Выпуклый анализ

В.Ю. Протасов *

МГУ, мех-мат, экономический поток.
Весенний семестр 2008-2009 г.

1 Лекция

10 февраля 2009

Выпуклые множества.

Основные свойства выпуклых множеств. Выпуклая комбинация, выпуклая оболочка. Размерность. Теорема Каратеодори. Компактность выпуклой оболочки.

Давно было замечено, что задача поиска минимума функции во многих случаях решается гораздо проще, если известно, что функция выпукла. Выпуклость облегчат решение многих проблем оптимизации. Происходит это из за специфических свойств выпуклых функций. В основном благодаря двум свойствам:

1) локальный минимум выпуклой функции является также ее глобальным минимумом;

2) для выпуклой функции f условие $f'(x) = 0$ является не только необходимым, но и достаточным для того, чтобы в точке x достигался минимум функции.

Везде далее X – линейное нормированное пространство.

Определение 1.1 *Множество $A \subset X$ называется выпуклым, если для любых его точек $x_1, x_2 \in A$ имеем $[x_1, x_2] \subset A$ или, что то же, $(1-t)x_1 + tx_2 \in A$ для всех $t \in [0,1]$.*

Легко показать, что пересечение любого семейства выпуклых множеств – выпукло. Кроме того, замыкание $\bar{A}$ и внутренность $\operatorname{int} A$ выпуклого множества A также выпуклы.

Внутренность. Допустим $x_1, x_2 \in \operatorname{int} A$. Это значит, что x_i лежит в A вместе с некоторым шаром радиуса r_i с центром x_i, $i = 0, 1$. Тогда для любого $t \in [0,1]$ точка $x = (1-t)x_1 + tx_2$ лежит в A вместе с шаром радиуса $(1-t)r_1 + tr_2$.

*Московский Государственный Университет, Механико-математический факультет, Воробьевы Горы, Москва, 119992, e-mail: v-protassov@yandex.ru

Замыкание. Если $x, y \in \bar{A}$, то найдутся последовательности $\{x_k\}, \{y_k\} \subset A$, сходящиеся к x и y соответственно; тогда для любого $t \in [0,1]$ имеем $(1-t)x_k + ty_k \in A$ при всех k, а значит точка $(1-t)x + ty$, будучи пределом последовательности точек $(1-t)x_k + ty_k$ при $k \to \infty$, принадлежит $\bar{A}$.

Для любых точек $x_1, \ldots, x_n \in X$ (среди которых могут быть одинаковые) и для любых чисел $t_1, \ldots, t_n \geq 0$ таких, что $\sum_{i=1}^n t_i = 1$ точка $x = \sum_{i=1}^n t_i x_i$ называется выпуклой комбинацией данных точек.

Предложение 1.1 *Если множество A выпукло, то для любых точек $x_1, \ldots, x_n \in A$ и для любых неотрицательных чисел $t_1, \ldots, t_n$ таких, что $\sum_{i=1}^n t_i = 1$ соответствующая выпуклая комбинация принадлежит A.*

`Доказательство` по индукции относительно числа точек n. При $n = 1$ – очевидно. Пусть $n \geq 2$, и известно, что утверждение верно для любой выпуклой комбинации не более, чем $n-1$ точки. Пусть $x = \sum_{i=1}^n t_i x_i$. Если $t_1 = 1$, то $x = x_1 \in A$. Если же $t_1 \neq 1$, то полагаем $\tilde{t}_k = \frac{t_k}{1-t_1}$ при $k = 2, \ldots, n$. Так как $\sum_{k=2}^n \tilde{t}_k = 1$, то $\tilde{x} = \sum_{k=2}^n \tilde{t}_k x_k$ – выпуклая комбинация, и по индукции $\tilde{x} \in A$. Тогда, в силу выпуклости A имеем $x = t_1 x_1 + (1-t_1)\tilde{x} \in A$.

□

Определение 1.2 *Выпуклой оболочкой* $\mathbf{co}\,(B)$ *произвольного множества $B \subset X$ называется множество всех выпуклых комбинаций его точек:*

$$\mathbf{co}\,(B) \quad = \quad \left\{ x = \sum_{i=1}^n t_i x_i \;\middle|\; x_1, \ldots, x_n \in A\,, \quad t_1, \ldots, t_n \geq 0\,, \quad \sum_{i=1}^n t_i = 1\,, \quad n \in \mathbb{N} \right\}$$

Среди всех выпуклых множеств, содержащих данное множество $B \subset X$, существует наименьшее (по включению), т.е., то, которое содержится в любом выпуклом множестве, содержащем B. В самом деле, семейство всех выпуклых множеств, содержащих B, не пусто (в него входит, например, всё пространство X). Пересечение всех множеств этого семейства будет выпуклым (т.к., пересечение любого семейства выпуклых множеств – выпукло), содержащим B и наименьшим по включению. Оказывается – это множество $\mathbf{co}\,(B)$.

Предложение 1.2 *Множество* $\mathbf{co}\,(B)$ *является наименьшим выпуклым множеством, содержащим B.*

`Доказательство`. В силу предложения 1.1, если выпуклое множество A содержит B, то оно содержит и любую выпуклую комбинацию точек множества B, значит, содержит и $\mathbf{co}\,(B)$. Таким образом, $\mathbf{co}\,(B)$ лежит в пересечении всех выпуклых множеств, содержащих B. Осталось показать что $\mathbf{co}\,(B)$ – выпукло, в этом случае оно совпадёт с пересечением всех выпуклых множеств, содержащих B. Пусть $x = \sum_{i=1}^n t_i\, x_i$ и $y = \sum_{j=1}^m s_j\, y_j$ – две произвольные точки множества $\mathbf{co}\,(B)$, т.е., две выпуклые комбинации точек B. Тогда для любого $t \in [0,1]$ имеем

$$(1-t)x + ty \;=\; (1-t)\sum_{i=1}^n t_i\, x_i \;+\; t\sum_{j=1}^m s_j\, y_j \;=\; \sum_{i=1}^n (1-t)\, t_i\, x_i \;+\; \sum_{j=1}^m t\, s_j\, y_j\,.$$

Заметим, что все числа $(1-t)\, t_i$ и $t\, s_j$ неотрицательны и

$$\sum_{i=1}^n (1-t)\, t_i + \sum_{j=1}^m t\, s_j \;=\; (1-t) \,+\, t \;=\; 1,$$

значит мы получили выпуклую комбинацию точек $x_1, \ldots, x_n, y_1, \ldots, y_m$, которая принадлежит $\mathbf{co}\,(B)$.

$\square$

Следствие 1.1 *Множество A является выпуклым тогда и только тогда когда $A = \mathbf{co}\,(A)$.*

Введём теперь понятие размерности выпуклого множества. Напомним, что подмножество $Y \subset X$ называется аффинным подпространством X, если для любых точек $x_1, x_2 \in Y$ $(x_1 \neq x_2)$ прямая, проходящая через $[x_1, x_2]$, лежит в Y, т.е., $(1-t)x_1 + tx_2 \in Y$ для любого $t \in \mathbb{R}$. Аналогично определяется аффинная комбинация точек $x_1, \ldots, x_n$ как сумма $\sum_{k=1}^{n} t_i x_i$, где $t_1, \ldots, t_n$ – произвольные действительные числа такие, что $\sum_{i=1}^{n} t_i = 1$. *Аффинная оболочка* $\mathbf{aff}\,(B)$ множества B – множество всех аффинных комбинаций точек из B. Так же доказывается аналог предложения 1.2: *аффинная оболочка множества B – есть минимальное по включению аффинное подпространство, содержащее B.* Напомним, что точки называются аффинно независимыми, если никакая из них не является аффинной комбинацией остальных, а размерность аффинного пространства равна размерности соответствующего линейного пространства, и на единицу меньше максимального числа аффинно независимых точек.

Если $\dim X = d$, то выпуклая оболочка любых $d{+}1$ аффинно независимых точек $x_1, \ldots, x_{d+1}$ называется d-мерным симплеском с вершинами $x_1, \ldots, x_{d+1}$. Точка $c = \sum_{i=1}^{d+1} \frac{1}{d+1} x_i$ – центр симплекса (центр тяжести, или точка пересечения медиан).

Лемма 1.1 *Выпуклая комбинация вершин симплекса с положительными коэффициентами является внутренней точкой симплекса.*

`Доказательство.` Пусть $\dim X = d$. Подходящим аффинным преобразованием переведем пространство X в аффинное пространство $V = \left\{ y \in \mathbb{R}^{d+1} \mid \sum_{k=0}^{d+1} y^k = 0 \right\}$, при этом каждая точка x_i перейдёт в конец базисного вектора $e_i \in \mathbb{R}^{d+1}$ (i-тая координата равна единице, остальные – нули). Также можно считать, что норма – евклидова, поскольку все нормы в конечномерном пространстве эквивалентны. Если $x = \sum_{i=1}^{d+1} t_i e_i = (t_1, \ldots, t_{d+1})$, то расстояние от этой точки до ближайшей координатной плоскости равно $r = \min t_i$, следовательно d-мерный шар с центром x радиуса r целиком лежит в симплексе. Таким образом, x – внутренняя точка симплекса.

$\square$

Поскольку выпуклая комбинация является также и аффинной комбинацией, имеем $\mathbf{co}\,(B) \subset \mathbf{aff}\,(B)$ для любого множества B.

Предложение 1.3 *Для любого $B \subset X$ имеем $\mathbf{co}\,(B) \subset \mathbf{aff}\,(B)$, причём в случае $\dim \mathbf{aff}\,(B) < \infty$ множество $\mathbf{co}\,(B)$ имеет непустую внутренность в пространстве $\mathbf{aff}\,(B)$.*

`Доказательство.` Если $\dim \mathbf{aff}\,(B) = d$, то множество B содержит $d+1$ аффинно-независимую точку. Эти точки являются вершинами d-мерного симплекса, который (лемма 1.1) имеет в d-мерном пространстве непустую внутренность, а, с другой стороны, содержится в $\mathbf{co}\,(B)$. Следовательно, $\mathbf{co}\,(B)$ имеет непустую внутренность.

$\square$

По определению *размерностью выпуклого множества* называется размерность его аффинной оболочки. Применив теперь следствие 1.1, получаем

Следствие 1.2 *Конечномерное выпуклое множество имеет непустую внутренность в своей аффинной оболочке.*

Следствие 1.3 *Если выпуклое множество $A \subset \mathbb{R}^d$ имеет пустую внутренность, то оно целиком помещается в аффинном подпространстве меньшей размерности.*

Условие конечномерности в предложении 1.3 и следствиях 1.2, 1.3 существенно, как показывает пример из следующего упражнения:

Упражнение 1 Приведите пример выпуклого подмножества A гильбертова пространства H, которое имеет пустую внутренность, но не лежит ни в каком замкнутом подпространстве пространства H.

`Указание.` Рассмотрим гильбертово пространство $l_2 = \{\, y = (y_1, y_2, \ldots,) \mid \sum_k y_k^2 < \infty \,\}$ и в нём "гильбертов кирпич": $K = \{\, y \in l_2 \mid |y_k| \le \frac{1}{k}\,,\ k \in \mathbb{N} \,\}$. Докажите, что K – выпуклое множество с пустой внутренностью, и $\mathbf{aff}\, K = l_2$.

Теорема 1.1 *(Каратеодори) Если $\dim \mathbf{co}\,(B) = d$, то любой элемент множества $\mathbf{co}\,(B)$ представляется в виде выпуклой комбинации не более, чем $d+1$ элементов множества B.*

`Доказательство.` Не ограничивая общности считаем, что $\mathbf{aff}\, B = X$, $\dim X = d$. Пусть $x = \sum_{i=1}^{n} t_i x_i$, где $x_i \in B$, $i = 1, \ldots, n$ и $n \ge d+2$. Рассмотрим векторы $x_2 - x_1, \ldots, x_n - x_1$. Их не менее $d+1$, и они принадлежат d-мерному линейному пространству, значит они линейно зависимы: $\sum_{i=2}^{n} \alpha_i (x_i - x_1) = 0$ для некоторых НЕРОН множителей $\alpha_2, \ldots, \alpha_n$. Раскрывая скобки и обозначая $\beta_1 = -\sum_{i=2}^{n} \alpha_i$ и $\beta_i = \alpha_i$ при $i = 2, \ldots, n$, получаем:

$$\sum_{i=1}^{n} \beta_i x_i \;=\; 0\,, \qquad \sum_{i=1}^{n} \beta_i \;=\; 0\,,$$

причём множители β_i – НЕРОН. Пусть $s = -\frac{t_j}{\beta_j} = \min\{\, -\frac{t_i}{\beta_i} \;\mid\; \beta_i < 0 \,\}$. Тогда в комбинации $x = \sum_{i=0}^{n} \big(t_i + s\,\beta_i\big)\, x_i$ сумма всех коэффициентов равна 1, они неотрицательны (следовательно, это – выпуклая комбинация) и j-тый коэффициент равен нулю. Мы представили точку x как выпуклую комбинацию из не более, чем $n-1$ точек. Уменьшая далее число точек, приходим к выпуклой комбинации из не более, чем $d+1$ точки.

$\square$

Упражнение 2 Верна ли обратная теорема: если любой элемент множества $\mathbf{co}\,(B)$ представляется в виде выпуклой комбинации не более, чем $d+1$ элементов множества B, то $\dim \mathbf{co}\,(B) = d$?

Перейдём к одном из важнейших следствий теоремы Каратеодори.

Теорема 1.2 *В конечномерном пространстве выпуклая оболочка компакта – компакт.*

`Доказательство.` Надо доказать, что из любой последовательности $\{z_k\} \subset \mathbf{co}\,(B)$ можно выбрать подпоследовательность, сходящуюся к некоторому элементу $x \in \mathbf{co}\,(B)$. Если $\dim X = d$, то согласно теореме 1.1 каждая точка z_k представляется в виде выпуклой комбинации $z_k = \sum_{i=1}^{d+1} t_{ki} z_{ki}$, где $z_{ki} \in B$ (если в выпуклой комбинации участвует меньше $d+1$ точек, то дополняем её произвольными точками с нулевыми коэффициентами). Поскольку B – компакт, выбираем из последовательности $\{z_{k1}\}_{k \in \mathbb{N}}$ подпоследовательность, сходящуюся к некоторой точке $z_1 \in B$, а из последовательности $\{t_{k1}\}_{k\in\mathbb{N}}$ – подпоследовательность, сходящуюся к некоторому числу $t_1 \in [0,1]$. Переходя к этой подпоследовательности, считаем что $z_{k1} \to z_1$ и $t_{k1} \to t_1$ при $k \to \infty$. Аналогично, переходя к подпоследовательностям, получаем, что $z_{ki} \to z_i \in B$ и $t_{ki} \to t_i \in [0,1]$ для любого $i = 2,\ldots,d+1$ при $k \to \infty$. Тогда $\sum_{i=1}^{d+1} t_i = 1$ и $\lim_{k\to\infty} z_k = \sum_{i=1}^{d+1} t_i z_i \in \mathbf{co}\,(B)$, что и требовалось доказать.

$\square$

Упражнение 3 Приведите пример замкнутого множества в $\mathbb{R}^d$, выпуклая оболочка которого не замкнута.

`Указание.` Рассмотрите следующее множество на плоскости $\mathbb{R}^2$:
объединение ветви гиперболы $\left\{ (x,y) \in \mathbb{R}^2 \mid xy = 1,\ x > 0 \right\}$ и начала координат $\{(0,0)\}$. .

Упражнение 4 Приведите пример компакта в бесконечномерном пространстве, выпуклая оболочка которого не замкнута (и, следовательно, не компактна).

2 Лекция

17 февраля 2009

Теоремы отдел$\bigcup$.

Первая и вторая теоремы отделимости. Функционал Минковского, теорема Хана-Банаха.
Крайние точки выпуклых множеств, теорема Крейна-Мильмана.

Определение 2.1 *Подмножества A и B нормированного пространства X называются отделимыми, если существует функционал $x^* \in X\backslash\{0\}$ такой, что $\sup_{a\in A}(x^*, a) \le \inf_{b\in B}(x^*, b)$. Они называются строго отделимыми, если $\sup_{a\in A}(x^*, a) < \inf_{b\in B}(x^*, b)$.*

Теорема 2.1 *(Первая теорема отделимости). Если A и B – непустые выпуклые подмножества нормированного пространства X, причём $\operatorname{int} B \ne \emptyset$, $(\operatorname{int} B) \cap A = \emptyset$, то A и B отделимы.*

Геометрически это означает, что два выпуклые множества при указанных условиях всегда можно разделить гиперплоскостью (коразмерности 1) так, чтобы множества оказались в разных замкнутых полупространствах.

Теорема 2.2 *(Вторая теорема отделимости). Если A и B – непустые выпуклые замкнутые подмножества нормированного пространства X, причём A – компакт и $A \cap B = \emptyset$, то A и B строго отделимы.*

Следствие 2.1 *Любая точка нормированного пространства X, не принадлежащая выпуклому замкнутому множеству $A \subset X$, строго отделима от A.*

Мы начнём с доказательства первой теоремы отделимости, причём, в силу центрального положения этой теоремы во всем курсе, мы приведём два разных доказательства: одно для случая $\dim X < \infty$, и одно для общего случая. Вначале заметим, что достаточно ограничиться случаем, когда B – открытое множество. Тогда условие теоремы будет выглядеть как $A \cap B = \emptyset$. В самом деле, если мы доказали теорему для этого случая, то, положив в исходной формулировке $B' = \operatorname{int} B$, получаем функционал x^*, для которого $\sup_{a\in A}(x^*, a) \le \inf_{b'\in B'}(x^*, b')$. Обозначим $r = \inf_{b'\in B'}(x^*, b')$ и покажем, что $\inf_{b\in B}(x^*, b) = r$. Тогда получится, что функционал x^* разделяет также A и B. Если предположить, например, что $\inf_{b\in B}(x^*, b) < r$, то найдётся $b \in B$, для которого $(x^*, b) = \tilde{r} < r$. Для произвольной точки $z \in B'$ точка $(1-t)b+tz$ также принадлежит B'. Поэтому $\inf_{b'\in B'}(x^*, b') \le (x^*\,,\,(1-t)b + tz) = (1-t)\tilde{r} + t\,(x^*, z)$, что меньше r при малых t.

`Первое доказательство теоремы 2.1 (конечномерный случай).` Пусть $X = \mathbb{R}^d$. Рассмотрим множество $C = A - B = \{x - y \mid x \in A, y \in B\}$ (разность по Минковскому множеств A и B). Ясно, что $0 \notin C$. Нужно доказать существование функционала $x^* \in \mathbb{R}^d$ такого, что $\sup_{z\in C}(x^*, z) \le 0$, т.е., нужно провести гиперплоскость, $\alpha = \{x \in \mathbb{R}^d\,,\ (x^*, x) = 0\}$, отделяющую множество C от начала координат O. Рассмотрим замыкание $\bar{C}$ множества C.

`Случай` $O \notin \bar{C}$. В этом случае существует точка $q \in \bar{C}$ ближайшая к точке O (поскольку функция $f(x) = |x|$ непрерывна и коэрцитивна на $\bar{C}$). Докажем, что плоскость $\alpha = \{x \in \mathbb{R}^d\,,\ (q, x) = 0\}$ – искомая. В самом деле, если $\sup_{z\in C}(-q, x) > 0$, то найдётся точка $z \in C$, для которой $(q, z) < 0$, т.е., векторы z и q образуют тупой угол. Тогда основание h высоты Oh треугольника Ozq лежит на отрезке $[z, q]$ (и, в силу выпуклости $\bar{C}$, принадлежит $\bar{C}$), и $|h| < |q|$, что противоречит выбору точки q.

`Случай` $O \in \bar{C}$. Так как $O \notin C$, то $O \in \partial C$. Следовательно, существует последовательность y_k, сходящаяся к O, состоящая из точек, не принадлежащих $\bar{C}$. Согласно предыдущему пункту, каждую из этих точек можно отделить от C некоторым функционалом x_k^*, для которого $\sup_{z\in C}(x_k^*, z) \le (x_k^*, y_k)$. Не ограничивая общности, считаем, что $|x_k^*| = 1$ для всех k. Из компактности сферы в $\mathbb{R}^d$ следует, что можно перейти к подпоследовательности x_k^*, сходящейся к некоторой точке $x^*\,, |x^*| = 1$. Для этой точки имеем

$$\sup_{z\in C}(x^*, z) \quad \le \quad \lim_{k\to\infty}(x_k^*, y_k) \quad \le \quad \lim_{k\to\infty}|y_k| \quad = \quad 0.$$

Значит, функционал x^* – искомый.

□

Для доказательства теоремы отделимости в общем случае нам понадобятся несколько дополнительных понятий.

Определение 2.2 *Если X – нормированное пространство, $U \subset X$ – выпуклое множество, причём $0 \in \operatorname{int} C$, то функционал $p_U(x) = \inf\big\{\alpha > 0 \bigm| \alpha^{-1}x \in U\big\}$ (при этом полагаем $p_U(0) = 0$) называется функционалом Минковского множества U.*

Функционал $p(x)$ называется *полуаддитивным* если $p(x+y) \le p(x) + p(y)$ для любых $x, y \in X$. Функционал называется *положительно-однородным*, если $p(\lambda x) = \lambda p(x)$ для любых $x \in X\,,\ \lambda \ge 0$. В частности, норма – полуаддитивна и положительно-однородна. Доказательство следующего предложения элементарно и оставляется читателю.

Предложение 2.1 *Функционал Минковского полуаддитивен и положительно однороден. Если множество U ограниченно и центрально-симметрично (т.е., $x \in U \Leftrightarrow -x \in U$), то этот функционал является нормой. Более того, если B – единичный шар в X, то $\|x\| = p_B(x)$.*

Таким образом, норма – это функционал Минковского единичного шара. Для каждого ограниченного центрально-симметричного выпуклого множества U, содержащего ноль в качестве внутренней точки, функционал Минковского p_U является нормой с единичным шаром U.

Упражнение 5 Вычислите функционал Минковского для следующих выпуклых множеств U:
а) $U = \{x \in \mathbb{R}^d \mid -1 \le x^i \le 1 \ , i = 1, \dots, d\}$ – d-мерный куб;
б) $U = \{x \in \mathbb{R}^d \mid \sum_{i=1}^d |x^i| \le 1\}$ – d-мерный октаэдр ("кросс-политоп");
в) $U = \{(x, y) \in \mathbb{R}^2 \mid y \ge x^2 - 1\}$ – внутренность параболы.

Следующий классический результат доказывался в курсе функционального анализа.

Теорема 2.3 *(Хан, Банах) Если в нормированном пространстве X задан полуаддитивный положительно-однородный функционал p, линейное подпространство $L \subset X$, и непрерывный линейный функционал $f : L \to \mathbb{R}$, обладающий свойством $(f, x) \le p(x)$, $x \in L$, то его можно продолжить до непрерывного линейного функционала на всём X с сохранением свойства $(f, x) \le p(x)$, $x \in X$.*

`Второе доказательство теоремы 2.1 (общий случай).` Положим $U = A - B - c$, где c – произвольная точка множества $(A - B)$. Заметим, что поскольку B открыто, то и множество $A - B$ открыто, как объединение открытых множеств $a - B$, $a \in A$. Таким образом, множество U открыто, выпукло, $0 \in U$ и $-c \notin U$. Функционал p_U полуаддитивен и положительно-однороден (предложение 2.1). Определим линейный функционал f на прямой $L = \{tc \mid t \in \mathbb{R}\}$ равенством $(f, tc) = -t$. Покажем, что $(f, x) \le p_U(x)$ при всех $x \in L$. В самом деле, при $t < 0$ имеем $(f, tc) = -t$, тогда как $p_U(tc) = p_U(-t(-c)) = -tp_U(-c) \ge -t$, так как $p_U(-c) \ge 1$, поскольку $-c \notin U$. При $t \ge 0$ имеем $(f, tc) = -t \le 0$, в то время, как $p_U(tc) \ge 0$. По теореме Хана-Банаха продолжаем функционал f на всё пространство X с сохранением свойства $(f, x) \le p_U(x)$. Поскольку для любой точки $x \in U$ имеем $p_U(x) \le 1$ и для любых точек $a \in A, b \in B$ выполнено $x = a - b - c \in U$, получаем $(f, a - b - c) \le p_U(a - b - c) \le 1$, откуда $(f, a) - (f, b) - (f, c) \le 1$. Так как $(f, c) = -1$, то $(f, a) - (f, b) \le 0$ для любых $a \in A, b \in B$. Итак, $\sup_{a \in A}(f, a) \le \inf_{b \in B}(f, b)$, что и требовалось.

$\square$

Упражнение 6 В пространстве X даны аффинная плоскость L и выпуклое множество $B \subset X$, внутренность которого не пуста и не пересекает L. Докажите, что существует гиперплоскость, содержащая L и не пересекающая $\operatorname{int} B$.

Замечание 2.1 В первом доказательстве теоремы 2.1 мы не пользовались условием $\operatorname{int} B \ne \emptyset$. Поэтому в конечномерном пространстве любые два непересекающихся выпуклых множества A и B всегда отделимы. Однако, в бесконечномерном пространстве это условие опустить нельзя. Пример предлагается построить в следующем упражнении.

Упражнение 7 Приведите пример двух непересекающихся выпуклых подмножеств нормированного пространства, которые нельзя отделить.

`Указание.` Рассмотрим гильбертово пространство l_2 и в нём множества $A = \{0\}$ (одна точка) и $B = \{ y \in l_2 \,|\sum_{k=1}^{\infty} y_k = 1\}$. Докажите, что для любого функционала $x^* \in l_2$, $x^* \ne 0$ имеем $\inf_{b \in B}(x^*, b) = -\infty$ и $\sup_{b \in B}(x^*, b) = +\infty$.

`Доказательство теоремы 2.2.` Обозначим через P открытый единичный шар в пространстве X. Покажем, что для достаточно малых r множества $A + rP$ и B не пересекаются. В противном случае найдутся последовательности точек $\{a_k\} \subset A$, $\{b_k\} \subset B$, для которых

$\|a_k - b_k\| \to 0$ при $k \to \infty$. Поскольку A компактно, переходя к подпоследовательности, считаем, что последовательность a_k сходится к некоторой точке $a \in A$. Но тогда и $b_k \to a$, $k \to \infty$, что, в силу замкнутости B влечёт $a \in B$. Таким образом, $A \cap B \neq \emptyset$, что противоречит условию. Итак, открытое выпуклое множество $A + rP$ не пресекает B, поэтому, в силу теоремы 2.1 найдётся функционал $x^* \in X^*$, разделяющий эти множества:

$$\sup_{a \in A, x \in P} (x^*, a + rx) \ \le \ \inf_{b \in B} (x^*, b).$$

Заметим, что $\sup_{a \in A, x \in P}(x^*, a + rx) = \sup_{a \in A}(x^*, a) + \sup_{x \in P}(x^*, rx) = \sup_{a \in A}(x^*, a) + r\|x^*\|$. Итак, $\sup_{a \in A}(x^*, a) \le -r\|x^*\| + \inf_{b \in B}(x^*, b)$. Таким образом, x^* строго разделяет A и B.

Упражнение 8 Приведите пример двух непересекающихся выпуклых замкнутых подмножеств плоскости, которые нельзя строго отделить.

Определение 2.3 *Точка выпуклого множества называется крайней, если она не является серединой никакого отрезка, лежащего в данном множестве.*

Множество крайних точек выпуклого множества A будем обозначать через $\mathbf{extr}\,(A)$. Крайние точки многогранника – суть его вершины. Все граничные точки евклидова шара – крайние. Цилиндр не имеет крайних точек.

Теорема 2.4 *(Крейн, Мильман) Выпуклый компакт в нормированном пространстве является замыканием выпуклой оболочки своих крайних точек.*

Мы докажем эту теорему только в случае $\dim X < \infty$.

`Доказательство.` Пусть $X = \mathbb{R}^d$, $A \subset \mathbb{R}^d$ – выпуклый компакт. Докажем сначала, что он имеет хотя бы одну крайнюю точку. Пусть a – произвольная точка, x – самая дальняя от a точка A. Тогда x – крайняя точка. В противном случае $x = \frac{1}{2}\bigl(x_1 + x_2\bigr)$ для некоторых $x_1, x_2 \in A$, причём $x_1 \neq x_2$. Так как медиана треугольника меньше максимальной из двух сторон, между которыми она заключена, $|x - a|$ меньше $\max\bigl\{\,|x - a_1|\,,\,|x - a_2|\,\bigr\}$, что невозможно.

Итак, множество $A' = \overline{\mathbf{co}\,(\mathbf{extr}\,(A))}$ не пусто, выпукло и компактно (как замкнутое подмножество компакта A). Если $A' \neq A$, то найдётся точка $z \in A$, не принадлежащая A'. Согласно теореме 2.2 существует гиперплоскость H, строго отделяющая z от A'. Проведём гиперплоскость H', параллельную H через точку z и выберем на ней произвольную точку $u \neq z$. Выберем шар достаточно большого радиуса R, касающийся H' в точке u и содержащий множество A'. Пусть a – его центр, x – самая дальняя от a точка A. Тогда x – крайняя точка, которая не принадлежит A' (поскольку находится вне шара, так как $|x - a| \ge |a - z| > R$). Полученное противоречие завершает доказательство.

$\square$

3 Лекция

24 февраля 2009

Выпуклые функции.

Эффективное множество, надграфик, собственные функции. Неравенство Йенсена. Минимумы и максимумы выпуклых функций. Локальная липшицевость выпуклых функций.

Пусть $\bar{\mathbb{R}} = \mathbb{R} \bigcup \{-\infty\} \bigcup \{+\infty\}$. *Эффективным множеством* функции f называется $\mathbf{dom}\, f = \{x \in X \mid f(x) \neq +\infty\}$. *Надграфиком* (epigraph) функции f называется множество

$$\mathbf{epi}\, f \quad = \quad \big\{\, (x, u) \in X \times \mathbb{R} \quad \big| \quad x \in \mathbf{dom}\, f \,,\ u \geq f(x) \,\big\}.$$

Определение 3.1 *Функция f называется выпуклой, если $\mathbf{epi}\, f$ – выпуклое множество.*

Функция f называется *собственной*, если $f(x) \neq -\infty$ для любого x и $\mathbf{dom}\, f \neq \emptyset$. Доказательство следующего предложения элементарно и оставляется читателю.

Предложение 3.1 *Для любой собственной функции следующие свойства равносильны:*
а) $\mathbf{epi}\, f$ – выпуклое множество;
б) $\mathbf{dom}\, f$ выпукло и $f\big((1-t)x + ty\big) \leq (1-t)f(x) + tf(y)$ для любых $x, y \in \mathbf{dom}\, f$ и $t \in [0, 1]$.

Замечание 3.1 Если в пункте б) одна из точек x, y (или обе) не принадлежат $\mathbf{dom}\, f$ то формально основное неравенство также выполнено, если только функция определена в точках $x, y, (1-t)x+ty$. В самом деле, если $t = 0$ или 1, то неравенство обращается в равенство, а если $t \in (0, 1)$, то правая часть равна $+\infty$ (поскольку одна из величин $f(x), f(y)$ равна $+\infty$, а другая не равна $-\infty$), и значит неравенство верно, не зависимо от того, какое значение принимает левая часть.

Если произвольное множество $A \subset X$ выпукло, то можно сказать, что функция $f : A \to \mathbb{R}$ выпукла, если она удовлетворяет неравенству $f\big((1-t)x+ty\big) \leq (1-t)f(x) + tf(y)$ для любых $x, y \in A$ и $t \in [0, 1]$. При этом мы доопределяем f на всём X как $f(x) = +\infty$, $x \notin A$. Теперь данное неравенство будет выполнено для всех точек x, y.

Везде далее, если не оговорено обратное, мы будем доопределять выпуклую функцию значением $+\infty$, и, соответственно, будем предполагать, что любая выпуклая функция определена на всём пространстве X.

Предложение 3.2 *(Неравенство Йенсена). Для любой выпуклой функции f, для любых точек $x_1, \ldots, x_n \in \mathbf{dom}\, f$ и для любых неотрицательных чисел $t_1, \ldots, t_n$ таких, что $\sum_{i=1}^n t_i = 1$ выполнено $f\big(\sum_{i=1}^n t_i x_i\big) \leq \sum_{i=1}^n t_i\, f(x_i)$.*

`Доказательство.` Применяем предложение 1.1 к выпуклому множеству $\mathbf{epi}\, f$.

$\square$

Изучим теперь основные свойства выпуклых функций. Начнём с такого вспомогательного факта.

Лемма 3.1 *Пусть $f : X \to \bar{\mathbb{R}}$ – выпуклая собственная функция, $x, x+h \in \mathbf{dom}\, f$ и $0 \le \alpha \le 1$. Тогда*

$$\frac{f(x+\alpha h) \;-\; f(x)}{\alpha} \quad \le \quad f(x+h) \;-\; f(x)\,. \tag{3.1}$$

`Доказательство.` Поскольку $x+\alpha h \;=\; (1-\alpha)x \;+\; \alpha(x+h)$, имеем $f(x+\alpha h) \;\le\; (1-\alpha)f(x) \;+\; \alpha f(x+h)$, откуда получаем требуемое.

□

Этот технический факт даёт несколько замечательных следствий, два из которых объясняют преимущества выпуклых функций в экстремальных задачах.

Предложение 3.3 *Для выпуклой собственной функции каждый локальный минимум $x \in \mathbf{dom}\, f$ является её абсолютным минимумом.*

`Доказательство.` Предположим, что найдется точка $x+h$, в которой $f(x+h) \;<\; f(x)$ (и следовательно $x+h \in \mathbf{dom}\, f$). Для достаточно маленьких $\alpha > 0$ имеем $f(x+\alpha h) \;\ge\; f(x)$, поэтому левая часть неравенства (3.1) неотрицательна, в то время как правая – отрицательна.

□

Предложение 3.4 *Если выпуклая собственная функция f в точке $x \in \mathbf{dom}\, f$ дифференцируема (т.е., имеет вариацию по Лагранжу) вдоль направления h, то $\delta_f(x,h) \;\le\; f(x+h) - f(x)$.*

`Доказательство.` Предел левой части неравенства (3.1) при $\alpha \to +0$ равен $\delta_f(x,h)$, предел правой равен $f(x+h) - f(x)$.

□

Следствие 3.1 *Если выпуклая собственная функция f в точке $x \in \mathrm{int}\,(\mathbf{dom}\, f)$ дифференцируема по Гато, то $f'(x)[h] \;\le\; f(x+h) \;-\; f(x)$ для любого h.*

Предложение 3.5 *Если выпуклая собственная функция f в точке $x \in \mathbf{dom}\, f$ дифференцируема (т.е., имеет вариацию по Лагранжу) вдоль каждого направления h, для которого $x+h \in \mathbf{dom}\, f$, и $\delta_f(x,h) \;=\; 0$, то x – точка абсолютного минимума.*

`Доказательство.` Следует из предложения 3.4.

□

Следствие 3.2 *Если функция f – выпуклая и собственная, а её производная по Гато в некоторой точке $x \in \mathrm{int}\,\big(\mathbf{dom}\, f\big)$ равна нулю, то $f \in \mathrm{absmin}\, f$.*

Итак, для выпуклых функций необходимое условие локального минимума является достаточным. Ситуация с максимумом выпуклых функций совершенно иная: максимум (если достигается вообще) достигается в крайней точке области определения.

Предложение 3.6 *Если множества $A \subset \mathbb{R}^d$, и $\mathbf{extr}\,(A)$ компактны, то для любой выпуклой функции $f : A \to \mathbb{R}$ имеем $\max_{x\in A} f(x) \;=\; \max_{x\in \mathbf{extr}\,(A)} f(x)$.*

`Доказательство.` По теореме 1.2 выпуклая оболочка множества $\mathbf{extr}\,(A)$ компактна, а по теореме Крейна-Мильмана (теорема 2.4) замыкание этой оболочки (т.е., она сама) совпадает с A. Следовательно, любая точка $x \in A$ является выпуклой комбинацией не более чем $d+1$ точки (теорема 1.1) множества $\mathbf{extr}\,(A)$: $x = \sum_{i=1}^{d+1} t_i x_i$. Тогда по неравенству Йенсена

$$f(x) \quad \le \quad \sum_{i=1}^{d+1} t_i f(x_i) \;\le\; \Big(\sum_{i=0}^{d+1} t_i\Big) \max_{i=1,\dots,d+1} f(x_i) \;=\; \max_{i=1,\dots,d+1} f(x_i) \;\le\; \sup_{y \in \mathbf{extr}\,(A)} f(y)\,.$$

□

Следующее свойство выпуклых функций: непрерывность в всех внутренних точках её эффективного множества. Мы установим нечто большее: выпуклая функция не только непрерывна, но и локально липшицева.

Определение 3.2 *Функция f локально липшицева в точке x, если существует $K > 0$ и шар U с центром x, в котором функция определена и $|f(x+h) - f(x)| \le K\|h\|$ для всех $x + h \in U$.*

Ясно, что из локальной липшицевости следует непрерывность.

Теорема 3.1 *Выпуклая функция в $\mathbb{R}^d$ локально липшицева в каждой внутренней точке своей эффективной области.*

`Доказательство.` Пусть $A = \mathbf{dom}\, f$, $x \in \operatorname{int} A$, шар радиуса R с центром x лежит в A. Покажем, что для любого $r < R$ существует константа K такая, что $|f(x+h) - (x)| \le K\,|h|$ для всех h таких, что $|h| \le r$. Если это не так, то найдётся последовательность h_i, для которой $|h_i| \le r$ при всех i и $\frac{|f(x+h_i)-f(x)|}{|h_i|} \to \infty$ при $i \to \infty$. Предполагая, что $|f(x+h_i) - f(x)| \ne 0$, положим

$$\beta_i \quad = \quad \operatorname{sign}\big(f(x+h_i) - f(x)\big)\,\frac{r}{|h_i|}\,; \qquad u_i \quad = \quad \beta_i\, h_i\,.$$

Таким образом, $|u_i| = r$ для всех i. Кроме того,

$$\frac{|f(x+u_i) \,-\, f(x)|}{|u_i|} \quad \ge \quad \frac{|f(x+h_i) \,-\, f(x)|}{|h_i|}\,.$$

Действительно, в случае $f(x+h_i) - f(x) > 0$ неравенство следует из (3.1) при $h = u_i, \alpha = \frac{1}{\beta_i}$, а случае $f(x+h_i) - f(x) < 0$ имеем цепочку неравенств:

$$\frac{|f(x+u_i) \,-\, f(x)|}{|u_i|} \quad \ge \quad \frac{|f(x-h_i) \,-\, f(x)|}{|-h_i|} \quad \ge \quad \frac{|f(x+h_i) \,-\, f(x)|}{|h_i|}\,.$$

Первое неравенство вновь следует из (3.1) при $h = u_i, \alpha = \frac{h_i}{r}$, а второе – из неравенства $f(x-h_i) + f(x+h_i) \ge 2f\big(\frac{x-h_i+x+h_i}{2}\big) = 2f(x)$, которое следует из выпуклости f. Далее, поскольку в любом случае $f(x+u_i) - f(x) > 0$ и $|u_i| = r$ для всех i, получаем, что $f(x+u_i) - f(x) \to \infty$, а значит $f(x+u_i) \to \infty$ при $i \to \infty$. В силу компактности, переходя к подпоследовательности, считаем, что u_i сходится к некоторой точке u. Так как $|u| = r < R$, то $x + u \in \operatorname{int} A$. Поэтому точку $x + u$ можно поместить в центр некоторого симплекса Δ, целиком лежащего в A. Пусть M – максимальное из значений функции f в вершинах данного симплекса. В силу предложения 3.6 имеем $f(z) \le M$ для любой точки $z \in \Delta$. Так как $x + u_i \to x + u$ и при этом $x + u \in \operatorname{int}\Delta$, то $x + u_i \in \Delta$, а значит $f(x+u_i) \le M$ для всех достаточно больших i. Но это невозможно, поскольку $f(x+u_i) \to \infty$ при $i \to \infty$.

□

Следствие 3.3 *Выпуклая функция в $\mathbb{R}^d$ непрерывна в каждой внутренней точке своей эффективной области.*

Пример 3.1 В граничных точках эффективной области выпуклая функция может быть не только разрывной, но и локально неограниченной. В качестве примера определим функцию f в замкнутом единичном круге на плоскости следующим образом: в любой внутренней точке круга $f(x) = 0$, а в любой граничной точке круга $f(x)$ – произвольное неотрицательное число. Это может показаться странным, но f будет выпуклой при любом выборе её значений на границе (докажите это!). В частности, мы можем определить её на границе так, что f будет стремиться к бесконечности при подходе к некоторой точке. Поэтому, в данной точке функция f не будет локально ограниченной, а значит, не будет и непрерывной.

Упражнение 9 Приведите пример выпуклой функции в бесконечномерном пространстве, разрывной в некоторой внутренней точке своей эффективной области.

4 Лекция

3 марта 2009

Критерии выпуклости. Теорема Каруша-Куна-Таккера

Критерии выпуклости. Монотонность производной. Неотрицательная определённость второй производной. Выпуклые задачи с ограничениями типа неравенств. Теорема Каруша-Куна-Таккера.

Как доказать выпуклость данной функции? Для этого есть несколько несложных признаков и критериев. Перечислим некоторые из них. Везде далее f – функция, заданная на выпуклом открытом подмножестве U нормированного пространства X.

а) если f выпукла и $\lambda \geq 0$, то функция λf выпукла;

б) сумма конечного числа выпуклых функций выпукла;

в) поточечный предел последовательности выпуклых функций – выпуклая функция;

г) поточечный максимум любой совокупности выпуклых функций – выпуклая функция;

д) если функция $f : Y \to \bar{\mathbb{R}}$ – выпуклая, а $h : X \to Y$ – линейное отображение, то функция $f(h(\cdot)) :\ X \to \bar{\mathbb{R}}$ –выпуклая;

е) если функция $f(x)$ выпуклая, то функция $e^{f(x)}$ также выпуклая;

ж) функция f является выпуклой тогда только тогда когда её ограничение на любую прямую – выпуклая функция.

Все эти утверждения доказываются непосредственно по определению выпуклой функции. Для функций одной переменной мы знаем два схожих критерия выпуклости: дифференцируемая функция выпукла тогда и только тогда когда её производная – неубывающая; дважды дифференцируемая функция выпукла тогда и только тогда когда её вторая производная неотрицательна. Распространяются ли эти критерии на выпуклые функции в произвольных нормированных пространствах ?

Определение 4.1 *Пусть множество $U \subset X$ открыто и выпукло. Отображение $F : X \to X^*$ называется монотонным, если $\big(F(x_2) - F(x_1)\, ,\, x_2 - x_1 \big) \ge 0$ для любых $x_1, x_2 \in U$.*

В частности, если $X = \mathbb{R}$, то это совпадет с определением неубывающей функции.

Теорема 4.1 *Если функция $f : X \to \mathbb{R}$ в каждой точке дифференцируема (по Фреше), то она является выпуклой тогда и только тогда когда её производная монотонна.*

`Доказательство.` Для любой пары точек $x_1, x_2 \in U$ рассмотрим функцию

$$\varphi(t) \;=\; f\big((1-t)x_1 \,+\, tx_2 \big).$$

Она является ограничением функции f на прямую $l \;=\; \big\{ (1-t)x_1 + tx_2 \;\big|\; t \in \mathbb{R} \big\}$. Для любого $t \in \mathbb{R}$ имеем

$$\varphi'(t) \;=\; \Big(f'\big((1-t)x_1 \,+\, tx_2 \big)\;,\; x_2 \,-\, x_1 \Big)\,. \tag{4.2}$$

Подставляя $t = 1$ и $t = 0$, получаем $\varphi'(1) - \varphi'(0) \;=\; \big(f'(x_2) - f'(x_1)\, ,\, x_2 - x_1 \big)$. Если f выпукла, то и её ограничение φ на прямую l выпукло, значит производная $\varphi'(t)$ – неубывающая функция, поэтому $\varphi'(1) - \varphi'(0) \ge 0$. Таким образом, $\big(f'(x_2) - f'(x_1)\, ,\, x_2 - x_1 \big) \ge 0$ для любых точек x_1, x_2, т.е., f' – монотонна.

Обратно, если f' монотонна, то для любой пары чисел $t_2 > t_1$, подставляя в равенство (4.2) и обозначая $\tilde{x}_1 = (1-t_1)x_1 + t_1 x_2\,,\ \tilde{x}_2 = (1-t_2)x_1 + t_2 x_2\,,$ имеем

$$\varphi'(t_2) \,-\, \varphi'(t_1) \;=\; \big(f'(\tilde{x}_2) \,-\, f'(\tilde{x}_1)\,,\, x_2 \,-\, x_1 \big) \;=\; \frac{1}{t_2 - t_1}\, \big(f'(\tilde{x}_2) \,-\, f'(\tilde{x}_1)\,,\, \tilde{x}_2 \,-\, \tilde{x}_1 \big) \;\ge\; 0$$

(в силу монотонности f'). Следовательно, ограничение функции f на прямую l имеет неубывающую производную, а значит – выпукло. Поэтому f – выпукла, так как выпукло её ограничение на любую прямую.

□

Теорема 4.2 *Если функция $f : X \to \mathbb{R}$ в каждой точке дважды дифференцируема, то она является выпуклой тогда и только тогда когда её вторая производная неотрицательна.*

Напомним, что вторая производная функции – это билинейная форма $f''(x)[h_1, h_2]$. Она неотрицательна (т.е., неотрицательно определена) если $f''(x)[h, h] \ge 0$ для любого $h \in X$. В случае $X = \mathbb{R}$ получаем известный из курса математического анализа критерий выпуклости гладкой функции: функция выпукла тогда и только тогда когда её вторая производная всюду неотрицательна.

`Доказательство.` Вновь для любой пары различных точек $x_1, x_2 \in U$ рассмотрим функцию $\varphi(t) \;=\; f\big((1-t)x_1 \,+\, tx_2 \big)$, которая является ограничением функции f на прямую $l = \big\{ (1-t)x_1 \,+\, tx_2 \;\big|\; t \in \mathbb{R} \big\}$. Имеем $\varphi''(0) \;=\; f''(x_1)[x_2 - x_1, x_2 - x_1]$. Таким образом, неотрицательность второй производной функции f равносильна неотрицательности второй производной ограничения этой функции на любую прямую l. Так как для функции одной переменной неотрицательность второй производной равносильна её выпуклости, то получаем: $f'' \ge 0$ тогда и только тогда когда ограничение f на любую прямую выпукло, т.е., когда функция f выпукла.

□

Если $X = \mathbb{R}^d$, то вторая производная $f''(x)$ есть квадратичная форма, которая задаётся матрицей Гесса: $\big(f''(x)\big)_{ij} = \frac{\partial^2 f}{\partial x^i\,\partial x^j}(x)\,,\ i,j = 1,\ldots,d$. Таким образом, выпуклость функции f равносильна тому, что матрица Гесса неотрицательно определена. Неотрицательную определённость можно проверить, например, по критерию Сильвестра (все угловые миноры неотрицательны) или, вычислив характеристический полином (все его корни – собственные числа матрицы – неотрицательны).

Обратимся теперь вновь к выпуклым экстремальным задачам. Из прошлой лекции мы знаем, что в безусловной задаче на минимум, т.е., задаче $f(x) \to \min\,,\ x \in A$ с выпуклой функцией f, заданной на выпуклой множестве A, локальный минимум всегда является абсолютным (предложение 3.3), а в случае дифференцируемой функции условие $f'(\hat{x}) = 0$ является не только необходимым, но и достаточным для того, чтобы точка $\hat{x} \in \operatorname{int} A$ была точкой минимума. Для задач с ограничениями типа неравенств есть метод множителей Лагранжа. Модификация этого метода для выпуклых задач устанавливается в следующей теореме.

Теорема 4.3 *(Каруш, Кун, Таккер). Пусть $f_0(x),...,f_m(x) : A \to \mathbb{R}$ – выпуклые функции, заданные на выпуклом множестве $A \subset X$. Рассмотрим следующую задачу*

$$\begin{array}{l} f_0(x) \ \to\ \min \\ \textit{при условиях} \quad x \in A\,,\ f_i(x) \ \le 0, \quad i = 1,...,m. \end{array} \tag{4.3}$$

Если $\hat{x}$ – решение задачи, то существуют НЕРОН множители $\lambda = (\lambda_0,...,\lambda_m)$, удовлетворяющие условиям:

а) (стационарность) $\min\limits_{x\in A} \mathcal{L}(x,\lambda) = \mathcal{L}(\hat{x},\lambda)$, где $\mathcal{L}(x,\lambda) = \sum\limits_{i=0}^{m} \lambda_i f_i(x)$ – лагранжиан;

б) (неотрицательность): $\lambda_i \ge 0,\ i = 0,...,m$;

в) (дополняющая нежёсткость): $\lambda_i f_i(\hat{x}) = 0, i = 1,...,m$.

Обратно, если для данной точки $\hat{x} \in A$ нашёлся набор множителей λ, удовлетворяющий условиям (а)-(в), причём $\lambda_0 > 0$, то $\hat{x}$ – решение задачи (4.3).

Если выполнено условие Слейтера: существует точка $\tilde{x} \in A$ в которой $f_i(\tilde{x}) < 0,\ i = 1,...,m$, то $\lambda_0 > 0$.

`Доказательство`. Без ограничения общности будем считать, что $f_0(\hat{x}) = 0$, в противном случае мы можем прибавить или отнять константу. Определим множества $B \subset R^{m+1}$ и $C \subset R^{m+1}$:

$$\begin{array}{rcl} B &=& \big\{ v = (v_0,\ldots,v_m) :\ v_i < 0,\ i = 0,\ldots,m \big\} \\ C &=& \big\{ u = (u_0,\ldots,u_m) :\ \exists\, x \in A \ :\ f_i(x) \le u_i\,,\ i = 0,\ldots,m \big\} \end{array}$$

Множество B выпукло и открыто. Проверим, что и множество C выпукло. Возьмём произвольное $t \in [0,1]$. Пусть $u,z \in C$. Это означает, что существуют такие $x,y \in A$, что

$$f_i(x) \ \le\ u_i,\ f_i(y) \ \le z_i\,, \qquad i = 0,\ldots,m\,.$$

Тогда

$$f_i\big((1-t)x + ty\big) \ \le\ (1-t)f_i(x) \ + t f_i(y) \ \le\ (1-t)u_i \ +\ t z_i\,, \qquad i = 0,\ldots,m\,.$$

Следовательно, $(1-t)u + tz \in C$. Мы доказали, что C выпукло. Далее, множества B и C не пересекаются. Действительно, если $u = (u_0,\ldots,u_m) \in B \cap C$, то существует такой x, что $f_i(x) \le u_i < 0$. Тогда x – допустимая точка в задаче (4.3), $f_0(x) < 0$, следовательно,

$\hat{x}$ – не решение. Значит, множества B и C отделимы (теорема 2.1). Вспоминая общий вид линейного функционала в $\mathbb{R}^{m+1}$, получаем, что существует вектор $\lambda \in \mathbb{R}^{d+1} \setminus \{0\}$, для которого $\sup_{v\in B} \big(\lambda, v\big) \ \le \ \inf_{u\in C} \big(\lambda, u\big)$. Заметим, сначала, что $\lambda_i \ge 0$ для всех $i = 0, \ldots, m$. В самом деле, если $\lambda_j < 0$ для некоторого j, то можем взять $v_j \to -\infty$ и $v_i \to 0$ для остальных i, тогда получим $\sup_{v\in B} \big(\lambda, v\big) \ = \ +\infty$. Это невозможно, поскольку для любой точки $\bar{u} \in C$ имеем $\inf_{u\in C} \big(\lambda, u\big) \ \le \ \big(\lambda, \bar{u}\big) \ < \ +\infty$. Итак, $\lambda \ge 0$. Так как при $v \to 0$ имеем $\big(\lambda, v\big) \to 0$, то $\sup_{v\in C} \big(\lambda, v\big) \ = \ 0$, откуда $\inf_{u\in C} \big(\lambda, u\big) \ \ge \ 0$. Однако, если взять $\hat{u} \ = \ \big(\, f_0(\hat{x}), \ldots, f_m(\hat{x})\,\big)$, то $\hat{u}_i \le 0$, и следовательно $(\lambda, \hat{u}) \le 0$. Таким образом,

$$\inf_{u\in C} \big(\lambda, u\big) \quad = \quad \big(\, \lambda \, , \, \hat{u} \big) \quad = \quad 0\, .$$

С другой стороны, для любого $x \in A$ имеем $u \ = \ \big(\, f_0(x), \ldots, f_m(x)\,\big) \ \in \ C$, поэтому $\mathcal{L}\,\big(\, x \, , \, \lambda\,\big) \ = \ \big(\, \lambda \, , \, u\big) \ \ge \ 0$. Таким образом, $\min_{x\in A} \mathcal{L}\,\big(\, x \, , \, \lambda\,\big) \ = \ \mathcal{L}\,\big(\, \hat{x} \, , \, \lambda\,\big) \ = 0$. Итак, набор $\lambda = (\lambda_0, \ldots, \lambda_m)$ удовлетворяет условиям (а) и (б) (стационарности и неотрицательности). Проверим условие (в) дополняющей нежёсткости. Заметим сначала, что $\lambda_i \ge 0$ и $\hat{u}_i \ \le \ 0$, а значит $\lambda_i \, \hat{u}_i \ \ \le \ 0$ для любого $i = 0, \ldots, m$. Поскольку $\sum_{i=0}^{m} \lambda_i \, \hat{u}_i \ = \ \big(\, \lambda \, , \, \hat{u}, \big) \ = \ 0$, получаем $\lambda_i \, \hat{u}_i \ \ = \ 0$ для каждого i.

Теперь установим достаточность. Если вектор $\lambda \ne 0$ таков, что в некоторой точке $\hat{x}$ выполнены условия (а) – (в), то

$$(\lambda \, , \, \hat{u}) \ = \ \lambda_0 \, f_0(\hat{x}) \ = \ \inf_{x\in A} \ \mathcal{L}(x \, , \, \lambda) \ \le \ \inf_{u\in C} (\lambda \, , \, u) \ \le \ (\lambda \, , \, \hat{u})\, .$$

Первое равенство выполнено в силу свойства (в), второе – в силу свойства (а), третье неравенство – поскольку для любого $u \in C$ найдётся $x \in A$, для которого $u_i \ge f_i(x)$, $i = 1, \ldots, m$, что в силу неотрицательности λ, влечёт $(\lambda, u) \ \ge \ \mathcal{L}(x, \lambda)$, четвёртое – по определению inf. Далее, прибавляя или вычитая константу из функции f_0, можно считать, что $f_0(\hat{x}) \ = \ 0$. Следовательно,

$$0 \ = \ \lambda_0 \, f_0(\hat{x}) \ = \ (\lambda, \hat{u}) \ = \ \min_{u\in C} (\lambda, u)\, . \tag{4.4}$$

Если $\lambda_0 > 0$, то $\hat{x}$ – решение задачи. Иначе найдётся $\bar{x}$, для которого $f_0(\bar{x}) \, < \, 0$ и $f_i(\bar{x}) \, \le \, 0$, $i = 1, \ldots, m$. Тогда для вектора $\bar{u} \ = \ \big(f_0(\bar{x}), \ldots, f_m(\bar{x})\big)$ имеем $(\lambda, \bar{u}) \, < \, 0$, что противоречит (4.4).

Наконец, если выполнено условие Слейтера: $f_i(\tilde{x}) \, < \, 0, \ i = 1, ..., m$ в некоторой точке $\tilde{x} \in A$, то в случае $\lambda_0 \, = \, 0$ для вектора $\tilde{u} \ = \ \big(f_0(\tilde{x}), \ldots, f_m(\tilde{x})\big)$ получим $(\lambda, \tilde{u}) \, < \, 0$, что противоречит (4.4). В самом деле, в силу условия НЕРОН $\lambda_j \, > \, 0$ для некоторого $j \ge 1$, а значит $(\lambda, \tilde{u}) \ = \ \sum_{i=1}^{m} \lambda_i \tilde{u}_i \ \le \ \lambda_j \tilde{u}_j \ < \ 0$.

□

Поскольку для дифференцируемой выпуклой функции g и точки $x \in \mathrm{int}\,(\mathbf{dom}\, g)$ условия $g'(x) = 0$ и $x \in \mathrm{absmin}\, f$ равносильны (предложение 3.5), получаем

Следствие 4.1 *При дополнительных условиях: $\hat{x} \in \mathrm{int}\, A$ и f_i дифференцируемы (по Гато) в точке $\hat{x}$, условие (а) можно заменить на равносильное условие $\mathcal{L}_x(\hat{x}, \lambda) \ = \ \sum\limits_{i=0}^{m} \lambda_i f_i'(\hat{x}) \ = \ 0.$*

Замечание 4.1 Задача (4.3) называется выпуклой задачей с ограничениями. Так как множество допустимых точек в этой задаче выпукло, то локальный минимум в ней всегда является абсолютным (предложение 3.5). Теорема 4.3 гарантирует, таким образом, что при условии $\lambda_0 \ne 0$ необходимые условия минимума в выпуклой задачей с ограничениями являются достаточными. Если множество A открыто (например, $A \, = \, X$), то условие Слейтера имеет геометрический смысл: множество допустимых точек имеет внутреннюю точку.

5 Лекция

10 марта 2009

Субдифференциал.

Свойства субдифференциала. Субдифференциал и производная.
Существование субдифференциала во внутренних точках эффективной области.
Критерий дифференцируемости выпуклой функции. Примеры.

Ещё одним преимуществом выпуклых экстремальных задач является то обстоятельство, что методы их решения распространяются на все выпуклые функции, даже не дифференцируемые. Для этого нужно обобщить понятие производной выпуклой функции.

Определение 5.1 *Пусть функция f определена в некоторой окрестности точки x. Функционал $x^* \in X^*$ называется субградиентом функции f в точке x, если $(x^*, h) \le f(x+h) - f(x)$ для любого $h \in X$, для которого f определена в точке $x + h$. Множество всех субградиентов функции f в точке x называется её субдифференциалом и обозначается $\partial f(x)$.*

Подчеркнём, что $\partial f(x)$ – не функционал, а множество функционалов. Из определения субдифферениала следует, что это множество всегда выпукло и замкнуто. В самом деле, для произвольного h множество элементов $x^* \in X^*$, удовлетворяющих неравенству $(x^*, h) \le f(x+h) - f(x)$ есть замкнутое полупространство. Поскольку пересечение любого числа выпуклых замкнутых множеств также выпукло и замкнуто, субдифференциал $\partial f(x)$, как пересечение данных полупространств по всем $h \in X$, является выпуклым и замкнутым. Далее мы увидим (предложение 5.2), что для выпуклой собственной функции множество $\partial f(x)$ ограничено.

Везде далее f – выпуклая собственная функция.

Начнём с того, что в случае дифференцируемой функции её производная совпадает с субдифференциалом.

Предложение 5.1 *Если f дифференцируема по Гато в точке $x \in \mathrm{int}\,(\mathbf{dom}\, f)$, то $\partial f(x) = \{f'(x)\}$.*

`Доказательство.` В силу следствия 3.1 $f'(x)[h] \le f(x+h) - f(x)$ для любого h, откуда $f'(x) \in \partial f(x)$. С другой стороны, для любого $x^* \in \partial f(x)$ и любого $h \in X$ имеем

$$\bigl(x^*, t\,h\bigr) \ \le\ f(x+th) \ -\ f(x) \ =\ t\,f'(x)[h] + o(t), \qquad \text{при } t \to 0\,.$$

Следовательно, $\bigl(x^*, h\bigr) \le f'(x)[h]$ для любого $h \in X$. Применяя это неравенство для элементов h и $-h$, получаем, что $\bigl(x^*, h\bigr) = f'(x)[h]$ для всех h, а значит $x^* = f'(x)$. Таким образом, $\partial f(x) = \{f'(x)\}$.

□

Итак, для дифференцируемых функций субдифференциал совпадает с производной. А для недифференцируемых ? Оказывается, что субдифференциал существует у любой выпуклой функции.

Теорема 5.1 *Для любой точки x в которой функция f непрерывна, имеем* $\partial f(x) \neq \emptyset$.

`Доказательство.` Функция f непрерывна в x, а значит ограничена в некоторой окрестности U этой точки. Итак, $M = \sup_{x \in U} f(x) < \infty$. Тогда открытое множество $\big\{ (u,v) \in (\mathbf{dom}\, f) \times \overline{\mathbb{R}} \ \big| \quad u \in U\,,\ v > M \big\}$ лежит в $\mathbf{epi}\, f$. Итак, надграфик $\mathbf{epi}\, f$ имеет непустую внутренность, причем точка $z = \big(x, f(x)\big)$ не принадлежит этой внутренности. В самом деле, точка $\big(x, f(x) - \varepsilon\big)$ стремится к z при $\varepsilon \to +0$ и не лежит в $\mathbf{epi}\, f$. Согласно теореме 2.1 выпуклые множества $\mathbf{epi}\, f$ и $\{z\}$ отделимы. Это значит, что существует непрерывный ненулевой функционал y^* на пространстве $X \times R$, для которого $\sup_{y \in \mathbf{epi}\, f} (y^*, y) \le (y^*, z)$. Из курса функционального анализа известно, что любой непрерывный линейный функционал y^* на прямом произведении $Y = X \times \mathbb{R}$ имеет вид $y^*[(u,v)] = (x^*, u) + c\, v$, где $x^* \in X^*$ и $c \in \mathbb{R}$. Таким образом,

$$\sup_{u \in \mathbf{dom}\, f\,,\ v \ge f(u)} \Big[\big(x^*, u\big) \ + \ c\, v \Big] \quad \le \quad \big(x^*, x\big) \ + \ c\, f(x)$$

Заметим, что $c \le 0$, в противном случае левая часть неравенства обращается в $+\infty$ при $v \to +\infty$. Следовательно, для любых $u \in \mathbf{dom}\, f$, $v \ge f(u)$ имеем

$$-\,c\Big(v \ - \ f(x) \Big) \quad \ge \quad \big(x^*\,,\, u - x \big)\,,$$

подставляя $v = f(u)$ и обозначая $u - x = h$, получаем

$$-\,c\Big(f(x+h) \ - \ f(x) \Big) \quad \ge \quad \big(x^*\,,\, h \big)\,,$$

Если $c = 0$, то $\big(x^*\,,\, h \big) \le 0$ для всех $h \in X$. Применяя это неравенство для элементов h и $-h$, получаем $\big(x^*\,,\, h \big) = 0$ для всех $h \in X$, следовательно $x^* = 0$, что невозможно. Таким образом, $c < 0$, и, поделив на число $-\,c$, получаем

$$f(x+h) \ - \ f(x) \quad \ge \quad \Big(-\frac{1}{c}\, x^*\,,\, h \Big)\,.$$

Следовательно, $-\frac{1}{c}\, x^* \in \partial f(x)$.

$\square$

Так как при $\dim X < \infty$ выпуклая функция непрерывна в любой точке $x \in \mathrm{int}\,(\mathbf{dom}\, f)$, получаем

Следствие 5.1 *При* $\dim X < \infty$ *выпуклая функция имеет непустой субдифференциал в любой точке* $x \in \mathrm{int}\,(\mathbf{dom}\, f)$.

Теорема 5.2 *(Критерий дифференцируемости выпуклой функции). В случае* $\dim X < \infty$ *выпуклая функция f дифференцируема в точке* $x \in \mathrm{int}\,(\mathbf{dom}\, f)$ *тогда и только тогда когда её субдифференциал $\partial f(x)$ состоит из одной точки.*

`Доказательство.` Достаточность следует из предложения 5.1. Докажем необходимость: если $\partial f(x)$ состоит из одной точки x^*, то f дифференцируема в x и $f'(x) = x^*$. Пусть $X = \mathbb{R}^d$. Не ограничивая общности считаем, что $x = 0$ и $f(x) = 0$. Таким образом, $\big(x^*, h\big) \le f(h)$ для всех h. Тогда выпуклая функция $g(h) = f(h) - \big(x^*, h\big)$ достигает минимума в точке 0. При этом $\partial g(0) = \{0\}$, в противном случае, если $\partial g(0)$ содержит некоторую точку $a^* \neq 0$, то $x^* + a^* \in \partial f(0)$, что невозможно, так как $\partial f(0) = \{x^*\}$. Докажем, что $g'(0) = 0$, из чего будет следовать, что $f'(0) = x^*$. Предположим обратное: существует константа $C > 0$, последовательность точек $u_k \in \mathbb{R}^d$, $|u_k| = 1\,, k \in \mathbb{N}$ и последовательность положительных чисел $\{\alpha_k\}$, стремящаяся к нулю, для которых $g(\alpha_k u_k) \ \ge \ C\,\alpha_k\,, \ k \in \mathbb{N}$.

Из леммы 3.1 следует, что для любого $t > \alpha_k$ имеем $g(tu_k) \ \ge \ \frac{t}{\alpha_k}\, f(\alpha_k u_k) \ \ge \ C\,t$ (для доказательства достаточно положить $x = 0$, $g(x) = 0$, $h = tu_k, \alpha = \alpha_k/t$ в неравенстве (3.1)). Поскольку $\alpha_k \to 0$, при любом $t > 0$ неравенство $g(tu_k) \ \ge \ C\,t$ выполнено для всех достаточно больших k. Пользуясь компактностью единичной сферы, переходя к подпоследовательности, считаем, что $u_k \to u$ при $k \to \infty$. Кроме того, g непрерывна в окрестности точки 0 (теорема 3.1). Следовательно, для малых $t > 0$ имеем $g(tu_k) \ \to \ g(tu)\,, \ k \to \infty$, а значит $g(tu) \ \ge \ C\,t\,.$ Тогда из той же леммы 3.1 следует, что $g(tu) \ \ge \ C\,t$ для всех $t > 0$. Для $t \le 0$ это неравенство тем более верно, поскольку $g(tu) \ \ge \ 0$, а $Ct \ \le \ 0$. Итак, $g(tu) \ \ge \ C\,t$ для всех $t \in \mathbb{R}$. Следовательно, в пространстве $\mathbb{R}^{d+1}$ прямая $\{\,(tu, tC) \mid t \in \mathbb{R}\,\}$ не пересекается с внутренностью надграфика $\mathbf{epi}\,g$. Поэтому, по теореме 2.1 эти множества можно отделить некоторым функционалом $(u^*, \alpha) \in \mathbb{R}^{d+1}$. Итак,

$$\sup_{h \in \mathbf{epi}\, g\,,\, v \ge g(h)} (u^*, h) \, + \, \alpha\, v \quad \le \quad \inf_{t \in \mathbb{R}} t(u^*, u) \, + \, t\,\alpha\, C\,.$$

Тогда $(u^*, u) + \alpha C \, = \, 0$, в противном случае правая часть неравенства равна $-\infty$. Кроме того, $\alpha \le 0$, иначе левая часть обратится к $+\infty$ при $v \to +\infty$. Если $\alpha = 0$, то $(u^*, y) \ \le \ 0$ при всех y, следовательно, $u^* = 0$, что невозможно, поскольку функционал – ненулевой. Значит, $\alpha < 0$. Тогда $(u^*, u) \ = \ -\,\alpha C \ \ne \ 0$, поэтому $u^* \ne 0$. Подставляя $v = g(h)$, получаем, что $(u^*, h) \, + \, \alpha\, g(h) \ \le \ 0$, откуда $\bigl(-\frac{1}{\alpha}\, u^*\,,\, h\bigr) \ \le \ g(h)$ для всех h. Поэтому $-\frac{1}{\alpha}\, u^* \ \in \partial g(0)$. Противоречие завершает доказательство.

$\square$

Пример 5.1 Функция $f : \mathbb{R} \to \mathbb{R}\,,\ f(x) \, = \, x$ при $x > 0\,,\ f(x) \, = \, 0$ при $x \ \le \ 0$ выпукла. Она дифференцируема везде, кроме нуля, поэтому $\partial f(x) \ = \ \{f'(x)\}$ при $x \ne 0$. Таким образом, $\partial f(x) \ = \ \{1\}$ при $x > 0$ и $\partial f(x) \ = \ \{0\}$ при $x < 0$. Найдём субдифференциал в точке $x = 0$. Точка a принадлежит $\partial f(0)$ тогда и только тогда когда $ah \ \le \ f(h) \, - \, f(0)$ для всех $h \in \mathbb{R}$. При $h > 0$ имеем $ah \le h$, откуда $a \le 1$. При $h \le 0$ имеем $ah \le 0$, откуда $a \ge 0$. Таким образом $\partial f(0) \ = \ [0, 1]$.

Пример 5.2 Функция $f \ : \ \mathbb{R} \ \to \ \overline{\mathbb{R}}$, $f(x) \ = \ x \ln x$ при $x \ > \ 0\,,\ f(x) \ = \ 0$ при $x \ = \ 0$, $f(x) \ = \ +\infty$ при $x \ < \ 0$ – выпукла. Однако $\partial f(0) \ = \ \emptyset$. В самом деле, если $a \ \in \ \partial f(0)$, то $ah \ \le \ h \ln h$ для любого $h \ > \ 0$, что невозможно. Противоречия со следствием 5.1 нет, поскольку $0 \notin \mathrm{int}\,(\mathbf{dom}\, f)$.

Пример 5.3 *(Субдифференциал нормы в нуле).* В нормированном пространстве X норма $f(x) \ = \ \|x\|$ является выпуклой функцией. Найдём $\partial f(0)$. Имеем $x^* \ \in \ \partial f(0) \ \Leftrightarrow \ (x^*, h) \ \le \ \|h\| \ \forall \ h \in X$. Разделив на $\|h\|$, получаем $\sup_{\|h\|=1} (x^*, h) \ \le \ 1$. Поскольку $\sup_{\|h\|=1} (x^*, h) \ = \ \|x^*\|$, получаем $\|x^*\| \ \le \ 1$. Таким образом, $\partial f(0) \ = \ \bigl\{\, x^* \ \in \ X^* \ \bigm| \ \|x^*\| \ \le \ 1\,\bigr\}$ – единичный шар в пространстве X^*. Например, для евклидовой нормы в $\mathbb{R}^d$, т.е., $|x| \ = \ \bigl[\sum_{i=1}^d (x^i)^2\bigr]^{1/2}$ её субдифференциал в нуле – единичный шар с центром в нуле.

Предложение 5.2 *Для любой выпуклой функции f в любой точке непрерывности субдифференциал $\partial f(x)$ – ограниченное выпуклое замкнутое множество.*

В доказательстве мы используем теорему Банаха-Штейнгауза, которая доказывается в курсе функционального анализа.

Предложение 5.3 *(Банах, Штейнгауз). Если множество функционалов $A^* \subset X^*$ ограничено в каждой точке, т.е., для любого $x \in X$ множество $\{(x^*, x) \mid x^* \in A^*\}$ ограничено, то A^* – ограниченное множество.*

`Доказательство предложения 5.2.` Выпуклость и замкнутость очевидна, докажем ограниченность. Поскольку $x \in \operatorname{int}(\mathbf{dom} f)$, точка x входит в $\mathbf{dom} f$ вместе с шаром некоторого радиуса r. Возьмём произвольную точку $h \in X$ и такое $t > 0$, что $\|th\| < r$. Для любого $x^* \in \partial f(x)$ имеем $(x^*, th) \leq f(x+th) - f(x)$ и $(x^*, -th) \leq f(x-th) - f(x)$, откуда $\frac{f(x)-f(x-th)}{t} \leq (x^*, h) \leq \frac{f(x+th)-f(x)}{t}$. Итак, множество $\partial f(x)$ ограниченно в каждой точке h, и по теореме Банаха-Штейнгауза оно ограничено.

□

Следствие 5.2 *В случае* $\dim X < \infty$ *субдифференциал выпуклой функции f в любой точке* $x \in \operatorname{int}(\mathbf{dom} f)$ *– выпуклый компакт.*

Ещё одно общее свойство производной и субдифференциала – это их локализованность. Производная функции в точке x зависит только от значений, которые функция принимает вблизи точки x, и не зависит от точек, далёких от x. Следующее вспомогательное утверждение устанавливает аналогичное свойство субдифференциала.

Лемма 5.1 *(о локализации субдифференциала). Пусть $f : X \to \overline{R}$ – выпуклая собственная функция и $x \in \mathbf{dom} f$. Если для некоторого $x^* \in X^*$ существует число $\varepsilon > 0$ такое, что для любого $h \in X$, $\|h\| < \varepsilon$ справедливо неравенство $(x^*, h) \leq f(x+h) - f(x)$, то $x^* \in \partial f(x)$.*

`Доказательство.` Для любого $h \in X$ возьмём такое $t > 0$, что $\|th\| < \varepsilon$. Имеем $(x^*, th) \leq f(x+th) - f(x)$, откуда $(x^*, h) \leq \frac{f(x+th)-f(x)}{t}$. Применив теперь лемму 3.1, получаем $(x^*, h) \leq f(x+h) - f(x)$. В силу произвольности h из этого следует, что $x^* \in \partial f(x)$.

□

Таким образом, субдифференциал однозначно определён значениями функции в сколь угодно малой окрестности точки x.

6 Лекция

17 марта 2009

Субдифференциальное исчисление.

Критерий минимума выпуклой функции. Теоремы Рокафеллара-Моро и Дубовицкого-Милютина.

Следующая теорема объясняет роль субдифференциалов в выпуклых экстремальных задачах.

Теорема 6.1 *Если f – выпуклая собственная функция и $x \in \mathbf{dom}\, f$, то $x \in \mathrm{absmin}\, f \Leftrightarrow 0 \in \partial f(x)$.*

`Доказательство.` Соотношение $0 \in \partial f(x)$ означает, что $0 = (0,h) \le f(x+h) - f(x)$ для любого $h \in X$, т.е., $x \in \mathrm{absmin}\, f$.

□

Следствие 6.1 *В условиях теоремы 4.3 условие (а) можно заменить на равносильное условие $0 \in \partial \mathcal{L}_x(\hat{x}, \lambda)$.*

Таким образом, для решения выпуклых экстремальных задач нужно уметь вычислять субдифференциалы выпуклых функций. Основу субдифференциального исчисления составляют теоремы Моро-Ракафеллара (о субдифференциале суммы) и Дубовицкого-Милютина (о субдифференциале максимума).

Напомним, что суммой по Минковскому двух множеств $A, B \subset X$ называется множество $A+B = \{\, a+b \mid a \in A,\, b \in B \,\}$. Лёгким упражнением для читателя будет показать, что если два множества выпуклы, то и их сумма выпукла.

Теорема 6.2 *(Моро, Рокафеллар). Если $f_1, f_2 : X \to \overline{\mathbb{R}}$ – собственные выпуклые функции, имеющие непустые субдифференциалы в точке $\hat{x}$, причём хотя бы одна из них в этой точке непрерывна, то $\partial(f_1 + f_2)(\hat{x}) = \partial f_1(\hat{x}) + \partial f_2(\hat{x})$.*

Пример 6.1 Если функции f_1, f_2 выпуклы, f_1 дифференцируема в точке $\hat{x}$, и $\partial f_2(\hat{x}) \neq \emptyset$, то $\partial(f_1 + f_2)(\hat{x}) = f_1'(\hat{x}) + \partial f_2(\hat{x})$. Например, если $f_1(x) = (a, x)$, $a \in X^*$, то $\partial(f_1 + f_2)(\hat{x}) = a + \partial f_2(\hat{x})$. Если $f_1 \equiv \mathrm{const}$, то $\partial(f_1 + f_2)(\hat{x}) = \partial f_2(\hat{x})$

`Доказательство.` Нам надо доказать, что множества $\partial(f_1 + f_2)(\hat{x})$ и $\partial f_1(\hat{x}) + \partial f_2(\hat{x})$ совпадают. Поэтому покажем, что верны прямое и обратное включения.

Обратное включение. Пусть $x^* \in \partial f_1(\hat{x}) + \partial f_2(\hat{x})$. Это значит, что найдутся $x_1^* \in \partial f_1(\hat{x})$ и $x_2^* \in \partial f_2(\hat{x})$, для которых $x^* = x_1^* + x_2^*$. Проверим, что последнее является субградиентом:

$$\big(f_1 + f_2\big)(\hat{x} + h) - \big(f_1 + f_2\big)(\hat{x}) - \big(x^*, h\big) =$$

$$= \big(f_1(\hat{x}+h) - f_1(\hat{x})\big) - \big(x_1^*, h\big) + \big(f_2(\hat{x}+h) - f_2(\hat{x})\big) - \big(x_1^*, h\big) \ge 0, \quad \text{так как}$$

$$\big(f_1(\hat{x}+h) - f_1(\hat{x})\big) - \big(x_1^*, h\big) \ge 0 \quad \text{и} \quad \big(f_2(\hat{x}+h) - f_2(\hat{x})\big) - \big(x_2^*, h\big) \ge 0.$$

Итак, $x^* \in \partial\big(f_1 + f_2\big)(\hat{x})$, что и требовалось доказать.

Прямое включение. Пусть $x^* \in \partial(f_1 + f_2)(\hat{x})$, надо доказать, что найдутся $x_1^* \in \partial f_1(\hat{x})$ и $x_2^* \in \partial f_2(\hat{x})$ такие, что $x^* = x_1^* + x_2^*$. Без ограничения общности можем считать, что $x^* = 0$. Действительно, рассмотрим $\tilde{f}_1(x) = f_1(x) - \big(x^*, x\big)$, тогда, поскольку $x^* \in \partial(f_1 + f_2)(\hat{x})$, имеем $0 \in \partial(\tilde{f}_1 + f_2)(\hat{x}) \Rightarrow 0 = \tilde{x}_1^* + x_2^*$, где $\tilde{x}_1^* \in \partial\tilde{f}_1(\hat{x})$, а $x_2^* \in \partial f_2(\hat{x})$. Следовательно, $x_1^* = \tilde{x}_1^* + x^* \in \partial f_1(\hat{x})$. Поэтому без ограничения общности считаем, что $f_1(\hat{x}) = f_2(\hat{x}) = 0$ и f_2 непрерывна в $\hat{x}$ (по условию одна из двух функций должна быть непрерывна в $\hat{x}$). Условие, что 0 – субградиент, означает, что $\big(f_1 + f_2\big)(x) \ge \big(f_1 + f_2\big)(\hat{x}) + \big(x^*, x - \hat{x}\big) = 0$ для любого x (так как первое слагаемое равно 0 по предположению, а второе – в силу условия $x^* = 0$. Пусть $A = \{\,(x, v) \mid x \in \mathbf{dom}\, f_1,\ v < -f_1(x)\,\}$, и $B = \mathbf{epi}\, f_2 = \{\,(x, u) \mid x \in \mathbf{dom}\, f_2,\ u \ge f_2(x)\,\}$.

Множества A и B – выпуклы, непусты и не пересекаются, поскольку из условия $(x,u) \in A$ следует, что $u < -f_1(x) \le f_2(x)$, а значит, $(x,u) \in \mathbf{epi}\, f_2 = B$ по определению надграфика. Поскольку функция f_2 непрерывна в точке $\hat{x}$, она ограничена в некоторой окрестности U этой точки: $M = \sup_{x \in U} f_2(x) < \infty$. Тогда открытое множество $\big\{ (u,v) \in (\mathbf{dom}\, f_2) \times \overline{\mathbb{R}} \;\big|\; u \in U\,,\ v > M \big\}$ лежит в $\mathbf{epi}\, f_2$. Таким образом, $\operatorname{int} B \ne \emptyset$. По теореме 2.1 множества A и B отделимы: существуют, не обращающиеся в ноль одновременно, $x^* \in X^*, \lambda \in \mathbb{R}$ такие, что для любых $(x_1, v) \in A$ и $(x_2, u) \in B$ выполнено: $\big(x^*\,,\, x_1\big) + \lambda\, v \le \big(x^*, x_2\big) + \lambda u$. Рассмотрим далее несколько случаев:

1) Пусть $x_1 = x_2 = \hat{x}, v < 0, u = f_2(x_2) = 0$. Тогда $\big(x^*, \hat{x}\big) + \lambda v \le \big(x^*, \hat{x}\big)$, следовательно, $\lambda v \le 0$, откуда $\lambda \ge 0$, так как $v < 0$.

2) Пусть $x_1 = \hat{x}$, x_2 – произвольное, $u = f_2(x_2)$. Тогда $\big(x^*, \hat{x}\big) + \lambda v \le \big(x^*, x_2\big) + \lambda f_2(x_2)$. Устремляя v к 0, в пределе получаем $\big(x^*, \hat{x}\big) \le \big(x^*, x_2\big) + \lambda f_2(x)$. Далее, проделываем доказательство аналогичное доказательству непустоты субдифференциала. Тогда в итоге мы получим, что $\lambda > 0$ и $-\frac{x^*}{\lambda} \in \partial f_2(\hat{x})$.

3). Пусть $x_2 = \hat{x}$, x_1 – произвольный элемент из $\mathbf{dom}\, f_1$. Тогда для любого $v < -f_1(x_1)$ выполнено $\big(x^*, x_1\big) + \lambda v \le \big(x^*, \hat{x}\big) + \lambda f_2(\hat{x})$. Устремляя v к $-f_1(x) - 0$, в пределе получаем неравенство $\big(x^*, x_1\big) - \lambda f(x_1) \le \big(x^*, \hat{x}\big)$, что можно переписать так: $-\big(x^*, x_1\big) + \lambda f(x_1) \ge \big(-x^*, \hat{x}\big)$, откуда $\frac{1}{\lambda} x^* \in \partial f_1(\hat{x})$. Значит, верно и прямое включение. Следовательно, два множества $\partial(f_1 + f_2)(\hat{x})$ и $\partial f_1(\hat{x}) + \partial f_2(\hat{x})$ совпадают.

$\square$

Теорема 6.3 *(Дубовицкий, Милютин) Если $f_i : X \to \overline{\mathbb{R}}$ – собственные выпуклые функции, конечные и непрерывные в точке $\hat{x}$, $f(x) = \max_{i=1,\dots,m} f_i(x)$, то*

$$\partial f(\hat{x}) \quad = \quad \mathbf{co} \left(\bigcup_{f_i(\hat{x}) = f(\hat{x})} \partial f_i(\hat{x}) \right).$$

`Доказательство.` Вначале рассмотрим случай, когда для любого i $f_i(\hat{x}) = f(\hat{x})$ для любого i. Покажем, что два указанные множества совпадают, доказав справедливость прямого и обратного включения.

Прямое включение. Докажем, что

$$\partial f(\hat{x}) \supset \mathbf{co} \left(\bigcup_{f_i(\hat{x}) = f(\hat{x})} \partial f_i(\hat{x}) \right).$$

Пусть $x^* \in \partial f_i(\hat{x})$, $h \in X$. По условию $f(\hat{x}+h) \ge f_i(\hat{x}+h) \ge f_i(\hat{x}) + \big(x^*, h\,\big) = f(\hat{x}) + \big(x^*, h\big)$. Таким образом, мы получили, что $x^* \in \partial f(\hat{x})$, отсюда, учитывая выпуклость субдифференциала $\partial f(\hat{x})$, мы находим:

$$\bigcup_{i=1}^{m} \partial f_i(\hat{x}) \subset \partial f(\hat{x}) \quad \Rightarrow \quad \partial f(\hat{x}) \supset \mathbf{co} \left(\bigcup_{f_i(\hat{x}) = f(\hat{x})} \partial f_i(\hat{x}) \right).$$

Обратное включение. Проверим, что если $x^* \in \partial f(\hat{x})$, то

$$x^* \in \mathbf{co} \left(\bigcup_{f_i(\hat{x}) = f(\hat{x})} \partial\, f_i(\hat{x}) \right).$$

Без ограничения общности считаем, что $x^* = 0$, в общем случае заменим функции $f_i(x)$ на $f_i(x) - \big(x^*, x\big)$. Также будем считать что $f_1(\hat{x}) = \ldots = f_m(\hat{x}) = f(\hat{x})$. Тогда для любого x справедливо неравенство $f(x) \geq f(\hat{x}) + \big(x^*, x - \hat{x}\big) = 0$.

Предположим, что для любого i существует x_i из X такой, что $f_i(x_i) < 0$ и $f_j(x_i) \leq 0$ для любого j. Возьмем

$$x = \frac{1}{m}\sum_{i=1}^{m} x_i \quad \Rightarrow \quad f_i(x) \leq \frac{1}{m}\sum_{j=1}^{m} f_i(x_j) \leq f_i(x_i) < 0 \quad \Rightarrow \quad f(x) < 0.$$

Получили противоречие. Таким образом, найдется такой индекс i, для которого система неравенств $f_i(x) < 0$ и $f_j(x) \leq 0\, j = 1, \ldots, m$ не имеет решения.

Существует $\varepsilon > 0$ такой, что для любого x, из шара $B(\hat{x}, \varepsilon)$ с центром $\hat{x}$ и радиусом ε все функции $f_i(x)$ конечны. Рассмотрим $A = B(\hat{x}, \varepsilon)$ и рассмотрим выпуклую экстремальную задачу:

$$\begin{cases} f_i(x) \;\to\; \min, \\ f_j(x) \leq 0, \quad x \in A, \quad j \neq i. \end{cases}$$

В силу предыдущих результатов, $f_i(x)$ не может быть отрицательным, поэтому решением этой задачи будет точка $\hat{x}$. Воспользуемся теоремой Каруша-Куна-Таккера (теорема 4.3), которая гарантирует существование вектора $\lambda = \big(\lambda_1, \ldots, \lambda_m\big) \neq 0$, $\lambda_i \geq 0$ такого, что для любого $x \in A$ имеем

$$\sum_{i=1}^{m} \lambda_i f_i(x) \geq \sum_{i=1}^{m} \lambda_i f_i(\hat{x}) = 0 \quad \Rightarrow \quad 0 \in \partial\left(\sum_{i=1}^{m} \lambda_i f_i\right)(\hat{x}) \quad \text{(в силу следствия 6.1)}$$

Таким образом,

$$\partial\left(\sum_{i=1}^{m} \lambda_i f_i(\hat{x})\right) = \sum_{i=1}^{m} \partial\left(\lambda_i f_i\right)(\hat{x}) = \sum_{i=1}^{m} \lambda_i\, \partial f_i\,(\hat{x}).$$

Следовательно, существует $x_i^* \in \partial f_i(\hat{x})$ такой, что $0 = \sum_{i=1}^m \lambda_i x_i^*$. Возьмем тогда

$$\lambda = \sum_{i=1}^{m} \lambda_i > 0, \qquad \alpha_i = \frac{\lambda_i}{\lambda}, \qquad i = 1, \ldots, m.$$

Поскольку $0 = \sum_{i=1}^m \alpha_i x_i^*$, $\sum_{i=1}^m \alpha_i = 1$ и $x_i^* \in \partial f_i(\hat{x})$, имеем

$$x_i^* \in \partial f_i(\hat{x}) \in \mathbf{co}\left(\bigcup_{j=1}^{m} \partial f_j(\hat{x})\right) \quad \Rightarrow \quad \sum_{i=1}^{m} \alpha_i x_i \in \mathbf{co}\left(\bigcup_{j=1}^{m} \partial f_j(\hat{x})\right).$$

Получили обратное включение для частного случая, когда все функции f_i принимают одинаковые значения в точке $\hat{x}$. Теперь рассмотрим общий случай. Введем следующие обозначения: $I = \big\{ i \;\big|\; f_i(\hat{x}) = f(\hat{x}) \big\}$ $J = \big\{ j \;\big|\; f_j(\hat{x}) < f(\hat{x}) \big\}$. В силу непрерывности всех функций f_i в $\hat{x}$, существует $\varepsilon > 0$ такой, что

$$\big|x - \hat{x}\big| < \varepsilon \quad \Rightarrow \quad \text{для любого } j \in J \quad f_j(x) < f(x), \quad f(x) = \max_{i \in I} f_i(x).$$

Тогда по лемме о локализации субдифференциала (лемма 5.1) имеем $\partial f(\hat{x}) = \partial(\max f_i)(\hat{x})$. Для последнего выражения нами уже все доказано.

□

Пример 6.2 . Рассмотрим задачу

$$f(x,y) \;=\; \sqrt{x^2+y^2} \;+\; |x| \;+\; \frac{3}{2}\,x \;+\; y^4 \;+\; \frac{1}{2}\,y \qquad \to \qquad \min\,.$$

Функция f выпукла. Кроме того, она непрерывна и коэрцитивна, поэтому её минимум достигается в некоторой точке $(\hat{x},\hat{y})$. В этой точке имеем $0 \in \partial f(\hat{x},\hat{y})$. Функция f дифференцируема везде кроме точек вида $(0,y)$ (в них не дифференцируем модуль), а в точке $(0,0)$ не дифференцируем и корень. Если начать разбор вариантов с "простейшего" случая $\hat{x} \neq 0$, когда f дифференцируема, то получим систему уравнений $f_x(\hat{x},\hat{y}) \,=\, f_y(\hat{x},\hat{y}) \,=\, 0$, которую, однако не удастся решить (она приводит к алгебраическому уравнению высокой степени). Как же найти минимум ? Рассмотрим случай $(\hat{x},\hat{y}) \,=\, 0$. В этой точке субдифференциал будет наиболее массивным, поэтому есть надежда что он будет содержать ноль. Имеем $\partial\bigl(\sqrt{x^2+y^2}\bigr)\,(0,0)$ – круг радиуса 1 с центром в нуле (пример 5.3), $\partial\bigl(\frac{3}{2}\,x \,+\, y^4 \,+\, \frac{1}{2}\,y\bigr)\,(0,0) \,=\, \bigl(\frac{3}{2}\,x \,+\, y^4 \,+\, \frac{1}{2}\,y\bigr)'\,(0,0) \,=\, \bigl(\frac{3}{2},\frac{1}{2}\bigr)$. Для вычисления субдифферециала модуля применим теорему Дубовицкого-Милютина. Поскольку $|x| \,=\, \max\{x,-x\}$, получаем $\partial|x|\,(0,0) \,=\, \mathbf{co}\,\bigl\{(-1,0)\,,\,(1,0)\,\bigr\}$ – отрезок с концами $(-1,0)$ и $(1,0)$. Теперь применяем теорему Моро-Рокафеллара. Множество $\partial f\,(0,0)$ – есть сумма круга $B(0,1)$, отрезка $[-1,1]$ оси OX и вектора $\bigl(\frac{3}{2},\frac{1}{2}\bigr)$. Изобразив это множество (оно является квадратом со стороной 2 с центром в точке $\bigl(\frac{3}{2},\frac{1}{2}\bigr)$ с построенными на его вертикальных сторонах во внешнюю сторону полукругами), получаем, что оно содержит ноль. Таким образом, $(0,0) \,\in\, \mathrm{absmin}$, и минимум равен нулю. Разбор остальных случаев оказался не нужен.

Пример 6.3 . В евклидовом пространстве $\mathbb{R}^d$ дано n точек $x_1,\ldots,x_n$, не лежащие на одной прямой. Какая точка плоскости является решением задачи $f(x) \,=\, \sum_{i=1}^{n}\,|x-x_i| \,\to\, \min$? Так как функция f непрерывна и коэрцитивна, она достигает минимума в некоторой точке $\hat{x}$. Положим $f_j(x) \,=\, |x-x_j|$. Рассмотрим два случая:

1) $\hat{x}$ не совпадает ни с одной из точек x_i. В этом случае каждая из функций f_i дифференцируема и $f'(\hat{x}) \,=\, u_i \,=\, \frac{\hat{x}\,-\,x_i}{|\hat{x}\,-\,x_i|}$ – единичный вектор, сонаправленный вектору $\hat{x}-x_i$. Точка $\hat{x}$ будет решением задачи в том и только том случае, когда $f'(\hat{x}) \,=\, \sum_{i=1}^{n}\,u_i \,=\, 0$.

2) $\hat{x} \,=\, x_j$ для некоторого j. В этом случае $\partial f_j(\hat{x}) \,=\, B(0,1)$ – шар радиуса 1 с центром в нуле (пример 5.3), и $\partial f_i(\hat{x}) \,=\, \{u_i\}$ для всех $i \,\neq\, j$, поскольку все функции f_i диффферецируемы в точке $\hat{x}$. Тогда по теореме Моро-Рокафеллара (теорема 6.2) имеем $\partial f(x_j) \,=\, B(0,1) \,+\, \sum_{i\neq j}^{n}\,u_i \,=\, B\bigl(\sum_{i\neq j}\,u_i\,,\,1\bigr)$. Тогда (теорема 6.1) $x_j \,\in\, \mathrm{absmin}$ в том и только том случае, когда шар с центром $\sum_{i\neq j}\,u_i$ и радиусом 1 содержит точку 0. Это означает, что $\bigl|\,\sum_{i\neq j}\,u_i\,\bigr| \,\le\, 1$. Итак,

для произвольных n точек $x_1,\ldots,x_n \,\in\, \mathbb{R}^d$, не лежащих на одной прямой, либо найдётся точка $\hat{x}$, не совпадающая ни с одной из них, для которой $\sum_{i=1}^{n}\,u_i \,=\, 0$, либо среди данных точек найдётся такая, что $\bigl|\,\sum_{i\neq j}\,u_i\,\bigr| \,\le\, 1$. В каждом случае данная точка является решением задачи $\sum_{i=1}^{n}\,|x-x_i| \,\to\, \min$.

Вернёмся теперь к теореме Каруша-Куна-Таккера (теорема 4.3) и посмотрим, что означают условия (а) – (в) в терминах субдифференциалов функций f_i, $i \,=\, 0,\ldots,m$. Назовём неравенство $f_i(x) \,\le\, 0$ *активным*, если $f_i(\hat{x}) \,=\, 0$ и *пассивным* если $f_i(\hat{x}) \,<\, 0$. Условие (в) дополняющей нежёсткости означает, что $\lambda_i \,=\, 0$ для всех пассивных неравенств. Обозначим $I(\hat{x}) \,=\, \{0\} \cap \{\,i \;|\; f_i(\hat{x}) \,=\, 0\}$, т.е., это множество состоит из нуля и индексов всех активных неравенств. Все множители λ_i неотрицательны (условие (б)), и по условию НЕРОН имеем

$\sum_{i=0}^{m} \lambda_i > 0$. Разделив все множители на их сумму, можем считать, что $\sum_{i=0}^{m} \lambda_i = 1$. Условие (a) равносильно условию $0 \in \partial \mathcal{L}_x(\hat{x}, \lambda)$ (следствие 6.1), которое, в силу теоремы Моро-Рокафеллара (теорема 6.2), означает, что $0 \in \sum_{i \in I(\hat{x})} \lambda_i \, \partial f_i(\hat{x})$. Это значит, что найдутся такие точки $z_i \in \partial f_i(\hat{x})$, $i \in I(\hat{x})$, что $\sum_{i \in I(\hat{x})} \lambda_i \, z_i = 0$, т.е., точка $\{0\}$ является выпуклой комбинацией точек $\{z_i\}$. А это, в свою очередь, означает, что $0 \in \mathbf{co}\left\{ \partial f_i(\hat{x}),\ i \in I(\hat{x}) \right\}$. Таким образом, доказана

Теорема 6.4 *В теореме Каруша-Куна-Таккера (теорема 4.3) условия (а) – (в) равносильны следующему условию: выпуклая оболочка субдифференциалов целевой функции f_0 и активных неравенств f_i в точке $\hat{x}$ содержит ноль.*

7 Лекция

24 марта 2009

Теоремы Радона и Хелли. Теорема об очистке и приложения. Теорема о минимаксе.

Теоремы Радона и Хелли. Теорема об очистке.
Приближения функций полиномами. Теорема Чебышова об альтернансе.
Теорема Какутани о неподвижной точке. Теорема о минимаксе.

Теорема 7.1 *(Радон) Конечное подмножество пространства $\mathbb{R}^d$, содержащее не менее $d+2$ точек, может быть разбито на два непересекающихся непустых множества, выпуклые оболочки которых пересекаются.*

`Доказательство`. Пусть x_i, $i = 1, \ldots, n$ – данные точки $n \geq d+2$. Рассмотрим векторы $x_2 - x_1, \ldots, x_n - x_1$. Их не менее $d+1$, поэтому они линейно зависимы: $\sum_{i=2}^{n} \alpha_i (x_i - x_1) = 0$ для некоторых НЕРОН множителей $\alpha_2, \ldots, \alpha_n$. Раскрывая скобки и обозначая $\beta_1 = -\sum_{i=2}^{n} \alpha_i$ и $\beta_i = \alpha_i$ при $i = 2, \ldots, n$, получаем:

$$\sum_{i=1}^{n} \beta_i x_i = 0, \qquad \sum_{i=1}^{n} \beta_i = 0,$$

причём множители β_i – НЕРОН. Поменяем нумерацию так, что $\beta_i \geq 0$ при $i = \{1, \ldots, m\}$ и $\beta_i < 0$ при $i = \{m+1, \ldots, n\}$. Тогда $\beta = \sum_{i \leq s} \beta_i = -\sum_{i > s} \beta_i$. Положим $B_1 = \{x_1, \ldots, x_m\}$ и $B_2 = \{x_{m+1}, \ldots, x_n\}$. Так как $\sum_{i=1}^{n} \beta_i x_i = 0$, то $\sum_{i=1}^{m} t_i x_i = \sum_{i=m+1}^{n} t_i x_i$, где $t_i = \frac{\beta_i}{\beta}$ при $i \leq m$ и $t_i = -\frac{\beta_i}{\beta}$ при $i > m$. Первая выпуклая комбинация принадлежит $\mathbf{co}\,(B_1)$, а вторая – $\mathbf{co}\,(B_2)$, поэтому $\mathbf{co}\,(B_1) \cap \mathbf{co}\,(B_2) \neq \emptyset$.

□

Теорема 7.2 *(Хелли).*

а) В пространстве $\mathbb{R}^d$ дано $n \geq d+1$ выпуклых множеств, любые $d+1$ из которых имеют общую точку. Тогда и все эти множества имеют общую точку.

б) Если в $\mathbb{R}^d$ дано произвольное бесконечное семейство компактных выпуклых множеств, любые $d+1$ из которых имеют общую точку, то все эти множества имеют общую точку.

В доказательстве пункта б) нам понадобится следующий факт, который мы примем без доказательства.

Лемма 7.1 *(о центрированной системе). Если семейство компактов в полном метрическом пространстве таково, что любое конечное подсемейство имеет непустое пересечение, то и все семейство имеет непустое пересечение.*

`Доказательство теоремы 7.2.` а) Индукция по числу множеств n. Для $n = d+1$ доказывать нечего. Пусть $n \geq d+2$, и теорема верна для любых $n-1$ выпуклых множеств. Обозначив данные множества как $A_j,\ j = 1, \ldots, n$, для каждого j возьмём произвольную точку $y_i \in \bigcap_{j \neq i} A_j$ (пересечение не пусто по предположению индукции). Поскольку $n \geq d+2$, применим теорему Радона к точкам $\{y_i\}_{i=1}^n$. С возможной перенумерацией точек, получаем, что существует натуральное число $m \leq n-1$, для которого множества $\mathbf{co}\,\{y_1, \ldots, y_m\}$ и $\mathbf{co}\,\{y_{m+1}, \ldots, y_n\}$ имеют общую точку y. Для любого $i = \{1, \ldots, m\}$ точка y_i принадлежит каждому из множеств $A_{m+1}, \ldots, A_n$, значит, в силу выпуклости этих множеств, каждое из них содержит $\mathbf{co}\,\{y_1, \ldots, y_m\}$, и следовательно, каждое из них содержит точку y. Итак, $y \in A_j$ для любого $j = m+1, \ldots, n$. Аналогично показываем, что $y \in A_j$ для $j = 1, \ldots, m$. Таким образом, $y \in \bigcap_{j=1,\ldots,n} A_j$.

б) Если все данные множества компактны, то, применяя пункт (а), получаем, что любое конечное подсемейство имеет непустое пересечение, а значит, в силу леммы о центрированной системе (лемма 7.1) все множества семейства пересекаются.

□

Упражнение 10 Несколько полупространств покрывают всё пространство $\mathbb{R}^d$. Докажите, что из них можно выбрать $d+1$ подпространство, которые также покрывают всё пространство.

Упражнение 11 На прямой дано N точек $x_1, \ldots, x_N$, даны N чисел $y_1, \ldots, y_N$, а также заданы число $\varepsilon > 0$ и натуральное $n \leq N-2$. Необходимо определить, существует ли алгебраический полином $p(x)$ степени не выше n, для которого $\big|p(x_k) - y_k\big| < \varepsilon$ для всех $k = 1, \ldots, N$. Докажите, что искомый полином существует тогда и только тогда когда он существует для любых $n+2$ точек данного набора.

`Указание.` Каждому полиному поставим в соответствие точку пространства $\mathbb{R}^{n+1}$, координаты которой – коэффициенты полинома. Множество точек пространства $\mathbb{R}^{n+1}$, для которых $\big|p(x_k) - y_k\big| < \varepsilon$ – выпукло.

Теорема Хелли имеет много важных следствий. Одно из них – теорема об очистке. Эта теорема имеет дело с минимумами выпуклых функций, которые заданы как поточечный максимум выпуклых функций.

Теорема 7.3 *(об очистке) Пусть T – метрический компакт, функция $f : T \times \mathbb{R}^d :\to \mathbb{R}$ такова, что для для каждого $t \in T$ функция $f(t, \cdot)$ выпукла, и для каждого $x \in \mathbb{R}^d$ функция $f(\cdot, x)$ непрерывна. Тогда существуют такие точки $t_1, \ldots, t_n$, где $n \leq d+1$, что*

$$\inf_{x \in \mathbb{R}^d} \left(\max_{t \in T} f(t, x) \right) = \inf_{x \in \mathbb{R}^d} \left(\max_{i=1,\ldots,n} f(t_i, x) \right). \tag{7.5}$$

Заметим, что любая функция на конечном множестве является непрерывной. Поэтому, если множество T конечно, то функция $f(t,x)$ всегда непрерывна по t. В этом случае теорема об очистке формулируется так: из любого конечного множества выпуклых функций $f_k(x)$, $k=1,\dots,N$ можно выбрать не более $d+1$ функции $f_{k_1},\dots,f_{k_n}$ таким образом, что $\min_{x\in\mathbb{R}^d}\max_{k=1,\dots,N} f_k(x) = \min_{x\in\mathbb{R}^d}\max_{i=1,\dots,n} f_{k_i}(x)$.

`Доказательство теоремы 7.3.` Обозначим левую часть равенства (7.5) через m. Рассмотрим сначала случай, когда множество T конечно. Для каждого $\varepsilon>0$ положим $A(t,\varepsilon) = \bigl\{x\in\mathbb{R}^d \bigm| f(t,x)\le m-\varepsilon\bigr\}$. Это множество выпукло. Если для любых $n\le d+1$ точек множества T соответствующие множества $A(t_i,\varepsilon)$ имеют общую точку, то, согласно пункту (а) теоремы Хелли (теорема 7.2) все множества $A(t,\varepsilon)$, $t\in T$ имеют некоторую общую точку x. В этой точке $f(t,x)\le m-\varepsilon$ для всех $t\in T$, а значит $\max_{t\in T} f(t,x)\le m-\varepsilon$, что невозможно, поскольку $\inf_{x\in\mathbb{R}^d}\max_{t\in T} f(t,x)=m$. Итак, для любого $\varepsilon>0$ найдутся не более $d+1$ точек $t_1,\dots,t_n$, для которых множества $A(t_i,\varepsilon)$, $i=1,\dots,n$ не пересекаются. Следовательно, $\inf_{x\in\mathbb{R}^d}\max_{i=1,\dots,n} f(t_i,x)>m-\varepsilon$. В силу произвольности ε это доказывает равенство (7.5).

Для доказательства теоремы в общем случае применяем тот же подход, при этом используем пункт (б) теоремы Хелли. Здесь, однако, нужно обойти одну сложность: множества $A(t,\varepsilon)$ могут быть не компактны. Поступим так. Каждому набору $\bar t$ точек $t_1,\dots,t_n$ поставим в соответствие точку $x(\bar t)\in\mathbb{R}^d$, в которой $\max_{i=1,\dots,n} f\bigl(t_i,x(\bar t)\bigr) < m-\varepsilon$. Если для некоторого $\varepsilon>0$ такой точки не нашлось, то для соответствующего набора точек $t_1,\dots,t_n$ разность между левой и правой частями (7.5) не превосходит ε. Если при этом ε можно взять сколь угодно малым, то вновь получаем, что равенство (7.5) доказано. В противном случае, для любого достаточно малого $\varepsilon>0$ для любого набора $\bar t$ существует точка $x(\bar t)$. Рассмотрим множество $V(\bar t)$, которое состоит из всех наборов $(t_1',\dots,t_n')$, для которых $\max_{i=1,\dots,n} f\bigl(t_i',x(\bar t)\bigr) < m-\varepsilon$. В силу непрерывности функции f по t, множество $V(\bar t)$ открыто и содержит $\bar t$. Следовательно, множества $\{V(\bar t)\,,\ \bar t\in T^n\}$ образуют открытое покрытие компактного множества T^n. Значит, из них можно выбрать конечное подпокрытие, отвечающее наборам $\bar t^1,\dots,\bar t^k$. Пусть $\tilde A = \mathbf{co}\{x(\bar t^1),\dots,x(\bar t^k)\}$. Это – выпуклый компакт. Для каждой точки x этого множества найдётся набор $\bar t$ такой, что $\max_{i=1,\dots,n} f\bigl(t_i,x(\bar t)\bigr) < m-\varepsilon$. Теперь положим $\tilde A(t,\varepsilon) = A(t,\varepsilon)\cap\tilde A$, $t\in T$. Для любого t это – выпуклый компакт. Применяя теперь пункт (б) теоремы Хелли к множествам $\tilde A(t,\varepsilon)$, $t\in T$, завершаем доказательство так же, как в предыдущем случае.

$\square$

Теорема об очистке играет исключительную роль а теории приближений. Она говорит о том, что для приближения функции на отрезке линейными комбинациями нескольких функций (например – многочленами), достаточно приблизить её в конечном числе точек этого отрезка.

Определение 7.1 *Дана функция $h\in C[a,b]$. Набор точек $\{t_i\}_{i=1}^k$, $a\le t_1<t_2<\dots<t_k\le b$ называется альтернансом этой функции, если $|h(t_i)| = \|h\|_{C[a,b]}$ для всех i, и знаки чисел $h(t_i)$ чередуются.*

Пусть $\mathcal{P}_n$ – пространство алгебраических полиномов степени не выше n. Это – линейно пространство размерности $n+1$. Для данной функции $g\in C[a,b]$ рассмотрим задачу

$$\begin{cases} \|\,g-p\,\| & \to \quad \min \\ p\in\mathcal{P}_n\,. \end{cases}$$

Так как функция $F(p) = \|g-p\|$ непрерывна и коэцитивна на конечномерном пространстве $\mathcal{P}_n$, она достигает своего минимума. Точка минимума p – *многочлен наилучшего приближения* для функции g на отрезке $[a,b]$ в метрике пространства C.

Теорема 7.4 *(Чебышов) Многочлен $\hat{p} \in P_n$ является многочленом наилучшего приближения на отрезке $[a,b]$ тогда и только тогда когда функция $h = g - \hat{p}$ имеет на этом отрезке альтернанс из $n+2$ точек.*

Таким образом, чтобы приблизить функцию многочленом степени n нужно приблизить её на подходящем множестве из $n+2$ точек.

Доказательство. Пусть функция h имеет альтернанс $t_1, \dots, t_{n+2}$. Пусть $\|h\| = \alpha$ и $h(t_k) = (-1)^k \alpha$. Предположим, что многочлен p приближает g на отрезке $[a,b]$ лучше, чем $\hat{p}$. Тогда величина $p(t_k) - \hat{p}(t_k)$ положительна при нечётных k и отрицательна при чётных k. Следовательно, разность $p - \hat{p}$ принимает на концах отрезка $[t_k, t_{k+1}]$ разные знаки, значит она обращается в ноль на каждом интервале (t_k, t_{k+1}), $k = 1, \dots, n+1$, т.е., имеет не менее $n+1$ нулей. Но так как она является полиномом степени n, то $p - \hat{p} \equiv 0$, а значит $p = \hat{p}$.

Обратно, докажем, что если $\hat{p}$ – многочлен наилучшего приближения, то функция h имеет альтернанс. Положим $f(t,p) = |g(t) - p(t)|$, $t \in [a,b]$, $p \in \mathcal{P}_n$. Эта функция непрерывна по t и выпукла по p. Применим к ней теорему об очистке (теорема 7.3). Существует такой набор точек $t_1, \dots, t_m$, где $m \le n+2$, для которого

$$\min_{p\in\mathcal{P}_n} \max_{i=1,\dots,m} f(t_i,p) = \min_{p\in\mathcal{P}_n} \max_{t\in[a,b]} f(t,p) = \min_{p\in\mathcal{P}_n} \big\| g - p \big\|_{C[a,b]} = \big\| g - \hat{p} \big\|_{C[a,b]} \qquad (7.6)$$

Докажем, что на самом деле $m = n+2$ и данный набор точек образует альтернанс. Ясно, что

$$\min_{p\in\mathcal{P}_n} \max_{i=1,\dots,m} f(t_i,p) \le \max_{i=1,\dots,m} \big| g(t_i) - \hat{p}(t_i) \big| \le \big\| g - \hat{p} \big\|_{C[a,b]}.$$

Следовательно, $\max_{i=1,\dots,m} \big| g(t_i) - \hat{p}(t_i) \big| = \big\| g - \hat{p} \big\|_{C[a,b]}$. Обозначив эту величину через α, получаем, что $|f(t_i,\hat{p})| = |g(t_i) - \hat{p}(t_i)| \le \alpha$ для всех i. Выбросив все точки, в которых данная величина меньше α, считаем, что $|f(t_i,\hat{p})| = \alpha$ для всех $i = 1, \dots, m$. Назовём точку t_i положительной, если $f(t_i,\hat{p}) = \alpha$ и отрицательной, если $f(t_i,\hat{p}) = -\alpha$. Пусть, для определённости, точка t_1 “положительна” и пусть t_q – наименьшая “отрицательная” точка. Возьмём произвольную точку $s_1 \in (t_{q-1}, t_q)$. Далее берем $s_2 \in (t_{r-1}, t_r)$, где t_r – наименьшая положительная точка, превосходящая t_k, и т.д. Получим последовательность точек $a = s_0 < s_1 < \ldots < s_k < s_{k+1} = b$. На каждом из отрезков $[s_j, s_{j+1}]$ лежат точки t_i одного знака, и эти знаки чередуются на отрезках. Заметим, что если точки $\{t_i\}_{i=1}^m$ не составляли альтернанс (т.е., если либо $m \le n+1$, либо в последовательности $\{t_i\}$ есть две точки подряд одного знака), то $k \le n$. Полином $q(t) = -(s_1 - t)\cdots(s_k - t)$ положителен во всех “отрицательных” точках, и отрицателен во всех “положительных”. Следовательно, при достаточно малых $\nu > 0$ имеем $|f(t_i, \hat{p} - \alpha q)| = |g(t_i) - \hat{p}(t_i) + \nu\, q(t_i)| < \nu$ для всех $i = 1, \dots, m$. В силу (7.6) это влечёт $\big\| g - \hat{p} + \nu\, q \big\| < \alpha$. Значит, полином $\hat{p} - \alpha q$ приближает функцию g лучше, чем полином $\hat{p}$, что противоречит условию.

$\square$

Упражнение 12 Найдите многочлен наилучшего приближения степени $n = 2$ для функции $f(x) = \sin \pi x$ на отрезке $[-1,1]$.

Упражнение 13 Найдите многочлен наилучшего приближения степени 2 для функции $f(x) = |x|$ на отрезке $[-1,1]$.

Теорема об очистке связана еще с одним важнейшим результатом выпуклого анализа – теоремой о минимаксе. Для её формулировки нам понадобится понятие полунепрерывной функции. Функция f, заданная на метрическом пространстве A называется *полунепрерывной сверху* в точке $a \in A$, если $\overline{\lim}_{x\to a} f(x) = f(a)$. Функция называется *полунепрерывной снизу*, если $\underline{\lim}_{x\to a} f(x) = f(a)$. Функция полунепрерывна сверху (снизу) на множестве, если она полунепрерывна сверху (соответственно, снизу) в каждой точке этого множества. Непрерывная функция является также полунепрерывной и сверху и снизу, но, вообще говоря, не наоборот. Так, функция $f(x) = 1,\ x \in (0,1]\,,\ f(0) = 0$ является полунепрерывной снизу, а функция $g(x) = 0,\ x \in (0,1]\,,\ g(0) = 1$ является полунепрерывной сверху на отрезке $[0,1]$. Но обе они разрывны в точке 0. Функция $f(x) = [x]$ (целая часть числа x) полунепрерывна сверху. Можно дать равносильное определение: функция является полунепрерывной сверху (снизу), если все её множества уровня $f(x) \ge b$ (соответственно $f(x) \le b$) замкнуты.

Полунепрерывные функции обладают двумя важнейшими свойствами:

1) Если функции $f_k(x)$ полунепрерывны сверху на компакте K, то функция $f(x) = \inf\limits_{k\in\mathbb{N}} f_k(x)$ полунепрерывна сверху.

2) Полунепрерывная сверху функция достигает своего максимального значения на компакте.

Так же формулируются свойства полунепрерывных снизу функций, с заменой inf на sup и минимума на максимум. Доказательства этих свойств мы оставляем в виде легких упражнений читателю. Функция $F(x)$ называется вогнутой, если функция $-F(x)$ выпукла. Для максимумов вогнутых функций выполнены те же свойства, как и для минимумов выпуклых.

Теорема 7.5 *(о минимаксе) Если функция $f(t,x)$, заданная не произведении конечномерных компактов $T \times X$, непрерывна по совокупности аргументов, выпукла по x при каждом фиксированном t и вогнута по t при каждом фиксированном x, то*

$$\min_{x\in X}\max_{t\in T} f(t,x) \quad = \quad \max_{t\in T}\min_{x\in X} f(t,x)\,. \tag{7.7}$$

Перед доказательством теоремы сделаем несколько замечаний. Во-первых, функция $\max_{t\in T} f(t,x)$ полунепрерывна снизу, как поточечный супремум непрерывных функций. Поэтому она достигает своего минимума на компакте. Поэтому мы ставим в левой части равенства min вместо inf. То же – со знаком max в правой части равенства.

Для любой функции двух переменных $f(t,x)$ без всяких предположений о её непрерывности и о компактности множеств T и X выполнено неравенство между минимумом от максимума и максимумом от минимума:

$$\inf_{x\in X}\sup_{t\in T} f(t,x) \quad \ge \quad \sup_{t\in T}\inf_{x\in X} f(t,x) \tag{7.8}$$

(здесь мы уже ставим inf и sup, поскольку минимальные и максимальные значения могут не достигаться.) Для доказательства заметим, что для любой пары фиксированных $\tilde{t}, \tilde{x}$ имеем $\sup_{t\in T} f(t,\tilde{x}) \ \ge\ f(\tilde{t},\tilde{x}) \ \ge\ \inf_{x\in X} f(\tilde{t},x)$. Взяв инфимум по $\tilde{x}$ от левой части неравенства и супремум правой части по $\tilde{t}$, приходим к (7.8). Это неравенство, вообще говоря, строгое. Разность между левой и правой частями обычно называют *дефектом двойственности* (duality gap). Оказывается, при условиях теоремы 7.5 дефект равен нулю. Эту теорему наиболее естественно доказывать с помощью теоремы об очистке. Но мы выберем более короткий путь, основанный на применении теоремы Какутани. Вначале сформулируем нужные нам утверждения.

Теорема 7.6 *(Теорема Брауэра о неподвижной точке). Любое непрерывное отображение выпуклого компактного подмножества пространства $\mathbb{R}^d$ в себя имеет неподвижную точку.*

Теорема 7.7 *(Какутани). Каждой точке x выпуклого компакта $B \subset \mathbb{R}^d$ поставлен в соответствие выпуклый компакт $\Phi(x) \subset B$. При этом, если $x_i \to x$ и $y_i \to y$ при $i \to \infty$, и $y_i \in \Phi(x_i)$ при всех i, то $y \in \Phi(x)$ (предел образов содержится в образе предела). Тогда отображение Φ имеет неподвижную точку x, для которой $x \in \Phi(x)$.*

`Доказательство.` Поместим B внутрь d-мерного симплекса Δ. Для каждой точки $x \in \Delta$ обозначим через $\psi(x)$ ближайшую к x точку множества B. Таким образом, $\psi(x) = x$ только при $x \in B$. Положим также $\Phi(x) = \Phi(\psi(x))$ для $x \in \Delta$. Непрерывность и выпуклость образов сохранится. Разобьём Δ на маленькие симплексы. Если v – вершина одного из них, то выберем точку $v' \in \Phi(v)$ и положим $\varphi_1(v) = v'$. Определим таким образом отображение φ_1 во всех вершинах симплексов, после чего продолжим его его внутрь каждого симплекса по линейности. Получили непрерывное отображение симплекса Δ в себя. По теореме Брауэра оно имеет неподвижную точку $x_1 = \sum_{i=0}^{d} \lambda_{1,i}\, x_{1,i}$, где $x_{1,i}$ – вершины симплекса, содержащего точку x_1. При этом $x_1 = \varphi_1(x_1) = \sum_{i=0}^{d} \lambda_{1,i}\, \varphi(x_{1,i})$. Взяв более мелкое разбиение, получим отображение φ_2 с неподвижной точкой $x_2 = \sum_{i=0}^{d} \lambda_{2,i}\, x_{2,i}$. Измельчая разбиение далее, получаем точки $x_3, x_4, \ldots$. Из соображений компактности, переходя к подпоследовательности, считаем что $x_k \to x$ при $k \to \infty$, а значит $x_{k,0}, \ldots, x_{k,d}$ сходятся к x, и что коэффициенты $\lambda_{k,0}, \ldots, \lambda_{k,d}$ стремятся к некоторым $\lambda_0, \ldots, \lambda_d \geq 0$, $\sum_{i=0}^{d} \lambda_i = 1$, точки $\varphi_k(x_{k,0}), \ldots \varphi_k(x_{k,d})$ стремятся к некоторым $y_0, \ldots, y_d$. Тогда $x = \sum_{i=0}^{d} \lambda_i\, y_i$. Из свойства отображения Φ следует, что $y_0, \ldots, y_d \in \Phi(x)$. Из выпуклости $\Phi(x)$ следует, что $x \in \Phi(x)$.

□

`Доказательство теоремы 7.5 о минимаксе.` Для любых $t \in T$, $x \in X$ положим

$$B_t = \left\{ y \in X \;\middle|\; f(t,y) = \min_{y' \in X} f(t,y') \right\}; \qquad A_x = \left\{ s \in T \;\middle|\; f(s,x) = \max_{s' \in T} f(s',x) \right\}.$$

Определим также $\Phi(t,x) = A_x \times B_t$. По теореме 7.7 найдётся неподвижная точка $(\hat{t}, \hat{x})$, для которой $(\hat{t}, \hat{x}) \in \Phi(\hat{t}, \hat{x})$, т.е., $\hat{t} \in A_{\hat{x}}$ и $\hat{x} \in B_{\hat{t}}$, и следовательно, обе части равенства (7.7) равны $f(\hat{t}, \hat{x})$.

□

8 Лекция

31 марта 2009

Двойственность. Сопряженные функции и поляры.

Двойственность. Преобразование Лежандра-Юнга-Фенхеля.
Теорема Фенхеля-Моро. Полярное преобразование.

Определение 8.1 *Линейные пространства X и Y называются пространствами в двойственности (ПВД), если определена билинейная форма $\langle .\,,\,. \rangle : X \times Y \to \mathbb{R}$ такая, что если $\langle x, y\rangle = 0$ для всех $y \in Y$, то $x = 0$, а если $\langle x, y\rangle = 0$ для всех $x \in X$, то $y = 0$.*

Пусть $\dim X = d < \infty$, определим отображение $\pi : Y \to X^*$ такое, что $\pi(y)[x] = \langle x, y\rangle$. Данное отображение линейно и инъективно. Поэтому $\dim Y \le \dim X^* = d$ (поскольку $\mathrm{dimX}^* = \dim X = d$). А так как Y конечномерно, то $\mathrm{dimX} \le \dim Y$. Аналогично, $\dim Y \le \dim X$. Итак, $\dim X = \dim Y$. Таким образом, если одно из пространств в двойственности конечномерно, тогда и второе конечномерно, а их размерности равны. Пусть $X = Y = \mathbb{R}^d$. Обозначив $a_{ij} = \langle e_i, e_j\rangle$, где $e_1, \dots, e_d$ – базис в $\mathbb{R}^d$, получаем, что отношение двойственности между пространствами X и Y задаётся матрицей $A = (a_{ij})$ по формуле $\langle x, y\rangle = (x, Ay)$. Эта матрица невырождена, иначе существовал бы ненулевой элемент y, для которого $Ay = 0$, а значит $\langle x, y\rangle = 0$ для любого x. Обратно, каждая невырожденная матрица определяет двойственность по формуле $\langle x, y\rangle = (x, Ay)$.

Следствие 8.1 *Если одно из пространств в двойственности конечномерно, то и второе конечномерно, а их размерности равны. При этом отношение двойственности задается невырожденной $d \times d$-матрицей A по формуле $\langle x, y\rangle = (x, Ay)$, где $d = \dim X = \dim Y$. Обратно, каждая невырожденная матрица задаёт отношение двойственности.*

Пример 8.1 Если X – нормированное пространство, $Y = X^*$, то форма $\langle x, y\rangle = (x, y)$ определяет отношение двойственности. В самом деле, если $(x, y) = 0$ для любого $x \in X$, то по определению линейного функционала $y = 0$. Если же $(x, y) = 0$ для любого $y \in X^*$ то $x = 0$. В противном случае точку x можно было бы строго отделить от нуля (теорема 2.2) некоторым функционалом $y \in X^*$, а значит $(x, y) \ne 0$.

В пространствах в двойственности определяется топология $\sigma(X, Y)$ следующим образом: $x_k \to x$ если $\langle x, y\rangle \to 0$ для любого $y \in Y$. Легко проверить, что для любого $y \in Y$ отображение $x \to \langle x, y\rangle$. непрерывно в топологии $\sigma(X, Y)$. Так, в примере 8.1 топология $\sigma(X, X^*)$ на пространстве X совпадает с со слабой топологией, а на пространстве X^* – со слабой *-топологией.

Предложение 8.1 *Если множество $A \subset X$ выпукло, то его замкнутость в сильной топологии X равносильна замкнутости в слабой топологии $\sigma(X, X^*)$.*

`Доказательство.` Последовательность, сходящаяся в сильной топологии, сходится и в слабой. Поэтому замкнутость в слабой топологии влечет замкнутость в сильной (и без условия выпуклости). Обратно, пусть A замкнуто в сильной топологии. Надо доказать, что произвольная слабо сходящаяся последовательность $x_k \in A$ имеет предел в множестве A. Если её предел x не принадлежит A, то его можно строго отделить от выпуклого замкнутого множества A (теорема 2.2) некоторым функционалом $y \in X^*$. Следовательно, (y, x_k) не стремится к (y, x) при $k \to \infty$, что противоречит слабой сходимости последовательности x_k.

$\square$

Для пространств в двойственности верно большинство утверждений, которые мы доказывали для нормированных пространств и их сопряженных. В частности, верна и теорема отделимости, которую мы не будем доказывать.

Теорема 8.1 *(отделимости для ПВД). Пусть X и Y - пространства в двойственности, $A \subset X$ – выпуклое $(\sigma(X,Y))$-замкнутое непустое множество. Тогда любая точка $\hat{x} \notin A$ строго отделима от A, т.е., существует $y \in Y$ такой, что $\sup\limits_{x\in A}\langle x, y\rangle < \langle \hat{x}, y\rangle$.*

Теперь мы определим двойственные объекты выпуклого анализа – сопряженные функции и сопряженные множества. Начнём с сопряженных функций.

Определение 8.2 *Пусть (X, Y) – ПВД. Для любой функции $f : X \to \bar{\mathbb{R}}$ сопряженной к ней функцией называется функция $f^* : Y \to \bar{\mathbb{R}}$, определяемая по формуле*

$$f^*(y) \quad = \quad \sup_{x\in X} \big(\langle x \, , \, y\rangle \; - \; f(x) \big) \, . \tag{8.9}$$

Преобразование $f \to f^*$ называется преобразованием Лежандра-Юнга-Фенхеля (ЛЮФ) функции f. Перед исследованием этого преобразования определим некоторые понятия. Рассмотрим пространства $(X \times \mathbb{R}, Y \times \mathbb{R})$. Они также являются пространствами в двойственности с билинейной формой $\langle\, (x, u)\, ,\, (y, v)\, \rangle \; = \; \langle x, y\rangle \, + \, uv$. В частности, если $X^* = Y$, то $Y \times \mathbb{R}$ сопряжено с $X \times \mathbb{R}$. Топология в пространстве $X \times \mathbb{R}$ задаётся через сходящиеся последовательности: $(x_k, t_k) \; \to \; (x, t)$, если $\langle x_k, y\rangle \; \to \; \langle x, y\rangle$ для любого $y \in Y$ и $t_k \; \to \; t$ при $k \to \infty$.

Определение 8.3 *Функция $f : X \to \bar{\mathbb{R}}$ называется замкнутой, если* **epi** *f замкнут в $X \times \mathbb{R}$.*

Предложение 8.2 *Функция является замкнутой тогда и только тогда когда она полунепрерывна снизу.*

`Доказательство.` Если f замкнута, то для любого $b \in \mathbb{R}$ множество уровня $\{x \in X \mid f(x) \le b\}$ замкнуто, поскольку оно является проекцией на X замкнутого множества $\mathbf{epi}\, f \,\cap\, \{\,(x,t) \in X \times \bar{\mathbb{R}} \mid t \le b\,\}$. Следовательно, f полунепрерывна снизу. Обратно, пусть f полунепрерывна снизу. Рассмотрим произвольную последовательность (x_k, t_k) элементов надграфика $\mathbf{epi}\, f$, сходящаяся к некоторой точке (x, t). Так как $f(x_k) \le t_k$ для любого k и $x_k \to x, t_k \to t$ при $k \to \infty$, то $f(x) \le \underline{\lim}_{k\to\infty} f(x_k)$ (полунепрерывность снизу!), откуда $f(x) \le \underline{\lim}_{k\to\infty} t_k \; = \; t$. Следовательно, $(x, t) \in \mathbf{epi}\, f$, что доказывает замкнутость f.

$\square$

Следствие 8.2 *Любая непрерывная функция $f : X \to \bar{\mathbb{R}}$ замкнута.*

Следствие 8.3 *Поточечный супремум замкнутых функций замкнут.*

Предложение 8.3 *Для любой функции f сопряженная функция f^* выпукла и замкнута.*

`Доказательство.` Для любого $x \in X$ функция $\varphi(y) = \langle x\,,\,y\rangle$ выпукла (поскольку она линейна) и замкнута (поскольку она непрерывна), значит выпукла и замкнута функция $\varphi(y) = \langle x\,,\,y\rangle - f(x)$, и следовательно функция $f^*(y)$, будучи супремумом таких функций по всем $x \in X$ (формула (8.9)) также выпукла и замкнута.

□

Для любых x, y выполнено *неравенство Юнга*:

$$f^*(y) + f(x) \quad \geq \quad \langle x\,,\,y\rangle\,. \tag{8.10}$$

В самом деле, $f^*(y) = \sup_{s\in X}\bigl(\,\langle s, y\rangle - f(s)\,\bigr) \geq \langle x, y\rangle - f(x)$.

Для повторного преобразования Лежандра имеем:

$$f^{**}(x) \quad = \quad \sup_{y\in Y}\bigl(\,\langle x\,,\,y\rangle - f^*(y)\,\bigr) \leq \sup_{y\in Y}\bigl(\,\langle x\,,\,y\rangle - \bigl(\,\langle x\,,\,y\rangle - f(x)\,\bigr) = \sup_{y\in Y} f(x) = f(x)\,.$$

Таким образом, получаем.

Предложение 8.4 *Для любой функции f имеем $f^{**} \leq f$.*

Возникает вопрос, для каких функций выполнено равенство? Согласно предложению 8.3, функция f^{**} выпукла и замкнута. Поэтому, функция f должна быть, как минимум, выпукла и замкнута. Оказывается, этих двух свойств достаточно.

Теорема 8.2 *(Фенхель, Моро). Если $f : X \to \bar{\mathbb{R}}$ – собственная выпуклая замкнутая функция, то $f^{**} = f$.*

Напомним, что функция называется собственной, если она не равна $-\infty$ ни в одной точке, и не равна $+\infty$ тождественно. Доказательству теоремы предпошлем лемму.

Лемма 8.1 *Если функция f – выпуклая, замкнутая и собственная, то и f^* – также собственная.*

`Доказательство.` Если $f^*(y) = -\infty$, то $\langle x, y\rangle - f(x) = -\infty$ при любом x, а значит $f \equiv +\infty$, что невозможно, поскольку f – собственная. Осталось установить существование $y \in Y$, для которого $f^*(y) < +\infty$. Возьмём $\hat{x} \in X$, для которого $f(\hat{x}) < +\infty$, и возьмем произвольное $u < f(\hat{x})$. Поскольку $(\hat{x}, u) \notin \mathbf{epi}\, f$, эту точку можно строго отделить от выпуклого замкнутого множества $\mathbf{epi}\, f$ (теорема 8.1). Это значит, что для некоторой пары $(y, \lambda) \in Y \times \mathbb{R}$ имеем

$$\langle \hat{x}\,, y\,\rangle + \lambda\, u \quad > \quad \sup_{(x,v)\in \mathbf{epi}\, f}\bigl(\langle x\,,\,y\rangle + \lambda\, v\,\bigr)\,.$$

Подставив $v = f(x)$, получим

$$\langle \hat{x}\,, y\,\rangle + \lambda\, u \quad > \quad \sup_{x\in \mathbf{dom}\, f}\bigl(\langle x\,,\,y\rangle + \lambda\, f(x)\,\bigr)\,.$$

При $x = \hat{x}$ имеем $\lambda\, u > \lambda f(\hat{x})$, откуда $\lambda < 0$. Далее

$$f^*\Bigl(-\frac{y}{\lambda}\Bigr) = \sup_{x\in X}\Bigl(\Bigl\langle x\,, -\frac{y}{\lambda}\Bigr\rangle - f(x)\Bigr) = \sup_{x\in \mathbf{dom}\, f}\Bigl(\Bigl\langle x\,, -\frac{y}{\lambda}\Bigr\rangle - f(x)\Bigr) =$$

$$= -\frac{1}{\lambda} \sup_{x \in \mathbf{dom}\, f} \Big(\langle x\,,\, y\,\rangle \;+\; \lambda\, f(x) \Big) \;<\; -\frac{1}{\lambda} \Big(\langle \hat{x}\,, y\,\rangle \;+\; \lambda\, u \Big) \;<\; +\infty\,,$$

следовательно, $f^*\big(-\frac{y}{\lambda}\big) \;<\; +\infty$.

$\square$

`Доказательство теоремы 8.2` – от противного. Пусть $f^{**}(\hat{x}) < f(\hat{x})$. Положим $u = f^{**}(\hat{x})$ и, воспользовавшись теоремой 8.1, строго отделим точку $(\hat{x}, u)$ от выпуклого замкнутого множества $\mathbf{epi}\, f$. Для некоторой пары $(y, \lambda) \in Y \times \mathbb{R}$ имеем

$$\langle \hat{x}\,, y\,\rangle \;+\; \lambda\, u \quad > \quad \sup_{(x,v) \in \mathbf{epi}\, f} \big(\langle x\,,\, y\rangle \;+\; \lambda\, v \big)\,.$$

Ясно, что $\lambda \le 0$, иначе, не меняя x, устремляем $v \to +\infty$ и получаем противоречие. Если $\lambda = 0$, то $\langle \hat{x}, y\rangle \;>\; \langle x, y\rangle$ для всех $x \in \mathbf{dom}\, f$. В силу леммы 8.1 существует $y_1 \in Y$ такой, что $f^*(y_1) < +\infty$. Для произвольного $t > 0$ имеем

$$f^*\big(y_1 + t y\,\big) \;=\; \sup_{x \in \mathbf{dom}\, f} \big(\langle x\,,\, y_1\rangle + t\,\langle x\,,\, y\rangle - f(x) \,\big) \;\le\; \sup_{x \in \mathbf{dom}\, f} \big(\langle x\,,\, y_1\rangle - f(x) \,\big) + \sup_{x \in \mathbf{dom}\, f} t\,\langle x\,,\, y\rangle \;=$$

$$=\; f^*(y_1) \;+\; t \sup_{x \in \mathbf{dom}\, f} \langle x\,,\, y\rangle\,.$$

Далее,

$$f^{**}(\hat{x}) \;\ge\; \sup_{t>0} \big(\langle \hat{x}\,,\, y_1 \,+\, t\,y\rangle \;-\; f^*(y_1 + t y) \big) \;\ge\; \big(\langle \hat{x}\,,\, y_1 \,+\, t\,y\rangle \;-\; f^*(y_1 \,+\, t\,y) \big) \;\ge$$

$$\ge \big(\langle \hat{x}\,,\, y_1\rangle \;+\; t\,\langle \hat{x}\,,\, y\,\rangle \;-\; f^*(y_1) \;-\; t \sup_{x \in \mathbf{dom}\, f} \langle x\,,\, y\rangle \big) \;=$$

$$=\; \Big(\langle \hat{x}\,,\, y_1\rangle \;-\; f^*(y) \;+\; t\,\big(\langle \hat{x}\,,\, y\rangle \;-\; \sup_{x \in \mathbf{dom}\, f} \langle x\,,\, y\rangle \;\big) \Big)\,.$$

При $t \to +\infty$, правая часть последнего соотношения стремится к $+\infty$ и $f^{**}(\hat{x}) = +\infty$. Противоречие. Остается вариант, когда $\lambda \;<\; 0$. Имеем

$$\langle \hat{x}, y\rangle \;+\; \lambda\, u \quad > \quad \sup_{x \in \mathbf{dom}\, f} \big(\langle x, y\rangle \;+\; \lambda f(x) \big)\,,$$

Откуда

$$\sup_{x \in \mathbf{dom}\, f} \left(\left\langle x\,,\, \frac{y}{|\lambda|} \right\rangle \;-\; f(x) \right) \quad < \quad \left\langle \hat{x}\,,\, \frac{y}{|\lambda|} \;\right\rangle \;-\; u\;.$$

Обозначим $\tilde{y} \;=\; \frac{y}{|\lambda|}$. Тогда

$$f^*(y) \quad < \quad \big(\langle \hat{x}\,,\, \tilde{y}\,\rangle \,-\, u \big)\,.$$

Таким образом,

$$f^{**}(\hat{x}) \quad \ge \quad \big(\langle \hat{x}\,,\, \tilde{y}\rangle \;-\; f^*(\tilde{y}) \big) \;>\; u \;=\; f^{**}(\hat{x})\,.$$

$\square$

Упражнение 14 Найдите сопряженные к функциям $f(x) = (Ax, x)$ и $f(x) \;=\; \sqrt{(Ax, x)}$, где A – положительно определенная матрица.

Упражнение 15 а) Найдите все квадратичные функции $f(x)$ одной переменной, для которых $f^* = f$;
б)* найдите все решения уравнения $f^* = f$ для функций одной переменной.

Преобразование Лежандра-Юнга-Фенхеля задаёт переход к двойственной функции. Для выпуклых замкнутых собственных функций это преобразование инволютивно, т.е., его двойное применение является тождественным преобразованием. Переход к двойственному выпуклому множеству называется *полярным преобразованием.*

Определение 8.4 *Пусть X и Y – ПДВ. Полярой множества $A \subset X$ называется множество $A^* = \left\{y \in Y \;\middle|\; \sup_{x\in A} \langle y, x\rangle \le 1\right\}$.*

Рассмотрим простейшие свойства поляры.

Предложение 8.5 *Поляра любого множества $A \subset X$ является выпуклым замкнутым подмножеством Y, содержащим ноль.*

`Доказательство.` Множество A^* является пересечением выпуклых замкнутых множеств (полупространств) $L_x = \{\, y \in Y \mid \langle y, x\rangle \le 1\,\}$ по всем $x \in X$, а значит – выпукло и замкнуто. Очевидно, поляра содержит ноль.

□

Ясно, что если $A \subset B$, то $B^* \subset A^*$. Положим $A^* = (A^*)^*$ – биполяра.

Теорема 8.3 *(о биполяре). Для любого множества $A \in X$ имеем $A \subset A^{**}$. Более того, биполяра A^{**} содержит замыкание выпуклой оболочки A. Равенство $A = A^{**}$ имеет место в точности когда множество A выпукло, замкнуто и содержит ноль.*

`Доказательство.` Если $x \in A$, то $\langle y, x\rangle \le 1$ для любого $y \in A^*$, а значит $x \in A^{**}$. Таким образом, $A \subset A^{**}$. Более того, так как биполяра A^{**}, будучи полярой множества A^*, выпукла и замкнута, то она содержит замыкание выпуклой оболочки множества A. Если $A = A^{**}$, то, в силу предложения 8.5, множество A выпукло, замкнуто и содержит ноль. Осталось доказать обратное. Пусть A выпукло, замкнуто и содержит ноль, однако не совпадает с A^{**}. Этот значит, что существует точка $\bar{x} \in A^{**}$, $\bar{x} \notin A$. По теореме 8.1 строго отделяем точку $\bar{x}$ от множества A: $\langle \bar{y}, \bar{x}\rangle > \sup_{x\in A} \langle \bar{y}, x\rangle$. Так как $O \in A$, правая часть неравенства неотрицательна, следовательно левая – положительна, а значит $\bar{y} \ne 0$. Кроме того, правая часть неравенства меньше $+\infty$. Поэтому, найдётся $t > 0$, для которого $\langle t\,\bar{y}, \bar{x}\rangle > 1 \ge \sup_{x\in A} \langle t\,\bar{y}, x\rangle$. Следовательно, $t\,\bar{y} \in A^*$, в то время, как $\langle t\,\bar{y}\,,\, \bar{x}\rangle > 1$, а значит $\bar{x} \notin A^{**}$, что противоречит предположению.

□

Непосредственно из определения вытекают следующие свойства полярного преобразования

Предложение 8.6 *а) для любых $A \subset X$ и $\lambda > 0$ имеем $(\lambda\, A)^* = \lambda^{-1} A^*$;*
б) для любого множества $A \subset X$ имеем $A^* = \left[\mathbf{co}\,(A)\right]^* = \left[\overline{\mathbf{co}\,(A)}\right]^*$;
в) для любых замкнутых множеств $A_1, A_2 \subset X$, содержащих ноль, имеем
$(A_1 \cup A_2)^* = A_1^* \cap A_2^*$ и $(A_1 \cap A_2)^* = \overline{\mathbf{co}}\{A_1^*, A_2^*\}$.

Пример 8.2 Пусть для определенности $X = Y = \mathbb{R}^d$. Полярой любой точки $a \ne 0$ является полупространство, ограниченное гиперплоскостью $(a, x) = 1$ и содержащее 0. В силу свойства (в) предложения 8.6 поляра конечного числа точек $a_1, \ldots, a_n$ является пересечением n подпространств, т.е, многогранным множеством. Следовательно, и поляра многогранника $A = \mathbf{co}\,\{a_1, \ldots, a_n\}$ является многогранным множеством. В силу упражнения 16 оно будет ограничено (т.е., будет многогранником) когда $0 \in \mathrm{int}\, A$. Итак, поляра многогранника, содержащего ноль в качестве внутренней точки, является многогранником. Число вершин многогранника A^* равно числу граней многогранника A, и наоборот, число вершин A равно числу граней A^*. Например, полярой куба в $\mathbb{R}^3$ является октаэдр, полярой тетраэдра – тетраэдр.

Упражнение 16 Докажите, что поляра множества $A \subset \mathbb{R}^d$ ограничена тогда и только тогда когда $0 \in \mathrm{int}\, A$.

Упражнение 17 Найдите поляру прямого кругового цилиндра (начало координат – в центре цилиндра). Найдите поляру прямого кругового конуса (начало координат – на оси конуса).

9 Лекция

7 апреля 2009

Линейное программирование. Постановка задачи и предварительные сведения.

Примеры задач линейного программирования.
Многогранные множества и их вершины. Случай отсутствия вершин.

Задачей линейного программирования (ЛП) называется задача

$$\begin{cases} (c,x) \ \to \ \inf \\ Ax \ \le b\,. \end{cases} \tag{9.11}$$

Здесь A – $n \times d$-матрица, $c \in \mathbb{R}^d$, $b \in \mathbb{R}^n$ – некоторый вектор. Неравенство $Ax \le b$ означает, что $(a_i, x) \le b_i$ для всех $i = 1, \ldots, n$, где a_i – вектора-строки матрицы A, b_i – координаты вектора b. Данная задача имеет и другие равносильные формулировки. Например, такую:

$$\begin{cases} (c,x) \ \to \ \inf \\ Ax \ \le \ b \\ x \ \ge \ 0\,, \end{cases} \tag{9.12}$$

или такую:

$$\begin{cases} (c,x) \ \to \ \inf \\ Ax \ = \ b \\ x \ \ge \ 0\,. \end{cases} \tag{9.13}$$

Неравенство $x \ge 0$ означает, что все координаты вектора x неотрицательны. Все три формулировки равносильны, каждая из них сводится к любой другой с помощью подходящей замены переменной (при этом, конечно, могут меняться матрица A, векторы b и c, и размерности n и d). Мы пока не будем доказывать их равносильность, а рассмотрим несколько практических примеров.

Пример 9.1 Задача оптимального планирования производства. Предприятие выпускает n видов продукции, и потребляет при этом m видов сырья. Для выпуска единицы i-того вида продукции затрачивается a_{i1} единиц первого вида сырья, a_{i2} единиц второго вида сырья, и т.д., $i = 1, \ldots, n$. Суммарный объём j-того вида сырья, находящегося в распоряжении предприятия равен b_j, $j = 1, \ldots, m$. Прибыль с производства i-того вида продукции равна r_i рублей. Задача: получить максимальную прибыль при условии, что расход сырья не превысит тот, который находится в нашем распоряжении. Пусть x_i – произведённый объём i-того вида продукции. Тогда общая прибыль равна $(r, x) = \sum_{i=1}^{n} r_i x_i$, где $r = (r_1, \ldots, r_n)$, $x = (x_1, \ldots, x_n)$. Обозначая $c = -r$, получаем задачу оптимального планирования производства:

$$\begin{cases} (c, x) \ \to \ \inf \\ \sum_{i=1}^{n} a_{ij} x_i \ \le \ b_j \,, \quad j \ = \ 1, \ldots, m; \\ x \ \ge \ 0\,. \end{cases} \tag{9.14}$$

Это ЛП задача в форме (9.12).

Пример 9.2 Задача оптимального оптимального использования посевной площади. В распоряжении фермера m участков земли площадей $b_1, \ldots, b_m$, на которых нужно посадить n возделываемых культур. Урожайность i-той культуры на j-том участке равна a_{ji}, $i = 1, \ldots, n$ $j = 1, \ldots, m$. Прибыль с единицы i-той культуры равна c_i. Наконец, общая площадь, отведенная под i-тую культуру, должна быть не меньше d_i. Цель – получить максимальную прибыль. Обозначая через x_{ji} площадь, отведенную под i-тую культуру на j-том участке, получаем задачу:

$$\begin{cases} -\sum_{i=1}^{n} c_i \left(a_{i1} x_{i1} + \ldots, a_{im} x_{im} \right) \ \to \ \inf \\ \sum_{i=1}^{n} a_{ji} x_{ji} \ = \ b_j \,, \quad j \ = \ 1, \ldots, m; \\ \sum_{j=1}^{m} a_{ji} x_{ji} \ \ge \ d_i \,, \quad i \ = \ 1, \ldots, n; \quad x \ \ge \ 0\,. \end{cases} \tag{9.15}$$

Это ЛП задача в форме (9.12).

Пример 9.3 Транспортная задача. Есть n карьеров песка, с которых надо обеспечивать песком m потребителей. С i-того карьера можно увезти a_i тонн песка в сутки, а j-тому потребителю нужно b_j тонн песка в сутки. При этом перевозка одной тонны песка с i-того карьера j-тому потребителю обходится в c_{ij} рублей. Задача: обеспечить всех потребителей, затратив при этом наименьшую возможную сумму денег:

$$\begin{cases} \sum_{i,j} c_{ij} x_{ij} \ \to \inf \\ \sum_{i=1}^{n} x_{ij} \ = \ b_j \,, \quad j \ = \ 1, \ldots, m; \\ \sum_{j=1}^{m} x_{ij} \ = \ a_i \,, \quad i \ = \ 1, \ldots, n; \\ x \ \ge \ 0\,. \end{cases} \tag{9.16}$$

Это ЛП задача в форме (9.13).

Рассмотрим ЛП задачу в форме (9.11). Её решение будем обозначать буквой S. При этом считаем, что $S = +\infty$, если система $Ax \le b$ не имеет решений, т.е., нет ни одной допустимой точки. Если при условиях $Ax \le b$ значение $\inf\,(c, x)$ может быть сколь угодно мало, то $S = -\infty$. Наконец, если значение (c, x) для допустимых точек x ограничено снизу, то $S = \inf\,(c, x)$. Решение задачи состоит в нахождении S. Главная идея решения состоит в том, что множество допустимых значений x, заданное линейными ограничениями $Ax \le b$ является многогранным множеством $G \subset \mathbb{R}^d$. А минимум линейной функции на многогранном множестве достигается в его вершине. Поэтому для решения нужно перебрать все вершины и выбрать наименьшее значение функции. Это, конечно, только идея, и для её реализации нужна предварительная работа.

Определение 9.1 *Многогранным множеством называется множество решений линейной системы* $(a_i, x) \le b_i$, $i = 1, \dots, n$. *Вершиной многогранного множества называется его крайняя точка. Многогранником называется ограниченное многогранное множество.*

Таким образом, многогранное множество всегда выпукло. Оно либо совпадает со всем пространством $\mathbb{R}^d$, либо является пересечением конечного числа полупространств. Везде далее мы рассматриваем непустые многогранные множества. Пусть многогранное множество G определено системой $Ax \le b$. Для данной точки $\bar{x} \in G$ положим $I(\bar{x}) = \{ i \mid (a_i, x) = \bar{x} \}$.

Предложение 9.1 *Точка* $\bar{x}$ *является вершиной многогранного множества* $G \subset \mathbb{R}^d$ *тогда и только тогда когда* $x \in G$ *и система* $(a_i, x) = b_i$, $i \in I(\bar{x})$ *имеет единственное решение.*

`Доказательство.` Предположим, что $\bar{x}$ – вершина. Это значит, что $(a_i, \bar{x}) = b_i$ при $i \in I(\bar{x})$ и $(a_i, \bar{x}) < b_i$ при $i \notin I(\bar{x})$. Если система уравнений $(a_i, \bar{x}) = b_i$ имеет решение $\bar{x} + h$, $h \ne 0$, то $(a_i, h) = 0$, $i \in I(\bar{x})$, а значит $\bar{x} - h$ – также решение этой системы. При достаточно малом h точка $x = \bar{x} - h$ удовлетворяет также всем неравенствам $(a_i, x) < b_i$, $i \notin I(\bar{x})$, а значит, принадлежит G. Следовательно, $\bar{x} \notin \mathbf{extr}\, G$, поскольку точка $\bar{x}$ является серединой отрезка $[\bar{x} - h\,,\, \bar{x} + h]$, концы которого принадлежат G. Значит, $\bar{x}$ – не вершина, что приводит к противоречию. Обратно, если $\bar{x}$ – единственное решение системы $(a_i, \bar{x}) = b_i$, $i \in I(\bar{x})$, но $\bar{x}$ – не вершина, то найдется $h \in \mathbb{R}^d$, для которого $\bar{x} \pm h \in G$, а значит $(a_i, \bar{x} \pm h) \le b_i$, $i \in I(\bar{x})$, откуда $(a_i, h) = 0$, $i \in I(\bar{x})$. Следовательно, $(a_i, \bar{x} + h) = b_i$, $i \in I(\bar{x})$, а значит $h = 0$.

$\square$

Обозначая через $|M|$ мощность множества M, получаем

Следствие 9.1 *Для любой вершины* $\bar{x}$ *имеем* $|I(\bar{x})| \ge d$.

Предложение 9.2 *Следующие свойства многогранного множества* $G = \{x \in \mathbb{R}^d \mid Ax \le b\}$ *равносильны:*

1) G не имеет вершин;
2) $\operatorname{rank} A < d$;
3) существует вектор $h \ne 0$, ортогональный всем a_i, $i = 1, \dots, m$;
4) G содержит прямую линию.

`Доказательство.` (1) $\Leftrightarrow$ (3). Если G не имеет вершин, то возьмём точку $\bar{x} \in G$, для которой число $|I(\bar{x})|$ максимально. Найдётся $h \ne 0$, для которого $\bar{x} \pm h \in G$, а значит $(a_i, h) = 0$ для $i \in I(\bar{x})$. Допустим, что $(a_j, h) \ne 0$ для некоторого j. Не ограничивая общности, считаем, что $(a_j, h) > 0$. Тогда полагаем $t = \min\limits_{j \notin I(\bar{x})\,,\,(a_i,h)>0} \frac{b_i - (a_i, \bar{x})}{(a_i, h)}$. Ясно, что $(a_j, \bar{x} + th) \le b_i$ для всех i, причём для некоторых $i \notin I(\bar{x})$ это неравенство обращается в равенство. Следовательно, $|I(\bar{x} + th)| > |I(\bar{x})|$, что противоречит предположению. Следовательно, $(a_i, h) = 0$ для всех i. Обратно, если нашёлся вектор $h \ne 0$, для которого $(a_i, h) = 0$ при всех i, то для любой точки $\bar{x} \in G$ имеем $\bar{x} \pm h \in G$, а значит $\bar{x}$ – не вершина.

(2) $\Leftrightarrow$ (3) – очевидно.

(4) $\Leftrightarrow$ (3). Если существует h ортогональный всем a_i, то для любого $\bar{x} \in G$ имеем $(a_i, \bar{x} + th) = (a_i, \bar{x}) \le b_i$ при всех $t \in \mathbb{R}$. Следовательно, прямая $\{\bar{x} + th \mid t \in \mathbb{R}\}$ лежит в G. Обратно, если некоторая прямая $\{\bar{x} + th \mid t \in \mathbb{R}\}$ лежит в G, то $(a_i, \bar{x}) + t(a_i, h) \le b_i$ для всех $t \in \mathbb{R}^d$, а значит $(a_i, h) = 0$ при всех $i = 1, \dots, m$.

□

Предложение 9.2 показывает, как распознавать многогранные множества, не имеющие вершин. Для этого достаточно, например, решить систему уравнений $Ah = 0$. Если она имеет ненулевое решение, то многогранное множество вершин не имеет, а если его решение – только $x = 0$, то вершины есть. Рассмотрим подробнее случай отсутствия вершин. Пусть $h_1, \dots, h_k$ – базис пространства решений уравнения $Ah = 0$. Возможны следующие случаи:

1) Если система $Ax \le b$ несовместна, т.е., если множество G пусто, то $S = +\infty$. Мы опишем позже, как эффективно распознавать этот случай.

2) Если система совместна, и $(c, h_j) \ne 0$ для некоторого j, то $S = -\infty$. В самом деле, взяв любую точку $x \in G$, получим, что $x + th_j \in G$ для каждого $t \in \mathbb{R}$. Тогда $(c, x + th_j) = (c, x) + t(c, h_j) \to -\infty$ при $t \to -\infty$, если $(c, h_j) > 0$, и при $t \to +\infty$, если $(c, h_j) < 0$.

3) Если $(c, h_j) = 0$ для всех $j = 1, \dots, k$, то рассматриваем ортогональную проекцию на подпространство H – ортогональное дополнение векторов $h_1, \dots, h_k$ в пространстве $\mathbb{R}^d$. Обозначим через x_h проекцию данной точки $x \in \mathbb{R}^d$. Так как каждый из векторов h_k ортогонален всем a_i и c, то все a_i лежат в H и $c \in H$. Кроме того $(a_i, x_h) = (a_i, x)$ для всех i и $(c, x_h) = (c, x)$. Таким образом, мы переходим к той же ЛП задаче, с теми же a_i, c и b, но на пространстве H вместо $\mathbb{R}^d$. Это пространство имеет меньшую размерность, чем d, и на нём множество $Ax \le b$ будет иметь вершины. В самом деле, если вершин нет, то согласно предложению 9.2 найдётся ненулевой $h \in H$, ортогональный всем a_i. Значит $h = \sum_{j=1}^{k} \alpha_j \, h_j$ для некоторых НЕРОН коэффициентов $\alpha_1, \dots, \alpha_k$, поскольку $h_1, \dots, h_k$ – базис пространства решений. Но так как $h \in H$, то h ортогонален всем h_k. Поэтому, домножив равенство $h = \sum_{j=1}^{k} \alpha_j \, h_j$ скалярно на h, получаем, что $(h, h) = \sum_{j=1}^{k} \alpha_j \, (h, h_j) = 0$, а значит $h = 0$. Таким образом, в данном случае мы переходим к ЛП задаче меньшей размерности, в которой допустимое множество G будет иметь вершины.

В дальнейшем будем всегда предполагать, что G имеет вершины.

10 Лекция

14 апреля 2009

Основная теорема линейного программирования. Симплекс-метод.

Основная теорема линейного программирования. Теорема о точечном конусе.
Алгоритм поиска первой вершины. Симплекс-метод в невырожденном случае.

Гранью называется пересечение многогранного множества G с любой его *опорной плоскостью*, т.е., с гиперплоскостью, имеющей общие точки с G, но не имеющей общих точек с $\operatorname{int} G$.

Грань – выпуклое множество, его размерность k может быть от нуля до $d-1$. Так нульмерная грань – это вершина. *Ребром* называется грань размерности 1. Таким образом, ребро – это либо прямая линия, либо луч, выходящий из вершины, либо отрезок, соединяющий две вершины. Далее под гранью будем всегда понимать *гипергрань*, т.е., грань размерности $d-1$, а все прочие будем называть k-мерными гранями, $k \le d-2$.

Теорема 10.1 *(Основная теорема ЛП). Пусть $G \ne \emptyset$ и $\operatorname{rank} A = d$ (т.е., G имеет вершины). Тогда либо $S = -\infty$, либо S достигается в одной из вершин G. Вершина является точкой минимума тогда и только тогда когда функция (c, x) не уменьшается при движении от этой вершины вдоль любого ребра, выходящего из неё.*

Для доказательства нам понадобятся некоторые факты из теории выпуклых конусов.

Определение 10.1 *Конусом с вершиной 0 называется подмножество K линейного пространства X, обладающее двумя свойствами:*
а) если $x \in K$, то $\lambda x \in K$ для любого $\lambda \ge 0$;
б) если $x, y \in K$, то $x + y \in K$.

Конус всегда содержит ноль, это следует из свойства (а) при $\lambda = 0$. Конус – выпуклый. В самом деле, если $x, y \in K$, то при любом $t \in [0,1]$ имеем $t(1-x), ty \in K$, а значит $(1-t)x + ty \in K$. Самый маленький конус – это $\{0\}$, самый большой – всё пространство X. Впредь не будем рассматривать тривиальный конус $\{0\}$. Также по умолчанию будем рассматривать только замкнутые конусы. В нормированном пространстве любой конус неограничен, так как он содержит луч. *Образующей* конуса называется луч $r = \bigl\{ tx \bigm| t \ge 0 \bigr\} \subset K$, каждая точка которого y обладает свойством: если $y = \frac{1}{2}\bigl(y_1 + y_2\bigr)$, то $y_1, y_2 \in r$. Конус называется *точечным*, если вершина – его крайняя точка. Таким образом, конус – точечный, если не существует элемента $a \ne 0$ такого, что $a \in K$ и $-a \in K$.

Упражнение 18 Докажите, что конус является точечным тогда и только тогда когда он не содержит прямой линии.

Теорема 10.2 *(о точечном конусе). Точечный конус в $\mathbb{R}^d$ является выпуклой оболочкой своих образующих.*

`Доказательство`. Пусть K' – пересечение конуса K с единичной сферой. Функция $f(x, y)$ – угол между векторами $x, y \in K'$, непрерывна на компакте $K' \times K'$, поэтому она достигает своего максимума. Этот максимум φ меньше, чем π, иначе конус – не точечный. Итак, угол между любыми двумя векторами из K не превосходит $\varphi < \pi$. Тогда существует точечный конус K_s, содержащий внутри себя K, т.е., $K \setminus \{0\} \subset \operatorname{int} K_s$. Например, $K_s = \bigl\{ tx \bigm| x \in \mathbf{co}\,\{K' + rB(0,1)\}\,,\, t \ge 0 \bigr\}$, где $B(0,1)$ – шар радиуса 1 с центром в нуле. При достаточно малом $r > 0$ получим, что угол между любыми двумя векторами из K_s меньше π, а значит K_s – точечный. По теореме отделимости, можно отделить конус K_s от нуля: существует вектор $z \ne 0$, для которого $\inf_{y \in K_s} (z, y) \ge 0$. Множество $P = \{x \in K \mid (z, x) = 1\}$ компактно. Замкнутость следует из того, что P – пересечение замкнутого множества K и гиперплоскости. Ограниченность следует из того, что $\inf_{x \in K, \|x\|=1, \|h\|=r} (z, x + h) \ge 0$, а значит

$$\inf_{x \in K, \|x\|=1} (z, x) \;\ge\; \sup_{\|h\|=r} (z, -h) \;=\; r\,\|z\|\,.$$

Таким образом, $(z, x) \ge r\,\|x\|\,\|z\|$ для любого $x \in K$, а значит $\|x\| \le \frac{1}{r\,\|z\|}$ для всех $x \in P$. По теореме Крейна-Мильмана (теорема 2.4) множество P есть выпуклая оболочка своих крайних точек. Так как для каждой крайней точки x множества P луч $\{ t\, x \mid t \geq 0 \}$ – образующая конуса K, то K – выпуклая оболочка своих образующих.

□

`Доказательство теоремы 10.1.` Так как число вершин G конечно, то можно выбрать вершину $\bar{x}$ с минимальным (среди всех вершин) значением функции (c, x). Положим $K = \{ x \in \mathbb{R}^d \mid (a_i, x) \leq b_i\,,\ i \in I(\bar{x}) \}$. Это – конус с вершиной $\bar{x}$, поскольку он состоит из точек x, для которых $(a_i, x - \bar{x}) \leq 0\,,\ i \in I(\bar{x})$. Этот конус — точечный, поскольку его вершина $\bar{x}$ – крайняя точка. Ясно, что $G \subset K$. Пусть $r_1, \ldots, r_m$ – векторы образующих конуса K, они же являются векторами ребер G, выходящих из вершины $\bar{x}$. Если $(c, r_j) \geq 0$ для всех j, то в силу теоремы 10.2 имеем $(c, x - \bar{x}) \geq 0$ для всех $x \in K$. Поэтому, $(c, x) \geq (c, \bar{x})$ для всех $x \in K$, а значит, и для всех $x \in G$. Если же $(c, r_j) < 0$ для некоторого j, то ребро, соответствующее вектору r_j, не содержит других вершин G, кроме $\bar{x}$. Иначе в этой вершине $\tilde{x} = \bar{x} + tr_j$ имеем $(c, \tilde{x}) = (c, \bar{x} + tr_j) < (c, \bar{x})$, что противоречит выбору вершины $\bar{x}$. Итак, данное ребро не содержит других вершин, значит $\bar{x} + tr_j \in G$ при всех $t \geq 0$. Поскольку $(c, \bar{x} + tr_j) \to -\infty$ при $t \to +\infty$, имеем $S = -\infty$.

□

Поиск первой вершины. Пока будем предполагать, что мы можем распознать случай, когда система $Ax \leq b$ несовместна (в этом случае сразу получаем ответ $S = +\infty$), и в случае совместной системы найти одно её решение $\bar{x}$. Если все неравенства $(a_i, \bar{x}) \leq b_i$ – строгие, то возьмём произвольное $h \in \mathbb{R}^d$. Хотя бы для одного j имеем $(a_j, h) \neq 0$, иначе G не имел бы вершин (предложение 9.2). При необходимости изменив знак, считаем, что $(a_j, h) > 0$. Полагаем $t = \min\limits_{j\colon (a_j,h)>0} \frac{b_j - (a_j, \bar{x})}{(a_j, h)}$. Тогда $(a_i, \bar{x} + th) \leq b_i$ для всех i, причём для некоторого i это неравенство обращается в равенство, таким образом $I(\bar{x} + th) \neq \emptyset$.

Предположим теперь, что мы получили точку $\bar{x}$, в которой из неравенств $(a_i, \bar{x}) \leq b_i$ ровно k обращаются в равенства. Если она не является вершиной, то найдётся $h \neq 0$, для которого $(a_i, h) = 0$ при $i \in I(\bar{x})$. Хотя бы для одного $j \notin I(\bar{x})$ имеем $(a_j, h) \neq 0$. При необходимости изменив знак, считаем, что $(a_j, h) > 0$. Полагаем $t = \min\limits_{j\colon (a_j,h)>0} \frac{b_j - (a_j, \bar{x})}{(a_j, h)}$. Тогда при $i \in I(\bar{x})$ имеем $(a_i, \bar{x} + th) = (a_i, \bar{x}) \leq b_i$. При $i \notin I(\bar{x})$ также имеем $(a_i, \bar{x} + th) \leq b_i$, причём для некоторого i это неравенство обращается в равенство. Следовательно, общее число равенств увеличится. Далее делаем следующую итерацию, и так увеличиваем число равенств, пока не получим вершину.

Симплекс-метод решения задачи ЛП. Для решения ЛП задачи $(c, x) \to \min\,,\ Ax \leq b$ сначала выясняем, есть ли вершины у множества допустимых точек $G = \{ x \in \mathbb{R}^d \mid Ax \leq b \}$. Если нет, то сводим задачу к задаче ЛП меньшей размерности. Этот процесс описан в предыдущей лекции. Теперь считаем, что G имеет вершины. Далее начинается *алгоритм симплекс-метода.* Ищем первую вершину x^1. Если функция (c, x) не уменьшается вдоль всех рёбер, выходящих из этой вершины, то x^1 – точка минимума (теорема 10.1). Если нашлось ребро, вдоль которого функция уменьшается, то: если на этом ребре нет других вершин, то $S = -\infty$, и задача решена; если же это ребро содержит другую вершину x^2, то $(c, x^2) < (c, x^1)$. Далее переходим к вершине x^2, в которой значение функции меньше, и т.д. Так получаем последовательность вершин $x^1, x^2, x^3, \ldots$ с уменьшением значения функции (c, x). Алгоритм завершается, когда либо мы приходим к вершине, в которой достигается минимум, либо получаем, что $S = -\infty$.

Естественный вопрос – как технически реализовать алгоритм ? Как перебирать все рёбра, выходящие из данной вершины ? Мы изложим идею практической реализации симплекс-метода в *невырожденном случае*, когда в каждой вершине $\bar{x} \in G$ сходятся ровно d рёбер, т.е., $|I(\bar{x})| = d$ для всех вершин $\bar{x}$.

Реализация симплекс-метода в невырожденном случае. Пусть $\bar{x}$ – данная вершина G. По предположению $|I(\bar{x})| = d$, не ограничивая общности, считаем, что $I(\bar{x}) = \{1, \ldots, d\}$. Обозначим через $\bar{A}$ $d \times d$-матрицу из строк $a_1, \ldots, a_d$, через $\bar{b}$ – вектор с координатами $b_1, \ldots, b_d$. Таким образом, $\bar{A}\,\bar{x} = \bar{b}$ и $(a_i, \bar{x}) < b_i$ при $i > d$. Поскольку $\bar{x}$ – вершина, уравнение $\bar{A}x = \bar{b}$ имеет единственное решение, матрица $\bar{A}$ – невырожденная, а значит невырождена и сопряжённая к ней матрица $\bar{A}^*$. Следовательно, уравнение $\bar{A}^*\, y = c$ имеет единственное решение. Найдём это решение $\bar{y}$ и рассмотрим два случая.

1) $\bar{y} \le 0$ (т.е., все координаты вектора $\bar{y}$ неположительны). В этом случае $\bar{x}$ – точка минимума. Докажем это. Пусть $y = (\bar{y}, 0, \ldots, 0) \in \mathbb{R}^n$ (дополнили вектор $\bar{y}$ с помощью $n - d$ нулей). Для любого $x \in G$ имеем

$$(c, x) \;=\; (\bar{A}^*\bar{y}, x) \;=\; (\bar{y}, \bar{A}x) \;=\; (y, Ax) \;\ge\; (y, b)\,,$$

поскольку $Ax \le b$ и $y \le 0$. С другой стороны,

$$(c, \bar{x}) \;=\; (\bar{A}^*\bar{y}, \bar{x}) \;=\; (\bar{y}, \bar{A}\bar{x}) \;=\; (\bar{y}, \bar{b}) \;=\; (y, b)\,.$$

Таким образом, $(c, x) \ge (c, \bar{x})$ для любого $x \in G$.

2) Вектор $\bar{y}$ имеет хотя бы одну положительную координату. Пусть, например, $\bar{y}_1 = \alpha > 0$. Решаем линейное уравнение $\bar{A}h = -e_1$, где $e_1 = (1, 0, \ldots, 0)$ – первый базисный вектор. Для любого $t > 0$ у вектора $A(\bar{x} + th) = A\bar{x} + t\,A\,h$. первая координата равна $b_1 - t$, j-тая координата равна b_j для $j = 2, \ldots, d$, для остальных j соответствующая координата меньше b_j при малых t. Для целевой функции имеем

$$(c, \bar{x}+th) \;=\; (c, \bar{x}) + t\,(c, h) \;=\; (c, \bar{x}) + t\,(\bar{A}^*\bar{y}, h) \;=\; (c, \bar{x}) + t\,(\bar{y}, \bar{A}h) \;=\; (c, \bar{x}) - t\,\alpha \;<\; (c, \bar{x})\,.$$

Таким образом, функция $(c, \bar{x}+th)$ линейно убывает по t. Если $Ah \le 0$, то $\bar{x}+th \in G$ при любом $t > 0$, а значит $S = -\infty$. Если же вектор Ah имеет хотя бы одну положительную координату, т.е., $(a_j, h) > 0$ для некоторых j, то полагаем $s \;=\; \min\limits_{j\,:\,(a_j,h)>0} \frac{b_j - (a_j, \bar{x})}{(a_j, h)}$. Тогда $\bar{x} + s\,h$ – следующая вершина. В самом деле, в этой точке выполнены все неравенства $(a_i, x) \le b_i\,,\; i = 1, \ldots, n$, и ровно d из них обращаются в равенства. Итак, мы нашли вершину, в которой значение функции (c, x) меньше, чем в предыдущей. Далее переходим к следующей вершине, и т.д.

Инициализация симплекс-метода. Вернёмся теперь в самое начало решения задачи ЛП. Для того, чтобы найти первую вершину многогранного множества G нужно сначала узнать, не будет ли множество G пустым. Если окажется, что $G = \emptyset$, то $S = +\infty$, а если $G \neq \emptyset$, то нужно найти хотя бы одну точку $x \in G$, с которой можно начать поиск первой вершины. Для того, чтобы решить эти задачи, решим сначала следующую вспомогательную задачу с $d + 1$ переменной:

$$\begin{cases} x_0 \;\to\; \inf \\ (a_i, x) \;\le b_i \;+\; x_0\,, \quad i \;=\; 1, \ldots, n \\ x_0 \;\ge\; 0\,. \end{cases} \tag{10.17}$$

Здесь $x = (x_1, \ldots, x_d)$, а x_0 – дополнительная переменная. Это задача ЛП. Вопрос о непустоте множества допустимых точек и о нахождении первой допустимой точки в этой задаче решается тривиально. Берём любой вектор $x \in \mathbb{R}^d$ и берём настолько большой x_0, чтобы все неравенства в (10.17) были выполнены. Получаем допустимую точку (x_0, x) в задаче (10.17). Далее мы ищем первую вершину и решаем эту задачу симплекс-методом. Если получаем $S > 0$, то в исходной задаче множество G допустимых точек пусто. Если же получаем $S = 0$, значит $x_0 = 0$, и получаем, что $G \neq \emptyset$. Более того, симплекс-метод дал нам вершину $(0, \bar{x})$, в которой достигается решение задачи. Для неё $A\bar{x} \le b$, а значит $\bar{x} \in G$. Таким образом, мы нашли допустимую точку в исходной задаче. Оказывается (упражнение 19) точка $\bar{x}$ будет вершиной многогранника G. С неё можно начать алгоритм симплекс-метода.

Упражнение 19 Докажите, что если точка $(0, \bar{x}_1, \ldots, \bar{x}_d)$ – вершина в задаче (10.17) дающая решение этой задачи, то $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_d)$ – вершина множества G допустимых точек в исходной задаче (9.11).

Двухфазный симплекс-метод. Решение любой ЛП задачи осуществляется двухфазным симплекс-методом. Первая фаза состоит в решении вспомогательной задачи (10.17). Она решается обычным симплекс-методом, для неё не нужно вспомогательных задач. Решение задачи (10.17) даёт нам первую вершину множества G. Далее осуществляем вторую фазу – решаем исходную задачу (9.11) симплекс-методом.

11 Лекция

21 апреля 2009

Выпуклое программирование. Метод центрированных сечений.

Постановка задачи выпуклого программирования.
Метод центрированных сечений. Метод центров тяжести. Метод эллипсоидов.

Задача выпуклого программирования состоит в нахождении минимума выпуклой функции при выпуклых ограничениях:

$$\begin{cases} f(x) \ \to \ \min \\ x \in G \ \Leftrightarrow \ f_j(x) \ \le \ 0\,, \quad j \,=\, 1, \ldots, n\,. \end{cases} \tag{11.18}$$

Все функции f, f_j – выпуклые, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, множество G компактно и имеет хотя бы одну внутреннюю точку. Требуется найти точку x_*, в которой относительное отклонение от минимума не превосходит заданного значения $\varepsilon > 0$. Это означает, что

$$\frac{f(x_*) \,-\, f_{\min}}{f_{\max} \,-\, f_{\min}} \ \le \ \varepsilon\,, \tag{11.19}$$

где $f_{\min}$ и $f_{\max}$ – соответственно, минимум и максимум функции f на G. Выпуклая задача с ограничениями решается, как мы знаем, с помощью теоремы Каруша-Куна-Таккера. Однако, при решении получающейся системы уравнений на $\hat{x}$ и $\hat{\lambda}$ возникают 2^n случаев: для каждого $j = 1, \dots, n$ либо $\hat{\lambda}_j = 0$, либо $f_j(\hat{x}) = 0$ (условия дополняющей нежёсткости). Таким образом, для нахождения точки минимума нужно, вообще говоря, разобрать 2^n случаев, что практически невозможно при большом числе ограничений n, скажем, при $n \ge 50$. В приложениях встречаются задачи, где число ограничений – несколько тысяч. В настоящее время существует несколько методов решения задачи (11.18). Мы рассмотрим только один из них: метод центрированных сечений, и опишем две его разновидности – метод центров тяжести и метод эллипсоидов.

Вначале докажем следующий вспомогательный результат. Через $\mathrm{Vol}\, G$ обозначаем объём множества G.

Лемма 11.1 *Пусть функция $f : G \to \mathbb{R}$ выпукла и непрерывна, G – выпуклый компакт. Точка $x_* \in \mathrm{int}\, G$ и множество $G' \subset G$ таковы, что*

1) $f(x) \ge f(x_*)$ для любого $x \in G \setminus G'$;

2) $\mathrm{Vol}\, G' \le \varepsilon^d \,\mathrm{Vol}\, G$.

Тогда точка x_* даёт минимум функции f с относительной погрешностью ε, т.е. удовлетворяет (11.19).

Таким образом, если точка x_* такова, что множество всех точек x, в которых значение функции f меньше, чем в точке x_*, имеет малый объём: $\mathrm{Vol}\, G' \le \varepsilon^d \,\mathrm{Vol}\, G$, то x_* – искомая точка, удовлетворяющая (11.19).

`Доказательство.` Пусть $x_{\min}$ – точка минимума функции f на G. Проведём гомотетию с центром $x_{\min}$ и коэффициентом $\frac{1}{\varepsilon}$. При этом множество G' перейдет в множество G_ε, объём которого не превосходит $\mathrm{Vol}\, G$. Следовательно, G не может лежать внутри G_ε. Значит, существует точка $\bar{x} \in G$, не принадлежащая внутренности G_ε. Следовательно, её прообраз – точка $(1-\varepsilon)x_{\min} + \varepsilon\bar{x}$ не принадлежит $\mathrm{int}\, G'$. Из непрерывности функции f следует, что $f\big((1-\varepsilon)x_{\min} + \varepsilon\bar{x}\big) \ge f(x_*)$. Пользуясь выпуклостью f заключаем, что

$$(1-\varepsilon)\, f(x_{\min}) \; + \; \varepsilon\, f(\bar{x}) \;\ge\; f\big((1-\varepsilon)x_{\min} \,+\, \varepsilon\bar{x}\big) \;\ge\; f(x_*)\,.$$

Отсюда получаем

$$\frac{f(x_*) \,-\, f_{\min}}{f(\bar{x}) \,-\, f_{\min}} \;\le\; \varepsilon\,.$$

Пользуясь теперь тем, что $f(\bar{x}) \le f_{\max}$, приходим к (11.19).

$\square$

Таким образом, для приближённого нахождения минимума достаточно локализовать точку минимума в малом объёме. В этом состоит идея метода центрированных сечений. Мы предполагаем, что функция f непрерывна, и в любой точке $x \in \mathrm{int}\, G$ мы можем вычислить хотя бы один элемент субдифференциала $\partial f(x)$. Например, найти производную $f'(x)$, если функция в этой точке дифференцируема.

Метод центров тяжести (А.Ю.Левин, Д.Ньюмен). Центром тяжести выпуклого множества G называется точка $\mathrm{gr}\,(G) \;=\; \frac{1}{\mathrm{Vol}\, G}\int_G x\; dx$. Обозначим G через G_1 и в качестве первой точки выберем $x^1 = \mathrm{gr}\, G_1$. Находим $a^1 \in \partial f(x^1)$ (при условии, что $a^1 \ne 0$, иначе $x^1 = x_*$ – точка минимума), отбросим ту часть G_1, которая лежит в полупространстве $\Pi_1 := \{x \mid (a^1, x - x^1) > 0\}$. Для любого $x \in G_1 \cap \Pi_1$ имеем

$$f(x) \,-\, f(x^1) \;\ge\; (a^1, x - x^1) \;>\; 0 \qquad \Longrightarrow \qquad f(x) \;>\; f(x^1)\,.$$

Получим множество $G_2 = G_1 \setminus \Pi_1$. Далее поступаем аналогично: вычисляем $x^2 = \mathrm{gr}\, G_2$, $a^2 \in \partial f(x^2)$, отбрасываем точки из $\Pi_2 := \{ x \mid (a^2, x - x^2) > 0 \}$, получаем G_3 и т.д.. Делаем так, пока не получим $\mathrm{Vol}\, G_N \le \varepsilon^d \mathrm{Vol}\, G$. Тогда выберем из точек $x^1, \ldots, x^N$ точку $x_* = x^j$ с наименьшим значением функции f. Для любого $x \in G \setminus G_N$ имеем $f(x) > f(x_*)$. В самом деле, поскольку точка x не принадлежит G_N, она была отброшена не некотором, скажем k-том, шаге алгоритма, следовательно $f(x) > f(x^k) \ge f(x_*)$. Итак, точка x_* и множество $G' = G_N$ удовлетворяют условиям леммы 11.1, следовательно, точка x_* – искомая.

Чтобы оценить, сколько понадобится шагов N, воспользуемся классическим результатом, который примем без доказательства.

Теорема 11.1 *(Грюнбаум, Хаммер). Пусть $U \subset \mathbb{R}^d$ — произвольное компактное выпуклое множество, $\mathrm{int}\, U \ne \emptyset$. x_0 — центр тяжести U, Γ — гиперплоскость, содержащая точку x_0, U', U'' — множества, на которые гиперплоскость Γ разбивает множество U, $U = U' \cup \Gamma \cup U''$. Тогда*

$$\max \left\{ \mathrm{Vol}\, U', \mathrm{Vol}\, U'' \right\} < (1 - e^{-1}) \mathrm{Vol}\, U.$$

Таким образом, $\mathrm{Vol}\, G_{k+1} / \mathrm{Vol}\, G_k < 1 - e^{-1} = 0.6321\ldots$ для каждой итерации k. Следовательно, после $N \le \frac{d \ln \varepsilon}{\ln(1 - e^{-1})} + 1$ итераций, получаем решение задачи.

Применение метода центров тяжести на практике осложнено тем, что поиск центра тяжести выпуклого множества – алгоритмически сложная задача. Следующая модификация метода центрированных сечений, легко реализуемая на практике – метод эллипсоидов.

Метод эллипсоидов (А.С.Немировский, Н.З.Шор). Идя метода состоит в том, чтобы изначально поместить множество G в некоторый эллипсоид, а затем брать центры эллипсоидов вместо центров тяжести. Начнём со вспомогательного факта (без доказательства).

Лемма 11.2 *Пусть $E_1 \subset \mathbb{R}^d$ — эллипсоид, Γ — гиперплоскость, проходящая через его центр, S_1 — полуэллипсоид, т.е. пересечение эллипсоида E_1 с одним из полупространств, на которые Γ разбивает $\mathbb{R}^d$. Тогда существует E_2 — эллипсоид, содержащий в себе S_1 такой, что $\mathrm{Vol}\, E_2 \le \lambda_d \mathrm{Vol}\, E_1$, где $\lambda_d = e^{-\frac{1}{2(d+1)}}$.*

Теперь изложим алгоритм метода эллипсоидов.

Нулевой шаг алгоритма: обозначим $G_0 = G$, E_0 — некоторый эллипсоид, содержащий в себе множество G_0.

На i-ом шаге $(1 \le i \le N)$: $G_{i-1} \subset E_{i-1}$, в качестве x^i возьмем центр эллипсоида E_{i-1}.

Если $x^i \notin \mathrm{int}\, G_{i-1}$, то в силу теоремы отделимости существует замкнутое полупространство Π_i, граница которого содержит x^i и $G_{i-1} \subset \Pi_i$. Обозначим $S_i = \Pi_i \cap E_{i-1}$. Тогда, в силу леммы 11.2, существует такой эллипсоид E_i, что $E_i \supset S_i$, $\mathrm{Vol}\, E_i \le \lambda_d \mathrm{Vol}\, E_{i-1}$, при этом $G_{i-1} \subset E_i$. Теперь полагаем $G_i = G_{i-1}$.

Если $x^i \in \mathrm{int}\, G_{i-1}$, то вычисляем $a^i \in \partial f(x^i)$, и в случае $a^i = 0$ заключаем, что точка $x_* = x^i$ искомая; в случае $a^i \ne 0$ обозначим $S_i = \left\{ x \in E_{i-1} \mid (a^i, x) \le (a^i, x^i) \right\}$ и построим эллипсоид $E_i \supset S_i$, для которого $\mathrm{Vol}\, E_i \le \mathrm{Vol}\, \lambda_d\, E_{i-1}$. Полагаем $G_i = G_{i-1} \cap S_i$. Когда объём эллипсоида E_N станет меньше $\varepsilon^d \mathrm{Vol}\, G_0$, полагаем $x_* = x^j$ – точка, для которой $f(x^j) = \min\limits_{i=1,\ldots,N} f(x^i)$ При этом, $\mathrm{Vol}\, E_N \le \lambda_d^N \mathrm{Vol}\, E_0$. Таким образом, после $N \le \frac{\ln\left(\frac{\mathrm{Vol}\, G_0}{\mathrm{Vol}\, E_0}\right) + d \ln \varepsilon}{\ln \lambda_d} + 1$ итераций, где $\lambda_d = e^{-\frac{1}{2(d+1)}}$, получаем решение задачи.