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 $\widehat{x}$ — локальный минимум максимума функций f_1 и f_2, где $f_1(\widehat{x}) = f_2(\widehat{x})$ и, скажем, $f_1'(\widehat{x}) \leqslant f_2'(\widehat{x})$, то справедливо включение $0 \in [f_1'(\widehat{x}), f_2'(\widehat{x})]$. Достаточным это условие не является, как показывает пример функции $f(x) = \max(-x, \sin x)$. Но оно становится достаточным, если в некоторой окрестности $\widehat{x}$ графики функций f_1 и f_2 лежат выше своих касательных в точке $\widehat{x}$.

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

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

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

Пусть $a = (a_1, \ldots, a_n)$ — вектор-строка из n действительных чисел. Для каждого $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ 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. $a \cdot (x_1 + x_2) = a \cdot x_1 + a \cdot x_1$ 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 $a = (l(e_1), \ldots, l(e_n))$, где $e_1 = (1, 0, \ldots, 0)^T, \ldots, e_n = (0, \ldots, 0, 1)^T$. 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 $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ 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 $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$. Величина $|x| = \sqrt{x_1^2 + \ldots + x_n^2}$ называется *длиной* или *евклидовой нормой* вектора x, а величина $d(x, y) = |x - y|$ — *расстоянием* между векторами x и y.

Пусть $\widehat{x} \in \mathbb{R}^n$ и $\delta > 0$. Множество $U_{\mathbb{R}^n}(\widehat{x}, \delta) = \{\, x \in \mathbb{R}^n \mid |x - \widehat{x}| < \delta \,\}$ называется *открытым шаром* с центром в точке $\widehat{x}$ 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 $[x_1, x_2] = \{\, x \in \mathbb{R}^n \mid (1 - \alpha)x_1 + \alpha x_2,\ 0 \leqslant \alpha \leqslant 1 \,\}$ 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 $\operatorname{co} A$.

Упражнение. Доказать, что $\operatorname{co} A$ — множество всех выпуклых комбинаций элементов из A, т. е. векторов вида $x = \alpha_1 x_1 + \ldots + \alpha_m x_m$, где $x_i \in A$, $\alpha_i \geqslant 0$, $i = 1, \ldots, m$, 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 $H_+(x^*, \gamma) = \{x \in \mathbb{R}^n \mid x^* \cdot x \leqslant \gamma\}$ и $H_-(x^*, \gamma) = \{x \in \mathbb{R}^n \mid x^* \cdot x \geqslant \gamma\}$.

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

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

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

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

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

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

Основной результат и приложение

Пусть U — открытое подмножество $\mathbb{R}^n$, $f_i \colon U \to \mathbb{R}$, $i = 1, \ldots, m$. Рассмотрим задачу

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

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

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

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

Теорема (основная). *Пусть в задаче* (P) *функции* f_i, $1 \leqslant i \leqslant m$, *дифференцируемы в* $\widehat{x} \in U$ *и* $f_1(\widehat{x}) = \ldots = f_m(\widehat{x})$. *Тогда условие* $0 \in \operatorname{co}\{\, f_1'(\widehat{x}), \ldots, f_m'(\widehat{x}) \,\}$ *необходимо, а если функции* f_i, $1 \leqslant i \leqslant m$, *выпуклы в* $\widehat{x}$, *то и достаточно для того, чтобы* $\widehat{x}$ *было локальным минимумом функции* f.

Доказательство. Необходимость. Пусть $\widehat{x}$ — локальный минимум функции f. Обозначим $A = \operatorname{co}\{\, f_1'(\widehat{x}), \ldots, f_m'(\widehat{x}) \,\}$ и допустим, что $0 \notin A$. Тогда по теореме отделимости найдется такой вектор $\overline{x} \in \mathbb{R}^m$, что $\sup_{y \in A} y \cdot \overline{x} < 0$ или что тоже $\sup\{ \sum_{i=1}^m \alpha_i f_i'(\widehat{x}) \cdot \overline{x} \mid \alpha_i \geqslant 0,\ 1 \leqslant i \leqslant m,\ \sum_{i=1}^m \alpha_i = 1 \,\} = \max_{1 \leqslant i \leqslant m}\{\, f_i'(\widehat{x}) \cdot \overline{x} \,\} < 0$. Отсюда и в силу дифференцируемости функций f_i, $i = 1, \ldots, m$, в точке $\widehat{x}$ следует, что для каждого $1 \leqslant i \leqslant m$ и всех достаточно малых $t > 0$ выполняется неравенство $f_i(\widehat{x} + t\overline{x}) = f_i(\widehat{x}) + t(f_i'(\widehat{x}) \cdot \overline{x} + o(t)/t) < f_i(\widehat{x})$, из которого вытекает, что $f(\widehat{x} + t\overline{x}) < f(\widehat{x})$ в противоречие с тем, что $\widehat{x}$ — локальный минимум f.

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

Отметим, что если функции f_i, $1 \leqslant i \leqslant m$, дополнительно, выпуклы на $\mathbb{R}^n$, то необходимость может быть выведена из (весьма непростой) теоремы Дубовицкого – Милютина 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 $\mathbb{R}^n$, $f_i\colon U \to \mathbb{R}$, $i = 0, 1, \ldots, m$.

Рассмотрим задачу

$$f_0(x) \to \min, \quad f_i(x) \leqslant 0, \quad i = 1, \dots, m. \tag{P_1}$$

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

Теорема (правило множителей Лагранжа для задачи (P_1)). *Пусть в задаче* (P_1) *все функции дифференцируемы в точке* $\widehat{x} \in U$. *Тогда, если* $\widehat{x}$ *является локальным минимумом в этой задаче, то найдется такой ненулевой набор множителей Лагранжа* $\overline{\lambda} = (\lambda_0, \dots, \lambda_m)$, *что* $\lambda_i \geqslant 0$, $i = 0, 1, \dots, m$, *и*

$$\mathcal{L}_x(\widehat{x}, \overline{\lambda}) = 0 \ \Leftrightarrow \ \sum_{i=0}^m \lambda_i f_i'(\widehat{x}) = 0.$$

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

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

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

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

Г. Г. Магарил-Ильяев, Московский государственный институт радиотехники, электроники и автоматики (технический университет)