

NONSMOOTH ANALYSIS ON THE PLANE. Part I

V. F. DEMYANOV

The problems of non-smooth analysis in the two-dimensional case are considered. In the classical ("smooth") mathematical analysis a domineering role is given to the gradient. In the non-smooth case the role of the gradient is played by the directional derivative. A special class of directionally differentiable functions – max-type functions – is discussed in detail.

Рассматриваются задачи негладкого анализа в двумерном случае. В классическом (гладком) математическом анализе основополагающую роль играет градиент. В негладком случае роль градиента играет производная по направлениям. Подробно обсуждается один класс дифференцируемых по направлениям функций – функции максимума.

НЕГЛАДКИЙ АНАЛИЗ НА ПЛОСКОСТИ. Часть I

В. Ф. ДЕМЬЯНОВ

Санкт-Петербургский государственный университет

ВВЕДЕНИЕ

В [1] обсуждались вопросы негладкого анализа функций одной вещественной переменной. Как и в классическом гладком математическом анализе, основные особенности и трудности при переходе к функциям многих переменных хорошо просматриваются уже в случае двух переменных. Поэтому изложение проводится для функций двух переменных. Для понимания дальнейшего знакомство с материалом статьи [1] необязательно. Негладкие функции возникают естественным образом при построении математических моделей, описывающих различные практические задачи.

Рассмотрим простейшую задачу обработки экспериментальных данных. В точках x_i , $i = 0, 1, 2, 3, 4$, заданы значения "наблюдений" y_i :

$$\begin{array}{cccccc} x_i & 0 & 1 & 2 & 3 & 4 \\ y_i & 0 & 1 & 2 & 3 & -1 \end{array}$$

Предполагая, что эти наблюдения описывают некоторую линейную зависимость, попробуем построить на двумерной плоскости xu прямую $y = y(x) = ax + b$, которая соответствует полученным наблюдениям. Поскольку точки $[x_i, y_i]$ не лежат на одной прямой, то требуется найти прямую, в некотором смысле наилучшим образом соответствующую имеющимся данным. Что значит наилучшим? На этот вопрос должен ответить экспериментатор, это не математический вопрос. Обычно выбирают какую-нибудь функцию параметров a и b и пытаются ее минимизировать. Среди инженеров для этой цели весьма популярен метод наименьших квадратов. Суть его в том, что строится функция

$$F_1(a, b) = \sum_{i=0}^4 (y(x_i) - y_i)^2 = \sum_{i=0}^4 (ax_i + b - y_i)^2$$

и находятся коэффициенты a и b , ее минимизирующие. В нашем случае минимум функции F_1 достигается при $a = 0$, $b = 1$, то есть прямая $y_1(x) = 1$ является наилучшей (для функции F_1).

Если в качестве минимизируемой функции выбрать суммарную ошибку

$$F_2(a, b) = \sum_{i=0}^4 |y(x_i) - y_i| = \sum_{i=0}^4 |ax_i + b - y_i|$$

и найти минимум этой функции, то окажется, что F_2 достигает наименьшего значения при $y_2(x) = x$ (то есть $a = 1, b = 0$).

Если же выбрать функцию наименьшего уклонения

$$F_3(a, b) = \max_i |ax_i + b - y_i|, \quad i \in \{0, 1, 2, 3, 4\},$$

и найти минимум этой функции, то ее минимум достигается при $y_3(x) = -\frac{1}{4}x + \frac{15}{8}$ (то есть $a = -\frac{1}{4}, b = \frac{15}{8}$). Решения $y_1(x), y_2(x), y_3(x)$ изображены на рис. 1 соответственно синим, красным и зеленым цветом.

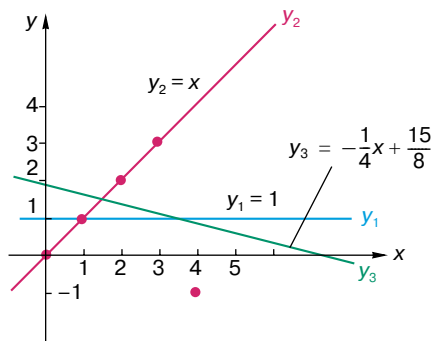


Рис. 1

Выбор конкретной функции в качестве критерия зависит от характера описываемого процесса. Так, если $|ax_i + b - y_i|$ есть штраф за нарушение графика, то, как следует из изложенного выше, выгоднее удовлетворить точки x_0, x_1, x_2, x_3 и игнорировать точку x_4 (и выбрать функцию F_2). Если же желаем минимизировать максимальное уклонение от графика, то наилучшим является решение $y_3(x)$, минимизирующее функцию F_3 .

Для нас здесь важно то, что, в то время как функция F_1 является непрерывно дифференцируемой, функции F_2 и F_3 недифференцируемые.

1. КЛАССИЧЕСКИЙ МАТЕМАТИЧЕСКИЙ АНАЛИЗ – ТЕАТР ОДНОГО АКТЕРА

Множество всех точек вида $z = [x, y]$ (где x и y – вещественные числа) называется двумерным пространством (или двумерной плоскостью) и обозначается \mathbf{R}^2 . Точки пространства \mathbf{R}^2 также называются векторами и направлениями (в зависимости от той роли, которую им приходится играть). Пусть на \mathbf{R}^2 задана функция $f(x, y) = f(z)$, принимающая вещественные значения.

Классический математический анализ имеет дело с дифференцируемыми функциями. Напомним

Определение. Функция f называется *дифференцируемой в точке* $z_0 = [x_0, y_0]$, если существует вектор $A = (a_1, a_2)$ такой, что имеет место разложение

$$f(z) = f(x, y) = f(z_0) + (A, z - z_0) + o(\|z - z_0\|). \quad (1)$$

Здесь

$$\frac{o(\|z - z_0\|)}{\|z - z_0\|} \xrightarrow{\|z - z_0\| \rightarrow 0} 0, \quad (2)$$

(A, B) обозначает скалярное произведение векторов A и B (если $A = [a_1, a_2], B = [b_1, b_2]$, то $(A, B) = a_1b_1 + a_2b_2$), $\|z\|$ – евклидова норма (длина) вектора z (если $z = [x, y]$, то $\|z\| = \sqrt{x^2 + y^2}$).

Если имеет место (1), то

$$a_1 = \lim_{\alpha \rightarrow 0} \frac{f(x_0 + \alpha, y_0) - f(x_0, y_0)}{\alpha} \equiv f'_x(x_0, y_0), \quad (3)$$

$$a_2 = \lim_{\alpha \rightarrow 0} \frac{f(x_0, y_0 + \alpha) - f(x_0, y_0)}{\alpha} \equiv f'_y(x_0, y_0). \quad (4)$$

Величина a_1 называется *частной производной функции f в точке z_0 по x* , а величина a_2 – *частной производной функции f в точке z_0 по y* . Таким образом, если функция f дифференцируема в точке z_0 , то она дифференцируема и по каждой из переменных x и y . К сожалению, существование частных производных еще не гарантирует дифференцируемости. Вектор $A = [f'_x(x_0, y_0), f'_y(x_0, y_0)]$ называется *градиентом* функции f в точке z_0 и обозначается $f'_z(z_0)$ или просто $f'(z_0)$. Негладкий анализ (НГА) имеет дело с функциями, которые не являются дифференцируемыми, а под гладким анализом теперь понимают (не совсем, строго говоря, обоснованно) классический математический анализ.

Пусть $g = [g_1, g_2] \in \mathbf{R}^2$. Говорят, что функция f *дифференцируема в точке z_0 по направлению g* , если существует конечный предел

$$\begin{aligned} f'(z_0, g) &= \lim_{\alpha \downarrow 0} \frac{f(z_0 + \alpha g) - f(z_0)}{\alpha} = \\ &= \lim_{\alpha \downarrow 0} \frac{f(x_0 + \alpha g_1, y_0 + \alpha g_2) - f(x_0, y_0)}{\alpha}. \end{aligned} \quad (5)$$

Здесь $\alpha \downarrow 0$ означает, что $\alpha \rightarrow 0$ и $\alpha > 0$. Величина $f'(z_0, g)$ называется *производной функции f в точке z_0 по направлению g* .

Положим $S_1 = \{g \in \mathbf{R}^2 \mid \|g\| = 1\}$. Это единичная окружность. Пусть функция f дифференцируема по всем направлениям. Направление $g_0 \in S_1$ называется *направлением наискорейшего спуска* функции f в точке z_0 , если

$$f'(z_0, g_0) = \min_{\|g\|=1} f'(z_0, g).$$

Аналогично направление $g^0 \in S_1$ называется *направлением наискорейшего подъема*, если

$$f'(z_0, g^0) = \max_{\|g\|=1} f'(z_0, g).$$

Без преувеличения можно сказать, что классический математический анализ — это театр одного актера и этим актером является градиент: понятие градиента (производной в одномерном случае) является наиболее важным понятием математического анализа. Отметим только некоторые задачи, которые решаются с использованием градиента. С его помощью можно:

1) найти производную функции f в точке z_0 по любому направлению g

$$f'(z_0, g) = (f'(z_0), g); \quad (6)$$

2) построить аппроксимацию функции f в окрестности точки z_0 (см. (1))

$$f(z) = f(z_0) + (f'(z_0), z - z_0) + o(\|z - z_0\|); \quad (7)$$

3) сформулировать необходимые условия минимума и максимума: для того чтобы в точке z^* функция f достигала своего наименьшего значения, необходимо, чтобы

$$f'(z^*) = [0, 0] = 0_2; \quad (8)$$

для того чтобы в точке z^{**} функция f достигала своего наибольшего значения, необходимо, чтобы

$$f'(z^{**}) = 0_2. \quad (9)$$

Здесь $0_2 = [0, 0]$. Отметим, что необходимые условия максимума и минимума совпадают. Условия (8) и (9) означают, что в точке максимума и минимума частные производные функции f обращаются в нуль. Если в точке z_0 оказалось $f'(z_0) = 0_2$, то точка z_0 называется *стационарной точкой функции f* ;

4) если точка z_0 еще не является стационарной (то есть $f'(z_0) \neq 0_2$), то можно найти направления наискорейшего спуска и подъема:

$$g_0 = -\frac{f'(z_0)}{\|f'(z_0)\|}, \quad g^0 = \frac{f'(z_0)}{\|f'(z_0)\|}. \quad (10)$$

Отметим, что направление наискорейшего спуска противоположно направлению наискорейшего подъема: $g_0 = -g^0$. Итак, существуют единственное направление наискорейшего спуска и единственное направление наискорейшего подъема. С помощью этих направлений можно построить численные методы для нахождения точек минимума или максимума. Из (6) и (7) вытекает также, что если $f'(z_0) \neq 0_2$, то прямая, проходящая через точку z_0 и перпендикулярная вектору $f'(z_0)$, делит плоскость на две полуплоскости: для любого направления g из той полуплоскости, куда смотрит вектор $f'(z_0)$, функция f возрастает (то есть при достаточно малых $\alpha > 0$ будет $f(z_0 + \alpha g) > f(z_0)$), а для любого g из

той полуплоскости, куда смотрит вектор $-f'(z_0)$, функция f убывает (рис. 2).

Принципиально важным является то, что указанные выше задачи решаются с использованием только трех чисел: значения функции $f(z_0)$ и двух частных производных $f'_x(z_0)$ и $f'_y(z_0)$. Существенно также, что имеется хорошо разработанное исчисление (правила вычисления градиентов).

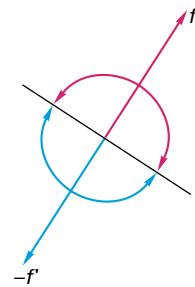


Рис. 2

2. ДИФФЕРЕНЦИРУЕМЫЕ ПО НАПРАВЛЕНИЯМ ФУНКЦИИ

Негладкий анализ имеет дело с недифференцируемыми функциями. Мы здесь ограничимся рассмотрением функций, дифференцируемых по направлениям. Класс этих функций достаточно широкий. Производная по направлениям (см. определение (5)) позволяет решать многие задачи, которые в гладком случае решались с помощью градиента. Так, необходимое условие минимума имеет вид

$$f'(z^*, g) \geq 0 \quad \forall g \in \mathbf{R}^2, \quad (11)$$

а необходимое условие максимума

$$f'(z^{**}, g) \leq 0 \quad \forall g \in \mathbf{R}^2. \quad (12)$$

Обратим внимание на то, что необходимые условия максимума (12) и минимума (11) оказались различными (в то время как в гладком случае необходимые условия минимума (8) и максимума (9) совпадали). Точка z^* , удовлетворяющая условию (11), называется *inf-стационарной*, а точка z^{**} , удовлетворяющая условию (12), называется *sup-стационарной*.

С помощью производной по направлениям можно построить аппроксимацию функции f в окрестности точки z_0 :

$$f(z) = f(z_0) + f'(z_0, z - z_0) + o(\|z - z_0\|).$$

В негладком случае может существовать несколько направлений наискорейшего спуска и несколько направлений наискорейшего подъема.

Пример 1. Пусть $f(z) = f(x, y) = |x| - |y|$, $z_0 = (0, 0)$. Эта функция не является дифференцируемой на прямых $x = y$ и $x = -y$. Возьмем направление $g = [g_1, g_2]$.

По формуле (5) находим производную по направлению g :

$$f'(z_0, g) = |g_1| - |g_2|.$$

Прямыми $g_1 = g_2$ и $g_1 = -g_2$ плоскость делится на четыре области (сами эти прямые не входят ни в одну из областей). На рис. 3 видно, что в областях I и III $|g_1| > |g_2|$, то есть $f'(z_0, g) > 0$. Следовательно, направления из областей I и III являются направлениями

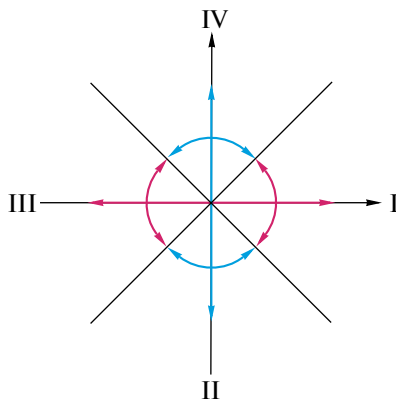


Рис. 3

подъема, в то время как в областях II и IV $|g_1| < |g_2|$, то есть $f'(z_0, g) < 0$, поэтому направления из областей II и IV являются направлениями спуска. При этом существуют два направления наискорейшего спуска: $g_0 = [0, 1]$ и $g'_0 = [0, -1]$ — и два направления наискорейшего подъема функции f в точке z_0 : $g^0 = [1, 0]$ и $g^{0n} = [-1, 0]$. Вспомним, что в гладком случае существовало только одно направление наискорейшего спуска и одно направление наискорейшего подъема.

3. ПОЧЕМУ НЕ СГЛАЖИВАНИЕ?

Один из распространенных и общепринятых способов борьбы с негладкостью состоит в замене негладкой функции гладкой (так называемое сглаживание).

Пример 2. Пусть, как и в примере 1, $f(x, y) = |x| - |y|$, $z_0 = [0, 0]$. Зафиксируем произвольные отличные от нуля числа $\alpha, \tau, \gamma, \mu$ и рассмотрим функцию

$$f_\epsilon(z) = \sqrt{(x - \alpha\epsilon)^2 + \alpha^2\tau^2\epsilon^2} - \sqrt{(y - \gamma\epsilon)^2 + \gamma^2\mu^2\epsilon^2}.$$

Здесь $\epsilon \neq 0$. Очевидно, что для любого z $f_\epsilon(z) \xrightarrow{\epsilon \rightarrow 0} f(z)$. Функция f_ϵ непрерывно дифференцируема, и ее градиент

$$f'_\epsilon(z) = \left(\frac{x - \alpha\epsilon}{\sqrt{(x - \alpha\epsilon)^2 + \alpha^2\tau^2\epsilon^2}}, \frac{-(y - \gamma\epsilon)}{\sqrt{(y - \gamma\epsilon)^2 + \gamma^2\mu^2\epsilon^2}} \right).$$

При этом в точке $z_0 = [0, 0]$

$$f'_\epsilon(z) = \left(\frac{-\text{sign } \alpha}{\sqrt{1 + \tau^2}}, \frac{\text{sign } y}{\sqrt{1 + \mu^2}} \right). \quad (13)$$

Здесь

$$\text{sign } a = \begin{cases} 1, & a > 0, \\ -1, & a < 0. \end{cases}$$

Из (10) и (13) заключаем, что любое наперед заданное направление $g \in S_1$ может быть (при соответствующем выборе параметров $\alpha, \tau, \gamma, \mu$) либо направлением наискорейшего спуска, либо направлением наискорейшего подъема для функции f_ϵ в точке z_0 .

Любое направление g из полуплоскости $(f'_\epsilon(z_0), g) > 0$ является направлением подъема, а любое направление g из полуплоскости $(f'_\epsilon(z_0), g) < 0$ является направлением спуска функции f_ϵ в точке z_0 , что не имеет ничего общего со свойствами функции $f(z)$.

Можно, конечно, возразить, что функция $f_\epsilon(z)$ неудачно подобрана. Однако никакая гладкая аппроксимация функции f не сможет обнаружить тот факт, что существуют два направления наискорейшего спуска и два направления наискорейшего подъема.

Следовательно, сглаживание не позволяет исследовать те свойства функции, которые связаны с понятием градиента. Тем не менее если построена гладкая аппроксимация, то ее можно использовать для исследования свойств исходной функции, не связанных с понятием градиента (например, если f_ϵ отличается от f не больше, чем на ϵ , и если найдена точка минимума функции f_ϵ , то значение функции f в этой точке отличается от минимального значения функции f не больше, чем на ϵ).

4. ФУНКЦИИ МАКСИМУМА

Из изложенного в разделе 2 следует, что в случае дифференцируемых по направлениям функций роль, аналогичную роли градиента в гладком анализе, выполняет производная по направлениям. Для решения многих теоретических задач это действительно так, однако при практическом использовании производных по направлениям возникают трудности, связанные с неконструктивным заданием производной по направлению: если в гладком случае мы находили только значения функции и ее частных производных в данной точке, то теперь необходимо уметь находить производную по любому направлению.

Для ряда классов дифференцируемых по направлениям функций удастся описать производную

по направлению более конструктивно. Одним из таких классов является класс функций максимума.

Пусть $f_i(z), i \in \{1, 2, \dots, N\}$, — непрерывно дифференцируемые функции. Положим

$$f(z) = \max_{i \in I} f_i(z),$$

где $I = \{1, 2, \dots, N\}$. Функция f уже не является непрерывно дифференцируемой, но она является дифференцируемой по направлениям:

$$f'(z, g) = \max_{i \in R(z)} (f'_i(z), g), \quad (14)$$

где $R(z) = \{i \in I \mid f'_i(z) = f'(z, g)\}$.

Выражение (14) можно переписать в виде

$$f'(z, g) = \max_{v \in \partial f(z)} (v, g), \quad (15)$$

где $\partial f(z) = \text{co}\{f'_i(z) \mid i \in R(z)\}$. Здесь $\text{co}\{A_i \mid i \in I\}$ — выпуклая оболочка векторов $\{A_i\}$ (в нашем случае выпуклый многоугольник, натянутый на точки $\{A_i\}$).

Обратим внимание на то, что в (14) и (15) множества $R(z)$ и $\partial f(z)$ не зависят от направления g , то есть для вычисления производной по направлениям (как функции направления g) необходимо вычислить и хранить только градиенты функций f_i (и то лишь для $i \in R(z)$).

Необходимое условие минимума (11) в данном случае (с учетом (15)) эквивалентно условию

$$0_2 \in \partial f(z^*), \quad (16)$$

а необходимое условие максимума (12) эквивалентно условию

$$f'_i(z^{**}) = 0_2 \quad \forall i \in R(z^{**}). \quad (17)$$

Если в точке z_0 необходимое условие минимума (16) еще не выполнено, то есть $0_2 \notin \partial f(z_0)$, то найдем

$$\min_{v \in \partial f(z)} \|v\| = \|v_0\| > 0.$$

Направление (рис. 4)

$$g_0 = -\frac{v_0}{\|v_0\|} \quad (18)$$

является направлением наискорейшего спуска функции f в точке z_0 . Такое направление единственно.

Если же в точке z_0 не выполнено необходимое условие максимума (17), то найдем

$$\max_{i \in R(z_0)} \|f'_i(z_0)\| = \|f'_{i_0}(z_0)\|. \quad (19)$$

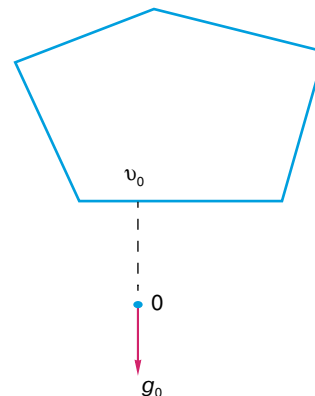


Рис. 4

Тогда направление

$$g^0 = \frac{f'_{i_0}(z_0)}{\|f'_{i_0}(z_0)\|}$$

является направлением наискорейшего подъема функции f в точке z_0 . Поскольку максимум в (19) может достигаться на нескольких индексах, то и направлений наискорейшего подъема тоже может оказаться несколько.

Пример 3. Пусть $f(z) = \max\{x + y, x - y\}$. Положим $f_1(z) = x + y, f_2(z) = x - y, I = \{1, 2\}$. Имеем $f(z) = \max_{i \in I} f_i(z)$. Возьмем $z_0 = [0, 0]$. Тогда $f(z_0) = f_1(z_0) = f_2(z_0) = 0$, то есть $R(z_0) = \{1, 2\} = I$. Так как $f'_1(z) = [1, 1], f'_2(z) = [1, -1]$, то (рис. 5, а) $\partial f(z_0) = \text{co}\{[1, 1], [1, -1]\}$. Поскольку $0_2 \notin \partial f(z_0)$, то условие (17) не выполнено.

Найдем

$$\min_{v \in \partial f(z_0)} \|v\| = \|v_0\|,$$

где $v_0 = [1, 0]$. Направление (см. (18)) $g_0 = [-1, 0]$ является направлением наискорейшего спуска. Условие (17) тоже не выполнено в точке z_0 . Поскольку

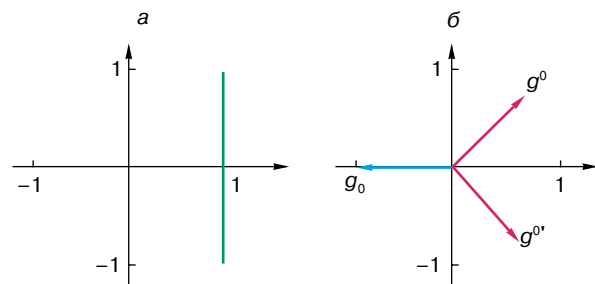


Рис. 5

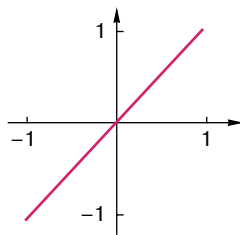


Рис. 6

$\max_{i \in R(z_0)} \|f'_i(z_0)\|$ достигается в двух точках $f'_1(z_0)$ и $f'_2(z_0)$, то направления

$$g^0 = \frac{f'_1(z_0)}{\|f'_1(z_0)\|} = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right],$$

$$g^{0,} = \frac{f'_2(z_0)}{\|f'_2(z_0)\|} = \left[\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right]$$

являются направлениями наискорейшего подъема (рис. 5, б).

Пример 4. Пусть $f(z) = \max\{x + y, -x - y\} = |x + y|$. Положим $f_1(z) = x + y$, $f_2(z) = -x - y$, $I = \{1, 2\}$. Имеем $f(z) = \max_{i \in I} f_i(z)$. Возьмем $z_0 = [0, 0]$. Имеем $f(z_0) = f_1(z_0) = f_2(z_0) = 0$, то есть $R(z_0) = \{1, 2\} = I$. Так как $f'_1(z) = [1, 1]$, $f'_2(z) = [-1, -1]$, то $\partial f(z_0) = \text{co}\{[1, 1], [-1, -1]\}$.

Поскольку (рис. 6) $0_2 \in \partial f(z_0)$, то в точке z_0 выполнено необходимое условие минимума, то есть точка z_0 inf-стационарная (в действительности точка z_0 является точкой минимума, но наша теория позволяет лишь утверждать, что z_0 – inf-стационарная точка). Необходимое условие максимума (17) в точке z_0 не выполнено. Направления

$$g^0 = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right] \text{ и } g^{0,} = \left[-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right]$$

являются направлениями наискорейшего подъема.

ЗАКЛЮЧЕНИЕ

Для функции максимума в каждой точке z существует выпуклое множество (в данном случае многоугольник) $\partial f(z)$, которое позволяет описать необходимые условия максимума и минимума, найти направления наискорейшего спуска и подъема, построить аппроксимацию функции, вычислить производную по направлениям. Множество $\partial f(z)$ называется субдифференциалом функции f в точке z . Оно играет роль, которую в гладком случае выполняет градиент. Оказывается, что аналогичное множество ∂f существует и для другого важного класса негладких функций – выпуклых функций.

ЛИТЕРАТУРА

1. Демьянов В.Ф. Обобщение понятия производной в негладком анализе // Соросовский Образовательный Журнал. 1996. № 5. С. 121–127.
2. Демьянов В.Ф., Малоземов В.Н. Введение в минимакс. М.: Наука, 1972. 368 с.
3. Рокафеллар Р. Выпуклый анализ. М.: Мир, 1973. 472 с.
4. Пшеничный Б.Н. Выпуклый анализ и экстремальные задачи. М.: Наука, 1980. 320 с.

* * *

Владимир Федорович Демьянов, доктор физико-математических наук, профессор, зав. кафедрой факультета прикладной математики – процессов управления Санкт-Петербургского государственного университета. Область научных интересов: оптимальное управление, математическое программирование, негладкий анализ, недифференцируемая оптимизация. Член редколлегии четырех международных математических журналов, автор более 100 работ, в том числе семи монографий, часть из которых переведена на английский, немецкий, польский и китайский языки.