М. Ю. АНДРАМОНОВ

МЕТОДЫ ГЛОБАЛЬНОЙ МИНИМИЗАЦИИ ДЛЯ НЕКОТОРЫХ КЛАССОВ ОБОБЩЕННО ВЫПУКЛЫХ ФУНКЦИЙ

Монография

Казань – 2001

ББК 22.18
УДК 519.6

Печатается по постановлению
Редакционно-издательского совета
Казанского математического общества

Андрамонов М. Ю. Методы глобальной минимизации для некоторых классов обобщенно выпуклых функций. – Казань: Издательство Казанского математического общества, 2001 г. – 190 с.

В работе построен ряд методов минимизации обобщенно выпуклых функций, включая квазивыпуклые, возрастающие выпуклые по лучам, звездные относительно бесконечности и липшицевы функции. Разработана схема двойственности, основанная на возрастающих функциях. Для решения задачи о гамильтоновом цикле предложена схема, основанная на марковских цепях и функциях типа минимума.

Книга предназначена для специалистов в области вычислительной математики, доступна аспирантам и студентам старших курсов.

Табл. 6. Библиогр. 101 назв.

Научный редактор – профессор А.М.Елизаров.
Рецензенты – профессор А.В. Лапин,
доцент В.П. Чуев.

Оглавление

Список обозначений

$\mathbb{R}^n$ – n-мерное евклидово пространство;

$\mathbb{R}^n_+$ – конус векторов из $\mathbb{R}^n$ с неотрицательными координатами;

$\mathbb{R}^n_{++}$ – множество векторов из $\mathbb{R}^n$ со строго положительными координатами;

$\|x\| = \sqrt{\sum_{i=1}^n x_i^2}$;

$\|x\|_1 = \sum_{i=1}^n |x_i|$;

$[x, y] = \sum_{i=1}^n x_i y_i$;

$[x; y]$ – отрезок с концами x и y;

B_ε – шар радиуса ε в $\mathbb{R}^n$ с центром в нуле;

$B(z, \varepsilon)$ – шар радиуса ε в $\mathbb{R}^n$ с центром в z;

$\langle x, y\rangle = \min_{i\in\mathcal{T}(y)}\ x_i y_i,\ \mathcal{T}(y) = \{i : y_i > 0\}$.

$$\frac{1}{x} = (u_i)_{i=1,\dots,n} \quad \text{с} \quad u_i = \frac{1}{x_i} \quad \text{для} \quad i \in \mathcal{T}(x);$$

$$u_i = 0 \quad \text{для} \quad i \notin \mathcal{T}(x).$$

$\partial_L f(x)$ – L-субдифференциал функции f в точке x.

Введение

В последнее время большое внимание привлекают задачи глобальной оптимизации обобщенно выпуклых функций. Существует ряд классов таких функций, например, квазивыпуклые, псевдовыпуклые и т. д. [32, 33, 73, 4, 5, 6]. Теория этих задач хорошо разработана, и в большинстве случаев задачи квазивыпуклого программирования могут быть решены эффективно. Они решаются методами субградиентного типа, которые достаточно изучены. Однако с применением абстрактного выпуклого анализа сфера обобщенной выпуклости значительно расширилась. Теперь такие функции, как липшицевы, возрастающие выпуклые по лучам, возрастающие звездные, могут также быть названы обобщенно выпуклыми. Подобные функции гораздо труднее минимизировать в силу многоэкстремальности. Был предложен ряд алгоритмов липшицева программирования, в частности, методы ветвей и границ и методы случайного поиска (см. [74] для подробного изложения, а также [53, 97]). Известные методы привели к решению задач с сотнями переменных, но их эффективность зависит от специальной структуры задачи.

Глобальная оптимизация, как теория, так и методы, становится все более важной областью изучения. Существует признанная практическая необходимость в методах, эффективно решающих задачи глобальной оптимизации (см. [54]). Однако, вообще говоря, по своей природе такие задачи чрезвычайно сложны для решения. Одной из причин является необходимость в средствах, предоставляющих глобальную информацию об изучаемых объектах (функциях и множествах). Следует отметить, что стандартные методы нелинейного программирования неприменимы в силу многоэкстремальности задач глобальной оптимизации. Наконец, вычислительные сложности глобальной оптимизации в силу ее комбинаторного характера делают разработку эффективных общих методов маловероятной. Тем не менее, несмотря на эти трудности, можно построить методы для некоторых специальных сильно структурированных задач, используя абстрактную выпуклость.

В данной монографии описаны новые обобщения классического метода секущих плоскостей выпуклой оптимизации, применимые к широкому спектру невыпуклых оптимизационных задач. Предложен также новый эффективный метод решения задачи квазивыпуклого программирования.

Простейшим примером задачи глобальной оптимизации является минимизация выпуклой функции на выпуклом множестве. Существует ряд методов решения такой задачи, основанных на глобальных или на локальных подходах, так как локальные экстремумы совпадают с глобальными. Локальные подходы по сути являются градиентными или субградиентными алгоритмами, в которых используется локальная аппроксимация функции, осуществляемая субдифференциалом. Развитие методов для общих негладких задач часто основано на обобщении субградиентного метода. Существуют, однако, методы выпуклой оптимизации, глобальные по своей природе. Первым таким примером является метод секущих плоскостей, основанный на представлении выпуклой функции в виде верхней грани аффинных минорант. Методы пучков (см., например [51]), также глобальны по своей природе.

В первой главе рассматривается обобщенный метод секущих плоскостей, примененный к некоторым структурированным задачам глобальной оптимизации, в которых целевая функция обладает обобщенными аффинными минорантами. Эти обобщенные аффинные функции являются опорными функциями в абстрактном выпуклом анализе.

Абстрактная выпуклость появилась недавно как область исследований с потенциально большими приложениями к глобальной оптимизации (см. [11, 95, 84, 70]).

Пусть имеется семейство функций на множестве X. Будем говорить, что функция $f : X \to \mathbb{R}^n_+$ абстрактно выпукла по отношению к множеству H, если существует множество $U \subseteq H$ такое, что

$$f(x) = \sup\{h(x) : h \in U\}.$$

Это соотношение обобщает представление выпуклой функции как верхней грани аффинных минорант. В этом случае H представляет собой множество обобщенно аффинных функций (в большинстве практических невыпуклых задач они нелинейны, и, возможно, разрывны). Важное средство, которое мы будем использовать из абстрактного выпуклого анализа – это обобщенный субдифференциал. Однако в отличие от субдифференциала Кларка из невыпуклого анализа он несет глобальную, а не локальную информацию о функции. Конечно, невозможно обсуждать общие задачи абстрактного выпуклого программирования с надеждой получения эффективных вычислительных методов. Следует отметить, что несколько алгоритмов были предложены для решения специальных классов задач глобальной оптимизации. Во многих случаях может быть показано, что данные алгоритмы укладываются в общую схему обобщенного метода секущих плоскостей, изложенного ниже. В частности, к ним относятся алгоритмы Младинео [68], Вуда [99, 100, 101], Пиявского-Шуберта [14, 94].

В первой главе рассмотрен специальный сильно структурированный, но тем не менее широкий класс абстрактно выпуклых функций: так называемые ВВЛ функции (возрастающие выпуклые по лучам), определенные на неотрицательном ортанте. Здесь функция является возрастающей по отношению к обычному отношению порядка в $\mathbb{R}^n$. Хорошо известно (см. [81]), что ВВЛ функция абстрактно выпукла по отношению к следующему семейству функций:

$$H = \{h(x) = \min_{i\in\mathcal{T}(\ell)} \ell_i x_i - c \,:\, \ell \in \mathbb{R}^n_+,\ c \in \mathbb{R},\ \ x \in \mathbb{R}^n_+\},$$

где $\mathcal{T}(\ell) = \{i \,:\, \ell_i > 0\}$. Данный класс функций весьма широк, что подтверждается тем, что любая полунепрерывная снизу на единичном симплексе функция может быть продолжена до ВВЛ функции. Мы представляем несколько примеров, чтобы показать широту данного класса функций. Специальная структура множества H привела к рассмотрению секущих углов как естественного обобщения понятия секущей плоскости выпуклой оптимизации. Метод секущих углов, предложенный нами, численно реализован, и начальные результаты приведены в главе 7. Многие задачи глобальной оптимизации с целевой функцией, определенной на неотрицательном ортанте $\mathbb{R}^n_+$, могут быть успешно решены применением методов абстрактной выпуклости, основанных на функциях типа $x \to \min_{i \in \mathcal{T}(l)} l_i x_i$ с $\mathcal{T}(l) = \{i : l_i > 0\}$ ([81, 21, 79]). В отличие от квадратичных функций функции типа минимума негладкие, но, по-видимому, это обстоятельство не очень важно для глобальной оптимизации. В то же время функция минимума есть конечная комбинация координатных функций и сохраняет некоторые свойства этих линейных функций. Такие функции легко строить. Оптимизационные задачи с функциями типа минимума могут быть легко переформулированы как специальные задачи частично целочисленного программирования [26], для которых есть ряд программных пакетов (CPLEX, LPsolve, LINsolve и т.д.).

Как отмечено выше, основная проблема детерминистских методов в глобальной оптимизации – это ее комбинаторный характер. Как указано в [53], задачи глобальной оптимизации по своему существу сложны для решения и должны быть вычислительно трудоемкими. Наше начальное тестирование было ограничено задачами малого размера, однако результаты являются обещающими и указывают на глобальный характер алгоритма (и комбинаторные свойства главной подзадачи). Мы включили примеры задач со многими локальными экстремумами, в которых алгоритм успешно сходится, даже стартуя из локального, но не глобального экстремума. Большее численное тестирование требует более мощного компьютерного обеспечения. Следует отметить, что используя метод из [21, 22], можно получить верхние и нижние оценки оптимального значения, что полезно для численной реализации.

В главе 2 рассматривается класс целевых функций, которые возрастают и имеют звездные по отношению к $+\infty$ множества уровня (мы их называем ВЗ функциями). Такие функции образуют решетку и имеют ряд хороших свойств, облегчающих их минимизацию. Задачи математического программирования с ВЗ функциями имеют различные приложения на практике, в частности, в математической экономике, что делает их очень важными. В частности, Интрилигатор [57] считает функции с данным свойством основным видом производственных функций.

В общем случае ВЗ функция имеет локальные экстремумы и не является ни выпуклой, ни вогнутой. Поэтому к ее минимизации нельзя применить методы локального поиска и методы вогнутой оптимизации. Поскольку ВЗ функция может быть приближена функциями типа минимума в силу ее свойств абстрактной выпуклости, задача минимизации ВЗ функции может быть сведена к последовательности минимаксных задач или к последовательности задач частично целочисленного программирования. Конечно, каждая такая задача трудна для решения, и сложность быстро растет по отношению к размерности начальной задачи. Однако, используя специальную технику сокращения размерности, можно получить хорошие численные результаты, если число переменных невелико. Может быть выгодным комбинировать методы, основанные на абстрактной выпуклости, с локальным поиском, получая гибридные методы. В этом случае абстрактное выпуклое программирование служит для нахождения приближенного решения, а локальный поиск улучшает его. Следует отметить, что многие оптимизационные задачи могут быть сведены к задачам с ВЗ целевыми функциями заменой переменных.

В главе 2 также показано, что во многих случаях задача оптимизации с произвольной липшицевой целевой функцией может быть сведена к минимизации ВВЛ функции. Это приводит к применению метода секущих углов к задачам липшицева программирования. Безусловно, ряд вычислительных трудностей появляется при данном подходе, в частности, можно получить плохо обусловленную целевую функцию. Тем не менее, во многих случаях применение нашего подхода было успешным, поскольку были найдены глобальные оптимальные решения для ряда задач невыпуклого программирования.

В главе 3 предложен эффективный метод минимизации квазивыпуклой функции, основанный на коническом проектировании. Начальная задача с ограничениями сводится к задаче безусловной минимизации, что облегчает нахождение оптимального решения. Предложенный алгоритм весьма эффективен для минимизации линейной функции на выпуклом множестве. Численные эксперименты показали, что метод конического проектирования превосходит по эффективности известные градиентные методы.

В главе 4 рассмотрена хорошо известная задача о гамильтоновом цикле (сокращенно ЗГЦ), которая может быть описана следующим образом: *ЗГЦ*: В ориентированном графе найти путь, проходящий через каждую вершину ровно один раз перед возвращением в начальную вершину, или определить, что такого пути не существует.

ЗГЦ, как правило, рассматривается как очень сложная задача с алгоритмической точки зрения. Хорошо известно, что обнаружение полиномиального алгоритма опровергнет гипотезу $P \neq NP$, которая считается справедливой. Однако есть хорошие эвристические алгоритмы для решения многих разновидностей ЗГЦ, большинство из этих эвристик основано на комбинаторных подходах. В данной монографии мы используем результаты неортодоксального подхода к ЗГЦ, разработанного в статьях Чена и Филара [41] и Филара и Красса [40]. Такой подход включает в себя вложение проблемы в контролируемую марковскую цепь и исследование индуцированной эргодической структуры. В итоге получается задача минимизации неопределенной квадратичной формы с известным оптимальным значением, равным нулю.

К сожалению, нет эффективных алгоритмов решения такой задачи для достаточно большого числа переменных. В данной монографии применяется подход к решению ЗГЦ с помощью марковских цепей и функций типа минимума. Мы представляем результаты численных экспериментов для графов умеренного размера. Предварительное тестирование показало эффективность подхода для различных классов графов.

В главе 5 мы рассматриваем схему двойственности, основанную на возрастающих функциях. Одна из важнейших задач оптимизации – минимизация нелинейной функции на допустимом множестве, заданном конечным числом линейных и нелинейных ограничений. Существует ряд алгоритмов решения такой задачи. Их можно разделить на две большие группы методов. Первая группа состоит из прямых методов, в которых на каждой итерации находится возможное направление и делается шаг вдоль него. Эти методы включают в себя метод условного градиента, методы приведенных градиентов, метод Зойтендейка, метод проекции градиента и др. (см. [67, 2] и библиографию). Однако они имеют ряд недостатков, касающихся сходимости и трудности выбора параметров. Поэтому во многих случаях предпочтительно применять двойственные методы, образующие вторую группу. Последняя включает в себя методы штрафных функций (см. [39]), метод центров ([56, 8]) и методы модифицированных функций Лагранжа (см. [67, 47] и ссылки в них). Они дают информацию о двойственной задаче и возможность найти нижние оценки оптимального значения целевой функции на допустимом множестве, что очень важно для минимизации невыпуклых функций. Часто эти алгоритмы называют методами последовательной безусловной оптимизации, так как они основаны на сведении исходной задачи с ограничениями к последовательности задач безусловной оптимизации. Иногда необходимо много раз решать вспомогательные задачи без ограничений. Однако существуют методы точных штрафных функций, в которых достаточно один раз решить вспомогательную задачу для получения оптимального решения исходной задачи.

В данной работе предлагается подход к построению вспомогательных функций для двойственных методов, основанный на возрастающих функциях. Мы доказываем эквивалентность исходной и вспомогательной задач, что ведет к построению семейства методов последовательной безусловной оптимизации. Мы показываем, что многие известные конкретные примеры штрафных функций или модифицированных функций Лагранжа укладываются в нашу общую схему.

При построении штрафных функций для ограничений в форме неравенств показано, что когда значения штрафных параметров достаточно велики, можно получить решение задачи с ограничениями с любой заранее заданной точностью. Затем мы развиваем тот же подход для ограничений-равенств.

Обычно предпочтительно использовать модифицированные функции Лагранжа, а не штрафные функции, так как вспомогательная задача может быть плохо обусловлена для больших значений штрафных параметров (для обсуждения модифицированных функций Лагранжа см. [30, 67] и библиографию). Мы предлагаем схему построения модифицированных функций Лагранжа, которая позволяет найти приближенное оптимальное решение исходной задачи, даже если у классической функции Лагранжа нет седловых точек. Одна из возможностей – комбинировать подход, основанный на модифицированных функциях Лагранжа, с методом центров и применить известные методы негладкой оптимизации для решения вспомогательной задачи.

В главе 6 мы предлагаем параметрический подход к задачам глобальной оптимизации специального вида. Очевидно, необходимо рассматривать специальные классы задач, для которых можно построить численные алгоритмы, успешно использующие структуру целевой функции и/или ограничений.

Задачи нахождения экономического равновесия (см. [85, 96, 45]) и решения нелинейных уравнений (см. [35, 74]) дают нам два важных класса, для которых можно разработать схемы, позволяющие найти глобальный оптимум за короткий отрезок времени. Для решения таких задач мы используем параметрический подход. Он отличается от методов продолжения по параметру [53, 20], туннельных методов [65] и методов гомотопии (см. [53, 96] и ссылки в них), которые наиболее часто используются для решения таких задач.

Мы предлагаем искать глобальный оптимум, надлежащим образом подбирая веса (параметры) в задаче. Как задача решения нелинейных уравнений, так и задача нахождения экономического равновесия обладают важным свойством. Если мы введем различные веса для функций задачи, глобальное решение не изменится. Однако другие стационарные точки изменятся, что позволяет преодолеть главное препятствие глобальной оптимизации – выйти из стационарной точки, не являющейся глобальным оптимумом.

Веса можно изменять детерминистским или стохастическим способами. Простейший способ – применить случайный поиск в пространстве весов, для чего имеется большое число алгоритмов (см. [97, 76, 91] и ссылки в них), или можно применить комбинаторный подход.

В данной работе предлагается ряд концептуальных схем для решения задач глобальной оптимизации специального вида, включая минимизацию взвешенной суммы квадратов и поиск экономического равновесия в модели, предложенной в [85], и представлены численные результаты.

Сходимость зависит от свойств используемого метода локального поиска и мер областей притяжения глобального и локальных минимумов (см.

обсуждение в [97]).

Мы обсуждаем ряд проблем, возникающих в нашем подходе. Одна из них – наличие стационарных неоптимальных точек для всех весов. Для данной ситуации нами предлагается алгоритм возмущений, преодолевающий это препятствие во многих случаях. При этом разработаны модификации алгоритмов для неограниченного допустимого множества, когда итерационная последовательность может уходить на бесконечность.

Наш подход может рассматриваться в общей схеме траекторных методов, ряд которых был предложен для решения задач нелинейного программирования и систем нелинейных уравнений (см. [97, 37]). Мы предполагаем, что имеется локальный алгоритм, позволяющий найти стационарную точку (это может быть метод наискорейшего спуска, метод пучков или любой другой алгоритм). При изменении весов осуществляется переход от одной стационарной точки к другой, и при некоторых предположениях сходимость к глобальному оптимуму может быть доказана.

В главе 7 мы рассматриваем результаты численных экспериментов для различных задач глобальной оптимизации.

Автор выражает благодарность научному редактору А.М. Елизарову за помощь в подготовке монографии, рецензентам А.В. Лапину и В.П. Чуеву за полезные замечания, а также А.М. Рубинову за ряд ценных советов.

Монография написана при поддержке грантов Российского фонда фундаментальных исследований, которая осуществлялась по проектам N 99-01-00173, N 01-01-06199. Автор благодарит РФФИ за оказанную поддержку.

Глава 1

Абстрактная выпуклость и метод секущих углов

В данной главе мы предлагаем обобщение метода секущих плоскостей выпуклой оптимизации, применимое к очень широким классам невыпуклых задач глобальной оптимизации. Построения основаны на абстрактной выпуклости.

1.1 Абстрактно выпуклые функции и субдифференциалы

Нами рассматриваются задачи глобальной оптимизации некоторых обобщенно выпуклых функций. Прежде всего, мы приведем известные определения и результаты теории абстрактного выпуклого анализа (см. [95, 84, 11, 70]), которые будут необходимы для построения численных алгоритмов.

Пусть L – множество вещественнозначных функций $h(x)$, определенных на множестве X, со следующим свойством (A): функция

$$h(x) = \ell(x) - c, \quad x \in X \tag{1.1}$$

не принадлежит L для всех $\ell \in L$ и всех ненулевых $c \in \mathbb{R}$. Далее мы будем использовать следующие определения (см. [81, 84, 11, 95, 70]).

Определение 1.1.1 *Пусть L – множество функций, удовлетворяющих условию (A). Функция h вида (1.1) с $\ell \in L$ и $c \in \mathbb{R}$ называется L-аффинной функцией. Мы будем обозначать множество L-аффинных функций H_L.*

Определение 1.1.2 *Пусть H есть множество конечных функций, определенных на множестве X. Функция $f : X \to \mathbb{R} \cup \{+\infty\}$, определенная на множестве X, называется абстрактно выпуклой по отношению к H (или H-выпуклой), если есть множество U такое, что для всех $x \in X$ имеем*

$$f(x) = \sup\{h(x) : h \in U\}.$$

Мы будем рассматривать только конечные абстрактно выпуклые функции.

Определение 1.1.3 *Пусть f – конечная функция, определенная на множестве X. Множество $s(f, H)$ всех L-аффинных минорант f называется опорным множеством. Так,*

$$s(f, H) = \{h \in H : (\forall x \in X) \quad h(x) \leq f(x)\}.$$

Тогда для H-выпуклой функции имеем $f(x) = \sup\{h(x) : h \in s(f, H)\}$.

Нас интересуют точки x, в которых достигается супремум.

Определение 1.1.4 *Пусть L – множество функций, определенных на X, со свойством (A), и f – H_L-выпуклая функция. Пусть $x_0 \in X$. Тогда множество*

$$\partial_L f(x_0) = \{\ell \in L : (\forall x \in X) \quad \ell(x) - \ell(x_0) \leq f(x) - f(x_0)\}$$

называется L-субдифференциалом функции f в точке x_0.

Следующее простое предложение устанавливает связь между субдифференциалами и опорными множествами.

Предложение 1.1 *Пусть f – H_L-выпуклая функция, $x_0 \in X$ и $\ell \in L$. Пусть $h(x) = \ell(x) - (\ell(x_0) - f(x_0))$. Тогда $\ell \in \partial_L f(x_0)$ тогда и только тогда, когда $h \in s(f, H_L)$.*

1.2 Схема обобщенного градиентного спуска

Пусть L – множество функций, определенных на открытом множестве, включающем компактное выпуклое множество X из гильбертова пространства.

Пусть L обладает свойством (A). Мы будем рассматривать H_L-выпуклую функцию f, определенную на X. Предположим, что f L-субдифференцируема в любой точке $x \in X$, т. е. $\partial_L(x) \neq \emptyset$ для всех $x \in X$. Из Предложения 1.1 следует, что для всех $x \in X$ множество

$$M(x) = \{h \in \mathrm{s}(f, H_L) : h(x) = f(x)\} \tag{1.2}$$

непусто. Рассмотрим следующий вариант обобщенного метода секущих плоскостей для задачи (см. [60, 51, 75])

$$f(x) \to \min, \quad x \in X \tag{1.3}$$

Обобщенный метод секущих плоскостей

Шаг 0. Пусть k := 0. Выбрать произвольно $x_o \in X$.

Шаг 1. Найти $\ell_k \in \partial_L f(x_k)$ и положить $h_k \in M(x_k)$, где $h_k(x) = \ell_k(x) - (\ell_k(x_k) - f(x_k))$.

Шаг 2. Найти глобальный оптимум задачи

$$\max_{0 \le i \le k} h_i(x) \to \min, \quad x \in X. \tag{1.4}$$

Пусть y^* – решение данной задачи.

Шаг 3. Положить $k := k+1, x_k = y^*$ и вернуться к Шагу 1.

Сходимость метода была установлена Палляшке и Ролевичем [70] в очень общем случае. Их доказательство очень сложное, поэтому даются более простые доказательства для интересующих нас случаев.

Чтобы изучить сходимость алгоритма, введем следующие величины (они были рассмотрены в [16] для выпуклого случая):

$$f_k(x) = \max_{0 \le i \le k} h_i(x) \quad \text{для} \quad x \in X,$$

$$\mu_k = f_k(x_k) \ \ (k = 0, 1, 2, \ldots); \qquad \lambda_k = f_{k-1}(x_k) \ \ (k = 1, 2 \ldots). \tag{1.5}$$

Укажем некоторые свойства функций f_k и чисел μ_k и λ_k:

1) $f_o(x) \le f_1(x) \le \ldots \le f_k(x) \ldots \le f(x) \quad$ для всех $\quad x \in X$.

2) $\mu_k = f(x_k) = h_k(x_k)$.

В самом деле,

$$f_k(x_k) \le f(x_k) = h_k(x_k) \le \max_{0 \le i \le k} h_i(x_k) = f_k(x_k).$$

Таким образом, $\mu_k = f_k(x_k) = f(x_k) = h_k(x_k)$.

3) Неравенство $\lambda_k \le \lambda_{k+1}$ выполняется для всех k.

   Действительно,

$$\lambda_k = f_{k-1}(x_k) = \min_{x \in X} f_{k-1}(x) = \min_{x \in X} \max_{0 \le i \le k-1} h_i(x)$$
$$\le \min_{x \in X} \max_{0 \le i \le k} h_i(x) = \min_{x \in X} f_k(x) = \lambda_{k+1}.$$

4) Из монотонности последовательности $\{\lambda_k\}$ следует, что $\lim_{k \to +\infty} \lambda_k$ существует.

5) $\lambda_k \le \min_{x \in X} f(x) \le \mu_k$ для всех $k = 1, 2, \ldots$.

   Действительно, для решения u задачи (1.3) имеем

$$\min_{x \in X} f(x) = f(u) \ge f_{k-1}(u) \ge \min_{x \in X} f_{k-1}(x) = \lambda_k.$$

   С другой стороны, $\mu_k = f(x_k) \ge \min_{x \in X} f(x)$.

Непосредственно из этих свойств следует, что если для некоторого k мы имеем $x_k = x_{k+1}$, то x_k – решение задачи (1.3). Действительно, мы можем взять $\ell_k = \ell_{k+1}$ в этом случае. Тогда

$$\lambda_{k+1} = f_k(x_{k+1}) = \max_{0 \le i \le k} h_i(x_{k+1}) = \max_{0 \le i \le k+1} h_i(x_{k+1}) =$$
$$= f_{k+1}(x_{k+1}) = \mu_{k+1}.$$

Из свойства 5) следует, что $f(x_k) = \min_{x \in X} f(x)$.

Мы опишем два набора условий, обеспечивающих сходимость алгоритма.

Теорема 1.1 *Пусть множество L состоит из вогнутых функций. Пусть последовательность $\{x_k\}$, генерируемая алгоритмом, бесконечна. Предположим, что производные по направлению $h_k'(x)$ вогнутых функций h_k равномерно ограничены на множестве X:*

$$\|h_k'(x)\| = \max_{\|u\| \le 1} |h_k'(x, u)| \le r < +\infty \quad \textit{для всех} \quad x \in X, k = 0, 1, \ldots \tag{1.6}$$

Тогда каждая предельная точка x последовательности (x_k) есть решение задачи (1.3).

Доказательство. Пусть $x = \lim_{j\to\infty} x_{k_j}$. Для каждого j рассмотрим вогнутую функцию h_i с $i \le k_{j-1}$ и найдем суперградиент a_i этой функции в точке x_{k_j}. Пусть $b_i(x) = h_i(x_{k_j}) + [a_i, x - x_{k_j}]$. Тогда $b_i(x) \ge h_i(x)$ для всех $x \in X$ и $b_i(x_{k_j}) = h_i(x_{k_j})$. Мы имеем

$$\begin{aligned}
\mu_{k_{j-1}} &= f_{k_{j-1}}(x_{k_{j-1}}) = \max_{0\le i\le k_{j-1}} h_i(x_{k_{j-1}}) \le \max_{0\le i\le k_{j-1}} b_i(x_{k_{j-1}}) \\
&= \max_{0\le i\le k_{j-1}} h_i(x_{k_j}) + [a_i, x_{k_{j-1}} - x_{k_j}] \\
&\le \max_{0\le i\le k_j-1} h_i(x_{k_j}) + \max_{0\le i\le k_{j-1}} \|a_i\|\|x_{k_{j-1}} - x_{k_j}\|.
\end{aligned}$$

Из условий теоремы легко следует, что $\|a_i\| \le r$. Пусть $\lambda = \lim_{k\to\infty} \lambda_k$. Поскольку $\|x_{k_{j-1}} - x_{k_j}\| \to 0$, имеем $\limsup \mu_{k_j} \le \lambda$. С другой стороны, из неравенства $\mu_{k_j} \ge \lambda_{k_j-1}$ следует $\liminf \mu_{k_j} \ge \lambda$. Таким образом, $\mu_{k_j} = f(x_{k_j}) \to \lambda$. Очевидно, что $\lambda = \min\{f(x') : x' \in X\}$. $\square$

Здесь и далее $\square$ означает завершение доказательства.

Замечание 1.1 Пусть $\nu_k = \min_{i\le k} \mu_i$. Последовательность ν_k убывает, и мы имеем

$$\nu_k = \min_{i\le k} \mu_i \ge \max_{i\le k} \lambda_i = \lambda_k.$$

Таким образом, алгоритм генерирует такие убывающую последовательность (ν_k) и возрастающую последовательность (λ_k), что

$$\nu_k \ge \min_{x\in X} f(x) \ge \lambda_k \quad \text{и} \quad \lim_{k\to\infty} \nu_k = \min_{x\in X} f(x) = \lim_{h\to\infty} \mu_k.$$

Докажем еще одну теорему сходимости.

Теорема 1.2 *Допустим, что:*

1) *множество L состоит из функций, непрерывных на X;*

2) *множество $M(X) = \cup_{x\in X} M(x)$, где $M(x)$ определено (1.2), равномерно секвенциально компактно (для любой последовательности $h_k \in M(X)$ существует равномерно сходящаяся подпоследовательность h_{k_i});*

3) субдифференциальное отображение $x \to \partial_L f(x)$ замкнуто.

Тогда если последовательность $\{x_k\}$, построенная алгоритмом, бесконечна, то каждая ее предельная точка является решением задачи (1.3).

Доказательство. Пусть $h_k \in M(x_k)$ – последовательность L - аффинных функций, используемых для построения последовательности x_k. Пусть x – предельная точка этой последовательности. Из условий 2) и 3) следует, что существует подпоследовательность k_j такая что $x_{k_j} \to x$ и h_{k_j} равномерно сходится к функции $h \in M(x)$. Пусть ε – произвольное положительное число. Поскольку h – непрерывная функция, мы имеем для достаточно больших j:

$$h(x) \le h(x_{k_j}) + \frac{\varepsilon}{2} \le h_{k_{j-1}}(x_{k_j}) + \varepsilon \le \max_{o \le i \le k_j - 1} h_i(x_{k_j}) + \varepsilon = \lambda_{k_j} + \varepsilon, \qquad (1.7)$$

где (λ_k) определено (1.5). Из (1.7) и свойств последовательности $\{\lambda_k\}$ следует, что $h(x) \le \min_{x' \in X} f(x')$. Так как $x \in X$, имеем $h(x) \ge \min_{x' \in X} f(x')$. □

Замечание 1.2 В ряде случаев субдифференциал может быть очень большим. В таких случаях можно использовать специальное подмножество $m(x)$ множества $\partial_L f(x)$ для построения последовательности $\{x_k\}$. В этом случае мы требуем замкнутость отображения m вместо $\partial_L f$.

Замечание 1.3 Можно применить схему обобщенного метода секущих плоскостей для некоторых классов разрывных функций. Например, мы можем взять как L класс двухступенчатых функций (см. [83]) вида:

$$\ell(x) = \begin{cases} 0, & v(x) \le a, \\ c, & v(x) > a, \end{cases}$$

где v – линейная функция, a – вещественное число, $c \ge 0$. Хорошо известно (см., например [95, 83]), что функция f H_L-выпукла тогда и только тогда, когда она полунепрерывна снизу и квазивыпукла. В данной ситуации мы можем применить схему обобщенного метода секущих плоскостей. Поскольку двухступенчатые функции разрывны, условия теоремы сходимости не выполняются. Если мы добавим дополнительные предположения (например, что итерационная точка лежит достаточно глубоко внутри некоторого вспомогательного множества), то алгоритмы сходятся, и мы получаем хорошо известные методы типа секущих плоскостей для квазивыпуклого программирования (см., например, [46]).

При численной реализации описанных алгоритмов имеются две основные трудности. Первая из них – вычисление элемента ℓ_k из $\partial_L f(x_k)$. В общем случае очень трудно найти численно элемент L-субдифференциала, однако это возможно для ряда важных частных случаев. Например, знание константы Липшица для липшицевой функции позволяет найти элемент из $\partial_L f(x_k)$, где L – множество минимумов линейных функций. В следующей части мы рассматриваем широкий класс функций, а именно, возрастающие выпуклые по лучам функции, для которых легко найти элемент $L-$субдифференциала. Вторая сложность – решение подзадачи на Шаге 2. Оно легко реализуется для линейных или выпуклых функций h_i, однако если h_i – минимумы аффинных функций или невыпуклые квадратичные функции, подзадача имеет комбинаторную природу и решается легко только для малых размерностей.

Существует ряд известных алгоритмов, которые могут рассматриваться как частные случаи обобщенного метода секущих плоскостей. Прежде всего, это метод секущих плоскостей выпуклой оптимизации (см. [60]). Однако есть методы глобальной оптимизации, укладывающиеся в нашу общую схему. В [68] рассмотрен специальный случай, в котором (если мы имеем задачу минимизации, а не максимизации) функция

$$f_k(x) = \max_{1 \le i \le k} \{f(x_i) - K\|x - x_i\|\}$$

минимизируется, где K – константа Липшица целевой функции. Очевидно, если f липшицева, она H_L-выпукла и L-субдифференцируема по отношению к множеству функций вида

$$\ell(x) = -K\|x - \overline{x}\|, \overline{x} \in \mathbb{R}^n.$$

Следовательно, алгоритм Р.Младинео [68] есть обобщенный метод секущих плоскостей для соответствующих функций $h(x) \equiv \ell(x) + c, c = const.$

Алгоритмы, построенные в [99, 100, 101], основаны на аппроксимации подграфика функции $f(x)$ объединением симплициальных конусов. В действительности они используют абстрактную выпуклость с множеством H, являющимся подмножеством множества минимумов аффинных функций.

Многие методы минимизации функции одной переменной, например, метод Пиявского-Шуберта и методы, на нем основанные (см.[14, 94, 48, 49, 74]), также являются специальными случаями общего метода, представленного выше. Поэтому можно доказать сходимость указанных алгоритмов по одной и той же технике.

Замечание 1.4 Мы можем рассматривать наш подход в рамках методов последовательного оценивания снизу, изученных в [54]. Метод связан с алгоритмами внешних аппроксимаций. Теорема 1.2 похожа на теорему сходимости методов внешней аппроксимации, рассмотренных в [55], но независима от нее.

Чтобы ускорить сходимость обобщенного метода секущих плоскостей, может быть очень полезно вводить гибридизацию, то есть использовать локальный поиск для нахождения локального минимума, а затем пытаться найти обобщенным методом секущих плоскостей локальный минимум с меньшим значением целевой функции. Также, если текущая итерационная точка достаточно близка к глобальному минимуму, локальный поиск в большинстве случаев найдет его с хорошей точностью. Таким образом, цель обобщенного метода секущих плоскостей – найти точку, достаточно близкую к глобальному минимуму.

Гибридный метод I

Шаг 0. Пусть $k := 0, x_0 \in X$. Положить $f_r := f(x_0), x_r := x_0$.

Шаг 1. Вычислить элемент $\ell_k \in \partial_L f(x_k)$ и взять $h_k \in M(x_k)$, где $h_k(x) = \ell_k(x) - (\ell_k(x_k) - f(x_k))$.

Шаг 2. Решить задачу:

$$\max\{h_0(y), \ldots, h_k(y)\} \to \min$$

$$y \in X$$

Пусть y^* – ее решение.

Шаг 3. Если $f(y^*) < f_r$, то вернуться к Шагу 4. Иначе перейти к Шагу 5.

Шаг 4. Найти локальный минимум $\overline{y}$ любым релаксационным методом, стартуя из y^*. Положить $f_r := f(\overline{y}); x_r := \overline{y}$. Положить $k := k+1,\ x_k := \overline{y}$ и перейти к Шагу 1.

Шаг 5. Положить $k := k+1, x_k := y^*, f_r := f_r, x_r := x_r$ и перейти к Шагу 1.

Теорема 1.3 *Пусть выполняются условия Теоремы 1.1 или Теоремы 1.2. Обозначим через* $\tilde{X}$ *множество всех локальных минимумов* f *на* X*, и пусть множество*

$$F = \{z : z = f(y) \quad \textit{для некоторого} \quad y \in \tilde{X}\}$$

конечно. Тогда глобальный минимум будет найден за конечное число итераций.

Доказательство. Так как F – конечное множество и

$$f^* = \min_{x \in X} f(x)$$

принадлежит F, мы можем отсортировать его элементы:

$$f_1 \geq f_2 \geq \ldots \geq f_m = f^*.$$

Если условие на Шаге 3 никогда не выполняется после конечного числа шагов, то есть $f(y^*) \geq f_r$, мы просто выполняем итерации обобщенного метода секущих плоскостей и имеем

$$f^* \geq \max\{h_0(x_{s+q}), h_1(x_{s+q}), \ldots, h_{s+q-1}(x_{s+q})\} \geq$$

$$\geq \max\{h_{s+1}(x_{s+q}), \ldots, h_{s+q-1}(x_{s+q})\}, \quad q = 1, 2, ...$$

Здесь s – число ранее добавленных ограничений. Повторяя доказательство для обобщенного метода секущих плоскостей, получаем, что любая сходящаяся подпоследовательность $\{x_{s+q}\}$ стремится к глобальному минимуму. Но это означает, что на некоторой итерации условие на Шаге 3 будет выполнено и мы найдем локальный минимум со значением, меньшим f_r. Так, если f_r было равно f_p для некоторого $p \leq m$, на некоторой итерации оно будет меньше или равно f_{p+1}. Так как F – конечное множество, через конечное число итераций получаем $f_r = f^*$. □

Если мы хотим применить нерелаксационный метод локального поиска (например, метод Шора [92]), можно использовать следующий

Гибридный метод II

Шаг 0. Пусть $k := 0; x_0 \in X$. Положить $S = \emptyset$.

Шаги 1-2. Как в Гибридном методе I.

Шаг 3. Найти локальный минимум $\overline{y}$ любым методом, стартуя с y^*.

Шаг 4. Если $\overline{y} \notin S$, то $k := k+1, x_k := \overline{y}, S := S \cup \{\overline{y}\}$ и перейти к Шагу 1.

Шаг 5. Положить $k := k+1, x_k := y^*, S := S$ и вернуться к Шагу 1.

Теорема 1.4 *Пусть выполняются условия Теоремы 1.1 или Теоремы 1.2. Для $\varepsilon > 0$ положим*

$$L_f^*(\varepsilon) = \{x \in X | f(x) \leq f^* + \varepsilon\}.$$

Пусть число локальных минимумов конечно и существует такое $\varepsilon > 0$, что локальный метод сходится к глобальному решению. Если начальная точка принадлежит $L_f^*(\varepsilon)$, то глобальный минимум будет найден за конечное число итераций.

Доказательство. Пусть $N = |\tilde{X}|$. Предположим, что мы знаем r $(0 \leq r < N)$ локальных минимумов, но не глобальный минимум. Тогда за конечное число итераций мы либо найдем новый локальный минимум, либо найдем ε-оптимальную точку (доказательство аналогично Теореме 1.3). Число локальных минимумов конечно, так что когда ε становится достаточно мало, мы попадаем в глобальный оптимум. □

1.3 Возрастающие выпуклые по лучам функции

Пусть $\mathbb{R}_+^n$ – конус всех n-мерных векторов с неотрицательными координатами. Рассмотрим множество L функций, определенных на множестве $X = \mathbb{R}_+^n$ формулой

$$\ell(x) = \min_{i \in \mathcal{T}(\ell)} \ell_i x_i, \tag{1.8}$$

где $(\ell_1, \ldots, \ell_n) \in \mathbb{R}_+^n$ и $\mathcal{T}(\ell) = \{i : \ell_i > 0\}$. Предположим, что минимум по пустому множеству равен нулю. Будем обозначать вектор $(\ell_1, \ldots \ell_n)$ тем же символом ℓ, что и функцию, генерируемую этим вектором по формуле (1.8). Мы будем обозначать также минимум в (1.8) $\langle \ell, x \rangle$, так что $\ell(x) = \langle \ell, x \rangle$.

Обычное скалярное произведение векторов x и y из $\mathbb{R}^n$ будет обозначаться $[x, y]$. Для $x \in \mathbb{R}^n_+$ мы будем использовать следующие обозначения :

$$\frac{1}{x} = (u_i)_{i=1,\dots,n} \quad \text{с} \quad u_i = \frac{1}{x_i} \quad \text{для} \quad i \in \mathcal{T}(x);$$

$$u_i = 0 \quad \text{для} \quad i \notin \mathcal{T}(x).$$

Предположим, что в пространстве $\mathbb{R}^n$ введено обычное отношение порядка: если $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, $y = (y_1, ..., y_n) \in \mathbb{R}^n$, то $x \geq y$ тогда и только тогда, когда $x_i \geq y_i$ для всех i. Нам понадобятся следующие определения.

Определение 1.3.1 *Функция f, определенная на $\mathbb{R}^n_+$, называется возрастющей, если из $x \geq y$ следует $f(x) \geq f(y)$. Функция f строго возрастает в точке $y \in \mathbb{R}^n_+$, если $x < y$ (то есть $x \leq y$ и $x \neq y$) означает $f(x) < f(y)$.*

Определение 1.3.2 *Конечная функция f, определенная на $\mathbb{R}^n_+$, называется выпуклой по лучам (ВЛ), если функция $f_x(\alpha) = f(\alpha x)$ выпукла на открытом луче $(0, +\infty)$ для всех $x \in \mathbb{R}^n_+$. Функция f называется вогнутой по лучам (ВГЛ), если f_x вогнута на $(0, +\infty)$.*

Следующие результаты доказаны в [81].

Теорема 1.5 *Функция $f : \mathbb{R}^n_+ \to \mathbb{R}$ является H_L-выпуклой тогда и только тогда, когда f возрастает и выпукла по лучам (ВВЛ).*

Теорема 1.6 *Пусть f – ВВЛ функция и $x \neq 0$. Тогда*

$$\partial_L f(x) \supseteq \left\{ \frac{v}{x} : v \in \partial f_x(1) \right\}.$$

Если, кроме того, f строго возрастает в точке x, то

$$\partial_L f(x) = \left\{ \frac{v}{x} : v \in \partial f_x(1) \right\},$$

где $f_x(\alpha) = f(\alpha x)$ для $\alpha > 0$. В частности, $f'(x, x)/x \in \partial_L f(x)$, где

$$f'(x, x) = \lim_{\alpha \to +\infty} \frac{1}{\alpha}[f(x + \alpha x) - f(x)]$$

– производная функции f в точке x по направлению x.

1.4 Примеры ВВЛ функций

Обозначим через Y класс ВВЛ функций, определенных на $\mathbb{R}^n_+$. Этот класс весьма широк. Известно (см. [1]), что каждая полунепрерывная снизу функция, определенная на единичном симплексе, может быть продолжена до функции из класса Y. Опишем некоторые свойства класса Y.

1) Если $f_1, f_2 \in Y$, то $f_1 + f_2 \in Y$.

2) Если $f \in Y$ и $\lambda > 0$, то $\lambda f \in Y$.

3) Если $f_1, f_2 \in Y$ и $f_1, f_2 \geq 0$, то $f_1 \times f_2 \in Y$.

4) Пусть $(f_\alpha) \in Y$ $(\alpha \in A)$, где A – произвольное множество индексов. Если $f(x) \equiv \sup_{\alpha \in A} f_\alpha(x) < +\infty$ для всех $x \in \mathbb{R}^n_+$,, то $f \in Y$.

5) Пусть $f_k \in Y$ и $f_k(x) \to f(x)$ для всех $x \in \mathbb{R}^n_+$. Если $f(x) < +\infty$ для всех x, то $f \in Y$.

Будем обозначать через $Y(x)$ класс ВВЛ функций, строго возрастающих в точке x. Очевидно, свойства 1) и 2) справедливы и для класса $Y(x)$. Более того, если хотя бы одна из функций f_1, f_2 строго возрастает, то и их сумма строго возрастает. Свойство 3) выполняется, только если $f_1 > 0, f_2 > 0$, а свойство 4) выполняется, если множество индексов A конечно.

Приведем теперь примеры ВВЛ функций и строго возрастающих ВЛ функций. Простейший пример – возрастающие положительно однородные степени $k \geq 1$ функции. Приведем несколько специфических примеров.

Пример 1. Линейная функция $f(x) = \sum_{i=1}^n \alpha_i x_i$ с $\alpha_i \geq 0$ для всех $i \in I = \{1, 2, ..., n\}$ есть ВВЛ функция. Если $\alpha_i > 0$ для всех $i \in I$, то f строго возрастает в каждой точке $x \neq 0$.

Пример 2. Функция Кобба-Дугласа

$$f(x) = Cx_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n}$$

с $\alpha_i > 0$ и $\sum_{i=1}^n \alpha_i \geq 1$ является ВВЛ функцией. Ясно, что эта функция строго возрастает в каждой точке $x \in \operatorname{int} \mathbb{R}^n_+$ и не возрастает строго в любой точке $x \notin \operatorname{int} \mathbb{R}^n_+$.

Заметим, что при $\sum_{i=1}^{n} \alpha_i = 1$ функция f вогнута.

Функции такого типа имеют многочисленные применения в математической экономике и используются во многих экономико-математических моделях, поэтому задача максимизации такой функции на ограниченном множестве важна с практической точки зрения.

Пример 3. Рассмотрим функцию вида

$$f(x) = (\alpha_1 x_1^p + \alpha_2 x_2^p + \dots \alpha_n x_n^p)^{\frac{s}{p}},$$

где $p > 0, s \geq 1, \alpha_i \geq 0$ для всех $i \in I$. Очевидно, f возрастает и положительно однородна степени s, так что она является ВВЛ функцией. Если $\alpha_i > 0$ для всех $i \in I$, то f строго возрастает для всех $x \neq 0$. Если $s = 1$ и $p < 1$, то f – вогнутая функция.

Мы можем построить более сложные примеры ВВЛ функций, использовав свойства класса Y.

Пример 4. Полином

$$f(x) = \alpha_o + \sum_{i \in I} \alpha_i x_i + \sum_{i_1, i_2 \in I} \alpha_{i_1 i_2} x_{i_1} x_{i_2} + \dots + \sum_{i_1, i_2, \dots, i_m \in I} \alpha_{i_1 i_2 \dots i_m} x_{i_1} x_{i_2} \dots x_{i_m}$$

с неотрицательными коэффициентами (кроме α_o, который может быть отрицательным) – ВВЛ функция на $\mathbb{R}^n_+$. Легко указать условия, когда данный полином строго возрастает. Например, если все коэффициенты, кроме α_o, положительны, то f строго возрастает в каждой точке $x \neq 0$.

В частности, квадратичная функция

$$f(x) = \sum_{i,j \in I} \alpha_{ij} x_i x_j$$

с $\alpha_{ij} \geq 0$ есть ВВЛ функция на $\mathbb{R}^n_+$.

Пример 5. Пусть

$$f(x) = \sup_{z \in Z} \sum_{i,j \in I} \alpha_{ij}(z) x_i x_j \quad (x \in \mathbb{R}^n_+)$$

с $\alpha_{ij}(z) \geq 0$. Здесь Z – произвольное множество индексов. Если $f(x)$ конечна для всех $x \in \mathbb{R}^n_+$, то f есть ВВЛ функция.

Пример 6. Пусть f – произведение неотрицательных линейных форм, то есть

$$f(x) = \prod_{p=1}^{m} (\alpha_{1p}x_1 + \alpha_{2p}x_2 + \ldots + \alpha_{np}x_n)$$

с $\alpha_{ip} \geq 0$ $(i = 1, ..., n; p = 1, ..., m)$. Тогда f – ВВЛ функция. Таким образом, задача мультипликативного линейного программирования (см. [61]) с неотрицательными коэффициентами может рассматриваться как специальный случай задачи минимизации ВВЛ функции.

1.5 Минимизация ВВЛ функций

Применим схему обобщенного метода секущих плоскостей к минимизации ВВЛ функций.

Пусть L – множество всех функций типа минимума вида (1.1). Для ВВЛ функции f будем находить элемент ℓ из L-субдифференциала в точке x, применяя Теорему 1.6. А именно, будем брать ℓ в форме

$$\ell(y) = f'(x,x) \min_{i \in \mathcal{T}(x)} \frac{y_i}{x_i}.$$

Очевидно, $\ell(x) = f'(x,x)$. Тогда мы можем представить обобщенный метод секущих плоскостей в следующем виде (будем называть получающийся алгоритм методом секущих углов, см. [22, 78]).

Метод секущих углов

Шаг 0. Пусть $k := 0$. Выбираем произвольно $x_0 \in X$.

Шаг 1. Вычисляем вектор ℓ_k с координатами ℓ_{ki} :

$$\ell_{ki} = \frac{f'(x_k, x_k)}{x_{ki}}, \quad \text{если} \quad x_{ki} \neq 0; \quad \ell_{ki} = 0 \quad \text{если} \quad x_{ki} = 0,$$

где x_{ki} – i-я координата вектора x_k .

Шаг 2. Определяем функцию h_k :

$$h_k(x) = \min_{i \in \mathcal{T}(x)} \ell_{ki}x_i - ((f'(x_k, x_k) - f(x_k)),$$

где ℓ_{ki} - i-я координата вектора ℓ_k .

Шаг 3. Находим глобальный оптимум y^* задачи

$$\max_{0 \le i \le k} h_i(x) \to \min, \tag{1.9}$$

$$x \in X.$$

Шаг 4. Полагаем $k := k+1, x_k := y^*$ и возвращаемся к Шагу 1.

Сходимость метода устанавливает

Теорема 1.7 *Пусть компактное подмножество X конуса $\mathbb{R}^n_+$ обладает следующими свойствами:*

1) *существует такое число r_1, что для всех $x = (x_1, x_2, \ldots, x_n)$ неравенство $x_i \ge r_1$ выполняется для всех i;*

2) $\sup_{x \in X} f'(x, x) = r_2 < +\infty$.

Тогда каждая предельная точка последовательности, построенной методом секущих углов, есть глобальный минимум функции f на множестве X .

Доказательство. Вытекает непосредственно из Теоремы 1.1. □

Замечание 1.5 Мы можем построить разновидность метода секущих углов, чтобы решить следующую задачу:

$$f(x) \to \max, \tag{1.10}$$

$$x \in X,$$

где X – компактное множество и f – функция, представимая как инфимум семейства L-аффинных функций. Здесь L – множество функций типа максимума вида $\ell(x) = \max_{i=1,2,\ldots,n} \ell_i x_i$ с неотрицательными векторами ℓ.

Функция f такого вида возрастает и вогнута по лучам (ВВГЛ). Подзадача в этом случае имеет вид:

$$\min_{i=0,\ldots,k} \left(\max_{j=1,\ldots,n} \ell_{ij} x_j + b_i \right) \to \max,$$

$$x \in X.$$

Мы можем применить подход, предложенный выше, для решения этой задачи, если X – нормальное компактное множество, то есть из $x \in X,\ 0 \le x' \le x$ следует $x' \in X$.

1.6 Решение подзадачи для ВВЛ функций

Рассмотрим подход, который позволяет построить численный метод для решения подзадачи, которая возникает в методе секущих углов для минимизации возрастающих выпуклых по лучам функций.

Напомним формулировку подзадачи:

$$\max_{i=\overline{0,k}}\{\langle \ell_i, x\rangle + b_i\} \to \min, \tag{1.11}$$

$$x \in X.$$

Мы можем переписать ее в следующей эквивалентной форме:

$$t \to \min, \tag{1.12}$$

$$\langle \ell_i, x\rangle + b_i \le t, \quad i = \overline{0,k},$$

$$x \in X.$$

Если обозначим координаты каждого вектора l_i как l_{ij}, подзадача также может быть представлена в следующем виде:

$$t \to \min, \tag{1.13}$$

$$\min_{j=\overline{1,n}} \ell_{ij} x_j + b_i \le t, \quad i = \overline{0,k},$$

$$x \in X.$$

Предположим для простоты, что все значения l_{ij} положительны. Если мы можем найти допустимую точку задачи (1.12), оптимальное t может быть найдено простым методом деления отрезка пополам. Таким образом, достаточно решить систему или определить, что у нее нет решений.

Рассмотрим два случая для допустимого множества:

1) X выпукло;

2) $X = \{x \in \mathbb{R}^n_+ : g(x) \ge 0,\ 0 \le x_i \le c,\ i = 1,..,n\}$, где g – возрастающая функция и c – положительная константа.

В первом случае, если X – многогранник, получаем последовательность задач решения систем линейных неравенств. Для произвольного выпуклого множества мы получаем последовательность задач решения систем линейных и выпуклых нелинейных неравенств.

Рассмотрим систему в полной форме:

$$
\begin{aligned}
\min\{\ell_{11}x_1; \ell_{12}x_2; ...; \ell_{1n}x_n\} &\le t - b_1,\\
\min\{\ell_{21}x_1; \ell_{22}x_2; ...; \ell_{2n}x_n\} &\le t - b_2,\\
&\dots\dots\dots\dots\dots\\
\min\{\ell_{k1}x_1; \ell_{k2}x_2; ...; \ell_{kn}x_n\} &\le t - b_k,\\
x &\in X.
\end{aligned}
$$

Применим к ее решению метод, похожий на динамическое программирование. Заметим, что если X – многогранник, мы имеем специальную задачу дизъюнктивного программирования, для которой есть общие результаты и методы (см. [26, 27, 29]).

Предположим, что $k > 1$ и $n > 1$. Будем называть задачу нахождения решения системы (1.12) с n переменными и k функциями типа минимума задачей $P(k, n)$. Покажем, как можно свести $P(k, n)$ к последовательности задач меньшей размерности.

Рассмотрим величины

$$
\theta_1 = \frac{t - b_1}{\ell_{11}};\ \theta_2 = \frac{t - b_2}{\ell_{21}}; \dots \theta_n = \frac{t - b_k}{\ell_{k1}}.
$$

Отсортируем их в возрастающем порядке (так как t теперь фиксировано, это можно сделать):

$$
\theta_{i_1} \le \theta_{i_2} \le \dots \theta_{i_k}.
$$

Предположим, что $\theta_{i_s} < x_1 \le \theta_{i_{s+1}}$. Для однородности положим $\theta_{i_0} = 0$, $\theta_{i_{k+1}} = c$, где c – верхняя граница для переменных (если множество X выпукло и замкнуто, мы можем найти ее методами выпуклого программирования). Все ограничения с номерами $i_{s+1}, ..., i_k$ удовлетворены, а остальные нарушены. Тогда мы получаем последовательность семейств подзадач:

$$
\begin{aligned}
\min\{\ell_{i_1 2}x_2; \ell_{i_1 3}x_3; \dots; \ell_{i_1 n}x_n\} &\le t - b_{i_1},\\
&\dots\dots\dots\dots\dots\\
\min\{\ell_{i_s 2}x_2; \ell_{i_s 3}x_3; \dots; \ell_{i_s n}x_n\} &\le t - b_{i_s},
\end{aligned}
$$

$$\theta_{i_s} < x_1 \le \theta_{i_{s+1}},$$
$$x \in X.$$

Таким образом, число переменных в функциях типа минимума убывает хотя бы на единицу в каждом случае. Число функций типа минимума не возрастает. Поэтому, применив процедуру последовательно для $x_2, x_3, ..., x_n$, мы можем в конце прийти к одной из трех ситуаций: $P(1,q)$, $P(0,q)$ или $P(s,0)$. Рассмотрим все три случая $(1 \le q \le n,\ 1 \le s \le k)$.

a) $P(1,q)$: это задача вида

$$\min\{\ell_1 x_1; ...; \ell_q x_q\} \le t - b,$$
$$x \in \tilde{X},$$

где $\tilde{X}$ получается из X добавлением новых верхних и нижних границ для $n - q$ переменных.

Мы просто рассматриваем все случаи:

$$\ell_j x_j \le t - b,$$
$$x \in \tilde{X},$$

для $j = \overline{1,q}$ или минимизируем вогнутую функцию $\min_j \ell_j x_j$, используя любой стандартный метод (см. [54]).

b) P(0,q): это задача вида

$$a_i \le x_i \le b_i, \quad i = \overline{1,n},$$
$$x \in X.$$

Если X выпукло, мы можем применить любой метод нахождения допустимой точки выпуклого множества [10]. Если X – множество вида

$$X = \{x \in \mathbb{R}^n : g(x) \ge 0\},$$

где g – возрастающая функция, просто берем $x_i = b_i$, $i = \overline{1,n}$. Если эта точка не принадлежит X, то у системы нет решений.

c) P(s,0): здесь мы имеем

$$\ell_{i_1 s} x_s \le t - b_{i_1},$$

$$\ell_{i_2 s} x_s \le t - b_{i_2},$$
$$\dots\dots\dots$$
$$\ell_{i_p s} x_s \le t - b_{i_p},$$
$$x \in \tilde{X}.$$

Таким образом,

$$x_s \le \min_{r=\overline{1,p}} \frac{t - b_{i_r}}{\ell_{i_r} s},$$

и мы должны искать точку из X с дополнительными верхними и нижними границами переменных.

Если $X = \{x \in \mathbb{R}^n_+ : \ g(x) \ge 0, 0 \le x_i \le c \ \ i = \overline{1,n}\}$, где g – возрастающая функция, все подзадачи имеют вид

$$g(x) \ge 0,$$

$$0 \le \alpha_i \le x_i \le \beta_i \le c \ \ i = \overline{1,n}.$$

Чтобы решить подзадачу, положим

$$x_i = \beta_i \ \ \forall \ i = \overline{1,n}.$$

Если $g(\beta_1, \beta_2, ..., \beta_n) \ge 0$, то решение существует, в противном случае решения нет.

Для данного случая можно легко найти верхние и нижние границы для оптимального t, так как можно просто взять в левой части системы $x_i = 0, \ i = \overline{1,n}$, для нижней границы и $x_i = c, \ \ i = \overline{1,n}$, для верхней границы.

Замечание 1.6 Если некоторые ℓ_{ij} равны нулю, ситуация сильно не изменитттся. В этом случае соответствующее θ равно плюс бесконечности, и мы рассматриваем соответствующее ограничение как выполненное. На следующем шаге размерность подзадачи убывает в любом случае, и процедура, указанная выше, по существу не меняется.

Глава 2

Минимизация возрастающих звездных и липшицевых функций

Рассмотрим широкий класс возрастающих невыпуклых функций, множества уровня которых звездны по отношению к бесконечности. Покажем, что эти функции (мы их называем ВЗ функциями) абстрактно выпуклы по отношению к множеству функций типа минимума, и используем этот факт для их минимизации.

Задачи математического программирования с такими целевыми функциями имеют различные применения на практике, особенно в математической экономике, что делает их очень важными. ВЗ функции часто используются в качестве производственных функций.

В дальнейшем нам понадобятся следующие определения.

Вещественнозначная функция f, определенная на конусе K n-мерного пространства $\mathbb{R}^n$, называется 0-звездной (или звездной относительно нуля), если

$$f(\lambda x) \leq \lambda f(x) \quad \forall x \in K \quad \forall \lambda \leq 1.$$

Этот термин был введен в несколько другой ситуации Палляшке и Ролевичем ([70]). Если f – 0-звездная, то ее лебеговы множества $\{x : f(x) \leq c\}$ звездны относительно нуля для всех $c > 0$. Напомним, что множество A звездно относительно нуля, если

$$x \in A,\ \lambda \in [0, 1] \implies \lambda x \in A.$$

Конечно, существуют не 0-звездные функции со звездными относительно нуля лебеговыми множествами. Полное описание таких функций дано в [87] (для случая, когда конус K совпадает со всем пространством).

Дополнение к звездному относительно нуля – так называемое звездное относительно бесконечности множество (см. [86]), то есть множество B со свойством

$$x \in B,\ \lambda \geq 1 \implies \lambda x \in B.$$

Выпуклые звездные по отношению к бесконечности множества изучались рядом авторов (см., например, работы [72, 28, 12] и ссылки в них).

Имея в виду эту терминологию, мы введем следующие определения: вещественнозначная функция f, определенная на конусе K, называется звездной по отношению к $+\infty$, если

$$f(\lambda x) \geq \lambda f(x) \quad \text{для всех} \quad x \in K,\ \lambda \in [0,1]. \tag{2.1}$$

Заметим, что $f(0) \geq 0$ для такой функции. Легко проверить, что f – звездная по отношению к бесконечности тогда и только тогда, когда

$$f(\mu x) \leq \mu f(x) \quad \text{для всех} \quad x \in K,\ \mu \geq 1. \tag{2.2}$$

Действительно, пусть f – звездная относительно $+\infty$. Пусть $\mu \geq 1$, $x \in K$ и $\mu x = x'$. Поскольку $x = \lambda x'$ с $\lambda = 1/\mu \leq 1$, имеем $f(x) \geq \lambda f(x')$. Мы доказали, что из (2.1) следует (2.2). То же рассуждение показывает, что из (2.2) следует (2.1).

Ниже мы будем рассматривать возрастающие звездные относительно $+\infty$ функции (сокращенно ВЗ функции), определенные на конусе $\mathbb{R}^n_+$ всех n-мерных векторов с неотрицательными координатами. Иногда мы будем рассматривать сужение таких функций на конус $\mathbb{R}^n_{++}$ всех векторов с положительными координатами.

Поскольку ВЗ функция f возрастает и $f(0) \geq 0$, имеем $f(x) \geq 0$ для всех $x \in \mathbb{R}^n_+$. Проверим, что ВЗ функция f непрерывна на конусе $\mathbb{R}^n_{++}$. Действительно, пусть $x \gg 0$ и $x_k \to x$. Пусть $\varepsilon > 0$. Тогда для достаточно большого k имеем $(1-\varepsilon)x \leq x_k \leq (1+\varepsilon)x$. Использовав свойства функции f и (2.1), (2.2), получим

$$(1-\varepsilon)f(x) \leq f((1-\varepsilon)x) \leq f(x_k) \leq f((1+\varepsilon)x) \leq (1+\varepsilon)f(x).$$

Непрерывность доказана.

Множество $\Phi(\text{ВЗ})$ ВЗ функций имеет ряд полезных свойств:

1) $\Phi(\text{ВЗ})$ – выпуклый конус: если f, g – ВЗ функции и λ, μ – положительные числа, то $\lambda f + \mu g$ – ВЗ функция;

2) $\Phi(\text{ВЗ})$ – условно полная решетка; точнее, для любого семейства $(f_\alpha)_{\alpha\in A}$ ВЗ функций поточечный инфимум $\inf_{\alpha\in A} f_\alpha$ также есть ВЗ функция. Если семейство $(f_\alpha)_{\alpha\in A}$ ограничено сверху, то есть существует ВЗ функция f, такая, что $f_\alpha \le f$ для всех $\alpha \in A$, то поточечный супремум $\sup_{\alpha\in A} f_\alpha$ также является ВЗ функцией.

3) Если семейство f_k ВЗ функций поточечно сходится к конечной функции f, то f также есть ВЗ функция.

Легко проверить, что возрастающая вогнутая по лучам (ВВГЛ) функция f, такая, что $f(0) \ge 0$, является ВЗ функцией. Действительно, имеем для каждого $x \in \mathbb{R}^n_+$ и $\lambda \in [0,1]$:

$$f(\lambda x) = f_x(\lambda) = f_x(\lambda + (1-\lambda)0) \ge \lambda f_x(1) + (1-\lambda)f(0) \ge \lambda f(x).$$

Дадим некоторые примеры ВЗ функций.

Пример 3.1 Возрастающая положительно однородная степени $0 < \delta \le 1$ функция f – ВЗ функция. Действительно, она ВВГЛ (если $\delta \ge 1$, то f – ВВЛ функция). В частности, функция Кобба-Дугласа

$$f(x) = Cx_1^{\alpha_1} \cdots x_n^{\alpha_n} \quad \text{с} \quad \alpha_i \ge 0, \sum_i \alpha_i \le 1 \tag{2.3}$$

есть ВЗ функция. Заметим, что если $\sum_i \alpha_i \ge 1$, то функция Кобба-Дугласа ВВЛ. Функция $f_p(x) = (\sum_i x_i^p)^{\frac{1}{p}}$ $(x \in \mathbb{R}^n_+)$ с $p > 0$ – также ВЗ функция. Заметим, что эта функция выпукла при $p \ge 1$ и вогнута при $p \le 1$.

Пример 3.2 Вогнутая возрастающая функция f, определенная на $\mathbb{R}^n_+$, с $f(0) \ge 0$ – ВЗ функция. Действительно, f – ВВГЛ функция. В частности, сумма двух функций вида (2.3) является вогнутой возрастающей.

Пример 3.3 Поточечный супремум f семейства ВВГЛ функций $(f_\alpha)_{\alpha\in A}$ есть ВЗ функция при условии, что $f(x) < +\infty$ для всех x. Заметим, что эта функция не обязательно ВВГЛ.

Пример 3.4 Пусть $f_1, \ldots, f_k$ – возрастающие положительно однородные степени $0 < \delta_k \le 1$ функции. Тогда их сумма, максимум и минимум – ВЗ функции.

2.1 ВЗ функции и ВПО функции

Существует тесная связь между ВЗ функциями и так называемыми ВПО функциями (ВПО означает *возрастающая* и *положительно однородная степени один*). Чтобы установить эту связь, нам понадобятся следующие определения.

Пусть f – функция, определенная на конусе $\mathbb{R}^n_+$. Функция $\hat{f}$, определенная на конусе

$$\mathbb{R}^{n+1}_* = \{(x,\lambda) : x \in \mathbb{R}^n_+,\ \lambda > 0\} \cup \{0,0\} \tag{2.4}$$

формулой

$$\hat{f}(x,\lambda) = \lambda f(\frac{x}{\lambda}), \quad (x \in K,\ \lambda > 0), \quad \hat{f}(0,0) = 0, \tag{2.5}$$

называется *положительно однородным расширением* функции f.

Следующий результат был доказан в [1]. Мы даем его доказательство для полноты изложения.

Теорема 2.1 *Функция f, определенная на $\mathbb{R}^n_+$, – ВЗ функция тогда и только тогда, когда ее положительно однородное расширение $\hat{f}(x,\lambda)$ возрастает по обеим переменным x, λ.*

Доказательство. Пусть f – ВЗ функция. Поскольку $f(x) \geq 0$ для всех $x \in \mathbb{R}^n_+$, имеем $\hat{f}(x,\lambda) = \lambda f(x/\lambda) \geq 0$. Рассмотрим теперь две точки (x_1, λ_1) и (x_2, λ_2) с $x_1, x_2 \in \mathbb{R}^n_+$, $x_1 \geq x_2$ и $\lambda_1 \geq \lambda_2 > 0$. Имеем

$$\hat{f}(x_1,\lambda_1) = \lambda_1 f(\frac{x_1}{\lambda_1}) \geq \lambda_1 f(\frac{x_2}{\lambda_1}) = \lambda_1 f(\frac{\lambda_2}{\lambda_1}\frac{x_2}{\lambda_2}) \geq \lambda_2 f(\frac{x_2}{\lambda_2}) = \hat{f}(x_2,\lambda_2).$$

Поэтому $\hat{f}$ – возрастающая функция. Допустим теперь, что $\hat{f}$ возрастает. Тогда $\hat{f}(x,\lambda) \geq \hat{f}(0,0) = 0$, в частности, $f(0) = \hat{f}(0,1) \geq 0$. Если $x_1 \geq x_2$, то $f(x_1) = \hat{f}(x_1, 1) \geq \hat{f}(x_2, 1) = f(x_2)$. Таким образом, f возрастает. Пусть $\lambda \in (0,1)$. Тогда $(\lambda x, \lambda) \leq (\lambda x, 1)$, следовательно,

$$\lambda f(x) = \hat{f}(\lambda x, \lambda) \leq \hat{f}(\lambda x, 1) = f(\lambda x).$$

Если $\lambda = 0$, то $0 = \lambda f(x) \leq f(0) = f(\lambda x)$. Отсюда следует, что f – ВЗ функция. $\square$

Замечание 2.1.1 То же рассуждение показывает, что верно следующее утверждение: функция f, определенная на $\mathbb{R}^n_+$, – 0-звездная и убывающая тогда и только тогда, когда ее положительно однородное расширение $\hat{f}$ убывает по обеим переменным.

Теорема 2.1 позволяет изучать ВЗ функции с помощью более простых ВПО функций.

Рассмотрим возрастающие положительно однородные степени один функции (ВПО функции). Простейшие примеры ВПО функций дают функции типа минимума и типа максимума:

$$l_-(x) = \min_{i \in \mathcal{T}(l)} l_i x_i := \langle l, x \rangle^- \tag{2.6}$$

и

$$l_+(x) = \max_{i \in I} l_i x_i := \langle l, x \rangle^+. \tag{2.7}$$

Здесь $l = (l_1, \ldots, l_n) \in \mathbb{R}^n_+$, $I = \{1, 2, \ldots, n\}$ и $\mathcal{T}(l) = \{i \in I : l_i > 0\}$. Предполагается, что минимум по пустому множеству равен нулю.

Мы обозначаем в данном параграфе $\min_{i \in \mathcal{T}(l)} l_i x_i$ через $\langle l, x \rangle^-$ и $\max_{i \in I} l_i x_i$ через $\langle l, x \rangle^+$.

Далее мы покажем, что каждая ВПО функция, определенная на $\mathbb{R}^n_+$, абстрактно выпукла по отношению к множеству $\mathcal{L}^-$ функций типа минимума вида (2.6) и что сужение каждой ВПО функции на конус $\mathbb{R}^n_{++}$ абстрактно вогнуто по отношению к множеству $\mathcal{L}^+$ функций типа максимума вида (2.7). Для этого рассмотрим субдифференциал $\partial p^-(y)$ ВПО функции p по отношению к $\mathcal{L}^-$ и супердифференциал $\partial p^+(y)$ функции p по отношению к $\mathcal{L}^+$. Здесь

$$\partial^- p(y) = \{l \in \mathbb{R}^n_+ : \langle l, x \rangle^- \leq p(x) \ \forall x \in \mathbb{R}^n_+, \ \langle l, y \rangle^- = p(y)\}.$$

$$\partial^+ p(y) = \{l \in \mathbb{R}^n_+ : \langle l, x \rangle^+ \geq p(x) \ \forall x \in \mathbb{R}^n_+, \ \langle l, y \rangle^+ = p(y)\}.$$

Для $l \in \mathbb{R}^n_+$ определим вектор $l^{-1} = \dfrac{1}{l}$ формулой

$$\frac{1}{l} = \begin{cases} \dfrac{1}{l_i}, & i \in \mathcal{T}(l), \\ 0, & i \notin \mathcal{T}(l). \end{cases}$$

Очевидно, $\langle l, \frac{1}{l} \rangle^- = \langle l, \frac{1}{l} \rangle^+ = 1$ для $l \neq 0$.

Предложение 2.1 *Пусть p – ВПО функция. Тогда множество $\partial p^-(y)$ непусто для всех $y \geq 0$; если $p(y) > 0$, то*

$$\partial^- p(y) = \{l : \mathcal{T}(l) \subset \mathcal{T}(y), l \geq \frac{p(y)}{y_l}, p(\frac{1}{l}) = 1\}, \tag{2.8}$$

где координата $(y_l)_i$ вектора y_l равна координате y_i вектора y, если $i \in \mathcal{T}(l)$, и равна нулю, если $i \notin \mathcal{T}(l)$.

Доказательство. Если $p(y) = 0$, то $0 \in \partial^- p(y)$ и поэтому $\partial^- p(y)$ непусто. Рассмотрим точку y с $p(y) > 0$. Пусть

$$A = \{l : \mathcal{T}(l) \subset \mathcal{T}(y), l \geq \frac{p(y)}{y_l}, p(\frac{1}{l}) \geq 1\}.$$

Легко проверить, что A совпадает с множеством в правой части (2.8). Действительно, пусть $l \in A$. Соотношения $\mathcal{T}(l) \subset \mathcal{T}(y)$ и $l \geq p(y)/y_l$ показывают, что $y_l \geq p(y)/l$. Поскольку p – ВПО функция, $y \geq y_l$ и $p(1/l) \geq 1$, имеем

$$p(y) \geq p(y_l) \geq p(\frac{p(y)}{l}) = p(y)p(\frac{1}{l}) \geq p(y).$$

Тогда $p(1/l) = 1$, и требуемое равенство доказано.

Итак, мы должны доказать, что $\partial^- p(y) = A$. Пусть $l \in A$. Вначале проверим, что $\langle l, x\rangle^- \leq p(x)$ для всех $x \in \mathbb{R}^n_+$. Допустим, напротив, что существует $z \in \mathbb{R}^n_+$, такое, что $\langle l, z\rangle^- = \min_{i \in \mathcal{T}(l)} l_i z_i > p(z)$. Тогда $l_i z_i > p(z)$ для всех $i \in \mathcal{T}(\ell)$, следовательно, $\mathcal{T}(l) \subset \mathcal{T}(z)$ и существует $\varepsilon > 0$, такое, что

$$z_i \geq \frac{p(z)}{l_i} + \frac{\varepsilon}{l_i} \quad \text{для всех} \quad i \in \mathcal{T}(l).$$

Тогда $z \geq (p(z) + \varepsilon)(1/l)$. Так как p возрастает, имеем

$$p(z) \geq p((p(z) + \varepsilon)\frac{1}{l}) = (p(z) + \varepsilon)p(\frac{1}{l}) \geq p(z) + \varepsilon.$$

Получили противоречие, следовательно, $\langle l, x\rangle^- \leq p(x)$ для всех $x \in \mathbb{R}^n_+$. Рассмотрим теперь вектор y. Так как $\mathcal{T}(l) \subset \mathcal{T}(y)$ и $l \geq p(y)/y_l$, получаем, что $l_i y_i \geq p(y)$ для всех $i \in \mathcal{T}(l)$. Таким образом, $\langle l, y\rangle^- \geq p(y)$. Неравенство $\langle l, y\rangle^- \leq p(y)$ было уже доказано. Поэтому $\langle l, y\rangle^- = p(y)$ и $\partial^- p(y) \supset A$.

Проверим, что множество A непусто. Действительно, пусть $l = p(y)/y$. Тогда $\mathcal{T}(l) = \mathcal{T}(y)$. Так как $p(y) \neq 0$, вектор $1/l = y/p(y)$ ненулевой и

$p(1/l) = 1$. Следовательно, $l \in A$. Мы доказали, что субдифференциал $\partial^- p(y)$ непуст.

Покажем, что $\partial^- p(y) \subset A$, если $p(y) > 0$. Пусть $l \in \partial^- p(y)$, то есть $\langle l, x\rangle^- \leq p(x)$ для всех x и $\min_{i \in \mathcal{T}(l)} l_i y_i = \langle l, y\rangle^- = p(y)$. Так как $p(y) > 0$, то $y_i > 0$ для $i \in \mathcal{T}(l)$, поэтому $\mathcal{T}(l) \subset \mathcal{T}(y)$. Неравенство $l_i y_i \geq p(y)$ для всех $i \in \mathcal{T}(l)$ показывает, что $l \geq p(y)/y_l$. Мы также имеем

$$p(\frac{1}{l}) \geq \langle l, \frac{1}{l}\rangle^- = 1.$$

Отсюда $l \in A$. □

Следствие 2.1.1 *Если $p(y) > 0$, то $l = \dfrac{p(y)}{y} \in \partial^- p(y)$.*

Следствие 2.1.2 *ВПО функция абстрактно выпукла по отношению к множеству $\mathcal{L}^-$ функций (2.6).*

Замечание 2.1.2 Рассмотрим ВПО функцию p от $n+1$ переменной, определенную на конусе $\mathrm{IR}^{n+1}_* = \{(x, \lambda) : x \in \mathrm{IR}^n_+, \lambda > 0\} \cup \{0, 0\}$. Очевидно, Предложение 2.1 и Следствие 2.1.2 выполняются также для этой функции.

Предложение 2.2 *Если ВПО функция p строго возрастает в точке y, то*

$$(l \in \partial^- p(y), \mathcal{T}(l) = \mathcal{T}(y)) \implies l = \frac{p(y)}{y}.$$

Доказательство. Из Предложения 2.1 следует, что необходимо проверить, что из соотношений

$$\mathcal{T}(l) = \mathcal{T}(y), \quad l \geq \frac{p(y)}{y_l}, \quad p(\frac{1}{l}) = 1 \tag{2.9}$$

вытекает $l = p(y)/y$. Возьмем такое l, что выполняется (2.9). Очевидно, $y \geq p(y)/l$. Допустим, что $y \neq p(y)/l$. Так как p строго возрастает в точке y, имеем

$$p(y) > p(\frac{p(y)}{l}) = p(y)p(\frac{1}{l}) \quad \text{и} \quad p(y) > 0.$$

Таким образом, $p(1/l) < 1$, получили противоречие. □

Приведем несколько примеров.

Пример 3.5 Пусть $p(x) = \max_{i\in I} x_i$ и $\mathbf{1} = (1, 1, \dots, 1)$. Тогда

$$\partial^- p(\mathbf{1}) = \{l : l \geq \mathbf{1}_l,\ \max_{i\in\mathcal{T}(l)} \frac{1}{l_i} = 1\} = \{l : l \geq \mathbf{1}_l,\ \min_{i\in\mathcal{T}(l)} l_i = 1\}.$$

Пример 3.6 Пусть $p(x) = \min_{i\in I} x_i$. Имеем

$$\partial^- p(\mathbf{1}) = \{l : l \geq \mathbf{1}_l,\ \min_{i\in\mathcal{T}(l)} \frac{1}{l_i} = 1\} = \{l : l \geq \mathbf{1}_l,\ \max_{i\in\mathcal{T}(l)} l_i = 1\}.$$

Таким образом, $\partial^- p(\mathbf{1}) = \{\mathbf{1}^K : K \subset I\}$, где $\mathbf{1}^K$ есть такой вектор, что его i-я координата равна единице, если $i \in K$, и равна нулю, если $i \notin K$. Иными словами, $\partial^- p(\mathbf{1})$ совпадает с множеством всех ненулевых вершин n-мерного куба $\{x : 0 \leq x \leq 1\}$.

Пример 3.7 Пусть $p(x) = (x_1 \cdot x_2 \cdots x_n)^{\frac{1}{n}}$. Легко проверить, что ВПО функция p строго возрастает в точке $\mathbf{1}$ и неравенство $\langle l, x\rangle^- \leq p(x)$ для всех x влечет за собой $\mathcal{T}(l) = I$. Из Предложения 2.2 следует, что $\partial^- p(\mathbf{1}) = \{\mathbf{1}\}$.

Рассмотрим теперь супердифференциал

$$\partial^+ p(y) = \{l \in \mathbb{R}^n_+ : \langle l, x\rangle^+ \geq p(x) \ \ \forall x \in \mathbb{R}^n_+,\ \langle l, y\rangle^+ = p(y)\}$$

ВПО функции p в точке $y \in \mathbb{R}^n_+$ по отношению к множеству $\mathcal{L}^+$. Приведем пример, который показывает, что супердифференциал $\partial^+ p(y)$ может быть пустым для некоторых точек $y \in \mathbb{R}^n_+$.

Пример 3.8 Пусть p – такая ВПО функция, что

1) существует ненулевой элемент y со свойством $p(y) = 0$ (в этом случае y – граничная точка конуса $\mathbb{R}^n_+$);

2) если $x^k = x_{(1)} + x^k_{(2)}$, где $x_{(1)}$ – фиксированный вектор из

$$\mathbb{R}^n_{++},\ \mathcal{T}(x^k_{(2)}) \subset \mathcal{T}(y)$$

и $\|x^k_{(2)}\| \to +\infty$, когда $k \to \infty$, то $p(x^k) \to +\infty$.

Предположим, что $l \in \partial^+ p(y)$. Тогда $\langle l, y\rangle^+ = \max_{i\in I} l_i y_i = p(y) = 0$, следовательно, $l_i = 0$ для $i \in \mathcal{T}(y)$. Значит, $\mathcal{T}(l) \subset (I \setminus \mathcal{T}(y))$. Пусть $x \gg 0$ – такой вектор, что $l_i x_i = 1$ для $i \in \mathcal{T}(l)$ и x_i для $i \notin \mathcal{T}(l)$ достаточно велико, чтобы $p(x) > 1$ (из 2) следует, что такой вектор существует). Так как $\langle l, x\rangle^+ = 1$, неравенство $\langle l, x\rangle^+ \geq p(x)$ не выполняется. Поэтому $\partial^+ p(y)$ пусто.

Рассмотрим, в частности, функцию $p(x) = (x_1 \cdot x_2 \cdots x_n)^{\frac{1}{n}}$. Из сказанного выше следует, что ее супердифференциал $\partial^+ p(y)$ пуст для каждой граничной точки y конуса $\mathbb{R}^n_+$. Тем не менее, если $y \gg 0$, то супердифференциал $\partial^+ p(y)$ непуст и содержит внутренние точки конуса $\mathbb{R}^n_+$.

Предложение 2.3 *Пусть p – ВПО функция и $y \gg 0$. Тогда $\partial^+ p(y) \cap \mathbb{R}^n_{++}$ непусто и*

$$\partial^+ p(y) \cap \mathbb{R}^n_{++} = \{l \gg 0 : l \leq \frac{p(y)}{y},\ p(\frac{1}{l}) = 1\}. \tag{2.10}$$

Доказательство. Мы будем использовать те же аргументы, что в доказательстве Предложения 2.1. Вначале покажем, что множество в правой части (2.10) совпадает со следующим множеством B:

$$B = \{l \gg 0 : l \leq \frac{p(y)}{y},\ p(\frac{1}{l}) \leq 1\}.$$

Действительно, если $l \in B$, то $y \leq p(y)/l$, следовательно,

$$p(y) \leq p(\frac{p(y)}{l}) = p(y)p(\frac{1}{l}).$$

Поскольку $p(y) > 0$, имеем $p(1/l) \geq 1$. С другой стороны, $p(1/l) \leq 1$. Поэтому требуемое равенство доказано. Пусть $l \in B$. Поскольку $l \leq p(y)/y$, имеем

$$\max_{i \in \mathcal{T}(y)} l_i y_i = \max_{i \in I} l_i y_i = \langle l, y\rangle^+ \leq p(y). \tag{2.11}$$

Проверим теперь, что $\langle l, x\rangle^+ \geq p(x)$ для всех $x \in \mathbb{R}^n_+$. Допустим, напротив, что существует такое z, что $\langle l, z\rangle^+ < p(z)$. Тогда $l_i z_i < p(z)$ для всех $i \in I$, поэтому $z < p(z)/l$. Существует $\varepsilon > 0$, такое, что $(1+\varepsilon)z < p(z)/l$. Получаем

$$(1+\varepsilon)p(z) = p((1+\varepsilon)z) \leq p(\frac{p(z)}{l}) = p(z)p(\frac{1}{l}).$$

Так как $p(z) > 0$, то $p(\frac{1}{l}) > 1+\varepsilon$. Получили противоречие, которое показывает, что $\langle l, x\rangle^+ \geq p(x)$ для всех x. В частности, $\langle l, y\rangle^+ \geq p(y)$. Соединяя это неравенство с (2.11), заключаем, что $\langle l, y\rangle^+ = p(y)$. Тогда $l \in \partial^+ p(y)$. Очевидно, элемент $l = p(y)/y$ принадлежит множеству B. Таким образом, супердифференциал $\partial^+ p(y)$ непуст.

Предположим теперь, что элемент $l \gg 0$ принадлежит $\partial^+ p(y)$. Тогда $1 = \langle l, 1/l\rangle^+ \geq p(1/l)$. Из равенства $\langle l, y\rangle^+ = p(y)$ следует, что $l \leq p(y)/y$. Следовательно, $l \in B$. $\square$

Следствие 2.1.3 *Сужение ВПО функции на конус* $\mathbb{R}^n_{++}$ *абстрактно вогнуто по отношению к классу* $\mathcal{L}^+$.

2.2 Субдифференциалы ВЗ функций

В этой части мы изучаем ВЗ функции, используя предыдущие результаты. Пусть f – ВЗ функция. Рассмотрим ее положительно однородное расширение $\hat{f}$ на конусе $\mathbb{R}^{n+1}_*$ (см. (2.4) для определения этого конуса).

Из Теоремы 2.1 следует, что $\hat{f}$ – ВПО функция на конусе $\mathbb{R}^{n+1}_*$, так что (см. Следствие 2.1.2 и Замечание 2.1.2) эта функция абстрактно выпукла по отношению к множеству $\hat{\mathcal{L}}^-$ всех функций $\hat{l}$, определенных на $\mathbb{R}^{n+1}_+$ формулой (2.6). Мы представляем вектор $\hat{l} \in \mathbb{R}^{n+1}_+$ в виде $\hat{l} = (l, c)$, где $l \in \mathbb{R}^n_+$, $c \geq 0$. Пусть $x \in \mathbb{R}^n_+$ и $\hat{x} = (x, 1)$. Имеем

$$f(x) = \hat{f}(\hat{x}) =$$

$$= \sup\{\min_{i \in \mathcal{T}(\hat{l})} \hat{l}_i \hat{x}_i : \hat{l} \in \hat{\mathcal{L}}^-, \hat{l} \leq \hat{f}\} =$$

$$= \sup\{\min(\min_{i \in \mathcal{T}(l)} l_i x_i, c) : \hat{l} = (l, c), l \in \mathcal{L}^-, c \geq 0, \hat{l} \leq f\}. \tag{2.12}$$

Из (2.12) следует, что f абстрактно выпукла по отношению ко множеству $\mathcal{H}^-$ функций $h : \mathbb{R}^n_+ \to \mathbb{R}_+$ вида $h(x) = \min(\langle l, x\rangle^-, c)$ с $l \in \mathbb{R}^n_+$ и $c \geq 0$.

Пусть f – ВЗ функция и $y \in \mathbb{R}^n_+$. Обозначим через $\bar{\partial}^- f(y)$ субдифференциал f в точке y по отношению ко множеству $\mathcal{H}^-$. Если $f(y) = 0$, то этот субдифференциал непуст, он содержит функцию $h = 0$, которая принадлежит $\mathcal{H}$. Пусть $f(y) > 0$. Тогда $\hat{f}(\hat{y}) = f(y) > 0$ и поэтому (см. Следствие 2.1.1)

$$\frac{\hat{f}(\hat{y})}{\hat{y}} = \frac{f(y)}{(y, 1)} \in \partial^- \hat{f}(y).$$

Из этого включения следует, что

$$f(y) = \hat{f}(\hat{y}) = \min(\min_{i \mathcal{T}(y)} \frac{f(y)}{y_i}, \frac{f(y)}{1}) = \min(\langle \frac{f(y)}{y}, y\rangle^-, f(y)) \tag{2.13}$$

и для каждого $x \in \mathbb{R}^n_+$

$$\min(\langle \frac{f(y)}{y}, x\rangle^-, f(y)) \leq \hat{f}(x, 1) = f(x). \tag{2.14}$$

Пусть

$$h(x) = \min(\langle \frac{f(y)}{y}, x\rangle^-, f(y)).$$

Очевидно, $h \in \mathcal{H}^-$. Формулы (2.13) и (2.14) показывают, что f – элемент субдифференциала $\bar{\partial}^- f(x)$ функции f в точке x по отношению к множеству $\mathcal{H}^-$. Отсюда выводим, что $\bar{\partial}^- f(x)$ непусто.

Дадим описание субдифференциала $\bar{\partial}^- f(x)$, считая, что $f(x) > 0$. Применив Предложение 2.1 и Замечание 2.1.2, заключаем, что

$$\partial^- \hat{f}(\hat{x}) = \{\hat{l} = (l, \mu) \in \mathbb{R}^{n+1}_+ : \mathcal{T}(\hat{l}) \subset \mathcal{T}(\hat{x}),\, \hat{l} \geq \frac{\hat{f}(\hat{x})}{\hat{x}_{\hat{l}}},\, \hat{f}(\frac{1}{\hat{l}}) = 1\}. \tag{2.15}$$

Поскольку

$$\hat{f}(\frac{1}{(l, \mu)}) = \frac{1}{\mu}\hat{f}(\frac{\mu}{l}, 1) = \frac{1}{\mu} f(\frac{\mu}{l}),$$

мы можем представить (2.15) в следующем виде:

$$\partial^- \hat{f}(\hat{x}) = \{(l, \mu) \in \mathbb{R}^n_+ \times \mathbb{R}_+ :\, \mathcal{T}(l) \subset \mathcal{T}(x),$$

$$l \geq \frac{f(x)}{x_l},\, \mu \geq f(x),\, f(\frac{\mu}{l}) = \mu\}.$$

Те же аргументы показывают, что сужение ВЗ функции на конус $\mathbb{R}^n_{++}$ абстрактно вогнуто по отношению к множеству $\mathcal{H}^+$ всех функций h типа максимума $h(x) = \max(\langle l, x\rangle^+, c)$ с $l \in \mathbb{R}^n_+$ и $c \geq 0$. Мы можем легко описать супердифференциал функции f в точке $y \gg 0$ по отношению к множеству $\mathcal{H}^+$. В частности, функция

$$h(x) = \max(\langle \frac{f(y)}{y}, x\rangle^+, f(y))$$

принадлежит супердифференциалу.

2.3 Алгоритм

Рассмотрим следующую задачу оптимизации:

$$f(x) \longrightarrow \min, \tag{2.16}$$

$$x \in X,$$

где f – ВЗ функция (в частности, это может быть ВПО функция), X – замкнутое выпуклое множество из $\mathbb{R}^n_+$. Мы предлагаем для решения этой задачи вариант обобщенного метода секущих плоскостей (метода Ф-пучков из [70]). Заметим, что для ВПО функции можно применить метод секущих углов, рассмотренный выше. Алгоритм может рассматриваться как обобщение метода секущих плоскостей выпуклой оптимизации [60] для задач с ВЗ целевой функцией.

Алгоритм минимизации ВЗ функций

Шаг 0. Полагаем $k := 0$. Выбираем произвольную начальную точку $x_o \in X$.

Шаг 1. Вычисляем элемент субдифференциала $(h_k, c_k) \in \partial_L f(x_k)$. Обозначим $f_k(x) = \min(\langle h_k, x\rangle^-, c_k)$.

Шаг 2. Решаем следующую подзадачу:

$$\max_{0 \le i \le k} f_i(x) \longrightarrow \min, \tag{2.17}$$

$$x \in X.$$

Шаг 3. Пусть y^* – решение задачи (2.17). Полагаем $k := k+1, x_k = y^*$ и возвращаемся к Шагу 1.

Сходимость алгоритма устанавливает

Теорема 2.2 *Пусть X – компактное множество и производные по направлению $f'_k(x)$ равномерно ограничены на множестве X:*

$$\|f'_k(x)\| = \max_{\|u\| \le 1} |f'_k(x,u)| \le R < +\infty \quad \textit{для всех} \quad x \in X, k = 0, 1, \ldots \tag{2.18}$$

Тогда каждая предельная точка x последовательности (x_k) – глобальный минимум задачи (2.16).

Доказательство. Очевидно, $f_k(x)$ – вогнутые функции. Так как производные по направлению ограничены, условия Теоремы 1.1 выполнены. Применив эту теорему, получим результат о сходимости. □

Если мы имеем задачу максимизации

$$f(x) \longrightarrow \max, \tag{2.19}$$

$$x \in X,$$

где f – ВЗ функция и $X \subset \mathbb{R}^n_{++}$, ситуация сходная. Вместо элементов субдифференциала мы должны брать элементы супердифференциала и вместо минимизации максимума $f_k(x)$ мы должны максимизировать минимум f_k. Выполняется тот же результат о сходимости (доказательство аналогично).

В общем случае трудно применить метод Ф-пучков для минимизации абстрактно выпуклой функции, так как нет явной формулы для субдифференциала. Главное достоинство ВЗ функций состоит в том, что элемент абстрактного субдифференциала можно вычислить. Это позволяет построить численно реализуемый алгоритм для их минимизации.

2.4 Решение подзадачи для ВЗ функций

Главная часть метода – решение подзадачи на Шаге 2. Представим подзадачу в следующем виде:

$$\begin{aligned} t &\to \min, \\ \min(\langle h_i, x\rangle^-, c_i) \le t \quad &\forall i = \overline{0, k}, \\ x &\in X. \end{aligned} \tag{2.20}$$

Так как $c_i \ge 0$ для всех i вследствие свойств ВЗ функций, оптимальное значение t лежит в интервале $[0, \max_{i=\overline{0,k}} \min(\langle h_i, x_0\rangle^-, c_i)]$. Применив дихотомию по отношению к t, получим последовательность систем:

$$\min(\langle h_i, x\rangle^-, c_i) \le \tilde{t} \quad \forall i = \overline{0, k},$$

$$s.t. \quad x \in X,$$

где $\tilde{t}$ – некоторое фиксированное число в интервале

$$[0, \max_{i=\overline{0,k}} \min(\langle h_i, x_0\rangle^-, c_i)].$$

Тогда, если для некоторого j мы имеем $c_j \le \tilde{t}$, соответствующее неравенство избыточно и может быть удалено. Напротив, если $c_j > \tilde{t}$, то соответствующая константа c_j избыточна и j-е ограничение может быть переписано без этой константы. Таким образом, мы получаем следующую систему для каждого $\tilde{t}$:

$$\langle h_j, x\rangle^- \le \tilde{t} \quad \forall j \in J,$$
$$s.t. \quad x \in X,$$

где $J = \{j \in \{0,1,\ldots,k\} : c_j > \tilde{t}\}$. Эта система была рассмотрена выше, где был предложен алгоритм ее решения, похожий на динамическое программирование. Применив этот алгоритм, мы можем получить решение подзадачи (2.20).

Другая возможность – свести систему (2.20) к задаче частично целочисленного булева линейного программирования. Это хорошо известная техника, основанная на введении большого положительного параметра M. Тогда каждое ограничение $\langle h_j, x\rangle^- \le \tilde{t}$ может быть представлено как $n+1$ линейное ограничение вида:

$$\begin{aligned} h_{j1}x_1 - My_{j1} &\le \tilde{t}, \\ h_{j2}x_2 - My_{j2} &\le \tilde{t}, \\ &\ldots\ldots\ldots \\ h_{jn}x_n - My_{jn} &\le \tilde{t}, \\ \sum_{i=1}^{n} y_{ji} &\le n-1, \end{aligned}$$

где $y_{ji} \in \{0,1\}$ для всех i, j. Если M достаточно велико, эта задача эквивалентна подзадаче (2.20). Данная техника увеличивает размерность задачи на n переменных со значениями $\{0,1\}$ на каждой итерации. Однако она позволяет применить существующие пакеты программного обеспечения для частично целочисленного программирования, чтобы решить подзадачу.

2.5 Липшицево программирование через ВВЛ функции

В последние годы большое внимание привлечено к задачам глобальной минимизации невыпуклых функций. Один из важнейших классов таких функций – липшицевы функции. Был построен ряд алгоритмов их минимизации, в основном это методы ветвей и границ или методы случайного поиска (см. [74] для подробного обсуждения, а также [53]).

В данной главе показано, что во многих случаях задача оптимизации произвольной липшицевой целевой функции может быть сведена к минимизации ВВЛ функции. Предложено обобщение метода секущих углов для задачи липшицева программирования.

2.6 Сведение задачи липшицева программирования к минимизации ВВЛ функции

Пусть $S = \{y \in \mathbb{R}^n_+ : \sum_{i=1}^n y_i = 1\}$ – единичный симплекс и $y \in S$. Мы предполагаем, что в пространстве $\mathbb{R}^n$ введена норма $\|\cdot\|_1 = \sum_{i=1}^n |x_i|$.

Пусть f – липшицева функция, определенная на S. Обозначим через $f'_-(y,u)$ нижнюю производную Дини функции f в точке y по направлению u:

$$f'_-(y,u) = \lim\inf_{\alpha\to+0} \frac{1}{\alpha}(f(y+\alpha u) - f(y)).$$

Рассмотрим функцию g, определенную на конусе $\mathbb{R}^n_+$ как

$$g(x) = \begin{cases} f\left(\dfrac{x}{\sum_{i=1}^n x_i}\right)(\sum_{i=1}^n x_i)^p, & x \neq 0, \\ 0 & x = 0. \end{cases} \tag{2.21}$$

Укажем некоторые свойства функции g:

1) $g(y) = f(y) \quad \forall y \in S$;

2) $g(\alpha x) = \alpha^p g(x) \quad \forall x \in \mathbb{R}^n_+,\ \alpha > 0$.

Пусть $p \geq 1$. Из 2) следует, что g выпукла по лучам. Таким образом, g – выпуклое по лучам расширение функции f.

Теорема 2.3 [80] *Пусть f – липшицева положительная функция, определенная на симплексе S, и*

$$L = \inf_{\substack{x\neq y \\ x,y\in S}} \frac{|f(x)-f(y)|}{\|x-y\|_1}$$

– константа Липшица функции f. Пусть

$$p \geq 2L/\min_{x\in S} f(x).$$

Тогда $g(x)$ – возрастающая функция.

Для доказательства теоремы покажем, что нижняя производная Дини функции g неотрицательна. Напомним следующие хорошо известные определения.

Если U – выпуклое множество, то конус $K(x,U)$ допустимых направлений в точке $x \in U$ по отношению к U есть

$$K(x,U) = \{u : \exists \alpha_0 > 0, \quad \text{такое, что} \quad x + \alpha u \in U \;\; \forall \alpha \in (0, \alpha_0]\}.$$

Легко проверить, что

$$K(y,S) = \{u : \sum_{i=1}^{n} u_i = 0,\; u_i \geq 0, \quad \text{если} \quad y_i = 0\},$$

$$K(x, \mathbb{R}^n_+) = \{v : v_i \geq 0, \quad \text{если} \quad x_i = 0\}.$$

Вычислим нижнюю производную Дини $g'_-(x,u)$ функции g в точке $x \in \mathbb{R}^n_+$ по направлению u.

Предложение 2.4 *Пусть* $x \in \mathbb{R}^n_+ \setminus \{0\}$ *и* $y = x/\sum_{i=1}^n x_i$. *Тогда для всех* $u \in K(x, \mathbb{R}^n_+)$ *имеем*

$$g'_-(x,u) = (\sum_{i=1}^{n} x_i)^{p-1} \cdot (f'_-(y, v(u)) + pf(y) \sum_{i=1}^{n} u_i), \tag{2.22}$$

где $v(u)$ *– вектор с координатами*

$$v_j = \sum_{k=1}^{n} (y_k u_j - y_j u_k) = u_j - y_j \sum_{k=1}^{n} u_k. \tag{2.23}$$

Доказательство. Пусть $H(x) = x/\sum_{i=1}^n x_i$ и $h(x) = f(Hx)$. Тогда $g(x) = h(x)(\sum_{i=1}^n x_i)^p$. Так как функция $x \mapsto (\sum_{i=1}^n x_i)^p$ дифференцируема и f липшицева, имеем

$$g'_-(x,u) = h'_-(x,u)(\sum_{i=1}^{n} x_i)^p + ph(x)(\sum_{i=1}^{n} x_i)^{p-1}(\sum_{i=1}^{n} u_i).$$

Поскольку отображение H непрерывно дифференцируемо и f липшицева, мы можем применить формулу для вычисления сложной функции, чтобы найти $h'_-(x,u)$. Имеем

$$h'_-(x,u) = f'_-(H(x), DH(x)u),$$

где $DH(x)$ – дифференциал отображения H. Легко проверить, что

$$DH(x)u = \frac{1}{(\sum_{i=1}^n x_i)^2} v'(u),$$

где координаты v'_j вектора $v'(u)$ имеют вид

$$v'_j = \sum_{k=1}^{n} x_k u_j - x_j u_k. \tag{2.24}$$

Тогда

$$h'(x,u) = f'_-(Hx, v')\frac{1}{(\sum_{i=1}^n x_i)^2}$$

и

$$g'_-(x,u) = f'_-(Hx, v')(\sum_{i=1}^n x_i)^{p-2} + ph(x)(\sum_{i=1}^n x_i)^{p-1}(\sum_{i=1}^n u_i).$$

Пусть $Hx = y$. Тогда $h(x) = f(y)$. Рассмотрим также вектор $v(u) = \dfrac{1}{(\sum_{i=1}^n x_i)}v'(u)$. Координаты этого вектора определены в (2.23). Имеем

$$g'_-(x,u) = (\sum_{i=1}^n x_i)^{p-1}(f'_-(y, v(u)) + pf(y)\sum_{i=1}^n u_i).$$

Предложение доказано. □

Продолжим доказательство теоремы. Пусть u_j совпадает с j-й координатой e_j. Тогда $v(e_j) \equiv v^j$ – вектор с координатами

$$v_i^j = -y_i \ \ (i \neq j), \qquad v_j^j = 1 - y_j.$$

Таким образом,

$$g'_-(x, u_j) = (\sum_{i=1}^n x_i)^{p-1}(f'_-(y, v^j) + pf(y)).$$

Получаем

$$g'_-(x, u_j) \geq 0 \Longleftrightarrow pf(y) \geq -f'_-(y, v^j).$$

Поскольку $\|v^j\|_1 = \sum_{k=1}^n |v_j^k| = 1 + \sum_{j=1}^n y_j = 2$, имеем

$$|f'_-(y, v^j)| \leq \max_{\|v\|_1 \leq 2} |f'_-(y, v)| \leq 2L,$$

где L – константа Липшица функции f по отношению к норме $\|\cdot\|_1$.

Пусть $c = \inf_{y \in S} f(y)$. Поскольку

$$pc \geq 2L, \tag{2.25}$$

получаем

$$\frac{\partial g(x)}{\partial x_j} \equiv g'_-(x, e_j) \geq 0 \quad \forall x \ \ \forall j.$$

Отсюда следует, что функция g возрастает на $\mathbb{R}^n_+\setminus\{0\}$. Действительно, применив теорему о среднем, имеем для всех $\alpha > 0$:

$$g(x+\alpha e_j) - g(x) \geq \inf_{0\leq\theta\leq\alpha} \frac{\partial g}{\partial x_j}(x+\theta e_j)\alpha \geq 0.$$

Пусть $z \geq 0$ и $x \in \mathbb{R}^n_+$. Пусть также $x_0 = x$, $x_1 = x + z_1 e_1$, $x_2 = x_1 + z_2 e_2, \dots, x_n = x_{n-1} + z_n e_n \equiv x + z$. Тогда $g(x+z) - g(x) = g(x_n) - g(x_1) = \sum_{k=0}^{n-1} g(x_k + z_{k+1}e_{k+1}) - g(x_k) \geq 0$. Значит, если (2.25) выполняется, то g возрастает. Теорема доказана. □

Следствие 2.6.1 *Если $p \geq \max\{1, 2L/c\}$ и $c = \min_{y\in S} f(y)$, то g – ВВЛ функция.*

Предложение 2.5 *Если $p > 2L/c$, то g строго возрастает для всех $x \neq 0$.*

Доказательство. Пусть $p_0 = 2L/c$. Обозначим через g_p функцию, определенную формулой (2.21) с $p \geq p_0$. Мы рассматриваем для простоты точки $x \in \mathbb{R}^n_+$ и $x - z_1 e_1 \in \mathbb{R}^n_+$ с $z_1 > 0$. Мы должны показать, что $g_p(x) > g_p(x - z_1 e_1)$ для всех $p > p_0$. Если $g_{p_0}(x) > g_{p_0}(x - z_1 e_1)$ и $p > p_0$, то также $g_p(x) > g_p(x - z_1 e_1)$. Пусть $g_{p_0}(x) = g_{p_0}(x - z_1 e_1)$. Тогда $g_{p_0}(x) = g_{p_0}(x - t e_1)$ для всех $t \in (0, z_1]$, следовательно, $(g'_{p_0})_-(x - \alpha e_1, e_1) = 0$ для всех $\alpha \in (0, z_1)$. Из (2.22) и (2.23) следует, что

$$f'_-(y(\alpha), v'(\alpha)) + p_0 f(y(\alpha)) = 0 \quad \forall \alpha \in (0, z_1),$$

где $y(\alpha)$ и $v'(\alpha)$ – некоторые векторы, не зависящие от p. Таким образом, если $p > p_0$, то $f'_-(y(\alpha), v'(\alpha)) + p_0 f(y(\alpha)) \neq 0$ для всех $\alpha \in (0, z_1)$. Так как $(g_p)'_-(x - \alpha e_1, e_1) \geq 0$, то $(g_p)'_-(x - \alpha e_1, e_1) > 0$ для всех $\alpha \in (0, z_1)$. Поскольку функция $\alpha \to g_p(x - \alpha e_1)$ липшицева, $g_p(\alpha)$ строго возрастает на $(0, z_1)$. Отсюда $g_p(x - z_1 e_1) < g_p(x)$. □

Пусть $g_y(\alpha) = g(\alpha y)$ для $\alpha > 0$ $(y \in S)$. Так как $g_y(\alpha) = \alpha^p g_y(\alpha) = \alpha^p f(y)$, то функция g_y дифференцируема в точке $\alpha = 1$ и

$$(g_y)'_-(1) = g'(y, y) = pf(y)$$

(мы можем получить тот же результат из (2.6)).

Пусть $p > 2L/c$. Из Теоремы 2.3 и Предложения 2.6 следует, что H-субдифференциал функции g в точке $y \in S$ состоит из единственного элемента $h_y(x) = \langle l_y, x\rangle + f(y) - \langle l_y, y\rangle$, где $l_y = pf(y)/y$.

2.7 Преобразование допустимого множества задачи липшицева программирования

Рассмотрим следующую задачу математического программирования:

$$\min \quad f(x),$$

$$g(x) \le 0, \tag{2.26}$$

$$x \ge 0, \tag{2.27}$$

где f, g – липшицевы функции на $\mathbb{R}^n$. Есть по крайней мере две возможности преобразования задачи липшицева программирования. Предположим, что задача оптимизации имеет вид

$$\begin{aligned} \min \quad & f(x), \\ g(x) &\le 0, \\ 0 \le x &\le C, \end{aligned} \tag{2.28}$$

где $C > 0$ – верхняя оценка для переменных. Тогда, введя дополнительные переменные

$$y_i = \frac{x_i}{Cn + q} \quad \forall i = \overline{1, n},$$

где $q > 0$, и добавив новую переменную y_{n+1}, можем переписать (2.28) в следующем виде:

$$\begin{aligned} \min \quad & f((Cn + q)y), \\ g((Cn + q)y) &\le 0, \\ 0 \le y_i \le \frac{C}{Cn + q} \quad & \forall i = \overline{1, n}, \\ \sum_{i=1}^{n+1} y_i &= 1, \\ y_{n+1} &\ge 0. \end{aligned} \tag{2.29}$$

Допустимое множество задачи (2.29) принадлежит единичному симплексу S.

Для вычислительных целей наличие дополнительной переменной может быть неудобным. Иногда можно применить способ без дополнительной переменной. Предположим, что задача оптимизации имеет вид

$$\min \quad f(x), \tag{2.30}$$

$$x \in X,$$

где множество $X \subset \mathbb{R}^n_+$ обладает свойством, что для любого $y \in S$ существуют $x \in X, \mu(y) > 0$, такие, что $\mu(y)y \in X$.

Рассмотрим функцию

$$F(y) = \min_{\substack{\lambda \geq 0 \\ \lambda y \in X}} f(\lambda y) \quad \forall y \in S.$$

Тогда задача (2.30) эквивалентна минимизации $F(y)$ на единичном симплексе.

Заметим, что $F(y)$ – маргинальная функция специального вида. Условия, гарантирующие, что $F(y)$ липшицева, можно найти, например, в [25].

2.8 Схема метода секущих углов для минимизации липшицевых функций

Рассмотрим задачу липшицева программирования в следующем виде:

$$\begin{gathered} \min \quad \left(M + f\left(\frac{y}{\sum_{i=1}^{n+1} y_i}\right)\right) \cdot (\sum_{i=1}^{n+1} y_i)^p \\ g(y) \leq 0, \\ \sum_{i=1}^{n+1} y_i = 1, \\ y \geq 0. \end{gathered} \tag{2.31}$$

Предположим, что параметры M и p выбраны таким образом, что целевая функция задачи (2.31) ВВЛ. Схема алгоритма следующая.

Метод секущих углов для липшицевых функций

Шаг 0. Полагаем $k := 0$. Выбираем произвольно допустимую точку $y_0 \in S$ и параметры $M > 0, p > 0$.

Шаг 1. Вычисляем вектор ℓ_k с координатами ℓ_{ki} :

$$\ell_{ki} = \frac{p(f(y_k) + M)}{y_{ki}}, \quad \text{если} \quad y_{ki} \neq 0; \quad \ell_{ki} = 0, \quad \text{если} \quad y_{ki} = 0,$$

где y_{ki} – i-я координата вектора y_k.

Шаг 2. Определяем функцию h_k:

$$h_k(y) = \langle \ell_k, y \rangle + (1-p)(f(y)+M).$$

Шаг 3. Находим глобальный оптимум задачи

$$\max_{0 \le i \le k} h_i(y) \to \min,$$
$$y \in S,$$
$$g(y) \le 0.$$

Пусть y^* – решение этой задачи.

Шаг 4. Полагаем $k := k+1, y_k := y^*$ и возвращаемся к Шагу 1.

В задаче (2.31) есть два параметра. Параметр $M > 0$ необходим, если значения целевой функции на допустимом множестве могут быть неположительными. Чтобы найти значение этого параметра, должна быть вычислена начальная нижняя оценка оптимального значения целевой функции. Второй параметр $p > 0$ зависит от константы Липшица и значения параметра M.

Заметим, что можно находить верхние и нижние оценки оптимального значения на каждой итерации метода секущих углов (см. выше):

$$\mu_k \ge f^* \ge \lambda_k \quad \text{для всех} \quad k,$$

где f^* – оптимальное значение целевой функции на допустимом множестве и

$$\mu_k = f(x_k), \quad \lambda_k = h_k(x_{k+1}) \quad \text{для всех} \quad k = 0, 1 \ldots .$$

Можно доказать, что

$$\lim_{k \to +\infty} \mu_k = \lim_{k \to +\infty} \lambda_k = f^*.$$

Это ведет к удобному критерию ε-оптимальности. Как только разность $\mu_k - \lambda_k$ становится меньше, чем $\varepsilon > 0$, где ε – требуемая точность, вычисления можно остановить.

2.9 Симплициальный метод для максимизации возрастающих функций

Важный класс задач глобальной оптимизации составляют задачи с монотонной целевой функцией. Такие задачи часто встречаются в приложениях, в частности, в математической экономике (см. [13]). Задачи оптимизации с возрастающими целевыми функциями могут иметь локальные экстремумы даже при очень простых ограничениях, что делает необходимой разработку методов глобальной оптимизации.

Есть много методов, позволяющих найти глобальный минимум. Важные их классы – методы внешней аппроксимации, методы внутренней аппроксимации, методы ветвления и отсечения и т. д. (см. [53, 90, 74]). Очень часто допустимое множество делится на прямоугольники или триангулируется (см. [74, 53]), и затем используются процедуры, похожие на методы ветвей и границ дискретной оптимизации. Эффективность зависит от типа разбиения и областей притяжения глобального максимума и других стационарных точек.

В этой части мы рассматриваем концептуальную схему алгоритма максимизации возрастающей функции на нормальном множестве ([90]). Такая задача важна для экономических приложений, но методов ее решения мало. Мы предлагаем способ вычисления верхних оценок целевой функции, ведущий к численно реализуемой схеме. Предлагается обобщение алгоритма, являющееся, по-существу, методом ветвления и отсечения (см. [53] и ссылки в данной книге). Такие методы стали очень популярны за последние годы и доказали численную эффективность.

Алгоритм использует разбиения единичного симплекса, образующиеся множества проектируются на границу допустимого множества. Простейшие разбиения – медианное и центральное. Верхние оценки могут быть вычислены, если целевая функция абстрактно вогнута (см. [11, 70, 95]) по отношению к более простым вспомогательным функциям (максимизируя эти функции) или непосредственно. Мы представляем предварительные результаты численных экспериментов для алгоритма, которые показывают, что он может быть эффективен для разнообразных задач глобальной оптимизации.

Рассмотрим следующую задачу глобальной оптимизации:

$$\max \quad f(x), \tag{2.32}$$

$$x \in X,$$

где X – нормальное множество из $\mathbb{R}^n_+$ и f – возрастающая непрерывная функция. Напомним, что функция f называется возрастающей, если

$$f(x) \geq f(y),$$

когда $x \geq y$, где $x, y \in \mathbb{R}^n_+$, то есть $x_i \geq y_i$ для всех i от 1 до n. Множество $Q \subseteq \mathbb{R}^n_+$ называется нормальным, если из $x \in Q,\ y \leq x$ следует $y \in Q$.

Обозначим

$$S = \{x \in \mathbb{R}^n_+ : \sum_{i=1}^{n} x_i = 1\}$$

единичный симплекс в $\mathbb{R}^n_+$. Алгоритм, который мы опишем в следующей части, использует разбиения единичного симплекса.

Если

$$S = S_1 \cup S_2 \cup \ldots \cup S_m,$$

где $S_i = \mathrm{co}\,\{v^i_1, v^i_2, \ldots, v^i_n\}$ и $v^i_j \in \mathbb{R}^n_+$ для всех $i = \overline{1,m}$ и всех $j = \overline{1,n}$, будем говорить, что $P = \{S_1, S_2, \ldots, S_m\}$ есть разбиение S. Если также $S_i \cap S_j = \emptyset$ при $i \neq j$, будем говорить, что P есть правильное разбиение S. Тогда, если мы имеем

$$\begin{array}{ll}
S_1 & = S_{11} \cup S_{12} \cup \ldots \cup S_{1m}, \\
S_2 & = S_{21} \cup S_{22} \cup \ldots \cup S_{2m}, \\
\ldots & \qquad \ldots\ldots\ldots\ldots \\
S_m & = S_{m1} \cup S_{m2} \cup \ldots \cup S_{mm},
\end{array}$$

будем говорить, что разбиение $\{S_{11}, S_{12}, \ldots, S_{mm}\}$ есть подразбиение P.

Пусть S_i – произвольный симплекс на S. Обозначим $m(S_i)$ его центр (сумму вершин, деленную на n). Для произвольного симплекса V обозначим

$$K(V) = \mathrm{cone}\,(V) \cap X = \{x \in X :\ x = \lambda y,\ y \in V,\ \lambda \geq 0\}.$$

Очевидно,

$$X = K(S) = \bigcup_{i=1}^{m} K(S_i).$$

Пусть

$$\partial X = \{x \in \mathbb{R}^n_+ : \nexists y \in X, y \geq x\}$$

– верхняя граница (множество максимальных элементов) X. Мы предполагаем, что множество X задано строго возрастающей непрерывной функцией g, то есть

$$X = \{x \in \mathbb{R}^n_+ : g(x) \le 1\},$$

где $g(x) > g(y)$, когда $x > y, x, y \in \mathbb{R}^n_+$. Здесь $x > y$ означает $x_i \ge y_i, i = \overline{1,n}$ и $x_j > y_j$ для некоторого j. Тогда

$$\partial X = \{x \in \mathbb{R}^n_+ : g(x) = 1\}.$$

Для каждых $x, y \in X$ обозначим

$$\min(x, y) = \{z \in \mathbb{R}^n_+ : z_i = \min\,(x_i, y_i) \quad \forall i = \overline{1,n}\},$$

$$\max(x, y) = \{z \in \mathbb{R}^n_+ : z_i = \max\,(x_i, y_i) \quad \forall i = \overline{1,n}\}.$$

Мы используем те же обозначения, если в скобках более двух векторов (так, $\max(x, y, z) = \max(\max(x, y), z)$).

Положим

$$\lambda(x) = \{\lambda \in \mathbb{R}_+ : g(\lambda x) = 1\}.$$

Допустим, что $g(0) < 1,\ \lim_{\|x\| \to +\infty} g(x) > 1$. Тогда $\lambda(x) > 0$ для всех $x \ne 0$. Так как g строго возрастает, функция λ определена на $\mathbb{R}^n_+ \backslash \{0\}$.

Предложение 2.6 *Функция* $\lambda(x)$ *убывает на* $\mathbb{R}^n_+ \backslash \{0\}$.

Доказательство. Пусть $x, y \in \mathbb{R}^n_+, y \ge x, x \ne 0, y \ne 0, x \ne y$. Тогда $y > x$. Из определения $\lambda(x)$ следует, что $g(\lambda(x)x) = 1$. Отсюда $g(\lambda(x)y) > 1$. Значит, $\lambda(y) < \lambda(x)$. Предложение доказано. □

2.10 Симплициальный алгоритм

Существует большое число вычислительных схем глобальной оптимизации, использующих триангуляцию. Мы рассматриваем метод для решения задачи (2.32), использующий конические проекции симплексов на границу допустимого множества. Это ведет к численно реализуемой схеме для допустимых множеств, заданных произвольной возрастающей функцией (которая может быть сложной функцией со многими экстремумами). Коническое проектирование требует специальных условий для верхних оценок оптимального значения. Алгоритм имеет следующий вид.

Симплициальный метод для максимизации возрастающих функций

Шаг 0. Полагаем $k := 0$. Выбираем произвольно начальное разбиение единичного симплекса $P_0 = \{S_1^0, S_2^0, \ldots, S_m^0\}$.

Шаг 1. Для каждого множества $K(S_i^k)$ находим верхнюю оценку

$$\hat{f}(S_i^k) \geq f(x) \quad \forall x \in K(S_i^k), i = \overline{1, m}.$$

Шаг 2. Вычисляем $\hat{f}_j = \max_i \hat{f}(S_i^k) = \hat{f}(S_j^k)$.

Шаг 3. Находим x_k как произвольную точку из $K(S_j^k)$.

Шаг 4. Обновляем разбиение и определяем P_{k+1} как подразбиение P_k.

Шаг 5. Полагаем $k := k+1$ и возвращаемся к Шагу 1.

Заметим, что данная концептуальная схема использует верхние оценки как главный критерий для нахождения x_k. Вместо этого можно на Шагах 1-3 вычислять центры симплексов S_i^k для каждого i:

$$c_i^k = m(S_i^k) \quad \forall i,$$

$$f(c_j^k) = \max_i f(c_i^k); \quad x_k = \lambda(c_j^k)c_j^k.$$

Таким образом, мы получаем "жадный алгоритм" глобальной оптимизации, основанный на значениях целевой функции.

Конкретная реализация алгоритма зависит от способа вычисления верхних оценок на Шаге 1 и способа разбиения P_k на Шаге 4. Докажем вначале теорему о сходимости, а затем приведем примеры выбора верхних оценок и разбиения.

Теорема 2.4 *Пусть выполняются следующие предположения:*

1) Диаметры множеств S_j^k стремятся к нулю, когда $k \to +\infty$. Здесь

$$diam(Y) = \sup_{x,y \in Y} \|x - y\| \quad \forall Y \subset \mathbb{R}^n.$$

Функция λ непрерывна и ограничена сверху, X ограничено.

2) $\hat{f}$ удовлетворяет условию

$$\hat{f}(Y) \to f(y),$$

когда $Y \to \{y\}$ по метрике Хаусдорфа (это условие выполнено, если $\hat{f}$ – непрерывная функция множества по отношению к метрике Хаусдорфа).

3) $\bigcup_{Y \in P_k} Y \supseteq X^*$, где $cone(X^*) \cap X$ есть множество глобальных максимумов целевой функции f на X.

Тогда для каждого $\varepsilon > 0$ за конечное число итераций будет найдена точка x_ε, такая, что

$$f(x_\varepsilon) \geq f(x) - \varepsilon \quad \forall x \in X.$$

Доказательство. Так как мы выбираем наибольшую верхнюю оценку, она дает верхнюю оценку оптимального значения. Пусть максимум достигается на множестве S_j^k. Без ограничения общности можно предположить, что $\mathrm{diam}\,(S_j^k) < \delta$, где $\delta > 0$ произвольно мало. Тогда по Предположениям 1-2 существует $\delta > 0$, такое, что для любого $\varepsilon > 0$ имеем $\hat{f}(S_j^k) \leq f(x) + \varepsilon$ для всех $x \in \mathrm{cone}\,(S_j^k) \cap X$. Тогда любая точка $x \in \mathrm{cone}\,(S_j^k) \cap X$ является ε-оптимальной. □

Можно рассматривать различные разбиения для сходимости. Простейшим и хорошо известным является медианное разбиение. Пусть

$$V = \mathrm{co}\,\{v_1, v_2, \ldots, v_n\}$$

есть симплекс с вершинами v_i. Положим

$$\|v_k - v_s\| = \max_{i,j} \|v_i - v_j\|.$$

Тогда разбиение следующее:

$$V = V_1 \cup V_2,$$

где

$$V_1 = \mathrm{co}\,(v_1, v_2, \ldots, v_{k-1}, (v_k + v_s)/2, v_{k+1}, \ldots, v_n),$$
$$V_2 = \mathrm{co}\,(v_1, v_2, \ldots, v_{s-1}, (v_k + v_s)/2, v_{s+1}, \ldots, v_n).$$

Для численных целей может быть полезным центральное разбиение, хотя оно не гарантирует сходимости (для сходимости, однако, его можно комбинировать с медианным разбиением). В этом случае

$$V = V_1 \cup V_2 \cup \ldots \cup V_n,$$

где $V_k = \mathrm{co}\,(v_1, \ldots, v_{k-1}, m(V), v_{k+1}, \ldots, v_n)$. Для других типов симплициальных разбиений см. работу [74] и ссылки в ней, где представлено детальное обсуждение.

Вычисление верхних оценок, вообще говоря, труднее. По крайней мере, возможны три подхода.

1) Если $X = S$, можно положить $\hat{f}(S_j^k) = f(\max(w_1, w_2, \ldots, w_n))$, где w_i – вершины симплекса S_j^k.

2) Пусть f – H-вогнутая функция (для детального представления теории абстрактного выпуклого анализа см. [11, 95, 70]), то есть

$$f(x) = \inf_{h \in H} h(x),$$

где $h(x)$ – некоторые простые функции, H – замкнутое множество. Тогда положим

$$H_s(x) = \min\{h_1(x), h_2(x), \ldots, h_s(x)\},$$

где f абстрактно вогнута по отношению к каждой h_i. Тогда, чтобы найти верхнюю оценку, достаточно найти максимум $H_s(x)$ на множестве $K(S_i^k)$ для каждого i. Это означает, что следует использовать метод секущих углов (см. [21] для его представления).

3) Можно использовать Предложение 2.6. Пусть $V = \mathrm{co}\,_{i=\overline{1,n}} w_i$ – произвольный симплекс из S. Тогда каждый элемент y из $K(V)$ может быть представлен как

$$y = (\sum_{i=1}^{n} \beta_i w_i) \times \lambda(\sum_{i=1}^{n} \beta_i w_i),$$

где $\sum_{i=1}^{n} \beta_i = 1, \beta_i > 0\ \forall i$. Так как λ – убывающая функция, то

$$(\sum_{i=1}^{n} \beta_i w_i) \times \lambda(\sum_{i=1}^{n} \beta_i w_i) \leq \max(w_1, \ldots, w_n) \times \lambda(\min(w_1, w_2, \ldots, w_n)).$$

Таким образом, значение целевой функции от правой части дает нам верхнюю оценку.

2.11 Метод ветвления и разбиения

Можно улучшить алгоритм, рассмотренный в предыдущей части, и построить метод ветвления и отсечения. Многие методы такого типа были предложены в последние годы (см. [74, 53]) и показали высокую эффективность для различных задач. Однако большинство методов было построено для общей задачи липшицева программирования. Свойства монотонности целевой функции позволяют упростить известные процедуры оптимизации и улучшить эффективность. Мы предлагаем алгоритм, использующий специальную структуру задачи (2.32).

Допустим, что можно найти верхнюю оценку количества стационарных точек на допустимом множестве (это может быть сделано, например, для целевых функций-полиномов). Предположим, что мы можем определить, является ли целевая функция выпуклой или вогнутой на симплексе. Для простоты положим $X = S$.

Симплициальный метод ветвления и отсечения

Шаг 1. Полагаем $k := 0$. Выбираем начальную точку $x_0 \in S$. Выбираем начальное разбиение P_0. Вычисляем $D(S)$ как верхнюю оценку числа стационарных точек на S. Полагаем $D := 0$, где D – количество найденных стационарных точек.

Шаг 2. Полагаем $x_r := x_0, f_r := f(x_0)$.

Шаг 3. Разбиение P_k может быть представлено как

$$P_k = \{V_k^1, V_k^2, \ldots, V_k^{i_k}\}.$$

Для каждого симплекса V_k^i вычисляем

$$\hat{f}(V_k^i) \geq f(x) \quad \forall x \in K(V_k^i);$$

$$\tilde{f}(V_k^i) = f(m(V_k^i));$$

$$\overline{f}(V_k^i) \leq f(x) \quad \forall x \in K(V_k^i).$$

Шаг 4. Вычисляем $\hat{f}(V_k^j) = \max \hat{f}(V_k^i), i = \overline{1, i_k}$.

Шаг 5. Если $f_r \geq \hat{f}(V_k^j)$, СТОП, x_r – оптимальная точка. Иначе переходим к Шагу 6.

Шаг 6. Если f выпукла на V_k^j, находим ее максимум x^* на V_k^j (сравнивая значения в вершинах). Полагаем $y := x^*$. Если $f(x^*) > f_r$, то $f_r := f(x^*), x_r := x^*$. Удалить V_k^j из P_k и перейти к Шагу 9.

Шаг 7. Если f вогнута на V_k^j, находим ее максимум x^* на V_k^j (методом выпуклого программирования). Полагаем $y := x^*$. Если $f(x^*) > f_r$, то $f_r := f(x^*), x_r := x^*$. Удалить V_k^j из P_k и перейти к Шагу 9.

Шаг 8. Найти стационарную точку $y \in V_k^j$. Если $f(y) > f_r$, то $f_r := f(y), x_r := y$.

Шаг 9. Если y – стационарная точка f на S, то $D := D + 1$. Если $D \geq D(S)$, СТОП, так как x_r – оптимальное решение.

Шаг 10. Для каждого симплекса V_k^i сделать сравнение $\hat{f}(V_k^i) \leq f_r$. Если неравенство выполняется, удалить V_k^i из P_k.

Шаг 11. Для каждой пары симплексов V_k^i, V_k^s сделать сравнения

$$\hat{f}(V_k^i) \leq \overline{f}(V_k^s); \tag{2.33}$$

$$\hat{f}(V_k^i) \leq \tilde{f}(V_k^s) \tag{2.34}$$

Если хотя бы одно из неравенств (2.33), (2.34) выполняется, удалить симплекс V_k^i из P_k.

Шаг 12. Если $P_k = \emptyset$, СТОП, так как x_r – оптимальное решение. Если V_j^k было удалено из P_k, то вернуться к Шагу 3 с $P_{k+1} := P_k \backslash W, k := k + 1$, где W – множество удаленных симплексов. Иначе разбиваем множество V_k^j. Полагаем

$$V_k^j = V_1^* \cup V_2^* \cup \ldots \cup V_q^*.$$

Тогда

$$P_{k+1} := (P_k \backslash \{V_k^j\}) \bigcup (\bigcup_{i=1}^{q} \{V_i^*\}).$$

Полагаем $k := k + 1$ и переходим к Шагу 3.

Данный алгоритм имеет ряд преимуществ. Прежде всего, он обеспечивает сокращение числа симплексов в разбиении (это стандартная процедура в методах ветвления и отсечения). Далее, возможно вставить методы локальной оптимизации в поиск глобального максимума (Шаги 6-7), так как вблизи глобального максимума они более эффективны. Более того, верхние оценки позволяют узнать точность наилучшего найденного решения.

Очевидно, можно использовать методы локального поиска для всех симплексов разбиения (не только V_k^j, а всех V_k^i). Можно также вычислять верхние оценки числа стационарных точек для каждого симплекса разбиения (и сравнивать их с числом найденных экстремумов). Эти модификации требуют дополнительных вычислений, но могут привести к более быстрому решению задачи (ввиду отсечения неоптимальных симплексов).

Установим сходимость алгоритма. Оказывается, если можно решать задачи локальной оптимизации на Шагах 6-7 точно, алгоритм находит точный глобальный оптимум (иначе он находит его приближенно).

Теорема 2.5 *Пусть выполняются следующие предположения:*

1) *Глобальный оптимум x^* единствен. Разбиение P_k – правильное разбиение для каждого k.*

2) *Целевая функция f локально выпукла или локально вогнута в окрестности x^*.*

3) *На Шаге 12 имеем*

$$diam\,(V_i^*) \le \gamma diam\,(V_j^k), \quad i = \overline{1,q}, \gamma < 1.$$

Предположим также, что выполняются условия Теоремы 2.4. Тогда решение x^* будет найдено за конечное число итераций.

Доказательство. По условиям теоремы диаметр V_j^k стремится к нулю. Покажем вначале, что множества V_j^k будут содержать x^* бесконечно много раз.

Пусть P_k – текущее разбиение и W – множество симплексов в нем, содержащих x^*. Обозначим

$$\beta = (f^* - \sup_{x \in X \setminus K(W)} f(x))/2.$$

Это значение положительно, так как f непрерывна и по Предположению 1. Пусть $V_i^k \notin W$ – другой симплекс из P_k. Без ограничения общности можно считать, что $\operatorname{diam} V_i^k \le \delta$, где $\delta > 0$ может быть произвольно мало. Тогда имеем

$$\hat{f}(V_i^k) \le f(x) + \beta < f^* \le \hat{f}(W_s) \quad \forall x \in V_i^k \quad \forall W_s \in W$$

для $\delta = \delta_1(\beta)$. Следовательно, симплекс из W дает наибольшую верхнюю оценку, и он будет выбран как V_j^k для некоторого k.

По Предположению 2 множество V_j^k будет принадлежать окрестности B_ε точки x^*, где f выпукла или вогнута, для некоторого $k = k^*$. Но это означает, что оптимальное решение будет найдено на Шаге 6 или Шаге 7 на итерации k^*. Теорема доказана. $\square$

2.12 Методы ветвей и границ для минимизации абстрактно выпуклых функций

Рассмотрим методы ветвей и границ для оптимизации абстрактно выпуклых функций (см. [70, 11, 95]). В частности, липшицевы функции (см. [74]) могут быть оптимизированы такими алгоритмами.

Простейшие формы таких методов основаны на разбиении допустимого множества (мы предполагаем, что оно – единичный симплекс). На каждом элементе разбиения в некоторой точке находится элемент абстрактного субдифференциала, что дает нижнюю оценку оптимального значения целевой функции.

Возможны различные стратегии. Простейшая из них – выбирать элемент разбиения с наименьшей нижней оценкой.

Алгоритм лучшей оценки

Шаг 0. Выбираем начальное разбиение P_0 допустимого множества $X = S$. Полагаем $k := 0$.

Шаг 1. Пусть $P_k = \{S_1^k, S_2^k, \dots, S_{N_k}^k\}$, где $X \supseteq \cup_{i=1}^{N_k} S_i^k$. Для каждого множества S_i^k выбираем конечное множество точек $X_i^k \subseteq S_i^k$.

Шаг 2. Для каждого множества X_i^k вычисляем элементы $h_i^k(y) \in M(y)$, где y принимает всевозможные значения в X_i^k. Полагаем

$$\hat{h}_i^k = \max_{y \in X_i^k} h_i^k(y).$$

Шаг 3. Минимизируем каждую функцию $\hat{h}_i^k$ на множестве S_i^k. Полагаем

$$\underline{h}_i^k = \min_{x \in S_i^k} \hat{h}_i^k(x).$$

Шаг 4. Вычисляем $\underline{h}_j^k = \min_i \underline{h}_i^k$.

Шаг 5. Разбиваем множество S_j^k. Пусть

$$S_j^k = \hat{S}^k \cup \tilde{S}^k.$$

Тогда

$$P_{k+1} := P_k \backslash \{S_j^k\} \cup \hat{S}^k \cup \tilde{S}^k.$$

Шаг 6. Для каждого множества X_i^k и всех $y \in X_i^k$ вычисляем $f(y)$. Полагаем $\hat{f} = \min_y f(y)$, то есть $\hat{f}$ – значение целевой функции в наилучшей найденной точке.

Шаг 7. Удаляем из разбиения все множества S_i^k, для которых $\hat{h}_i^k \geq \hat{f}$.

Шаг8. Полагаем $k := k + 1$ и возвращаемся к Шагу 1.

Эффективность алгоритма зависит от Шагов 2 - 4. Мы предлагаем выполнить несколько итераций метода секущих углов для определения X_i^k (начальная точка может быть выбрана произвольно из S_i^k). Это даст нижние оценки, сколь угодно близкие к оптимальному значению. Вместо разбиения S_j^k на две части можно его разбивать на произвольное количество частей, метод изменится несущественно.

Есть по крайней мере две возможности разбиения множества S_j^k. Предположим, что допустимое множество выпукло.

a) Пусть S_j^k – произвольное выпуклое множество. Рассмотрим задачу

$$\begin{aligned} \lambda &\to \max, \\ x \pm \lambda e_i \in S_j^k, \quad i &= \overline{1, n}, \\ x &\in S_j^k. \end{aligned}$$

Пусть $(\overline{x}, \overline{\lambda})$ – решение этой задачи. Определим гиперплоскость $\{x : (a, x) = (a, \overline{x})\}$, где $a \in \mathbb{R}_+^n$, $\|a\| = 1$. Тогда разбиение определяется следующим образом:

$$\hat{S}^k = S_j^k \cap \{x : (a, x) \leq (a, \overline{x})\},$$

$$\tilde{S}^k = S_j^k \cap \{x : (a, x) \geq (a, \overline{x})\}.$$

b) Пусть S_j^k – многогранник, то есть $S_j^k := \mathrm{co}_{i=1}^{m} v_i$. Вычислим

$$\|v_r - v_s\| = \max_{i,j} \|v_j - v_i\|.$$

Пусть $w = (v_r + v_s)/2$. Выберем вектор b, $\|b\| = 1$. Положим

$$\hat{S}^k = S_j^k \cap \{x : (b, x) \le (b, w)\},$$

$$\tilde{S}^k = S_j^k \cap \{x : (b, x) \ge (b, w)\}.$$

Допустим, что выполняются следующие условия:

1) На каждой итерации k выполняются неравенства

$$\mathrm{diam}\,(\hat{S}^k) \le \gamma_1 \mathrm{diam}\,(S_j^k), \tag{2.35}$$

$$\mathrm{diam}\,(\hat{S}^k) \le \gamma_2 \mathrm{diam}\,(S_j^k) \tag{2.36}$$

для $0 < \gamma_1, \gamma_2 < 1$;

2) Все абстрактные субградиенты h_i^k непрерывны.

Тогда мы имеем, что $\mathrm{diam}\,(S_j^k) \le \varepsilon$ для некоторого k. Минимальное значение максимума абстрактных субградиентов на S_j^k не превосходит f^*. Поэтому, в силу непрерывности и абстрактной выпуклости f, мы можем найти глобальный минимум с любой заранее заданной точностью. Действительно, для некоторой точки $y \in S_j^k$ имеем

$$f(y) = \hat{h}_j^k(y) \le \hat{h}_j^k(\tilde{x}) + \delta \le f^* + \delta,$$

где $\tilde{x}$ – точка из S_j^k, в которой достигается минимум $\hat{h}_j^k$. Мы имеем $\delta > 0$, и δ может стать сколь угодно малым при уменьшении ε.

Можно сохранять диаметры множеств S_i^k постоянными и проводить итерации метода секущих углов на каждом элементе разбиения. Сходимость алгоритма к глобальному оптимуму следует из сходимости метода секущих углов.

Для повышения эффективности (число множеств в P_k может стать очень большим) полезно использовать гибридные методы, сочетая ветвление с локальной оптимизацией.

Метод ветвей и границ с локальной оптимизацией

Шаг 0. Полагаем $k := 0$. Выбираем начальное разбиение P_0 и начальную верхнюю оценку D_0 числа стационарных точек на X. Полагаем $D := 0$, где D – число найденных стационарных точек.

Шаг 1. Пусть $P_k = \{S_1^k, S_2^k, \ldots, S_{N_k}^k\}$. Для каждого множества S_i^k вычисляем стационарную точку x_i^k функции $f(x)$ на S_i^k.

Шаг 2. Проверяем, является ли x_i^k стационарной точкой $f(x)$ на X (для всех i). Если да, полагаем $D := D + 1$.

Шаг 3. Если $D \geq D_k$, СТОП, все стационарные точки известны.

Шаг 4. Для каждой точки x_i^k вычисляем $f(x_i^k)$ и элементы $\hat{h}_i^k \in M(x_i^k)$.

Шаг 5. Для каждого i вычисляем

$$y_i^k = \mathrm{Arg}\ \min{}_{x \in S_i^k} \hat{h}_i^k(x).$$

Шаг 6. Пусть $\hat{h}_j^k(y_j^k) \leq \hat{h}_i^k(y_i^k)$ для всех i. Разбиваем множество $S_j^k = \hat{S}_j^k \cup \tilde{S}_j^k$. Полагаем $P_{k+1} := P_k \backslash \{S_j^k\} \cup \{\hat{S}_j^k\} \cup \{\tilde{S}_j^k\}$.

Шаг 7. Обновляем верхнюю оценку D_k. Полагаем $k := k + 1$ и возвращаемся к Шагу 1.

Преимущество данного метода в том, что не обязательно вычислять стационарные точки для всех S_i^k, достаточно найти их для $\hat{S}_j^k$ и $\tilde{S}_j^k$ (начиная с первой итерации). Остальные стационарные точки известны из предыдущих итераций.

Если значение $\hat{h}_i^k(y_i^k)$ больше, чем $f(x_l^k)$ или $f(y_l^k)$ для некоторого $l \neq i$, множество S_i^k может быть удалено из разбиения, так как оно не содержит глобальный минимум. Также, если $f(x)$ выпукла на S_i^k (или S_i^k – симплекс, и $f(x)$ вогнута на нем), можно найти x_i^k как глобальный минимум $f(x)$ на S_i^k и удалить S_i^k из разбиения.

Верхняя оценка D_k числа стационарных точек может быть найдена для большого числа целевых функций, в частности, для многомерных полиномов.

Чтобы найти элемент абстрактного субдифференциала липшицевой функции, достаточно знать константу Липшица. Можно взять (см. также [68]), как обычно,

$$\hat{h}_i^k(y) = f(x_i^k) - L\|y - x_i^k\|.$$

Предположим, что существует окрестность A_δ глобального оптимума, такая, что метод локального поиска находит глобальный оптимум, стартуя из любой точки $x_0 \in A_\delta$. Пусть также выполнены неравенства (2.35), (2.36). Тогда, если диаметры симплексов становятся достаточно малы, точка из A_δ, следовательно, и глобальный оптимум будут найдены за конечное число итераций. Конечно, применение абстрактного выпуклого анализа (см. [70, 11, 95]) должно быть подтверждено численной эффективностью алгоритмов. Различные структурированные задачи глобальной оптимизации были решены эффективно этой техникой (см. [24]).

Глава 3

Квазивыпуклое программирование с коническим проектированием

3.1 Основные понятия и определения

В данной главе рассматривается задача минимизации квазивыпуклых функций, являющихся важным классом обобщенно выпуклых функций (см. [32, 33, 5, 73, 46]) и имеющих многочисленные применения в экономике. Предлагается алгоритм субградиентного типа (см. [51, 92, 67, 18, 10]).

Пусть D – замкнутое выпуклое множество из $\mathbb{R}^n$.

Определение 3.1.1 [4, 6, 7] *Вектор $g \in \mathbb{R}^n$ называется обобщенно-опорным к множеству D в точке x, если для любого $y \in D$*

$$[g, y - x] \leq 0.$$

Если неравенство выполняется строго для $y \neq x$, вектор называется строго обобщенно-опорным.

Предложение 3.1 *Пусть D, D' – два множества из $\mathbb{R}^n$, $D \subset D'$. Тогда конус обобщенно-опорных векторов к D' принадлежит конусу обобщенно-опорных векторов к D.*

Это означает, что если требуется найти обобщенно-опорный вектор к множеству, достаточно найти обобщенно-опорный вектор к множеству, его содержащему.

Определение 3.1.2 *Гиперплоскость $\Gamma = \{x \in \mathbb{R}^n : [q, x - y] = 0\}$ называется обобщенно-опорной гиперплоскостью к $D \subset \mathbb{R}^n$ в $x \in \mathbb{R}^n$, если q или $-q$ является обобщенно-опорным вектором к D в x.*

Если q – строго обобщенно-опорный вектор, Γ называется строго обобщенно-опорной гиперплоскостью. Определим некоторые
множества для $\varepsilon > 0$ и произвольной точки $y \in \mathbb{R}^n$. Пусть f – произвольная функция, определенная на R^n. Тогда

$$E_\varepsilon(y) = \{x \in D : f(x) \leq f(y) - \varepsilon\},$$

$$\tilde{E}_\varepsilon(y) = \{x \in D : f(x) < f(y) - \varepsilon\},$$

$$W_\varepsilon(y) = \{g \in \mathbb{R}^n : [g, x - y] \leq 0 \quad \forall x \in E_\varepsilon(f, y)\}.$$

Если $\varepsilon = 0$, пишем просто $E(y), U(y), W(y)$. Элементы множества $W_\varepsilon(y)$ называются ε-обобщенно-опорными векторами для f по отношению к множеству D (см. [4, 5, 6]).

Определение 3.1.3 *Вектор $g \in \mathbb{R}^n$ называется ε-подходящим направлением в точке x по отношению к множеству D для функции f, если существует $\alpha > 0$, такое, что $x + \alpha g \in E_\varepsilon(f, y)$. Если $\varepsilon = 0$, мы называем такой вектор подходящим направлением.*

Можно дать аналогичное определение ε-обобщенно-опорных векторов.

Определение 3.1.4 *Вектор $g \in \mathbb{R}^n$ называется ε-обобщенно-опорным вектором в точке y, если для любого ε-подходящего направления p в точке y мы имеем $[g, p] \leq 0$.*

Дадим еще несколько определений.

Определение 3.1.5 *Функция f, определенная на выпуклом множестве D, называется квазивыпуклой, если для любого $\alpha \in \mathbb{R}$ множество $L_f(\alpha) = \{y \in D : f(y) \leq \alpha\}$ выпукло. Эквивалентно, если для любых x, y выполняется следующее неравенство:*

$$f(\lambda x + (1 - \lambda) y) \leq \max\{f(x), f(y)\} \quad \forall \lambda \in [0, 1],$$

то функция квазивыпукла.

Определение 3.1.6 *Функция f, определенная и дифференцируемая по направлениям на открытом выпуклом множестве $D \subseteq \mathbb{R}^n$, называется псевдовыпуклой, если для любой точки $x \in D$ из $f'(x, y - x) \geq 0$ следует $f(y) \geq f(x)$ для весх $y \in D$.*

Здесь $f'(x,d)$ – производная функции f в точке x по направлению d.

Пусть на D определена выпуклая функция.

Определение 3.1.7 *Вектор $g'(y) \in \mathbb{R}^n$ называется субградиентом функции g в точке y, если для любого $x \in \mathbb{R}^n$*

$$g(x) - g(y) \geq [g'(y), x - y].$$

Лемма 3.1 *Для любого $y \in D$ вектор $g'(y)$ является ε-обобщенно - опорным вектором для множества $E_\varepsilon(y)$.*

3.2 Сведение задачи квазивыпуклого программирования к задаче безусловной минимизации методом конического проектирования

Пусть мы имеем следующую задачу оптимизации:

$$\min \quad f(x),$$

$$x \in D,$$

где D – замкнутое выпуклое множество из $\mathbb{R}^n$ с непустой внутренностью, f – квазивыпуклая непрерывная функция. Обозначим $D^* = \operatorname{Arg}\min_{x \in D} f(x)$. Предположим, что $D^* \neq \emptyset$ и под "min" будем понимать глобальный минимум. Мы сведем начальную задачу к задаче безусловной минимизации коническим проектированием (см. также [5, 9]).

Пусть известны точка $z \in intD$ и вектор q, строго обобщенно-опорный в этой точке к D^*. Предположим также, что для некоторой точки $v \in \mathbb{R}^n$ выполняются следующие неравенства:

$$[q, x^* - v] \geq 0 \quad \forall x^* \in D^*, \tag{3.1}$$

$$[q, z - v] > 0.$$

Определим гиперплоскость $U = \{u \in \mathbb{R}^n : [q, u] = [q, v]\}$.

Предложение 3.2 *Гиперплоскость U – обобщенно-опорная в точке v к множеству D^*.*

Доказательство. Действительно, из (3.1) следует, что вектор $-q$ является обобщенно-опорным вектором в v по отношению к D^*, и поэтому U – обобщенно-опорная гиперплоскость по определению. □

Определим следующие множества:

$$I(u,z) = \{t \in [0,1] : u + t(z-u) \in D\},$$

$$I^*(u,z) = \{t^* \in [0,1] : f(u + t^*(z-u)) = \min_{t \in I(u,z)} f(u + t(z-u))\}.$$

На гиперплоскости U определим функцию

$$\Phi(u) = \min_{t \in I(u,z)} f(u + t(z-u)).$$

Пусть также

$$U^* = \{u^* \in U : \Phi(u^*) = \min_{u \in U} \Phi(u)\}.$$

Предложение 3.3 *Для любых $u \in U, x^* \in D^*$ выполняется неравенство*

$$f(z) \geq \Phi(u) \geq f(x^*).$$

Доказательство. Правая часть следует из того факта, что

$$\Phi(u) = \min_{t \in I(u,z)} f(u + t(z-u)) \geq f(x^*),$$

а левая выполняется, так как $1 \in I(u,z)$. □

Определение 3.2.1 *Будем говорить, что точки $x \in D$ и $u \in U$ соответствуют друг другу, если существует $t^* \in I^*(u,z)$, такое, что*

$$u + t^*(z-u) = x.$$

Если функция f достигает минимума на D, то для каждого $u \in U$ функция $f(u + t(z-u))$ достигает минимума на $I(u,z)$. Поэтому каждой $u \in U$ соответствует точка $x \in D$. Обратное неверно.

Следующая теорема утверждает, что для любой $x^* \in D^*$ найдется соответствующая точка $u \in U$, в которой функция Φ достигает минимума на U.

Теорема 3.1 *Пусть $x^* \in D^*$. Тогда:*

1) Существует $u^* \in U$, *соответствующая* $x^* \in D^*$, *и*

$$u^* = \frac{z[q, v - x^*] - x^*[q, v - z]}{[q, z - x^*]}; \tag{3.2}$$

2) Число $t^* = [q, x^* - v]/[q, z - v]$ *принадлежит* $I^*(u^*, z)$*;*

3) $0 \leq t^* < 1$*;*

4) Функция Φ *достигает минимума на* U *в* u^**;*

5) $\min_{x \in D} f(x) = \min_{u \in U} \Phi(u)$.

Доказательство. Используя коническое проектирование, получаем, что (3.2) справедливо, так как есть единственная точка на луче $z + \alpha(x - z)$, принадлежащая U. Умножая равенство (3.2) на q, находим, что $[u^*, q] = [v, q]$. Следовательно, $u^* \in U$. Легко проверить, что $[q, z] > [q, u^*]$. Более того, так как q – строго обобщенно-опорный вектор в z к множеству D^*, мы имеем $[q, x^* - z] < 0$. Из этого и определения $I^*(u, z)$ следует, что $t^* < 1$. Простым расчетом, принимая во внимание (3.2), устанавливаем, что

$$u^* + t^*(z - u^*) = x^*.$$

Так как $x^* \in D$, число $t^* \in I(u^*, z)$. Покажем, что $t^* \in I^*(u^*, z)$.

По определению Φ мы имеем

$$\Phi(u^*) = \min_{t \in I(u^*, z)} f(u^* + t^*(z - u^*)) \leq f(u^* + t(z - u^*)) = f(x^*).$$

Но по Предложению 3.3 $\Phi(u^*) \geq f(x^*)$. Поэтому $\Phi(u^*) = f(x^*)$ и

$$\min_{t \in I(u^*, z)} f(u^* + t(z - u^*)) = f(u + t^*(z - u^*)).$$

Равенство означает, что $u^* \in U^*$, $t^* \in I^*(u^*, z)$. Теорема доказана. □

Теорема 3.2 *Пусть* $u^* \in U$ *и* x^* *соответствует* u^*. *Тогда* $x^* \in D^*$.

Доказательство. Допустим, что $x^* \notin D^*$. Пусть x' – произвольная точка из D^*. Тогда найдется соответствующая точка $u' \in U$. Но

$$\Phi(u^*) = f(x^*) > f(x') = \Phi(u').$$

Это противоречит условиям теоремы. Теорема доказана. □

Для выбора точек z, v и вектора q возможны различные стратегии. В частности, их можно выбрать следующим образом.

Методами выпуклого программирования находятся произвольная точка $z \in intD$ и строго обобщенно-опорный вектор q в этой точке. Например, если f псевдовыпукла и дифференцируема, мы можем взять $q = f'(z)$, если градиент ненулевой. Затем выбираем $\mu \in R$ и задаем

$$v = z - \mu q, \tag{3.3}$$

где $\mu \geq \|q\|^{-2} \max_{x \in D}[q, z - x]$. Так как $D^* \subseteq D$, вектор q – обобщенно-опорный к D^*. При этом $\mu > 0$. Учитывая (3.3), получаем

$$[q, v - x^*] = [q, z - x^*] - \mu\|q\|^2 \leq [q, z - x^*] - \max_{x \in D}[q, z - x] \leq 0.$$

Это означает, что $-q$ является обобщенно-опорным вектором к D^* в точке v.

Чтобы строить итерационные методы, может быть полезен следующий результат.

Лемма 3.2 *Пусть для некоторой последовательности* $\{u_k\}$ *выполняется условие*

$$\lim_{k \to +\infty} \Phi(u_k) = \min_{u \in U} \Phi(u).$$

Тогда для последовательности соответствующих точек из D *мы имеем*

$$\lim_{k \to +\infty} f(x_k) = \min_{x \in D} f(x).$$

Если есть точка u^*, *такая, что* $\|u - u^*\| \to 0$, *функция* f *непрерывна на* D *и каждой* $u \in U$ *соответствует ровно одна* $x \in D$, *то существует точка* x^*, *такая, что* $\|x^* - x\| \to 0$.

Доказательство. Имеем

$$\lim_{k \to +\infty} \Phi(u_k) = \lim_{k \to +\infty} f(x_k) = f(x^*) = \Phi(u^*),$$

где x^* соответствует u^*. Докажем сильную сходимость $\{x_k\}$. Пусть $x_k \in D$ соответствует $u_k \in U$, $x^* \in D^*$ соответствует $u^* \in U^*$. Тогда

$$\|x_k - x^*\| = \|u_k + t_k(z - u_k) - u^* - t^*(z - u^*)\| =$$

$$= (1 - t_k)\|u_k - u^*\| + |t_k - t^*|\|z - u^*\|.$$

Докажем, что $|t_k - t^*| \to 0$ (так как каждой точке u соответствует ровно одна точка x, значение t^* единственно). Предположим противное. Тогда существует подпоследовательность $\{t_l\}, l \in L$, такая, что выполняется условие

$$|t_l - t^*| > \varepsilon \quad \forall l \in L$$

Подпоследовательность ограничена, так что из нее можно выделить сходящуюся подпоследовательность L_2. Пусть $\lim_{l \in L_2} t_l = \bar{t}$. Мы имеем $|\bar{t} - t^*| \geq \varepsilon$. Обозначим $\overline{x} = u^* + \bar{t}(z - u^*)$. Тогда для любого $l \in L_2$

$$\|x_l - \overline{x}\| = \|u_l + t_l(z - u_l) - u^* - \bar{t}(z - u^*)\| =$$

$$= \|u_l - u^*\|(1 - t_l) + |t_l - \bar{t}|\|z - u^*\|.$$

Так как $\lim_{l \in L_2} x_l = \overline{x}$, выполнено равенство $\lim_{l \in L_2} f(x_l) = f(\overline{x})$. Но таким образом мы получили противоречие, так как $f(x_k)$ стремится к $f(x^*)$ и $f(\overline{x}) > f(x^*)$, поскольку соответствующая точка единственна. Лемма доказана. □

Следствие 3.2.1 *Пусть f – линейная функция, то есть $f(x) = [c, x]$ $\forall x \in \mathbb{R}^n$. Пусть также $q = c$ и существует u^*, такая, что $\lim_{k \to +\infty} \|u_k - u^*\| = 0$. Тогда $\lim_{k \to +\infty} \|x_k - x^*\| = 0$, где x_k соответствует u_k, x^* соответствует u^*.*

Доказательство. В силу линейности функции f она непрерывна. Поэтому достаточно показать, что каждой $u \in U$ соответствует единственная точка $x \in D$. Но это следует из линейности целевой функции, так как функция $f(u + t(z - u))$ монотонна на $[0, 1]$ и минимум на отрезке достигается в единственной точке. Следствие доказано. □

3.3 Свойства маргинальных функций

Изучим некоторые свойства маргинальных функций. Пусть

$$\Phi(u) = \min_{x \in a(u)} f(x). \tag{3.4}$$

Если точечно-множественное отображение параметризовано, то есть

$$a(u) = \bigcup_{z \in z(u) \subseteq Z} b(u,z) \quad z \in \mathbb{R}^s \quad z(u) : U \to Z,$$

где b – отображение из $U \times Z$ в D, формула (3.4) принимает вид

$$\Phi(u) = \min_{z \in Z(u)} f(b(u,z)) = \min_{z \in Z(u)} \tilde{f}(u,z),$$

и Φ может быть представлена в виде (3.4).

Такие свойства маргинальных функций, как дифференцируемость по направлениям и обобщенная выпуклость, изучались рядом авторов (см. [34] и ссылки в этой работе). Предположим, что $a(u)$ имеет выпуклые компактные образы, то есть для любого $u \in U$ множество $a(u)$ выпукло и компактно. Обозначим

$$Im(a) = \cup_{u \in U} a(u), \quad U^* = \text{Arg min}_{u \in U} \Phi(u),$$

$$U_\varepsilon(\overline{u}) = \{u \in U : \Phi(u) \le \Phi(\overline{u}) - \varepsilon\}.$$

Теорема 3.3 *Пусть* $Im(a) \subseteq D, Im(a) \cap D^* \neq \emptyset$. *Тогда* $f(x^*) = \Phi(u^*)$ *для всех* $x^* \in D^*, u^* \in U^*$. *Более того, если* $u^* \in U^*$, *то* $a(u^*) \cap D^* \neq \emptyset$ *и* $Arg\ min_{x \in a(u^*)} f(x) \subset D^*$.

Доказательство. По условиям теоремы $Im(a) \cap D^* \neq \emptyset$, так что существует точка $\overline{u} \in U$, такая, что $a(\overline{u}) \cap D^* \neq \emptyset$. Тогда существует точка $\overline{x} \in a(\overline{u}) \in D^*$, и $f(\overline{x}) = f(x^*)$ по построению. Значение Φ всегда равно значению f в некоторой точке D. Но поскольку

$$\Phi(\overline{u}) = \min_{x \in a(\overline{u})} f(x) = f(x^*) = \Phi(u^*),$$

первая часть теоремы доказана.

Пусть теперь $u^* \in U$. Тогда $\Phi(u^*) = f(x^*)$ из указанного выше. Допустим, что $a(u^*) \cap D^* = \emptyset$. Тогда $f(x) > f(x^*)$ для всех $x \in a(u^*)$. В силу компактности и выпуклости a выполняется соотношение

$$\min_{x \in a(u^*)} f(x) > f(x^*) = \Phi(u^*),$$

ведущее к противоречию. Наконец, так как $\text{Arg min}_{x \in a(u^*)} f(x) \subseteq D$ и $f(x) = f(x^*)$ для $x \in D^*$, имеем $\text{Arg min}_{x \in a(u^*)} f(x) \in D^*$. Теорема доказана. □

Следующая теорема связывает лебеговы множества целевой и маргинальной функций.

Теорема 3.4 *Пусть* $L_f(c) = \{x \in D : f(x) \le c\}, L_\Phi(c) = \{u \in U : \Phi(u) \le c\}$. *Тогда*

$$L_\Phi(c) = a^{-1}(Im(a) \cap L_f(c)) = \bigcup a^{-1}(v), v \in Im(a) \cap L_f(c),$$

где a^{-1} *– обратное отображение к* a.

Доказательство. Вначале покажем, что $L_\Phi(c) \subseteq a^{-1}(Im(a) \cap L_f(c))$. Пусть $u \in a^{-1}(Im(a) \cap L_f(c))$. Тогда $f(x) \le c$. По определению маргинальной функции имеем, что $\Phi(u) \le c$ или $u \in L_\Phi(c)$.

Обратно, пусть $\overline{u} \in L_\Phi(c)$. Тогда в соответствующей точке x (хотя бы одна такая существует) выподняется $f(x) = \Phi(\overline{u})$. Так как x – соответствующая точка, то $x \in Im(a)$, и поскольку $\Phi(\overline{u}) = f(x) \le c$, то $x \in L_f(c)$. По определению обратного отображения, выполняется включение $\overline{u} \in a^{-1}(Im(a) \cap L_f(c))$, которое завершает доказательство теоремы. □

Введем следующее семейство множеств:

$$R(u) = \{x \in a(u) : f(x) = \Phi(u) \quad \forall u \in U\}.$$

Для каждого фиксированного $\overline{u}$ множество $R(\overline{u})$ состоит из точек, соответствующих $\overline{u}$.

Теорема 3.5 *Пусть для любых* u_1, u_2 *и любого* $\lambda \in [0,1]$ *выполняется условие*

$$a(\lambda u_1 + (1-\lambda)u_2) \cap C(u_1, u_2) \ne \emptyset,$$

где $C(u_1, u_2) = \{x \in \mathbb{R}^n : x = \alpha x_1 + (1-\alpha)x_2, x_i \in R(u_i), i = 1, 2, \alpha \in [0,1]\}$. *Тогда, если функция* f *квазивыпукла на множестве* D, *функция* Φ *квазивыпукла на множестве* U.

Доказательство. Зафиксируем соответствующие точки $u_1, u_2 \in U$ и $x_1 \in R(u_1), x_2 \in R(u_2)$. По определению,

$$\Phi(\lambda u_1 + (1-\lambda)u_2) = \min_{x \in a(\lambda u_1 + (1-\lambda)u_2)} f(x).$$

Без ограничения общности мы можем предположить, что

$$a(\lambda u_1 + (1-\lambda)u_2) \cap \mathrm{co}\,(x_1, x_2) \ne \emptyset,$$

то есть существует

$$\overline{x} \in a(\lambda u_1 + (1-\lambda)u_2), \beta \in [0,1], \overline{x} = \beta x_1 + (1-\beta)x_2.$$

Тогда

$$\Phi(\lambda u_1 + (1-\lambda)u_2) = \min_{x \in a(\lambda u_1 + (1-\lambda)u_2)} f(x) \leq f(\overline{x}) =$$
$$= f(\beta x_1 + (1-\beta)x_2) \leq \max\{f(x_1); f(x_2)\}.$$

Так как $x_i \in R_i$, получаем $f(x_i) = \Phi(u_i), i = 1, 2$, и теорема доказана. □

Данная теорема дает достаточные условия квазивыпуклости для фиксированной целевой функции. Удобно, однако, определить класс отображений, которые дают квазивыпуклую маргинальную функцию Φ для любой квазивыпуклой функции f.

Теорема 3.6 *Пусть для любых точек $u_1, u_2 \in U$ и любых точек $x_1 \in a(u_1), x_2 \in a(u_2)$ и числа $\lambda \in [0,1]$*

$$a(\lambda u_1 + (1-\lambda)u_2) \cap [x_1; x_2] \neq \emptyset. \tag{3.5}$$

Здесь $[x1; x2]$ – отрезок между x_1 и x_2. Тогда функция Φ квазивыпукла для любой квазивыпуклой непрерывной функции f.

Доказательство. Используем Теорему 3.5. Пусть f – произвольная квазивыпуклая функция на D. Тогда для каждого $u \in U$ множество $R(u)$ непусто. Для любых $u_1, u_2 \in U$ мы имеем $R(u_1) \subseteq a(u_1), R(u_2) \subseteq a(u_2)$ и

$$a(\lambda u_1 + (1-\lambda)u_2) \cap C(u_1, u_2) \neq \emptyset.$$

Это значит, что выполнены условия Теоремы 3.5, и утверждение доказано. □

Обратное также верно.

Теорема 3.7 *Пусть для любой непрерывной квазивыпуклой функции f функция Φ квазивыпукла. Тогда отображение удовлетворяет условию 3.5.*

Доказательство. Предположим, что это неверно. Тогда найдутся точки $u_1, u_2 \in U, x_1 \in a(u_1), x_2 \in a(u_2)$ и число $\lambda \in [0,1]$, такие, что

$$a(\lambda u_1 + (1-\lambda)u_2) \cap \mathrm{co}\,(x_1, x_2) = \emptyset.$$

Заметим, что при $\lambda = 0$ и $\lambda = 1$ это условие не выполняется в силу выбора x_1, x_2.

Построим функцию F следующим образом:

$$F = dist(x, \mathrm{co}\,(x_1, x_2))$$

для любого $x \in \mathbb{R}^n$, где $dist$ означает расстояние от точки до множества. Эта функция выпукла и непрерывна, следовательно, функция $\Psi(u) = \min_{t \in I(u,z)} F(t)$ должна быть квазивыпукла. Но это невозможно, поскольку $\Psi(u_1) = 0, \Psi(u_2) = 0$ и $\Psi(\lambda u_1 + (1 - \lambda) u_2) > 0$. Данное противоречие доказывает теорему. $\square$

Применим Теорему 3.6, чтобы построить обратное отображение для функции Φ. Заметим, что мы имеем

$$a^{-1}(x) = z + (x - z)\frac{[q, v] - [q, z]}{[q, x] - [q, z]}.$$

Теорема 3.8 *Пусть* $\min_{x \in D} f(x) \le b < f(z)$. *Тогда*

$$L_\Phi(b) = U \cap (z + cone\,(L_f(b) \cap V - z)). \tag{3.6}$$

Доказательство. По предыдущим результатам,

$$L_\Phi(b) = a^{-1}(Im(a) \cap L_f(b)) = a^{-1}(V \cap L_f(b)).$$

Так как $b < f(z), z \cap L_f(b) = \emptyset$, имеем

$$L_\Phi(b) = a^{-1}(V \cap L_f(b)) = \bigcup_{x \in V \cap L_f(b)} [z + (x - z)\frac{[q, v] - [q, z]}{[q, x] - [q, z]}]. \tag{3.7}$$

Поскольку $[q, z] > [q, x] \ge [q, v]$, дробь строго положительна. Тогда значение в скобках принадлежит множеству $z + \mathrm{cone}\,(L_f(b) \cap V - z)$. Непосредственным подсчетом убеждаемся, что правая часть (3.7) совпадает с $U \cap (z + \mathrm{cone}\,(L_f(b) \cap V - z))$. $\square$

3.4 Метод минимизации функции конического проектирования

Установим связь между обобщенно-опорными векторами для функции Φ и f.

Теорема 3.9 *Пусть* Φ *– функция конического проектирования,* $\overline{u} \in U, \overline{x} \in D, \overline{x} \neq z$ *– соответствующая точка. Пусть в точке* $\overline{x}$ *известен обобщенно-опорный вектор* g*, такой, что*

$$[g, z - \overline{x}] = 0.$$

Тогда вектор $\overline{g} = g - [g, e]e$ *– обобщенно-опорный для функции* Φ *в* $\overline{u}$*, где* $e = q/\|q\|$.

Доказательство. Возьмем точку $x \in D$. Тогда по условию $[g, x - \overline{x}] \leq 0$. Кроме того, существуют точка $u \in U$ и число $t \in [0, 1]$, такие, что $x = u + t(z - u)$. Тогда $\|g, u + t(z-u) - \overline{u} - \overline{t}(z - \overline{u}\| = (1-t)(g, u - \overline{u}) + |t - \overline{t}|[g, z - \overline{u}] \leq 0$. Так как $\overline{x} = \overline{u} + \overline{t}(z - \overline{u})$, имеем $z - \overline{x} = (1 - \overline{t})(z - \overline{u})$ и $[g, z - \overline{u}] = 0$. Следовательно, $(1 - t)[\overline{g}, u - \overline{u}] \leq 0$ и $[\overline{g}, u - \overline{u}] \leq 0$. Так как x выбрано произвольно и каждому $x \in D$ соответствует $u \in U$, теорема доказана. □

Мы применяем к минимизации функции конического проектирования метод обобщенного градиентного спуска с расходящимся шагом (он может быть применен, так как $\Phi(u)$ квазивыпукла).

Алгоритм обобщенного градиентного спуска с коническим проектированием

Шаг 0. Определяется обобщенно-опорная гиперплоскость $U = \{x \in \mathbb{R}^n : [q, x] = \underline{q}\}$.

Выбирается z из внутренности D и u_0 из U. Полагаем $k := 0$.

Шаг 1. Вычисляется точка x_k, соответствующая u_k:

$$x_k = u_k + t_k(z - u_k),$$

$$t_k = \arg\min{}_{t \in I(u_k, z)} f(u_k + t(z - u_k)).$$

Шаг 2. В x_k находится обобщенно-опорный вектор g_k для функции f по отношению к D, такой, что $[g_k, z - u_k] = 0$.

Шаг 3. Полагаем

$$u_{k+1} = u_k - \lambda_k \frac{g_k - [g_k, e]e}{\|g_k - [g_k, e]e\|},$$

где $\lim_{k \to +\infty} \lambda_k = 0, \ \ \sum_{k=0}^{+\infty} \lambda_k = +\infty$.

Шаг 4. Полагаем $k := k+1$ и возвращаемся к Шагу 1.

Легко показать, что метод эквивалентен обобщенному градиентному спуску и потому сходится. Действительно, вектор $g_k - [g_k, e]e$ - обобщенно-опорный для множества $U_0(u_k)$ в точке u_k. Так как шаг выбирается из расходящегося ряда, по теореме из [92] мы имеем

$$\lim_{k\to+\infty} \hat{f}(x_k) = \min_{x\in D} f(x) = \lim_{k\to+\infty} \hat{\Phi}(u_k) = \min_{u\in U} \Phi(u),$$

$$\hat{f}(x_k) = \min_{1\le i\le k} f(x_i), \quad \hat{\Phi}(u_k) = \min_{1\le i\le k} \Phi(u_i).$$

Метод может применяться для минимизации линейной функции на выпуклом множестве. Тогда g_k можно выбирать как опорный вектор к допустимому множеству. Можно применить в методе ε-обобщенно-опорные вектора. Тогда при некоторых предположениях найдется ε-оптимальная точка.

В заключение главы рассмотрим еще одну возможность решения задачи квазивыпуклого программирования.

Пусть $D = \{x \in \mathbb{R}^n : F(x) \le 0\}$, где F – квазивыпуклая функция. Допустим, что известна точка $\overline{z} \in \mathbb{R}^n$, такая, что $f^* \le f(\overline{z}) \le f^* + \varepsilon$. Пусть также в точке $\overline{z}$ известен обобщенно-опорный вектор q к множеству $Q = \{x \in \mathbb{R}^n : f(x) \le f(\overline{z})\}$. Определим гиперплоскость

$$U = \{x \in \mathbb{R}^n : [q, x] = [q, \overline{x}], \quad \overline{x} \in \mathbb{R}^n\}.$$

Очевидно, она обобщенно-опорная к множеству Q (и, следовательно, к D^*), если выполняются условия

$$[q, \overline{x} - \overline{z}] < 0; \quad [q, y - \overline{x}] \ge 0 \quad \forall y \in Q.$$

Тогда мы можем ввести следующую вспомогательную функцию на U:

$$\Psi(u) = \min_{t\in I(u,z)} F(u + t(\overline{z} - u)),$$

где

$$I(u, z) = \{t \in [0, 1] : f(u + t(\overline{z} - u)) \le f^* + \varepsilon, \ \varepsilon > 0\}.$$

Множество $I(u, z) \supset \{1\}$ и поэтому непусто. Мы ставим следующую задачу безусловной оптимизации:

$$\min_{u\in U} \ \Psi(u), \tag{3.8}$$

Лемма 3.3 *Пусть* u^* *– решение задачи(3.8),* x^* *– соответствующая точка из множества* $\{x \in \mathbb{R}^n : f(x) \leq f^* + \varepsilon\}$*, то есть* $x^* = u^* + t^*(\overline{z} - u^*), t^* \in I(u^*, \overline{z}), t^* = arg\ min_{t \in I(u^*, z)} F(u^* + t(z - u^*))$*. Тогда* $x^* \in D^*_\varepsilon = \{x \in D : f(x) \leq f^* + \varepsilon\}$.

Доказательство. По построению, так как x^* – соответствующая точка, мы имеем $f(x^*) \leq f^* + \varepsilon$. Это следует из определения множества $I(u^*, \overline{z})$ и определения соответствующих точек. Более того, полагая D и D^* непустыми, мы имеем

$$\min \quad F(x) \leq 0$$

$$x \in \mathbb{R}^n, \ f(x) \leq f^* + \varepsilon.$$

Задача (3.8) эквивалентна минимизации функции максимума из ограничений на множестве точек, для которых $f(x) \leq f^* + \varepsilon$. Решив ее, получим ε-оптимальную точку исходной задачи. Если f^* неизвестно, можно использовать любое число и получить ε-оптимальную точку уменьшением оценки f^*. $\square$

Глава 4

Решение задачи о гамильтоновом цикле через марковские цепи и функции типа минимума

4.1 Формулировка

Мы рассматриваем задачу о гамильтоновом цикле (ЗГЦ) с несколько неортодоксальной точки зрения, основанной на включении в марковский процесс решения (МПР), разработанном Филаром и Крассом в [40] и Ченом и Филаром в [41].

В этой части мы даем формулировку ЗГЦ в виде задачи глобальной оптимизации, которая использовалась для численных экспериментов. Следует отметить, что результаты монографии стимулировались более ранними попытками Филара, Оберие и Пардалоса ([43]).

Формулировка ЗГЦ, анализируемая здесь, следующая: пусть дан ориентированный граф G с N вершинами; найти в нем простой цикл из N вершин, то есть гамильтонов цикл, или определить, что он не существует.

Вряд ли целесообразно представлять полную библиографию по этой классической задаче. Вместо этого мы отсылаем читателя к монографии [71].

В данной работе ЗГЦ сформулирована как нахождение глобального минимума (со значением целевой функции, равным 0) надлежащим образом построенной задачи квадратичного программирования:

$$\min \;\; x^{'}Qx, \tag{4.1}$$

$$Ax = b, \tag{4.2}$$

$$x \geq 0. \tag{4.3}$$

Часть 4.2 мотивирует получение данной задачи, доказательство эквивалентности которой ЗГЦ подробно изложено в [40]. Для нас более удобно непосредственно описать структуру матриц Q, A и вектора b, не входя в детали их происхождения.

Обозначения. Допустим, что граф G не имеет петель, и пусть $A(i)$ – множество ребер, выходящих из вершины i. Предположим, что

$$n_i = |A(i)| \geq 1 \quad \forall i \in E = \{1, 2, \dots, N\},$$

то есть существует хотя бы одно ребро, выходящее из каждой вершины, иначе ЗГЦ была бы тривиальной. Ребро, выходящее из вершины i, образует упорядоченную пару (i, a).

Когда очевидно, каково начало ребра, мы будем обозначать его только по его концу. Например, если текущая вершина i и рассматривается ребро (i, a), оно будет обозначаться просто a. Пусть $\varepsilon \in (0, 1)$ фиксировано. Определим

$$d_N(\varepsilon) = 1 + \sum_{k=2}^{N-2} (1-\varepsilon)^{k-2}.$$

Далее, для любых $i, j \in E$ и $a \in A(i)$ определим коэффициенты $p_{ija}(\varepsilon)$:

$$p_{ija}(\varepsilon) = \begin{cases} 1, & \text{если } i = 1 \text{ и } a = j; \\ 0, & \text{если } i = 1 \text{ и } a \neq j; \\ 1, & \text{если } i > 1 \text{ и } a = j = 1; \\ \varepsilon, & \text{если } i > 1, a \neq j, \text{ и } j = 1; \\ 1-\varepsilon, & \text{если } i > 1, a = j, \text{ и } j > 1; \\ 0, & \text{если } i > 1, a \neq j, \text{ и } j > 1. \end{cases}$$

Интерпретация этих коэффициентов дана ниже в разделе 4.2.

Матрица A и вектор b из (4.3) определяются следующей системой линейных ограничений:

$$\begin{array}{ll} (C1) & \sum_{i \in E} \sum_{a \in A(i)} (\delta_{ij} - p_{ija}(\varepsilon)) x_{ia} = 0, \ j \in E; \\ (C2) & \sum_{i \in E} \sum_{a \in A(i)} x_{ia} = 1; \\ (C3) & \sum_{a \in A(i)} x_{ia} = 1/d_N(\varepsilon); \\ (C4) & x_{ia} \geq 0; \ i \in E, a \in A(i). \end{array}$$

Так, A есть m-мерная матрица $m = (N+2)\times(\sum_{i=1}^{N} n_i)$ и $b^{'} = (0, 0, \ldots, 1, 1/d_N(\varepsilon))$ – $(N+2)$-мерный вектор. Здесь δ_{ij} – символ Кронекера. Матрица Q есть $(\sum_{i=1}^{N} n_i)\times(\sum_{i=1}^{N} n_i)$-мерная блочно-диагональная матрица, i-й диагональный блок которой есть $n_i \times n_i$ матрица $Q_i = J_i - I_i$, где J_i – матрица из одних единиц и I_i – единичная матрица соответствующей размерности.

В соответствии с теоретическими результатами раздела 4.2, если x удовлетворяет условиям $(C1)-(C4)$ и $x^{'}Qx = 0$, то положительные компоненты x определяют гамильтонов цикл в G.

Пример 1 взят из ([41]). Рассмотрим полный граф G с четырьмя вершинами (без петель). При $\varepsilon = 0.1$ элементы задачи квадратичного программирования (4.3) следующие:

$$\begin{pmatrix} Q_1 & 0 & 0 & 0 \\ 0 & Q_2 & 0 & 0 \\ 0 & 0 & Q_3 & 0 \\ 0 & 0 & 0 & Q_4 \end{pmatrix},$$

где для $i = 1, 2, 3, 4$ имеем

$$Q_i = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Вектор $b^{'} = (0, 0, 0, 0, 1, 0.2695)$ и матрица коэффициентов A при данном значении ε есть

$$\begin{bmatrix} 1 & 1 & 1 & -1 & -0.1 & -0.1 & -1 & -0.1 & -0.1 & -1 & -0.1 & -0.1 \\ -1 & 0 & 0 & 1 & 1 & 1 & 0 & -0.9 & 0 & 0 & -0.9 & 0 \\ 0 & -1 & 0 & 0 & -0.9 & 0 & 1 & 1 & 1 & 0 & 0 & -0.9 \\ 0 & 0 & -1 & 0 & 0 & -0.9 & 0 & 0 & -0.9 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Легко проверить, что глобальный минимум задачи квадратичного программирования, приведенной выше, достигается в точке

$$x^{'} = (0.2695, 0, 0, 0, 0.2695, 0, 0, 0, 0.2425, 0.2183, 0, 0),$$

положительные компоненты которой определяют гамильтонов цикл: $1 \to 2 \to 3 \to 4 \to 1$.

Конечно, задача глобальной оптимизации (4.3), в принципе, очень трудная задача, так как Q – неопределенная матрица. Мы используем для решения (4.3) функции типа минимума (см. Главы 1-2). Заметим, что для x, допустимой для квадратичной задачи, имеем

$$x^{'}Qx = 0 \tag{4.4}$$

тогда и только тогда, когда

$$\sum_{i=1}^{N} \sum_{a \neq b} x_{ia} x_{ib} = 0, \tag{4.5}$$

что выполняется, только если

$$\sum_{i=1}^{N} \min\{x_{ia}, x_{ib}\} = 0. \tag{4.6}$$

Поэтому задача определения, равен ли глобальный минимум (4.3) нулю, может быть сведена к следующей задаче глобальной оптимизации:

$$\min_{x} \ \max_{i} \ \min_{a \neq b} \ \ \min\{x_{ia}, x_{ib}\} \tag{4.7}$$

при ограничениях $(C1) - (C4)$.

В общем случае решение задачи (4.7) трудно с алгоритмической точки зрения. Однако в нашем случае (4.7) можно решить нахождением любого допустимого решения задачи частично целочисленного программирования:

$$\min \ \sum_{i=1}^{N} \sum_{a \in A(i)} c_{ia} x_{ia}; \tag{4.8}$$

$$(C1) - (C4); \tag{4.9}$$

$$\begin{array}{ll} (C5) & x_{ia} \le M y_{ia} \quad i \in E, a \in A(i); \\ (C6) & y_{ia} + y_{ib} \le 1, \quad i \in E, a, b \in A(i), a \neq b; \\ (C7) & y_{ia} \in \{0,1\}; \quad i \in E, a \in A(i), \end{array}$$

где $M \ge 1/d_N(\varepsilon)$ и c_{ia} – некоторые числа, не все равные нулю.

Во многих случаях формулировка задач с помощью функций типа минимума предпочтительнее квадратичного программирования, так как сохраняются некоторые свойства линейности. Негладкость целевой функции не является недостатком, так как производные полезны в локальной оптимизации, а для решения ЗГЦ нужна глобальная. В следующей части мы покажем, что при надлежащем выборе ε, M и c_{ia} задача (4.9) быстро решается для задач умеренного размера.

Следует отметить, что коэффициент c_{ia} можно рассматривать как стоимость транспортировки вдоль ребра ia. Поэтому задачу коммивояжера (см. [63]) можно решить, использовав модель (4.9). Действительно, переменная y_{ia} равна единице, если ребро ia включается в маршрут, и нулю в противном случае. Таким образом, ограничения задают гамильтонов цикл, а целевая функция дает его стоимость, так что мы переформулировали задачу коммивояжера.

4.2 Включение ЗГЦ в марковский процесс решения

В настоящем разделе мы следуем подходу Филара и Красса (см. [40]) и Чена и Филара (см. [41]) к ЗГЦ.

Рассмотрим объект, движущийся по направленному пути графа G, причем его движение "контролируется" функцией f, отображающей вершины N в ребра A. Эта функция индуцирует $N \times N$-мерную марковскую матрицу $P(f)$ со значениями 0-1, положительные значения которой соответствуют ребрам, выбранным f в соответствующих вершинах. Предположим, что движение продолжается бесконечно, и рассмотрим $P(f)$ как марковскую цепь с ее стационарным распределением, содержащимся в матрице пределов сумм Чезаро:

$$P^*(f) := \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} P^{t-1}(f),$$

где $P^0(f) := I_N$.

В [40] и [41] изучалось отношение между эргодическими классами и временными состояниями таких марковских цепей и возможными циклами в графе.

Чтобы уточнить рассуждения, сделанные выше, формально введем МПР состояния-действия как кортеж $\Gamma = \{E, A, r, p\}$, где

$$E = \{1, 2, \ldots, N\}$$

– множество *состояний*, $A = \cup_i A(i)$ с $A(i) = \{1, 2 \ldots, n_i\}$ – множество *действий*, возможных в состоянии i для каждого $i \in E, r = \{r(i,a)|a \in A(i), i \in S\}$ – множество возможных (непосредственных) премий и $p = \{p_{iaj}|a \in A(i), i, j \in S\}$ – множество (одношаговых) вероятностей перехода. *Стационарная политика* f в Γ есть множество N вероятностных векторов

$$f(i) = (f(i,1), f(i,2), \ldots, f(i,n_i)),$$

где $f(i,k)$ означает вероятность выбора действия k в состоянии i, когда оно встречается. Обозначим множество всех стационарных политик через $C(S)$. *Детерминистская политика* f – это политика, при которой единственное действие выбирается с вероятностью 1 в каждом состоянии, и мы пишем $f(i) = k$ для $i \in E$.

Для любых f, начального распределения $\gamma, j \in E$ и $a \in A(j)$ определим

$$x'_{ja}(f) = \frac{1}{T}\sum_{i=1}^{T}\sum_{i=1}^{N}\gamma_i P(f)(X_i = j, Y_i = a|X_1 = i). \qquad (4.10)$$

Далее обозначим через $X(f)$ множество предельных точек векторов $\{x'(f)|T = 1, 2, \ldots\}$, где $x'(f)$ есть $\sum_{i=1}^{N}|A(i)|$-мерный вектор с компонентами, определяемыми (4.10). Если $X(f) = \{x'(f)\}$ – одна точка, то значения $x_{ia}(f)$ из $x(f)$ могут быть интерпретированы как *долговременные ожидаемые частоты состояния-действия*, порождаемые f. Подобным образом долговременные ожидаемые частоты нахождения в состоянии $j \in E$ при f даются формулой

$$x_j(f) = \sum_{a \in A(j)} x_{ja}(f). \qquad (4.11)$$

МПР называется процессом с одной цепью, если для любой детерминистской политики f марковская цепь, индуцированная $P(f)$, имеет один эргодический класс плюс (возможно пустое) множество временных состояний.

Рассмотрим следующую задачу линейного программирования (LP1):

$$\max \sum_{i \in A}\sum_{a \in A(i)} r_{ia}x_{ia}, \qquad (4.12)$$

$$
\begin{array}{ll}
(L) & \sum_{i\in E}\sum_{a\in A(i)}(\delta_{ij}-p_{iaj})x_{ia}=0,\;\; j\in E;\\
(L2) & \sum_{i\in E}\sum_{a\in A(i)}x_{ia}=1;\\
(L3) & x_{ia}\geq 0,\; i\in E, a\in A(i),
\end{array}
$$

где δ_{ij} – символ Кронекера. Пусть X – допустимое множество данной задачи и $C(S)$ обозначает класс стационарных стратегий МПР с одной цепью. Рассмотрим отображение $T : X \to C(S)$, где $T(x) = f_x$ определяется как

$$
f_x(i,a)=\begin{cases} x_{ia}/x_i, & \text{если } x_i=\sum_{a\in A(i)}x_{ia}>0;\\ 1, & \text{если } x_i=0 \text{ и } a=a_1;\\ 0, & \text{если } x_i=0 \text{ и } a\neq a_1\end{cases}
$$

для каждого $i \in E$ и $a \in A(i)$, где a_1 означает первое возможное действие в данном состоянии согласно некоторому упорядочению. Рассмотрим также отображение $\hat{T} : C(S) \to X$, где $\hat{T}(f) = x(f)$ определено в соответствии с (4.10) и (4.11) как

$$
x_{ia}(f)=p^*(f)f(i,a),\;\; i\in E, a\in A(i). \tag{4.13}
$$

В данном равенстве $p_i^*(f)$ – i-я компонента единственного вектора фиксированных вероятностей (вектора стационарного распределения) $P(f)$. Преобразования T и $\hat{T}$ изучались рядом авторов (см., например, [36], [59]). Рассмотрим теперь ориентированный взвешенный граф G с множеством вершин $V = \{1, \ldots, N\}$, множеством ребер A и весами c_{ij}, связанными с ребрами (i, j). Пусть $G_1 \subseteq G$ – ориентированный подграф G с тем же множеством вершин V и множеством ребер $A_1 = \{(i, j)|$ для каждого $i \in V$ существует только один $j \in V$, такой, что одно ребро выходит из каждой вершины $G_1\}$. Первый МПР, который мы ассоциируем с G, – процесс $\Gamma = \{E, A, R, p\}$, $E = \{1, 2, \ldots, N\}$ – множество вершин $G, A(i) = \{j \in E|(i, j) \in A\}$ для каждого $i \in E$ и $A = \cup_{i=1}^N A(i), r = \{r(i, j) = -c_{ij}|j \in A(i), i \in E\}$, и $p = \{p(j|i, a)|a \in A(i), i, j \in E\}$ с $p(j|i, a) = \delta_{ij}$. Допустим, что 1 – начальное состояние. Будем говорить, что детерминистская политика f в Γ порождает гамильтонов цикл, если подграф G_1 с множеством ребер $\{(1, f(1)),(2, f(2)), \ldots,(N, f(N))\}$ есть гамильтонов цикл в G. Если подграф G_1 имеет циклы длиной меньше N, будем говорить, что f имеет *подцикл* в G. Если подграф G_1 содержит цикл длины k, мы говорим, что

f имеет k-подцикл. Сказанное выше можно проиллюстрировать на полном графе из четырех вершин без петель. Например, политика f, такая, что $f(1) = 2, f(2) = 1, f(3) = 4$ и $f(4) = 3$, индуцирует следующий подграф $G_f = \{(1,2),(2,1),(3,4),(4,3)\}$, содержащий два 2-подцикла. Заметим, что f индуцирует также марковскую цепь с матрицей вероятностей перехода

$$P(f) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

которая имеет два эргодических класса, соответствующих подциклам в G_f.

Вообще любая стационарная политика $\pi \in C(S)$ порождает матрицу вероятностей перехода

$$P(f) = p_{ij}(f),$$

где для всех имеем $i, j \in S$

$$p_{ij}(f) = \sum_{a \in A(i)} p(i,a,j)\pi_{i,a}.$$

Если для каждого $F \in C(S)$ марковская цепь $P(f)$ содержит только один эргодический класс (плюс, возможно, временные состояния), МПР графа Γ называется МПР *с одной цепью*. По ряду технических причин МПР с одной цепью легче анализировать. Мы видели из предыдущего примера, что прямое включение G в Γ индуцирует эргодическую структуру со многими цепями. Эта и другие технические трудности исчезли бы, если бы Γ был МПР с одной цепью. Ввиду указанного выше, в [40] и [41] закон движения в Γ был возмущен и были введены вероятности $p(\varepsilon) = p_{iaj}(\varepsilon)|(i,a,j) \in S \times A(i) \times S\}$, где для любого $\varepsilon \in (0,1)$ мы определяем

$$p_{iaj}(\varepsilon) = \begin{cases} 1, & \text{если } i = 1 \text{ и } a = j, \\ 0, & \text{если } i = 1 \text{ и } a \neq j, \\ 1, & \text{если } i > 1 \text{ и } a = j = 1, \\ \varepsilon, & \text{если } i > 1, a \neq j, \text{ и } j = 1, \\ 1-\varepsilon, & \text{если } i > 1, a = j, \text{ и } j > 1, \\ 0, & \text{если } i > 1, a \neq j, \text{ и } j > 1. \end{cases}$$

Заметим, что 1 означает начальную вершину. Для каждой пары вершин i, j (не равных 1), соответствующих (детерминистскому) ребру (i, j), наше возмущение заменяет это ребро парой "стохастических ребер" $(i, 1)$ и (i, j) с весами ε и $1-\varepsilon$ соответственно ($\varepsilon \in (0, 1)$). Стохастическое возмущение имеет ту интерпретацию, что решение двигаться вдоль ребра (i, j) принимается с вероятностью $1-\varepsilon$ и с вероятностью ε мы возвращаемся в начальную вершину 1. Заметим, что ε-возмущенный процесс $\Gamma(\varepsilon) = \{E, A, r, p(\varepsilon)\}$, очевидно, стремится к Γ, когда $\varepsilon \to 0$. Этот процесс имеет следующие свойства, которые можно найти в [41].

Лемма 4.1 *1) МПР $\Gamma(\varepsilon)$ является МПР с одной цепью.*

2) Рассмотрим марковскую цепь, индуцированную стационарной политикой f на $\Gamma(\varepsilon)$, и пусть $E_1 \subseteq E$ – эргодический класс этой цепи. Тогда 1 $\in E_1$.

Заметим, что с каждой $f \in C(D)$ мы можем ассоциировать подграф G_f графа G, определенный как

$$\text{ребро}(i, a) \in G_f \Leftrightarrow f(i) = a.$$

Следующая теорема была доказана Ченом и Филаром [41].

Теорема 4.1 *Пусть Q – блочно-диагональная матрица из раздела 5.1 и пусть $M(\varepsilon)$ – многогранное множество, определенное соотношениями $(C1)-(C4)$. Тогда справедливы следующие два утверждения:*

1. Предположим, что $f \in C(P)$ соответствует гамильтонову циклу в G. Тогда $x(f)$ – глобальный оптимум задачи квадратичного программирования

$$\min\{x^{'}Qx | x \in M(\varepsilon)\} \tag{4.14}$$

и $x^{'}(f)Qx(f) = 0$.

2. Обратно, пусть глобальный оптимум задачи квадратичного программирования равен 0, и предположим, что он достигается в x^*. Тогда $f_{x^*} = T(x^*)$ – детерминистская политика, соответствующая гамильтонову циклу в G.

Замечание 4.2.1 *Из предыдущей теоремы следует, что гамильтоновы циклы в G характеризуются глобальными минимумами задачи квадратичного программирования с неопределенной матрицей. Если глобальный минимум в (4.3) положителен, гамильтонова цикла в G нет.*

4.3 Численные эксперименты

Чтобы проверить применимость к ЗГЦ подхода, основанного на марковских цепях, был решен ряд тестовых примеров. Все эксперименты проводились с помощью программы CPLEX 4.0 для частично целочисленного программирования и модели с функциями типа минимума. Программы были написаны на языке программирования C++ и выполнены в операционной системе AIX 3.2.

Тестовые примеры генерировались обычно одним из двух способов: или гамильтонов цикл был задан и больше ребер было добавлено к каждой вершине, чтобы получить граф, содержащий гамильтонов цикл, или фиксированное число ребер генерировалось для каждой вершины, образуя регулярный граф. Кроме того, старая задача обойти ходом шахматного коня доску, побывав в каждой клетке ровно по одному разу и вернувшись в начальную позицию, была взята как тестовый пример. Несколько примеров, не содержащих гамильтоновых циклов, также были рассмотрены.

Максимальное число вершин в графе было равно 100 и максимальное число ребер – 400. Важно отметить, что для нахождения гамильтонова цикла нужно было явно указывать целевую функцию, чтобы облегчить процедуру ветвей и границ программы CPLEX. Если примеры решались с целевой функцией, равной нулю, решения были найдены только для очень малых размерностей. Однако для нетривиальных целевых функций гамильтоновы циклы для графов со 100 вершинами были найдены.

Численные эксперименты показали, что наилучший выбор целевой функции был следующий:

$$f(x, y) = \sum_{i=1}^{N} \sum_{a \in A(i)} c_{ia} y_{ia},$$

где N – число вершин и c_{ia} – положительные числа, для которых разности

Таблица 4.1: Результаты для произвольных графов

Число вершин	Число ребер	Число вершин дерева	Время вычислений(с)
15	50	75	0.05
20	100	125	2.5
35	100	573	15.5
50	200	875	45.5
50	300	1254	50.0
60	400	2182	122.5
60	300	1728	87.5
80	200	2029	63.0
100	200	886	121.5
100	300	1117	145.5

$c_{i_1a_1} - c_{i_2a_2}$ велики для любой пары разных ребер $(i_1, a_1), (i_2, a_2)$. Например, во многих случаях было выгодно задавать $c_j = j^3$, где индекс j соответствует порядковому числу ребра j среди всех ребер в графе, отсортированных некоторым образом.

Для некоторых примеров допустимое решение было найдено несколько быстрее, если было добавлено ограничение

$$(C8) \qquad\qquad \textstyle\sum_{a \in A(i)} y_{ia} = 1, \quad \forall i \in E,$$

хотя среднее время вычисления не изменилось.

Наилучшие результаты не были получены со стандартными параметрами программы CPLEX. Нужно было использовать ветвление по переменной с наименьшей недопустимостью или по оценкам задач линейного программирования (в сочетании с поиском по наилучшей оценке целевой функции).

Одна из важнейших численных проблем в предложенном подходе – выбор параметра ε. Для малых размерностей (10 - 25) было выгодно выбирать ε между 0.1 и 0.25. Для больших размерностей нужно было постепенно уменьшать ε, чтобы гарантировать сходимость. Так, для $N = 60$ наилучшее значение ε было равно 0.07 и для $N = 100$ оно было равно 0.04. В большинстве случаев было легче найти гамильтонов цикл, чем обнаружить, что он не

Таблица 4.2: Результаты для регулярных графов

Число вершин	Число ребер	Число вершин дерева	Время вычислений(с)
10	50	13	0.05
20	100	434	4.5
20	200	247	8.5
40	200	518	14.0
50	200	715	24.0
50	500	1038	38.0
60	300	1690	75.0
80	400	2478	139.5
100	200	1072	79.0
100	300	2593	122.0

Таблица 4.3: Ходом шахматного коня, доска 6×6

4	15	34	27	6	17
35	26	5	16	33	28
12	3	14	29	18	7
25	36	11	32	21	30
10	13	2	23	8	19
1	24	9	20	31	22

существует, вероятно по той причине, что все дерево решений должно быть обследовано, чтобы доказать недопустимость задачи. Однако для простых случаев, таких, как наличие вершины, у которой имеется только одна соседняя, недопустимость была обнаружена очень быстро (анализировалось только 50 - 100 вершин дерева).

Рассмотрим конкретный пример со случайно сгенерированным графом из 10 вершин. Матрица инцидентности графа дана в Таблице 5. Гамильтонов цикл

$$1 \to 4 \to 7 \to 2 \to 5 \to 6 \to 8 \to 3 \to 9 \to 10 \to 1$$

Таблица 4.4: Ходом шахматного коня, доска 8×8

34	47	20	61	32	49	18	59
21	62	33	48	19	60	31	50
46	35	4	13	8	29	58	17
63	22	7	10	5	16	51	30
36	45	14	3	12	9	28	57
23	64	11	6	15	54	41	52
44	37	2	25	42	39	56	27
1	24	43	38	55	26	53	40

Таблица 4.5: Пример

0	1	0	1	1	1	0	0	0	0
1	0	0	0	1	1	0	0	0	0
1	1	0	0	0	0	0	1	1	1
1	1	1	0	1	1	1	1	0	0
1	1	0	0	0	1	0	1	1	0
0	1	1	1	0	0	0	1	1	1
1	1	1	0	1	0	0	1	1	1
1	1	1	0	1	0	1	0	1	0
1	0	1	0	1	1	1	1	0	1
1	1	0	0	1	1	1	0	1	0

был найден после анализа 1613 вершин дерева, что потребовало 12 секунд. Наилучшие результаты для данной задачи были получены при ветвлении, основанном на оценках задач линейного программирования.

В Таблице 1 мы даем типичные результаты для произвольных графов и в Таблице 2 – для регулярных графов.

Результаты для регулярных графов были значительно лучше по скорости вычислений. В первой колонке приведено число вершин, и во второй – число ребер. В третьей колонке мы даем число вершин дерева программы CPLEX, которые были исследованы для нахождения первого допустимого решения (дающего гамильтонов цикл). Наконец, в последней колонке приведено время вычисления в секундах. Заметим, что общее время вычислений менялось от 0.02 минуты для графов с 10 вершинами до 20 минут для некоторых графов со 100 вершинами.

В Таблицах 3 - 4 мы даем ответы в задаче об обходе доски ходом шахматного коня для досок размером 6×6 и 8×8. В каждой клетке приведен ее номер в маршруте. Данная задача вначале была самым трудным тестом. Однако, используя технику декомпозиции (находя гамильтоновы подциклы и соединяя их), удалось решить задачу достаточно быстро . Общее число просмотренных вершин дерева решений изменялось от 2000 до 40000, и время вычисления было между одной и пятнадцатью минутами, в зависимости от параметров программы CPLEX.

Наши основные выводы следующие. Подход, основанный на марковских цепях и функциях типа минимума, применим к решению задачи о гамильтоновом цикле для графов умеренной размерности, особенно для регулярных графов. Необходимо приспосабливать процедуру ветвей и границ к каждому специфическому виду задачи (наиболее важен выбор направления движения по дереву решений и переменной, по которой идет ветвление). Необходима нетривиальная целевая функция для получения хороших оценок при решении задачи частично целочисленного программирования. Для решения практических задач наиболее целесообразно использовать простую технику декомпозиции, генерируя подциклы и затем соединяя их.

Глава 5

Подход к построению обобщенных штрафных функций

5.1 Обобщенные штрафные функции для ограничений в форме неравенств

Рассмотрим следующую задачу условной минимизации:

$$\begin{gathered} \min \;\; f_0(x), \\ f_i(x) \le 0, \quad i = \overline{1,m}, \quad x \in X, \end{gathered} \tag{5.1}$$

где f_i – непрерывные функции, определенные на $\mathbb{R}^n$, и X – замкнутое множество в $\mathbb{R}^n$. Обозначим множество оптимальных точек через X^* (предположим, что оно непусто) и минимальное значение целевой функции на допустимом множестве через f^*. Предположим, что f_0 положительна. Тогда значение f^* также положительно (мы не предполагаем, однако, что f^* известно). Будем решать вместо (5.1) следующую вспомогательную задачу:

$$\begin{gathered} \min \;\; g(x,d), \\ x \in X, \end{gathered} \tag{5.2}$$

где $g(x,d)$ – функция, определенная на $\mathbb{R}^n$, которая зависит от векторного параметра $d \in \mathbb{R}^{m+1}_+$. Именно, предположим, что $g(x,d)$ имеет вид

$$g(x,d) = G(d_0 f_0(x), d_1 f_1(x), \dots, d_m f_m(x)),$$

где G – непрерывная функция, определенная на $\mathbb{R}^{m+1}$. Заметим, что эта общая схема для одного ограничения была рассмотрена в [3], где функция

g называлась *точной вспомогательной функцией.* Однако мы не предполагаем, что седловая точка функции Лагранжа

$$L(x,\lambda) = f_0 + \sum_{i=1}^{m} \lambda_i f_i(x)$$

существует, и наши условия эквивалентности исходной и вспомогательной задач применимы к гораздо более широкому классу задач. Приведенная схема ранее была рассмотрена в более общей форме в [88, 89], где был предложен другой подход.

Пусть выполняются следующие предположения:

1) $G(v, y_1, \ldots, y_m) = v$ для всех $v \geq 0$ и для всех y, таких, что $y_i \leq 0, \quad i = \overline{1, m}$;

2) $G(v, y_1, \ldots, y_m) \to +\infty$, когда $v \geq 0$, и $\max_{i=\overline{1,m}} y_i \to +\infty$;

3) $G(y)$ – возрастающая функция на $\mathbb{R}_+^{m+1}$.

Приведем примеры вспомогательных функций, удовлетворяющих этим условиям:

1) $G_1(y) = y_0 + \sum_{i=1}^m \max\{y_i, 0\}$;

2) $G_2(y) = (y_0^p + \sum_{i=1}^m (\max\{y_i, 0\})^p)^{1/p}; \quad (p > 0)$

3) $G_3(y) = \max\{y_0, y_1, \ldots, y_m\}$;

4) $G_4(y) = y_0 + \sum_{i=1}^m (\exp(\max\{y_i, 0\}) - 1)$;

5) $G_5(y) = \max\{y_0, \exp(\max\{y_1, 0\}) - 1, \ldots, \exp(\max\{y_m, 0\}) - 1\}$;

Функции первого типа очень хорошо известны и изучались многими авторами (см., например, [39]). Функции второго типа были рассмотрены в [3], где требовалось существование седловой точки функции Лагранжа. Функции третьего типа позволяют построить метод центров, вначале предложенный без штрафных параметров в [56] и затем изученный в [8].

Предположим, что множество оптимальных решений (5.2) непусто (например, лебеговы множества $g(x, d)$ ограничены). Обозначим через $x^*(d)$ любое оптимальное решение (5.2) (конечно, оно зависит от параметра d).

Лемма 5.1 *Для любого $\varepsilon > 0$ существует $M > 0$, такое, что из $d_0 = 1$, $\min_{i=\overline{1,m}} d_i > M$ следует $f_i(x^*(d)) \le \varepsilon$ для всех $i = \overline{1,m}$.*

Доказательство. Предположим, что это неверно. Тогда существует $\overline{\varepsilon} > 0$, такое, что для любого $M > 0$ существует вектор $\overline{d} = \overline{d}(M) = (\overline{d}_1, \overline{d}_2, \ldots, \overline{d}_m)$, для которого $\overline{d}_i > M$ для всех $i = \overline{1,m}$ и

$$f_j(x^*(\overline{d})) > \overline{\varepsilon} \quad \exists j \in \{1, 2, \ldots, m\}.$$

Но

$$g(x^*(\overline{d}), \overline{d}) = G(f_0(x^*(\overline{d})), d_1 f_1(x^*(\overline{d})), \ldots, d_m f_m(x^*(\overline{d}))).$$

Так как

$$\max_{i=\overline{1,m}} d_i f_i(x^*(\overline{d})) \ge d_j f_j(x^*(\overline{d})) \ge M\overline{\varepsilon}$$

и M может быть выбрано сколь угодно большим, можно использовать Предположение 2, из которого следует, что для некоторого $M > 0$ будет выполнено неравенство $g(x^*(\overline{d}), \overline{d}) > f^*$. Но, с другой стороны,

$$g(x^*(\overline{d}), \overline{d})) \le g(x^*, \overline{d}), \ \ \forall\, x^* \in X^*,$$

так как $X^* \subseteq X$. Однако

$$g(x^*, \overline{d}) = G(f_0(x^*), \overline{d}_1 f_1(x^*), \ldots, \overline{d}_m f_m(x^*)) = f_0(x^*) = f^*$$

по Предположению 1 и в силу того, что x^* – допустимая точка. Получили противоречие, и лемма доказана. □

Лемма показывает, что когда мы увеличиваем значения параметров d_i, которые можно рассматривать как штрафные коэффициенты, оптимальные точки вспомогательной задачи стремятся к допустимому множеству.

Введем следующие точечно-множественные отображения:

$$A(\varepsilon) = \{x \in X : f_i(x) \le \varepsilon \quad \forall i = \overline{1,m}\},$$

$$D(\varepsilon) = \{x \in A(\varepsilon) : f^* \ge f_0(x) \ge f^* - \varepsilon\}.$$

Заметим, что $A(0)$ – допустимое множество (5.1) и $D(0) = X^*$. Пусть B – единичный шар в $\mathbb{R}^n$.

Теорема 5.1 *Пусть отображение A полунепрерывно сверху в нуле. Тогда для любого $\varepsilon > 0$ существует $\overline{d} = (1, \overline{d}_1, \overline{d}_2, \ldots, \overline{d}_m) \in \mathbb{R}_+^{m+1}$, такое, что $f_0(x^*(\overline{d})) \ge f^* - \varepsilon$.*

Доказательство. В силу полунепрерывности сверху, для любого $\gamma > 0$ существует $\Delta > 0$, такое, что

$$A(\lambda) \subseteq A(0) + \gamma B \quad \forall \lambda \in [0, \Delta].$$

По Лемме 5.1, когда значение $\min_{i=\overline{1,m}} \overline{d}_i$ достаточно велико, мы имеем $x^*(\overline{d}) \in A(\Delta)$. Тогда $x^*(\overline{d}) \in A(0) + \gamma B$ и существует точка $\tilde{x} \in A(0)$, такая, что $\|x^*(\overline{d}) - \tilde{x}\| < \gamma$. Но $f_0(\tilde{x}) \geq f^*$, так как $\tilde{x}$ допустима. Значение γ может быть выбрано сколь угодно малым, и в силу непрерывности f_0 теорема доказана. $\square$

Лемма 5.2 *Оптимальное значение задачи (5.2) не превосходит оптимального значения задачи (5.1) для любого $d \in \mathbb{R}_+^{m+1}$, такого, что $d_0 = 1$.*

Доказательство. По определению,

$$\min_{x \in X} g(x^*(d), d) = G(f_0(x^*(d)), d_1 f_1(x^*(d)), \ldots, d_m f_m(x^*(d))),$$

где $x^*(d)$ – оптимальное решение (5.2), соответствующее параметру d. Пусть x^* – произвольная точка из X^*. Тогда

$$G(f_0(x^*(d)), d_1 f_1(x^*(d)), \ldots, d_m f_m(x^*(d))) \leq$$

$$\leq G(f_0(x^*), d_1 f_1(x^*), \ldots, d_m f_m(x^*)).$$

Но значение в правой части равно $f_0(x^*) = f^*$ по Предположению 1 и в силу свойства $x^* \in X$. Лемма доказана. $\square$

Теорема 5.2 *Предположим, что отображение D полунепрерывно сверху в нуле. Тогда для любого $\gamma > 0$ существует $\overline{d} = (1, \overline{d}_1, \ldots, \overline{d}_m) \in \mathbb{R}_+^{m+1}$, такое, что $\|x^* - x^*(\overline{d})\| \leq \gamma$ для некоторого $x^* \in X^*$.*

Доказательство. Для любого $\gamma > 0$ существует $\Delta > 0$, такое, что

$$D(\lambda) \subseteq D(0) + \gamma B \ \ \forall \lambda \in [0, \Delta].$$

Если значение $\min_{i=\overline{1,m}} \overline{d}_i$ достаточно велико, мы имеем по Теореме 5.1 и Лемме 5.2 $x^*(\overline{d}) \in D(\Delta)$. Тогда $x^*(\overline{d}) \in D(0) + \gamma B$, то есть $x^*(\overline{d}) \in X^* + \gamma B$, что означает $\|x^*(\overline{d}) - x^*\| \leq \gamma$ для некоторого $x^* \in X^*$. Теорема доказана. $\square$

Замечание 5.1 *Вместо рассмотрения A как отображения, зависящего от одного скалярного параметра, можно рассмотреть следующее отображение:*

$$\tilde{A}(\varepsilon) = \{x \in X : f_i(x) \leq \varepsilon_i \ \ \forall i = \overline{1,n}\},$$

где $\varepsilon = \{\varepsilon_1, \dots, \varepsilon_m\} \in \mathbb{R}^m$. Полунепрерывность сверху данного отображения также достаточна для выполнения Теоремы 5.1.

5.2 Ограничения-равенства

Рассмотрим задачу условной минимизации с ограничениями в виде равенств:

$$\begin{gathered} \min \ \ h_0(x), \\ h_i(x) = 0, \quad i = \overline{1,m} \quad x \in X, \end{gathered} \tag{5.3}$$

где h_i – непрерывные функции, определенные на $\mathbb{R}^n$ и X – замкнутое множество в $\mathbb{R}^n$. Предположим, что допустимое множество (5.3) непусто и h_0 положительна. Обозначим множество оптимальных точек задачи через Q^* (допустим, что оно непусто) и минимальное значение целевой функции на допустимом множестве через h^*.

Аналогично разделу 6.1 мы решаем вместо (5.3) следующую вспомогательную задачу:

$$\begin{gathered} \min \ \ q(x,d), \\ x \in X, \end{gathered} \tag{5.4}$$

где $q(x,d)$ – функция, определенная на $\mathbb{R}^n$, которая зависит от векторного параметра $d \in \mathbb{R}^{m+1}_+$. Функция q определяется как

$$q(x,d) = H(d_0 h_0(x), d_1 h_1(x), \dots, d_m h_m(x)),$$

где H – непрерывная функция, определенная на $\mathbb{R}^{m+1}$.

Предположим, что выполняются следующие предположения:

1) $G(v, 0, 0, \dots, 0) = v \quad \forall v \geq 0$;

2) $G(v, y_1, \dots, y_m) \to +\infty$, когда $v \geq 0$ и $\|y\| \to +\infty$, где $y = (y_1, \dots, y_m)$;

3) $G(v, y_1, \dots, y_m) \geq v$ для всех $y \in \mathbb{R}^m$ и для всех $v \geq 0$.

Примеры вспомогательных функций, удовлетворяющих данным условиям, следующие:

1) $H_1(y) = y_0 + \sum_{i=1}^m \max\{y_i^2, 0\}$;

2) $H_2(y) = (y_0^p + \sum_{i=1}^m (\max\{|y_i|, 0\})^p)^{1/p}$ $(p > 0)$;

3) $H_3(y) = \max\{y_0, y_1^2, \dots, y_m^2\}$;

4) $H_4(y) = y_0 + \sum_{i=1}^m \max\{|y_i|, 0\}$.

Функции первого и четвертого типа – классические примеры штрафных функций, и по ним есть обширная библиография (см. [67] и ссылки в данной книге).

Обозначим $x^*(d)$ любое оптимальное решение (5.4).

Лемма 5.3 *Для любого* $\varepsilon > 0$ *существует* $M > 0$*, такое, что из* $d_0 = 1$, $\min_{i=\overline{1,m}} d_i > M$ *следует* $-\varepsilon \leq h_i(x^*(d)) \leq \varepsilon$ $\forall i = \overline{1, m}$.

Доказательство. Предположим, что это неверно. Тогда существует $\overline{\varepsilon} > 0$, такое, что для любого $M > 0$ существует вектор $\overline{d} = \overline{d}(M) = (\overline{d}_1, \overline{d}_2, \dots, \overline{d}_m)$, для которого $\overline{d}_i > M$ для всех $i = \overline{1, m}$ и

$$\text{либо} \quad h_j(x^*(\overline{d})) > \overline{\varepsilon}, \quad \text{либо} \quad h_j(x^*(\overline{d})) < -\overline{\varepsilon}$$

для некоторого $j \in \{1, 2, \dots, m\}$. Но

$$q(x^*(\overline{d}), \overline{d}) = H(h_0(x^*(\overline{d})), d_1 h_1(x^*(\overline{d})), \dots, d_m h_m(x^*(\overline{d}))).$$

Поскольку

$$\max_{i=\overline{1,m}} \overline{d}_i |h_i(x^*(\overline{d}))| \geq |\overline{d}_j h_j(x^*(\overline{d}))| \geq M\overline{\varepsilon}$$

и M может быть выбрано сколь угодно большим, можно использовать Предположение 2, из которого следует, что для некоторого $M > 0$ будет выполнено неравенство $q(x^*(\overline{d}), \overline{d}) > h^*$. Но, с другой стороны,

$$q(x^*(\overline{d}), \overline{d})) \leq q(x^*, \overline{d}) \quad \forall x^* \in Q^*,$$

так как $Q^* \subseteq X$. Однако

$$q(x^*, \overline{d}) = H(h_0(x^*), \overline{d}_1 h_1(x^*), \ldots, \overline{d}_m h_m(x^*)) = h_0(x^*) = h^*$$

по Предположению 1 и потому, что $h_i(x^*) = 0$ для всех $i = \overline{1, m}$. Получили противоречие, и лемма доказана. □

Определим точечно-множественные отображения вида

$$P(\varepsilon) = \{x \in X : -\varepsilon \leq h_i(x) \leq \varepsilon \ \ \forall i = \overline{1, m}\},$$

$$S(\varepsilon) = \{x \in P(\varepsilon) : h^* \geq h_0(x) \geq h^* - \varepsilon\}.$$

Очевидно, $P(0)$ – допустимое множество (5.3) и $S(0) = Q^*$.

Теорема 5.3 *Пусть отображение P полунепрерывно сверху в нуле. Тогда для любого $\varepsilon > 0$ существует $\overline{d} = (1, \overline{d}_1, \overline{d}_2, \ldots, \overline{d}_m) \in \mathbb{R}_+^{m+1}$, такое, что*

$$h(x^*(\overline{d})) \geq h^* - \varepsilon.$$

Доказательство. Аналогично доказательству Теоремы 5.1. □

Лемма 5.4 *Оптимальное значение в задаче (5.4) не превосходит оптимального значения в задаче (5.3) для всех $d \in \mathbb{R}_+^{m+1}$, таких, что $d_0 = 1$.*

Доказательство. По определению,

$$\min_{x \in X} q(x^*(d), d) = H(h_0(x^*(d)), d_1 h_1(x^*(d)), \ldots, d_m h_m(x^*(d))),$$

где $x^*(d)$ – оптимальное решение (5.4), соответствующее параметру d. Пусть x^* – произвольная точка из Q^*. Тогда по Предположению 1

$$H(h_0(x^*), d_1 h_1(x^*), \ldots, d_m h_m(x^*)) = H(h_0(x^*), 0, 0, \ldots, 0) = h^*,$$

и получаем

$$H(h_0(x^*(d)), d_1 h_1(x^*(d)), \ldots, d_m h_m(x^*(d))) \leq h^*.$$

Лемма доказана. □

Теорема 5.4 *Предположим, что S полунепрерывно сверху в нуле. Тогда для любого $\gamma > 0$ существует $\overline{d} = (1, \overline{d}_1, \ldots, \overline{d}_m) \in \mathbb{R}_+^{m+1}$, такое, что $\|x^* - x^*(\overline{d})\| \leq \gamma$ для некоторого $x^* \in Q^*$.*

Доказательство. Для любого $\gamma > 0$ существует $\Delta > 0$, такое, что

$$S(\lambda) \subseteq S(0) + \gamma B \ \ \forall \lambda \in [0, \Delta].$$

Если значение $\min_{i=\overline{1,m}} \overline{d}_i$ достаточно велико, мы имеем по Теореме 5.3 и Лемме 5.4 $x^*(\overline{d}) \in S(\Delta)$. Тогда $x^*(\overline{d}) \in S(0) + \gamma B$, то есть $x^*(\overline{d}) \in Q^* + \gamma B$, что означает $\|x^*(\overline{d}) - x^*\| \leq \gamma$ для некоторого $x^* \in Q^*$. В силу произвольности γ теорема доказана. □

5.3 Построение модифицированных функций Лагранжа с помощью возрастающих функций

В данном разделе мы рассматриваем комбинацию подходов, основанных на классической функции Лагранжа и на штрафных функциях, в которых используется возрастающая функция. Мы рассматриваем задачу (5.1) и определяем обобщенную функцию Лагранжа, такую, что, если ее седловая точка существует, она дает решение задачи с ограничениями.

Обобщенная функция Лагранжа задается формулой

$$U(x, d, r) = G(L(x, d), r_1 f_1(x), \ldots, r_m f_m(x)),$$

где $L(x, d) = f_0(x) + \sum_{i=1}^m d_i f_i(x)$ – классическая функция Лагранжа. Поэтому нам требуется решить вспомогательную задачу

$$\sup_{d \in \mathbb{R}_+^m} \inf_{x \in X} U(x, d, \overline{r}) \tag{5.5}$$

при фиксированном векторе штрафных параметров $\overline{r}$. Если седловая точка не существует, мы ищем приближенную седловую точку по отношению к d и x.

Пусть выполняются следующие предположения:

1) f_0 положительна на $\mathbb{R}^n$;

2) Все функции f_i ограничены снизу на $\mathbb{R}^n$;

3) $G(y_0, y_1, \ldots, y_m) \to +\infty$, когда $\max_{i=\overline{1,m}} y_i \to +\infty$, и y_0 ограничено снизу;

4) $G(v, y_1, \ldots, y_m) = v$ для всех $v \geq 0$ и всех $y \in \mathbb{R}^m$, таких,что $y_i \leq 0$ для всех $i \in \{1, 2, \ldots, m\}$;

5) $0 \geq G(v, y_1, \ldots, y_m) \geq v$ для всех $v < 0$ и всех $y \in \mathbb{R}^m$, таких, что $y_i \leq 0$ для всех $i \in \{1, 2, \ldots, m\}$;

6) G – возрастающая функция на $\mathbb{R}^{m+1}$.

Есть по крайней мере два примера функций, удовлетворяющих этим условиям. Первая из них

$$G_1(y) = y_0 + \sum_{i=1}^{m} \max\{y_i, 0\};$$

$$U_1(x, d, r) = L(x, d) + \sum_{i=1}^{m} r_i \max\{f_i(x), 0\}.$$

Функция такого типа была предложена Хестенесом и Пауэллом для ограничений в виде равенств и затем изучалась в [77] для ограничений в виде неравенств, и потому хорошо известна.

Другой пример следующий:

$$G_2(y) = \max\{y_0, y_1, \ldots, y_m\};$$

$$U_2(x, d, r) = \max\{L(x, d), r_1 f_1(x), \ldots, r_m f_m(x)\}.$$

Насколько нам известно, эта вспомогательная функция новая и позволяет комбинировать метод центров (см. [56, 8, 47]) с классическими методами множителей Лагранжа. В этом случае мы имеем минимаксную задачу, для которой есть различные алгоритмы недифференцируемой оптимизации (см. [92]). Теоретический подход к нахождению седловых точек недифференцируемых функций был предложен в [34].

Замечание 5.2 *Похожая схема была рассмотрена в [3]. Однако наши условия гораздо более общие; мы не предполагаем существования седловой точки $L(x, d)$.*

Более точно, в нашем случае верна следующая

Теорема 5.5 *Пусть седловая точка* $(\overline{x}, \overline{d})$ *для* $L(x, d)$ *существует,* $f_i(\overline{x}) \leq 0$ *для всех* $i = \overline{1, m}$ *и выполняется условие дополняющей нежесткости*

$$\overline{d}_i f_i(\overline{x}) = 0 \quad \forall i = \overline{1, m}.$$

Тогда $(\overline{x}, \overline{d})$ *– седловая точка для функции* $U(x, d, \overline{r})$ *для любого фиксированного вектора* $\overline{r}$ *штрафных параметров.*

Доказательство. Заметим, во-первых, что если $(\overline{x}, \overline{d})$ – седловая точка, то $\overline{x}$ – оптимальное решение задачи (5.1), и поэтому $f_0(\overline{x}) = f^*$. Также

$$U(\overline{x}, \overline{d}, \overline{r}) = f_0(\overline{x}) = f^*$$

по Предположению 4 и в силу условия дополняющей нежесткости.

С другой стороны, по определению седловой точки имеем

$$L(\overline{x}, \overline{d}) \leq L(x, \overline{d}) \quad \forall x \in X$$

и

$$L(\overline{x}, \overline{d}) \geq L(\overline{x}, d) \quad \forall d \in \mathbb{R}^m_+.$$

Но G – возрастающая функция, и если мы подставим все эти значения в первый аргумент G, то получим

$$\begin{aligned} U(x, \overline{d}, \overline{r}) &= G(L(x, \overline{d}), \overline{r}_1 f_1(x), \ldots, \overline{r}_m f_m(x)) \geq \\ &\geq G(L(\overline{x}, \overline{d}), \overline{r}_1 f_1(x), \ldots, \overline{r}_m f_m(x)). \end{aligned} \tag{5.6}$$

Все значения $\overline{r}_i f_i(x)$ ограничены снизу по Предположению 2 и потому, что G возрастает, значит,

$$G(L(\overline{x}, \overline{d}), \overline{r}_1 f_1(x), \ldots, \overline{r}_m f_m(x)) \geq G(L(\overline{x}, \overline{d}), -C, \ldots, -C)$$

для достаточно большой положительной константы C. Но значение в правой части равно $f_0(\overline{x})$. Следовательно,

$$G(L(\overline{x}, \overline{d}), \overline{r}_1 f_1(x), \ldots, \overline{r}_m f_m(x)) \geq U(\overline{x}, \overline{d}, \overline{r}) \quad \forall x \in X,$$

и применив (5.6), получим

$$U(x, \overline{d}, \overline{r}) \geq U(\overline{x}, \overline{d}, \overline{r}) \quad \forall x \in X.$$

Покажем, что также

$$U(\overline{x}, d, \overline{r}) \leq U(\overline{x}, \overline{d}, \overline{r}) \quad \forall d \in \mathbb{R}^m_+.$$

По определению U имеем

$$U(\overline{x}, d, \overline{r}) = G(L(\overline{x}, d), \overline{r}_1 f_1(\overline{x}), \ldots, \overline{r}_m f_m(\overline{x})) \leq$$
$$\leq G(L(\overline{x}, \overline{d}), \overline{r}_1 f_1(\overline{x}), \ldots, \overline{r}_m f_m(\overline{x})) = f^* = U(\overline{x}, \overline{d}, \overline{r}) \quad \forall d \in \mathbb{R}^m_+.$$

Поэтому, действительно, $(\overline{x}, \overline{d})$ – седловая точка для U, и теорема доказана. □

Замечание 5.3 *Обратное неверно. Хорошо известно, что модифицированная функция Лагранжа $U_1(x, d, r)$ имеет седловую точку во многих случаях, когда классическая функция Лагранжа $L(x, d)$ ее не имеет (см. [67]).*

Даже если седловой точки нет, оптимальное значение задачи (5.5) равно f^* при слабых предположениях.

Теорема 5.6 *Мы имеем* $\sup_{d \in \mathbb{R}^m_+} \inf_{x \in X} U(x, d, \overline{r}) = f^*$.

Доказательство. Покажем вначале, что

$$\sup_{d \in \mathbb{R}^m_+} \inf_{x \in X} U(x, d, \overline{r}) \geq f^*. \tag{5.7}$$

По определению супремума,

$$\sup_{d \in \mathbb{R}^m_+} \inf_{x \in X} U(x, d, \overline{r}) \geq \inf_{x \in X} U(x, 0_m, \overline{r}) \geq \inf_{x \in X} f_0(x) = f^*.$$

Здесь 0_m – нулевой вектор в $\mathbb{R}^m$. Одна часть теоремы доказана.

Покажем теперь, что

$$\sup_{d \in \mathbb{R}^m_+} \inf_{x \in X} U(x, d, \overline{r}) \leq f^*.$$

Пусть x^* – оптимальное решение (5.1). Тогда $x^* \in X$ и $f_i(x^*) \leq 0$ для всех $i \in \{1, 2, \ldots, m\}$. Пусть также $\tilde{d}$ – произвольный вектор из $\mathbb{R}^m_+$. Тогда

$$\inf_{x \in X} U(x, \tilde{d}, \overline{r}) \leq U(x^*, \tilde{d}, \tilde{r}).$$

Если значение $L(x^*, \tilde{d})$ меньше или равно нулю, то правая часть неположительна и поэтому меньше, чем f^*. В противном случае получаем

$$U(x^*, \tilde{d}, \tilde{r}) = L(x^*, \tilde{d}) \le f_0(x^*).$$

Последнее неравенство выполняется, так как x^* допустима для задачи (5.1). Наконец, поскольку $\tilde{d}$ выбран произвольно, имеем

$$\sup_{d \in \mathbb{R}^m_+} \inf_{x \in X} U(x, d, \overline{r}) \le f_0(x^*) = f^*.$$

Теорема доказана. □

Установим теперь соответствие между оптимальными точками начальной задачи с ограничениями и седловыми точками вспомогательной задачи.

Теорема 5.7 *Пусть* $(\overline{x}, \overline{d})$ *– седловая точка задачи (5.5). Тогда* $\overline{x}$ *– оптимальное решение (5.1).*

Доказательство. Покажем вначале, что $\overline{x}$ – допустимая точка. Так как это седловая точка, по Предположениям 4 - 6 и Теореме 5.6 имеем

$$f^* = U(\overline{x}, \overline{d}, \overline{r}) = G(L(\overline{x}, \overline{d}), \overline{r}_1 f_1(\overline{x}), \dots, \overline{r}_m f_m(\overline{x})) \ge$$

$$\ge G(L(\overline{x}, d), \overline{r}_1 f_1(\overline{x}), \dots, \overline{r}_m f_m(\overline{x})) \quad \forall d \in \mathbb{R}^m_+.$$

Тогда по Предположению 4

$$f^* \ge L(\overline{x}, d) = f_0(\overline{x}) + \sum_{i=1}^{m} d_i f_i(\overline{x}) \tag{5.8}$$

для всех неотрицательных d. Если мы предположим, что $\overline{x}$ недопустима, получим противоречие, так как можно взять $d_k = 0$ для всех k, таких, что $f_k(\overline{x}) \le 0$, и устремить их к плюс бесконечности для всех k, таких, что $f_k(\overline{x}) > 0$, которые должны существовать по предположению. Увеличивая эти компоненты d, получаем нарушение неравенства (5.8). Значит, $\overline{x}$ допустима.

Но если мы положим $d = 0_m$, имеем $f^* \ge f_0(\overline{x})$, что означает, что $\overline{x}$ также оптимальна. Теорема доказана. □

Глава 6

Параметрический подход к задачам глобальной оптимизации специального вида

6.1 Формулировка задачи

Рассмотрим следующее параметрическое семейство задач математического программирования:

$$\min \quad f(x, \rho), \tag{6.1}$$

$$x \in X,$$

где f – непрерывная функция, определенная на $\mathbb{R}^n, X$ – компактное множество и ρ – векторный параметр, $\rho = (\rho_1, \ldots, \rho_m) \in \mathbb{R}^m$.

Предположим, что f неотрицательна для всех $x \in X, \rho \in P$, где $P \subset \mathbb{R}^m$ – компактное выпуклое множество, и существует точка x^*, такая, что

$$f(x^*, \rho) = 0 \quad \forall \rho \in P.$$

Очевидно, это означает

$$f(x^*, \rho^*) = \min_{x \in X} \min_{\rho \in P} f(x, \rho) = 0 \tag{6.2}$$

для любого фиксированного вектора $\rho^* \in P$. Наша задача – найти такую пару (x^*, ρ^*). В данной главе мы называем вектор ρ "весами", поскольку в экономических моделях он соответствует весам (значимостям) разных экономических агентов.

Есть по крайней мере два важных частных случая такой задачи. Первый из них – решение системы нелинейных уравнений

$$f_i(x) = 0, \quad i = \overline{1, k}, \quad x \in X. \tag{6.3}$$

Если мы определим $f(x, \rho) = \sum_{i=1}^{k} \rho_i f_i^2(x)$, то наши предположения выполнятся для любого множества $P \in \mathbb{R}_+^n$ при условии, что f_i непрерывны и множество решений (6.3) непусто (конечно, можно использовать другие нормы вместо евклидовой).

Второй пример – задача нахождения экономического равновесия. Ее математическая модель была разработана А. Рубиновым в [85]. В [85] было также показано, что для нахождения равновесия необходимо решить уравнение

$$\varphi_1(p, \rho) - \varphi_2(p, \rho) = 0, \tag{6.4}$$

причем известно, что $\varphi_1(p, \rho) \geq \varphi_2(p, \rho)$ для всех p, ρ. Здесь $\varphi_1(p, \rho)$, $\varphi_2(p, \rho)$ – маргинальные функции специального вида. Определяя

$$f(x, \rho) = \varphi_1(p, \rho) - \varphi_2(p, \rho),$$

получаем важный частный случай начальной общей задачи.

Заметим, что обычно функция f невыпукла, что не позволяет применять локальные методы для нахождения (x^*, ρ^*). Поэтому мы имеем задачу глобальной минимизации специального вида.

Предположим, однако, что для каждого вектора весов $\rho \in P$ имеется локальный алгоритм $M_\rho : X \to X$, то есть каждой точке $y \in X$ соответствует следующая итерационная точка $M_\rho(y)$, и этот метод гарантирует сходимость к стационарной точке $f(x, \rho)$ на X для любого фиксированного ρ. Если $X = \mathbb{R}^n$ и $f(x, \rho)$ дифференцируема, это может быть метод наискорейшего спуска.

Все алгоритмы, рассматриваемые ниже, состоят из "больших" и "малых" итераций. "Большая" итерация – это один шаг алгоритма глобальной оптимизации, и соответствующие итерационные точки обозначаются y_k. "Малая" итерация – это итерация локального алгоритма M_ρ, и ее итерационные точки обозначаются z_i. Однако z_0 обычно совпадает с y_k для некоторого k, так как итерационная точка "большой" итерации является также начальной точкой последовательности "малых" итераций.

Предположим, что для любого вектора весов $\hat{\rho}$ и любой стационарной точки x функции $f(x, \hat{\rho})$ имеем $M_{\hat{\rho}}(x) = x$, то есть стационарная точка $f(x, \hat{\rho})$ есть неподвижная точка алгоритма $M_{\hat{\rho}}$. Конечно, это условие выполняется для любого решения начальной задачи.

Предположим также, что для каждого $\rho \in P$ имеем одну точку глобального минимума функции $f(x, \rho)$ на X, которую обозначим G_ρ, и $N \geq 0$ других стационарных точек (некоторые из них могут совпадать), обозначаемых $S_\rho^1, S_\rho^2, \ldots, S_\rho^N$. Конечно, если мы найдем G_ρ для некоторых весов, то G_ρ – решение начальной задачи, так как $f(G_\rho, \rho) = 0$ по предыдущему предположению. Предположим для простоты, что алгоритм M_ρ позволяет найти точно стационарную точку S_ρ^k для некоторого k, $1 \leq k \leq N$, для любого веса $\rho \in P$. Идея в том, чтобы изменять веса ρ, переходя из стационарной точки с одним значением ρ к следующей стационарной точке, соответствующей другому ρ, и в итоге найти глобальный минимум.

6.2 Алгоритмы

Чтобы найти решение (x^*, ρ^*), мы предлагаем следующие схемы:

Концептуальный Алгоритм 1

Шаг 0. Полагаем $k := 0$. Выбираем произвольно $y_0 \in X$. Выбираем вектор начальных весов ρ_k.

Шаг 1. Находим локальным методом M_{ρ_k} с начальной точкой y_k стационарную точку задачи

$$\begin{gathered} \min \quad f(x, \rho_k), \\ x \in X. \end{gathered} \tag{6.5}$$

Обозначим ее y^*.

Шаг 2. Если $f(y^*, \rho_k) = 0$, СТОП. Иначе выбираем новые веса ρ_{k+1}, полагаем $k := k + 1, y_k := y^*$ и возвращаемся к Шагу 1.

Простейший выбор весов – случайным образом с равномерным распределением на замкнутом множестве $P \subset \mathbb{R}_+^m$.

Концептуальный Алгоритм 2

Шаг 0 - Шаг 1. Как и в предыдущем алгоритме.

Шаг 2. Если $f(y^*, \rho_k) = 0$, СТОП. Иначе переходим к Шагу 3.

Шаг 3. Если $f(y^*, \rho_k) \le \gamma f(y_k, \rho_k)$, то полагаем $k := k + 1$, $y_k := y^*$ и возвращаемся к Шагу 1 с новыми весами (*хороший шаг*).

Шаг 4. Полагаем $k := k + 1$; $y_k := y_{k-1}$, выбираем новые веса ρ_k и возвращаемся к Шагу 1.

Здесь $\gamma \in (0, 1)$ – параметр, определяющий требуемое убывание целевой функции.

Нам понадобятся следующие определения.

Определение 6.2.1 *Пусть $\rho \in P$. Множество*

$$A_\rho = \{z \in X : \lim_{i\to+\infty} z_i = G_\rho,\ z_0 = z,\ z_{i+1} = M_\rho(z_i) \quad \forall i\}$$

называется областью притяжения глобального минимума для вектора весов ρ.

Заметим, что для задач нахождения экономического равновесия и минимизации взвешенной суммы квадратов $f(G_\rho, \rho) = 0$ для любого вектора ρ со строго положительными координатами (см. [85]). Это значит, что для решения исходной задачи достаточно найти G_ρ для некоторых положительных весов.

Определение 6.2.2 *Пусть $\rho \in P$. Множество*

$$B_\rho^j = \{z \in X : \lim_{i\to+\infty} z_i = S_\rho^j,\ z_0 = z,\ z_{i+1} = M_\rho(z_i) \quad \forall i\}$$

называется областью притяжения стационарной точки S_ρ^j для вектора весов ρ.

Обозначим также

$$R(x) = \{\rho : x \in A_\rho\}, \quad R^j(x) = \{\rho : x \in B_\rho^j\}.$$

Пусть μ_x – лебегова мера на X и μ_ρ – лебегова мера на P. Так как локальный метод M_ρ сходится для любого ρ, имеем

$$X = A_\rho \bigcup (\bigcup_{j=1}^{N} B_\rho^j) \quad \forall \rho \in P,$$

$$P = R(x) \bigcup (\bigcup_{j=1}^{N} R^j(x)) \quad \forall x \in X.$$

Представим некоторые результаты о сходимости наших концептуальных алгоритмов.

Теорема 6.1 *Предположим, что для всех* $x \in X$ *выполняется условие* $\mu_\rho(R(x)) > \alpha_\rho > 0$. *Предположим также, что мы выбираем веса* ρ *случайно с равномерным распределением на* $P \subset \mathbb{R}^m$. *Пусть выборы весов для разных* k *независимы. Тогда для Концептуального Алгоритма 1 предел*

$$\lim_{k \to +\infty} P\{y_i \neq G_\rho \quad \forall i = 1, 2, \ldots, k\} = 0.$$

Доказательство. Так как веса ρ распределены равномерно на множестве P, для любого $\Omega \subseteq P$ вероятность

$$P(\rho_k \in \Omega) = \int_\Omega \frac{1}{\mu_\rho(P)} d\mu_\rho = \frac{\mu_\rho(\Omega)}{\mu_\rho(P)}$$

на каждой k-й итерации. Тогда для каждой итерационной точки y_k, по Алгоритму 1, мы имеем

$$P(\rho_k \in R(y_k)) = \frac{\mu_\rho(R(y_k))}{\mu_\rho(P)} \geq \frac{\alpha_\rho}{\mu_\rho(P)} > 0.$$

Так как выборы весов ρ_k для разных k независимы, получаем

$$P(\rho_0 \notin R(y_0), \rho_1 \notin R(y_1), \ldots, \rho_k \notin R(y_k)) \leq (1 - \frac{\alpha_\rho}{\mu_\rho(P)})^k$$

на k-й итерации. Можно всегда предположить,что $\alpha_\rho < \mu_\rho(P)$. Так как значение в правой части стремится к нулю, вероятность в левой части также стремится к нулю. Но $G_\rho \neq y_i$ для всех $i \leq k$, только если $\rho_i \notin R(y_i)$ для всех i от 0 до $k-1$, иначе $y_i \in A_{\rho_i}$ и точка y_{i+1} – глобальный минимум. Теорема доказана. □

Обозначим

$$F_{\rho,\gamma}(x) = \{\rho : f(M_\rho(x), \rho) \leq \gamma f(x, \rho)\},$$

где $0 < \gamma < 1$. Здесь $M_\rho(x)$ – следующая итерационная точка, построенная локальным методом M с весами ρ в целевой функции.

Теорема 6.2 *Предположим, что веса выбираются, как в предыдущей теореме. Пусть существует $\varepsilon > 0$, такое, что если $f(z, \rho) < \varepsilon$ для некоторого $\rho \in P$, то $z \in A_\rho$ для всех $\rho \in P$. Предположим также, что для всех $x \in X$ имеем*

$$\mu_\rho(F_{\rho,\gamma}(x)) \geq \beta > 0.$$

Тогда для последовательности, построенной Алгоритмом 2,

$$\lim_{k \to +\infty} P\{y_i \neq G_\rho \quad \forall i = 1, 2, \ldots, k\} = 0.$$

Доказательство. Обозначим $f_0 = f(y_0, \rho_0)$. Это значение конечно и положительно по предположению. Пусть

$$s = \frac{\log(\varepsilon/f_0)}{\log \gamma}.$$

Это значение конечно и положительно для достаточно малого ε, так как γ строго положительно. Предположим теперь, что итерационные точки y_i не принадлежат A_{ρ_i} для всех $i = \overline{1, k}$. Это означает, что мы можем сделать не более $[s]$ хороших шагов, где скобки означают округление к ближайшему меньшему целому. Действительно, иначе мы бы имели

$$f(y_k, \rho_k) \leq \gamma^s f_0 \leq \varepsilon,$$

что означает, что $y_k \in A_\rho$ для всех ρ.

Вероятность сделать хороший шаг на каждой итерации положительна, и она не меньше, чем $\delta \equiv \beta/\mu_\rho(P)$ (это повторяет первую часть доказательства Теоремы 6.1). Поэтому, вспоминая схему Бернулли, получаем для достаточно большого k, что

$$P(N \leq [s]) \leq \sum_{i=0}^{[s]} C_k^i \delta^i (1 - \delta)^{k-i},$$

где N – число хороших шагов. Значение в правой части стремится к нулю, так как s фиксировано и не зависит от k. Но если число хороших шагов не превосходит $[s]$, это именно тот случай, когда $f(y_i, \rho_i) > \varepsilon$ для всех i. Поэтому вероятность не найти глобальный минимум за k итераций стремится к нулю, и теорема доказана. $\square$

Теорема 6.3 *[23] Предположим, что для всех $x \in X$ выполняется равенство*

$$\bigcup_{\rho \in P} M_\rho(x) = X.$$

Предположим также, что существует $\varepsilon > 0$, такое, что для всех $\rho \in P$ мы имеем $B(G_\rho, \varepsilon) \subset A_\rho$, где $B(G_\rho, \varepsilon)$ – шар с центром G_ρ радиуса ε. Наконец, пусть следующее соотношение выполняется для любого множества $\tilde{P} \subset P$ и любого $x \in X$:

$$\mu_\rho(\tilde{P}) \geq C\mu_x(\bigcup_{\rho \in \tilde{P}} M_\rho(x)).$$

Здесь $C > 0$. Тогда для последовательности, построенной Алгоритмом 1,

$$\lim_{k \to +\infty} P\{y_i \neq G_\rho \quad \forall i = 1, 2, \ldots, k\} = 0.$$

Доказательство. Пусть дана итерационная точка y_j для произвольного j. Обозначим

$$Q(y_j) = \{\rho \in P : M_\rho(y_j) \in B(G_\rho, \varepsilon)\}.$$

Для достаточно малого ε справедливо равенство

$$\bigcup_{\rho \in Q(y_j)} M_\rho(y_j) = B(G_\rho, \varepsilon).$$

Включение левой части в правую вытекает непосредственно из определения. Покажем, что

$$\bigcup_{\rho \in Q(y_j)} M_\rho(y_j) \supseteq B(G_\rho, \varepsilon).$$

Возьмем точку $y \in B(G_\rho, \varepsilon)$. Допустим, что она не принадлежит $\bigcup_{\rho \in Q(y_j)} M_\rho(y_j)$. Так как отображение M_ρ сюръективно, должно выполняться $y \in M_{\tilde{\rho}}(y_j)$ для некоторого $\tilde{\rho}$, не принадлежащего $Q(y_j)$. Но это противоречит определению $Q(y_j)$.

Заметим, что если выберем вес $\rho_j \in Q(y_j)$, следующая итерационная точка будет принадлежать A_ρ для всех ρ и, следовательно, мы найдем глобальный минимум на следующей итерации.

Используя неравенство в утверждении теоремы, получаем

$$\mu_\rho(Q(y_j)) \geq C\mu_x(\bigcup_{\rho \in Q(y_j)} M_\rho(y_j)) \geq C\mu_x(B(G_\rho, \varepsilon)) \geq \sigma > 0$$

для некоторой константы σ. Вероятность

$$P(\rho_j \in Q(y_j)) \geq \frac{\sigma}{\mu_\rho(P)} \quad \forall\, i = 0, 1, \ldots.$$

Поэтому вероятность

$$P(\rho_0 \notin Q(y_0), \rho_1 \notin Q(y_1), \ldots, \rho_k \notin Q(y_k)) \leq (1 - \frac{\sigma}{\mu_\rho(P)})^k,$$

и она стремится к нулю. Таким образом, вероятность не найти оптимальную точку за k итераций стремится к нулю, когда k стремится к бесконечности. Теорема доказана. □

Представим простой пример, который показывает, как можно численно реализовать Алгоритм 1.

Пример

Пусть

$$f_1(x) = (x-1), \quad f_2(x) = (x+1)(x-1), \quad x \in \mathbb{R}.$$

Требуется решить систему нелинейных уравнений

$$f_i(x) = 0, \quad i = 1, 2,$$

с решением $x^* = 1$. Вспомогательная задача минимизации суммы квадратов следующая:

$$min \quad [\rho_1(x-1)^2 + \rho_2(x+1)^2(x-1)^2\ ],$$
$$x \in \mathbb{R}.$$

Предположим, что мы ищем стационарные точки для этой задачи, используя алгоритм локальной минимизации без производных.

Возьмем начальные веса $\rho_1 = 1$; $\rho_2 = 10$. Для этих значений есть локальный минимум $y_1 = -0.9472$. Предположим, что мы попали в него (иначе мы бы нашли глобальный минимум).

В следующий раз выбираем $\rho_1 = 1$ и $\rho_2 = 5$. Локальный поиск с малым начальным шагом, стартуя из y_1, приводит в локальный минимум $y_2 = -0.8873$. Теперь, если мы выберем веса $\rho_1 = 1,\ \rho_2 = 1$ и стартуем с y_2, то найдем глобальный минимум.

Алгоритмы 1 и 2 имеют общий недостаток: для стационарной точки соответствующая область притяжения может содержать саму эту точку для всех весов. Например, если $f'(y_k, \rho) = 0$ для всех $\rho \in P$, то при применении любого градиентного метода будет повторяться одна и та же точка y_k для всех весов и всех k. Поэтому необходимо модифицировать концептуальную схему, допуская возмущение текущей итерационной точки, чтобы выходить из стационарных неоптимальных точек.

Концептуальный Алгоритм 3 (метод возмущений)

Шаг 0. Полагаем $k := 0$. Выбираем произвольно $y_0 \in X$ и параметр $\varepsilon > 0$. Выбираем вектор начальных весов ρ_k. Полагаем $S := \emptyset$. Фиксируем целое положительное число K.

Шаг 1. Как в Алгоритме 1.

Шаг 2. Если $f(y^*, \rho_k) = 0$, СТОП. Иначе переходим к Шагу 3.

Шаг 3. Если $y^* \notin S$, то выбираем новые веса ρ_{k+1}, полагаем $k := k + 1, y_k := y^*, S := S \cup \{y^*\}$ и возвращаемся к Шагу 1. Иначе переходим к Шагу 4.

Шаг 4. Полагаем $m := 0$. Выбираем $\tilde{y}$ как случайный вектор из шара $B(y^*, \varepsilon), \varepsilon > 0$.

Шаг 5. Находим локальным методом M_{ρ_k} с начальной точкой $\tilde{y}$ стационарную точку целевой функции f с прежними весами ρ_k. Обозначим ее $\hat{y}$.

Шаг 6. Если $\hat{y} \notin S$, то выбираем новые веса ρ_{k+1}, полагаем $k := k + 1, y_k := \hat{y}, S := S \cup \{\hat{y}\}$ и возвращаемся к Шагу 1. Иначе переходим к Шагу 7.

Шаг 7. Полагаем $m := m + 1$. Если $m > K$, меняем веса и переходим к Шагу 4. Иначе переходим к Шагу 4 с прежними весами.

Для изучения сходимости алгоритма введем следующие множества:

$$Y_\rho^{i,j} \equiv \{x \in B(S_\rho^i, \varepsilon) : x \in B_\rho^j\}, \quad \forall j = \overline{1, N},$$

$$U(i, j, \rho) = \mu_x(Y_\rho^{i,j}), \quad \forall j = \overline{1, N},$$

$$Y_\rho^{i,0} \equiv \{x \in B(S_\rho^i, \varepsilon) : x \in A_\rho\},$$

$$U(i, 0, \rho) = \mu_x(Y_\rho^{i,0}).$$

Здесь $B(S_\rho^i, \varepsilon)$ – шар с центром S_ρ^i радиуса ε. Предположим, что $\varepsilon > 0$ достаточно мало.

Множество $Y_\rho^{i,j}$ есть множество точек из окрестности стационарной точки S_ρ^i, принадлежащих области притяжения стационарной точки S_ρ^j. Тогда вероятность перейти из окрестности i-й стационарной точки в j-ю стационарную точку равна $\mu_x(Y_\rho^{i,j})/\mu_x(B_\varepsilon)$, где B_ε – шар радиуса ε.

Пусть веса ρ выбраны из конечного множества P случайно и независимо с равномерным распределением. Предположим для простоты, что все точки S_ρ^j и G_ρ различны. Если точка y^* равна S_ρ^j для некоторого ρ, мы имеем состояние j. Если точка y^* – глобальный оптимум, то имеем состояние нуль.

Выбор весов и сделанные предположения означают, что построена конечная марковская цепь. Найдем вероятности перехода $\tilde{p}_{ij}$ для нее. Предположим, что веса из P перенумерованы от 1 до $|P|$, где $|P|$ – число элементов (сложность) конечного множества P. Получаем

$$\tilde{p}_{ij} = \sum_{s=1}^{|P|} \frac{U(i,j,\rho_s)}{\mu_x(B_\varepsilon)} \cdot \frac{1}{|P|} = \frac{\sum_{s=1}^{|P|} U(i,j,\rho_s)}{|P| \cdot \mu_x(B_\varepsilon)}$$

(индекс s означает порядковый номер веса P, а не номер итерации).

Предложение 6.1 *Пусть не существует такое подмножество* $I \subseteq \{1, 2, \ldots, N\}$, *где* N *– максимальное число стационарных точек при данном* ρ, *что*

$$\tilde{p}_{ij} = 0 \ \ \forall i \in I, \ \ \forall j \notin I, \tag{6.6}$$

$$\tilde{p}_{i0} = 0 \ \ \forall i \in I. \tag{6.7}$$

Тогда с вероятностью 1 за конечное число итераций Алгоритм 3 остановится в нулевом состоянии, то есть будет найден глобальный оптимум.

Доказательство. Оно следует из свойств конечных марковских цепей (см. [58]). Условие (6.6) означает, что есть единственное абсорбирующее состояние, и оно будет достигнуто за конечное число итераций с вероятностью 1. □

Сходимость алгоритма зависит от комбинаторных свойств, связанных с положением стационарных точек целевой функции $f(x, \overline{\rho})$ для различных фиксированных весов $\overline{\rho}$. Изучение сходимости в случае выбора весов из конечного множества может быть сделано с использованием теории графов. Если мы определим специальный граф с M вершинами, соответствующими глобальному минимуму G_ρ, и MN вершинами, соответствующими другим стационарным точкам (M здесь – число различных весов, N – максимальное число стационарных точек), то для сходимости достаточна связность графа. Ребро (i,j) графа означает, что вероятность перейти из вершины i в вершину j положительна.

Рассмотрим теперь детерминистский выбор весов. Пусть мы имеем s различных фиксированных векторов $p_i, i = \overline{1,s}$, которые образуют множество $P \subseteq \mathbb{R}^n$, такое, что

$$\|p_i - p_j\| \geq \varepsilon > 0 \ \ \forall i,j = \overline{1,s}, \ i \neq j.$$

Простейший алгоритм имеет следующий вид.

Концептуальный Алгоритм 4

Шаг 0. Полагаем $k := 0; i := 1$. Выбираем произвольную точку $y_0 \in X$. Полагаем $\rho_0 = p_1$.

Шаг 1. Используя локальный метод M_{ρ_k}, находим стационарную точку задачи

$$\min \ \ f(x, \rho_k), \tag{6.8}$$

$$x \in X$$

с начальной точкой y_k. Обозначим ее y^*.

Шаг 2. Если $f(y^*, \rho_k) \leq \gamma f(y_k, \rho_k)$, то полагаем $y_k := y^*, i := i, \rho_{k+1} := \rho_k, k := k+1$ и возвращаемся к Шагу 1 (это *хороший шаг*). Иначе переходим к Шагу 3.

Шаг 3. Полагаем $i := i+1$. Если $i > s$, то $i := 1$. Полагаем $k := k+1, \ y_k := y_{k-1}, \ \rho_k := p_i$ и возвращаемся к Шагу 1.

В данном алгоритме мы меняем веса циклически. Если последний по порядку вес не дает спуск по целевой функции, то выбираем начальные веса вновь.

Обозначим

$$Z(x,\rho) = \lim_{j \to +\infty} M_\rho(y_j), \ y_0 = x;$$

$$D_{\rho,\gamma} = \{x \in X : f(Z(x,\rho),\rho) \leq \gamma f(x,\rho)\}.$$

Множество $Z(x,\rho)$ всегда состоит из одной точки, так как по предыдущим предположениям метод M сходится для любого ρ и любой начальной точки x. Множество $D_{\rho,\gamma}$ состоит из всех точек, из которых хороший шаг будет выполнен при весах ρ. Следующее необходимое условие сходимости очевидно.

Предложение 6.2 *Для сходимости к глобальному минимуму необходимо, чтобы*

$$\bigcup_{\rho\in P} D_{\rho,\gamma} \supseteq X\backslash(\bigcup_{\rho} G_\rho).$$

Действительно, иначе существует точка $x \in X$, из которой нельзя сделать хороший шаг и которая неоптимальна. Тогда последовательность $\{y_k\}$ не может сойтись к G_ρ для любого $\rho \in P$.

Теорема 6.4 *Предположим, что существует $\varepsilon > 0$, такое, что из $f(z,\rho) < \varepsilon$ следует $z \in A_\rho$ для всех ρ. Пусть также*

$$\bigcup_{\rho\in P} D_{\rho,\gamma} = X\backslash(\bigcup_{\rho} G_\rho).$$

Тогда за конечное число "больших" итераций оптимальное решение G_ρ начальной задачи будет найдено.

Доказательство. Пусть $f_0 = f(y_0,\rho_0)$ и $s = \log(\varepsilon/f_0)/\log\gamma$. Тогда для сходимости мы должны сделать хотя бы $]s[= [s]+1$ хороших шагов. Предположим, что после сколь угодно большого числа итераций мы можем сделать только $q < [s]$ хороших шагов. Но поскольку $\cup_{\rho\in P} D_{\rho,\gamma} = X\backslash(\cup_\rho G_\rho)$, это означает, что мы не можем сделать хороший шаг из некоторой итерационной точки y_k, что противоречит условию, если y_k неоптимальна. Теорема доказана. □

Основная проблема в алгоритме – угадать выбор весов, обеспечивающих выполнение условий теоремы. Обозначим

$$Q(x) = \{\rho \in P : f(Z(x,\rho),\rho) \leq \gamma f(x,\rho)\}.$$

Тогда, если мы хотим использовать конечное множество весов, то должны выбирать их так, чтобы для каждого $x \in X$ был хотя бы один вес $\rho \in Q(x)$.

Один из возможных выборов следующий. Пусть для каждого $x \in X$, такого, что $f(x, \rho) > \varepsilon$, множество $Q(x)$ открыто и содержит шар радиуса $\delta > 0$ с центром в некоторой точке $\overline{\rho}(x)$. Тогда мы можем выбирать веса так, чтобы

$$\inf_i \sup_{p \in P} \|p_i - p\| < \delta.$$

В этом случае для каждого $\rho \in Q(x)$ существует значение $j \in \{1, \ldots, s\}$, для которого ρ_j также принадлежит $Q(x)$. Поэтому для каждой точки x, которая не ε-оптимальна, после конечного числа итераций будет сделан хороший шаг. Одна из возможных модификаций Алгоритма 4, реализующая эту идею, – следующая схема, в которой предлагается выбирать самый "удаленный" вес от ранее рассмотренных.

Модифицированный Концептуальный Алгоритм 4

Шаг 0. Полагаем $k := 0$. Выбираем произвольную точку $y_0 \in X$. Выбираем начальные веса ρ_0. Полагаем $\Omega = \{\rho_0\}$.

Шаги 1-2. Как в Алгоритме 4.

Шаг 3. Решаем следующую задачу:

$$\max \ d(z, \Omega),$$

$$z \in P,$$

где $d(z, \Omega)$ – расстояние от точки z до множества Ω. Пусть $\overline{z}$ – решение этой задачи.

Шаг 4. Полагаем $\Omega := \Omega \cup \{\overline{z}\}$, $\rho_k := \overline{z}$, $k := k + 1$ и возвращаемся к Шагу 1.

Если для каждого $x \in X$ существует открытое множество весов, гарантирующих хороший шаг, и меры всех таких множеств ограничены снизу положительной константой, данная модификация также позволяет найти глобальный оптимум.

6.3 Случай неограниченного допустимого множества

В предыдущем разделе мы предполагали, что локальный метод M сходится к стационарной точке из любой начальной. Однако если допустимое множество неограничено, итерационные точки могут уходить на бесконечность. Обозначим

$$I_\rho = \{z \in X : \overline{\lim}_{i\to+\infty}\|z_i\| = +\infty,\ z_0 = z,\ z_{i+1} = M_\rho(z_i)\}, \tag{6.9}$$

$$J_\rho(x) = \{\rho \in P : x \in I_\rho\}. \tag{6.10}$$

Множество I_ρ называется областью притяжения плюс бесконечности. Если существует точка $y \in I_\rho$ для всех ρ, Алгоритм 1 может расходиться. Однако, если для любой точки можно изменить веса так, что она не принадлежит больше области притяжения плюс бесконечности, можно построить сходящийся метод.

Концептуальный Алгоритм 5

Шаг 0. Полагаем $k := 0; i := 1$. Выбираем произвольную точку $y_0 \in X$. Выбираем начальные веса ρ_0.

Шаг 1. Используя локальный метод M_{ρ_k} с начальной точкой y_k, находим стационарную точку задачи

$$\min \quad f(x, \rho_k), \tag{6.11}$$

$$x \in X,$$

Обозначим ее y^*. Если стационарная точка не может быть найдена, так как итерационные точки стремятся к бесконечности, то переходим к Шагу 3.

Шаг 2. Если $f(y^*, \rho_k) = 0$, СТОП. Иначе выбираем новые веса ρ_{k+1}, полагаем $k := k + 1,\ y_k := y^*$ и возвращаемся к Шагу 1 (*хороший шаг*).

Шаг 3. Выбираем новые веса ρ_{k+1}, полагаем $k := k + 1,\ y_k := y_{k-1}$ и возвращаемся к Шагу 1.

Изучим сходимость алгоритма. Мы используем обозначения части 7.1 для областей притяжения глобального и локальных минимумов.

Теорема 6.5 *Предположим, что для всех $x \in X$ имеем*

$$\mu_\rho(R(x)) > \alpha_\rho > 0.$$

Предположим также, что выбираем веса ρ случайно и независимо из множества $P \subseteq \mathbb{R}^m$. Тогда для Алгоритма 5

$$\lim_{k \to +\infty} P\{y_i \neq G_\rho \ \ \forall i = 1, 2, \dots, k\} = 0. \tag{6.12}$$

Доказательство. Если $\rho_i \in J_\rho(y_i)$, мы не выполняем хороший шаг, и "малые" итерации останавливаются. Вероятность

$$P(\rho_i \in R(y_i)) \geq \frac{\alpha_\rho}{\mu_\rho(P)} > 0.$$

Повторив доказательство Теоремы 7.1, получаем равенство (6.12). Теорема доказана. □

Можно модифицировать Алгоритм 2, чтобы учесть, что итерационные точки могут стремиться к бесконечности. Построение аналогично.

Чтобы построить численно реализуемый метод, можем ограничивать норму итерационных точек. Если она превосходит данный уровень, мы делаем нулевой шаг, иначе – хороший шаг. Предположим, что $\sup_{\rho \in P} \|G_\rho\| \leq C > 0$. Выберем параметр $K > C$ и рассмотрим следующую схему.

Концептуальный Алгоритм 6

Шаг 0. Полагаем $k := 0$. Выбираем начальную точку $y_0 \in X$ такую, что $\|y_0\| < K$. Выбираем начальные веса ρ_0.

Шаг 1. Полагаем $z_0 := y_k, \ i := 0$.

Шаг 2. Если $\|z_i\| < K$, то переходим к Шагу 3. Иначе переходим к Шагу 4.

Шаг 3. Если z_i – стационарная точка, то переходим к Шагу 5. Иначе полагаем $i := i + 1, z_i := M_{\rho_k}(z_{i-1})$ и переходим к Шагу 2.

Шаг 4. Выбираем новые веса ρ_{k+1}, полагаем $k := k + 1, \ y_k := y_0$ и возвращаемся к Шагу 1 (*нулевой шаг*).

Шаг 5. Если $f(z_i, \rho_k) = 0$, СТОП. Иначе полагаем $k := k + 1, y_k := z_i$. Выбираем новые веса ρ_k и возвращаемся к Шагу 1.

Сходимость Алгоритма 6 зависит от параметра K. Если он слишком мал, то не будут делаться хорошие шаги и мы никогда не найдем стационарную точку.

Теорема 6.6 *Пусть веса ρ выбраны как в Теореме 6.1 и пусть ее условия выполняются. Предположим, что константа*

$$K > \sup_{\rho \in P} \|G_\rho\|.$$

Тогда для Алгоритма 6 выполняется равенство (6.12).

Доказательство. Мы по-существу имеем схему Алгоритма 1 с той разницей, что процесс может быть остановлен до нахождения стационарной точки. Поэтому алгоритм сходится. □

6.4 Алгоритм с возможностью возвращения

Одна из возможных модификаций Алгоритма 2 – позволить методу возвращаться в предыдущую итерационную точку, если за данное число итераций не было хороших шагов. Предлагается следующая схема.

Концептуальный Алгоритм 7 (метод с возвращениями)

Шаг 0. Полагаем $k := 0$. Выбираем начальную точку $y_0 \in X$. Выбираем начальные веса ρ_0 и целое число $m > 0$.

Шаг 1. Находим стационарную точку y^*, начиная с y_k. Если $f(y^*, \rho) = 0$, СТОП. Иначе переходим к Шагу 2.

Шаг 2. Если $f(y^*, \rho_k) \leq \gamma f(y_k, \rho_k)$, то полагаем $k := k+1, y_k := y^*$ и возвращаемся к Шагу 1 (*хороший шаг*). Иначе переходим к Шагу 3.

Шаг 3. Полагаем $i := 0$. Выбираем случайно вектор весов $\hat\rho$ и находим стационарную точку задачи

$$\min_{x \in X} \; f(x, \hat\rho), \tag{6.13}$$

с начальной точкой y_k. Обозначим ее $\hat y$. Если $f(\hat y, \hat\rho) \leq \gamma f(y_k, \rho_k)$, то полагаем $k := k+1, y_k := \hat y, \rho_k := \hat\rho$ и возвращаемся к Шагу 1 (*хороший шаг*). Иначе переходим к Шагу 4.

Шаг 4. Полагаем $i := i+1$. Если $i \leq m$, то переходим к Шагу 3 (*нулевой шаг*). Иначе переходим к Шагу 5.

Шаг 5. Если $k \leq 1$, то $y_k := y_0$. Иначе полагаем $y_{k+1} := y_{k-1}, k := k+1$ (*шаг назад*) и возвращаемся к Шагу 1 с новыми весами ρ_k.

В этой схеме веса меняются, если мы делаем нулевой шаг или шаг назад. Достоинство алгоритма в том, что он позволяет присутствие неоптимальных стационарных точек для всех ρ. Действительно, для такой точки после m попыток будет сделан шаг назад, что позволяет сохранить сходимость к глобальному оптимуму.

Вероятность сделать шаг назад из итерационной точки y_k равна

$$p_b = (1 - \frac{\mu_\rho(F_{\rho,\gamma}(y_k))}{\mu_\rho(P)})^m,$$

и если y_k – неоптимальная стационарная точка, она равна 1 (после m итераций). Алгоритм соответствует марковской цепи с отталкивающими состояниями, соответствующими стационарным неоптимальным точкам.

Вероятность перехода из состояния i в состояние j с $i, j \geq 1$ равна

$$\tilde{p}_{ij} = \frac{\mu_\rho(B_\rho^{ij})}{\mu_\rho(P)},$$

где

$$B_\rho^{ij} = \{\rho \in P : \lim_{k\to+\infty} z_k = S_\rho^j,\ z_0 = S_\rho^i,\ z_k = M_\rho(z_{k-1})\},$$

и вероятность поглощения равна

$$\tilde{p}_{i0} = \frac{\mu_\rho(A_\rho^i)}{\mu_\rho(P)},$$

где

$$A_\rho^i = \{\rho \in P : \lim_{k\to+\infty} z_k = G_\rho,\ z_0 = S_\rho^i,\ z_k = M_\rho(z_{k-1})\}.$$

Следовательно, для сходимости необходимо и достаточно отсутствие циклов, то есть таких множеств $I \subseteq \{1, 2, \ldots, N\}$, что

$$\mu_\rho(B_\rho^{ij}) = 0$$

для всех $i \in I, j \notin I$ и $\mu_\rho(A_\rho^i) = 0$. Если вероятность перехода для каждого состояния i в другое состояние j положительна, глобальный оптимум будет найден с вероятностью 1.

Чтобы улучшить численные результаты, полезно использовать методы (см. [97, 76]), когда вместо одной начальной точки y_0 берется целое множество начальных точек. Концептуальные схемы, однако, по сути не изменяются, как и результаты о сходимости.

6.5 Случай бесконечного числа стационарных точек

Для простоты мы предполагали выше, что число стационарных точек конечно для каждого веса ρ. Рассмотрим более общий случай. Пусть для каждого $\rho \in P$ мы имеем $N > 0$ связных множеств S^j_ρ стационарных точек и одно связное множество G_ρ глобальных оптимумов (используем ту же нотацию, что и в части 6.1, где множества S^j_ρ и G_ρ состояли из одной точки). Назовем множество S^j_ρ j-й стационарной областью, а множество G_ρ – областью глобальных оптимумов.

Обозначим

$$V^j(x) = \{\rho \in P : \lim_{i\to+\infty} z_i \in S^j_\rho,\ z_i = M(z_{i-1}),\ z_0 = x\}, \tag{6.14}$$

$$W(x) = \{\rho \in P : \lim_{i\to+\infty} z_i \in G_\rho,\ z_i = M(z_{i-1}),\ z_0 = x\}. \tag{6.15}$$

Множество $V^j(x)$ есть множество весов, для которых мы попадаем в j-ю стационарную область, начиная с x, и $W(x)$ – множество весов, для которых мы попадаем в область глобальных оптимумов, если начальная точка x.

Вероятность перейти из x в j-ю стационарную область равна

$$p_j = \mu_\rho(V^j(x))/\mu_\rho(P).$$

Вероятность найти глобальный оптимум равна

$$p_0 = \mu_\rho(W(x))/\mu_\rho(P).$$

Предположим, что $\mu_\rho(V^j(x)) \geq \lambda > 0$ для всех $\rho \in P, j \in I$ и всех $x \in X$. Предположим также, что существует стационарная область S^m_ρ, такая, что $\mu_\rho(W(x)) \geq \lambda > 0$ для всех $\rho \in P$ и всех $x \in S^m_\rho$. Это означает, что мы можем перейти из некоторой стационарной области в глобальный минимум с ненулевой вероятностью.

Теорема 6.7 *При предыдущих предположениях для Алгоритма 1 выполняется равенство (6.12).*

Доказательство. Вероятность найти точку из G_ρ за N "больших" итераций не меньше λ^N, а это фиксированное число. По нашим предположениям оно не зависит от итерационной точки. Поэтому (6.12) выполняется. □

Если значение λ мало, может потребоваться большое число итераций для нахождения оптимального решения. Если множество $W(x)$ непусто, но его мера нуль, Алгоритм 1 может расходиться. Важная и трудная задача – найти вес, гарантирующий сходимость, если он единствен и множество P содержит бесконечное число элементов.

Одна из возможностей следующая. Предположим, что для каждого $x \in X$ мы можем минимизировать $f(x, \rho)$ как функцию ρ. Например, для взвешенной суммы квадратов мы имеем линейную функцию для любого фиксированного x. Если P – многогранник, имеем задачу линейного программирования.

Концептуальный Алгоритм 8

(минимизация по отношению к весам)

Шаг 0. Полагаем $k := 0$. Выбираем начальную точку $y_0 \in X$. Выбираем начальные веса ρ_0.

Шаг 1. Находим стационарную точку y^* с начальной точкой y_k. Если $f(y^*, \rho) = 0$, СТОП. Иначе переходим к Шагу 2.

Шаг 2. Полагаем $y_k := y^*$. Решаем следующую задачу:

$$\min \quad f(y_k, \rho), \tag{6.16}$$

$$\rho \in P.$$

Пусть ρ^* – оптимальное решение этой задачи.

Шаг 3. Если $f(y_k, \rho^*) \le \gamma f(y_k, \rho_k)$, то полагаем $k := k + 1, \rho_k := \rho^*, y_k := y_{k-1}$ и возвращаемся к Шагу 1. Иначе выбираем ρ_{k+1} случайно, полагаем $k := k + 1, y_k := y_{k-1}$ и возвращаемся к Шагу 1.

К сожалению, недостаточно минимизировать ρ для сходимости, так как после этого итерационная точка может остаться прежней стационарной точкой, и в результате не будет глобальной сходимости. Однако с помощью предыдущего алгоритма можно ускорить Алгоритм 2. Выбор ρ на Шаге 3 не может нарушить сходимость в силу присутствия параметра γ.

6.6 Возможное обобщение

Условия сходимости могут быть не выполнены во многих случаях, и по этой причине мы вводим в анализ алгоритмов следующие понятия, которые позволяют получить более слабые условия сходимости.

Пусть $Z_0 \equiv \cup_\rho G_\rho$, то есть Z_0 – множество оптимальных решений исходной задачи. Обозначим

$$W^{(0)}(x) = \{\rho \in P : \lim_{i\to+\infty} z_i \in Z_0,\ z_i = M(z_{i-1}),\ z_0 = x\}.$$

Определение 6.6.1 *Пусть $\varepsilon > 0$ – положительный параметр. Точка $y \in X$ называется ε-субоптимальной точкой первого порядка, если выполняется следующее неравенство.*

$$\frac{\mu_\rho(W^{(0)}(y))}{\mu_\rho(P)} > \varepsilon > 0.$$

Множество всех ε-субоптимальных точек первого порядка обозначается Z_1.

Определение 6.6.2 *Точка y называется ε-субоптимальной точкой k-го порядка, если выполняется неравенство*

$$\frac{\mu_\rho(W^{(k-1)}(y))}{\mu_\rho(P)} > \varepsilon > 0.$$

Множество всех ε-субоптимальных точек k-го порядка обозначается Z_k. Здесь

$$W^{(k)}(x) = \{\rho \in P : \lim_{i\to+\infty} z_i \in Z_k,\ z_i = M(z_{i-1}),\ z_0 = x\}.$$

По сути, мы даем рекурсивное определение, так как множество $W^{(k)}(x)$ определяется через Z_k, а множество Z_{k+1} для $k \geq 0$ – через $W^{(k)}(x)$.

Теорема 6.8 *Пусть $X = \cup_{j=1}^K Z_j$, где $K < +\infty$. Предположим, что веса выбраны, как в Теореме 6.1. Тогда выполняется (6.12).*

Доказательство. Так как выборы весов независимы, вероятность найти глобальный минимум за K шагов не меньше ε^K. Вероятность не найти глобальный минимум за $K \times M$ шагов, где M – положительное число, не превосходит $(1-\varepsilon^K)^M$. Так как K фиксировано, это значение стремится к нулю и выполняется (6.12). Теорема доказана. □

Если P состоит из конечного числа весов, ε-субоптимальность означает, что существует хотя бы один вес, который гарантирует, что мы найдем глобальный минимум на следующей итерации. ε-субоптимальность k-го порядка означает, что найдутся k векторов весов, таких, что выбирая на i-й итерации i-й вектор весов, мы найдем глобальный минимум через k итераций. Конечно, если множество X состоит из ε-субоптимальных точек порядка не выше $K > 0$, достаточно рассмотреть всевозможные последовательности K векторов весов. Тогда за конечное число итераций глобальный минимум будет найден.

6.7 Приближенные стационарные точки

В концептуальных схемах, представленных выше, предполагалось, что можно найти стационарную точку точно для любого веса. Это предположение нереалистично, так как известные локально сходящиеся методы позволяют найти стационарную точку с конечной точностью $\varepsilon > 0$. Это может нарушить сходимость. Например, если для функции одной переменной мы ищем стационарную точку приближенно, все итерационные точки могут оставаться с одной стороны от нее, тогда как глобальный минимум находится с другой стороны.

Пусть мы имеем $N > 0$ стационарных неоптимальных точек и один глобальный оптимум для каждого ρ. Предположим, что для каждого вектора весов ρ и каждой начальной точки x мы можем найти локальным методом итерационную точку z_k, такую, что $\|z_k - S_\rho^j\| < \delta$ или $\|z_k - G_\rho\| < \delta$ для $\delta > 0$, которое может быть сколь угодно мало.

Предложим следующие схемы.

Концептуальный Алгоритм 9

Шаг 0. Полагаем $k := 0$. Выбираем начальную точку $y_0 \in X$. Выбираем начальные веса ρ_0 и положительный параметр $\delta > 0$.

Шаг 1. Находим локальным методом M_{ρ_k} с начальной точкой y_k точку z, такую, что $\|z - S_\rho^j\| < \delta$ для некоторого j.

Шаг 2. Выбираем y_{k+1} как случайный вектор из шара $B(z, 2\delta)$.

Шаг 3. Полагаем $k := k + 1$, выбираем новые веса ρ_k и возвращаемся к Шагу 1.

Другая модификация следующая:

Концептуальный Алгоритм 10

Шаги 0 - 1. Как в Алгоритме 9.

Шаг 2. Полагаем i:=0. Если $f(z, \rho_k) = 0$, СТОП. Если $f(z, \rho_k) \le \gamma f(y_k, \rho_k)$, то полагаем $k := k+1,\ y_k := z$ и возвращаемся к Шагу 1 с прежними весами. Иначе переходим к Шагу 3.

Шаг 3. Полагаем $i := i + 1$. Если $i < K$, где $K > 0$ – фиксированное положительное число, то меняем веса, полагаем $k := k + 1, y_k := y_{k-1}$ и возвращаемся к Шагу 1 (*нулевой шаг*). Иначе переходим к Шагу 4.

Шаг 4. Полагаем $k := k + 1$, выбираем y_k как случайный вектор из шара $B(z, 2\delta)$ и возвращаемся к Шагу 1.

Оба алгоритма похожи на метод возмущений. Их цель, однако, иная. Метод возмущений может быть использован, чтобы выйти из "плохой" стационарной точки, тогда как последние два алгоритма предлагаются, чтобы решить проблему, что невозможно найти стационарные точки точно и только точки из их окрестностей могут быть найдены.

Установить сходимость для случая приближенных стационарных точек гораздо труднее. Причина в том, что малые возмущения около стационарных точек могут изменять области притяжения.

Для изучения сходимости мы должны рассматривать не только области притяжения стационарных точек, но и следующие множества:

$$Y_\rho^{i,j} \equiv \{x \in B(S_\rho^i, \delta) : x \in B_\rho^j\}, \quad \forall j = \overline{1, N},$$

$$U(i, j, \rho) = \mu_x(Y_\rho^{i,j}), \quad \forall j = \overline{1, N},$$

$$Y_\rho^{i,0} \equiv \{x \in B(S_\rho^i, \delta) : x \in A_\rho\},$$
$$U(i, 0, \rho) = \mu_x(Y_\rho^{i,0}).$$

Мы имеем марковскую цепь, в которой каждая точка S_ρ^j соответствует j-му состоянию и состояние нуль соответствует глобальному оптимуму. Состояние нуль – абсорбирующее состояние (см. [58]), так как метод останавливается в оптимальной точке.

Пусть веса ρ выбираются из множества P случайно с равномерным распределением. Вероятность перехода p_{ij} из состояния i в состояние j можно оценить следующим образом:

$$p_{ij} \geq U(i, j, \rho)/\mu_x(B_{2\delta}), \quad j \neq 0,$$
$$p_{i0} \geq U(i, 0, \rho)/\mu_x(B_{2\delta}).$$

Здесь $\mu_x(B_{2\delta})$ – лебегова мера шара с радиусом 2δ в $\mathbb{R}^n$.

Заметим, что мы выбираем точку y_k из $B(z, 2\delta)$, а не из $B(z, \delta)$, чтобы сохранить области притяжения стационарной точки. Если все значения в правой части положительны, глобальный оптимум будет найден с вероятностью 1. Конечно, численная реализация будет зависеть от свойств конкретной задачи и ее целевой функции. Условия сходимости очень общие и они могут позволить разработать эффективную комбинацию локальных и глобальных методов.

Глава 7

Результаты численных экспериментов

7.1 Результаты для метода секущих углов

Чтобы проверить эффективность метода секущих углов, был проведен ряд численных экспериментов. Большинство из них проводилось с возрастающими выпуклыми по лучам функциями одного из следующих классов:

1) $f(x) = [Qx, x]$, где Q – матрица, не обязательно положительно полуопределенная; Q выбиралась как положительная матрица, что гарантирует, что соответствующая квадратичная функция ВВЛ.

2) Функция Кобба-Дугласа

$$f(x) = \prod_{i=1}^{s} x^{\alpha_i}, \ \sum_{i=1}^{s} \alpha_i \geq 1.$$

   Заметим, что задачи минимизации функций Кобба-Дугласа были изучены в [93], где был предложен другой алгоритм.

3) $f(x) = \max[Q_i x, x], \ i = \overline{1, k}$. Здесь f – максимум из квадратичных невыпуклых функций.

Функция ограничений g выбиралась или как квадратичная вогнутая функция с положительными коэффициентами, или как максимум линейных функций. Таким образом, допустимое множество часто выбиралось как в задаче обратного выпуклого программирования (см. [54, 53, 74]). Современная вычислительная техника позволяет решать задачи с размерностью до нескольких десятков переменных. Мы надеемся увеличить размерность, используя специальные свойства комбинаторной подзадачи. Программа была написана на языке TurboC для DOS/Windows и реализована также в операционной системе Unix. Значение верхней границы для переменных колебалось от 5 до 20. Начальная точка выбиралась случайно или как локальный минимум целевой функции, найденный локальным методом. Конечно, итерация метода секущих углов требует больших затрат времени, и, более того, нет необходимости применять его в окрестности глобального минимума, поскольку можно просто применить локальный поиск. Для улучшения алгоритма был разработан гибридный метод, и он оказался гораздо более эффективным. Локальный поиск был реализован как метод Шора с расходящимся шагом (см. [92]). Этот метод предпочтителен, так как его итерация требует очень мало времени и он работает также с негладкими функциями. Однако он чувствителен к выбору начального шага, и результаты заметно зависели от начального шага. Во всех случаях глобальный минимум был найден с точностью $\varepsilon = 0.001$. Если начальной точкой был локальный минимум, из одной из последующих точек осуществлялся спуск к глобальному минимуму. Рассмотрим некоторые конкретные примеры. Они имеют локальные, но не глобальные минимумы.

Пример 1.

$$x_1x_2x_3x_4x_5x_6x_7 + \sqrt{x_1x_2x_3x_4x_5x_6x_7} \to \min,$$

$$x_1^2 + 4x_2^2 + 2x_3^2 + 3x_4^2 + 2x_5^2 + 3x_6^2 + x_7^2 \geq 24,$$

$$1 \leq x_1 \leq 10,\ 1 \leq x_2 \leq 10,\ 1 \leq x_3 \leq 10,$$

$$1 \leq x_4 \leq 10,\ 1 \leq x_5 \leq 10, 1 \leq x_6 \leq 10,$$

$$1 \leq x_7 \leq 10.$$

Оптимальное решение этой задачи – точка $(1; \sqrt{3}; 1; 1; 1; 1; 1)$. Наилучшее решение, найденное гибридным алгоритмом, следующее:

$$x_1 = 1.000083;\ x_2 = 1.732257;\ x_3 = 1.000035;$$

$$x_4 = 1.000007;\ x_5 = 1.000042;\ x_6 = 1.000031;$$

$$x_7 = 1.000018.$$

Чтобы найти эту точку, гибридный метод потребовал 21 итерацию. Абсолютная точность по норме и относительная точность по целевой функции равны соответственно 0.000206 и 10^{-7}.

Пример 2.

$$x_1^2 x_2^2 + 3x_1x_2 + x_1 + 5x_2 + 2 \to \min$$

$$x_1^2 + x_2^2 \geq 1,$$

$$0 \leq x_1 \leq 1,$$

$$0 \leq x_2 \leq 1.$$

Оптимальное решение этой задачи – точка $(1;0)$. Наилучшее решение, найденное гибридным алгоритмом, следующее:

$$x_1 = 0.99981000;\ x_2 = 0.001000.$$

Чтобы найти эту точку, метод секущих углов потребовал 12 итераций.

Пример 3.

$$\max\{x_1, 2x_2, 3x_3, 4x_4, 5x_5\} \to \min,$$

$$x_1 + x_2 + x_3 + x_4 + x_5 \geq 1,$$

$$0 \leq x_i,\ i = \overline{1,5}.$$

Оптимальное решение этой задачи – точка

$$(60/137; 30/137; 20/137; 15/137; 12/137).$$

Наилучшее решение, найденное гибридным алгоритмом, следующее:

$$x_1 = 0.437950;\ x_2 = 0.218980;\ x_3 = 0.145980$$

$$x_4 = 0.10949;\ x_5 = 0.08759.$$

Чтобы найти эту точку, гибридный метод потребовал 80 итераций.

Пример 4.

$$\sum_{i=1}^{19} x_i x_{i+1} \to \min,$$

$$\sum_{i=1}^{n} x_i = 1,$$

$$0 \leq x_i, i = \overline{1,20}.$$

Оптимальное решение этой задачи – точка $z, z_{20} = 1, z_i = 0, i \neq 20$. Методом секущих углов эта задача была решена точно за 20 итераций.

Пример 5.

$$\exp(x_1) + 2\exp(x_2) + 3\exp(x_3) \to \min,$$

$$x_1 + x_2 + x_3 \geq 1,$$

$$0 \leq x_1 \leq 1,\ 0 \leq x_2 \leq 1,\ 0 \leq x_3 \leq 1.$$

Оптимальное решение этой задачи – точка $(1; 0; 0)$. Наилучшее решение, найденное методом секущих углов, следующее:

$$x_1 = 0.999230;\ x_2 = 0.001000;\ x_3 = 0.001010.$$

Чтобы найти эту точку, алгоритм потребовал 22 итерации.

Пример 6.

$$x_1^2 + x_2^2 + x_3^2 + 2x_1^2x_2^2 + 2x_1^2x_3^2 \to \min,$$

$$x_1 + x_2 + x_3 \geq 1,$$

$$0 \leq x_i \leq 1,\ i = \overline{1,3}.$$

Наилучшее решение, найденное гибридным алгоритмом, следующее:

$$x_1 = 0.26369;\ x_2 = 0.35694;\ x_3 = 0.37936.$$

Значение целевой функции в этой точке 0.37858. Чтобы найти ее, потребовалось 80 итераций.

Пример 7.

$$\max\{2x_1^2x_2; 5x_2x_3^2; 2x_1^2 + x_2^2 + 3x_3^2\} \to \min,$$

$$x_1 + x_2 + x_3 \geq 1,$$

$$0 \leq x_i \leq 1,\ i = \overline{1,3}.$$

Наилучшее решение, найденное гибридным алгоритмом, следующее:

$$x_1 = 0.25253;\ x_2 = 0.57021;\ x_3 = 0.17725.$$

Значение целевой функции 0.5469449. Чтобы найти эту точку, гибридный метод потребовал 80 итераций. Время решения не увеличивалось очень резко по отношению к размерности. Наилучшей стратегией для подзадачи было использовать метод имитации замораживания, вероятно, потому, что число локальных минимумов невелико. Метод секущих углов продемонстрировал следующее поведение. Он работает эффективно с негладкими функциями и функциями Кобба-Дугласа. Причина, возможно, в том, что функции типа минимума (см. [81]) хорошо их аппроксимируют. Для квадратичных функций результаты были удовлетворительными, хотя иногда требовалось много итераций, вероятно, в силу плохой обусловленности задач. Худший случай для метода секущих углов – линейная целевая функция, поскольку функции типа минимума плохо ее аппроксимируют. Можно сказать, что гибридный метод (метод секущих углов в сочетании с локальным поиском методом Шора) оказался очень эффективен для минимизации возрастающих выпуклых по лучам функций, когда число переменных относительно невелико. Он нечувствителен к ограничениям и давал близкие результаты для линейных и нелинейных ограничений. Применение метода ветвей и границ и более полное изучение подзадачи позволили применить метод к более широким классам задач и увеличить их размерность. Легко построить параллельные алгоритмы для решения подзадачи, и это – одно из направлений дальнейших исследований.

7.2 Результаты для звездных функций

Ряд численных экспериментов был проведен для тестовых примеров, большинство из которых имеет многочисленные локальные, но не глобальные минимумы. Оказалось, что ряд практических инженерных задач имеют ВЗ целевую функцию и выпуклое допустимое множество. Во многих случаях можно свести исходную задачу к минимизации ВЗ функции.

Рассмотрим следующие тестовые примеры, взятые из книг [44, 52].

Пример 1 [52].

$$0.01x_1^2 + x_2^2 \to \min,$$

$$x_1 x_2 \geq 25,$$

$$x_1^2 + x_2^2 \geq 25,$$

$$2 \leq x_1 \leq 50,$$

$$0 \leq x_2 \leq 50.$$

Целевая функция в данном и следующем примерах – не ВЗ функция, но это неотрицательная квадратичная функция, поэтому вместо нее минимизировался ее квадратный корень, который является ВПО (а поэтому ВЗ) функцией. Оптимальное решение этой задачи – точка $(\sqrt{250}; \sqrt{2.5})$. Наилучшее найденное решение

$$x_1 = 15.811402; x_2 = 1.581137.$$

Абсолютная точность по норме и относительная точность по целевой функции равны соотвественно 0.000014 и 0.000003. Чтобы найти это решение, алгоритм потребовал 18 итераций.

Пример 2 [52].

$$x_1^2 + x_2^2 + x_3^2 \to \min,$$

$$x_1^2 + x_2^2 - 1 \geq 0,$$

$$1 \leq x_1 \leq 10,$$

$$-10 \leq x_2 \leq 10,$$

$$-10 \leq x_3 \leq 10.$$

Так как допустимое множество этой задачи не принадлежит $\mathbb{R}^n_+$, было необходимо разделить его на несколько частей и минимизировать целевую функцию на каждой из них. Оптимальное решение этой задачи – точка $(1; 0; 0)$. Наилучшее найденное решение

$$x_1 = 1.000000; x_2 = 0.000000; x_2 = 0.000000.$$

Абсолютная точность по норме равна 0. Чтобы найти это решение, алгоритм потребовал 20 итераций. Заметим, что за 12 итераций было найдено оптимальное решение с точностью 0.001.

В следующих примерах, взятых из практических инженерных задач, целевая функция представляет собой сумму ВЗ и линейной функций. Она имеет вид

$$f(x) = g(x) + [c, x],$$

где g – ВЗ функция и $c \in \mathbb{R}^n$. В этом случае мы вводили новую переменную $v = [c, x]$ для линейной части, и в результате целевая функция становилась ВЗ на $\mathbb{R}^n_+$. Однако, так как в оптимальном решении v могло быть отрицательным, было необходимо несколько изменить подзадачу, включив v в функции типа минимума. А именно, мы генерировали подзадачи вида

$$\begin{aligned} t &\to \min, \\ \min(\langle h_i, x\rangle + v, c_i + v) \le t &\quad \forall i = \overline{0,k}, \\ x \in X, &\quad [c, x] = v, \end{aligned} \tag{7.1}$$

которые могут быть решены как задачи частично целочисленного программирования или задачи дизъюнктивного программирования (см. [26]). Важно отметить, что алгоритм позволяет проверить оптимальность наилучшей найденной точки. Если две последовательные итерационные точки совпадают, это означает, что они – глобальный минимум целевой функции (см. [21]). Оптимальность была подтверждена в примерах 3 и 4 этим способом.

Пример 3 [44].

$$x_1^{0.6} + x_2^{0.6} - 6x_1 - 4u_1 + 3u_2 \to \min,$$

$$x_2 - 3x_1 - 3u_1 = 0,$$

$$x_1 + 2u_1 \le 4,$$

$$x_2 + 2u_2 \le 4,$$

$$x_1 \le 3,$$

$$u_2 \le 1.$$

Оптимальная точка задачи – $(4/3; 4; 0; 0)$. Наилучшее найденное решение

$$x_1 = 1.333333; x_2 = 3.999999;$$

$$u_1 = 0.000000; u_2 = 0.000000.$$

Абсолютная точность по норме равна 0.000001, относительная точность по целевой функции – $3 \cdot 10^{-7}$. Чтобы найти это решение, алгоритм потребовал 12 итераций.

Пример 4 [44].

$$x_1^{0.6} + 2x_2^{0.6} + 2u_1 - 2x_2 - u_2 \to \min,$$

$$x_2 - 3x_1 - 3 = 0,$$

$$x_1 + 2u_1 \leq 4,$$

$$x_2 + u_2 \leq 4,$$

$$x_1 \leq 3,$$

$$u_2 \leq 2.$$

Алгоритм нашел следующее оптимальное решение:

$$x_1 = 0.000000; x_2 = 2.999999;$$

$$u_1 = 0.000000; u_2 = 0.999999.$$

Абсолютная точность по норме – 0.000001, относительная точность по целевой функции – $2 \cdot 10^{-7}$. Чтобы найти это решение, алгоритм потребовал 11 итераций.

Пример 5 [44].

$$x_1^{0.6} + x_2^{0.6} + x_3^{0.4} + 2u_1 + 5u_2 - 4x_3 - u_3 \to \min,$$

$$x_2 - 3x_1 - 3u_1 = 0,$$

$$x_3 - 2x_2 - 2u_2 = 0,$$

$$4u_1 - u_3 = 0,$$

$$x_1 + 2u_1 \leq 4,$$

$$x_2 + u_2 \leq 4,$$

$$x_3 + u_3 \leq 6,$$

$$x_1 \leq 3,$$

$$u_2 \leq 2,$$

$$x_3 \leq 4.$$

Оптимальное решение – точка $(2/3; 2; 4; 0; 0; 0)$. Наилучшее найденное решение

$$x_1 = 0.666666; x_2 = 2.000000; x_3 = 3.999999,$$

$$u_1 = 0.000000; u_2 = 0.000000; u_3 = 0.000000.$$

Абсолютная точность по норме равна 0.000001, относительная точность по целевой функции есть 10^{-7}. Чтобы найти это решение, алгоритм потребовал 27 итераций.

Вероятно, для подобных задач выгодно использовать гибридные методы. В случае ВВЛ функций они позволили значительно увеличить размерность решаемых задач ([21]).

Пример 6 [52].

$$2 - x_1x_2x_3 \to \max,$$

$$x_1 + 2x_2 + 2x_3 - x_4 = 0,$$

$$0 \le x_i \le 1 \quad i = \overline{1,3},$$

$$0 \le x_4 \le 2.$$

Оптимальное решение задачи – точка $(2/3; 1/3; 1/3; 2)$. Наилучшее найденное решение

$$x_1 = 0.667039; x_2 = 0.333240;$$

$$x_3 = 0.333240; x_4 = 2.000000.$$

Так как произведение переменных – не ВЗ функция, мы минимизировали его кубический корень, который является ВПО функцией. Абсолютная точность по норме равна 0.000372, и относительная точность по целевой функции – $7 \cdot 10^{-8}$. Чтобы найти это решение, алгоритм потребовал 57 итераций.

Подзадача с функциями типа минимума решалась с помощью программного пакета CPLEX MIP 4.0-5.0 (после ее сведения к задаче частично целочисленного программирования). Программа была написана в среде Borland C++ для Windows 95 и реализована в операционной системе AIX 3.2 на IBM RS 6000. Начальная точка выбиралась с помощью выпуклой оптимизации. Важно отметить, что алгоритм применим не только к минимизации, но и к максимизации возрастающих положительно однородных функций.

7.3 Результаты численных экспериментов для липшицевых функций

Применимость нашего подхода проверялась решением ряда тестовых задач с липшицевыми целевыми функциями, которые обычно выбирались из одного из следующих классов:

1) $f(x) = [Qx, x]$, где Q – симметричная матрица, не обязательно положительно полуопределенная. Обычно Q выбиралась как диагональная матрица с положительными и отрицательными элементами. Таким образом была рассмотрена задача минимизации разности выпуклых функций (см. [53] и ссылки в данной работе).

2) Дифференцируемые функции.

Допустимое множество было многогранником, что позволяло применять программный пакет CPLEX 5.0 для решения комбинаторной подзадачи. Если многогранник не принадлежал $\mathbb{R}^n_+$, находились нижние границы переменных, и допустимое множество сдвигалось в $\mathbb{R}^n_+$ простым линейным преобразованием. Программа была написана на языке программирования C++ для операционных систем Unix и AIX 3.2. Параметры M и p выбирались вручную. Было важно, чтобы эти параметры не принимали больших значений. Значение p должно быть в интервале между 2 и 8, и значение M не должно превосходить нескольких тысяч, чтобы найти оптимальное решение (иначе алгоритм неустойчив). Мы фиксировали p и увеличивали его, если оптимальное решение не было найдено за заданное число итераций. В большинстве случаев минимум был найден с точностью $\varepsilon = 0.0001$. Вероятно, эффективность зависит от свойств целевой функции. Рассмотрим три примера задач липшицева программирования, решенных методом секущих углов. Все они имеют локальные, но не глобальные минимумы.

Пример 1 [15].

$$\min \quad f(x) = x_1^2 + x_2^2 - \cos(18x_1) - \cos(18x_2),$$

$$-1 \le x_1 \le 1, \quad -1 \le x_2 \le 1.$$

Оптимальное решение задачи – точка $(0; 0)$. У задачи около 50 локальных минимумов. Наилучшая найденная методом секущих углов точка $(-0.001336; -0.001$

Чтобы найти ее, потребовалось 18 итераций и 10 секунд. Абсолютная точность по целевой функции – 0.0004. Наилучшие результаты были получены при p, равном семи. Значение M слабо влияло на сходимость, оно могло быть от нескольких десятков до нескольких тысяч.

Пример 2 [38].

$$\max \quad f(x) = x_1^2 + x_2^2,$$

$$4x_1 + 7x_2 \geq 28,$$

$$x_1 - 5x_2 \leq 5,$$

$$1 \leq x_1 \leq 3, \ 0 \leq x_2 \leq 2.$$

Оптимальное решение этой задачи (с точностью до шести цифр после десятичной точки) – точка $(6.481481; 0.296296)$. Наилучшее решение, найденное методом секущих углов,

$$x_1 = 6.481481, \quad x_2 = 0.296296.$$

Точность по норме – по крайней мере 10^{-7}. Алгоритм потребовал 3 итерации и 0.01 секунды для нахождения оптимального решения.

Пример 3 [31].

$$\min \quad f(x) = (x_2 - \frac{5.1}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6)^2 + 10(1 - \frac{1}{8\pi})\cos(x_1) + 10,$$

$$-5 \leq x_1 \leq 10, \quad 0 \leq x_2 \leq 15.$$

Глобальный оптимум достигается в трех точках:

$$(-3.142; 12.275), \ (3.142, 2.275), \ (9.425, 2.425).$$

Минимальное значение целевой функции равно 0.3979. Наилучшее решение, найденное методом секущих углов, есть $x_1 = 3.104749; x_2 = 2.353947$ со значением целевой функции 0.3995. Потребовались 18 итераций со значением M, равным 25, и значением p, равным 7. Заметим, что используя локальный метод, начиная с наилучшей найденной точки, удавалось найти решение с точностью 10^{-6}. Поэтому для данного примера целесообразно применять гибридный метод, сочетая метод секущих углов с локальным поиском. Для этого примера наилучшее значение нижней оценки равнялось 0.3946, и оно было получено при $M = 1$ и $p = 8$. Часто скорость сходимости нижних оценок λ_k к оптимальному значению была медленной. Это зависит от структуры целевой функции и ограничений.

Следующие примеры были решены методом ветвей и границ с применением абстрактной выпуклости. Данная техника позволила найти приближенные решения, начиная с которых локальный поиск находил глобальный оптимум. Абстрактные субградиенты помогали попасть в достаточно малую окрестность глобального минимума.

Пример 4.

$$\max \quad f(x) = 2(x_1-1)^2 + 3(x_2-2)^2 + 5(x_3-3)^2 + (x_4-2)^2 + 2(x_5-3)^2,$$

$$0 \le x_i \le 5,\ i = \overline{1,5}.$$

Оптимальное решение – точка $(5; 5; 0; 5; 0)$. Наилучшее решение, найденное методом ветвей и границ:

$$x_1 = 4.984131,\ x_2 = 4.997559,\ x_3 = 0.0024411,$$

$$x_4 = 4.997559,\ x_5 = 0.0024411.$$

Относительная точность по целевой функции – 0.0025. Алгоритм потребовал 200 итераций для нахождения данного решения.

Пример 5.

$$\max \quad f(x) = \frac{(x_1+x_2-2)^2}{x_1+3x_2} + \frac{(3x_1-x_2)^2}{2x_1+x_2},$$

$$0.5 \le x_i \le 5,\ i = \overline{1,2}.$$

Оптимальное решение – $(0.5; 1.5)$. Наилучшее решение, найденное методом ветвей и границ:

$$x_1 = 0.4828101, \quad x_2 = 1.468550.$$

Абсолютная точность по целевой функции равна 0.0007. Алгоритм потребовал 100 итераций для нахождения данного решения.

Пример 6.

$$\max \quad f(x) = 2\sin^2 2x + 3\sin^2 3x,$$

$$1 \le x_i \le 5,\ i = \overline{1,2}.$$

Оптимальное решение – $(3\pi/2; \pi)$. Наилучшее решение, найденное методом ветвей и границ:

$$x_1 = 4.723145, \quad x_2 = 3.145508.$$

Абсолютная точность по целевой функции – 0.001339. Когда локально сходящийся метод стартовал из наилучших найденных методом ветвей и границ точек (Примеры 4 - 6), оптимальные решения были найдены.

Основные выводы следующие. Метод секущих углов позволяет найти хорошее приближение глобального минимума, которое может быть использовано локальным методом. Когда он вставлялся в процедуру ветвей и границ для нахождения оценок, решения были найдены эффективно для задач умеренного размера с не слишком сложными целевыми функциями (имеющими относительно небольшое число локальных минимумов и не слишком малую область притяжения глобального минимума). Сложность комбинаторной подзадачи быстро растет с ростом размерности. Полученные за последнее время результаты о локальных минимумах подзадачи, так же, как и прогресс в дизъюнктивном программировании (см. [27, 29]), позволяют надеяться решить задачи липшицева программирования большего размера.

7.4 Численные результаты для симплициального метода

Чтобы проверить применимость метода, был решен ряд тестовых задач. Они включали задачи со следующими целевыми функциями:

1) Функция $f(x) = [Qx, x]$, где Q – матрица с положительными коэффициентами, не обязательно положительно полуопределенная. Как пример была взята следующая целевая функция:

$$f(x) = \sum_{i=1}^{n-1} x_i x_{i+1} + x_1 x_n.$$

2) Функции типа минимума вида

$$f(x) = \min a_i x_i, \quad a_i > 0 \quad \forall i = \overline{1, n}.$$

Таблица 7.1: Численные результаты

n	Тип	Точность	Время(с)
3	QUAD	0.0	2.0
5	QUAD	10^{-6}	3.0
10	QUAD	10^{-6}	7.5
20	QUAD	10^{-6}	27.0
3	MIN	10^{-5}	1.5
5	MIN	10^{-4}	45.0
3	SQRT	10^{-4}	2.5
5	SQRT	10^{-2}	180.0

3) Функции

$$f(x) = \sum_{i=1} a_i\sqrt{x_i}, \quad a_i > 0 \quad \forall i = \overline{1,n}.$$

Допустимое множество было единичным симплексом в $\mathbb{R}^n_+$. Было выбрано медианное разбиение. Размерность пространства изменялась от 2 до 10 (для квадратичных задач – до 20). Программа была написана на языке Fortran 90 для Pentium MMX 200. Максимальное число итераций колебалось от 100 до 10000. Результаты показали, что для ряда задач квадратичного программирования (даже с 20 переменными) $200-300$ итераций были достаточны, чтобы найти решение с точностью 10^{-6}. Для задач третьего типа эффективность зависела от размерности пространства. Для трех переменных потребовалось 200 итераций, для пяти переменных – 1000 итераций и для десяти переменных – более 10000 итераций. Однако эти результаты приведены для чистого симплициального метода. Использование ветвления и отсечения может резко сократить объем вычислений. Время вычислений колебалось от 0.1 секунды для 200 итераций до 0.5 часа для 10000 итераций. В Таблице 7.1 представлены численные результаты для некоторых тестовых функций. Они показывают, что алгоритм хорошо работает с квадратичными функциями, а для функций типа минимума и для квадратных корней сходимость медленнее. В первой колонке приведено число переменных, во второй – тип задачи, в третьей колонке – точность найденного решения $f(x) - f^*$ и в последней колонке – время вычислений. Конечно, можно провести больше экспериментов, которые включали бы ветвление и отсечение и более эффективное разбиение симплексов. Тем не менее, предварительные результаты показали применимость нашего подхода.

7.5 Результаты для квазивыпуклых функций

Приведем результаты решения тестовых примеров алгоритмом из Главы 2. Представим список тестовых задач

Пример 1 [17].

$$\min \quad f(y) = -2\xi_1 - 3\xi_2 + 2\xi_2^2$$

$$f_i(y) \le 0, \;\; i = 1,2, \quad 0 \le \xi_j \le 10, \;\; j = 1,2,$$

где

$$f_1(y) = \xi_1 + 4\xi_2 - 4, \quad f_2(y) = \xi_1 + \xi_2 - 2.$$

Начальная точка $y_0 = (0;0)$. Решение достигается на границе допустимого множества в точке $y^* = (1.75; 0.25)$. Значение целевой функции $f^* = -4.125$.

Пример 2 [17].

$$\min \quad f(x) = 7(\xi_1 - 6)^2 + (\xi_2 - 4)^2;$$

$$x \in D = \{f_i(x) \le 0, \;\; i = 1,2, \;\; 0 \le \xi_j \le 10 \;\; j = 1,2\},$$

где $f_1(x) = -\xi_1\xi_2 + 1; f_2(x) = \xi_1^2 + \xi_2^2 - 9$. Начальная точка – $(1;1)$. Минимальное значение целевой функции равно 97.30945, оптимальная точка $(2.7955, 1.0880)$.

Пример 3 (Розен-Судзуки) [47]. Минимизировать функцию

$$f(y) = \xi_1^2 + \xi_2^2 + \xi_2^3 + \xi_4^2 - 5\xi_1 - 5\xi_2 - 21\xi_3 + 7\xi_1,$$

$$x \in D = \{x \in \mathbb{R}^n : f_i(x) \le 0, i = 1,2,3\},$$

где

$$f_1(x) = \xi_1^2 + \xi_2^2 + \xi_3^2 + \xi_4^2 + \xi_1 - \xi_2 + \xi_3 - \xi_4 - 8,$$

$$f_2(x) = \xi_1^2 + 2\xi_2^2 + \xi_3^2 + 2\xi_4^2 - \xi_1 + \xi_4 - 10,$$

$$f_3(x) = 2\xi_1^2 + \xi_2^2 + 3\xi_3^2 + 2\xi_1 - \xi_2 - 5.$$

Оптимальное значение целевой функции равно -44.8682. Это значение достигается в точке

$$y^* = (0.273015; 1.04192; 2.06424; -0.738074).$$

Пример 4 [19]. Найти безусловный минимум функции

$$f(y) = \max_{1 \le i \le 5} \ f_i(y),$$

где

$$f_i(y) = b_i \sum_{j=1}^{10} (\xi_i - c_{ji}), \quad i = \overline{1,5},$$

$$b = (1; 5; 10; 2; 4; 3; 1.7; 2.5; 6; 3.5),$$

$$c = (c_{ji}) = \begin{bmatrix} 0 & 2 & 1 & 1 & 3 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 2 & 4 & 2 & 2 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 \\ 0 & 1 & 1 & 2 & 0 & 0 & 1 & 2 & 1 & 0 \\ 0 & 3 & 2 & 2 & 1 & 1 & 1 & 1 & 0 & 0 \end{bmatrix}.$$

Минимальное значение целевой функции, равное 22.00016, достигается в точке

$$y^* = (1.12434; 9.7945; 1.4770; 0.92023; 1.12427).$$

Начальная точка $y_0 = (0; 0; 0; 0; 0)$. Эта задача решалась многими специалистами и является классическим примером задачи негладкой оптимизации. Хорошо известно, что для нее очень эффективны методы обобщенного градиентного спуска с растяжением пространства ([92]).

Пример 5 (MAXQUAD) [64]. Найти минимум функции на $\mathbb{R}^n$:

$$f(y) = \max_{i=\overline{1,5}} [\sum_{j,k=1}^{10} a_{ijk}\xi_j\xi_k - \sum_{j=1}^{10} b_{ij}x_j],$$

где

$$a_{ijk} = \exp(i/j)\cos(ij)\sin(k), \quad b_{ij} = \exp(j/i)\sin(ij).$$

Известный минимум задачи $f^* = -0.84140$. Задача плохо обусловлена вблизи точки минимума:

$$\xi_1^* = -1.263721E-1, \ \xi_2^* = -3.445566E-2, \ \xi_3^* = -6.727508E-3,$$

$$\xi_4^* = 2.62061E-2,\ \xi_5^* = 0.723512E-2,\ \xi_6^* = -2.782221E-1,$$

$$\xi_7^* = 7.482760E-2,\ \xi_8^* = 1.385525E-1,\ \xi_9^* = 8.399986E-2,$$

$$\xi_{10}^* = 3.855654E-2.$$

Начальная точка – нуль.

Для решения данных задач была написана программа на языке C. Вычисления проводились в Turbo C для IBM PC. При необходимости параметры подбирались по ходу задачи.

Главные факторы, определяющие скорость нахождения минимума, – выбор начальной точки z и выбор шага. Если шаг выбирался из расходящегося ряда, наиболее важным был начальный шаг λ_0. Кроме того, эффективность зависела от скорости вычисления $\Phi(u)$, то есть от алгоритма одномерной минимизации. В экспериментах применялся метод золотого сечения. Вектор g_k выбирался как $f'(x_k)$, если соответствующая точка была внутри допустимого множества. Если она была на границе, g_k выбирался как выпуклая комбинация градиентов целевой функции и ограничений, чтобы выполнялось условие $[g_k, z-u_k] = 0$.

Что касается начальной точки, она выбиралась так, чтобы любое направление $u-z, u \in U$ было направлением убывания функции f. Иначе функция Φ может быть константой в некоторых областях, и трудно найти обобщенно-опорные векторы. Естественно выбирать $q = f'(z)$, если f дифференцируема. Заметим, что скорость сходимости иногда увеличивалась при использовании одномерной минимизации функции Φ.

Для Примера 1 найдена точная точка минимума, равная

$$x^* = (1.75, 0.25).$$

Начальная точка $z = (1;0)$, начальное λ_0 выбиралось в интервале $[0.1;10]$.

Для Примера 2 решение было найдено быстрее всего для $z = (2;1)$ и для $\lambda_0 = 0.15$. Наилучшее найденное решение

$$x^* = (2.7955; 1.0882).$$

Значение целевой функции 97.30945. В этой задаче выбор начальной точки был существен. Например, при $z = (1;1)$ сходимость замедлялась.

Выбор начального шага был определяющим для Примера 3. Наилучшее значение целевой функции было $f_r = -44.8393$, и оно было достигнуто в точке

$$x^* = (0.273438; 1.041149; 2.06379; -0.737780).$$

Было необходимо выбирать малое λ_0.

В Примере 4, как известно, целевая функция плохо обусловлена, и методы без растяжения пространства неэффективны. Наилучшая найденная точка

$$x = (1.124336; 9.79448; 1.47701; 0.92023; 1.124239)$$

со значением целевой функции, равным 22.000184. Сходимость, однако, была медленной, начиная с точек, значение целевой функции в которых не превосходило оптимальное значение на 0.01.

Последняя задача была решена лишь при удачном выборе параметров. В одном из экспериментов за 2112 вычислений целевой функции метод конического проектирования нашел точку с координатами

$$\xi_1^* = -1.263709E-1,\ \xi_2^* = -3.44493E-2,\ \xi_3^* = -6.727506-3,$$

$$\xi_4^* = 2.6202E-2,\ \xi_5^* = 0.723589E-2,\ \xi_6^* = -2.748973E-1,$$

$$\xi_7^* = 7.481360E-2,\ \xi_8^* = 1.385525E-1,\ \xi_9^* = 8.4000034E-2,$$

$$\xi_{10}^* = 3.855242E-2.$$

Значение целевой функции равно -0.84093.

Основной вывод состоит в том, что метод конического проектирования превосходит обобщенный градиентный спуск по времени и точности. Однако оба метода работают недостаточно хорошо с плохо обусловленными целевыми функциями и ограничениями. Кроме того, сходимость зависит от λ_0 и z. Уменьшая начальный шаг, мы увеличиваем как точность, так и время вычислений. Вероятно, следует включить растяжение пространства в схему минимизации $\Phi(u)$, когда задача плохо обусловлена.

Заключение

В данной работе получен ряд новых результатов в области абстрактного выпуклого анализа. Построен ряд численно реализуемых разновидностей обобщенного метода секущих плоскостей, которые позволяют решать задачи абстрактно выпуклого программирования со сложными целевыми функциями. Эти алгоритмы включают следующие. Во-первых, метод секущих углов для минимизации возрастающих выпуклых по лучам функций, которые играют важную роль в математической экономике. Во-вторых, алгоритм минимизации возрастающих звездных относительно бесконечности функций, которые являются основным классом производственных функций. Наконец, метод секущих углов был обобщен для произвольных липшицевых целевых функций. Численные эксперименты показали эффективность подхода для задач умеренной размерности. Были построены методы ветвей и границ, использующие метод секущих углов для вычисления оценок. Они могут значительно повысить численную эффективность.

Построен метод конического проектирования для задач квазивыпуклого программирования. Он превосходит по эффективности метод обобщенного градиентного спуска и основан на сведении задачи математического программирования к задаче безусловной минимизации с помощью специальной маргинальной функции.

Предложена схема нахождения гамильтонова цикла, базирующаяся на марковских цепях и функциях типа минимума. Задача сведена к минимизации функции типа минимума на многограннике. Негладкость функций типа минимума не является препятствием для численной реализации, сделанной с использованием программы CPLEX. Численные эксперименты показали хорошие результаты для графов не очень большого размера (до 100-200 вершин).

Разработана схема двойственности, основанная на возрастающих функциях. Доказана эквивалентность исходной и вспомогательной задач. Предложенная схема позволяет строить обобщенные функции Лагранжа и комбинировать метод центров с методами модифицированных функций Лагранжа, что может быть удобным для негладкой оптимизации.

Рассмотрен параметрический подход для решения задач глобальной оптимизации специального вида. Эти задачи включают решение нелинейных уравнений и нахождение экономического равновесия в некоторых моделях. При изменении весов в задаче осуществляется спуск, и находится глобальный минимум. На этой основе предложен ряд алгоритмов для ограниченного или неограниченного допустимого множества.

Литература

[1] Абасов Т.М., Рубинов А.М. Об одном классе H-выпуклых функций // Докл. АН России, Серия матем. - 1994. - Т.48. - N1. - С.95 - 97.

[2] Васильев Ф.П. Численные методы решения экстремальных задач. - М.: Наука, 1988. - 549с.

[3] Евтушенко Ю., Жадан В. Точные вспомогательные функции в задачах оптимизации // Журнал вычисл. матем. и матем. физики. - 1990. - Т.30, N1. - С.31 - 42.

[4] Заботин Я.И. О вычислении конусов обобщенно опорных функционалов // Исследования по прикладной математике. - Казань: Изд-во Казан. ун-та. - 1977. - N4. - С.3 - 10.

[5] Заботин Я.И., Анастасьев Е.В. Подход к решению задач нелинейного программирования // Тезисы докладов VI Научной конференции "Методы математического программирования и программное обеспечение". - Свердловск, 1989. - С.93.

[6] Заботин Я.И., Кораблев А.И., Хабибуллин Р.Ф. О минимизации квазивыпуклых функционалов // Известия вузов. Математика. - 1972. - N10. - С.27 - 33.

[7] Заботин Я.И., Кораблев А.И., Хабибуллин Р.Ф. Условия экстремума функционала при наличии ограничений // Кибернетика. - 1973. - N6. - С.65 - 70.

[8] Заботин Я.И. Минимаксный метод решения задач математического программирования // Известия вузов. Математика. - 1975. - N10. - С.36 - 44.

[9] Крейнин М.И., Анастасьев Е.В. Метод сведения задачи математического программирования к задаче безусловной минимизации // Тезисы докладов, X Симпозиум "Методы программного обеспечения для решения задач оптимального планирования", Нарва-Йыэсуу. - 1988. - С.43 - 44.

[10] Куликов А.Н., Фазылов В.Р. Выпуклая оптимизация с заданной точностью // Журнал вычисл. матем. и матем. физики. - 1990. - N5. - С.663 - 671.

[11] Кутателадзе С.С., Рубинов А.М. Двойственность Минковского и ее приложения. - Новосибирск: Наука, 1976. - 254с.

[12] Левин В.Л. Полуконические множества, полуоднородные функции и новая схема двойственности в выпуклом анализе // Докл. АН России. - 1997. - Т.354. - С.597 - 599.

[13] Макаров В.Б., Рубинов А.М. Математическая теория экономической динамики и равновесия. - М.: Наука, 1973. - 335с.

[14] Пиявский С.А. Алгоритм нахождения абсолютного экстремума функции // Журнал вычисл. матем. и матем. физики. - 1972. - N2. - С.57 - 67.

[15] Растригин Л.А. Системы экстремального управления. - М.: Наука, 1974. - 630с.

[16] Рубинов А.М. Минимизация норм на компактных множествах // Вестник Ленинградского университета. - 1965. - Т.20, N1. - С.140 - 142.

[17] Скоков В.А. Некоторый вычислительный опыт решения задач нелинейного программирования // Математические методы решения экономических задач. - М.: Наука, 1980. - С.51 - 69.

[18] Фазылов В.Р. Один общий метод отыскания точки выпуклого множества // Известия вузов. Математика. - 1983. - N6. - С.43 - 57.

[19] Шор Н.З., Шабашова Л.П. О решении минимаксных задач методом обобщенного градиентного спуска с растяжением пространства // Кибернетика. - 1972. - N1. - С.82 - 88.

[20] Allgower E.L., Georg K. Numerical Continuation Methods. - Berlin, Springer Verlag, 1990.

[21] Andramonov M.Yu., Rubinov A.M., Glover B.M. Cutting angle methods in global optimization // Applied Mathematics Letters. - May 1999.

[22] Andramonov M.Yu., Rubinov A.M., Glover B.M. Cutting angle method for minimizing increasing convex-along-rays functions // Research Report, University of Ballarat. - 1997. - 27p.

[23] Andramonov M.Yu. A parametric approach to global optimization problems of a special kind // Optimization: Contributions from Australasia. - Dordrecht, Kluwer Academic Publishers. - 1999.

[24] Andramonov M.Yu. A survey of methods of abstract convex programming, Journal of Statistics and Management Systems, to appear.

[25] Aubin J.-P, Ekeland I. Applied Nonlinear Analysis. - New York: Wiley & Sons. - 1990.

[26] Balas E. Disjunctive programming // Annals of Discrete Mathematics. - 1979. - No. 5. - P.3 - 51.

[27] Balas E., Tama J., Tind J. Sequential convexification in reverse convex and disjunctive programming // Mathematical Programming. - 1989. - V.44, No.3. - P.337 - 350.

[28] Barbara A., Crouzeix J.P. Concave gauge functions and applications // Zeitschrift fur Operations Research. - 1994. - V.40. - P.43 - 74.

[29] F.Beaumont Algorithm for disjunctive programming problems // European Journal of Operational Research. - 1990.

[30] Berteskas D.P. Constrained Optimization and Lagrange Multiplier Methods. - New York: Academic Press, 1982.

[31] Branin F.H., Hoo S.K. A method for finding multiple extrema of a function of n variables // Numerical Methods of Nonlinear Optimization, Lootsma, F.A. (ed.). - London: Academic Press, 1972. - P.231 - 237.

[32] Crouzeix J.P. A review of continuity and differentiability properties of quasi-convex functions on $\mathbb{R}^n$ // Optimization and Optimal Control, A. Auslender, W. Oettli, J. Stoer (eds.). - Berlin: Springer Verlag, 1981. - P.9 - 20.

[33] Crouzeix J.P., Ferland J.A. Criteria for quasi-convexity and pseudo-convexity: Relationships and Comparisons // Mathematical Programming. - 1982. - V. 23. - P.193 - 205.

[34] Demyanov V.F., Rubinov A.M. Constructive Non-smooth Analysis. - Frankfurt: Peter Lang, 1995.

[35] Dennis J.E., Schnabel R.B. Numerical Methods for Nonlinear Equations and Unconstrained Optimization. - Englewood Cliffs, NJ, Prentice Hall, 1983.

[36] Derman C. Finite State Markovian Decision Processes. - New York: Academic Press, 1970.

[37] Diener I. Trajectory nets connecting all critical points of a smooth function // Mathematical Programming. - 1986. - V.36, No.3. - P.340 - 352.

[38] Ehnbat R. An algorithm for maximizing a convex function over a simple set // Journal of Global Optimization. - 1996. - No.8. - P.379 - 391.

[39] Fiacco A.V., McCormick G.P. Nonlinear Programming: Sequential Unconstrained Minimization Techniques. - New York: Wiley & Sons, 1968.

[40] Filar J.A., Krass D. Hamiltonian cycles and Markov chains // Mathematics of Operations Research. - 1995. - V. 19, No. 1. - P.223 - 237.

[41] Ming Chen, Filar J.A. Hamiltonian Cycles, Quadratic Programming, and Ranking of Extreme Points // C.Floudas and P.Pardalos (eds.), Global Optimization. - Princeton University Press, 1992.

[42] Filar J.A., Ke Liu Hamilton Cycle Problem and Singularly Perturbed Markov Decision Process // Research Report 1996/1. - University of South Australia, School of Mathematics. - 1996.

[43] Filar J.A., Oberije M.G.M., Pardalos P.M. Hamiltonian Cycle Problem, Controlled Markov Chains and Quadratic Programming // Research Report, University of South Australia. - 1993.

[44] Floudas C.A., Pardalos P. A Collection of Test Problems for Constrained Global Optimization Algorithms // Lecture Notes in Computer Science. - Berlin: Springer-Verlag, 1990. - V.455.

[45] Garcia C.B., Zangwill W.I. Pathways to Solutions, Fixed Points and Equilibria. - Englewood Cliffs NJ, Prentice Hall, 1981.

[46] Gromicho J. Quasi-convex Optimization and Location Theory. - Amsterdam, Thesis Publishers. - 1995.

[47] Grossmann Ch., Kaplan A. Straffunktionen und modifizierte Lagrange funktionen in der nichtlinearen Optimierung. - Leipzig: Teubner Texte zur Mathematik, 1979.

[48] Hansen P., Jaumard B., Lu S.H. Global optimization of univariate Lipschitz functions: II. New algorithms and computational comparison // Mathematical Programming. - 1992. - V.55. - P.273 - 292.

[49] Hansen P., Jaumard B., Lu S.H. An analytical approach to global optimization // Mathematical Programming. - 1991. - V.52. - P.227 - 254.

[50] Held M., Wolf P., Crowder H.P. Validations of subgradient optimization // Mathematical Programming. - 1974. - No.6. - P.62 - 88.

[51] Hiriart-Urruty J.B., Lemarechal C. Convex Analysis and Minimization Algorithms. - Berlin: Springer-Verlag, 1993. - V. II.

[52] Hock W., Schittkowski K. Test Examples for Nonlinear Programming Codes // Lecture Notes in Economic and Mathematical Systems. - Berlin: Springer-Verlag, 1981. - V.177.

[53] Horst R., Pardalos P. (eds) Handbook on Global Optimization. - Dordrecht: Kluwer Academic Publishers, 1996.

[54] Horst R., Tuy H. Global Optimization. - Berlin: Springer-Verlag, 1990.

[55] Horst R., Tuy H. On an outer approximation concept in global optimization // Optimization. - 1989. - V.20. - P.255 - 264.

[56] Huard P. Resolution of mathematical programming with nonlinear constraints by the method of centers // Nonlinear Programming (J. Abadie ed.). - Amsterdam: North Holland Publishing Company, 1967. - P.207 - 219.

[57] Intriligator M. Mathematical Optimization and Economic Theory. - Englewood Cliffs NJ, Prentice Hall, 1971.

[58] Isaacson D.L., Madsen R.W. Markov Chain Theory and Applications. - New York: Wiley & Sons, 1976.

[59] Kallenberg L.C.M. Linear Programming and Finite Markovian Control Problems // Mathematical Center Tracts. - Amsterdam. - 1983. - V.148.

[60] Kelley J. The cutting plane method for solving convex programs // SIAM Journal. - 1960. - V. 8, No.4. - P.703 - 712.

[61] Konno H.J., Kuno T. Generalized linear multiplicative and fractional programming // Annals of Operations Research. - 1990. - V.25. - P.147 - 162.

[62] Konnov I.V. Relaxation methods and Variational Inequalities. - Berlin: Springer Verlag, 2000.

[63] Lawler E.L., Lenstra J.K., Rinnooy Kan A.H.G., Shmoys D.B. The Traveling Salesman Problem. A Guided Tour of Combinatorial optimization. - Chichester: Wiley & Sons, 1985.

[64] Lemarechal C. Numerical experiments in nonsmooth optimization. Progress in Non-differentialble optimization, E. A. Nurminski (ed.), Laxenburg, IIASA. - 1982. - P.61 - 64.

[65] Levy A.V., Montalvo A. The tunneling algorithms for the global minimization of functions // SIAM Journal on Scientific and Statistical Computation. - 1985. - V.6. - P.15 - 29.

[66] Luenberger D.G. Linear and Nonlinear Programming. - Reading Massachusets: Addison Wesley, 1984.

[67] Minoux M. Programmation Mathematique. Theorie et Algorithmes. - Paris: Bordas et C.N.F.T.-E.N.S.T, 1989.

[68] Mladineo R.H. An algorithm for finding the global maximum of a multimodal, multivariate function // Mathematical Programming. - 1986. - V.34. - P.188 - 200.

[69] Nurminski E.A. Separating plane algorithms for convex optimization // Mathematical Programming. - 1997. - V. 76. - P.373 - 391.

[70] Pallaschke D., Rolewicz S. Foundations of Mathematical Optimization. - Dordrecht, Kluwer Academic Publishers. - 1997.

[71] Papadimitriou C., Steiglitz K. Combinatorial Optimization: Algorithms and Complexity. - Prentice Hall: New Jersey, 1982.

[72] Penot J.P. Duality for radiant and shady problems, manuscript.

[73] Penot J.P., Volle M. On quasi-convex duality // Mathematics of Operations Research. - 1990. - V.15, No.4. - P.597 - 624.

[74] Pinter J. Global Optimization in Action. Continuous and Lipschitz Optimization: Algorithms, Implementations and Applications. - Dordrecht: Kluwer Academic Publishers, 1996.

[75] Plastria F. Lower subdifferentiable functions and their minimization by cutting planes // Journal of Optimization Theory and Applications. - 1985. - V.46. - P.37 - 53.

[76] Rinnooy Kan A.H.G., Timmer G.T. Stochastic global optimization methods. Part II Multi-level methods // Mathematical Programming. - 1987. - V.39. - P.57 - 78.

[77] Rockafellar R.T. The multiplier method of Hestenes and Powell applied in convex programming // JOTA. - 1973. - V. 12, No.6. - P.555 - 562.

[78] Rubinov A.M. Abstract Convexity and Global Optimization. - Dordrecht, Kluwer Academic Publishers. - 2000.

[79] Rubinov A.M., Andramonov M.Yu. Minimizing increasing star-shaped functions based on abstract convexity // Journal of Global Optimization. - 1999. - V.15. - P.19 - 39.

[80] Rubinov A.M., Andramonov M.Yu. Lipschitz programming via increasing convex-along-rays functions // Optimization Methods and Software. - 1999. - V.10. - P. 763 - 781.

[81] Rubinov A.M., Glover B.M. Increasing convex along rays functions with applications to global optimization // Research Report 21/96, University of Ballarat. - 1996.

[82] Rubinov A.M., Glover B.M. Increasing convex along-rays-functions with applications to global optimization // Journal of Optimization Theory and Applications. - 1999. - V.102. - P.615 - 642.

[83] Rubinov A.M., Glover B.M. Quasi-convexity via two step functions // Proceedings of 5th International Symposium on Generalized Convexity, Luminy-Marseille. - Kluwer Academic Publishers, 1996.

[84] Rubinov A.M., Glover B.M., Jeyakumar V. A general approach to dual characterizations of inequality systems with applications // Journal of Convex Analysis. - 1995. - V.2. - No. 1/2. - P.309 - 344.

[85] Rubinov A.M. On some problems of nonsmooth optimization in economic theory // Nonsmooth Optimization: Methods and Applications, F. Giannessi (ed.) - Amsterdam: Gordon and Breach Sci. Publ, 1992.

[86] Rubinov A.M., Glover B.M. Characterizations of optimality for homogeneous programming problems with applications // Recent Advances in Nonsmooth Optimization, D.-Z. Du, L. Qi and R. S. Womersley (eds.) - 1995. - P.351-380.

[87] Rubinov A.M., Vladimirov A.A. Convex-along-rays functions and star-shaped sets // Research Report 97/8, University of Ballarat. - 1997.

[88] Rubinov A.M., Glover B.M., Yang X.Q. Extended Lagrange and Penalty functions in Continuous Optimization // Research Report 19/97, University of Ballarat. - 1997.

[89] Rubinov A.M., Glover B.M., Yang X.Q. Decreasing functions with application to penalization // SIAM J. Optimization. - 2000. - V.10. - P.289 - 313.

[90] Rubinov A.M., Tuy H., Mays H. Algorithm for a nonconvex global optimization problem // Research Report, University of Ballarat. - 1999.

[91] Schoen F. Stochastic techniques for global optimization: A survey of recent advances // Journal of Global Optimization. - 1991. - V.1. - P.207 - 228.

[92] Shor N.Z. Methods of Minimizing Nondifferentiable Functions. - Berlin: Springer-Verlag, 1985.

[93] Shor N.Z. Dual estimates in multiextremal problems // Journal of Global Optimization. - 1995. - V.7. - P.75-91.

[94] Shubert B.O. A sequential method seeking the global maximum of a function // SIAM Journal of Numerical Analysis. - 1972. - V.9. - P.379 - 388.

[95] Singer I. Abstract Convex Analysis. - New York: Wiley & Sons, 1997.

[96] Talman A.J.J., Van der Laan G. The Computation and Modelling of Economic Equilibria. - Amsterdam: North Holland, 1987.

[97] Törn A., Zilinskas A.V. Global Optimization // Lecture Notes in Computer Science. - Berlin: Springer Verlag, 1990. - V.350.

[98] Wolf P. A method of conjugate subgradients for minimizing non-differentiable functions // Mathematical Programming. - 1975. - Study 3. - P.145 - 173.

[99] Wood G.R. Multidimensional bisection applied to global optimization // Computers and Mathematics with Applications. - 1991. - V.21. - P.161 - 172.

[100] Wood G.R. The bisection method in higher dimensions // Mathematical Programming. - 1992. - V. 55. - P.319 - 337.

[101] Zhang B., Wood G.R., Baritompa W.P. Multidimensional bisection: the performance and the context // Journal of Global Optimization. - 1993. - V.3. - P.337 - 358.