Лекции по методам оптимизации

А. В. Плясунов

10 января 2006 г.

Оглавление

Предисловие

02.09.2005

Самая главная формула в этой науке такова:

$$\begin{aligned}\text{Методы оптимизации} &\equiv \text{Теория оптимизации} \equiv\\ &\equiv \text{Теория экстремальных задач} \equiv\\ &\equiv \text{Математическое программирование}\end{aligned}$$

А теперь, когда вы уже почти всё знаете, откуда появилась эта дисциплина. Её актуальность связана с появлением ЭВМ, и знакомой вам теорией *NP-полных задач*, которая возникла в связи с феноменом *переборных задач.* Эти задачи можно решить с помощью перебора всех возможных вариантов, число которых растет *экспоненциально* в зависимости от *размеров* задачи.

Для большинства таких задач неизвестно действительно *эффективных* алгоритмов решения. Они вам известны под названием *NP-полных* или *NP-трудных задач.*

Визитной карточкой большей части интересных экстремальных задач, возникающих в чистой и прикладной математике, является их принадлежность этим классам.

Приложения

- ▷ в исследовании операций: оптимизация технико-экономических систем, транспортные задачи, управление и животноводство;
- ▷ в численном анализе: аппроксимация, решение линейных и нелинейных задач и тому подобное;
- ▷ в автоматике: оптимальное управление системами, управление производством, роботы и так далее;
- ▷ в технике: управление размерами и оптимизация структур, оптимальное планирование сложных технических систем, например, информационных систем, сетей трубопроводов, компьютеров и тому подобное;

- ▷ в прикладной математике:
    - ○ теория игр: невозможна без понятия седловой точки, методов решения экстремальных задач;
    - ○ в численном анализе: распространение основных конечномерных алгоритмов на функциональные пространства дает инструмент для изучения уравнений в частных производных или задач оптимального управления;
    - ○ в комбинаторной оптимизации многие базовые алгоритмы: в задачах о потоках, на графах, на матроидах, целочисленного и булевого программирования и многих других, используют необходимые условия экстремума, понятия двойственности, дополнительности и так далее, полученные в математическом программировании.

Цели лекционного курса

- Изучение ряда базовых алгоритмов, которые используются для решения *конечномерных задач оптимизации.*
- Получение (приобретение) теоретических и концептуальных представлений, достаточных для понимания, оценки этих алгоритмов и, если необходимо, создания новых.

Глава 1

Теория экстремальных задач

ОПР 1.1 (Экстремальной (оптимизационной) задачи).

Экстремальная задача (P) — *это задача нахождения минимума функции* $f(\vec{x})$ *при условии, что*

$$\vec{x} \in S \subseteq \mathbb{R}^n \text{ или } S \subseteq \mathbb{Z}^n \text{ или } S \subseteq \mathfrak{B}^n \text{ и } \varphi_i(\vec{x}) \leqslant 0 - i = 1, 2, \ldots, m$$

($\mathfrak{B}$ — *бинарное множество, то есть* $\mathfrak{B} = \{0, 1\}$), *где* $\vec{x} = (x_1, x_2, \ldots, x_n)$ — вектор переменных, *а* f — целевая функция задачи.

Условия $\varphi_i(\vec{x}) \leqslant 0$, $i = 1, 2, \ldots, m$, $\vec{x} \in S$ — *называются* ограничениями задачи.

ОПР 1.2 (Допустимого решения).

Любой вектор $\vec{x}$, *удовлетворяющий ограничениям задачи, называется* допустимым решением задачи.

Обозначим

$$Q(P) = \left\{\vec{x} \in \mathbb{R}^n \middle| \varphi_i(x) \leqslant 0, i = 1, 2, \ldots, m, \vec{x} \in S\right\}$$

— множество допустимых решений *задачи* P.

ОПР 1.3 (Оптимального решения).

Любое допустимое решение задачи, на котором достигается минимум целевой функции f *на множестве* $Q(P)$, *называется* оптимальным решением (глобальным минимумом).

Замечание 1.3.1 (Приведение к экстремальной задаче).

- ▷ Ограничение-равенство $g(x) = 0$ эквивалентно двум неравенствам:

$$g(x) \leqslant 0 \text{ и } -g(x) \leqslant 0.$$

- ▷ Задача максимизации функции g на множестве Q сводится к задаче минимизации функции $f = -g$ на этом же множестве.

1.4 Классификация задач

- ▷ В зависимости от природы множества S задачи оптимизации классифицируются как:
  - *дискретные* (*комбинаторные*) — S конечно или счетно;
  - *целочисленные* — $\vec{x} \in S \subseteq \mathbb{Z}^n$;
  - *булевы* — $\vec{x} \in S \subseteq \mathfrak{B}^n$;
  - *вещественные* (*непрерывные*) — $\vec{x} \in S \subseteq \mathbb{R}^n$;
  - *бесконечномерные* — S подмножество гильбертова пространства.
- ▷ Если множество S совпадает с основным пространством $\mathbb{R}^n$, $\mathbb{Z}^n$ или $\mathfrak{B}^n$, а ограничения φ_i отсутствуют ($m = 0$), то задачу P называют *задачей безусловной оптимизации*. В противном случае говорят о *задаче условной оптимизации*.
- ▷ Если принять во внимание свойства целевой функции f и ограничений φ_i, то возникает более тонкое деление конечномерных экстремальных задач на классы:
  - *непрерывное математическое программирование*, то есть f и φ_i — непрерывные, произвольные и нелинейные, а S — связное, компактное подмножество $\mathbb{R}^n$;
  - *дискретное математическое программирование*, то есть f и φ_i — нелинейные, а S — дискретное множество;
  - *нелинейное целочисленное программирование*, то есть f и φ_i — нелинейные, а $S \subseteq \mathbb{Z}^n$;
  - *непрерывная нелинейная оптимизация без ограничений*, то есть f — непрерывная, произвольная и нелинейная функция; $m = 0$, $S = \mathbb{R}^n$;
  - *целочисленная нелинейная оптимизация без ограничений*, то есть f — произвольная и нелинейная функция; $m = 0$, $S = \mathbb{Z}^n$;
  - *выпуклое программирование*, то есть f и φ_i — произвольные и выпуклые, а S — выпуклое множество из $\mathbb{R}^n$;
  - *линейное программирование*, то есть f и φ_i — произвольные и линейные, а $S = \left\{\vec{x} \in \mathbb{R}^n \middle| \mathtt{A}\vec{x} \leqslant \vec{b}\right\}$;
  - *целочисленное линейное программирование*, то есть f и φ_i — произвольные и линейные, а $S \subseteq \mathbb{Z}^n$;

Пример 1.4.1 (Определения классификации задачи).

- ▷ Если f и φ_i — произвольные и нелинейные функции, а S — произвольное подмножество $\mathbb{R}^n$, то получим задачу нелинейного программирования и так далее.

1.5 Теорема Фаркаша-Минковского

ОПР 1.5.1 (Ортогонального подпространства)**.**
$L \subseteq \mathbb{R}^n$ — линейное подпространство $\mathbb{R}^n$, *если*

$$\forall \vec{x},\, \vec{y} \in L\colon \forall \alpha,\, \beta \in \mathbb{R}\colon \alpha\vec{x} + \beta\vec{y} \in L.$$

Подпространство ортогональное к L — *это подпространство вида*

$$L^{\perp} \overset{def}{=} \left\{ \vec{y} \in \mathbb{R}^n \middle| \forall \vec{x} \in L\colon (\vec{y},\, \vec{x}) = 0 \right\}$$

(то есть скалярное произведение двух любых векторов из разных подпространств — нуль).

ОПР 1.5.2 (Подпространства оператора)**.**
Пусть $\mathtt{A}$ — *матрица (оператор) порядка* $(m \times n)$, *тогда* подпространство оператора — *это*

$$L(\mathtt{A}) \overset{def}{=} \left\{ \vec{z} \in \mathbb{R}^m \middle| \exists \vec{x} \in \mathbb{R}^n\colon \mathtt{A}\vec{x} = \vec{z} \right\}.$$

Аналогично определяется подпространство, ортогональное подпространству оператора*:*

$$L(\mathtt{A})^{\perp} \leftrightharpoons \left\{ \vec{y} \in \mathbb{R}^m \middle| \forall \vec{z} \in L(\mathtt{A})\colon (\vec{y},\, \vec{z}) = 0 \right\} = \\ = \left\{ \vec{y} \in \mathbb{R}^m \middle| \forall \vec{x} \in \mathbb{R}^n\colon (\vec{y},\, \mathtt{A}\vec{x}) = 0 \right\} = \left\{ \vec{y} \in \mathbb{R}^m \middle| \vec{y}\mathtt{A} = 0 \right\}.$$

Лемма 1.5.3 (Связь совместных и несовместных систем уравнений)**.**

▷ Пусть

Система уравнений

$$\begin{cases} \vec{y}\mathtt{A} & = 0, \\ (\vec{y},\, \vec{b}) & > 0 \end{cases}$$

— несовместна.

▷ Тогда

Совместна система уравнений $\mathtt{A}\vec{x} = \vec{b}$.

▷ **Доказательство**.

- Пусть система уравнений $\mathtt{A}\vec{x} = \vec{b}$ тоже несовместна; из определения операции ортогонализации $(\perp)$ имеем

$$L(\mathtt{A}) = \left(L(\mathtt{A})^{\perp}\right)^{\perp},$$

следовательно,

$$L(\mathtt{A}) = \left\{ \vec{z} \in \mathbb{R}^m \middle| \forall \vec{y} \in L(\mathtt{A})^{\perp}\colon (\vec{z},\, \vec{y}) = 0 \right\} = \\ = \left\{ \vec{z} \in \mathbb{R}^m \middle| \forall \vec{y} \middle| \vec{y}\mathtt{A} = 0\colon (\vec{z},\, \vec{y}) = 0 \right\}.$$

- Так как $\vec{b} \notin L(\mathtt{A})$, то найдется такой $\vec{y}$, что $\vec{y}\mathtt{A} = 0$ и $(\vec{b},\, \vec{y}) \neq 0$. Следовательно, либо $\vec{y}$, либо $-\vec{y}$ — решение системы уравнений $\vec{y}\mathtt{A} = 0$, $(\vec{y},\, \vec{b}) > 0$. Противоречие с условиями леммы, следовательно предположение неверно и система уравнений $\mathtt{A}\vec{x} = \vec{b}$ совместна.

□

Теорема 1.5.4 (Фаркаша-Минковского)**.**

▷ Система уравнений $\mathtt{A}\vec{x} = \vec{b}$, $\vec{x} \geqslant 0$ разрешима в том и только в том случае, когда неравенство $(\vec{b},\, \vec{y}) \leqslant 0$ выполняется для всех решений системы уравнений $\vec{y}\mathtt{A} \leqslant 0$.

▷ **Доказательство**.

- Необходимость: пусть $\exists \vec{x} \big| \mathtt{A}\vec{x} = \vec{b}$, $\vec{x} \geqslant 0$ и пусть $\vec{y}$ — произвольное решение системы $\vec{y}\mathtt{A} \leqslant 0$, тогда

$$(\vec{b},\, \vec{y}) = (\mathtt{A}\vec{x},\, \vec{y}) = (\vec{x},\, \vec{y}\mathtt{A}) \leqslant (\vec{x},\, 0) = 0.$$

- Достаточность: пусть неравенство $(\vec{b},\, \vec{y}) \leqslant 0$ выполняется для всех решений системы неравенств $\vec{y}\mathtt{A} \leqslant 0$, индукцией по числу столбцов матрицы $\mathtt{A}$ покажем, что совместна система уравнений $\mathtt{A}\vec{x} = \vec{b}$, $\vec{x} \geqslant$ $\geqslant 0$.

  ✓ Для $n = 1$ доказать утверждение самостоятельно.

  ✓ Пусть $n > 1$ и для матриц с числом столбцов строго меньшим n утверждение доказано. Поскольку

$$\forall \vec{y}\colon \vec{y}\mathtt{A} \leqslant 0 \Rightarrow (\vec{y},\, \vec{b}) \leqslant 0,$$

то несовместна система уравнений

$$\vec{y}\mathtt{A} \leqslant 0,\, (\vec{y},\, \vec{b}) > 0$$

и, следовательно, несовместна система уравнений

$$\vec{y}\mathtt{A} = 0,\, (\vec{y},\, \vec{b}) > 0.$$

Отсюда и леммы 1.5.3 получим, что совместна система уравнений $\mathtt{A}\vec{x} = \vec{b}$, покажем, что тогда

$$\exists \vec{x} \big| \mathtt{A}\vec{x} = \vec{b},\, \vec{x} \geqslant 0:$$

итак,

$$\sum_{j=1}^{n} \vec{A}_j \cdot x_j = \vec{b}, \text{ где } \vec{A}_j = (a_{1j},\, a_{2j},\, \ldots,\, a_{mj})^{\mathsf{T}}.$$

— j-ый столбец матрицы $\mathtt{A}$. Пусть $x_j \geqslant 0$ для $j = 1, 2, \ldots, k$ и $x_j < 0$ для $j = k+1, k+2, \ldots, n$, рассмотрим вектор

$$\vec{c} = \sum_{j=1}^{k} \vec{A}_j x_j + \vec{A}_n x_n = \vec{b} - \sum_{j=k+1}^{n-1} \vec{A}_j x_j; \tag{1.1}$$

допустим, что найдется такой $\vec{y}$, что:

$$(\vec{y}, \vec{A}_j) \leqslant 0,\ j = 1, 2, \ldots, n-1,\ (\vec{y}, \vec{c}) > 0. \tag{1.2}$$

Отсюда и определения (1.1) вектора $\vec{c}$ получим, что

$$\sum_{j=1}^{k} (\vec{y}, \vec{A}_j) \cdot x_j + (\vec{y}, \vec{A}_n) \cdot x_n = (\vec{y}, \vec{c}) > 0,$$

таким образом

$$(\vec{y}, \vec{A}_n) \cdot x_n > \sum_{j=1}^{k} -(\vec{y}, \vec{A}_j) \cdot x_j \geqslant 0.$$

Учитывая что $x_n < 0$, имеем

$$(\vec{y}, \vec{A}_n) \leqslant 0. \tag{1.3}$$

С другой стороны, из определения (1.1) вектора $\vec{c}$ получим

$$(\vec{y}, \vec{b}) = (\vec{y}, \vec{c}) + \sum_{j=k+1}^{n-1} (\vec{y}, \vec{A}_j) \cdot x_j > 0,$$

отсюда и условий (1.2) и (1.3) следует, что вектор $\vec{y}$ — решение системы уравнений:

$$\vec{y}\mathtt{A} \leqslant 0,\ (\vec{y}, \vec{b}) > 0.$$

Получили противоречие (эта система несовместна по условию). Таким образом система (1.2) не имеет решения, но по условию разрешима система $\vec{y}\mathtt{A} \leqslant 0$, следовательно $(\vec{y}, \vec{c}) \leqslant 0$ для всех решений системы

$$(\vec{y}, \vec{A}_j) \leqslant 0,\ j = 1, 2, \ldots, n-1.$$

Тогда по предположению индукции найдется вектор $\vec{z} \geqslant 0$:

$$\sum_{j=1}^{n-1} \vec{A}_j \cdot z_j = \vec{c},$$

отсюда и определения (1.1) на стр. 15 вектора $\vec{c}$ получим

$$\sum_{j=1}^{n-1} \vec{A}_j \cdot z_j = \vec{b} - \sum_{j=k+1}^{n-1} \vec{A}_j \cdot x_j,$$

что приводит к соотношению

$$\sum_{j=1}^{k} \vec{A}_j \cdot z_j + \sum_{j=k+1}^{n-1} \vec{A}_j \cdot (x_j + z_j) + \vec{A}_j \cdot 0 = \vec{b}.$$

Итак, число отрицательных компонент нового решения

$$\vec{x} = (z_1, z_2, \ldots, z_k, x_{k+1} + z_{k+1}, \ldots, x_{n-1} + z_{n-1}, 0)$$

системы равенств $\mathtt{A}\vec{x} = \vec{b}$ строго меньше, чем у вектора $\vec{x}$, следовательно, через конечное число шагов получим неотрицательное решение системы равенств $\mathtt{A}\vec{x} = \vec{b}$.

□

Замечание 1.5.4.1 (Другими словами).

▷ Теореме Фаркаша–Минковского можно придать другую форму: система уравнений $\mathtt{A}\vec{x} \leqslant \vec{b}$ разрешима в том и только в том случае когда неравенство $(\vec{b}, \vec{y}) \geqslant 0$ выполняется для всех решений системы уравнений $\vec{y}\mathtt{A} = 0,\ \vec{y} \geqslant \vec{0}$.

Действительно, система уравнений $\mathtt{A}\vec{x} \leqslant \vec{b}$ разрешима $\Leftrightarrow$ разрешима система уравнений $\mathtt{A}\vec{x}_1 - \mathtt{A}\vec{x}_2 + \mathtt{E}\vec{u} = \vec{b},\ \vec{x}_1,\ \vec{x}_2,\ \vec{u} \geqslant \vec{0}$. По теореме Фаркаша-Минковского последняя система уравнений разрешима $\Leftrightarrow$ когда неравенство $(\vec{b}, \vec{y}) \leqslant \vec{0}$ выполняется для всех решений системы уравнений $\vec{y}\mathtt{A} \leqslant 0,\ -\vec{y}\mathtt{A} \leqslant 0,\ \mathtt{E}\vec{y} \leqslant \vec{0}$ $\Leftrightarrow$ неравенство $(\vec{b}, \vec{y}) \geqslant 0$ выполняется для всех решений системы уравнений $\vec{y}\mathtt{A} = \vec{0},\ \vec{y} \geqslant \vec{0}$.

Следствие 1.5.5 (Теорема Гордана).

▷ Имеет место одно и только одно из следующих двух условий:

1. разрешима система уравнений $\mathtt{A}\vec{x} < \vec{0}$;
2. существует такой $\neq \vec{0}$ вектор $\vec{y}$, что $\vec{y}\mathtt{A} = \vec{0},\ \vec{y} \geqslant \vec{0}$.

▷ **Доказательство**.

- Действительно, система уравнений $\mathtt{A}\vec{x} < \vec{0}$ разрешима $\Leftrightarrow$ разрешима система уравнений $\mathtt{A}\vec{x} \leqslant (-1, -1, \ldots, -1)^\top$. По теореме Фаркаша-Минковского разрешимость последней системы эквивалентна выполнению следующего условия: если вектор $\vec{y}$ решение системы $\vec{y}\mathtt{A} = \vec{0},\ \vec{y} \geqslant \vec{0}$, то выполняется неравенство $-\vec{y} \geqslant \vec{0}$. То есть не существует ненулевого вектора $\vec{y}$ такого, что $\vec{y}\mathtt{A} = \vec{0},\ \vec{y} \geqslant \vec{0}$.

- Теперь пусть существует ненулевой вектор $\vec{y}$ такой, что $\vec{y}\mathtt{A} = \vec{0}$, $\vec{y} \geqslant \vec{0}$. Но тогда не выполняется неравенство $-\vec{y} \geqslant \vec{0} \Rightarrow$ не выполнено условие теоремы Фаркаша-Минковского $\Rightarrow$ система

$$\mathtt{A}\vec{x} \leqslant (-1,\, -1,\, \ldots,\, -1)^{\top}$$

неразрешима $\Rightarrow$ неразрешима система $\mathtt{A}\vec{x} < 0$.

□

09.09.2005

09.09.2005

1.6 Необходимые условия экстремума

Далее будем работать со следующей задачей нелинейного программирования:

ОПР 1.6.1 (Задачи нелинейного программирования)**.**
Задачей нелинейного программирования *называется задача нахождения минимума функции $f(\vec{x})$ при условии, что*

$$\forall i = 1,\, 2,\, \ldots,\, m\colon \varphi_i(\vec{x}) \leqslant 0.$$

Здесь f и $\varphi_i\colon \mathbb{R}^n \to \mathbb{R}$ и они $\in \mathfrak{C}^1$, это непрерывно дифференцируемые функции.

ОПР 1.6.2 (Возможного направления)**.**
Направление $\vec{s}$ *(произвольный ненулевой вектор) в точке $\vec{x} \in Q$ называется* возможным*, если существует такое число $\bar{\beta}$, что*

$$\forall \beta \in [0,\, \bar{\beta}]\colon \vec{x} + \beta\vec{s} \in Q.$$

Напомню, что множество K называется конусом*, если $\forall \lambda \big|\; f(\lambda) \geqslant 0$ и $\forall \vec{x} \in K$ имеем $\lambda \cdot \vec{x} \in K$. Очевидно, что множество возможных направлений в точке $\vec{x}$ образует конус, который обозначим как $K_f(\vec{x})$.*

ОПР 1.6.3 (Активного ограничения)**.**
Ограничение φ_i называется активным *в точке $\vec{x}$, если $\varphi_i(\vec{x}) = 0$. Обозначим через $I(\vec{x})$ множество номеров ограничений активных в данной точке.*

Лемма 1.6.4 (Условие возможности направления)**.**

▷ Пусть

Вектор $\vec{s} \neq \vec{0}$ удовлетворяет системе

$$\big(\varphi_i'(\vec{x}),\, \vec{s}\big) + \sigma \leqslant 0,\; i \in I(\vec{x}),$$

при некотором $\sigma > 0$.

▷ Тогда

Направление $\vec{s}$ является возможным в точке $\vec{x}$.

▷ **Доказательство**.

- Можно считать, что $I(\vec{x}) \neq \varnothing$, так как иначе $\vec{x}$ — внутренняя точка Q (напомним, что Q — множество, которое удовлетворяем всем ограничениям по определению) и тогда любое направление $\vec{s} \neq 0$ является возможным.
- Если $i \notin I(\vec{x})$, то малое перемещение не нарушает строгое ограничение $\varphi_i(\vec{x}) < 0$, то есть найдется подходящее $\bar{\beta}$.

  Пусть теперь $i \in I(\vec{x})$ (то есть $\varphi_i(\vec{x}) = 0$) и допустим, что $\forall \beta > 0\colon \varphi_i(\vec{x} + \beta\vec{s}) > 0$, тогда получаем, что при $\beta \to 0$:

$$\frac{\varphi_i(\vec{x} + \beta\vec{s})}{\beta} = \frac{\big(\varphi_i(\vec{x} + \beta\vec{s}) - \varphi_i(\vec{x})\big)}{\beta} \to \big(\varphi_i'(\vec{x}),\, \vec{s}\big) \geqslant 0.$$

  Получили противоречие, так как по условию леммы $(\varphi_i'(\vec{x}),\, \vec{s}) < 0$.

□

Замечание 1.6.4.1 (Формулировка в терминах конусов).

▷ Пусть

$$K_<(\vec{x}) = \{\vec{s} \,\big|\, \vec{s} \neq \vec{0} \text{ и } (\vec{\varphi}_i'(\vec{x}), \vec{s}) < 0 - \forall i \in I(\vec{x})\},$$

из этого определения следует, что множество $K_<(\vec{x})$ — конус, поэтому лемма 1.6.4 утверждает, что конус $K_<(\vec{x})$ является подмножеством конуса $K_f(\vec{x})$, оправдывая название первого из них, как внутренней аппроксимации второго — конуса возможных направлений.

Конус

$$K_d(\vec{x}) = \{\vec{s} \neq \vec{0} \,\big|\, (f_i'(\vec{x}), \vec{s}) < 0\}$$

называется *конусом направлений убывания функции f*.

Теорема 1.6.5 (Геометрическая форма необходимых условий оптимальности)**.**

▷ Для того, чтобы точка $\vec{x} \in Q$ являлась точкой локального минимума функции f на множестве Q необходимо, чтобы для любого решения $(\vec{s},\, \sigma)$ системы

$$\begin{cases} \forall i \in I(\vec{x})\colon (\varphi_i'(\vec{x}),\, \vec{s}) + \sigma \leqslant 0, \\ (f'(\vec{x}),\, \vec{s}) + \sigma \leqslant 0, \end{cases} \tag{1.4}$$

выполнялось условие

$$\sigma \leqslant 0. \tag{1.5}$$

▷ **Доказательство.**

◦ Пусть $\vec{x}$ — локальный минимум функции f на множестве Q и $\exists(\vec{s},\, \sigma)$ такая, что выполняются условия (1.4) и $\sigma > 0$. Отсюда и из леммы 1.6.4 получаем, что $\vec{s}$ — возможное направление в точке $\vec{x}$.

Так как $\vec{f}'(\vec{x})$ — непрерывная функция и $\big(f'(\vec{x}),\, \vec{s}\big) \leqslant -\sigma < 0$, то существует достаточно малое число $\beta > 0$ такое, что

$$\big(\vec{f}'(\vec{x}+\beta\vec{s}),\, \vec{s}\big) < 0,\ \vec{x}+\beta\vec{s} \in Q.$$

Тогда по теореме о среднем получаем

$$f(\vec{x}+\beta\vec{s}) - f(\vec{x}) = \beta\cdot\big(\vec{f}'(\vec{x}+\theta\beta\vec{s}),\, \vec{s}\big) < 0,\ \text{где } 0 \leqslant \theta \leqslant 1,$$

то есть $f(\vec{x})$ больше $f(\vec{x}+\beta\vec{s})$, что немножко противоречит с тем, что $\vec{x}$ — локальный минимум функции f.

□

Замечание 1.6.5.1 (В терминах конусов).

▷ Теорема 1.6.5 эквивалентна утверждению, что для локальной оптимальности точки $\vec{x}$ необходимо выполнение условия

$$K_d(\vec{x}) \cap K_<(\vec{x}) = \varnothing.$$

Название теоремы связано с тем, что в ней необходимые условия формулируются в виде требования пустоты пересечения двух геометрических объектов — двух конусов. Эту теорему можно сформулировать в более сильном варианте:

$$K_d(\vec{x}) \cap \overline{K}_f(\vec{x}) = \varnothing,$$

который, однако, неконструктивен, так как отсутствует аналитическое описание замыкания конуса возможных направлений. Впоследствии мы увидим, что наличие дополнительной информации о задаче приводит к нужному описанию этого конуса. Эта информация будет формулироваться в виде так называемых *условий регулярности.*

Теорема 1.6.6 (Необходимые условия Фритца-Джона)**.**

▷ Пусть

$\vec{x}^*$ — локальный экстремум задачи 1.1 на стр. 11, функции f и φ_i — $\forall i = 1,\, 2,\, \ldots,\, m$ непрерывны и непрерывно дифференцируемы.

▷ Тогда

Найдутся такие не все равные нулю множители $\lambda_i \geqslant 0$ для всех $i = 0,\, 1,\, \ldots,\, m$, что

$$\lambda_0\cdot\vec{f}'(\vec{x}^*) + \sum_{i\in I(\vec{x})} \lambda_i\cdot\vec{\varphi}_i'(\vec{x}^*) = \vec{0}^n \text{ и } \forall i = 1,\, 2,\, \ldots,\, m\colon \lambda_i\varphi_i(\vec{x}^*) = 0. \quad (1.6)$$

▷ **Доказательство.**

◦ Перепишем неравенство (1.5) на стр. 18 теоремы 1.6.5 ($\sigma \leqslant 0$) в следующем виде:

$$(0,\, \vec{s}) + 1\cdot\sigma \leqslant 0.$$

Из условия непрерывной дифференцируемости получаем, что если вектор $(\vec{s},\, \sigma)$ удовлетворяет системе уравнений (1.4) на стр. 18:

$$\big(s_1,\, s_2,\, \ldots,\, s_n,\, \sigma\big)\cdot\begin{pmatrix}\vec{f}'(\vec{x}^*) & \ldots & \vec{\varphi}_i'(\vec{x}^*) & \ldots\\ 1 & \ldots & 1 & \ldots\end{pmatrix} \leqslant 0,$$

то выполняется неравенство (1.5)

$$\big(s_1,\, s_2,\, \ldots,\, s_n,\, \sigma\big)\cdot\big(0,\, 0,\, \ldots,\, 0,\, 1\big)^{\mathsf{T}} \leqslant 0.$$

По теореме Фаркаша-Минковского 1.5.4 на стр. 14 получается, что существует неотрицательное решение системы уравнений:

$$\begin{pmatrix}\vec{f}'(\vec{x}^*) & \ldots & \vec{\varphi}_i'(\vec{x}^*) & \ldots\\ 1 & \ldots & 1 & \ldots\end{pmatrix}\cdot\big(\lambda_0,\, \lambda_1,\, \ldots,\, \lambda_i,\, \ldots\big)^{\mathsf{T}} = \big(0,\, 0,\, \ldots,\, 0,\, 1\big)^{\mathsf{T}};$$

следовательно

$$\lambda_0\cdot\vec{f}'(\vec{x}^*) + \sum_{i\in I(\vec{x}^*)} \lambda_i\vec{\varphi}_i'(\vec{x}^*) = \vec{0}^n,$$

$$\lambda_0 + \sum_{i\in I(x^*)} \lambda_i = 1.$$

Положим $\forall i \notin I(x^*)\colon \lambda_i = 0$ и получим

$$\lambda_0\cdot\vec{f}'(\vec{x}^*) + \sum_{i=1}^{m} \lambda_i\cdot\vec{\varphi}_i'(\vec{x}^*) = \vec{0}^n \text{ и } \forall i = 1,\, 2,\, \ldots,\, m\colon \lambda_i\varphi_i(\vec{x}^*) = 0.$$

□

Замечание 1.6.6.1 (Исторические названия переменных).

- ▷ Необходимые условия Фритца-Джона дают уже чисто алгебраическую характеризацию локальных экстремумов как решений системы уравнений. Компоненты решения $\lambda_i \geqslant 0$, $i = 0, 1, \ldots, m$, системы (1.6) называют *множителями Лагранжа локального оптимума* $\vec{x}^*$. Соотношения $\lambda_i \varphi_i(\vec{x}^*) = 0$ называют *соотношениями дополняющей нежёсткости*, в следствии которых множители λ_i соответствующие неактивным ограничениям $i \notin I(\vec{x}^*)$ равны нулю.

  Множитель λ_i характеризует *чувствительность* данного локального оптимума $\vec{x}^*$ относительно малых изменений величины $\varphi_i(\vec{x}^*)$.

Перейдём к выводу классических условий оптимальности Куна-Таккера, которые совпадают с необходимыми условиями Фритца-Джона затем лишь исключением, что в них $\lambda_0 \neq 0$.

Теорема 1.6.7 (Необходимые условия оптимальности Куна-Таккера)**.**

- ▷ Пусть
  - ∘ $\vec{x}^*$ — локальный экстремум задачи 1.1 на стр. 11;
  - ∘ функции f, φ_i, $i = 1, 2, \ldots, m$, — непрерывны и непрерывно дифференцируемы;
  - ∘ вектора $\vec{\varphi}_i'(\vec{x}^*)$, $i \in I(\vec{x}^*)$ — линейно независимы.
- ▷ Тогда

  Тогда найдутся такие множители $\lambda_i \geqslant 0$ — $\forall i = 1, 2, \ldots, m$, что

$$-\vec{f}'(\vec{x}^*) = \sum_{i \in I(\vec{x})} \lambda_i \cdot \vec{\varphi}_i'(\vec{x}^*),$$

$$\forall i = 1, 2, \ldots, m\colon \lambda_i \cdot \varphi_i(\vec{x}^*) = 0.$$

- ▷ **Доказательство**.
  - ∘ Из теоремы Фритца-Джона 1.6.6 на стр. 19 следует, что найдутся такие, не все равные нулю множители $\lambda_i \geqslant 0$, $i = 1, 2, \ldots, m$, что

$$\lambda_0 \cdot \vec{f}'(\vec{x}^*) + \sum_{i \in I(\vec{x})} \lambda_i \cdot \vec{\varphi}_i'(\vec{x}^*) = \vec{0}^n,$$

$$\forall i = 1, 2, \ldots, m\colon \lambda_i \varphi_i(\vec{x}^*) = 0.$$

  Также мы получили равенство

$$\lambda_0 + \sum_{i \in I(\vec{x}^*)} \lambda_i = 1.$$

  - ∘ Покажем, что $\lambda_0 \neq 0$: пусть $\lambda_0 = 0$, из последнего равенства получаем, что существует $\lambda_i \neq 0$, следовательно $\left\{\vec{\varphi}_i'(\vec{x}^*)\right\}_{i \in I(\vec{x}^*)}$ — линейно зависимые вектора. Получили противоречие с условием, значит $\lambda_0 > 0$.

□

Замечание 1.6.7.1 (На языке конусов).

- ▷ Для понимания смысла, использованного в теореме условия регулярности, что градиенты $\vec{\varphi}_i'(\vec{x}^*)$, $i \in I(\vec{x}^*)$ линейно независимы, введём конус

$$K_{\leqslant}(\vec{x}) = \{\vec{s} \neq \vec{0} \,|\, (\vec{\varphi}_i'(\vec{x}),\, \vec{s}) \leqslant 0,\ \forall i \in I(\vec{x})\},$$

  который назовём *внешней аппроксимацией конуса возможных направлений*. Из определения указанных конусов следует, что $K_f(\vec{x}) \subseteq K_{\leqslant}(\vec{x})$.

  Верно следующее утверждение: если $K_<(\vec{x}) \neq \varnothing$, то $\overline{K}_<(\vec{x}) = K_{\leqslant}(\vec{x})$. Действительно, пусть конус $K_<(\vec{x})$ не пуст, тогда найдётся $\vec{s}$ такой, что

$$(\vec{\varphi}_i'(\vec{x}),\, \vec{s}) < \vec{0} — \forall i \in I(\vec{x}).$$

  Возьмём произвольный элемент $\vec{s}$ из внешней аппроксимации конуса возможных направлений, то есть

$$(\vec{\varphi}_i'(\vec{x}),\, \vec{s}) \leqslant 0 — \forall i \in I(\vec{x}).$$

  Очевидно, что для любого $\lambda \in [0, 1)$:

$$\lambda \cdot \vec{s} + (1 - \lambda) \cdot \vec{s} \in K_<(\vec{x}).$$

  Таким образом $\vec{s}$ — предел последовательности направлений из $K_<(\vec{x})$ при λ стремящимся к единице снизу. Учитывая, что

$$K_<(\vec{x}) \subseteq K_{\leqslant}(\vec{x}),$$

  получим требуемое равенство

$$\overline{K}_<(\vec{x}) = K_{\leqslant}(\vec{x}).$$

  Так как

$$K_<(\vec{x}) \subseteq K_f(\vec{x}) \subseteq K_{\leqslant}(\vec{x}),$$

  то при условии $K_<(\vec{x}) \neq \varnothing$ получим исчерпывающее аналитическое описание замыкания конуса возможных направлений

$$\overline{K}_f(\vec{x}) = K_{\leqslant}(\vec{x}) = \{\vec{s} \neq \vec{0} \,|\, (\vec{\varphi}_i'(\vec{x}),\, \vec{s}) \leqslant \vec{0} — \forall i \in I(\vec{x})\}.$$

Теперь, основываясь на этой информации, можно дать интерпретацию условий регулярности теоремы 1.6.7: линейная независимость градиентов активных ограничений $\vec{\varphi}_i'(\vec{x}^*)$— $i \in I(\vec{x}^*)$ означает, что не существует таких ненулевых коэффициентов λ_i, $i \in I(\vec{x}^*)$, что

$$\sum_{i \in I(\vec{x}^*)} \lambda_i \cdot \vec{\varphi}_i'(\vec{x}^*) = 0.$$

Отсюда при помощи теоремы Гордана 1.5.5 на стр. 16 (следствие теоремы Фаркаша-Минковского) выводим, что найдется такой вектор $\vec{s}$, что

$$(\vec{\varphi}_i'(\vec{x}^*),\, \vec{s}) < 0 — \forall i \in I(\vec{x}^*).$$

То есть конус $K_<(\vec{x}^*)$ является не пустым, следовательно

$$\overline{K}_f(\vec{x}^*) = \{\vec{s} \neq 0 \,|\, (\vec{\varphi}_i'(\vec{x}^*),\, \vec{s}) \leqslant 0 — \forall i \in I(\vec{x}^*)\}.$$

ОПР 1.6.8 (Задачи выпуклого программирования)**.**
Задача 1.1 на стр. 11 называется задачей выпуклого программирования, *если функции f и φ_i для $i = 1, 2, \ldots, m$ — выпуклые. Соответственно, получаем что множество Q — выпукло.*

По прежнему считаем, что f, $\varphi_i \in \mathfrak{C}^1$.

ОПР 1.6.9 (Условия регулярности и условия регулярности Слейтора)**.**
Условие регулярности *для выпуклого случая:*

$$\forall i = 1, 2, \ldots, m\colon \exists \vec{x}^i \in Q \,\big|\, \varphi_i(\vec{x}^i) < 0,$$

что эквивалентно условию регулярности Слейтора

$$\exists \widetilde{\vec{x}} \in Q \,\big|\, \forall i = 1, 2, \ldots, m\colon \varphi_i(\widetilde{\vec{x}}) < 0.$$

Лемма 1.6.10 (Условие выпуклости функции)**.**

▷ Функция f, дифференцируемая на выпуклом множестве Q, выпукла в том и только в том случае, когда для любых $\vec{x}, \vec{y} \in Q$: $\big(\vec{f}'(\vec{x}),\, \vec{y} - \vec{x}\big) \leqslant f(\vec{y}) - f(\vec{x})$.

▷ <u>Доказательство</u>.

∘ $\forall \vec{x} \neq \vec{y} \in Q\colon \forall \alpha \,\big|\, 0 < \alpha \leqslant 1$:

$$f\big(\vec{x} + \alpha(\vec{y} - \vec{x})\big) \leqslant f(\vec{x}) + \alpha \cdot \big(f(\vec{y}) - f(\vec{x})\big)$$

— из выпуклости, перепишем неравенство таким образом

$$\|\vec{y} - \vec{x}\| \cdot \frac{f(\vec{x} + \beta \cdot \vec{s}) - f(\vec{x})}{\beta} \leqslant f(\vec{y}) - f(\vec{x}),$$

где $\vec{s} = (\vec{y} - \vec{x})/\|\vec{y} - \vec{x}\|$, $\beta = \alpha \cdot \|\vec{y} - \vec{x}\|$. Устремим β к нулю и в пределе получим

$$\frac{\partial f}{\partial s} \cdot \big\|\vec{y} - \vec{x}\big\| \leqslant f(\vec{y}) - f(\vec{x}),$$

но

$$\frac{\partial f}{\partial s} \cdot \big\|\vec{y} - \vec{x}\big\| = \big(\vec{f}'(\vec{x}),\, \vec{s}\big) \cdot \big\|\vec{y} - \vec{x}\big\| = \big(\vec{f}'(\vec{x}),\, \vec{y} - \vec{x}\big).$$

Итак, мы доказали неравенство

$$\big(\vec{f}'(\vec{x}),\, \vec{y} - \vec{x}\big) \leqslant f(\vec{y}) - f(\vec{x}).$$

∘ В обратную сторону: пусть $\forall \vec{x}, \vec{y} \in Q\colon \big(\vec{f}'(\vec{x}),\, \vec{y} - \vec{x}\big) \leqslant f(\vec{y}) - f(\vec{x})$, но по условию $\forall \alpha \in [0, 1]\colon \vec{z} = \alpha \cdot \vec{x} + (1 - \alpha) \cdot \vec{y} \in Q$. Умножим неравенство $\big(\vec{f}'(\vec{z}),\, \vec{x} - \vec{z}\big) \leqslant f(\vec{x}) - f(\vec{z})$ на α, а неравенство $\big(\vec{f}'(\vec{z}),\, \vec{y} - \vec{z}\big) \leqslant f(\vec{y}) - f(\vec{z})$ на $(1 - \alpha)$ и сложим их:

$$0 = \big(\vec{f}'(\vec{z}),\, \alpha \cdot (\vec{x} - \vec{z}) + (1 - \alpha) \cdot (\vec{y} - \vec{z})\big) \leqslant \\ \leqslant \alpha \cdot f(\vec{x}) + (1 - \alpha) \cdot f(\vec{y}) - f(\vec{z}),$$

следовательно

$$f\big(\alpha \cdot \vec{x} + (1 - \alpha) \cdot \vec{y}\big) \leqslant \alpha \cdot f(\vec{x}) + (1 - \alpha) \cdot f(\vec{y}).$$

□

Замечание 1.6.10.1 (На языке конусов).

▷ Теперь нетрудно показать, что условие регулярности Слейтера также гарантирует непустоту конуса $K_<(\vec{x})$ и, следовательно, $\overline{K}_f(\vec{x}) = K_{\leqslant}(\vec{x})$. Действительно, из леммы 1.6.10 и условия Слейтера следует, что для любого $i \in I(\vec{x})$

$$0 > \varphi_i(\tilde{\vec{x}}) = \varphi_i(\tilde{\vec{x}}) - \varphi_i(\vec{x}) \geqslant (\varphi_i'(\vec{x}),\, \tilde{\vec{x}} - \vec{x}).$$

Таким образом вектор $\vec{s} = (\tilde{\vec{x}} - \vec{x}) \in K_<(\vec{x})$ и требуемое равенство следует из доказанного ранее утверждения.

Лемма 1.6.11 (Достаточные условия возможного направления)**.**

▷ <u>Пусть</u>

$$Q = \big\{\vec{x} \,\big|\, \forall i = 1, 2, \ldots, m\colon \varphi_i(\vec{x}) = (\vec{a}_i,\, \vec{x}) - b_i \leqslant 0\big\}.$$

▷ Тогда

Условия
$$(\vec{a}_i,\, \vec{s}) \leqslant 0 — i \in I(\vec{x}^*)$$
необходимы и достаточны для того, чтобы направление $\vec{s}$ было возможным в точке $\vec{x}^* \in Q$.

▷ **Доказательство**.

- Пусть $\beta > 0$, рассмотрим
$$\varphi_i(\vec{x}^* + \beta\vec{s}) = (\vec{a}_i,\, \vec{x}^* + \beta\vec{s}) - b_i = (\vec{a}_i,\, \vec{x}^*) - b_i + \beta\cdot(\vec{a}_i,\, \vec{s}),$$
но по условию
$$\forall i \notin I(\vec{x}^*)\colon (\vec{a}_i,\, \vec{x}^*) - b_i \leqslant 0 \Rightarrow \forall i \notin I(\vec{x}^*)\colon \varphi_i(\vec{x}^* + \beta s) \leqslant 0$$
для достаточно малых β. Дальше перепишем
$$\forall i \in I(\vec{x}^*)\colon \varphi_i(\vec{x}^* + \beta\cdot\vec{s}) = \big(\vec{a}_i,\, \vec{x}^* + \beta\cdot\vec{s}\big) - b_i = \beta(\vec{a}_i,\, \vec{s}),$$
следовательно
$$\forall\beta > 0\colon \vec{x}^* + \beta\cdot\vec{s} \in Q \Leftrightarrow \forall i \in I(\vec{x}^*)\colon (\vec{a}_i,\, \vec{s}) \leqslant 0.$$

□

Эта лемма позволяет элиминировать условие Слейтера в задаче выпуклого программирования в случае линейных ограничений. Помним: функции f и φ_i — выпуклые и непрерывно-дифференцируемые; а множество допустимых решений Q удовлетворяет условию Слейтера.

Теорема 1.6.12 (Условия оптимальности Куна-Таккера: выпуклый случай)**.**

▷ Пусть

Точка $\vec{x}^*$ — локальный минимум задачи 1.1 на стр. 11.

▷ Тогда

Найдутся такие множители $\lambda_i \geqslant 0,\ i = 1, 2, \ldots, m$, что
$$-\vec{f}'(\vec{x}^*) = \sum_{i\in I(\vec{x})} \lambda_i\cdot\vec{\varphi}_i'(\vec{x}^*),\ \ \forall i = 1, 2, \ldots, m\colon \lambda_i\varphi_i(\vec{x}^*) = 0.$$

▷ **Доказательство**.

- Из теоремы 1.6.6 на стр. 19 следует, что найдутся такие, не все равные нулю множители $\lambda_i \geqslant 0$, $i = 0, 1, \ldots, m$, что
$$\lambda_0\cdot\vec{f}'(\vec{x}^*) + \sum_{i\in I(\vec{x})} \lambda_i\cdot\vec{\varphi}_i'(\vec{x}^*) = \vec{0}^n, \tag{1.7}$$
$$\forall i = 1, 2, \ldots, m\colon \lambda_i\varphi_i(\vec{x}^*) = 0.$$
Также мы получили равенство
$$\lambda_0 + \sum_{i\in I(\vec{x}^*)} \lambda_i = 1.$$
Покажем, что $\lambda_0 > 0$: пусть $\lambda_0 = 0$, тогда из предыдущего равенства следует, что $\sum\limits_{i\in I(\vec{x}^*)} \lambda_i = 1$; из условия Слейтера получаем, что
$$\exists\vec{z} \in Q\,\big|\, \forall i = 1, 2, \ldots, m\colon \varphi_i(\vec{z}) < 0.$$

- Из леммы 1.6.10 на стр. 23 получаем, что
$$\forall i \in I(\vec{x}^*)\colon \big(\vec{\varphi}_i'(\vec{x}^*),\, \vec{z} - \vec{x}^*\big) \leqslant \varphi_i(\vec{z}) - \varphi_i(\vec{x}^*) = \varphi_i(\vec{z}) < 0.$$
Умножим (1.7) скалярно на вектор $\vec{s} = \vec{z} - \vec{x}^*$, тогда получим
$$\sum_{i\in I(\vec{x}^*)} \lambda_i\cdot\big(\vec{\varphi}_i'(\vec{x}^*),\, \vec{s}\big) = 0, \tag{1.8}$$
но $\sum_{i\in I(\vec{x}^*)} \lambda_i = 1$ и $\forall i \in I(\vec{x}^*)\colon \lambda_i \geqslant 0$, $\big(\varphi_i'(\vec{x}^*),\, \vec{s}\big) < 0$, откуда получается, что
$$\sum_{i\in I(\vec{x}^*)} \lambda_i\cdot\big(\vec{\varphi}_i'(\vec{x}^*),\, \vec{s}\big) < 0.$$

Получили противоречие с формулой (1.8), следовательно $\lambda_0 > 0$.

□

16.09.2005

Далее рассмотрим задачу *с линейными ограничениями*, то есть
$$Q = \big\{\vec{x}\,\big|\,\forall i = 1, 2, \ldots, m\colon \varphi_i(\vec{x}) = (\vec{a}_i,\, \vec{x}) - b_i \leqslant 0\big\};$$
условие Слейтера не используется, $f \in \mathfrak{C}^1$ — выпуклая.

Теорема 1.6.13 (Условия оптимальности Куна-Таккера: линейный случай)**.**

▷ Пусть

Точка $\vec{x}^*$ — локальный минимум задачи 1.1 на стр. 11.

▷ Тогда

Найдутся такие множители $\lambda_i \geqslant 0$, $i = 1, 2, \ldots, m$, что

$$-\vec{f}'(\vec{x}^*) = \sum_{i \in I(\vec{x})} \lambda_i \cdot \vec{a}_i \text{ и } \forall i = 1, 2, \ldots, m\colon \lambda_i \varphi_i(\vec{x}^*) = 0.$$

▷ **Доказательство**.

- Из условия локального минимума найдется такое $\varepsilon > 0$, что

$$\forall \vec{y} \in B(\vec{x}^*, \varepsilon) \cap Q\colon f(\vec{x}^*) < f(\vec{y}),$$

где $B(\vec{x}^*, \varepsilon)$ — шар с радиусом ε. Пусть $\vec{z} \in Q$, но $\vec{z} \neq \vec{x}^*$, тогда:

1. Для малых $\alpha > 0$ выполнено

$$\alpha \cdot \vec{z} + (1 - \alpha) \cdot \vec{x}^* = \vec{x}^* + \alpha \cdot (\vec{z} - \vec{x}^*) \in B(\vec{x}^*, \varepsilon) \cap Q,$$

следовательно

$$\big(\vec{f}'(\vec{x}^*),\, \vec{z} - \vec{x}^*\big) = \lim_{\alpha \to 0} \frac{f\big(\vec{x}^* + \alpha(\vec{z} - \vec{x}^*)\big) - f(\vec{x}^*)}{\alpha} \geqslant 0. \qquad (1.9)$$

2. Формула (1.9) выполняется для любого возможного направления $\vec{s}$, так как для подходящего $\vec{z} \in Q$ выполнено $\vec{s} = \vec{z} - \vec{x}^*$.
3. Из предыдущего пункта и леммы 1.6.11 на стр. 24 следует, что

$$\forall \vec{s} \neq 0,\ \forall i \in I(\vec{x}^*)\colon (\vec{a}_i, \vec{s}) \leqslant 0,$$

а отсюда, следует неравенство $(-f'(\vec{x}^*), \vec{s}) \leqslant 0$. Дальше применяем теорему Фаркаша-Минковского 1.5.4 на стр. 14, но это уже вы сами сделайте.

□

УПР 1.6.13.1 (К теореме)**.**

▷ Провести соответствующие выкладки в теореме самостоятельно.

Теорема 1.6.14 (Правило множителей Лагранжа)**.**

▷ Пусть

- $\vec{x}^*$ — локальный экстремум задачи 1.1 на стр. 11;
- функции f и φ_i для $i = 1, 2, \ldots, m$ — непрерывны и непрерывно дифференцируемы;
- вектора $\vec{\varphi}_i'(\vec{x}^*)$, $i = 1, 2, \ldots, m$ — линейно независимы.

▷ Тогда

Найдутся такие множители λ_i, $i = 1, 2, \ldots, m$, что

$$\vec{f}'(\vec{x}^*) = \sum_{i=1}^{m} \lambda_i \cdot \vec{\varphi}_i'.$$

1.7 Критерии оптимальности

Теорема 1.7.1 (Куна-Таккера в локальной форме)**.**

▷ Точка $\vec{x}^* \in Q$ — оптимальное решение задачи выпуклого программирования в том и только в том случае, когда существуют такие числа $\lambda_i \geqslant 0$, $i = 1, 2, \ldots, m$, что

$$-\vec{f}'(\vec{x}^*) = \sum_{i \in I(\vec{x})} \lambda_i \cdot \vec{\varphi}_i'(\vec{x}^*) \text{ и } \forall i = 1, 2, \ldots, m\colon \lambda_i \varphi_i(\vec{x}^*) = 0.$$

▷ **Доказательство**.

- Необходимость: следует из теоремы 1.6.12 на стр. 25.
- Достаточность: множество допустимых решений Q — выпукло, следовательно, $\forall \vec{y} \in Q$ вектор $\vec{s} = \vec{y} - \vec{x}^*$ — возможное направление в точке $\vec{x}^*$. Значит,

$$\forall i \in I(\vec{x}^*)\colon \big(\vec{\varphi}_i'(\vec{x}^*),\, \vec{s}\big) \leqslant 0,$$

но f — выпуклая функция и из леммы 1.6.10 на стр. 23 имеем

$$\forall \vec{y} \in Q\colon f(\vec{y}) - f(\vec{x}^*) \geqslant \big(\vec{f}_i'(\vec{x}^*),\, \vec{y} - \vec{x}^*\big) = \big(\vec{f}_i'(\vec{x}^*),\, \vec{s}\big) = \\ = \left(-\sum_{i \in I(\vec{x})} \lambda_i \cdot \vec{\varphi}_i'(\vec{x}^*),\, \vec{s}\right) = \sum_{i \in I(\vec{x})} \big(-\lambda_i\big)\big(\vec{\varphi}_i'(\vec{x}^*),\, \vec{s}\big) \geqslant 0.$$

□

Если все ограничения линейны, то условие Слейтера в предыдущей теореме можно опустить.

Теорема 1.7.2 (Условие оптимальности)**.**

▷ Точка $\vec{x}^* \in Q$ — оптимальное решение задачи выпуклого программирования с линейными ограничениями в том и только в том случае, когда существуют такие числа $\lambda_i \geqslant 0$, $i = 1, 2, \ldots, m$, что

$$-\vec{f}'(\vec{x}^*) = \sum_{i=1}^{m} \lambda_i \vec{a}_i \text{ и } \lambda_i \varphi_i(\vec{x}^*) = 0,\ i = 1, 2, \ldots, m.$$

▷ **Доказательство**.

- Необходимость: следует из теоремы 1.6.13 на стр. 26.

- Достаточность: аналогично теореме 1.6.14 на стр. 27.

□

ОПР 1.7.3 (Функции Лагранжа)**.**

Функцию

$$L\big(\vec{x},\,\vec{\lambda}\big) = f(\vec{x}) + \sum_{i=1}^{m} \lambda_i \varphi_i(\vec{x}),$$

определенную при всех $\vec{x}$ *и* $\vec{\lambda}$, *назовем* функцией Лагранжа *для задачи 1.1 на стр. 11.*

ОПР 1.7.4 (Седловой точки функции Лагранжа)**.**

Пара $\big(\vec{x}^*,\,\vec{\lambda}^*\big)$ *называется* седловой точкой функции Лагранжа, *если* $\forall \vec{x} \in \mathbb{R}^n$, $\forall \vec{\lambda} \geqslant 0$:

$$L\big(\vec{x}^*,\,\lambda\big) \leqslant L\big(\vec{x}^*,\,\lambda^*\big); \tag{1.10}$$

$$L\big(\vec{x}^*,\,\lambda^*\big) \leqslant L\big(\vec{x},\,\lambda^*\big). \tag{1.11}$$

Пусть для задачи выпуклого программирования будет выполнено условие Слейтера, тогда верно следующее утверждение:

Теорема 1.7.5 (Куна-Таккера в нелокальной форме)**.**

▷ Допустимое решение $\vec{x}^* \in Q$ является оптимальным тогда и только тогда, когда существует такой вектор $\vec{\lambda}^*$, что пара $(\vec{x}^*,\,\vec{\lambda}^*)$ является седловой точкой функции Лагранжа.

▷ **Доказательство**.

- Необходимость: пусть $\vec{x}^* \in Q$ — оптимальное решение, тогда из теоремы 1.6.12 на стр. 25 имеем существует такое вектор $\vec{\lambda}^* \geqslant 0$, что

$$\vec{f}'(\vec{x}^*) + \sum_{i \in I(\vec{x})} \lambda_i^* \cdot \vec{\varphi}_i'(\vec{x}^*) = 0;$$

$$\lambda_i^* \cdot \varphi_i(\vec{x}^*) = 0,\ i = 1,\,2,\,\ldots,\,m.$$

✓ Фиксируем $\vec{\lambda}$, тогда функция

$$\vec{x} \to L\big(\vec{x},\,\vec{\lambda}\big) = f(\vec{x}) + \sum_{i \in I(\vec{x})} \lambda_i \cdot \varphi_i(\vec{x})$$

выпукла, отсюда по лемме 1.6.10 на стр. 23, получаем

$$L\big(\vec{x},\,\vec{\lambda}^*\big) \geqslant L\big(\vec{x}^*,\,\vec{\lambda}^*\big) + \left(\frac{\partial L(\vec{x},\,\vec{\lambda})}{\partial \vec{x}} \bigg|_{(\vec{x}^*,\,\vec{\lambda}^*)},\ \vec{x} - \vec{x}^* \right) =$$

$$\text{так как } \frac{\partial L(\vec{x},\,\vec{\lambda})}{\partial \vec{x}} \bigg|_{(\vec{x}^*,\,\vec{\lambda}^*)} = \vec{f}'(\vec{x}^*) + \sum_{i=1}^{m} \lambda_i^* \vec{\varphi}_i'(\vec{x}^*) = 0$$

$$= L\big(\vec{x}^*,\,\vec{\lambda}^*\big),$$

получили неравенство (1.11) определения седловой точки.

✓ Фиксируем $\vec{x}$, тогда функция

$$\vec{\lambda} \to L\big(\vec{x},\,\vec{\lambda}\big) = f(\vec{x}) + \sum_{i=1}^{m} \lambda_i \varphi_i(\vec{x})$$

линейна; таким образом, из леммы 1.6.10 на стр. 23 и равенства $\frac{\partial L(\vec{x},\,\vec{\lambda})}{\partial \vec{\lambda}} \Big|_{(\vec{x}^*,\,\vec{\lambda}^*)} = \varphi(\vec{x}^*)$ получаем

$$L\big(\vec{x}^*,\,\vec{\lambda}^*\big) \geqslant L\big(\vec{x}^*,\,\vec{\lambda}\big) + \big(\varphi(\vec{x}^*),\,\vec{\lambda}^* - \vec{\lambda}\big) =$$
$$= L\big(\vec{x}^*,\,\vec{\lambda}\big) + \big(\varphi(\vec{x}^*),\,\vec{\lambda}^*\big) - \big(\varphi(\vec{x}^*),\,\vec{\lambda}\big) \geqslant L\big(\vec{x}^*,\,\vec{\lambda}\big)$$

(последнее неравенство следует из дополняющей нежёсткости $\big(\varphi(\vec{x}^*),\,\vec{\lambda}^*\big) = 0$ и неравенств $\vec{\lambda} \geqslant \vec{0}$, $\varphi(\vec{x}^*) \leqslant 0$).

Таким образом мы получили неравенство (1.10) из определения 1.7.4 на стр. 29 и, следовательно, $(\vec{x}^*,\,\vec{\lambda}^*)$ — седловая точка функции Лагранжа.

- Достаточность: пусть $(\vec{x}^*,\,\vec{\lambda}^*)$ — седловая точка функции L, из неравенства (1.11) на стр. 29 получаем:

$$\forall \vec{x} \in R^n\colon L\big(\vec{x}^*,\,\vec{\lambda}^*\big) \leqslant L\big(\vec{x},\,\vec{\lambda}^*\big) \Rightarrow$$
$$\Rightarrow L\big(\vec{x}^*,\,\vec{\lambda}^*\big) = \min_{\vec{x} \in R^n} L\big(\vec{x},\,\vec{\lambda}^*\big). \tag{1.12}$$

А из неравенства (1.10) на стр. 29:

$$\forall \vec{\lambda} \geqslant \vec{0}\colon L\big(\vec{x}^*,\,\vec{\lambda}\big) \leqslant L\big(\vec{x}^*,\,\vec{\lambda}^*\big) \Rightarrow$$
$$\Rightarrow \forall \vec{\lambda} \geqslant \vec{0}\colon \sum_{i=1}^{m} \lambda_i \varphi_i(\vec{x}^*) \leqslant \sum_{i=1}^{m} \lambda_i^* \varphi_i(\vec{x}^*) < \infty, \tag{1.13}$$

откуда следует, что $\vec{x}^*$ — допустимое решение[I)], значит

$$\sum_{i=1}^{m} \lambda_i^* \varphi_i(\vec{x}^*) \leqslant 0, \text{ но } \sum_{i=1}^{m} \lambda_i^* \varphi_i(\vec{x}^*) \geqslant 0^{\text{II)}}.$$

[I)] От противного: пусть $\exists i_0 \big| \varphi_{i_0}(\vec{x}^*) > 0$, тогда $\sum_{i=1}^{m} \lambda_i \varphi_i(\vec{x}^*) \to +\infty$ при $\lambda_{i_0} \to +\infty$, но это противоречит с неравенством (1.13).

Так как равенство $\sum_{i=1}^{m} \lambda_i^* \varphi_i(\vec{x}^*) = 0$ эквивалентно условиям дополняющей нежёсткости $\forall i\colon \lambda_i^* \varphi_i(\vec{x}^*) = 0$, то отсюда и равенства (1.12) получим

$$\forall \vec{x} \in Q\colon f(\vec{x}^*) \leqslant f(\vec{x}) + \sum_{i=1}^{m} \lambda_i^* \varphi_i(\vec{x}) \leqslant f(\vec{x}),$$

значит, $\vec{x}^*$ — оптимальное решение.

□

[II] Положить в неравенстве (1.13) $\forall i\colon \lambda_i = 0$.

Глава 2

Теория двойственности

ОПР 2.1 (Двойственной задачи).

Рассмотрим задачу оптимизации (P): *нахождение минимума функции* $f(\vec{x})$, *при ограничениях*

$$\varphi_i(\vec{x}) \leqslant 0 - i = 1,\, 2,\, \ldots,\, m,$$

функции f *и* φ_i *произвольны. Пусть* $g(\vec{x}) = \sup_{\vec{\lambda}\geqslant 0} L(\vec{x}, \vec{\lambda})$, *где* L — *функция Лагранжа:*

$$L\big(\vec{x},\, \vec{\lambda}\big) = f(\vec{x}) + \sum_{i=1}^{m} \lambda_i \varphi_i(\vec{x}),$$

тогда обозначим

$$g(\vec{x}) = \begin{cases} f(\vec{x}), & \vec{x} \in Q, \\ +\infty, & \vec{x} \notin Q. \end{cases}$$

Задача (P) *эквивалентна следующей задаче: нахождение минимума функции* $g(\vec{x})$, *при ограничениях*

$$i = 1,\, 2,\, \ldots,\, m\colon \varphi_i(\vec{x}) \leqslant 0.$$

Пусть

$$h(\vec{\lambda}) = \inf_{\vec{x}\in\mathbb{R}^n} L\big(\vec{x},\, \vec{\lambda}\big),$$

рассмотрим задачу D: *нахождение максимума функции* $h(\vec{\lambda})$, *при условии* $\vec{\lambda} \geqslant 0$. *Задача* D *называется* двойственной задачей *к прямой (или исходной) задаче* P. *Переменные* $\vec{\lambda}_1,\, \vec{\lambda}_2,\, \ldots,\, \vec{\lambda}_m$ *назовем* двойственными, *а переменные* $\vec{x}_1,\, \vec{x}_2,\, \ldots,\, \vec{x}_n$ — прямыми.

Если $\vec{x} \in Q$, $\vec{\lambda} \geqslant 0$, *то* $\vec{x}$ — *допустимое решение прямой задачи, а* $\vec{\lambda}$ — *допустимое решение двойственной задачи.*

Лемма 2.2 (Слабая теорема двойственности).

- ▷ $\forall \vec{x} \in Q,\, \forall \vec{\lambda} \geqslant 0\colon h(\vec{\lambda}) \leqslant f(\vec{x})$.
- ▷ **Доказательство**.
  - ∘ Так как $\forall i\colon \lambda_i \cdot \varphi_i(\vec{x}) \leqslant 0$ и $\vec{\lambda} \geqslant \vec{0}$, то

$$h(\vec{\lambda}) = \inf_{\vec{x}\in\mathbb{R}^n} L\big(\vec{x},\, \vec{\lambda}\big) \leqslant L\big(\vec{x},\, \vec{\lambda}\big) = f(\vec{x}) + \sum_{i=1}^{m} \lambda_i \cdot \varphi_i(\vec{x}) \leqslant f(\vec{x}).$$

□

Лемма 2.3 (Оптимальные решения прямой и двойственной задач).

- ▷ Пусть

  $\overline{\vec{x}} \in Q$, $\overline{\vec{\lambda}} \geqslant 0$ и $f(\overline{\vec{x}}) = h(\overline{\vec{\lambda}})$.
- ▷ Тогда

  $\overline{\vec{x}}$ и $\overline{\vec{\lambda}}$ — оптимальные решения задачи (P) и (D) из определения 2.1.
- ▷ **Доказательство**.
  - ∘ Из леммы 2.2 имеем:

$$\forall \vec{x} \in Q\colon f(\overline{\vec{x}}) = h(\overline{\vec{\lambda}}) \leqslant f(\vec{x}) \Rightarrow \overline{\vec{x}} \text{ — оптимальное решение задачи } (P),$$
$$\forall \vec{\lambda} \geqslant 0\colon h(\vec{\lambda}) \leqslant f(\overline{\vec{x}}) = h(\overline{\vec{\lambda}}) \Rightarrow \overline{\vec{\lambda}} \text{ — оптимальное решение задачи } (D).$$

□

Теорема 2.4 (Допустимые точки как седловая точка функции Лагранжа).

- ▷ Пусть

  $\overline{\vec{x}}$, $\overline{\vec{\lambda}}$ — допустимые решения прямой и двойственной задачи.
- ▷ Тогда

  $f(\overline{\vec{x}}) = h(\overline{\vec{\lambda}})$ тогда и только тогда, когда пара $\big(\overline{\vec{x}},\, \overline{\vec{\lambda}}\big)$ — седловая точка функции Лагранжа, причем $L\big(\overline{\vec{x}},\, \overline{\vec{\lambda}}\big) = f(\overline{\vec{x}}) = h(\overline{\vec{\lambda}})$.
- ▷ **Доказательство**.
  - ∘ Необходимость:

✓ По формуле (1.10) на стр. 29 получаем:

$$L(\overline{\vec{x}}, \vec{\lambda}) \leqslant f(\overline{\vec{x}}) = h(\overline{\vec{\lambda}}) \leqslant L(\vec{x}, \overline{\vec{\lambda}}),$$

из первого неравенства получаем $f(\overline{\vec{x}}) = \sup_{\vec{\lambda}\geqslant 0} L(\overline{\vec{x}}, \vec{\lambda})$, а из второго $h(\vec{\lambda}) = \inf_{\vec{x}\in R^n} L(\vec{x}, \vec{\lambda})$.

✓ Аналогично для формулы (1.11) на стр. 29:

$$L(\overline{\vec{x}}, \overline{\vec{\lambda}}) \leqslant f(\overline{\vec{x}}) = h(\overline{\vec{\lambda}}) \leqslant L(\vec{x}, \overline{\vec{\lambda}}),$$

следовательно, $(\overline{\vec{x}}, \overline{\vec{\lambda}})$ — седловая точка функции L и $L(\overline{\vec{x}}, \overline{\vec{\lambda}}) = f(\overline{\vec{x}}) = h(\overline{\vec{\lambda}})$.

◦ Достаточность: пусть $(\overline{\vec{x}}, \overline{\vec{\lambda}})$ — седловая точка функции Лагранжа L и $L(\overline{\vec{x}}, \overline{\vec{\lambda}}) = f(\overline{\vec{x}}) = h(\overline{\vec{\lambda}})$, тогда:

✓ Используя формулу (1.10) на стр. 29 получаем, что

$$\forall \vec{\lambda} \geqslant 0 \colon L(\overline{\vec{x}}, \vec{\lambda}) \leqslant L(\overline{\vec{x}}, \overline{\vec{\lambda}}),$$

следовательно,

$$\sup_{\vec{\lambda}\geqslant 0} L(\overline{\vec{x}}, \vec{\lambda}) = L(\overline{\vec{x}}, \overline{\vec{\lambda}}) = f(\overline{\vec{x}}).$$

Первое из этих равенств тривиально, а второе следует из условия $\overline{\vec{x}} \in Q$[I)].

✓ Аналогично, используя формулу (1.11) на стр. 29 получаем

$$\forall \vec{x} \in \mathbb{R}^n \colon L(\overline{\vec{x}}, \overline{\vec{\lambda}}) \leqslant L(\vec{x}, \overline{\vec{\lambda}}),$$

таким образом:

$$L(\overline{\vec{x}}, \overline{\vec{\lambda}}) = \inf_{\vec{x}} L(\vec{x}, \overline{\vec{\lambda}}) = h(\overline{\vec{\lambda}}).$$

$\overline{\vec{x}}$ и $\overline{\vec{\lambda}}$ — допустимые решения, тогда по лемме 2.2 на стр. 34 получаем:

$$L(\overline{\vec{x}}, \overline{\vec{\lambda}}) = h(\overline{\vec{\lambda}}) \leqslant f(\overline{\vec{x}}) = L(\overline{\vec{x}}, \overline{\vec{\lambda}}).$$

□

23.09.2005

I) Напомним, что это значит, что если $\varphi_i(\overline{\vec{x}}) < 0$, то $\overline{\vec{\lambda}}_i = 0$.

Следствие 2.5 (Эквивалентные утверждения для оптимальных решений).

▷ Следующие утверждения эквивалентны:

1. Пара $(\overline{\vec{x}}, \overline{\vec{\lambda}})$ — седловая точка функции Лагранжа;
2. $f(\overline{\vec{x}}) = h(\overline{\vec{\lambda}})$;
3. $\min\limits_{\vec{x}} \sup\limits_{\vec{\lambda}\geqslant 0} L(\vec{x}, \vec{\lambda}) = \max\limits_{\vec{\lambda}\geqslant 0} \inf\limits_{\vec{x}} L(\vec{x}, \vec{\lambda})$.

Следствие 2.6 (Связь седловых точек функции Лагранжа).

▷ Пусть

$\vec{x}^*$, $\overline{\vec{x}} \in Q$ и $\vec{\lambda}^*$, $\overline{\vec{\lambda}} \geqslant 0$.

▷ Тогда

Если пары $(\overline{\vec{x}}, \overline{\vec{\lambda}})$ и $(\vec{x}^*, \vec{\lambda}^*)$ — седловые точки функции Лагранжа, то пары $(\overline{\vec{x}}, \vec{\lambda}^*)$ и $(\vec{x}^*, \overline{\vec{\lambda}})$ — также седловые точки функции Лагранжа, причем

$$L(\vec{x}^*, \overline{\vec{\lambda}}) = L(\overline{\vec{x}}, \vec{\lambda}^*) = L(\overline{\vec{x}}, \overline{\vec{\lambda}}) = L(\vec{x}^*, \vec{\lambda}^*).$$

УПР 2.7 (Эквивалентность двойственной задачи задаче выпуклого программирования).

▷ Двойственная задача (D) из определения 2.1 эквивалентна задаче выпуклого программирования.

Следствие 2.8 (Условие оптимальности для задачи выпуклого программирования с условием Слейтора).

▷ Пусть

Задача (P) — задача выпуклого программирования 1.6.8 на стр. 23 и выполняется условие Слейтера 1.6.9 на стр. 23.

▷ Тогда

Допустимое решение $\overline{\vec{x}}$ прямой задачи является оптимальным тогда и только тогда, когда существует такое $\overline{\vec{\lambda}} \geqslant 0$, что $f(\overline{\vec{x}}) = h(\overline{\vec{\lambda}})$.

▷ **Доказательство**.

◦ Необходимость: пусть $\overline{\vec{x}}$ — оптимальное решение задачи (P), по теорема Куна-Таккера в нелокальной форме 1.7.5 на стр. 29

$$\exists \overline{\vec{\lambda}} \geqslant 0 \,\big|\, (\overline{\vec{x}}, \overline{\vec{\lambda}}) \text{ — седловая точка функции Лагранжа,}$$

а из следствия 2.5 на стр. 36 имеем $f(\overline{\vec{x}}) = h(\overline{\vec{\lambda}})$.

◦ Достаточность: следует из леммы 2.3 на стр. 34.

□

Замечание 2.8.1 (Для линейных ограничений).

▷ Пусть

Ограничения линейны, то есть $\varphi_i(x) = (a_i,\, x) - b_i \leqslant 0$.

▷ Тогда

Условие Слейтера 1.6.9 на стр. 23 излишне и при обосновании следствия 2.8 на стр. 36 вместо теоремы 1.7.5 на стр. 29 лучше использовать теорему 2.4 на стр. 34.

Глава 3

Линейное программирование

3.1 Двойственная задача

ОПР 3.1.1 (Задачи линейного программирования в канонической форме)**.**
Такой задачей называется нахождение минимума функции $w(\vec{x}) = (\vec{c}, \vec{x})$, *при ограничениях*

$$\begin{cases} \mathtt{A}\vec{x} = \vec{b}, \\ \vec{x} \geqslant \vec{0}. \end{cases}, \tag{3.1}$$

где $\vec{c} = (c_j)$, $\vec{x} = (x_j) \in \mathbb{R}^n$, $\mathtt{A} = (a_{ij})$ *— матрица размерности* $(m \times n)$, $\vec{b} = (b_i) \in \mathbb{R}^m, m \leqslant n$ *и* $\operatorname{rang}(\mathtt{A}) = m$. *Это можно переписать так:*

$$\mathtt{A}\vec{x} = \vec{b} \equiv (\vec{a}_i, \vec{x}) = b_i, \, i = 1, 2, \ldots, m,$$
$$\mathtt{A}\vec{x} = \vec{b} \equiv \sum_{j=1}^{n} \vec{A}_j \cdot x_j = \vec{b}.$$

Задача 3.1.1 эквивалентна задаче нахождения минимума $(\vec{c}, \vec{x})$ при ограничениях

$$\begin{cases} (\vec{a}_i, \vec{x}) - b_i \leqslant 0 & : \vec{\lambda}^1_i \geqslant 0, \\ -(\vec{a}_i, \vec{x}) + b_i \leqslant 0 & : \vec{\lambda}^2_i \geqslant 0, \\ -x_j \leqslant 0 & : \vec{\mu}_j \geqslant 0. \end{cases},$$

соответственно ее функция Лагранжа:

$$L\big(\vec{x}, \vec{\lambda}^1, \vec{\lambda}^2, \vec{\mu}\big) = (\vec{c}, \vec{x}) + (\vec{\lambda}^1, \mathtt{A}\vec{x} - \vec{b}) + (\vec{\lambda}^2, -\mathtt{A}\vec{x} + \vec{b}) + (\vec{\mu}, -\vec{x}) =$$
$$= \Big(\vec{c} + \big(\vec{\lambda}^1 - \vec{\lambda}^2\big)\mathtt{A} - \vec{\mu}, \vec{x}\Big) - \big(\vec{\lambda}^1 - \vec{\lambda}^2, \vec{b}\big).$$

Следовательно, целевая функция двойственной задачи имеет вид:

$$h(\vec{\lambda}^1, \vec{\lambda}^2, \vec{\mu}) = \inf_{\vec{x}} L\big(\vec{x}, \vec{\lambda}^1, \vec{\lambda}^2, \vec{\mu}\big) =$$
$$= \begin{cases} -(\vec{b}, \vec{\lambda}^1 - \vec{\lambda}^2), & \text{если } \vec{c} + (\vec{\lambda}^1 - \vec{\lambda}^2) \cdot \mathtt{A} - \vec{\mu} = \vec{0}, \\ -\infty, & \text{иначе.} \end{cases}$$

Двойственная задача

$$h(\vec{\lambda}^1, \vec{\lambda}^2, \vec{\mu}) \to \sup_{\vec{\lambda}^1 \geqslant 0, \vec{\lambda}^2 \geqslant 0, \vec{\mu} \geqslant 0}$$

эквивалентна задаче линейного программирования:

$$-(\vec{b}, \vec{\lambda}^1 - \vec{\lambda}^2) \to \max$$
$$\vec{c} + (\vec{\lambda}^1 - \vec{\lambda}^2) \cdot \mathtt{A} - \vec{\mu} = \vec{0} \equiv \vec{c} + (\vec{\lambda}^1 - \vec{\lambda}^2) \cdot \mathtt{A} \geqslant 0.$$

Умножим ограничения на (-1), обозначим $\vec{y} = -(\vec{\lambda}^1 - \vec{\lambda}^2)$ и получим задачу

$$(\vec{b}, \vec{y}) \to \max, \; \vec{y}\mathtt{A} \leqslant \vec{c}. \tag{3.2}$$

Замечание 3.1.1.1 (Свойства таких задач).

▷ Для задач (3.1) и (3.2) выполняются все утверждения:

- лемма 2.2 на стр. 34 (слабая теорема двойственности);
- лемма 2.3 на стр. 34 (оптимальные решения прямой и двойственной задач);
- теорема 2.4 на стр. 34 (допустимые точки как седловая точка функции Лагранжа);
- следствие 2.5 на стр. 36 (эквивалентные утверждения для оптимальных решений);
- следствие 2.6 на стр. 36 (связь седловых точек функции Лагранжа);
- следствие 2.8 на стр. 36 без условия Слейтера (условие оптимальности для задачи выпуклого программирования).

Теорема 3.1.2 (Совпадение двойственной задачи для двойственной задачи с исходной)**.**

▷ Задача двойственная к задаче (3.2) на стр. 40 совпадает с исходной задачей (3.1) на стр. 39.

▷ **Доказательство**.

- Задача (3.2) эквивалентна задаче нахождения минимума $-(\vec{b}, \vec{y})$, при условиях:
$$\vec{y}\mathtt{A} \leqslant \vec{c},$$
а условие $\vec{y}\mathtt{A} \leqslant \vec{c}$ эквивалентно системе неравенств
$$(\vec{y}, \vec{A}_j) - \vec{c}_j{}^{1)} \quad : x_j \geqslant 0.$$
Дальше функция Лагранжа имеет такой вид
$$L(\vec{y}, \vec{x}) = -(\vec{b}, \vec{y}) + (\vec{x}, \vec{y}\mathtt{A} - \vec{c}) = \\ = -(\vec{b}, \vec{y}) + (\mathtt{A}\vec{x}, \vec{y}) - (\vec{x}, \vec{c}) = (\mathtt{A}\vec{x} - \vec{b}, \vec{y}) - (\vec{c}, \vec{x}),$$
целевая функция двойственной задачи
$$h(\vec{x}) = \inf_{\vec{y}} L(\vec{y}, \vec{x}) = \begin{cases} -(\vec{c}, \vec{x}), & \text{если } \mathtt{A}\vec{x} = \vec{b}, \\ -\infty, & \text{иначе.} \end{cases}$$
- Задача $\max_{\vec{x}\geqslant 0} h(\vec{x})$ эквивалентна следующей задаче линейного программирования: нахождение максимума $-(\vec{c}, \vec{x})$, при условиях
$$\mathtt{A}\vec{x} = \vec{b},\ \vec{x} \geqslant 0$$
или минимума $(\vec{c}, \vec{x})$, при тех же условиях
$$\mathtt{A}\vec{x} = \vec{b}, \vec{x} \geqslant 0.$$

□

Следствие 3.1.3 (Общий вид перехода к двойственной задачи).

▷ В итоге мы получаем следующую схему:

Прямая задача		**Двойственная задача**
$w(\vec{x}) = \sum_{j=1}^{n} c_j \cdot x_j \to \min$		$z(\vec{y}) = \sum_{i=1}^{m} b_i \cdot y_i \to \max$
$(\vec{a}_i, \vec{x}) \geqslant b_i$	$i \in I_1$	$y_i \geqslant 0$
$(\vec{a}_i, \vec{x}) = b_i$	$i \in I_2$	y_i — своб.
$x_j \geqslant 0$	$j \in J_1$	$(\vec{y}, \vec{A}_j) \leqslant c_j$
x_j — своб.	$j \in J_2$	$(\vec{y}, \vec{A}_j) = c_j$.

УПР 3.1.3.1 (Рецепт получение конечной схемы).

▷ Техника получения этой схемы: либо повторить выкладки, приведшие к задаче (3.2) на стр. 40, либо воспользоваться сводимостью общей задачи линейного программирования к задаче линейного программирования в канонической форме и применить готовый рецепт (задача (3.2) на стр. 40).

[1)] То есть сопоставили каждому ограничению двойственную переменную (множитель Лагранжа).

3.2 Понятие базисного допустимого решения

ОПР 3.2.1 (Базиса, базисных и не базисных переменных).

Базис — *любой набор векторов* $(\vec{A}_{\sigma(1)}, \vec{A}_{\sigma(2)}, \ldots, \vec{A}_{\sigma(m)})$ *из* m *линейно независимых столбцов матрицы системы ограничений* $\mathtt{A}$; *такую матрицу* $B = [\mathtt{A}_{\sigma(1)}, \mathtt{A}_{\sigma(2)}, \ldots, \mathtt{A}_{\sigma(m)}]$ *также назовем* базисной.

Пусть $S = \{\sigma(1), \sigma(2), \ldots, \sigma(m)\}$, *а* $S' = \{1, 2, \ldots, n\} \setminus S$, *тогда считаем, что* $\mathtt{A} = [\mathtt{B}, \mathtt{N}]$, *где*
$$\mathtt{N} = [\vec{A}_j]_{j\in S'},\ \vec{x} = [\vec{x}_B, \vec{x}_N];$$
причём $\vec{x}_B = (\vec{x}_{\sigma(1)}, \vec{x}_{\sigma(2)}, \ldots, \vec{x}_{\sigma(m)})$ — базисные, *а* $\vec{x}_N = (\vec{x}_j)_{j\in S'}$ — не базисные переменные.

Умножим систему ограничений (3.1) на стр. 39 на $\mathtt{B}^{-1}$ *и получим вид*
$$\mathtt{E}\vec{x}_B + \mathtt{B}^{-1}\mathtt{N}\vec{x}_N = \mathtt{B}^{-1}\vec{b} \tag{3.3}$$

ОПР 3.2.2 (Базисного решения).

Решение $[\vec{x}_B, \vec{x}_N] = [\mathtt{B}^{-1}\vec{b}, 0]$ *системы уравнений (3.1) на стр. 39 назовем* базисным *(соответствующим базису* $\mathtt{B}$*).*

Лемма 3.2.3 (Условие базисного решения системы).

▷ Вектор $\vec{x}$ — базисное решение системы (3.1) на стр. 39 тогда и только тогда, когда множество столбцов матрицы ограничений $\mathtt{A}$ с индексами из множества $S(\vec{x}) = \{j \mid x_j \neq 0\}$ — линейно независимо.

▷ **<u>Доказательство</u>**.

(⇒) Пусть $\vec{x}$ — базисное решение, $\mathtt{B}$ — соответствующий его базис, тогда $S \supseteq S(\vec{x})$.

(⇐) Пусть множество столбцов с индексами из $S(\vec{x})$ линейно независимо. Если $|S(\vec{x})| = m$, то $S(\vec{x})$ — базис; если нет, то из условия $\operatorname{rang}(\mathtt{A}) = m$ следует, что множество $S(\vec{x})$ можно дополнить до подходящего базиса $\mathtt{B}$, которому будет соответствовать решение $\vec{x}$.

□

ОПР 3.2.4 (Базисного допустимого решения).

Базисным допустимым решением *(б.д.р.) называется любой элемент множества* Q, *являющийся базисным решением системы уравнений (3.1) на стр. 39*

Замечание 3.2.4.1 (Неравенство для базисного вектора).

▷ Решение соответствующее базису $\mathtt{B}$ — базисно-допустимое, следовательно, $\mathtt{B}^{-1}\vec{b} \geqslant \vec{0}$.

ОПР 3.2.5 (Крайней точки).

Вектор $\vec{x} \in Q$ — крайняя точка *или* вершина *множества* Q, *если не существует допустимых решений* $\vec{x}^1 \neq \vec{x}^2$ *из* Q *таких, что* $\vec{x} = \alpha \cdot \vec{x}^1 + (1-\alpha) \cdot \vec{x}^2$, *где* $0 < \alpha < 1$.

Теорема 3.2.6 (Совпадение крайней точки с базисным допустимым решением).

▷ Вектор $\vec{x}$ — б.д.р. тогда и только тогда, когда $\vec{x}$ — крайняя точка множества Q.

▷ **Доказательство**.

($\Leftarrow$) Пусть $\vec{x}$ — крайняя точка, но не б.д.р, тогда из леммы 3.2.3 на стр. 42 следует, что

$$\exists \vec{y} \neq \vec{0} \,\big|\, \mathtt{A}\vec{y} = 0, \tag{3.4}$$

при этом можем считать, что из условия $x_j = 0$ следует, что $y_j = 0$, то есть

$$\{j \,|\, y_j \neq 0\} \subseteq \{j \,|\, x_j > 0\}. \tag{3.5}$$

Поскольку $\vec{x} \in Q$, то из условия (3.4) следует, что $\forall t \in \mathbb{R}$: $\vec{z}(t) = \vec{x} + t\vec{y}$ — решение системы (3.1) на стр. 39, тогда из (3.5) получаем, что для любого малого $t \in \mathbb{R}$:

$$\vec{z}(t) \in Q \Rightarrow \exists \varepsilon > 0 \,\big|\, \vec{x}^1 = \vec{x} + \varepsilon\vec{y} \in Q,\ \vec{x}^2 = \vec{x} - \varepsilon\vec{y} \in Q \Rightarrow$$
$$\Rightarrow \vec{x}^1 \neq \vec{x}^2, \text{ но } \vec{x} = \frac{1}{2} \cdot \vec{x}^1 + \frac{1}{2} \cdot \vec{x}^2.$$

Получили противоречие, значит $\vec{x}$ — базисное допустимое решение.

($\Rightarrow$) Пусть $\vec{x}$ — б.д.р., но

$$\exists 0 < \alpha < 1 \,\big|\, \exists \vec{x}^1 \neq \vec{x}^2,\ \vec{x}^1, \vec{x}^2 \in Q \,\big|\, \vec{x} = \alpha\vec{x}^1 + (1-\alpha)\vec{x}^2,$$

тогда по условию $\mathtt{A}\vec{x}^1 = \vec{b} = \mathtt{A}\vec{x}^2$, то есть $\mathtt{A} \cdot (\vec{x}^1 - \vec{x}^2) = 0$, но

$$\mathtt{A} \cdot (\vec{x}^1 - \vec{x}^2) = 0 \Leftrightarrow \{\vec{A}_j \,|\, x_j^1 \neq x_j^2\} \text{ — линейно зависимо } \Rightarrow$$
$$\Rightarrow \{\vec{A}_j \,|\, \alpha x_j^1 + (1-\alpha)x_j^2 > 0\} \text{ — линейно зависимо } \Rightarrow$$
$$\Rightarrow \{\vec{A}_j \,|\, x_j > 0\} \text{ — линейно зависимо }.$$

Получили противоречие, то есть если $\vec{x}$ — б.д.р., то $\vec{x}$ — крайняя точка множества Q.

□

3.3 Критерий разрешимости

Теорема 3.3.1 (Критерий разрешимости).

▷ Задача линейного программирования (3.1) на стр. 39 разрешима тогда и только тогда, когда множество допустимых решений не пусто и целевая функция ограничена на нем.

▷ **Доказательство**.

- Необходимость: понятно из определения разрешимости.
- Достаточность: покажем, что для любого $\vec{x}^0 \in Q$ существует базисное допустимое решение

$$\overline{\vec{x}} \,\big|\, w(\vec{x}^0) \leqslant w(\overline{\vec{x}}).$$

Для этого пусть

$$\overline{\vec{x}} \in Q^0 = \{\vec{x} \in Q \,\big|\, w(\vec{x}) \leqslant w(\vec{x}^0)\} \neq \varnothing \tag{3.6}$$

и имеет минимальное число ненулевых компонент (будем обозначать $\mathrm{supp}(\overline{\vec{x}})$); докажем, что $\overline{\vec{x}}$ — б.д.р. Допустим противное, тогда множество

$$\{\vec{A}_j \,\big|\, \overline{\vec{x}}_j > 0\} \text{ линейно зависимо } \Rightarrow \exists \vec{y} \neq \vec{0} \,\big|\, \mathtt{A}\vec{y} = \vec{0}$$

и если $\overline{\vec{x}}_j = 0$, то и $y_j = 0$. Пусть $w(\vec{y}) \leqslant 0$ (если необходимо, то возьмем $-\vec{y}$), положим $\vec{x}(t) = \overline{\vec{x}} + t\vec{y}$, тогда выполняется следующее свойство:

$$\forall \text{ малого } t \in \mathbb{R}\colon \vec{x}(t) \in Q.$$

Рассмотрим возможные случаи:

1. Пусть $\forall j\colon y_j \geqslant 0$, следовательно, $\forall t \geqslant 0\colon \vec{x}(t) \geqslant 0$, значит, $\forall t \geqslant 0\colon \vec{x}(t) \in Q$, отсюда и неравенства (3.6)

$$w(\vec{x}(t)) = w(\overline{\vec{x}}) + tw(\vec{y}) \geqslant_{(\text{по условию})} \text{const}.$$

Учитывая, что $w(\vec{y}) \leqslant 0$ и $t \geqslant 0$ произвольно, получаем $w(\vec{y}) = 0$ и, следовательно,

$$\forall t\colon w(\vec{x}(t)) = w(\overline{\vec{x}}).$$

Так как из условия $y_j > 0$ следует $\overline{\vec{x}}_j > 0$, то

$$\forall \text{ малого по абсолютной величине } t < 0\colon \vec{x}(t) \in Q.$$

Найдем такое $\overline{t}$ наибольшее по абсолютной величине из условия:

$$\forall j\colon y_j > 0 \Rightarrow \overline{\vec{x}}_j + \overline{t}y_j \geqslant 0,$$

которое эквивалентно условию:

$$\forall j\colon y_j > 0 \Rightarrow \overline{x}_j \geqslant (-\overline{t})y_j \Leftrightarrow (-\overline{t}) = \min_{y_j>0} \frac{\overline{x}_j}{y_j} \Leftrightarrow \overline{t} = -\min_{y_j>0} \frac{\overline{x}_j}{y_j}.$$

Итак: $\vec{x}(\overline{t}) \in Q$ и $w\big(\vec{x}(\overline{t})\big) = w(\overline{\vec{x}}) \leqslant w(\vec{x}^0) \Rightarrow \vec{x}(\overline{t}) \in Q^0$, получили противоречие, так как $\operatorname{supp}\big(\vec{x}(\overline{t})\big) = \operatorname{supp}\big(\overline{\vec{x}}\big) - 1$.

2. Пусть $\exists j \big| \, y_j < 0$, тогда

$$\forall \text{ достаточно малых } t \geqslant 0\colon \vec{x}(t) \in Q.$$

Найдем наибольшее такое $\overline{t}$ из условия:

$$\forall j\colon y_j < 0 \Rightarrow \overline{x}_j + \overline{t}y_j \geqslant 0,$$

которое эквивалентно условию:

$$\forall j\colon y_j < 0 \Rightarrow \overline{x}_j \geqslant \overline{t}(-y_j) \Leftrightarrow \overline{t} = \min_{y_j<0} \frac{\overline{x}_j}{-y_j}.$$

Итак, $\vec{x}(\overline{t}) \in Q$ и так как $\overline{t} > 0$, $d \leqslant 0$, то $w\big(\vec{x}(\overline{t})\big) = w\big(\overline{\vec{x}}\big) + d\overline{t} \leqslant$ $\leqslant w\big(\vec{x}^0\big)$, следовательно, $\vec{x}(\overline{t}) \in Q^0$. Получили противоречие, так как $\operatorname{supp}\big(\vec{x}(\overline{t})\big) = \operatorname{supp}\big(\overline{\vec{x}}\big) - 1$.

Так как по условию $Q \neq \varnothing$, то множество базисных допустимых решений задачи не пусто, но поскольку оно конечно, то существует $\vec{x}^*$ — б.д.р. такое, что

$$\forall \text{ б.д.р. } \vec{x}\colon w\big(\vec{x}^*\big) \leqslant w(\vec{x}).$$

Из ранее доказанного следует, что $\vec{x}^*$ — оптимальное решение.

□

Следствие 3.3.1.1 (Существование базисных допустимых решений).

▷ Если множество допустимых решений задачи линейного программирования не пусто, то существуют базисные допустимые решения.

▷ **Доказательство**.

- Следует из доказательства теоремы 3.3.1, если взять $w(\vec{x}) \equiv 0$

□

Следствие 3.3.1.2 (Существование оптимального базисного решения).

▷ Если задача линейного программирования разрешима, то существует оптимальное базисное решение.

30.09.2005

3.4 Теоремы двойственности

3.4.1 Первая теорема двойственности

30.09.2005

Рассмотрим задачу ЛП в канонической форме.

Теорема 3.4.1.1 (Первая теорема двойственности)**.**

▷ Прямая и двойственная к ней задачи либо одновременно разрешимы, либо одновременно неразрешимы.

▷ При этом

В первом случае оптимальные значения целевых функций этих задач совпадают, а во втором случае по крайней мере одна из задач неразрешима в силу несовместности ее ограничений.

▷ **Доказательство**.

- Пусть $\overline{\vec{x}}$ — оптимальное решение задачи (3.1) на стр. 39, из следствия 2.8 на стр. 36 (условие оптимальности для задачи выпуклого программирования с условием Слейтора) следует, что найдется оптимальное решение $\overline{\vec{y}}$ задачи (3.2) на стр. 40:

$$w(\overline{\vec{x}}) = h(\overline{\vec{y}});$$

то есть задача (3.2) на стр. 40 разрешима и значения целевых функций на оптимальных решениях совпадают.

Пусть теперь разрешима задача (3.2) на стр. 40 и $\overline{\vec{y}}$ — ее некоторое оптимальное решение, запишем ее в каноническом виде, тогда в силу доказанного выше, ее двойственная задача разрешима. Но по теореме 3.1.2 на стр. 40 (совпадение двойственной задачи двойственной задачи с исходной) эта двойственная и есть задача (3.1) на стр. 39. Следовательно, она разрешима и существует ее оптимальное решение $\overline{\vec{x}} \big| \, w(\overline{\vec{x}}) = h(\overline{\vec{y}})$.

- Покажем, что верно обратное утверждение: пусть совместны ограничения прямой и двойственной задач и $\overline{\vec{y}}$ — некоторое допустимое решение задачи (3.2) на стр. 40, тогда по лемме 2.2 на стр. 34 (слабая теорема двойственности)

$$\forall \vec{x} \in Q\colon w(\vec{x}) \geqslant h(\overline{\vec{y}}).$$

Из теоремы 3.3.1 на стр. 44 (критерий разрешимости задачи линейного программирования) тогда следует, что разрешима прямая задача (3.1) на стр. 39. По доказанному ранее разрешима и двойственная задача, следовательно, разрешимость задач (3.1) на стр. 39

и (3.2) на стр. 40 эквивалентна совместности их ограничений. Значит, если данные задачи неразрешимы, то хотя бы одна из них неразрешима из-за несовместности их ограничений.

□

Замечание 3.4.1.1.1 (Совместность ограничений при разрешимости двойственных задач).

▷ Так как совместность часть разрешимости, то из разрешимости задач прямой и двойственной следует совместность их ограничений.

3.4.2 Вторая теорема двойственности

Теорема 3.4.2.1 (Вторая теорема двойственности или теорема о дополняющей нежесткости)**.**

▷ Допустимые решения $\overline{\vec{x}}$ и $\overline{\vec{y}}$ соответственно прямой и двойственной задачи оптимальны тогда и только тогда, когда выполняются условия:

$$\forall i \in I\colon y_i \cdot ((\vec{a}_i,\, \vec{x}) - b_i) = 0$$
$$\forall j \in J\colon (c_j - (\vec{y},\, \vec{A}_j))x_j = 0.$$

▷ **Доказательство**.

∘ Смотрите учебное пособие [1] („Методы оптимизации“ Н. И. Глебов и др.)

□

3.5 Симплекс-таблица

Пусть $\overline{\vec{x}}$ — б.д.р., $\mathtt{B} = (\vec{A}_{\sigma(1)},\, \vec{A}_{\sigma(2)},\, \ldots,\, \vec{A}_{\sigma(m)})$ — соответствующая базисная матрица, тогда

$$\mathtt{E}\vec{x}_B + \mathtt{B}^{-1}\mathtt{N}\vec{x}_N = \mathtt{B}^{-1}\vec{b}, \tag{3.7}$$

следовательно

$$\mathtt{E}\vec{x}_B = \mathtt{B}^{-1}\vec{b} - \mathtt{B}^{-1}\mathtt{N}\vec{x}_N,$$

и

$$w = \vec{c}_B\mathtt{B}^{-1}\vec{b} + (\vec{c}_N - \vec{c}_B\mathtt{B}^{-1}\mathtt{N})\vec{x}_N, \tag{3.8}$$

где $\vec{x}_B = (x_{\sigma(1)},\, x_{\sigma(2)},\, \ldots,\, x_{\sigma(m)})$ — базисные, а $\vec{x}_N = (x_j)_{j \in S'}$ — небазисные переменные. Дальше берём

$$-w + \sum_{j \in S'} z_{0j}x_j = z_{00} \tag{3.9}$$

$$\forall i = 1,\, 2,\, \ldots,\, m\colon x_{\sigma(i)} + \sum_{j \in S'} z_{ij}x_j = z_{i0}, \tag{3.10}$$

где

▷ $z_{00} = -\vec{c}_B\mathtt{B}^{-1}\vec{b} = -w(\overline{\vec{x}})$,

▷ $\forall j = 1,\, 2,\, \ldots,\, n\colon z_{0j} = c_j - \vec{c}_B\mathtt{B}^{-1}\vec{A}_j$,

▷ $(z_{10},\, z_{20},\, \ldots,\, z_{m0})^\top = \mathtt{B}^{-1}\vec{b}$,

▷ $\forall j = 1,\, 2,\, \ldots,\, n\colon (z_{1j},\, z_{2j},\, \ldots,\, z_{mj})^\top = \mathtt{B}^{-1}\vec{A}_j$.

и строим таблицу:

		x_1	x_2	$\ldots$	x_j	$\ldots$	x_n
$-w$	z_{00}	z_{01}	z_{02}	$\ldots$	z_{0j}	$\ldots$	z_{0n}
$x_{\sigma(1)}$	z_{10}	z_{11}	z_{12}	$\ldots$	z_{1j}	$\ldots$	z_{1n}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
$x_{\sigma(i)}$	z_{i0}	z_{i1}	z_{i2}	$\ldots$	z_{ij}	$\ldots$	z_{in}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
$x_{\sigma(m)}$	z_{m0}	z_{m1}	z_{m2}	$\ldots$	z_{mj}	$\ldots$	z_{mn}

где

▷ z_{0j} — *оценки замещения*,

▷ z_{ij} — *коэффициенты замещения*,

▷ z_{i0} — значения базисных компонент текущего б.д.р.

ОПР 3.5.1 (Прямо допустимой симплекс таблицы)**.**
Симплекс-таблица называется прямо допустимой*, если*

$$\forall i = 1,\, 2,\, \ldots,\, m\colon z_{i0} \geqslant 0.$$

Базис $\mathtt{B}$*, которому эта таблица соответствует, также называется* прямо допустимым.

ОПР 3.5.2 (Двойственно допустимой симплекс таблицы).

Симплекс-таблица называется двойственно допустимой, если

$$\forall j = 1,\, 2,\, \ldots,\, n\colon z_{0j} \geqslant 0,$$

базис $\mathtt{B}$, *которому эта таблица соответствует, называется* двойственно допустимым.

Замечание 3.5.3 (Выполнение условий в зависимости от знаков величин).

▷ В зависимости от знаков величин

$$z_{ij},\, z_{0j} - j \in S',\, i = 1,\, 2,\, \ldots,\, m$$

выполняется хотя бы одно из условий:

1. $\forall j \in S^{'}\colon z_{0j} \geqslant 0$;
2. $\exists s \in S^{'}\colon z_{0s} < 0$ и $\forall i = 1,\, 2,\, \ldots,\, m\colon z_{is} \leqslant 0$;
3. $\exists s \in S^{'}\colon z_{0s} < 0$ и $\exists r\colon 1 \leqslant r \leqslant m \mid z_{rs} > 0$.

Лемма 3.5.4 (Признак оптимальности).

▷ Пусть

Симплекс-таблица прямо и двойственно допустима.

▷ Тогда

Текущее базисное допустимое решение $\overline{\vec{x}}$ является оптимальным решением задачи (3.1) на стр. 39.

▷ **Доказательство**.

∘ Пусть $\vec{x}$ — произвольное допустимое решение, так как $z_{0j} \geqslant 0$ и $x_j \geqslant 0$ для $j \in S'$, то из формулы (3.9) на стр. 48 следует, что

$$w(\vec{x}) = -z_{00} + \sum_{j\in S'} z_{0j}x_j \geqslant -z_{00} = w(\overline{\vec{x}})$$

□

07.10.2005

ОПР 3.5.5 (Семейства векторов).

Пусть $s \in S'$. *Для анализа случаев 2 и 3 введём* параметризованное семейство векторов $\vec{x}(t)$, $t \geqslant 0$:

$$\begin{cases} x_{\sigma(i)}(t) = \overline{x}_{\sigma(i)} - z_{is}t, \\ x_s(t) = t, \\ \forall j \in S' \setminus s\colon x_j(t) = 0, \end{cases} \tag{3.11}$$

ОПР 3.5.6 (Грани множества допустимых решений).

Пусть $S' \cup S = \{1,\, 2,\, \ldots,\, n\}$, *множество решений системы уравнений*

$$\mathtt{A}\vec{x} = b,\, \forall j \in S'\colon x_j = 0,\, \forall j \in S\colon x_j \geqslant 0$$

— *называется* гранью множества допустимых решений *(3.1) на стр. 39.*

ОПР 3.5.7 (Размерности грани).

Величина $n - m - |S'|$ — размерность грани *(в данном случае* $m + |S'|$ *— ранг системы уравнений).*

ОПР 3.5.8 (Грани размерности 0 и 1).

- *Если* $\overline{\vec{x}}$ *— б.д.р., то* $|S'| = n - m$, *то* $\overline{\vec{x}}$ — грань размерности 0.
- *Если* $|S'| = n - m - 1$, *то получим* грань размерности 1, *то есть ограниченное или неограниченное* ребро.

Замечание 3.5.9 (Семейство векторов из множеств номеров).

▷ Если

S — множество номеров базисных переменных, а S', соответственно, небазисных, то б.д.р. $\overline{\vec{x}}$ — единственное решение системы уравнений

$$\mathtt{A}\vec{x} = b,\, \forall j \in S'\colon x_j = 0.$$

▷ Тогда

Семейство векторов (3.11) есть множество решений системы

$$\mathtt{A}\vec{x} = b,\, \forall j \in S' \setminus s\colon x_j = 0.$$

При $t = 0$ решение $\vec{x}(0)$ совпадает с исходным б.д.р. $\overline{\vec{x}}$. При увеличении $t \geqslant 0$ точка $\vec{x}(t)$ движется в пределах множества решений системы $\mathtt{A}\vec{x} = b$. Пока сохраняется неотрицательным знак величин $x_{\sigma(i)}(t) = \overline{x}_{\sigma(i)} - z_{is}t \geqslant 0$, при всех $i = 1,\, 2,\, \ldots,\, m$, то движение происходит по ребру

$$\mathtt{A}\vec{x} = b,\, \forall j \in S' \setminus s\colon x_j = 0,\, \forall j \in S \cup \{s\}\colon x_j \geqslant 0,$$

в пределах множества допустимых решений задачи (3.1) на стр. 39.

ОПР 3.5.10 (Элементарного преобразования б.д.р.).

Преобразование исходного б.д.р. $\overline{\vec{x}}$ *по формулам (3.11) назовём его* элементарным преобразованием. *Предыдущие описание движения позволяет интерпретировать элементарное преобразование как движение по ребру, которое начинается из вершины* $\overline{\vec{x}}$.

Лемма 3.5.11 (О неразрешимости).

▷ Пусть

Для номера s оценка замещения $z_{0s} < 0$ и для всех индексов i коэффициенты замещения z_{is} неположительны.

▷ Тогда

В задаче (3.1) на стр. 39 не существует оптимального решения.

▷ **Доказательство**.

- Так как $\forall i\colon z_{is} \leqslant 0$, то из (3.11) на стр. 49 имеем

$$\forall t \geqslant 0\colon \forall j = 1,\, 2,\, \ldots,\, n\colon x_j(t) \geqslant 0.$$

Таким образом множество точек $\vec{x}(t)$, $t \geqslant 0$, образует неограниченное ребро многогранного множества Q, выходящее из вершины $\overline{x}$.

- Но из (3.9) на стр. 48 следует, что

$$w\big(\vec{x}(t)\big) = -z_{00} + z_{0s}t \to -\infty \text{ при } t \to +\infty.$$

Значит, в задаче (3.1) на стр. 39 не существует оптимального решения.

□

Замечание 3.5.11.1 (Движение по бесконечному ребру).

▷ В этой лемме анализируется случай, когда элементарное преобразование приводит к движению по бесконечному ребру, на котором целевая функция убывает.

Лемма 3.5.12 (О существовании лучшей вершины).

▷ Пусть

Оценка замещения $z_{0s} < 0$ и существуют базисные переменные с коэффициентами замещения $z_{is} > 0$.

▷ Тогда

Элементарное преобразование приведёт либо в вершину с меньшим значением целевой функции, либо вершина останется прежней, но изменится её базис.

▷ **Доказательство**.

- Пусть r — индекс базисной переменной такой, что

$$\overline{t} = \frac{\overline{x}_{\sigma(r)}}{z_{rs}} = \frac{z_{r0}}{z_{rs}} = \min_{z_{is}>0,\, i\geqslant 1} \frac{z_{i0}}{z_{is}}.$$

- Возможно два случая: $\overline{t} > 0$ и $\overline{t} = 0$.

✓ Рассмотрим первый случай, в нём

$$\forall i\colon \forall t < \overline{t}\colon x_{\sigma(i)}(t) = \overline{x}_{\sigma(i)} - z_{is}t > 0.$$

В силу выбора индекса $r\colon x_{\sigma(r)}(\overline{t}) = 0$ и для $\forall t > \overline{t}$ имеем

$$x_{\sigma(r)}(\overline{t}) < 0;$$

итак, множество векторов $\vec{x}(t)$, $0 \leqslant t \leqslant \overline{t}$ образует ограниченное ребро множества допустимых решений Q. Покажем, что вектор $\vec{x}(\overline{t})$ — б.д.р: по определению

$$(z_{1s},\, z_{2s},\, \ldots,\, z_{ms})^\top = \mathtt{B}^{-1}\vec{A}_s \Rightarrow \vec{A}_s = \mathtt{B}\cdot(z_{1s},\, z_{2s},\, \ldots,\, z_{ms})^\top \Rightarrow$$
$$\Rightarrow \vec{A}_s = \sum_{i=1}^{m} z_{is}\vec{A}_{\sigma(i)}.$$

Так как $z_{rs} > 0$, то от противного легко доказывается, что столбцы матрицы

$$\mathtt{B}' = [\vec{A}_{\sigma(1)},\, \vec{A}_{\sigma(2)},\, \ldots,\, \vec{A}_{\sigma(r-1)},\, \vec{A}_s,\, \vec{A}_{\sigma(r+1)},\, \ldots,\, \vec{A}_{\sigma(n)}]$$

линейно независимы. Следовательно $\mathtt{B}'$ — базис и $\vec{x}(\overline{t})$ — б.д.р., а так как из (3.9) на стр. 48 следует, что

$$\forall t,\, 0 \leqslant t \leqslant \overline{t},\, w\big(\vec{x}(t)\big) = -z_{00} + z_{0s}t \leqslant -z_{00},$$

то $\vec{x}(\overline{t})$ — искомое б.д.р. с меньшим значением целевой функции.

✓ Рассмотрим случай $\overline{t} = 0$: в силу выбора индекса $r\colon x_{\sigma(r)}(\overline{t}) = x_{\sigma(r)}(0) = \overline{x}_{\sigma(i)} = z_{r0} = 0$. Так как $z_{rs} > 0$, то как и ранее нетрудно показать, что столбцы матрицы

$$\mathtt{B}' = [\vec{A}_{\sigma(1)},\, \vec{A}_{\sigma(2)},\, \ldots,\, \vec{A}_{\sigma(r-1)},\, \vec{A}_s,\, \vec{A}_{\sigma(r+1)},\, \ldots,\, \vec{A}_{\sigma(n)}]$$

линейно независимы. Следовательно $\mathtt{B}'$ — другой базис вершины $\overline{\vec{x}}$.

□

Замечание 3.5.12.1 (Геометрическая интерпретация).

▷ Эта лемма анализирует варианты возникающие при движении по ограниченному ребру.

ОПР 3.5.13 (Элементарного преобразования симплекс-таблицы)**.**
Преобразование симплекс-таблицы (соответственно, базиса), при котором происходит замена одного из базисных столбцов на другой столбец матрицы $\mathtt{A}$ *из числа небазисных, называется* элементарным преобразованием *симплекс-таблицы (соответственно, базиса).*

Пусть $z_{rs} \neq 0$, тогда базис

$$\mathtt{B} = [\vec{A}_{\sigma(1)},\, \vec{A}_{\sigma(2)},\, \ldots,\, \vec{A}_{\sigma(m)}]$$

заменяется на новый базис

$$[\vec{A}_{\sigma(1)},\, \vec{A}_{\sigma(1)},\, \ldots,\, \vec{A}_{\sigma(r-1)},\, \vec{A}_s,\, \vec{A}_{\sigma(r+1)},\, \ldots,\, \vec{A}_{\sigma(m)}].$$

$\alpha_i' = (z_{i0}',\, z_{i1}',\, \ldots,\, z_{in}'),\ i = 0,\, 1,\, \ldots,\, m$ строки новой симплекс-таблицы:

$$\begin{cases} \alpha_i' = \alpha_i - \frac{z_{is}}{z_{rs}} \cdot \alpha_r,\ i \neq r, \\ \alpha_r' = \frac{1}{z_{rs}} \alpha_r. \end{cases} \tag{3.12}$$

При этом: r-я строка, s-й столбец и элемент z_{rs} называются *ведущими*.
Соотношения (3.12) эквивалентны следующим преобразованиям

$$\begin{cases} z_{ij}' = z_{ij} - \frac{z_{is}}{z_{rs}} z_{rj},\ i \neq r, \\ z_{rj}' = \frac{1}{z_{rs}} z_{rj}. \end{cases}$$

3.6 Симплекс метод

3.6.1 Алгоритм: симплекс-метод

0. Построить симплекс-таблицу соответствующую заданному базисному допустимому решению, при этом таблица, естественно, будет прямо допустимой, то есть
$$\forall i = 1,\, 2,\, \ldots,\, m\colon z_{i0} \geqslant 0.$$

1. Если симплекс-таблица двойственно допустима, то есть
$$\forall j = 1,\, 2,\, \ldots,\, n\colon z_{0j} \geqslant 0,$$
то *КОНЕЦ* (получено оптимальное решение).

2. Иначе, выбрать ведущий столбец $s \,\big|\, z_{os} < 0,\ s \geqslant 1$.

3. Если $\left\{i \,\middle|\, z_{is} > 0,\ i \geqslant 1\right\} \neq \varnothing$, то выбрать ведущую строку r по правилу:
$$\frac{z_{r0}}{z_{rs}} = \min\left\{\frac{z_{i0}}{z_{is}} \,\middle|\, z_{is} > 0,\ i \geqslant 1\right\},$$
иначе *КОНЕЦ* (задача оказалась неразрешимой из-за неограниченности целевой функции).

4. Преобразовать симплекс-таблицу, положить $\sigma(r) := s$ и перейти на шаг 1.

Замечание 3.6.1.1 (Сохранение прямой допустимости при преобразованиях).

▷ Элементарное преобразование симплекс таблицы на шаге 3.6.1 сохраняет её прямо допустимость в силу выбора ведущей строки.

Замечание 3.6.1.2 (Случай невырожденных задач).

▷ Анализ сходимости симплекс-метода значительно упрощается, если предполагать невырожденность задачи. Так как в этом случае элементарное преобразование б.д.р. всегда приводит к новому б.д.р. со строго меньшим значением целевой функции (лемма 3.5.12 на стр. 51). Поэтому, раз число вершин конечно, то и алгоритм конечен.

Для почти всех известных вариантов симплекс метода построены примеры требующие экспоненциального числа шагов относительно размерности задачи. Однако, с точки зрения вычислительной практики, симплекс метод — самый эффективный метод решения задач линейного программирования.

Для произвольной задачи линейного программирования конечность симплекс метода может быть гарантирована только при уточнении правила выбора ведущего столбца и ведущей строки. Наиболее известный и простой вариант конечного симплекс метода — это лексикографический симплекс метод (см. учебное пособие Глебова Н.И. ([1]))

14.10.2005

3.7 Метод искусственного базиса

ОПР 3.7.1 (Искусственных переменных)**.**
Пусть $\vec{x} = (\vec{x}_1,\, \vec{x}_2,\, \ldots,\, \vec{x}_n)$, $\overline{\vec{x}} = (x_{n+1},\, x_{n+2},\, \ldots,\, x_{n+m})$, *рассмотрим вспомогательную задачу:*

$$\xi = x_{n+1} + x_{n+2} + \ldots + x_{n+m} \to \min$$
$$\mathtt{A}\vec{x} + \mathtt{A}'\overline{\vec{x}} = \vec{b}\ (\forall i = 1,\, 2,\, \ldots,\, m\colon \vec{a}_i\vec{x} + x_{n+i} = b_i)$$
$$\vec{x},\, \overline{\vec{x}} \geqslant 0$$

Переменные $\overline{\overline{x}} = (x_{n+1}, \ldots, x_{n+m})$ *называются* искусственными.

Считаем, что правые части $\forall i = 1, 2, \ldots, m\colon b_i$ *неотрицательны. Отсюда и условия* $\det \mathtt{A} = 1 \neq 0$ *имеем, что вектор* $\vec{z}^0 = [0, \vec{b}] \in \mathbb{R}^{n+m}$, *с компонентами* $\forall j = 1, 2, \ldots, n\colon z_j = 0$, $\forall i = 1, 2, \ldots, m\colon z_{n+i} = b_i$ — *б.д.р. вспомогательной задачи с базисом* $\mathtt{B} = \mathtt{A}$.

Из Критерия разрешимости имеем: вспомогательная задача разрешима и $\min \xi \geqslant 0$.

ОПР 3.7.2 (Симплекс метод для искусственного базиса)**.**

Применим симплекс метод к вспомогательной задаче, взяв на 0*-ом шаге симплекс таблицу соответствующую б.д.р.* $\vec{z}^0 = [0, \vec{b}]$.

Замечание 3.7.3 (Случаи решения вспомогательной задачи)**.**

▷ Возможны следующие случаи решения вспомогательной задачи:

а). $\min \xi > 0$ $\Leftrightarrow$задача (3.1) на стр. 39 не имеет допустимых решений (очевидно).

б). $\min \xi = 0$ $\Rightarrow \forall i = 1, 2, \ldots, m\colon x^*_{n+i} = 0$, значит, вектор $\vec{x}^*$ — допустимое решение задачи (3.1) на стр. 39, а следовательно, $\vec{x}^*$ — б.д.р. задачи (3.1).

3.7.4 Алгоритм симплекс метода с искусственным базисом

▷ Пусть вектор $\vec{z}^* = [\vec{x}^*, \overline{\overline{x}}^*]$ — оптимальное решение вспомогательной задачи. $\{\vec{A}_j \big| \, j \in I \setminus I'\} \cup \{\vec{E}_i \big| \, i \in I'\}$ — его базис, $\mathtt{B} = [A_{j \in I \setminus I'}, E_{i \in I'}]$ — базисная матрица[II]), следовательно,

$|I \setminus I'| + |I'| = m$ — так как ранг системы ограничений $\mathtt{A}\vec{x} + \mathtt{A}'\overline{\overline{x}} = \vec{b}$ равен m.

Пусть $\vec{Z}_k = [z_{1k}, \ldots, z_{mk}, \mu_{m+1k}, \ldots, \mu_{nk}]$, $k = 1, 2, \ldots, n,$ — столбец коэффициентов замещения оптимальной с.-т. Из их определения получаем

$$\vec{Z}_k = \mathtt{B}^{-1}\vec{A}_k = [\vec{A}_{j \in I \setminus I'}, \vec{E}_{i \in I'}]^{-1}\vec{A}_k \Rightarrow \vec{A}_k = [\vec{A}_{j \in I \setminus I'}, \vec{E}_{i \in I'}]\vec{Z}_k \Rightarrow$$
$$\Rightarrow \vec{A}_k = \sum_{j \in I \setminus I'} z_{jk}\vec{A}_j + \sum_{i \in I'} \mu_{ik}\vec{E}_i$$

▷ Возможны следующие случаи:

[II]) $I \setminus I' \subseteq \{1, 2, \ldots, n\}$, $I' \subseteq \{1, \ldots, m\}$.

I. $I' = \varnothing$: значит $|I \setminus I'| = m \Rightarrow \{\vec{A}_j \big| \, j \in I\}$ — базис б.д.р. $\vec{x}^*$, следовательно столбцы $\vec{Z}_k$, $k = 1, 2, \ldots, n,$ — коэффициенты замещения именно в этом базисе. Необходимо вычеркнуть из оптимальной с.-т. столбцы соответствующие переменным $x_{n+1}, x_{n+2}, \ldots, x_{n+m}$ и пересчитать 0-строку:

$$z_{00} = -(\vec{c}, \vec{x}^*), \forall k \in I\colon z_{0k} = 0, \ \forall k \notin I\colon z_{0k} = c_k - \sum_{j \in I} c_j z_{jk}.$$

II. $I' \neq \varnothing$ и $\exists r \in I'$, $\exists s \notin I \setminus I' \big| \, \mu_{rs} \neq 0$.

Значит, множество векторов

$$\{\vec{A}_j \big| \, j \in J''\} \cup \{\vec{E}_i \big| \, i \in I''\},$$

где

$$J'' = I \setminus I' \cup \{s\},\ I'' = I' \setminus \{r\}$$

— базис вершины $\vec{z}^* = (\vec{x}^*, 0)$[III]), следовательно, делаем элементарное преобразование с.-т. с ведущим элементом $\mu_{rs} \neq 0$, а значит, получим с.-т. соответствующую новому базису. Новая с.-т. остаётся оптимальной, так как в нулевом столбце все величины не изменяют своих значений включая z_{00}.

$$x^*_{n+r} = z_{r0} = 0,$$
$$z'_{i0} = z_{i0} - \frac{z_{is}}{z_{rs}} z_{r0} = z_{i0}$$

Поэтому элементы 0-столбца не меняются и остаются неотрицательными, но в этом базисе стало больше исходных переменных и меньше искусственных, на 1.

Для нового базиса проверяем условия I, II, III. Если выполняется II, то повторяем описанные действия. Так как множество I' конечно, то рано или поздно либо произойдёт случай I, либо III.

III. $I' \neq \varnothing$ и $\forall r \in I'$, $\forall s = 1, 2, \ldots, n\colon \mu_{rs} = 0$.

Это означает, что переменные $\forall i \in I'\colon x_{n+i}$, нельзя вывести из базиса, заменив их на подходящие исходные переменные. Покажем, что исходные ограничения $\vec{a}_i\vec{x} = b_i$ системы (3.1) на стр. 39 с номерами из множества $i \in I'$ являются избыточными.

[III]) легко доказывается от противного.

Действительно, $\vec{x}$ — допустимое решение (3.1) на стр. 39 тогда и только тогда, когда $(\vec{x}, 0)$ — решение системы уравнений:

$$x_{n+1} + x_{n+2} + \ldots + x_{n+m} = 0,$$
$$\forall k \in I \setminus I' \colon x_k + \sum_{j \in J \setminus \left(I \setminus I'\right)} z_{kj} x_j + \sum_{i \in I \setminus I'} z_{kn+i} x_{n+i} = z_{k0},$$
$$\forall i \in I' \colon x_{n+i} + \sum_{j \in J \setminus \left(I \setminus I'\right)} \mu_{ij} x_j + \sum_{l \in I \setminus I'} z_{kn+l} x_{n+l} = 0, \vec{x}, \overline{\vec{x}} \geqslant 0$$

или учитывая первое равенство, неотрицательность переменных и условие III: $\vec{x}$ — допустимое решение (3.1) на стр. 39 тогда и только тогда, когда

$$\forall i \in I \setminus I' \colon x_i + \sum_{j \in J \setminus \left(I \setminus I'\right)} z_{ij} x_j = z_{i0}, \ \vec{x} \geqslant 0,$$

которая эквивалентна системе уравнений

$$\begin{cases} \forall i \in I \setminus I' \colon \vec{a}_i \vec{x} = b_i, \\ \vec{x} \geqslant 0, \end{cases}$$

где матрица, образованная вектор-строками $\forall i \in I \setminus I' \colon \vec{a}_i$ имеет ранг $m - |I'|$.

Это ещё можно сформулировать так. Уравнения с номерами $i \in I'$ (в последней с.-т.) содержат исходные переменные с нулевыми коэффициентами и они приобрели такой вид при элементарных преобразованиях предшествующих с.-т. с ненулевыми ведущими элементами. При этом эти уравнения преобразовывались с помощью уравнений с номерами не принадлежащих множеству I' последней с.-т. Это означает, что уравнения из множества I' последней с.-т. являются нетривиальными линейными комбинациями уравнений из множества $I \setminus I'$ и их можно исключить.

3.7.5 Комментарий

- В с.-т. i-ая строка, при $i \geqslant 1$, соответствует i-му ограничению задачи (3.1) на стр. 39. Номер же базисной переменной в i-ой строке обозначается как $\sigma(i)$. Здесь принято соглашение, что исходные x_j переменные перенумерованы так, чтобы $\sigma(i) = i$ для $i \in I \setminus I'$. Такая перенумерация возможна так как ограничения разбиваются ровно на две части. Ограничения с номерами i из множества $I' \subseteq I$, содержащие вспомогательные переменные x_{n+i}, в качестве базисных. И ограничения с номерами i из множества

$I \setminus I'$, содержащие исходные переменные $x_{\sigma(i)}$ в качестве базисных. Вспомогательные переменные вводились так, чтобы в i-ом ограничении была переменная x_{n+i}.

3.8 Анализ чувствительности

3.8.1 Цель исследования

- Как влияет на оптимальное решение задачи (3.1) на стр. 39 возмущение входных данных задачи. Необходимость таких исследований вытекает из невозможности всегда иметь достоверную информацию о входных данных.

Лемма 3.8.2 (Оптимальное решение двойственной задачи)**.**

- Пусть

  $\vec{x}^*$ — оптимальное б.д.р. задачи (3.1) на стр. 39 и $\mathtt{B}$, $\mathtt{T}$ — соответствующие ему базис и симплекс таблица.

- Тогда

  Вектор $y^* = \vec{c}_B \mathtt{B}^{-1}$ — оптимальное решение двойственной задачи.

- **Доказательство**.
  - По условию с.-т. $\mathtt{T}$ прямо и двойственно оптимальна тогда и тогда, когда
  $$\forall\ i, \forall\ j \colon z_{i0} \geqslant 0,\ z_{0j} \geqslant 0 \Leftrightarrow z_{i0} = \mathtt{B}^{-1}\vec{b} \geqslant 0 \text{ и } z_{0j} = c_j - \vec{c}_B \mathtt{B}^{-1} \vec{A}_j \geqslant 0$$
  - Из второго неравенства следует, что вектор $\vec{y}^* = \vec{c}_B \mathtt{B}^{-1}$ удовлетворяет неравенству:
  $$\vec{c}_N - \vec{y}^* \mathtt{N} \geqslant \vec{0}.$$
  Но по определению $\vec{y}^* \colon \vec{y}^* \mathtt{B} = \vec{c}_B$, следовательно:
  $$\vec{y}^* \mathtt{A} \leqslant \vec{0}.$$
  А это означает, что $\vec{y}^*$ — допустимое решение двойственной задачи.
  - Утверждение леммы следует из равенств леммы 2.3 на стр. 34:
  $$w\left(\vec{x}^*\right) = \vec{c}_B \vec{x}^*_B = \vec{c}_B \mathtt{B}^{-1} \vec{b} = \vec{y}^* \vec{b} = \vec{z}\left(\vec{y}^*\right).$$

□

Из леммы 3.8.2 получим следующее представление оценок замещения

$$\forall j \colon z_{0j} = c_j - \vec{c}_B \mathtt{B}^{-1} \vec{A}_j = c_j - \vec{y}^* \vec{A}_j. \tag{3.13}$$

Далее предполагается, что имеется с.-т., отвечающая оптимальному решению задачи.

3.8.3 Возмущение целевой функции

3.8.3.1 Возмущение коэффициентов целевой функции

▷ Пусть

j — небазисная переменная и $\overline{c}_j = c_j + \delta$, где δ — возмущение.

▷ Тогда

$$\overline{z}_{0j} = \overline{c}_j - \vec{y}^* \vec{A}_j = c_j + \delta - \vec{y}^* \vec{A}_j = z_{0j} + \delta.$$

▷ Следовательно

Решение $\vec{x}^*$ остается оптимальным б.д.р. тогда и только тогда, когда

$$\overline{z}_{0j} \geqslant 0 \Leftrightarrow z_{0j} + \delta \geqslant 0 \Leftrightarrow z_{0j} \geqslant -\delta > -\infty.$$

3.8.3.2 Диапазон устойчивости для базисной переменной

▷ Пусть

$\sigma(i)$ — индекс базисной переменной и $\overline{c}_{\sigma(i)} = c_{\sigma(i)} + \delta$, где δ — возмущение и $\sigma(i) = i$.

▷ Тогда

Мы можем записать, что $\overline{c}_B = c_B + \delta e_i$, следовательно, новая оценка небазисной переменной с номером j вычисляется по формулам:

$$\overline{z}_{0j} = c_j - (c_B + \delta e_i)\mathtt{B}^{-1}\vec{A}_j = c_j - \vec{c}_B \mathtt{B}^{-1}\vec{A}_j - \delta z_{ij} = z_{0j} - \delta z_{ij}.$$

Решение остается оптимальным тогда и только тогда, когда

$$\overline{z}_{0j} = z_{0j} - \delta z_{ij} \geqslant 0 \text{ для всех небазисных } j.$$

▷ Следовательно

Диапазон устойчивости для базисной переменной задается неравенствами:

$$\max_{j: z_{ij} > 0} \frac{z_{0j}}{z_{ij}} \leqslant \delta \leqslant \min_{j: z_{ij} < 0} \frac{z_{0j}}{z_{ij}}.$$

▷ Таким образом

При изменении δ в указанных пределах базис остается оптимальным, но значение целевой функции меняется в соответствии с формулой

$$\vec{z}(\delta) = (\vec{c}_B + \delta e_i)\vec{x}_B^* = \vec{c}_B \vec{x}_B^* + \delta e_i \vec{x}_B^* = \vec{z}^* + \delta \vec{x}_i^*.$$

21.10.2005

3.8.4 Возмущение правых частей

3.8.4.1 Возмущение правых частей

▷ Пусть

$\vec{b}(\delta) = \vec{b} + \delta \vec{e}_r$.

▷ Тогда

Текущий базис сохранит оптимальность, если он останется прямо допустимым при изменении δ. То есть решение системы

$$\mathtt{B}\vec{x}(\delta)_B = \vec{b}(\delta)$$

должно удовлетворять неравенствам $\vec{x}_B(\delta) \geqslant 0$.

▷ Имеем

$$\vec{x}_B(\delta) = \mathtt{B}^{-1}(\vec{b} + \delta \vec{e}_r) = \mathtt{B}^{-1}\vec{b} + \delta \mathtt{B}^{-1}\vec{e}_r = \vec{x}_B^* + \delta \mathtt{B}^{-1}\vec{e}_r \geqslant 0 \Rightarrow$$
$$\Rightarrow \forall i\colon x_i^* + \delta \beta_{ir} \geqslant 0,$$

где $\mathtt{B}^{-1}\vec{e}_r = (\beta_{1r}, \beta_{2r}, \ldots, \beta_{mr})^\top$.

▷ Следовательно

- Диапазон устойчивости задается неравенствами:

$$\max_{i: \beta_{ir} > 0} \frac{-x_i^*}{\beta_{ir}} \leqslant \delta \leqslant \min_{i: \beta_{ir} < 0} \frac{x_i^*}{-\beta_{ir}}.$$

- Значение целевой функции меняется в соответствии с формулой

$$\vec{z}(\delta) = \vec{b}(\delta)\vec{y}^* = (\vec{b} + \delta \vec{e}_r)\vec{y}^* = \vec{z}^* + \delta y_r^*.$$

3.8.5 Возмущение матрицы ограничений

Так как в оптимальном решении значения небазисных компонент нулевые, то изменение коэффициентов a_{ij} небазисных столбцов не влияет на допустимость решения $\vec{x}^*$, но оно может стать неоптимальным.

3.8.5.1 Возмущение матрицы ограничений

▷ Пусть

j — номер небазисной переменной и $a_{kj}(\delta) = a_{kj} + \delta$

▷ <u>Тогда</u>

Для того, чтобы решение $\vec{x}^*$ осталось оптимальным необходимо, чтобы осталась неотрицательной оценка замещения $z_{0j}(\delta)$.

▷ <u>То есть</u>

Учитывая лемму 3.8.2 на стр. 58 получаем:

$$z_{0j}(\delta) = c_j - \vec{y}^*(\vec{A}_j + \delta\vec{e}_k) = c_j - \vec{y}^*\vec{A}_j - \vec{y}^*\delta\vec{e}_k = z_{0j} - y_k^*\delta \geqslant 0.$$

▷ <u>Итак</u>

$$z_{0j} - y_k^*\delta \geqslant 0 \Leftrightarrow \delta \leqslant z_{0j}/y_k^*.$$

3.8.5.2 Диапазоны устойчивости

▷ <u>Пусть</u>

r — номер базисной переменной и $a_{kr}(\delta) = a_{kr} + \delta_{kr}$, $\mathtt{B}(\delta_{kr})$ — возмущенная базисная матрица:

$$\mathtt{B}(\delta_{kr}) = \mathtt{B} + \delta_{kr}\vec{e}_k\vec{e}_r^{\mathsf{T}} = \mathtt{B}(\mathtt{I} + \delta_{kr}\mathtt{B}^{-1}\vec{e}_k\vec{e}_r^{\mathsf{T}}).$$

Обратная матрица имеет следующий вид:

$$\mathtt{B}(\delta_{kr})^{-1} = \left(\mathtt{I} - \frac{\delta_{kr}}{1 + \delta_{kr}\beta_{rk}}\mathtt{B}^{-1}\vec{e}_k\vec{e}_r^{\mathsf{T}}\right)\mathtt{B}^{-1}.$$

▷ <u>Положим</u>

$$\beta(\delta_{kr}) = \frac{1}{1/\delta_{kr} + \beta_{rk}}.$$

▷ <u>Тогда</u>

$$\mathtt{B}(\delta_{kr})^{-1} = \left(\mathtt{I} - \beta(\delta_{kr})\mathtt{B}^{-1}\vec{e}_k\vec{e}_r^{\mathsf{T}}\right)\mathtt{B}^{-1}. \tag{3.14}$$

I. *Прямая допустимость.*

Решение $\vec{x}(\delta_{kr})$ системы уравнений $\mathtt{B}(\delta_{kr})\vec{x} = \vec{b}$ прямо допустимо, если $\vec{x}(\delta_{kr}) \geqslant 0$, то есть

$$\vec{x}(\delta_{kr}) = \mathtt{B}(\delta_{kr})^{-1}\vec{b} \geqslant 0.$$

Подставим в это соотношение формулу (3.14) и получим:

$$\begin{gathered}\vec{x}(\delta_{kr}) = \mathtt{B}(\delta_{kr})^{-1}\vec{b} = \mathtt{B}^{-1}\vec{b} - \beta(\delta_{kr})\mathtt{B}^{-1}\vec{e}_k\vec{e}_r^{\mathsf{T}}\mathtt{B}^{-1}\vec{b} = \\ = \vec{x}_B^* - \beta(\delta_{kr})\mathtt{B}^{-1}\vec{e}_k\vec{e}_r^{\mathsf{T}}\vec{x}_B^* = \vec{x}_B^* - \beta(\delta_{kr})\mathtt{B}_k^{-1}x_r^* \geqslant 0 \equiv \\ \equiv \forall i = 1, 2, \ldots, m\colon x_i^* - \beta(\delta_{kr})\beta_{ik}x_r^* \geqslant 0.\end{gathered}$$

Имеем следующий диапазон на значения величины $\beta(\delta_{kr})$:

$$\max_{\{i:x_r^*\beta_{ik}<0\}} \frac{x_i^*}{x_r^*\beta_{ik}} \leqslant \beta(\delta_{kr}) \leqslant \min_{\{i:x_r^*\beta_{ik}>0\}} \frac{x_i^*}{x_r^*\beta_{ik}}.$$

II. *Двойственная допустимость.*

Рассмотрим решение $\vec{y}(\delta_{kr})$ системы уравнений $\vec{y}\mathtt{B}(\delta_{kr}) = c_B$. Базис сохраняет двойственную допустимость, если возмущение δ_{kr} таково, что для всех небазисных j выполнено:

$$z_{0j} = c_j - \vec{y}(\delta_{kr})\vec{A}_j \geqslant 0. \tag{3.15}$$

Но

$$\begin{gathered}\vec{y}(\delta_{kr}) = c_B\mathtt{B}(\delta_{kr})^{-1} = c_B\left(\mathtt{I} - \beta(\delta_{kr})\mathtt{B}^{-1}\vec{e}_k\vec{e}_r^{\mathsf{T}}\right)\mathtt{B}^{-1} = \\ = c_B\mathtt{B}^{-1} - \beta(\delta_{kr})c_B\mathtt{B}^{-1}\vec{e}_k\vec{e}_r^{\mathsf{T}}\mathtt{B}^{-1} = \vec{y}^* - \beta(\delta_{kr})\vec{y}^*\vec{e}_k\vec{e}_r^{\mathsf{T}}\mathtt{B}^{-1} = \\ = \vec{y}^* - \beta(\delta_{kr})y_k^*\vec{e}_r^{\mathsf{T}}\mathtt{B}^{-1}.\end{gathered}$$

Подставим в формулу (3.15) вместо $\vec{y}(\delta_{kr})$ полученное выражение и получим:

$$\begin{gathered}z_{0N}(\delta_{kr}) = c_N - \left(\vec{y}^* - \beta(\delta_{kr})y_k^*\vec{e}_r^{\mathsf{T}}\mathtt{B}^{-1}\right)N = \\ = c_N - \vec{y}^*N + \beta(\delta_{kr})y_k^*\vec{e}_r^{\mathsf{T}}\mathtt{B}^{-1}N = \\ = z_{0N} + \beta(\delta_{kr})y_k^*\left(z_{rm+1}, z_{rm+2}, \ldots, z_{rj}, \ldots, z_{rn}\right) \geqslant 0.\end{gathered}$$

Таким образом получили интервал, в котором должна содержаться величина $\beta(\delta_{kr})$:

$$\max_{\{i:y_k^*z_{rj}>0\}} -\frac{z_{0j}}{y_k^*z_{rj}} \leqslant \beta(\delta_{kr}) \leqslant \min_{\{i:y_k^*z_{rj}<0\}} -\frac{z_{0j}}{y_k^*z_{rj}}.$$

Теорема 3.8.5.3 (О возмущении базисных коэффициентов)**.**

▷ При малых возмущениях δ_{kr} изменения в $\vec{x}_B^*$, $\vec{y}^*$, z_{0N} нелинейны относительно величины δ_{kr}, но эти изменения являются линейными относительно величины

$$\beta(\delta_{kr}) = \frac{1}{1/\delta_{kr} + \beta_{rk}}.$$

ОПР 3.8.6 (Маргинального значения)**.**

Под ценой, маргинальным значением *или* теневой ценой *ограничения с номером i будем понимать величину, которая задает скорость изменения целевой функции при изменении правой части ограничения b_i.*

28.10.2005

3.9 Лексикографический двойственный симплекс-метод

Лемма 3.9.1 (Соотношение между решениями двойственных систем).

▷ Пусть

$\mathtt{B}$ – базисная матрица, $\overline{\vec{x}}(\mathtt{B})$ – решение системы уравнений

$$\mathtt{B}\vec{x}_B = \vec{b}, \vec{x}_N = 0,$$

а $\overline{\vec{y}}(\mathtt{B})$, соответственно, системы уравнений

$$\vec{y}\mathtt{B} = \vec{c}.$$

▷ Тогда

$$(\vec{c}, \overline{\vec{x}}(\mathtt{B})) = (\vec{b}, \overline{\vec{y}}(\mathtt{B})).$$

Пусть $\mathtt{B}$ — двойственно допустимый базис, $S' = \{\tau(1), \tau(2), \ldots, \tau(\ell)\}$, $\ell = n - m$ — множество номеров небазисных переменных, а S — базисных переменных. Перепишем соотношения (3.9) и (3.10) на стр. 48:

$$x_i = z_{i0} + \sum_{j \in S'} z_{ij}(-x_j), \tag{3.16}$$

добавим к системе уравнений (3.16) тождественные соотношения $x_i = x_i$ для небазисных переменных:

$$x_i = (-1)(-x_i), i \in S'. \tag{3.17}$$

Симплекс-таблица будет состоять из коэффициентов правых частей соотношений (3.16) и (3.17):

	1	$-x_{m+1}$	$\ldots$	$-x_n$
x_0	z_{00}	z_{01}	$\ldots$	$z_{0\ell}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	z_{i0}	z_{i1}	$\ldots$	z_{il}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{m+1}	0	-1	$\ldots$	0
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
x_n	0	0	$\ldots$	-1

$$\beta_j = (z_{0j}, z_{1j}, \ldots, z_{nj})^\top, j = 0, 1, \ldots, \ell.$$

Тогда система (3.16), (3.17) эквивалентна векторному уравнению:

$$(x_0, x_1, \ldots, x_n)^\top = \beta_0 + \sum_{j \in S'} \beta_j(-x_j). \tag{3.18}$$

Если $z_{rs} \neq 0$, $r \in S$, $s \geqslant 1$, то возможно элементарное преобразование базиса

$$x_r \to x_{\tau(s)}.$$

Выразим переменную $x_{\tau(s)}$ из r-го уравнения системы (3.16) на стр. 63:

$$x_{\tau(s)} = \frac{1}{z_{rs}} \cdot \left(z_{i0} + \sum_{j \neq s} z_{rj}(-x_{\tau(j)}) - x_r \right)$$

и заменим ее в правой части (3.18) на стр. 64:

$$(x_0, x_1, \ldots, x_n)^\top = \left(\beta_0 - \frac{z_{r0}}{z_{rs}} \beta_s \right) + \\ + \sum_{j \neq s} \left(\beta_j - \frac{z_{rj}}{z_{rs}} \beta_s \right)(-x_{\tau(j)}) + \left(\frac{-1}{z_{rs}} \beta_s \right)(-x_r).$$

Итак, элементарное преобразование базиса $x_r \to x_{\tau(s)}$ приводит к с.-т. со столбцами:

$$\begin{cases} \beta'_j = \beta_j - \frac{z_{rj}}{z_{rs}} \cdot \beta_s, \ j \neq s, \\ \beta'_s = \left(\frac{-1}{z_{rs}} \right) \beta_s. \end{cases} \tag{3.19}$$

ОПР 3.9.2 (Лексикографического сравнения).

а). *Пусть* $\vec{\beta} = (\beta_0, \beta_1, \ldots, \beta_n)^\top \in \mathbb{R}^{n+1}$ *— вектор-столбец, тогда он* лексикографически больше *нуля* ($\vec{\beta} \succ \vec{0}$), *если* $\beta_p > 0$, *где* $p = \min\{i \mid \beta_i \neq 0\}$.

б). *Пусть* $\beta', \beta'' \in \mathbb{R}^{n+1}$, *тогда вектор* β' лексикографически больше *вектора* β'' ($\beta' \succ \beta''$), *если* $\beta' - \beta'' \succ 0$.

ОПР 3.9.3 (Нормальной симплекс таблицы).

Симплекс-таблицу будем называть нормальной, *если каждый ее столбец* $\vec{\beta}_j$, $j = 1, 2, \ldots, \ell$ *лексикографически больше нуля.*

3.9.4 Лексикографический двойственный симплекс-метод

0. Начать с нормальной симплекс-таблицы.

1. Если симплекс-таблица прямо допустима, то есть $\forall i = 1, 2, \ldots, n$: $z_{i0} \geqslant 0$, то КОНЕЦ (получено оптимальное решение)

2. Иначе выбрать ведущую строку с индексом $r \big| z_{r0} < 0,\ r \geqslant 1$.

3. Если $\{j \big| z_{rj} < 0, j \geqslant 1\} \neq \varnothing$, то выбрать ведущий столбец s по правилу:

$$\frac{1}{|z_{rs}|}\beta_s = \operatorname{lexmin}\left\{\frac{1}{|z_{rj}|}\beta_j \Big| z_{rj} < 0,\ j \geqslant 1\right\},$$

иначе КОНЕЦ (задача неразрешима из-за несовместности ограничений задачи (3.1) на стр. 39).

4. Преобразовать симплекс-таблицу, положить $\tau(s) := r$ и перейти на шаг 1.

Замечание 3.9.4.1 (Свойства преобразований).

- ▷ Преобразование (3.19) на стр. 64 симплекс-таблицы на шаге 4 сохраняет ее нормальность.
- ▷ Так как $z_{r0} < 0, z_{rs} < 0$ и $\beta_s \succ 0$, то

$$\beta_0 - \frac{z_{r0}}{z_{rs}}\beta_s \prec \beta_0.$$

Итак, базисы не могут повторятся, следовательно, метод конечен.

- ▷ Решения $\vec{\overline{x}}(\mathtt{B})$ и $\vec{\overline{y}}(\mathtt{B})$ назовем *соответствующими базису* $\mathtt{B}$.

3.9.5 Идеи симплекс-методов

- ▷ Идея прямого симплекс-метода: двигаться по б.д.р. до тех пор пока решение $\vec{\overline{y}}(\mathtt{B})$, соответствующее текущему базису, не окажется допустимым решением двойственной задачи. Тогда в силу леммы 3.9.1 на стр. 63 и леммы 2.3 на стр. 34 решения $\vec{\overline{x}}(\mathtt{B})$ и $\vec{\overline{y}}(\mathtt{B})$ оптимальные решения.
- ▷ Идея двойственного симплекс-метода: двигаться по решениям системы уравнений $\vec{y}\mathtt{B} = \vec{c}$, которые являются допустимыми решениями двойственной задачи, до тех пор, пока решение $\vec{\overline{x}}(\mathtt{B})$, соответствующее текущему базису, не окажется допустимым решением прямой задачи.

Глава 4

Целочисленное линейное программирование

ОПР 4.1 (Задачи ЦЛП (IP))**.**
Задача ЦЛП (IP) это задача:

$$\begin{cases} (\vec{c}, \vec{x}) \to \min \\ \mathtt{A}\vec{x} = \vec{b}, \\ \vec{x} \geqslant 0, \\ x_j \text{ — \textit{целое}, } j = 1, 2, \ldots, n. \end{cases} \tag{4.1}$$

ЛП-релаксация *задачи (IP) — задача ЛП (3.1) на стр. 39.*

4.2 Методы отсечения

4.2.1 Метод отсечения

- ▷ Пусть $\vec{x}^0$ — оптимальное решение ЛП-релаксации задачи (IP).
- ▷ Если $\vec{x}^0 \in \mathbb{Z}^n$, то $\vec{x}^0$ — оптимальное решение задачи (IP).
- ▷ Иначе добавить к ЛП-релаксации новое ограничение (*отсечение*) следующего вида:
  - ∘ $\vec{x}^0$ этому ограничению не удовлетворяет (отсекается);
  - ∘ все допустимые решения задачи ЦЛП остаются допустимыми решениями новой задачи ЛП.

- ▷ Решаем новую задачу ЛП и повторяем указанные шаги до получения решения задачи ЦЛП, либо обнаружения ее неразрешимости.

Пусть линейная функция $d_0 - \sum_j d_j x_j$ принимает целые неотрицательные значения на множестве допустимых решений задачи (IP) и $h \neq 0$ (если h — целое, то требование неотрицательности избыточно). Рассмотрим следующее линейное уравнение

$$\xi = d_0 - \sum_j d_j x_j. \tag{4.2}$$

Тогда для любого допустимого решения $\vec{x}$ задачи (IP) имеют место следующие соотношения:

$$h\xi + \sum_j h d_j x_j = h d_0,$$

так как x_j и ξ неотрицательны, то

$$\lfloor h \rfloor \xi + \sum_j \lfloor h d_j \rfloor x_j \leqslant \lfloor h d_0 \rfloor.$$

Подставим вместо ξ выражение (4.2):

$$\sum_j (\lfloor h d_j \rfloor - \lfloor h \rfloor d_j) x_j \leqslant \lfloor h d_0 \rfloor - \lfloor h \rfloor d_0.$$

Если к задаче (IP) добавить ограничения:

$$u = (\lfloor h d_0 \rfloor - \lfloor h \rfloor d_0) - \sum_j (\lfloor h d_j \rfloor - \lfloor h \rfloor d_j) x_j \geqslant 0, \ u \text{ — целое},$$

то получим эквивалентную задачу ЦЛП.

04.11.2005

4.3 Первый алгоритм Гомори

4.3.1 Описание первого алгоритма Гомори

0. Начать с нормальной симплекс-таблицы (для задачи (3.1) на стр. 39), положить $\nu := 0$.

1. Если симплекс-таблица прямо допустима и все элементы z_{i0}, $i = 1, \ldots, n$, целые, то КОНЕЦ (получено оптимальное решение задачи (3.1) на стр. 39).

2. Если симплекс-таблица прямо допустима, то выбрать такое минимальное $p \geqslant 1$, что z_{p0} — нецелое, положить $\nu := \nu+1$. Строку с номером p назовем *производящей*. Этой строке соответствует уравнение
$$x_p = z_{p0} - \sum_{j=1}^{\ell} z_{pj} x_{\tau(j)},$$
по которому строится дополнительное ограничение согласно описанному ранее способу при $h = 1$ (роль ξ играет x_p):
$$x_{n+\nu} = -f_{p0} - \sum_{j=1}^{\ell} (-f_{pj}) x_{\tau(j)} \geqslant 0,$$
где f_{pj} — дробная часть числа z_{pj} ($z_{pj} = \lfloor z_{pj} \rfloor + f_{pj}$, $0 \leqslant f_{pj} < 1$). К симплекс-таблице добавляется $(n+1)$-ая строка (*отсечение Гомори*), соответствующая дополнительному ограничению с базисной переменной $x_{n+\nu}$.

3. Выбрать ведущую строку r: $z_{r0} < 0$, $r \geqslant 1$.

4. Если $\{j \,|\, z_{rj} < 0,\, j \geqslant 1\} \neq \varnothing$, то выбрать ведущий столбец s:
$$\frac{1}{|z_{rs}|}\beta_s = \operatorname{lexmin}\left\{ \frac{1}{|z_{rj}|}\beta_j \,\Big|\, z_{rj} < 0,\, j \geqslant 1 \right\},$$
иначе КОНЕЦ (текущая задача ЛП, а следовательно, и исходная задача ЦЛП, неразрешима ввиду несовместности ее ограничений).

5. Преобразовать симплекс-таблицу; положить $\tau(s) := n + \nu$ и отбросить $(n+1)$-ю строку, если таковая имелась, иначе $\tau(s) := r$; перейти на шаг 1.

Пусть $\vec{x}^0 = (z_{10}, z_{20}, \ldots, z_{n0})^\intercal$ — б.д.р. (задачи (3.1) на стр. 39), соответствующее текущей с.-т. в момент введения отсечения Гомори (является оптимальным решением ЛП-релаксации), так как z_{p0} — нецелое, то $f_{p0} > 0$, следовательно,
$$x_{n+\nu}(\vec{x}^0) = -f_{p0} < 0.$$
Итак, $\vec{x}^0$ отсекается.

Пусть
$$x_{n+\nu} = -f_{p0} - \sum_{j=1}^{\ell} (-f_{pj}) x_{\tau(j)} \geqslant 0,$$
отсечение Гомори (шаг 2). Так как $-f_{p0} < 0$, то $x_{n+\nu}$ – единственная отрицательная базисная переменная. Следовательно, с.-т. двойственно допустима, но не прямо допустима.

Итак, в качестве ведущей будет выбрана $n + 1$ строка (единственность). При элементарном преобразовании с.-т. переменная $x_{n+\nu}$ станет небазисной. Ведущая строка превращается в тождество
$$x_{n+\nu} = (-1)(-x_{n+\nu})$$
и на шаге 5 удаляется из с.-т. Поэтому максимальное число учитываемых отсечений не превосходит числа небазисных переменных.

4.3.2 Конечность первого алгоритма Гомори

1. Пусть известна некоторая (условная) нижняя граница M для оптимального значения целевой функции x_0.

2. Пусть функция x_0 целочисленна на множестве допустимых решений задачи ЦЛП. (Тогда нулевая строка с.-т. может (и будет) использоваться в качестве производящей.)

Итерация алгоритма (шаги с 1 по 5) это LD-итерация (не вводится отсечение) + итерация Гомори (вводится отсечение). Элементы и столбцы симплекс-таблицы, полученной после выполнения первых t итераций, обозначим z_{ij}^t и β_j^t соответственно (z_{ij}^0 — элементы начальной симплекс-таблицы).

Теорема 4.3.2.1 (Конечность алгоритма Гомори)**.**

▷ Первый алгоритм Гомори конечен.

▷ <u>**Доказательство**</u>.

- Пусть при решении задачи алгоритмом Гомори выполняется бесконечная последовательность итераций, тогда из описания LD-метода имеем
$$\beta_0^0 \succ \beta_0^1 \succ \beta_0^2 \succ \ldots \succ \beta_0^t \succ \beta_0^{t+1} \succ \ldots. \tag{4.3}$$

  Из конечности LD-метода имеем, что число LD-итераций конечно, а число итераций Гомори бесконечно. Пусть $t_\nu + 1$, $\nu = 1, 2, \ldots$ — порядковые номера этих итераций. Из (4.3) имеем
$$z_{00}^0 \geqslant z_{00}^1 \geqslant z_{00}^2 \geqslant \ldots \geqslant z_{00}^t \geqslant z_{00}^{t+1} \geqslant \ldots. \tag{4.4}$$

- Рассмотрим подпоследовательность
$$z_{00}^{t_1},\, z_{00}^{t_2},\, \ldots,\, z_{00}^{t_\nu},\, \ldots, \tag{4.5}$$

состоящую из элементов z_{00} с.-т., которые являются входными для итераций Гомори. Пусть $z_{00}^{t_\nu}$ — нецелое, тогда на итерации $t_\nu+1$ нулевая строка будет производящей и

$$z_{00}^{t_\nu+1} = z_{00}^{t_\nu} - z_{0s}^{t_\nu} \cdot \frac{f_{00}^{t_\nu}}{f_{0s}^{t_\nu}}.$$

Так как $z_{0s}^{t_\nu} \geqslant 0$ (с.-т. двойственно допустима) и $z_{0s}^{t_\nu} \geqslant f_{0s}^{t_\nu}$, то

$$z_{00}^{t_\nu+1} \leqslant z_{00}^{t_\nu} - f_{00}^{t_\nu} = \lfloor z_{00}^{t_\nu} \rfloor < z_{00}^{t_\nu}.$$

Таким образом, каждый интервал $(z,\, z+1)$, где z — целое, содержит не более одного члена из последовательности (4.5).

Отсюда, монотонности и ограниченности снизу этой последовательности следует, что она стабилизируется на некотором целом значении. Тогда

$$\exists T_0 \Big| \forall t \geqslant T_0\colon z_{00}^t = \overline{z}_{00}.$$

Тогда в силу (4.3) (строгого лексикографического убывания нулевого столбца):

$$z_{10}^{T_0} \geqslant z_{10}^{T_0+1} \geqslant \ldots \geqslant z_{10}^{t} \geqslant z_{10}^{t+1} \geqslant \ldots. \tag{4.6}$$

- Изучим подпоследовательность состоящую из элементов z_{10} симплекс-таблицы, которые являются входными для итераций Гомори с номерами $t_\nu+1$ строго большими T_0. Пусть $z_{10}^{t_\nu}$ — нецелое, тогда на итерации $t_\nu+1$ первая строка будет производящей и

$$z_{10}^{t_\nu+1} = z_{10}^{t_\nu} - z_{1s}^{t_\nu} \cdot \frac{f_{10}^{t_\nu}}{f_{1s}^{t_\nu}}, \tag{4.7}$$

где $0 < f_{10}^{t_\nu} < 1$ и $0 < f_{1s}^{t_\nu} < 1$.

Покажем, что $z_{1s}^{t_\nu} \geqslant 0$, допустим противное: $z_{1s}^{t_\nu} < 0$; тогда из лексикографической положительности столбца $\beta_s^{t_\nu}$ следует неравенство $z_{0s}^{t_\nu} > 0$. Но из формулы

$$z_{00}^{t_\nu+1} = z_{00}^{t_\nu} - z_{0s}^{t_\nu} \cdot \frac{f_{10}^{t_\nu}}{f_{1s}^{t_\nu}}$$

и условий $0 < f_{10}^{t_\nu} < 1$, $0 < f_{1s}^{t_\nu} < 1$ следует, что $z_{00}^{t_\nu+1} < z_{00}^{t_\nu}$. Противоречие.

Таким образом из (4.7), учитывая неравенства $z_{1s}^{t_\nu} \geqslant 0$ и $\frac{z_{1s}^{t_\nu}}{f_{1s}^{t_\nu}} \geqslant 1$, следует

$$z_{10}^{t_\nu+1} \leqslant z_{10}^{t_\nu} - f_{10}^{t_\nu} = \lfloor z_{10}^{t_\nu} \rfloor < z_{10}^{t_\nu}.$$

Это означает, что каждый интервал $(z,\, z+1)$, где z — целое, содержит не более одного члена из последовательности (4.6). Так как симплекс-таблицы с номерами t_ν прямо допустимы, то $z_{10}^{t_\nu} \geqslant 0$.

Итак, последовательность (4.6) на стр. 71 монотонно убывающая и ограниченна снизу, т.е. она стабилизируется на некотором целом неотрицательном значении (прямо допустимость). Тогда

$$\exists T_1 \geqslant T_0 \Big| \forall t \geqslant T_1\colon z_{10}^t = \overline{z}_{10}.$$

- Аналогичные утверждения можно провести и для оставшихся компонент, вплоть до n-ой.

Следовательно, существует такой номер T_n, что для всех $t \geqslant T_n$ и для всех $i = 1, 2, \ldots, n$ выполнено:

$$z_{i0}^t = \overline{z}_{i0},\ \text{где } \overline{z}_{i0} \in \mathbb{Z}^+.$$

Подобное утверждение противоречит предположению о бесконечности числа итераций.

□

Глава 5

Метод ветвей и границ

11.11.2005

5.1 Принцип метода

- ▷ Простой перебор: наилучшее текущее решение (рекорд) сравнивается с очередным допустимым решением.

- ▷ Метод ветвей и границ (МВГ): рекорд сравнивается с целым множеством допустимых решений.

  МВГ — метод, который сводит поиск оптимального решения посредством последовательного разбиения множества допустимых решений на все более мелкие подмножества и последующего сравнения этих подмножеств с рекордом.

Рассмотрим задачу:

$$\min\{f(\vec{x}) \big| \vec{x} \in Q\}.$$

Известно разбиение множества Q на конечное число атомарных множеств. Содержательно то, что атомарное множество это подмножество Q, на котором исходная задача легко решается (точно или приближенно).

Пусть $\vec{x}(d)$ — наилучшее достижимое решение задачи на атомарном множестве d, для которой существует малотрудоёмкий алгоритм вычисления. Например, когда множество Q — конечно, роль атомарных могут играть одноточечные множества. Если Q — гиперкуб, а целевая функция непрерывна, то в качестве атомарных можно выбрать достаточно маленькие гиперкубики.

При описании алгоритма МВГ рассматриваем лишь такие подмножества Q, которые есть объединение атомарных множеств (разложимые подмножества). Будем предполагать, что независимо от природы множества Q число атомарных множеств конечно.

ОПР 5.2 (Функции ветвления)**.**

Функцию $b(d)$, определенную на разложимых подмножествах множества Q и ставящую в соответствие множеству d определенное разбиение его на несобственные разложимые подмножества, будем называть функцией ветвления.

ОПР 5.3 (Нижней границы)**.**

Вещественную функцию $H(d)$, определенную на разложимых подмножествах множества Q и такую, что

$$0 < H(d) \leqslant \min_{\vec{x} \in d} f(\vec{x})$$

назовем нижней границей.

Функция $H(d)$ — невозрастающая, то есть такая, что $H(d_1) \geqslant H(d_2)$, если $d_1 \subseteq d_2$.

Алгоритм, реализующий метод ветвей и границ, состоит из конечной последовательности однотипных шагов. На каждом шаге рассматривается некоторое разбиение $t_1, t_2, \ldots, t_L$ множества еще неотброшенных допустимых решений и некоторый рекорд $\vec{x}^0$.

5.4 Алгоритм метода ветвей и границ

- ▷ На первом шаге имеем $t_1 = Q$, $\vec{x}^0$ — произвольный элемент множества Q. Пусть к очередному шагу имеется разбиение $t_1, t_2, \ldots, t_L$ и рекорд $\vec{x}^0$.

- ▷ Шаг начинается с проверки элементов разбиения (не обязательно всех) на предмет выяснения:

  1. содержит ли оно решение лучше рекорда;
  2. какое решение в подмножестве является наилучшим.

  Предположим, что проверяется множество t_ℓ; это множество считается проверенным и отбрасывается, если выполняется одно из двух условий:

  1. $H(t_\ell) \geqslant f(\vec{x}^0)$,
  2. функция $\vec{x}(d)$ определена на множестве t_ℓ. При этом, если реализуется второй случай и
  $$f(\vec{x}(t_\ell)) < f(\vec{x}^0),$$
  то устанавливается новое значение рекорда $\vec{x}^0 = \vec{x}(t_\ell)$.

  Если отброшенными оказываются все элементы разбиения, то алгоритм заканчивает работу и $\vec{x}^0$ — решение, найденное в результате его работы.

- ▷ Пусть $t'_1, t'_2, \ldots, t'_{L'}$, $0 < L' \leqslant L$ — множества, не отброшенные в результате проверок. Выберем среди них некоторое „перспективное“ подмножество t_{ℓ_0}. Применим к нему функцию ветвления $b(d)$, в результате чего получим его разбиение $d_1, d_2, \ldots, d_N$ и в целом новое разбиение

$$t'_1, t'_2, \ldots, t'_{\ell_0-1}, d_1, d_2, \ldots, d_N, t'_{\ell_0+1}, \ldots, t'_{L'}$$

множества неотброшенных решений.

- ▷ После этого начинается следующий шаг.

5.5 Конечность метода ветвей и границ

- ▷ Конечность алгоритма — это следствие следующих фактов:

1. Конечность числа атомарных множеств.
2. На каждом шаге алгоритма хотя бы один элемент разбиения либо отбрасывается, либо разбивается на подмножества, каждое из которых состоит из меньшего числа атомарных множеств.
3. Атомарные множества всегда отбрасываются.

5.6 Оптимальность решения

- ▷ Что можно сказать о полученном в результате работы алгоритма рекорде $\vec{x}^0$? Ответ зависит от природы функции $\vec{x}(d)$. Ограничимся анализом двух случаев и покажем, что полученный в результате работы алгоритма рекорд $\vec{x}^0$ является оптимальным решением.

1. Данная функция вычисляет оптимальное решение на любом из атомарных множеств.

   Предположим, что это не так и пусть $f(\vec{x}^0) > f(\vec{x}^*)$, где $\vec{x}^*$ — оптимальное решение. Тогда на некотором шаге алгоритма элемент $\vec{x}^*$ был отброшен вместе с некоторым множеством t. Но множество t не может быть отброшено по первому условию, поскольку в этом случае будут выполняться противоречащие исходному предположению неравенства
$$f(\vec{x}^0) \leqslant f(\vec{x}^0_\ell) \leqslant H(t) \leqslant f(\vec{x}^*),$$
где $\vec{x}^0_\ell$ — рекорд, действующий на момент проверки множества t. Значит на множестве t определена функция $\vec{x}(d)$ и это множество отброшено по второму условию. Но тогда рекорд x^0_1 должен быть заменен на решение $\vec{x}(t)$ такое, что $f(\vec{x}(t)) = f(\vec{x}^*)$.

   Отсюда получаем что $f(\vec{x}^0) \leqslant f(\vec{x}^*)$. Это противоречит исходному предположению и доказывает оптимальность решения $\vec{x}^0$.

2. Функция находит на любом из атомарных множеств лишь некоторое приближённое решение.

   Будем считать, что для всякого атомарного множества d выполняется неравенство
$$f(\vec{x}(d)) \leqslant (1+\varepsilon) f(\vec{x}^*(d)),$$
где $\vec{x}^*(d)$ — оптимальное решение задачи на множестве d. Покажем, что выполняется неравенство
$$f(\vec{x}^0) \leqslant (1+\varepsilon) f(\vec{x}^*),$$
где $\vec{x}^*$ — оптимальное решение задачи. Предположим противное и пусть
$$f(\vec{x}^0) > (1+\varepsilon) f(\vec{x}^*),$$
тогда $f(\vec{x}^0) > f(\vec{x}^*)$ и, следовательно, решение $\vec{x}^*$ на некотором шаге алгоритма было отброшено вместе с некоторым множеством d.

   Поскольку $f(\vec{x}^0) > f(\vec{x}^*)$, то множество d не могло быть отброшено по первому условию. Если оно отброшено по второму условию, то можем написать
$$f(\vec{x}^0) \leqslant f(\vec{x}(d)) \leqslant (1+\varepsilon) f(\vec{x}^*(d)) = (1+\varepsilon) f(\vec{x}^*),$$
но полученное неравенство противоречит предположению, что доказывает справедливость требуемого.

При разработке алгоритма МВГ, исходя из специфики рассматриваемой задачи, необходимо конкретизировать и определить следующие составные элементы общей схемы:

- ▷ атомарные множества решений;
- ▷ способ задания подмножеств решений;
- ▷ функцию ветвления;
- ▷ способ вычисления нижней границы;
- ▷ функцию выбора наилучшего решения;
- ▷ правило выбора перспективного элемента разбиения.

Рассмотрим задачу, в которой целевая функция удовлетворяет условию Липшица. Ограничимся задачей минимизации на параллелепипеде:

$$f(\vec{x}) \to \inf_{\vec{x}\in Q},$$

где $Q = \{\vec{x} = (x_1,\, x_2,\, \ldots,\, x_n) \big| \forall i = 1,\, 2,\, \ldots,\, n\colon a_i \leqslant x_i \leqslant b_i\}$

Соглашения:

$$a_i,\, b_i \in \mathbb{R},\, a_i < b_i,$$

$$|f(\vec{x}) - f(\vec{y})| \leqslant L\|\vec{x} - \vec{y}\|_\infty \text{ для всех } \vec{x},\, \vec{y} \in Q, \tag{5.1}$$

где $L = \mathrm{const} > 0$ — константа Липшица,

$$\|\vec{x} - \vec{y}\|_\infty = \max_{1 \leqslant i \leqslant n} |x_i - y_i|.$$

Из (5.1) следует, что функция f — непрерывна и, следовательно, достигает минимального значения f^* на параллелепипеде.

Выберем на отрезке $[a_j, b_j]$ (на оси j) следующие точки:

$$x_j^1 = a_j + \frac{h}{2},\, x_j^2 = x_j^1 + h,\, \ldots,\, x_j^{i+1} = x_j^i + h,\, \ldots,$$
$$x_j^{m_j - 1} = x_j^1 + (m_j - 2)h,\; x_j^{m_j} = \min\{x_j^1 + (m_j - 1)h,\, b_j\},$$

где $h = 2\varepsilon/L$ — шаг метода, а m_j — натуральное число:

$$x_j^{m_j - 1} < b_j - \frac{h}{2} \leqslant x_j^1 + (m_j - 1)h$$

На параллелепипеде Q введем сетку

$$Q_h = \{\vec{x}^{i_1\, i_2\, \ldots\, i_n} = (x_1^{i_1},\, x_2^{i_2},\, \ldots,\, x_j^{i_j},\, \ldots,\, x_n^{i_n})\},$$

где j-ая координата $x_j^{i_j}$ точки $\vec{x}^{i_1\, i_2\, \ldots\, i_n}$ принимает одно из следующих значений:

$$x_j^1,\, x_j^2,\, \ldots,\, x_j^{m_j}.$$

Пусть

$$F_h = \min_{Q_h} f(\vec{x}^{i_1\, i_2\, \ldots\, i_n}).$$

Теорема 5.7 (Оценка для функции с условием Липшица)**.**

▷ Для любой функции $f(\vec{x})$, которая удовлетворяет условию Липшица (5.1), справедлива оценка

$$f^* \leqslant F_h \leqslant f^* + \varepsilon. \tag{5.2}$$

▷ **Доказательство**.

- Множество

$$Q_{i_1\, i_2\, \ldots\, i_n} = \left\{\vec{x} \in \mathbb{R}^n \Big| \|\vec{x} - \vec{x}^{i_1 i_2 \ldots i_n}\|_\infty \leqslant \frac{h}{2}\right\}$$

является кубом с центром в точке $\vec{x}^{i_1\, i_2\, \ldots\, i_n}$ и гранями параллельными осям координат, у которого длина ребра равна h.

Множество всех кубов $Q_{i_1\, i_2\, \ldots\, i_n}$ с центрами $\vec{x}^{i_1\, i_2\, \ldots\, i_n} \in Q_h$ покрывают весь параллелепипед Q. Значит для любой точки $\vec{x} \in Q$ найдётся куб $Q_{i_1\, i_2\, \ldots\, i_n}$ содержащий эту точку. Отсюда и условия (5.1), имеем: для любого $\vec{x} \in Q$ выполняется

$$f(\vec{x}) \geqslant f(\vec{x}^{i_1\, i_2\, \ldots\, i_n}) - L\|\vec{x} - \vec{x}^{i_1\, i_2\, \ldots\, i_n}\|_\infty \geqslant F_h - L\frac{h}{2} = F_h - \varepsilon \Rightarrow$$
$$\Rightarrow f^* \leqslant F_h \leqslant f^* + \varepsilon.$$

□

18.11.2005

5.2 Модификация алгоритма

▷ Рассмотрим произвольный гиперкуб $\Gamma[\vec{x}^c,\, h]$ с центром $\vec{x}^c$ и длиной стороны h :

$$\Gamma[\vec{x}^c,\, h] = \left\{\vec{x} \in \mathbb{R}^n \Big| \|\vec{x} - \vec{x}^c\|_\infty \leqslant \frac{h}{2}\right\}.$$

Из доказательства теоремы 5.7 следует, что $\forall \vec{x} \in \Gamma[\vec{x}^c,\, h]$:

$$f(\vec{x}) = f(\vec{x}) - f(\vec{x}^c) + f(\vec{x}^c) \leqslant -L|\vec{x} - \vec{x}^c| + f(\vec{x}^c) \leqslant f(\vec{x}^c) - Lh/2.$$

▷ Естественно, тогда определить нижнюю границу на гиперкубе $\Gamma[\vec{x}^c,\, h]$ равенством $H(\Gamma[\vec{x}^c,\, h]) = f(\vec{x}^c) - Lh/2$.

▷ Функцию выбора наилучшего решения определим на гиперкубах со стороной h (играют роль атомарных множеств решений) не превосходящей $2\varepsilon/L$, положив

$$\vec{x}\big(\Gamma[\vec{x}^c,\, h]\big) = \vec{x}^c.$$

▷ Подмножества решений будем задавать в виде набора гиперкубов.

▷ Получаем алгоритм:

- На первом шаге имеем $t_1 = \Gamma[\vec{x}^R,\, \Delta]$, где первый рекорд $\vec{x}^R$ — центр гипрекуба со стороной $\Delta^R = \Delta$.
- Пусть к очередному шагу имеется разбиение

$$t_1 = \Gamma[\vec{x}^1,\, h_1],\, t_2 = \Gamma[\vec{x}^2,\, h_2],\, \ldots,\, t_L = \Gamma[\vec{x}^L,\, h_L]$$

и рекорд $\vec{x}^R$ — центр гиперкуба со стороной Δ^R.

- Очередной шаг начинается с проверки гиперкуба с номером ℓ. Он считается проверенным и отбрасывается, если выполняется одно из следующих условий:
  1. $H(\Gamma[\vec{x}^\ell,\, h_\ell]) \leqslant f(\vec{x}^R)$;
  2. сторона гиперкуба h_ℓ не превосходит величины $2\varepsilon/L$.

  При этом, если реализуется второй случай и
  $$f(\vec{x}^\ell) < f(\vec{x}^R),$$
  то устанавливается новое значение рекорда $\vec{x}^R = \vec{x}^\ell$ и величины $\Delta^R = h_\ell$.

5.3 Дополнительное правило

В процессе выполнения возможны два случая:

I. Текущий рекорд хуже.

Если $f(\vec{x}^\ell) < f(\vec{x}^R)$, то пересчитываем рекорд и среди оставшихся гиперкубов разбиения отбрасываем те, которые содержатся в гиперкубе
$$\Gamma\left[\vec{x}^R,\, \frac{2\big(f(\vec{x}^R) - f(\vec{x}^\ell)\big)}{L}\right].$$

По определению нижней границы имеем:
$$\forall \vec{x} \in \Gamma[\vec{x}^c,\, h]\colon f(\vec{x}) \geqslant H\big(\Gamma[\vec{x}^c,\, h]\big) = f(\vec{x}^c) - Lh/2,$$
поэтому для любой точки $\vec{x}$ из данного гиперкуба имеем
$$f(\vec{x}) \geqslant H\left(\Gamma\left[\vec{x}^R,\, \frac{2\big(f(\vec{x}^R) - f(\vec{x}^\ell)\big)}{L}\right]\right) =$$
$$= f(\vec{x}^R) - \frac{L}{2}\cdot\frac{2\big(f(\vec{x}^R) - f(\vec{x}^\ell)\big)}{L} = f(\vec{x}^\ell).$$

II. Текущий рекорд лучше.

Если $f(\vec{x}^R) \leqslant f(\vec{x}^\ell)$, то среди оставшихся гиперкубов разбиения отбрасываем те, которые содержаться в гиперкубе
$$\Gamma\left[\vec{x}^\ell,\, \frac{f(\vec{x}^\ell) - f(\vec{x}^R)}{L}\right],$$
так как для любой точки $\vec{x}$ из данного гиперкуба имеем $f(\vec{x}) \geqslant f(\vec{x}^R)$.

Если отброшены все элементы разбиения, то алгоритм заканчивает работу и $\vec{x}^R$ — требуемое решение.

Если есть неотброшеные множества, тогда выбираем „перспективное“ подмножество $\Gamma[\vec{x}^\ell,\, h_\ell]$. Функция ветвления разбивает его на 2^n одинаковых подкубов со стороной $h_\ell/2$.

После этого начинается следующий шаг.

5.4 Конечность метода ветвей и границ

▷ Конечность алгоритма следствие следующих фактов:

1. Гиперкуб $\Gamma[\vec{x}^R,\, \Delta]$ является компактным множеством.
2. На каждом шаге алгоритма хотя бы один элемент разбиения либо отбрасывается, либо разбивается на подмножества, каждое из которых состоит из меньшего числа атомарных множеств.
3. Атомарные множества всегда отбрасываются.

Замечание 5.5 (Случай не смены рекорда)**.**

▷ Если в процессе работы алгоритма не происходит смены рекорда по правилу 2, то полученный рекорд — оптимальное решение задачи. В противном случае ε — приближенное решение.

Замечание 5.6 (Правила ветвления)**.**

▷ Наиболее распространены следующие два правила ветвления:

I. „*в ширину*“, когда ветвятся все или по очереди вершины одного уровня и затем переходят к следующему уровню;

II. „*в глубину*“, когда ветвится лишь одна вершина уровня (обычно с лучшим значением рекорда, а если его нет, то нижней границы) до конца ветки.

▷ На практике используются смешанные варианты.

Глава 6

Численные методы нелинейного программирования

ОПР 6.1 (Численных методов).

Задача поиска безусловного минимума:

$$f(\vec{x}) \to \min_{\vec{x} \in \mathbb{R}^n}.$$

Для решения используются численные методы*, в которых текущее приближение вычисляется по формуле*

$$\vec{x}^{k+1} = \vec{x}^k + \vec{\alpha}_k \vec{p}^k,$$

где $\vec{p}^k$ — направление спуска*,* $\vec{\alpha}_k$ — длина шага *вдоль этого направления.*

ОПР 6.2 (Релаксационных методов).

Методы, в которых последовательность векторов $\vec{x}^0, \vec{x}^1, \ldots, \vec{x}^k, \ldots,$ *удовлетворяет условию*

$$f(\vec{x}^0) \geqslant f(\vec{x}^1) \geqslant \ldots \geqslant f(\vec{x}^k) \geqslant \ldots$$

называются релаксационными.

ОПР 6.3 (Скорости сходимости).

Пусть $\vec{x}^*$ *— точка минимума функции* $f(\vec{x})$*, бывают следующие* скорости сходимости:

▷ линейна*, если для* $k = 0, 1, \ldots,$ *выполнено:*

$$\|\vec{x}^{k+1} - \vec{x}^*\| \leqslant q\|\vec{x}^k - \vec{x}^*\|, \; 0 < q < 1,$$

Так же говорят, что метод сходится со скоростью геометрической прогрессии*, так как*

$$\|\vec{x}^k - \vec{x}^*\| \leqslant q^k\|\vec{x}^0 - \vec{x}^*\|.$$

▷ сверхлинейна*, если*

$$\|\vec{x}^{k+1} - \vec{x}^*\| \leqslant q_k\|\vec{x}^k - \vec{x}^*\|, \text{ где } q_k \to 0 \text{ при } k \to \infty.$$

▷ квадратична*, если*

$$\|\vec{x}^{k+1} - \vec{x}^*\| \leqslant C\|\vec{x}^k - \vec{x}^*\|^2, \; C \geqslant 0.$$

ОПР 6.4 (Порядка метода).

Методы нулевого порядка *— это методы, использующие только значения самой целевой функции.*

Методы первого порядка *— это методы, использующие помимо значений целевой функции и её производные.*

И так далее.

ОПР 6.5 (Градиентных методов).

Все итерационные процессы, в которых направление движения на каждом шаге совпадает с антиградиентом (градиентом) функции, называются градиентными методами*:*

$$\vec{x}^{k+1} = \vec{x}^k - \vec{\alpha}_k \vec{f}'(x_k), \; \vec{\alpha}_k \geqslant \vec{0}.$$

6.6 Классификация методов по длине шага

▷ Методы отличаются способами выбора длины шага $\vec{\alpha}_k$:

- Метод с *постоянным шагом*: $\vec{\alpha}_k = \vec{\alpha}$.
- Метод с *дроблением шага*: на каждом шаге проверяется неравенство

$$f\big(\vec{x}^k - \vec{\alpha}_k \vec{f}'(\vec{x}^k)\big) - f(\vec{x}^k) \leqslant -\varepsilon\vec{\alpha}_k\|\vec{f}'(\vec{x}^k)\|^2,$$

  где $0 < \varepsilon < 1$.
- Метод *наискорейшего спуска*: при выборе $\vec{\alpha}_k$ минимизируется по $\vec{\alpha}$ функция $f(\vec{x}^k - \vec{\alpha}\vec{f}'(\vec{x}^k))$:

$$\vec{\alpha}_k = \arg\min_{\vec{\alpha} \geqslant 0} f\big(\vec{x}^k - \vec{\alpha}\vec{f}'(\vec{x}^k)\big).$$

Теорема 6.7 (Достаточные условия сходимости).

▷ Пусть

- функция $f \in C(\mathbb{R}^n)$, ограничена снизу, то есть $f(\vec{x}) \geqslant \vec{f}^* > -\infty$;
- выполняется условие Липшица для градиента $\vec{f}'(\vec{x})$:

$$\|\vec{f}'(\vec{x}) - \vec{f}'(\vec{y})\| \leqslant L\|\vec{x} - \vec{y}\|$$

- длина шага α удовлетворяет условию $0 < \alpha < 2/L$.

▷ Тогда

$\vec{f}'(\vec{x}^k) \to 0$ при $k \to \infty$ и $f(\vec{x}^{k+1}) \leqslant f(\vec{x}^k)$ при любом выборе начального приближения x_0.

▷ **Доказательство**.

- Воспользуемся формулой конечных приращений

$$f(\vec{x} + \vec{y}) = f(\vec{x}) + \int_0^1 (\vec{f}'(\vec{x} + \tau\vec{y}), \vec{y})d\tau.$$

Прибавим и вычтем из правой части величину $(\vec{f}'(\vec{x}), \vec{y})$, получим:

$$f(\vec{x} + \vec{y}) = f(\vec{x}) + (\vec{f}'(\vec{x}), \vec{y}) + \int_0^1 (\vec{f}'(\vec{x} + \tau\vec{y}) - \vec{f}'(\vec{x}), \vec{y})d\tau.$$

Подставим вместо $\vec{x}$ и $\vec{y}$, соответственно, $\vec{x}^k$ и $-\vec{\alpha}\vec{f}'(\vec{x}^k)$, получим:

$$f(\vec{x}^{k+1}) \leqslant_{\left(\substack{\text{вводим модуль}\\ \text{под интеграл}}\right)} f(\vec{x}^k) + (\vec{f}'(\vec{x}^k), -\alpha\vec{f}'(\vec{x}^k)) + \\ + \int_0^1 \left| \Big(\vec{f}'\big(\vec{x}^k - \tau\alpha\vec{f}'(\vec{x}^k)\big) - \vec{f}'(\vec{x}^k), -\alpha\vec{f}'(\vec{x}^k)\Big) \right| d\tau \leqslant$$

$$\leqslant_{(\text{ неравенство Коши-Буняковского}^{\text{I)}})} f(\vec{x}^k) - \alpha\|\vec{f}'(\vec{x}^k)\|^2 + \\ + \int_0^1 \|\vec{f}'(\vec{x}^k - \tau\alpha\vec{f}'(\vec{x}^k)) - \vec{f}'(\vec{x}^k)\| \cdot \|\alpha\vec{f}'(\vec{x}^k)\|d\tau \leqslant$$

$$\leqslant_{(\text{ условие Липшица })} f(\vec{x}^k) - \alpha\|\vec{f}'(\vec{x}^k)\|^2 + \\ + \int_0^1 L\|\tau\alpha\vec{f}'(\vec{x}^k)\| \cdot \|\alpha\vec{f}'(\vec{x}^k)\|d\tau =$$

$$= f(\vec{x}^k) - \alpha\|\vec{f}'(\vec{x}^k)\|^2 + L\alpha^2\|\vec{f}'(\vec{x}^k)\|^2 \int_0^1 \tau d\tau = \\ = f(\vec{x}^k) - \alpha(1 - L\alpha/2)\|\vec{f}'(\vec{x}^k)\|^2 = f(\vec{x}^k) - \alpha\|\vec{f}'(\vec{x}^k)\|^2.$$

Но у нас $\alpha > 0$, следовательно

$$f(\vec{x}^{k+1}) \leqslant f(\vec{x}^k).$$

- Индукцией легко доказать

$$\forall s\colon f(\vec{x}^{s+1}) \leqslant f(\vec{x}^0) - \alpha \sum_{k=0}^{s} \|\vec{f}'(\vec{x}^k)\|^2,$$

учитывая ограниченность функции $f(\vec{x}) \geqslant f^* > -\infty$ получим

$$\sum_{k=0}^{s} \|\vec{f}'(\vec{x}^k)\|^2 \leqslant \big(f(\vec{x}^0) - f(\vec{x}^{s+1})\big)/\alpha \leqslant \big(f(\vec{x}^0) - f^*\big)/\alpha.$$

А это возможно тогда и только тогда, когда $\vec{f}'(\vec{x}^k) \to 0$ при $k \to \infty$.

□

25.11.2005

6.8 Градиентные методы

ОПР 6.8.1 (Сильно выпуклой функции).
Дифференцируемая функция f называется сильно выпуклой *с константой $\ell > 0$, если для любых $\vec{x}$ и $\vec{y}$ из $\mathbb{R}^n$ справедливо*

$$f(\vec{x} + \vec{y}) \geqslant f(\vec{x}) + (\vec{f}'(\vec{x}), \vec{y}) + \ell\|\vec{y}\|^2/2. \tag{6.1}$$

Лемма 6.8.2 (Существование глобального минимума для сильно выпуклой функции).

I) $|(\vec{a}, \vec{b})| \leqslant \|\vec{a}\| \cdot \|\vec{b}\|$.

▷ Пусть

Функция f является сильно выпуклой с константой $\ell > 0$.

▷ Тогда

Она имеет глобальный минимум на $\mathbb{R}^n$.

▷ **Доказательство**.

∘ Воспользуемся неравенством Коши-Буняковского и перепишем условие (6.1):

$$f(\vec{x}+\vec{y}) \geqslant f(\vec{x}) - \|\vec{f}'(\vec{x})\| \cdot \|\vec{y}\| + \ell\|\vec{y}\|^2/2,$$

вынесем величину $\ell\|\vec{y}\|/2$ за скобки:

$$f(\vec{x}+\vec{y}) \geqslant f(\vec{x}) + \frac{\ell\|\vec{y}\|}{2} \cdot \big(\|\vec{y}\| - 2\|\vec{f}'(\vec{x})\|/\ell\big).$$

∘ Следовательно, $\forall \vec{y}\ \|\vec{y}\| > r = 2\|\vec{f}'(\vec{x})\|/\ell$ имеем неравенство

$$f(\vec{x}+\vec{y}) > f(\vec{x}),$$

из которого следует, что минимум существует и достигается на шаре $B(\vec{x}, r)$.

□

Лемма 6.8.3 (Неравенство для сильно выпуклой функции).

▷ Пусть

Функция f является сильно выпуклой с константой $\ell > 0$ и $\vec{x}^*$ — ее глобальный минимум.

▷ Тогда

Для любого $\vec{x} \in \mathbb{R}^n$ выполняется неравенство:

$$\|\vec{f}'(\vec{x})\|^2 \geqslant 2\ell \cdot \big(f(\vec{x}) - f(\vec{x}^*)\big).$$

▷ **Доказательство**.

∘ Подставим в неравенство (6.1) вместо $\vec{y}$ вектор $\vec{x}^* - \vec{x}$, получим

$$f(\vec{x}) - f(\vec{x}^*) + (\vec{f}'(\vec{x}),\, \vec{x}^* - \vec{x}) + \ell\|\vec{x}^* - \vec{x}\|^2/2 \leqslant 0.$$

∘ По определению скалярного произведения имеем

$$\Big(\vec{f}'(\vec{x})/\sqrt{2\ell} + \sqrt{\ell/2}(\vec{x}^* - \vec{x}),\, \vec{f}'(\vec{x})/\sqrt{2\ell} + \sqrt{\ell/2}(\vec{x}^* - \vec{x})\Big) =$$
$$= \big\|\vec{f}'(\vec{x})/\sqrt{2\ell} + \sqrt{\ell/2}(\vec{x}^* - \vec{x})\big\|^2 \geqslant 0,$$

поэтому раскроем это скалярное произведение и получим

$$\|\vec{f}'(\vec{x})\|^2/2\ell + (\vec{f}'(\vec{x}),\, \vec{x}^* - \vec{x}) + \|\vec{x}^* - \vec{x}\|^2/2 \geqslant 0 \geqslant$$
$$\geqslant f(\vec{x}) - f(\vec{x}^*) + (\vec{f}'(\vec{x}),\, \vec{x}^* - \vec{x}) + \ell\|\vec{x}^* - \vec{x}\|^2/2.$$

∘ После приведения подобных членов получим требуемое неравенство:

$$\|\vec{f}'(\vec{x})\|^2 \geqslant 2\ell\big(f(\vec{x}) - f(\vec{x}^*)\big).$$

□

Теорема 6.8.4 (Условия линейной сходимости).

▷ Пусть

Функция f дифференцируема в $\mathbb{R}^n$, является сильно выпуклой, выполняется условие Липшица для градиента $\vec{f}'(\vec{x})$:

$$\|\vec{f}'(\vec{x}) - \vec{f}'(\vec{y})\| \leqslant L\|\vec{x} - \vec{y}\|$$

и длина шага α удовлетворяет условию $0 < \alpha < 2/L$.

▷ Тогда

$\vec{x}^k \to \vec{x}^*$ при $k \to \infty$ и $\|\vec{x}^k - \vec{x}^*\| \leqslant Cq^k$, $0 \leqslant q < 1$.

▷ **Доказательство**.

∘ У функции f существует глобальный минимум $\vec{x}^*$ (лемма 6.8.2 на стр. 84). При доказательстве теоремы 6.7 на стр. 82 получили:

$$f(\vec{x}^{k+1}) \leqslant f(\vec{x}^k) - \alpha(1 - L\alpha/2)\|\vec{f}'(\vec{x}^k)\|^2$$

отсюда, учитывая неравенство из леммы 6.8.3

$$\|\vec{f}'(\vec{x}^k)\|^2 \geqslant 2\ell\big(f(\vec{x}^k) - f(\vec{x}^*)\big)$$

получим

$$f(\vec{x}^{k+1}) \leqslant f(\vec{x}^k) - \ell\alpha(2 - L\alpha)\big(f(\vec{x}^k) - f(\vec{x}^*)\big)$$

или, вычитая $f(\vec{x}^*)$ из обеих частей, имеем:

$$f(\vec{x}^{k+1}) - f(\vec{x}^*) \leqslant \big(1 - \ell\alpha(2 - L\alpha)\big) \cdot \big(f(\vec{x}^k) - f(\vec{x}^*)\big). \tag{6.2}$$

Положим $q_1 = 1 - \ell\alpha(2 - L\alpha)$, тогда из (6.2) получим

$$f(\vec{x}^{k+1}) - f(\vec{x}^*) \leqslant q_1^{k+1}\big(f(\vec{x}^0) - f(\vec{x}^*)\big). \tag{6.3}$$

- Функция $f \neq \mathrm{const}$[II]), следовательно найдётся начальная точка $\vec{x}^0$ такая, что
$$f(\vec{x}^0) > f(\vec{x}^*),$$
тогда при $k = 0$ из неравенства (6.2) имеем:
$$0 \leqslant f(\vec{x}^1) - f(\vec{x}^*) \leqslant q_1\big(f(\vec{x}^0) - f(\vec{x}^*)\big),$$
следовательно $q_1 \geqslant 0$. Так как $q_1 < 1$, то из (6.3) следует, что
$$f(\vec{x}^k) \to f(\vec{x}^*).$$

- Подставим в неравенство
$$f(\vec{x} + \vec{y}) \geqslant f(\vec{x}) + (\vec{f}'(\vec{x}), \vec{y}) + \ell\|\vec{y}\|^2/2$$
вместо $\vec{y}$ вектор $(\vec{x}^k - \vec{x}^*)$ и $\vec{x}^*$ вместо $\vec{x}$. Учитывая, что $\vec{x}^*$ — глобальный минимум, и, следовательно, $\vec{f}'(\vec{x}^*) = 0$, получим
$$\big(f(\vec{x}^k) - f(\vec{x}^*)\big) \geqslant \ell\|\vec{x}^k - \vec{x}^*\|^2/2.$$

Из этого неравенства и неравенства (6.3) получаем:
$$\|\vec{x}^k - \vec{x}^*\|^2 \leqslant 2q_1^k\big(f(\vec{x}^0) - f(\vec{x}^*)\big)/\ell.$$

- Перепишем это неравенство в следующем виде: $\|\vec{x}^k - \vec{x}^*\| \leqslant Cq^k$, где
$$C = \sqrt{2\big(f(\vec{x}^0) - f(\vec{x}^*)\big)/\ell},\ q = \sqrt{q_1},$$
таким образом метод имеет линейную оценку скорости сходимости. Из этого неравенства также получим, что $\vec{x}^k \to \vec{x}^*$ при $k \to \infty$.

□

6.9 Метод Ньютона

6.9.1 Идея метода Ньютона

▷ Пусть f — дважды непрерывно дифференцируемая функция и есть алгоритм вычисления её вторых производных.

[II]) От противного: пусть $f = \mathrm{const}$, тогда $\vec{f}'(\vec{x}) = 0$ и из неравенства $f(\vec{x}+\vec{y}) \geqslant f(\vec{x}) + (\vec{f}'(\vec{x}), \vec{y}) + \ell\|\vec{y}\|^2/2$ имеем $\ell\|\vec{y}\|^2/2 \leqslant 0$. Противоречие с условием $\ell > 0$.

▷ *Идея метода*: заменить функцию f в окрестности текущего приближения $\vec{x}^k$ её квадратичной аппроксимацией:
$$q(\vec{x}) = f(\vec{x}^k) + \vec{f}'(\vec{x}^k)(\vec{x} - \vec{x}^k) + \frac{1}{2}(\vec{x} - \vec{x}^k)^\top f''(\vec{x}^k)(\vec{x} - \vec{x}^k)$$
и выбираем в качестве нового приближения $\vec{x}^{k+1}$ точку минимума функции $q(\vec{x})$ (если она существует).

Предположим, что матрица $\mathtt{f}''(\vec{x}^k)$ — положительно определённая, значит функция $q(\vec{x})$ — сильно выпукла, следовательно у неё единственный минимум и поэтому его можно найти как решение системы уравнений $q'(\vec{x}^{k+1}) = 0$, которая по определению $q(\vec{x})$ эквивалентна следующей системе линейных уравнений
$$\vec{f}'(\vec{x}^k) = -\mathtt{f}''(\vec{x}^k)(\vec{x}^{k+1} - \vec{x}^k).$$
Отсюда и получаем необходимую итерационную формулу
$$\vec{x}^{k+1} = \vec{x}^k - \big(\mathtt{f}''(\vec{x}^k)\big)^{-1}\vec{f}'(\vec{x}^k).$$

Заметим, что здесь как направление, так и длина шага заданы.

Лемма 6.9.2 (Неравенство для сильно выпуклой дважды непрерывно дифференцируемой функции)**.**

▷ Пусть

f — дважды непрерывно дифференцируемая функция.

▷ Тогда

Если f — сильно выпуклая функция с константой ℓ, то выполняется следующее неравенство:
$$\|[\mathtt{f}''(\vec{x})]^{-1}\| \leqslant \ell^{-1}.$$

▷ **Доказательство**.

- Воспользуемся следующей формулой конечных приращений функции f:
$$f(\vec{x} + \vec{y}) - f(\vec{x}) = (\vec{f}'(\vec{x}),\, \vec{y}) + \frac{\big(\mathtt{f}''(\vec{x} + \tau_2\vec{y})\vec{y},\, \vec{y}\big)}{2},$$
где $0 \leqslant \tau_2 \leqslant 1$.

- Из неравенства (6.1) на стр. 84 (определение сильно выпуклой функции) следует что
$$\frac{\mathtt{f}''(\vec{x}+\tau_2\vec{y})\vec{y},\,\vec{y}}{2} = f(\vec{x}+\vec{y}) - f(\vec{x}) - (\vec{f}'(\vec{x}),\,\vec{y}) \geqslant \frac{\ell\|\vec{y}\|^2}{2},$$
заменяя $\vec{y}$ на $t\vec{y}$ получим:
$$\big(\mathtt{f}''(\vec{x}+\tau_2 ty)ty,\,ty\big) \geqslant \ell\|t\vec{y}\|^2,$$
следовательно,
$$t^2\big(\mathtt{f}''(\vec{x}+\tau_2 ty)\vec{y},\,\vec{y}\big) \geqslant t^2\ell\|\vec{y}\|^2.$$
Поделив на t^2 и, устремляя t к нулю, получим
$$\big(\mathtt{f}''(\vec{x})\vec{y},\,\vec{y}\big) \geqslant \ell\|\vec{y}\|^2.$$

- Пусть $\vec{y} = \big(\mathtt{f}''(\vec{x})\big)^{-1}z$, тогда
$$\big(z,\,(\mathtt{f}''(\vec{x}))^{-1}z\big) \geqslant \ell\|(\mathtt{f}''(\vec{x}))^{-1}z\|^2,$$
а используя неравенство Коши-Буняковского получаем:
$$\|z\|\cdot\|(\mathtt{f}''(\vec{x}))^{-1}z\| \geqslant \ell\|(\mathtt{f}''(\vec{x}))^{-1}z\|^2 \Rightarrow \forall z\colon \ell\|(\mathtt{f}''(\vec{x}))^{-1}z\| \leqslant \|z\|.$$
Это означает, что
$$\|[\mathtt{f}''(\vec{x})]^{-1}\| \leqslant \ell^{-1}.$$

□

Пусть последовательность $\{\vec{x}^k\}$ получена с помощью метода Ньютона и точка $\vec{x}^*$ — глобальный минимум функции f.

Теорема 6.9.3 (Условие квадратичной скорости сходимости)**.**

▷ Пусть

Дважды непрерывно дифференцируемая функция f сильно выпукла с константой $\ell > 0$, вторая производная удовлетворяет условию Липшица: для любых $\vec{x},\vec{y} \in R^n$
$$\|\mathtt{f}''(\vec{x}) - \mathtt{f}''(\vec{y})\| \leqslant L\|\vec{x}-\vec{y}\|,$$
и $q = L\|\vec{f}'(\vec{x}^0)\|/2\ell^2 < 1$.

▷ Тогда

$\vec{x}^k \to \vec{x}^*$ при $k \to \infty$ и метод Ньютона имеет квадратичную скорость сходимости
$$\|\vec{x}^k - \vec{x}^*\| \leqslant \frac{2\ell}{L}q^{2^k}.$$

▷ **Доказательство**.

- Воспользуемся следующей формулой конечных приращений:
$$g(\vec{x}+\vec{y}) = g(\vec{x}) + (\vec{g}'(\vec{x}),\,\vec{y}) + \int_0^1 \big(\vec{g}'(\vec{x}+\tau\vec{y}) - \vec{g}'(\vec{x}),\,\vec{y}\big)d\tau,$$
но подставим вместо g производную функции f, получаем
$$\vec{f}'(\vec{x}+\vec{y}) = \vec{f}'(\vec{x}) + (\mathtt{f}''(\vec{x}),\,\vec{y}) + \int_0^1 \big(\mathtt{f}''(\vec{x}+\tau\vec{y}) - \mathtt{f}''(\vec{x}),\,\vec{y}\big)d\tau.$$

- В этой формуле использованы следующие соглашения:
$$(\mathtt{f}''(\vec{x}),\,\vec{y}) = \Big(\big((\vec{f}'(\vec{x}))'_{x_1},\,\vec{y}\big),\,\big((\vec{f}'(\vec{x}))'_{x_2},\,\vec{y}\big),\,\ldots,\,\big((\vec{f}'(\vec{x}))'_{x_n},\,\vec{y}\big)\Big)$$
и
$$\int_0^1 \big(\mathtt{f}''(\vec{x}+\tau\vec{y}) - \mathtt{f}''(\vec{x}),\,\vec{y}\big)d\tau =$$
$$= \Bigg(\int_0^1 \Big(\big(f'(\vec{x}+\tau\vec{y}) - f'(\vec{x})\big)'_{x_1},\,\vec{y}\Big)d\tau,\,\ldots,$$
$$\int_0^1 \Big(\big(f'(\vec{x}+\tau\vec{y}) - f'(\vec{x})\big)'_{x_n},\,\vec{y}\Big)d\tau\Bigg).$$
Перепишем формулу:
$$\vec{f}'(\vec{x}+\vec{y}) - \vec{f}'(\vec{x}) - (\mathtt{f}''(\vec{x}),\,\vec{y}) = \int_0^1 \big(\mathtt{f}''(\vec{x}+\tau\vec{y}) - \mathtt{f}''(\vec{x}),\,\vec{y}\big)d\tau.$$

- Вектор слева обозначим через $\vec{F}$, вектор справа через $\vec{Z}$, вектор под интегралом через $\vec{Z}(\tau) = \big(Z_1(\tau),\,Z_2(\tau),\,\ldots,\,Z_n(\tau)\big)$. Понятно, что $\vec{Z} = \int_0^1 \vec{Z}(\tau)d\tau$ и $\|\vec{F}\| = \|\vec{Z}\|$; оценим норму $\vec{Z}$
$$\|\vec{Z}\|^2 = \sum_j Z_j^2 = \sum_j Z_j \int_0^1 Z_j(\tau)\,d\tau = \int_0^1 \sum_j Z_j Z_j(\tau)\,d\tau \leqslant$$

$$\leqslant \int_0^1 \|\vec{Z}\| \cdot \|\vec{Z}(\tau)\| \, d\tau = \|\vec{Z}\| \int_0^1 \|\vec{Z}(\tau)\| \, d\tau \leqslant$$

$$\leqslant \|\vec{Z}\| \int_0^1 \left\| \mathtt{f}''(\vec{x} + \tau\vec{y}) - \mathtt{f}''(\vec{x}) \right\| d\tau \leqslant L\|\vec{Z}\| \cdot \|\vec{y}\|^2/2,$$

то есть

$$\|\vec{F}\| = \|\vec{Z}\| \leqslant L\|\vec{y}\|^2/2 \Rightarrow \left\| \vec{f'}(\vec{x} + \vec{y}) - \vec{f'}(\vec{x}) - (\mathtt{f}''(\vec{x}), \vec{y}) \right\| \leqslant L\|\vec{y}\|^2/2.$$

- Подставим $\vec{x} = \vec{x}^k$ и $\vec{y} = -\left[\mathtt{f}''(\vec{x}^k)\right]^{-1} \cdot \vec{f'}(\vec{x}^k)$:

$$\|\vec{f'}(\vec{x}^{k+1})\| \leqslant \frac{L}{2} \left\| [f''(\vec{x}^k)]^{-1} \right\|^2 \cdot \left\| \vec{f'}(\vec{x}^k) \right\|^2,$$

тогда применяя лемму 6.9.2 на стр. 88, получим

$$\|\vec{f'}(\vec{x}^{k+1})\| \leqslant \frac{L}{2\ell^2} \|\vec{f'}(\vec{x}^k)\|^2,$$

итерируем это неравенство по k и получаем

$$\|\vec{f'}(\vec{x}^{k+1})\| \leqslant \frac{2\ell^2}{L} \left(\underbrace{\frac{L\|\vec{f'}(\vec{x}^0)\|}{2\ell^2}}_{q} \right)^{2^{k+1}},$$

- Но f — сильно выпуклая функция, следовательно:

$$f(\vec{x} + \vec{y}) \geqslant f(\vec{x}) + (\vec{f'}(\vec{x}), \vec{y}) + \ell\|\vec{y}\|^2/2,$$

заменим $\vec{y}$ на $\vec{y} - \vec{x}$ и получим:

$$f(\vec{y}) \geqslant f(\vec{x}) + (\vec{f'}(\vec{x}), \vec{y} - \vec{x}) + \ell\|\vec{y} - \vec{x}\|^2/2,$$

меняем $\vec{x}$ на $\vec{y}$, а $\vec{y}$ на $\vec{x}$:

$$f(\vec{x}) \geqslant f(\vec{y}) + (\vec{f'}(\vec{y}), \vec{x} - \vec{y}) + \ell\|\vec{x} - \vec{y}\|^2/2.$$

Теперь сложим оба неравенства:

$$(\vec{f'}(\vec{x}) - \vec{f'}(\vec{y}), \vec{x} - \vec{y}) \geqslant \ell\|\vec{x} - \vec{y}\|^2.$$

- Подставим $\vec{y} = \vec{x}^*, \vec{x} = \vec{x}^{k+1}$, и, учитывая равенство $\vec{f'}(\vec{x}^*) = 0$ получим:

$$\ell\|\vec{x}^{k+1} - \vec{x}^*\|^2 \leqslant \left(\vec{f'}(\vec{x}^{k+1}), \vec{x}^* - \vec{x}^{k+1}\right) \leqslant$$
$$\leqslant \left\|\vec{f'}(\vec{x}^{k+1})\right\| \cdot \left\|\vec{x}^* - \vec{x}^{k+1}\right\| \leqslant \frac{2\ell^2}{L} q^{2^{k+1}} \left\|\vec{x}^* - \vec{x}^{k+1}\right\|$$

то есть

$$\left\|\vec{x}^{k+1} - \vec{x}^*\right\| \leqslant \frac{2\ell}{L} q^{2^{k+1}}$$

□

Глава 7

Прямые и двойственные методы решения экстремальных задач

02.12.2005

7.1 Принципы методов решения экстремальных задач

▷ *Прямые методы* решения экстремальных задач с ограничениями имеют дело непосредственно с рассматриваемой задачей (называемой исходной или прямой задачей в противоположность двойственной задаче).

В прямых методах порождается последовательность допустимых решений, на которых значение целевой функции монотонно убывает.

Плюс: на любом шаге имеем допустимое приближенное решение;

Минус: непросто гарантировать глобальную сходимость.

▷ Общий принцип *двойственных методов* заключается в замене исходной задачи на решение последовательности экстремальных задач без ограничений.

Плюс: проще гарантировать глобальную сходимость;

Минус: допустимое решение обычно получается лишь по завершении работы метода.

7.2 Метод штрафных функций

Метод штрафов — пример двойственного метода. Основная идея метода заключается в сведении исходной задачи

$$f(\vec{x}) \to \min_{\vec{x}\in Q} \tag{7.1}$$

$$Q = \{\vec{x} \in R^n \mid \varphi_i(\vec{x}) \leqslant 0,\ i = 1, 2, \ldots, m\} \tag{7.2}$$

к последовательности задач минимизации

$$F_k(\vec{x}) = f(\vec{x}) + P_k(\vec{x}) \to \min_{\vec{x}\in R^n},\ k = 1, 2, \ldots \tag{7.3}$$

где $P_k(\vec{x})$ — штрафная функция множества Q.

ОПР 7.2.1 (Штрафной функции).
Функция $P_k(\vec{x})$ называется штрафной функцией *множества Q, если выполняются условия*

1. $\forall k = 1, 2, \ldots,\ \forall \vec{x} \in \mathbb{R}^n\colon P_k(\vec{x}) \geqslant 0;$

2. *Предел*

$$\lim_{k\to\infty} P_k(\vec{x}) = \begin{cases} 0, & \textit{если } \vec{x} \in Q \\ +\infty, & \textit{если } \vec{x} \notin Q. \end{cases}$$

Пример 7.2.2 (Штрафных функций).

▷ Для любого множества Q можно указать сколь угодно много штрафных функций. Пусть $[a]_+ = \max(0, a)$ и h функция от двух переменных. Например: $h(k, y) = k[y]_+$, $h(k, y) = k[y]_+^2$, $h(k, y) = (1 + [y]_+)^k - 1$, где k — целое, а $y \in R$.

▷ Штрафные функции:

$$k \cdot \sum_{i=1}^{m} [\varphi_i(\vec{x})]_+,\ k \cdot \sum_{i=1}^{m} [\varphi_i(\vec{x})]_+^2,\ \sum_{i=1}^{m} \big(1 + [\varphi_i(\vec{x})]_+\big)^k - 1.$$

Это разнообразие позволяет подобрать наиболее удобный вид минимизируемой функции $F_k(\vec{x})$ и применить более простые методы безусловной оптимизации. Для упрощения дальнейших рассуждений будем считать, что функция штрафа имеет следующий вид: $P_k(\vec{x}) = kH(\vec{x})$, где функция $H(\vec{x})$ может быть, например, функцией вида

$$\sum_{i=1}^{m} [\varphi_i(\vec{x})]_+^2 \text{ или } \sum_{i=1}^{m} [\varphi_i(\vec{x})]_+.$$

Величина k называется *коэффициентом штрафа*.

7.2.3 Описание метода штрафных функций

- ▷ Опишем $(\ell+1)$-ю итерацию метода штрафов. Пусть $\vec{x}(k_\ell)$ — решение задачи без ограничений (7.3) на шаге ℓ.
- ▷ **Шаг** $\ell+1$.

  Если $H(\vec{x}(k_\ell)) \leqslant \varepsilon$, то $\vec{x}(k_\ell)$ — хорошее приближение для оптимального решения, и вычисления заканчиваются. Иначе выбираем коэффициент штрафа $k_{\ell+1} > k_\ell$ (например, $k_{\ell+1} = 10k_\ell$) и решаем задачу (7.3) для нового значения коэффициента штрафа. Получим новое приближение $\vec{x}(k_{\ell+1})$.

Замечание 7.2.3.1 (Возможность решения методом градиентов).

- ▷ Задачу (7.3) можно решить методом градиентов. При этом в качестве начальной точки можно использовать ℓ-ое приближение $\vec{x}(k_\ell)$.

Итак, вместо точного решения $\vec{x}^*$ будем искать приближенное решение с погрешностью, не превосходящей ε. Отметим, что, вообще говоря, приближенное решение $\vec{x}(k_\ell)$ может и не принадлежать Q.

Дальнейшее изложение уже не зависит от того, каким именно методом будет найдена точка $\vec{x}(k_\ell)$, и следовательно, ограничимся предположением о существовании такого метода и перейдем к исследованию сходимости метода штрафных функций.

7.2.4 Метод внешних штрафов

Итак задача (7.3) имеет вид

$$F_k(\vec{x}) = f(\vec{x}) + kH(\vec{x}) \to \min_{\vec{x}\in R^n}, k = 1, 2, \ldots$$

где k — *коэффициент штрафа*, а

$$H(\vec{x}) = \sum_{i=1}^{m} [\varphi_i(\vec{x})]_+^2 \text{ или } H(\vec{x}) = \sum_{i=1}^{m} [\varphi_i(\vec{x})]_+.$$

7.2.4.1 Соглашения

- ▷ Будем считать, что

  I. функция $H\colon \mathbb{R}^n \to \mathbb{R}$ — непрерывна и $\forall \vec{x} \in \mathbb{R}^n\colon H(\vec{x}) \geqslant 0$;

  II. $H(\vec{x}) = 0 \Leftrightarrow \vec{x} \in Q = \big\{\vec{x} \,\big|\, \forall i = 1, 2, \ldots, m\colon \varphi_i(\vec{x}) \leqslant 0\big\}$;

  III. f — непрерывная функция;

  IV. множество Q замкнуто.

Теорема 7.2.4.2 (Условия для сходимости).

- ▷ Пусть

  Выполняется одно из двух условий:

  1. $f(\vec{x}_k) \to +\infty$ для любой такой последовательности $\{\vec{x}_k\} \in Q$, что $\|\vec{x}_k\| \to +\infty$;
  2. Q ограничено и $H(\vec{x}_k) \to +\infty$ для любой такой последовательности $\{\vec{x}_k\}$, что $\|\vec{x}_k\| \to +\infty$.

- ▷ Тогда

  1. последовательность $\vec{x}(k)$ имеет хотя бы одну предельную точку и любая предельная точка этой последовательности есть оптимальное решение задачи;
  2. $H\big(\vec{x}(k)\big) \to 0$

- ▷ **Доказательство**.
  - ○ Покажем, что оптимальное решение задачи существует.
    - ✓ Пусть выполняется 1: если Q ограничено, то всё очевидно, иначе существует последовательность $\{\vec{x}_k\} \in Q$ такая, что $\|\vec{x}_k\| \to$ $\to +\infty \Rightarrow f(\vec{x}_k) \to +\infty$.

      Множества Лебега вида $M(\vec{x}_k) = \big\{\vec{x} \,\big|\, \vec{x} \in Q,\ f(\vec{x}) \leqslant f(\vec{x}_k)\big\}$ — ограничены и замкнуты, значит существует оптимальное решение.
    - ✓ В случае 2 утверждение очевидно.
  - ○ Пусть $\vec{x}^*$ — оптимальное решение задачи,

    $$k_1 < k_2 < \ldots < k_\ell < k_{\ell+1} < \ldots$$

    — коэффициенты штрафа используемые алгоритмом, а $\vec{x}^\ell$ — соответствующие приближения ($\vec{x}^\ell = \vec{x}(k_\ell)$).

    Из $k_{\ell+1} > k_\ell$ следует, что

    $$F_{\ell+1}(\vec{x}^{\ell+1}) = f(\vec{x}^{\ell+1}) + k_{\ell+1}H(\vec{x}^{\ell+1}) > f(\vec{x}^{\ell+1}) + k_\ell H(\vec{x}^{\ell+1}),$$

    а так как $\vec{x}^\ell$ — минимум $F_\ell(\vec{x})$, то

    $$f(\vec{x}^{\ell+1}) + k_\ell H(\vec{x}^{\ell+1}) \geqslant f(\vec{x}^\ell) + k_\ell H(\vec{x}^\ell) \Rightarrow$$
    $$\Rightarrow \forall \ell\colon F_{\ell+1}(\vec{x}^{\ell+1}) > F_\ell(\vec{x}^\ell). \tag{7.4}$$

- По определению $\vec{x}^\ell$ и $\vec{x}^{\ell+1}$ имеем

$$f(\vec{x}^\ell) + k_\ell H(\vec{x}^\ell) \leqslant f(\vec{x}^{\ell+1}) + k_\ell H(\vec{x}^{\ell+1}),$$
$$f(\vec{x}^{\ell+1}) + k_{\ell+1} H(\vec{x}^{\ell+1}) \leqslant f(\vec{x}^\ell) + k_{\ell+1} H(\vec{x}^\ell),$$

суммируем их и приводим подобные:

$$(k_{\ell+1} - k_\ell) H(\vec{x}^{\ell+1}) \leqslant (k_{\ell+1} - k_\ell) H(\vec{x}^\ell).$$

Отсюда и неравенства $k_{\ell+1} > k_\ell$ легко видеть:

$$\forall \ell\colon H(\vec{x}^{\ell+1}) \leqslant H(\vec{x}^\ell), \tag{7.5}$$

но с другой стороны для $\forall \ell$ имеем:

$$f(\vec{x}^\ell) \leqslant f(\vec{x}^\ell) + k_\ell H(\vec{x}^\ell) \leqslant f(\vec{x}^*) + k_\ell H(\vec{x}^*).$$

Отсюда и равенства $H(\vec{x}^*) = 0 (\vec{x}^* \in Q)$ вытекает

$$\forall \ell f(\vec{x}^\ell) \leqslant F_\ell(\vec{x}^\ell) \leqslant f(\vec{x}^*). \tag{7.6}$$

- Покажем, что последовательность $\{\vec{x}^\ell\}$ является ограниченной. Рассуждаем от противного: предположим, что она не ограничена,
  - ✓ Если выполняется условие 1 теоремы, то $f(\vec{x}^\ell) \to +\infty$ при $\ell \to \to +\infty$: противоречие с неравенством (7.6).
  - ✓ Если выполнено условие 2 теоремы, то $H(\vec{x}^\ell) \to +\infty$ при $\ell \to \to +\infty$: противоречие с неравенством $\forall \ell\colon H(\vec{x}^\ell) \leqslant H(\vec{x}^1)$, которое является следствием неравенства (7.5).
- Итак, во всех возможных случаях существует сходящаяся подпоследовательность. Для упрощения будем считать, что ею является последовательность $\{\vec{x}^\ell\}$.

  Пусть $\widehat{\vec{x}}$ ее предел, тогда
  - ✓ Из непрерывности f имеем $\lim_{\ell\to+\infty} f(\vec{x}^\ell) = f(\widehat{\vec{x}})$
  - ✓ Из неравенства (7.6) получаем, что $f(\widehat{\vec{x}}) \leqslant f(\vec{x}^*)$.
  - ✓ Из (7.4) следует, что $F_\ell(\vec{x}^\ell)$ монотонно возрастает.
  - ✓ Из (7.6) видим, что $\forall \ell\colon F_\ell(\vec{x}^\ell) \leqslant f(\vec{x}^*)$.

  Поэтому последовательность $\{F_\ell(\vec{x}^\ell)\}$ имеет предел F^* и

$$\lim_{\ell\to+\infty} F_\ell(\vec{x}^\ell) = F^* \leqslant f(\vec{x}^*).$$

  Отсюда и определения $F_k(\vec{x})$ имеем

$$f(\vec{x}^\ell) + k_\ell H(\vec{x}^\ell) \to F^* \text{ при } \ell \to +\infty \Rightarrow \lim_{\ell\to+\infty} k_\ell H(\vec{x}^\ell) = F^* - f(\widehat{\vec{x}}).$$

- Так как величина $F^* - f(\widehat{\vec{x}})$ конечна, $H(\vec{x}^\ell) \geqslant 0$, $k_\ell \to +\infty$, следовательно, $H(\vec{x}^\ell) \to 0$, при $\ell \to +\infty$. Отсюда и следующих условий $x_\ell \to \widehat{\vec{x}}$ функция H непрерывна. Имеем: $H(\widehat{\vec{x}}) = 0$, следовательно $\widehat{\vec{x}} \in Q$, но тогда $f(\vec{x}^*) \leqslant f(\widehat{\vec{x}})$, а мы ранее показали, что $f(\widehat{\vec{x}}) \leqslant f(\vec{x}^*)$, значит

$$f(\widehat{\vec{x}}) = f(\vec{x}^*) \Rightarrow \widehat{\vec{x}} \text{ — оптимальное решение.}$$

□

7.2.5 Метод внутренних штрафов или метод барьерных функций

Основной минус метода внешних штрафов: промежуточные приближения $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_\ell, \ldots$ (коэффициенты штрафа: $k_1, k_2, \ldots, k_\ell, \ldots$) не являются допустимыми решениями задачи, то есть оптимум аппроксимируется снаружи, следовательно, естественно рассмотреть методы штрафа аппроксимирующие оптимум изнутри.

Как и в предыдущем случае идея метода заключается в сведении исходной задачи (7.1) на стр. 94 к последовательности задач минимизации:

$$F_k(\vec{x}) = f(\vec{x}) + a_k B(\vec{x}) \to \min_{\vec{x}\in R^n}, \; k = 1, 2, \ldots \tag{7.7}$$

где $B(\vec{x})$ — подходящая *функция штрафа* или *барьерная функция*, $a_k > 0$ — *коэффициент штрафа* или *барьерный коэффициент*, $a_k \to 0$ при $k \to \infty$. Пусть ∂Q – граница множества Q.

ОПР 7.2.5.1 (Барьерной функции)**.**

Функция $B(\vec{x})$ называется барьерной функцией *для множества Q, если $B(\vec{x})$ определена, конечна и неотрицательна во всех точках из* $\operatorname{Int} Q$ *и*

$$\lim_{\vec{x}\to\partial Q} B(\vec{x}) = +\infty.$$

Можно считать, что $\forall \vec{x} \in \partial Q\colon B(\vec{x}) = +\infty$.

Пример 7.2.5.2 (Барьерных функций)**.**

$$-\sum_{i=1}^m \varphi_i(\vec{x})^{-1}, \; \sum_{i=1}^m |\varphi_i(\vec{x})|^{-1}, \; \sum_{i=1}^m |\varphi_i(\vec{x})|^{-2}.$$

Замечание 7.2.5.2.1 (Возможность решения методом градиентов).

▷ Задачу (7.7) можно решать методом градиентов:

$$\vec{x}^{k+1} = \vec{x}^k - \alpha_k f'(x_k), \alpha_k \geqslant 0.$$

▷ Если $\vec{x}^k \in \operatorname{Int} Q$, то при достаточно малом $\alpha_k \vec{x}^{k+1} \in \operatorname{Int} Q$, то есть если $\vec{x}^0 \in \operatorname{Int} Q$, то все приближения $\vec{x}^k$ — допустимые решения (7.1) на стр. 94 и (7.2) на стр. 94

7.2.5.3 Описание метода внутренних штрафов

▷ Пусть $\vec{x}^k$ — решение задачи (7.7) на шаге k и $\vec{x}^k \in \operatorname{Int} Q$.

▷ **Итерация** $(k+1)$**.**

Решаем задачу (7.7) для нового значения коэффициента штрафа a_{k+1} с решением $\vec{x}^k$ в качестве начального приближения. Так как $\vec{x}^k \in \operatorname{Int} Q$, то $\vec{x}^{k+1} \in \operatorname{Int} Q$.

Если $a_{k+1}B(\vec{x}^{k+1}) \leqslant \varepsilon$, то СТОП. Иначе переходим к следующей итерации.

7.2.5.4 Соглашения

▷ Будем считать, что

1. Q замкнуто;
2. $\operatorname{Int} Q \neq \varnothing$;
3. любая точка $\vec{x} \in Q$ есть предел последовательности точек из внутренности Q.

Пусть f — непрерывная функция на всем $\mathbb{R}^n$, а барьерная функция $B(\vec{x})$ непрерывна на множестве $\operatorname{Int} Q$.

Теорема 7.2.5.5 (Условия сходимости)**.**

▷ Пусть

Выполняется одно из двух условий:

1. $f(\vec{x}_k) \to +\infty$ для любой такой последовательности $\{\vec{x}_k\} \in Q$, что $\|\vec{x}_k\| \to +\infty$;
2. Q ограничено.

▷ Тогда

1. последовательность $\vec{x}^k$ имеет хотя бы одну предельную точку и любая предельная точка этой последовательности есть оптимальное решение задачи;
2. $a_k B(\vec{x}^k) \to 0$

▷ Доказательство.

∘ По теореме Вейерштрасса в задаче (7.1) на стр. 94 и (7.2) на стр. 94 существует оптимальное решение $\vec{x}^* \in Q$.

Пусть $\vec{x}^1, \vec{x}^2, \ldots, \vec{x}^k, \ldots$ — приближения, которые получены алгоритмом, а $a_1, a_2, \ldots, a_k, \ldots$ — соответствующие барьерные коэффициенты, тогда $\forall k$ имеем

$$f(\vec{x}^*) \leqslant f(\vec{x}^k) \leqslant f(\vec{x}^k) + a_k B(\vec{x}^k) = F_k(\vec{x}^k). \tag{7.8}$$

∘ $\forall \varepsilon > 0$ найдется $\tilde{\vec{x}} \in \operatorname{Int} Q \,|\, f(\tilde{\vec{x}}) \leqslant f(\vec{x}^*) + \varepsilon$, следовательно

$$f(\vec{x}^*)+\varepsilon+a_k B(\tilde{\vec{x}}) \geqslant f(\tilde{\vec{x}})+a_k B(\tilde{\vec{x}}) \geqslant F_k(\vec{x}^k) \Rightarrow \lim_{k\to\infty} F_k(\vec{x}^k) \leqslant f(\vec{x}^*)+\varepsilon.$$

Так как это верно $\forall \varepsilon > 0$, то из (7.8) и монотонного убывания $\{F_k(\vec{x}^k)\}_k$ имеем:

$$\lim_{k\to\infty} F_k(\vec{x}^k) = f(\vec{x}^*),$$
$$\lim_{k\to\infty} f_k(\vec{x}^k) = f(\vec{x}^*),$$
$$\lim_{k\to\infty} a_k B(\vec{x}^k) = 0.$$

∘ Как и в теореме 7.2.4.2 на стр. 96, можно считать, что последовательность $\vec{x}^k$ — ограничена (следует из 1 и 2) и пусть $\vec{x}^k \to \hat{\vec{x}}$. Из непрерывности f следует, что $f(\hat{\vec{x}}) = f(\vec{x}^*)$, а так как Q — замкнуто и $\forall k\colon \vec{x}^k \in Q$, то имеем $\hat{\vec{x}} \in Q$. Следовательно, любая предельная точка последовательности приближений $\vec{x}^k$ является оптимальным решением задачи.

□

УПР 7.2.5.6 (Неравенства для функций разных методов)**.**

▷ Доказать следующие неравенства:

1. $\forall k\colon F_{k+1}(\vec{x}^{k+1}) < F_k(\vec{x}^k)$;
2. $\forall k\colon B(\vec{x}^k) \leqslant B(\vec{x}^{k+1})$;
3. $\forall k\colon f(\vec{x}^{k+1}) \leqslant f(\vec{x}^k)$.

09.12.2005

7.3 Метод Келли

Метод Келли (*метод секущих плоскостей*) является прямым и используется для решения задач выпуклого программирования вида:

$$\min f(\vec{x})$$

при условии, что

$$\varphi_i(\vec{x}) \leqslant 0, i = 1, 2, \ldots, m$$

и функции f, $\varphi_i \colon \mathbb{R}^n \to \mathbb{R}$ — выпуклые и $\in C^1$.

Вводя дополнительные переменную и ограничение, сделаем функционал задачи линейным:

$$\min y,$$
$$f(\vec{x}) \leqslant y,$$
$$\varphi_i(\vec{x}) \leqslant 0,\ i = 1, 2, \ldots, m.$$

Без ограничения общности считаем, что $f(\vec{x}) = (\vec{c}, \vec{x})$ и ограничимся изучением выпуклых задач вида:

$$(\vec{c}, \vec{x}) = \sum_{j=1}^{n} c_j x_j \to \min$$

при условии, что

$$\varphi_i(\vec{x}) \leqslant 0,\ i = 1, 2, \ldots, m.$$

Также считаем, что в задаче существует оптимальное решение $\vec{x}^*$, которое содержится в многогранном множестве $Q_0 = \{\vec{x} \in \mathbb{R}^n | \mathtt{A}\vec{x} = \vec{b}, \vec{x} \geqslant 0\}$, а также $Q \subseteq Q_0$.

7.3.1 Алгоритм метода Келли

▷ Итерация k:

Шаг 1. Решаем задачу ЛП

$$(\vec{c}, \vec{x}) \to \min,\ \vec{x} \in Q^k,$$

где Q^k — текущее многогранное приближение множества Q, причём $Q \subseteq Q^k$.

Пусть $\vec{x}^k$ — оптимальное решение этой задачи. Если $\vec{x}^k$ — допустимое решение исходной задачи, то оно является его оптимальным решением (так как $Q \subseteq Q^k$) и алгоритм заканчивает работу; в противном случае переходим к следующему шагу.

Шаг 2. Найдём номер ограничения i_k, для которого величина $\varphi_{i_k}(\vec{x}^k) >$ > 0 максимальна. Перейдём к выполнению следующей итерации с

$$Q^{k+1} = Q^k \cap \bigl\{\vec{x} \bigm| \varphi_{i_k}(\vec{x}^k) + (\varphi'_{i_k}(\vec{x}^k),\ \vec{x} - \vec{x}^k) \leqslant 0\bigr\}.$$

7.3.2 Корректность определения множества Q^{k+1}

▷ Корректность определения множества Q^{k+1}

1. Так как ограничение $\varphi_{i_k}(\vec{x}^k) + (\vec{\varphi}'_{i_k}(\vec{x}^k),\ \vec{x} - \vec{x}^k) \leqslant 0$ линейно, то множество Q^{k+1} является многогранным.
2. В силу выпуклости множества Q^k и функции φ_{i_k} из леммы 1.6.10 на стр. 23 имеем

$$\forall \vec{x} \in Q^k \colon \varphi_{i_k}(\vec{x}^k) + (\vec{\varphi}'_{i_k}(\vec{x}^k),\ \vec{x} - \vec{x}^k) \leqslant \varphi_{i_k}(\vec{x}),$$

но для $\vec{x} \in Q \colon \varphi_{i_k}(\vec{x}) \leqslant 0$, следовательно $Q \subseteq Q^{k+1}$. Другими словами, ограничение

$$\varphi_{i_k}(\vec{x}^k) + (\vec{\varphi}'_{i_k}(\vec{x}^k),\ \vec{x} - \vec{x}^k) = 0$$

является отсечением, которое отсекает точку $\vec{x}^k$.

Если алгоритм останавливается через конечное число шагов, то текущее приближение — оптимальное решение задачи. Рассмотрим случай, когда последовательность $\{\vec{x}^k\}$ бесконечна.

Теорема 7.3.3 (Совпадение любой предельной точки с оптимальным решением).

▷ Любая предельная точка последовательности $\{\vec{x}^k\}$, порождённая методом секущих плоскостей, есть оптимальное решение задачи.

▷ **Доказательство**.

- Последовательность значений $\{(\vec{c}, \vec{x}^k)\}$ монотонно неубывающая и ограничена сверху (так как существует оптимальное решение). Поэтому без ограничения общности считаем, что последовательность $\{\vec{x}^k\}$ ограничена и, значит, найдётся сходящаяся подпоследовательность. Для упрощения обозначений считаем, что это и есть наша последовательность $\{\vec{x}^k\}_{k \in N}$, пусть $\overline{\vec{x}}$ — её предел.
- Выберем произвольное ограничение с номером i и рассмотрим подпоследовательность элементов $\{\vec{x}^k\}_{k \in T}$, где $T \subseteq N$, для которых секущая плоскость порождалась с помощью i-го ограничения. Возможны два случая:
  1. Найдётся номер k_0 такой, что $\forall k \geqslant k_0 \colon \varphi_i(\vec{x}^k) \leqslant 0$, тогда

$$\varphi_i(\vec{x}^k) \to \varphi_i(\overline{\vec{x}}) \leqslant 0.$$

2. Подпоследовательность $\{\vec{x}^k\}_{k\in T}$ бесконечна, тогда

$$\forall k' > k\colon \varphi_i(\vec{x}^k) + (\vec{\varphi}_i'(\vec{x}^k),\, \vec{x}^{k'} - \vec{x}^k) \leqslant 0,$$

следовательно,

$$\varphi_i(\vec{x}^k) \leqslant \|\vec{\varphi}_i'(\vec{x}^k)\| \cdot \|\vec{x}^{k'} - \vec{x}^k\|.$$

Так как $\|\vec{x}^{k'} - \vec{x}^k\| \to_{(\text{из-за сходимости последовательности})} 0$ и $\varphi_i \in C^1(\mathbb{R}^n)$, то

$$\|\vec{\varphi}_i'(\vec{x}^k)\| \to \|\vec{\varphi}_i'(\overline{\vec{x}})\|.$$

Таким образом

$$\varphi_i(\vec{x}^k) \to \varphi_i(\overline{\vec{x}}) \leqslant 0.$$

Следовательно, в силу произвольного выбора i, $\overline{\vec{x}}$ — допустимое решение задачи, но

$$\forall k\colon (\vec{c},\, \vec{x}^k) \leqslant (\vec{c},\, \vec{x}^*) \Rightarrow (\vec{c},\, \overline{\vec{x}}) \leqslant (\vec{c},\, \vec{x}^*),$$

то есть $\overline{\vec{x}}$ — оптимальное решение задачи.

$\square$

Глава 8

Методы нулевого порядка

8.1 Метод покоординатного спуска

8.1.1 Область применения

- *Область применения*: минимизируемая функция либо не обладает нужной гладкостью, либо является гладкой, но вычисление производных довольно трудоемко.
- Наиболее эффективен, когда функция является гладкой. Метод не требует знания градиента, но сходимость можно гарантировать лишь для гладких функций.

Рассмотрим задачу

$$f(\vec{x}) \to \inf_{\vec{x} \in \mathbb{R}^n} \tag{8.1}$$

8.1.2 Метод покоординатного спуска

- Пусть
  - $\vec{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ — i-ый единичный координатный вектор;
  - $\vec{x}_0$ — начальное приближение;
  - $\alpha_0 > 0$ — начальная длина шага;
  - $\vec{x}_k \in \mathbb{R}^n$ — текущее приближение;
  - $\alpha_k > 0$ — текущая длина шага;
  - вектор $\vec{p}_k$ — текущее направление движения;
  - $\lambda | 0 < \lambda < 1$ — фиксированное число.

- Выбор направления движения

$$\vec{p}_k = \vec{e}_{i_k},\ i_k = k - n[k/n] + 1, \tag{8.2}$$

где $[k/n]$ – целая часть, числа k/n. Условие (8.2) гарантирует циклический перебор векторов $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n$:

$$\vec{p}_0 = \vec{e}_1, \ldots, \vec{p}_{n-1} = \vec{e}_n, \vec{p}_n = \vec{e}_1, \ldots, \vec{p}_{2n-1} = \vec{e}_n, \vec{p}_{2n} = \vec{e}_1, \ldots$$

- Итерация $(k+1)$.

Вычислить $f(\vec{x})$ в точке $\vec{x} = \vec{x}_k + \alpha_k \vec{p}_k$ и если

$$f(\vec{x}_k + \alpha_k \vec{p}_k) < f(\vec{x}_k), \tag{8.3}$$

то положим

$$\vec{x}_{k+1} = \vec{x}_k + \alpha_k \vec{p}_k,\ \alpha_{k+1} = \alpha_k;$$

иначе попробуем вычислить $f(\vec{x})$ в точке $\vec{x} = \vec{x}_k - \alpha_k \vec{p}_k$ и если

$$f(\vec{x}_k - \alpha_k \vec{p}_k) < f(\vec{x}_k), \tag{8.4}$$

то положить

$$\vec{x}_{k+1} = \vec{x}_k - \alpha_k \vec{p}_k,\ \alpha_{k+1} = \alpha_k.$$

Итерацию $(k+1)$-ую назовем *удачной*, если выполняется хотя бы одно из неравенств (8.3) или (8.4). Если $(k+1)$-ая итерация неудачная, то положим: $\vec{x}_{k+1} = \vec{x}_k$:

$$\alpha_{k+1} = \begin{cases} \lambda\alpha_k, & \text{если } i_k = n,\ \vec{x}_k = \vec{x}_{k-n+1}, \\ \alpha_k, & \text{если } i_k \neq n \text{ или } \vec{x}_k \neq \vec{x}_{k-n+1} \text{ или } 0 \leqslant k \leqslant n-1. \end{cases} \tag{8.5}$$

В (8.5) условие $i_k = n$, $\vec{x}_k = \vec{x}_{k-n+1}$ означает, что при последовательном переборе направлений $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n$ (n последних итераций) не оказалось ни одной удачной, следовательно шаг α_k дробится. В этом случае $\forall i = 1, 2, \ldots, n$ выполняются неравенства:

$$f(\vec{x}_k + \alpha_k \vec{e}_i) \geqslant f(\vec{x}_k),\ f(\vec{x}_k - \alpha_k \vec{e}_i) \geqslant f(\vec{x}_k). \tag{8.6}$$

Если в данном цикле из n итераций реализовалась хотя бы одна удачная итерация, то тогда на последней итерации цикла $i_k = n$, но $\vec{x}_k \neq \vec{x}_{k-n+1}$, следовательно длина шага α_k не дробится и сохраняется еще на протяжении n итераций следующего цикла (так как дробление возможно только на последней итерации цикла).

Теорема 8.1.3 (Условие существования предельной точки и её оптимальности).

▷ Пусть

Функция $f(\vec{x})$ выпукла на $\mathbb{R}^n$ и $f \in C^1(\mathbb{R}^n)$, а начальное приближение таково, что множество $M(\vec{x}_0) = \{\vec{x} \in \mathbb{R}^n \big| f(\vec{x}) \leqslant f(\vec{x}_0)\}$ ограничено.

▷ Тогда

Последовательность $\vec{x}_k$ имеет хотя бы одну предельную точку и любая предельная точка этой последовательности есть оптимальное решение задачи.

▷ **Доказательство**.

- По теореме Вейерштрасса в задаче (8.1) на стр. 105 существует оптимальное решение $\vec{x}^* \in \mathbb{R}^n$, а с другой стороны из (8.2)-(8.5) следует:

$$\forall k = 0, 1, \ldots : f(\vec{x}_{k+1}) \leqslant f(\vec{x}_k) \Rightarrow$$
$$\Rightarrow \{\vec{x}_k\} \in M(\vec{x}_0) \text{ и } \exists \lim_{k\to\infty} f(\vec{x}_k) \geqslant f^* = f(\vec{x}^*).$$

- Покажем, что $\lim_{k\to\infty} \alpha_k = 0$: допустим противное: $\forall k \geqslant k_0 \colon \alpha_k = \alpha > 0$ (то есть процесс дробления конечен) и пусть

$$M_\alpha = \left\{\vec{u} \big| \vec{u} = (\vec{x}_{k_0} + \alpha r \vec{e}_i) \in M(\vec{x}_0),\ i = 1, \ldots, n,\ r = 0, \pm 1, \pm 2, \ldots\right\}$$

— сетка с шагом α. Понятно, что начиная с номера k_0 любой цикл из n итераций содержит хотя бы одну удачную итерацию. На каждой удачной итерации переходим от текущей точки сетки $\vec{x}_k$ к соседней $\vec{x}_{k+1}$.

Так как при этом $f(\vec{x}_k) > f(\vec{x}_{k+1})$, то посетим каждую точку сетки не более одного раза, значит сетка M_α содержит бесконечное множество разных точек. Противоречие с ограниченностью множества $M(\vec{x}_0)$, следовательно, процесс дробления длины шага α_k бесконечен, а это означает, что

$$\lim_{k\to\infty} \alpha_k = 0.$$

- Пусть $k_1 < k_2 < \ldots < k_m < \ldots$ — номера итераций, на которых дробится длина шага и выполняются неравенства (8.6). Поскольку $\{\vec{x}_k\} \in M(\vec{x}_0)$ и множество $M(\vec{x}_0)$ ограничено, то без ограничения общности можно считать, что $\exists \lim_{m\to+\infty} \vec{x}_{k_m}$ и равен $\widehat{\vec{x}}$.

Из формулы конечных приращений и неравенств (8.6) имеем:

$$\forall m\colon \big(f'(\vec{x}_{k_m} + \theta_{k_m}\alpha_{k_m}\vec{e}_i),\ \alpha_{k_m}\vec{e}_i\big) = f(\vec{x}_{k_m} + \alpha_{k_m}\vec{e}_i) - f(\vec{x}_{k_m}) \geqslant 0 \Rightarrow$$
$$\Rightarrow f'_{\vec{x}_i}(\vec{x}_{k_m} + \theta_{k_m}\alpha_{k_m}\vec{e}_i) \geqslant 0,$$

где $i = 1, 2, \ldots, n$, $0 \leqslant \theta_{k_m} \leqslant 1$.

- Аналогично получим, что

$$\forall m\colon \big(f'(\vec{x}_{k_m} - \overline{\theta}_{k_m}\alpha_{k_m}\vec{e}_i),\ -\alpha_{k_m}\vec{e}_i\big) = f(\vec{x}_{k_m} - \alpha_{k_m}\vec{e}_i) - f(\vec{x}_{k_m}) \geqslant 0.$$

И, следовательно,

$$\forall m,\ \forall i = 1, 2, \ldots, n\colon f'_{\vec{x}_i}(\vec{x}_{k_m} - \overline{\theta}_{k_m}\alpha_{k_m}\vec{e}_i) \leqslant 0,$$

где $0 \leqslant \overline{\theta}_{k_m} \leqslant 1$.

- По условию частные производные $f'_{\vec{x}_i}(\vec{x})$ — непрерывные функции на $\mathbb{R}^n$, поэтому из условий

$$\lim_{m\to\infty} \alpha_{k_m} = 0,\ \lim_{m\to\infty} \vec{x}_{k_m} = \widehat{\vec{x}},$$
$$\forall m,\ \forall i = 1, 2, \ldots, n\colon f'_{\vec{x}_i}(\vec{x}_{k_m} + \theta_{k_m}\alpha_{k_m}\vec{e}_i) \geqslant 0,$$
$$\forall m,\ \forall i = 1, 2, \ldots, n\colon f'_{\vec{x}_i}(\vec{x}_{k_m} - \overline{\theta}_{k_m}\alpha_{k_m}\vec{e}_i) \leqslant 0,$$

имеем:

$$\forall i = 1, 2, \ldots, n\colon f'_{\vec{x}_i}(\widehat{\vec{x}}) = 0.$$

Это означает, что градиент функции f в точке $\widehat{\vec{x}}$ равен 0, то есть $f'(\widehat{\vec{x}}) = 0$. Для выпуклой функции f это необходимое условие локальной оптимальности является достаточным. Следовательно $\widehat{\vec{x}}$ является оптимальным решением задачи.

□

Литература

[1] Глебов Н.И. и др. "Методы оптимизации НГУ, 2000;

[2] Карманов В. Г. "Математическое программирование М.: Наука;

[3] Ларин Р.М. и др. "МЕТОДЫ ОПТИМИЗАЦИИ. Примеры и задачи НГУ, 2003;

[4] Мину М. "Математическое программирование. Теория и алгоритмы М.: Наука, 1990;

[5] Моисеев Н.Н. и др. "Методы оптимизации М.: Наука, 1978;

[6] Васильев Ф.П. "Методы оптимизации М.: Факториал Пресс, 2002;

[7] Васильев Ф.П. "Численные методы решения экстремальных задач М.: Наука, 1980.