

ЭКОНОМИКО-
МАТЕМАТИЧЕСКАЯ
БИБЛИОТЕКА

В.С. МИХАЛЕВИЧ
А.И. КУКСА

МЕТОДЫ
ПОСЛЕДОВАТЕЛЬНОЙ
ОПТИМИЗАЦИИ

В. С. МИХАЛЕВИЧ,
А. И. КУКСА

МЕТОДЫ
ПОСЛЕДОВАТЕЛЬНОЙ
ОПТИМИЗАЦИИ
В ДИСКРЕТНЫХ СЕТЕВЫХ
ЗАДАЧАХ ОПТИМАЛЬНОГО
РАСПРЕДЕЛЕНИЯ РЕСУРСОВ



МОСКВА «НАУКА»
ГЛАВНАЯ РЕДАКЦИЯ
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
1983

Методы последовательной оптимизации в дискретных сетевых задачах оптимального распределения ресурсов. Михалеви ч В. С., Кукса А. И.— М.: Наука. Главная редакция физико-математической литературы, 1983.— 208 с.

Книга посвящена актуальным приложениям численных методов решения дискретных задач оптимизации. Последовательно изложены современные методы решения задач оптимизации распределения ресурсов на сетях, основанные на идее последовательной оптимизации, и проанализирована эффективность таких методов.

Для специалистов в области прикладной математики и кибернетики, а также специалистов в области экономики, применяющих в своей работе математические методы.

Рис. 16. Табл. 16. Библ. 119 назв.

Предисловие	5
Глава I. Дискретные сетевые задачи оптимального распределения ресурсов	11
1.1. Элементы языка сетевых моделей	11
1.2. Сетевое планирование с ограниченными ресурсами	15
1.3. Программно-целевой метод управления и сетевые задачи распределения ресурсов	19
1.4. Детерминированные задачи оптимального планирования в параллельных вычислительных системах	25
Библиографический комментарий	29
Глава II. Теория сложности экстремальных комбинаторных задач и сетевые задачи распределения ресурсов и составления расписаний	30
2.1. Элементы прикладной теории сложности	30
2.2. Простейшая модель вычислений: машины Тьюринга	34
2.3. Классы P и NP , языки и задачи	38
2.4. NP -полнота задачи выполнимости булевой формулы	42
2.5. NP -полнота простейших сетевых задач теории расписаний	49
2.6. NP -полнота задач с несколькими разнородными процессорами и независимыми цепями операций	60
2.7. Сложность обобщенной задачи составления расписания с векторными потребностями в ресурсах	66
Библиографический комментарий	68
Глава III. Вариации длительности кратчайших расписаний. Приближенные алгоритмы с оценкой погрешности	69
3.1. Алгоритмы с оценками	69
3.2. Вариации длительности приоритетных расписаний	71
3.3. Один класс приближенных асимптотически оптимальных алгоритмов в задачах типа «станки — детали»	91
Библиографический комментарий	102
Глава IV. Динамическое программирование	104
4.1. Алгоритм динамического программирования	104
4.2. О сложности алгоритмов динамического программирования	118
Библиографический комментарий	125

Глава V. Последовательный анализ вариантов	126
5.1. Задача с независимыми операциями	126
5.2. Метод последовательного анализа вариантов (общее описание)	127
5.3. Алгоритм последовательного анализа вариантов в задаче с независимыми операциями	130
5.4. Оценки трудоемкости в среднем алгоритма последовательного анализа вариантов	131
Библиографический комментарий	145
Глава VI. Метод ветвей и границ	146
6.1. Основные понятия и структура алгоритмов	146
6.2. Методы теории двойственности для решения оценочных задач	149
6.3. Двойственные оценки в задачах на адиклических сетях операций	159
6.4. Специальные методы решения оценочных задач	166
6.5. Практические алгоритмы ветвей и границ	182
Библиографический комментарий	196
Литература	198
Предметный указатель	206

Дискретные сетевые задачи оптимального распределения ресурсов являются обобщением соответствующих задач теории расписаний. Как известно, теория расписаний имеет дело с совокупностями действий (операций, работ и т. п.) и ресурсами (станками, процессорами и т. п.), необходимыми для осуществления этих действий. Экстремальные задачи распределения ресурсов возникают в связи с ограничениями на общие количества ресурсов или на сроки достижения конечных результатов совокупности действий. В публикациях по теории расписаний традиционно уделяется большое внимание задачам типа «станки — детали». В этих задачах технологический маршрут совокупности деталей представляется ориентированным графом операций, состоящим из нескольких не связанных между собой цепей. Отношение «операции — ресурсы» описывается указанием имени станка, выполняющего данную операцию. Каждый станок одновременно выполняет не более одной операции. Однако многие важные практические задачи не вписываются в указанную схему. Представляет интерес изучение более общих операционных моделей сложных составных действий, в которых:

- каждая операция описывается действительным вектором, компонентами которого являются длительность операции и уровни потребления ресурсов;

- технологические ограничения описываются произвольным ориентированным графом без петель и контуров;

- общие уровни наличия нескольких видов ресурсов заданы как функции времени.

Задачи с указанными особенностями условимся называть сетевыми задачами распределения ресурсов и составления расписаний. В сущности, все известные задачи теории расписаний могут быть рассмотрены как сетевые (с пустым, в частности, отношением предшествования операций). Не претендуя на полноту, упомянем лишь наиболее обширные области применения задач рассматриваемого класса.

Традиционной областью применения сетевых задач распределения ресурсов и составления расписаний является сетевое планирование с ограниченными ресурсами. Экстремальные задачи распределения ресурсов в сетевом планировании были осознаны и сформулированы еще в конце 50-х годов. В теории сетевого планирования они занимают одно из центральных мест. Однако практическое решение этих задач в автоматизированных системах СПУ связано с рядом трудностей. Они вызваны в основном недостаточной разработкой вопросов создания и автоматизированного ведения баз нормативных данных, типовых технологических схем планируемых процессов, недостаточной степенью автоматизации процессов подготовки данных, оперативного и наглядного отображения результатов счета. Требования оперативности ослабляются в тех случаях, когда речь идет о перспективном крупномасштабном планировании, разовом решении задач с целью выбора планов развития на большой отрезок времени, когда цена оптимального решения сравнительно велика. Одно из важных применений такого рода — использование сетевых задач оптимального распределения ресурсов в математическом обеспечении программно-целевого планирования народнохозяйственных программ.

В середине 60-х годов в СССР начали широко обсуждаться методы математического обеспечения процедур программно-целевого планирования. В частности, была предложена модель оценки выполнимости сформулированной программы, исследование которой сводится к сетевой задаче распределения ресурсов.

Новые области использования сетевых задач распределения ресурсов стимулируются развитием вычислительной техники и расширением сфер ее применения. В конце 60-х годов — начале 70-х годов в связи с появлением и технической реализацией идеи мультипроцессорирования были сформулированы задачи построения оптимальных мультипроцессорных расписаний, направленные на максимальное использование ресурсов развитой ЭВМ. В последнее десятилетие в этой интенсивно расширяющейся области исследований были получены интересные теоретические результаты. В связи с интенсивной разработкой сложных АСУ самого различного назначения поставлены задачи оптимальной организации внутренних процессов систем обработки данных, оперирующих большими массивами информации. Речь идет о распределении ресурсов вычислительной системы (физических устройств ЭВМ, реентерабельных программ и т. д.) при обработке информационно-связанной совокупности программ. В отличие от мультипроцессорных расписаний, задачи планирования вычислений при решении задач АСУ имеют ряд особенностей. А именно, объекты планирования (прикладные программы АСУ) характеризуются большей детерминированностью поведения, сравнительно низким уровнем требований к быстродействию процессоров и высоким уровнем требований к каналам обмена, часто должны решаться в заранее заданные сроки.

Настоящая монография посвящена численным методам решения дискретных сетевых задач оптимального распределения ограниченных ресурсов. Структура ее следующая. Вводная глава I «Дискретные сетевые задачи оптимального распределения ресурсов» посвящена содержательному описанию указанных задач. Она предназначена как пользователям, желающим распознать и сформулировать «свои» задачи на языке теории оптимизации, так и специалистам по методам дискретной оптимизации, для которых важно понять особенности возникающих задач, с тем чтобы использовать их для повышения эффективности соответствующих численных методов.

В главе II «Теория сложности экстремальных комбинаторных задач и сетевые задачи распределения ресурсов и составления расписаний» после краткого экскурса в современную прикладную теорию сложности вычислений приводятся результаты, показывающие, что даже простейшие сетевые задачи теории расписаний по крайней мере столь же сложны, как и ряд других широко известных экстремальных комбинаторных задач (например, задача минимизации дизъюнктивных нормальных форм функций алгебры логики, задача определения максимального внутренне устойчивого множества вершин графа). Как известно, большинство встречающихся на практике оптимизационных задач входит в класс так называемых *NP*-полных задач. Теория *NP*-полноты основана на понятии полиномиальной сводимости задач. Основной практический вывод этой теории состоит в том, что если хотя бы одна из *NP*-полных задач имеет эффективный полиномиальный алгоритм, то остальные тоже могут быть решены посредством алгоритма с полиномиальной оценкой трудоемкости. Однако, имея очевидное теоретическое значение, сам факт *NP*-полноты некоторой новой задачи оптимизации практически мало что дает исследователю-разработчику эффективных алгоритмов, а тем более пользователю. «Хорошие», эффективные алгоритмы в новых сложных задачах большой размерности являются не столько следствием хороших формальных схем вычислений, сколько результатом искусства разработчика, математика-вычислителя и системного программиста, их умения использовать особенности задач в целях ускорения вычислений и просто их опыта. В еще большей степени это справедливо, когда речь идет о задачах дискретного программирования, каковыми являются сетевые задачи распределения нескладируемых ресурсов. Существующий опыт решения этих задач говорит о том, что, несмотря на все ухищрения, практические алгоритмы оптимизации, например алгоритмы ветвей и границ, имеют экспоненциальную трудоемкость. В связи с высокой трудоемкостью оптимизации в строгом смысле опре-

деленный интерес представляют методы приближенной оптимизации и, в частности, методы эвристического типа. Последние, как правило, имеют полиномиальную оценку трудоемкости, и основной интерес в их анализе представляет отыскание гарантированной величины погрешности по функционалу. В главе III «Вариации длительности кратчайших расписаний. Приближенные алгоритмы с оценкой погрешности» представлен ряд наиболее общих результатов такого рода.

Собственно методы и алгоритмы последовательной оптимизации представлены в главах IV—VI. В главе IV «Динамическое программирование» алгоритм динамического программирования иллюстрируется на примере сетевой задачи определения кратчайшего расписания, хотя та же общая схема пригодна и для функционалов более общего вида. Приводятся асимптотические оценки трудоемкости алгоритмов динамического программирования в экстремальных задачах на ациклических графах. В главе V «Последовательный анализ вариантов» алгоритм последовательного анализа вариантов иллюстрируется на примере задачи объемного календарного планирования с аддитивно-сепарабельной целевой функцией. Приводятся оценки трудоемкости в среднем алгоритмов последовательного анализа вариантов, основанные на аналогии результатов соответствующих процедур выбора со статистиками крайних значений. И наконец, наиболее обширная глава VI «Метод ветвей и границ» посвящена, по-видимому, самому перспективному в практическом отношении методу решения задач рассматриваемого класса — методу ветвей и границ. Алгоритмы ветвей и границ обладают рядом структурных особенностей, которые делают их пригодными для решения задач большой размерности. Они позволяют строить приближенные эффективные алгоритмы с апостериорной оценкой погрешности. Как известно, качество алгоритмов этого класса существенно зависит от способов вычисления оценок (границ). Сетевые задачи оптимального распределения ресурсов обладают рядом особенностей, которые могут быть эффектив-

но использованы для вычисления оценок. Систематически изучены два направления, позволяющие эффективно вычислять соответствующие оценки. Одно из них использует специальные функции, определенные на сети операций. Второе основано на использовании современных идей теории двойственности и негладкой оптимизации в дискретном программировании. Здесь же приводятся некоторые результаты численных экспериментов.

Каждая глава завершается комментариями, в которых делаются ссылки на использованные источники и родственные результаты.

В выборе содержания монографии и ее разделов мы руководствовались главным образом двумя критериями. В выборе моделей задач мы стремились к наиболее общим формальным моделям оптимизационных задач, нашедшим отражение в специальной литературе, учитывая, что даже такие модели все еще существенно упрощают реальные проблемы. В выборе методов решения задач мы ориентировались на общие схемы упорядоченного перебора вариантов, названные нами методами последовательной оптимизации. Даже с такими ограничениями круг исследователей, в той или иной степени занимавшихся проблематикой монографии, все еще очень широк. И мы заранее благодарны тем из них, кто любезно укажет нам на наши упущения.

В. С. Михалевич, А. И. Кукса

ДИСКРЕТНЫЕ СЕТЕВЫЕ ЗАДАЧИ ОПТИМАЛЬНОГО РАСПРЕДЕЛЕНИЯ РЕСУРСОВ

1.1. Элементы языка сетевых моделей

В очень общих терминах задачи, о которых будет идти речь в этой книге, заключаются в следующем. Задана некоторая совокупность действий, подлежащих выполнению. В ходе выполнения каждого действия загружаются или расходуются определенные средства. Как правило, оказывается, что общего уровня наличных средств недостаточно для одновременного выполнения всех действий. Требуется упорядочить процесс выполнения действий во времени таким образом, чтобы при соблюдении ряда дополнительных условий минимизировать заданную функцию потерь. Выполнение совокупности действий рассматривается как процесс, состояние которого изменяется во времени и может варьироваться в определенных пределах посредством манипулирования заданными средствами. Иными словами, порядок использования средств является управляющей переменной процесса, и речь идет об оптимальном использовании заданных средств во времени с учетом эффективности процесса в целом. Частные задачи указанного типа традиционно рассматривались в детерминированной теории расписаний. Для более точного описания общих задач уточним содержание используемых в дальнейшем терминов.

Время. Время является независимой переменной в описании процесса выполнения действий и их совокупностей. В дальнейшем главным образом будет рассматриваться случай, когда время принимает только целочисленные неотрицательные значения. Отметим, что абсолютное значение единицы измерения времени зависит от объекта планирования и может изменяться в очень широких пределах. Так, планирование внутренних процессов вычислительных систем может осуществляться с точностью до наносекунд, планирование промышленного производства — с точностью до суток, а планирование

национальных программ развития — с точностью до нескольких лет.

Ресурсы. Термин «ресурсы» обозначает средства для достижения результатов отдельных действий и их совокупностей. По ряду причин удобно различать два типа ресурсов: складываемые ресурсы, называемые иногда ресурсами типа «материалы», и нескладываемые ресурсы, или ресурсы типа «мощность». Примерами ресурсов первого типа являются: топливо, сырье, полуфабрикаты, детали и т. п. В экономических приложениях среди ресурсов этого типа иногда выделяют стоимость как интегральный показатель, характеризующий затраты всех остальных видов ресурсов (складываемых и нескладываемых). Складываемые ресурсы непосредственно расходуются в процессе выполнения того или иного действия. Как правило, не будучи использованными в данный момент времени, они могут быть использованы в более поздние моменты времени. Обычно складываемые ресурсы задаются общим объемом или функцией поставок во времени.

Примерами нескладываемых ресурсов являются: трудовые ресурсы, приборы, машины и комплексы технических средств, рабочие помещения и т. п. В отличие от складываемых, нескладываемые ресурсы в процессе выполнения соответствующих действий сами не расходуются, а производят сами или в сочетании с другими ресурсами некоторый расходимый фактор (машино-смены, например), не допускающий складывания. Таким образом, простой нескладываемого ресурса в определенном смысле связаны с потерями. Обычно нескладываемые ресурсы заданы как функции времени. По существу, отношение конкретного вида ресурса к тому или иному типу во многих случаях определяется не физической природой этого ресурса, а способом его использования при реализации того или иного действия. Существенно отличными, однако, являются математические методы для решения экстремальных задач распределения складываемых и нескладываемых ресурсов.

Операции, работы и системы операций. Операция — это физическое действие или процесс, осуществляемый в заданное время посредством заданного набора ресурсов. Каждая операция рассматривается как элементарное, атомарное действие. Она описывается лишь в ее отношении к внешнему окружению, т. е. в форме указания требуемых ресурсов и ее результатов, в то время как

внутренние процессы операции не являются предметом анализа. Осуществление операции состоит в том, что на время ее выполнения операции предоставляются необходимые ресурсы. Характеризуя процесс обслуживания, в дальнейшем будем полагать, что выражения «ресурсы обслуживают операции», «ресурсы используются на операциях», «операции потребляют ресурсы» и т. п. имеют один и тот же смысл. Ресурсы, обслуживающие данную операцию, не используются в то же самое время на других операциях. Формально каждая операция может быть описана своим именем или номером, списком имен используемых ею ресурсов и их количеством, а также длительностью. Длительность операции может зависеть от абсолютного (календарного) времени ее начала.

Как правило, операции являются элементами более общих целенаправленных действий или их совокупностей. В зависимости от содержания такие целенаправленные совокупности операций называются работами или системами (комплексами) операций. Обычно операции представляют собой определенные звенья, этапы некоторых физических, технологических, производственных процессов. Часто оказывается, что совокупности операций, образующих работы или системы операций, зависимы в том смысле, что в соответствии с требованиями реального процесса, отображаемого данной системой операций, одни операции должны следовать во времени за другими. Иначе говоря, на множестве операций, образующих систему операций, определено бинарное отношение предшествования. В свою очередь, работы или системы операций могут объединяться в более общие, содержащие их системы или комплексы операций и т. д. Таким образом, объединение операций в некоторую систему связано с определенной интерпретацией их как целенаправленных совокупностей действий, имеющих общие результаты. Иногда для обозначения общности результатов отдельных групп операций используют термин «событие».

Расписания. Термин «расписание» используется для обозначения порядка выполнения операций или работ во времени. В случае непрерываемых операций расписанием операции является пара чисел (календарное время начала операции, длительность операции). Соответственно, расписание работы или системы операций достаточно описывать списком времен начала входящих в них операций и их длительностями. Расписание называется допустимым, если:

1. В каждый момент времени суммарные уровни использования ресурсов каждого вида по всем операциям, выполняющимся в данное время, не превосходят общих уровней наличия соответствующих ресурсов.

2. Выполнены технологические ограничения, т. е. для каждой пары «связанных» операций последующая начинается не ранее окончания предыдущей.

3. Выполнены специальные условия. Специальные условия чаще всего регламентируют сроки начала или завершения отдельных операций, работ, сроки достижения определенных событий, взаимное расположение во времени отдельных групп операций и т. п.

Цели и задачи планирования. Наиболее известные цели календарного планирования сводятся к следующим:

А. Простейшая цель состоит в минимизации общего времени выполнения всех операций заданной системы операций при заданных общих уровнях наличия ресурсов каждого типа и вида. Родственная экстремальная задача состоит в минимизации максимального общего уровня потребления ресурсов при условии, что сроки завершения всех операций не превосходят заданной величины.

Б. Во многих практических задачах требуется минимизировать некоторую неубывающую по срокам завершения отдельных операций аддитивно-сепарабельную функцию потерь.

В. Наиболее общий тип целей в задачах календарного планирования отражает два ряда факторов. Первый фактор учитывает условные потери вследствие того, что операции или события осуществляются в те или иные сроки; второй фактор учитывает потери вследствие того, что ресурсы в некоторые моменты времени используются слишком интенсивно или недостаточно интенсивно. Последний фактор особенно важен в случае трудовых ресурсов.

Таким образом, в наиболее общем виде сетевые задачи оптимального распределения (использования) ресурсов определяются указанием:

- системы взаимосвязанных операций или работ;
- номенклатуры и уровней наличия ресурсов во времени;
- перечня специальных ограничений;
- цели планирования

и состоят в построении некоторого допустимого расписания, на котором заданная функция потерь достигает минимального значения.

В следующих разделах данной главы будут охарактеризованы наиболее известные области применения задач указанного типа и конкретизированы соответствующие задачи.

1.2. Сетевое планирование с ограниченными ресурсами

В практике народного хозяйства системы сетевого планирования и управления (СПУ) используются как инструмент для решения организационных задач планирования и управления комплексами работ на основе построения, анализа и оптимизации сетевых моделей работ. В настоящее время имеется значительный опыт разработки и использования систем СПУ в различных областях народного хозяйства. Наиболее целесообразными областями использования СПУ являются:

а) Целевые разработки сложных систем, в которых принимают участие организации и предприятия различных ведомств, включая научно-исследовательские и опытно-конструкторские работы, проектирование, опытное производство, испытания и т. д.

б) Государственные, межведомственные и региональные программы, например, охраны окружающей среды, развития региона и т. д.

в) Основная деятельность таких организаций и предприятий, как ОКБ, НИИ, проектные институты, предприятия опытного и мелкосерийного производства.

г) Строительство и монтаж промышленных и гражданских объектов (заводов, электростанций, шахт, мостов, дорог, жилых и гражданских зданий и т. п.).

д) Подготовка и освоение производства новых видов продукции.

е) Реконструкция и ремонт промышленных и гражданских объектов.

ж) Подготовка и проведение крупных организационных мероприятий (съездов, конференций, кампаний по предупреждению стихийных бедствий).

з) Разведка и освоение новых месторождений полезных ископаемых.

и) Ремонт промышленного оборудования и средств транспорта.

к) Материально-техническое снабжение крупных промышленных и гражданских объектов.

Как отмечалось, общей основой всех систем СПУ является использование сетевых моделей сложных состав-

ных работ — комплексов работ. Работа (операция) — основной элемент комплекса. Наиболее важная характеристика работы связана с понятием ее объема. Реальный смысл понятия объема работы может быть весьма различным: трудоемкость в человеко-днях или машино-сменах, тонно-километры, продолжительность в единицах времени, стоимость и т. д. В выбранных единицах объем представляется чаще всего скалярной величиной W . Объем работы во многих случаях точно известен до начала ее выполнения; в других случаях объем зависит от условий реализации (например, трудоемкость земляных работ может зависеть от применяемых средств механизации, а также от природных условий, времени года и т. п.); наконец, этот объем может быть случайной величиной из-за присущей самой работе неопределенности (трудоемкость поискового научного исследования и т. п.). Так как в общем случае работа есть процесс, происходящий во времени, можно говорить об объеме работы $W(t)$, выполненном к моменту времени t . Величина $W(t)$ является характеристикой состояния работы в момент времени t и не обязательно совпадает с какой-либо мерой труда, затраченного на достижение этого состояния. Формально выполнение работы состоит в изменении $W(t)$ от 0 до W . Скорость изменения функции $W(t)$ есть скорость ведения данной работы. Можно говорить о скорости ведения работы в данный момент времени t или о средней скорости за данный промежуток времени. В общем случае скорость выполнения работы является функцией времени и зависит от ряда других факторов (интенсивность потребления ресурсов, метеорологические условия и т. д.). В наиболее простых и распространенных случаях скорость выполнения работы принимают постоянной; при этом основной характеристикой работы становится ее продолжительность, играющая в этом частном случае роль объема.

Потребность отдельной работы в складываемом ресурсе l -го вида может быть задана числом r^l , равным общему количеству этого вида ресурса (в натуральных единицах), необходимому для выполнения работы в полном объеме. Более полная информация может быть дана с помощью функции, выражающей потребное количество ресурса l -го вида в зависимости от состояния работы. Такая функция, называемая графиком потребления ресурса данной работой, не убывает с возрастанием объема и равна r^l , когда объем достигает W . Потребность отдель-