Лекция 1. 10 января 2020 г.

Подготовила: Иванова А.

Ключевые слова: целевая функция, допустимое множество, оракул, аналитическая сложность, выпуклое множество, выпуклая функция

Общая формулировка задачи оптимизации

Обозначим через x вещественный вектор размерности n: $x = \left(x^{(1)}, \ldots, x^{(n)}\right)^T \in \mathbb{R}^n$. Пусть S – некоторое множество из пространства $\mathbb{R}^n$, а $f_0(x), \ldots, f_m(x)$ вещественнозначные функции от x. Тогда задачу оптимизации (минимизации) можно записать в следующем общем виде:

$$\begin{aligned} & \min f_0(x) \\ s.t.\quad & f_j(x) \,\&\, 0, \quad j = 1, \ldots, m, \\ & x \in S, \end{aligned} \tag{1}$$

где в качестве & берется $\leq, \geq$ либо $=$.

При этом в дальнейшем будем называть:

- $x = (x^{(1)}, \ldots, x^{(n)}) \in \mathbb{R}^n$ – оптимизируемые параметры (переменные);
- $f_0 : \mathbb{R}^n \to \mathbb{R}$ – целевая функция;
- $f(x) = (f_1(x), \ldots, f_m(x))^T$ – вектор функциональных ограничений;
- S базовое допустимое множество;
- $Q = \{x \in S : f_j(x) \leq 0, \, j = 1, \ldots, m\}$ – допустимое множество.

Задача: найти решение x^*, которое доставляет наименьшее значение целевой функции $f_0(x)$ среди всех x из допустимого множества.

Приведем некоторые типы задач минимизации:

- Условные задачи: $Q \subset \mathbb{R}^n$.
- Безусловные задачи: $Q \equiv \mathbb{R}^n$.
- Гладкие задачи: все функции f_j дифференцируемы.
- Негладкие задачи: существует по крайней мере одна недифференцируемая компонента $f_k(x)$.
- Задачи с линейными ограничениями : все функциональные ограничения являются линейными функциями: $f_j(x) \equiv \langle a_j, x \rangle + b_j, \quad j = 1, \ldots, m$.

Классификация задач основанная на свойствах допустимого множества:

- Допустимая: $Q \neq \emptyset$.
- Строго допустимая: если существует такой вектор $x \in \mathrm{int} Q$, что $f_j(x) < 0$ для всех ограничений-неравенств и $f_j(x) = 0$ для всех ограничений-равенств (условие Слейтера).

Классификация различных типов решения задачи:

- Точка x^* называется оптимальным глобальным решением задачи, если $f_0(x^*) \leq f_0(x)$ для всех $x \in Q$ (глобальный минимум). В этом случае $f_0(x^*)$ называется (глобальным) оптимальным значением задачи.
- Точка x^* называется оптимальным локальным решением задачи, если для всех $x \in \text{int}\bar{Q} \subset Q$ выполнено неравенство $f_0(x^*) \leq f_0(x)$ (локальный минимум).

Эффективность численных методов

Мы будем говорить об эффективности метода $\mathcal{M}$ для какого-то класса задач $\mathcal{P}$ (однотипные задачи с близкими характеристиками). Поэтому мы предполагаем, что наш метод с самого начала не имеет полной информации о решаемой задачи. При этом заранее известная методу "часть"задачи $\mathcal{P}$ называется моделью Σ решаемой задачи. Обычно в модель включают формулировку задачи, описание свойств функциональных компонент.

При этом численный метод должен накапливать информацию о решаемой задачи. Для этого введем понятие оракула $\mathcal{O}$. Оракул можно представить в виде некоторого устройства, которое отвечает на последовательные вопросы численного метода. При этом метод $\mathcal{M}$ пытается решить задачу $\mathcal{P}$, собирая и анализируя ответы оракула.

Тогда, эффективность метода определяется его эффективностью на наихудшем представителе $\mathcal{P}_{\sqsupseteq}$ при условии, что модель и оракул зафиксированы. Стоит отметить, что определенная задача может быть наихудшей только для данного конкретного метода $\mathcal{M}$.

Отметим, что для большинства численных задач найти их точное решение невозможно. Поэтому в качестве решения задачи мы будем понимать следующее: решить задачу $\mathcal{P}$ означает найти ее приближенное решение с заранее заданной точностью $\varepsilon > 0$. Поэтому определим $\mathcal{J}_\varepsilon$ для некоторого критерия останова, способное оценить качество предлагаемого решения.

Теперь мы можем определить класс решаемых задач следующим образом $\mathcal{F} \equiv (\Sigma, \mathcal{O}, \mathcal{J}_\varepsilon)$.

Любой метод $\mathcal{M}$ можно представить в виде итеративной процедуры в следующем виде:

Общая итеративная схема

Вводные данные: x_0 – начальная точка, $\varepsilon > 0$ – требуемая точность.
Настройка: Полагаем $k = 0$ (счетчик итераций) и $I_{-1} = \emptyset$ (накапливаемая информационная модель решаемой задачи).
Основной цикл:

1. Задаем вопрос оракулу $\mathcal{O}$ в точке x_k.
2. Пересчитываем информационную модель:
$$I_k = I_{k+1} \cup (x_k, \mathcal{O}(x_k)).$$
3. Применяем правила метода $\mathcal{M}$ для анализа модели I_k и формируем точку x_{k+1}.
4. Проверяем критерий остановки $\mathcal{J}_\varepsilon$. Если ответ положительный, то генерируем ответ $\bar{x}$. В противном случае полагаем $k := k + 1$ и переходим на шаг 1.

Введем понятие сложности для представленной выше схемы.

Аналитическая сложность. Число обращений к оракулу, необходимое для решения задачи $\mathcal{P}$ с точностью ε.

Арифметическая сложность. Это общее число всех вычислений (включая как работу оракула, так и работу метода), необходимых для решения задачи $\mathcal{P}$ с точностью ε.

В большинстве случаев вторую оценку можно получить из первой, поэтому мы в основном будем говорить именно об аналитической сложности класса задач.

Так же введем еще одно предположение необходимое для для получения большинства результатов теории сложности задач оптимизации. Это предположение **концепции черного ящика**:

1. Единственной информацией, получаемой в ходе работы итеративного метода, являются ответы оракула.

2. Ответы оракула являются локальными: небольшое изменение задачи, произведенное достаточно далеко от тестовой точки x и согласованное с описанием данного класса задач, не обязано привести к изменению исходного ответа в точке x.

Отметим, что (1) является функциональной моделью. Обычно для такой модели стандартные предположения связаны с гладкостью функциональных компонент. И в соответствии со степенью гладкости можно пользоваться разными типами оракулов:

- Оракул нулевого порядка: возвращает значение функции $f(x)$.
- Оракул первого порядка: возвращает значение функции $f(x)$ и ее градиент $f'(x)$.
- Оракул второго порядка: возвращает значение функции $f(x)$, ее градиент $f'(x)$ и матрицу гессиана $f''(x)$.

Оценки вычислительной сложности задач глобальной оптимизации

Рассмотрим следующую задачу:

$$\min_{x \in B_n} f(x). \tag{2}$$

Данная задача является задачей условной оптимизации без функциональных ограничений. Допустимым множеством является n-мерный куб B_n в пространстве R^n:

$$B_n = \left\{ x \in R^n \,|\, 0 \le x^{(i)} \le 1,\ i = 1, \dots, n \right\}.$$

Введем l_∞-норму в $\mathbb{R}^n$: $||x||_\infty = \max\limits_{1 \le i \le n} |x^{(i)}|$. Предположим, что целевая функция $f(x)$ *липшицева* относительно этой нормы на B_n с константой Липшица L, то есть

$$|f(x) - f(y)| \le L||x - y||_\infty, \quad x,\, y \in B_n. \tag{3}$$

Для решения данной задачи рассмотрим следующий простейший метод

Метод перебора $G(p)$.

Вводные данные: $p \ge 1$ – целочисленный параметр.

1. Формируем $(p+1)^n$
$$x_{(i_1,\dots,i_n)} = \left(\tfrac{i_1}{p}, \tfrac{i_2}{p}, \dots, \tfrac{i_n}{p} \right)^T$$
где $(i_1, \dots, i_n) \in \{0, \dots, p\}^n$.
2. Среди всех точек $x_{(i_1,\dots,i_n)}$ находим точку $\bar{x}$ с наименьшим значением целевой функции.
3. Представим пару $(\bar{x}, f(\bar{x}))$ как результат работы метода.

Таким образом наш метод просто перебирает точки равномерной сетки, сформированной внутри куба. Получаем метод нулевого порядка с отсутствием какого бы то ни было влияния накопленной информации на формирование последовательности пробных точек.

Сформулируем теорему об оценке эффективности этого алгоритма:

Теорема 1. Обозначим через f^* оптимальное значение целевой функции. Тогда

$$f(\bar{x}) - f^* \leq \tfrac{L}{2p}.$$

Доказательство. Пусть точка x^*- глобальное решение задачи. Тогда найдется такой мультииндекс $(i_1, \ldots, i_n)$, что

$$x \equiv x_{(i_1,\ldots,i_n)} \leq x^* \leq x_{(i_1+1,\ldots,i_n+1)} \equiv y$$

(соотношение $x \leq y$ для векторов x, $y \in R^n$ означает, что $x^{(i)} \leq y^{(i)}$ для всех индексов $i = 1, \ldots, n$). Заметим, что $y^{(i)} - x^{(i)} = 1/p$ при всех $i = 1, \ldots, n$ и $x_*^{(i)} \in [x^{(i)}, y^{(i)}]$, $i = 1, \ldots, n$. Пусть $\hat{x} = (x+y)/2$. Зададим координаты точки $\tilde{x}$ следующим образом:

$$\tilde{x}^{(i)} = \begin{cases} y^{(i)}, & \text{если } x_*^{(i)} \geq \hat{x}^{(i)} \\ x^{(i)}, & \text{в противном случае.} \end{cases}$$

Понятно, что $|\tilde{x}^{(i)} - x_*^{(i)}| \leq \frac{1}{2p}$, $i = 1, \ldots, n$. Поэтому

$$||\tilde{x} - x_*||_\infty = \max_{1 \leq i \leq n} |\tilde{x}^{(i)} - x_*^{(i)}| \leq \tfrac{1}{p}.$$

Поскольку точка $\tilde{x}$ принадлежит сформированной сетке, можно утверждать, что

$$f(\bar{x}) - f(x_*) \leq f(\tilde{x}) - f(x_*) \leq L||\tilde{x} - x_*||_\infty \leq \tfrac{L}{2p}.$$

□

Под решением задачи будем иметь в виду такую точку $\bar{x} \in B_n$, что

$$f(\bar{x}) - f^* \leq \varepsilon. \tag{4}$$

Тогда сформулируем следующее утверждение:

Следствие 1. Аналитическая сложность класса задач минимизации (2), (3), (4) для метода G не превосходит

$$\mathcal{A}(G) = \left(\lfloor \tfrac{L}{2\varepsilon} \rfloor + 2\right)^n$$

вызовов оракула.

Доказательство. Выберем $p = \lfloor \frac{L}{2\varepsilon} \rfloor + 1$. Тогда $p \geq \frac{L}{2\varepsilon}$, и в силу Теоремы 1 получаем

$$f(\bar{x}) - f^* \leq \tfrac{L}{2p} \leq \varepsilon.$$

Осталось заметить, что было просмотрено $(p+1)^n$ пробных точек. □

Таким образом величина $\mathcal{A}(G)$ устанавливает верхнюю границу сложности для рассматриваемого класса задач.

Теперь оценим нижнюю границу аналитической сложности. Отметим особенности получения подобных оценок:

- Они основаны на применении концепции черного ящика.
- Полученные оценки верны для всех мыслимых итеративных методов. Таким образом, устанавливается нижняя оценка для аналитической сложности рассматриваемого класса задач.
- Очень часто эти оценки выводятся с помощью сопротивляющегося оракула.

Сопротивляющийся оракул создает наихудшую задачу для каждого конкретного метода. Он начинает свою работу с пустой задачи и начинает отвечать на вопросы метода наихудшем образом (важно, что ответы должны быть согласованы с как с предыдущими ответами так и с характеристиками данного класса задач). Тогда после завершения работы такого оракула возможна реконструкция задачи, которая полностью соответствует информации, собранной тестируемым методом оптимизации.

Покажем на примере работу сопротивляющегося оракула. Рассмотрим класс $\mathcal{C}$ оптимизации:

Модель: $\min\limits_{x \in B_n} f(x)$, где $f(x)$ является l_∞-липщецевой на B_n.

Оракул: черный ящик нулевого порядка.

Приближенное решение: $\bar{x} \in B_n \,:\, f(\bar{x}) - f^* \leq \varepsilon$.

Теорема 2. Пусть $\varepsilon < \frac{1}{2}L$. Тогда аналитическая сложность класса $\mathcal{C}$ составляет по крайней мере $\left(\lfloor \frac{L}{2\varepsilon} \rfloor\right)^n$ вызовов оракула.

Доказательство. Положим $p = \lfloor \frac{L}{2\varepsilon} \rfloor \leq 1$. Пусть существует некоторый метод, которому требуется $N < p^n$ вызовов оракула для того, чтобы решить любую задачу из класса $\mathcal{C}$. Применим для этого метода следующий сопротивляющийся оракул: **сообщается, что** $f(x) = 0$ **в любой тестовой точке** x**.**

В этом случае метод может обнаружить только $\bar{x} \in B_n \,:\, f(\bar{x}) = 0$. Однако заметим, что существует такая точка $\hat{x} \in B_n$, что

$$\hat{x} + \frac{1}{p} e \in B_n, \quad e = (1, \ldots, 1)^T \in \mathbb{R}^n,$$

и при этом не существует ни одной тестовой точки внутри куба

$$B = \left\{ x \,|\, \hat{x} \leq x \leq \hat{x} + \frac{1}{p} e \right\}.$$

Обозначим $x_* = \hat{x} + \frac{1}{2p} e$ и рассмотрим функцию $\bar{f}(x) = \min\{0, L||x - x_*||_\infty - \varepsilon\}$. Ясно, что эта функция является l_∞-липшицевой и что значение ее глобального минимума равно $-\varepsilon$. Более того функция $\bar{f}(x)$ отлична от нуля только внутри куба $B' = \{x \,:\, ||x - x_*||_\infty \leq \varepsilon / L\}$. Поскольку $2p \leq L/\varepsilon$, то

$$B' \subseteq B = \left\{ x \,:\, ||x - \hat{x}||_\infty \leq \frac{1}{2p} \right\}.$$

Таким образом функция $\bar{f}(x)$ равна нулю во всех тестовых точках нашего метода.Так как точность полученного ответа никак не лучше чем ε, мы получаем, что за число обращений к оракулу меньше чем p^n нельзя гарантировать что достигнутая абсолютная точность будет лучше, чем заранее заданное $\varepsilon > 0$. ☐

Получаем, что если $\varepsilon \leq O(L/n)$, то нижние и верхние оценки эффективности совпадают с точностью до мультипликативной абсолютной константы.

Зачем нужна выпуклость?

- Локальный оптимум является глобальным
- Необходимое условие оптимальности является достаточным

Выпуклое множество

Пусть даны две точки x_1, x_2. Тогда отрезок, соединяющий их определяется следующим образом:

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

Определение 1. Пусть U — вещественное векторное пространство, Q — подмножество U. Множество Q называется выпуклым, если для любых двух точек x, y из множества Q и любого $0 \leq \alpha \leq 1$ точка $\alpha x + (1 - \alpha) y$ также принадлежит множеству Q.

Пример 2. Ниже приведены примеры выпуклых множеств:

1. *(Тривиальные выпуклые множества).* Все пространство U , множество из одного элемента $\{a\}$, $a \in U$ и пустое множество $\emptyset$.

2. *(Множество решений системы линейных ограничений).* Пусть в пространстве U задано (произвольное) скалярное произведение $\langle \cdot, \cdot \rangle$. Пусть I — произвольное (не обязательно конечное и не обязательно счетное) индексное множество, и пусть для каждого индекса $i \in I$ заданы вектор $a_i \in U$ и скаляр $b_i \in \mathbb{R}$. Рассмотрим систему линейных неравенств $\langle a_i, x \rangle \leq b_i, \quad \forall\, i \in I$, где $x \in U$. Тогда соответствующее множество всевозможных решений этой системы, т. е. множество

$$Q := \{x \in U \,:\, \langle a_i, x \rangle \leq b_i \,\text{для всех}\, i \in I\}$$

является выпуклым.

3. *(Шар в произвольной норме).* Пусть в пространстве U задана (произвольная) норма $||\cdot||$. Тогда (замкнутый) шар радиуса $r > 0$ с центром в точке $a \in U$, т.е. множество

$$\bar{B}_{||\cdot||}(a, r) := \{x \in U \,:\, ||x - a|| \leq r\}\,,$$

является выпуклым.

4. *(Множество положительно полуопределенных матриц).* В пространстве симметричных матриц $\mathbb{S}^n$ множество положительно полуопределенных матриц $\mathbb{S}^n_+$ является выпуклым.

Операции, сохраняющие выпуклость

Утверждение 1. (Операции, сохраняющие выпуклость множеств).

Следующие операции сохраняют выпуклость множеств.

1. *(Пересечение)* Пересечение любого числа выпуклых множеств выпукло.

   Пусть U — вещественное векторное пространство. Пусть I — произвольное (не обязательно конечное и не обязательно счетное) индексное множество, и пусть для каждого индекса $i \in I$ задано множество Q_i в пространстве U. Если каждое из множеств Q_i для $i \in I$ является выпуклым, тогда и их пересечение

$$\cap_{i \in I} Q_i = \{x \,:\, x \in Q_i \text{ для всех } i \in I\}\,.$$

   также является выпуклым множеством в пространстве U.

   Приведем доказательство для двух множеств. Если $x_1 \in Q_1 \cap Q_2$, $x_2 \in Q_1 \cap Q_2$, то $[x_1, x_2] \subset Q_1$ и $[x_1, x_2] \subset Q_2$, поэтому $[x_1, x_2] \subset Q_1 \cap Q_2$.

2. *(Образ при аффинном преобразовании)* Пусть U и V — вещественные векторные пространства, и Q — множество в пространстве U. Пусть $A \,:\, U \to V$ — аффинное преобразование из пространства U в пространство V , т.е. преобразование вида $A(x) = Lx + a$, где $L \,:\, U \to V$— линейное преобразование и $a \in V$. Если множество Q является выпуклым, тогда и его образ при аффинном преобразовании A, т.е. множество

$$A(Q) := \{A(x) \,:\, x \in Q\}\,,$$

   является выпуклым множеством в пространстве V.

   Если y_1, $y_2 \in A(Q)$, то $y_1 = Lx_1 + a$ и $y_2 = Lx_2 + a$ для некоторых $x_1\, x_2 \in Q$. Поэтому для $y(\alpha) = \alpha y_1 + (1-\alpha) y_2$, $0 \leq \alpha \leq 1$, выполнено соотношение

$$y(\alpha) = \alpha(Lx_1 + a) + (1-\alpha)(Lx_2 + a) = L(\alpha x_1 + (1-\alpha)x_2) + a.$$

   Таким образом, $y(\alpha) \in A(Q)$.

3. *(Прообраз при аффинном преобразовании)* Пусть U и V — вещественные векторные пространства, и S — множество в пространстве U. Пусть $A \,:\, U \to V$ — аффинное преобразование из пространства U в пространство V , т.е. преобразование вида $A(x) = Lx + a$, где $L \,:\, U \to V$— линейное преобразование и $a \in V$. Если множество S является выпуклым, тогда его обратный образ при аффинном преобразовании A, т.е. множество

$$A^{-1}(S) := \{x \,:\, A(x) \in S\}\,,$$

является выпуклым множеством в пространстве U.

Если $x_1,\, x_2 \in A^{-1}(S)$, *то* $x_1 = Ly_1 + a$ *и* $x_2 = Ly_2 + a$ *для некоторых* $y_1\, y_2 \in S$. *Поэтому для* $x(\alpha) = \alpha x_1 + (1-\alpha)x_2$, $0 \le \alpha \le 1$, *выполнено соотношение*

$$A(x(\alpha)) = A(\alpha x_1 + (1-\alpha)x_2) + a = \alpha(Ax_1 + a) + (1-\alpha)(Ax_2 + a) = \alpha y_1 + (1-\alpha)y_2 \in S.$$

4. *(Прямое произведение)* Пусть $U_1, \ldots, U_n$—вещественные векторные пространства,и $Q_1, \ldots, Q_n$ — множества в пространствах $U_1, \ldots, U_n$ соответственно. Если каждое из множеств $Q_1, \ldots, Q_n$ является выпуклым, тогда и их прямое (декартово) произведение

$$Q_1 \times \ldots \times Q_n := \{(x_1, \ldots, x_n) \,:\, x_1 \in Q_1; \ldots; x_n \in Q_n\}$$

является выпуклым множеством в пространстве $U_1 \times \ldots \times U_n$.

Приведем доказательство для двух множеств. Если $z_1 = (x_1, x_2)$, $x_1 \in Q_1$, $x_2 \in Q_2$ *и* $z_2 = (y_1, y_2)$, $y_1 \in Q_1$, $y_2 \in Q_2$, *то*

$$\alpha z_1 + (1-\alpha)z_2 = ((\alpha x_1 + (1-\alpha)y_1)_1, (\alpha x_2 + (1-\alpha)y_2)_2),$$

где $(\cdot)_1 \in Q_1$ *и* $(\cdot)_2 \in Q_2$.

5. *(Линейная комбинация)* Пусть U — вещественное векторное пространство, и $Q_1, \ldots, Q_k$— множества в пространстве U. Пусть также $\alpha_1, \ldots, \alpha_k$ —скаляры (произвольного знака). Если каждое из множеств $Q_1, \ldots, Q_k$ является выпуклым, тогда и их линейная комбинация

$$\alpha_1 Q_1 + \ldots + \alpha_k Q_k = \left\{ \sum_{i=1}^{k} \alpha_i x_1 \,:\, x_i \in Q_i \,\text{для всех}\, 1 \le i \le k \right\}$$

также является выпуклым множеством в пространстве U.

Приведем доказательство для двух множеств. Рассмотрим два выпуклых множества Q_x, Q_y. *Определим множество*

$$Q = \{z \,|\, z = \alpha_1 x + \alpha_2 y,\, x \in Q_x,\, y \in Q_y,\, \alpha_1,\, \alpha_2 \in \mathbb{R}\}\,.$$

Возьмем две точки из этого множества: $z_1 = \alpha_1 x_1 + \alpha_2 y_1$ *и* $z_2 = \alpha_1 x_2 + \alpha_2 y_2$ *и покажем, что отрезок между ними также принадлежит* Q, *т. е. что* $\theta z_1 + (1-\theta)z_2$, $\theta \in [0,1]$ *принадлежит* Q.

$$\theta z_1 + (1-\theta)z_2 = \theta(\alpha_1 x_1 + \alpha_2 y_1) + (1-\theta)(\alpha_1 x_2 + \alpha_2 y_2) = \alpha_1(\underbrace{\theta x_1 + (1-\theta)x_2}_{\in Q_x}) + \alpha_2(\underbrace{\theta y_1 + (1-\theta)y_2}_{\in Q_y}) = \underbrace{\alpha_1 x + \alpha_2 y}_{\in Q}.$$

Выпуклая оболочка

Определение 2. Пусть U — вещественное векторное пространство, M — множество в U. Выпуклой оболочкой множества M, обозначаемой $\mathrm{Conv}(M)$, называется пересечение всевозможных выпуклых множеств в пространстве U, содержащих M:

$$\mathrm{Conv}(M) := \cap \{Q \subseteq U \,:\, Q \,\text{выпуклое и}\, M \subseteq Q\}\,.$$

Другими словами, выпуклая оболочка $\mathrm{Conv}(M)$ — это наименьшее выпуклое множество, содержащее M: множество $\mathrm{Conv}(M)$ является выпуклым, $M \subseteq \mathrm{Conv}(M)$, и для любого другого выпуклого множества $M \subseteq Q$ справедливо $\mathrm{Conv}(M) \subseteq Q$.

Определение 3. Пусть U — вещественное векторное пространство и $x_1,\, x_2,\, \ldots,\, x_m \in U$. Тогда **выпуклой комбинацией** точек $x_1,\, x_2,\, \ldots,\, x_m$ называется точка $x = \alpha_1 x_1 + \alpha_2 x_2 + \, \ldots + \, \alpha_m x_m$, при условии, что $\alpha_1 + \alpha_2 + \, \ldots + \, \alpha_m = 1$ и $\alpha_i \geqslant 0$, $i = 1,\, \ldots,\, m$.

Утверждение 2. Множество Q в векторном пространстве является выпуклым, если и только если Q содержит любую выпуклую комбинацию своих точек.

Доказательство.

Будем считать, что $Q \neq = \emptyset$, поскольку иначе утверждение тривиально.

В одну сторону утверждение очевидно: если множество Q содержит любую выпуклую комбинацию точек из Q, то вместе с любой парой точек $x, y \in Q$ оно также содержит и выпуклую комбинацию $\alpha x + (1-\alpha)y$ для всех $0 \leq \alpha \leq 1$.

В другую сторону докажем утверждение по индукции. Из базового определения выпуклости множества Q содержит всевозможные выпуклые комбинации размера 2. Предположим теперь, что Q содержит всевозможные выпуклые комбинации размера m, где $m \geq 2$, и докажем, что Q также содержит всевозможные выпуклые комбинации размера $m+1$.

Пусть $x_1, x_2, \ldots, x_{m+1} \in U$, и пусть $\alpha_1, \ldots, \alpha_m \geq 0$, причем $\alpha_1 + \alpha_2 + \ldots + \alpha_{m+1} = 1$. Не ограничивая общности, будем считать, что $\alpha_{m+1} < 1$ (иначе это выпуклая комбинация размера $< m+1$). Тогда легко видеть, что

$$z := \sum_{i=1}^{m+1} \alpha_i x_i = (1-\alpha_{m+1})y + \alpha_{m+1}x_{m+1}, \quad \text{где } y := \sum_{i=1}^{m} \frac{\alpha_i}{1-\alpha_{m+1}} x_i.$$

Поскольку $\sum\limits_{i=1}^{m} = 1 - \alpha_{m+1}$, и веса при x_i в формуле для y неотрицательные, то y является выпуклой комбинацией m точек из множества Q, и, по предположению индукции, $y \in Q$. Наконец, согласно установленному тождеству, z представляет собой выпуклую комбинацию двух точек $y \in Q$ и $x_{m+1} \in Q$, и поэтому $z \in Q$. $\square$

Определение 4. *(Внутреннее описание выпуклой оболочки).* Пусть U — вещественное векторное пространство, M — множество в U. Тогда выпуклая оболочка $\mathrm{Conv}(M)$ — это множество, в точности состоящее из всевозможных выпуклых комбинаций точек множества M:

$$\mathrm{Conv}(M) = \cup_{m=1}^{\infty} \left\{ \sum_{i=1}^{m} \alpha_i x_i \,|\, x_i \in M,\, \theta_i \geqslant 0,\, i = 1, \ldots, m,\, \sum_{i=1}^{m} \alpha_i = 1 \right\}$$

Утверждение 3. *(Простейшие свойства выпуклой оболочки).* Пусть A и B — множества в вещественном векторном пространстве. Справедливы следующие свойства:

1. $A \subseteq \mathrm{Conv}(A)$.
2. Если $A \subseteq B$, то $\mathrm{Conv}(A) \subseteq \mathrm{Conv}(B)$.
3. Множество A выпукло, если и только если $\mathrm{Conv}(A) = A$.

Выпуклая функция

Определение 5. Пусть U — вещественное векторное пространство, Q — непустое выпуклое множество в U. Функция $f : Q \to \mathbb{R}$ называется *выпуклой*, если для любых $x, y \in Q$ и любого $0 < \alpha < 1$ выполняется

$$f(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha) f(y) \tag{5}$$

Если данное неравенство выполняется как строгое при $x \neq x$ и $0 \leqslant \lambda \leqslant 1$, то функция называется **строго выпуклой** на U.

Замечание 1. Заметим, что в этом определении подразумевается, что для любых двух допустимых точек $x, y \in Q$ функцию f возможно «вычислить» в любой промежуточной точке отрезка $[x, y]$. Именно поэтому в определении требуется, чтобы область определения Q функции f являлась выпуклым множеством.

Абсолютно аналогично вводится понятие вогнутой функции. Единственное отличие по сравнению с определением выпуклой функции состоит в том, что неравенство $\leq$ заменяется на $\geq$.

Замечание 2. Из определения легко видеть, что функция f является (строго) выпуклой тогда и только тогда, когда функция $-f$ является (строго) вогнутой.

Пример 3. Ниже приведены примеры выпуклых функций:

1. *(Аффинная функция).* Пусть в пространстве U задано (произвольное) скалярное произведение $\langle\cdot,\cdot\rangle$. Пусть $f : U \to \mathbb{R}$— аффинная функция
$$f(x) := \langle a, x\rangle + b,$$
где $a \in U$ и $b \in \mathbb{R}$. Заметим, что для этой функции неравенство (6) переходит в равенство. Таким образом, аффинная функция является одновременно и выпуклой, и вогнутой (но не строго).

2. *(Произвольная норма).* Пусть в пространстве U задана (произвольная) норма $||\cdot||$. Тогда функция $f : U \to \mathbb{R}$, заданная формулой
$$f(x) := ||x||,$$
является выпуклой.

Определение 6. Пусть U — вещественное векторное пространство, Q — непустое выпуклое множество в U. Функция $f : Q \to \mathbb{R}$ называется *μ-сильно выпуклой (сильно выпуклой)*, если для любых $x, y \in Q$ и любого $0 < \alpha < 1$ и какого-то $\mu \geqslant 0$ выполняется

$$f(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha)f(y) - \mu\alpha(1-\alpha)\|x-y\|^2. \tag{6}$$

Простейшие свойства выпуклых функций

Утверждение 4. *(Неравенство Йенсена).* Пусть U — вещественное векторное пространство, Q — непустое выпуклое множество в U, и $f : Q \to \mathbb{R}$— выпуклая функция. Пусть также $x_1, \ldots, x_k$ — точки во множестве Q и $\alpha_1, \ldots, \alpha_k$ — неотрицательные коэффициенты, суммируются в единицу: $\alpha_i \geq 0$ и $\sum\limits_{i=1}^{k} \alpha_i = 1$. Тогда справедливо следующее неравенство:
$$f\left(\sum_{i=1}^{k} \alpha_i x_i\right) \leq \sum_{i=1}^{k} \alpha_i f(x_i)$$
причем равенство достигается тогда и только тогда, когда функция f является аффинной или когда все точки совпадают: $x_1 = \ldots = x_k$.

Определение 7. *(Надграфик).*Пусть U — вещественное векторное пространство, Q — непустое выпуклое множество в U. Надграфиком функции $f : Q \to \mathbb{R}$ называется множество

$$\text{Epi }(f) = \{(x,t) \in Q \times \mathbb{R} : f(x) \leq t\}$$

Утверждение 5. *(Определение выпуклости через надграфик).* Пусть U — вещественное векторное пространство, Q — непустое выпуклое множество в U. Функции $f : Q \to \mathbb{R}$ является выпуклой тогда и только тогда, когда ее надграфик Epi (f) является выпуклым множеством в пространстве $U \times \mathbb{R}$.

Утверждение 6. *(Выпуклость множества линий уровня).* Пусть U — вещественное векторное пространство, Q — непустое выпуклое множество в U, и пусть функция $f : Q \to \mathbb{R}$ — выпуклая функция. Тогда для любого $\alpha \in \mathbb{R}$ R соответствующее множество линий уровня

$$\text{Lev }_f(\alpha) := \{x \in Q \,:\, f(x) \leq \alpha\}$$

Критерии выпуклости и сильной выпуклости

Утверждение 7. *(Условие выпуклости первого порядка).* Пусть Dom$f$ является открытым множеством, и функция f дифференцируема всюду на Dom$f$. Функция f является выпуклой тогда и только тогда, когда Dom$f$ является выпуклым множеством и

$$f(y) \geq f(x) + \nabla\langle\nabla f(x), y - x\rangle$$

для всех $x, y \in \text{Dom}f$.

μ-сильно выпукла тогда и только тогда, когда $\forall x, y \in Dom f$:

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

Утверждение 8. *(Дифференциальное условие оптимальности для выпуклой функции).* Пусть $f : U \to R \cup \{+\infty\}$ — выпуклая функция, $\mathrm{Dom} f$ является открытым множеством, и пусть $x^* \in \mathrm{Dom} f$. Тогда x^* является глобальным минимумом функции f , если и только если $\nabla f(x^*) = 0$. Другими словами, любая стационарная точка автоматически является глобальным минимумом функции f.

Утверждение 9. *(Условие выпуклости второго порядка).* Пусть $\mathrm{Dom} f$ является открытым множеством, и функция f дважды дифференцируема на $\mathrm{Dom} f$. Функция f является выпуклой тогда и только тогда, когда $\mathrm{Dom} f$ является выпуклым множеством и

$$D^2 f(x)[h,h] =: \langle \nabla^2 f(x) h, h \rangle \geq 0$$

для всех $x \in \mathrm{Dom} f$ и всех $h \in U$. Если $U = \mathbb{R}^n$, то это эквивалентно положительной полуопределенности гессиана:

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

для всех $x \in \mathrm{Dom} f$.

μ-сильно выпукла тогда и только тогда, когда $\forall x, y \in Dom f$:

$$D^2 f(x)[h,h] =: \langle \nabla^2 f(x) h, h \rangle \geq \mu \|y\|^2.$$

Если $U = \mathbb{R}^n$, то это эквивалентно

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

Пример 4. *(Одномерные выпуклые функции).* Следующие функции являются выпуклыми:

- (Экспонента) $\exp(x)$ выпукла на $\mathbb{R}$
- (Минус логарифм) $-\ln x$ выпукла на $\mathbb{R}_{++}$
- (Степенная функция)
  - x^{2p} для $p \in \{1, 2, \ldots\}$ на $\mathbb{R}$
  - x^p для $p \geq 1$ на $\mathbb{R}_+$
  - $-x^p$ для $0 \leq p \leq 1$ на $\mathbb{R}_+$
  - $1/x^p$ для $p > 0$ на $\mathbb{R}_{++}$
- $x \ln x$ на $\mathbb{R}_+$

Операции, сохраняющие выпуклость

Утверждение 10. (Операции, сохраняющие выпуклость множеств).

Следующие операции сохраняют выпуклость множеств.

1. *(Положительная взвешенная сумма).* Пусть U — вещественное векторное пространство и $f_1, \ldots, f_k : U \to R \cup \{\infty\}$ — функции. Пусть также $c_1, \ldots, c_k$ — неотрицательные коэффициенты. Рассмотрим взвешенную сумму функций $f_1, \ldots, f_k$ с коэффициентами $c_1, \ldots, c_k$, т. е. функцию $\phi : U \to R \cup \{\infty\}$, определенную по формуле

$$\phi(x) := \sum_{i=1}^{k} c_i f_i(x).$$

   Если каждая из функций $f_1, \ldots, f_k$ является выпуклой, тогда и $\phi(x)$ будет выпуклой функцией.

2. *(Аффинная подстановка аргумента).* Пусть U и V — вещественные векторные пространства, и $f : V \to R \cup \{\infty\}$ — функция. Пусть $A : U \to V$ — аффинное преобразование. Рассмотрим функцию, получающуюся из функции f с помощью аффинной подстановки аргумента, т.е. $\phi : U \to R \cup \{\infty\}$, определенную по формуле
$$f(x) := f(A(x)).$$
Если функция f выпуклая, тогда и функция ϕ также будет выпуклой.

3. *(Поточечный супремум).* Пусть U — вещественное векторное пространство. Пусть I — произвольное (не обязательно конечное и не обязательно счетное) индексное множество, и пусть для каждого индекса $i \in I$ задана функция и $f_i : U \to R \cup \{\infty\}$. Рассмотрим поточечный супремум функций f_i, т. е. функцию и $f : U \to R \cup \{\infty\}$, определенную по формуле
$$f(x) := \sup_{i \in I} f_i(x).$$
Если каждая из функций f_i является выпуклой, тогда и функция f также будет выпуклой.

4. *(Монотонная суперпозиция).* Пусть U — вещественное векторное пространство, и $f_1, \ldots, f_n : U \to R$ — функции. Пусть также $g : \mathbb{R}^n \to R \cup \{+\infty\}$ — функция. Рассмотрим функцию, являющуюся суперпозицией функции g и $f_1, \ldots, f_n$, т.е. функцию $\phi : U \to R \cup \{+\infty\}$, определенную по формуле
$$\phi(x) := g(f_1(x), \ldots f_n(x)).$$
Если каждая из функций $f_1, \ldots, f_n$ является выпуклой, а функция g является выпуклой и монотонно неубывающей, т.е. $g(y) \leq g(y')$ для всех $y \preceq y'$, тогда функция ϕ также будет выпуклой.