АКАДЕМИЯ НАУК УКРАИНСКОЙ ССР
ОРДЕНА ЛЕНИНА ИНСТИТУТ КИБЕРНЕТИКИ

На правах рукописи

Н.З. ШОР

МЕТОДЫ МИНИМИЗАЦИИ НЕДИФФЕРЕНЦИРУЕМЫХ ФУНКЦИЙ И ИХ ПРИЛОЖЕНИЯ

(Специальность 008 – вычислительная математика)

А в т о р е ф е р а т

КИЕВ – 1970

АКАДЕМИЯ НАУК УКРАИНСКОЙ ССР
ОРДЕНА ЛЕНИНА ИНСТИТУТ КИБЕРНЕТИКИ

На правах рукописи

Н.З. Шор

МЕТОДЫ МИНИМИЗАЦИИ НЕДИФФЕРЕНЦИРУЕМЫХ ФУНКЦИЙ И ИХ ПРИЛОЖЕНИЯ

(Специальность 008 – вычислительная математика)

А в т о р е ф е р а т
Диссертации, представленной на соискание
ученой степени доктора физико-математических
наук

КИЕВ – 1970

Работа выполнена в Ордена Ленина Институте кибернетики Академии наук УССР.

Официальные оппоненты:

член-корр.АН СССР, доктор физ.-мат. наук Т.М. ЭНЕЕВ
доктор физ.-мат. наук, профессор В.В. ИВАНОВ
доктор физ.-мат. наук Ф.Л. ЧЕРНОУСЬКО

Ведущее предприятие – Вычислительный центр АН СССР, г. Москва

Автореферат разослан "__11__" ____мая_______1970 г.
Защита диссертации состоится "____" _____________1970 г. на заседании Ученого совета Института кибернетики АН УССР по присуждению ученых степеней, г. Киев-28, Проспект науки 109.

С диссертацией можно ознакомиться в библиотеке Института.

УЧЕНЫЙ СЕКРЕТАРЬ СОВЕТА ИК АН УССР
кандидат техн.наук /К.Д. Жук/

Наше время характеризуется внедрением математических методов и математического стиля мышления в широкие области человеческой деятельности. Особенно большую роль в приложениях играют математические методы оптимизации, которые успешно применяются для выбора наиболее рациональных вариантов технико-экономических решений в задачах планирования, проектирования и исследования операций, управления технологическими процессами, выбора наилучших схем функционирования сложных динамических систем и т.п. Кроме того, задачи оптимизации очень тесно связаны с задачами решения различного рода систем уравнений и неравенств, отыскания спектра операторов, так что эти области методологически обогащают друг друга. И здесь мы имеем необозримое поле приложений: задачи механики, теоретической физики, квантовой химии и т.п.

Следует отметить, что развитие вычислительной техники также во многом стимулировало интерес к проблемам оптимизации, так как оно с одной стороны давало возможность реализовывать сложные алгоритмы и решать задачи довольно высокой размерности, обычно встречающиеся на практике; с другой стороны появилась возможность проводить в короткие сроки проверку эффективности новых методов и алгоритмов.

В последние 30 лет наблюдается очень быстрый рост числа публикаций по методам оптимизации. Именно эти годы – период становления таких важных разделов теории оптимальных решений как линейное и выпуклое программирование, последовательный статистический анализ, теория игр, динамическое и стохастическое программирование, теория оптимального управления динамическими системами. Усилился интерес к дискретным задачам оптимизации (дискретное программирование, экстремальные задачи на графах и пр.).

Следует сказать, что эти новые разделы при своем развитии существенно опирались на достижения классической математики. Так, в основе теории математического программирования лежит идея использования множителей Лагранжа и изучения возникающих при этом ситуаций двойственности, теория оптимального управления динамическими системами возникла как естественное обобщение вариационного исчисления, в основе методов возможных направлений лежит идея наискорейшего спуска, восходящая к Коши, симплекс-метод можно рассматривать как обобщение схемы исключения Гаусса. Однако существенно изменился характер исследований. Наряду с продолжением исследований по обоснованию необходимых и достаточных условий экстремальности все большее место занимают работы, связанные с разработкой новых алгоритмов, строгим математическим обоснованием сходимости, оценки их точности и эффективности. Это вызвано в первую очередь практическими потребностями. Дело в том, что многие численные методы, хорошо зарекомендовавшие себя при решении определенного круга задач, оказываются неприспособленными или малоэффективными для решения близких по формулировке задач из другой области.

Градиентные методы в силу своей “идейной” и алгоритмической простоты являются, пожалуй, наиболее часто употребляемыми при решении задач численной оптимизации непрерывно дифференцируемых функций. Однако они оказываются медленно сходящимися в том случае, когда скорости изменения градиента при движении в различных направлениях резко различаются (функции “овражного” типа) и не приспособлены для решения задач минимизации функций, не всюду дифференцируемых.

В XIX и в начале XX века, в период расцвета аналитических методов, вопросы минимизации недифференцируемых функций не привлекали большого внимания математиков и рассматривались лишь эпизодически (например, в рамках теории чебышевских приближений, в теории линейных неравенств). В настоящее время в связи с развитием математического программирования эти вопросы представляют большой интерес.

Проблемы минимизации функций не всюду дифференцируемых, естественным образом появляются при решении самых разнообразных задач. Перечислим некоторые наиболее важные классы таких задач.

1/ Решение систем нелинейных уравнений и неравенств.

Пусть требуется решить систему уравнений и неравенств:

$$f_i(x) = 0\,; \qquad i = 1,...m\,;$$
$$x \in E_n\,;$$
$$\varphi_j(x) \le 0\,; \qquad j = 1,...l\,;$$

Эта задача сводится к минимизации функции

$$\Phi(x) = \max_{\substack{1\le i\le m \\ 1\le j\le l}} \left[|f_i(x)|, \varphi_j(x) \right].$$

Функция $\Phi(x)$, вообще говоря, не является всюду дифференцируемой, даже в предположении, что $f_i(x)$ и $\varphi_i(x)$ являются гладкими.

2/ Общая задача минимаксного типа:

найти $\inf \Phi(x)$,

где $\Phi(x) = \sup\limits_{\omega\in\Omega} f(x,\omega)$.

3/ Решение задач оптимизации с использованием схем разложения.

Простейший класс таких задач формулируется следующим образом:

найти $\min\limits_{x\in E_n}(F(x)$,

где $F(x) = \min\limits_{y\in E_m} f(x,y)$; $x \in E_n$.

Если даже функция $f(x,y)$ является непрерывно дифференцируемой по совокупности переменных (x,y), функция

$F(x)$ может быть не всюду дифференцируемой.

4/ Реализация метода штрафных функций для решения задач математического программирования.

Для получения высокой точности решения обычно используются штрафные функции с разрывным градиентом, так как при использовании гладких штрафных функций исходная задача с ограничениями лишь приближенно сводится к задаче без ограничений и для получения достаточной точности в этом случае приходится строить процессы с последовательным изменением «штрафного» параметра.

5/ Задачи минимизации кусочно-линейных функций.

6/ Задачи оптимального управления.

Широкий класс задач оптимального управления, используя принцип максимум, можно свести к решению двухточечных краевых задач для систем дифференциальных уравнений с разрывными правыми частями. Решение этих краевых задач сводится к задачам минимизации функций, не везде дифференцируемых.

С другой стороны, задачи оптимального управления с помощью разностной аппроксимации сводятся к задачам математического программирования, которые часто решаются с использованием метода штрафных функций, схем разложения; здесь опять возникают задачи минимизации недифференцируемых функций.

В диссертации дано изложение работ автора, посвященных разработке и обоснованию методов минимизации недифференцируемых функций, вопросам ускорения сходимости градиентных методов (и их аналогов в случае недифференцируемых функций), а также приложениями этих методов к решению задач математического программирования.

Предлагаемые методы и алгоритмы были получены в результате исследований, связаных с поиском эффективных способов решения большого числа практических задач из области оптимального планирования, проектирования и исследования операций.

Диссертация состоит из введения, трех глав (I. Обобщенный градиентный спуск, II. Использование операций растяжения пространства для ускорения сходимости методов градиентного типа, III. Применение методов минимизации недифференцируемых функций для решения задач математического программирования) и заключения.

В первой главе излагаются теоретические вопросы, связанные с исследованием сходимости различных вариантов метода обобщенного градиентного спуска, применяемого для минимизации выпуклых (вообще говоря, недифференцируемых) функций. Этот метод был предложен автором в 1962 году применительно к решению сетевых транспортных задач [1].

В кандидатской диссертации [2] исследовались вопросы точности определения минимума выпуклой функции при применении метода обобщенного градиентного спуска с постоянным малым шаговым множителем. В дальнейшем обоснование сходимости метода обобщенного градиентного спуска было проведено в работах Б.Т.Поляка [3], Ю.М.Ермольева [4] и автора [5–6].

Определение 1. Пусть задана выпуклая функция $f(x)$ определённая в n-мерном евклидовом пространстве E_n. Обобщённым градиентом функции $f(x)$ в точке x_0 называется вектор $g_f(x_0)$, удовлетворяющий следующему неравенству для произвольного $x \in E_n$:

$$f(x) - f(x_0) \geq (g_f(x_0), x - x_0) .$$

Такое название является естественным, так как в случае, когда $f(x)$ является дифференцируемой в точке x_0 вектор $g_f(x_0)$ определяется однозначно и совпадает с градиентом функции $f(x)$ в точке x_0.

В общем случае, вектор $g_f(x_0)$ определяется неоднозначно. Пусть $G_f(x_0)$ – множество обобщенных градиентов в точке x_0. Это множество для произвольного $x_0 \in E_n$ является ограниченным, замкнутым и выпуклым. В дальнейшем под $g_f(x_0)$ будет пониматься произвольный представитель множества $G_f(x_0)$. Необходимым и достаточным условием того, что в точке x_0 достигается минимум функции $f(x)$, является то, что множество $G_f(x_0)$ содержит нулевой вектор.

Определение 2. Пусть $\eta \in E_n$, $\eta \neq 0$.

Обобщенной производной по направлению η от функции $f(x)$ в точке x_0 назовем число

$$g_f^{\eta}(x_0) = \frac{(g_f(x_0), \eta)}{\| \eta \|}.$$

$g_f^{\eta}(x_0)$ удовлетворяет неравенству

$$f(x+\eta) - f(x) \geq g_f^{\eta}(x) \| \eta \|.$$

Отметим, что для случая выпуклых функций, определённых в банаховом пространстве, аналогом обобщенного градиента является опорный функционал, свойства которого подробно рассмотрены в работах [7–9].

Радемейстеру принадлежит теорема о том, что выпуклая функция почти везде дифференцируема [10]. Это означает, что если имеется произвольное измеримое множество в пространстве E_n и с отличной от нуля и конечной n-мерной мерой и в этом множестве случайным образом в соответствии с равномерным распределением выбирается точка, то с вероятностью 1 выпуклая функция $f(x)$ будет дифференцируемой в этой точке и, следовательно, непрерывно дифференцируемой.

Таким образом, почти везде обобщенный градиент выпуклой функции определяется однозначно и совпадает с градиентом. Если предположить возможность вычисления значения выпуклой функции $f(x)$ в произвольной точке x, то вычисление градиента с заданной точностью в точке x_0, где $f(x)$ дифференцируема, не предоставляет трудности благодаря свойству монотонности производных по направлению. Для многих частных классов задач (например, задачи минимаксного типа) вычисление обобщенного градиента даже в тех точках, где минимизируя функция недифференцируема, не представляет принципиальных трудностей. Исходя из этого при описании конкретных алгоритмов автор предполагает, что имеется возможность вычисления обобщенного градиента функции $f(x)$ в произвольной точке x_0.

Всё же вопрос о вычислении обобщенного градиента $g_f(x_0)$ с заданной точностью в точке x_0, где $f(x)$ не является дифференцируемой, представляет определенный интерес. К сожалению, в общем случае этот вопрос решается отрицательно. Автор показал (Лемма 1.2), что если вся информация о функции $f(x)$ получается только за счёт вычисления значения $f(x)$ в произвольных точках, то не существует алгоритма, который определял бы для произвольной выпуклой функции $f(x)$ такой вектор $\tilde{g}_f^{\delta}(x_0)$ для заданного произвольного $\delta > 0$, что

$$\min_{g \in G_f(x_0)} \| \tilde{g}_f^{\delta}(x_0) - g \| \leq \delta .$$

Пусть ставится задача минимизации функции $f(x)$. Методом обобщенного градиентного спуска мы назовем процедуру построения последовательности $\{x_k\}$ следующего вида

$$x_{k+1} = x_k - h_k(x_k) g_f(x_k), \; k = 0,1,...., \tag{1}$$

(x_0 – нулевое приближение задается до начала процесса спуска).

Формула (1) по внешнему виду ничем не отличается от формулы, задающей спуск в обычных градиентных методах. Однако существенное различие будет состоять в способе регулировки шага. Дело в том, что обычно применяемые для непрерывных дифференцируемых функций методы регулировки шага:

а/ $h_k = const$ (в методе с постоянным шаговым множителем);

б/ выбор h_k из условия $\min\limits_{h_k} f(x_k - h_k g_f(x_k))$ (как в методе наискорейшего спуска), для произвольной выпуклой функции, вообще говоря, неприменимы.

Далее для обоснования методов обобщенного градиентного спуска (сокращённо: О.Г.С.) доказываются следующие утверждения:

Лемма 1.3. Пусть при использовании методов обобщенного градиентного спуска $h_k(x_k) = \dfrac{h}{\| g_f(x_k) \|}$; $h > 0$.

Тогда для любого $\varepsilon > 0$ и $x^* \in \Omega$ (Ω – множество минимумов функции $f(x)$) найдутся такие $k = \bar{k}$ и $x = \bar{x}$, что будет выполняться следующее свойство:

$$f(x_{\bar{k}}) = f(\bar{x}), \text{ причем}$$

$$\| \bar{x} - x^* \| \le \frac{h}{2}(1+\varepsilon).$$

Теорема 1.1 Для произвольного $\delta > 0$ можно найти такое h_δ, что при применении ОГС с шагом $h = h_\delta$ при любом $x_0 \in E_n$ либо найдется k^* такое, что $x_{k^*} \in \Omega$, либо найдется такая подпоследовательность $k_1 < k_2 < k_3 < \ldots$,что $f(x_{k_i}) - \min\limits_{x \in E_n} f(x) < \delta$ ($\Omega \ne \varnothing$).

Теорема 1.2 Если функция имеет множество минимумов Ω, содержащее сферу радиуса большего, чем $\frac{h}{2}$, то при использовании метода ОГС с $h_k(x_k) = \frac{h}{\| g_f(x_k) \|}$ найдется такое k^*, что $x_{k^*} \in \Omega$.

Имеется несколько вариантов доказательства теоремы о сходимости метода ОГС. Все они основаны на изучении поведения последовательности $\{\rho_k\}$, где $\rho_k = \min_{x \in \Omega} \| x_k - x \|$, ([3], [4], [6]), подобно тому, как это делалось в [2].

В диссертации приводится принадлежащий автору вариант доказательства теоремы о сходимости ОГС.

Теорема 1.3. Пусть $f(x)$ – выпуклая функция, определенная в E_n, множество минимумов которой Ω ограничено, и задана последовательность положительных чисел $\{h_k\}$, $k = 1, 2, \ldots$, обладающая следующими свойствами:

$$\lim_{k \to \infty} h_k = 0; \quad \sum_{k=1}^{\infty} h_k = \infty .$$

Тогда последовательность $\{x_k\}$, $k = 0, 1, \ldots$, образованная по формуле:

$$x_{k+1} = x_k - h_k \frac{g_f(x_k)}{\| g_f(x_k) \|},$$

при произвольном $x_0 \in E_n$ обладает одним из следующих свойств:

либо найдется такое $\bar{k}$, что $x_{\bar{k}} \in \Omega$;

либо

$$\lim_{k \to \infty} \min_{x \in \Omega} \| x_k - x \| = 0, \ \lim_{k \to \infty} f(x_k) = \min_{x \in E_n} f(x) .$$

Приводится модификация этой теоремы, когда шаговый множитель умножается на ненормированный градиент.

Теорема 1.4. Если при условиях теоремы 1.3 мы образуем последовательность $\{x_k\}$ по формуле:

$$x_{k+1} = x_k - h_{k+1} g_f(x_k),\ k = 0,1,\dots,\ x_0 \in E_n \text{ – произвольная точка,}$$

$$\lim_{k\to\infty} h_k = 0,\ \sum_{k=1}^{\infty} h_k = \infty,$$

то либо для некоторого $\bar{k}$ $g_f(x_{\bar{k}}) = 0$ и $f(x_{\bar{k}}) = \min_{x\in E_n} f(x)$, либо

$$\lim_{k\to\infty} \min_{x\in\Omega} \| x_k - x \| = 0,\ \lim_{k\to\infty} f(\mathrm{x}_k) = \min_{x\in E_n} f(x).$$

При некоторых дополнительных предположениях о функции $f(x)$ можно применить регулировку шагового множителя, обеспечивающую сходимость к минимуму со скоростью геометрической прогрессии.

Теорема 1.5. Пусть $x^*(x) \in \Omega$ удовлетворяет условию: $\| x - x^*(x) \| = \min_{y\in\Omega} \| x - y \|$ и выполняется следующее неравенство:

$$(g_f(x), x - x^*(x)) \ge \cos\varphi \cdot \| g_f(x) \| \cdot \| x - x^*(x) \| \qquad (2)$$

где φ – фиксированное число, $\dfrac{\pi}{4} \le \varphi \le \dfrac{\pi}{2}$ при всех $x \in E_n$.

Тогда, если при заданном x_0 мы выбираем величину $h = h_1 \ge \| x_0 - x^*(x_0) \| \cdot \cos\alpha$ и определим $\{h_k\}$ в соответствии с рекуррентной формулой:

$$h_{k+1} = h_k \cdot \sin\varphi,\ k = 1,2,\dots,$$

и $\{x_k\}$ по формуле: $x_{k+1} = x_k - h_k \dfrac{g_f(x_k)}{\| g_f(x_k) \|}$, $k = 0,1,\dots,$

то либо при некотором $\bar{k}$ $g_f(x_k) = 0$, либо для всех $k = 1,2,\dots$ будет выполняться неравенство:

$$\| x_k - x^*(x_k) \| \le \frac{h_{k+1}}{\cos\gamma}.$$

Постоянная $\cos\varphi$ в выражении (2) тесно связана с параметрами, характеризующими степень “вытянутости” гиперповерхности уровня функции $f(x)$. С использованием этой связи доказывается следующая теорема:

Теорема 1.6. Пусть выпуклая функция $f(x)$ определена на E_n, x^* – единственная точка минимума и задана точка $x_0 \in E_n$, числа $\delta \geq \sqrt{2}$ и $h = h_1 \geq \dfrac{\| x_0 - x^* \|}{\delta}$.

Рассмотрим множество $Y = \left\{ y : \| y - x^* \| \leq \delta h \right\}$.

Если для любой пары точек x, $z \in Y$ и такой, что $f(x) = f(z) \neq f(x^*)$ выполняется условие:

$$\frac{\| x - x^* \|}{\| z - z^* \|} \leq \delta ,$$

то последовательность $\{x_k\}$, образованная с помощью рекуррентных формул:

$$x_{k+1} = x_k - h_k \frac{g_f(x_k)}{\| g_f(x_k) \|} ,$$

$$h_{k+1} = h_k \frac{\sqrt{\delta^2 - 1}}{\delta} ,$$

сходится к x^* со скоростью геометрической прогрессии:

$$\| x - x^* \| \leq h_{k+1} \delta ,$$

за исключением того случая, когда для некоторого $\bar{k}$ $g_f(x_{\bar{k}}) = 0$, т.е. $x_{\bar{k}} = x^*$.

Показывается, что в частном случае, когда $f(x) = (Ax, x)$, где A - положительно определенный оператор, $\underline{\lambda}$ – наименьшее, $\bar{\lambda}$ -

наибольшее собственные значения оператора A, $\cos\varphi$ в неравенстве (2) можно взять меньшим или равным $\dfrac{2\sqrt{\underline{\lambda}\overline{\lambda}}}{\underline{\lambda}+\overline{\lambda}}$, при этом $\sin\varphi \geq \dfrac{\overline{\lambda}-\underline{\lambda}}{\overline{\lambda}+\underline{\lambda}}$.

Этот результат сопоставляется с результатом Л.В.Канторовича [11] о скорости сходимости метода наискорейшего спуска для квадратичных функционалов. Показывается, если мы имеем возможность получить достаточно точную оценку $\hat{\rho} \geq \rho$ величины $\rho = \dfrac{\overline{\lambda}}{\underline{\lambda}}$, то метод ОГС в форме, данной в условии 1.5 , может оказаться при выборе $\sin\varphi = \dfrac{\hat{\rho}-1}{\hat{\rho}+1}$ вполне пригодным для решения линейных систем большой размерности.

Рассматривается еще один вариант метода ОГС, в котором шаговый множитель остается в течении определенного числа шагов постоянным, а затем уменьшается в 2 раза.

Теорема 1.7. Пусть для выпуклой функции $f(x)$ выполняются условия теоремы 1.6 и $\delta \geq 2$. Рассмотрим при заданном x_0 следующий итеративный процесс:

$$x_{k+1} = x_k - h_k \frac{g_f(x_k)}{\| g_f(x_k) \|},$$

где $h_{k+1} = h_0 \cdot 2^{-\left[\frac{k+1}{N}\right]}$.

При достаточно большом h_0 и $N \geq 3\delta^2 + 1$

$$\| x_k - x^* \| \leq 2\delta h_{k+1}.$$

Затем в условиях применимости теоремы 1.5 исследуются вопросы уточнения в процессе счета оценок параметров h и φ, что позволяет строить модификации алгоритмов ОГС, когда h и φ неизвестны.

Вторая глава посвящена вопросам ускорения сходимости методов градиентного типа с использованием операции растяжения пространства.

Во многих случаях, применяя методы градиентного типа, мы сталкиваемся с их медленной сходимостью. Теоретические оценки скорости сходимости, полученные для наиболее простых модификаций градиентных методов, являются во многих случаях весьма неутешительными, и этот пессимизм подтверждается практическими расчетами.

Легко видеть, что если ограничиться на каждом шаге движением строго в направлении, противоположном градиенту (или его аналогам – обобщенному градиенту, квазиградиенту и т.п.), то с помощь одной лишь “удачной” процедуры регулировки шага трудно добиться в достаточно общем случае значительного ускорения сходимости. В самом деле, медленная скорость сходимости связана с тем, что антиградиент может образовывать угол, близкий к $\frac{\pi}{2}$, с направлением на точку минимума. В такой ситуации расстояние до точки минимума убывает медленно, а значит и скорость уменьшения шага не может быть слишком большой, чтобы гарантировать сходимость к минимуму, а это и приводит к медленной сходимости. Эти эвристические соображения хорошо подтверждаются результатами исследований по методу обобщенного градиентного спуска.

С другой стороны, имеется простая возможность изменения углов между направлением градиента и направлением на точку минимума – использование линейных неортогональных преобразований пространства. Возникает идея построения в процессе последовательных приближений линейных операторов, изменяющих метрику пространства, и выбора направления градиента для функции, определённой в пространстве с изменённой метрикой. Соответствующее направление в первоначальном пространстве, вообще говоря, не будет совпадать с направлением градиента. Так появляется класс градиентных алгоритмов с изменённой метрикой.

Среди работ этого направления отмечаются работы Давидона [12], Флетчера и Пауэла [13], Э.Л.Акима и Т.М.Энеева [14], Б.Н.Пшеничного [15] и др.

Интересные алгоритмы, обладающие ускоренной сходимостью получены на основе переноса схемы метода сопряженных градиентов на неквадратичные функционалы (Даниэль [16], Б.Т.Поляк [17] и др.)

Другое направление, связанное с ускорением сходимости градиентных методов, – методы “овражного” поиска экстремумов (И.М.Гельфманд, М.Л.Цетлин [18]) и другие полуэвристические методы (например метод “вращающихся координат” Розенброка [19]), которые успешно применяются для решения сложных задач минимизации, для которых градиентные методы оказываются неэффективными.

Если мы имеем дело с функциями не всюду дифференцируемыми, то приведенные выше методы ускорения сходимости неприменимы. Скажем, если мы минимизируем кусочно-линейную функцию, то гессиан её во всех точках, где он существует, равен 0, и, следовательно, его значение не дает никакой информации о точке минимума. Не применимы в этом случае также методы основанные на определении минимума в данном направлении (например метод Розеброка или сопряженных градиентов), так как имеется опасность, что процесс минимизации остановится, не дойдя до точки минимума

Таким образом, требуется новые идеи и новые методы. Для ускорения сходимости методов градиентного типа в случае не гладких функций автор предложил использовать операцию растяжения пространства. В диссертации рассматриваются две разновидности алгоритмов с растяжением пространства.

Первая из них связана с операцией растяжения пространства в направлении градиента (или его аналогов: обобщенного градиента, квазиградиента и т.п); вторая – с операцией растяжения пространства в направлении разности градиентов в двух последовательных точках минимизирующей последовательности.

Пусть задан вектор $\xi \in E_n$, $\|\xi\| = 1$, и число $\alpha \geq 0$. Каждый вектор $x \in E_n$ однозначно представим в следующем виде:

$$x = \gamma_\xi(x)\xi + d_\xi(x) \tag{3}$$

при условии

$$(\xi, d_\xi(x)) = 0\,. \tag{4}$$

Из (3), (4) получаем:

$$\gamma_\xi(x) = (x, \xi)\,;\; d_\xi(x) = x - (x, \xi)\xi\,. \tag{5}$$

Определение. Оператором растяжения пространства E_n в направлении ξ с коэффициентом α назовём оператор $R_\alpha(\xi)x$, действующий следующим образом на вектор, представленный в форме (3):

$$R_\alpha(\xi)x = \alpha\gamma_\xi(x)\xi + d_\xi(x)\,. \tag{6}$$

Из этого определения сразу следует:

1/ $R_\alpha(\xi)x = \alpha(x, \xi)\xi + [x - (x, \xi)\xi] = (\alpha - 1)(x, \xi)\xi + x$; (7)

2/ оператор $R_\alpha(\xi)$ – линейный симметричный оператор;

3/ $R_{\alpha\beta}(\xi) = R_\alpha(\xi) \cdot R_\beta(\xi)$; (8)

4/ при $\alpha > 0$

$$R_\alpha^{-1}(\xi) = R_{\frac{1}{\alpha}}(\xi)\,; \tag{9}$$

5/ оператор $R_0(\xi)$ является оператором проектирования на подпространство, ортогональное вектору ξ:

$$R_0(\xi)x = d_\xi(x)\,; \tag{10}$$

6/ оператор $R_\alpha(\xi)$ при $n \geq 2$ имеет 2 собственных числа $\lambda_1 = \alpha$; $\lambda_2 = 1$. Первому из них соответствует подпространство собственных векторов, порожденное вектором ξ , второму – подпространство собственных векторов, состоящее из векторов, ортогональных ξ ;

7/ пусть координаты вектора ξ в некоторой ортонормированной системе координат $S=\{e_1,...,e_n\}$ равны $\xi^{(1)},...,\xi^{(n)}$. Тогда в силу (7) в этой системе координат преобразованию $R_\alpha(\xi)$ соответствует матрица $R_\alpha(\xi)$ с элементами $\{r_{ij}\}$, вычисленными по следующей формуле:

$$r_{ij}=\begin{cases}(\alpha-1)\xi^{(i)}\xi^{(j)} \text{ } \textit{для } i\neq j\\ (\alpha-1)\xi^{(i)^2}+1 \text{ } \textit{для } i=j\end{cases} \tag{11}$$

8/ в силу формулы (7) вычисление вектора $R_\alpha(\xi)x$ требует $(2n+1)$ операций умножения. Вычисление матриц вида $R_\alpha(\xi)A$ или $AR_\alpha(\xi)$ при заданной матрице A, ξ и α требует $n(2n+1)$ операций умножения.

9/ пусть x – произвольный ненулевой вектор из E_n.

Тогда

$$\frac{\| R_\alpha(\xi)\|}{\| x\|}=\sqrt{1+\frac{(\alpha^2-1)(\xi,x)^2}{\| x^2\|}}\ . \tag{12}$$

Затем рассматривается класс алгоритмов минимизации выпуклых функций, на каждом шаге которых движение в направлении обобщенного антиградиента сочетается с операцией растяжения пространства аргументов в этом же направлении. Алгоритмы этого класса мы будем называть алгоритмами обобщенного градиентного спуска с растяжением пространства (сокращенно: ОГСРП–алгоритмы).

Предполагается, что имеется алгоритм, позволяющий точно вычислять обобщенный градиент $g_f(x)$ минимизируемой выпуклой функции $f(x)$ в произвольной точке $x\in E_n$, а также алгоритмы вычисления последовательностей положительных чисел h_k и α_k (шаговых множителей и коэффициентов растяжения пространства),

$k=1,2,\dots$, задано начальное приближение x_0 и начальная матрица $B_0 = A_0^{-1}$ (в частности, можно брать $B_0 = E$).

При этих условиях определяется бесконечно-шаговый процесс, (k+1)-й шаг которого описывается следующим образом ($k = 0,1,2,\dots$):

Вычисляем:

1/ $g_f(x_k)$; (если $g_f(x_k) = 0$, происходит остановка вычислений, т.к. x_k дает точку минимума).

$$2\,/\; g_{\varphi_k}(y_k) = B_k^* g_f(x_k), \tag{13}$$

где

$\varphi_k = f(B_k y)$; $y_k = A_k x_k$; B_k^* – оператор, сопряженный оператору B_k.

$$3/\; \xi_{k+1} = \frac{g_{\varphi_k}(y_k)}{\| g_{\varphi_k}(y_k) \|}; \tag{14}$$

4/ h_{k+1};

5/ α_{k+1};

$$6/\; x_{k+1} = x_k - B_k h_{k+1} \xi_{k+1}; \tag{15}$$

$$7/\; B_{k+1} = A_k^{-1} = B_k R_{\frac{1}{\alpha_{k+1}}}(\xi_{k+1}); \tag{16}$$

8/ переходим к (*k*+*2*)-му шагу.

Заметим, что если $B_0 = E$ и $\alpha_k = 1$, $k = 1,2,\dots$, то метод ОГСРП становится эквивалентным обобщенному градиентному спуску.

Далее в диссертации формулируются некоторые достаточно общие условия сходимости метода ОГСРП и оценивается скорость сходимости. Затем строятся конкретные алгоритмы, для которых эти условия будут выполняться.

Теорема 2.1. Пусть $f(x)$ – выпуклая функция $x \in E_n$, $\lim\limits_{\|x\|\to\infty} f(x) = \infty$; Ω – множество точек минимума $f(x)$, и в процессе выполнения алгоритма ОГСРП применительно к этой функции выполняются следующие условия для $k = 0,1,2,...$:

$$\min_{x^* \in \Omega} \| A_k (x_k - x^*) \| < r \,;$$

$$\alpha_k \geq 1 + \varepsilon \,,$$

где r, ε – некоторые положительные константы.

Тогда $\lim_{k \to \infty} f(x_k) = \min_{x \in E_n} f(x)$

Теорема 2.2. Пусть в процессе реализации алгоритма ОГСРП выполняются следующие условия:

1/ $\min_{x^* \in \Omega} \| A_k (x_k - x^*) \| \leq r$, $k = 1,2,...$

2/ $1 < \alpha_k \leq \alpha^*$, $k = 1,2,...$

Тогда для некоторой бесконечной подпоследовательности индексов $\{k_i\}$, $i = 1,2,...$, $k_{i+1} > k_i$, выполняется неравенство:

$$f(x_{k_i}) - m^* \leq c(\prod_{j=1}^{k_i} \alpha_j)^{-\frac{1}{n}} \,;$$

(здесь $r > 0$, $c > 0$, $\alpha^* > 1$ – положительные числа, m^* – значение функции $f(x)$ в точке минимума).

Теорема 2.3. Пусть соблюдаются условия теоремы 2.2 и $\alpha_k = \alpha > 1$ ($k = 1,2,...$). Тогда для произвольного δ, $1 > \delta > 0$, найдется такое число c_δ, что для всех $k = 1,2,...$

$$\min_{0 \leq i \leq k} (f(x_i) - m^*) \leq c_\delta \alpha^{-\frac{k(1-\delta)}{n}} \,.$$

Затем приводятся конкретные способы выбора шаговых множителей и последовательностей коэффициентов расширения, при которых условия теорем 2.2 и 2.3 выполняются при определенных довольно слабых предположениях о поведении одномерных функций, получающихся вдоль лучей, проходящих через точку минимума.

Теорема 2.4. Пусть множество Ω точек минимума выпуклой функции $f(x)$ ограничено и $f(x)$ принимает на Ω значение m^*.

Введем обозначения:

$\rho(x)=\min_{x^*\in\Omega}\|x-x^*\|$, $x^*(x)$ – точка, для которой $\|x-x^*(\text{x})\|=\rho(\text{x})$.

Пусть для всех x, удовлетворяющих условию $\rho(\text{x})\le r$, $x\in\Omega$ и $x^*\in\Omega$ выполняется следующие неравенства:

$$N[f(x)-m^*]\le|g_f^{x-x^*(x)}(x)|\cdot\|x-x^{**}\|\le M[f(x)-m^*] \qquad (17)$$

где x^{**} – ближайшая к x точка минимума, лежащая на луче $y=x+t(x^*-x)$, $t\ge 0$.

Тогда, если при применении ОГСРП:

$$(1) \qquad \rho(x_0)\le r; \qquad (18)$$

$$(2) \qquad h_{k+1}=\frac{2MN}{M+N}\cdot\frac{f(x_k)-m^*}{\|g_{\varphi_k}(y_k)\|}; \qquad (19)$$

$$(3) \qquad \alpha_{k+1}\le\frac{M+N}{M-N}, \qquad (20)$$

то $\|A_{k+1}(x_{k+1}-x^*(x))\|\le\|A_k(x_k-x^*(x))\|$.

Здесь M, *N*, *r* – положительные константы.

Использование теоремы 2.4 становится возможным при априорном знании значения функции в точке минимума. В теореме 2.6 сформулированы признаки, позволяющие определять в процессе спуска меньше или больше выбранное нами число значения функции в точке минимума.

Теорема 2.6. Пусть выпуклая функция $f(x)$ обладает следующим свойством:

Существует постоянная $M>1$, такая, что если $f[(1-\alpha)x_1+\alpha x_2]$, ($0\le\alpha\le1$) строго убывает по α, то выполняется неравенство:

$$g_f^{x_1-x_2}(x_1)\cdot\|x_1-x_2\|\le M[f(x_1-f(x_2)]. \qquad (21)$$

Тогда, если при применении алгоритма ОНСРП с

$$\alpha_{k+1}=\frac{M+1}{M-1}, \qquad (22)$$

$$h_{k+1}=\frac{2M}{M+1}\cdot\frac{f(x_k)-\mathrm{m}}{\|g_{\varphi_k}(y_k)\|} \qquad (23)$$

m выбрано большим или равным m^*, то последовательность $\{h_k\}$ является ограниченной и для произвольного $\varepsilon>0$ найдется $\overline{k}$ такое, что $f(\overline{x})<m+\varepsilon$; (счет прекращается, если на некотором шаге $f(x_k)\le m$); если m выбрано меньше m^*, то последовательность $\{h_k\}$ является неограниченной.

Эта теорема служит основанием для предложенной далее модификации алгоритма ОГСРП, когда значение функции в точке минимума неизвестно.

Для многих задач минимизации (например, задач минимизации, возникающих при решении систем нелинейных уравнений и неравенств) требование выпуклости не является естественным. В связи с этим возникает необходимость распространения алгоритмов с растяжением пространства на более широкий класс функций.

Определение. Непрерывную функцию $F(x)$, $x\in E_n$, назовем π-дифференцируемой, если:

1/ в произвольной точке $x_0\in E_n$ эта функция имеет производные $F'_\eta(x_0)$ по любому направлению (здесь η, $\|\eta\|=1$ - вектор, задающий направление) ;

2/ в каждой точке x_0 существует по крайней мере один вектор $g^\pi(x_0)$ такой, что для любого η, $\|\eta\|=1$,

$$\min(F'_\eta(x_0),-F'_{-\eta}(x_0))\le(\mathrm{g}^\pi(x_0),\eta)\le\max(F'_\eta(x_0),-F'_{-\eta}(x_0)) \qquad (24)$$

Определение. Вектор $g^{\pi}(x_0)$, удовлетворяющий условию (24) для всех η, $\|\eta\|=1$, будет называться π-градиентом функции $F(x)$ в точке x_0. Показывается, что класс π-дифференцируемых функции включает в себя классы выпуклых и вогнутых функций и введенный Б.Н.Пшеничным в [9] класс квазидифференцируемих функций.

Для нахождения минимума π-дифференцируемых функций предлагаются алгоритмы π-градиентного спуска с растяжением пространства (π-ГСРП алгоритмы), описание которых почти полностью повторяет описание ОГСРП-алгоритмов. Изменения состоят лишь в том, что вместо обобщенных градиентов вычисляются π -градиенты.

Для π-ГСРП алгоритмов при известном m^* - значении функции в точке минимума - доказываются теоремы, аналогичные теоремам 2.1 - 2.6.

Затем описывается применение этих результатов к решению систем нелинейных уравнений :

$$\psi_i(x)=0,\ i=1,\ldots,n;\ x=(t_1,\ldots,t_2) \tag{25}$$

Для ускорения сходимости в малой окрестности решения метод π-ГСРП удобно сочетать при определенных условиях с методом ортогонадизации градиентов, который можно рассматривать как предельный вариант метода π-ГСРП.

Метод ортогонализации градиентов решения системы уравнении состоит из последовательности этапов, каждый из которых в свою очередь состоит из n последовательных шагов.

Перед началом каждого r-го этапа мы имеем начальное для этого этапа приближение $x_0^{(r)}$.

1 шаг. Вычисляем $\psi_1(x_0)$ и градиент $g_{\psi_1}(x_0)$, и находим

$$x_1^{(r)} = x_0^{(r)} - \frac{|\psi_1(x_0^{(r)}|}{\| g_{\psi_1}(x_0^{(r)}) \|^2} g_{\psi_1}(x_0^{(r)}) sign\psi_1(x_0^{(r)}) . \qquad (26)$$

<u>k+1 шаг.</u> Вычисляем $\psi_{k+1}^{(r)}(x_k)$ и градиент $g_{\psi_{k+1}}(x_k^{(r)})$, затем вектор $\varphi_{k+1}^{(r)}$ - проекцию вектора $g_{\psi_{k+1}}(x_k^{(r)})$ на подпространство, ортогональное векторам $g_{\psi_1}(x_0^{(r)}),..., g_{\psi_k}(x_{k-1}^{(r)})$ и находим

$$x_{k+1}^{(r)} = x_k^{(r)} = \frac{|\psi_{k+1}(x_k^{(r)}|}{\| \varphi_{k+1}^{(r)} \|^2} \varphi_{k+1}^{(r)} sign\psi_{k+1}(x_k^{(r)}) . \qquad (27)$$

Подученное таким образом $x_n^{(r)}$ служит начальным приближением для следующего этапа:

$$x_n^{(r)} = x_0^{(r+1)} .$$

Теорема 2.11. Пусть система уравнений (25) имеет решение $x^* = 0$. Функции $\psi_i(x)$, $i = 1,...,n$, непрерывно дифференцируемы и их градиенты $g_{\psi_i}(x)$ удовлетворяют условию Липшица в некоторой области $S_\delta = \{x : \| x \| < \delta\}$:

$$\| g_{\psi_i}(x) - g_{\psi_i}(x') \| \le L \| x - x' \| . \qquad (28)$$

Кроме того, $\{g_{\psi_i}(0)\}$, $i = 1,...,\mathrm{n}$, образуют систему линейных независимых векторов, так что якобиан $\det I(x) = \det \left\{ \frac{\partial \psi_i}{\partial t_i}(x) \right\}$, $i, j = 1,...,\mathrm{n}$, отличен от 0 в точке 0.

Тогда найдется такое $\varepsilon > 0$, что если $\| x_0^{(r)} \| \le \varepsilon$, то $\| x_n^{(r)} \| = \| x_0^{(r+1)} \| \le k | x_0^{(r)} \|^2$, где k – некоторое положительное число.

В алгоритмах с растяжением пространства сложной проблемой является вопрос выбора шаговых множителей $\{h_k\}$. При известном m^* этот вопрос решается с помощью формул вида (19). Если m^*

неизвестно, то выбор h_k может осуществляться для гладких функций из условия минимума $f(x)$ по направлению, как в методе наискорейшего спуска. Однако сочетание метода наискорейшего спуска с растяжением пространства в направлении градиента не дает эффекта ускорения сходимости. Оказалось, что для получения эффекта ускорения сходимости метода наискорейшего спуска растяжение пространства целесообразно производить в направлении разности градиентов, вычисленных в двух последовательных точках минимизирующей последовательности. Далее приводится описание класса алгоритмов с растяжением пространства в направлении разности двух последовательных градиентов (или их аналогов).

Пусть $f(x)$ – непрерывно дифференцируемая функция, $x \in E_n$, $\lim\limits_{\|x\| \to \infty} f(x) = \infty$. Рассмотрим следующий алгоритм минимизации $f(x)$ (r -алгоритм)

Вначале имеем: $x_0 \in E_n$; $B_0 = E$; $\tilde{g}_0 = 0$.

После k -го шага: x_k; B_k; $\tilde{g}_k$, где B_k -матрица размерности $n \times n$, $x_k \in E_n$, $\tilde{g}_k \in E_n$.

Опишем ($k+1$)-й шаг:

1/ $g_f(x_k)$;

2/ $B_k^* g_f(x_k) = g_k^*$;

3/ $g_k^* - \tilde{g}_k = r_k$;

4/ $\xi_{k=1} = \dfrac{r_k}{\| r_k \|}$;

5/ $B_{k+1} = B_k \cdot R_{\beta_{n+1}}(\xi_{k+1})$;

6/ $\tilde{g}_{k+1} = R_{\beta_{n+1}}(\xi_{k+1}) g_k^* = B_{k+1}^* g_f(x_k)$;

7/ $x_{k+1} = x_k - h_{k+1} B_{k+1} \tilde{g}_{k+1}$.

Приведённое выше описание определяет класс итеративных процессов вычисления последовательности x_k, $k = 1, 2, \ldots$. Конкретный процесс получается при фиксации x_0 и алгоритмов вычисления последовательностей $\{\beta_k\}$ и $\{h_k\}$.

Рассматриваются также модификации r-алгоритмов, в которых после определенного числа шагов матрица B_k "восстанавливается" , то есть заменяется матрицей *E*.

Для частного алгоритма $\beta_k = 0$ и с "восстановлением" матрицы B_k после каждых *n* итераций доказывается следующая

Теорема 2.12. Пусть функция $f(x)$ дважды непрерывно дифференцируема в некоторой окрестности S точки x^* , причем в этой окрестности матрица вторых производных (гессиан) $H(x)$ удовлетворяет условию Липшица.

$$\| H(x) - H(x^{'}) \| \leq L \| x - x^{'} \|; \; x, x^{'} \in S\,. \tag{29}$$

Кроме того,

$$g(x^{"}) = 0\,, \tag{30}$$

$H(x^{'})$-положительно определенная матрица.

Тогда найдется такая окрестность точки x^* $S^{'} \leq S$, что если $x_0 \in S$, $\| x_0 - x^* \| = R_0$, то задается такое число k, не зависящее от x_0, что $\| x_n - x^* \| \leq kR_0^2$, где x_n - точка, полученная после *n* шагов работы выше приведенного алгоритма, *n* – размерность пространства, в котором определена $f(x)$; $f(x_{k+1}) = \min\limits_{h_{k+1}}(x_k - h_{k+1} B_{k+1} \tilde{g}_{k+1})$.

r-алгоритмы с выбором шагового множителя, как в методе наискорейшего спуска, применительно к гладким функциям близки по своим свойствам к методам сопряженных градиентов [17] . Однако, в отличие от последних они могут применяться с небольшими модификациями для минимизации недифференцируемых функций.

Суть модификации состоит в том, что h_k выбирается несколько большим, чем при условии минимума в данном направлении. Экспериментальные расчеты по минимизации недифференцируемых функций показали хорошую эффективность r -алгоритмов.

Третья глава диссертации состоит из нескольких разделов, каждый из которых охватывает определенный круг задач математического программирования. Для решения этих задач используются методы, которые были исследованы в первых двух главах, или их модификации.

В первом разделе рассмотрены вопросы применения предложенных в диссертации алгоритмов минимизации к решению общей задачи выпуклого программирования:

$$\text{найти } \min f_0(x) \tag{31}$$

при ограничениях:

$$f_i(x) \le 0\,;\; x \in E_n\,;\; i = 1, 2, \ldots, m\,. \tag{32}$$

Если ограничения удовлетворяют условию Слейтера [20] то задача (31) – (32) при использовании метода штрафных функций [21] сводится к задаче отыскания безусловного минимума выпуклой функции. Рассмотрены вопросы применения алгоритмов ОГС, ОГСРП и r-алгоритмов для решения этой задачи, а также для решения минимаксных задач, сводящихся к задачам выпуклого программирования.

Следующий раздел посвящен применению метода ОГС к решению задач большого объема с использованием схем разложения [21].

Для задач выпуклого программирования рассмотрены две основные схемы разложения: декомпозиция множества переменных и декомпозиция множества ограничивающих неравенств.

Пусть заданы выпуклые функции $f_\nu(x, y)$, $\nu = 1, 2, \ldots,$ зависящие от двух групп переменных и сформулирована задача выпуклого программирования:

найти

$$\min f_0(x, y) \tag{33}$$

при ограничениях

$$f_i(x, y) \le 0\,;\ i = 1, 2, ..., n\,,\ x \in E_e\,;\ y \in E_m \tag{34}$$

Зафиксируем $x = \bar{x}$ и рассмотрим задачу нахождения

$$\min_{y \in E_m} f_0(\bar{x}, y) \tag{35}$$

при ограничении

$$f_i(\bar{x}, y) \le 0\,. \tag{36}$$

Если при произвольном x задача ((35), (36)) имеет решение и неравенства (36) удовлетворяют условию Слейтера, то обозначив $\Phi(\bar{x})$ оптимальное значение $f_0(\bar{x}, y)$, получим сведение задачи (33)-(34) к задаче минимизации $\Phi(\bar{x})$. Это и есть схема декомпозиции множества переменных. Основой для применения метода ОГС к решению этой задачи является следующая

Лемма 3.1. Если $f_\alpha(x, y)$, $\alpha = 0, 1, ..., n$ – выпуклые по совокупности векторных переменных x, y функции, то функция $\Phi(x)$ является выпуклой на некотором выпуклом множестве $W \le E_e$. Если для некоторого $x \in W$ выполняется условие Слейтера, то обобщенный градиент функции $\Phi(x)$ в точке x вычисляется по формуле:

$$g_\Phi(\bar{x}) = \mathrm{g}^x_{L_u}(\bar{x}, y(\bar{x}))\,,$$

где $L_u(x, y) = L(x, y, u) = f_0(x, y) + \sum_{i=1}^{n} u_i f_i(x, y)$;

$y(\bar{x})$ – одно из оптимальних значений y в задаче (35)-(36), $u = \{u_i\}$ – множители Лагранжа задачи (35)-(36), полученные в соответствии с теоремой Куна-Такера, $\mathrm{g}^x_{L_u}(\bar{x}, y(\bar{x}))$ - проекция такого обобщенного градиента функции $L_u(z) = L_u(x, y)$ на подпространство $E_e^{(x)}$, у которого проекция на подпространство $E_m^{(y)}$ равна 0 (обобщенный градиент берется в точке $\bar{z} = (\bar{x}, y(\bar{x}))$.

С использованием леммы 3.1 мы получаем алгоритм ОГС минимизации $\Phi(x)$, дающий возможность получать оптимальное решение задачи (33)-(34) в схеме декомпозиции множества переменных.

Подробно рассматриваются алгоритмы решения задач линейного программирования с использованием этой схемы разложения и примеры практических задач, для решения которых использовались эти алгоритмы.

Затем исследуется схема декомпозиции множества ограничений, в некотором смысле двойственная предыдущей схеме, для задач выпуклого программирования. Предлагается алгоритм с использованием ОГС, дающий возможность получать в рамках этой схемы разложения значения множителей Лагранжа задачи (33) – (34).

Рассмотрена также специальная схема разложения для решения задач линейного программирования, матрица которых не содержит отрицательных элементов.

Подробно рассматривается важный в практическом отношении класс задач линейного программирования - задачи распределительного типа. После общей формулировки задач распределительного типа (прямой и двойственной) рассмотрена схема разложения по множеству переменных в пространстве двойственных переменных. Обсуждается вопрос получения оптимального плана прямой задачи по решению двойственной, предлагаются специальные алгоритмы. Затем приводятся примеры решения важных задач отраслевого планирования, сводящихся к задачам распределительного типа. Результаты многих решенных задач были использованы планирующими органами со значительным экономическим эффектом.

Стандартные программы для ЭВМ "М-220" были разработаны ст. инженером ИК АН УССР Г.И.Горбач [23].

Следующий раздел посвящен методам решения задач многоэтапного стохастического программирования. Эти задачи являются удобной математической формализацией большого числа проблем, возникающих при решении вопросов перспективного планирования, многостадийного проектирования, управления сложными протяженными во времени процессами в условиях неопределенности. Первоначально такого рода задачи были сформулированы Р.Беллманом на языке динамического программирования [24], однако методы динамического программирования позволили решать лишь простейшие задачи такого типа.

В настоящее время многоэтапные стохастические задачи принятия решений рассматриваются в рамках стохастического программирования. Особенно много работ посвящено так называемой двухэтапной задаче линейного стохастического программирования ([25], [26], [27] и др.). В этих работах предлагаются методы, основанные на сведении двухэтапной задачи к специальной задаче линейного программирования большого объема, при этом возникают серьёзные вычислительные трудности.

В диссертации описывается предложенный для решения задач двухэтапного стохастического программирования в совместной статье автора и Ю.М.Ермольева [28] метод обобщенного стохастического градиента, являющийся стохастическим аналогом метода ОГС. Этот метод позволяет эффективно решать достаточно сложные задачи двухэтапного стохастического программирования. Далее в диссертации предлагается новый подход к формализации задач многоэтапного стохастического программирования и рассматриваются возможности применения для решения этих задач методом ОГС и обобщенных стохастических градиентов.

В последнем разделе третьей главы описываются результаты численных экспериментов по решению ряда задач минимизации методами градиентного типа с использованием операции растяжении пространства.

В качестве примеров гладких функции для минимизации брались, в основном, функции, которые встречались в других работах по методам оптимизации и обладают ярко выраженными овражными особенностями. Кроме того, приведено несколько примеров минимизации недифференцируемых функций также с овражными особенностями.

Результаты расчетов показывают существенное ускорение сходимости, вызванное использованием операции растяжения пространства и хорошо согласуются с теоретическими результатами гл.II. Стандартные программы для различных модификаций методов градиентного типа с растяжением пространства, с помощью которых проведены эти расчеты, были разработаны ст. инженером ИК АН УССР В.И.Билецким для ЭВМ М-220 [29], аспиранткой Днепропетровского Госуниверситета Л.П.Шабашовой для ЭВМ "Минск-22" и студентом-дипломантом МФТИ Н.Г.Журбенко для ЭВМ "БЭСМ-6".

В заключении кратко сформулированы основные результаты работы и рассматриваются некоторые проблемы минимизации недифференцируемых функций, еще ждущие своего решения.

Перечислим основные результаты автора, отраженные в диссертации.

1. Предложен общий метод минимизации выпуклых (не обязательно дифференцируемых) функций – метод обобщенного градиентного спуска. Для различных способов регулировки шага доказаны теоремы о сходимости и теоремы, определяющие скорость сходимости.

2. Предложен новый общий класс методов градиентного типа с растяжением пространства для ускорения сходимости при минимизации функции, не обязательно дифференцируемых. Проведено теоретическое исследование скорости сходимости некоторых конкретных алгоритмов из этого класса.

3. Подробно исследованы возможности применения этих алгоритмов к решению систем нелинейных уравнений.

4. Доказаны теоремы о квадратической скорости сходимости некоторых предельных вариантов методов с растяжением пространства.

5. Проведено исследование применимости предложенных алгоритмов к решению общих задач математического программирования и задач минимаксного типа.

6. Предложены новые алгоритмы решения линейных и нелинейных задач при использовании схем декомпозиции.

7. Подробно исследованы итеративные алгоритмы решения задач распределительного типа, имеющие важные практические применения.

8. Предложены новые формулировки и методы решения задач многоэтапного стохастического программирования. Проведено исследование метода обобщенных стохастических градиентов применительно к задачам этого типа, в частности, для двухэтапных задач стохастического программирования.

9. Под руководством автора разработаны алгоритмы и стандартные программы решения важных практических задач оптимального планирования, проектирования и исследования операций с использованием методов, описанных в диссертации.

Основные результаты, нашедшие отражение в диссертации, докладывались на Международном математическом конгрессе в г.Москве, на Всесоюзной конференции по математическому программированию г. Новосибирск, 1969г.), на 1-ой и 2-ой зимних школах по математическому программированию и смежным вопросам (г.Дрогобыч, 1968, 1969 гг.), на 3-ей Всесоюзной школе по методам оптимизации (г.Тирасполь,1969 г.) и на ряде других совещаний и семинаров союзного и республиканского значения, приведены в работах [1–2], [5–6], [23], [28–44].

ЛИТЕРАТУРА

1. Н.З. ШОР. Применение метода градиентного спуска для решения сетевой транспортной задачи. “Материалы научного семинара по теоретическим и прикладным вопросам кибернетики и исследования операций”, 1962, вып.1
2. Н.З. ШОР. О структуре алгоритмов численного решения задач оптимального планирования и проектирования, Диссертация, 1964.
3. Б.Т. ПОЛЯК. Один общий решения экстремальных задач. ДАН СССР, т. 174, №1, 1967.
4. Ю.М. ЕРМОЛЬЕВ. Методы решения нелинейных экстремальных задач, ж. "Кибернетика", №4,1966.
5. Н.З. ШОР. О скорости сходимости обобщенного градиентного спуска, ж. "Кибернетика", №8, 1968.
6. Н.З. ШОР. Обобщенный градиентный спуск, Труды первой Зимней школы по математическому программированию, г. Дрогобыч, вып.3, Москва,1969.
7. А.Я. ДУБОВИЦКИЙ, А.А. МИЛЮТИН. Задачи на экстремум при наличии ограничений. Журнал вычислительной математики и математической физики,1965, т.5, №3.
8. Б.Н. ПШЕНИЧНЫЙ. Выпуклое программирование в нормированных пространствах. Кибернетика, №5, 1965.
9. Б.Н. ПШЕНИЧНЫЙ. Необходимые условия экстремума. Изд."Наука", М., 1969.
10. Г. БУЗЕМАН. Выпуклые поверхности. Изд."Наука", М., 1964.

11. Л.В. КАНТОРОВИЧ. Об одном эффективном методе решения экстремальных задач для квадратичного функционала, ДАН СССР, №7 ,48, 1945.
12. W. DAVIDON. Variance algorithm for minimization. The computer journal, 1968, vol.10, No.4.
13. R. FLETCHER, M. POWELL. A rapidly convergent descent method for minimization, Comp. journ, vol.6 (1963).
14. Э.Л. АКИМ, Т.М. ЭНЕЕВ. Определение параметров движения космического летательного аппарата по данным траекторных измерений, Космические исследования, т.1, №1 (1963).
15. Б.Н. ПШЕНИЧНЫЙ. Об одном алгоритме спуска. Журнал выч.мат. и матем.физики, т.8, №3 (1968).
16. J.W. DANIEL. Convergence of the conjugate gradient method with computationally convinient modifications, Numer. Math., No.2, 1967.
17. Б.Т. ПОЛЯК. Метод сопряженных градиентов в задачах на экстремум. Журн. выч. матем. и матем.физики, №4, (1969).
18. И.М. ГЕЛЬФАНД, М.Л. ЦЕТЛИН. Принцип нелокального поиска в задачах автоматической оптимизации, ДАН СССР, 137, №2 (1961).
19. Х. РОЗЕНБРОК, С. СТОРИ. Вычислительные методы для инженеров-химиков. ИзД."Мир", М.,1968.
20. Г. ЗОЙТЕНДЕЙК. Методы возможных направлений, ИЛ, М.,1963.
21. Е.Г. ГОЛЬШТЕЙН, Д.Б. ЮДИН. Новые направления в линейном программировании, изд. "Советское радио", М.,1966.
22. И.И. ЕРЕМИН. О методе "штрафов" в выпуклом программировании, ж. "Кибернетика", №4, К.,1967.

23. Н.З. ШОР, Г.И. ГОРБАЧ. Решение задач распределительного типа методом обобщенного градиентного спуска, Труды семинара НС АН УССР но кибернетике "Теория оптимальных решений", №1, 1968.
24. Р. БЕЛЛМАН. Динамическое программирование, ИЛ, М., 1960.
25. G.B. DANTZIG. Linear programming under uncertainty. Manag. Sci., 1, No.2 (1955).
26. G. DANTZIG, A. MADANSKY. On the solution of two-stage linear programs under uncertainty. Proc of the 4-th Berkley Symposium on Mathem. St.and Probability, 1961.
27. R.VAN SLYKE, R.WETS. L-Shaped linear programs with applications to optimal control and stochastic programming. SIAM J. on Appl. Math., July, 1969.
28. Ю.М. ЕРМОЛЬЕВ, Н.З. ШОР. Метод случайного поиска для задач двухэтапного стохастического программирования, ж."Кибернетика", №1, Киев, 1968.
29. Н.З. ШОР, В.И. БИЛЕЦКИЙ. Метод растяжения пространства для ускорения сходимости в задачах овражного типа. Труды семинара НС АН УССР по кибернетике "Теория оптимальных решений", вып.2, Киев, 1969.
30. А.А. БАКАЕВ, В.С. МИХАЛЕВИЧ, С.В. БРАНОВИЦКАЯ, Н.З. ШОР. "Методика и опыт решения сетевых транспортных задач большого объема на ЭЦВМ", Сб. "Математические методы и проблемы производства", 1963.
31. А.А. БАКАЕВ, Н.И. РОСИНА, Н.З. ШОР. Задачи планирования грузопотоков на транспортной сети, Сб. МИИТ, изд, "Транспорт",

вып. 237, 1967.

32. Н.З. ШОР, Н.И. РОСИНА. Схема разложения задач линейного и выпуклого программирования и ее применение для решения задач планирования перевозок, Сб. докладов 1 Всесоюзной конференции по оптимизации и моделированию транспортных сетей, Киев,1967.
33. Н.З. ШОР, М.Б. ЩЕПАКИН. Об одной задаче стохастического программирования. Труды семинара НС по кибернетике АН УССР "Теория оптимальных решений", вып.2, Киев, 1967.
34. Н.З. ШОР, М.Б. ЩЕПАКИН. Алгоритм решения двухэтапной задачи стохастического программирования, ж."Кибернетика", Киев, 1968, №3.
35. Ю.М. ЕРМОЛЬЕВ, Н.З. ШОР. О минимизации недиффереци-руемых функций, ж. "Кибернетика", Киев, 1967.
36. Н.З. ШОР. Применение обобщенного градиентного спуска в блочном программировании, ж, "Кибернетика", №3, Киев, 1967.
37. В.С. МИХАЛЕВИЧ, Ю.М. ЕРМОЛЬЕВ, В.В. ШКУРБА, Н.З. ШОР. Сложные системы и решение экстремальных задач, ж. "Кибернетика", №5,Киев,1967.
38. Н.З. ШОР, М.Б. ЩЕПАКИН. Задачи многоэтапного стохастического программирования в параметрической форме. Труды семинара НС АН УССР по кибернетике "Теория оптимальных решений", №8, Киев, 1969.
39. Н.З. ШОР, Л.В. ИВАНОВА. Об одном итеративном методе решения задач линейного программирования и матричных игр, Труды семинара НС АН УССР по кибернетике "Теория оптимальных решений", №3, Киев, 1969.

40. Н.З. ШОР. Использование операций растяжения пространства в задачах минимизации выпуклых функций, ж. "Кибернетика", №1, К., 1970.
41. Н.З. ШОР. О скорости сходимости метода обобщенного градиентного спуска с растяжением пространства, ж. "Кибернетика" №2, К.,1970.
42. Н.З. ШОР, М.Б. ЩЕПАКИН. О некоторых подходах к решению задач многоэтапного стохастического программирования, Труды II зимней школы по математическому программированию, г. Дрогобыч, (1969), Москва, 1970.
43. Н.З. ШОР. Многоэтапное стохастическое выпуклое программирование. Труды семинара НС по кибернетике АН УССР "Теория оптимальных решений", вып.I, К., 1967.
44. Н.З. ШОР, М.Б. ЩЕПАКИН, С.Я. АТУТОВА. Алгоритм решения задачи многоэтапного линейного стохастического программирования. Труды семинара НС по кибернетике АН УССР "Теория оптимальных решений", вып.1, К., 1967

Н.З. ШОР

МЕТОДЫ МИНИМИЗАЦИИ НЕДИФФЕРЕНЦИРУЕМЫХ ФУНКЦИЙ И ИХ ПРИЛОЖЕНИЯ

(Специальность 008 – вычислительная математика)

А в т о р е ф е р а т

Киев-*1970*

БФ 08271. Подписано к печати 6.V 1970 г. Изд. № 2-182. Зак. Л68
Объем 1,8 учетн.-изд.л. Тираж 180 экз.

Лаборатория офсетной печати ИК АН УССР
Киев-28, проспект Науки, 109.