Оглавление

Введение

Одной из задач, которые всегда и во всех сферах своей деятельности приходится решать человеку, является принятие решений. Постоянная необходимость поиска решений сопровождает всю историю человека.

В любые времена люди, приступая к своим делам, раздумывали над их организацией и выбирали те или другие способы осуществления задуманного мероприятия. Обычно решения принимались без глубокого анализа на основе опыта, интуиции и здравого смысла.

И сегодня при решении бытовых вопросов и принятии решений человек обходится без специальных расчётов, опираясь на те же рычаги. Ошибки в выборе наилучших вариантов здесь приводят к ущербам, касающимся только одного человека или небольшого круга людей.

Однако в производственной деятельности при управлении сложными технологическими процессами здравого смысла и интуиции при поиске наилучших решений уже не хватает и требуется более обоснованный и ответственный подход, ибо ошибки в этом случае могут привести к тяжёлым последствиям.

Для анализа решений на этапе выбора их целесообразно привлекать математическую науку, что позволит ускорить процесс выбора и избежать негативных последствий от неудачных решений.

Наибольшие трудности возникают, когда необходимо планировать новые мероприятия, опыта проведения которых ещё не существует, и опереться здравому смыслу не на что. Здесь только расчёты помогут избежать ошибок в выборе решений.

Очевидно, чем сложнее производственная система и чем больше вкладывается в развитие её материальных ресурсов, тем менее допустимы решения, принимаемые волевым путём без опоры на научные методы и расчёты.

Это в первую очередь относится к энергетике, наиболее сложной и капиталоёмкой производственной системе, от надёжности и экономичности которой зависит работа всех отраслей промышленности, сельского хозяйства и условия жизни людей.

Современные математические методы, облегчающие людям поиск наилучших решений, и являются предметом науки «Исследование операций». Это сравнительно молодая наука. Её возникновение относится к периоду второй мировой войны, когда в армиях США и Англии были сформированы научные группы для принятия решений по организации боевых операций.

В настоящее время исследование операций – одна из самых быстро развивающихся наук, которая находит всё более обширные области применения: промышленность, сельское хозяйство, транспорт, торговля и т.д.

Для решения практических задач в исследовании операций используется целый ряд разделов прикладной математики, среди которых математические методы оптимизации, теория случайных процессов, теория массового обслуживания, теория игр и многие другие.

Из всего многообразия методов здесь будут рассмотрены только наиболее простые и распространённые методы исследования операций.

Тема 1. Задачи исследования операций и их решение

1.1. Предмет исследования операций

Сегодня трудно найти такую область практической деятельности человека, где бы не применялись в разной степени математические модели и методы исследования операций.

Чтобы получить представление о специфике этой науки, рассмотрим ряд типичных для неё задач. Эти задачи, хотя и рассматриваются в упрощённой постановке и взяты из разных предметных областей, имеют практический интерес и дают представление о том, каков предмет и каковы цели исследования операций.

1. Планирование развития энергосистемы. Имеется ряд площадок для сооружения электростанций, необходимых для снабжения электроэнергией новых потребителей. Известны удельные затраты на сооружение электростанций и транспорт электроэнергии. Надо так выбирать тип и мощность блоков, чтобы покрыть дефицит мощности при минимальных затратах.
2. Определение оптимального режима тепловой электрической станции. Имеется ТЭС с заданным составом работающих блоков и известной суммарной нагрузкой, определяемой диспетчером энергосистемы. Необходимо так распределить нагрузку между блоками ТЭС, чтобы обеспечить диспетчерский график выдачи мощности с минимальным расходом топлива.
3. Раскрой металлических полос. При производстве комплектных распределительных устройств требуются отрезки шин нескольких размеров, которые нарезаются из алюминиевых или медных полос различной длины. Необходимо так раскроить имеющиеся полосы, чтобы укомплектовать максимальное число устройств, и свести до минимума обрезки.
4. Планирование производства. Предприятие выпускает трансформаторы разных типов, продажа которых даёт различную прибыль. При производстве используются различные материалы: медный или алюминиевый провод, трансформаторное железо, изоляция и другие материалы, запасы которых известны. Необходимо так спланировать

производство, чтобы при имеющихся ресурсах обеспечить максимальную прибыль.

5. План снабжения предприятий сырьем. Имеется ряд предприятий, использующих известные виды сырья, и есть несколько сырьевых баз, которые могут поставлять это сырьё предприятиям. Базы связаны с потребителями имеющимися путями сообщения разных типов известными тарифами. Требуется найти такой план снабжения предприятий сырьём, в котором затраты на транспорт были бы минимальными.
6. Задачи о пищевом рационе. Имеется несколько видов продуктов питания с известной стоимостью каждого. Требуется составить из этих продуктов пищевой рацион, который должен содержать определённое количество различных составляющих (белки, жиры, углеводы и т.п.), необходимых для сбалансированного питания. Рацион должен обеспечить минимальные затраты.
7. Задача о перехвате самолёта противника. Нарушено воздушное пространство страны. Группа истребителей поднята в воздух для перехвата одиночного самолёта противника. Цель операции – сбить самолёт.

Подобных примеров можно привести гораздо больше, но и этих доста точно, чтобы отметить общие черты наиболее популярного класса задач исследования операций. В каждом из примеров речь идёт о мероприятии, имеющем определённую цель. При этом известны некоторые условия и параметры, характеризующие обстановку, с учётом которых требуется принять такое решение, при котором мероприятие будет осуществлено наиболее эффективным способом.

Под эффективностью операции понимают степень выполнения поставленной цели. Чтобы судить об эффективности операции и сравнивать между собой различные планы организации её, нужно иметь определённый численный критерий или показатель эффективности. В рассмотренных примерах это в основном затраты, реже – прибыль.

Некоторые операции, например рассмотренные в последнем примере, содержат элементы случайности. В этих случаях исход операции и эффективность её оценивается вероятностью достижения цели, например, поражения самолёта противника.

Во всех рассмотренных примерах требуется обеспечить минимум или максимум показателя эффективности.

Часто реальные задачи исследования операций наряду с известными, так называемыми детерминированными параметрами, имеющими определённые количественные значения, описываются и неопределёнными, заранее неизвестными факторами.

Наличие неопределённых факторов переводит задачу в область принятия решений в условиях неопределённости. При неизвестных условиях операции нет возможности также успешно оптимизировать решение, как и в случае с детерминированной информацией, и для ожидания в этих условиях хороших решений нет оснований. Один из специалистов по исследованию операций Т.Л. Саати говорил с иронией о своём предмете: «Исследование операций представляет собой искусство давать плохие ответы на практические вопросы, на которые даются ещё худшие ответы другими методами».

Задачи принятия решений в условиях неопределённости встречаются довольно часто. Обычно причиной неопределённости являются погодные условия, покупательский спрос, действия конкурента на рынке и т.п. Методы исследования операций и предназначены для принятия обоснованных решений во всех рассмотренных ситуациях и им подобных.

1.2. Основные понятия и методы исследования операций

Рассмотрим некоторые термины, основные понятия и принципы исследования операций.

Операцией называют всякое мероприятие или систему действий, объединённое единым замыслом и направленное на достижение определённой цели. Операция всегда управляемое мероприятие, некоторые параметры которого можно выбирать на этапе планирования мероприятия.

Всякий определённый набор зависящих от нас параметров называется решением. Оптимальными называются решения, в которых обеспечивается максимум или минимум критерия, количественно определяющего эффективность операции.

Для применения математических методов исследования, необходима математическая формализация операции путём формирования математической модели. Как известно, математической моделью операции называют совокупность переменных и систем уравнений и неравенств, которые описывают основные свойства исследуемого объекта и его внутренние и внешние связи, а также целевой функции, формализующей критерий оптимизации.

Требования к математической модели достаточно противоречивы. С одной стороны она должна быть достаточно полной и учитывать все основные свойства рассматриваемой системы. Однако стремление учесть и несущественные обстоятельства может привести к переусложнению модели.

С другой стороны модель должна быть достаточно простой, удобной для анализа математическими методами. Однако подчинение этому требованию может переупростить модель и результаты решения её не будут иметь практического значения.

Искусство составления математических моделей заключается в поиске своеобразной узкой «тропы» между «болотом переусложнения» и «западнями переупрощения». Опыт в составлении моделей приобретается постепенно.

Построение математической модели – наиболее ответственная часть, требующая подчас глубокого понимания свойств моделируемой системы и сущности технологических процессов, происходящих в ней.

Математические модели, применяемые в задачах исследования операций, можно разделить на два класса: аналитические и статистические. В аналитических моделях свойства системы и зависимости между параметрами описываются с учётом принимаемых допущений формулами в виде линейных и нелинейных алгебраических уравнений. С помощью аналитических моделей удаётся получать удовлетворительные решения для операций с детерминированной информацией.

В сложных операциях, где переплетаются действия многих факторов, включая и случайные, целесообразно использование методов статистического моделирования. Суть этих методов состоит в том, что процесс развития операций «проигрывается» на ЭВМ с учётом всех случайностей, имитируемых с помощью генераторов «псевдослучайных» чисел.

Решение практических задач исследования операций обычно требует комплексного участия специалистов технологов, компетентных в описании предметной области; математиков, хорошо ориентирующихся в современных методах исследования операций; экономистов, способных сформулировать критерии оптимизации, и специалистов по информационным технологиям, ориентирующихся в программном обеспечении современных ЭВМ.

Окончательное принятие решений является компетенцией ответственного лица, обычно заказчика и постановщика задачи, которому предоставлено право окончательного выбора решения. Сокращённо его называют ЛПР – лицо, принимающее решение. При окончательном выборе ЛПР или группа лиц наряду с рекомендациями, вытекающими их анализа математической модели, могут учитывать ряд свойств, которые не удалось формализовать на этапе составления модели.

По результатам анализа решения модели ЛПР может поставить перед исследователями задачу об уточнении или изменении модели. В этом случае процесс исследования операций носит итерационный характер.

В результате решения детерминированной задачи исследования операций находят совокупность параметров, называемых элементами решения, в качестве которых могут фигурировать числа, векторы, функции, состав оборудования, тип его и т.п. Например, в задаче планирования перевозок однотипных грузов от поставщиков к потребителям элементами решения будут числа , определяющие объём груза, перевозимого от i-го поставщика к j-му потребителю. Совокупность чисел, объединяемая в вектор, и образует решение.

Кроме элементов решения, которые являются переменными математической модели, в любой задаче исследования операций имеются ещё и заданные условия, выполнение которых определяет допустимую область решений, называемую «множеством возможных решений».

Задача определения оптимального решения заключается в поиске на множестве возможных решений такого, которое эффективнее других с позиций какого-либо критерия. Такая задача исследования операций относится к категории обратных задач. Обратные задачи направлены на поиск решения, в котором показатель эффективности достигает своего минимального или максимального значения.

Прямые задачи позволяют определить, что будет, если в заданных условиях мы примем какое-либо решение X из числа возможных. В первую очередь интересует величина критерия эффективности.

Естественно, прямые задачи значительно проще обратных. Понятно, что для решения обратной задачи надо уметь решать и прямую.

В предлагаемом курсе будут рассматриваться в основном обратные задачи, формулировка которых в самом общем виде приводится ниже.

Пусть имеется некоторая операция, на которую можно влиять, выбирая решение X, состоящее из группы параметров. Полагаем, что эффективность операции определяется одним показателем W.

Рассмотрим «детерминированный» случай, когда известны условия операции. В этом случае все, влияющие на операции, факторы делятся на две группы:

заданные условия выполнения операции, образующие вектор A; элементы решения X, выбираемые исследователем.

Первая группа факторов определяет допустимую область решений D. Показатель эффективности – критерий зависит от обеих групп факторов и определяется функцией W = W(A,X).

Обратная задача формулируется следующим образом.

При заданном множестве условий A найти такое решение X* , которое обеспечивает экстремум показателя эффективности $W^* = \underset{x \in D}{\mathrm{extr}}\{W(A, X)\}$.

Таким образом, задача исследования операций формулируется как типичная математическая задача нахождения максимума или минимума функции F(X) на множестве допустимых решений, определяемом вектор-функцией G(X)≥0. Эта задача принадлежит к классу задач математического программирования.

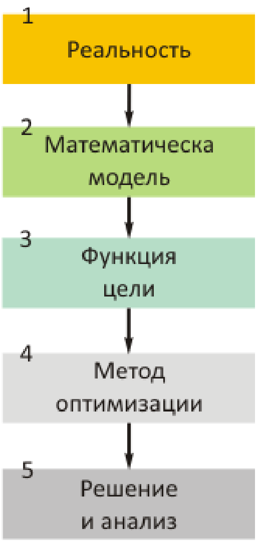


Рисунок 1.1 – Алгоритм решения задач исследования операций

Характерной особенностью задач исследования операций является алгоритм решения, показанный на рисунке 1.1. Задачи ИО всегда порождаются реальной человеческой деятельностью, что и отражается в блоке 1.

Для использования современных методов математики и технических средств обработки информации обязательна математическая формализация, описывающая основные свойства объекта исследования и его связи.

Важное значение имеет также определение целей функционирования системы и формирование функции, позволяющей качественно оценить эффективность решения. При составлении модели учитывается ориентация на выбранный математический метод оптимизации. И, наконец, выполняется решение задачи на ЭВМ с последующим анализом его и реализацией на реальном объекте.

1.3. Многокритериальные задачи и системный анализ

При планировании мероприятий по развитию больших и сложных технических или любых других систем часто возникают затруднения с оценкой эффективности и формализацией её в виде функции цели.

Оказывается одним критерием оценить многообразие целей функционирования системы удаётся далеко не всегда.

Например, планируется работа энергетической системы. Какие цели необходимо учитывать при выборе решения? Прежде всего хотелось бы максимально увеличить выработку электроэнергии на электростанциях системы, получить наибольший чистый доход от продажи тепловой и электрической энергии, снизить себестоимость, обеспечить показатели качества отпускаемой продукции, свести до минимума ущерба от перерывов электроснабжения. При глубоком осмысливании задачи могут возникнуть ещё и другие критерии.

Многокритериальность, из которых одни критерии следует обратить в максимум, а другие в минимум, характерна для большинства задач исследования операций.

В этих условиях невозможно найти решение, удовлетворяющее одновременно всем критериям. Решение, обеспечивающее максимальную выработку на тепловых электростанциях, не обращает в минимум вредное влияние выбросов на экологию. Не является, например, корректным лозунг времён плановой экономики «Больше продукции при меньших затратах».

Как выбрать решение в таких ситуациях. Чаще всего решения принимаются на основе использования предшествующего опыта и интуиции. Обычно обеспечить выполнение разных целей невозможно и приходится жертвовать полным обеспечением одних во имя выполнения других, т.е. принимать компромиссные решения.

В проблеме поиска компромиссных решений важную роль играет понятие множества Парето - оптимальных решений. Это множество используется в тех случаях, когда разные критерии не сопоставимы и улучшение решения по одному критерию обязательно ухудшает хотя бы один другой критерий. Каждое решение, принадлежащее множеству Парето, лучше других по одним критериям и хуже по другим.

Для примера рассмотрим операцию, которая оценивается двумя противоречивыми критериями, один из которых – затраты S минимизируется, а другой – прибыль W требуется максимизировать.

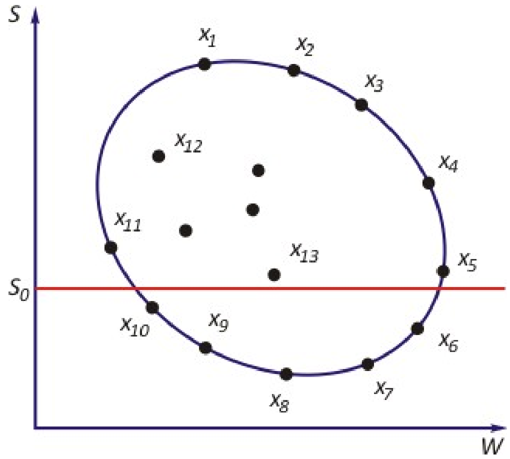


Рисунок 1.2 – Область допустимых решений

Пусть всё множество допустимых решений образует область, на которой показано несколько возможных решений x_i в координатах W и S (рисунок 1.2).

Решения, лучшие по прибыли лежат на правой границе допустимой области, образуя множество M_1, в которое входят варианты x_1, x_2,..., x_7.

Решения, лучшие по затратам, лежат на нижней границе, образуя множество M_2 с вариантами x_5, x_6, x_7,..., x_{11}.

Парето-множество образуется пересечением $П = M_1 \cap M_2$ и включает решения x_5, x_6, x_7, среди которых лучшее по прибыли – x_5, лучшее по затратам – x_7 и решения с промежуточными показателями.

Совершенно очевидно, что дальнейший поиск на основе каких-либо математических или эвристических процедур следует проводить на множестве Парето.

Задача лица, принимающего решение, отдать предпочтение одному из этих вариантов x_5, x_7 или выбрать компромиссный – x_6, взяв на себя ответственность за окончательный выбор.

Рассмотрим некоторые возможные формальные процедуры, позволяющие выбирать решения на множестве Парето.

Самая естественная процедура заключается в формировании некоторого обобщённого показатели эффективности, в котором каждый критерий W_i участвует со своим весовым коэффициентом α_i, отражающим его важность:

$$W = \alpha_1 W_1 + \alpha_2 W_2 + \ldots = \sum_{(i)} \alpha_i W_i$$

Этот показатель формируется по правилам вычисления скалярного произведения двух векторов, что и послужило основанием для названия процедуры методом скаляризации. Коэффициенты важности α_i определяются обычно экспертным путём.

Существует ещё один путь, называемый методом предпочтительного критерия, позволяющий привести задачу к однокритериальной. Здесь выделяется один главный критерий, например W_i, а все остальные учитываются в виде ограничений, дополняющих математическую модель.

Например, в рассмотренном примере (см. рисунок 1.2) можно максимизировать прибыль, наложив ограничения на затраты $S \leq S_0$, что позволит однозначно выбрать в качестве наилучшего решение x_6.

Для поиска компромиссного решения может использоваться ещё один приём, названный методом последовательных уступок. Здесь все критерии W_1, W_2 располагаются в порядке убывания важности, например W_1, W_2 и т.д. Сначала ищется наилучшее решение по критерию W_1, например, $\dot{W}_1 = \max W_1$.

Затем назначается на основе опыта, исходя их практических соображений, с учётом малой точности исходных данных, некоторая уступка ΔW_1, которую можно допустить для того, чтобы оптимизировать второй критерий W_2:

$$\operatorname{extr} W_2 \text{ при } W_1 \geq \dot{W}_1 - \Delta W_1$$

При любом методе поиска наилучшего решения на множестве Парето задача сводится к оптимизации по одному критерию, что открывает широкие возможности для применения соответствующих математических методов поиска оптимальных решений, разработанных для однокритериальных задач.

Однако и переход к однокритериальной оптимизации не упрощает задачу выбора оптимальных решений для сложных технико-экономических систем, к которым относится и энергетика. Такие системы с позиций управления относят к большим системам.

Анализ больших систем возможен только на основе системного подхода.

Системный подход опирается на признание объективного характера всеобщей связи, причинной обусловленности явлений, господства необходимости, правильного сочетания необходимости и случайности [8]. Методология системного подхода опирается на следующие главные положения:

1. окружающая действительность должна изучаться в единстве, целостности и развитии;

2. наши знания окружающей действительности всегда относительны, ибо они отражают неполную информацию о событиях, явлениях и свойствах системы;
3. в природе и в обществе всегда прослеживаются причинно-следственные связи, которые являются проявлением объективных законов развития природы.

Системный подход реализуется путём сочетания следующих принципов исследования:

- система должна рассматриваться как единое целое;
- система всегда, с одной стороны, находится в окружении других систем разных типов и испытывает на себе их влияние, с другой стороны, она находится на некотором уровне иерархии систем данного типа и её управление также обладает свойствами иерархичности;
- в основе оптимизации должны лежать предварительно и достаточно чётко сформулированные цели функционирования системы;
- оптимизационная модель системы должна учитывать все определяющие свойства системы, а также связи с другими системами и окружающей средой;
- полученное решение в ходе его реализации должно корректироваться и дополняться с учётом вновь появляющихся обстоятельств и уточнений информации.

Инструментом реализации системного подхода при исследовании больших систем является системный анализ. Системный анализ предполагает, что для получения решения необходимо выполнить следующие основные этапы исследования.

1. Постановка задачи, заключающаяся в выборе исследуемой системы, локализации её, т.е. определении границ и связей с соседними системами, формулировке целей управления.
2. Составление математической модели, т.е. формализация внутренних свойств системы и её внешних связей, формирующих систему ограничений G(X) и целевой функционал F(X).
3. Выбор метода решения модели, ориентация на который должна учитываться и при составлении математической модели.
4. Решение модели на ЭВМ и определение по полученным результатам возможных оптимальных путей развития системы в зависимости от управляющих возможностей.

5. Планирование оптимального развития системы постановщиком задачи – лицом принимающим решение или группой лиц с учётом полученных в результате решения рекомендаций, а также тех обстоятельств, которые не были формализованы при разработке математической модели.

В процессе системного анализа важное значение имеет согласованность математической модели и методов её решения. Разработка модели и выбор метода решения всегда выполняются совместно и согласованно. Вопросы составления математических моделей для энергетических систем и их объектов будут рассматриваться позже в специальном курсе «Модели оптимального развития энергосистем и САПР».

Изложение сведений по методам оптимизации будет проведено в следующих главах учебного пособия.

Задача однокритериальной оптимизации математически записывается предельно компактно

$$F(X) \to \text{extr при } G(X) \geq 0$$

где X – вектор неизвестных параметров,

F – целевой функционал,

G(X) – вектор-функция ограничений, определяющая допустимую область решений.

Методы оптимизации в математической науке разработаны достаточно полно и составляют целый раздел прикладной математики, называемый «Математическое программирование», хотя используемый здесь термин и не имеет никакого отношения к теории программирования.

Изучение методов оптимизации начнём с простейшей линейной оптимизации.

Тема 2. Линейное программирование

2.1. Примеры задач линейного программирования

В линейном программировании функции F(X) и g(X) описываются линейными выражениями, например:

$$F(X) = c_1 x_1 + c_2 x_2 + c_3 x_3 + \ldots,$$
$$g_1(X) = a_{11} x_1 + a_{12} x_1 + \ldots + a_{1n} x_n - b_1 > 0 \quad (< 0, = 0)$$

(2.1)

Многие практические задачи достаточно точно могут быть описаны как линейные.

Пример 1. **Планирование рациона.** Для обеспечения сбалансированного питания необходимо обеспечить суточные потребности организма в белках, жирах и углеводах. На выбор имеются четыре продукта питания: $П_1$, $П_2$, $П_3$, $П_4$. Известны стоимость каждого продукта C_i, (i = 1, 2, 3, 4) и суточная потребность в белке B_1, углеводах B_2 и жирах B_3. Содержание этих составляющих в каждом продукте a_{i1}, a_{i2}, a_{i3}, a_{i4} образует матрицу

$$A = \begin{bmatrix} a_{11} & a_{21} & a_{31} & a_{41} \\ a_{12} & a_{22} & a_{32} & a_{42} \\ a_{13} & a_{23} & a_{33} & a_{43} \end{bmatrix}$$

Необходимо определить такой рацион, который обеспечит потребности в белках, жирах и углеводах при минимальных затратах.

Составим математическую модель:

1. обозначим через x_1, x_2, x_3, x_4 неизвестные объемы продуктов питания,
2. опишем ограничения:

$$G(X) \begin{cases} a_{11}x_1 + a_{21}x_2 + a_{31}x_3 + a_{41}x_4 \ge B_1 \\ a_{12}x_1 + a_{22}x_2 + a_{32}x_3 + a_{42}x_4 \ge B_2 \\ a_{13}x_1 + a_{23}x_2 + a_{33}x_3 + a_{43}x_4 \ge B_3 \end{cases}$$

(2.2)

3. целевая функция:

$$F(X) = c_1 x_1 + c_2 x_2 + c_3 x_3 + c_4 x_4 \to \min$$

В задаче требуется найти такое неотрицательное решение ($x_i \ge 0$), которое удовлетворяет ограничениям и обеспечивает минимум F(X).

Пример 2. **Планирование производства.** На предприятии, выпускающем провода марки АС-70, АС-95, АС-120 и АС-150 имеются две группы установок: первого типа в количестве N1 и второго – N2. Каждая установка может выпускать любой провод из этой номенклатуры, но с разной производительностью a_{ij}, заданной в таблице 2.1.

Таблица 2.1

	АС-70	**АС-95**	**АС-120**	**АС-150**
1	a_{11}	a_{12}	a_{13}	a_{14}
2	a_{21}	a_{22}	a_{23}	a_{24}

где i – номер группы;

j – виды марок.

Известна прибыль от реализации 1 км каждого типа провода: c_1, c_2, c_3, c_4.

Задан плановый выпуск каждого типа провода: b_1, b_2, b_3, b_4.

Считается также, что все установки должны быть в работе и заданный план можно перевыполнять.

Необходимо распределить производство проводов между установками таким образом, чтобы обеспечить максимальную прибыль.

Математическая модель:

1. x_{ij} – число станков i-го типа, выделяемых на выпуск каждого j-го провода:

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \end{bmatrix}$$

2. Условие по количеству установок:

$$\begin{array}{l} x_{11} \quad x_{12} \quad x_{13} \quad x_{14} = N_1 \\ x_{21} \quad x_{22} \quad x_{23} \quad x_{24} = N_2 \end{array}$$

(2.3)

и условие выполнения плана:

$$\begin{array}{l} a_{11}x_{11} + a_{21}x_{21} \ge b_1 \\ a_{12}x_{12} + a_{22}x_{22} \ge b_2 \\ a_{13}x_{13} + a_{23}x_{23} \ge b_3 \\ a_{14}x_{14} + a_{24}x_{24} \ge b_4 \end{array}$$

(2.4)

3. Целевая функция:

$$\begin{array}{l} F(X) = c_1\left(a_{11}x_{11} + a_{21}x_{21}\right) + \left(a_{12}x_{12} + a_{22}x_{22}\right) + \\ + c_3\left(a_{13}x_{13} + a_{23}x_{23}\right) + c_4\left(a_{14}x_{14} + a_{24}x_{24}\right) = \\ = \beta_{11}x_{11} + \beta_{12}x_{12} + \beta_{13}x_{13} + \beta_{14}x_{14} + \beta_{21}x_{21} + \beta_{22}x_{22} + \beta_{23}x_{23} + \beta_{24}x_{24} \end{array}$$

(2.5)

Требуется найти такое неотрицательное решение, которое удовлетворяет ограничениям и обеспечивает максимум F(X).

При составлении математической модели учет и описание свойств системы и ее связей производится ограничениями g(X) в виде равенств и неравенств. В задаче линейного программирования, имеющей канонический вид, ограничения представляются в форме равенств. Переход от неравенств к равенствам осуществляется путем введения фиктивных переменных.

Неравенства типа ≤ требуют ввода положительной переменной; например, неравенство $2x_1 + x_2 \le 5$ преобразуется в равенство $2x_1 + x_2 + y = 5$, где $y \ge 0$.

Неравенства типа ≥ требуют ввода положительной переменной, но с отрицательным знаком; например, $2x_1 + x_2 \ge 5$ преобразуется в $2x_1 + x_2 - y = 5$, где $y \ge 0$.

С вводом фиктивных переменных система условий задачи принимает вид $A \cdot X = B$, но число неизвестных при этом возрастает.

В рассмотренных примерах целевая функция имеет разные типы экстремума. В дальнейшем будут рассматриваться методы поиска минимума. Если метод решения ориентирован только на поиск минимума целевой функции, то в случае решения задачи $F(X) \to \max$ достаточно изменить знаки в целевой функции на противоположные и найти решение задачи при $F(-X) \to \min$.

2.2. Общая задача линейного программирования

Общая задача линейного программирования заключается в поиске среди неотрицательных решений системы линейных алгебраических уравнений

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \end{cases}$$

(2.6)

такого, которое обеспечивает минимум линейной функции

$$F(X) = c_1x_1 + c_2x_2 + \ldots + c_nx_n \to \min$$

В матричном виде:

$$A \cdot X = B,$$
$$F(X) = C^T \cdot X \to \min$$

(2.7)

здесь:

$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$ – матрица условий задачи,

$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ – вектор неизвестных,

$B = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$ – вектор правой части,

$C = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$ – вектор коэффициентов целевой функции.

2.3. Алгебра линейного программирования

Напомним некоторые сведения из линейной алгебры, необходимые для понимания линейного программирования.

Операции над матрицами

Пусть A – матрица размерностью m×n (m – строк и n – столбцов).

Транспонирование A^T. Для нахождения транспонированной матрицы у исходной матрицы строки заменяются столбцами и наоборот. Пример:

$$A = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{bmatrix} \Rightarrow A^T = \begin{bmatrix} a_1 & a_4 & a_7 \\ a_2 & a_5 & a_8 \\ a_3 & a_6 & a_9 \end{bmatrix}$$

Обратная матрица:

$$A^{-1} = \frac{1}{A}, \;\; A \cdot A^{-1} = E = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Минор M_{ij} элемента a_{ij} определителя D n-го порядка – это определитель (n–1)-го порядка, получающийся из D вычёркиванием i-й строки и j-го столбца.

Алгебраическое дополнение A_{ij} элемента a_{ij} – это его минор, взятый со знаком $(-1)^{i+j}$:

$$A_{ij} = (-1)^{i+j} M_{ij}$$

Число миноров k-го порядка для матрицы условий задачи A размером m×n определяется числом сочетаний $C_m^k \cdot C_n^k$:

$$N_k = C_m^k \cdot C_n^k = \frac{m!}{k!(m-k)} \cdot \frac{n!}{k!(n-k)}$$

Пример. Определить число миноров 1-го и 2-го порядков для матрицы А

$A = \begin{bmatrix} 2 & 5 & 4 \\ 6 & 7 & 1 \end{bmatrix}$ $N_1 = C_2^1 \cdot C_3^1 = 6$ – это сами элементы матрицы А;

$N_2 = C_2^2 \cdot C_3^2 = \frac{2!}{2! \cdot 0!} \cdot \frac{3!}{2!(3-2)!} = 3$. Вот эти три минора $\begin{bmatrix} 2 & 5 \\ 6 & 7 \end{bmatrix}, \begin{bmatrix} 2 & 4 \\ 6 & 1 \end{bmatrix}, \begin{bmatrix} 5 & 4 \\ 7 & 1 \end{bmatrix}$.

Система алгебраических уравнений

В матричной форме система линейных алгебраических уравнений (СЛАУ) имеет вид

$$AX = B$$

где

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}, \ X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \ B = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

Эта же система в алгебраической форме имеет вид:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \end{cases}$$

В зависимости от соотношения m и n возможны следующие случаи:

1. m = n и определитель матрицы $|A| \neq 0$. В этом случае система совместна и имеет единственное решение. Решение можно найти по правилу Крамера: $x_i = \frac{|A_i|}{|A|}$, где A_i – исходная матрица, в которой i-ый столбик заменен правой частью
$$A_i = \begin{bmatrix} a_{11} & \cdots & b_1 & \cdots & a_{1n} \\ \vdots & & \vdots & & \\ a_{n1} & \cdots & b_n & \cdots & a_{nn} \end{bmatrix}$$
Для решения могут использоваться и многочисленные другие методы. При $|A| = 0$ система несовместна, т.е. не имеет решений.
2. m > n. Как правило, система несовместна так как уравнений больше, чем неизвестных.
3. m < n. Система совместна, если ранг исходной матрицы и ранг расширенной матрицы совпадают (теорема Кронеккера – Капелли).

Как известно, рангом матрицы называют наивысший порядок отличных от нуля миноров этой матрицы. Ранг определяет число линейно независимых строк (столбцов) в матрице, если их рассматривать как вектора. Строка a_1, a_2, …, a_m

линейно независима, если можно подобрать такие числа $\lambda_1, \lambda_2, \ldots, \lambda_m$ не равные нулю одновременно, что

$$\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_m a_m = 0$$

Если это равенство имеет место только при всех $\lambda_i = 0$, то строки $a_1, a_2, \ldots, a_m$ линейно независимы. Например, строки $a_1 = (1, 2, 3)$ и $a_2 = (2, 4, 6)$ линейно зависимы, так как при $\lambda_1 = 2$ и $\lambda_2 = -1$ выполняется условие $\lambda_1 a_1 + \lambda_2 a_2 = 0$.

Если матрица A условий задачи состоит из линейно независимых уравнений, то её ранг $r(A) = m$. Расширенная матрица B получается из A добавлением столбца правых частей $в_i$. При $r(A) = r(B)$ система совместна и имеет бесконечное множество решений. Найти решения можно, если матрицу A разделить на блоки, выделив базисный минор $A_Б$, и оставшийся блок A_C. Соответственно делится и вектор X на базисные $X_Б$ и свободные X_C переменные (рисунок 2.1).

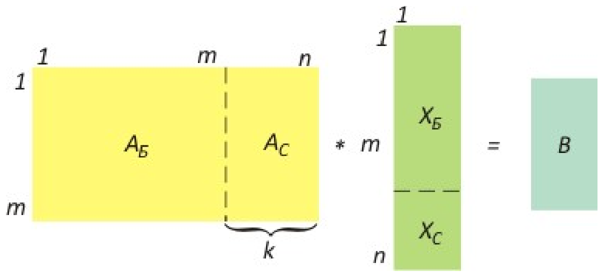


Рисунок 2.1 – Область допустимых решений

Тогда по правилам выполнения операций с блочными матрицами можно записать $A_Б \times X_Б + A_C \times X_C = B$. Откуда можно получить выражения, позволяющие линейно выразить базисные переменные через свободные $X_Б = A_C^{-1}(B - A_C X_C)$.

Свободные переменные X_C могут принимать любые положительные значения, что позволяет однозначно определять и значения базисных переменных, т.е. находить полное решение СЛАУ.

Число таких решений бесконечно. Среди них существуют и недопустимые, в которых одна или несколько базисных становятся отрицательными.

Решение, в котором $X_C = 0$, называют базисным.

Поскольку в качестве свободных могут использоваться любые k неизвестных из n, то число базисных решений определяется числом сочетаний из n по k:

$$N_Б = C_n^k = \frac{k!}{n!(n-k)!}$$

2.4. Геометрия линейного программирования

Рассмотрим геометрический смысл задачи линейного программирования.

В алгебраической форме задача формулируется следующим образом.

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \end{cases}$$

(2.8)

$$F(X) = c_1x_1 + c_2x_2 + \ldots + c_nx_n \rightarrow \min$$

Рассмотрим одно уравнение: $a_{11}x_1+a_{12}x_2+\ldots+a_{1n}x_n = b_1$. Геометрическим местом точек x_i, являющихся решением уравнения, будет некая гиперплоскость, вырождающаяся при n = 2 в прямую, а при n = 3 в плоскость. Эта гиперплоскость делит пространство на два полупространства. Решением системы неравенств является пересечение полупространств, образующее выпуклую область. Выпуклая область – такая область, в которой для двух точек, принадлежащих области, отрезок, их соединяющий, полностью принадлежит ей (рисунок 2.2).

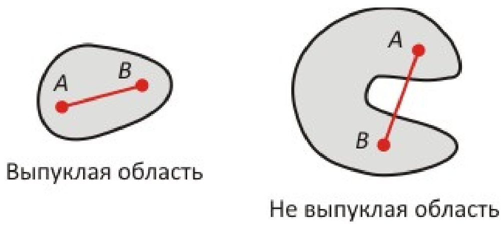


Рисунок 2.2 – Области решений

Пример. Определить допустимую область для системы неравенств.

$$\begin{cases} x_1 + x_2 \geq 2 \\ x_1 - x_2 \leq 2 \\ x_1 \leq 3 \\ x_2 \leq 3 \end{cases}$$

(2.9)

На рисунке 2.3 показаны допустимые полуплоскости для всех неравенств, отмеченных номерами. Допустимая область образует выпуклый многоугольник ABCDE.

В общем случае, геометрическим местом точек, являющихся решением системы линейных уравнений или неравенств будет выпуклый многогранник, образованный пересечением линейных гиперплоскостей.

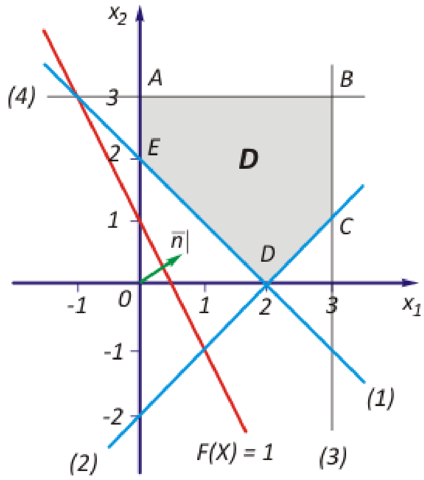


Рисунок 2.3 – Допустимые полуплоскости

Определим теперь геометрический смысл целевой функции F(X).

Рассмотрим две гиперплоскости разных значений целевой функции и получившуюся систему из 2-х уравнений:

$$\begin{cases} c_1x_1 + c_2x_2 + \ldots + c_nx_n = F_1 \\ c_1x_1 + c_2x_2 + \ldots + c_nx_n = F_2 \end{cases}$$

Проверим совместность этой системы по теореме Кронеккера-Капелли.

Ранг основной матрицы системы равен r = 1, т.к. любой определитель второго порядка равен 0. В расширенной же матрице всегда можно найти определитель второго порядка, включающий добавленный столбец со значениями F_1 и F_1, отличный от 0. Значит ранг расширенной матрицы равен 2.

Таким образом, ранги r(A) и r(B) не равны и система не имеет решений. Следовательно, гиперплоскости разных значений целевой функции не пересекаются, т.е. они образуют семейство параллельных гиперплоскостей.

В связи с этим решение задачи линейного программирования находится там, где допустимая область касается гиперплоскости наименьшего значения целевой функции.

В рассмотренном выше примере для допустимой области (см. рисунок 2.3) введем целевую функцию $F(X) = 2x_1+x_2 \rightarrow \min$. При n = 2 целевая функция F представляется семейством параллельных прямых. Направление прямой F(X) = const можно определить по нормали, координаты которой равны коэффициентами при неизвестных, т.е. вектор нормали n = (2,1). Минимум целевой функции достигается в точке E при $x_1 = 0$, $x_2 = 2$.

Таким образом, решение задачи линейного программирования всегда лежит на границе допустимой области и, как правило, в вершине многогранника, иногда на ребре, а в редких случаях – на грани.

Как искать решение, лежащее на границе допустимой области? Методы поиска, основанные на производных, здесь не работают, так как на границе не существует понятия производной. В 1947 году американский математик Данциг предложил симплекс-метод, который основан на последовательном обходе вершин многогранника допустимых решений, начиная с произвольно выбранной, в направлении снижения целевой функции.

Алгебраически каждая вершина многогранника соответствует какому-либо базисному решению, в котором соответствующие ему свободные переменные равны нулю. Переход к соседней вершине осуществляется путем замены только одной свободной переменной на соответствующую базисную и перевода этой базисной в список свободных.

2.5. Идея симплекс-метода

Идею и суть симплекс-метода рассмотрим на примере следующей задачи линейного программирования

$$\begin{cases} x_2 + 2x_4 + 3x_5 = 7 \\ x_3 - x_4 - 3x_5 = 2 \\ x_1 + x_4 + x_5 = 2 \end{cases}$$

(2.10)

$$F(X) = 3 - x_4 + x_5 \to \min$$

Все уравнения задачи линейно независимы, ранги матрицы системы и расширенной равны 3, следовательно, система совместна. Примем в качестве свободных переменных x_4 и x_5.

Выразим базисные переменные x_1, x_2, x_3 через свободные:

$$\begin{cases} x_1 = 2 - x_4 - x_5 \\ x_2 = 7 - 2x_4 - 3x_5 \\ x_3 = 2 + x_4 + 4x_5 \end{cases}$$

(2.11)

$$F(X) = 3 - x_4 + x_5 \to \min$$

Базисное решение: $x_4 = x_5 = 0$, $x_1 = 2$, $x_2 = 7$, $x_3 = 2$, F(X) = 3 является допустимым.

Проанализируем возможности уменьшения целевой функции. При этом имеет смысл рассматривать изменение x_4, x_5 только в сторону их увеличения, так как в линейном программировании переменные не могут быть отрицательными.

Перед x_4 в формуле F(X) стоит знак «–». Следовательно, для снижения F(X) выгодно увеличивать x_4. Переменную $x_5 = 0$ изменять невыгодно, так как любое ее увеличение приведет только к росту целевой функции.

Оценим поведение базисных переменных x_1, x_2, x_3 при росте x_4. Очевидно, что при этом x_3 возрастает, а x_1 и x_2 уменьшаются и могут перейти в недопустимую область, где $x_i<0$.

Проанализируем x_1 и x_2, чтобы определить, которая из них скорее станет равной нулю при увеличении x_4. При $x_5 = 0$ зависимости этих базисных от x_4 будут $x_1 = 2–x_4$ и $x_2 = 7–2x_4$, откуда следует, что $x_2 = 0$ при $x_4 = 3,5$, а $x_1 = 0$ при $x_4 = 2$. Значит рост x_4 будет ограничиваться значением $x_4 = 2$, при котором базисная переменная x_1 станет равной 0, а x_2 останется положительной равной 3.

В новом решении $x_5 = x_1 = 0$, $x_4 = 2$, $x_2 = 3$, $x_3 = 4$ состав свободных изменился. Выразим базисные через новый набор свободных x_5 и x_1:

$$\begin{cases} x_4 = 2 - x_1 - x_5 \\ x_2 = 7 - 2(2 - x_1 - x_5) - 3x_5 \quad = 3 + 2x_1 - x_5 \\ x_3 = 2 + 2 - x_1 - x_5 + 4x_5 \qquad = 4 - x_1 + 3x_5 \end{cases} \tag{2.12}$$

$$F(X) = 3 - 2 + x_1 + x_5 + x_5 = 1 + x_1 + 2x_5$$

При этом новое базисное решение: $x_1 = x_5 = 0$, $x_2 = 3$, $x_3 = 4$, $x_4 = 2$, $F(X) = 1$.

Дальнейший анализ получившегося базисного решения нужно проводить таким же образом. В данном случае свободные переменные входят в целевую функцию с положительными знаками, следовательно, дальнейших возможностей по уменьшению целевой функции нет. Значит, полученное решение является оптимальным.

2.6. Алгебра симплекс-метода

Основная задача линейного программирования в матричной форме:

$$\begin{gathered} A \cdot X = B \\ F = C^T \cdot X \Rightarrow \min \end{gathered} \tag{2.13}$$

Выберем произвольно первое базисное решение, приняв в качестве свободных первые $k = n–m$ неизвестных. Будем считать, что оно допустимо. Выразим базисные переменные и целевую функцию F через свободные. Для этого разделим матрицу A и вектора X и B на соответствующие блоки (рисунок 2.4).

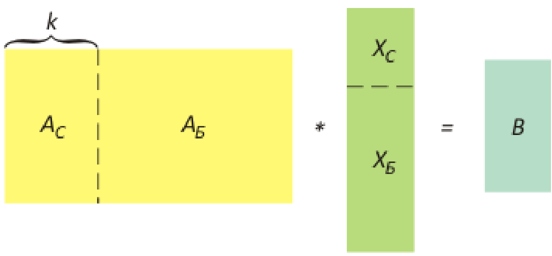


Рисунок 2.4 – Блоки матрицы

По правилам операций над блочными матрицами запишем

$$A_C \cdot X_C + A_Б \cdot X_Б = B$$

откуда

$$X_Б = A_Б^{-1}\left(B - A_C \cdot X_C\right)$$

Аналогичное выражение для целевой функции

$$F = C_Б^T \cdot A_Б^{-1} \cdot B - \left(A_Б^{-1} \cdot A_C \cdot C_Б^T - C_K^T\right) \cdot X_C$$

Запишем матричные выражения для базисных переменных и целевую функцию в алгебраической форме:

$$\begin{cases} x_{k+1} = \beta_1 - \left(\alpha_{11}x_1 + \ldots + \alpha_{1j}x_j + \ldots + \alpha_{1k}x_k\right) \\ \vdots \\ x_{k+i} = \beta_i - \left(\alpha_{i1}x_1 + \ldots + \alpha_{ij}x_j + \ldots + \alpha_{ik}x_k\right) \\ \vdots \\ x_{k+m} = \beta_1 - \left(\alpha_{m1}x_1 + \ldots + \alpha_{mj}x_j + \ldots + \alpha_{mk}x_k\right) \end{cases}$$

(2.14)

$$F = \gamma_0 - \left(\gamma_1 x_1 + \gamma_2 x_2 + \ldots + \gamma_k x_k\right)$$

Эту форму записи называют **стандартной**.

Рассмотрим базисное решение, в котором все свободные $x_1 = \ldots = x_j = \ldots = x_k = 0$. Значения базисных переменных $x_{k+i} = \beta_i$. Полагаем, что все $\beta_i > 0$, т.е. базисное решение $x_{k+j} = \beta_i > 0$ допустимо. При этом целевая функция $F(X) = \gamma_0$.

Рассмотрим внимательнее полученное базисное решение и оценим пути уменьшения целевой функции. Здесь возможны следующие ситуации:

1. $\gamma_j<0$ для всех $j = 1, \ldots, k$. В этом случае увеличение любой свободной переменной от нуля приводит к росту целевой функции. Следовательно, рассматриваемое базисное решение является оптимальным.
2. $\gamma_j>0$ (хотя бы для одного j), а все $\alpha_{ij}<0$. При этом рост x_j, приводящий к снижению целевой функции, сопровождается ростом всех базисных переменных, которые сохраняют свою допустимость. При этом $F \to -\infty$. Этот случай носит чисто теоретический характер. В практических задачах подобная ситуация возможна только при некорректной математической модели.
3. $\gamma_j>0$ для одного или нескольких j и $\alpha_{ij}>0$ для нескольких i. В этом случае увеличение x_j сопровождается уменьшением тех базисных, для которых $\alpha_{ej}>0$, а зависимость от x_j определяется уравнением $x_{k+e} = \beta_e - \alpha_{ej}x_j$. Каждая из этих базисных подходит к границе допустимой области, определяемой условием $x_{k+e} = 0$ при значении $x_j = \beta_e/\alpha_{ej}$. Из всех этих значений выберем минимальное и присвоим ему индекс i.

$$\min\left\{\frac{\beta_e}{\alpha_{ej}}\right\} = \frac{\beta_i}{\alpha_{ij}}$$

Это значение определяет допустимую для роста величину x_j, при которой из всех убывающих базисных первая становится равной нулю.

Коэффициент α_{ij} имеет важное значение, и его называют генеральным. При этом снижение целевой функции определяется произведением

$$\Delta F_j = \gamma_j \min\left\{\frac{\beta_e}{\alpha_{ej}}\right\}$$

В том случае, если есть несколько $\gamma_j > 0$, генеральный элемент должен для ускорения решения выбираться из следующего условия:

$$\max \Delta F_j = \max_j\left\{\gamma_j \min_j\left\{\frac{\beta_e}{\alpha_{ej}}\right\}\right\}$$

(2.15)

После поиска генерального элемента, определяющего по сути пару переменных x_{k+i} и x_j, делается переход к новому базисному решению. В новом наборе свободных переменных x_{k+i} заменяет x_j, которая переходит в список базисных на место x_{k+i}, т.е. эти две переменные меняются местами в списках свободных и базисных.

Таким образом, изменившийся набор свободных включает

$$x_1, \ldots x_{j-1},\ x_{k+i},\ x_{j+1}, \ldots, x_k$$

Найдем выражение базисных и формы F через новый набор свободных переменных. Для этого из выражения для исключаемой базисной

$$x_{k+i} = \beta_i - \left(\alpha_{i1}x_1 + \ldots + \alpha_{ij}x_j + \ldots + \alpha_{ik}x_k\right)$$

(2.16)

найдём значение свободной x_j

$$x_j = \frac{\beta_i}{\alpha_{ij}} - \left(\frac{\alpha_{i1}}{\alpha_{ij}}x_1 + \ldots + \frac{1}{\alpha_{ij}}x_{k+i} + \ldots + \frac{\alpha_{ik}}{\alpha_{ij}}x_k\right)$$

(2.17)

Подставим это выражение во все остальные базисные и проведем необходимые преобразования

$$x_{k+e} = \beta_e - \left(\alpha_{e1}x_1 + \ldots + \alpha_{ej}x_j + \ldots + \alpha_{ek}x_k\right) =$$
$$= \beta_e - \left[\alpha_{e1}x_1 + \ldots + \alpha_{ej}\left(\frac{\beta_i}{\alpha_{ij}} - \frac{\alpha_{i1}}{\alpha_{ij}}x_1 - \ldots - \frac{1}{\alpha_{ij}}x_{k+i} - \ldots - \frac{\alpha_{ik}}{\alpha_{ij}}x_k\right) + \ldots + \alpha_{ek}x_k\right] =$$
$$= \left(\beta_e - \alpha_{ej}\frac{\beta_i}{\alpha_{ij}}\right) - \left\{\left(\alpha_{e1} - \alpha_{ej}\frac{\alpha_{i1}}{\alpha_{ij}}\right)x_1 + \ldots + \left(-\alpha_{ej}\frac{1}{\alpha_{ij}}\right)x_{k+i} + \ldots + \left(\alpha_{ek} - \alpha_{ej}\frac{\alpha_{ik}}{\alpha_{ij}}\right)x_k\right\}$$

(2.18)

Аналогично преобразуется и целевая функция, но при этом в формуле вместо α_{ej} участвует γ_j, а вместо β_e–γ_0

$$F = \left(\gamma_0 - \gamma_j\frac{\beta_i}{\alpha_{ij}}\right) - \left\{\left(\gamma_1 - \gamma_j\frac{\alpha_{i1}}{\alpha_{ij}}\right)x_1 + \ldots + \left(-\gamma_j\frac{1}{\alpha_{ij}}\right)x_{k+i} + \ldots + \left(\gamma_k - \gamma_j\frac{\alpha_{ik}}{\alpha_{ij}}\right)x_k\right\}$$

(2.19)

Рассмотрим полученное базисное решение. Все свободные принимаются равными нулю $x_1 = \ldots = x_{k+i} = \ldots = x_k = 0$.

При этом базисные переменные будут равны:

$$x_j = \frac{\beta_i}{\alpha_{ij}},\ x_{k+e} = \left(\beta_e - \alpha_{ej}\frac{\beta_i}{\alpha_{ij}}\right)$$

и целевая функция

$$F = \left(\gamma_0 - \gamma_j\frac{\beta_i}{\alpha_{ij}}\right)$$

Проверим решение на допустимость. Базисная переменная $x_j>0$, так как в исходном базисном решении $\beta_i>0$ и $\alpha_{ij}>0$.

Рассмотрим остальные базисные x_{k+e}. Здесь характер изменения определяется знаком α_{ej}. При $\alpha_{ej}<0$ базисная переменная x_{k+e} возрастает, оставаясь положительной. При $\alpha_{ej}>0$ значение базисной x_{k+e} уменьшается. Для анализа преобразуем зависимость x_{k+e} к виду

$$x_{k+e} = \beta_e - \alpha_{ej}\frac{\beta_i}{\alpha_{ij}} = \alpha_{ej}\left(\frac{\beta_e}{\alpha_{ej}} - \frac{\beta_i}{\alpha_{ij}}\right)$$

по которому следует, что x_{k+e} будет оставаться положительной, так как генеральный элемент выбирается по условию $\frac{\beta_i}{\alpha_{ij}} = \min\left\{\frac{\beta_e}{\alpha_{ej}}\right\}$.

Таким образом, полученное в результате проведенных преобразований новое базисное решение остаётся допустимым.

Оценим теперь изменение целевой функции, равной в новом базисе

$$F = \left(\gamma_0 - \gamma_j\frac{\beta_i}{\alpha_{ij}}\right)$$

Здесь все коэффициенты, определяющие приращение функции, положительны, что приводит к снижению функции на величину $\Delta F = \gamma_j \frac{\beta_i}{\alpha_{ij}}$

Переход к новому базису достаточно просто формализуется с помощью специальных симплекс-таблиц.

2.7. Правила работы с симплекс-таблицей

Симплекс-таблица строится для допустимого базисного решения. В ней столбцы соответствуют свободным переменным, строки – базисным и линей ной форме, а в верхние половинки образовавшихся клеток таблицы записываются коэффициенты, которые соответствуют выражению базисных переменных и формы через свободные.

Таблица 2.2

	β	**$-x_1$**	**...**	**$-x_j$**	**...**	**$-x_k$**
x_{k+1}	β_1	α_{11}	...	α_{1j}	...	α_{1k}
...	...	...		...		...
x_{k+i}	β_i	α_{i1}	...	α_{ij}	...	α_{ik}
...	...	...		...		...
x_{k+m}	β_m	α_{m1}	...	α_{mj}	...	α_{mk}
F	γ_0	γ_1	...	γ_j	...	γ_k

Переход к новому базисному решению осуществляется по следующим правилам.

1. Определяется генеральный элемент по условию $\gamma_j>0$, $\alpha_{ej}>0$ и при
$$\frac{\beta_i}{\alpha_i} = \min\left\{\frac{\beta_e}{\alpha_{ej}}\right\}$$
Генеральный элемент подчеркивается. Определяется обратная величина
$$\lambda = \frac{1}{\alpha_{ij}}$$
и записывается в нижнюю половинку клетки с генеральным элементом.
2. Все элементы i-й строки умножаются на λ и результаты записываются в нижние половинки клеток.
3. Все элементы j-го столбца умножаются на –λ и результаты записываются в нижние половинки клеток.
4. Как-либо помечаются числа в верхних половинках клеток i-й строки и в нижних половинках клеток j-го столбца.
5. Нижние половинки остальных клеток определяются как произведение выделенных элементов строки и столбца, на пересечении которых лежит клетка.

После этого осуществляется переход к новой таблице, в которой x_{k+j} и x_j меняются местами. Верхние половинки клеток новой таблицы определяются по следующим правилам:

- i-я строка и j-й столбец новой таблицы заполняются содержимым нижних половинок соответствующих клеток старой таблицы;
- все остальные верхние половинки заполняются суммами содержимого верхних и нижних половинок соответствующих клеток старой таблицы.

Пример. Решить задачу линейного программирования

$$\begin{cases} x_1 + x_2 \le 6 \\ x_3 + x_4 \le 6 \\ 6x_1 + 13x_3 = 30 \\ 24x_2 + 13x_4 = 96 \end{cases}$$

(2.20)

$$F = 4x_1 + 47x_2 + 13x_3 + 26x_4 \to \min$$

Для перехода к канонической форме введём фиктивные переменные x_5, x_6 и перейдем к равенствам:

$$\begin{cases} x_1 + x_2 + x_5 = 6 \\ x_3 + x_4 + x_6 = 6 \\ 6x_1 + 13x_5 = 30 \\ 24x_2 + 13x_4 = 96 \end{cases}$$

(2.21)

Все уравнения линейно независимы, поэтому ранг системы r = m = 4. Число неизвестных: n = 6, число свободных k = 6–4 = 2.

Рассмотрим базисное решение, в котором в качестве свободных примем x_1 и x_2. Проверим его на допустимость при $x_1 = x_2 = 0$. Определим значения базисных переменных:

$$x_3 = \frac{30}{13},\ \ x_4 = \frac{96}{13},\ \ x_5 = 6,\ \ x_6 = 6 - x_3 - x_4 = 6 - \frac{30+96}{13} = -\frac{48}{13}$$

т.е. принятое базисное решение недопустимо.

Рассмотрим другой набор свободных, включающий x_1 и x_4.

В этом случае при $x_1 = x_4 = 0$ все базисные переменные положительны, т.е.

$$x_3 = \frac{30}{13},\ \ x_4 = \frac{96}{13},\ \ x_5 = 6 - 4 = 2,\ \ x_6 = 6 - x_3 = 6 - \frac{30}{13} = \frac{48}{13}$$

Таким образом, рассмотренное базисное решение является допустимым. Составим для него симплекс-таблицу, предварительно выразив базисные переменные и форму через свободные:

$$
\begin{cases}
x_2 = \dfrac{96-13x_4}{24} = 4 - \dfrac{13}{24}x_4 \\
x_3 = \dfrac{30-6x_4}{13} = \dfrac{30}{13} - \dfrac{6}{13}x_1 \\
x_5 = 6 - x_1 - x_2 = 6 - x_1 - 4 + \dfrac{13}{24}x_4 = 2 - x_1 + \dfrac{13}{24}x_4 \\
x_6 = 6 - x_3 - x_4 = 6 - \dfrac{30}{13} + \dfrac{6}{13}x_1 - x_4 = \dfrac{48}{13} + \dfrac{6}{13}x_1 - x_4
\end{cases}
$$

(2.22)

$$
F = 4x_1 + 47\left(4 - \frac{13}{24}x_4\right) + (30 - 6x_1) + 26x_4 = 218 - 2x_1 + \frac{13}{24}x_4
$$

Коэффициенты заносят в таблицу в соответствии со стандартной формой записи полученных выражений, т.е. с изменением знаков. Запишем условия (2.22) в виде стандартной таблицы (таблица 2.3).

В соответствии с правилами работы с таблицей, найдем генеральный элемент. Он лежит в столбце, соответствующем свободной x_1, поскольку $y_1 = 2$, т.е. больше нуля. Конкурентами для перехода в свободные являются базисные x_3 и x_5. Для них β/α равны, соответственно, 5 и 2. По минимальному отношению выбираем x_5, что и определяет положение разрешающего элемента α = 1. Обратная величина α = 1 записывается в нижнюю часть клетки (таблица 2.4).

Теперь элементы разрешающей строки умножаются на λ и записываются в нижние половинки этой строки.

Элементы разрешающего столбца, умноженные на –λ, записываются аналогично.

Таблица 2.3

	β	$-x_1$	$-x_4$
x_2	4	0	13/24
x_3	30/13	6/13	0
x_5	2	1	-13/24
x_6	48/13	-6/13	1
F	218	2	-13/24

Таблица 2.4

	β	$-x_1$	$-x_4$
x_2	4 0	0 0	13/24 0
x_3	30/13 -12/13	6/13 -6/13	0 1/4
x_5	2 2	1 1	-13/24 -13/24
x_6	48/13 12/13	-6/13 -6/13	1 -1/4
F	218 -4	2 -2	-13/24 26/24

Выделение соответствующих элементов выполнено кружками. В таблице 2.4 приведены результаты перемножения выделенных элементов.

Вторая симплекс-таблица (таблица 2.5) соответствует новому базисному решению, в котором $x_5 = x_4 = 0$, $x_2 = 4$, $x_5 = 18/13$, $x_1 = 2$, $x_6 = 60/13$ и F = 214.

Таблица 2.5

	β	$-x_1$	$-x_4$	β/α
x_2	4 -3	0 1	13/24>0 -52/24	96/13
x_3	18/13 72/13	-6/13 -24/13	1/4>0 4	72/13
x_5	2 3	1 -1	-13/24 52/24	
x_6	60/13 -54/13	6/13 18/13	3/4>0 -3	80/13
F	214 -3	-2 1	13/24>0 -52/24	

минимальное отношение β/α
$\lambda = 4$

Генеральный элемент

Работаем с новой таблицей также, как с предыдущей, и составляем третью симплекс-таблицу (таблица 2.6). Она соответствует оптимальному решению, т.к. в ней все $\gamma_j<0$.

Таблица 2.6

	β	$-x_5$	$-x_3$
x_2	1	1	-52/24
x_4	72/13	-24/13	4
x_1	5	0	52/24
x_6	6/13	24/13	-3
F	211	-1	-52/24

В оптимальном решении $x_3 = x_5 = 0$, $x_1 = 5$, $x_2 = 1$, $x_6 = 6/13$ и $F = 211$.

Поскольку в рассмотренной задаче число свободных переменных $k = 2$, ее можно решить графически.

Примем в качестве независимых переменных x_1 и x_2 и преобразуем исходную задачу к системе неравенств. Исходная система уже включает 2 неравенства:

$$\begin{cases} x_1 + x_2 \le 6 \\ x_3 + x_4 \le 6 \\ 6x_1 + 13x_3 = 30 \\ 24x_2 + 13x_4 = 96 \end{cases}$$

(2.23)

Исключим x_3 и x_4 из уравнений системы и по условиям $x_3 \ge 0$ и $x_4 \ge 0$ получим следующие неравенства:

$$x_3 = \frac{30 - 6x_1}{13} \geq 0, \quad 30 \geq 6x_1, \quad x_1 \leq 5;$$

$$x_4 = \frac{96 - 24x_2}{13} \geq 0, \quad 96 \geq 24x_2, \quad x_2 \leq 4;$$

Преобразуем второе неравенство исходной системы:

$$x_3 + x_4 = \frac{30 - 6x_1 + 96 - 24x_2}{13} \leq 6; \quad x_1 + 4x_2 \geq 8;$$

Аналогично преобразуем и выражение для целевой функции

$$F = 4x_1 + 47x_2 + 13x_3 + 26x_4 =$$
$$= 4x_1 + 47x_2 + 30 - 6x_1 + 2(96 - 24x_2) = 222 - 2x_1 - x_2 \Rightarrow \vec{n}\{-2;-1\}$$

Таким образом, исходная задача описывается следующей системой

$$\begin{cases} x_1 + x_2 \leq 6 \\ x_1 \leq 5 \\ x_2 \leq 4 \\ x_1 + 4x_2 \geq 8 \end{cases}$$

(2.24)

$$F(X) = 222 - 2x_1 - x_2 \rightarrow \min$$

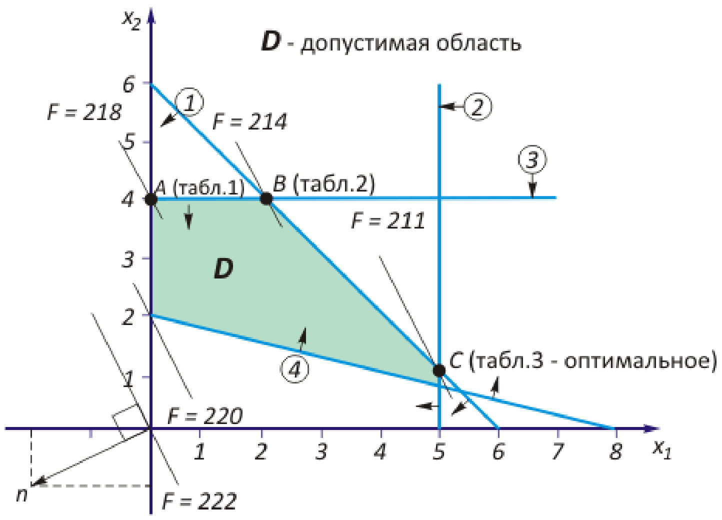


Рисунок 2.5 – График пересечения полуплоскостей

Допустимая область D лежит в первом квадранте и формируется как пересечение полуплоскостей, являющихся решением соответствующих неравенств. Семейство параллельных линий F = const определяется по нормали $\vec{n}\{-2;-1\}$ (рисунок 2.5)

Решение симплекс-методом соответствует обходу вершин, начиная с вершины A, соответствующей таблице 2, по направлению снижения целевой функции.

2.8. Применение ЭВМ

Правила работы с симплекс-таблицей легко формализуются. Рассмотрим алгоритм перехода к новому базису, заключающемуся в пересчёте коэффициентов исходной таблицы. Полагаем, что в памяти ЭВМ таблица представлена двумерным массивом А (рисунок 2.6), имеющем размерность A(m+1, k+1).

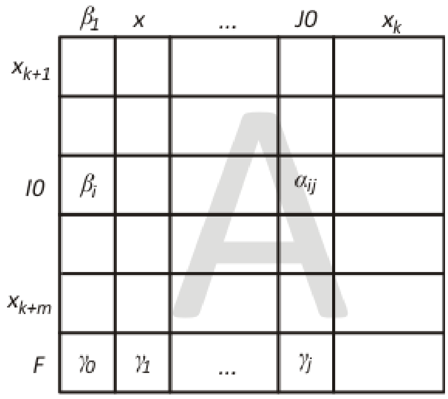


Рисунок 2.6 – Двумерный массив А

Будем считать, что генеральный элемент α_{ij} уже найден и находится в строке I0 и столбце J0 массива, т.е. α_{ij} = A(I0, J0). Блок-схема алгоритма показана на рисунке 2.7. Здесь введены следующие обозначения: LM = λ, K = k+1, M = m+1.

В первом цикле по i организуется пересчет коэффициентов разрешающего столбца.

В следующем вложенном цикле по i, исключая разрешающую строку, и по j, исключая разрешающий столбец, идет пересчет коэффициентов всех остальных элементов таблицы.

В последнем цикле по j пересчитывается разрешающая строка.

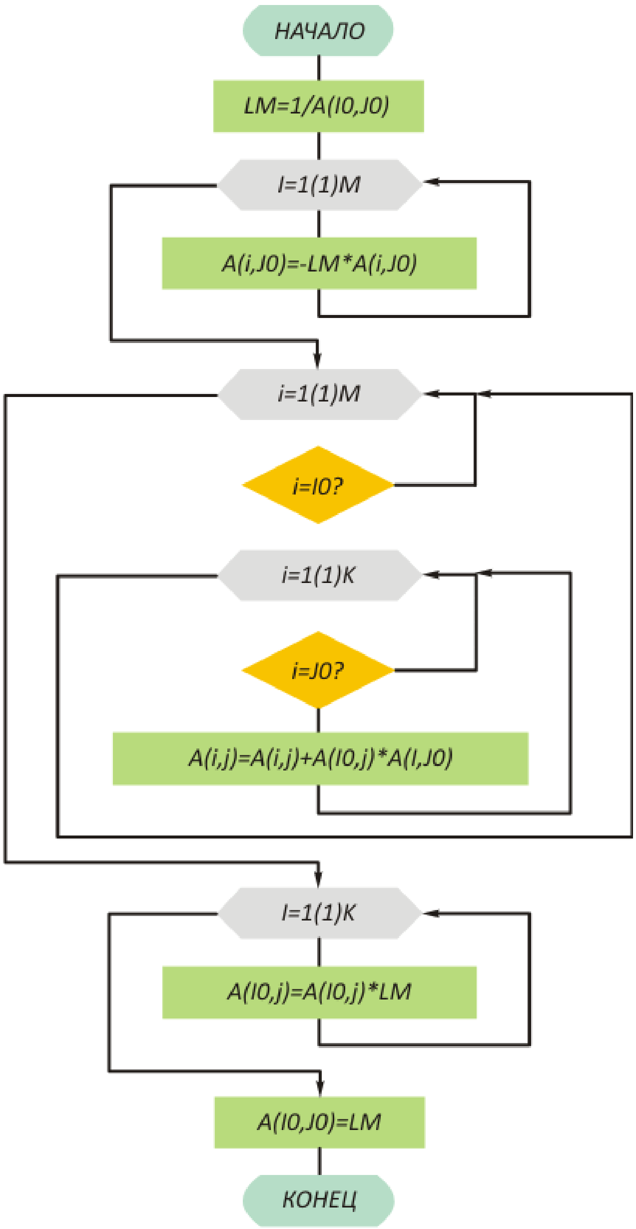


Рисунок 2.7 – Блок-схема алгоритма

В учебном процессе можно использовать две программы, разработанные на кафедре. Для решения линейных оптимизационных моделей может использоваться программа Simple.exe. Программа работает с любым видом ограничений, вводимых с помощью клавиш <, > и =. Переход к канонической форме условий задачи осуществляется программно.

На рисунке 2.8 приведен пример решения задачи оптимального распределения топлива между энергоустановками с заданной выработкой тепла. Основные свойства системы определяются ограничениями по располагаемому топливу и условиям выработки планируемых объемов тепла. В связи с этим математичская модель включает 2 неравенства и 3 равенства

$$x_1 + x_2 + x_3 \le B_1$$

$$x_4 + x_5 + x_6 \le B_2$$

$$x_1 q_1 \eta_{1,1} + x_4 q_2 \eta_{1,2} = Q_1$$

$$x_2 q_1 \eta_{2,1} + x_5 q_2 \eta_{2,2} = Q_2$$

$$x_3 q_1 \eta_{3,1} + x_6 q_2 \eta_{3,2} = Q_3$$

Целевая функция определяется стоимостью топлива

$$F(X) = c_1 (x_1 + x_2 + x_3) + c_2 (x_4 + x_5 + x_6) = \min$$

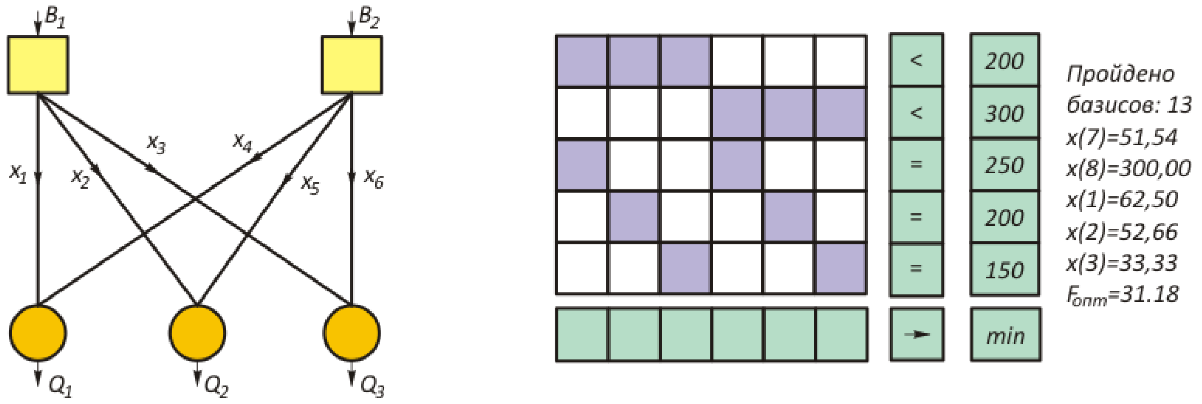


Рисунок 2.8 – Пример решения задачи оптимального распределения топлива

Учитывая слабую заполненность матрицы А условий задачи, в программе предусматривается возможность ввода только ненулевых элементов. В этом случае работа начинается с расстановки ненулевых элементов в ячейках матрицы А, шаблон которой выводится на экран после ввода чисел n и m, определяющих размерность задачи. Осуществляется это перемещением с помощью стандартных клавиш навигации курсора затемненной ячейки матрицы на место ненулевого элемента с последующим нажатием Enter. Выход из режима разметки после фиксации положения всех ненулевых элементов осуществляется по Esc.

Работа с программой проводится в режиме диалога с возможностью многовариантных расчётов путём корректировки исходных данных, но без изменения размерности задачи.

По результатам решения на экран выводятся значения базисных переменных, целевой функции и количество пройденных вершин многогранника допустимых решений. На рисунке 2.8 показан пример решения сформулированной задачи для произвольного значения исходных данных.

Для решения задач линейного программирования могут использоваться и другие программы, реализующие алгоритм симплекс-метода. На рисунке 2.9 приведен пример решения транспортной задачи с помощью программы

Simple.exe. Программа практически не имеет ограничений по размерности решаемых задач. Исходная матрица условий задачи вводится своими ненулевыми элементами в таблицу, сформированную в соответствии с размерностью задачи. Навигация внутри таблицы осуществляется с помощью мышки и полос прокрутки.

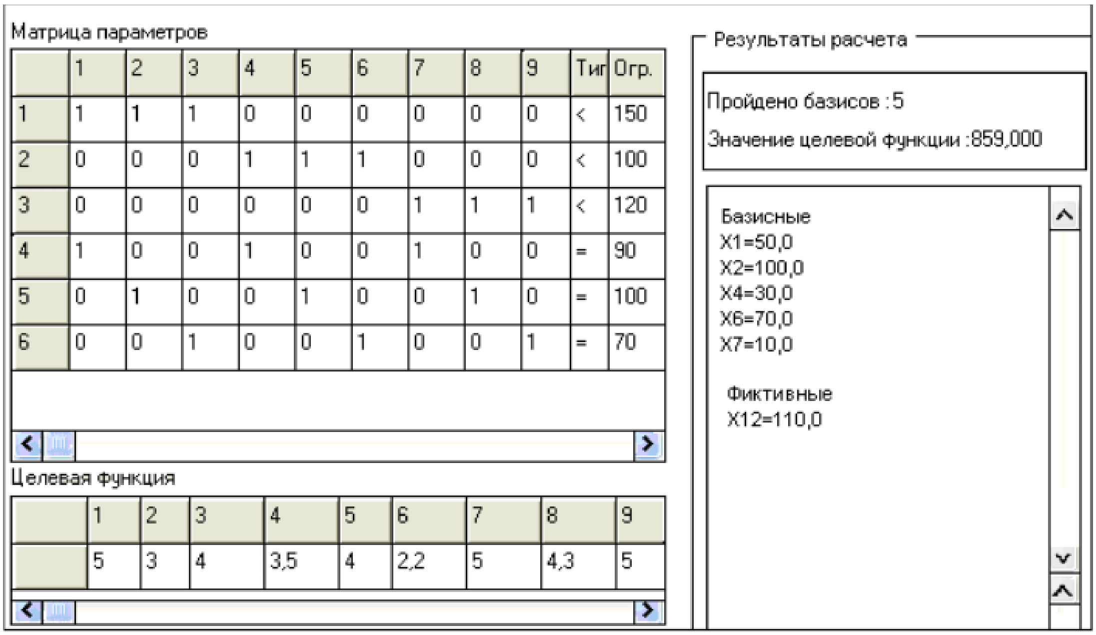


Рисунок 2.9 – Пример решения транспортной задачи с помощью программы Simple.exe

2.9. Поиск допустимого базисного решения

Реализация алгоритма симплекс-метода начинается с определения допустимого базисного решения, в котором при нулевых свободных переменных все базисные переменные положительны (не отрицательны).

Желательно, чтобы при поиске допустимого базиса по возможности более просто получались и выражения базисных переменных через свободные. Для этой цели могут использоваться разные методы решения в зависимости от типа системы условий исходной задачи.

Если ограничения задачи линейного программирования преобразуются к виду $A \cdot X \le B$ и все $b_i>0$, то в этом случае для перехода к равенствам в каждое неравенство вводят фиктивные переменные $A \cdot X + Y = B$.

Эти фиктивные переменные и принимаются за базисные, а исходные X за свободные. Выражения для базисных определяются исходной системой $Y = B - A \cdot X$, а базисное решение является допустимым. Целевая функция при этом не меняется $F = C^T X$.

Если ограничения заданы в виде A·X = B и в каждом уравнении есть неизвестная xi, которая встречается только в этом уравнении, то в матрице условий задачи A можно выделить блок, являющийся единичной матрицей (рисунок 2.10). Неизвестные, соответствующие этой матрице, можно принять в качестве базисных.

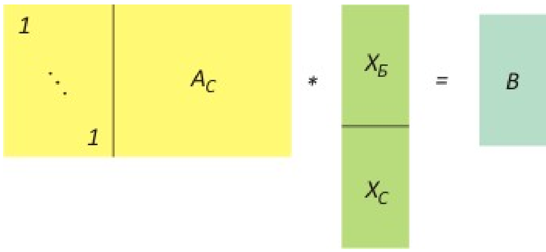


Рисунок 2.10 – Блоки матрицы

Записав операции над блочными матрицами получим

$$E \cdot X_Б + A_C \cdot X_C = B$$

откуда можно найти выражения для базисных переменных и формы

$$X_Б = E^{-1}\left(B - A_C \cdot X_C\right) = B - A_C \cdot X_C$$

$$F = C_Б^T \cdot X_C = C_Б^T\left(B - A_C \cdot X_C\right) + C_C^T \cdot X_C = C_Б^T \cdot B + \left(C_C^T - C_Б^T \cdot A_C\right) \cdot X_C$$

(2.25)

Для выделения базисных переменных в общей задаче A·X = B и представления их через свободные может использоваться метод исключения Жордана.

Для поиска допустимого базиса в этом случае может использоваться метод расширенной задачи, которая формируется путём ввода дополнительных переменных z_i в каждое уравнение. Эти переменные и принимаются за базисные Z = B – A·X. В допустимом решении все z_i должны стать равными нулю. Для того, чтобы обеспечить это условие в рамках алгоритма симплекс-метода, переменные z_i вводятся в целевую функцию с большим штрафным коэффициентом:

$$F = C^T X + Z \cdot k_{\varnothing} = C^T X + k_{\varnothing}\left(B - A \cdot X\right)$$

(2.26)

Применение симплексного метода к расширенной задаче обеспечивает поиск решения, в котором все z_i = 0. Если в решении хотя бы одна переменная остается большей нуля, т.е. $z_i > 0$, то исходная задача не совместна и не имеет решения.

2.10. Понятие двойственности в линейном программировании

В математике бывают ситуации, когда для некоторой задачи можно по определённым правилам построить другую, из решения которой автоматически получается решение исходной.

В линейном программировании каждая задача имеет двойственную задачу и по решению одной можно находить решение другой.

Исходная задача 1:

$$A \cdot X \leq B$$

$$f = C^T X \to \max$$

Двойственная задача 2:

$$A^T Y \geq C$$

$$Z = B^T Y \to \min$$

Сопоставляя математическую запись задач 1 и 2, можно отметить следующие особенности:

- если в задаче 1 имеем m неравенств для n неизвестных, то в задаче 2 n противоположных по знаку неравенств для m неизвестных,
- коэффициенты системы условий задачи 2 формируются путем транспонирования матрицы A задачи 1,
- если в задаче 1 определяется максимум F, то в задаче 2 – минимум,
- коэффициенты целевой функции одной задачи являются правыми частями ограничений другой,
- по теореме двойственности $f_{max} = Z_{min}$.

Обычно эти задачи имеют экономический смысл и могут использоваться для поиска путей повышения эффективности производства.

В качестве примера рассмотрим упрощенную задачу распределения ресурсов, которая решается при планировании производства трансформаторов на максимум прибыли.

Предположим, на предприятии производят два типа трансформаторов, которые реализуются на рынке по цене C_1 = 600 руб/шт, C_2 = 500 руб/шт.

При изготовлении используются три основных вида ресурсов – железо, медь и изоляция, запасы которых известны и равны соответственно $в_1$ = 400, $в_2$ = 300 и $в_3$ = 50 кг. Определены расходы ресурсов на один трансформатор каждого типа, которые приведены в таблице 2.7.

Таблица 2.7

	Железо	Медь	Изоляция	цена
T_1	2,0	2,5	1	600
T_1	3	4,75	0,5	500
запас	400	300	50	

Составим математическую модель исходной задачи.

В качестве неизвестных примем: x_1 – объём выпуска T_1 и x_2 – объём выпускаT_1. Модель учитывает ограничения по ресурсам и целевую функцию:

$$\begin{cases} 2x_1 + 3x_2 \le 400 \\ 2,5x_1 + 2x_2 \le 300 \\ x_1 + 0,5x_2 \le 50 \end{cases}$$

(2.28)

$$F = 600x_1 + 500x_2 \to \max$$

Решение: $x_1 = x_5 = 0$, $x_1 = 5$, $x_2 = 100$, $x_3 = 100$, $x_4 = 50$ и $F = 50000$ руб.

Таким образом, в оптимальном плане выгодно производить трансформаторы T_2 в количестве 100 штук. Остаток железа – 100 кг, меди – 50 кг, изоляция в дефиците и используется полностью.

Двойственная задача позволяет определить условную (теневую) цену единицы каждого ресурса, чтобы минимизировать затраты на имеющееся сырьё.

Введём переменные: y_1 – стоимость железа; y_2 – цена меди; y_3 – цена изоляции.

Сформируем двойственную задачу с учётом отмеченных выше особенностей:

$$\begin{cases} 2y_1 + 2,5y_2 + y_3 \ge 600 \\ 3y_1 + 2y_2 + 0,5y_3 \ge 500 \end{cases}$$

$$z = 400y_1 + 250y_2 + 50y_3 \to \min$$

Решение этой задачи:

$$y_3 = 1000,\ \ y_4 = 400,\ \ y_1 = y_2 = y_5 = 0$$
$$z = 50\ 000 \text{ руб.}$$

Полученные теневые цены позволяют выявить степень дефицитности каждого ресурса. При этом найденная цена определяется приращением Δf при увеличении запаса дефицитного ресурса на единицу.

Здесь используются следующие свойства двойственности оценок y_i:

- положительная оценка y_i может быть только у ресурса, который в оптимальном плане полностью исчерпан;
- величина y_i для каждого дефицитного вида ресурса показывает, насколько можно увеличить f, если ввести в производство дополнительно единицу ресурса $в_i$.

В полученном решении теневая цена изоляции 1000 руб/кг определяется приращением прибыли, которую можно получить на каждый дополнительный кг изоляции, позволяющий получить еще 2 трансформатора по 500 руб за штуку.

Решение двойственных задач позволяет оценивать пути повышения рентабельности производства.

2.11. Целочисленное программирование

Задачи, в которых все составляющие вектора X или часть из них могут быть только целыми, называются целочисленными. Алгоритмы их решения гораздо сложнее алгоритмов решения непрерывных задач линейного программирования.

На первый взгляд наиболее простым методом решения целочисленных задач является метод округления, по которому на первом этапе решается симплекс-методом общая задача, не учитывающая требование целочисленности переменных, а затем получившиеся базисные переменные, не являющиеся целыми, округляются до ближайших целочисленных значений.

Однако при этом полученное решение может значительно отличаться от оптимального. Поэтому для решения задач целочисленного программирования разработаны специальные более сложные методы.

Некоторые вопросы, связанные с решением таких задач, рассмотрим на примере следующей целочисленной задачи

$$\begin{cases} x_1 + 4x_2 \le 19 \\ 2,5x_1 + x_2 \le 11 \end{cases}$$

(2.29)

$$F = 2x_1 + x_2 \to \max$$

Графическое решение этой задачи показано на рисунке 2.11.

Оптимальное решение задачи без учёта целочисленности лежит в вершине 1, координаты которой приведены в таблице 2.8., и не удовлетворяют требованиям целочисленности. Естественное округленное значение, соответствующее точке 2 оказывается недопустимым. Оптимальное решение лежит в точке 4.

На этом простом примере видно, что округление оптимальных непрерывных значений до целых может приводить к недопустимым решениям и для поиска оптимального целочисленного решения может потребоваться перебор большого числа комбинаций возможных округленных значений. Поэтому для поиска оптимального целочисленного решения разработаны специальные методы.

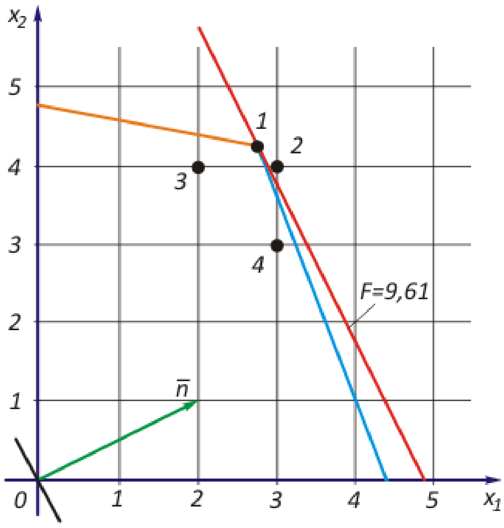


Рисунок 2.11 – Графическое решение задачи

Таблица 2.8

Точка	**x_1**	**x_2**	**F**	**Решение**
1	2,78	4,06	9,61	Непрерывное
2	$x_1 = 3$	$x_2 = 4$	–	Недопустимое
3	$x_1 = 2$	$x_2 = 4$	8	Допустимое
4	$x_1 = 3$	$x_2 = 3$	9	Оптимальное

Наиболее популярен метод отсекающих плоскостей (метод Гомори). Метод основан на поиске непрерывного решения, проверке его на целочисленность и формировании при необходимости дополнительных специальных ограничений, отсекающих найденную нецелочисленную вершину. Решение расширенной задачи подвергается аналогичной обработке.

Другой метод, называемый методом ветвей и границ, основан на направленном переборе целочисленных вершин, расположенных вблизи непрерывного решения. При этом всё множество целочисленных вершин делится на два подмножества, в каждом из которых рассматриваются округленные значения только для одной базисной x_i.

Диаграмма ветвления показана на рисунке 2.12.

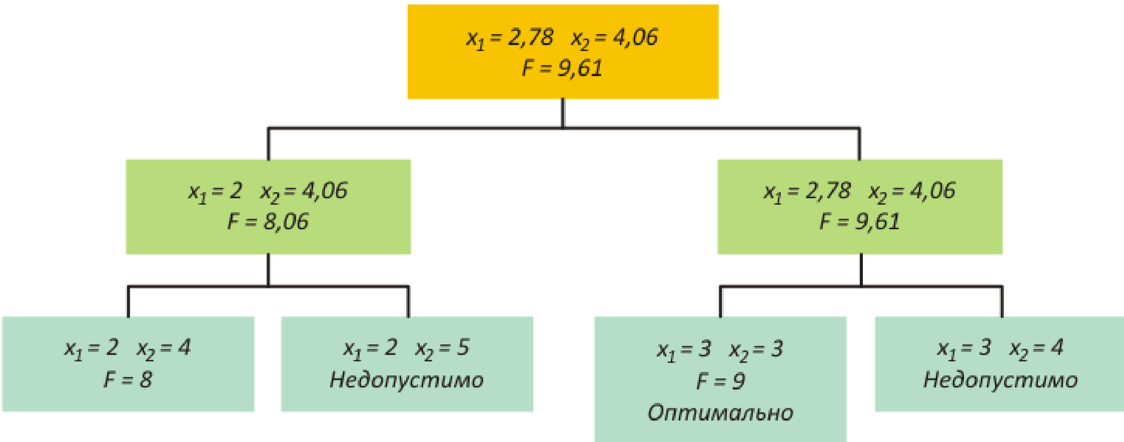


Рисунок 2.12 – Диаграмма ветвления

Метод ветвей и границ даёт возможность решать задачи с любым числом целочисленных переменных.

2.12. Транспортная задача

Одной из наиболее простых и распространённых задач линейного программирования является транспортная.

Смысл задачи предельно ясен: имеется несколько поставщиков и потребителей определенного продукта. Заданы известные объемы продукта у поставщиков и потребности у потребителей. Известны также расстояния по существующим трассам от каждого потребителя до каждого поставщика. Задача заключается в определении таких направлений и объемов транспорта продукта, при которых обеспечивается минимальный грузопоток, определяющий общие транспортные расходы.

Составим математическую модель для простейшей задачи, включающей двух поставщиков с запасами D_1 и D_2, и двух потребителей с потребностью B_1 и B_2 (рисунок 2.13). Известны расстояния между объектами $l_1, \ldots, l_4$.

В зависимости от соотношения объёмов продукта у поставщиков и потребителей различают закрытую задачу, когда $\sum D_i = \sum B_j$, и открытую, когда $\sum D_i > \sum B_j$.

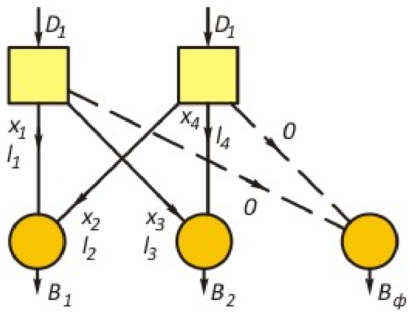


Рисунок 2.13 – Пример транспортной задачи

Открытую задачу можно свести к закрытой, если ввести фиктивного потребителя с потребностью $B_ф = \sum D_i - \sum B_j$, длины путей к которому равны нулю.

Составим математическую модель.

Введем переменные: x_1, x_2, x_3, x_4 – неизвестные объемы продукта, транспортируемого по соответствующим путям.

Опишем основные свойства системы, включающие условия баланса продуктов у потребителей и поставщиков:

$$\begin{cases} x_1 + x_2 = B_1 \\ x_3 + x_4 = B_2 \\ x_1 + x_3 \le D_1 \\ x_2 + x_4 \le D_2 \end{cases}$$

(2.30)

В качестве целевой функции принимаем грузопоток

$$F(X) = \ell_1 \cdot x_1 + \ell_2 \cdot x_2 + \ell_3 \cdot x_3 + \ell_4 \cdot x_4 \to \min$$

Полученную задачу линейного программирования приведём к каноническому виду, т.е. перейдем от неравенств к равенствам:

$$\begin{cases} x_1 + x_2 = B_1 \\ x_3 + x_4 = B_2 \\ x_1 + x_3 + x_5 = D_1 \\ x_2 + x_4 + x_6 = D_2 \\ x_5 + x_6 = B_ф \end{cases}$$

(2.31)

Для системы условий задачи имеем n = 6, m = 5.

Проанализируем полученную систему условий задачи с целью определения ранга. Одно из уравнений системы является линейно зависимым, так как для них можно найти условие, например,

$$У_1 \cdot 1 + У_2 \cdot 1 + У_3 \cdot (-1) + У_4 \cdot (-1) + У_5 \cdot 1 = 0$$

где $У_i$ – уравнение с номером i.

Значит ранг системы R = 4, и в оптимальном решении должно быть только 4 базисных переменных. В общем случае R = N + M для открытой задачи и R = N + M – 1 для закрытой задачи, где N, M – число поставщиков и потребителей.

Для решения транспортной задачи может использоваться симплекс-метод. Однако существует несколько более простых методов решения её, из которых рассмотрим два метода: **распределительный метод** и **метод потенциалов**.

2.13. Распределительный метод

Суть метода рассмотрим на примере, в котором участвуют 3 потребителя и 3 поставщика. Запасы, потребности в тоннах и расстояния в км заданы в таблице 2.9.

Решение начинается с составления допустимого базисного решения, называемого опорным планом. Он должен иметь 5 базисных, т.е. не равных нулю переменных.

Таблица 2.9

потребители (столбцы), *поставщики* (строки)

	40	85	55
50	2	1	5
60	3	4	3
70	4	6	6

Таблица 2.10

потребители (столбцы), *поставщики* (строки)

	40	85	55
50	2 / 40	1 / 10	5
60	3	4 / 60	3
70	4	6 / 15	6 / 55

Задача является закрытой, так как ∑D = ∑B и R = n+m = 3+3–1 = 5.

Решение начинается с составления допустимого базисного решения, называемого опорным планом. Он должен иметь 5 базисных, т.е. не равных нулю переменных.

Составить опорный план можно разными методами. Применим так называемый метод северо-западного угла. Левую верхнюю клетку (путь от первого поставщика до первого потребителя) загружаем по максимуму, т.е. планируем перевозить по этому пути столько, сколько надо потребителю или сколько может дать поставщик. Если первому потребителю не хватило продукта – загружаем соседнюю клетку и т.д. Полученный опорный план представлен в таблице 2.10. Значение целевой функции

$$F = 2 \cdot 40 + 1 \cdot 10 + 4 \cdot 60 + 6 \cdot 15 + 6 \cdot 55 = 750 \text{ т} \cdot \textit{км}$$

Даже поверхностный взгляд на опорный план показывает, что неудачно загружено направление длиной 1 км. Очевидно, этот путь надо загрузить более полно.

Для оценки ожидаемого эффекта загрузим эту клетку на 1 т, что вызовет необходимость разгрузки на 1 т клетки слева, чтобы запас у поставщика не изменился. Поскольку потребности потребителей также не должны изменяться, поэтому нужно загрузить на 1 т клетку 3 км и разгрузить на 1 т клетку 4 км. Таким образом, в перераспределении поставок была рассмотрена цепочка, показанная на рисунке 2.14.

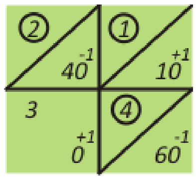


Рисунок 2.14 – Цепочка поставок

Цепь – всегда замкнутый многоугольник с четным числом вершин, только одна из которых не имеет поставки.

Для оценки эффекта найдем изменение целевой функции при перераспределении поставок на 1 т в рассматриваемой цепочке, характеризующее клетку без поставки: $\Delta F = -1 \cdot 2 + 1 \cdot 1 - 1 \cdot 4 + 1 \cdot 3 = -2$ т·км.

Этот показатель для цепочки называют характеристикой, обозначаемой $E_{2,1} = -2$ т·км.

Перераспределение по всем клеткам цепочки проводится на одинаковую величину, а объем перераспределения определяется минимальной поставкой в клетке, где поставки уменьшаются. Очевидно после перераспределения там окажется 0. В нашем случае уменьшаются поставки в клетках 2/40 и 4/60. Следовательно, объем перераспределения поставок – 40 т (рисунок 2.15). Целевая функция при этом изменится на $\Delta F = E \cdot \min\{x_{-1}\} = -2 \cdot 40 = -80$ т·км и для полученного плана составит $F = 750 - 80 = 670$ т·км.

Таблица 2.11

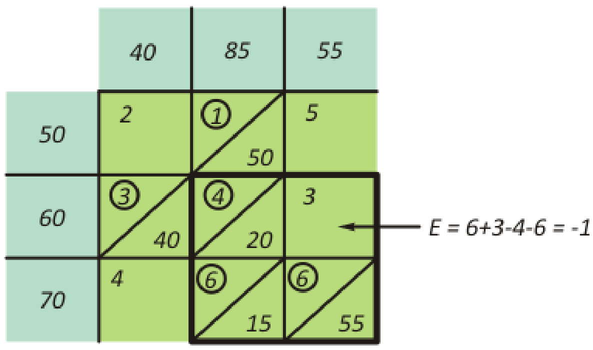
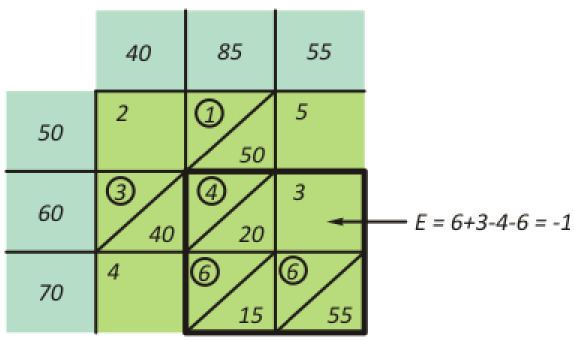


Рисунок 2.15 – Перераспределение цепочек

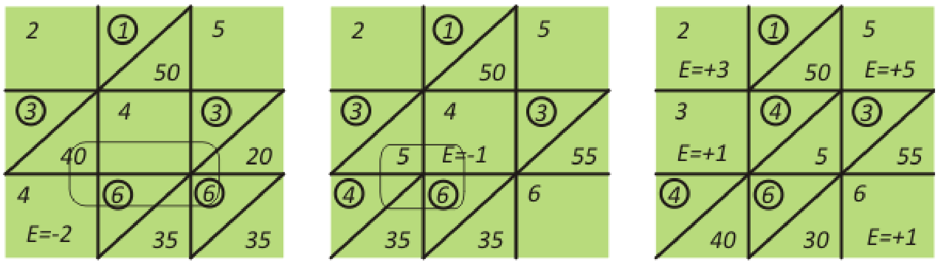
Улучшенный новый план по результатам проведённого перераспределения показан в таблице 2.11.

Таблица 2.11



Аналогично произведем дальнейшую оптимизацию плана. Результаты последовательного улучшения плана поставок показаны в таблицах 2.12, 2.13 и 2.14.

Таблица 2.12 Таблица 2.13 Таблица 2.14



(*) F=670-1·20=650 F=650-2·35=580 F=580-1·5=575

Чтобы убедиться в оптимальности полученного плана, необходимо определить характеристики для всех «пустых» клеток, для которых они должны быть положительными.

При реализации распределительного метода возникают следующие самостоятельные задачи, связанные с построением опорного плана и определением цепочек для «пустых» клеток. Опорный план должен иметь число «занятых» клеток (кружков), равное рангу системы, и эти кружки должны образовывать вычёркиваемую комбинацию.

Эта комбинация проверяется последовательным вычеркиванием кружков, которые являются единственными в строке или в столбце. На рисунке 2.16 показаны два опорных плана, из которых второй не отвечает этим требованиям.

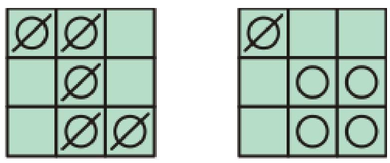


Рисунок 2.16 – Опорные планы

Формирование опорного плана может проводиться одним из следующих методов:

- метод северо-западного угла;
- метод минимального элемента в столбике;
- метод минимального элемента в строке;
- метод минимального элемента в таблице.

Из них наиболее эффективным является последний метод из этого списка.

При определении цепочек следует иметь ввиду, что в больших задачах они могут быть много сложнее, чем в рассмотренном примере. На рисунке 2.17 показаны возможные виды цепочек.

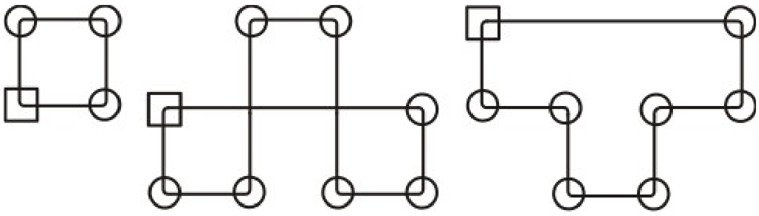


Рисунок 2.17 – Возможные виды цепочек

Алгоритм определения цепочек строится на основе проверки вычеркиваемой комбинации кружков (клеток с поставками).

При определении цепочки для какой-либо «пустой» клетки искусственно записывают в неё кружок, затем последовательно вычеркивают все кружки, являющиеся единственными в своей строке или столбце. Оставшиеся от вычеркивания кружки образуют вершины многоугольника цепочки.

В качестве примера найдем цепочку с использованием вычеркиваемой комбинации для плана, получившегося после второго улучшения в рассмотренном ранее примере (рисунок 2.18), где показана последовательность вычеркивания.

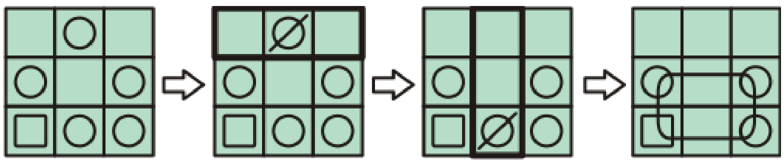


Рисунок 2.18 – Последовательность вычеркивания комбинаций

При составлении опорного плана может оказаться, что число базисных переменных меньше ранга R. Этот случай называют «вырождением». Для устранения его в опорный план вводят искусственную поставку, равную нулю, выбрав для неё любую клетку, не вызывающую образования цепочки.

Характеризуя в целом распределительный метод, следует отметить его достоинство – простоту, дающую возможность вручную решать сложные задачи. Недостаток метода заключается в необходимости рассмотрения большого числа цепочек.

2.14. Метод потенциалов

Этот метод позволяет определять характеристики клеток без построения цепочек. Для этого используются потенциалы – некоторые показатели поставщиков v_i и потребителей u_j, которые определяются по следующему правилу: сумма потенциалов поставщика и потребителя должна быть равна показателю клетки, которая их связывает и по которой запланирован транспорт продукта,

$$v_i + u_j = \bar{\beta}_{ij}$$

(2.32)

где $\bar{\beta}_{ij}$ – показатель перевозок клетки с кружком, т.е. с запланированной поставкой, которая связывает i-го поставщика и j-го потребителя.

Один из потенциалов в опорном плане может приниматься произвольно. Найдем потенциалы (таблица 2.15) для первого опорного плана в рассмотренном ранее примере.

Таблица 2.15

	40	85	55	v_i
50	2	1	5	0
60	3	4	3	3
70	4	6	6	5
u_j	2	1	1	

Потенциал первого поставщика выберем произвольно $v_1 = 0$. Остальные потенциалы найдем по условию (2.32) в следующей последовательности:

$$u_1 = \bar{\beta}_{11} - v_1 = 2 - 0 = 2;\ u_2 = \bar{\beta}_{21} - v_1 = 1 - 0 = 1$$

$$v_2 = \bar{\beta}_{22} - u_2 = 4 - 1 = 3;\ v_3 = \bar{\beta}_{32} - u_2 = 6 - 1 = 5$$

$$u_3 = \bar{\beta}_{33} - v_3 = 6 - 5 = 1$$

Потенциалы имеют определённый экономический смысл, определяемый затратами 3-х субъектов: поставщика, потребителя и транспортной компании. При этом потенциалы можно рассматривать как транспортные издержки компании β_{ij}, которые в принятом плане компенсируются потребителем и поставщиком в объемах, соответственно, u_j и v_i.

Определим способ расчета характеристики с помощью потенциалов. Ранее была получена характеристика клетки, лежащей на пересечении 2-й строки и 1-го столбца $E_{2,1} = -2$. Найдем ее с помощью потенциалов:

$$E_{2,1} = \beta_{2,1} + \bar{\beta}_{1,2} - \bar{\beta}_{1,1} - \bar{\beta}_{2,2} = \beta_{2,1} + v_1 + u_2 - (v_1 + u_1) - (v_2 + u_2)$$
$$E_{2,1} = \beta_{2,1} - (v_2 + u_1)$$

(2.33)

Таким образом, характеристика клетки без поставки определяется как разность показателя перевозок и суммы потенциалов поставщика и потребителя, которых связывает рассматриваемая клетка:

$$E_{2,1} = 3 - (3 + 2) = -2$$

Использование потенциалов позволяет более просто, без построения цепочек найти клетки, в которых целесообразно ввести поставку. Однако для связанной с этим последующей корректировки плана будет необходимо найти цепочку и провести перераспределение поставок в её вершинах.

Тема 3. Основы нелинейного программирования

3.1. Особенности задач линейного программирования

Математически задача нелинейного программирования формулируется следующим образом:

$$F(X) \to extr, \ \ G(X) \geq 0 \tag{3.1}$$

где $X = \{x_1, x_2, ..., x_n\}$ – вектор переменных,

$F(X)$ – целевая функция,

$G(X) = \{g_1(x), g_2(x), ..., g_m(x)\}$ – вектор-функция ограничений, которые могут иметь форму равенств и неравенств.

В нелинейном программировании $F(X)$ или хотя бы одна функция $g_j(X)$ должна быть нелинейной. Появление нелинейностей накладывает целый ряд особенностей на задачу. Основные из них:

- Допустимая область может быть невыпуклой, как например для системы из двух следующих ограничений (рисунок 3.1)

$$\begin{cases} g_1(X) = x_1^2 + x_2^2 - 5 \leq 0 \\ g_2(X) = x_1 x_2 - 1 \leq 0 \end{cases} \tag{3.2}$$

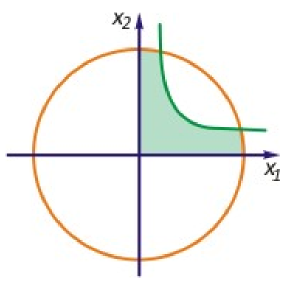


Рисунок 3.1 – График допустимой области

- Решение задачи может лежать как внутри, так и на границе допустимой области.
- Целевая функция может иметь несколько экстремумов, среди которых различают локальные А, С и глобальные В (рисунок 3.2).

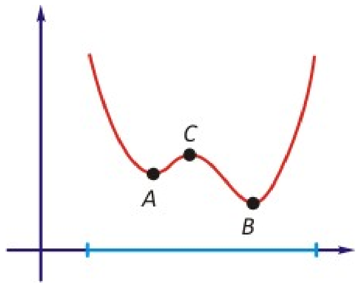


Рисунок 3.2 – График целевой функции

В отличии от линейного программирования в нелинейном существуют десятки методов, ориентированных на надёжную работу в конкретных условиях, и практически не существует универсального метода, который позволяет решать любые задачи. Ниже будут рассмотрены лишь основные методы. При изложении материала иллюстрация методов будет проводиться для функции двух переменных $F(x_1, x_2)$, представленной на плоскости в координатах x_1, x_2 линиями равных значений $F(x_1, x_2) = const$.

Суть такого представления показана на рисунке 3.3 для функции

$$F(X) = 10 + (x_1 - 4)^2 + (x_2 - 2)^2 \to \min$$

(3.3)

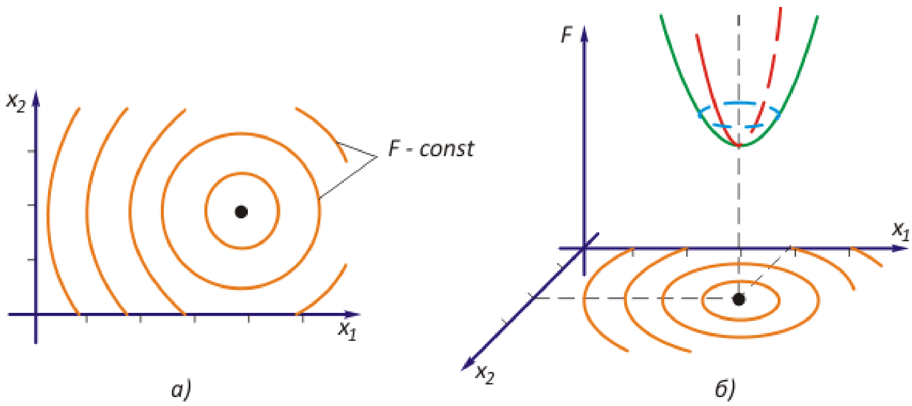


Рисунок 3.3 – График функии двух переменных

Основы нелинейного программирования начнём с изучения методов оптимизации нелинейной функции $F(X) \to \min$, где на допустимую область не накладывается никаких условий.

3.2. Методы безусловной оптимизации

Итак, рассмотрим задачу $F(X) \to \min$, где $X = \{x_1, ..., x_n\}$.

Из общего курса математики известно, что аналитическое решение задачи определяется равенством нулю всех частных производных

$$\frac{\partial F}{\partial x_i} = 0, \quad i = 1(1)n$$

и положительным значением второго дифференциала $d^2F \geq 0$.

Однако, для решения практических задач обычно используются итеративные методы, ориентированные на реализацию их на ЭВМ. Эти методы делятся на два основных типа – методы перебора и методы возможных направлений.

Метод перебора

При переборе используется только вычисление значения функции F(X) в различных точках некоторой сетки, определяемых условием:

$$x_{1i} = x_{1(0)} + i \cdot \Delta x,$$
$$x_{2j} = x_{2(0)} + j \cdot \Delta x$$

(3.4)

Функция F(X) вычисляется в каждой точке сетки и затем путем сравнения выбирается ее минимальное значение (рисунок 3.4).

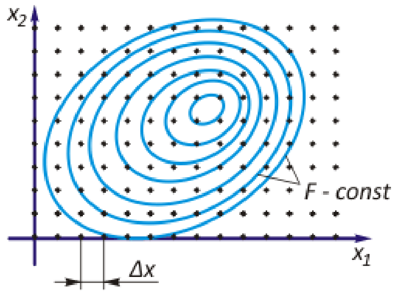


Рисунок 3.4 – Функция F(X)

Достоинство метода: за счет сканирования всей допустимой области могут решаться задачи с несколькими экстремумами. Недостаток: большой объем расчетов.

Для сокращения объема вычислений можно использовать метод направленного перебора, который заключается в расчете F(X) в соседних точках и переходе в ту их них, где значение функции минимально. Но в этом случае в задаче с несколькими экстремумами можно не найти глобальный.

Методы перебора используются редко. Поэтому наибольшее распространение получили методы возможных направлений. Это разновидность итерационных методов, которые начинаются с произвольно выбранного начального приближения.

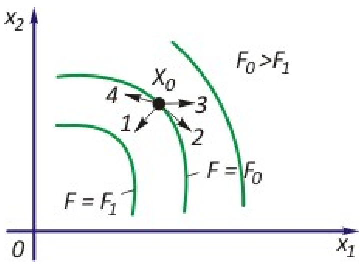


Рисунок 3.5 – Варианты направления для перехода исходной точки

В исходной точке X_0 рассматривается несколько направлений (рисунок 3.5) для перехода в новую точку X_1. Возможными называют направления, которые ведут в сторону уменьшения целевой функции (направления 1 и 4). Перемещение допускается по любому возможному направлению в текущей точке пропорционально некоторой константе t, называемой шагом, т.е.

$$X_1 = X_0 + \vec{S}\Big|_{X_0} \cdot t$$

(3.5)

В точке X_1 аналогично выбирается возможное направление и делается очередной шаг. Общее уравнение итерационного процесса по методу возможных направлений

$$X_{k+1} = X_k + \vec{S}\Big|_{X_k} \cdot t$$

(3.6)

Величина шага влияет на сходимость вычислительного процесса, определяемую числом итераций: при малых t – процесс сходится, но медленно; при больших t – процесс может расходиться.

Между этими крайностями существует оптимальный шаг, который на принятом направлении приводит в точку минимального значения F, где это направление касается линии F = const (рисунок 3.6)

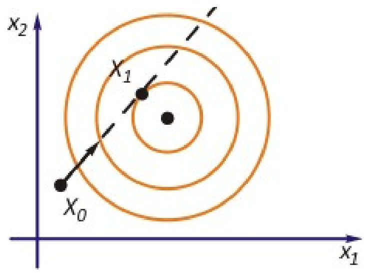


Рисунок 3.6 – Точка минимального значения F

Для отыскания оптимального шага $t_{опт}$ можно на принятом направлении $\vec{S}$ выражение X(t) подставить в целевую функцию F(X(t)) = f(t)→min и после преобразований получить новую функцию, зависящую только от шага f(t), а затем найти ее минимальное значение.

Пример: $F(X) = 10 + (x_1 - 4)^2 + (x_2 - 2)^2 \rightarrow min$, при исходном приближении и заданном возможном направлении: $X_0 = \{0,0\}$, $\vec{S}\Big|_{X_0} = (1,1)$.

Значения переменных в конце шага:

$$x_1 = 0 + 1 \cdot t$$
$$x_2 = 0 + 1 \cdot t$$

Подставим эти значения в целевую функцию:

$$f(t)=10+(t-4)^2+(t-2)^2 \to \min$$

Запишем условие минимума функции:

$$\frac{df}{dt}=2\cdot(t-4)+2\cdot(t-2)=4\cdot t-12=0,\ \text{откуда}\ t_{onm}=3$$

$$X_1=X_{(0)}+\vec{S}\Big|_{X_0}\cdot 3=\{0,0\}+\{1,1\}\cdot 3=\{3,3\}$$

Траектория перехода показана на рисунке 3.7.

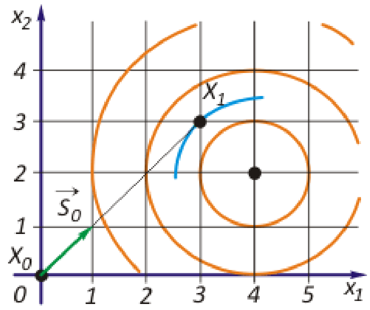


Рисунок 3.7 – Траектория перехода

Зависимость F(t) на принятом возможном направлении далеко не всегда удается найти в аналитическом виде. В этом случае функцию F(t) можно аппроксимировать, например, кривой второго порядка

$$F(t)=a\cdot t^2+b\cdot t+c$$

(3.7)

что позволяет определить оптимальный шаг

$$\frac{dF}{dt}=2at+b=0 \quad\Rightarrow\quad t_{onm}=-\frac{b}{2a}$$

(3.8)

Поиск трех неизвестных коэффициентов a, b, c требует формирования трех уравнений. Для этого рассматриваются 3 точки на принятом направлении для разных значений t, равных, например, t = 0, t = 1, t = 2, и для каждой точки определяется значение целевой функции:

$$\left.\begin{array}{l}
t=0: X=X_0+\vec{S}\Big|_{X_0}\cdot t=X_0,\quad F_0=F(X_0)\\
t=1: X=X_0+\vec{S}\Big|_{X_0}\cdot 1=X_1,\quad F_1=F(X_1)\\
t=2: X=X_0+\vec{S}\Big|_{X_0}\cdot 2=X_2,\quad F_2=F(X_2)
\end{array}\right\}$$

(3.9)

Очевидно, для рассмотренных шагов эти значения должны совпадать с рассчитанными по аппроксимирующей кривой

$$\begin{cases} c = F_0 \\ a + b + c = F_1 \\ 4a + 2b + c = F_2 \end{cases}$$

(3.10)

Решение этой системы позволяет найти все неизвестные a, b, c и $t_{опт}$ следующим образом. Исключим c из системы и получим систему из двух уравнений:

$$\begin{cases} a + b = F_1 - F_0 \\ 4a + 2b = F_2 - F_0 \end{cases}$$

(3.11)

Домножив первое уравнение на –4, затем на –2 и сложив со вторым найдём неизвестные a и b:

$$\begin{cases} -2b = F_2 - 4F_1 + 3F_0 \\ 2a = -2F_1 + F_2 + F_0 \end{cases} \quad \begin{cases} -b = \frac{1}{2}\left(F_2 - 4F_1 + 3F_0\right) \\ 2a = -2F_1 + F_2 + F_0 \end{cases}$$

Для оптимального шага получим следующее выражение:

$$t_{опт} = \frac{-b}{2a} = \frac{1}{2} \cdot \frac{F_2 - 4F_1 + 3F_0}{F_2 - 2F_1 + F_0}$$

(3.12)

В рассмотренном выше примере найдем оптимальный шаг этим методом:

$$t = 0, \quad X_0 = \{0,0\}, \quad F_0 = 30,$$
$$t = 1. \quad X_0 = \{1,1\}, \quad F_0 = 20,$$
$$t = 2, \quad X_0 = \{2,2\}, \quad F_0 = 14;$$
$$t_{опт} = \frac{1}{2} \cdot \frac{14 - 4 \cdot 20 + 3 \cdot 30}{14 - 2 \cdot 20 + 30} = \frac{24}{2 \cdot 4} = 3$$

Методы, в которых определяется оптимальный шаг, называются методами скорейшего поиска. Эти методы широко используются в энергетике, так как позволяют за меньшее число итераций находить оптимальные решения, хотя требуют большего количества вычислений на итерации.

В зависимости от способа определения возможного направления различают несколько методов, которые и рассмотрим далее:

- Метод покоординатного спуска
- Градиентный метод
- Метод Ньютона
- Минимизация квадратичной формы
- Методы нулевого порядка

Метод покоординатного спуска

В качестве возможных направлений здесь используются орты, т.е. единичные векторы по всем неизвестным. Оптимизация производится поочереднотолько по одному орту, начиная с e_1. На рисунке 3.8 показана траектория оптимизации с использованием скорейшего спуска. В полученной точке X_1 спуск проводится по $\vec{e}_2$ и т.д. После оптимизации по $\vec{e}_n$ первый цикл заканчивается.

Очередной цикл проводится аналогично.

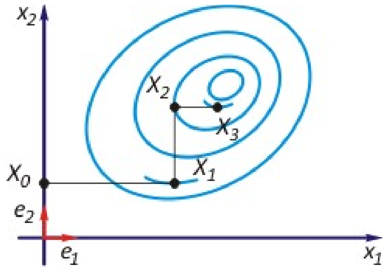


Рисунок 3.8 – Траектория оптимизации с использованием скорейшего спуска

Достоинство метода: простой алгоритм, по которому на каждом этапе производится оптимизация только по одной переменой при фиксированных значениях остальных. При этом могут использоваться различные методы одномерной оптимизации по одной переменной, в том числе и метод перебора.

Недостатком метода является не всегда хорошая сходимость. Особенно плохая сходимость отмечается для функций типа «овраг».

Градиентный метод

В этом методе в качестве возможного направления в текущей точке принимается градиент $\vec{S} = \nabla F$ при поиске максимума, или антиградиент $\vec{S} = -\nabla F$ при поиске минимума функции. Градиент, как известно, это вектор, направленный в сторону наибольшего возрастания функции, т.е. по перпендикуляру к касательной линии F = *const*. На рисунке 3.9. показано направление антиградиента.

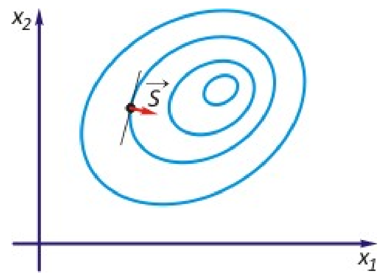


Рисунок 3.9 – Направление антиградиента

Основное уравнение градиентного метода при поиске минимума функции

$$X_{k+1} = X_k - \nabla F\big|_{X_k} \cdot t$$

(3.13)

При постоянном небольшом шаге t траектория поиска проходит перпендикулярно линиям F = const. При реализации скорейшего спуска поиск $t_{опт}$ может осуществляться следующим методом, основанным на определении градиентов в двух точках: в исходной при t = 0 и в конце некоторого пробного шага t_0.

Действительно, на направлении антиградиента целевая функция F зависит только от величины шага t (рисунок 3.10). Но вид F(t), а тем более dF/dt в большинстве случаев невозможно определить.

Для оценки зависимости dF/dt = φ(t) и делается пробный шаг t_0, в конце которого определяется величина $\left.\frac{dF}{dt}\right|_{t_0}$, соответствующая отрезку cd. Значение $\left.\frac{dF}{dt}\right|_{t_0}$ в исходной точке определяется отрезком ab.

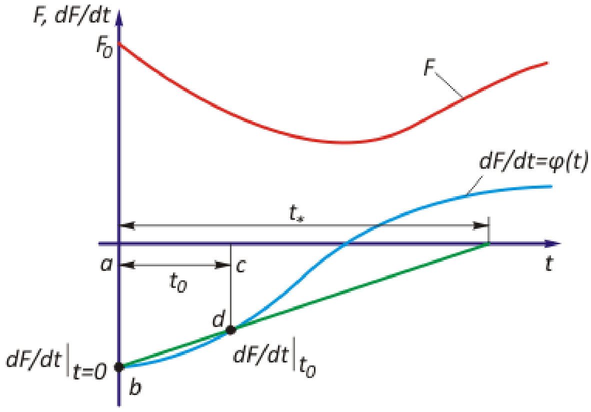


Рисунок 3.10 – Графическое отображение метода

Таким образом, появляется возможность аппроксимировать неизвестную зависимость φ(t) прямой, проходящей через две точки. Из подобия треугольников можно определить шаг t*, при котором эта прямая проходит через нуль, определяя шаг, не совпадающий с оптимальным, но близкий к нему

$$t_* = t_0 \frac{ab}{ab - cd} = t_0 \frac{\left.\frac{dF}{dt}\right|_{t=0}}{\left.\frac{dF}{dt}\right|_{t=0} - \left.\frac{dF}{dt}\right|_{t_0}}$$

(3.14)

Производные dF/dt могут быть найдены с учётом неявной зависимости X(t), определяемой основным уравнением градиентного метода.

Например, производная в текущей точке X_k

$$\left.\frac{dF}{dt}\right|_{t=0} = \left.\frac{dF}{dX}\right|_{t=0} \cdot \left.\frac{dX}{dt}\right|_{x_k} = \nabla F|_{x_k} \cdot \left(-\nabla F|_{x_k}\right) = -\sum_{(i)} \left.\frac{\partial F}{\partial x_i}\right|_{t=0} \cdot \left.\frac{\partial F}{\partial x_i}\right|_{t=0}$$

(3.15)

Аналогично найдем производную в конце пробного шага

$$\left.\frac{dF}{dt}\right|_{t_0} = \left.\frac{dF}{dx}\right|_{t_0} \cdot \left.\frac{dX}{dt}\right|_{x_k} = \nabla F|_{t_0} \cdot \left(-\nabla F|_{x_k}\right) = -\sum_{(i)} \left.\frac{\partial F}{\partial x_i}\right|_{t_0}$$

(3.16)

В этих выражениях учтено, что произведение градиентов определяется по правилу вычисления скалярного произведения векторов.

Ниже приведён алгоритм градиентного метода с выбором оптимального шага (рисунок 3.11).

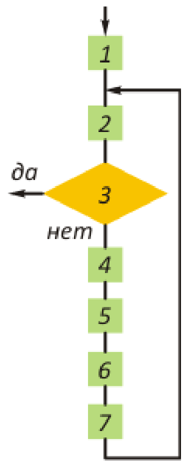


Рисунок 3.11 – Алгоритм градиентного метода с выбором оптимального шага

Блоки алгоритма выполняют следующие функции.

1. Исходное приближение $X = X_0$.
2. Определение градиента в текущей точке $\nabla F|_X$.
3. Проверка на заданную точность $|\nabla F| < eps$.
4. Выполнение пробного шага $X_1 = X - \nabla F \cdot t_0$.
5. Определение градиента в конце пробного шага $\nabla F|_{X_1}$.
6. Определение $t' = t_0 \dfrac{\nabla F|_X \cdot \nabla F|_X}{\nabla F|_X \cdot \nabla F|_X - \nabla F|_{X_1} \cdot \nabla F|_X}$.
7. Выполнение рабочего шага $X = X - \nabla F|_X \cdot t'$.

При реализации алгоритма наибольшие затраты времени связаны с определением градиента целевой функции. Как известно, составляющими вектор градиента ∇F являются частные производные по всем компонентам вектора X;

$$\nabla F = \left(\frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial x_2}, \ldots, \frac{\partial F}{\partial x_n}\right)$$

Все составляющие $\partial F/\partial x_i$ определяются легко, если функция F задана аналитически. В практических задачах такое встречается крайне редко. В этом случае используются численные методы определения производных, основанные на конечных приращениях (рисунок 3.12).

Как известно, производная в точке x_{i0} определяется наклоном касательной к кривой, т.е. пропорциональна tg α.

Определение через конечные приращения подменяет касательную секущей

$$\frac{\partial F}{\partial x_i} = \frac{\Delta F}{\Delta x} = \frac{F_2 - F_1}{\Delta x} = tg\,\beta \tag{3.17}$$

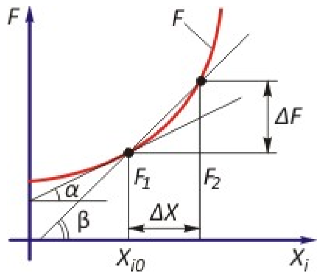


Рисунок 3.12 – Метод конечных приращений

Для уменьшения погрешностей используется метод центрированных приращений, при которых приращение делается вправо и влево от рабочей точки (рисунок 3.13)

$$\frac{\partial F}{\partial x_i} = \frac{F_+ - F_-}{2\Delta x} = tg\,\delta \tag{3.18}$$

При этом угол δ значительно ближе по величине к углу α, определяющему точное значение производной.

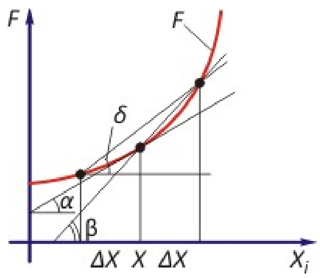


Рисунок 3.13 – Метод центрированных приращений

Метод Ньютона

Относится к методам второго порядка, более сложным по сравнению с градиентным, так как рассматриваемый метод строится на основе производных второго порядка. Как любой итерационный метод и метод Ньютона начинается с исходного приближения X_0.

Рассмотрим суть метода на примере функции только одной неизвестной f(x)→min.

Разложим f(x) в ряд Тейлора в точке x_0 с учётом первых слагаемых

$$f(x) \cong f(x_0) + \left.\frac{df}{dx}\right|_{x_0} \cdot \Delta x + \frac{1}{2} \cdot \frac{d^2 f}{dx^2} \cdot \Delta x^2 = \varphi(\Delta x) \to \min$$

Такое разложение эквивалентно замене исходной функции параболой в окрестности x_0 (рисунок 3.14).

Запишем условие минимума функции φ(Δx)

$$\frac{d\varphi(\Delta x)}{d(\Delta x)} = \left.\frac{df}{dx}\right|_{x_0} + \left.\frac{d^2 f}{dx^2}\right|_{x_0} \cdot \Delta x = 0$$

откуда найдем отклонение Δx

$$\Delta x = -\left.\frac{d^2 f}{dx^2}\right|_{x_0}^{-1} \cdot \left.\frac{df}{dx}\right|_{x_0}$$

Новое значение переменной, при котором имеет место минимум функции,

$$x_1 = x_0 + \Delta x = x_0 - \left.\frac{d^2 f}{dx^2}\right|_{x_0}^{-1} \cdot \left.\frac{df}{dx}\right|_{x_0}$$

(3.20)

В новой точке x_1 выполняются те же вычисления: f(x) заменяется разложением в ряд Тейлора и находится очередная точка x_2 и т.д.

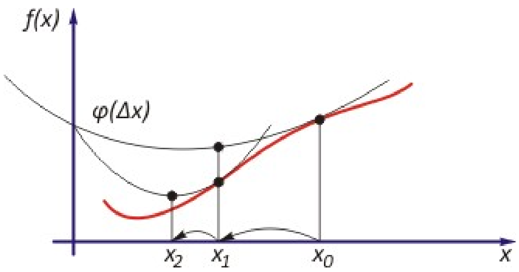


Рисунок 3.14 – Метод Ньютона

Общее уравнение метода:

$$x_{k+1} = x_k - \left.\frac{d^2 f}{dx^2}\right|_{x_k}^{-1} \cdot \left.\frac{df}{dx}\right|_{x_k}$$

(3.21)

Основное уравнение метода Ньютона для функции многих переменных:

$$X_{k+1} = X_k - G^{-1}(X_k) \cdot \nabla F\Big|_{X_k}$$

(3.22)

где X – вектор неизвестных, ∇F – вектор-градиент, G – матрица вторых производных (матрица Гессе), равная

$$G = \begin{bmatrix} \frac{\partial^2 F}{\partial x_1^2} & \frac{\partial^2 F}{\partial x_1 \cdot \partial x_2} & \cdots \\ \frac{\partial^2 F}{\partial x_2 \cdot \partial x_1} & \frac{\partial^2 F}{\partial x_2^2} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

Матрица Гессе является квадратной, симметричной относительно главной диагонали и имеет обратную матрицу.

Для функции нескольких переменных разложение в ряд Тейлора заменяет её в точке X_0 функцией второго порядка (эллипсоидом), и поиск минимума опускает точку X в центр эллипсоида.

Метод имеет хорошую сходимость, т.е. малое количество итераций. Платой за это является большой объем вычислений, связанный с формированием матрицы Гессе и её обращением.

Метод Ньютона используется в программах оптимизации режима энергосистем. При этом, учитывая слабую заполненность матрицы G ненулевыми элементами приращение Δx находят без обращения матрицы Гессе путём решения системы уравнений

$$G(X_k) \cdot \Delta x = \nabla F|_{x_k}$$

Минимизация квадратичной формы

На практике часто встречаются задачи, в которых целевые функции имеют квадратичную форму

$$F(X) = \frac{1}{2} X^T A X - B^T X \to \min$$

(3.24)

Поиск минимума функции здесь сводится к поиску решения системы линейных алгебраических уравнений AX = B.

Действительно, условие минимума функции $\nabla F(X) = AX - B = 0$, что и определяет решение задачи. Проверим справедливость этого положения на примере функции второго порядка.

Итак, имеем квадратичную форму F(X), в которой

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \; X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \; B = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Целевая функция, преобразованная в алгебраическую форму, будет иметь вид

$$F(X)=\frac{1}{2}\begin{bmatrix}x_1 & x_2\end{bmatrix}\begin{bmatrix}a_{11}x_1+a_{12}x_2\\ a_{12}x_1+a_{22}x_2\end{bmatrix}-\begin{bmatrix}b_1 & b_2\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}=$$
$$=\frac{1}{2}\left(a_{11}x_1^2+a_{12}x_2x_1+a_{21}x_1x_2+a_{22}x_2^2\right)-b_1x_1-b_2x_2=$$
$$=\frac{1}{2}\left(a_{11}x_1^2+a_{12}x_2x_1+a_{22}x_2^2\right)-b_1x_1-b_2x_2$$

для которой условия минимума функции

$$\left.\begin{aligned}\frac{\partial F}{\partial x_1}=a_{11}x_1+a_{12}x_2-b_1=0\\ \frac{\partial F}{\partial x_2}=a_{12}x_1+a_{22}x_2-b_2=0\end{aligned}\right\}$$

или в матричной форме AX – B = 0.

Методы нулевого порядка

Мы рассмотрели методы, связанные с вычислением производных, которые в зависимости от порядка производных относят к методам первого или второго порядка. Эти методы отличаются малым числом итераций, но большим объёмом вычислений на каждом приближении.

Существуют методы нулевого порядка, не требующие вычисления производных, основанные только на вычислении целевой функции. Особенность этих методов – большое число итераций при малом объёме вычислений на каждой.

При высоком быстродействии современных ЭВМ большое число итераций становится некритичным, а простые расчёты на шаге позволяют создавать более универсальные алгоритмы, способные решать более широкий круг задач нелинейного программирования.

Поэтому далее рассмотрим два метода нулевого порядка: метод случайного поиска и метод деформированного многогранника.

1. Метод случайного поиска

В этом методе возможное направление перехода из исходной точки $X_{(0)}$, в которой вычислено значение целевой функции F_0, определяется случайным образом.

Для этого вокруг точки $X_{(0)}$ рассматривается n-мерный куб (при n = 2 – квадрат) с ребром δ (рисунок 3.15). Координаты новой точки X_1 определяются с помощью генератора псевдослучайных чисел $0 \le \gamma \le 1$ с равномерным распределением.

Для n = 2 осуществляется генерация двух случайных чисел, равных, например, $\gamma_1 = 0{,}6$ и $\gamma_2 = 0{,}9$. Координаты случайной точки определяются по выражениям

$$x_{1(1)} = x_{1(0)} + (\gamma_1 - 0,5) \cdot \delta,$$
$$x_{2(1)} = x_{2(0)} + (\gamma_2 - 0,5) \cdot \delta$$
(3.25)

В полученной точке $X_{(1)}$ считается функция F_1 и сравнивается с F_0. Если $F_1 < F_0$, то $X_{(0)}$ переносится в $X_{(1)}$, вокруг нее вновь рассматривается куб с ребром δ, выбирается случайное направление и т.д.

В случае, когда $F_1 > F_0$ (см. рисунок 3.15), делается реверс неудачного направления при тех же γ_1 и γ_2

$$x^p_{1(1)} = x_{1(0)} - (\gamma_1 - 0,5) \cdot \delta,$$
$$x^p_{2(1)} = x_{2(0)} - (\gamma_2 - 0,5) \cdot \delta$$
(3.26)

В этом случае в новой точке $X_{(1)}{}^p$ значение $F_1{}^p$ вероятнее всего будет меньше F_0, куда и переместится текущая точка.

При небольшом размере куба и значительном удалении X_0 от точки минимума вероятность случайного выбора возможного направления составляет 0,5. По мере приближения к решению ребро куба уменьшается обычно в 2 раза. Основанием для этого является результат реверса направления, при котором окажется $F_1{}^p > F_0$.

Траектория спуска к оптимуму носит явно выраженный случайный характер и напоминает Броуновское движение.

Достоинство метода: простые вычисления на шаге и несложная логика в организации поиска траектории спуска к решению.

Недостатки: большое число итераций и сложная траектория.

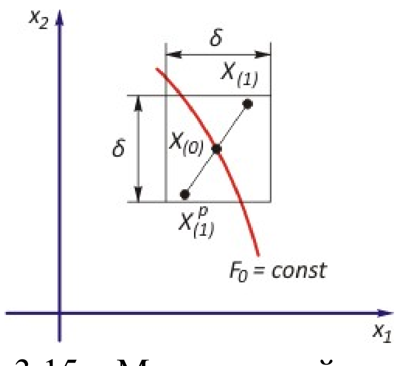


Рисунок 3.15 – Метод случайного поиска

2. Метод деформируемого многогранника

В этом методе вокруг исходной точки $X_{(0)}$ намечаются n + 1 равноудаленные точки, обозначенные на рисунке 3.16 номерами 1, 2, 3. В каждой точке считается целевая функция F(X) и определяются вершины с максимальным значением функции F_h и с минимальным значением F_l.

Затем производится деформация исходного многогранника (при n = 2 треугольника) и вершина F_h переносится через центр тяжести других вершин X_C в точку X_r с некоторым коэффициентом α:

$$X_r - X_C = \alpha \cdot (X_C - X_h)$$

(3.27)

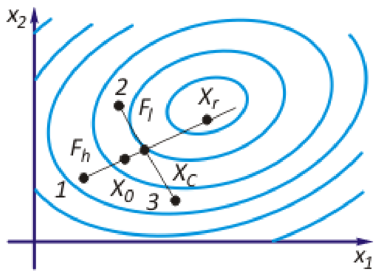


Рисунок 3.16 – Метод деформируемого многогранника

Координаты X_r определяются по формуле

$$X_r = X_C \cdot (1+\alpha) - \alpha \cdot X_h$$

(3.28)

При α = 1 деформацию называют отражением. В полученной точке X_r считается целевая функция F_r. Если $F_r < F_l$, то α увеличивают обычно до значения α = 1,5 и деформацию называют растяжением. При $F_r > F_l$ α уменьшают обычно до α = 0,5, что приводит к сжатию многогранника.

После этих операций вершина h переносится в точку X_r и новый многогранник подвергается деформации по тому же алгоритму.

По мере приближения к оптимуму появляется необходимость уменьшать размер многогранника, сжимая его в 2 раза относительно вершины, в которой отмечено минимальное значение функции.

3.3. Методы учета ограничений в форме равенств

Рассмотрим более сложную задачу нелинейного программирования, в которой обеспечивается поиск F(X)→min на допустимой области, определяемой системой ограничений в форме равенств G(X) = 0. Здесь вектор-функция G(X) ограничений состоит из m функций $g_j(X) = 0$.

Задача оптимизации имеет смысл, когда число уравнений m меньше числа неизвестных n. В этом случае существует множество решений, среди которых нужно найти решение, соответствующее минимуму функции.

Если m = n, то система ограничений может иметь единственное решение или быть несовместной, когда оптимизация теряет смысл.

Рассмотрим некоторые методы условной оптимизации.

Метод прямой оптимизации

Метод применяется, когда $g_j(X)$ описываются простыми, желательно линейными функциями, что позволяет выразить аналитически какие-либо m переменных из общего числа неизвестных, объединив их в новый вектор Y, через оставшиеся k = n–m, которые называют свободными или независимыми $X_{св}$. Найденные выражения $Y(X_{св})$ подставим в F(X), что позволит после преобразований получить новую по виду функцию $F(X_{св})$, для которой необходимо найти минимум.

Решение полученной задачи безусловной оптимизации любым из рассмотренных методов позволит найти k неизвестных, а после подстановки их в аналитические выражения определить и остальные неизвестные.

Метод приведенного градиента

Метод используется в том случае, если объективно существующие зависимости Y(X), определяемые системой ограничений G(X) = 0 не удается выразить в явном виде.

В математике известно правило определения производных с учетом не явно выраженных функций.

Запишем исходную задачу с учётом разделения исходного вектора неизвестных на составляющие Y и X_C:

$$
\begin{aligned}
&F(X) = F(Y, X_C) \to \min \\
&G(X) = G(Y, X_C) = 0
\end{aligned}
$$

(3.29)

Запишем производную $\partial F/\partial X_C$ с учетом неявной зависимости $Y(X_C)$

$$\frac{\partial F}{\partial X_C} = \left(\frac{\partial F}{\partial X_C}\right) + \left(\frac{\partial F}{\partial Y}\right) \cdot \frac{\partial Y}{\partial X_C}$$

(3.30)

где производные в скобках определяются с учетом явной зависимости F от X_C и Y. Производную $\partial Y/\partial X_C$ найдем из ограничения $G(Y, X_C) = 0$, которое в результате дифференцирования определяет условие

$$\left(\frac{\partial G}{\partial X_C}\right) + \left(\frac{\partial G}{\partial Y}\right) \cdot \frac{\partial Y}{\partial X_C} = 0$$

(3.31)

позволяющее получить

$$\frac{\partial Y}{\partial X_C} = -\left(\frac{\partial G}{\partial Y}\right)^{-1} \cdot \left(\frac{\partial G}{\partial X_C}\right)$$

(3.32)

Здесь ∂G/∂Y – квадратная матрица (m×m), для которой существует обратная.

После подстановки (3.32) в (3.30) получим

$$\frac{\partial F}{\partial X_C} = \left(\frac{\partial F}{\partial X_C}\right) - \left(\frac{\partial F}{\partial Y}\right) \cdot \left(\frac{\partial G}{\partial Y}\right)^{-1} \cdot \left(\frac{\partial G}{\partial X_C}\right) = \nabla F_П$$

Полученный градиент, составляющие которого определяются независимыми переменными, и называется приведённым градиентом.

Этот градиент может использоваться в процедуре градиентного метода:

$$X_{k+1} = X_k - \nabla F_П \big|_{X_k} \cdot t_k$$

(3.34)

Геометрически приведённый градиент $\nabla F_П$ является проекцией градиента ∇F на поверхность ограничений, а точнее – на плоскость, касательную нелинейной поверхности в текущей точке (рисунок 3.17).

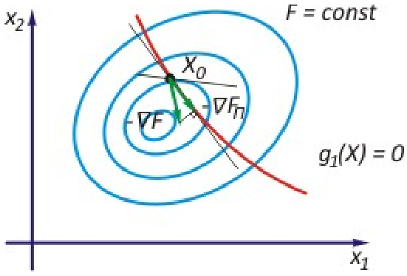


Рисунок 3.17

В точке условного минимума функции $\nabla F_П = 0$.

Метод неопределенных множителей Лагранжа

В точке, являющейся решением задачи F(X) → min при G(X) = 0, приведённый градиент равен 0, т.е.

$$\left(\frac{\partial F}{\partial X_C}\right) - \left(\frac{\partial F}{\partial Y}\right) \cdot \left(\frac{\partial G}{\partial Y}\right)^{-1} \cdot \left(\frac{\partial G}{\partial X_C}\right) = 0$$

(3.35)

Введём следующее обозначение, определяющее вектор λ так называемых неопределенных множителей Лагранжа,

$$\lambda = -\left(\frac{\partial F}{\partial Y}\right) \cdot \left(\frac{\partial G}{\partial Y}\right)^{-1}$$

С учётом λ перепишем условие (3.35) в виде системы из двух условий

$$\begin{cases} \left(\frac{\partial F}{\partial X_C}\right) + \lambda \cdot \left(\frac{\partial G}{\partial X_C}\right) = 0 \\ \left(\frac{\partial F}{\partial Y}\right) + \lambda \cdot \left(\frac{\partial G}{\partial Y}\right) = 0 \end{cases}$$

(3.36)

которые должны выполняться в точке, являющейся решением задачи.

Эту систему можно составить формально на основе условия минимума некоторой функции Лагранжа, определённой следующим образом.

$$L(X_C, Y, \lambda) = F(X_C, Y) + \lambda \cdot G(X_C, Y) = F(X_C, Y) + \sum_{j=1}^{m} \lambda_j \cdot g_j(X_C, Y) \to \min$$

(3.37)

Действительно, условие минимума функции Лагранжа

$$\begin{cases} \frac{\partial L}{\partial x_i} = \frac{\partial F}{\partial x_i} + \sum_{(j)} \lambda_j \cdot \frac{\partial g_j}{\partial x_i} = 0 \\ \frac{\partial L}{\partial y_i} = \frac{\partial F}{\partial y_i} + \sum_{(j)} \lambda_j \cdot \frac{\partial g_j}{\partial y_i} = 0 \\ \frac{\partial L}{\partial \lambda_j} = g_j(X, Y) = 0 \end{cases}$$

(3.38)

Полученная система из k + m + m = n + m уравнений позволяет найти все n неизвестных исходного вектора X и m множителей Лагранжа.

Оказывается, функцию Лагранжа можно составить и без разделения исходного вектора X на составляющие:

$$L(X, \lambda) = F(X) + \sum_{j=1}^{m} \lambda_j \cdot g_j(X) \to \min$$

(3.39)

Условие минимума этой функции определяется следующей системой:

$$\begin{cases} \frac{\partial L}{\partial x_i} = \frac{\partial F}{\partial x_i} + \sum_{(j)} \lambda_j \cdot \frac{\partial g_j}{\partial x_i} = 0 \\ \frac{\partial L}{\partial \lambda_j} = g_j(X) = 0 \end{cases}$$

(3.40)

Величина множителей Лагранжа λ_j может иметь практический интерес, если ограничения представляются в форме $g_j(X) - b_j = 0$. В этом случае множители Лагранжа можно рассматривать как производные $\lambda_j = \frac{\partial L}{\partial b_i}$, что позволяет оценивать по найденному в ходе решения (3.40) их значению возможность уменьшения целевой функции за счет изменения b_j.

Ниже приводится пример решения задачи условной оптимизации рассмотренными методами.

Пример решения задачи нелинейного программирования

Рассмотрим пример решения следующей задачи

$$F(x_1, x_2) = 5x_1^2 + x_1x_2 + 3x_2^2 + 10 \to \min$$

$$g_1(x_1, x_2) = x_1 + x_2 - 10 = 0$$

1. Метод прямой оптимизации

   В качестве независимой примем x = x_1. Тогда зависимой будет x_2 = 10–x_1. Подставим выражение для x_2 в целевую функцию

   $$F(x_1) = 5x_1^2 + x_1(10 - x_1) + 3(10 - x_1)^2 + 10 \to \min$$

   Условие минимума ее:

   $$\frac{dF}{dx_1} = 10x_1 + 10 - 2x_1 + 6(10 - x_1)(-1) = 14x_1 - 50 = 0$$

   Откуда

   $$x_1 = \frac{50}{14} = 3{,}57 \text{ и } x_2 = 10 - x_1 = \frac{90}{14} = 6{,}43, \quad F = 166$$

2. Метод приведенного градиента при t = 0,1

   Воспользуемся формулой (3.33) для приведенного градиента

   $$\nabla F_П = \left(\frac{\partial F}{\partial X_C}\right) - \left(\frac{\partial F}{\partial Y}\right)\cdot\left(\frac{\partial G}{\partial Y}\right)^{-1}\cdot\left(\frac{\partial G}{\partial X_C}\right) = 0$$

   В решаемой задаче x_1 является составляющей X_C, а x_2 относится к Y. С учетом этого

   $$\left(\frac{\partial F}{\partial X_C}\right) = 10x_1 + x_2; \left(\frac{\partial G}{\partial Y}\right) = 1; \left(\frac{\partial F}{\partial Y}\right) = x_1 + 6x_2; \left(\frac{\partial G}{\partial X_C}\right) = 1$$

   С учетом этого приведенный градиент

   $$\nabla F_П = 10x_1 + x_2 - (x_1 + 6x_2)\cdot 1\cdot 1 = 9x_1 - 5x_2$$

   В качестве исходного приближения примем

   $$x_1 = 0, x_2 = 10$$

   Основное уравнение градиентного метода

   $$x_{1(k+1)} = x_{1(k)} - \nabla F_П|\cdot t$$

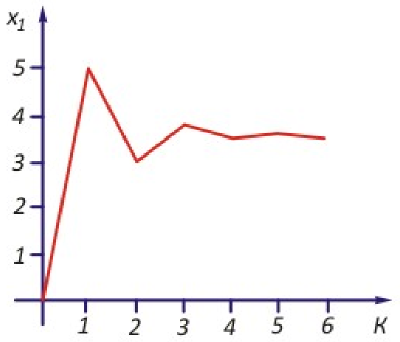
   $$x_2 = 10 - x_1$$

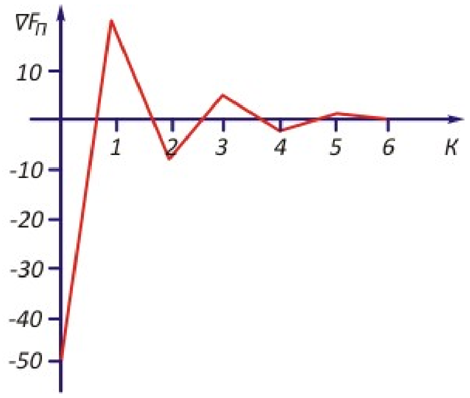
   Результаты расчёта сведены в таблице 3.1, а ход итерационного процесса показан на рисунке 3.18. Видно, что процесс сходится, так как $|\nabla F|$ стремится к нулю.

Таблица 3.1

K	x1	(x)	x2
0	0	-50	10
1	5	20	5
2	3	-8	7
3	3,8	3.2	6,2
и т. д.			

Рисунок 3.18 – Ход итерационного процесса





3. Метод множителей Лагранжа.

Составим функцию Лагранжа

$$L(X,\lambda) = F(X) + \lambda_1 \cdot g_1(X) = 5x_1^2 + x_1 x_2 + 3x_2^2 + 10 + \lambda(x_1 + x_2 - 10) \to \min$$

Условие минимума её определяется следующей системой

$$\begin{cases} \dfrac{\partial L}{\partial x_1} = 10x_1 + x_2 + \lambda = 0 \\ \dfrac{\partial L}{\partial x_2} = x_1 + 6x_2 + \lambda = 0 \\ \dfrac{\partial L}{\partial \lambda} = x_1 + x_2 - 10 = 0 \end{cases}$$

Решим полученную систему уравнений, выразив λ из первых двух уравнений и приравняв их получим

$$\begin{cases} 10x_1 + x_2 = x_1 + 6x_2 \\ x_1 + x_2 - 10 = 0 \end{cases}$$

откуда нетрудно найти решение, в котором $x_1 = \dfrac{50}{14}$, $x_2 = \dfrac{90}{14}$ и $-\lambda = 10 \cdot \dfrac{50}{14} + \dfrac{90}{14} = \dfrac{590}{14}$. Найденное значение –λ определяет величину производной $\dfrac{\partial L}{\partial \lambda}$, которая позволяет оценить изменение функции L, а также и F при изменении b. Например, при $\Delta b = 1$ целевая функция возрастет на $\frac{590}{14}$.