

том переоценки на начало 1950 г. пересчитывались основные фонды на 1.I.1950 г., среднегодовыми общими коэффициентами переоценки — выбытие и сальдо передач, общими коэффициентами пересчета — ввод.

Достоинство этой схемы заключается в том, что появляется возможность сравнения результатов. В самом деле путем прямой переоценки получается база (1.I.1950 г.), а в дальнейшем добавляются (соответственно вычитаются) пересчитанные приросты, сальдо балансовых передач и выбытия. В конце концов, мы подходим к 1960 г. и сравниваем объем основных фондов на 1.I.1960 г., полученный в результате проделанного пересчета с официальной цифрой переоценки на 1.I.1960 г. Этим методом были пересчитаны основные промышленно-производственные фонды на 13 предприятиях (металлургических, машиностроительных, текстильных и лесозаготовительных). Сравнение расчетной цифры основных фондов на 1.I.1960 г. с официальной цифрой переоценки на 1.I.1960 г., принимаемой за 100, дало следующие размеры ошибок: до  $\pm 3\%$  у семи предприятий, от  $\pm 3$  до  $\pm 5\%$  у двух предприятий, от  $\pm 5$  до  $\pm 7\%$  у двух предприятий,  $-8,1\%$  на одном предприятии и  $+11,3\%$  — на одном предприятии.

Эти данные показывают, что несмотря на все упрощения и условности точность пересчета оказалась достаточно приемлемой, а кроме того, косвенно свидетельствуют о небольшой, по-видимому, погрешности в пересчете отраслей промышленности.

В 1964 г. в статистическом сборнике «Промышленность СССР» ЦСУ СССР опубликовало динамику пересчета по отраслям промышленности, что дало возможность сравнения.

Абсолютные величины индексов 1959 г. по нашему пересчету отличаются от индексов ЦСУ СССР (принимая последние за 100) следующим образом: по черной металлургии —  $4\%$ , топливной промышленности —  $1\%$ , электроэнергетике —  $5\%$ , машиностроению и металлообработке  $+1\%$ , лесной, бумажной и деревообрабатывающей —  $1\%$ , промышленности строительных материалов —  $11\%$ , химической —  $5\%$ , легкой —  $2\%$ , пищевой —  $2\%$ .

О разнице в динамике основных фондов по всей промышленности не приходится говорить, поскольку абсолютные величины основных фондов промышленности в целом в разработке ЦСУ СССР, содержащей отраслевую расшифровку и использованной для пересчета и разработки, по данным которой ЦСУ СССР публикует динамику основных фондов промышленности в целом, несопоставимы.

Произведенный пересчет явился в сущности первой работой такого рода; тем более, что сейчас заканчивается второй пересчет основных промышленно-производственных фондов по большему количеству отраслей промышленности. Хочется надеяться, что деловое обсуждение изложенных принципов и методов поможет дальнейшему совершенствованию цифровой базы нашей экономической науки.

Поступила в редакцию  
24 II 1967

## МЕТОДЫ МИНИМИЗАЦИИ ФУНКЦИЙ МНОГИХ ПЕРЕМЕННЫХ \*

(Обзор)

Б. Т. ПОЛЯК

(Москва)

### 1. Введение

Отыскание экстремума (всюду в дальнейшем для определенности минимума) функции  $n$  переменных без ограничений, вероятно, одна из старейших задач прикладной математики. Более того это — одна из важнейших задач. Вот лишь несколько примеров чисто математических проблем, приводящих к конечномерной задаче на безусловный минимум:

а) оценка неизвестных параметров по экспериментальным данным методом наименьших квадратов или из принципа максимального правдоподобия. В частности, такую задачу часто приходится решать в математической экономике при определении параметров математической модели по статистическим данным прошлых лет;

б) решение системы алгебраических или трансцендентных уравнений эквивалентно минимизации невязки этих уравнений. Так, решение линейной системы  $Ax = b$  с положительно определенной матрицей  $A$  эквивалентно минимизации квадратичной функции  $f(x) = \frac{1}{2}(Ax, x) - (b, x)$ , а с произвольной матрицей  $A$  — функции  $f(x) = \|Ax - b\|^2$ ;

в) вариационные задачи (т. е. бесконечномерные задачи минимизации) при решении методом Рунге сводятся к последовательности конечномерных;

г) задачи оптимального управления путем применения принципа максимума Понтрягина сводятся к подбору начальных значений сопряженной системы, минимизирующих отклонение от заданной точки на правом конце траектории;

д) решая задачи минимизации при наличии ограничений методом штрафных функций, приходим к последовательности задач на безусловный экстремум.

Число примеров можно было бы многократно умножить за счет прикладных физических, технических и экономических задач. Если раньше в таких задачах математика применялась в основном для расчета тех или иных состояний, режимов и т. д., то теперь все чаще возникает задача оптимизации некоторого критерия — функции цели — по тем или иным управляемым переменным. Среди всего многообразия задач оптимизации две можно считать основными: задачу линейного программирования (т. е. минимизацию конечномерной линейной функции при линейных ограничениях) и задачу минимизации функции  $n$  переменных без ограничений. Первая из этих задач, будучи на несколько веков моложе второй (линейное программирование возникло лишь в 1939 г.), в настоящее время решительно обгоняет вторую по числу посвященных ей работ.

Так, при наличии многих книг по линейному программированию на русском языке нет ни одной обзорной статьи по методам минимизации

\* В основу статьи положен обзорный доклад автора на конференции по математическому оптимальному программированию (Новосибирск, 31 мая — 4 июня 1965 г.).

нелинейной функции без ограничений. Может быть, это происходит отчасти потому, что некоторые математики считают последнюю задачу слишком легкой. В самом деле, здесь нет никаких сложных математических проблем. Однако все, кто сталкивался на практике с задачами минимизации (а такие задачи приходится решать десяткам организаций), знают, насколько они сложны с вычислительной точки зрения. Простейшие, широко известные методы минимизации обычно совершенно непригодны для реальных задач: в частности, они сходятся слишком медленно и часто практически останавливаются вдали от точки минимума. Поэтому задача разработки методов минимизации отнюдь не может считаться решенной.

В настоящей статье сделана попытка дать по возможности полный обзор известных методов минимизации, а также оценить их с вычислительной точки зрения. При этом основное внимание уделяется методам, пригодным для достаточно точного отыскания минимума гладких функций, имеющих единственную точку минимума (локальные минимумы отсутствуют) и являющихся выпуклыми в окрестности этой точки.

Минимизация негладких функций, отыскание абсолютного минимума и специальные методы, пригодные для решения узких классов задач, — эти вопросы изложены менее подробно. Наконец, стохастические задачи (т. е. минимизация функций, значения которых вычисляются со случайной ошибкой) и комбинаторные задачи (в которых переменные принимают конечное множество значений) вообще не рассматриваются.

Оценка методов дается на основании практического опыта расчетов в Вычислительном центре МГУ, а также многочисленных бесед с лицами, занимающимися решением задач минимизации. Всем им автор глубоко признателен. Во многих случаях, однако, практический опыт решения слишком мал (а иногда и полностью отсутствует), так что оценка является умозрительной и в достаточной степени субъективной.

## 2. КЛАССИФИКАЦИЯ И КРИТЕРИИ ОЦЕНКИ МЕТОДОВ

Методы минимизации являются в большей части итерационными, т. е. в них строится последовательность приближений  $x^0, x^1, \dots, x^k, \dots$ , сходящаяся к точке минимума. Впрочем, в некоторых частных случаях (например для минимизации квадратичной функции) возможны и точные методы, т. е. такие, в которых решение находится за конечное число шагов. Кроме того, иногда рассматривают непрерывные аналоги итерационных методов, в которых строится не дискретная последовательность  $x^0, x^1, \dots, x^k, \dots$ , а траектория  $x(t)$ ,  $0 \leq t \leq \infty$ , описываемая системой обыкновенных дифференциальных уравнений. Практически каждый итерационный метод имеет свой непрерывный аналог. Такие непрерывные методы могут использоваться при решении задачи минимизации на аналоговых машинах. Среди специфических непрерывных методов существует ряд весьма интересных (см., например, [1, 2]). Если же вычисления ведутся на цифровой машине, то интегрирование обыкновенных дифференциальных уравнений производится каким-либо конечно-разностным методом, что эквивалентно построению итерационной последовательности. По этой причине не будем рассматривать непрерывных методов. Наконец, в ряде методов (см. раздел 5) строится не последовательность точек, а последовательность вложенных областей, которым принадлежит точка минимума.

Итерационный метод называется  $m$ -шаговым, если при построении очередной итерации используется  $m$  предыдущих. Большинство методов является одношаговыми, однако нам встретятся и двухшаговые методы (иногда и нульшаговые), а также методы, в которых используются все предыдущие итерации.

Далее можно делить методы на монотонные и немонотонные. Монотонные методы (к ним принадлежит большинство известных методов) — это такие методы, в которых значение функции монотонно убывает от итерации к итерации (иногда их называют релаксационными).

С вычислительной точки зрения наиболее важны две характеристики методов минимизации гладких функций. Это, во-первых, так называемый порядок метода, равный наивысшему порядку производных функции  $f(x)$ , участвующих в вычислениях. Например, во многих методах используются лишь значения функции  $f(x)$ , но не ее производных. Это методы нулевого порядка. В градиентных методах требуется вычисление производных  $f(x)$ , т. е. они являются методами первого порядка. Метод Ньютона, в котором используется матрица вторых производных, имеет порядок два. Методы более высокого порядка для многомерных задач практически не применяются.

Остановимся на другой важной характеристике метода. Во всяком методе при построении следующей итерации исходит из того или иного предположения о поведении функции  $f(x)$ . Так, в градиентном методе функция по существу локально аппроксимируется линейной — такие методы будем называть линейными. В методе сопряженных градиентов и Ньютона функция аппроксимируется квадратичной. Такие методы — будем называть их квадратичными — дают решение за конечное число шагов для случая квадратичной функции.

Наконец, в данной статье будем рассматривать только детерминированные методы. Нам кажется, что внесение элемента случайности обычно лишь ухудшает эффективность метода и что существующее иногда мнение [3] о преимуществах случайного поиска для многомерных задач минимизации является ошибочным (доказательство этого см. в [4]). За недостатком места, однако, мы здесь не можем обсудить этот вопрос подробнее.

Каковы критерии, по которым можно сравнивать различные методы? Это прежде всего объем вычислений, требующихся для отыскания решения с заданной точностью. Эта величина зависит от числа итераций, требуемых для этой цели, и от трудоемкости вычислений на каждую итерацию. Последнюю величину часто можно выразить числом вычислений значения функции на одну итерацию (Островский [5] предложил специальное название «горнер» для этой единицы). Правда, не всегда можно оценить, например, сложность вычисления градиента (или матрицы вторых производных) в «горнерах». В самом деле, если мы умеем вычислять лишь значения функции в точке, то для этой цели требуется  $n + 1$  горнеров ( $n$  — число переменных). Если же функция задана посредством формул, которые можно явно продифференцировать, то вычисление градиента часто лишь немногим более трудоемко, чем вычисление значения функции. В частности, вычисление значения квадратичной функции  $Ax, x$  требует даже большего числа действий, чем вычисление ее градиента  $2Ax$ .

Кроме этого основного критерия, при сравнении методов приходится принимать в расчет и другие соображения: область сходимости, устойчивость по отношению к погрешностям, требуемый объем памяти машины, сложность программирования и т. д.

### 3. ЛИНЕЙНЫЕ МЕТОДЫ

Все описываемые здесь методы применимы для достаточно гладких функций (имеющих непрерывный градиент) и сходятся к стационарной точке функции при любом начальном приближении (практически к локальному минимуму). Более подробно мы не будем рассматривать вопросы сходимости, отсылая читателей к цитируемой ниже литературе.

Типичным для данной группы методов является градиентный. Мы рассмотрим сначала классический вариант метода, а затем его различные модификации.

Введем векторно-матричные обозначения, которых будем всюду придерживаться. Пусть  $x$  —  $n$ -мерный вектор с координатами  $x_1, \dots, x_n$ ;  $f(x)$  — минимизируемая функция  $n$  переменных;  $f'(x)$  — ее градиент, т. е. вектор с координатами  $\partial f / \partial x_1, \dots, \partial f / \partial x_n$ ,  $\|x\| = (x, x)^{1/2}$ , где  $(x, y) = x_1 y_1 + \dots + x_n y_n$  — скалярное произведение векторов  $x$  и  $y$ ;  $f''(x)$  — матрица  $n \times n$  с элементами  $\partial^2 y / \partial x_i \partial x_j$ . Через  $y = Ax$  обозначим умножение матрицы  $A$  на вектор  $x$ , т. е.  $y_i = \sum_{j=1}^n a_{ij} x_j$ ,  $i = 1, 2, \dots, n$ . Греческие буквы  $\alpha, \beta, \dots$  будут применяться для обозначения числовых величин.

В этих обозначениях градиентный метод имеет следующий вид:

$$x^{k+1} = x^k - \alpha_k f'(x^k). \quad (1)$$

Здесь  $x^k$  —  $k$ -е приближение (начальное приближение —  $x^0$ ),  $\alpha_k \geq 0$ . В координатной записи тот же метод принимает вид

$$x_i^{k+1} = x_i^k - \alpha_k \frac{\partial f(x^k)}{\partial x_i}, \quad i = 1, 2, \dots, n. \quad (2)$$

Иначе говоря, в этом методе делается шаг по антиградиенту  $-f'(x)$ , т. е. по направлению скорейшего убывания функции  $f(x)$  в линейном приближении. В теории рассматриваются обычно два варианта выбора длины шага  $\alpha_k$ :

а) простой градиентный метод, в котором

$$\alpha_k = \alpha > 0; \quad (3)$$

б) метод скорейшего спуска, в котором  $\alpha_k$  выбирается из условия минимума  $f(x)$  по направлению  $-f'(x^k)$ , т. е.

$$f(x^k - \alpha_k f'(x^k)) = \min_{\alpha \geq 0} f(x^k - \alpha f'(x^k)). \quad (4)$$

Метод скорейшего спуска впервые рассматривал Коши (1830 г.). В более поздних работах [6—11] доказывается сходимость градиентного метода при различных предположениях, а также рассматривается ряд модификаций метода. Доказано, в частности, что для непрерывно дифференцируемых функций, растущих на бесконечности, для обоих вариантов метода (в первом для достаточно малого  $\alpha$ ) при любом начальном приближении  $f'(x^k) \rightarrow 0$ . Если точка минимума — единственная стационарная точка, то к ней сходится последовательность  $x^k$ .

При практических вычислениях обычно применяются следующие методы выбора длины шага:

а) ведется счет с постоянным  $\alpha_k \equiv \alpha$ . При этом на каждом шаге проверяется условие монотонности  $f(x^{k+1}) \leq f(x^k)$ . Если оно оказывается нарушенным, то  $\alpha$  дробится (например вдвое) до тех пор, пока монотонность не восстановится. Время от времени можно пробовать увеличивать  $\alpha$ . И. М. Глазман [12] отметил, что такой метод, вообще говоря, может и не сходиться, и предложил более сложную процедуру, обеспечивающую сходимость. Однако при практических вычислениях нужда в таком усложнении никогда не возникала;

б) приближенно реализуется метод скорейшего спуска. Обычно одномерная задача на минимум решается путем построения квадратичной

аппроксимации по значениям функции в трех точках на выбранном направлении. Точка минимума параболы берется в качестве  $x^{k+1}$ . Более точное отыскание минимума по направлению обычно нецелесообразно;

в) иногда применяют градиентный метод с постоянным шагом, т. е.  $\alpha_k = \lambda / \|f'(x^k)\|$ . Иначе говоря, длина шага равна  $\lambda = \text{const}$ . Эта длина дробится при нарушении условия монотонности.

Сравним эти три метода. В последнем из них каждый шаг вблизи минимума сопровождается дроблением, т. е. приходится несколько раз вычислять значение функции. В методе б) также на каждом шагу примерно 3 раза вычисляется значение функции. В методе же а) такое вычисление производится обычно лишь одно (величина  $\alpha$  не имеет тенденции к дроблению: при достаточно малом  $\alpha$  условие монотонности не нарушается). С точки зрения скорости сходимости все три метода примерно одинаковы (оценка скорости сходимости для а) и б) для квадратичных (и близких к ним) функций одинакова). Итак, метод а) во многих случаях наиболее эффективный.

Выясним теперь, какова скорость сходимости градиентного метода и от чего она зависит. Пусть функция  $f(x)$  имеет точку минимума  $x^*$  и в этой точке существует положительно определенная матрица  $f''(x^*)$ . Тогда, оказывается, в окрестности  $x^*$  оба варианта градиентного метода (3), (4) сходятся со скоростью геометрической прогрессии со знаменателем  $q = (M - m) / (M + m)$ , где  $M$  — наибольшее, а  $m$  — наименьшее собственное значение матрицы  $f''(x^*)$  (см. [7, 10, 11]). Обозначим  $p = M / m$  ( $p$  называется числом обусловленности матрицы  $f''(x^*)$ ), тогда  $q = (p - 1) / (p + 1) \approx 1 - (2 / p)$  для больших  $p$ . Таким образом, если  $p$  близко к единице, что соответствует случаю, когда  $f(x)$  меняется примерно одинаково по всем направлениям вблизи минимума, то  $q$  мало, т. е. градиентный метод сходится быстро вблизи минимума. Если же собственные значения матрицы  $f''(x^*)$  сильно различаются, то  $p$  велико, а  $q$  близко к единице, т. е. скорость сходимости мала. В этом случае в окрестности минимума  $f(x)$  сильно меняется по одним направлениям (отвечающим собственным векторам матрицы  $f''(x^*)$ , для которых собственные значения велики) и слабо меняется по другим (отвечающим малым собственным значениям). В терминологии И. М. Гельфанда и М. Л. Цетлина [13] первые направления называются несущественными, вторые существенными, а сами функции такого типа — хорошо организованными. Линии уровня таких функций сильно вытянуты (рис. 1), они имеют характерный «овраженный» вид.

Результаты последовательных итераций метода скорейшего спуска для квадратичной функции двух переменных такого типа показаны на рис. 2. Наглядно видны причины медленной сходимости: направление антиградиента сильно отличается от направления к точке минимума.

Оказывается, в подавляющем большинстве прикладных задач встречаются именно функции овражного вида, для которых градиентный метод сходится чрезвычайно медленно. Способам ускорения сходимости в этом случае посвящена значительная часть статьи, в частности, весь раздел 4.

Прежде чем их описывать, остановимся на методах координатного спуска, весьма близких к градиентному. В этих методах спуск из очередной точки производится по направлению одной из координатных осей. Иногда (особенно в линейной алгебре — см. [14, гл. 3]) эти методы называются релаксационными, а различные их варианты — методами Гаусса — Зейделя, Саусвелла и т. д. Сходимость некоторых координатных методов доказана в [15, 16]. Последовательность, в которой выбираются координатные оси, может быть различной. Обычно они берутся в фиксированном циклическом порядке (чаще всего поочередно). Иногда выбирается та ось, для

которой величина  $|\partial f / \partial x_i|$  максимальна. Последний способ вряд ли целесообразен, так как, во-первых, в каждой точке выполняется много лишних вычислений для определения всех частных производных, а, во-вторых, при таком методе некоторые координаты могут практически никогда не попасть в число выбранных.

Что касается способов выбора длины шага, то они такие же, как и в градиентном методе, причем применимы те же соображения об их эффек-

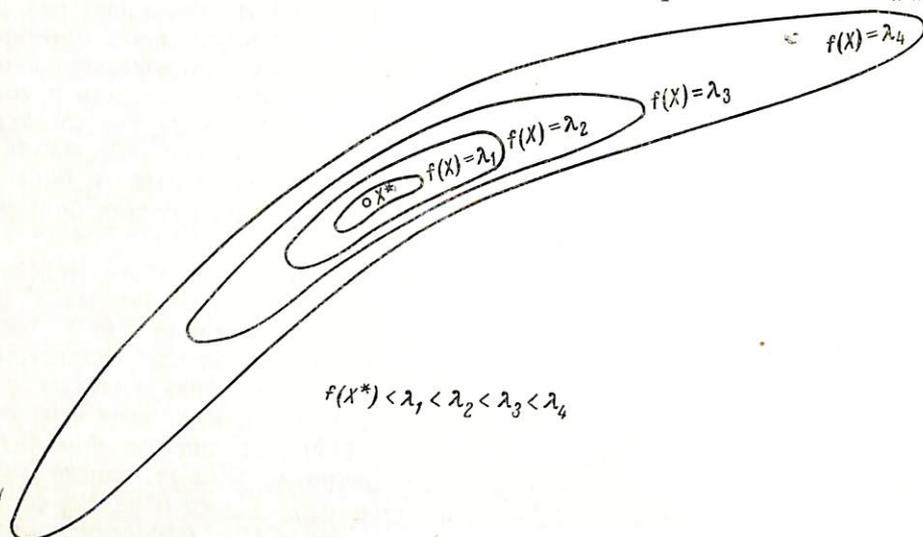


Рис. 1

тивности. Наиболее употребителен (и обычно наиболее целесообразен) метод покоординатного скорейшего спуска, в котором приближенно (путем квадратичной аппроксимации по трем точкам) ищется минимум  $f(x)$  по очередной координате.

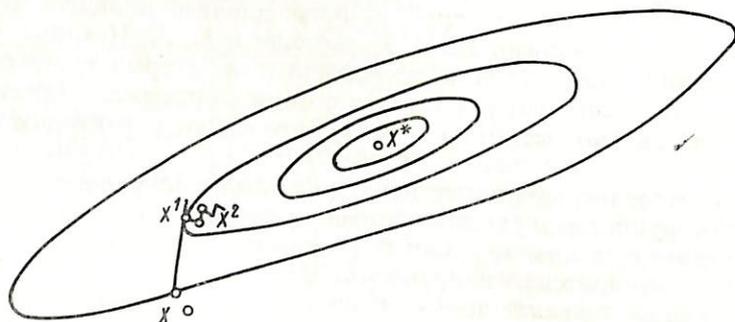


Рис. 2

Метод покоординатного скорейшего спуска нулевого порядка, т. е. в нем используются лишь значения  $f(x)$  в точках. Существует и несколько других методов нулевого порядка, но некоординатного типа. Так, в [17] (см. также [18, стр. 178—183]) описано так называемое симплекс-планирование. Выбирается  $n + 1$  точка  $x^0, x^1, \dots, x^n$ , образующая правильный симплекс. Вычисляются  $f(x^0), f(x^1), \dots, f(x^n)$  и выбирается  $j$ , для которого  $f(x^j)$  максимально. Строится новый симплекс, отличающийся от старого лишь одной вершиной:  $x^j$  заменяется на  $x^{n+1}$ :

$$x^{n+1} = \frac{2}{n}(x^0 + \dots + x^{j-1} + x^{j+1} + \dots + x^n) - x^j$$

(т. е.  $x^{n+1}$  симметрично с  $x^j$  относительно грани, противоположащей  $x^j$ ). Если окажется, что в новом симплексе максимум достигается в  $x^{n+1}$ , то возвращаемся к исходному симплексу, заменив  $x^j$  на вершину, в которой значение  $f(x)$  максимально среди оставшихся вершин. Наконец, если какая-либо точка сохраняется в  $n + 1$ -м последовательном симплексе, то строится новый симплекс вокруг этой точки с вдвое меньшим ребром. Отметим еще, что в работе [19] описана модификация симплексного метода, в которой симплекс не является обязательно правильным, а величина шага и условия дробления симплекса несколько иные. Сходные идеи лежат в основе «метода аффинного градиента» С. С. Лазрова, описание которого можно найти в [20]. Сходимость всех этих методов не доказана, а эффективность их неясна.

Возвращаясь к методу покоординатного скорейшего спуска, отметим, что он обладает в ряде случаев серьезным преимуществом по сравнению с градиентным методом. А именно, для функций, в которых «существенные» и «несущественные» направления близки к координатным, он сходится гораздо быстрее. В частности, для квадратичной функции вида

$$f(x) = \sum_{i=1}^n \gamma_i (x_i - \beta_i)^2 \quad \text{координатный метод дает точный минимум за}$$

$n$  шагов, тогда как градиентные методы могут сходиться очень медленно (если  $\max \gamma_i / \min \gamma_i$  велико).

Мы опишем сейчас модификацию градиентного метода — градиентный метод с подбором масштабов, свободный от этого недостатка. Как отмечалось выше, градиентный метод сходится тем быстрее, чем ближе линии уровня  $f(x)$  к сферам. Сделаем такое преобразование координат  $x'_i = \lambda_i x_i$ , которое уменьшит вытянутость линий уровня. Для этого вычислим в какой-либо точке величины  $\partial^2 f / \partial x_i^2$  (для квадратичной функции они не зависят от точки) и выберем  $\lambda_i = \sqrt{\partial^2 f / \partial x_i^2}$ . При таком преобразовании ли-

нии уровня функции  $f(x) = \sum_{i=1}^n \gamma_i (x_i - \beta_i)^2$   $\gamma_i > 0$ , станут сферами, а для

функций общего вида станут менее вытянутыми. Поэтому в новых координатах  $x'_i$  градиентный метод будет сходиться быстрее. В старых координатах метод примет вид

$$x_i^{k+1} = x_i^k - \alpha_k \lambda_i \frac{\partial f(x^k)}{\partial x_i}. \quad (5)$$

Один шаг такого метода с точки зрения скорости сходимости эквивалентен примерно  $n$  шагам координатного метода. В то же время  $n$  шагов покоординатного скорейшего спуска требуют примерно  $3n$  вычислений  $f(x)$ , тогда как в методе (5) нужно лишь немногим более  $n$  вычислений. Величины  $\lambda_i$  удобно подсчитывать в процессе покоординатного скорейшего спуска. В самом деле  $\partial^2 f / \partial x_i^2 \approx \Delta_i f / (\Delta x_i)^2$ , где  $\Delta_i f$ ,  $\Delta x_i$  — приращение  $f(x)$  и  $x_i$  при спуске по  $i$ -й координате.

Таким образом, наиболее целесообразной является следующая комбинация методов. Сначала применяются масштабные множители  $\lambda_i$ . Когда эти в процессе которого вычисляются, переходим на градиентный метод в форме (5). При этом время от времени  $\lambda_i$  нужно пересчитывать. Такая процедура часто дает очень значительное ускорение сходимости по сравнению с простым градиентным методом. Численные результаты для одной конкретной задачи можно найти в [21].

Отметим еще несколько простых способов ускорения сходимости. Их общая идея — использовать результаты предыдущих итераций для отыскания «дна оврага», т. е. направлений, по которым функция изменяется медленнее всего.

Наиболее известны такие приемы для задач линейной алгебры, т. е. для минимизации квадратичной функции [14, гл. IX]. Однако они могут быть перенесены и на общий случай.

Так, в методе Л. А. Люстерника рекомендуется в градиентном методе с постоянным шагом вычислять отношение  $\|f'(x^k)\| / \|f'(x^{k-1})\|$ . Когда оно стабилизируется (к некоторому  $q < 1$ ,  $q \approx 1$ ), нужно сделать большой шаг по градиенту, а именно

$$x^{k+1} = x^k - \frac{q}{1-q} f'(x^k), \quad (6)$$

независимо от возрастания  $f(x)$ , а затем продолжать спуск градиентным методом.

В методе А. А. Абрамова предлагается сделать шаг по направлению, задаваемому точками  $x^k$  и  $x^{k+2}$  в градиентном методе.

В методе Франкела каждый следующий шаг делается по направлению, являющемуся линейной комбинацией антиградиента в очередной точке и предыдущего направления движения, т. е.

$$x^{k+1} = x^k - \alpha f'(x^k) + \beta(x^k - x^{k-1}). \quad (7)$$

Этот метод естественно назвать «методом тяжелого шарика», так как (7) описывает движение тела, обладающего инерцией, на которое действует сила  $f(x)$  при наличии трения. Сходимость этого метода и оценка скорости сходимости для неквадратичных задач исследованы в [22]. Величина  $\alpha$  примерно та же, что и в градиентном методе, а  $\beta$  подбирается экспериментально ( $\beta < 1$ ,  $\beta$  близко к 1). Отметим, что метод (7) не монотонный. Метод такого типа применялся в [23].

В «овражном методе» И. М. Гельфанда и М. Л. Цетлина [13] предлагается провести градиентный спуск (до его заметного замедления) из двух начальных точек, а затем сделать большой шаг по направлению, задаваемому этими «спущенными» точками. Еще один подобный метод описан в [24].

Все эти приемы имеют много близких черт. Они немногим сложнее градиентного метода и построены на его основе. Они дают некоторое ускорение сходимости, тем большее, чем меньше размерность «дна оврага». Однако и они часто сходятся слишком медленно, так что для точного отыскания минимума приходится применять более мощные (но и более трудоемкие) методы квадратичной аппроксимации, описываемые в разделе 4.

Подведем итог. 1) Линейные методы сходятся для широкого класса функций и для любых начальных приближений. 2) Объем вычислений на одну итерацию не слишком велик — порядка  $n$  вычислений значения  $f(x)$ . 3) Эти методы сходятся тем медленнее, чем более вытянуты линии уровня функции. 4) Простые способы ускорения сходимости в ряде случаев дают эффект, но не всегда достаточный.

#### 4. КВАДРАТИЧНЫЕ МЕТОДЫ

Описанные выше линейные методы и способы их ускорения обычно дают возможность получить значение функции, достаточно близкое к минимальному. Если же целью является отыскание достаточно точного приближения для самой точки минимума, то при их помощи это не всегда

удается. Автору известно много случаев, когда при минимизации функции совсем небольшого числа переменных исследователи приходили к выводам о наличии большого числа локальных минимумов, так как спуск из различных начальных точек градиентным методом давал совсем разные результаты. В действительности же минимум был один, но его не могли отыскать достаточно точно.

Квадратичные методы позволяют найти точку минимума гораздо более точно. Основная идея этих методов, заключающаяся в том, что вблизи минимума нужно аппроксимировать функцию квадратичной, а затем найти точку минимума этой квадратичной формы, была впервые высказана, по-видимому, статистиками — Боксом и другими — в их работах по «планированию экспериментов», т. е. по проблеме минимизации при наличии случайных ошибок (см., например, [18]).

Рассмотрим различные варианты методов, отличающиеся прежде всего количеством используемых при вычислениях производных функций  $f(x)$ .

**Методы второго порядка.** Пусть функция  $f(x)$  имеет вторые производные  $\partial^2 f / \partial x_i \partial x_j$ ,  $i, j = 1, 2, \dots, n$ . Матрицу  $n \times n$  с этими элементами обозначим  $f''(x)$ . Аппроксимируем функцию  $f(x)$  в окрестности точки  $x^k$  первыми тремя членами ряда Тейлора

$$f(x) \approx f(x^k) + (f'(x^k), x - x^k) + \frac{1}{2} (f''(x^k) (x - x^k), x - x^k)$$

и найдем  $x^{k+1}$  — точку минимума этой квадратичной функции. Получим:

$$x^{k+1} = x^k - [f''(x^k)]^{-1} f'(x^k). \quad (8)$$

Иначе говоря, для получения следующего приближения нужно в точке  $x^k$  вычислить  $f'(x^k)$  и  $f''(x^k)$  и решить систему линейных уравнений  $f''(x^k)z = f'(x^k)$ . Если обозначить ее решение через  $z^k$ , то  $x^{k+1} = x^k - z^k$ . Отметим, что метод (8) полностью совпадает с методом Ньютона для решения нелинейных уравнений, примененным к системе уравнений  $f'(x) = 0$ , поэтому и (8) будем называть методом Ньютона для минимизации  $f(x)$ . Таким образом, к (8) можно применить известные теоремы о сходимости метода Ньютона, полученные Л. В. Канторовичем и другими авторами [25, гл. XVIII]. Для задачи минимизации  $f(x)$  эти результаты принимают следующий вид. Пусть  $f(x)$  дважды непрерывно дифференцируема в окрестности точки минимума  $x^*$ , причем матрица  $f''(x^*)$  положительно определена. Тогда для всех  $x^0$ , достаточно близких к  $x^*$ , метод (8) сходится к  $x^*$  с квадратичной скоростью (т. е.  $\|x^{k+1} - x^*\| \leq \gamma \|x^k - x^*\|^2$ ). Таким образом, метод (8) в отличие от градиентных методов сходится лишь при наличии достаточно хорошего начального приближения (для  $f(x)$  выпукла), причем скорость сходимости очень высока и не зависит от овражного характера  $f(x)$  (т. е. от обусловленности  $f''(x^*)$ ). Разумеется, для квадратичной функции  $f(x)$  метод (8) даст решение за один шаг. Все это, как увидим, характерно и для других квадратичных методов.

Основная вычислительная работа на каждом шаге метода Ньютона связана с вычислением матрицы вторых производных  $f''(x^k)$ . Метод целесообразно применять, лишь если эта матрица может быть вычислена явно (например, если  $f(x)$  задается формулами, которые можно продифференцировать). Конечноразностная же аппроксимация  $f''(x)$  нецелесообразна, так как ниже будут описаны другие квадратичные методы, не использующие  $f''(x^k)$ . Кроме того, на каждом шаге метода нужно решать систему линейных уравнений  $f''(x^k)z = f'(x^k)$ . Это удобно делать методом квадратного корня [14, стр. 165—168]. В самом деле этот метод очень экономичен для систем с симметричной матрицей (каковой является  $f''(x^k)$ ), а кроме

того, при решении этим методом автоматически проверяется, является ли матрица положительно-определенной. Если  $f''(x^k)$  не такова, то метод Ньютона расходится.

Чтобы не вычислять  $f''(x^k)$  на каждом шаге, иногда применяют модифицированный метод Ньютона, в котором  $f''(x)$  вычисляется лишь в начальной точке

$$x^{k+1} = x^k - [f''(x^0)]^{-1}f'(x^k). \quad (9)$$

Разумеется, такое упрощение покупается за счет более медленной сходимости: метод (9) сходится лишь со скоростью геометрической прогрессии. Иногда применяют и комбинацию методов, в которой матрица  $f''(x)$  вычисляется один раз за несколько итераций.

Как уже отмечалось, метод Ньютона сходится лишь для очень хорошего начального приближения. Существует несколько модификаций метода, которые сходятся в более широкой области. Так, в [26] предлагается следующая комбинация градиентного метода и метода Ньютона:

$$x^{k+1} = x^k - (a_k I + f''(x^k))^{-1}f'(x^k).$$

Здесь  $I$  — единичная матрица, а величина  $a_k$  выбирается специально в зависимости от результатов предыдущих итераций, причем  $a_k \rightarrow 0$  при приближении к точке минимума.

В работе Л. В. Канторовича [27] предложен метод

$$x^{k+1} = x^k - a_k [f''(x^k)]^{-1}f'(x^k), \quad (10)$$

где  $0 < a_k \leq 1$ . Л. В. Канторович ссылается на А. В. Граве, применявшего этот метод в одномерном случае. Чем меньше  $a_k$ , тем больше область сходимости, однако, тем меньше скорость сходимости. В ряде работ [28—31] предлагаются различные способы выбора коэффициентов  $a_k$ . В частности, можно выбирать  $a_k$ , как в методе скорейшего спуска.

**Методы первого порядка.** В этом разделе будут рассмотрены методы, которые дают точное решение задачи на минимум для случая квадратичной функции за конечное число шагов, но в которых используется лишь градиент функции, а не матрица вторых производных.

Наиболее прямолинейный метод такого рода является полным аналогом метода Ньютона, в котором вычисление  $f''(x)$  производится при помощи конечных разностей. Сходимость такого метода доказана в [32]. Однако этот метод чрезвычайно трудоемкий, так как на каждом шаге приходится вычислять значение градиента в  $n + 1$ -й точке  $x^k, x^k + \varepsilon e_1, \dots, x^k + \varepsilon e_n$  ( $e_1, \dots, e_n$  — координатные орты;  $\varepsilon$  — шаг разностной схемы).

Значительно более экономным является метод, предложенный Ф. Вольфом [33]. Он заключается в следующем. Выбирается  $n + 1$  базисная точка  $x^0, x^1, \dots, x^n$  и в них вычисляется  $f'(x^0), f'(x^1), \dots, f'(x^n)$ . Далее ищутся

такие коэффициенты  $\lambda_0, \lambda_1, \dots, \lambda_n$ ,  $\sum_{i=0}^n \lambda_i = 1$ , что линейная комбинация градиентов с этими коэффициентами обращается в нуль. Для этого

нужно решить систему  $n + 1$  линейных уравнений с  $n + 1$  переменными  $\lambda_0, \lambda_1, \dots, \lambda_n$ :

$$\sum_{i=0}^n \lambda_i f'(x^i) = 0, \quad \sum_{i=0}^n \lambda_i = 1. \quad (11)$$

Полученные значения коэффициентов нужно взять в качестве барицентрических координат новой точки  $x^{n+1}$ :

$$x^{n+1} = \sum_{i=0}^n \lambda_i x^i. \quad (12)$$

Нетрудно видеть, что, если  $f(x)$  — положительно определенная квадратичная функция, то  $x^{n+1}$  будет ее точкой минимума. В общем же случае  $x^{n+1}$  вводится в систему базисных точек, а одна из старых базисных точек (обычно  $x^0$  или та, в которой  $f(x)$  максимально) выводится из нее. Для выбранной таким образом  $n+1$ -й базисной точки расчет повторяется. Отметим, что для  $n=1$  этот метод совпадает с методом секущих для решения уравнения  $f'(x) = 0$ . Для  $n=2$  такое обобщение метода секущих, как указывает А. М. Островский [5, приложение IV], предложил еще Гаусс.

Таким образом, в методе Вольфа требуется  $n+1$  вычисление градиента для начала расчетов (если предварительно применялся градиентный метод, то можно взять последние точки в этом методе) и всего лишь одно дополнительное вычисление на каждом шаге, т. е. существенно меньше, чем в конечноразностном варианте метода Ньютона. Однако метод Вольфа обладает одной неприятной особенностью. Дело в том, что если базисные точки окажутся лежащими в подпространстве меньшей размерности, то система (11) будет вырождена. Однако в процессе счета базисные точки действительно имеют тенденцию к вырождению. Для борьбы с вырождением в [34] предложена довольно сложная процедура, в которой число используемых точек значительно больше, чем в методе Вольфа. Однако наш опыт практических расчетов по методу Вольфа не подтвердил опасности вырождения. По-видимому, дело заключается в следующем. Базисные точки, как правило, растягиваются по «дну оврага». Таким образом, по существенным направлениям квадратичная аппроксимация, которая строится в методе Вольфа, достаточно точна. Что же касается несущественных направлений, то по ним  $f(x)$  изменяется столь сильно, что достаточно небольшого различия в этих переменных для построения довольно точной квадратичной аппроксимации.

Сходимость метода Вольфа не исследована. По-видимому, он сходится примерно при тех же условиях, что и метод Ньютона. Чтобы расширить область сходимости метода, мы применяли следующие приемы. Во-первых, желательнее предотвратить слишком большие шаги ( $|\lambda_i| \geq 20$ ), то делалось, что  $\lambda_i$  (система (11)) слишком велики (так, чтобы барицентрические координаты новой точки удовлетворяли условию  $|\lambda_i| \leq 20$ ). Кроме того, во вновь полученной точке всегда проверялось условие  $f(x^{n+1}) \leq \max_{0 \leq i \leq n} f(x^i)$ .

Если оно оказывалось невыполненным, то из точки  $x^{n+1}$  производился спуск градиентным методом до выполнения этого условия. Эта «спущенная» точка и вводилась в базисную систему.

Опишем еще некоторые квадратичные методы первого порядка.

Некоторые варианты метода Вольфа имеются в [35].

В методе Давидона [36] строится последовательность

$$x^{h+1} = x^h - a_h H_h f'(x^h).$$

Здесь  $H_h$  — матрица, которая пересчитывается на каждом шаге с использованием  $H_{h-1}$  и нового градиента. Для квадратичной функции  $f(x) = (Ax, x) - (b, x)$  через  $n$  шагов оказывается  $H_n = A^{-1}$ . Мы не приводим точных формул метода, так как они весьма громоздки.

Опишем еще один очень интересный метод — метод сопряженных градиентов. Этот метод, предложенный Хестенсом и Штифелем для задач линейной алгебры, дает точное решение задачи минимизации квадратичной формы  $f(x) = 1/2(Ax, x) - (b, x)$  за  $n$  шагов (см., например, [14]).

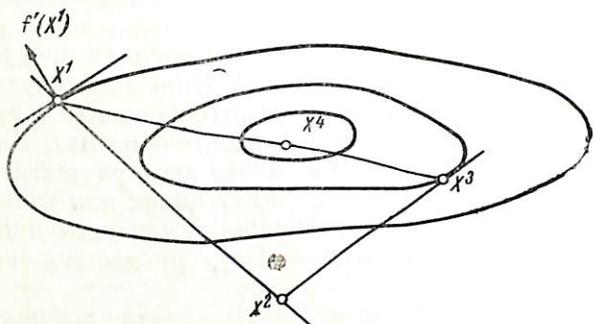


Рис. 3

Формулы метода для этого случая таковы (приводим ту их форму, которая нам наиболее удобна):

$$\begin{aligned} x^{k+1} &= x^k - \alpha_k p^k, \\ p^0 &= Ax^0 - b, \\ p^k &= Ax^k - b + \beta_k p^{k-1}, \\ \alpha_k &= \frac{\|Ax^k - b\|^2}{(Ap^k, p^k)}, \\ \beta_k &= \frac{\|Ax^k - b\|^2}{\|Ax^{k-1} - b\|^2}. \end{aligned}$$

Флетчер и Ривз [37] заметили, что, поскольку  $Ax - b = f'(x)$ , этот метод можно записать в таком виде:

$$\begin{aligned} x^{k+1} &= x^k - \alpha_k p^k, \quad p^0 = f'(x^0), \quad p^k = f'(x^k) + \beta_k p^{k-1}, \\ f(x^k - \alpha_k p^k) &= \min_{\alpha \geq 0} f(x^k - \alpha p^k), \quad \beta_k = \frac{\|f'(x^k)\|^2}{\|f'(x^{k-1})\|^2}. \end{aligned} \quad (13)$$

В такой форме метод применим и для неквадратичных функций (хотя его сходимость пока не доказана). Метод сопряженных градиентов весьма похож на метод «тяжелого шарика» (7). В самом деле, (13) можно переписать иначе:

$$x^{k+1} = x^k - \alpha_k f'(x^k) + \frac{\alpha_k \beta_k}{\alpha_{k-1}} (x^k - x^{k-1}), \quad (14)$$

что только способом выбора коэффициентов отличается от (7). Однако в отличие от метода (7) метод (13) (или (14)) дает точное решение задачи на минимум квадратичной функции за конечное число шагов. Для неквадратичных функций скорость сходимости метода сопряженных градиентов, как показали эксперименты, значительно выше, чем метода тяжелого шарика. Метод сопряженных градиентов для неквадратичных функций целесообразно время от времени «обновлять»: после ряда итераций (порядка  $n$ ) в точке  $x^k$  вновь выбирать  $p^k = f'(x^k)$ .

Наконец, еще один квадратичный метод первого порядка — метод параллельных касательных — предложен в работе [38]. Идея этого метода для  $n = 2$  видна из рис. 3. Здесь точки  $x^1, x^2$  произвольны; прямая  $x^2x^3$  параллельна касательной к линии уровня в точке  $x^1$  (т. е. ортогональна  $f'(x^1)$ );  $x^3$  — точка минимума  $f(x)$  на прямой  $x^2x^3$ ;  $x^4$  — точка минимума  $f(x)$  на прямой  $x^1x^4$ . Для случая квадратичной функции двух переменных  $x^4$  есть решение задачи.

Сравним теперь различные методы, описанные в этом разделе. Наиболее эффективными представляются метод Вольфа (с описанными выше модификациями) и метод сопряженных градиентов, которые мы и будем сравнивать. И в том, и в другом методе каждый новый шаг требует однократного вычисления градиента. В методе Вольфа требуется существенно больший объем памяти: нужно запомнить  $n$  предыдущих градиентов, а также решать систему линейных уравнений (14), тогда как в методе

сопряженных градиентов достаточно запоминать лишь предыдущий вектор  $p^{k-1}$ . Метод Вольфа (даже с описанными выше модификациями) сходится лишь в достаточно малой окрестности минимума, тогда как метод сопряженных градиентов в силу своей монотонности сходится практически всегда (вдали от минимума он работает примерно как градиентный метод). Единственным соображением в пользу метода Вольфа является возможность (недостаточно проверенная экспериментально) несколько более быстрой сходимости.

**Методы нулевого порядка.** Перейдем к описанию методов, дающих решение задачи минимизации квадратичной функции за конечное число шагов, но не требующих вычисления  $f'(x)$  или  $f''(x)$  (а лишь вычисления  $f(x)$ ). Этот случай особенно важен, так как в большинстве практических задач мы имеем возможность только вычислять  $f(x)$  в любой точке.

Разумеется, для каждого из описанных выше методов нулевого порядка, второго порядка можно построить аналогичные методы нулевого порядка, заменяя соответствующие производные конечными разностями (см., например, [39], [40]). Приводимые ниже методы являются, однако, более экономными по сравнению с таким прямолинейным подходом.

Начнем с метода «барицентрических координат» С. С. Лаврова [20], являющегося некоторым аналогом метода Вольфа. Выбирается  $n + 1$  базисная точка  $x^0, x^1, \dots, x^n$ . Затем в этих точках вычисляются значения функции  $f(x)$ ; обозначим  $f(x^i) = f_i$ . То же проделывается для середин всех отрезков, соединяющих пары базисных точек; обозначим  $f((x^i + x^j) / 2) = f_{ij}$ ,  $f_{ii} = f_i$ . После этого решается система линейных (относительно  $\lambda, \lambda_0, \lambda_1, \dots, \lambda_n$ ) уравнений

$$\sum_{j=0}^n f_{ij} \lambda_j + \lambda = f_{ii}, \quad i = 0, 1, \dots, n; \quad \sum_{j=0}^n \lambda_j = 1 \quad (15)$$

и строится точка  $x^{n+1} = \sum_{i=0}^{n+1} \lambda_i x^i$ . Нетрудно видеть, что, если  $f(x)$  — квадратичная функция, то точка  $x^{n+1}$  будет являться ее точкой минимума. Если же  $f(x)$  — неквадратичная функция, то процесс продолжается: одна из базисных точек (обычно  $x^0$  или та, для которой  $f(x)$  максимально) заменяется на  $x^{n+1}$ , вновь вычисляются  $f_{ij}$  (достаточно это сделать для  $i = n + 1$ , т. е. вычислить  $f(x)$  всего лишь в  $n + 1$ -й точке, так как остальные  $f_{ij}$  сохраняются), решается новая система (15) (которая отличается от предыдущей лишь одной строкой матрицы  $f_{ij}$ , что может быть эффективно использовано при решении) и т. д. По поводу сходимости этого метода можно повторить все, что было сказано о методе Вольфа: необходимость теоретического обоснования, опасность вырождения, его отсутствие при практических вычислениях, требование хорошего начального приближения. Для расширения области сходимости можно применять те же модификации, что и для метода Вольфа: ограничение величин  $|\lambda_i|$ , проверка условия монотонности для  $f(x^{n+1})$  и т. д. Отметим еще, что для начального базиса имеет смысл проверять условие выпуклости  $f_{ij} \leq 1/2(f_i + f_j)$ . Если оно не выполнено для какой-либо пары  $i, j$ , то метод применять нельзя — он будет расходиться.

Все остальные известные методы нулевого порядка [41—44], сводят задачу минимизации  $f(x)$  к последовательности одномерных минимизаций. Существует довольно много вариантов таких методов, отличающихся лишь способом выбора последовательности направлений, по которым производится минимизация. Опишем лишь один из них [41, 42], довольно типич-

ный и достаточно простой (назовем его методом последовательных минимизаций). Выбираются начальная точка  $x^0$  и линейно независимые направления  $p^1, \dots, p^n$  (обычно просто координатные орты). Затем последовательно ищутся точки  $x^1, x^2, \dots, x^{n+1}$ , где  $x^i = x^{i-1} + \alpha_i p^i$ ,  $i = 1, 2, \dots, n$ ,  $x^{n+1} = x^n + \alpha_{n+1} p^0$ , а  $\alpha_i$  выбирается из условия минимума  $f(x^i)$ . После этого строится  $p^{n+1} = x^{n+1} - x^1$ ,  $p^1$  заменяется на  $p^{n+1}$ ,  $x^1$  на  $x^{n+1}$  и весь цикл повторяется. Можно доказать, что через  $n$  таких циклов для случая квадратичной функции мы получим точку минимума. Для неквадратичных функций процесс является бесконечным, причем время от времени целесообразно «обновлять» систему направлений.

Отметим еще методы [45, 46], которые также сводятся к последовательной минимизации по специально подбираемым направлениям, но не дают точного минимума для квадратичной функции.

Сравним теперь методы барицентрических координат и последовательных минимизаций. С точки зрения требуемого объема памяти эти методы примерно одинаковы: в первом нужно запоминать симметричную матрицу  $f_{ij}$ , а во втором — систему направлений  $p^1, \dots, p^n$ . Для отыскания минимума квадратичной функции в первом методе требуется  $[n(n+1)/2]$  вычислений значения  $f(x)$ , а во втором  $\sim 3n^2$  (считая, что минимизация по направлению требует вычисления  $f(x)$  в трех точках). Однако первый метод сходится лишь для достаточно хорошего начального приближения (правда, описанные модификации расширяют область сходимости), тогда как второй — практически всегда. Итак, метод барицентрических координат по сравнению с методом последовательных минимизаций обычно дает выигрыш по времени счета, но имеет более узкую область сходимости.

Подытожим теперь все сказанное о квадратичных методах. 1) Квадратичные методы применимы для достаточно гладких функций (дважды непрерывно дифференцируемых) при наличии хорошего начального приближения. 2) Они дают точный минимум для квадратичной функции за конечное число шагов и сходятся тем быстрее, чем ближе функция к квадратичной (независимо от обусловленности  $f''(x^*)$ ). 3) Объем вычислений в них весьма значительный: около  $n^2$  вычислений значения  $f(x)$  для квадратичного случая.

В связи с этим квадратичные методы целесообразно применять в комбинации с линейными. Сначала для далекого приближения применяется тот или иной линейный метод, который дает вначале быстрое уменьшение функционала и не требует большой вычислительной работы. При попадании в квадратичную окрестность минимума он сильно замедляется, однако здесь уже обычно оказываются применимыми квадратичные методы, более трудоемкие, но дающие более быструю сходимость. Примеры, подтверждающие эффективность такого подхода, можно найти в [21, 47, 48]. Некоторые результаты по численному сравнению методов имеются в [49].

## 5. МИНИМИЗАЦИЯ НЕГЛАДКИХ ФУНКЦИЙ

Рассмотрим здесь лишь задачу минимизации непрерывной выпуклой, но вообще говоря недифференцируемой функции. Случай разрывных функций и непрерывных, но не выпуклых функций не будет затронут. Отметим прежде всего, что отсутствие производной делает невозможным применение почти всех описанных выше методов. Так, для функции двух переменных  $f(x) = 10|x_1 - x_2| + |x_1 + x_2|$  при любом  $x_1 = x_2$  метод по координатного спуска не работает (по любому из координатных направлений функция возрастает), однако точка минимума единственна  $x_1 = x_2 = 0$ . Градиентный метод или любые методы, использующие гра-



точке, то опорный в нем единственный и совпадает с градиентом. В одномерном случае опорной является прямая, проходящая через точку  $\{x, f(x)\}$  и лежащая всюду ниже кривой  $y = f(x)$  (рис. 4). Существует ряд простых правил вычисления опорного функционала. Так, если  $f(x) = \max_{1 \leq i \leq N} \{g_i(x)\}$ ,  $g_i(x)$  — выпуклые дифференцируемые функции, то  $c = \sum \lambda_i g_i'(x)$ ,  $\lambda_i \geq 0$ ,  $\sum \lambda_i = 1$ , где сумма берется по тем  $i$ , для которых в точке  $x$   $g_i(x) = f(x)$ . В частности, можно взять  $c = g_i'(x)$  для любого такого  $i$ . Для произвольной выпуклой функции  $f(x)$  для любой точки  $x$  найдется последовательность  $x^n \rightarrow x$ , причем  $f'(x^n)$  существуют и можно взять  $c = \lim_{n \rightarrow \infty} f'(x^n)$ . Опишем

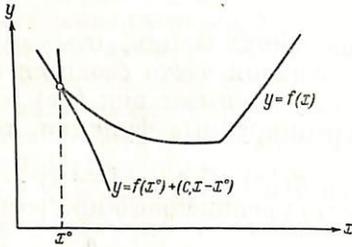


Рис. 4

теперь метод минимизации, предложенный в работах [58, 59] и использующий опорные функционалы. Пусть известно, что точка минимума  $f(x)$  находится в некотором множестве  $M^0$  (например, это может быть множество вида  $|x_i| \leq R$ ,  $i = 1, \dots, n$ ,  $R$  достаточно велико). Найдём  $x^0$  — центр тяжести  $M^0$ , найдём  $c^0$  — опорный к  $f(x)$  в  $x^0$  и построим  $M^1 = \{x \in M^0, (c^0, x - x^0) \leq 0\}$ . Тогда нетрудно видеть, что точка минимума  $f(x)$  будет находиться в  $M^1$  и процесс можно повторить: найти  $x^1$  — центр тяжести  $M^1$ , построить  $M^2 = M^1 \cap \{x: (c^1, x - x^1) \leq 0\}$  и т. д. Как доказано в [58, 59], такой метод сходится со скоростью геометрической прогрессии, знаменатель которой не зависит от свойств  $f(x)$ , а лишь от размерности пространства. В этом методе мы получаем не только последовательность  $x^h$ , сходящуюся к точке минимума, но и сжимающуюся последовательность областей  $M^h$ , заключающую точку минимума. Серьезным недостатком данного метода является сложность нахождения центра тяжести множества  $M^h$  и сложность запоминания самих множеств  $M^h$  (они образованы все увеличивающимся числом гиперплоскостей). В работе [58] приведены некоторые приемы борьбы с этими трудностями (например, предлагается заключать  $M^h$  в некоторый симплекс), не являющиеся, однако, вполне удовлетворительными. По-видимому, данный метод может представлять практический интерес лишь для задач совсем небольшой ( $n = 2 \div 3$ ) размерности.

Метод Келли [60] построен на весьма близкой идее. Однако в качестве вспомогательной задачи на каждом шаге в нем решается задача линейного программирования, а не ищется центр тяжести многогранника. Пусть по-прежнему точка минимума  $f(x)$  заведомо находится в  $M^0$ . Будем предполагать, что  $M^0$  — многогранник, т. е. задается при помощи линейных ограничений  $M^0 = \{x: (a^i, x) \leq b_i\}$ . Возьмем произвольное  $x^0 \in M^0$ , найдём  $c^0$  — опорный к  $f(x)$  в  $x^0$  и решим задачу линейного программирования  $\min x_{n+1}, f(x^0) + (c^0, x - x^0) \leq x_{n+1}, x \in M^0$  (здесь  $x_{n+1}$  — дополнительная переменная). Пусть ее решение есть  $x^1, x_{n+1}^1$ , тогда будем решать задачу  $\min x_{n+1}, f(x^0) + (c^0, x - x^0) \leq x_{n+1}, f(x^1) + (c^1, x - x^1) \leq x_{n+1}, x \in M$  и т. д. Каждая следующая задача получается из предыдущей добавлением нового линейного ограничения, поэтому их удобно решать двойственным симплекс-методом. Недостатком метода является необходимость решать линейные задачи со все большим числом ограничений. По-видимому, этого можно избежать, отбрасывая каждый раз все ограничения, которые не обращаются в равенство в очередной точке  $x^h$ , однако сходимость такого варианта метода не доказана. Кроме того, метод имеет тенденцию к вырождению — векторы  $c^h$  становятся почти линейно зависимыми, и решать задачу линейного программирования все труднее. Достоинством метода является

ся, в частности, возможность оценить точность решения. В самом деле, если  $f^*$  — точное значение минимума, то  $x_{n+1}^k \leq f^* \leq f(x^k)$ . Какова скорость сходимости метода — неясно. Для гладких функций целесообразно применить модификацию метода, предложенную Ф. Вольфом [61]: скорость сходимости при этом увеличивается, а для квадратичной функции можно получить точное решение за конечное число шагов.

Опишем теперь метод, являющийся обобщением градиентного на случай негладкой функции. Образует последовательность

$$x^{k+1} = x^k - \lambda_k c^k, \tag{16}$$

где  $\lambda_k \geq 0$ , а  $c^k$  — опорный к  $f(x)$  в точке  $x^k$ . Если выбирать длину шага постоянной  $\|\lambda_k c^k\| \equiv \rho$ , то, как показал Н. З. Шор [62], предложивший этот метод, можно найти минимум с точностью порядка  $\rho$ . Ю. М. Ермольев [63] и Б. Т. Поляк [64] предложили уменьшать длину шага таким образом, чтобы  $\|\lambda_k c^k\| = \rho_k$ , где  $\rho_k \rightarrow 0$ ,  $\sum_{k=0}^{\infty} \rho_k = \infty$  (например,  $\rho_k = 1 / (k + 1)$ ),

и доказали сходимость такого метода. Важно отметить, что оба варианта метода являются немонотонными: не обязательно  $f(x^{k+1}) \leq f(x^k)$ . Метод чрезвычайно прост. Не совсем ясно, однако, какова его скорость сходимости и от чего она зависит.

В заключение следует сказать, что большинство методов, описанных в этом разделе, недостаточно проверены на практике, так что выводы об их эффективности делать рано.

### 6. УЧЕТ СПЕЦИФИКИ ЗАДАЧ

До сих пор мы предполагали, что минимизируемая функция  $f(x)$  есть некая «функция общего вида». В конкретных задачах, используя специфический вид  $f(x)$ , можно получить значительно более эффективные методы, чем в общем случае.

Прежде всего переменные  $x_1, \dots, x_n$  могут входить по группам, слабо связанным между собой, так что задача распадается на ряд задач минимизации меньшей размерности.

Так, в тривиальном случае  $f(x) = f_1(x_1) + \dots + f_n(x_n)$  минимизация  $f(x)$  сводится к независимой минимизации функций одной переменной  $f_1(x_1), \dots, f_n(x_n)$ .

Рассмотрим несколько более сложный случай

$$f(x) = f_1(x_1, x_2) + f_2(x_2, x_3) + \dots + f_{n-1}(x_{n-1}, x_n).$$

Тогда можно применить метод динамического программирования [52]. Обозначая  $\varphi_i(x_i) = \min_{x_{i+1}, \dots, x_n} [f_i(x_i, x_{i+1}) + \dots + f_{n-1}(x_{n-1}, x_n)]$ , получаем, что  $\min_{x_1, \dots, x_n} f(x) = \min_{x_1} \varphi_1(x_1)$ , а функции  $\varphi_i(x_i)$  связаны рекуррентными соотношениями:

$$\begin{aligned} \varphi_{n-1}(x_{n-1}) &= \min_{x_n} f_{n-1}(x_{n-1}, x_n), \\ \varphi_{i-1}(x_{i-1}) &= \min_{x_i} [f_{i-1}(x_{i-1}, x_i) + \varphi_i(x_i)]. \end{aligned} \tag{17}$$

Чтобы получить отсюда численный метод минимизации, нужно выбрать тот или иной способ аппроксимации функций  $\varphi_i(x_i)$ . Наиболее распространенным в вычислительной практике является следующий «метод блуждающей трубки». Выбираются начальное приближение  $x^0 = (x_1^0, \dots, x_n^0)$  и шаг  $\delta_0$ . Далее функции  $\varphi_i(x_i)$  вычисляются только в точках  $x_i^0, x_i^0 \pm \delta_0$  (иногда на большей сетке, например  $x_i^0, x_i^0 \pm \delta_0, x_i^0 \pm 2\delta_0$ ) с помощью

рекуррентных соотношений (17). Когда  $\varphi_i(x_i)$  построены, ищется следующее приближение  $x^1 = (x_1^1, \dots, x_n^1)$ , реализующее минимум при данных  $\varphi_i(x_i)$ , т. е.

$$\varphi_1(x_1^1) = \min_{x_1} \varphi_1(x_1), \quad f_1(x_1^1, x_2^1) + \varphi_2(x_2^1) = \min_{x_2} [f_1(x_1^1, x_2) + \varphi_2(x_2)]$$

и т. д. Если оказывается, что  $x^1 = x^0$ , то шаг  $\delta_0$  дробится пополам:  $\delta_1 = \delta_0 / 2$ , если  $x' \neq x^0$ , то  $\delta_1 = \delta_0$  и цикл повторяется. Нетрудно видеть связь этого метода с методом покоординатного спуска. В самом деле, в одном из вариантов метода покоординатного спуска (см. раздел 3) производится вычисление значения функции в  $2n$  точках вокруг  $x^0$  вида  $x_1^0, \dots, x_{i-1}^0, x_i^0 \pm \delta_0, x_{i+1}^0, \dots, x_n^0$  и делается шаг в ту из них, в которой  $f(x)$  минимально. Здесь же по существу ищется минимум  $f(x)$  по  $3^n$  точкам вокруг  $x^0$  вида  $x_1^0 + \varepsilon_1 \delta_0, \dots, x_i^0 + \varepsilon_i \delta_0, \dots, x_n^0 + \varepsilon_n \delta_0$ , где  $\varepsilon_i$  пробегает значения  $0, \pm 1$ .

Однако, благодаря специфическому виду  $f(x)$ , такое вычисление минимума по  $3^n$  точкам производится весьма экономно. Существуют и многие другие вычислительные схемы динамического программирования, мы на них не будем останавливаться.

Приведем еще один пример, в котором специфический вид функции позволяет применить специальные методы [65]. Пусть  $f(x) = \varphi(p(x))$ , где  $p(x) = (p_1(x), \dots, p_m(x))$  — векторная функция  $n$  переменных  $x_1, \dots, x_n$ , а  $\varphi(y)$  — числовая функция  $m$  переменных  $y_1, \dots, y_m$ . Например, в методе наименьших квадратов приходится минимизировать выражения вида  $f(x) = (a_1 - p_1(x))^2 + \dots + (a_m - p_m(x))^2$ . В таком случае целесообразно функцию  $p(x)$  аппроксимировать линейно, а функцию  $\varphi(x)$  — квадратично, т. е.

$$\begin{aligned} f(x + \bar{x}) &= \varphi(p(x + \bar{x})) \approx \varphi(p(x) + p'(x)\bar{x}) \approx \\ &\approx \varphi(p(x)) + (\varphi'(p(x)), p'(x)\bar{x}) + 1/2(\varphi''(p(x))p'(x)\bar{x}, p'(x)\bar{x}) = \\ &= f(x) + (p'(x)^* \varphi'(p(x)), \bar{x}) + 1/2(p'(x)^* \varphi''(p(x))p'(x)\bar{x}, \bar{x}). \end{aligned}$$

Отсюда получается метод, являющийся промежуточным между методами градиентным и Ньютона:

$$x^{h+1} = x^h - [p'(x^h)^* \varphi''(p(x^h))p'(x^h)]^{-1} p'(x^h)^* \varphi'(p(x^h)).$$

В частности, в методе наименьших квадратов [65, 66] очередное приближение  $x^{h+1}$  находится из условия минимума квадратичной функции

$$\begin{aligned} &(a_1 - p_1(x^h) - p_1'(x^h)(x - x^h))^2 + \dots \\ &\dots + (a_m - p_m(x^h) - p_m'(x^h)(x - x^h))^2. \end{aligned}$$

Другой специальный прием для метода наименьших квадратов приведен в [67].

Следующий вопрос, который нужно решить при анализе конкретной задачи, — это вопрос о правильном выборе переменных. Можно делать различные преобразования переменных, и в зависимости от этого минимизируемая функция будет становиться «плохой» или «хорошей» с точки зрения методов минимизации. В частности, мы уже отмечали важность выравнивания масштабов переменных для сходимости градиентных методов. Приведем один конкретный пример, в котором переход к новым переменным позволил резко облегчить минимизацию [47, 48]. В задачах подбора параметров, описывающих протекание реакции в химическом реакторе, неизвестные параметры  $A_i, E_i$  входят в комбинации  $A_i \exp(-E_i/T(l))$ , где  $T(l)$  — температура в данной точке реактора. Если бы было  $T(l) = T = \text{const}$ , то эти два параметра можно было бы заменить од-

ним  $R_i = A_i \exp(-E_i/T)$ . Линии уровня минимизируемой функции на плоскости  $A_i, E_i$  имели бы вид  $A_i \exp(-E_i/T) = \text{const}$ . В действительности  $T(l)$  меняется, но слабо, поэтому линии уровня имеют приблизительно тот же вид, т. е. они сильно вытянуты и изогнуты. Минимизировать такую функцию чрезвычайно трудно. Если ввести новую переменную  $B_i = A_i \exp(-E_i/T_{\text{ср}})$ , где  $T_{\text{ср}}$  — некоторая средняя температура, и рассматривать функцию от переменных  $B_i, E_i$ , то линии уровня распрямляются, а вытянутость их остается лишь по направлениям, близким к координатным осям. Численный эксперимент подтвердил, что минимизация в новых переменных гораздо эффективнее.

Отметим в заключение несколько специальных приемов, позволяющих в некоторых случаях ускорять сходимость. Пусть  $f(x) = f_1(x) + f_2(x)$ . Тогда целесообразно время от времени перемежать движение по градиенту  $f(x)$  отысканием минимума на двумерном подпространстве, натянутом на  $f_1'(x)$  и  $f_2'(x)$ . Аналогично, если  $f(x)$  такова, что известна другая функция  $\varphi(x)$ , минимум которой совпадает с минимумом  $f(x)$ , то целесообразно перемежать минимизацию по этим двум функциям.

## 7. МНОГОЭКСТРЕМАЛЬНЫЕ ЗАДАЧИ

До сих пор мы рассматривали методы, пригодные для отыскания единственного минимума функции (или одного из локальных минимумов, если их несколько). Задача отыскания абсолютного минимума функции, имеющей много локальных минимумов, существенно более трудна. Теория таких методов почти не развита. Нет также никакой удовлетворительной классификации всего многообразия многоэкстремальных задач.

Поэтому мы ограничимся описанием нескольких приемов, применяемых в вычислительной практике для отыскания абсолютного минимума. Ни один из этих приемов не гарантирует, что область абсолютного минимума будет найдена (например, в случае глубоких и узких «провалов» функции). Неизвестна также сравнительная эффективность методов.

Многие из описываемых ниже методов (а также некоторые дополнительные) приведены в обзоре Д. Б. Юдина [68]. Известен также ряд методов случайного поиска (см., например, [3]), которые в многоэкстремальных задачах в отличие от одноэкстремальных могут оказаться целесообразными. Опишем кратко еще три группы методов.

Первая группа построена на сочетании какого-либо метода локального спуска с тем или иным правилом выбора совокупности начальных точек. Так, можно выбирать начальные точки на некоторой сетке, можно выбирать их случайно, можно после спуска в локальный минимум делать из него большой шаг тем или иным образом и т. д. (см., например, [69]). При этом спуск из каждой начальной точки производится тем грубее, чем больше ожидаемое число локальных минимумов. В предельном случае, если функция совсем «плохая» (очень быстро изменяющаяся), целесообразно совсем не проводить спуск, а просто вычислять значения функции на сетке, либо детерминированной (обычно равномерной), либо случайной (метод Монте-Карло). Подчеркнем, что для сколько-нибудь «хорошей» функции эти методы совершенно не эффективны.

Другая группа методов — это те из описанных выше методов отыскания единственного минимума, которые способны «проскакивать» небольшие локальные минимумы. Так, И. Б. Моцкус [70] предложил модифицировать обычный координатный спуск, делая в нем по каждой координате шаг до абсолютного минимума (отыскивать абсолютный минимум одномерной функции — задача реальная). Метод «тяжелого шарика» (7) чрезвычайно чувствителен к локальным минимумам с малой «областью притяжения».

В самом деле, если  $x^k$  близко к локальному минимуму  $f'(x^k) \approx 0$ , но  $x^k - x^{k-1}$  велико, то и  $x^{k+1} - x^k$  будет велико и мы можем оказаться вне области притяжения этого минимума. Наконец, весьма эффективным представляется «овражный метод» И. М. Гельфанда и М. Л. Цетлина [13], так как в нем делаются большие «овражные шаги», причем величина такого шага и его направление выбираются независимо от значения функции в «спущенных» точках (см. выше). Поэтому метод обычно выводит из всякой области локального минимума. Практика применения этого метода описана в [71].

Наконец, возможен и статистический подход к многоэкстремальным задачам (он подробно развит в [70]). Предположим, что значение функции в точке можно рассматривать как случайную величину, распределенную по известному закону (оказывается, что для многих практических задач логарифмически нормальный закон распределения удовлетворительно описывает функцию). Разобьем область, в которой ищется минимум, на две части. Тогда в соответствии с общими правилами последовательного анализа мы можем после нескольких испытаний (т. е. вычислений значения функции в точке) принять или отвергнуть гипотезу о том, что точка минимума находится в первой части. Затем оставшаяся часть вновь разбивается на две и испытания продолжаются. Этот метод, по-видимому, довольно трудоемок. Кроме того, функция распределения, которая используется в нем, часто бывает неизвестна.

#### ЛИТЕРАТУРА

1. Л. Н. Фицнер. Автоматическая оптимизация сложных систем пространственного распределения. Автоматика и телемеханика, 1964, т. 25, № 5.
2. С. И. Альбер, Я. И. Альбер. Применение метода дифференциального спуска для решения нелинейных систем. Журн. вычисл. матем. и матем. физ., 1967, т. 7, № 1.
3. Л. А. Растрингин. Случайный поиск в задачах оптимизации многопараметрических систем. Рига, «Зинатне», 1965.
4. С. М. Мовшович. Случайный поиск и градиентный метод в задачах оптимизации. Технич. киберн., 1966, № 6.
5. А. М. Островский. Решение уравнений и систем уравнений. Изд-во иностр. лит., 1963.
6. H. V. Curry. The method of steepest descent for non-linear minimization problem. Quart. Appl. math., 1944, v. 2, № 3.
7. J. B. Crockett, H. Chernoff. Gradient methods of maximization. Pacif. J. Math., 1955, v. 5, № 1.
8. A. D. Booth. An application of the method of steepest descent to the solution of systems of non-linear simultaneous equations. Quart. J. Mech. Appl. Math., 1949, v. 2, № 2.
9. H. A. Spang. III. A review of minimization techniques for non-linear functions. SIAM Rev., 1962, v. 4, № 4.
10. A. A. Goldstein. Cauchy's method of minimization. Number. Math., 1962, v. 4, № 2.
11. Б. Т. Поляк. Градиентные методы минимизации функционалов. Ж. вычисл. матем. и матем. физ., 1963, т. 3, № 4.
12. И. М. Глазман. О градиентной релаксации для неквадратичных функционалов. Докл. АН СССР, 1964, т. 154, № 5.
13. И. М. Гельфанд, М. Л. Цетлин. Принципы нелокального поиска в системах автоматической оптимизации. Докл. АН СССР, 1961, т. 137, № 2.
14. Д. К. Фаддеев, В. Н. Фаддеева. Вычислительные методы линейной алгебры. М., Физматгиз, 1960.
15. D'Esopo. A convex programming procedure. Naval Res. Log. Quart., 1959, v. 6, № 1.
16. V. Schechter. Iteration methods for non-linear problems. Trans. Amer. Math. Soc., 1962, v. 104, № 1.
17. W. Spendley, G. R. Hext, F. R. Himsforth. Sequential application of simplex design in optimization and evolutionary operations. Technometrics, 1962, v. 4, № 4.
18. В. В. Налимов, Н. А. Чернова. Статистические методы планирования экстремальных экспериментов. М., «Наука», 1965.

19. J. A. Nelder, R. Mead. A simplex method for function minimization. *Computer J.*, 1965, v. 7, № 4.
20. С. С. Лавров. Применение барицентрических координат для решения некоторых вычислительных задач. *Ж. вычисл. матем. и матем. физ.*, 1964, т. 4, № 5.
21. Б. Т. Поляк, Б. И. Шостаковский. Одна задача на максимум функции нескольких переменных. В сб. *Вычислительные методы и программирование*, вып. V. М., изд-во МГУ, 1966.
22. Б. Т. Поляк. О некоторых способах ускорения сходимости итерационных методов. *Ж. вычисл. матем. и матем. физ.*, 1964, т. 4, № 5.
23. Б. М. Щедрин, Н. В. Белов, Н. П. Жидков. Метод материальной точки в структурном анализе кристаллов. *Докл. АН СССР*, 1967, т. 170, № 5.
24. J. H. Gouda. Roklovy algoritmus pro urteni minima funkce nekolika promennych. *Aplikace matematiky*, 1966, v. 11, № 4.
25. Л. В. Канторович, Г. П. Акилов. *Функциональный анализ в нормированных пространствах*. М., Физматгиз, 1959.
26. D. W. Marquardt. An algorithm for least square estimation of non-linear parameters. *J. SIAM*, 1963, v. 11, № 2.
27. Л. В. Канторович. О методе Ньютона. *Тр. Матем. ин-та им. В. А. Стеклова АН СССР*, 1949, т. 28.
28. Э. Л. Аким, Т. М. Энеев. Определение параметров движения космического летательного аппарата по данным траекторных измерений. *Космические исследования*, 1963, т. 1, № 1.
29. В. А. Матвеев. Метод приближенного решения систем нелинейных уравнений. *Ж. вычисл. матем. и матем. физ.*, 1964, т. 4, № 6.
30. М. Н. Яковлев. К решению нелинейных уравнений методом итераций. *Докл. АН СССР*, 1964, т. 156, № 3.
31. И. Н. Силин, Л. В. Беляевская. Минимизация функций многих переменных с использованием вторых производных. *Препринт ОИЯИ 2674*, Дубна, 1966.
32. L. Bittner. Mehrpunktverfahren zur Auflösung von Gleichungssystemen. *ZAMM*, 1963, т. 43, № 3.
33. R. Wolfe. The secant method for simultaneous non-linear equations. *Comm. ACM*, 1959, v. 2, № 12.
34. W. M. Kincaid. A two-point method for numerical solution of systems of simultaneous equations. *Quart. Appl. Math.*, 1961, v. 18, № 4.
35. I. G. P. Barnes. An algorithm for solving non-linear equations based on the secant method. *Computer J.*, 1965, v. 8, № 1.
36. R. Fletcher, M. I. D. Powell. A rapidly convergent descent method for minimization. *Computer J.*, 1963, v. 6, № 2.
37. R. Fletcher, C. M. Reeves. Function minimization by conjugate gradients. *Computer J.*, 1964, v. 7, № 2.
38. B. V. Shah, R. J. Buchler, O. Kempthorne. Some algorithms for minimizing a function of several variables. *J. SIAM*, 1964, v. 12, № 1.
39. В. Поляк. О некоторых методах нахождения стационарных точек функций нескольких переменных. *Изв. АН ЭстССР*, т. 16, сер. физ.-мат., 1967, № 1.
40. В. Поляк. О сходимости некоторых методов нахождения стационарных точек функций нескольких переменных. *Изв. АН ЭстССР*, т. 16, сер. физ.-мат., 1967, № 2.
41. И. Ш. Пинскер, Б. М. Цетлин. Нелинейная задача оптимизации. *Автоматика и телемеханика*, 1962, т. 23, № 12.
42. M. I. D. Powell. An efficient method of finding the minimum of a function of several variables without calculating derivatives. *Computer J.*, 1964, v. 7, № 1.
43. R. Fletcher. Function minimization without evaluating derivatives — a review. *Computer J.*, 1965, v. 8, № 1.
44. И. Н. Силин. Минимизация функций многих переменных методом сопряженных прямых. *Препринт ОИЯИ Р-1228*, Дубна, 1963.
45. Rosenbrock. An automatic method for finding the greatest or least value of a function. *Computer J.*, 1960, v. 3, № 3.
46. R. Hoocke, T. A. Jeeves. Direct search solution of numerical and statistical problems. *J. ACM*, 1961, v. 8, № 2.
47. И. В. Гирсанов, В. М. Платонов, Б. Т. Поляк, Е. А. Фейгин. Вычисление параметров кинетических уравнений по экспериментальным данным. *Тр. Всес. конф. по хим. реакторам*, Новосибирск, 1965.
48. Б. Т. Поляк, В. А. Скоков. Определение параметров кинетических уравнений по экспериментальным данным. В сб. *Вычислительные методы и программирование экспериментальным данным*. В сб. *Вычислительные методы и программирование*, вып. IX, М., изд-во МГУ, 1967.
49. M. J. Voh. A comparison of several current optimization methods and the use of transformations in constrained problems. *Comput. J.*, 1966, v. 9, № 1.
50. А. М. Бер, Е. Н. Белов, Б. Т. Поляк. Некоторые задачи оптимизации сетей. В сб. *Вычислительные методы и программирование*, вып. V. М., изд-во МГУ, 1966.

51. J. W a r g a. Minimizing certain convex functions. J. SIAM, 1963, v. 11, № 3.
52. Р. Беллман. Динамическое программирование. М., Изд-во иностр. лит., 1960.
53. P. C r o l a k, L. C o o p e r. An extension of Fibonacci search to several variables. Comm. ACM, 1963, v. 6, № 10.
54. N. S u g i e. An extension of Fibonacci searching to multidimensional cases. IEEE Trans. Autom. Contr., 1964, v. 9, № 1.
55. D. J. W i l d e. A multivariable dichotomous optimum-seeking methods. IEEE Trans. Autom. Contr., 1965, v. 10, № 1.
56. Н. Д. Уайлд. Методы поиска экстремума. М., «Наука», 1967.
57. А. И. Кузовкин, В. М. Тихомиров. О количестве вычислений для нахождения минимума выпуклой функции. Экон. и матем. методы, 1967, т. 3, вып. 1.
58. А. Ю. Левин. Об одном алгоритме минимизации выпуклых функций. Докл. АН СССР, 1965, т. 160, № 6.
59. D. J. N e w m a n n. Location of the maximum on unimodal surfaces. J. ACM, 1965, v. 12, № 3.
60. H. J. K e l l e y. The cutting plane method for solving convex programs. J. SIAM, 1960, v. 8, № 4.
61. P. W o l f e. Accelerating the cutting plane method for nonlinear programming. J. SIAM, 1961, v. 9, № 3.
62. Н. З. Шор. Применение метода градиентного спуска для решения сетевых транспортных задач. Материалы научн. семинаров по теории и прикл. вопросам кибернетики. Эконом. кибернетика и исследование операций, вып. I. Киев, 1962.
63. Ю. М. Ермольев. Методы решения нелинейных экстремальных задач. Кибернетика, 1966, № 4.
64. Б. Т. Поляк. Об одном общем методе решения экстремальных задач. Докл. АН СССР, 1967, т. 174, № 1.
65. С. Н. Соколов, И. Н. Сплин. Нахождение минимумов функционалов методом линеаризации. Препринт ОИЯИ Д 810, Дубна, 1961.
66. М. К. Гавурци, Ю. Б. Фарфоровская. Об одном итеративном методе разыскания минимума суммы квадратов. Журн. вычисл. матем. и матем. физ., 1966, т. 6, № 6.
67. M. J. D. P o w e l l. A method for minimizing a sum of squares of non-linear functions without calculating derivatives. Computer J., 1965, v. 7, № 4.
68. Д. Б. Юдин. Методы количественного анализа сложных систем. I. Техническая кибернетика, 1965, № 1.
69. И. Н. Бочаров, А. А. Фельдбаум. Автоматический оптимизатор для поиска минимального из нескольких минимумов (глобальный оптимизатор). Автоматика и телемеханика, 1962, т. 23, № 3.
70. И. Б. Моцкус. Многоэкстремальные задачи в проектировании. М., «Наука», 1967.
71. И. М. Гельфанд, Е. Б. Вул, С. А. Гинзбург, Ю. Г. Федоров. Метод оврагов в задачах рентгеноструктурного анализа. М., «Наука», 1966.

Поступила в редакцию  
6 VI 1966

## НОВЫЙ МЕТОД РЕШЕНИЯ ОДНОЙ ЗАДАЧИ РАСПРЕДЕЛЕНИЯ

ХОАНГ ТУЙ, НГУЕН КУАНГ ТХАЙ

(Вьетнам)

1. При распределении заданий между несколькими станками (или предприятиями) часто встречается следующая задача, которая была впервые рассмотрена Л. В. Канторовичем [1]:

$$z \rightarrow \max \quad (1)$$

при условиях:

$$(A) \begin{cases} \sum_{j=1}^n x_{ij} + x_{i0} = a_i, & i = 1, 2, \dots, m, \end{cases} \quad (2)$$

$$\begin{cases} \sum_{i=1}^m b_{ij} x_{ij} = b_j z, & j = 1, 2, \dots, n, \end{cases} \quad (3)$$

$$\begin{cases} x_{ij} \geq 0; & i = 1, 2, \dots, m; \quad j = 0, 1, 2, \dots, n, \end{cases} \quad (4)$$

где  $a_i > 0$  — время, которым располагает  $i$ -й станок;  $b_{ij} \geq 0$  — количество изделий  $j$ -го вида, которое может изготовить станок в единицу времени (причем предполагается, что  $\max_i b_{ij} > 0$  для каждого  $j$ );  $b_j > 0$  — число

изделий  $j$ -го вида, входящих в каждый комплект готовой продукции;  $x_{ij}$  — время, использованное  $i$ -м станком для производства изделий  $j$ -го вида;  $z$  — число выпускаемых комплектов.

Как указал Данциг в [2], эта задача была позже изучена также Марковицем [3]. Цель настоящей заметки — изложить новый метод решения этой задачи, который является довольно простым и, в отличие от известных до сих пор методов, не требует специального рассмотрения случая вырожденности.

Докажем сначала некоторые предварительные теоремы.

2. Рассмотрим таблицу  $T$  с  $m$  строками  $i = 1, 2, \dots, m$  и  $n + 1$  столбцами  $j = 0, 1, 2, \dots, n$ . Обозначим через  $(i, j)$  клетку, находящуюся в пересечении  $i$ -й строки и  $j$ -го столбца. Пусть  $G$  — набор клеток, такой, что  $b_{ij} > 0$  для всех  $(i, j) \in G, j \neq 0$ . Предположим, что задача (A) при дополнительном ограничении  $x_{ij} = 0$  для  $(i, j) \notin G$  имеет хотя бы одно допустимое решение, и пусть  $x = [x_{ij}]$  — ее оптимальное решение. Тогда матрицу  $[x_{ij}]$  будем называть максимальным распределением по  $G$ , а соответствующее ей значение  $z$  — мощностью набора  $G$ .

Определим активные столбцы и строки по следующему индуктивному правилу: 1) столбец  $j = 0$  активен; 2) если  $j$ -й столбец активен и  $x_{ij} > 0$ , то  $i$ -я строка активна; 3) если  $i$ -я строка активна и  $(i, j) \in G$ , то  $j$ -й столбец активен.

**Теорема 1.** Существует по крайней мере один неактивный столбец.