Методы оптимизации в машинном обучении, ВМК+Физтех, осень 2017

Семинар 4: Выпуклые множества и функции

1 Выпуклые множества

1.1 Определение и основные примеры

Определение 1.1 (Выпуклые множества)**.** Пусть U — вещественное векторное пространство, Q — подмножество U. Множество Q называется *выпуклым*, если для любых двух точек x, y из множества Q и любого $\lambda \in [0, 1]$ точка $\lambda x + (1 - \lambda)y$ также принадлежит множеству Q.

Другими словами, множество Q называется выпуклым, если для каждой пары точек $x, y \in Q$, множество Q также содержит весь *отрезок* $[x, y] := \{\lambda x + (1-\lambda)y : 0 \le \lambda \le 1\}$. Точка вида $\lambda x + (1-\lambda)y$ для $\lambda \in [0, 1]$ называется *выпуклой комбинацией* точек x, y.

Пример 1.1 (Тривиальные выпуклые множества)**.** Все пространство U, множество из одного элемента $\{a\}$ (где $a \in U$) и пустое множество $\emptyset$ являются выпуклыми.[1]

Пример 1.2 (Множество решений системы линейных ограничений)**.** Пусть в пространстве U задано (произвольное) скалярное произведение $\langle \cdot, \cdot \rangle$. Пусть $\mathcal{A}$ — произвольное (не обязательно конечное и не обязательно счетное) индексное множество, и пусть для каждого индекса $\alpha \in \mathcal{A}$ заданы вектор $a_\alpha \in U$ и скаляр $b_\alpha \in \mathbb{R}$. Рассмотрим систему линейных неравенств

$$\langle a_\alpha, x \rangle \le b_\alpha \qquad \text{для всех } \alpha \in \mathcal{A}, \tag{1.1}$$

где $x \in U$. Тогда соответствующее множество всевозможных решений системы (1.1), т. е. множество

$$Q := \{x \in U : \langle a_\alpha, x \rangle \le b_\alpha \text{ для всех } \alpha \in \mathcal{A}\}$$

является выпуклым.

Действительно, пусть $x, y \in Q$ и $\lambda \in [0, 1]$. Тогда для любого $\alpha \in \mathcal{A}$ выполняется

$$\langle a_\alpha, \lambda x + (1 - \lambda)y \rangle = \lambda \langle a_\alpha, x \rangle + (1 - \lambda)\langle a_\alpha, y \rangle \le \lambda b_\alpha + (1 - \lambda)b_\alpha = b_\alpha.$$

Таким образом, $\lambda x + (1 - \lambda)y \in Q$.

Заметим, что в этом примере абсолютно не принципиально, что в системе линейных ограничений (1.1) используется именно неравенство $\le$. Нетрудно видеть, что если некоторые (возможно, даже

[1] Последнее следует из того, что не существует такой пары точек $x, y \in \emptyset$, для которой отрезок $[x, y]$ не принадлежит множеству — для пустого множества нельзя предъявить даже точку $x \in \emptyset$.

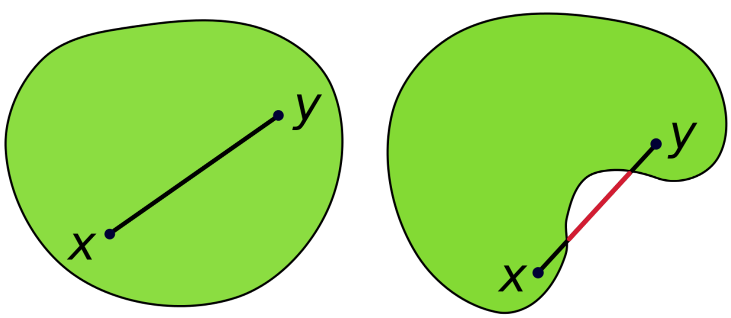

Рис. 1: Иллюстрация к определению выпуклого множества из Википедии. Слева: выпуклое множество; справа: невыпуклое множество.

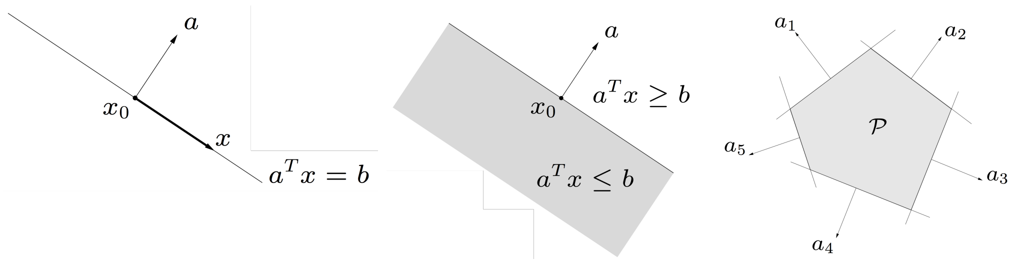

Рис. 2: Иллюстрация к примерам. Слева направо: гиперплоскость, полупространство, полиэдр.

все) из неравенств $\leq$ заменить на $\geq$ или $=$ (или $<$, или $>$), то множество Q по-прежнему останется выпуклым. Таким образом, *множество решений произвольной системы линейных ограничений является выпуклым.*

Рассмотрим наиболее популярные частные случаи. Во всех примерах мы работаем в пространстве $U := \mathbb{R}^n$ со стандартным скалярным произведением.

(a) *Гиперплоскость*, т. е. множество вида

$$\{x \in \mathbb{R}^n : \langle a, x \rangle = b\},$$

где $a \in \mathbb{R}^n \setminus \{0\}$ и $b \in \mathbb{R}$, является выпуклым.

(b) *Полупространство*, т. е. множество вида

$$\{x \in \mathbb{R}^n : \langle a, x \rangle \leq b\},$$

где $a \in \mathbb{R}^n \setminus \{0\}$ и $b \in \mathbb{R}$, является выпуклым.

(c) *Полиэдр*, т. е. множество вида

$$\{x \in \mathbb{R}^n : Ax \preceq b,\ Cx = d\},$$

где $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ и $C \in \mathbb{R}^{p \times n}$, $d \in \mathbb{R}^p$, является выпуклым. (Здесь символ $\preceq$ нужно понимать как поэлементное неравенство $\leq$.) Полиэдр представляет собой пересечение конечного числа полупространств и гиперплоскостей.

(d) Рассмотрим особый случай полиэдра:

$$\Delta_n := \left\{ x \in \mathbb{R}^n : x \succeq 0,\ \sum_{i=1}^{n} x_i = 1 \right\}.$$

Это множество называется *стандартным симплексом* и является выпуклым.

Пример 1.3 (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

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

является выпуклым.

Действительно, пусть $x, y \in B_{\|\cdot\|}(a, r)$ и $\lambda \in [0, 1]$. Тогда

$$\|\lambda x + (1-\lambda)y - a\| = \|\lambda(x-a) + (1-\lambda)(y-a)\| \leq \lambda\|x-a\| + (1-\lambda)\|y-b\| \leq \lambda r + (1-\lambda)r = r.$$

Здесь неравенство следует из неравенства треугольника для нормы и условия $\lambda \in [0, 1]$. Таким образом, $\lambda x + (1-\lambda)y \in B_{\|\cdot\|}(a, r)$.

Рассмотрим наиболее популярные частные случаи. Во всех примерах мы работаем в пространстве $U := \mathbb{R}^n$.

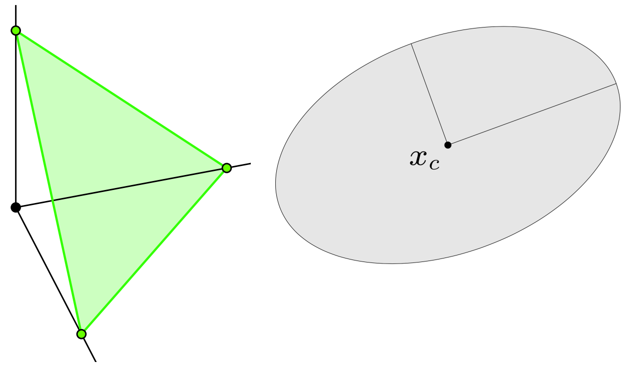

Рис. 3: Иллюстрация к примерам. Слева направо: стандартный симплекс, эллипсоид.

(a) (Евклидов шар) Пусть $\|x\| := \|x\|_2$ — евклидова норма. В этом случае рассмотренный выше абстрактный шар переходит в множество

$$B_2(a, r) := \{x \in \mathbb{R}^n : \|x - a\|_2 \leq r\},$$

которое называется (замкнутым) *евклидовым шаром* (с центром в точке a и радиусом r) и является выпуклым.

(b) (Эллипсоид) Пусть $P \in \mathbb{S}^n_{++}$, и пусть $\|x\| := \|x\|_P := \langle Px, x\rangle$. В этом случае рассмотренный выше абстрактный шар переходит в множество

$$\mathcal{E} := \{x \in \mathbb{R}^n : \langle P(x - a), x - a\rangle \leq r^2\},$$

которое называется *эллипсоидом* (с центром в точке a и радиусом r) и является выпуклым.[2]

(c) (Гипероктаэдр) Пусть $\|x\| := \|x\|_1 := \sum_{i=1}^n |x_i|$ — ℓ_1-норма. В этом случае рассмотренный выше абстрактный шар переходит в множество

$$B_1(a, r) := \left\{x \in \mathbb{R}^n : \sum_{i=1}^n |x_i - a_i| \leq r\right\},$$

которое называется *гипероктаэдром* (с центром в точке a и радиусом r) и является выпуклым.

(d) (Гиперкуб) Пусть $\|x\| := \|x\|_\infty := \max_{1 \leq i \leq n} |x_i|$ — ℓ_1-норма. В этом случае рассмотренный выше абстрактный шар переходит в множество

$$B_\infty(a, r) := \{x \in \mathbb{R}^n : a_i - r \leq x_i \leq a_i + r \text{ для всех } 1 \leq i \leq n\ \},$$

которое называется *гиперкубом* (с центром в точке a и радиусом r) и является выпуклым.

Замечание 1.1. Аналогично можно показать, что и *открытый* шар, т. е. множество $\{x \in U : \|x - a\| < r\}$, также является выпуклым. Однако *сфера*, т. е. множество $\{x \in U : \|x - a\| = r\}$, уже *не* является выпуклым (почему?).

Пример 1.4 (Множество положительно полуопределенных матриц)**.** В пространстве симметричных матриц $\mathbb{S}^n$ множество положительно полуопределенных матриц $\mathbb{S}^n_+$ является выпуклым.

[2] Другое популярное представление эллипсоида — это афинный образ единичного евклидового шара: $\mathcal{E} = \{Qx + a : \|x\|_2 \leq 1\}$, где $Q \in \mathbb{R}^{n \times n}$ — невырожденная матрица (такая, что $Q^T PQ = I_n$). Это представление можно получить из приведенного ранее, сделав замену переменной.

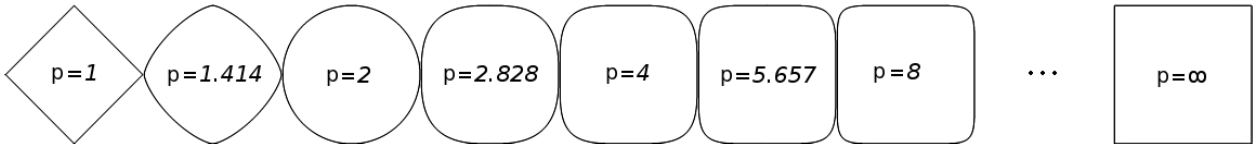

Рис. 4: Линии уровня $f(x) = \|x\|_p$ для различных значений p.

Действительно, пусть $X, Y \in \mathbb{S}^n_+$ — две положительно полуопределенные матрицы и $\lambda \in [0,1]$. Тогда для любого $u \in \mathbb{R}^n$ верно

$$\langle (\lambda X + (1-\lambda)Y)u, u\rangle = \lambda \langle Xu, u\rangle + (1-\lambda)\langle Yu, u\rangle \geq 0.$$

Здесь последнее неравенство следует из положительной полуопределенности X и Y и условия $\lambda \in [0,1]$. Таким образом, матрица $\lambda X + (1-\lambda)Y$ также будет положительно полуопределенной.

Замечание 1.2. Аналогично можно показать, что и внутренность множества $\mathbb{S}^n_+$, т. е. множество всех (строго) положительно определенных матриц $\mathbb{S}^n_{++}$, также является выпуклым.

1.2 Операции, сохраняющие выпуклость множеств

Утверждение 1.1 (Операции, сохраняющие выпуклость множеств)**.** *Следующие операции сохраняют выпуклость множеств.*

(a) (Пересечение) Пусть U — вещественное векторное пространство. Пусть $\mathcal{A}$ — произвольное (не обязательно конечное и не обязательно счетное) индексное множество, и пусть для каждого индекса $\alpha \in \mathcal{A}$ задано множество Q_α в пространстве U. Если каждое из множеств Q_α является выпуклым, тогда и их пересечение

$$\bigcap_{\alpha \in \mathcal{A}} Q_\alpha := \{x : x \in Q_\alpha \text{ для всех } \alpha \in \mathcal{A}\}$$

также является выпуклым множеством в пространстве U.

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

$$Q_1 \times \cdots \times Q_n := \{(x_1, \dots, x_n) : x_i \in Q_i \text{ для всех } 1 \leq i \leq n\}$$

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

(c) (Проекция) Пусть $U_1, \dots, U_n$ — вещественные векторные пространства, и $Q_1, \dots, Q_n$ — множества в пространствах $U_1, \dots, U_n$ соответственно. Пусть также $1 \leq i \leq n$. Если каждое из множеств $Q_1, \dots, Q_n$ является выпуклым, тогда и проекция их прямого произведения $Q_1 \times \cdots \times Q_n$ на i-ую ось, т. е. множество

$$\{x_i : (x_1, \dots, x_n) \in Q_1 \times \dots Q_n\},$$

является выпуклым множеством в пространстве U_i.

(d) (Линейная комбинация)[3] *Пусть U — вещественное векторное пространство, и $Q_1, \dots, Q_k$ — множества в пространстве U. Пусть также $\alpha_1, \dots, \alpha_k$ — скаляры (произвольного знака). Если каждое из множеств $Q_1, \dots, Q_k$ является выпуклым, тогда и их линейная комбинация*

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

[3] Данная операция в литературе часто встречается под названием *сумма Минковского*.

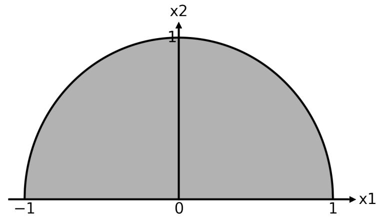

Рис. 5: Полукруг в пространстве $\mathbb{R}^2$.

также является выпуклым множеством в пространстве U.

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

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

является выпуклым множеством в пространстве V.

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

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

является выпуклым множеством в пространстве U.

Пример 1.5. Рассмотрим в пространстве $\mathbb{R}^2$ *полукруг* (рис. 5)

$$\{x \in \mathbb{R}^2 : x_1^2 + x_2^2 \le 1,\ x_2 \ge 0\}.$$

Согласно утверждению 1.1, это множество является выпуклым как пересечение единичного круга $\{x \in \mathbb{R}^2 : x_1^2 + x_2^2 \le 1\}$ и полупространства $\{x \in \mathbb{R}^2 : x_2 \ge 0\}$.

Пример 1.6. Будем работать в пространстве $\mathbb{R}^2$. Пусть для каждого $\alpha \in [-\pi, \pi]$ задано полупространство

$$Q_\alpha := \{x \in \mathbb{R}^2 : (\cos\alpha)x_1 + (\sin\alpha)x_2 \le 1\},$$

порожденное касательной к единичной окружности в точке, соответствующей углу α. Рассмотрим пересечение всех полупространств Q_α, т. е. множество

$$Q := \bigcap_{\alpha \in [-\pi,\pi]} Q_\alpha.$$

Согласно утверждению 1.1, множество Q является выпуклым как пересечение (произвольного числа) выпуклых множеств (полупространств). Нетрудно понять, что множество Q, на самом деле, является единичным кругом, который, как мы уже знаем, действительно, является выпуклым (см. рис. 6).

Пример 1.7. Рассмотрим в пространстве $\mathbb{R}^3$ *прямой круговой цилиндр* (рис. 7)

$$\{x \in \mathbb{R}^3 : x_1^2 + x_2^2 \le 1,\ -1 \le x_3 \le 1\}.$$

Согласно утверждению 1.1, это множество является выпуклым как прямое произведение единичного круга $\{x \in \mathbb{R}^2 : x_1^2 + x_2^2 \le 1\}$ в пространстве $\mathbb{R}^2$ и отрезка $[-1, 1]$ в пространстве $\mathbb{R}$.

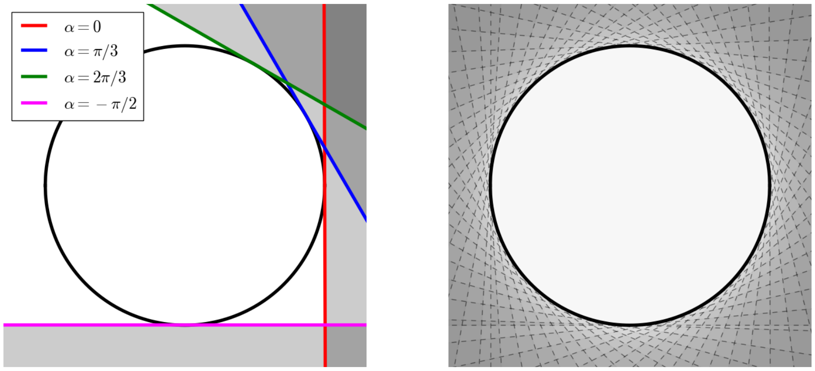

Рис. 6: Единичный круг как пересечение полупространств, образованных всевозможными касательными.

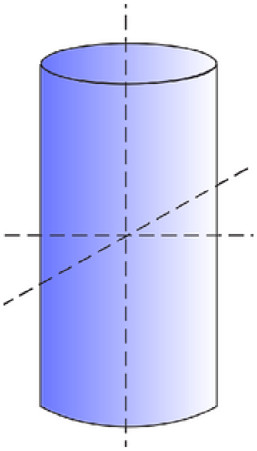

Рис. 7: Прямой круговой цилиндр.

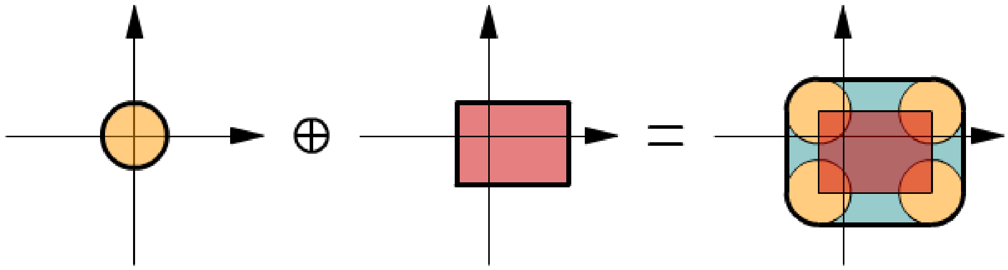

Рис. 8: Сумма круга и прямоугольника представляет собой прямоугольник большего размера с закругленными углами.

Пример 1.8. Рассмотрим в пространстве $\mathbb{R}^3$ следующий эллипсоид:

$$\{x \in \mathbb{R}^3 : x_1^2 + x_2^2 + 5x_3^2 \leq 1\}.$$

Проекцией этого множества на плоскость (x_1, x_2) будет множество

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

представляющее собой единичный круг.

Пример 1.9. Будем работать в пространстве $\mathbb{R}^2$. Пусть $Q_1 := \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq 1\}$ — единичный круг с центром в нуле и $Q_2 := \{x \in \mathbb{R}^2 : -1 \leq x_1 \leq 4,\ -3 \leq x_2 \leq 1\}$ — прямоугольник. Сумма множеств Q_1 и Q_2 будет представлять собой увеличенный прямоугольник Q_2 с закругленными углами (рис. 8). Согласно утверждению 1.1, множество $Q_1 + Q_2$ будет выпуклым.

Пример 1.10. Рассмотрим в пространстве $\mathbb{R}^n$ *эллипсоид*

$$\{Lx + a : x \in \mathbb{R}^n,\ \|x\|_2 \leq 1\},$$

где $L \in \mathbb{R}^{n \times n}$ — невырожденная матрица и $a \in \mathbb{R}^n$. Согласно утверждению 1.1, это множество является выпуклым как образ единичного евклидового шара $B_2(0,1)$ при аффинном преобразовании $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^n$, определенным по формуле $\mathcal{A}(x) = Lx + a$.

Пример 1.11 (Множество решений LMI)**.** Пусть $A_1, \ldots, A_k$ и B — матрицы в пространстве $\mathbb{S}^n$. Рассмотрим *линейное матричное неравенство (LMI)*

$$x_1 A_1 + \cdots + x_k A_k \preceq B,$$

где $x_1, \ldots, x_k \in \mathbb{R}$. Рассмотрим множество всевозможных решений этого линейного матричного неравенства, т. е. множество

$$\{x \in \mathbb{R}^k : x_1 A_1 + \cdots + x_k A_k \preceq B\}.$$

Согласно утверждению 1.1, это множество является выпуклым в пространстве $\mathbb{R}^k$ как обратный образ множества положительно полуопределенных матриц $\mathbb{S}^n_+$ при аффинном преобразовании $\mathcal{A} : \mathbb{R}^k \to \mathbb{S}^n$, заданного как

$$\mathcal{A}(x) := B - x_1 A_1 - \cdots - x_n A_n.$$

(Почему это преобразование аффинное?)

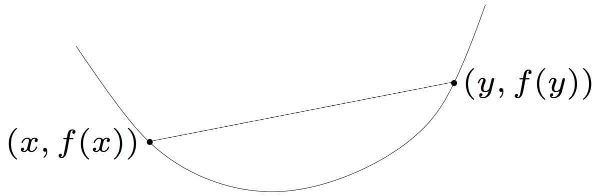

Рис. 9: Иллюстрация к определению выпуклой функции. Хорда лежит целиком над графиком функции.

2 Выпуклые функции

2.1 Определение и примеры

Определение 2.1 (Выпуклые функции)**.** 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 $\lambda \in [0, 1]$ выполняется

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y). \tag{2.1}$$

Если это неравенство является строгим для всех $x \neq y$ и $0 < \lambda < 1$, то функция f называется *строго выпуклой*.

Замечание 2.1. 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 $\geq$.

Определение 2.2 (Вогнутые функции)**.** Пусть U — вещественное векторное пространство, Q — непустое выпуклое множество в U. Функция $f : Q \to \mathbb{R}$ называется *вогнутой*, если для любых $x, y \in Q$ и любого $\lambda \in [0, 1]$ выполняется

$$f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda)f(y).$$

Если неравенство (2.1) является строгим для всех $x \neq y$ и $0 < \lambda < 1$, то функция f называется *строго вогнутой*.

Замечание 2.2. Из определения легко видеть, что функция f является (строго) выпуклой тогда и только тогда, когда функция $-f$ является (строго) вогнутой.

Пример 2.1 (Аффинная функция)**.** Пусть в пространстве U задано (произвольное) скалярное произведение $\langle \cdot, \cdot \rangle$. Пусть $f : U \to \mathbb{R}$ — *аффинная функция*

$$f(x) := \langle a, x \rangle + \beta,$$

где $a \in U$ и $\beta \in \mathbb{R}$. Заметим, что для этой функции неравенство (2.1) переходит в равенство. Таким образом, аффинная функция является одновременно и выпуклой, и вогнутой (но не строго).

Пример 2.2 (Произвольная норма)**.** Пусть в пространстве U задана (произвольная) норма $\|\cdot\|$. Тогда функция $f : U \to \mathbb{R}$, заданная формулой

$$f(x) := \|x\|,$$

является выпуклой.

Действительно, пусть $x, y \in U$ и $\lambda \in [0, 1]$. Тогда

$$f(\lambda x + (1-\lambda)y) = \|\lambda x + (1-\lambda)y\| \le \lambda\|x\| + (1-\lambda)\|y\| = \lambda f(x) + (1-\lambda)f(y).$$

Здесь неравенство следует из неравенства треугольника для нормы и условия $\lambda \in [0, 1]$.

Замечание 2.3. Рассматриваемая функция $x \mapsto \|x\|$ не является строго выпуклой (почему?).

2.2 Простейшие свойства выпуклых функций

Утверждение 2.1 (Неравенство Йенсена)**.** *Пусть U — вещественное векторное пространство, Q — непустое выпуклое множество в U, и $f : Q \to \mathbb{R}$ — выпуклая функция. Пусть также $x_1, \dots, x_k$ — точки во множестве Q и $\lambda_1, \dots, \lambda_k$ — неотрицательные коэффициенты, суммирующиеся в единицу: $\lambda_i \ge 0$ и $\sum_{i=1}^{n} \lambda_i = 1$. Тогда справедливо следующее неравенство:*

$$f\left(\sum_{i=1}^{k} \lambda_i x_i\right) \le \sum_{i=1}^{k} \lambda_i f(x_i),$$

причем равенство достигается тогда и только тогда, когда функция f является аффинной или когда все точки совпадают: $x_1 = \dots = x_k$.

Другими словами, неравенство Йенсена говорит о том, для выпуклой функции значение функции от выпуклой комбинации точек не превосходит соответствующей выпуклой комбинации значений функции.

Замечание 2.4. Неравенство Йенсена также обобщается и на случай выпуклой комбинации бесконечного (счетного или несчетного) числа точек. В случае счетного числа точек $x_1, x_2, \dots \in Q$ соответствующие суммы переходят в бесконечные суммы $\sum_{i=1}^{\infty} \lambda_i x_i$ и $\sum_{i=1}^{\infty} \lambda_i f(x_i)$, где $\lambda_1, \lambda_2, \dots \ge 0$ и $\sum_{i=1}^{\infty} \lambda_i = 1$. В наиболее общей форме (для несчетного числа точек) неравенство Йенсена формулируется в терминах вероятностных интегралов или математических ожиданий. Действительно, требование о том, что веса в выпуклой комбинации должны быть неотрицательными и суммироваться в единицу, в общем случае означает, что на множестве Q должно быть задано вероятностное распределение, по которому выполняется усреднение точек множества. Сформулируем неравенство Йенсена для случайных величин. Пусть функция f выпуклая, и X — случайная величина, принимающая значения во множестве Q. Тогда справедливо неравенство

$$f(\mathbb{E}X) \le \mathbb{E}f(X),$$

при условии, что соответствующие математические ожидания существуют.

Оказывается, что выпуклые функции и выпуклые множества тесно связаны. В частности, исследование выпуклости заданной функции всегда может быть сведено к исследованию выпуклости специального множества, ассоциированного с функцией, которое называется *надграфиком* (рис. 10).

Определение 2.3 (Надграфик)**.** Пусть U — вещественное векторное пространство, Q — непустое множество в U. *Надграфиком* функции $f : Q \to \mathbb{R}$ называется множество

$$\mathrm{Epi}(f) := \{(x, t) \in Q \times \mathbb{R} : f(x) \le t\}.$$

(Это множество лежит в пространстве $U \times \mathbb{R}$.)

Следующее утверждение можно считать альтернативным определением выпуклости функции.

Утверждение 2.2 (Определение выпуклости через надграфик)**.** *Пусть U — вещественное векторное пространство, Q — непустое выпуклое множество в U. Функция $f : Q \to \mathbb{R}$ является выпуклой тогда и только тогда, когда ее надграфик* $\mathrm{Epi}(f)$ *является выпуклым множеством в пространстве $U \times \mathbb{R}$.*

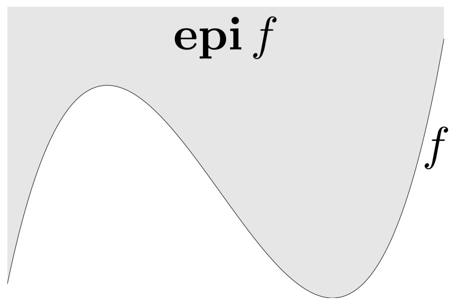

Рис. 10: Иллюстрация к определению надграфика функции.

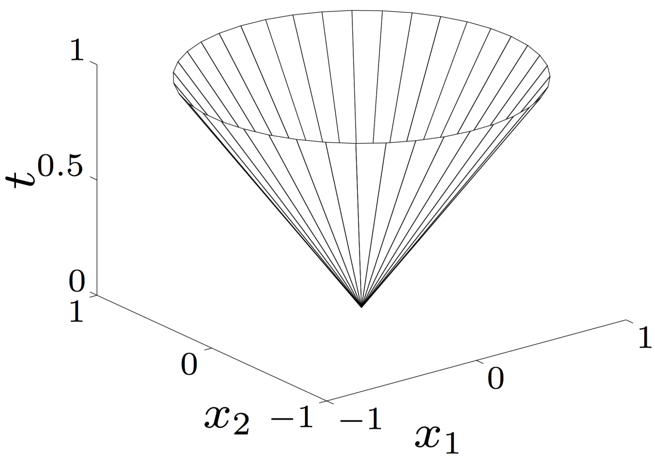

Рис. 11: Конус Лоренца для $x \in \mathbb{R}^2$.

Пример 2.3 (Конус нормы)**.** Пусть в пространстве U задана (произвольная) норма $\|\cdot\|$. Рассмотрим множество

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

представляющее собой надграфик функции $x \mapsto \|x\|$. Это множество называется *конусом нормы*. Согласно утверждению 2.2, множество множество K является выпуклым.

В случае, когда $U = \mathbb{R}^n$ и $\|x\| = \|x\|_2$ (евклидова норма), абстрактное множество K переходит в множество

$$\{(x, t) \in \mathbb{R}^n \times \mathbb{R}_+ : \|x\|_2 \le t\} = \{(x, t) \in \mathbb{R}^n \times \mathbb{R}_+ : \|x\|_2 \le t\}.$$

Это множество называется *конусом Лоренца* (рис. 11). Альтернативные названия: *конус второго порядка* или *конус мороженого*.

Следующее утверждение показывает, что у выпуклой функции все линии уровня являются выпуклыми множествами.

Утверждение 2.3 (Выпуклость множества линий уровня)**.** *Пусть U — вещественное векторное пространство, Q — непустое множество в U, и пусть $f : Q \to \mathbb{R}$ — выпуклая функция. Тогда для любого $\alpha \in \mathbb{R}$ соответствующее множество линий уровня*

$$\operatorname{Lev}_f(\alpha) := \{x \in Q : f(x) \le \alpha\}$$

является выпуклым.

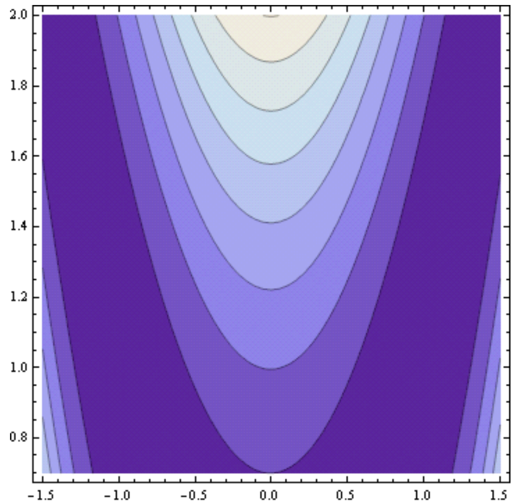

Рис. 12: Линии уровня функции Розенброка.

Из этого утверждения сразу же следует, что в выпуклой задаче оптимизации множество оптимальных решений является выпуклым.

Следствие 2.1. *Пусть U — вещественное векторное пространство, Q — непустое множество в U, и пусть $f : Q \to \mathbb{R}$ — выпуклая функция. Обозначим $f^* := \inf_{x \in Q} f(x)$. Тогда множество*

$$X^* := \{x \in Q : f(x) = f^*\}$$

является выпуклым.

Пример 2.4. С помощью утверждения 2.3 иногда можно устанавливать *невыпуклость* функции. Например, функция Розенброка, линии уровня которой приведены на рис. 12, не может быть выпуклой, потому что имеет невыпуклые линии уровня.

Замечание 2.5. Функции, которые обладают указанным выше свойством, т. е. что множество линий уровня $\mathrm{Lev}_f(\alpha)$ является выпуклым для любого $\alpha \in \mathbb{R}$, называются *квазивыпуклыми.* Как показывает это утверждение, любая выпуклая функция является квазивыпуклой. Однако обратное утверждение не верно: существуют квазивыпуклые функции, которые не являются выпуклыми (см. рис. 13). Нам не понадобится понятие квазивыпуклой функции в этом курсе.

2.3 Расширение выпуклой функции на все пространство

При работе с выпуклыми функциями $f : Q \to \mathbb{R}$ оказывается удобным считать, что функция задана не только на своей «истинной» области определения Q, но также и за ее пределами. В этом случае говорят о расширении выпуклой функции на все пространство, и считают, что за пределами своей «истинной» области определения функция принимает значение $+\infty$.

Определение 2.4 (Эффективная область определения)**.** Пусть U — векторное пространство, и $f : U \to \mathbb{R}^*$ — функция, принимающая значения во множестве расширенных вещественных чисел $\mathbb{R}^* := \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$. Будем называть *эффективной областью определения* функции f множество всех точек, в которых функция принимает конечные значения:

$$\mathrm{Dom}\, f := \{x \in U : |f(x)| < +\infty\}.$$

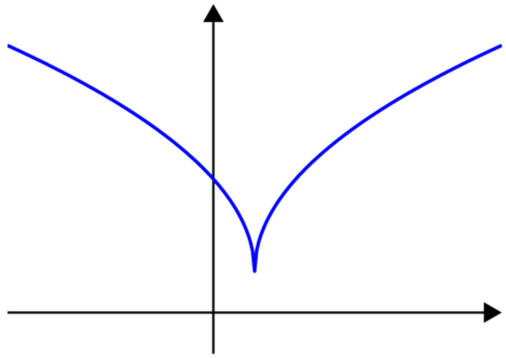

Рис. 13: Пример квазивыпуклой функции, которая не является выпуклой (из Википедии).

Определение 2.5 (Выпуклые и вогнутые расширеннозначные функции)**.** Пусть U — векторное пространство. Пусть функция $f : U \to \mathbb{R} \cup \{+\infty\}$ задана на всем пространстве и принимает расширенные вещественные значения. Функция f называется *выпуклой*, если для любых $x, y \in U$ и любых $\lambda \in (0, 1)$ выполнено

$$f(\lambda x + (1 - \lambda)y) \le \lambda f(x) + (1 - \lambda)f(y). \tag{2.2}$$

Аналогично, если $f : U \to \mathbb{R} \cup \{-\infty\}$, и для всех $x, y \in U$ и $\lambda \in (0, 1)$ указанное выше неравенство выполнено с противоположным знаком, то функция f называется вогнутой.

Замечание 2.6. Согласно определению, выпуклая расширеннозначная функция может принимать только одно расширенное значение — значение $+\infty$. Аналогично, вогнутая вещественнозначная функция может принимать только $-\infty$. Функции, которые в некоторых точках принимают $+\infty$, а в некоторых $-\infty$, не рассматриваются.

Замечание 2.7 (Операции в $\mathbb{R}^*$)**.** Операции во множестве расширенных вещественных чисел подчиняются следующим естественным правилам.

(a) Операции с вещественными числами понимаются в обычном смысле.

(b) (Порядок) Любое вещественное число строго меньше $+\infty$, а также $-\infty < +\infty$. Аналогично для $-\infty$.

(c) (Сумма) Сумма $+\infty$ и любого вещественного числа, а также двух $+\infty$ равна $+\infty$. Аналогично для $-\infty$. Сумма $+\infty$ и $-\infty$ не определена.

(d) (Произведение) Произведение $+\infty$ и положительного вещественного числа, а также произведение $+\infty$ и $+\infty$ равно $+\infty$. Произведение $+\infty$ и отрицательного вещественного числа, а также произведение $+\infty$ и $-\infty$ равно $-\infty$. Аналогично для $-\infty$. Произведение «бесконечности» и нуля не определено.

Замечание 2.8. Определение 2.5 автоматически накладывает условие на выпуклость (эффективной) области определения функции f. Действительно, пусть $x, y \in \operatorname{Dom} f$ и $\lambda \in [0, 1]$. Поскольку $x, y \in \operatorname{Dom} f$, то $f(x)$ и $f(y)$ являются конечными. Отсюда следует, что правая часть в неравенстве (2.2) также конечна. Но это возможно лишь в том случае, когда и левая часть в неравенстве (2.2) конечна. Значит, $f(\lambda x + (1 - \lambda)y) < +\infty$ и $\lambda x + (1 - \lambda)y \in \operatorname{Dom} f$. Таким образом, расширенное определение выпуклости 2.5 эквивалентно введенному до этого обычному определению выпуклости 2.1 (в котором функция задана на множестве $Q := \operatorname{Dom} f$).

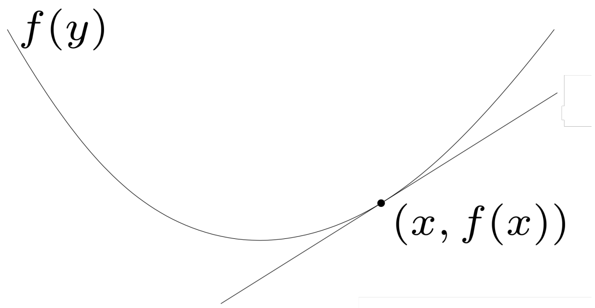

Рис. 14: Иллюстрация к условию выпуклости первого порядка. График функции лежит всюду выше касательной, проведенной к графику в любой точке x.

Пример 2.5 (Индикатор выпуклого множества)**.** Пусть U — векторное пространство, и пусть Q — выпуклое множество в U. Рассмотрим функцию $\delta_Q : U \to \mathbb{R} \cup \{+\infty\}$, заданную формулой

$$\delta_Q(x) := \begin{cases} 0, & \text{если } x \in Q, \\ +\infty, & \text{если } x \notin Q. \end{cases}$$

Эта функция называется *индикатором* множества Q и является выпуклой (почему?).

2.4 Дифференциальные критерии выпуклости

Утверждение 2.4 (Условие выпуклости первого порядка)**.** *Пусть* $\operatorname{Dom} f$ *является открытым множеством, и функция* f *дифференцируема всюду на* $\operatorname{Dom} f$*. Функция* f *является выпуклой тогда и только тогда, когда* $\operatorname{Dom} f$ *является выпуклым множеством и*

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

для всех $x, y \in \operatorname{Dom} f$*.*

Утверждение 2.5 (Дифференциальное условие оптимальности для выпуклой функции)**.** *Пусть* $f : U \to \mathbb{R} \cup \{+\infty\}$ *— выпуклая функция, и пусть* x^* *— некоторая* <u>*внутренняя*</u> *точка множества* $\operatorname{Dom} f$*. Точка* x^* *является глобальным минимумом функции* f *тогда и только тогда, когда* $\nabla f(x^*) = 0$*. Другими словами, любая стационарная точка автоматически является глобальным минимумом функции* f*.*

Доказательство. Согласно условию оптимальности первого порядка, для всех $x \in \operatorname{Dom} f$ справедлива оценка

$$f(x) \geq f(x^*) + \langle \nabla f(x^*), x - x^* \rangle = f(x^*).$$

□

Пример 2.6. Это утверждение позволяет для выпуклых дифференцируемых функций не задумываться о том, достигается ли глобальный минимум или нет. Например, вспомним задачу регрессии наименьших квадратов $f : \mathbb{R}^n \to \mathbb{R}$:

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) := \frac{1}{2} \|Ax - b\|_2^2 \right\}.$$

Эта функция является всюду дифференцируемой ($\operatorname{Dom} f = \mathbb{R}^n$), и, значит, поиск ее минимума эквивалентен решению системы уравнений

$$\nabla f(x) = A^T(Ax - b) = 0.$$

Согласно установленному утверждению, решения задачи — это в точности все решения этой системы линейных уравнений, и только они. Таким образом, можно просто решать систему линейных уравнений и не переживать о том, что таким образом могут быть найдены какие-то стационарные точки, которые не являются глобальными решениями задачи. (Для невыпуклых функций так делать нельзя!)

Утверждение 2.6 (Условие выпуклости второго порядка)**.** *Пусть* $\operatorname{Dom} f$ *является открытым множеством, и функция* f *дважды дифференцируема на* $\operatorname{Dom} f$*. Функция* f *является выпуклой тогда и только тогда, когда* $\operatorname{Dom} f$ *является выпуклым множеством и*

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

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

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

для всех $x \in \operatorname{Dom} f$*.*

Пример 2.7 (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 $p \in \{1, 2, \dots\}$ 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 $\mathbb{R}_{++}$
- $x \ln x$ выпукла на $\mathbb{R}_+$

Доказывается через условие второго порядка.

Пример 2.8 (Квадратичная функция)**.** Пусть $A \in \mathbb{S}^n$, $b \in \mathbb{R}^n$. Квадратичная функция

$$f(x) := \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle$$

является выпуклой тогда и только тогда, когда $A \succeq 0$ и вогнутой тогда и только тогда, когда $A \preceq 0$. Это следует из условия второго порядка: $\nabla^2 f(x) = A$.

Пример 2.9 (Логарифм суммы экспонент)**.** Функция $f : \mathbb{R}^n \to \mathbb{R}$

$$f(x) := \ln \left(\sum_{i=1}^n e^{x_i} \right)$$

является выпуклой. Ее гессиан равен

$$\nabla^2 f(x) = \operatorname{Diag}\{\pi(x)\} - \pi(x) \pi(x)^T,$$

где $\pi(x) := \exp(x) / \langle 1_n, \exp(x) \rangle$ (поэлементно). Гессиан оказывается положительно полуопределенным:

$$\langle \nabla^2 f(x) u, u \rangle = \langle \operatorname{Diag}\{\pi\} u, u \rangle - \langle \pi, u \rangle^2 = \sum_{i=1}^n \pi_i u_i^2 - \left(\sum_{i=1}^n \pi_i u_i \right)^2 \geq 0.$$

Последнее неравенство следует из неравенства Коши-Буняковского и того факта, что $\sum_{i=1}^n \pi_i = 1$.

Пример 2.10 (Минус логарифм определителя)**.** Функция

$$f(X) := -\ln \operatorname{Det}(X)$$

является выпуклой на $\mathbb{S}^n_{++}$. Действительно, рассмотрим

$$D^2 f(X)[H, H] = \langle X^{-1} H X^{-1}, H \rangle.$$

Покажем, что $D^2 f(X)[H, H] \geq 0$ для всех $X \in \mathbb{S}^n_{++}$ и всех $H \in \mathbb{S}^n$. Поскольку X является симметричной положительно определенной матрицей, можно рассмотреть ее корень $X^{1/2}$. Тогда

$$D^2 f(X)[H, H] = \langle X^{-1/2} H X^{-1/2}, X^{-1/2} H X^{-1/2} \rangle = \|X^{-1/2} H X^{-1/2}\|_F^2 \geq 0.$$

Следующее свойство позволяет с помощью производных доказать выпуклость на внутренности множества, а затем расширить это понятие на все множество — если функция непрерывна на множестве.

Утверждение 2.7 (Полезное свойство расширения на замыкание)**.** *Пусть функция f является выпуклой всюду на* <u>*внутренности*</u> $\operatorname{Dom} f$*, и непрерывной всюду на* $\operatorname{Dom} f$*. Тогда f является выпуклой на всем* $\operatorname{Dom} f$*.*

2.5 Операции, сохраняющие выпуклость функций

Утверждение 2.8 (Операции, сохраняющие выпуклость функций)**.** *Следующие операции сохраняют выпуклость функций.*

(a) (Положительная взвешенная сумма) Пусть U — вещественное векторное пространство и $f_1, \ldots, f_k : U \to \mathbb{R} \cup \{+\infty\}$ — функции. Пусть также $w_1, \ldots, w_k$ — положительные коэффициенты. Рассмотрим взвешенную сумму функций $f_1, \ldots, f_k$ с коэффициентами $w_1, \ldots, w_k$, т. е. функцию $\phi : U \to \mathbb{R} \cup \{+\infty\}$, определенную по формуле

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

Если каждая из функций $f_1, \ldots, f_k$ является выпуклой, тогда и ϕ будет выпуклой функцией.

(b) (Аффинная подстановка аргумента) Пусть U и V — вещественные векторные пространства, и $f : V \to \mathbb{R} \cup \{+\infty\}$ — функция. Пусть $\mathcal{A} : U \to V$ — аффинное преобразование из пространства U в пространство V, т. е. преобразование вида $\mathcal{A}(x) = Lx + a$, где $L : U \to V$ — линейное преобразование и $a \in V$. Рассмотрим функцию, получающуюся из функции f с помощью аффинной подстановки аргумента, т. е. функцию $\phi : U \to \mathbb{R} \cup \{+\infty\}$, определенную по формуле

$$\phi(x) := f(\mathcal{A}(x)).$$

Если функция f выпуклая, тогда и функция ϕ также будет выпуклой.

(c) (Поточечный супремум) Пусть U — вещественное векторное пространство. Пусть $\mathcal{A}$ — произвольное (не обязательно конечное и не обязательно счетное) индексное множество, и пусть для каждого индекса $\alpha \in \mathcal{A}$ задана функция $f : U \to \mathbb{R} \cup \{+\infty\}$. Рассмотрим поточечный супремум функций f_α, т. е. функцию $\phi : U \to \mathbb{R} \cup \{+\infty\}$, определенную по формуле

$$\phi(x) := \sup_{\alpha \in \mathcal{A}} f_\alpha(x).$$

Если каждая из функций f_α является выпуклой, тогда и функция ϕ также будет выпуклой.

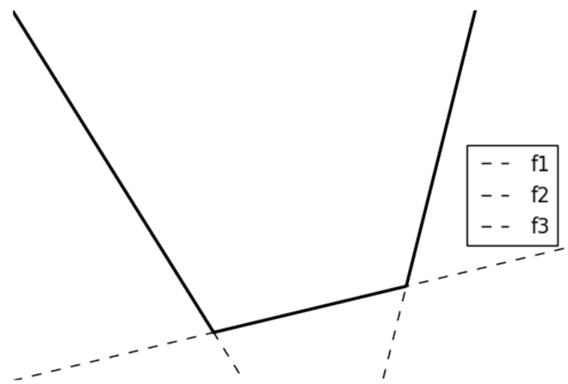

Рис. 15: Иллюстрация к примеру поточечный максимум. Так как линейная функция — выпуклая, максимум из линейных функций — выпуклая функция.

(d) *(Монотонная суперпозиция) Пусть U — вещественное векторное пространство, и $f_1, \ldots, f_n : U \to \mathbb{R}$ — функции. Пусть также $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ — функция. Рассмотрим функцию, являющуюся суперпозицией функции g и $f_1, \ldots, f_n$, т. е. функцию $\phi : U \to \mathbb{R} \cup \{+\infty\}$, 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 <u>монотонно неубывающей</u>, т. е. $g(y) \leq g(y')$ для всех $y \preceq y'$, тогда функция ϕ также будет выпуклой.

(e) *(Частичная минимизация) Пусть U и V — вещественные векторные пространства, и $f : U \times V \to \mathbb{R} \cup \{+\infty\}$ — функция. Рассмотрим функцию $\phi : U \to \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$, определенную по формуле*

$$\phi(x) := \inf_{y \in V} f(x, y).$$

Если функция f является выпуклой (как функция одновременно двух переменных x и y), тогда ϕ является выпуклой функцией (при условии, что ϕ ни в одной точке не принимает значение $-\infty$).

Пример 2.11 (Взвешенная сумма и аффинная подстановка аргумента)**.** Пусть $a, b \in \mathbb{R}^k$ и $c \in \mathbb{R}^k_+$. Функция

$$f(x) := \sum_{i=1}^{k} c_i \exp(\langle a_i, x \rangle + b_i)$$

является выпуклой как взвешенная сумма экспонент с аффинной подстановкой аргумента.

Пример 2.12 (Поточечный максимум)**.** Пусть $f_1, \ldots, f_k$ — выпуклые функции. Тогда

$$\phi(x) := \max\{f_1(x), \ldots, f_k(x)\}$$

будет выпуклой функцией, потому что максимум — это частный случай супремума.

Пример 2.13 (Минимальное и максимальное собственные значения)**.** Работаем в пространстве симметричных матриц $\mathbb{S}^n$. Функция $\lambda_{\max}(X)$ является выпуклой, а функция $\lambda_{\min}(X)$ является вогнутой. Это следует из представления

$$\lambda_{\max}(X) := \max\{\langle Xu, u \rangle : u \in S_2^{n-1}\}, \qquad \lambda_{\min}(X) := \min\{\langle Xu, u \rangle : u \in S_2^{n-1}\},$$

где $S_2^{n-1} := \{u \in \mathbb{R}^n : \|u\|_2 = 1\}$ — евклидова сфера в пространстве $\mathbb{R}^n$.

Замечание 2.9 (Опорная функция множества)**.** Пусть M — произвольное (не обязательно выпуклое) непустое множество. Тогда *опорная функция* этого множества

$$S_M(y) := \sup_{x \in M} \langle y, x \rangle$$

является выпуклой как поточечный супремум от аффинных функций.

Пример 2.14 (Норма в степени)**.** Пусть $\|\cdot\|$ — произвольная норма. Тогда функция

$$f(x) := \|x\|^p$$

является выпуклой при $p \geq 1$ на всем пространстве. Здесь используется композиция выпуклой монотонно возрастающей одномерной функции x^p и выпуклой функции $\|x\|$.

Пример 2.15 (Расстояние до выпуклого множества)**.** Пусть Q — выпуклое множество, $\|\cdot\|$ — произвольная норма, и пусть $x \in U$, Тогда функция

$$f(x) := \rho(x, Q) := \inf_{y \in Q} \|x - y\|$$

является выпуклой как частичная минимизация $\|x-y\|+\delta_Q(y)$ по y (почему $\|x-y\|$ выпукла совместно по (x, y)?).

Пример 2.16. Пусть f — выпуклая функция. Тогда функция

$$\phi(x) := \inf_y \{f(y) : Ay = x\}$$

также выпуклая как частичная минимизация по y функции

$$g(x, y) := \begin{cases} f(y), & \text{если } Ay = x, \\ +\infty, & \text{иначе,} \end{cases}$$

которая является совместно выпуклой по (x, y).