i, y_b, z_{i+1}) = \min_{y_{i+1} \in \mathcal{Y}_{i+1}} F^{Di}_{i+1}(x_i, y_b, y_{i+1}, z^H_{i+1}); \quad (10.117)$$

после чего условия (10.116) можно представить так:

$$Z^*_i = \{z^H_i: F_i(x_b, y_b, z^H_i) \leq \tilde{F}_i(x_i, y_b, z_{i+1})\}, \quad i = 1, 2, \dots, N-1.$$

С введением данной функции условия (10.112) могут быть записаны в следующем виде:

$$\tilde{F}_i(x_i, y_b, z_{i+1}) \leq \tilde{F}_{i+1}(x_{i+k}, y_{i+k}, z_{i+k+1}), \quad (10.118)$$

$$i = 1, 2, \dots, N-1; \quad k = 1, 2, \dots, N-i.$$

Приведенные выше условия решения задачи определения значений z^H_i ($i = 1, 2, \dots, N-1$) требуют знания функции $\tilde{F}_i(x_i, y_b, z_{i+k})$, каждая точка на которой является результатом решения задачи определения \bar{y}_{i+1} из условий минимизации функции $F_{i+1}(x_i, y_i, y_{i+1}, z_{i+1})$, т.е. задачи на модели $(i+1)$ -го уровня при некоторых фиксированных y_i . Следует отметить, что определение \bar{y}_i может быть произведено на той же модели, т.е. путем одновременной минимизации $F_{i+1}(x_i, y_i, y_{i+1}, z_{i+1})$ по y_i и y_{i+1} . Это, однако, связано с практическим совмещением решения задач двух уровней, что нежелательно. Отсюда следует вывод о необходимости минимизации числа необходимых вычислений

$\tilde{F}_{i+1}(x_i, y_i, z_{i+1})$ при построении z^H_i . Число таких вычислений должно быть меньше, чем требуется для определения \bar{y}_i из минимизации

$\tilde{F}_{i+1}(x_i, y_i, y_{i+1}, z_{i+1})$ по y_i .

Точки, в которых производятся вычисления функции

$\tilde{F}_{i+1}(x_i, y_i, z_{i+1})$ обозначим через \hat{y}_i , а множество таких точек — через \hat{Y}_i . Причем, с какой бы частотой не располагались точки \hat{y}_i на \mathcal{Y}_i никогда нет уверенности в том, что в некоторых из них, например, $\hat{y}_i \notin \mathcal{Y}_i$ для функции $F_i(x_i, y_i, z_i^H)$, значения z_i^H которой определяются из выражения (10.113) при

$$Z_i^H = \{ z_i^H : F_i(x_i, y_i, z_i^H) \leq \tilde{F}_i(x_i, y_i, z_{i+1}), y_i \in \hat{Y}_i \};$$

не будет наблюдаться

$$F_i(x_i, \hat{y}_i, z_i^H) > \tilde{F}_{i+1}(x_i, \hat{y}_i, z_{i+1})$$

Другими словами, в точках $\hat{y}_i \notin \hat{Y}_i$ возможно нарушение условия (10.112) или его аналога (10.118). Отсюда следует вывод, что при построении z_i^H знания отдельных точек $\tilde{F}_i(x_i, y_i, z_{i+1}), y_i \in \mathcal{Y}_i$ недостаточно, так как при этом отсутствует гарантия выполнения условия, о том, что функции нижних границ моделей более верхних уровней должны минорировать аналогичные функции моделей следующих за ними уровней. А это условие было определено как основное, обеспечивающее наличие в Z_i элемента, соответствующего рациональным $\bar{y}_{i+k} \in \mathcal{Y}_{i+k} (k = 1, 2, \dots, N - i)$.

Используем идею привлечения для нахождения абсолютного минимума функции многих переменных констант Липшица.

При этом учтем сделанное ранее замечание о том, что в процессе непосредственного решения консультационных задач используются непрерывные модели. Это условие соответствует непрерывности

$F_i(x_i, y_i, z_i)$ при $y_i \in \mathcal{Y}_i (i = 1, 2, \dots, N)$ и определяет существование констант Липшица на всех \mathcal{Y}_i .

Введем в рассмотрение точки, которые будем называть далее липшицевыми. Обозначим эти точки $l_i \in \mathcal{P}_i$, где \mathcal{P}_i — множество липшицевых точек на D_i (плоскость с координатами (F_i, y_i) и определим как точки пересечения плоскостей, проходящих через точки $\{y_i, \tilde{F}_i(x_i, y_i, z_{i+1})\}, y_i \in \hat{Y}_i$ под углами $\pm \operatorname{arctg} L_j$, где L_j — константы Липшица:

$$L_j = \max_{k \in [i, N-1]} \max_{y^j \in \mathcal{Y}_i} \frac{\partial F_k(x_i, y_i, y_{i+1}, \dots, y_{1+k}, z_{1+k})}{\partial y^j} \quad (10.119)$$

$j = 1, 2, \dots, N_{y_i}$, здесь N_{y_i} — размерность вектора y_i .

Каждую j -ю липшицеву точку будем определять парой $(^j y_{li}, K(^j y_{li}))$:

$$l_i = \{^j y_{li}, K(^j y_{li})\}, i=1, 2, \dots, N-1.$$

При этом множество $\mathcal{Y}_{li} = \{^j y_{li}\}$ определяется множеством \hat{Y}_i :

$$\mathcal{Y}_{li} = \mathcal{Y}_{li}(\hat{Y}_i)$$

и, соответственно,

$$\mathcal{L}_i = \mathcal{L}_i(\hat{Y}_i)$$

Очевидно, что функция $F_i(x_i, y_i, z_i^H)$, минорирующая семейство точек $\mathcal{L}_i(\hat{Y}_i)$, минорирует и функцию $\tilde{F}_i(x_i, y_i, z_{i+1})$ при

$$y_i \in \prod_{j=1}^{N_i} [y^{jn_i}, y^{jn_i}], \text{ где } y^{jn_i} \text{ и } y^{jn_i} \text{ — правая и левая границы,}$$

соответственно, изменения значения координат j -й компоненты измерения множества $\tilde{\mathcal{Y}}_i$ (рис. 10.30).

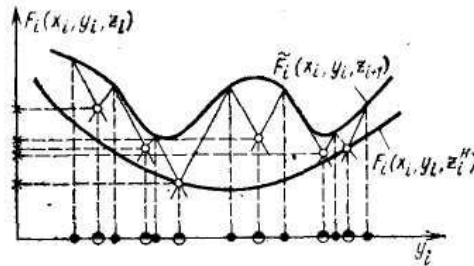


Рис. 10.30. Определение липшицевых точек (● — соответствует \hat{Y}_i ; ● — соответствует Y_{li} ; ○ — соответствует l_i)

Поэтому условия (10.112) или (10.118), присутствующие в определении задачи по нахождению Z_i^H , могут быть заменены на следующие:

$$Z_i^* = \{z_i^H: F_i(x_i, y_i, z_i^H) \leq (K(^j y_{li})); \forall ^j y_{li} \in \mathcal{Y}_{li}\} \quad (10.120)$$

Таким образом, задача нахождения значений z_i^H , учитывающая дискретность представления минорируемых функций, может быть определена условиями (10.113), (10.114), (10.120). Ее решение позволяет по конечному числу проработок рекомендаций

рассматриваемого i -го уровня на последующих детализирующих уровнях определять наиболее оптимистические значения реакций этих уровней. Будем рассматривать решение этой задачи как реализацию некоторого оператора KZN :

$$KZN: \{i, F, \mathcal{L}\} \Rightarrow z^H,$$

где множество \mathcal{L} формируется оператором $KLIP$:

$$KLIP: \{i, L, \hat{Y}, \mathcal{Y}, \tilde{F}\} \Rightarrow \mathcal{L}$$

в основе выполнения которого лежит решение задачи формирования липшицевых точек путем пересечения плоскостей, проводимых через точки $\{y_i, \tilde{F}_i(x_i, y_i, z_{i+1})\}_{y_i \in \hat{Y}_i}$, $\hat{Y}_i \subset \mathcal{Y}_i$ под углами $\pm \operatorname{arctg} L_j$, где L_j определяют с помощью выражения (10.119).

Теперь рассмотрим оператор, реализующий определение программы детализации формируемых рекомендаций. Для этого запишем очевидное условие:

$$K(y_{li}) \leq \tilde{F}_i(x_{li}, y_i, z_{i+1}) \quad (i = 1, 2, \dots, N-1),$$

свидетельствующее о том, что в любой липшицевой точке значение координаты, соответствующей критерию, не больше, чем значение функции $\tilde{F}_i(x_i, y_i, z_{i+1})$ в этой же точке. Из данного условия следует, что «поднятие» функции $\tilde{F}_i(x_{li}, y_i, z_i^H)$ без нарушения условия (10.120) требует проведения вычислений функции $\tilde{F}_i(x_i, y_i, z_{i+1})$ в точках $y_i \in \mathcal{Y}_{li}$. При этом точка, ранее бывшая липшицевой, переходит в состав \hat{Y}_i и порождает ряд новых липшицевых точек, каждая из которых лежит «выше» ее (рис. 10.31).

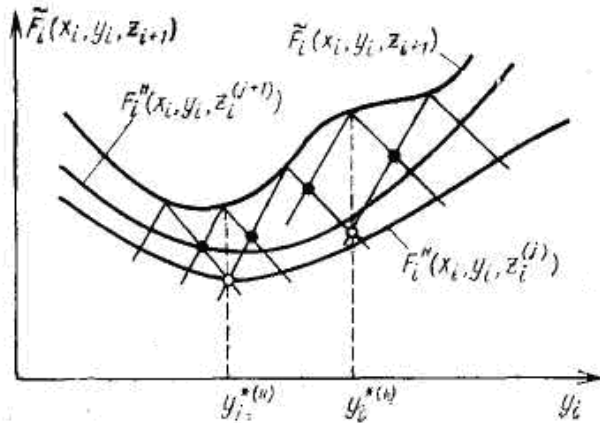


Рис. 10.31. Уточнение функции нижних границ (● — липшицевые точки, порожденные вычислением функции $\tilde{F}_i(x_i, y_i, z_{i+1})$ в точке $y_i^{*(n)}$ и точке $y_i^{*(m)}$)

Исходя из сказанного, процедуру «поднятия» функций $F_i(x_i, y_i, z_i^n)$ можно организовать следующим образом.

Полученная в результате решения задачи (10.113), (10.114), (10.120) функция $F_i(x_i, y_i, z_i^n)$ должна проходить, как правило, через N_{Z_i} или большее число липшицевых точек (N_{Z_i} — размерность вектора z_i).

Обозначим такие точки через $\tilde{\mathcal{L}}_i \subseteq \mathcal{L}_i$

$$|\tilde{\mathcal{L}}_i| \geq N_{Z_i} \quad (10.121)$$

а их проекции на \mathcal{Y}_i — через \hat{Y}_i . При этом очевидна зависимость \hat{Y}_i от множества детализированных рекомендаций \hat{Y}_i , т. е. $\hat{Y}_i = \hat{Y}_i(\hat{Y}_i)$.

Проведение вычислений функции $\tilde{F}_i(x_i, y_i, z_{i+1})$ в любых $|\hat{Y}_i| - N_{Z_i} + 1$ точках из \hat{Y}_i нарушает условие (10.121), что является свидетельством возможности изменения z_i^n в сторону увеличения значений функции $F_i(x_i, y_i, z_i^n)$. Поэтому процедуру уточнения $F_i(x_i, y_i, z_i^n)$ будем рассматривать как выполнение:

- детализации рекомендаций $y_i \in \hat{Y}_i$, т. е. определение значений $\tilde{F}_i(x_i, y_i, z_{i+1})$ в этих точках;
- соответствующей коррекции множеств \hat{Y}_i и \hat{Y}_i ;
- непосредственного определения значений z_i^n , из решения задачи (10.113), (10.114), (10.120) при скорректированном \hat{Y}_i .

Очевидно, что в данной процедуре детализации подлежат лишь рекомендации $y_{li} \in \tilde{Y}_i$ содержащиеся в множестве конкурентоспособных рекомендаций, полученном на предыдущем шаге. Поэтому множество \tilde{Y}_i может быть представлено в виде:

$$\tilde{Y}_i = \dot{Y}_i (Y^{\text{кон}}_b, \dot{Y}_i).$$

Оно тождественно программе детализации рекомендаций рассматриваемого уровня с целью уточнения функции нижних границ $F_i(x_b, y_i, z_i^H)$, введенной в рассмотрение выше и обозначенной π_b , т. е. $\pi_i \equiv \dot{Y}_i$

Формирование этой программы, заключающееся в выделении из множества липшицевых точек \mathcal{L}_i тех, которые находятся в ε -окрестности функции $F_i(x_b, y_i, z_i^H)$, при $y_i \in \mathcal{Y}_i$, будем возлагать на некоторый оператор *KPR*:

$$KPR: \{i, F, Z^H, \mathcal{Y}, \varepsilon\} \Rightarrow \pi.$$

Отметим, что приближение функции $F_i(x_b, y_i, z_i^H)$ к $\tilde{F}_i(x_b, y_i, z_{i+1})$ ограничено. В частности, при достижении касания функций

$F_i(x_b, y_i, z_i^H)$ и $\tilde{F}_i(x_b, y_i, z_{i+1})$ в числе точек, большем или равном, чем N_{Z_i} , дальнейшее приближение первой из них ко второй невозможно.

Обозначим

$$Y_i^+ = \{y_i: F_i(x_b, y_i, z_i^H) = \tilde{F}_i(x_b, y_i, z_{i+1}), y_i \in \mathcal{Y}_i\}$$

и будем считать, что, если

$$|Y_i^+| \geq N_{Z_i}, \tag{10.122}$$

то соответствующее значение z_i^H обеспечивает максимальную степень приближения $F_i(x_b, y_i, z_i^H)$ к $\tilde{F}_i(x_b, y_i, z_{i+1})$. Область \mathcal{Y}_i , на которой наблюдается условие (10.122), будем обозначать \mathcal{Y}_i^+ .

Дальнейшее приближение $F_i(x_b, y_i, z_i^H)$ к $\tilde{F}_i(x_b, y_i, z_{i+1})$ на \mathcal{Y}_i^+ возможно лишь на его подмножествах либо при изменении z_{i+1}^H . Первая из этих возможностей реализуется при уменьшении Γ_i^b — гарантированно достижимых значений критерия, когда происходит сжатие $Y^{\text{кон}}_b$, выступающего в роли исходного множества альтернатив для следующего шага. Например, если на некотором k -м шаге определена область $\mathcal{Y}_i^{(k)} \supseteq Y_i^+$, то при $\mathcal{Y}_i^{(k+1)} \subset \mathcal{Y}_i^{(k)}$ согласно определению Y_i^+ условие (10.122) может быть нарушено и тогда $\mathcal{Y}_i^{(k+1)} \not\supseteq Y_i^+$. Вторая из указанных возможностей связана с детализацией рекомендаций $(i+1)$ -го уровня, в результате чего

возможно изменение z_{i+1}^H , приводящее к уменьшению числа касаний, определяющих мощность \mathcal{Y}_i^+ .

Теперь определим оператор $RKON$, предназначенный для формирования множеств конкурентоспособных рекомендаций. В основе его выполнения лежит условие (10.91), позволяющее задать его в виде:

$$RKON: \{i, \mathcal{Y}, \Gamma^B, z^H\} \Rightarrow \{Y^{кон}, \Delta\}.$$

Покажем, что условия, положенные в основу реализации рассмотренных операторов, обеспечивают выполнение условия (10.96) и, как следствие, сходимость метода.

Вначале покажем, что гарантированный диапазон оценки рекомендаций каждого уровня $\Delta_i = \Delta(x_i, \mathcal{Y}_i, Z_i)$ содержит в себе значения критерия, соответствующего рациональным рекомендациям в целом, т. е.

$$F_N(x_i, \bar{y}_0, z_N) \in \Delta(x_i, \mathcal{Y}_i, Z_i) \quad i = 1, 2, \dots, N. \quad (10.123)$$

Предположим, что существует $J \subseteq \{1, 2, \dots, N-1\}$ такое, что $F_N(x_i, \bar{y}_0, z_N) \notin \Delta(x_j, y_j, Z_j)$ при $j \in J$. Это согласно определению гарантированного диапазона означает, что

$$F_N(x_i, \bar{y}_0, z_N) < F_j(\bar{x}_j, y_j^*, z_j^H), \quad j \in J,$$

или

$$F_N(x_i, \bar{y}_0, z_N) > F_j(\bar{x}_j, y_j^{**}, z_j^B), \quad j \in J,$$

где $\bar{x}_j = \{x_i, \bar{y}_1, \bar{y}_2, \dots, \bar{y}_{j-1}\}$.

Из данных неравенств на основании определения y^* и y^{**} можно получить:

$$F_N(x_1, \bar{y}_0, z_N) < F_j(x_j, y_j, z_j^H) \quad \text{при} \quad \forall y_j \in \mathcal{Y}_j, \quad j \in J;$$

$$F_N(x_1, \bar{y}_0, z_N) > F_j(x_j, y_j, z_j^B) \quad \text{при} \quad \forall y_j \in \mathcal{Y}_j, \quad j \in J$$

или, что то же самое,

$$F_N^{DJ}(x_1, \bar{y}_1, \bar{y}_2, \dots, \bar{y}_N) < F_j(x_j, y_j, z_j^H) \quad \text{при} \quad \forall y_j \in \mathcal{Y}_j, \quad j \in J;$$

$$F_N^{DJ}(x_1, \bar{y}_1, \bar{y}_2, \dots, \bar{y}_N) > F_j(\bar{x}_j, y_j, z_j^B) \quad \text{при} \quad \forall y_j \in \mathcal{Y}_j, \quad j \in J.$$

Полученные условия противоречат условиям (10.111), (10.112), соответственно. Это определяет сделанное предположение о нарушении (10.123) как неверное и свидетельствует, что данное условие в действительности всегда имеет место.

Теперь непосредственно рассмотрим условие (10.96), которое, в силу независимости компонент вектора y_0 , может быть представлено в

виде $\bar{y}'_i \in Y^{\text{КОН}}_i$ для $\forall i \in [1, N]$, где $\bar{y}'_i \in y_0$. Предположим, что условие (10.96) не имеет места, т. е. существует $J \subseteq [1, N] \in$ такое, что $\bar{y}'_i \in Y^{\text{КОН}}_i$. Это означает, что

$$\delta(\bar{x}_j, \bar{Y}'_j, Z_j) \cap \Delta(\bar{x}_j, \mathcal{Y}_j, Z_j) = \emptyset, j \in J$$

Отсюда следует, что

$$F_j(\bar{x}_j, y^*_j, z^*_j) < \Gamma^b_j, j \in J$$

или

$$F_j(\bar{x}_j, y^{**}_j, z^b_j) < \Gamma^h_j, j \in J$$

Учитывая соотношения (10.98), (10.99), усилим полученные неравенства:

$$F_N(x_1, \bar{y}_0, z_N) > \Gamma^b_j, j \in J;$$

$$F_N(x_1, \bar{y}_0, z_N) < \Gamma^h_j, j \in J.$$

Из данных неравенств, учитывая, что $\Delta(x_j, \mathcal{Y}_j, Z_j) = [\Gamma^h_j, \Gamma^b_j]$, следует $F_N(x_1, y_0, z_N) \notin \Delta(\bar{x}_j, \mathcal{Y}_j, Z_j)$ при $j \in J$. Это противоречит доказанному условию (10.123) и свидетельствует об истинности условия (10.106).

Покажем, что условия (10.106), (10.123) остаются в силе и при определении $Y^{\text{КОН}}_i$ из условий (10.101), где $\Gamma^b_j = \min \{F_i(x_i, y^{**}_j, z^b_j),$

$F_i(x_i, y_i, y_{i+1}, y^{**}_j, z^b_j); j \in [1, N]\}$; $y_k \in \mathcal{Y}_k$; $k = i, i + 1, \dots, N$. Для этого подставим в использовавшиеся выше неравенства вместо $F_i(x_i, y^{**}_j, z^b_j)$ значения Γ^b_j . Все эти неравенства, учитывая определения Γ^b_j , только усилятся. Отсюда следует правомочность распространения условий (10.123) и (10.106) и на $Y^{\text{КОН}}_i$, определяемых по правилу (10.101).

Особо отметим, что каждый элемент $y_i \in Y^{\text{КОН}}_i (i=1, 2, \dots, N)$ является потенциально рациональным и, следовательно, не может быть исключен из рассмотрения. Данное утверждение основывается на том, что на моделях каждого i -го уровня по имеющейся информации (множествам допустимых значений $\mathcal{Y}_j (j > i)$ или соответствующим им z^h_i и z^b_i , а также результатам моделирования в отдельных точках на моделях следующих за i -м уровнем, по которым формируется Γ^b_i, Γ^h_i) определить оценки функционала конкретнее, чем диапазоном Δ , не представляется возможным. Оценка каждого y_i в этих же условиях не может быть названа конкретнее, чем принадлежность диапазону $[F_i(x_i, y_i, z^h_i); F_i(x_i, y_i, z^b_i)]$. Следовательно, каждую рекомендацию, у которой диапазон оценок пересекается с Δ , можно рассматривать как потенциально рациональную. Именно это условие положено в основу

определения Y^{KOH}_i , являющегося минимальным подмножеством \mathcal{Y}_i , содержащим $\bar{y}_i \in \bar{Y}_0$.

10.1.3.5. Структура алгоритма формирования и последовательного сужения множеств конкурентоспособных рекомендаций

Пусть в некоторый момент определена совокупность консультационных операций, математические модели которых образуют элементарную тандемную модель с признаком одноименности K . Это является условием, требующим и допускающим проведение согласования результатов выполнения рассматриваемых совокупностей консультационных операций между собой. Уровни образовавшейся тандемной модели, представляемые операторами $F_i: \{x_i, y_i, z_i\} \Rightarrow K, (i = 1, 2, \dots, N)$, являются основными исходными данными для рассматриваемого алгоритма, который носит пошаговый характер. Целью каждого k -го шага, выполняемого на i -м уровне детализации, является определение множества конкурентоспособных рекомендаций $Y^{KOH(k)}_i$ из множества исходных для этого шага альтернатив $\mathcal{Y}^{(k)}_i$. Такими альтернативами для первого шага являются рекомендации из \mathcal{Y}_{i0} , определяемые рассматриваемой концепцией консультируемой проблемы, а для последующих шагов — множество конкурентоспособных рекомендаций, полученное на предыдущем шаге, т. е. $\mathcal{Y}^{(k)}_i = Y^{KOH(k-1)}_i, i=1, 2, \dots, N; k > 1$.

Определяющим для выделения $Y^{KOH(k)}_i$ из $\mathcal{Y}^{(k)}_i$ и формирования гарантированного диапазона Δ являются значения z^H_i и z^B_i ,

задающие Z_i соответствующее $\prod_{j=i+1}^N \mathcal{Y}_j$. Вместо значения z^B_i в

процессе формирования Y^{KOH}_i более эффективным является использование минимального из гарантированно достижимых значений критерия, обозначаемого Γ^B_i . Поэтому в дальнейшем в качестве исходных данных для определения Y^{KOH}_i принимают z^H_i и Γ^B_i . Эти значения определяются на каждом шаге результатами детализации некоторой совокупности рекомендаций $\bar{y}_i \in Y^{KOH}_i$, каждая из которых находится при выполнении текущего шага из условия максимальной конкретизации Y^{KOH}_i . Ранее такая совокупность была обозначена $\pi_i = \pi_i(Y^{KOH}_i, \hat{Y}_i)$. Рекомендации \bar{y}_i (совместно с x_i) определяют директивные данные x_{i+1} для следующего уровня детализации. На основе полученных данных в этом случае

выполняется процедура, аналогичная рассматриваемой, а результаты в виде значений $\Gamma_{i+1}^b, \Gamma_{i+1}^h$ сообщаются на i -й уровень, где значения Γ_{i+1}^b используются для понижения верхней границы Γ_i^b , а Γ_{i+1}^h — для порождения новых липшицевых точек.

В итоге всю совокупность действий, выполняемых на i -м уровне детализации, можно представить в виде некоторого оператора:

$$KVS: \{i, x, \hat{Y}, \mathcal{Y}, \alpha\} \Rightarrow \{Y^{koh}, \Gamma^h, \Gamma^b\},$$

где α_i — параметр, определяющий схему организации взаимодействия i -го и $(i - 1)$ -го уровней. В рассматриваемом случае $\alpha \sim \bar{\Delta}_i$, т. е. передача информации на $(i - 1)$ -й уровень производится по достижению заданной точности определения значений критерия. Для описания этих действий вводятся в рассмотрение следующие операторы:

- *KLIP* — реализующий формирование множества липшицевых точек по значениям вычислений функции $\tilde{F}_i(x_i, y_i, z_{i+1})$ в точках

$$y_i \in \hat{Y}_i \subseteq \mathcal{Y}_i;$$

$$KLIP: \{i, L, \hat{Y}, \mathcal{Y}, \tilde{F}\} \Rightarrow \mathcal{L};$$

- *KZN* — реализующий определение значений z_i^H по множеству липшицевых точек:

$$KZN: \{i, F, \mathcal{L}\} \Rightarrow z^H;$$

- *KPR* — реализующий выделение из множества липшицевых точек L_i тех, которые находятся в ε -окрестности функции $F_i(x_i, y_i, z_i^H)$ на \mathcal{Y}_i :

$$KPR: \{i, F, z^H, \mathcal{Y}, \varepsilon\} \Rightarrow \pi;$$

- *RKON* — реализующий выделение из множества альтернативных рекомендаций \mathcal{Y}_i конкурентоспособных:

$$RKON: \{i, \mathcal{Y}, \Gamma^b, Z^H\} \Rightarrow \{Y^{koh}, \Delta\}.$$

Логическая взаимосвязь приведенных операторов, составляющая рассматриваемый алгоритм, показана на рис. 10.32.

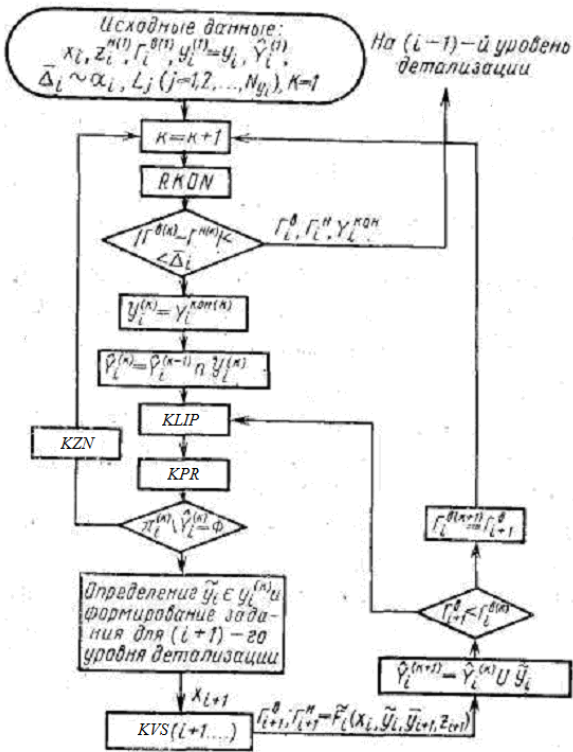


Рис. 10.32. Структура алгоритма формирования конкурентоспособных рекомендаций на i -м уровне детализации формируемых рекомендаций

Здесь исходными данными являются: $z_i^{H(1)}, \Gamma_i^{B(1)}$,

$$y_i^{(1)} = y_{i0}, \hat{y}_i^{(1)}, \bar{\Delta}, \quad (i = 1, 2, \dots, N),$$

$$x_1 = x_0, z_N = z_0, L_j \quad (j = 1, 2, \dots, N_{y_i}), \quad k=1$$

Основные действия алгоритма следующие:

1) установление номера приближений, приводящих к изменению $y_i^{кон(k)}$:

$$k = k + 1;$$

2) выделение $y_i^{кон(k)}$ из $y_i^{(k)}$ как результат выполнения оператора $RKON$;

3) окончание работы алгоритма на i -м уровне, когда $|\bar{\Delta}^{(k)} - \Gamma^{B(k)} - \Gamma^{H(k)}|$. Передача данных: $\Gamma^{B(k)}, \Gamma^{H(k)}$ на $(i - 1)$ -й уровень детализации. Если $i = 1$, то окончание процесса согласования;

4) определение множества проработанных на следующих уровнях детализации рекомендаций из $\mathcal{Y}_i^{(k)}$:

$$\hat{Y}_i^{(k)} = \hat{Y}_i^{(k-1)} \cap \mathcal{Y}_i^{(k)};$$

5) определение множества липшицевых точек (порождаемых рекомендациями $\hat{Y}_i^{(k)}$) как результат выполнения оператора *KLIP*;

6) определение программы проработки рекомендаций $y_i \in \mathcal{Y}_i^{(k)}$ (множества липшицевых точек, лежащих в ε -окрестности функции $F_i(x_i, y_i, z_i^H)$) как результат выполнения оператора *KPR*;

7) переход к 12 (к коррекции функции нижних границ), если $\pi_i^{(k)} \setminus \hat{Y}_i^{(k)} = \emptyset$;

8) определение очередной рекомендации $\tilde{y}_i \in \mathcal{Y}_i^{(k)}$ ($\tilde{y}_i \notin \hat{Y}_i^{(k)}$) для детальной проработки и формирование задания $x_{i+1} = \{x_i, \tilde{y}_i\}$ для $(i + 1)$ -го уровня, определяя в том числе и параметр α_{i+1} ;

9) проведение детализации рекомендации \tilde{y}_i на $(i + 1)$ -м уровне, как результат выполнения оператора *KVS* $(i + 1, \dots)$;

10) пополнение множества сформированных рекомендаций;

$$\hat{Y}_i^{(k+1)} = \hat{Y}_i^{(k)} \cup \tilde{y}_i$$

11) переписывание $\Gamma_i^{b(k+1)} = \Gamma_{i+1}^{b(k)}$ и переход к 1), если $\Gamma_{i+1}^b < \Gamma_i^{(k)}$, в противном случае переход к 5);

12) определение значения $z_i^{H(k+1)}$ как результат выполнения оператора *KZN*; переход к 1).

Итак, в данном разделе рассмотрены методические вопросы взаимодействия консультационных модулей применительно к вертикальному согласованию результатов их автономного функционирования. В основе изложенного лежит использование многоуровневых математических моделей консультируемых проблем, представленных так называемыми тандемными моделями, характерной чертой которых является наличие связей идентификации между их различными уровнями.

Предложенный метод согласования рекомендаций вертикально связанных консультационных модулей отличается тем, что идентификация моделей каждого уровня проводится на множестве конкурентоспособных рекомендаций, которое в процессе согласования последовательно сжимается вокруг рациональной совокупности рекомендаций. При этом на каждом шаге согласования даются верхняя и нижняя оценки значений критерия для рассматриваемых конкурентоспособных рекомендаций. Величина определяемого этими

оценками диапазона является наиболее реальной основой для прекращения процесса согласования в отличие от часто используемых в качестве такой основы разностей результатов (по критерию или параметрам рекомендаций), которые при «пологих» или «крутых» зависимостях критериев от параметров рекомендаций являются весьма непоказательными.

В заключение отметим, что в неавтоматизированном режиме решение многоуровневых задач формирования рекомендаций в общем случае возможно лишь без учета циклических связей между используемыми на каждом уровне детализации математическими моделями. Это объясняется тем, что единственным способом реализации такого рода связей является решение задачи идентификации моделей верхних менее достоверных уровней по результатам численных экспериментов на более достоверных моделях нижележащих уровней. Решение этой задачи весьма трудоемко, что делает необходимым применение средств вычислительной техники. Кроме того, организация рассматриваемых циклических связей делает необходимым проведение решения консультационных задач на моделях различных уровней и задач идентификации в виде решения единой комплексной задачи на основе общей для них программно-информационной базы, что возможно лишь в условиях САК.

10.1.4. Формирование КМ на базе управляющего модуля САК

10.1.4.1. Состав и структура алгоритмизирующих элементов управляющего модуля

Рассмотрим управляющий модуль (УМ) САК, основным назначением которого является обеспечение гибкого и оперативного формирования консультационных модулей в процессе автоматизированного формирования рекомендаций. Состав УМ модуля определяют:

- инвариантные процедуры формирования и функционирования КМ как элементы целеполагающей компоненты «процесс формирования рекомендаций»;

- ЛФР, непосредственно участвующее в процессе формирования КМ для выполнения заданной операции и далее, управляющие процессом автоматизированного выполнения этой операции. Данное ЛФР представляет в УМ элемент компоненты «организационная система»;

- инструкции для ЛФР, в качестве примера которых могут рассматриваться комментарии к излагаемым далее диалоговым

процедурам. Эти инструкции являются элементами компоненты «методическое обеспечение»;

- организационные ограничения по доступу отдельных ЛФР, ов к прикладному программно-информационному обеспечению, описывающему консультируемую проблему. Эти ограничения как элемент компоненты «организационное обеспечение» далее находят свое отражение в информационном обеспечении УМ;

- виртуальное вычислительное оборудование, на базе которого производится формирование КМ. Его составляют элементы таких компонент, как программное и информационное обеспечение, а также комплекс технических средств. Рассмотрим состав и структуру этого оборудования подробнее.

Структура алгоритмизирующего слоя УМ представляет собой ациклический граф, где каждой его вершине ставится в соответствие определенный класс консультационных процедур в их инвариантной постановке. Такого рода процедуры описаны в п. 4, где было выявлено их содержание и взаимосвязи.

Структура алгоритмизирующих элементов может быть представлена в виде дерева с вершинами E_{ij} , где через E_{ij} обозначен j -й алгоритмизирующий элемент i -го уровня. Данная структура определяется основными действиями, связанными с формированием, а в ряде случаев и с модификацией КМ. Последовательность этих действий не является заранее запрограммированной, что позволяет переходить от выполнения одного действия к другому, руководствуясь принципом достаточности исходных данных для его выполнения. При этом переход к выполнению действий, не обеспеченных исходными данными, управляющим модулем не допускается. Такой принцип управления по данным обеспечивает существенно большую гибкость при реализации процесса формирования КМ в условиях многовариантной проработки формируемых рекомендаций, когда анализ различных вариантов формируемых рекомендаций может требовать различных структур действий, особенно учитывая возможность существенного (по дням) распределения этих действий во времени.

Проиллюстрируем алгоритмизирующие элементы УМ диалоговыми процедурами, сопровождающими его функционирование. При этом второстепенные сообщения, носящие сервисный характер, в описании будем опускать.

В начале каждого сеанса работы (функционирование элемента $E_{1,1}$) система идентифицирует каждого вступающего с ней в контакт ЛФР, а, запрашивая:

1. ЗАДАЙТЕ СВОЙ ПАРОЛЬ (КОД) —*-----

(Здесь и далее каждое сообщение пронумеровывается для удобства изложения).

Если пароль код (код), в качестве которого, в частности, может использоваться фамилия ЛФР,а, задан верно, т. е. данное ЛФР имеет право работать с системой, то далее оно должно ответить на следующие сообщение.

2. ВАМ ПРЕДОСТАВЛЯЮТСЯ СЛЕДУЮЩИЕ ВОЗМОЖНОСТИ:

1. ФОРМИРОВАНИЕ ИСХОДНЫХ ДАННЫХ
2. ОПРЕДЕЛЕНИЕ СТАТУСА КМ
3. ФОРМИРОВАНИЕ КМ
4. ВЫПОЛНЕНИЕ КОНСУЛЬТАЦИОННЫХ ОПЕРАЦИЙ
5. ФОРМИРОВАНИЕ СРЕДЫ ВЗАИМОДЕЙСТВИЯ С ДРУГИМИ

КМ

В ответ на это сообщение ЛФР должно задать номер одной из названных возможностей. В случае задания «пустого ввода» происходит завершение работы в рассматриваемом режиме, и управление передается вышестоящему алгоритмизирующему элементу. В данном случае, в частности, ЛФР завершает работу с системой. Если данных для выполнения выбранной процедуры недостаточно (например, выбрано выполнение консультационной операции, для которой еще не сформирован КМ), то тогда здесь и далее в аналогичных ситуациях появляется сообщение:

3. ДЛЯ ВЫПОЛНЕНИЯ ЗАДАНИЯ НЕДОСТАТОЧНО ДАННЫХ
(*текст, поясняющий, каких именно данных не хватает*)

После этого сообщения ЛФР по команде «пустой ввод» возвращается к сообщению 2). Описанные сообщения соответствуют функционированию алгоритмизирующего элемента $E_{1,1}$.

Формирование исходных данных возлагается на алгоритмизирующий элемент $E_{2,1}$. Его функционирование включает в себя выполнение следующих действий, отраженных в сообщении, которое появляется в начале его работы

4. ВАМ ПРЕДОСТАВЛЯЮТСЯ СЛЕДУЮЩИЕ ВОЗМОЖНОСТИ:

1. ПРОСМОТР ПЕРЕЧНЯ СФОРМИРОВАННЫХ КОНЦЕПЦИЙ
2. ФОРМИРОВАНИЕ НОВОЙ КОНЦЕПЦИИ в ОБД
3. МОДИФИКАЦИЯ КОНЦЕПЦИИ
4. УДАЛЕНИЕ КОНЦЕПЦИЙ из ОБД
5. КОПИРОВАНИЕ ДАННЫХ

Первая возможность носит сервисный характер, в частности, при ее выборе появляется сообщение:

5. ВАМИ СФОРМИРОВАНЫ КОНЦЕПЦИИ

<аннотация на 1-ю сформированную концепцию> (заносятся ЛФР,ом при формировании или модификации концепции) <номер варианта ОБД, где эта концепция хранится>

.

.

.

Здесь же ЛФР имеет возможность пометить ту или иную концепцию знаком \mathfrak{A} , в результате чего ему будет выведено содержимое этой концепции, т. е. перечень переменных с определенными границами изменения непрерывных параметров и значениями дискретных параметров (см. также сообщение 44).

Возможности 2, 3, 4, 5 реализуются алгоритмизирующими элементами $E_{3,1}$, $E_{3,2}$, $E_{3,3}$, $E_{3,4}$, соответственно. Так, при выборе второй возможности в сообщении 4), т. е. при функционировании $E_{3,1}$, появляется сообщение:

6. ЗАДАЙТЕ ИМЕНА ПЕРЕМЕННЫХ И ИХ ЗНАЧЕНИЯ

*

*

В ответ на это сообщение ЛФР должен задать имена переменных, определяющих вектор y . При этом для непрерывных переменных он также должен после звездочек * задать границы рассматриваемого диапазона изменений этих переменных, а для дискретных — после первой звездочки поставить «Д», а после второй — значение дискретной переменной. После завершения выполнения этой процедуры, которое ЛФР должно определить введением «К», система самостоятельно присвоит тот или иной номер варианта в ОБД. В результате появляется сообщение:

**7. ЗАДАННАЯ ВАМИ КОНЦЕПЦИЯ ХРАНИТСЯ В ОБД В
ВАРИАНТЕ <номер варианта>
ЗАДАЙТЕ АННОТАЦИЮ**

После ввода ЛФР,ом аннотации концепция считается окончательно сформированной.

При выборе третьей и четвертой альтернатив в сообщении 4), т. е. в начале функционирования алгоритмизирующих элементов $E_{3,2}$ и $E_{3,3}$, появляется сообщение:

8. ЗАДАЙТЕ НОМЕР ВАРИАНТА В ОБД

При этом ЛФР должно указать номер варианта, где хранится ранее сформированная им концепция. Причем, если заданный номер не входит в перечень доступных ему номеров, в частности, приводимых в сообщении 5), то появится сообщение, что вариант недоступен данному ЛФР,у.

В случае функционирования элемента $E_{3,2}$ на экране появляется сообщение

9. МОДИФИЦИРУЕМАЯ КОНЦЕПЦИЯ

<имя 1-й переменной>*<значение левой границы>*<значение правой границы> <имя 2-й переменной>*Д*<значение дискретной переменной>

При этом ЛФР имеет возможность изменить ранее заданные им значения переменных, а также удалить ту или иную переменную из описания концепции или ввести дополнительную переменную.

(Для краткости описания способ выполнения этой процедуры опустим).

В случае функционирования элемента $E_{3,3}$ появляется сообщение:

10. УНИЧТОЖАЕТСЯ КОНЦЕПЦИЯ

Отметим, что режимы модификации и уничтожения концепции выбираются, как правило, если статус КМ, формируемого для анализа той или иной концепции, определяется как невыполнимый.

Если в сообщении 4) была выбрана пятая возможность, то это означает начало функционирования элемента $E_{3,4}$. Этот элемент предназначен для копирования информации из одного варианта ОБД в другой. Данная возможность необходима, в первую очередь, для обмена информацией между ЛФР,ом. Ее реализация обеспечивается следующим сообщением.

11. ДЛЯ СЧИТЫВАНИЯ ИНФОРМАЦИИ ЗАДАЙТЕ НОМЕР ВАРИАНТА

ОТКУДА СЧИТЫВАТЬ-----*-----

КУДА СЧИТЫВАТЬ-----*-----

Справа от звездочки на первой строке ЛФР должно указать номер варианта, где ему, например, подготовлена информация от другого ЛФР,а. Этот номер должен быть ему сообщен этим ЛФР,ом (см., например, сообщение 47). Во второй строке ЛФР должно указать номер доступного ему варианта ОБД.

Таким образом, на основе использования алгоритмизирующего элемента $E_{2,1}$ ЛФР может сформировать вариант данных в ОБД, содержащий как описания исходных данных, при которых должна

быть выполнена текущая консультационная операция, так и описание предлагаемой им консультационной концепции. Другими словами, в результате функционирования элемента $E_{2,1}$ в ОБД так или иначе формируется вариант данных, содержащий:

а) перечень параметров консультационной проблеме и диапазоны возможных изменений каждого из них. Данная информация представляет собой введенное выше в рассмотрение множество концептуально неотличимых консультационных рекомендаций. Обоснованное выделение из этого множества одной или группы конкурентоспособных рекомендаций составляют суть всех последующих действий в процессе выполнения соответствующей консультационной операции;

б) перечень параметров, значения которые переданы ему другим ЛФР,ом, а также сами эти значения. В общем случае все эти параметры описывают реакции, действующие на рассматриваемую консультационную операцию со стороны других выполняемых в процессе формирования рекомендаций операций. Значения этих реакций могут быть заданы как «точками», так и диапазонами их возможных изменений.

Рассмотренная в качестве примера методика формирования исходных данных основана на использовании диалоговых процедур, организованных по принципу «меню». Возможны и другие, более сложные методики, основывающиеся, в частности, на использовании формальных правил генерирования альтернативных концепций рекомендаций. Введение их в состав возможностей системы означает добавление в состав алгоритмизирующего слоя системы соответствующих элементов (в данном случае элементов третьего уровня). Причем такое расширение как в данном случае, так и в рассматриваемых далее не влияет на функционирование других алгоритмизирующих элементов.

В дальнейшем будем считать, что для каждой консультационной концепции, рассматриваемой в рамках выполнения текущей консультационной операции, в ОБД на базе элемента $E_{2,1}$ формируется вариант данных, описывающих как саму эту концепцию, так и соответствующие ограничения.

Следующим действием в процессе формирования КМ является определение статуса формируемого модуля и, если это необходимо, то и доопределение решаемых задач. Необходимость такого доопределения состоит в том, что в ряде случаев ЛФР должно иметь возможность исследовать множество допустимых формируемых рекомендаций в режиме параметрического анализа. Для этого ему

должна быть предоставлена возможность свести задачу к расчетному статусу и провести расчеты в отдельных, выбираемых им точках множества допустимых формируемых рекомендаций. Если же он предпочтет решать задачу оптимизации, то ему необходимо задать критерии, по значениям которых он должен оценивать альтернативы формируемых рекомендаций. Выполнение рассмотренных действий возложим на элемент $E_{2,2}$.

Методика определения статуса операций на базе использования оператора *KON* подробно рассматривалась в п. 10.1.2.2. Будем считать, что реализация этой методики возлагается на элемент $E_{3,5}$. Его функционирование заключается в том, что при определении ЛФР, о той или иной консультационной концепции он выдает сообщение либо о несоответствии данной концепции директивным данным ($|Y| = 0$), либо об одноэлементности множества допустимых рекомендаций данной концепции ($|Y|=1$), либо о множественности таких рекомендаций ($|Y| > 1$). В зависимости от полученного сообщения ЛФР строит свои дальнейшие действия.

Здесь же заметим (см. п. 10.1.2.2), что в ряде случаев, когда исходные для выполнения консультационной операции данные не фиксированы, а ограничены, при определении статуса КМ необходимо решать соответствующие оптимизационные задачи. Формулирование этих задач и переход к их решению (элемент $E_{4,1}$) должны происходить, как правило, без участия ЛФР.

В результате функционирования элемента $E_{3,5}$, использующего в качестве исходных данных как и информацию о математической модели консультационной проблемы, так и данные из соответствующего варианта ОБД, появляется сообщение:

12. СТАТУС ФОРМИРУЕМОГО КМ ОПРЕДЕЛЕН

В случае нерасчетного статуса появляется сообщение

13. СТАТУС КМ НЕРАСЧЕТНЫЙ

ВЫ МОЖЕТЕ

1. ИЗМЕНИТЬ КОНЦЕПЦИЮ
2. СФОРМУЛИРОВАТЬ НОВУЮ КОНЦЕПЦИЮ
3. ОТКАЗАТЬСЯ ОТ РАССМАТРИВАЕМОЙ КОНЦЕПЦИИ
4. ПЕРЕЙТИ К РЕШЕНИЮ ЗАДАЧИ СОГЛАСОВАНИЯ

При выборе первой возможности управление переходит к элементу $E_{3,2}$, при выборе второй возможности — к элементу $E_{3,1}$, при выборе третьей возможности — к элементу $E_{3,3}$, а при выборе четвертой — к элементу $E_{3,12}$.

Если система определила статус формируемого КМ как оптимизационный, то появляется следующее сообщение, свидетельствующее о передаче управления элементу $E_{3,7}$.

Если статус формируемого КМ расчетный, то начинает работать элемент $E_{3,6}$.

В начале работы элемента $E_{3,7}$ появляется сообщение:

**14. СТАТУС ЗАДАЧИ: ОПТИМИЗАЦИОННЫЙ
РЕКОМЕНДАЦИЯ БУДЕТ ФОРМИРОВАТЬСЯ:**

1. ПУТЕМ ДОЗАДАНИЯ ИСХОДНЫХ ДАННЫХ
2. ПУТЕМ ОПТИМИЗАЦИИ

Выбор 1-й альтернативы означает, что ЛФР хочет свести задачу к расчетной и решать ее в режиме параметрического анализа. В этом случае на основе исполнения оператора *KON*, в частности определения ΔJ , определяются переменные, задав значения которых исходную задачу можно свести к расчетной. ЛФР,у в данном случае предоставляется перечень переменных из $P(L)$, задавая значения, части из которых он тем самым формирует ΔJ^* (см. п. 10.1.2.4). Для реализации этой возможности служит сообщение:

**15. ДЛЯ ПЕРЕЧИСЛЕННЫХ НИЖЕ ПЕРЕМЕННЫХ
ЗАДАЙТЕ ЗНАЧЕНИЕ ДЛЯ (число) ПЕРЕМЕННЫХ**

<имя 1-й переменной> *-----

<имя 2-й переменной> *-----

.

.

.

Здесь справа от звездочки в строке с переменной, которую ЛФР хочет добавить к исходным, оно задает ее значение или помечает символом и, определяя тем самым, что это значение оно задаст позже. После этого кадр повторяется, однако число задаваемых переменных уменьшается на единицу и из предлагаемого списка исключаются переменные, которые могут быть вычислены по известным уже данным. В результате постановка исходной оптимизационной задачи изменяется и задача становится расчетной.

Если в ответ на сообщение 14 ЛФР выбрало вторую возможность, то появляется сообщение:

**16. СТАТУС ЗАДАЧИ: ОПТИМИЗАЦИОННЫЙ
ЗАДАЙТЕ КРИТЕРИИ ОЦЕНКИ ВАРИАНТОВ:**

Действия ЛФР,а в данном случае состоят в том, что, начиная с первой свободной строки, оно вводит по одному имена переменных, значения которых будут использоваться для оценки различных

вариантов формируемых рекомендаций. Если будет введен символ «?», то на экране появятся все переменные, которые для данной задачи могут быть выбраны в качестве критериев, и ЛФР может отметить необходимые ему переменные. Если ЛФР задало только один критерий, то задача оптимизации является скалярной, в противном случае — векторной.

Таким образом, в результате работы элемента $E_{2, 2}$ формируется постановка консультационной задачи, которая согласуется с математической моделью консультируемой проблемы, представленной в САК в виде ППП. На следующем этапе требуется сформировать или выбрать методику решения этой задачи. При этом под формированием методики ниже подразумевается построение диалоговых процедур основывающихся как на использовании формальных алгоритмов, так и участии ЛФР,а в процессе решения поставленной задачи.

Рассмотрим теперь элемент $E_{2, 3}$, на базе которого могут решаться задачи как векторной, так и скалярной оптимизации, а также расчетные задачи. Очевидна «вложенность» этих задач, элемент $E_{3, 8}$ будем отождествлять с задачами формирования рекомендаций по векторному критерию. Передача управления элементу $E_{3,8}$ производится в случае, если в ответ на сообщение 16) ЛФР задало более одного критерия (если критерий один, то управление передается элементу $E_{4,1}$ а если задача расчетная — то элементу $E_{5,1}$). Методики формирования рекомендаций при нескольких критериях в своей основе содержат два вида действий, первое из которых заключается в управлении переходами от одной недоминирующей рекомендации к другой, а второе — в решении задачи скалярной оптимизации. Соответственно, выполнение данных действий возложим на функциональные элементы следующего уровня за $E_{3, 8}$, т. е. на $E_{4,2}$ и $E_{4,1}$.

Управление переходами от одной недоминирующей рекомендации к другой в общем случае может быть основано как на формальных условиях выбора, так и непосредственно отдано ЛФР,у. Во втором случае функционирование элемента $E_{4, 2}$ заключается в предоставлении ЛФР,у информации об очередной полученной недоминирующей рекомендации (параметры этой рекомендации и соответствующие значения критериев) и назначении им значений параметров управления переходом.

В роли таких параметров могут выступать, например, либо значения весовых коэффициентов, либо значения всех, кроме некоторой одной, компоненты вектора критериев.

Выбрать методику построения недоминирующих элементов позволяет следующее сообщение:

17. УКАЖИТЕ МЕТОДИКУ ПОСТРОЕНИЯ НЕДОМИНИРУЮЩИХ ЭЛЕМЕНТОВ:

1. СВЕРТКА КРИТЕРИЕВ
2. ФИКСАЦИЯ КРИТЕРИЕВ

В результате выбора первой альтернативы появляется следующий запрос:

18. УКАЖИТЕ ФОРМУЛУ СВЕРТКИ КРИТЕРИЕВ

<имя первой переменной, выбранной критерием>

<имя второй переменной, выбранной критерием>

<коэффициенты>

В ответ на этот запрос ЛФР должно явно в алгебраическом виде, используя синтаксис языка высокого уровня (например, С), задать формулу свертки критериев $F=F(K, a)$, определив предварительно обозначения компонент.

В результате выбора второй альтернативы построение точек множества Парето происходит путем фиксации всех критериев, кроме одного, и появляется запрос:

19. ОПРЕДЕЛИТЕ СВОБОДНЫЙ КРИТЕРИЙ, ПОМЕТИВ ЕГО СИМВОЛОМ «К»

(имя 1-го критерия) *-----

(имя 2-го критерия) *-----

.

.

Здесь ЛФР должно среди всех ранее определенных критериев выделить главный, а для всех остальных, которые должны быть зафиксированы, на мониторе появляется запрос:

20. ЗАДАЙТЕ ЗНАЧЕНИЯ ФИКСИРУЕМЫХ КРИТЕРИЕВ

(1-й фиксируемый критерий) *-----

(2-й фиксируемый критерий) *-----

.

.

«Ц» — цикл, «С» — по списку, «А» — алгоритм

В позициях справа от звездочек ЛФР может задать либо конкретную числовую информацию, либо один из символов: «Ц», «С», «А». Символ «Ц» означает, что задача построения множества недоминирующих рекомендаций будет решаться при значениях фиксируемых критериев, изменяющихся в цикле; символ «С» — при их изменении по некоторому списку; символ «А» означает, что задача выбора и, соответственно, переход от одной недоминирующей точки к

другой будет осуществляться по некоторому алгоритму выбора, реализуемому соответствующей программой.

Если ЛФР указало перебор в цикле или по списку значений, то запросы на конкретные данные расчета аналогичны сообщениям 35)–37).

В случае использования формальных алгоритмов формирования множества недоминирующих рекомендаций система запрашивает:

21. ЗАДАЙТЕ ИМЯ АЛГОРИТМА ВЫБОРА *-----

и ЛФР должно задать имя алгоритма, которым оно хочет воспользоваться.

Как уже отмечалось, в настоящее время предлагается ряд подходов к решению задач выбора формируемых рекомендаций. Один из таких подходов (метод ПРИНН), основанный на агрегированном учете неопределенностей, наиболее соответствующий, по мнению автора, реальным условиям формирования рекомендаций, рассматривается п. 10.2.

Функционирование элемента $E_{4, i}$ может быть проиллюстрировано следующими сообщениями:

22. РАБОТАЕТ РЕЖИМ ОПТИМИЗАЦИИ

**ЗАДАЙТЕ НАЧАЛЬНОЕ ПРИБЛИЖЕНИЕ И ГРАНИЦЫ
ДЛЯ ВАРЬИРУЕМЫХ ПЕРЕМЕННЫХ**

**ПРОБЕЛЫ ВОСПРИНИМАЮТСЯ КАК ОТСУТСТВИЕ
ОГРАНИЧЕНИЙ (имя первой переменной)**

·
·
·

НАЧАЛЬНОЕ ПРИБЛИЖЕНИЕ: ----*-----

НИЖНЯЯ ГРАНИЦА: -----*-----

ВЕРХНЯЯ ГРАНИЦА: -----*-----

В выделенных полях ЛФР,ом задаются численные значения или номер варианта ОБД. Для начального приближения значение должно обязательно быть задано, а отсутствие числа в полях для границ означает, что ограничения на варьируемую переменную соответственно снизу или (и) сверху не накладываются.

После ответа на приведенное сообщение система запрашивает

**23. УКАЖИТЕ МЕТОД РЕШЕНИЯ ЗАДАЧИ
ОПТИМИЗАЦИИ**

****-----****

Здесь в поле, выделенном звездочками, должен быть задан шифр метода оптимизации. Если заданного метода нет в таблице методов оптимизации (см. п. 10.1.4.2), то выдается сообщение:

НЕВЕРНО ЗАДАН МЕТОД ОПТИМИЗАЦИИ

Указав в поле между звездочками символ «?», ЛФР получает на экране перечень всех допустимых методов оптимизации с их краткой аннотацией, после чего запрос метода повторяется.

После того, как задан метод оптимизации, запрашивается информация, необходимая для работы выбранного метода оптимизации.

24. ЗАДАЙТЕ ЗНАЧЕНИЯ ПАРАМЕТРОВ МЕТОДА ОПТИМИЗАЦИИ

(название первого параметра метода) * -----

(название второго параметра метода) *-----

·
·
·

Справа от звездочки для всех параметров задаются значения. Следующий запрос позволяет ЛФР,у задать параметры решения задачи оптимизации:

РАБОТАЕТ РЕЖИМ ОПТИМИЗАЦИИ

ЗАДАЙТЕ ПАРАМЕТРЫ РЕШЕНИЯ ЗАДАЧИ ОПТИМИЗАЦИИ

ТОЧНОСТЬ РЕШЕНИЯ ЗАДАЧИ *-----

ЧАСТОТА СОХРАНЕНИЯ ИНФОРМАЦИИ *-----

ЧАСТОТА ВЫДАЧИ ИНФОРМАЦИИ НА МОНИТОРЕ *-----

В выделенных полях в соответствующих строках ЛФР,ом задаются соответствующие значения, после чего появляется сообщение

ОПТИМИЗАЦИОННЫЙ КМ

ФОРМИРУЕТСЯ

Данное сообщение означает, что управление переходит к элементам реализующего слоя системы (см. п. 10.1.4.3). В частности, к элементам, осуществляющим генерацию рабочей программы, соответствующей той информации, которую ЛФР задало на этапе функционирования алгоритмизирующего слоя.

Когда рабочая программа вычислена целевой функции (критерия оценки) и ограничений сформирована, появляется сообщение:

ОПТИМИЗАЦИОННЫЙ КМ СФОРМИРОВАН

На этом формирование консультационного модуля завершается и ЛФР,у предоставляются следующие возможности:

25. 1. ПОСМОТРЕТЬ ПОЛНУЮ ПОСТАНОВКУ ЗАДАЧИ

2. ПЕРЕЙТИ К РЕШЕНИЮ ЗАДАЧИ ОПТИМИЗАЦИИ

3. ЗАКОНЧИТЬ РАБОТУ В РЕЖИМЕ ОПТИМИЗАЦИИ

Если он выбирает первую возможность, то появляется информация о:

- фиксированных переменных задачи,
- варьируемых переменных задачи,
- параметрах задачи оптимизации,
- настроечных параметрах метода.

Если ЛФР выбирает режим решения задачи оптимизации, то появляется сообщение:

РАБОТАЕТ ПРОГРАММА ОПТИМИЗАЦИИ

В процессе выполнения программы оптимизации система через каждые N шагов (где N — заданная ЛФР,ом частота выдачи информации) выдает получающиеся результаты оптимизации: текущее значение варьируемых переменных, значение критерия, а также дополнительную информацию в зависимости от конкретного метода. Кроме того, во время этого выполнения ЛФР, осуществив, например, «пустой ввод», может остановить ход процесса оптимизации и получить доступ к следующим возможностям, которые сообщаются в результате этого прерывания:

26. РАБОТАЕТ РЕЖИМ ОПТИМИЗАЦИИ

ВАМ ПРЕДОСТАВЛЯЮТСЯ СЛЕДУЮЩИЕ ВОЗМОЖНОСТИ:

1. ПОЛУЧИТЬ СПРАВОЧНУЮ ИНФОРМАЦИЮ
2. РАСПЕЧАТАТЬ РЕЗУЛЬТАТЫ, ПОЛУЧЕННЫЕ К НАСТОЯЩЕМУ МОМЕНТУ
3. ИЗМЕНИТЬ ПАРАМЕТРЫ ВЫДАЧИ ИНФОРМАЦИИ
4. ИЗМЕНИТЬ ТОЧНОСТЬ РЕШЕНИЯ ЗАДАЧИ
5. ИЗМЕНИТЬ ПАРАМЕТРЫ МЕТОДА ОПТИМИЗАЦИИ
6. ПРОДОЛЖИТЬ ПРОЦЕСС ОПТИМИЗАЦИИ
7. ЗАДАТЬ ДРУГОЕ НАЧАЛЬНОЕ ПРИБЛИЖЕНИЕ
ВАРИРУЕМЫХ ПЕРЕМЕННЫХ
8. ВЫБРАТЬ ДРУГОЙ МЕТОД ОПТИМИЗАЦИИ
9. ЗАКОНЧИТЬ РАБОТУ ПРОГРАММЫ ОПТИМИЗАЦИИ

В результате, если выбрано действие, связанное с изменением того или иного параметра задачи оптимизации, система сообщает текущее в данный момент значение этого параметра и предоставляет возможность его изменения. Например, если выбрана возможность под номером 8, то появляется сообщение

27. ИЗМЕНИТЕ МЕТОД ОПТИМИЗАЦИИ

ИСПОЛЬЗУЕТСЯ МЕТОД — * <название использовавшегося метода>

ЗАДАЙТЕ ИМЯ НОВОГО МЕТОДА — *-----

При этом ЛФР справа от звездочки должно ввести имя метода, на который он должен сменить ранее использовавшийся метод. Тогда задача оптимизации будет решаться новым методом с прерванной точки. Таким образом, в процессе функционирования консультационного модуля ЛФР,у предоставляется возможность модифицировать этот модуль.

По достижению заданной точности решения задачи оптимизации выдается сообщение:

28. ПРОЦЕСС РЕШЕНИЯ ЗАДАЧИ ОПТИМИЗАЦИИ ЗАВЕРШЕН
РЕЗУЛЬТАТЫ РЕШЕНИЯ ЗАДАЧИ

·
·
·

ЧТО ВЫ ХОТИТЕ ДЕЛАТЬ ДАЛЬШЕ:

1. ЗАВЕРШИТЬ РАБОТУ
2. ЗАНЕСТИ РЕЗУЛЬТАТЫ В БАНК
3. ПЕРЕФОРМУЛИРОВАТЬ ЗАДАЧУ
4. СОХРАНИТЬ КМ КАК ТИПОВОЙ

Теперь рассмотрим процедуры подготовки к решению расчетных задач, что возлагается на элемент $E_{5,1}$. Возможности решения такого рода задач должны включать не только проведение «точечных» расчетов, но и расчетов в цикле, по списку значений исходных параметров и ряд других возможностей. Пример таких возможностей отражен в следующем сообщении, соответствующем началу работы элемента $E_{5,1}$, основным назначением которого является формирование консультационных модулей расчетного статуса.

29. СТАТУС ЗАДАЧИ: РАСЧЕТНЫЙ
ВАМ ПРЕДОСТАВЛЯЮТСЯ СЛЕДУЮЩИЕ ВОЗМОЖНОСТИ:

1. ИЗМЕНЕНИЕ ИСХОДНЫХ ДАННЫХ *-----
2. ПОЛНЫЙ РАСЧЕТ ЗАДАЧИ *-----
3. ЧАСТИЧНЫЙ РАСЧЕТ ЗАДАЧИ *-----
4. РАСЧЕТ В ЦИКЛЕ *-----
5. РАСЧЕТ СПИСКОМ *-----
6. ПРОСМОТР РЕЗУЛЬТАТОВ *-----
7. ЗАНЕСЕНИЕ РЕЗУЛЬТАТОВ В ОБД *-----
8. СОХРАНЕНИЕ МОДУЛЯ КАК ТИПОВОГО *-----
9. ПОЛУЧЕНИЕ ДОПОЛНИТЕЛЬНОЙ ИНФОРМАЦИИ *-----

Если указано изменение исходных данных, то появляется следующее сообщение:

30. ИСХОДНЫЕ ПЕРЕМЕННЫЕ ЗАДАЧИ:

<имя переменной>-----М:-----*---

РАЗМЕРНОСТЬ: <физические единицы> *ЗНАЧЕНИЕ: <число>

.
. .
.

В выделенном поле для нового значения должно быть задано число или «БД — номер варианта», что означает выборку значения из соответствующего варианта БД. Пустой ввод сохраняет старое значение переменной.

Если в ответ на сообщение 29) указан частичный расчет, то следующее сообщение позволяет указать непосредственно те переменные, по которым ЛФР хочет провести расчеты:

31. РАСЧЕТ ПО ОТДЕЛЬНЫМ ПЕРЕМЕННЫМ

имя первой переменной *-----

.
. .
.

ПЕРЕМЕННЫЕ, ЗНАЧЕНИЯ КОТОРЫХ ВЫ ХОТИТЕ ПОЛУЧИТЬ, ПОМЕТЬТЕ СИМВОЛОМ «□»

При этом на экран выводятся все переменные, которые могут быть определены при решении сформулированной ранее задачи.

Дополнительные возможности расчетного режима обеспечивает следующий запрос:

32. УКАЖИТЕ, ТРЕБУЕТСЯ ЛИ ПОШАГОВАЯ ВЫДАЧА ИНФОРМАЦИИ *-----

Если ЛФР выберет эту возможность, то вычисления каждой переменной будут сопровождаться следующим сообщением;

33. РЕЗУЛЬТАТЫ ВЫЧИСЛЕНИЯ <имя переменной>

ЗНАЧЕНИЕ: <число>* РАЗМЕРНОСТЬ: <физические единицы>

.
. .
.

При пустом вводе система переходит к обработке следующего модуля, «К» в нижней строке с позиции курсора — завершение процесса вычислений и переход к сообщению 29).

Если при постановке задачи задание значений некоторых переменных было отложено, то в процессе вычислений может появиться запрос следующего вида:

34. ВАМ СЛЕДУЕТ ВЫПОЛНИТЬ ОТЛОЖЕННОЕ РАНЕЕ ЗАДАНИЕ ЗНАЧЕНИЯ ДЛЯ ПЕРЕМЕННОЙ:

<имя первой переменной> ----*----

.
. .
. . .

Если в ответ на сообщение 29 был выбран расчет в цикле или списком, то для того, чтобы определить переменную, которая должна быть аргументом расчета в цикле или списком, поступает следующее сообщение:

35. ИСХОДНЫЕ ПЕРЕМЕННЫЕ РАСЧЕТА В ЦИКЛЕ

<имя первой переменной> *-----

.
. .
. . .

ПОМЕТЬТЕ ПЕРЕМЕННУЮ - АРГУМЕНТ СИМВОЛОМ «Ц»

Здесь справа от звездочки в строке с исходной переменной, значение которой должно меняться при расчете в цикле (списком), отмечается символом «Ц» («С»).

Примечание. При запросе расчета списком первая и последняя строки сообщения изменяются соответствующим образом.

При запросе расчета в цикле следующее сообщение позволяет задать конкретные параметры проведения вычислений:

36. ДАННЫЕ ДЛЯ РАСЧЕТА В ЦИКЛЕ:

ПЕРЕМЕННАЯ ЦИКЛА: <имя переменной>

ЗАДАЙТЕ СЛЕДУЮЩИЕ ЗНАЧЕНИЯ:

НАЧ. ЗНАЧЕНИЕ: -----*

ШАГ: -----*

КОЛ-ВО ШАГОВ: -----*

При этом в выделенных полях должны быть заданы числовые значения.

При запросе расчета списком следующее сообщение позволяет задать конкретную информацию для проведения расчетов:

37. ДАННЫЕ ДЛЯ РАСЧЕТА СПИСКОМ

ПЕРЕМЕННАЯ СПИСКА: <имя переменной>

КОЛ-ВО ЭЛЕМЕНТОВ СПИСКА (НЕ БОЛЕЕ 50): -----

ЗНАЧЕНИЯ: -----

При этом в соответствующих строках задаются количество элементов списка и непосредственные значения через запятую или пробел. В ходе проведения расчетов в цикле или списком появляются соответствующие сообщения, а после нормального или по условию, заданному в п. 36), завершения опять появится сообщение 29), но список альтернатив пополнится следующей:

ПОСТРОЕНИЕ ГРАФИКА *-----

Если ЛФР выбрало построение графика, то последовательность следующих запросов обеспечивает выбор устройств для вывода графика и включение их в состав формируемого консультационного модуля и задание особенностей рисования этого графика.

Выбор в ответ на сообщение 29) 7-й альтернативы вызывает появление следующего запроса:

38. УКАЖИТЕ НОМЕР ВАРИАНТА В БАНКЕ ДАННЫХ
_*-----

При выборе в ответ на сообщение 29) 8-й альтернативы появляется следующий запрос

**УКАЖИТЕ НЕОБХОДИМОСТЬ СОХРАНЕНИЯ
ФОРМИРУЕМОГО КМ ***

Если справа от звездочки указать «ДА», то состав сформированного КМ будет зафиксирован.

Заметим, что рассмотренная процедура формирования расчетных консультационных модулей учитывает, что в процессе вычислений может потребоваться решать системы уравнений. В этом случае от ЛФР,а требуется задать метод решения этих уравнений. В рассматриваемом примере в качестве таких методов используются методы оптимизации. В случае, когда расчетная задача требует решения систем уравнений, появляется сообщение:

**39. СТАТУС ЗАДАЧИ: РАСЧЕТНЫЙ (ОБРАТНАЯ ЗАДАЧА)
СРЕДИ ПЕРЕЧИСЛЕННЫХ НИЖЕ ПЕРЕМЕННЫХ
ОТМЕТЬТЕ СИМВОЛАМИ «НП» <число> ВАРЬИРУЕМЫХ
ПЕРЕМЕННЫХ**

<имя 1-й переменной> *-----
<имя 2-й переменной> *-----

·
·
·

Здесь необходимо справа от звездочки в строке с переменной, за счет варьирования которой можно получить решение, набрать «НП» (независимая переменная). После этого кадр повторяется, однако число

выбираемых переменных уменьшается на единицу и, возможно, меняется перечень переменных для выбора. Данная процедура выполняется на базе оператора *KON* и не допускает, чтобы в качестве варьируемых выбрались взаимозависимые переменные.

Далее от ЛФР,а требуется задать вид невязки, которая далее выступает в роли критерия при решении систем уравнений методами оптимизации.

40. УКАЖИТЕ ВИД КРИТЕРИЯ ОЦЕНКИ

1. СРЕДНЕКВАДРАТИЧНЫЙ
2. СРЕДНЕЕ АБСОЛЮТНОЕ ОТКЛОНЕНИЕ
3. ОТНОСИТЕЛЬНОЕ ОТКЛОНЕНИЕ

Для переменных, помеченных как варьируемые, сообщение позволяет задать нижнюю и (или) верхнюю границы варьирования и начальное приближение, используемые для решения задачи оптимизации. Дальнейшие сообщения аналогичны тем, что и для формирования КМ оптимизационного статуса.

Следующей процедурой, которая должна быть выполнена в процессе формирования консультационных модулей, является определение информационных связей между формируемым модулем и другими консультационными модулями в САК. Эта процедура выполняется на базе оператора *KON*. Ее реализация может быть произведена, как только зафиксирована в ОБД та или иная консультационная концепция. Рассматриваемый пример соответствует именно такому подходу. При этом, как только завершается задание перечня переменных, определяющих вектор y для формируемой концепции, начинают функционировать элементы $E_{3,9}$, $E_{3,10}$, на которые возлагаются функции по выявлению горизонтальных и вертикальных информационных связей соответственно.

Исходными данными для выполнения этой процедуры являются перечни переменных, определяющих концепции, исследуемые различными консультационными модулями, организационные взаимосвязи между ЛФР,ми, сформировавшими эти концепции [эти взаимосвязи отражены в таблице ЛФР,ов (см. п. 10.1.4.2), а также математическая модель консультируемой проблемы.

Функционирование элемента $E_{3,9}$, на который возложено выявление горизонтальных связей, основывается на использовании оператора *KON*. Функционирование элемента $E_{3,10}$, на который возложено выявление вертикальных информационных связей, базируется на выявлении среди моделей операций (детализирующую рассматриваемую) тех, которые составляют с моделью этой операции уровни тандемной модели. Одним из способов формирования

тандемных моделей также является использование оператора *KON* (см. п. 10.1.2.2).

В результате функционирования этих элементов определяется «степень свободы» выбора рекомендаций *y*. В частности, теперь все компоненты этого вектора могут быть разделены на те, значения которых не зависят от результатов выполнения других консультационных операций, и на те, значения которых должны быть согласованы. Соответствующее «помечивание» переменных производится в содержащих рассматриваемые концепции вариантах ОБД. Так, если в сообщении 5) ЛФР пометило знаком и ту или иную концепцию, то появляется сообщение:

41. <имя *i*-й переменной> * <значение левой границы> * <значение правой границы> * <характеристика переменной>

·
·
·

При этом в поле «характеристика переменной» будет указано либо «СВОБОДНАЯ», либо «ТРЕБУЕТ СОГЛАСОВАНИЯ С КОНЦЕПЦИЕЙ, СФОРМИРОВАННОЙ <пароль ЛФР,а>, ХРАНЯЩЕЙСЯ В <число> ВАРИАНТЕ ОБД» Руководствуясь этой информацией, ЛФР должно выбирать значения требующих согласования рекомендаций, уже учитывая, что оно не полностью свободно в своем выборе.

После окончания автономного выполнения консультационной операции, признаком чего является занесение в ОБД либо конкретных значений переменных *y*, либо близость правой и левой границ возможных изменений по всем компонентам *y*, появляется сообщение:

42. ВАМ НЕОБХОДИМО СОГЛАСОВАТЬ РЕЗУЛЬТАТЫ РЕШЕНИЯ ЗАДАЧИ СО СЛЕДУЮЩИМИ ЛФР,МИ

<пароль ЛФР,а> — ВАРИАНТ <номер варианта> *-----

ПО <число> (данное число соответствует *St*, а предоставляемые параметры, по которым возможно проведение согласования, определяются как ΦSt (см. п. 10.1.2.2) ПАРАМЕТРАМ

<наименование 1-го параметра>

·
·
·

Приведенное сообщение соответствует задаче горизонтального согласования. Задачам вертикального согласования соответствует сообщение:

43. РЕШЕНЫ ЗАДАЧИ, РЕЗУЛЬТАТЫ КОТОРЫХ ПОЗВОЛЯЮТ УТОЧНИТЬ РАНЕЕ ЗАДАННЫЕ ВАМИ ЗНАЧЕНИЯ ПРОГНОЗНЫХ ПАРАМЕТРОВ:

<1-й прогнозный параметр>

.
. .
.

<шифр ЛФР, решившего 1-ю задачу>

.
. .
.

Следующая группа алгоритмизирующих элементов связана с формированием среды взаимодействия КМ. При этом часть этой среды, а именно, модели, трансформирующие результаты функционирования одних КМ в исходные данные для других, формируется элементом $E_{5,1}$, формирование оставшейся части — методик обмена информацией между взаимодействующими КМ в процессе согласования результатов их автономного функционирования — возложим на элементы $E_{3,11}$ (для горизонтального согласования) и $E_{3,12}$ (для вертикального согласования).

Выбрать режим решения задачи согласования позволяет следующее сообщение:

**44. УКАЖИТЕ РЕЖИМ ПРОВЕДЕНИЯ СОГЛАСОВАНИЯ
ДИАЛОГОВЫЙ *-----
ЭЛЕКТРОННАЯ ПОЧТА *-----**

Вне зависимости от выбранного режима после ответа ЛФР,а на предыдущее сообщение запрашивается:

**45. ЗАДАЙТЕ МЕТОДИКУ СОГЛАСОВАНИЯ
*-----***

В ответ ЛФР должно задать между звездочками имя выбираемой методики. Если решается задача горизонтального согласования, то предполагается, что выбранная методика «устраивает» обоих ЛФР,ов.

Решение задачи горизонтального согласования, возлагаемое на элемент $E_{3,11}$, заключается в минимизации потерь в эффективности формируемых рекомендаций каждой операции с точки зрения их локальных критериев. В общем случае потери эффективности рекомендаций одной операции приводят к увеличению эффективности другой. Поэтому в основу методики функционирования элемента $E_{3,11}$ положим построение кривой компромисса в координатах,

определяемых компонентами критериев, используемых при оценке эффективности рекомендаций каждого из рассматриваемых КМ. При этом предполагается, что построение этой кривой производится путем взаимных уступок, а построение каждой точки кривой компромисса представляет собой определение рациональных сформированных рекомендаций при исходных данных, переданных от взаимодействующей операции, т. е. требует функционирования элемента $E_{3,8}$, который является составным элементом $E_{2,3}$ (именно этим обстоятельством он ранее был отнесен к элементам третьего уровня).

Функционирование элемента $E_{3,11}$ может быть проиллюстрировано следующими сообщениями:

46. ЗАДАЙТЕ УДОВЛЕТВОРЯЮЩЕЕ ВАС ЗНАЧЕНИЕ

<имя 1-го согласуемого параметра> *-----

.
.
.

В ответ на данный запрос ЛФР должно в позициях после звездочки задать свои предложения по значениям параметров. Эти значения далее «пересчитываются» в значения параметров, которыми оперирует другое ЛФР, и он получает соответствующее сообщение.

Если при ответе на сообщение 44) выбран режим электронной почты, то система формирует соответствующее сообщение на базе ответа на сообщение 46) и помещает его в «почтовый ящик» указанного в сообщении 42) ЛФР,а. Как только ЛФР начинает очередной сеанс работы, ему на экран поступает сообщение:

47. РАБОТАЕТ РЕЖИМ ЭЛЕКТРОННОЙ ПОЧТЫ

ПОСТУПИЛИ СЛЕДУЮЩИЕ СООБЩЕНИЯ:

ОТПРАВИТЕЛЬ *

СООБЩЕНИЕ <пароль ЛФР, пославшего сообщение>

<текст сообщения, включающий предложения по значениям согласуемых параметров>

Рассматривая полученные значения согласуемых параметров как исходные данные, ЛФР должно выполнить возложенную на него консультационную операцию с этими данными. Для этого ему надо передать управление элементу $E_{3,8}$ и решить задачу. Помимо этого оно должно, действуя вышеизложенным способом, дать другому ЛФР свои предложения по согласуемым параметрам. Таким образом обмениваясь информацией и оценивая каждый раз, во что выливается движение навстречу друг другу, ЛФР,ы должны в итоге выработать совместно согласованные сформированные рекомендации.

50. ЗАДАЙТЕ СПОСОБ ОПРЕДЕЛЕНИЯ ПРОГНОЗНЫХ ПАРАМЕТРОВ

В позиции между звездочками требуется набрать имя программы, осуществляющей вычисления значений прогнозных параметров в зависимости от результатов решения задач последующих уровней детализации консультируемой проблемы.

Далее следует сообщение

51. ЗАДАЙТЕ ЗАДАНИЕ ДЛЯ МОДУЛЕЙ НИЖНЕГО УРОВНЯ <пароль 1-го подчиненного ЛФР,а>

АННОТАЦИЯ:

ДАННЫЕ ---*---

В ответ на это сообщение ЛФР должно составить аннотацию, а также указать в той же строке, что и слово «ДАННЫЕ», после звездочки номер варианта ОБД, откуда соответствующее ЛФР должно брать исходные данные для решения задачи. В другом случае оно должно на каждой строке ввести имена переменных и их значения, которые будут восприниматься в нижележащем модуле как директивные исходные данные. После того, как задание для первого подчиненного ЛФР,а сформировано, формулируется задание для второго ЛФР,а и т. д.

Если в строке, где помещено слово «ДАННЫЕ», в позициях после звездочки будет определено имя формального алгоритма, то процедура формирования исходных данных для нижележащих консультационных модулей будет производиться на основе этого алгоритма. При этом ЛФР будет получать лишь информационные сообщения:

52. ДЛЯ ЛФР,а <пароль 1-го ЛФР,а>

СФОРМУЛИРОВАНО ЗАДАНИЕ

<1-я переменная> -----* <значение>

.
. .
.

АННОТАЦИЯ:

Ознакомившись с этим сообщением и составив аннотацию, ЛФР, осуществив «пустой ввод», как бы санкционирует выполнение сформированного алгоритмом задания. Если есть необходимость, то ЛФР может скорректировать задание и лишь затем осуществить его пересылку.

Сформулированное задание, попав в «почтовый ящик» соответствующего ЛФР,а, предстает перед ним в виде следующего сообщения

53. ВЫ ДОЛЖНЫ ВЫПОЛНИТЬ ЗАДАНИЕ

<текст аннотации>

ДАННЫЕ НАХОДЯТСЯ В ВАРИАНТЕ <номер варианта>
ОБД

Представленные алгоритмизирующие элементы управляющего модуля САК охватывают в структурированном виде основные действия, требующие выполнения в процессе формирования консультационных модулей.

10.1.4.2. Состав и структура информационного обеспечения УМ

Информационное обеспечение (ИО) представляет собой совокупность данных, используемых в процессе формирования КМ. Основное требование, предъявляемое к ИО, состоит в обеспечении независимости (инвариантности) УМ от конкретного прикладного программного и информационного обеспечения (ППИО). Данная независимость может быть получена путем введения:

1. Фиксированной информационной структуры, содержащей требуемые для работы алгоритмов УМ сведения о ППИО.

2. Предварительного этапа обработки ППИО, заключающегося в преобразовании ППИО во внутреннее представление УМ. В результате выполнения этой процедуры информационная структура заполняется конкретными данными о ППИО. Данный этап далее будем называть этапом настройки.

3. Некоторой канонической формы описания ППИО, содержащей в себе всю информацию, необходимую для работы алгоритмов формирования КМ. Такая канонизация является необходимой для автоматизации проведения этапа настройки УМ на конкретное ППИО.

Рассмотрим состав и структуру ИО. Настройка УМ на конкретное ППИО, сводящаяся к заполнению этой структуры данными о том или ином используемом ППИО, представляет собой самостоятельный режим работы УМ и в данной работе не рассматривается.

Основными функциями ИО являются обеспечение информационной связи УМ с внешними элементами (ЛФР,ми, ППИО, стандартным программным обеспечением КТС, в роли которого далее рассматривается операционная система (ОС), а также между различными консультационными модулями).

Проблема организации связи между ЛФР,ми, ППИО и операционной системой заключается в том, что каждый из названных элементов имеет свой «алфавит» описания данных. Так, модули ППП описываются идентификаторами переменных, а ЛФР,ы оперируют

принятыми в каждой конкретной предметной (проблемной) области (отрасли) наименованиями этих переменных. Причем, так как разработка модулей ППП, как правило, ведется независимо, зачастую в различных организациях, то в результате одна и та же переменная, присутствующая в различных автономно программируемых частях пакета, может описываться различными идентификаторами. В связи с этим на информационное обеспечение возлагаются функции задания взаимно однозначного соответствия между идентификаторами, описывающими одну и ту же переменную модели, а также проблемным (отраслевым) наименованием этой переменной.

Будем обозначать для некоторой переменной $p^j \in P$:

$DS(p^j)$ — ее системный идентификатор;

$DP(p^j)$ — ее проблемное (предметное, отраслевое) наименование;

$\mathcal{L}(p^j)$ — множество модулей ППП, содержащих переменную p^j ;

$Df(p^j, \mathcal{L}_i(p^j))$ — идентификатор, соответствующий переменной p^j в i -м элементе $\mathcal{L}(p^j)$.

Тогда функции информационного обеспечения УМ как связующего звена между ЛФР,ми, элементами ППИО и операционной системой можно представить как установление отношений:

$$p^j \Leftrightarrow DP(p^j) \Leftrightarrow DS(p^j) \Leftrightarrow \{Df(p^j, \mathcal{L}_i(p^j))\}_{i \in \mathcal{L}(p^j)}; \quad \forall p^j \in P.$$

Приведенное отношение определяет проблемные (отраслевые) наименования переменных модели и их системные идентификаторы как разнозначные. В качестве базовых понятий описания ППП предлагается принять проблемные (предметные, отраслевые) наименования переменных модели. Преимущества такого выбора состоят в том, что проблемные (предметные, отраслевые) наименования в любом случае должны являться основой диалога вычислительной системы с ЛФР,ми.

Очевидным недостатком принятия проблемных (предметных, отраслевых) наименований в качестве базовых для описания модели является необходимость их стандартизации. С целью снижения влияния этого недостатка в состав информационного обеспечения может вводиться так называемая таблица понятий, представляющая собой описания некоторых дополнительных структур, которые могут быть образованы из переменных предметной области.

Различаются понятия объединения и эквивалентности. Первое понятие подразумевает объединение некоторого списка переменных, имеющих общий физический смысл, под одним именем (проблемным, предметным или отраслевым наименованием). Оно позволяет обращаться к переменным, входящим в некоторое понятие, как по их

собственным именам, так и через имя понятия. Так, например, понятие «Геометрия крыла» может содержать переменные «Удлинение крыла», «Размах крыла» и т. п. Понятие эквивалентности показывает, что при определенных условиях разные переменные должны восприниматься как идентичные и обрабатываться соответствующим образом с единым значением.

Таким образом, выполнение ИО функций связующего звена между ППИО, операционной системой и ЛФР,ми может быть обеспечено путем включения в ИО таблиц соответствия, совокупность которых называется далее словарем. ***Словарь вместе с таблицей понятий будем называть описанием предметной области ППИО.***

Следующей функцией ИО является обеспечение информационной связи между различными КМ. Учитывая возможную разнесенность во времени формирования и функционирования этих модулей, необходимо обеспечить сохранение результатов каждой из них, а также возможность их использования при формировании и функционировании других модулей. Выполнение указанной функции может быть сведено к введению в состав ИО упорядоченного массива, в котором фиксируются результаты функционирования различных КМ, а ранее занесенные в этот массив данные используются как исходные.

Согласно рассмотренной выше процедуре автоматизированного формирования рекомендаций на базе консультационных модулей необходимо:

1) наличие в ИО совокупности указанных массивов, идентифицируемых именем (паролем) ЛФР,а, а также номером (шифром) варианта рассматриваемой консультационной концепции;

2) обеспечение обмена информацией между различными массивами.

Совокупность таких массивов будем называть оперативной базой данных (ОБД), а каждый из этих массивов — вариантом ОБД.

Как уже отмечалось, для функционирования САК необходимо знание организационной структуры коллектива ЛФР,ов. С целью фиксации этой структуры в состав информационной системы вводится таблица ЛФР,ов. Данная таблица содержит имена пользователей и структуру их организационной подчиненности, представленную в виде матрицы инцидентности. В этой же таблице для каждого ЛФР,а определена как отражение организационного обеспечения та часть ППИО, к которой оно имеет доступ.

Многовариантный ОБД и таблицу ЛФР,ов будем объединять понятием «описание решаемых задач».

Следующий вид информации, хранящейся в ИО, представляет собой данные о ППИО — основном источнике исходных данных для формирования КМ. Сюда входит:

- 1) информация о конкретном перечне пакетов и баз данных, доступных УМ, называемая описанием ППИО;
- 2) информация о «содержимом» каждого конкретного пакета, называемая описанием модулей пакетов.

Описание ППИО включает таблицы моделей (ППП), таблицы модулей ППЧМ и таблицу ПБД. Таблица моделей представляет собой описание сегментов исходной модели консультируемой проблемы. Причем каждый из этих сегментов рассматривается как самостоятельный ППП. Таблица модулей ППЧМ представляет собой описание тематических разделов методов и методик, доступных УМ, например, раздел оптимизации, раздел решения систем нелинейных алгебраических уравнений и т. д.

Таблица баз данных предназначена для организации доступа ЛФР, в ко всем возможностям «подключенных» к УМ информационно-справочных систем и баз данных. Она должна обеспечить выборку, преобразование и перенос данных между оперативной базой данных и некоторой произвольной базой данных. Таблица содержит для каждой из доступных баз данных конкретный формат подготовки информации для ввода-вывода, согласно которому осуществляется преобразование информации при передаче управления соответствующей системе управления базами данных и наоборот.

Представленная структура информационного обеспечения охватывает всю ту информацию, которая необходима для алгоритмов формирования КМ и организации процесса автоматизированного формирования рекомендаций на базе консультационных модулей.

10.1.4.3. Состав и структура программного обеспечения УМ

Ниже рассматривается вариант реализации программного обеспечения (ПО) УМ, ориентированный на функционирование в среде ОС, выступающей в данном случае (совместно с КТС) в роли исполнительного слоя виртуального вычислительного оборудования УМ.

К задачам, решаемым ПО УМ, относятся:

- 1) централизованное (супервизорное) управление работой УМ;
- 2) организация мультизадачной среды совместно с режимом разделения времени для одновременного обслуживания произвольного числа ЛФР, ов;

3) гибкое распределение дефицитных (время центрального процессора, основная память и т. д.) и других ресурсов вычислительного оборудования и операционной системы;

4) динамическое управление состоянием терминалов и сеансами связи;

5) сохранение состояния УМ в контрольной точке для последующего продолжения работы с этой точки через произвольный интервал времени («замораживание» состояния);

6) автоматическое сохранение последовательности состояний УМ с возможностью последующего полного восстановления работоспособности в случае сбоев вычислительных средств или ошибок операционной системы, что обеспечивает высокую степень надежности УМ в процессе работы;

7) управление информационным обеспечением и обеспечение его целостности и защиты от несанкционированного доступа;

8) управление прикладным программно-информационным обеспечением, используемым в процессе формирования и функционирования КМ.

Рассматриваемое ПО имеет иерархическую многоуровневую организацию, где все программные модули (за исключением некоторых модулей нижнего уровня), входящие в его состав, допускают их параллельное использование (являются рентабельными). Это позволяет значительно сократить объем основной памяти, требуемой для работы ПО УМ в мультизадачной среде.

Модули нижнего уровня, входящие в состав ПО, предназначены для выполнения базисных (основных) элементарных операций, обеспечивающих работу УМ в целом. Поток запросов, направляемый к модулям нижнего уровня от модулей более верхних уровней, преобразуется ими в соответствующие запросы к управляющей программе ОС, обеспечивая тем самым доступ к ресурсам ОС и функционирование УМ в ее среде.

Модули синхронизации процессов, входящие в состав нижнего уровня ПО, предназначены для выполнения следующих базисных операций:

1) синхронизации программных событий. Сюда входят две операции: ожидание одного или нескольких программных событий каким-либо процессом (в этом случае данный процесс блокируется); сигнализация о завершении одного или нескольких программных событий (в этом случае соответствующие ожидающие процессы разблокируются);

2) синхронизации доступа к ресурсам. Эти операции требуются для защиты ресурсов в мультизадачной среде, если эти ресурсы изменяются каким-либо процессом, и, следовательно, не должны быть доступны другим процессам в это время. Сюда входят две операции: запрос одного или нескольких ресурсов процессом и его блокировка до тех пор, пока требуемые ресурсы не будут освобождены другими процессами, а также освобождение одного или нескольких ресурсов.

Организация синхронизации процессов модулями ПО без использования соответствующих возможностей управляющей программы ОС позволяет уменьшить затраты основной памяти, так как заблокированные процессы могут быть временно удалены из основной памяти во вспомогательную, а соответствующие освободившиеся участки основной памяти — использованы по другому назначению.

Модули управления процессами предназначены для выполнения следующих базисных операций:

1) создание процесса. Результатом выполнения этой операции является создание самостоятельного процесса, который начинает развиваться (выполняться) параллельно со всеми другими процессами, протекающими в САК. Совокупность всех параллельно развивающихся процессов образует мультизадачную среду системы

2) останов процесса. Любой процесс может быть остановлен (заблокирован) другим процессом, если это необходимо;

3) запуск процесса. Эта операция противоположна операции останова процесса и применяется, если заблокированный процесс должен быть разблокирован;

4) уничтожение процесса. Если процесс уничтожается, то он прекращает свое развитие, а все принадлежащие ему ресурсы освобождаются. Любой процесс может быть уничтожен другим процессом, если это необходимо, а также любой процесс может быть уничтожен самим собой, если он выполнил требуемые функции и больше не нужен.

Модули управления процессами обращаются к управляющей программе ОС и используют возможности управления загрузочными модулями и управления задачами.

Модули управления основной памятью необходимы для выполнения запросов на выделение областей основной памяти. Эти модули выполняют следующие базисные операции:

1) выделение одной или нескольких областей основной памяти требуемого размера из имеющейся свободной области основной памяти и передача их в монопольное владение процессу;

2) освобождение одной или нескольких областей основной памяти, если процесс, ими владеющий, больше не нуждается в них (области основной памяти, принадлежащие процессу, освобождаются автоматически модулями управления процессами, если этот процесс уничтожается);

3) реорганизация основной памяти с целью устранения фрагментации. Эта операция выполняется всякий раз, если запрос на выделение основной памяти не может быть удовлетворен, но свободная память имеется в требуемом объеме.

Модули управления внешней памятью позволяют процессам осуществлять доступ к информационному обеспечению, так как все это обеспечение хранится во внешней памяти. Модули управления внешней памятью используют страничную организацию. Это позволяет, во-первых, уменьшить количество фактических обращений к внешней памяти (количество операций ввода-вывода), а во-вторых, обрабатывать данные сколь угодно сложной структуры, используя технологию обработки списков. К операциям управления внешней памятью относятся:

1) чтение блока данных из внешней памяти в основную по указанному адресу внешней памяти;

2) запись блока данных из основной памяти во внешнюю по указанному адресу внешней памяти;

3) выделение области внешней памяти требуемого размера для нового блока (область внешней памяти не передается в монопольное владение процессу, ее запросившему, а доступна для всех процессов);

4) освобождение области внешней памяти;

5) реорганизации внешней памяти с целью устранения фрагментации.

Логика выполнения операций выделения, освобождения и реорганизации внешней памяти аналогична логике этих же операций для основной памяти;

6) управление режимами обмена для увеличения эффективности доступа к внешней памяти (выбор алгоритма вытеснения страниц, управление режимом «обязательного обновления страниц», накопление статистики и т. д.).

Для выполнения фактических операций ввода-вывода модули управления внешней памятью используют базисный прямой метод доступа (BDAM) ОС. С точки зрения ОС, внешняя память представляется в виде наборов данных, расположенных на томах прямого доступа.

Модули управления терминалами позволяют выполнять операции динамического управления терминалами системы и операции по передаче данных. К этим операциям относятся следующие:

1) динамическое подключение физического монитора (конкретного устройства ввода-вывода). Подключенный монитор становится недоступным для других заданий ОС и может быть использован только для ведения сеанса связи с УМ и сформированными КМ;

2) динамическое отклонение физического терминала. В этом случае конкретный терминал становится недоступным УМ, но может использоваться другими заданиями ОС. Подобное динамическое управление терминалами позволяет другим заданиям ОС работать с теми терминалами, которые в данный момент не обслуживаются УМ, и тем самым увеличить эффективность их использования, так как время решения некоторых задач системой, когда наличие терминалов не требуется, может быть довольно продолжительным;

3) управление очередями сообщений, передаваемых от терминалов (запросы пользователей) к терминалам (ответы пользователям на их запросы);

4) управление форматированием сообщений. Использование средств форматирования сообщений позволяет улучшить «читабельность» информации, выводимой на терминал, а также представляет пользователю возможности автономного редактирования информации, вводимой с терминала.

Для выполнения фактических операций ввода-вывода модули управления терминалами используют существующие методы доступа ОС.

Основным назначением модулей управления состоянием системы является обеспечение надежности ее работы. Соответствующими операциями являются:

1) останов системы. В этом случае текущее состояние системы сохраняется во внешней памяти («замораживается»), все сеансы связи прекращаются и система заканчивает свою работу;

2) рестарт системы. Текущее состояние системы восстанавливается из внешней памяти (это состояние фиксируется во время выполнения операции останова системы) и она продолжает свою работу. Рестарт системы может быть выполнен через любой промежуток времени после останова, при этом система сохраняет полную работоспособность;

3) автоматическое сохранение последовательности состояний.

Так как состояние системы изменяется в процессе формирования и функционирования КМ, то сохранение текущей информации позволяет

гарантировать полное восстановление состояний системы после сбоя вычислительных средств или ошибок операционной системы. Операции, связанные с сохранением последовательности состояний, выполняются всякий раз, когда поступает запрос извне.

Модули управления состоянием систем обращаются к управляющей программе ОС при выполнении операций останова и рестарта системы и используют возможность контрольной точки рестарта. Для автоматического сохранения последовательности состояний используется базисный последовательный метод доступа (BSAM) ОС, а соответствующая информация записывается в последовательный набор данных.

Модуль инициализации системы выполняет единственную операцию, которая приводит систему в состояние готовности. Во время функционирования системы этот модуль не используется. После приведения системы в состояние готовности (загрузка ПО в основную память и т. д.) модуль инициализации порождает единственный процесс и передает управление диспетчеру. Процесс, порожденный модулем инициализации, начинает свое развитие и порождает совокупность других процессов, которые обеспечивают требуемую логику функционирования.

Диспетчер также входит в нижний уровень ПО. Он предназначен для распределения такого ресурса, как время центрального процессора между процессами, протекающими в системе. Поскольку рассматриваемое ПО предназначено для работы в однопроцессорном вычислительном средстве, то мультизадачная среда может быть реализована только за счет поперечного диспетчирования всех готовых к выполнению (т. е. незаблокированных в текущий момент времени) процессов.

Диспетчер получает управление всякий раз, когда выполняется какая-либо операция, связанная с управлением процессами или синхронизацией процессов, или если истекает квант времени, отведенный процессу, фактически выполняющемуся в данный момент времени. В любом случае диспетчер выделяет новый квант времени следующему готовому процессу, обеспечивая тем самым параллельное развитие всех готовых процессов, а следовательно, поддерживает мультизадачную среду. Если нет одного готового процесса, диспетчер переводит систему в состояние ожидания до тех пор, пока не произойдет внешнее программное событие, которого ожидает какой-либо процесс.

Следующая группа программных модулей образует средний уровень ПО. Эти модули также предназначены для выполнения

основных (базисных) операций, обеспечивающих формирование и функционирование КМ, но являющихся более сложными по сравнению с операциями нижнего уровня. Они не включаются в нижний уровень по следующим двум причинам:

1) модули среднего уровня обращаются для выполнения требуемых операций к модулям нижнего уровня, поэтому включение их в нижний уровень нарушило бы иерархичность уровней ПО — условие, обеспечивающее расширяемость ПО;

2) параллельное выполнение модулей нижнего уровня не допускается, а модули среднего уровня могут выполняться параллельно.

Поток запросов к модулям среднего уровня преобразуется ими в запросы к модулям нижнего уровня, однако некоторые запросы могут быть направлены к модулям нижнего уровня непосредственно.

Модули управления ППИО предназначены для выполнения следующих операций:

1) подготовка и модификация проблемного программного обеспечения. Сюда входят обработка запросов на редактирование исходных текстов программ и обработка запросов на динамический вызов обрабатывающих программ ОС (таких, как компиляторы, редакторы связей и т. д.). В последнем случае выполнение обрабатывающей программы в рамках УМ происходит как развитие соответствующего процесса;

2) подготовка и модификация проблемного информационного обеспечения. Сюда входят обработка запросов на редактирование текстов паспортов программ, запросов на выполнение автономных подготовительных операций с оперативными банками данных, запросов на заполнение словарей и т. д.;

3) доступ к ППИО. Выполнение операции доступа к проблемному ПО предполагает создание списка входных и выходных параметров для конкретного прикладного модуля в соответствии с информацией, выбираемой из его паспорта, а также просчет этого прикладного модуля. При этом просчет модуля реализуется как развитие соответствующего процесса. Операции доступа к прикладному ИО включают в себя в основном действия со списками (поиск в линейных списках, модификация линейных списков, поиск в сетевых списках и т. д.).

Необходимо также заметить, что ППИО не входит непосредственно в рассматриваемое ПО, но модули управления ППИО в совокупности с другими модулями среднего уровня позволяют адаптировать УМ к решению задач в конкретной проблемной области, определяемой соответствующим ППИО.

Модули сопряжения разнородного ППИО, входящие в средний уровень ПО, позволяют выполнять преобразования представлений прикладного программного обеспечения. Необходимость подобных преобразований обусловлена тем, что прикладное ПО может быть представлено в виде программных модулей, написанных на различных алгоритмических языках, а также на различных диалектах какого-либо алгоритмического языка. Использование преобразования представлений данных минимизирует количество изменений, вносимых в прикладное ПО.

Модули управления сеансами связи выполняют следующие операции:

1) модификация существующих «сценариев» диалога с пользователем, а также создание новых «сценариев». Введение «сценариев» как алгоритмически управляемой последовательности кадров для интерактивного взаимодействия с пользователем позволяет, во-первых, упростить интерфейс между терминалом и процессом, ведущим диалог, а во-вторых, предусматривает возможность гибкой модификации развития диалога с учетом требований конкретного ЛФР;

2) автоматическое заполнение кадров информацией, представляемой соответствующими процессами, и вывод очередного кадра на терминал (при движении сообщений к пользователю);

3) ввод кадра с терминала, выделение информации из полей этого кадра и передача ее обрабатывающему процессу (при движении сообщений от пользователя);

4) управление архивом «сценариев». Эта операция связана с хранением функционально взаимосвязанных кадров во внешней памяти, а также с извлечением кадров, требуемых во время сеанса.

Использование кадров позволяет также сделать модули верхних уровней независимыми от типа конкретного терминала, на котором ведется сеанс связи в УМ.

Следующую группу модулей ПО составляют модули верхнего уровня. Они предназначены для реализации инвариантных к проблемной области процедур формирования и функционирования КМ.

Модули верхнего уровня соответствуют представленному выше алгоритмизирующему слою УМ. Основными из них являются:

1) модули, реализующие оператор KON и его составляющие операторы (KL, KJ, KST), структура которых представлена в 10.1.2;

2) модули, реализующие оператор KVS и его составляющие операторы (KZN, KLIP, KPR, RKON), структура которых представлена в 10.1.3;

- 3) модули, формирующие конкретную вычислительную схему реализации расчетных моделей [оператор KFOR (см. п. 10.1.2.5)];
- 4) модули, обеспечивающие постановку и решение задач оптимизации и выбора рациональных рекомендаций.

10.2. Функционирование консультационных модулей

10.2.1. Содержание процедур функционирования консультационных модулей

Процесс автоматизированного формирования рекомендаций на базе САК, рассматриваемой как организационно-технической системой (ОТС), может быть описан как формирование и функционирование образовавшихся при этом ОТС меньшего «масштаба» — консультационных модулей (КМ). Основным назначением каждого КМ является выполнение соответствующей консультационной операции. Содержание задач, возлагаемых при этом на КМ, непосредственно вытекает из сути формирования рекомендаций.

Хотя под консультированием понимается весь процесс создания прообраза, прототипа предполагаемых рекомендаций, мы будем использовать этот термин в более узком понимании — как выбор консультационных параметров для разрабатываемых рекомендаций. С этих позиций цель формирования рекомендаций состоит в том, чтобы сузить до конкретных значений исходное множество возможных изменений консультационных параметров. Получаемые в результате формирования рекомендаций значения консультационных параметров должны удовлетворять двум требованиям: непротиворечивости и обоснованности.

Непротиворечивость означает, что сочетание значений консультационных параметров не противоречит объектным связям, существующим между ними, а также между ними и так называемыми внешними параметрами. Эти связи задаются математической моделью, описывающей класс консультируемой проблемы, к которому относится консультируемая проблема по отношению к которой формируются рекомендации.

Требование **обоснованности** преемственно по отношению к требованию оптимальности формируемых рекомендаций, но шире и реалистичнее его. Если консультируемая проблема допускает исчерпывающую оценку ее качества по единственному количественному показателю (критерию), причем известна точная

зависимость критерия от консультационных параметров и в отведенный для формирования рекомендаций срок может быть решена математическая задача оптимизации критерия по консультационным параметрам, то обоснованность сформированных рекомендаций сводится к их оптимальности. Если же хотя бы одно из перечисленных условий не имеет места, любой результат сформированных рекомендаций, даже хорошо обоснованный, не является оптимальным в строгом смысле этого слова. Рассмотрим сказанное подробнее.

В целом каждая задача формирования рекомендаций может быть охарактеризована:

-перечнем параметров двух групп. В первую группу входят **параметры**, называемые **консультационными**, для которых изначально известны диапазоны их рассматриваемых (возможных) изменений, а в результате функционирования процесса формирования рекомендаций требуется определить их конкретные значения. Во вторую группу входят **директивные и нормативные данные (внешние параметры)**, значения которых заданы либо однозначно, либо множеством возможных значений;

-математической моделью, содержащей связи между отдельными группами как консультационных, так и внешних параметров. Удовлетворение этим связям рассматривается как непротиворечивость значений консультационных параметров как друг другу, так и внешним параметрам;

-правилами обоснования или выбора наиболее обоснованных значений консультационных параметров из множества их рассматриваемых значений.

Одним из таких правил, применительно к формированию рекомендации для сложных консультируемых проблем, является реализация процесса выбора значений всего множества консультационных параметров по частям. При этом процедура выбора значений параметров каждой отдельной группы определяется как консультационная операция.

Каждая консультационная операция состоит в выборе наиболее обоснованных значений консультационных параметров, характеризующих отдельные элементы и подпроблемы консультационной проблемы или отдельные аспекты ее функционирования. В этом плане консультационные операции являются аналогичными процессу формирования рекомендаций в целом и отличаются от него лишь своими «масштабами». Тогда, учитывая сказанное ранее, **каждая консультационная операция**

может быть охарактеризована перечнем параметров, моделью и правилом выбора.

Функционирование КМ как ОТС, выполняющей отдельную консультационную операцию, состоит в том, что ЛФР оперирует значениями отдельных параметров, наблюдая за значениями других параметров, получаемыми с помощью математической модели, определяет на основе используемых правил выбора стратегию своих дальнейших действий.

*Параметры, участвующие в процессе выполнения консультационной операции, целесообразно разделить на **четыре класса.***

В *первый класс* определенных исходных данных попадают **внешние параметры**, значения которых с *необходимой консультационной точностью являются однозначно заданными*. Сюда относятся физические константы, характеристики применяемых материалов, однозначные требования КЗ, характеристики окончательно утвержденных к текущему моменту сформированных рекомендаций, параметры используемых законов распределения случайных величин и т. п.

Во *второй класс* неопределенных исходных данных входят **внешние параметры**, в отношении которых ЛФР,у может быть *указан лишь диапазон их возможных значений*, причем ЛФР лишено возможности непосредственно повлиять на то, какие значения этих параметров будут реализовываться фактически, в процессе функционирования консультируемой проблемы. К ним относятся имеющие значительный разброс условия функционирования консультируемой проблемы, интервальные требования КЗ, недостаточно определенные в силу объективных причин характеристики перспективных материалов, параллельно разрабатываемых комплектов изделий и систем, с которыми предстоит взаимодействовать консультируемой проблеме и другие параметры, характеризующие уровень потребности в формируемых рекомендациях, консультационные параметры других параллельно протекающих консультационных операций, по которым еще не принято окончательное решение, поправочные коэффициенты, характеризующие неточность используемых математических моделей и др.

К *третьему классу выбираемых консультационных параметров относятся те характеристики консультируемой проблемы*, назначение конкретных значений которых составляет конечную цель данной консультационной операции, а также связанные с

ними характеристики консультируемой проблемы, значения которых должны лежать в заданных диапазонах. На начальной стадии автономного функционирования консультационного модуля ЛФР может задать лишь границы диапазонов рассматриваемых им значений этих параметров.

Четвертый класс показателей эффективности, или критериев, составляют параметры, которые в совокупности оценивают качество сформированных рекомендаций, разрабатываемых в рамках данной консультационной операции, причем повышению качества соответствует монотонное изменение значений этих параметров (увеличение или уменьшение).

Пользуясь математическими терминами, можно сказать, что ***консультационная операция представляет собой задачу формирования рекомендаций, описываемую с помощью следующих компонентов:***

- ***множества неопределенностей X*** , представляющего собой совокупность возможных значений параметров второго класса;
- ***множества допустимых рекомендаций Y*** , представляющего собой совокупность допустимых значений параметров третьего класса;
- ***векторной функции эффективности $f(x, y)$*** , представляющей собой зависимость значений параметров четвертого класса от значений параметров первых трех классов [в обозначении функции эффективности зависимость от параметров первого класса явно не указана ввиду константного характера этих параметров], (эта зависимость реализуется через математическую модель консультируемой проблемы).

Такая задача является задачей формирования рекомендаций в условиях неопределенности.

Если ***множество неопределенностей пусто*** (неопределенных исходных данных нет), а ***функция эффективности скалярна*** (используется единственный критерий эффективности), эта задача переходит в более простую и хорошо изученную ***задачу оптимизации,*** которая описывается ***множеством допустимых рекомендаций Y и функцией эффективности $f(y)$.***

Задача оптимизации обладает очень привлекательным свойством: она формально замкнута, т. е. требование найти решение (сформировать рекомендацию) $y \in Y$, доставляющее наименьшее (или наибольшее, в зависимости от смысла функции эффективности) значение функции эффективности $f(y)$ на множестве допустимых решений (рекомендаций) Y :

$$\bar{y} = \underset{y \in Y}{\text{arc min}} f(y).$$

вполне достаточно для того, чтобы получить однозначный результат без всякого участия ЛФР, чисто математическим путем, что свидетельствует о формальной замкнутости задач оптимизации (рис. 10.33).

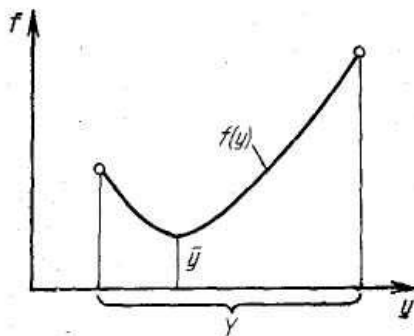


Рис. 10.33. Формальная замкнутость задач

Это свойство иногда побуждает стремиться построить консультационные операции в САК на основе одних только задач оптимизации. Такое стремление следует признать ошибочным. Ведь чтобы реализовать «чисто оптимизационный» консультационный процесс, нужно исключить из рассмотрения параметры второго класса, а это значит (см. описание параметров этого класса) отказаться от формирования рекомендаций для многоцелевых консультируемых проблем, рассчитанных на различные задачи и условия применения, абсолютизировать неточные исходные предпосылки, исключить неточности в используемых математических моделях, а также взаимосвязь консультационных операций в консультационном процессе, заменив ее директивной пересылкой сформированных рекомендаций. Кроме этого, чтобы реализовать «чисто оптимизационный» консультационный процесс, необходимо свести учет всего многообразия технико-экономических требований и аспектов эффективности консультируемой проблемы к единственному числовому показателю.

Сказанное, конечно, не отрицает целесообразности использования оптимизационных задач в отдельных (но не во всех без исключения) консультационных операциях, а также важности разработки

математических методов оптимизации, которые составляют фундамент задач формирования рекомендаций в условиях неопределенности.

В чем же принципиальное отличие, с точки зрения ЛФР, задач формирования рекомендаций в условиях неопределенности от задач оптимизации? В том, что они значительно шире в содержательно-постановочном плане, но в то же время формально не замкнуты, т. е. их описание в виде триады $(X, Y, f(x, y))$ не позволяет получить однозначный результат без дальнейшего участия ЛФР. Это свойство задач формирования рекомендаций в условиях неопределенности иллюстрируется на рис. 10.34—10.36.

На рис. 10.34 показан случай, когда функция эффективности скалярна, но имеется множество неопределенностей, состоящее из трех элементов: x_1, x_2, x_3 . В этом случае для каждого из этих элементов можно найти свою оптимальную рекомендацию $\bar{y}_1, \bar{y}_2, \bar{y}_3$:

$$\bar{y}_x = \operatorname{arc} \min_{y \in Y} f(x, y), \quad x \in X = \{x_i\}_{i=1,2,3}.$$

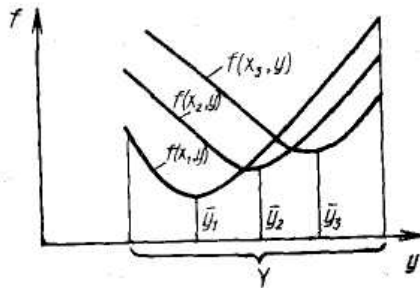


Рис. 10.34. Формальная незамкнутость задачи формирования рекомендаций в условиях неопределенности, вызванная наличием множества неопределенностей $X = \{x_1, x_2, x_3\}$.

Какую же рекомендацию принять в качестве окончательной (одну из них или совсем иную) с учетом того, что, по смыслу множества неопределенностей, на практике может реализовываться любая, заранее неизвестная из x_1, x_2, x_3 , а то и все вместе? Ответить на этот вопрос формальным путем, без участия ЛФР, невозможно.

На рис. 10.35 показан случай, когда в задаче формирования рекомендаций отсутствует множество неопределенностей, но функция эффективности содержит два показателя эффективности: $f^1(y), f^2(y)$.

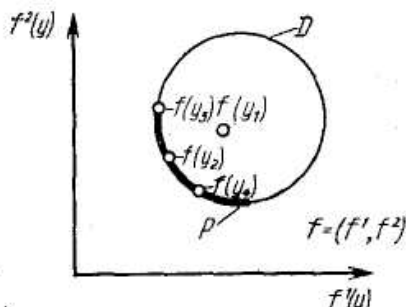


Рис. 10.35. Формальная незамкнутость задачи формирования рекомендаций в условиях неопределенности, вызванная векторным характером функции эффективности.

Область D состоит из точек, соответствующих значениям векторной функции эффективности $f(y) = (f^1(y), f^2(y))$, отвечающим различным допустимым рекомендациям $y \in Y$. Ясно, что рекомендации, которым соответствуют точки, не лежащие на линии P , нецелесообразны, так как для каждой из них можно указать рекомендацию, которая по обоим показателям эффективности лучше нее [например, рекомендация y_2 лучше y_1 , так как $f^1(y_2) \leq f^1(y_1)$, $f^2(y_2) \leq f^2(y_1)$]. Для таких рекомендаций сравнение и выбор выполняются формально, без участия ЛФР. Таким формальным путем может быть выделено некоторое множество «наилучших» рекомендаций, которому отвечает линия P (множество Парето). Однако однозначно указать «наилучшую» из рекомендаций, соответствующих точкам P , формальным путем невозможно, так как в каждой паре соответствующих P рекомендаций по одному из показателей эффективности лучшей является одна из них, а если рассматривать по другому показателю, то другая.

Общий случай наличия в задаче формирования рекомендаций, как множества неопределенностей, так и векторной функции эффективности, показан на рис. 10.36, который иллюстрирует формальную незамкнутость задачи формирования рекомендаций в условиях неопределенности в общем случае.

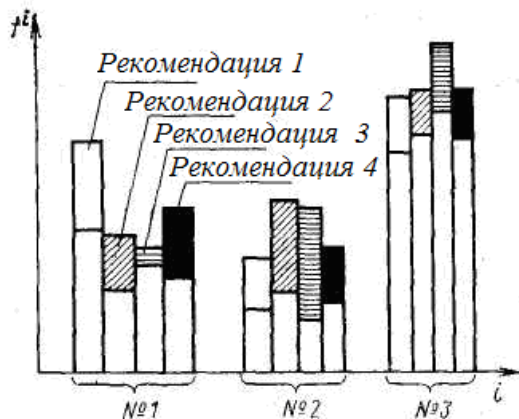


Рис. 10.36. Формальная незамкнутость задачи формирования рекомендаций в условиях неопределенности.

На этой диаграмме по оси ординат откладываются значения различных показателей эффективности (№1—№3) для четырех вариантов рекомендаций (1—4). Каждому варианту рекомендаций соответствует диапазон значений показателя эффективности, возникший из-за наличия множества неопределенностей X :

$$\min_{x \in X} f^i(x, y_j) \leq f^i(y_j) \leq \max_{x \in X} f^i(x, y_j); \quad i = 1, 2, 3; \quad j = 1, 2, 3, 4.$$

При этом видно, что формально установить, какая из рекомендаций «лучше», невозможно: для окончательного однозначного формирования рекомендаций необходимо привлечение ЛФР.

Возникает вопрос: возможно ли формально замкнуть задачу формирования рекомендаций в условиях неопределенности и тем самым сделать ее результат не зависящим от субъективных особенностей ЛФР, а его получение — процедурой, целиком возлагаемой «на плечи» ЭВМ. Теория консалтинга дает на этот вопрос отрицательный ответ. В соответствии с этим понятие «наилучшей» рекомендации лишается в условиях неопределенности сугубо объективного содержания: не только практически, но и теоретически невозможно доказать (за редким исключением), что сформированная рекомендация лучше некоторой другой. Хорошо известные дискуссии между сторонниками различных вариантов консультационных рекомендаций, которые разрешаются в практике консультационных организаций на основе чисто человеческих качеств участников

дискуссии или «волевым» вмешательством начальства, являются отражением этого факта. Поэтому принцип поиска «оптимальной» рекомендации должен быть заменен при функционировании консультационных модулей принципом поиска «наиболее обоснованной» рекомендации.

Обоснованность означает уверенность ЛФР в том, что в рамках внешних условий, наложенных на организацию формирования рекомендаций (специфика консультационной проблемы, лимит времени, ограниченность ресурсов и др.), им использована одна из наиболее эффективных для этих условий методология формирования рекомендаций.

В рамках принципа обоснованности задача ЛФР состоит в квалифицированном выборе одной из эффективных для данных условий формирования рекомендаций методик формирования рекомендаций, овладении ее понятийным аппаратом и непосредственном выполнении функций, возлагаемых ею в процессе формирования рекомендаций на человека.

Эффективность методики означает, что не может быть указана методика, которая в данных условиях формирования рекомендаций привела бы к заведомо лучшему результату. С этой точки зрения в определенных условиях даже методика, состоящая в формировании рекомендаций наугад, является эффективной.

Резюмируя, можно сказать, что функционирование консультационных модулей состоит в выборе наиболее обоснованных сформированных рекомендаций, при котором:

1) в частном случае, когда задача формирования рекомендаций является задачей оптимизации, методика приводит к получению оптимальных рекомендаций;

2) в частном случае, когда задача формирования рекомендаций является задачей многокритериальной оптимизации, методика приводит к получению Парето-оптимальной рекомендации (см. рис. 10.35, дуга *P*);

3) методика позволяет лицу, формирующему рекомендацию активно влиять на выбор рекомендации в соответствии с изменением его информированности и целевой ориентированности, а также обеспечивает однозначное формирование рекомендации при любом уровне неопределенности.

Первые два требования совершенно очевидны, что же касается третьего то его содержание, отражающее основное отличие принципа обоснованности от принципа оптимальности, может показаться сомнительным. В самом деле, не является ли «активное влияние»

синонимом субъективизма при формировании рекомендаций, не должна ли методика сигнализировать о невозможности уверенного формирования рекомендаций при имеющейся информации вместо того, чтобы даже при недостаточной информативности формировать однозначную рекомендацию.

Это не соответствует действительности, так как в условиях неопределенности просто отсутствует возможность альтернативного поведения, не прерывающего консультационный процесс: если бы была возможность собрать дополнительную информацию или даже превратить решаемую задачу в полностью формализованную задачу оптимизации, это было бы сделано; если же это невозможно и тем не менее требуется выработать консультационную рекомендацию, это можно сделать лишь при выполнении третьего требования.

Может возникнуть еще один вопрос: раз субъективизм и недостаточная информативность налицо, не все ли равно: сформировать рекомендации по научной методике или просто наугад (альтернатива именно такова, так как осмысленный выбор с позиций «здорового смысла» есть уже формирование по некоторой методике). В том, что это не так убеждает простейший пример, показанный в табл. 10.2.

Таблица 10.2

Номер варианта	Номер критерия		
	1	2	3
1	0,300	0,500	0,400
2	0,295	0,800	0,450
3	0,700	0,495	0,700
4	0,900	0,600	0,395

В нем рассмотрены четыре варианта рекомендаций, которые оценены по трем критериям, и необходимо выбрать «наилучший» из них. Эта задача является задачей выбора рекомендаций в условиях неопределенности и формальных оснований предпочесть одну из них, нет. При выборе рекомендации наугад первый вариант будет выбран лишь с вероятностью 0,25. Между тем, если рассматривать данную задачу как консультационную, в которой все измерения и расчеты производятся с ограниченной точностью, то можно заметить, что при уменьшении значений критериев для первого варианта на 0,006 т. е. на 1—2 %, он становится строго лучше, чем любые другие варианты, и любая разумная методика выбора однозначно укажет его в качестве наилучшего.

Из сказанного видно, что использование принципа обоснованности при выполнении консультационных операций в САК требует выработки у ЛФР,а своеобразного подхода к формированию рекомендаций в условиях неопределенности, овладения специфическим понятийным аппаратом. Этот аппарат, различный для различных методик, отражает конкретные формы участия ЛФР,а в формировании рекомендаций, уровень влияния неопределенности на сформированные рекомендации. В настоящее время можно предложить ряд методик формирования рекомендаций в условиях неопределенности, облегчающих ЛФР,у формировать рекомендации в условиях неопределенности и повышающих его обоснованность. В данной работе излагается одна из таких методик (метод ПРИНН), которая, по мнению автора, наиболее приспособлена для использования ее в качестве основы для организации функционирования консультационных модулей. Эта методика, подробно изложенная в разделе 10, требует от ЛФР овладения несколькими основными понятиями, используя которые он сможет оперативно формировать рекомендации при всестороннем учете влияния неопределенности, варьируя степень своего участия в формировании рекомендаций по собственному усмотрению, начиная от минимальной, эквивалентной решению задач оптимизации.

10.2.2. Основные понятия метода учета неопределенности при формировании рекомендаций (метод ПРИНН)

Исследуем функционирование консультационного модуля, состоящее из генерирования множества альтернативных вариантов формируемых рекомендаций и выбора наилучшей из них в условиях неопределенности. Не затрагивая проблемы генерирования вариантов, решаемой на практике на основе интуиции и опыта ЛФР,ов с использованием различных методов прямого и направленного перебора, уделим основное внимание выбору наилучшего варианта сформированной рекомендации. Этот выбор в системах автоматизированного консультирования уже не может основываться только на здравом смысле ЛФР: слишком велико число вариантов, которые могут быть сгенерированы в автоматизированном режиме, и слишком многоаспектны и противоречивы требования к их эффективности и условия их оценки.

Из рассмотренного в п. 10.2.1 принципа обоснованности формируемых рекомендаций следует необходимость выработки у ЛФР своеобразного подхода к формированию рекомендаций в условиях

неопределенности, овладения им специфическим понятийным аппаратом. Последний неодинаков для различных методик, отражает конкретные формы участия ЛФР,а в формировании рекомендаций и уровень влияния неопределенности на сформированную рекомендацию. К настоящему времени известен ряд методик, облегчающих ЛФР,у формирование рекомендаций в условиях неопределенности и повышающих его обоснованность. Ниже будет изложена одна из таких методик (метод ПРИНН), которая, по мнению автора, наиболее приспособлена для использования в САК. Эта методика требует от ЛФР,а овладения несколькими основными понятиями, используя которые он сможет оперативно формировать рекомендации при всестороннем учете влияния неопределенности, варьируя степень своего участия в формировании рекомендаций по собственному усмотрению, начиная от минимальной, эквивалентной решению задач оптимизации. Рассмотрим эти понятия, характерные для предлагаемого метода.

Первым из таких понятий является понятие *обобщенных потерь*.

Потерями будем называть показатели эффективности, значения которых желательно минимизировать, например, затраты, расход топлива, масса конструкции.

Если показатель эффективности желательно максимизировать, его легко преобразовать в виду потерь, заменив, например, на величину, дополняющую его значение до максимально возможного. Так, вместо целевой отдачи какой-либо консультируемой проблемы можно рассматривать величину недополученной целевой отдачи по сравнению с заведомо рекордным значением. Таким образом, ЛФР может все показатели эффективности сформулировать в виде потерь.

Существует строго математическое доказательство, что задачу формирования рекомендаций с множеством неопределенностей X_0 и векторной функцией эффективности можно различными способами преобразовать в эквивалентную задачу формирования рекомендаций с новым, расширенным множеством неопределенностей X и уже скалярной функцией эффективности f . Один из способов такого преобразования, называемый *линейной сверткой*, состоит в том, что в задачу формирования рекомендаций вводятся дополнительные переменные — *неопределенные коэффициенты сравнительной значимости* различных показателей эффективности α_i , $i = 1, \dots, n$, где n — число показателей эффективности. Их значения неизвестны и удовлетворяют лишь условиям неотрицательности и нормировки:

$$\alpha_i \geq 0, \quad i = 1, \dots, n, \quad (10.124)$$

$$\sum_{i=1}^n \alpha_i = 1. \quad (10.125)$$

Эти условия добавляются к условиям, описывающим множество X_0 , составляя вместе с ними описание множества X . Функция эффективности при этом задается в виде

$$f(x, y) = \sum_{i=1}^n \alpha_i f^i(x, y). \quad (10.126)$$

В данном соотношении вектор x содержит и переменные α_i ($i = 1, 2, \dots, n$) как параметры второго класса (см. п. 10.2.1) в новой задаче формирования рекомендаций.

Функция (10.126) в обобщенном виде отражает все исходные показатели эффективности (потери). Так, если какой-либо из коэффициентов α_i равен 1, то эта функция совпадает с соответствующим показателем эффективности. Поэтому естественно называть функцию (10.126) обобщенными потерями. На практике удобно измерять исходные показатели эффективности в долях их максимально возможного значения (что мы и будем подразумевать в дальнейшем). При этом

$$0 \leq f^i(x, y) \leq 1, \quad i = 1, 2, \dots, n$$

и

$$0 \leq f(x, y) \leq 1,$$

т. е. обобщенные потери измеряются в той же шкале, что и исходные потери. Кроме того, если $f^i = C$, $i = 1, \dots, n$, то и $f = C$.

Если для двух вариантов рекомендаций y_1, y_2

$$\forall x \in X, \quad f^i(x, y_1) \geq f^i(x, y_2), \quad i = 1, \dots, n,$$

то

$$f(x, y_1) \geq f(x, y_2).$$

Это **свойство** можно назвать **монотонностью обобщенных потерь**. Оно показывает, что рекомендация, лучшая по каждому из показателей эффективности, будет лучшей и по обобщенным потерям. Вообще, эквивалентность задачи формирования рекомендаций с обобщенными потерями исходной задаче означает, что любое парето-оптимальное (т. е. соответствующее некоторой точке на линии P , рис. 10.35) решение исходной задачи может быть получено как оптимальная рекомендация по обобщенным потерям и наоборот.

Графическое пояснение этого факта дано на рис. 10.37 для задачи формирования рекомендаций, показанной на рис. 10.35.

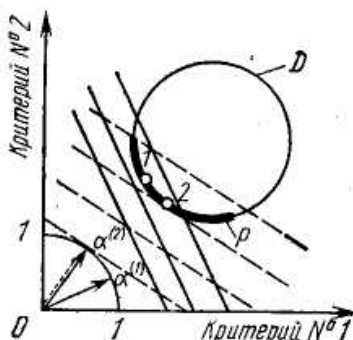


Рис. 10.37. Эквивалентность задач формирования рекомендаций при переходе к обобщенным потерям по линейной свертке

Семейство линий уровня обобщенных потерь как функции исходных потерь по соотношению (10.126) есть семейство прямых, нормальных к вектору α . На рис. 10.37 показаны два таких семейства, отвечающих векторам $\alpha^{(1)}$ и $\alpha^{(2)}$. Рекомендация, оптимальная по критерию обобщенных потерь, должна, очевидно, соответствовать точке области D , через которую проходит линия уровня обобщенных потерь с минимальной отметкой. Видно, что таковыми являются точки 1 и 2 для векторов $\alpha^{(1)}$ и $\alpha^{(2)}$ соответственно. Понятно, что для любой точки линии P легко указать значение вектора α , при котором она окажется оптимальной по соответствующим обобщенным потерям и наоборот, любой вектор α , удовлетворяющий условию (10.125), выделит в качестве оптимальной лишь одну из точек линии P .

Эквивалентность перехода к задаче формирования рекомендаций с обобщенными потерями может быть неполной, если, например, множество D невыпукло. Так, точка 3 на рис. 7.38, принадлежащая линии P , не является оптимальной по обобщенным потерям ни при каких значениях вектора α (точки 1 и 2 более предпочтительны, чем точка 3).

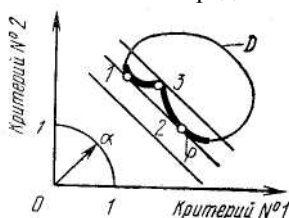


Рис. 10.38. Неполная эквивалентность задач формирования рекомендаций при переходе к обобщенным потерям по линейной свертке при невыпуклости множества D

Следует специально отметить, что описанный переход к обобщенным потерям не приводит к потере информации об исходных показателях эффективности, к замене их значений одним числом. Такое может случиться лишь при неверной трактовке обобщенных потерь, когда коэффициенты α_i рассматривают не как неопределенные переменные, относительно которых известны лишь условия (10.125), а как конкретные числа, значения которых пытаются задать.

Несколько иной метод перехода к обобщенным потерям, также обладающим всеми отмеченными выше свойствами, положен в основу «свертки Гермейера», при которой

$$\begin{aligned} \alpha_i &\geq 0 \\ \max_{i=1, \dots, n} \alpha_i &= 1 \end{aligned} \quad (10.127)$$

и

$$f(x, y) = \max_{i=1, \dots, n} \alpha_i f^i(x, y), \quad (10.128).$$

Если в случае линейной свертки обобщенные потери можно трактовать как средневзвешенное значение потерь, то при свертке Гермейера — *как предельное значение взвешенных потерь*. Такая трактовка менее привычна, зато при свертке Гермейера эквивалентность задач формирования рекомендаций с вектором потерь и с обобщенными потерями имеет место всегда, лишь бы потери были неотрицательны. Такая эквивалентность иллюстрируется с помощью рис. 10.39, где линии уровня обобщенных потерь имеют вид семейства «углов».

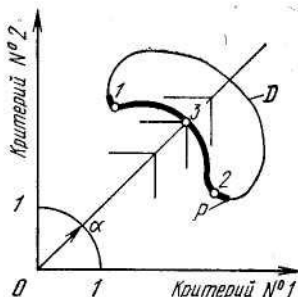


Рис. 10.39. Эквивалентность задач формирования рекомендаций при переходе к обобщенным потерям по свертке Гермейера

Итак, задачу формирования рекомендаций в условиях неопределенности можно без ограничения общности описывать триадой $(X, Y, f(x, y))$, где X — множество неопределенностей (включая

неопределенные коэффициенты сравнительной значимости показателей эффективности); Y — множество допустимых рекомендаций; $f(x, y)$ — функция обобщенных потерь (скалярная). В дальнейшем будем подразумевать именно такую ее постановку, формальная незамкнутость которой иллюстрируется рис. 10.34.

Замыкание задачи формирования рекомендаций в условиях неопределенности происходит за счет привлечения ЛФР, а для выбора окончательной однозначной рекомендации. Основными понятиями метода ПРИНН являются *способ учета неопределенности* и *n-обобщенные потери*. Поясним их содержание.

В процессе формирования рекомендаций в условиях неопределенности на базе метода ПРИНН ЛФР, y предоставляется информация об эффективности рассматриваемых вариантов рекомендаций при различных реализациях неопределенных факторов (рис. 10.40).

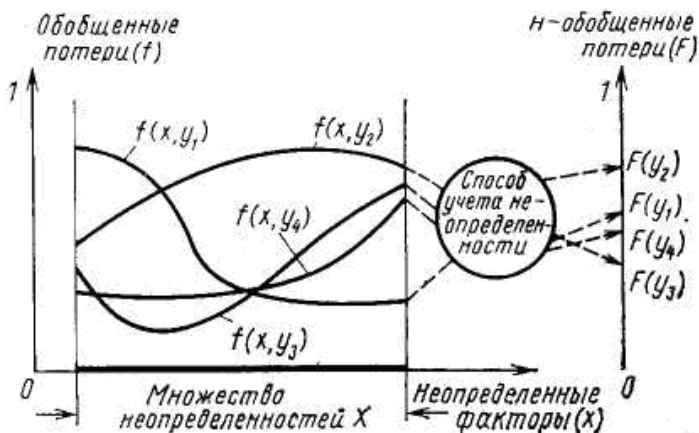


Рис. 10.40. Выбор рекомендаций в условиях неопределенности на основе вычисления n-обобщенных потерь

В отдельных случаях, когда один вариант рекомендации оказывается лучше другого при всех реализациях неопределенных факторов (например, y_3, y_4 по сравнению с y_2), выбор наилучшей рекомендации не представляет труда. Однако в общем случае он является трудноразрешимой проблемой, решаемой целиком на субъективной основе (например, выбор лучшего из вариантов y_1, y_3, y_4). Эта проблема существенно упростилась бы, если бы ЛФР сумело оценить эффективность рекомендации y агрегированно на всем множестве

неопределенностей, т. е. заменить обобщенные потери $f(x, y)$ как определенную на X функцию от x одним числом $F(y)$, которое выражало бы обобщенные потери для рекомендации y , вычисленные с учетом всего множества неопределенностей X . **Такие потери сокращенно будем называть н-обобщенными**. Рассчитав н-обобщенные потери, не составляло бы труда указать наилучшую рекомендацию \bar{y} (например, $\bar{y} = y_3$, рис. 10.40):

$$\bar{y} = \arg \min_{y \in Y} F(y).$$

В практике формирования рекомендаций будем использовать **критерии-аналоги** н-обобщенных потерь. Чаще всего они принимают форму оценок «в среднем», когда

$$F(y) = \frac{1}{n} \sum_{x \in X} f(x, y), \quad X = \{x_i\}$$

или «гарантированных оценок»:

$$F(y) = \max_{y \in Y} f(x, y)$$

(например, средняя дальность полета, коэффициент загрузки самолета, максимальная грузоподъемность и т. п.).

При любом фиксированном y функция $f(x, y)$ есть функция, заданная на множестве X , поэтому **правило, сопоставляющее ей число $F(y)$, является функционалом**. Если ЛФР может заранее указать конкретный способ учета неопределенности, т. е. функционал s , то

$$F(y) = s[f(x, y)],$$

и формирование рекомендаций сводится к оптимизационной задаче. В этом случае, хотя неопределенность в виде множества X присутствует в консультационном модуле, процесс формирования рекомендаций утрачивает специфические черты, свойственные задачам формирования рекомендаций в условиях неопределенности. В дальнейшем мы будем рассматривать случаи (наиболее характерные для практики формирования рекомендаций), когда изначально невозможно указать единственный адекватный способ учета неопределенности, а можно говорить лишь о том, что объективно существует множество допустимых способов учета неопределенности, обозначаемое далее S . Для описания множества S проще всего было бы предложить ЛФР, y предварительно сформулировать интересующие его способы учета неопределенности, т. е. задать множество S перечислением элементов. В отдельных задачах так и приходится поступать, но как общий метод этот подход не может быть принят, так

как он аналогичен перебору вариантов при решении оптимизационных задач.

Можно было бы сформировать множество S из наиболее часто используемых способов учета неопределенности (к ним можно отнести критерии Вальда, Гурвица, Лапласа, Сэвиджа и др., и то лишь в случаях неопределенности критериев), но совершенно неясно, в какой мере эти функционалы, предложенные в разное время разными авторами для решения совершенно различных задач, могут отражать интересы ЛФР,а, выполняющего конкретную консультационную операцию.

Предлагаемый способ разрешения указанной проблемы состоит в том, чтобы сочетать со сведениями, получаемыми от ЛФР,а, аксиоматическое описание множества S . При этом система аксиом должна отражать естественные требования, налагаемые на любой разумный способ учета неопределенности. Ясно, что различным системам аксиом будут отвечать различные множества S , поэтому, задавая их, мы уже вносим в процесс формирования рекомендаций долю субъективизма, но, как было установлено выше, в этом проявляется объективная особенность формирования рекомендаций в условиях неопределенности. При этом возможные вредные последствия можно уменьшить за счет выбора наиболее естественной системы аксиом, достаточно хорошо отражающих отношение ЛФР,а к учету неопределенности, а также сохранения за ЛФР,ом возможности дополнительного включения в используемое множество S интересующих его способов учета неопределенности и исключения излишних.

Далее будет предложена достаточно естественная система аксиом, при которой доказано, что любой допустимый способ учета неопределенности сводится к вычислению средней оценки по множеству неопределенностей, однако осредняются не исходные обобщенные потери, а некоторые другие, вспомогательные потери; результат осреднения пересчитывается из шкалы вспомогательных потерь в шкалу исходных обобщенных потерь, задавая тем самым

n -обобщенные потери, соответствующие данному способу учета неопределенности. Пересчет из исходных потерь t во вспомогательные G и обратно производится по «тарировочной кривой» $G(t)$ — порождающей функции, различной для различных способов учета неопределенности. Эта функция непрерывна и строго монотонна, кроме того, $G(0) = 0$, $G(1) = 1$ (рис. 10.41). Если известна порождающая функция $G(t)$, то n -обобщенные потери,

соответствующие определяемому ею способу учета неопределенности, будем рассчитывать по формуле

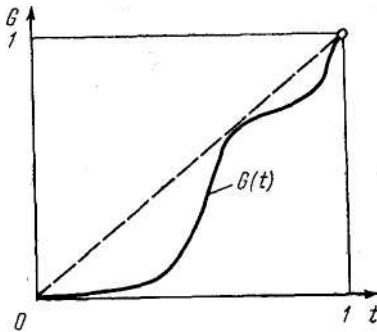


Рис. 10.41. Типичный вид порождающей функции, задающей некоторый допустимый способ учета неопределенности

$$F(y) = G^{-1} \left(\frac{1}{N} \sum_{i=1}^N G(f(x_i, y)) \right) \quad (10.129)$$

для множества неопределенностей X , состоящего из N элементов x_i , $i = 1, \dots, N$, или

$$F(y) = G^{-1} \left(\int_X D(f(x, y)) dx \right) \quad (10.130)$$

для множества неопределенностей X , являющегося областью меры Ω пространства конечного числа переменных.

В качестве примера рассмотрим три способа учета неопределенности A , B и C , заданные соответственно порождающими функциями

$G = t$, $G = t^2$ и $G = \sqrt[4]{t}$ (рис. 10.42).

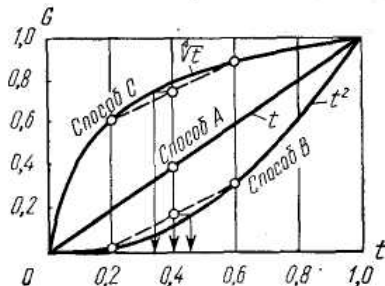


Рис. 10.42. Примеры порождающих функций

Первый способ учета характеризуется совпадением вспомогательных и исходных потерь. При способе учета неопределенности B вспомогательные потери меньше исходных, а при способе C — больше. На первый взгляд способы B и C соответствуют более оптимистичному и более пессимистичному учетам неопределенностей по сравнению со способом A . Чтобы показать, что это не так, воспользуемся рассматриваемыми способами для оценки одним числом неопределенности X , принимающей два значения: 0,2 и 0,6. Из соотношения (10.129) следует, что для этого нужно найти аргумент функции $G(t)$, при котором она принимает значение

$$[G(0,2) + G(0,6)]/2.$$

Соответствующие построения проведены на рис. 10.42, откуда видно, что искомая оценка (n -обобщенные потери) составляет при способе учета неопределенности A — 0,4, при способе B — 0,47, при способе C — 0,32.

Таким образом, способ, ранее названный пессимистическим, дает лучшее значение n -обобщенных потерь, чем более оптимистические. Это показывает, что на направление смещения значения n -обобщенных потерь относительно средней оценки влияет не абсолютное значение порождающей функции, а ее кривизна: при вогнутой функции оценка значения неопределенной величины оказывается пессимистической, а при выпуклой — оптимистической.

Итак, множество допустимых способов учета неопределенности S в методе ПРИНН задается совокупностью всех непрерывных строго монотонных функций $G(t)$, графики которых проходят через точки $(0; 0)$ и $(1; 1)$. Важнейшим свойством такого множества является его инвариантность к конкретным условиям задач формирования рекомендаций в условиях неопределенности. Можно сказать, что в виде элементов этого множества (порождающих функций) удастся выделить инвариантные компоненты учета неопределенности, в то время как их предметные компоненты, отражающие реальные условия решаемых задач, находят выражение в структуре формул (10.129), (10.130), включающих конкретные обобщенные потери и конкретное множество неопределенностей данной задачи. Достоинством описанного множества S является его наглядность и возможность простой трактовки правила расчета n -обобщенных потерь как осреднения вспомогательных потерь.

Для практического использования множества S , содержащего бесконечное число элементов, требуется заменить это множество его конечным подмножеством T , причем оптимально, т. е. так, чтобы T с наибольшей полнотой представляло все множество S . Ввиду

инвариантности S , его конечное подмножество T также будет инвариантно, что позволяет назвать его **типовым набором способов учета неопределенности**. Это понятие является одним из центральных в методе ПРИНН.

Для пояснения основной идеи формирования типового набора способов учета неопределенности проведем аналогию между рассматриваемой проблемой и исследованием некоторой функции $R(z)$ на множестве Z — **отрезке числовой оси**. Предположим, что консультант не имеет никакой информации о $R(z)$, но может вычислить ее значение в любой точке $z \in Z$. Поскольку вычислить значения $R(z)$ во всем бесконечном числе точек Z невозможно, консультант первоначально рассчитает ее значения в некотором конечном числе точек

$Z_K = \{z_i\}_{i=1,2,\dots,K}$, причем постарается разместить эти точки в Z равномерно, т. е. так, чтобы они наилучшим образом позволяли прозондировать это множество. Типовой набор способов учета неопределенности играет относительно множества S ту же роль, что и Z_K относительно Z . Чтобы перенести на него требование равномерности распределения в представляемом им множестве, обратим внимание на следующее свойство Z_K . Для любой точки $z \in Z$ можно рассчитать ее «расстояние» до Z_K как расстояние до ближайшей к z точки Z_K :

$$l_{z_k}(z) = \min_{z' \in Z_k} \rho(z, z'),$$

где $\rho(z, z')$ — расстояние между точками $z, z' \in Z$. Максимальное из таких расстояний, рассчитанных для всех точек $z \in Z$, задает погрешность представления множества Z подмножеством Z_K :

$$\varepsilon_{z_k} = \max_{z \in Z} l_{z_k}(z),$$

так как любая точка $z \in Z$ может быть заменена одной из точек Z_K , отстоящей от нее на расстоянии, не большем ε_{z_k} . Очевидно, что из всех подмножеств Z , содержащих данное число точек, Z_K обладает наименьшей погрешностью представления Z . Таким образом, для построения типового набора T , наиболее полно представляющего множество S допустимых способов учета неопределенности, следует ввести в S понятие «расстояния» между двумя любыми способами учета неопределенности, а затем найти набор из заданного числа элементов множества S , обеспечивающий минимум погрешности ε_T представления S с помощью T :

$$\varepsilon_T = \max_{G \in S} \min_{G' \in T} \Delta(G, G') \quad (10.131)$$

здесь $\Delta(G, G')$ — расстояние между способами учета неопределенности $G, G' \in S$. За это расстояние естественно принять площадь фигуры, ограниченной графиками функций $G(t)$ и $G'(t)$ (рис. 10.43), т. е.

$$\Delta(G, G') = \int_0^1 |G(t) - G'(t)| dt \quad (10.132)$$

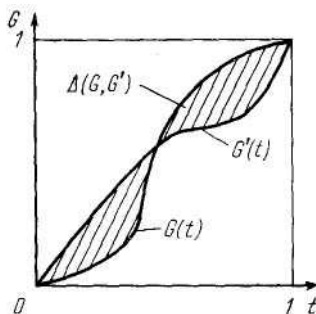


Рис. 10.43. К определению «расстояния» между способами учета неопределенности

После такого задания расстояния $\Delta(G, G')$ построение типового набора способов учета неопределенности становится хотя и достаточно сложной, но обычной задачей математического программирования. Ее решение позволило получить погрешность представления множества допустимых способов учета неопределенности типовыми наборами в зависимости от числа таких наборов (m), показанную на рис. 10.44.

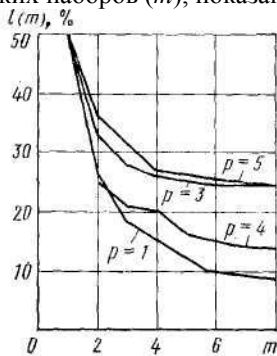


Рис. 10.44. Погрешность представления множества допустимых способов учета неопределенности типовыми наборами (при аппроксимации порождающих функций n -звенными ломаными)

При этом можно заметить, что, начиная с 7—8 типовых способов учета неопределенности, кривая погрешности представления ими всех допустимых способов учета неопределенности выходит на участок насыщения, на котором увеличение числа типовых способов лишь незначительно увеличивает точность. Сравнительный анализ представленных результатов позволил рекомендовать в качестве наиболее предпочтительного типовой набор из семи порождающих функций (см. табл. 10.3 и рис. 10.45), которым для удобства пользования присвоены индивидуальные наименования: наихудшая, наилучшая, средняя, осторожная, оптимистическая, релейная, нивелирующая.

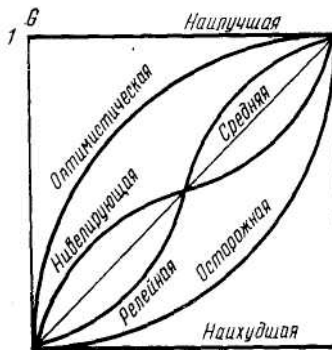


Рис. 10.45. Порождающие функции типового набора допустимых способов учета неопределенности ($m = 7$)

Таблица 10.3

Название порождающей функции	Вид порождающей функции	Н-обобщенные потери для $X = [0, 1], f(x) = x$
Наихудшая	$t^k, k \rightarrow \infty$	1,00
Наилучший	$t^k, k \rightarrow \infty$	0,00
Средняя	t	0,50
Осторожная	t^4	0,61
Оптимистическая	$\sqrt[4]{t}$	0,41
Релейная	$\frac{1}{2} (1 + \sqrt[5]{2t - 1})$	0,70
Нивелирующая	$\frac{1}{2} (1 + (2t - 1)^5)$	0,85

При формировании представленного набора учитывались следующие соображения: отразить все типы линий, полученных в расчетах на ЭВМ, ненамного отклониться от полученной на ЭВМ оптимальной аппроксимации множества и при этом включить в набор три общепринятых подхода к учету неопределенности (средняя, наилучшая, наихудшая), подобрать достаточно простые выражения для уравнений порождающих функций, входящих в набор. Ошибка при замене любой порождающей функции наиболее близкой к ней функцией из предложенного набора не превосходит 14 %

10.2.3. Процедура формирования рекомендации при функционировании консультационного модуля

Рассматривая локальное функционирование консультационного модуля, следует полагать, что *изначально заданы три множества: неопределенностей X , допустимых рекомендаций (стратегий) Y , способов учета неопределенностей S (или типовой набор способов учета неопределенности T), а также функция обобщенных потерь $f(x, y)$, $x \in X$, $y \in Y$. Эти компоненты описывают состояние системы «консультируемая проблема и ЛФР» в начале процесса формирования рекомендаций. Если отвлечься от ограничений по времени, отведенному на формирование рекомендаций, мощности используемого КТС, то идеальный консультационный процесс выглядел бы так. Для каждого допустимого способа учета неопределенности $s \in S$ рассчитывается оптимальная рекомендация $\tilde{y}_s \in Y$. Ее эффективность оценивается при любых способах учета неопределенности $z \in S$, поскольку неизвестно, какой из них является наиболее адекватным данной задаче. Так определяется функция $F_z(\tilde{y}_s)$, $s \in S$, $z \in S$. Зависимости \tilde{y}_s , $F_z(\tilde{y}_s)$, $s \in S$, $z \in S$ анализируются ЛФР,ом. Оно окончательно выбирает способ учета неопределенности и порожденную им рекомендацию (формирует рекомендацию) или изменяет исходные компоненты задачи, описывающие как консультируемую проблему ($X, Y, f(x, y)$), так и его самого (S), после чего процесс повторяется.*

Такая процедура обеспечивает полноценный учет влияния неопределенных факторов на формирование рекомендации и в то же время позволяет ЛФР,у пользоваться агрегированной, обобщенной оценкой этого влияния в виде n -обобщенных потерь. Преимущества такого подхода видны на следующем примере. Предположим, что имеет место исключительно благоприятный для формирования рекомендации случай, когда при любых неопределенных факторах

$x \in X$ оптимизация обобщенных потерь приводит к одной и той же рекомендации \tilde{y} . Тогда ЛФР, использующий указанную процедуру, сразу же получит от ЭВМ один-единственный рекомендуемый вариант рекомендации $\tilde{y}_s = \tilde{y}$ для любых допустимых способов учета неопределенности, в то время как если бы оно вело учет неопределенности непосредственно по обобщенным потерям и множеству неопределенностей, ему пришлось бы анализировать изменение $f(x, y)$ на всем множестве $X \times Y$. Можно сказать, что учет неопределенности с помощью n -обобщенных потерь позволяет заранее, до передачи информации ЛФР,у, исключить из рассмотрения ту часть неопределенной информации, которая является несущественной для оценки эффективности сформированной рекомендации.

Описанная идеальная процедура нереализуема в чистом виде ввиду бесконечности множества S , однако при его замене типовым набором способов учета неопределенности T она может быть осуществлена. Формой представления ее результатов является **матрица выбора**. Число ее строк и столбцов равно числу способов учета неопределенности в типовом наборе T , i -я строка соответствует варианту рекомендации, оптимальной с позиций i -го способа учета неопределенности из типового набора, а j -й столбец — j -му способу учета неопределенности из этого же набора.

Элемент матрицы выбора F_i^j — это n -обобщенные потери, рассчитанные по j -му способу учета неопределенности для варианта сформированной рекомендации, оптимальной при i -м способе учета неопределенности. Таким образом, матрица выбора

$$W = \|F_i^j\|_{i,j=1,\dots,m}$$

задает перекрестную оценку в шкале n -обобщенных потерь наилучших рекомендаций по всему типовому набору способов учета неопределенности. Она предоставляет ЛФР,у богатый материал для анализа, позволяя сконцентрировать внимание на рекомендациях, обладающих наилучшими характеристиками при различных аспектах учета неопределенности, и дать их сравнительную оценку при всех этих аспектах. Еще раз подчеркнем, что неправомерно ставить вопрос — а где гарантия, что среди этих рекомендаций действительно находится «самая лучшая». В рассматриваемой ситуации формирования рекомендаций понятие «самая лучшая» рекомендация не определена. Можно лишь утверждать, что среди рекомендаций, составляющих матрицу выбора, нет «заведомо худшей», так как все они выделены, исходя из допустимых (следовательно, разумных)

способов учета неопределенности, и, кроме того, сделано все возможное, чтобы с максимальной полнотой учесть неопределенные факторы: множество способов учета неопределенности описано математически строго и полно, а его замена конечным типовым набором произведена наилучшим образом. ***Тем самым совокупность рекомендаций, вошедших в матрицу выбора, хорошо обоснована, а это — самое большое, что можно сделать в соответствии с принципом обоснованности.***

Неформальный содержательный анализ ЛФР,ом матрицы выбора безусловно полезен, однако нет уверенности, что исходя из него оно сможет оперативно выбрать однозначную рекомендацию. Поэтому ***целесообразно повторно применить описанный метод для процедуры формирования рекомендаций на основе матрицы W .*** Эту процедуру тоже можно рассматривать как формирование рекомендаций в условиях неопределенности, в котором эффективность каждого варианта рекомендации оценивается ***вектором критериев — соответствующей строкой матрицы выбора***, т. е. n -обобщенными потерями при различных способах учета неопределенности из типового набора.

В такой задаче вновь можно рассчитать матрицу выбора W' . Встает вопрос: не будет ли эта матрица повторением исходной матрицы W , поскольку в ней используется тот же типовой набор способов учета неопределенности в применении к вариантам рекомендаций, однажды уже выделенным из множества допустимых рекомендаций с помощью этого набора. Нет не будет, так как исходной информацией для построения матрицы выбора W были обобщенные потери и множество неопределенностей X , а для матрицы W' ею являются n -обобщенные потери, записанные по столбцам матрицы W . Таким образом, при оценке по некоторому способу учета неопределенности эффективности варианта рекомендации в матрице W' , учитываются оценки его эффективности по всем другим способам учета неопределенности. Это приводит к сближению оценок эффективности варианта рекомендации по различным способам учета неопределенности, а как следствие, — к тому, что некоторые из вариантов рекомендаций, входящих в матрицу W , могут уже не войти в матрицу W' , так как с позицией соответствующего способа учета неопределенности они перестали быть оптимальными. ***Если в результате этого в матрице W' остался единственный вариант рекомендации, он и рекомендуется для окончательного выбора. Если же в матрицу W' вошло несколько вариантов рекомендаций, следует на ее основе повторить описанный выше процесс***, т. е. рассчитать матрицу выбора W'' , затем,

если потребуется, $W^{(3)}$ и так далее, вплоть до матрицы выбора $W^{(p)}$, *содержащей единственный вариант рекомендации.*

Смысл указанной процедуры состоит в формализации естественного критерия выбора рекомендаций: за наилучшую принимают ту рекомендацию, которая хороша со всех «точек зрения». Если первоначально «взгляды» на наилучшую рекомендацию расходятся, то впоследствии происходит согласование позиций, при котором в каждой точке зрения учитываются оценки, данные с других позиций, после чего выбор повторяют, и так до получения совпадающего результата.

Проведение описанной процедуры приводит к следующим результатам. Во-первых, объективно определяется уровень влияния неопределенности (УВН) на рекомендацию, характеризуемой числом «присоединенных задач», или этапов согласования, потребовавшихся для завершения консультационной операции. При $УВН \leq 3$ выбор можно считать достаточно надежным. Во-вторых, выявляется наилучшая рекомендация, а также (для сравнения с ней) несколько наиболее конкурентоспособных вариантов рекомендаций, входивших в промежуточные матрицы выбора.

Таким образом, к началу функционирования КМ определено множество неопределенностей X , множество допустимых вариантов сформированных рекомендаций Y , а также сформирована модель, позволяющая рассчитать вектор показателей эффективности для каждого варианта рекомендации $y \in Y$ при различных реализациях неопределенных факторов $x \in X$ и выделить из всех этих вариантов парето-оптимальные.

Локальное функционирование консультационного модуля начинается с исследования возможности уменьшения выявляемой неопределенности за счет привлечения дополнительной информации и, если позволяют условия осуществлять процесс формирования рекомендаций, каждая такая возможность используется. Затем начинается *собственно процесс формирования наиболее рациональной рекомендации.* Для этого формируется консультационная установка, т. е. показатели эффективности распределяются, в случае надобности, по группам сравнительной важности; для допустимых вариантов сформированных рекомендаций рассчитываются значения показателей эффективности при различных реализациях неопределенных фактов. Получаемые при этом данные автоматически анализируются рассмотренным выше способом и рассчитываются наиболее рациональная рекомендация, ее УВН и несколько близких к ней по эффективности рекомендаций. Далее мо-

жет быть скорректирована консультационная установка, дополнен банк рассматриваемых сформированных рекомендаций, изменена исходная постановка консультационной задачи, что может потребовать переформирования консультационного модуля, и т. д., после чего процедура выбора повторяется. Все эти операции легко осуществляются в диалоге со специально разработанным комплексом ПРИНН, реализованном на персональных ЭВМ.

10.2.4. Формальное (математическое) обоснование описания многоцелевой системы консультирования

10.2.4.1. Модель многоцелевой системы консультирования

Под многоцелевой системой консультирования (МСК) будем понимать определенный вид систем, который вводится в рассмотрение для того, чтобы формализованно описать и исследовать особенности формирования рекомендаций в условиях неустранимой неопределенности.

Расчленим множество элементов многоцелевой системы консультирования на внешнее множество X и множество элементов (стратегий) Y . Совокупность m любых элементов (центров) $y_j \in Y$, $j = 1, \dots, m$ назовем стратегией A :

$$A = \{y_j\}_{j=1,2,\dots,m}$$

Определим на X целочисленную распределяющую функцию, возможно неоднозначную, принимающую значения $1, 2, \dots, m$. Тогда каждому центру $y_j \in A$, $j = 1, \dots, m$ сопоставляется в X его область Дирихле E_j , в точках которой распределяющая функция принимает значение, равное j :

$$E_j = \{x \in X / E(x) = j\}, \quad j = 1, \dots, m.$$

Очевидно, что объединение областей Дирихле задает внешнее множество X , т. е.

$$\bigsqcup_{j=1}^m E_j = X,$$

и в случае однозначной $E(x)$ они не пересекаются:

$$E_j \cap E_k = \emptyset \quad \forall j, k = j, k = 1, \dots, m.$$

Таким образом, множество элементов системы структурировалось в набор множеств $X, A, \{E_j\}_{j=1,2,\dots,m}$. На нем определена характеристика (показатель) эффективности системы $f(X, A, \{E_j\}_{j=1,2,\dots,m})$. Учитывая, что области Дирихле однозначно задаются распределяющей

функцией $E(x)$, можно записать, что $f=f(X, A, E(x))$. Значения эффективности определяются значениями на отдельных элементах прямого произведения множеств $X, A, \{E_j\}_{j=1,2,\dots,m}$ и правилом их «свертывания» при объединении этих элементов и составленных из них множеств. «Свертывание» по элементам областей Дирихле приводит к зависимости эффективности МСК от некоторой интегральной скалярной характеристики области Дирихле $\mu(E_j)$. Поэтому введем **функцию локальной эффективности** (вообще говоря, **функционал**):

$$f = f(x, y, \mu(E_{E(x)})), \quad x \in X, \quad y \in Y. \quad (10.133)$$

Здесь и далее использование $E(x)$ в качестве индекса подчеркивает, что его значение определяется выбранной распределяющей функцией. Функция (10.133) описывает эффективность реагирования элемента y МСК на реализацию x внешних факторов при условии распределения их между элементами системы, задаваемого распределяющей функцией $E(x)$.

Функция (10.133) определена для любых $y \in Y, \mu(E_{E(x)}) \geq 0$, но не для любых $x \in X$. Область ее определения в X будем обозначать через $J(y)$, т. е. считать зависящей от соответствующего элемента $y \in X$. В этом случае необходимо потребовать

$$E_{E(x)} \subset J(y_{E(x)}), \quad (10.134)$$

$$\bigcap_{j=1}^m J(y_j) = X$$

Важное значение в различных приложениях имеют различные частные варианты задания функции локальной эффективности (10.133):

$$a) J(y) = X, \quad \forall y \in Y, \quad (10.135)$$

т. е. функция (10.133) определена на всем множестве X ;

$$б) f = \rho(x, y),$$

т. е. функционал (10.133) обращается в функцию;

$$в) f = \rho(y_j, \mu_j), \quad \text{где } \mu_j = \mu(E_j),$$

и т. п.

Функция (10.133) и правило «свертывания» ее значений по элементам внешнего множества и стратегии определяют характеристику МСК, которую будем называть **показателем эффективности**:

$$f = f(X, A, E(x)). \quad (10.136)$$

Правило «свертывания» по элементам внешнего множества может состоять, например, в том, что за значение показателя эффективности принимается наибольшее из значений функции локальной эффективности (скалярной) на элементах внешнего множества. МСК с

таким правилом «свертывания» назовем *гарантирующей* (ГМСК), с иными — *интегральной* (ИМСК).

Введение выражения (10.136) позволяет ставить задачу формирования оптимальной МСК, характеристики которой выбраны из условий векторной оптимизации (для определенности, минимизации) показателя эффективности. При этом будем рассматривать следующие случаи:

1) **задача оптимального распределения:** внешнее множество X и стратегия A заданы, необходимо найти оптимальную распределяющую функцию $E(x)$:

$$f(X, A, \bar{E}(x)) = \min_{E(x)} f(X, A, E(x)); \quad (10.137)$$

2) **задача оптимизации стратегии:** внешнее множество X , множество стратегий Y и число центров m стратегий заданы, требуется определить оптимальную стратегию A и распределяющую функцию $E(x)$:

$$f(X, \bar{A}, \bar{E}(x)) = \min_{\substack{E(x) \\ A \subset Y \\ m = \text{const}}} f(X, A, E(x)); \quad (10.138)$$

3) **общая оптимизация:** внешнее множество X и множество стратегий Y заданы, требуется найти оптимальную стратегию A и распределяющую функцию $E(x)$:

$$f(X, \bar{A}, \bar{E}(x)) = \min_{\substack{E(x) \\ A \subset Y}} f(X, A, E(x)); \quad (10.139)$$

Рассматриваемые в модели МСК стратегии A содержат m автономных элементов $y_j, j=1, \dots, m$, взаимодействие которых заключается в перераспределении между собой элементов, входящих во внешнее множество. Это распределение устанавливает функции $E(x)$ и, таким образом, области Дирихле E_j определяют в X область реагирования каждого центра стратегии. В целом каждая стратегия A определяет собой некоторый определенный вариант многоцелевой консультируемой проблемы.

В консалтинге внешнее множество часто задает совокупность выполняемых консультируемой проблемой задач, и, как правило, консультируемая проблема u может выполнять не любые задачи $x \in X$ (скажем, болт не может воспринимать нагрузки больше, чем разрушающие). Эта особенность вводится в модель МСК путем задания областей определения $J(y)$ показателей эффективности (10.133). Выполнение условий (10.134) гарантирует, что стратегия A действительно обеспечит выполнение всех задач $x \in X$.

Функции (10.133), описывающие эффективность реагирования элементов $y \in A \subset Y$ на элемент $x \in X$, зависят от интегральной характеристики (меры) соответствующей области Дирихле. Это расширяет сферу приложения модели, так как, например, в экономических задачах эффективность (стоимость) изделия определяется не только его характеристиками и параметрами выполняемой задачи, но и размером партии выпускаемых изделий (серийностью), а эта последняя зависит от всей совокупности выполняемых данным типом изделий задач (попутно заметим, что в этом случае функция локальной эффективности имеет вид $\rho(y_i, \mu_i)$).

Целесообразность постановки трех сформулированных задач оптимизации очевидна. Приведем несколько модификаций математической модели МСК, которые направлены в сторону упрощения исходной постановки.

Скалярной МСК будем называть МСК, в которой вектор эффективности имеет лишь одну компоненту, в общем же случае будем говорить о векторной МСК.

При исследовании сформированных рекомендаций в условиях неопределенности достаточно рассматривать скалярную МСК, так как неопределенность критерия, связанная с многокритериальностью, находит полное отражение во внешнем множестве X путем включения в него множества значений весовых коэффициентов из свертки критериев.

Одноцентрковой назовем МСК, в которой стратегии A содержат лишь по одному элементу множества стратегий, т. е. $m = 1$.

Рассмотрение скалярной одноцентрковой МСК, внешнее множество которой состоит из единственного элемента $X = x^*$, приводит к обычной задаче оптимизации. Действительно, в этом случае функция эффективности записывается в виде

$$f = f(x^*, y, 1) = F(y).$$

В этом случае задача распределения теряет смысл, а задача оптимизации стратегии совпадает с задачей общей оптимизации и сводится к определению элемента $y \in Y$, минимизирующего $F(y)$:

$$F(\bar{y}) = \min_{x \in Y} F(y).$$

10.2.4.2. Модель выбора формируемых рекомендаций в условиях неопределенности

Рассмотрим следующий ряд некоторых многоцелевых систем консультирования. Системой выбора типового набора способов учета неопределенности назовем ГМСК $\langle S, S, \triangleright \rangle$. Ее внешнее множество и

множество стратегий есть множество функционалов, описывающих учет неопределенности, функция локальной эффективности $l(\varphi, \psi)$, $\varphi, \psi \subset S$ есть расстояние от элемента внешнего множества φ до элемента множества стратегий ψ , а значение $l = l(S, A, E(\varphi))$, $\varphi \in S$ есть максимальное из расстояний от элементов внешнего множества до ближайших элементов множества стратегий. Ясно, что оптимальная m -элементная стратегия $A\mathcal{F}_m \subset S$ есть оптимальная l -сеть в S , $l = l(S, A\mathcal{F}_m, E(\varphi))$. Эту стратегию назовем типовым набором способов учета неопределенности, а отдельные ее элементы $\mathcal{F}_m^j \in A\mathcal{F}_m, j = 1, \dots, m$ — типовыми способами учета неопределенности. Варьируя m , можно построить $\bar{l}(m)$, значения которой равны значениям показателя эффективности $l(S, A\mathcal{F}_m, E(\varphi))$ на оптимальной m -элементной стратегии $A\mathcal{F}_m$ при оптимальной распределяющей функции. Эту функцию назовем характеристикой возможностей типизации учета неопределенностей.

Для фиксированного m введем группу МСК $\langle X, Y, \mathcal{F}_{mf}^j \rangle$, $\mathcal{F}_{mf}^j \in A\mathcal{F}_m, j = 1, \dots, m$. Здесь X — внешнее множество, Y — множество стратегий, f — функция локальной эффективности, а показатель эффективности F получен суперпозицией f и одного из типовых способов учета неопределенности из m -элементного типового набора $M\mathcal{F}_m$. Таким образом, этот типовой представитель используется в качестве правила «свертывания» при формировании показателя эффективности МСК. Обозначим через $A^{X,m,l}$ оптимальные A -элементные стратегии, а через $F_{M,j}^{m,j}$ — соответствующие им значения показателей эффективности МСК. Ясно, что $A_M^{X,m,l}$ есть результат оптимального выбора M -элементной рекомендации, если для учета неопределенности использован j -й типовой представитель из m -элементного типового набора способов учета неопределенностей, а значение $F_{M,j}^{m,j}$ измеряет эффективность такого выбора.

Наконец, обозначим через $F_{M,\kappa}^{m,j}, j, \kappa = 1, \dots, m$ значение показателя эффективности МСК $\langle X, Y, \mathcal{F}_{mf}^j \rangle$ на стратегии $A_M^{X, m, \kappa}$. Оно измеряет эффективность M -элементной стратегии, оптимальной для способа учета неопределенности $\mathcal{F}_m^j \subset A\mathcal{F}_m$, для другого способа учета неопределенности $\mathcal{F}_m^\kappa \in A\mathcal{F}_m$. Очевидно, что

$$F_{M,j}^{m,j} \leq F_{M,\kappa}^{m,j} \quad \forall \kappa \neq j, \kappa, j = 1, \dots, m.$$

Введем квадратную матрицу выбора W_M^m :

$$W_M^m = \| F_{M,\kappa}^{m,j} \|_{\kappa, j=1, \dots, m}$$

где j — номер строки, а κ — номер столбца. Из предыдущего ясно, что j -я строка матрицы выбора соответствует M -элементной стратегии, оптимальной для j -го представителя m -элементного типового набора способов учета неопределенности, а столбец κ соответствует элементу

k из этого набора. Число, стоящее на их пересечении, определяет эффективность выбора стратегии (определяемой соответствующей строкой) при способе учета неопределенности, определяемом соответствующим столбцом. Учитывая, что m -элементный типовой набор с точностью $\bar{l}(m)$ представляет всевозможные способы учета неопределенности, можно установить, что матрица W^m_M отражает с этой точностью влияние неопределенности на оптимальную M -элементную рекомендацию. Ее может использовать ЛФР для того, чтобы выбрать одну из стратегий $A_M^{X,m,j}$, $j = 1, \dots, m$ в качестве окончательной рекомендации для исходной задачи. Эта матрица предоставляет ему в наглядной форме обладающую полнотой агрегированную информацию как об оптимальных возможностях консультируемой проблемы, так и о степени влияния на них неопределенных факторов. При этом ЛФР может изменять степень агрегирования, задавая соответствующие значения m , и управлять в допустимых пределах степенью адаптации к неопределенным факторам, устанавливая соответствующие значения M .

Выбор стратегии $A_M^{X,m,j}$ может быть, по желанию ЛФР, выполнен формализованно, без его вмешательства. Для этого достаточно рассмотреть присоединенную задачу формирования рекомендаций в условиях неопределенности критериев, в которой множество стратегий составляют стратегии $A_M^{k,m,j}$, а соответствующая строка матрицы выбора рекомендаций задает вектор критериев оптимизации каждой стратегии. Строя матрицу выбора рекомендаций для этой задачи, можно уменьшить число различных стратегий, входящих в матрицу выбора рекомендаций. Если оно все еще превышает единицу, необходимо построить аналогичным образом следующую присоединенную задачу и так до тех пор, пока в рассмотрении не останется единственная рекомендация.

10.2.4.3. Описание множества допустимых способов учета неопределенности

Основной особенностью метода ПРИНН является учет неопределенных факторов в задаче не непосредственно, а через влияние, оказываемое ими на эффективность допустимых рекомендаций. Степень этого влияния может оцениваться по-разному (от «пессимистического» до «оптимистического» варианта). При этом суть метода ПРИНН состоит в том, чтобы не пытаться найти «правильный» способ такой оценки, так как в сложных задачах он не может быть уверенно указан (особенно априори), а построить представительное множество способов учета неопределенности

Прежде всего нужно отметить, что способ учета неопределенности ЛФР имеет две составляющие: проблемно-ориентированную и инвариантную. Формирование первой из них происходит в терминах конкретной задачи формирования рекомендаций, требует глубокого проникновения в ее суть и завершается введением понятий или параметров, являющихся входными для второй составляющей, которая оперирует с ними, используя общематематические термины.

Проследим это деление на примере обычных задач оптимизации. ***Здесь проблемно-ориентированная составляющая заключается в выборе показателя эффективности и установлении его зависимости от допустимых параметров рекомендации. Инвариантная же составляющая представляет собой требование выбора таких параметров рекомендации, при которых показатель эффективности принимает экстремальное значение.*** Ввиду кажущейся естественности этого требования, самостоятельная роль инвариантной составляющей обычно не подчеркивается, а она весьма велика. ***Если проблемно-ориентированная составляющая поставляет инвариантной информации о конкретной задаче, то инвариантная указывает, каким условиям эта информация должна удовлетворять.*** Например, в обычной задаче оптимизации показатель эффективности должен быть единственным, и если на проблемно-ориентированном уровне этого нельзя добиться, не искажая смысла решаемой задачи, то скалярный оптимизационный подход не может быть применен. В таком случае нужно принимать иную инвариантную составляющую, например, векторную оптимизацию.

Выделение инвариантной составляющей позволяет преобразовать структуру задачи формирования рекомендаций в условиях неопределенности. Объединим модель консультируемой проблемы (КП) с проблемно-ориентированной составляющей способа учета неопределенности ЛФР в ***единую модель «КП + ЛФР»***, которую впредь будем называть просто моделью. Тогда задачу формирования рекомендаций при фиксированном способе учета неопределенности составят модель и инвариантная составляющая этого способа. Учтем теперь, что ЛФР не может указать определенного способа учета неопределенности, а вынуждено оперировать множеством таких способов. С одной стороны, это порождает множество моделей, но поскольку в модель входит множество неопределенностей X , вместо множества моделей ***можно рассмотреть единую модель с модифицируемым множеством неопределенностей.*** Для этого достаточно проблемно-ориентированную составляющую способа учета неопределенности

ЛФР задать некоторыми числовыми параметрами, значения которых считать неопределенными,

С другой стороны, присущая ЛФР множественность способов учета неопределенности приведет к появлению множества инвариантных составляющих способов учета неопределенности ЛФР. Это множество инвариантных составляющих будет одинаковым для различных задач и ЛФР (хотя в разных задачах ЛФР могут предпочитать разные его элементы). Впредь будем его называть просто **множеством способов учета неопределенности** и по-прежнему обозначать символом S , так же как на модель «КП + ЛФР» мы распространим прежние символы X , Y и $f(x, y)$. Таким образом, изменения, вносимые ЛФР при формировании рекомендации в модель консультационной проблемы и в собственное отношение к неопределенности, найдут отражение в изменении модели $X, Y, f(x, y)$. Множество S опишем **аксиоматически**, определив тем самым множество инвариантных составляющих способов учета неопределенности. Этим мы определим требования к модели. Причем **критерий** того, насколько удачно введено множество S , будет лишь один: **степень охвата задач формирования рекомендаций, которые можно адекватно описать в рамках этих требований**.

Рассмотрим оценку эффективности некоторой стратегии A с учетом множества неопределенностей X на основе значений функции локальной эффективности $f(x, A) \equiv f_A(x)$, $x \in X$, определяемой моделью «КП + ЛФР». При этом будем полагать

$$f_A(x) \in [0, 1] \forall x \in X. \quad (10.140)$$

Дадим следующее определение: **способ учета неопределенности** s — это **правило**, однозначно сопоставляющее любому подмножеству $X_\alpha \subset X$ с определенной на нем функцией $f_A(x)$ некоторое число $F_A(X_\alpha)$:

$$\forall X_\alpha \subset X < X_\alpha \subset X, f_A(x), x \in X_\alpha > \xrightarrow{s} F_A(X_\alpha). \quad (10.141)$$

Способ учета неопределенности есть одновременно **функционал** и **функция** от множества. При **фиксированном** X_α — это функционал над множеством определенных на X функций $f_A(x)$, при **фиксированной на X функции** $f_A(x)$ — это функция от подмножеств X . **Число** $F_A(X_\alpha)$ будем называть **n -обобщенными потерями**, т. е. обобщенными потерями для стратегии A , вычисленными с учетом влияния множества неопределенностей X_α .

Перейдем к аксиоматическому описанию множества допустимых инвариантных составляющих способа учета неопределенности. Впредь, говоря о способе учета неопределенности, будем иметь в виду

только его инвариантную составляющую. Введем ограниченный набор **свойств** элементов множества S , достаточно естественных для задач формирования рекомендаций :

1) **нормировка**

$$F_A(X) \in [0, 1];$$

2) **осреднение**

$$\forall X_\alpha \subset X, \text{ если } f_A(x) = C \forall x \in X_\alpha \text{ то } F_A(X) = C ;$$

3) **монотонность**

$$\forall X_1, X_2 \subset X, X_1 \cap X_2 = \emptyset,$$

$$\text{если } F_A(X_1) > F_A(X_2),$$

$$\text{то } \forall X_3 \subset X, X_1 \cap X_3 = X_2 \cap X_3 = \emptyset, F_A(X_1 \cup X_3) > F_A(X_2 \cup X_3);$$

4) **устойчивость** — в пределах множества S допускается не нарушающее условий нормировки бесконечно малое варьирование значений $F_A(X_\omega)$.

Первое из перечисленных свойств означает, что поскольку обобщенные потери ограничены пределами 0 и 1 независимо от реализаций неопределенных факторов $x \in X$, учет всей совокупности этих факторов не должен выводить за указанные пределы.

Второе свойство фиксирует несомненный факт, что если неопределенные факторы не оказывают влияния на обобщенные потери, то н-обобщенные потери должны совпадать с обобщенными при любом допустимом способе учета неопределенности.

Свойство монотонности означает, что если н-обобщенные потери для одного множества неопределенностей больше, чем для другого, и каждое из них объединяется с некоторым третьим множеством, то н-обобщенные потери для первого объединенного множества больше, чем для второго.

Поясним смысл этого свойства. Пусть имеются две партии деталей. Каждая деталь характеризуется некоторым значением показателя качества, а каждая партия — некоторым интегральным значением этого показателя. По значению этого показателя первая партия хуже второй. Свойство монотонности означает, что если в каждую партию добавить детали одинакового качества, то первая партия все равно останется хуже второй.

Наконец, четвертое свойство позволит использовать вводимое аксиоматически множество S в реальных задачах, для которых характерно приближенное вычисление параметров, в том числе и н-обобщенных потерь.

Используя приведенные выше свойства, докажем существование так называемой **функции построения** $\Phi(u, v); 0 \leq u, v \leq 1; 0 \leq \Phi \leq 1$, такой,

что если X_1, X_2 — любые непересекающиеся подмножества внешнего множества X , то

$$F(X_1 \cup X_2) = \Phi(F(X_1), F(X_2)),$$

т. е. н-обобщенные потери для множества $X_1 \cup X_2$ зависят лишь от н-обобщенных потерь для составляющих его множеств X_1 и X_2 .

Теорема 1.

$$\begin{aligned} \exists \Phi = \Phi(u, v): \forall X_1, X_2 \subset X, X_1 \cap X_2 = \emptyset \\ F_A(X_1 \cup X_2) = \Phi(F_A(X_1), F_A(X_2)); \\ u, v, \in [0, 1]. \end{aligned} \quad (10.142)$$

Доказательство. Из требований монотонности и устойчивости следует, что каждый входящий в S способ учета неопределенности может быть описан некоторой числовой функцией, зависящей от числовых переменных, так называемой функцией построения $\Phi(u, v)$, $0 \leq u, v, \Phi \leq 1$. Покажем это. Пусть для некоторых множеств $X_1, X_2 \subset X$

$$F_y(X_1) = F_y(X_2). \quad (10.143)$$

Тогда для любого $X_3 \subset X$

$$F_y(X_1 \cup X_3) = F_y(X_2 \cup X_3). \quad (10.144)$$

Предположим, что это не так:

$$F_y(X_1 \cup X_3) > F_y(X_2 \cup X_3).$$

Произведем допускаяемое требованием устойчивости бесконечное малое варьирование $F_y(X_1)$ или $F_y(X_2)$ на величину, меньшую разности $F_y(X_1 \cup X_3) - F_y(X_2 \cup X_3)$, так, чтобы исходное равенство заменялось неравенством $F'_y(X_1) < F'_y(X_2)$ (штрихом обозначен результат варьирования). При этом сохранится неравенство $F'_y(X_1 \cup X_3) > F'_y(X_2 \cup X_3)$, а следовательно, нарушится условие монотонности.

Из выражений (10.143) и (10.144) следует, что если $X_1, X_2, X_\alpha, X_\beta \subset X$ — любые множества, такие, что $F_y(X_1) = F_y(X_\alpha) = u$, а $F_y(X_2) = F_y(X_\beta) = v$, то

$$F_y(X_1 \cup X_2) = F_y(X_\alpha \cup X_\beta)$$

Это означает, что н-обобщенные потери для множества $X_1 \cup X_2$ зависят лишь от н-обобщенных потерь для составляющих его множеств X_1, X_2 , т. е.

$$F_y(X_1 \cup X_2) = \Phi(F_y(X_1), F_y(X_2)). \quad (10.145)$$

По свойству нормировки аргументы и значения функции (10.143) заключены между 0 и 1.

Следствие. Функция построения симметрична и ассоциативна относительно своих аргументов:

$$\begin{aligned} \Phi(u, v) = \Phi(v, u); \\ \Phi(u, \Phi(v, z)) = \Phi(v, \Phi(u, z)). \end{aligned}$$

Доказательство. Следует из симметричности операции объединения множеств X_1, X_2 в условиях (10.142).

Отнесем к множеству допустимых способов учета неопределенности S лишь достаточно универсальные и простые способы учета неопределенности. Эти свойства проявятся в том, что описывающие их функции построения должны быть определены при любых значениях аргументов, заключенных между 0 и 1, и иметь непрерывные частные производные 1-го порядка. Для этого потребуем от элементов множества S удовлетворения еще **двум условиям**: универсальности и гладкости.

Условие универсальности состоит в том, что функция построения определена для любых $u, v \in [0, 1]$.

Условие гладкости состоит в том, что функция построения на $[0, 1] \times [0, 1]$ имеет непрерывные частные производные 1-го порядка.

Используя все наложенные на элементы S условия, можно установить ряд дополнительных свойств допустимых способов учета неопределенности.

Лемма 1.

$$\begin{aligned} \forall u \in [0, 1]; \\ \varphi(u) &\equiv \max_{v \in [0,1]} \Phi(u, v) = \Phi(u, 1); \\ \psi(u) &\equiv \min_{v \in [0,1]} \Phi(u, v) = \Phi(u, 0) \end{aligned}$$

строго монотонно возрастающие функции и

$$\max_{v \in [0,1]} \Phi(u, v) > \min_{v \in [0,1]} \Phi(u, v).$$

Доказательство. Из гладкости $\Phi(u, v)$ следует ее непрерывность на множестве $[0, 1] \times [0, 1]$, а поэтому она принимает при любом $u \in [0, 1]$ экстремальные значения по v на отрезке $[0, 1]$. Ввиду соотношения (10.141)

$$\forall v_1 > v_2 \quad \Phi(u, v_1) > \Phi(u, v_2),$$

т. е. функция построения строго монотонна по своим аргументам, а следовательно,

$$\begin{aligned} \max_{v \in [0,1]} \Phi(u, v) &= \Phi(u, 1); \\ \min_{v \in [0,1]} \Phi(u, v) &= \Phi(u, 0). \end{aligned}$$

Тогда

$$\Phi(u, 1) > \Phi(u, 0),$$

что завершает доказательство.

Теорема 2. Существует такое число $l \in [0, 1]$ — *метка*, что

$$\forall u \in [0, 1], \Phi(u, l) = u.$$

Доказательство. Обозначим $\Phi(0, 1) = \kappa$. Из теоремы 1 следует, что

$$0 \leq \kappa \leq 1.$$

Ввиду симметричности $\Phi(u, v)$ (по следствию из теоремы 1)

$$\Phi(1, 0) = \Phi(0, 1) = \kappa.$$

Из леммы 1 при $u = 1$ имеем

$$\min_{v \in [0,1]} \Phi(1, v) = \Phi(1, 0) = \kappa.$$

Ввиду строгой монотонности $\psi(u)$ можно записать

$$\forall u \in [0, 1] \quad \min_{v \in [0,1]} \Phi(u, v) \leq \min_{v \in [0,1]} \Phi(1, v) = \kappa. \quad (10.146)$$

Аналогично, ввиду строгой монотонности $\Phi(u, v)$

$$\max_{v \in [0,1]} \Phi(0, v) = \Phi(0, 1) = \kappa$$

и из строгой монотонности $\varphi(u)$

$$\forall u \in [0, 1] \quad \max_{v \in [0,1]} \Phi(u, v) \geq \Phi(0, v) = \kappa. \quad (10.147)$$

Тогда при $u = \kappa$ из условий (10.146), (10.147) следует

$$\min_{v \in [0,1]} (\kappa, v) \leq \kappa;$$

$$\max_{v \in [0,1]} \Phi(\kappa, v) \geq \kappa$$

и ввиду непрерывности функции $\Phi(\kappa, v)$ по v на отрезке $[0, 1]$ существует такое $l \in [0, 1]$, что

$$\Phi(\kappa, l) = \kappa.$$

Далее из условия (10.147) следует, что $\forall u \in [0, 1]$ значения функции $\Phi(u, v)$ при изменении v от 0 до 1 заключены в пределах, содержащих κ , поэтому

$$\forall u \in [0, 1] \quad \exists z_u \in [0, 1]: \Phi(u, z_u) = \kappa.$$

Ввиду ассоциативности функции построения

$$\Phi(\kappa, l) = \Phi(\Phi(u, z_u), l) = \Phi(\Phi(u, l), z_u) = \kappa = \Phi(u, z_u).$$

Но так как $\Phi(u, v)$ строго монотонна по своим аргументам, равенство

$$\Phi(\Phi(u, l), z_u) = \Phi(u, z_u)$$

возможно тогда и только тогда, когда

$$\Phi(u, l) = u,$$

что и требовалось доказать.

Теорема 3.

$$\begin{aligned} \forall u, v \in [0, 1]; & \quad (10.148) \\ G(\Phi(u, v)) &= G(u) + G(v) - G(l), \end{aligned}$$

где $G(t)$ — дифференцируемая, строго монотонно возрастающая на отрезке $[0, 1]$ функция — так называемая *поряджающая функция*.

Доказательство. Пусть $z = l + \Delta z$, где l — метка, а $\Delta z \rightarrow 0$. Для любых $u, v \in [0, 1]$ обозначим:

$$C = \Phi(C, z).$$

Тогда определено число

$$C_l = \Phi(C, z).$$

По свойству гладкости функция $\Phi(u, v)$, как имеющая непрерывные частные производные, дифференцируема в любой точке множества $[0, 1] \times [0, 1]$. Поэтому в некоторой окрестности точки (C, l)

$$C = \Phi(C, l + \Delta z) = \Phi(C, l) + \Phi'_v(C, l)\Delta z + \alpha, \quad (10.149)$$

где $\alpha / (\Delta z) \rightarrow 0$ при $\Delta z \rightarrow 0$.

Рассмотрим в этой окрестности разность

$$\Delta C = C_l - C = \Phi'_v(C, l)\Delta z + \alpha. \quad (10.150)$$

Покажем, что в интервале $[0, 1]$ существует бесконечно близкое к u число u' , такое, что

$$\Phi(u', v) = C_l, \quad u' = u + \Delta u, \quad \Delta u \rightarrow 0. \quad (10.151)$$

Действительно, по свойству ассоциативности $\Phi(u, v)$

$$C_l = \Phi(C, z) = \Phi(\Phi(u, v), z) = \Phi(\Phi(u, z), v).$$

Из близости z к l и непрерывности функции построения следует близость $\Phi(u, z)$ к $\Phi(u, l) = u$. Если обозначить $\Phi(u, z)$ через u' , то $u' \rightarrow u$. Из условия (10.151) можно получить соотношение, аналогичное (10.149):

$$\begin{aligned} \Delta u = u' - u &= \Phi(u, z) - u = \Phi(u, l + \Delta z) - u = \\ &= \Phi(u, l) + \Phi'_v(u, l)\Delta z + \beta - u, \end{aligned} \quad (10.152)$$

где $\beta / (\Delta z) \rightarrow 0$ при $\Delta z \rightarrow 0$.

Переходя в выражениях (10.150), (10.152) к пределу при $\Delta z \rightarrow 0$, получим

$$\frac{dC}{dz} = \Phi'_v(C, l), \quad \frac{du}{dz} = \Phi'_v(u, l),$$

или

$$\frac{dC}{\Phi'_v(C, l)} = \frac{du}{\Phi'_v(u, l)} \quad (10.153)$$

При этом заметим, что $\Phi'_v(u, l) > 0, \forall u \in [0, 1]$ ввиду строгой монотонности $\Phi(u, v)$.

Обозначим

$$\Phi'_v(t, l) = g(t),$$

где $g(t)$ — непрерывная и положительная функция. Тогда равенство (10.153) можно записать в виде

$$\frac{dC}{g(C)} = \frac{du}{g(u)} \quad (10.154)$$

Проинтегрируем выражение (10.154) от некоторой начальной точки (u_0, C_0) до текущей точки (u, C) , т. е.

$$\int_{C_0}^C \frac{dC}{g(C)} = \int_{u_0}^u \frac{du}{g(u)} \quad (10.155)$$

Введем первообразную

$$G(t) = \int_0^t \frac{dt}{g(t)},$$

которая ввиду $g(t) > 0$ существует, непрерывно и монотонно возрастает. Тогда соотношение (10.155) можно записать в виде

$$G(C) - G(C_0) = G(u) - G(u_0).$$

Примем $u_0 = l$. В этом случае $C_0 = \Phi(l, v) = v$ и окончательно

$$G(\Phi(u, v)) = G(u) + G(v) - G(l).$$

Следствие. Множество S способов учета неопределенности, удовлетворяющих свойствам 1—4, а также универсальности и гладкости, описывается множеством порождающих функций, удовлетворяющих условиям:

$$G(0) = 0, G(1) = 1.$$

Данное следствие вытекает из того, что по соотношению (10.148) семейство порождающих функций $\{\kappa_0 + \kappa_1 G(t)\}_{\kappa_0, \kappa_1}$, описывает одну и ту же функцию построения. Следовательно, из этого семейства достаточно включить в S лишь функцию, удовлетворяющую указанным условиям. Такая функция существует в каждом указанном семействе, так как система уравнений:

$$\begin{aligned} \kappa_0 + \kappa_1 G(0) &= 0; \\ \kappa_0 + \kappa_1 G(1) &= 1 \end{aligned}$$

всегда имеет решение ввиду строгой монотонности порождающих функций:

$$\begin{vmatrix} 1 & G(0) \\ 1 & G(1) \end{vmatrix} = G(1) - G(0) \neq 0$$

Описанное множество S весьма представительно. В частности, ему принадлежит функция построения $\Phi = u + v$, описывающая учет неопределенности в среднем (порождается функцией $G = t$); функция

$\Phi = \sqrt[s]{u^s + v^s}$ — так называемая «обобщенная средняя Эйлера»

(порождается функцией $G=t^s$); функция $\Phi=uv$, отражающая «надежностный» подход к учету неопределенности (порождается функцией $G = \ln t$). В это подмножество не входят, правда, функции $\Phi = \max (u, v)$, $\Phi = \min (u, v)$, задающие экстремальные способы учета неопределенности. Эти функции недифференцируемы при $u=v$, однако, они являются предельными для входящих в подмножество обобщенных средних Эйлера при $s \rightarrow \pm\infty$. Следует помнить, что эти две функции следует включить во множество S , так как они широко применяются при учете неопределенности. Математическим основанием для этого являются результаты работы Борисова В.И., показывающие, что требование строгой монотонности порождающей функции может быть заменено требованием монотонности.

Теорема 3 является необходимым условием для функции построения. По ее подобию можно сформулировать следующее достаточное условие.

Теорема 4. Для того чтобы функция $\Phi (u, v)$ была функцией построения, достаточно, чтобы она удовлетворяла следующему функциональному уравнению:

$$G(\Phi (u, v)) = G (u) + G (v) - G (l),$$

где $G (l)$ — непрерывная порождающая функция, имеющая обратную, а l — некоторое фиксированное число.

Доказательство. Учитывая теорему 1, данное доказательство можно свести к доказательству симметричности и ассоциативности $\Phi (u, v)$.

Ввиду существования обратной функции $t = G^{-1}(y)$ из соотношения (10.148) можно получить

$$\Phi (u, v) = G^{-1} [G (u) + G (v) - G (l)].$$

Тогда

$$\Phi (v, u) = G^{-1} [G (v) + G (u) - G (l)] = \Phi (u, v),$$

т. е. симметричность доказана.

Далее

$$G (\Phi (u, \Phi (v, z))) = G (u) + G (\Phi (v, z)) - G (l) = G(u) + G(v) + G(z) - G(l) - G(l) = G(v) + G(\Phi(u, z)) - G(l) = G(v, \Phi(u, z)),$$

откуда

$$\Phi (u, \Phi (v, z)) = \Phi (v, \Phi (u, z)),$$

и ассоциативность также доказана.

Получим явное выражение для способа учета неопределенности, порождаемого заданной функцией $G(t)$. Будем рассматривать множество неопределенностей X *двух* видов: *состоящее из конечного числа элементов N и являющееся замкнутой ограниченной областью n -мерного Евклидова пространства.*

Пусть X состоит из конечного числа элементов: $X = \{x_i\}_{i=1,2,\dots,N}$.

Рассматривая это множество как результат последовательного объединения его элементов и используя соотношение (10.142), имеем

$$F_A(X) = \Phi(f(x_1, A), \Phi(f(x_2, A), \Phi(\dots, \Phi(f(x_{N-1}, A), f(x_N, A)), \dots))).$$

Тогда, последовательно используя соотношение (10.145), получим

$$G(F_A(X)) = \Phi(f(x_1, A), \Phi(f(x_2, A), \Phi(\dots, \Phi(f(x_{N-1}, A), f(x_N, A)), f(x_N, A)), \dots)) =$$

$$= G(f(x_1, A), G(f(x_2, A), \Phi(\dots, \Phi(f(x_{N-1}, A), f(x_N, A)), \dots))) - G(l) =$$

$$= \dots, G(f(x_1, A)) + G(f(x_2, A)) + \dots + G(f(x_N, A)) -$$

$$- (N - 1)G(l) = \sum_{i=1}^N G(f(x_i, A)) - (N - 1)G(l).$$

Потребуем, чтобы пара $G(t), l$, определяющая допустимый способ учета неопределенности, удовлетворяла свойству осреднения. Это накладывает (при $f(x_i, A) = C = F_A(X)$) следующее условие:

$$G(F_A(X)) = \sum_{i=1}^N G(F_A(X)) - (N - 1)G(l),$$

откуда, ввиду строгой монотонности $G(t), l = F_A(X)$. Тогда окончательно получим

$$G(F_A(X)) = \frac{1}{N} \sum_{i=1}^N G(f(x_i, A))$$

что позволяет искомым функционал записать в виде

$$F_A(X) = G^{-1} \left(\frac{1}{N} \sum_{i=1}^N G(f(x_i, A)) \right) \quad (10.156)$$

Пусть X — замкнутая ограниченная область Евклидова пространства меры Ω . Разобьем ее на N подобластей одинаковой меры $\Delta_i = \Delta = \Omega/N$, $i = 1, \dots, N$, выберем в каждой из них произвольную точку $x_i \in X$. Полученное конечное множество точек X_N с известной точностью характеризует все множество X . Значение F_N показателя эффективности рекомендации, вычисленной с учетом X_N вместо X , удовлетворяет равенству

$$G(F, N) = \frac{1}{N} \sum_{i=1}^N G(f(x_i, A)) = \frac{1}{\Omega} \sum_{i=1}^N G(f(x_i, A)) \frac{\Omega}{N} = \frac{1}{\Omega} \sum_{i=1}^N G(f(x_i, A)) \Delta.$$

С возрастанием N множество X_N все полнее характеризует X , так что ввиду непрерывности $G(t)$

$$\lim_{N \rightarrow \infty} G(F_N) = G(F_A(X)),$$

а с учетом предыдущего равенства

$$\lim_{N \rightarrow \infty} G(F_N) = \frac{1}{\Omega} \lim_{\substack{N \rightarrow \infty \\ \Delta \rightarrow 0}} \sum_{i=1}^N G(x_i, A) \Delta = \frac{1}{\Omega} \int_X G(f(x), A) dx.$$

Итак, окончательно,

$$G(F_A(X)) = \frac{1}{\Omega} \int_X G(f(x), A) dx, \quad (10.157)$$

или

$$F(X) = G^{-1} \left(\frac{1}{\Omega} \int_X G(f(x), A) dx \right). \quad (10.158)$$

10.2.4.4. Формирование оптимального типового набора способов учета неопределенности

Базируясь на результатах пп. 10.2.4.2 и 10.2.4.3, приступим к формированию типовых наборов способов учета неопределенности, т. е. к замене множества всех допустимых порождающих функций из квадрата $0 \leq G \leq 1$, $0 \leq t \leq 1$ конечным (и небольшим) числом типовых порождающих функций. Эту замену будем производить таким образом, чтобы минимизировать ошибку, возникающую при использовании типовой порождающей функции вместо «истинной».

Формирование типовых наборов способов учета неопределенности состоит в размещении оптимальных ε -сетей в множестве допустимых порождающих функций $S = \{G(t), G(0) = 0, G(1) = 1\}$. Число элементов такой сети определяет число способов учета неопределенности, используемых ЛФР при формировании рекомендаций, а величина ε характеризует, насколько полно эти способы отражают все многообразие различных аспектов влияния неопределенности на эффективность рекомендаций.

Построение ε -сетей во множестве S возможно лишь после введения **метрики**, т. е. понятия «расстояния» Δ между двумя допустимыми способами учета неопределенности, характеризуемыми порождающими функциями $G_1(t)$ и $G_2(t)$. Эта метрика должна отражать понимание близости между способами учета неопределенности, но в то же время удовлетворять **известным аксиомам метрики: тождества, симметрии и треугольника**. Ясно, что естественной мерой близости между способами учета неопределенности является различие в оценке эффективности формируемых рекомендаций, вычисленной с их использованием. Учитывая инвариантный характер множества S , правилу определения расстояния в нем также следует придать инвариантный характер. Отсюда следует, что расстояние

должно быть определено в рамках некоторой эталонной задачи. В качестве такой задачи примем расчет n -обобщенных потерь при функции локальной эффективности $f(x)=x$ и множестве неопределенностей $X = [0, 1]$.

Самым естественным мерилом близости способов учета неопределенности в этой задаче была бы абсолютная величина разности между значениями n -обобщенных потерь, вычисленными по различным способам учета неопределенности:

$$A(G_1; G_2) = |G_1^{-1} \left(\int_0^1 G_1(x) dx \right) - G_2^{-1} \left(\int_0^1 G_2(x) dx \right)|. \quad (10.159)$$

Однако в этом случае не выполняется аксиома тождества (рис. 10.46): расхождение между различными функциями $G_1(t)$ и $G_2(t)$ может оказаться равным нулю.

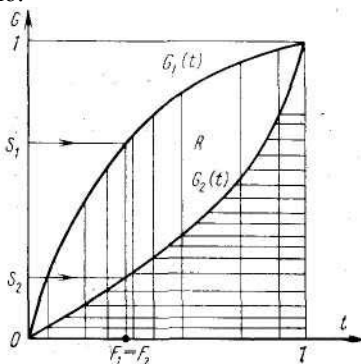


Рис. 10.46. К определению расстояния между допустимыми способами учета неопределенности

Действительно, в выражении (10.158) интегралы определяют площади S_1 и S_2 криволинейных трапеций, ограниченных графиками функций $G_1(t)$ и $G_2(t)$, а тогда n -обобщенные потери F_1 и F_2 могут совпадать. Поэтому от рассмотрения разности между «истинными» n -обобщенными потерями целесообразно перейти к рассмотрению разности «субъективных» n -обобщенных потерь $G_1(F_1)$ и $G_2(F_2)$, т.е. считать

$$\Delta(G_1, G_2) = \left| \int_0^1 G_1(x) dx - \int_0^1 G_2(x) dx \right| = \left| \int_0^1 (G_1(x) - G_2(x)) dx \right|. \quad (10.160)$$

В этом случае расстояние между порождающими функциями $G_1(t)$ и $G_2(t)$ определяется как площадь фигуры R и отлично от нуля. Однако, если допускается пересечение графиков функций $G_1(t)$ и $G_2(t)$, правило (10.159) не удовлетворяет аксиоме тождества. Для получения окончательного правила определения расстояния, удовлетворяющего всем необходимым аксиомам, примем

$$\Delta(G_1, G_2) = \int_0^1 |G_1(x) - G_2(x)| dx,$$

т. е. определим расстояние как площадь фигуры, ограниченной графиками функций $G_1(t)$, $G_2(t)$. Такое расстояние достаточно естественно: если $G_1(t) > (<) G_2(t)$ на $[0, 1]$, оно совпадает с разностью субъективных n -обобщенных потерь. В то же время, для любых $G_1(t), G_2(t) \in S$ можно записать:

$$\Delta(G_1, G_2) = 0 \Leftrightarrow G_1(t) = G_2(t),$$

$$\forall t \in [0, 1];$$

$$\Delta(G_1, G_2) = \Delta(G_2, G_1);$$

$$\Delta(G_1, G_3) \leq \Delta(G_1, G_2) + \Delta(G_2, G_3).$$

Два первых свойства очевидны. Докажем третье. Для этого запишем $|G_1(x) - G_3(x)| = |G_1(x) - G_2(x) + G_2(x) - G_3(x)| \leq |G_1(x) - G_2(x)| + |G_2(x) - G_3(x)|$, откуда

$$\int_0^1 |G_1(x) - G_3(x)| dx \leq \int_0^1 |G_1(x) - G_2(x)| dx + \int_0^1 |G_2(x) - G_3(x)| dx,$$

т. е.

$$\Delta(G_1, G_3) \leq \Delta(G_1, G_2) + \Delta(G_2, G_3).$$

При полученных условиях построение ε -сети в S сводится к задаче оптимизации стратегии гарантирующей многоцелевой системы $\langle S, S, \Delta(G_1, G_2) \rangle$. В этой многоцелевой системе множество порождающих функций является и множеством неопределенности, и множеством стратегий. Отыскиваем n -элементную стратегию $A \subset S$, обеспечивающую минимум функционалу, который определяет наибольшее расстояние от любых порождающих функций $G(t) \in S$ до ближайших к ним порождающих функций из стратегии A . Заменяем порождающие функции кусочно-линейными аппроксимациями. Для этого разобьем отрезок $[0, 1]$ оси t на $n + 1$ равную часть точками t_i , $i = 0, \dots, n + 1$, $t_0 = 0$, $t_{n+1} = 1$, $t_i > t_k$ при $i > k$ и обозначим значения типовой и «истинной» порождающих функций в этих точках соответственно через y^i и x^i , $i = 0, \dots, n + 1$:

$$y^0 = x^0 = 0, \quad y^{n+1} = x^{n+1} = 1.$$

Тогда «истинная» порождающая функция определяется n -мерным вектором $x = (x^1, \dots, x^n)$.

Допустимые «истинные» порождающие функции являются точками области X n -мерного пространства, описываемой условиями $0 \leq x^i \leq x^{i+1} \leq 1, i = 1, \dots, n - 1$. Аналогично типовые порождающие функции описываются n -мерными векторами $y = (y^1, \dots, y^n)$, принадлежащими области Y n -мерного пространства, задаваемой условиями:

$$0 \leq y^i \leq y^{i+1} \leq 1, i = 1, \dots, n - 1$$

В соответствии с моделью МСК область X является внешним множеством МСК, область Y — множеством стратегий, а искомый набор m типовых порождающих функций определяет m -элементную стратегию $A = \{y_j\}_{j=1, \dots, m} \subset Y$. В качестве функции локальной эффективности примем интегральную ошибку при замене «истинной» порождающей функции $x(t)$ на типовую $y(t)$:

$$f(x, y) = \int_0^1 |x(t) - y(t)| dt.$$

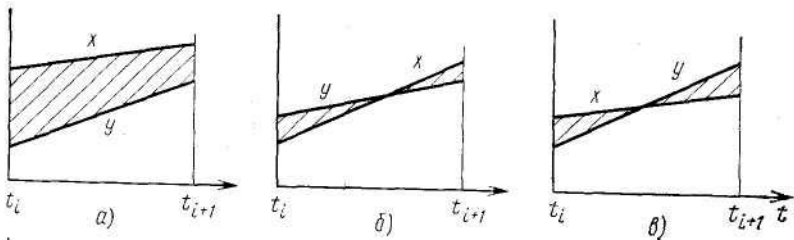
С учетом введенной аппроксимации

$$f(x, y) = \sum_{i=0}^n \int_{t_i}^{t_{i+1}} |x(t) - y(t)| dt = \sum_{i=0}^n K(x^i, y^i, x^{i+1}, y^{i+1}), \quad (10.161)$$

где

$$K(x^i, y^i, x^{i+1}, y^{i+1}) = \int_{t_i}^{t_{i+1}} |x(t) - y(t)| dt \quad (10.162).$$

Для вычисления интеграла (10.162) рассмотрим различные варианты взаимного расположения типовой и «истинной» порождающих функций на отрезке $[t_i, t_{i+1}]$ (рис. 10.47).



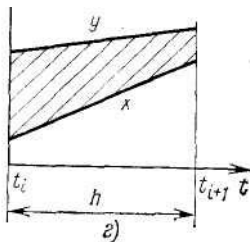


Рис. 10.47. Варианты взаимного расположения «истинной» и типовой порождающих функций:

- а) $x^i > y^i, x^{i+1} > y^{i+1}$; б) $x^i > y^i, x^{i+1} \geq y^{i+1}$;
 в) $x^i > y^i, x^{i+1} \leq y^{i+1}$; г) $x^i \geq y^i, x^{i+1} \leq y^{i+1}$

Значение вычисляемого интеграла для каждого варианта, очевидно, равно площади заштрихованной фигуры. Соответствующие выкладки проводятся достаточно просто, их результаты, определяющие $K(x^i, y^i, x^{i+1}, y^{i+1})$, сведены в табл. 10.4.

Таблица 10.4

К	$x^i > y^i$	$x^i \leq y^i$
$x^{i+1} \geq y^{i+1}$	$\frac{h}{2} (x^i + x^{i+1} - y^i - y^{i+1})$	$\frac{h}{2} \frac{(y^i - x^i)^2 + (x^{i+1} - y^{i+1})^2}{y^i - x^i + x^{i+1} - y^{i+1}}$
$x^{i+1} \leq y^{i+1}$	$\frac{h [(y^i - x^i)^2 + (x^{i+1} - y^{i+1})^2]}{2(x^i - y^i + y^{i+1} - x^{i+1})}$	$\frac{h}{2} (y^i + y^{i+1} - x^i - x^{i+1})$

Показатель эффективности МСК естественно выбрать в виде

$$f(X, A, E(x)) = \max_{j=1, \dots, m} \min_{x \in T_j} f(x, y_j), y_j \in A.$$

Тогда он оценивает гарантированную ошибку при наилучшей замене «истинной» порождающей функции на функцию из типового набора.

Итак, *задача определения типового набора функций учета неопределенности сведена к оптимизации стратегии ГМСК при функции локальной эффективности вида $\rho(x, y)$.*

Она решается общим алгоритмом оптимизации МСК. Прежде чем переходить к описанию полученных результатов, исследуем эту задачу с использованием наглядных графических приемов. Это позволяет

определить ее качественную специфику, а также получить эффективное начальное приближение и использовать в общем алгоритме оптимизации МСК быстро работающие алгоритмы улучшения и распределения.

Примем $n = 1$, полагая, что в точках $t = 0$ и $t = 1$ порождающие функции могут принимать любые значения от 0 до 1. Это относительно разбиению отрезка $[0, 1]$ оси t на три неравные части, две из которых (крайние) неограниченно уменьшаются (рис. 10.48).

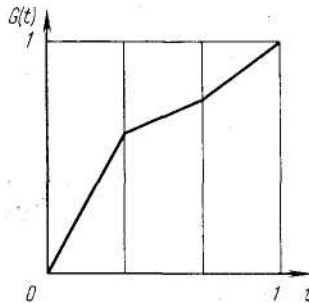


Рис. 10.48. Аппроксимация порождающей функции трехзвенной ломаной

В этом случае соотношение (10.161) совпадает с $K(x^1, y^1, x^2, y^2)$ и задача сводится к оптимальному «покрытию» треугольника, изображенного на рис. 10.49, фигурами, определяемыми линиями уровня функции $K(x^1, y^1, x^2, y^2)$ (рис. 10.50).

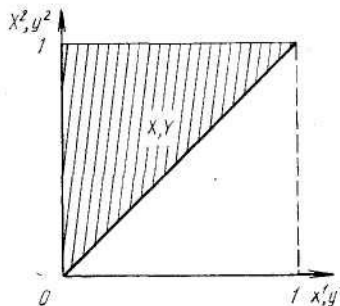


Рис. 10.49. Внешнее множество X и множество Y при формировании типового набора допустимых способов учета неопределенности (аппроксимация порождающей функции при $n = 1, 2$)

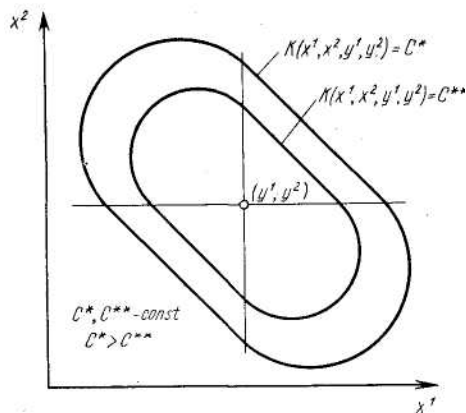


Рис. 10.50. Линия уровня функции локальной эффективности при $n = 1$

Отметим, что при деформации плоскости (x^1, x^2) — сжатии в 2 раза в направлении биссектрис 2-го и 4-го квадрантов — эти линии уровня близки к окружностям (рис. 10.51), поэтому окончательно можно рассматривать оптимальное покрытие рассматриваемого равнобедренного треугольника с гипотенузой $\sqrt{2}$ и высотой $\sqrt{2}/4$ кругами.

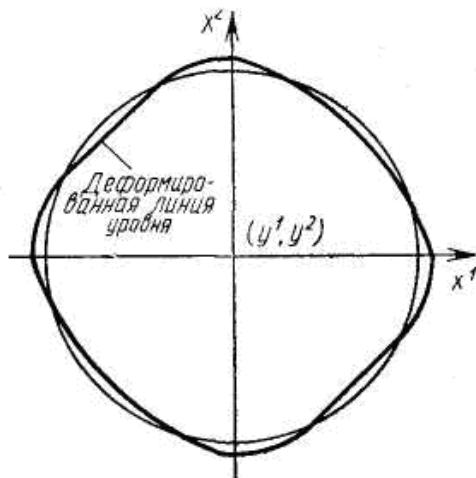


Рис. 10.51. Линия уровня функции локальной эффективности при $n = 1$ на деформированной плоскости

Результаты решения такой задачи показаны на рис. 10.52 и 10.53.

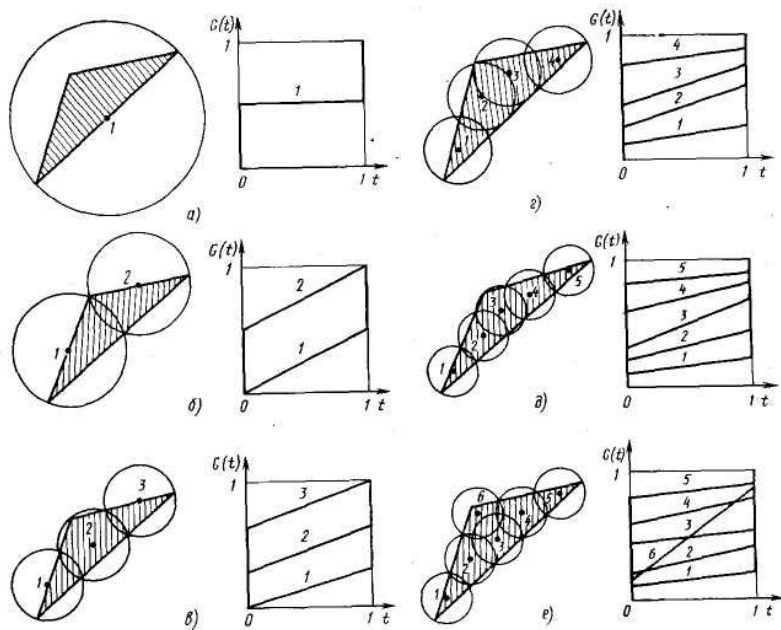


Рис. 10.52. Типовые наборы способов учета неопределенности при линейной аппроксимации порождающей функции:

а) $m = 1$; б) $m = 2$; в) $m = 3$; г) $m = 4$; д) $m = 5$; е) $m = 6$

Пусть $n = 2$, т. е. отрезок $[0, 1]$ оси разбит на три равные части. В этом случае выражение (10.161) примет вид:

$$f(x, y) = K(0, 0, x^1, y^1) + K(x^1, y^1, x^2, y^2) + K(x^2, y^2, 1, 1).$$

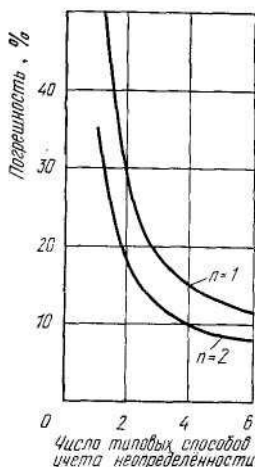


Рис. 10.53. Погрешность представления множества допустимых способов учета неопределенности типовыми наборами (при аппроксимации порождающих функций $n = 1, 2$)

Линии уровня этой функции на плоскости (x^1, x^2) при фиксированных значениях y^1, y^2 показаны на рис. 10.54 — они могут быть получены одна из другой преобразованием подобия с центром в точке (y^1, y^2) .

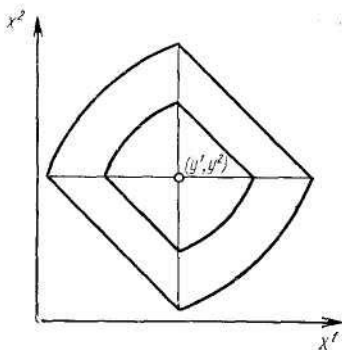


Рис. 10.54. Линии уровня функции локальной эффективности при $n=2$

Внешнее множество X в этом случае представляет собой тот же треугольник, показанный на рис. 10.49. Множество стратегий совпадает с ним, и исходную задачу можно представить как задачу

«покрытия» X заданным числом фигур, изображенных на рис. 10.54, при их минимальных размерах. Результаты решения этой задачи, полученные геометрическим путем, для числа элементов стратегии от 1 до 6 показаны на рис. 10.53 и 10.55.

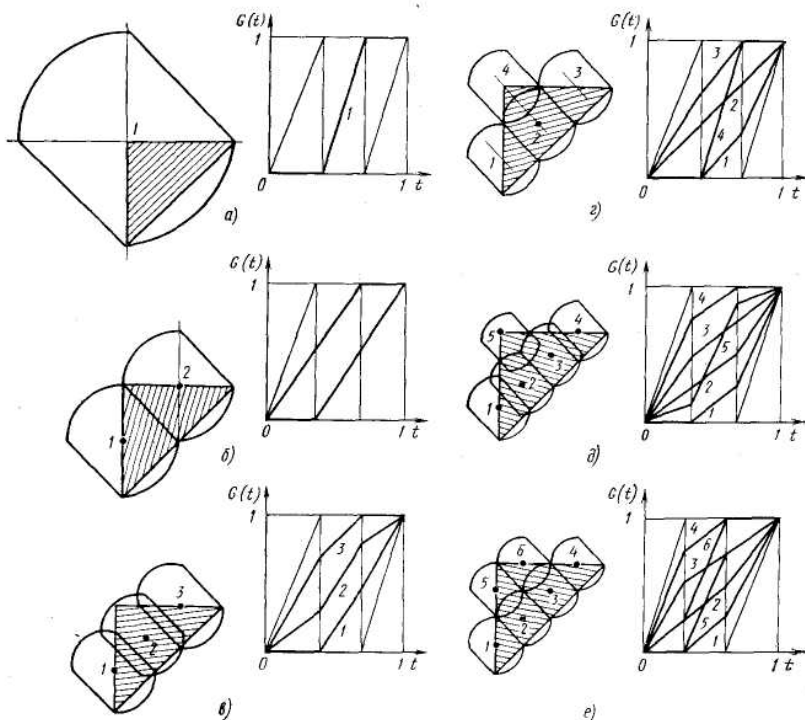


Рис. 10.55. Типовые наборы способов учета неопределенности при аппроксимации порождающих функций 3-звенными ломаными:

а) $m = 1$; б) $m = 2$; в) $m = 3$; г) $m = 4$; д) $m = 5$; е) $m = 6$

Приведенные результаты позволяют приблизительно построить характеристику возможностей типизации учета неопределенности $l(m)$ (рис. 10.53) и получить начальные приближения для эффективного использования общего алгоритма оптимизации МСК. Результаты такой типизации показаны на рис. 10.56 и на рис. 10.44.

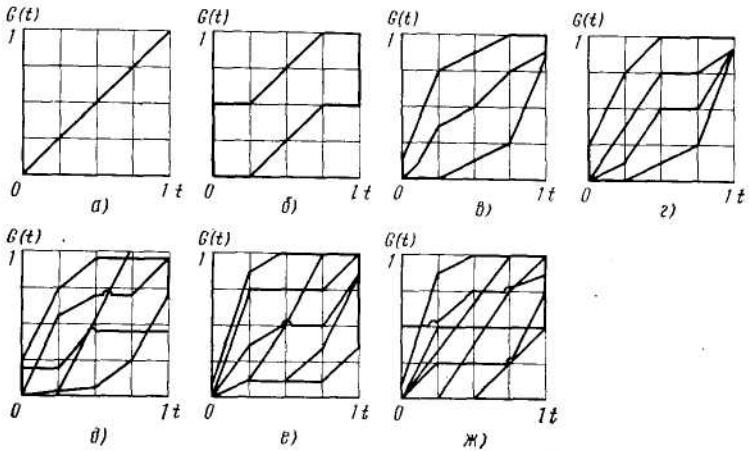


Рис. 10.56. Типовые наборы способов учета неопределенности при аппроксимации порождающих функции 4-звенными ломаными:

а) $m = 1$; б) $m = 2$; в) $m = 3$; г) $m = 4$; д) $m = 5$; е) $m = 6$; ж) $m = 7$

Заметим, что линия $p=5$ почти совпадает с линией $p=3$. Это объясняется тем, что при проведении расчетов полагалось, что в точках $t = 0$ и $t = 1$ порождающие функции могут принимать лишь соответственно значения 0 и 1, и тем самым при оптимизации типовых функций учета неопределенности опускались два «угла» квадрата $[0, 1] \times [0, 1]$ площадью 0,1 каждый. Наиболее точными мы считаем результаты, полученные для $p = 4$. В этом случае, как и при $n = 1$, допускалось, что при $t = 0$, $t = 1$ порождающие функции могут принимать любые значения от 0 до 1, т. е. отрезок $[0, 1]$ оси t делился как бы на 9 частей, из которых две крайние стягивались в точки. Линия $p = 4$ (рис. 10.44) лежит ниже линии $p = 5$ именно за счет того, что оптимизация типовых функций учета неопределенности производилась для полного внешнего множества X' . Решение для $n = 4$ и было принято как окончательное, аппроксимация которого приведена в табл. 10.3 и на рис. 10.45.

10.2.5. Расчет *n*-обобщенных потерь

10.2.5.1. Соотношения для расчета *n*-обобщенных потерь

Как указано в п.10.2.2 [см. соотношения (10.129), (10.130)] и доказано в п. 10.2.4.3, при $X \subset R_n$

$$G(F_A(X)) = \frac{1}{\Omega_X} \int G(f(x, A)) dx, \quad (10.163)$$

или

$$F_A(X) = G^{-1} \left(\frac{1}{\Omega_X} \int G(f(x, A)) dx \right). \quad (10.164)$$

Например, если в задаче формирования рекомендаций имеются независимые неопределенные параметры $x_i, i = 1, \dots, n$ и $a_i \leq x_i \leq b_i, i = 1, \dots, n$, то

$$G(F_y(X)) = \frac{1}{\prod_{i=1}^n (b_i - a_i)} \int_{a_1}^{b_1} \int_{a_2}^{b_2} \dots \int_{a_n}^{b_n} G(f(y, x_1, \dots, x_n)) \times dx_1 dx_2 \dots dx_n.$$

Формулы (10.163), (10.164) могут быть значительно упрощены, в частности, освобождены от многомерных интегралов. Преобразуем, например, формулу (10.163). Для этого разобьем отрезок $[m, M]$:

$$m = \min_{x \in X} f(x, A); \quad M = \max_{x \in X} f(x, A)$$

на N частей точками $\mu_i, \mu_{i+1} \geq \mu_i$, где $i = 0, \dots, N$. При этом обозначим

$$s_i = \int_{x \in X} dx.$$

Тогда

$$\mu_{i-1} \leq f(x, A) \leq \mu_i; \\ s^i = \bar{s}/\Omega;$$

$$G(F_A(X)) = \frac{1}{\Omega} \int_{x \in X} G(f(x, A)) dx = \frac{1}{\Omega} \sum_{i=1}^N \int_{x \in X} G(f(x, A)) dx \approx \\ \approx \frac{1}{\Omega} \sum_{i=1}^N G(\bar{f}_i(A)) s_i = \sum_{i=1}^N G(\bar{f}_i(A)) \bar{s}_i,$$

$$\mu_0 = m, \quad \mu_i = M, \quad \mu_{i-1} \leq \mu_i,$$

$$\bar{f}_i(y) \in [\mu_{i-1}, \mu_i].$$

Если $\xi_i \in [0, 1]$ — некоторая средняя точка, то

$$f_i(A) = \mu_{i-1} + \xi_i(\mu_i - \mu_{i-1})$$

и

$$G(F_A(X)) = \sum_{i=1}^N G[\mu_{i-1} + \xi_i(\mu_i - \mu_{i-1})] \tilde{s}_i,$$

Можно представить теперь вычисление $F_A(x)$ как задачу формирования рекомендаций с обобщенными потерями:

$$F(\xi, A) = G^{-1} \left\{ \sum_{i=1}^N G[\mu_{i-1} + \xi_i(\mu_i - \mu_{i-1})] \tilde{s}_i \right\}, \quad \xi = (\xi_1, \dots, \xi_N)$$

в условиях неопределенности значений ξ_i , $i = 1, \dots, N$, используя формулу (10.163). При этом необходимо воспользоваться тем же способом учета неопределенности, что и в исходной задаче, т. е. той же функцией $G(t)$. Тогда

$$\begin{aligned} G(F_A(X)) &= \frac{1}{\Omega_\xi} \int_{\xi \in X} \sum_{i=1}^n G[\mu_{i-1} + \xi_i(\mu_i - \mu_{i-1})] \bar{s}_i d\xi = \\ &= \int_0^1 \int_0^1 \dots \int_0^1 \sum_{i=1}^N G[\mu_{i-1} + \xi_i(\mu_i - \mu_{i-1})] \times \\ &\times \bar{s}_i d\xi_1 \dots d\xi_N = \sum_{i=1}^N \bar{s}_i \int_0^1 G[\mu_{i-1} + \xi_i(\mu_i - \mu_{i-1})] d\xi_i = \\ &= \sum_{i=1}^N \frac{\bar{s}_i}{\mu_i - \mu_{i-1}} \int_{\mu_{i-1}}^{\mu_i} G(t) dt. \end{aligned}$$

Здесь $\Omega_\xi = 1$ — мера N -мерного куба, $0 \leq \xi_i \leq 1$, где $i = 1, \dots, N$. Если $N = 1$, т. е. распределение значений функции локальной эффективности на множества X не учитывается, то

$$G(F_A(X)) = \frac{1}{M - m} \int_m^M G(t) dt, \quad (10.165)$$

или

$$G(F_A(X)) = \frac{I(M) - I(m)}{M - m},$$

где

$$I(t) = \int_0^t G(t) dt$$

Метод расчета n -обобщенных потерь по формуле (10.165) будем называть *Mm*-методом.

Представим выражения (10.129), (10.130) в несколько иной форме. Для этого введем в рассмотрение приведенную функцию обобщенных потерь $f^*(x, A)$ и приведенные n -обобщенные потери F^* , которые определим соотношениями:

$$f^*(x, A) = G(f(x, A)), \quad F^*_A(X) = G(F_A(X)).$$

Тогда соотношения (10.129), (10.130) примут вид

$$F^*_A(X) = \frac{1}{N} \sum_{i=0}^n f^*(x_i, A); \quad (10.166)$$

$$F^*_A(X) = \frac{1}{\Omega} \int_X f^*(x, A) dx. \quad (10.167)$$

В этом случае приведенные n -обобщенные потери рассчитывают как среднее значение приведенных обобщенных потерь по множеству неопределенности X , причем правило перерасчета обобщенных и n -обобщенных потерь в приведенные определяется способом учета неопределенности [порождающей функцией $G(t)$].

Агрегированный учет неопределенности предполагает необходимость вычисления многомерных интегралов по множеству неопределенности X от приведенных обобщенных потерь, которые рассчитываются для фиксированных порождающих функций из типового набора. Для вычисления таких интегралов на ЭВМ имеются соответствующие стандартные программы. Однако, когда различные виды неопределенности воздействуют на обобщенные потери независимо (т. е. множество X можно представить как прямое произведение множества значений неопределенных факторов различного вида), целесообразно от одновременного учета всех видов неопределенности перейти к их последовательному учету. Действительно, пусть $\Omega, \Omega_1, \Omega_2, \dots, \Omega_\kappa$ — мера множества $X, X_1, X_2, \dots, X_\kappa$ соответственно, при $X = X_1 \times X_2 \times \dots \times X_\kappa$ а $f^*(x) \equiv f^*(x_1, x_2, x_3, \dots, x_\kappa)$ ($x \in X; x_i \in X_i, i = 1, \dots, \kappa$) — непрерывная на X функция приведенных обобщенных потерь.

Тогда

$$\begin{aligned} \frac{1}{\Omega} \int_X f^*(x) dx &= \frac{1}{\Omega_1 \Omega_2 \dots \Omega_\kappa} \int_{X_1} \int_{X_2} \dots \int_{X_\kappa} f^*(x_1, x_2, \dots, x_\kappa) \times \\ &\times dx_\kappa \dots dx_1 = \frac{1}{\Omega_1} \int_{X_1} \left[\frac{1}{\Omega_2} \int_{X_2} \left[\dots \left[\frac{1}{\Omega_\kappa} \int_{X_\kappa} f^*(x_1, \dots, x_\kappa) \times \right. \right. \right. \\ &\quad \left. \left. \left. \times dx_\kappa \right] dx_{\kappa-1} \right] \dots \right] dx_1. \end{aligned}$$

В соответствии с выражением (10.167):

$$\frac{1}{\Omega_k X_k} \int f^*(x) \equiv f^*(x_1, x_2, x_3, \dots, x_k) dx_k \equiv f^*(x_1, \dots, x_{k-1}),$$

где $f^*(x_1, \dots, x_{k-1})$ — приведенные n -обобщенные потери с учетом множества неопределенности X_k . Далее,

$$\frac{1}{\Omega_{k-1} X_{k-1}} \int f^*(x_1, \dots, x_{k-1}) dx_{k-1} \equiv f^*(x_1, \dots, x_{k-2}).$$

где $f^*(x_1, \dots, x_{k-2})$ — приведенные n -обобщенные потери с учетом уже двух множеств неопределенности X_k и X_{k-1} и т. д.

Наконец,

$$\frac{1}{\Omega_1 X_1} \int f^*(x_1) dx_1 \equiv F^*$$

где F^* — приведенные n -обобщенные потери с учетом всех множеств неопределенности X_k, X_{k-1}, \dots, X_1 , т. е. в целом множества X .

При раздельном учете каждого вида неопределенности может быть в полной мере использована его специфика. Учитывая это, дадим способы расчета n -обобщенных потерь для различных видов неопределенности.

10.2.5.2. Расчет n -обобщенных потерь для основного типового набора способов учета неопределенности при различных ее видах

Неопределенность критериев. Этот вид неопределенности состоит в том, что эффективность рекомендации $y \in Y$ оценивается n показателями эффективности $\varphi^1(y), \varphi^2(y), \dots, \varphi^n(y)$. Дадим методы расчета n -обобщенных потерь для основного (семизаэлементного) набора способов учета неопределенности. Для наилучшего способа учета неопределенности

$$F(y) = \max_{s=1, \dots, n} \varphi^s(y);$$

для наилучшего

$$F(y) = \min_{s=1, \dots, n} \varphi^s(y);$$

для среднего ($\kappa = 1$), осторожного ($\kappa = 4$) и оптимистического $\kappa = 0,25$ способа учета неопределенности

$$F = \left[\frac{(n-1)!}{\prod_{i=1}^n (\kappa + i)} \sum_{s=1}^n \frac{(\varphi^s)^{\kappa+n-1}}{\prod_{i=1}^n (\varphi^s - \varphi^i)} \right]^{\frac{1}{\kappa}},$$

а для релейного ($\kappa = 0,2$) и нивелирующего ($\kappa = 5$)

$$F = 0,5 + \left[\frac{(n-1)!}{\prod_{i=1}^n (\kappa + i)} \sum_{s=1}^n \frac{(\varphi^s - 0,5)^{\kappa+n-1}}{\prod_{\substack{i=1 \\ i \neq s}}^n (\varphi^s - \varphi^i)} \right]^{\frac{1}{\kappa}}.$$

Полученные результаты сведены в табл. 10.5, используя которые можно получить n -обобщенные потери, а следовательно, строить матрицы выбора, минуя вычисление многомерных интегралов. На рис. 10.57 приведена номограмма для вычисления n -обобщенных потерь при двух критериях оптимальности ($n = 2$).

Таблица 10.5

Способ учета неопределенности	n -обобщенные потери
Наихудший	$\max_{s=1, \dots, m} \varphi_s$
Наилучший	$\min_{s=1, \dots, m} \varphi_s$
Средний	$\frac{1}{m} \sum_{i=1}^m \varphi_i$
Осторожный	$\left[\frac{(m-1)!}{\prod_{i=1}^{m-1} (i+0,25)} \sum_{s=1}^m \frac{\varphi_s^{m+3}}{\prod_{\substack{i=1, \\ i \neq s, \\ m-0,75}}^m (\varphi_s - \varphi_i)} \right]^{0,25}$
Оптимистический	$\left[\frac{(m-1)!}{\prod_{i=1}^{m-1} (i+0,25)} \sum_{s=1}^m \frac{\varphi_s^{m-0,75}}{\prod_{\substack{i=1 \\ i \neq s}}^m (\varphi_s - \varphi_i)} \right]^4$

Релейный	$0,5 + \left[\frac{(m-1)}{\prod_{i=1}^{m-1} (i+0,2)} \sum_{s=1}^m \frac{(\varphi_s - 0,5)^{m-0,8}}{\prod_{\substack{i=1 \\ i \neq s}}^m (\varphi_s - \varphi_i)} \right]^5$
Нивелирующий	$0,5 \left[\frac{(m-1)!}{\prod_{i=1}^m (5+i)} \sum_{s=1}^m \frac{(\varphi_s - 0,5)^{m+4}}{\prod_{\substack{i=1 \\ i \neq s}}^m (\varphi_s - \varphi_i)} \right]^{0,2}$

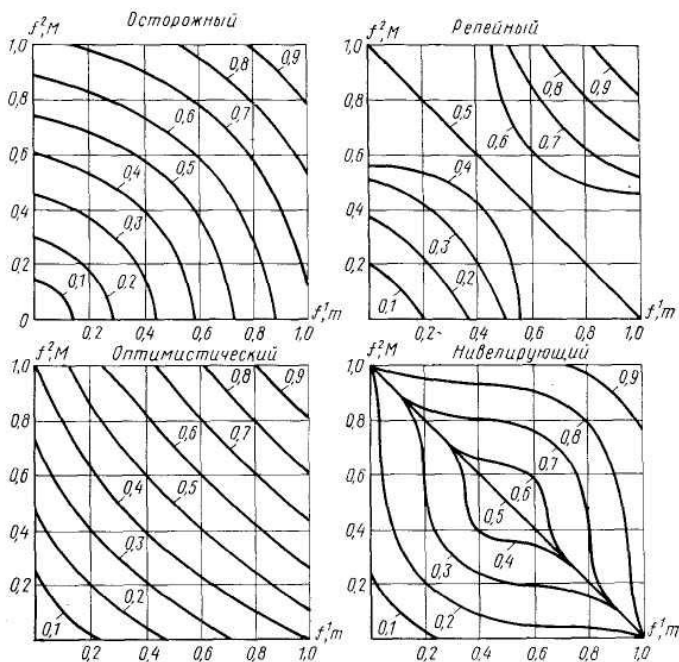


Рис. 10.57. Номограмма для вычисления n -обобщенных потерь при отдельных способах учета неопределенности по Mm — методу или при оценке рекомендации двумя критериями f^1, f^2 .

Полученные выше соотношения обладают особенностью: в случае равенства значений некоторых критериев оптимальности их знаменатели обращаются в нуль, что затрудняет их вычислительную реализацию. Поэтому рассмотрим иной (эквивалентный) метод

агрегированного учета неопределенности критериев, основанный не на линейной свертке для определения парето-оптимальных вариантов рекомендации, а на свертке Ю. Б. Гермейера.

В этом случае

$$F = G_q^{-1} \left[\sum_{j=1}^{n-1} \frac{\bar{s}_j}{\varphi^{j+1} - \varphi^j} \int_{\varphi^j}^{\varphi^{j+1}} G_q(t) dt \right],$$

где $G_q(t)$ — порождающие функции, соответствующие типовому набору способов учета неопределенности;

$$\begin{aligned} \bar{s}_i &= \frac{1}{n} \left[(i+1) \prod_{l=i+2}^n \left(1 - \frac{\varphi^l - \varphi^{i+1}}{\varphi^l} \right) - i \prod_{l=i+1}^n \right. \\ &\quad \left. \times \left(1 - \frac{\varphi^l - \varphi^i}{\varphi^l} \right) \right]_{i=2, \dots, n-1}; \\ \bar{s}_1 &= 2 \prod_{l=3}^n \left(1 - \frac{\varphi^l - \varphi^2}{\varphi^l} \right). \end{aligned}$$

Соответствующие рассматриваемому случаю значения интеграла приведены в табл. 10.6.

Таблица 10.6.

Способ учета неопределенности	н-обобщенные потери
Наихудший	M
Наилучший	m
Средний	$\frac{1}{2}(M + m)$
Осторожный	$([M^5 - m^5] / [5(M - m)])^{0,25}$
Оптимистический	$([M^{1,25} - m^{1,25}] / [1,25(M - m)])^4$
Релейный	$0,5 \left[1 + \frac{(2M - 1)^{1,2} - (2m - 1)^{1,2}}{2,4(M - m)} \right]^5$
Нивелирующий	$0,5 \left[1 + \frac{(2M - 1)^6 - (2m - 1)^6}{12(M - m)} \right]^{0,2}$

В ряде задач критерии оптимальности не являются для ЛФР однопорядковыми: одни из них представляются более важными, чем другие. При этом ЛФР не может указать количественно, насколько

одни критерии важнее других. Дадим способ агрегированного учета этого вида неопределенности критериев.

Пусть f^1, \dots, f^m — критерии, выстроенные в лексико-графическом порядке в k групп по убыванию важности: важность критериев группы с большим номером выше, чем с меньшим; внутри же группы важность не ранжирована. Группы будем задавать перечислением номеров входящих в них критериев, т. е. множествами I_1, \dots, I_k

$$\square_{j=1}^k I_j \neq \{1, \dots, m\}; I_j \cap I_s = \emptyset$$

$$j, s = 1, \dots, k, j \neq s.$$

При этом первоначально вычисляют

$$M = \max \left\{ \frac{\sum_{\mu=s+1}^k \sum_{i \in I_\mu} f^i + \max_{i \in I_s} f^i}{\sum_{\mu=s+1}^k n_\mu + 1}, \max_{i \in I_k} f^i \right\}_{s=1, \dots, k-1};$$

$$m = \min \left\{ \frac{\sum_{\mu=s+1}^k \sum_{i \in I_\mu} f^i + \min_{i \in I_s} f^i}{\sum_{\mu=s+1}^k n_\mu + 1}, \min_{i \in I_k} f^i \right\}_{s=1, \dots, k-1},$$

а затем по приближенной формуле находят n -обобщенные потери:

$$F = G^{-1}_q \left(\frac{1}{M - m} \int_m^M G_q(t) dt \right)$$

где значения интеграла берут по табл. 10.6.

Итак, алгоритмически расчет M и m для лексико-графических групп критериев осуществляют следующим образом. Для каждой группы вычисляют: сумму критериев (массив \sum , их наибольшее и наименьшее значение (массивы \max и \min), определяют число самих критериев (массив n). Эти массивы располагают по убыванию важности групп критериев. Затем вычисляют частные, где числитель есть сумма ряда последовательных членов массива \sum и следующего за последним из них члена массива \max (начиная с нуля членов массива \sum). Знаменатель есть сумма соответствующих

привлекаемым членам массива \sum членов массива n плюс 1. Из этих частных наибольшее есть M . Если в этой процедуре массив \max заменить массивом \min , то наименьшее из частных есть m .

Неопределенность модели и исходных данных. Неопределенность модели необходимо рассматривать в двух планах: широком и узком.

В широком плане это понятие является пограничным с философскими понятиями познаваемости и моделирования. В настоящей работе неопределенность модели понимается в узком плане как параметрическая неопределенность и даже более конкретно: как неточность в вычислении значений показателей эффективности в используемой при формировании рекомендаций математической модели. Учитывая, что отдельные показатели эффективности вводятся в обобщенные потери, как правило, линейной операцией, неопределенность модели описывается n -мерным вектором поправочных коэффициентов:

$$x_i, \quad i = 1, \dots, n; \quad \alpha_i \leq x_i \leq \beta_i. \quad (10.168)$$

Здесь пределы изменения коэффициентов $\alpha_i, \beta_i, i = 1, \dots, n$ характеризуют точность модели. Соотношения (10.168) задают множество неопределенностей X .

Окончательно получим следующие формулы для вычисления элементов матрицы выбора при различных способах учета неопределенности:

- наихудший

$$F_1 = \max_{j=0, \dots, 2^{n-1}} f_j ;$$

- наилучший

$$F_2 = \min_{j=0, \dots, 2^{n-1}} f_j ;$$

- средний ($\kappa = 1$), осторожный ($\kappa = 4$), оптимистический ($\kappa = 0,25$)

$$F = \left[\frac{\sum_{j=0}^{2^n-1} (-1)^n j (f|j)^{\kappa+n}}{\prod_{i=1}^n (\beta_i - \alpha_i) (\kappa + i) d_i} \right]^{1/\kappa} ;$$

- релейный ($\kappa = 0,2$), нивелирующий ($\kappa = 5$)

$$F = 0,5 + \left[\frac{\sum_{j=0}^{2^n-1} (-1)^{n_j} (f|_j - 0,5)^{\kappa+n}}{\prod_{i=1}^n (\beta_i - \alpha_i) (\kappa + i) \alpha_i} \right]^{1/\kappa}.$$

Здесь n_j — число нулей в двоичном изображении индекса j ; $f|_j$ — значение f , куда вместо i -й переменной подставлено α_i , если i -й разряд двоичного изображения числа j равен 0, или β_i , если этот разряд равен 1, $i = 1, \dots, n$. Попросту говоря, в числителях суммируются возведенные в степень $\kappa + n$ значения f в вершинах области X , взятые со знаком плюс, если левых границ в вершине четное число, и минус, если оно нечетное. Как и при учете неопределенности критериев, полученные соотношения имеют особенность: при равенстве $\alpha_i = \beta_i$ знаменатели обращаются в нуль. Однако в этом случае переменная x_i не является неопределенной и просто исключается из числа неопределенных, так что указанная особенность не осложняет вычисление n -обобщенных потерь.

В случае неопределенности исходных данных для вычисления n -обобщенных потерь необходимо использовать соотношения (10.130), (10.164) или приближенное выражение (10.165). Однако вычисление M и m при большой размерности множества X является весьма сложной задачей. Рассмотрим способ применения выражения (10.165) с учетом вычислительной неопределенности в значениях M и m , связанной с невозможностью их точного вычисления.

Пусть на множестве X выполняется неравенство

$$|f_A(x) - f_A(\xi)| \leq L \|x - \xi\| \quad \forall x, \xi \in X.$$

Если на X задана некоторая сеть, на элементах которой функция $f_A(x)$ принимает значения f_i , $i = 1, \dots, N$ и ε — диаметр X , то

$$F = (G^q)^{-1} \left(\frac{I^q(M_B) - I^q(m_B)}{M_B - m_B} \right),$$

где

$$M_B = (G^q)^{-1} \left(\frac{I^q(\min(A - L\varepsilon; 1)) - I^q(A)}{\min(A + L\varepsilon; 1) - A} \right);$$

$$m_B = (G^q)^{-1} \left(\frac{I^q(B) - I^q(\max(0; B - L\varepsilon))}{B - \max(0; B - L\varepsilon)} \right);$$

$$A = \min_{i=1, \dots, N} f_i;$$

$$B = \max_{i=1, \dots, N} f_i;$$

$$I^q(t) = \int_0^t G^q(t) dt.$$

Пусть, помимо ε , известно значение параметра ε_c для ε_c -сети, на которой вычислены значения M_B, m_B . Тогда

$$M_B = (G^q)^{-1} \left(\frac{I(\min(A + L\varepsilon; 1; B - L\varepsilon_c)) - I(A)}{\min(A + L\varepsilon; 1; B + L\varepsilon_c) - A} \right);$$

$$m_B = (G^q)^{-1} \left(\frac{I(B) - I(\max(0; B - L\varepsilon; A - L\varepsilon_c))}{B - \max(0; B - L\varepsilon; A - L\varepsilon_c)} \right).$$

10.2.6. Алгоритмы выбора формируемых рекомендаций при неопределенности внешних условий и задач функционирования консультируемой проблемы

10.2.6.1 Алгоритм точечной аппроксимации

Формирование рекомендаций при неопределенности внешних условий и задач функционирования консультируемой проблемы в математическом плане состоит в оптимизации стратегий многоцелевой системы (см. п. 10.2.4.1). Для решения этой задачи предложены аналитические и численные алгоритмы. В развитие этих решений рассмотрим ряд более эффективных алгоритмов.

Для решения задачи оптимизации стратегий интегральных МСК разработан класс «алгоритмов дискретизации», основанных на построении последовательности стратегий, сходящихся к искомой оптимальной, в соответствии со все более полным представлением множества конечным числом точек.

Будем рассматривать функции эффективности в виде $f(x, y, \mu)$ и считать их непрерывными и дифференцируемыми.

Введем конечное множество Z_N элементов внешнего множества X :

$$Z_N = \{x_i\} \subset X, \quad i = 1, \dots, N.$$

По определенному правилу будем добавлять к нему элементы X , строя последовательность множеств Z_{N+1} , Z_{N+2} и т. д.

Обозначим через ξ_i^N совокупность точек X , которая находится к элементу $x_i \in X_N$ ближе, по метрике X , чем к остальным элементам Z_N .

Очевидно, что множества ξ_i^N не пересекаются, а их объединение равно X . Меры множеств ξ_i^N обозначим через Δx_i . Правило построения множества Z_N состоит в том, что с возрастанием N максимальная величина $\max_{i=1, \dots, N} \Delta x_i$ убывает пропорционально $1/N$, а максимальный из

диаметров d_i множеств ξ_i^N — пропорционально $1/\sqrt[N]{N}$. Этого можно добиться, например, добавляя случайным образом новые элементы множества Z_N , при равномерной на X плотности вероятности выбора новых элементов. Таким образом, к

$$\max_{i=1, \dots, N} \Delta x_i \leq \kappa_1 / N; \quad \max d_i \leq \kappa_2 / \sqrt[N]{N}. \quad (10.169)$$

Рассмотрим следующую сумму:

$$S_N = \sum_{i=1}^N f(x_i, y_{E(x_i)}, \mu_{E(x_i)}) \Delta x_i, \quad (10.170)$$

где

$$\Delta x_i = \int_{\xi_i^N} dx; \quad \mu_j = \sum_{i/E(x_i)=j} \Delta x_i.$$

Можно показать, что для непрерывных функций эффективности, удовлетворяющих условию Липшица:

$$\lim_{N \rightarrow \infty} S_N = F(X, A, E(x)). \quad (10.171)$$

Действительно:

$$\begin{aligned} \Delta &\equiv |F(X, A, E(x)) - S_N| = \\ &= \left| \int_X f(x, y_E(x), \mu_E(x)) dx - \sum_{i=1}^N f(x_i, y_E(x_i), \mu_E(x_i)) \Delta x_i \right| = \\ &= \left| \sum_{i=1}^N [f(\theta_i, y_E(x_i), \mu_E(x_i)) - f(x_i, y_E(x_i), \mu_E(x_i))] \Delta x_i \right|, \end{aligned}$$

где θ_i — некоторые средние точки множеств ξ_j^N .

Ввиду выполнения условий Липшица

$$|f(\theta_i, y_E(x_i), \mu_E(x_i)) - f(x_i, y_E(x_i), \mu_E(x_i))| \leq L \|\theta_i - x_i\| \leq L d_i.$$

Тогда

$$\Delta \leq \sum_{i=1}^N |f(\theta_i, y_E(x_i), \mu_E(x_i)) - f(x_i, y_E(x_i), \mu_E(x_i))| \Delta x_i \leq \\ \leq L \sum_{i=1}^N d_i \Delta x_i.$$

Учитывая выражение (10.169), запишем

$$\Delta \leq L \sum_{i=1}^N \kappa_1 \kappa_2 \left| N^{1+\frac{1}{n}} = \kappa_1 \kappa_2 \sqrt[n]{N} \right|.$$

Теперь видно, что для любого

$$\varepsilon > 0 \text{ при } N > (\kappa_1 \kappa_2 \varepsilon)^n \Delta < \varepsilon,$$

т. е. равенство (10.171) выполняется.

Заменим решение задачи оптимизации параметров исходной МСК с внешним множеством X оптимизацией последовательностей МСК с конечными внешними множествами Z_N и критериями оптимальности (10.170).

Рассмотрим две МСК с внешними множествами Z_N и Z_{N+1} , отличающиеся наличием в последней из них дополнительного элемента x_{N+1} .

Обозначим через $A^* = \{y_j^*\}_{j=1, \dots, m}$, $E^*(x)$ решение задачи оптимизации параметров для МСК с внешним множеством Z_N .

При этом выражение (10.170) примет вид:

$$S_N = \sum_{i=1}^N f(x_i, y_{E^*(x_i)}^*, \mu_{E^*(x_i)}^*) \Delta x_i^*.$$

Пусть $A = \{y_j\}_{j=1, \dots, m}$ — решение задачи оптимизации параметров МСК с внешним множеством Z_{N+1} при заданной распределяющей функции:

$$E(x) = \begin{cases} E^*(x) & \text{при } x \in Z_N; \\ \kappa & \text{при } f(x_{N+1}, y_i^*, \mu_i^*) \leq f(x_{N+1}, y_i^*, \mu_i^*), \\ & i = 1, \dots, m; i \neq \kappa \text{ при } x = x_{N+1}. \end{cases}$$

Для этого решения, используя выражение (10.170), можно записать

$$S_{N+1} = \sum_{i=1}^N f(x_i, y_{E^*(x_i)}^*, \mu_{E^*(x_i)}^*) \Delta x_i + f(x_{N+1}, y_\kappa, \mu_\kappa) \Delta x_\kappa.$$

Ввиду оптимальности A^* и A и открытости Y

$$\frac{\partial S_N}{\partial y_j^s} = \sum_{i=1}^N \frac{\partial f(x_i, y_{E^*(x_i)}^*, \mu_{E^*(x_i)}^*)}{\partial y_j^s} \Delta x_i^* = 0:$$

$$s = 1, \dots, p, j = 1, \dots, m; \tag{10.172}$$

$$\frac{\partial S_{N+1}}{\partial y_j^s} = \sum_{i=1}^N \frac{f(x_i, y_{E^*}(x_i), \mu_{E^*}(x_i))}{\partial y_i^s} \Delta x_i + \frac{\partial f(x_{N+1}, y_R, \mu_R)}{\partial y_j^s} \Delta x_R = 0, \quad (10.173)$$

$s = 1, \dots, p, j = 1, \dots, m$

или, перегруппировав слагаемые в выражениях (10.172), (10.173) в соответствии с распределением точек $x_i, i = 1, \dots, N + 1$ по областям Дирихле, получим

$$\sum_{i|E^*(x_i)=l} \frac{\partial f(x_i, y_i^*, \mu_i^*)}{\partial y^s} \Delta x_i^* = 0; \quad (10.174)$$

$$\sum_{i|E^*(x_i)=l} \frac{\partial f(x_i, y_i, \mu_i)}{\partial y^s} \Delta x_i = 0;$$

$j = 1, \dots, m, j \neq \kappa;$ (10.175)

$$\sum_{i|E^*(x_i)=\kappa} \frac{\partial f(x_i, y_R, \mu_R)}{\partial y^s} \Delta x_i + \frac{\partial f(x_{N+1}, y_R, \mu_R)}{\partial y^s} \Delta x_R = 0. \quad (10.176)$$

Очевидно, что при $N \rightarrow \infty$

$$\|y_j - y_j^*\| \rightarrow 0, j=1, \dots, m.$$

поэтому

$$y_i = y_i^* + \Delta y_i, \quad \|\Delta y_i\| \rightarrow 0. \quad (10.177)$$

Используя дифференцируемость функции эффективности, из формул (10.174) — (10.176) найдем условия, определяющие Δy_j с точностью до бесконечно малых высшего порядка малости:

$$\sum_{i|E^*(x_i)=l} \frac{\partial f(x_i, y_i^*, \mu_j)}{\partial y^s} \Delta x_i +$$

$$+ \sum_{r=1}^p \sum_{i|E^*(x_i)=l} \frac{\partial^2 f(x_i, y_i^*, \mu_j)}{\partial y^s \partial y^r} \Delta x_i \Delta y_i^r = 0;$$

$j = 1, \dots, m, j \neq k, s = 1, \dots, p;$

(10.178)

$$\begin{aligned}
 & \sum_{i \in E^*} \frac{\partial f(x_i, y_k^*, \mu_k)}{\partial x_i} \Delta x_i + \frac{\partial f(x_{N+1}, y_k^*, \mu_k)}{\partial y^s} \Delta x_k + \\
 & + \sum_{r=1}^p \sum_{i \in E^*} \frac{\partial^2 f(x_i, y_k^*, \mu_k)}{\partial y^s \partial y^r} \Delta x_i \Delta y_k^r + \\
 & + \sum_{r=1}^p \frac{\partial f(x_{N+1}, y_k^*, \mu_k)}{\partial y^s} \Delta x_k \Delta y_k^r = 0.
 \end{aligned} \tag{10.179}$$

Услови (10.178), (10.179) дают необходимое число уравнений для определения компонент Δy_j^r , перемещений элементов ряда A . При этом система разбивается на m независимых подсистем, каждая из которых задает вектор перемещения соответствующих элементов y_j .

Из уравнений (10.178) видно, что при $j \neq k$ величина перемещения будет бесконечно малой высшего порядка малости относительно Δy_k .

Действительно, Δy_j^r , определяемые из уравнений (10.178), имеют тот же порядок малости, что и изменения Δx_i , так как при $\mu_j = \mu_j^*$ и $\Delta x_i = \Delta x_i^*$ имеют лишь тривиальное нулевое решение [поскольку свободный член обращается в нуль, в силу условий (10.174)]. Система (10.179) при $\mu_k = \mu_k^*$ и $\Delta x_i = \Delta x_i^*$ имеет ненулевое решение ввиду наличия свободного члена $df(x_{N+1}, y_k^*, \mu_k^*) \Delta x_k / dy^s$. По той же причине первое слагаемое в уравнении (10.178) является бесконечно малой высшего порядка малости относительно второго слагаемого.

Итак, с точностью до бесконечно малых высшего порядка можно считать, что

$$\begin{aligned}
 y_j &= y_j^*, j=1, \dots, m, j \neq k; \\
 y_k &= y_k^* + \Delta y_k,
 \end{aligned}$$

где Δy_k — решение системы уравнений;

$$\begin{aligned}
 & \sum_{r=1}^p \left[\frac{\partial f(x_{N+1}, y_k^*, \mu_k^*)}{\partial y^s} \Delta x_k^* + \sum_{i \in E^*} \frac{\partial^2 f(x_i, y_k^*, \mu_k^*)}{\partial y^s \partial y^r} \Delta x_i^* \right] \Delta y_k^r = \\
 & = - \frac{\partial f(x_{N+1}, y_k^*, \mu_k^*)}{\partial y^s} \Delta x_k^*, \quad j = 1, 2, \dots, m; \quad j \neq k; \quad s = 1, 2, \dots, p.
 \end{aligned}$$

Последовательность действий в рассматриваемом алгоритме следующая. Начальная система центров $\{y_{ij}^0\}$ задается произвольной, точку x_{n+1} выбираем следующим образом. Используя датчик

случайных чисел, получим n равномерно распределенных случайных чисел $\xi_1, \xi_2, \dots, \xi_n$. Далее примем

$$x_{n+1}^j = a^j + \xi_i(b^j - a^j), \quad i=1, \dots, n.$$

Распределяющую функцию $E(x)$ и связанную с этим целочисленную функцию $j(i)$ вычислим путем сравнения значений функции $r(x_i, y_j)$ для различных центров y_j и выбора номера центра, соответствующего наименьшему значению функции. Аналогично определим меры областей Дирихле точек x_1, \dots, x_{n+1} .

Произведем перебор равномерно распределенной в X системы точек и подсчитаем число попадающих в область Дирихле различных точек x_i . Приращения координат элементов $\Delta y_j^s, s=1, \dots, p; i=1, \dots, m$ вычислим путем решения соответствующих систем линейных уравнений.

Целесообразно предусмотреть два условия окончания счета: по значению оптимизируемой функции и по приращению координат центров:

$$|f_{n+1} - f_n| \leq \varepsilon_f, \quad \max_{j=1, \dots, m} \|y_j - y'_j\| \leq \varepsilon_y.$$

Поскольку сложность реализации алгоритма на ЭВМ практически не возрастает с увеличением размерности множеств X и Y (увеличивается лишь время счета), соответствующие программы достаточно просто разрабатываются на случай больших размерностей.

10.2.6.2. Алгоритм абсолютной оптимизации

Наряду с приведенными выше алгоритмами, область применимости которых ограничена специальными условиями, необходима разработка надежно действующего универсального алгоритма, гарантирующего отыскание абсолютного оптимума, который предъявлял бы минимальные требования к свойствам множеств X и Y и функции локальной эффективности.

Алгоритм оптимизации МСК, удовлетворяющий перечисленным требованиям, описан ниже. В его основу положена идея о том, что минорируемость (в частности, липшицевость) является достаточно необременительным свойством функции, позволяющим строить эффективные алгоритмы оптимизации. Используя эту идею для общего алгоритма оптимизации МСК, возможно: во-первых, распространить конкретные оценки и вычислительные схемы на случай многоцелевой системы, являющейся обобщением обычной задачи оптимизации; во-вторых, поставить надежность отыскания абсолютного оптимума в зависимость от вычислительных ресурсов, так что если последние полностью соответствуют требуемым, то отыскание абсолютного

оптимума гарантируется, если же они меньше, то вычислительный процесс доводится до конца, однако есть вероятность, что полученное решение соответствует лишь относительному оптимуму.

Прежде чем перейти к описанию общего алгоритма, напомним, что задача оптимизации МСК (с некоторыми упрощениями) заключается в следующем. Даны множества X и Y и определенная на подмножестве J их прямого произведения функция $\rho(x, y) \geq 0$, такая, что любое сечение множества J по x непусто (обозначим сечение J по x через $J(x)$, а J по y — через $J(y)$):

$$J(x) \neq \emptyset, \quad \forall x \in X.$$

Кроме того, задано целое число m . Требуется найти такое m -элементное подмножество $A = \{Y_{i=1}, \dots, m\} \subset Y$ (стратегию) и такую определенную на X распределяющую функцию $E(x)$, имеющую область значений $\{1, 2, \dots, m\}$ и удовлетворяющую условию

$$x \in J(Y_{E(x)}) \quad \forall x \in X$$

которые доставляют минимум функционалу

$$F(X, A, E(x)) = \sum_{x \in X} \rho(x, y_{E(x)}). \quad (10.180)$$

или

$$F(X, A, E(x)) = \max_{x \in X} \rho(x, y_{E(x)}). \quad (10.181)$$

В первом случае говорится об интегральной многоцелевой системе, а суммирование понимается в смысле метрики X , во втором случае — о гарантирующей многоцелевой системе.

Как было нами установлено, распределяющая функция оптимальной МСК при стратегии A задается условием

$$\begin{aligned} E(x) = k / \rho(x, y_k) \leq \rho(x, y_u), \\ x \in J(y_k), \quad x \in J(y_u), \quad y_k, y_u \in A. \end{aligned} \quad (10.182)$$

Поэтому выражения (10.180), (10.181) могут быть представлены в виде

$$F_A = \sum_{x \in X} L_A(x), \quad (10.183)$$

или

$$F_A = \max_{x \in X} L_A(x), \quad (10.184)$$

где

$$L_A(x) = \min_{\substack{y_j \in A \\ x \in J(y_j)}} \rho(x, y_j). \quad (10.185)$$

Таким образом, общий алгоритм оптимизации МСК направлен на решение следующей задачи. Даны ограниченные замкнутые множества

X и Y векторов конечно-мерных пространств размерности n и p соответственно, числа $m, \varepsilon > 0$ и определенная на J функция $\rho(x, y) \geq 0$, удовлетворяющая условию Липшица:

$$|\rho(x_2, y_2) - \rho(x_1, y_1)| \leq L_X \|x_2 - x_1\| + L_Y \|y_2 - y_1\|;$$

$$\forall x_1, x_2 \in X, \quad \forall y_1, y_2 \in Y. \tag{10.186}$$

В условии (10.186) норма вычисляется в чебышевской метрике, т. е. если

$$x_k = (x_k^i)_{i=1, \dots, n}, \quad y_k = (y_k^s)_{s=1, \dots, p},$$

то

$$\left. \begin{aligned} \|x_2 - x_1\| &= \max_{i=1, \dots, n} |x_2^i - x_1^i|; \\ \|y_2 - y_1\| &= \max_{s=1, \dots, p} |y_2^s - y_1^s|. \end{aligned} \right\} \tag{10.187}$$

Требуется найти m -элементную стратегию $A = \{y_j\}_{j=1, \dots, m} \in Y$, доставляющую с ошибкой не более ε минимум функций (10.183) (для ИМКС) или (10.184) (для ГМСК).

Решение этой задачи всегда существует. Действительно, ее можно рассматривать как задачу оптимизации функции F_A на множестве Y' элементов A — прямом произведении m множеств Y :

$$Y' = \prod_{m} Y$$

Так как Y ограничено и замкнуто, то Y' также ограничено и замкнуто. Функция F_A непрерывна на Y , поскольку получена суперпозицией конечного числа непрерывных функций (10.183)—(10.185), а функция $\rho(x, y)$ — липшицева, т.е. непрерывна. Тогда по теореме Вейерштрасса F_A принимает на Y' свое наименьшее значение. Соответствующий элемент $A \in Y'$ является решением задачи при любом ε .

Основываясь на общем подходе к отысканию абсолютного минимума функции, изложим идею алгоритма абсолютной оптимизации МСК. Представляя собой замкнутые ограниченные множества в метрических пространствах, X и Y компактны. Следовательно, в них могут быть размещены конечные сети с параметрами $\varepsilon_x, \varepsilon_y$ соответственно. Обозначая такого рода сети через $X_{\varepsilon_x}, Y_{\varepsilon_y}$, можно оптимизировать МСК с конечными множествами $X_{\varepsilon_x}, Y_{\varepsilon_y}$, и функцией локальной эффективности $\rho(x, y)$, определенной в их прямом произведении. При надлежащем выборе $\varepsilon_x, \varepsilon_y$ результат оптимизации будет отличаться от оптимальной стратегии МСК с множествами X и Y

не более чем на ε , следовательно, явится решением поставленной задачи.

По всей видимости, число элементов $X_{\varepsilon_x}, Y_{\varepsilon_y}$ окажется значительным, это определяет большие требования к объему памяти и трудоемкости выполнения алгоритма. Их можно уменьшить, если учесть, что нас интересуют лишь те элементы сетей, которые оказывают влияние на оптимальное решение. Они могут быть выявлены с использованием функции G , а точнее сказать, оценок, основанных на липшицевости $\rho(x, y)$ и конкретной структуре F_A , после чего остальные элементы сетей могут быть опущены или агрегированы.

Из соображений экономии вычислительных ресурсов процесс «отсеивания» элементов сетей целесообразно организовать последовательно. Вначале строятся сети с большими значениями параметра, оптимизируется соответствующая им МСК с конечными множествами и на основе результатов оптимизации вычисляются оценки влияния отдельных элементов сетей на искомое оптимальное решение. Полученные оценки используются для перехода к новым сетям с меньшим значением параметра, уже не содержащим элементов, в известном смысле близких к заведомо не влияющим на решение. Эту операцию назовем *модификацией*. Затем снова оптимизируется МСК с конечными множествами для модифицированных сетей и так далее, пока не будет получено решение для сетей с параметрами, не превосходящими $\varepsilon_x, \varepsilon_y$.

Структура описанного алгоритма показана на рис. 10.58.

Перейдем к проработке отдельных его частей. Несложно получить оценки отличия оптимальных решений исходной МСК с множествами X и Y и показателем эффективности F''_A от вспомогательной МСК с замкнутыми ограниченными множествами $X^* \subset X, Y^* \subset Y$ и показателем эффективности E_A при одинаковой функции локальной эффективности $\rho(x, y)$. Обозначим соответственно оптимальные стратегии через \tilde{A} и A^* , а оптимальные значения показателей эффективности — через \tilde{F} и F^* . Положим для простоты $J = X \times Y$. Рассмотрим ГМСК. Значение показателя эффективности как исходной, так и вспомогательной МСК будем рассчитывать по формуле (10.184). Таким образом,

$$\tilde{F} = \max_{x \in X} L_{\tilde{A}}(x), F^* = \max_{x \in X} L_{A^*}(x).$$

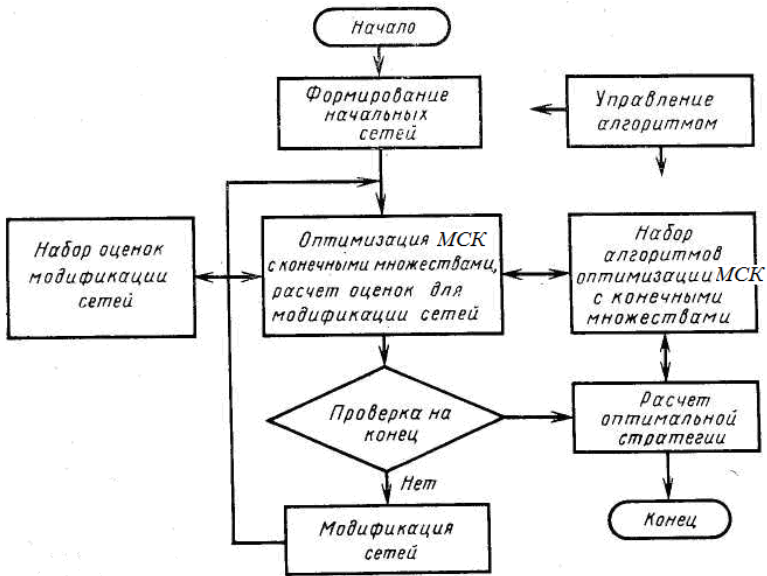


Рис. 10.58. Структура общего алгоритма оптимизации МСК

Для оптимального значения показателя эффективности ГМСК получена оценка:

$$F^* - L_Y \delta_Y \leq \tilde{F} \leq F^* + L_X \delta_X. \quad (10.188)$$

Следовательно, оптимизируя вместо исходной ГМСК вспомогательную ГМСК с конечными множествами X^* , Y^* , которые являются сетями соответственно с параметрами δ_x , δ_y , можно принять полученную стратегию A^* за оптимальную для исходной ГМСК, допустив при этом проигрыш A :

$$\Delta \leq \max_{x \in X} L_{A^*}(x) - F^* + L_Y \delta_Y, \quad (10.189)$$

но очевидно, что

$$\max_{x \in X} L_{A^*}(x) \leq \max_{x \in X^*} L_{A^*}(x) + L_X \delta_X = F^* + L_X \delta_X$$

Поэтому

$$\Delta \leq L_X \delta_X + L_Y \delta_Y. \quad (10.190)$$

Обозначив через Ω меру множества X :

$$\Omega = \sum_{x \in X} 1,$$

а через ω_s — меру области Дирихле D_s :

$$\omega_s = \sum_{x \in D_s} 1, \quad \sum_{s=1}^N \omega_s = \Omega, \quad (10.191)$$

окончательно получим

$$\left| F^H_A - \sum_{s=1}^N L_A(\xi_s) \omega_s \right| \leq L_X \epsilon \Omega$$

Таким образом, целесообразно ввести показатель эффективности вспомогательной МСК в виде

$$F^B_A = \sum_{s=1}^N L_A(\xi_s) \omega_s \quad (10.192)$$

Тогда для показателей эффективности исходной и вспомогательной МСК будет иметь место оценка

$$|F^H_A - F^B_A| \leq L_X \epsilon \Omega,$$

где ϵ — параметр ϵ -сети X^* в X . Используя его прежнее обозначение, получим

$$|F^H_A - F^B_A| \leq L_X \delta_x \Omega. \quad (10.193)$$

После простых преобразований можно показать, что

$$F^* - (L_X \delta_x + L_Y \delta_y) \Omega \leq \tilde{F} + L_X \delta_x \Omega. \quad (10.194)$$

Таким образом, оптимизируя вместо исходной МСК вспомогательную с конечными множествами X^* и Y^* (сетями в X и Y с параметрами δ_x , δ_y) и показателем эффективности (10.192), можно принять полученную стратегию A^* за оптимальную для исходной ИМСК; при этом проигрыш по показателю эффективности A ограничен значением

$$\Delta \leq \sum_{x \in X} L_{A^*}(x) - F^* + (L_X \delta_x + L_Y \delta_y) \Omega. \quad (10.195)$$

Но из условия (10.193) при $A = A^*$ имеем

$$F^H_{A^*} = \sum_{x \in X} L_{A^*}(x) \leq F^B_{A^*} + L_X \delta_x \Omega \equiv F^* + L_X \delta_x \Omega$$

Учитывая полученное соотношение, неравенство (10.195) можно усилить:

$$\Delta \leq (2L_X \delta_x + L_Y \delta_y) \Omega \equiv \Delta^*. \quad (10.196)$$

Оценки, полученные для ГМСК (10.190) и ИМСК (10.196), позволяют найти параметры сетей X^* , Y^* , δ_x , δ_y , обеспечивающие отыскание оптимальной стратегии с заданной точностью по показателям эффективности ϵ . Эти параметры удовлетворяют условию

$$\Delta^* \leq \epsilon. \quad (10.197)$$

Используя оценки (10.190), (10.196), исследуем возможность отсеивания или агрегирования элементов ε -сетей при их модификации.

Рассмотрим ГМСК. Пусть произведена оптимизация вспомогательной ГМСК при параметрах сетей $\varepsilon_x, \varepsilon_y$. Из выражения (10.170) следует, что пары (x, y) , для которых

$$\rho(x, y) > F^* + L_X \delta_x,$$

не влияют на оптимальную стратегию исходной ГМСК. Следовательно, если

$$\min_{x \in X} \rho(x, y) > F^* + L_X \delta_x, \quad (10.198)$$

то точка $y \in X$ может быть исключена из рассмотрения.

Обозначим через A_y m -элементную стратегию, содержащую y . Если

$$\min_{A_y \in X} F_{A_y} > F^* + L_X \delta_x, \quad (10.199)$$

то точка $y \in X$ может быть исключена из рассмотрения.

Если

$$\begin{aligned} \max_A L_A(x) < F^* - L_Y \delta_y; \\ F'_A \leq F^* + L_X \delta_x \end{aligned} \quad (10.200)$$

то точка $x \in X$ также может быть исключена из рассмотрения.

Условия (10.198)— (10.200) могут быть использованы при модификации сетей X^*, Y^* (при замене более «густыми» сетями), но для этого они должны быть определены в терминах этих сетей.

Пусть ε, η — точки сетей X^*, Y^* соответственно. Можно показать, что если

$$\begin{aligned} \max_{A \subset Y^*} L_A(\xi) \leq F^* - L_X \delta_x - 2L_Y \delta_y; \\ F_A \leq F^* + L_X \delta_x + 2L_Y \delta_y, \end{aligned} \quad (10.201)$$

то из рассмотрения можно исключить точку $\xi \in X^*$ с ее δ_x -окрестностью в X , а если

$$\min_{\xi \subset X^*} \rho(\varepsilon, \eta) > F^* + 2L_X \delta_x + L_Y \delta_y \quad (10.202)$$

или

$$\min_{A_\eta \subset Y^*} F_{A_\eta}^B > F^* + 2L_X \delta_x + L_Y \delta_y,$$

то из рассмотрения можно исключить точку $\eta \in Y^*$ с ее δ_y -окрестностью в Y .

Перейдем к рассмотрению ИМСК. Если $\xi \in X^*, \eta \in Y^*$ и

$$\min_{s=1, \dots, N} \omega_s \min_{\xi \subset X^*} \rho(\varepsilon, \eta) > F^* + L_X \delta_x \Omega + (L_X \delta_x + L_Y \delta_y), \quad (10.203)$$

то элемент $\eta \in Y^*$ с его δ_y -окрестностью в Y можно исключить из рассмотрения.

Нужно отметить, что условие (10.203) отсеивания элементов довольно «грубое», так как при уменьшении параметра сети δ_y происходит уменьшение $\min_{s=1, \dots, N} \omega_s$, т. е. выполнение условия (10.203)

становится все более затруднительным. Этот недостаток отсутствует у следующего условия, являющегося аналогом условия (10.202).

Если $\xi \in X^*$, $\eta \in Y^*$ и

$$\min_{A_\eta \in Y^*} F^B_{A_\eta} > F^* + (2L_X\delta_x + L_Y\delta_y)\Omega, \quad (10.204)$$

то элемент $\eta \in Y^*$ с его δ_y -окрестностью в Y можно исключить из рассмотрения.

Для ИМСК невозможно исключить из рассмотрения элементы X , как это делается для ГМСК, однако здесь существует возможность их агрегирования. Если некоторая область $D \subset X$ заведомо входит в одну область Дирихле оптимального решения, нет необходимости размещать в ней несколько элементов сети, а достаточно включить ее в рассмотрение в виде одного элемента, приняв

$$\rho(\tilde{\xi}, y) = \frac{1}{\omega_D} \sum_{x \in D} \rho(x, y) \quad (10.205)$$

где ω_D — мера этой области.

Можно показать, что если выполнено условие

$$|\rho(\xi_s, y_k) - \rho(\xi_s, y_p)| > 2(L_X\delta_x + L_Y\delta_y) \forall y_k, y_p \in \tilde{Y}, \quad (10.206)$$

то точка ξ_s агрегирует ее δ_x -окрестность в X , а значение $\rho(\xi_s, y)$ вычисляется по формуле (10.205) при $\tilde{\xi} = \xi_s$. Отметим, что для выполнения условия (10.206) необходимо

$$L_Y \|y_p - y_k\| > 2(L_X\delta_x + L_Y\delta_y) \forall y_p, y_k \in \tilde{Y}. \quad (10.207)$$

Действительно, если $L_Y \|y_p - y_k\| \leq 2(L_X\delta_x + L_Y\delta_y)$, то, ввиду неравенства (10.186)

$$|\rho(\xi_s, y_k) - \rho(\xi_s, y_p)| \leq L_Y \|y_p - y_k\|,$$

условие (10.206) выполняться не будет.

Полученные оценки используются при выборе параметров сетей X^* , Y^* , обеспечивающих заданную точность оптимизации МСК, а также для уменьшения размерности задачи при переходе к сетям с меньшими значениями параметров (модификации сетей).

Можно указать оценки для взаимного расположения элементов оптимальной стратегии. Обозначим через \tilde{F}_m оптимальное значение показателя эффективности МСК с m -элементной стратегией.

Пусть A_m — некоторая m -элементная стратегия, причем каждые q ее элементов ($q = 2, \dots, m$), например y_1, y_2, \dots, y_q , отстоят друг от друга более чем на z :

$$\|y_j - y_s\| \geq z, \quad j, s = 1, \dots, q.$$

Рассмотрим $m - q + 1$ -элементную стратегию A_v , $v = m - q + 1$, полученную из A_m отбрасыванием $q - 1$ элемента y_1, \dots, y_{q-1} .

Тогда для ГМСК

$$F_{A_m} = \sum_{x \in X} L_{A_m}(x) \geq \sum_{x \in X} L_{A_v}(x) - L_Y z = F_{A_v} - L_Y z$$

Аналогично для ИМСК

$$F_{A_m} = \sum_{x \in X} L_{A_m}(x) \geq \sum_{x \in X} L_{A_v}(x) - L_Y z \Omega = F_{A_v} - L_Y z \Omega.$$

Итак, вообще для МСК

$$F_{A_m} \geq F_{A_v} - \lambda \geq \tilde{F}_v - \lambda,$$

где

$$\lambda = \begin{cases} L_Y z & \text{для ГМСК} \\ L_Y z \Omega & \text{для ИМСК} \end{cases}$$

Если A_m — оптимальная стратегия, то $F_{A_m} = \tilde{F}_m$.

Таким образом, для того чтобы m -элементная стратегия, q элементов которой отстоят друг от друга не более чем на z , была оптимальна, должно соблюдаться условие

$$\lambda \geq \tilde{F}_{m-q+1} - \tilde{F}_m; \quad \lambda = \begin{cases} L_Y z & \text{для ГМСК} \\ L_Y z \Omega & \text{для ИМСК} \end{cases} \quad (10.208)$$

Полученное условие (10.208) позволяет ограничить переборы при проверке различных стратегий из Y на оптимальность. Заметим, что ввиду очевидного неравенства $\tilde{F}_m \geq \tilde{F}_k$ при $k > m$ условие (10.208) можно усилить следующим образом.

Если z таково, что

$$\lambda < \tilde{F}_{m-q+1} - \tilde{F}_m, \quad (10.209)$$

в частности,

$$\lambda < \tilde{F}_{m-q+1} - \tilde{F}_{m-1}, \quad \lambda = \tilde{F}_{m-q+1} - F_A$$

(где A — любая m -элементная стратегия), то стратегия, из m элементов которой q расположены не далее чем на расстоянии z друг от друга, неоптимальна.

Если положить в условии (10.208) $q = m$, то оказывается, что если z таково, что $\lambda < \tilde{F}_1 - \tilde{F}_m$, в частности,

$$\left. \begin{aligned} \lambda < \tilde{F}_1 - \tilde{F}_{m-1} \\ \lambda < \tilde{F}_1 - F_A \end{aligned} \right\} \quad (10.210)$$

(где A — любая m -элементная стратегия), то элементы m -элементной оптимальной стратегии отстоят друг от друга на расстоянии не меньшем z .

Если положить $q=2$ и воспользоваться условием (10.209), можно получить оценку для целесообразного значения δ_y -параметра сети Y^* . Действительно, если некоторый элемент сети входит в оптимальную стратегию, то в нее не может войти элемент, отстоящий от него менее чем на z , определяемое условием:

$$\lambda < \tilde{F}_{m-1} - \tilde{F}_m,$$

Поэтому нецелесообразно выбирать $\delta_y < z/2$.

Наконец, получим оценку для сравнения показателей эффективности МСК двух близких стратегий. Пусть $A = \{y_i\}_{i=1, \dots, m}$, а стратегия A' отличается от A лишь последним элементом η , причем

$$\|y_m - \eta\| \leq \psi.$$

Тогда имеем:

для ГМСК

$$F_A \geq F_{A'} - L_Y \psi;$$

для ИМСК

$$F_A \geq F_{A'} - L_Y \psi \Omega.$$

Отсюда видно, что если в стратегиях $A, A'_q, q \leq m$ совпадает $m - q$ элементов, а остальные q соответственно отличаются друг от друга не более чем на ψ_1, \dots, ψ_q , то

$$\lambda < \beta(F_{m-1} - F_A),$$

где

$$\beta = \begin{cases} L_Y \sum_{i=1}^q \psi_i & \text{для ГМСК} \\ L_Y \sum_{i=1}^q \psi_i \Omega & \text{для ИМСК} \end{cases} \quad (10.211)$$

Полученные оценки позволяют с заданной точностью переходить от оптимизации МСК с бесконечными множествами X и Y к оптимизации МСК с конечными множествами X_K, Y_K :

$$X_K = \{x_s\}_{s=1, \dots, N}$$

$$Y_K = \{y_l\}_{l=1, \dots, n}$$

Для оптимизации такой МСК разработано несколько алгоритмов:

- комбинаторный с ограничением перебора;
- оценок на основе линейного программирования;
- распределения;
- улучшения.

Они отличаются временем счета и надежностью отыскания абсолютного оптимума, поэтому их совместное использование в едином программном комплексе позволяет решить задачу оптимизации МСК, наиболее эффективно используя располагаемые вычислительные ресурсы. Опишем каждый из них.

10.2.6.3. Комбинаторный алгоритм оптимизации с ограничением перебора

Допустим, что даны число элементов оптимизируемой стратегии m и число z , которые, используя условия (10.210), определяют минимальную взаимную близость элементов в оптимальной стратегии. Обозначим

$$G(y_p) = \{y_i : \|y_i - y_p\| < z, p=1, \dots, M, p \neq i, i=1, \dots, M\}.$$

Тогда оптимальными могут оказаться лишь стратегии $A = \{y_j\}_{j=1, \dots, m}$, удовлетворяющие условию

$$A \cap \left(\bigcap_{j=1}^m G(y_j) \right) = \emptyset$$

Используя известные свойства операций над множествами, получим

$$\left. \begin{aligned} \bigcup_{l=1}^m (A \cap G(y_l)) = \emptyset, \quad y_l \in A; \\ A \cap G(y_j) = \emptyset, \quad y_j \in A; \\ y_s \in G(y_j) \quad \forall y_j, y_s \in A. \end{aligned} \right\} \quad (10.212)$$

Первое условие в (10.212) порождает следующий эффективный алгоритм перебора m -элементных стратегий, содержащих оптимальную стратегию. При этом рассматривается множество номеров $I_l = \{1, 2, \dots, M\}$. Один из его элементов j_l принимается за номер первого элемента стратегии. Затем из I_l исключаются номера j такие,

что

$$G(y_j) \ni j_i; j \leq j_i.$$

Второе условие в (10.212) служит для того, чтобы исключить из перебора стратегии, отличающиеся лишь порядком следования элементов. В перебираемых стратегиях номер в Y последующего элемента всегда больше, чем у предыдущего. Тем самым, I_1 преобразуется в I_2 . Один из элементов I_2 — j_2 , принимается за номер второго элемента стратегии. Из I_2 исключаются номера j такие, что

$G(y_j) \ni j_2$, тем самым формируется I_3 и т. д. до получения I_m , т. е.

$$I_1 = \{1, 2, \dots, M\};$$

$$I_{k+1} = I_k \setminus \{j: G(y_j) \ni j_k; j \leq j_k; j_k \in I_k\}_{k=1 \dots m-1};$$

$$A = \{y_{j_s}, j_s \in I_s, s = 1, \dots, m.$$

Если $I_k = \emptyset$, $k \leq m$, то осуществляется выбор иного элемента в I_{k-1} . Когда это невозможно (все элементы уже перебраны), процесс прекращается.

Если $I_m \neq \emptyset$, то каждый его элемент является m -м элементом стратегии A . После перебора из всех выбирается новый элемент I_{m-1} в качестве $m - 1$ -го элемента стратегии и т. д. до перебора всех элементов I_1 . Для каждой сформированной таким образом стратегии A подсчитывается F_A , и таким образом определяется оптимальная стратегия A :

$$F \equiv F_A \equiv \min_A F_A$$

а заодно подсчитываются оценки (10.201), (10.202), необходимые для последующей модификации сетей $X_{k\kappa}$, $Y_{k\kappa}$.

Для еще большего ограничения перебора используют условие (10.203). Пусть F_T — текущая верхняя оценка \bar{F} , например, наименьшее из уже подсчитанных значений F_A , а A_1 — некоторая сформированная алгоритмом стратегия. Для того чтобы стратегия A , q членов которой не более чем на ψ_i ($i=1, \dots, q$) отличаются от соответствующих членов A_1 , не была оптимальной, достаточно по условию (10.211), чтобы

$$F_{A_1} - \beta > F_T$$

так как тогда $F_A \geq F_{A_1} - \beta > F_T > F$. Отсюда следует, что для этой стратегии

$$\sum_{i=1}^q \psi_i < y = \begin{cases} (F_{A_1} - F_T) / F_Y & \text{для ГМСК} \\ (F_{A_1} - F_T) / (L_Y \Omega) & \text{для ИМСК} \end{cases} \quad (10.213)$$

Проверка условия (10.213) позволяет отсеивать заведомо неоптимальные стратегии, не вычисляя для них значения показателя эффективности.

Алгоритм позволяет найти абсолютный минимум F_A и оптимальную стратегию, однако (в худшем случае) может возникнуть потребность полного перебора.

10.2.6.4. Алгоритм оценок на основе линейного программирования

В отличие от предыдущего настоящий алгоритм позволяет гарантированно получить лишь оценки сверху и снизу для оптимального значения показателя эффективности МСК (хотя в отдельных случаях дает, как и предыдущий, \bar{F} и \bar{A}). Однако он значительно экономнее в вычислительном плане, так как сводится к решению обычной задачи линейного программирования. Кроме того, для того чтобы вести процесс модификации сетей для оптимизации исходной МСК, вполне достаточно оценок, обеспечиваемых этим алгоритмом.

Введем переменные:

$$\left. \begin{aligned} 0 \leq u_i \leq 1, \quad i = 1, \dots, M; \\ 0 \leq v_{si} \leq 1, \quad s = 1, \dots, N, \quad i = 1, \dots, M, \end{aligned} \right\} \quad (10.214)$$

удовлетворяющие условиям:

$$\sum_{i=1}^M u_i = m; \quad (10.215)$$

$$\sum_{i=1}^M v_{si} = m-1, \quad s=1, \dots, N, \quad i=1, \dots, M; \quad (10.216)$$

$$v_{si} \leq u_i, \quad s=1, \dots, N,$$

и функции:

$$L(x_s) = \sum_{i=1}^M (u_i - v_{si}) \rho(x_s, y_i), \quad s=1, \dots, N; \quad (10.217)$$

$$F_{ГМСК} \geq L(x_s), \quad s=1, \dots, N; \quad (10.218)$$

$$F_{ИМСК} = \sum_{s=1}^N L(x_s). \quad (10.219)$$

Рассмотрим следующие оптимизационные задачи: найти значения $\{\tilde{u}_i, \tilde{v}_{si}\}_{s=1, \dots, N}^{i=1, \dots, M}$, при которых достигается минимум $F_{ГМСК}$ или $F_{ИМСК}$ при выполнении условий (10.214)—(10.218).

Эти задачи являются задачами линейного программирования с матрицей ограничений весьма простой структуры. Решение таких задач — стандартная проблема вычислительной математики, для чего имеется достаточно много программ, входящих в математическое обеспечение всех современных ЭВМ.

Исследуем эти задачи. Первыми определяющими их условиями являются

$$F_{ГМСК} \geq \min_{s=1, \dots, N} L(x_s) \geq 0; F_{ИМСК} \geq L(x_s) \geq 0,$$

т. е. критерии ограничены снизу, поэтому задачи имеют решения.

Если $u=(u_1, \dots, u_m)$ — фиксированный вектор, удовлетворяющий условиям (10.213), (10.214), определим значения $\{v_{si}\}_{s=1, \dots, N}^{i=1, \dots, M}$, минимизирующие $F_{ГМСК}$, $F_{ИМСК}$. Значения $\{v_{si}\}_{i=1, \dots, M}$ для различных u независимы между собой, поэтому их можно определить из следующей вспомогательной задачи:

$$v_{si} = u_i - z_{si}, \quad i=1, \dots, M; \quad (10.220)$$

$$L(x_s) = \sum_{i=1}^M z_{si} \rho(x_s, y_i) \rightarrow \min; \quad (10.221)$$

$$\sum_{i=1}^M z_{si} = 1; \\ 0 \leq z_{si} \leq u_i, \quad i=1, \dots, M. \quad (10.222)$$

Простейшая задача линейного программирования, определяемая условиями (10.220)—(10.222), имеет очевидное решение. Числа $\rho(x_s, y_i)$, $i=1, \dots, M$ упорядочиваются по возрастанию, т. е. каждому номеру i сопоставляется номер k числа $\rho(x_s, y_i)$ в этой последовательности, так что если,

$$k_s(i_1) > k_s(i_2)$$

то

$$\rho(x_s, y_{i_1}) \geq \rho(x_s, y_{i_2}) \quad \forall i_1, i_2 \in \{1, \dots, M\}. \quad (10.223)$$

Затем определяется номер k_s^* такой, что

$$\sum_{k_s^*}^{k_s^*} u_{is}(k_s) \leq 1, \quad \sum_{k_{s-1}}^{k_{s+1}^*} u_{is}(k_s) = 1, \quad (10.224)$$

где $i_s(k)$ — обратная функция $k_s(i)$. Оптимальные значения:

$$z_{si} = \begin{cases} u_i & \text{при } k_s(i) \leq k_s^*; \\ 0 & \text{при } k_s(i) > k_s^* + 1; \\ 1 - \sum_{k_s=1}^{k_s^*} u_{i_s}(k_s) & \text{при } k_s(i) = k_s^* + 1, \end{cases} \quad (10.225)$$

которым соответствуют значения:

$$v_{si} = \begin{cases} 0 & \text{при } k_s(i) \leq k_s^*; \\ u_i & \text{при } k_s(i) > k_s^* + 1; \\ \sum_{k_s=1}^{k_s^*} u_{i_s}(k_s)^{-1} & \text{при } k_s(i) = k_s^* + 1. \end{cases} \quad (10.226)$$

Из выражения (10.226) следует, что оптимальное значение v_{si} будет не более, чем при одном значении i (при данном s), отличным от 0 или u_i . При этом соотношению (10.225) соответствует следующее оптимальное значение (10.220):

$$\begin{aligned} L(x_s) &= \sum_{k_s=1}^{k_s^*} u_{i_s}(k_s) \rho(x_s, y_{i_s}(k_s)) + \\ &+ \left(1 - \sum_{k_s=1}^{k_s^*} u_{i_s}(k_s) \right) \rho(x_s, y_{i_s}(k_s^* + 1)) = \rho(x_s, y_{i_s}(k_s^* + 1)) - \\ &- \sum_{k_s=1}^{k_s^*} u_{i_s}(k_s) (\rho(x_s, y_{i_s}(k_s^* + 1)) - \rho(x_s, y_{i_s}(k_s))). \end{aligned} \quad (10.227)$$

Существует оптимальное решение исходной задачи, для которого выполняются соотношения (10.225)—(10.227).

Допустим, что условия (10.214)–(10.219) дополнены требованиям целочисленности u_i , $i = 1, \dots, M$. Такие задачи назовем суженными, поскольку множество векторов u_i здесь уже, и соответственно оптимальные значения критериев будут не меньше, чем в задачах, определяемых условиями (10.224)—(10.229).

В суженных задачах, очевидно,

$$\sum_{k_s=1}^{k_s^*} u_{i_s}(k_s) = u_{i_s}(k_s^*) = 1;$$

$$u_{i_s}(k_s) = 0 \quad \forall k_s \neq k_s^*;$$

$$\sum_{k_s=1}^{k_s^*} u_{i_s}(k_s) - 1 = 0.$$

Тогда из соотношения (10.227) с учетом условия (10.223) можно записать

$$L(x_s) = u_i(k_s^*)\rho(x_s, y_{i_s}, (k_s^*)) = \min_{\substack{i=1, \dots, M \\ u_i=1}} \rho(x_s, y_i) \quad (10.228)$$

Соотношение (10.228) совместно с выражениями (10.218), (10.219) позволяет установить взаимоднозначное соответствие между m -элементными стратегиями A множества Y_K и булевыми векторами $u = (u_1, \dots, u_M)$, удовлетворяющими выражениями (10.218), (10.219):

$$u_i = \begin{cases} 0, & \text{если } y_i \in \bar{A}; \\ 1, & \text{если } y_i \in A, \quad i = 1, \dots, M. \end{cases} \quad (10.229)$$

Причем значения $F_{ГМСК}$, $F_{ИМСК}$ в суженных задачах равны соответствующим значениям показателей эффективности $ГМСК$ и $ИМСК$. Это следует из того, что по условиям (10.228) и (10.229)

$$L_u(x_s) = L(x_s),$$

а в суженной задаче, учитывая неравенство (10.228), а также требование минимизации $F_{ГМСК}$, можно получить

$$F_{ГМСК} = \max_{s=1, \dots, M} L(x_s) = \max_{x \in X_k} L(x).$$

Как отмечалось, оптимальное значение критериев в суженных задачах не меньше, чем в исходных задачах, определяемых условиями (10.214)—(10.219), поэтому окончательно можно установить, что если $\bar{F}_{ГМСК}$, $\bar{F}_{ИМСК}$ — оптимальные значения критериев в исходных задачах, то

$$\bar{F} \geq \bar{F}_{ИМСК} \text{ для } ИМСК$$

$$\bar{F} \geq \bar{F}_{ГМСК} \text{ для } ГМСК$$

Обозначая

$$\bar{F}_{\text{лп}} = \begin{cases} \bar{F}_{\text{ИМСК}} & \text{для ИМСК;} \\ \bar{F}_{\text{ГМСК}} & \text{для ГМСК,} \end{cases}$$

запишем

$$F \geq \bar{F}_{\text{лп}} \quad (10.230)$$

Итак, решая задачу, определяемую условиями (10.214)—(10.219), с критерием $\bar{F}_{\text{ГМСК}}$ или $\bar{F}_{\text{ИМСК}}$ в зависимости от типа оптимизируемой МСК можно получить оценку снизу для оптимального значения показателя эффективности МСК. Если \bar{u} — оптимальное решение рассматриваемой задачи — булево, то

$$\bar{F} = \bar{F}_{\text{лп}} \quad (10.230)$$

и соответствующая \bar{u} по условию (10.229) стратегия \bar{A} — оптимальна. Если же \bar{u} небулево, можно получить ближайший к нему булевый вектор \bar{u}_s , например,

$$\begin{aligned} \bar{u}_s &= (\bar{u}_{1\sigma} \dots \bar{u}_{M\sigma}); \\ \sum_{i=1}^M u_{i\sigma} &= m; \\ \min_{i/u_{i\sigma}=1} &\geq \max_{i/u_{i\sigma}=0} \bar{u}_i. \end{aligned}$$

В этом случае $F_{\bar{A}\sigma} \geq F$, где \bar{A}_σ — стратегия, соответствующая по условию (10.229) \bar{u}_s .

Заметим, что если \bar{u} — булево, то $\bar{u}_s = \bar{u}$.

Итак, рассматривая полученные условия совместно с выражением (10.226), получим

$$F_{\text{лп}} \leq F \leq F_{\bar{A}_\sigma} \quad (10.232)$$

т. е. решение соответствующей задачи линейного программирования, определяемой условиями (10.214)-(10.219), позволяет получить двустороннюю оценку для показателя эффективности оптимальной МСК.

10.2.6.5. Алгоритмы распределения и улучшения

Алгоритм распределения, так же как и алгоритм улучшения, не гарантирует отыскания абсолютного решения, так как основан на локальном улучшении некоторого начального варианта. Однако его преимуществом является сравнительная простота и малое время счета.

Поэтому его целесообразно использовать как отдельно, так и совместно с алгоритмом оценок на основе линейного программирования.

Пусть $\{E^0_{j|j=1, \dots, m}$ — начальный набор областей Дирихле. Перебором для любого $y \in Y_K$ можно рассчитать

$$F^0_j(y) = \begin{cases} \max_{x \in E^0_j} \rho(x, y) & \text{для ГМКС} \\ \sum_{x \in E^0_j} \rho(x, y) & \text{для ИМКС} \end{cases} \quad (10.233)$$

и вычислить $F^0_j, y^0_j, j = 1, \dots, m$, т. е.

$$\bar{F}^0_j = \min_{y \in Y_K} F^0_j(y) = F^0_j(\bar{y}^0_j), \quad j=1, \dots, m.$$

Для стратегии $A^0 = \{\bar{y}^0_{j|j=1, \dots, m}$ найдем оптимальный набор областей Дирихле $\{E^1_{j|j=1, \dots, m}$, используя очевидное условие: $x \in E^1_j$, если $L_{A^0}(x) = \rho(x, \bar{y}^0_j)$. При этом, учитывая выражения (10.183), (10.184), можно получить

$$F_{A^0} = \begin{cases} \max_{j=i, \dots, m} F_j^1(\bar{y}^0_j) & \text{для ГМКС} \\ \sum_{j=1}^m F_j^1(\bar{y}^0_j) & \text{для ИМКС} \end{cases}$$

Затем процесс повторяется: для $\{E^1_{j|j=1, \dots, m}$ отыскивается $A^1 = \{\bar{y}^1_{j|j=1, \dots, m}$ и т.д.

В общем случае

$$F^s_j(y) = \begin{cases} \max_{x \in E^s_j} \rho(x, y) & \text{для ГМСК} \\ \sum_{x \in E^s_j} \rho(x, y) & \text{для ИМСК} \end{cases}$$

$$\left. \begin{aligned} A^s &= \{\bar{y}^s_{j|j=1, \dots, m}, F^s_j(\bar{y}^s_j) = F^s_j = \min_{y \in Y_K} F^s_j(y); \\ E^{s+1}_j &= \{x \in X_K : L_{A^s}(x) = \rho(x, \bar{y}^s_j)\}_{j=1, \dots, m}; \end{aligned} \right\} \quad (10.234)$$

$$F_{A^s} = \begin{cases} \max_{j=i, \dots, m} F_j^{s+1}(\bar{y}_j^s) & \text{для ГМКС} \\ \sum_{j=1}^m F_j^{s+1}(\bar{y}_j^s) & \text{для ИМКС, } s = 0, 1, \dots \end{cases}$$

Поскольку из связей (10.234) можно получить неравенство

$$F^{s+1}_j(\bar{y}^{s+1}_j) \leq F^{s+1}_j(\bar{y}^s_j),$$

совместное рассмотрение которого с выражениями (10.183), (10.184) дает

$$F_{A^{s+1}} \leq \begin{cases} \max_{j=i, \dots, m} F_j^{s+1}(\bar{y}_j^{s+1}) & \text{для ГМКС} \\ \sum_{j=1}^m F_j^{s+1}(\bar{y}_j^{s+1}) & \text{для ИМКС,} \end{cases}$$

то

$$F_{A^{s+1}} \leq F_{A^s}, s = 0, 1, \dots$$

Таким образом, алгоритм обеспечивает монотонное изменение показателя эффективности МСК. Поскольку он ограничен снизу и изменяется на замкнутом ограниченном (конечном) множестве, алгоритм сходится к некоторой неуплучшаемой стратегии A , а условие остановки будет $A^{s+1} = A^s$.

Для уменьшения перебора используем тот факт, что поскольку на каждом шаге при изменении стратегии заменяется лишь один ее элемент, из перебора исключаются элементы $y \in Y$, расположенные не далее чем на z от элементов уже рассмотренных стратегий A . Аналогично из перебора исключаются элементы $y \in Y$, расположенные не далее чем на z от элементов уже рассмотренных стратегий, где z удовлетворяет условию

$$z < \begin{cases} \frac{F_{m-1}^0 - F_m^0}{L_Y} & \text{для ГМСК} \\ \frac{F_{m-1}^0 - F_m^0}{L_Y \Omega} & \text{для ИМСК,} \end{cases} \quad (10.235)$$

являющемуся следствием из условий (10.208), (10.209).

Алгоритм улучшения несколько более трудоемок, чем рассмотренный выше алгоритм распределения, но и несколько эффективнее его. Обозначим через $F_{A,j}$ значение показателя эффективности МСК на области Дирихле j -го элемента стратегии A . По условиям (10.183), (10.184)

$$F_A \leq \begin{cases} \max_{j=i, \dots, m} F_{A,j} & \text{для ГМСК} \\ \sum_{j=1}^m F_{A,j} & \text{для ИМСК,} \end{cases}$$

Пусть $A^0 = \{y_j^0\}_{j=1, \dots, m}$ — некоторая начальная стратегия. Для любого элемента $y \in Y_K$ можно вычислить

$$\Delta^0(y) = \min_{\substack{j=1, \dots, m \\ y+y^0}} (-F_{A^0/y} + F_{(A^0 \setminus y^0_j) \cup_{y,j}}) = -F_{A^0/y^0} + F_{(A^0 \setminus y^0_{j^0}) \cup_{y^0_{j^0}}}. \quad (10.236)$$

Произведем перебор $y \in Y_K$, для которых вычисляется выражение (10.236). Если для некоторого $y \in Y_K$ оказывается

$$\Delta^0(\tilde{y}) < 0 \text{ и } F_{(A^0 \setminus y^0_{j^0}) \cup_{\tilde{y}}} \leq F_{A^0}, \quad (10.237)$$

то A_0 замещается новой стратегией:

$$A^1 = (A^0 \setminus y^0_{j^0}) \square \tilde{y},$$

т. е. элемент y_{j^0} заменяется элементом \tilde{y} , и процесс продолжается. Он прекращается, когда

$$\Delta^p(\tilde{y}) \geq 0 \quad \forall y \in Y_K.$$

Таким образом, в общем случае

$$\Delta^s(y) = \min_{\substack{j=1, \dots, m \\ y^s_j \neq y}} (F_{(A^s \setminus y^s_{j^s}) \cup_{y,j}} - F_{A^s, j}) = F_{(A^s \setminus y^s_{j^s}) \cup_{y^s_{j^s}, j^s}};$$

$$A^s(\tilde{y}^s) < 0 \quad \text{при } F_{A^s}^{s+1} \leq F_{A^s}$$

$$A^{s+1} = (A^s \setminus y^s_{j^s}) \cup \tilde{y}^s, \quad s=0, 1, \dots$$

С учетом этого условие (10.237) означает

$$F_{A^{s+1}} \leq F_{A^s}, \quad s=0, 1, \dots,$$

что доказывает, аналогично алгоритму распределения, сходимость, алгоритма улучшения. При этом имеют место те же условия уменьшения перебора. Алгоритм улучшения «сильнее» алгоритма распределения, так как, используя его, можно улучшить решение, полученное алгоритмом распределения, но не наоборот.

Перейдем к конструированию рационального управления общим алгоритмом оптимизации МСК (рис. 10.59).

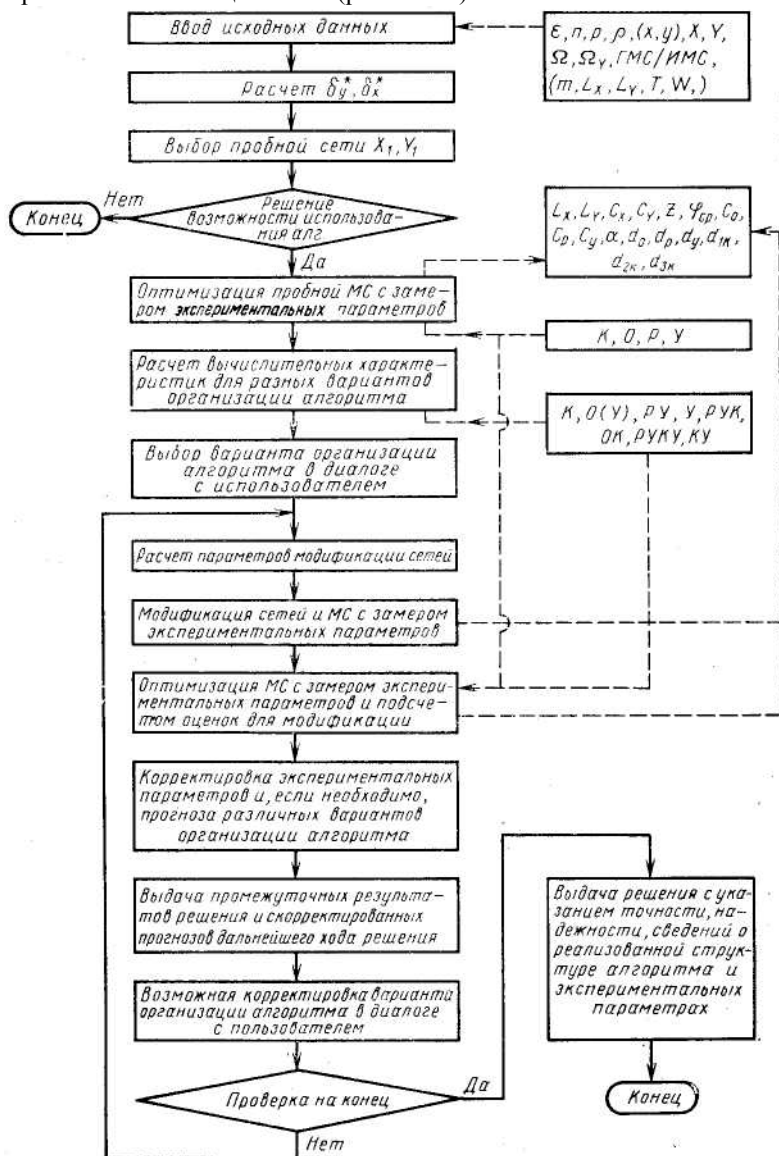


Рис. 10.59. Блок-схема общего алгоритма оптимизации МСК

Его основными функциональными составляющими являются:

- выявление и уточнение коэффициентов, характеризующих вычислительные особенности решаемой задачи;
- прогнозирование вычислительных характеристик различных вариантов организации решения задачи и выбор в возможном диалоге с пользователем варианта, наиболее соответствующего выделенным вычислительным ресурсам;
- расчет оценок, необходимых для модификации сетей в множествах X и Y ;
- модификация сетей;
- оптимизация МСК с конечным внешним множеством и множеством стратегий.

Структурные составляющие (блоки) алгоритма объединяют, в отдельных случаях, несколько функциональных составляющих.

Исходными данными для алгоритма являются описание оптимизируемой МСК, т. е. множеств X , Y (в том числе n , p , Ω , Ω_Y) и функции локальной эффективности $\rho(x, y)$, а также типа МСК и для ИМСК, порождающей функции $G(t)$. Кроме того, могут быть заданы постоянные Липшица L_X , L_Y , число элементов стратегии m , точность решения z , машинное время T и объем памяти W .

Если они не заданы, то L_X , L_Y устанавливаются в процессе решения, а m , ε , T и W уточняются в диалоге по мере прогнозирования вычислительных характеристик решения и получения промежуточных результатов.

После ввода исходных данных производится формирование пробной сети X_I , Y_I и оптимизация соответствующей пробной МСК с использованием всех наличных алгоритмов оптимизации МСК с конечными множествами. Пробные сети формируют возможно более грубыми, содержащими малое число элементов, с тем, чтобы выполнение указанной операции требовало сравнительно малого машинного времени и заведомо было обеспечено вычислительными ресурсами. Цель формирования и оптимизации пробной МСК состоит в том, чтобы получить приближенные значения коэффициентов, характеризующих вычислительные особенности задачи. К ним относятся: L_X , L_Y (если они не заданы); C_x , C_y — связывающие параметры сетей с числом элементов в них; C_0 , C_p , C_y — определяющие трудоемкость различных алгоритмов оптимизации; d_0 , d_p , d_y , d_{1K} , d_{2K} , d_{3K} — задающие объемы памяти, необходимые для различных алгоритмов оптимизации; α , z , y_{cp} — характеризующие возможность отбрасывания элементов сетей при модификации.

Одновременно подсчитывают (хотя и очень грубые) оценки для оптимального значения показателя эффективности, которые уже используют на следующем шаге для уменьшения переборов.

После получения перечисленных характеристик прогнозируют время счета и объем памяти для различных вариантов организации решения. Каждый вариант представляет собой ряд последовательных модификаций сетей в множествах X и Y , к каждой из которых применяют алгоритмы оптимизации МСК с конечными множествами. Число модификаций и параметры сетей устанавливаются в соответствии с анализом, проведенным выше, так что различные варианты отличаются друг от друга набором и последовательностью применяемых алгоритмов оптимизации. Целесообразно рассмотреть 8 вариантов алгоритмов. Дадим их описание, обозначая: K — комбинаторный алгоритм, O — алгоритм оценок, P — алгоритм распределения, U — алгоритм улучшения.

Вариант K — используется комбинаторный алгоритм. При $m > 2$ этот вариант требует заведомо больше времени, чем все остальные, зато гарантирует отыскание с заданной точностью абсолютно оптимального решения.

Вариант $O(Y)$ — используется алгоритм оценок. При $m > 2$ требует меньше времени, чем вариант K , но значительно больше, чем при использовании алгоритмов распределения и улучшения. Гарантирует получение двусторонней оценки для показателя эффективности абсолютно оптимального решения. Если его решение целочисленно, то тем самым определяется абсолютно оптимальное решение; если нецелочисленно, то ближайшее целочисленное решение отыскивается алгоритмом улучшения. Важным параметром варианта является $\delta_{ym} \geq \delta_j^*$ — параметр сети в Y , начиная с которого следует переходить к алгоритму улучшения.

Вариант PY — используется алгоритм распределения вплоть до δ_j^* - сети в Y , после чего однократно применяется алгоритм улучшения. Это самый быстрый вариант, однако полученное им решение лишь неуклучшаемо, оно может не быть абсолютным оптимумом. О степени близости к абсолютному оптимуму можно судить по оценкам, полученным после оптимизации пробной МСК. Его рекомендуется использовать, если известно начальное приближение стратегии, улучшение которого заведомо приведет к абсолютному оптимуму, а также для быстрого получения ориентировочных результатов в задаче.

Вариант U — используется алгоритм улучшения. Время счета несколько больше, чем при варианте PY , зато полученное решение

может быть лучше (по показателю эффективности), чем в варианте РУ. В остальном справедливо все, сказанное о предыдущем варианте.

Вариант РУК — до получения сети в Y с параметром δ_{yn} используется алгоритм распределения, затем однократно применяется алгоритм улучшения. Полученное решение используется для оценок, уменьшающих перебор в комбинаторном алгоритме, который применяется вплоть до получения заданной точности решения. Здесь обеспечивается отыскание абсолютного минимума, а машинное время может оказаться меньше, чем в варианте К, если затраты времени на работу алгоритмов распределения и улучшения будут перекрыты экономией за счет уменьшения переборов.

Вариант ОК — аналогичен варианту РУК с той лишь разницей, что оценки для уменьшения перебора получаются использованием до δ_{yn} -сети в Y алгоритма оценок.

Вариант КУ — до δ_{yn} -сети в Y используется комбинаторный алгоритм, а затем алгоритм улучшения. По затратам времени этот вариант промежуточный между вариантами К и У, причем чем больше затраченное время, а оно регулируется параметром δ_{yn} , тем выше вероятность, что будет найдено не просто неуплощаемое, а абсолютно оптимальное решение.

Вариант РУКУ — до δ_{yn} -сети в Y используется вариант РУК, а затем алгоритм улучшения. Справедливо все, сказанное о предыдущем варианте, причем время счета может быть меньше, чем в варианте КУ, если оно в варианте РУК меньше, чем в варианте К.

Другие варианты сочетания алгоритмов оптимизации МСК с конечными множествами, на наш взгляд, нерациональны.

Рассчитанные вычислительные характеристики этих вариантов сравнивают с выделенными ресурсами, желательно в диалоге с пользователем, и производят выбор варианта организации решения задачи. При этом можно корректировать вычислительные ресурсы, устанавливать требуемую точность решения ϵ .

Дальнейшая работа алгоритма является циклической. На каждом шаге циклической процедуры производится модификация сетей в X , Y и оптимизация полученной МСК алгоритмом, соответствующим принятому варианту организации решения задачи. Одновременно замеряются значения экспериментальных коэффициентов, характеризующих вычислительные особенности задачи. На этой основе корректируются ранее принятые значения этих коэффициентов, уточняется прогноз вычислительных характеристик различных вариантов организации решения и, в случае необходимости, в диалоге с пользователем корректируется первоначальный вариант организации

решения. Пользователю сообщается результат оптимизации МСК на текущем шаге с целью возможного уточнения значения m и внесения изменений в исходные данные и организацию процесса решения. Затем начинается следующий шаг итерационного процесса или, если выполнены условия окончания, пользователю выдаются результаты решения.

Литература

1. European Directory of Management Consultants. 1995. London: FEACO-AP Information services, 1995.
2. Guidelines for the Use of Consultants by World Bank Borrowers and by the World Bank as Executing Agency. Washington (D.C.): The World Bank, 1992.
3. Hurley N. Management Consultancy Manual: Operating a Successful Management Consultancy Assignment. Ankara: SMIDO, 1990.
4. Kubr M. How to select and use consultants: A client's guide. Geneva: ILO, 1993.
5. Maister D. Professional Service Firm Management. Boston: Maister Associates. Inc., 1990.
6. Кудинов А. О рынке консалтинговых услуг. // www.bcg.ru.
7. Монахова Е. Управленческое консультирование конца XX века. // www.pcweek.ru/kis.
8. Посадский А.П., Хайниш С.В. Консультационные услуги в России. - М.: Финстатинформ, 1995, - 171 с.
9. Уткин Э.А. Консалтинг. – М.: ЭКМОС, 1998, - 256 с.
10. Интернет-сервер Гарвардской школы бизнеса: www.hbs.edu
11. BS7799: Информационная безопасность начинается с менеджмента. // Банковские технологии. №8, 1998, - С. 73-75.
12. Международный стандарт ИСО 9000. Системы менеджмента качества. Основные положения и словарь. 2-е изд. 2000-12-15. ISO - 2000.
13. Международный стандарт ИСО 9001. Системы менеджмента качества. Требования. 3-е изд. 2000-12-15. ISO – 2000.
14. Международный стандарт ИСО 9004. Системы менеджмента качества. Руководство по улучшению деятельности. 2-е изд. ISO – 2000.
15. ISO 9000 Introduction and Support Package: Guidelines on the Process Approach to quality management systems. ISO/TC 176/SC 2/N 544R. 17 May, 2001.
16. ISO 9000 Introduction and Support Package: Guidance on the Documentation Requirements of ISO 9001:2000. ISO/TC 176/SC 2/N 544R. 13 March, 2001.

17. Давид Марка, Клемент МакГоуэн. Методология структурного анализа и проектирования. Пер. с англ. М.: 1993, 240 с., ISBN 5-7395-0007-9
18. INTEGRATION DEFINITION FOR FUNCTION MODELING (IDEF0). Draft Federal Information Processing Standards Publication 183, 1993, December 2. www.idef.com
19. Р50.1.028-2001. Методология функционального моделирования. М.: Госстандарт России, 2000. www.cals.ru
20. Менеджмент качества и международные стандарты ИСО 9000 версии 2000 г. Материалы семинара в рамках Программы ИСО для развивающихся стран. Минск, Июль 2001 г. 79 с.
21. Framework for Managing Process Improvement. Vol.1. Electronic College of Process Innovation. DoD USA. May, 1994.
22. С. Бобровски - ORACLE 7 и вычисления клиент - сервер, Москва, "ЛОРИ", 1996.
23. М. Ричарде и др. - Oracle 7.3, Энциклопедия пользователя, Киев, "DiaSoft", 1997.
24. ORACLE7 SERVER - CONCEPT MANUAL.
25. ORACLE7 SERVER - SQL LANGUAGE REFERENCE MANUAL -Part Number 778-70-1292, December 1992.
26. С. Бемер, Г.Фратер - MS Access...для пользователя, Киев, "ВНУ", 1994.
27. Р. Дженнингс - Access 95 в подлиннике, 2 тома, СПб, "ВНУ", 1997.
28. Р. Дженнингс - Microsoft Access 97 в подлиннике, 2 тома, Санкт-Петербург, "ВНУ", 1997.
29. Керри Н. Праг, Мишель Р. Ирвин - Access 97, Библия пользователя, Киев-Москва, "Диалектика", 1997.
30. А. Горев, С. Макашаринов - Microsoft Visual FoxPro 3.0, Новые возможности для программиста, Санкт-Петербург, "Питер Пресс", 1995.
31. Б. Сосински - Разработка приложений в среде Visual FoxPro 5, Киев-Москва, "Диалектика", 1997.
32. Цыпкин Я. З. Основы теории автоматических систем. — М.: Наука. Гл. ред. физ.-мат. лит., 1977. — 560 с.
33. Лукашин Ю. 77. Адаптивные методы краткосрочного прогнозирования. — М.: Статистика, 1979. — 254 с.
34. Марпл-мл. С. Л. Цифровой спектральный анализ и его приложения: Пер. с англ. — М.: Мир, 1990. — 584 с.

35. Дрейпер Н., Смит Г. Прикладной регрессионный анализ: В 2-х книгах: Пер. с англ. — М.: Финансы и статистика, 1986. — Кн. 1. — 366 с.

36. Дрейпер Н., Смит Г. Прикладной регрессионный анализ: В 2-х книгах: Пер. с англ. — М.: Финансы и статистика, 1987. — Кн. 2. — 351 с.

37. Бриллинджер Д. Временные ряды. Обработка данных и теория: Пер. с англ. — М.: Мир, 1980. — 536 с.

38. Ахметшин А. М., Киргизов И. А. Прогнозирование кратковременных экономических рядов как задача адаптивной экстраполяции функций с финитным спектром // Прац Міжнар. конф. з індуктивного моделювання «МКІМ — 2002». — Львів 2002. — С. 29—33.

39. Клик Ю. А., Плотникова Л. И. Спектральный анализ функций. — К.: УМКВО, 1992. — 108 с.

40. Kasabov N. K. Foundations of neural networks, fuzzy systems, and knowledge engineering. — Cambridge, Mass.: MIT Press, 1996. — 550 p.

41. Беллман Р. Введение в теорию матриц: Пер. с англ. — М.: Наука. Гл. ред. физ.-мат. лит., 1976. — 352 с.

42. Михалевич В.С., Волкович В.Л. Вычислительные методы исследования и проектирования сложных систем.-М.: Наука, 1982.-288 с.

43. Моисеев Н.Н., Иванилов Ю.П., Столярова Е.М. Методы оптимизации. М. -Наука, 1978.-352 с.

44. Математическое моделирование экономических процессов в сельском хозяйстве: Уч. пособие /Под ред. А.М. Гатаулина. - М.: Агропромиздат, 1990.-432 с.

45. Гранберг А.Г. Динамическое моделирование народного хозяйства -М: Экономика, 1985.-240с.

46. Бакаев О.О., Кочура С.Н., Кальна-Дубинюк Т.П. Один із можливих підходів до визначення періоду дослідження розвитку економічних систем. //Вісник С.-Г. науки. -1978.- №10.-С. 97-102.

47. Кальная-Дубинюк Т.П. Динамический подход к моделированию развития сложных экономических систем // Применение математических методов в экономических исследованиях и планировании. - К.: Институт кибернетики АН УССР, 1979. - С. 40-47.

48. Кальна-Дубинюк Т.П. Застосування методів послідовного аналізу варіантів для наукового обґрунтування керівних рішень в агробізнесі // Удосконалення госпрозрахункових відносин за нових умов господарювання: Зб. наук. пр - К.: НАУ, 1996.- С 121-125

49. Ястремський О.І. Основи теорії економічного ризику: Начальний посібник для студентів екон. спец. вищ. навч. закладів. - К.: АртЕк, 1999 - 248 с.
50. Кальна-Дубінюк Т.П. Моделювання інвестиційної діяльності і агропідприємств // Агроінком - 1999.- № 1-2.- С. 42-44.
51. Кальна-Дубінюк Т. П. Агroeкономічна динаміка та методологічні аспекти її моделювання: Зб. наук. праць.- К.: НАУ. 2000, вип.24.- С. 300-308.
52. Кальна-Дубінюк Т.П. Динамічне моделювання економічних процесів у галузях АНК. Типова програма для сільськогосподарських вищих навчальних закладів із спеціальності 7.050202 "Аграрний менеджмент" - ЕС, 1996.-8 с. 3.
53. Кальна-Дубінюк Т.П. Рейтингова система оцінки знань студентів сільськогосподарських вищих і середніх навчальних закладів. Метод, вказівки / УМК по середній спеціальній освіті. - К., 1993.-15 с.
54. Пономаренко О.І., Пономаренко В.О. Системні методи в економіці, менеджмент та бізнесі: Навч. посібник.- К. Либідь, 1995. -240 с.
55. Воркуев Б.Л. Модели макро- и микроэкономики. -М.: Экономический факультет, ТЕИС, 1999.- 235с.
56. Математические методы в социально-экономических исследованиях: Сб. науч.ст./ Под ред. С.М.Ермакова, В.Б.Меласа.- Санкт-Петербург, ТОО ТК «Петрополис», 1996. - 138с.
57. Методичні рекомендації до виконання практичних і лабораторних робіт з теорії прийняття рішень // Упорядн. О.Ф.Волошин, С.О.Мащенко. - К.: ВПЦ "Київський університет", 2001. - 46с.
58. Алексеев А. В. Проблема построения нечетких ситуационных моделей управления. В кн.: Методы и системы принятия решений — Рига: РПИ 1979 с. 12—19.
59. Алексеев А. В. Интерпретация и определение функций принадлежности нечетких множеств. — Там же, с. 42—50.
60. Алексеев А. В. Лингвистические модели принятия решений в нечетких ситуационных системах управления. — В кн.: Методы принятия решений в условиях неопределенности. — Рига: РПИ, 1980, с. 19—25.
61. Алексеев А. В., Борисов А. Н. Модели нечетких статистических и динамических отношений в ситуационных моделях управления организационными системами. — Управление сложными системами. — Рига- РПИ 1978 вып 5 с. 81—88.

62. Борисов А. Н. Некоторые обучающиеся алгоритмы диагностики систем с размытыми классами состояний. — *Техническая кибернетика*. — Рига, 1970.
63. Борисов А. Н., Алексеев А. В. Нечеткие алгоритмы в ситуационных моделях управления организационными системами. — В кн.: *Методика построения систем ситуационного управления* / Науч. совет АН СССР по комплексной проблеме «Кибернетика». — М., 1978, с. 3—10.
63. Борисов А. Н., Аппен Е. П. Оценка возможных характеристик при анализе альтернатив. — В кн.: *Методы принятия решений в условиях неопределенности*. — Рига: РПИ, 1980, с. 94—100.
64. Борисов А. Н., Голендер В. Е. Оптимальное разделение размытых образов — *Методы и средства технической кибернетики*. — Рига: РПИ, 1969, вып. 5, с. 32—38.
65. Борисов А. Н., Корнеева Г. В. Лингвистический подход к построению моделей принятия решений в условиях неопределенности. — В кн.: *Методы принятия решений в условиях неопределенности*. — Рига: РПИ, 1980, с. 4—12.
66. Борисов А. Н., Крумберг О. А. Анализ решений при выборе технологических объектов. — Там же, с. 127—134.

Научно-практическое издание

Кононюк Анатолий Ефимович

Консалтинтология

Общая теория консалтинга

Книга 4

Авторская редакция

Подписано в печать 21.12.2010 г.

Формат 60x84/16.

Усл. печ. л. 16,5. Тираж 300 экз.

Издатель и изготовитель:

Издательство «Освита Украины»

04214, г. Киев, ул. Героев Днепра, 63, к. 40

Свидетельство о внесении в Государственный реестр
издателей ДК №1957 от 23.04.2009 г.

Тел./факс (044) 411-4397; 237-5992

E-mail: osvita2005@ukr.net, www.rambook.ru

Издательство «Освита Украины» приглашает
авторов к сотрудничеству по выпуску изданий,
касающихся вопросов управления, модернизации,
инновационных процессов, технологий, методических
и методологических аспектов образования
и учебного процесса в высших учебных заведениях.

Предоставляем все виды издательских
и полиграфических услуг.

□