Выпуклость функций

Опр: 1. f выпукла на (a,b), если $\forall x,y\in(a,b)\wedge\forall\alpha\in[0,1],\ f(\alpha x+(1-\alpha)y)\le\alpha f(x)+(1-\alpha)f(y)$.

Rm: 1. Геометрически выпуклость $\Leftrightarrow$ любая хорда лежит выше графика $\Leftrightarrow$ тангенсы углов наклона хорд не убывают, если их располагать вдоль графика.

Теорема 1. Пусть f - дифференцируема на интервале (a,b), тогда следующие утверждения эквивалентны:

(1) f - выпукла на (a,b);

(2) f' - не убывает на (a,b);

(3) $f(x)\ge f(y)+f'(y)(x-y)$, $\forall x,y,\in(a,b)$;

$\square$

(3) $\Rightarrow$ (1): Возьмем точки $x_1,x,x_2\in(a,b)\colon x_1<x<x_2$. Докажем, что $\dfrac{f(x)-f(x_1)}{x-x_1}\le\dfrac{f(x_2)-f(x)}{x_2-x}$.

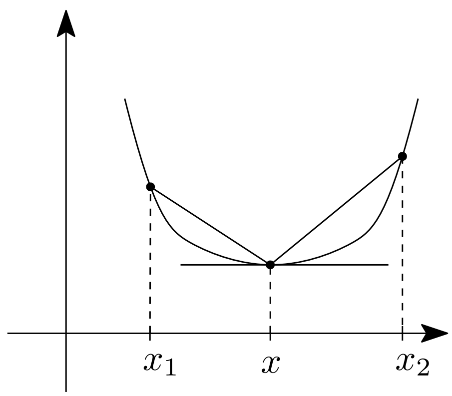

Рис. 1: Сравнение наклона хорд с наклоном касательной в точке x.

Рассмотрим точки x и $x_1\Rightarrow$ по условию $f(x_1)\ge f(x)+f'(x)(x_1-x)\Rightarrow f'(x)(x-x_1)\ge f(x)-f(x_1)\Rightarrow$ $\Rightarrow$ поскольку $x>x_1\Rightarrow f'(x)\ge\dfrac{f(x)-f(x_1)}{x-x_1}$.

Рассмотрим точки x_2 и $x\Rightarrow$ по условию $f(x_2)\ge f(x)+f'(x)(x_2-x)\Rightarrow f(x_2)-f(x)\ge f'(x)(x_2-x)\Rightarrow$ $\Rightarrow$ поскольку $x_2>x\Rightarrow\dfrac{f(x_2)-f(x)}{x_2-x}\ge f'(x)$.

Таким образом мы получаем, что $\dfrac{f(x)-f(x_1)}{x-x_1}\le f'(x)\le\dfrac{f(x_2)-f(x)}{x_2-x}$. $\blacksquare$

Следствие 1. Пусть f дважды дифференцируема на интервале (a,b). Тогда f - выпукла $\Leftrightarrow f''\ge0$.

$\square$ f' - не убывает $\Leftrightarrow f''\ge0$. $\blacksquare$

Неравенство Йенсена

Теорема 2. (Неравенство Йенсена) Пусть f выпукла на (a,b), $\alpha_1,\dots,\alpha_n\in[0,1]$: $\sum\limits_{k=1}^{n}\alpha_k=1$, $x_1,\dots,x_n\in(a,b)$, тогда $f(\alpha_1x_1+\ldots+\alpha_nx_n)\le\alpha_1f(x_1)+\ldots+\alpha_nf(x_n)$.

□ Индукцией по n (все $\alpha_i\ge0$):
База: $n=1\Rightarrow\alpha_1=1$, $f(x_1)\le f(x_1)$; $n=2\Rightarrow\alpha_1+\alpha_2=1$, $f(\alpha_1x_1+\alpha_2x_2)\le\alpha_1f(x_1)+\alpha_2f(x_2)$ - по определению выпуклости.
Шаг: Пусть $S_n=\sum\limits_{k=1}^{n}\alpha_k$ и утверждение доказано для n, докажем для $n+1$ (случай, когда $\alpha_1=\ldots=\alpha_n=0$, $\alpha_{n+1}=1$ очевиден):

$$f(\alpha_1x_1+\ldots+\alpha_nx_n+\alpha_{n+1}x_{n+1})=f\Big((\alpha_1+\ldots+\alpha_n)\Big(\frac{\alpha_1}{S_n}x_1+\ldots+\frac{\alpha_n}{S_n}x_n\Big)+\alpha_{n+1}x_{n+1}\Big)$$

Из-за выпуклости f и $\sum\limits_{k=1}^{n+1}\alpha_k=S_n+\alpha_{n+1}=1$ получим:

$$f(\alpha_1x_1+\ldots+\alpha_{n+1}x_{n+1})=f\Big(S_n\Big(\frac{\alpha_1}{S_n}x_1+\ldots+\frac{\alpha_n}{S_n}x_n\Big)+\alpha_{n+1}x_{n+1}\Big)\le S_nf\Big(\frac{\alpha_1}{S_n}x_1+\ldots+\frac{\alpha_n}{S_n}x_n\Big)+\alpha_{n+1}f(x_{n+1})$$

где $\beta_1=\dfrac{\alpha_1}{\alpha_1+\ldots+\alpha_n}\ge0,\dots,\beta_n=\dfrac{\alpha_n}{\alpha_1+\ldots+\alpha_n}\ge0$, $\sum\limits_{k=1}^{n}\beta_n=1\Rightarrow$ воспользуемся предположением индукции:

$$S_nf\Big(\frac{\alpha_1}{S_n}x_1+\ldots+\frac{\alpha_n}{S_n}x_n\Big)+\alpha_{n+1}f(x_{n+1})\le S_n\Big(\beta_1f(x_1)+\ldots+\beta_nf(x_n)\Big)+\alpha_{n+1}f(x_{n+1})=$$

$$=S_n\cdot\frac{\alpha_1}{S_n}f(x_1)+\ldots+S_n\cdot\frac{\alpha_n}{S_n}f(x_n)+\alpha_{n+1}f(x_{n+1})=\alpha_1f(x_1)+\ldots+\alpha_{n+1}f(x_{n+1})$$

■

Пример (ср. арифм-ое $\ge$ ср. геом-ое): Пусть $x_i>0\Rightarrow\dfrac{x_1+\ldots+x_n}{n}\ge\sqrt[n]{x_1\cdot\ldots\cdot x_n}$.

□

$$\frac{x_1+\ldots+x_n}{n}\ge\sqrt[n]{x_1\cdot\ldots\cdot x_n}\Leftrightarrow\ln\Big(\frac{x_1+\ldots+x_n}{n}\Big)\ge\frac{\ln(x_1)+\ldots+\ln(x_n)}{n}\Leftrightarrow$$

$$\Leftrightarrow-\ln\Big(\frac{x_1+\ldots+x_n}{n}\Big)\le-\frac{\ln(x_1)}{n}-\ldots-\frac{\ln(x_n)}{n}$$

Функция $f(x)=-\ln x$ - выпукла: $f'(x)=-\dfrac{1}{x}$, $f''(x)=\dfrac{1}{x^2}>0\Rightarrow$ неравенство выше справедливо по неравенству Йенсена. ■

Пример (энтропия): $f(x)=x\ln x$, $x>0$; $f(x)$ - выпукла: $(x\ln x)''=(\ln x+1)'=\dfrac{1}{x}>0$. Функция энтропии дает характеристику, насколько неопределенна ситуация: когда все варианты равновероятны $\Rightarrow$ это ситуация с минимальной определенностью $\Rightarrow$ максимальная энтропия; когда только один вариант возможен $\Rightarrow$ это ситуация с максимальной определенностью $\Rightarrow$ минимальная энтропия.

Опр: 2. Функция энтропии: $H(p) = -\sum\limits_{i=1}^{n} p_i \ln p_i$.

Если $\forall i = \overline{1,n}$, $p_i = \frac{1}{n} \Rightarrow H(p) = -\frac{n}{n}\ln\frac{1}{n} = \ln n$.

Предположим, что мы кодируем слова, используя 0 и 1, тогда для слов длины n пришлось бы сделать 2^n кодов. Длина слова, которое можно закодировать равна $\ln 2^n = n$. Есть набор слов, спрашивается: какой длины коды будем писать $\Rightarrow$ возникает логарифм.

Сравним функцию $H(p)$ и $\ln n$. Что больше? Запишем $-\dfrac{1}{n}\ln n$ в следующем виде и применим неравенство Йенсена:

$$-\frac{1}{n}\ln n = \left(\frac{p_1 + \ldots + p_n}{n}\right)\ln\left(\frac{p_1 + \ldots + p_n}{n}\right) = f\left(\frac{p_1 + \ldots + p_n}{n}\right) \leq \frac{1}{n}p_1 \ln p_1 + \frac{1}{n}p_2 \ln p_2 + \ldots + \frac{1}{n}p_n \ln p_n$$

Домножаем на $(-n)$ получим:

$$-\sum_{i=1}^{n} p_i \ln p_i = H(p) = H(\{p_1, \ldots, p_n\}) \leq \ln n = H(\{\tfrac{1}{n}, \ldots, \tfrac{1}{n}\})$$

Таким образом, энтропия всегда меньше или равна энтропии в самой неопределенной ситуации. Энтропия минимальна, когда все слагаемые нулевые: $p_i = 0 \vee p_i = 1$, где мы определяем $0 \cdot \ln 0 = 0$. В таком случае, один из исходов - достоверный.

Строгая выпуклость

Опр: 3. Функция f на интервале (a,b) называется строго выпуклой, если

$$\forall x_1, x_2 \in (a,b), x_1 \neq x_2 \wedge \forall \alpha \in (0,1) \colon f(\alpha x_1 + (1-\alpha)x_2) < \alpha f(x_1) + (1-\alpha) f(x_2)$$

Утв. 1. Строгая выпуклость $\Leftrightarrow \forall x_1, x_2, x \in (a,b)$, $x_1 < x < x_2$, $\dfrac{f(x)-f(x_1)}{x-x_1} < \dfrac{f(x_2)-f(x)}{x_2-x}$.

□ Аналогичное доказательству для простой выпуклости. ■

Теорема 3. Пусть f - дифференцируема на интервале (a,b), тогда следующие утверждения эквивалентны:

(1) f - строго выпукла на (a,b);

(2) f' - возрастает на (a,b);

(3) $f(x) > f(y) + f'(y)(x-y)$, $\forall x \neq y \in (a,b)$;

□
(1) $\Leftrightarrow$ (2): Здесь уже не получится доказать по аналогии, поскольку строгие неравенства перейдут в нестрогие. Возьмем две точки $x_1, x_2 \in (a,b)$, $x_1 < x_2$. Возьмем любую точку $x \in (x_1, x_2)$. Мы знаем, что $\dfrac{f(x)-f(x_1)}{x-x_1} < \dfrac{f(x_2)-f(x)}{x_2-x}$. По теореме Лагранжа

$$\exists\, c_1 \in (x_1, x) \colon f'(c_1) = \frac{f(x)-f(x_1)}{x-x_1} \wedge \exists\, c_2 \in (x, x_2) \colon f'(c_2) = \frac{f(x_2)-f(x)}{x_2-x}$$

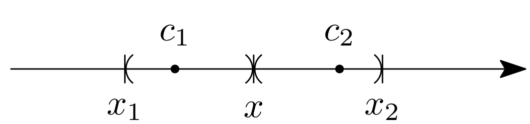

Рис. 2: Применение теоремы Лагранжа на отрезках (x_1, x) и (x, x_2).

Поскольку функция выпукла $\Rightarrow$ ее производная не убывает $\Rightarrow f'(x_1) \leq f'(c_1) \wedge f'(c_2) \leq f'(x_2)$, тогда:

$$\forall x_1, x_2 \in (a,b),\ x_1 < x_2,\ f'(x_1) \leq f'(c_1) = \frac{f(x)-f(x_1)}{x-x_1} < \frac{f(x_2)-f(x)}{x_2-x} = f'(c_2) \leq f(x_2)$$

Доказательство верно и в обратную сторону.

(2) $\Rightarrow$ (3): Рассмотрим несколько случаев:

1) Пусть $x > y \Rightarrow$ по теореме Лагранжа: $\exists\, c \in (y,x) \colon \dfrac{f(x)-f(y)}{x-y} = f'(c)$, по пункту (2):

$$c > y \Rightarrow f'(c) > f'(y) \Rightarrow f(x) - f(y) > f'(y)(x-y) \Rightarrow f(x) > f(y) + f'(y)(x-y)$$

2) Пусть $y > x \Rightarrow$ по теореме Лагранжа: $\exists\, c \in (x,y) \colon \dfrac{f(y)-f(x)}{y-x} = f'(c)$, по пункту (2):

$$c < y \Rightarrow f'(c) < f'(y) \Rightarrow f(y) - f(x) < f'(y)(y-x) \Rightarrow f(x) > f(y) + f'(y)(x-y)$$

(3) $\Rightarrow$ (1): Возьмем точки $x_1, x, x_2 \in (a,b)\colon x_1 < x < x_2$. Докажем, что $\dfrac{f(x)-f(x_1)}{x-x_1} < \dfrac{f(x_2)-f(x)}{x_2-x}$.

Рассмотрим точки x и $x_1 \Rightarrow$ по условию $f(x_1) > f(x) + f'(x)(x_1 - x) \Rightarrow f'(x)(x-x_1) > f(x) - f(x_1) \Rightarrow$ $\Rightarrow$ поскольку $x > x_1 \Rightarrow f'(x) > \dfrac{f(x)-f(x_1)}{x-x_1}$.

Рассмотрим точки x_2 и $x \Rightarrow$ по условию $f(x_2) > f(x) + f'(x)(x_2 - x) \Rightarrow f(x_2) - f(x) > f'(x)(x_2-x) \Rightarrow$ $\Rightarrow$ поскольку $x_2 > x \Rightarrow \dfrac{f(x_2)-f(x)}{x_2-x} > f'(x)$.

Таким образом мы получаем, что $\dfrac{f(x)-f(x_1)}{x-x_1} < f'(x) < \dfrac{f(x_2)-f(x)}{x_2-x}$. ■

Следствие 2. Пусть f дважды дифференцируема на (a,b). Если $f'' > 0$, то f строго выпукла.

Rm: 2. Необходимо выделить ряд замечаний:

(1) Выпуклые функции иногда называют выпуклыми вниз (потому что они провисают);

(2) Если $(-f)$ выпукла, то f называется вогнутой функцией или выпуклой вверх;

(3) Если f определена в окрестности $\mathcal{U}(a)$ и при $x < a, f$ выпукла (вогнута), при $x > a, f$ вогнута (выпукла), то есть f меняет выпуклость в точке a, то говорят, что a - точка перегиба f;

Метод касательных или метод Ньютона

Во многих прикладных задачах, важную роль играет поиск нуля функции: $f(x) = 0$. Способы поиска нуля функции:

1) Метод бисекций: делим пополам и выбираем ту половину на концах которой знаки разные, скорость сходимости к нулю функции $\dfrac{1}{2^n}$;

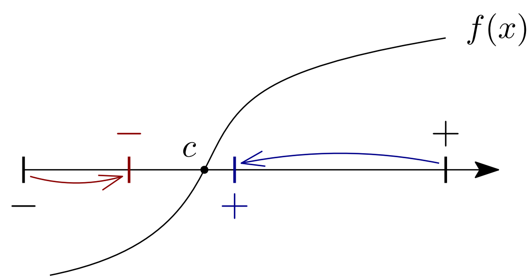

Рис. 3: Метод бисекций в поиске нуля функции $f(x)$.

2) Метод Ньютона: строим касательные в точках пересечения предыдущей касательной с осью x, скорость сходимости к нулю функции $\dfrac{1}{2^{2^n}}$;

Метод Ньютона

Пусть задана функция f на интервале (a, b), $f'(x) \geq \alpha > 0$, $\beta \geq f''(x) > 0$ (то есть функция строго монотонна и строго выпукла) и $\exists\, c \in (a, b)\colon f(c) = 0$. Поскольку функция возрастает, то такое c только одно.

Пусть $x_1 > c$. Будем устраивать последовательность, которая сходится к c:

$$x_n \to x_{n+1}\colon y = f(x_n) + f'(x_n)(x - x_n) \Rightarrow f(x_n) + f'(x_n)(x_{n+1} - x_n) = 0 \Rightarrow x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

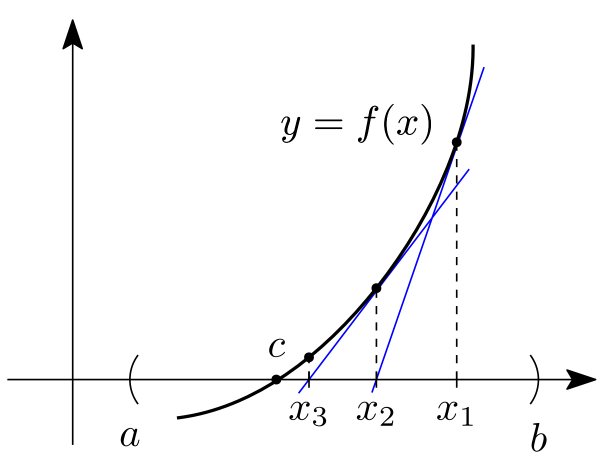

Рис. 4: Сходимость в методе Ньютона.

Rm: 3. Почему $x_n \to c$?

(1) Из-за строгой выпуклости $f(x_n) > 0 \Rightarrow$ так как $f(x) > 0$ только справа от $c \Rightarrow x_n > c$;

(2) x_n - убывает: $x_n > c \Rightarrow f(x_n) > 0 \land f'(x_n) > 0 \Rightarrow \dfrac{f(x_n)}{f'(x_n)} > 0 \Rightarrow x_{n+1} = x_n - \dfrac{f(x_n)}{f'(x_n)} < x_n$;

Следовательно, $x_n \to a \geq c \Rightarrow$ перейдем к пределу:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \Rightarrow a = a - \frac{f(a)}{f'(a)} \Rightarrow f(a) = 0$$

Но у нас есть только одна точка, где $f(x) = 0$ и эта точка $c \Rightarrow a = c$.

Rm: 4. Как быстро эта последовательность стремится к c?

Рассмотрим следующую разность:

$$|x_{n+1} - c| = \left| x_n - \frac{f(x_n)}{f'(x_n)} - c \right| = \left| (x_n - c) - \frac{f(x_n)}{f'(x_n)} \right|$$

Применим разложение Тейлора с остаточным членом в форме Лагранжа в точке x_n:

$$f(t) = f(x_n) + f'(x_n)(t - x_n) + \frac{f''(u)(t - x_n)^2}{2}$$

Подставим в разложение $t = c$:

$$0 = f(x_n) + f'(x_n)(c - x_n) + \frac{f''(u)(c - x_n)^2}{2} \Rightarrow \frac{f(x_n)}{f'(x_n)} = -(c - x_n) - \frac{f''(u)}{2f'(x_n)}(c - x_n)^2$$

Подставим это разложение в исходную разность и вспомним, что $f'(x) \geq \alpha > 0 \land \beta \geq f''(x) > 0$:

$$|x_{n+1} - c| = \left| \frac{f''(u)}{2f'(x_n)}(c - x_n)^2 \right| \leq \frac{\beta}{2\alpha}|x_n - c|^2 = M|x_n - c|^2$$

Таким образом мы получили:

$$|x_{n+1} - c| \leq M|x_n - c|^2 \leq M^2|x_{n-1} - c|^4 \leq M^{2^n}|x_1 - c|^{2^n}$$

Пусть мы изначально выбрали точку так, что $M|x_1 - c| < \dfrac{1}{2} \Rightarrow |x_{n+1} - c| \leq \dfrac{1}{2^{2^n}}$.