Глава 3. НЕЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

3.1. Задача математического программирования

В предыдущей главе мы познакомились с линейным программированием. Приведенные примеры показывают, что многие практические проблемы можно формулировать математически как задачу линейного программирования. Однако имеются проблемы, в которых связи заведомо не являются линейными. Таковы, например, задачи увеличения масштабов производства, перехода на новую технологию, оптовой торговли и т.д. Поэтому ясна необходимость изучения нелинейных моделей и методов их анализа. Модели эти, конечно, сложнее линейных и разработанные для них методы менее эффективны, чем методы решения линейных задач.

Нелинейные задачи мы уже рассматривали в главе I. Однако это были задачи без ограничений или с ограничениями типа равенств. Здесь мы остановимся на случае ограничений типа неравенств, которые изучаются в рамках математического программирования.

Общей задачей математического программирования называется задача отыскания максимума функции $f(x)$ при ограничениях

$$g_i(x) \geq 0, i = \overline{1,m}, \tag{3.1}$$

$$x \in R, \tag{3.2}$$

где R — некоторое непустое подмножество n-мерного евклидова пространства.

Если ввести множество

$$X = \left\{x \middle| x \in R, g_i(x) \geq 0, i = \overline{1,m}\right\} \tag{3.3}$$

то кратко эту задачу можно записать как задачу определения величины (в предположении существования максимума)

$$\hat{v} = \max_{x \in X} f(x) = f(x^0). \tag{3.4}$$

Вообще говоря, в задаче могут быть одновременно ограничения типа равенств и неравенств. Однако такую задачу можно преобразовать к указанному виду. Действительно, ограничение вида равенства $g(x) = 0$ можно заменить двумя ограничениями вида неравенств $g(x) \geq 0, -g(x) \geq 0$, а ограничение вида $g(x) \leq 0$ можно заменить на $-g(x) \geq 0$, т.е. система ограничений (3.1) имеет общий вид. Разделение ограничений на (3.1) и (3.2) не носит принципиального характера, так как каждое из них может быть записано в одном из этих двух видов, но иногда оказывается полезным. Так в задаче линейного программирования записывается отдельно условие неотрицательности $x \geq 0$, хотя это просто частный вид линейного ограничения. Кстати, нетрудно видеть, что задача линейного программирования есть частный случай задачи (3.4). Действительно, если $f(x)$ и $g_i(x)$ линейные функции, R — неотрицательный ортант, а (3.1) переписать в виде обратных неравенств, то получим задачу линейного программирования в стандартной форме.

Точку x^0 называют решением, или точкой глобального максимума, или оптимальным планом, множество X — допустимым множеством или множеством допустимых планов (вообще говоря, оно может быть и пустым), его точки — допустимыми точками или допустимыми планами.

Для задачи (3.4) можно построить функцию Лагранжа (по аналогии со случаем ограничений типа равенств):

$$F(x,y) = f(x) + \sum_{i=1}^{m} y_i g_i(x), \tag{3.5}$$

где $y = (y_i,...,y_m)$ — вектор множителей Лагранжа. Однако в общем случае метод множителей Лагранжа в обычном виде для ее решения неприменим. Во-первых, применительно к задаче с ограничениями вида (3.1), (3.2) метод множителей Лагранжа нуждается в некоторой модификации, для описания которой нам понадобятся новые важные понятия. Во-вторых, и в модифицированном виде метод множителей Лагранжа справедлив лишь для некоторых частных случаев общей задачи математического программирования. Важнейшими такими случаями являются линейное программирование, для которого, как будет показано, метод Лагранжа уже фактически был обоснован и использован, и выпуклое программирование, которое будет далее рассматриваться. Кроме того, мы познакомимся с основными численными методами решения задач математического программирования.

3.2. Седловые точки и двойственность

Важную роль в математическом программировании играет понятие седловой точки, которая определяется следующим образом.

Определение. Пара (x^0, y^0) называется седловой точкой функции $F(x,y)$ на прямом произведении множеств $X \times Y$, если

$$F(x,y^0) \le F(x^0,y^0) \le F(x^0,y) \quad \forall x \in X, y \in Y \tag{3.6}$$

или эквивалентно

$$\max_{x\in X} F(x,y^0) = F(x^0,y^0) = \min_{y\in Y} F(x^0,y). \qquad 3.7)$$

Основными свойствами седловых точек является взаимозаменяемость и эквивалентность. Если (x^1, y^1) и (x^2, y^2) — седловые точки, то (x^1, y^2) и (x^2, y^1) — также седловые точки (взаимозаменяемость), при этом

$$F(x^1, y^1)=F(x^1, y^2)=F(x^2, y^1)=F(x^2, y^2)$$

(эквивалентность). Доказательство этих фактов непосредственно вытекает из определения, и мы его оставляем в качестве упражнения.

Аргументы x и y функции $F(x, y)$ мы будем считать векторами соответственно n–мерного и m–мерного евклидовых пространств. Таким образом, из определения седловой точки следует, что в этой точке по одной группе переменных функция достигает максимума, а по другой — минимума. Хотя в этом состоит существенное отличие седловых точек от точек максимума или минимума функций, условия для их определения аналогичны условиям экстремума. Так, если (x^0, y^0) является внутренней точкой произведения $X{\times}Y$, функция $F(x, y)$ непрерывно дифференцируема по всем переменным, то в седловой точке (x^0, y^0) необходимо выполнение условий

$$\frac{\partial F(x^0, y^0)}{\partial x_j} = \frac{\partial F(x^0, y^0)}{\partial y_i} = 0, \quad j = \overline{1,n}, \quad i = \overline{1,m}. \tag{3.8}$$

Среди решений системы уравнений (3.8) могут быть и точки максимума, и точки минимума, и седловые точки, и другие критические точки (например, точки перегиба).

Возникает вопрос: если в седловой точке функция $F(x, y)$ достигает максимума по x и минимума по y, то нельзя ли находить эти точки последовательным применением операций поиска максимума и минимума функций. Если мы берем минимум функции $F(x, y)$ по y, то получаем функцию от x:

$$\varphi(x) = \min_{y \in Y} F(x, y).$$

У этой функции можно брать максимум по x, в результате получается величина (в предположении достижимости верхней и нижней граней)

$$v_1 = \max_{x \in X} \varphi(x) = \max_{x \in X} \min_{y \in Y} F(x, y). \tag{3.9}$$

Величина v_1 называется максимином функции $F(x, y)$, а задача ее определения называется максиминной задачей.

Аналогично определяется минимакс функции $F(x, y)$:

$$v_2 = \min_{y \in Y} \Psi(y) = \min_{y \in Y} \max_{x \in X} F(x, y), \tag{3.10}$$

где

$$\Psi(y) = \max_{x \in X} F(x, y).$$

В общем случае, когда соответствующие верхняя и нижняя грани не обязательно достигаются, можно определить величины

$$v = \sup_{x \in X} \inf_{y \in Y} F(x, y), \tag{3.11}$$

$$\tilde{v} = \inf_{y \in Y} \sup_{x \in X} F(x, y), \tag{3.12}$$

которые превращаются в v_1 и v_2 при достижимости всех граней. Вообще говоря, существенным является достижимость только верхней грани в (3.11) и нижней грани в (3.12), эти грани называются наружными; если они достигаются, то соответствующие точки реализации верхней грани в (3.11) и нижней грани в (3.12) называются решениями максиминной и минимаксной задач. Теперь наш вопрос можно переформулировать так: имеют ли отношение решения максиминной и минимаксной задач к седловым точкам? Оказывается имеют и самое непосредственное. Но прежде чем показать это, докажем одну простую, но важную лемму.

Лемма. *Величины v и $\tilde{v}$, определенные выражениями* (3.11), (3.12), *всегда связаны соотношением*

$$v \le \tilde{v},$$

т.е. имеет место неравенство

$$\sup_{x \in X} \inf_{y \in Y} F(x, y) \le \inf_{y \in Y} \sup_{x \in X} F(x, y). \tag{3.13}$$

Доказательство. Из определений верхней и нижней граней следуют следующие неравенства

$$\inf_{y\in Y} F(x,y) \le F(x,y') \qquad \forall y' \in Y,$$

$$v = \sup_{x\in X}\inf_{y\in Y} F(x,y) \le \sup_{x\in X} F(x,y') \qquad \forall y' \in Y,$$

откуда $v \le \inf_{y\in Y}\sup_{x\in X} F(x,y) = \tilde{v}$, что и требовалось доказать.

В (3.13) может быть как равенство, так и строгое неравенство. Это видно из следующих простых примеров.

Пример 1. $F(x,y) = (x-y)^2, \quad 0 \le x \le 1, \quad 0 \le y \le 1.$

Здесь

$$v = \max_{0\le x\le 1}\min_{0\le y\le 1}(x-y)^2 = 0,$$

$$\tilde{v} = \min_{0\le y\le 1}\max_{0\le x\le 1}(x-y)^2 = \min\{\min_{0\le y\le \frac{1}{2}}(1-y)^2, \min_{\frac{1}{2}\le y\le 1} y^2\} = \frac{1}{4},$$

т.е. имеет место неравенство $v < \tilde{v}$.

Пример 2. $F(x,y) = (x-y), \quad 0 \le x \le 1, \quad 0 \le y \le 1.$

Здесь

$$v = \max_{0\le x\le 1}\min_{0\le y\le 1}(x-y) = \max_{0\le x\le 1}(x-1) = 0,$$

$$\tilde{v} = \min_{0\le y\le 1}\max_{0\le x\le 1}(x-y) = \min_{0\le y\le 1}(1-y) = 0,$$

т.е. имеет место равенство $v = \tilde{v}$.

Оказывается, что существование седловых точек связано с двумя моментами: достижимостью наружных граней в (3.11) и (3.12), т.е. существованием решений x^0 и y^0 максиминной и минимаксной задач, и выполнением равенства $v = \tilde{v}$. Если выполняется и то и другое, то пара (x^0, y^0) образует седловую точку и любая седловая точка состоит из решений задач (3.11) и (3.12); естественно, их может быть несколько. Если хотя бы один из указанных моментов отсутствует, то у функции *F(x, y)* седловых точек нет. Докажем эти важные факты.

Теорема (необходимые и достаточные условия существования седловой точки). *Для того чтобы функция F(x, y) имела седловую точку на* X×Y, *необходимо и достаточно выполнения равенства*

$$\max_{x\in X}\inf_{y\in Y} F(x,y) = \min_{y\in Y}\sup_{x\in X} F(x,y). \qquad (3.14)$$

Доказательство. *Необходимость.* Пусть существует точка (x^0, y^0), удовлетворяющая (3.7), тогда

$$\sup_{x\in X}\inf_{y\in Y} F(x,y) \le \sup_{x\in X} F(x,y^0) = F(x^0,y^0) = \inf_{y\in Y} F(x^0,y),$$

откуда

$$\sup_{x\in X}\inf_{y\in Y} F(x,y) = \max_{x\in X}\inf_{y\in Y} F(x,y) = \inf_{y\in Y} F(x^0,y) = F(x^0,y^0).$$

Аналогично

$$\inf_{y\in Y}\sup_{x\in X} F(x,y) \ge \inf_{y\in Y} F(x^0,y) = F(x^0,y^0) = \sup_{x\in X} F(x,y^0),$$

откуда

$$\inf_{y\in Y}\sup_{x\in X}F(x,y)=\min_{y\in Y}\sup_{x\in X}F(x,y)=\sup_{x\in X}F(x,y^0)=F(x^0,y^0).$$

Тем самым доказано (3.14), причем x^0 и y^0 — решения задач (3.11), (3.12).

Достаточность. Так как равенство (3.14) подразумевает достижимость наружных верхней и нижней граней, то существуют x^0 и y^0 такие, что

$$\inf_{y\in Y}F(x^0,y)=\max_{x\in X}\inf_{y\in Y}F(x,y),$$

$$\sup_{x\in X}F(x,y^0)=\min_{y\in Y}\sup_{x\in X}F(x,y),$$

причем

$$\inf_{y\in Y}F(x^0,y)=\sup_{x\in X}F(x,y^0).$$

По определению верхней и нижней граней

$$\inf_{y\in Y}F(x^0,y)\le F(x^0,y^0)\le\sup_{x\in X}F(x,y^0)=\inf_{y\in Y}F(x^0,y),$$

т.е. справедливо (3.7) и (x^0, y^0) — седловая точка.

Приведенные выше примеры показывают, что седловая точка может существовать и не существовать. В примере 2 седловая точка существует, это (1,1); в примере 1 седловой точки нет.

В математическом программировании понятие седловой точки используют для построения теории двойственности. Вернемся к общей задаче математического программирования (3.3), (3.4). Введем функцию Лагранжа (3.5) и рассмотрим максиминную задачу

$$v=\max_{x\in R}\inf_{y\in Y}F(x,y), \tag{3.15}$$

где Y — неотрицательный ортант, т.е. $Y=\{y|\ y\ge 0\}$.

Теорема. *Задачи* (3.3), (3.4) и (3.5), (3.15) *эквивалентны, т.е.* $\hat v=v$ *и решение* x^0 *одной задачи* (*если оно существует*) *является решением другой*.

Доказательство. Рассмотрим функцию

$$\varphi(x)=\inf_{y\in Y}F(x,y),\qquad x\in R.$$

Если $x\in X$, то $g_i(x)\ge 0,\ \ i=\overline{1,m}$, и $F(x,y)\ge f(x)$. Так как $F(x,0)=f(x)$, $\varphi(x)\equiv f(x)$ на множестве X. Если $x\in R\backslash X$, то хотя бы для одного i выполняется $g_i(x)<0$. Выбирая последовательность векторов $y^k\in Y$ с неограниченно возрастающей данной компонентой $y_i^k\to\infty$ и нулевыми остальными компонентами, получаем, что $\varphi(x)=-\infty$ на множестве $R\backslash X$. Значит, если $X=\varnothing$, то обе задачи не имеют решения. Если $X\ne\varnothing$, то

$$v=\sup_{x\in X}\varphi(x)=\sup_{x\in X}f(x)=\hat v$$

и точки максимума функций $\varphi(x)$ и $f(x)$, если верхние грани достигаются, совпадают. Теорема доказана.

Рассмотрим теперь минимаксную задачу для функции Лагранжа:

$$\hat v=\min_{y\in Y}\sup_{x\in R}F(x,y). \tag{3.16}$$

Определение. Задача (3.5), (3.16) называется двойственной к задаче (3.3), (3.4); соотношение $\tilde{v} = \hat{v}$, если оно выполняется, называется соотношением двойственности, а теоремы, устанавливающие (при некоторых предположениях) справедливость этого соотношения, называются теоремами двойственности.

Так как выше было показано, что $\hat{v} = v$ и $v \le \tilde{v}$ (последнее для любых функций $F(x, y)$), всегда справедливо $\hat{v} \le \tilde{v}$, но соотношение двойственности не обязательно должно выполняться, и действительно иногда оно справедливо, а иногда и нет. Выполнение соотношения двойственности связано с существованием седловой точки у функции Лагранжа.

Теорема. *Если функция Лагранжа* (3.5) *имеет седловую точку на* $R \times X$, *то выполняется соотношение двойственности, причем если* (x^0, y^0) — *седловая точка функции Лагранжа, то* x^0 — *решение задачи* (3.3), (3.4), а y^0 — *решение двойственной задачи* (3.5), (3.16). *Обратно, если задачи* (3.3), (3.4) *и* (3.5), (3.16) *имеют решения* x^0 *и* y^0 *и выполнено соотношение двойственности, то* (x^0, y^0) — *седловая точка функции Лагранжа* (3.5).

Доказательство данной теоремы непосредственно вытекает из теоремы об эквивалентности задач (3.3), (3.4) и (3.5), (3.15) и теоремы о необходимых и достаточных условиях существования седловой точки.

Факт существования седловой точки у функции Лагранжа представляет не просто теоретический интерес. На нем основываются методы решения задач математического программирования, в первую очередь, метод множителей Лагранжа. Действительно, если (x^0, y^0) — седловая точка функции Лагранжа (3.5), то по определению

$$F(x^0, y^0) = \max_{x \in R} F(x\ , y^0),$$

т.е.

$$f(x^0) + \sum_{i=1}^{m} y_i^0 g_i(x^0) = \max_{x \in R}\Big(f(x) + \sum_{i=1}^{m} y_i^0 g_i(x)\Big).$$

В простейшем случае, когда x^0 — внутренняя точка R, это означает, что существует такой вектор множителей Лагранжа y^0, что

$$\frac{\partial F(x^0, y^o)}{\partial x_j} = \frac{\partial f(x^0)}{\partial x_j} + \sum_{i=1}^{m} y_i^0 \frac{\partial g_i(x^0)}{\partial x_j} = 0, \qquad j = \overline{1,n}\,. \tag{3.17}$$

Уравнения (3.17) представляют собой обычный метод множителей Лагранжа для нахождения экстремума функции с ограничениями на переменные. Однако тут заложены два предположения. Одно, менее существенное, что x^0 — внутренняя точка множества R. Если это не так, то условия оптимальности претерпевают некоторые изменения, мы это продемонстрируем в дальнейшем на примере задачи выпуклого программирования. Другое предположение, принципиальное, о существовании седловой точки у функции Лагранжа. Если она не существует (а такое действительно может быть), то метод Лагранжа уже не обоснован. Достаточно представительными случаями существования седловой точки у функции Лагранжа, да и то при дополнительных предположениях, являются задача линейного программирования и задача выпуклого программирования.

Для задачи линейного программирования справедливо следующее утверждение.

Рассмотрим задачу линейного программирования в стандартной форме

$$\max \sum_{j=1}^{n} c_j x_j$$

при ограничениях

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i, i = \overline{1,m},$$

$$x_j \ge 0, j = \overline{1,n}.$$

В данном случае

$$f(x) = \sum_{j=1}^{n} c_j x_j, \quad g_i(x) = b_i - \sum_{j=1}^{n} a_{ij} x_j, i = \overline{1,m}, \quad R = \{x | x \ge 0\}.$$

Функция Лагранжа имеет вид

$$F(x,y) = \sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{m} y_i (b_i - \sum_{j=1}^{n} a_{ij} x_j) = \sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{m} b_i y_i - \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_j y_i.$$

Для построения двойственной задачи вводим функцию

$$\Psi(y) = \sup_{x \in R} F(x,y) = \sup_{x \ge 0} \left[\sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{m} b_i y_i - \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_j y_i \right] =$$

$$= \sum_{i=1}^{m} b_i y_i + \sup_{x \ge 0} \left[\sum_{j=1}^{n} (c_j - \sum_{i=1}^{m} a_{ij} y_i) x_j \right].$$

Нетрудно видеть, что

$$\sup_{x \ge 0} \left[\sum_{j=1}^{n} (c_j - \sum_{i=1}^{m} a_{ij} y_i) x_j \right] = \begin{cases} 0, & c_j - \sum_{i=1}^{m} a_{ij} y_i \le 0, j = \overline{1,n}, \\ \infty, & \textit{в противном случае,} \end{cases}$$

значит,

$$\Psi(y) = \begin{cases} \sum_{i=1}^{m} b_i y_i, & \textit{если} \quad \sum_{i=1}^{m} a_{ij} y_i \ge c_j, j = \overline{1,n}, \\ \infty, & \textit{если} \quad \exists j : \sum_{i=1}^{m} a_{ij} y_i \le c_j. \end{cases}$$

Так как двойственная задача состоит в минимизации функции $\Psi(y)$ на неотрицательном ортанте, то достаточно ограничиться множеством, где $\Psi(y) < \infty$, если оно не пусто. Поэтому окончательно двойственная задача формулируется следующим образом:

$$\min \sum_{i=1}^{m} b_i y_i$$

при ограничениях

$$\sum_{i=1}^{m} a_{ij} y_i \ge c_j, \quad j = \overline{1,n},$$

$$y_i \ge 0, \quad i = \overline{1,m}.$$

Получили пару двойственных задач, которые рассматривались во второй главе. Аналогично можно построить двойственную задачу к задаче линейного программирования в любой форме. При таком подходе достаточно освоить общий принцип построения двойственных задач и не нужно запоминать конкретный их вид в каждом отдельном случае.

Для задачи линейного программирования справедливо следующее утверждение.

Теорема. *Для того чтобы* x^0 *было решением задачи линейного программирования, необходимо и достаточно существования* y^0 *такого, что пара* (x^0, y^0) *является седловой точкой функции Лагранжа на прямом произведении неотрицательных ортантов.*

Достаточность доказана для общей задачи математического программирования, значит, в частности, и для линейного программирования. Необходимость следует из первой теоремы двойственности в линейном программировании, так как если прямая задача имеет решение x^0, то и двойственная имеет решение y^0 и пара (x^0, y^0) является седловой точкой.

3.3. Выпуклое программирование

Для того чтобы сформулировать задачу выпуклого программирования, необходимо напомнить определения выпуклой и вогнутой функций.

Определение. Функция $f(x)$ называется выпуклой на выпуклом множестве X, если для любых точек $x', x'' \in X$ и произвольного числа $\lambda \in [0,1]$ выполняется неравенство

$$f(\lambda x'+(1-\lambda)x'') \le \lambda f(x')+(1-\lambda)f(x'').$$

Функция $f(x)$ называется вогнутой на выпуклом множестве X, если для любых точек $x', x'' \in X$ и произвольного числа $\lambda \in [0,1]$ выполняется обратное неравенство

$$f(\lambda x'+(1-\lambda)x'') \ge \lambda f(x')+(1-\lambda)f(x'').$$

Определение. Задачей выпуклого программирования называется частный случай общей задачи математического программирования (3.3), (3.4), когда целевая функция $f(x)$ и функции ограничений $g_i(x)$ являются вогнутыми на выпуклом множестве R.

Так как задача максимизации функции эквивалентна задаче минимизации этой функции со знаком минус, ограничение $g_i \ge 0$ эквивалентно $-g_i \le 0$, и из вогнутости функции $f(x)$ следует выпуклость $-f(x)$ и наоборот (докажите это в качестве упражнения), то задачей выпуклого программирования называется также задача минимизации выпуклой функции $f(x)$ при ограничениях $g_i(x) \le 0$, $i=\overline{1,m}$, $x \in R$, где $g_i(x)$ — выпуклые функции, R — выпуклое множество. Это просто другая форма той же задачи. Однако следует обратить внимание, что если задача выпуклого программирования формулируется как задача на максимум, то целевая функция обязательно вогнута, а если на минимум, то — выпукла; если ограничения записываются в виде $g_i(x) \ge 0$, то функции ограничений вогнуты, если $g_i(x)\leq 0$, то — выпуклы. Эта жесткая связь обусловлена наличием определенных свойств у задачи выпуклого программирования, которые в противном случае (при нарушении такой связи) могут и не выполняться. Два основных таких свойства сформулированы в следующей лемме.

Лемма. *Множество допустимых планов задачи выпуклого программирования* (3.3) *является выпуклым. Любой локальный максимум вогнутой функции $f(x)$ на X является глобальным.*

Доказательство. Пусть $x', x''\in X$, т.е. $x', x''\in R$ и $g_i(x')\geq 0$, $g_i(x'')\geq 0$, $i=\overline{1,m}$, и $\lambda\in[0, 1]$.Тогда по определению выпуклого множества $x=\lambda x'+(1-\lambda)x''\in R$ и по определению вогнутой функции

$$g_i(x)=g_i(\lambda x'+(1-\lambda)x'')\geq \lambda g_i(x')+(1-\lambda)g_i(x'')\geq 0,\ i=\overline{1,m},$$

т.е. $x\in X$. Выпуклость множества X доказана.

Далее, пусть x^* — локальный максимум функции $f(x)$ на X, т.е. $f(x^*)\geq f(x)$ для любого x из пересечения X с некоторой (возможно малой) окрестностью ω точки x^*. Нам надо показать, что x^* является и глобальным максимумом, т.е. что $f(x^*)\geq f(x)\ \forall x\in X$. Предположим противное, т.е. что существует $\hat{x}\in X$ такое, что $f(\hat{x})>f(x^*)$. Возьмем $\tilde{x}_\lambda=\lambda x^*+(1-\lambda)\hat{x}$, где $0<\lambda<1$, тогда в силу вогнутости функции $f(x)$ имеем

$$f(\tilde{x}_\lambda)\geq\lambda\ f(x^*)+(1-\lambda)f(\hat{x})>\lambda\ f(x^*)+(1-\lambda)f(x^*)=f(x^*).$$

При достаточно малом $\lambda>0$, очевидно, $\tilde{x}_\lambda\in X\cap\omega,\ \ f(\tilde{x}_\lambda)>f(x^*),$ пришли к противоречию. Лемма доказана.

Если бы функции ограничений $g_i(x)$ были выпуклы, то для определяемого (3.3) множества X выпуклость уже бы не следовала; в этом случае можно доказать выпуклость множества $X'=\left\{x\middle| x\in R, g_i(x)\leq 0, i=\overline{1,m}\right\}$. Если бы целевая функция $f(x)$ была выпукла, то утверждение леммы относительно ее максимумов стало бы неверным, однако аналогичное утверждение можно было бы доказать для минимумов.

Бывают функции, которые выпуклы по одной группе переменных и вогнуты по другой. Такие функции представляют собой один из основных классов функций, у которых заведомо существует седловая точка. Приведем без доказательства следующую теорему.

Теорема (о существовании седловой точки у вогнуто-выпуклой функции). *Пусть X и Y — выпуклые, замкнутые, ограниченные подмножества конечномерных евклидовых пространств, а функция $F(x, y)$ непрерывна по x и y, вогнута по x $\forall y$ и выпукла по y $\forall x$, тогда $F(x, y)$ имеет на $X\times Y$ седловую точку.*

Рассмотрим для задачи выпуклого программирования функцию Лагранжа (3.5), определенную на прямом произведении $R\times Y$, где $Y=\{y\,|\,y\geq 0\}$. Эта функция в силу вогнутости $f(x)$, $g_i(x)$ является вогнутой по x $\forall y$ и линейной, а следовательно, выпуклой по y $\forall x$. Однако нельзя утверждать на основании предыдущей теоремы, что она имеет седловую точку, так как множество R не обязательно является замкнутым и ограниченным, а Y заведомо неограничено. Тем не менее можно ожидать, что при некоторых условиях, заменяющих невыполняющиеся в данном случае условия этой теоремы, седловая точка все же будет существовать. Наиболее известной теоремой, относящейся к этому вопросу, является теорема Куна-Таккера, которая представляет собой центральный результат теории двойственности в выпуклом программировании. Эта теорема устанавливает связь между существованием решения задачи выпуклого программирования и седловой точки у функции Лагранжа при выполнении некоторого условия регулярности. Наиболее простым вариантом такого условия является условие Слейтера: задача выпуклого программирования удовлетворяет условию Слейтера, если существует такая точка $x \in R$, что $g_i(x)>0$, $i=\overline{1,m}$.

Теорема Куна-Таккера. *Если задача выпуклого программирования удовлетворяет условию Слейтера, то необходимое и достаточное условие оптимальности плана $x^0 \in X$ – это существование такого $y^0 \geq 0$, что пара (x^0, y^0) является седловой точкой функции Лагранжа* (3.5) *на $R \times Y$.*

Доказательство. *Достаточность* нами доказана для общей задачи математического программирования, следовательно, она справедлива и для задачи выпуклого программирования без всяких дополнительных условий (типа условия Слейтера).

Докажем *необходимость*. Для этого рассмотрим в $(m+1)$-мерном пространстве два множества

$$A = \left\{ z = \begin{pmatrix} z_1 \\ z_2 \\ \dots \\ z_{m+1} \end{pmatrix} \middle| z_i \leq g_i(x), i=\overline{1,m}, z_{m+1} \leq f(x) \right\},$$

$$B = \left\{ z = \begin{pmatrix} z_1 \\ z_2 \\ \dots \\ z_{m+1} \end{pmatrix} \middle| z_i > 0, i=\overline{1,m}, z_{m+1} > f(x^0) \right\},$$

где x^0 — решение задачи выпуклого программирования, x — произвольная точка множества R.

Покажем, что множество A является выпуклым. Действительно, если $z', z'' \in A$, то $z_i' \le g_i(x')$, $z_i'' \le g_i(x'')$, $i = \overline{1,m}$, $z'_{m+1} \le f(x')$, $z''_{m+1} \le f(x'')$, $x', x'' \in R$. Рассмотрим $z^\alpha = \alpha \cdot z' + (1-\alpha) z''$, $0 \le \alpha \le 1$.

Для него выполняются неравенства

$$z_i^\alpha \le \alpha\, g_i(x') + (1-\alpha) g_i(x'') \le g_i(\alpha\, x' + (1-\alpha) x'') = g_i(x^\alpha),$$

где

$$x^\alpha = \alpha\, x' + (1-\alpha) x'' \in R,\ i = \overline{1,m};$$

$$z_{m+1}^\alpha \le \alpha\, f(x') + (1-\alpha) f(x'') \le f(\alpha x' + (1-\alpha) x'') = f(x^\alpha),$$

т.е. по определению множества A вектор $z^\alpha \in A$, значит A — выпуклое множество.

Множество B также является выпуклым и представляет собой внутренность ортанта с вершиной в точке $(f(x^0), 0, \ldots, 0)$.

В силу оптимальности x^0 множества A и B не имеют общих векторов, т.е. $A \cap B = \varnothing$. Действительно, если существует $z \in A \cap B$, то для некоторого $x \in R$ выполняются условия

$$g_i(x) \ge z_i,\ \ i = \overline{1,m},\ \text{т.е. } x \in X,$$

и

$$f(x) \ge z_{m+1} > f(x^0),$$

пришли к противоречию.

Значит, по теореме о разделяющей гиперплоскости существует такой ненулевой вектор $a = \begin{pmatrix} a_1 \\ a_2 \\ \ldots \\ a_{m+1} \end{pmatrix}$, что

$$\langle a, z^1 \rangle \le \langle a, z^2 \rangle \qquad \forall z^1 \in A, z^2 \in B. \tag{3.18}$$

По определению множества B, очевидно, $a \ge 0$, так как если $a_i < 0$, то, выбирая последовательность векторов $z^k \in B$ с $z^k_i \to \infty$ и остальными компонентами равными нулю, получим $\langle a, z^k \rangle \to -\infty$, что противоречит неравенству (3.18). Далее, так как точка $(f(x^0), 0, \ldots, 0)$ является предельной для множества B, то имеем из (3.18) с учетом определения множества A

$$\sum_{i=1}^{m} a_i g_i(x) + a_{m+1} f(x) \le a_{m+1} f(x^0) \tag{3.19}$$

для любой точки $x \in R$. Отсюда следует, что $a_{m+1} > 0$, поскольку, если $a_{m+1} = 0$, то из (3.19) вытекает

$$\sum_{i=1}^{m} a_i\, g_i(x) \le 0 \quad \forall x \in R \quad ,$$

что противоречит условию Слейтера, так как хотя бы одно $a_i > 0$ и при выполнении условия Слейтера существует $x \in R$ такое, что $g_i(x) > 0, i = \overline{1,m}$, т.е. $\sum_{i=1}^{m} a_i g_i(x) > 0$. Итак, $a \geq 0$ и $a_{m+1} > 0$. Положим $y_i^0 = \frac{a_i}{a_{m+1}}, i = \overline{1,m}$, тогда $y^0 = (y_1^0, \ldots, y_m^0) \geq 0$ и из (3.19) получаем

$$\sum_{i=1}^{m} y_i^0 g_i(x) + f(x) \leq f(x^0) \quad \forall x \in R. \tag{3.20}$$

Полагая в (3.20) $x = x^0$, получим $\sum_{i=1}^{m} y_i^0 g_i(x^0) \leq 0$, но, с другой стороны, $\sum_{i=1}^{m} y_i^0 g_i(x^0) \geq 0$, так как $y_i^0 \geq 0$, $g_i(x^0) \geq 0$, $i = \overline{1,m}$, следовательно,

$$\sum_{i=1}^{m} y_i^0 g_i(x^0) = 0. \tag{3.21}$$

Из (3.20) и (3.21) имеем

$$F(x, y^0) \leq F(x^0, y^0) \quad \forall x \in R.$$

Учитывая, что $\sum_{i=1}^{m} y_i g_i(x^0) \geq 0 \quad \forall y \geq 0$, имеем $F(x^0, y^0) \leq F(x^0, y) \quad \forall y \geq 0$. Значит (x^0, y^0) — седловая точка функции $F(x, y)$ на $R \times Y$. Теорема доказана.

То, что условие Слейтера является существенным, показывает следующий пример.

Пример. $f(x) = x$, $m = 1$,

$g(x) = -x^2$, $R = R^1$.

Очевидно, $x = 0$ есть решение задачи, но у функции $F(x, y) = x - yx^2$ седловой точки нет.

Теорема Куна-Таккера обосновывает сведение задачи выпуклого программирования к задаче поиска седловой точки функции Лагранжа. Задача поиска седловой точки аналогична по сложности задаче поиска экстремума функции, поэтому для того, чтобы такое сведение имело практический смысл, необходимо, чтобы получающаяся экстремальная задача была в чем-то проще исходной. Такое упрощение связано со следующим обстоятельством. Исходная задача представляет собой поиск максимума на множестве X, которое имеет вид (3.3). Разбиение ограничений, описывающих допустимое множество X, на $x \in R$ и $g_i(x) \geq 0$, как уже говорилось, является условным, так как любое ограничение можно записать и в том, и в другом виде. Поэтому обычно под R понимается множество простого вида, это либо все пространство R^n, либо неотрицательный ортант, либо параллелепипед, т.е. границы множества R задаются простыми ограничениями типа $a \leq x \leq b$.

Сложность же задачи выпуклого программирования определяется системой ограничений $g_i(x) \geq 0$, $i = \overline{1,m}$. Так как седловая точка функции Лагранжа ищется на произведении множеств $R \times Y$, где Y также является множеством простого вида (неотрицательным ортантом), то смысл теоремы Куна-Таккера состоит в сведении задачи поиска экстремума функции со сложными ограничениями на переменные к задаче поиска экстремума новой функции с простыми ограничениями на переменные. Если множество R совпадает со всем пространством R^n, то условия экстремума имеют вид (3.17), причем очень существенным моментом является то, что эти условия являются не только необходимыми для того, чтобы функция $F(x, y^0)$ достигала в точке x^0 максимума по x, но и достаточными. Это является важным свойством вогнутых функций (а $F(x, y)$ для задачи выпуклого программирования вогнута по x).

Теорема. *Для того чтобы непрерывно дифференцируемая вогнутая функция $\varphi(x)$ имела в точке x^0 максимум на R^n, необходимо и достаточно, чтобы $\varphi'(x^0)=0$* (докажите в качестве упражнения).

Таким образом, для нахождения седловой точки функции Лагранжа $F(x, y)$ на произведении $R^n \times Y$, а следовательно, и для нахождения решения x^0 задачи выпуклого программирования при $R=R^n$ надо решить систему уравнений (3.17). Но в этой системе n уравнений, а неизвестных, вообще говоря, $n+m$, так как помимо n-мерного вектора x^0 нам неизвестен и m–мерный вектор соответствующих ему множителей Лагранжа y^0. Однако, если внимательно просмотреть доказательство теоремы Куна-Таккера, то мы увидим очень важное свойство, которое выполняется для седловой точки функции Лагранжа, а именно, условие (3.21). Так как $y_i^0 \geq 0$ и $g_i(x^0) \geq 0,\ i=\overline{1,m}$, то из него вытекает, что либо $y_i^0=0$, либо $g_i(x^0)=0$, либо то и другое одновременно. Это свойство аналогично второй теореме двойственности линейного программирования. Ограничения, которые выполняются в некоторой точке как равенства, называются активными. Таким образом, только те множители Лагранжа y_i^0 могут быть отличны от нуля, которые соответствуют активным в точке x^0 ограничениям. С учетом этого свойства для нахождения решения задачи выпуклого программирования получаем следующий способ.

Введем множество

$$I(x) = \{i \mid 1 \leq i \leq m, g_i(x) = 0\}.$$

Множество $I(x)$ представляет собой совокупность индексов активных в точке x ограничений. Тогда можно утверждать в силу сказанного выше, что

$$y_i^0=0 \quad \forall i \notin I(x^0).$$

Поэтому для определения x^0 и ненулевых y_i^0 имеем систему уравнений

$$\begin{cases} \dfrac{\partial f(x^0)}{\partial x_j} + \sum\limits_{i \in I(x^0)} y_i^0 \dfrac{\partial g_i(x^0)}{\partial x_j} = 0,\ j=\overline{1,n}, \\ g_i(x^0)=0, i \in I(x^0). \end{cases} \tag{3.22}$$

Число неизвестных и число уравнений здесь совпадают, поэтому, если мы найдем решение системы уравнений (3.22), мы тем самым решим исходную задачу выпуклого программирования. Система (3.22) представляет собой необходимые и достаточные условия оптимальности для задачи выпуклого программирования в случае, когда $R=R^n$.

Рассмотрим, как трансформируются эти условия в случаях, когда R не совпадает со всем пространством R^n.

Сначала остановимся на случае, когда R является неотрицательным ортантом, т.е. $R=\{x|\ x\geq 0\}$. Тут для каждой компоненты x^0_j решения задачи выпуклого программирования x^0 возможны две альтернативы: либо $x^0_j > 0$, либо $x^0_j = 0$. Для первой альтернативы условия оптимальности совпадают со случаем $R=R^n$, т.е. имеют вид (3.17) при данном j. Однако для второй альтернативы это не так. Если $x^0_j = 0$ является компонентой точки максимума функции $F(x, y^0)$ по x в области $x\geq 0$, то можно утверждать, что необходимо (а для вогнутой функции и достаточно) только выполнение неравенства

$$\frac{\partial F(x^0, y^0)}{\partial x_j} \leq 0,$$

т.е. функция $F(x, y^0)$ должна не возрастать по x_j при движении внутрь области $x\geq 0$. Но так как $x_j^0 = 0,$ то можно записать

$$\frac{\partial F(x^0, y^0)}{\partial x_j} x_j^0 = 0.$$

Объединяя обе альтернативы, необходимые и достаточные условия оптимальности для задачи выпуклого программирования в случае $R=\{x|\ x\geq 0\}$ можно представить в виде

$$\begin{cases} \dfrac{\partial F(x^0, y^0)}{\partial x_j} \leq 0,\ j=\overline{1,n}, \\ \displaystyle\sum_{j=1}^{n} \frac{\partial F(x^0, y^0)}{\partial x_j} x_j^0 = 0. \end{cases} \tag{3.23}$$

Так как $x^0_j \geq 0$, то из (3.23) следует, что либо $x_j^0 = 0,$ либо

$$\frac{\partial F(x^0, y^0)}{\partial x_j} = 0\ \forall j.$$

Таким образом, как и в случае $R=R^n$, мы имеем n уравнений, которые вместе с уравнениями для активных ограничений определяют, в принципе, x^0 и y^0.

Теперь перейдем к более общим случаям.

Хотя в качестве множества R, как говорилось, обычно выбирается достаточно простое множество, чтобы не рассматривать каждый случай отдельно, рассмотрим общую ситуацию, когда R — произвольное выпуклое множество. Рассмотрение такой ситуации полезно еще и потому, что оно позволяет несколько по-другому подойти к задаче выпуклого программирования. Действительно, если в R включить все ограничения $g_i(x)\geq 0$, то оно будет просто совпадать с множеством X и задача выпуклого программирования состоит в максимизации вогнутой функции $f(x)$ на выпуклом множестве R (или X). Поэтому, не конкретизируя вида ограничений и не вводя функции Лагранжа, можно напрямую рассмотреть функцию $f(x)$ на X. При этом очень полезным является следующее понятие.

Определение. Направление g является допустимым в точке x, принадлежащей выпуклому множеству X, если $\exists\ \varepsilon_0 > 0$ такое, что $\forall\ \varepsilon \le \varepsilon_0$ выполняется $x+\varepsilon g\in X$.

Так как определение допустимого направления не зависит, очевидно, от длины вектора g, то будем его в дальнейшем считать единичным, т.е. $|g|=1$.

Допустимое направление — это такое направление, небольшое перемещение вдоль которого не выводит из допустимого множества, т.е. не нарушает ограничений задачи математического программирования. Но если точка x^0 является решением задачи математического программирования, т.е. доставляет максимум функции $f(x)$ на X, то при перемещении вдоль допустимого направления, по определению максимума, функция $f(x)$ не может возрастать. Изменение функции вдоль некоторого направления определяется ее производной по направлению.

Определение. Производной функции $f(x)$ по направлению, задаваемому единичным вектором g, называется величина

$$\frac{\partial f(x)}{\partial g} = \lim_{\varepsilon\to 0^+} \frac{f(x+\varepsilon g)-f(x)}{\varepsilon}, \qquad (3.24)$$

если предел в (3.24) существует.

С помощью производных по направлению получаем другую форму условий оптимальности: для того чтобы x^0 было решением задачи математического программирования, необходимо, а для задачи выпуклого программирования и достаточно, чтобы для любого допустимого в x^0 направления g выполнялось

$$\frac{\partial f(x^0)}{\partial g} \le 0. \qquad (3.25)$$

В частности, условие (3.25) применительно к функции Лагранжа на R дает условие

$$\frac{\partial F(x^0, y^0)}{\partial g} \le 0,$$

где g — любое допустимое направление в точке $x^0\in R$.

3.4. Графический метод нелинейного программирования

Задачи нелинейного программирования обладают свойствами, которые усложняют процесс их решения по сравнению с задачами линейного программирования:

1. Множество допустимых планов (мы будем обозначать его буквой G) может иметь очень сложную структуру. Например, быть невыпуклым или несвязным.

2. Глобальный максимум (минимум) может достигаться как внутри множества *G,* так и на его границах (где он, вообще говоря, будет не совпадать ни с одним из локальных экстремумов).

3. Целевая функция *f(x)* может быть не дифференцируемой, что затрудняет применение классических методов математического анализа.

В силу названных факторов задачи нелинейного программирования настолько разнообразны, что для них не существует общего метода решения.

В этом параграфе мы рассмотрим графический метод решения задач нелинейного программирования. Его алгоритм такой же, как и для решения задач линейного программирования. Для решения задач нелинейного программирования существенно важно знать:

1) выпукло или не выпукло множество допустимых решений задачи;

2) является ли целевая функция выпуклой или вогнутой, или она не относится ни к тому ни к другому классу.

Если количество переменных в неравенствах, задающих область допустимых планов задач, равно 2, то ее можно изобразить на координатной плоскости. Каждое неравенство определяет некоторую полуплоскость. Пересечение данных полуплоскостей *G* является областью допустимых планов задач. Поведение целевой функции в рамках двумерной иллюстрации может быть охарактеризовано с помощью линии уровня.

Линией уровня функции называется множество точек из области ее определения, в которых функция принимает одно и тоже фиксированное значение. Градиент функции — это вектор, указывающий направление наиболее быстрого возрастания функции (мы будем его обозначать $\vec{c}$).

Таким образом, с геометрической точки зрения задача максимизации сводится к определению такой точки области *G*, через которую проходят линии уровня соответствующие наибольшему из возможных значений. Для чего необходимо сначала построить линию уровня для некоторого произвольного значения функции. Затем осуществить ее параллельное движение (перпендикулярно $\vec{c}$) до тех пор, пока не достигнем такой точки области допустимых планов *G,* из которой смещение в направлении вектора $\vec{c}$ было бы невозможно.

Заметим, что решение задачи поиска минимума *f(x)* осуществляется аналогично, лишь движение по линиям уровня должно производиться в направлении обратному градиенту целевой функции, то есть по вектору ($-\vec{c}$). Таким образом, решением может быть одна точка.

Возможна ситуация неограниченности целевой функции *f(x)* на множестве *G*, то есть сколько бы мы ни перемещались по линиям уровня в направлении вектора $\vec{c}$ ее значение будет возрастать.

Возможен и случай касания линии уровня грани (линии) множества *G*, соответственно, все точки, лежащие на этой грани, будут являться оптимальными планами.

Заметим, что аналогичным образом могут быть построены интерпретации задач нелинейного программирования для случая трехмерного пространства R^3, где множеству G будет соответствовать некоторый ограниченный или неограниченный многогранник, а поведение целевой функции будет характеризоваться поверхностями (плоскостями) уровня.

Этот метод может быть применен не только к задачам с двумя или тремя переменными и ограничениями в виде неравенств, но и к задачам, у которых $i - k = 2$ (i — количество ограничений, k — количество неизвестных).

Действительно, можно выбрать две произвольные переменные и, используя систему уравнений (неравенств), выразить через них остальные переменные.

Задача. *На предприятии имеется два вида ресурсов. Определите оптимальное распределение величин затрачиваемых ресурсов на производство некоторого продукта, если цена ресурса первого вида 3 единицы, второго – 4 единицы, а всего на производство выделено 24 единицы. Известно, что из количества x первого ресурса и y второго ресурса можно получить* $\sqrt{x^2 + y^2}$ *единиц продукта.*

Р е ш е н и е. Пусть x — количество ресурсов первого вида, y — количество ресурсов второго вида. Математическая модель задачи: на множестве ограничений

$$G: \begin{cases} 3x + 4y \le 24, \\ x \ge 0, \\ y \ge 0. \end{cases}$$

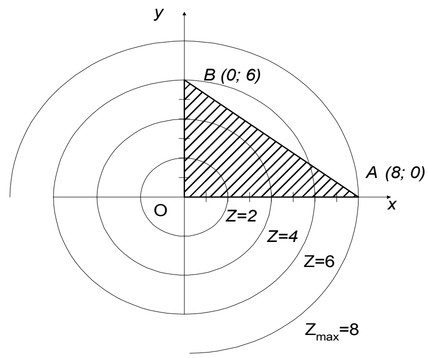

Рис. 1

Множество допустимых решений заштриховано на рис. 1. Если целевой функции придавать фиксированные значения 1, 2, 3,..., то будем получать окружности с центром в начале координат и радиусом 1, 2, 3, … Начертим ряд окружностей (линии уровня целевой функции). Из рисунка видно, что функция $z = \sqrt{x^2 + y^2}$ достигает наибольшего значения, равного 8, в точке $A(8; 0)$, т.е. $z_{max}=z(8; 0)=8$. Значит, количество первого ресурса должно равняться 8, а использование второго ресурса нерационально.

3.5. Численные методы нелинейного программирования

Хотя теорема Куна-Таккера и вытекающие из нее условия на производные дают характеристику решений задачи математического программирования (по крайней мере выпуклого), они еще не дают конструктивных методов нахождения этих решений для сколь-либо сложных случаев. Рассмотрим некоторые из наиболее распространенных конструктивных (численных) методов, которые так или иначе используют полученные выше условия оптимальности.

Наиболее сильный метод решения экстремальных задач — метод множителей Лагранжа. Но он разработан для случая, когда множество условий задается системой уравнений, а не системой неравенств. Метод проекции и метод условного градиента применимы лишь для задач минимизации на выпуклых множествах, причем метод условного градиента применим лишь для множеств, задаваемых линейными ограничениями, поскольку в этих случаях для выбора направления спуска достаточно решить задачу линейного программирования, а метод проекции градиента применяют на множестве такого вида, что задача отыскания проекции некоторой точки является достаточно простой с точки зрения ее численной реализации, так как решение этой задачи и определяет направление спуска. Метод Ньютона целесообразно применять в том случае, если целевая функция строго выпуклая и достаточно гладкая в окрестности точки x^0. Метод штрафных функций применим к задачам со сложными ограничениями.

1. Градиентные методы

Рассмотрим сначала задачу максимизации функции $f(x)$ без ограничений, т.е. в случае, когда X совпадает со всем пространством R^n. Градиент функции $f(x)$ будем по-прежнему обозначать $f'(x)$. Условие оптимальности в этом случае имеет вид

$$f'(x)=0, \tag{3.26}$$

однако непосредственное решение системы уравнений (3.26) может оказаться чересчур сложным, поэтому на практике поступают следующим образом. Выбирая произвольную начальную точку $x^{(0)}$, строят итеративный процесс

$$x^{(k+1)} = x^{(k)} + \alpha_k f'(x^{(k)}), \qquad k=0, 1, 2 \tag{3.27}$$

Число α_k называют длиной шага, или просто шагом. Если все α_k равны между собой, то имеем процесс с постоянным шагом.

Процесс (3.27), лежащий в основе градиентных методов, представляет собой движение в сторону возрастания функции $f(x)$, так как если $f'(x^{(k)})\neq 0$, то всегда можно выбрать α_k, так, что $f(x^{(k+1)})>f(x^{(k)})$. Существуют разные способы выбора α_k. Вообще говоря, наилучшим является выбор такого α_k, при котором обеспечивается максимальный рост функции $f(x)$. Такое α_k находится из условия

$$f(x^{(k)}+\alpha_k\, f'(x^{(k)}))=\max_{\alpha\geq 0} f(x^{(k)}+\alpha_k\, f'(x^{(k)})). \qquad (3.28)$$

Градиентный метод поиска экстремума (3.27) с выбором шага по способу (3.28) называется метод скорейшего подъема (или спуска для задачи на минимум). Такой метод требует наименьшего числа итераций, но зато на каждом шаге приходится решать дополнительную задачу поиска экстремума (3.28) (правда, в одномерном случае). На практике часто довольствуются нахождением любого α_k, обеспечивающего рост функции. Для этого берут произвольное α_k и проверяют условие роста, если оно не выполняется, то дробят α_k до тех пор, пока это условие не будет выполнено (такое достаточно малое α_k при $f'(x^{(k)})\neq 0$ существует всегда).

Процесс (3.27), очевидно, останавливается, когда выполнено условие (3.26). При этом, если функция $f(x)$ вогнута, то найденная стационарная точка будет решением задачи максимизации. В противном случае необходимо провести дополнительное исследование функции $f(x)$ в окрестности найденной точки. Однако, даже если она будет точкой максимума, в невыпуклом случае трудно определить локальный это максимум или глобальный. Поэтому градиентные методы обеспечивают нахождение глобального экстремума только для вогнутых (выпуклых) функций, а в общем случае дают лишь локальные экстремумы (при этом можно попытаться найти глобальный экстремум, применяя итеративный процесс многократно с разными начальными точками).

Если рассматривается задача максимизации $f(x)$ при ограничениях, т.е. когда X не совпадает с R^n, то непосредственное применение процесса (3.27) может привести к нарушению ограничений, даже если начальная точка $x^{(0)}\in X$. Однако эту трудность можно преодолеть, например, если получаемую по формуле (3.27) очередную точку проектировать на множество X. Если обозначить операцию проектирования точки x на множестве X через $P_x(x)$, то соответствующий итеративный процесс имеет вид

$$x^{(k+1)}=P_x(x^{(k)}+\alpha_k\, f'(x^{(k)})). \qquad (3.29)$$

Полученный метод носит название метода проекции градиента. Шаг α_k в методе (3.29) может выбираться различными способами (например, как в методе скорейшего подъема). Стационарная точка этого процесса является решением задачи $\max\limits_{x \in X} f(x)$ в случае вогнутой функции $f(x)$, а в общем случае требуется дополнительное исследование.

Недостатком метода проекции градиента является необходимость проведения операции проектирования, которая в общем случае эквивалентна некоторой задаче поиска экстремума. Однако, когда X является шаром, параллелепипедом, гиперплоскостью, полупространством или ортантом, задача проектирования решается просто и в явном виде.

Еще одной разновидностью градиентных методов является метод условного градиента, который также предназначен для решения экстремальных задач с ограничениями. Суть его состоит в решении вспомогательной задачи максимизации на множестве X линейной функции $\langle f'(x^{(k)}), x-x^{(k)} \rangle$, представляющей собой главную часть приращения функции $f(x)$ в точке $x^{(k)}$. Эта вспомогательная задача может быть непростой, но если X задается линейными ограничениями, то она представляет собой задачу линейного программирования, которая решается за конечное число шагов стандартными методами (например, симплекс-методом). Если решение вспомогательной задачи $\hat{x}^{(k)}$ найдено, то следующее приближение для исходной задачи строится по формуле

$$x^{(k+1)} = x^{(k)} + \alpha_k (\hat{x}^{(k)} - x^{(k)}). \qquad (3.30)$$

Если множество X выпуклое, то $x^{(k+1)} \in X$. Шаг α_k выбирается из условия максимального роста функции $f(x)$ или любым другим способом, обеспечивающим рост $f(x)$. На практике обычно решают вспомогательную задачу не точно, а приближенно. В процессе (3.30) направление движения не совпадает с градиентом функции $f(x)$ в точке $x^{(k)}$, но определяется им, так как его компоненты берутся в качестве коэффициентов линейной целевой функции вспомогательной задачи.

Задача. *Найти минимальное значение функции $f= -x_1^2+ x_2^2$ при условии $x_1^2+ x_2^2 \le 1$.*

Р е ш е н и е. Решим данную задачу методом проекции градиента, завершая вычисления при выполнении одного из условий

$$\| f'(x^{(k)}) \| \le 0{,}01\,, \quad \| x^{(k-1)} - x^{(k)} \| \le 0{,}01.$$

В качестве начального приближения возьмем точку $x^{(0)}=(0;0{,}5) \in X$. Градиент $f'(x)=(-1;2x_2)$ удовлетворяет условию Лившица

$\|f'(x')-f'(x'')\| \le L\|x'-x''\|$ с константой $L=2$, так как $\|f'(x')-f'(x'')\| = \sqrt{(-1+1)^2+(2x_2'-2x_2'')} = 2|x_2'-x_2''| \le 2\|x'-x''\|$.

Поэтому в (3.29) можно положить $\alpha_k \equiv \alpha \in (0;1),\ k = 0, 1, 2\ldots$, например $\alpha_k = 0{,}75$.

<u>Ш а г 1.</u> Так как $f'(x^{(0)})$=(-1;1), по формуле (3.29) находим

$$x^{(0)}=P_x[x^{(0)}-\alpha f'(x^{(0)})]=P_x[(0;0{,}5)-0{,}75(-1;1)]=P_x[(0{,}75;-0{,}25)].$$

Точка (0,75;-0,25) принадлежит допустимому множеству, так как $0{,}75^2+0{,}25^2=0{,}625 < 1$, поэтому $x^{(1)}$=(0,75; -0,25).

Требуемая точность не достигнута, так как $\|x^{(0)}-x^{(1)}\| = 1{,}6 > 0{,}01$.

<u>Ш а г 2</u>. Как и на предыдущем шаге находим $f'(x^{(1)})$=(-1; -0,5),

$x^{(2)}=P_x$[(0,75; -0,25)-0,75(-1; -0,5)]=P_x[(1,5; 0,125)].

Точка (1,5; 0,15) допустимому множеству не принадлежит, потому что $1{,}5^2+0{,}125^2=2{,}266 > 1$. Так как допустимое множество представляет собой замкнутый шар радиуса *1* с центром в точке *O* в пространстве R^2, то

$$x^{(2)}=P_x[(1{,}5;\ 0.125)]=\frac{(1{,}5;\ 0{,}125)}{\sqrt{1{,}5^2+0{,}125^2}}=(0{,}9965;\ 0{,}08304).$$

Требуемая точность снова не достигнута, так как $\|x^{(1)}-x^{(2)}\| = 0{,}298 > 0{,}01$.

Результаты остальных шагов метода проекции градиента приведены в следующей таблице:

k	$x^{(k)}$	$\|x^{(k-1)}-x^{(k)}\|$	$f'(x^{(k)})$	$x^{(k)}-\alpha f'(x^{(k)})$
2	*(0,9965; 0,0830)*	*0,289*	*(-1; 0,1661)*	*(1,7465; -0,0415)*
3	*(0,9997; -0,0238)*	*0,107*	*(-1; -0,0476)*	*(1,7497; 0,0119)*
4	*(0,99998; 0,00769)*	*0,031*	*(-1; 0,0136)*	*(1,74998; -0,00339)*
5	*(0,99998; -0,00194)*	*0,0087*	точность	достигнута

Из таблицы следует, что $x^0 \approx x^{(5)}=(0{,}099998;\ -0{,}00194)$, $f^0 \approx f(x^{(5)}) = -1$. Отметим, что точное решение рассматриваемой задачи $f_{min}=f(1;\ 0)= -1$.

Существуют и другие варианты градиентных методов.

2. Методы возможных направлений

Идея методов возможных направлений, близкая к идее градиентных методов для задач с ограничениями, состоит в следующем: на каждой итерации определяется допустимое направление на множестве X, вдоль которого функция $f(x)$ возрастает (такое направление называется возможным направлением возрастания функции $f(x)$), и по нему совершается шаг. Фактически в методе проекции градиента и в методе условного градиента мы находим такие направления. Однако там исходным было определение градиента, а допустимое направление определялось по нему однозначно. В методах же возможных направлений исходным пунктом является описание всех допустимых направлений и выбор из них такого, вдоль которого функция $f(x)$ возрастает и желательно скорейшим образом.

Рассмотрим вариант метода возможных направлений применительно к задаче максимизации $f(x)$ на множестве (3.3), где $R=R^n$. Пусть мы имеем k-е приближение $x^{(k)}$ к решению этой задачи и для построения следующего приближения поставим следующую вспомогательную задачу: максимизировать u при ограничениях

$$\left\langle f'(x^{(k)}), a\right\rangle \geq u, \quad \left\langle g'(x^{(k)}), a\right\rangle \geq u, \quad i \in I_k, \quad \left|a^j\right| \leq 1, \quad j=\overline{1,n},$$

где

$$I_k = \left\{i \mid 1 \leq i \leq m,\ g_i(x^{(k)}) = 0\right\}, \qquad a = (a^1, \ldots, a^n).$$

Эта задача представляет собой задачу линейного программирования в $(n+1)$–мерном пространстве векторов (a, u). Множество допустимых планов замкнуто, ограничено и непусто, так как $a=0$, $u=0$ является допустимым планом. Значит, вспомогательная задача имеет решение (a^k, u^k), причем $u^k \geq 0$. Если $u^k>0$, то нетрудно показать, что направление a^k является возможным направлением возрастания функции $f(x)$, т.е. точка $x^{(k+1)}=x^{(k)}+\alpha_k a^k$ при достаточно малом α_k принадлежит множеству X и обеспечивает большее значение функции $f(x)$, чем $x^{(k)}$. Выбор пары (a^k, u^k) с возможно большим значением u^k при этом означает выбор допустимого направления, наиболее близкого к градиенту функции $f(x)$, т.е. возможного направления с наибольшим ростом функции. Если $u^k = 0$, то получается стационарная точка процесса, которая для задачи выпуклого программирования дает решение, а в общем случае требует дополнительного исследования.

Задача. *Найти минимальное значение функции* $f(x)=(x_1-7)^2+(x_2-5)^2$

при условиях $\begin{cases} -x_1 + x_2 - 3 \le 0, \\ 2x_1 + x_2 - 15 \le 0, \\ x_1,\ x_2 \ge 0. \end{cases}$

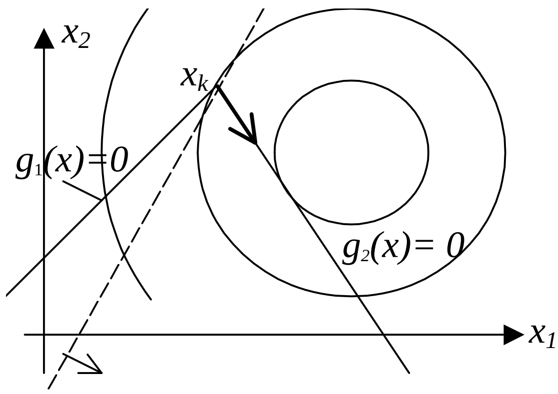

Рис. 2

Р е ш е н и е. Точку $x^{(0)}$ необходимо выбрать таким образом, чтобы она принадлежала допустимому множеству. Из нее можно сделать шаг алгоритма безусловного спуска до пересечения с границей области (если $x^{(k)}$ принадлежит границе области, то ограничения, которые выполняются, называются активными). Для определения возможного направления спуска ищут направление, удовлетворяющее активным ограничениям и составляющее минимальный угол с направлением антиградиента.

Возьмем $x^{(0)}= (4;7)$, т.к. она будет принадлежать допустимому множеству. Тогда первые два ограничения активные, а градиент

$$\nabla f(x^0)=(2\times4 - 14;2\times7 - 10) = (-6;4).$$

Для определения возможного направления спуска необходимо решить задачу линейного программирования