Общая задача нелинейного программирования

Ибо самое странное, заставляющее быть настороже в этой среде, или мире, в котором мы вынуждены существовать, состоит в том, что мир — в пределах неумолимо очерченного горизонта — всегда представляет нам разнообразные возможности действия, и нам не остается ничего, кроме как выбирать из этого разнообразия, осуществляя таким образом на практике свою свободу.

Х. Ортега-и-Гассет. *Человек и люди*

Здесь в соответствии со схемой предыдущего параграфа рассматривается общая задача нелинейного программирования с гладкими функциями.

1. Постановка задачи.

Общей задачей нелинейной условной оптимизации мы будем называть условную задачу минимизации

$$f_0(x) \to \min, \quad x \in \Omega, \tag{1}$$

где Ω выделяется как ограничениями типа равенств

$$f_i(x) = 0, \quad i = 1, ..., k, \tag{2}$$

так и ограничениями типа неравенств

$$g_j(x) \le 0, \quad j = 1, ..., l \tag{3}$$

(см. рис. 1).

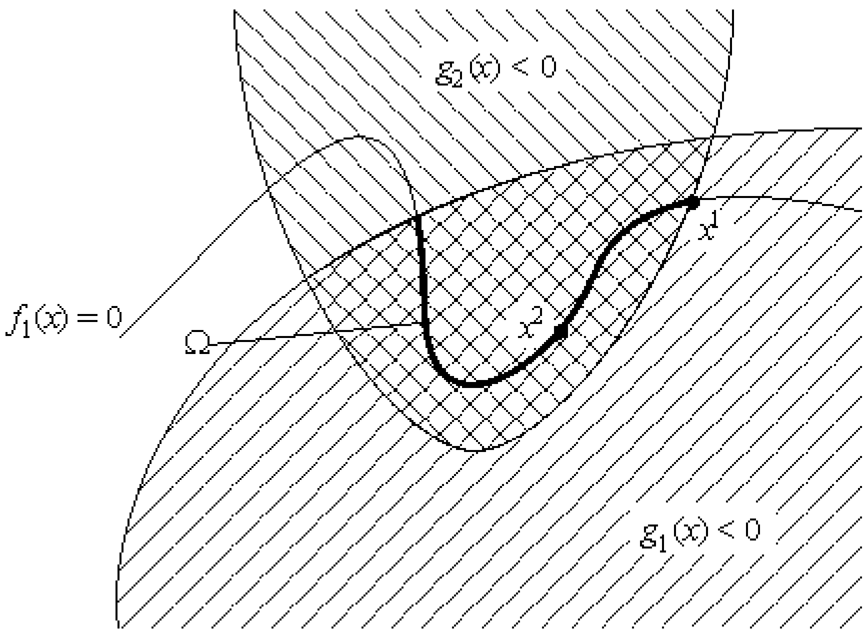

Рис. 1.

Определяются *допустимые точки*, *локальный* и *глобальный*, *строгий* и *нестрогий минимумы*. Так же мы будем использовать обозначения f и g для функций из $\mathbf{R}^m$ в $\mathbf{R}^k$ и $\mathbf{R}^l$, соответственно, определяемые координатами f_i и g_j. Поэтому задачу (1)–(3) можно записывать в виде

$$f_0(x) \to \min, \quad f(x) = \Theta, \quad g(x) \le \Theta. \tag{4}$$

(напомним, что неравенство $g(x) \le \Theta$ означает покоординатные неравенства). Нам также потребуется следующее обозначение: $a_+ = (a + |a|)/2$; таким образом, $a_+ = a$, если $a \ge 0$ и $a_+ = 0$, если $a \le 0$. Для векторов $a \in \mathbf{R}^m$ операция "+" определяется покоординатно: $a_+ = ((a_1)_+, \ldots, (a_m)_+)$. Кроме того, через $J(x)$ будет обозначаться множество индексов так называемых *активных ограничений*: $J(x) = \{j \in \{1, \ldots, l\}: g_j(x) = 0\}$ — это номера ограничений, которые в данной точке существенны. Например, на рис. 22 $J(x^1) = \{2\}$, а $J(x^2) = \varnothing$.

2. Теорема (обобщенное правило множителей Лагранжа)

Пусть $f_0, f, g \in C^1$, *а* x^* *— локальное решение задачи (4). Тогда найдутся такие* $\lambda^*_0 \in \mathbf{R}$, $\lambda^* \in \mathbf{R}^k$, $\mu^* \in \mathbf{R}^l$, *не равные одновременно нулю, такие, что* $\mu^*_j \ge 0$ *при* $j \in J(x^*)$ *и*

$$\lambda^*_0 f'_0(x^*) + \sum_{i=1}^{k} \lambda^*_i f'_i(x^*) + \sum_{j \in J(x^*)} \mu^*_j g'_j(x^*) = \Theta. \tag{5}$$

Д о к а з а т е л ь с т в о. Возьмем $r > 0$ таким, чтобы $x^* = \operatorname{argmin}_{x \in \Omega_r} f(x)$, где $\Omega_r = \Omega \cap B(x^*, r)$. Для любого $n \in \mathbf{N}$ определим функцию $f^n: \mathbf{R}^m \to \mathbf{R}$ равенством

$$f^n(x) = f_0(x) + n(\|f(x)\|^2 + \|g_+(x)\|^2) + \|x - x^*\|^2.$$

З а д а ч а 7.1. Докажите, что $f^n \in C^1$, в частности, покажите, что $(\|g_+(x)\|^2)' = 2g_+(x)g'(x)$.

Поскольку f^n непрерывно дифференцируемы и поэтому непрерывны, функции f^n на шаре $B(x^*, r)$ достигают минимума; обозначим его через x^n. В частности,

$$f^n(x^n) \le f^n(x^*), \tag{6}$$

т. е.

$$f_0(x^n) + n(\|f(x^n)\|^2 + \|g_+(x^n)\|^2) + \|x^n - x^*\|^2 \le f^n(x^*).$$

Отсюда, поскольку, как легко видеть, $f^n(x^*) = f_0(x^*)$,

$$\|f(x^n)\|^2 + \|g_+(x^n)\|^2 \le \frac{1}{n}[f_0(x^*) - \|x^n - x^*\|^2 - f_0(x^n)],$$

Выражение в квадратных скобках ограничено, так как $x^n \in B(x^*, r)$. Поэтому $\|f(x^n)\|^2 + \|g_+(x^n)\|^2 \to 0$ при $n \to \infty$ и, следовательно, $f(x^n) \to \Theta$ и $g_+(x^n) \to \Theta$ при $n \to \infty$.

Пусть теперь $\{x^n{}_i\}$ — произвольная сходящаяся, скажем к y, подпоследовательность. Тогда, как легко видеть, во-первых, $y \in \Omega$ и, во-вторых, $f^{ni}(x^{ni}) \to f_0(y) + \|y - x^*\|^2$ при $n \to \infty$. Поэтому, подставляя в (6) n_i вместо n и переходя к пределу в получившемся неравенстве при $i \to \infty$, получим

$$f_0(y) + \|y - x^*\|^2 \le f_0(x^*).$$

С другой стороны, так как $x^* = \operatorname{argmin} f(x)$, $f_0(x^*) \le f_0(y)$. Отсюда $\|y - x^*\|^2 \le 0$, т. е. $y = x^*$. Таким образом, любая сходящаяся подпоследовательность последовательности $\{x^n\}$ сходится к x^*. Последнее, учитывая компактность последовательности, гарантирует ее сходимость к тому же пределу.

Далее, в силу доказанного можно считать, что при всех n точка x^n лежит внутри шара $B(x^*, r)$. Поэтому в силу теоремы Ферма $(f^n)'(x^n) = \Theta$, т. е.

$$f'_0(x^n) + 2\sum_{i=1}^{k} f_i(x^n)f'_i(x^n) + 2\sum_{j=1}^{l} [g_j(x^n)]_+ g'_j(x^n) + 2(x^n - x^*) = \Theta. \tag{7}$$

Положим теперь

$$s^n = \left(1 + 4n^2 \sum_{i=1}^{k} [f_i(x^n)]^2 + 4n^2 \sum_{j=1}^{l} [g_j(x^n)]^2\right)^{-1/2},$$

$\lambda^n_0 = s^n$, $\lambda^n_i = 2nf_i(x^n)s^n$, $\mu^n_j = 2n[g_j(x^n)]_+s^n$. Поскольку $\lambda^n_0, \lambda^n_j, \mu^n_j \in [-1, 1]$ ($n \in \mathbf{N}$; $i = 1, \ldots, k$; $j = 1, \ldots, l$), можно считать, не ограничивая общности, что найдутся такие λ^*_0, λ^*_i и μ^*_j, что

$$\lambda^n_0 \to \lambda^*_0,\ \ \lambda^n_i \to \lambda^*_i,\ \ \mu^n_j \to \mu^*_j \ \text{ при } \ n \to \infty.$$

Умножив (7) на s^n и переходя в получившемся равенстве к пределу при $n \to \infty$, получаем

$$\lambda^*_0 f'_0(x^*) + \sum_{i=1}^{k} \lambda^*_i f'_i(x^*) + \sum_{j=1}^{l} \mu^*_j g'_j(x^*) = \Theta. \tag{8}$$

Остается заметить, что, во-первых, так как $|\lambda^n_0|^2 + \sum_{i=1}^{k} |\lambda^n_i|^2 + \sum_{j=1}^{l} |\mu^n_j|^2 = 1$, не все из $\lambda^*_0, \lambda^*_i, \mu^*_j$ равны нулю, во-вторых, поскольку $\mu^n_j = 2n[g_j(x^n)]_+s^n \geq 0$, неотрицательны и μ^*_j и, наконец, в-третьих, если $j \notin J(x^*)$ (т. е. $g_j(x^*) < 0$), то $g_j(x^n) < 0$ при больших n; поэтому $[g_j(x^*)]_+ = 0$, что влечет равенство $\mu^*_j = 0$ и, следовательно, равенство

$$\sum_{j=1}^{l} \mu^*_j g'_j(x^*) = \sum_{j \in J(x^*)} \mu^*_j g'_j(x^*).$$

Таким образом (8) — это (5).

3. Регулярный случай

Так же, как и в случае ограничений-равенств в случае общей задачи нелинейной оптимизации, необходимый признак, задаваемый теоремой 2, информативен только в случае, если $\lambda^*_0 \neq 0$. В этой ситуации, так же как и в предыдущем параграфе можно разделить (5) на λ^*_0 и, следовательно, считать его равным единице. Это позволяет ввести *функцию Лагранжа* L: $\mathbf{R}^m\times\mathbf{R}^k\times\mathbf{R}^k \to \mathbf{R}$ (*в регулярном случае*) равенством

$$\mathrm{L}(x, \lambda, \mu) = f_0(x) + (\lambda, f(x)) + (\mu, g(x)).$$

Условие регулярности в случае общей задачи выглядит сложнее. Именно, допустимая точка x называется *регулярной*, если векторы $f'_1(x),..., f'_k(x)$ линейно независимы и для некоторого ненулевого вектора $h \in \mathbf{R}^m$

$$(f'_i(x),h) = 0 \text{ при } i = 1, ..., k$$

и

$$(g'_j(x),h) < 0 \text{ при } j \in J(x).$$

Геометрически, эти условия означают, что, во-первых, вектор h является касательным к многообразию, выделяемому ограничениями-равенствами (т. е. ортогонален всем градиентам $f'_i(x)$), и, во-вторых, он образует с градиентами $g'_j(x)$ активных ограничений (указывающими, очевидно, вовне множества Ω) тупой угол (см. рис. 2).

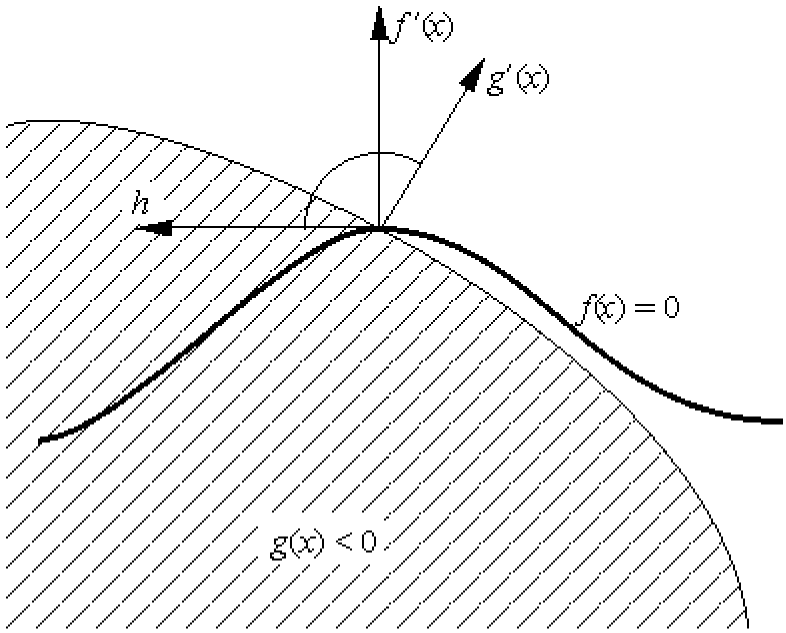

Рис. 2.

4. Теорема (обобщенное правило множителей Лагранжа в регулярном случае)

Пусть $f_0, f, g \in C^1$, *а* x^* *— регулярное локальное решение задачи (4). Тогда найдутся* $\lambda^* \in \mathbf{R}^k$, $\mu^* \in \mathbf{R}^l$ *не равные одновременно нулю, такие, что* $\mu^*_j \geq 0$ *при* $j \in J(x^*)$ *и*

$$L'_x(x^*, \lambda^*, \mu^*) = 0, \tag{9}$$

$$(\mu^*, g(x^*)) = 0. \tag{10}$$

Д о к а з а т е л ь с т в о. Пусть λ_0^{**}, λ^{**} и μ^{**} — величины, существование которых утверждается в теореме 7.2. Покажем, что $\lambda_0^{**} \neq 0$. В предположении противного умножим (5) скалярно на вектор h, фигурирующий в определении регулярности x^*:

$$\sum_{i=1}^{k} \lambda_i^{**}(f'_i(x^*), h) + \sum_{j \in J(x^*)} \mu_j^{**}(g'_j(x^*), h) = \Theta. \tag{11}$$

В силу регулярности $(f'_i(x^*), h) = 0$ и $(g'_j(x), h) < 0$, а по теореме 7.2 $\mu_j^{**} \geq 0$ при $j \in J(x^*)$. Поэтому (11) влечет равенство $\mu^{**} = \Theta$. Но тогда (5) означает, что

$$\sum_{i=1}^{k} \lambda_i^{**} f'_i(x^*) = \Theta,$$

что вместе с линейной независимостью $f'_i(x^*)$ дает равенство $\lambda^{**} = \Theta$. Противоречие.

Таким образом $\lambda_0^{**} \neq 0$. Положим теперь $\lambda^*_i = \lambda_i^{**}/\lambda_0^{**}$ ($i = 1, ..., k$),

$$\mu^*_j = \begin{cases} \mu_j^{**}/\lambda_0^{**} \text{ при } j \in J(x^*), \\ 0 \text{ при } j \notin J(x^*). \end{cases}$$

Очевидно теперь, (10) выполняется автоматически, а (9) тривиально получается из (5).

5. Достаточные условия, существование, единственость

Ситуация с задачей (1)–(3) несколько отличается от изучавшихся ранее, поскольку минимумы здесь могут быть двух типов. В случае, когда в точке минимума x^* нет активных ограничений, т. е. $J(x^*) = \varnothing$, ситуация полностью аналогична описанной в предыдущем параграфе, поскольку ограничения-неравенства в окрестности точки x^* можно опустить (см. рис. 3). В случае же, когда $J(x^*) \neq \varnothing$ минимум может достигаться на границе множества Ω и точка x^* может не быть стационарной точкой (см. рис. 3).

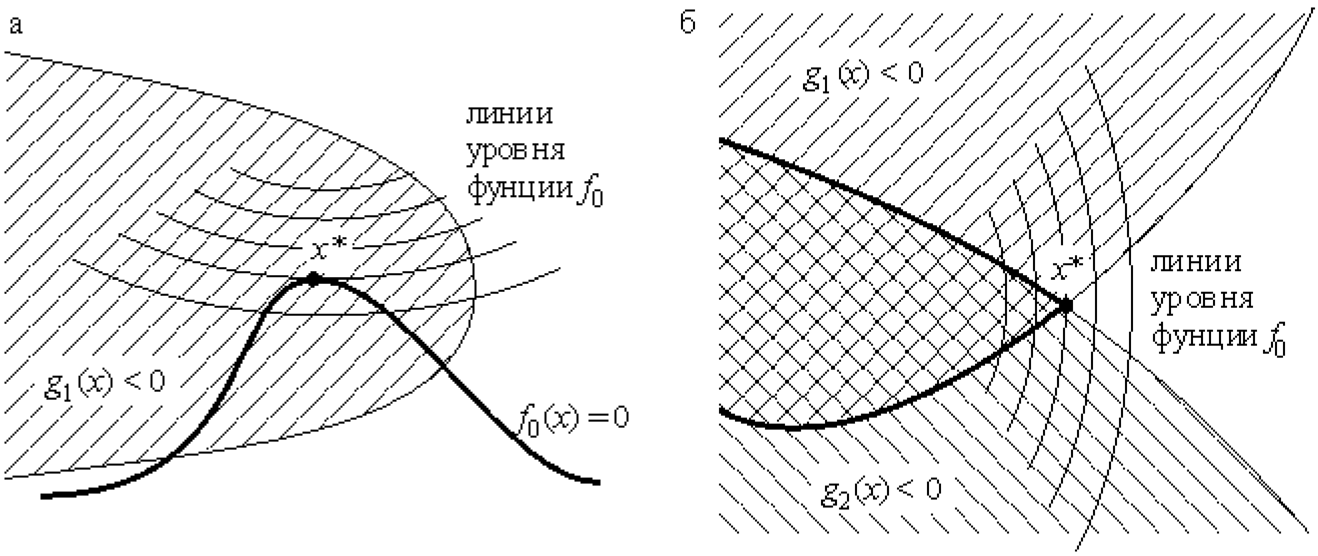

Рис. 3.

Мы ограничимся лишь формулировками результатов, поскольку в идейном плане они не доставляют ничего нового. Аналогом теоремы 6.9 может служить следующая

Теорема (достаточные условия минимума). *Пусть* $f_0, f, g \in C^2$, а x^* — *допустимая точка такая, что для некоторых* $\lambda^* \in \mathbf{R}^k$ *и* $\mu^* \in \mathbf{R}^l$, *одновременно не равных нулю, выполнены условия (9)–(10) и при любых* $h \in \mathbf{R}^m$, $h \neq \Theta$ *таких, что*

$$(f'_i(x^*), h) = 0 \ \text{ при } \ i = 1, \ldots, k;$$

$$(g'_j(x^*),h) = 0 \text{ при } j \in J(x^*),\ \mu^*_j > 0;$$

$$(g'_j(x^*),h) \geq 0 \text{ при } j \in J(x^*),\ \mu^*_j = 0$$

выполнено неравенство $(L''_{xx}(x^*,\lambda^*, \mu^*)h, h) > 0$. *Тогда* x^* — *локальное решение задачи (1)–(3)*.

З а д а ч а 7.2. Докажите.

Что касается утверждений о существовании и единственности, то ситуация с результатами, которые могут быть доказаны изученными выше приемами, полностью аналогична.

З а д а ч а 7.3. Сформулируйте и докажите аналоги утверждений задач 6.7 и 6.8 для задачи (1)–(3).

6. Об ограничениях-равенствах

С целью упрощения изложения, начиная с этого момента, мы будем рассматривать задачу условной оптимизации, содержащую только ограничения-неравенства. Это, с одной стороны, не снижает общности изложения, поскольку ограничения-равенства сводятся к ограничениям-неравенствам: ограничение

$$f(x) = \Theta$$

эквивалентно ограничениям

$$f(x) \leq \Theta, \quad -f(x) \leq \Theta.$$

С другой стороны, задачи с ограничениями-равенствами мы уже достаточно подробно рассматривали выше.

Здесь, правда, следует помнить об одном важном обстоятельстве. При решении задач с ограничениями-неравенствами часто бывает важна (существенно используется) выпуклость функций, фигурирующих в задаче, вернее, выпуклость минимизируемой функции f_0 и выпуклость множества Ω допустимых точек. Последнее гарантируется выпуклостью функций g_j в ограничениях-неравенствах (см. задачу 7.4 ниже). Множества же, выделяемые ограничениями-равенствами, не бывают выпуклыми, за исключением случая аффинных функций f_i. Поэтому методы решения задач условной оптимизации, существенно использующие "выпуклость задачи" могут применяться к задачам с ограничениями-равенствами с большой осторожностью.

Напомним, что множество $\Omega \subset \mathbf{R}^m$ называется *выпуклым*, если $\forall(x, y \in \Omega, \lambda \in [0, 1])[\lambda x + (1 - \lambda)y \in \Omega]$.

З а д а ч а 7.4. Докажите, что если функции g_i: $\mathbf{R}^m \to \mathbf{R}$ ($i = 1, ..., l$) выпуклы, то множество $\Omega = \{x \in \mathbf{R}^m\colon g_i(x) \le 0, i = 1, ..., l\}$ выпукло.

З а д а ч а 7.5. Докажите, что если функция g: $\mathbf{R}^m \to \mathbf{R}$ выпукла, вместе с функцией $-g$, то она аффинна: $g(x) = (b, x) + c$.

7.7. Еще один достаточный признак условного минимума.

Напомним, что точка (x^*, λ^*) называется *седловой точкой функции* L: $\Omega_1 \times \Omega_2 \to \mathbf{R}$ ($\Omega_1 \subset \mathbf{R}^m$, $\Omega_2 \subset \mathbf{R}^l$), если при всех $(x, \lambda) \subset \Omega_1 \times \Omega_2$ выполнены неравенства

$$L(x^*, \lambda) \le L(x^*, \lambda^*) \le L(x, \lambda^*)$$

Теорема. *Пусть* (x^*, λ^*) *— седловая точка функции Лагранжа* L: $\mathbf{R}^m \times \mathbf{R}^l_+ \to \mathbf{R}$ *задачи (1), (3) (здесь* $\mathbf{R}^l_+ = \{\lambda \in \mathbf{R}^l\colon \lambda \ge \Theta\}$*),* $\lambda^* \ge \Theta$, x^* *— допустимая точка. Тогда* x^* *— глобальное решение этой задачи.*

Д о к а з а т е л ь с т в о. Пусть x — произвольная допустимая точка, т. е. $g_i(x) \le 0$ ($i = 1, ..., l$). Тогда

$$f_0(x^*) + (\lambda^*, g(x^*)) = \mathsf{L}(x^*, \lambda^*) \le \mathsf{L}(x, \lambda^*) = f_0(x) + (\lambda^*, g(x)). \qquad (12)$$

Покажем, что

$$(\lambda^*, g(x^*)) = 0. \qquad (13)$$

Тогда из (12), поскольку $\lambda^* \geq \Theta$, а $g(x) \leq \Theta$, будет следовать нужное неравенство $f_0(x^*) \leq f_0(x)$. Для доказательства (13) заметим, что по условию теоремы $\mathsf{L}(x^*, \Theta) \leq \mathsf{L}(x^*, \lambda^*)$, т. е.

$$(\lambda^*, g(x^*)) \geq 0. \tag{14}$$

Равенство (13) вытекает теперь из (14), т.к. $\lambda^* \geq \Theta$, а $g(x^*) \leq \Theta$.

Тот факт, что доказанная теорема дает лишь достаточный признак условного минимума демонстрируетсмя в следующей задаче.

З а д а ч а 7.6. Покажите, что одномерная задача минимизации функции $f(x) = -x$ при ограничениях $x \leq 1$, $-x \leq 0$ разрешима, но ее функция Лагранжа не имеет седловых точек.

З а д а ч а 7.7. Если в предыдущей теореме (x^*, λ^*) — локальная седловая точка, то можно ли утверждать, что x^* — локальное решение задачи (1), (3)? (Ответ: нельзя).

Однако, *если функции* f_0 *и* g_i ($i = 1, ..., l$) *выпуклы, а множество* Ω *содержит хотя бы одну внутреннюю точку, то условия доказанной теоремы являются необходимыми и достаточными* (это утверждение называется **теоремой Куна — Таккера**; оно является центральным фактом теории выпуклого программирования).

Перейдем к описанию методов решения задач с ограничениями-неравенствами. Многие из описываемых ниже методов либо существенно используют выпуклость объектов, фигурирующих в задаче, либо вообще неработоспособны на невыпуклых задачах.

8. Методы возможных направлений

Эти методы основываются на следующем простом понятии. Вектор (направление) z в допустимой точке x назовем *возможным* для задачи (1), (3), если малое перемещение в этом неправлении уменьшает значение функции f_0 и не выводит из множества допустимых точек, т. е. если при достаточно малых s точка $x^s = x + sz$ допустима и $f(x^s) < f(x)$. Если теперь на каждом шаге сдвигаться в возможном направлении на достаточно малое расстояние, то мы очевидно получим релаксационную последовательность, которая во многих случаях будет сходиться к решению задачи. Методы этого типа называются *методами возможных направлений*.

З а д а ч а 7.8*. Докажите, что если $(f_0(x), z) < 0$, то малое перемещение из точки x в направлении z уменьшает значение функции f_0 (ср. с доказательством теоремы Ферма).

З а д а ч а 7.9*. Докажите, что если точка x допустима и $(g_i(x), z) < 0$ при $i \in J(x)$, то малое перемещение из точки x в направлении z не выводит из множества допустимых точек.

В силу этих задач возможное направление z^n в очередной точке x^n, в котором функция f_0 будет убывать быстрее всего, можно искать, решая задачу линейного программирования (о методах решения линейных задач см. список литературы)

$$(f'_0(x^n),z) \to \min, \tag{15}$$

$$(g'_i(x^n),z) \le 0, \quad i \in J(x^n), \tag{16}$$

$$z_i - 1 \le 0, \quad -z_i - 1 \le 0, \quad i = 1, ..., m. \tag{17}$$

Здесь "нормализующие" ограничения (16) введены для того, чтобы искать вектор возможных направлений на ограниченном множестве, т. е. для того, чтобы линейная задача (15)–(16) была разрешима. К сожалению, решение задачи (15)–(16) не всегда дает возможное направление (см. рис. 4). Причиной этого являются нестрогие неравенства в (16) (задачи же со строгими неравенствами часто оказываются неразрешимыми, поскольку точная нижняя грань значений функции f_0 часто достигается на границе множества допустимых точек, которая в случае строгих ограничений-неравенств не принадлежит этому множеству). Эту трудность можно обойти разными способами. Например, на каждом шаге можно "возвращать" точку во множество Ω с помощью какой-либо процедуры "проектирования" см., например, следующий пункт). Можно также поступать следующим образом.

Для любой допустимой точки x и произвольного $\varepsilon > 0$ определим множество $J_\varepsilon(x)$ индексов так называемых *ε-активных ограничений*:

$$J_\varepsilon(x) = \{i \in \{1, ..., m\}: g_i(x) \geq -\varepsilon\}.$$

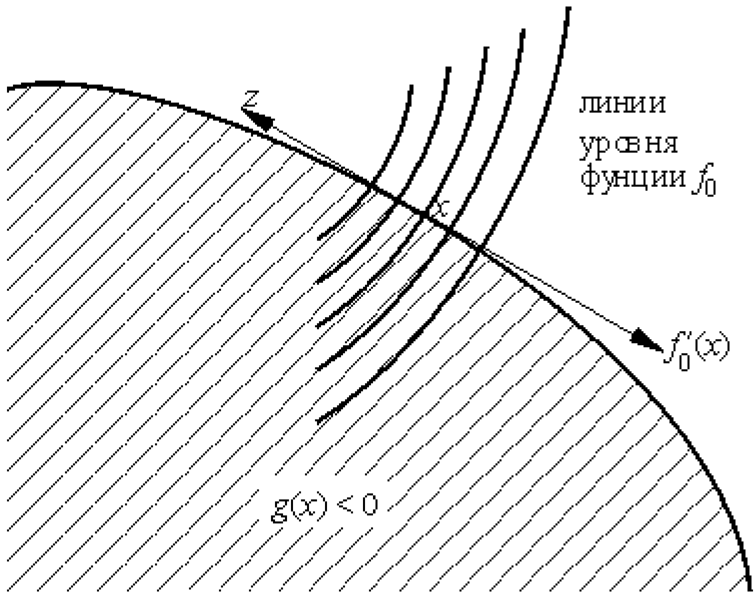

Рис. 4.

Таким образом, если ε мало́ и $i \in J_\varepsilon(x)$, то i-ое ограничение в точке x почти обращается в равенство. В одном из вариантов метода возможных направлений выбирают последовательность $\varepsilon_n \to 0$ и из очередной точки x^n следующий шаг совершют в направлении z^n:

$$x^{n+1} = x^n + \alpha^n z^n, \qquad (18)$$

где z^n есть решение следующей линейной задачи

$$\xi \to \min$$

с ограничениями

$$(f'_0(x^n), z) - \xi \leq 0,$$

$$(g'_i(x^n), z) - \xi \leq 0, \quad i \in J_{\varepsilon n}(x^n)$$

и нормализующими ограничениями (16). Подчеркнем, что в этой задаче неизвестными являются z и ξ. Длина шага в (18) выбирается таким образом, чтобы точка x^{n+1} не выходила из множества допустимых точек.

9. Методы проекции градиента

Проекцией $P_\Omega x$ *точки* $x \in \mathbf{R}^m$ *на множество* $\Omega \subset \mathbf{R}^m$ называется любая ближайшая к x точка множества Ω:

$$\|x - P_\Omega x\| \le \|x - y\| \text{ при всех } y \in \Omega.$$

З а д а ч а 7.10. Докажите, что если Ω замкнуто и выпукло, то для любой точки проекция сужествует и единственна.

З а д а ч а 7.11. Приведите пример, когда проекция: а) не существует; б) не единственна.

В тех случаях, когда проекцию точки на множество допустимых точек задачи (1), (3) найти достаточно легко (например, когда Ω — линейное подпространство, полупространство, шар, $\mathbf{R}^m_+$и т. д.) используют *метод проекции градиента*:

$$x^{n+1} = P_\Omega(x^n - \alpha^n f'_0(x^n))$$

(см. рис. 5), являющийся прямым обобщением градиентного метода.

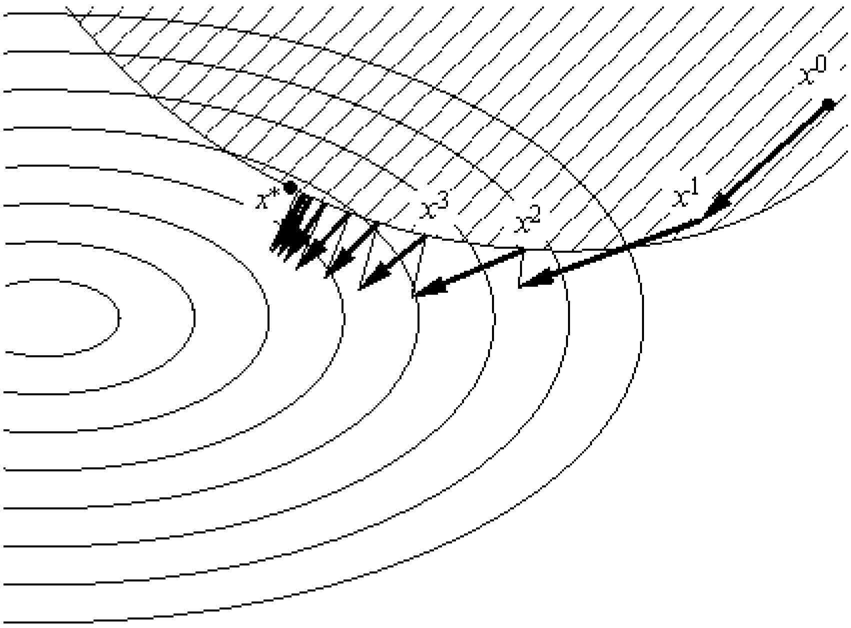

Рис. 5.

Можно доказать, например, что *если функция* $f_0 \in C^1$ *сильно выпукла и* f' *удовлетворяет условию Липшица, а множество* Ω *замкнуто и ограничено, то методпроекции градиента сходится со скоростью геометрической прогрессии*.

10. Методы линеаризации

Суть этих методов, как следует из названия, состоит в линеаризации минимизируемой функции и ограничений в очередной точке x^n строящейсярелаксационной последовательности и объявлении следующим значением x^{n+1} решения получающейся линейной задачи, т. е. задачи

$$(f'_0(x^n), x - x^n) \to \min, \tag{19}$$

$$g_i(x^n) + (g'_i(x^n), x - x^n) \le 0, \quad i = 1, ..., l. \tag{20}$$

Чтобы эта (линейная) задача была разрешима либо добавляют штраф в минимизируемую функцию, заменяя (19), например, задачей

$$(f'_0(x^n), x - x^n) + \delta\|x - x^n\|^2 \to \min,$$

либо добавляя к (20) простые ограничения, которые делают множество допустимых точек этой задачи ограниченным, например, (линейные) ограничения

$$x_i - x^n_i - \alpha^n \le 0, \quad -x_i + x^n_i - \alpha^n \le 0 \quad (i = 1, ..., m).$$

Часто для уменьшения объема вычислительной работы среди ограничений (20) оставляют только ε-активные.

11. Методы штрафов

Основная идея здесь заключается в переходе от задачи (1), (3) к задаче безусловной оптимизации, в которой "наказывается" либо удаление от множества Ωдопустимых точек (*внешний штраф*), либо приближение изнутри множества Ω к его границе (*внутренний штраф*). Например, это можно сделать так. Для любого $s > 0$ определим функцию $F_s^{out}: \mathbf{R}^m \to \mathbf{R}$ равенством (индекс "out" станет понятен чуть ниже)

$$F_s^{out}(x) = f_0(x) + s\sum_{i=1}^{l}(g_i)^2_+(x). \tag{21}$$

З а д а ч а 7.12 (ср. с задачей 6.11). Пусть $f_0, g \in C^1$, а x^* — единственное решение задачи (1), (3). Пусть, кроме того, множество $\{x \in \mathbf{R}^m: \|g_i(x)\| \le \varepsilon, i =$

1, ..., l} при некотором $\varepsilon \geq 0$ непусто и ограничено. Докажите, что (безусловная) задача

$$\mathcal{F}_s^{\text{out}}(x) \to \min \tag{22}$$

при всех s разрешима и ее решение x^s сходится к x^* при $s \to \infty$.

Геометрическая трактовка замены задачи (1), (3) задачей (22) изображена на рис. 6.

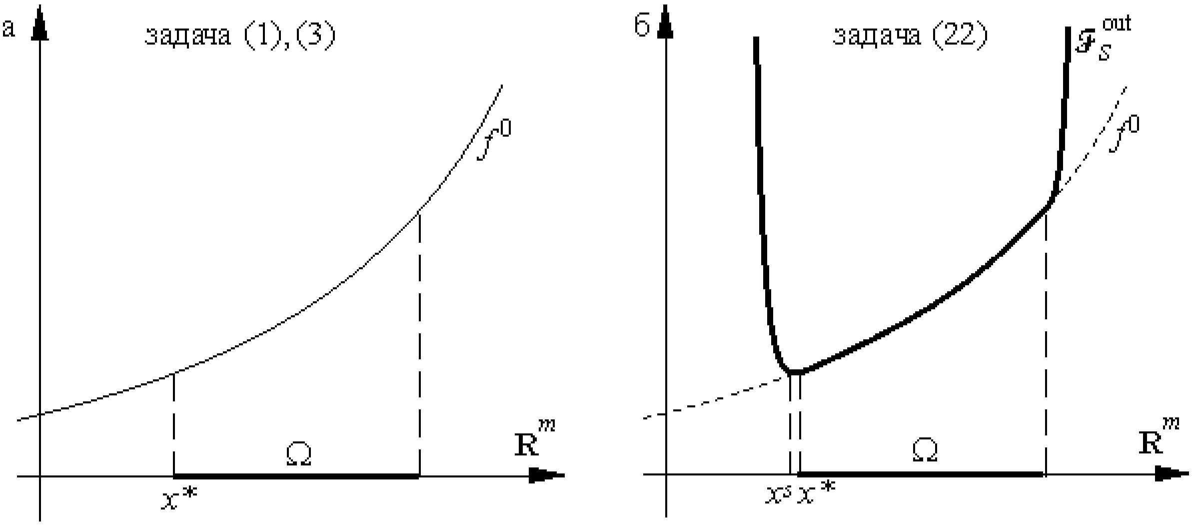

Рис. 6.

Задачу (22) решают тем или иным методом решения безусловных задач, увеличивая на каждом шаге штраф s. Как и в случае задач с ограничениями-равенствами, основным недостатком метода штрафов является рост числа обусловленности задачи (22) с ростом s.

Описанный класс методов называют *методами внешних штрафов*, поскольку минимизируемая функция изменяется вне множества Ω. На несколько другой идее основываются так называемые *методы внутренних штрафов* или *барьеров*. Образно его можно описать так: у границы множества Ω возводятся барьеры, не позволяющие приближаться к его границе. Например, вместо задачи (1), (3) можно рассмотреть задачу

$$\mathcal{F}_s^{\text{in}}(x) \to \min, \tag{23}$$

где функция $\mathcal{F}_s^{\text{in}}: \Omega_g \to \mathbf{R}$ определяется равенством (ср. с ((21))

l 1 .

$$F_s^{in}(x) = f_0(x) + s \sum_{i=1} g_i(x)$$

Сравнение геометрических интерпретаций задач (1), (3) и (23) изображено на рис. 7.

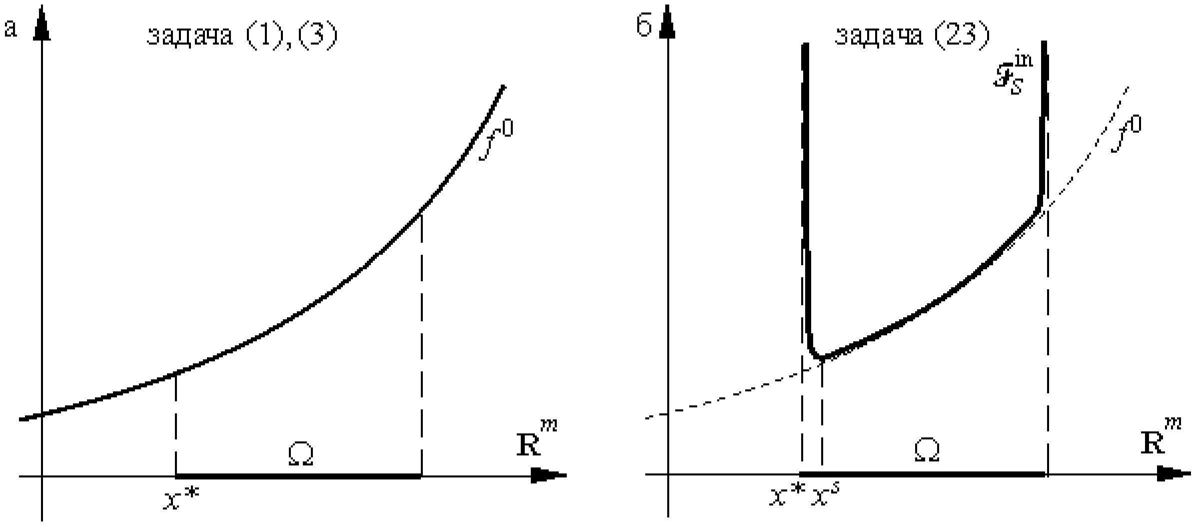

Рис. 7.

З а д а ч а 7.13. Сформулируйте и докажите аналог задачи 7.5 для метода барьеров.

Задача (23) решается методами, развитыми для безусловных задач; при этом "крутизна барьера" характеризуемая числом s, возрастает на каждом шаге.

З а д а ч а 7.14. Попытайтесь выписать аналоги штрафных функций F_s^{out}и F_s^{in}для общей задачи нелинейного программирования (1)–(3), содержащей как ограничения-неравенства, так и ограничения-равенства.

12. Двойственные методы

Суть методов из этого весьма обширного класса в применении к задачам с ограничениями-неравенствами, так же как и в задачах с ограничениями-равенствами, заключается в замене условной задачи (1), (3) задачей поиска стационарной (часто седловой) точки функции Лагранжа. В выпуклом случае задача (1), (3) и задача о поиске седловой точки функции Лагранжа в силе теоремы Куна – Таккера эквивалентны.

Например, применение простейшего (градиентного) метода к последней задаче приводит к следующему *методу Эрроу – Гурвица – Удзавы*:

$$x^{n+1} = x^n - \alpha^n L'_x(x^n, \mu^n),$$

$$\mu^{n+1} = [\mu^n + \alpha^n L'_\mu(x^n, \mu^n)]_+.$$

Для поиска седловой точки в двойственных методах применяют почти весь спектр описанных выше методов минимизации — методы Ньютона, квазиньютоновские методы, методы сопряженных направлений и т. д.