Ю. Е. Нестеров

ЭФФЕКТИВНЫЕ МЕТОДЫ в нелинейном программировании

Предисловие

Выбор наилучших при данных условиях решений - несомненно один из основных принципов практической деятельности В областях, где соответствующие проблемы допускают формальную постановку, а список таких областей весьма представителен и постоянно расширяется, реализация этого принципа связана, как правило, с привлечением методов численной оптимизации. Эти методы сегодня являются важной частью инструментария прикладной математики.

Теория и численные методы оптимизации в последние десятилетия развивались весьма интенсивно - под воздействием и внутренних стимулов, и потребностей расширяющихся приложений. Все более увеличивается интерес к "трудным" задачам - негладким, большой размерности и т.д., и заметно смещаются акценты в самом характере исследований. Если раньше в центре внимания были такие вопросы, как необходимые и достаточные условия оптимума, унификация способов обоснования сходимости, локальный асимптотический анализ поведения конкретных методов, то теперь все больше внимания уделяется синтезу методов, эффективно или оптимально решающих все задачи того или иного типа.

Эта тенденция заслуживает более подробного рассмотрения. В не столь далекое время "право на существование" методу с теоретической точки зрения давал уже факт его сходимости, и основные усилия специалистов были направлены на обоснование сходимости тех или иных методов при по возможности более слабых предположениях о решаемой задаче. Между тем ясно, что сходимость - это чисто асимптотическое свойство вычислительной процедуры, само по себе ничего не говорящее о поведении этой процедуры на конечном времени, а только это последнее и представляет практический интерес. Возникающие вопросы скорости сходимости решались, по существу, качественно - локальным асимптотическим анализом поведения метода в окрестности решения, да и то лишь для "хороших" задач. Возникал своего рода порочный круг: ответ на вопрос о конструктивной характеризации асимптотического свойства - сходимости - давался в терминах асимптотического же описания поведения погрешности при неограниченном увеличении числа итераций. Если результаты такого рода и допускали конструктивные формулировки, то в них, как правило, фигурировали достаточно тонкие и трудно отслеживаемые характеристики решаемой задачи.

Намеченный подход не позволял снабдить рекомендуемые методы никакими "явными", выражающимися через наблюдаемые характеристики задачи и действующими "на любом конечном времени", а не только в асимптотике, оценками погрешности. Тем самым весь комплекс вопросов, связанных с анализом эффективности методов, оставался вне поля зрения теории, и вопросы эти решались чисто эмпирически. Некоторыми плодами этого подхода, скажем, симплекс-методом решения задач линейного программирования, мы с успехом пользуемся и по сей день; однако разнообразие задач оптимизации, возникающих в приложениях, таково, что эмпирические рекомендации, сделанные на основании (неизбежно далекого от полноты) вычислительного опыта не имеют особой цены или вовсе оказываются несостоятельными, как только поток однотипных задач, на которые настраивались" рекомендации, сменяется другим. Еще один недостаток эмпирического измерения эффективности - значительная трудоемкость "переранжирования" методов при расширении их арсенала.

Все описанные трудности стимулировали интерес к теоретическому исследованию проблемы эффективности Обычно она ставится следующим образом: дается класс оптимизационных задач. Требуется указать метод решения задач этого класса, погрешность которого при решении любой задачи не превосходит некоторой предъявляемой явно функции от числа итераций и параметров, задающих класс (числа переменных, характеристик гладкости минимизируемых функций и т.п.). Эта функция (она должна стремиться к нулю с ростом числа итераций) и является характеристикой эффективности соответствующего метода. В сочетании с оценкой вычислительной трудоемкости одной итерации (эта оценка, как правило, очевидным образом извлекается из описания метода) указанная характеристика эффективности позволяет дать гарантированные оценки расхода вычислительных ресурсов, необходимых для решения задач данного класса рассматриваемым методом.

Изучение эффективности различных методов и сравнение их друг с другом по этому показателю позволяет ранжировать их и, не обращаясь к вычислительным экспериментам, дать теоретически обоснованные рекомендации по выбору методов в зависимости от типа решаемых задач. Более того, для основных классов задач выпуклой оптимизации (из известных на сегодня классов задач математического программирования лишь выпуклые задачи, как оказывается, могут решаться сколь-нибудь эффективно) удалось решить и проблему оптимальных методов - иначе говоря, найти потенциальные границы эффективности "всех мыслимых" численных методов и предъявить методы, "в существенном" реализующие эти границы.

Следует иметь в виду, что два указанных подхода к построению рекомендаций по выбору численных методов оптимизации не исключают друг друга. Если речь идет о задачах, для которых гарантированные теоретические оценки построения приближенного решения нужной точности укладываются в отведенные вычислительные ресурсы, то можно полностью положиться на теоретические рекомендации. Если же, как это часто случается, решаемая задача в указанном смысле слишком сложна, так что приходится экспериментировать с методами, модифицировать их, приспосабливать к ее специфике, то для успеха обычно требуется большая доза эмпирики. Однако и в этом случае полезно знать, чего можно добиться "теоретически гарантированным" образом.

Сказанное достаточно полно отражает общие тенденции в теории численной оптимизации. Как же эти тенденции отражены в монографической литературе? Имеется много отечественных и переводных книг "классической" ориентации, посвященных условиям оптимальности, описанию, обоснованию и анализу асимптотического поведения традиционных численных методов. В то же время проблематика, связанная с эффективными и оптимальными методами, представлена существенно слабее. Пожалуй, нельзя указать монографию, полностью посвященную этому кругу вопросов и суммирующую накопленные результаты, хотя соответствующие исследования продвинуты достаточно далеко и несомненно заслуживают систематического изложения.

Предалагемая вниманию читателя монография представляется вполне успешной попыткой восполнить этот пробел. Она посвящена изложению современного состояния теории численных методов выпуклой оптимизации, причем во главу угла в ней ставятся именно вопросы эффективности. Применительно к описанию основ используемого аппарата этот подход проявляется прежде всего в оригинальной, принадлежащей автору теории лексикографического дифференцирования. Нужды приложений давно уже потребовали расширения классического дифференциального исчисления на негладкие функции; имеется несколько вариантов такого рода обобщений (дифференциалы по Пшеничному, Демьянову, Кларку и т.п.). Особенность предлагаемого подхода в том, что здесь строится настоящее исчисление - производная, как и в классическом смысле, оказывается линейным отображением (а не множеством таких отображений) и дифференциал результата некоторой операции над функциями вполне конструктивно выражается через дифференциалы операндов по правилам, естественным образом обобщающим классические. В то же время класс отображений, на которые распространяется предложенная теория, оказывается весьма широким (он содержит гладкие в классическом смысле и все выпуклые функции, замкнут относительно стандартных алгебраических операций и суперпозиции). Развиваемая во второй главе техника лексикографического дифференцирования позволяет автоматически преобразовывать программу, вычисляющую функцию, в программу, вычисляющую ее вместе с производной,- и в этом отношении доставляет "информационное обеспечение" методам негладкой оптимизации.

Основное место в монографии занимают описание, обоснование и оценки трудоемкости эффективных численных методов решения задач выпуклого программирования. Рассмотрены основные классы таких задач - негладкие, гладкие с данными характеристиками гладкости, сильно выпуклые, задачи линейного и квадратичного программирования. Применительно к классам нелинейных задач описаны не просто эффективные, а оптимальные методы. Здесь стоит отметить, что в монографии впервые описывается принадлежащее автору окончательное решение проблемы оптимальных методов решения гладких выпуклых (и сильно выпуклых) задач большой размерности. Базовая версия предлагаемого метода по сложности одной итерации такая же, как у простейшего градиентного метода с постоянным шагом, тогда как трудоемкость отыскания решения заданной точности для нее в типичных ситуациях есть примерно квадратный корень из трудоемкости как градиентного, так и других традиционных методов гладкой выпуклой оптимизации.

Представляются весьма интересными и результаты, связанные с эффективными методами в линейном и квадратичном программировании. Здесь предложены оригинальные алгоритмы, у которых оценка трудоемкости (в числе арифметических операций) на указанных классах задач логарифмически зависит от требуемой точности решения и всего лишь кубически - от размеров задачи; напомним, что уже решение систем линейных уравнений методом Гаусса приводит к той же кубической зависимости числа операций от размеров системы. В связи с указанными результатами стоит отметить, что улучшение оценок "арифметической" (в числе операций) трудоемкости итеративных методов линейного программирования - это, пожалуй, наиболее бурно развивающийся в самые последние годы раздел в теории численной оптимизации; методы с кубическими оценками для задач линейного программирования (существенно отличающиеся от строящихся в монографии) были предложены в 1987 году, а для задач квадратичного программирования подобных методов ранее не было.

Несколько слов о характере изложения. Монография замкнута в себе и не предполагает от читателя знания математики в объеме, превышающем стандартный вузовский минимум. Заинтересованный читатель найдет в тексте полные доказательства всех утверждений, тогда как читатель с более практическими склонностями обнаружит, что все описанные в монографии численные методы доведены до явных, допускающих непосредственную программную реализацию, формул. Полагаю, что рекомендуемая книга будет с интересом и пользой прочитана самой широкой аудиторией лиц, интересующихся нынешним состоянием теории и численных методов оптимизации.

Профессор, доктор технических наук

Д. Б. Юдин

Введение

В математическом программировании, как и во многих других областях вычислительной математики, конечным результатом развития общей теории являются методы решения определенных классов задач. При этом важным показателем зрелости теории выступает перечень свойств, наличие которых у данного метода признается желательным. Расширение этого перечня неизбежно приводит к пересмотру сложившихся представлений о возможностях традиционных алгоритмов, стимулирует разработку новых вычислительных процедур.

К концу 70-х годов в теории оптимизации постепенно сформировалось новое направление, связанное с оценкой эффективности метода по его глобальной скорости сходимости. Если раньше исследователей прежде всего интересовал вопрос о том, сходится ли данный метод и какова оценка его скорости сходимости в окрестности точки минимума, то с этого времени стали все чаще появляться работы, в которых выводились оценки скорости сходимости существующих алгоритмов, "работавшие" с первой же итерации метода. Важным вкладом в развитие этого направления стало выявление А.С.Немировским и Д.Б.Юдиным нижних оценок сложности традиционных классов экстремальных задач. Созданная ими теория сложности дала мощный импульс для разработки новых вычислительных алгоритмов по двум причинам. Во-первых, были установлены потенциальные границы для скорости сходимости методов оптимизации на различных классах экстремальных задач, сформулировано понятие оптимального метода. А во-вторых, среди существовавших к тому времени алгоритмов лишь два (метод субградиентного спуска и метод центров тяжести) можно было отнести к оптимальным. К настоящему времени арсенал оптимальных методов удалось существенно расширить.

Подтверждением продуктивности нового направления стало обоснование Л.Г.Хачияном полиномиальной разрешимости задачи линейного программирования (ЛП), опиравшееся на глобальную оценку скорости сходимости метода эллипсоидов. С помощью методики Л.Г.Хачияна любой итеративный метод решения задачи ЛП, сходящийся со скоростью геометрической прогрессии, знаменатель которой зависит лишь от числа переменных, может быть превращен в полиномиальный алгоритм ЛП. С такой интерпретацией связан бурный прогресс в построении полиномиальных алгоритмов ЛП, наблюдавшийся в последние годы.

Результаты, связанные с проблематикой оптимальных алгоритмов нелинейного программирования, полиномиальных алгоритмов ЛП, в силу их новизны еще не успели найти отражения в монографической литературе. В настоящей монографии делается попытка восполнить этот пробел.

Остановимся вкратце на содержании монографии.

Первая глава носит вводный характер и содержит необходимые для дальнейшего сведения из выпуклого анализа. Рассматриваются традиционные классы гладких выпуклых, равномерно выпуклых и квазивыпуклых функций, исследуются их свойства. Вводится новый класс квазиоднородных функций, основная характеристика которых оказывается инвариантной относительно линейной замены переменных.

Во второй главе описывается аппарат дифференцирования негладких функций, позволяющий конструктивно вычислять дифференциальные характеристики, необходимые для методов негладкой оптимизации. Предлагается новый способ введения аналога производной негладкой функции, при котором таким аналогом оказывается линейный оператор - лексикографическая производная. Для однозначного определения лексикографической производной в точке достаточно зафиксировать произвольный базис в пространстве переменных. Класс функций, у которых можно вычислять лексикографические производные, - лексикографически гладкие функции, является линейным пространством, содержит все дифференцируемые, выпуклые и квазидифференцируемые по Демьянову-Рубинову функции и все суперпозиции лексикографически гладких функций. При этом для пересчета лексикографических производных таких функций при операции суперпозиции удается выписать формулы по своей простоте сравнимые с формулами классического дифференциального исчисления.

Лексикографические производные негладких функций естественным образом согласуются с ранее известными дифференциальными объектами. Так, лексикографический градиент гладкой функции совпадает с ее градиентом по Гато, лексикографический градиент выпуклой функций является крайней точкой ее субдифференциала, а лексикографический градиент квазивыпуклой функции - опорным функционалом к ее множеству уровня. Исследуются свойства лексикографически гладких функций, формулируются необходимые условия безусловного экстремума. Устанавливается справедливость формулы Ньютона-Лейбница для лексикографических производных по любому фиксированному базису в пространстве переменных. Демонстрируются возможности лексикографических производных как инструмента анализа негладких функций.

В третьей главе рассматриваются методы минимизации негладких выпуклых и квазивыпуклых функций: метод субградиентного спуска, метод центров тяжести, метод эллипсоидов. Обосновываются оценки их скорости сходимости. Описывается версия метода растяжения пространства в направлении градиента, применимого для минимизации квазиоднородных функций

Четвертая глава посвящена описанию оптимальных методов решения гладких экстремальных задач Рассматриваются оптимальные методы минимизации сильно выпуклых функций, обладающих липшицевым градиентом, метод минимизации функций с гельдеровым градиентом, автоматически настраивающийся на параметры гладкости целевой функции, метод минимизации негладких сверток гладких функций, обладающий оценкой скорости сходимости, характерной для гладкого случая, методы решения условных задач.

В пятой главе рассматриваются итеративные методы решения задач линейного и квадратичного программирования. Наряду с обоснованием полиномиальности метода внутенних штрафных функций, исследуются вычислительные схемы и новых методов: метода параллельных траекторий (прямого и двойственного) и полиномиального метода квадратичного программирования, основанного на использовании оптимальных алгоритмов минимизации гладких функций. Трудоемкость решения соответствующей задачи у лучших из этих методов оценивается произведением логарифма требуемой точности на полином третьей степени от размерности задачи.

В приложении, написанном А.С.Немировским, приведены основные результаты из теории сложности, относящиеся к рассмотренным постановкам задач.

ГЛАВА 1. ВЫПУКЛЫЕ МНОЖЕСТВА И ВЫПУКЛЫЕ ФУНКЦИИ

1.1. Выпуклые множества

Начнем с основных определений.

<u>Определение 1.1.1.</u> Множество X из пространства R^n называется выпуклым, если вместе с любыми двумя точками x и y из X весь отрезок

$$[x, y] \equiv \{ z \mid z = \alpha x + (1 - \alpha) y, \ 0 \le \alpha \le 1 \}$$

также принадлежит множеству X.

Простейшими примерами выпуклых множеств в пространстве R^n являются положительный ортант и евклидов шар:

$$R^n_+ \equiv \{ x \in R^n \mid x^{(i)} \ge 0, \ i = 1, \ldots, n \},$$

$$B(x, r) \equiv \{ z \in R^n \mid \| z - x \| \le r \}.$$

В дальнейшем нам будет удобно считать пустое множество выпуклым.

Нетрудно видеть, что если два множества X и Y являются выпуклыми, то и любое из множеств

$$X \cap Y,$$

$$X + Y \equiv \{ z \mid z = x + y, \ x \in X, \ y \in Y \},$$

$$\mathrm{Con}(X) \equiv \{ z \mid z = \alpha x, \ x \in X, \ \alpha \ge 0 \},$$

$$\mathrm{Conv}(X, Y) \equiv \{ z \mid z \in [x, y], \ x \in X, \ y \in Y \},$$

$$\mathrm{Lin}(X, Y) \equiv \{ z \mid z = \alpha x + (1 - \alpha) y, \ x \in X, \ y \in Y, \ \alpha \in R \},$$

также будет выпуклым.

Назовем проекцией точки x на множество X такую точку $p \equiv \pi(X, x)$, принадлежащую замыканию множества X, что

$\| x - p \| \leq \| x - y \|$

для любых y из X.

Теорема 1.1.1. Для любой точки x из R^n существует единственная проекция $\pi(X, x)$ этой точки на выпуклое множество X. При этом для всех y из замыкания множества X будет справедливо неравенство

$$\langle x - \pi(X, x), \pi(X, x) - y \rangle \geq 0. \qquad (1.1.1)$$

Доказательство. Зафиксируем произвольную точку x' из R^n. Пусть $\{ x_k \}$ - произвольная последовательность точек из X, минимизирующая функцию $\varphi(x) \equiv \equiv \| x - x' \|^2$ на множестве X:

$$\varphi(x_k) \to \varphi^* \equiv \inf \{ \| x - x' \|^2 \mid x \in X \}.$$

Заметим, что для любых номеров k, m справедливо неравенство

$$\varphi^* \leq \varphi(0.5\, x_k + 0.5\, x_m) =$$
$$= \| 0.5 (x' - x_k) + 0.5 (x' - x_m) \|^2 =$$
$$= 0.5 \| x' - x_k \|^2 + 0.5 \| x' - x_m \|^2 -$$
$$- 0.25 \| x_k - x_m \|^2 \equiv 0.5 (\varphi(x_k) + \varphi(x_m)) -$$
$$- 0.25 \| x_k - x_m \|^2$$

Из этого неравенства сразу получаем, что любая последовательность точек, минимизирующая функцию $\varphi(x)$, является фундаментальной. А это, в свою очередь, означает, что на замыкании множества X функция $\varphi(x)$ достигает минимума в единственной точке.

Нам осталось доказать справедливость неравенства (1.1.1). Зафиксируем произвольную точку $y \in X$ и число α из $(0, 1]$. Тогда

$$\varphi(\pi(X, x)) \leq \varphi(\pi(X, x) + \alpha(y - \pi(X, x))),$$

т.е.

$$\| \pi(X, x) - x' \|^2 \leq$$

$$< \| \pi(X, x) - x' + \alpha(y - \pi(X, x)) \|^2 =$$
$$= \| \pi(X, x) - x' \|^2 + 2\alpha \langle \pi(X, x) - x', y - \pi(X, x) \rangle +$$
$$+ \alpha^2 \| y - \pi(X, x) \|^2$$
Приводя подобные члены и сокращая на $\alpha > 0$, приходим к неравенству
$$0 < 2 \langle \pi(X, x) - x', y - \pi(X, x) \rangle + \alpha \| y - \pi(X, x) \|^2,$$
откуда в силу произвольности α получаем неравенство (1.1.1). В свою очередь, из справедливости неравенства (1.1.1) для всех точек y из множества X сразу следует справедливость его и для всех точек из замыкания множества X. □

Здесь и далее знаком □ отмечается конец доказательства.

<u>Следствие 1.1.1.</u> Для любых x, y из R^n выполняется неравенство:
$$\langle x - y, \pi(X, x) - \pi(X, y) \rangle >$$
$$> \| \pi(X, x) - \pi(X, y) \|^2 .$$

Д о к а з а т е л ь с т в о. Нетрудно видеть, что
$$\langle x - y, \pi(X, x) - \pi(X, y) \rangle -$$
$$- \| \pi(X, x) - \pi(X, y) \|^2 =$$
$$= \langle x - \pi(X, x) - y + \pi(X, y), \pi(X, x) - \pi(X, y) \rangle =$$
$$= \langle x - \pi(X, x), \pi(X, x) - \pi(X, y) \rangle +$$
$$+ \langle y - \pi(X, y), \pi(X, y) - \pi(X, x) \rangle .$$
Осталось заметить, что в силу неравенства (1.1.1) оба слагаемых в последней формуле неотрицательны. □

<u>Следствие 1.1.2.</u> Функция $\rho(X, x) = \| x - \pi(X, x) \|$ является липшицевой функцией от x:
$$| \rho(X, x) - \rho(X, y) | < \| x - y \|$$
для любых x, y из R^n.

Д о к а з а т е л ь с т в о. Действительно, в силу следствия 1.1.1 для любых x, y из R^n

$$| \rho(X, x) - \rho(X, y) |^2 \leq$$
$$\leq \| x - y - (\pi(X, x) - \pi(X, y)) \|^2 =$$
$$= \| x - y \|^2 - 2 \langle x - y, \pi(X, x) - \pi(X, y) \rangle +$$
$$+ \| \pi(X, x) - \pi(X, y) \|^2 \leq \| x - y \|^2 .$$

□

<u>Следствие 1.1.3.</u> Для любой точки x', не принадлежащей замыканию множества X, найдется гиперплоскость

$$\Gamma(a, \lambda) \equiv \{ y \in R^n \mid \langle a, y \rangle = \lambda \}, \ a \neq 0,$$

отделяющая эту точку от множества X, т.е. такая, что

$$\langle a, x' \rangle > \lambda > \langle a, y \rangle$$

для любого y из X.

Д о к а з а т е л ь с т в о. Заметим, что при сделанных предположениях $x' \neq \pi(X, x')$. Положим

$$a = x' - \pi(X, x'), \quad \lambda = \langle a, \pi(X, x') \rangle .$$

Тогда

$$\| a \|^2 \equiv \langle x' - \pi(X, x'), x' - \pi(X, x') \rangle > 0 .$$

Поэтому

$$\langle a, x' \rangle = \langle x' - \pi(X, x'), x' \rangle >$$
$$> \langle x' - \pi(X, x'), \pi(X, x') \rangle \equiv \lambda .$$

Далее в силу неравенства (1.1.1)

$$\lambda \equiv \langle x' - \pi(X, x'), \pi(X, x') \rangle >$$
$$> \langle x' - \pi(X, x'), y \rangle \equiv \langle a, y \rangle . \quad □$$

<u>Теорема 1.1.2.</u> Если точка x_0 является граничной точкой выпуклого множества X, то существует гиперплоскость $\Gamma(a, \lambda)$ такая, что

а) $x_0 \in \Gamma(a, \lambda)$;

б) $\Gamma(a, \lambda)$ отделяет точку x_0 от множества X

(такую гиперплоскость обычно называют опорной гиперплоскостью к множеству X в точке x_0).

Д о к а з а т е л ь с т в о. Рассмотрим последовательность точек $\{ x_k \}$, не принадлежащих замыканию множества X и сходящихся к граничной точке x_0 В силу следствия 1.1.3 этой последовательности можно поставить в соответствие последовательность векторов $\{ a_k \}$, $\| a_k \| = 1$, и последовательность чисел $\{ \lambda_k \}$ такие, что при всех y из X справедливо неравенство

$$\langle a_k , y \rangle < \lambda_k \equiv \langle a_k , \pi(X , x_k) \rangle . \qquad (1.1.2)$$

Не ограничивая общности, можно считать последовательность векторов $\{ a_k \}$ сходящейся. Обозначим предел этой последовательности через a^*. Нетрудно видеть, что последовательность точек $\{ \pi(X , x_k) \}$ также сходится (к точке x_0). Поэтому и последовательность чисел $\{ \lambda_k \}$ будет сходится к некоторому числу λ^*. Таким образом, переходя в неравенстве (1.1.2) к пределу, получаем

$$\langle a , y \rangle < \lambda^* \equiv \langle a^* , x_0 \rangle . \quad \square$$

<u>Теорема 1.1.3.</u> Пусть выпуклые замкнутые множества X и Y таковы, что $X \cap Y = \emptyset$, причем одно из этих множеств ограничено. Тогда существует гиперплоскость $\Gamma(a , \lambda)$, строго их разделяющая, т.е.

$$\langle a , x \rangle > \lambda > \langle a , y \rangle$$

для любых $x \in X$, $y \in Y$.

Д о к а з а т е л ь с т в о. Пусть для определенности ограниченным является множество X. Рассмотрим функцию $\varphi(x) = \rho(Y , x)$, $x \in X$. В силу следствия 1.1.2 эта функция будет непрерывной на X, и, следовательно, по теореме Вейерштрасса найдется точка $x_* \in X$, в которой $\varphi(x)$ достигает своего минимума на X. Положим

$$y_* = \pi(Y , x_*) , \quad a = x_* - y_* ,$$

$$\lambda = \langle a , 0.5\, x_* + 0.5\, y_* \rangle .$$

Из способа построения точки x_* следует, что $x_* = \pi(X, y_*)$. Кроме того, из сделанных предположений ясно, что $x_* \neq y_*$. Поэтому в силу теоремы 1.1.1 имеем

$$\langle a, x \rangle = \langle x_* - y_*, x \rangle \geq \langle x_* - y_*, x_* \rangle =$$
$$= \lambda + 0.5 \| x_* - y_* \|^2 > \lambda$$

для всех x из X;

$$\langle a, y \rangle = \langle x_* - y_*, y \rangle \leq \langle x_* - y_*, y_* \rangle =$$
$$= \lambda - 0.5 \| x_* - y_* \|^2 < \lambda$$

для всех y из Y. □

В заключение этого параграфа приведем удобный критерий ограниченности выпуклых множеств.

Теорема 1.1.4. Пусть выпуклое множество X не является ограниченным. Тогда существует направление $s \in R^n$ такое, что для любой точки x', принадлежащей относительной внутренности множества X, луч

$$L(x', s) = \{ z \mid z = x' + \alpha s, \alpha \geq 0 \}$$

содержится в множестве X. Если же x' - произвольная точка из X, то луч $L(x', s)$ содержится в замыкании множества X.

Доказательство. Для упрощения рассуждений будем считать, что множество X имеет непустую внутренность. Из условия теоремы следует, что найдется последовательность точек $\{ x_k \}_{k=1}^{\infty}$, $x_k \in X$, таких, что $\| x_k \| \to \infty$ при $k \to \infty$. При этом мы можем считать, что $\| x_k \| \geq 1$ для всех $k \geq 1$. Положим

$$s_k = x_k / \| x_k \|, \quad k \geq 1.$$

Все точки s_k принадлежат единичному шару с центром в нуле. Поэтому множество предельных точек последовательности $\{ s_k \}$ не пусто. Пусть s_* - одна из предельных точек этой последовательности. Не ограничивая общности, будем считать, что такая точка единственная.

Пусть x' - точка из множества X такая, что шар $B(x_0, \varepsilon)$ содержится в X при некотором $\varepsilon > 0$. Покажем, что $L(x', s^*) \subseteq X$. Для этого зафиксируем произвольное $R > 0$ и положим

$$y' = x' + R s^*,$$

$$\varphi_k = \min\{ \| y' + \alpha (y' - x_k) \|^2 \mid \alpha > 0 \}.$$

Нетрудно видеть, что найдется номер N такой, что при всех $k > N$ справедливо неравенство $\langle s_*, x_k - y' \rangle > 0$. Следовательно, при всех $k > N$

$$\varphi_k = \| y' + \alpha_k^* (y' - x_k) \|^2,$$

где $\alpha_k^* = \langle s_*, x_k - y' \rangle / \| y' - x_k \|^2 > 0$,

т.е.

$$\varphi_k = R^2 \| s^* \|^2 - R^2 \langle s_*, x_k - y' \rangle^2 / \| y' - x_k \|^2.$$

Из последней формулы получаем, что $\varphi_k \to 0$ при $k \to \infty$.

Таким образом, мы показали, что при некотором $k > N$ выполнится неравенство $\varphi_k < \varepsilon^2$. Но это означает, что точка

$$z' = y' + \alpha_k^* (y' - x_k) \in B(x_0, \varepsilon) \subseteq X,$$

и, следовательно,

$$y' = (1 + \alpha_k^*)^{-1} z' + \alpha_k (1 + \alpha_k^*)^{-1} x_k \in X.$$

Пусть теперь x' - произвольная точка из множества X. Рассмотрим последовательность точек

$$y'_k = x' + \alpha_k (x_k - x'), \ k = N, \ldots,$$

где $\alpha_k = \langle s^*, y' - x' \rangle / \langle s_*, x_k - x' \rangle$. Нетрудно видеть, что $\alpha_k \in (0, 1)$ при всех $k > N$. Поэтому $y'_k \in$ $\in X$, $k > N$. В то же время

$$\| y' - y'_k \| = \| R s^* - \alpha_k (x_k - x') \| =$$
$$= \| R s^* - R \langle s^*, x_k - x' \rangle^{-1} (x_k - x') \| \to 0$$

при $k \to \infty$. □

Следствие 1.1.4. Для ограниченности выпуклого множества X достаточно, чтобы оно было ограничено по любому лучу $L(x', s)$, выходящему из некоторой точки x', принадлежащей относительной внутренности множества X. Если к тому же множество X замкнуто, то точка x' может быть любой точкой множества X.

1.2. Выпуклые функции

Обозначим через $\operatorname{dom} f$ множество точек в R^n, в которых функция $f(x)$ определена (это множество часто называют эффективной областью функции $f(x)$). Введем также понятие надграфика функции $f(x)$:

$$\operatorname{epi} f \equiv \{ (x, t) \mid f(x) \leq t \}.$$

Определение 1.2.1. Функция $f(x)$ называется выпуклой, если ее надграфик - выпуклое множество.

Удобно считать, что во всех точках x, $x \notin \operatorname{dom} f$, функция $f(x)$ принимает значение $+\infty$, причем для этого значения арифметические операции и операции сравнения выполняются по следующим правилам: при всех t из R

$$t < \infty; \quad \max\{t, \infty\} = \infty; \quad \infty \leq \infty;$$
$$t + \infty = \infty; \quad \infty + \infty = \infty;$$
$$t \times \infty = \infty \text{ при } t > 0; \quad 0 \times \infty = 0.$$

Такие соглашения позволяют сформулировать следующую характеристическую лемму.

Лемма 1.2.1. Функция $f(x)$ является выпуклой в том и только в том случае, если при всех $x, y \in R^n$ и любом α из $[0, 1]$ выполняется неравенство

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha)f(y) \qquad (1.2.1)$$

Следствие 1.2.1 (неравенство Иенсена). Если функция $f(x)$ является выпуклой, то для любого набора точек x_1, ..., x_k из R^n и чисел $\alpha_i \in [0, 1]$, $\sum_{i=1}^k \alpha_i = 1$, справедливо неравенство

$$f\left(\sum_{i=1}^k \alpha_i x_i\right) \le \sum_{i=1}^k \alpha_i f(x_i).$$

Приведем некоторые очевидные свойства выпуклых функций.

1. Если функции $f_1(x)$ и $f_2(x)$ являются выпуклыми, то и их взвешенная сумма $\alpha f_1(x) + \beta f_2$ с положительными весами также будет выпуклой функцией, причем

$$\mathrm{dom}(\alpha f_1 + \beta f_2) = \mathrm{dom}\, f_1 \cap \mathrm{dom}\, f_2$$

2. Если функции $f_1(x)$ и $f_2(x)$ являются выпуклыми, то и функция $\max\{f_1(x), f_2(x)\}$ также будет выпуклой функцией, причем

$$\mathrm{dom}\,\max\{f_1, f_2\} = \mathrm{dom}\, f_1 \cap \mathrm{dom}\, f_2.$$

3. Множество уровня $\mathcal{L}(f, t) \equiv \{x \mid f(x) \le t\}$ выпуклой функции $f(x)$ выпукло при любом t из R.

В некоторых случаях бывает удобно пользоваться другим критерием выпуклости.

Лемма 1.2.2. Функция $f(x)$ является выпуклой тогда и только тогда, когда при всех x, y из R^n и $\lambda \ge 0$ справедливо неравенство

$$f(y + \lambda(y - x)) \ge f(y) + \lambda(f(y) - f(x)). \qquad (1.2.2)$$

Доказательство. Пусть при всех x, y из R^n и $\alpha \in [0,1]$ справедливо неравенство (1.2.1). Зафиксируем произвольные u, v из R^n и $\lambda \ge 0$. Положим

$$y = u + \lambda(u - v); \quad \alpha = \lambda(1+\lambda)^{-1}; \quad x = v.$$

Нетрудно видеть, что $u = (1-\alpha)y + \alpha x$. Поэтому

$$f(u) \le (1-\alpha) f(y) + \alpha f(x) =$$
$$= (1+\lambda)^{-1} f(u + \lambda(u - v)) + \lambda(1+\lambda)^{-1} f(v).$$

т.е.

$$f(u + \lambda(u - v)) \geq f(u) + \lambda(f(u) - f(v)).$$

Обратно, пусть неравенство (1.2.2) справедливо при всех x, y из R^n и $\lambda \geq 0$. Зафиксируем произвольные $u, v \in R^n$ и α из $(0, 1]$. Положим

$$y = \alpha u + (1 - \alpha) v; \quad x = v; \quad \lambda = \alpha^{-1} - 1.$$

Заметим, что в этом случае $u = y + \lambda(y - x)$. Поэтому

$$f(u) \equiv f(y + \lambda(y - x)) \geq f(y) + \lambda(f(y) - f(x)) =$$
$$= \alpha^{-1} f(\alpha u + (1 - \alpha) v) - (\alpha^{-1} - 1) f(v),$$

т.е. $f(\alpha u + (1 - \alpha) v) \leq \alpha f(u) + (1 - \alpha) f(v)$.

□

Остановимся на наиболее важных свойствах выпуклых функций.

Теорема 1.2.1. Выпуклая функция непрерывна в любой внутренней точке своей эффективной области. Более того, для любого ограниченного замкнутого множества $Q \subset \operatorname{int}(\operatorname{dom} f)$ и любой точки $x_0 \in \operatorname{int} Q$ при всех y из Q справедливо неравенство

$$| f(y) - f(x_0) | \leq \varepsilon^{-1} M \| y - x_0 \|, \tag{1.2.3}$$

где $M = \max\{ f(x) \mid x \in Q \} - \min\{ f(x) \mid x \in Q \}$;

$\varepsilon = \max\{ r \mid B(x_0, r) \in Q \}$.

Доказательство. Ясно, что первая часть утверждения теоремы вытекает из второй при условии, что функция $f(x)$ ограничена на множестве Q. Покажем, что это так. В силу компактности Q нам достаточно доказать ограниченность функции $f(x)$ в окрестности любой точки $x_0 \in \operatorname{int}(\operatorname{dom} f)$. Действительно, при достаточно малом $\nu > 0$ точка x_0 будет внутренней точкой множества

$$S_\nu = x_0 + \operatorname{Conv}\{ s_i, i = 1, \ldots, 2n \},$$

где $s_i = \nu e_i$ при $i = 1, \ldots, n$, $s_i = -\nu e_i$ при

$i = n+1, \ldots, 2n$, e_i - i-й координатный вектор пространства R^n. Поэтому в силу неравенства Иенсена

$$\max\{f(x) \mid x \in S_y\} \le \max\{f(s_i) \mid i = 1, \ldots, 2n\} < \infty.$$

Пусть теперь y - произвольная точка из Q. Выберем точку z из Q так, чтобы $y = (1-\alpha)x_0 + \alpha z$ при некотором $\alpha \in [0, 1]$ и $\| z - x_0 \| \ge \varepsilon$. Тогда в силу леммы 1.2.1

$$f(y) - f(x_0) \le \alpha (f(z) - f(x_0)) \le \alpha M.$$

Заметим, что

$$\alpha = \| y - x_0 \| / \| z - x_0 \| \le \varepsilon^{-1} \| y - x_0 \|.$$

Поэтому

$$f(y) - f(x_0) \le \varepsilon^{-1} M \| y - x_0 \|.$$

С другой стороны, точку z можно выбрать таким образом, чтобы $y = x_0 + \lambda(x_0 - z)$, $\lambda \ge 1$, и $\| z - x_0 \| \ge \varepsilon$. Тогда в силу леммы 1.2.2

$$f(y) - f(x_0) \ge \lambda (f(x_0) - f(z)) \ge -\lambda M.$$

Заметим, что

$$\lambda = \| y - x_0 \| / \| z - x_0 \| \le \varepsilon^{-1} \| y - x_0 \|.$$

Объединяя полученные неравенства, приходим к неравенству (1.2.3). □

З а м е ч а н и я. 1. При доказательстве теоремы мы нигде не пользовались евклидовостью нормы пространства R^n. Утверждение теоремы остается в силе и для любой другой нормы.

2. Формулировку теоремы можно изменить таким образом, чтобы неравенство (1.2.3) осталось верным и для случая $\mathrm{int}(\mathrm{dom}\, f) = \emptyset$. Для этого надо лишь перейти к относительной внутренности множества $\mathrm{dom}\, f$.

Следствие 1.2.2. Выпуклая функция $f(x)$ удовлетворяет условию Липшица на любом выпуклом замкнутом ограниченном подмножестве множества $\operatorname{int}(\operatorname{dom} f)$.

Доказательство. Действительно, пусть $Q \subset \operatorname{int}(\operatorname{dom} f)$. Тогда найдется $\varepsilon > 0$ такое, что множество $Q' = Q + B(0, \varepsilon) \subset \operatorname{int}(\operatorname{dom} f)$. Поэтому в силу теоремы 1.2.1 функция $f(x)$ будет удовлетворять условию Липшица на множестве Q с константой
$L = \varepsilon^{-1} [\max \{ f(x) \mid x \in Q' \} - \min \{ f(x) \mid x \in Q' \}]$.
□

Следствие 1.2.3. Пусть $\operatorname{dom} f$ - замкнутое множество. Тогда если при некотором t множество $\mathcal{L}(f ; t)$ принадлежит $\operatorname{int}(\operatorname{dom} f)$, то это множество будет замкнутым.

Действительно, в силу следствия 1.2.2 функция $f(x)$ будет непрерывна на $\mathcal{L}(f, t)$, и следовательно, это множество замкнуто. □

Теорема 1.2.2. Пусть множество $\mathcal{L}(f, t)$ замкнуто при всех t из R. Тогда, если найдется число t_0, для которого множество уровня $\mathcal{L}(f, t_0)$ выпуклой функции $f(x)$ ограничено, то такие множества будут ограниченными и при всех t из R.

Доказательство. Для всех $t' < t_0$ утверждение теоремы тривиально. Пусть $t' > t_0$. В силу следствия 1.1.4 нам достаточно доказать ограниченность множества $\mathcal{L}(f, t')$ по любому лучу, выходящему из произвольной точки $x_0 \in \mathcal{L}(f, t_0)$. Действительно, пусть s - произвольное направление в R^n. Рассмотрим луч
$L(x_0, s) = \{ y \mid y = y(\lambda) = x_0 + \lambda s, \lambda > 0 \}$
В силу ограниченности множества $\mathcal{L}(f, t_0)$ и непрерывности функции $f(x)$ на множестве $\mathcal{L}(f, t')$ найдется точка $x_1 = y(\lambda_1)$ такая, что $f(x_0) < f(x_1) < t'$. Тогда в силу леммы 1.2.2 при всех $\lambda > \lambda_1$ будет спра-

ведливо неравенство

$$f(y(\lambda)) > f(x_1) + (\lambda - \lambda_1)(f(x_1) - f(x_0))/(\lambda_1 - \lambda_0).$$

Таким образом, при

$$\lambda > \lambda_1 + (\lambda_1 - \lambda_0)(t' - f(x_1))/(f(x_1) - f(x_0))$$

точки $y(\lambda)$ заведомо выйдут из множества $\mathcal{M}(f, t')$.

□

Требование замкнутости множеств $\mathcal{M}(f, t)$ в теореме 1.2.2 существенно, что видно из следующего примера.

Пример 1.2.1. Рассмотрим функцию $f(x^{(1)}, x^{(2)})$, заданную на множестве $\{ x \in R^2 \mid x^{(1)} \geq 0 \}$ следующим образом:

$f(x) = \| x \| - x^{(2)}$ при $x^{(1)} > 0$;

$f(x) = 2 \mid x^{(2)} \mid$ при $x^{(1)} = 0$.

Нетрудно убедиться в том, что $\mathcal{M}(f, 0) \equiv \{ 0 \}$, однако все множества $\mathcal{M}(f, t)$ при $t > 0$ не ограничены. □

Ведем понятие производной функции $f(x)$ по направлению:

$$f'(x_0, s) = \lim_{\alpha \to +0} \alpha^{-1} [f(x_0 + \alpha s) - f(x_0)].$$

Будем говорить, что функция $f(x)$ дифференцируема в точке x_0 по направлению s, если этот предел существует.

Теорема 1.2.3. Выпуклая функция дифференцируема по направлениям в любой внутренней точке своей эффективной области.

Д о к а з а т е л ь с т в о. Зафиксируем произвольное направление $s \in R^n$, $s \neq 0$. Пусть $\varepsilon > 0$ такое, что $x_0 + \alpha s \in \mathrm{dom}\, f$ при всех α из $[0, \varepsilon]$. Рассмотрим функцию

$$\varphi(\alpha) = \alpha^{-1} [f(x_0 + \alpha s) - f(x_0)],\ \alpha \in (0, \varepsilon].$$

Пусть $\alpha_1, \alpha_2 \in (0, \varepsilon]$, $\alpha_1 > \alpha_2$. Тогда в силу леммы 1.2.2 имеем

$$\varphi(\alpha_1) = \alpha_1^{-1}[f(x_0 + \alpha_1 s) - f(x_0)] =$$
$$= \alpha_1^{-1}[f(x_0 + \alpha_2 s + \alpha_2^{-1}(\alpha_1 - \alpha_2) \times$$
$$\times (x_0 + \alpha_2 s - x_0)) - f(x_0)] \geq$$
$$\geq \alpha_1^{-1}[f(x_0 + \alpha_2 s) + \alpha_2^{-1}(\alpha_1 - \alpha_2) \times$$
$$\times (f(x_0 + \alpha_2 s) - f(x_0)) - f(x_0)] =$$
$$= \alpha_2^{-1}[f(x_0 + \alpha_2 s) - f(x_0)] = \varphi(\alpha_2).$$

Таким образом, функция $\varphi(\alpha)$ убывает при $\alpha \to +0$. Следовательно, существует предел $\lim\limits_{\alpha \to +0} \varphi(\alpha) \equiv f'(x_0, s)$.

□

<u>Теорема 1.2.4.</u> Пусть точка x_0 принадлежит внутренности эффективной области выпуклой функции $f(x)$. Тогда функция $f'(x, s)$ является выпуклой положительно однородной (первой степени) функцией своего второго аргумента, причем при всех y из R^n будет справедливо неравенство

$$f(y) \geq f(x_0) + f'(x_0, y - x_0).$$

Д о к а з а т е л ь с т в о. Действительно, при любом $s \in R^n$ и $\lambda > 0$

$$f'(x_0, \lambda s) = \lim_{\alpha \to +0} \alpha^{-1}[f(x_0 + \lambda \alpha s) - f(x_0)] =$$
$$= \lambda \lim_{\alpha \to +0} \alpha^{-1}[f(x_0 + \alpha s) - f(x_0)] = \lambda f'(x_0, s).$$

Далее, для любых $s_1, s_2 \in R^n$ и $\lambda \in [0, 1]$

$$f'(x_0, \lambda s_1 + (1 - \lambda) s_2) =$$
$$= \lim_{\alpha \to +0} \alpha^{-1}[f(x_0 + \alpha(\lambda s_1 + (1 - \lambda) s_2)) - f(x_0)] \leq$$
$$\leq \lim_{\alpha \to +0} \alpha^{-1}\{\lambda[f(x_0 + \alpha s_1) - f(x_0)] +$$
$$+ (1 - \lambda)[f(x_0 + \alpha s_2) - f(x_0)]\} =$$
$$= \lambda f'(x_0, s_1) + (1 - \lambda) f'(x_0, s_2).$$

И, наконец, для любого $\alpha \in (0, 1]$

$$f(y) = f(\alpha y + (1-\alpha)x_0 + (1-\alpha)(y - x_0)) =$$
$$= f(\alpha y + (1-\alpha)x_0 + \alpha^{-1}(1-\alpha)(\alpha y + (1-\alpha)x_0 - x_0)) \geq$$
$$\geq f(\alpha y + (1-\alpha)x_0) + \alpha^{-1}(1-\alpha)[f(\alpha y + (1-\alpha)x_0) - f(x_0)].$$

Переходя в последнем соотношении к пределу при $\alpha \to +0$, получаем искомое неравенство. □

Выпуклые функции, вообще говоря, не являются дифференцируемыми (простейший пример - $\| x \|$). Однако для них удается ввести дифференциальные характеристики, аналогичные по своим свойствам градиенту гладкой функции.

<u>Определение 1.2.2.</u> Вектор $g \in R^n$, для которого при всех y из R^n выполняется неравенство

$$f(y) \geq f(x_0) + \langle g, y - x_0 \rangle$$

называется субградиентом функции $f(x)$ в точке x_0 Множество всех субградиентов функции $f(x)$ в точке x_0 называется субдифференциалом этой функции в x_0 и обозначается через $\partial f(x_0)$.

Укажем сразу на важное следствие из этого определения

Если $\partial f(x_0) \neq \emptyset$, то функция $f(x)$ полунепрерывна снизу в точке x_0.

Действительно, для любой последовательности $\{ y_k \}$, такой, что $y_k \to x_0$ при $k \to \infty$, имеем

$$\sup\lim_{k \to \infty} f(y_k) \geq f(x_0) + \lim_{k \to \infty} \langle g, y_k - x_0 \rangle = f(x_0).$$

<u>Теорема 1.2.5.</u> Субдифференциал выпуклой функции $f(x)$ в любой внутренней точке эффективной области является непустым выпуклым замкнутым ограниченным множеством.

Д о к а з а т е л ь с т в о. Сначала докажем непустоту множества $\partial f(x_0)$. Рассмотрим надграфик

$$\mathrm{epi}\, f = \{ (x, t) \mid f(x) \leq t \}.$$

Точка $(x_0, f(x_0))$ является граничной точкой этого множества. Поэтому в силу теоремы 1.1.2 в этой точке существует опорный к $\operatorname{epi} f$ вектор (a, α), такой, что $\langle (a, \alpha), (x, t) \rangle \le \lambda = \langle (a, \alpha), (x_0, f(x_0)) \rangle$ для всех (x, t) из $\operatorname{epi} f$. В частности, для всех точек $(x, f(x))$ будет справедливо неравенство

$$\langle a, x \rangle + \alpha f(x) \le \langle a, x_0 \rangle + \alpha f(x_0),$$

или

$$-\alpha f(x) \ge -\alpha f(x_0) + \langle a, x - x_0 \rangle .$$

Заметим, что при всех $t \ge f(x_0)$ точка $(x_0, t) \in \operatorname{epi} f$ Поэтому

$$\langle a, x_0 \rangle + \alpha t \le \langle a, x_0 \rangle + \alpha f(x_0)$$

при всех $t \ge f(x_0)$ Отсюда следует, что $\alpha \le 0$. И, наконец, из локальной липшицевости функции $f(x)$ в окрестности точки x_0 (см. следствие 1.2.2) нетрудно получить, что $\alpha \ne 0$ Таким образом, вектор $-\alpha^{-1} a \in \partial f(x_0)$. Ограниченность множества $\partial f(x_0)$ непосредственно вытекает из неравенства (1.2.3). Осталось заметить, что выпуклость и замкнутость субдифференциала $\partial f(x_0)$ сразу следует из его определения. □

<u>Следствие 1.2.4.</u> Выпуклая функция ограничена снизу на любом выпуклом ограниченном множестве.

Доказательство. Действительно, пусть выпуклая функция $f(x)$ определена на выпуклом ограниченном множестве $X \subseteq \operatorname{dom} f$. Предположим, что $\operatorname{int} X \ne \emptyset$ (если это не так, то все рассуждения надо вести в относительной внутренности множества X). Тогда в силу теоремы 1.2.5 $\partial f(x) \ne \emptyset$ для любой точки $x \in \operatorname{int} X$. Поэтому при всех y из X имеем

$$f(y) \ge f(x') + \langle g_x, y - x \rangle \ge$$
$$\ge f(x) - \| g_x \| \operatorname{diam} X ,$$

где g_x - произвольный вектор из $\partial f(x)$. □

Итак, для любых точек $x, y \in R^n$ и вектора $g_x \in \partial f(x)$ справедливо неравенство

$$f(y) \ge f(x) + \langle g_x , y - x \rangle$$

Если к тому же $g_y \in \partial f(y) \ne \emptyset$, то

$$f(x) \ge f(y) + \langle g_y , x - y \rangle .$$

Складывая эти два неравенства, получаем

$$\langle g_x - g_y , x - y \rangle \ge 0 . \qquad (1.2.4)$$

Нетрудно убедиться в том, что величина, стоящая в левой части неравенства (1.2.4), ограничивает "прогиб" функции на отрезке $[x , y]$. Действительно, зафиксируем произвольное $\alpha \in [0 , 1]$ и положим $z = \alpha x + (1 - \alpha) y$. Тогда

$$f(z) \ge f(x) + \langle g_x , z - x \rangle = f(x) + (1 - \alpha)\langle g_x , y - x \rangle,$$

$$f(z) \ge f(y) + \langle g_y , z - y \rangle = f(y) + \alpha \langle g_y , x - y \rangle .$$

Умножая первое из этих неравенств на α , второе - на $1 - \alpha$ и складывая их, получаем

$$0 \le \alpha f(x) + (1 - \alpha) f(y) - f(\alpha x + (1 - \alpha) y) \le$$
$$\le \alpha (1 - \alpha) \langle g_x - g_y , x - y \rangle . \qquad (1.2.5)$$

Справедливо следующее достаточное условие выпуклости функции $f(x)$.

Теорема 1.2.6. Если в каждой точке x эффективной области функции $f(x)$ существует вектор g_x такой, что

$$f(y) \ge f(x) + \langle g_x , y - x \rangle$$

для всех $y \in R^n$, то функция $f(x)$ выпукла.

Д о к а з а т е л ь с т в о. Выберем произвольные x , $y \in \mathrm{dom}\, f$ и α из $[0 , 1]$. Тогда в точке $z = \alpha x + (1 - \alpha) y$ найдется вектор g_z такой, что

$$f(y) \ge f(z) + \langle g_z , y - z \rangle = f(z) + \alpha \langle g_z , y - x \rangle ,$$

$$f(x) \ge f(z) + \langle g_z , x - z \rangle = f(z) + (1 - \alpha)\langle g_z , x - z \rangle$$

Умножая эти неравенств на α и $1 - \alpha$ соответственно и складывая их, получаем

$$\alpha f(x) + (1 - \alpha) f(y) \geq f(z) = f(\alpha x + (1 - \alpha)y).$$

□

Приведем примеры субдифференциалов простейших функций.

Пример 1.2.2. Рассмотрим функцию $f(x) = \| x \|$ Субдифференциал этой функции в нуле будет состоять из векторов g таких, что $\| x \| \geq \langle g, x \rangle$ для любых x из R^n Несложно убедиться в том, что это множество есть

$$\{ g \in R^n \mid \| g \| = 1 \} \equiv \partial f(0)$$

Если же $x \neq 0$, то $\partial f(x) = \{ x / \| x \| \}$. □

Пример 1.2.3. Пусть $f(x) = \langle c, x \rangle$. В этом случае субградиенты g функции $f(x)$ в точке x должны удовлетворять неравенству

$$\langle c - g, y \rangle \geq \langle c - g, x \rangle$$

при всех y из R^n. Но это возможно лишь при $g = c$. Таким образом, $\partial f(x) = \{ c \}$. □

Между производными по направлению и субдифференциалом выпуклой функции существует простая связь. Прежде чем приводить формулировку соответствующего утверждения, докажем две вспомогательные леммы.

Лемма 1.2.3. Если выпуклые функции $f_1(x)$ и $f_2(x)$ таковы, что $f_1(x_0) = f_2(x_0)$, где $x_0 \in \operatorname{int}(\operatorname{dom} f_1) \cap \cap \operatorname{int}(\operatorname{dom} f_2)$, и $f_1(x) \geq f_2(x)$ при всех x из R^n, то $\partial f_2(x_0) \subseteq \partial f_1(x_0)$.

Действительно, если $g \in \partial f_2(x_0)$, то при любом y из R^n имеем

$$f_1(y) \geq f_2(y) \geq f_2(x_0) + \langle g, y - x_0 \rangle =$$

$$= f_1(x_0) + \langle g, y - x_0 \rangle. \quad \square$$

Лемма 1.2.4. Пусть $f(x)$ - выпуклая положительно однородная функция с $\operatorname{dom} f \equiv R^n$. Тогда для любого x из R^n

1) $f(x) \equiv \max\{ \langle g, x \rangle \mid g \in \partial f(0) \}$;

2) $\partial f(x) \equiv \mathrm{Argmax}\{ \langle g, x \rangle \mid g \in \partial f(0) \}$;

3) $f(x) = \langle g_x, x \rangle$, где g_x - произвольный вектор из $\partial f(x)$.

Доказательство. Зафиксируем произвольную точку x из R^n. Тогда для любого вектора g_x из $\partial f(x)$ имеем

$f(2x) \equiv 2 f(x) \geq f(x) + \langle g_x, x \rangle$,

$0 = f(0) \geq f(x) + \langle g_x, -x \rangle$.

Таким образом, $f(x) = \langle g_x, x \rangle$ для любого g_x из $\partial f(x)$. Далее, для любого y из R^n

$f(y) \geq f(x) + \langle g_x, y - x \rangle = f(0) + \langle g_x, y \rangle$.

Следовательно, $\partial f(x) \subseteq \partial f(0)$. Теперь заметим, что

$\langle g_x, x \rangle = f(x) \geq \langle g, x \rangle$

для любого g из $\partial f(0)$. Но это означает, что

$\partial f(x) \equiv \mathrm{Argmax}\{ \langle g, x \rangle \mid g \in \partial f(0) \}$

и

$f(x) \equiv \max\{ \langle g, x \rangle \mid g \in \partial f(x) \} =$

$= \max\{ \langle g, x \rangle \mid g \in \partial f(0) \}$. □

Теорема 1.2.7. Если точка x_0 принадлежит внутренности эффективной области выпуклой функции $f(x)$, то

1) $f'(x_0, s) \equiv \max\{ \langle g, s \rangle \mid g \in \partial f(x_0) \}$;

2) $\partial_s f'(x_0, 0) \equiv \partial f(x_0)$.

Доказательство. Заметим, что для любого $g \in \partial f(x_0)$

$$f'(x_0, s) = \lim_{\alpha \to +0} \alpha^{-1} [f(x_0 + \alpha s) - f(x_0)] \geq \langle g, s \rangle.$$

Поэтому $\partial f(x_0) \subseteq \partial_s f'(x_0, 0)$. С другой стороны, в силу теоремы 1.2.4 $f(x) \geq f(x_0) + f'(x_0, x - x_0)$. Поэтому, воспользовавшись леммой 1.2.3, получаем

$\partial f(x_0) \equiv \partial_s f'(x_0 , 0)$.

Осталось заметить, что в силу леммы 1.2.4

$f'(x_0 , s) \equiv \max \{ \langle g , s \rangle \mid g \in \partial_s f'(x_0 , 0) \} =$

$= \max \{ \langle g , s \rangle \mid g \in \partial f(x_0) \}$. □

Следствие 1.2.5. Если выпуклая функция $f(x)$ дифференцируема в точке x_0 , то $\partial f(x_0) \equiv \{ f'(x_0) \}$.

Действительно, если функция $f(x)$ дифференцируема в точке x_0 , то $f'(x_0 , s) \equiv \langle f'(x_0) , s \rangle$. Поэтому, в силу теоремы 1.2.6

$\partial f(x_0) \equiv \partial_s f'(x_0 , 0) \equiv \{ f'(x_0 \}$. □

Теорема 1.2.8. Если у выпуклой функции $f(x)$ существуют две точки $x , y \in \operatorname{int} \operatorname{dom} f$, такие, что при любом $\alpha \in [0, 1]$

$f(\alpha x + (1 - \alpha) y) = \alpha f(x) + (1 - \alpha) f(y)$,

то при всех $\alpha \in (0 , 1)$

$\partial f(\alpha x + (1 - \alpha) y) = \partial f(x) \cap \partial f(y) \neq \emptyset$.

Доказательство. Зафиксируем произвольное $\alpha \in (0 , 1)$. Из теоремы 1.2.7 следует, что

$f'(\alpha x + (1 - \alpha) y , x - y) = f(x) - f(y) =$

$= \max \{ \langle g , x - y \rangle \mid g \in \partial f(\alpha x + (1 - \alpha) y) \}$;

$f'(\alpha x + (1 - \alpha) y , y - x) = f(y) - f(x) =$

$= \max \{ \langle g , y - x \rangle \mid g \in \partial f(\alpha x + (1 - \alpha) y) \}$.

Таким образом, $\langle g , x - y \rangle = f(x) - f(y)$ для любого вектора g из $\partial f(\alpha x + (1 - \alpha) y)$. Следовательно, для любого $z \in \operatorname{dom} f$, $g \in \partial f(\alpha x + (1 - \alpha) y)$ имеем

$0 \le f(z) - f(\alpha x + (1 - \alpha)y) - \langle g, z - \alpha x - (1 - \alpha)y\rangle =$

$= f(z) - f(x) - \langle g , z - x \rangle$,

т.е. $\partial f(\alpha x + (1 - \alpha) y) \subseteq \partial f(x)$. Аналогично можно показать, что $\partial f(\alpha x + (1 - \alpha) y) \subseteq \partial f(y)$. Следовательно,

$\emptyset \neq \partial f(\alpha x + (1-\alpha) y) \subseteq \partial f(x) \cap \partial f(y)$.

Докажем обратное включение. Пусть теперь $g \in \partial f(x) \cap \partial f(y)$. Тогда для любого $z \in \operatorname{dom} f$ имеем

$f(z) \geq f(x) + \langle g, z - x \rangle, \quad f(z) \geq f(y) + \langle g, z - y \rangle$.

Складывая эти неравенства с весами α и $1-\alpha$ соответственно, получаем

$f(z) \geq \alpha f(x) + (1-\alpha) f(y) + \langle g, z - \alpha x - (1-\alpha) y \rangle =$
$= f(\alpha x + (1-\alpha) y) + \langle g, z - \alpha x - (1-\alpha) y \rangle$,

то есть $g \in \partial f(\alpha x + (1-\alpha) y)$ при любом $\alpha \in [0, 1]$. □

В заключение мы обоснуем формулы преобразований субдифференциалов при различных операциях над выпуклыми функциями.

Лемма 1.2.5. Пусть функция $F(u)$, $u \in R^m$, является локально липшицевой и дифференцируемой по направлениям в точке $f(x_0) = (f^{(1)}(x_0), \ldots, f^{(m)}(x_0))$, а функции $f^{(1)}(x), \ldots, f^{(m)}(x)$ дифференцируемы по направлению s в точке x_0. Тогда функция $\varphi(u) = F(f(x))$ дифференцируема по направлению s в точке x_0, причем справедлива формула

$\varphi'(x_0, s) = F'(f(x_0), f'(x_0, s))$,

где $f'(x_0, s) = ((f^{(1)})'(x_0, s), \ldots, (f^{(m)})'(x_0, s))$.

Доказательство. При достаточно малых $\alpha > 0$

$\varphi(x_0 + \alpha s) = F(f(x_0 + \alpha s)) = F(f(x_0) +$
$+ \alpha f'(x_0, s) + o_n(\alpha)) = F(f(x_0) + \alpha f'(x_0, s)) +$
$+ o_1(\alpha) = F(f(x_0)) + \alpha F'(f(x_0), f'(x_0, s)) +$
$+ o_1(\alpha) = \varphi(x_0) + \alpha F'(f(x_0, f'(x_0, s)) + o_1(\alpha)$.

Следовательно, предел $\lim\limits_{\alpha \to +0} \alpha^{-1} [\varphi(x_0 + \alpha s) - \varphi(x_0)]$ существует и равен $F'(f(x_0), f'(x_0, s))$. □

Теорема 1.2.9. Пусть $F(u)$, $u \in R^m$, $f^{(1)}(x_0)$,

$\dots , f^{(m)}(x_0)$ – выпуклые на соответствующих пространствах функции. Предположим, что $\partial F(u) \in R^m_+$ при всех $u \in R^m$. Тогда функция

$$\varphi(x) = F(f^{(1)}(x), \dots, f^{(m)}(x))$$

будет выпуклой функцией, причем для ее субдифференциала будет справедлива формула

$$\varphi(x) = \mathrm{Conv}\,\{ g \mid g = \textstyle\sum_{i=1}^m \lambda^{(i)} g_i,\ g_i \in \partial f^{(i)}(x),$$
$$\lambda = (\lambda^{(1)}, \dots, \lambda^{(m)}) \in \partial F(f^{(1)}(x), \dots, f^{(m)}(x)) \}.$$

Доказательство. Действительно, для любого y из R^m

$$\varphi(y) = F(f^{(1)}(y), \dots, f^{(m)}(y)) \geq$$
$$\geq F(f^{(1)}(x), \dots, f^{(m)}(x)) + \textstyle\sum_{i=1}^m \lambda^{(i)} \times$$
$$\times (f^{(i)}(y) - f^{(i)}(x)) \geq \varphi(x) + \textstyle\sum_{i=1}^m \lambda^{(i)} \langle g_i, y - x\rangle,$$

где $\lambda = (\lambda^{(1)}, \dots, \lambda^{(m)}) \in \partial F(f^{(1)}(x), \dots, f^{(m)}(x))$, $g_i \in \partial f^{(i)}(x)$. Следовательно, в силу теоремы 1.2.6 функция $\varphi(x)$ выпукла. Далее, в силу леммы 1.2.5 имеем

$$\varphi'(x, s) = F'(f(x) . f'(x, s)) =$$
$$= \max \{ \textstyle\sum_{i=1}^m \lambda^{(i)} (f^{(i)})'(x, s) \mid \lambda \in \partial F(f(x) \} =$$
$$= \max \{ \textstyle\sum_{i=1}^m \lambda^{(i)} \max \{\langle g_i, s\rangle \mid g_i \in \partial f^{(i)}(x)\} \mid$$
$$\lambda \in \partial F(f(x)) \} =$$
$$= \max \{\langle \textstyle\sum_{i=1}^m \lambda^{(i)} g_i, s\rangle \mid g_i \in \partial f^{(i)}(x),\ \lambda \in \partial F(f(x))\} =$$
$$= \max \{ \langle g, s \rangle \mid g \in \mathrm{Conv}\,\{ \textstyle\sum_{i=1}^m \lambda^{(i)} g_i \mid$$
$$g_i \in \partial f^{(i)}(x),\ \lambda \in \partial F(f(x)) \} \}. \quad \square$$

Следующие правила дифференцирования непосредственно вытекают из теоремы 1.2.9:

если функции $f^{(1)}(x), \dots, f^{(m)}(x)$ выпуклы, то для субдифференциалов функций

$$\varphi_1(x) = \textstyle\sum_{i=1}^m \alpha_i f^{(i)}(x), \quad \alpha_i \geq 0,$$
$$\varphi_2(x) = \max \{ f^{(i)}(x) \mid i = 1, \dots, m \}$$

справедливы формулы

$\partial\varphi_1(x) = \sum_{i=1}^{m} \alpha_i\, \partial f^{(i)}(x)$,

$\partial\varphi_2(x) = \mathrm{Conv}\,\{\, \partial f^{(i)}(x) \mid f^{(i)}(x) = \varphi_2(x) \,\}$.

Для вычисления субградиентов выпуклых функций, при вычислении которых используются "невыпуклые" операции приходится применять более сложный аппарат лексикографического дифференцирования, который будет описан в гл.2.

1.3. Равномерно выпуклые функции

В §1.2 были изучены основные свойства выпуклых функций. Однако класс выпуклых функций довольно широк и, как мы увидим в дальнейшем, не допускает построения достаточно эффективных методов минимизации. В этом и следующем параграфе будут рассмотрены подклассы выпуклых функций, обладающих более благоприятными для построения численных методов свойствами.

Определение 1.3.1. Функцию $f(x)$ будем называть равномерно выпуклой, если существует неотрицательная функция $\delta(t)$, $\delta(0) = 0$, $\delta(t) > 0$ при $t > 0$, такая, что для всех $x, y \in \mathrm{dom}\, f$ и любого $\alpha \in [0, 1]$ выполняется неравенство

$$f(\alpha x + (1-\alpha) y) \le \alpha f(x) + (1-\alpha) f(y) - \alpha(1-\alpha)\,\delta(\| x - y \|). \tag{1.3.1}$$

Функцию $\delta(t)$ назовем модулем выпуклости функции $f(x)$, а функцию

$$\mu(t) = \inf\,\{\, \alpha^{-1}(1-\alpha)^{-1}\,[\,\alpha f(x) + (1-\alpha) f(y) - f(\alpha x + (1-\alpha) y)\,] \mid x, y \in X,\ \| x - y \| = t,\ 0 < \alpha < 1 \,\}$$

– точным модулем выпуклости функции $f(x)$

Ясно, что $\mu(t) \geq \delta(t)$ для любого модуля выпуклости $\delta(t)$. Отметим, что функции $\delta(t)$ и $\mu(t)$ определены для всех $t \in [0, \mathrm{diam}(\mathrm{dom} f))$.

Перечислим простейшие свойства равномерно выпуклых функций.

1. Пусть функция $f(x)$ является равномерно выпуклой с модулем выпуклости $\delta(t)$, а функция $\varphi(x)$ выпукла на $\mathrm{dom} f$. Тогда функция $f(x) + \varphi(x)$ также будет равномерно выпуклой с тем же модулем выпуклости $\delta(t)$. Отсюда, в частности, следует, что равномерно выпуклые функции не обладают лучшими дифференциальными свойствами по сравнению с выпуклыми функциями.

2. Если $\varphi(x)$ - линейная функция, а $\mu(t)$ - точный модуль выпуклости функции $f(x)$, то этот модуль останется точным и для функции $f(x) + \varphi(x)$.

3. Пусть функции $f_k(x)$, $k = 1, \ldots, m$, равномерно выпуклы с модулями выпуклости $\delta_k(t)$ и $\cap_{k=1}^{m} \mathrm{dom} f_k \neq \emptyset$ Тогда функция $\varphi(x) = \sum_{k=1}^{m} \alpha_k f_k(x)$, $\alpha_k > 0$, также будет равномерно выпуклой с модулем выпуклости $\delta(t) = \sum_{k=1}^{m} \alpha_k \delta_k(t)$.

<u>Лемма 1.3.1.</u> Пусть функция $f(x)$ является равномерно выпуклой и $\delta(t)$ - ее модуль выпуклости. Тогда, если в точке $x \in \mathrm{dom} f$ множество $\partial f(x) \neq \emptyset$, то для любого $y \in \mathrm{dom} f$ справедливо неравенство

$$f(y) \geq f(x) + \langle g_x, y - x \rangle + \delta(\| x - y \|), \quad (1.3.2)$$

где g_x - произвольный элемент из $\partial f(x)$. Если же и $\partial f(y) \neq \emptyset$, то

$$\langle g_x - g_y, x - y \rangle \geq 2\delta(\| x - y \|), \quad (1.3.3)$$

где g_y - произвольный вектор из $\partial f(y)$.

Д о к а з а т е л ь с т в о. Действительно, из неравенства (1.3.1) имеем

$$\delta(\| x - y \|) \le \alpha^{-1} [f(y) - f(\alpha x + (1 - \alpha)y)] +$$
$$+ (1 - \alpha)^{-1} [f(x) - f(\alpha x + (1 - \alpha)y)] \le$$
$$\le \alpha^{-1} [f(y) - f(\alpha x + (1 - \alpha)y)] + \langle g_x , x - y \rangle .$$

В точке x субдифференциал $\partial f(x)$ не пуст. Поэтому функция $f(x)$ полунепрерывна снизу в этой точке. Переходя в полученном неравенстве к пределу по $\alpha \to 1 - 0$, приходим к неравенству (1.3.2). Если теперь $g_y \in \partial f(y) \ne \emptyset$, то

$$f(x) \ge f(y) + \langle g_y , x - y \rangle + \delta(\| x - y \|) .$$

Складывая это неравенство с (1.3.2), получаем (1.3.3). □

Докажем одно важное свойство точного модуля выпуклости функции $f(x)$.

Теорема 1.3.1. Пусть $\mu(t)$ - точный модуль выпуклости функции $f(x)$, а $C \ge 1$ и $t \ge 0$ таковы, что функция $\mu(t)$ определена в точке Ct. Тогда

$$\mu(Ct) \ge C^2 \mu(t) .$$

Доказательство. Пусть сначала $1 \le C \le 2$. Выберем произвольное $\varepsilon > 0$. По определению функции $\mu(t)$ найдутся точки $x_1, x_2 \in \mathrm{dom}\, f$ и число $\alpha \in (0, 0.5)$ такие, что

$$\| x_1 - x_2 \| = Ct ,$$

$$\mu(Ct) + \varepsilon \ge \alpha^{-1} (1 - \alpha)^{-1} [\alpha f(x_1) +$$
$$+ (1 - \alpha) f(x_2) - f(x_3)] \ge \mu(Ct) , \qquad (1.3.4)$$

где $x_3 = \alpha x_1 + (1 - \alpha) x_2$. Обозначим через $\beta = C^{-1}$, $\beta \in (0.5, 1)$. Выберем точку $x_4 = \beta x_1 + (1 - \beta) x_2$. Тогда $\| x_2 - x_4 \| = t$, $x_3 = \gamma x_4 + (1 - \gamma) x_2$, где $\gamma = \alpha / \beta$, и из (1.3.4) имеем

$$\mu(Ct) + \varepsilon \ge \alpha^{-1} (1 - \alpha)^{-1} [\alpha f(x_1) +$$
$$+ (1 - \alpha) f(x_2) - \gamma f(x_4) - (1 - \gamma) f(x_2)] +$$
$$+ \alpha^{-1} (1 - \alpha)^{-1} [\gamma f(x_1) + (1 - \gamma) f(x_2) - f(x_3)] \ge$$

$\geq \beta^{-1}(1-\alpha)^{-1}[\beta f(x_1)+(1-\beta)f(x_2)-f(x_4)]+$
$+(\beta-\alpha)(1-\alpha)^{-1}\beta^{-2}\mu(t) \geq (1-\beta)(1-\alpha)^{-1}\mu(Ct)+$
$+(\beta-\alpha)(1-\alpha)^{-1}C^2\mu(t)$
или
$$(\beta-\alpha)(1-\alpha)^{-1}\mu(Ct) \geq (\beta-\alpha)(1-\alpha)^{-1}C^2\mu(t)-\varepsilon.$$
Заметим, что $0 < (1-\alpha)(\beta-\alpha)^{-1} < (\beta-0.5)^{-1}$. Поэтому в силу произвольности ε получаем $\mu(Ct) \geq C^2\mu(t)$ для $C \in (1, 2)$.

Пусть теперь $C > 2$. Тогда найдется число $m \geq 1$ такое, что $C = a^m$ для некоторого $a \in (1, 2)$. Теперь, пользуясь уже доказанным, получаем
$$\mu(Ct) = \mu(a^m t) \geq a^2\mu(a^{m-1}t) \geq \ldots \geq$$
$$\geq a^{2m}\mu(t) = C^2\mu(t). \quad \square$$

Теорема 1.3.2. Пусть $f(x)$ - равномерно выпуклая функция. Тогда

1) функция $f(x)$ ограничена снизу;

2) при любом t множества $\mathcal{M}(f, t)$ ограничены;

3) любая минимизирующая последовательность точек $\{x_k\}$ таких, что
$x_k \in \operatorname{dom} f$, $f(x_k) \to f^* = \inf\{f(x) \mid x \in \operatorname{dom} f\}$
сходится к единственной точке x^*;

4) для всех x из $\operatorname{dom} f$ справедливо неравенство
$$\mu(\|x-x^*\|) \leq f(x)-f^*. \qquad (1.3.5)$$

Доказательство. Докажем сначала ограниченность множеств $\mathcal{M}(f, t)$ при любом $t \in R$. Нетривиальным является лишь случай $t > f^*$. Пусть точки x', y' принадлежат относительной внутренности множества $\mathcal{M}(f, t)$, $x' \neq y'$. Покажем, что найдется число $\alpha > 0$ такое, что точка $y(\alpha) = y' + \alpha(y'-x') \notin \mathcal{M}(f, t)$.

Действительно, пусть $f(y(\alpha)) \leq t$ при всех $\alpha > 0$.

Тогда, пользуясь равномерной выпуклостью функции $f(x)$ и теоремой 1.3.1, получаем

$$f(y') = f(\alpha(1+\alpha)^{-1} x' + (1+\alpha)^{-1} y(\alpha)) <$$
$$< \alpha(1+\alpha)^{-1} f(x') + (1+\alpha)^{-1} f(y(\alpha)) -$$
$$- \alpha(1+\alpha)^{-2} \mu((1+\alpha) \| x' - y' \|) <$$
$$< \alpha(1+\alpha)^{-1} f(x') + (1+\alpha)^{-1} f(y(\alpha)) - \alpha \mu(\| x' - y' \|) <$$
$$< t - \alpha \mu(\| x' - y' \|),$$

что невозможно при достаточно больших α. Таким образом, множество $\mathcal{M}(f, t)$ ограничено по любому лучу, выходящему из точки x', и, следовательно, в силу следствия 1.4.1 ограничено. Утверждение 1) теоремы теперь непосредственно вытекает из следствия 1.2.4.

Пусть теперь $\{x_k\}$ - произвольная минимизирующая последовательность. Тогда для любых номеров k, l имеем

$$0.5 f(x_k) + 0.5 f(x_m) > f(0.5 x_k + 0.5 x_m) +$$
$$+ 0.25 \mu(\| x_k - x_m \|) > f^* + 0.25 \mu(\| x_k - x_m \|).$$

Таким образом, последовательность $\{x_k\}$ является фундаментальной и сходится к некоторой точке x^*.

Зафиксируем теперь произвольное $\alpha \in (0, 1)$. Тогда

$$\alpha f(x) + (1-\alpha) f(x_k) > f^* + \alpha(1-\alpha)\mu(\| x - x_k \|).$$

Переходя в этом неравенстве к пределу по $k \to \infty$ и сокращая результат на α, имеем

$$f(x) > f^* + (1-\alpha) \mu(\| x - x^* \|),$$

откуда в силу произвольности α получаем неравенство (1.3.5). □

Сформулируем достаточные условия равномерной выпуклости.

<u>Лемма 1.3.2.</u> Пусть $f(x)$ - выпуклая функция, определенная на отрезке $[a, b] \subset R$, и пусть для любых точек x и y из $[a, b]$, таких, что $x < y$ и существуют произ-

водные $f'(x)$, $f'(y)$, выполнено неравенство
$f'(y) - f'(x) \geq \varkappa(y - x)$,
где $\varkappa(t)$ - монотонная на $[0, b - a]$ функция. Тогда при любом $\alpha \in [0, 1]$ справедливо неравенство

$$\alpha f(a) + (1 - \alpha) f(b) - f(\alpha a + (1 - \alpha) b) \geq$$

$$\geq \alpha(1 - \alpha) \int_0^{b-a} \varkappa(t)\, dt. \qquad (1.3.6)$$

Доказательство. Зафиксируем $\alpha \in (0, 1)$ и введем точку $c = \alpha a + (1 - \alpha) b$. Выберем произвольное ξ, $0 < \xi < 1$. Из выпуклости функции $f(x)$ следует, что она удовлетворяет условию Липшица на отрезке $[c + \xi(a - c), c + \xi(b - c)]$, содержащемся в интервале (a, b) (см. следствие 1.2.2), а значит, $f(x)$ абсолютно непрерывна на этом отрезке. Поэтому справедливы равенства

$$f(c + \xi(a - c)) - f(c) = \int_0^{c+\xi(a-c)} f'(x)\, dx, \qquad (1.3.7)$$

$$f(c + \xi(b - c)) - f(c) = \int_0^{c+\xi(b-c)} f'(x)\, dx. \qquad (1.3.8)$$

Введем параметр t, $0 \leq t \leq \xi(b - a)$, и положим $x(t) = c - (1 - \alpha)t$, $y(t) = c + \alpha t$. Умножая (1.3.7),(1.3.8) соответственно на α и $(1 - \alpha)$ и складывая, получаем

$$\alpha f(c + \xi(a - c)) + (1 - \alpha) f(c + \xi(b - c)) - f(c) =$$

$$= \alpha(1 - \alpha) \int_0^{\xi(b-a)} [\, f'(y(t)) - f'(x(t))\,]\, dt \geq$$

$$\geq \alpha(1 - \alpha) \int_0^{\xi(b-a)} \xi(t)\, dt. \qquad (1.3.9)$$

Функция $f(x)$ ограничена на отрезке $[a, b]$, а значит, в силу своей выпуклости и полунепрерывна сверху в точках a и b. Поэтому, переходя в (1.3.9) к пределу при $\xi \to 1-0$, получаем (1.3.6). □

Теорема 1.3.3. Пусть у выпуклой функции $f(x)$ субдифференциал $\partial f(x) \neq \emptyset$ в любой точке $x \in \mathrm{dom}\, f$. Тогда, если существует строго положительная при $t > 0$ функция $\xi(t)$, такая, что при всех $x, y \in \mathrm{dom}\, f$ выполнено неравенство

$$\langle g_x - g_y, x - y \rangle \geq \xi(\| x - y \|), \tag{1.3.10}$$

где g_x, g_y - некоторые векторы из $\partial f(x)$, $\partial f(y)$ соответственно, то функция $f(x)$ будет равномерно выпуклой с модулем выпуклости

$$\delta(t) = \int_0^t \tau^{-1} \xi(\tau)\, d\tau .$$

Доказательство. Выберем произвольные x, $y \in \mathrm{dom}\, f$, $x \neq y$, и $\alpha \in (0, 1)$. Положим $r = \| x - y \|$. Введем параметр t, $0 < t < r$, и положим

$$s(t) = x + t r^{-1} (y - x) .$$

Введем функцию $\varphi(t) = f(s(t))$. Заметим, что $\varphi(t)$ - выпуклая на $[0, r]$ функция. Причем, если g_t - вектор из субдифференциала функции $f(x)$ в точке $s(t)$, то во всех точках t, в которых у функции $\varphi(t)$ существует производная, выполняется равенство

$$\varphi'(t) = r^{-1} \langle g_t, y - x \rangle . \tag{1.3.11}$$

Пусть функция $\varphi(t)$ дифференцируема в точках t_1 и t_2, $0 < t_1 < t_2 < r$, и l_1, l_2 - векторы из субдифференциалов функции $f(x)$ в точках $s(t_1)$, $s(t_2)$ соответственно. Учитывая (1.3.11), получаем

$$\varphi'(t_1) - \varphi'(t_2) =$$
$$= \langle l_1 - l_2 ,\ s(t_1) - s(t_2) \rangle / \| s(t_1) - s(t_2) \| \ge$$
$$\ge \xi(\| s(t_1) - s(t_2) \|) / \| s(t_1) - s(t_2) \| =$$
$$= (t_2 - t_1)^{-1} \xi(t_2 - t_1) .$$

Осталось воспользоваться леммой 1.3.2:

$$\alpha f(x) + (1 - \alpha) f(y) - f(\alpha x + (1 - \alpha) y) =$$
$$= \alpha \varphi(r) + (1 - \alpha) \varphi(0) - \varphi(\alpha r) \ge$$
$$\ge \alpha(1 - \alpha) \int_0^r t^{-1} \xi(t)\, dt . \quad \square$$

Теорема 1.3.3 во многих случаях позволяет упростить процесс проверки равномерной выпуклости конкретных функций. Продемонстрируем это на примере функций вида $f(x) = \varphi(\| x \|)$.

Теорема 1.3.4. Пусть функция $\xi(t)$ определена при $t \ge 0$, $\xi(0) = 0$, $\xi(Ct) \ge C \xi(t)$ для всех $C \ge 1$, $t \ge 0$. Тогда функция $f(x) = \int_0^{\|x\|} \xi(t)\, dt$ является равномерно выпуклой на всем R^n с модулем выпуклости $\delta(t) = \int_0^t \xi(0.5\, \tau)\, d\tau$. Если, кроме того, функция $t^{-1} \xi(t)$ выпукла на R_+, то можно взять $\delta(t) = 2 \int_0^t \xi(0.5\, \tau)\, d\tau$, а если вогнута на R_+, то $\delta(t) = 0.5 \int_0^t \xi(\tau)\, d\tau$.

Д о к а з а т е л ь с т в о. Пусть $x \in R^n$, $x \ne 0$. Тогда из неубывания и неотрицательности функции $\xi(t)$ следует, что

$$f(y) \geq \int_0^{\|x\|} \xi(\tau)\,d\tau + \xi(\|x\|)(\|y\| - \|x\|) \geq$$

$$\geq f(x) + \langle \xi(\|x\|)\,\|x\|^{-1} x,\ y - x \rangle$$

для всех y из R^n. Таким образом, вектор

$$g(x) = \xi(\|x\|)\,\|x\|^{-1} x \in \partial f(x).$$

Нетрудно также видеть, что $g(0) = 0 \in \partial f(0)$.

Проверим теперь выполнение условий теоремы 1.3.3. Поскольку функция $t^{-1}\xi(t)$ монотонно возрастает, то при всех $x, y \neq 0$ имеем

$$\langle g(x) - g(y),\ x - y \rangle =$$
$$= 0.5\,[(\xi(\|x\|)\,\|x\|^{-1} - \xi(\|y\|)\,\|y\|^{-1})\times$$
$$\times(\|x\|^2 - \|y\|^2) +$$
$$+ \|x - y\|^2 (\xi(\|x\|)\,\|x\|^{-1} + \xi(\|y\|)\,\|y\|^{-1}]; \tag{1.3.12}$$

$$\langle g(x) - g(y),\ x - y \rangle \geq$$
$$\geq 0.5\,\|x - y\|^2\, 2\,\xi(0.5\,\|x\| + 0.5\,\|y\|)/$$
$$/(\|x\| + \|y\|) \geq$$
$$\geq \|x - y\|\,\xi(0.5\,\|x - y\|). \tag{1.3.13}$$

Если функция $t^{-1}\xi(t)$ выпукла при $t > 0$, то, воспользовавшись неравенством

$$\xi(\|x\|\ \ \|x\|^{-1} + \xi(\|y\|)\,\|y\|^{-1}) \geq$$
$$\geq 4\,[\xi(0.5\,\|x\| + 0.5\,\|y\|)/(\|x\| + \|y\|)],$$

с помощью (1.3.12) получим более точное, чем (1.3.13), неравенство

$$\langle g(x) - g(y),\ x - y \rangle \geq$$
$$= 2\,\|x - y\|^2\,\xi(0.5\,\|x\| + 0.5\,\|y\|)/$$
$$/(\|x\| + \|y\|) \geq$$
$$\geq 2\,\|x - y\|\,\xi(0.5\,\|x - y\|). \tag{1.3.14}$$

Наконец, если функция $\varphi(t) = t^{-1}\xi(t)$ вогнута при

$t > 0$, то

$$\|x\| \, (\|x\| + \|y\|)^{-1} \, \varphi(\|x\| + \|y\|) \le \varphi(\|x\|),$$

$$\|y\| \, (\|x\| + \|y\|)^{-1} \, \varphi(\|x\| + \|y\|) \le \varphi(\|y\|).$$

Складывая эти неравенства, получаем

$$\xi(\|x\| \ \|x\|^{-1} + \xi(\|y\|) \, \|y\|^{-1}) \ge$$

$$\ge \xi(\|x\| + \|y\|) / (\|x\| + \|y\|).$$

Отсюда и из (1.3.14) следует, что

$$\langle g(x) - g(y), x - y \rangle \ge$$

$$= 0.5 \|x - y\| \, \xi(\|x\| + \|y\|)(\|x\| + \|y\|)^{-1} \ge$$

$$\ge 0.5 \, \|x - y\| \, \xi(\|x - y\|). \qquad (1.3.15)$$

Нетрудно видеть, что неравенства (1.3.13),(1.3.14), (1.3.15) выполняются и в случае, когда одна из точек x или y равна нулю. Для завершения доказательства осталось воспользоваться теоремой 1.3.3. □

<u>Пример 1.3.1</u>. Пусть $f(x) = \|x\|^p$, $p \ge 2$. Тогда $\xi(t) = p\,t^{p-1}$. Поэтому $\delta(t) = t^2$ при $p = 2$, $\delta(t) = 0.5\,t^p$ при $2 < p < 3$ и $\delta(t) = 4\,(0.5\,t)^p$ при $p \ge 3$. Более детальные исследования показывают, что точный модуль выпуклости этой функции есть $\mu(t) = 4\,(0.5\,t)^p$ при всех $p \ge 2$. □

Наиболее важными представителями равномерно выпуклых функций являются сильно выпуклые функции.

<u>Определение 1.3.2.</u> Функцию $f(x)$ будем называть сильно выпуклой, если существует константа $m = m(f) > 0$ такая, что для всех $x, y \in \mathrm{dom}\, f$ и любого $\alpha \in [0, 1]$ выполняется неравенство

$$f(\alpha x + (1 - \alpha) y) \le \alpha f(x) + (1 - \alpha) f(y) -$$

$$- 0.5\,\alpha\,(1 - \alpha)\,m\,\|x - y\|^2.$$

Константа $m(f)$ называется константой сильной выпуклости функции $f(x)$.

Другими словами, сильно выпуклые функции - это равномерно выпуклые функции с модулем выпуклости $\delta(t) = 0.5 \times m(f) t^2$.

Простейшим примером сильно выпуклой функции является квадратичная функция, порожденная положительно определенной симметрической матрицей. Действительно, пусть A - симметрическая матрица, такая, что $\langle A s , s \rangle \geq m \| s \|^2$ для всех s из R^n , $m > 0$. Рассмотрим функцию

$$f(x) = 0.5 \langle A x , x \rangle - \langle b , x \rangle ,$$

где b - произвольный вектор из R^n . Тогда для любых x , y из R^n и $\alpha \in [0 , 1]$ имеем

$$f(\alpha x + (1 - \alpha) y) = 0.5 \langle A (\alpha x + (1 - \alpha) y) , \alpha x + (1 - \alpha) y \rangle + \langle b , \alpha x + (1 - \alpha) y \rangle =$$
$$= 0.5 \alpha^2 \langle A x , x \rangle + 0.5 (1 - \alpha)^2 \langle A y , y \rangle +$$
$$+ \alpha(1 - \alpha) \langle A x , y \rangle + \alpha \langle b, x \rangle + (1 - \alpha) \langle b, y \rangle =$$
$$= 0.5 \alpha \langle A x , x \rangle + 0.5 (1 - \alpha) \langle A y , y \rangle -$$
$$- 0.5 \alpha(1 - \alpha)\langle A(x - y), x - y\rangle + \alpha\langle b,x\rangle + (1 - \alpha)\langle b,y\rangle =$$
$$= \alpha f(x) + (1 - \alpha)f(y) - 0.5 \alpha(1 - \alpha)\langle A(x - y),x - y\rangle \leq$$
$$\leq \alpha f(x) + (1 - \alpha) f(y) - 0.5 \alpha(1 - \alpha) m \| x - y \|^2 .$$

Таким образом, константа сильной выпуклости квадратичной функции равна минимальному собственному значению порождающей ее матрицы.

Разумеется, все свойства равномерно выпуклых функций распространяются и на класс сильно выпуклых функций. Однако ввиду важности последнего нам будет удобно сформулировать некоторые из них в отдельной теореме.

<u>Теорема 1.3.5.</u> Пусть функция $f(x)$ субдифференцируема в каждой точке своей эффективной области. Для того чтобы функция $f(x)$ была сильно выпуклой с константой m необходимо и достаточно, чтобы при любых x, y из $\mathrm{dom}\, f$ вы-

полнялось любое из неравенств

$$f(y) \geq f(x) + \langle g_x , y - x \rangle + 0.5\, m \, \| x - y \|^2 , \qquad (1.3.16)$$

$$\langle g_x - g_y , x - y \rangle \geq m \, \| x - y \|^2 , \qquad (1.3.17)$$

где g_x , g_y – элементы субдифференциалов функции $f(x)$ в точках x и y соответственно.

Доказательство. Действительно, пусть функция $f(x)$ сильно выпукла с константой m . Тогда формулы (1.3.16),(1.3.17) получаются из формул (1.3.2), (1.3.3) подстановкой $\delta(t) = 0.5\, m\, t^2$.

Обратно, пусть при всех $x , y \in \mathrm{dom}\, f$ справедливо неравенство (1.3.16). Тогда

$$f(x) \geq f(y) + \langle g_y , x - y \rangle + 0.5\, m \, \| x - y \|^2 .$$

Складывая это неравенство с (1.3.16), получаем (1.3.17).

И, наконец, если неравенство (1.3.17) выполняется при всех $x, y \in \mathrm{dom}\, f$, то в силу теоремы 1.3.3 функция $f(x)$ будет равномерно выпуклой с модулем

$$\delta(t) = \int_0^t m\, \tau^2 / \tau \, d\tau = 0.5\, m\, t^2 . \qquad \square$$

Следствие 1.3.1. Если точка y из $\mathrm{dom}\, f$ такова, что $f(y) \leq f(x_0)$, то справедливо неравенство

$$\| y - x^* \| \leq [\, 2\, m^{-1} (f(x_0) - f^*) \,]^{1/2} ,$$

где x^* – точка минимума функции $f(x)$, $f^* = f(x^*)$.

Действительно, $0 \in \partial f(x^*)$. Поэтому, в силу неравенства (1.3.16)

$$0.5\, m \, \| y - x^* \|^2 \leq f(y) - f(x^*) \leq f(x_0) - f^* .$$

$\square$

1.4. Гладкие выпуклые функции

В этом параграфе мы получим основные неравенства, описывающие поведение гладких выпуклых функций.

Определение 1.4.1. Будем говорить, что дифференцируемая на R^n функция $f(x)$ принадлежит классу $C^{1,\nu}(L)$, $0 \le \nu \le 1$, если для любых точек x, y из R^n справедливо неравенство

$$\| f'(x) - f'(y) \| \le L \| x - y \|^{\nu}.$$

Таким образом, класс $C^{1,1}(L)$ состоит из функций, градиент которых удовлетворяет условию Липшица с константой L, классы $C^{1,\nu}(L)$, $0 < \nu < 1$, - из функций, градиент которых удовлетворяет условию Гельдера с параметром ν. Под классом $C^{1,0}(L)$ мы будем понимать класс выпуклых функций с равномерно ограниченными на R^n субградиентами. Несложно показать, что классы $C^{1,\nu}(L)$ с $\nu > 1$ рассматривать бессмысленно. так как они содержат только линейные функции, которые входят во все классы $C^{1,\nu}(L)$ с $\nu \ge 0$, $L \ge 0$.

Непосредственным следствием определения является следующее утверждение: если функция $f_1(x) \in C^{1,\nu}(L_1)$, $f_2(x) \in C^{1,\nu}(L_2)$, то при любых α, $\beta \ge 0$ функция $f_3(x) = \alpha f_1(x) + \beta f_2(x)$ принадлежит классу $C^{1,\nu}(L_3)$, где $L_3 = \alpha L_1 + \beta L_2$.

Теорема 1.4.1. Пусть функция $f(x) \in C^{1,\nu}(L)$, $0 \le \nu \le 1$. Тогда для любых x, $y \in R^n$ справедливы неравенства

$$| \langle f'(x) - f'(y), x - y \rangle | \le L \| x - y \|^{1+\nu},$$

$$| f(y) - f(x) - \langle f'(x), y - x\rangle | \le L(1+\nu)^{-1} \| x - y \|^{1+\nu}$$

Доказательство. Действительно,

$$| \langle f'(x) - f'(y), x - y \rangle | \le$$
$$\le \| f'(x) - f'(y) \| \times \| x - y \| \le L \| x - y \|^{1+\nu};$$

$$f(y) = f(x) + \int_0^1 \langle f'(x + t(y - x)), y - x \rangle dt =$$

$$= f(y) = f(x) + \langle f'(x), y - x \rangle dt +$$

$$+ \int_0^1 \langle f'(x + t(y - x)) - f'(x), y - x \rangle dt.$$

Поэтому

$$| f(y) - f(x) - \langle f'(x), x - y \rangle | \le$$

$$\le \left| \int_0^1 \langle f'(x + t(y - x)) - f'(x), y - x \rangle dt \right| \le$$

$$\le L \| y - x \|^{1+\nu} \int_0^1 t^{\nu} dt = L(1+\nu)^{-1} \| y - x \|^{1+\nu}$$

$\square$

<u>Следствие 1.4.1.</u> Если $f(x) \in C^{1,\nu}(L)$, $0 < \nu \le 1$, то найдется точка $y \in R^n$ такая, что

$$f(y) \le f(x) - \nu(1+\nu)^{-1} L^{-1/\nu} \| f'(x) \|^{1+1/\nu}. \qquad (1.4.1)$$

В частности, можно положить

$$y = x - L^{-1/\nu} \| f'(x) \|^{1/\nu - 1} f'(x).$$

В самом деле, при таком выборе точки y, в силу теоремы 1.4.1 имеем

$$f(y) \le f(x) - \langle f'(x), y - x \rangle + L(1+\nu)^{-1} \| x - y \|^{1+\nu} =$$
$$= f(x) - L^{-1} \| f'(x) \|^{1+1/\nu} +$$
$$+ L(1 + \nu)^{-1} [L^{-1/\nu} \| f'(x) \|^{1/\nu}]^{1+\nu} =$$
$$= f(x) - \nu(1 + \nu)^{-1} L^{-1/\nu} \| f'(x) \|^{1+1/\nu}. \quad \square$$

В случае, когда у функции $f(x)$ существует точка минимума x^*, из неравенства (1.4.1) получаем

$$f(x) - f^* \ge \nu(1 + \nu)^{-1} L^{-1/\nu} \| f'(x) \|^{1+1/\nu}. \qquad (1.4.2)$$

Заметим, что в силу теоремы 1.4.1, при любом $x_0 \in R^n$ множество

$$\{ (x, t) \in R^{n+1} \mid t \geq f(x_0) + \langle f'(x_0), x - x_0 \rangle + L(1+\nu)^{-1} \| x - x_0 \|^{1+\nu} \}$$

принадлежит надграфику функции $f(x)$ Для выпуклых функций это означает, что при любых x, y, z из R^n справедливо неравенство

$$f(x) + \langle f'(x), z - x \rangle + L(1+\nu)^{-1} \| z - x \|^{1+\nu} \geq f(y) + \langle f'(y), z - y \rangle \quad (1.4.3)$$

Из неравенства (1.4.3) можно извлечь два полезных следствия.

<u>Теорема 1.4.2.</u> Если выпуклая функция $f(x)$ принадлежит классу $C^{1,\nu}(L)$, $0 < \nu \leq 1$, то при всех x, y из R^n справедливы неравенства

$$f(x) \geq f(y) + \langle f'(y), x - y \rangle + \nu(1+\nu)^{-1} L^{-1/\nu} \| f'(x) - f'(y) \|^{1+1/\nu}; \quad (1.4.4)$$

$$\langle f'(x) - f'(y), x - y \rangle \geq 2\nu(1+\nu)^{-1} L^{-1/\nu} \| f'(x) - f'(y) \|^{1+1/\nu}. \quad (1.4.5)$$

Д о к а з а т е л ь с т в о. Перепишем неравенство (1.4.3) в следующей форме:

$$f(x) \geq f(y) + \langle f'(y), x - y \rangle + \langle f'(y) - f'(x), z - x \rangle + L(1+\nu)^{-1} \| z - x \|^{1+\nu} \quad (1.4.6)$$

Заметим, что в этом неравенстве точка z может быть любой. Выберем

$$z = x + L^{-1/\nu} \| f'(y) - f'(x) \|^{1/\nu - 1} (f'(y) - f'(x))$$

(эта точка максимизирует правую часть неравенства (1.4.6) по $z \in R^n$). Подставляя выбранное значение z в неравенство (1.4.6), получаем (1.4.4).

Запишем теперь неравенство (1.4.4), поменяв местами точки x и y:

$$f(y) \ge f(x) + \langle f'(x), y - x \rangle +$$
$$+ \nu(1 + \nu)^{-1} L^{-1/\nu} \| f'(x) - f'(y) \|^{1+1/\nu}$$

Складывая это неравенство с (1.4.4), получаем (1.4.5). □

Из всех классов $C^{1,\nu}(L)$ с $\nu > 0$ в теории оптимизации наиболее часто используется класс $C^{1,1}(L)$. Для функций из этого класса утверждения теорем 1.4.1, 1.4.2 являются не только необходимыми, но и достаточными.

Теорема 1.4.3. Для вхождения выпуклой дифференцируемой функции $f(x)$ в класс $C^{1,1}(L)$ необходимо и достаточно, чтобы при всех x, y из R^n выполнялось любое из неравенств

$$0 \le \langle f'(x) - f'(y), x - y \rangle \le 0.5 L \|x - y\|^2; \qquad (1.4.7)$$

$$f(x) + \langle f'(x), y - x \rangle \le f(y) \le$$
$$\le f(x) + \langle f'(x), y - x \rangle + 0.5 L \|x - y\|^2; \qquad (1.4.8)$$

$$f(y) \ge f(x) + \langle f'(x), y - x \rangle +$$
$$+ 0.5 L^{-1} \| f'(x) - f'(y) \|^2; \qquad (1.4.9)$$

$$\langle f'(x) - f'(y), x - y \rangle \ge$$
$$\ge L^{-1} \| f'(x) - f'(y) \|^2. \qquad (1.4.10)$$

Доказательство. При доказательстве теорем 1.4.1, 1.4.2 выписанные неравенства обосновывались в такой последовательности:

$$f(x) \in C^{1,1}(L) \Rightarrow (1.4.7) \Rightarrow (1.4.8) \Rightarrow$$
$$\Rightarrow (1.4.9) \Rightarrow (1.4.10).$$

Осталось заметить, что в силу (1.4.10)

$$L^{-1} \| f'(x) - f'(y) \|^2 \le \langle f'(x) - f'(y), x - y \rangle \le$$
$$\le \| f'(x) - f'(y) \| \times \| x - y \|,$$

т.е. $\| f'(x) - f'(y) \| \le L \| x - y \|$. Таким образом, выполнение любого из неравенств (1.4.7)-(1.4.10) эквивалентно принадлежности выпуклой функции классу $C^{1,1}(L)$. □

Укажем также для класса $C^{1,1}(L)$ вид неравенств (1.4.1),(1.4.2):

$$f(y) \le f(x) - 0.5\, L^{-1} \| f'(x) \|^2 \qquad (1.4.11)$$

для $y = x - L^{-1} f'(x)$;

$$f(x) - f^* \ge 0.5\, L^{-1} \| f'(x) \|^2 . \qquad (1.4.12)$$

Любое из неравенств (1.4.7)-(1.4.10) является точным на классе $C^{1,1}(L)$, т.е. существует функция $f(x) \in C^{1,1}(L)$ и точки $x, y \in R^n$, для которых эти неравенства превращаются в равенства. Действительно, возьмем $f(x) = 0.5 \langle A x, x \rangle$, где $A = A^T$ - неотрицательно определенная матрица. Тогда $f(x) \in C^{1,1}(L)$ с $L = \| A \|$. Обозначим через s собственный вектор матрицы A, отвечающий ее максимальному собственному значению L. В этом случае для любого вектора x из R^n справедливо тождество

$$f'(x + s) \equiv A(x + s) \equiv A x + A s \equiv f'(x) + L s .$$

Нетрудно убедиться в том, что при выборе $y = x + s$ любое из неравенств (1.4.7)÷(1.4.10) обратится в равенство. Возьмем, например, неравенство (1.4.9)

$$f(y) = f(x) + \langle f'(x), y - x \rangle + 0.5 \langle A(y - x), y - x \rangle =$$
$$= f(x) + \langle f'(x), y - x \rangle + 0.5 \langle A s, s \rangle =$$
$$= f(x) + \langle f'(x), y - x \rangle + 0.5\, L \| s \|^2 =$$
$$= f(x) + \langle f'(x), y - x \rangle + 0.5\, L^{-1} \| f'(x) - f'(y) \|^2 .$$

□

Для дважды дифференцируемой функции $f(x)$ вхождение в класс $C^{1,1}(L)$ можно вывести из следующей теоремы.

<u>Теорема 1.4.4.</u> Если дважды дифференцируемая на R^n функция $f(x)$ такова, что

$$\sup \{ \| f''(x) \| \mid x \in R^n \} \equiv L < \infty ,$$

то $f(x) \in C^{1,1}(L)$.

Д о к а з а т е л ь с т в о. Действительно, в силу теоремы о среднем для любых $x, y \in R^n$ найдется величина $\theta \in [0, 1]$ такая, что

$$\| f'(x) - f'(y) \| = \| f''(y + \theta(x - y))(x - y) \| .$$

Поэтому

$$\| f'(x) - f'(y) \| \le$$
$$\le \| f''(y + \theta(x - y)) \| \times \| x - y \| \le L \| x - y \| .$$

□

В заключение этого параграфа выведем необходимые и достаточные условия для вхождения функций в класс $\mathcal{F}(L, m)$ - класс сильно выпуклых с константой $m > 0$ на R^n функций, содержащихся также и в классе $C^{1,1}(L)$.

<u>Теорема 1.4.5.</u> Для того чтобы дифференцируемая на R^n функция $f(x)$ принадлежала классу $\mathcal{F}(L, m)$ необходимо и достаточно, чтобы при всех x, y из R^n выполнялось любое из неравенств

$$m \| x - y \|^2 \le \langle f'(x) - f'(y), x - y \rangle \le$$
$$\le L \| x - y \|^2 ; \qquad (1.4.13)$$

$$f(x) + \langle f'(x), y - x \rangle + 0.5\, m \| x - y \|^2 \le$$
$$\le f(y) \le f(x) + \langle f'(x), y - x \rangle + 0.5\, L \| x - y \|^2 ;$$
$$(1.4.14)$$

$$f(y) \ge f(x) + \langle f'(x), y - x \rangle + 0.5\, m \| x - y \|^2 +$$
$$+ 0.5\, (L - m)^{-1} \| f'(x) - f'(y) - m(x - y) \|^2 ; \qquad (1.4.15)$$

$$\langle f'(x) - f'(y), x - y \rangle \ge$$
$$\ge m L (m + L)^{-1} \| x - y \|^2 + (L + m)^{-1} \| f'(x) - f'(y) \|^2 .$$
$$(1.4.16)$$

Д о к а з а т е л ь с т в о. Необходимость и достаточность неравенств (1.4.13),(1.4.14) для включения в $\mathcal{F}(L, m)$ следует из теоремы 1.3.5 и теоремы 1.4.3. Докажем такое же утверждение и для неравенств (1.4.15),(1.4.16). Предположим сначала, что $L > m$.

Итак, пусть $f(x) \in \mathcal{F}(L, m)$. Это означает, что для любого $y \in R^n$ параболоид

$$\{ (z, t) \in R^{n+1} \mid t \geq f(y) + \langle f'(y), z - y \rangle + 0.5 L \| z - y \|^2 \}$$

лежит в надграфике функции $f(x)$, который, в свою очередь, содержится в параболоиде

$$\{ (z, t) \in R^{n+1} \mid t \geq f(x) + \langle f'(x), z - x \rangle + 0.5 m \| z - x \|^2 \}$$

при всяком x из R^n. Другими словами, для любых x, y, z из R^n справедливо неравенство

$$f(y) + \langle f'(y), z - y \rangle + 0.5 L \| z - y \|^2 \geq$$
$$\geq f(x) + \langle f'(x), z - x \rangle + 0.5 m \| z - x \|^2 . \quad (1.4.17)$$

После тождественных преобразований неравенство (1.4.17) может быть записано в следующем виде:

$$f(y) \geq f(x) + \langle f'(x), y - x \rangle + 0.5 m \| y - x \|^2 +$$
$$+ \langle f'(x) - f'(y) - m(x - y), z - y \rangle - 0.5 (L - m) \| z - y \|^2 .$$

Максимизируя правую часть последнего неравенства по $z \in R^n$ (этот максимум достигается при $z = y + m^{-1} (f'(x) - f'(y)) - x + y$), получаем неравенство (1.4.15).

Далее, пусть неравенство (1.4.15) выполняется при всех x, y из R^n. Поменяв в нем точки x и y местами, получаем

$$f(x) \geq f(y) + \langle f'(y), x - y \rangle + 0.5 m \| x - y \|^2 +$$
$$+ 0.5 (L - m)^{-1} \| f'(x) - f'(y) - m (x - y) \|^2 .$$

Складывая это неравенство с (1.4.15), имеем

$$\langle f'(x) - f'(y), x - y \rangle \geq m \| x - y \|^2 +$$
$$+ (L - m)^{-1} \| f'(x) - f'(y) - m (x - y) \|^2 =$$
$$= m L (L - m)^{-1} \| x - y \|^2 + (L - m)^{-1} \| f'(x) - f'(y) \|^2 -$$
$$- 2 m (L - m)^{-1} \langle f'(x) - f'(y), x - y \rangle ,$$

откуда сразу получаем неравенство (1.4.14).

И, наконец, пусть неравенство (1.4.16) выполняется при всех $x, y \in R^n$ Но тогда при всех x, y из R^n

$$(L + m) \| f'(x) - f'(y) \| \times \| x - y \| \geq$$
$$\geq mL \| x - y \|^2 + \| f'(x) - f'(y) \|^2 ,$$

т.е.

$$(\| f'(x) - f'(y) \| - L \| x - y \|) \times$$
$$\times (\| f'(x) - f'(y) \| - m \| x - y \|) \leq 0 .$$

Это возможно лишь в том случае, если при любых $x, y \in R^n$

$$m \| x - y \| \leq \| f'(x) - f'(y) \| \leq L \| x - y \| . \quad (1.4.18)$$

Из неравенства (1.4.18) заключаем, что $f(x) \in C^{1,1}(L)$. Подставляя левую часть неравенства (1.4.18) в (1.4.16), получаем

$$\langle f'(x) - f'(y), x - y \rangle \geq$$
$$\geq mL (L + m)^{-1} \| x - y \|^2 + m^2 (L + m)^{-1} \| x - y \|^2 =$$
$$= m \| x - y \|^2 ,$$

т.е. в силу теоремы 1.3.5 функция $f(x)$ является сильно выпуклой на R^n с константой сильной выпуклости m. Таким образом, $f(x) \in \mathfrak{F}(L, m)$.

В заключение заметим, что если $L = m$, то в силу (1.4.14)

$$f(y) \equiv f(x) + \langle f'(x), y - x \rangle + 0.5\, m \| x - y \|^2$$

для любых x, y из R^n, и все неравенства (1.4.13)-(1.4.16) будут выполнены как равенства (в правой части неравенства (1.4.15) будет отсутствовать последний член, так как $f'(x) - f'(y) - m(x - y) \equiv 0$). □

1.5. Квазивыпуклые функции

Как мы увидим в дальнейшем, для обоснования многих методов минимизации требование выпуклости целевой функции является излишним. Так, например, для получения оценок скорости сходимости метода субградиентного спуска, методов отсечений нужна лишь выпуклость множеств $\mathcal{M}(f, t)$. Функции, обладающие таким свойством, называются квазивыпуклыми. Дадим точное определение.

<u>Определение 1.5.1.</u> Функция $f(x)$, определенная на выпуклом множестве X, называется квазивыпуклой на этом множестве, если для любых двух точек x, y из X и произвольного α, $0 < \alpha < 1$, выполняется неравенство

$$f(\alpha x + (1 - \alpha) y) < \max\{f(x), f(y)\}.$$

Нетрудно убедиться в том, что всякая выпуклая функция будет также и квазивыпуклой. Если $f(x)$ - квазивыпуклая на X функция, то и функция $\lambda f(x)$, $\lambda > 0$, квазивыпукла. Если $f_1(x)$ и $f_2(x)$ - квазивыпуклые функции, то и функция $f_3(x) = \max\{f_1(x), f_2(x)\}$ будет квазивыпуклой. Однако, в отличие от выпуклых функций, сумма квазивыпуклых функций, вообще говоря, квазивыпуклой функцией не является. Для некоторых функций удобным средством проверки их квазивыпуклости является следующая лемма.

<u>Лемма 1.5.1.</u> Если $f_1(x)$ - выпуклая на X, а $f_2(x)$ - вогнутая на X функции, причем $f_1(x) > 0$, а $f_2(x) > 0$ для всех x из X, то функция $\varphi(x) = f_1(x) / f_2(x)$ будет квазивыпуклой на X.

Доказательство. Для любых x, y из X и $\alpha \in [0, 1]$ имеем

$\varphi(\alpha x + (1 - \alpha) y) =$

$= f_1(\alpha x + (1-\alpha) y) / f_2(\alpha x + (1-\alpha) y) \le$
$\le [\alpha f_1(x) + (1-\alpha) f_1(y)] / [\alpha f_2(x) + (1-\alpha) f_2(y)] =$
$= [\alpha f_2(x) \varphi(x) + (1-\alpha) f_2(y) \varphi(y)] /$
$/ [\alpha f_2(x) + (1-\alpha) f_2(y)] \le$
$\le \max\{\varphi(x), \varphi(y)\}$. □

З а м е ч а н и е. Если $f_2(x)$ - линейная функция, то в условии леммы требование $f_1(x) > 0$ можно опустить.

С помощью леммы 1.5.1 нетрудно установить, например, квазивыпуклость функции $q(A) = \| A \| / m(A)$ - числа обусловленности матрицы A как функции от ее элементов на множестве симметрических положительно определенных матриц Важным примером квазивыпуклой функции является функция

$$f(x) = \max\left\{ \left| \frac{\langle a_i, x \rangle}{\langle b_i, x \rangle} - c_i \right| , i = 1, \dots, N \right\},$$

$$x \in \{ x \mid \langle b_i, x \rangle > 0, i = 1, \dots, N \}.$$

Эта функция связана с задачей нахождения наилучшей дробно-линейной аппроксимации.

Докажем характеристическое свойство квазивыпуклых функций.

Теорема 1.5.1. Функция $f(x)$ является квазивыпуклой на выпуклом множестве X тогда и только тогда, когда при любых t из R множество $\mathcal{M}(f, t) \cap X$ выпукло.

Д о к а з а т е л ь с т в о. Пусть функция $f(x)$ является квазивыпуклой на X. Тогда для любых $x, y \in \mathcal{M}(f, t) \cap X$ и $\alpha \in [0, 1]$ имеем

$f(\alpha x + (1-\alpha) y) \le \max\{f(x), f(y)\} \le t$,

следовательно, точка $\alpha x + (1-\alpha) y \in \mathcal{M}(f, t) \cap X$.

Пусть теперь множество $\mathcal{M}(f, t) \cap X$ является выпуклым при всех $t \in R$. Выберем произвольные точки x, y из X и положим $t = \max\{f(x), f(y)\}$. Тогда при любых α из $[0, 1]$ точка $\alpha x + (1 - \alpha) y$ принадлежит множеству $\mathcal{M}(f, t) \cap X$ Другими словами,

$f(\alpha x + (1 - \alpha) y) \le t = \max\{f(x), f(y)\}$ □

Для квазивыпуклых функций в отличие от функций выпуклых. ограниченность и непустота множества $\mathcal{M}(f, t)$ при некотором t из R не влечет за собой ограниченность такого множества при любом t из R Тем не менее. справедлив следующий результат.

Теорема 1.5.2. Пусть непрерывная функция $f(x)$ является квазивыпуклой на замкнутом выпуклом множестве X . Тогда, если при некотором t_0 из R множество $\mathcal{M}(f, t_0) \cap X$ непусто и ограничено, то задача минимизации функции $f(x)$ на X корректна: для любой последовательности точек $\{x_k\} \subset X$, такой, что

$f(x_k) \to f^* = \min\{f(x) \mid x \in X\}$,

множество $X^* = \mathrm{Argmin}\{f(x) \mid x \in X\} \ne \emptyset$ и расстояние $\rho(x_k, X^*) \to 0$ при $k \to \infty$.

Доказательство. Из ограниченности множества $\mathcal{M}(f, t_0) \cap X$, непрерывности функции $f(x)$ и замкнутости множества X сразу получаем, что $X^* \ne \emptyset$ Заметим также, что $X^* \subseteq \mathcal{M}(f, t_0) \cap X$, поэтому $d = \mathrm{diam}\, X^* < \infty$

Пусть теперь последовательность $\{x_k\}$ такова , что $f(x_k) \to f^*$ при $k \to \infty$. Если эта последовательность ограничена, то в силу непрерывности $f(x)$ любая ее предельная точка принадлежит множеству X^* . Предположим, что последовательность $\{x_k\}$ не ограничена (в условиях теоремы это может произойти лишь в случае $t_0 = f^*$). Зафиксируем произвольную точку $x^* \in X^*$. Не ограничивая общности будем считать, что $\| x_k - x^* \| \to \infty$, $\| x_k - x^* \| > 3d$ при всех

$k > 0$. Рассмотрим последовательность точек

$y_k = x^* + 2d(x_k - x^*) / \| x_k - x^* \|$

Заметим, что $y_k \in [x^*, x_k]$ Поэтому в силу квазивыпуклости функции $f(x)$ имеем

$f(y_k) \le \max\{ f(x^*), f(x_k) \} = f(x_k)$.

Отсюда получаем, что $f(y_k) \to f^*$ при $k \to \infty$ При этом $\| y_k - x^* \| = 2d$. Следовательно, существует точка y^* такая, что $\| y^* - x^* \| = 2d$ и $f(y^*) = f^*$, что противоречит способу выбора константы d. □

Квазивыпуклые функции в отличие от выпуклых не обладают, вообще говоря, свойствами непрерывности, субдифференцируемости и дифференцируемости по направлениям. В качестве их дифференциальных характеристик обычно используют элементы конуса

$\mathcal{D}f(x) = \{ g \mid \langle g, x - y \rangle \ge 0 \ \forall y \in X:$

$f(y) \le f(x) \}$.

Заметим, что это множество всегда непусто ($0 \in \mathcal{D}f(x)$), замкнуто и является выпуклым конусом. Элементарными следствиями определения множества $\mathcal{D}f(x)$ являются следующие свойства:

1) если $\mathcal{D}f(x) \ne \{0\}$ и $\langle g, y - x \rangle > 0$ для какого-либо вектора g из $\mathcal{D}f(x)$ некоторой точки y из X, то $f(y) > f(x)$;

2) если $y \in X$ и $f(y) < f(x)$, то $\langle g, x - y \rangle > 0$ для любого вектора $g \in \mathcal{D}f(x)$, $g \ne 0$

Рассмотрим квазивыпуклую функцию $f(x)$, заданную на открытом выпуклом множестве $X \subseteq R^n$.

<u>Теорема 1.5.3.</u> 1. Если функция $f(x)$ дифференцируема в точке $x \in X$ и $f'(x) \ne 0$, то

$\mathcal{D} f(x) \equiv \mathrm{Con}(f'(x))$.

2. Если функция $f(x)$ выпукла на X, то для любого x из X, такого, что $0 \notin \partial f(x)$, справедливо тождество
$\mathcal{D} f(x) \equiv \mathrm{Con}(\partial f(x))$.

Доказательство. Выберем произвольную точку x_0 из X.

1. Для любой точки y из $\mathcal{M}(f, f(x_0))$ и α из отрезка $[0, 1]$ имеем
$$f(x_0) \geq f(\alpha y + (1 - \alpha) x) =$$
$$= f(x_0) + \alpha \langle f'(x_0), y - x_0 \rangle + o(\alpha).$$
Поэтому $\langle f'(x_0), x_0 - y \rangle \geq 0$ для всех y из множества $\mathcal{M}(f, f(x_0))$, и, следовательно,
$$\mathrm{Con}(f'(x_0)) \subseteq \mathcal{D} f(x_0).$$

Предположим теперь, что существует вектор g из $\mathcal{D} f(x_0)$, $g \neq t f'(x_0)$ ни при каком $t \geq 0$ Ясно, что в этом случае найдется направление s такое, что
$$\langle f'(x_0), s \rangle < 0 < \langle g, s \rangle. \qquad (1.5.1)$$
Левое неравенство (1.5.1) означает, что найдется достаточно малое $\alpha_0 > 0$, при котором выполнится неравенство
$$f(x_0 + \alpha_0 s) = f(x_0) + \alpha_0 \langle f'(x_0), s \rangle + o(\alpha_0) < f(x_0).$$
Следовательно, $x_0 + \alpha_0 s \in \mathcal{M}(f, f(x_0))$, что противоречит правому неравенству (1.5.1).

2. Пусть $g \in \partial f(x_0)$. Тогда в силу выпуклости функции $f(x)$ для любого $t \geq 0$ и любой точки y из множества $\mathcal{M}(f, f(x_0))$ имеем
$$\langle t g, x_0 - y \rangle \geq t (f(x_0) - f(y)) \geq 0.$$
Следовательно, $\mathrm{Con}(\partial f(x_0)) \subseteq \mathcal{D} f(x_0)$.

Предположим теперь, что существует вектор g_1 из $\mathcal{D} f(x_0)$ такой, что множества $\mathrm{Con}(g_1)$ и $\partial f(x_0)$ не пересекаются. Точка x_0 - внутренняя точка множества X. Поэтому в силу теоремы 1.2.5. ее субдифференциал является замкнутым ограниченным множеством. Следовательно, по теореме 1.1.3 найдется вектор $s \neq 0$ и число α такие, что при всех g из $\partial f(x_0)$ и произвольных $t > 0$ справедливы неравенства

$$\langle g, s \rangle < \alpha < t \langle g_1, s \rangle .$$

Правое из этих неравенств означает, что $\alpha < 0$ и $\langle g_1, s \rangle > 0$.

Теперь заметим, что при достаточно малых $t > 0$

$$f(x_0 + ts) = f(x_0) + f'(x_0, ts) + o(t) =$$
$$= f(x_0) + t \max \{ \langle g, s \rangle \mid g \in \partial f(x_0) \} + o(t) <$$
$$< f(x_0) + \alpha t + o(t) .$$

Таким образом, найдется число t_0, при котором

$$f(x_0 + t_0 s) < f(x_0) .$$

Функция $f(x)$ выпукла на открытом множестве X. Следовательно, в силу следствия 1.2.2 она является локально липшицевой в замкнутой окрестности точки $x_0 + t_0 s$. Последнее означает, что все точки $x_0 + t_0 s + y$, $\| y \| < \varepsilon$, содержатся в множестве $\mathcal{M}(f, f(x_0))$ при достаточно малом $\varepsilon > 0$.

Вектор $g_1 \in \mathcal{D}f(x_0)$. Поэтому при любых y, $\| y \| < \varepsilon$, справедливо неравенство

$$0 < \langle g_1, x_0 - (x_0 + t_0 s + y) \rangle =$$
$$= - t_0 \langle g_1, s \rangle - \langle g_1, y \rangle .$$

В силу произвольности y ($\| y \| < \varepsilon$) отсюда получаем, что $t_0 \langle g_1, s \rangle < - \varepsilon \| g_1 \|$, а это противоречит полученному ранее неравенству $\langle g_1, s \rangle > 0$. □

Условие $0 \notin \partial f(x)$ в теореме 1.5.3 существенно, что видно из следующего примера.

Пример 1.5.1. Рассмотрим функцию $f(x)$, $x \in R^n$, заданную формулой

$f(x) = \| x \| - x^{(1)}$.

Эта функция выпукла, причем

$X^* \equiv \mathcal{M}(f, 0) \equiv \{ x \in R^n \mid x^{(i)} = 0 , i = 2, \ldots, n, x^{(1)} \geq 0 \}$.

Нетрудно видеть, что

$\mathcal{D}f(0) = \{ g \in R^n \mid g^{(1)} \leq 0 \}$.

В то же время

$\partial f(0) = \{ g \in R^n \mid \| g - e_1 \| \leq 1 \}$,

и, следовательно,

$\mathrm{Con}\, (\partial f(0)) = \{ g \in R^n \mid g^{(1)} < 0 \} \cup \{ 0 \} \neq \mathcal{D}f(0)$.

Рассмотрим теперь функцию $\varphi(x) = (f(x))^2$. Эта функция выпукла, дифференцируема в точке $x = 0$, причем $\varphi'(0) = 0$. В то же время $\mathcal{M}(\varphi, 0) = \mathcal{M}(f, 0)$, $\mathcal{D}\varphi(0) = \mathcal{D}f(0)$, а $\mathrm{Con}\, (\varphi'(0)) = \{ 0 \}$. □

Как уже отмечалось выше, для квазивыпуклых функций операция максимума не выводит нас из этого класса.

<u>Теорема 1.5.4.</u> Пусть функции $f_1(x), \ldots, f_m(x)$ являются квазивыпуклыми на выпуклом множестве X. Тогда функция $\varphi(x) = \max \{ f_i(x) \mid i = 1, \ldots, m \}$ также будет квазивыпуклой на X , причем для любого x из X справедливо включение

$$\mathcal{D}\varphi(x) \supset \mathrm{Conv} \{ \mathcal{D} f_i(x) \mid i \in I(x) \} \equiv \equiv \sum_{i \in I(x)} \mathcal{D} f_i(x) , \qquad (1.5.2)$$

где $I(x) = \{ i \mid f_i(x) = \varphi(x) \}$.

Д о к а з а т е л ь с т в о. Квазивыпуклость функции $\varphi(x)$ очевидным образом следует из определения. Выберем произвольную точку $x_0 \in X$. Пусть $i \in I(x_0)$. Тогда для любого x из множества $\mathcal{M}(\varphi, \varphi(x_0))$

$$f_i(x) \leq \varphi(x) \leq \varphi(x_0) = f_i(x_0) .$$

Поэтому для любого $g_i \in \mathcal{D} f_i(x_0)$

$\langle g_i , x_0 - x \rangle \geq 0$.

Следовательно, $\mathcal{D}f_i(x_0) \subseteq \mathcal{D}\varphi(x_0)$. Но конус $\mathcal{D}\varphi(x_0)$ является выпуклым в силу своего определения. Поэтому и $\sum_{i \in I(x)} \mathcal{D}f_i(x)$ также принадлежит множеству $\mathcal{D}\varphi(x_0)$. $\square$

Приведем пример, показывающий, что в теореме 1.5.4 включение (1.5.2) нельзя заменить на тождество.

<u>Пример 1.5.2.</u> Рассмотрим две выпуклые функции:

$f_1(x) = \| x - e_1 \| - 1$, $f_2(x) = \| x + e_1 \| - 1$.

Нетрудно видеть, что

$\mathcal{D}f_1(0) = \{ g \in R^n \mid g = t e_1 , t \geq 0 \}$,

$\mathcal{D}f_2(0) = \{ g \in R^n \mid g = - t e_1 , t \geq 0 \}$.

Положим $\varphi(x) = \max \{ f_1(x) , f_2(x) \}$. Заметим, что $\varphi(x) > 0$ при любом $x \neq 0$, $\varphi(0) = 0$. Поэтому

$\mathcal{M}(\varphi , 0 \} \equiv \{ 0 \}$, $\mathcal{D} \varphi(0) \equiv R^n$.

В то же время

$\mathcal{D}f_1(0) + \mathcal{D}f_2(0) \equiv \{ g \in R^n \mid g = t e_1 , t \in R^n \}$.

$\square$

Как уже отмечалось выше, элементы множества $\mathcal{D}f(x)$ обладают одним важным свойством, определившим их широкое применение в методах минимизации: если y - произвольная точка из множества $\mathcal{M}(f , f(x)) \cap X$, то

$$\langle g , x - y \rangle \geq 0 \qquad (1.5.3)$$

для любого вектора g из $\mathcal{D}f(x)$. Таким образом, направление $- g$, $g \neq 0$, составляет в точке x острый угол с направлением на любую точку из множества $\mathcal{M}(f, f(x)) \cap X$, и в частности с направлениями на точку минимума.

Неравенство (1.5.3) можно уточнить, воспользовавшись мажорантой роста целевой функции по направлениям из точки y.

Итак, пусть y - фиксированная точка из множества X, а функция $f(x)$ непрерывна и квазивыпукла на R^n. Введем функцию

$$\omega(t) = \max\{ f(x) - f(y) \mid \| x - y \| = t \}.$$

Заметим, что $\omega(t) \equiv \omega(f, y, t)$ (в некоторых случаях во избежание недоразумений мы будем использовать последнее обозначение). Функция $\omega(t)$ является непрерывной неубывающей функцией от t, $t \geq 0$, причем $\omega(0) = 0$. Заметим, что для любого x из X

$$f(x) - f(y) \leq \omega(\| x - y \|). \qquad (1.5.4)$$

Оказывается, неравенства (1.5.3),(1.5.4) можно соединить в одно более сильное неравенство.

<u>Теорема 1.5.5.</u> Для любого $x \in X$, такого, что $f(x) > f(y)$ и $Df(x) \neq \{0\}$, выполняется неравенство

$$f(x) - f(y) \leq \omega(\langle g, x - y\rangle / \| g \|), \qquad (1.5.5)$$

где g - произвольный ненулевой вектор из $Df(x)$.

Д о к а з а т е л ь с т в о. Зафиксируем произвольную точку x, удовлетворяющую предположениям теоремы. Пусть g - произвольный ненулевой вектор из $Df(x)$. Положим

$$z = y + \| g \|^{-2} \langle g, x - y\rangle g.$$

Заметим, что $\langle g, z - x\rangle = 0$. Поэтому $f(z) \geq f(x)$. Следовательно, в силу неравенства (1.5.4)

$$f(x) - f(y) \leq f(z) - f(y) \leq \omega(\| z - y \|) = \omega(\langle g, x - y\rangle / \| g \|). \quad \square$$

<u>Следствие 1.5.1.</u> Для любого $x \in X$, такого, что $Df(x) \neq \{0\}$, выполняется неравенство

$$f(x) - f^* \leq \omega(f, x^*, \langle g, x - x^*\rangle / \| g \|),$$

где x^* - произвольная точка из X^*, g - произвольный ненулевой вектор из $Df(x)$.

Докажем утверждение, усиливающее следствие 1.5.1 для выпуклых функций.

Теорема 1.5.6. Пусть функция $f(x)$ выпукла на всем пространстве R^n. Тогда для всех $x \in X \setminus X^*$ и любого y из R^n справедливо неравенство

$$f(y) - f(x^*) \leqslant \omega(f, x^*,$$
$$|\langle g, x - x^* \rangle + f(y) - f(x)| / \| g \|), \qquad (1.5.6)$$

где g - произвольный ненулевой элемент из $\mathcal{D}f(x)$.

Доказательство. Зафиксируем точку $x \in X \setminus X^*$ и точку y из R^n. Положим

$$z = x^* + t g,$$

где $t = [\langle g, x - x^* \rangle + f(y) - f(x)] / \| g \|^2$. Тогда

$$f(y) = f(x) + \langle g, z - x \rangle \leqslant f(z).$$

Поэтому

$$f(y) - f(x^*) \leqslant f(z) - f(x^*) \leqslant \omega(\| z - x^* \|) =$$
$$= \omega(|t| \| g \|) = \omega(|\langle g, x - x^* \rangle + f(y) - f(x)| / \| g \|).$$

□

Следствие 1.5.2. Если функция $f(x)$ выпукла на всем пространстве R^n, то для всех $x \in X \setminus X^*$ и всех α из $(0, 1]$ выполняется неравенство

$$f(x) - f(x^*) \leqslant$$
$$\leqslant \alpha^{-1} \omega([\langle g, x - x^* \rangle - (1 - \alpha)(f(x) - f^*)] / \| g \|).$$

Действительно, при любом α из $(0, 1]$ существует точка $y \in X$ такая, что

$$f(y) - f(x^*) = \alpha (f(x) - f(x^*)).$$

Осталось воспользоваться неравенством (1.5.6). □

В оставшейся части этого параграфа мы введем понятие равномерно квазивыпуклой функции и приведем их основные свойства.

Определение 1.5.2. Функция $f(x)$, определенная на выпуклом множестве $X \subseteq R^n$, называется равномерно квазивыпуклой на X, если существует неотрицательная функция $\delta(t)$,

$\delta(0) = 0$, $\delta(t) > 0$ при $t > 0$, такая, что при всех x, y из X и любого α из $[0, 1]$ выполняется неравенство

$$f(\alpha x + (1 - \alpha) y) \le \max\{f(x), f(y)\} -$$
$$- \alpha(1 - \alpha)\, \delta(\| x - y \|). \qquad (1.5.7)$$

Функцию $\delta(t)$ назовем модулем квазивыпуклости функции $f(x)$ на X, а функцию

$$\mu(t) = \inf\{\alpha^{-1}(1 - \alpha)^{-1} [\max\{f(x), f(y)\} -$$
$$- f(\alpha x + (1 - \alpha) y)] \mid$$
$$x, y \in X, \| x - y \| = t, 0 < \alpha < 1\},$$

будем называть точным модулем квазивыпуклости функции $f(x)$ на X. Ясно, что $\mu(t) \ge \delta(t)$ для любого модуля квазивыпуклости $\delta(t)$.

Нетрудно видеть, что любая функция, являющаяся равномерно выпуклой с модулем $\delta(t)$, будет равномерно квазивыпуклой с тем же модулем.

<u>Лемма 1.5.2.</u> Если функции $f_1(x)$ и $f_2(x)$ являются равномерно квазивыпуклыми на выпуклом множестве X с модулями квазивыпуклости $\delta_1(t)$ и $\delta_2(t)$ соответственно, то функция $\varphi(x) = \min\{f_1(x), f_2(x)\}$ также будет равномерно квазивыпуклой на X с модулем квазивыпуклости

$$\delta(t) = \min\{\delta_1(t), \delta_2(t)\}.$$

Действительно, для любых x, y из X имеем

$$\varphi(\alpha x + (1 - \alpha) y) =$$
$$= \max\{f_1(\alpha x + (1 - \alpha) y), f_2(\alpha x + (1 - \alpha) y)\} \le$$
$$\le \max\{\max\{f_1(x), f_1(y)\} - \alpha(1 - \alpha)\, \delta_1(\| x - y \|),$$
$$\max\{f_2(x), f_2(y)\} - \alpha(1 - \alpha)\, \delta_2(\| x - y \|)\} \le$$
$$\le \max\{\max\{f_1(x), f_1(y)\} - \alpha(1 - \alpha)\, \delta(\| x - y \|),$$
$$\max\{f_2(x), f_2(y)\} - \alpha(1 - \alpha)\, \delta(\| x - y \|)\} =$$
$$= \max\{f_1(x), f_1(y), f_2(x), f_2(y)\} -$$

$- \alpha(1 - \alpha)\ \delta(\| x - y \|)\} =$

$= \max\{\varphi(x), \varphi(y)\} - \alpha(1 - \alpha)\ \delta(\| x - y \|)\}$.

□

<u>Лемма 1.5.3.</u> Пусть функция $f_1(x)$ является равномерно выпуклой на выпуклом множестве X с модулем выпуклости $\delta_1(t)$, а функция $f_2(x)$ вогнута на X. Тогда, если при всех x из X функция $f_1(x) > 0$, а функция $f_2(x) > 0$ и если

$\sup\{ f_2(x) \mid x \in X\} \equiv M < \infty$,

то функция $\varphi(x) = f_1(x) / f_2(x)$ будет равномерно квазивыпуклой на X с модулем квазивыпуклости

$\delta(t) = M^{-1}\ \delta_1(t)$.

Д о к а з а т е л ь с т в о. Выберем произвольные x, y из X. Тогда

$\varphi(\alpha x + (1 - \alpha) y) =$

$= f_1(\alpha x + (1 - \alpha) y) / f_2(\alpha x + (1 - \alpha) y) <$

$< f_1(\alpha x + (1 - \alpha) y) / [\alpha f_2(x) + (1 - \alpha) f_2(y)] <$

$< [\alpha f_1(x) + (1 - \alpha) f_1(y) - \alpha(1 - \alpha)\ \delta_1(\| x - y \|)] /$

$/ [\alpha f_2(x) + (1 - \alpha) f_2(y)] <$

$< [\alpha f_2(x)\ \varphi(x) + (1 - \alpha) f_2(y)\ \varphi(y)] /$

$/ [\alpha f_2(x) + (1 - \alpha) f_2(y)] -$

$- \alpha(1 - \alpha) M^{-1}\ \delta_1(\| x - y \|)] <$

$< \max\{\varphi(x), \varphi(y)\} - \alpha(1 - \alpha) M^{-1}\ \delta_1(\| x - y \|)$.

□

В формулировке леммы 1.5.3, как и в формулировке леммы 1.5.1, в случае линейной функции $f_2(x)$ требование $f_1(x) > 0$ можно опустить.

Отметим, что равномерно квазивыпуклые функции в отличие от функций равномерно выпуклых не всегда ограничены снизу. Так, например, функция $f(x) = x$, $x \in R$, является равномерно квазивыпуклой с точным модулем $\mu(t) = t$, однако $\inf\{ f(x) \mid x \in R \} = -\infty$.

Теорема 1.5.7. Пусть функция $f(x)$ является равномерно квазивыпуклой на выпуклом множестве $X \subseteq R^n$, а $\delta(t)$ – ее модуль квазивыпуклости. Тогда, если функция $f(x)$ дифференцируема по направлениям в точке $x_0 \in X$, то для любого y из $\mathcal{M}(f, f(x_0)) \cap X$ справедливо неравенство
$f'(x_0, y - x_0) \le -\delta(\| y - x_0 \|)$.

Доказательство. По определению

$$f'(x_0, y - x_0) = \lim_{\alpha \to +0} \alpha^{-1} [f(x_0 + \alpha(y - x_0)) - f(x_0)] \le$$

$$\le \lim_{\alpha \to +0} \alpha^{-1} [\max\{ f(y), f(x_0) \} -$$

$$- \alpha(1 - \alpha) \delta(\| y - x_0 \|) - f(x_0)] =$$
$$= -\delta(\| y - x_0 \|). \quad \square$$

Следствие 1.5.3. Пусть равномерно квазивыпуклая на выпуклом множестве X функция $f(x)$ дифференцируема по направлениям в некоторой точке $x^* \in X$. Тогда, если для всех x из X выполняется неравенство $f'(x^*, x - x^*) \ge 0$, то x^* – единственная точка минимума функции $f(x)$ на X.

Действительно, если найдется $y \in \mathcal{M}(f, f(x^*)) \cap X$, то по теореме 1.5.7
$f'(x^*, y - x^*) \le -\delta(\| y - x^* \|) < 0$,
что противоречит сделанным предположениям. $\square$

В следующей теореме дается оценка скорости роста точного модуля квазивыпуклости $\mu(t)$.

Теорема 1.5.8. Пусть $\mu(t)$ – точный модуль квазивыпуклости функции $f(x)$ на выпуклом множестве $X \subseteq R^n$. Тогда для любых $C \ge 1$, $t \ge 0$, таких, что функция $\mu(t)$ определена в точке $C t$, выполнено неравенство

$\mu(Ct) \geq C\mu(t)$. (1.5.8)

Д о к а з а т е л ь с т в о. Зафиксируем числа $t > 0$ и $C \in (1, 3)$ такие, что Ct принадлежит области определения функции $\mu(t)$. Докажем сначала неравенство

$$\mu(Ct) \geq (C - 3/4\ (C-1)^2)\,\mu(t). \qquad (1.5.9)$$

Выберем произвольное $\varepsilon > 0$. Тогда по определению 1.5.2 найдутся точки $x_1, x_2 \in X$ и число $\alpha \in (0, 1)$ такие, что

$$\mu(Ct) + \varepsilon \geq \alpha^{-1}(1-\alpha)^{-1}[\max\{f(x_1), f(x_2)\} - f(\alpha x_1 + (1-\alpha)x_2)]. \qquad (1.5.10)$$

Возможны три случая.

1. Пусть $0 < \alpha \leq 0.5\, C^{-1}(C-1)$. Положим

$x_3 = C^{-1}x_1 + C^{-1}(C-1)x_2$.

Тогда $\alpha x_1 + (1-\alpha)x_2 \in [x_3, x_2]$. Учитывая, что $\| x_3 - x_2 \| \leq t$, получаем

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \max\{f(x_2), f(x_3)\} - \alpha C(1-\alpha C)\mu(t) \leq \max\{f(x_1), f(x_2)\} - \alpha C(1-\alpha C)\mu(t).$$

Поэтому

$$\mu(Ct) + \varepsilon \geq (1-\alpha)^{-1}C(1-\alpha C)\mu(Ct) \geq (C - 3/4\ (C-1)^2)\,\mu(t),$$

откуда в силу произвольности ε получаем (1.5.9).

2. Пусть $0.5\, C^{-1}(C-1) \leq \alpha \leq 0.5\, C^{-1}(C+1)$. Тогда точка $\alpha x_1 + (1-\alpha)x_2 \in [x_4, x_5]$, где

$x_4 = 0.5\, C^{-1}(C+1)x_1 + 0.5\, C^{-1}(C-1)x_2$;

$x_5 = 0.5\, C^{-1}(C-1)x_1 + 0.5\, C^{-1}(C+1)x_2$.

Заметим, что

$\| x_1 - x_4 \| = \| x_2 - x_5 \| = 0.5\, C^{-1}(C-1)\,t \leq t$;

$\| x_4 - x_5 \| = t$.

Поэтому $x_5 \in [x_3, x_2]$, и можно записать

$$f(x_5) < \max\{ f(x_2), f(x_3) \} - 0.25\,(C-1)(3-C)\,\mu(t) <$$
$$< \max\{ f(x_1), f(x_2) \} - 0.25\,(C-1)(3-C)\,\mu(t).$$

Аналогичное неравенство можно записать и для значения функции $f(x)$ в точке x_4. Поэтому в силу равномерной квазивыпуклости функции $f(x)$

$$\max\{ f(x_1), f(x_2) \} - 0.25\,(C-1)(3-C)\,\mu(t) >$$
$$> \max\{ f(x_4), f(x_5) \} > f(\alpha x_1 + (1-\alpha) x_2) +$$
$$+ (\alpha C - 0.5(C-1))(0.5(C+1) - \alpha C)\,\mu(t).$$

Подставив последнее неравенство в (1.5.10), после несложных преобразований получим

$$\mu(Ct) + \varepsilon > [C^2 - 0.5\,\alpha^{-1}(1-\alpha)^{-1}(C-1)^2]\,\mu(t) >$$
$$> (C - 3/4\,(C-1)^2)\,\mu(t),$$

что в силу произвольности ε дает (1.5.9).

3. Случай $0.5\,C^{-1}(C-1) < \alpha < 1$ рассматривается так же, как и случай 1.

Итак, неравенство (1.5.9) доказано. Пусть теперь $t > 0$, $C > 1$ и $C\,t$ принадлежит области определения функции $\mu(t)$. Выберем целое положительное число n таким образом, чтобы $b = C^{1/n} < 2$. Заметим, что в этом случае в силу неравенства (1.5.9) $\mu(b^{k+1} t) > \mu(b^k t)$ при всех $k = 0, \ldots, n-1$. Поэтому, еще раз воспользовавшись неравенством (1.5.9), можно записать

$$b^{-1} t^{-1} \mu(bt) > t^{-1}\mu(t) - 3/4\, b^{-1} t^{-1} (b-1)^2\, \mu(t) >$$
$$> t^{-1}\mu(t) - 3/4\, t^{-1} (b-1)^2\, \mu(C\,t);$$
$$b^{-2} t^{-1} \mu(b^2 t) > b^{-1} t^{-1} \mu(bt) - 3/4\, b^{-2} t^{-1} (b-1)^2\, \mu(bt) >$$
$$> b^{-1} t^{-1} \mu(b\,t) - 3/4\, t^{-1} (b-1)^2\, \mu(C\,t);$$

. . . .

$$C^{-1} t^{-1} \mu(C\,t) >$$
$$> b^{n-1} t^{-1} \mu(b^{n-1} t) - 3/4\, t^{-1} (b-1)^2\, \mu(C\,t)$$

Складывая эти неравенства, получаем

$C^{-1}\,t^{-1}\,\mu(C\,t) \geq t^{-1}\,\mu(t) - .3\,n/4\,t^{-1}\,(b-1)^2\,\mu(C\,t)$.

Заметим, что

$$\lim_{n\to\infty} n\,(b-1)^2 \equiv \lim_{n\to\infty} n\,(C^{1/n}-1)^2 = 0 .$$

Поэтому, устремляя в последнем неравенстве n к бесконечности, получаем неравенство (1.5.8). □

Пример функции $f(x) = x$, $x \in R^n$, показывает, что неравенство (1.5.8) дает неулучшаемую оценку скорости роста точного модуля квазивыпуклости.

Следствие 1.5.4. Пусть для равномерно квазивыпуклой функции $f(x)$ множество $\mathcal{M}(f, t_0) \cap X$ непусто и ограничено при некотором $t_0 \in R$. Тогда множество $\mathcal{M}(f, t) \cap X$ ограничено при всех t из R.

Д о к а з а т е л ь с т в о. Пусть множество $\mathcal{M}(f, t_0) \cap X$ непусто и ограничено. Тогда для всех $t < t_0$ множества $\mathcal{M}(f, t) \cap X \subset \mathcal{M}(f, t_0) \cap X$ также ограничены. Предположим теперь, что при некотором $t > t_0$ множество $\mathcal{M}(f, t) \cap X$ ограниченным не является. Это означает, что найдется последовательность точек $\{x_k\}$ такая, что $t_0 < f(x_k) < t$ и $\| x_k \| \to \infty$ при $k \to \infty$. Выберем произвольную точку x из $\mathcal{M}(f, t_0) \cap X$. Тогда начиная с некоторого номера N точки $0.5\,x + 0.5\,x_k \notin \mathcal{M}(f, t_0) \cap X$. Поэтому для всех $k > N$ справедливо неравенство

$$t > f(x_k) = \max\{ f(x_k),\ f(x) \} \geq$$
$$\geq f(0.5\,x_k + 0.5\,x) + 0.25\,\mu(\| x_k - x_0 \|) \geq$$
$$\geq t_0 + 0.25\,\mu(\| x_k - x_0 \|),$$

где $\mu(t)$ - точный модуль квазивыпуклости функции $f(x)$. Но в силу теоремы 1.5.8 $\mu(t) \to \infty$ при $t \to \infty$. Поэтому, устремляя $k \to \infty$, приходим к противоречию. □

Следствие 1.5.5. Пусть $\delta(t)$ - модуль квазивыпуклости функции $f(x)$. Тогда функция

$$\delta^*(t) = \sup\{\tau^{-1}\delta(\tau) \mid 0 < \tau \le t\}\, t$$

также будет модулем квазивыпуклости функции $f(x)$.

Действительно, если $\mu(t)$ - точный модуль квазивыпуклости функции $f(x)$, то при любом τ, $0 < \tau \le t$, имеем

$$\tau^{-1}\delta(\tau) \times t \le t\,\tau^{-1}\mu(\tau) \le \mu(t).$$

Поэтому $\delta^*(t) \le \mu(t)$ и, следовательно, является модулем квазивыпуклости функции $f(x)$. □

Докажем для равномерно квазивыпуклых функций аналог теоремы 1.3.2.

<u>Теорема 1.5.9.</u> Пусть $f(x)$ - равномерно квазивыпуклая на выпуклом множестве X функция. Тогда если

$$f^* = \inf\{f(x) \mid x \in X\} > -\infty,$$

то 1) множества $\mathcal{M}(f, t) \cap X$ ограничены при всех $t \in R$;

2) любая минимизирующая последовательность точек $\{x_k\}$, такая, что $x_k \in X$, $f(x_k) \to f^*$, сходится к единственной точке x^*;

3) при всех $x \in X$ справедливо неравенство

$$\mu(\| x - x^* \|) \le 4\,(f(x) - f^*), \tag{1.5.11}$$

где $\mu(t)$ - точный модуль квазивыпуклости функции $f(x)$.

Д о к а з а т е л ь с т в о. Зафиксируем произвольную точку $x_0 \in X$. И пусть $x, y \in \mathcal{M}(f, f(x_0)) \cap X$. Тогда

$$0.25\,\mu(\| x - y \|) \le \max\{f(x), f(y)\} -$$
$$- f(0.5x + 0.5y) \le f(x_0) - f^* < \infty,$$

т.е. все множества $\mathcal{M}(f, t) \cap X$ ограничены.

Пусть теперь $\{x_k\}$ - минимизирующая последовательность. Тогда для любых номеров k, m

$$0.25\,\mu(\| x_k - x_m \|) \le \max\{f(x_k), f(x_m)\} -$$

$- f(0.5 x_k + 0.5 x_m) < \max \{ f(x_k) , f(x_m) \} - f^*$.
Таким образом, любая минимизирующая последовательность точек является фундаментальной и сходится к единственной точке x^*.

И, наконец, пусть x - произвольная точка из X, а $\{ x_k \}$ - минимизирующая последовательность. Тогда

$$\max \{ f(x_k) , f(x) \} > f(0.5 x_k + 0.5 x) + \\ + 0.25 \mu(\| x_k - x \|) > f^* + 0.25 \mu(\| x_k - x \|) . \tag{1.5.12}$$

Переходя в (1.5.12) к пределу по $k \to \infty$ получаем (1.5.11).
□

Условие $f^* > - \infty$ использовалось в теореме 1.5.9 только для доказательства ограниченности множества $\mathcal{M}(f , f(x_0)) \cap X$. Однако ограниченность этого множества может быть выведена и из некоторых дополнительных условий на модуль равномерной квазивыпуклости функции $f(x)$.

<u>Теорема 1.5.10.</u> Пусть функция $f(x)$ является равномерно квазивыпуклой на выпуклом множестве $X \subset R^n$ с модулем квазивыпуклости $\delta(t)$. Тогда, если найдется точка $x_0 \in X$, в которой функция $f(x)$ дифференцируема и

$$\sup \{ t^{-1} \delta(t) \mid t > 0 \} > \| f'(x_0) \| ,$$

(например, возможно $f'(x_0) = 0$ или $t^{-1} \delta(t) \to \infty$ при $t \to \infty$), то множества $\mathcal{M}(f , t) \cap X$ ограниченн при всяком $t \in R$.

Д о к а з а т е л ь с т в о. В силу следствия 1.5.4 нам достаточно доказать ограниченность множества $\mathcal{M}(f , f(x_0))$. Предположим, что это множество не ограничено. Рассмотрим последовательность чисел $\{t_k\}$ таких, что

$$\lim_{k \to \infty} t_k^{-1} \delta(t_k) = \sup \{ t^{-1} \delta(t) \mid t > 0 \},$$

и выберем последовательность точек $\{ x_k \} \subset \mathcal{M}(f , f(x_0))$

$\cap X$ такую, что $\| x_k - x_0 \| = t_k$ (это возможно в силу предположения о неограниченности множества $\mathcal{M}(f, f(x_0))$ $\cap X$). Тогда по теореме 1.5.7

$$\lim_{k\to\infty} t_k^{-1} \delta(t_k) = \lim_{k\to\infty} \| x_k - x_0 \|^{-1} \delta(\| x_k - x_0 \|) \leq$$

$$\leq \| f'(x_0) \| ,$$

что противоречит условию теоремы. □

Перейдем к обоснованию достаточных условий равномерной квазивыпуклости.

<u>Лемма 1.5.4.</u> Пусть $\varkappa(t)$ - неубывающая на $[0, a]$ функция, $\varkappa(0) = 0$, и пусть функция $g(u)$ абсолютно непрерывна на $[0, a]$, причем $g(a) = 0 \geq g(a)$. Тогда, если

$$g'(u)(u - v) \geq \varkappa(|u - v|)\, |u - v| \qquad (1.5.13)$$

для всех $u, v \in [0, a]$, таких, что $g(u) \geq g(v)$, то

$$g(\alpha a) \leq -\alpha(1-\alpha) \int_0^a \varkappa(t)\, dt$$

для любого α из $(0, 1)$.

Д о к а з а т е л ь с т в о. Сразу заметим, что из (1.5.13) следует квазивыпуклость функции $g(u)$ на $[0, a]$. Действительно, если, например, $0 \leq u_1 < v < u_2 \leq a$ и $g(u_2) \leq g(u_1) < g(v)$, то найдется такая точка v^*, $v < v^* < u_2$, что $g(v^*) \geq g(u_1)$ и $g'(v^*) < 0$, т.е. $g'(v^*) \times \mathrm{sign}(v^* - u_1) < 0$, что противоречит неравенству (1.5.13).

Возьмем некоторое $x \in [\inf\{ g(u) \mid 0 \leq u \leq a \}, 0]$ и положим

$$u_1(x) = \min\{ u \mid 0 \leq u \leq a,\ g(u) = x \} ;$$

$$u_2(x) = \begin{cases} a, \text{ если } x > g(a), \\ \max\{ u \mid 0 < u < a,\ g(u) = x \}, \text{ если } x < g(a); \end{cases}$$

$\varphi(x) = u_2(x) - u_1(x)$.

Положим $g^* = \inf\{ x \mid æ(\varphi(x)) > 0 \}$. Если $g^* = 0$, то $æ(t) \equiv 0$ при $0 < t < a$, и утверждение леммы следует из квазивыпуклости функции $g(u)$. Пусть $g^* < 0$. Зафиксируем x' такое, что $0 > x' > g^*$. Пусть $0 < u < u_1(x')$ и существует $g'(u)$. Положив в (1.5.13) $v = u_2(x')$, получим

$$g'(u) < -æ(u_2(x') - u) < -æ(u_2(x') - u_1(x')) = -æ(\varphi(x')) < 0$$

Таким образом, функция $g(u)$ строго убывает на $[0, u_1(x')]$, а обратная к ней функция $u_1(x)$ удовлетворяет на $[x', 0]$ условию Липшица с константой $1/æ(\varphi(x'))$, причем $u'_1(x) = 1 / g'(u_1(x))$ почти всюду на $(g^*, 0]$. Проводя аналогичные рассуждения для $u_2(x)$, получим, что

$u'_2(x) = 0$, если $x > g(a)$,

$u'_2(x) = 1 / g'(u_2(x))$ почти всюду на $(g^*, 0]$.

Ясно также, что

$u_1(x) \to u_1(g^*)$ при $x \to g^* + 0$,

$u_2(x) \to u_2(g^*)$ при $x \to g^* + 0$.

Таким образом, $\varphi(x) \to \varphi(g^*)$ при $x \to g^* + 0$.

Итак, $\varphi(x)$ - абсолютно непрерывная, неубывающая на $[g^*, 0]$ функция. Положив в (1.5.13) $u = u_1(x)$, $v = a$, если $x < g(a)$, и $v = u_2(x)$, если $g^* < x < g(a)$, получим

$$u'_1(x) > -1 / æ(\varphi(x)) \tag{1.5.14}$$

почти всюду на $(g^*, 0]$. Аналогично

$u'_2(x) < 1 / æ(\varphi(x))$

почти всюду на $(g^{*} , 0]$, а значит,

$\varphi'(x) = u'_2(x) - u'_1(x) < 2 / æ(\varphi(x))$

почти всюду на $(g^{*} , 0]$. Это неравенство можно переписать в виде

$$\frac{d}{d x} \left(\int_{\varphi(x)}^{a} æ(t)\, dt \right) > - 2 ,$$

откуда получим

$$\int_{\varphi(x)}^{a} æ(t)\, dt < - 2 x \qquad (1.5.15)$$

при $g^{*} < x < 0$.

Рассмотрим вспомогательную функцию

$$S(u) = \begin{cases} - \int_0^{u} æ(a - 2 t)\, dt & \text{при } 0 < u < 0.5 a , \\ - \int_0^{a-u} æ(a - 2 t)\, dt & \text{при } 0.5 a < u < a. \end{cases}$$

Положим

$v_1(x) = \min \{ u \mid 0 < u < a , S(u) = x \}$;

$v_2(x) = \max \{ u \mid 0 < u < a , S(u) = x \} = a - v_1(x)$;

$\xi(x) = v_2(x) - v_1(x) = a - 2 v_1(x)$;

$$S^{*} = \min \{ S(u) \mid 0 < u < a \} = - 0.5 \int_0^{a} æ(t)\, dt.$$

Нетрудно видеть, что функции $v_1(x)$, $v_2(x)$ и $\xi(x)$ абсолютно непрерывны на $(S^{*} , 0]$, причем

$v'_1(x) = 1 / S'(v_1(x)) =$

$= - 1 / æ(a - 2 v_1(x)) = - 1 / æ(\xi(x)) \qquad (1.5.16)$

почти всюду на $(S^{*} , 0]$.

Покажем, что $g(u) < S(u)$ для всех $0 < u < a$.

Для этого достаточно показать, что $g^* < S^*$ и $v_1(x) > u_1(x)$, $v_2(x) < u_2(x)$ при всех $x \in (g^*, 0]$. Из (1.5.15) мы получим

$$g^* < -0.5 \int_{\varphi(g^*)}^{a} æ(t)\,dt .$$

Покажем, что $æ(\varphi) = 0$ для любого $\varphi \in [0, \varphi(g^*)]$. Предположим, что существует $\varphi_0 < \varphi(g^*)$ такое, что $æ(\varphi_0) > 0$. Рассмотрим два отрезка:

$[u_1(g^*), u_1(g^*) + 0.5(\varphi(g^*) - \varphi_0)]$,

$[u_2(g^*) - 0.5(\varphi(g^*) - \varphi_0), u_2(g^*)]$.

Возможны два случая.

1. $g(u) \equiv g^*$ на одном из отрезков, например, на первом. Тогда, взяв в неравенстве (1.5.13) за u внутреннюю точку этого отрезка, а за v точку $u_2(g^*)$, получим

$0 = æ(|u - u_2(g^*)|) > \xi(\varphi_0)$,

что противоречит $æ(\varphi_0) > 0$.

2. $g(u)$ не является постоянной ни на одном из отрезков. Тогда в силу квазивыпуклости функции $g(u)$ найдутся точки v_1, v_2 такие, что $u_1(g^*) < v_1 < v_2 < u_2(g^*)$, $v_2 - v_1 > \varphi_0$, $g(v_1) < g^*$, $g(v_2) < g^*$. Если, например, $g(v_1) > g(v_2)$, то $\varphi(g(v_1)) > v_2 - v_1 > \varphi_0$, что противоречит определению g^*.

Таким образом, $æ(\varphi) = 0$ для всех $\varphi \in [0, \varphi(g^*)]$, и поэтому

$$g^* < -0.5 \int_{\varphi(g^*)}^{a} æ(t)\,dt = -0.5 \int_{0}^{a} æ(t)\,dt = S^*.$$

Покажем теперь, что

$$u'_1(x) > v'_1(x) \tag{1.5.17}$$

почти всюду на $(S^*, 0]$. Действительно, из (1.5.16) следует, что $\xi'(x) = 2 / æ(\xi(x))$, т.е.

$$\int_{\xi(x)}^{a} æ(t)\,dt = -2x > \int_{\varphi(x)}^{a} æ(t)\,dt .$$

При этом $æ(\varphi(x)) > 0$, а значит, $\xi(x) < \varphi(x)$ и $æ(\xi(x)) < æ(\varphi(x))$. Поэтому (1.5.17) сразу следует из (1.5.16) и (1.5.14). Теперь заметим, что $u_1(0) = v_1(0) = 0$. Таким образом, в силу (1.5.17) $v_1(x) > u_1(x)$ при всех $x \in (S^*, 0]$. Аналогично показывается, что $v_2(x) < u_2(x)$ при $x \in (S^*, 0]$.

Итак, $g(u) < S(u)$ для всех u, $0 < u < a$. Поэтому при $\alpha \in (0, 0.5]$ имеем

$$-g(\alpha a) > -S(\alpha a) = \int_0^{\alpha a} æ(a - 2t)\,dt =$$

$$= \alpha \int_0^{a} æ(a - 2\alpha t)\,dt > \alpha \int_0^{a} æ(a - t)\,dt =$$

$$= \alpha \int_0^{a} æ(t)\,dt > \alpha(1-\alpha) \int_0^{a} æ(t)\,dt .$$

Случай $0.5 < \alpha < 1$ рассматривается аналогично. □

<u>Теорема 1.5.11.</u> Пусть функция $f(x)$ дифференцируема по направлениям в каждой точке выпуклого множества X. Тогда, если для любых точек x, y из X, $f(y) < f(x)$, выполнено неравенство

$$f'(x, x-y) > \delta(\|x-y\|),$$

где $\delta(t)$ - строго возрастающая при $t > 0$ функция, то функция $f(x)$ будет равномерно квазивыпуклой на X с модулем квазивыпуклости

$$\delta(t) = \int_0^t \tau^{-1}\, \gamma(t)\, dt .$$

Доказательство. Зафиксируем произвольные $x, y \in X$, $x \neq y$. Пусть $f(x) > f(y)$. Рассмотрим фуйкцию

$g(u) = f(x + u(y - x) / \| y - x \|) - f(x)$,

где $u : 0 < u < \| y - x \|$. Нетрудно видеть, что для функции $g(u)$ выполнены все условия леммы 1.5.4 с $\varkappa(t) = t^{-1}\,\gamma(t)$. Поэтому

$f(\alpha x + (1 - \alpha) y) - \max\{ f(x), f(y) \} =$

$= f(\alpha x + (1 - \alpha) y) - f(x) = g((1 - \alpha)\| y - x \|) <$

$$< \alpha(1 - \alpha) \int_0^{\| y-x \|} \tau^{-1}\, \gamma(\tau)\, d\tau . \quad \square$$

1.6. Квазиоднородные функции

Характеристическим свойством квадратичной функции $f(x)$ является выполнение при всех x, y из R^n следующего соотношения:

$$f(y) \equiv f(x) + \langle f'(x), y - x \rangle + 0.5 \langle f'(x) - f'(y), x - y \rangle . \qquad (1.6.1)$$

Характеристическое неравенство для квазиоднородных функций можно интерпретировать как ослабление тождества (1.6.1).

<u>Определение 1.6.1.</u> Дифференцируемая на R^n выпуклая функция $f(x)$ называется квазиоднородной, если существует константа $p(f) > 0$ такая, что при всех x, y из R^n выполняется неравенство

$$f(y) \geq f(x) + \langle f'(x), y - x\rangle +$$
$$+ p(f)\langle f'(x) - f'(y), x - y\rangle . \qquad (1.6.2)$$

Величину $p(f)$ будем называть степенью квазиоднородности функции $f(x)$.

Докажем характеристическую лемму для квазиоднородных функций.

Лемма 1.6.1. Функция $f(x)$ является квазиоднородной тогда и только тогда, когда найдется константа $p(f)$, $p(f) > 0$ такая, что при всех x, y из R^n и α из $[0, 1]$ справедливо неравенство

$$f(\alpha x + (1 - \alpha) y) \leq \alpha f(x) + (1 - \alpha) f(y) -$$
$$- \alpha(1 - \alpha) p(f)\langle f'(x) - f'(y), x - y\rangle . \qquad (1.6.3)$$

Доказательство. Пусть функция $f(x)$ является квазиоднородной. Зафиксируем произвольные точки x, y из R^n и число $\alpha \in [0, 1]$. Положим $z = \alpha x + (1 - \alpha) y$. Тогда

$$f(x) - f(z) \geq \langle f'(z), x - z\rangle + p(f)\langle f'(x) - f'(z), x - z\rangle = (1 - \alpha)[\langle f'(z), x - y\rangle + p(f)\langle f'(x) - f'(z), x - y\rangle];$$

$$f(y) - f(z) \geq \langle f'(z), y - z\rangle + p(f)\langle f'(y) - f'(z), y - z\rangle = -\alpha[\langle f'(z), x - y\rangle + p(f)\langle f'(y) - f'(z), x - y\rangle].$$

Поэтому

$$\alpha f(x) + (1 - \alpha) f(y) - f(z) =$$
$$= \alpha[f(x) - f(z)] + (1 - \alpha)[f(y) - f(z)] \geq$$
$$\geq \alpha(1 - \alpha)[\langle f'(z), x - y\rangle + p(f)\langle f'(x) - f'(z), x - y\rangle - \langle f'(z), x - y\rangle - p(f)\langle f'(y) - f'(z), x - y\rangle] = \alpha(1 - \alpha) p(f)\langle f'(x) - f'(y), x - y\rangle.$$

Пусть теперь функция $f(x)$ удовлетворяет при всех $x, y \in R^n$ и $\alpha \in [0, 1]$ неравенству (1.6.3). Тогда

$\alpha(1-\alpha)^{-1}[\, f(x) - f(\alpha x + (1-\alpha) y)\,] + f(y) -$
$- f(\alpha x + (1-\alpha) y) \geq \alpha\, p(f) \langle f'(x) - f'(y),\, x - y\rangle$.

Устремляя в этом неравенстве α к единице, получаем неравенство (1.6.2). □

Приведем несколько элементарных свойств квазиоднородных функций.

1. Если функция $f(x)$ является квазиоднородной, то при любом $\alpha > 0$ функция $\alpha f(x)$ также будет квазиоднородной, причем $p(\alpha f) = p(f)$.

2. Если функция $f_1(x)$ квазиоднородна, а $f_2(x)$ - линейна, то функция $f_1(x) + f_2(x)$ также квазиоднородна и $p(f_1 + f_2) = p(f_1)$.

3. Если функции $f_1(x)$ и $f_2(x)$ являются квазиоднородными, то при любых $\alpha, \beta > 0$ функция $\alpha f_1(x) + \beta f_2(x)$ также будет квазиоднородной функцией, причем $p(\alpha f_1 + \beta f_2) \geq \min\{ p(f_1), p(f_2) \}$.

4. Если функция $f(x)$, $x \in R^n$, квазиоднородна и A - линейный оператор, действующий из R^m в R^n, то функция $\varphi(y) = f(Ay)$ будет квазиоднородной функцией с той же степенью квазиоднородности $p(\varphi) = p(f)$.

Лемма 1.6.2. При всех x, y из R^n справедливы неравенства

$$f(y) \leq f(x) + \langle f'(x),\, y - x\rangle + (1 - p(f)) \langle f'(x) - f'(y),\, x - y\rangle,$$
$$f(\alpha x + (1-\alpha) y) \geq \alpha f(x) + (1-\alpha) f(y) - \alpha(1-\alpha)(1 - p(f)) \langle f'(x) - f'(y),\, x - y\rangle,$$

где α такое, что $0 \leq \alpha \leq 1$.

Доказательство. Действительно,

$$f(x) \geq f(y) + \langle f'(y),\, x - y\rangle + p(f) \langle f'(x) - f'(y),\, x - y\rangle =$$

$= f(y) + (1 - p(f)) \langle f'(y), x - y \rangle +$
$+ p(f) \langle f'(x), x - y \rangle =$
$= f(y) + (1 - p(f)) \langle f'(y) - f'(x), x - y \rangle +$
$+ \langle f'(x), x - y \rangle$.

Далее, положим $z = \alpha x + (1 - \alpha) y$. Тогда в силу только что доказанного неравенства

$\alpha f(x) + (1 - \alpha) f(y) - f(z) =$
$= \alpha [f(x) - f(z)] + (1 - \alpha) [f(y) - f(z)] \leq$
$\leq \alpha(1 - \alpha) [\langle f'(z), x - y\rangle + (1 - p(f)) \langle f'(x) -$
$- f'(z), x - y \rangle - \langle f'(z), x - y \rangle - (1 - p(f)) \langle f'(y) -$
$- f'(z), x - y\rangle] = \alpha(1-\alpha)(1 - p(f))\langle f'(x) - f'(y), x - y\rangle$.

□

Следствие 1.6.1. Если функция $f(x)$ квазиоднородна и x^* - точка ее безусловного минимума, то при любом $x \in R^n$

$p(f) \langle f'(x), x - x^* \rangle \leq f(x) - f(x^*) \leq$
$\leq (1 - p(f)) \langle f'(x), x - x^* \rangle$.

Лемма 1.6.3. Если функция $f(x)$ квазиоднородна, то при любых точках $x, y \in R^n$ и числе $t \in [0, 1]$ выполняются неравенства

$f(x + t(y - x)) - f(x) - t \langle f'(x), y - x \rangle \geq$
$\geq t^{1/p(f)} [f(y) - f(x) - \langle f'(x), y - x \rangle],$ (1.6.4)

$f(x + \alpha(y - x)) - f(x) - \alpha\langle f'(x), y - x \rangle \leq$
$\leq t^{1/(1-p(f))} [f(y) - f(x) - \langle f'(x), y - x \rangle]$.

(1.6.5)

Если же $t \geq 1$, то знаки обоих неравенств меняются на противоположные.

Д о к а з а т е л ь с т в о. Зафиксируем произвольную точку x_1 из R^n. Рассмотрим функцию

$\varphi(x) = f(x) - f(x_1) - \langle f'(x_1), x - x_1 \rangle$.

Как уже отмечалось выше, функция $\varphi(x)$ будет квазиоднородной, причем $p(\varphi) = p(f)$. Кроме того, по построению $\varphi'(x_1) = 0$. Зафиксируем произвольную точку $y \in R^n$ и сформируем функцию $g(t) = t^{-1/p(f)}\varphi(x_1 + t(y - x_1))$. Заметим, что в силу равенства $\varphi'(x_1) = 0$ и следствия 1.6.1 имеем

$$g'(t) = t^{-1/p(f)} \langle \varphi'(x_1 + t(y - x_1)), y - x_1 \rangle -$$
$$- 1/p(f)\, t^{-1/p(f)-1}\varphi(x_1 + t(y - x_1)) < 0.$$

Следовательно, функция $g(t)$ убывает и поэтому при $t < 1$

$$\varphi(y) = g(1) < g(t) = t^{-1/p(f)}\varphi(x_1 + t(y - x_1)).$$

Если же $t > 1$, то

$$\varphi(y) = g(1) > g(t) = t^{-1/p(f)}\varphi(x_1 + t(y - x_1)).$$

Тем самым неравенство (1.6.4) доказано. Доказательство неравенства (1.6.5) аналогично вышеприведенному. Надо лишь ввести функцию $g(t) = t^{-1/(1-p(f))}\varphi(x_1 + t(y - x_1))$ и показать, что $g'(t) > 0$. □

<u>Следствие 1.6.2.</u> Если для некоторой точки $x \in R^n$ и направления $s \neq 0$ выполняется соотношение

$$f(x + s) = f(x) + \langle f'(x), s \rangle,$$

то для любой точки $y \in R^n$ и $t \in R$

$$f(y + ts) = f(y) + t\langle f'(y), s \rangle$$

Доказательство. Действительно, в силу леммы 1.6.3,

$$f(x + ts) = f(x) + t\langle f'(x), s \rangle \qquad (1.6.6)$$

при всех $t > 0$. Кроме того, в силу теоремы 1.2.7 $f'(x + s) = f'(x)$. Поэтому, положив $x' = x + s$, $s' = -s$, получим

$$f(x' + s') = f(x') + \langle f'(x'), s' \rangle.$$

Следовательно, в силу леммы 1.6.3 при всех $t > 0$

$$f(x' + ts') = f(x') + t\langle f'(x'), s' \rangle,$$

т.е.

$$f(x + s - ts) = f(x + s) - t\langle f'(x), s \rangle =$$

$= f(x) + (1 - t) \langle f'(x), s \rangle$.

Таким образом, мы показали, что соотношение (1.6.6) выполнено при всех $t \in R$.

Зафиксируем произвольную точку y из R^n и число τ. Тогда в силу выпуклости функции $f(x)$ при всех $t \in R$ выполняется неравенство

$$f(y + \tau s) + \langle f'(y + \tau s), x + t s - y - \tau s \rangle \le$$
$$\le f(x + t s) = f(x) + t \langle f'(x), s \rangle,$$

т.е.

$$f(y + \tau s) + \langle f'(y + \tau s), x - y - \tau s \rangle -$$
$$- f(x) \le t \langle f'(x) - f'(y + \tau s), s \rangle.$$

В силу произвольности числа t получаем

$$\langle f'(x), s \rangle = \langle f'(y + \tau s), s \rangle.$$

Последнее означает, что функция $\xi(t) = f(y + t s)$ есть линейная функция от t. □

Простейшими примерами квазиоднородных функций являются квадратичная функция ($p(f) = 0.5$) и линейная функция ($p(f) = +\infty$). Можно показать, что среди нелинейных функций у квадратичных степень квазиоднородности наибольшая.

Лемма 1.6.4. У любой нелинейной квазиоднородной функции степень квазиоднородности не превосходит 0.5 .

Доказательство. Если $f(x)$ не есть линейная функция, то в силу неравенства 1.2.5 найдутся точки x, y такие, что $\langle f'(x) - f'(y), x - y \rangle > 0$. Теперь из определения квазиоднородной функции и леммы 1.6.2 получаем

$$p(f) \langle f'(x) - f'(y), x - y \rangle \le f(y) - f(x) -$$
$$- \langle f'(x), y - x \rangle \le (1 - p(f)) \langle f'(x) - f'(y), x - y \rangle.$$

□

Как показывает пример линейной функции, квазиоднородные функции могут и не быть ограниченными снизу. Однако, сделав такое предположение, можно доказать теорему о существовании и "единственности" точки минимума.

Теорема 1.6.1. Если квазиоднородная функция $f(x)$ ограничена снизу, то ее безусловный минимум достигается в точках некоторого множества X^*, которое обладает следующими свойствами:

1) если точки $x_1^*, x_2^* \in X^*$, то и точка $x_1^* + \alpha(x_2^* - x_1^*) \in X^*$ при любом α из R;

2) для любых двух точек x_1^*, x_2^* из X^* и произвольной точки x из R^n справедливы соотношения

$$\langle f'(x), x_1^* - x_2^* \rangle = 0, \tag{1.6.7}$$

$$f(x + x_1^* - x_2^*) = f(x), \tag{1.6.8}$$

$$f'(x + x_1^* - x_2^*) = f'(x) \tag{1.6.9}$$

Доказательство. Заметим, что если множество $X^* \neq \emptyset$, то $f(x^*) = f^*$, $f'(x^*) = 0$ для любой точки $x^* \in X^*$. Поэтому утверждение 1) теоремы непосредственно вытекает из следствия 1.6.2. Далее, для любых точек $x_1^*, x_2^* \in X^*$, произвольной точки x из R^n и любого $t \in R$ имеем

$$\langle f'(x), x - x_1^* + t(x_2^* - x_1^*) \rangle \geq 0.$$

Отсюда в силу произвольности t сразу получаем (1.6.7). Соотношение (1.6.8), в свою очередь, сразу следует из (1.6.7). И наконец, соотношение (1.6.9) вытекает из (1.6.8) и теоремы 1.2.8.

Итак, нам осталось доказать непустоту множества X^*. Если существует хотя бы одна ограниченная минимизирующая последовательность, то множество X^* не пусто. Предположим, что все минимизирующие последовательности неограничены. Рассмотрим произвольную последовательность точек $\{x_k\}_{k=0}^{\infty}$, таких, что $f(x_k) \to f^*$, $\| x_k \| \to \infty$ при $k \to \infty$. Не ограничивая общности, можно считать, что $x_0 = 0$, $\| x_k \| > 1$ и значения функции $f(x)$ в точках x_k монотонно убывают. Сформируем последовательность точек

$y_k = x_k / \| x_k \|$, $k = 1, 2, \ldots$.

Положим $t_k = \| x_k \| > 1$. Заметим, что $x_k = t_k y_k$. Поэтому в силу леммы 1.6.3

$$f(0) > f(x_k) > f(0) + t_k \langle f'(0), y_k \rangle +$$
$$+ t_k^{1/(1-p(f))} [f(y_k) - f(0) - \langle f'(0), y_k \rangle] ,$$

т.е.

$$0 > \langle f'(0), y_k \rangle + t_k^{p(f)/(1-p(f))} [f(y_k) -$$
$$- f(0) - \langle f'(0), y_k \rangle] ,$$

что возможно лишь в том случае, если

$$f(y_k) - f(0) - \langle f'(0), y_k \rangle \to 0$$

при $k \to \infty$. У последовательности точек $\{ y_k \}_{k=1}^{\infty}$ существует по крайней мере одна предельная точка y^*. Причем, как мы показали , $f(y^*) = f(0) + \langle f'(0), y^* \rangle$. В силу следствия 1.6.2 $f(y + t y^*) = f(y) + t \langle f'(y), y^* \rangle$ для любой точки y из R^n и числа $t \in R$. Но функция $f(x)$ ограничена снизу. Поэтому $\langle f'(y), y^* \rangle = 0$ и $f(y + t y^*) = f(y)$. Таким образом, каждая неограниченная минимизирующая последовательность дает нам некоторое направление, вдоль которого функция $f(x)$ не меняется. Обозначим линейную оболочку всех таких направлений через $\mathcal{L}$. Заметим, что для любой точки $x \in R^n$ и направления $s \in \mathcal{L}$ справедливо $\langle f'(x), s \rangle = 0$. Поэтому $f(x + s) = f(x)$ для любого s из $\mathcal{L}$ Последнее означает, что по любой минимизирующей последовательности $\{ x_k \}_{k=0}^{\infty}$ можно сформировать минимизирующую последовательность $\{ x'_k \}_{k=0}^{\infty}$, такую, что $f(x'_k) = f(x_k)$ при всех $k > 0$ и $\langle x'_k , s \rangle = 0$ для любого s из $\mathcal{L}$ (например, $x'_k = \pi(x_k + L , 0))$. Но тогда, повторяя вышеприведенные рассуждения, можно показать, что существует вектор y^* такой, что $f(t y^*) = f(0)$ при всех $t \in R$ и $\langle y^*, s \rangle$ для любого s из $\mathcal{L}$. Последнее противоречит определению подпространства $\mathcal{L}$. $\square$

В заключение приведем два примера квазиоднородных функций.

1. Если функция $f(x) \in C^{1,1}(L)$ и является сильно выпуклой на R^n с константой $m > 0$, то эта функция будет квазиоднородной с $p(f) = 0.5\, m\, L^{-1}$

2. Функция $f(t) = |t|^p$, $t \in R$, $p > 1$, является квазиоднородной с $p(f) = 0.5\, p^{-1}$. Следовательно, и любая функция вида $\sum_{k=1}^{m} | \langle a_k, x \rangle - b_k |^p$ будет квазиоднородной с той же степенью квазиоднородности.

Метод минимизации квазиоднородных функций будет рассмотрен в §3.5.

ГЛАВА 2. МАТЕМАТИЧЕСКИЙ АППАРАТ ДИФФЕРЕНЦИРОВАНИЯ НЕГЛАДКИХ ФУНКЦИЙ

2.1. Проблема дифференцирования негладких функций

Интерес к проблеме "дифференцирования" негладких функций связан, в первую очередь, с развитием нелинейного программирования. Действительно, при разработке теории и методов минимизации негладких функций возникает необходимость в удобном обобщении понятия градиента - таком, которое бы, с одной стороны, обеспечивало сохранение некоторых характерных (с точки зрения теории экстремума) свойств градиента гладкой функции, а с другой - позволяло бы вывести достаточно простые правила вычисления таких "градиентов".

Основные усилия исследователей в этой области были сосредоточены на поиске решения первой части сформулированной проблемы. Причем в качестве образца был выбран класс негладких выпуклых функций. Такой выбор был продиктован тем, что у выпуклой функции есть очень удобная дифференциальная характеристика - ее субдифференциал. С помощью субдифференциала выпуклой функции можно найти (по крайней мере, в принципе) производную функции $f(x)$ по направлению (см. теорему 1.2.7):

$$f'(x_0, s) \equiv \lim_{\alpha \to +0} \alpha^{-1} [f(x_0 + \alpha s) - f(x_0)] =$$

$$= \max \{ \langle u, s \rangle \mid u \in \partial f(x_0) \}, \qquad (2.1.1)$$

проверить необходимое условие минимума

$$0 \in \partial f(x_0), \qquad (2.1.2)$$

найти направление спуска.

Важная роль субдифференциала выпуклой функции побудила многих авторов исследовать классы негладких функций, для которых при вычислении производной по направлению и проверке условий оптимальности можно было бы пользоваться аналогами формул (2.1.1),(2.1.2). Так, Б.Н.Пшеничный предложил рассматривать функции, у которых для каждого x_0 существует множество $\partial_p f(x_0)$, такое, что производную по направлению функции $f(x)$ в точке x_0 можно найти по формуле (2.1.1). Для таких функций были получены содержательные необходимые условия экстремума. Важным подклассом квазидифференцируемых по Пшеничному функций является класс слабо выпуклых функций, исследованный Е.А.Нурминским.

Ф.Кларк ввел понятие верхней производной по направлению

$$f^c(y, s) = \sup\lim_{\substack{x \to y, \\ \alpha \to +0}} \alpha^{-1} [f(x + \alpha s) - f(x)].$$

Несложно показать, что функция $f^c(y, s)$ при всяком фиксированном y является выпуклой однородной (первой степени) функцией от s. Поэтому она субдифференцируема по s в нуле и в силу леммы 1.2.4

$$f^c(y, s) = \max [\langle g, s \rangle \mid g \in \partial_c f(y)],$$

где через $\partial_c f(y)$ обозначено множество $\partial_s f(y, 0)$. Множество $\partial_c f(y)$ называется субдифференциалом Кларка функции $f(x)$ в точке y. Отметим, что значение верхней производной по направлению может не совпадать со значением обычной производной по направлению,даже если последняя существует.

В.Ф.Демьяновым и А.М.Рубиновым было введено понятие квазидифференцируемой функции - это функция, у которой в каждой точке y можно указать два множества $\partial_1 f(y)$ и $\partial_2 f(y)$, такие, что

$$f'(y, s) = \max [\langle g, s \rangle \mid g \in \partial_1 f(y)] +$$
$$+ \min [\langle g, s \rangle \mid g \in \partial_2 f(y)].$$

Класс квазидифференцируемых функций является линейным пространством. Кроме того, для таких функций существуют формулы квазидифференциального исчисления, с помощью которых по последовательности арифметических операций, задающих функцию $f(x)$, можно записать правила формирования множеств $\partial_1 f$ и $\partial_2 f$. Таким образом, квазидифференциальное исчисление дает принципиальную возможность получить для функции, заданной в виде суперпозиции элементарных арифметических операций, множества $\partial_1 f$ и $\partial_2 f$, которые определяют поведение исходной функции в окрестности точки негладкости.

В отличие от работ, посвященных необходимым условиям экстремума, работ по методике вычисления аналогов градиента негладких функций не так много. Понятно, что дифференциальное исчисление для какого-либо класса негладких функций можно построить лишь с помощью соответствующей теоремы о суперпозиции, в которой показывается, как преобразуются дифференциальные характеристики внутренних функций-аргументов под воздействием внешней функции. Так, например, для выпуклых функций хорошо известны формулы пересчета субградиентов для операций, сохраняющих выпуклость (см. теорему 1.2.9). Однако в тех случаях, когда выпуклая функция "собирается" с помощью "невыпуклых" операций, применение этих формул может дать неверный результат.

В отдельных работах делались попытки построить дифференциальное исчисление для субдифференциалов Кларка. Однако в этом направлении удалось получить лишь ослабленный вариант теоремы о суперпозиции - доказывалось, что субдифференциал Кларка суперпозиции функций принадлежит выпуклой оболочке субдифференциалов Кларка внутренних функций, взятых с некоторыми весами.

И, наконец, остановимся на квазидифференциальном исчислении. Эта методика дает нам, с одной стороны, удобное средство полного описания локального поведения негладкой функции. С другой стороны, она позволяет для функций, заданных как конечные суперпозиции элементарных операций, организовать процесс "вычисления" множеств $\partial_1 f$ и $\partial_2 f$ в том или ином виде. Однако формулы квазидифференциального исчисления предполагают сопровождение каждой операции, участвующей в вычислении значения функции, соответствующей операцией над множествами в многомерном пространстве,и поэтому труднореализуемы.

Практически все перечисленные выше способы переноса понятия дифференцируемости на негладкий случай можно трактовать как обобщения свойства субдифференцируемости выпуклых функций. Вследствие этого в качестве аналога градиента в этих определениях всегда выступает целое множество элементов. В настоящей главе рассматривается другой способ введения понятия дифференцируемости для негладкой функции, при котором таким аналогом оказывается вектор. Для однозначного определения такого вектора в точке достаточно зафиксировать произвольный базис в пространстве переменных.

2.2. Класс лексикографически гладких функций

Всюду ниже под словом "функция" понимается отображение одного конечномерного пространства в другое.

Введем следующие обозначения:

D_n - множество локально-липшицевых функций, которые дифференцируемы по любому направлению в R^n;

$U \equiv U(n, m)$ - последовательность точек $U_1, \dots, U_m$ в пространстве R^n;

$A_n[x']$ - оператор линеаризации функции f из D_n в точке $x' \in R^n$: если $\varphi = A_n[x'] f$, то это означает, что $\varphi(s) = f'(x', s)$;

Отметим свойство линейности оператора A_n:

$A_n[x'](f_1 + f_2) \equiv A_n[x'] f_1 + A_n[x'] f_2$,

$A_n[x'] \alpha f_1 \equiv \alpha A_n[x'] f_1$

при любых x' из R^n, α из R, f_1 и f_2 из D_n.

Определение 2.2.1. Функцию $f(x)$ будем называть лексикографически гладкой (l-гладкой) в точке $x_0 \in R^n$, если для любой последовательности точек $U(n, m)$ все члены функциональной последовательности $\{f_k[U]\}$:

$f_0[U] = A_n[x_0] f$,

$f_k[U] = A_n[U_k] f_{k-1}[U]$, $k = 1, \dots, m$,

принадлежат множеству D_n. При этом последовательность функций $\{f_k[U]\}$ будем называть последовательностью линеаризаций, а функцию $f_k[U]$ - k-й линеаризацией функции $f(x)$ в точке x_0 по последовательности направлений $U \equiv U(n, m)$.

Функцию, которая является l-гладкой в каждой точке R^n, будем называть l-гладкой функцией на R^n. Множество таких функций обозначим через $\mathfrak{F}_n$.

Нетрудно видеть, что в силу линейности оператора $A_n[x]$ и линейной структуры множества D_n класс функций $\mathfrak{F}_n$ является линейным пространством.

Перечислим некоторые важные классы функций, содержащиеся в классе $\mathfrak{F}_n$.

1. Всякая дифференцируемая на R^n функция принадлежит

классу $\mathfrak{F}_n$.

Действительно, если функция $f(x)$ дифференцируема в точке x_0, то независимо от выбора последовательности направлений $U(n, m)$ при всех k, $0 < k < m$, справедливо тождество

$$f_k [U] (x) \equiv f'(x_0) x \in D_n ,$$

где $f'(x_0)$ - матрица Якоби функции $f(x)$ в точке x_0.

2. Любая выпуклая на R^n функция принадлежит классу $\mathfrak{F}_n$.

Это утверждение нетрудно вывести из того, что в силу теоремы 1.2.4 линеаризация выпуклой функции есть выпуклая функция, а любая выпуклая на R^n функция в силу теоремы 1.2.3 дифференцируема по направлениям.

3. Любая квазидифференцируемая по Демьянову-Рубинову на R^n функция принадлежит классу $\mathfrak{F}_n$.

Напомним, что по определению линеаризация такой функции фактически есть разность двух выпуклых однородных функций. Сопоставляя этот факт с утверждением 2, нетрудно убедиться в справедливости утверждения 3.

Основной теоремой, из которой будет следовать представительность класса $\mathfrak{F}_n$, является теорема об l - гладкости суперпозиции l -гладких функций. Прежде чем перейти к ее дозательству, сформулируем следующий результат.

<u>Лемма 2.2.1.</u> Пусть функция $f: R^n \to R^m$ принадлежит множеству D_n, а функция $F \in D_m$. Тогда их суперпозиция - функция $\varphi(x) = F(f(x))$ - также принадлежит множеству D_n, причем при любом $y \in R^n$ справедлива формула

$$A_n[y] \varphi(x) \equiv A_m[f(y)] F (A_n[y] f (x)). \quad (2.2.1)$$

Действительно, в силу леммы 1.2.5 функция $\varphi(x)$ дифференцируема по направлениям и

$$\varphi'(y , x) = F'(f(y) , f'(y , x)) .$$

Осталось заметить, что формула (2.2.1) есть просто перезапись последней формулы в других обозначениях. □

Теорема 2.2.1. Если функция $f\colon R^n \to R^m$ является l-гладкой в точке x', а функция $F(z)$, $z \in R^m$, является l-гладкой в точке $z' = f(x')$, то их суперпозиция - функция $\varphi(x) = F(f(x))$ - также будет l-гладкой в точке x', причем для линеаризаций этих функций в точке x' будет справедлива формула

$$\varphi_k [U](x) \equiv F_k [V] (f_k [U](x)), \tag{2.2.2}$$

$k = 0, \dots, p$,

где $U \equiv U(n, p)$ - произвольная последовательность направлений в R^n, последовательность направлений $V \equiv V(m, p) \equiv \{V_k\}$ такова, что $V_k = f_{k-1}[U](U_k)$, $k = 1, \dots, p$, а функции $\varphi_k[U]$, $F_k[V]$, $f_k[U]$ - это k-е линеаризации функций $\varphi(x)$, $F(z)$, $f(x)$ в точках x', z', x' по последовательностям $U(n, p)$, $V(m, p)$, $U(n, p)$ соответственно.

Доказательство. Заметим, что в силу леммы 2.2.1 соотношение (2.2.2) выполнено при $k = 0$. Доказательство будем вести по индукции.

Пусть формула (2.2.2) справедлива при некотором $k > 0$. Тогда из l-гладкости функций F, f и утверждения леммы (1.2.5) следует, что функция $\varphi_k[U](x)$ дифференцируема в каждой точке пространства R^n Поэтому в силу индуктивного предположения и леммы 2.2.1 имеем

$$\varphi_{k+1}[U](x) = A_n[U_{k+1}]\,\varphi_k[U](x) =$$
$$= A_m[f_k[U](U_{k+1})]\,F_k[V](A_n[U_{k+1}]\,f_k[U](x)) =$$
$$= F_{k+1}[V](f_{k+1}[U](x)). \quad \square$$

2.3. Лексикографические производные и их свойства

Приведем две леммы, описывающие основные свойства оператора линеаризации и последовательности линеаризаций функции $f(x)$.

<u>Лемма 2.3.1.</u> Пусть функция $f \in D_n$. Тогда

1) при любых y из R^n справедлтво тождество
$A_n^2[y]f \equiv A_n[y]f$;

2) при любых y из R^n функция $A_n[y]f$ является положительно однородной:
$A_n[y]f(\xi x) = \xi A_n[y]f(x)$
при всех x из R^n, $\xi > 0$;

3) если функция f является положительно однородной, то при любых $y \in R^n$, $\xi > 0$ справедливо тождество
$A_n[\xi y]f \equiv A_n[y]f$;

4) если функция f является положительно однородной, то при всех $x, y \in R^n$ и произвольном ξ из R справедлива формула

$$A_n[y]f(x + \xi y) = A_n[y]f(x) + \xi f(y). \tag{2.3.1}$$

Д о к а з а т е л ь с т в о. Утверждения 1) - 3) этой леммы проверяются непосредственно с помощью определения оператора A_n. Докажем справедливость формулы (2.3.1).

$$A_n[y]f(x) = \lim_{\alpha \to +0} \alpha^{-1}[f(y + \alpha x) - f(y)] =$$

$$= \lim_{\alpha \to +0} \alpha^{-1}[f((1 - \alpha\xi)y + \alpha(x + \xi y)) - f(y)] =$$

$$= \lim_{\alpha \to +0} \alpha^{-1}(1 - \alpha\xi)[f(y + \alpha^{-1}(1 - \alpha\xi)^{-1}(x + \xi y)) -$$

$$- f(y)] - \xi f(y) = A_n[y]f(x + \xi y) - \xi f(y).$$

□

Обозначим через $L_k(U)$, $k = 1, \dots, n$, линейное подпространство, натянутое на первые k векторов последовательности $U \equiv U(n, n)$. Положим $L_0(U) \equiv \{0\}$.

Лемма 2.3.2: если функция f является l-гладкой в точке x_0, то для ее линеаризаций справедливы следующие равенства:

1) $f_k[U](\xi x) = \xi f_k[U](x)$

при всех x из R^n, $\xi > 0$;

2) $f_k[U](x + \xi y) = f_k[U](x) + \xi f_k[U](y)$

при любых x из R^n, y из $L_k(U)$ и ξ из R;

3) $f_k[U](x) \equiv f_{k-1}[U](x)$ при всех x из $L_{k-1}(U)$.

Доказательство. Утверждение 1) леммы 2.3.2 сразу следует из утверждения 1) леммы 2.3.1. Доказательство утверждения 2) леммы 2.3.2 будем проводить по индукции.

При $k = 1$ множество $L_1(U) \equiv \{y \mid y = \beta U_1, \beta \in R\}$. Поэтому в силу (2.3.1) имеем

$$f_1[U](x + \xi y) = f_1[U](x + \xi \beta U_1) =$$
$$= A_n[U_1] f_0[U](x + \xi \beta U_1) =$$
$$= A_n[U_1] f_0[U](x) + \xi \beta f_0[U](U_1) =$$
$$= f_1[U](x) + \xi A_n[U_1] f_0[U](\beta U_1) \equiv$$
$$\equiv f_1[U](x) + \xi f_1[U](y).$$

Пусть теперь утверждение 2) выполнено при некотором $k > 0$. Выберем произвольные x из R^n и y из $L_{k+1}(U)$. Тогда $y = z + \beta U_{k+1}$, где $z \in L_k(U)$. Далее, пользуясь формулой (2.3.1), линейностью оператора A_n и индуктивным предположением, получаем

$$f_{k+1}[U](x + \xi y) = f_{k+1}[U]((x + \xi z) +$$
$$+ \xi \beta U_{k+1}) = A_n[U_{k+1}] f_k[U]((x + \xi z) +$$
$$+ \xi \beta U_{k+1}) = A_n[U_{k+1}] f_k[U](x + \xi z) +$$
$$+ \xi \beta f_k[U](U_{k+1}) = A_n[U_{k+1}] f_k[U](x) +$$
$$+ \xi A_n[U_{k+1}] f_k[U](z) + \xi \beta f_k[U](U_{k+1}) =$$
$$= f_{k+1}[U](x) + \xi (A_n[U_{k+1}] f_k[U](z) +$$

$+ \beta f_k [U] (U_{k+1})) = f_{k+1} [U](x) +$

$+ \xi A_n [U_{k+1}] f_k [U](y) =$

$= f_{k+1} [U](x) + \xi f_{k+1} [U](y).$

Таким образом, утверждение 2) доказано.

Наконец, в силу утверждения 2) леммы 2.3.2 при произвольном x из $L_k(U)$ имеем

$$f_{k+1} [U](x) = \lim_{\alpha \to +0} \alpha^{-1} [f_k [U](U_{k+1} + \alpha x) -$$

$$- f_k [U](U_{k+1})] = \lim_{\alpha \to +0} \alpha^{-1} [\alpha f_k [U](x)] =$$

$= f_k [U](x).$ □

Следствие 2.3.1. Если $x = \sum_{i=1}^{k} \xi_i U_i$, то справедлива формула

$$f_k [U](x) = \sum_{i=1}^{k} \xi_i f_k [U](U_i). \qquad (2.3.2)$$

Формула (2.3.2) является непосредственным следствием утверждений 2),3) леммы 2.3.2. □

Теперь мы можем дать определения дифференциальных характеристик лексикографически гладкой функции $f(x)$.

Определение 2.3.1. Лексикографической производной k-го сужения функции $f(x)$ (l-k -производная) по последовательности U в точке x_0 будем называть матрицу Якоби любой линейной функции $c(x)$, $x \in R^n$, такой, что $c(x) \equiv \equiv f_k [U](x)$ при всех x из $L_k(U)$. Для любой такой матрицы будем использовать обозначение $g_k(f, U, x_0)$. В тех случаях, когда речь идет о функции $f\colon R^n \to R$, вместо слова "производная" будет использоваться слово "градиент".

Ясно, что в силу следствия 2.3.1 определение 2.3.1 корректно. Заметим, что матрица $g_k(f, U, x_0)$, вообще говоря, определена неоднозначно. Однако в одном важном частном случае при некоторых k можно гарантировать единственность

l-k - производной.

Теорема 2.3.1. Если функция f является l - гладкой, то для любой последовательности точек $U(n, m)$, такой, что $L_m(U) \equiv R^n$, найдется номер k_0, $1 \leqslant k_0 \leqslant m$, начиная с которого функция $f_k[U]$ становится линейной. Причем при всех i, $k_0 \leqslant i \leqslant m$, l-i -производные i-ых сужений функции $f(x)$ определены однозначно и совпадают друг с другом.

Доказательство. Из условия теоремы следует существование номера k такого, что $L_k(U) \equiv R^n$. Но тогда из формулы (2.3.2) вытекает, что функция $f_k[U]$ является линейной на всем пространстве R^n. Осталось заметить, что результат линеаризации линейной функции совпадает с исходной функцией. □

Обозначим через $\mathcal{K}_n$ множество линейных на R^n функций.

Определение 2.3.2. Наименьший из номеров k, таких, что $f_k[U] \in \mathcal{K}_n$, будем называть дефектом функции $f(x\cdot)$ по последовательности U в точке x_0:

$$d(f, U, x_0) = \min\{k \mid f_k[U] \in \mathcal{K}_n\}.$$

Если функции $f_k[U]$ не являются линейными ни при каких k, $1 \leqslant k \leqslant m$, то положим $d(f, U, x_0) = \infty$.

Замечания. 1. Для дифференцируемой в точке x_0 функции дефект $d(f, U, x_0)$ равен нулю независимо от способа выбора последовательности U.

2. Если U - базис в пространстве R^n, а функция $f(x)$ l-гладкая в точке x_0, то $d(f, U, x_0) \leqslant n$ (это следует из теоремы 2.3.1).

Определение 2.3.3. Матрицу Якоби линейной функции $f_k[U](x)$, где $k \geqslant d(f, U, x_0)$, будем называть лексикографической производной (l-производной) функции $f(x)$ по последовательности U в точке x_0.

Для матрицы l-производной будем использовать обозначение $g(f, U, x_0)$.

В дальнейшем будут использоваться в основном l-производные функции $f(x)$. Введенные ранее l-k-производные потребуются лишь при выводе формул лексикографического дифференциального исчисления.

Лексикографические производные функции $f(x)$, вычисленные в точке x_0 по различным последовательностям U и V, вообще говоря, отличаются друг от друга. Случаи их заведомого совпадения отражены в следующей лемме.

Лемма 2.3.3. Пусть $d(f, U, x_0), d(f, V, x_0) < \infty$. Равенство $g(f, U, x_0) = g(f, V, x_0)$ заведомо выполняется, если:

1) $U_i = V_i$ для всех $i < \min\{ d(f, U, x_0), d(f, V, x_0) \}$;
2) $Ort(U) \equiv Ort(V)$, где $Ort(U)$ - система ортонормированных векторов в пространстве R^n, полученная в результате применения к последовательности U процесса ортогонализации Грамма-Шмидта с нормировкой и удалением промежуточных нулей.

Доказательство. Утверждение 1) леммы 2.3.3 является очевидным следствием определений.

Для доказательства утверждения 2) достаточно будет показать, что

$$g(f, U, x_0) \equiv g(f, Ort(U), x_0) \tag{2.3.3}$$

Заметим, что если $U_{k+1} \in L_k(U)$, то в силу утверждения 2) леммы 2.3.2

$$f_{k+1}[U](x) = \lim_{\alpha \to +0} \alpha^{-1} (f_k[U](U_{k+1} + \alpha x) -$$

$$- f_k[U](U_{k+1})) = \lim_{\alpha \to +0} \alpha^{-1} (\alpha f_k[U](x)) =$$

$= f_k [U] (x)$.

Таким образом, без ограничения общности можно считать, что U - система ненулевых линейно независимых векторов. Для завершения доказательства осталось учесть два обстоятельства

1. Умножение векторов последовательности U на произвольные положительные числа не изменит матрицы $g(f, U, x_0)$ (см. утверждение 3) леммы 2.3.1).

2. Для произвольного y из $L_k (U)$

$$A_n [U_{k+1} + y] f_k [U](x) =$$
$$= \lim_{\alpha \to +0} \alpha^{-1} (f_k [U](U_{k+1} + y + \alpha x) -$$
$$- f_k[U](U_{k+1} + y)) = \lim_{\alpha \to +0} \alpha^{-1} (f_k [U](U_{k+1} + \alpha x) -$$
$$- f_k [U] (U_{k+1})) = f_{k+1} [U] (x) .$$

Следовательно, процесс ортогонализации Грамма-Шмидта, примененный к последовательности U , не изменит $g(f, U, x_0)$. Формула (2.3.3), а значит, и лемма 2.3.3 доказаны. □

Таким образом, множество лексикографических производных функции $f(x)$ в точке x_0 полностью исчерпывается l-производными по всевозможным ортонормированным базисам. В дальнейшем, как правило, мы будем рассматривать только лексикографические производные по базисам пространства R^n, закрепив за такими базисами обозначение S.

Приведем три леммы, отражающие некоторые свойства l -градиентов лексикографически гладких функций.

<u>Лемма 2.3.4.</u> Пусть $d(f, U, x_0) < \infty$. Тогда

$$\lim_{\alpha \to +0} \alpha^{-1} [f(x_0 + \alpha U_1) - f(x_0)] = \langle g(f, U, x_0) , U_1 \rangle,$$

$$\lim_{\alpha \to +0} \alpha^{-1} [f_k [U](\alpha U_{k+1}) - f_k [U] (0)] =$$

$= \langle g(f, U, x_0), U_{k+1} \rangle, \quad k = 0, \dots, d(f, U, x_0)$.

Доказательство. Пусть $m = d(f, U, x_0)$. Тогда в силу формулы (2.3.1), утверждения 3) леммы 2.3.2 и определения l-градиента имеем

$$\lim_{\alpha \to +0} \alpha^{-1} [f(x_0 + \alpha U_1) - f(x_0)] = f_0 [U] (U_1) =$$

$$= f_1 [U] (U_1) = \dots = f_m [U] (U_1) =$$

$$= \langle g(f, U, x_0), U_1 \rangle .$$

Аналогично

$$\lim_{\alpha \to +0} \alpha^{-1} [f_k [U](\alpha U_{k+1}) - f_k [U] (0)] =$$

$$= f_k [U] (U_{k+1}) = f_{k+1} [U] (U_{k+1}) = \dots =$$

$$= f_m [U] (U_{k+1}) = \langle g(f, U, x_0), U_{k+1} \rangle . \quad \square$$

Лемма 2.3.5. Пусть $f(x)$ - выпуклая на R^n функция, S - базис в R^n. Определим последовательность множеств $\{ B_k \}$ следующим образом:

$B_0 = \partial f(x_0)$,

$B_k = \mathrm{Argmax} \{ \langle S_k, z \rangle \mid z \in B_{k-1} \}, \quad k = 1, \dots, n$.

Тогда вектор $g(f, S, x_0)$ является единственной точкой множества B_n.

Доказательство. Напомним, что для однородной функции $f(x)$ тождество

$f(u) \equiv \max \{ \langle u, z \rangle \mid z \in Q \}$

выполнено тогда и только тогда, когда $\partial f(0) \equiv \mathrm{Conv}\, Q$.

Покажем по индукции, что

$$\partial f_k [S] (0) \equiv B_k, \quad k = 0, \dots, n. \qquad (2.3.4)$$

Действительно, при $k = 0$

$$f_0 [S] (x) = \lim_{\alpha \to +0} \alpha^{-1} (f(x_0 + \alpha x) - f(x_0)) =$$

$$= \max \{\langle x, z \rangle \mid z \in \partial f(x_0)\} = \max \{\langle x, z \rangle \mid z \in B_0 \},$$

т. е. $\partial f_0 [S] (0) \equiv B_0$.

Далее, пусть для некоторого k справедлива формула (2.3.4). Тогда

$$f_{k+1}[S](x) =$$
$$= \lim_{\alpha\to+0} \alpha^{-1}(f_k [S] (S_{k+1} + \alpha x) - f_k [S] (S_{k+1})) =$$
$$= \max \{ \langle x , z \rangle \mid z \in \partial f_k [S] (S_{k+1}) \} .$$

Но функция $f_k [S]$ является выпуклой и положительно однородной. Поэтому

$$\partial f_k [S] (S_{k+1}) =$$
$$= \operatorname{Argmax} \{ \langle S_{k+1} , z \rangle \mid z \in \partial f_k [S] (0) \} =$$
$$= \operatorname{Argmax} \{ \langle S_{k+1} , z \rangle \mid z \in B_k \} = B_{k+1} .$$

Таким образом, формула (2.3.4) доказана.

По теореме 2.3.1 найдется номер k, $0 \leqslant k \leqslant n$, при котором функция $f_k [S]$ станет линейной. А это значит, что в силу (2.3.5) множество B_k будет состоять из единственной точки, которая и является l-градиентом функции $f(x)$ по базису S в точке x_0. □

<u>Лемма 2.3.6.</u> Пусть лексикографически гладкая функция $f(x)$ является квазивыпуклой на всем пространстве R^n. Тогда для любой точки x_0 и произвольного базиса S в R^n справедливо включение

$$\{ y \mid f(y) < f(x_0) \} \subseteq$$
$$\subseteq \{ y \mid \langle g(f, S, x_0) , x_0 - y \rangle > 0 \} ,$$

т. е. ненулевой лексикографический градиент квазивыпуклой функции, вычисленный по произвольному базису пространства R^n, является опорным функционалом к ее линии уровня.

Д о к а з а т е л ь с т в о. Рассмотрим последовательность линеаризаций $\{ f_k [S] \}$ функции $f(x)$ в точке x_0 по базису S. Покажем по индукции, что все функции $f_k [S]$ являются квазивыпуклыми.

Действительно, для любых x, y из R^n и ξ, $0 \le \xi \le 1$,

$$\xi f_0 [S] (x) + (1 - \xi) f_0 [S] (y) =$$
$$= \xi \lim_{\alpha \to +0} \alpha^{-1} [f(x_0 + \alpha x) - f(x_0)] +$$
$$+ (1 - \xi) \lim \alpha^{-1} [f(x_0 + \alpha y) - f(x_0)] =$$
$$= \lim_{\alpha \to +0} \alpha^{-1} [\xi f(x_0 + \alpha x) + (1 - \xi) f(x_0 + \alpha y) -$$
$$- f(x_0)] \le \lim_{\alpha \to +0} \alpha^{-1}[\max \{ f(x_0 + \alpha x), f(x_0 + \alpha y) \} -$$
$$- f(x_0)] = \lim_{\alpha \to +0} \alpha^{-1} [\max \{ f(x_0 + \alpha x) - f(x_0),$$
$$f(x_0 + \alpha y) - f(x_0) \}] = \max \{f_0 [S](x), f_0 [S](y)\}.$$

Далее, предположив, что функция $f_k [S] (x)$ является квазивыпуклой при некотором $k > 0$ и повторив вышеприведенные рассуждения, получим, что функция $f_{k+1} [S] (x)$ также является квазивыпуклой. Таким образом, мы показали, что все функции $\{ f_k [S] \}$ являются квазивыпуклыми.

Покажем теперь, что при всех k, $0 \le k \le n-1$, справедливы включения

$$\{ y \mid f(y) \le f(x_0) \} \subseteq \{ x_0 \} + \{ z \mid f_0 [S](z) \le 0 \}; \tag{2.3.5}$$

$$\{ z \mid f_k [S] (z) \le 0 \} \subseteq \{ z \mid f_{k+1} [S] (z) \le 0 \} . \tag{2.3.6}$$

Действительно, в силу квазивыпуклости функции $f(x)$ из неравенства $f(y) \le f(x)$ вытекает, что

$$f_0[S](y - x_0) = \lim_{\alpha \to +0} \alpha^{-1} [f(x_0 + \alpha(y - x_0)) - f(x_0)] \le$$
$$\le \lim_{\alpha \to +0} \alpha^{-1}[\max \{ f(y) , f(x_0) \} - f(x_0)] = 0 ,$$

а значит, справедливо включение (2.3.5).

Доказательство включения (2.3.6) проводится аналогично по индукции с учетом квазивыпуклости функций $f_k[\,S\,](x)$.

Для завершения доказательства леммы осталось заметить, что в силу включений (2.3.5),(2.3.6) и теоремы 2.3.1

$$\{ y \mid f(y) < f(x_0) \} \subseteq \{ x_0 \} + \{ z \mid f_n[\,S\,](z) < 0 \} \equiv$$
$$\equiv \{ x_0 \} + \{ z \mid \langle g(f, S, x_0), z \rangle < 0 \} \equiv$$
$$\equiv \{ y \mid \langle g(f, S, x_0), x_0 - y \rangle > 0 \}. \quad \square$$

2.4. Основная теорема лексикографического дифференциального исчисления

Одно из основных преимуществ лексикографических дифференциальных характеристик заключается в том, что для них существуют правила вычисления, по своей простоте приближающиеся к правилам вычисления дифференциальных характеристик гладких функций. Эти правила можно вывести из следующей теоремы.

Теорема 2.4.1. Пусть функция $f : R^n \to R^m$ является l-гладкой в точке x_0, а функция $F(z)$, $z \in R^m$, является l-гладкой в точке $z_0 = f(x_0)$. Тогда функция $\varphi(x) = F(f(x))$ является лексикографически гладкой в точке x_0 и ее l-производная по произвольному базису S вычисляется по формуле

$$g(\varphi, S, x_0) = g_n(F, U, z_0)\, g(f, S, x_0), \quad (2.4.1)$$

где точки последовательности $U = U(m, n)$ задаются соотношениями

$$U_k = g(f, S, x_0)\, S_k, \; k = 1, \dots, n. \quad (2.4.2)$$

Доказательство. По теореме 2.2.1 функция

$\varphi(x)$ является лексикографически гладкой и ее лексикографической производной будет матрица Якоби функции

$$\varphi_n[S](x) = F_n[U](f_n[S](x)), \qquad (2.4.3)$$

где последовательность U задается формулами

$$U = U(m, n): \quad U_k = f_{k-1}[S](S_k), \; k = 1, \ldots, n. \qquad (2.4.4)$$

С другой стороны, из теоремы 2.3.1 и из определения лексикографического градиента следует что

$$f_n[S](x) = g(f, S, x_0)\, x.$$

Кроме того, из утверждения 3) леммы 2.3.2 следует, что

$$f_{k-1}[S](S_k) = f_n[S](S_k), \quad k = 1, \ldots, n.$$

Поэтому формулы (2.4.3),(2.4.4) можно заменить на следующие:

$$\varphi_n[S](x) = F_n[U](g(f, S, x_0)\, x), \qquad (2.4.5)$$

где $U = U(m, n)$ такова, что $U_k = g(f, S, x_0)\, S_k$. Таким образом, формула (2.4.2) получена. Осталось обосновать формулу (2.4.1).

Система векторов S является базисом в R^n. Поэтому из формулы (2.4.5) следует, что при любом x из R^n справедливо включение

$$g(f, S, x_0) \times x \in L_n(U)$$

А это значит, что в силу определения l-n производной функции F по последовательности U

$$F_n[U](g(f, S, x_0)\, x) =$$
$$= g_n(F, U, z_0)\, g(f, S, x_0)\, x.$$

Осталось подставить полученное выражение в формулу (2.4.3) и вычислить матрицу Якоби правой части. □

<u>Следствие 2.4.1.</u> Пусть функция $f: R^n \to R^m$ является l-гладкой в точке x_0, а функция $F(z): R^m \to R$, является выпуклой на R^m. Тогда функция $\varphi(x) = F(f(x))$ является лексикографически гладкой в точке x_0 и ее l-производная по произвольному базису S вычисляется по формуле

$g(\varphi, S, x_0) = g^T g(f, S, x_0)$,

где вектор g - это произвольная точка последнего из множеств $\{B_k\}$:

$B_0 = \partial F(x_0)$,

$B_k = \mathrm{Argmax}\, \{ \langle g(f, S, x_0) S_k , z \rangle \mid z \in B_{k-1} \}$,

$k = 1, \ldots , n$.

Доказательство этого следствия сводится к обоснованию формул, характеризующих l-n-градиент выпуклой функции. Мы не будем его приводить здесь, так как оно почти дословно совпадает с доказательством леммы 2.3.5. □

Таким образом, в силу теоремы 2.3.1 для вычисления лексикографического градиента функции $\varphi(x) = F(f(x))$ из "предыстории" достаточно иметь соответствующую l-производную внутренней функции f . Именно это обстоятельство позволяет вывести простые правила пересчета лексикографических градиентов в тех случаях, когда внешняя функция F принадлежит некоторому достаточно широкому набору базовых элементарных функций. В свою очередь, это означает, что имеется возможность вычислять лексикографические градиенты для функций, являющихся конечными суперпозициями элементарных базовых функций.

Приведем правила вычисления l-градиентов для некоторых типичных ситуаций.

Будем говорить, что вектор x больше вектора y ($x > y$), если найдется номер i такой, что $x^{(j)} = y^{(j)}$ для всех $j < i$, а $x^{(i)} > y^{(i)}$ Всюду далее символ S будет использоваться как для обозначения базиса в R^n, так и для обозначения $n \times n$ - матрицы, столбцами которой являются векторы $S_1, \ldots , S_n$.

Итак, пусть функция $f(x) : R^n \to R^m$ является l-гладкой в точке x_0. Рассмотрим функцию $\varphi(x) = F(f(x))$. Приведем правила вычисления лексикографического градиента функции $\varphi(x)$, конкретизировав вид функции $F(z)$.

1. $F(z)$ - дифференцируемая функция. Тогда

$g(\varphi, S, x_0) = F'(f(x_0))\, g(F, S, x_0)$.

2. $m = 1$, $F(z) = |z|$. Образуем $(n + 1)$ - мерный вектор y:

$y = (f(x_0), g(F, S, x_0) S)$.

Тогда

$$g(\varphi, S, x_0) = \begin{cases} g(f, S, x_0), & \text{если } y > 0, \\ -g(f, S, x_0), & \text{если } 0 > y, \\ 0, & \text{если } y \equiv 0. \end{cases}$$

Обоснование этой и всех последующих формул представляет собой элементарное упражнение по использованию следствия из теоремы 2.4.1.

3. $m = 2$, $F(z) = \max(z^{(1)}, z^{(2)})$. Образуем два $(n + 1)$ - мерных вектора y_1 и y_2:

$y_1 = (f_1(x_0), g(f_1, S, x_0) S)$,

$y_2 = (f_2(x_0), g(f_2, S, x_0) S)$.

Тогда

$$g(\varphi, S, x_0) = \begin{cases} g(f_1, S, x_0), & \text{если } y_1 > y_2, \\ g(f_2, S, x_0), & \text{если } y_2 > y_1, \\ \text{любой из них, если } y_1 \equiv y_2. \end{cases}$$

Случай, когда у функции максимума число аргументов больше двух, рассматривается аналогично.

4. $m = 2$, $F(z) = \min(z^{(1)}, z^{(2)})$. Образуем два $(n + 1)$ - мерных вектора y_1 и y_2:

$y_1 = (f_1(x_0), g(f_1, S, x_0) S)$,

$y_2 = (f_2 (x_0) , g(f_2, S, x_0) S)$.

Тогда

$$g(\varphi , S, x_0) = \begin{cases} g(f_1, S, x_0) , & \text{если} \quad y_2 > y_1 , \\ g(f_2, S, x_0) , & \text{если} \quad y_1 > y_2 , \\ \text{любой из них, если} & y_1 \equiv y_2 . \end{cases}$$

Случай, когда у функции минимума число аргументов больше двух,рассматривается аналогично.

5. $m = 2$, $F(z) = [(z^{(1)})^2 + (z^{(2)})^2]^{1/2}$. Образуем $(n + 1)$ - мерный вектор y с координатами

$y^{(1)} = \varphi(x_0)$,

$y^{(i+1)} = [(g(f_1, S, x_0)S_i)^2 + (g(f_2, S, x_0)S_i)^2]^{1/2}$,

$k = 1, \ldots , n$.

Тогда, если $y \equiv 0$, то $g(\varphi , S, x_0) = 0$. Если же $y \neq$ $\neq 0$, то

$g(\varphi , S, x_0) = \alpha\, g(f_1, S, x_0) + \beta\, g(f_2, S, x_0)$,

где

$\alpha = (g(f_1, S, x_0) S_k) / y_k$;

$\beta = (g(f_2, S, x_0) S_k) / y_k$;

k - номер первой отличной от нуля координаты вектора y.

Аналогичные формулы нетрудно получить и для l_p - норм с $p > 1$ и с $m > 2$.

Отметим, что специфика правил 2-5 оправдывает до некоторой степени выбор термина "лексикографическое дифференцирование".

В следующем параграфе мы покажем, как можно вычислять лексикографические производные более сложных функций.

2.5. Примеры лексикографических производных

В настоящем разделе будут получены формулы для вычисления лексикографических производных некоторых важных функций. Формулы для l-i-производных этих функций, необходимые для пересчета лексикографических производных с помощью теоремы 2.4.1, не выписываются, так как они могут быть получены аналогичным способом.

2.5.1. Производная оператора проектирования на выпуклое множество

Пусть Q - выпуклое замкнутое множество в R^n, x_0 - точка из множества Q. Обозначим через $\mathcal{L}(Q, x_0)$ касательный конус ко множеству Q в точке x_0:

$$\mathcal{L}(Q, x_0) = \overline{\{ s \mid s = \alpha (y - x_0) ,\ \alpha > 0 ;\ y \in Q \}} .$$

Напомним, что точка $p = \pi(K, x_0)$ из выпуклого замкнутого конуса K является проекцией точки x_0 на этот конус тогда и только тогда, когда при всех s из K справедливо неравенство

$$\langle x_0 - p , p - s \rangle \geq 0 .$$

Отображение $\pi(Q , x_0)$ является функцией из R^n в R^n. Ниже мы покажем, что эта функция является лексикографически гладкой и выведем формулу для ее l-производной.

<u>Теорема 2.5.1.</u> Функция $p(x) \equiv \pi(Q , x)$ является дифференцируемой по всем направлениям в каждой точке пространства R^n. При этом оператор линеаризации $A_n [x_0]$ действует на функцию $p(x)$ следующим образом:

$$A_n [x_0] p (x) = \pi(Q_1 , x), \tag{2.5.1}$$

где $Q_1 = \{ s \in \mathcal{L}(Q, p(x_0)) \mid \langle x_0 - p(x_0), s \rangle = 0 \}$.

Доказательство. Возможны три случая.

1. Пусть $x_0 \in \mathrm{int}(Q)$. Тогда $x_0 = p(x_0)$, и поэтому $Q_1 = \mathcal{L}(Q, x_0) \equiv R^n$.

Следовательно, в этом случае соотношение (2.5.1) эквивалентно тривиальному равенству $A_n[x_0]\, p(x) \equiv x$, которое с очевидностью выполняется.

2. Пусть $x_0 \in \partial Q$. Тогда $x_0 = p(x_0)$, и поэтому $Q_1 = \mathcal{L}(Q, x_0)$. Обоснуем этот вариант формулы (2.5.1). Заметим, что в силу следствия 1.1.2 функция $p(x)$ липшицева на R^n с константой единица. Кроме того, в силу теоремы 1.1.1 при всех x из R^n, y из Q и $\alpha > 0$ справедливо неравенство

$$\langle x_0 + \alpha x - p(x_0 + \alpha x), p(x_0 + \alpha x) - y \rangle \geq 0. \tag{2.5.2}$$

Зафиксируем произвольный $x \neq 0$. Обозначим

$$w(\alpha) = \alpha^{-1} [p(x_0 + \alpha x) - p(x_0)], \tag{2.5.3}$$

где $0 < \alpha < 1$. Ясно, что все $w(\alpha)$ принадлежат некоторому шару - компактному множеству. Поэтому множество предельных точек функции $w(\alpha)$ при $\alpha \to 0$ не пусто. Обозначим любую из таких точек через w^*. Подставим в (2.5.2) $y = x_0 + \alpha(z - x_0)$, где z - точка из Q. Теперь, разделив (2.5.2) на α^2 и перейдя к пределу при $\alpha \to +0$, имеем

$$\langle x - w^*, w^* + x_0 - z \rangle \geq 0$$

при всех z из Q. А это значит, что

$$\langle x - w^*, w^* - s \rangle \geq 0 \tag{2.5.4}$$

при всех s из $\mathcal{L}(Q, x_0)$. Кроме того, в силу (2.5.3) и определения касательного конуса $w^* \in \mathcal{L}(Q, x_0)$. Учитывая этот факт в совокупности с (2.5.4), получаем, что точка

w^* является проекцией точки x на выпуклый конус $\mathcal{L}(Q, x_0)$, а такая точка в силу теоремы 1.1.1 всегда единственна.

Таким образом, мы доказали, что в рассматриваемом случае функция $p(x)$ дифференцируема по направлениям и что $A_n[x_0]p(x) = w^* \equiv \pi(\mathcal{L}(Q, x_0), x)$.

3. Пусть $x_0 \notin Q$. Как и в случае 2, зафиксируем некоторое w^* - любую предельную точку траектории $w(\alpha)$. Из (2.5.2) получаем

$$\langle x_0 - p(x_0), w^* \rangle \geq 0. \qquad (2.5.5)$$

Кроме того,

$$\langle x_0 - p(x_0), p(x_0) - p(x_0 + \alpha x) \rangle \geq 0.$$

А это в совокупности с (2.5.5) означает, что

$$\langle x_0 - p(x_0), w^* \rangle = 0. \qquad (2.5.6)$$

Напомним также, что при любых x_0 вектор w^* принадлежит конусу $\mathcal{L}(Q, p(x_0))$, т. е. вектор w^* всегда принадлежит множеству Q_1.

Если Q_1 состоит из единственной точки - нуля, то дифференцируемость по направлениям функции $p(x)$ и формула (2.5.1) доказаны.

Предположим, что Q_1 состоит не только из нуля. Пусть, кроме того, найдется вектор s из Q_1 такой, что

$$\| x - s \| < \| x - w^* \|.$$

Это значит, что при всех $\xi \in R$

$$\| x + \xi(x_0 - p(x_0)) - s \| <$$
$$< \| x + \xi(x_0 - p(x_0)) - w^* \|. \qquad (2.5.7)$$

Рассмотрим множество точек

$$y(\beta) = \pi(Q, p(x_0) + \beta s), \ \beta > 0.$$

По доказанному в пункте 2 функция $p(x)$ является дифференцируемой по направлениям в точке $p(x_0)$, причем

$$A_n[p(x_0)]p(s) = \pi(\mathcal{L}(Q, p(x_0)), s) = s.$$

Следовательно,

$y(\beta) - p(x_0) = \beta s + o(\beta)$.

Подставив это выражение в (2.5.7) с $\xi = \beta^{-1}$, получим

$\| \beta^{-1} (x_0 + \beta x - p(x_0) - \beta w^*) \| >$

$> \| \beta^{-1} (x_0 + \beta x - y(\beta)) - \beta^{-1} o(\beta) \|$

Таким образом, при достаточно малых β

$$\| x_0 + \beta x - p(x_0) - \beta w^* \| >$$

$$> \| x_0 + \beta x - y(\beta) \| \qquad (2.5.8)$$

С другой стороны, найдется последовательность чисел β_k такая, что

$p(x_0) + \beta_k w^* = p(x_0 + \beta_k x) + o(\beta_k)$.

Но тогда в силу (2.5.8) при достаточно малых β_k

$\| x_0 + \beta_k x - p(x_0 + \beta_k x) \| > \| x_0 + \beta_k - y(\beta_k) \|$,

что противоречит определению функции $p(\ )$. □

<u>Следствие 2.5.1.</u> Функция $p(x) \equiv \pi(Q , x)$ является лексикографически гладкой на R^n.

Действительно, из теоремы 2.5.1 следует, что линеаризация функции $p(x)$ есть функция проекции на выпуклое множество Q_1. А эта функция дифференцируема по любому направлению, что и требуется в определении l-гладкой функции.

С помощью теоремы 2.5.1 нетрудно получить формулы для лексикографической производной функции $p(x)$.

<u>Теорема 2.5.2.</u> Пусть x_0 - произвольная точка, а S - базис в R^n. Определим последовательность множеств $\{ B_k \}$ следующим образом:

$B_0 = \mathcal{L}(Q , p(x_0)\cdot) ; \quad S_0 = x_0 ;$

$B_{k+1} = \mathcal{L}(B_{k-1} , \pi(B_{k-1} , S_k)) \cap$

$\cap \{ s \mid \langle S_k - \pi(B_{k-1} , S_k) , s \rangle = 0 \} , \qquad (2.5.9)$

$k = 1 , \dots , n$.

Тогда множество B_n является линейным подпространством, а матрица $g(p, S, x_0)$ - матрицей проектора на это подпространство.

Доказательство. Из теоремы 2.5.1 с очевидностью следует, что

$p_k[S](x) \equiv \pi(B_k, x), k = 0, \dots, n.$

Осталось заметить, что в силу следствия 2.5.1 функция $p(x)$ является лексикографически гладкой и, следовательно, по теореме 2.3.1, функция $p_n[S](x)$ является линейной функцией. В свою очередь, это означает, что множество B_n есть линейное подпространство, а матрица $g(p, S, x_0)$ является матрицей проектора на подпространство B_n □

2.5.2. Лексикографический градиент матричной нормы.

Пусть X - матрица размера $n \times n$. Обозначим через $r(X)$ ее матричную норму:

$r(X) = \max\{ \| X s \| \mid \| s \| = 1 \}.$

Напомним, что функция $r(X)$ является выпуклой функцией от X. Следовательно, она будет и l-гладкой. Кроме того, эта функция является положительно однородной функцией X. Выведем формулы для лексикографического градиента функции $r(X)$, считая матрицу X элементом векторного пространства R^m, где $m = n^2$. При этом будем отождествлять пространство R^m с пространством $n \times n$ - матриц.

Скалярное произведение в пространстве R^m обозначим через $\langle \cdot , \cdot \rangle_m$. Нетрудно убедиться в справедливости формулы

$$\langle X u, v \rangle = \langle X, v u^T \rangle_m \qquad (2.5.10)$$

при любых u, v из R^n, X из R^m

Докажем две вспомогательные леммы.

Лемма 2.5.1. Субдифференциал $\partial r(0)$ функции $r(X)$ в точке $X = 0$ задается формулой

$$\partial r(0) \equiv \operatorname{Conv}\{ Y \mid Y = z s^T, \ z \in R^n, \ s \in R^n, \ \| z \| = \| s \| = 1 \}$$

Доказательство. Из формулы (2.5.10) следует, что

$$r(X) = \max\{ \| X s \| \mid \| s \| = 1 \} \equiv$$
$$\equiv \max\{ \langle X s, z \rangle \mid \| s \| = \| z \| = 1 \} \equiv$$
$$\equiv \max\{ \langle X, Y \rangle_m \mid Y = z s^T, \ \| z \| = \| s \| = 1 \}.$$

Осталось сопоставить это тождество с утверждением 1) леммы 1.2.4. □

Пусть L - произвольное линейное подпространство в R^n. Обозначим через $R(X \mid L)$ коническую оболочку множества $\operatorname{Argmax}\{ \langle X s, s \rangle \mid s \in L, \ \| s \| = 1 \}$.

Лемма 2.5.2. Множество $R(X \mid L)$ есть линейное подпространство.

Доказательство. Положим $r = \langle X s, s \rangle$, где s - произвольный вектор из $R(X \mid L)$. Нам достаточно доказать, что если для двух s_1, s_2 из L

$\langle X s_1, s_1 \rangle = r \| s_1 \|^2$, $\langle X s_2, s_2 \rangle = r \| s_2 \|^2$,

то $\langle X(s_1 + s_2), s_1 + s_2 \rangle = r \| s_1 + s_2 \|^2$.

Действительно,

$$r \| s_1 + s_2 \|^2 \geq \langle X(s_1 + s_2), s_1 + s_2 \rangle =$$
$$= 2\langle X s_1, s_1 \rangle + 2\langle X s_2, s_2 \rangle - \langle X(s_1 - s_2), s_1 - s_2 \rangle =$$
$$= 2r \| s_1 \|^2 + 2r \| s_2 \|^2 - \langle X(s_1 - s_2), s_1 - s_2 \rangle \geq$$
$$\geq r(2\| s_1 \|^2 + 2\| s_2 \|^2 - \| s_1 - s_2 \|^2) =$$
$$= r \| s_1 + s_2 \|^2$$

Таким образом, все неравенства выполняются как равенства. □

Правила вычисления l-градиента функции $r(X)$ даются следующей теоремой.

<u>Теорема 2.5.3.</u> Пусть $X \neq 0$ - $n \times n$-матрица, $U = U(m, m)$ - последовательность линейно независимых $n \times n$ - матриц, $U_0 = X$. Зададим последовательность линейных подпространств в R^n:

$R_0 = R^n$,

$R_{k+1} = R(X^T U_k \mid R_k)$, $k = 0, \dots, m - 1$.

Тогда справедлива формула

$g(R, U, X) = r^{-1} (X) s s^T X$,

где s - единственный (с точностью до знака) вектор с единичной евклидовой нормой из подпространства R_m.

Д о к а з а т е л ь с т в о. Положим $r = r(X)$. Напомним, что в силу леммы 2.3.5 лексикографический градиент функции $r(X)$ будет единственным элементом последнего из множеств B_k:

$B_0 = \partial r(X)$,

$B_k = \mathrm{Argmax} \{ \langle U_k , Z \rangle_m \mid Z \in B_{k-1} \}$,

$k = 1, \dots, m$.

Напомним также, что $U_0 = X$, и поэтому

$$\partial r(X) = \mathrm{Argmax} \{ \langle U_0 , Z \rangle_m \mid Z \in \partial r(0) \} . \qquad (2.5.11)$$

Таким образом, из (2.5.11), лемм 2.5.1, 2.5.2 и предположения $X \neq 0$ имеем

$B_0 = \mathrm{Conv}\, \mathrm{Argmax} \{ \langle X , z s^T \rangle_m \mid \| s \| = \| z \| = 1\} =$
$= \mathrm{Conv} [z s^T \mid (z, s) \in \mathrm{Argmax} \{ \langle X s, z \rangle \mid$
$\| s \| = \| z \| = 1 \}] =$
$= \mathrm{Conv} [r^{-1} X s s^T \mid s \in \mathrm{Argmax} \{ \| X s \| \mid \| s \| = 1\}] =$
$= \mathrm{Conv} [r^{-1} X s s^T \mid s \in R_1 , \| s \| = 1]$.

Продолжим доказательство формулы

$$B_k = \mathrm{Conv} [r^{-1} X s s^T \mid s \in R_{k+1} , \| s \| = 1] ,$$
$$k = 0, \dots , m - 1 , \qquad (2.5.12)$$

по индукции. Пусть эта формула справедлива при некотором $k > 0$. Тогда в силу определения множеств B_k, R_k и индуктивного предположения имеем

$$B_{k+1} = \mathrm{Argmax}\{ \langle U_{k+1}, Z \rangle_m \mid Z \in B_k \} =$$
$$= \mathrm{Conv}[r^{-1} X s s^T \mid$$
$$s \in \mathrm{Argmax}\{ \langle U_{k+1}, X z z^T \rangle_m \mid z \in R_{k+1}, \ \| z \| = 1 \}] =$$
$$= \mathrm{Conv}[r^{-1} X s s^T \mid$$
$$s \in \mathrm{Argmax}\{ \langle X^T U_{k+1} z, z \rangle \mid z \in R_{k+1}, \ \| z \| = 1 \}] =$$
$$= \mathrm{Conv}[r^{-1} X s s^T \mid s \in R(X^T U_{k+1} \mid R_{k+1}), \ \| s \| = 1] =$$
$$= \mathrm{Conv}[r^{-1} X s s^T \mid s \in R_{k+2}, \ \| s \| = 1].$$

Таким образом, формула (2.5.12) справедлива при всех $k = 0, \ldots, m - 1$. Заметим, что в силу соотношения (2.5.11) и леммы 2.3.5 множество B_{m-1} при $X \neq 0$ будет состоять из одной точки. А это в силу справедливости формулы (2.5.12) при $k = m - 1$ доказывает теорему. □

Приведем без доказательства следствие из теоремы 2.5.3, показывающее, как вычислять лексикографический градиент в точке $X = 0$.

<u>Следствие 2.5.2.</u> Пусть $U = U(m, m)$ - последовательность линейно независимых $n \times n$ - матриц. Положим

$$R_0 = R^n, \ R_{k+1} = R(U_1^T U_k \mid R_k), \ k = 0, \ldots, m - 1.$$

Тогда справедлива формула

$$g(r, U, 0) = r^{-1}(U_1) s s^T U_1,$$

где s - единственный (с точностью до знака) вектор с единичной евклидовой нормой из подпространства R_m.

2.5.3. Лексикографические градиенты некоторых характеристик симметрических матриц

Покажем, как вычислять лексикографические градиенты максимального собственного числа, минимального собственного числа и числа обусловленности симметрической матрицы.

Введем следующие обозначения:

$L(X) = \max \{ \langle X s, s \rangle \mid \| s \| = 1 \}$,

$l(X) = \min \{ \langle X s, s \rangle \mid \| s \| = 1 \}$,

$q(X) = L(X) / l(X)$,

где X - симметричная $n \times n$ - матрица. Множество таких матриц будем отождествлять с пространством R^m, $m = (n + 1) n / 2$. Нетрудно убедиться в том, что функции $L(X)$ и $l(X)$ являются соответственно выпуклой и вогнутой на R^m. А значит, они являются l-гладкими на всем R^m. Следовательно, функция $q(X)$ является l-гладкой в тех точках, где $l(X) \neq 0$. Отметим также, что на открытом множестве

$\mathfrak{M} \equiv \{ X \mid l(X) > 0 \}$

в силу леммы 1.5.1 функция $q(X)$ является квазивыпуклой.

Формулы вычисления лексикографических градиентов функций $L(X)$, $l(X)$ и $q(X)$ приводятся в следующей теореме.

Теорема 2.5.4. Пусть $X \neq 0$ - симметрическая $n \times n$-матрица, $U = U(m, m)$ - последовательность линейно независимых $n \times n$ - матриц. Зададим две последовательности подпространств в R^n:

$R'_0 = \mathrm{Argmax} \{ \langle X s, s \rangle \mid \| s \| = 1 \}$,

$R'_{k+1} = \mathrm{Argmax} \{ \langle U_{k+1} s, s \rangle \mid s \in R'_k \}$,

$k = 0, \dots, m - 2$,

$R''_0 = \mathrm{Argmin} \{ \langle X s, s \rangle \mid \| s \| = 1 \}$,

$R''_{k+1} = \text{Argmin} \{ \langle U_k s, s \rangle \mid s \in R''_k \}$

$k = 0, \ldots, m-2$.

Тогда

$g(L, U, X) = s_1 s_1^T$, $g(l, U, X) = s_2 s_2^T$,

где s_1, s_2 - единственные (с точностью до знака) векторы с единичной евклидовой нормой подпространств R'_m и R''_m соответственно. Если при этом $l(X) \neq 0$, то

$g(q, U, X) = (s_1 s_1^T - q(X) s_2 s_2^T) / l(X)$.

Мы не будем приводить доказательство этой теоремы, так как оно во многом повторяет доказательство теоремы 2.5.3, отличаясь от него лишь в деталях. Отметим, что при выводе формулы для $g(l, U, X)$ использовалось тождество $L(-X) \equiv - l(X)$, а для получения формулы для $g(q, U, X)$ применялось правило 1 дифференцирования суперпозиции l-гладких функций из §2.4. □

2.6. Необходимые условия экстремума

Всюду в этом параграфе рассматриваются функции $f(x)$: $R^n \to R$ Поэтому символом $g(f, S, x)$ здесь обозначается лексикографический градиент такой функций, который является вектором-столбцом.

2.6.1 Необходимые условия - векторная форма

Основным утверждением, из которого будут следовать необходимые условия экстремума в векторной форме, является следующая стандартная лемма.

Лемма 2.6.1. Если функция $f(x)$ из класса D_n имеет в точке x^* локальный минимум, то $f'(x^*, s) \geq 0$ для всех s из R^n.

Мы не будем приводить доказательства этой леммы в силу ее тривиальности. Утверждение следующей теоремы фактически является переформулировкой утверждения леммы 2.6.1 в терминах дифференциальных характеристик лексикографически гладких функций.

Теорема 2.6.1. Если функция $f(x)$ из $\mathfrak{F}_n$ имеет в точке x^* локальный минимум, то для любого базиса S пространства R^n

$$S^T g(f, S, x^*) \geq 0. \qquad (2.6.1)$$

Доказательство. Для сокращения выкладок положим

$$g = g(f, S, x^*), \quad y = S^T g(f, S, x^*).$$

В силу определения лексикографического неравенства нам достаточно показать, что первая ненулевая компонента вектора y будет положительной.

Пусь $y_1 = y_2 = \ldots = y_k = 0$, $y_{k+1} \neq 0$. Покажем, что в этом случае

$$f_i[S](z) \geq 0 \qquad (2.6.2)$$

для всех z из R^n, $i = 0, \ldots, k$.

Действительно, в силу леммы 2.6.1

$$f_0[S](z) = A_n[x^*] f(z) \geq 0$$

при всех z из R^n. Далее, если (2.6.2) выполнено при $i = j - 1$, то в силу леммы 2.3.4

$$f_j[S](z) = \lim_{\alpha \to +0} \alpha^{-1} [f_{j-1}[S](S_j + \alpha z) -$$

$$- f_{j-1}[S](S_j)] =$$

$$= \lim_{\alpha \to +0} \alpha^{-1} [\, f_{j-1} [\, S \,] (\, S_j + \alpha z \,) - y_j \,] =$$

$$= \lim_{\alpha \to +0} \alpha^{-1} [\, f_{j-1} [\, S \,] (\, S_j + \alpha z \,) \,] \geq 0$$

при всех z из R^n. Таким образом, соотношение (2.6.2) доказано. Теперь из (2,6.2) в силу леммы 2.3.4 получаем
$y_{k+1} = \langle S_{k+1} , g \rangle = f_k[\, S \,] (\, S_{k+1} \,) \geq 0$.
Но по предположению $y_{k+1} \neq 0$. Следовательно, $y_{k+1} > 0$. □

С помощью неравенств типа (2.6.1) удается выписать необходимые и достаточные условия квазидифференцируемости функций по Пшеничному. Напомним, что функция $f(\, x \,)$ называется квазидифференцируемой по Пшеничному в точке x_0, если существует выпуклое множество $\partial_P f(\, x_0 \,)$, такое, что
$A_n [\, x_0 \,] f (\, z \,) = \max \{ \langle z , s \rangle \mid s \in \partial_P f(\, x_0 \,) \}$.

<u>Теорема 2.6.2.</u> Лексикографически гладкая функция $f(\, x \,)$ является квазидифференцируемой по Пшеничному в точке x_0 тогда и только тогда, когда для любых базисов S' и S'' в пространстве R^n выполняется соотношение

$$S'^T (\, g(\, f, S', x_0 \,) - g(\, f, S'', x_0 \,) \,) \succeq 0 . \qquad (2.6.3)$$

Доказательство. Для сокращения выкладок положим

$g_1 = g(\, f, S', x_0 \,)$, $y_1 = S'^T g_1$,
$g_2 = g(\, f, S'', x_0 \,)$, $y_2 = S''^T g_2$.

Пусть функция $f(\, x \,)$ является квазидифференцируемой по Пшеничному в точке x_0. Нетрудно показать (это делается так же, как в лемме 2.3.5), что в этом случае вектор g_1 является единственным элементом последнего из множеств B_k :
$B_0 = \partial_P f (\, x_0)$,
$B_{k+1} = \mathrm{Argmax} \{ \langle z, S'_{k+1} \rangle \mid z \in B_k \}$, $k = 0, \ldots , n - 1$.
Предположим, что $y_1^{(i)} = y_2^{(i)}$ при всех $i = 1, \ldots , k - 1$,

и $y_1^{(k)} \neq y_2^{(k)}$. Это означает, что $g_2 \in B_i$, $i = 1, \ldots, k-1$, и $g_2 \notin B_k$, т.е. $y_1^{(k)} > y_2^{(k)}$. Необходимость доказана.

Пусть функция $f(x)$ является лексикографически гладкой в точке x_0 и пусть для любых двух базисов S' и S'' в R^n всегда выполняется хотя бы одно из соотношений (2.6.3). Зададим множество Q следующим образом:

$$Q \equiv \{ y \mid \langle S_1, y - g(f, S, x_0) \rangle \geq 0$$
для всех базисов S }.

Из соотношений (2.6.3) следует, что множество Q содержит все лексикографические градиенты функции $f(x)$ по всевозможным базисам в R^n. Пусть теперь S_1 - произвольное направление в R^n, S - базис в R^n, первым элементом которого является S_1. Тогда в силу определения множества Q и леммы 2.3.4 имеем

$$\max \{ \langle S_1, z \rangle \mid z \in q \} = \langle S_1, g(f, S, x_0) \rangle = A_n [x_0](S_1). \quad \square$$

2.6.2. Необходимые условия - субдифференциальная форма

Обозначим через W_n множество всех базисов в R^n.

Определение 2.6.1. Пусть функция $f(x)$ является лексикографически гладкой в точке x_0. Замыкание выпуклой оболочки множества

$$\{ g(f, S, x_0) \mid S \in W_n \} \qquad (2.6.4)$$

будем называть лексикографическим субдифференциалом функции $f(x)$ в точке x_0, закрепив за ним обозначение $\partial_l f(x_0)$.

Множество $\partial_l f(x_0)$ выпукло и замкнуто в силу своего определения.

Лемма 2.6.2. Множество $\partial_l f(x_0)$ ограничено.

Доказательство. Из определения l-гладкой функции следует, что найдется окрестность точки x_0, в которой функция $f(x)$ является липшицевой с некоторой константой L. Это значит, что для любых x, y из R^n

$$| f_0[S](x) - f_0[S](y) | \equiv$$

$$\equiv | \lim_{\alpha \to +0} \alpha^{-1} [f(x_0 + \alpha x) - f(x_0 + \alpha y)] | \leq$$

$$\leq L \| x - y \| .$$

Аналогично доказывается, что из липшицевости функции $f_k[S]$ следует липшицевость функции $f_{k+1}[S]$. Таким образом, функция

$$f_n[S](x) \equiv \langle g(f, S, x_0), x \rangle$$

является липшицевой. Другими словами, все лексикографические градиенты функции $f(x)$ в точке x_0 равномерно ограничены константой L. Это, в свою очередь, означает, что норма всех элементов множества $\partial_l f(x_0)$ ограничена той же константой L. □

Замечания. 1. Требование замкнутости в определении множества $\partial_l f(x_0)$ существенно, так как несложно привести пример l-гладкой функции, для которой множество (2.6.2) замкнутым не является.

2. Для выпуклой функции $f(x)$ лексикографический субдифференциал совпадает с обычным субдифференциалом.

Определение 2.6.2. Пусть функция $f(x)$ является l-гладкой в точке x_0. Функцию

$$f^l(x_0, x) \equiv \max \{ \langle x, s \rangle \mid s \in \partial_l f(x_0) \}$$

будем называть верхней лексикографической аппроксимацией функции $f(x)$ в этой точке.

Ясно, что функция $f^l(x_0, x)$ является положительно однородной выпуклой функцией, конечной (в силу леммы 2.6.2) при любом x из R^n Причем

$$\partial f^{l}(x_0, 0) \equiv \partial_l f(x_0) \tag{2.6.5}$$

<u>Лемма 2.6.3.</u> При всех x из R^n справедливо неравенство

$$f^{l}(x_0, x) \geqslant f'(x_0, x).$$

Доказательство. Зафиксируем произвольный ненулевой вектор x из R^n И пусть S - базис в R^n, содержащий этот вектор в качестве первого направления. Тогда в силу определения функции $f^{l}(x_0, x)$ имеем

$$f'(x_0, x) = \lim_{\alpha \to +0} \alpha^{-1} [f(x_0 + \alpha x) - f(x_0)] =$$

$$= \langle g(f, S, x_0), x \rangle \leqslant f^{l}(x_0, x). \qquad \square$$

Непосредственным следствием этой леммы является

<u>Теорема 2.6.3.</u> Если функция $f(x)$ из класса $\mathfrak{F}_n$ имеет в точке x^* локальный минимум, то справедливо включение

$$0 \in \partial_l f(x^*). \tag{2.6.6}$$

Доказательство. Из условия теоремы и леммы 2.6.1 следует, что при всех s из R^n справедливо неравенство $0 \leqslant f'(x^*, s)$. Предположим, что ноль не принадлежит лексикографическому субдифференциалу функции $f(x)$ в точке x^*. Тогда в силу теоремы 1.1.3 найдутся вектор s и положительное число λ, такие, что $\langle g, s \rangle \geqslant$ $\geqslant \lambda > 0$ для любого вектора g из $\partial_l f(x^*)$. Но в силу леммы 2.6.3 это означает, что

$$f'(x^*, -s) \leqslant f^{l}(x^*, -s) =$$

$$= \max \{ \langle g, -s \rangle \mid s \in \partial_l f(x^*) \} \leqslant -\lambda.$$

Таким образом, мы пришли к противоречию, которое и доказывает теорему. $\square$

В заключение этого раздела мы рассмотрим связь лексикографического субдифференциала и субдифференциала Кларка.

Напомним, что субдифференциалом Кларка $\partial_c f(x_0)$ локально липшицевой в точке z функции $f(x)$ называется субдифференциал положительно однородной выпуклой функции

$$f^c(z, x) = \operatorname*{suplim}_{\substack{y \to z, \\ \alpha \to +0}} \alpha^{-1} [f(y + \alpha x) - f(y)],$$

которую будем называть верхней аппроксимацией Кларка.

Справедлива следующая теорема

Теорема 2.6.4. Если функция $f(x)$ является лексикографически гладкой в точке x_0, то при всех x из R^n справедливо неравенство

$$f^l(x_0, x) \leq f^c(x_0, x),$$

и, следовательно, имеет место включение

$$\partial_l f(x_0) \subset \partial_c f(x_0). \tag{2.6.7}$$

Доказательство. Пусть S – произвольный базис в R^n. Докажем поточечную монотонность верхних аппроксимаций Кларка для последовательности линеаризаций $\{ f_k [S] \}$:

$$(f_0 [S])^c (0, x) \leq f^c(x_0, x). \tag{2.6.8}$$

$$(f_{k+1} [S])^c (0, x) \leq (f_k [S])^c (0, x). \tag{2.6.9}$$

Мы покажем, что неравенства (2.6.8), (2.6.9) справедливы для любых x из R^n при всех k, $0 \leq k \leq n-1$.

Действительно,

$$(f_0 [S])^c (0, x) =$$

$$= \operatorname*{suplim}_{\substack{y \to 0, \\ \beta \to +0}} \beta^{-1} [f_0 [S](y + \beta x) - f_0 [S](y)] =$$

$$= \operatorname*{suplim}_{\substack{y \to 0, \\ \beta \to +0}} \beta^{-1} \{ \lim_{\alpha \to +0} \alpha^{-1} [f(x_0 + \alpha(y + \beta x)) - f(x_0)] -$$

$$- \lim_{\alpha \to +0} \alpha^{-1} [f(x_0 + \alpha y) - f(x_0)] \} \leq$$

$$\leq \operatorname*{suplim}_{\substack{y \to 0; \\ \alpha, \beta \to +0}} (\alpha \beta)^{-1} [f(x_0 + \alpha y + \alpha \beta x) - f(x_0 + \alpha y)] \equiv$$

$$\equiv \operatorname*{suplim}_{\substack{z \to x_0 \\ \xi \to +0}} \xi^{-1}[\, f(z + \xi x) - f(z)\,] \equiv f^c(x_0, x) .$$

Справедливость неравенства (2.6.9) устанавливается с помощью аналогичных рассуждений по индукции.

Таким образом, мы показали, что для любого базиса S при всех x из R^n справедливо неравенство

$$\langle g(f, S, x_0), x \rangle \leq f^c(x_0, x) .$$

Осталось воспользоваться определением верхней лексикографической аппроксимации. □

Покажем на конкретном примере, что включение (2.6.7) может быть строгим.

Пример 2.6.1. Зададим функцию одного переменного $f(t)$, $|t| \leq 1$, следующим образом:

$$f(t) = \max \{ t_k^2 - 2 \mid |t| - t_k \mid \;\; ; \;\; k = 1,2,\dots \} ,$$

где последовательность положительных чисел t_k строго монотонно сходится к нулю. Функция $f(t)$ будет, очевидно, липшицевой и лексикографически гладкой на всей вещественной прямой.

Нетрудно подобрать числа t_k так, чтобы функция $f(t)$ оказалась неотрицательной на интервале $[-1, +1]$. Но в этом случае функция $f(t)$ будет дифференцируемой в нуле и ее единственный лексикографический градиент в этой точке равен нулю. В то же время, субдифференциалом Кларка в точке $t = 0$ является отрезок $[-2, +2]$. □

Таким образом, пример 2.6.1 показывает, что с помощью включения (2.6.6) можно осуществить более строгий отбор кандидатов на экстремум, чем с помощью субдифференциала Кларка.

Важным достоинством субдифференциала Кларка является возможность обобщения с его помощью теоремы о среднем. Докажем аналогичное утверждение для лексикографического субдифференциала.

<u>Теорема 2.6.5.</u> Пусть функция $f(x)$ является лексикографически гладкой на всем пространстве R^n. Тогда для любых двух точек x, y из R^n найдется точка $z = (1 - t)x + t y$, $0 < t < 1$, такая, что

$$f(y) - f(x) = \langle g, y - x \rangle ,$$

где g - некоторый вектор из $\partial_l f(z)$.

Доказательство. Зафиксируем произвольные точки x, y в R^n, $x \neq y$. Положим $s = y - x$. Рассмотрим функцию

$$\varphi(z) = f(z) - f(x) + (f(y) - f(x)) \langle s, z - x \rangle / \| s \|^2 .$$

Заметим, что $\varphi(y) = \varphi(x) = 0$. Покажем, что на отрезке $[x, y]$ найдется строго внутренняя точка z, такая, что производные функции φ по направлениям s и $-s$ в ней совпадают по знаку.

Действительно, пусть z^* - точка максимума, а z_* - точка минимума функции $\varphi(z)$ на отрезке $[x, y]$. Если ни одна из этих точек не лежит внутри отрезка $[x, y]$, то, как нетрудно убедиться, $\varphi(z) = \mathrm{const}$ на всем отрезке и ее производные по направлениям s, $-s$ в любой внутренней точке этого множества равны нулю.

Пусть какая-либо из точек (например, z_*) лежит строго внутри отрезка $[x, y]$. Но тогда в силу леммы 2.6.1 производные функции $\varphi(z)$ в точке z_* по направлениям s и $-s$ неотрицательны. Таким образом, в любом случае существует строго внутренняя точка z_* отрезка $[x, y]$, для которой найдется число ξ, $0 < \xi < 1$, обеспечивающее выполнение равенства

$$\xi A_n[z_*] \varphi(s) = (1 - \xi) A_n[z_*] \varphi(-s) .$$

Отсюда, учитывая определение функции $\varphi(z)$, получаем

$$f(y) - f(x) = \xi\, A_n[z_*]\, f(s) - (1-\xi)\, A_n[z_*]\, f(-s). \quad (2.6.10)$$

Но из определения l-гладкой функции следует, что найдутся базисы S', S'' пространства R^n такие, что

$$A_n[z_*]\, f(s) = \langle g(f, S', z_*),\ s \rangle,$$

$$A_n[z_*]\, f(-s) = \langle g(f, S'', z_*),\ -s \rangle.$$

Поэтому из (2.6.10) получаем

$$f(y) - f(x) = \langle \xi\, g(f, S', z_*) + (1-\xi)\, g(f, S'', z_*),\ s \rangle.$$

Осталось воспользоваться определением множества $\partial_l f(x)$.

□

И, наконец, отметим еще одно свойство субдифференциала Кларка. Это свойство полунепрерывности сверху. Отсутствие этого свойства у лексикографического субдифференциала демонстрирует пример 2.6.1. Однако, как видно из этого же примера, полунепрерывность сверху субдифференциала Кларка обусловлена привлечением в него векторов, не имеющих прямого отношения к аппроксимации исходной функции в окрестности исследуемой точки.

2.6.3. О методе субградиентного спуска

Метод обобщенного субградиентного спуска состоит в следующем. Пусть $\{ h_k \}$ - априорно заданная последовательность чисел, такая, что

$$h_k > 0, \quad h_k \to 0, \quad \sum_{k=1}^{\infty} h_k = \infty. \quad (2.6.11)$$

Метод обобщенного субградиентного спуска заключается в последовательном пересчете точек $\{ x_k \}$ начиная с некоторого начального приближения x_0:

$$x_0 \in R^n,$$

$$x_{k+1} = x_k - h_k g_k, \quad k = 0, 1, \ldots, \quad (2.6.12)$$

где g_k - некоторые аналоги градиентов негладкой функции

$f(x)$ в точках x_k. При некоторых предположениях о свойствах функции $f(x)$ (дифференцируемость либо выпуклость) удается доказать сходимость метода (2.6.11), (2.6.12) к стационарной точке (этому посвящен §2.3.2). Если же функция $f(x)$ не является дифференцируемой или не обладает какими-либо специальными свойствами типа выпуклости, то метод (2.6.11),(2.6.12) может и не сходиться к стационарной точке функции $f(x)$. Покажем на конкретном примере, что это действительно так.

Пример 2.6.2. Пусть последовательность шагов $\{h_k\}$ удовлетворяет условиям (2.6.11) и, кроме того, является строго монотонно убывающей. Зададим последовательность точек $\{x_k\}$ следующим образом:

x_0 - произвольная точка из R,

$$x_{k+1} = x_k - (-1)^k h_k, \quad k = 0, 1, \dots . \tag{2.6.13}$$

Нетрудно видеть, что система отрезков $[x_k, x_{k+1}]$ является стягивающейся. Обозначим через t_* единственную общую точку этих отрезков. Отметим, что все точки x_k с четными номерами лежат справа от точки t_*, а с нечетными - слева.

Нам потребуется вспомогательная функция

$$\xi(t) = [0.5(t - t_*)^2 + t - t_*] / (h_0 + 1).$$

Легко проверить, что на любом из отрезков $[x_k, x_{k+1}]$ производная функции $\xi(t)$ по модулю не превосходит единицы. И, наконец, зададим последовательность чисел $\{s_k\}$:

$$s_k = x_k + (-1)^k e_k, \tag{2.6.14}$$

где положительные числа e_k достаточно малы (требования к их малости будут приведены немного позже).

Выпишем целевую функцию $f(x)$.

$$f(x) = \max\{(h_0 + 1)^{-1}(x - t_*);$$
$$\xi(s_k) - |x - s_k|, \ k = 0, 1, \dots\}.$$

Заметим, что функция $f(x)$ не превосходит функции $\xi(x)$ и всюду больше функции $(x - t_*) / (h_0 + 1)$. Поэтому она дифференцируема в точке $x = t_*$ и ее производная равна $(h_0 + 1)^{-1} \neq 0$. С другой стороны, ясно, что, выбрав в (2.6.14) числа e_k достаточно малыми, можно добиться того, чтобы функция $f(x)$ оказалась непрерывно дифференцируема в некоторой окрестности каждой из точек x_k и чтобы ее производная в этих точках равнялась $(-1)^k$. Но тогда метод (2.6.11), (2.6.12), запущенный из точки x_0 пройдет в точности по точкам x_k и сойдется к точке t_*. □

Таким образом, мы получили, что метод (2.6.11), (2.6.12), примененный к липшицевой функции и идущий по точкам дифференцируемости этой функции может сходится к точке с ненулевой производной даже в тех случаях, когда в качестве направления движения выбирается правильный градиент текущей точки.

Этот вывод остается верным для всех обобщений понятия градиента гладкой функции,и в частности для лексикографических градиентов. Разница заключается лишь в оценке конечного результата - точки t_*. Так, например, если с точки зрения теории лексикографического дифференцирования точка t_* не является стационарной точкой функции $f(x)$, то для субдифференциала Кларка оказывается справедливым включение $0 \in \partial_c f(t_*) \equiv [-1, 1]$.

5 Зак. 2406

2.7. Некоторые теоремы из анализа

В этот параграф включены результаты, демонстрирующие возможности лексикографических производных как инструмента анализа негладких функций. Перечень приводимых примеров носит в достаточной степени произвольный характер и может быть без труда расширен.

Как уже отмечалось выше (см. лемму 2.3.3), множество лексикографических производных функции $f(x)$ исчерпывается l-производными по всевозможным ортонормированным базисам пространства R^n. Оказывается, любая лексикографически гладкая функция в некотором смысле полностью описывается системой своих l-производных по любому фиксированному базису пространства R^n. А именно, справедлива следующая теорема

<u>Теорема 2.7.1.</u> Пусть $x(t)$ - непрерывно дифференцируемая вектор-функция, $x(t) : [0, T] \to R^n$, $f(x)$ - лексикографически гладкая на R^n функция. Тогда справедлива формула Ньютона-Лейбница

$$f(x(T)) - f(x(0)) = \int_0^T g(f, S, x(t))\, x'(t)\, dt. \qquad (2.7.1)$$

Интеграл в формуле (2.7.1) и всюду далее понимается в смысле Лебега. В дальнейшем мы предполагаем базис S фиксированным

Прежде чем доказывать теорему 2.7.1, докажем вспомогательную лемму.

<u>Лемма 2.7.1.</u> Пусть функция $f(x)$ принадлежит классу $\mathcal{F}_n$. Тогда при любых x_0 и y из R^n справедливы тождества

$$f_k[S](y) = \lim_{\alpha \to +0} \ldots \lim_{\alpha_0 \to +0} \{$$

$$[\, f(x_0 + \alpha_0 S_1 + \alpha_0\alpha_1 S_2 + \ldots + \alpha_0 \ldots \alpha_{k-1} S_k + \alpha_0 \ldots \alpha_k y) -$$

$- f(x_0 + \alpha_0 S_1 + \alpha_0\alpha_1 S_2 + \ldots + \alpha_0\ldots\alpha_{k-1}S_k)]/(\alpha_0\ldots\alpha_k)\}$,

$k = 0, \ldots, n$. (2.7.2)

Д о к а з а т е л ь с т в о. Заметим, что

$$f_0[S](x) \equiv \lim_{\alpha_0 \to +0} \alpha_0^{-1}[f(x_0 + \alpha_0 x) - f(x)],$$

$$f_1[S](x) \equiv \lim_{\alpha_1 \to +0} \alpha_1^{-1}[f_0[S](S_1 + \alpha_1 x) -$$

$$- f_0[S](S_1)] =$$

$$= \lim_{\alpha_1 \to +0} \alpha_1^{-1}\{\lim_{\alpha_0 \to +0} \alpha_0^{-1}[f(x_0 + \alpha_0 S_1 + \alpha_0\alpha_1 x) - f(x_0)] -$$

$$- \lim_{\alpha_0 \to +0} \alpha_0^{-1}[f(x_0 + \alpha_0 S_1) - f(x_0)]\} =$$

$$= \lim_{\alpha_1 \to +0} \lim_{\alpha_0 \to +0} \alpha_0^{-1}\alpha_1^{-1}[f(x_0 + \alpha_0 S_1 + \alpha_0\alpha_1 x) -$$

$$- f(x_0 + \alpha_1 S_1)]$$

И так далее. □

Теперь мы можем перейти к доказательству теоремы 2.7.1.

Д о к а з а т е л ь с т в о. Введем последовательность вспомогательных функций $\{\varphi_k\}$:

$$\varphi_0[(\alpha), S](t) \equiv$$

$$\equiv [f(x(t) + \alpha_0 S_1 + \alpha_0\alpha_1 S_2 + \ldots$$

$$+ \alpha_0\ldots\alpha_{n-1}S_n + \alpha_0\ldots\alpha_n x'(t)) -$$

$$- f(x(t) + \alpha_0 S_1 + \alpha_0\alpha_1 S_2 + \ldots -$$

$$+ \alpha_0\ldots\alpha_{n-1}S_n)]/(\alpha_0\ldots\alpha_n),$$

$$\varphi_{k+1}[(\alpha), S](t) = \lim_{\alpha_k \to +0} \varphi_k[(\alpha), S](t),\ k = 0, \ldots, n,$$

где $(\alpha) \equiv (\alpha_0, \ldots, \alpha_n) > 0$, $t \in [0, T]$.

По индукции нетрудно установить, что

$$\varphi_k[(\alpha), S](t) \equiv$$

$$\equiv [f_{k-1}[S](S_k + \alpha_k S_{k+1} + \ldots + \alpha_k\ldots\alpha_{n-1}S_n +$$

$$+ \alpha_0\ldots\alpha_n x'(t)) - f_{k-1}[S](S_k + \alpha_k S_{k+1} + \ldots +$$

$+ \alpha_k \ldots \alpha_{n-1} S_n)] / (\alpha_k \ldots \alpha_n),$

$$k = 1, \ldots, n, \tag{2.7.3}$$

где последовательность линеаризаций $f_k[S]$ вычисляется в точке $x(t)$. Напомним, что функция $f(x)$ является локально липшицевой в каждой точке пространства R^n Это эквивалентно тому, что функция $f(x)$ является липшицевой на каждом ограниченном множестве. Обозначим константу Липшица функции $f(x)$ для ε -окрестности кривой $x(t)$, $0 \leq t \leq T$, через L. Тогда все линеаризации $f_k[S]$, вычисленные в любой из точек $x(t)$, будут удовлетворять условию Липшица с константой L. Сопоставляя этот факт с тождеством (2.7.3), получаем, что все функции $\varphi_k[(\alpha), S](t)$, $0 \leq k \leq n+1$, ограничены в совокупности на отрезке $[0, T]$:

$$\| \varphi_k [(\alpha), S](t) \| \leq$$
$$\leq L \max \{ \| x(t) \| \mid 0 \leq t \leq T \} < \infty$$

В свою очередь, это означает, что функции $\varphi_k[(\alpha), S](t)$ суммируемы на отрезке $[0, T]$ Теперь мы можем вычислить значение интеграла

$$\mathfrak{J} = \int_0^T g(f, S, x(t))\, x'(t)\, dt,$$

учитывая, что в силу леммы 2.7.1

$$\varphi_{n+1}[(\alpha), S](t) \equiv g(f, S, x(t))\, x'(t)$$

Воспользовавшись теоремой Лебега, получаем

$$\mathfrak{J} = \int_0^T \varphi_{n+1}[(\alpha), S](t)\, dt =$$

$$= \int_0^T \lim_{\alpha_n \to +0} \varphi_n[(\alpha), S](t)\, dt =$$

$$= \lim_{\alpha_n \to +0} \int_0^T \varphi_n [(\alpha) , S] (t) dt = \ldots =$$

$$= \lim_{\alpha_n \to +0} \ldots \lim_{\alpha_0 \to +0} \int_0^T \varphi_0 [(\alpha), S] (t) dt .$$

Нетрудно видеть, что в силу непрерывной дифференцируемости функции $x(t)$

$$\varphi_0 [(\alpha), S] (t) =$$
$$= [f(x(t) + \alpha_0 S_1 + \ldots + \alpha_0 \ldots \alpha_{n-1} S_n + \alpha_0 \ldots \alpha_n x'(t)) -$$
$$- f(x(t) + \alpha_0 S_1 + \ldots + \alpha_0 \ldots \alpha_{n-1} S_n)] / (\alpha_0 \ldots \alpha_n) =$$
$$= [f(x(t + \alpha_0 \ldots \alpha_n) + \alpha_0 S_1 + \ldots + \alpha_0 \ldots \alpha_{n-1} S_n) -$$
$$- f(x(t) + \alpha_0 S_1 + \ldots + \alpha_0 \ldots \alpha_{n-1} S_n)]/(\alpha_0 \ldots \alpha_n) + o(1) .$$

Поэтому интеграл $\mathfrak{J}$ преобразуется следующим образом:

$$\mathfrak{J} = \lim_{\alpha_n \to +0} \ldots \lim_{\alpha_0 \to +0} \{$$

$$\int_0^T [f(x(t + \alpha_0 \ldots \alpha_n) + \alpha_0 S_1 + \ldots + \alpha_0 \ldots \alpha_{n-1} S_n) -$$

$$- f(x(t) + \alpha_0 S_1 + \ldots + \alpha_0 \ldots \alpha_{n-1} S_n)]/(\alpha_0 \ldots \alpha_n) dt \} =$$

$$= \lim_{\alpha_n \to +0} \ldots \lim_{\alpha_0 \to +0} \{ [\int_0^{T + \alpha_0 \ldots \alpha_n} f(x(t) + \alpha_0 S_1 + \ldots + \alpha_0 \ldots \alpha_{n-1} S_n) dt -$$

$$- \int_0^{\alpha_0 \ldots \alpha_n} f(x(t) + \alpha_0 S_1 + \ldots + \alpha_0 \ldots \alpha_{n-1} S_n) dt]/(\alpha_0 \ldots \alpha_n)\} =$$

$$= \lim_{\alpha_n \to +0} \ldots \lim_{\alpha_0 \to +0} \{ \int_0^1 f(x(T + t \alpha_0 \ldots \alpha_n) + \alpha_0 S_1 + \ldots +$$
$$+ \alpha_0 \ldots \alpha_{n-1} S_n) dt -$$

$$- \int_0^1 f(x(t \alpha_0 \ldots \alpha_n) + \alpha_0 S_1 + \ldots + \alpha_0 \ldots \alpha_{n-1} S_n) dt\} =$$

$$= \lim_{\alpha_n \to +0} \dots \lim_{\alpha_1 \to +0} \{f(x(T)) - f(\dot{x}(0))\} = f(x(T)) - f(x(0)).$$

□

Тривиальным следствием теоремы 2.7.1 и леммы 2.7.1 является

Лемма 2.7.2. Для линеаризаций лексикографически гладкой функции $f(x)$ в точке x_0 по базису S имеет место следующее представление:

$$f_k[S](x + y) = f_k[S](x) + \lim_{\alpha_k \to +0} \dots \lim_{\alpha_0 \to +0} \{$$

$$\int_0^1 g(f, S, x_0 + \alpha_0 S_1 + \dots + \alpha_0 \dots \alpha_{k-1} S_k +$$

$$+ \alpha_0 \dots \alpha_k (x + t y)) dt \} y.$$

В частности,

$$g(f, S, x_0) y = \lim_{\alpha_k \to +0} \dots \lim_{\alpha_0 \to +0} \{$$

$$\int_0^1 g(f, S, x_0 + \alpha_0 S_1 + \dots + \alpha_0 \dots \alpha_{k-1} S_k +$$

$$+ t \alpha_0 \dots \alpha_k y) dt \} y.$$

С помощью лексикографических производных удается сформулировать условия потенциальности негладких операторов.

Теорема 2.7.2. Пусть отображение $F(x) : R^n \to R^n$ принадлежит классу $\mathfrak{F}_n$. Тогда для его потенциальности (в смысле Гато) необходимо и достаточно, чтобы при любых x, s, h из R^n выполнялось равенство

$$\int_0^1 \int_0^1 \langle g(F, S, x + \xi h + t s) h, s \rangle d\xi dt =$$

$$= \int_0^1 \int_0^1 \langle g(F, S, x + \xi h + t s) s , h \rangle \, d\xi \, dt. \tag{2.7.4}$$

Д о к а з а т е л ь с т в о. Пусть отображение $F(x)$ потенциально, т.е. пусть существует дифференцируемая по Гато функция $f(x)$, такая, что при любых x из R^n справедливо равенство $f'(x) = F(x)$. Тогда при любых x, h, s из R^n имеем

$$f(x + h + s) - f(x + h) - f(x + s) + f(x) =$$
$$= [f(x + h + s) - f(x + h)] - [f(x + s) - f(x)] =$$
$$= \int_0^1 \langle F(x + h + ts), s \rangle \, dt - \int_0^1 \langle F(x + ts), s \rangle \, dt =$$
$$= \int_0^1 \langle F(x + h + t s) - F(x + t s) , s \rangle \, dt =$$
$$= \int_0^1 \int_0^1 \langle g(F, S, x + \xi h + t s) h , s \rangle \, d\xi \, dt .$$

Группируя эту разность по-другому, можно получить, что

$$f(x + h + s) - f(x + h) - f(x + s) + f(x) =$$
$$= [f(x + h + s) - f(x + s)] - [f(x + h) - f(x)] =$$
$$= \int_0^1 \int_0^1 \langle g(F, S, x + \xi h + t s) s , h \rangle \, d\xi \, dt .$$

Необходимость доказана.

Прежде чем доказывать достаточность, заметим, что из равенства (2.7.4) с очевидностью следует, что при всех x, s, h из R^n справедливо равенство

$$\int_0^1 \int_0^t \langle g(F, S, x + \xi h + t s) h , s \rangle \, d\xi \, dt =$$

$$= \int_0^1 \int_0^t \langle g(F, S, x + \xi h + t s) s, h \rangle d\xi \, dt. \quad (2.7.5)$$

Предположим, что равенство (2.7.4), а значит, и (2.7.5) выполнено при всех x, s, h из R^n. Введем функцию

$$f(x) = \int_0^1 \langle F(x_0 + t(x - x_0)), x - x_0 \rangle dt .$$

Тогда для любых двух точек x и y будет справедливо тождество

$$f(y) - f(x) \equiv \int_0^1 \langle F(x_0 + t(y - x_0)), y - x \rangle dt +$$

$$+ \int_0^1 \langle F(x_0 + t(y - x_0)) - F(x_0 + t(x - x_0)), x - x_0 \rangle dt$$

Обозначим второй интеграл в этом тождестве через $\mathfrak{J}$. Теперь, воспользовавшись формулой Ньютона - Лейбница, равенством (2.7.5) и изменив порядок интегрирования, получим

$$\mathfrak{J} = \int_0^1 \int_0^t \langle g(F, S, x_0 + t(x - x_0) + \xi(y - x))(y - x),$$

$$x - x_0 \rangle d\xi \, dt =$$

$$= \int_0^1 \int_0^t \langle g(F, S, x_0 + t(x - x_0) + \xi(y - x)))(x - x_0),$$

$$y - x \rangle d\xi \, dt =$$

$$= \int_0^1 \int_\xi^1 \langle g(F, S, x_0 + t(x - x_0) + \xi(y - x)))(x - x_0),$$

$$y - x \rangle dt \, d\xi =$$

$$= \int_0^1 \langle F(x_0 + (x - x_0) + \xi (y - x)) - F(x_0 + \xi (y - x_0)), y - x \rangle d\xi$$

Таким образом,

$$f(y) - f(x) = \int_0^1 \langle F(x + \xi (y - x)), y - x \rangle d\xi .$$

Следовательно, производная Гато функции f в каждой точке x существует и равна $f(x)$. □

Следствие 2.7.1. Если у лексикографически гладкого отображения $F(x)$ при любых x из R^n матрица $g(F, S, x)$ является симметрической, то это отображение потенциально.

Следующая теорема показывает, как надо дифференцировать интегралы от негладких функций, зависящие от параметра.

Теорема 2.7.3. Пусть функция $f(x, y)$ определена на прямом произведении множеств X и Y, где X - открытое множество в пространстве R^n, а Y - измеримое множество в пространстве R^m Предположим, что

1) при любом фиксированном x из X функция $f(x, y)$ суммируема по y на множестве Y ;

2) при всех y из Y функции $f(x, y)$ являются липшицевыми на X функциями с константами Липшица $L(y)$, причем функция $L(y)$ суммируема на Y ;

3) при любом фиксированном y из Y функция $f(x, y)$ является лексикографически гладкой на X

Тогда функция

$$f(x) = \int_Y f(x , y) dy$$

будет лексикографически гладкой на множестве X, и для ее l -производной будет справедлива формула

$$g(f, S, x) = \int_Y g(f, S, x)(y)\, dy, \qquad (2.7.6)$$

где $g(f, S, x)(y)$ - лексикографическая производная функции $f(x, y)$, вычисленная в точке x по базису S пространства R^n при фиксированном векторе y из Y.

Доказательство. Зафиксируем произвольную точку x_0 из множества X. Докажем по индукции, что при всех x из R^n и $k = > 0$ справедлива следующая формула:

$$f_k[S](x) = \int_Y f_k[S](x, y)\, dy. \qquad (2.7.7)$$

Пусть $k = 0$. Зафиксируем произвольное направление x' в R^n. Тогда

$$f_0[S](x, y) = \lim_{\alpha \to +0} \alpha^{-1} [f(x_0 + \alpha x, y) - f(x_0, y)],$$

т. е. функция $f_0[S](x', y)$ является пределом функций

$$\varphi_\alpha(y) = \alpha^{-1} [f(x_0 + \alpha x', y) - f(x_0, y)],$$

причем по условию 2 теоремы при любых y из Y

$$\| \varphi_\alpha(y) \| < L(y) \| x' \|$$

и $L(y)$ - суммируемая функция. Поэтому в силу теоремы Лебега функция $f_0[S](x', y)$ будет суммируемой по y и

$$\int_Y f_0[S](x', y)\, dy =$$

$$= \lim_{\alpha \to +0} \alpha^{-1} \int_Y [f(x_0 + \alpha x', y) - f(x', y)]\, dy =$$

$$= \lim_{\alpha \to +0} \alpha^{-1} [f(x_0 + x') - f(x_0)] = f_0[S](x').$$

Заметим, что при $k = 0$ соотношение

$$\| f_k[S](x', y) - f_k[S](x'', y) \| < L(y) \| x' - x'' \| \qquad (2.7.8)$$

выполняется при любых x', x'' из X и y из Y.

Рассуждая аналогично, нетрудно установить по индукции справедливость соотношений (2.7.7), (2.7.8) для всех $k = 0$.

..., n . Осталось заметить, что в силу теоремы 2.3.1 для любых z из R^n

$$g(f , S , x_0) z = F_n [S] (z) =$$

$$= \int_Y f_n[S] (z , y) dy = \int_Y g(f , S , x_0) (y) z dy =$$

$$= [\int_Y g(f , S , x_0) (y) dy] z ,$$

откуда и вытекает справедливость (2.7.6). □

И, наконец, последняя теорема дает критерий почленного дифференцирования последовательности негладких функций.

Теорема 2.7.4. Пусть каждая функция $f_k(x)$ является лексикографически гладкой на открытом ограниченном множестве Q И пусть последовательность l -производных $g(f_k , S, x)$ равномерно сходится на Q , а последовательность $\{ f_k(x) \}$ сходится хотя бы в одной точке x_0 из Q Тогда последовательность функций $\{ f_k(x) \}$ сходится к некоторой предельной функции $f(x)$ равномерно на Q , причем эту последовательность можно дифференцировать почленно, т.е. всюду на Q функция $f(x)$ имеет лексикографическую производную $g(f, S, x)$, являющуюся пределом лексикографических производных функций $f_k(x)$.

Доказательство. Заметим, что в силу теоремы 7.6.1 для любого $i > 0$

$$\| f_{k+i}(x) - f_k(x) \| < \| f_{k+i}(x_0) - f_k(x_0) \| +$$

$$+ \| f_{k+i}(x) - f_{k+i}(x_0) - (f_k(x) - f_k(x_0)) \| =$$

$$= \| f_{k+i}(x_0) - f_k(x_0) \| +$$

$$+ \| \int_0^1 g(f_{k+i} , S, x_0 + t(x - x_0)) -$$

$$- g(f_k, S, x_0 + t(x - x_0)) dt (x - x_0) \| .$$

Но в силу условий теоремы числовая последовательность

$\{ f_k(x_0) \}$ сходится, а последовательность производных $\{ g(f_k . S. x) \}$ сходится равномерно. Поэтому и функциональная последовательность $f_k(x)$ сходится равномерно на множестве Q. Обозначим ее предел через $f(x)$

Пусть x_0 - произвольная внутренняя точка множества Q. Зафиксируем любое направление x в R^n. Тогда найдется число $\varepsilon > 0$ такое, что при всех α из отрезка $[0 , \varepsilon]$ точки $x_0 + \alpha x$ принадлежат множеству Q. Рассмотрим последовательности функций

$$\varphi_i^0 (\alpha) = \alpha^{-1} [f_i(x_0 + \alpha x) - f_i(x_0)] . \tag{2.7.9}$$

$$\varphi_i^k (\alpha) = \alpha^{-1} [(f_i)_{k-1} [S] (S_k + \alpha x) - (f_i)_{k-1} [S] (S_k)]. \tag{2.7.10}$$

Покажем, что эти функциональные последовательности сходятся равномерно на множестве $[0, \varepsilon]$ при $i \to \infty$.

Действительно, для $k = 0$ и любых $m, i > 0$ в силу теоремы 2.7.1

$$\| \varphi_{m+i}^0 (\alpha) - \varphi_m^0 (\alpha) \| =$$

$$= \| \int_0^1 [g(f_{m+i} . S . x_0 + t \alpha x) -$$

$$- g(f_m . S, x_0 + t \alpha x)] dt \times x \| .$$

Если $k > 0$, то, пользуясь леммой 2.7.2, имеем

$$\| \varphi_{m+i}^k (\alpha) - \varphi_m^k (\alpha) \| = \| \{ \lim_{\alpha_k \to +0} \ldots \lim_{\alpha_0 \to +0}$$

$$\int_0^1 g(f_{m+i} . S. x_0 + \alpha_0 S_1 + \ldots + \alpha_0 \ldots \alpha_{k-1} S_k +$$

$$+ \alpha_0 \ldots \alpha_k(S_{k+1} + t \alpha x)) dt -$$

$$- \int_0^1 g(f_m . S , x_0 + \alpha_0 S_1 + \ldots + \alpha_0 \ldots \alpha_{k-1} S_k +$$

$+ \alpha_0 \ldots \alpha_k (S_{k+1} + t \alpha x)) dt \} x \|$.

Таким образом, равномерная сходимость производных $g(f_i , S, \dot{x})$ на множестве Q обеспечивает и равномерную сходимость функций $\varphi_i^k (\alpha)$ на множестве $[0, \dot{\varepsilon}]$ при $i \to \infty$. Но это означает, что в соотношениях (2.7.9), (2.7.10) можно перейти к пределу при $\alpha \to +0$. Причем этот предел будет существовать и его можно переставить с пределом при $i \to +\infty$. В результате получим

$$(f)_0[S] (x) = \lim_{\alpha \to +0} \alpha^{-1}[f(x_0 + \alpha x) - f(x_0)] =$$

$$= \lim_{\alpha \to +0} \quad \lim_{k \to \infty} \alpha^{-1}[f_k(x_0 + \alpha x) - f_k(x_0)] =$$

$$= \lim_{k \to \infty} \quad \lim_{\alpha \to +0} \alpha^{-1}[f_k(x_0 + \alpha x) - f_k(x_0)] =$$

$$= \lim_{k \to \infty} (f_k)_0 [S](x) .$$

Точно так же из (2.7.10) получаем

$$(f)_m [S] (x) = \lim_{k \to \infty} (f_k)_m [S] (x) \qquad \square$$

2.8. Лексикографическое дифференцирование в бесконечномерных пространствах

Теория лексикографического дифференцирования с небольшими изменениями переносится и на бесконечномерный случай. Действительно, в настоящей главе лишь одно утверждение по существу использует конечномерность пространства переменных - это теорема 2.3.1. В бесконечномерном пространстве подобное утверждение уже не может иметь места. Поэтому определение лексикографических производных в таких пространствах приходится давать исходя из более слабых свойств линеаризаций негладкой функции.

Итак, пусть E_1 - банахово пространство, в котором существует базис $S = \{ S_1, S_2, \ldots \}$. В этом случае каждый элемент x из E_1 однозначно раскладывается в ряд
$x = \sum_{k=1}^{\infty} x^{(k)} S_k$,
сходящийся по норме пространства E_1. И пусть E_2 - произвольное банахово пространство. Введем обозначения:

$\mathcal{D}(X)$ - множество функций, действующих из E_1 в E_2, которые являются локально липшицевыми и дифференцируемыми по всем направлениям пространства E_1 в точках множества X;

$A[x_0]$ - оператор линеаризации функции f из $\mathcal{D}(x_0)$ в точке x_0. Запись $\varphi = A[x_0] f$ означает, что $\varphi(x) = f'(x_0, x)$.

<u>Определение 2.8.1.</u> Функцию $f(x)$ из $\mathcal{D}(x_0)$ будем называть лексикографически гладкой в точке x_0 из E_1, если для любой последовательности направлений $U = \{ U_1, U_2, \ldots \} \subset E_1$ все члены функциональной последовательности $\{ f_k [U] \}$:
$f_0 [U] = A[x_0] f$;
$f_k [U] = A[U_k] f_{k-1}$, $k = 1,2,\ldots$,
принадлежат множеству $\mathcal{D}(E_1)$.

Класс функций, являющихся l-гладкими в каждой точке множества X, обозначим через $\mathcal{F}(X)$.

Можно показать, что класс функций $\mathcal{F}(E_1)$ является линейным пространством, содержащим все локально липшицевы выпуклые функции, и что суперпозиция l-гладких функций остается l-гладкой функцией.

Отметим, что оператор линеаризации и сами линеаризации функции $f(x)$ в бесконечномерном пространстве обладают всеми свойствами, приведенными в леммах 2.3.1, 2.3.2. Поэтому определение $(l\text{-}k)$-производных функции $f(x)$ в точке x_0 по последовательности направлений U остается почти без изменений: это произвольный ограниченный линейный оператор $C(x)$, такой, что

$$f_i [U](x) \equiv C(x)$$

при всех x из $L_i(U)$, $i = 1, \ldots, k$. Однако здесь в отличие от конечномерного случая допускается значение $k = \infty$. Способ же определения l-производной естественным образом вытекает из следующей теоремы

<u>Теорема 2.8.1.</u> Если функция $f(x)$ принадлежит классу $\mathfrak{F}(x_0)$ и S - базис пространства E_1, то существует единственный линейный ограниченный оператор $g \equiv g(f, S, x_0)$, такой, что для любого x из E_1

$$g\ x \equiv \lim_{k\to\infty} f_k [S](x). \tag{2.8.1}$$

Доказательство. Определим оператор $g = g(f, S, x_0) : E_1 \to E_2$ следующим образом:

если

$$x = \Sigma_{k=1}^{\infty} x^{(k)} S_k ,$$

то положим

$$g\ x = \Sigma_{k=1}^{\infty} x^{(k)} f_{k-1}[S](S_k). \tag{2.8.2}$$

Покажем, что это определение корректно.

Введем обозначения

$$x'_N = \Sigma_{k=1}^{N} x^{(k)} S_k ,$$

$$y'_N = \Sigma_{k=1}^{N} x^{(k)} f_{k-1} [S](S_k).$$

Последовательность точек x'_N является фундаментальной в силу того, что она сходится к точке x по норме пространства E_1. Покажем, что последовательность точек y'_N также является фундаментальной. Для этого заметим, что из локальной липшицевости функции $f(x)$ с константой L вытекает локальная липшицевость функций $f_k[S](x)$ с той же константой L. Поэтому, воспользовавшись свойством линеаризаций 3 из леммы 2.3.2, получаем

$$\| y'_{N+p} - y_N \| = \| \sum_{k=N+1}^{N+p} x^{(k)} f_{k-1}[S](S_k) \| =$$
$$= \| f_{N+p}[S] (\sum_{k=N+1}^{N+p} x^{(k)} S_k) \| \leq L \| x'_{N+p} - x'_N \| .$$

Следовательно, последовательность точек $\{ y'_N \}$ является фундаментальной и сходится к некоторой точке y' по норме пространства E_2. Таким образом, формула (2.8.2) корректно определяет линейный оператор $g \equiv g(f, S, x_0)$. Покажем, что этот оператор ограниченный:

$$\| g x \| = \| \lim_{k \to \infty} y'_k \| = \lim_{k \to \infty} \| y'_k \| =$$
$$= \lim_{k \to \infty} \| \sum_{i=1}^{k} x^{(i)} f_{i-1}[S](S_i) \| =$$
$$= \lim_{k \to \infty} \| f_k[S](x'_k) \| \leq \lim_{k \to \infty} L \| x'_k \| = L \| x \| .$$

Осталось доказать справедливость формулы (2.8.1):

$$g x \equiv y' = \lim_{k \to \infty} y'_k = \lim_{k \to \infty} f_k[S](x'_k) .$$

Поэтому

$$\| g x - \lim_{k \to \infty} f_k[S](x) \| =$$
$$= \| \lim_{k \to \infty} (f_k[S](x'_k) - f_k[S](x)) \| \leq$$
$$\leq L \lim_{k \to \infty} \| x'_k - x \| = 0 . \quad \square$$

Линейный оператор $g(f, S, x_0)$, введенный в теореме 2.8.1, будем называть лексикографической производной функции $f(x)$ в точке x_0 по базису S. При таком способе определения лексикографических производных будет справедлив естественный аналог теоремы 2.4.1.

Теорема 2.8.2. Пусть функция $f : E_1 \to E_2$ является l-гладкой в точке x_0, а функция $F(z)$, $z \in E_2$, является l-гладкой в точке $z_0 = f(x_0)$. Тогда функция $\varphi(x) = F(f(x))$ является лексикографически гладкой в точке x_0 и ее l-производная по произвольному базису S пространства E_1 вычисляется по формуле

$$g(\varphi, S, x_0) = g_{\infty}(F, U, z_0)\, g(f, S, x_0),$$

где точки последовательности U задаются соотношениями

$$U_k = g(f, S, x_0)\, S_k, \quad k = 1, \dots .$$

После внесения подобных изменений остаются справедливыми большинство свойств лексикографических производных, рассмотренных в настоящей главе.

ГЛАВА 3 МЕТОДЫ МИНИМИЗАЦИИ НЕГЛАДКИХ ФУНКЦИЙ

3.1. Постановка задачи

В настоящей главе рассматриваются итеративные методы решения следующей экстремальной задачи:

$$\inf \{ f(x) \mid x \in X \}, \tag{3.1.1}$$

где X - выпуклое замкнутое множество в R^n, $f(x)$ - квазивыпуклая, вообще говоря, негладкая функция, определенная на всем пространстве R^n. Часто множество X задается в виде $X = Q \cap \{ x \in R^n \mid F(x) \leq 0 \}$, где Q - выпуклое замкнутое множество в R^n простой структуры; $F(x)$ - вектор-функция, компонентами которой являются квазивыпуклые функции $f^{(1)}(x), \ldots, f^{(m)}(x)$. Помимо квазивыпуклого случая для некоторых методов решения задачи (3.1.1) будут рассмотрены их модификации, рассчитанные на выпуклость функций $f_i(x)$.

Всегда, если не оговорено противное, будем предполагать, что:

1) функции $f(x)$, $f_1(x)$, ..., $f_m(x)$ в задаче (3.1.1) непрерывны, квазивыпуклы и определены на всем пространстве R^n;

2) множество $X^* = \operatorname{Argmin} \{ f(x) \mid x \in X \}$ не пусто;

3) для любого вектора $x \in X \setminus X^*$ ($x \in X_k \equiv \{ y \in R^n \mid f^{(k)}(y) > 0 \}$) найдется ненулевой вектор g, принадлежащий множеству $D f(x)$ (вектор $l^{(k)}$, принадлежащий множеству $D f^{(k)}(x)$).

Чтобы в дальнейшем избежать неоднозначности в обозначениях, поставим в соответствие каждой точке x из X произвольный вектор $g(x)$ из $D f(x)$, причем, если $x \notin X^*$, то выберем $g(x) \neq 0$. Аналогично, каждой точке $x \in X_k$ поставим в соответствие ненулевой вектор $l^{(k)}(x)$ из $D f^{(k)}(x)$, $k = 1, \ldots, m$. Если речь идет о выпуклой функции $f(x)$ (или $f^{(k)}(x)$), то под $g(x)$ ($l^{(k)}(x)$) понимается субградиент соответствующей функции в точке x.

Далее для произвольной функции $a(x)$, определенной на последовательности точек $\{ x_k \}_{k=0}^{\infty}$, будем обозначать $a_k = a(x_k)$. Для произвольной скалярной последовательности $\{ b_k \}_{k=0}^{\infty}$ определим ее "рекордные" значения следующим образом: $b_N^* = \min \{ b_k \mid 0 \leq k \leq N \}$.

В настоящей главе, помимо установления факта сходимости того или иного метода, выводятся гарантированные оценки его скорости сходимости на определенных классах экстремальных задач, причем оценка скорости сходимости, как правило, выписывается для значений целевой функции в "рекордных" точках траектории соответствующего метода. Оценки скорости сходимости всех итеративных методов, рассматриваемых ниже, выводятся по единой двухэтапной схеме. На первом этапе устанавливается оценка скорости стремления к нулю величин v_k^*, где

$v(x) = \langle g(x), x - x^* \rangle / \| g(x) \|$,

x^* - произвольная фиксированная точка из множества X^*. При этом используются лишь самые общие свойства функциональных компонент задачи (3.1.1) типа выпуклости их множеств уровня. Более точные сведения о классе решаемых задач используются на втором этапе, который состоит в преобразовании оценок скорости стремления к нулю величин v_k^* в оценки скорости сходимости значений f_k^* к f^*. Для этой цели используется следствие 1.5.1, в силу которого

$$f(x_k) - f^* \leq \omega(v_k).$$

где $\omega(t) \equiv \omega(f, x^*, t)$ - мажоранта роста целевой функции в окрестности точки минимума x^* Поэтому, если для рассматриваемого класса целевых функций $\mathfrak{F}$ можно указать общую мажоранту $\eta(t)$ их роста в ограниченной окрестности любой точки минимума ($\omega(f, x^*, t) \leq \eta(t)$, $0 \leq t \leq R$, для любой функции $f(x) \in \mathfrak{F}$ и точки $x^* \in X^*$), то оценка скорости сходимости данного метода на этом классе задач является простым следствием неравенства

$$f_k^* - f^* \leq \eta(v_k^*).$$

Приведем примеры мажорант $\eta(t)$ для различных классов целевых функций.

1. Пусть для целевых функций из класса $\mathfrak{F}$ известна общая мажоранта их модулей непрерывности

$$\sup \{ | f(x) - f(y) | \mid x, y \in B(x^*, R), \|x - y\| \leq t \} \leq \eta(t)$$

для любого t, $0 \leq t \leq R$. Тогда $\omega(f, x^*, t) \leq \eta(t)$, $0 \leq t \leq R$, для любой функции f из $\mathfrak{F}$.

2. В частности, если функции f из класса $\mathfrak{F}$ липшицевы на шаре $B(x^*, R)$ с константой C, то

$$\omega(f, x^*, t) \leq Ct, \quad 0 \leq t \leq R.$$

3. Если функции из класса $\mathfrak{F}$ на шаре $B(x^*, R)$ выпуклы и ограничены константой C, то

$$\omega(f, x^*, t) \leq CR^{-1}t, \quad 0 \leq t \leq R.$$

4. Если $\mathfrak{F} = C^{1,\nu}(L)$, то в силу теоремы 1.4.1 для любых x, y из R^n

$$f(y) \leq f(x) + \langle f'(x), y - x \rangle + (1 + \nu)^{-1} L \|y - x\|^{1+\nu}$$

и, следовательно,

$$f(y) \leq f(x) + \|f'(x)\| \times \|y - x\| + (1 + \nu)^{-1} L \|y - x\|^{1+\nu}.$$

Поэтому

$$\omega(f , x , t) \leq \eta(t) = \| f'(x^*) \| t + (1 + \nu)^{-1} L t^{1+\nu} .$$

В частности, при $\| f'(x^*) \| = 0$

$$\omega(f , x^* , t) \leq \eta(t) = (1 + \nu)^{-1} L t^{1+\nu} ,$$

откуда при $\nu = 1$

$$\omega(f , x^* , t) \leq \eta(t) = 0.5 L t^2 .$$

Перейдем к исследованию поведения конкретных методов решения задачи (3.1.1).

3.2. Субградиентный метод

Пусть необходимо решить задачу минимизации квазивыпуклой функции на множестве простой структуры (в задаче (3.1.1) $X \equiv Q$). Рассмотрим следующий метод:

$x_0 \in R^n$;

$$x_{k+1} = \pi(Q , x_k - h_k g(x_k) / \| g(x_k) \|) , \qquad (3.2.1)$$

где $\{ h_k \}_{k=0}^{\infty}$ - некоторая последовательность шагов, задаваемая априорно. В методе (3.2.1) в зависимости от имеющейся априорной информации о целевой функции $f(x)$ можно применять различные стратегии выбора шагов h_k. Если такая информация отсутствует, то последовательность шагов $\{ h_k \}_{k=0}^{\infty}$ можно задать одним из двух способов:

$$h_k = h > 0 ; \qquad (3.2.2)$$

$$h_k > 0 , \quad h_k \to 0 , \quad \sum_{k=0}^{\infty} h_k = \infty . \qquad (3.2.3)$$

При доказательстве теорем о субградиентном методе нам потребуются две вспомогательные леммы о числовых последовательностях.

Лемма 3.2.1. Если для последовательности чисел $\{ a_k \}_{k=0}^{\infty}$ при всех $k > 0$ выполняются неравенства

$0 < a_{k+1} < a_k + b_k$,

где числа $b_k > 0$ неотрицательны и $\sum_{k=0}^{\infty} b_k < \infty$, то последовательность $\{ a_k \}_{k=0}^{\infty}$ сходится.

Доказательство. Действительно, в силу условия леммы последовательность $\{ a_k \}_{k=0}^{\infty}$ ограничена Обозначим какую-нибудь ее предельную точку через a^* . Тогда, очевидно,

$$\sup\lim_{k \to \infty} a_k < a^* ,$$

откуда заключаем, что такая точка у последовательности $\{ a_k \}_{k=0}^{\infty}$ единственная. □

Лемма 3.2.2. Пусть $h(t)$ - монотонно убывающая при $t > 0$ функция. Тогда для любых целых чисел p и q $0 <$ $< q < p$, выполняются неравенства

$$\int_q^{p+1} h(t)\, dt < \sum_{k=p}^{q} h(k) < h(q) + \int_q^{p} h(t)\, dt$$

Доказательство очевидно.

Теорема 3.2.1. Если метод (3.2.1), (3.2.3) применяется для решения задачи минимизации квазивыпуклой функции на множестве простой структуры, то:

1) независимо от существования решения задачи (3.1.1)

$$\lim_{k \to \infty} f_k^* = f^* ;$$

2) если множество X^* непусто и либо множество X^* ограничено, либо последовательность $\{ x_k \}_{k=0}^{\infty}$ ограничена, то

$$\lim_{k \to \infty} \rho(x_k , X^*) = 0 \quad , \quad \lim_{k \to \infty} f(x_k) = f^* ;$$

3) если ряд $\sum_{k=0}^{\infty} h_k^2$ сходится и множество X^* не

пусто, то существует точка $x^* \in X^*$ такая, что $x_k \to x^*$ при $k \to \infty$.

Доказательство. 1. Нам, естественно, достаточно доказать, что для любой точки $x' \in R^n$ справедливо неравенство

$$\inf_{k \to \infty} \lim f_k^* \leq f(x'). \quad (3.2.4)$$

Зафиксируем произвольную точку $x' \in Q$. Предположим сначала, что существует положительная константа α, такая, что

$$v'_k \equiv \| g(x_k) \|^{-1} \langle g(x_k), x_k - x' \rangle \geq \alpha$$

при любом $k \geq 0$. Заметим, что найдется номер $N > 0$, начиная с которого $h_k \leq \alpha$ при всех $k \geq N$. Поэтому для всех $k \geq N$

$$\| x_{k+1} - x' \|^2 \leq \| x_k - x' \|^2 -$$
$$- 2 h_k v'_k + h_k^2 \leq \| x_k - x' \|^2 - h_k \alpha,$$

что противоречит расходимости ряда $\sum_{k=1}^{\infty} h_k$. Таким образом, мы показали, что

$$\inf_{k \to \infty} \lim v'^*_k \leq 0$$

Теперь, если найдется номер k такой, что $v'_k \leq 0$, то в силу квазивыпуклости функции $f(x)$ значение $f(x_k)$ не превосходит $f(x')$ и (3.2.4) доказано. Если же $v'_k > 0$ и $f(x_k) > f(x')$ при всех k, то в силу теоремы 1.5.5

$$f(x_k) - f(x') \leq \omega(f, x', v'_k),$$

где функция $\omega(f, x', t)$ непрерывна и $\omega(0) = 0$. Переходя в последнем неравенстве к пределу по подпоследовательности, реализующей нижний предел величин v'_k, получаем неравенство (3.2.4).

2. Пусть множество X^* ограничено. Положим

$\pi^*(x) = \pi(X^*, x)$;

$\upsilon^*(x) = \| g(x) \|^{-1} \langle g(x), x - \pi^*(x) \rangle$;

$\omega^*(t) = \sup \{ f(x) - f^* \mid \| x - \pi^*(x) \| \le t \}$.

В силу компактности множества X^* функция $\omega^*(t)$ определена корректно, причем $\omega^*(t) \to 0$ при $t \to 0$. Кроме того, при всех $x \notin X^*$ выполнено

$$f(x) - f^* \le \omega^*(\upsilon^*(x)). \tag{3.2.5}$$

Рассмотрим множество номеров

$\mathfrak{J} = \{ j \mid \upsilon^*(x_j) \le h_j \}$.

Заметим, что для любого $k \ge 0$

$\rho(x_{k+1}, X^*) \le \| x_{k+1} - \pi^*(x_k) \|^2 =$

$= \rho(X^*, x_k) - 2 h_k \upsilon^*_k + h^2_k$.

Возможны два случая.

А. Множество $\mathfrak{J}$ конечно. Тогда начиная с некоторого номера $N \ge 0$ (если $\mathfrak{J} = \emptyset$, то $N = 0$) последовательность $\{ \rho(X^*, x_k) \}_{k=0}^{\infty}$ монотонно убывает. Из утверждения 1 настоящей теоремы и из теоремы 1.5.2 следует, что у этой последовательности существует сходящаяся к нулю подпоследовательность, а значит, и вся она сходится к нулю.

Б. Множество $\mathfrak{J}$ бесконечно. Тогда для любых двух соседних номеров $k, l \in \mathfrak{J}$ и номера i, $l < i \le k$,

$$\rho^2(X^*, x_l) \le \rho^2(X^*, x_{i+1}) \le$$
$$\le \rho^2(X^*, x_i) \le \rho^2(X^*, x_k) + h^2_k. \tag{3.2.6}$$

Заметим, что $\upsilon^*_k \to 0$ при $k \in \mathfrak{J}$, $k \to \infty$. Поэтому из (3.2.5) заключаем, что $f(x_k) \to f^*$ при $k \in \mathfrak{J}$, $k \to \infty$. Следовательно, в силу теоремы 1.5.2 $\rho(x_k, X^*) \to 0$ при $k \in \mathfrak{J}$, $k \to \infty$. Осталось воспользоваться соотношениями (3.2.6).

Пусть теперь последовательность $\{ x_k \}_{k=0}^{\infty}$ ограничена, т.е. $\{ x_k \}_{k=0}^{\infty} \subset B(0, D)$ с некоторым $D > 0$. Заметим, что в силу утверждения 1 теоремы в этом случае заведомо $f^* > -\infty$, $X^* \neq \emptyset$. Введем функцию

$$\varphi(x) = \max\{ f(x) - f^*,\ \| x \| - D \}.$$

Функция $\varphi(x)$ является квазивыпуклой и ее множество минимумов есть $X^* \cap B(0, D)$. Если для минимизации функции $\varphi(x)$ применить метод (3.2.1),(3.2.3), то он построит ту же самую последовательность точек $\{ x_k \}_{k=0}^{\infty}$, что и при минимизации исходной функции $f(x)$. Но тогда в силу доказанного выше

$$\rho(X^* \cap B(0, D),\ x_k) \to 0$$

при $k \to \infty$, а значит, $\rho(X^*, x_k) \to 0$ при $k \to \infty$.

3. Из условий этого пункта следует, что для любой точки $x^* \in X^*$ при всех $k \geq 0$ выполняется неравенство

$$\| x_{k+1} - x^* \|^2 \leq \| x_k - x^* \|^2 + h_k^2.$$

Поэтому в силу леммы 3.2.1 для любого x^* последовательность $\{ \| x_k - x^* \| \}_{k=0}^{\infty}$ сходится. Но последовательность $\{ x_k \}_{k=0}^{\infty}$ ограничена,и все ее предельные точки принадлежат множеству X^*. Выбирая в качестве x^* одну из таких точек, получаем, что $\| x_k - x^* \| \to 0$ при $k \to \infty$. □

Докажем теорему о скорости сходимости методов (3.2.1),(3.2.2) и (3.2.1),(3.2.3). Положим

$$r = \| x_0 - x^* \|, \quad \omega(t) = \omega(f, x^*, t).$$

Теорема 3.2.2. Справедливы следующие оценки скорости сходимости:

1) при всех $N \geq 0$ для метода (3.2.1),(3.2.2)

$$f_N^* - f^* \leq \omega(0.5\, h + 0.5\, h^{-1} (N+1)^{-1} r^2),$$

для метода (3.2.1),(3.2.3)

$$f_N^* - f^* \leq \omega(0.5\, [\, r^2 + \Sigma_{k=0}^{N} h_k^2 \,] / \Sigma_{k=0}^{N} h_k);$$

2) если в методах (3.2.1),(3.2.2) и (3.2.1),(3.2.3) заранее задано число шагов N, которое разрешено сделать ме-

тоду, то наилучшая оценка погрешности конечного результата достигается на наборе $h_k = r (N + 1)^{-1/2}$, $k = 0, \dots, N$:
$f_N^* - f^* \leq \omega(r (N + 1)^{-1/2})$;

3) если в методе (3.2.1), (3.2.3) ряд $\sum_{k=0}^{\infty} h_k^2$ расходится, то

$$\inf_{k \to \infty} \lim (f_k^* - f^*) [\omega(h_k)]^{-1} \leq 1 .$$

Доказательство. Зафиксируем произвольный номер $N > 0$. Для любого k, $0 \leq k \leq N$, имеем

$$\| x_{k+1} - x^* \|^2 \leq \| x_k - x^* \|^2 -$$
$$- 2 h_k v_k + h_k^2 \leq \| x_k - x' \|^2 - 2 h_k v_k^* + h_k^2 .$$

Просуммировав эти неравенства по k от 0 до N, получим

$$v_N^* \leq 0.5 [r^2 + \sum_{k=0}^{N} h_k^2] / \sum_{k=0}^{N} h_k . \qquad (3.2.7)$$

Из полученного неравенства и следствия 1.5.1 следует справедливость утверждения п.1 настоящей теоремы. Заметим, что условия (3.2.3) являются достаточными условиями стремления к нулю правой части неравенства (3.2.7) при $N \to \infty$.

Если теперь проминимизировать правую часть неравенства (3.2.7) как квазивыпуклую функцию (см.лемму 1.5.1) от $N + 1$ -го переменного $h_0, \dots, h_N$ на положительном ортанте R_+^{N+1}, то получим оптимальные значения

$$h_k^* = r (N + 1)^{-1/2} , \quad k = 0, \dots, N ,$$

откуда сразу следует утверждение 2 теоремы.

Для доказательства справедливости утверждения 3 достаточно установить, что для любого N найдется номер $k = k(N)$ такой, что $v(x_k) \leq h_k$. Действительно, предположим, что это не так. Тогда существует номер N, начиная с которого выполняется неравенство

$$v(x_k) > h_k$$

при $k > N$. Но тогда

$$\| x_{k+1} - x^* \|^2 \leq \| x_k - x^* \|^2 -$$
$$- 2 h_k v_k + h_k^2 \leq \| x_k - x' \|^2 - h_k^2$$
для всех $k \geq N$, что противоречит расходимости ряда $\sum_{k=0}^{N} h_k^2$. □

Покажем на примере утверждения 2 теоремы 3.2.2, как записываются соответствующие оценки скорости сходимости методов (3.2.1),(3.2.2) и (3.2.3),(3.2.5) для некоторых классов целевых функций.

1. Функция $f(x)$ квазивыпукла и липшицева на шаре $B(x_0, r)$ с константой $C > 0$. Тогда
$$f_N^* - f^* \leq C r (N + 1)^{-1/2}.$$

2. Функция $f(x)$ выпукла и ограничена на шаре $B(x_0, r)$ константой $C > 0$. Тогда
$$f_N^* - f^* \leq C (N + 1)^{-1/2}.$$

3. Функция $f(x)$ выпукла и принадлежит классу $C^{1,\nu}(L)$. Тогда
$$f_N^* - f^* \leq (1 + \nu)^{-1} L (N + 1)^{-(1+\nu)/2}.$$
В частности, если $f(x) \in C^{1,1}(L)$, то
$$f_N^* - f^* \leq 0.5 L (N + 1)^{-1}.$$

Сравнивая полученные оценки скорости сходимости субградиентного метода с нижними оценками сложности (см. Приложение), нетрудно убедиться в том, что этот метод при правильном выборе шаговых множителей является оптимальным на классе негладких выпуклых задач в случае, если размерность пространства переменных достаточно велика и допустимое множество ограничено.

В дальнейшем мы не будем подробно расписывать оценки скорости сходимости исследуемых методов для конкретных классов целевых функций, а ограничимся только записью таких оценок через функцию $\omega(f, x^*, t)$.

Приведем простое следствие из теоремы 3.2.2.

Следствие 3.2.1. Если множество X^* имеет внутреннюю точку, то метод (3.2.1),(3.2.2) за конечное число шагов построит точку x_k, содержащуюся в множестве X^*.

Доказательство. Не ограничивая общности, можем считать, что зафиксированная (при определении функции $\omega(t)$) точка $x^* \in \operatorname{int} X^*$. Тогда, как нетрудно убедиться, существует $\varepsilon > 0$, такое, что $\omega(t) \equiv 0$ для всех $t \in [0, \varepsilon]$. Из неравенства (3.2.7) следует, что $v_N^* \to 0$ при $N \to \infty$. Поэтому найдется номер k такой, что $v_k^* < \varepsilon$ и, следовательно, $f_k^* = f^*$. □

Если в методах (3.2.1),(3.2.2) и (3.2.1),(3.2.3) задано число шагов N, а оценка для величины r неизвестна, то, полагая

$$h_k = a(N+1)^{-1/2}, \ k = 0, \dots, N,$$

где a - положительная константа, получаем

$$f_N^* - f^* \leq \omega(\min\{ r, 0.5 [r^2 + a^2] a^{-1} (N+1)^{-1/2} \}),$$

т.е. порядок скорости сходимости этих методов при такой стратегии выбора шаговых множителей сохраняется. Если же число шагов в методе (3.2.1),(3.2.3) заранее не задается, то можно выбрать

$$h_k = a(k+1)^{-1/2}, \ k \geq 0.$$

В этом случае в силу леммы 3.2.2

$$v_N^* \leq 0.5 [r^2 + \textstyle\sum_{k=0}^N h_k^2] / \sum_{k=0}^N h_k =$$

$$= 0.5 a^{-1} [r^2 + a^2(1 + \ln(N+1))] (N+1)^{-1/2},$$

то есть оценка скорости сходимости метода (3.2.1), (3.2.3) несколько ухудшится. Однако при некоторых дополнительных предположениях можно доказать оптимальную оценку скорости сходимости метода (3.2.1),(3.2.3) и при таком способе выбора шаговых множителей.

Теорема 3.2.3. Пусть последовательность точек $\{x_k\}_{k=0}^{\infty}$, вырабатываемая методом (3.2.1),(3.2.3) с $h_k = a(k+1)^{-1/2}$, $k \geq 0$, ограничена:

$\{x_k\}_{k=0}^{\infty} \subset B(x^*, D)$, $D > 0$.

Тогда при любом $N \geq 0$ справедлива оценка

$f_N^* - f^* \leq \omega(\min\{r, C(N+2)^{-1/2}\})$,

где $C = 0.5a^{-1}(D^2 + a^2(1 + (1 + 2^{-1.5})\ln 3))$

Доказательство. Зафиксируем номер $N \geq 1$. Для любого k, $0 \leq k \leq N$, можно записать

$\| x_{k+1} - x^* \|^2 \leq \| x_k - x^* \|^2 - 2h_k.v_N^* + h_k^2$

Просуммируем эти неравенства по $k = l, \ldots, N$, где $0 \leq l \leq N$. Получим

$v_N^* \leq 0.5[\| x_l - x^* \|^2 + \Sigma_{k=l}^N h_k^2] / \Sigma_{k=l}^N h_k =$

Заметим, что в силу леммы 3.2.2

$$\Sigma_{k=l}^N h_k^2 \leq 1 + \int_l^N (t+1)^{-1}\,dt = 1 + \ln[(N+1)/(l+1)].$$

$$\Sigma_{k=l}^N h_k \geq 2((N+2)^{1/2} - (l+1)^{1/2})$$

Выберем число $l =]0.5(N+1)[- 1$, где $]a[$ - целая часть числа a. Тогда

$2((N+2)^{1/2} - (l+1)^{1/2}) \geq$

$\geq (2 - 2^{1/2})(N+2)^{1/2}$,

$N + 1 \leq 3(l+1)$.

Итак, окончательно

$v_N^* \leq (1 + 2^{-1/2})a^{-1}(D^2 + a(1 + \ln 3))(N+2)^{-1/2}$

Осталось воспользоваться следствием 1.5.1. □

Заметим, что последовательность $\{x_k\}_{k=0}^{\infty}$ будет ограниченной, если, например, ограничено множество Q.

Если известна дополнительная информация о целевой функции $f(x)$, то в методе (3.2.1),(3.2.3) можно использовать и другие стратегии выбора шагов h_k.

Пусть, например, мы знаем оптимальное значение функции f^* и некоторую непрерывную монотонную функцию $\eta(t)$, такую, что $\eta(t) \to 0$ при $t \to 0$ и $\omega(f, x^*, t) \leq \eta(t)$ при всех $t \in [0, 1]$. Будем выбирать шаг h_k как решение следующего уравнения:

$$f(x_k) - f^* = \eta(h_k) \tag{3.2.8}$$

Так, например, если функция $f(x)$ липшицева на шаре $B(x^*, r)$ с константой $C > 0$, из уравнения (3.2.8) шаг h_k определяется следующим образом:

$$h_k = C^{-1}(f(x_k) - f^*).$$

Если функция $f(x) \in C^{1,1}(L)$, то стратегия выбора шага h_k будет такой:

$$h_k = [2L^{-1}(f(x_k) - f^*)]^{1/2}$$

Докажем теорему о скорости сходимости метода (3.2.1), (3.2.8).

<u>Теорема 3.2.4.</u> Если последовательность точек $\{x_k\}_{k=0}^{\infty}$ вырабатывается методом (3.2.1),(3.2.8), то

1) $\lim\limits_{k \to \infty} f(x_k) = f^*$, $\lim\limits_{k \to \infty} \rho(x_k, X^*) = 0$;

2) для всех $k \geq 0$ справедливо неравенство

$$f_k^* - f^* \leq \eta(r(k+1)^{-1/2})$$

Д о к а з а т е л ь с т в о. Из уравнения (3.2.8) следует, что

$$\eta(h_k) = f(x_k) - f^* \leq \omega(v_k) \leq \eta(v_k),$$

откуда $h_k \leq v_k$ при всех $k \geq 0$. Поэтому для любого $k \geq 0$ справедливо неравенство

$$\| x_{k+1} - x^* \|^2 \leq \| x_k - x^* \|^2 - 2h_k v_k + h_k^2 \leq$$
$$\leq \| x_k - x^* \|^2 - h_k^2$$

Следовательно, последовательность $\{x_k\}_{k=0}^{\infty}$ ограничена и

$\sum_{k=0}^{\infty} h_k^2 \le r^2$. Отсюда получаем, что $h_k \to 0$ при $k \to \infty$, $h_k^* \le r(k+1)^{-1/2}$ □

В заключение рассмотрим модификацию субградиентного метода (3.2.1),(3.2.3), предназначенную для решения задачи минимизации квазивыпуклой функции при наличии функциональных ограничений.

Применим для решения задачи (3.1.1) следующий метод:

0. Выбираем начальную точку $x_0 \in R^n$ и параметр точности ε

1. k -я итерация ($k \ge 0$). (3.2.9)

Полагаем

$x_{k+1} = \pi(Q, x_k - h_k s_k / \| s_k \|)$,

где

$$s_k = \begin{cases} g(x_k), & \text{если } f^{(j)}(x_k) \le \varepsilon \\ & \quad \text{для всех } j,\ 1 \le j \le m, \\ g^{(j)}(x_k), & \text{если найдется } j \equiv j(k), \\ & \quad \text{такой, что } f^{(j)}(x_k) > \varepsilon \end{cases}$$

Итерация закончена.

Теорема 3.2.5. Если в методе (3.2.9) последовательность шагов $\{h_k\}$ удовлетворяет условию (3.2.3) и на некотором шаге с номером N выполнились все неравенства

$\omega(f, x^*, t_N) < \varepsilon$,

$\omega(f^{(j)}, x^*, t_N) < \varepsilon$, $j = 1, \ldots, m$,

где

$t_N = 0.5 [r^2 + \sum_{k=0}^{N} h_k^2] / \sum_{k=0}^{N} h_k$.

то найдется номер k, $0 \le k \le N$, такой, что

$f(x_k) - f^* \le \varepsilon$, $f^{(j)}(x_k) \le \varepsilon$, $j = 1, \ldots, m$.

Доказательство. Пусть условия теоремы выполнились на шаге с номером N. Положим

$I = \{ i \mid f^{(j)}(x_i) \le \varepsilon, j = 1, \ldots, m \}$,

$J = \{0, \ldots, N\} \setminus I$,

$v_i = \langle g(x_i), x_i - x^* \rangle \| g(x_i) \|^{-1}$,

$u_i = \langle l^{(j(i))}(x_i), x_i - x^* \rangle \| l^{(j(i))}(x_i) \|^{-1}$.

Тогда, если $i \in I$, то

$$\| x_{i+1} - x^* \|^2 \le \| x_i - x^* \|^2 - 2 h_i v_i + h_i^2 .$$

Если же $i \in \mathfrak{J}$, то

$$\| x_{i+1} - x^* \|^2 \le \| x_i - x^* \|^2 - 2 h_i u_i + h_i^2 .$$

Складывая эти неравенства по i от 1 до k, получаем

$$\textstyle\sum_{i \in I} h_i v_i + \sum_{i \in \mathfrak{J}} h_i u_i \le 0.5 [r^2 + \sum_{i=0}^{N} h_i^2] . \qquad (3.2.10)$$

Заметим, что в силу теоремы 1.5.5 и способа определения множества $\mathfrak{J}$ при всех $i \in \mathfrak{J}$ справедливо неравенство

$$\varepsilon \le f^{(j(i))}(x_i) \le f^{(j(i))}(x_i) - f^{(j(i))}(x^*) \le \omega(f^{(j(i))}, x^*, u_i) . \qquad (3.2.11)$$

Далее, если найдется номер $k \in I$ такой, что $f(x_k) \le f^*$, то утверждение теоремы очевидно. Если же таких номеров нет, то для всех $i \in \mathfrak{J}$ справедливо $f(x_i) > f^*$, и, следовательно, в силу теоремы 1.5.5

$$f(x_i) - f^* \le \omega(f, x^*, v_i) . \qquad (3.2.12)$$

Из условий на функциональные компоненты задачи (3.1.1) вытекает, что функции $\omega(f, x^*, t)$, $\omega(f^{(j)}, x^*, t)$, $j = 1, \ldots, m$, монотонно возрастают по t. Поэтому у них существуют монотонно возрастающие обратные функции

$$\omega^{-1}(\cdot, \cdot, t) \equiv æ(\cdot, \cdot, t) .$$

Подставляя теперь неравенства (3.2.11), (3.2.12), записанные через обратные функции, в (3.2.10), получаем

$$t_N \textstyle\sum_{i=0}^{N} h_i \ge \sum_{i \in I} h_i\, æ(f, x^*, f(x_i) - f^*) + \sum_{i \in \mathfrak{J}} h_i\, æ(f^{(j(i))}, x^*, \varepsilon) . \qquad (3.2.13)$$

Отсюда в силу предположений теоремы следует непустота множества I. Кроме того, в силу тех же предположений

$$t_N \textstyle\sum_{i=0}^{N} h_i \ge \sum_{i \in I} h_i\, æ(f, x^*, f(x_i) - f^*) +$$

$+ \sum_{i \in J} h_i t_N$.

т.е.

$$t_N \sum_{i \in I} h_i > \sum_{i \in I} h_i \varkappa(f , x^* , f(x_i) - f^*) ,$$

откуда заключаем, что найдется номер $k \in I$ такой, что

$$\varkappa(f , x^* , f(x_k) - f^*) < t_N$$

и, значит,

$$f(x_k) - f^* < \omega(f , x^* , t_N) < \varepsilon . \qquad \square$$

Нетрудно убедиться в том, что в методе (3.2.9) оптимальная стратегия выбора шаговых множителей $\{ h_k \}_{k=0}^{\infty}$ такая же, как и в методе (3.2.1),(3.2.3). По той же схеме, что и в случае простых множеств, для метода (3.2.9) можно получить соответствующие оценки его скорости сходимости при различных предположениях о свойствах функциональных компонент задачи (3.1.1).

3.3. Общая схема методов отсечений. Метод центров тяжести

Методы отсечений предназначены для решения задачи условной минимизации квазивыпуклых функций вида (3.1.1). Сходимость этих методов и оценка скорости сходимости являются следствием стремления к нулю объема некоторых тел в R^n, локализующих оптимальное решение исходной задачи. Само по себе наличие выпуклого тела малого объема, содержащего точку минимума, не дает, конечно, никаких оснований для вывода о близости значений функции в точках этого тела к оптимальному Однако ситуация меняется, если локализующее тело есть пересечение полупространств вида

$\{ x \mid \langle g(x_k), x_k - x \rangle > 0 \}$

Действительно, пусть в пространстве R^n задана квазивыпуклая функция $f(x)$ и последовательность точек $\{ x_k \}_{k=0}^N$. Рассмотрим множество

$$S = \{ x \mid \langle g(x_k), x_k - x \rangle > 0, k = 0, \ldots, N \}.$$

Нетрудно видеть, что $x^* \in S$ и величина

$$v^* = \min \{ \| g(x_k) \|^{-1} \langle g(x_k), x_k - x^* \rangle \mid k = 0, \ldots, N \}$$

есть радиус шара максимального объема с центром в точке x^*, содержащегося в теле S. В силу следствия 1.5.1 с помощью величины v^* можно оценить невязку $f_N^* - f^*$. В то же время величина $(v^*)^n$ есть с точностью до константы объем максимального шара с центром в x^*, содержащегося в теле S. В свою очередь, если тело S ограничено, то объем такого шара не превосходит как объема самого тела S, так и объема любого эллипсоида, содержащего тело S. Описываемые в этой главе методы отсечений в процессе работы видоизменяют тело S таким образом, чтобы одна из перечисленных его характеристик монотонно убывала. Приведем общую схему методов отсечений, в которую укладываются все эти методы.

Рассмотрим произвольную последовательность точек $\{ x_k \}_{k=0}^{\infty}$, $x_k \in X \setminus X^*$. Пусть множество X ограничено. Обозначим

$$S_N = \{ x \in X \mid \langle g(x_k), x_k - x \rangle > 0, k = 0, \ldots, N \},$$

$$v_k = \| g(x_k) \|^{-1} \langle g(x_k), x_k - x^* \rangle .$$

<u>Лемма 3.3.1.</u> Для любого $N > 0$ справедлива оценка

$$v_N^* < r \min \{ 1, C_1 [V(S_N)]^{1/n} \}, \tag{3.3.1}$$

где $r = \| x_0 - x^* \|$; $C_1 = V(B(x^*, r) \cap X)^{-1/n}$; $V(\ )$ - объем соответствующего тела в R^n.

Д о к а з а т е л ь с т в о. Легко проверяется, что при любых $\alpha \in [0, 1]$ выполняется включение

$$\Omega(\alpha) = \{ z \mid z = x^* + \alpha(y - x^*),\ y \in B(x^*, r) \cap X \} \subseteq B(x^*, \alpha r) \cap X$$

Поэтому

$$V(B(x^*, \alpha r) \cap X) \geq V(\Omega(\alpha)) = \\ = \alpha^n V(B(x^*, r) \cap X) \qquad (3.3.2)$$

Заметим, что

$$v_N^* \leq r, \quad B(x^*, v_N^*) \cap X \subseteq S_N$$

Выберем $\alpha_0 = r^{-1} v_N^*$. Теперь, подставив в (3.3.2) значение $\alpha = \alpha_0$, получаем неравенство (3.3.1). □

Отметим, что множества S_k в силу своего определения всегда содержат множество X^*.

Если решается задача выпуклого программирования, то при построении последовательности множеств, локализующих множество X^*, можно производить и более глубокие отсечения.

Действительно, пусть

$$U_N = \{ x \in X \mid \langle g(x_k), x_k - x \rangle \geq f_k - f_k^*,\ k = 0, \dots, N \},$$

$$u_k = \| g(x_k) \|^{-1} [\langle g(x_k), x_k - x^* \rangle - f_k + f_k^*].$$

Справедлива следующая лемма

<u>Лемма 3.3.2.</u> Для любого $N \geq 0$

$$u_N^* \leq r \min \{ 1, C_1 [V(U_N)]^{1/n} \},$$

где величины r и C_1 те же, что и в лемме 3.3.1.

Доказательство этой леммы аналогично доказательству леммы 3.3.1, и мы его приводить не будем. Напомним, что связь между величинами u_N^* и f_N^* устанавливается теоремой 1.5.6.

Если для задачи выпуклого программирования известно оптимальное значение функции f^*, то последовательность локализующих множеств может быть задана еще одним способом.

Зафиксируем произвольное $\alpha \in [0, 1]$ и положим

$U_N^*(\alpha) = \{ x \in X \mid \langle g(x_k), x_k - x \rangle \geq$
$\geq \alpha (f_k - f^*), k = 0, \ldots, N \}$
$u_k^* = \| g(x_k) \|^{-1} [\langle g(x_k), x_k - x^* \rangle - \alpha (f_k - f^*)]$

<u>Лемма 3.3.3</u> Для любого $N \geq 0$

$f_N^* - f^* \leq$
$\leq (1 - \alpha)^{-1} \omega(f, x^*, r \min \{ 1, C_1 [V(U_N^*(\alpha))]^{1/n} \})$,
где величины r и C_1 определены в лемме 3.3.1

Эта лемма доказывается точно так же, как и лемма 3.3.1. При выводе окончательного неравенства в лемме 3.3.3 надо воспользоваться следствием 1.5.2.

Таким образом, в методах отсечений необходимо выбирать очередные точки последовательности $\{ x_k \}_{k=0}^{\infty}$ так, чтобы объем множеств S_k (в квазивыпуклом случае) либо объемы множеств U_k, $U_k^*(\alpha)$ стремились к нулю. При этом, как следует из вышеприведенных лемм, скорость убывания объемов локализующих множеств определяет скорость сходимости соответствующего метода.

Приведем схему самого первого из предложенных методов отсечений.

Метод центров тяжести. (3.3.3)

0. Полагаем $S_0 = X$.

1. k-я итерация ($k \geq 0$)

   а) Вычисляем точку x_k - центр тяжести множества S_k;

   б) полагаем

$S_{k+1} = \{ x \in S_k \mid \langle g(x_k), x_k - x \rangle \geq 0 \}$

Итерация закончена.

Основным результатом, на который опирается обоснование оценки скорости сходимости метода центров тяжести, является следующая геометрическая лемма, которую мы приведем без доказательства.

Лемма 3.3.4. Пусть Q - выпуклое замкнутое ограниченное тело в R^n с центром тяжести z, g - произвольный ненулевой вектор, $Q_+ = \{ x \in Q \mid \langle g, z - x \rangle \geq 0 \}$. Тогда
$V(Q_+) \leq (1 - [n(n+1)^{-1}]^n) V(Q) \leq (1 - e^{-1}) V(Q)$.

Теорема 3.3.1. Если множество X в задаче (3.1.1) есть выпуклое замкнутое ограниченное тело, а последовательность $\{x_k\}_{k=0}^{\infty}$ построена методом (3.3.3), то при всех $k \geq 0$ справедлива оценка

$$f_k^* - f^* \leq \omega(f, x^*, \min\{ r, R(1 - e^{-1})^{(k+1)/n} \}),$$

где $r = \| x_0 - x^* \|$, $R = \min\{ t \mid X \subseteq B(x^*, t) \}$.

Доказательство. Действительно, из леммы 3.3.4 и способа построения множеств S_k следует, что

$$V(S_k) \leq (1 - e^{-1})^k V(X)$$

Поэтому в силу леммы 3.3.1

$$v_k^* \leq \min\{ r, r[V(S_{k+1}) / V(B(x^*, r) \cap X)]^{1/n} \}.$$

Осталось заметить, что функция

$$\xi(r) = r^n V(B(x^*, r) \cap X)^{-1}$$

является неубывающей функцией от r (это было показано при доказательстве леммы 3.3.1). Поэтому

$$\xi(r) \leq \xi(R) = R^n V(X)^{-1}. \quad \square$$

Следствие 3.3.1. Метод центров тяжести является оптимальным методом на классе липшицевых квазивыпуклых экстремальных задач.

Для доказательства этого утверждения достаточно сопоставить оценку скорости сходимости метода центров тяжести с нижними оценками сложности класса липшицевых квазивыпуклых задач, приведенными в Приложении.

Несмотря на свою теоретическую оптимальность, метод центров тяжести никак не может быть отнесен к практически реализуемым алгоритмам. Дело в том, что вспомогательная задача по отысканию центра тяжести даже самых простых выпуклых множеств в пространстве R^n чрезвычайно сложна Так, например, известно, что задача отыскания центра тяжести пересечения n - мерного куба с гиперплоскостью есть NP -полная задача (для решения таких задач пока не найдено полиномиальных алгоритмов) В связи с этим представляют интерес методы, которые хотя и уступают по оценке скорости сходимости оптимальным алгоритмам, но зато имеют существенно более низкую стоимость реализации одной итерации. К таким алгоритмам относится метод описанных эллипсоидов, который мы рассмотрим в следующем параграфе.

3.4. Метод описанных эллипсоидов

Изложим идею метода описанных эллипсоидов. Пусть на k -м шаге имеется эллипсоид, содержащий множество S_{k-1}. Возможны два случая. Если центр эллипсоида не содержится в допустимом множестве X, то, вычислив в этой точке линейный функционал, отделяющий ее от множества X (можно, например, взять субградиент любого нарушенного ограничения), мы сможем отбросить половину эллипсоида, не содержащую точек множества X, а значит, и множество S_{k-1}. Если же центр эллипсоида принадлежит множеству X, то, вычислив в этой точке субградиент целевой функции, мы сможем отбросить половину эллипсоида, не содержащую множество S_k. В обоих случаях, описывая вокруг оставшейся половины эллипсоид минимального объема, мы получаем новый эллипсоид, объемом меньший, чем исходный, который содержит множество S_k (либо S_{k-1}, если центр старого эллипсоида не принадлежал множеству X). Можно показать, что объем таких эллипсоидов стремится к нулю. А значит, стремится к нулю и объем множеств S_k, что влечет за собой в силу леммы 3.3.1 сходимость метода.

В выпуклом случае аналогичным способом можно строить последовательность эллипсоидов, содержащих множества U_k либо, если известно оптимальное значение целевой функции, последовательность эллипсоидов, содержащих множества $U_k^*(\alpha)$.

Прежде чем привести формальное описание метода описанных эллипсоидов, докажем одну геометрическую лемму.

Лемма 3.4.1. Пусть $G = G^T$ - положительно определенная $n \times n$ -матрица, $H = G^{-1}$, $A = \{ x \in R^n \mid \langle G x, x \rangle \le \lambda^2 \}$, $\lambda > 0$. Предположим, что величина α и вектор g связаны соотношением

$$-n^{-1} \langle H g, g \rangle^{1/2} \lambda \le \alpha \le \langle H g, g \rangle^{1/2} \lambda$$

Тогда множество $A_+ = \{ x \in A \mid \langle g, x \rangle \ge \alpha \}$ содержится в эллипсоиде

$$A^* = \{ x \in R^n \mid \langle G^* (x - y), x - y \rangle \le (\lambda^*)^2 \},$$

где

$$G^* = G + \frac{2(\lambda + n \xi)}{(n - 1)(\lambda - \xi)} \cdot \frac{g g^T}{\langle H g, g \rangle},$$

$$y = \frac{\lambda + n \xi}{n + 1} \cdot \frac{H g}{\langle H g, g \rangle^{1/2}}, \quad \lambda^* = n \left[\frac{\lambda^2 - \xi^2}{n^2 - 1} \right]^{1/2},$$

$$\xi = \alpha \langle H g, g \rangle^{1/2}$$

Объем эллипсоида A^* вычисляется по фурмуле

$$V(A^*) = \left\{ \frac{n}{\lambda} \left[\frac{\lambda^2 - \xi^2}{n^2 - 1} \right]^{1/2} \right\}^n \left[\frac{(n - 1)(\lambda - \xi)}{(n + 1)(\lambda + \xi)} \right] V(A)$$

Поясним геометрический смысл леммы. В ней приводятся формулы построения эллипсоида A^*, задаваемого матрицей G^* и правой частью λ^*, с центром в точке y. Эллипсоид A^*

является эллипсоидом минимального объема, содержащим множество A_+ - часть исходного эллипсоида A, центр которого находился в нуле и который был задан матрицей G и правой частью λ. Соотношения, связывающие величины H, g, λ и α, обеспечивают непустоту множества A_+ и отличие нового центра y от нуля.

Д о к а з а т е л ь с т в о. Зафиксируем произвольный вектор x из A_+. Из определения матрицы G^* и вектора y получаем

$$G^* y = \frac{\lambda + n\xi}{n+1}\left(1 + \frac{2(\lambda + n\xi)}{(n-1)(\lambda - \xi)}\right)\frac{g}{\langle Hg, g\rangle^{1/2}} =$$

$$= \frac{(\lambda + n\xi)(\lambda + \xi)}{(n-1)(\lambda - \xi)} \cdot \frac{g}{\langle Hg, g\rangle^{1/2}} .$$

$$\langle G^*(x - y), x - y\rangle = \langle G^* x, x\rangle - 2\langle G^* y, x\rangle +$$

$$+ \langle G^* y, y\rangle = \langle Gx, x\rangle +$$

$$+ \frac{2(\lambda + n\xi)}{(n-1)(\lambda - \xi)} \cdot \frac{\langle g, x\rangle^2}{\langle Hg, g\rangle} - 2\frac{(\lambda + n\xi)(\lambda + \xi)}{(n-1)(\lambda - \xi)} \times$$

$$\times \frac{\langle g, x\rangle}{\langle Hg, g\rangle^{1/2}} + \frac{(\lambda + n\xi)^2(\lambda + \xi)}{(n^2-1)(\lambda - \xi)} = \langle Gx, x\rangle +$$

$$+ \frac{2(\lambda + n\xi)}{(n-1)(\lambda - \xi)} \cdot \left[\frac{\langle g, x\rangle}{\langle Hg, g\rangle^{1/2}} - \lambda\right] \times$$

$$\times \left[\frac{\langle g, x\rangle}{\langle Hg, g\rangle^{1/2}} - \xi\right] + \frac{\lambda^2 - n^2\xi^2}{n^2 - 1}$$

Заметим, что

$$\xi \le \langle g, x\rangle \langle Hg, g\rangle^{-1/2} \le \langle Gx, x\rangle^{1/2} \le \lambda$$

Следовательно,

$$\langle G^*(x - y), x - y\rangle \le \langle Gx, x\rangle + (n^2 - 1)^{-1} \times$$

$$\times (\lambda^2 - n^2\xi^2) \le n^2(n^2-1)^{-1}(\lambda^2 - \xi^2) = (\lambda^*)^2 .$$

Докажем справедливость формулы вычисления объема эллипсоида A^*. Действительно,

$$\det G^* = \frac{(n+1)(\lambda + \xi)}{(n-1)(\quad - \xi)} \det G .$$

Поэтому

$$V(A^*) / V(A) = (\lambda^* / \lambda)^n [\det G / \det G^*]^{1/2} =$$

$$= \{ \frac{n}{\lambda} [\frac{\lambda^2 - \xi^2}{n^2 - 1}]^{1/2} \}^n [\frac{(n-1)(\lambda - \xi)}{(n+1)(\lambda + \xi)}] . \quad \square$$

Приведем общую схему метода описанных эллипсоидов. Конкретные схемы этого метода, обеспечивающие уменьшение объемов множеств S_k, U_k либо $U^*_k(\alpha)$, могут быть получены из общей схемы с применением соответствующей стратегии выбора параметров α_k. Нетрудно заметить, что формулы пересчета центров эллипсоидов, матриц и правых частей совпадают с формулами леммы 3.4.1 за тем лишь исключением, что вместо матриц G пересчитываются обратные ι ним матрицы H.

Метод описанных эллипсоидов (3.4.1)

0. Выбираем точку $x_0 \in X$ и такое число $R > 0$, что $\| x_0 - x^* \| < R$;

полагаем $y_0 = x_0$, $s_0 = g(x_0)$, $\lambda_0 = R$.

1. k -я итерация ($k > 1$).

а) Выбираем параметр $\alpha_k > 0$;

б) полагаем

$$\xi_k = \alpha_k \langle H_k s_k , s_k \rangle^{-1/2} ,$$

$$y_{k+1} = \frac{\lambda_k + n \xi_k}{n+1} \cdot \frac{H_k s_k}{\langle H_k s_k , s_k \rangle^{1/2}} ,$$

$$H_{k+1} = H_k - \frac{2(\lambda_k + n \xi_k)}{(n+1)(\lambda_k + \xi_k)} \cdot \frac{H_k s_k s_k^T H_k}{\langle H_k s_k , s_k \rangle} ,$$

$$\lambda_{k+1} = n [(\lambda_k^2 - \xi_k^2) / (n^2 - 1)]^{1/2} ;$$

в) если $y_{k+1} \in X \cap B(x_0 , R)$, то полагаем $x_{k+1} = y_{k+1}$, $s_{k+1} = g(x_{k+1})$;

если $y_{k+1} \notin B(x_0 , R)$, то полагаем

$x_{k+1} = x_k$, $s_{k+1} = (y_{k+1} - x_0)/\| y_{k+1} - x_0 \|$,

если $y_{k+1} \in B(x_0, R) \setminus X$, то полагаем
$x_{k+1} = x_k$, $s_{k+1} = l^{(i)}(y_{k+1})$,
где i - номер любого из нарушенных ограничений в задаче (3.1.1).

Итерация закончена.

Остановимся на стратегиях выбора параметров α_k в методе (3.4.1).

1. Если $y_k \in X \cap B(x_0, R)$, то параметр α_k может быть выбран следующим образом:

$\alpha_k = 0$, если функция $f(x)$ квазивыпукла. В этом случае построенные эллипсоиды содержат множества S_k;

$\alpha_k = f_k - f_k^*$, если функция $f(x)$ выпукла. В этом случае эллипсоиды содержат множества U_k;

$\alpha_k = \alpha(f_k - f^*)$, $\alpha = \text{const} \in [0, 1)$, если функция $f(x)$ выпукла и известно оптимальное значение f^*. В этом случае эллипсоиды содержат множества $U_k^*(\alpha)$;

2. Если $y_k \notin B(x_0, R)$, то выбираем
$\alpha_k = \| y_k - x_0 \| - R$.

3. Если $y_k \in B(x_0, R) \setminus X$ и $s_k = l^{(i)}(y_k)$, то выбираем

$$\alpha_k = \begin{cases} 0, & \text{если функция } f^{(i)}(x) \text{ квазивыпукла}, \\ f^{(i)}(y_k), & \text{если функция } f^{(i)}(x) \text{ выпукла}. \end{cases}$$

Докажем теорему о скорости сходимости метода (3.4.1) для задачи квазивыпуклого программирования ($\alpha_k \equiv 0$).

Теорема 3.4.1. Если последовательность точек $\{x_k\}_{k=0}^{\infty}$ строится по методу (3.4.1) с $\alpha_k = 0$ и $\text{int}\, X \neq \emptyset$, то для всех $k > 0$ справедлива оценка

$f_k^* - f^* < \omega(f, x^*, \min\{r, [C_1 q(n)^k]^{1/n}\})$, где

$r = \| x_0 - x^* \|$,

$C_1 = 0.5\, r\, V(\, B(\, x_0 \,,\, R\,)\,) \,/\, V(\, B(\, x_0 \,,\, r\,) \cap X\,)$,

$$q(\, n\,)^2 = [\, \frac{n^2}{n^2 - 1}\,] \cdot \frac{n - 1}{n + 1} < 1 \,.$$

Д о к а з а т е л ь с т в о. Заметим, что метод (3.4.1) фактически решает задачу минимизации функции $f(\, x\,)$ на множестве $X \cap B(\, x_0 \,,\, R\,)$. Поэтому можно считать, что $X = X \cap B(\, x_0 \,,\, R\,)$.

Рассмотрим последовательность матриц $\{\, G_k \,\}_{k=0}^{\infty}$:

$$G_0 = I \,, \quad G_{k+1} = G_k + \frac{2}{n - 1} \cdot \frac{s_k\, s_k^T}{\langle\, H_k\, s_k \,,\, s_k \,\rangle} \,, \tag{3.4.2}$$

$k = 0,\, 1,\, \ldots$.

Нетрудно убедиться в том, что при любом k матрица G_k является обратной к матрице H_k , построеннной методом (3.4.1). Заметим также, что формула прересчета (3.4.2) совпадает с формулой пересчета матриц из леммы 3.4.1.

Покажем, что при любом $k \geq 0$ выполняется включение

$$S_k \subseteq A_k = \{\, x \in R^n \mid \langle\, G_k\, (\, x - y_k\,) \,,\, x - y_k \,\rangle \leq \lambda_k^2 \,\} \,. \tag{3.4.3}$$

Доказательство будем вести по индукции. При $k = 0$ это включение выполнено. Пусть оно выполнено и при некотором $k \geq 0$. Возможны два случая.

1. $y_{k+1} \in X$. Тогда $x_{k+1} = y_{k+1}$ и, пользуясь леммой 3.4.1, получаем

$$S_{k+1} = \{\, x \in S_k \mid \langle\, g(\, y_{k+1}\,) \,,\, y_{k+1} - x \,\rangle \geq 0 \,\} \subseteq$$
$$\subseteq \{\, x \in A_k \mid \langle\, g(\, y_{k+1}\,) \,,\, y_{k+1} - x \,\rangle \geq 0 \,\} \subseteq A_{k+1} \,.$$

2. $y_{k+1} \notin X$. Тогда $x_{k+1} = x_k$, $s_{k+1} = l^{(i)}(\, y_{k+1}\,)$ для некоторого i , $1 \leq i \leq m$. Поэтому

$S_{k+1} = S_k = S_k \cap X \subseteq A_k \cap X \subseteq \{\, x \in R^n \mid$

$f^{(i)}(x) < 0 \} \subseteq A_k \cap$

$\cap \{ x \in R^n \mid \langle l^{(i)}(y_{k+1}), y_{k+1} - x \rangle > 0 \} \subseteq A_{k+1}.$

Итак, включение (3.4.3) выполнено при всех $k > 0$. Из леммы 3.4.1 следует, что

$V(A_k) = q(n)^k V(B(x_0, R))$.

Далее, из способа построения множеств S_k ясно, что

$V(S_k) = 0.5 V(A_k)$.

Осталось воспользоваться леммой 3.1.1. □

Обоснование оценки скорости сходимости других версий метода описанных эллипсоидов производится аналогично. Необходимо лишь на заключительной стадии доказательства воспользоваться леммой 3.3.2, если строящиеся эллипсоиды содержат множества U_k, и леммой 3.3.3, если строящиеся эллипсоиды содержат множества $U_k^*(\alpha)$.

Заметим, что показатель $q(n)$ в оценке скорости сходимости метода описанных эллипсоидов при больших n ведет себя как $1 - 0.5\, n^{-2}$. Поэтому метод описанных эллипсоидов, в отличие от метода центров тяжести, не является оптимальным методом решения квазивыпуклых задач. Однако трудоемкость одной итерации у этого метода довольно мала - порядка $O(n^2)$ арифметических итераций.

3.5. Метод растяжения пространства

В этом параграфе будет рассмотрен метод безусловной минимизации негладких функций, удовлетворяющих специальному условию согласованности роста значения функции относительно минимума и производной по направлению. Примером таких функций являются гладкие квазиоднородные функции (см. §1.6). Для простоты изложения ограничимся только выпуклыми функциями.

Итак, пусть для выпуклой функции $f(x)$ найдется точка $x^* \in X^*$ и константы $M \ge N \ge 1$, такие, что для всех x из R^n выполняется неравенство

$$N (f(x) - f^*) \le \langle g(x), x - x^* \rangle \le M (f(x) - f^*), \tag{3.5.1}$$

где $g(x)$ - произвольный вектор из $\partial f(x)$. Например, для однородных функций в неравенстве (3.5.1) $M = N = p$, где p - степень однородности, для квадратичных функций $M = N = 2$, для квазиоднородной функции $f(x)$ можно взять $M = p(f)^{-1}$, $N = (1 - p(f))^{-1}$.

Заметим, что константы M, N в неравенстве (3.5.1) не зависят от обусловленности функции $f(x)$ и не изменяются при линейном преобразовании пространства переменных. Для выпуклых функций всегда $N \ge 1$

<u>Лемма 3.5.1.</u> Если функция $f(x)$ обладает свойством (3.5.1), то при любом $x \in R^n$ и $\alpha \in [0, 1]$ выполняется неравенство

$$\alpha^M (f(x) - f^*) \le f(x^* + \alpha(x - x^*)) \le$$
$$\le \alpha^N (f(x) - f^*). \tag{3.5.2}$$

Д о к а з а т е л ь с т в о. Докажем правую часть неравенства (3.5.2). Рассмотрим функцию

$$\xi(\alpha) = \alpha^{-N} [f(x^* + \alpha(x - x^*)) - f^*)].$$

Эта функция непрерывна при $\alpha \in [0, 1]$ и дифференцируема по направлению. Воспользуемся неравенством (3.5.1) и оценим ее правую производную

$$\xi'_+(\alpha) = \alpha^{-2N} [\alpha^N \max \{ \langle g, x - x^* \rangle \mid$$
$$g \in \partial f(x^* + \alpha(x - x^*)) \} -$$
$$- N \alpha^{-N-1} (f(x^* + \alpha(x - x^*)) - f^*] \ge$$

$\geq \alpha^{-N-1} [\alpha \langle g(x^* + \alpha(x - x^*)) , x - x^* \rangle -$
$- N (f(x^* + \alpha(x - x^*)) - f^*] \geq 0.$

Таким образом, функция $\xi(\alpha)$ возрастает при $\alpha \in (0, 1)$ и

$\xi(\alpha) = \alpha^{-N} [f(x^* + \alpha(x - x^*)) - f^*)] \leq$
$\leq \xi(1) = f(x) - f^*.$ □

Следствие 3.5.1. Для всех x, $\| x - x^* \| \leq r$, выполняется неравенство

$$f(x) - f^* \leq r^N \omega(f, x^*, r) \| x - x^* \|^N. \quad (3.5.3)$$

Пусть для функции $f(x)$, удовлетворяющей условию (3.5.1), известны константы M, N и минимальное значение функции f^*. Рассмотрим следующий метод.

Метод растяжения пространства . (3.5.4)

0. Выбираем точку $x_0 \in R^n$. Полагаем $H_0 = I$ - единичная $n \times n$ -матрица, $\gamma = (M + N) / (M - N)$.

1. k -я итерация ($k \geq 0$)

Полагаем

$$\lambda_k = 2 M N (M + N)^{-1} (f(x_k)- f^*) \langle H_k g_k , g_k \rangle^{-1} ,$$

$$x_{k+1} = x_k - \lambda_k H_k g_k .$$

$$H_{k+1} = H_k - (1 - \gamma^{-2}) \frac{H_k g_k g_k^T H_k}{\langle H_k g_k , g_k \rangle}$$

где $g_k = g(x_k)$.

Итерация закончена.

Теорема 3.5.1. Пусть последовательность точек $\{ x_k \}_{k=0}^{\infty}$ вырабатывается методом (3.5.4). Тогда $f(x_k) \to f^*$ при $k \to \infty$ и справедлива следующая оценка скорости сходимости:

$$f_k^* - f^* \leq [\frac{(k + 1)(\gamma^2 - 1)}{n(\gamma^{2(k+1)/n} - 1)}]^{N/2} \cdot \omega(f , x^* , r) .$$

где $r = \| x_0 - x^* \|$; $\gamma = (M + N) / (M - N)$.

Д о к а з а т е л ь с т в о. Рассмотрим последовательность матриц $\{ G_k \}$ таких, что

$$G_0 = I\ ,\quad G_{k+1} = G_k + (\gamma^2 - 1)\,\frac{g_k\, g_k^T}{\langle H_k\, g_k\, ,\, g_k\rangle}\ ,$$

$k = 0,\ 1,\ \ldots$.

Нетрудно видеть, что $G_k H_k = H_k G_k = I$ для всех $k \geq 0$

Покажем, что при всех k выполнено неравенство

$$\langle G_{k+1}\,(x_{k+1} - x^*)\, ,\, x_{k+1} - x^*\rangle \leq$$

$$\leq \langle G_k\,(x_k - x^*)\, ,\, x_k - x^*\rangle \tag{3.5.5}$$

Положим

$\Delta G_k = G_{k+1} - G_k\ ,\quad \delta_k = x_{k+1} - x_k$

$$d_k = \langle G_{k+1}\,(x_{k+1} - x^*)\, ,\, x_{k+1} - x^*\rangle -$$
$$- \langle G_k\,(x_k - x^*)\, ,\, x_k - x^*\rangle = \langle G_{k+1}\,\delta_k\, ,\, \delta_k\rangle +$$
$$+ 2\,\langle G_{k+1}\,\delta_k\, ,\, x_k - x^*\rangle + \langle \Delta G_k\,(x_k - x^*)\, ,\, x_k - x^*\rangle .$$

Заметим, что

$$G_{k+1}\,\delta_k = -\,\gamma^2\,\lambda_k^2\,g_k\ ,\quad \langle g_k\, ,\, \delta_k\rangle =$$
$$= -\,2\,M\,N\,(M + N)^{-1}\,(\,f(x_k) - f^*\,)$$

Поэтому

$$d_k = \gamma^2\left(\frac{2\,M\,N}{M + N}\right)^2 \frac{f(x_k) - f^*}{\langle H_k\,g_k\, ,\, g_k\rangle} -$$
$$-\,2\,\gamma^2\,\frac{2\,M\,N}{M + N}\cdot\frac{f(x_k) - f^*}{\langle H_k\,g_k\, ,\, g_k\rangle}\,\langle g_k\, ,\, x_k - x^*\rangle +$$
$$+\,(\gamma^2 - 1)\,\frac{\langle g_k\, ,\, x_k - x^*\rangle^2}{\langle H_k\,g_k\, ,\, g_k\rangle} = \langle H_k\,g_k\, ,\, g_k\rangle^{-1}\,\times$$
$$\times\,\{\,\gamma^2\,[\langle g_k, x_k - x^*\rangle - 2\,M\,N\,(M + N)^{-1}(f(x_k) - f^*)]^2 -$$
$$-\,\langle g_k\, ,\, x_k - x^*\rangle^2\,\}\ .$$

Из этого соотношения, используя равенство $\gamma = (M + N)\,/\,(M - N)$ и неравенство (3.5.1), получаем, что $d_k \leq 0$ для всех $k \geq 0$. Из (3.5.5) получаем

$$\langle G_k\,(x_k - x^*)\, ,\, x_k - x^*\rangle \leq r^2 \tag{3.5.6}$$

Запишем теперь выражение для следа матрицы H_k

$$\mathrm{Trace}\ H_{k+1} = \mathrm{Trace}\ H_k - (1 - \gamma^{-2})\,\|\,H_k\,g_k\,\|^2\,/\,\|\,g_k\,\|^2 =$$

$= \ldots = n - (1 - \gamma^{-2}) \sum_{i=0}^{k} \| H_i g_i \|^2 / \| g_i \|^2 > 0.$

Положим

$v_i = \langle g_i , x_i - x^* \rangle / \| g_i \|$

Тогда из предыдущего неравенства и неравенства (3.5.5) получаем

$$\sum_{i=0}^{k} v_i^4 \le \sum_{i=0}^{k} \| g_i \|^{-4} \langle H_i g_i , g_i \rangle^2 \times$$
$$\times \langle G_i (x_i - x^*) , x_i - x^* \rangle^2 \le$$
$$\le r^4 \sum_{i=0}^{k} \| g_i \|^{-4} \langle H_i g_i , g_i \rangle^2 \le$$
$$\le r^4 \sum_{i=0}^{k} \| g_i \|^{-2} \| H_i g_i \|^2 \le$$
$$\le n \gamma^2 (\gamma^2 - 1)^{-1} r^4 .$$

Отсюда, воспользовавшись следствием 1.5.1. получаем, что $f(x_k) \to f^*$ при $k \to \infty$

Далее,

$$\det G_{k+1} = \det [G_k (I + (\gamma^2 - 1) \frac{H_k g_k g_k^T}{\langle H_k g_k , g_k \rangle}] =$$
$$= \gamma^2 \det G_k = \ldots = \gamma^{2(k+1)} .$$

Положим $b_k = \| g_k \|^{-2} \langle H_k g_k , g_k \rangle$. Тогда

$$\gamma^{2(k+1)} = n [\det G_{k+1}]^{1/n} \le \mathrm{Trace}\, G_{k+1} =$$
$$= n + (\gamma^2 - 1) \sum_{i=0}^{k} b_i^{-1} \le n + (b_k^*)^{-1} (\gamma^2 - 1)(k + 1)$$

Таким образом,

$$b_k^* \le n^{-1} (\gamma^2 - 1)(k + 1) [\gamma^{2(k+1)/n} - 1]^{-1}$$

для всех $k \ge 0$. Следовательно,

$$(v_k^*)^2 \le b_k^* r^2 \le$$
$$\le n^{-1} (\gamma^2 - 1)(k + 1) [\gamma^{2(k+1)/n} - 1]^{-1} r^2$$

Для завершения доказательства осталось воспользоваться следствиями 3.5.1 и 1.5.1. □

ГЛАВА 4. МЕТОДЫ РЕШЕНИЯ ЭКСТРЕМАЛЬНЫХ ЗАДАЧ С ГЛАДКИМИ КОМПОНЕНТАМИ

4.1. Оптимальные методы безусловной минимизации функций с липшицевым градиентом

В этом параграфе рассматриваются итеративные методы решения следующей экстремальной задачи:

$$\min \{ f(x) \mid x \in R^n \}, \tag{4.1.1}$$

где $f(x)$ - выпуклая функция из класса $\mathcal{F}(L, m)$.

Прежде всего остановимся на способе получения оценок скорости сходимости, который будет использоваться в настоящем параграфе.

Пусть $\{ x_k \}_{k=0}^{\infty}$ - последовательность точек, вырабатываемая методом $\mathfrak{M}$, $\Psi_k(x)$, $x \in R^n$, - последовательность функций таких, что для любого x из R^n

$$\lim_{k \to \infty} \Psi_k(x) = 0.$$

Предположим, что при любом $k > 0$ и $x \in R^n$ справедливо неравенство

$$f(x_k) \le f(x) + \Psi_k(x). \tag{4.1.2}$$

Тогда очевидно,

$$f(x_k) - f^* \le \Psi_k(x^*) \to 0 \text{ при } k \to \infty \tag{4.1.3}$$

Таким образом, если метод $\mathfrak{M}$ обеспечивает выполнение неравенства (4.1.2), то в силу (4.1.3) оценка стремления к нулю числовой последовательности $\{ \Psi_k(x^*) \}$ дает глобальную оценку скорости сходимости метода $\mathfrak{M}$.

Способ (4.1.2),(4.1.3) при соответствующем выборе функций Ψ_k позволяет получать оценки скорости сходимости некоторых известных методов, например метода градиентного спуска (см. §4.4). Однако более важным является тот факт, что попытка построить метод минимизации, непосредственно учитывающий структуру неравенства (4.1.2), приводит к оптимальным на $\mathcal{F}(L, m)$ алгоритмам.

Для упрощения изложения всюду далее будем считать, что параметры m и L класса $\mathcal{F}(L, m)$ известны.

Рассмотрим последовательность функций $\{ \varphi_k(x) \}$, $x \in R^n$, задаваемую следующими рекуррентными соотношениями:

$$\varphi_0(x) = f(x_0) + 0.5\, A_0 \| x - x_0 \|^2 ,$$
$$\varphi_{k+1}(x) = (1 - \alpha_k)\, \varphi_k(x) +$$
$$+ \alpha_k (f(y_k) + \langle f'(y_k), x - y_k \rangle +$$
$$+ 0.5\, m \| x - y_k \|^2), \quad k = 0, 1, \ldots , \tag{4.1.4}$$

где y_k - произвольные точки из R^n, α_k - произвольные числа из отрезка $[0, 1]$, A_0 - положительная константа.

Положим

$$\Psi_0 = \varphi_0(x) - f(x) = f(x_0) - f(x) +$$
$$+ 0.5\, A_0 \| x - x_0 \|^2 ,$$
$$\Psi_{k+1} = (1 - \alpha_k)\, \Psi_k , \; k = 0, 1, \ldots$$

Покажем, что при всех $x \in R^n$ и $k \geq 0$ выполняется неравенство

$$\varphi_k(x) \leq f(x) + \Psi_k(x) .$$

Действительно, $\varphi_0(x) \leq f(x) + \Psi_0(x)$. Далее, пусть это неравенство выполнено при некотором $k \geq 0$. Тогда, если $f \in \mathcal{F}(L, m)$, то

$$\varphi_{k+1}(x) = (1 - \alpha_k)\, \varphi_k(x) + \alpha_k (f(y_k) +$$
$$+ \langle f'(y_k), x - y_k \rangle + 0.5\, m \| x - y_k \|^2) \leq$$
$$\leq (1 - \alpha_k)(f(x) + \Psi_k(x)) + \alpha_k f(x) =$$
$$= f(x) + (1 - \alpha_k)\, \Psi_k(k) = f(x) + \Psi_{k+1}(x) .$$

Таким образом, при любом $k \geq 0$ и произвольном x из R^n справедливо неравенство

$$\varphi_k(x) \le f(x) + \Psi_k(x) = f(x) + \lambda_k \Psi_0(x), \qquad (4.1.5)$$
где $\lambda_0 = 1$, $\lambda_k = \Pi_{j=0}^{k-1} (1 - \alpha_j)$, $k = 0, 1, \dots$.

Следовательно, для того, чтобы воспользоваться способом оценки (4.1.2),(4.1.3), достаточно выбрать точки x_k, y_k и числа $\alpha_k \in [0, 1]$ так, чтобы при любом $k \ge 0$ и $x \in \in R^n$ выполнялось неравенство
$$f(x_k) \le \varphi_k(x) . \qquad (4.1.6)$$
В этом случае в силу (4.1.5) будет справедлива оценка
$$f(x_k) - f^* \le \lambda_k \Psi_0(x^*) =$$
$$= \lambda_k [f(x_0) - f^* + 0.5 A_0 \| x_0 - x^* \|^2] \qquad (4.1.7)$$

Покажем, как можно добиться выполнения неравенства (4.1.6).

Заметим, что в силу формул пересчета (4.1.4) функции φ_k допускают следующее каноническое представление:
$$\varphi_k(x) = \varphi_k + 0.5 A_k \| x - v_k \|^2 ,$$
где φ_k - минимальное значение функции $\varphi_k(x)$ на R^n; v_k - ее точка минимума, числа A_k пересчитываются по формуле $A_{k+1} = (1 - \alpha_k) A_k + \alpha_k m$, $k = 0, 1, \dots$.

<u>Лемма 4.1.1.</u> Если при некотором $k \ge 0$ значение функции $f(x_k) \le \varphi_k$, то при любом $y_k \in R^n$ справедливо неравенство
$$f(y_k) + \langle f'(y_k) , \alpha_k v_k + (1 - \alpha_k) x_k - y_k \rangle -$$
$$- 0.5 A_k^{-1} \alpha_k^2 (1 - \alpha_k)^{-1} \| f'(y_k) \|^2 \le \varphi_{k+1} .$$

Доказательство. Для упрощения выкладок опустим индексы у всех объектов, относящихся к k-му шагу и пометим плюсом в нижнем индексе все объекты, относящиеся к $(k + 1)$-му шагу.

Используя предположения леммы и выпуклость функции f, получаем
$$\varphi_+ = (1 - \alpha) \varphi(v_+) + \alpha(f(y) + \langle f'(y), v_+ - y \rangle +$$

$$+ 0.5\, m \, \| v_+ - y \|^2) \ge (1 - \alpha)\, \varphi + 0.5\, A\, (1 - \alpha) \times$$
$$\times \| v_+ - v \|^2 + \alpha \, (f(y) - \langle f'(y), v_+ - y \rangle) \ge$$
$$\ge (1 - \alpha)\, f(x) + 0.5\, A\, (1 - \alpha) \| v_+ - v \|^2 +$$
$$+ \alpha \, (f(y) - \langle f'(y), v_+ - y \rangle) \ge$$
$$\ge f(y) + \langle f'(y), \alpha v + (1 - \alpha) x - y \rangle +$$
$$+ \alpha \langle f'(y), v_+ - v \rangle + 0.5\, A\, (1 - \alpha) \| v_+ - v \|^2 \ge$$
$$\ge f(y) + \langle f'(y), \alpha v + (1 - \alpha) x - y \rangle -$$
$$- 0.5\, A^{-1} \alpha^2 (1 - \alpha)^{-1} \| f'(y) \|^2 . \quad \square$$

В силу теоремы 1.4.3 для любой функции $f(x)$ из класса $C^{1,1}(L)$ справедливо неравенство

$$f(x) \le f(y) + \langle f'(y), x - y \rangle + 0.5\, L \| x - y \|^2,$$

где x, y - произвольные точки из R^n. В частности, для $x_+ = y - L^{-1} f'(y)$ имеем

$$f(x_+) \le f(y) - 0.5\, L^{-1} \| f'(y) \|^2 .$$

Сопоставив это неравенство с утверждением леммы 4.1.1, нетрудно заметить, что для выполнения неравенства $f(x_{k+1}) \le \varphi_{k+1}$ достаточно выбрать точки x_{k+1}, y_k и число α_k из следующих условий:

$$L \alpha_k^2 = (1 - \alpha_k) A_k ;$$

$$\langle f'(y_k), \alpha_k v_k + (1 - \alpha_k) x_k - y_k \rangle \ge 0 , \tag{4.1.8}$$

$$f(x_{k+1}) \le f(y_k) - 0.5\, L^{-1} \| f'(y_k) \|^2 . \tag{4.1.9}$$

В частности, можно взять

$$y_k = \alpha_k v_k + (1 - \alpha_k) x_k , \tag{4.1.10}$$

$$x_{k+1} = y_k - L^{-1} f'(y_k) \tag{4.1.11}$$

В этом случае мы приходим к следующему методу:

Метод $\mathfrak{M}(m, L, x_0, A)$ (4.1.12)

0. Полагаем $A_0 = A$, где $A > 0$, $A \ge m$,
$\varphi_0(x) = f(x_0) + 0.5\, A_0 \| x - x_0 \|^2$

1. k -я итерация ($k \ge 0$)

 a) Вычисляем $\alpha_k > 0$ из уравнения

$L\ \alpha_k^2 = (1 - \alpha_k)\ A_k$,

б) полагаем

$v_k = \mathrm{argmin}\ \{\ \varphi_k(x)\ |\ x \in R^n\ \}$,

$y_k = \alpha_k\, v_k + (1 - \alpha_k)\, x_k$,

$A_{k+1} = \alpha_k\, m + (1 - \alpha_k)\, A_k$,

$x_{k+1} = y_k - L^{-1}\, f'(y_k)$,

$\varphi_{k+1}(x) = (1 - \alpha_k)\ \varphi_k(x) + \alpha_k\ (\ f(y_k) +$
$+ \langle f'(y_k),\ x - y_k \rangle + 0.5\, m\ \| x - y_k \|^2\ )$.

Итерация закончена.

Оценим скорость сходимости метода (4.1.12).

Теорема 4.1.1. Если метод $\mathfrak{M}(m, L, x_0, A)$ с $A > 0$, $A > m$ применяется для минимизации функции $f \in \mathcal{F}(L, m)$, то при любом $k > 0$ в случае $m > 0$ справедлива оценка

$$f(x_k) - f^* \leq \frac{4\, m\, [\ f(x_0) - f^* + 0.5\, A \| x_0 - x^* \|^2]}{[(A^{1/2} + m^{1/2})\, Q^k - (A^{1/2} - m^{1/2})\, Q^{-k}]^2}\ , \tag{4.1.13}$$

где $Q = 1 + 0.5\, q^{1/2}$, $q = m\, L^{-1}$

В случае $m = 0$ при условии существования точки x^* оценка (4.1.13) переходит в следующую:

$$f(x_k) - f^* \leq \frac{4\, L\ [\ f(x_0) - f^* + 0.5\, A \| x_0 - x^* \|^2]}{[\ 2\, L^{1/2} + k\, A^{1/2}\ ]^2} \tag{4.1.14}$$

Доказательство. В силу построения метода (4.1.12), для получения оценки его скорости сходимости можно использовать неравенство (4.1.7). А это значит, что достаточно оценить скорость стремления к нулю скалярных величин λ_k : $\lambda_0 = 1$; $\lambda_k = \Pi_{j=1}^{k-1}\, (1 - \alpha_j)$, $k = 1, 2, \ldots$, где числа α_j определяются из уравнения

$L\ \alpha_j^2 = (1 - \alpha_j)\ A_j$,

а величины A_j пересчитываются по формуле

$A_{j+1} = \alpha_j m + (1 - \alpha_j) A_j$.

Нетрудно убедиться в справедливости формулы

$A_k = m + \lambda_k (A_0 - m)$.

Поэтому, учитывая, что $\lambda_{k+1} = (1 - \alpha_k) \lambda_k$, получаем уравнение для величин λ_k :

$$L \lambda_k^{-2} (\lambda_k - \lambda_{k-1})^2 = \lambda_k^{-1} \lambda_{k+1} (m + \lambda_k (A - m)) . \tag{4.1.15}$$

Положим $\mu_k = \lambda_k^{-1/2}$ Тогда уравнение (4.1.15) примет вид

$$1 - \mu_k^2 \mu_{k+1}^{-2} = \mu_{k+1}^{-1} (q \mu_k^2 + L^{-1} (A - m))^{1/2}$$

т.е.

$$\mu_{k+1}^2 - \mu_k^2 = \mu_{k+1} (q \mu_k^2 + L^{-1} (A - m))^{1/2}$$

Отсюда в силу возрастания величин μ_k следует, что

$$\mu_{k+1} > \mu_k + 0.5 (q \mu_k^2 + L^{-1} (A - m))^{1/2} . \tag{4.1.16}$$

Теперь, воспользовавшись неравенством (4.1.16) и начальным условием $\mu_0 = 1$, по индукции нетрудно показать, что при всех $k > 0$ (в случае $m > 0$) справедлива оценка

$$\mu_k > 0.5 m^{-1/2} [(A^{1/2} + m^{1/2}) Q^k - (A^{1/2} - m^{1/2}) Q^{-k}] \tag{4.1.17}$$

В случае $m = 0$ справедлив предельный (при $m \to 0$) вариант оценки (4.1.17)

$$\mu_k > 1 + 0.5 k A^{1/2} L^{-1/2} . \tag{4.1.18}$$

Осталось заметить, что оценки (4.1.17),(4.1.18) в силу соотношения $\lambda_k = \mu_k^{-2}$ переходят в оценки скорости стремления к нулю величин λ_k . Подставляя эти оценки в неравенство (4.1.7), получаем (4.1.13),(4.1.14). □

Следствие 4.1.1. При всех $m > 0$ метод $\mathfrak{M}(m, L, x_0, A)$ является оптимальным на классе функций $\mathcal{F}(L, m)$.

Д о к а з а т е л ь с т в о. Из неравенства (4.1.17) легко получить, что

$$\mu_k > 0.5 (\exp (q^{1/2} k / 3) + 0.5 k A^{1/2} L^{-1/2}) >$$

$> 0.5 \max \{ \exp (q^{1/2} k / 3) , 1 + 0.5 k A^{1/2} L^{-1/2} \}$

Следовательно, для величин λ_k справедливо неравенство

$$\lambda_k = \mu_k^{-2} < 4 \min \{ \exp (- 2 q^{1/2} k / 3), 4 (k + 2)^{-2} \}$$

Подставляя это неравенство в (4.1.7) и учитывая, что

$$f(x_0) < f^* + 0.5 L \| x_0 - x^* \|^2 ,$$

получаем

$$f(x_k) - f^* < 4 L \| x_0 - x^* \|^2 \times$$
$$\times \min \{ \exp (- 2 q^{1/2} k / 3) ,4 (k + 2)^{-2} \}.$$

Таким образом, метод (4.1.12) обладает оптимальной оценкой скорости сходимости (см.Приложение). Осталось заметить, что на каждой итерации этот метод лишь один раз вычисляет значение функции и градиент.

Приведем алгоритмическую форму метода (4.1.12), в которой учитывается возможность канонического представления функций $\varphi_k(x)$.

0. Полагаем $A_0 = A$, где $A > 0$, $A > m$; $v_0 = x_0$

1. k -я итерация ($k > 0$)

а) Вычисляем $\alpha_k > 0$ из уравнения

$$L \alpha_k^2 = (1 - \alpha_k) A_k ;$$

б) полагаем

$$y_k = \alpha_k v_k + (1 - \alpha_k) x_k ,$$
$$A_{k+1} = \alpha_k m + (1 - \alpha_k) A_k ,$$
$$x_{k+1} = y_k - L^{-1} f'(y_k) ,$$
$$v_{k+1} = (1 - \alpha_k) A_k A_{k+1}^{-1} v_k + \alpha_k m A_{k+1}^{-1} y_k -$$
$$- \alpha_k A_{k+1}^{-1} f'(y_k) .$$

Итерация закончена.

При выводе метода (4.1.12) были использованы наиболее простые способы (4.1.10),(4.1.11), обеспечивающие выполнение условий (4.1.8),(4.1.9). Заметим, что существует много способов выбора точек x_k , y_k , удовлетворяющих этим соотношениям. В частности, можно взять

$$y_k = \mathrm{argmin} \{ f(x) \mid x = x_k + t(v_k - x_k),\ t \in R \},$$

$$x_{k+1} = \mathrm{argmin} \{ f(x) \mid x = y_k - t f'(y_k),\ t \geq 0 \}.$$

В этом случае приходим к следующему методу

Метод $\mathfrak{M}_1(m, L, x_0, A)$ (4.1.19)

0. Полагаем $A_0 = A$, где $A > 0$, $A \geq m$; $v_0 = x_0$

1. k -я итерация ($k \geq 0$).

а) Вычисляем $\alpha_k > 0$ из уравнения

$$L \alpha_k^2 = (1 - \alpha_k) A_k.$$

б) полагаем

$$y_k = \mathrm{argmin} \{ f(x) \mid x = x_k + t(v_k - x_k),\ t \in R \},$$

$$x_{k+1} = \mathrm{argmin} \{ f(x) \mid x = y_k - t f'(y_k),\ t \geq 0 \},$$

$$A_{k+1} = \alpha_k m + (1 - \alpha_k) A_k,$$

$$v_{k+1} = (1 - \alpha_k) A_k A_{k+1}^{-1} v_k + \alpha_k m A_{k+1}^{-1} y_k -$$

$$- \alpha_k A_{k+1}^{-1} f'(y_k).$$

Итерация закончена.

Мы не будем выписывать здесь схемы других методов, соответствующих различным стратегиям удовлетворения условий (4.1.8),(4.1.9). Отметим лишь, что для любого метода такого типа удается получить оптимальную оценку скорости сходимости. Причем обоснование этой оценки сводится, как и у метода (4.1.12), к анализу скорости убывания величин λ_k.

4.2. Метод сопряженных градиентов

Методы сопряженных градиентов относятся к числу наиболее распространенных методов безусловной минимизации. В настоящее время известно более ста методов такого типа Однако ни для одного из них пока не удалось получить оптимальные оценки скорости сходимости при решении задачи (4.1.1). Более того, недавно построены специальные примеры, на которых отдельные версии метода сопряженных градиентов работают не лучше, чем простейший метод градиентного спуска. В то же время, как известно, методы сопряженных градиентов являются оптимальными методами для решения задачи минимизации квадратичной функции. В связи с этим возникает вопрос: можно ли модифицировать схему метода сопряженных градиентов таким образом, чтобы, с одной стороны, сохранились все достоинства этого метода, проявляющиеся при минимизации функций, близких к квадратичным, а с другой стороны, - чтобы полученный метод был оптимальным на классе $\mathcal{F}(L, m)$.

В этом параграфе мы постараемся ответить на этот вопрос. Для этого проанализируем поведение одной версии метода сопряженных градиентов при решении квадратичной задачи с помощью методики оценки скорости сходимости, описанной в предыдущем параграфе.

Рассмотрим следующий метод.

$$\text{Метод } \mathfrak{M}_2(x_0) \tag{4.2.1}$$

0. Полагаем $y_{-2} = y_{-1} = x_0$

1. k -я итерация ($k \geq 0$)

Полагаем

$$y_k = \operatorname{argmin} \{ f(x) \mid x = x_k + t(y_{k-2} - x_k), \ t \in R \}, \tag{4.2.2}$$

$$x_{k+1} = \operatorname{argmin} \{ f(x) \mid x = y_k - t f'(y_k), \ t \geq 0 \}. \tag{4.2.3}$$

Итерация закончена.

Пусть метод (4.2.1) применяется для решения задачи

$$\min \{ f(x) = 0.5 \langle B(x - x^*), x - x^* \rangle \mid x \in R^n \}, \tag{4.2.4}$$

где $B = B^T$ - неотрицательно определенная $n \times n$ -матрица; $f \in \mathcal{F}(L, m)$ Обозначим

$\mathcal{L}_k = \{ x \mid x = x_0 + \sum_{j=0}^{k-1} \lambda_j f'(y_j), \ \lambda_j \in R \}$.

Теорема 4.2.1. Пусть последовательности точек $\{ x_k \}$, $\{ y_k \}$ сформированы методом (4.2.1) при решении задачи (4.2.4). Тогда

1) при любом $k > 0$ выполняется равенство

$$y_k = \operatorname{argmin} \{ f(x) \mid x \in \mathcal{L}_k \}, \tag{4.2.5}$$

и, следовательно, за число шагов, не превышающее n, этот метод построит точное решение задачи (4.2.4);

2) справедливы следующие оценки:

$$f(x_k) - f^* \le \frac{4m [f(x_0) - f^* + 0.5 A \| x_0 - x^* \|^2]}{[(A^{1/2} + m^{1/2}) Q^k - (A^{1/2} - m^{1/2}) Q^{-k}]^2} \tag{4.2.6}$$

в случае $m > 0$ и

$$f(x_k) - f^* \le \frac{4L [f(x_0) - f^* + 0.5 A \| x_0 - x^* \|^2]}{[2 L^{1/2} + k A^{1/2}]^2} \tag{4.2.7}$$

в случае $m = 0$. В оценках (4.2.6), (4.2.7) $Q = 1 + 0.5 \times q^{1/2}$, $q = m L^{-1}$, A - любое положительное число, $A > m$.

Доказательство. Соотношение (4.2.5) будем доказывать по индукции Из схемы метода (4.2.1) следует, что

$y_0 = \operatorname{argmin} \{ f(x) \mid x \in \mathcal{L}_0 \equiv \{ x_0 \} \}$,

$y_1 = \operatorname{argmin} \{ f(x) \mid x \in \mathcal{L}_1 \}$

Предположим, что соотношение (4.2.5) выполнено при всех k, $0 \le k \le m$, где $m \ge 1$. Из способов (4.2.2),(4.2.3) выбора точек y_k, x_k вытекает, что

$$\langle f'(x_{m+1}), f'(y_m) \rangle = 0, \tag{4.2.8}$$

$$< f'(y_{m+1}) , x_{m+1} - y_{m-1} > = 0 \qquad (4.2.9)$$

Кроме того, из (4.2.2) и из квадратичности функции f следует, что

$$f'(y_{m+1}) = (1 + t') f'(x_{m+1}) - t' f'(y_{m-1}) . \qquad (4.2.10)$$

Сопоставляя соотношения (4.2.8) - (4.2.10) и пользуясь индуктивным предположением, получаем

$$< f'(y_{m+1}), f'(y_m) > = t' < f'(y_{m-1}), f'(y_m) > = 0 ,$$

$$< f'(y_{m+1}), y_m - y_{m-1} > = < f'(y_{m+1}) , x_{m+1} +$$

$$+ t'' f'(y_m) - y_{m-1} > = 0 .$$

А значит,

$$y_{m+1} = \operatorname{argmin} \{ f(x) \mid x = y_m + t' f'(y_m) +$$

$$+ t'' (y_m - y_{m-1}) , \ t', t'' \in R \} . \qquad (4.2.11)$$

В свою очередь, пользуясь индуктивным предположением и соотношением (4.2.11), можно стандартным способом доказать справедливость равенства (4.2.5) для точки y_{m+1} (см.[37, с.70-71]). Таким образом, соотношение (4.2.5) доказано.

Теперь заметим, что в силу конечномерности пространства R^n число итераций, на которых происходит расширение многообразий $\mathcal{L}_k$, не может быть больше n. Если же на некоторой итерации с номером k оказалось, что $\mathcal{L}_k = \mathcal{L}_{k+1}$, то это означает, что линейная оболочка векторов $f'(x_0), \ldots ,$ $f'(y_{k-1})$ содержит нуль. Отсюда в силу линейности оператора $f'(x)$ получаем, что $\mathcal{L}_{k+1} \cap X^* \neq \emptyset$, и, следовательно, $y_{k+1} \in X^*$

Для доказательства второго утверждения введем последовательность вспомогательных функций $\{ \varphi_k(x) \}$, которые определяются следующими рекуррентными соотношениями:

$$\varphi_0(x) = f(x_0) + 0.5 A_0 \| x - x_0 \|^2 , \qquad (4.2.12)$$

$$\varphi_{k+1}(x) = (1 - \alpha_k) \varphi_k(x) + \alpha_k (f(y_k) +$$

$$+ < f'(y_k) , x - y_k > + 0.5 m \| x - y_k \|^2) . \qquad (4.2.13)$$

В формулах (4.2.12),(4.2.13) $A_0 = A$ - любое положительное число, не меньшее, чем m, числа $\alpha_k > 0$ вычисляются из уравнения

$$L\ \alpha_k^2 = (1 - \alpha_k)\ A_k,$$

а числа A_k пересчитываются по формуле

$$A_{k+1} = \alpha_k\ m + (1 - \alpha_k)\ A_k .$$

Отметим, что при любом $k \geq 0$ точки x_k и точки

$$v_k = \operatorname{argmin}\{\ \varphi_k(x)\ |\ x \in R^n\ \}$$

лежат в многообразии $\mathcal{L}_k$. С помощью рассуждений, приведенных в начале §4.1, нетрудно установить справедливость неравенства

$$\varphi_k = \operatorname{argmin}\{\ \varphi_k(x)\ |\ x \in R^n\ \} \leq$$
$$\leq f^* + \lambda_k [\ f(x_0) - f^* + 0.5\ A_0\ \|\ x_0 - x^*\ \|^2\ ],$$

где $\lambda_0 = 1$, $\lambda_k = \Pi_{j=0}^{k-1} (1 - \alpha_j)$.

Покажем, что в методе (4.2.1) $f(x_k) \leq \varphi_k$.

Действительно, $f(x_0) = \varphi_0$. Далее, по лемме 4.2.1. если $f(x_k) \leq \varphi_k$, то

$$\varphi_{k+1} \geq f(y_k) + \langle f'(y_k),\ \alpha_k v_k + (1 - \alpha_k)\ x_k - y_k \rangle -$$
$$- 0.5\ A_k^{-1}\ \alpha_k^2\ (1 - \alpha_k)^{-1}\ \|\ f'(y_k)\ \|^2 =$$
$$= f(y_k) - 0.5\ L^{-1}\ \|\ f'(y_k)\ \|^2 +$$
$$+ \langle f'(y_k),\ \alpha_k v_k + (1 - \alpha_k)\ x_k - y_k \rangle \geq$$
$$\geq f(x_{k+1}) + \langle f'(y_k),\ \alpha_k v_k + (1 - \alpha_k)\ x_k - y_k \rangle .$$

Но $y_k = \operatorname{argmin}\{\ f(x)\ |\ x \in \mathcal{L}_k\ \}$. Поэтому

$$\langle f'(y_k),\ z - y_k \rangle = 0$$

для любого $z \in \mathcal{L}_k$, и в частности для

$$z = \alpha_k v_k + (1 - \alpha_k)\ x_k$$

Таким образом, мы установили, что для метода (4.2.1) при всех $k \geq 0$ справедливо неравенство

$$f(x_k) \leq \varphi_k \leq f^* + \lambda_k [\ f(x_0) - f^* +$$
$$+ 0.5\ A_0\ \|\ x_0 - x^*\ \|^2\ ].$$

Осталось воспользоваться оценками для величин λ_k, полученными при доказательстве теоремы 4.1.1. □

При выводе оценки скорости сходимости метода (4.2.1) характеристическое свойство метода сопряженных градиентов (4.2.5) использовалось лишь для доказательства неравенства

$$\langle f'(y_k),\ \alpha_k v_k + (1 - \alpha_k) x_k - y_k \rangle \geq 0 . \qquad (4.2.13)$$

Однако, если метод (4.2.1) применять для минимизации неквадратичных функций, то гарантировать выполнение равенства (4.2.5) заведомо не удастся. В этом случае соотношение (4.2.13) может оказаться нарушенным, а значит, может ухудшиться и скорость сходимости метода. Поэтому при переносе метода (4.2.1) на общий нелинейный случай необходимо принять специальные меры для того, чтобы обеспечить выполнение неравенства (4.2.13). Так, например, можно модифицировать метод (4.2.1), соединив его с методом (4.1.19). Одна из возможных комбинаций этих методов выглядит следующим образом.

Метод $\mathfrak{M}_3(m, L, x_0, A)$ (4.2.14)

0. Полагаем $A_0 = A$, где $A > 0$, $A \geq m$;

$y_{-2} = y_{-1} = v_0 = x_0$.

1. k-я итерация ($k \geq 0$).

а) Полагаем

$y'_k = \arg\min \{ f(x) \mid x = x_k + t(y_{k-2} - x_k),\ t \in R \}$,

$y_k = \arg\min \{ f(x) \mid x = y'_k + t(v_k - y'_k),\ t \in R \}$;

б) вычисляем $\alpha_k > 0$ из уравнения

$L \alpha_k^2 = (1 - \alpha_k) A_k$;

в) полагаем

$x_{k+1} = \arg\min \{ f(x) \mid x = y_k - t f'(y_k),\ t \geq 0 \}$,

$A_{k+1} = \alpha_k m + (1 - \alpha_k) A_k$,

$v_{k+1} = (1 - \alpha_k) A_k A_{k+1}^{-1} v_k + \alpha_k m A_{k+1}^{-1} y_k -$
$- \alpha_k A_{k+1}^{-1} f'(y_k)$.

Итерация закончена.

Нетрудно показать, что для квадратичных функций метод (4.2.14) генерирует последовательность точек, удовлетворяющую соотношению (4.2.5). При минимизации неквадратичных функций для метода (4.2.14) справедливы оценки (4.1.13), (4.1.14). В то же время, трудоемкость одной итерации метода (4.2.14) лишь в полтора раза превышает трудоемкость итерации каждого из методов (4.1.19),(4.2.1).

4.3. Адаптивный метод безусловной минимизации для функций с гельдеровым градиентом

Рассмотрим следующую экстремальную задачу

$$\min \{ f(x) \mid x \in R^n \}, \qquad (4.3.1)$$

где $f(x)$ - функция из класса $C^{1,\nu}(L)$, $0 < \nu < 1$ (см. §1.4). Мы считаем, что оптимальное множество X^* задачи (4.3.1) не пусто. Из результатов, приведенных в Приложении, ясно, что если размерность пространства n достаточно велика, то никакой метод решения задачи (4.3.1), стартующий из фиксированной точки $x' \in R^n$, не может гарантировать получение ε-приближения к минимуму произвольной функции $f(x)$ из класса $C^{1,\nu}(L)$, $\rho(X^*(f), x') < R$, с трудоемкостью, меньшей, чем

$$O([\varepsilon L^{-1} R^{-1-\nu}]^{-1/\mu})$$

вычислений значения функции и градиента, где $\mu = 0.5 \times$
$\times (3\nu + 1)$.

Опишем метод, реализующий нижнюю оценку трудоемкости на рассматриваемом классе задач. Отметим, что в число параметров этого метода не входят показатель гладкости ν и константа L, т.е. адаптация метода к реальной гладкости конкретной целевой функции производится автоматически.

Метод $\mathcal{M}(x', N, R)$ (4.3.2)

0. Полагаем

$v_0 = x'$, $x_0 = x'$, $a_0 = 1$, $h_0 = R \| f'(x_0) \|^{-1}$.

1. k -я итерация ($k = 0, \ldots, N$).

а) Полагаем

$a_{k+1} = 0.5 [1 + (4 a_k^2 + 1)^{1/2}]$,

$y_k = (1 - a_{k+1}^{-1}) x_k + a_{k+1}^{-1} v_k$;

б) последовательным перебором значений $i = 0, 1, \ldots$ находим первый номер $i = i(k)$, при котором для $h = 2^{-i} h_k$ выполнится неравенство

$f(y_k) - f(y_k - h f'(y_k)) >$

$> 0.5 h \| f'(y_k) \|^2 - 0.5 h^{-1} a_{k+1}^{-2} N^{-1} R^2$;

в) полагаем

$h_{k+1} = 2^{-i(k)} h_k$,

$x_{k+1} = y_k - h_{k+1} f'(y_k)$.

$v_{k+1} = v_k - h_{k+1} a_{k+1} f'(y_k)$

Итерация закончена.

<u>Теорема 4.3.1.</u> Если последовательность точек $\{ x_k \}_{k=0}^N$ построена методом $\mathcal{M}(x', N, R)$ при решении задачи (4.3.1) с функцией $f(x)$ из класса $C^{1,\nu}(L)$, $0 < \nu < 1$, то

$f(x_N) - f^* < 4 L R^{1+\nu} N^{-(1+3\nu)/2} (1 + r R^{-1})^2$, (4.3.3)

где $r = \| x' - x^* \|$. При этом суммарное число вычислений значения функции за N итераций не превзойдет величины порядка

$O(N + \ln \{ L R^\nu \| f'(x') \|^{-1} \})$

(вектор градиента вычисляется на итерации только один раз).

Доказательство. Пусть сначала параметр ν в условии теоремы меньше единицы. Покажем корректность правила б) выбора шага h. Действительно, в силу теоремы 1.4.1 при любом $h > 0$ справедливо неравенство

$$f(y_k - h f'(y_k)) \le f(y_k) + \langle f'(y_k), (y_k - h f'(y_k)) - y_k \rangle + (1 + \nu)^{-1} L \| h f'(y_k) \|^{1+\nu} = f(y_k) - h \| f'(y_k) \|^2 + (1 + \nu)^{-1} L h^{1+\nu} \times \| f'(y_k) \|^{1+\nu}.$$

Поэтому

$$f(y_k - h f'(y_k)) - f(y_k) + 0.5 h \| f'(y_k) \|^2 \le - 0.5 h \| f'(y_k) \|^2 + (1 + \nu)^{-1} L h^{1+\nu} \times \| f'(y_k) \|^{1+\nu}. \quad (4.3.4)$$

Воспользуемся неравенством

$$\alpha u^{1+\nu} - \beta u^2 \le (1 - \nu) [(1 + \nu) \beta^{-1}]^{(1+\nu)/(1-\nu)} \times [0.5 \alpha]^{2/(1-\nu)},$$

которое справедливо при любых $\beta > 0$, $\alpha, u > 0$. Подставим в это неравенство следующие значения:

$$\alpha = (1 + \nu)^{-1} L h^{1+\nu}, \quad \beta = 0.5 h, \quad u = \| f'(y_k) \|$$

После несложных преобразований получим

$$- 0.5 h \| f'(y_k) \|^2 + (1 + \nu)^{-1} L h^{1+\nu} \times \| f'(y_k) \|^{1+\nu} \le 0.5 (1 - \nu) (1 + \nu)^{-1} L^{2/(1-\nu)} \times h^{(1+\nu)/(1-\nu)}$$

Таким образом, в силу неравенства (4.3.4) процесс одномерного поиска в пункте б) прервется не позднее, чем произойдет выполнение неравенства

$$0.5 (1 - \nu) (1 + \nu)^{-1} L^{2/(1-\nu)} h^{(1+\nu)/(1-\nu)} \le 0.5 [h a_{k+1}^2 N]^{-1} R^2,$$

т.е.

$$(L h)^{2/(1-\nu)} \le (1 + \nu) (1 - \nu)^{-1} [a_{k+1}^2 N]^{-1} R^2.$$

Но процесс одномерного поиска в пункте 6) организован таким образом, что после невыполнения критерия прерывания текущее значение h перед очередной проверкой уменьшается вдвое. Отсюда заключаем, что справедливо следующее неравенство:

$$h_k \geq h_{k+1} \geq 0.5\, L^{-1} \times$$
$$\times \left[(1 + \nu)(1 - \nu)^{-1} a_{k+1}^{-2} N^{-1} R^2 \right]^{(1-\nu)/2} \qquad (4.3.5)$$

Перейдем к доказательству неравенства (4.3.3). Прежде всего заметим, что

$$a_{k+1}^2 \equiv a_k^2 + a_{k+1} .$$

Поэтому

$$a_{k+1} \geq a_k ,$$
$$2\, a_{k+1} (\, a_{k+1} - a_k \,) \geq a_{k+1}^2 - a_k^2 = a_{k+1} .$$

Таким образом, $a_{k+1} \geq a_k + 0.5$ и, следовательно, при всех $k \geq 0$

$$a_k \geq 1 + 0.5\, k . \qquad (4.3.6)$$

Покажем, что при всех $k \geq 0$ выполняется неравенство

$$2\, h_{k+1} a_{k+1}^2 (\, f(\, x_{k+1} \,) - f^* \,) + \| \, v_{k+1} - x^* \, \|^2 \leq$$
$$\leq 2\, h_k a_k^2 (\, f(\, x_k \,) - f^* \,) + \| \, v_k - x^* \, \|^2 + N^{-1} R^2 .$$
$$(4.3.7)$$

Действительно,

$$\| \, v_{k+1} - x^* \, \|^2 - \| \, v_k - x^* \, \|^2 - h_{k+1}^2 a_{k+1}^2 \times$$
$$\times \| \, f'(\, y_k \,) \, \|^2 =$$
$$= -\, 2\, h_{k+1} a_{k+1} \langle f'(\, y_k \,) ,\, v_k - x^* \rangle =$$
$$= -\, 2\, h_{k+1} a_{k+1} \langle f'(\, y_k \,) ,\, y_k - x^* \rangle +$$
$$+\, 2\, h_{k+1} a_{k+1} (a_{k+1} - 1) \langle f'(\, y_k \,) ,\, x_k - y_k \rangle \leq$$
$$\leq -\, 2\, h_{k+1} a_{k+1} (\, f(\, y_k \,) - f^* \,) +$$
$$+\, 2\, h_{k+1} a_{k+1} (a_{k+1} - 1) (\, f(\, x_k \,) - f(\, y_k \,) \,) =$$
$$= 2\, h_{k+1} a_k^2 (\, f(\, x_k \,) - f^* \,) -$$
$$-\, 2\, h_{k+1} a_{k+1}^2 (\, f(\, y_k \,) - f^* \,) \leq$$
$$\leq 2\, h_k a_k^2 (\, f(\, x_k \,) - f^* \,) - 2\, h_{k+1} a_{k+1}^2 (\, f(\, y_k \,) - f^* \,)$$

Поэтому в силу критерия прерывания процесса одномерного поиска имеем

$$\| v_{k+1} - x^* \|^2 - \| v_k - x^* \|^2 -$$
$$- 2 h_k a_k^2 (f(x_k) - f^*) \le$$
$$\le - 2 h_{k+1} a_{k+1}^2 [f(y_k) - f^* -$$
$$- 0.5 h_{k+1} \| f'(y_k) \|^2] \le$$
$$\le - 2 h_{k+1} a_{k+1}^2 [f(x_{k+1}) - f^*] + N^{-1} R^2$$

Таким образом, неравенство (4.3.7) доказано.

Объединяя теперь неравенства (4.3.5)-(4.3.7), получаем

$$k N^{-1} R^2 + 2 h_0 (f(x') - f^*) + \| x' - x^* \|^2 \ge$$
$$\ge 2 h_k a_k^2 (f(x_k) - f^*) + \| v_k - x^* \|^2 \ge$$
$$\ge L^{-1} R^{1-\nu} [(1 + \nu)(1 - \nu)^{-1}]^{(1-\nu)/2} \times$$
$$\times a_k^{1+\nu} N^{-(1-\nu)/2} (f(x_k) - f^*) \ge$$
$$\ge L^{-1} R^{1-\nu} (1 + 0.5 k)^{1+\nu} N^{-(1-\nu)/2} (f(x_k) - f^*)$$

Подставим в это неравенство значение $k = N$:

$$0.25 L^{-1} R^{1-\nu} N^{(1+3\nu)/2} (f(x_N) - f^*) \le$$
$$\le R^2 + 2 R \| f'(x_0) \|^{-1} (f(x') - f^*) + r^2 \le$$
$$\le R^2 (1 + r R^{-1})^2 ,$$

откуда получаем неравенство (4.3.3).

Оценим теперь суммарное количество вычислений значения функции $f(x)$, которое обозначим через S. Ясно, что

$$S = N + \Sigma_{k=0}^{N} i(k) .$$

Заметим, что в силу неравенства (4.3.5) и

$$0.5 L^{-1} R^{1-\nu} N^{-3(1-\nu)/2} \le h_N = 2^{- \Sigma_{k=0}^{N} i(k)} R \| f'(x') \|^{-1} ,$$

откуда получаем требуемую оценку.

Пусть теперь $\nu = 1$. Заметим, что в приведенных рассуждениях неравенство $\nu < 1$ использовалось только для обоснования конечности одномерного поиска и вывода нижней оценки для величин h_k. В случае $\nu = 1$ в силу неравенства (4.3.4) критерий прерывания одномерного поиска в методе (4.3.2) заведомо сработает, если $h < L^{-1}$. Поэтому

$$h_k > 0.5 L^{-1}.$$

Остальная часть доказательства при $\nu = 1$ остается без изменений. □

Метод (4.3.2) можно превратить в "бесконечношаговый", если неравенство в критерии прерывания одномерного поиска на шаге 6) заменить на следующее.

$$f(y_k) - f(y_k - h f'(y_k)) >$$
$$> 0.5 h \| f'(y_k) \|^2 - 0.5 h^{-1} a_{k+1}^{-2} (k + 1)^{-1} R^2 .$$

Повторяя почти дословно доказательство теоремы 4.3.1, можно показать, что оценка скорости сходимости такого метода останется "почти" оптимальной:

$$f(x_k) - f^* < \text{const } L R^{1+\nu} (1 + r R^{-1})^2 k^{-(1+3\nu)/2} \ln k.$$

При этом порядок суммарного числа вычислений значения функции за k итераций не изменится.

4.4. Минимизация составных функций. Условные задачи

Рассмотрим следующую экстремальную задачу

$$f(x) = F(f(x)) \to \min \mid x \in Q , \qquad (4.4.1)$$

где Q - выпуклое замкнутое множество пространства R^n; $f(x)$ - вектор-функция из R^n в R^m с выпуклыми компонентами $f^{(i)}(x) \in C^{1,1}(L^{(i)})$; $F(u)$ - выпуклая монотонная функция на R^m: если $u, v \in R^m$ и $u > v$, то $F(u) > F(v)$. Так, например, возможно, что

$$F(u) = \max \{ u^{(i)} \mid i = 1 , \dots , m \} ,$$

$$F(u) = [\,\sum_{i=1}^{m} (u_+^{(i)})^2\,]^{1/2},$$

$$F(u) = [\,\sum_{i=1}^{m} (u^{(i)})^2\,]^{1/2} + \sum_{i=1}^{m} u^{(i)},$$

$$F(u) = \begin{cases} 0, & \text{если } u \le 0, \\ +\infty & \text{в противном случае.} \end{cases}$$

Нетрудно видеть, что в силу сделанных предположений функция $f(x)$ в задаче (4.4.1) будет выпуклой. Более того, при всех x, y из R^n справедливо неравенство

$$F(f(x) + f'(x)(y - x)) \le f(y) \le$$
$$\le F(f(x) + f'(x)(y - x) + 0.5\,\| y - x \|^2 L), \tag{4.4.2}$$

где $L = (L^{(1)}, \ldots, L^{(m)}) \in R^m$

Заметим, что функция $f(x)$ может оказаться и негладкой. Поэтому для решения задачи (4.4.1) нельзя непосредственно применять методы минимизации гладких функций, описанные в §4.1. Однако в том случае, если функция $F(u)$ достаточно просто устроена можно попытаться учесть ее структуру непосредственно в схеме соответствующего метода минимизации. Здесь существует общий прием, предполагающий возможность эффективного решения вспомогательных экстремальных задач, в которых участвует функция $F(u)$. Покажем, как можно ввести такую конструкцию в схему простейшего градиентного метода с постоянным шагом. Для вывода оценки скорости сходимости этого метода воспользуемся подходом, изложенным в §4.1.

Пусть в задаче (4.4.1) функции $f^{(i)}(x) \in \mathcal{F}(L^{(i)}, m^{(i)})$, причем $m^{(i)} / L^{(i)} \ge q > 0$ для любого $i = 1, \ldots, m$. Для простоты изложения будем считать, что векторы m и L известны. Зафиксируем произвольную точку $x_0 \in Q$ и положим

$$\lambda_k = (1 - q)^k, \quad \Psi_0(x) \equiv f(x_0).$$

$$\Psi_k(x) = (1 - \lambda_k) f(x) + \lambda_k \Psi_0(x)$$

для всех $k > 0$. Покажем, как можно построить последовательность точек $\{ x_k \}_{k=0}^{\infty}$, для которой при любом x из Q справедливо соотношение

$$f(x_k) < \Psi_k(x) \qquad (4.4.3)$$

Действительно, это неравенство выполнено при $k = 0$. Пусть оно выполнено и при некотором $k > 0$. Тогда

$$\begin{aligned}
&\Psi_{k+1}(x) = (1 - \lambda_{k+1}) f(x) + \lambda_{k+1} \Psi_0(x) = \\
&= q f(x) + (1 - q) \Psi_k(x) > q f(x) + \\
&+ (1 - q) f(x_k) = \\
&= q F(f(x)) + (1 - q) F(f(x_k)) > \\
&> q F(f(x_k) + f'(x_k)(x - x_k) + \\
&+ 0.5 \| x - x_k \|^2 m) + (1 - q) F(f(x_k)) > \\
&> F(f(x_k) + q f'(x_k) (x - x_k) + \\
&+ 0.5 q \| x - x_k \|^2 m) > \\
&> F(f(x_k) + f'(x_k) [q (x - x_k)] + \\
&+ 0.5 \| q (x - x_k) \|^2 L).
\end{aligned}$$

Поэтому

$$\begin{aligned}
&\min \{ \Psi_{k+1}(x) \mid x \in Q \} > \\
&> \min \{ F(f(x_k) + f'(x_k) [q (x - x_k)] + \\
&+ 0.5 \| q (x - x_k) \|^2 L) \mid x \in Q \} = \\
&= \min \{ F(f(x_k) + f'(x_k) (y - x_k) + \\
&+ 0.5 \| y - x_k \|^2 L) \mid y \in (1 - q) x_k + q Q \} > \\
&> \min \{ F(f(x_k) + f'(x_k) (y - x_k) + \\
&+ 0.5 \| y - x_k \|^2 L) \mid y \in Q \}. \qquad (4.4.4)
\end{aligned}$$

Таким образом, если выбрать точку x_{k+1} по любому из следующих правил:

$$\begin{aligned}
&x_{k+1} = \operatorname{argmin} \{ F(f(x_k) + f'(x_k) (y - x_k) + \\
&+ 0.5 \| y - x_k \|^2 L) \mid y \in (1 - q) x_k + q Q \}, \qquad (4.4.5)
\end{aligned}$$

$$x_{k+1} = \text{argmin} \{ F(f(x_k) + f'(x_k) (y - x_k) + + 0.5 \| y - x_k \|^2 L) \mid y \in Q \} , \qquad (4.4.6)$$

то в силу неравенств (4.4.2),(4.4.6) получим, что соотношение (4.4.3) выполнено и для номера $k + 1$.

Приведенные рассуждения доказывают следующую теорему.

Теорема 4.4.1. Если в задаче (4.4.1) $f^{(i)}(x) \in \in \mathcal{F}(L^{(i)}, m^{(i)})$, $m^{(i)} / L^{(i)} > q > 0$, $i = 1, \dots , m$, то для последовательности точек $\{ x_k \}_{k=0}^{\infty}$ ($x_0 \in Q$, $f(x_0) < \infty$), построенной любым из методов (4.4.5),(4.4.6), справедлива оценка скорости сходимости

$$f(x_k) - f(x^*) < (1 - q)^k [f(x_0) - f(x^*)] .$$

Нетрудно заметить, что в случае $F(u) = u$, $u \in R$, метод (4.4.6) совпадает с методом проекции градиента с постоянным шагом

$$x_0 \in Q , \quad x_{k+1} = \pi(Q , x_k - L^{-1} f'(x_k)) .$$

При обосновании методов (4.4.5),(4.4.6) мы не использовали никаких дополнительных свойств функции $F(u)$ помимо монотонности и выпуклости. Поэтому указанные методы применимы, в частности, и для решения задачи (4.4.1) с разрывной функцией $F(u)$. Положим

$$F(u) = \begin{cases} u^{(1)} , \text{ если } u^{(i)} < 0 , i = 2, \dots , m , \\ +\infty \quad \text{в противном случае.} \end{cases}$$

В этом случае задача (4.4.1) эквивалентна следующей экстремальной задаче:

$$f_1(x) \to \min \mid x \in Q , f^{(i)}(x) < 0 , i = 2, \dots , m .$$

Приведем схему метода (4.4.6), в которой учтена специфика рассматриваемой функции $F(u)$.

0) Выбираем начальную точку $x_0 \in Q$, такую, что $f^{(i)}(x_0) < 0$, $i = 2, \ldots, m$.

1) k-я итерация ($k \geq 0$). (4.4.7)

В качестве точки x_{k+1} выбираем решение следующей вспомогательной задачи:

$$f^{(1)}(x_k) + \langle (f^{(1)})'(x_k), x - x_k \rangle + 0.5\, L^{(1)} \| x - x_k \|^2 \to \min$$

при условиях:

$$x \in Q,$$

$$f^{(i)}(x_k) + \langle (f^{(i)})'(x_k), x - x_k \rangle + 0.5\, L^{(i)} \| x - x_k \|^2 \leq 0, \quad i = 2, \ldots, m.$$

Итерация закончена.

Теорема 4.4.1 дает для этого метода такую оценку скорости сходимости

$$f^{(1)}(x_k) - f^{(1)}(x^*) \leq (1 - q)^k [f^{(1)}(x_0) - f^{(1)}(x^*)]$$

Отметим, что если в задаче (4.4.1) $Q \equiv R^n$, то для решения вспомогательной задачи в методе (4.4.7) можно сначала решить двойственную задачу с дробно - квадратичным целевым функционалом

$$\langle f(x_k), u \rangle - 0.5 \langle L, u \rangle^{-1} \| [f'(x_k)]^T u \|^2 \to \max$$

при условиях $u \in R^m_+$, $u^{(1)} = 1$, (4.4.8)

а затем положить

$$x_{k+1} = x_k - \langle L, u_k^* \rangle^{-1} [f'(x_k)]^T u_k^*,$$

где u_k^* - решение задачи (4.4.8).

Пусть теперь в задаче (4.4.1) функция $F(u)$ удовлетворяет следующему дополнительному условию:

$$F(u + t v) \leq F(u) + t F(v) \qquad (4.4.9)$$

для любых $u \in R^m$, $v \in R^m_+$, $t \geq 0$. Можно показать, что если $F(u)$ - положительно однородная (степени единица) выпуклая функция с $\operatorname{dom} F = R^m$, то свойство (4.4.9) для нее выполнено. С помощью методики, изложенной в §4.1, выведем аналог метода (4.1.12) с $m = 0$ для решения задачи (4.4.1), (4.4.9). Для упрощения выкладок обозначим

$$\Phi(x, y) = F(f(y) + f'(y)(x - y))$$

Ясно, что при любых x, y функция $f(x) \geq \Phi(x, y)$.

Зафиксируем произвольную точку $x_0 \in Q$ и число $A_0 > 0$. Положим

$$\varphi_0 = f(x_0), \quad v_0 = x_0,$$

$$\Psi_0(x) = f(x_0) - f(x) + 0.5 A_0 \| x - v_0 \|^2.$$

Предположим, что к k-й итерации построены точки $x_i \in Q$, $v_i \in R^n$ и числа $A_i > 0$, $\alpha_i \in (0, 1)$, φ_i, такие, что при любом i, $0 \leq i \leq k$, и $x \in Q$ справедливы неравенства

$$f(x_i) \leq \varphi_i + 0.5 A_i \| x - v_i \|^2 \leq$$
$$\leq f(x^*) + \Psi_i(x), \tag{4.4.10}$$

где при $i > 0$ функция

$$\Psi_i(x) = \Pi_{j=0}^{i-1} (1 - \alpha_j) \Psi_0(x).$$

Нетрудно видеть, что при $i = 0$ неравенство (4.4.10) выполнено. Как и в обосновании метода (4.1.12), нам необходимо вывести правила пересчета точек x_k, v_k и чисел A_k, φ_k, α_k, при которых из справедливости (4.4.10) для $i = k$ следовала бы справедливость этого неравенства и при $i = k + 1$.

Итак, пусть неравенство (4.4.10) справедливо при $i = k$. Зафиксируем произвольное $\alpha \in (0, 1)$ и положим

$$y_k = (1 - \alpha) x_k + \alpha v_k,$$

$$Q_k(\alpha) = \{ y \mid y = (1 - \alpha) x_k + \alpha x, \ x \in Q \}.$$

Тогда для любого $x \in Q$

$$f(x) + (1 - \alpha) \Psi_k(x) = \alpha f(x) +$$
$$+ (1 - \alpha) [f(x) + \Psi_k(x)] \geq \alpha \Phi(x, y_k) +$$

$$+ (1 - \alpha)\, \Phi(x_k , y_k) + 0.5\, (1 - \alpha)\, A_k \, \| x - v_k \|^2 \geq$$
$$\geq \Phi(\alpha x + (1 - \alpha) x_k , y_k) +$$
$$+ 0.5\, (1 - \alpha)\, A_k \, \| x - v_k \|^2 .$$

Обозначим правую часть последнего неравенства через $\xi(x)$. Нетрудно видеть, что $Q_k(\alpha) \subset Q \subset Q_k(\alpha^{-1})$. Поэтому

$$\xi^* = \min \{ \xi(x) \mid x \in Q \} = \min \{ \Phi(y , y_k) +$$
$$+ 0.5\, \alpha^{-2} (1 - \alpha)\, A_k \, \| y - y_k \|^2 \mid y \in Q(\alpha) \} \geq$$
$$\geq \min \{ \Phi(y , y_k) + 0.5\, \alpha^{-2} (1 - \alpha)\, A_k \, \| y - y_k \|^2$$
$$\mid y \in Q \} \equiv \xi_* .$$

Выберем $\alpha > 0$ из условия

$$(1 - \alpha)\, A_k = \alpha^2 F(L) . \qquad (4.4.11)$$

И пусть точка $T(L , y_k)$ - решение экстремальной задачи

$$\Phi(y, y_k) + 0.5\, F(L)\, \| y - y_k \|^2 \to \min \mid y \in Q . \qquad (4.4.12)$$

Тогда в силу неравенств (4.4.9),(4.4.2)

$$f(T(L , y_k)) \leq \xi_* \leq \xi^* .$$

Кроме того, в силу сильной выпуклости целевой функции задачи (4.4.12) для любых y из Q

$$\xi_* + 0.5\, F(L)\, \| y - T(L , y_k) \|^2 \leq$$
$$\leq \Phi(y, y_k) + 0.5\, F(L)\, \| y - y_k \|^2 .$$

Поэтому в силу соотношения (4.4.11) для всех $x \in Q(\alpha^{-1})$ (и тем более из Q)

$$\xi_* + 0.5\, (1 - \alpha)\, A_k \times$$
$$\times \| x - [v_k - \alpha^{-1} (y_k - T(L , y_k))] \|^2 \leq \xi(x) .$$

Выберем

$$\varphi_{k+1} = \xi_* , \quad A_{k+1} = (1 - \alpha)\, A_k ,$$
$$x_{k+1} = T(L, y_k),$$
$$v_{k+1} = v_k + \alpha^{-1} (x_{k+1} - y_k) .$$

Тогда в силу последнего неравенства получаем, что соотношение (4.4.10) будет выполнено и при $i = k + 1$.

Приведенные рассуждения обосновывают скорость сходимости следующего метода.

Метод $\mathfrak{M}(L , A , x_0)$ (4.4.13)

0. Полагаем $A_0 = A$, $v_0 = x_0$.

1. k -я итерация ($k \geq 0$).

а) Вычисляем $\alpha_k > 0$ из уравнения

$F(L) \alpha_k^2 = (1 - \alpha_k) A_k$;

б) полагаем

$y_k = \alpha_k v_k + (1 - \alpha_k) x_k$,

$x_{k+1} = T(L , y_k)$,

$v_{k+1} = v_k + \alpha_k^{-1} (x_{k+1} - y_k)$,

$A_{k+1} = (1 - \alpha_k) A_k$.

Итерация закончена.

Оценку скорости сходимости метода (4.4.13) дает

Теорема 4.4.2. Если метод $\mathfrak{M}(L, A, x_0)$ применяется для решения задачи (4.4.1) с функциями $f^{(i)} \in C^{1,1}(L^{(i)})$, $i = 1, \ldots , m$, и функцией $F(u)$, $u \in R^m$, удовлетворяющей условию (4.4.9), то при любом $k \geq 0$ справедлива оценка

$$f(x_k) - f^* \leq \frac{f(x_0) - f^* + 0.5 A \| x_0 - x^* \|^2}{[1 + A^{1/2} F(L)^{-1/2} k]^2}$$

Доказательство. Из вышеприведенных рассуждений следует, что для точек последовательности $\{ x_k \}_{k=0}^{\infty}$, вырабатываемой методом (4.4.13), справедливо неравенство

$f(x_k) - f(x^*) \leq \Psi_k(x^*) = \lambda_k \Psi_0(x^*)$,

где

$\lambda_k = \Pi_{i=0}^{k-1} (1 - \alpha_k) = A_k A^{-1}$

Положим $a_k = \alpha_k^{-1}$. Тогда в силу схемы метода (4.4.13) имеем

$A_{k+1} = (1 - \alpha_k) A_k = F(L) \alpha_k^2 = F(L) a_k^{-2}$

откуда получаем уравнение для величин a_k.

$a_k^{-2} = (1 - a_k^{-1}) a_{k-1}^{-2}$.

Следовательно, $a_k > a_{k-1}$ и

$$a_k = (a_k + a_{k-1}) (a_k - a_{k-1}) < 2 a_k (a_k - a_{k-1})$$

Таким образом,

$$a_k > a_{k-1} + 0.5 > a_0 + 0.5 k$$

Поэтому

$$\lambda_{k+1} = A^{-1} A_{k+1} = A^{-1} F(L) a_k^{-2} <$$
$$< A^{-1} F(L) [a_0 + 0.5 k]^{-2}$$

Заметим, что $F(L) A^{-1} = a_0 (a_0 - 1)$ и, следовательно,

$$A F(L)^{-1} (a_0^2 + a_0 k + 0.25 k^2) =$$
$$= 1 + A F(L)^{-1} [a_0 (k + 1) + 0.25 k^2] >$$
$$> 1 + 0.25 A F(L)^{-1} (k + 1)^2 .$$

Из этого неравенства, учитывая вид функции $\Psi_0(x)$, получаем нужную оценку. □

При реализации метода (4.4.13) наиболее трудоемкой операцией является вычисление точки минимума функции

$$F(f(y_k) + f'(y_k)(x - y_k)) + 0.5 F(L) \| x - y_k \|^2$$

по x на множестве Q. Во многих практически важных случаях множество Q - многогранник, а функция $F(u)$ однородна и допускает представление

$$F(u) = \max \{ \langle s , u \rangle \mid s \in G \} ,$$

где G - множество в пространстве R^m, задаваемое системой линейных неравенств. Примерами таких функций являются

$$F(u) = \max \{ u^{(i)} \mid i = 1 , \dots , m \} ,$$
$$F(u) = \sum_{i=1}^{m} u_+^{(i)} .$$

В этом случае вместо задачи (4.4.12) удобно решать задачу, двойственную к ней. А именно, пусть

$$Q = \{ x \in R^n \mid B x \le b \} ,$$

где B - $n \times N$ -матрица. И пусть (s^* , v^*) - решение следующей задачи:

$$\langle s , f(y_k) \rangle + \langle v , B y_k - b \rangle -$$
$$- 0.5 F(L)^{-1} \| B^T v + [f'(y_k)]^T s \|^2 \rightarrow \max$$

при условиях

$$s \in G , v \in R^N_+ . \tag{4.4.14}$$

Тогда решение $T(L , y_k)$ задачи (4.4.12) определяется по формуле

$$T(L , y_k) = y_k - F(L)^{-1} (B^T v^* + [f'(y_k)]^T s^*) .$$

Отметим, что задача (4.4.14) является задачей квадратичного программирования, для решения которой существуют достаточно эффективные алгоритмы. Некоторые из них будут рассмотрены в следующей главе.

ГЛАВА 5. ИТЕРАТИВНЫЕ МЕТОДЫ РЕШЕНИЯ ЗАДАЧ ЛИНЕЙНОГО И КВАДРАТИЧНОГО ПРОГРАММИРОВАНИЯ

5.1. Полиномиальные алгоритмы линейного программирования. Постановка задачи

Традиционным методом решения задач линейного программирования (ЛП) является симплексный алгоритм. Общепринятая эмпирическая оценка числа арифметических операций, необходимых такому методу для решения задачи ЛП, имеет кубический порядок роста в зависимости от размерности задачи. Однако к настоящему времени построено много тестовых задач на которых число итераций симплекс-метода имеет экспоненциальный порядок роста. Факт существования таких задач лишь в небольшой степени отразился на "практической" репутации симплекс-метода. Напротив, в теории методов решения задач ЛП такая ситуация привела к возникновению двух направлений исследования. В работах, относящихся к первому направлению, выясняется поведение симплекс-метода "в среднем" на классах задач ЛП с некоторыми фиксированными вероятностными характеристиками значений их коэффициентов. В исследованиях по второму направлению разрабатываются новые алгоритмы ЛП, принципиально отличающиеся от симплексного. Отметим, что многочисленные попытки построения метода решения общей задачи ЛП над полем действительных чисел с полиномиальной (по размерам задачи) оценкой числа арифметических операций, пока не увенчались успехом. Наилучшим полученным к настоящему времени результатом является полиномиальный метод стоимости порядка $m\, n^3 \times \ln n$ операций (m - число неравенств, n - число переменных) для случая, когда в каждое неравенство задачи входит не более двух переменных (случай вхождения трех переменных в неравенства так же труден, как общий). Кроме того, построены алгоритмы, которые при фиксированном числе переменных решают задачу ЛП за линейное по второму размеру число операций $C(n)m$, где $C(n)$, однако, растет быстрее экспоненты.

Начиная с 70-х годов в рамках проблематики алгоритмической сложности возникла так называемая битовая постановка проблемы построения полиномиального алгоритма решения задачи ЛП:

$$\min \{ \langle c, x \rangle \mid x \in R^n, x \geq 0, Ax = b \}, \tag{5.1.1}$$

где $c \in R^n$; $b \in R^m$; A - $m \times n$ матрица. В этой постановке входные коэффициенты задачи (5.1.1) считаются целыми, а размеры задачи оцениваются не парой (m, n), а тройкой параметров (m, n, l), где l - максимальное число двоичных знаков, отводимых на позиционную запись одного коэффициента. Требуется построить алгоритм с полиномиальным по битовой размерности (m, n, l) числом операций, которые выполнялись бы над числами полиномиальной по указанным параметрам битовой длины.

Положительный ответ на рассматриваемый вопрос получен Л.Г.Хачияном, который показал, что любой метод $\mathcal{M}$ приближенного решения задачи (5.1.1) с полиномиальной по m, n и $\ln 1/\varepsilon$ (ε - требуемая точность решения задачи) арифметической стоимостью может быть превращен в полиномиальный в битовом смысле алгоритм линейного программирования, если только исходный приближенный метод не является "патологическим" для его реализации достаточно вести вычисления с полиномиальной по указанным параметрам разрядностью. Более точно, был предложен некоторый алгоритм "округления", позволяющий по достаточно хорошему ε-приближенному решению задачи линейного программирования найти ее точное решение с затратой $m^2 n$ арифметических операций. Требования к "достаточно хорошему" приближенному решению формулируются следующим образом: оно должно иметь в каждой компоненте не более L знаков перед запятой, удовлетворять ограничениям задачи с точностью L знаков после запятой и обеспечивать достижение минимума целевой функции с точностью L знаков после запятой, где

$$L = \mathrm{const} \ln (m \Delta) , \tag{5.1.2}$$

а Δ - произвольная величина, мажорирующая модули всевозможных миноров матрицы ограничений задачи. Из неравенства Адамара следует, что $L \leqslant \mathrm{const}\, m (1 + \ln m)$, причем для соответствующей погрешности приближенного решения справедливо неравенство $\ln 1/\varepsilon \leqslant \mathrm{const}\, L$. Из этих двух неравенств ясно, что любой удовлетворяющий сформулированным выше требованиям метод действительно перестраивается с учетом алгоритма округления в точный полиномиальный алгоритм. Отметим, что операции алгоритма "округления" также могут выполняться с точностью L знаков.

Л.Г.Хачияном в качестве приближенного метода $\mathcal{M}$ был использован метод описанных эллипсоидов (см. §3.4) с некоторой модификацией, использующей "раздутие" описанных эллипсоидов для гашения ошибок округления на итерации. С помощью этого метода было показано, что для решения задачи ЛП достаточно выполнить $O(n m^3 L)$ арифметических операций над L - разрядными числами, где L определено в (5.1.2).

После доказательства полиномиальной разрешимости задачи линейного программирования с помощью метода эллипсоидов появилось несколько других приближенных методов, с помощью которых можно получить тот же результат (метод Кармаркара, метод проекций А.С.Немировского, методы Ренегара, Вадьи и др.). Однако теперь уже целью является не только улучшение теоретических оценок, но и построение методов, способных на практике конкурировать с симплекс-методом. Ниже описывается несколько новых полиномиальных (в битовом смысле) алгоритмов решения задач линейного и квадратичного программирования. Теоретически наиболее эффективные из этих методов имеют кубическую зависимость числа арифметических операций от размеров решаемой задачи.

5.2. Двойственные алгоритмы решения систем линейных неравенств. Вспомогательные результаты

Рассмотрим задачу нахождения произвольного решения системы линейных неравенств в следующей постановке,

$$A x = b \ , \quad x \in \mathcal{K} = \{ x \in R^n \mid |x^{(i)}| < 1 \} . \qquad (5.2.1)$$

Здесь A - $m \times n$ матрица. Не ограничивая общности будем считать, что матрица A имеет полный ранг.

Предположим, что система (5.2.1) является θ - устойчивой, т.е. что многообразие $\mathcal{L} \equiv \{ x \mid A x = b \}$ пересекается с множеством $(1 - \theta) \mathcal{K}$. Это предположение, в частности, обеспечивает существование точки x' такой, что $A x' = b$ и $x' + \theta \mathcal{K} \subset \mathcal{K}$. Можно показать, что предположение об θ-устойчивости системы (5.2.1) эквивалентно требованию конечности длины информации L с $\theta = O(2^{-L})$.

Рассмотрим функцию

$$f(t) = - | t | - \ln(1 - | t |) .$$

Эта функция является сильно выпуклой, неотрицательной и дважды дифференцируемой на $(-1, 1)$, причем $f(0) = 0$. Заметим также, что

$$f'(t) = t(1 - |t|)^{-1}; \quad f''(t) = (1 - |t|)^{-2}$$

Введем сопряженную функцию

$$\varphi(\tau) = \max\{\tau t - f(t) \mid |t| < 1\}.$$

Нетрудно видеть, что максимум внутренней функции достигается при значении $t(\tau) = \tau(1 + |\tau|)^{-1}$. Поэтому

$$\varphi(\tau) = \tau^2 (1 + |\tau|)^{-1} + |\tau| (1 + |\tau|)^{-1} -$$
$$- \ln(1 + |\tau|) = |\tau| - \ln(1 + |\tau|).$$

Функция $\varphi(\tau)$ будет выпуклой, неотрицательной, дважды непрерывно дифференцируемой на всей прямой, $\varphi(0) = 0$. Кроме того,

$$\varphi'(\tau) = \tau(1 + |\tau|)^{-1}; \quad \varphi''(\tau) = (1 + |\tau|)^{-2}.$$

Заменим задачу нахождения произвольного решения системы (5.2.1) на следующую экстремальную задачу:

$$\min\{F(x) \equiv \sum_{i=1}^{n} f(x^{(i)}) \mid x \in \mathcal{L}\}. \tag{5.2.2}$$

В силу θ - устойчивости системы (5.2.1) такая задача корректна. Сформулируем задачу, двойственную к (5.2.2). Для этого выпишем функцию Лагранжа

$$L(x, y) = \langle y, Ax - b\rangle - F(x)$$

и положим

$$\Phi(y) = \max\{L(x, y) \mid x \in \mathcal{K}\}.$$

Тогда задача (5.2.2) будет эквивалентна следующей:

$$\min\{\Phi(y) \mid y \in R^m\}. \tag{5.2.3}$$

Выпишем явные формулы для функции $\Phi(y)$ и ее производных:

$$\Phi(y) = -\langle b, y\rangle + \max\{\langle y, Ax\rangle - F(x) \mid$$
$$x \in \mathcal{K}\} = \sum_{i=1}^{n} \varphi(\langle a_i, y\rangle) - \langle b, y\rangle \equiv$$
$$\equiv \sum_{i=1}^{n} [\,|\langle a_i, y\rangle| - \ln(1 + |\langle a_i, y\rangle|)\,] - \langle b, y\rangle;$$

$$\Phi'(y) = \sum_{i=1}^{n} \frac{\langle a_i, y\rangle}{1 + |\langle a_i, y\rangle|} a_i - b;$$

$$\Phi''(y) = \sum_{i=1}^{n} \frac{a_i\ a_i^T}{(1 + |\langle a_i\ ,\ y \rangle|)^2}$$

Двойственные методы решения задачи (5.2.1) сначала строят удовлетворительное приближение y' к решению задачи (5.2.3), а затем по нему формируют решение исходной задачи. Корректность такого подхода видна, в частности, из следующей формулы:

$$\Phi'(y) \equiv A\, x(y) - b\ ,$$

где

$$x(y) = (x^{(1)}(y)\ ,\ \dots\ ,\ x^{(n)}(y))^T \in R^n\ ,$$

$$x^{(i)}(y) = \langle a_i\ ,\ y \rangle / (1 + | \langle a_i\ ,\ y \rangle |)\ ,$$

$$i = 1,\ \dots\ ,\ n\ .$$

Заметим, что $x(y) \in \mathcal{K}$ при любых y из R^m.

Для обоснования скорости сходимости методов минимизации функции $\Phi(y)$ нам потребуется несколько вспомогательных утверждений. Пусть Φ^* - минимальное значение функции $\Phi(y)$ на R^m.

Лемма 5.2.1. Если задача (5.2.1) θ-устойчива, то минимум функции $\Phi(y)$ в задаче (5.2.3) достигается в некоторой точке y^*, причем

$$\Phi(y^*) \geq - n\, [\, \ln (1/\theta) + \theta - 1\,]\ , \qquad (5.2.4)$$

$$\langle b\ ,\ y^* \rangle \leq (1 - \theta)\, n / \theta\ . \qquad (5.2.5)$$

Д о к а з а т е л ь с т в о. Заметим, что в силу определения функции $\Phi(y)$ справедливо равенство $\Phi^* + F^* = 0$, где F^* - оптимальное значение функции $F(x)$ в задаче (5.2.2). Мы предположили, что существует точка $x' \in \mathcal{L} \cap (1 - \theta)\mathcal{K}$. Поэтому

$$F^* \leq F(x') \leq \max \{\, F(x) \mid x \in (1 - \theta)\, \mathcal{K}\, \} =$$

$$= n \max \{\, f(t) \mid\ | t | \leq 1 - \theta\, \} =$$

$$= n\, [\, \ln (1/\theta) + \theta - 1\,]\ .$$

Далее, в силу θ-устойчивости задачи (5.2.1)

$$\sum_{i=1}^{n} |\langle a_i, y\rangle| - \langle b, y\rangle =$$
$$= \max\{\sum_{i=1}^{n} \lambda^{(i)} \langle a_i, y\rangle - \langle b, y\rangle \mid \lambda \in \mathcal{K}\} \geq$$
$$\geq \max\{\langle A\lambda - b, y\rangle \mid \lambda \in x' + \theta\mathcal{K}\} =$$
$$= \max\{\langle A\lambda, y\rangle \mid \lambda \in \theta\mathcal{K}\} =$$
$$= \theta \sum_{i=1}^{n} |\langle a_i, y\rangle| .$$

Поэтому

$$0 = \langle \Phi'(y^*), y^*\rangle =$$
$$= \sum_{i=1}^{n} |\langle a_i, y^*\rangle|^2 (1 + |\langle a_i, y^*\rangle|)^{-1} -$$
$$- \langle b, y^*\rangle = \sum_{i=1}^{n} [\, |\langle a_i, y^*\rangle| - |\langle a_i, y^*\rangle| /$$
$$/(1 + |\langle a_i, y^*\rangle|)\,] - \langle b, y^*\rangle \geq$$
$$\geq \sum_{i=1}^{n} |\langle a_i, y^*\rangle| - \langle b, y^*\rangle - n \geq$$
$$\geq (1-\theta)^{-1}\langle b, y^*\rangle - \langle b, y^*\rangle - n \quad \square$$

Всюду далее используются следующие обозначения Пусть $d \in R^n$, $d > 0$. Положим

$$D(d) = \mathrm{diag}(d),$$
$$G(d) = A\, D(d)^2 A^T,$$
$$H(d) = [\, G(d)\,]^{-1},$$
$$S(d, y_0, \alpha) = \{ y \in R^m \mid \langle G(d)(y - y_0), y - y_0\rangle \leq \alpha^2 \}.$$

Для y из R^m введем функцию $d(y)$ такую, что

$$d(y)^{(i)} = (1 + |\langle a_i, y\rangle|)^{-1}, \quad i = 1, \ldots, n.$$

Ясно, что в таких обозначениях $G(d(y)) \equiv \Phi''(y)$.

Будем говорить, что функция $\Phi(y)$ является (d, α, τ)-согласованной в точке y_0, если для любого $y \in S(d, y_0, \alpha)$ справедливы неравенства

$$(1+\tau)^{-1} d \leq d(y) \leq (1-\tau)^{-1} d$$

Заметим, что из (d, α, τ) - согласованности функции $\Phi(y)$ в точке y_0 очевидным образом выводятся неравенства

$(1 + \tau)^{-2} G(d) \le \Phi''(y) \le (1 - \tau)^{-2} G(d)$,

$(1 - \tau)^2 H(d) \le [\Phi''(y)]^{-1} \le (1 + \tau)^2 H(d)$,

справедливые при всех y из $S(d, y_0, \alpha)$. Другими словами, функция $\Phi(y)$ является сильно выпуклой на множестве $S(d, y_0, \alpha)$ с константой $(1 + \tau)^{-2}$ (в метрике, задаваемой матрицей $G(d)$), и ее градиент удовлетворяет условию Липшица на том же множестве и в той же метрике с константой $(1 - \tau)^{-2}$.

Нетрудно видеть, что функция $\Phi(y)$ является $(d(y), 0, 0)$ - согласованной в любой точке y из R^m. Менее тривиальный случай описывается в следующей лемме.

Лемма 5.2.2. Если функция $\Phi(y)$ является $(d, 0, \tau)$ - согласованной в некоторой точке y_0, $0 \le \tau < 1$, то она также будет и $(d, \alpha, \alpha + \tau)$ - согласованной в той же точке при любом α из множества $[0, 1 - \tau)$.

Доказательство. Действительно, из включения $y \in S(d, y_0, \alpha)$ следует, что

$|\langle a_i, y - y_0 \rangle| \le \alpha / d^{(i)}$, $i = 1, \dots, n$

Кроме того, из $(d, 0, \tau)$ -согласованности функции $\Phi(y)$ в точке y_0 имеем

$(1 + \tau)^{-1} d \le d(y_0) \le (1 - \tau)^{-1} d$

Отсюда получаем

$(1 + \alpha + \tau)^{-1} d \le d(y) \le (1 - \alpha - \tau)^{-1} d$ □

Следствие 5.2.1. Функция $\Phi(y)$ будет $(d(y_0), \alpha, \alpha)$ -согласованной в любой точке $y_0 \in R^m$ при всяком α из $[0, 1)$.

Лемма 5.2.3. Если функция $\Phi(y)$ является (d, α, τ) -согласованной в точке y_0, то при любых y_1, y_2 из $S(d, y_0, \alpha)$ справедливы неравенства

$$\Phi(y_1) \le \Phi(y_2) + \langle \Phi'(y_2), y_1 - y_2 \rangle +$$
$$+ 0.5\,(1-\tau)^{-2} \langle G(d)(y_1 - y_2), y_1 - y_2 \rangle . \qquad (5.2.6)$$
$$\langle \Phi'(y_1) - \Phi'(y_2), y_1 - y_2 \rangle \ge$$
$$\ge 0.5\,(1+\tau^2)^{-1} [\langle G(d)(y_1 - y_2), y_1 - y_2 \rangle +$$
$$+ (1-\tau^2)^2 \langle H(d)(\Phi'(y_1) - \Phi'(y_2)),$$
$$\Phi'(y_1) - \Phi'(y_2) \rangle] . \qquad (5.2.7)$$

Справедливость этого утверждения сразу следует из сильной выпуклости функции $\Phi(y)$ на множестве $S(d, y_0, \alpha)$ в метрике, задаваемой матрицей $G(d)$, с константой $(1 + \tau)^{-2}$ и ограниченности спектра матриц вторых производных на том же множестве и в той же метрике константой $(1 - \tau)^{-2}$ (см. теорему 1.4.5). □

Лемма 5.2.4. Если функция $\Phi(y)$ является $(d, 0, \tau)$-согласованной в точке y_0 и
$$\langle H(d)\,\Phi'(y_0), \Phi'(y_0) \rangle \le (1 + \tau)^{-2} ,$$
то точка
$$x_*(d, y_0) = x(y_0) - D(d)^2 A^T H(d)\,\Phi'(y_0)$$
будет решением задачи (5.2.1).

Д о к а з а т е л ь с т в о. Заметим, что
$$A\, x_*(d(y_0), y_0) = b .$$
Кроме того,
$$\Sigma_{i=1}^{n} (x_*(d, y_0)^{(i)} - x(y_0)^{(i)})^2 /$$
$$/ (1 - | x(y_0)^{(i)} |)^2 \equiv$$
$$\equiv \| D(d(y_0))^{-1}(x_*(d, y_0) - x(y_0)) \|^2 \le$$
$$\le (1 + \tau)^2 \| D(d)^{-1} (x_*(d, y_0) - x(y_0)) \|^2 \equiv$$
$$\equiv (1 + \tau)^2 \langle H(d)\,\Phi'(y_0), \Phi'(y_0) \rangle \le 1 .$$
Поэтому $x_*(d, y_0) \in \mathcal{K}$. □

Утверждение леммы 5.2.4 дает удобный критерий прерывания, который используется во всех описываемых ниже двойственных методах. В свою очередь, выполнение условий леммы

5.2.4 в некоторых случаях удается вывести из следующего факта.

Лемма 5.2.5. Если $y^* \in S(d, y_0, \nu)$ и функция $\Phi(y)$ является (d, ν, τ)-согласованной в точке y_0, то

$$\langle H(d)\Phi'(y_0), \Phi'(y_0)\rangle \le \nu^2/(1-\tau)^4.$$

Доказательство. Действительно,

$$\Phi'(y^*) = 0.$$

Поэтому в силу неравенства (5.2.7) имеем

$$\langle H(d)\Phi'(y_0), \Phi'(y_0)\rangle \le$$
$$\le (1-\tau)^{-4}\langle G(d)(y^* - y_0), y^* - y_0\rangle \le$$
$$\le \nu^2/(1-\tau)^4. \quad \square$$

Теперь мы можем перейти к обоснованию двойственных методов решения задачи (5.2.1).

5.3. Двойственный метод Ньютона

Посмотрим, каково будет изменение функции $\Phi(y)$ при смещении из точки y_0 в точку y_+ по ньютоновскому направлению:

$$y_+ = y_0 - \alpha \frac{[\Phi''(y_0)]^{-1}\Phi'(y_0)}{\langle[\Phi''(y_0)]^{-1}\Phi'(y_0), \Phi'(y_0)\rangle^{1/2}}.$$

В этом случае $y_+ \in S(d(y_0), y_0, \alpha)$. Поэтому в силу следствия 5.2.1 и неравенства (5.2.6) имеем

$$\Phi(y_+) \le \Phi(y_0) + \langle \Phi'(y_0), y_+ - y_0\rangle +$$
$$+ 0.5(1-\alpha)^{-2}\langle \Phi''(y_0)(y_+ - y_0), y_+ - y_0\rangle =$$
$$= \Phi(y_0) - \alpha\langle[\Phi''(y_0)]^{-1}\Phi'(y_0), \Phi'(y_0)\rangle^{1/2} +$$
$$+ 0.5\,\alpha^2(1-\alpha)^{-2} \qquad (5.3.1)$$

Из этого неравенства можно сделать следующий вывод: если точка $x_*(d(y_0), y_0)$ не является решением задачи (5.2.1), то в этом случае в силу лемм 5.2.2, 5.2.4

$$\langle [\Phi''(y_0)]^{-1} \Phi'(y_0), \Phi'(y_0)\rangle > 1 ,$$

и поэтому с помощью ньютоновского шага удается получить точку y_+, значение целевой функции в которой (при правильно выбранном α) на абсолютную константу меньше, чем в точке y_0. Таким образом, мы приходим к следующему методу.

Метод $\mathcal{N}(\alpha)$

0. Полагаем $y_0 = 0 \in R^m$.

1. k -я итерация ($k \geq 0$).

   а) Вычисляем матрицу $[\Phi''(y_k)]^{-1}$. Если

$$\langle [\Phi''(y_k)]^{-1} \Phi'(y_k), \Phi'(y_k) \rangle \leq 1 ,$$

то полагаем $x_* = x_*(d(y_k), y_k)$ и останавливаемся.

   б) Полагаем

$$y_{k+1} = y_k - \alpha \frac{[\Phi''(y_k)]^{-1} \Phi'(y_k)}{\langle[\Phi''(y_k)]^{-1} \Phi'(y_k), \Phi'(y_k)\rangle^{1/2}}$$

Итерация закончена.

Теорема 5.3.1. Если задача (5.2.1) θ -устойчива, то метод $\mathcal{N}(\alpha)$ с $\alpha \in (0, 0.5)$ не более чем за N итераций, где

$$N = 2(1 - \alpha)^2 [\alpha(2 - \alpha)(1 - 2\alpha)]^{-1} n \ln(1/\theta) ,$$

выдаст в качестве ответа точку x_* - решение задачи (5.2.1). Близкое к оптимальному значение $\alpha = 0.25$.

Доказательство. Заметим, что если на некоторой итерации метода $\mathcal{N}(\alpha)$ сработал его критерий прерывания, то в силу леммы 5.2.4 в качестве ответа будет выдано решение задачи (5.2.1).

Пусть теперь за N итераций метода этот критерий ни разу не сработал. Но тогда в силу неравенства (5.3.1) и способа прерывания метода $\mathcal{N}(\alpha)$ при всех k, $0 \le k \le N - 1$, выполнялось неравенство

$$\Phi(y_k) - \Phi(y_{k+1}) \ge$$
$$\ge \alpha \langle [\Phi''(y_k)]^{-1} \Phi'(y_k), \Phi'(y_k) \rangle^{1/2} -$$
$$- 0.5 \alpha^2 (1 - \alpha)^{-2} \ge 0.5 \alpha (2 - \alpha)(1 - 2\alpha)(1 - \alpha)^{-2}.$$

Суммируя эти неравенства по всем k, $0 \le k \le N - 1$, и пользуясь леммой 5.2.1, получаем

$$0.5 \alpha (2 - \alpha)(1 - 2\alpha)(1 - \alpha)^{-2} N \le$$
$$\le \Phi(y_0) - \Phi(y_N) \le - \Phi^* \le n \ln(1/\theta).$$

Осталось заметить, что в силу включения $\alpha \in (0, 0.5)$ в этом неравенстве множитель перед N положителен. □

В теореме 5.3.1 мы обосновали вариант метода Ньютона с постоянным нормированным шагом. Понятно, что аналогичное обоснование можно провести и для других версий метода Ньютона, использующих различные процедуры одномерного поиска.

Если для обращения матрицы $\Phi''(y_k)$ на каждой итерации метода $\mathcal{N}(\alpha)$ пользоваться стандартными методами линейной алгебры, то для выполнения такой итерации потребуется порядка $O(m^2 n + m^3)$ арифметических операций. Поэтому общая трудоемкость решения θ-устойчивой задачи в этом случае не превысит величины порядка $O(m^2 n^2 \ln(1/\theta))$. Оказывается, можно модифицировать метод $\mathcal{N}(\alpha)$ так, что отпадает необходимость полностью обращать на каждой итерации матрицу $\Phi(y_k)$. В результате у модифицированного метода оценка трудоемкости решения задачи (2.1) снижается до величины порядка $O(m^2 n^{1.5} \ln(1/\theta))$. Перейдем к описанию этой модификации.

Зафиксируем произвольные константы $\alpha > 0$ и $\tau \in (0, 1)$. Рассмотрим две последовательности векторов $\{y_k\}_{k=0}^{\infty} \subset R^m$ и $\{d_k\}_{k=0}^{\infty} \subset R^n$, такие, что

1 . $y_0 = 0$; $d_0^{(i)} = 1$, $i = 1,\dots,n$.

2 . Очередные векторы y_{k+1} и d_{k+1} выбираются из следующих условий:

а) y_{k+1} может быть произвольной точкой из множества $S(d_k, y_k, \alpha)$;

б) $$d_{k+1}^{(j)} = \begin{cases} d(y_{k+1})^{(j)}, & \text{если } j \in \mathcal{J}(k), \\ d_k^{(j)}, & \text{в противном случае;} \end{cases}$$

здесь

$$\mathcal{J}(k) = \{ j \mid d_k^{(j)} > (1 + \tau) d(y_{k+1})^{(j)} \} \cup$$
$$\cup \{ j \mid d_k^{(j)} < (1 - \tau) d(y_{k+1})^{(j)} \} .$$

Лемма 5.3.1. При любом $N > 1$ справедлива оценка
$$\sum_{k=0}^{N-1} |\mathcal{J}(k)| < \alpha \tau^{-1} n^{1/2} N ,$$
где $|\mathcal{J}(k)|$ - мощность множества $\mathcal{J}(k)$.

Доказательство. В силу условия а) при любом $N > 1$ имеем

$$N \alpha^2 > \sum_{i=0}^{N-1} \langle G(d_i)(y_{i+1} - y_i), y_{i+1} - y_i \rangle \equiv$$
$$\equiv \sum_{i=0}^{N-1} \sum_{j=1}^{n} [d_i^{(j)}]^2 \langle a_j, y_{i+1} - y_i \rangle^2 =$$
$$= \sum_{j=1}^{n} \sum_{i=0}^{N-1} [d_i^{(j)}]^2 \langle a_j, y_{i+1} - y_i \rangle^2 .$$

Обозначим внутренние суммы этого выражения через S_j . Зафиксируем произвольный номер j, $1 < j < n$. Нас будет интересовать величина $m(j)$ - число изменений j-й координаты векторов d_i , $i = 0, \dots, N - 1$. Введем целочисленную функцию $s(l)$, $l = 0, \dots, m(j)$, значение которой соответствуют номеру точки y_i , при формировании которой произошел l-й пересчет, $s(0) = 0$. Заметим, что при всех q, $s(l) < q < s(l + 1) - 1$, справедливо тождество $d_q^{(j)} \equiv \equiv d_{s(l)}^{(j)}$. Поэтому

$$S_j > \sum_{l=0}^{m(j)-1} [d_{s(l)}^{(j)}]^2 \sum_{q=s(l)}^{s(l+1)-1} \langle a_j, y_{q+1} - y_q \rangle^2 >$$
$$> \sum_{l=0}^{m(j)-1} [d_{s(l)}^{(j)}]^2 \langle a_j, y_{s(l+1)} - y_{s(l)} \rangle^2 /$$

$/ (s(l+1) - s(l)) \geq$

$\geq \sum_{l=0}^{m(j)-1} [d^{(j)}_{s(l)}]^2 [|< a_j , y_{s(l+1)} >| -$

$- |< a_j , y_{s(l)} >|]^2 / (s(l+1) - s(l))$

В силу способа пересчета векторов d_i

$[d^{(j)}_{s(l)}]^{-1} = |< a_j , y_{s(l)} >| + 1 ,$

$[d^{(j)}_{s(l+1)}]^{-1} = |< a_j , y_{s(l+1)} >| + 1$

Таким образом,

$S_j \geq \sum_{l=0}^{m(j)-1} [1 - d^{(j)}_{s(l)} / d^{(j)}_{s(l+1)}]^2 / (s(l+1) - s(l))$.

Из критерия пересчета векторов d_i следует, что существует две возможности:

либо $d^{(j)}_{s(l)} / d^{(j)}_{s(l+1)} \geq 1 + \tau$,

либо $d^{(j)}_{s(l)} / d^{(j)}_{s(l+1)} \leq 1 - \tau$

В любом случае $| 1 - d^{(j)}_{s(l)} / d^{(j)}_{s(l+1)} | \geq \tau$. Поэтому

$S_j \geq \tau^2 \sum_{l=0}^{m(j)-1} (s(l+1) - s(l))^{-1}$.

Заметим, что сумма по l чисел $s(l+1) - s(l)$ не превосходит N , причем число слагаемых в сумме равно $m(j)$.

Воспользовавшись теперь соотношением

$\min\{ \sum_{l=1}^{m} t_l^{-1} \mid t_l > 0 , \sum_{l=1}^{m} t_l \leq T \} = m^2 / T.$

получаем, что для любого j , $1 \leq j \leq n$,

$S_j \geq \tau^2 m(j)^2 N^{-1}$

Отсюда следует, что

$\sum_{j=1}^{n} m(j)^2 \leq \tau^{-2} N \sum_{j=1}^{n} S_j \leq N^2 \alpha^2 \tau^{-2}$

$\sum_{k=0}^{N-1} | J(k) | \equiv \sum_{j=1}^{n} m(j) \leq$

$\leq [n \sum_{j=1}^{n} m(j)^2]^{1/2} \leq N n^{1/2} \alpha \tau^{-1}$ □

Теперь мы можем описать модифицированную версию метода $\mathcal{N}(\alpha)$.

Метод $\mathcal{N}_1(\alpha , \tau)$

0 Полагаем

$y_0 = 0 \in R^m, \; H_0 = [A A^T]^{-1}, \; d_0 = (1,\dots,1)^T \in R^n$

1. k -я итерация ($k > 0$).

а) проверяем выполнение неравенства

$\langle H_k \Phi'(y_k), \Phi'(y_k) \rangle < (1 + \tau)^{-2}$

Если оно выполнено, то полагаем $x_* = x_*(d_k , y_k)$ и останавливаемся;

б) полагаем

$$y_{k+1} = y_k - \alpha \frac{H_k \Phi'(y_k)}{\langle H_k \Phi'(y_k), \Phi'(y_k) \rangle^{1/2}}$$

в) формируем множество

$$\mathfrak{J}(k) = \{ j \mid d_k^{(j)} > (1 + \tau) d(y_{k+1})^{(j)} \} \cup$$
$$\cup \{ j \mid d_k^{(j)} < (1 - \tau) d(y_{k+1})^{(j)} \} ;$$

г) производим пересчет вектора d_k :

$$d_{k+1}^{(j)} = \begin{cases} d(y_{k+1})^{(j)} , & \text{если } j \in \mathfrak{J}(k) , \\ d_k^{(j)} & \text{в противном случае;} \end{cases}$$

д) с помощью формул одноранговой коррекции производим пересчет матрицы H_k :

$$H_{k+1} = [H_k^{-1} + \textstyle\sum_{j \in \mathfrak{J}(k)} [(d_{k+1}^{(j)})^2 - (d_k^{(j)})^2] a_j a_j^T]^{-1} .$$

Итерация закончена.

<u>Теорема 5.3.2.</u> Если задача (5.2.1) θ - устойчива, то метод $\mathcal{N}_1(\alpha, \tau)$ с параметрами $\alpha , \tau > 0$ такими, что

$\alpha + \tau < 1$, $\alpha (1 + \tau)^2 < 2 (1 - \alpha - \tau)^2$,

не более чем за N итераций, где

$N = n \ln(1/\theta)[\alpha (1 + \tau)^{-2} - 0.5 \alpha^2 (1 - \alpha - \tau)^{-2}]^{-1}$,

выдаст в качестве ответа точку x_* - решение задачи (5.2.1). При этом за любые k итераций, $0 < k < N$, число одноранговых коррекций матрицы H_k не превзойдет $k n^{1/2} \alpha \tau^{-1}$. Таким образом, общая трудоемкость получения решения задачи (5.2.1) методом $\mathcal{N}_1(\alpha, \tau)$ оценивается величиной порядка $O(m^2 n^{1.5} \ln(1/\theta))$.

Д о к а з а т е л ь с т в о. Прежде всего заметим, что в силу способа пересчета матрицы H_k и вектора d_k справедливы соотношения:

$H_k = H(d_k)$,

$(1 + \tau)^{-1} d_k \le d(y_k) \le (1 - \tau)^{-1} d_k$.

Отсюда, в частности, следует, что во всех точках y_k функция $\Phi(y)$ будет $(d_k, 0, \tau)$ -согласованной (а значит, в силу леммы 5.2.2 и $(d_k, \alpha, \alpha + \tau)$ - согласованной). Поэтому в силу леммы 5.2.4 критерий прерывания метода $\mathcal{N}_1(\alpha, \tau)$ корректен и даст в случае его выполнения решение задачи (5.2.1). Предположим теперь, что этот критерий не сработал за первые N итераций. Тогда из $(d_k, \alpha, \alpha + \tau)$ -согласованности функции $\Phi(y)$ в точках y_k и неравенства (5.3.1) получаем

$$\Phi(y_{k+1}) \le \Phi(y_k) + \langle \Phi'(y_k), y_{k+1} - y_k \rangle +$$
$$+ 0.5 (1 - \alpha - \tau)^{-2} \langle G(d_k) (y_{k+1} - y_k), y_{k+1} - y_k \rangle =$$
$$= \Phi(y_k) - \alpha \langle H_k \Phi'(y_k), \Phi'(y_k) \rangle^{1/2} +$$
$$+ 0.5 \alpha^2 (1 - \alpha - \tau)^{-2} \le$$
$$\le \Phi(y_k) - \alpha [(1 + \tau)^{-2} - 0.5 \alpha (1 - \alpha - \tau)^{-2}]$$

при всех $k \ge 0$. Поэтому в силу леммы 5.2.1 число таких итераций не может быть больше, чем

$$n \ln(1/\theta)[\alpha (1 + \tau)^{-2} - 0.5 \alpha^2 (1 - \alpha - \tau)^{-2}]^{-1}$$

Осталось заметить, что способ пересчета векторов y_k и d_k удовлетворяет предположениям леммы 5.3.1. Поэтому за любые k итераций; $0 \le k \le N - 1$, число одноранговых коррекций матрицы H_k не превзойдет величины $(k + 1) n^{1/2} \alpha \tau^{-1}$

□

Отметим, что в методе $\mathcal{N}_1(\alpha, \tau)$ можно использовать различные стратегии выбора параметров α и τ. В простейшем случае их можно задавать абсолютными константами, не зависящими от размерности задачи. В то же время можно поступить и по-другому. Так, например, если выбрать параметр α равным $O(n^{-1/2})$, то оценка общего числа итераций метода возрастет и будет иметь порядок $O(n^{1.5}\ln(1/\theta))$. Однако в этом случае на каждой итерации будет производиться в среднем постоянное число одноранговых коррекций. Поэтому порядок общей оценки трудоемкости метода при таком выборе параметров не изменится.

5.4. Двойственный метод параллельных траекторий

Рассмотрим траекторию

$y^*(\xi) = \operatorname{argmin}\{\Phi(y) \mid \langle b, y\rangle = \xi\}$, $\xi > 0$.

Нетрудно видеть, что $y^*(0) = 0$. Можно показать, что при некотором $\xi^* > 0$ траектория $y^*(\xi)$ проходит через точку минимума функции $\Phi(y)$: $y^*(\xi^*) = y^*$. В этом и следующем параграфах будут рассмотрены методы, основанные на "отслеживании" точек траектории $y^*(\xi)$.

<u>Лемма 5.4.1.</u> Если функция $\Phi(y)$ является $(d, 0, \tau)$-согласованной в точке $y_0 \in R^m$, то

$\langle H(d)b, b\rangle^{1/2} \geq (1+\tau)^{-1} n^{-1/2} \langle b, y_0\rangle$.

Справедливость этого утверждения следует из неравенства Коши-Буняковского и того факта, что

$(1+\tau)^{-2}\langle G(d)y_0, y_0\rangle \leq \langle G(d(y_0))y_0, y_0\rangle \leq n$

□

Зафиксируем произвольную точку $y_0 \in R^m$. Предположим, что функция $\Phi(y)$ является $(d; \alpha, \tau)$-согласованной в этой точке.

Лемма 5.4.2. Пусть $\langle y_0, b \rangle = \xi > 0$ и точка $y^*(\xi)$ принадлежит множеству $S(d, y_0, \varkappa\alpha)$ при некотором $\varkappa \in (0, 1)$. Положим

$$y(\delta) = y_0 + \delta\alpha \frac{H(d)b}{\langle H(d)b, b\rangle^{1/2}};$$

$$\xi(\delta) = \langle y(\delta), b\rangle,$$

где параметр δ удовлетворяет неравенству

$$0 < \delta < (1 - \varkappa)[1 + 0.5(1 + \tau)^2(1 - \tau)^{-2}]^{-1}$$

Тогда точка $y^*(\xi(\delta)) \in S(d, y_0, \alpha)$, причем

$$\langle G(d)(y^*(\xi(\delta)) - y(\delta)), y^*(\xi(\delta)) - y(\delta)\rangle <$$
$$< \alpha^2[\varkappa + 0.5\delta(1 + \tau)^2(1 - \tau)^{-2}]^2 < \alpha^2(1 - \delta)^2$$

Доказательство. Положим

$$y'(\delta) = y^*(\xi) + \delta\alpha \frac{H(d)b}{\langle H(d)b, b\rangle^{1/2}}$$

$$y'_*(\delta) = \operatorname{argmin}\{\Phi(y) \mid \langle b, y - y'(\delta)\rangle = 0,$$
$$y \in S(d, y_0, \alpha)\}.$$

Заметим, что при всех $\delta \in [0, 1 - \varkappa]$ точка $y'(\delta)$ заведомо принадлежит множеству $S(d, y_0, \alpha)$. Поэтому в силу леммы 5.2.3

$$(1 + \tau)^{-2}\langle G(d)(y'(\delta) - y'_*(\delta)), y'(\delta) - y'_*(\delta)\rangle <$$
$$< \langle \Phi'(y'(\delta)) - \Phi(y'_*(\delta)), y'(\delta) - y'_*(\delta)\rangle <$$
$$< \langle \Phi'(y'(\delta)), y'(\delta) - y'_*(\delta)\rangle.$$

Введем матрицу

$$P = I - b\,b^T H(y_0) / \langle H(y_0)b, b\rangle.$$

Нетрудно видеть, что

$$P^T(y'(\delta) - y'_*(\delta)) = y'(\delta) - y'_*(\delta), \quad Pb = 0,$$

т.е. $P\Phi'(y^*(\xi)) = 0$. Следовательно,

$$(1 + \tau)^{-2}\langle G(d)(y'(\delta) - y'_*(\delta)), y'(\delta) - y'_*(\delta)\rangle <$$
$$< \langle P\Phi'(y'(\delta)), y'(\delta) - y'_*(\delta)\rangle <$$
$$< \langle G(d)(y'(\delta) - y'_*(\delta)), y'(\delta) - y'_*(\delta)\rangle^{1/2} \times$$

$\times \langle H(d) P \Phi'(y'(\delta)), P \Phi'(y'(\delta)) \rangle^{1/2}$

Таким образом,

$$(1+\tau)^{-4} \langle G(d)(y'(\delta) - y'_*(\delta)), y'(\delta) - y'_*(\delta) \rangle \le$$
$$\le \langle H^P(d) \Phi'(y'(\delta)), \Phi'(y'(\delta)) \rangle ,$$

где $H^P(d) = P^T H(d) P \equiv H(d) - H(d) b b^T H(d)/\langle H(d) b, b\rangle$.

Заметим, что

$$\Phi'(y'(\delta)) = \Phi'(y^*(\xi)) + \delta [\int_0^1 G(d(y^*(\xi) + \theta \delta s)) d\theta] s,$$

где $s = \alpha H(d) b / \langle H(d) b, b \rangle^{1/2}$. Обозначив интегральную матрицу (заключенную в квадратные скобки) через G', получим

$$\langle H^P(d) \Phi'(y'(\delta)), \Phi'(y'(\delta)) \rangle =$$
$$= \langle H^P(d)(\Phi'(y^*(\xi)) + \delta G's), \Phi'(y^*(\xi)) + \delta G's) \rangle =$$
$$= \delta^2 \langle H^P(d) G's, G's \rangle =$$
$$= \delta^2 [\langle H(d) G's, G's\rangle - \langle H(d) b, G's\rangle^2 / \langle H(d) b, b\rangle] \le$$
$$\le \delta^2 [(1-\tau)^{-2} \langle G's, s \rangle - \alpha^{-2} \langle G's, s \rangle^2] \le$$
$$\le 0.25\, \alpha^2 \delta^2 (1-\tau)^{-4}.$$

Объединяя с полученным ранее неравенством, имеем

$$\langle G(d)(y'(\delta) - y'_*(\delta)), y'(\delta) - y'_*(\delta) \rangle^{1/2} \le$$
$$\le 0.5\, \alpha\, \delta\, (1+\tau)^2 (1-\tau)^{-2} \qquad (5.4.1)$$

при всех $\delta \in [0, 1 - æ]$. Из неравенства (5.4.1), в частности, следует, что

$$\langle G(d)(y(\delta) - y'_*(\delta)), y(\delta) - y'_*(\delta) \rangle^{1/2} \le$$
$$\le \langle G(d)(y'(\delta) - y'_*(\delta)), y'(\delta) - y'_*(\delta) \rangle^{1/2} +$$
$$+ \langle G(d)(y'(\delta) - y(\delta)), y'(\delta) - y(\delta) \rangle^{1/2} \le$$
$$\le 0.5\, \alpha\, \delta\, (1+\tau)^2 (1-\tau)^{-2} + \alpha\, æ.$$

Таким образом, если выбрать δ из интервала

$[0, (1 - æ) / (1 + 0.5 (1+\tau)^2 (1-\tau)^{-2})]$,

то точки $y'_*(\delta)$ будут внутренними точками множества

$S(d, y_0, \alpha)$, и, следовательно, при таких δ справедливо тождество $y'_*(\delta) \equiv y^*(\xi(\delta))$. □

<u>Лемма 5.4.3.</u> Пусть при некотором $\xi_0 > 0$ точка $y^*(\xi_0)$ принадлежит множеству $S(d, y_0, \alpha)$ И пусть точка y' из $S(d, y_0, \alpha)$ такова, что $\langle y', b \rangle = \xi_0$ Тогда для точки

$$y_+ = y' - \gamma H^P(d) \Phi'(y'),$$

где $\gamma = (1 - \tau^2)^2 (1 + \tau^2)^{-1}$, справедлива оценка

$$\langle G(d)(y_+ - y^*(\xi_0)), y_+ - y^*(\xi_0) \rangle <$$
$$< 4\tau^2 (1 + \tau^2)^{-2} \langle G(d)(y' - y^*(\xi_0)), y' - y^*(\xi_0) \rangle$$

Доказательство. Действительно, в силу леммы 5.2.3 имеем

$$\langle G(d)(y_+ - y^*(\xi_0)), y_+ - y^*(\xi_0) \rangle =$$
$$= \langle G(d)(y' - y^*(\xi_0)), y' - y^*(\xi_0) \rangle -$$
$$- 2\gamma \langle G(d) H^P(d) \Phi'(y'), y' - y^*(\xi_0) \rangle +$$
$$+ \gamma^2 \langle G(d) H^P(d) \Phi'(y'), H^P(d) \Phi'(y') \rangle =$$
$$= \langle G(d)(y' - y^*(\xi_0)), y' - y^*(\xi_0) \rangle -$$
$$- 2\gamma \langle \Phi'(y'), y' - y^*(\xi_0) \rangle + \gamma^2 \langle H^P(d) \Phi'(y'),$$
$$\Phi'(y') \rangle = \langle G(d)(y' - y^*(\xi_0)), y' - y^*(\xi_0) \rangle -$$
$$- 2\gamma \langle \Phi'(y') - \Phi'(y^*(\xi_0)), y' - y^*(\xi_0) \rangle +$$
$$+ \gamma^2 \langle H^P(d)(\Phi'(y') - \Phi'(y^*(\xi_0))), \Phi'(y') -$$
$$- \Phi'(y^*(\xi_0)) \rangle < \langle G(d)(y' - y^*(\xi_0)), y' - y^*(\xi_0) \rangle -$$
$$- 2\gamma \langle \Phi'(y') - \Phi'(y^*(\xi_0)), y' - y^*(\xi_0) \rangle +$$
$$+ \gamma^2 \langle H(d)(\Phi'(y') - \Phi'(y^*(\xi_0))), \Phi'(y') - \Phi'(y^*(\xi_0)) \rangle <$$
$$< (1 - \gamma/(1 + \tau^2)) \langle G(d)(y' - y^*(\xi_0)), y' - y^*(\xi_0) \rangle +$$
$$+ \gamma [\gamma - (1 - \tau^2)^2 (1 + \tau^2)^{-1}] \langle H(d)(\Phi'(y') -$$
$$- \Phi'(y^*(\xi_0))), \Phi'(y') - \Phi'(y^*(\xi_0)) \rangle$$ □

Из приведенных результатов ясно, как построить метод нахождения решения задачи (5.2.1), основанный на отслеживании траектории $y^*(\xi)$, $\xi > 0$. Действительно, если бы нам удалось построить последовательность точек y_k таких, что:

а) для величин $\xi_k = \langle b, y_k \rangle$ выполняется неравенство $\xi_{k+1} \geq (1 + \mu n^{-1/2}) \xi_k$;

б) $y^*(\xi_k) \in S(d(y_k), y_k, \nu)$,

где $0 < \nu < (1 - \nu)^2$, то в силу неравенства (5.2.5) и леммы 5.2.5 не более чем за $\mu^{-1} n^{1/2} (\ln n + \ln 1/\theta)$ шагов нашелся бы номер i, для которого

$\langle H(d(y_i)) \Phi'(y_i), \Phi'(y_i) \rangle < 1$,

и, значит, в силу леммы 5.2.4 точка $x_*(d(y_i), y_i)$ была бы решением системы (5.2.1).

Леммы 5.4.1-5.4.3 показывают, как можно добиться выполнения условий а),б). Действительно, из леммы 5.4.2 следует, что при движении из y_k в $y_k(\delta)$ расстояние между точками $y_k(\delta)$ и $y^*(\xi(\delta))$ в метрике задаваемой матрицей $G(d(y_k))$, растет не слишком сильно. Поэтому в силу леммы 5.4.3 его можно вернуть на прежний уровень с помощью одного шага метода Ньютона в гиперпространстве, задаваемом уравнением $\langle b, y - y_k(\delta) \rangle = 0$. При этом значение линейной формы $\langle b, y \rangle$ увеличится в силу леммы 5.4.1 в $(1 + \alpha \delta n^{-1/2})$ раз. Для обоснования этого процесса нужна лишь согласованность в выборе параметров α и δ.

Итак, выпишем точную схему такого метода

Метод $D_0(\alpha, \delta)$

0. Полагаем $y_0 = 0 \in R^m$

1. k-я итерация ($k \geq 0$).

а) Если $\langle H(d(y_k)) \Phi'(y_k), \Phi'(y_k) \rangle < 1$, то полагаем $x_* = x_*(d(y_k), y_k)$ и останавливаемся;

8 Зак. 2406

б) в противном случае полагаем

$$y'_k = y_k + \delta\ \alpha\ \frac{H(\ d(\ y_k\ )\ )b}{\langle\ H(\ d(\ y_k\ )\ b,\ b\ \rangle^{1/2}}\ ,$$

$$y_{k+1} = y'_k - \gamma\ H^P\ (\ d(\ y_k\ )\ )\ \Phi'(\ y'_k\ )\ ,$$

где $\gamma = (1 - \alpha^2)^2\ (1 + \alpha^2)^{-1}$.

Итерация закончена.

Теорема 5.4.1. Если для решения θ - устойчивой задачи (5.2.1) применяется метод $\mathcal{D}'_0(\ \alpha\ ,\ \delta\ )$ с параметрами α, δ такими, что

$$\alpha \in (0,\ 1/6)\ ,\quad \delta = 2\ \alpha\ ,$$

то не более чем за

$$N = O(\ n^{1/2}\ (\ \ln n + \ln 1/\theta\ )\ )$$

итераций в качестве ответа x_* будет выдано допустимое решение системы (5.2.1). При этом, если вычисление матрицы $H(\ d(\ y_k\ )\ )$ производить с помощью стандартных методов линейной алгебры, то общее число арифметических операций метода не превзойдет величины порядка

$$O(\ n^{1.5}\ m^2\ [\ \ln n + \ln 1/\theta\ ]\ ).$$

Д о к а з а т е л ь с т в о. Покажем, что на всех итерациях метода выполняется соотношение

$$\langle\ G_k(\ y^*(\ \xi_k\ ) - y_k\ ),\ y^*(\ \xi_k\ ) - y_k\ \rangle < \ \varkappa^2\ \alpha^2, \qquad (5.4.2)$$

где

$$G_k = G(\ d(\ y_k\ )\ ),\ \ \xi_k = \langle\ b,\ y_k\ \rangle,\ \ \varkappa = 2\ \alpha\ (1 + \alpha^2)^{-1}$$

Действительно, в силу выбора точки y_0 при $k = 0$ (5.4.2) выполнено. Пусть теперь оно выполнено при некотором $k > 0$. Заметим, что в силу выбора параметров α, δ условия леммы 5.4.2 выполнены. Следовательно,

$$\langle\ G_k(y^*(\ \xi_{k+1}\ ) - y'_k),\ y^*(\ \xi'_k\ ) - y'_k\ \rangle < \ \alpha^2\ (1 - \delta\ )^2$$

(мы учли, что $\langle\ b,\ y'_k\ \rangle = \langle\ b,\ y_{k+1}\ \rangle = \xi_{k+1}$). Кроме того, из леммы 5.4.3

$\langle G_{k+1}(y^*(\xi_{k+1}) - y_{k+1}), y^*(\xi'_{k+1}) - y'_{k+1} \rangle^{1/2} \le$
$\le æ^2 \langle G_k(y'_k - y^*(\xi_{k+1})), y'_k - y^*(\xi_{k+1}) \rangle^{1/2}$.
Поэтому, принимая во внимание, что $\delta = 2\alpha$ и $æ < 1$, получаем
$\langle G_k (y_{k+1} - y_k), y_{k+1} - y_k \rangle^{1/2} \le$
$\le \langle G_k (y_{k+1} - y^*(\xi_{k+1})), y_{k+1} - y^*(\xi_{k+1}) \rangle^{1/2} +$
$+ \langle G_k (y'_k - y^*(\xi_{k+1})), y'_k - y^*(\xi_{k+1}) \rangle^{1/2} +$
$+ \langle G_k (y'_k - y_k), y'_k - y_k \rangle^{1/2} \le$
$\le \alpha(1 - \delta)(1 + æ) + \alpha\delta = \alpha(1 + (1 - \delta)æ) \le \delta$.
Таким образом, учитывая $(d(y_k), \delta, \delta)$-согласованность функции $\Phi(y)$ в точке y_k и предыдущие неравенства, имеем
$\langle G_{k+1}(y^*(\xi_{k+1}) - y_{k+1}), y^*(\xi'_{k+1}) - y'_{k+1} \rangle^{1/2} \le$
$\le (1 - \delta)^{-1} \langle G_k(y^*(\xi_{k+1}) - y_{k+1}), y^*(\xi'_{k+1}) -$
$- y'_{k+1} \rangle^{1/2} \le (1 - \delta)^{-1} \alpha (1 - \delta) æ = æ\alpha$.

Итак, мы показали, что соотношение (5.4.2) выполнено при всех $k \ge 0$ Заметим, что в силу леммы 5.2.1 для величин ξ_k, $k \ge 0$, справедлива оценка
$\xi_{k+1} \ge (1 + \delta\alpha n^{-1/2})^k \xi_1 =$
$= (1 + \delta\alpha n^{-1/2})^k \delta\alpha \langle H_0 b, b \rangle^{1/2} \ge$
$\ge (1 + \delta\alpha n^{-1/2})^k \delta\alpha$.
Воспользовавшись теперь неравенством (5.2.5), получаем, что среди первых $N = O(n^{1/2} [\ln n + \ln(1/\theta)])$ итераций найдется итерация с номером k такая, что $\xi_k \le \langle b, y^* \rangle \le \le \xi_{k+1}$. Но тогда в силу леммы 5.4.2 будет выполнено включение $y^* \in S(d(y_k), y_k, æ\alpha)$.

Заметим, что $æ\alpha < (1 - æ\alpha)^2$. Следовательно, в силу леммы 5.2.5 в этот момент сработает критерий прерывания метода, причем в силу леммы 5.2.4 точка x_* окажется решением задачи (5.2.1). □

Трудоемкость метода $\mathcal{D}_0(\alpha, \delta)$ можно снизить с помощью приема, описанного в §5.3. В результате можно получить метод с оценкой трудоемкости $O(n m^2 [\ln n + \ln(1/\theta)])$. Приведем общую схему такого метода.

Метод $\mathcal{D}_1(\alpha, \delta, \tau)$

0. Полагаем

$\varkappa = 2(\alpha + \tau)[1 + (\alpha + \tau)^2]^{-1}$,

$\gamma = [1 - (\tau + \alpha)^2]^2 [1 + (\tau + \alpha)^2]^{-1}$,

$y_0 = 0 \in R^m$, $H_0 = [A A^T]^{-1}$,

$d_0 = (1, \dots, 1)^T \in R^n$.

1. k-я итерация ($k \geq 0$):

а) проверяем выполнение неравенства

$\langle H_k \Phi'(y_k), \Phi'(y_k) \rangle \leq (\varkappa\alpha)^2 (1 - \tau - \varkappa\alpha)^{-2}$;

если оно выполнено, то полагаем $x_* = x_*(d_k, y_k)$ и останавливаемся;

б) в противном случае полагаем

$$y'_k = y_k + \delta \alpha \frac{H_k b}{\langle H_k b, b \rangle^{1/2}}$$

$y_{k+1} = y'_k - \gamma H_k^p \Phi'(y'_k)$;

в) формируем множество

$$\mathcal{J}(k) = \{ j \mid d_k^{(j)} \geq (1 + \tau) d(y_{k+1})^{(j)} \} \cup$$
$$\cup \{ j \mid d_k^{(j)} \leq (1 - \tau) d(y_{k+1})^{(j)} \}$$

г) производим пересчет вектора d_k :

$$d_{k+1}^{(j)} = \begin{cases} d(y_{k+1})^{(j)}, & \text{если } j \in \mathcal{J}(k), \\ d_k^{(j)} & \text{в противном случае;} \end{cases}$$

д) с помощью формул одноранговой коррекции производим пересчет матрицы H_k :

$H_{k+1} = [H_k^{-1} + \sum_{j \in \mathcal{J}(k)} [(d_{k+1}^{(j)})^2 - (d_k^{(j)})^2] a_j a_j^T]^{-1}$.

Итерация закончена.

Теорема 5.4.2. Если для решения θ - устойчивой задачи (5.2.1) применяется метод $\mathcal{D}_1(\alpha, \delta, \tau)$ с параметрами $\alpha \in (0, 1/12)$, $\delta = 4\alpha$, $\tau = \alpha$,
то не более, чем за
$N = O(n^{1/2} [\ln n + \ln 1/\theta])$
итераций в качестве ответа x_* будет выдано допустимое решение системы (5.2.1). При этом за любые k итераций, $0 \leq k \leq N$, суммарное число одноранговых коррекций матрицы H_k не будет больше чем $2\alpha\tau^{-1} k n^{1/2}$. Таким образом, общая трудоемкость решения задачи (5.2.1) методом $\mathcal{D}_1(\alpha, \delta, \tau)$ оценивается величиной порядка
$O(n m^2 [\ln n + \ln(1/\theta)])$.

Доказательство. Как и в методе $\mathcal{N}_1(\alpha, \tau)$, в методе $\mathcal{D}_1(\alpha, \delta, \tau)$ в силу способа пересчета матрицы H_k и вектора d_k справедливы соотношения:
$H_k = H(d_k)$,
$(1 + \tau)^{-1} d_k \leq d(y_k) \leq (1 - \tau)^{-1} d_k$
Отсюда получаем, что при любом β, $0 \leq \beta < 1 - \tau$, функция $\Phi(y)$ будет $(d_k, \beta, \beta + \tau)$-согласованной во всех точках y_k. Покажем, что на всех итерациях метода выполняется соотношение
$\langle G_k(y^*(\xi_k) - y_k), y^*(\xi_k) - y_k \rangle \leq æ^2 \alpha^2$, (5.4.3)
где $G_k = [H_k]^{-1}$; $\xi_k = \langle b, y_k \rangle$; $æ = 2(\alpha + \tau) \times [1 + (\alpha + \tau)^2]^{-1}$. Действительно, в силу выбора точки y_0 при $k = 0$ (5.4.3) выполнено. Пусть теперь оно выполнено и при некотором $k \geq 0$. Заметим, что в силу выбора параметров α, δ, τ и определения $æ$ условия леммы 5.4.2 на длину шага δ выполнены. Следовательно,
$\langle G_k(y^*(\xi_{k+1}) - y'_k), y^*(\xi'_k) - y'_k \rangle \leq \alpha^2(1 - \delta)^2$
Кроме того, из леммы 5.4.3

$$\langle G_{k+1} (y^*(\xi_{k+1}) - y_{k+1}), y^*(\xi'_{k+1}) - y'_{k+1} \rangle^{1/2} \le$$
$$\le æ^2 \langle G_k (y'_k - y^*(\xi_{k+1})), y'_k - y^*(\xi_{k+1}) \rangle^{1/2}$$

Поэтому, принимая во внимание, что $æ < 1$, получаем

$$\langle G_k (y_{k+1} - y_k), y_{k+1} - y_k \rangle^{1/2} \le$$
$$\le \langle G_k (y_{k+1} - y^*(\xi_{k+1})), y_{k+1} - y^*(\xi_{k+1}) \rangle^{1/2} +$$
$$+ \langle G_k (y'_k - y^*(\xi_{k+1})), y'_k - y^*(\xi_{k+1}) \rangle^{1/2} +$$
$$+ \langle G_k (y'_k - y_k), y'_k - y_k \rangle^{1/2} \le$$
$$\le \alpha (1 - \delta)(1 + æ) + \alpha \delta = \alpha(1 + (1 - \delta) æ) < 2 \alpha.$$

Таким образом, учитывая $(d_{k+1}, 0, \tau)$ - согласованность функции $\Phi(y)$ в точке y_{k+1} и предыдущие неравенства, имеем

$$(1 - \tau)\langle G_{k+1}(y^*(\xi_{k+1}) - y_{k+1}), y^*(\xi'_{k+1}) - y'_{k+1} \rangle^{1/2} \le$$
$$\le \langle G(d(y_{k+1}))(y^*(\xi_{k+1}) - y_{k+1}), y^*(\xi'_{k+1}) - y'_{k+1} \rangle^{1/2} \le$$
$$\le (1 - 2 \alpha - \tau)^{-1} \langle G_k (y^*(\xi_{k+1}) - y_{k+1}),$$
$$y^*(\xi'_{k+1}) - y'_{k+1} \rangle^{1/2} \le (1 - 2 \alpha - \tau)^{-1} \alpha (1 - \delta) æ.$$

Осталось заметить, что $\tau = \alpha$, $\delta = 4 \alpha$ и, значит,

$$(1 - 2 \alpha - \tau) (1 - \tau) > 1 - \delta.$$

Итак, мы показали, что соотношение (5.4.3) выполнено при всех $k \ge 0$ Как и в теореме 5.4.1, для величин ξ_k, $k \ge 0$, будет справедлива оценка

$$\xi_{k+1} \ge (1 + \delta \alpha n^{-1/2})^k \xi_1 \ge (1 + \delta \alpha n^{-1/2})^k \delta \alpha.$$

Воспользовавшись теперь неравенством (5.2.5), получаем, что среди первых $N = O(n^{1/2} [\ln n + \ln(1/\theta)])$ итераций найдется итерация с номером k такая, что

$$\xi_k \le \langle b, y^* \rangle \le \xi_{k+1}$$

Но тогда в силу леммы 5.4.2 будет выполнено включение

$$y^* \in S(d_k , y_k , æ \alpha)$$

и в силу леммы 5.2.5 в этот момент сработает критерий прерывания метода.

Заметим, что

$\varkappa\ \alpha\ (1 - \varkappa\ \alpha)^{-2} < (1 + \tau)^{-1}$.

Поэтому вследствие леммы 5.2.4 точка x_* будет решением задачи (5.2.1). Осталось заметить, что способ пересчета векторов y_k и d_k удовлетворяет предположениям леммы 5.3.1. Поэтому за любые k итераций, $0 < k < N - 1$, число одноранговых коррекций матрицы H_k не превысит $2\,(k + 1) \times \times\, n^{1/2}\ \alpha\ \tau^{-1}$ □

Отметим, что в сформулированной теореме мы не ставили целью обосновать оптимальные значения для параметров α, δ и τ. Приведенная стратегия их определения была выбрана лишь из соображений упрощения доказательства.

5.5. Прямой метод параллельных траекторий

Рассмотрим задачу линейного программирования в следующей постановке:

$$\max \{\ t \mid A x = t\, b\ ,\ x \in K\ \} \equiv t^* , \tag{5.5.1}$$

где A - невырожденная $m{\times}n$ -матрица, $t \in R$, $K = \{\ x \in R^n \mid \mid x^{(i)} \mid < 1\ ,\ i = 1,\ \ldots,\ n\ \}$. Как и в §5.2, введем барьерную функцию для множества K

$$F(\,x\,) = \sum_{i=1}^{n} [\ - \mid x^{(i)} \mid - \ln\,(1 - \mid x^{(i)} \mid\,)\ ] .$$

Положим

$$E(d,\ x_0,\ r) = \{\ x \in R^n \mid \|\ D(\ d\ )^{-1}\,(\ x - x_0\ )\ \| < r\ \} .$$

Для x из R^n введем функцию $s(\ x\ ) :\ R^n \to R^n$, такую, что $s(\ x\ )^{(i)} = 1 - \mid x^{(i)} \mid$, $i = 1,\ \ldots,\ n$. Ясно, что в таких обозначениях $D(\ s(\ x\ )\ )^{-2} \equiv F''(\ x\ )$.

Будем говорить, что векторная функция $s(\ x\ )$ является $(\ d,\ r,\ \tau\ )$ -согласованной в точке x_0, если для любого

$x \in E(d, x_0, r)$ справедливо неравенство

$(1 - \tau) d \le s(x) \le (1 + \tau) d$.

Заметим, что из (d, r, τ) - согласованности функции $s(x)$ в точке x_0 очевидным образом выводится неравенство

$(1 + \tau)^{-2} D(d)^{-2} \le F''(x) \le (1 - \tau)^{-2} D(d)^{-2}$,

выполняющееся при всех x из $E(d, x_0, r)$.

Нетрудно видеть, что функция $s(x)$ является $(s(x), 0, 0)$ -согласованной в любой точке x из R^n. Справедлив также следующий результат.

Лемма 5.5.1. Если функция $s(x)$ является $(d, 0, \tau)$ -согласованной в некоторой точке x_0, $0 \le \tau \le 1$, то она также будет и $(d, r, r + \tau)$ -согласованной в той же точке при любом r из множества $[0, 1 - \tau)$.

Доказательство. Действительно, из включения $x \in E(d, x_0, r)$ следует, что

$| x^{(i)} - x_0^{(i)} | \le r d^{(i)}$, $i = 1, \dots, n$.

Кроме того, из $(d, 0, \tau)$ -согласованности функции $s(x)$ в точке x_0 имеем

$(1 - \tau) d \le s(x_0) \le (1 + \tau) d$.

Отсюда получаем

$(1 - r - \tau) d \le s(x) \le (1 + r + \tau) d$. □

Следствие 5.5.1. Функция $s(x)$ будет $(s(x_0), r, r)$ -согласованной в любой точке $x_0 \in R^n$ при всяком r из $[0, 1)$

Лемма 5.5.2. Если функция $s(x)$ является (d, r, τ) -согласованной в точке x_0 и $r \in [0, 1 - \tau)$, то справедливо включение $E(d, x_0, r) \subset K$

Доказательство. В силу условия леммы для любого x из множества $E(d, x_0, r)$ имеем

$| x^{(i)} - x_0^{(i)} | \le r d^{(i)} \le$

$< r (1 - \tau)^{-1} s(x_0) < 1 - | x_0^{(i)} |$ □

Лемма 5.5.3. Если функция $s(x)$ является (d, r, τ)-согласованной в точке x_0, то при любых x_1, x_2 из $E(d, x_0, r)$ справедливо неравенство

$$\langle F'(x_1) - F'(x_2), x_1 - x_2 \rangle \geq$$
$$\geq 0.5 (1 + \tau^2)^{-1} [\| D(d)^{-1} (x_1 - x_2) \|^2 +$$
$$+ (1 - \tau^2)^2 \| D(d) (F'(x_1) - F'(x_2)) \|^2] \geq$$
$$\geq (1 + \tau)^{-2} \| D(d)^{-1} (x_1 - x_2) \|^2$$

Это утверждение следует из сильной выпуклости функции $F(x)$ на множестве $E(d, x_0, r)$ с константой $(1 + \tau)^{-2}$ в метрике, задаваемой диагональной матрицей $D(d)$, и ограниченности сверху спектра матриц вторых производных на том же множестве и в той же метрике константой $(1 - \tau)^{-2}$ (см. теорему 1.4.5)

Определим траекторию

$$x^*(t) = \operatorname{argmin} \{ F(x) \mid A x = t b \},$$

где $t \in [0, t^*]$ Ясно, что $x(0) = 0$ В этом и следующем параграфах будут рассмотрены методы, основанные на "отслеживании" точек траектории $x^*(t)$.

Введем в пространстве R^m траекторию $y_*(t)$

$$A^T y_*(t) = F'(x^*(t))$$

Между точками траекторий $x^*(t)$ и $y_*(t)$ существует простая зависимость

$$\langle a_i , y_*(t) \rangle = (x^*)^{(i)}(t) / [1 - | (x^*)^{(i)}(t) |],$$
$$(x^*)^{(i)}(t) = \langle a_i , y_*(t) \rangle / [1 + | \langle a_i , y_*(t) \rangle |],$$
$$i = 1, \dots, n,$$

где a_i - столбцы матрицы A

Нетрудно видеть, что

$$F''(x^*(t)) (x^*(t))'_t \equiv A^T (y_*(t))'_t$$

т.е.

$$(x^*(t))'_t \equiv D(s(x^*(t)))^2 A^T (y_*(t))'_t$$

Поэтому

$$G(s(x^*(t))) (y_*(t))'_t \equiv A (x^*(t))'_t = b ,$$

т.е. $(y_*(t))'_t = H(s(x^*(t)))b$.

Лемма 5.5.4. При любых $t \in (0, t^*)$ справедливо неравенство

$$t^* - t \leq n < y_*(t) , b >^{-1} .$$

Доказательство. Положим $z = t x^*(t^*) / t^*$. Тогда $A z = t$ и поэтому

$$< F'(x^*(t)) , z - x^*(t) > \geq 0 ;$$

т.е.

$$< F'(x^*(t)), x^*(t) > \leq t / t^* < F'(x^*(t)), x^*(t^*) > \leq$$
$$\leq t (t^*)^{-1} \sum_{i=1}^n | (x^*)^{(i)} (t) | / [1 - | (x^*)^{(i)} (t) |] .$$

Следовательно,

$$t^{-1} (t^* - t) \leq$$
$$\leq \{ \sum_{i=1}^n | (x^*)^{(i)} (t) | / [1 - | (x^*)^{(i)} (t) |] \} / (\sum_{i=1}^n ((x^*)^{(i)} (t))^2 / [1 - | (x^*)^{(i)} (t) |]) -$$
$$- 1 = \sum_{i=1}^n | (x^*)^{(i)} (t) | / < F'(x^*(t)) , x^*(t) >.$$

Осталось заметить, что

$$< F'(x^*(t)) , x^*(t) > = t < y_*(t) , b > \qquad \square$$

Зафиксируем произвольную точку $x_0 \in R^n$ Предположим, что функция $s(x)$ является (d, r, τ) -согласованной в этой точке.

Лемма 5.5.5. Пусть $A x_0 = t_0 b$, $t_0 > 0$ и точка $x^*(t_0)$ принадлежит множеству $E(d\ x_0, \varkappa r)$ при некотором $\varkappa$ из $(0 , 1)$. Обозначим

$$x(\delta) = x_0 + \delta r \frac{D(d)^2 A^T H(d) b}{< H(d) b, b >^{1/2}} ,$$

$$t(\delta) = t_0 + \delta r < H(d) b, b >^{-1/2} ,$$

где параметр δ удовлетворяет соотношениям

$$0 < \delta < (1 - \varkappa)\,[1 + (1 + \tau)^2 (1 - \tau)^{-2}]^{-1}.$$

Тогда точка $x^*(t(\delta)) \in E(d, x_0, r)$, причем справедливы неравенства

$$\| D(d)^{-1} (x^*(t(\delta)) - x(\delta)) \| <$$
$$< r\,[\varkappa + \delta (1 - \tau)^{-4}] < r\,(1 - \delta);$$
$$\langle y_*(t(\delta)), b \rangle > \langle y_*(t_0), b \rangle +$$
$$+ \delta (1 - \tau)^2 r \langle H(d) b, b \rangle^{1/2}.$$

Д о к а з а т е л ь с т в о Положим

$$x'(\delta) = x^*(t_0) + \delta r \frac{D(d)^2 A^T H(d) b}{\langle H(d) b, b \rangle^{1/2}},$$

$$x'_*(\delta) = \operatorname{argmin} \{ F(x) \mid A x = t(\delta),\ x \in E(d, x_0, r) \}.$$

Заметим, что при всех $\delta \in [0, 1 - \varkappa]$ точка $x'(\delta)$ заведомо принадлежит множеству $E(d, x_0, r)$. Поэтому в силу леммы 5.5.3

$$(1 + \tau)^{-2} \| D(d)^{-1} (x'(\delta) - x'_*(\delta)) \|^2 <$$
$$< \langle F'(x'(\delta)) - F'(x'_*(\delta)),\ x'(\delta) - x'_*(\delta) \rangle <$$
$$< \langle F'(x'(\delta)),\ x'(\delta) - x'_*(\delta) \rangle.$$

Введем матрицу

$$P = I - A^T H(d) A D(d)^2.$$

Нетрудно видеть, что

$$P^T (x'(\delta) - x'_*(\delta)) = x'(\delta) - x'_*(\delta).$$

Следовательно,

$$(1 + \tau)^{-2} \| D(d)^{-1} (x'(\delta) - x'_*(\delta)) \|^2 <$$
$$< \langle P F'(x'(\delta)),\ x'(\delta) - x'_*(\delta) \rangle <$$
$$< \| D(d)^{-1} (x'(\delta) - x'_*(\delta)) \| \times \| D(d) P F'(x'(\delta)) \|$$

Таким образом,

$$(1 + \tau)^{-4} \| D(d)^{-1} (x'(\delta) - x'_*(\delta)) \|^2 <$$
$$< \langle H^P(d) F'(x'(\delta)),\ F'(x'(\delta)) \rangle,$$

где $H^P(d) = P^T D(d)^2 P$. Заметим, что

$F'(\, x'(\delta)\,) =$

$$= F'(\, x^*(t_0)\,) + \delta \left[\, \int_0^1 D(\, s(x^*(t_0) + \theta\, \delta s)\,)^{-2}\, d\theta \,\right] s\,,$$

где $s = r\, D(d)^2 A^T H(d)\, b \,/\, \langle H(d)\, b,\, b\rangle^{1/2}$ Обозначим интегральную матрицу (заключенную в квадратные скобки) через G' Нетрудно видеть, что $P\, F'(\, x^*(\, t_0\,)\,) = 0$ G' - диагональная матрица, $G' \le (1 - \tau)^{-2} D(\, d\,)^{-2}$ Поэтому

$\langle H^P(\, d\,)\, F'(\, x'(\delta)\,)\,,\; F'(\, x'(\delta)\,)\,\rangle =$

$= \langle H^P(d)(\, F'(x^*(t_0)) + \delta\, G's\,),\, F'(x^*(t_0)) + \delta\, G's) \rangle =$

$= \delta^2 \langle H^P(d)\, G' s,\, G' s \rangle \le \delta^2 \langle D(d)^2\, G' s,\, G' s \rangle \le$

$\le \delta^2 (1 - \tau)^{-4} \langle D(d)^{-2} s,\, s \rangle = r^2\, \delta^2 (1 - \tau)^{-4}$

Объединяя последнее соотношение с полученным ранее неравенством, имеем

$\| D(d)^{-1} (x'(\delta) - x'_*(\delta)) \| \le \delta\, r\, (1 + \tau)^2 (1 - \tau)^{-2}$

при всех $\delta \in [\, 0,\, 1 - æ\,]$. Из последнего неравенства, в частности, следует, что

$\| D(\, d\,)^{-1} (\, x(\delta) - x_*(\delta)\,) \| \le$

$\le \| D(\, d\,)^{-1} (\, x'(\delta) - x'_*(\delta)\,) \| +$

$+ \| D(\, d\,)^{-1} (\, x'(\delta) - x(\delta)\,) \| \le$

$\le \delta\, r\, (1 + \tau)^2 (1 - \tau)^{-2} + r\, æ$

Таким образом, если выбрать δ из интервала

$[\, 0,\, (1 - æ) \,/\, (1 + (1 + \tau)^2 (1 - \tau)^{-2})\,]$,

то точки $x'_*(\,\delta\,)$ будут внутренними точками множества $E(\, d,\, x_0\,,\, r\,)$ и, следовательно, при таких δ справедливо $x'_*(\,\delta\,) \equiv x^*(\, t(\delta)\,)$.

Нам осталось обосновать оценку скорости роста величины $\xi(\, t\,) \equiv \langle y_*(\, t\,)\,,\, b\,\rangle$ Заметим, что

$$\xi(\, t(\delta)\,) = \xi(\, t_0\,) + \int^{\delta} \xi'(\, t(\,\theta\,)\,)\, t'(\,\theta\,)\, d\theta$$

Но по доказанному выше при всех $\theta < \delta$ точка $x^*(t(\theta))$ принадлежит множеству $E(d , x_0, r)$. Поэтому

$$\xi (t(\theta)) = \langle H(s(x^*(t(\theta))) b , b \rangle \geq$$
$$\geq (1 - \tau)^2 \langle H(d) b , b \rangle .$$

Таким образом,

$$\xi (t(\delta)) \geq \xi (t_0) + \delta (1 - \tau)^2 \langle H(d) b, b\rangle t'(0) =$$
$$= \xi (t_0) + \delta (1 - \tau)^2 r \langle H(d) b, b \rangle^{1/2} . \quad \square$$

<u>Лемма 5.5.6.</u> Пусть при некотором $t > 0$ точка $x^*(t)$ принадлежит множеству $E(d, x_0 , r)$. И пусть точка x' из $E(d, x_0 , r)$ такова, что $A x' = t b$. Тогда для точки

$$x_+ = x' - h H^P(d) F'(x'),$$

где

$$h = (1 - \tau^2)^2 (1 + \tau^2)^{-1};$$
$$H^P(d) = D(d)^2 - D(d)^2 A^T H(d) A D(d)^2 ,$$

справедливо равенство $A x_+ = t b$, причем

$$\| D(d)^{-1} (x_+ - x^*(t)) \| \leq$$
$$\leq 2 \tau (1 + \tau^2)^{-1} \| D(d)^{-1} (x' - x^*(t)) \| .$$

Д о к а з а т е л ь с т в о. Действительно в силу леммы 5.5.3 имеем

$$\| D(d)^{-1} (x_+ - x^*(t)) \|^2 =$$
$$= \| D(d)^{-1} (x' - x^*(t)) \|^2 -$$
$$- 2 h \langle D(d)^{-2} H^P(d) F'(x'), x' - x^*(t) \rangle +$$
$$+ h^2 \langle D(d)^{-2} H^P(d) F'(x') , H^P(d) F'(x') \rangle =$$
$$= \| D(d)^{-1} (x' - x^*(t)) \|^2 -$$
$$- 2h \langle F'(x'), x' - x^*(t) \rangle + h^2\langle H^P(d) F'(x'), F'(x') \rangle =$$
$$= \| D(d)^{-1} (x' - x^*(t)) \|^2 -$$
$$- 2 h \langle F'(x') - F'(x^*(t)) , x' - x^*(t) \rangle +$$
$$+ h^2 \langle H^P(d)(F'(x') - F'(x^*(t))), F'(x') - F'(x^*(t)) \rangle \leq$$
$$\leq \| D(d)^{-1} (x' - x^*(t)) \|^2 -$$
$$- 2 h \langle F'(x') - F'(x^*(t)) , x' - x^*(t) \rangle +$$

$$+ h^2 \langle D(d)^2(F'(x') - F'(x^*(t))), F'(x') - F'(x^*(t)) \rangle \le$$

$$\le (1 - h/(1 + \tau^2)) \| D(d)^{-1} (x' - x^*(t)) \|^2 +$$

$$+ h [h - (1 - \tau^2)^2 (1 + \tau^2)^{-1}] \times$$

$$\times \langle D(d)^2 (F'(x') - F'(x^*(t))), F'(x') - F'(x^*(t)) \rangle \qquad \square$$

Из приведенных результатов ясно, как построить итеративный метод решения задачи (5.5.1), основанный на отслеживании траектории $x^*(t)$, $t > 0$ (прямой метод параллельных траекторий). Действительно, если есть точка x_0 такая, что

$A x_0 = t_0 b$, $x^*(t_0) \in E(s(x_0) , x_0 , æ r)$,

а параметры æ и r достаточно малы то в силу лемм 5.5.5 5.5.6 можно увеличить значение t до величины $t' > t$ и построить точку x' такую, что

$A x' = t' b$, $x^*(t') \in E(s(x') , x' , æ r)$.

Для обоснования этого процесса нужна лишь согласованность в выборе соответствующих параметров.

Итак, выпишем точную схему этого метода

Метод $\mathcal{P}_0(r , \delta)$

0. Полагаем $x_0 = 0 \in R^n$, $h = (1 - r^2)^2 (1 + r^2)^{-1}$

1. k -я итерация ($k \ge 0$).

Полагаем

$$x'_k = x_k + \delta r \frac{D(s(x_k))^2 A^T H(s(x_k)) b}{\langle H(s(x_k)) b, b \rangle^{1/2}} ,$$

$$x_{k+1} = x'_k - h H^P (s(x_k)) F'(x'_k) ,$$

где $H^P(d) = D(d)^2 - D(d)^2 A^T H(d) A D(d)^2$

Итерация закончена

Теорема 5.5.1. Пусть для решения задачи (5.5.1) применяется метод $\mathcal{P}_0(r, \delta)$ с параметрами r, δ такими, что

$r \in (0, 1/6)$, $\delta = 2 r$.

Тогда все точки последовательности $\{ x_k \}_{k=0}^{\infty}$, построенной методом $\mathcal{P}_0(r, \delta)$, обладают следующими свойствами:

1) $x_{k+1} \in E(s(x_k) , x_k , \delta) \subset K$;

2) для последовательности величин $\{ t_k \}_{k=0}^{\infty}$ такой, что

$t_0 = 0$; $t_{k+1} = t_k + \delta \, r \langle H(s(x_k)) b, b \rangle^{-1/2}$,

справедливо соотношение $A x_k = t_k b$;

3) при любом $k \geq 0$ выполнено включение

$x^*(t_k) \in E(s(x_k) , x_k , \text{æ} \, r)$,

где $\text{æ} = 2 r (1 + r^2)^{-1}$

При этом последовательность $\{ t_k \}_{k=0}^{\infty}$ сходится к t^* и справедлива следующая оценка скорости сходимости:

$(t^* - t_k) \langle [A A^T]^{-1} b, b \rangle^{1/2} \leq$

$\leq n \, C(r, \delta) (1 + q n^{-1/2})^{-k+1}$,

где $C(r , \delta) = \delta^{-1} r^{-1} (1 - r)^{-2}$; $q = \delta \, r (1 - r)^2 \times$

$\times (1 + \text{æ} \, r)^{-1}$.

Д о к а з а т е л ь с т в о. Для упрощения выкладок положим

$x_k^* = x^*(t_k)$; $d_k = s(x_k)$;

$D_k = D(s(x_k))$; $H_k^p = H^p (s(x_k))$.

Заметим, что в силу следствия 5.5.1 функция $s(x)$ будет (d_k , r , r) -согласованной в любой точке x_k .

Докажем утверждение 1) теоремы. При $k = 0$ оно выполнено. Далее, пусть оно выполнено и при некотором $k \geq 0$ Тогда из схемы метода, лемм 5.5.5, 5.5.6 и способа выбора параметров следует

$\| D_k^{-1} (x_{k+1} - x_k) \| \leq \| D_k^{-1} (x_{k+1} - x_{k+1}^*) \| +$

$+ \| D_k^{-1} (x_{k+1}^* - x'_k) \| + \| D_k^{-1} (x'_k - x_k) \| \leq$

$\leq (1 + \text{æ})(1 - \delta) r + \delta \, r < 2 r = \delta$.

Таким образом, $x_{k+1} \in E(d_k , x_k , \delta) \subset K$ при всех k

Утверждение 2) теоремы 5.5.1 непосредственно следует из схемы метода и того факта, что $A x'_k = A x_{k+1}$

Перейдем к доказательству утверждения 3). В силу выбора точки x_0 при $k = 0$ это включение выполнено. Пусть теперь оно выполнено при некотором $k \geq 0$. Заметим, что параметры r и δ удовлетворяют условиям леммы 5.5.5. Поэтому

$$\| D_k^{-1} (x_{k+1}^* - x'_k) \| \leq r (1 - \delta)$$

Кроме того, из леммы 5.5.6

$$\| D_k^{-1} (x_{k+1}^* - x_{k+1}) \| \leq \text{æ} \| D_k^{-1} (x_{k+1}^* - x'_k) \|$$

Таким образом, учитывая (d_k , δ , δ) - согласованность функции $s(x)$ в точке x_k и предыдущие неравенства, имеем

$$\| D_{k+1}^{-1} (x_{k+1}^* - x_{k+1}) \| \leq$$

$$\leq (1 - \delta)^{-1} \| D_k^{-1} (x_{k+1}^* - x_{k+1}) \| \leq$$

$$\leq (1 - \delta)^{-1} r (1 - \delta) \text{æ} = \text{æ} r .$$

Нам осталось получить оценку скорости сходимости величин t_k. Для этого заметим, что

$$\langle y_*(t_k) , b \rangle \leq$$

$$\leq \langle G(d_k) y_*(t_k), y_*(t_k) \rangle^{1/2} \cdot \langle H(d_k) b, b \rangle^{1/2} \leq$$

$$\leq (1 + \text{æ} r) \langle G(s(x^*(t_k))) y_*(t_k), y_*(t_k) \rangle^{1/2} \times$$

$$\times \langle H(d_k) b, b \rangle^{1/2} \leq (1 + \text{æ} r) n^{1/2} \langle H(d_k) b, b \rangle^{1/2} .$$

Таким образом, в силу леммы 5.5.5

$$\langle y_*(t_{k+1}), b \rangle \geq$$

$$\geq (1 + \delta \, r(1 - r)^2 (1 + \text{æ} r)^{-1} n^{-1/2}) \langle y_*(t_k), b \rangle$$

В свою очередь, в силу той же леммы

$$\langle y_*(t_1), b\rangle \geq \delta \, r (1 - r)^2 \langle H(0) b, b \rangle^{1/2}$$

Осталось воспользоваться леммой 5.5.4. $\square$

При выполнении итерации метода $\mathcal{P}_0(r, \delta)$ наиболее трудоемкой операцией является вычисление произведений матриц

$$H(s(x_k)) = [A D(s(x_k))^2 A^T]^{-1} ;$$

$$H^P(s(x_k)) = D(s(x_k))^2 - D(s(x_k))^2 \times$$

$$\times A^T H(s(x_k)) A D(s(x_k))^2$$

на соответствующий вектор Если для этого пользоваться стандартными методами линейной алгебры, то трудоемкость такой итерации будет иметь порядок $O(n m^2)$ (здесь и далее мы считаем, что величина порядка $O(m)$ больше, чем $O(n^{1/2})$).

В оставшейся части этого параграфа мы подвергнем метод $\mathcal{P}_0(r, \delta)$ некоторой модификации (сравните с §5.3) с тем, чтобы избавиться от необходимости заново вычислять на каждой итерации матрицу $H(s(x_k))$. В результате будет получен метод решения задачи (5.5.1) с оценкой трудоемкости $O(m^2 n \times \times [\ln \varepsilon^{-1} + \ln n])$.

Зафиксируем произвольные константы $r > 0$ и $\tau \in (0, 1)$. Рассмотрим две последовательности векторов $\{ x_k \}_{k=0}^{\infty}, \{ d_k \}_{k=0}^{\infty} \subset R^n$, такие, что

1). $x_0 = 0$; $d_0^{(i)} = 1$, $i = 1,\dots,n$.

2). Очередные векторы x_{k+1} и d_{k+1} выбираются из следующих условий:

а) x_{k+1} может быть произвольной точкой из множества $E(d_k , x_k , r)$;

б)
$$d_{k+1}^{(j)} = \begin{cases} d(x_{k+1})^{(j)}, & \text{если } j \in \mathfrak{J}(k), \\ d_k^{(j)} & \text{в противном случае;} \end{cases}$$

здесь

$$\mathfrak{J}(k) = \{ j \mid s(x_{k+1})^{(j)} > (1 + \tau) d_k^{(j)} \} \cup \cup \{ j \mid s(x_{k+1})^{(j)} < (1 - \tau) d_k^{(j)} \}.$$

<u>Лемма 5.5.7.</u> При любом $N > 1$ справедлива оценка

$$\Sigma_{k=0}^{N-1} | \mathfrak{J}(k) | \le r \tau^{-1} n^{1/2} N,$$

где $| \mathfrak{J}(k) |$ - мощность множества $\mathfrak{J}(k)$.

Д о к а з а т е л ь с т в о. В силу условия а) при любом $N > 1$ имеем

$$N r^2 > \Sigma_{i=0}^{N-1} \| D(d_i)^{-1} (x_{i+1} - x_i) \|^2 \equiv$$

$$\equiv \sum_{i=0}^{N-1} \sum_{j=1}^{n} [(d_i^{(j)})^{-1} (x_{i+1}^{(j)} - x_i^{(j)})]^2 =$$
$$= \sum_{j=1}^{n} \sum_{i=0}^{N-1} [(d_i^{(j)})^{-1} (x_{i+1}^{(j)} - x_i^{(j)})]^2$$

Обозначим внутренние суммы этого выражения через S_j. Зафиксируем произвольный номер j, $1 \le j \le n$. Нас будет интересовать величина $m(j)$ – число изменений j-й координаты векторов d_i, $i = 0, \ldots, N - 1$. Введем целочисленную функцию $s(l)$, $l = 0, \ldots, m(j)$, значение которой соответствует номеру точки x_i, при формировании которой произошел l-й пересчет, $s(0) = 0$. Заметим, что при всех q, $s(l) \le q \le s(l + 1) - 1$, справедливо тождество $d_q^{(j)} \equiv \equiv d_{s(l)}^{(j)}$. Поэтому

$$S_j \ge \sum_{l=0}^{m(j)-1} [d_{s(l)}^{(j)}]^{-2} \sum_{q=s(l)}^{s(l+1)-1} (x_{q+1}^{(j)} - x_q^{(j)})^2 \ge$$
$$\ge \sum_{l=0}^{m(j)-1} [d_{s(l)}^{(j)}]^{-2} (x_{s(l+1)}^{(j)} - x_{s(l)}^{(j)})^2/(s(l+1) - s(l)) \ge$$
$$\ge \sum_{l=0}^{m(j)-1} [d_{s(l)}^{(j)}]^{-2} [\mid x_{s(l+1)}^{(j)} \mid - \mid x_{s(l)}^{(j)} \mid]^2 /$$
$$/(s(l+1) - s(l)) .$$

В силу способа пересчета векторов d_i

$$d_{s(l)}^{(j)} = 1 - \mid x_{s(l)}^{(j)} \mid , \quad d_{s(l+1)}^{(j)} = 1 - \mid x_{s(l+1)}^{(j)} \mid$$

Таким образом,

$$S_j \ge \sum_{l=0}^{m(j)-1} [1 - d_{s(l+1)}^{(j)} / d_{s(l)}^{(j)}]^2 /(s(l+1) - s(l)) .$$

Из критерия пересчета векторов d_i следует, что существует две возможности:

либо $d_{s(l+1)}^{(j)} / d_{s(l)}^{(j)} \ge 1 + \tau$,

либо $d_{s(l+1)}^{(j)} / d_{s(l)}^{(j)} \le 1 - \tau$

В любом случае $\mid 1 - d_{s(l+1)}^{(j)} / d_{s(l)}^{(j)} \mid \ge \tau$. Поэтому

$$S_j \ge \tau^2 \sum_{l=0}^{m(j)-1} (s(l+1) - s(l))^{-1}$$

Заметим, что сумма по l чисел $s(l+1) - s(l)$ не превосходит N, причем число слагаемых в сумме равно $m(j)$. Воспользовавшись теперь соотношением

$$\min\{ \sum_{i=1}^{m} t_i^{-1} \mid t_i \ge 0 , \sum_{i=1}^{m} t_i \le T \} = m^2 / T,$$

получаем, что для любого j, $1 \le j \le n$,

$S_j > \tau^2 m(j)^2 N^{-1}$

Отсюда сразу следует, что

$\sum_{j=1}^{n} m(j)^2 < \tau^{-2} N \sum_{j=1}^{n} S_j < N^2 \alpha^2 \tau^{-2}$,

$\sum_{k=0}^{N-1} | \mathcal{J}(k) | \equiv \sum_{j=1}^{n} m(j) <$

$< [n \sum_{j=1}^{n} m(j)^2]^{1/2} < N n^{1/2} \alpha \tau^{-1}$. □

Теперь можно привести общую схему модифицированного прямого метода параллельных траекторий.

Метод $\mathcal{P}_1(r , \delta , \tau)$

0. Полагаем

$h = [1 - (\tau + r)^2]^2 [1 + (\tau + r)^2]^{-1}$,

$x_0 = 0 \in R^n$, $H_0 = [A A^T]^{-1}$,

$d_0 = (1,\ldots,1)^T \in R^n$

1. k -я итерация ($k > 0$).

а) Полагаем

$$x'_k = x_k + \delta r \frac{D_k^2 A^T H_k b}{\langle H_k b, b \rangle^{1/2}} ,$$

$x_{k+1} = x'_k - h H_k^P F'(x'_k)$,

где $D_k = \mathrm{diag}(d_k)$; $H_k^P = D_k^2 - D_k^2 A^T H_k A D_k^2$

б) формируем множество

$\mathcal{J}(k) = \{ j \mid s(x_{k+1})^{(j)} > (1 + \tau) d_k^{(j)} \} \cup$

$\cup \{ j \mid s(x_{k+1})^{(j)} < (1 - \tau) d_k^{(j)} \}$;

в) производим пересчет вектора d_k

$$d_{k+1}^{(j)} = \begin{cases} s(x_{k+1})^{(j)} , & \text{если } j \in \mathcal{J}(k) , \\ d_k^{(j)} & \text{в противном случае} \end{cases}$$

г) с помощью формул одноранговой коррекции производим пересчет матрицы H_k

$H_{k+1} = [H_k^{-1} + \sum_{j \in \mathcal{J}(k)} [(d_{k+1}^{(j)})^2 - (d_k^{(j)})^2] a_j a_j^T]^{-1}$

Итерация закончена.

<u>Теорема 5.5.2.</u> Пусть для решения задачи (5.5.1) приме-

няется метод $\mathcal{P}_1(r, \delta, \tau)$ с параметрами r, δ, τ, такими, что

$r \in (0, 1/12)$, $\delta = 4r$, $\tau = r$.

Тогда все точки последовательности $\{x_k\}_{k=0}^{\infty}$, построенной методом $\mathcal{P}_1(r, \delta, \tau)$, обладают следующими свойствами:

1) $x_{k+1} \in E(d_k, x_k, 2r) \subset K$

2) для последовательности величин $\{t_k\}_{k=0}^{\infty}$ такой, что $t_0 = 0$ $t_{k+1} = t_k + \delta r \langle H_k b, b\rangle^{-1/2}$, справедливо соотношение $A x_k = t_k b$;

3) при любом $k > 0$ выполнено включение
$x^*(t_k) \in E(d_k, x_k, æ r)$,
где $æ = 2(r + \tau)[1 + (r + \tau)^2]^{-1}$

4) последовательность $\{t_k\}_{k=0}^{\infty}$ сходится к t^* и справедлива следующая оценка скорости сходимости:

$$(t^* - t_k)\langle [A A^T]^{-1} b, b\rangle^{1/2} \leq$$
$$\leq n\, C(r, \delta, \tau)(1 + q n^{-1/2})^{-k+1},$$

где $C(r, \delta, \tau) = \delta^{-1} r^{-1}(1 - r - \tau)^{-2}$; $q = \delta r \times$
$\times (1 - r - \tau)^2 (1 + æ(r + \tau))^{-1}$.

Если при этом пересчет матриц H_k на шаге г) производить с помощью формул одноранговой коррекции, то суммарная трудоемкость таких пересчетов за первые N итераций, где N - любое натуральное число, не превысит $2 r \tau^{-1} n^{1/2} (N + 1)$

Д о к а з а т е л ь с т в о. Как и в методе $\mathcal{P}_0(r, \delta)$, в методе $\mathcal{P}_1(r, \delta, \tau)$ в силу способа пересчета матрицы H_k и вектора d_k справедливы соотношения

$H_k = H(\dot{d}_k)$, $(1 - \tau) d_k \leq s(x_k) \leq (1 + \tau)^{-1} d_k$

Отсюда получаем, что при любом r, $0 < r < 1 - \tau$, функция $s(x)$ будет $(d_k, r, r + \tau)$-согласованной во всех точках x_k.

Утверждения 1),2) доказываются точно так же как и в теореме 5.5.1.

Докажем утверждение 3). Доказательство будем вести по индукции. При $k = 0$ это утверждение верно. Пусть оно выполнено и при некотором $k > 0$. Параметры r, δ, τ в теореме 5.5.1 выбраны так, что условия леммы 5.5.5 выполнены. Поэтому

$$\| D_k^{-1} (x_{k+1}^* - x'_k) \| < r (1 - \delta)$$

Кроме того, из леммы 5.5.6

$$\| D_k^{-1} (x_{k+1}^* - x_{k+1}) \| < æ \| D_k^{-1} (x_{k+1}^* - x'_k) \| .$$

Таким образом, учитывая $(d_{k+1}, 0, \tau)$ - согласованность функции $s(x)$ в точке x_{k+1}, $(d_k, 2r, 2r + \tau)$ - согласованность функции $s(x)$ в точке x_k и предыдущие неравенства, имеем

$$(1 + \tau)^{-1} \| D_{k+1}^{-1} (x_{k+1}^* - x_{k+1}) \| <$$
$$< \| D(s(x_{k+1}))^{-1} (x_{k+1}^* - x_{k+1}) \| <$$
$$< (1 - 2r - \tau)^{-1} \| D_k^{-1} (x_{k+1}^* - x_{k+1}) \| <$$
$$< (1 - 2r - \tau)^{-1} r (1 - \delta) æ = æ r .$$

Осталось заметить, что $\tau = r$, $\delta = 4 r$ и, значит,

$$(1 - 2 r - \tau) (1 + \tau)^{-1} > 1 - \delta$$

Утверждение 4) доказывается так же, как и в теореме 5.5.1, с той лишь разницей, что здесь функция $s(x)$ будет $(d_k, r, r + \tau)$-согласованной в точке x_k

Осталось заметить, что способ пересчета векторов x_k и d_k удовлетворяет в силу утверждения 1) теоремы 5.5.2 предположениям леммы 5.5.7. Поэтому за любые k итераций, $0 < k < N - 1$, число одноранговых коррекций матрицы H_k не превысит величины $2 (k + 1) n^{1/2} r \tau^{-1}$ □

Методы $\mathcal{P}_0(r, \delta)$ и $\mathcal{P}_1(r, \delta, \tau)$ могут применяться для нахождения допустимого решения системы линейных неравенств. А именно, пусть требуется найти произвольную точку x, такую, что

$$A x = b \quad , \quad x \in K \tag{5.5.2}$$

Предположим, что система (5.5.2) является θ - устойчивой. Сформируем задачу

$$\max \{ t \mid A x = t b , \ x \in K \}$$

и применим для ее решения любой из методов $\mathcal{P}_0(r, \delta)$, $\mathcal{P}_1(r, \delta, \tau)$, введя в них следующий критерий.

Критерий прерывания (5.5.3)

Если $t_k + \delta r \langle H_k b, b \rangle^{-1/2} > 1$, то делаем еще одну итерацию соответствующего метода с

$$\delta = r^{-1} (1 - t_k) \langle H_k b, b \rangle^{1/2},$$

и после этого останавливаемся, положив $x_* = x_{k+1}$.

Теорема 5.5.3. Если для решения θ -устойчивой системы неравенств (5.5.2) применяется любой из методов $\mathcal{P}_0(r, \delta)$, $\mathcal{P}_1(r, \delta, \tau)$ (один из возможных способов выбора параметров этих методов указан в теоремах 5.5.1, 5.5.2), снабженный критерием прерывания (5.5.3), то не более чем за N итераций, где

$$N = O(n^{1/2} (\ln n + \ln \theta^{-1})),$$

в качестве ответа x_* будет выдано допустимое решение системы (5.5.1), такое, что

$$x^*(1) \in E(d_k , x_* , æ r) .$$

При этом суммарное число арифметических операций для первого метода не превзойдет величины порядка

$$O(n^{1.5} m^2 [\ln n + \ln \theta^{-1}]) ,$$

а для второго метода - величины порядка

$$O(n m^2 [\ln n + \ln \theta^{-1}]).$$

Доказательство. Из θ - устойчивости задачи (5.5.2) следует, что $t^* - 1 > \theta (1 - \theta)^{-1}$. Пусть критерий прерывания (5.5.3) не сработал на первой итерации.

Из этого заключаем, что

$$1 < \delta\, r \langle H_k b, b \rangle^{-1/2} \equiv \delta\, r \langle [A A^T]^{-1} b, b \rangle^{-1/2}$$

Поэтому в силу теорем 5.5.1, 5.5.2 для обоих методов будет справедлива оценка скорости сходимости вида

$$t^* - t_k < C_1 n \exp\{ - C_2 n^{-1/2} k \},$$

где C_1, C_2 - абсолютные константы. Следовательно, не более чем за $N = O(n^{1/2} (\ln n + \ln \theta^{-1}))$ итераций заведомо выполнится неравенство $t_N > 1$. Критерий прерывания (5.5.3) срабатывает в тот момент, когда $t_k < 1 < t_{k+1}$. Правило выбора параметра δ на $(k+1)$-й итерации гарантирует выполнение равенства $A x_{k+1} = b$. Включение точки $x^*(1)$ в множество $E(d_k, x_*, æ r)$ обосновано в теоремах 5.5.1, 5.5.2. □

5.6. Метод барьеров для задачи квадратичного программирования

Рассмотрим задачу квадратичного программирования в следующей постановке

$$f(x) = 0.5 \langle Q x, x \rangle - \langle c, x \rangle \to \min \mid x \in K \cap \mathcal{L}, \tag{5.6.1}$$

где $Q \geq 0$ - симметрическая неотрицательно определенная $n\times n$-матрица; $c \in R^n$. Введем штрафную функцию

$$\Psi(t, x) = t f(x) + F(x),$$

где барьер $F(x)$ определен в §5.2.

Будем говорить, что дважды дифференцируемая функция $\varphi(x)$ допускает (B, r, τ)-аппроксимацию в точке x', где B - симметрическая неотрицательно определенная $n\times n$

-матрица. $r > 0$, $0 < \tau < 1$, если для любого y из множества

$$V(B, x', r) = \{ x \in R^n \mid \langle B(x - x'), x - x' \rangle < r^2 \}$$

справедливы соотношения

$$(1 + \tau)^{-2} B < \varphi''(x) < (1 - \tau)^{-2} B$$

<u>Лемма 5.6.1.</u> Если вектор-функция $s(x)$ является (d, r, τ)-согласованной в точке x', то при любом t, таком, что

$$(1 + \tau)^{-2} t' < t < (1 - \tau)^{-2} t',$$

где $t' > 0$, функция $\Psi(t, x)$ допускает (B, r, τ)-аппроксимацию в точке x' с $B = t' Q + D(d)^{-2}$

Д о к а з а т е л ь с т в о. Заметим, что при любом $t > 0$ множество

$$V([t Q + D(d)^{-2}], x', r) \subseteq E(d, x', r)$$

Поэтому в силу (d, r, τ)-согласованности функции $s(x)$ в точке x' при любом $x \in V([t Q + D(d)^{-2}], x', r)$ имеем

$$\Psi''(t, x) = t Q + D(s(x))^{-2} > t Q + (1 + \tau)^{-2} D(d)^{-2} >$$
$$> (1 + \tau)^{-2} [t' Q + D(d)^{-2}];$$
$$\Psi''(t, x) < t Q + (1 - \tau)^{-2} D(d)^{-2} <$$
$$< (1 - \tau)^{-2} [t' Q + D(d)^{-2}] \quad \square$$

Положим

$$x^*(t) = \operatorname{argmin} \{ \Psi(t, x) \mid x \in K \cap \mathcal{L} \}, \ t > 0.$$

<u>Лемма 5.6.2.</u> Пусть функция $\Psi(t, x')$ допускает (B, r, τ)-аппроксимацию в точке x' и $x^*(t) \in V(B, x', æ r)$, где $æ \in [0, 1)$. Тогда при всех Δ, таких, что

$$0 < \Delta < t (1 - æ) r (1 - \tau)^2 (1 + \tau)^{-2} n^{-1/2}, \qquad (5.6.2)$$

точка $x^*(t + \Delta) \in V(B, x', r)$.

Д о к а з а т е л ь с т в о. Пусть

$$x'(\Delta) = \operatorname{argmin} \{ \Psi(t + \Delta, x) \mid x \in V(B, x', r) \cap \mathcal{L} \}.$$

Заметим, что $\Psi''(t, x) < \Psi''(t + \Delta, x)$. Поэтому функция $\Psi(t + \Delta, x)$ будет сильно выпуклой в метрике, задаваемой матрицей B на множестве $V(B, x', r)$, с константой $(1+\tau)^{-2}$, и, следовательно,

$$(1+\tau)^{-2} \langle B(x'(\Delta) - x^*(t)), x'(\Delta) - x^*(t) \rangle \le$$
$$\le \langle \Psi'(t+\Delta, x^*(t)) - \Psi'(t+\Delta, x'(\Delta)),$$
$$x^*(t) - x'(\Delta) \rangle \le$$
$$\le \langle \Psi'(t+\Delta, x^*(t)), \quad x^*(t) - x'(\Delta) \rangle =$$
$$= \langle \Delta f'(x^*(t)), \quad x^*(t) - x'(\Delta) \rangle =$$
$$= -\Delta t^{-1} \langle F'(x^*(t)), \quad x^*(t) - x'(\Delta) \rangle \le$$
$$\le \Delta t^{-1} \langle [F''(x^*(t))]^{-1} F'(x^*(t)), F'(x^*(t)) \rangle^{1/2} \times$$
$$\times \langle F''(x^*(t))(x^*(t) - x'(\Delta)), x^*(t) - x'(\Delta) \rangle^{1/2}$$

Заметим, что

$$\langle F''(x^*(t))(x^*(t) - x'(\Delta)), x^*(t) - x'(\Delta) \rangle \le$$
$$\le \langle \Psi''(x^*(t))(x^*(t) - x'(\Delta)), x^*(t) - x'(\Delta) \rangle \le$$
$$\le (1-\tau)^{-2} \langle B(x^*(t) - x'(\Delta)), x^*(t) - x'(\Delta) \rangle$$

Кроме того, при любых $x \in \operatorname{int} K$

$$\langle [F''(x)]^{-1} F'(x), F'(x) \rangle = \| x \|^2 \le n$$

Поэтому окончательно получаем

$$\langle B(x'(\Delta) - x'), x'(\Delta) - x' \rangle^{1/2} \le$$
$$\le \langle B(x'(\Delta) - x^*(t)), x'(\Delta) - x^*(t) \rangle^{1/2} +$$
$$+ \langle B(x' - x^*(t)), x' - x^*(t) \rangle^{1/2} \le$$
$$\le \Delta t^{-1} n^{1/2} (1+\tau)^2 (1-\tau)^{-2} + æ r$$

Таким образом, если Δ удовлетворяет неравенству (5.6.2), то

$$\langle B(x'(\Delta) - x'), x'(\Delta) - x' \rangle \le r^2,$$

и, следовательно, $x'(\Delta) \equiv x^*(t+\Delta)$ □

<u>Лемма 5.6.3.</u> Пусть функция $\Psi(t', x')$ допускает (B, r, τ)-аппроксимацию в точке x', $Ax' = b$ и $x^*(t') \in V(B, x', r)$. Тогда для точки

$$x_+ = x' - h P(A, B) \Psi'(t', x'),$$

где $h = (1-\tau^2)^2 (1+\tau^2)^{-1}$; $P(A, B) = B^{-1} -$

$- B^{-1} A^T [A B^{-1} A^T]^{-1} A B^{-1}$, справедливо неравенство
$\langle B(x_+ - x^*(t')), x_+ - x^*(t') \rangle < 4 \tau^2 (1 + \tau^2)^{-2} r^2$

Доказательство. Действительно, спектр матриц вторых производных в метрике, задаваемой матрицей B, у функции $\Psi(t', x)$ будет ограничен на эллипсоиде $V(B, x', r)$ сверху и снизу соответственно константами $L = (1 - \tau)^{-2}$ и $l = (1 + \tau)^{-2}$. Поэтому

$$\langle B(x_+ - x^*(t')), x_+ - x^*(t') \rangle =$$
$$= \langle B(x' - x^*(t')), x' - x^*(t') \rangle -$$
$$- 2h \langle B P(A, B) \Psi'(t', x'), x' - x^*(t') \rangle +$$
$$+ h^2 \langle B P(A, B) \Psi'(t', x'), P(A, B) \Psi'(t', x') \rangle$$

Заметим, что

$$\langle B P(A, B) \Psi'(t', x'), x' - x^*(t') \rangle =$$
$$= \langle \Psi'(t', x') - \Psi'(t', x^*(t')), x' - x^*(t') \rangle$$
$$\langle B P(A, B) \Psi'(t', x'), P(A, B) \Psi'(t', x') \rangle =$$
$$= \langle P(A, B) [\Psi'(t', x') - \Psi'(t', x^*(t'))],$$
$$\Psi'(t', x') - \Psi'(t', x^*(t')) \rangle <$$
$$< \langle B^{-1} [\Psi'(t', x') - \Psi'(t', x^*(t'))],$$
$$\Psi'(t', x') - \Psi'(t', x^*(t')) \rangle$$

Следовательно,

$$\langle B(x_+ - x^*(t')), x_+ - x^*(t') \rangle <$$
$$< \langle B(x' - x^*(t')), x' - x^*(t') \rangle -$$
$$- 2h \langle \Psi'(t', x') - \Psi'(t', x^*(t')), x' - x^*(t') \rangle +$$
$$+ h^2 \langle B^{-1} [\Psi'(t', x') - \Psi'(t', x^*(t'))],$$
$$\Psi'(t', x') - \Psi'(t', x^*(t')) \rangle$$

Но

$$\langle \Psi'(t', x') - \Psi'(t', x^*(t')), x' - x^*(t') \rangle >$$
$$> l L (l + L)^{-1} \langle B(x_+ - x^*(t')), x_+ - x^*(t') \rangle +$$
$$+ (l + L)^{-1} \langle B^{-1} [\Psi'(t', x') - \Psi'(t', x^*(t'))],$$
$$\Psi'(t', x') - \Psi'(t', x^*(t')) \rangle.$$

где $l = (1 + \tau)^{-2}$; $L = (1 - \tau)^{-2}$. Таким образом,

$$\langle B (x_+ - x^*(t')) , x_+ - x^*(t') \rangle \le$$
$$\le (1 - 2 h l L (l + L)^{-1}) \langle B(x' - x^*(t')), x' - x^*(t') \rangle +$$
$$+ h(h - 2(l + L)^{-1}) \langle B^{-1}[\Psi'(t', x') - \Psi'(t', x^*(t'))],$$
$$\Psi'(t', x') - \Psi'(t', x^*(t')) \rangle$$

Осталось заметить, что

$$1 - 2 h l L (l + L)^{-1} = (L - l)^2 (L + l)^{-2} =$$
$$= 4 \tau^2 (1 + \tau^2)^{-2} \qquad \square$$

Лемма 5.6.4. Если функция $\Psi(t' , x')$ допускает (B, r, τ) -аппроксимацию в точке x' , $A x' = b$, $t' > 0$, и точка $x^*(t')$ принадлежит множеству $V(B, x', r)$, то $f(x') - f(x^*(t')) \le t^{-1} [(1 - \tau)^{-2} r^2 + n]$.

Д о к а з а т е л ь с т в о. Заметим, что

$$f(x') - f(x^*(t')) \le \langle f'(x') , x' - x^*(t') \rangle =$$
$$= t^{-1} \langle \Psi'(x') - F'(x') , x' - x^*(t') \rangle ;$$
$$\langle F'(x') , x^*(t') - x' \rangle \le \sum_{i=1}^{n} | x'^{(i)} | \le n$$

Далее, из ограниченности сверху спектра матриц вторых производных функции $\Psi(t', x)$ в метрике, задаваемой матрицей B на эллипсоиде $V(B, x', r)$, константой $(1 - \tau)^{-2}$ следует, что

$$\langle \Psi'(x') , x' - x^*(t') \rangle =$$
$$= \langle \Psi'(x') - \Psi'(x^*(t')) , x' - x^*(t') \rangle \le$$
$$\le (1 - \tau)^{-2} \langle B (x' - x^*(t')) , x' - x^*(t') \rangle \le$$
$$\le (1 - \tau)^{-2} r^2 \qquad \square$$

Теперь мы можем перейти к обоснованию методов решения задачи (5.6.1), базирующихся на методе барьеров

Оценка скорости сходимости метода барьеров, применяемого для решения задачи (5.6.1), выводится из следующей леммы

Лемма 5.6.5. При всяком $t > 0$ справедлиео неравенство $f(x^*(t)) - f(x^*) \le n t^{-1}$, где x^* - решение задачи (5.6.1).

Доказательство. Действительно,

$f(x^*(t)) - f(x^*) \le \langle f'(x^*(t)), x^*(t) - x^* \rangle =$
$= t^{-1} \langle F'(x^*(t)), x^* - x^*(t) \rangle \le$
$\le t^{-1} \sum_{i=1}^{n} | x^{*(i)}(t) | \le n t^{-1}$ □

Предположим, что система условий задачи (5.6.1) является θ-устойчивой, $0 < \theta \le 1$. Приведем общую схему метода $\mathcal{M}_0(r)$

Метод $\mathcal{M}_0(r)$

0 Любым из методов §5 строим начальную точку x_0, такую, что

а) $A x_0 = b$;

б) $x^*(0) \in V(F''(x_0), x_0, æ r)$,

где $æ = 2 r [1 - r - r^2 - r^3]^{-1}$;

полагаем

$t_0 = 0.5 r [(1 + r)^{-2} r - (1 - æ r)^{-2} æ r] \times$
$\times [\sum_{i=1}^{n} | (f'(x_0))^{(i)} |]^{-1}$;
$h = (1 - r^2)^2 (1 + r^2)^{-1}$;
$q = r (1 - æ) (1 - r)^2 (1 + r)^{-2}$.

1. k-я итерация ($k \ge 0$).

Полагаем

$x_{k+1} = x_k - h P(A, \Psi''(t_k, x_k)) \Psi'(t_k, x_k)$,
$t_{k+1} = (1 + q n^{-1/2}) t_k$.

Итерация закончена.

<u>Теорема 5.6.1.</u> Траектория точек $\{x_k\}_{k=0}^{\infty}$ и последовательность $\{t_k\}_{k=0}^{\infty}$, построенные методом $\mathcal{M}_0(r)$, $0 < r < 1/5$, обладают следующими свойствами:

1) при любом $k \ge 0$ выполняются включения

$x^*(t_k) \in V(\Psi''(t_k, x_k), x_k, r)$, (5.6.3)
$x^*(t_k) \in V(\Psi''(t_{k+1}, x_{k+1}), x_{k+1}, æ r)$, (5.6.4)

где $æ = 2 r [1 - r - r^2 - r^3]^{-1}$;

2) последовательность $\{ f(x_k) \}_{k=0}^{\infty}$ сходится к $f(x^*)$, причем при $k \geq 0$ справедлива оценка

$$f(x_k) - f(x^*) \leq (n+1)\, C_1 (1 + q\, n^{-1/2})^{-k},$$

где $q = r(1 - æ)(1 - r)^2 (1 + r)^{-2}$; $C_1 = 4\, r^{-1} \times$
$\times \sum_{i=1}^{n} | f'(x_0)^{(i)} | \, [(1 + r)^{-2} r - (1 - æ\, r)^{-2}\, æ r]^{-1}$

Доказательство. Предположим, что включение (5.6.3) справедливо при некотором $k \geq 0$. Тогда, воспользовавшись леммой 5.6.3 для

$$t' = t_k,\ x' = x_k,\ \tau = r,\ B = \Psi''(t_k, x_k),$$

получаем

$$\langle F''(x_k)(x_{k+1} - x_k), x_{k+1} - x_k \rangle^{1/2} \leq$$
$$\leq \langle B(x_{k+1} - x_k), x_{k+1} - x_k \rangle^{1/2} \leq$$
$$\leq \langle B(x_{k+1} - x^*(t_k)), x_{k+1} - x^*(t_k) \rangle^{1/2} +$$
$$+ \langle B(x_k - x^*(t_k)), x_k - x^*(t_k) \rangle^{1/2} \leq$$
$$\leq 2\, r^2 (1 + r^2)^{-1} + r \equiv r'$$

Поэтому, учитывая, что при заданном в условии теоремы интервале изменения параметра r справедливо неравенство

$$t_{k+1} \leq (1 - r')^{-2}\, t_k,$$

имеем

$$\Psi''(t_{k+1}, x_{k+1}) \leq (1 - r')^{-2} B$$

Следовательно,

$$\langle \Psi''(t_{k+1}, x_{k+1})(x_{k+1} - x^*(t_k)), x_{k+1} - x^*(t_k) \rangle \leq$$
$$\leq (1 - r')^{-2} \langle B(x_{k+1} - x^*(t_k)), x_{k+1} - x^*(t_k) \rangle \leq$$
$$\leq 4\, r^4 [(1 - r')(1 + r^2)]^{-2} = æ^2 r^2$$

т.е. включение (5.6.4) для этого номера k выполнено

Пусть теперь включение (5.6.4) выполнено при некотором $k \geq 0$. Заметим, что функция $\Psi(t_k, x)$ допускает в точке x_{k+1} (B, r, r) - аппроксимацию с $B = \Psi''(t_{k+1}, x_{k+1})$. Поэтому в силу леммы 5.6.2 включение (5.6.3) выполнено для номера $k + 1$.

Таким образом, для доказательства утверждения 1) теоремы достаточно показать, что включение (5.6.3) справедливо при $k = 0$. Выберем

$$x' = \operatorname{argmin}\{ \Psi(t_0, x) \mid x \in \mathcal{L} \cap V(\Psi''(t_0, x_0), x_0, r) \subset K \}.$$

Заметим, что функция $\Psi(t_0, x)$ допускает (B, r, r)-аппроксимацию в точке x_0 с $B = \Psi''(t_0, x_0)$. Поэтому

$$(1 + r)^{-2} \langle B(x' - x_0), x' - x_0 \rangle \le$$
$$\le \langle \Psi'(t_0, x_0) - \Psi'(t_0, x'), x_0 - x' \rangle \le$$
$$\le \langle \Psi'(t_0, x_0), x_0 - x' \rangle = t_0 \langle f'(x_0), x_0 - x' \rangle +$$
$$+ \langle F'(x_0), x_0 - x' \rangle$$

Нетрудно видеть, что

$$\langle f'(x_0), x_0 - x' \rangle \le 2 \sum_{i=1}^{n} | (f(x_0))^{(i)} | ;$$

$$\langle F'(x_0), x_0 - x' \rangle = \langle F'(x_0) - F'(x^*(0)), x_0 - x' \rangle \le$$
$$\le \langle [F''(x_0)]^{-1} (F'(x_0) - F'(x^*(0))), F'(x_0) - F'(x^*(0)) \rangle^{1/2} \times$$
$$\times \langle F''(x_0)(x_0 - x'), x_0 - x' \rangle^{1/2} \le$$
$$\le \langle [F''(x_0)]^{-1} (F'(x_0) - F'(x^*(0))), F'(x_0) - F'(x^*(0)) \rangle^{1/2} \times$$
$$\times \langle B(x_0 - x'), x_0 - x' \rangle^{1/2}.$$

По способу построения точки x_0 точка $x^*(0)$ принадлежит множеству $V(F''(x_0), x_0, æ r)$. Поэтому

$$\langle [F''(x_0)]^{-1} (F'(x_0) - F'(x^*(0))), F'(x_0) - F'(x^*(0)) \rangle \le$$
$$\le (1 - æ r)^{-4} \langle F''(x_0)(x_0 - x^*(0)), x_0 - x^*(0) \rangle \le$$
$$\le (1 - æ r)^{-4} (æ r)^2.$$

Положим $R = \langle B(x_0 - x'), x_0 - x' \rangle^{1/2}$. Тогда из предыдущих соотношений имеем

$$(1 + r)^{-2} R^2 \le 2 t_0 \sum_{i=1}^{n} | (f'(x_0))^{(i)} | +$$
$$+ (1 - æ r)^{-2} æ r R.$$

В силу выбора t_0 правая часть последнего неравенства не превышает величины

$$(1 - æ r)^{-2} æ r R + r [(1 + r)^{-2} r - (1 - æ r)^{-2} æ r].$$

Откуда заключаем, что $R < r$, и, следовательно, $x' \equiv \equiv x^*(0)$. Таким образом, включение (5.6.3) при $k = 0$, а значит, и утверждение 1) теоремы 5.6.1 доказано.

Утверждение 2) теоремы вытекает непосредственно из лемм 5.6.4, 5.6.5. □

Замечания. 1. В силу теоремы 5.5.3 число арифметических операций, необходимое для построения начальной точки x_0 методами из §5.5, оценивается сверху величиной порядка $O(n^{1.5} m^2 [\ln n + \ln \theta^{-1}])$ для метода $\mathcal{P}_0(r, \delta)$ и величиной порядка $O(n m^2 [\ln n + \ln \theta^{-1}])$ для метода $\mathcal{P}_1(r, \delta, \tau)$.

2. Если для нахождения начальной точки x_0 применяется метод $\mathcal{P}_0(r, \delta)$, то общая трудоемкость получения решения задачи (5.6.1) с погрешностью ε для метода $\mathcal{M}_0(r)$ оценивается сверху величиной порядка

$$O(n^{1.5} m^2 (\ln \theta^{-1} + \ln n) + n^{3.5} (\ln \varepsilon^{-1} + \ln n)),$$

где θ - параметр устойчивости системы линейных уравнений.

3. Оптимальное с точки зрения оценок трудоемкости значение параметра r для метода $\mathcal{M}_0(r)$ близко к 0.13.

Как и все базовые версии методов из §5.3 - 5.5, метод $\mathcal{M}_0(r)$ можно модифицировать таким образом, что пропадет необходимость заново вычислять на каждой итерации матрицу $P(A, \Psi''(t_k; x_k))$. В результате удается получить метод решения θ-устойчивой задачи (5.6.1) с оценкой трудоемкости $O(n m^2 (\ln \theta^{-1} + \ln n) + n^3 (\ln \varepsilon^{-1} + \ln n))$, где ε - требуемая точность решения задачи.

Приведем общую схему такого метода.

Метод $\mathcal{M}_1(r, \tau)$

0. С помощью метода $\mathcal{P}_1(r, 4r, \tau)$ находим начальную точку x_0, такую, что:

а) $A x_0 = b$;

б) $x^*(0) \in V(F''(x_0), x_0, \varkappa r)$,

где $\varkappa = 2(r + \tau)[(1 - \tau)(1 + (r + \tau)^2]^{-1}$

Полагаем

$t_0 = 0.5\, r\, [(1 + r)^{-2} r - (1 - \varkappa r)^{-2} \varkappa r] \times$
$\times [\sum_{i=1}^{n} |(f'(x_0))^{(i)}|]^{-1}$.

$d_0 = s(x_0)$; $T_0 = t_0$.

$B_0 = [T_0 Q + D(s(x_0))^{-2}]^{-1}$,

$H_0 = [A B_0 A^T]^{-1}$,

$h = (1 - (r + \tau)^2)^2 (1 + (r + \tau)^2)^{-1}$,

$q = (r + \tau)(1 - \varkappa)(1 - r - \tau)^2 (1 + r + \tau)^{-2}$

1. k-я итерация ($k \geq 0$).

а) полагаем

$x_{k+1} = x_k - h P_k \Psi'(t_k, x_k)$,

где $P_k = B_k - B_k A^T H_k A B_k$;

$t_{k+1} = (1 + q n^{-1/2}) t_k$;

б) если $t_{k+1} > (1 - \tau)^{-2} T_k$, то полагаем

$T_{k+1} = (1 - \tau)^{-2} T_k$, $t_{k+1} = T_{k+1}$,

$\mathcal{J}(k) = \varnothing$, $d_{k+1} = s(x_{k+1})$,

$B_{k+1} = [T_{k+1} Q + D(s(x_{k+1}))^{-2}]^{-1}$,

$H_{k+1} = [A B_{k+1} A^T]^{-1}$;

в) в противном случае полагаем

$T_{k+1} = T_k$,

$\mathcal{J}(k) = \{ j \mid s(x_{k+1})^{(j)} > (1 + \tau) d_k^{(j)} \} \cup$
$\cup \{ j \mid s(x_{k+1})^{(j)} < (1 - \tau) d_k^{(j)} \}$;

производим пересчет вектора d_k:

$$d_{k+1}^{(j)} = \begin{cases} s(x_{k+1})^{(j)}, & \text{если } j \in \mathfrak{J}(k), \\ d_k^{(j)}; & \text{если нет} \end{cases}$$

формируем последовательности матриц $\{ B_{k,i} \}_{i=0}^{m(k)}$ и $\{ H_{k,i} \}_{i=0}^{m(k)}$, где $m(k)$ - число элементов множества $\mathfrak{J}(k)$ по следующим правилам:

$B_{k,0} = B_k$, $H_{k,0} = H_k$;

$B_{k,i+1} = B_{k,i} + \alpha_{k,i} B_{k,i} e_{s(k,i)} e_{s(k,i)}^T B_{k,i}$,

$H_{k,i+1} = H_{k,i} + \beta_{k,i} H_{k,i} u_{k,i} u_{k,i}^T H_{k,i}$.

$i = 0, \dots, m(k) - 1$,

где $s(k, i)$ - i-й элемент в множестве $\mathfrak{J}(k)$;

$$\alpha_{k,i} = \frac{\Delta_{k,i}}{1 + \Delta_{k,i} \langle B_{k,i} e_{s(k,i)}, e_{s(k,i)} \rangle}$$

$$\Delta_{k,i} = (d_{k+1}^{(j)})^{-2} - (d_k^{(j)})^{-2};$$

$$u_{k,i} = A B_{k,i+1} e_{s(k,i)};$$

$$\beta_{k,i} = \frac{\alpha_{k,i}}{1 + \alpha_{k,i} \langle H_{k,i} u_{k,i}, u_{k,i} \rangle}$$

полагаем $B_{k+1} = B_{k,m(k)}$; $H_{k+1} = H_{k,m(k)}$

Итерация закончена.

Теорема 5.6.2. Пусть для решения θ -устойчивой задачи (5.6.1) применяется метод $\mathfrak{M}_1(r, \tau)$ с параметрами r τ, такими, что

$r \in (0, 0.1)$, $\tau = 2r$

Тогда все точки последовательности $\{ x_k \}_{k=0}^{\infty}$, построенной методом $\mathfrak{M}_1(r, \tau)$, обладают следующими свойствами:

1) при любом $k > 0$ выполняются включения

$$x^*(t_k) \in V(B_k^{-1}, x_k, r), \qquad (5.6.5)$$

$$x^*(t_k) \in V(B_{k+1}^{-1}, x_{k+1}, \varkappa r), \qquad (5.6.6)$$

где $\varkappa = 2(r + \tau) [(1 - \tau)(1 + (r + \tau)^2]^{-1}$

2) последовательность $\{ f(x_k) \}_{k=0}^{\infty}$ сходится к $f(x^*)$.

причем при $k > 0$ справедлива следующая оценка скорости сходимости:

$$f(x_k) - f(x^*) < (n + 1) C_1 (1 + q n^{-1/2})^{-k},$$

где $q = (r + \tau)(1 - æ)(1 - r - \tau)^2 (1 + r + \tau)^{-2}$;

$$C_1 = 4 r^{-1} \sum_{i=1}^{n} | f'(x_0)^{(i)} | \times$$

$$\times [(1 + r)^{-2} r - (1 - æ r)^{-2} æ r]^{-1}$$

3) при всех $k > 0$

$$x_{k+1} \in E(d_k, x_k, 2r) \subset K.$$

При этом суммарная трудоемкость выполнения первых N итераций, где N - любое натуральное число, не превысит величины порядка $O(n^3 + n^{2.5} N)$.

Доказательство. Нетрудно видеть, что в методе $\mathcal{M}_1(r, \tau)$ поддерживается выполнение соотношений:

$$(1 - \tau) d_k < s(x_k) < (1 + \tau) d_k,$$

$$T_k < t_k < (1 - \tau)^{-2} T_k,$$

$$B_k^{-1} = T_k Q + D(d_k)^{-2},$$

$$H_k = [A B_k A^T]^{-1},$$

$$P_k = P(A, B_k^{-1}),$$

$$A x_k = b.$$

Отсюда, в частности, с помощью лемм 5.5.1 и 5.6.1, можно получить, что функция $\Psi(t, x)$ допускает $(B_k^{-1}, r, r + \tau)$-аппроксимацию в точках x_k при t, таких, что

$$T_k < t < (1 - \tau)^{-2} T_k.$$

Докажем утверждение 1) теоремы. Справедливость включения (5.6.5) при $k = 0$ обосновывается так же, как и в теореме 5.6.1, в силу того, что

$$T_0 = t_0, \quad F''(x_0) = D_0^{-2},$$

и, значит,

$$\Psi''(t_0, x_0) = T_0 Q + D(d_0)^{-2}$$

Пусть теперь включение (5.6.5) справедливо при некотором $k > 0$. Тогда, воспользовавшись леммой 5.6.3 для

$t' = t_k$, $x' = x_k$, $\tau = r + \tau$, $B = B_k^{-1}$,

получаем

$$\langle D(d_k)^{-2}(x_{k+1} - x_k), x_{k+1} - x_k \rangle^{1/2} <$$
$$< \langle B_k^{-1}(x_{k+1} - x_k)), x_{k+1} - x_k \rangle^{1/2} <$$
$$< \langle B_k^{-1}(x_{k+1} - x^*(t_k)), x_{k+1} - x^*(t_k) \rangle^{1/2} +$$
$$+ \langle B_k^{-1}(x_k - x^*(t_k)), x_k - x^*(t_k) \rangle^{1/2} <$$
$$< 2r(r + \tau)(1 + (r + \tau)^2)^{-1} + r < 2r = \tau$$

(таким образом, доказав (5.6.5), мы докажем и утверждение 3) теоремы 5.6.2). Поэтому, учитывая, что при заданном в условии теоремы интервале изменения параметра r и способе выбора τ всегда имеет место неравенство

$$T_{k+1} < (1 - \tau)^{-2} T_k ,$$

получаем

$$B_{k+1}^{-1} < (1 - \tau)^{-2} B_k^{-1}$$

Следовательно,

$$\langle B_{k+1}^{-1}(x_{k+1} - x^*(t_k)), x_{k+1} - x^*(t_k) \rangle <$$
$$< (1 - \tau)^{-2} \langle B_k^{-1}(x_{k+1} - x^*(t_k)), x_{k+1} - x^*(t_k) \rangle <$$
$$< 4r^2(r + \tau)^2 [(1 - \tau)(1 + (r + \tau)^2)]^{-2} = æ^2 r^2 ,$$

т.е. включение (5.6.6) для этого номера k выполнено.

Пусть теперь включение (5.6.6) выполнено при некотором $k > 0$. Как уже отмечалось, функция $\Psi(t, x)$ допускает в точке x_{k+1} $(B_{k+1}^{-1}, r, r + \tau)$-аппроксимацию. Поэтому в силу леммы 5.6.2 включение (5.6.6) выполнено и для номера $k + 1$. Таким образом, утверждение 1) теоремы 5.6.2 доказано.

Утверждение 2) непосредственно вытекает из лемм 5.6.4, 5.6.5.

Для завершения доказательства осталось оценить суммарную трудоемкость выполнения первых N итераций метода.

На каждой итерации этого метода основной объем вычислений приходится на п.п. б) и в) Заметим, что в п. б) операции по вычислению матриц B_{k+1} и H_{k+1} выполняются лишь один раз за $O(n^{1/2})$ итераций. Поэтому суммарный объем вычислений за первые N итераций, производящихся в этом пункте, имеет порядок $O(n^3 + n^{2.5} N)$ Суммарную арифметическую стоимость выполнения п. в) можно оценить с помощью леммы 6.5.7. Действительно, в силу утверждения 3) теоремы 5.6.2 условия этой леммы выполнены. Поэтому общее число арифметических операций, выполняемых в этом пункте, не превысит величины порядка $O(n^{2.5} N)$. □

Следствие 5.6.1. Общая трудоемкость получения решения задачи (5.6.1) с погрешностью ε для метода $\mathcal{M}_1(r, \tau)$ оценивается сверху величиной порядка

$$O(n m^2 (\ln \theta^{-1} + \ln n) + n^3 (\ln \varepsilon^{-1} + \ln n)),$$

где θ - параметр устойчивости системы линейных неравенств.

Близкий к оптимальному способ выбора параметров r, τ в методе $\mathcal{M}_1$ следующий: $r = 0.05$, $\tau = 0.01$.

5.7. Другие полиномиальные алгоритмы линейного и квадратичного программирования

Описанные итеративные методы решения задач линейного и квадратичного программирования на кубе могут применяться и для решения задач на областях другой структуры. Для этого необходимо лишь правильно выбрать соответствующую барьерную функцию $F(x)$. Так, например, для стандартного симплекса в качестве барьера можно взять

$$F_0(x) = -\sum_{i=1}^{n} \ln x^{(i)}, \tag{5.7.1}$$

для многогранника

$\{ x \mid \langle a_i , x \rangle \leq b^{(i)} , i = 1,\dots,m \}$,

$m > n$, имеющего непустую внутренность, можно взять

$$F_1(x) = - \sum_{i=1}^{n} \ln [b^{(i)} - \langle a_i , x \rangle] \qquad (5.7.2)$$

Основными свойствами барьерной функции, делающими ее пригодной для применения описанных методов являются:

а) существование, невырожденность и положительная определенность матриц вторых производных во внутренности выпуклого множества G ;

б) выполнение для любой точки $x' \in \inf G$ включения

$\{ x \mid \langle F''(x') (x - x'), x - x' \rangle \leq r^2 \} \subset \operatorname{int} G$

при некотором $r > 0$;

в) "согласованность" матриц вторых производных, т.е. выполнение неравенства

$(1 - \alpha r) F''(x') \leq F''(x) \leq (1 + \alpha r) F''(x')$

при некотором $\alpha > 0$ для любых $r \in [0, 1)$ и любых точек x' из $\operatorname{int} G$, таких, что

$\langle F''(x') (x - x'), x - x' \rangle \leq r^2$ ·

г) ограниченность "ньютоновского приращения"

$\langle [F''(x)]^{-1} F'(x) , F'(x) \rangle \leq \beta$

для всех x из $\operatorname{int} G$ при некотором $\beta > 0$

Множество областей, допускающих построение таких барьеров довольно велико. Так, например, в него входят области, задаваемые квадратичными ограничениями. Если для области G удалось построить барьер $F(x)$, обладающий свойствами а)-г), то для минимизации линейных или квадратичных функций на этом множестве можно применять как метод барьеров, так и прямой метод параллельных траекторий. При этом число итераций этих методов, необходимое для получения решения с точностью ε , будет ограничено величиной порядка $O(\beta^{1/2} \times$

$\times \ln 1/\varepsilon$) , где β - параметр барьера из свойства г). Отметим, что параметр барьеров (5.7.1), (5.7.2) равен n,

Приведем схему прямого метода параллельных траекторий для задачи

$\langle b, x \rangle \to \min \mid x \in G$,

где множество G ограничено и имеет непустую внутренность. Предположим, что у множества G существует барьер $F(x)$, удовлетворяющий свойствам а)-г).

Метод $\mathcal{P}_2(h)$.

0. Выбираем в качестве x_0 минимум функции $F(x)$ на множестве G.

1. k -я итерация ($k > 0$).

Полагаем

$$x'_k = x_k - h \frac{[F''(x_k)]^{-1} b}{\langle [F''(x_k)]^{-1} b , b \rangle^{1/2}} ,$$

$$x_{k+1} = x'_k + \operatorname{argmin} \{ \langle F'(x'_k) , s \rangle + $$
$$+ 0.5 \langle F''(x_k) s, s \rangle \mid \langle b , s \rangle = 0 \} .$$

Итерация закончена

Мы не будем приводить здесь доказательства оценки скорости сходимости метода $\mathcal{P}_2(h)$, так как оно почти дословно повторяет рассуждения §5.5. Отметим лишь, что при правильно выбранном $h > 0$ метод $\mathcal{P}_2(h)$ осуществляет отслеживание траектории

$$x^*(t) = \operatorname{argmin} \{ \langle b , x \rangle + t F(x) \mid x \in G \}.$$

Шаг $x_k \to x'_k$ делается по касательному направлению к этой траектории, а шаг $x'_k \to x_{k+1}$ позволяет вернуться в достаточно малую окрестность траектории.

Заметим, что все методы, рассмотренные в §5.4 - 5.6, основаны на отслеживании соответствующих траекторий с помощью выполнения на каждой итерации одного шага метода Ньютона. При этом, например, в методе барьеров штрафной параметр, отвечающий за скорость сходимости метода, удается увеличивать на каждой итерации только в $1 + O(n^{-1/2})$ раз. При увеличении скорости роста штрафного параметра приходится увеличивать и число вспомогательных шагов, необходимых для возврата точки в окрестность отслеживаемой траектории. Оценим эффективность двух схем таких "многошаговых" методов барьеров, один из которых использует на каждой итерации метод градиентного спуска (см. §4.4), а другой - оптимальный метод минимизации сильно выпуклых функций с липшицевым градиентом из §4.1.

Рассмотрим задачу квадратичного программирования в следующей постановке:

$$f(x) = 0.5 \langle Qx, x \rangle - \langle b, x \rangle \to \min \mid x \in G, \qquad (5.7.3)$$

где Q - симметрическая неотрицательно определенная $n \times n$ -матрица. Для упрощения изложения будем считать, что множество G имеет вид

$$G = \{ x \in R^n_+ \mid Ax = Ae, \langle e, x \rangle = n \},$$

где A - невырожденная $m \times n$ - матрица; $e = (1, \ldots, 1)^T \in R^n$.

Положим

$$\Psi(t, x) = t f(x) + F_0(x),$$

$$x^*(t) = \arg\min \{ \Psi(t, x) \mid x \in G \}.$$

Ясно, что в силу определения множества G точка $x^*(0) = e$.

<u>Лемма 5.7.1.</u> При любых $t_1, t_2 > 0$ справедливо тождество

$$t_2 \langle Q(x^*(t_1) - x^*(t_2)), x^*(t_1) - x^*(t_2) \rangle +$$
$$+ \sum_{i=1}^{n} [(x^*(t_1))^{(i)} - (t_2/t_1)^{1/2} (x^*(t_2))^{(i)}]^2 /$$
$$/[(x^*(t_1))^{(i)} (x^*(t_2))^{(i)}] \equiv n [(t_2/t_1)^{1/2} - 1]^2.$$

Доказательство. Действительно, при любом $t > 0$ и любом h, таком, что $Ah = 0$, $\langle e, h \rangle = 0$, имеем

$\langle\, t\, f'(\, x^*(\, t\,)\,) + F'(\, x^*(\, t\,)\,)\, .\, h\, \rangle = 0$,

т.е.

$$t^{-1} \sum_{i=1}^{n} h^{(i)} / (\, x^*(t)\,)^{(i)} = \langle\, Q\, x^*(t) - b,\, h\, \rangle \qquad (5.7.4)$$

Вычитая равенства (5.7.4) при $t = t_1$ и $t = t_2$ друг из друга, получаем

$$\sum_{i=1}^{n} \{\, [\, t_1\, (\, x^*(\, t_1\,)\,)^{(i)}\,]^{-1} -$$

$$- [\, t_2\, (\, x^*(\, t_2\,)\,)^{(i)}\,]^{-1}\, \}\, h^{(i)} =$$

$$= \langle\, Q\, (\, x^*(\, t_1\,) - x^*(\, t_2\,)\,)\, ,\, h\, \rangle$$

Подставляя теперь в последнее равенство $h = x^*(t_1) - x^*(t_2)$, имеем

$$\langle\, Q\, (\, x^*(t_1) - x^*(t_2)\,)\, .\, x^*(t_1) - x^*(t_2)\, \rangle =$$

$$\langle\, \sum_{i=1}^{n} \{\, [\, t_1(\, x^*(t_1)\,)^{(i)}\,]^{-1} - [\, t_2(\, x^*(t_2)\,)^{(i)}\,]^{-1}\, \} \times$$

$$\times [\, (\, x^*(t_1)\,)^{(i)} - (\, x^*(t_2)\,)^{(i)}\,] = n\, (\, t_1^{-1} + t_2^{-1}\,) -$$

$$- \sum_{i=1}^{n} \{\, (\, x^*(t_2)\,)^{(i)}\, [\, t_1\, (\, x^*(t_1)\,)^{(i)}\,]^{-1} +$$

$$+ (\, x^*(t_1)\,)^{(i)}\, [\, t_2\, (\, x^*(t_2)\,)^{(i)}\,]^{-1}\, \}\, .$$

откуда после несложных преобразований получаем требуемое тождество. □

<u>Следствие 5.7.1.</u> Если $t_1 > 0$, $t_2 = (1 + \alpha)^2$, $\alpha > 0$, то справедливы неравенства

$$(\, 2 + 2\, \alpha + n\, \alpha^2\,)^{-1}\, x^*(t_1) < x^*(\, t_2\,) <$$

$$< (1 + \alpha)^{-2}\, (\, 2 + 2\, \alpha + n\, \alpha^2\,)\;\; x^*(\, t_1\,) \quad ,$$

$$t_2 \langle\, Q\, (\, x^*(t_1) - x^*(t_2)\,),\, x^*(t_1) - x^*(t_2)\, \rangle +$$

$$+ \sum_{i=1}^{n} [\, (\, x^*(t_1)\,)^{(i)} - (\, x^*(t_2)\,)^{(i)}\,]^2 /$$

$$/[\, (\, x^*(t_1)\,)^{(i)}\, (\, x^*(t_2)\,)^{(i)}\,] \;\; < \;\; n\, [\, t_2/t_1 - 1\,]\, .$$

Поставим в соответствие каждому $a > 0$, $q \in (\, 0\, ,\, 1\,)$ и $M > 1$ функцию $\varphi(\, t\,) \equiv \varphi(\, a,\, q,\, M,\, t\,)$, определенную следующим образом:

$$\varphi(t) = \begin{cases} -\ln(Ma) - (Ma)^{-1}(t - Ma) + 0.5(Ma)^{-2}(t - Ma)^2, \\ \quad \text{если } t > Ma, \\ -\ln(t), \quad \text{если } qa < t < Ma, \\ -\ln(qa) - (qa)^{-1}(t - qa) + 0.5(qa)^{-2}(t - qa)^2, \\ \quad \text{если } 0 < t < qa. \end{cases}$$

Нетрудно видеть, что функция $\varphi(t)$ есть гладкое продолжение параболами функции $-\ln t$ за границы отрезка $[qa, Ma]$. Для всякого $x_0 > 0$ сформируем функцию

$F(x_0, q, M, x) = \sum_{i=1}^{n} \varphi(x_0^{(i)}, q, M, x^{(i)})$.

Ясно, что для всех x, $q x_0 < x < M x_0$, справедливо тождество $F(x_0, q, M, x) \equiv F_0(x)$. Кроме того, в силу определения функции $\varphi(t)$ соотношения

$$M^{-2} F_0''(x_0) < F''(x_0, q, M, x) < q^{-2} F_0''(x_0) \quad (5.7.5)$$

справедливы для всех x, таких, что $q x_0 < x < M x_0$.

Теперь мы можем перейти к описанию схем "многошаговых" методов барьеров.

Метод $\mathcal{M}_2(\alpha, \tau)$

0 Полагаем

$q = (1 - \tau)(2 + 2\alpha + n\alpha^2)^{-1}$,

$M = (1 + \tau)(1 + \alpha)^{-2}(2 + 2\alpha + n\alpha^2)$,

$\delta = q M^{-2} \tau (1 + \alpha + \tau)^{-1}$,

$h = 2 q^2 M^2 (q^2 + M^2)^{-1}$,

$x_0 = e$, $t_0 = 0.5(1 + \alpha)^{-2} \tau^2 n^{-3} \| f'(e) \|_\infty^{-1}$.

1 k-я итерация $(k > 0)$

а) Полагаем

$t_{k+1} = (1 + \alpha)^2 t_k$,

$f_{k+1}(x) = t_{k+1} f(x) + F(x_k, q, M, x)$;

б) формируем матрицу P_k – проектор на множество $\mathcal{L} = \{x \mid Ax = 0, \langle e, x \rangle = 0\}$ в метрике, задаваемой матрицей $B_k = t_{k+1} Q + F_0''(x_k)$;

в) производим минимизацию функции $f_{k+1}(x)$ на множестве $x_k + \mathcal{L}$ с помощью метода градиентного спуска:

полагаем $x_{k,0} = x_k$ и продолжаем процесс
$x_{k,i+1} = x_{k,i} - h P_k B_k^{-1} f'_{k+1}(x_{k,i})$, $i = 0, 1,$
до тех пор, пока на некотором шаге с номером $i = l$ не будет выполнено условие
$\langle P_k B_k^{-1} P_k^T f'_{k+1}(x_{k,i}), f'_{k+1}(x_{k,i}) \rangle < \delta^2$

г) полагаем $x_{k+1} = x_{k,i}$.
Итерация закончена.

Теорема 5.7.1. Все точки последовательности $\{ x_k \}_{k=0}^{\infty}$, построенной методом $\mathcal{M}_2(\alpha , \tau)$, $0 < \tau < 1$, являются допустимыми точками задачи (5.7.3). При этом число арифметических операций, необходимых методу $\mathcal{M}_2(\alpha, \tau)^*$, $0 < \alpha < 1$, для получения решения задачи (5.7.3) с точностью ε, не превышает величины порядка
$O([\alpha^{-1} m^2 n + \alpha^7 m n^5 \ln^2 n] \ln (t_0 \varepsilon)^{-1})$.

При оптимальном значении $\alpha = O(n^{-1/2} (m/\ln^2 n)^{1/8})$ оценка трудоемкости метода $\mathcal{M}_2(\alpha, \tau)$ имеет вид
$O([m^2 n^{3/2} (m / \ln^2 n)^{-1/8}] \ln (t_0 \varepsilon)^{-1})$.

Доказательство. 1. Докажем, что при всех $i \geq 0$ справедливо неравенство

$$\langle B_i (x_i - x^*(t_i)), x_i - x^*(t_i) \rangle \leq \tau^2 \tag{5.7.6}$$

Доказательство будем вести по индукции.

Заметим, что в силу теоремы 1.3.5 функция $F_0(x)$ является сильно выпуклой на G с константой сильной выпуклости n^{-2}. Поэтому в силу неравенства (1.3.17) при $i = 0$ имеем

$$\langle B_0 (x_0 - x^*(t_0)), x_0 - x^*(t_0) \rangle =$$
$$= t_1 \langle Q (e - x^*(t_0)), e - x^*(t_0) \rangle + \| e - x^*(t_0) \|^2 \leq$$
$$\leq (1 + \alpha)^2 n^2 \langle \Psi'(t_0, e) - \Psi'(t_0, x^*(t_0)), e -$$
$$- x^*(t_0) \rangle =$$

$= (1 + \alpha)^2 n^2 \langle t_0 f'(e) + F_0'(e), e - x^*(t_0) \rangle =$
$= (1 + \alpha)^2 n^2 t_0 \langle f'(e), e - x^*(t_0) \rangle \le$
$\le 2 (1 + \alpha)^2 n^3 t_0 \| f'(e) \|_\infty \le \tau^2 .$

Таким образом, неравенство (5.7.6) при $i = 0$ выполнено.

Пусть (5.7.6) выполнено и при некотором $i = k > 0$. В этом случае из неравенства (5.7.6) получаем

$$\Sigma_{i=1}^{n} [x_k^{(i)} - (x^*(t_k))^{(i)}]^2 / (x_k^{(i)})^2 \le \tau^2 .$$

Поэтому в силу следствия 5.7.1

$$x^*(t_{k+1}) \le (1 + \alpha)^{-2}(2 + 2 \alpha + n \alpha^2) x^*(t_k) \le M x_k .$$

$$x^*(t_{k+1}) \ge (2 + 2 \alpha + n \alpha^2)^{-1} x^*(t_k) \ge q x_k . \quad (5.7.7)$$

Заметим, что функция $f_{k+1}(x)$ совпадает с функцией $\Psi(t_{k+1}, x)$ на параллелепипеде $\{ x \mid q x_k \le x \le M x_k \}$ Поэтому минимум функции $f_{k+1}(x)$ на множестве $x_k + \mathcal{L}$ достигается в точке $x^*(t_{k+1})$.

Из способа определения функции $f_{k+1}(x)$ и неравенства (5.7.5) следует, что при всех x из R^n справедливо неравенство

$$M^{-2} B_k \le f''_{k+1}(x) \le q^{-2} B_k . \quad (5.7.8)$$

Следовательно, функция $f_{k+1}(x)$ является сильно выпуклой в метрике, задаваемой матрицей B_k. Поэтому вспомогательный процесс минимизации функции $f_{k+1}(x)$ на шаге в) метода $\mathcal{M}_2(\alpha, \tau)$ будет сходиться к точке $x^*(t_{k+1})$. Таким образом, критерий прерывания вспомогательного процесса корректен.

Заметим, что в силу неравенства (5.7.8)

$M^{-2} \langle B_k (x_{k+1} - x^*(t_{k+1})), x_{k+1} - x^*(t_{k+1}) \rangle \le$
$\le \langle f'_{k+1}(x_{k+1}), x_{k+1} - x^*(t_{k+1}) \rangle \le$
$\le \langle P_k B_k^{-1} P_k^T f'_{k+1}(x_{k+1}), f'_{k+1}(x_{k+1}) \rangle^{1/2} \times$
$\times \langle B_k (x_{k+1} - x^*(t_{k+1})), x_{k+1} - x^*(t_{k+1}) \rangle^{1/2} ,$

т.е.

$$\langle B_k (x_{k+1} - x^*(t_{k+1})),\ x_{k+1} - x^*(t_{k+1}) \rangle \le$$
$$\le M^4 \langle P_k B_k^{-1} P_k^T f'_{k+1}(x_{k+1}),\ f'_{k+1}(x_{k+1}) \rangle . \qquad (5.7.9)$$

Следовательно, в силу (5.7.7), (5.7.9) и критерия прерывания вспомогательного процесса

$$\sum_{i=1}^{n} [x_{k+1}^{(i)} - (x^*(t_{k+1}))^{(i)}]^2 / (x_k^{(i)})^2 =$$
$$= \langle F''_0(x_k) (x_{k+1} - x^*(t_{k+1})),\ x_{k+1} - x^*(t_{k+1}) \rangle \le$$
$$\le \langle B_k (x_{k+1} - x^*(t_{k+1})),\ x_{k+1} - x^*(t_{k+1}) \rangle \le$$
$$\le M^4 \langle P_k B_k^{-1} P_k^T f'_{k+1}(x_{k+1}),\ f'_{k+1}(x_{k+1}) \rangle \le$$
$$\le M^4 \gamma^2 .$$

Отсюда и из (5.7.7) получаем, что

$$x_{k+1} \ge x^*(t_{k+1}) - M^2 \gamma x_k \ge (q - M^2 \gamma) x_k$$

Поэтому $B_{k+1} \le (1 + \alpha)^2 (q - M^2 \gamma)^{-2} B_k$ и, следовательно,

$$\langle B_{k+1} (x_{k+1} - x^*(t_{k+1})),\ x_{k+1} - x^*(t_{k+1}) \rangle \le$$
$$\le (1 + \alpha)^2 (q - M^2 \gamma)^{-2} \langle B_k (x_{k+1} - x^*(t_{k+1})),$$
$$x_{k+1} - x^*(t_{k+1}) \rangle \le (1 + \alpha)^2 (q - M^2 \gamma)^{-2} M^4 \gamma^2 = \tau^2 .$$

Таким образом, неравенство (5.7.6) выполнено при $i = k + 1$, и, значит, оно справедливо при всех $i \ge 0$. В частности, из этого неравенства вытекает допустимость всех точек x_k

2. Оценим число шагов вспомогательного процесса. Из неравенства (5.7.8) и оценки скорости сходимости метода градиентного спуска (см. §4.4) получаем

$$q^2 \langle P_k B_k^{-1} P_k^T f'_{k+1}(x_{k,i}),\ f'_{k+1}(x_{k,i}) \rangle \le$$
$$\le f_{k+1}(x_{k,i}) - f_{k+1}(x^*(t_{k+1})) \le$$
$$\le [1 - q^2 M^{-2}]^i R_k^2 ,$$

где $R_k^2 = \langle B_k (x_k - x^*(t_{k+1})),\ x_k - x^*(t_{k+1}) \rangle$. Заметим, что в силу неравенств (5.7.6), (5.7.7)

$$x^*(t_{k+1}) \le M x_k \quad , \quad x^*(t_k) \le (1 + \tau) x_k .$$

Поэтому в силу следствия 5.7.1

$$\langle B_k (x_k - x^*(t_{k+1})),\ x_k - x^*(t_{k+1}) \rangle^{1/2} \le$$
$$\le \tau + \langle B_k(x^*(t_k) - x^*(t_{k+1})),\ x^*(t_k) - x^*(t_{k+1}) \rangle^{1/2} ;$$

$< B_k(x^*(t_k) - x^*(t_{k+1})),\ x^*(t_k) - x^*(t_{k+1}) > =$
$= t_{k+1}< Q(x^*(t_{k+1}) - x^*(t_k)),\ x^*(t_{k+1}) - x^*(t_k) > +$
$+ \sum_{i=1}^{n} [(x^*(t_{k+1}))^{(i)} - (x^*(t_k))^{(i)}]^2 / (x_k^{(i)})^2 \leq$
$\leq t_{k+1}< Q(x^*(t_{k+1}) - x^*(t_k)),\ x^*(t_{k+1}) - x^*(t_k) > +$
$+ (1 + \tau) M \sum_{i=1}^{n} [(x^*(t_{k+1}))^{(i)} - (x^*(t_k))^{(i)}]^2 /$
$/[(x^*(t_{k+1}))^{(i)} (x^*(t_k))^{(i)}] \leq n (1 + \tau) \alpha(2 + \alpha) M.$

Таким образом,

$R_k \leq \tau + [n (1 + \tau) \alpha (2 + \alpha) M]^{1/2} = O(n)$

Следовательно, число вспомогательных шагов $l = l(k)$ оценивается следующим образом:

$l(k) \leq \ln 1/\gamma\ [2 \ln R_k - 2 \ln q] /\ln (1 - q^2 M^{-2}) \leq$
$\leq \text{const}\ M^2 q^{-2} \ln^2 n = O(n^4 \alpha^8 \ln^2 n).$

Осталось учесть, что трудоемкость выполнения одного вспомогательного шага имеет порядок $O(m n)$, трудоемкость подготовительных операций на каждой итерации - $O(m^2 n)$, а общее число итераций метода не превышает величины порядка $O(\ln (t_0 \varepsilon)^{-1} / \ln (1 + \alpha)) = O(\alpha^{-1} \ln (t_0 \varepsilon)^{-1})$. □

Таким образом, применение в методе барьеров для возврата на траекторию градиентного метода позволило уменьшить трудоемкость исходного метода в $O((m / \ln^2 n)^{1/8})$ раз. Более существенный выигрыш получается, если в методе барьеров использовать оптимальный метод минимизации сильно выпуклых функций из §4.1. Приведем схему такого метода.

Метод $\mathcal{M}_3(\alpha, \tau)$

0. Полагаем

$q = (1 - \tau) (2 + 2 \alpha + n \alpha^2)^{-1}$,
$M = (1 + \tau) (1 + \alpha)^{-2} (2 + 2 \alpha + n \alpha^2)$,
$\gamma = q M^{-2} \tau (1 + \alpha + \tau)^{-1}$,
$x_0 = e$, $t_0 = 0.5 (1 + \alpha)^{-2} \tau^2 n^{-3} \| f'(e) \|_\infty^{-1}$.

1. k - я итерация ($k > 0$)

а) Полагаем

$t_{k+1} = (1 + \alpha)^2 t_k$,

$f_{k+1}(x) = t_{k+1} f(x) + F(x_k, q, M, x)$;

б) формируем матрицу P_k - проектор на множество $\mathcal{L} = \{ x \mid Ax = 0, \langle e, x \rangle = 0 \}$ в метрике, задаваемой матрицей $B_k = t_{k+1} Q + F''_0(x_k)$;

в) производим минимизацию функции $f_{k+1}(x)$ на множестве $x_k + \mathcal{L}$ с помощью оптимального метода минимизации сильно выпуклых функций:

полагаем

$A_{k,0} = M^2$, $x_{k,0} = x_k$, $v_{k,0} = x_k$;

на i -м шаге вспомогательного процесса:

вычисляем $\alpha_{k,i}$ из уравнения

$M^2 \alpha^2_{k,i} = (1 - \alpha) A_{k,i}$.

полагаем

$y_{k,i} = \alpha_{k,i} v_{k,i} + (1 - \alpha_{k,i}) x_{k,i}$,

$A_{k,i+1} = \alpha_{k,i} q^2 + (1 - \alpha_{k,i}) A_{k,i}$,

$x_{k,i+1} = y_{k,i} - M^{-2} P_k B_k^{-1} f'_{k+1}(y_{k,i})$,

$v_{k,i+1} = (1 - \alpha_{k,i}) A_{k,i} A^{-1}_{k,i+1} v_{k,i} +$

$+ \alpha_{k,i} q^2 A^{-1}_{k,i+1} y_{k,i} - \alpha_{k,i} A^{-1}_{k,i+1} P_k B_k^{-1} f'_{k+1}(y_{k,i})$;

продолжаем вспомогательный процесс до тех пор, пока на некотором шаге с номером $i = l$ не будет выполнено условие $\langle P_k B_k^{-1} P_k^T f'_{k+1}(x_{k,l}), f'_{k+1}(x_{k,l}) \rangle < \delta^2$;

г) полагаем $x_{k+1} = x_{k,l}$.

Итерация закончена.

Теорема 5.7.2. Все точки последовательности $\{ x_k \}_{k=0}^{\infty}$, построенной методом $\mathcal{M}_3(\alpha, \tau)$, $0 < \tau < 1$, являются допустимыми точками задачи (5.7.3). При этом число арифметических операций, необходимых методу $\mathcal{M}_3(\alpha, \tau)$, $0 <$

$< \alpha < 1$, для получения решения задачи (5.7.3) с точностью ε , не превышает величины порядка

$O([\alpha^{-1} m^2 n + \alpha^3 m n^3 \ln^2 n] \ln (t_0 \varepsilon)^{-1})$.

При оптимальном значении $\alpha = O(n^{-1/2} (m/\ln^2 n)^{1/4}))$ оценка трудоемкости метода $\mathcal{M}_3(\alpha, \tau)$ имеет вид

$O([m^2 n^{3/2} (m / \ln^2 n)^{-1/4}] \ln (t_0 \varepsilon)^{-1})$

Мы не будем приводить доказательство этой теоремы, так как оно почти дословно повторяет доказательство теоремы 5.7.1, отличаясь от него лишь тем, что вместо оценки скорости сходимости градиентного метода в нем используется оценка скорости сходимости оптимального метода.

Таким образом, применение в методе барьеров для возврата на траекторию оптимального метода минимизации сильно выпуклых функций позволяет уменьшить трудоемкость исходного метода в $O((m / \ln^2 n)^{1/4})$ раз. Отметим, что модифицированная стратегия пересчета приближенных проекторов, описанная в §5.6, дает более существенный выигрыш в трудоемкости. Однако в многошаговых методах появляется возможность управлять скоростью роста штрафного параметра, соизмеряя ее с реальной трудоемкость минимизации вспомогательных функций. Учитывая, что оптимальные методы из §4.1 на практике обычно сходятся существенно быстрее, чем им предписывает теоретическая оценка скорости сходимости, можно ожидать, что реальная эффективность метода $\mathcal{M}_3$ при разумной стратегии выбора параметров будет выше, чем у метода $\mathcal{M}_1$.

ПРИЛОЖЕНИЕ. НЕКОТОРЫЕ РЕЗУЛЬТАТЫ ИЗ ТЕОРИИ СЛОЖНОСТИ ЭКСТРЕМАЛЬНЫХ ЗАДАЧ

П.1. Сложность экстремальных задач: постановка вопроса

Содержательно сложность некоторого класса экстремальных задач - это нижняя граница трудоемкости всевозможных методов, гарантирующих заданную погрешность решения любой задачи класса. Дадим необходимые формальные определения (ограничиваясь степенью общности, диктуемой нуждами основного текста; более общие постановки и результаты можно найти в [19]).

Пусть $\mathfrak{F} = \{ f \}$ - некоторое семейство выпуклых функций на R^n; каждую функцию семейства отождествим с задачей безусловной минимизации ("задачей f")

$$f(x) \to \min \mid x \in R^n. \qquad (\text{П.1})$$

Будем рассматривать лишь такие семейства, что все задачи $f \in \mathfrak{F}$ разрешимы; подобные $\mathfrak{F}$ назовем классами задач.

Для $x \in R^n$ и $f \in \mathfrak{F}$ определим погрешность точки x в качестве приближенного решения задачи f как

$$\varepsilon(x, f) = f(x) - \min \{ f(y) \mid y \in R^n \}$$

("погрешность по функционалу").

Теперь определим понятие "метод первого порядка решения задач класса $\mathfrak{F}$". Пусть, решая задачу $f \in \mathfrak{F}$, мы можем вычислять (в любой желаемой точке $x \in R^n$) значение $f(x)$ и субградиент $\partial f(x)$ минимизируемой функции; другими источниками информации о ней мы не располагаем. Как в этих предположениях может выглядеть самая общая процедура минимизации? Очевидно, следующим образом. Процедура работает по шагам; на первом из них вычисляются $f(\,)$, $\partial f(\,)$ в некоторой точке x_1, на втором - в точке x_2, и т.д; i-я точка может при этом зависеть от накопленной к соответствующему шагу информации

$f(x_1), \ldots, f(x_{i-1}), \partial f(x_1), \ldots, \partial f(x_{i-1})$.

В некоторый момент процедура должна остановиться и сформировать результат своей работы на f. Наша стратегия поведения - метод решения задач из $\mathfrak{F}$ - есть, таким образом, не что иное, как набор правил формирования очередных точек x_i, момента остановки и результата в зависимости от накопленной к соответствующим шагам информации. Подчеркнем, что в таком определении понятия "метод" на составляющие метод правила не налагается никаких ограничений сверх "физической реализуемости" - правила, применяемые на i-м шаге, имеют в качестве аргументов

$f(x_1), \ldots, f(x_{i-1}), \partial f(x_1), \ldots, \partial f(x_{i-1})$.

Пусть $\mathfrak{M}$ - метод первого порядка решения задач из $\mathfrak{F}$. Если $f \in \mathfrak{F}$, то $\mathfrak{M}$, решая f, либо останавливается после некоторого числа шагов $l(\mathfrak{M}, f)$, и в этом случае определен результат $x(\mathfrak{M}, f)$ применения $\mathfrak{M}$ к f, либо не останавливается; в последнем случае положим $l(\mathfrak{M}, f) = +\infty$. Величины

$l(\mathfrak{M}, \mathfrak{F}) = \sup\{ l(\mathfrak{M}, f) \mid f \in \mathfrak{F} \}$,

$\varepsilon(\mathfrak{M}, \mathfrak{F}) = \sup\{ \varepsilon(x(\mathfrak{M}, f)) \mid f \in \mathfrak{F},$
$l(\mathfrak{M}, f) < \infty \}$

назовем соответственно трудоемкостью и погрешностью метода $\mathfrak{M}$ на классе $\mathfrak{F}$. Чтобы избежать тривиальных оговорок, будем далее рассматривать лишь методы трудоемкости, не меньшей 1.

Наконец, сложностью класса $\mathfrak{F}$ назовем функцию

$N(\varepsilon) = \min\{ l(\mathfrak{M}, \mathfrak{F}) \mid \mathfrak{M}$ - метод первого порядка решения задач из $\mathfrak{F}$, $\varepsilon(\mathfrak{M}, \mathfrak{F}) < \varepsilon \}$.

(это - минимальная из трудоемкостей методов, решающих все задачи класса с погрешностью, не большей чем ε).

Ниже будут приведены явные двусторонние оценки сложности наиболее часто рассматриваемых в теории выпуклой оптимизации классов задач. Сопоставление этих оценок с оценками трудоемкости тех или иных процедур, решающих соответствующие задачи с заданной погрешностью, позволяет судить о возможностях "потенциального улучшения" процедур, и в частности оправдать сделанные в основном тексте утверждения об оптимальности методов.

Прежде чем переходить к конкретным оценкам, отметим следующее. В литературе часто рассматриваются "идеализированные" численные методы, для которых считается доступным (и притом "бесплатным") решение некоторых вспомогательных задач (обычно- задач одномерной минимизации). Не вдаваясь в обсуждение содержательной правомерности такого рода идеализаций, модифицируем определение метода и сложности с тем, чтобы охватить и этот их тип Именно, фиксируем натуральное π и разрешим методу на i-м шаге вычислять f и ∂f не в точке x_i , а во всех точках (π - 1)-мерного аффинного многообразия $X_i \subseteq R^n$. После очевидных модификаций данных выше описаний мы приходим к определению "методов первого порядка решения задач $\mathcal{F}$ с (π - 1)-мерным поиском" (коротко ($\pi, \mathcal{F}$) методов), их трудоемкости и погрешности на задаче и на классе и сложности $N_\pi(\varepsilon)$ класса $\mathcal{F}$ относительно таких методов. При этом "метод" и "сложность" в первоначальном определении отвечают $\pi = 1$.

В последующем изложении $c_i > 0$ - абсолютные постоянные.

П.2. Сложность задач выпуклой минимизации большой размерности

Фиксируем $\tau \in [0, 1]$, $R, L > 0$ и рассмотрим класс $\mathcal{F}$ всех выпуклых непрерывно-дифференцируемых функций f на n-мерном евлидовом пространстве R^n, таких, что

(i) f достигает минимума в шаре радиуса R с центром в 0;

(ii) ∇f гельдерово с показателем τ и константой L:

$\| \nabla f(x) - \nabla f(y) \| \leq L \| x - y \|^{\tau}$

($\forall\, x, y \in R^n$; $0^0 = 1$ по определению).

<u>Теорема П.1.</u> Для сложности описанного класса задач при любом $\pi \in N$ справедливы неравенства

$$N_{\pi}(\varepsilon) \geq \min \{ (n-1)(\pi+1)^{-1}, \quad c (L R^{1+\tau} \varepsilon^{-1})^{\sigma} \}, \tag{П.2}$$

$$N(\varepsilon) \leq C (L R^{1+\tau} \varepsilon^{-1})^{\sigma}. \tag{П.3}$$

Здесь $\sigma = (0.5 + 1.5\tau)^{-1}$; C, c - положительные абсолютные постоянные.

Доказательство Оценка (П.3) следует из результатов §4.3.

Займемся доказательством (П.2). Подходящим масштабированием (общим для всех задач класса подобным преобразованием аргумента и умножением минимизируемых функций на постоянную) можно свести дело к случаю $L = 1$, $R = 1$, что мы и предположим.

Фиксируем $\varepsilon > 0$, и пусть $N_{\pi}(\varepsilon) = N$, причем

$$N(\pi + 1) + 1 < n; \tag{П.4}$$

нам достаточно доказать, что в этом случае

$$N \geq c\, \varepsilon^{-\sigma} \tag{П.5}$$

при подходящей положительной абсолютной константе c

1^0. Очевидно, при некотором c_1 и любых τ, n можно указать выпуклую C^∞-функцию ξ на R^n, удовлетворяющую (ii) с $L = 1/2$ и такую, что $\xi(x) = 0$ при $\|x\| < 1/2$ и $\xi(x) > c_1$ при $\|x\| = 1$

2^0. Пусть φ - выпуклая функция на R^n, липшицева с константой 1, и пусть $T\varphi$ - функция, чей надграфик получается из надграфика φ сложением с шаром радиуса 1 (в R^{n+1}) Следующий результат хорошо известен.

Лемма П.1. Пусть φ и ψ - липшицевы с константой 1 функции на R^n. Тогда

1) $T\varphi$ - липшицева с константой 1 и выпукла;

2) если $\varphi \geq \psi$, то $T\varphi \geq T\psi$;

3) если φ и ψ совпадают в ρ-окрестности некоторой точки x, $\rho > 1$, то $T\varphi$ и $T\psi$ также совпадают в некоторой окрестности x;

4) $\varphi - 1 \geq T\varphi \geq \varphi - 2$;

5) для любых $x, y \in R^n$

$$\| \nabla(T\varphi)(x) - \nabla(T\varphi)(y) \| \leq c_0 \| x - y \|.$$

3^0 Пусть $\mathfrak{M} - (\pi, \mathfrak{F})$ - метод, решающий все задачи класса с погрешностью, не большей, чем ε, при трудоемкости, не большей N. Не ограничивая общности, можем считать, что число шагов $\mathfrak{M}$ на любой задаче из $\mathfrak{F}$ есть $N+1$, а результат работы лежит в $(N+1)$-м поисковом многообразии X_{N+1}, имеющем размерность 1 и содержащем 0. Нам будет удобно считать $\mathfrak{M}$ применимым к любым непрерывно-дифференцируемым выпуклым функциям (трудоемкость и погрешность метода мы, разумеется, будем оценивать лишь на функциях из $\mathfrak{F}$).

Лемма П.2. Пусть ρ, λ положительны, l натурально и не превосходит $N+1$. Тогда найдется функция f_l вида

$$f_l(x) = \xi(x) + \rho\, g_l(\lambda x), \qquad (\text{П}.6_l)$$

где $g_l = T\varphi_l$, а

$$\varphi_l(x) = \max\{ e_i^T x - 3i + 3 \mid 1 \le i \le l \}, \qquad (\text{П}.7_l)$$

обладающая следующими свойствами:

$a(l)$: векторы $\{ e_i \}_{i=1}^{l}$ образуют ортонормированную систему;

$b(l)$: если X_i - i-е поисковое многообразие $\mathfrak{M}$ на f_l, то при $1 \le i \le j \le l$ вектор e_j ортогонален X_i.

Д о к а з а т е л ь с т в о. Доказательство проведем индукцией по l.

База $l = 1$. Пусть X_1 - первое поисковое многообразие $\mathfrak{M}$ (оно ни от чего не зависит), X_1^+ - линейная оболочка X_1. Тогда $\dim X_1^+ \le \pi < n$ (последнее - по (П.4)). Поэтому найдется единичный вектор e_1, ортогональный X_1^+. Определим функцию φ_1 по (П.7_1), затем f_1 - по (П.6_1); очевидно, $a(1)$ и $b(1)$ при этом выполнятся.

Шаг индукции $l \Rightarrow l + 1$. Пусть

$e_1, \dots, e_l, X_1, \dots, X_l, f_l, g_l, \varphi_l$

есть объекты, отвечающие l в силу индуктивного предположения. Определим X_{l+1} как $(l + 1)$-е поисковое многообразие $\mathfrak{M}$ на задаче f_l. Рассмотрим линейную оболочку E множеств $X_1, \dots, X_{l+1}$ и векторов $e_1, \dots, e_l$; ее размерность k не превосходит $\pi N + N + 1$, ибо $l \le N$, а размерности линейных оболочек многообразий X_i не превосходят π при $i \le N$ и 1 при $i = N + 1$. По (П.4) $k < n$; поэтому найдется единичный вектор e_{l+1}, ортогональный E. Определим φ_{l+1}, g_{l+1} и f_{l+1} по (П.7_{l+1}) и (П.6_{l+1}) соответственно построенным $e_1, \dots, e_{l+1}$, и проверим, что эти объекты - требуемые утверждением леммы для $l + 1$. Достаточно убедиться, что $X_1, \dots, X_{l+1}$ - первые $l + 1$ поисковые многообразия $\mathfrak{M}$ при решении f_{l+1}. Для этого достаточно убедиться, что если x - точка из $X = \cup_{i \le l} X_i$, то f_l и f_{l+1} совпадают друг с другом в некоторой окрестности x. В самом деле, $X_1, \dots,$

X_{l+1} по построению есть первые $l + 1$ поисковые многообразия метода $\mathfrak{M}$ на f_l; поскольку вдоль первых l из них (при совпадении f_l с f_{l+1} в окрестности X) f_l и f_{l+1} неразличимы по получаемой методом информации, то они же суть поисковые многообразия $\mathfrak{M}$ и на f_{l+1}.

Итак, пусть $x \in X$; положим $z = \lambda x$. Очевидно, для любого y

$\varphi_l(y) \le \varphi_{l+1}(y) = \max\{\varphi_l(y), \psi(y)\}$,

где $\psi(y) = e_{l+1}^T y - 3l$, тогда как $\varphi_l(z) \ge -3l + 3$ в силу (7_l) (мы учли, что e_l ортогонально $X_1, \ldots, X_l$ и, следовательно, x, а тогда и z). Но и e_{l+1} ортогонально x (и z); поэтому $\psi(z) \le -3l \le \varphi_l(z) - 3$. И функция φ_l, и функция ψ липшицевы с константой 1; поэтому неравенство $\varphi_l > \psi$ выполнено в 3/2-окрестности точки z; в той же окрестности, стало быть, выполнено равенство $\varphi_l = \varphi_{l+1}$. По лемме П.1 последнее означает, что $g_l \equiv T\varphi_l = T\varphi_{l+1} \equiv g_{l+1}$ в некоторой окрестности z, так что $f_l = f_{l+1}$ в некоторой окрестности x (ввиду связи x и z) □

4°. Продолжим доказательство теоремы П.1. Чтобы доказать (П.5), положим

$$\rho = \min\{c_1/(15N + 6), 0.5\,(4(N + 1)^{1/2}(3N + 1))^{-(1+\tau)} \times c_0^{-\tau}\}, \quad \lambda = 4(N + 1)^{1/2}(3N + 1). \qquad (П.8)$$

По лемме П.2, примененной к $l = N + 1$, найдется функция $f \equiv f_{N+1}$, для траектории $\mathfrak{M}$ на которой выполнены a($N + 1$), b($N + 1$). В частности, коль скоро результат x применения $\mathfrak{M}$ к f лежит в X_{N+1} (ввиду наших соглашений о $\mathfrak{M}$), мы имеем

$$f(x) \ge \rho\, g_{N+1}(\lambda x) \ge -\rho(3N + 2) \qquad (П.9)$$

(ибо e_{N+1} ортогонально x и ввиду п.4) леммы П.1).

С другой стороны, пусть

$$z = -0.5\,(N + 1)^{-1/2} \sum_{i=1}^{N+1} e_i .$$

Из (П.6_{N+1}), (П.7_{N+1}) ясно, что

$$f(z) < -\rho\,(0.5\,\lambda\,(N+1)^{-1/2}+1)$$

(мы учли (П.7_{N+1}) и п.4) леммы П.1, а также равенство ξ нулю в 1/2 -окрестности 0, очевидно, содержащей z). Из определения ρ и λ отсюда следует

$$f(x) - f(z) > (3N+1)\,\rho,$$

и после несложных дальнейших вычислений

$$\varepsilon(\mathfrak{M}, f) > c_2\,N^{-1/6} \qquad (П.10)$$

Проверим, что $f \in \mathfrak{F}$: это завершит доказательство, ибо левая часть (П.10) тогда обязана быть не меньше ε (происхождение $\mathfrak{M}$), и (П.5) окажется немедленным следствием (П.10).

Чтобы проверить, что f удовлетворяет (ii) с $L = 1$, достаточно убедиться, что $h(x) = \rho\,g_{N+1}(\lambda x)$ удовлетворяет тому же соотношению с $L = 0.5$ (определение ξ). Функция ∇g ограничена по норме константой 1 и липшицева с константой c_0 (лемма П.1); поэтому она гельдерова с показателем τ и константой c_0^{τ}. Отсюда немедленно извлекается, что градиент функции h гельдеров с показателем τ и константой $c_0^{\tau}\,\rho\,\lambda^{1+\tau}$; эта последняя величина не превосходит 0.5 по выбору λ и ρ.

Для доказательства включения $f \in \mathfrak{F}$ осталось проверить, что минимум функции f достигается в единичном шаре. Для этого, в свою очередь, достаточно убедиться, что $f(x) > f(0)$ при $\|x\| = 1$. При таком x имеем ввиду леммы П.1 и очевидных неравенств

$$h(x) > -\rho\,(3N + \max\{\lambda e_i^T x \mid i < N+1\} + 2) >$$
$$> -\rho\,(3N + \lambda\,(N+1)^{-1/2} + 2) > -\rho\,(15N+6);$$

последняя величина (по определению ρ) не меньше $-c_1$. Тем самым $f(x) > 0$. В то же время $f(0) < -1$ по лемме П.1. □

П.3. Сложность задач минимизации сильно выпуклых функций большой размерности

Фиксируем $Q > 2$, L, $F > 0$ и обозначим через $\mathcal{F}$ класс всех непрерывно-дифференцируемых функций f на n-мерном евклидовом пространстве R^n, таких, что

(iii) f сильно выпукла с модулем сильной выпуклости Q, т.е. для любых $x, y \in R^n$

$$Q^{-1} L \| x - y \|^2 \le \langle \nabla f(x) - \nabla f(y), x - y \rangle \le$$
$$\le L \| x - y \|^2 ;$$

(iv) $f(0) - \min \{ f(x) \mid x \in R^n \} \le F$

Если f дважды непрерывно дифференцируема, то (iii) означает, что собственные числа гессиана f лежат в отрезке $[L Q^{-1}, L]$.

<u>Теорема П.2.</u> Для сложности описанного класса задач при любых $\pi \in N$ и $\varepsilon \in (0, F)$ справедливы неравенства

$$N_{\pi}(\varepsilon) \ge \min \{ (n-1)/(\pi + 1), 0.4\, Q^{1/2} \ln F/\varepsilon \} \qquad (П.11)$$

и

$$N_{\pi}(\varepsilon) \le C\, Q^{1/2} \ln 2F/\varepsilon \}. \qquad (П.12)$$

где C - абсолютная константа.

Доказательство. Оценка (П.12) следует из результатов §4.1.

Докажем нижнюю оценку (П.11). Подходящим масштабированием дело сводится к случаю

$L = 1$, $F = 1$,

что мы и предположим.

Фиксируем $\varepsilon \in (0, 1)$, и пусть $N_\pi(\varepsilon) = N$, причем

$$(\pi + 1) N + 2 < n; \tag{П.13}$$

нам достаточно показать, что в этой ситуации

$$N_\pi(\varepsilon) > 0.4\, Q^{1/2} \ln F/\varepsilon. \tag{П.14}$$

1^0. Обозначим через $\mathcal{N}$ множесто всех пар (α, β), у которых $\alpha = \mathrm{diag}\{\alpha_i\}_{i=1}^{N+1}$ - диагональная $(N+1)\times(N+1)$ -матрица с диагональными элементами из $[Q^{-1}, 1]$, а β - $(N + 1)$ -мерный вектор, такой, что

$$\beta^T \alpha^{-1} \beta < 2.$$

В условиях (П.13) имеем $(N + 1) < n$, так что определены $n\times n$ -матрица α^* с такими же первыми $(N + 1)$ диагональными элементами, что и у α, и остальными диагональными элементами, равными 1, и n-мерный вектор β^*, у которого первые $(N + 1)$ координат те же, что и у β, а остальные равны 0.

Пусть $\mathcal{U}$ - множество всех ортогональных $n\times n$ -матриц, а $\mathcal{K}(\alpha, \beta)$ - семейство всех квадратичных форм на R^n вида

$$f_{U,\alpha,\beta}(x) = 0.5\, x^T U^T \alpha^* U x - x^T U^T \beta^*$$

Очевидно,

$$(\alpha, \beta) \in \mathcal{N} \Rightarrow \mathcal{K}(\alpha, \beta) \subseteq \mathcal{F} \tag{П.15}$$

и

$$U, V \in \mathcal{U} \Rightarrow f_{UV,\alpha,\beta}(x) \equiv f_{U,\alpha,\beta}(Vx). \tag{П.16}$$

Кроме того, если

$$f(x) = 0.5\, x^T A x - b^T x \in \mathcal{K}(\alpha, \beta),\ A = A^T,$$

то можно указать

$$m = n - (N + 1) > \pi N + 1 \tag{П.17}$$

-мерное подпространство $E^*(f) \subseteq R^n$, такое, что A тождественно на $E^*(f)$ и b ортогонально $E^*(f)$; в частности, подпространства Крылова формы f

$E_0(f) = \{0\}$,

$E_i(f) = \mathrm{Lin}\{b, Ab, \ldots, A^{i-1}b\}$, $i > 0$,

ортогональны $E^+(f)$.

Ясно также, что при $V, U \in \mathcal{U}$ и всех i имеем

$E_i(f_{UV,\alpha,\beta}) = V^T E_i(f_{U,\alpha,\beta})$,

$$E^+(f_{UV,\alpha,\beta}) = V^T E^+(f_{U,\alpha,\beta}) \qquad (\text{П.18})$$

2^0. Обозначим через $\mathfrak{M}$ (π, $\mathfrak{F}$) -метод трудоемкости N, решающий все задачи из $\mathfrak{F}$ с погрешностью, не превосходящей ε; можно считать, что число шагов метода $\mathfrak{M}$ на любой задаче из $\mathfrak{F}$ есть $N + 1$, а результат работы лежит в $(N + 1)$ -м поисковом многообразии X_{N+1}, являющемся одномерным подпространством R^n.

Фиксируем $(\alpha, \beta) \in \mathcal{N}$.

Лемма П.3. Каково бы ни было натуральное $l \leq N + 1$, найдутся $U(l) \in \mathcal{U}$ и не более чем πl -мерное подпространство $E^l \subseteq E^+(f_{U(l),\alpha,\beta})$, такие, что для первых l поисковых многообразий $\mathfrak{M}$ на $f^l \equiv f_{U(l),\alpha,\beta}$ имеем

$$X_i \subseteq E_i(f_{U(l),\alpha,\beta}) + E^l, \quad 1 \leq i \leq l. \qquad (\text{П.}19_l)$$

Доказательство проведем индукцией по l.

База $l = 1$ Пусть X_1 - первое поисковое многообразие метода $\mathfrak{M}$. Размерность его линейной оболочки не превышает $\pi \leq n$ ((П.13)), так что можно найти такое $U(1) \in \mathcal{U}$, что $U^T(1)\, E^+(f_{I,\alpha,\beta})$ содержит не более чем π -мерное подпространство E^1, содержащее X_1; здесь I - единичная матрица. По (П.18)

$U^T(1)\, E^+(f_{I,\alpha,\beta}) = E^+(f_{U(1),\alpha,\beta})$,

и (19_1) выполнено.

Шаг индукции $l \Rightarrow l + 1$. Пусть $U(l)$, f^l, E^l, $X_1, \ldots, X_l$ - объекты, отвечающие l в силу предположения индукции. Определим X_{l+1} как $(l + 1)$ -е поисковое многообразие метода $\mathfrak{M}$ на f^l, и пусть X^+_{l+1} - линейная оболочка этого многообразия, а H - ортогональное дополнение к $E_l(f^l)$. Пусть Y - ортогональная проекция X^+_{l+1} на H. В силу наших соглашений о $\mathfrak{M}$ имеем $\dim Y \leq \pi$ при $l < N$ и $\dim Y \leq 1$ при $l = N$, во всех случаях

$$\dim \{ E^l + Y \} \leq \min \{ \pi (l + 1), \dim E^+(f^l) \} \equiv k$$

Как мы уже отмечали, $E_l(f^l)$ ортогонально $E^+(f_l)$, так что $E^+(f^l) \subseteq H$; тем самым $E^l + Y \subseteq H$. Поэтому найдется k -мерное подпространство $E^{l+1} \subseteq H$, содержащее $E^l + Y$. Пусть V^+ - ортогональный оператор в H, тождественный на E^l и такой, что $E^{l+1} \subseteq (V^+)^T E^+(f^l)$. Продолжим V^+ с H на все R^n тождественным на $E_l(f^l)$ образом; получим ортогональный оператор V на R^n, такой, что

V тождественно на $E_l(f^l) + E^l$; (П.20$_a$)

$V^T E^+(f^l)$ содержит E^{l+1}; (П.20$_b$)

кроме того, по построению

$$X_{l+1} \subseteq E_l(f^l) + E^{l+1},$$

$$E^l \subseteq E^{l+1},$$

$$\dim E^{l+1} \leq (l + 1) \pi \qquad \text{(П.20}_c\text{)}$$

Положим

$$U(l+1) = U(l) V,$$

$$f^{l+1}(x) = f_{U(l+1), \alpha, \beta}(x) = f_{U(l), \alpha, \beta}(V x)$$

Убедимся, что объекты f^{l+1}, $U(l+1)$, E^{l+1} - искомые. Имея в виду (П.19$_l$) и (П.20), достаточно установить, что $X_1, \ldots, X_{l+1}$ - первые $(l + 1)$ поисковое многообразие $\mathfrak{M}$ на f^{l+1}. В свою очередь, для этого достаточно проверить, что

$$x \in E^l + E_{l-1}(f^l) \Rightarrow f^l(x) = f^{l+1}(x),$$

$$\nabla f^l(x) = \nabla f^{l+1}(x) \qquad \text{(П.21)}$$

Действительно, $X_1, \ldots, X_{l+1}$ по построению есть первые $l + 1$ поисковые многообразия метода $\mathfrak{M}$ на f^l и для всех точек у первых l из них выполнена посылка импликации (П.21) (ввиду (П.19$_l$)). Если эта импликация истинна, то сказанное означает неразличимость для метода $\mathfrak{M}$ задач f^l и f^{l+1} по информации, полученной методом на первых l шагах работы, так что первые $l+1$ поисковые многообразия метода $\mathfrak{M}$ на f^{l+1} соответственно те же, что и на f^l.

Проверка (П.21) выглядит следующим образом. Пусть x удовлетворяет посылке этого соотношения; тогда

$f^{l+1}(x) = f^l(Vx)$ и $\nabla f^{l+1}(x) = V^T \nabla f^l(Vx)$.

Оператор V тождественен на $E_{l-1}(f^l) + E^l$; это дает первое из равенств в (П.21) и, кроме того, включение

$\nabla f^l(Vx) = \nabla f^l(x) \in E^l + E_l(f^l)$

(ибо отображение $y \to 0.5 \nabla f^l(y)$ тождественно на E^l и переводит $E_{l-1}(f^l)$ в $E_l(f^l)$). Упомянутое включение ввиду тождественности V (и, значит, V^T) на $E^l + E_l(f^l)$ дает второе равенство в (П.21). □

3^0 Продолжим доказательство теоремы. Для его завершения остается заметить, что по лемме П.3, примененной с $l = N+1$, результат применения метода $\mathfrak{M}$ к некоторой задаче $f \in \mathcal{K}(\alpha, \beta)$ лежит в $E^+(f) + E_N(f)$, так что в силу происхождения $\mathfrak{M}$

$\varepsilon > \varepsilon(\mathfrak{M}, f) > \min\{f(x) \mid x \in E^+(f) + E_N(f)\} -$
$- \min\{f(x) \mid x \in R^n\} =$
$= \min\{0.5\,(p(\alpha)\beta)^T \alpha\,(p(\alpha)\beta) \mid$
$p \in \mathcal{P}(N-1)\}$,

где $\mathcal{P}(N-1)$ - пространство многочленов на оси степени, не большей, чем $N-1$ (последнее равенство очевидно). Иными словами, мы показали, что если $N = N_\pi(\varepsilon)$ удовлетворяет (П.13), то

$\varepsilon > 0.5 \max\{ \min\{ \langle (I - \alpha p(\alpha))^2 \gamma_{\alpha,\beta},$

$\gamma_{\alpha,\beta} > | p \in \mathcal{P}(N-1) \} \mid (\alpha, \beta) \in \mathcal{N} \}$.
где $\gamma_{\alpha,\beta} = \alpha^{-1/2} \beta$, так что $\gamma^{T}_{\alpha,\beta} \gamma_{\alpha,\beta} < 2$

Последнее соотношение можно переписать в виде

$$\varepsilon \geq \max \{ \min \{ \int_{\Delta} (1 - t\,p(t))^2 \, d\nu(t) \mid p \in \mathcal{P}(N-1) \} \mid \nu \in \Phi \}, \tag{П.22}$$

где Φ - пространство вероятностных мер на $\Delta = [Q^{-1}, 1]$ с не более чем $(N+1)$-точечным носителем.

Пусть теперь Ψ - пространство всех вероятностных борелевских мер на Δ Выпукло-вогнутая функция

$$G(\nu, p) = \int_{\Delta} (1 - t\,p(t))^2 \, d\nu(t)$$

имеет на $\Psi \times \mathcal{P}(N-1)$ седловую точку (ν^*, p^*); ясно, что

$$G(\nu^*, p^*) = \min \{ \max \{ (1 - t\,p(t))^2 \mid t \in \Delta \} \mid p \in \mathcal{P}(N-1) \}.$$

Отсюда видно, что $1 - t\,p^*(t)$ непостоянно, а ν^* сосредоточена на множестве максимумов модуля этого многочлена на Δ; последнее состоит из не более чем $N+1$ точки, так что фактически $\nu^* \in \Phi$ Поэтому (П.22) дает

$$\varepsilon \geq \min \{ G(\nu^*, p) \mid p \in \mathcal{P}(N-1) \} = G(\nu^*, p^*)$$

Правая часть последнего неравенства, как хорошо известно (см., скажем, [19, гл.7]), равна

$$\mathrm{ch}^{-2} \{ N \,\mathrm{arcch}\, (Q+1)/(Q-1) \}$$

так что

$$\varepsilon \geq \mathrm{ch}^{-2} \{ N \,\mathrm{arcch}\, (Q+1)/(Q-1) \} \geq ((Q^{1/2} - 1) / (Q^{1/2} + 1))^{2N},$$

откуда немедленно следует (П.14). □

П.4. Асимптотика сложности общих выпуклых задач по погрешности

Фиксируем $F > 0$ и выпуклое компактное тело G в R^n и рассмотрим класс $\mathfrak{F}$ всех задач выпуклых задач f, таких, что

(v) $\mathrm{Argmin}\,\{ f(x) \mid x \in R^n \} \subseteq G$;

(vi) $\max\,\{ f(x) \mid x \in G \} - \min\,\{ f(x) \mid x \in G \} \leq F$.

<u>Теорема П.3.</u> Пусть в R^n можно указать два подобных с коэффициентом $\alpha \geq 1$ параллелепипеда с соответственно параллельными ребрами, из которых меньший содержится в G, а больший содержит G. Тогда сложность $N_1(\varepsilon)$ класса $\mathfrak{F}$ допускает оценку снизу

$N_1(\varepsilon) \geq c\, n \ln\{ F/\alpha\varepsilon \},\ 0 < \varepsilon < F/2\alpha$,

где c - положительная абсолютная константа; кроме того, всегда справедлива оценка сложности сверху

$N_1(\varepsilon) \leq 3\, n \ln\{ 2F/\varepsilon \},\ 0 < \varepsilon < F/2$.

Эта теорема доказана в [19, гл. 4]. Укажем, что верхняя оценка сложности доставляется методом центров тяжести (см. §3.3). Прокомментируем сформулированный результат. Из него следует, что в асимптотике по $\nu = \varepsilon/F \to 0$ сложность с точностью до абсолютной мультипликативной постоянной есть $n \ln 1/\nu$. Момент ν_G выхода на асимптотику зависит от значения α, т.е. определяется аффинными свойствами тела G. Для параллелепипеда G асимптотика устанавливается сразу - $\alpha = 1$, $\nu_G = 0.5$. Если G - евклидов шар, то $\alpha = n^{1/2}$, и можно взять $\nu_G = n^{-1}$. Укажем, что в случае шара G по теореме П.1 (примененной с $\pi = 0$, радиусом G в качестве R и $L = F/R$) сложность N_1 в "доасимптотическом" диапазоне $\nu \geq n^{-1/2}$ с точностью до абсолютной мультипликативной константы есть. ν^{-2} . В левом конце диапазона сложность с указанной точностью равна n. Доасимптотический диапазон отделяется от области действия стандартной асимптотики $n \ln 1/\nu$ интервалом

$$n^{-1/2} > \nu > n^{-1} ,$$

в котором сложность, очевидно, заключена между $O(n)$ и $O(n \ln n)$. Мы видим, что теоремы П.1 и П.3 доставляют практически законченную информацию о поведении сложности N_1 как функции n и ε на "простых" множествах - параллелепипедах и шарах.

Укажем еще, что для любого выпуклого компактного тела в R^n можно взять $\nu_G = 0.5\, n^{-2}$, так что стандартная асимптотика сложности класса $\mathfrak{F}$ независимо от геометрии G заведомо действует при $\varepsilon/F < 0.5\, n^{-2}$. Сколько-нибудь законченные результаты о сложности N_π с $\pi > 1$ в ситуации П.4 неизвестны.

А.С.Немировский

БИБЛИОГРАФИЧЕСКИЙ КОММЕНТАРИЙ

К ГЛАВЕ 1

§1, 2. Материал этих параграфов достаточно стандартен. Сведения из выпуклого анализа приводятся в большинстве монографий, посвященных методам решения экстремальных задач (см., например, [2,10,13,37]). Систематическое изложение основ выпуклого анализа дано в [41], см. также [16,39].

§3. Равномерно выпуклые функции были введены и исследованы Б.Т.Поляком в [15]. Одна из главных областей применения таких функций - это методы регуляризации, основанные на идеях А.Н.Тихонова (см.[3]). Содержание параграфа в основном соответствует работе [5].

§4. Неравенство (1.4.10) получено Е.Г.Гольштейном и Н.В.Третьяковым [7], а неравенство (1.4.16) - А.С.Антипиным [1]. Обобщения этих неравенств ((1.4.4), (1.4.9) и (1.4.15)) являются, по-видимому, новыми.

§5. Квазивыпуклые и равномерно квазивыпуклые функции были введены Б.Т.Поляком в [34]. Основные результаты параграфа были опубликованы в [6].

§6. Квазиоднородные функции близки к функциям с ограниченной степенью роста, рассмотренным Н.З.Шором [49].

К ГЛАВЕ 2

Одно из первых определений "дифференцируемости" для негладких невыпуклых функций принадлежит Б.Н.Пшеничному [38]. Важный подкласс квазидифференцируемых по Пшеничному функций рассмотрел Е.А.Нурминский [33]. Обобщение понятия дифференцируемости, основанное на использовании верхней производной по направлению, введено Ф.Кларком [53], см. также [32,56]. Близкое определение обобщенно-дифференцируемых функций использовал Н.З.Шор [49]. Квазидифференциальное исчисление развито в работах В.Ф.Демьянова и А.М.Рубинова [9,10].

Лексикографические правила дифференцирования негладких функций предложены в [31]. Однако в этой работе использовалось не совсем удачное обобщение понятия дифференцируемости, что позволило обосновать такие правила только для кусочно-линейных функций. Общая теория лексикографического дифференцирования построена в работах [23,24], результаты которых и изложены в этой главе. Доказательство теоремы 2.7.1 принадлежит А.С.Немировскому.

К ГЛАВЕ 3

§2. Субградиентный метод с постоянным шагом предложен Н.З.Шором [47]. Правило (3.2.3) для выбора шаговых множителей предложено Ю.М.Ермольевым [11] и Б.Т.Поляком [35]. Подробное исследование субградиентного метода проведено в работах [12,36,54].

§3. Метод центров тяжести предложен независимо А.Ю.Левиным [14] и Д.Ньюменом [57]. Доказательство леммы 3.3.4 приведено в [8]. Оценка скорости сходимости и оптимальность метода центров тяжести доказана А.С.Немировским и Д.Б.Юдиным [19]. Недавно появился новый метод отсечений - метод вписанных эллипсоидов [42]. В отличие от метода центров тяжести, реализация одной итерации этого метода обходится в полиномиальное число арифметических операций.

§4. Метод описанных эллипсоидов предложен независимо в работах [48,50].

§5. Метод растяжения пространства предложен в [49]. Приводимая оценка скорости сходимости этого метода (см. теорему 3.5.1) улучшает известную ранее.

Содержание настоящей главы в основном соответствует работе [21].

К ГЛАВЕ 4

§1,2. Проблема построения оптимального метода гладкой выпуклой оптимизации возникла в связи с результатами работ А.С.Немировского и Д.Б.Юдина, посвященных нижним оценкам сложности экстремальных задач [19]. Там же приводятся оптимальные алгоритмы гладкой минимизации для широкого спектра бесконечномерных пространств и "субоптимальный" алгоритм для гильбертова пространства (на каждой итерации приходилось решать вспомогательную задачу двумерной минимизации). Первый оптимальный метод минимизации выпуклых функций с липшицевым градиентом в гильбертовом пространстве предложен в [20], см. также [22]. Результаты параграфов 1. 2 опубликованы в [26].

§3. Описываемый метод идейно близок к методам, предложенным в [18].

§4. Подход, связанный с использованием на каждой итерации основного метода результатов решения линеаризованных вспомогательных подзадач, разработан Б.Н.Пшеничным [40]. В несколько измененном виде эта идеология применялась и в оптимальных методах [19], см. также [18,20]. Схемы метода градиентного спуска, близкие к описанной в этом параграфе, рассматривались в [1,19].

§1. Пример задачи ЛП, на которой симплекс-метод делает экспоненциально большое число шагов, приведен в [59]. Оценки среднего числа шагов симплекс-метода даны в [4,52,51,61]. Полиномиальная разрешимость задачи линейного программирования доказана Л.Г.Хачияном в [43] с помощью метода эллипсоидов, см. также [44,45]. Подробное обсуждение вопросов, связанных с проблемой полиномиальной разрешимости задач ЛП содержится в [46].

§2-6. После того как Н.Кармаркар [58] в 1984г. предложил новый полиномиальный алгоритм ЛП, основанный на применении проективного преобразования, в этой области наметился бурный прогресс. Оценка трудоемкости базовой версии нового алгоритма имела порядок $O(n^4 L)$ (такую же трудоемкость дает метод эллипсоидов). Однако с помощью специального приема, описанного в [58], эту оценку удавалось снизить до $O(n^{3.5} L)$ (аналогичный прием используется и в настоящей главе). За работой Кармаркара последовало много работ, в которых строились полиномиальные алгоритмы ЛП, так или иначе использующие проективное преобразование (см.,например, [17, 62]). Однако следующим существенным продвижением в этой области стал результат Дж.Ренегара [60], показавшего, что оценку $O(n^{3.5} L)$ на задачах ЛП дает обычный метод логарифмических центров, использующий на каждой итерации один шаг метода Ньютона. Оценка метода Ренегара была снижена на полпорядка П.Вадьей [63], которому удалось ввести в этот метод ускорение кармаркаровского типа. Одновременно методы с такой же трудоемкостью были получены в [25-28], причем методы, предложенные в [27,28] базировались на методе логарифмических барьеров и применялись к задаче квадратичного программирования. Независимо оценку $O(n^3 L \ln n)$ для метода логарифмических барьеров в случае задачи ЛП получил К.Гонзага [55]. В §2-6 описываются методы линейного и квадратичного программирования, опубликованные в [25-28].

§7. Результаты этого параграфа являются, по-видимому, новыми. Относительно полиномиальности метода логарифмических барьеров для структуризованных типов нелинейных задач выпуклого программирования см.[30].

Список литературы

1. Антипин А.С. Методы нелинейного программирования, основанные на прямой и двойственной модификации функции Лагранжа. - Препринт. - М.,1979. - 65с. - (ВНИИСИ).
2. Васильев Ф.П. Численные методы решения экстремальных задач. - М.: Наука, 1980.
3. Васильев Ф.П. Методы решения экстремальных задач. - М.: Наука, 1981.
4. Вершик А.М., Спорышев П.В. Оценка среднего числа шагов симплекс-метода и задачи асимптотический интегральной геометрии //ДАН СССР. - 1983. - Т.250, №5. - С,35-47.
5. Владимиров А.А., Нестеров Ю.Е., Чеканов Ю.Н. О равномерно выпуклых функционалах //Вестник Московск. Ун-та. Сер. вычислит. матем. и кибернетика. - 1978. - №3. - С.12-23.
6. Владимиров А.А., Нестеров Ю.Е., Чеканов Ю.Н. О равномерно квазивыпуклых функционалах //Вестник Московск. Ун-та. Сер.вычислит.матем. и кибернетика.- 1978. - №4.-С.18-27.
7. Гольштейн Е.Г., Третьяков Н.В. Градиентный метод минимизации и алгоритмы выпуклого программирования, связанные с модифицированной функцией Лагранжа //Экономика и математич. методы. - 1975. - Т.11,вып.4. - С.730-742.
8. Грюнбаум Б. Этюды по комбинаторной геометрии и теории выпуклых тел. - М.: Наука, 1971.
9. Демьянов В.Ф., Рубинов А.М. О квазидифференцируемых функционалах //ДАН СССР. - 1980. - Т.250, №1. - С.21-25.
10. Демьянов В.Ф., Васильев Л.В. Недифференцируемая оптимизация. - М.: Наука, 1981.
11. Ермольев Ю.М. Методы решения нелинейных экстремальных задач //Кибернетика. - 1966. - №4. - С.1-17.

12. Ермольев Ю.М. Методы стохастического программирования - М.:Наука, 1976.

13. Карманов В.Г. Математическое программирование. - М.: Наука, 1975.

14. Лёвин А.Ю. Об одном алгоритме минимизации выпуклых функций //ДАН СССР. - 1965. - Т.160, №6. - С.1244-1247.

15. Левитин Е.С., Поляк Б.Т. О сходимости минимизирующих последовательностей в задаче на условный экстремум //ДАН СССР. - 1966. - Т.168, №5. - С.997-1000.

16. Экланд И., Темам Р. Выпуклый анализ и вариационные проблемы. - М.: Мир, 1979.

17. Немировский А.С. Об одном алгоритме типа Кармаркара // Известия АН СССР. Техническая кибернетика.- 1987. - Вып. 1. - С.105-118.

18. Немировский А.С.,Нестеров Ю.Е. Оптимальные методы выпуклого программирования //Журн. вычислит. математики и математ. физики. - 1985. - Т.25, вып.3. - С.356 - 369.

19. Немировский А.С., Юдин Д.Б. Сложность задач и эффективность методов оптимизации. - М. Наука, 1979.

20. Нестеров Ю.Е. Метод решения задачи выпуклого программирования со скорости сходимости $O(k^{-2})$ // ДАН СССР - 1983. - Т.269, №3 - С.543-547.

21. Нестеров Ю.Е. Методы минимизации негладких выпуклых и квазивыпуклых функций //Экономика и математ. методы. - 1984. - Т.11,вып.3. - С.519-531.

22. Нестеров Ю.Е. Об одном классе методов безусловной минимизации выпуклой функции, обладающих высокой скоростью сходимости //Журн.вычислит.математики и математ.физики.- 1984 - №7. - С.1090-1093.

23. Нестеров Ю.Е. Лексикографические принципы дифференциро-

вания негладких функций.- Препринт. - М.,1986. - 66с - (ЦЭМИ АН СССР).

24. Нестеров Ю.Е. Техника негладкого дифференцирования //Известия АН СССР. Техническая кибернетика. - 1987 - №1. - С.199-208.

25. Нестеров Ю.Е. Метод линейного программирования с кубической трудоемкостью //Экономика и математ. методы. - 1988 - Т.24,вып.1. - С.174-176.

26. Нестеров Ю.Е. Об одном подходе к построению оптимальных методов минимизации гладких выпуклых функций //Экономика и математ. методы. - 1988. - Т.24,вып.3. - С.509-517.

27. Нестеров Ю.Е. Полиномиальные методы в линейном и квадратичном программировании //Известия АН СССР. Техническая кибернетика. - 1988 - №3. - С.3-6.

28. Нестеров Ю.Е. Полиномиальные итеративные методы линейного и квадратичного программирования //Вопросы кибернетики. - 1988. - С.321- 343.

29. Нестеров Ю.Е. Двойственные полиномиальные алгоритмы линейного программирования //Кибернетика. - 1989. - №1 - С.47-61.

30. Нестеров Ю.Е., Немировский А.С. Полиномиальные методы барьеров в выпуклом программировании //Экономика и математ. методы. - 1988. - Т.24,вып.6. - С.122-131.

31. Нестеров Ю.Е., Черкасский Б.В. Лексикографические правила дифференцирования негладких функций // ДАН СССР. - 1987. - Т.297, №2. - С.441-445.

32. Норкин В.И. Обобщенно-дифференцируемые функции //Кибернетика. - 1980. - №1. - С.9-11.

33. Нурминский Е.А. Численные методы решения детерминированных и стохастических задач. - Киев: Наукова Думка, 1979.

34. Поляк Б.Т. Теоремы существования и сходимость минимизирующих последовательностей для задач на экстремум при наличии ограничений //ДАН СССР. - 1966. - Т.166, №2. - С.287-290.
35. Поляк Б.Т. Один общий метод решения экстремальных задач //ДАН СССР. - 1967. - Т.174, №1. - С.33-36.
36. Поляк Б.Т. Минимизация негладких функционалов // Журн. вычислит. математики и математ. физики. - 1969. - Т.9, №3. - С.509-521.
37. Поляк Б.Т. Введение в оптимизацию.-М.: Наука, 1984.
38. Пшеничный Б.Н. Необходимые условия экстремума. - М.: Наука, 1969.
39. Пшеничный Б.Н. Выпуклый анализ и экстремальные задачи. - М.: Наука, 1980.
40. Пшеничный Б.Н., Данилин Ю.М. Численные методы в экстремальных задачах. - М.: Наука, 1975.
41. Рокафеллар Р.Т. Выпуклый анализ. - М.: Мир, 1973.
42. Тарасов С.В., Хачиян Л.Г., Эрлих И.И. Метод вписанных эллипсоидов //ДАН СССР.- 1988.- Т.296, №5. - С.124-129.
43. Хачиян Л.Г. Полиномиальные алгоритмы в линейном программировании //ДАН СССР.- 1979. - Т.244, №5. - С.1093-1096.
44. Хачиян Л.Г. Полиномиальные алгоритмы в линейном программировании//Журн. вычислит. математики и математ. физики. - 1980. - Т.20, №1. - С.51-69.
45. Хачиян Л.Г. О точном решении систем линейных неравенств и задач линейного программирования //Журн. вычислит. математики и математ.физики.- 1982. - Т.22, №4. - С.67-84.
46. Хачиян Л.Г. Сложность задач линейного программирования.- М.:Знание, 1987.
47. Шор Н.З. Применение обобщенного градиентного спуска в

блочном программировании // Кибернетика. - 1967. - №3. - С.17-22.

48. Шор Н.З. Метод отсечения с растяжением пространства для решения задач выпуклого программирования //Кибернетика.- 1977. - №1. - С.42-50.

49. Шор Н.З. Методы минимизации недифференцируемых функций и их приложения. - Киев: Наукова Думка, 1979.

50. Юдин Д.Б., Немировский А.С. Информационная сложность и эффективные методы решения выпуклых экстремальных задач //Экономика и математ. методы. - 1976. - Т.12, вып.2. - С.357-369.

51. Adler I., Meggido N., Todd M.J. New results on average behavior of simplex algorithms//Bulletin of the American Math. Soc. - 1984 - V.11, №2. - P.11-32.

52. Borgwardt K.-H. The average number of steps required by the simplex method is polynomial. Zeitschrift fur Operation Research, Ser.A-B, 26. 1982.

53. Clarke F.H. Generalized gradients and applications // Trans.Amer.Math.Soc. - 1975. - V.205. - P.247-262.

54. Goffin J.L. On the convergence rate of subgradient optimization methods //Math.Progr. - 1977. - V.13, №.3. - P. 329-347.

55. Gonzaga C.C. An algorithm for solving linear programming problems in $O(n^3 L)$ operations. - Tech.Report. - 1987. - 45p. - California, Berkley (Dept. of Electrical Engineaning and Computer Sciences, 94720).

56. Mifflin R. Semismooth and semiconvex functions in constrained optimization //SIAM J. Control and Optimization. - 1977. - V.15,N.6. - P.959-972.

57. Newman D.J. Location of the maximum on unimodal surfaces

//J.ACM. - 1965. - V.12.N3. - P.395-398.

58. Karmarkar N. A new polinimial-time algorithm for linear programming //Combinatorica. - 1984. - N.4. - P.373-395.
59. Klee V., Minty G.J. How good is the simplex algirithm // Inequalities 3/ed. O.Shisha. - New Jork: Academic Press. - 1972. - P.159-179.
60. Renegar G.J. A. polinomial - time algorithm, based on Newton's method for linear programming // Math. Progr. - 1988. - V.40.N1. - P.59-93.
61. Smale S. The problem of average speed of the simplex method //"Mathematical Programming. The State of Art" /eds Bachnem A., Grotschel M., Korte B. - Berlin: Shpringer Verlag., 1983.
62. Todd M.J., Burrell B.P. An extention of Karmarkar's algorithm for linear programming, using dual variables //Algoritmica. - 1986. - N.1. - P.409-424.
63. Vaidya P. An algorithm for linear programming which requires $O(((m+n)n^2+(m+n)^{1.5}n)L)$ arithmetic operations. ATT Bell Laboratories, Murray Hill. - New Jork. - 1987.

О Г Л А В Л Е Н И Е

Ju.E.Nesterov Effective methods in nonlinear programming

The book presents in high intelligible form recent results concerning effective algorithms for linear and nonlinear optimization problems. Among the most important results which have not yet been highlighted in earlier monographs are the following:

- *The theory of lexicographic differentiation of nonsmooth functions*. In terms of this theory, the set of differentiable functions is rather wide: it contains in particular all smooth and all convex functions and is closed with respect to algebraic operations as well as to superposition of its elements. In the theory, the generalized differentials are, same as with the classical approach, linear operators rather than sets. The differential can easily be computed from analytical or algorithmic representation of the function by using calculus rules which are similar to the classical ones. All basic results of classical analysis (such as, for instance, the Newton-Leibnitz formula) are valid in this theory.

- *Optimal methods for smooth convex optimization*. These are new methods whose theoretical rate of convergence is essentially higher than that for the known methods. Moreover, this rate of convergence attains the theoretical potential bounds for the classes of problems under consideration while from the computational viewpoint these methods are no more complicated than usual gradient descent method.

- *Polynomial-time methods for Linear Programming and Quadratic Programming*. Recently the theory of polynomial-time methods for LP has become one of the fast progressing areas in Computer Science. The amount of operations required by these methods per a bit of accuracy of the resulting approximate solution is a polynomial in the problem size. For the best known polynomial-time methods, this amount is proportional to the cube of the dimension of the LP problem. The author describes a number of new LP methods with such "record" estimates and then generalizes the methods to obtain polynomial-time QP algorithms. This generalization does not lead to any increase in the complexity of the algorithms.

The monograph is expected to be of interest for those working in the area of Mathematical Programming.

Contents

Научное издание

Нестеров Юрий Евгеньевич

ЭФФЕКТИВНЫЕ МЕТОДЫ В НЕЛИНЕЙНОМ ПРОГРАММИРОВАНИИ

Заведующая редакцией *Г. И. Козырева*
Редактор издательства *Н. Г. Давыдова*
Обложка художника *С. Н. Голубева*
Художественный редактор *Н. С. Шеин*
Технический редактор *Т. Н. Зыкина*
Корректор *Г. Г. Казакова*

ИБ № 2168

Набрано на ПЭВМ. Подписано в печать с оригинала-макета 15.08.89. Т-23796. Формат 60х88/16. Бумага офс. № 2. Гарнитура машинописная. Печать офсетная. Усл. печ. л. 18,62. Усл. кр.-отт. 18,87. Уч.-изд. л. 11,18. Тираж 2030 экз. Изд. № 22192. Заказ №2406 . Цена 2 р. 20 к.
Издательство "Радио и связь". 101000 Москва, Почтамт, а/я 693

Московская типография № 4 Союзполиграфпрома при Государственном комитете СССР по делам издательств, полиграфии и книжной торговли.
129041 Москва, Б. Переяславская ул., д. 46

2 р. 20 к.

Ю.Е. Нестеров

ЭФФЕКТИВНЫЕ МЕТОДЫ в нелинейном программировании

Теория лексико-графического дифференцирования

Оптимальные методы минимизации гладких функций

Полиномиальные алгоритмы линейного и квадратичного программирования