Математический анализ 1. Лекция 2.9
Выпуклые функции.
Скалярные функции, заданные неявно

30 ноября 2023 г.

Выпуклые множества и функции.

Определение. 1 (напоминание). Множество $D \subset \mathbb{R}^n$ называется **выпуклым**, если вместе с каждыми двумя точками $\mathbf{x}$ и $\mathbf{y}$ оно содержит соединяющий их отрезок:

$$\mathbf{x}, \mathbf{y} \in D \Rightarrow \alpha\mathbf{x} + (1-\alpha)\mathbf{y} \in D \text{ при всех } \alpha \in [0,1].$$

2. Скалярная функция $f : X \to \mathbb{R}$, заданная на выпуклом множестве $X \subset \mathbb{R}^n$, называется:

- **выпуклой**, если

  $$f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \leqslant \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y})$$

  для всех $\mathbf{x}, \mathbf{y} \in X$ и $\alpha \in [0,1]$;
- **вогнутой**, если

  $$f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \geqslant \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y})$$

  для всех $\mathbf{x}, \mathbf{y} \in X$ и $\alpha \in [0,1]$.
- **строго выпуклой**, если

  $$f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) < \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y})$$

  для всех $\mathbf{x}, \mathbf{y} \in X$, $\mathbf{x} \neq \mathbf{y}$ и $\alpha \in (0,1)$;
- **строго вогнутой**, если

  $$f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) > \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y})$$

  для всех $\mathbf{x}, \mathbf{y} \in X$, $\mathbf{x} \neq \mathbf{y}$ и $\alpha \in (0,1)$.

Замечание

1. Функция f – выпуклая тогда и только тогда, когда функция $(-f)$ – вогнутая.
2. Функция f – строго выпуклая тогда и только тогда, когда функция $(-f)$ – строго вогнутая.

Простейшими примерами строго выпуклой и нестрого выпуклой функции служат соответственно $f(\mathbf{x}) = |\mathbf{x}|^2$ и $g(\mathbf{x}) = \max\{|\mathbf{x}|^2, 1\}$,
а строго вогнутой и нестрого вогнутой функций – те же функции, взятые со знаком минус.

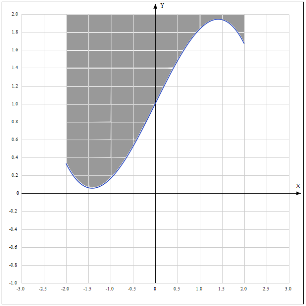
Определение. Для любой скалярной функции $f : X \to \mathbb{R}$, где $X = D(f) \subset \mathbb{R}^n$,

- ее *надграфиком* называется множество

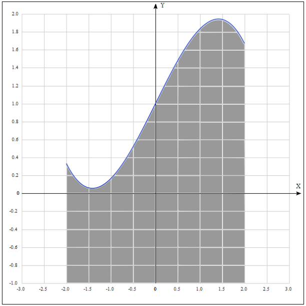
$$\operatorname{epi} f = \{(\mathbf{x}, y) \in X \times \mathbb{R} : y \geqslant f(\mathbf{x})\},$$

- ее *подграфиком* называется множество

$$\operatorname{hyp} f = \{(\mathbf{x}, y) \in X \times \mathbb{R} : y \leqslant f(\mathbf{x})\}.$$



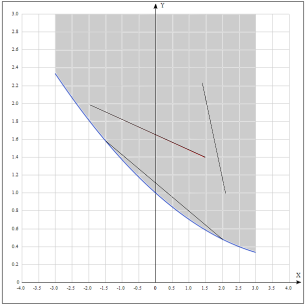
epi f



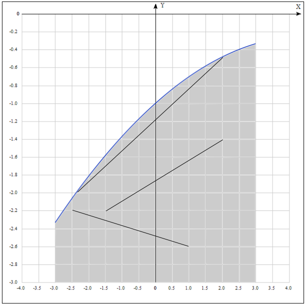
hyp f

Теорема

- *Скалярная функция $f : X \to \mathbb{R}$, заданная на выпуклом множестве $X \subset \mathbb{R}^n$, является*
  - *выпуклой тогда и только тогда, когда ее надграфик выпуклый;*
  - *вогнутой тогда и только тогда, когда ее подграфик выпуклый.*



$\mathrm{epi}\, f$ выпуклое множество, функция f выпуклая.



$\mathrm{hyp}\, f$ выпуклое множество, функция f вогнутая.

Определение. Квадратичная форма

$$(A\mathbf{h}, \mathbf{h})_{\mathbb{R}^n} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} h_j h_i$$

называется **неотрицательно определенной**, если

$$(A\mathbf{h}, \mathbf{h})_{\mathbb{R}^n} \geqslant 0 \quad \text{при всех} \quad \mathbf{h} \in \mathbb{R}^n,$$

и называется **неположительно определенной**, если

$$(A\mathbf{h}, \mathbf{h})_{\mathbb{R}^n} \leqslant 0 \quad \text{при всех} \quad \mathbf{h} \in \mathbb{R}^n.$$

Эти свойства называются также соответственно неотрицательностью $A \geqslant 0$ и неположительностью $A \leqslant 0$ матрицы A.

Очевидно, что свойства $A \geqslant 0$ и $-A \leqslant 0$ эквивалентны.

Для симметричной матрицы A свойство $A \geqslant 0$ эквивалентно неотрицательности всех ее собственных значений, а свойство $A \leqslant 0$ эквивалентно неположительности всех ее собственных значений.

В частности, для диагональной матрицы A свойство $A \geqslant 0$ эквивалентно неотрицательности ее диагональных элементов.

Теорема (критерии выпуклости и вогнутости функций)

Пусть даны открытое выпуклое множество $D \subset \mathbb{R}^n$ и функция $f \in C^2(D)$. Тогда:

1. *Функция f **выпукла** на D тогда и только тогда, когда 2-й дифференциал $d^2 f(\mathbf{x})(\mathbf{h})$ – неотрицательно определенная квадратичная форма (переменных $\mathbf{h}$) при всех $\mathbf{x} \in D$.*
2. *Если 2-й дифференциал $d^2 f(\mathbf{x})(\mathbf{h})$ – положительно определенная квадратичная форма (переменных $\mathbf{h}$) при всех $\mathbf{x} \in D$, то функция f **строго выпукла** на D (обратное, вообще говоря, неверно).*
3. *Функция f **вогнута** на D тогда и только тогда, когда 2-й дифференциал $d^2 f(\mathbf{x})(\mathbf{h})$ – неположительно определенная квадратичная форма (переменных $\mathbf{h}$) при всех $\mathbf{x} \in D$.*
4. *Если 2-й дифференциал $d^2 f(\mathbf{x})(\mathbf{h})$ – отрицательно определенная квадратичная форма (переменных $\mathbf{h}$) при всех $\mathbf{x} \in D$, то функция f **строго вогнута** на D (обратное, вообще говоря, неверно).*

Вместо 2-го дифференциала можно говорить о знаке матрицы Гессе в точках $\mathbf{x} \in D$.

Теорема

Пусть функция f выпукла (вогнута) на открытом выпуклом множестве D и непрерывна на замыкании $\overline{D}$ множества D. Тогда множество $\overline{D}$ выпукло, а функция f выпукла (соответственно, вогнута) на множестве $\overline{D}$.

Отметим, что строго выпуклая на D функция может быть нестрого выпуклой на $\overline{D}$.

Пример. Рассмотрим функцию

$$f(\mathbf{x}) = (A\mathbf{x}, \mathbf{x})_{\mathbb{R}^n} + (\mathbf{b}, \mathbf{x})_{\mathbb{R}^n} + c$$

с симметричной матрицей A и некоторыми $\mathbf{b} \in \mathbb{R}^n$ и числом c. Она

выпукла на $\mathbb{R}^n$ тогда и только тогда, когда $A \geqslant 0$,

строго выпукла на $\mathbb{R}^n$ тогда и только тогда, когда $A > 0$,

вогнута на $\mathbb{R}^n$ тогда и только тогда, когда $A \leqslant 0$,

строго вогнута на $\mathbb{R}^n$ тогда и только тогда, когда $A < 0$.

Экономический пример. Рассмотрим функцию Кобба-Дугласа

$$f(x,y) = Ax^{\alpha}y^{\beta} \text{ при } x \geqslant 0, y \geqslant 0,$$

где $A > 0, \alpha > 0, \beta > 0$ – параметры. Требуется исследовать ее на выпуклость/вогнутость.

Поскольку функция f имеет непрерывные 2-е производные в каждой внутренней точке своей области определения (т.е. при $x > 0$, $y > 0$), то достаточно исследовать на знакоопределенность ее матрицу Гессе с помощью критерия Сильвестра.
Матрица Гессе функции f в точке (x, y) имеет вид

$$H_f(x,y) = \begin{pmatrix} A\alpha(\alpha-1)x^{\alpha-2}y^{\beta} & A\alpha\beta x^{\alpha-1}y^{\beta-1} \\ A\alpha\beta x^{\alpha-1}y^{\beta-1} & A\beta(\beta-1)x^{\alpha}y^{\beta-2} \end{pmatrix}.$$

Ее главные угловые миноры таковы

$$\Delta_1 = A\alpha(\alpha-1)x^{\alpha-2}y^{\beta},$$
$$\Delta_2 = A^2\alpha\beta x^{2\alpha-2}y^{2\beta-2}(1-\alpha-\beta).$$

С учетом положительности параметров и переменных выводим:

1. При $\alpha+\beta<1$ для всех $(x,y)\in\mathbb{R}^2_+=\{(x,y):x>0,y>0\}$ имеем
$$\Delta_1<0,\Delta_2>0.$$
Значит, 2-й дифференциал $d^2f(x,y)(h_1,h_2)$ функции f – отрицательно определенная квадратичная форма в каждой точке $(x,y)\in\mathbb{R}^2_+$. Поэтому, во-первых, функция f строго вогнута на множестве $\mathbb{R}^2_+$ и, значит, во-вторых (по последней теореме), вогнута на всей своей области определения $D(f)=\overline{\mathbb{R}}^2_+=\{(x,y):x\geqslant 0,\ y\geqslant 0\}$.
2. При $\alpha+\beta>1$ для всех $(x,y)\in\mathbb{R}^2_+$ имеем
$$\Delta_2<0.$$
Значит, 2-й дифференциал $d^2f(x,y)(h_1,h_2)$ функции f – знакопеременная квадратичная форма в каждой точке $(x,y)\in\mathbb{R}^2_+$. Поэтому функция f не выпукла и не вогнута ни на одном выпуклом открытом подмножестве в $\mathbb{R}^2_+$.
3. При $\alpha+\beta=1$ можно 2-й дифференциал функции f в любой точке $(x,y)\in\mathbb{R}^2_+$ преобразовать так (**проверьте!**):
$$d^2f(x,y)(h_1,h_2)=-A\alpha\beta x^{\alpha-2}y^{\beta-2}\left(yh_1-xh_2\right)^2\leqslant 0,$$
т.е. это неположительно определенная форма. Поэтому функция f вогнута на множестве R^2_+, а, значит, и на всей своей области определения $D(f)=\overline{\mathbb{R}}^2_+$.

Теоремы об экстремумах выпуклой функции

Определение (точек локального экстремума на множестве). Для функции $f : X \to \mathbb{R}$, $X = D(f) \in \mathbb{R}^n$ точка $\mathbf{x}_0 \in X$ называется:
1) **точкой строгого локального максимума на** X, если $f(\mathbf{x}_0) > f(\mathbf{x})$ для всех $\mathbf{x}$ из пересечения X и некоторой проколотой окрестности $\mathcal{U}$ точки $\mathbf{x}_0$;
2) **точкой строгого локального минимума на** X, если $f(\mathbf{x}_0) < f(\mathbf{x})$ для всех $\mathbf{x}$ из пересечения X и некоторой проколотой окрестности $\mathcal{U}$ точки $\mathbf{x}_0$;
3) **точкой (нестрогого) локального максимума на** X, если $f(\mathbf{x}_0) \geqslant f(\mathbf{x})$ для всех $\mathbf{x}$ из пересечения X и некоторой окрестности $\mathcal{U}$ точки $\mathbf{x}_0$;
4) **точкой (нестрогого) локального минимума на** X, если $f(\mathbf{x}_0) \leqslant f(\mathbf{x})$ для всех $\mathbf{x}$ из пересечения X и некоторой окрестности $\mathcal{U}$ точки $\mathbf{x}_0$.

Теорема (1-я теорема об экстремумах выпуклой функции)

Пусть даны выпуклое множество D и функция $f : D \to \mathbb{R}$. Тогда:

1. *если функция f (строго) выпукла на D, то всякая точка $\mathbf{x} \in D$ (строгого) локального минимума функции f на множестве D есть ее точка (строгого)* ***глобального минимума*** *на D $\Leftrightarrow$*
2. *если функция f (строго) вогнута на D, то всякая точка $\mathbf{x} \in D$ (строгого) локального максимума функции f есть ее точка (строгого)* ***глобального максимума*** *на D.*

Теорема (2-я теорема об экстремумах выпуклой функции)

Пусть даны выпуклое открытое множество D и функция $f \in C^1(D)$. Тогда:

1. *если функция f (строго) выпукла на D, то всякая стационарная точка $\mathbf{x} \in D$ функции f есть ее точка (строгого)* ***глобального минимума*** *на D $\Leftrightarrow$*
2. *если функция f (строго) вогнута на D, то всякая стационарная точка $\mathbf{x} \in D$ функции f есть ее точка (строгого)* ***глобального максимума*** *на D.*

Теорема (3-я теорема об экстремумах выпуклой функции)

Максимум выпуклой функции и минимум вогнутой функции на выпуклом множестве D могут достигаться только на границе множества D.

Пример

Найдем точку глобального максимума функции $f(x,y) = x^{0.1}y^{0.5} - x - 5y$ на множестве $\mathbb{R}^2_+$, если она существует.

Приравнивая к нулю частные производные, находим стационарные точки:

$$\begin{cases} 0.1x^{-0.9}y^{0.5} - 1 = 0 \\ 0.5x^{0.1}y^{-0.5} - 5 = 0 \end{cases} \Rightarrow x = y = \frac{1}{10^{5/2}}.$$

Поскольку $0.1 + 0.5 < 1$, то по доказанному выше функция $f(x,y) = x^{0.1}y^{0.5}$ строго вогнута на множестве $\mathbb{R}^2_+$. Добавление линейной функции не влияет на (строгую) выпуклость/вогнутость (**упражнение:** почему?). Значит, функция f строго вогнутая, и, следовательно, точка $\left(\frac{1}{10^{5/2}}, \frac{1}{10^{5/2}}\right)$ есть точка ее строгого глобального максимума на множестве $\mathbb{R}^2_+$.

Замечание

Функция полезности обычно достаточно гладкая и вогнутая.

Теорема о неявной функции: простейший случай

В линейной алгебре изучаются системы **линейных уравнений**. Мы начинаем изучение **нелинейных уравнений и систем уравнений**.

Рассмотрим скалярную функцию F, заданную на некотором множестве D на плоскости, и множество точек на плоскости, которое задается уравнением

$$F(x, y) = 0, \quad (x, y) \in D.$$

В общем случае зависимость между переменными x и y, которая определяется этим уравнением, не является функциональной.

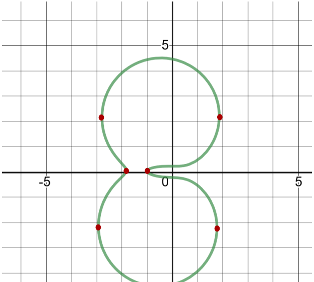
Пример 1. Уравнение $f(y) - x = 0$ для построения обратной функции $y = f^{(-1)}(x)$.

Пример 2. Нелинейное уравнение с параметром $x^2 + y^2 = c$, где c – параметр: анализ на доске.

Пример 3. На рисунке изображена кривая, заданная уравнением

$$\left(x^2 + y^2 + x\right)^2 - \left(x^2 + xy + 20y^2\right) + 1 = 0.$$

В окрестности каждой точки, кроме шести выделенных, уравнение задает функциональную зависимость y от x.



Теорема

Пусть выполнены условия:

1. *скалярная функция F определена и имеет непрерывные частные производные в некоторой окрестности $\mathcal{O}$ точки $(x_0, y_0) \in \mathbb{R}^2$,*
2. *точка (x_0, y_0) такова, что $F(x_0, y_0) = 0$,*
3. *$F'_y(x_0, y_0) \neq 0$.*

Тогда при достаточно малых $\delta > 0$, $\varepsilon > 0$ для каждого $x \in (x_0 - \delta, x_0 + \delta)$ существует единственное решение уравнения

$$F(x, y) = 0, \quad \text{где} \quad |y - y_0| < \varepsilon,$$

которое определяет функцию $y = f(x)$ такую, что $y_0 = f(x_0)$.
Эта функция f обращает уравнение в тождество
$F(x, f(x)) \equiv 0$ для всех $x \in (x_0 - \delta, x_0 + \delta)$.
Функция f имеет непрерывную производную

$$f'(x) = -\left.\frac{F'_x(x, y)}{F'_y(x, y)}\right|_{y=f(x)}$$

для всех $x \in (x_0 - \delta, x_0 + \delta)$ (вывод: на доске).

Кроме того, если $F \in C^k(\mathcal{O})$, то $f \in C^k(x_0 - \delta, x_0 + \delta)$ для любого $k \geqslant 1$.

Замечания.

- Условие $F'_y(x_0, y_0) \neq 0$ и достаточная малость δ существенны: на доске.
- Производную неявно заданной скалярной функции одной вещественной переменной можно вычислять и непосредственно в силу заданного уравнения.
  Пример: $x^2 + y^2 = 1 \;\Rightarrow$
  $$x^2 + y^2(x) \equiv 1 \;\Rightarrow\; 2x + 2y(x)y'(x) = 0 \;\Rightarrow\; y'(x) = -\frac{x}{y(x)}$$
  (здесь предполагается, что уравнение $x^2 + y^2 = 1$ локально задает функциональную зависимость $y = y(x)$).
  К этому же результату, естественно, приводит и формула из теоремы (поскольку она так и выводится):
  $$y'(x) = -\frac{(x^2 + y^2 - 1)'_x}{(x^2 + y^2 - 1)'_y} = -\frac{x}{y(x)}.$$
- Последний пункт теоремы дает существование производных высших порядков неявно заданной функции. **Два способа вычисления:**
  $$x^2 + y^2 = 1 \;\Rightarrow\; y' = -\frac{x}{y} \Rightarrow y'' = -\frac{y - xy'}{y^2} = -\frac{x^2 + y^2}{y^3} = -\frac{1}{y^3} \quad \text{или}$$
  $$x^2 + y^2 = 1 \;\Rightarrow\; 2yy' + 2x = 0 \Rightarrow (y')^2 + yy'' + 1 = 0 \Rightarrow \ldots \Rightarrow y'' = -\frac{1}{y^3}.$$