

О минимуме максимума гладких функций

Г. Г. Магарил-Ильяев

В заметке приводятся необходимые и достаточные условия минимума функции, являющейся максимумом конечного числа гладких функций многих переменных. Для функций одного переменного «из картинок» сразу видно, что если \hat{x} — локальный минимум максимума функций f_1 и f_2 , где $f_1(\hat{x}) = f_2(\hat{x})$ и, скажем, $f'_1(\hat{x}) \leq f'_2(\hat{x})$, то справедливо включение $0 \in [f'_1(\hat{x}), f'_2(\hat{x})]$. Достаточным это условие не является, как показывает пример функции $f(x) = \max(-x, \sin x)$. Но оно становится достаточным, если в некоторой окрестности \hat{x} графики функций f_1 и f_2 лежат выше своих касательных в точке \hat{x} .

Все эти наблюдения *mutatis mutandis* имеют место и в более общей ситуации. Основной результат заключается в том, что если функция f есть максимум функций f_1, \dots, f_m , определенных в окрестности некоторой точки $\hat{x} \in \mathbb{R}^n$ и $f_1(\hat{x}) = \dots = f_m(\hat{x})$, то принадлежность нуля выпуклой оболочке производных (градиентов) функций f_1, \dots, f_m в \hat{x} является необходимым условием локального минимума f в точке \hat{x} , и достаточным, если график каждой из функций f_i , $1 \leq i \leq m$, лежит выше некоторой гиперплоскости, проходящей через \hat{x} .

Для точной формулировки результата нужны некоторые начальные сведения о пространстве \mathbb{R}^n , понятие производной функции на \mathbb{R}^n , понятие выпуклого множества, выпуклой оболочки множества, а также теорема Минковского об отделимости точки от выпуклого замкнутого множества в \mathbb{R}^n . Ниже все эти факты (которые вполне стандартны) напоминаются, но если они известны читателю, то следующий параграф можно пропустить.

ПРЕДВАРИТЕЛЬНЫЕ СВЕДЕНИЯ

Пространство \mathbb{R}^n это совокупность всех упорядоченных наборов $x =$
 $= \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ из n действительных чисел (если $n = 1$, то это просто мно-

жество действительных чисел, и мы пишем \mathbb{R} вместо \mathbb{R}^1), называемых *векторами* или *вектор-столбцами*, а числа x_i , $1 \leq i \leq n$ — *координатами вектора* x . Ради экономии места, элементы \mathbb{R}^n будем записывать так

$x = (x_1, \dots, x_n)^T$, где T обозначает *транспонирование* (перевод строки в столбец). В \mathbb{R}^n естественным образом вводится операция (покоординатно) сложения векторов и операция (покоординатного) умножения вектора на число.

Пусть $a = (a_1, \dots, a_n)$ — вектор-строка из n действительных чисел. Для каждого $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ обозначим $a \cdot x = \sum_{i=1}^n a_i x_i$. Ясно, что отображение $x \mapsto a \cdot x$ есть линейная функция (или, говорят, линейный функционал) на \mathbb{R}^n , т.е. $a \cdot (x_1 + x_2) = a \cdot x_1 + a \cdot x_2$ и $a \cdot \alpha x = \alpha a \cdot x$ для любых $x, x_1, x_2 \in \mathbb{R}^n$ и $\alpha \in \mathbb{R}$. Легко видеть, что и любой линейный функционал l на \mathbb{R}^n задается подобным образом с $a = (l(e_1), \dots, l(e_n))$, где $e_1 = (1, 0, \dots, 0)^T, \dots, e_n = (0, \dots, 0, 1)^T$. Таким образом, если обозначить через $(\mathbb{R}^n)^*$ множество, элементы которого суть те же наборы из n действительных чисел, но расположенные в строку (с аналогичными операциями сложения и умножения на числа), то его можно отождествить с пространством всех линейных функционалов на \mathbb{R}^n , которое называют *сопряженным* или *двойственным* пространством к \mathbb{R}^n .

Далее, каждому $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ можно сопоставить линейный функционал $a \mapsto a \cdot x$ на $(\mathbb{R}^n)^*$ и тогда совершенно аналогично устанавливается, что сопряженное (двойственное) пространство к $(\mathbb{R}^n)^*$ можно отождествить с \mathbb{R}^n .

Пусть $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$. Величина $|x| = \sqrt{x_1^2 + \dots + x_n^2}$ называется *длиной* или *евклидовой нормой* вектора x , а величина $d(x, y) = |x - y|$ — *расстоянием* между векторами x и y .

Пусть $\hat{x} \in \mathbb{R}^n$ и $\delta > 0$. Множество $U_{\mathbb{R}^n}(\hat{x}, \delta) = \{x \in \mathbb{R}^n \mid |x - \hat{x}| < \delta\}$ называется *открытым шаром* с центром в точке \hat{x} радиуса δ .

Множество $G \subset \mathbb{R}^n$ называется *открытым*, если с каждой своей точкой оно содержит и некоторый шар с центром в этой точке.

Любое открытое множество, содержащее данную точку, называется *окрестностью* этой точки.

Множество $F \subset \mathbb{R}^n$ называется *замкнутым*, если его дополнение $\mathbb{R}^n \setminus F$ — открытое множество.

Пусть $x_1, x_2 \in \mathbb{R}^n$. Множество $[x_1, x_2] = \{x \in \mathbb{R}^n \mid (1 - \alpha)x_1 + \alpha x_2, 0 \leq \alpha \leq 1\}$ называется *отрезком* (соединяющим точки x_1 и x_2). В случае, когда $n = 1, 2$ и 3 — это обычный отрезок.

Непустое множество $A \subset \mathbb{R}^n$ называется *выпуклым*, если с любыми двумя своими точками оно содержит и отрезок, соединяющий эти точки.

Если A — произвольное непустое подмножество \mathbb{R}^n , то наименьшее выпуклое множество, содержащее A называется *выпуклой оболочкой* A и обозначается со A .

Упражнение. Доказать, что со A — множество всех выпуклых комбинаций элементов из A , т.е. векторов вида $x = \alpha_1 x_1 + \dots + \alpha_m x_m$, где $x_i \in A$, $\alpha_i \geq 0$, $i = 1, \dots, m$, и $\sum_{i=1}^m \alpha_i = 1$.

Пусть $x^* \in (\mathbb{R}^n)^*$, $x^* \neq 0$ и $\gamma \in \mathbb{R}$. Множество $H = H(x^*, \gamma) = \{x \in \mathbb{R}^n \mid x^* \cdot x = \gamma\}$ называется *гиперплоскостью*. Гиперплоскость порождает два полупространства $H_+(x^*, \gamma) = \{x \in \mathbb{R}^n \mid x^* \cdot x \leq \gamma\}$ и $H_-(x^*, \gamma) = \{x \in \mathbb{R}^n \mid x^* \cdot x \geq \gamma\}$.

Пусть A и B — непустые подмножества \mathbb{R}^n . Говорят, что эти множества *отделимы*, если существует такая гиперплоскость, что A и B принадлежат различным полупространствам, порожденным этой гиперплоскостью.

Данное геометрическое определение отделимости равносильно, очевидно, следующему алгебраическому: множества A и B отделимы, если существует такой ненулевой элемент $x^* \in (\mathbb{R}^n)^*$, что

$$\sup_{a \in A} x^* \cdot a \leq \inf_{b \in B} x^* \cdot b.$$

Если неравенство строгое, то говорят, что множества A и B *строго отделимы*.

ТЕОРЕМА (МИНКОВСКОГО ОБ ОТДЕЛИМОСТИ). Пусть A — непустое выпуклое замкнутое подмножество \mathbb{R}^n и $b \notin A$. Тогда множество A и точка b строго отделимы.

Все перечисленные выше понятия (начиная с длины вектора и кончая теоремой Минковского) дословно переносятся на пространство $(\mathbb{R}^n)^*$ (двойственным к которому является пространство \mathbb{R}^n).

Пусть U — окрестность точки $\hat{x} \in \mathbb{R}^n$. Говорят, что функция $f: U \rightarrow \mathbb{R}$ дифференцируема в точке \hat{x} , если существует линейный функционал на \mathbb{R}^n , т. е. вектор $a = (a_1, \dots, a_n) \in (\mathbb{R}^n)^*$ такой, что для всех $h \in \mathbb{R}^n$, для которых $\hat{x} + h \in U$ справедливо представление

$$f(\hat{x} + h) = f(\hat{x}) + a \cdot h + r(h),$$

где $|r(h)|/|h| \rightarrow 0$ при $h \rightarrow 0$, или пишут $r(h) = o(h)$. Вектор a , определяемый этим представлением однозначно, называется *производной функции f в точке \hat{x}* и обозначается $f'(\hat{x})$.

Беря в качестве h векторы $(h_1, 0, \dots, 0)^T, \dots, (0, \dots, 0, h_n)^T$, получим, что $a_i = \partial f(\hat{x})/\partial x_i$ — частная производная функции f по x_i в точке \hat{x} , $i = 1, \dots, n$, и таким образом, $f'(\hat{x}) = (\partial f(\hat{x})/\partial x_1, \dots, \partial f(\hat{x})/\partial x_n)$.

ОСНОВНОЙ РЕЗУЛЬТАТ И ПРИЛОЖЕНИЕ

Пусть U — открытое подмножество \mathbb{R}^n , $f_i: U \rightarrow \mathbb{R}$, $i = 1, \dots, m$. Рассмотрим задачу

$$f(x) = \max_{1 \leq i \leq m} f_i(x) \rightarrow \min, \quad x \in U, \quad (P)$$

закключающуюся в нахождении таких точек $\hat{x} \in U$, где функция f достигает минимума. Нас будут интересовать локальные минимумы f . Точка

\hat{x} — локальный минимум функции f , если существует такая окрестность $U_0 \subset U$ точки \hat{x} , что $f(x) \geq f(\hat{x})$ для всех $x \in U_0$.

Перед формулировкой теоремы дадим одно определение. Пусть U — окрестность точки $\hat{x} \in \mathbb{R}^n$ и $f: U \rightarrow \mathbb{R}$. Скажем, что f *выпукла в точке* \hat{x} , если найдутся такие $x^* \in (\mathbb{R}^n)^*$ и окрестность $U_0 \subset U$ точки \hat{x} , что $f(x) - f(\hat{x}) \geq x^* \cdot (x - \hat{x})$ для всех $x \in U_0$. Другими словами, в окрестности U_0 график функции f лежит выше гиперплоскости $y = x^* \cdot (x - \hat{x}) + f(\hat{x})$.

ТЕОРЕМА (ОСНОВНАЯ). Пусть в задаче (P) функции $f_i, 1 \leq i \leq m$, дифференцируемы в $\hat{x} \in U$ и $f_1(\hat{x}) = \dots = f_m(\hat{x})$. Тогда условие $0 \in \text{co}\{f'_1(\hat{x}), \dots, f'_m(\hat{x})\}$ необходимо, а если функции $f_i, 1 \leq i \leq m$, выпуклы в \hat{x} , то и достаточно для того, чтобы \hat{x} было локальным минимумом функции f .

ДОКАЗАТЕЛЬСТВО. Необходимость. Пусть \hat{x} — локальный минимум функции f . Обозначим $A = \text{co}\{f'_1(\hat{x}), \dots, f'_m(\hat{x})\}$ и допустим, что $0 \notin A$. Тогда по теореме отделимости найдется такой вектор $\bar{x} \in \mathbb{R}^m$, что $\sup_{y \in A} y \cdot \bar{x} < 0$ или что тоже $\sup\{\sum_{i=1}^m \alpha_i f'_i(\hat{x}) \cdot \bar{x} \mid \alpha_i \geq 0, 1 \leq i \leq m, \sum_{i=1}^m \alpha_i = 1\} = \max_{1 \leq i \leq m} \{f'_i(\hat{x}) \cdot \bar{x}\} < 0$. Отсюда и в силу дифференцируемости функций $f_i, i = 1, \dots, m$, в точке \hat{x} следует, что для каждого $1 \leq i \leq m$ и всех достаточно малых $t > 0$ выполняется неравенство $f_i(\hat{x} + t\bar{x}) = f_i(\hat{x}) + t(f'_i(\hat{x}) \cdot \bar{x}) + o(t)/t < f_i(\hat{x})$, из которого вытекает, что $f(\hat{x} + t\bar{x}) < f(\hat{x})$ в противоречие с тем, что \hat{x} — локальный минимум f .

Достаточность. Пусть $0 \in A$. Тогда, очевидно, $\sup_{y \in A} y \cdot x \geq 0$ для всех $x \in \mathbb{R}^m$, что, как показано выше, равносильно неравенству $\max_{1 \leq i \leq m} \{f'_i(\hat{x}) \cdot x\} \geq 0$. Для каждого $1 \leq i \leq m$ по условию существует такой элемент $x_i^* \in (\mathbb{R}^n)^*$, что $f_i(x) - f_i(\hat{x}) \geq x_i^* \cdot (x - \hat{x})$ для x , близких к \hat{x} . Так как функция f_i дифференцируема в \hat{x} , то $x_i^* = f'_i(\hat{x})$. Действительно, для любого $h \in \mathbb{R}^n$ и достаточно малых $t > 0$ имеем $f_i(\hat{x} + th) - f_i(\hat{x}) \geq tx_i^* \cdot h$, или $t^{-1}(f_i(\hat{x} + th) - f_i(\hat{x})) \geq x_i^* \cdot h$. Переходя к пределу при $t \rightarrow 0$, получаем, что $f'_i(\hat{x}) \cdot h \geq x_i^* \cdot h$. Но h произвольно и поэтому $x_i^* = f'_i(\hat{x})$. Тогда для x , достаточно близких к \hat{x} , учитывая, что $f_1(\hat{x}) = \dots = f_m(\hat{x})$, будем иметь $f(x) - f(\hat{x}) = \max_{1 \leq i \leq m} \{f_i(x) - f_i(\hat{x})\} \geq \max_{1 \leq i \leq m} \{f'_i(\hat{x}) \cdot (x - \hat{x})\} \geq 0$, т. е. \hat{x} — локальный минимум функции f .

Отметим, что если функции $f_i, 1 \leq i \leq m$, дополнительно, выпуклы на \mathbb{R}^n , то необходимость может быть выведена из (весьма непростой) теоремы Дубовицкого — Милотина о субдифференциале максимума выпуклых функций (см. [1]). Постановка, которая рассмотрена здесь, представляется вполне естественной, но автор затрудняется дать какие-либо ссылки по этому поводу.

Приведем здесь одно приложение данного результата к экстремальным задачам. Пусть U — открытое подмножество $\mathbb{R}^n, f_i: U \rightarrow \mathbb{R}, i = 0, 1, \dots, m$.

Рассмотрим задачу

$$f_0(x) \rightarrow \min, \quad f_i(x) \leq 0, \quad i = 1, \dots, m. \quad (P_1)$$

Функция $\mathcal{L}: U \times \mathbb{R}^{m+1} \rightarrow \mathbb{R}$, $\mathcal{L}(x, \bar{\lambda}) = \sum_{i=0}^m \lambda_i f_i(x)$, где $\bar{\lambda} = (\lambda_0, \dots, \lambda_m)$, называется *функцией Лагранжа задачи* (P_1) , а вектор $\bar{\lambda}$ — набором *множителей Лагранжа*. Справедлива следующая

ТЕОРЕМА (ПРАВИЛО МНОЖИТЕЛЕЙ ЛАГРАНЖА ДЛЯ ЗАДАЧИ (P_1)). Пусть в задаче (P_1) все функции дифференцируемы в точке $\hat{x} \in U$. Тогда, если \hat{x} является локальным минимумом в этой задаче, то найдется такой ненулевой набор множителей Лагранжа $\bar{\lambda} = (\lambda_0, \dots, \lambda_m)$, что $\lambda_i \geq 0$, $i = 0, 1, \dots, m$, и

$$\mathcal{L}_x(\hat{x}, \bar{\lambda}) = 0 \Leftrightarrow \sum_{i=0}^m \lambda_i f'_i(\hat{x}) = 0.$$

ДОКАЗАТЕЛЬСТВО. Заметим сначала, что можно считать, что $f_1(\hat{x}) = \dots = f_m(\hat{x}) = 0$. В самом деле, отбросим те ограничения, для которых $f_i(\hat{x}) < 0$. Тогда \hat{x} будет локальным минимумом и в новой задаче. Если для нее доказано утверждение, то дополнив найденный набор множителей Лагранжа нулевыми компонентами, соответствующие тем номерам, где $f_i(\hat{x}) < 0$, получим утверждение в общем случае.

Если \hat{x} — локальный минимум в задаче (P_1) , то \hat{x} — локальный минимум и в такой задаче

$$f(x) = \max\{f_0(x) - f_0(\hat{x}), f_1(x), \dots, f_m(x)\} \rightarrow \min, \quad x \in U. \quad (i)$$

Действительно, если это не так, то в любой окрестности \hat{x} найдется точка \tilde{x} такая, что $f(\tilde{x}) < 0$ (поскольку $f(\hat{x}) = 0$). Это означает, что \tilde{x} — допустимая точка в задаче (P_1) и $f(\tilde{x}) < f(\hat{x})$, в противоречие с тем, что \hat{x} — локальный минимум. Итак, \hat{x} — локальный минимум в задаче (i) . Тогда по основной теореме найдутся такие $\lambda_i \geq 0$, $i = 0, 1, \dots, m$, $\sum_{i=0}^m \lambda_i = 1$, что $\sum_{i=0}^m \lambda_i f'_i(\hat{x}) = 0$, а это и есть утверждение теоремы.

СПИСОК ЛИТЕРАТУРЫ

- [1] Магарил-Ильяев Г. Г., Тихомиров В. М. *Выпуклый анализ и его приложения*. М.: Книжный дом «ЛИБРИКОМ», 2011. 3-е изд.