Выпуклость: выпуклые множества, выпуклые функции. Условие Поляка - Лоясиевича. Сильная выпуклость

Даня Меркулов

Оптимизация для всех! ЦУ

Выпуклые множества

Аффинные множества

Пусть x_1, x_2 два вектора в $\mathbb{R}^n$. Тогда прямая, проходящая через них, определяется следующим образом:

$$x = \theta x_1 + (1 - \theta)x_2, \theta \in \mathbb{R}$$

Множество A называется **аффинным**, если для любых x_1, x_2 из A прямая, проходящая через них, также лежит в A, т.е.

$$\forall \theta \in \mathbb{R}, \forall x_1, x_2 \in A : \theta x_1 + (1 - \theta)x_2 \in A$$

Example

- $\mathbb{R}^n$ - аффинное множество.

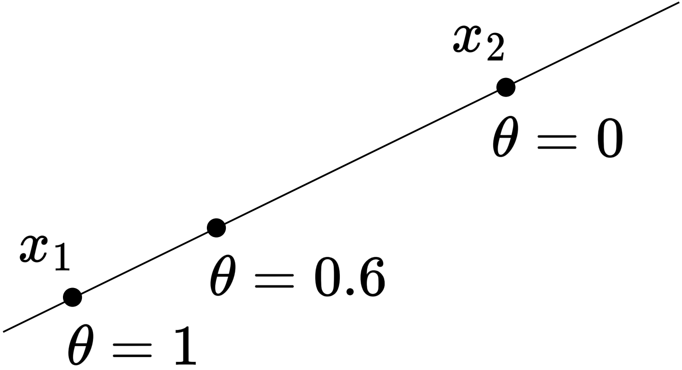


Рис. 1: Иллюстрация прямой между двумя векторами x_1 и x_2

Аффинные множества

Пусть x_1, x_2 два вектора в $\mathbb{R}^n$. Тогда прямая, проходящая через них, определяется следующим образом:

$$x = \theta x_1 + (1-\theta)x_2, \theta \in \mathbb{R}$$

Множество A называется **аффинным**, если для любых x_1, x_2 из A прямая, проходящая через них, также лежит в A, т.е.

$$\forall \theta \in \mathbb{R}, \forall x_1, x_2 \in A : \theta x_1 + (1-\theta)x_2 \in A$$

Example

- $\mathbb{R}^n$ - аффинное множество.
- Множество решений $\{x \mid \mathbf{A}x = \mathbf{b}\}$ также является аффинным множеством.

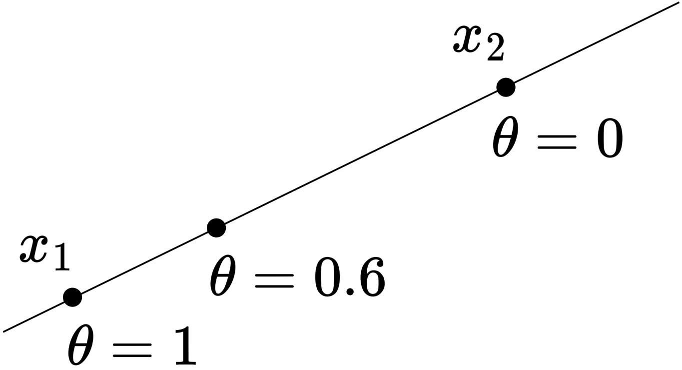


Рис. 1: Иллюстрация прямой между двумя векторами x_1 и x_2

Конус

Множество S называется **конусом**, если:

$$\forall x \in S,\ \theta \geq 0 \quad \rightarrow \quad \theta x \in S$$

Если точка принадлежит конусу, то и весь луч, проходящий из начала координат через эту точку, также принадлежит этому конусу.

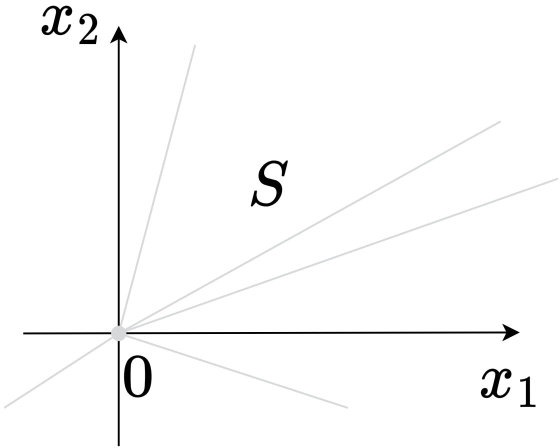


Рис. 2: Иллюстрация конуса

Выпуклый конус

Множество S называется **выпуклым конусом**, если:

$$\forall x_1, x_2 \in S,\ \theta_1, \theta_2 \geq 0 \quad \to \quad \theta_1 x_1 + \theta_2 x_2 \in S$$

Выпуклый конус это конус, который также является выпуклым множеством.

Example

- $\mathbb{R}^n$

Выпуклый конус

Множество S называется **выпуклым конусом**, если:

$$\forall x_1, x_2 \in S,\ \theta_1, \theta_2 \geq 0 \ \rightarrow \ \theta_1 x_1 + \theta_2 x_2 \in S$$

Выпуклый конус это конус, который также является выпуклым множеством.

Example

- $\mathbb{R}^n$
- Аффинные множества, содержащие 0

Выпуклый конус

Множество S называется **выпуклым конусом**, если:

$$\forall x_1, x_2 \in S,\ \theta_1, \theta_2 \geq 0 \ \rightarrow \ \theta_1 x_1 + \theta_2 x_2 \in S$$

Выпуклый конус это конус, который также является выпуклым множеством.

Example

- $\mathbb{R}^n$
- Аффинные множества, содержащие 0
- Луч

Выпуклый конус

Множество S называется **выпуклым конусом**, если:

$$\forall x_1, x_2 \in S,\ \theta_1, \theta_2 \geq 0 \ \ \rightarrow \ \ \theta_1 x_1 + \theta_2 x_2 \in S$$

Выпуклый конус это конус, который также является выпуклым множеством.

Example

- $\mathbb{R}^n$
- Аффинные множества, содержащие 0
- Луч
- $\mathbf{S}^n_+$ - множество симметричных положительно полуопределенных матриц

Выпуклый конус

Множество S называется **выпуклым конусом**, если:

$$\forall x_1, x_2 \in S,\ \theta_1, \theta_2 \geq 0 \ \ \rightarrow \ \ \theta_1 x_1 + \theta_2 x_2 \in S$$

Выпуклый конус это конус, который также является выпуклым множеством.

Example

- $\mathbb{R}^n$
- Аффинные множества, содержащие 0
- Луч
- $\mathbf{S}^n_+$ - множество симметричных положительно полуопределенных матриц

Выпуклый конус

Множество S называется **выпуклым конусом**, если:

$$\forall x_1, x_2 \in S,\ \theta_1, \theta_2 \geq 0 \ \ \rightarrow \ \ \theta_1 x_1 + \theta_2 x_2 \in S$$

Выпуклый конус это конус, который также является выпуклым множеством.

Example

- $\mathbb{R}^n$
- Аффинные множества, содержащие 0
- Луч
- $\mathbf{S}^n_+$ - множество симметричных положительно полуопределенных матриц

Выпуклый конус является выпуклым множеством, содержащим все конические комбинации точек в множестве.

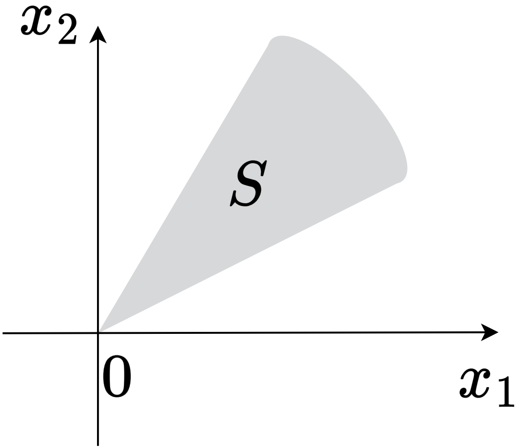


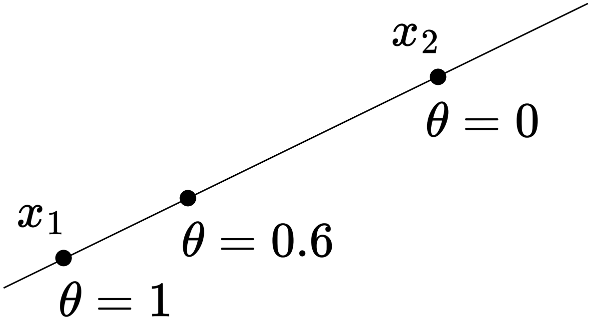
Рис. 3: Иллюстрация выпуклого конуса

Отрезок

Пусть x_1, x_2 два вектора в $\mathbb{R}^n$.
Тогда отрезок между ними определяется следующим образом:

$$x = \theta x_1 + (1 - \theta) x_2, \ \theta \in [0, 1]$$

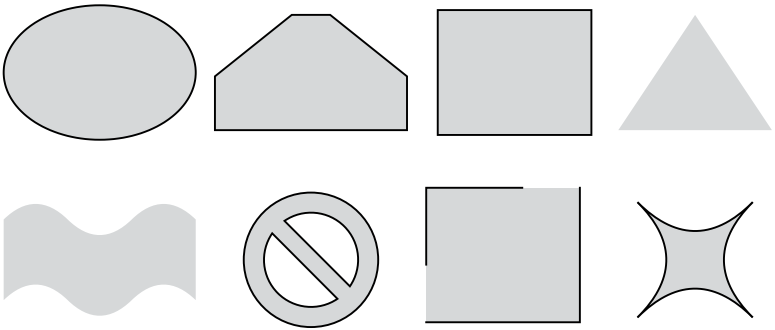
Выпуклое множество содержит отрезок между любыми двумя точками в множестве.



Выпуклое множество

Множество S называется **выпуклым**, если для любых x_1, x_2 из S отрезок между ними также лежит в S, т.е.

$$\forall \theta \in [0,1], \ \forall x_1, x_2 \in S : \theta x_1 + (1-\theta) x_2 \in S$$



Example

Пустое множество и множество из одного вектора являются выпуклыми по определению.

Example

Любое аффинное множество, луч или отрезок являются выпуклыми множествами.

Выпуклая комбинация

Пусть $x_1, x_2, \dots, x_k \in S$, тогда точка $\theta_1 x_1 + \theta_2 x_2 + \dots + \theta_k x_k$ называется **выпуклой комбинацией** точек $x_1, x_2, \dots, x_k$ если $\sum\limits_{i=1}^{k} \theta_i = 1, \; \theta_i \geq 0$.

Выпуклая оболочка

Множество всех выпуклых комбинаций точек из S называется **выпуклой оболочкой** множества S.

$$\mathbf{conv}(S) = \left\{ \sum_{i=1}^{k} \theta_i x_i \mid x_i \in S, \sum_{i=1}^{k} \theta_i = 1, \ \theta_i \geq 0 \right\}$$

* Множество $\mathbf{conv}(S)$ является наименьшим выпуклым множеством, содержащим S. * Множество S является выпуклым тогда и только тогда, когда $S = \mathbf{conv}(S)$.

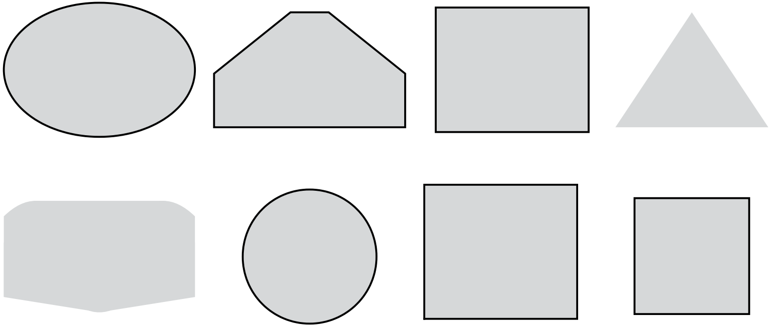


Рис. 5: Верх: выпуклые оболочки выпуклых множеств. Низ: выпуклая оболочка невыпуклых множеств.

Сумма Минковского

Сумма Минковского двух множеств векторов S_1 и S_2 в евклидовом пространстве образуется путем сложения каждого вектора из S_1 с каждым вектором из S_2.

$$S_1 + S_2 = \{\mathbf{s_1} + \mathbf{s_2} \,|\, \mathbf{s_1} \in S_1,\ \mathbf{s_2} \in S_2\}$$

Также можно определить линейную комбинацию множеств.

Example

Рассмотрим пространство $\mathbb{R}^2$. Определим:

$$S_1 := \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq 1\}$$

Это единичная окружность, с центром в начале координат. И:

$$S_2 := \{x \in \mathbb{R}^2 : -4 \leq x_1 \leq -1, -3 \leq x_2 \leq -1\}$$

Это прямоугольник. Сумма множеств S_1 и S_2 образуетувеличенный прямоугольник S_2 с закругленными углами. Полученное множество будет выпуклым.

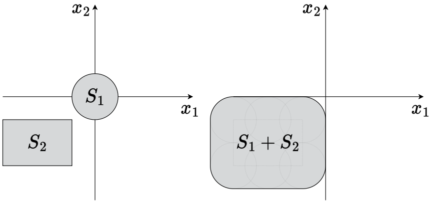


Рис. 6: $S = S_1 + S_2$

Проверка выпуклости

На практике очень важно понимать, является ли конкретное множество выпуклым или нет. Для этого используются два подхода в зависимости от контекста.

- По определению.

Проверка выпуклости

На практике очень важно понимать, является ли конкретное множество выпуклым или нет. Для этого используются два подхода в зависимости от контекста.

- По определению.
- Показать, что S получается из простых выпуклых множеств с помощью операций, сохраняющих выпуклость.

Проверка выпуклости по определению

$$
x_1, x_2 \in S,\ 0 \leq \theta \leq 1 \ \ \rightarrow \ \ \theta x_1 + (1-\theta) x_2 \in S
$$

Example

Доказать, что множество симметричных положительно определенных матриц $\mathbf{S}^n_{++} = \{\mathbf{X} \in \mathbb{R}^{n \times n} \mid \mathbf{X} = \mathbf{X}^\top,\ \mathbf{X} \succ 0\}$ является выпуклым.

Операции, сохраняющие выпуклость

Линейная комбинация выпуклых множеств является выпуклым множеством.

Пусть есть два выпуклых множества S_x, S_y, тогда множество

$$S = \left\{s \mid s = c_1 x + c_2 y, \ x \in S_x, \ y \in S_y, \ c_1, c_2 \in \mathbb{R}\right\}$$

Возьмем два вектора из S: $s_1 = c_1 x_1 + c_2 y_1, s_2 = c_1 x_2 + c_2 y_2$ и докажем, что отрезок между ними $\theta s_1 + (1 - \theta)s_2, \theta \in [0, 1]$ также принадлежит S

$$\theta s_1 + (1 - \theta)s_2$$

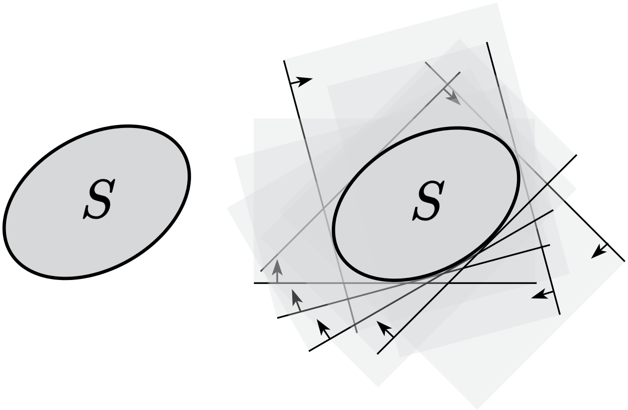
$$\theta(c_1 x_1 + c_2 y_1) + (1 - \theta)(c_1 x_2 + c_2 y_2)$$

$$c_1(\theta x_1 + (1 - \theta)x_2) + c_2(\theta y_1 + (1 - \theta)y_2)$$

$$c_1 x + c_2 y \in S$$

Пересечение любого (!) числа выпуклых множеств является выпуклым

Если пересечение пустое или содержит одну точку, свойство доказывается по определению. В противном случае возьмем две точки и отрезок между ними. Эти точки должны лежать во всех пересекающихся множествах, и поскольку они все выпуклые, отрезок между ними лежит во всех множествах и, следовательно, в их пересечении.



Образ выпуклого множества при аффинном отображении является выпуклым

$$
S \subseteq \mathbb{R}^n \text{ выпукло } \rightarrow \; f(S) = \{ f(x) \mid x \in S \} \text{ выпукло } \quad (f(x) = \mathbf{A}x + \mathbf{b})
$$

Примеры аффинных множеств: расширение, проекция, транспонирование, множество решений линейного матричного неравенства $\{x \mid x_1 A_1 + ... + x_m A_m \preceq B\}$. Здесь $A_i, B \in \mathbf{S}^p$ симметричные матрицы $p \times p$.

Также обратим внимание, что прообраз выпуклого множества при аффинном отображении также является выпуклым.

$$
S \subseteq \mathbb{R}^m \text{ выпукло } \rightarrow \; f^{-1}(S) = \{ x \in \mathbb{R}^n \mid f(x) \in S \} \text{ выпукло } \; (f(x) = \mathbf{A}x + \mathbf{b})
$$

Пример

Пусть $x \in \mathbb{R}$ - случайная величина с заданным вероятностным распределением $\mathbb{P}(x = a_i) = p_i$, где $i = 1, \ldots, n$, и $a_1 < \ldots < a_n$. Тогда вектор вероятностей $p \in \mathbb{R}^n$ принадлежит вероятностному симплексу, т.е.

$$P = \{p \mid \mathbf{1}^T p = 1, p \succeq 0\} = \{p \mid p_1 + \ldots + p_n = 1, p_i \geq 0\}.$$

Определить, являются ли следующие множества вероятностных векторов p выпуклыми:

- $\mathbb{P}(x > \alpha) \leq \beta$

Пример

Пусть $x \in \mathbb{R}$ - случайная величина с заданным вероятностным распределением $\mathbb{P}(x = a_i) = p_i$, где $i = 1, \ldots, n$, и $a_1 < \ldots < a_n$. Тогда вектор вероятностей $p \in \mathbb{R}^n$ принадлежит вероятностному симплексу, т.е.

$$P = \{p \mid \mathbf{1}^T p = 1, p \succeq 0\} = \{p \mid p_1 + \ldots + p_n = 1, p_i \geq 0\}.$$

Определить, являются ли следующие множества вероятностных векторов p выпуклыми:

- $\mathbb{P}(x > \alpha) \leq \beta$
- $\mathbb{E}|x^{201}| \leq \alpha \mathbb{E}|x|$

Пример

Пусть $x \in \mathbb{R}$ - случайная величина с заданным вероятностным распределением $\mathbb{P}(x = a_i) = p_i$, где $i = 1, \ldots, n$, и $a_1 < \ldots < a_n$. Тогда вектор вероятностей $p \in \mathbb{R}^n$ принадлежит вероятностному симплексу, т.е.

$$P = \{p \mid \mathbf{1}^T p = 1, p \succeq 0\} = \{p \mid p_1 + \ldots + p_n = 1, p_i \geq 0\}.$$

Определить, являются ли следующие множества вероятностных векторов p выпуклыми:

- $\mathbb{P}(x > \alpha) \leq \beta$
- $\mathbb{E}|x^{201}| \leq \alpha \mathbb{E}|x|$
- $\mathbb{E}|x^2| \geq \alpha \mathbb{V}x \geq \alpha$

Выпуклые функции

Неравенство Йенсена

Функция $f(x)$, **определенная на выпуклом множестве** $S \subseteq \mathbb{R}^n$, называется **выпуклой** на S, если:

$$f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda)f(x_2)$$

для любых $x_1, x_2 \in S$ и $0 \le \lambda \le 1$.
Если вышеуказанное неравенство выполняется строгим неравенством для $x_1 \neq x_2$ и $0 < \lambda < 1$, то функция называется строго выпуклой на S.

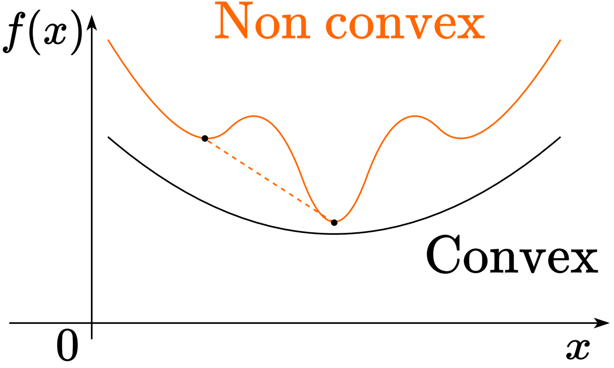


Рис. 8: Разница между выпуклой и невыпуклой функцией

Неравенство Йенсена

Theorem

Пусть $f(x)$ выпуклая функция на выпуклом множестве $X \subseteq \mathbb{R}^n$ и пусть $x_i \in X, 1 \leq i \leq m$, произвольные точки из X. Тогда

$$f\left(\sum_{i=1}^{m} \lambda_i x_i\right) \leq \sum_{i=1}^{m} \lambda_i f(x_i)$$

для любого $\lambda = [\lambda_1, \ldots, \lambda_m] \in \Delta_m$ - вероятностного симплекса.

Доказательство

1. Во-первых, обратим внимание, что точка $\sum_{i=1}^{m} \lambda_i x_i$ как выпуклая комбинация точек из выпуклого множества X принадлежит X.

Неравенство Йенсена

Theorem

Пусть $f(x)$ выпуклая функция на выпуклом множестве $X \subseteq \mathbb{R}^n$ и пусть $x_i \in X, 1 \leq i \leq m$, произвольные точки из X. Тогда

$$f\left(\sum_{i=1}^{m} \lambda_i x_i\right) \leq \sum_{i=1}^{m} \lambda_i f(x_i)$$

для любого $\lambda = [\lambda_1, \ldots, \lambda_m] \in \Delta_m$ - вероятностного симплекса.

Доказательство

1. Во-первых, обратим внимание, что точка $\sum_{i=1}^{m} \lambda_i x_i$ как выпуклая комбинация точек из выпуклого множества X принадлежит X.
2. Мы докажем это индукцией. Для $m = 1$, утверждение очевидно, и для $m = 2$, оно следует из определения выпуклой функции.

Неравенство Йенсена

3. Предположим, что оно верно для всех m до $m = k$, и мы докажем его для $m = k + 1$. Пусть $\lambda \in \Delta k + 1$ и

$$x = \sum_{i=1}^{k+1} \lambda_i x_i = \sum_{i=1}^{k} \lambda_i x_i + \lambda_{k+1} x_{k+1}.$$

Предположим, что $0 < \lambda_{k+1} < 1$, иначе, оно сводится к рассмотренным ранее случаям, тогда мы имеем

$$x = \lambda_{k+1} x_{k+1} + (1 - \lambda_{k+1}) \bar{x},$$

где $\bar{x} = \sum_{i=1}^{k} \gamma_i x_i$ и $\gamma_i = \frac{\lambda_i}{1 - \lambda_{k+1}} \geq 0, 1 \leq i \leq k$.

Неравенство Йенсена

3. Предположим, что оно верно для всех m до $m = k$, и мы докажем его для $m = k + 1$. Пусть $\lambda \in \Delta k + 1$ и

$$x = \sum_{i=1}^{k+1} \lambda_i x_i = \sum_{i=1}^{k} \lambda_i x_i + \lambda_{k+1} x_{k+1}.$$

Предположим, что $0 < \lambda_{k+1} < 1$, иначе, оно сводится к рассмотренным ранее случаям, тогда мы имеем

$$x = \lambda_{k+1} x_{k+1} + (1 - \lambda_{k+1}) \bar{x},$$

где $\bar{x} = \sum_{i=1}^{k} \gamma_i x_i$ и $\gamma_i = \frac{\lambda_i}{1 - \lambda_{k+1}} \geq 0, 1 \leq i \leq k$.

4. Поскольку $\lambda \in \Delta_{k+1}$, то $\gamma = [\gamma_1, \ldots, \gamma_k] \in \Delta_k$. Следовательно, $\bar{x} \in X$ и по выпуклости $f(x)$ и гипотезе индукции:

$$f\left(\sum_{i=1}^{k+1} \lambda_i x_i\right) = f\left(\lambda_{k+1} x_{k+1} + (1 - \lambda_{k+1}) \bar{x}\right) \leq \lambda_{k+1} f(x_{k+1}) + (1 - \lambda_{k+1}) f(\bar{x}) \leq \sum_{i=1}^{k+1} \lambda_i f(x_i)$$

Таким образом, исходное неравенство выполняется для $m = k + 1$.

Примеры выпуклых функций

- $f(x) = x^p, \ p > 1, \ x \in \mathbb{R}_+$
- $f(x) = \|x\|^p, \ p > 1, x \in \mathbb{R}^n$
- $f(x) = e^{cx}, \ c \in \mathbb{R}, x \in \mathbb{R}$
- $f(x) = -\ln x, \ x \in \mathbb{R}_{++}$
- $f(x) = x \ln x, \ x \in \mathbb{R}_{++}$
- Сумма k наибольших координат $f(x) = x_{(1)} + ... + x_{(k)}, \ x \in \mathbb{R}^n$
- $f(X) = \lambda_{max}(X), \ X = X^T$
- $f(X) = -\log \det X, \ X \in S^n_{++}$

Надграфик

Для функции $f(x)$, определенной на $S \subseteq \mathbb{R}^n$, множество:

$$
\text{epi } f = \{[x, \mu] \in S \times \mathbb{R} : f(x) \leq \mu\}
$$

называется **надграфиком** функции $f(x)$.

> ℹ Выпуклость надграфика = выпуклость функции
>
> Для того чтобы функция $f(x)$, определенная на выпуклом множестве X, была выпуклой на X, необходимо и достаточно, чтобы надграфик функции f был выпуклым множеством.

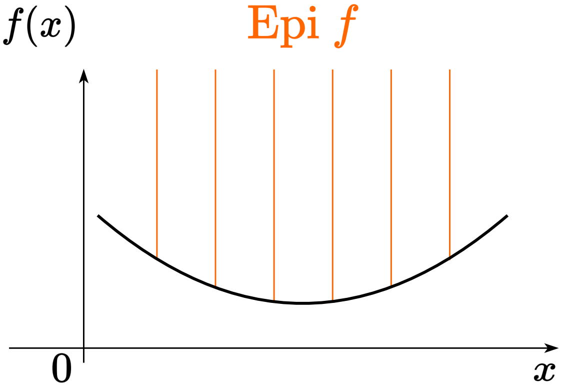


Рис. 9: Надграфик функции

Выпуклость надграфика = выпуклость функции

1. **Необходимость**: Предположим, что $f(x)$ выпукла на X. Возьмем любые две произвольные точки $[x_1, \mu_1] \in \text{epi} f$ и $[x_2, \mu_2] \in \text{epi} f$. Также возьмем $0 \leq \lambda \leq 1$ и обозначим $x_\lambda = \lambda x_1 + (1-\lambda)x_2, \mu_\lambda = \lambda \mu_1 + (1-\lambda)\mu_2$. Тогда,

$$\lambda \begin{bmatrix} x_1 \\ \mu_1 \end{bmatrix} + (1-\lambda) \begin{bmatrix} x_2 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} x_\lambda \\ \mu_\lambda \end{bmatrix}.$$

Из выпуклости множества X следует, что $x_\lambda \in X$. Кроме того, поскольку $f(x)$ выпуклая функция,

$$f(x_\lambda) \leq \lambda f(x_1) + (1-\lambda) f(x_2) \leq \lambda \mu_1 + (1-\lambda)\mu_2 = \mu_\lambda$$

Неравенство выше означает, что $\begin{bmatrix} x_\lambda \\ \mu_\lambda \end{bmatrix} \in \text{epi} f$. Таким образом, надграфик функции f является выпуклым множеством.

Выпуклость надграфика = выпуклость функции

2. **Достаточность**: Предположим, что надграфик функции f, epi$f$, является выпуклым множеством. Тогда, из принадлежности точек $[x_1, \mu_1]$ и $[x_2, \mu_2]$ надграфику функции f, следует, что

$$\begin{bmatrix} x_\lambda \\ \mu_\lambda \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ \mu_1 \end{bmatrix} + (1 - \lambda) \begin{bmatrix} x_2 \\ \mu_2 \end{bmatrix} \in \text{epi} f$$

для любого $0 \le \lambda \le 1$, т.е. $f(x_\lambda) \le \mu_\lambda = \lambda \mu_1 + (1 - \lambda)\mu_2$. Но это верно для всех $\mu_1 \ge f(x_1)$ и $\mu_2 \ge f(x_2)$, в частности, когда $\mu_1 = f(x_1)$ и $\mu_2 = f(x_2)$. Следовательно, мы приходим к неравенству

$$f(x_\lambda) = f(\lambda x_1 + (1 - \lambda)x_2) \le \lambda f(x_1) + (1 - \lambda) f(x_2).$$

Поскольку точки $x_1 \in X$ и $x_2 \in X$ могут быть выбраны произвольно, $f(x)$ является выпуклой функцией на X.

Пример: конус нормы

Пусть норма $\|\cdot\|$ определена в пространстве U. Рассмотрим множество:

$$K := \{(x, t) \in U \times \mathbb{R}^+ : \|x\| \leq t\}$$

которое представляет собой надграфик функции $x \mapsto \|x\|$. Это множество называется конусом нормы. Согласно утверждению выше, множество K является выпуклым. Код для рисунков

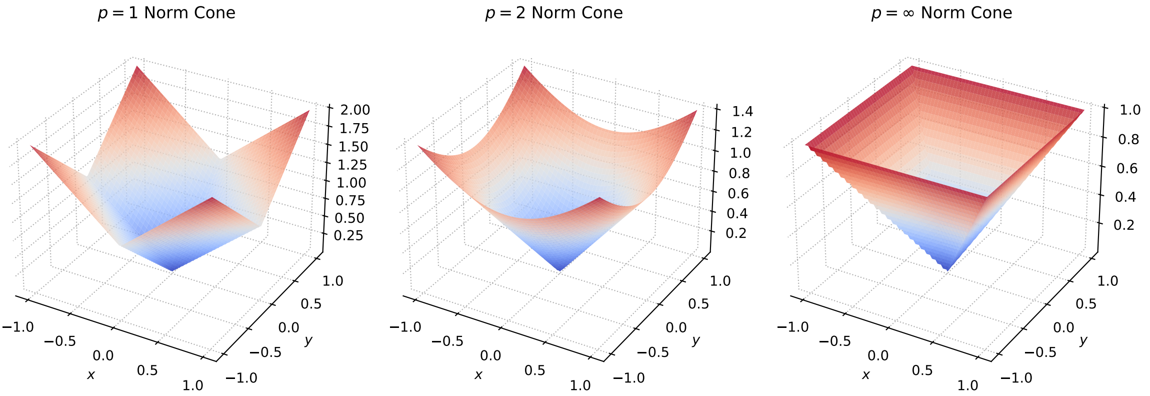


Рис. 10: Конусы нормы для разных p - норм

Множество подуровня

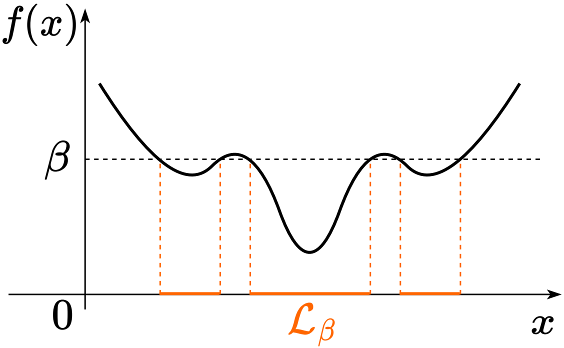


Рис. 11: Множество подуровня функции с уровнем β

Для функции $f(x)$, определенной на $S \subseteq \mathbb{R}^n$, следующее множество:

$$
\mathcal{L}_\beta = \{x \in S : f(x) \leq \beta\}
$$

называется **множеством подуровня** или множеством Лебега функции $f(x)$.

Множество подуровня

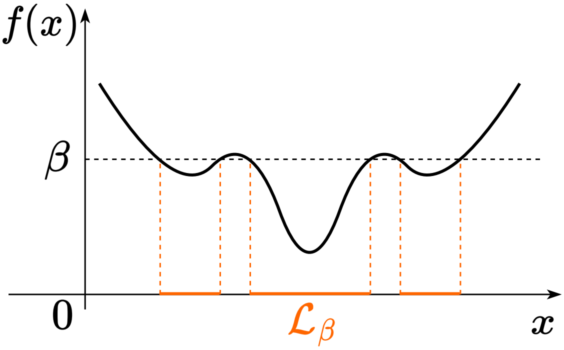


Рис. 11: Множество подуровня функции с уровнем β

Для функции $f(x)$, определенной на $S \subseteq \mathbb{R}^n$, следующее множество:

$$
\mathcal{L}_\beta = \{x \in S : f(x) \leq \beta\}
$$

называется **множеством подуровня** или множеством Лебега функции $f(x)$.

Обратите внимание, что если функция $f(x)$ выпукла, то ее множества подуровня выпуклы для любого $\beta \in \mathbb{R}$. Однако **обратное неверно**. (Рассмотрим функцию $f(x) = \sqrt{|x|}$)

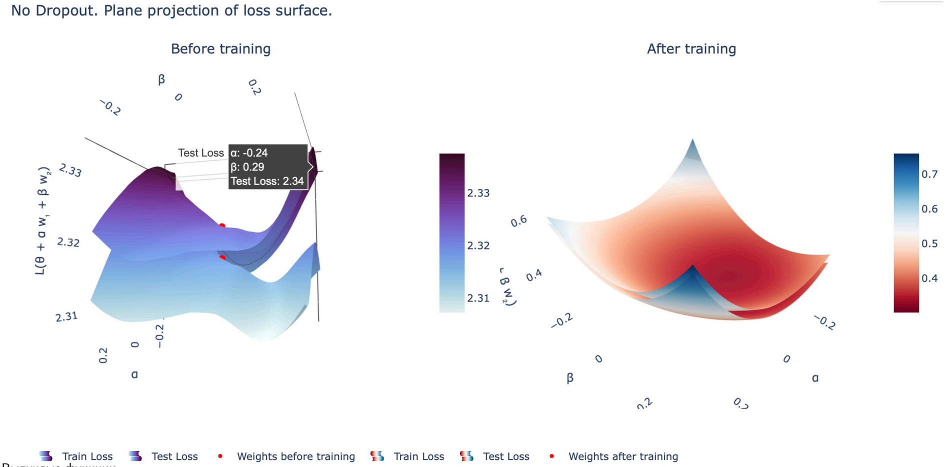
Сведение к прямой

$f : S \to \mathbb{R}$ выпукла тогда и только тогда, когда S выпукло и функция $g(t) = f(x + tv)$ определена на $\{t \mid x + tv \in S\}$ и выпукла для любого $x \in S, v \in \mathbb{R}^n$, что позволяет проверять выпуклость скалярной функции для установления выпуклости векторной функции.

Сведение к прямой

$f : S \to \mathbb{R}$ выпукла тогда и только тогда, когда S выпукло и функция $g(t) = f(x + tv)$ определена на $\{t \mid x + tv \in S\}$ и выпукла для любого $x \in S, v \in \mathbb{R}^n$, что позволяет проверять выпуклость скалярной функции для установления выпуклости векторной функции.

Если существует направление v для которого $g(t)$ не выпукло, то f не выпукла.



Операции, сохраняющие выпуклость

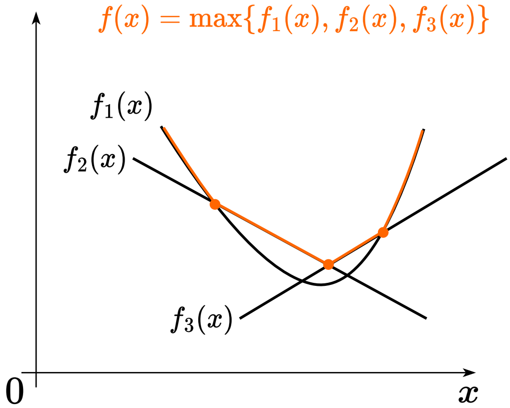


Рис. 12: Поточечный максимум (супремум) выпуклых функций выпуклый

- Поточечный максимум (супремум) любого числа функций: Если $f_1(x), \ldots, f_m(x)$ выпуклы, то $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ выпукла.

Операции, сохраняющие выпуклость

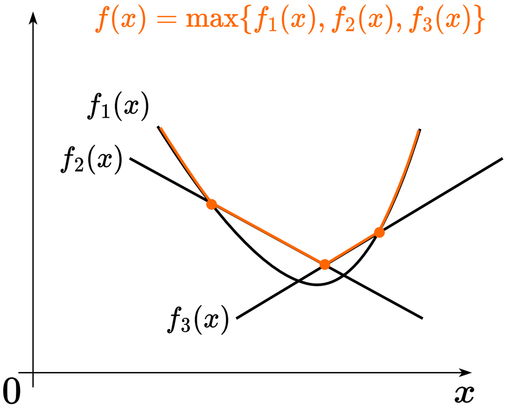


Рис. 12: Поточечный максимум (супремум) выпуклых функций выпуклый

- Поточечный максимум (супремум) любого числа функций: Если $f_1(x), \ldots, f_m(x)$ выпуклы, то $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ выпукла.
- Неотрицательная сумма выпуклых функций: $\alpha f(x) + \beta g(x), (\alpha \geq 0, \beta \geq 0)$.

Операции, сохраняющие выпуклость

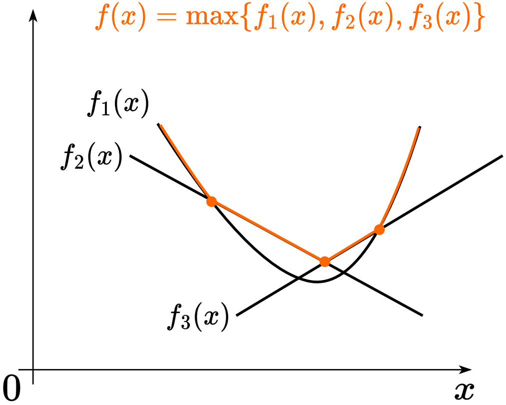


Рис. 12: Поточечный максимум (супремум) выпуклых функций выпуклый

- Поточечный максимум (супремум) любого числа функций: Если $f_1(x), \ldots, f_m(x)$ выпуклы, то $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ выпукла.
- Неотрицательная сумма выпуклых функций: $\alpha f(x) + \beta g(x), (\alpha \geq 0, \beta \geq 0)$.
- Композиция с аффинной функцией $f(Ax + b)$ выпукла, если $f(x)$ выпукла.

Операции, сохраняющие выпуклость

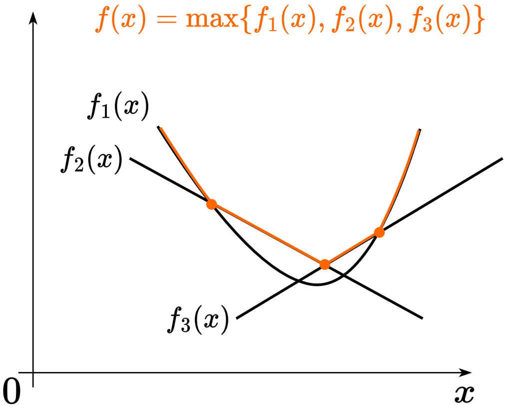


Рис. 12: Поточечный максимум (супремум) выпуклых функций выпуклый

- Поточечный максимум (супремум) любого числа функций: Если $f_1(x), \ldots, f_m(x)$ выпуклы, то $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ выпукла.
- Неотрицательная сумма выпуклых функций: $\alpha f(x) + \beta g(x), (\alpha \geq 0, \beta \geq 0)$.
- Композиция с аффинной функцией $f(Ax + b)$ выпукла, если $f(x)$ выпукла.
- Если $f(x, y)$ выпукла по x для любого $y \in Y$: $g(x) = \sup\limits_{y \in Y} f(x, y)$ также выпукла.

Операции, сохраняющие выпуклость

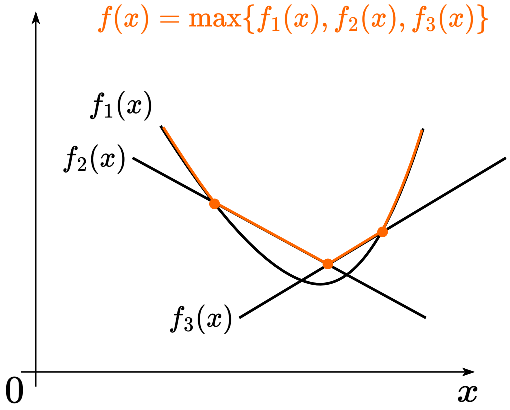


Рис. 12: Поточечный максимум (супремум) выпуклых функций выпуклый

- Поточечный максимум (супремум) любого числа функций: Если $f_1(x), \ldots, f_m(x)$ выпуклы, то $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ выпукла.
- Неотрицательная сумма выпуклых функций: $\alpha f(x) + \beta g(x), (\alpha \geq 0, \beta \geq 0)$.
- Композиция с аффинной функцией $f(Ax + b)$ выпукла, если $f(x)$ выпукла.
- Если $f(x, y)$ выпукла по x для любого $y \in Y$: $g(x) = \sup\limits_{y \in Y} f(x, y)$ также выпукла.
- Если $f(x)$ выпукла на S, то $g(x, t) = tf(x/t)$ - выпукла с $x/t \in S, t > 0$.

Операции, сохраняющие выпуклость

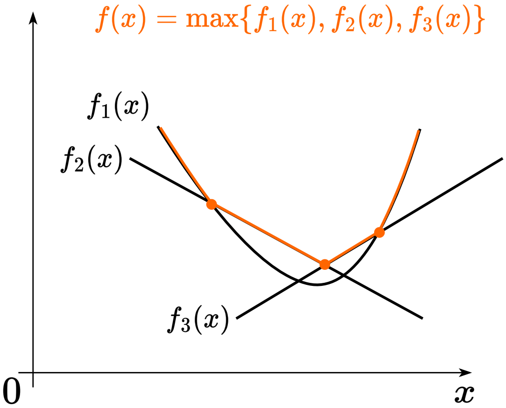


Рис. 12: Поточечный максимум (супремум) выпуклых функций выпуклый

- Поточечный максимум (супремум) любого числа функций: Если $f_1(x), \ldots, f_m(x)$ выпуклы, то $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ выпукла.
- Неотрицательная сумма выпуклых функций: $\alpha f(x) + \beta g(x), (\alpha \geq 0, \beta \geq 0)$.
- Композиция с аффинной функцией $f(Ax + b)$ выпукла, если $f(x)$ выпукла.
- Если $f(x, y)$ выпукла по x для любого $y \in Y$: $g(x) = \sup\limits_{y \in Y} f(x, y)$ также выпукла.
- Если $f(x)$ выпукла на S, то $g(x, t) = tf(x/t)$ - выпукла с $x/t \in S, t > 0$.
- Пусть $f_1 : S_1 \to \mathbb{R}$ и $f_2 : S_2 \to \mathbb{R}$, где $\text{range}(f_1) \subseteq S_2$. Если f_1 и f_2 выпуклы, и f_2 возрастает, то $f_2 \circ f_1$ выпукла на S_1.

Функция максимального собственного значения матрицы является выпуклой

Example

Покажите, что $f(A) = \lambda_{max}(A)$ - выпукла, если $A \in S^n$.

Критерии сильной выпуклости

Дифференциальный критерий выпуклости первого порядка

Дифференцируемая функция $f(x)$ определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$ выпукла тогда и только тогда, когда $\forall x, y \in S$:

$$f(y) \geq f(x) + \nabla f^T(x)(y - x)$$

Пусть $y = x + \Delta x$, тогда критерий запишется в виде:

$$f(x + \Delta x) \geq f(x) + \nabla f^T(x)\Delta x$$

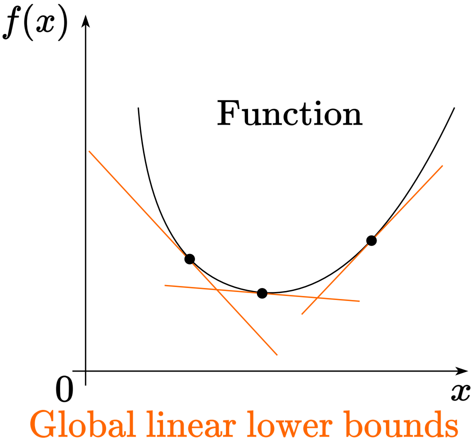


Рис. 13: Выпуклая функция больше или равна линейной аппроксимации Тейлора в любой точке

Дифференциальный критерий выпуклости второго порядка

Дважды дифференцируемая функция $f(x)$ определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$ выпукла тогда и только тогда, когда $\forall x \in \mathbf{int}(S) \neq \emptyset$:

$$\nabla^2 f(x) \succeq 0$$

Другими словами, $\forall y \in \mathbb{R}^n$:

$$\langle y, \nabla^2 f(x) y \rangle \geq 0$$

Сильная выпуклость

$f(x)$, **определенная на выпуклом множестве** $S \subseteq \mathbb{R}^n$, называется μ-сильно выпуклой (сильно выпуклой) на S, если:

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2) - \frac{\mu}{2}\lambda(1-\lambda)\|x_1 - x_2\|^2$$

для любых $x_1, x_2 \in S$ и $0 \leq \lambda \leq 1$ для некоторого $\mu > 0$.

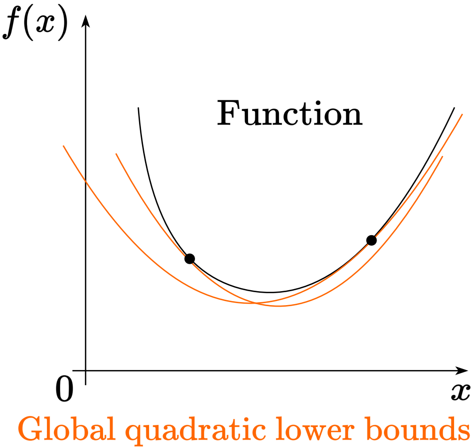


Рис. 14: Сильно выпуклая функция не меньше некоторой параболы в любой точке

Дифференциальный критерий сильной выпуклости первого порядка

Дифференцируемая $f(x)$ определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$ является μ-сильно выпуклой тогда и только тогда, когда $\forall x, y \in S$:

$$f(y) \geq f(x) + \nabla f^T(x)(y - x) + \frac{\mu}{2}\|y - x\|^2$$

Дифференциальный критерий сильной выпуклости первого порядка

Дифференцируемая $f(x)$ определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$ является μ-сильно выпуклой тогда и только тогда, когда $\forall x, y \in S$:

$$f(y) \geq f(x) + \nabla f^T(x)(y - x) + \frac{\mu}{2}\|y - x\|^2$$

Пусть $y = x + \Delta x$, тогда критерий запишется в виде:

$$f(x + \Delta x) \geq f(x) + \nabla f^T(x)\Delta x + \frac{\mu}{2}\|\Delta x\|^2$$

Дифференциальный критерий сильной выпуклости первого порядка

Дифференцируемая $f(x)$ определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$ является μ-сильно выпуклой тогда и только тогда, когда $\forall x, y \in S$:

$$f(y) \geq f(x) + \nabla f^T(x)(y - x) + \frac{\mu}{2}\|y - x\|^2$$

Пусть $y = x + \Delta x$, тогда критерий запишется в виде:

$$f(x + \Delta x) \geq f(x) + \nabla f^T(x)\Delta x + \frac{\mu}{2}\|\Delta x\|^2$$

Theorem

Пусть $f(x)$ дифференцируемая функция на выпуклом множестве $X \subseteq \mathbb{R}^n$. Тогда $f(x)$ сильно выпукла на X с константой $\mu > 0$ тогда и только тогда, когда

$$f(x) - f(x_0) \geq \langle \nabla f(x_0), x - x_0 \rangle + \frac{\mu}{2}\|x - x_0\|^2$$

для всех $x, x_0 \in X$.

Доказательство. Необходимость

Пусть $0 < \lambda \leq 1$. Согласно определению сильно выпуклой функции,

$$f(\lambda x + (1-\lambda)x_0) \leq \lambda f(x) + (1-\lambda)f(x_0) - \frac{\mu}{2}\lambda(1-\lambda)\|x - x_0\|^2$$

Доказательство. Необходимость

Пусть $0 < \lambda \leq 1$. Согласно определению сильно выпуклой функции,

$$
f(\lambda x + (1-\lambda)x_0) \leq \lambda f(x) + (1-\lambda)f(x_0) - \frac{\mu}{2}\lambda(1-\lambda)\|x - x_0\|^2
$$

или эквивалентно,

$$
f(x) - f(x_0) - \frac{\mu}{2}(1-\lambda)\|x - x_0\|^2 \geq \frac{1}{\lambda}[f(\lambda x + (1-\lambda)x_0) - f(x_0)] =
$$

Доказательство. Необходимость

Пусть $0 < \lambda \leq 1$. Согласно определению сильно выпуклой функции,

$$f(\lambda x + (1-\lambda)x_0) \leq \lambda f(x) + (1-\lambda)f(x_0) - \frac{\mu}{2}\lambda(1-\lambda)\|x - x_0\|^2$$

или эквивалентно,

$$f(x) - f(x_0) - \frac{\mu}{2}(1-\lambda)\|x - x_0\|^2 \geq \frac{1}{\lambda}[f(\lambda x + (1-\lambda)x_0) - f(x_0)] =$$

$$= \frac{1}{\lambda}[f(x_0 + \lambda(x - x_0)) - f(x_0)] = \frac{1}{\lambda}[\lambda\langle\nabla f(x_0), x - x_0\rangle + o(\lambda)] =$$

Доказательство. Необходимость

Пусть $0 < \lambda \leq 1$. Согласно определению сильно выпуклой функции,

$$f(\lambda x + (1-\lambda)x_0) \leq \lambda f(x) + (1-\lambda)f(x_0) - \frac{\mu}{2}\lambda(1-\lambda)\|x - x_0\|^2$$

или эквивалентно,

$$f(x) - f(x_0) - \frac{\mu}{2}(1-\lambda)\|x - x_0\|^2 \geq \frac{1}{\lambda}[f(\lambda x + (1-\lambda)x_0) - f(x_0)] =$$

$$= \frac{1}{\lambda}[f(x_0 + \lambda(x - x_0)) - f(x_0)] = \frac{1}{\lambda}[\lambda\langle\nabla f(x_0), x - x_0\rangle + o(\lambda)] =$$

$$= \langle\nabla f(x_0), x - x_0\rangle + \frac{o(\lambda)}{\lambda}.$$

Таким образом, переходя к пределу при $\lambda \downarrow 0$, мы приходим к исходному утверждению.

Доказательство. Достаточность

Предположим, что неравенство в теореме выполняется для всех $x, x_0 \in X$. Возьмем $x_0 = \lambda x_1 + (1-\lambda)x_2$, где $x_1, x_2 \in X$, $0 \le \lambda \le 1$. Согласно неравенству, следующие неравенства выполняются:

Доказательство. Достаточность

Предположим, что неравенство в теореме выполняется для всех $x, x_0 \in X$. Возьмем $x_0 = \lambda x_1 + (1-\lambda)x_2$, где $x_1, x_2 \in X$, $0 \leq \lambda \leq 1$. Согласно неравенству, следующие неравенства выполняются:

$$
f(x_1) - f(x_0) \geq \langle \nabla f(x_0), x_1 - x_0 \rangle + \frac{\mu}{2}\|x_1 - x_0\|^2,
$$

$$
f(x_2) - f(x_0) \geq \langle \nabla f(x_0), x_2 - x_0 \rangle + \frac{\mu}{2}\|x_2 - x_0\|^2.
$$

Доказательство. Достаточность

Предположим, что неравенство в теореме выполняется для всех $x, x_0 \in X$. Возьмем $x_0 = \lambda x_1 + (1-\lambda)x_2$, где $x_1, x_2 \in X$, $0 \leq \lambda \leq 1$. Согласно неравенству, следующие неравенства выполняются:

$$f(x_1) - f(x_0) \geq \langle \nabla f(x_0), x_1 - x_0 \rangle + \frac{\mu}{2}\|x_1 - x_0\|^2,$$

$$f(x_2) - f(x_0) \geq \langle \nabla f(x_0), x_2 - x_0 \rangle + \frac{\mu}{2}\|x_2 - x_0\|^2.$$

Умножая первое неравенство на λ и второе на $1-\lambda$ и складывая их, учитывая, что

$$x_1 - x_0 = (1-\lambda)(x_1 - x_2), \quad x_2 - x_0 = \lambda(x_2 - x_1),$$

Доказательство. Достаточность

Предположим, что неравенство в теореме выполняется для всех $x, x_0 \in X$. Возьмем $x_0 = \lambda x_1 + (1-\lambda)x_2$, где $x_1, x_2 \in X$, $0 \leq \lambda \leq 1$. Согласно неравенству, следующие неравенства выполняются:

$$
f(x_1) - f(x_0) \geq \langle \nabla f(x_0), x_1 - x_0 \rangle + \frac{\mu}{2}\|x_1 - x_0\|^2,
$$

$$
f(x_2) - f(x_0) \geq \langle \nabla f(x_0), x_2 - x_0 \rangle + \frac{\mu}{2}\|x_2 - x_0\|^2.
$$

Умножая первое неравенство на λ и второе на $1-\lambda$ и складывая их, учитывая, что

$$
x_1 - x_0 = (1-\lambda)(x_1 - x_2), \quad x_2 - x_0 = \lambda(x_2 - x_1),
$$

и $\lambda(1-\lambda)^2 + \lambda^2(1-\lambda) = \lambda(1-\lambda)$, мы получаем

$$
\begin{aligned}
\lambda f(x_1) + (1-\lambda) f(x_2) - f(x_0) - \frac{\mu}{2}\lambda(1-\lambda)\|x_1 - x_2\|^2 \geq \\
\langle \nabla f(x_0), \lambda x_1 + (1-\lambda)x_2 - x_0 \rangle = 0.
\end{aligned}
$$

Таким образом, неравенство из определения сильно выпуклой функции выполняется. Важно отметить, что $\mu = 0$ соответствует случаю выпуклой функции и соответствующему дифференциальному критерию.

Дифференциальный критерий сильной выпуклости второго порядка

Дважды дифференцируемая функция $f(x)$ определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$ называется μ-сильно выпуклой тогда и только тогда, когда $\forall x \in \mathbf{int}(S) \neq \emptyset$:

$$\nabla^2 f(x) \succeq \mu I$$

Другими словами:

$$\langle y, \nabla^2 f(x) y \rangle \geq \mu \|y\|^2$$

Дифференциальный критерий сильной выпуклости второго порядка

Дважды дифференцируемая функция $f(x)$ определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$ называется μ-сильно выпуклой тогда и только тогда, когда $\forall x \in \mathbf{int}(S) \neq \emptyset$:

$$\nabla^2 f(x) \succeq \mu I$$

Другими словами:

$$\langle y, \nabla^2 f(x) y \rangle \geq \mu \|y\|^2$$

Theorem

Пусть $X \subseteq \mathbb{R}^n$ выпуклое множество, с $\text{int}X \neq \emptyset$. Кроме того, пусть $f(x)$ дважды непрерывно дифференцируемая функция на X. Тогда $f(x)$ сильно выпукла на X с константой $\mu > 0$ тогда и только тогда, когда

$$\langle y, \nabla^2 f(x) y \rangle \geq \mu \|y\|^2$$

для всех $x \in X$ и $y \in \mathbb{R}^n$.

Доказательство. Необходимость

Целевое неравенство тривиально, когда $y = \mathbf{0}_n$, поэтому мы предполагаем $y \neq \mathbf{0}_n$.

Предположим, что x является внутренней точкой множества X. Тогда $x + \alpha y \in X$ для всех $y \in \mathbb{R}^n$ и достаточно малых α. Поскольку $f(x)$ дважды дифференцируема,

$$f(x + \alpha y) = f(x) + \alpha \langle \nabla f(x), y \rangle + \frac{\alpha^2}{2} \langle y, \nabla^2 f(x) y \rangle + o(\alpha^2).$$

Доказательство. Необходимость

Целевое неравенство тривиально, когда $y = \mathbf{0}_n$, поэтому мы предполагаем $y \neq \mathbf{0}_n$.

Предположим, что x является внутренней точкой множества X. Тогда $x + \alpha y \in X$ для всех $y \in \mathbb{R}^n$ и достаточно малых α. Поскольку $f(x)$ дважды дифференцируема,

$$f(x + \alpha y) = f(x) + \alpha \langle \nabla f(x), y \rangle + \frac{\alpha^2}{2} \langle y, \nabla^2 f(x) y \rangle + o(\alpha^2).$$

На основании первого дифференциального критерия сильной выпуклости:

$$\frac{\alpha^2}{2} \langle y, \nabla^2 f(x) y \rangle + o(\alpha^2) = f(x + \alpha y) - f(x) - \alpha \langle \nabla f(x), y \rangle \geq \frac{\mu}{2} \alpha^2 \|y\|^2.$$

Это неравенство сводится к целевому неравенству после деления обеих сторон на α^2 и перехода к пределу при $\alpha \downarrow 0$.

Если $x \in X$ но $x \notin \text{int}X$, рассмотрим последовательность $\{x_k\}$ такую, что $x_k \in \text{int}X$ и $x_k \to x$ при $k \to \infty$. Тогда, мы приходим к целевому неравенству после перехода к пределу.

Доказательство. Достаточность

Используя формулу Тейлора с остаточным членом Лагранжа и целевое неравенство, мы получаем для $x + y \in X$:

$$f(x+y) - f(x) - \langle \nabla f(x), y \rangle = \frac{1}{2} \langle y, \nabla^2 f(x + \alpha y) y \rangle \geq \frac{\mu}{2} \|y\|^2,$$

где $0 \leq \alpha \leq 1$. Следовательно,

Доказательство. Достаточность

Используя формулу Тейлора с остаточным членом Лагранжа и целевое неравенство, мы получаем для $x + y \in X$:

$$f(x+y) - f(x) - \langle \nabla f(x), y \rangle = \frac{1}{2}\langle y, \nabla^2 f(x + \alpha y) y \rangle \geq \frac{\mu}{2}\|y\|^2,$$

где $0 \leq \alpha \leq 1$. Следовательно,

$$f(x+y) - f(x) \geq \langle \nabla f(x), y \rangle + \frac{\mu}{2}\|y\|^2.$$

Следовательно, согласно первому дифференциальному критерию сильной выпуклости, функция $f(x)$ сильно выпукла с константой μ. Важно отметить, что $\mu = 0$ соответствует случаю выпуклой функции и соответствующему дифференциальному критерию.

Выпуклая и вогнутая функция

Example

Покажите, что $f(x) = c^\top x + b$ выпукла и вогнута.

Простейшая сильно выпуклая функция

Example

Покажите, что $f(x) = x^\top Ax$, где $A \succeq 0$ - выпукла на $\mathbb{R}^n$. Является ли она сильно выпуклой?

Выпуклость и непрерывность

Пусть $f(x)$ - выпуклая функция на выпуклом множестве $S \subseteq \mathbb{R}^n$. Тогда $f(x)$ непрерывна $\forall x \in \mathbf{ri}(S)$.

Собственная выпуклая функция

Функция $f : \mathbb{R}^n \to \mathbb{R}$ называется **собственной выпуклой функцией**, если она никогда не принимает значения $-\infty$ и не равна ∞ тождественно.

Индикаторная функция

$$\delta_S(x) = \begin{cases} \infty, & x \in S, \\ 0, & x \notin S, \end{cases}$$

является собственной выпуклой функцией.

Выпуклость и непрерывность

Пусть $f(x)$ - выпуклая функция на выпуклом множестве $S \subseteq \mathbb{R}^n$. Тогда $f(x)$ непрерывна $\forall x \in \mathbf{ri}(S)$.

Собственная выпуклая функция

Функция $f : \mathbb{R}^n \to \mathbb{R}$ называется **собственной выпуклой функцией**, если она никогда не принимает значения $-\infty$ и не равна ∞ тождественно.

Индикаторная функция

$$\delta_S(x) = \begin{cases} \infty, & x \in S, \\ 0, & x \notin S, \end{cases}$$

является собственной выпуклой функцией.

Замкнутая функция

Функция $f : \mathbb{R}^n \to \mathbb{R}$ называется **замкнутой**, если для каждого $\alpha \in \mathbb{R}$, множество подуровня замкнуто. Эквивалентно, если надграфик замкнут, то функция f замкнута.

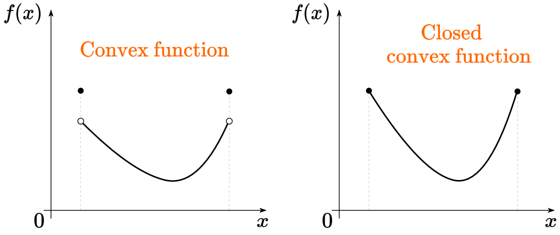


Рис. 15: Выпуклые функции могут иметь разрывы на границе своей области определения.

Факты о выпуклости

- $f(x)$ называется (строго, сильно) вогнутой, если функция $-f(x)$ - (строго, сильно) выпукла.
- Неравенство Йенсена для выпуклых функций:

$$f\left(\sum_{i=1}^{n} \alpha_i x_i\right) \leq \sum_{i=1}^{n} \alpha_i f(x_i)$$

для $\alpha_i \geq 0; \quad \sum\limits_{i=1}^{n} \alpha_i = 1$ (вероятностный симплекс)

Для непрерывного случая:

$$f\left(\int\limits_S x p(x) dx\right) \leq \int\limits_S f(x) p(x) dx$$

Если интегралы существуют и $p(x) \geq 0, \quad \int\limits_S p(x) dx = 1.$

- Если функция $f(x)$ и множество S выпуклы, то любой локальный минимум $x^* = \arg\min\limits_{x \in S} f(x)$ будет глобальным. Сильная выпуклость гарантирует единственность решения.

Другие формы выпуклости

- Логарифмическая выпуклость: $\log f$ выпукла; Логарифмическая выпуклость влечет выпуклость.
- Логарифмическая вогнутость: $\log f$ вогнута; **не** замкнута относительно сложения!
- Экспоненциальная выпуклость: $[f(x_i + x_j)] \succeq 0$, для $x_1, \ldots, x_n$
- Операторная выпуклость: $f(\lambda X + (1-\lambda)Y)$
- Квазивыпуклость: $f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\}$
- Псевдовыпуклость: $\langle \nabla f(y), x - y \rangle \geq 0 \longrightarrow f(x) \geq f(y)$
- Дискретная выпуклость: $f : \mathbb{Z}^n \to \mathbb{Z}$; "выпуклость + теория матроидов."

Условие Поляка-Лоясиевича. Линейная сходимость градиентного спуска без выпуклости

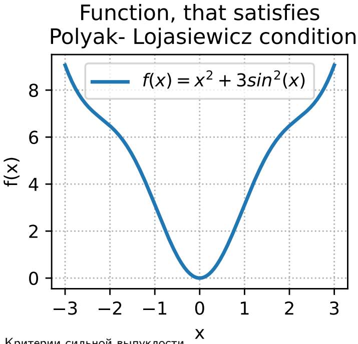
Неравенство PL выполняется, если выполняется следующее условие для некоторого $\mu > 0$,

$$\|\nabla f(x)\|^2 \geq 2\mu(f(x) - f^*)\forall x$$

При выполнении условия PL алгоритм градиентного спуска имеет линейную сходимость.

Следующие функции удовлетворяют условию PL, но не являются выпуклыми. Ссылка на код

$$f(x) = x^2 + 3\sin^2(x)$$



Условие Поляка-Лоясиевича. Линейная сходимость градиентного спуска без выпуклости

Неравенство PL выполняется, если выполняется следующее условие для некоторого $\mu > 0$,

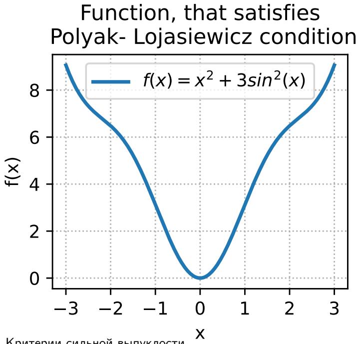
$$\|\nabla f(x)\|^2 \geq 2\mu(f(x) - f^*) \forall x$$

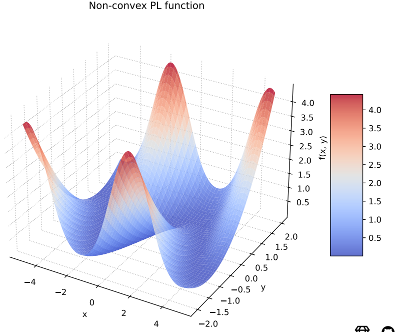
При выполнении условия PL алгоритм градиентного спуска имеет линейную сходимость.

Следующие функции удовлетворяют условию PL, но не являются выпуклыми. Ссылка на код

$$f(x) = x^2 + 3\sin^2(x)$$

$$f(x, y) = \frac{(y - \sin x)^2}{2}$$





Выпуклость в машинном обучении

Метод наименьших квадратов aka линейная регрессия

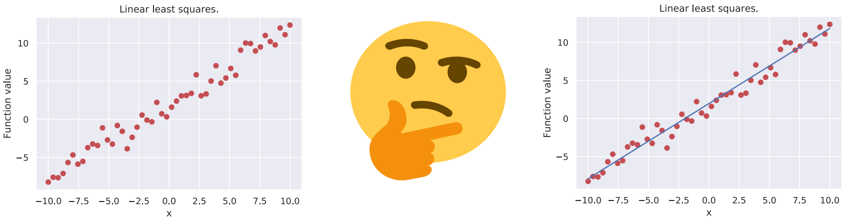


Рис. 18: Иллюстрация

В задаче линейной регрессии у нас есть измерения $X \in \mathbb{R}^{m \times n}$ и $y \in \mathbb{R}^m$ и мы ищем вектор $\theta \in \mathbb{R}^n$ такой, что $X\theta$ близок к y. Близость определяется как сумма квадратов разностей:

$$\sum_{i=1}^{m} (x_i^\top \theta - y_i)^2 = \|X\theta - y\|_2^2 \to \min_{\theta \in \mathbb{R}^n}$$

Например, рассмотрим набор данных, содержащий m пользователей, каждый из которых представлен n признаками. Каждая строка $x_i^\top$ матрицы признаков X соответствует признакам пользователя i, а соответствующий элемент y_i вектора откликов y представляет собой измеряемую величину, которую мы хотим предсказать на основе $x_i^\top$, например, расходы на рекламу. Предсказание значения осуществляется по формуле $x_i^\top \theta$.

Метод наименьших квадратов aka линейная регрессия [1]

1. Является ли эта задача выпуклой? Сильно выпуклой?

Метод наименьших квадратов aka линейная регрессия [1]

1. Является ли эта задача выпуклой? Сильно выпуклой?
2. Что вы думаете о сходимости градиентного спуска для этой задачи?

[1] Посмотрите на пример реальных данных линейной задачи метода наименьших квадратов

l_2-регуляризованный метод наименьших квадратов

В случае недоопределенной задачи может возникнуть желание восстановить сильную выпуклость целевой функции, добавив l_2-штраф, также известный как регуляризация Тихонова, l_2-регуляризация или демпфирование весов.

$$\|X\theta - y\|_2^2 + \frac{\mu}{2}\|\theta\|_2^2 \to \min_{\theta \in \mathbb{R}^n}$$

Примечание: С этой модификацией целевая функция снова становится μ-сильно выпуклой.

Посмотрите на код

Наиболее важная разница между выпуклостью и сильной выпуклостью

$$
f(x) = \frac{1}{2m}\|Ax - b\|_2^2 + \frac{\mu}{2}\|x\|_2^2 \to \min_{x \in \mathbb{R}^n}, \quad A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m
$$

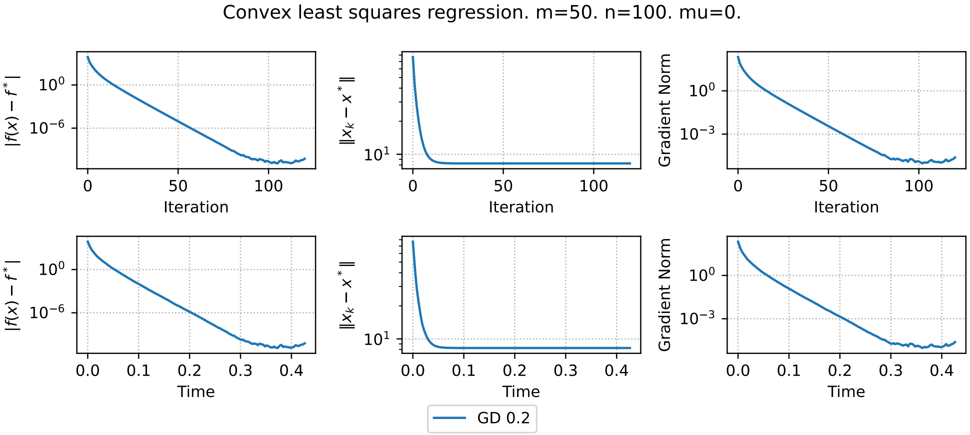


Рис. 19: Выпуклая задача не имеет сходимости по аргументу

Наиболее важная разница между выпуклостью и сильной выпуклостью

$$
f(x) = \frac{1}{2m}\|Ax - b\|_2^2 + \frac{\mu}{2}\|x\|_2^2 \to \min_{x \in \mathbb{R}^n}, \quad A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m
$$

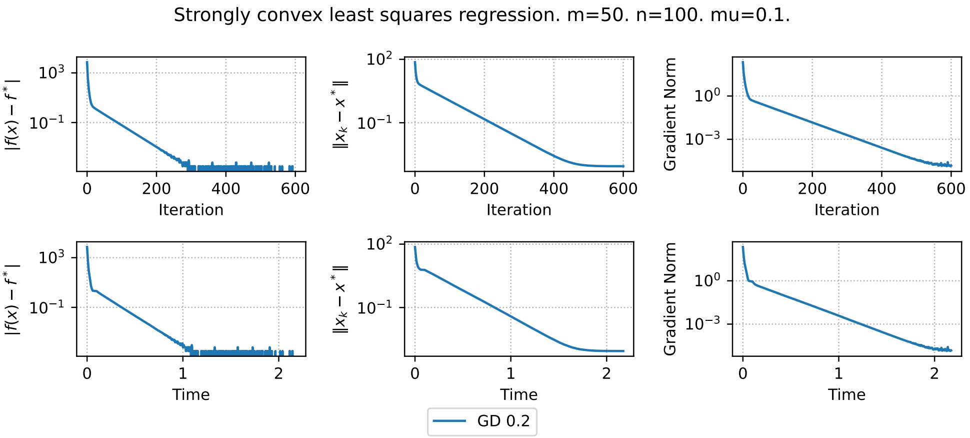


Рис. 20: Но если добавить даже небольшую регуляризацию, вы гарантируете сходимость по аргументу

Наиболее важная разница между выпуклостью и сильной выпуклостью

$$
f(x) = \frac{1}{2m}\|Ax - b\|_2^2 + \frac{\mu}{2}\|x\|_2^2 \to \min_{x \in \mathbb{R}^n}, \quad A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m
$$

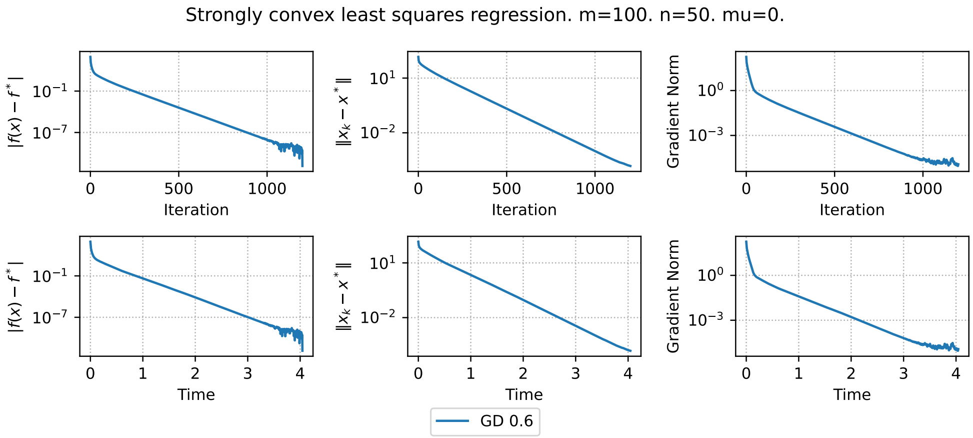


Рис. 21: Другой способ обеспечить сходимость в предыдущей задаче - поменять местами значения размерности задачи

Для сходимости к решению с высокой точностью необходима сильная выпуклость (или выполнение условия Поляка-Лоясиевича).

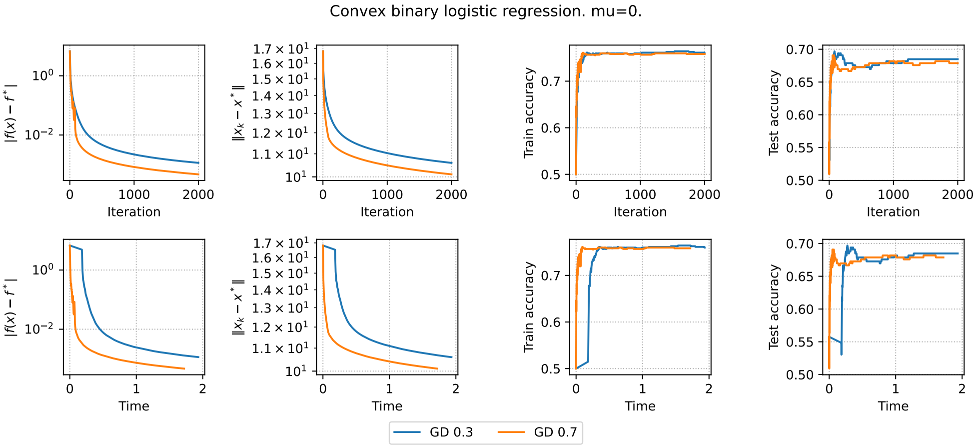


Рис. 22: Лишь небольшая точность может быть достигнута с сублинейной сходимостью

Для сходимости к решению с высокой точностью необходима сильная выпуклость (или выполнение условия Поляка-Лоясиевича).

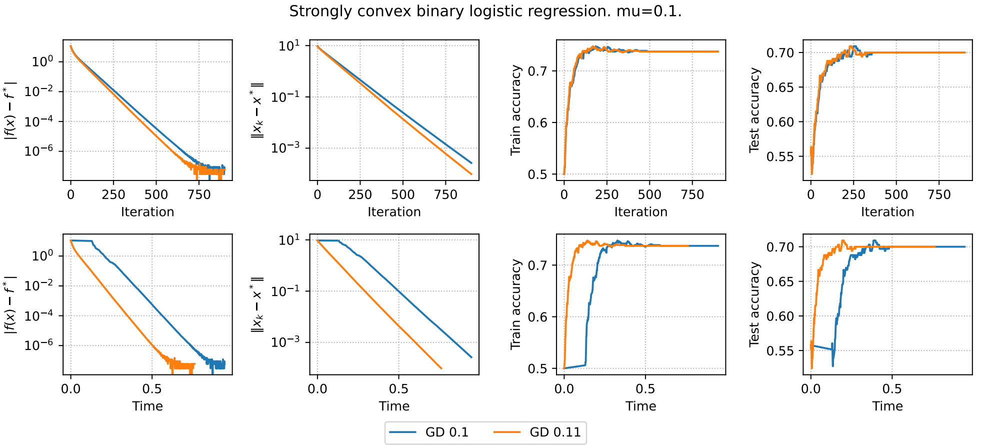


Рис. 23: Сильная выпуклость обеспечивает линейную сходимость

Любой локальный минимум является глобальным минимумом для глубоких линейных сетей [2]

Рассмотрим следующую задачу оптимизации:

$$\min_{W_1,\ldots,W_L} L(W_1,\ldots,W_L) = \frac{1}{2}\|W_L W_{L-1}\cdots W_1 X - Y\|_F^2,$$

где

$X \in \mathbb{R}^{d_x \times n}$ - матрица данных/входных данных,

$Y \in \mathbb{R}^{d_y \times n}$ - матрица меток/выходных данных.

Theorem

Пусть $k = \min(d_x, d_y)$ - “ширина” сети, и определим

$$V = \{(W_1,\ldots,W_L) \mid \mathrm{rank}(\Pi_i W_i) = k\}.$$

Тогда каждая критическая точка $L(W)$ в V является глобальным минимумом, в то время как каждая критическая точка в дополнении V^c является седловой точкой.

[2] Глобальные условия оптимальности для глубоких нейронных сетей