Определения и формулировки

1. Положительно определённая матрица.

   Матрица $A \in \mathbb{S}$ называется положительно (отрицательно) определённой, если для $\forall x \neq 0: \ x^T A x > (<) 0$. Обозначение: $A \prec 0$ $(A \succ 0)$.
   Аналогично определяется полуопределённость, только там неравенства нестрогие.

2. Евклидова норма вектора.

   $$||x||_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}$$

   Данная норма соответствует расстоянию в реальном мире. Иначе называется 2-норма (см. p-норма вектора)

3. Неравенство треугольника для нормы.

   Норма должна удовлетворять следующим свойствам:

   1. $||\alpha x|| = |\alpha| ||x||$, $\alpha \in \mathbb{R}$
   2. $||x|| = 0 \ \Rightarrow \ x = 0$
   3. $||x + y|| \leq ||x|| + ||y||$ – неравенство треугольника

4. p-норма вектора.

   $$||x||_p = \left(\sum_{i=0}^{n} |x_i|^p\right)^{\frac{1}{p}}$$

   Важные частные случаи:

   - Норма Чебышева: $||x||_\infty = \max\limits_i |x_i|$
   - Манхэттенское расстояние или $L1$ норма: $||x||_1 = \sum_{i=0}^{n} |x_i|$

5. Как выглядит единичный шар в p - норме на плоскости для $p = 1, 2, \infty$?

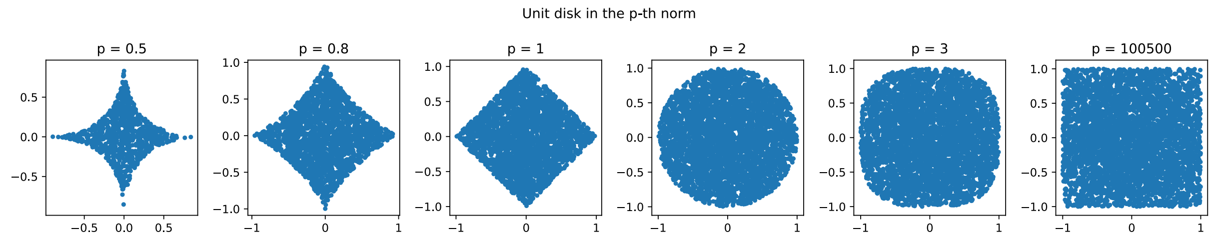


Figure 1: Шары в разных нормах

6. Норма Фробениуса для матрицы.

$$\|A\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{\frac{1}{2}}$$

7. Спектральная норма матрицы.

$$\|A\|_2 = \sup_{x\neq 0}\frac{\|Ax\|_2}{\|x\|_2} = \sigma_1(A) = \sqrt{\lambda_{max}(A^\top A)}$$

Где $\sigma_1(A)$ – старшее сингулярное значение A, $\lambda_{max}(A^\top A)$ – наибольшее собственное значение $A^\top A$.

8. Скалярное произведение двух векторов.

Пусть $x, y \in \mathbb{R}^n$, тогда их скалярное произведение это

$$\langle x, y\rangle = x^T y = \sum_{i=1}^{n} x_i y_i = y^T x = \langle y, x\rangle$$

9. Скалярное произведение двух матриц, согласованное с нормой Фробениуса.

Пусть $X, Y \in \mathbb{R}^{m\times n}$, тогда их скалярное произведение это

$$\langle X, Y\rangle = tr(X^T Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}Y_{ij} = tr(Y^T X) = \langle Y, X\rangle$$

Связь с нормой Фробениуса: $\langle X, X\rangle = ||X||_F^2$

10. Собственные значения матрицы. Спектр матрицы.

Скаляр λ является собственным значением для матрицы A, если существует вектор q, такой что $Aq = \lambda q$. В таком случае q называют собственным вектором.
Спектр матрицы – совокупность её собственных значений.

11. Связь спектра матрицы и её определенности.

Матрица положительно (неотрицательно) определена $\iff$ её спектр (все её собственные значения) положителен (неотрицателен).

12. Спектральное разложение матрицы.

Спектральное разложение матрицы, или разложение матрицы на основе собственных векторов, — это представление квадратной матрицы A в виде произведения трёх матриц $A = V\Lambda V^{-1}$, где V — матрица, столбцы которой являются собственными векторами матрицы A, Λ — диагональная матрица с соответствующими собственными значениями на главной диагонали. В таком виде могут быть представлены только матрицы, обладающие полным набором собственных векторов.
Тогда $A^n = V\Lambda^n V^{-1}$.

13. Сингулярное разложение матрицы.

$A \in \mathbb{R}^{m \times n}$, $rank\ A = r$.

$$A = U\Sigma V^T$$

$U \in \mathbb{R}^{m \times r}$, $U^TU = I$, $V \in \mathbb{R}^{n \times r}$, $V^TV = I$, Σ is a diagonal matrix with

$$\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r)$$

such that

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$$

Столбцы U, V - левые и правые собственные векторы A, σ_i - сингулярные значения.

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

14. Связь определителя и собственных чисел для квадратной матрицы.

Если у матрицы A собственные значения $\lambda_1, \lambda_2, \dots, \lambda_n$, то её определитель равен:

$$\det(A) = \lambda_1 \cdot \lambda_2 \cdot \dots \cdot \lambda_n$$

15. Связь следа и собственных чисел для квадратной матрицы.

Если у матрицы A собственные значения $\lambda_1, \lambda_2, \dots, \lambda_n$, то её след равен:

$$\text{tr}(A) = \lambda_1 + \lambda_2 + \dots + \lambda_n$$

16. Градиент функции $f(x) : \mathbb{R}^n \to \mathbb{R}$.

$\nabla f(x)$, вектор частных производных функции f.

17. Гессиан функции $f(x) : \mathbb{R}^n \to \mathbb{R}$.

$$f''(x) = \nabla^2 f(x) = \frac{\partial^2 f}{\partial x_i \partial x_j} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix}$$

18. Якобиан функции $f(x) : \mathbb{R}^n \to \mathbb{R}^m$.

$$J_f = f'(x) = \frac{df}{dx^T} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_n} & \frac{\partial f_2}{\partial x_n} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

19. Формула для аппроксимации Тейлора первого порядка $f^I_{x_0}(x)$ функции $f(x) : \mathbb{R}^n \to \mathbb{R}$ в точке x_0.

Для дифференцируемой f:

$$f^I_{x_0}(x) = f(x_0) + \nabla f(x_0)^T (x - x_0)$$

20. Формула для аппроксимации Тейлора второго порядка $f^{II}_{x_0}(x)$ функции $f(x) : \mathbb{R}^n \to \mathbb{R}$ в точке x_0.

Для дважды дифференцируемой f:

$$f^{II}_{x_0}(x) = f(x_0) + \nabla f(x_0)^T (x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0)$$

21. Определение дифференцируемости функции в точке через производную как линейный оператор.

Пусть $x \in S$ - внутренняя точка множества S, и пусть $D : U \to V$ - линейный оператор. Функция f называется дифференцируемой в точке x с производной D если для всех достаточно малых $h \in U$ верно следующее:

$$f(x + h) = f(x) + D[h] + o(\|h\|)$$

Если для любого линейного оператора $D : U \to V$ функция f не является дифференцируемой в точке x с производной D, тогда мы говорим, что f не дифференцируема в точке x.

22. Связь дифференциала функции df и градиента ∇f для функции $f(x) : \mathbb{R}^n \to \mathbb{R}$.

$$df(x) = \langle \nabla f(x), dx \rangle$$

23. Связь второго дифференциала функции $d^2 f$ и гессиана $\nabla^2 f$ для функции $f(x) : \mathbb{R}^n \to \mathbb{R}$.

$$d(df) = d^2 f(x) = \langle \nabla^2 f(x) dx_1, dx \rangle = \langle H_f(x) dx_1, dx \rangle$$

24. Формула для приближенного вычисления производной функции $f(x) : \mathbb{R}^n \to \mathbb{R}$ по k-ой координате с помощью метода конечных разностей.

$$\frac{\partial f}{\partial x_k}(x) \approx \frac{f(x + \varepsilon e_k) - f(x)}{\varepsilon}, \quad e_k = (0, \ldots, \underset{k}{1}, \ldots, 0)$$

Время работы: $2dT$, где вызов $f(x)$ занимает T,$x \in \mathbb{R}^d$

25. Пусть $f = f(x_1(t), \ldots, x_n(t))$. Формула для вычисления $\frac{\partial f}{\partial t}$ через $\frac{\partial x_i}{\partial t}$ (Forward chain rule).

$$\frac{\partial f}{\partial t} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \frac{\partial x_i}{\partial t}$$

26. Пусть L - функция, возвращающая скаляр, а v_k - функция, возвращающая вектор $x \in \mathbb{R}^t$. Формула для вычисления $\frac{\partial L}{\partial v_k}$ через $\frac{\partial L}{\partial x_i}$ (Backward chain rule).

$$\frac{\partial L}{\partial v_k} = \sum_{i=1}^{t} \frac{\partial L}{\partial x_i} \frac{\partial x_i}{\partial v_k}$$

27. Афинное множество. Афинная комбинация. Афинная оболочка.

Множество A называется аффинным если для любых x_1, x_2 из A прямая, проходящая через x_1, x_2, тоже лежит в A. То есть:

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

Пример аффинного множества: $\mathbb{R}^n$
Пусть $x_1, x_2, \ldots, x_k \in S$. Тогда точка $\theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_k x_k$ называется аффинной комбинацией, если

$$\forall i \in \{1, \ldots, k\} : \theta_i \in \mathbb{R}, \quad \sum_{i=1}^{k} \theta_i = 1$$

Аффинная оболочка – множество всех возможных аффинных комбинаций элементов множества.

$$\mathbf{aff}(S) = \left\{ \sum_{i=1}^{k} \theta_i x_i \middle| k > 0, x_i \in S, \theta_i \in \mathbb{R}, \sum_{i=1}^{k} \theta_i = 1 \right\}$$

28. Выпуклое множество. Выпуклая комбинация. Выпуклая оболочка.

Множество S называется выпуклым если для любых x_1, x_2 из S отрезок между x_1, x_2 тоже лежит в S. То есть:

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

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, если

$$\forall i \in \{1, \dots, k\}:\ \theta_i \geq 0, \quad \sum_{i=1}^{k} \theta_i = 1$$

Выпуклая оболочка – множество всех возможных выпуклых комбинаций элементов множества.

$$\mathrm{conv}(S) = \left\{ \sum_{i=1}^{k} \theta_i x_i \,\middle|\, k > 0,\ x_i \in S,\ \theta_i \geq 0,\ \sum_{i=1}^{k} \theta_i = 1 \right\}$$

29. Конус. Выпуклый конус. Коническая комбинация. Коническая оболочка.

Множество S называется конусом если для любого x из S луч, проходящий из 0 через x, тоже лежит в S. То есть:

$$\forall \theta \geq 0,\ \forall x \in S:\ \theta x \in S$$

Множество S называется выпуклым конусом если для любых $x_1, x_2 \in S$ их коническая комбинация тоже лежит в S. То есть:

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

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 конической комбинацией, если

$$\forall i \in \{1, \dots, k\}:\ \theta_i \geq 0$$

Коническая оболочка - множество всех возможных конических комбинаций элементов множества.

$$\mathrm{cone}(S) = \left\{ \sum_{i=1}^{k} \theta_i x_i \,\middle|\, k > 0,\ x_i \in S,\ \theta_i \geq 0 \right\}$$

30. Внутренность множества.

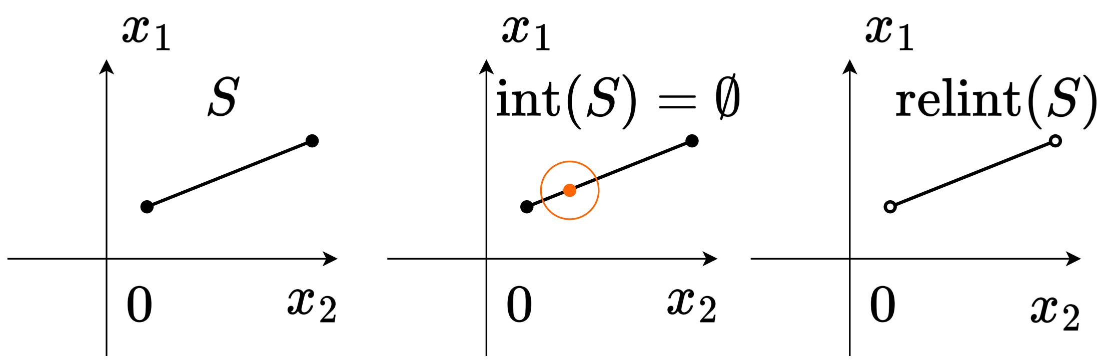
Внутренность множества - совокупность всех точек множества, содержащих вместе с собой в множестве некоторую окрестность вокруг себя.

31. Относительная внутренность множества.

Относительная внутренность множества - внутренность множества в его аффинной оболочке. Может быть полезной при работе с множествами меньшей размерности чем пространство, в котором они находятся.

$$\mathrm{relint}(S) = \{x \in S \mid \exists \varepsilon > 0,\ N_\varepsilon(x) \cap \mathrm{aff}(S) \subseteq S\}$$

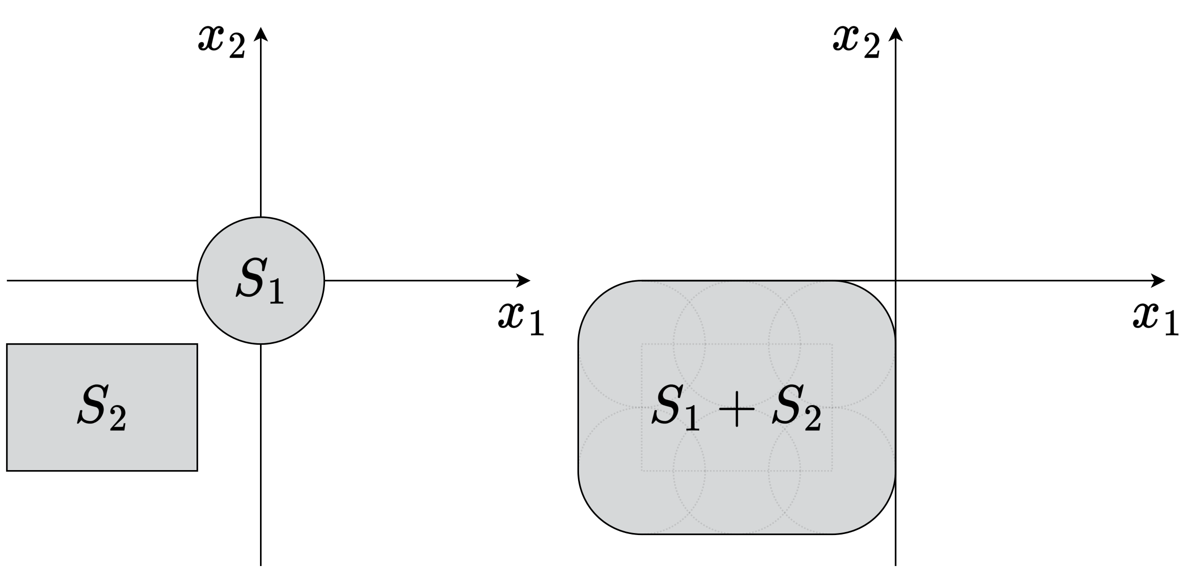
$N_\varepsilon(x)$ – шар радиуса ε с центром в x, $\mathrm{aff}(S)$ – аффинная оболочка S
Пример: отрезок на плоскости имеет пустую внутренность, но его относительная внутренность – тот же отрезок без концов.



32. Сумма Минковского.

Сумма Минковского – евклидово пространство, формирующееся сложением каждого вектора из S_1 с каждым вектором из S_2:

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



33. *Любые 2 операции с множествами, сохраняющие выпуклость.*

1. *Линейная комбинация:*

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

2. Пересечение любого числа выпуклых множеств

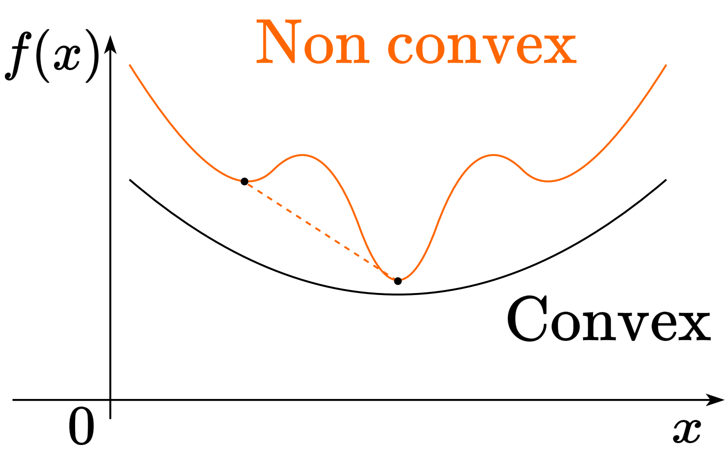
3. Образ множества в аффинном преобразовании:

$$S \subseteq \mathbb{R}^n \text{ convex} \to f(S) = \{f(x) | x \in S\} \text{ convex} \quad (f(x) = Ax + b)$$

34. Выпуклая функция.

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

$$\forall x_1, x_2 \in S, \quad \forall \lambda \in [0, 1]$$
$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda) f(x_2)$$



35. Строго выпуклая функция.

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

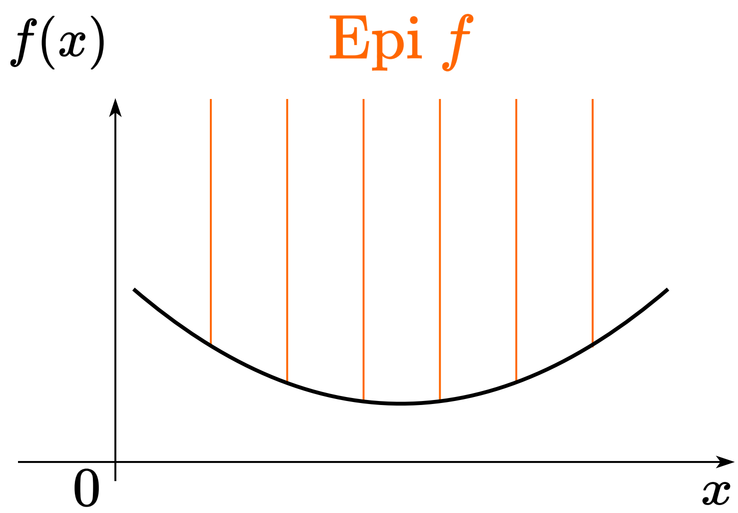
$$\forall x_1, x_2 \in S : x_1 \neq x_2, \quad \forall \lambda \in (0, 1)$$
$$f(\lambda x_1 + (1 - \lambda)x_2) < \lambda f(x_1) + (1 - \lambda) f(x_2)$$

36. Надграфик функции $f(x) : \mathbb{R}^n \to \mathbb{R}$.

Для функции, определённой на $S \subseteq \mathbb{R}^n$, множество:

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

называется надграфиком функции $f(x)$

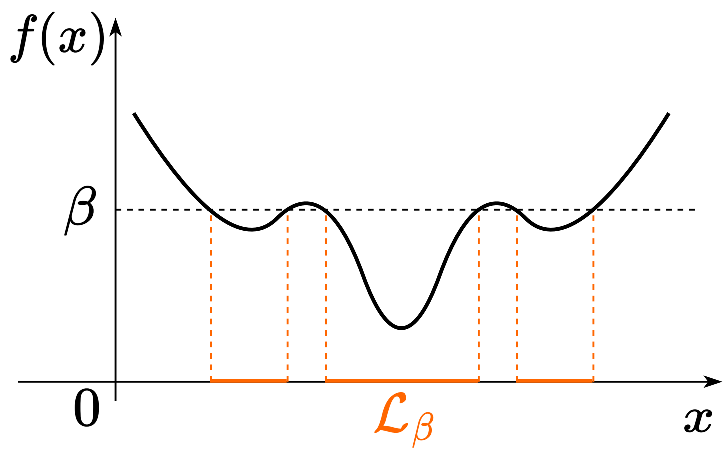


37. Множество подуровней функции $f(x) : \mathbb{R}^n \to \mathbb{R}$.

Для функции, определённой на $S \subseteq \mathbb{R}^n$, множество:

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

называется множеством подуровней или множеством Лебега функции $f(x)$
Если функция выпукла, то множество её подуровней выпукло. Обратное - не верно ($f(x) = \sqrt{|x|}$).



38. Дифференциальный критерий выпуклости первого порядка.

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

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

39. Дифференциальный критерий выпуклости второго порядка.

   Дважды дифференцируемая функция определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$ выпукла тогда и только тогда когда для любой внутренней точки x $\forall x \in \textbf{int}(S) \neq \emptyset$:

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

40. Связь выпуклости функции и её надграфика.

   Функция выпукла тогда и только тогда, когда её надграфик - выпуклое множество.

41. μ-сильно выпуклая функция.

   Функция определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$ называется сильно выпуклой если $\forall x_1, x_2 \in S$, $0 \leq \lambda \leq 1$ и $\mu > 0$:

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

42. Дифференциальный критерий сильной выпуклости первого порядка.

   Дифференцируемая функция определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$ является сильно выпуклой тогда и только тогда, когда существует $\mu > 0$: $\forall x, y \in S$:

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

43. Дифференциальный критерий сильной выпуклости второго порядка.

   Дважды дифференцируемая функция определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$ является сильно выпуклой тогда и только тогда, когда существует $\mu > 0$

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

44. Любые 2 операции с функциями, сохраняющие выпуклость.

   1. Сумма выпуклых функций с не отрицательными коэффициентами является выпуклой функцией.
   2. Композиция выпуклой функции с афинной выпукла: $g(x) = f(Ax + b)$
   3. Поточечный максимум любого числа выпуклых функций есть выпуклая функция.

45. Любые 2 нетривиальных свойства сопряженного множества.

1. Сопряженное множество всегда замкнуто, выпукло и содержит ноль.
2. Для произвольного множества $S \subseteq \mathbb{R}^n$: $S^{**} = \overline{\mathbf{conv}(S \cup \{0\})}$
3. Если $S_1 \subseteq S_2$, то $S_2^* \subseteq S_1^*$.
4. $\left(\bigcup\limits_{i=1}^m S_i\right)^* = \bigcap\limits_{i=1}^m S_i^*$.
5. Если S замкнуто, выпукло и включает 0, то $S^{**} = S$.
6. $S^* = \left(\overline{S}\right)^*$.

46. Является ли задача линейных наименьших квадратов для переопределенной линейной системы выпуклой/сильно выпуклой?

Рассмотрим задачу минимизации функции:

$$\|Ax - b\|^2 \to \min_{x \in \mathbb{R}^d},$$

где матрица $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $m > n$ (стоячая). Легко заметить, что гессиан минимизируемой функции $A^T A$ - невырожденная матрица размера $n \times n$. Она является положительно определенной. То есть задача в классической постановке является сильно выпуклой. Т.е. содержит единственный локальный минимум (единственное решение).

47. Является ли задача линейных наименьших квадратов для недоопределенной линейной системы выпуклой/сильно выпуклой?

Рассмотрим задачу минимизации функции:

$$\|Ax - b\|^2 \to \min_{x \in \mathbb{R}^d},$$

где матрица $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $m < n$ (лежачая). Легко заметить, что гессиан минимизируемой функции $A^T A$ - вырожденная матрица размера $n \times n$. Однако, она является положительно полуопределенной. То есть задача в классической постановке является выпуклой, но не сильно выпуклой. Т.е. содержит бесконечное количество локальных минимумов, каждый из которых - глобальный. Стоит отметить, что добавление ℓ_2 регуляризации к минимизируемой функции изменит задачу, однако, будет гарантировать сильную выпуклость.

48. Сопряжённое множество.

Пусть $S \subseteq \mathbb{R}^n$ - произвольное непустое множество. Тогда его сопряженное множество определяется как:

$$S^* = \{y \in \mathbb{R}^n \mid \langle y, x \rangle \geq -1 \ \ \forall x \in S\}$$

49. Сопряжённый конус.

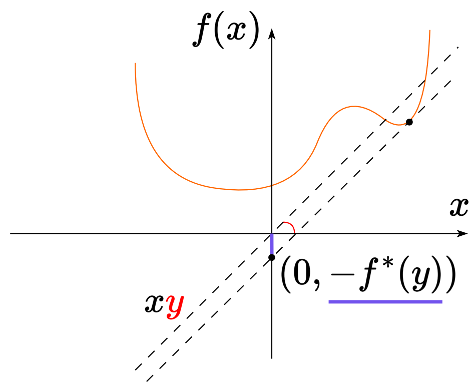
Сопряженным конусом к конусу K называется множество K^* такое, что:

$$K^* = \{y \mid \langle x, y\rangle \geq 0 \quad \forall x \in K\}$$

50. Сопряженная функция.

Для функции $f : \mathbb{R}^n \to \mathbb{R}$, сопряженной к ней называется функция f^*, причем область её определения можно считать теми y, для которых max конечен.

$$f^*(y) = \max_x \left[y^T x - f(x)\right]$$



51. Связь сильной выпуклости функции и гладкости сопряженной функции.

Пусть f - замкнутая и выпуклая. Тогда f - сильно выпуклая с константой выпуклости $\mu \Leftrightarrow \nabla f^*$ - липшицев с параметром $\frac{1}{\mu}$.

52. Сопряжённая норма. Сопряжённая норма к векторной p-норме.

Сопряжённой нормой $\|\cdot\|_*$ к норме $\|\cdot\|$ называется норма, определённая как:

$$\|y\|_* = \sup_{\|x\| \leq 1} \langle y, x\rangle,$$

где $x, y \in \mathbb{R}^n$.
Для p-нормы сопряжённой является q-норма, где p и q связаны соотношением:

$$\frac{1}{p} + \frac{1}{q} = 1, \quad p, q \geq 1.$$

Например:

- Для $p = 1$ сопряжённой является $q = \infty$.
- Для $p = 2$ сопряжённая норма также является 2-нормой.
- Для $p = \infty$ сопряжённой является $q = 1$.

53. Субградиент. Субдифференциал.

Субградиент функции f в точке x — это вектор g, удовлетворяющий условию:

$$f(y) \geqslant f(x) + g^T(y - x), \quad \forall y.$$

Множество всех субградиентов в точке x называется субдифференциалом и обозначается как $\partial f(x)$.

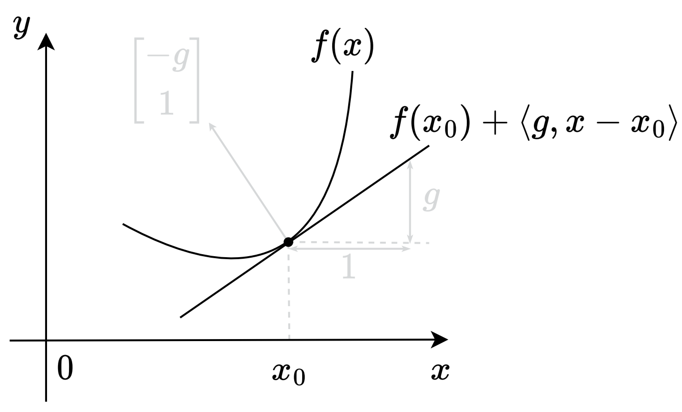
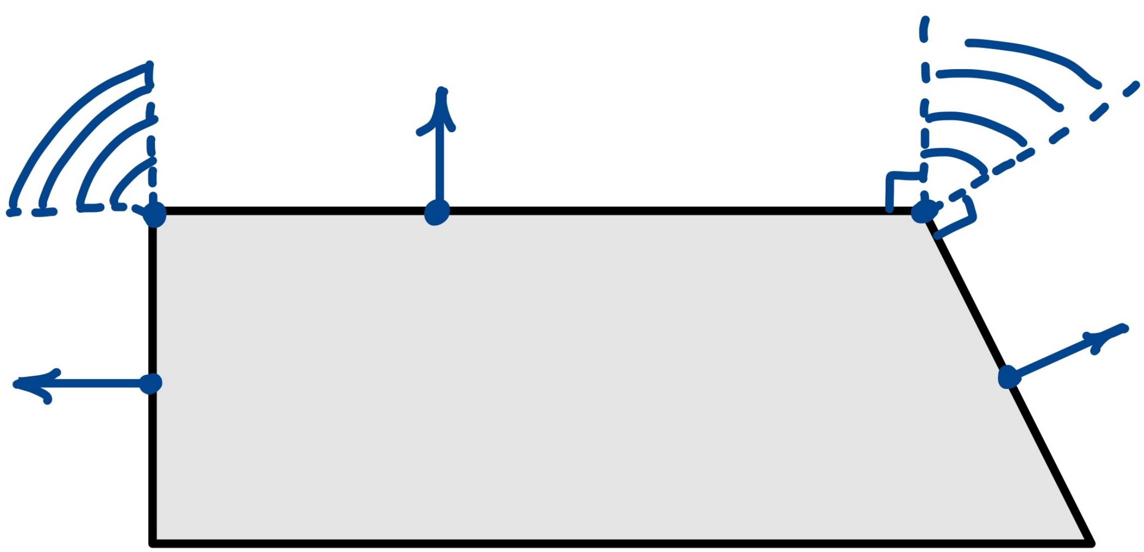


Figure 2: Субдифференциал функции ReLU.

54. Нормальный конус.

Для $x \in S$, $\partial I_S(x) = \mathcal{N}_S(x)$, **нормальный конус** к S в точке x, напомним.

$$\mathcal{N}_S(x) = \{g \in \mathbb{R}^n : g^T x \geq g^T y \text{ для любого } y \in S\}$$



55. Теорема Моро — Рокафеллара.

Пусть $f_i(x)$ — выпуклые функции, определённые на выпуклых множествах S_i, $i = \overline{1,n}$. Если выполнено условие $\bigcap\limits_{i=1}^{n} \mathbf{ri}(S_i) \neq \emptyset$, то функция $f(x) = \sum\limits_{i=1}^{n} a_i f_i(x)$, $a_i > 0$ имеет субдифференциал $\partial_S f(x)$ на множестве $S = \bigcap\limits_{i=1}^{n} S_i$, и его можно выразить следующим образом:

$$\partial_S f(x) = \sum_{i=1}^{n} a_i \partial_{S_i} f_i(x).$$

Это означает, что субдифференциал линейной комбинации выпуклых функций равен взвешенной сумме их субдифференциалов, взятых на пересечении соответствующих множеств.

56. Теорема Дубовицкого — Милютина.

Пусть $f_i(x)$ — выпуклые функции, определённые на открытом выпуклом множестве $S \subseteq \mathbb{R}^n$, и $x_0 \in S$. Пусть $f(x)$ определяется как покоординатный максимум этих функций:

$$f(x) = \max_i f_i(x).$$

Тогда субдифференциал $f(x_0)$ выражается следующим образом:

$$\partial_S f(x_0) = \mathbf{conv}\left\{ \bigcup_{i \in I(x_0)} \partial_S f_i(x_0) \right\},$$

где множество $I(x)$ определяется как индексы функций, достигающих максимума:

$$I(x) = \{i \in [1:m] : f_i(x) = f(x)\}.$$

Это утверждение говорит, что субдифференциал максимума выпуклых функций представляет собой выпуклую оболочку объединения субдифференциалов тех функций, которые достигают максимума в данной точке.

57. Теорема Вейерштрасса.

Пусть $S \subset \mathbb{R}^n$ - компакт, а $f(x)$ - непрерывная функция на S. Значит, точка глобального минимума функции $f(x)$ на S существует.

58. Теорема Тейлора.

$f : \mathbb{R}^n \to \mathbb{R}$ - непрерывная, дифференцируемая функция и $p \in \mathbb{R}^n$, тогда теорема Тейлора гласит:

$$f(x+p) = f(x) + \nabla f(x+tp)^T p$$

Для некоторого $t \in (0,1)\backslash$
Более того, если f - дважды дифференцируема, то:

$$f(x+p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x+tp) p$$

Для некоторого $t \in (0,1)$

59. Необходимые условия локального экстремума.

Если x^* - локальный экстремум и f непрерывная дифференцируема в открытой окрестности x^*, то:

$$\nabla f(x^*) = 0$$

60. Достаточные условия локального экстремума.

Если $\nabla^2 f$ непрерывна в открытой окрестности x^* и

$$\nabla f(x^*) = 0$$

$$\nabla^2 f(x^*) \succ 0$$

То x^* - локальный минимум $f(x)$.\ Для локального максимума аналогично, только

$$0 \succ \nabla^2 f(x^*)$$

61. Принцип Ферма для минимума функции.

Пусть $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, тогда x^* является глобальным минимумом f тогда и только тогда, когда

$$0 \in \partial f(x^*)$$

62. Общая задача математического программирования. Функция Лагранжа.

$$\begin{cases} f_0(x) \to \min\limits_{x \in \mathbb{R}^d} \\ f_i(x) \leq 0, \quad i = 1, \ldots, m, \\ h_i(x) = 0, \quad i = 1, \ldots, p. \end{cases}$$

Функция Лагранжа:

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x),$$

63. Теорема Каруша - Куна - Таккера в форме необходимых условий решения задачи математического программирования.

Пусть x_* - решение задачи с нулевым зазором двойственности

$$\begin{cases} f_0(x) \to \min\limits_{x \in \mathbb{R}^d} \\ f_i(x) \leq 0, \quad i = 1, \ldots, m, \\ h_i(x) = 0, \quad i = 1, \ldots, p. \end{cases}$$

Функция Лагранжа:

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x),$$

Тогда найдутся такие векторы~λ^* и~ν^*, что выполнены условия

$$\begin{cases} \nabla f_0(x_*) + \sum\limits_{i=1}^{m} \lambda_i^* \nabla f_i(x_*) + \sum\limits_{i=1}^{p} \nu_i^* \nabla h_i(x_*) = 0 \\ f_i(x_*) \leq 0, \quad i = 1, \ldots m \\ h_i(x_*) = 0, \quad i = 1, \ldots p \\ \lambda_i^* \geq 0, \quad i = 1, \ldots m \\ \lambda_i^* f_i(x_*) = 0, \quad i = 1, \ldots m \end{cases}$$

64. Условие Слейтера.

1. Если задача выпуклая (т.е., говоря о задаче минимизации, оптимизируемая функция f_0 и ограничения вида неравенство f_i – выпуклые, ограничения вида равенства h_i – аффинные)
2. И существует точка x такая, что $h(x) = 0$ и $f_i(x) < 0$ (ограничения вида равенства активные, а ограничения вида неравенства выполняются строго)

То тогда задача имеет нулевой зазор двойственности и условия ККТ становятся необходимыми и достаточными.

65. Задача выпуклого программирования.

Задача выпуклого программирования — это задача оптимизации, в которой целевая функция является выпуклой функцией и область допустимых решений выпукла. В форме ниже функции $f_0, \ldots, f_m$ - выпуклые, а функции h_i - афинные.

$$\begin{cases} f_0(x) \to \min\limits_{x \in \mathbb{R}^d} \\ f_i(x) \leq 0, \quad i = 1, \ldots, m, \\ h_i(x) = 0, \quad i = 1, \ldots, p. \end{cases}$$

66. Двойственная функция в задаче математического программирования.

Предположим, что $D = \bigcap_{i=0}^{m} \mathbf{dom}\, f_i \cap \bigcap_{i=0}^{p} \mathbf{dom}\, h_i$ непустое. Определим двойственную функцию $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ как минимум лагранжиана по x : для $\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p$

$$g(\lambda, \nu) = \inf_{x \in D} L(x, \lambda, \nu) = \inf_{x \in D} f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

Так как двойственная функция это поточечный инфинум семейства аффинных функций от (λ, ν), она вогнутая, даже если изначальная задача не выпуклая.

67. Двойственная задача для задачи математического программирования.

Пусть p^* - оптимальное решение изначальной задачи. Пусть $\hat{x}$ достижимая точка для изначальной задачи, т.е. $f_i(\hat{x}) \leq 0$ and $h_i(\hat{x}) = 0, \lambda \geq 0$. Тогда имеем:

$$L(\hat{x}, \lambda, \nu) = f_0(\hat{x}) + \underbrace{\lambda^T f(\hat{x})}_{\leq 0} + \underbrace{\nu^T h(\hat{x})}_{=0} \leq f_0(\hat{x})$$

Тогда

$$g(\lambda, \nu) = \inf_{x \in D} L(x, \lambda, \nu) \leq L(\hat{x}, \lambda, \nu) \leq f_0(\hat{x})$$

$$g(\lambda, \nu) \leq p^*$$

Двойственной задачей называется

$$g(\lambda, \nu) \to \max_{\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p}$$

$$s.t. \lambda \geq 0$$

68. Сильная двойственность. Зазор двойственности.

Пусть p^* -решение прямой задачи, d^* - решение двойственной задачи. Зазором двойственности называется

$$p^* - d^* \geq 0$$

Сильная двойственность возникает, если зазор равен нулю

$$p^* = d^*$$

69. Локальный анализ чувствительности с помощью множителей Лагранжа.

Перейдем к возмущенной версии задачи:

$$f_0(x) \to \min_x$$

$$f_i(x) \le u_i, \quad i = 1, \dots, m$$
$$h_i(x) = v_i, \quad i = 1, \dots, p,$$

Обозначим $p^*(u, v)$ - оптимальное решение этой задачи. Если имеет место сильная двойственность, то выполнено:

$$p^*(u, v) \ge p^*(0, 0) - {\lambda^*}^T u - {\nu^*}^T v$$

Если множители Лагранжа λ_i^*, ν_i^* большие, то небольшое изменение ограничений приведет к существенному изменению оптимального решения. То есть соответствующие ограничения очень сильно влияют на задачу.\ Если множители Лагранжа маленькие, то соответствующие ограничения мало влияют на задачу.

$$\lambda_i^* = -\frac{\partial p^*(0,0)}{\partial u_i} \quad \nu_i^* = -\frac{\partial p^*(0,0)}{\partial v_i}$$

70. Задача линейного программирования. Задача линейного программирования в стандартной форме.

Все задачи с линейным функционалом и линейными ограничениями считаются задачами линейного программирования. Стандартная форма:

$$\min_{x \in \mathbb{R}^n} c^T x$$

$$s.t. Ax = b$$

$$x_i \ge 0, i = 1, \dots, n$$

71. Возможные случаи двойственности в задаче линейного программирования.

Двойственная задача:

$$\max_{\nu \in \mathbb{R}^m} -b^T \nu$$

$$s.t. - A^T \nu \le c$$

1. Если либо у прямой, либо у двойственной задачи есть конечное решение, то и у другой тоже, и целевые переменные равны.
2. Если либо прямая, либо двойственная задача неограничена, то вторая из них невыполнима.

72. Симплекс метод.

Симплекс метод решает следующую задачу:

$$\min_{x \in \mathbb{R}^n} c^\top x$$

$$s.t. Ax \leq b$$

Шаги выполнения симплекс метода:

1. **Поиск начальной базисной допустимой точки:** Выберем начальную базисную (она является решением системы $A_B x = b_B$, где B - базис размера n пространства, а матрица A обычно имеет больше n ограничений) допустимую ($Ax_0 \leq b$) точку x_0 (искать ее будем через двухфазный симплексметод). Если такая точка не найдена, задача не имеет допустимого решения.

2. **Проверка оптимальности:**
   - Разложение вектора c в данном базисе B с коэффициентами λ_B:

     $$\lambda_B^\top A_B = c^\top \quad \text{или} \quad \lambda_B^\top = c^\top A_B^{-1}$$

   - Если все компоненты λ_B неположительны, текущий базис является оптимальным. Иначе далее меняем вершину симплекса.

3. **Определение переменной для удаления из базиса:**

4. **Вычисление шага вдоль выбранного направления d:**
   - Для всех $j \notin B$ считаем шаг:

     $$\mu_j = \frac{b_j - a_j^\top x_B}{a_j^\top d}$$

   - Новая вершина, которую добавим в базис:

     $$t = \arg\min_j \{\mu_j \mid \mu_j > 0\}$$

5. **Обновление базиса:**

6. **Повторение:**
   - Далее повторяем шаги 2-5 до достижения оптимального решения или установления, что задача не имеет допустимого решения.

73. Нахождение первоначальной угловой точки с помощью двухфазного симплекс метода.

1. Рассмотрим задачу (Phase 1):

$$\min_{\xi \in \mathbb{R}^m, y \in \mathbb{R}^n, z \in \mathbb{R}^n} \sum_{i=1}^{m} \xi_i$$

$$s.t. Ay - Az \leq b + \xi$$

$$y \geq 0, z \geq 0, \xi \geq 0$$

   Для нее есть допустимая угловая точка $z = 0, y = 0, \xi_i = \max(0, -b_i)$. Начиная с нее, решим задачу симплекс методом и получим точку оптимума, в которой $\xi = 0$ и выполнены указанные ограничения.
2. Решение задачи Phase 1 является допустимым базисом задачи Phase 2:

$$\min_{y \in \mathbb{R}^n, z \in \mathbb{R}^n} c^\top (y - z)$$

$$s.t. Ay - Az \leq b$$

$$y \geq 0, z \geq 0$$

3. Заметим, что оно так же будет являться допустимым базисом и угловой точкой для исходной задачи:

$$\min_{x \in \mathbb{R}^n} c^\top x$$

$$s.t. Ax \leq b$$

4. Так и нашли первоначальную угловую точку для исходной задачи.

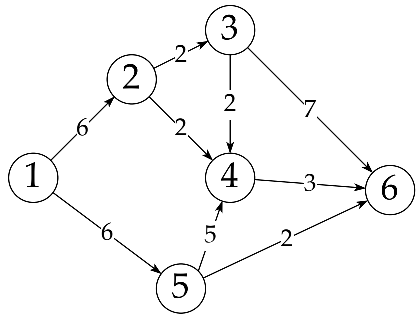
74. Сходимость симплекс метода.

   В худшем случае симплекс метод сходится экспоненциально от размерности задачи, но на практике в среднем алгоритм работает сильно лучше (полиномиально). Задача, на которой симплекс метод работает экспоненциальное время, называется примером Klee Minty.

75. Теорема о связи задач max-flow и min-cut (надо суметь описать обе задачи).

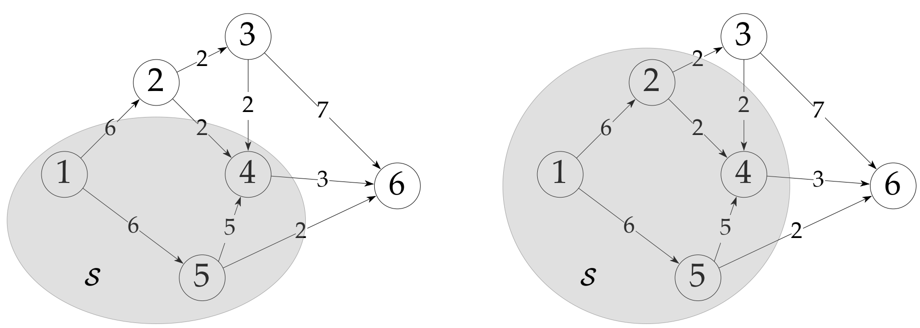
Задача Max-Flow (максимальный поток): Дано ориентированное взвешенное графовое представление сети, где узлы — это вершины графа, а ребра имеют пропускную способность (capacity). Требуется найти максимальный поток из источника (source) в сток (sink), при условии:

1. Поток на каждом ребре не превышает его пропускную способность.
2. Сохраняется закон сохранения потока в промежуточных узлах (входящий поток равен исходящему, за исключением источника и стока).



Задача Min-Cut (минимальный разрез): Для той же сети требуется найти разрез — разделение вершин графа на два множества (одно включает источник, другое — сток), при котором суммарная пропускная способность ребер, пересекающих разрез, минимальна.

$\mathcal{S} = \{1, 4, 5\}$ $\qquad$ $\mathcal{S} = \{1, 2, 4, 5\}$



Теорема Max-Flow Min-Cut: Максимальный поток из источника в сток равен минимальной пропускной способности разреза между источником и стоком.
Формально:

$$\text{MAXFLOW} = \text{MINCUT}.$$

Эта теорема утверждает, что задачи поиска максимального потока и минимального разреза являются двойственными: решение одной задачи предоставляет решение другой.

76. *Линейная сходимость последовательности.*

> Пусть есть последовательность $\{\|x_k - x^*\|_2\}$ в $\mathbb{R}$, сходящаяся к 0.
> *Линейная* сходимость при $q \in (0, 1)$ (скорость сходимости) и $C \in (0, \infty)$ (константа сходимости) определяется одним из двух способов:
>
> $$\|x_{k+1} - x^*\| \le Cq^k \text{ или } \|x_{k+1} - x^*\| \le q\|x_k - x^*\|$$
>
> Чем меньше q, тем быстрее сходится последовательность.
> По-другому, говорят, что последовательность x_k сходится к числу L. Мы говорим, что эта последовательность *линейно* сходится к L, если $\exists$ число $\mu \in (0, 1)$, такое, что
>
> $$\lim_{k\to\infty} \frac{|x_{k+1} - L|}{|x_k - L|} = \mu$$
>
> и μ называется скоростью сходимости.

77. Сублинейная сходимость последовательности.

> Если последовательность r_k сходится к нулю, но не обладает *линейной* сходимостью, то говорят, что она сходится сублинейно. Иногда мы можем рассматривать следующий класс сублинейной сходимости:
>
> $$|x_{k+1} - x^*|_2 \le Ck^q,$$
>
> где $q < 0$ и $0 < C < \infty$.

78. Сверхлинейная сходимость последовательности.

> Мы определяем сверхлинейную сходимость как сходимость последовательности, которая быстрее любой *линейной* сходимости. Иногда рассматривают более специальный класс. Тогда говорят, что сверхлинейная сходимость при $q > 1$, $C > 0$ определяется следующим образом:
>
> $$|x_{k+1} - x^*| \le C|x_k - x^*|^q$$

79. Квадратичная сходимость последовательности.

> Квадратичная сходимость является частным случаем сверхлинейной сходимости, когда $q = 2$. Она определяется следующим образом:
>
> $$|x_{k+1} - x^*| \le C|x_k - x^*|^2$$
>
> Или по-другому:
>
> $$\lim_{k\to\infty} \frac{|x_{k+1} - L|}{|x_k - L|^2} = \mu$$
>
> где $\mu > 0$.

80. Тест корней для определения скорости сходимости последовательности.

Пусть $(r_k)_{k=m}^\infty$ - последовательность неотрицательных чисел, сходящаяся к нулю, и пусть $\alpha := \limsup_{k\to\infty} r_k^{1/k}$. (Заметим, что $\alpha \geq 0$.)

1. Если $0 \leq \alpha < 1$, то $(r_k)_{k=m}^\infty$ сходится линейно с константой α.
2. В частности, если $\alpha = 0$, то $(r_k)_{k=m}^\infty$ сходится сверхлинейно.
3. Если $\alpha = 1$, то $(r_k)_{k=m}^\infty$ сходится сублинейно.
4. Случай $\alpha > 1$ невозможен.

81. Тест отношений для определения скорости сходимости последовательности.

Пусть $r_{k_{k=m}}^\infty$ - последовательность строго положительных чисел, сходящаяся к нулю. Пусть

$$q = \lim_{k\to\infty} \frac{r_{k+1}}{r_k}$$

1. Если существует q и $0 \leq q < 1$, то $r_{k_{k=m}}^\infty$ имеет линейную сходимость с константой q.
2. В частности, если $q = 0$, то $r_{k_{k=m}}^\infty$ имеет сверхлинейную сходимость.
3. Если q не существует, но $q = \lim_{k\to\infty} \sup_k \frac{r_{k+1}}{r_k} < 1$, то $r_{k_{k=m}}^\infty$ имеет линейную сходимость с константой, не превышающей q.
4. Если $\lim_{k\to\infty} \inf_k \frac{r_{k+1}}{r_k} = 1$, то $r_{k_{k=m}}^\infty$ имеет сублинейную сходимость.
5. Случай $\lim_{k\to\infty} \inf_k \frac{r_{k+1}}{r_k} > 1$ невозможен.

82. Унимодальная функция.

Функция $f(x)$ называется унимодальной на $[a,b]$, если существует $x^* \in [a,b]$, такое, что

1. $f(x_1) > f(x_2)$ для всех $a \leq x_1 < x_2 < x^*$
2. $f(x_1) < f(x_2)$ для всех $x^* < x_1 < x_2 \leq b$

83. Метод дихотомии.

Наша цель - решить следующую задачу: $\min_{x\in[a,b]} f(x)$ Мы делим отрезок на две равные части и выбираем ту, которая содержит решение задачи, используя значения функции, опираясь на ключевое свойство, описанное выше. Наша цель после одной итерации метода - уменьшить область поиска решения в два раза (в среднем). Метод описан на рисунках ниже.

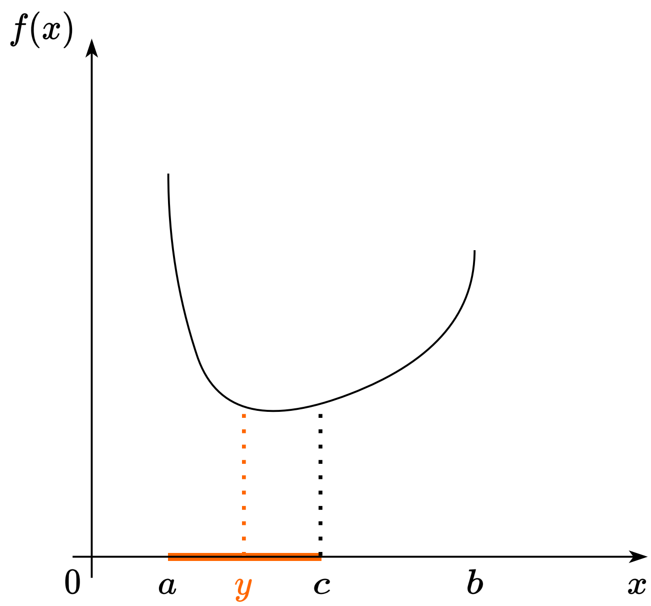


Figure 3: Диаграмма метода дихотомии

Длина отрезка на $(k+1)$-ой итерации:

$$\Delta_{k+1} = b_{k+1} - a_{k+1} = \frac{1}{2^k}(b-a)$$

Для унимодальных функций:

$$|x_{k+1} - x^*| \leq \frac{\Delta_{k+1}}{2} \leq \frac{1}{2^{k+1}}(b-a) \leq (0.5)^{k+1} \cdot (b-a)$$

Заметим, что на каждой итерации мы обращаемся к оракулу не более чем два раза, поэтому число вычислений функции равно $N = 2 \cdot k$, что подразумевает:

$$|x_{k+1} - x^*| \leq (0.5)^{\frac{N}{2}+1} \cdot (b-a) \leq (0.707)^N \frac{b-a}{2}$$

84. Метод золотого сечения.

Общая идея: хотим поделить отрезок на 3 части так, чтобы потом когда одна из частей отпадет на следующей итерации одно из нужных значений функций будет уже известно.

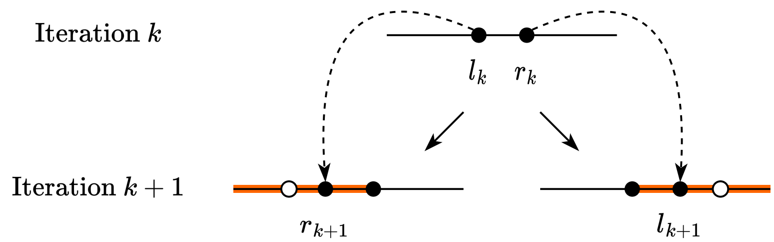


Figure 4: Иллюстрация метода золотого сечения

```
def golden_search(f, a, b, epsilon):
    tau = (sqrt(5) + 1) / 2
    y = a + (b - a) / tau**2
    z = a + (b - a) / tau
    while b - a > epsilon:
        if f(y) <= f(z):
            b = z
            z = y
            y = a + (b - a) / tau**2
        else:
            a = y
            y = z
            z = a + (b - a) / tau
    return (a + b) / 2
```

85. Метод параболической интерполяции (без точных формул).

Идея метода: берем 3 точки, по этим 3 точкам однозначно строим параболу, находим ее минимум, и из этих 4 точек оставляем 3 так, чтобы между первой и третьей находился минимум.

```
def parabola_search(f, x1, x2, x3, epsilon):
    f1, f2, f3 = f(x1), f(x2), f(x3)
    while x3 - x1 > epsilon:
        u = x2 - ((x2 - x1)**2*(f2 - f3) - (x2 - x3)**2*(f2 - f1))/(2*((x2 - x1)*(f2 - f3)
        fu = f(u)

        if x2 <= u:
            if f2 <= fu:
                x1, x2, x3 = x1, x2, u
                f1, f2, f3 = f1, f2, fu
            else:
                x1, x2, x3 = x2, u, x3
                f1, f2, f3 = f2, fu, f3
        else:
            if fu <= f2:
                x1, x2, x3 = x1, u, x2
                f1, f2, f3 = f1, fu, f2
            else:
                x1, x2, x3 = u, x2, x3
                f1, f2, f3 = fu, f2, f3
    return (x1 + x3) / 2
```

Сходится сверхлинейно, но метод довольно неустойчивый. Если $f(x)$ не похожа на параболу, нам конец. Если она обратна параболе, то мы и вовсе уйдём искать максимум.

86. Условие достаточного убывания для неточного линейного поиска.

Неточный линейный поиск:

$$x_{k+1} = x_k - \alpha \nabla f(x_k) \alpha = \text{argmin } f(x_{k+1})$$

Хотим приближенно найти α. Сведем задачу к поиску минимума следующей функции:

$$\phi(\alpha) = f(x_k - \alpha \nabla f(x_k)), \alpha \geq 0$$

Приблизим ее через первые 2 члена ряда Тейлора:

$$\phi(\alpha) \approx f(x_k) - \alpha \nabla f(x_k)^\top \nabla f(x_k)$$

Тогда условием достаточного убывания (Armijo condition) является:

$$f(x_k - \alpha \nabla f(x_k)) \leq f(x_k) - c_1 \cdot \alpha \nabla f(x_k)^\top \nabla f(x_k), c_1 \in (0, 1)$$

Иллюстрация для понимания:

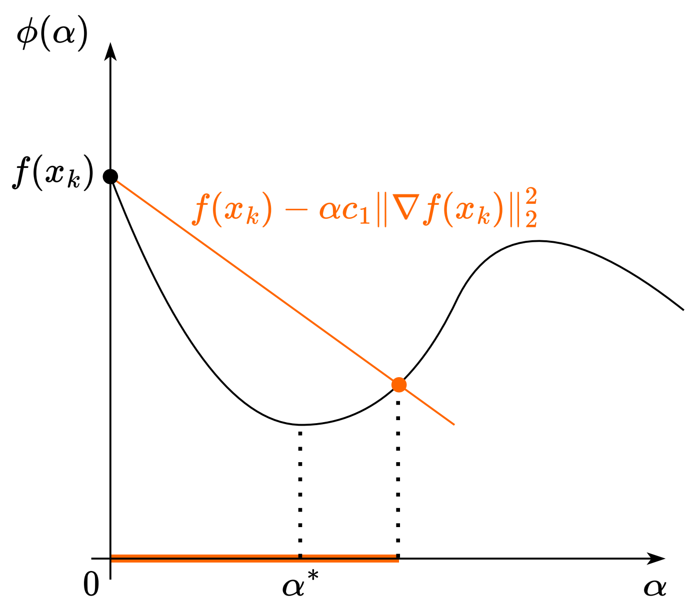


Figure 5: Иллюстрация условия достаточного убывания

87. Условия Гольдштейна для неточного линейного поиска.

Определим ϕ_1 и ϕ_2 следующим образом ($c_1 > c_2$)

$$\phi_1(\alpha) = f(x_k) - c_1\alpha\|\nabla f(x_k)\|^2$$

$$\phi_2(\alpha) = f(x_k) - c_2\alpha\|\nabla f(x_k)\|^2$$

Тогда условие Гольдштейна заключается в том, что $\phi_1(\alpha) \le \phi(\alpha) \le \phi_2(\alpha)$.
Иллюстрация для понимания:

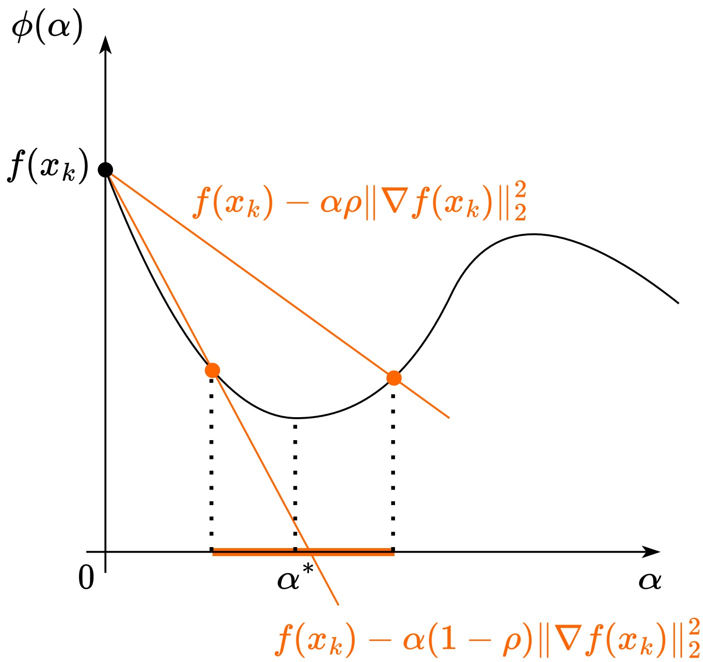


Figure 6: Иллюстрация условий Гольдштейна

88. Условие ограничения на кривизну для неточного линейного поиска.

$$-\nabla f(x_k - \alpha \nabla f(x_k))^\top \nabla f(x_k) \geq c_2 \nabla f(x_k)^\top (-\nabla f(x_k)),$$

где $c_2 \in (c_1, 1)$, и c_1 взято из условия достаточного убывания.
Иллюстрация для понимания:

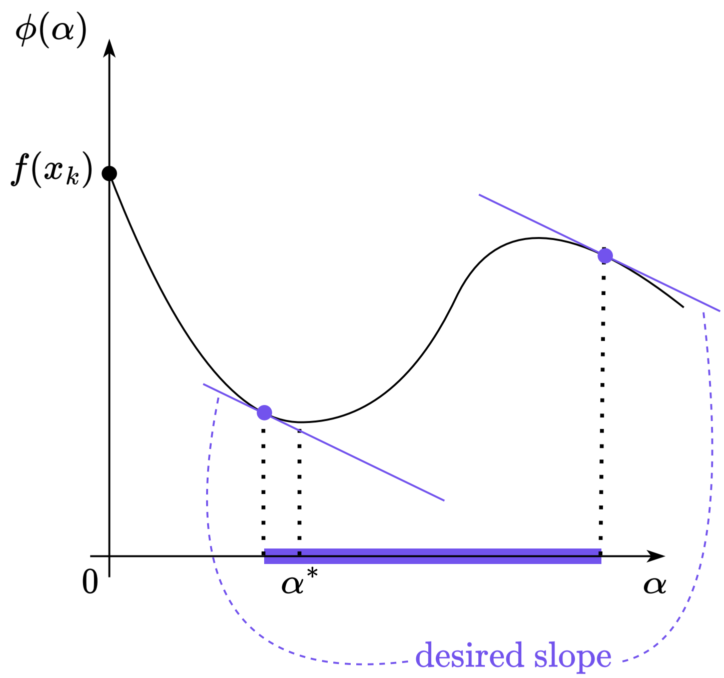


Figure 7: Иллюстрация условия ограничения на кривизну

Теоремы с доказательствами

1. Критерий положительной определенности матрицы через знаки собственных значений матрицы.

$A \succeq (\succ) 0 \iff$ все собственные значения матрицы $A \geq (>) 0$

$\rightarrow$ Пусть некоторые собственные значения λ отрицательны, и x - соответствующий ему собственный вектор. Тогда:

$Ax = \lambda x, x^T Ax \geq 0 \rightarrow x^T Ax = \lambda x^T x$, $x^T x \geq 0 \rightarrow \lambda \geq 0$ - противоречие

$\leftarrow$ Помним, что положительная определённость задаётся для симметричных матриц. Для симметричной матрицы можем выбрать собственные векторы v_i, образующие ортогональный базис ($i \neq j : v_i^T v_j = 0$ - выкиываем часть слогаемых из суммы в доказательстве). Тогда для $x \in \mathbb{R}^n$

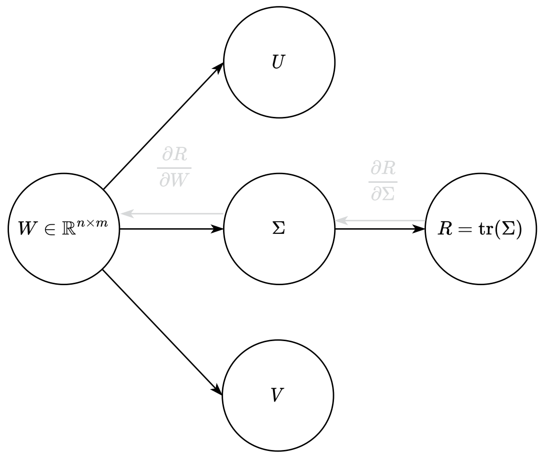
$$x^T Ax = (\alpha_1 v_1 + \dots + \alpha_n v_n)^T A(\alpha_1 v_1 + \dots + \alpha_n v_n) = \sum \alpha_i^2 v_i^T A v_i = \sum \alpha_i^2 v_i^T \lambda v_i$$

Так как $\lambda_i \geq 0$, то и вся сумма неотрицательна.

2. Связь $\dfrac{\partial L}{\partial W}$ и $\dfrac{\partial L}{\partial \Sigma}$, если $W = U\Sigma V^T \in \mathbb{R}^{m \times n}$, при этом

$$U^T U = I, \quad V^T V = I, \quad \Sigma = \text{diag}(\sigma_1, \dots, \sigma_{\min(m,n)})$$

> ℹ
> $$\frac{\partial L}{\partial W} = U \frac{\partial L}{\partial \Sigma} V^T$$



Пусть у нас есть прямоугольная матрица $W \in \mathbb{R}^{m \times n}$, имеющая сингулярное разложение:

$$W = U\Sigma V^T, \quad U^T U = I, \quad V^T V = I, \quad \Sigma = \text{diag}(\sigma_1, \dots, \sigma_{\min(m,n)})$$

1. Отметим:

$$\begin{aligned} W &= U\Sigma V^T \\ dW &= dU\Sigma V^T + Ud\Sigma V^T + U\Sigma dV^T \\ U^T dWV &= U^T dU\Sigma V^T V + U^T U d\Sigma V^T V + U^T U\Sigma dV^T V \\ U^T dWV &= U^T dU\Sigma + d\Sigma + \Sigma dV^T V \end{aligned}$$

2. Заметим, что $U^T U = I \to dU^T U + U^T dU = 0$. Но также $dU^T U = (U^T dU)^T$, из чего фактически следует, что матрица $U^T dU$ антисимметрична:

$$(U^T dU)^T + U^T dU = 0 \quad \to \quad \text{diag}(U^T dU) = (0, \dots, 0)$$

Та же логика может быть применена к матрице V и

$$\text{diag}(dV^T V) = (0, \dots, 0)$$

3. В то же время матрица $d\Sigma$ диагональная, а это значит (посмотрите на 1.), что

$$\text{diag}(U^T dWV) = d\Sigma$$

Здесь с обеих сторон мы имеем диагональные матрицы.

4. Теперь мы можем разложить дифференциал функции потерь как функцию от Σ - такие задачи возникают в ML, где нужно ограничить ранг матрицы:

$$
\begin{aligned}
dL &= \left\langle \frac{\partial L}{\partial \Sigma}, d\Sigma \right\rangle \\
&= \left\langle \frac{\partial L}{\partial \Sigma}, \text{diag}(U^T dW V) \right\rangle \\
&= \text{tr}\left(\frac{\partial L}{\partial \Sigma}^T \text{diag}(U^T dW V) \right)
\end{aligned}
$$

5. Поскольку внутри произведения находятся диагональные матрицы, след диагональной части матрицы будет равен следу всей матрицы:

$$
\begin{aligned}
dL &= \text{tr}\left(\frac{\partial L}{\partial \Sigma}^T \text{diag}(U^T dW V) \right) \\
&= \text{tr}\left(\frac{\partial L}{\partial \Sigma}^T U^T dW V \right) \\
&= \left\langle \frac{\partial L}{\partial \Sigma}, U^T dW V \right\rangle \\
&= \left\langle U \frac{\partial L}{\partial \Sigma} V^T, dW \right\rangle
\end{aligned}
$$

3. Базовые операции, сохраняющие выпуклость множеств: пересечение бесконечного числа множеств, линейная комбинация множеств, образ афинного отображения.

> - Пересечение любого (!) количества выпуклых множеств — выпуклое множество.
> - *Линейная* комбинация выпуклых множеств выпукла.
> - Образ выпуклого множества после применения афинного отображения — выпуклое множество.

Пересечение бесконечного числа множеств

Пересечение любого (!) количества выпуклых множеств — выпуклое множество.

Если итоговое пересечение пустое или содержит одну точку, то свойство выпуклости выполняется по определению. Иначе возьмем 2 точки и отрезок между ними. Эти точки должны лежать во всех пересекаемых множествах. Так как все пересекаемые множества выпуклы, отрезок между этими двумя точками лежит во всех множествах. А значит, отрезок лежит и в их пересечении.

***Линейная* комбинация множеств**

Линейная комбинация выпуклых множеств выпукла.

Пусть есть 2 выпуклых множества S_x, S_y, рассмотрим их линейную комбинацию

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

Возьмем две точки 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

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

$$\theta(c_1x_1 + c_2y_1) + (1-\theta)(c_1x_2 + c_2y_2)$$
$$c_1(\theta x_1 + (1-\theta)x_2) + c_2(\theta y_1 + (1-\theta)y_2)$$
$$c_1x + c_2y \in S$$

Образ афинного отображения

Образ выпуклого множества после применения афинного отображения — выпуклое множество.

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

Доказательство

При $\theta \in [0,1]; x, y \in S, S$ — выпуклое. Тогда и $\theta x + (1-\theta)y \in S$. В то же время $f(\theta x + (1-\theta)y) = \theta Ax + \theta b + (1-\theta)Ay + (1-\theta)b = \theta Ax + (1-\theta)Ay + b = \theta f(x) + (1-\theta)f(y)$. В итоге мы доказали, что образ $f(S)$ — тоже выпуклый, так как $\forall \theta \in [0,1], x, y \in S$ выполняется $\theta f(x) + (1-\theta)f(y) \in f(S)$.

Примеры афинных функций: растяжение, сжатие, проекция, транспонирование, множество решений линейного матричного неравенства $\{x \mid x_1A_1 + \ldots + x_mA_m \preceq B\}$. Здесь $A_i, B \in \mathbf{S}^p$ — симметричные матрицы $p \times p$.

Заметим также, что прообраз выпуклого множества при аффинном отображении также является выпуклым.

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

4. Неравенство Йенсена для выпуклой функции и выпуклой комбинации точек.

> Пусть $f(x)$ – выпуклая функция, определённая на выпуклом множестве $S \subseteq \mathbb{R}^n$. Тогда для точек $x_1, \ldots, x_m \in S$ выполнено неравенство:
>
> $$f\left(\sum_{i=1}^m \lambda_i x_i\right) \leq \sum_{i=1}^m \lambda_i f(x_i)$$
>
> $\lambda = [\lambda_1, \ldots, \lambda_m] \in \Delta_m$.

1. Заметим, что $\sum\limits_{i=1}^m \lambda_i x_i$ является выпуклой комбинацией элементов S и лежит в S.
2. Доказательство по индукции. Для $m = 1$ очевидно, для $m = 2$ следует из определения выпуклой функции.
3. Пусть неравенство верно для $m = 1, \ldots, k$, докажем для $m = k+1$. Пусть $\lambda \in \Delta_{k+1}$, $x = \sum\limits_{i=1}^{k+1} \lambda_i x_i = \lambda_{k+1}x_{k+1} + \sum\limits_{i=1}^{k} \lambda_i x_i$. При $\lambda_i = 0$ либо 1 выражение сводится к уже рассмотренным случаям, далее полагаем $0 < \lambda_i < 1$:

$$x = \lambda_{k+1}x_{k+1} + (1-\lambda_{k+1})\sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}} x_i = \lambda_{k+1}x_{k+1} + (1-\lambda_{k+1})\hat{x}$$

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

4. Так как $\lambda \in \Delta_{k+1}$, то $\gamma = [\gamma_1, \dots, \gamma_k] \in \Delta_k$. Значит, $\hat{x} \in S$, из выпуклости $f(x)$ и предположения индукции следует:

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

5. Выпуклость надграфика как критерий выпуклости функции.

> Чтобы функция $f(x)$, определенная на выпуклом множестве X, была выпуклой на X, необходимо и достаточно чтобы надграфик f был выпуклым множеством.

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 $f(x)$ (здесь $\mu \in$, $x \in S$).

Необходимость

Предположим, что $f(x)$ выпукла на X. Возьмем 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 X следует, что $x_\lambda \in X$. Более того, так как $f(x)$ – выпуклая функция, то

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

Из неравенства выше по определению надграфика следует, что $\begin{bmatrix} x_\lambda \\ \mu_\lambda \end{bmatrix} \in \text{epi}f$. Следовательно, надграфик f – выпуклое множество.

Достаточность

Предположим, что надграфик f, epi$f$, выпуклое множество. Тогда, исходя из того что $[x_1, \mu_1] \in \text{epi}f$ и $[x_2, \mu_2] \in \text{epi}f$, получаем

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

для любого $0 \leq \lambda \leq 1$.

Следовательно, из определения надграфика, подставив значение μ_λ, получаем, что $f(x_\lambda) \leq \mu_\lambda = \lambda\mu_1 + (1-\lambda)\mu_2$.

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

Но это верно для всех $\mu_1 \geq f(x_1)$ и $\mu_2 \geq f(x_2)$, в том числе и при $\mu_1 = f(x_1)$ и $\mu_2 = f(x_2)$. Тогда мы получаем неравенство:

$$f(x_\lambda) \leq \lambda f(x_1) + (1-\lambda)f(x_2)$$

Так как $x_1 \in X$ и $x_2 \in X$ выбирались произвольно, $f(x)$ - выпуклая функция на X.

6. Дифференциальный критерий сильной выпуклости первого порядка.

> 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 $x, x_0 \in X$.

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,

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

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

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

Таким образом, переходя к пределу при $\lambda \to 0$, мы приходим к первоначальному утверждению.

Достаточность

Предположим, что неравенство в теореме выполнено 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 из условия теоремы, выполняются следующие неравенства:

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

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

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 что $\lambda(1-\lambda)^2 + \lambda^2(1-\lambda) = \lambda(1-\lambda)$, получаем:

$$\lambda f(x_1) + (1-\lambda) f(x_2) - f(x_0) - \frac{\mu}{2}\lambda(1-\lambda)\|x_1 - x_2\|^2 \geq$$

$$\geq \langle \nabla f(x_0), \lambda x_1 + (1-\lambda)x_2 - x_0 \rangle = 0.$$

Таким образом, неравенство из определения сильно выпуклой функции выполнено. Важно отметить, что при $\mu = 0$ получаем случай выпуклости и соответствующий дифференциальный критерий.

7. Дифференциальный критерий сильной выпуклости второго порядка.

> Пусть $X \subseteq \mathbb{R}^n$ — выпуклое множество с непустой внутренностью. Пусть также $f(x)$ — 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 $y \in \mathbb{R}^n$.
> Другая форма записи:
> $$\nabla^2 f(x) \succcurlyeq \mu I$$

Целевое неравенство тривиально, когда $y = 0_n$, поэтому предположим, что $y \neq 0_n$.

Необходимость

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

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

Основываясь на критерии первого порядка сильной выпуклости, имеем

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

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

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.

Достаточность

Формула Тейлора с остаточным членом Лагранжа второго порядка $\forall x, y : x, x + y \in X$ найдется α такая, что:

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

где $0 < \alpha < 1$.

Используя формулу Тейлора с остаточным членом Лагранжа и неравенство из условия, 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,

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

Таким образом, по критерию первого порядка сильной выпуклости, функция $f(x)$ является сильно выпуклой с константой μ. Важно отметить, что $\mu = 0$ соответствует случаю выпуклости и соответствующему дифференциальному критерию.

8. Теорема о построении сопряженного множества к многогранному множеству.

> Пусть $x_1, \ldots, x_m \in \mathbb{R}^n$. Сопряжённое к многогранному множеству:
>
> $$S = \mathbf{conv}(x_1, \ldots, x_k) + \mathbf{cone}(x_{k+1}, \ldots, x_m)$$
>
> будет многогранным множеством:
>
> $$S^* = \left\{p \in \mathbb{R}^n \mid \langle p, x_i \rangle \geq -1, i = \overline{1,k}; \langle p, x_i \rangle \geq 0, i = \overline{k+1,m}\right\}$$

- Пусть $S = X, S^* = Y$. Возьмём произвольный $p \in X^*$, тогда $\langle p, x_i \rangle \geq -1, i = \overline{1,k}$. В то же время, для любого $\theta > 0, i = \overline{k+1,m}$:

$$\langle p, x_i \rangle \geq -1 \to \langle p, \theta x_i \rangle \geq -1$$

$$\langle p, x_i \rangle \geq -\frac{1}{\theta} \to \langle p, x_i \rangle \geq 0.$$

Таким образом, $p \in Y \to X^* \subset Y$.

- В обратную сторону: пусть $p \in Y$. Для любого $x \in X$:

$$x = \sum_{i=1}^m \theta_i x_i \qquad \sum_{i=1}^k \theta_i = 1, \theta_i \geq 0$$

Тогда:

$$\langle p, x \rangle = \sum_{i=1}^m \theta_i \langle p, x_i \rangle = \sum_{i=1}^k \theta_i \langle p, x_i \rangle + \sum_{i=k+1}^m \theta_i \langle p, x_i \rangle \geq \sum_{i=1}^k \theta_i(-1) + \sum_{i=1}^k \theta_i \cdot 0 = -1.$$

Значит, $p \in X^* \to Y \subset X^*$.

9. Вывод сопряженной функции к норме.

> Пусть $f(x) = \|x\|$. Тогда сопряженной функцией к норме является функция $f^\star(y) = \mathbb{0}_{\|y\|_\star \leq 1}$

Доказательство По определению сопряженной функции:

$$f^*(y) = \sup_x \{\langle y, x \rangle - f(x)\} = \sup_x \{\langle y, x \rangle - \|x\|\}$$

Рассмотрим случай $\|y\|_* > 1$. По определению двойственной нормы,

$$\|y\|_* = \sup_{\|x\| \leq 1} \langle y, x \rangle > 1$$

Это означает, что существует некоторый $x^\dagger$, такой, что $\|x^\dagger\| \leq 1$, но $\langle y, x^\dagger \rangle > 1$. Теперь рассмотрим вектор $\bar{x} = tx^\dagger$, где $t \in \mathbb{R}^+$. Значение сопряженной функции является супремумом, поэтому мы имеем следующее соотношение:

$$\begin{aligned} f^*(y) &\geq \langle y, \bar{x} \rangle - \|\bar{x}\| = \langle y, tx^\dagger \rangle - t\|x^\dagger\| \\ &= t(\langle y, x^\dagger \rangle - \|x^\dagger\|) \to \infty \text{ with } t \to \infty \end{aligned}$$

Таким образом, $\|y\|_* > 1$ не принадлежит $\text{dom } f^*$. Рассмотрим случай $\|y\|_* \leq 1$. Из определения двойственной нормы следует:

$$\langle y, x \rangle \leq \|y\|_* \|x\| \leq \|x\|$$

Равенство выполняется, когда $x = 0$. Поэтому

$$f^*(y) = \sup_x \{\langle y, x \rangle - \|x\|\} = 0$$

10. Вывод субдифференциала нормы.

> Пусть V - конечномерное евклидово пространство, и x_0 V. Пусть $\|\cdot\|$ - произвольная норма в V (не обязательно индуцированная скалярным произведением), и пусть $\|\cdot\|_*$ - соответствующая сопряженная норма. Тогда,
>
> $$\partial\|\cdot\|(x_0) = \begin{cases} B_{\|\cdot\|_*}(0,1), \text{если } x_0 = 0, \\ \{s \in V : \|s\|_* \leq 1; \langle s, x_0 \rangle = \|x_0\|\} = \{s \in V : \|s\|_* = 1; \langle s, x_0 \rangle = \|x_0\|\}, & \text{иначе.} \end{cases}$$
>
> Где $B_{\|\cdot\|_*}(0,1)$ - замкнутый единичный шар, с центром в нуле относительно сопряженной нормы. Другими словами, вектор $s \in V$ с $\|s\|_* = 1$ является субградиентом нормы $\|\cdot\|$ в точке $x_0 \neq 0$ тогда и только тогда, когда неравенство из определения двойственной нормы $\langle s, x_0 \rangle \leq \|x_0\|$ становится равенством.

Пусть $s \in V$. По определению, $s \in \partial\|\cdot\|(x_0)$ тогда и только тогда, когда

$$\langle s, x \rangle - \|x\| \leq \langle s, x_0 \rangle - \|x_0\|, \text{ для всех } x \in V,$$

или эквивалентно,

$$\sup_{x \in V} \{\langle s, x \rangle - \|x\|\} \leq \langle s, x_0 \rangle - \|x_0\|.$$

По определению супремума, последнее эквивалентно

$$\sup_{x \in V} \{\langle s, x \rangle - \|x\|\} = \langle s, x_0 \rangle - \|x_0\|.$$

Важно отметить, что выражение в левой части - это супремум из определения сопряженной функции для нормы, которая, как известно, равна

$$\sup_{x \in V}\{\langle s, x\rangle - \|x\|\} = \begin{cases} 0, & \text{если } \|s\|_* \le 1, \\ +\infty, & \text{иначе} \end{cases}$$

Таким образом, уравнение эквивалентно $\|s\|_* \le 1$ и $\langle s, x_0\rangle = \|x_0\|$.

Следовательно, остается заметить, что для $x_0 \neq 0$ неравенство $\|s\|_* \le 1$ должно превратиться в равенство, так как, когда $\|s\|_* < 1$, определение двойственной нормы подразумевает $\langle s, x_0\rangle \le \|s\|_* \|x_0\| < \|x_0\|$.

11. Связь субградиента сопряженной функции и субградиента функции.

> Если f замкнута и выпукла, то $f^{**} = f$. Также,
>
> $$x \in \partial f^*(y) \Leftrightarrow y \in \partial f(x) \Leftrightarrow x \in \arg\min_z \left[f(z) - y^T z\right]$$
>
> Кроме того, если f строго выпуклая, то
>
> $$\nabla f^*(y) = \arg\min_z \left[f(z) - y^T z\right]$$

Мы покажем, что $x \in \partial f^*(y) \Leftrightarrow y \in \partial f(x)$, предполагая, что f выпуклая и замкнутая.

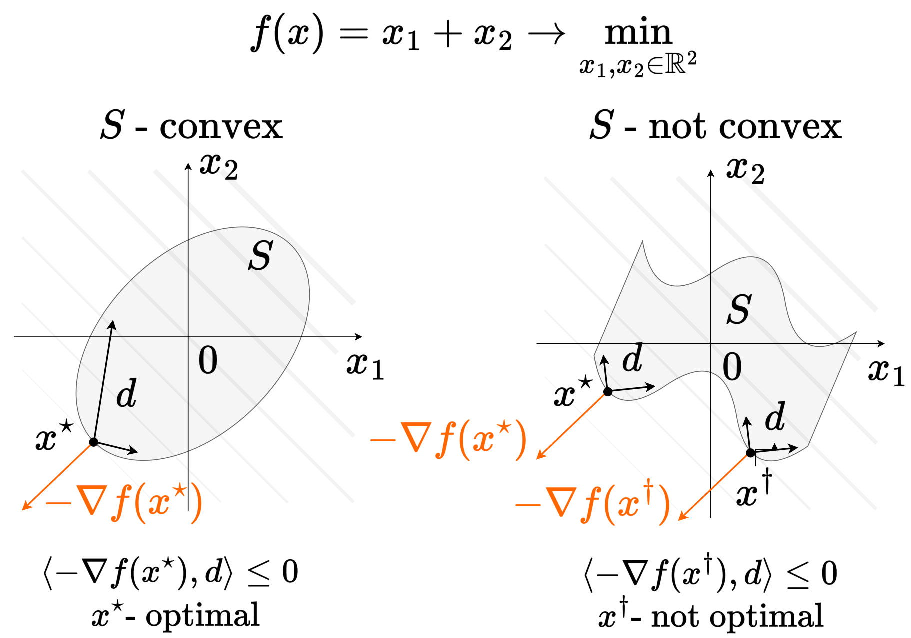
- **Доказательство** $\Leftarrow$: Предположим, что $y \in \partial f(x)$. Тогда $x \in M_y$, множество максимизаторов $y^T z - f(z)$ над z. Но

  $$f^*(y) = \max_z\{y^T z - f(z)\} \quad \text{и} \quad \partial f^*(y) = \mathrm{cl}(\mathrm{conv}(\bigcup_{z \in M_y} \{z\})).$$

  Таким образом, $x \in \partial f^*(y)$\$.

- **Доказательство** $\Rightarrow$: Из того, что мы показали выше, если $x \in \partial f^*(y)$, то $y \in \partial f^*(x)$, но $f^{**} = f$.
- Ясно, что $y \in \partial f(x) \Leftrightarrow x \in \arg\min_z\{f(z) - y^T z\}$.
- Наконец, если f строго выпуклая, то мы знаем, что $f(z) - y^T z$ имеет единственный минимизатор над z, и это должен быть $\nabla f^*(y)$.

12. Субдифференциальное условие оптимальности для условных выпуклых задач.



Пусть $f : \mathbb{R}^n \to \mathbb{R}$ — выпуклая функция, а $S \subset \mathbb{R}^n$ — некоторое (выпуклое) множество допустимых точек. Рассмотрим задачу оптимизации

$$\min_{x \in S} f(x).$$

Тогда точка x^* является решением этой задачи тогда и только тогда, когда

$$0 \in \partial f(x^*) + \mathcal{N}_S(x^*),$$

где $\partial f(x^*)$ — субдифференциал функции f в точке x^*, а $\mathcal{N}_S(x^*)$ — нормальный конус к множеству S в точке x^*.
Если же f дополнительно дифференцируема, то условие оптимальности принимает вид

$$\nabla f(x^*)^T(y - x^*) \geq 0 \quad \text{для всех } y \in S.$$

Рассмотрим задачу

$$\min_{x \in S} f(x).$$

1. **Переход к неограниченной задаче:**
Введём *индикаторную функцию* множества S, то есть

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

Тогда исходная задача эквивалентна безусловной (без явных ограничений) задаче

$$\min_x \Big(f(x) + I_S(x) \Big).$$

2. **Условие оптимальности через субградиент:**
Из общего субдифференциального условия оптимальности следует, что точка x является решением $\min_x\{f(x) + I_S(x)\}$ тогда и только тогда, когда

$$0 \in \partial\big(f(x) + I_S(x)\big).$$

3. **Свойство субградиента суммы:**
   Поскольку f выпуклая и I_S — тоже выпуклая (это индикаторная функция выпуклого множества S), имеем
   $$\partial(f(x) + I_S(x)) = \partial f(x) + \partial I_S(x).$$
   Но $\partial I_S(x) = \mathcal{N}_S(x)$, то есть *нормальный конус* к множеству S в точке x. Следовательно,
   $$0 \in \partial f(x) + \partial I_S(x) \quad \Longleftrightarrow \quad 0 \in \partial f(x) + \mathcal{N}_S(x).$$

4. **Интерпретация условия $0 \in \partial f(x) + \mathcal{N}_S(x)$:**
   Это означает, что существует субградиент $g \in \partial f(x)$ такой, что
   $$-g \in \mathcal{N}_S(x).$$

5. **Частный случай дифференцируемой функции f:**
   Если f дифференцируема, то $\partial f(x) = \{\nabla f(x)\}$. Условие
   $$0 \in \{\nabla f(x)\} + \mathcal{N}_S(x) \quad \Longleftrightarrow \quad -\nabla f(x) \in \mathcal{N}_S(x).$$
   По определению нормального конуса,
   $$-\nabla f(x) \in \mathcal{N}_S(x) \quad \Longleftrightarrow \quad -\nabla f(x)^T x \geq -\nabla f(x)^T y \ \text{ для всех } y \in S,$$
   что переписывается как
   $$\nabla f(x)^T (y - x) \geq 0 \ \text{ для всех } y \in S.$$
   Это и есть классическое *первое условие оптимальности* для дифференцируемых выпуклых задач оптимизации с ограничениями.

13. Необходимые условия безусловного экстремума.

> Если в x^* достигается локальный минимум и f непрерывно дифференцируема в открытой окрестности, то
> $$\nabla f(x^*) = 0$$

Предположим обратное. Пусть $\nabla f(x^*) \neq 0$. Рассмотрим вектор $p = -\nabla f(x^*)$ и заметим, что
$$p^T \nabla f(x^*) = -\|\nabla f(x^*)\|^2 < 0$$
Так как ∇f непрерывна в окрестности x^*, то существует скаляр $T > 0$ такой, что
$$p^T \nabla f(x^* + tp) < 0, \ \text{для любого } t \in [0, T]$$
Для любого $\bar{t} \in (0, T]$, мы можем воспользоваться теоремой Тейлора:
$$f(x^* + \bar{t}p) = f(x^*) + \bar{t}p^T \nabla f(x^* + tp), \ \text{для некоторого } t \in (0, \bar{t})$$
Следовательно, $f(x^* + \bar{t}p) < f(x^*)$ для любого $\bar{t} \in (0, T]$. Мы нашли направление, идя вдоль которого из x^* функция f убывает. Тогда x^* – не точка локального минимума. Получили противоречие.

14. Достаточные условия безусловного экстремума.

> Пусть $\nabla^2 f$ непрерывна в открытой окрестности x^* и
>
> $$\nabla f(x^*) = 0 \quad \nabla^2 f(x^*) \succ 0.$$
>
> Тогда x^* – точка локального минимума f.

Так как гессиан непрерывен и положительно определен в x^*, то мы можем выбрать радиус $r > 0$ такой, что $\nabla^2 f(x)$ остается положительно определенной для всех x в открытом шаре $B = \{z \mid \|z - x^*\| < r\}$. Взяв любой ненулевой вектор p, для которого выполняется $\|p\| < r$, мы получаем $x^* + p \in B$, а также по формуле Тейлора:

$$\begin{aligned} f(x^* + p) &= f(x^*) + p^T \nabla f(x^*) + \frac{1}{2} p^T \nabla^2 f(z) p \\ &= f(x^*) + \frac{1}{2} p^T \nabla^2 f(z) p \end{aligned}$$

где $z = x^* + tp$ для некоторого $t \in (0, 1)$. Так как $z \in B$, мы получаем $p^T \nabla^2 f(z) p > 0$, и следовательно $f(x^* + p) > f(x^*)$. Таким образом x^* – точка локального минимума.

15. Субдифференциальная форма теоремы Каруша Куна Таккера (доказательство). Необходимые условия ККТ для произвольной задачи математического программирования (только формулировка).

> Пусть X - линейное нормированное пространство, а $f_j : X \to \mathbb{R}$, $j = 0, 1, \ldots, m$, - выпуклые собственные (никогда не принимающие значения $-\infty$, а также не тождественно равные ∞) функции. Рассмотрим задачу
>
> $$\begin{aligned} f_0(x) &\to \min_{x \in X} \\ \text{s.t.} f_j(x) &\leq 0, \; j = 1, \ldots, m \end{aligned}$$
>
> Пусть $x^* \in X$ - минимум в задаче выше, а функции f_j, $j = 0, 1, \ldots, m$, непрерывны в точке x^*. Тогда существуют числа $\lambda_j \geq 0$, $j = 0, 1, \ldots, m$, такие, что
>
> $$\sum_{j=0}^{m} \lambda_j = 1,$$
>
> $$\lambda_j f_j(x^*) = 0, \quad j = 1, \ldots, m,$$
>
> $$0 \in \sum_{j=0}^{m} \lambda_j \partial f_j(x^*).$$

Доказательство.

1. Рассмотрим функцию

$$f(x) = \max\{f_0(x) - f_0(x^*), f_1(x), \ldots, f_m(x)\}.$$

Точка x^* является глобальным минимумом этой функции. Действительно, если бы в некоторой точке $x_e \in X$ выполнялось неравенство $f(x_e) < 0$, то из этого следовало бы, что $f_0(x_e) < f_0(x^*)$ и $f_j(x_e) < 0$, $j = 1, \ldots, m$, что противоречит минимальности x^* в задаче выше.

2. Тогда из теоремы Ферма в субдифференциальной форме следует, что

$$0 \in \partial f(x^*).$$

3. По теореме Дубовицкого-Милютина имеем

$$\partial f(x^*) = \text{conv}\left(\bigcup_{j \in I} \partial f_j(x^*)\right),$$

4. Поэтому существует $g_j \in \partial f_j(x^*)$, $j \in I$, такой, что

$$\sum_{j \in I} \lambda_j g_j = 0, \quad \sum_{j \in I} \lambda_j = 1, \quad \lambda_j \geq 0, \quad j \in I.$$

Осталось задать $\lambda_j = 0$ для $j \notin I$.

Для задачи математического программирования в общем виде

$$
\begin{aligned}
f_0(x) &\to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } f_i(x) &\leq 0, \; i = 1, \dots, m \\
h_i(x) &= 0, \; i = 1, \dots, p
\end{aligned}
$$

можно сформулировать лагранжиан:

$$
L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)
$$

Пусть x^*, (λ^*, ν^*) - решение задачи математического программирования с нулевым зазором двойственности (оптимальное значение для прямой задачи p^* равно оптимальному значению для двойственной задачи d^*). Пусть также функции f_0, f_i, h_i дифференцируемы.

- $\nabla_x L(x^*, \lambda^*, \nu^*) = 0$
- $\nabla_\nu L(x^*, \lambda^*, \nu^*) = 0$
- $\lambda_i^* \geq 0, i = 1, \dots, m$
- $\lambda_i^* f_i(x^*) = 0, i = 1, \dots, m$
- $f_i(x^*) \leq 0, i = 1, \dots, m$

Некоторые условия регулярности

Эти условия необходимы для того, чтобы сделать условия ККТ необходимыми. Некоторые из них даже превращают необходимые условия в достаточные (например, условия Слейтера). Более того, если у вас есть регулярность, вы можете записать необходимые условия второго порядка $\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle \geq 0$ с *полуопределенным* гессианом лагранжиана.

- **Условие Слейтера.** Если для выпуклой задачи (т.е. в предположении минимизации, f_0, f_i выпуклы и h_i аффинны) существует точка x такая, что $h(x) = 0$ и $f_i(x) < 0$ (существование строго внутренней точки бюджетного множества), то мы имеем нулевой зазор двойственности и условия ККТ становятся необходимыми и достаточными.
- **Условие линейности ограничений.** Если f_i и h_i - аффинные функции, то других условий не требуется.
- **Условие линейной независимости ограничений.** Градиенты активных ограничений неравенств и градиенты ограничений равенств линейно независимы в точке x^*.

- Другие примеры см. в wiki.

16. Формулировка симплекс метода для задачи линейного программирования в стандартной форме. Теорема о проверке оптимальности решения.

Если все элементы λ_B неположительны и базис B допустимый, тогда базис B оптимален. Здесь λ_B это коэффициенты при разложении c по базису B: $\lambda_B^T A_B = c^T \Rightarrow \lambda_B^T = c^T A_B^{-1}$.

Формулировка симплекс метода для задачи линейного программирования в стандартной форме. Теорема о проверке оптимальности решения

Задача линейного программирования:

Пусть $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$, тогда задача формулируется так:

$$
\begin{aligned}
\min_{x \in \mathbb{R}^n} \quad & c^T x \\
\text{s.t.} \quad & Ax \leq b
\end{aligned}
$$

Идейное описание симплекс метода:

1. Убедится, что точка, в которой мы находимся, является угловой
2. Проверить оптимальность точки
3. Если необходимо, сменить угол (то есть сменить базис)
4. Повторять до схождения

Шаги выполнения симплекс метода:

1. **Поиск начальной базисной допустимой точки:**
   - Выберем начальную базисную (она является решением системы $A_B x = b_B$, где B - базис размера n пространства, а матрица A обычно имеет больше n ограничений) допустимую ($Ax_0 \leq b$) точку x_0 (искать ее будем через двухфазный симплексметод). Если такая точка не найдена, задача не имеет допустимого решения.
2. **Проверка оптимальности:**
   - Разложение вектора c в данном базисе B с коэффициентами λ_B:
   $$\lambda_B^\top A_B = c^\top \quad \text{или} \quad \lambda_B^\top = c^\top A_B^{-1}$$
   - Если все компоненты λ_B неположительны, текущий базис является оптимальным. Иначе далее меняем вершину симплекса.
3. **Определение переменной для удаления из базиса:**
   - Если в разложении λ_B есть положительные координаты, продолжаем оптимизацию. Пусть $\lambda_B^k > 0$. Необходимо исключить k из базиса. Рассчитаем направляющий вектор d, идя вдоль которого изменим вершину следующим образом: во-первых, для векторов всех ограничений из базиса, которые мы оставляем, направление должно быть им ортогонально, и, во-вторых, вдоль него значение, связанное с нашим ограничением, должно убывать:
   $$\begin{cases} A_{B \setminus \{k\}} d = 0 \\ a_k^\top d < 0 \end{cases}$$
4. **Вычисление шага вдоль выбранного направления d:**
   - Для всех $j \notin B$ считаем шаг:
   $$\mu_j = \frac{b_j - a_j^\top x_B}{a_j^\top d}$$
   - Новая вершина, которую добавим в базис:
   $$t = \arg\min_j \{\mu_j \mid \mu_j > 0\}$$
5. **Обновление базиса:**

- Обновляем базис и текущее решение:

$$B' = B \backslash \{k\} \cup \{t\},$$
$$x_{B'} = x_B + \mu_t d = A_{B'}^{-1} b_{B'}$$

- Изменение базиса приводит к уменьшению значения целевой функции:

$$c^\top x_{B'} = c^\top (x_B + \mu_t d) = c^\top x_B + \mu_t c^\top d$$

6. **Повторение:**
   - Далее повторяем шаги 2-5 до достижения оптимального решения или установления, что задача не имеет допустимого решения.

Теорема о проверке оптимальности решения:

Если все элементы λ_B неположительны и базис B достижим, тогда базис B оптимален.

Здесь λ_B это коэффициенты при разложении c по базису B: $\lambda_B^T A_B = c^T \Rightarrow \lambda_B^T = c^T A_B^{-1}$.

Доказательство:

Предположим противное (что этот базис не оптимален), пусть $\exists x^* : \; Ax^* \leq b$ и при этом $c^T x^* < c^T x_B$. Так как для всей матрицы A и вектора b неравенство верно, то и для подматрицы оно верно:

$$A_B x^* \leq b_B$$

Так как все элементы λ_B неположительны, то домножим строки на соответствующие элементы и сложим:

$$\lambda_B^T A_B x^* \geq \lambda_B^T b_B$$

$$c^T x^* \geq \lambda_B^T b_B = \lambda_B^T A_B x_B = c^T x_B$$

Противоречие.

17. Доказательство работы теста корней

> Пусть $(r_k)_{k=m}^{\infty}$ последовательность неотрицательных чисел, сходящихся к нулю, и пусть $\alpha := \limsup_{k \to \infty} r_k^{1/k}$. (причем $\alpha \geq 0$.)
>
> (a) Если $0 \leq \alpha < 1$, то $(r_k)_{k=m}^{\infty}$ сходится линейно с константой α.
>
> (b) В частности, если $\alpha = 0$, то $(r_k)_{k=m}^{\infty}$ сходится сверхлинейно.
>
> (c) Если $\alpha = 1$, то $(r_k)_{k=m}^{\infty}$ сходится сублинейно.
>
> (d) Случай $\alpha > 1$ невозможен.

Доказательство.

1. Покажем, что если $(r_k)_{k=m}^\infty$ сходится линейно с константой $0 \leq \beta < 1$, то обязательно $\alpha \leq \beta$.

   Действительно, по определению константы линейной сходимости для любого $\varepsilon > 0$, удовлетворяющего условию $\beta + \varepsilon < 1$, существует $C > 0$, такое что $r_k \leq C(\beta + \varepsilon)^k$ для всех $k \geq m$.

   Отсюда $r_k^{1/k} \leq C^{1/k}(\beta + \varepsilon)$ для всех $k \geq m$. Переходя к пределу при $k \to \infty$ и используя $C^{1/k} \to 1$, получаем $\alpha \leq \beta + \varepsilon$. Учитывая произвольность ε, следует, что $\alpha \leq \beta$.

2. Таким образом, в случае $\alpha = 1$ последовательность $(r_k)_{k=m}^\infty$ не может иметь линейную сходимость согласно вышеуказанному результату (доказанному от противного). Поскольку, тем не менее, $(r_k)_{k=m}^\infty$ сходится к нулю, она должна сходиться сублинейно.

3. Теперь рассмотрим случай $0 \leq \alpha < 1$. Пусть $\varepsilon > 0$ — произвольное число, такое что $\alpha + \varepsilon < 1$.

Согласно свойствам lim sup, существует $N \geq m$, такое что $r_k^{1/k} \leq \alpha + \varepsilon$ для всех $k \geq N$.

Следовательно, $r_k \leq (\alpha + \varepsilon)^k$ для всех $k \geq N$. Поэтому последовательность $(r_k)_{k=m}^\infty$ сходится линейно с параметром $\alpha + \varepsilon$ (не имеет значения, что неравенство выполняется только начиная с числа N).

Из произвольности ε следует, что константа линейной сходимости последовательности $(r_k)_{k=m}^\infty$ не превышает α.

Поскольку, как показано выше, константа линейной сходимости не может быть меньше α, это означает, что константа линейной сходимости последовательности $(r_k)_{k=m}^\infty$ равна именно α.

4. Покажем, что случай $\alpha > 1$ невозможен.

   Действительно, предположим, что $\alpha > 1$. Тогда из определения lim sup следует, что для любого $N \geq m$ существует $k \geq N$, такое что $r_k^{1/k} \geq 1$, и, в частности, $r_k \geq 1$.

   Но это означает, что у r_k есть подпоследовательность, ограниченная от нуля. Следовательно, последовательность $(r_k)_{k=m}^\infty$ не может сходиться к нулю, что противоречит условию. \end{enumerate}

18. Метод дихотомии и золотого сечения для унимодальных функций. Скорость сходимости.

> ℹ Методы локализации решения для скалярной минимизации. Сходятся линейно.

Метод дихотомии

Решаем следующую задачу:

$$f(x) \to \min_{x \in [a,b]}$$

где $f(x)$ — унимодальная функция.

Мы хотим на каждом шаге вдвое сокращать область, в которой ищем минимум. Для этого будем пользоваться основным свойством унимодальных функций:

$$\forall a \leq x_1 < x_2 \leq b :$$

$$f(x_1) \leq f(x_2) \Rightarrow x_* \in [a, x_2]$$

$$f(x_1) \geq f(x_2) \Rightarrow x_* \in [x_1, b]$$

где x_* — точка, в которой достигается минимум

Алгоритм:

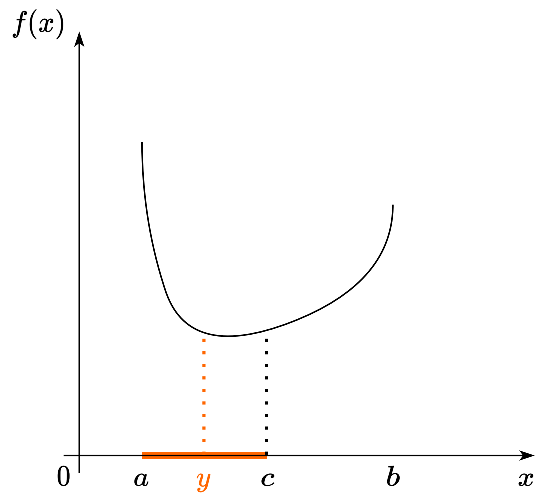


Figure 8: Алгоритм дихотомии

Можно заметить, что на каждой итерации требуется не более 2-х вычислений значения функции.

Сходимость метода дихотомии

Длина отрезка на $k+1$ итерации:

$$\Delta_{k+1} = b_{k+1} - a_{k+1} = \frac{1}{2^k}(b-a)$$

Если будем выбирать середину отрезка как выход $k+1$ итерации:

$$|x_{k+1} - x_*| \leq \frac{\Delta_{k+1}}{2}$$

Подставим полученное ранее выражение для длины отрезка:

$$|x_{k+1} - x_*| \leq \frac{1}{2^{k+1}}(b-a)$$

$$|x_{k+1} - x_*| \leq (0.5)^{k+1}(b-a)$$

Получили выражение для сходимости по итерациям. Отсюда также можно выразить необходимое количество итераций для достижения точности ε:

$$K = \left\lceil \log_2 \frac{b-a}{\varepsilon} - 1 \right\rceil$$

Теперь получим выражение для сходимости по количеству вычислений значения функции. Знаем, что на каждой итерации вычисляем значение не более 2-х раз, значит количество вычислений значения функции возьмём $N = 2k$:

$$|x_{k+1} - x_*| \leq (0.5)^{\frac{N}{2}+1}(b-a)$$

$$|x_{k+1} - x_*| \leq (0.707)^N \frac{b-a}{2}$$

Метод золотого сечения

Идея такая же, как и в методе дихотомии, но хотим уменьшить количество вычислений значения функции. Для этого будем вычислять значения в точках золотого сечения. Так на каждой итерации нам нужно будет вычислять значение только в одной точке, так как для нового отрезка в одной из точек золотого сечения значение будет уже посчитано:

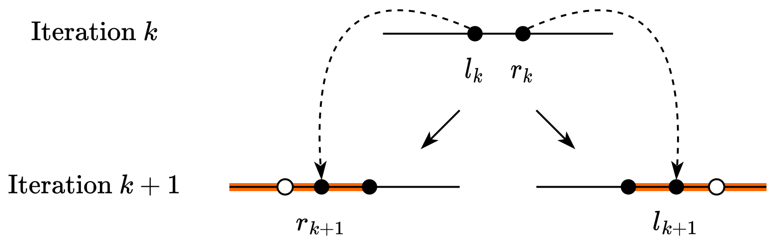


Figure 9: Золотое сечение

Алгоритм:

```
def golden_search(f, a, b, epsilon):
    tau = (sqrt(5) + 1) / 2
    y = a + (b - a) / tau**2
    z = a + (b - a) / tau
    while b - a > epsilon:
        if f(y) <= f(z):
            b = z
            z = y
            y = a + (b - a) / tau**2
        else:
            a = y
            y = z
            z = a + (b - a) / tau
    return (a + b) / 2
```

Сходимость метода золотого сечения

На каждой итерации длина отрезка будет уменьшаться в $\tau = \frac{\sqrt{5}+1}{2}$ раз. Тогда оценка сходимости (и по итерациям, и по вычислениям значений функции):

$$|x_{k+1} - x_*| \leq \frac{b_{k+1} - a_{k+1}}{2} = \left(\frac{1}{\tau}\right)^{N-1} \frac{b-a}{2} \approx 0.618^k \frac{b-a}{2}$$

Получили сходимость по итерациям хуже, чем у дихотомии, так как отрезки уменьшаются слабее на каждой итерации. Но по количеству вычислений значения функции сходимость у метода золотого сечения быстрее.