Элементы конечномерного выпуклого анализа

Локуциевский Л.В.

Дорогие студенты, перед вами черновые записи лекций, прочитанных мною в осеннем семестре 2017 г. на механико-математическом факультете МГУ. Я постарался изложить материал в максимально ясной и удобной форме. Например, я переделал некоторые доказательства, так что они стали легче, чем рассказанные на лекциях. Тем не менее, я почти уверен, что некоторые места могли получиться не до конца понятными, или запутанными. К тому же, в результате, в данные записи могли вкрасться не только опечатки (которых, я уверен, много), но и даже ошибки. Я буду очень! признателен, если вы сообщите мне обо всех найденных проблемах.

Содержание

1 Выпуклые множества и операции над ними

Основные объекты исследования выпуклого анализа – это *выпуклые множества* и *выпуклые функции*. Наиболее удобное определение выпуклой функции прямо опирается на выпуклость надграфика функции, поэтому мы начнем с изучения основных свойств выпуклых множеств.

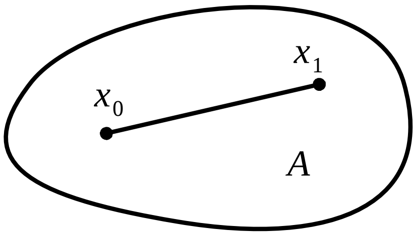

Рис. 1: Выпуклое множество

Определение 1. Множество $C \subset \mathbb{R}^d$ называется выпуклым, если любой отрезок $[x_0, x_1]$ с концами из C целиком лежит в C (см. рис. 1).

Определим теперь основные операции над выпуклыми множествами, которые сохраняют свойство выпуклости (аналогично тому, как сложение и умножение сохраняют свойство гладкости).

Для проверки выпуклости множества удобно пользоваться тем, что для любых двух точек $x_0, x_1 \in \mathbb{R}^d$ отрезок $[x_0, x_1]$ задается формулой

$$[x_0, x_1] = \{x_\lambda,\ \ \lambda \in [0; 1]\},\ \text{где}\ x_\lambda \stackrel{\text{def}}{=} (1-\lambda)x_0 + \lambda x_1 = x_0 + \lambda(x_1 - x_0)$$

Список операций. Пусть множества $C, D \subset \mathbb{R}^n$ выпуклы. Из определения выпуклого множества непосредственно следует, что перечисленные ниже множества являются выпуклыми:

- Если $\mu \in \mathbb{R}$, то гомотетия $\mu C = \{\mu x,\ x \in C\}$ множества C тоже является выпуклым множеством, что легко следует из определения. Также выпуклым является конус

$$\operatorname{cone} C = \bigcup_{\mu \geqslant 0} \mu C.$$

- Выпуклой является *сумма по Минковскому* множеств C и D:

$$C + D \overset{\text{def}}{=} \{x + y,\ x \in C \text{ и } y \in D\}$$

Действительно, если $z_0, z_1 \in C + D$, то $z_0 = x_0 + y_0$ и $z_1 = x_1 + y_1$, где $x_0, x_1 \in C$ и $y_0, y_1 \in D$. Поэтому

$$z_\lambda = (1-\lambda)z_0 + \lambda z_1 = (1-\lambda)x_0 + \lambda x_1 + (1-\lambda)y_0 + \lambda y_1 = x_\lambda + y_\lambda \in C + D.$$

- Таким образом, если $C_1, \dots, C_k$ – выпуклые множества из $\mathbb{R}^d$, а $\mu_1, \dots, \mu_k$ – действительные числа, то следующее множество выпукло:

$$\mu_1 C_1 + \dots + \mu_k C_k.$$

- Пересечение $C \cap D$ выпуклых множеств выпукло. Объединение же выпуклых множеств может быть невыпуклым.

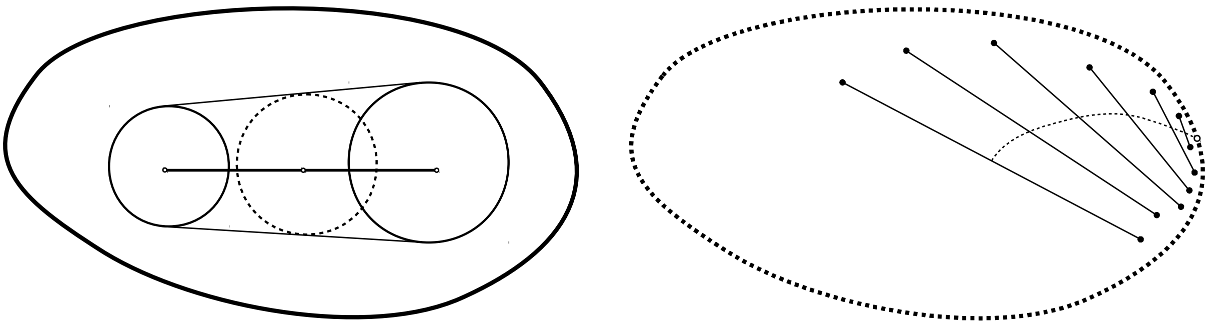

Рис. 2: Схемы доказательства выпуклости множеств $\operatorname{int} C$ (слева) и $\operatorname{cl} C$ (справа).

- Внутренность $\operatorname{int} C$ и замыкание $\operatorname{cl} C$ выпуклого множества выпуклы. Схемы доказательств изображены на рис. 2. Отметим, что $\operatorname{int}\operatorname{int} C = \operatorname{int} C$ и $\operatorname{cl}\operatorname{cl} C = \operatorname{cl} C$ независимо от выпуклости множества C.

- Если $E \subset \mathbb{R}^d$ – не обязательно выпуклое множество, то выпуклыми являются его *аффинная* и *выпуклая оболочки*:

$$\operatorname{aff} E = \left\{ \sum_{i=1}^{N} \lambda_i x_i,\ \text{где } N \in \mathbb{N},\ x_i \in E \text{ и } \sum_{i=1}^{N} \lambda_i = 1 \right\};$$

$$\operatorname{conv} E = \left\{ \sum_{i=1}^{N} \lambda_i x_i, \text{ где } N \in \mathbb{N}, \ x_i \in E, \sum_{i=1}^{N} \lambda_i = 1 \text{ и } \lambda_i \geqslant 0 \right\}.$$

Отметим, что если $E \neq \emptyset$, то $\operatorname{aff} E$ является аффинным подпространством[1] в $\mathbb{R}^d$.

- Последняя нужная нам операция – это взятие *относительной внутренности* выпуклого множества. Устроена она следующим образом: поскольку любое множество лежит в своей аффинной оболочке, то можно рассмотреть внутренность этого множества в ограничении на его аффинную оболочку:

$$\operatorname{ri} C = \operatorname{int}|_{\operatorname{aff} C} C \qquad \text{(relative interior)}$$

Например, относительной внутренностью отрезка на плоскости будет интервал с теми же концами, тогда как его обычная внутренность пуста.

Выпуклая оболочка. В определениях аффинной и выпуклой оболочек мы использовали линейные комбинации с ограничениями на используемые коэффициенты λ_i. Такие линейные комбинации очень часто встречаются в выпуклом анализе и носят специальные названия:

Определение 2. Линейная комбинация $\sum_{i=1}^N \lambda_i x_i$ точек $x_i \in \mathbb{R}^d$ с коэффициентами $\lambda_i \in \mathbb{R}$ называется *аффинной комбинацией*, если $\sum_{i=1}^N \lambda_i = 1$. В свою очередь, аффинная комбинация называется *выпуклой*, если к тому же все коэффициенты неотрицательны, $\lambda_i \geqslant 0$.

Замечание 1. В выпуклой комбинации $\lambda_i \leqslant 1$ для всех i.

Выпуклые множества (как и выпуклые функции) обладают рядом очень сильных геометрических свойств, которыми не обладает произвольное подмножество (или функция) в $\mathbb{R}^d$. Начнем с естественного факта:

***Предложение* 1.** *Если множество C выпукло, то любая выпуклая комбинация точек из C снова лежит в C.*

Доказательство. Доказательство проведем индукцией по числу слагаемых N в выпуклой комбинации $\sum_{i=1}^N \lambda_i x_i$. Для $N = 1, 2$ утверждение тривиально. Сделаем шаг индукции. Предположим сначала, что $\lambda_N = 1$. Тогда все остальные коэффициенты $\lambda_1, \dots, \lambda_{N-1}$ должны быть нулями и $\sum_{i=1}^N \lambda_i x_i = x_N \in C$. Пусть теперь $\lambda_N < 1$. Тогда

$$\sum_{i=1}^{N} \lambda_i x_i = (1 - \lambda_N) \sum_{i=1}^{N-1} \frac{\lambda_i}{1 - \lambda_N} x_i + \lambda_N x_N.$$

Поскольку $x_N \in C$, остается проверить, что сумма $\sum_{i=1}^{N-1} \frac{\lambda_i}{1-\lambda_N} x_i$ тоже лежит в C. Действительно, она является выпуклой комбинацией точек $x_1, \dots, x_{N-1}$, так как $\frac{\lambda_i}{1-\lambda_N} \geqslant 0$ и

$$\sum_{i=1}^{N-1} \frac{\lambda_i}{1 - \lambda_N} = \frac{\sum_{i=1}^{N-1} \lambda_i}{1 - \lambda_N} = \frac{1 - \lambda_N}{1 - \lambda_N} = 1.$$

[1] Т.е. подпространством, не обязательно проходящим через начало координат.

Поэтому $\sum_{i=1}^{N-1} \frac{\lambda_i}{1-\lambda_N} x_i$ лежит в C по индуктивному предположению. $\square$

***Предложение* 2.** *Пусть E – произвольное множество. Тогда множество $\operatorname{conv} E$ выпукло.*

Доказательство. Действительно, если $x, y \in \operatorname{conv} E$, то $x = \sum_i \lambda_i x_i$ и $y = \sum_j \mu_j y_j$, где $x_i, y_j \in E$, $\lambda_i \geqslant 0$, $\mu_j \geqslant 0$ и $\sum_i \lambda_i = \sum_j \mu_j = 1$. Мы утверждаем, что $z = (1-\nu)x + \nu y$ лежит в $\operatorname{conv} E$ для любого числа $\nu \in [0; 1]$. Действительно, z есть линейная комбинация элементов x_i и y_j из E с неотрицательными коэффициентами $(1-\nu)\lambda_i$ и $\nu\mu_j$, сумма которых есть

$$\sum_i (1-\nu)\lambda_i + \sum_j \nu\mu_j = (1-\nu)\sum_i \lambda_i + \nu \sum_j \mu_j = (1-\nu) + \nu = 1.$$

$\square$

***Следствие* 1.** *Множество C является выпуклым, если и только если $\operatorname{conv} C = C$.*

Доказательство. Множество $\operatorname{conv} C$ выпукло. Поэтому если $C = \operatorname{conv} C$, то C тоже выпукло.

Обратно. Очевидно $C \subset \operatorname{conv} C$, так как в определении выпуклой комбинации можно рассмотреть комбинации из одного элемента. Обратное включение $\operatorname{conv} C \subset C$ есть в точности предложение 1 $\square$

***Следствие* 2.** $\operatorname{conv}\operatorname{conv} E = \operatorname{conv} E$.

***Следствие* 3.** $\operatorname{conv} E$ *– наименьшее выпуклое множество, содержащее E.*

Доказательство. Если $E \subset C$ и C выпукло, то $\operatorname{conv} E \subset C$ по предложению 1. То есть множество $\operatorname{conv} E$ должно лежать в любом выпуклом множестве, содержащем E. Остается заметить, что само множество $\operatorname{conv} E$ выпукло и, следовательно, является наименьшим. $\square$

Размерность. Аналогично доказывается, что $\operatorname{aff} E$ – наименьшее аффинное подпространство, содержащее E. Очевидно, любое аффинное подпространство (в том числе $\operatorname{aff} E$) выпукло.

Определение 3. Если непустое множество C выпукло, то его размерностью называется размерность его аффинной оболочки:

$$\dim C \stackrel{\text{def}}{=} \dim \operatorname{aff} C$$

На первый взгляд такое определение несколько странно, так как бывают очень сложные множества, имеющие даже нецелые размерности (например, множество Кантора лежит на прямой, но имеет размерность $\log 2/\log 3$). Однако, с выпуклыми множествами никаких неприятностей быть не может по следующим причинам. Если множество C лежит в некотором аффинном подпространстве A, то размерность естественно определять так, чтобы размерность C не превосходила размерности A. Равенство же достигается, например, если внутренность C в A непуста – в этом случае множество C содержит шар радиуса ε, размерность которого равна $\dim A$.

Таким образом, очевидно, необходимо считать, что $\dim C \leqslant \dim \operatorname{aff} C$. Остается доказать, что C имеет в $\operatorname{aff} C$ непустую внутренность.

Относительная внутренность. Одно из важнейших качеств выпуклых множеств состоит в том, что они почти неотличаются от своей относительной внутренности. Начнем с

***Предложение* 3.** *Всякое непустое выпуклое множество имеет непустую относительную внутренность:*

$$C \text{ – выпукло и } C \neq \emptyset \qquad \Longrightarrow \qquad \operatorname{ri} C \neq \emptyset.$$

Доказательство. Поскольку $C \subset \operatorname{aff} C$, то при доказательстве мы можем ограничиться рассмотрением точек только из подпространства $\operatorname{aff} C \neq \emptyset$, вместо того, чтобы использовать точки всего пространства $\mathbb{R}^d$. Поэтому, если мы докажем предложение для случая $\operatorname{aff} C = \mathbb{R}^d$, то общий случай $\operatorname{aff} C \subset \mathbb{R}^d$ получится автоматически (этим приемом мы будем пользоваться очень часто). Таким образом, достаточно доказать, что если $\operatorname{aff} C = \mathbb{R}^d$, то $\operatorname{int} C \neq \emptyset$.

Поскольку $\dim \operatorname{aff} C = d$, то в C найдутся точки $x_1, \dots, x_{d+1}$, количеством $d+1$, не лежащие в одной $(d-1)$-мерной гиперплоскости. Рассмотрим d-мерный симплекс Δ с вершинами в этих точках. Он имеет непустую внутренность, $\operatorname{int} \Delta \neq \emptyset$ (см. упражнение 1.1). Поскольку $\Delta \subset C$, то $\operatorname{int} \Delta \subset \operatorname{int} C$ и $\operatorname{int} C \neq \emptyset$. □

***Следствие* 4.** *Если непустое выпуклое множество $C \subset \mathbb{R}^d$ имеет пустую внутренность, то оно целиком лежит в аффинном подпространстве размерности меньше d.*

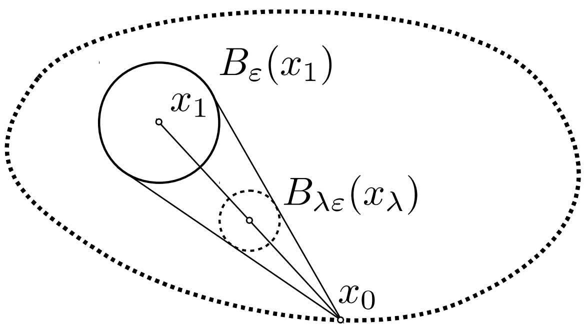

Рис. 3: Иллюстрация к доказательству предложения 4.

Всякую точку выпуклого множества можно приблизить точками из его относительной внутренности:

***Предложение* 4.** *Пусть $C \subset \mathbb{R}^d$ – выпуклое множество. Если $x_0 \in C$ и $x_1 \in \operatorname{ri} C$, то полуинтервал $(x_0; x_1]$ лежит в $\operatorname{ri} C$.*

Доказательство. Если множество C пусто, то доказательство тривиально. Если $C \neq \emptyset$, то без ограничения общности $\operatorname{aff} C = \mathbb{R}^d$, и потому $\operatorname{ri} C = \operatorname{int} C \neq \emptyset$ и $x_1 \in \operatorname{int} C$. Для достаточно маленького числа $\varepsilon > 0$ открытый шар $B_{x_1}(\varepsilon)$ с центром в точке x_1 и радиуса ε лежит в C. Рассмотрим выпуклую оболочку точки x_0 и шара $B_{x_1}(\varepsilon)$ (см. рис. 3). Она является объединением шаров $B_{x_\lambda}(\lambda\varepsilon)$, каждый из которых должен лежать в C. Поэтому точки x_λ являются внутренними при $\lambda > 0$. □

***Предложение* 5.** *Всякое выпуклое множество $C \subset \mathbb{R}^d$ содержится в замыкании своей относительной внутренности (см. упражнение 1.3):*

$$C \subset \operatorname{cl} \operatorname{ri} C.$$

Доказательство. Пусть $x_0 \in C$ и $x_1 \in \operatorname{ri} C$. Тогда $x_\lambda \in \operatorname{ri} C$ при $\lambda > 0$ по предложению 4. Осталось заметить, что $x_\lambda \to x_0$ при $\lambda \to 0$. □

***Следствие* 5.** *Выпуклое множество C замкнуто, если и только если $C = \operatorname{cl}\operatorname{ri} C$.*

Доказательство. Очевидно, если $C = \operatorname{cl}\operatorname{ri} C$, то множество C замкнуто, так как замкнуто множество $\operatorname{cl}\operatorname{ri} C$. Поэтому интерес представляет только обратное утверждение.

Пусть C замкнуто. Тогда $C \subset \operatorname{cl}\operatorname{ri} C$ по предложению 3. Обратное включение доказывается просто:

$$\operatorname{ri} C \subset C \qquad \Rightarrow \qquad \operatorname{cl}\operatorname{ri} C \subset \operatorname{cl} C = C.$$

□

Определение 4. Выпуклое множество C называется относительно открытым, если оно открыто в своей аффинной оболочке $\operatorname{aff} C$, т.е. если $C = \operatorname{ri} C$.

***Предложение* 6.** *Если выпуклые множества C_1 и C_2 относительно открыты, то их сумма $C_1 + C_2$ тоже относительно открыта.*

Доказательство. Пусть $C = C_1 + C_2$. Требуется доказать, что $C \subset \operatorname{ri} C$. Пусть $x \in C$, тогда $x = y + z$, где $y \in C_1$ и $z \in C_2$. Используя параллельный перенос множеств C_1 и C_2, будем без ограничения считать, что $y = 0$ и $z = 0$ (а, значит, и $x = 0$). Также без ограничения общности будем считать, что линейное подпространство $\operatorname{aff} C = \operatorname{aff} C_1 + \operatorname{aff} C_2$ совпадает со всем пространством $\mathbb{R}^d$. Введем на $\mathbb{R}^d$ такое скалярное произведение, что $\operatorname{aff} C_1 \perp \operatorname{aff} C_2$. Покажем, что B_ε – шар с центром в 0 и радиуса $\varepsilon > 0$ – содержится в C, если ε достаточно мало. Поскольку множества C_1 и C_2 относительно открыты, то для некоторого $\varepsilon > 0$ шары $B^1_\varepsilon = B_\varepsilon \cap \operatorname{aff} C_1$ и $B^2_\varepsilon = B_\varepsilon \cap \operatorname{aff} C_2$ лежат в C_1 и C_2, соответственно. Заметим, что B^1_ε есть в точности ортогональная проекция шара B_ε на $\operatorname{aff} C_1$, и аналогично B^2_ε. Поскольку подпространства $\operatorname{aff} C_1$ и $\operatorname{aff} C_2$ ортогональны и в сумме дают все пространство $\mathbb{R}^d$, то шар B_ε содержится в сумме своих проекций B^1_ε и B^2_ε, и, следовательно, лежит в $C_1 + C_2$. □

На самом деле, можно доказать и более сильное утверждение: для любых выпуклых множеств C_1 и C_2 выполнено $\operatorname{ri}(C_1 + C_2) = \operatorname{ri} C_1 + \operatorname{ri} C_2$. Однако, простое доказательство этого утверждения опирается на теорему отделимости, которую мы докажем в следующем параграфе (см. упражнение 2.4).

Отметим, что аналогичное утверждение про замыкание неверно: можно построить пару замкнутых выпуклых множеств, сумма которых невыпукла (см. упражнение 1.9).

Теорема Каратеодори. Любая точка $x \in \operatorname{conv} E$ по определению выпуклой оболочки является выпуклой комбинацией точек из E. Однако, априори, количество слагаемых может быть сколь угодно большим. Теорема Каратеодори утверждает, что для любой точки $x \in \operatorname{conv} E$ найдется выпуклая комбинация с не более, чем $d + 1$ слагаемым:

***Теорема* 1** (Каратеодори)**.** *Пусть $E \subset \mathbb{R}^d$ – произвольное множество. Тогда для любой точки $x \in \operatorname{conv} E$ найдутся набор из $d + 1$ точки $x_1, \dots, x_{d+1} \in E$ и числа $\lambda_i \geqslant 0$, $\sum_{i=1}^{d+1} \lambda_i = 1$, такие, что $x = \sum_{i=1}^{d+1} \lambda_i x_i$*

Теорема Каратеодори имеет следующую геометрическую интерпретацию (см. упражнение 1.10): если $E \subset \mathbb{R}^d$, то любая точка из выпуклой оболочки $\operatorname{conv} E$ лежит в некотором (невырожденном) симплексе с вершинами из E. Напомним, что n-мерный (невырожденный) симплекс – это выпуклая оболочка конечного набора из $n+1$ точек, не лежащих в одном аффинном подпространстве размерности $n-1$ или меньше (такой набор точек называется *аффинно независимым*). Например, выпуклая оболочка четырех точек в $\mathbb{R}^3$ является трехмерным симплексом (тетраэдром), если они не лежат в одной плоскости.

Доказательство теоремы 1. По определению выпуклой оболочки множества, любая точка $x \in \operatorname{conv} E$ может быть представлена как выпуклая комбинация точек из E. Пусть $x = \sum_{i=1}^{n} \lambda_i x_i$ – такое представление с минимально возможным числом слагаемых. Очевидно, $\lambda_i > 0$ – иначе какое-то слагаемое можно было бы выбросить.

Докажем от противного, что $n \leqslant d+1$. Пусть $n \geqslant d+2$. Рассмотрим разности $x_1 - x_n$, $x_2 - x_n$, ..., $x_{n-1} - x_n$. Эти векторы линейно зависимы, так как их не меньше $d+1$. Поэтому для некоторых не равных одновременно 0 чисел $a_1, \dots, a_{n-1}$ выполнено $\sum_{i=1}^{n-1} a_i (x_i - x_n) = 0$. Обозначая $a_n = -\sum_{i=1}^{n-1} a_i$, получаем:

$$\sum_{i=1}^{n} a_i x_i = 0 \quad \text{и} \quad \sum_{i=1}^{n} a_i = 0.$$

Домножим полученное представление 0 в виде линейной комбинации на $t > 0$ и вычтем из исходной выпуклой комбинации:

$$x = \sum_{i=1}^{n} (\lambda_i - t a_i) x_i \tag{1}$$

Здесь сумма коэффициентов равна 1, и поэтому линейная коминация (1) является аффинной. Более того, поскольку $\lambda_i > 0$ при всех i, то при малых t все коэффициенты положительны, и, следовательно, данная аффинная комбинация является выпуклой. Поскольку $\sum_{i=1}^{n} a_i = 0$, то некоторые числа a_i положительны. Поэтому при достаточно больших t некоторые коэффициенты в (1) становятся отрицательными. Возьмем минимальное t, при котором впервые обнулился какой-либо коэффициент в (1):

$$t = \min_{i: a_i > 0} \frac{\lambda_i}{a_i}.$$

Тогда при данном t один коэффициент в (1) обнулился, а остальные – неотрицательны. Выбросив нулевое слагаемое, мы получим выпуклую комбинацию с меньшим числом слагаемых, чем исходная. Противоречие.

Таким образом, в минимальной выпуклой комбинации $x = \sum_{i=1}^{n} \lambda_i x_i$, $x_i \in E$, число слагаемых не больше, чем $d+1$. Если число слагаемых оказалось меньше $d+1$, то мы дополним эту сумму слагаемыми вида $0 \cdot x_0$, где $x_0 \in E$ – любая точка.

□

Нижеследующее следствие из теоремы Каратеодори имеет огромное значение (возможно, даже большее, чем сама теорема):

Следствие 6. *Выпуклая оболочка компактного множества в $\mathbb{R}^d$ компактна.*

Доказательство. Пусть $E \subset \mathbb{R}^d$ – компакт и $x^n \in \operatorname{conv} E$ – произвольная последовательность. Докажем, что из нее можно выбрать сходящуюся. По теореме Каратеодори

$$x^n = \sum_{i=1}^{d+1} \lambda_i^n x_i^n.$$

Воспользуемся компактностью E и компактностью симплекса

$$\Delta = \{\lambda = (\lambda_1, \dots, \lambda_{d+1}) : \lambda_i \geqslant 0, \sum \lambda_i = 1\}.$$

Мы имеем $d+1$ последовательность x_i^n, $i = 1, \dots, d+1$, из E и одну последовательность $\lambda^n = (\lambda_1^n, \dots, \lambda_{d+1}^n)$ из Δ. По очереди выберем из них сходящиеся подпоследовательности, и будем без ограничения общности считать, что они сами сходятся. Тогда $x_i^n \to \hat{x}_i \in E$ и $\lambda_n \to \hat{\lambda} \in \Delta$. Поэтому

$$\lim_{n\to\infty} x_n = \lim_{n\to\infty} \sum_{i=1}^{d+1} \lambda_i^n x_i^n = \sum_{i=1}^{d+1} \hat{\lambda}_i \hat{x}_i \in \operatorname{conv} E,$$

что и требовалось. □

Упражнения:

1.1 Доказать, что если точки x_1, ..., x_{d+1} из $\mathbb{R}^d$ не лежат в одном аффинном пространстве размерности $d-1$ или меньше, то множество $\Delta = \operatorname{conv}\{x_1, \dots, x_{d+1}\}$ (называемое d-мерным симплексом) имеет непустую внутренность.

1.2 Пусть $C \neq \emptyset$ – ограниченное выпуклое подмножество $\mathbb{R}^d$. Доказать, что $0 \in \operatorname{ri} C$, если и только если $-\lambda C \subset C$ для некоторого $\lambda > 0$.

1.3 Привести пример такого невыпуклого множества E, что $\operatorname{int} E \neq \emptyset$ и $\operatorname{cl} \operatorname{int} E \neq E$.

1.4 Привести пример замкнутого множества, выпуклая оболочка которого не замкнута.

1.5 Верны ли включения $\operatorname{conv} \operatorname{cl} E \subset \operatorname{cl} \operatorname{conv} E$ и $\operatorname{cl} \operatorname{conv} E \subset \operatorname{conv} \operatorname{cl} E$?

1.6 Пусть C_1 и C_2 – выпуклые замкнутые множества. Докажите, что если $\operatorname{ri} C_1 \subset C_2$ и $\operatorname{ri} C_2 \subset C_1$, то $C_1 = C_2$.

1.7 Показать, что сумма по Минковскому замкнутого и компактного множеств – замкнута.

1.8 Показать, что сумма по Минковскому открытого множества и любого – открыта.

1.9 Показать, что сумма по Минковскому двух замкнутых выпуклых множеств может быть не замкнута.

1.10 Доказать, что если $E \subset \mathbb{R}^d$, то любая точка $x \in \operatorname{conv} E$ лежит в относительной внутренности некоторого (невырожденного) симплекса размерности $\leqslant d$ и с вершинами из E.

1.11 Пусть $A : \mathbb{R}^d \to \mathbb{R}^{d'}$ – линейное отображение и $C \subset \mathbb{R}^d$ – выпуклое множество. Докажите, что если для любого вектора $\xi \in \ker A$ и любой точки $x \in C$ пересечение $C \cap (x + \mathbb{R}\xi)$ либо ограничено, либо совпадает с $x + \mathbb{R}\xi$, то $\operatorname{cl}(AC) = A(\operatorname{cl} C)$.

2 Теоремы отделимости и теорема Каратеодори

Формулировка теорем отделимости. Без сомнений, самым главным результатом выпуклого анализа является теорема отделимости, которая, по сути, утверждает, что любые два непересекающихся выпуклых множества C_1 и C_2 можно разделить гиперплоскостью (размерности $d-1$) так, что множества окажутся по разные стороны от нее (см. рис. 4). Другими словами, если взять ковектор $p \in \mathbb{R}^{d*}$, ортогональный разделяющей гиперплоскости, то она запишется в виде $\{x \in \mathbb{R}^n : \langle p, x \rangle = c\}$ при подходящем выборе константы c. Для множеств C_1 и C_2 мы получим

$$\sup_{x \in C_1} \langle p, x \rangle \leqslant c \leqslant \inf_{x \in C_2} \langle p, x \rangle,$$

поскольку множества C_1 и C_2 лежат по разные стороны от этой гиперплоскости.

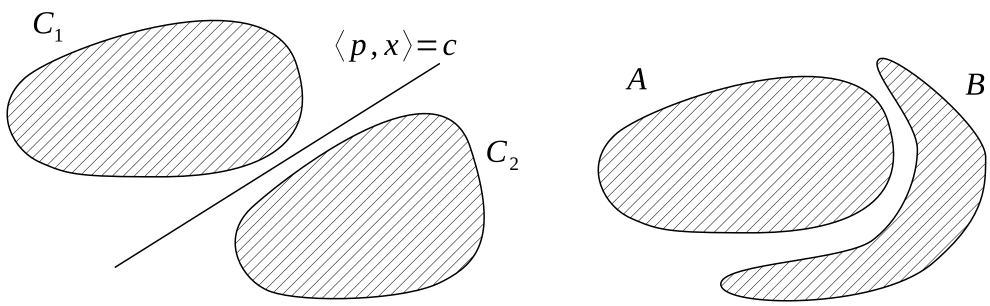

Рис. 4: Геометрический смысл теоремы отделимости.

Мы будем говорить, что гиперплоскость *строго отделяет* множества C_1 и C_2, если предыдущее неравенство является строгим. В этом случае между множествами можно поместить не одну гиперплоскость, а целую полосу, зажатую между параллельными гиперплоскостями.

Мы будем говорить, что гиперплоскость *собственно отделяет* множества C_1 и C_2, если найдутся точки $x_1 \in C_1$ и $x_2 \in C_2$ такие, что $\langle p, x_1 \rangle \neq \langle p, x_2 \rangle$. Собственная отделимость исключает вырожденный случай, когда оба выпуклых множества имеют размерность меньше d и содержатся в одной гиперплоскости (про которую, формально, можно сказать, что она их разделяет).

Возможность разделять выпуклые множества, по своему существу, есть геометрический результат. Его удобно формулировать двумя не совсем эквивалентными способами. Для удобства формулировок договоримся считать

$$\sup \emptyset = -\infty \quad \text{и} \quad \inf \emptyset = +\infty.$$

Теорема 2 (первая теорема отделимости)**.** *Пусть C_1 и C_2 – выпуклые подмножества $\mathbb{R}^d$. Тогда если их относительные внутренности не пересекаются,*

$$\operatorname{ri} C_1 \cap \operatorname{ri} C_2 = \emptyset,$$

то их можно собственно отделить: найдется такой ненулевой ковектор $p \in \mathbb{R}^{d*}$, что

$$\sup_{x \in C_1} \langle p, x \rangle \leqslant \inf_{x \in C_2} \langle p, x \rangle$$

и $\langle p, x_1\rangle \neq \langle p, x_2\rangle$ для некоторых $x_1 \in C_1$ и $x_2 \in C_2$.

Теорема 3 (вторая теорема отделимости)**.** *Пусть C_1 и C_2 – выпуклые подмножества $\mathbb{R}^d$, причем одно из них замкнуто, а другое – компактно и непусто. Тогда если C_1 и C_2 не пересекаются, $C_1 \cap C_2 = \emptyset$, то их можно строго отделить: найдется такой ненулевой ковектор $p \in \mathbb{R}^{d*}$, что*

$$\sup_{x\in C_1} \langle p, x\rangle < \inf_{x\in C_2} \langle p, x\rangle.$$

Заметим, что если множества C_1 и C_2 удовлетворяют условиям второй теоремы отделимости, то они заведомо удовлетворяют условиям и первой теоремы отделимости. Поэтому первая теорема отделимости применима в большем количестве ситуаций, чем вторая. Более того, первая теорема отделимости сформулирована, в каком-то смысле, в максимальном виде, а именно: если относительные внутренности множеств пересекаются, $\operatorname{ri} C_1 \cap \operatorname{ri} C_2 \neq \emptyset$, то формально «разделяющая» гиперплоскость должна содержать оба множества C_1 и C_2 (см. упражнение 2.5). Такая гиперплоскость может существовать только в вырожденном случае $\dim \operatorname{aff}(C_1 + C_2) < d$. С другой стороны, результат второй теоремы отделимости сильнее: если выполнены ее условия, то между множествами C_1 и C_2 можно поместить полосу, зажатую между двумя параллельными гиперплоскостями (что, вообще говоря, неверно в условиях первой теоремы отделимости).

Доказательство второй теоремы отделимости. Доказательство этих теорем мы проведем следующим образом: мы докажем сначала вторую теорему отделимости, выделив в ней главную геометрическую идею, а затем получим первую теорему, опираясь на результат второй и на теорему Каратеодори.

Доказательство второй теоремы отделимости. Без ограничения общности будем считать, что множество C_1 компактно, а C_2 – замкнуто. Также будем считать, что $C_2 \neq \emptyset$ (иначе заявленное в условии теоремы неравенство тривиально выполнено для любого $p \neq 0$).

Начнем с главного случая, когда C_1 – одноточечное множество, $C_1 = \{x_0\}$ и $x_0 \notin C_2$. Зафиксируем на $\mathbb{R}^d$ произвольное скалярное произведение $\langle\cdot,\cdot\rangle$ и выберем в C_2 точку x_1, ближайшую к x_0. Такая точка существует в силу замкнутости C_2. Очевидно, $x_0 \neq x_1$. Далее, проведем через любую точку x_2 внутри отрезка $[x_0; x_1]$ ортогональную ему гиперплоскость H (см. рис. 5).

Мы утверждаем, что H строго отделяет x_0 от C_2, то есть они лежат в разных открытых полупространствах относительно H. Докажем это от противного. Если в C_2 есть некоторая точка y_0 из того же полупространства, что и x_0, то отрезок $[x_1; y_0]$ лежит в C_2, так как C_2 выпукло. Найдем на этом отрезке точку, которая находится ближе к x_0, чем x_1 (что и приведет к противоречию). Отрезок $[x_1; y_0]$ пересекает H в некоторой точке $y_1 \in C_2$. При этом y_1 не может совпадать с x_2, так как x_2 ближе к x_0, чем x_1, что невозможно. Таким образом, точки x_1, x_2 и y_1 образуют прямоугольный треугольник с прямым углом при x_2. Обозначим через y_2 основание высоты, проведенной из прямого угла x_2 на гипотенузу x_1y_1. Очевидно, $y_2 \in C_2$. При этом $\operatorname{dist}(x_2, y_2) < \operatorname{dist}(x_1, x_2)$, так как отрезки $[y_2; x_2]$ и $[x_1; x_2]$ являются катетом и гипотенузой в прямоугольном треугольнике $\Delta x_1x_2y_2$. Воспользуемся неравенством треугольника:

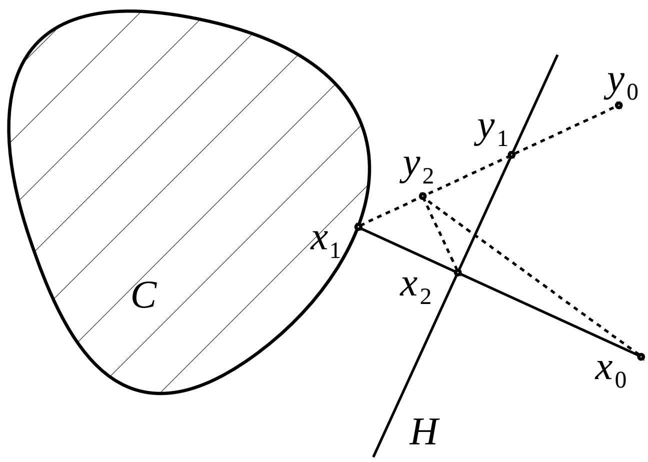

Рис. 5: Строгое отделение точки x_0 от замкнутого множества C.

$$\mathrm{dist}(x_0, y_2) \leqslant \mathrm{dist}(x_0, x_2) + \mathrm{dist}(x_2, y_2) < \mathrm{dist}(x_0, x_2) + \mathrm{dist}(x_1, x_2) = \mathrm{dist}(x_0, x_1).$$

Итак, мы пришли к противоречию, так как y_2 ближе к x_0, чем x_1, но при этом x_1 является ближайшей точкой к x_0 в C_2.

Таким образом, гиперплоскость H строго разделяет x_0 и C_2. В качестве ковектора p, заявленного в условии теоремы, возьмем любой ненулевой ковектор, ортогональный H.

Разберем теперь общий случай, когда C_1 – произвольное компактное множество. Составим разность (по Минковскому) множеств C_1 и C_2:

$$C = C_1 - C_2 = \{x - y, \quad x \in C_1,\ y \in C_2\}.$$

Условие $C_1 \cap C_2 = \emptyset$ равносильно условию $0 \notin C$. Осталось заметить, что множество C выпукло, так как выпуклы множества C_1 и C_2, и замкнуто, так как C_1 замкнуто, а C_2 компактно (см. упражнение 1.7). Применение к C уже полученного результата об отделении замкнутого выпуклого множества от 0 завершает доказательство. □

Компактно-выпуклые исчерпания. Напомним, что *исчерпанием* множества A называется счетное число подмножеств $B_1 \subset B_2 \subset \ldots \subset A$, таких, что $A = \bigcup_{n=1}^{\infty} B_n$. Для доказательства первой теоремы отделимости нам потребуется исчерпание открытого выпуклого множества компактными выпуклыми подмножествами.

Следствие 7. *Пусть $C \subset \mathbb{R}^d$ – открытое выпуклое множество. Тогда существует его исчерпание $K_1 \subset K_2 \subset \ldots C$ компактными выпуклыми подмножествами.*

Доказательство. Поскольку любое открытое множество в $\mathbb{R}^d$ имеет компактное исчерпание, то, пользуясь теоремой Каратеодори 1 (точнее, ее следствием 6), легко построить компактно-выпуклое исчерпание – достаточно просто овыпуклить любое компактное исчерпание.

Более того, компактно-выпуклое исчерпание любого открытого выпуклого множества можно построить и явно. Например,

$$K_n = \{x \in C : \operatorname{dist}(x, \partial C) \geqslant \frac{1}{n} \text{ и } |x| \leqslant n\}.$$

Множества K_n, очевидно, замкнуты и ограничены, а, значит, компактны. Осталось заметить, что K_n есть пересечение шара $B_n = \{x \in \mathbb{R}^n : |x| \leqslant n\}$ и множества $L_n = \{x \in C : \operatorname{dist}(x, \partial C) \geqslant \frac{1}{n}\}$. Шар B_n, очевидно, выпукл. Множество L_n тоже выпукло. Действительно, если $x_0, x_1 \in L_n$, то окрестности $U_{\frac{1}{n}}(x_0)$ и $U_{\frac{1}{n}}(x_1)$ лежат в C. Поскольку множество C выпукло, то в C лежит и их выпуклая оболочка $\cup_{\lambda \in [0;1]} U_{\frac{1}{n}}(x_\lambda)$, т.е. $x_\lambda \in L_n$. Поэтому выпукло их пересечение $B_n \cap L_n = K_n$. □

Доказательство первой теоремы отделимости. Теперь у нас все готово, чтобы получить

Доказательство первой теоремы отделимости. Если хотя бы одно из двух множеств пусто, то заявленное в условии теоремы неравенство тривиально выполняется для любого $p \neq 0$. Поэтому будем считать, что $C_1 \neq \emptyset$ и $C_2 \neq \emptyset$.

Разберем сначала главный случай, когда множество C_1 является одноточечным, $C_1 = \{0\}$ и $0 \notin C_2$, а выпуклое множество C_2 – открыто. Пусть $K_1 \subset K_2 \subset K_3 \subset \ldots \subset C_2$ – исчерпание C_2 компактными выпуклыми множествами. Тогда $0 \notin K_n$, и по доказанной второй теореме отделимости, существует такой ковектор $p_n \neq 0$, что $\inf_{x \in K_n} \langle p_n, x\rangle > 0$.

Без ограничения общности $|p_n| = 1$. Тогда все ковекторы p_n лежат на единичной сфере $\{|p| = 1\}$ (которая компактна в $\mathbb{R}^{d*}$), и, выбирая, если потребуется, сходящуюся подпоследовательность, будем считать, что последовательность p_n имеет предел p_0. Очевидно, $p_0 \neq 0$, так как $|p_0| = 1$. Покажем, что ковектор p_0 – искомый. Действительно, для любой точки $x \in C_2$ найдется такой номер N, что $x \in K_n$ для любого $n > N$. Поэтому $\langle p_n, x\rangle > 0$ для любого $n > N$. Переходя к пределу, получаем:

$$\langle p_0, x\rangle = \lim_{n\to\infty} \langle p_n, x\rangle \geqslant 0.$$

Таким образом, ковектор p_0 отделяет 0 от C_2. Поскольку $\operatorname{int} C_2 \neq \emptyset$, то $\langle p_0, C_2\rangle \neq \mathrm{const}$, и, следовательно, ковектор p_0 отделяет 0 и C_2 собственно.

Разберем теперь общий случай $\operatorname{ri} C_1 \cap \operatorname{ri} C_2 = \emptyset$. Покажем, что выпуклые множества $\operatorname{ri} C_1$ и $\operatorname{ri} C_2$ можно (собственно) отделить гиперплоскостью – этого достаточно, так как $C_1 \subset \operatorname{cl}\operatorname{ri} C_1$ и $C_2 \subset \operatorname{cl}\operatorname{ri} C_2$ по предложению 3. Рассмотрим разность $C = \operatorname{ri} C_1 - \operatorname{ri} C_2$. По условию, $0 \notin C$. Для того, чтобы отделить C_1 от C_2, достаточно отделить 0 от C. Выпуклое множество C является относительно открытым как разность относительно открытых множеств (предложение 6). Разберем все возможные случаи.

- Если $\operatorname{aff} C = \mathbb{R}^d$, то множество C открыто и 0 можно отделить от C по уже доказанному.

- Если аффинная оболочка $\operatorname{aff} C$ меньше, чем $\mathbb{R}^d$ (т.е. $\dim C < d$), то возможны два простых случая:

  – Если $0 \notin \operatorname{aff} C$, то в качестве (собственно) разделяющей гиперплоскости подойдет любая гиперплоскость, содержащая $\operatorname{aff} C$ и не содержащая 0.

– Если $0 \in \operatorname{aff} C$, то по уже доказанному 0 можно собственно отделить от C в $\operatorname{aff} C$ аффинным подпространством $L \subset \operatorname{aff} C$ размерности $\dim \operatorname{aff} C - 1$. В этом случае в качестве (собственно) разделяющей гиперплоскости в $\mathbb{R}^d$ подойдет любая гиперплоскость, содержащая L и не содержащая $\operatorname{aff} C$.

□

Следствие 8. *Пусть C – выпуклое подмножество $\mathbb{R}^d$. Если точка x_0 не лежит во внутренности $\operatorname{int} C$, то x_0 и C можно отделить.*

Доказательство. Если $\dim \operatorname{aff} C = d$, то $\operatorname{int} C \neq \emptyset$ и разделяющая гиперплоскость существует по первой теореме отделимости. Если же $\dim \operatorname{aff} C < d$, то в качестве разделяющей гиперплоскости подойдет любая, содержащая $\operatorname{aff} C$. □

Упражнения:

2.1 Докажите, что если множество $C \subset \mathbb{R}^d$ выпукло, то $\operatorname{ri} C = \operatorname{ri} \operatorname{cl} C$.

2.2 Докажите, что если множество $C \subset \mathbb{R}^d$ выпукло, то его граница совпадает с границей его замыкания: $\partial C = \partial \operatorname{cl} C$. Приведите контрпример для случая, когда множество не выпукло.

2.3 Пусть $C \subset \mathbb{R}^d$ - выпуклое множество. Докажите, что если $x_0 \in \operatorname{ri} C$ можно отделить от C гиперплоскостью, то только несобственным образом. Равносильно: разделяющая гиперплоскость содержит $\operatorname{aff} C$. Равносильно: разделяющий ковектор p перпендикулярен $\operatorname{aff} C$. Равносильно: $C \subset \{x \in \mathbb{R}^d : \langle p, x\rangle = c_0\}$.

2.4 Докажите (используя теоремы отделимости), что если C_1 и C_2 – выпуклые подмножества $\mathbb{R}^d$, то $\operatorname{ri}(C_1 + C_2) = \operatorname{ri} C_1 + \operatorname{ri} C_2$. Попробуйте также доказать это утверждение, не используя теорему отделимости.

2.5 Пусть C_1 и C_2 – выпуклые множества в $\mathbb{R}^d$. Доказать, что если $\operatorname{ri} C_1 \cap \operatorname{ri} C_2 \neq \emptyset$, то C_1 и C_2 нельзя собственно разделить, а если дополнительно $\dim(C_1 + C_2) = d$, то их нельзя разделить даже несобственным образом.

3 Простейшие свойства выпуклых функций

Соглашения о ∞. При работе с выпуклыми функциями удобно использовать расширенную числовую прямую $\bar{\mathbb{R}} = \mathbb{R} \cup \{\infty, -\infty\}$. Операции сложения, вычитания, умножения и сравнения определены естественным образом (деление на бесконечность мы никогда использовать не будем). Неочевидными являются только выражения типа $\infty - \infty$ или $0 \cdot \infty$. Мы будем считать разность $\infty - \infty$ некорректной, а произведение $0 \cdot \infty$ считать равным 0:

$$\text{выражение } \infty - \infty \text{ некорректно}; \qquad 0 \cdot \infty \stackrel{\text{def}}{=} 0.$$

Также мы будем придерживаться соглашений, уже предложенных в параграфе 2:

$$\inf \emptyset = \infty; \qquad \sup \emptyset = -\infty.$$

Надграфик и эффективная область. *Надграфиком* функции $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ называется множество $\mathrm{epi}f \subset \mathbb{R}^{d+1}$ точек, лежащих над графиком функции f:

$$\mathrm{epi}f = \{(x, a) : x \in \mathbb{R}^d,\ a \in \mathbb{R} \text{ и } f(x) \leqslant a\}.$$

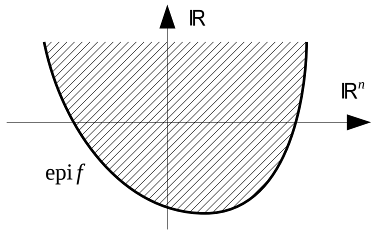

Рис. 6: Выпуклая функция

Также надграфик функции иногда называют *эпиграфом*, откуда и идет обозначение $\mathrm{epi}f$.

Определение 5. Функция $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ называется выпуклой, если ее надграфик $\mathrm{epi}f$ является выпуклым множеством (см. рис. 6).

Таким образом, мы всегда предполагаем, что выпуклая функция определена в каждой точке пространства $\mathbb{R}^d$, но, возможно, принимает в некоторых точках бесконечные значения. В пространстве $\mathbb{R}^{d+1}$ направление, отвечающее значениям функций, является выделенным и мы будем называть его вертикальным. Исходное пространство $\mathbb{R}^d$ мы будем отождествлять с горизонтальным подпространством в $\mathbb{R}^{d+1}$ вида $\{(x, 0), x \in \mathbb{R}^d\}$.

Определение 6. *Эффективной областью* выпуклой функции f называется множество $\mathrm{dom}f \subset \mathbb{R}^d$, в точках которого функция f принимает значения, меньшие ∞:

$$\mathrm{dom}f = \{x \in \mathbb{R}^d : f(x) < \infty\}.$$

Из определения легко видеть, что множество $\mathrm{dom}f$ всегда выпукло – действительно, множество $\mathrm{dom}f$ является проекцией множества $\mathrm{epi}f$, а линейные отображения, очевидно, сохраняют выпуклость. Иногда выпуклые функции задают по-другому: считают, что функция f выпукла, если она определена и конечна на некотором выпуклом подмножестве $C \subset \mathbb{R}^d$ и ее надграфик на C является выпуклым. Перейти к нашим обозначениям в этом случае очень просто: достаточно просто доопределить выпуклую функцию значением ∞ вне множества C.

Неравенство Йенсена. Определение выпуклой функции можно формулировать множеством эквивалентных способов:

Предложение 7. *Пусть $f : \mathbb{R}^d \to \bar{\mathbb{R}}$. Тогда следующие утверждения эквивалентны:*

0. *Функция f выпукла.*

1. *Для любой прямой l в $\mathbb{R}^d$ ограничение f на l является выпуклой функцией.*

2. *Для любых двух точек* $x_0, x_1 \in \operatorname{dom} f$ *и любого числа* $\lambda \in [0;1]$ *выполнено*

$$f(x_\lambda) \leqslant (1-\lambda)f(x_0) + \lambda f(x_1), \quad \text{где} \quad x_\lambda = (1-\lambda)x_0 + \lambda x_1.$$

3. *Для любых двух точек* $x_0 \in \operatorname{dom} f$ *и* x_1, $f(x_1) \neq -\infty$, *и любого числа* $\lambda \geqslant 1$ *выполнено*

$$f(x_\lambda) \geqslant (1-\lambda)f(x_0) + \lambda f(x_1).$$

3′. *Для любых двух точек* x_0, $f(x_0) \neq -\infty$, *и* $x_1 \in \operatorname{dom} f$ *и любого числа* $\lambda \leqslant 0$ *выполнено*

$$f(x_\lambda) \geqslant (1-\lambda)f(x_0) + \lambda f(x_1).$$

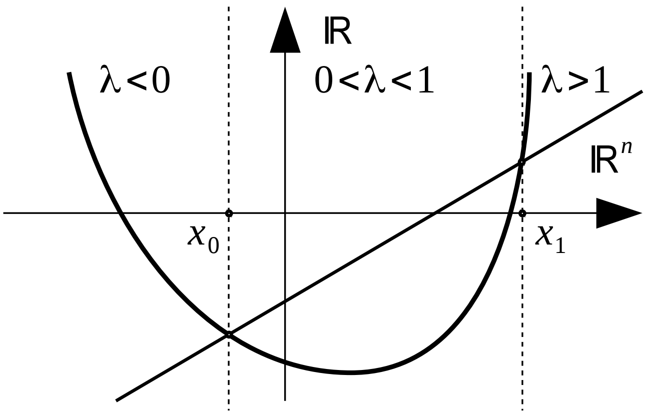

Рис. 7: Иллюстрация к предложению 7.

Геометрический смысл пункта 2 заключается в том, что значения выпуклой функции в точках отрезка $[x_0; x_1]$ лежат под секущей прямой, проходящей через точки $(x_0, f(x_0))$ и $(x_1, f(x_1))$ (см. рис. 7). Геометрический смысл утверждений 3 и 3′ заключается в том, что значения выпуклой функции в точках прямой, проходящей через x_0 и x_1, вне отрезка $[x_0; x_1]$ лежат над секущей (см. рис. 7).

Доказательство 7. Сначала убедимся в том, что все написанные выражения в пунктах 2, 3 и 3′ корректны. В пункте 2 это так, поскольку $\lambda \geqslant 0$ и $(1-\lambda) \geqslant 0$, и, значит, выражение в правой части может быть некорректно, только если $f(x_0) = \infty$ и $f(x_1) = -\infty$ или наоборот. Однако, $x_0, x_1 \in \operatorname{dom} f$ по условию пункта 2, и, значит, $f(x_0) \neq \infty$ и $f(x_1) \neq \infty$. Аналогично проверяется корректность выражений в пунктах 3 и 3′.

(0 $\Rightarrow$ 1) Надграфиком ограничения f на прямую l является пересечение $\operatorname{epi} f$ с двумерной вертикальной плоскостью, проходящей через l. Поскольку и плоскость, и $\operatorname{epi} f$ являются выпуклыми множествами, то таковым является и их пересечение.

(1 $\Rightarrow$ 2) Проведем через точки x_0 и x_1 прямую l и обозначим через L вертикальную двумерную плоскость, проходящую через l. По условию п. 1 множество $\operatorname{epi} f \cap L$ выпукло. Пусть $a_0 > f(x_0)$ и $a_1 > f(x_1)$, $a_0, a_1 \in \mathbb{R}$ (напомним $f(x_{0,1}) \neq \infty$). Поскольку $(x_0, a_0) \in \operatorname{epi} f \cap L$ и $(x_1, a_1) \in \operatorname{epi} f \cap L$, то $(x_\lambda, a_\lambda) \in \operatorname{epi} f \cap L$. Поэтому $f(x_\lambda) \leqslant (1-\lambda)a_0 + \lambda a_1$. Устремляя $a_0 \to f(x_0)$ и $a_1 \to f(x_1)$, получаем искомое.

(2 ⇒ 0) Предположим, выполняется условие п. 2. Покажем, что надграфик выпукл. Пусть (x_0, a_0) и (x_1, a_1) лежат в $\mathrm{epi} f$. Тогда x_0 и x_1 лежат в $\mathrm{dom} f$. Поэтому

$$a_\lambda \geqslant (1-\lambda) f(x_0) + \lambda f(x_1) \geqslant f(x_\lambda),$$

т.е. $(x_\lambda, a_\lambda) \in \mathrm{epi} f$.

(2 ⇒ 3) Если $x_\lambda \notin \mathrm{dom} f$, то п. 3 выполняется автоматически. Пусть $x_\lambda \in \mathrm{dom} f$. Воспользуемся тем, что $x_1 = (1-\frac{1}{\lambda})x_0 + \frac{1}{\lambda}x_\lambda$, и положим $y_0 = x_0$, $y_1 = x_\lambda$ и $\mu = \frac{1}{\lambda}$. Тогда $y_\mu = x_1$ и $f(x_1) \leqslant (1-\frac{1}{\lambda})f(x_0) + \frac{1}{\lambda}f(x_\lambda)$. Домножая полученное неравенство на λ и собирая слагаемые с нужных сторон, получаем искомое.

(3 ⇒ 2) Действуем абсолютно аналогично: если $f(x_\lambda) = -\infty$, то нужное неравенство выполнено автоматически. В противном случае воспользуемся п. 3 для $y_0 = x_0$, $y_1 = x_\lambda$ и $y_\mu = x_1$.

(3 ⇔ 3′) Утверждения 3 и 3′ получаются друг из друга с помощью замены $y_0 = x_1$, $y_1 = x_0$ и $\mu = 1-\lambda$.

□

Также очень удобным является следующее обобщение предложения 7:

***Предложение* 8** (Неравенство Йенсена)**.** *Пусть f – выпуклая функция. Тогда если $x_1, \ldots, x_n \in \mathrm{dom} f$ – любые точки, а $\lambda_1, \ldots, \lambda_n$ – неотрицательные числа и $\sum_{i=1}^n \lambda_i = 1$, то*

$$f(\sum_{i=1}^n \lambda_i x_i) \leqslant \sum_{i=1}^n \lambda_i f(x_i).$$

Доказательство. Поскольку $(x_i, a_i) \in \mathrm{epi} f$ для любых $a_i > f(x_i)$, то $(\sum_i \lambda_i x_i, \sum_i \lambda_i a_i) \in \mathrm{epi} f$ по предложению 1. Поэтому $f(x) \leqslant \sum_i \lambda_i a_i$ по определению надграфика. В силу произвольности выбора $a_i > f(x_i)$ получаем искомое. □

Собственные функции. Оказывается, что если выпуклая функция принимает значение $-\infty$ хоть в одной точке, то она устроена относительно просто:

***Следствие* 9.** *Если f – выпуклая функция и $f(x_0) = -\infty$ для некоторого $x_0 \in \mathbb{R}^d$, то $f(x) = -\infty$ для всех $x \in \mathrm{ri}\,\mathrm{dom} f$.*

Доказательство. Пусть $x \in \mathrm{ri}\,\mathrm{dom} f$. Поскольку $x_0 \in \mathrm{dom} f$, то найдется $x_1 \in \mathrm{dom} f$, $x_1 \neq x$, что $x \in [x_0; x_1]$. То есть $x = x_\lambda$ для некоторого $\lambda \in [0; 1)$. Применяя п.2 предложения 7, получаем искомое. □

Однако, вне $\mathrm{ri}\,\mathrm{dom} f$ такая выпуклая функция может быть устроена достаточно плохо.

Пример 1. Рассмотрим функцию, равную $-\infty$ во внутренних точках единичного круга на плоскости и равную $+\infty$ вне этого круга. Тогда в граничных точках круга значения функции можно задать совершенно произвольным образом – функция в любом случае получится выпуклой. Связан этот феномен с незамкнутостью надграфика (см. также упражнение 3.1).

Так или иначе, наибольший интерес представляют выпуклые функции, не принимающие значение $-\infty$ ни в одной точке:

Определение 7. Функция $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ называется *собственной*, если $\mathrm{dom} f \neq \emptyset$ и $f(x) \neq -\infty$ при любом $x \in \mathbb{R}^d$.

Использование собственных функций очень удобно – не надо беспокоиться о том, что какое-то выражение может оказаться некорректным. Например, если функция f собственная, то в неравенстве Йенсена можно не проверять, что $x_i \in \mathrm{dom} f$:

***Предложение* 9** (Неравенство Йенсена для собственных функций)**.** *Если функция f – выпуклая и собственная, то неравенство Йенсена выполняется для любых выпуклых комбинаций $\sum_i \lambda_i x_i$ точек $x_i \in \mathbb{R}^d$.*

Доказательство. Будем без ограничения общности считать, что $\lambda_i \neq 0$ для всех i. Если $x_i \in \mathrm{dom} f$ для всех i, то нужное неравенство мы уже доказали в предложении 8. Если же $x_i \notin \mathrm{dom} f$ для некоторого i, то правая часть неравенства определена корректно и равна $+\infty$. □

***Теорема* 4.** *Пусть f – выпуклая функция на $\mathbb{R}^d$. Предположим, что функция f ограничена на* $\mathrm{int}\,\mathrm{dom} f$, $|f(x)| \leqslant M$ *для* $x \in \mathrm{int}\,\mathrm{dom} f$. *Тогда для любых* $x_0, x_1 \in \mathrm{int}\,\mathrm{dom} f$ *выполнено*

$$|f(x_1) - f(x_0)| \leqslant \frac{2M}{\rho}|x_1 - x_0|,$$

где $\rho > 0$ – минимум из расстояний от точек x_0 и x_1 до границы $\mathrm{dom} f$.

Доказательство. Рассмотрим точки x_λ при $\lambda < 0$. Очевидно, $\mathrm{dist}(x_\lambda, x_0) = |\lambda||x_1 - x_0|$, и, значит, при $|\lambda| < \frac{\rho}{|x_1 - x_0|}$ точка x_λ лежит в $\mathrm{int}\,\mathrm{dom} f$. Далее, $f(x_\lambda) \geqslant (1-\lambda)f(x_0) + \lambda f(x_1)$ по предложению 7. Перенося $f(x_0)$ налево, деля на $-\lambda > 0$ и устремляя $\lambda \to -\frac{\rho}{|x_1 - x_0|}$, получаем

$$f(x_0) - f(x_1) \leqslant \frac{f(x_\lambda) - f(x_0)}{-\lambda} \leqslant \frac{2M}{-\lambda} \to \frac{2M}{\rho}|x_1 - x_0|.$$

Обратное неравенство получается аналогично. □

***Лемма* 1.** *Собственная выпуклая функция на $\mathbb{R}^d$ ограничена в окрестности любой внутренней точки эффективной области.*

Доказательство. Пусть $x_0 \in \mathrm{int}\,\mathrm{dom} f$. Доказать ограниченность сверху функции f в окрестности x_0 просто. Для этого выберем любой симплекс $\Delta = \mathrm{conv}\{x_1, \dots, x_{d+1}\}$ с вершинами $x_1, \dots, x_{d+1}$ из $\mathrm{dom} f$ так, что $x_0 \in \mathrm{int}\,\Delta$. Если $M = \max_{i=1,\dots,d+1} f(x_i)$, то по неравенству Йенсена $f(x) \leqslant \sum_{i=1}^{d+1} \lambda_i f(x_i) \leqslant M \sum_{i=1}^{d+1} \lambda_i = M$ для любой точки $x \in \Delta$.

Пусть теперь B – (замкнутый) шар с центром в x_0, на котором функция f ограничена сверху. Докажем ограниченность снизу f на B от противного: предположим, найдется такая последовательность $x_n \in B$, что $f(x_n) \to -\infty$. Рассмотрим луч l_n, выходящий из x_n и проходящий через центр x_0. Он пересекает сферу ∂B в точке $y_n = l_n \cap \partial B$. Поскольку $\mathrm{dist}(y_n, x_0) \geqslant \mathrm{dist}(x_n, x_0)$, то $y_n = (1-\lambda)x_0 + \lambda x_n$ для $\lambda \leqslant -1$. Без ограничения общности $f(x_n) - f(x_0) < 0$, поэтому (по предложению 7)

$$f(y_n) \geqslant f(x_0) + \lambda(f(x_n) - f(x_0)) \geqslant f(x_0) - (f(x_n) - f(x_0)) = 2f(x_0) - f(x_n).$$

Правая часть стремится к $+\infty$ при $n \to \infty$, что невозможно, так как $y_n \in B$, а функция f на B ограничена сверху. □

Из ограниченности выпуклой функции в окрестности внутренних точек $\operatorname{dom} f$ и из теоремы 4 вытекает

Теорема 5. *Собственная выпуклая функция f на $\mathbb{R}^d$ является локально липшицевой (и, значит, непрерывной) на $\operatorname{int}\operatorname{dom} f$.*

Напомним, что функция называется *локально липшицевой* на некотором множестве, если у любой точки из этого множества найдется окрестность, в которой функция будет липшицевой. При этом константа Липшица может отличаться для разных окрестностей.

Доказательство теоремы 5. Пусть $x_0 \in \operatorname{int}\operatorname{dom} f$ и $B_\varepsilon(x_0)$ – шар с центром в x_0 и радиуса ε, на котором функция f ограничена константой $M \geqslant 0$. Введем новую функцию $g(x)$, совпадающую с $f(x)$, если $\operatorname{dist}(x, x_0) \leqslant \varepsilon$, и равную $+\infty$, если $\operatorname{dist}(x, x_0) > \varepsilon$. Эта функция является выпуклой, так как ее надграфик есть пересечение двух выпуклых множеств $\operatorname{epi} f$ и $\{(x,a) | \, a \in \mathbb{R}, \operatorname{dist}(x,x_0) \leqslant \varepsilon\}$. Поэтому по теореме 4 функция g является липшицевой на $\operatorname{dom} g = B_{\frac{\varepsilon}{2}}(x_0)$ с константой $4M/\varepsilon$. Такой же является и функция f на $B_{\frac{\varepsilon}{2}}(x_0)$. □

Следствие 10. *Пусть f – выпуклая собственная функция. Тогда $f \in C(\operatorname{ri}\operatorname{dom} f)$.*

Внутренность надграфика. Надграфик собственных выпуклых функций обладает (помимо выпуклости) рядом удобных свойств. Одно из них связано со структурой его относительной внутренности.

Следствие 11. *Пусть f – выпуклая функция на $\mathbb{R}^d$. Тогда*

$$\operatorname{ri}\operatorname{epi} f = \{(x,a) : x \in \operatorname{ri}\operatorname{dom} f \quad \text{и} \quad a > f(x)\}.$$

Доказательство. Обозначим через A множество в правой части утверждаемого равенства. Тогда требуется доказать, что $A \subset \operatorname{ri}\operatorname{dom} f$ и $\operatorname{ri}\operatorname{dom} f \subset A$. Будем считать, что $\operatorname{dom} f \neq \emptyset$ (иначе все тривиально).

Без ограничения общности $\operatorname{aff}\operatorname{dom} f = \mathbb{R}^d$, то есть $\operatorname{int}\operatorname{dom} f \neq \emptyset$. Сначала докажем, что $\operatorname{int}\operatorname{epi} f \subset A$. Действительно, если точка $(x,a) \in \operatorname{epi} f$ не лежит в A, значит, $a = f(x)$ или $x \in \partial \operatorname{dom} f$. В любом из этих случаев точка (x,a), очевидно, не может лежать в $\operatorname{int}\operatorname{epi} f$.

Теперь в обратную сторону: $A \subset \operatorname{int}\operatorname{epi} f$. Если f – несобственная функция, то $f|_{\operatorname{int}\operatorname{dom} f} \equiv -\infty$ по следствию 9. В этом случае множество $A = \operatorname{int}\operatorname{dom} f \times \mathbb{R}$ открыто. Поскольку $A \subset \operatorname{epi} f$ (всегда), то $A \subset \operatorname{int}\operatorname{epi} f$. Пусть теперь f – собственная функция и $(x_0, a_0) \in A$. Поскольку $x_0 \in \operatorname{int}\operatorname{dom} f$, то функция f непрерывна в окрестности x_0. Поскольку $f(x_0) < a_0$, то для малого $\varepsilon > 0$ в малой окрестности U точки x_0 выполнено $f(x) < a_0 - \varepsilon$, $x \in U$. Поэтому точки (x,a), где $x \in U$ и $a > a_0 - \varepsilon$, лежат в $\operatorname{epi} f$, и, значит, точка (x_0, a_0) является внутренней точкой $\operatorname{epi} f$. □

Замкнутые функции. Еще одно важное свойство выпуклой функции – замкнутость ее надграфика. Очевидно, не все выпуклые функции обладают замкнутым надграфиком. Поэтому введем

Определение 8. Функция $f : \mathbb{R}^d \to \bar{f}$ называется *замкнутой*, если ее надграфик $\operatorname{epi} f$ является замкнутым множеством.

***Предложение* 10.** *Функция f является замкнутой, если и только если она полунепрерывна снизу в любой точке $x_0 \in \mathbb{R}^d$:*

$$\liminf_{k\to\infty} f(x_k) \geqslant f(x_0) \quad \text{для любой последовательности} \quad x_k \to x_0.$$

Доказательство. (Необходимость). Если $f(x_k) < \infty$ лишь для конечного набора индексов k, то нижний предел равен ∞ и неравенство выполнено автоматически. В противном случае выберем из последовательности x_k такую подпоследовательность, что $f(x_k)$ сходится к нижнему пределу. Сохраняя нумерацию, будем считать, что существует $\lim_{k\to\infty} f(x_k) = a_0$. Поскольку точки $(x_k, f(x_k))$ лежат в надграфике $\mathrm{epi} f$, который замкнут, то $(x_0, a_0) \in \mathrm{epi} f$, то есть $f(x_0) \leqslant a_0$, что и требовалось.

(Достаточность). Пусть $(x_k, a_k) \in \mathrm{epi} f$, и $x_k \to x_0$ и $a_k \to a_0$ при $k \to \infty$. Докажем, что $(x_0, a_0) \in \mathrm{epi} f$. Действительно, $f(x_k) \leqslant a_k$, поэтому $\liminf_{k\to\infty} f(x_k) \leqslant a_0$. Значит, $f(x_0) \leqslant a_0$, что и требовалось. □

Отметим, что эффективная область $\mathrm{dom} f$ замкнутой функции f может быть незамкнутой, например, это так для функции $f(x) = 1/x$ при $x > 0$ и $f(x) = +\infty$ при $x \leqslant 0$.

Замечание 2. Любая собственная выпуклая функция f полунепрерывна снизу в точках из $\mathrm{ri\,dom} f$. Действительно, если $x_0 \in \mathrm{ri\,dom} f$, то в малой окрестности x_0 есть только точки из $\mathrm{ri\,dom} f$ и точки, не лежащие в $\mathrm{dom} f$. Мы знаем, что $f \in C(\mathrm{ri\,dom} f)$, а значения в точках не из $\mathrm{dom} f$ равны ∞ и не могут нарушить свойство полунепрерывности снизу.

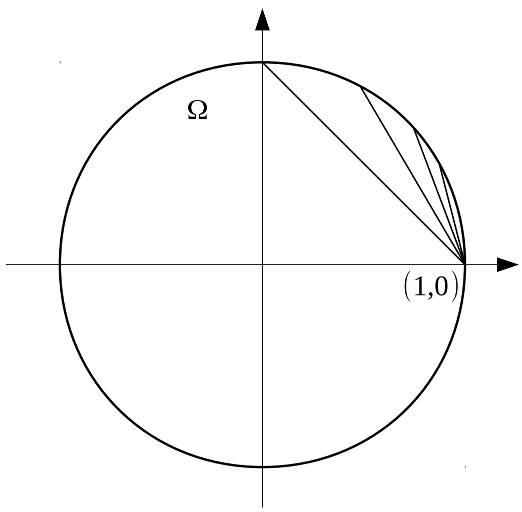

Рис. 8: Пример выпуклой замкнутой функции, не являющейся непрерывной на своей эффективной области.

Итак, выпуклая, собственная, замкнутая функция является непрерывной во внутренних точках эффективной области и полунепрерывной снизу в граничных точках. Более того, если $d = 1$, то такая функция непрерывна на своей эффективной области (см. упражнение 3.4). Однако, даже в двумерном случае такая функция может быть разрывна в граничных точках эффективной области:

Пример 2. Рассмотрим серию функций f_n, $n \in \mathbb{N}$, на плоскости $\mathbb{R}^2$ с одной и той же эффективной областью $\mathrm{dom} f_n = \Omega = \{x^2 + y^2 \leqslant 1\}$:

$$f_n(x) = \max\{0, n^2(x-1) + ny\}.$$

Другими словами, левее прямой $y = -n(x-1)$ функция f_n равна 0 на круге Ω, и линейно растет правее этой прямой, достигая на Ω максимума $M_n = \frac{n^3+n-n^2\sqrt{n^2+1}}{\sqrt{n^2+1}} \leqslant 1$ в точке $x_n = \frac{n}{\sqrt{n^2+1}}$, $y_n = \frac{1}{\sqrt{n^2+1}}$. Каждая из функций f_n, очевидно, выпукла, замкнута и даже непрерывна на Ω.

Рассмотрим функцию $g(x) = \sup_n f_n(x)$. Отметим, что для любой точки $x \in \Omega$ выполнено $f_n(x) = 0$ для достаточно больших n. Поэтому для каждой точки x супремум превращается в максимум по конечному множеству.

Функция g ограничена, так как $M_n \leqslant 1$. Убедимся в том, что функция g выпукла и замкнута, но разрывна на Ω. Она является выпуклой и замкнутой, так как $\operatorname{epi} g = \bigcap_n \operatorname{epi} f_n$. При этом $g(x_n, y_n) \geqslant M_n \to \frac{1}{2}$ и $x_n \to 1$, $y_n \to 0$. Поскольку $g(1,0) = 0$, то функция g разрывна в точке $(1,0)$ (см. рис. 8).

Упражнения:

3.1 Описать все несобственные выпуклые функции с замкнутым надграфиком.

3.2 Пусть $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ – выпуклая функция и $x_0, x_1 \in \operatorname{dom} f$. Докажите, что имеет место альтернатива: либо $f(x_\lambda) = (1-\lambda)f(x_0) + \lambda f(x_1)$ для всех $\lambda \in [0;1]$, либо $f(x_\lambda) < (1-\lambda)f(x_0) + \lambda f(x_1)$ для всех $\lambda \in (0;1)$.

3.3 Докажите, что собственная выпуклая функция f на $\mathbb{R}^d$ ограничена на любом компактном множестве $K \subset \operatorname{ri}\operatorname{dom} f$.

3.4 Докажите, что собственная выпуклая замкнутая функция на $\mathbb{R}^1$ непрерывна на $\operatorname{dom} f$.

3.5 (i) Пусть f_1 и f_2 – выпуклые замкнутые функции на $\mathbb{R}^d$. Докажите, что если $f_1(x) \geqslant f_2(x)$ для всех $x \in \operatorname{ri}\operatorname{dom} f_1$ и $f_2(x) \geqslant f_1(x)$ для всех $x \in \operatorname{ri}\operatorname{dom} f_2$, то $f_1(x) = f_2(x)$ для всех x. (ii) Останется ли утверждение верным, если заменить оба неравенства на противоположные?

3.6 Докажите, что функция расстояния $f(x) = \operatorname{dist}(x, C)$ от точки до выпуклого множества C является выпуклой.

3.7 Докажите следующую теорему существования: если функция в конечномерном пространстве имеет ограниченную эффективную область и замкнута, то она достигает своего минимума.

3.8 Докажите, что если функция $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ непрерывна, то ее график замкнут. Верно ли обратное утверждение?

3.9 Верно ли, что если собственная функция f выпукла, то выпуклы множества $f^{-1}(-\infty; c]$ для всех $c \in \mathbb{R}$? Верно ли обратное утверждение: если все множества $f^{-1}(-\infty; c]$ выпуклы при всех $c \in \mathbb{R}$, то собственная функция f выпукла?

4 Субдифференциал

Опорные гиперплоскости. Для выпуклых функций понятие производной не совсем естественно, так как выпуклые функции часто являются недифференцируемыми. Однако,

***Предложение* 11.** *Предположим, что выпуклая собственная функция f дифференцируема в точке x_0. Тогда для любого вектора ξ выполнено*

$$f(x_0+\xi) \geqslant f(x_0) + \langle f'(x_0), \xi\rangle$$

(напомним, $f'(x_0) \in \mathbb{R}^{d*}$ и $\langle f'(x_0), \xi\rangle = f'(x_0)[\xi]$).

Доказательство. Пусть $x_1 = x_0+\xi$, тогда $x_\lambda = x_0+\lambda\xi \in [x_0; x_1]$ при $\lambda \in [0;1]$. Согласно п.3 предложения 7 значение $f(x_0+\xi)$ лежит над секущей через точки $(x_0, f(x_0))$ и $(x_\lambda, f(x_\lambda))$, которая стремится к касательной прямой $(x_0, f(x_0)) + \mathbb{R}(\xi, \langle f'(x_0), \xi\rangle)$ в точке x_0 при $\lambda \to +0$. □

Таким образом, если производная в какой-либо точке существует, то надграфик выпуклой функции должен целиком лежать над касательной гиперплоскостью. Это свойство впуклых функций лежит в основе понятия *субдифференциала*, обобщающего понятие производной:

Определение 9. Субдифференциалом выпуклой функции f на $\mathbb{R}^d$ в точке x_0 называют множество $\partial f(x_0) \subset \mathbb{R}^{d*}$, состоящее из таких ковекторов p, что

$$\forall \xi\ f(x_0+\xi) \geqslant f(x_0) + \langle p, \xi\rangle \quad \Leftrightarrow \quad \forall x_1\ f(x_1) \geqslant f(x_0) + \langle p, x_1 - x_0\rangle. \tag{2}$$

Субдифференциальное исчисление оказывается очень удобным. Например, для выпуклых задач минимизации, имеем:

***Предложение* 12.** *Точка x_0 является глобальным минимумом выпуклой функции f, если и только если $0 \in \partial f(x_0)$.*

Геометрический смысл субдифференциала заключается в следующем. Пусть $p \in \mathbb{R}^{d*}$. Проведем через точку $(x_0, f(x_0))$ гиперплоскость в $\mathbb{R}^{d+1}$, ортогональную ковектору $(p, -1)$. Тогда $p \in \partial f(x_0)$, если надграфик $\mathrm{epi} f$ целиком лежит в верхнем полупространстве относительно этой гиперплоскости. Дадим точную формулировку:

Определение 10. Гиперплоскость H называется *опорной* к множеству C, если C целиком лежит в одном из двух (замкнутых) полупространств относительно H и при этом гиперплоскость H пересекается с замыканием $\mathrm{cl}\, C$. При этом если $x_0 \in H \cap \mathrm{cl}\, C$, то говорят, что гиперплоскость H является опорной к C в точке x_0. Если же множество C не лежит целиком в H, то опорную гиперплоскость называют *собственной.*

Второе условие в определении опорной гиперплоскости означает следующее: если гиперплоскость H удовлетворяет первому условию, то ее всегда можно подвинуть параллельно самой себе так, чтобы она удовлетворяла еще и второму условию. Это всегда можно сделать, и при том единственным способом, если $C \neq \emptyset$.

***Предложение* 13.** *Пусть f – выпуклая собственная функция и $x_0 \in \mathrm{dom} f$. Каждому элементу $p \in \partial f(x_0)$ отвечает единственная невертикальная опорная гиперплоскость H к $\mathrm{epi} f$ в точке $(x_0, f(x_0))$, а каждой невертикальной опорной гиперплоскости H к $\mathrm{epi} f$ в точке $(x_0, f(x_0))$ отвечает единственный элемент $p \in \partial f(x_0)$. Соответствие устроено следующим образом: гиперплоскость H ортогональна ковектору $(p, -1)$.*

Под вертикальным направлением в пространстве $\mathbb{R}^{d+1}$ (в котором лежит надграфик f) мы понимаем направление, отвечающее значениям функций. Другими словами, если гиперплоскость H в $\mathbb{R}^{d+1}$ ортогональна ковектору $(p,\lambda)\neq 0$, $p\in\mathbb{R}^{d*}$, $\lambda\in\mathbb{R}^*=\mathbb{R}$, то она является вертикальной, если $\lambda=0$.

Доказательство предложения 13. $(\to)$ Пусть $p\in\partial f(x_0)$. Рассмотрим в $\mathbb{R}^{d+1}$ гиперплоскость H, ортогональную ковектору $(p,-1)$ и проходящую через точку $(x_0,f(x_0))$, т.е. $H=\{x_0+\xi, f(x_0)+\langle p,\xi\rangle\}$, $\xi\in\mathbb{R}^d$. Тогда по определению субдифференциала $f(x_0+\xi)\geqslant f(x_0)+\langle p,\xi\rangle$, то есть надграфик f лежит выше H, что и требовалось.

$(\leftarrow)$ Пусть H – невертикальная гиперплоскость, опорная к $\operatorname{epi} f$ в $(x_0,f(x_0))$. Обозначим через $\psi=(p,\lambda)\neq 0$, $p\in\mathbb{R}^{d*}$, $\lambda\in\mathbb{R}^*=\mathbb{R}$, ортогональный ковектор к H. Ковектор ψ определен однозначно с точностью до умножения на ненулевое число. Поскольку гиперплоскость H не является вертикальной, то $\lambda\neq 0$, и без ограничения общности $\lambda=-1$. В этом случае верхнее полупространство H^+ относительно H задается условием

$$H^+=\{(x,a): x\in\mathbb{R}^d, a\in\mathbb{R}, \text{ и } \langle p,x\rangle-a\leqslant\langle p,x_0\rangle-f(x_0)\}.$$

Поскольку $\operatorname{epi} f\subset H^+$ (в нижнем полупространстве надграфик, очевидно, лежать не может), то $\langle p,x\rangle-f(x)\leqslant\langle p,x_0\rangle-f(x_0)$, что и требовалось. □

Итак, для нахождения субдифференциала достаточно построить опорную невертикальную гиперплоскость к надграфику. Удобно иметь простое условие, гарантирующее, что опорная гиперплоскость не является вертикальной:

***Предложение* 14.** *Пусть f – выпуклая собственная функция на $\mathbb{R}^d$ и H – собственная опорная гиперплоскость к $\operatorname{epi} f$ в $(x_0,f(x_0))$. Тогда если $x_0\in\operatorname{ri}\operatorname{dom} f$, то гиперплоскость H не является вертикальной.*

Доказательство. Действительно, если гиперплоскость H была бы вертикальной, то ее вертикальная проекция на $\mathbb{R}^d$ была бы гиперплоскостью в $\mathbb{R}^d$ и разделяла бы x_0 и $\operatorname{dom} f$. Поскольку $x_0\in\operatorname{ri}\operatorname{dom} f$, то проекция должна была бы содержать $\operatorname{dom} f$ целиком. В этом случае сама гиперплоскость H содержала бы надграфик $\operatorname{epi} f$, что запрещено условием. □

Локализация субдифференциала. Субдифференциал наследует многие свойства производной. Например, также как и производная, субдифференциал в точке определяется по значениям функции в сколь угодно малой окрестности точки и без учета членов типа o-малое:

***Теорема* 6** (о локализации субдифференциала)**.** *Пусть f – выпуклая собственная функция на $\mathbb{R}^d$, $x_0\in\operatorname{dom} f$ и $p\in\mathbb{R}^{d*}$. Предположим, что для любого ξ выполнено*

$$f(x_0+\xi)\geqslant f(x_0)+\langle p,\xi\rangle+r(\xi)$$

где $r:\mathbb{R}^d\to\mathbb{R}$ такая функция, что для любого $\xi\in\mathbb{R}^d$ выполнено $\frac{r(\lambda\xi)}{\lambda}\to 0$ при $\lambda\to 0$ (например, $r(\xi)=o(|\xi|)$). Тогда $p\in\partial f(x_0)$.

Доказательство. Предположим противное: пусть для некоторого $\xi\in\mathbb{R}^d$ оказалось:

$$f(x_0+\xi)=f(x_0)+\langle p,\xi\rangle-a,$$

где $a > 0$. Если $x_1 = x_0 + \xi$, то $x_\lambda = x_0 + \lambda\xi$, и для $\lambda \in [0;1]$ получаем

$$f(x_0 + \lambda\xi) \leqslant (1-\lambda)f(x_0) + \lambda f(x_1) = f(x_0) + \lambda\langle p, \xi\rangle - \lambda a.$$

По условию, $f(x_0 + \lambda\xi) \geqslant f(x_0) + \lambda\langle p, \xi\rangle + r(\lambda\xi)$. Значит, $-\lambda a \geqslant r(\lambda\xi)$. Деля на λ и устремляя $\lambda \to +0$, получаем $-a \geqslant 0$, что невозможно, так как по предположению $a > 0$. Противоречие. □

***Следствие* 12.** *Пусть f – выпуклая функция. Если для некоторого p неравенство 2 выполняется для всех x_1 из окрестности x_0, то оно выполняется для всех x_1, т.е. $p \in \partial f(x_0)$.*

Доказательство. Достаточно в теореме о локализации субдифференциала положить $r(\xi) = 0$ в окрестности 0 и доопределить любым удобным образом вне этой окрестности, например, $r(\xi) = f(x_0 + \xi) - f(x_0) - \langle p, \xi\rangle$. □

***Следствие* 13.** *Любой локальный минимум выпуклой функции является и ее глобальным минимумом.*

Доказательство. Достаточно применить предыдущее следствие к нулевому ковектору. □

***Следствие* 14.** *Пусть f и g – выпуклые собственные функции. Если $f(x) = g(x)$ для всех x из некоторой окрестности x_0, то $\partial f(x_0) = \partial g(x_0)$.*

Структура субдифференциала. Хорошо известно, что производная гладкой выпуклой функции одной переменной монотонна. Это же свойство наследует и субдифференциал выпуклой функции многих переменных:

***Предложение* 15.** *Субдифференциал выпуклой функции монотонен, т.е. для любых точек x_0 и x_1 и для любой выпуклой функции f выполнено*

$$\langle \partial f(x_1) - \partial f(x_0), x_1 - x_0\rangle \geqslant 0.$$

(равносильно: $\forall p_0 \in \partial f(x_0)\, \forall p_1 \in \partial f(x_1)$ выполнено $\langle p_1 - p_0, x_1 - x_0\rangle \geqslant 0$).

Доказательство. По определению субдифференциала: $f(x_1) \geqslant f(x_0) + \langle p_0, x_1 - x_0\rangle$ и $f(x_0) \geqslant f(x_1) + \langle p_1, x_0 - x_1\rangle$. Складывая эти неравенства, получаем искомое. □

Нетрудно показать, что

***Предложение* 16.** *Субдифференциал выпуклой функции в точке является выпуклым замкнутым множеством.*

Доказательство. Действительно, если зафиксировать вектор ξ, то неравенство из 2 будет выполняться для ковекторов p из замкнутого полупространства $F_\xi = \{p : \langle p, \xi\rangle \geqslant f(x_0 + \xi) - f(x_0)\} \subset \mathbb{R}^{d*}\}$. Очевидно, $\partial f(x_0) = \bigcap_\xi F_\xi$, поэтому субдифференциал будет выпуклым и замкнутым как пересечение выпуклых замкнутых множеств. □

Очевидно, что если $f \not\equiv \infty$, то в любой точке $x \notin \operatorname{dom} f$ субдиференциал пуст, $\partial f(x) = \emptyset$. Если же есть точка x такая, что $f(x) = -\infty$, то либо $f(y) \neq -\infty$ и $\partial f(y) = \emptyset$, либо $f(y) = -\infty$ и $\partial f(y) = \mathbb{R}^{d*}$. Поэтому субдифференциал несобственных функций устроен очень просто. Реальный интерес субдифференциал представляет только для собственных функций. Поэтому, конечно, в первую очередь нас интересует следующий вопрос: насколько много точек, в которых субдифференциал выпуклой функции непуст:

Теорема 7. *Пусть f – выпуклая собственная функция на $\mathbb{R}^d$. Тогда $\partial f(x) = \emptyset$, если $x \notin \operatorname{dom} f$, и $\partial f(x) \neq \emptyset$, если $x \in \operatorname{ri}\operatorname{dom} f$.*

Доказательство. Очевидно, что если $x \notin \operatorname{dom} f$ и $f \not\equiv +\infty$, то $\partial f(x) = \emptyset$. Поэтому в проверке нуждается только второе утверждение.

Пусть $x \in \operatorname{ri}\operatorname{dom} f$ и, значит, $f(x) \neq \pm\infty$. Точка $(x_0, f(x_0))$ не лежит в относительной внутренности надграфика $\operatorname{epi} f$ по следствию 11. Поэтому по первой теореме отделимости найдется собственно разделяющая их в $\mathbb{R}^{d+1}$ гиперплоскость H. Гиперплоскость H, очевидно, является собственной опорной гиперплоскостью к $\operatorname{epi} f$. Она не может быть вертикальной по предложению 14, так как $x \in \operatorname{ri}\operatorname{dom} f$. Поэтому она определяет субдифференциал в точке x_0 (по предложению 13). □

Таким образом, любая выпуклая функция имеет непустой субдифференциал почти во всех точках эффективной области. Однако, вообще говоря, субдифференциал может быть пустым множеством в точках относительной границы эффективной области $(\operatorname{dom} f) \setminus (\operatorname{ri}\operatorname{dom} f)$. Рассмотрим типичный

Пример 3. Пусть $f(x) = -\sqrt{x}$ при $x \geqslant 0$ и $f(x) = +\infty$ при $x < 0$. Тогда функция f выпукла, но $\partial f(0) = \emptyset$.

Также несложно привести пример такой выпуклой функции, что ее субдифференциал в некоторой точке был бы неограниченным множеством. Например, если $\dim\operatorname{dom} f < d$, и $p \in \partial f(x)$ для какой-то точки $x \in \operatorname{dom} f$, то $p + q \in \partial f(x)$ для любого ковектора q, ортогонального $\operatorname{aff}\operatorname{dom} f$. Поскольку множество таких ковекторов образует линейное подпространство в $\mathbb{R}^{d*}$ размерности $d - \dim\operatorname{dom} f$ (и, следовательно, неограничено), то неограничен и субдифференциал $\partial f(x_0)$.

Предложение 17. *Пусть f – выпуклая собственная функция и $K \subset \operatorname{int}\operatorname{dom} f$ – компактное множество. Тогда субдифференциал f на K ограничен. Другими словами, множество $\bigcup_{x \in K} \partial f(x)$ ограничено.*

Доказательство. Предположим противное: для некоторой последовательности $x_n \in K$ нашлась неограниченная последовательность $p_n \in \partial f(x_n)$. Поскольку множество K компактно, то из последовательности x_n можно выбрать сходящуюся. Без ограничения общности $x_n \to x_0 \in K$. Разделим неравенство 2 для x_n и p_n на $|p_n|$:

$$\frac{1}{|p_n|} f(x) \geqslant \frac{1}{|p_n|} f(x_n) + \langle \frac{p_n}{|p_n|}, x - x_0 \rangle$$

для всех $x \in \operatorname{dom} f$. Поскольку ковекторы $\frac{p_n}{|p_n|}$ лежат на единичной сфере, то есть хотя бы одна предельная точка q, $|q| = 1$. Переходя, если потребуется к подпоследовательности, будем считать, что $\frac{p_n}{|p_n|} \to q$. Поскольку f непрерывна на $\operatorname{int}\operatorname{dom} f$ (теорема 5), то она ограничена на K, и, переходя к пределу в полученном равенстве, получаем $\langle q, x - x_0 \rangle \leqslant 0$ для всех $x \in \operatorname{dom} f$, и, следовательно, $x_0 \notin \operatorname{int}\operatorname{dom} f$. Противоречие.

Геометрическое объяснение таково: если субдифференциал неограничен, то есть сколь угодно близкие к вертикальному направлению опорные гиперплоскости к надграфику f в точках K. А, значит, выбирая подпоследовательности, можно найти и вертикальную опорную гиперплоскость к надграфику, которая может существовать, только если предельная точка не лежит в $\operatorname{int}\operatorname{dom} f$. □

Предложение 18. *Следующие два утверждения для любой собственной выпуклой функции f на $\mathbb{R}^d$ эквивалентны:*

- *Субдифференциал $\partial f(x_0)$ является непустым ограниченным множеством.*
- *Точка x_0 лежит в $\operatorname{int}\operatorname{dom} f$.*

Доказательство. ($\Longrightarrow$) Мы знаем, что если $\partial f(x_0) \neq \emptyset$, то $x_0 \in \operatorname{dom} f$. Покажем, что если $x_0 \in \partial \operatorname{dom} f$, то субдиференциал $\partial f(x_0)$ является либо пустым, либо неограниченным множеством. Поскольку $x_0 \notin \operatorname{int}\operatorname{dom} f$, то по следствию 8 их можно отделить: найдется ковектор $q \neq 0$ такой, что $\langle q, x_1 - x_0\rangle \leqslant 0$ при всех $x_1 \in \operatorname{dom} f$. С другой стороны, если $\partial f(x_0) \neq \emptyset$, то для $p \in \partial f(x_0)$ выполнено $f(x_0) + \langle p, x_1 - x_0\rangle \leqslant f(x_1)$ для всех $x_1 \in \operatorname{dom} f$. Домножая первое неравенство на $\lambda \geqslant 0$ и складывая со вторым, убеждаемся, что $p + \lambda q \in \partial f(x_0)$, и, значит, множество $\partial f(x_0)$ неограничено, если непусто.

Геометрическая суть дела такова: мы имеем две опорные гиперплоскости к $\operatorname{epi} f$ в $(x_0, f(x_0))$ – одна из них вертикальна и отделяет x_0 от $\operatorname{dom} f$, а вторая определяется элементом из субдифференциала. Постепенно поворачивая вторую к первой, мы получаем целое семейство невертикальных опорных гиперплоскостей, каждая из которых определяет субдифференциал в x_0.

($\Longleftarrow$) Немедленно следует из предложения 17, если положить $K = \{x_0\}$. □

Следствие 15. *Если $x \in \operatorname{int}\operatorname{dom} f$, то $\partial f(x)$ – выпуклое компактное множество.*

Следствие 16. *Пусть функция f является собственной и выпуклой. Если точка x лежит на границе $\operatorname{dom} f$, то субдифференциал $\partial f(x)$ либо пуст, либо неограничен.*

Дифференцируемость выпуклых функций. Как мы знаем, если выпуклая функция дифференцируема в точке, то ее производная в этой точке лежит в субдифференциале (предложение 11). Можно также показать, что в этом случае других элементов в субдифференциале нет. Верно даже намного более сильное утверждение:

Теорема 8. *Пусть f – выпуклая собственная функция на $\mathbb{R}^d$. Тогда следующие два утверждения эквивалентны:*

- *Существует производная $f'(x_0)$ в точке $x_0 \in \operatorname{dom} f$.*
- *Субдифференциал $\partial f(x_0)$ в точке $x_0 \in \operatorname{dom} f$ состоит из одного элемента.*

В любом из этих случаев $\partial f(x_0) = \{f'(x_0)\}$.

Таким образом, если выпуклая функция дифференцируема, то найти субдифференциал очень просто – он совпадает с производной. О том, как считать субдифференциал недифференцируемой выпуклой функции, мы поговорим в параграфе 6.

Доказательство теоремы 8. ($\Longrightarrow$) В предложении 11 мы уже доказали, что $f'(x_0) \in \partial f(x_0)$. Докажем, что других элементов в $\partial f(x_0)$ нет. Пусть $p \in \partial f(x_0)$. Тогда для любого вектора ξ и $\lambda > 0$ получаем

$$f(x_0)+\lambda\langle p,\xi\rangle \leqslant f(x_0+\lambda\xi)=f(x_0)+\lambda\langle f'(x_0),\xi\rangle+o(\lambda).$$

Сократим слева и справа на $f(x_0)$ и поделим на λ. Тогда после взятия предела при $\lambda\to 0$ получим $\langle p,\xi\rangle \leqslant \langle f'(x_0),\xi\rangle$. Заменяя ξ на $-\xi$, получим обратное неравенство.

($\Longleftarrow$) Пусть $\partial f(x_0)=\{p\}$. Докажем, что $\exists f'(x_0)=p$. Заменяя $f(x)$ на $f(x)-\langle p,x\rangle$, будем без ограничения общности считать, что $\partial f(x_0)=\{0\}$, т.е. $f(x)\geqslant f(x_0)$ для всех x. Предположим, что производной в x_0 не существует или она не равна 0. В этом случае построим элемент из $\partial f(x_0)$, отличный от 0. По предположению можно выбрать такую последовательность точек $x_k\to x_0$, что $|f(x_k)-f(x_0)|\geqslant C|x_k-x_0|$ для некоторой константы $C>0$. Поскольку $f(x_k)\geqslant f(x)$, получаем, что $f(x_k)\geqslant f(x_0)+C|x_k-x_0|$.

Обозначая $\xi_k=\frac{x_k-x_0}{|x_k-x_0|}$ и $a_k=|x_k-x_0|$, получим неравенство $f(x_0+a_k\xi_k)\geqslant f(x_0)+Ca_k$. Воспользуемся тем фактом, что при движении по лучу x_0x_k от точки x_0 наклон секущей может только возрастать (предложение 7). Поэтому для $t\geqslant a_k$ выполнено

$$\frac{f(x_0+t\xi_k)-f(x_0)}{t}\geqslant\frac{f(x_0+a_k\xi_k)-f(x_0)}{a_k}\quad\Longrightarrow\quad f(x_0+t\xi_k)\geqslant f(x_0)+Ct.$$

Заметим, что это неравенство выполнено также и для всех отрицательных t, так как $f(x)\geqslant f(x_0)$.

Последовательность направлений ξ_k имеет предельную точку, так как все векторы ξ_k лежат на единичной сфере. Без ограничения общности $\xi_k\to\xi_0$, $|\xi_0|=1$. Переходя к пределу в предыдущем неравенстве при $k\to\infty$, получаем, что

$$f(x_0+t\xi_0)\geqslant f(x_0)+Ct$$

для всех $t\geqslant\lim a_k=0$ и $t<0$, то есть для всех t. Таким образом, прямая $l=\{(x_0+t\xi_0,f(x_0)+Ct)\}$ не пересекается со внутренностью надграфика f (см. следствие 11). Поэтому по первой теореме отделимости их можно отделить. Разделяющая гиперплоскость H в $\mathbb{R}^{d+1}$ не может быть вертикальной, так как $x_0\in\operatorname{int}\operatorname{dom}f$. Поэтому H определяет элемент из $\partial f(x_0)$. Поскольку гиперплоскость H не может быть горизонтальной (так как обязана содержать наклонную прямую l), то она определяет ненулевой элемент в $\partial f(x_0)$, что противоречит условию.

□

Хорошо известно, что если функция имеет частные производные в области, и они непрерывны, то функция дифференцируема. Для выпуклых функций аналогичное утверждение намного сильнее: достаточно потребовать существования частных производных лишь в одной точке.

***Теорема* 9.** *Пусть f – выпуклая собственная функция на $\mathbb{R}^d$. Тогда если в точке x_0 существуют частные производные по d линейно независимым направлениям, то функция f дифференцируема в x_0. Например:*

$$\forall i=1,\dots d\ \exists\,\frac{\partial}{\partial x^i}f(x_0)\qquad\Longrightarrow\qquad\exists f'(x_0)$$

Доказательство. Без ограничения общности линейно можно считать, что независимые направления в условии – это как раз базисные векторы. Существование производных $f'_{x^i}(x_0)$ гарантирует, что $x_0 \in \operatorname{int}\operatorname{dom} f$. Значит, субдифференциал $\partial f(x_0)$ непуст. Покажем, что он содержит ровно один элемент. Действительно, рассмотрим ограничение f на прямую l_1 в $\mathbb{R}^d$, проходящую через x_0 и параллельную первому базисному вектору e_1. Надграфик $f|_{l_1}$ есть пересечение $\operatorname{epi} f$ с двумерной вертикальной плоскостью, содержащей l_1. Поэтому любая опорная гиперплоскость к $\operatorname{epi} f$ в $(x_0, f(x_0))$ будет и опорной гиперплоскостью к $\operatorname{epi} f|_{l_1}$. По теореме 8 она должна содержать касательную прямую $m_1 = \{(x_0 + \lambda e_1, f(x_0) + \lambda f'_{x^1}(x_0)), \lambda \in \mathbb{R}\}$. Действуя аналогично, получаем, что любая опорная гиперплоскость к $\operatorname{epi} f$ в $(x_0, f(x_0))$ должна содержать d прямых $m_i = \{(x_0 + \lambda e_i, f(x_0) + \lambda f'_{x^i}(x_0)), \lambda \in \mathbb{R}\}$, $i = 1, \dots, d$, а такая гиперплоскость в $\mathbb{R}^{d+1}$ единственна. □

Как и производная гладкой функции, субдифференциал непрерывно зависит от точки. А именно:

Предложение 19. *Пусть f – выпуклая собственная функция на $\mathbb{R}^d$, $x_k \to x_0$, $p_k \in \partial f(x_k)$ и $p_k \to p_0$. Тогда если функция f полунепрерывна в x_0, то $p_0 \in \partial f(x_0)$.*

Доказательство. Действительно, для любой точки x и любого числа k выполнено $f(x) \geqslant f(x_k) + \langle p_k, x - x_k \rangle$. Переходя к пределу при $k \to \infty$, поскольку $\liminf f(x_k) \geqslant f(x_0)$, имеем:

$$f(x) \geqslant \liminf_{k\to\infty} f(x_k) + \lim_{k\to\infty} \langle p_k, x - x_k \rangle \geqslant f(x_0) + \langle p, x - x_0 \rangle.$$

□

Следствие 17. *График субдифференциала $\operatorname{graph}\partial f = \{(x,p), p \in \partial f(x)\} \subset \mathbb{R}^d \times \mathbb{R}^{d*}$ выпуклой замкнутой собственной функции f является замкнутым.*

Доказательство. Замкнутая функция полунепрерывна снизу во всех точках по предложению 10. □

Еще одно важное свойство заключается в том, что если субдифференциал функции состоит из одного ковектора для каждой точки из некоторой области, то функция является непрерывно дифференцируемой в этой области:

Теорема 10. *Пусть f – выпуклая собственная функция на $\mathbb{R}^d$, и $\Omega \subset \mathbb{R}^d$ – некоторое множество. Тогда если субдифференциал $\partial f(x)$ для всех $x \in \Omega$ состоит из одной точки, то производная f' определена и непрерывна на Ω.*

Доказательство. Очевидно, $\Omega \subset \operatorname{dom} f$, и, поскольку в любой точке $x \in \Omega$ субдифференциал $\partial f(x)$ состоит из одной точки (а, значит, непуст и ограничен), то $\Omega \subset \operatorname{int}\operatorname{dom} f$ по предложению 17. По теореме 8 функция f имеет производную $f'(x)$ в каждой точке $x \in \Omega$ и $\partial f(x) = \{f'(x)\}$. Докажем, что производная непрерывна. Пусть $x_k \to x_0$. Тогда по предложению 17 последовательность $f'(x_k)$ ограничена и, следовательно, имеет хотя бы одну предельную точку. При этом любая ее предельная точка обязана лежать в $\partial f(x_0)$ по предложению 19. Поскольку субдифференциал $\partial f(x_0)$ состоит из одной точки $f'(x_0)$, то необходимо $f'(x_k) \to f'(x_0)$, что и требовалось. □

Определим теперь, насколько много точек дифференцируемости имеет выпуклая функция. Очевидно, что если точка лежит на границе эффективной области, то субдифференциал в ней либо пуст, либо неограничен. Поэтому если выпуклая функция дифференцируема в какой-то точке, то такая точка должна лежать во внутренности эффективной области.

***Теорема* 11** (Александрова). *Выпуклая собственная функция* $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ *дифференцируема почти всюду на* $\operatorname{int}\operatorname{dom} f$.

Доказательство. Пусть $x_0 \in \operatorname{int}\Omega$. Тогда по теореме 5 функция f является липшицевой в некоторой окрестности $U_{\varepsilon_0}(x_0)$, $\varepsilon > 0$ точки x_0. Поэтому по теореме Радемахера функция f дифференцируема в почти всех точках множества $U_{\varepsilon_0}(x_0)$.

Покроем теперь открытое множество $\operatorname{int}\operatorname{dom} f$ счетным числом таких окрестностей (например, с центрами в точках $x_k \in \mathbb{Q}^d$ с рациональными координатами и рациональных радиусов $\varepsilon_j \in \mathbb{Q}$). На каждой такой окрестности $U_{\varepsilon_j}(x_k)$ множество точек недифференцируемости f имеет меру 0. Остается заметить, что счетное объединение множеств меры 0 имеет меру 0. □

***Следствие* 18.** *Если* $\Omega \subset \operatorname{int}\operatorname{dom} f$ *– множество точек дифференцируемости выпуклой собственной функции* $f : \mathbb{R}^d \to \bar{\mathbb{R}}$, *то* $f \in C^1(\Omega)$ *и мера множества точек недифференцируемости* $\operatorname{int}\operatorname{dom} f \ \Omega$ *равна 0.*

Доказательство. По теореме 10 $f \in C^1(\Omega)$. Вторая же часть следствия есть в точности теорема Александрова. □

Упражнения:

4.1 Докажите, что если $\partial f(x) \neq \emptyset$, то функция f полунепрерывна снизу в x.

4.2 Предположим, f – выпуклая собственная функция на $\mathbb{R}^d$ и $k = \dim\operatorname{dom} f < d$. Пусть $\varphi : \operatorname{aff}\operatorname{dom} f \to \mathbb{R}^k$ – линейный изоморфизм. Тогда f можно рассмотреть как функцию на $\mathbb{R}^k$, а именно: рассмотрим функцию $g = f \circ \varphi^{-1}$, которая тоже будет выпуклой функцией. Пусть $L^\perp$ обозначает множество ковекторов, перпендикулярных подпространству L. Докажите, что для $x_0 \in \operatorname{dom} f$ выполнено

$$\partial f(x_0) = (\operatorname{aff}\operatorname{dom} f)^\perp \oplus \varphi^*[\partial g(\varphi(x_0))].$$

4.3 Докажите, что в любой точке $x_0 \in \operatorname{ri}\operatorname{dom} f$ субдифференциал $\partial f(x_0)$ есть сумма линейного подпространства $(\operatorname{aff}\operatorname{dom} f)^\perp$ и компактного выпуклого множества, аффинная оболочка которого трансверсальна $(\operatorname{aff}\operatorname{dom} f)^\perp$.

4.4 Вычислить субдифференциал любой нормы в 0.

5 Выпуклый принцип Лагранжа

В этом параграфе мы докажем принцип Лагранжа для выпуклой задачи – его обычно называют теоремой Каруша-Куна-Таккера. В противоположность принципу Лагранжа для гладкой задачи, который дает лишь необходимое условие минимума, принцип Лагранжа для выпуклой задачи является еще и достаточным.

Критерий разрешимости системы неравенств. Основная геометрическая идея заложена в нижеследующей теореме.

***Теорема* 12.** *Пусть $A \subset \mathbb{R}^d$ – выпуклое множество, а $f_1, \dots, f_n$ – собственные выпуклые функции на $\mathbb{R}^d$. Предположим, что $\operatorname{ri} A \subset \operatorname{dom} f_i$ для любого i. Тогда есть альтернатива:*

- *либо существует такая точка $\hat{x} \in A$, что $f_i(\hat{x}) < 0$ для всех i;*
- *либо найдется такой набор не равных одновременно нулю, неотрицательных чисел $\lambda_1, \dots, \lambda_n$, что $\sum_{i=1}^n \lambda_i f_i(x) \geqslant 0$ для всех $x \in A$.*

Числа $\lambda_i \geqslant 0$ как раз и являются множителями Лагранжа, как мы увидим в теореме Каруша-Куна-Таккера.

Доказательство теоремы 12. В одну сторону доказательство тривиально: если точка $\hat{x} \in A$ такова, что $f_i(\hat{x}) < 0$ при всех i, то для любых неотрицательных чисел $\lambda_i \geqslant 0$ получаем, что $\sum_i \lambda_i f_i(\hat{x}) \leqslant 0$. Если же хотя бы одно из чисел λ_i строго положительно, то эта сумма становится строго отрицательной.

Докажем обратное утвеждение. Пусть $A \neq \emptyset$ (иначе все тривиально). Предположим, что для любой точки $x \in A$ надется номер i такой, что $f_i(x) \geqslant 0$. Нам необходимо доказать существование набора $\lambda_i \geqslant 0$, указанного в условии теоремы. Для этого мы построим два специальных непересекающихся выпуклых множества и воспользуемся теоремой отделимости – набор λ_i будет ковектором, ортогональным разделяющей гиперплоскости.

Итак, пусть

$$C = \{(\mu^1, \dots, \mu^n) \in \mathbb{R}^n : \text{ найдется } x \in A, \text{ что } f_i(x) \leqslant \mu^i \text{ для всех } i\}.$$

Отметим, что $C \subset \mathbb{R}^n$, тогда как $A \subset \mathbb{R}^d$. Множество C, очевидно, непусто, так как, например, содержит точки вида $(f_1(x), \dots, f_n(x))$ для любого $x \in A \cap \operatorname{dom} f_1 \cap \dots \cap \operatorname{dom} f_n$.

Множество C выпукло. Действительно, если $\mu_0, \mu_1 \in C$, то найдутся такие точки $x_0, x_1 \in A$, что $\mu_0^i \geqslant f_i(x_0)$ и $\mu_1^i \geqslant f_i(x_1)$. Для $\gamma \in [0;1]$ обозначим $\mu_\gamma = (1-\gamma)\mu_0 + \gamma\mu_1$ и $x_\gamma = (1-\gamma)x_0 + \gamma x_1$. Тогда по неравенству Йенсена для выпуклой функции f_i получаем

$$\mu_\gamma^i \geqslant (1-\gamma) f_i(x_0) + \gamma f_i(x_1) \geqslant f_i(x_\gamma).$$

Поэтому точка μ_γ лежит в C.

Второе выпуклое множество – это отрицательный ортант в $\mathbb{R}^n$:

$$D = \{(\mu_1, \dots, \mu_n) \in \mathbb{R}^n : \ \mu_i \leqslant 0 \text{ для всех } i\}.$$

Множества C и $\operatorname{int} D$ не пересекаются – если $\mu \in C \cap \operatorname{int} D$, то найдется такая точка x, что $f_i(x) \leqslant \mu^i < 0$, что невозможно по сделанному предположению. Поэтому по первой теореме отделимости (теорема 2) существует разделяющая C и D гиперплоскость. А именно: найдется такой набор не равных одновременно 0 чисел $\lambda_1, \dots, \lambda_n$, что

$$\sup_{\mu \in D} \langle \lambda, \mu \rangle \leqslant \inf_{\mu \in C} \langle \lambda, \mu \rangle. \tag{3}$$

Покажем, что набор λ_i – искомый. Во-первых, $\lambda_i \geqslant 0$ – иначе бы супремум в левой части (3) равнялся $+\infty$. Во-вторых, поскольку $\lambda_i \geqslant 0$, то $\langle \lambda, \mu \rangle \leqslant 0$ для всех $\mu \in D$, и, значит, $\sup_{\mu \in D} \langle \lambda, \mu \rangle \leqslant 0$. Более того, $\sup_{\mu \in D} \langle \lambda, \mu \rangle = 0$, так как $0 \in D$.

Таким образом, $\inf_{\mu \in C} \langle \lambda, \mu \rangle \geqslant 0$, то есть $\langle \lambda, \mu \rangle \geqslant 0$ для всех $\mu \in C$. Как уже отмечалось, множество C содержит точки вида $(f_1(x), \dots, f_n(x))$ для любого $x \in A \cap \operatorname{dom} f_1 \cap \dots \cap \operatorname{dom} f_n$. Последнее множество содержит $\operatorname{ri} A$ по условию теоремы. Таким образом, мы доказали, что выпуклая функция $g = \sum_i \lambda_i f_i$ неотрицательна на относительной внутренности A:

$$g(x) \geqslant 0 \quad \text{для всех} \quad x \in \operatorname{ri} A.$$

Отсюда легко следует нужная нам неотрицательность функции g на всем множестве A (см. нижеследующую лемму). □

Лемма 2. *Пусть $A \subset \mathbb{R}^d$ – выпуклое множество, g – выпуклая собственная функция на $\mathbb{R}^d$ и $\operatorname{ri} A \subset \operatorname{dom} g$. Тогда если функция g неотрицательна на $\operatorname{ri} A$, то она неотрицательна на всем множестве A.*

Доказательство. Действительно, если есть такая точка $x_0 \in A$, что $g(x_0) < 0$, то для любой точки $x_1 \in \operatorname{ri} A$ получаем, что, с одной стороны, точки x_γ при $\gamma \in (0; 1]$ лежат в $\operatorname{ri} A$ (см. предложение 4), а, с другой стороны, $g(x_\gamma) \leqslant (1-\gamma) g(x_0) + \gamma g(x_1)$. Поскольку $x_1 \in \operatorname{dom} g$, то $g(x_1) \neq \infty$, и, значит, $g(x_\gamma) < 0$ при достаточно малых γ, что противоречит условию леммы. □

Теорема Каруша-Куна-Таккера. Докажем теперь принцип Лагранжа для выпуклой задачи. Пусть $f_0, f_1, \dots, f_n$ – выпуклые функции на $\mathbb{R}^d$ и $A \subset \mathbb{R}^d$ – выпуклое множество. *Выпуклой задачей* называется задача минимизации

$$f_0(x) \to \min$$

при условиях

$$f_i(x) \leqslant 0 \text{ при } i = 1, \dots, n \quad \text{и} \quad x \in A.$$

Отметим, что предыдущие условия задают выпуклое множество допустимых точек, на котором необходимо минимизировать выпуклую функцию f_0 – поэтому задачу и называют выпуклой. Важно также сказать, что функция f_0 занимает особое положение по отношению ко всем остальным функциям f_i. Необходимо их не путать – некоторые утверждения мы будем делать про все функции f_i, а некоторые – про функции f_i с индексами $i \geqslant 1$.

***Теорема* 13** (Каруша-Куна-Таккера)**.** *Пусть $A \subset \mathbb{R}^d$ – выпуклое множество, а f_i, $i = 0, \dots, n$б – выпуклые собственные функции на $\mathbb{R}^d$. Предположим, что $\operatorname{ri} A \subset \operatorname{dom} f_i$ для всех $i \geqslant 0$. Тогда*

- *Если $\hat{x} \in A$ – решение выпуклой задачи, то найдется такой набор не равных одновременно 0 чисел λ_i, $i = 0, \dots, n$, что*

  (i) (минимум лагранжиана) Функция Лагранжа, или лагранжиан

  $$\mathscr{L}(x) = \sum_{i=0}^{n} \lambda_i f_i(x)$$

  достигает в $\hat{x}$ своего глобального минимума на всем множестве A;

(ii) (условие неотрицательности) Числа λ_i неотрицательны для всех i;

(iii) (условие дополняющей нежесткости) Выполнено $\lambda_i f_i(\hat{x}) = 0$ для $i \geqslant 1$.

- *Предположим, для некоторой точки $\hat{x} \in A$, удовлетворяющей всем неравенствам $f_i(\hat{x}) \leqslant 0$, $i \geqslant 1$, нашелся такой набор не равных одновременно 0 чисел λ_i, что выполняются все три предыдущих условия (i), (ii) и (iii). Тогда если $\lambda_0 \neq 0$, то $\hat{x}$ – глобальный минимум выпуклой задачи.*

Числа λ_i называются *множителями Лагранжа* соответствующей выпуклой задачи.

Замечание 3. Если функция Лагранжа $\mathscr{L}$ дифференцируема в $\hat{x}$, то условие (i) минимума лагранжиана равносильно привычному условию $\mathscr{L}'(\hat{x}) = 0$, аналогично гладкому случаю. Однако, в выпуклом случае это условие является не только необходимым, но и достаточным, если $\lambda_0 \neq 0$.

Как всегда, принцип Лагранжа дает $d+n$ условий (d штук из $\mathscr{L}'(x) = 0$ и еще n из $\lambda_i f_i(x) = 0$) на $d+n+1$ неизвестную $x \in \mathbb{R}^d$, $\lambda \in \mathbb{R}^{n+1}$: естественно, множители Лагранжа λ_i можно все одновременно умножить на одно и то же положительное число – при этом ничего не изменится.

Доказательство теоремы 13. Докажем сначала первую часть. Будем считать без ограничения общности, что $f_0(\hat{x}) = 0$. Тогда если $x \in A$ и $f_i(x) \leqslant 0$ при $i \geqslant 1$, то $f_0(x) \geqslant f_0(\hat{x}) = 0$. Поэтому по критерию о разрешимости системы выпуклых неравенств, найдется такой ненулевой набор $\lambda_i \geqslant 0$, $i = 0, \dots, n$, что $\mathscr{L}(x) \geqslant 0$ для всех $x \in A$. Покажем, что набор λ_i – искомый.

Условие (ii) неотрицательности выполнено автоматически по теореме 12.

Условие (iii) дополняющей нежесткости выполнено по следующей причине. С одной стороны, $\lambda_0 f_0(\hat{x}) = 0$ и $\lambda_i f_i(\hat{x}) \leqslant 0$ при $i \geqslant 1$. Поэтому $\mathscr{L}(\hat{x}) \leqslant 0$. А, с другой стороны, $\mathscr{L}(x) \geqslant 0$ для всех $x \in A$ и, в том числе, для $\hat{x}$. Поэтому необходимо $\lambda_i f_i(\hat{x}) = 0$ при $i \geqslant 0$.

Таким образом, $\mathscr{L}(\hat{x}) = 0$ по условию дополняющей нежесткости, и условие (i) минимума Лагранжиана выполнено автоматически, так как $\mathscr{L}(x) \geqslant 0$ для всех $x \in A$.

Докажем теперь вторую часть теоремы. Поскольку $\lambda_0 \neq 0$, то без ограничения общности будем считать, что $\lambda_0 = 1$. Пусть $x \in A$ и $f_i(x) \leqslant 0$ при $i \geqslant 1$. Покажем, что $f(\hat{x}) \leqslant f(x)$:

$$f_0(\hat{x}) \overset{(a)}{=} f_0(\hat{x}) + \sum_{i=1}^{n} \lambda_i f_i(\hat{x}) = \mathscr{L}(\hat{x}) \overset{(b)}{\leqslant} \mathscr{L}(x) = f_0(x) + \sum_{i=1}^{n} \lambda_i f_i(x) \overset{(c)}{\leqslant} f_0(x).$$

((a) по дополняющей нежесткости, (b) по условию минимума лагранжиана и (c) по условию неотрицательности). □

Условие Слейтера. Заметим, что в обоих пунктах теоремы 13 утверждается (или требуется) существование набора λ_i, удовлетворяющего трем условиям (i), (ii) и (iii). Приведенное ниже *условие Слейтера* является достаточным для того, чтобы в таком наборе λ_0 не могло равняться 0 автоматически:

***Предложение* 20** (условие Слейтера)**.** *Предположим, что ненулевой набор λ_i для точки $\hat{x} \in A$ удовлетворяет условиям (i) минимума лагранжиана, (ii) неотрицательности и (iii) дополняющей нежесткости. Тогда если существует такая точка $\tilde{x} \in A$, что $f_i(\tilde{x}) < 0$ для всех $i \geqslant 1$, то $\lambda_0 \neq 0$.*

Доказательство. Действительно, если $\lambda_0 = 0$, то по условию дополняющей нежесткости $\mathscr{L}(\hat{x}) = 0$. С другой стороны, по условию неотрицательности $\lambda_i \geqslant 0$ и одно из них (отличное от λ_0) положительно. Поэтому $\mathscr{L}(\tilde{x}) < 0$ и мы получили противоречие с условием минимума лагранжиана. □

Разберем один занятный пример использования выпуклого принципа Лагранжа.

Пример 4. Пусть дана гладкая вогнутая возрастающая функция $h : [0; +\infty) \to \mathbb{R}$ (т.е. $h' > 0$ и $h'' < 0$), $n \in \mathbb{N}$ и числа $a_i > 0$, $i = 1, \dots, n$. Требуется найти такие неотрицательные числа $x_1, \dots, x_n$, дающие в сумме 1, что сумма $\sum_{i=1}^{n} a_i h(x_i)$ максимальна.

Другими словами, необходимо решить выпуклую задачу

$$-\sum_{i=1}^{n} a_i h(x_i) \to \min \quad \text{при условиях} \quad -x_i \leqslant 0 \text{ и } \sum_{i=1}^{n} x_1 \leqslant 1.$$

Заметим, что минимум достигается, поскольку мы минимизируем непрерывную функцию на компакте. Мы утверждаем, что в точке минимума выполнено $\sum_i x_i = 1$ и существует такое число $\gamma > 0$, что если $a_i < \gamma$, то $x_i = 0$, а если $a_i \geqslant \gamma$, то $a_i h'(x_i) = \gamma h'(0)$ (x_i однозначно находится из этого уравнения, т.к. производная h' убывает).

Для этой задачи выполнено условие Слейтера, поэтому $\lambda_0 = 1$ и любой набор x_i, удовлетворяющий теореме Каруша-Куна-Таккера, является оптимальным. Согласно этой теореме, найдутся такие числа $\mu \geqslant 0$ и $\lambda_i \geqslant 0$, что $\mu(\sum_i x_i - 1) = 0$, $\lambda_i x_i = 0$ и $\mathscr{L}' = 0$, где

$$\mathscr{L} = -\sum_i a_i h(x_i) + \mu\Big(\sum_i x_i - 1\Big) - \sum_i \lambda_i x_i.$$

Дифференцируя по x_i, получаем, что $\mu - \lambda_i = a_i h'(x_i)$.

Обозначим $\gamma = \mu / h'(0)$. Если $a_i < \gamma$, то $\mu - \lambda_i < \gamma h'(x_i) \leqslant \mu$, т.к. $h'(0) \geqslant h'(x_i) > 0$. Поэтому $\lambda_i > 0$, и, значит, $x_i = 0$ по условию дополняющей нежесткости.

Покажем, что если $a_i \geqslant \gamma$, то $\lambda_i = 0$. Действительно, если $\lambda_i > 0$, то $x_i = 0$ и $\mu - \lambda_i = a_i h'(0) \geqslant \gamma h'(0) = \mu$, что невозможно при $\lambda_i > 0$. Поэтому если $a_i \geqslant \gamma$, то $\lambda_i = 0$, и, значит, $\mu = a_i h'(x_i)$, т.е. $h'(x_i) = \mu / a_i$ или $a_i h'(x_i) = \gamma h'(0)$.

Осталось показать, что $\sum_i x_i = 1$. По условию дополняющей нежесткости достаточно проверить, что $\mu \neq 0$. Действительно, если $\mu = 0$, то $a_i > \gamma = 0$ для всех i. Поэтому $a_i h'(x_i) = \gamma h'(0) = 0$, что невозможно, т.к. $h' > 0$.

Упражнения:

5.1 Докажите лемму Фаркаша: пусть $f_0, f_1, \dots, f_n$ – линейные (однородные) функции на $\mathbb{R}^d$, тогда если для любого $x \in \mathbb{R}^d$ из неравенств $f_i(x) \geqslant 0$ при $i \geqslant 1$ следует неравенство $f_0(x) \geqslant 0$, то найдутся такие неотрицательные числа $\lambda_1, \dots, \lambda_n$, что $f_0 = \sum_{i=1}^{n} \lambda_i f_i$.

5.2 Остается ли верной теорема о разрешимости системы выпуклых неравенств, если отказаться от требования собственности?

5.3 Остается ли верной модификация теоремы 12 о разрешимости системы выпуклых неравенств, если в первом пункте заменить строгое неравенство $f_i(x) < 0$ нестрогим $f_i(x) \leqslant 0$, а во втором – наоборот: нестрогое $\sum_i \lambda_i f_i(x) \geqslant 0$ на строгое $\sum_i \lambda_i f_i(x) > 0$.

6 Основные выпуклые функции

Для определения выпуклости функции существует мощный аппарат, позволяющий быстро и просто проверять, является ли та или иная функция выпуклой – не применять же каждый раз определение. Для проверки непрерывности в математическом анализе обычно используется следующая логика: есть набор стандартных функций, про которые непрерывность известна, а также набор операций (типа сложения, умножения, деления, сложной функции и т.п.), которые сохраняют непрерывность или, как, например, в случае деления, сохраняют непрерывность при некоторых условиях. По той же логике вычисляется и производная – производные стандартных функций известны, и есть правила дифференцирования сложения, умножения и т.п. Аналогично и в выпуклом анализе есть набор стандартных выпуклых функций и операций, которые сохраняют выпуклость. К выпуклым операциям также прилагаются правила вычисления субдифференциала.

Мы начнем с введения основных выпуклых функций, которые так или иначе появляются почти в любой выпуклой задаче.

Индикаторная функция.

Определение 11. *Индикаторная функция* множества $C \subset \mathbb{R}^d$ обозначается через[2] $\delta_C(x) = \delta(x|C) : \mathbb{R}^d \to \bar{\mathbb{R}}$. По определению, $\delta_C(x) = 0$ для всех $x \in C$ и $\delta_C(x) = +\infty$ для всех $x \notin C$.

Предложение 21. *Индикаторная функция δ_C является (i) выпуклой, если и только если множество C выпукло, (ii) замкнутой, если и только если множество C замкнуто и (iii) собственной, если и только если $C \neq \emptyset$.*

Доказательство. Надграфик δ_C в $\mathbb{R}^{d+1}$ есть $C \times [0; +\infty)$ и потому выпукл, если и только если выпукло множество C и замкнут, если и только если C замкнуто. Поскольку $\delta_C(x) \neq -\infty$ ни для каких C и x, то функция δ_C является собственной, если и только если найдется такая точка x, что $\delta_C(x) \neq \infty$, т.е. если $C \neq \emptyset$. □

Предложение 22. *Если множество C выпукло, то для любого $x \in \operatorname{ri} C$ выполнено $\partial\delta_C(x) = (\operatorname{aff} C)^\perp$, где через $(\operatorname{aff} C)^\perp \subset \mathbb{R}^{d*}$ обозначено линейное подпространство ковекторов, ортогональных $\operatorname{aff} C$.*

Доказательство. Если $\operatorname{aff} C = \mathbb{R}^d$, то в любой внутренней точке $x \in \operatorname{int} C$ существует производная $\delta'_C(x) = 0$, и поэтому $\partial\delta_C(x) = \{0\}$ по теореме 8. В общем случае достаточно воспользоваться результатом упражнения 4.2. □

Опорная функция. Индикаторная функция позволяет точно описать произвольное множество $C \subset \mathbb{R}^d$ как функцию на $\mathbb{R}^d$. Однако, такое описание чересчур прямолинейно. Намного более изящно можно описать множество $C \subset \mathbb{R}^d$ с помощью некоторой функции на $\mathbb{R}^{d*}$ – такое описание часто оказывается намного полезнее прямого:

Определение 12. *Опорная функция* множества $C \subset \mathbb{R}^d$ обозначается через[3] $s_C(p) = s(p|C) : \mathbb{R}^{d*} \to \bar{\mathbb{R}}$. По определению, $s_C(p) \overset{\text{def}}{=} \sup_{x \in C}\langle p, x\rangle$ (напомним, $\sup\emptyset = -\infty$).

[2] Символ δ выбран по аналогии с δ-функцией Дирака.
[3] Символ s выбран от английского support.

Опорная функция, очевидно, положительно однородна, т.е.

$$s_C(\lambda p) = \lambda s_C(p), \quad \text{если} \quad \lambda > 0.$$

Описание множества с помощью его опорной функции имеет множество преимуществ. Однако, надо помнить, что при таком описании теряется часть информации о множестве, если оно не выпукло или не замкнуто, а именно:

***Предложение* 23.** $s_C = s_{\mathrm{cl\,conv}\, C}$.

Доказательство. Очевидно, $C \subset \mathrm{cl\,conv}\, C$, поэтому $s_C(p) \leqslant s_{\mathrm{cl\,conv}\, C}(p)$ для любого p. С другой стороны, если зафиксировать $p \in \mathbb{R}^{d*}$, то $C \subset A_p \stackrel{\mathrm{def}}{=} \{x : \langle p, x\rangle \leqslant s_C(p)\}$. Множество A_p есть замкнутое полупространство. Поскольку множество A_p выпукло, то $\mathrm{conv}\, C \subset A_p$, а поскольку множество A_p замкнуто, то $\mathrm{cl\,conv}\, C \subset A_p$. Поэтому $\langle p, x\rangle \leqslant s_C(p)$ для всех $x \in \mathrm{cl\,conv}\, C$, т.е. $s_{\mathrm{cl\,conv}\, C}(p) \leqslant s_C(p)$, что и требовалось. □

С другой стороны, если множество C выпукло и замкнуто, то оно может быть однозначно восстановлено по своей опорной функции с помощью взятия субдифференциала (см. следствие 24). Для того, чтобы это доказать, сначала докажем, что опорная функция является выпуклой замкнутой собственной функцией, а потом вычислим ее субдифференциал.

***Предложение* 24.** *Опорная функция s_C является (i) выпуклой при любом C, (ii) замкнутой при любом C, (iii) собственной, если и только если $C \neq \emptyset$. Если дополнительно $\mathrm{cl\,conv}\, C \neq \mathbb{R}^d$, то $s(p) \neq \infty$ для некоторого $p \neq 0$.*

Доказательство. Пусть $x \in \mathbb{R}^d$. Обозначим через $l_x : \mathbb{R}^{d^*} \to \mathbb{R}$ линейную функцию $l_x(p) = \langle p, x\rangle$. Тогда по определению s_C есть поточечный супремум функций l_x по всем $x \in C$: $s_C(p) = \sup_{x\in C} l_x(p)$. Надграфик s_c есть пересечение надграфиков функций l_x по всем $x \in C$. Поскольку надграфики функций l_x являются выпуклыми замкнутыми множествами, то таким же является и надграфик s_C и пункты (i) и (ii) доказаны.

Докажем теперь пункт (iii). Очевидно, $s_\emptyset \equiv -\infty$ – несобственная функция. Если $C \neq \emptyset$, то $s_C(0) = 0$, поэтому функция s_C является собственной. Если же дополнительно $\mathrm{cl\,conv}\, C \neq \mathbb{R}^d$, то, отделяя любую точку вне $\mathrm{cl\,conv}\, C$ от него самого, мы найдем ковектор $p_0 \neq 0$ такой, что $\sup_{x\in \mathrm{cl\,conv}\, C}\langle p_0, x\rangle < \infty$, и, значит, $s_C(p_0) < \infty$. □

Функция Минковского.

Определение 13. *Функция Минковского* множества C обозначается через[4] $\mu_C(x) = \mu(x|C) : \mathbb{R}^d \to \mathbb{R}$. По определению, $\mu_C(x) \stackrel{\mathrm{def}}{=} \inf\{\lambda \geqslant 0 : x \in \lambda C\}$ (напомним, $\inf \emptyset = +\infty$).

Таким образом, $\mu_C(x)$ – это минимальное[5] число $\lambda > 0$, в которое надо растянуть множество C, чтобы данная точка x в него попала.

Бывает удобно другое представление функции Минковского:

$$\text{если} \quad x \neq 0, \quad \text{то} \quad \mu_C(x) = \inf\{\lambda > 0 : \frac{x}{\lambda} \in C\}.$$

Для $x = 0$ данное равенство верно, только если $0 \in C$ или $C = \emptyset$.

---

[4]Символ μ выбран в честь Минковского

[5]или инфимум, если минимального нет

Функция Минковского μ_C, очевидно, является положительно однородной:

$$\mu_C(\lambda x) = \lambda \mu_C(x), \qquad \text{если} \qquad \lambda > 0.$$

Предложение 25. *Функция Минковского является (i) выпуклой, если множество C выпукло, (ii) собственной, если $C \neq \emptyset$.*

Доказательство. (i). Поскольку функция Минковского μ_C положительно однородна, то достаточно проверить, что если множество C выпукло, то $\mu(x_0 + x_1) \leqslant \mu(x_0) + \mu(x_1)$ для всех x_0 и x_1. Пусть x_0 и x_1 отличны от 0 (иначе все тривиально). Пусть $\lambda_i > 0$, $i = 0, 1$, – такие, что $y_i = \frac{x_i}{\lambda_i} \in C$. Отметим, что

$$\frac{x_0 + x_1}{\lambda_0 + \lambda_1} = \frac{\lambda_0}{\lambda_0 + \lambda_1} y_0 + \frac{\lambda_1}{\lambda_0 + \lambda_1} y_1 \in C,$$

так как коэффициенты при y_0 и y_1 неотрицательны и в сумме дают 1, а множество C выпукло. Таким образом, $\mu_C(x_0 + x_1) \leqslant \lambda_0 + \lambda_1$. Поскольку λ_i можно выбрать сколь угодно близким к $\mu_C(x_i)$, получаем искомое.

(ii). Очевидно, что $\mu_C(x) \in [0; +\infty]$ и $\mu_C(x) \neq -\infty$. Поэтому функция μ_C является собственной, если хотя бы для одной точки x выполнено $\mu_C(x) < \infty$ – т.е. когда множество C непусто. □

Свойство замкнутости для функции Минковского ведет себя довольно коварно. Несложно привести пример такого замкнутого выпуклого множества C, что функция μ_C не замкнута (см. упражнение 6.8). Однако, такое может происходить, только если $0 \notin C$. Наибольший интерес для функции Минковского представляют как раз множества, содержащие начало координат. Причина заключается в следующем:

Предложение 26. *Для любого непустого выпуклого множества C выполнено $\mu_C = \mu_{\mathrm{conv}\{0\} \cup C}$.*

Доказательство. Поскольку $C \subset \mathrm{conv}\{0\} \cup C$, то $\mu_{\mathrm{conv}\{0\} \cup C} \leqslant \mu_C$. Обратное неравенство следует из того, что

$$\mathrm{conv}\{0\} \cup C = \bigcup_{0 \leqslant \lambda \leqslant 1} \lambda C.$$

Поэтому, если $x/\lambda \in \mathrm{conv}\{0\} \cup C$ для $\lambda > 0$, то $x/\lambda' \in C$ для некоторого $0 < \lambda' \leqslant \lambda$, и, значит, $\mu_C(x) \leqslant \mu_{\mathrm{conv}\{0\} \cup C}(x)$, что и требовалось. □

Предложение 27. *Предположим, что множество $C \subset \mathbb{R}^d$ выпукло, замкнуто и $0 \in C$. Тогда функция Минковского μ_C выпукла и замкнута.*

Доказательство. (ii). Докажем, что функция Минковского μ_C полунепрерывна снизу, что равносильно замкнутости (по предложению 10). Итак, пусть $x_n \to x_0$. Необходимо доказать, что $\mu(x_0) \leqslant \liminf_{n \to \infty} \mu_C(x_n)$. Будем считать, что $\liminf \mu_C(x_n) < \infty$ и $x_0 \neq 0$ (иначе нужное неравенство выполняется тривиально). Будем считать, что существует $\lim \mu_C(x_n) < \infty$ (перейдем к подпоследовательности, если потребуется).

Разберем сначала случай, когда $\lim_{n \to \infty} \mu_C(x_n) > 0$. Тогда $\mu_C(x_n) > 0$ для почти всех n, и в силу замкнутости C выполнено $x_n / \mu_C(x_n) \in C$. Переходя к пределу, опять же в силу замкнутости C получаем, что $x_0 / \lim \mu_C(x_n) \in C$.

Поэтому $\mu_C(x_0) \leqslant \lim \mu_C(x_n)$. Отметим, что в разобранном случае условие $0 \in C$ не нужно.

Разберем теперь случай $\lim \mu_C(x_n) = 0$. Поскольку мы условились, что $x_0 \neq 0$, и, значит, $x_n \neq 0$, то в этом случае существует такая последовательность ненулевых чисел $\lambda_n \to 0$, что $x_n/\lambda_n \in C$. Воспользуемся тем, что $0 \in C$ и множество C выпукло. Соединим 0 и x_n/λ_n отрезком $[0; x_n\lambda_n]$, который должен целиком лежать в C. Поскольку $\frac{1}{\lambda_n} \to \infty$, то для любого числа $M > 0$ выполнено $Mx_n \in [0; x_n\lambda_n] \subset C$ для достаточно больших n. Поскольку $Mx_n \to x$, а множество C замкнуто, то $Mx \in C$ для всех $M > 0$, то есть $\mu_C(x) = 0$, что и требовалось. □

Отметим, что требование выпуклости C в предыдущем предложении, на самом деле, излишне. Для замкнутости функции μ_C достаточно потребовать от множества C замкнутости и звездности: чтобы вместе с любым элементом $x \in C$ множество C содержало целый отрезок $[0; x] \subset C$. Однако, нас будет интересовать лишь выпуклый случай, который гарантирует звездность.

Пример 5. Вычислим функцию Минковского множества $C \subset \mathbb{R}^2$, $C = \{(x, y) : y \geqslant x^2\}$. Очевидно, $\mu_C(0,0) = 0$. Если $y < 0$ или $y = 0$, но $x \neq 0$, то $\mu_C(x, y) = \infty$. Если же $y > 0$, то $\mu_C(x, y) = x^2/y$. Поскольку множество C выпукло, замкнуто и содержит начало координат, то функция μ_C выпукла и замкнута. При этом, во-первых, ее эффективная область $\operatorname{dom} \mu_C = \{(x, y) : y > 0\} \cup \{(0,0)\}$ не замкнута, и, во-вторых, ограничение $\mu_C|_{\operatorname{dom} \mu_C}$ не является непрерывной функцией, так как $\mu_C(x, x^2) = 1$ при $x \neq 0$, но $\mu_C(0,0) = 0$ (сравните с примером 2).

Функция Минковского очень похожа на опорную функцию. На самом деле, если множество C выпукло, замкнуто и содержит 0, то его функция Минковского совпадает с опорной функцией *поляры* множества C (см. предложение 52). Этот факт удобнее всего получать в рамках теории двойственности выпуклых объектов, которую мы подробно изложим в параграфе 8. Там же мы с легкостью вычислим субдифференциалы и функции Минковского, и опорной функции (см. предложение 51 и следствие 26).

Гладкие выпуклые функции. Помимо рассмотренных базовых выпуклых функций, выпуклыми являются многие гладкие функции:

Предложение 28. *Пусть $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ – собственная функция, множество $\operatorname{dom} f$ открыто и $f \in C^1(\operatorname{dom} f)$. Тогда функция f выпукла, если и только если $\operatorname{dom} f$ – выпуклое множество и производная f' монотонна, т.е. для любых $x_0, x_1 \in \operatorname{dom} f$ выполнено*

$$\langle f'(x_1) - f'(x_0), x_1 - x_0 \rangle \geqslant 0.$$

Доказательство. (Необходимость). Если функция f выпукла, то очевидно, что множество $\operatorname{dom} f$ выпукло. По теореме 8 $\partial f(x) = \{f'(x)\}$, а по предложению 15 субдифференциал монотонен.

(Достаточность). Сначала рассмотрим главный случай, когда функция f одномерна, т.е. $f : \mathbb{R} \to \bar{\mathbb{R}}$. Пусть $x_0, x_1 \in \operatorname{dom} f$. Тогда для $x_\lambda = (1-\lambda)x_0 + \lambda x_1$ выполнено $x_\lambda \in \operatorname{dom} f$ при $\lambda \in [0;1]$, так как множество $\operatorname{dom} f$ выпукло по условию. Вычислим разность значений f в x_0 и x_λ через производную:

$$f(x_\lambda) - f(x_0) = \int_{x_0}^{x_\lambda} f'(x)\,dx.$$

Растянем теперь отрезок $[x_0; x_\lambda]$ на весь отрезок $[x_0, x_1]$:

$$x = x_0 + \lambda(y - x_0), \;\; y \in [x_0; x_1]; \qquad dx = \lambda dy.$$

Заметим, что $x \leqslant y$, так как $\lambda \in [0; 1]$, и, значит, $f'(x) \leqslant f'(y)$ по условию монотонности. Поэтому

$$f(x_\lambda) - f(x_0) = \int_{x_0}^{x_\lambda} f'(x)\,dx \leqslant \lambda \int_{x_0}^{x_1} f'(y)\,dy = \lambda(f(x_1) - f(x_0)).$$

Перенося $f(x_0)$ направо, получаем $f(x_\lambda) \leqslant (1-\lambda)f(x_0) + \lambda f(x_1)$, что и требовалось.

Разберем теперь общий случай, $f : \mathbb{R}^d \to \bar{\mathbb{R}}$. Для доказательства выпуклости f нам достаточно доказать ее выпуклость при ограничении на любую прямую (см. предложение 7). Пусть $x_0 \in \operatorname{dom} f$ и l – прямая, проходящая через x_0 и имеющая направление $\xi \in \mathbb{R}^d$, $l = \{x_0 + \lambda\xi, \lambda \in \mathbb{R}\}$. Ограничение $f|_l$ можно задать следующим образом: $\varphi(\lambda) = f(x_0 + \lambda\xi)$. Функция φ одномерна, поэтому достаточно проверить, что она удовлетворяет условиям теоремы. Эффективная область ограничения $\varphi = f|_l$ есть пересечение $l \cap \operatorname{dom} f$, которое, очевидно, выпукло и открыто. Функция φ из класса C^1:

$$\varphi'(\lambda) = \langle f'(x_\lambda), \xi\rangle.$$

Осталось проверить монотонность ее производной: если $\lambda_1 > \lambda_0$, то, с учетом того, что $x_{\lambda_1} - x_{\lambda_0} = (\lambda_1 - \lambda_0)\xi$, получаем

$$\begin{aligned}\varphi'(\lambda_1) - \varphi'(\lambda_0) &= \langle f'(x_{\lambda_1}) - f'(x_{\lambda_0}), \xi\rangle = \\ &= \frac{1}{\lambda_1 - \lambda_0}\langle f'(x_{\lambda_1}) - f'(x_{\lambda_0}), (\lambda_1 - \lambda_0)\xi\rangle \geqslant 0,\end{aligned}$$

что и требовалось. □

***Предложение* 29.** *Пусть $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ – собственная функция, множество $\operatorname{dom} f$ открыто и $f \in C^2(\operatorname{dom} f)$. Тогда функция f выпукла, если и только если билинейная форма $f''(x)$ неотрицательно определена при любом $x \in \operatorname{dom} f$.*

Доказательство. Сначала разберем одномерный случай. В этом случае для дважды дифференцируемой функции f неотрицательность второй производной равносильна монотонности первой, и мы получаем предложение 28.

В многомерном случае проведем через любую точку $x_0 \in \operatorname{dom} f$ прямую l с любым направляющим вектором $\xi \in \mathbb{R}^d$, $l = \{x_0 + \lambda\xi, \lambda \in \mathbb{R}\}$. Тогда для $\varphi(\lambda) = f(x_0 + \lambda\xi)$ получаем

$$\varphi''(\lambda) = f''(x_\lambda)[\xi, \xi].$$

Поэтому вторая производная φ'' неорицательна, если и только если билинейная форма $f''(x_\lambda)$ в точке x_λ неотрицательна на направлении ξ. Поскольку прямую l можно провести через любую точку $x_0 \in \operatorname{dom} f$ и под любым направлением $\xi \in \mathbb{R}^d$, то $\varphi'' \geqslant 0$ для любой прямой l, если и только если форма $f''(x_0)$ неотрицательна в любой точке $x_0 \in \operatorname{dom} f$. □

Упражнения:

6.1 Пусть C – непустое, выпуклое, компактное множество. Докажите, что s_C является полунормой на $\mathbb{R}^{d*}$, если и только если $-C = C$.

6.2 Пусть C – непустое, выпуклое, компактное множество. При каких условиях опорная функция s_C будет нормой?

6.3 Доказать, что если множество C выпукло и $0 \in \operatorname{ri} C$, то $\mu_C = \mu_{\operatorname{cl} C}$.

6.4 Привести примеры (i) такого выпуклого множества C, что $0 \in C$, но $\mu_C \neq \mu_{\operatorname{cl} C}$ и (ii) такого множества C, что $0 \in \operatorname{int} C$, но $\mu_C \neq \mu_{\operatorname{cl} C}$.

6.5 Описать все такие функции $g : \mathbb{R}^d \to \mathbb{R}^{d*}$, что g и $-g$ монотонны, т.е. $\langle g(x)-g(y), x-y\rangle = 0$ для всех $x, y \in \mathbb{R}^d$.

6.6 Пусть $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ – положительно однородная неотрицательная функция и $f(0) = 0$. Доказать, что $f = \mu_{\{x : f(x) \leqslant 1\}}$.

6.7 Пусть $\|\cdot\|$ – какая-либо норма на $\mathbb{R}^d$ и $B = \{x : \|x\| \leqslant 1\}$ – единичный шар в этой норме. Докажите, что B – выпуклое, компактное множество, $0 \in \operatorname{int} B$ и $\mu_B(x) = \|x\|$.

6.8 Привести пример такого выпуклого замкнутого множества $C \subset \mathbb{R}^d$, что его функция Минковского μ_C не замкнута.

6.9 Докажите, что если множество C выпукло и компактно, что функция μ_C выпукла и замкнута.

7 Выпуклые операции

С выпуклыми функциями $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ очень удобно работать через их надграфик $\operatorname{epi} f \subset \mathbb{R}^{d+1}$. Многие операции с выпуклыми функциями также удобнее всего определять через теоретико-множественные операции над надграфиками. Однако, в результате часто могут получаться множества, которые надграфиками не являются, и которые необходимо преобразовать в функцию.

Построение функции по множеству. Пусть $A \subset \mathbb{R}^{d+1}$ – произвольное множество. A является надграфиком некоторой функции $f : \mathbb{R}^{d+1} \to \mathbb{R}$, если пересечение A с любой вертикальной прямой либо пусто, либо представляет собой замкнутую полупрямую $[a_0; +\infty)$, либо всю прямую. Алгебраически этот факт можно записать следующим образом. Пусть $x \in \mathbb{R}^d$, обозначим $A_x = \{a \in \mathbb{R} : (x, a) \in A\} \subset \mathbb{R}$. Тогда множество A является надграфиком, если

$$\forall x \in \mathbb{R}^d \qquad A_x = [\inf A_x; +\infty)\ \{-\infty\}.$$

Если же множество $A \subset \mathbb{R}^{d+1}$ произвольно, то есть естественный способ определить функцию $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ по A:

Определение 14. Пусть $A \subset \mathbb{R}^{d+1}$. Тогда функция $\operatorname{func} A : \mathbb{R}^d \to \bar{\mathbb{R}}$ задается следующим образом (см. рис. 9):

$$\operatorname{func} A(x) \stackrel{\text{def}}{=} \inf A_x.$$

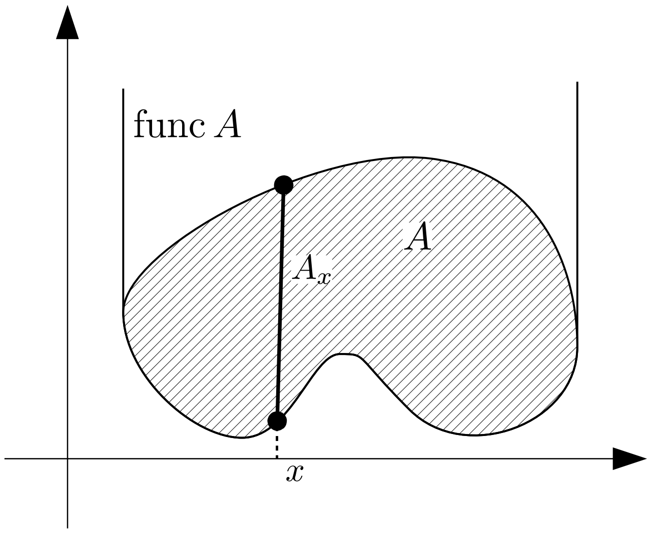

Рис. 9: Надграфик функции $\operatorname{func} A$.

***Предложение* 30.** *Функция* $\operatorname{func} A$ *есть максимальная из таких функций* $f : \mathbb{R}^d \to \bar{\mathbb{R}}$, *что* $A \subset \operatorname{epi} f$. *Равносильно: если* $A \subset \operatorname{epi} f$, *то* $\operatorname{epi} \operatorname{func} A \subset \operatorname{epi} f$.

Доказательство. Действительно, если $A \subset \operatorname{epi} f$, то для любой точки $(x, a) \in A$ выполнено $f(x) \leqslant a$. Поэтому $f(x) \leqslant \inf A_x = \operatorname{func} A(x)$. □

Включение множеств обращается при операции func:

***Предложение* 31.** *Если* $A \subset B$, *то* $\operatorname{func} A \geqslant \operatorname{func} B$.

Операция func обратна операции построения надграфика. Тривиально доказывается

***Предложение* 32.** *Для любой функции* $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ *выполнено* $f = \operatorname{func} \operatorname{epi} f$.

Построенная функция $\operatorname{func} A$, с одной стороны, может быть не выпуклой, а, с другой стороны, может оказаться выпуклой, даже если исходное множество A было невыпуклым. Однако,

***Предложение* 33.** *Если множество* $A \subset \mathbb{R}^{d+1}$ *выпукло, то функция* $\operatorname{func} A$ *является выпуклой.*

Доказательство. Допустим множество A выпукло. Возьмем две точки из надграфика: $(x_{0,1}, a_{0,1}) \in \operatorname{epi} \operatorname{func} A$. Тогда, по определению функции $\operatorname{func} A$, для любого $\varepsilon > 0$ найдутся такие числа $b_{0,1} \leqslant a_{0,1} + \varepsilon$, что $(x_{0,1}, b_{0,1}) \in A$. Поскольку A выпукло, то для любого $\lambda \in [0; 1]$ имеем $(x_\lambda, b_\lambda) \in A$, и, значит, $(x_\lambda, a_\lambda + \varepsilon) \in \operatorname{epi} \operatorname{func} A$. В силу произвольности $\varepsilon > 0$ получаем $(x_\lambda, a_\lambda) \in \operatorname{epi} \operatorname{func} A$, что и требовалось. □

Аналогичное утверждение для замкнутости неверно. Например, множество $A = \{(x, -x^{-2}) \in \mathbb{R}^2, x \in \mathbb{R}, x \neq 0\}$ замкнуто, $\operatorname{func} A(x) = -x^{-2}$ для $x \neq 0$ и $\operatorname{func} A(0) = +\infty$, т.е. функция $\operatorname{func} A$ не замкнута.

С помощью операции func легко описать геометрическую структуру функции Минковского. Поместим множество $C \subset \mathbb{R}^d$ в пространство $\mathbb{R}^{d+1}$ на высоте 1, $(C, 1) = \{(x, 1) : x \in C\}$, и рассмотрим выпуклый конус над $(C, 1)$, $\operatorname{cone}(C, 1) = \{(\lambda x, \lambda) : \lambda \geqslant 0, x \in C\}$. Тогда

$$\mu_C = \operatorname{func} \operatorname{cone}(C, 1).$$

Овыпукление функции. Из любой не выпуклой функции $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ легко построить выпуклую, просто взяв выпуклую оболочку ее надграфика:

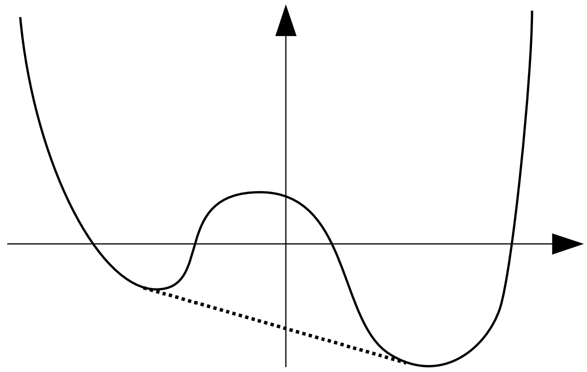

Рис. 10: Овыпукление функции

Определение 15. *Овыпуклением* (или выпуклой оболочкой) функции $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ называется функция $\operatorname{conv} f$, задаваемая равенством (см. рис. 10)

$$\operatorname{conv} f = \operatorname{func} \operatorname{conv} \operatorname{epi} f.$$

Здесь важно отметить, что если взять выпуклую оболочку надграфика некоторой функции, то получившееся множество, вообще говоря, может не являться надграфиком ни одной функции (см. упражнение 7.1), и, значит, априори нет гарантии, что построенная функция $\operatorname{conv} f$ является выпуклой. Тем не менее,

***Предложение* 34.** *Функция* $\operatorname{conv} f$ *является выпуклой. Равенство* $\operatorname{conv} f = f$ *выполнено, если и только если функция* f *выпукла.*

Доказательство. Множество $\operatorname{conv} \operatorname{epi} f$ является выпуклым, поэтому построенная по нему функция $\operatorname{conv} f$ тоже является выпуклой (по предложению 33).

Если функция f выпукла, то множество $\operatorname{epi} f$ выпукло и $\operatorname{conv} \operatorname{epi} f = \operatorname{epi} f$. Поэтому $\operatorname{conv} f = f$. Обратное утверждение легко следует из первой части: если $f = \operatorname{conv} f$, то функция f выпукла, так как выпукла функция $\operatorname{conv} f$. □

Операция conv наследует свойства операции func. Например,

***Предложение* 35.** *Если* $f \leqslant g$, *то* $\operatorname{conv} f \leqslant \operatorname{conv} g$.

Доказательство. Так как $f \leqslant g$, то $\operatorname{epi} f \supset \operatorname{epi} g$, поэтому $\operatorname{conv} \operatorname{epi} f \supset \operatorname{conv} \operatorname{epi} g$. Следовательно, $\operatorname{conv} f \leqslant \operatorname{conv} g$ по предложению 31. □

Рассмотрим теперь другое наследуемое свойство операции func. Поскольку $\operatorname{func} A$ – это максимальная функция, надграфик которой содержит множество A, то

***Следствие* 19.** $\operatorname{conv} f$ *– это максимальная выпуклая функция, не превосходящая* f.

Доказательство. Действительно, $\operatorname{epi} f \subset \operatorname{epi} \operatorname{conv} f$, поэтому $\operatorname{conv} f \leqslant f$. С другой стороны, если выпуклая функция g не превосходит f, то $\operatorname{epi} f \subset \operatorname{epi} g$ и $\operatorname{conv} \operatorname{epi} f \subset \operatorname{conv} \operatorname{epi} g = \operatorname{epi} g$, т.е. надграфик $\operatorname{epi} g$ содержит множество $\operatorname{conv} \operatorname{epi} f$. Осталось заметить, что функция $\operatorname{conv} f = \operatorname{func} \operatorname{conv} \operatorname{epi} f$ – это максимальная из функций, надграфик которых содержит множество $\operatorname{conv} \operatorname{epi} f$. □

Овыпукление функции можно задать и на алгебраическом языке:

Предложение 36.

$$\mathrm{conv} f(x) = \inf\{ \sum_{i=1}^{N} \lambda_i f(x_i), \ \ N \in \mathbb{N}\},$$

где инфимум берется по всем представлениям x в виде выпуклой комбинации $x = \sum_{i=1}^{N} \lambda_i x_i$ точек x_i из $\mathrm{dom} f$.

Доказательство. Обозначим инфимум в правой части утверждаемого равенства через $g(x)$ и докажем, что $\mathrm{conv} f = g$. Согласно следствию 19, для этого достаточно убедиться в том, что (i) $g \leqslant f$, (ii) функция g выпукла и (iii) если h – выпуклая функция и $h \leqslant f$, то $h \leqslant g$.

Пункт (i) получается очень просто: если $f(x) = \infty$, то, очевидно, $g(x) \leqslant f(x)$, если же $f(x) < \infty$, то $x \in \mathrm{dom} f$ и в определении $g(x)$ можно положить $N = 1$, $\lambda_1 = 1$ и $x_1 = x$.

Для доказательства пункта (ii) покажем, что для всех $x, y \in \mathrm{dom}\, g$ и $v \in [0, 1]$ выполнено $(1-v)g(x)+vg(y) \geqslant g((1-v)x+vy)$. По определению g, для любого $\varepsilon > 0$ найдется такое представление x в виде выпуклой комбинации $x = \sum_i \lambda_i x_i$ точек $x_i \in \mathrm{dom} f$, что $g(x) \geqslant \sum_i \lambda_i f(x_i) - \varepsilon$. Аналогично для $y = \sum_j \mu_j y_j$, $y_j \in \mathrm{dom} f$, выполнено $g(y) \geqslant \sum_j \mu_j f(y_j) - \varepsilon$. Умножая эти неравенства на $1 - v$ и v и складывая, получаем

$$(1-v)g(x) + vg(y) \geqslant (1-v)\sum_i \lambda_i f(x_i) + v\sum_j \mu_j f(y_j) - \varepsilon \geqslant g((1-v)x + vy) - \varepsilon$$

(последнее неравенство выполнено, так как $x = (1-v)\sum_i \lambda_i x_i + v\sum_j \mu_j y_j$ есть представление x в виде выпуклой комбинации точек x_i и y_j). Осталось заметить, что $\varepsilon > 0$ произвольно.

Ну и, наконец, докажем пункт (iii). Пусть функция h выпукла и $h \leqslant f$. Очевидно, $\mathrm{dom}\, h \supset \mathrm{dom} f$. Для любой точки $x \in \mathbb{R}^d$ и для любого ее представления $x = \sum_i \lambda_i x_i$ в виде выпуклой комбинации точек $x_i \in \mathrm{dom} f$ по неравенству Йенсена получаем

$$h(x) \leqslant \sum_i \lambda_i h(x_i) \leqslant \sum_i \lambda_i f(x_i).$$

Как уже было сказано, для любого $\varepsilon > 0$ выпуклое преставление $x = \sum_i \lambda_i x_i$ можно выбрать так, что $g(x) \geqslant \sum_i \lambda_i f(x_i) - \varepsilon$. Поэтому $h(x) \leqslant g(x) + \varepsilon$ для любого $\varepsilon > 0$, что и требовалось. □

Замыкание функции. По незамкнутой функции легко построить замкнутую, взяв замыкание ее надграфика:

Определение 16. *Замыканием* функции $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ называется функция $\mathrm{cl} f$, задаваемая равенством

$$\mathrm{cl} f = \mathrm{func}\, \mathrm{cl}\, \mathrm{epi} f.$$

Априори, функция, построенная по замкнутому множеству, сама может быть незамкнутой. Тем не менее, если в качестве множества выступает замыкание надграфика некоторой функции, то такого происходить не может:

***Предложение* 37.** $\operatorname{cl}\operatorname{epi} f = \operatorname{epi}\operatorname{cl} f$.

Доказательство. Включение в одну сторону $\operatorname{cl}\operatorname{epi} f \subset \operatorname{epi}\operatorname{cl} f$ напрямую следует из определения функции $\operatorname{cl} f$.

Докажем обратное включение. Пусть $(x_0, a_0) \in \operatorname{epi}\operatorname{cl} f$. По определению $\operatorname{cl} f$ для любого $\varepsilon > 0$ найдется такое число $a \leqslant a_0 + \varepsilon$, что $(x_0, a) \in \operatorname{cl}\operatorname{epi} f$. Поэтому найдется такая последовательность $(x_n, a_n) \in \operatorname{epi} f$, что $x_n \to x_0$ и $a_n \to a$. Отбросив начальные члены последовательности, будем считать, что $a_n \leqslant a+\varepsilon = a_0+2\varepsilon$. Поэтому $(x_n, a_0+2\varepsilon) \in \operatorname{epi} f$, и, значит, $(x_0, a_0+2\varepsilon) \in \operatorname{cl}\operatorname{epi} f$. Поскольку множество $\operatorname{cl}\operatorname{epi} f$ замкнуто, то, устремляя ε к 0, получаем, что $(x_0, a_0) \in \operatorname{cl}\operatorname{epi} f$, что и требовалось. □

***Следствие* 20.** *Функция $\operatorname{cl} f$ является замкнутой. Равенство $\operatorname{cl} f = f$ выполнено, если и только если функция f замкнута.*

Доказательство. Действительно, множество $\operatorname{epi}\operatorname{cl} f = \operatorname{cl}\operatorname{epi} f$ замкнуто, поэтому замкнута функция $\operatorname{cl} f$. Очевидно также, что

$$f = \operatorname{cl} f \quad \Leftrightarrow \quad \operatorname{epi} f = \operatorname{epi}\operatorname{cl} f = \operatorname{cl}\operatorname{epi} f \quad \Leftrightarrow \quad \text{функция } f \text{ замкнута.}$$

□

Также мы автоматически получаем:

***Предложение* 38.** *Функция $\operatorname{cl} f$ является максимальной замкнутой функцией, не превосходящей f.*

Доказательство. Очевидно, $\operatorname{cl} f \leqslant f$. Пусть g – замкнутая функция и $g \leqslant f$. Тогда $\operatorname{epi} f \subset \operatorname{epi} g$. Значит, $\operatorname{epi}\operatorname{cl} f = \operatorname{cl}\operatorname{epi} f \subset \operatorname{cl}\operatorname{epi} g = \operatorname{epi} g$. Поэтому $\operatorname{cl} f \geqslant g$. □

Тривиально из 31 следует, что

***Предложение* 39.** *Если $f \leqslant g$, то $\operatorname{cl} f \leqslant \operatorname{cl} g$.*

Так же, как и овыпукление, замыкание функции можно задать и на алгебраическом языке:

***Предложение* 40.**

$$\operatorname{cl} f(x) = \min\{f(x), \liminf_{\xi\to 0} f(x+\xi)\}.$$

Доказательство. Обозначим правую часть через $g(x)$ и докажем, что $\operatorname{cl} f = g$. Для этого достаточно проверить, что (i) $g \leqslant f$, (ii) функция g замкнута и (iii) если h – замкнутая функция и $h \leqslant f$, то $h \leqslant g$.

Пункт (i) тривиален. Пункт (iii) тоже доказывается просто: поскольку замкнутость функции равносильна ее полунепрерывности снизу (см. предложение 10), то из $h \leqslant f$ получаем $h(x) \leqslant \liminf_{\xi\to 0} h(x+\xi) \leqslant \liminf_{\xi\to 0} f(x+\xi)$. Поэтому $h(x) \leqslant g(x)$.

Для доказательства пункта (ii) покажем, что $g(x) \leqslant \liminf_{\xi\to 0} g(x+\xi)$. Зафиксируем $\varepsilon > 0$. По определению g, для каждого вектора ξ найдется такой вектор $\eta(\xi)$ сколь угодно малой длины (например, $|\eta(\xi)| < |\xi|$), что $g(x+\xi) \geqslant f(x+\xi+\eta(\xi)) - \varepsilon$. Устремляя $\xi \to 0$, получаем

$$\liminf_{\xi\to 0} g(x+\xi) \geqslant \liminf_{\xi\to 0} f(x+\xi+\eta(\xi)) - \varepsilon \geqslant g(x) - \varepsilon.$$

Поскольку $\varepsilon > 0$ произвольно, получаем нужное неравенство. □

Для выпуклых функций можно вывести и более удобное алгебраическое описание операции cl (см. упражнение 7.4).

Мы доказали в предложении 10, что замкнутая функция полунепрерывна снизу во всех своих точках. Из алгебраического описания замыкания функции следует более сильное утверждение:

Предложение 41. *Функция f полунепрерывна снизу в точке x, если и только если $f(x) = \mathrm{cl} f(x)$.*

Доказательство. Действительно, из алгебраического описания $\mathrm{cl} f$ следует, что $f(x) = \mathrm{cl} f(x)$, если и только если $\liminf_{\xi\to 0} f(x+\xi) \geqslant f(x)$, что и есть определение полунепрерывности снизу в x. □

Операция замыкания функции не может нарушить ее выпуклости:

Предложение 42. *Если функция f выпукла, то ее замыкание $\mathrm{cl} f$ тоже является выпуклой функцией.*

Доказательство. Замыкание выпуклого множества выпукло, поэтому множество $\mathrm{cl}\,\mathrm{epi} f$ выпукло. Поэтому функция $\mathrm{cl} f = \mathrm{func}\,\mathrm{cl}\,\mathrm{epi} f$ выпукла (см. предложение 33). □

Напомним, что в случае множеств замыкание выпуклой оболочки множества всегда содержит выпуклую оболочку замыкания множества, но равенства между ними, вообще говоря, нет. Это включение трансформируется в соответствующее неравенство для функций:

Предложение 43. $\mathrm{cl}\,\mathrm{conv} f \leqslant \mathrm{conv}\,\mathrm{cl} f$.

Доказательство. Очевидно, $\mathrm{conv} f \leqslant f$. Поэтому $\mathrm{cl}\,\mathrm{conv} f \leqslant \mathrm{cl} f$. Функция $\mathrm{cl}\,\mathrm{conv} f$ является выпуклой, так как функция $\mathrm{conv} f$ выпукла. Функция $\mathrm{conv}\,\mathrm{cl} f$ – это максимальная выпуклая функция, не превосходящая $\mathrm{cl} f$. Поэтому $\mathrm{cl}\,\mathrm{conv} f \leqslant \mathrm{conv}\,\mathrm{cl} f$. □

Обратное неравенство доказать аналогичным образом не получится, так как функция $\mathrm{conv}\,\mathrm{cl} f$ может быть незамкнутой. Более того, равенства здесь, вообще говоря, нет (см. упражнение 7.2).

Естественно вместе с замыканием надграфика попробовать рассмотреть и его внутренность (или относительную внутренность), однако для выпуклых функций ничего путного не получится. Легко проверить, что если f – выпуклая функция, то $f(x) = \mathrm{func}\,\mathrm{ri}\,\mathrm{epi} f(x)$ для всех $x \in \mathrm{ri}\,\mathrm{dom} f$, и мы просто отбросили значения функции на относительной границе $\mathrm{dom} f$

Поточечный максимум. В этом параграфе до этого момента мы изучали только унарные операции. Перейдем теперь к бинарным. Начнем с операции пересечения надграфиков:

Определение 17. Поточечный максимум двух функций f_1 и f_2 обозначается символом $\vee$:

$$f_1 \vee f_2 = \mathrm{func}(\mathrm{epi} f_2 \cap \mathrm{epi} f_2) \qquad \Longleftrightarrow \qquad (f_1 \vee f_2)(x) \stackrel{\mathrm{def}}{=} \max\{f_1(x), f_2(x)\}.$$

Предложение 44. *Если функции f_1 и f_2 или обе выпуклы, или обе замкнуты, то такой же является функция $f_1 \vee f_2$.*

Доказательство. Очевидно, $\mathrm{epi}(f_1 \vee f_2) = \mathrm{epi} f_1 \cap \mathrm{epi} f_2$. Остается заметить, что пересечение выпуклых множеств выпукло, а замкнутых – замкнуто. $\square$

Операция $\vee$, очевидно, коммутативна и ассоциативна, так как коммутативна и ассоциативна операция пересечения множеств. Поэтому можно определить поточечный супремум семейства функций f_a:

$$\mathrm{epi} \bigvee_a f_a \overset{\text{def}}{=} \bigcap_a \mathrm{epi} f_a \qquad \Leftrightarrow \qquad (\bigvee_a f_a)(x) \overset{\text{def}}{=} \sup_a f_a(x).$$

Поточечный супремум выпуклых (или замкнутых) функций, очевидно, является выпуклой (соответственно, замкнутой) функцией.

Аналогично определяется операция над функциями, отвечающая объединению надграфиков. Однако, объединение надграфиков может быть не выпуклым, поэтому после объединения еще необходимо взять выпуклую оболочку:

$$f_1 \wedge f_2 \overset{\text{def}}{=} \mathrm{func\,conv}(\mathrm{epi} f_1 \cup \mathrm{epi} f_2).$$

Операция $\wedge$ также коммутативна и ассоциативна, поэтому можно определить

$$\bigwedge_a f_a = \mathrm{func\,conv} \bigcup_a f_a.$$

Несложно доказать, что

$$(\bigwedge_a f_a)(x) = \inf\{\sum_{i=1}^{n} \lambda_i f_{a_i}(x_i), n \in \mathbb{N}, x = \sum_{i=1}^{n} \lambda_i x_i\}.$$

Здесь $\lambda_i \geqslant 0$, $\sum_{i=1}^{n} \lambda_i = 1$, и a_i – произвольные индексы. А если все функции f_a выпуклы, то индексы a_i можно считать различными.

Инфимальная конволюция. Мы знаем, что сумма (по Минковскому) двух выпуклых множеств является выпуклой функцией.

Определение 18. *Инфимальной конволюцией* двух функций f_1 и f_2 называется функция

$$f_1 \square f_2 = \mathrm{func}(\mathrm{epi} f_1 + \mathrm{epi} f_2).$$

Очевидно, из выпуклости f_1 и f_2 следует выпуклость $f_1 \square f_2$. Алгебраически операцию $\square$ можно записать так:

Предложение 45.

$$f_1 \square f_2(x) = \inf_y (f_1(x-y) + f_2(y)).$$

Следствие 21. *Пусть $C, D \subset \mathbb{R}^d$. Тогда $\delta_C \square \delta_D = \delta_{C+D}$.*

Из алгебраической записи следует коммутативность и ассоциативность инфимальной конволюции. Поэтому можно определить

$$(f_1 \square f_2 \square \ldots \square f_n)(x) = \inf\{\sum_{i=1}^{n} f_i(x_i) : x_i \in \mathbb{R}^d \text{ и } \sum_{i=1}^{n} x_i = x\}.$$

Сумма выпуклых множеств выпукла, поэтому инфимальная конволюция выпуклых функций является выпуклой функцией. Однако, сумма замкнутых множеств может быть незамкнутой, и аналогичная трудность есть и у инфимальной конволюции:

Пример 6. Пусть $f : \mathbb{R}^2 \to \bar{\mathbb{R}}$, f – индикаторная функция множества $C = \{(x,y) : x \geqslant 0; y \geqslant 0; xy \geqslant 1\}$, а $g(x,y) = f(-x,y)$, т.е. g – индикаторная функция множества $D = \{(x,y) : x \leqslant 0; y \geqslant 0; xy \leqslant -1\}$. Обе функции f и g выпуклы и замкнуты, но их инфимальная конволюция $f \,\square\, g$ есть индикаторная функция незамкнутого множества $C + D = \{(x,y) : y > 0\}$.

Еще один неприятный факт заключается в том, что инфимальная конволюция собственных функций может оказаться несобственной.

Пример 7. Пусть $f, g : \mathbb{R} \to \bar{\mathbb{R}}$, $f(x) = x$ и $g(x) = -x$. Тогда $f \,\square\, g \equiv -\infty$.

Как мы увидим в дальнейшем, операция «инфимальной конволюции» двойственна операции обыкновенного сложения функций.

Линейные операции. Очевидным образом определяются операции $f_1 + f_2$ и λf при $\lambda \geqslant 0$. Обе эти операции сохраняют и свойство выпуклости, и свойство замкнутости. Однако, сумма собственных функций может быть несобственной. Более того, если функции f_1 и f_2 не замкнуты, то, вообще говоря, $\mathrm{cl}(f_1 + f_2) \neq \mathrm{cl} f_1 + \mathrm{cl} f_2$ (см. упражнение 7.7).

По любому линейному преобразованию можно несколькими способами построить (унарное) преобразование выпуклых функций. Например, при $\lambda > 0$ мы можем растянуть надграфик f в λ раз. Получим

$$f\lambda \stackrel{\text{def}}{=} \mathrm{func}(\lambda\, \mathrm{epi} f),$$

то есть, если $\lambda > 0$, то $(f\lambda)(x) = \lambda f(x/\lambda)$. Операция $f \mapsto f\lambda$ тоже уважает и выпуклость, и замкнутость.

По данному линейному преобразованию $A : \mathbb{R}^d \to \mathbb{R}^{d'}$ строятся две естественные унарные операции, сохраняющие выпуклость. Во-первых, с его помощью можно, как обычно, перетаскивать любые функции в обратную сторону с $\mathbb{R}^{d'}$ на $\mathbb{R}^d$:

$$\text{для } f : \mathbb{R}^{d'} \to \bar{\mathbb{R}} \text{ имеем } (fA)(x) \stackrel{\text{def}}{=} f(Ax) \text{ для } x \in \mathbb{R}^d.$$

Отметим, что, согласно определениям, $f(\lambda A) \neq (f\lambda)A$. Точнее, $(f\lambda)A = \lambda f(\frac{A}{\lambda})$.

Операция $f \mapsto fA$, очевидно, сохраняет выпуклость: для $\lambda \in [0;1]$ и $x_0, x_1 \in \mathrm{dom} f$ получаем

$$\begin{aligned}(fA)((1-\lambda)x_0 + \lambda x_1) = f(A((1-\lambda)x_0 + \lambda x_1)) = f((1-\lambda)Ax_0 + \lambda Ax_1) \geqslant \\ \geqslant (1-\lambda)f(Ax_0) + \lambda f(Ax_1) = (1-\lambda)(fA)(x_0) + \lambda(fA)(x_1);\end{aligned}$$

и замкнутость: если $x_n \to x_0$, то $Ax_n \to Ax_0$, и поэтому

$$\liminf_{n\to\infty}(fA)(x_n) = \liminf_{n\to\infty} f(Ax_n) \geqslant f(Ax_0) = (fA)(x_0).$$

Во-вторых, можно определить унарную операцию по оператору A, применив A к надграфику f. Здесь следует проявить аккуратность, так как $\mathrm{epi} f \subset \mathbb{R}^{d+1}$. Положим $\tilde{A} : \mathbb{R}^{d+1} \to \mathbb{R}^{d'+1}$, $\tilde{A} : (x, a) \mapsto (Ax, a)$. Тогда

$$\text{для } f : \mathbb{R}^d \to \bar{\mathbb{R}} \text{ имеем } Af \stackrel{\text{def}}{=} \mathrm{func}\, \tilde{A}\, \mathrm{epi} f.$$

Алгебраическая запись этого определения такая:

$$(Af)(y) \stackrel{\text{def}}{=} \inf\{f(x) : Ax = y\}.$$

***Предложение* 46.** *Если функция $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ выпукла, то выпукла и функция $Af : \mathbb{R}^{d'} \to \mathbb{R}$.*

Доказательство. Действительно, образ $\tilde{A}\,\mathrm{epi} f$ выпуклого множества $\mathrm{epi} f$ под действием линейного отображения $\tilde{A}$, очевидно, выпукл. Поэтому нужный результат следует из предложения 33. □

Однако, замкнутость здесь может нарушаться. Например, если $A : \mathbb{R}^2 \to \mathbb{R}$ – это проекция, $A(x,y) = x$, то для индикаторной функции множества $C = \{(x,y) : x \geqslant 0; y \geqslant 0; xy \geqslant 1\}$ имеем $A\delta_C = \delta_{\{x>0\}}$.

Инфимальную конволюцию можно выразить через линейные операции. Действительно, если для $f_1, f_2 : \mathbb{R}^d \to \bar{\mathbb{R}}$ считать, что f_1 имеет аргумент x_1, а f_2 – x_2, то можно записать $(f_1 \Box f_2)(x) = \inf_{x_1+x_2=x}(f_1(x_1) + f(x_2))$. Формально мы использовали следующие линейные операции: линейные проекции $P_1, P_2 : \mathbb{R}^d \oplus \mathbb{R}^d \to \mathbb{R}^d$, заданные формулами $P_1(x_1,x_2) = x_1$ и $P_2(x_1,x_2) = x_2$. Тогда

$$f_1 \Box f_2 = (P_1 + P_2)(f_1 P_1 + f_2 P_2).$$

Таким образом, операция $f \mapsto Af$ может нарушать свойства замкнутости и собственности функции, аналогично тому, как их может нарушать операция инфимальной конволюции.

Еще одна линейная операция – это сдвиг аргумента на фиксированный вектор $\xi \in \mathbb{R}^d$, т.е. $f(x) \mapsto f(x - \xi)$. Эту операцию тоже легко определить через уже имеющиеся:

$$f(x - \xi) = (f \Box \delta_{\{\xi\}})(x).$$

Операция сдвига, очевидно, сохраняет и выпуклость, и замкнутость, и собственность функции.

Упражнения:

7.1 Построить пример такой функции f, что $\mathrm{epi}\,\mathrm{conv} f \neq \mathrm{conv}\,\mathrm{epi} f$.

7.2 Привести пример такой замкнутой функции f, что функция $\mathrm{conv} f$ не замкнута. Этот же пример показывает, что, вообще говоря, $\mathrm{cl}\,\mathrm{conv} f \neq \mathrm{conv}\,\mathrm{cl} f$.

7.3 Докажите, что если выпуклая функция f является собственной, то ее замыкание $\mathrm{cl} f$ также является собственной функцией.

7.4 Пусть f – выпуклая функция. Докажите, что если $x_1 \in \mathrm{ri}\,\mathrm{dom} f$, то для любой точки x_0 существует предел $\lim_{\lambda\to+0} f(x_\lambda)$ и он равен $\mathrm{cl} f(x_0)$.

7.5 Зафиксируем произвольное выпуклое множество $C \subset \mathbb{R}^d$. Выразите функцию расстояния $d(x, C)$ от точки до множества через стандартные выпуклые функции и операции.

7.6 Докажите, что если линейное отображение $A : \mathbb{R}^d \to \mathbb{R}^d$ обратимо, то $Af = fA^{-1}$.

7.7 Привести пример двух таких выпуклых функций f и g, что $\mathrm{cl}(f+g) \neq \mathrm{cl} f + \mathrm{cl}\, g$. Доказать, что если $\mathrm{ri}\,\mathrm{dom} f \cap \mathrm{ri}\,\mathrm{dom}\, g \neq \emptyset$, то все же $\mathrm{cl}(f+g) = \mathrm{cl} f + \mathrm{cl}\, g$.

8 Двойственность выпуклых объектов

Геометрия двойственности. Из курса линейной алгебры хорошо известно, что линейные объекты допускают как прямое, так и двойственное описание. Например, линейное подпространство можно задать как линейную оболочку образующих его векторов, а можно задать с помощью системы линейных уравнений – т.е. с помощью системы ортогональных ему ковекторов.

Эта двойственность продолжается и на выпуклые объекты. Ключевая геометрическая идея заложена в нижеследующей теореме:

***Теорема* 14** (Минковского о геометрии выпуклой двойственности)**.** *Любое выпуклое замкнутое множество совпадает с пересечением замкнутых полупространств, его содержащих.*

Доказательство. Очевидно, что любое множество $A \subset \mathbb{R}^d$ содержится в пересечении полупространств, его содержащих. Покажем, что если множество A выпукло и замкнуто, то оно совпадает с этим пересечением. Если $A = \emptyset$, то все тривиально. Пусть $A \neq \emptyset$ и $x_0 \notin A$. Построим полупространство, содержащее A, но не содержащее x_0. По второй теореме отделимости, x_0 и A можно строго отделить, т.е. найдется такая гиперплоскость H, что A содержится в одном замкнутом полупространстве относительно H, а x_0 в этом полупространстве не содержится, что и требовалось. □

Однако, пользоваться выпуклой двойственностью в такой геометрической форме не очень удобно – не хватает удобного алгебраического языка. Мы разработаем такой язык для трех типов выпуклых объектов: (i) для выпуклых функций, (ii) для выпуклых конусов и (iii) для выпуклых множеств, содержащих начало координат.

Проблема алгебраизации теоремы 14 заключается в том, что полупространства в $\mathbb{R}^d$ (или, равносильно, ориентированные гиперплоскости в $\mathbb{R}^d$) не образуют линейного пространства, а образуют гладкое многообразие. Для того, чтобы ввести на нем линейную структуру, необходимо выбрать карту, т.е. отбросить некоторые элементы.

Теорема Фенхеля-Моро. Начнем с выпуклых функций на $\mathbb{R}^d$, которые описываются своим надграфиком в $\mathbb{R}^{d+1}$. Мы хотим алгебраизировать результат теоремы 14, а именно: для данной выпуклой функции f на $\mathbb{R}^d$ мы каноническим образом построим некоторую функцию f^* на $\mathbb{R}^{d*}$ (называемую сопряженной к f) так, что вторая сопряженная f^{**} совпадет с f.

Самый естественный способ описать выпуклое множество с помощью сопряженного пространства – это использовать опорную функцию. Рассмотрим для произвольной выпуклой функции $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ опорную функцию ее надграфика $s_{\mathrm{epi}\, f}$, которая является однородной выпуклой функцией в $\mathbb{R}^{d*} \times \mathbb{R}$. Спустить функцию $s_{\mathrm{epi}\, f}$ в $\mathbb{R}^{d*}$ несложно: в силу положительной однородности, полную информацию несут значения $s_{\mathrm{epi}\, f}(p, b)$, при $b = 0, \pm 1$.

Очевидно, при $b = 1$, если $\mathrm{dom} f \neq \emptyset$, получаем

$$s_{\mathrm{epi}\, f}(p, 1) = \sup_{x \in \mathrm{dom} f,\, a \geqslant f(x)} (\langle p, x\rangle + a) = +\infty,$$

что не информативно. При $b = 0$ получаем

$$s_{\mathrm{epi}\, f}(p, 0) = \sup_{x \in \mathrm{dom} f} \langle p, x\rangle = s_{\mathrm{dom} f}(p).$$

Последняя формула имеет лишь опосредованное отношение к самой функции f, а описывает лишь ее эффективное множество $\mathrm{dom} f$. Самый интересный результат получается (естественно) при $b=-1$ и лежит в основе определения сопряженной функции:

$$s_{\operatorname{epi} f}(p,-1)=\sup_{x\in\operatorname{dom} f,\, a\geqslant f(x)}(\langle p,x\rangle - a)=\sup_{x\in\operatorname{dom} f}(\langle p,x\rangle - f(x)).$$

Очевидно, в последнем супремуме можно заменить включение $x\in\operatorname{dom} f$ на $x\in\mathbb{R}^d$, так как это ничего не меняет. Итак,

Определение 19. *Сопряженной функцией* к функции $f:\mathbb{R}^d\to\bar{\mathbb{R}}$ называется функция $f^*:\mathbb{R}^{d*}\to\bar{\mathbb{R}}$, заданная формулой

$$f^*(p)\overset{\text{def}}{=}s_{\operatorname{epi} f}(p,-1)=\sup_{x\in\mathbb{R}^d}(\langle p,x\rangle - f(x)).$$

Преобразование $f\mapsto f^*$ называют *преобразованием Лежандра-Юнга-Фенхеля*[6].

Пример 8. Вычислим функцию, сопряженную к индикаторной функции. Пусть $C\subset\mathbb{R}^d$ – произвольное множество. Поскольку $\delta_C(x)=\infty$ для $x\notin C$ и $\delta_C(x)=0$ для $x\in C$, то

$$\delta_C^*(p)=\sup_{x\in\mathbb{R}^d}(\langle p,x\rangle - \delta(x))=\sup_{x\in C}\langle p,x\rangle = s_C(p).$$

Таким образом, $\delta_C^*=s_C$. Поэтому опорную функцию иногда обозначают δ_C^*. Вычислить сопряженную функцию к опорной уже несколько сложнее, и мы сделаем это с помощью теоремы Фенхеля-Моро о дважды сопряженной функции. Об этой теореме будет подробно рассказано ниже.

Определение сопряженной функции несет в себе одновременно два геометрических смысла: с его помощью можно описать, с одной стороны, все невертикальные гиперплоскости, лежащие под графиком функции, а, с другой стороны, все точки, лежащие над каждой невертикальной гиперплоскостью из некоторого семейства. Оба этих геометрических смысла станут ясны по ходу доказательства нижеследующей теоремы двойственности для выпуклых функций:

***Теорема* 15** (Фенхеля-Моро)**.** *Пусть $f:\mathbb{R}^d\to\bar{\mathbb{R}}$. Тогда $f^{**}\leqslant f$, а если f является выпуклой, замкнутой, собственной, то $f^{**}=f$.*

Отметим, что вторая сопряженная функция f^{**} снова определена на пространстве $\mathbb{R}^d=\mathbb{R}^{d**}$.

Доказательство теоремы 15. Ясно, что надграфик $\operatorname{epi} f$ не может лежать под невертикальной гиперплоскостью в $\mathbb{R}^{d+1}$, а может лежать только над такой гиперплоскостью. Опишем алгебраически все невертикальные гиперплоскости в $\mathbb{R}^{d+1}$, над которыми лежит $\operatorname{epi} f$.

Пусть H – гиперплоскость в $\mathbb{R}^{d+1}$, а (p,λ) – ортогональный ему ненулевой ковектор в $\mathbb{R}^{(d+1)*}=\mathbb{R}^{d*}\times\mathbb{R}$, определенный с точностью до умножения на константу. Гиперплоскость H не является вертикальной, если $\lambda\neq 0$. В этом случае будем считать $\lambda=-1$.

Произвольная гиперплоскость H в $\mathbb{R}^{d+1}$ может, вообще говоря, не содержать начала координат. Поэтому для того, чтобы однозначно задать невертикальную

[6]В литературе иногда в названии этого преобразования не упоминают случайную часть из этих фамилий (встречаются все возможные варианты).

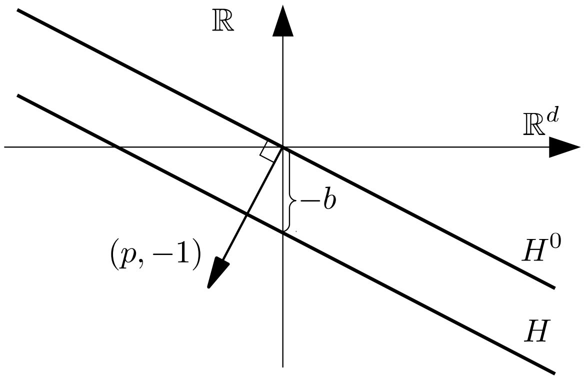

Рис. 11: Биекция между невертикальными гиперплоскостями в $\mathbb{R}^d \times \mathbb{R}$ и точками в $\mathbb{R}^{d*} \times \mathbb{R}$.

гиперплоскость H, помимо ковектора $p \in \mathbb{R}^{d*}$ необходимо еще знать число $b \in \mathbb{R}$, на которое необходимо поднять H, чтобы получившаяся гиперплоскость H^0 проходила через начало координат (см. рис. 11). Итак, $0 \in H^0$ и $H^0 \perp (p,-1)$, поэтому

$$H^0 = \{(x, \langle p, x\rangle),\ x \in \mathbb{R}^d\} \subset \mathbb{R}^{d+1}.$$

Значит,

$$H = H^0 - (0,b) = \{(x, \langle p, x\rangle - b),\ x \in \mathbb{R}^d\} \subset \mathbb{R}^{d+1}.$$

Таким образом, множество невертикальных гиперплоскостей в $\mathbb{R}^{d+1}$ отождествляется с пространством $\mathbb{R}^{d*} \times \mathbb{R}$ с помощью пары $p \in \mathbb{R}^{d*}$ и $b \in \mathbb{R}$. Для краткости, невертикальную гиперплоскость в $\mathbb{R}^{d+1}$, отвечающую паре (p,b) будем по ходу доказательства обозначать через $H(p,b)$.

Определим теперь, при каких $p \in \mathbb{R}^{d*}$ и $b \in \mathbb{R}$ соответствующая гиперплоскость $H(p,b)$ лежит под графиком функции f. Для этого необходимо и достаточно, чтобы

$$\langle p, x\rangle - b \leqslant f(x) \quad \text{для всех} \quad x \in \mathbb{R}^d,$$

или, равносильно,

$$b \geqslant \sup_{x \in \mathbb{R}^d} (\langle p, x\rangle - f(x)) = f^*(p) \quad \Longleftrightarrow \quad (p,b) \in \operatorname{epi} f^*.$$

Таким образом, мы доказали, что все гиперплоскости $H(p,b)$, лежащие под графиком f, отвечают точкам (p,b) из надграфика f^*.

Вычислим теперь, как устроено пересечение всех верхних полупространств относительно невертикальных гиперплоскостей $H(p,b)$ при $(p,b) \in \operatorname{epi} f^*$. Итак, точка $(x,a) \in \mathbb{R}^{d+1}$ лежит над гиперплоскостью $H(p,b)$, если

$$a \geqslant \langle p, x\rangle - b.$$

Следовательно, точка (x,a) лежит над всеми гиперплоскостями $H(p,b)$ при $(p,b) \in \operatorname{epi} f^*$, если предыдущее неравенство выполнено для всех $(p,b) \in \operatorname{epi} f^*$. Заметим, что $b \geqslant f^*(p)$, поэтому проверять его можно только при $b = f^*(p)$. Итак,

$$a \geqslant \langle p, x\rangle - f^*(p) \quad \text{для всех} \quad p \in \mathbb{R}^{d*},$$

или, равносильно,

$$a \geqslant \sup_{p\in\mathbb{R}^{d*}} (\langle p, x\rangle - f^*(p)) = f^{**}(x) \qquad \Longleftrightarrow \qquad (x, a) \in \operatorname{epi} f^{**}.$$

Таким образом, множество $\operatorname{epi} f^{**} \subset \mathbb{R}^{d+1}$ есть пересечение всех верхних полупространств относительно невертикальных гиперплоскостей в $\mathbb{R}^{d+1}$, содержащих $\operatorname{epi} f$. Поэтому, очевидно, $\operatorname{epi} f \subset \operatorname{epi} f^{**}$, и, значит, $f^{**} \leqslant f$.

Пусть теперь f – выпуклая замкнутая собственная функция. Для того, чтобы доказать равенство $f = f^{**}$ достаточно получить обратное включение $\operatorname{epi} f^{**} \subset \operatorname{epi} f$. Для этого будем действовать аналогично доказательству теоремы 14. Пусть $(x_0, a_0) \notin \operatorname{epi} f$. Докажем, что $(x_0, a_0) \notin \operatorname{epi} f^{**}$. Для этого покажем, что точку (x_0, a_0) можно строго отделить от $\operatorname{epi} f$ невертикальной гиперплоскостью. Так как $\operatorname{epi} f$ – выпуклое замкнутое подмножество $\mathbb{R}^{d+1}$, то выполняются условия второй теоремы отделимости 3, которая гарантирует, что строго отделяющая гиперплоскость $\hat{H}$ существует, однако, априори, она может оказаться вертикальной. Покажем, что если гиперплоскость $\hat{H}$ вертикальна, то ее можно немного повернуть таким образом, что она перестанет быть вертикальной, но по-прежнему будет строго отделять (x_0, a_0) от $\operatorname{epi} f$.

Если гиперплоскость $\hat{H}$ вертикальна, то $\hat{H} \perp (p_0, 0)$ для некоторого $p_0 \in \mathbb{R}^{d*}$, т.е.

$$\sup_{(x,a)\in\operatorname{epi} f} (\langle p_0, x\rangle + 0\cdot a) < \langle p_0, x_0\rangle + 0\cdot a_0 \quad \Leftrightarrow \quad \langle p_0, x\rangle \leqslant b_0 \ \ \forall x \in \operatorname{dom} f,$$

где $b_0 < \langle p_0, x_0\rangle$. Следовательно, ковектор p_0 строго отделяет x_0 от $\operatorname{dom} f$.

Выберем теперь любую невертикальную гиперплоскость $H(p_1, b_1)$, лежащую под графиком f. Такая обязательно найдется, так как функция f – собственная. Действительно, подойдет, например, любая опорная гиперплоскость в любой точке $x_1 \in \operatorname{ri}\operatorname{dom} f$, определяемая любым элементом p_1 из субдифференциала $\partial f(x_1)$ (последний непуст по теореме 7). Итак,

$$f(x) \geqslant \langle p_1, x\rangle - b_1 \quad \text{для всех} \quad x \in \operatorname{dom} f.$$

Будем поворачивать гиперплоскость $H(p_1, b_1)$ в сторону вертикальной гиперплоскости $\hat{H}$, а именно: рассмотрим гиперплоскости $H_\lambda \stackrel{\text{def}}{=} H(p_1 + \lambda p_0, b_1 + \lambda b_0)$ при $\lambda > 0$. Поскольку $\langle p_0, x\rangle - b_0 \leqslant 0$ для всех $x \in \operatorname{dom} f$, то при $\lambda > 0$ получаем

$$f(x) \geqslant \langle p_1 + \lambda p_0, x\rangle - (b_1 + \lambda b_0) \quad \text{для всех} \quad x \in \operatorname{dom} f.$$

Поэтому гиперплоскость H_λ лежит под $\operatorname{epi} f$ при любом $\lambda > 0$.

С другой стороны, точка (x_0, a_0) лежит под гиперплоскостью H_λ при данном λ, если выполняется неравенство

$$a_0 < \langle p_1 + \lambda p_0, x_0\rangle - (b_1 + \lambda b_0) = \langle p_1, x_0\rangle - b_1 + \lambda(\langle p_0, x_0\rangle - b_0).$$

Заметим, что $\langle p_0, x_0\rangle - b_0 > 0$. Поэтому при достаточно большом λ неравенство будет выполняться, и мы найдем невертикальную гиперплоскость H_λ, строго отделяющую точку (x_0, a_0) от $\operatorname{epi} f$, что и требовалось. □

Пример 9. Теперь мы готовы к тому, чтобы вычислить сопряженную функцию к опорной. Пусть $C \subset \mathbb{R}^d$ – произвольное множество. Пусть сначала $C \neq \emptyset$. Мы знаем, что $s_C = s_{\operatorname{cl}\operatorname{conv} C}$ (см. предложение 23). Поэтому $s_{\operatorname{cl}\operatorname{conv} C} = \delta^*_{\operatorname{cl}\operatorname{conv} C}$.

Осталось сказать, что функция $\delta_{\mathrm{cl\,conv}\,C}$ является выпуклой, замкнутой и собственной, так как множество $\mathrm{cl\,conv}\,C$ выпукло, замкнуто и непусто. Поэтому по теореме Фенхеля-Моро $\delta^{**}_{\mathrm{cl\,conv}\,C} = \delta_{\mathrm{cl\,conv}\,C}$, т.е. $s^*_C = \delta_{\mathrm{cl\,conv}\,C}$. Если же $C = \emptyset$, то получаем тот же результат, так как $s_\emptyset \equiv -\infty$ и $s^*_\emptyset \equiv \infty = \delta_\emptyset$.

Выпуклая двойственность для функций дает множество очень сильных результатов. Например (ср. упражнение 8.4):

***Предложение* 47.** *Пусть f_1 и f_2 – выпуклые замкнутые собственные функции, $f_1(0) = f_2(0) = 0$. Если f_1 и f_2 положительно однородны, то*

$$f_1 = f_2 \quad \Longleftrightarrow \quad \partial f_1(0) = \partial f_2(0).$$

Доказательство. Из определения сопряженной функции легко следует, что $f_i^* = \delta_{\partial f_i(0)}$, $i = 1,2$. Также мы доказали, что $\delta^*_A = s_A$ для любого множества A. Поэтому $f_i = f_i^{**} = s_{\partial f_i(0)}$. □

Опишем теперь, как устроена сопряженная функция f^* для произвольной функции f.

***Предложение* 48.** *Пусть $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ – произвольная функция. Тогда функция f^* является замкнутой, выпуклой и либо собственной, либо $f^* \equiv \infty$, либо $f^* \equiv -\infty$. Если дополнительно функция $\mathrm{conv} f$ является собственной, то функция f^* также является собственной.*

Доказательство. Пусть $x \in \mathbb{R}^d$ и $a \in \mathbb{R}$. Обозначим через $l_{x,a} : \mathbb{R}^{d*} \to \mathbb{R}$ – аффинную функцию

$$l_{x,a}(p) = \langle p, x\rangle - a.$$

Тогда, по определению сопряженной функции, f^* есть поточечный супремум аффинных функций $l_{x,a}$ по всем $(x,a) \in \mathrm{epi} f$. Надграфик f^* есть пересечение надграфиков $l_{x,a}$. Поскольку аффинные функции $l_{x,a}$ выпуклы и замкнуты, то такой же является функция f^*.

Далее: очевидно, что если $\mathrm{epi} f \neq \emptyset$, то $f^*(p) \neq -\infty$ ни для какого p. Поэтому если $\mathrm{epi} f \neq \emptyset$, то либо функция f^* является собственной, либо $f^* \equiv \infty$. Если же $\mathrm{epi} f = \emptyset$, то $f^* \equiv -\infty$.

Ну и, наконец, если функция $\mathrm{conv} f$ – собственная, то f^* не равна $-\infty$ ни для какого p. Осталось убедиться в том, что множество $\mathrm{epi} f^*$ непусто. Другими словами, необходимо найти невертикальную гиперплоскость, лежащую под $\mathrm{epi} f$. Для этого можно, например, воспользоваться теоремой 7, которая гарантирует непустоту субдифференциала выпуклой собственной функции $\mathrm{conv} f$ в точках $\mathrm{ri\,dom\,conv} f$. □

Обозначим через $\mathscr{C}(\mathbb{R}^d)$ множество выпуклых собственных замкнутых функций на $\mathbb{R}^d$.

***Теорема* 16.** *Операция сопряжения $f \mapsto f^*$ является биекцией между $\mathscr{C}(\mathbb{R}^d)$ и $\mathscr{C}(\mathbb{R}^{d*})$.*

Доказательство. Если $f \in \mathscr{C}(\mathbb{R}^d)$, то $f^* \in \mathscr{C}(\mathbb{R}^d)$ по предложению 48. Значит, отображение $f \mapsto f^*$ переводит множество $\mathscr{C}(\mathbb{R}^d)$ в $\mathscr{C}(\mathbb{R}^{d*})$. Это отображение является биекцией, так как имеет обратное – действительно, $f^{**} = f$. □

***Предложение* 49.** $f^* = (\mathrm{conv} f)^* = (\mathrm{cl} f)^* = (\mathrm{cl\,conv} f)^*$.

Доказательство. Равенства $f^* = (\mathrm{conv} f^*)$ и $f^* = (\mathrm{cl} f)^*$ следуют из того, что (замкнутое) полупространство содержит $\mathrm{epi} f$ тогда и только тогда, когда оно содержит $\mathrm{conv}\,\mathrm{epi} f$ и $\mathrm{cl}\,\mathrm{epi} f$. Последнее равенство $(\mathrm{conv} f)^* = (\mathrm{cl}\,\mathrm{conv} f)^*$ получается подстановкой $g = \mathrm{conv} f$. □

Вычислим теперь вторую сопряженную к выпуклой, но не обязательно замкнутой функции. Следующее утверждение тоже иногда называют теоремой Фенхеля-Моро:

***Теорема* 17.** *Пусть $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ – выпуклая собственная функция. Тогда $f^{**} = \mathrm{cl} f$.*

Доказательство. Действительно, $f^* = (\mathrm{cl} f)^*$, а функция $\mathrm{cl} f$ является собственной (см. упражнение 7.3) и, очевидно, замкнутой. Поэтому

$$f^{**} = ((\mathrm{cl} f)^*)^* = (\mathrm{cl} f)^{**} = \mathrm{cl} f.$$

□

Поляра. Сформулируем на алгебраическом языке двойственность для выпуклых множеств, содержащих начало координат. Как и в случае двойственности для выпуклых функций, нам потребуется избавиться от части заведомо не нужных полупространств. Основное соображение заключается в следующем: если множество A содержит начало координат и лежит в полупространстве H^+, то и начало координат лежит в H^+. Поэтому ограничимся только такими полупространствами H^+, что $0 \in \mathrm{int}\, H^+$. Такое полупространство можно представить в виде

$$H^+ = \{x \in \mathbb{R}^d : \ \langle p, x\rangle \leqslant 1\},$$

где $p \in \mathbb{R}^{d*}$ – ортогональный границе H^+ ковектор, имеющий нужные направление и длину.

Определение 20. *Полярой* множества $A \subset \mathbb{R}^d$ называется множество

$$A^\circ \overset{\text{def}}{=} \{p \in \mathbb{R}^{d*} : s_A(p) \leqslant 1\} = \{p \in \mathbb{R}^{d*} : \ \langle p, x\rangle \leqslant 1 \ \ \forall x \in A\} \subset \mathbb{R}^{d*}.$$

Двойственность для выпуклых множеств, содержащих начало координат, формулируется следующим образом:

***Теорема* 18** (о биполяре)**.** *Пусть A – выпуклое замкнутое подмножество $\mathbb{R}^d$ и $0 \in A$. Тогда $A^{\circ\circ} = A$.*

Доказательство. Итак, поляра A° описывает все полупространства H^+, содержащие A и такие, что $0 \in \mathrm{int}\, H^+$, а множество $A^{\circ\circ}$ есть пересечение этих полупространств. Поэтому по сравнению с теоремой 14 мы выкинули только полупространства, содержащие начало координат на своей границе. Таким образом, нам достаточно показать, что любое такое полупространство есть пересечение полупространств, содержащих 0 в своей внутренности.

Пусть H^+ – полупространство и 0 лежит на его границе. Тогда для некоторого ковектора $p \neq 0$ имеем $H^+ = \{x : \langle p, x\rangle \geqslant 0\}$. Пусть $\lambda < 0$. Рассмотрим полупространства

$$H^+_\lambda = \{x : \ \langle p, x\rangle \geqslant \lambda\} \supset H^+.$$

Тогда $0 \in \mathrm{int}\, H^+_\lambda$ при $\lambda < 0$ и $\bigcap_{\lambda<0} H^+_\lambda = H^+$, что и требовалось.

□

Аналогично предложению 48 доказывается, что поляра A° является выпуклым замкнутым множеством и $0 \in A^\circ$ независимо от A. Действительно, поляра A° есть пересечение полупространств $\{p : \langle p, x\rangle \leqslant 1\} \subset \mathbb{R}^{d*}$ по всем $x \in A$. Поэтому

***Теорема* 19.** *Операция взятия поляры $A \mapsto A^\circ$ является биекцией между множеством всех выпуклых замкнутых содержащих начало координат подмножеств $\mathbb{R}^d$ и множеством таких же подмножеств в $\mathbb{R}^{d*}$.*

По этой теореме несложно вычислить биполяру произвольного множества (см. упражнение 8.7).

Двойственные конусы. Займемся теперь двойственностью выпуклых конусов. Напомним, что множество $K \subset \mathbb{R}^d$ называется выпуклым конусом, если оно выпукло, непусто и вместе с любой точкой $x \in K$ содержит целиком луч $\lambda x \in K$, $\lambda \geqslant 0$.

Вычислим поляру конуса. Пусть $p \in \mathbb{R}^d$. Если $\langle p, x\rangle > 0$ для какого-то $x \in K$, то $\sup_{x\in K}\langle p, x\rangle = +\infty$. Поэтому, если $\langle p, x\rangle \leqslant 1$ для всех x из конуса K, то $\langle p, x\rangle \leqslant 0$ для всех $x \in K$.

Определение 21. *Сопряженным конусом* к конусу $K \subset \mathbb{R}^d$ называется его поляра:

$$K^* \stackrel{\text{def}}{=} K^\circ = \{p \in \mathbb{R}^{d*} : s_K(p) \leqslant 0\} = \{p \in \mathbb{R}^{d*} : \ \langle p, x\rangle \leqslant 0 \ \ \forall x \in K\} \subset \mathbb{R}^{d*}.$$

Тривиально проверяется, что сопряженный конус K^* действительно является конусом.

Таким образом, если выкинуть 0 из K^*, то оставшиеся ковекторы $p \in K^*$ описывают ровно те гиперплоскости $H(p) = \{x : \langle p, x\rangle = 0\}$, для которых $K \subset H^-(p) = \{x : \langle p, x\rangle \leqslant 0\}$. Отметим, что если L – линейное подпространство (рассматриваемое как конус), то $L^* = L^\perp$.

Выпуклая двойственность для конусов формулируется так:

***Теорема* 20.** *Если K – выпуклый замкнутый конус, то $K^{**} = K$.*

Доказательство. Действительно, по теореме о биполяре, $K^{**} = K^{\circ\circ} = K$. □

Сопряженный конус K^* является выпуклым и замкнутым независимо от исходного конуса K, так как таковой является поляра любого множества. Поэтому

***Теорема* 21.** *Операция взятия сопряжения $K \mapsto K^*$ является биекцией между множеством выпуклых замкнутых конусов в $\mathbb{R}^d$ и множеством выпуклых замкнутых конусов в $\mathbb{R}^{d*}$.*

Упражнения:

8.1 Докажите, что результат теоремы 14 о геометрии выпуклой двойственности верен только для выпуклых замкнутых множеств.

8.2 Остается ли верной теорема 14 о геометрии выпуклой двойственности, если вместо замкнутых полупространств рассмотреть открытые?

8.3 Доказать, что любое относительно открытое выпуклое множество совпадает с пересечением открытых полупространств, его содержащих.

8.4 Попробуйте доказать предложение 47, не используя теорему Фенхеля-Моро.

8.5 Пусть $f = f^{**}$. Что можно сказать об f?

8.6 Докажите теорему Минковского о том, что любая выпуклая замкнутая собственная функция $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ есть поточечный супремум аффинных функций $l_{p,b}(x) = \langle p, x\rangle - b$, $p \in \mathbb{R}^{d*}$, $b \in \mathbb{R}$, ее не превосходящих.

8.7 Докажите, что для любого множества A выполнено $A^{\circ\circ} = \mathrm{cl}\,\mathrm{conv}(A \cup \{0\})$.

8.8 Пусть $C \subset \mathbb{R}^3$ – правильный многогранник и 0 – его центр. Докажите, что поляра C° есть двойственный к C многогранник: тетраэдр $\leftrightarrow$ тэтраэдр, куб $\leftrightarrow$ октаэдр, додекаэдр $\leftrightarrow$ икосаэдр.

8.9 Пусть $A \subset \mathbb{R}^d$ – открытое выпуклое множество и $0 \notin A$. Докажите, что если A не содержит аффинных подпространств размерности 1 или больше, то найдутся такой ковектор p и число $c > 0$, что $\langle p, \frac{x}{|x|}\rangle > c$ для всех $x \in A$.

9 Субдифференциальное исчисление

Субдифференциал сопряженной функции. Из определения сопряженной функции немедленно следует, что для любых p и x выполнено $f^*(x) \geqslant \langle p, x\rangle - f(x)$. Если перенести $f(p)$ направо, получится *неравенство Юнга*:

$$f(x) + f^*(p) \geqslant \langle p, x\rangle.$$

Пример 10. Самый известный пример использования неравенства Юнга такой:

$$\frac{x^a}{a} + \frac{p^b}{b} \geqslant px$$

для любых неотрицательных чисел x, p, положительных a, b и $\frac{1}{a} + \frac{1}{b} = 1$. Проще всего доказательство получить так: обозначим $f(x) = \frac{|x|^a}{a}$, $x \in \mathbb{R}$. Тогда f – выпуклая замкнутая функция и

$$f^*(p) = \sup_x (px - |x|^a/a).$$

Супремум достигается в точке, где производная равна 0, т.е. $p = |x|^{a-1}\operatorname{sign} x$. Поэтому $x = |p|^{\frac{1}{a-1}} \operatorname{sign} p$ и $f^*(p) = \frac{|p|^b}{b}$.

Как видно из этого примера, вычисление сопряженной функции имеет прямое отношение к субдифференциалу самой функции. Эту связь удобнее всего записать в терминах равенства в неравенстве Юнга:

***Предложение* 50.** *Пусть $x \in \mathbb{R}^d$ и $p \in \mathbb{R}^{d*}$. Тогда независимо от выпуклости f имеем*

$$p \in \partial f(x) \qquad \Longleftrightarrow \qquad f(x) + f^*(p) = \langle p, x\rangle.$$

Доказательство. Если $p \in \partial f(x)$, то для любого y выполнено $f(y) - \langle p, y\rangle \geqslant f(x) - \langle p, x\rangle$, или $\langle p, y\rangle - f(y) \leqslant \langle p, x\rangle - f(x)$. Поэтому получаем неравенство

$$f^*(p) = \sup_y(\langle p, y\rangle - f(y)) \leqslant \langle p, x\rangle - f(x),$$

которое обратно неравенству Юнга (слагаемые те же, но знак противоположный). Поэтому в неравенстве Юнга должен стоять знак равенства.

Обратно: предположим, для некоторых x и p выполнено $f(x)+f^*(p) = \langle p, x\rangle$. По неравенству Юнга для любого y выполнено $f(y) + f^*(p) \geqslant \langle p, y\rangle$. Вычитая из неравенства равенство, получаем $f(y) - f(x) \geqslant \langle p, y - x\rangle$, т.е. $p \in \partial f(x)$. $\square$

Это предложение дает немедленный ответ на следующий вопрос: когда в неравенстве $\frac{x^a}{a} + \frac{p^b}{b} \geqslant px$ возможно равенство. Естественно, только когда $p = f'(x)$, что равносильно $x^a = p^b$.

Заметим, что доказанное предложение несимметрично относительно f и f^*, тогда как само неравенство Юнга симметрично, если функция f замкнута, так как в этом случае работает теорема Фенхеля-Моро. На самом деле, из неравенства Юнга действительно следует очень удобный симметричный результат о связи субдифференциалов f и f^*:

***Теорема* 22** (о связи субдифференциалов функции и ее сопряженной)**.** *Предположим, собственная выпуклая функция $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ полунепрерывна снизу в точке x. Тогда*

$$p \in \partial f(x) \quad \Longleftrightarrow \quad x \in \partial f^*(p).$$

Доказательство. По теореме Фенхеля-Моро $f^{**} = \mathrm{cl} f$, но функция f полунепрерывна в x, поэтому $f(x) = \mathrm{cl} f(x) = f^{**}(x)$. Дважды воспользовавшись предложением 50, получаем

$$p \in \partial f(x) \;\Leftrightarrow\; f(x) + f^*(p) = \langle p, x\rangle \;\Leftrightarrow\; f^{**}(x) + f^*(p) = \langle p, x\rangle \;\Leftrightarrow\; x \in \partial f^*(p).$$

$\square$

***Следствие* 22.** *Пусть $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ – собственная выпуклая функция. Тогда*

$$p \in \partial f(x) \quad \Longrightarrow \quad x \in \partial f^*(p).$$

Доказательство. Формально мы отказались от полунепрерывности f в x. Однако, если субдифференциал f в точке x непуст, то функция f обязана быть полунепрерывной снизу в x. $\square$

Теорема 22 оказывается очень полезной при вычислении субдифференциала некоторых функций. Например, на нее можно посмотреть под таким углом:

***Следствие* 23.** *Предположим, что выпуклая собственная функция $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ полунепрерывна снизу в точке x. Тогда*

$$\partial f(x) = \{p \in \mathbb{R}^{d*} : x \in \partial f^*(p)\}.$$

Таким образом, мы получили инструмент, позволяющий вычислять субдифференциал самой функции по ее сопряженной и, естественно, наоборот.

Преобразование Лежандра. На самом деле, мы построили очень мощный аппарат, называемый обычно *преобразованием Лежандра*. Мы разберем самый простую ситуацию – пусть обе выпуклые функции f и f^* является гладкими. В этом случае мы можем устроить переход от пространства $\mathbb{R}^d$ к пространству $\mathbb{R}^{d*}$ с помощью замены: $p(x) = f'(x)$. Тогда обратная замена задается равенством $x(p) = f^{*\prime}(p)$. Якобианы этих замен обратны: $\frac{\partial p}{\partial x} = (\frac{\partial x}{\partial p})^{-1}$, и их несложно вычислить:

$$\frac{\partial p}{\partial x} = f''(x) \qquad \text{и} \qquad \frac{\partial x}{\partial p} = f^{*\prime\prime}(p).$$

Поэтому, в частности, $f''(x) = (f^{*\prime\prime}(p))^{-1}$ для x и p, удовлетворяющих равенству $f(x) + f^*(p) = \langle p, x\rangle$.

Отметим, что даже в случае, когда $\operatorname{dom} f = \mathbb{R}^d$ и $f'' > 0$, вообще говоря, неверно, что $\operatorname{dom} f^* = \mathbb{R}^{d*}$. Простейший пример $f = \sqrt{1+x^2}$, $x \in \mathbb{R}$, дает $\operatorname{dom} f^* = [-1; 1]$. Проблема эта возникает потому, что хоть $f'' > 0$ на всем пространстве, тем не менее f'' нельзя отделить от нуля, а именно: $0 < f'' \leqslant 1/2^{\frac{3}{2}}$. Получается, что вторая производная сопряженной функции $f^{*\prime\prime} = f''^{-1}$ имеет обратные ограничения $2^{\frac{3}{2}} \leqslant f^{*\prime\prime} < \infty$ и может становиться сколь угодно большой. Ограничения сверху на вторую производную исходной функции приводят к ограничениям снизу на вторую производную сопряженной функции, и наоборот. Подобного рода соображения в итоге приводят к теории двойственности понятий строгой выпуклости функции и гладкости ее сопряженной. Подробнее см. параграф 11.

Преобразование Лежандра оказывается также очень удобным в случае, когда выпуклая функция зависит от параметров. Предположим, что $f(x, a)$, $x \in \mathbb{R}^d$, $a \in \mathbb{R}^m$, выпукла по x и $f^*(p, a)$ – ее сопряженная по x при данном a, т.е. $p(x, a) = f'_x(x, a)$. Разрешая это уравнение относительно x, находим обратную замену $x = x(p, a)$, и, следовательно, $f^*(p, a) = \langle p, x(p, a)\rangle - f(x(p, a), a)$. В этом случае

$$f_a^{*\prime} = \langle p, x'_a\rangle - \langle f'_x, x'_a\rangle - f'_a = -f'_a.$$

В свое время Гамильтон заметил, что это соображение оказывается невероятно полезным при исследовании уравнения Эйлера-Лагранжа и позволяет преобразовать обыкновенное дифференциальное уравнение второго порядка к очень удобной канонической (гамильтоновой) форме, обладающей огромным количеством алгебраических и геометрических свойств.

Двойственность опорной функции и функции Минковского. Как мы выяснили, субдифференциалы функции и ее сопряженной тесно связаны. Вычислив субдиффиренциал самой функции, мы немедленно найдем субдифференциал ее сопряженной, и наоборот. Начнем с основных функций: индикаторной δ_C, опорной s_C и функции Минковского μ_C.

Мы уже выяснили, что $\delta_C^* = s_C$. Поэтому и субдифференциал опорной функции находится очень просто. Отметим, что если $C \subset \mathbb{R}^d$, то опорная функция s_C определена на сопряженном пространстве $\mathbb{R}^{d*}$, и поэтому ее субдифференциал в точке есть подмножество $\mathbb{R}^{d**} = \mathbb{R}^d$. Для опорной функции наибольшее значение имеет субдифференциал в 0. Однако, его несложно вычислить и в остальных случаях:

***Предложение* 51.** *Пусть $C \neq \emptyset$ – выпуклое замкнутое множество. Тогда для любого ковектора $p \in \mathbb{R}^{d*}$ выполнено*

$$\partial s_C(p) = \arg\max_{x\in C}\langle p, x\rangle.$$

Доказательство. Функция δ_C является собственной, выпуклой и замкнутой, и $\delta_C^* = s_C$. Поэтому $s_C^* = \delta_C$ по теореме Фенхеля-Моро, и по предложению 50 мы получаем

$$x \in \partial s_C(p) \qquad \Leftrightarrow \qquad s_C(p) = \langle p, x\rangle - \delta_C(x).$$

Поскольку $s_C = \delta_C^*$, то это равенство возможно, если и только если

$$x \in \arg\max_{y}(\langle p, y\rangle - \delta_C(y)) = \arg\max_{y\in C}\langle p, y\rangle,$$

что и требовалось. □

Если же множество C не является выпуклым или замкнутым, то субдифференциал опорной функции s_C легко находится с помощью предложения 23. Например,

***Следствие* 24.** *Для любого непустого множества C выполнено $\partial s_C(0) = \operatorname{cl}\operatorname{conv} C$.*

Доказательство. Поскольку $s_C = s_{\operatorname{cl}\operatorname{conv} C}$, то, применяя предложение 51 к выпуклому замкнутому множеству $\operatorname{cl}\operatorname{conv} C$, получаем

$$\partial s_C(0) = \partial s_{\operatorname{cl}\operatorname{conv} C}(0) = \arg\max_{x\in\operatorname{cl}\operatorname{conv} C}\langle 0, x\rangle = \operatorname{cl}\operatorname{conv} C.$$

□

Теперь вычислим субдифференциал функции Минковского с помощью опорной функции поляры множества.

***Предложение* 52.** *Пусть $A \subset \mathbb{R}^d$. Предположим, что опорная функция s_A множества A неотрицательна: $\forall p \in \mathbb{R}^{d*}$ $s_A(p) \geqslant 0$ (например, $0 \in A$). Тогда $s_A = \mu_{A^\circ}$.*

Доказательство. Очевидно, $s_A(0) = \mu_{A^\circ}(0) = 0$. Для $p \neq 0$ вычислим

$$\mu_{A^\circ}(p) = \inf\{\lambda > 0 : p \in \lambda A^\circ\} = \inf\{\lambda > 0 : \sup_{x\in A}\langle p, x\rangle \leqslant \lambda\} = \\ = \inf\{\lambda > 0 : s_A(p) \leqslant \lambda\} = s_A(p).$$

Последнее равенство выполнено, так как $s_A(p) \geqslant 0$ по условию. □

***Следствие* 25.** *Если множество C выпукло, замкнуто и $0 \in C$, то $\mu_C = s_{C^\circ}$.*

Доказательство. Действительно, $s_{C^\circ} = \mu_{C^{\circ\circ}}$ и $C^{\circ\circ} = C$ по теореме о биполяре. □

***Следствие* 26.** *Если множество C выпукло, замкнуто и $0 \in C$, то*

$$\partial\mu_C(x) = \arg\max_{p\in C^\circ}\langle p, x\rangle,$$

Например, $\partial\mu_C(0) = C^\circ$.

Доказательство. Действительно, $\mu_C = s_{C^\circ}$. Остается вычислить субдифференциал опорной функции по предложению 51. □

Унарные линейные операции. Итак, для построения субдифференциального исчисления нам необходимо провести вычисления лишь для половины операций – остальные получаются с помощью следствия 23. Следующее предложение получается тривиально из определения сопряженной функции:

Предложение 53. *Пусть $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ и $c \in \mathbb{R}$. Тогда $(f + c)^* = f^* - c$.*

Займемся теперь сдвигом аргумента:

Предложение 54. *Пусть $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ – некоторая функция и $p_0 \in \mathbb{R}^d$. Обозначим $g(x) = f(x) + \langle p_0, x\rangle$. Тогда*

$$g^*(p) = f^*(p - p_0).$$

Обратно: если $\xi_0 \in \mathbb{R}^d$ и $h(x) = f(x - \xi_0)$, то

$$h^*(p) = f^*(p) + \langle p, \xi_0\rangle.$$

Доказательство. С помощью прямого вычисления получаем:

$$g^*(p) = \sup_x(\langle p, x\rangle - g(x)) = \sup_x(\langle p, x\rangle - \langle p_0, x\rangle - f(x)) = f^*(p - p_0);$$

$$h^*(p) = \sup_x(\langle p, x\rangle - h(x)) = \sup_x(\langle p, x\rangle - f(x - \xi_0)) = \\ = \sup_y(\langle p, y + \xi_0\rangle - f(y)) = f^*(p) + \langle p, \xi_0\rangle.$$

□

Следствие 27. *Если $g(x) = f(x - \xi_0) + \langle p_0, x\rangle$, то $g^*(p) = f^*(p - p_0) + \langle p, \xi_0\rangle$.*

Следствие 28. *Если $g(x) = f(x) + \langle p_0, x\rangle$, то $\partial g(x) = \partial f(x) + p_0$.*

Доказательство. Доказательство этого факта можно получить, проведя простую прямую выкладку. Однако, мы предпочтем продемонстрировать работу неравенства Юнга. Итак,

$$p \in \partial f(x) \quad \Leftrightarrow \quad f(x) + f^*(p) = \langle p, x\rangle$$

согласно предложению 50. Поэтому

$$p \in \partial g(x) \;\Leftrightarrow\; g(x) + g^*(p) = \langle p, x\rangle \;\Leftrightarrow\; f(x) + \langle p_0, x\rangle + f^*(p - p_0) = \langle p, x\rangle \;\Leftrightarrow \\ \Leftrightarrow\; f(x) + f^*(p - p_0) = \langle p - p_0, x\rangle \;\Leftrightarrow\; p - p_0 \in \partial f(x),$$

что и требовалось. □

Таким образом, операция сдвига аргумента полностью двойственна операции добавления к функции линейной. Для умножения на константу получаем

Предложение 55. *Пусть $\lambda > 0$. Тогда $(\lambda f)^* = f^*\lambda$ и $(f\lambda)^* = \lambda f^*$.*

Доказательство.

$$(\lambda f)^*(p) = \sup_x(\langle p, x\rangle - \lambda f(x)) = \lambda \sup_x(\langle \frac{p}{\lambda}, x\rangle - f(x)) = \lambda f^*(p/\lambda) = (f^*\lambda)(p);$$

$$(f\lambda)^*(p) = \sup_x(\langle p, x\rangle - \lambda f(\frac{x}{\lambda})) = \sup_y(\langle p, \lambda y\rangle - \lambda f(y)) = \lambda f^*(p).$$

□

Следствие 29. *$\partial(\lambda f)(x) = \lambda\partial f(x)$, и поэтому $\partial(f\lambda)(x) = \partial f(x/\lambda)$.*

Для линейной замены $x \mapsto Ax$ получаем

Предложение 56. *Пусть $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ – некоторая функция и $A : \mathbb{R}^d \to \mathbb{R}^{d'}$ – линейное отображение. Тогда $(Af)^* = f^*A^*$. Если $g : \mathbb{R}^{d'} \to \bar{\mathbb{R}}$ – выпуклая собственная функция, то $(A^*g^*)^* = (\mathrm{cl}\, g)A$.*

Доказательство. Будем обозначать символом x точки из $\mathbb{R}^d$, а символом y – точки из $\mathbb{R}^{d'}$, чтобы не путать эти пространства. Пусть $q \in \mathbb{R}^{d'*}$. С помощью прямой выкладки находим

$$(Af)^*(q) = \sup_y(\langle q, y\rangle - \inf_{x:Ax=y} f(x)) = \sup_{x,y:Ax=y}(\langle q, Ax\rangle - f(x)) = \\ = \sup_x(\langle A^*q, x\rangle - f(x)) = f^*(A^*q),$$

и первое равенство доказано.

Для того, чтобы получить второе равенство, подставим в первое g^* вместо f и A^* вместо A. Получим $(A^*g^*)^* = g^{**}A^{**} = (\mathrm{cl}\, g)A$, т.к. $A^{**} = A$ всегда и $g^{**} = \mathrm{cl}\, g$ для собственной выпуклой функции g по теореме Фенхеля-Моро. □

Предложение 57. *Пусть $g : \mathbb{R}^{d'} \to \bar{\mathbb{R}}$ – некоторая функция и $A : \mathbb{R}^d \to \mathbb{R}^{d'}$ – линейный оператор. Тогда если оператор A сюрьективен, то $(gA)^* = A^*g^*$.*

Доказательство. Вычислим сначала $(gA)^*(p)$ для $p \in \mathbb{R}^{d*}$:

$$(gA)^*(p) = \sup_{x\in\mathbb{R}^d}(\langle p, x\rangle - g(Ax)).$$

Легко видеть, что для любого $h \in \ker A$ выполнено $g(A(x+h)) = g(Ax)$. Поэтому если $\langle p, h\rangle \neq 0$ для какого-то $h \in \ker A$, то супремум в правой части равен ∞. Если же $p \in (\ker A)^\perp = \mathrm{Im}\, A^*$, то существует ковектор q такой, что $A^*q = p$. Отметим, что оператор A^* инъективен, поэтому такой ковектор q единственен, и мы будем его обозначать $(A^*)^{-1}p$. Итак, если $p \in \mathrm{Im}\, A^*$, то

$$(gA)^*(p) = \sup_{x\in\mathbb{R}^d}(\langle (A^*)^{-1}p, Ax\rangle - g(Ax)).$$

Оператор A сюрьективен, поэтому Ax пробегает все пространство $\mathbb{R}^{d'}$, и, следовательно, справа стоит просто $g^*((A^*)^{-1}p)$. Итак,

$$(gA)^*(p) = \begin{cases} g^*((A^*)^{-1}p), & \text{если } p \in \mathrm{Im}\, A^*; \\ \infty, & \text{иначе.} \end{cases}$$

Вычислим теперь $(A^*g^*)(p)$:

$$(A^*g^*)(p) = \inf_{q:A^*q=p} g^*(q) = \begin{cases} g^*((A^*)^{-1}p), & \text{если } p \in \mathrm{Im}\, A^*; \\ \infty, & \text{иначе;} \end{cases} = (gA)^*(p).$$

□

***Следствие* 30.** *В условиях предыдущего предложения работает правило сложной функции:* $\partial(gA)(x) = A^*\partial g(Ax)$.

Доказательство. $p \in \partial(gA)(x)$, если и только если $(gA)(x) + (gA)^*(p) = \langle p, x\rangle$. Таким образом, оба слагаемых в правой части конечны, поэтому $p \in \operatorname{Im} A^*$ и $(gA)^*(p) = (A^*g^*)(p) = g^*((A^*)^{-1}p)$, т.е.

$$g(Ax) + g^*((A^*)^{-1}p) = \langle p, x\rangle = \langle (A^*)^{-1}p, Ax\rangle,$$

и поэтому $(A^*)^{-1}p \in \partial g(Ax)$. Обратное доказательство получается аналогично. $\square$

***Следствие* 31.** *Если* $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ *– собственная выпуклая замкнутая функция и* $A : \mathbb{R}^d \to \mathbb{R}^{d'}$ *– инъективный оператор, то*

$$\partial(Af)(y) = \begin{cases} (A^*)^{-1}\partial f^*(A^{-1}y), & \text{если } y \in \operatorname{Im} A; \\ \emptyset, & \text{иначе.} \end{cases}$$

Доказательство. Поскольку оператор A инъективен, то $\operatorname{epi} Af = A \operatorname{epi} f$, и, следовательно, функция Af является выпуклой, замкнутой и собственной. Поэтому, по теореме 22 о связи субдифференциалов функции и ее сопряженной, получаем

$$\partial(Af)(y) = \{q : y \in \partial(Af)^*(q)\} = \{q : y \in \partial(f^*A^*)(q)\} = \{q : y \in A\partial f^*(A^*q)\}.$$

Таким образом, если $y \notin \operatorname{Im} A$, то $\partial(Af)(y) = \emptyset$. В противном случае,

$$\partial(Af)(y) = \{q : A^{-1}y \in \partial f^*(A^*q)\} = \\ = (A^*)^{-1}\{p : A^{-1}y \in \partial f^*(p)\} = (A^*)^{-1}\partial f(A^{-1}y).$$

$\square$

Субдифференциал суммы и инфимальной конволюции. Как мы покажем чуть ниже, операции $+$ и $\square$ двойственны друг другу. Поэтому достаточно вычислить субдифференциал одной, чтобы получить субдифференциал другой.

***Теорема* 23** (Моро-Рокафеллара)**.** *Пусть* $f_1, f_2 : \mathbb{R}^d \to \bar{\mathbb{R}}$ *– выпуклые собственные функции и* $x_0 \in \mathbb{R}^d$. *Тогда* $\partial(f_1 + f_2)(x_0) \supset \partial f_1(x_0) + \partial f_2(x_0)$. *Если же дополнительно относительные внутренности эффективных областей* f_1 *и* f_2 *пересекаются,* $\operatorname{ri}\operatorname{dom} f_1 \cap \operatorname{ri}\operatorname{dom} f_2 \neq \emptyset$, *то включение превращается в равенство:* $\partial(f_1 + f_2)(x_0) = \partial f_1(x_0) + \partial f_2(x_0)$.

Доказательство. Первая часть теоремы доказывается напрямую по определению субдифференциала. Пусть $p_1 \in \partial f_1(x_0)$ и $p_2 \in \partial f_2(x_0)$. Тогда для любого $x \in \mathbb{R}^d$ получаем:

$$f_1(x) + f_2(x) \geqslant \langle p_1, x - x_0\rangle + f_1(x_0) + \langle p_2, x - x_0\rangle + f_2(x_0) = \\ = \langle p_1 + p_2, x - x_0\rangle + (f_1 + f_2)(x_0),$$

т.е. $p_1 + p_2 \in \partial(f_1 + f_2)(x_0)$, и включение $\partial(f_1 + f_2)(x_0) \supset \partial f_1(x_0) + \partial f_2(x_0)$ доказано.

Перейдем ко второй части теоремы. Здесь мы для любого $p \in \partial(f_1 + f_2)(x_0)$ должны доказать существование такого разложения $p = p_1 + p_2$, что $p_i \in \partial f_i(x_0)$, $i = 1, 2$. Теоремы отделимости дают мощный инструмент для доказательства существования ковекторов. Необходимо лишь подобрать «правильные» не пересекающиеся выпуклые множества.

Разберем сначала самый важный случай: $p = 0$. В этом случае требуется найти такой ковектор p_0, что $p_0 \in \partial f_1(x_0)$ и $-p_0 \in \partial f_2(x_0)$. Заменяя, если потребуется, f_1 на $f_1 + \mathrm{const}$, будем считать, что $f_1(x_0) + f_2(x_0) = 0$. Тогда $f_1(x) + f_2(x) \geqslant f_1(x_0) + f_2(x_0) + \langle p, x - x_0\rangle = 0$ для всех $x \in \mathbb{R}^d$. Рассмотрим два множества: $A_1 = \operatorname{epi} f_1 = \{(x, a) : a \geqslant f_1(x)\}$ и $A_2 = \{(x, a) : a \leqslant -f_2(x)\}$. Множество A_2 есть отражение надграфика $\operatorname{epi} f_2$ относительно горизонтальной гиперплоскости $a = 0$. Функции f_1 и f_2 выпуклы, поэтому выпуклы множества A_1 и A_2.

Воспользуемся характеризацией относительной внутренности надграфика выпуклой функции (см. следствие 11), согласно которой $\operatorname{ri} \operatorname{epi} f_i = \{(x, a) : x \in \operatorname{ri} \operatorname{dom} f_i,\ a > f_i(x)\}$. Поэтому

$$\operatorname{ri} A_1 = \{(x, a) : x \in \operatorname{ri} \operatorname{dom} f_1 \text{ и } a > f_1(x)\};$$
$$\operatorname{ri} A_2 = \{(x, a) : x \in \operatorname{ri} \operatorname{dom} f_2 \text{ и } a < -f_2(x)\}.$$

Из этого описания немедленно получаем, что относительные внутренности A_1 и A_2 не пересекаются. Действительно, если $(x, a) \in \operatorname{ri} A_1 \cap \operatorname{ri} A_2$, то $f_1(x) < a < -f_2(x)$, что невозможно, т.к. $f_1 + f_2 \geqslant 0$.

Таким образом, выпуклые множества A_1 и A_2 можно собственно разделить по первой теореме отделимости: найдутся такие не равные одновременно 0 ковектор $p_0 \in \mathbb{R}^d$ и число $b_0 \in \mathbb{R}$, что

$$\sup_{(x,a)\in A_1} (\langle p_0, x\rangle + a b_0) \leqslant \inf_{(x,a)\in A_2} (\langle p_0, x\rangle + a b_0). \tag{4}$$

Подчеркнем, что разделение собственное.

Пара (p_0, b_0) определена с точностью до умножения на положительную константу, поэтому будем считать, что $b_0 = \pm 1$ или $b_0 = 0$. Разберем по очереди все три случая.

- Если $b_0 = 0$, то ковектор $p_0 \neq 0$ собственно разделяет $\operatorname{dom} f_1$ и $\operatorname{dom} f_2$, что невозможно, так как относительные внутренности этих множеств пересекаются по условию теоремы.

- Если $b_0 = 1$, то супремум в левой части полученного неравенства (4) равен $+\infty$ (можно устремить $a \to +\infty$), а инфимум в правой равен $-\infty$ (можно устремить $a \to -\infty$), и мы приходим к противоречию.

Таким образом, необходимо $b_0 = -1$. Поэтому для любых $x \in \operatorname{dom} f_1$ и $y \in \operatorname{dom} f_2$ при крайних значениях $a = f_1(x)$ (слева в (4)) и $a = -f_2(y)$ (справа в (4)) получаем

$$\langle p_0, x\rangle - f_1(x) \leqslant \langle p_0, y\rangle + f_2(y).$$

Исследуем это неравенство при разных значениях x и y. Положим сначала $y = x_0$. Тогда, с учетом $f_1(x_0) + f_2(x_0) = 0$, получим $f_1(x) \geqslant \langle p_0, x - x_0\rangle + f_1(x_0)$

для любого $x \in \mathrm{dom} f_1$, т.е. $p_0 \in \partial f_1(x_0)$. Положим теперь $x = x_0$. Получим $f_2(y) \geqslant \langle p_0, x_0 - y\rangle + f_2(x_0)$ для любого $y \in \mathrm{dom} f_2$, т.е. $-p_0 \in \partial f_2(x_0)$, что и требовалось.

Осталось разобрать общий случай: $p \in \partial(f_1 + f_2)(x_0)$, $p \neq 0$. Сделаем замену $\tilde{f}_1(x) = f_1(x) - \langle p, x\rangle$. Тогда $0 \in \partial(\tilde{f}_1 + f_2)(x_0)$. По доказанному ранее, найдется такой ковектор p_0, что $p_0 \in \partial\tilde{f}_1(x_0)$ и $-p_0 \in \partial f_2(x_0)$. Тогда для $p_1 = p + p_0$ и $p_2 = -p_0$ получаем $p_1 \in \partial f_1(x_0)$, $p_2 \in \partial f_2(x_0)$ и $p_1 + p_2 = p$, что и требовалось. □

***Следствие* 32.** *Пусть f_1 и f_2 – выпуклые собственные функции, и $\mathrm{dom} f_1$ пересекается с внутренностью $\mathrm{int}\,\mathrm{dom} f_1$. Тогда $\partial(f_1+f_2)(x_0) = \partial f_1(x_0)+\partial f_2(x_0)$ для любой точки x_0.*

Доказательство. Действительно, пусть $x_0 \in \mathrm{dom} f_1 \cap \mathrm{int}\,\mathrm{dom} f_2$. Выберем $x_1 \in \mathrm{ri}\,\mathrm{dom} f_1$. Тогда точки $x_\lambda = (1 - \lambda)x_0 + \lambda x_1$ стремятся к x_0 при $\lambda \to 0$ и лежат в $\mathrm{ri}\,\mathrm{dom} f_1$ при всех $\lambda \neq 0$. Таким образом, при достаточно малых λ выполнено $x_\lambda \in \mathrm{ri}\,\mathrm{dom} f_1 \cap \mathrm{int}\,\mathrm{dom} f_2$ и к сумме $f_1 + f_2$ применима теорема Моро-Рокафеллара. □

Покажем теперь двойственность операций сложения и инфимальной конволюции:

***Предложение* 58.** *Пусть $f_1, f_2 : \mathbb{R}^d \to \bar{\mathbb{R}}$ – выпуклые собственные функции. Тогда*

$$(f_1 \square f_2)^* = f_1^* + f_2^* \quad \text{и} \quad (\mathrm{cl} f_1 + \mathrm{cl} f_2)^* = (f_1^* \square f_2^*)^{**}.$$

Доказательство. Проведем прямое вычисление:

$$\begin{aligned}(f_1 \square f_2)^*(p) = \sup_x(\langle p, x\rangle - \inf_{x_1+x_2=x}(f_1(x_1) + f_2(x_2))) = \\ = \sup_{x_1,x_2}(\langle p, x_1 + x_2\rangle - f_1(x_1) - f_2(x_2)) = f_1^*(p) + f_2^*(p),\end{aligned}$$

и первое равенство доказано. Подставляя в него f_1^* и f_2^* вместо f_1 и f_2, получаем $(f_1^* \square f_2^*)^* = f_1^{**} + f_2^{**} = \mathrm{cl} f_1 + \mathrm{cl} f_2$ по теореме Фенхеля-Моро. Взяв сопряжение от обоих частей, получаем второе заявленное равенство. □

Отметим, что, вообще говоря, $(f_1^* \square f_2^*)^{**} \neq \mathrm{cl}(f_1^* \square f_2^*)$, так как функция $f_1^* \square f_2^*$ может быть несобственной:

Пример 11. Пусть $f_1, f_2 : \mathbb{R}^2 \to \bar{\mathbb{R}}$, $f_1(1, y) = f_2(-1, y) = 0$ и $f_1(x, y) = f_2(x, y) = \infty$ в остальных случаях. Тогда $f_1^*(p, 0) = -f_2^*(p, 0) = p$ и $f_1^*(p, q) = f_2^*(p, q) = \infty$ при $q \neq 0$. Далее, $(f_1^* \square f_2^*)(p, 0) = -\infty$ и $(f_1^* \square f_2^*)(p, q) = \infty$ при $q \neq 0$. Поэтому

$$-\infty \equiv (f_1 + f_2)^* \neq \mathrm{cl}(f_1^* \square f_2^*).$$

Тем не менее, равенство $(\mathrm{cl} f_1 + \mathrm{cl} f_2)^* = \mathrm{cl}(f_1^* \square f_2^*)$ можно встретить в разных книжках по выпуклому анализу. Дело заключается в различии определений операции замыкания cl. В отличие от данного в этой книжке определения операции cl (через замыкание надграфика), часто операцию cl для несобственных функций переопределяют так, что $\mathrm{cl} f \equiv -\infty$ для любой несобственной функции, отличной от тождественной бесконечности. При таком определении операции cl автоматически $f^{**} = \mathrm{cl} f$ даже для несобственных выпуклых функций, и равенство $(\mathrm{cl} f_1 + \mathrm{cl} f_2)^* = \mathrm{cl}(f_1^* + f_2^*)$ становится верным. Однако, мы избрали другой подход в определении операции cl по двум причинам: во-первых, чтобы не забывать про присутствующую тонкость с несобственными функциями, а, во-вторых, чтобы сохранить геометричность изложения – замыкание функции получается исключительно из замыкания ее надграфика. Поэтому

***Следствие* 33.** *Если функция $f_1^* \square f_2^*$ является собственной, то* $(\mathrm{cl} f_1 + \mathrm{cl} f_2)^* = \mathrm{cl}(f_1^* \square f_2^*)$.

Можно получить и точную двойственность операций $+$ и $\square$ при некоторых дополнительных предположениях (см. упражнение 9.4).

***Предложение* 59.** *Пусть f_1, f_2 – выпуклые собственные функции. Предположим, $\mathrm{ri}\,\mathrm{dom} f_1^* \cap \mathrm{ri}\,\mathrm{dom} f_2^* \neq \emptyset$ и инфимальная конволюция $f_1 \square f_2$ полунепрерывна снизу в x. Тогда*

$$\partial(f_1 \square f_2)(x) = \bigcup_{x_1+x_2=x} (\partial\, \mathrm{cl} f_1(x_1) \cap \partial\, \mathrm{cl} f_2(x_2)).$$

Доказательство. Доказательство проведем прямой выкладкой. Итак, по следствию 23 получаем:

$$\partial(f_1 \square f_2) = \{p : x \in \partial(f_1 \square f_2)^*(p)\} = \{p : x \in \partial(f_1^* + f_2^*)(p)\}.$$

Поскольку $\mathrm{ri}\,\mathrm{dom} f_1^* \cap \mathrm{ri}\,\mathrm{dom} f_2^* \neq \emptyset$, то, по теореме Моро-Рокафеллара, $\partial(f_1^* + f_2^*) = \partial f_1^* + \partial f_2^*$. Поэтому

$$\partial(f_1 \square f_2)(x) = \{p : x \in \partial f_1^*(p) + \partial f_2^*(p)\} = \\ = \{p : \exists x_1, x_2 : x_1 + x_2 = x,\ x_1 \in \partial f_1^*(p),\ x_2 \in \partial f_2^*(p)\}.$$

Поскольку функции f_i^*, $i = 1, 2$ являются выпуклыми, замкнутыми и собственными, то $x_i \in \partial f_i^*(p)$ равносильно $p \in \partial f_i^{**}(x_i) = \partial\, \mathrm{cl} f_i(x_i)$. Итак,

$$\partial(f_1 \square f_2)(x) = \{p : \exists x_1, x_2 : x_1 + x_2 = x,\ p \in \partial\, \mathrm{cl} f_1(x_1),\ p \in \partial\, \mathrm{cl} f_2(x_2)\} = \\ = \bigcup_{x_1+x_2=x} (\partial\, \mathrm{cl} f_1(x_1) \cap \partial\, \mathrm{cl} f_2(x_2)).$$

□

Теорема Дубовицкого-Милютина. Нам осталось вычислить субдифференциал пары операций – $\wedge$ и $\vee$. Напомним, что $(f_1 \vee f_2)(x) = \max\{f_1(x), f_2(x)\}$.

***Теорема* 24** (Дубовицкого-Милютина)**.** *Пусть $f = f_1 \vee \ldots \vee f_n$, где $f_i : \mathbb{R}^d \to \bar{\mathbb{R}}$ – выпуклые собственные функции. Тогда если $x_0 \in \bigcap_i \mathrm{ri}\,\mathrm{dom} f_i$ и $f_i(x_0) = f(x_0)$ для всех i, то*

$$\partial f(x_0) = \mathrm{conv} \bigcup_{i=1}^{n} \partial f_i(x_0).$$

Доказательство. Включение в одну сторону получается очень просто: если $p \in \partial f_i(x_0)$ для какого-то i, то

$$f(x) \geqslant f_i(x) \geqslant f_i(x_0) + \langle p, x - x_0 \rangle = f(x_0) + \langle p, x - x_0 \rangle,$$

т.е. $p \in \partial f(x_0)$. Таким образом, $\partial f_i(x_0) \subset f(x_0)$ для всех i. Поскольку субдифференциал всегда выпукл, то $\operatorname{conv} \bigcup_i \partial f_i(x_0) \subset \operatorname{conv} \partial f(x_0) = \partial f(x_0)$, и включение в одну сторону доказано.

Для доказательства включения в обратную сторону нам потребуется для всякого $p \in \partial f(x_0)$ найти его представление в виде выпуклой комбинации $p = \sum_i \lambda_i p_i$, где $p_i \in \partial f_i(x_0)$. Для этого воспользуемся теоремой 12 о критерии разрешимости системы выпуклых неравенств на выпуклом множестве.

Обозначим $g_i(x) = f_i(x) - f_i(x_0) - \langle p, x - x_0 \rangle$. Мы утверждаем, что не существует такой точки $\hat{x}$, что $g_i(\hat{x}) < 0$ при всех i. Действительно, в противном случае при всех i выполнялось бы $f_i(\hat{x}) < f_i(x_0) + \langle p, \hat{x} - x_0 \rangle$, а, значит, $f(\hat{x}) = \max_i \{f_i(\hat{x})\} < f(x_0) + \langle p, \hat{x} - x_0 \rangle$, т.к. $f_i(x_0) = f(x_0)$. Но это невозможно, т.к. p лежит в субдифференциале f в x_0.

Итак, система неравенств $g_i(x) < 0$, $i \in \{1, \dots, n\}$, несовместна. Постараемся уменьшить, насколько это возможно, число неравенств в системе, но таким образом, чтобы сохранить ее несовместность. Поступим следующим образом: назовем непустой набор индексов $I \subset \{1, \dots, n\}$ полным, если система неравенств $g_i(x) < 0$ при $i \in I$ несовместна. Среди всех непустых полных наборов выберем какой-либо с минимально возможным числом индексов и обозначим его за J.

Для применения теоремы 12 необходимо еще выбрать такое выпуклое множество A, что $\operatorname{ri} A \subset \operatorname{dom} g_i$ при всех $i \in J$. Положим $A = \bigcap_{i \in J} \operatorname{dom} f_i$. Значит, $\operatorname{ri} A \subset \operatorname{dom} g_i$ для всех $i \in J$, т.к. $\operatorname{dom} g_i = \operatorname{dom} f_i$ при любом i. Таким образом, набор функций g_i, $i \in J$, на множестве A удовлетворяет условиям теоремы 12, которая утверждает, что система неравенств $g_i(x) < 0$, $i \in J$, $x \in A$, несовместна тогда и только тогда, когда найдутся такие не равные одновременно 0 числа $\lambda_i \geqslant 0$, $i \in J$, что $\sum_{i \in J} \lambda_i g_i(x) \geqslant 0$ для всех $x \in A$. Множители Лагранжа λ_i определены с точностью до умножения на произвольную положительную константу, поэтому мы можем без ограничения общности считать, что $\sum_i \lambda_i = 1$. Более того, ни одно из чисел λ_i, $i \in J$, не может равняться 0, так как в противном случае по теореме 12 набор J можно было бы уменьшить, выкинув из него те индексы i, для которых $\lambda_i = 0$.

Отметим, что для $x \notin A$ эта сумма $\sum_{i \in J} \lambda_i g_i(x)$ равна бесконечности. Действительно, если $x \notin \operatorname{dom} g_j$ для некоторого $j \in J$, то $g_j(x) = \infty$ и $\sum_{i \in J} \lambda_i g_i(x) = \infty$, так как $\lambda_j > 0$. Таким образом, полученное неравенство $\sum_{i \in J} \lambda_i g_i(x) \geqslant 0$ выполняется вообще для всех x, т.е.

$$\sum_{i \in J} \lambda_i f_i(x) \geqslant \sum_{i \in J} \lambda_i f_i(x_0) + \sum_{i \in J} \lambda_i \langle p, x - x_0 \rangle = \sum_{i \in J} \lambda_i f_i(x_0) + \langle p, x - x_0 \rangle,$$

т.к. $\sum_{i \in J} \lambda_i = 1$. Следовательно, $p \in \partial(\sum_{i \in J} \lambda_i f_i)(x_0)$, и, значит,

$$p \in \partial(\sum_{i \in J} \lambda_i f_i)(x_0) = \sum_{i \in J} \lambda_i \partial f_i(x_0) \subset \operatorname{conv} \bigcup \partial_i f_i(x_0),$$

что и требовалось $\square$

Следствие 34. *Если $f = f_1 \wedge \dots \wedge f_n$, где $f_i : \mathbb{R}^d \to \mathbb{R}$ – выпуклые собственные функции, и $x_0 \in \bigcap_i \operatorname{int} f_i$, то*

$$\partial f(x_0) = \operatorname{conv} \bigcup_{i: f_i(x_0) = f(x_0)} \partial f_i(x_0).$$

Доказательство. Обозначим через $I \subset \{1, \dots, n\}$ набор всех таких индексов i, что $f_i(x_0) = f(x_0)$. В этом случае для остальных индексов i выполняется $f_i(x) < f(x_0)$ для x из малой окрестности x_0, так как $x_0 \in \operatorname{int} \operatorname{dom} f_i$, а любая выпуклая функция на $\mathbb{R}^d$ непрерывна во внутренних точках своей эффективной области. Поэтому $f(x) = \max_{i \in I}\{f_i(x)\}$ по крайней мере для точек x из некоторой окрестности x_0. Но субдифференциал выпуклой функции в точке определяется любой ее малой окрестностью по теореме 6 о локализации субдифференциала (см. следствие 14). Поэтому применение теоремы Дубовицкого-Милютина к $\max_{i \in I}\{f_i(x)\}$ дает нужный результат. $\square$

Упражнения:

9.1 Пусть $K \subset \mathbb{R}^d$ – непустой выпуклый конус. Докажите, что $p \in \partial \delta_K(x)$ если и только если $x \in K$, $p \in K^*$ и $\langle p, x \rangle = 0$.

9.2 Пусть $g : \mathbb{R}^{d'} \to \bar{\mathbb{R}}$ – выпуклая собственная функция и $A : \mathbb{R}^d \to \mathbb{R}^{d'}$ – линейный оператор. Докажите, что если $\operatorname{Im} A \cap \operatorname{ri} \operatorname{dom} g \neq \emptyset$, то $(gA)^* = A^* g^*$.

9.3 Приведите пример двух таких выпуклых замкнутых функций, что $\partial(f_1 \vee f_2)(x_0) \neq \operatorname{conv}(\partial f_1(x_0) \cup \partial f_2(x_0))$ для некоторой точки x_0, в которой их значения совпадают $f_1(x_0) = f_2(x_0)$.

9.4 Докажите, что если f_1, f_2 – выпуклые собственные функции и $\operatorname{ri} \operatorname{dom} f_1 \cap \operatorname{ri} \operatorname{dom} f_2 \neq \emptyset$, то $(f_1 + f_2)^* = (f_1^* \square f_2^*)$.

9.5 На плоскости $\mathbb{R}^2$ заданы прямая l и две точки $x_0 \neq x_1$, не лежащие на ней. Найти такую точку $x \in \mathbb{R}^2$, что сумма трех расстояний от нее до прямой и двух точек (глобально) минимальна.

9.6 Доказать, что $\sqrt{x^2 + y^2} + |x| + \frac{3}{2}x + y^4 + \frac{1}{2}y \geqslant 0$.

10 Двойственность выпуклых задач

Общая двойственность семейств выпуклых задач. Вычисление сопряженной функции к функции $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ есть, по сути, решение задачи

$$f(x) - \langle p, x \rangle \to \min_x,$$

где $p \in \mathbb{R}^{d*}$ входит в минимизируемую функцию линейно и выступает параметром. Предположим теперь, что и сама функция f зависит от параметра $y \in \mathbb{R}^{d'}$. В этом случае мы получаем семейство задач

$$f(x, y) - \langle p, x \rangle \to \min_x, \tag{5}$$

зависящих линейно от параметра $p \in \mathbb{R}^{d*}$ и, вообще говоря, нелинейно от параметра $y \in \mathbb{R}^{d'}$. Оказывается, для таких семейств определено двойственное семейство задач, зависящее от тех же параметров p и y, но теперь уже линейно по y и, вообще говоря, нелинейно по p:

Определение 22. *Двойственным семейством* к семейству (5) называется семейство задач минимизации по $q \in \mathbb{R}^{d'*}$ с параметрами y и p

$$f^*(p,q) - \langle q, y\rangle \to \min_q .$$

Очевидно, что функция f^* является выпуклой, поэтому двойственная задача всегда выпукла. Обозначим значения прямой и двойственной задач через[7] $S_p(y)$ и $D_y(p)$ (линейный параметр выступает в качестве индекса):

$$S_p(y) = \inf_x (f(x,y) - \langle p, x\rangle); \qquad D_y(p) = \inf_q (f^*(p,q) - \langle q, y\rangle).$$

Отметим, что если f – выпуклая функция, то и функция $S_p(y)$ тоже выпукла по y. Если обозначить $g(x,y) = f(x,y) - \langle p, x\rangle$, то для линейной проекции $A(x,y) = x$ получаем $S_p = Ag$. Функция $D_y(p)$ всегда выпукла по p.

Оказывается, что решения прямой и двойственной задач, а также их значения тесно связаны. Самый простой результат заключается в следующем:

***Предложение* 60** (Слабая двойственность)**.** *Независимо от f, выполнено $S_p(y) + D_y(p) \geqslant 0$.*

Доказательство. Мы знаем, что $S_p^{**}(y) \leqslant S_p(y)$. Вычислим явно

$$S_p^*(q) = \sup_y (\langle q, y\rangle - S_p(y)) = \sup_{x,y} (\langle q, y\rangle + \langle p, x\rangle - f(x,y)) = f^*(p,q).$$

Поэтому

$$S_p(y) \geqslant S_p^{**}(y) = \sup_q (\langle q, y\rangle - f^*(p,q)) = -D_y(p).$$

□

Итак, двойственное семейство задач есть не что иное, как вычисление второй сопряженной функции S^{**} (с точностью до знака). Таким образом, получается простое мнемоническое правило для описания логики двойственной задачи:

двойственная задача есть задача о вычислении второй сопряженной функции.

В случае, когда $S_p(y) + D_y(p) = 0$, говорят о сильной двойственности прямого и двойственного семейств задач. Однако, сильная двойственность не всегда имеет место (особенно в невыпуклом случае), и тогда говорят о наличии *разрыва двойственности*, который, по определению, равен положительному числу $S_p(y) + D_y(p) > 0$.

Для выпуклого же случая мы часто можем доказать отсутствие разрыва. Более того, в выпуклом случае, решив прямую задачу, мы немедленно решаем двойственную, и наоборот:

***Теорема* 25** (о связи прямой и двойственной задач)**.** *Зафиксируем некоторые значения параметров $y_0 \in \mathbb{R}^{d'}$ и $p_0 \in \mathbb{R}^{d*}$. Рассмотрим для любых $x_0 \in \mathbb{R}^d$ и $q_0 \in \mathbb{R}^{d'}$ следующие утверждения:*

1. *x_0 есть решение прямой задачи, и $q_0 \in \partial_y S_{p_0}(y_0)$.*

[7] S от straight, т.е. прямая задача, и D от dual, т.е. двойственная задача.

2. *q_0 есть решение двойственной задачи, и $x_0 \in \partial_p D_{y_0}(p_0)$.*

3. *x_0 есть решение прямой задачи, q_0 есть решение двойственной задачи и отсутствует разрыв двойственности, т.е. $S_{p_0}(y_0) + D_{y_0}(p_0) = 0$.*

4. *$f(x_0, y_0) + f^*(p_0, q_0) = \langle p_0, x_0\rangle + \langle q_0, y_0\rangle$.*

Тогда при любой функции f имеем $1 \Leftrightarrow 3 \Leftrightarrow 4 \Rightarrow 2$, а если дополнительно функция f является выпуклой, замкнутой и собственной, то $1 \Leftrightarrow 2 \Leftrightarrow 3 \Leftrightarrow 4$.

Таким образом, если мы решили прямую задачу, т.е. при всех значениях параметров y и p нашли оптимальное решение и вычислили значение задачи $S_p(y)$, то решение сопряженной задачи есть субдифференциал от $S_p(y)$ по y, и наоборот. Трудности в этом случае могут возникнуть только при тех значениях параметров, при которых субдифференциал $S_p(y)$ пуст. Если же функция $S_p(y)$ дифференцируема по y, то $S'_p(y)$ есть решение двойственной задачи. Аналогично, решив двойственную, мы немедленно решим прямую.

Доказательство теоремы 25. $(1 \Rightarrow 4)$. Из $q_0 \in \partial_y S_{p_0}(y_0)$ получаем

$$S_{p_0}(y) \geqslant S_{p_0}(y_0) + \langle q_0, y - y_0\rangle. \tag{6}$$

Поскольку $S_{p_0}(y)$ есть инфимум выражения $f(x, y) - \langle p_0, x\rangle$ по всем x, то мы получаем

$$f(x, y) - \langle p_0, x\rangle \geqslant S_{p_0}(y_0) + \langle q_0, y - y_0\rangle, \tag{7}$$

а поскольку x_0 – решение прямой задачи, то $S_{p_0}(y_0) = f(x_0, y_0) - \langle p_0, x_0\rangle$, то

$$f(x, y) - \langle p_0, x\rangle \geqslant f(x_0, y_0) - \langle p_0, x_0\rangle + \langle q_0, y - y_0\rangle, \tag{8}$$

а это неравенство есть не что иное, как определение субдифференциала f, т.е. $(p_0, q_0) \in \partial f(x_0, y_0)$. Поэтому $f(x_0, y_0) + f^*(p_0, q_0) = \langle p_0, x_0\rangle + \langle q_0, y_0\rangle$.

$(4 \Rightarrow 1)$. Прочитаем предыдущее доказательство в обратную сторону: если $f(x_0, y_0) + f^*(p_0, q_0) = \langle p_0, x_0\rangle + \langle q_0, y_0\rangle$, то $(p_0, q_0) \in \partial f(x_0, y_0)$, и, значит, неравенство (8) выполняется при всех x и y. Подставляя в него сначала $y = y_0$, получим, что x_0 есть решение прямой задачи, и, следовательно, $S_{p_0}(y_0) = f(x_0, y_0) - \langle p_0, x_0\rangle$. Таким образом, неравенство (8) превращается в (7). Неравенство (7) по-прежнему выполняется для всех x и y. Поскольку $S_{p_0}(y)$ есть инфимум выражения $f(x, y) - \langle p_0, x\rangle$ по всем x, то мы получаем неравенство (6), которое совпадает с определением субдифференциала $\partial_y S_{p_0}(y_0)$.

$(4 \Rightarrow 2)$. Мы знаем, что $f^{**}(x_0, y_0) \leqslant f(x_0, y_0)$, поэтому $f^*(p_0, q_0) + f^{**}(x_0, y_0) \leqslant f(x_0, y_0) + f^*(p_0, q_0) = \langle p_0, x_0\rangle + \langle q_0, y_0\rangle$. Полученное неравенство обратно неравенству Юнга для f^* и f^{**}. Поэтому должно быть равенство. Остается применить уже полученное доказательство $(4 \Rightarrow 1)$, заменив f на f^*.

$(3 \Rightarrow 4)$. Достаточно заметить, что если x_0 и p_0 – решения прямой и двойственной задач соответственно, то

$$S_{p_0}(y_0) + D_{y_0}(p_0) = f(x_0, y_0) - \langle p_0, x_0\rangle + f^*(p_0, q_0) - \langle q_0, y_0\rangle, \tag{9}$$

и правая часть должна равняться 0, что и требовалось.

$(4 \Rightarrow 3)$. Мы уже знаем, что из пункта 4 следуют пункты 1 и 2, поэтому x_0 и q_0 заведомо являются решениями прямой и двойственной задач соответственно. Остается заметить, что в этом случае выполняется равенство (9), и, значит, отсутствует разрыв двойственности, что и требовалось. □

Задачи линейного программирования. Вычислим двойственную задачу для задачи линейного программирования. Сформулируем сначала прямую задачу. Пусть $K_1 \subset \mathbb{R}^d$ и $K_2 \subset \mathbb{R}^{d'}$ – выпуклые замкнутые конусы (например, положительные ортанты). Задача линейного программирования имеет вид:

$$\langle c, x\rangle \to \min_x \quad \text{при условиях} \quad x \in K_1 \ \text{и} \ b - Ax \in K_2.$$

Здесь $c \in \mathbb{R}^{d*}$ и $b \in \mathbb{R}^{d'}$ – данные векторы, $A : \mathbb{R}^d \to \mathbb{R}^{d'}$ – данный линейный оператор и $x \in \mathbb{R}^d$ – неизвестный вектор.

Построим двойственную задачу. Избавиться от ограничений очень просто: положим

$$f(x, y) = \begin{cases} 0, & \text{если } x \in K_1 \text{ и } y - Ax \in K_2, \\ \infty, & \text{иначе.} \end{cases}$$

Тогда прямая задача перепишется в виде

$$f(x, b) + \langle c, x\rangle \to \min_x .$$

Поэтому $y = b$ и $p = -c$, и двойственная задача имеет вид

$$f^*(-c, q) - \langle q, b\rangle \to \min_q .$$

Вычислим сопряженную функцию

$$f^*(p, q) = \sup_{x,y}(\langle p, x\rangle + \langle q, y\rangle - f(x, y)) = \sup_{x \in K_1, y - Ax \in K_2}(\langle p, x\rangle + \langle q, y\rangle).$$

Обозначим $r = A^*q + p$, т.е. $p = r - A^*q$, и $z = y - Ax$, т.е. $y = Ax + z$. Тогда

$$f^*(p, q) = \sup_{x \in K_1, z \in K_2}(\langle r, x\rangle + \langle q, z\rangle).$$

Поэтому, если $r \notin K_1^*$, то $f^*(p, q) = \infty$, так как если $\langle r, x\rangle > 0$ для какого-то $x \in K_1$, то $\langle r, \lambda x\rangle \to \infty$ при $\lambda \to \infty$ и $\lambda x \in K_1$. Аналогично, если $q \notin K_2^*$. Если же $r \in K_1^*$ и $q \in K_2^*$, то $f^*(p, q) = 0$. Таким образом,

$$f^*(p, q) = \begin{cases} 0, & \text{если } q \in K_2^* \text{ и } A^*q + p \in K_1^*; \\ \infty, & \text{иначе.} \end{cases}$$

Итак, двойственная задача имеет вид:

$$-\langle q, b\rangle \to \min_q \quad \text{при условиях} \quad q \in K_2^* \ \text{и} \ A^*q - c \in K_1^*.$$

Все возможные классические типы двойственности при различных комбинациях неравенств $x \lesseqgtr 0$ и $Ax \lesseqgtr y$ получаются при нужном выборе конусов K_1 и K_2 – это могут быть положительный и отрицательный ортанты (отвечают $\geqslant$ и $\leqslant$), все пространство (отвечает отсутствию ограничений) и 0 (отвечает =). Первые два конуса двойственны друг другу, последние два тоже.

Двойственность задач по Фенхелю. Двойственность выпуклых задач по Фенхелю отличается от той, что мы разработали, но легко из нее следует. Задача ставится следующим образом:

$$f(x) + g(Ax) \to \min_x, \tag{10}$$

где $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ и $g : \mathbb{R}^{d'} \to \mathbb{R}$ – функции, а $A : \mathbb{R}^d \to \mathbb{R}^{d'}$ – линейный оператор.

Параметр y в этой задаче отсутствует, поэтому его необходимо ввести искусственно. От того, как мы его введем, вообще говоря, зависит то, какое двойственное семейство мы получим. Обозначим $F(x, y) = f(x) + g(Ax + y)$ и рассмотрим семейство

$$F(x, y) - \langle p, x\rangle \to \min_x .$$

Исходная задача получается при $y = 0$ и $p = 0$. Вычислим сопряженную функцию:

$$F^*(p, q) = \sup_{x,y}(\langle p, x\rangle + \langle q, y\rangle - f(x) - g(Ax + y)).$$

Обозначим $z = Ax + y$, т.е. $y = z - Ax$. Тогда

$$F^*(p, q) = \sup_{x,z}(\langle p - A^*q, x\rangle + \langle q, z\rangle - f(x) - g(z)) = f^*(p - A^*q) + g^*(q).$$

Таким образом, двойственное семейство имеет вид

$$f^*(p - A^*q) + g^*(q) - \langle q, y\rangle \to \min_q .$$

При исходных значениях параметров $y = 0$ и $p = 0$ получаем задачу, обычно называемую двойственной по Фенхелю к задаче (10):

$$f^*(-A^*q) + g^*(q) \to \min_q .$$

Упражнения:

10.1 Постройте двойственное семейство задач к семейству транспортных задач $\sum_{i,j} c_{ij}x_{ij} \to \min$ при условиях $\sum_i x_{ij} = b_j$ для всех i, $\sum_j x_{ij} = a_i$ для всех i и $x_{ij} \geqslant 0$ для всех i и j. Здесь $i = 1, \dots, n$, $j = 1, \dots, m$, $x = (x_{ij}) \in \mathbb{R}^{nm}$ – неизвестные, а $(a, b, c) \subset \mathbb{R}^{n+m+nm}$ – параметры семейства.

10.2 Докажите, что двойственное семейство задач к семейству $f_0(x) \to \min_x$, $f_i(x) \leqslant y_i$, $i = 1, \dots, n$, $x \in A$, есть семейство задач $-\inf_{x\in A} \mathscr{L}(\lambda, x, y) \to \min_\lambda$, $\lambda_i \geqslant 0$, где $\lambda = -q$ и $\mathscr{L} = f_0(x) + \sum_{i=1}^n \lambda_i(f_i(x) - y_i)$.

10.3 Докажите, что если в предыдущем упражнении все функции выпуклы и выполнено условие Слейтера, то отсутствует разрыв двойственности.

11 Двойственность строгой выпуклости и гладкости

Строгая выпуклость и гладкость. По аналогии со строгой монотонностью можно определить понятие строгой выпуклости: функция

$f : \mathbb{R}^d \to \bar{\mathbb{R}}$ называется *строго выпуклой* на множестве $S \subset \operatorname{dom} f$, если для всех $x_0, x_1 \in S$, $x_0 \neq x_1$, и $\lambda \in (0;1)$ выполняется строгое неравенство

$$f(x_\lambda) < (1-\lambda)f(x_0) + \lambda f(x_1).$$

Понятие строгой выпуклости очень близко к понятию гладкости сопряженной функции:

Теорема 26. *Пусть $f : \mathbb{R}^d \to \mathbb{R}$ – собственная, выпуклая, замкнутая функция. Пусть $S \subset \operatorname{dom} f$ и $S^* \subset \operatorname{dom} f^*$ – такие множества, что $S \supset \bigcup_{p \in S^*} \partial f^*(p)$ и $S^* \supset \bigcup_{x \in S} \partial f(x)$. Тогда*

- *Если функция f строго выпукла на S, то производная $f^{*\prime}$ существует и непрерывна на $S^* \cap \operatorname{ri} \operatorname{dom} f^*$;*
- *Если производная $f^{*\prime}$ существует и непрерывна на S^*, то функция f строго выпукла на $S \cap \operatorname{ri} \operatorname{dom} f$.*

Доказательство. Для доказательства первой части теоремы нам достаточно показать, что для любого $p \in S^*$ субдифференциал $\partial f^*(p)$ состоит ровно из одного элемента. Действительно, если мы докажем этот факт, то по теореме 10 мы немедленно получим нужный результат.

Субдифференциал $\partial f^*(p)$, как минимум, непуст при любом $p \in \operatorname{ri} \operatorname{dom} f^*$ по теореме 7. Покажем, что субдифференциал $\partial f^*(p)$ для $p \in S^*$ не может содержать двух различных элементов. Действительно, если $x_0, x_1 \in \partial f^*(p)$, $x_0 \neq x_1$, то $x_0, x_1 \in S$ по условию теоремы и $p \in \partial f(x_0)$, $p \in \partial f(x_1)$. Другими словами, надграфик f имеет общую опорную гиперплоскость H в точках x_0 и x_1. Поэтому отрезок $[(x_0, f(x_0)), (x_1, f(x_1))]$ должен лежать в H, и, следовательно, $f(x_\lambda) = (1-\lambda)f(x_0) + \lambda f(x_1)$ при $\lambda \in [0;1]$, что противоречит строгой выпуклости f.

Докажем вторую часть теоремы. Предположим противное: пусть для некоторых $x_0, x_1 \in S \cap \operatorname{ri} \operatorname{dom} f$ и $\hat{\lambda} \in (0;1)$ выполнено равенство $f(x_{\hat{\lambda}}) = (1-\hat{\lambda})f(x_0) + \hat{\lambda} f(x_1)$. Тогда это равенство должно выполняться для всех $\lambda \in [0;1]$ (см. упражнение 3.2). Таким образом, отрезок $[(x_0, f(x_0)), (x_1, f(x_1))]$ не пересекается с относительной внутренностью надграфика f (см. следствие 11) и его можно собственно отделить от $\operatorname{epi} f$ некоторой опорной гиперплоскостью H. Гиперплоскость H не может быть вертикальной, так как $[x_0; x_1] \subset \operatorname{ri} \operatorname{dom} f$, и потому определяет некоторый общий во всех точках x_λ субдифференциал $p \in \partial f(x_\lambda)$. Поскольку $p \in \partial f(x_0)$ и $x_0 \in S$, то по условию теоремы $p \in S^*$. Более того, $x_\lambda \in \partial f^*(p)$ при всех $\lambda \in [0;1]$, что невозможно, так как функция f^* дифференцируема в $p \in S^*$. □

Следствие 35. *Пусть $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ – собственная замкнутая выпуклая функция. Если $\operatorname{dom} f$ – открытое множество и $f \in C^1(\operatorname{dom} f)$, то функция f^* является строго выпуклой на $\operatorname{ri} \operatorname{dom} f^*$.*

Пример 12. Обратное, вообще говоря, неверно. Пусть $f(x) = 1/x$ при $x > 0$ и $f(x) = \infty$ при $x \leqslant 0$. Тогда $f^*(p) = -2\sqrt{-p}$ при $p \leqslant 0$ и $f^*(p) = \infty$ при $p > 0$, и f^* не дифференцируема в 0.

В формулировке этой теоремы прослеживается некоторая несимметричность. Если функция строго выпукла на S, то мы можем утверждать гладкость сопряженной только на $S^* \cap \operatorname{ri} \operatorname{dom} f^*$, и аналогично в обратную сторону. На самом деле, полной двойственности между свойствами строгой выпуклости и гладкости, все-таки, нет. Например, если, скажем, $\operatorname{int}\operatorname{dom} f \neq \emptyset$ и $f \in C^1(\operatorname{int}\operatorname{dom} f)$, то мы не можем утверждать, что сопряженная функция является строго выпуклой на всей своей эффективной области, и наоборот. Дело заключается в том, что при $d \geqslant 2$ на относительной границе эффективной области могут появляться сложные эффекты:

Пример 13. Рассмотрим выпуклое замкнутое множество $C \subset \mathbb{R}^2$, $C = \{(x, y) : x \geqslant \frac{1}{2}y^2\}$ и выпуклую функцию $h(x, y) = -2\sqrt{y}$ при $y \geqslant 0$ и $h(x, y) = \infty$ при $y < 0$. Контрпример нам даст выпуклая замкнутая собственная функция $f(x, y) = \mu_C(x, y) + h(x, y)$ – мы докажем, что функция $f^*(p, q)$ определена на открытом множестве и является на нем гладкой, при этом сама функция f не является строго выпуклой.

Итак, $\mu_C(x, y) = y^2/2x$, если $x, y > 0$, $\mu_C(0, 0) = 0$ и $\mu_C(x, y) = \infty$ во всех остальных случаях. Поэтому

$$f(x, y) = \begin{cases} y^2/2x - 2\sqrt{y}, & \text{если } x > 0,\ \ y \geqslant 0; \\ 0, & \text{если } x = y = 0; \\ \infty, & \text{в остальных случаях.} \end{cases}$$

Функция f является выпуклой и замкнутой, так как таковыми являются функции h и μ_C. Однако, функция f не является строго выпуклой на $\operatorname{dom} f$, так как $f(x, 0) = 0$ при всех $x \geqslant 0$. Вычислим сопряженную функцию f^*. Поскольку $\mu_C = s_{C^\circ}$, то $\mu_C^* = s_{C^\circ}^* = \delta_{C^\circ}$. Поляра C° легко находится: $C^\circ = \{(p, q) : p \leqslant -\frac{1}{2}q^2\}$. Поэтому $\mu_C^*(p, q) = 0$ при $p + \frac{1}{2}q^2 \leqslant 0$ и $\mu_C^*(p, q) = \infty$, иначе. Вычислим теперь

$$h^*(p, q) = \sup_{x \in \mathbb{R}, y \geqslant 0} (px + qy + 2\sqrt{y}).$$

Поэтому, если $p \neq 0$ или $q \geqslant 0$, то $h^*(p, q) = \infty$. Если же $p = 0$ и $q < 0$, то $h^*(0, q) = -1/q$. Итак,

$$f^*(p, q) = (\mu_C^* \square h^*)(p, q) = \inf_{\alpha, \beta} (\mu_C^*(p - \alpha, q - \beta) + h^*(\alpha, \beta)).$$

Поскольку в этом выражении мы ищем инфимум, то $\alpha = 0$ и $\beta \leqslant 0$. Поэтому

$$f^*(p, q) = \inf\{\frac{-1}{\beta} : \beta < 0 \text{ и } p + \frac{1}{2}(q - \beta)^2 \leqslant 0\}.$$

Поэтому если $p > 0$, то $f^*(p, q) = \infty$. Если же $p \leqslant 0$, то

$$\inf\{\frac{-1}{\beta} : \beta < 0 \text{ и } q - \sqrt{-2p} \leqslant \beta \leqslant q + \sqrt{-2p}\}.$$

Таким образом,

$$f^*(p, q) = \begin{cases} \frac{1}{\sqrt{-2p} - q}, & \text{если } p + \frac{1}{2}q^2 < 0; \\ \infty, & \text{иначе.} \end{cases}$$

Следовательно, $\operatorname{dom} f^*$ – открытое множество и $f^* \in C^\infty(\operatorname{dom} f^*)$, однако, ее сопряженная $f^{**} = f$ является строго выпуклой только на относительной внутренности $\operatorname{ri}\operatorname{dom} f$ и не является строго выпуклой на всей эффективной области.

Для того, чтобы получить точную двойственность на этом пути, мы немного усилим оба определения.

Сильная выпуклость. Если гладкая выпуклая функция имеет положительную вторую производную, то ее называют сильно выпуклой. «Степень выпуклости» функции по некоторому направлению определяется значением второй производной на этом направлении – чем оно больше, тем быстрее меняется производная функции, а график имеет бóльшую кривизну. Для измерения «степени выпуклости» функции необходимо задать в пространстве $\mathbb{R}^d$ какую-либо норму $\|\cdot\|$, относительно которой можно измерять кривизну. Итак,

Определение 23. Пусть $a > 0$. Собственная выпуклая функция $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ называется *a-сильно выпуклой* относительно нормы $\|\cdot\|$, если для любых $x, h \in \mathbb{R}^d$ и $p \in \partial f(x)$ выполняется

$$f(x+h) \geqslant f(x) + \langle p, h\rangle + \frac{a}{2}\|h\|^2$$

При проверке этого неравенства можно ограничиться случаем $x, x+h \in \mathrm{dom} f$, так как если $x \notin \mathrm{dom} f$, то $\partial f(x) = \emptyset$, а если $x + h \notin \mathrm{dom} f$, то $f(x+h) = \infty$, и в обоих этих случаяхнеравенство выполняется автоматически.

Геометрически это неравенство означает, что точка $(x+h, f(x+h))$ должна лежать не просто над опорной гиперплоскостью в точке $(x, f(x))$, а выше некоторой параболы, касающейся опорной гиперплоскости и имеющей «кривизну» (старший коэффициент) a.

Часто a-сильную выпуклость функции определяют по-другому, опираясь на неравенство

$$f(x_\lambda) \leqslant (1-\lambda)f(x_0) + \lambda f(x_1) - \frac{a}{2}\lambda(1-\lambda)\|x_1 - x_0\|^2. \tag{11}$$

Связь сильной выпуклости и этого неравенства дана в нижеследующем предложении:

***Предложение* 61.** *Пусть f – собственная выпуклая функция. Тогда если f является a-сильно выпуклой, то неравенство* (11) *заведомо выполняется при $\partial f(x_\lambda) \neq \emptyset$. Обратно: если неравенство* (11) *выполняется для всех x_0, x_1 и λ таких, что $\partial f(x_0) \neq \emptyset$ или $\partial f(x_1) \neq \emptyset$, то функция f является a-сильно выпуклой.*

Доказательство. Докажем сначала первую часть. Поскольку $\partial f(x_\lambda) \neq \emptyset$, то выберем произвольный элемент $p \in \partial f(x_\lambda)$. Тогда, согласно определению сильной выпуклости, выполняются неравенства

$$f(x_0) \geqslant f(x_\lambda) - \lambda\langle p, x_1 - x_0\rangle + \frac{a}{2}\lambda^2\|x_1 - x_0\|,$$
$$f(x_1) \geqslant f(x_\lambda) + (1-\lambda)\langle p, x_1 - x_0\rangle + \frac{a}{2}(1-\lambda)^2\|x_1 - x_0\|.$$

Домножая эти неравенства на $1-\lambda$ и λ (соответственно) и складывая, получаем (11) после приведения подобных членов.

Теперь докажем вторую часть. Пусть $x, h \in \mathbb{R}^d$ и $p \in \partial f(x)$. Тогда для $x_0 = x$, $x_1 = x + h$ и $\lambda \in [0; 1]$ получаем

$$f(x_0) + \lambda\langle p, x_1 - x_0\rangle \leqslant f(x_\lambda) \leqslant (1-\lambda)f(x_0) + \lambda f(x_1) - \frac{a}{2}\lambda(1-\lambda)\|x_1 - x_0\|^2.$$

Перенося $f(x_0)$ направо и деля на λ при $\lambda \neq 0$, получаем $f(x_1) - f(x_0) + \frac{a}{2}(1-\lambda)\|x_1 - x_0\|^2 \geqslant \langle p, x_1 - x_0\rangle$. Осталось взять предел при $\lambda \to 0$. □

Двойственная норма. Для определения понятия сильной гладкости, нужно измерить скорость изменения производной функции. Поскольку для гладкой функции $f'(x) \in \mathbb{R}^{d*}$, то нам потребуется ввести норму на сопряженном пространстве $\mathbb{R}^{d*}$. Делается это стандартным образом, а именно: если $\|\cdot\|$ – некоторая норма на пространстве $\mathbb{R}^d$, то она естественным образом определяет двойственную норму $\|\cdot\|_*$ на сопряженном пространстве $\mathbb{R}^{d*}$ по формуле

$$\|p\|_* \stackrel{\text{def}}{=} \sup_{\|x\|\leqslant 1} \langle p, x\rangle.$$

Из определения легко получается, что $\langle p, x\rangle \leqslant \|x\|\|p\|_*$ для всех $x \in \mathbb{R}^d$ и $p \in \mathbb{R}^{d*}$. Также, в определении сопряженной нормы супремум берется по компактному множеству $\{\|x\| \leqslant 1\}$, поэтому непрерывная функция $\langle p, \cdot\rangle$ достигает максимума в некоторой точке $\hat{x}$ и, очевидно, $\|\hat{x}\| = 1$. Другими словами, для любого $p \in \mathbb{R}^{d*}$ найдется такая точка $\hat{x} \in \mathbb{R}^d$, $\|\hat{x}\| = 1$, что $\langle p, \hat{x}\rangle = \|p\|_*$.

Из теоремы Фенхеля-Моро несложно получить, что $\|\cdot\|_{**} \equiv \|\cdot\|$. Для этого обозначим $g(x) = \frac{1}{2}\|x\|^2$ и вычислим g^*:

$$g^*(p) = \sup_x(\langle p, x\rangle - \frac{1}{2}\|x\|^2).$$

С одной стороны, $\langle p, x\rangle \leqslant \|x\|\|p\|_*$, поэтому $g^*(p) \leqslant \sup_x(\|x\|\|p\|_* - \frac{1}{2}\|x\|^2)$ и полученный супремум уже не зависит явно от x, а зависит только от длины $\lambda = \|x\|$. Поэтому $g^*(p) \leqslant \sup_{\lambda\geqslant 0}(\lambda\|p\|_* - \frac{1}{2}\lambda^2) = \frac{1}{2}\|p\|_*^2$ (достигается при $\lambda = \|p\|_*$). С другой стороны, при $x = \|p\|_*\hat{x}$ получаем $\langle p, x\rangle - \frac{1}{2}\|x\|^2 = \frac{1}{2}\|p\|_*^2$. Поэтому $g^*(p) = \frac{1}{2}\|p\|_*^2$. Проводя аналогичные рассуждения для $\frac{1}{2}\|p\|_*^2$, получаем $g^{**}(x) = \frac{1}{2}\|x\|_{**}^2$. Осталось сказать, что функция g, очевидно, выпуклая и определена на всем пространстве $\mathbb{R}^d$. Поэтому она непрерывна на $\mathbb{R}^d$ (по предложению 5) и, следовательно, замкнута. Поэтому по теореме Фенхеля-Моро $g^{**} = g$, что и требовалось.

Единичные шары в нормах $\|\cdot\|$ и $\|\cdot\|_*$ тоже тесно связаны (см. упражнение 11.3).

Сильная гладкость.

Определение 24. Пусть $\beta > 0$. Мы будем говорить, что функция $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ называется *β-сильно гладкой* относительно нормы $\|\cdot\|$, если множество $\mathrm{dom} f$ открыто, $f \in C^1(\mathrm{dom} f)$ и производная $f' : \mathrm{dom} f \to \mathbb{R}^{d*}$ является β-липшицевой, т.е. для любых $x_0, x_1 \in \mathrm{dom} f$ выполнено

$$\|f'(x_1) - f'(x_0)\|_* \leqslant \beta\|x_1 - x_0\|.$$

Для дальнейшего нам потребуется следующая характеризация сильно гладких функций:

Лемма 3. *Пусть $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ – собственная выпуклая функция. Тогда она является β-сильно гладкой, если и только если множество $\mathrm{dom} f$ открыто, и для всех $x \in \mathrm{dom} f$, $h \in \mathbb{R}^d$ и $p \in \partial f(x)$ неравенство*

$$f(x + h) \leqslant f(x) + \langle p, h\rangle + \frac{\beta}{2}\|h\|^2 \qquad (12)$$

выполняется при $x + h \in \mathrm{dom} f$.

Доказательство. (Необходимость). Поскольку $f \in C^1(\mathrm{dom} f)$, то необходимо $p = f'(x)$. Обозначим $\varphi(\lambda) = f(x+\lambda h)$. Итак, требуется доказать, что $\varphi(1) - \varphi(0) - \langle f'(x), h\rangle \leqslant \frac{\beta}{2}\|h\|^2$. Вычислим по формуле Ньютона-Лейбница:

$$\begin{aligned}\varphi(1) - \varphi(0) - \langle f'(x), h\rangle &= \int_0^1 \varphi'(\lambda)\,d\lambda - \langle f'(x), h\rangle = \\ &= \int_0^1 \langle f'(x+\lambda h), h\rangle\,d\lambda - \langle f'(x), h\rangle = \int_0^1 \langle f'(x+\lambda h) - f'(x), h\rangle\,d\lambda \leqslant \\ &\leqslant \int_0^1 \|h\|\|f'(x+\lambda h) - f'(x)\|_*\,d\lambda \leqslant \int_0^1 \lambda\beta\|h\|^2 = \frac{\beta}{2}\|h\|^2.\end{aligned}$$

(Достаточность). Сначала покажем, что $f \in C^1(\mathrm{dom} f)$. Субдифференциал f в любой точке $x \in \mathrm{dom} f$ непуст по предложению 7. Поэтому по предложению 8 остается доказать, что субдифференциал $\partial f(x)$ не может содержать различных элементов. Пусть $p_0, p_1 \in \partial f(x)$. Для произвольного h и достаточно малого λ получаем неравенства

$$f(x+\lambda h) \geqslant f(x) + \lambda\langle p_0, h\rangle \quad \text{и} \quad f(x+\lambda h) \leqslant f(x) + \lambda\langle p_1, h\rangle + \frac{\beta}{2}\lambda^2\|h\|^2$$

(первое – из определения субдифференциала, а второе – по условию). Вычитая эти неравенства, получаем $\lambda\langle p_1 - p_0, h\rangle + \frac{\beta}{2}\lambda^2\|h\|^2 \geqslant 0$. Деля на λ и устремляя $\lambda \to 0$, получаем $\langle p_1 - p_0, h\rangle \geqslant 0$ для всех h, что возможно, только если $p_1 = p_0$.

Докажем теперь липшицевость производной f'. Если бы функция f была хотя бы дважды дифференцируемой, сделать это было бы легко, просто оценив норму второй производной. Однако, априори, функция f может не иметь второй производной. Поступим следующим образом. Пусть $\omega(x) \geqslant 0$ – шапочка, то есть ω – неотрицательная, бесконечно-гладкая функция с компактным носителем $\mathrm{supp}\,\omega \subseteq \{\|x\| \leqslant 1\}$ и $\int_{\mathbb{R}^d} \omega(x)\,dx = 1$. Обозначим $\omega_k(x) = k^d\omega(kx)$ и $f_k = f * \omega_k$ (свертка). Интегрирование в свертке $(f * \omega_k)(x)$ ведется по $(\frac{1}{k})$-окрестности точки $x \in \mathrm{dom} f$ и поэтому корректно определено при достаточно больших k.

Теперь уже f_k есть бесконечно-гладкая функция, и $f_k' = f' * \omega_k$. Пусть $x_0, x_1 \in \mathrm{dom} f$. Докажем неравенство $\|f'(x_1) - f'(x_0)\|_* \leqslant \beta\|x_1 - x_0\|$ предельным переходом по k. Обозначим $x = x_0$, $h = x_1 - x_0$ и

$$\begin{aligned}\Lambda_k &\overset{\text{def}}{=} f_k(x+h) - f_k(x) - \langle f_k'(x), h\rangle = \\ &= \int\limits_{\|y\| \leqslant \frac{1}{k}} \big(f(x+h-y) - f(x-y) - \langle f'(x-y), h\rangle\big)\omega_k(y)\,dy.\end{aligned}$$

Данный интеграл заведомо определен, если сумма $\|h\|$ и $\frac{1}{k}$ меньше расстояния от x_0 до границы эффективной области. Итак,

$$0 \leqslant \Lambda_k = \frac{1}{2}f_k''(x)[h,h] + o(\|h\|^2) \leqslant \frac{\beta}{2}\|h\|^2.$$

Подставив λh вместо h в это неравенство и устремив λ к 0, получаем, что для всех $x \in \mathrm{dom} f$, $h \in \mathbb{R}^d$ и достаточно больших k выполнено

$$0 \leqslant f_k''(x)[h,h] \leqslant \beta\|h\|^2.$$

Следовательно, $f_k''(x)$ – неотрицательно определенная матрица при $\frac{1}{k} <$ $\mathrm{dist}(x, \partial\,\mathrm{dom} f)$, и

$$f_k''(x)[h,h'] \leqslant \sqrt{f_k''(x)[h,h]}\sqrt{f_k''(x)[h',h']} \leqslant \beta\|h\|\|h'\|.$$

Поэтому

$$\langle f_k'(x+h) - f_k'(x), h'\rangle = \int_0^1 f_k''(x+\lambda h)[h,h']\,d\lambda \leqslant \beta\|h\|\|h'\|.$$

Проверим корректность написанного равенства: отрезок $[x; x+h]$ компактен и, значит, лежит в открытом множестве $\mathrm{dom} f$ вместе с некоторой своей ε-окрестностью. Поэтому при $k > \frac{2}{\varepsilon}$ функция f_k определена, по крайней мере, в $\frac{\varepsilon}{2}$-окрестности отрезка $[x_0; x_1]$.

Таким образом, $\|f_k'(x+h) - f_k'(x)\|_* \leqslant \beta\|h\|$. Устремив k к бесконечности (при фиксированных x и h) получаем искомое:

$$\|f'(x+h) - f'(x)\| \leqslant \beta\|h\|.$$

□

Следствие 36. *Собственная выпуклая замкнутая функция $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ является β-сильно гладкой, если и только если неравенство (12) выполняется для всех x и h.*

Доказательство. (Необходимость). Если β-сильно гладкая функция замкнута, то она обязана быть конечной на всем пространстве, и поэтому $\mathrm{dom} f = \mathbb{R}^d$.

(Достаточность). Если неравенство (12) выполняется для всех x и h, то, взяв $x \in \mathrm{dom} f$ и любой вектор $h \in \mathbb{R}^d$, получаем, что $f(x+h) < \infty$, т.е. $\mathrm{dom} f = \mathbb{R}^d$. □

Теорема о двойственности сильной выпуклости и сильной гладкости.

Теорема 27. *Пусть $\|\cdot\|$ – некоторая норма на $\mathbb{R}^d$ и $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ – выпуклая, замкнутая, собственная функция. Тогда следующие два факта равносильны при $\alpha > 0$:*

- *Функция f является α-сильно выпуклой относительно нормы $\|\cdot\|$;*
- *Функция f^* является $\frac{1}{\alpha}$-сильно гладкой относительно нормы $\|\cdot\|_*$.*

Эта теорема демонстрирует один из самых естественных способов сформулировать двойственность понятий гладкости и строгой выпуклости. Здесь возможны и другие подходы (см., например, упражнения в конце параграфа).

Заметим, что по теореме Фенхеля-Моро $f = f^{**}$, поэтому, подставляя f^* вместо f в эту теорему, мы получаем, что f является β-сильно гладкой и $\mathrm{dom} f = \mathbb{R}^d$, если и только если f^* является α-сильно выпуклой.

Еще отметим, что сочетание требований замкнутости функции и ее β-сильной гладкости довольно ограничительно, например, по следующей причине:

Предложение 62. *Если собственная выпуклая функция $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ является и β-сильно гладкой, и замкнутой, то она конечна на всем пространстве, т.е. $\mathrm{dom} f = \mathbb{R}^d$.*

Доказательство. Покажем, что эффективная область $\mathrm{dom} f$ не может иметь граничных точек. Действительно, пусть $x_0 \in \partial\, \mathrm{dom} f$. Тогда $x_0 \notin \mathrm{dom} f$, так как множество $\mathrm{dom} f$ открыто. Выберем любую точку $x_1 \in \mathrm{dom} f$. Тогда $x_\lambda \to x_0$ при $\lambda \to 0$. Поскольку $x_\lambda \in \mathrm{dom} f$, то существуют производные $f'(x_\lambda)$, и они обязаны быть ограничены при $\lambda \in (0; 1]$:

$$\|f'(x_\lambda) - f'(x_1)\|_* \leqslant \beta\|x_\lambda - x_0\| \leqslant \beta\|x_1 - x_0\|.$$

Итак, производные $f'(x_\lambda)$ ограничены, и поэтому при $\lambda \to 0$ у них должна быть хотя бы одна предельная точка p_0, которая обязана лежать в $\partial f(x_0)$ в силу замкнутости f (см. предложение 19). Таким образом, субдифференциал $\partial f(x_0)$ непуст. Поэтому, как минимум, $f(x_0) < \infty$, т.е. $x_0 \in \mathrm{dom} f$. Противоречие. $\square$

Доказательство теоремы 27. Зафиксируем $x \in \mathbb{R}^d$ и $p \in \mathbb{R}^{d*}$ и обозначим $g(h) = f(x+h) - f(x) - \langle p, h\rangle$. Мы знаем, что

- Функция f является α-сильно выпуклой, если и только если $g(h) \geqslant \frac{\alpha}{2}\|h\|^2$ при всех h, x и $p \in \partial f(x)$.
- Функция f является β-сильно гладкой, если и только если $g(h) \leqslant \frac{\beta}{2}\|h\|^2$ при всех h, x и $p \in \partial f(x)$.

Вычислим сопряженную функцию g^*:

$$\begin{aligned} g^*(u) &= \sup_h(\langle u, h\rangle - f(x+h) + f(x) + \langle p, h\rangle) = \sup_h(\langle p+u, h\rangle - f(x+h)) + f(x) = \\ &= \sup_y(\langle p+u, y-x\rangle - f(y)) + f(x) = f^*(p+u) + f(x) - \langle p, x\rangle - \langle x, u\rangle. \end{aligned}$$

Поэтому если $p \in \partial f(x)$, то неравенство Юнга превращается в равенство $f(x) + f^*(p) = \langle p, x\rangle$ (см. предложение 50), и

$$g^*(u) = f^*(p+u) - f^*(p) - \langle x, u\rangle.$$

Отметим, что функция $g^*(u)$ имеет ту же структуру, что и $g(h)$, только построена по функции f^* и определена для $u \in \mathbb{R}^{d*}$.

Теперь мы готовы доказать теорему. Предположим сначала, что функция f является α-сильно выпуклой. Пусть $p \in \mathbb{R}^{d*}$ и $x \in \partial f^*(p)$. Тогда $p \in \partial f(x)$ по теореме 22 о связи субдифференциалов функции и ее сопряженной. Поэтому $g \geqslant \frac{\alpha}{2}\|\cdot\|^2$. Поэтому $g^* \leqslant (\frac{\alpha}{2}\|\cdot\|^2)^*$. Осталось заметить, что, поскольку $(\frac{1}{2}\|\cdot\|^2)^* = \frac{1}{2}\|\cdot\|_*^2$, то $(\frac{\alpha}{2}\|\cdot\|^2)^* = \frac{1}{2\alpha}\|\cdot\|_*^2$. Таким образом, $g^* \leqslant \frac{1}{2\alpha}\|\cdot\|_*^2$, т.е. функция f^* является $(\frac{1}{\alpha})$-сильно гладкой.

Предположим теперь, что функция f является β-сильно гладкой. Пусть $p \in \mathbb{R}^{d*}$ и $x \in \partial f^*(p)$. Тогда $p \in \partial f(x)$. Поэтому $g \leqslant \frac{\beta}{2}\|\cdot\|^2$. Поэтому $g^* \geqslant (\frac{\beta}{2}\|\cdot\|^2)^* = \frac{1}{2\beta}\|\cdot\|_*^2$, т.е. функция f^* является $(\frac{1}{\beta})$-сильно гладкой. $\square$

Упражнения:

11.1 Пусть $C \subset \mathbb{R}^d$ – выпуклое компактное множество и $0 \in \mathrm{int}\, C$. Докажите, что множество C строго выпукло, если и только если множество C° имеет гладкую границу (множество называется строго выпуклым, если любая его граничная точка является крайней, т.е. не может быть помещена внутрь отрезка из этого множества).

11.2 Пусть f – выпуклая замкнутая функция. Верно ли, что если функция f строго выпукла на $\mathrm{dom} f$, то множество $\mathrm{dom} f^*$ открыто и $f^* \in C^1(\mathrm{dom} f^*)$?

11.3 Докажите, что поляра единичного шара в норме $\|\cdot\|$ есть единичный шар в норме $\|\cdot\|_*$.

12 Топологические свойства выпуклых множеств

Выпуклые множества обладают очень хорошими топологическими свойствами. Наиболее важным результатом топологического характера в выпуклом анализе является теорема Брауэра о неподвижной точке (и ее обобщение – теорема Какутани), а также теорема Хелли о непустоте пересечения выпуклых компактов.

Гомеоморфность шару. Мы начнем с доказательства того, что, с топологической точки зрения, любое выпуклое множество неотличимо от шара:

***Теорема* 28.** *Пусть A и B – выпуклые компактные подмножества $\mathbb{R}^d$. Тогда если $\dim A = \dim B$, то существует такой гомеоморфизм $f : \mathbb{R}^d \to \mathbb{R}^d$, что $f(A) = B$.*

Прежде, чем доказывать эту теорему, выведем из нее несколько исключительно полезных свойств компактных выпуклых множеств.

***Следствие* 37.** *Всякое выпуклое компактное множество $A \subset \mathbb{R}^d$ гомеоморфно замкнутому шару размерности $\dim A$, его относительная граница A $\mathrm{ri}\, A$ гомеоморфна сфере размерности $\dim A - 1$, а его относительная внутренность – открытой сфере размерности $\dim A - 1$.*

Доказательство. Без ограничения общности $\dim A = d$. Положим B – единичный замкнутый шар в $\mathbb{R}^d$. Тогда отображение f из теоремы 28 есть искомый гомеоморфизм и $A \simeq B$, $\partial A \simeq \partial B$ и $\mathrm{int}\, A \simeq \mathrm{int}\, B$. □

***Следствие* 38.** *Если A и B – компактные выпуклые подмножества $\mathbb{R}^d$ и $\dim A = \dim B$, то множество A гомеоморфно B, $A \simeq B$, гомеоморфны их границы $\partial A \simeq \partial B$ и относительные внутренности $\mathrm{ri}\, A \simeq \mathrm{ri}\, B$.*

Таким образом, все компактные выпуклые подмножества $\mathbb{R}^d$ одной размерности с топологической точки зрения неотличимы от замкнутого шара. Аналогичное утверждение верно и для относительно открытых выпуклых множеств – с топологической точки зрения они неотличимы от открытого шара той же размерности (см. упражнение 12.1). Однако, замкнутые не компактные выпуклые подмножества $\mathbb{R}^d$ могут быть устроены сложнее. Например, множество $\{(x, y) : xy \geqslant 1, x > 0, y > 0\} \subset \mathbb{R}^2$ гомеоморфно кругу без одной граничной точки.

Доказательство теоремы 28. Разберем сначала главный случай: $\dim A = \dim B = d$, т.е. $\mathrm{int}\, A \neq \emptyset$ и $\mathrm{int}\, B \neq \emptyset$. Используя параллельный перенос, подвинем множества A и B таким образом, чтобы $0 \in \mathrm{int}\, A$ и $0 \in \mathrm{int}\, B$. Рассмотрим отображение $f : \mathbb{R}^d \to \mathbb{R}^d$,

$$f_{AB}(x) = \frac{\mu_A(x)}{\mu_B(x)} x \text{ для } x \neq 0 \text{ и } f_{AB}(0) = 0,$$

где μ_A и μ_B – функции Минковского множеств A и B соответственно.

Докажем, что f_{AB} – искомый гомеморфизм. Для этого нам необходимо проверить несколько фактов:

1. *Отображение f_{AB} задано корректно*, так как $\mu_B(x) \neq 0$ при $x \neq 0$ (так как $0 \in \operatorname{int} B$) и $\mu_A(x) \neq +\infty$ (так как множество A ограничено).

2. *Отображение f_{AB} биективно.* Действительно, мы утверждаем, что отображение f_{BA}, заданное равенствами $f_{BA}(x) = \frac{\mu_B(x)}{\mu_A(x)} x$ для $x \neq 0$ и $f_{BA}(0) = 0$, обратно к нему. Отображение f_{BA} заданно корректно по тем же причинам, что и отображение f_{AB}. Для $x \neq 0$ получаем

$$f_{BA}(f_{AB}(x)) = \frac{\mu_B(f_{AB}(x))}{\mu_A(f_{AB}(x))} f_{AB}(x) = \frac{\mu_B(\frac{\mu_A(x)}{\mu_B(x)} x)}{\mu_A(\frac{\mu_A(x)}{\mu_B(x)} x)} \frac{\mu_A(x)}{\mu_B(x)} x = x,$$

   так как функция Минковского положительно однородна: $\mu_{A,B}(\lambda x) = \lambda \mu_{A,B}(x)$ для $\lambda > 0$. При $x = 0$, очевидно, получаем $f_{BA}(f_{AB}(0)) = 0$.

3. *Отображения f_{AB} и f_{BA} непрерывны.* Действительно, функции $\mu_A(x)$ и $\mu_B(x)$ непрерывны, так как они являются выпуклыми и опеределены на всем пространстве: $\operatorname{dom} \mu_A = \operatorname{dom} \mu_B = \mathbb{R}^d$. Поскольку $\mu_B(x) \neq 0$ при $x \neq 0$, то функция $f_{AB}(x)$ непрерывна при $x \neq 0$. Покажем ее непрерывность в 0: если $x \to 0$, то $\mu_A(x) \to 0$, а отношение $\frac{x}{\mu_B(x)} \in B$ ограничено, поэтому $f_{AB}(x) \to 0$. Непрерывность f_{BA} доказывается аналогично.

Таким образом, отображение f_{AB} является биекцией и непрерывно, и обратное к нему тоже непрерывно. Следовательно, f_{AB} – гомеоморфизм. Заметим, что $f_{AB}(A) \subset B$ и $f_{BA}(B) \subset A$ по построению. Поэтому $f_{AB}(A) = B$, что и требовалось.

Осталось разобрать общий случай: $\dim A = \dim B = k < d$. Представим $\mathbb{R}^d$ как сумму $\mathbb{R}^d = \mathbb{R}^k \oplus \mathbb{R}^{d-k}$. Поскольку $\dim \operatorname{aff} A = \dim \operatorname{aff} B = k$, то, используя параллельный перенос и поворот, можно добиться того, что $\operatorname{aff} A = \operatorname{aff} B = \mathbb{R}^k$. Поскольку и A, и B имеют непустые внутренности в $\mathbb{R}^k$, то по уже доказанному существует такой гомеоморфизм $\tilde{f} : \mathbb{R}^k \to \mathbb{R}^k$, что $\tilde{f}(A) = B$. Определим $f : \mathbb{R}^d \to \mathbb{R}^d$ так: на компоненте $\mathbb{R}^k$ отображения f и $\tilde{f}$ совпадают, а на компоненте $\mathbb{R}^{d-k}$ отображение f тождественно:

$$y \in \mathbb{R}^k, \; z \in \mathbb{R}^{d-k} \quad \Rightarrow \quad f(y, z) = (\tilde{f}(y), z).$$

Построенное отображение f, очевидно, является искомым. $\square$

Теоремы Брауэра и Какутани. *Теорему Брауэра* удобнее всего воспринимать как теорему существования: она гарантирует разрешимость уравнения $\Phi(x) = x$ в случае если Φ является отображением выпуклого множества в себя:

***Теорема* 29** (Брауэра)**.** *Пусть B – компактное выпуклое множество в конечномерном пространстве $\mathbb{R}^d$. Тогда любое непрерывное отображение Φ множества B в себя обязано иметь неподвижную точку:*

$$\varphi \in C(B \to B) \quad \Longrightarrow \quad \exists \hat{x} \in B : \varphi(\hat{x}) = \hat{x}.$$

Доказательство. Используя теорему 28, будем без ограничения общности считать, что B – единичный шар в $\mathbb{R}^d$.

Доказательство проведем от противного: предположим, $\varphi(x) \in B$ для любого $x \in B$, но не существует такой точки $x \in B$, что $\Phi(x) = x$. Рассмотрим луч $l(x)$, выходящий из точки $\varphi(x)$ и проходящий через x. Такой луч определен для всех x, так как $x \neq \varphi(x)$. Двигаясь по лучу $l(x)$ от $\varphi(x)$ к x и дальше, мы дойдем до пересечения $l(x)$ с границей шара ∂B. Пересечение $l(x) \cap \partial B$ состоит из одной точки[8]. Обозначим ее через $\theta(x)$. Таким образом, θ – это отображение шара B на сферу ∂B. Отображение θ непрерывно, так как отображение f непрерывно. При этом отображение θ оставляет точки на границе на месте: если $x \in \partial B$, то луч $l(x)$ пересекает ∂B в x.

Таким образом, если $x \neq \Phi(x)$ для всех $x \in B$, то мы построили непрерывное отображение θ, которое переводит точки шара B на его границу ∂B, оставляя при этом точки границы на месте. Но такого непрерывного отображения существовать не может! Докажем это строго. Для этого рассмотрим серию отображений $i_\lambda : \partial B \to B$: отображение i_λ сжимает сферу ∂B к центру B в $\lambda \in [0;1]$ раз. Композиции $\theta \circ i_\lambda$

$$\theta \circ i_\lambda : \partial B \xrightarrow{i_\lambda} B \xrightarrow{\theta} \partial B$$

задают непрерывную деформацию сферы ∂B при $\lambda \in [0;1]$. Проследим за ней при λ, меняющемся от 1 к 0. При $\lambda = 1$ отображение $\theta \circ i_1$ есть тождественное отображение, которое оставляет все точки сферы ∂B на месте; а при $\lambda = 0$ отображение $\theta \circ i_0$ есть отображение всей сферы в одну точку (так как $i_0(\partial B)$ есть центр сферы). Такой непрерывной деформации существовать не может, так как сфера не стягиваема. □

Теорема Брауэра имеет прямое обобщение на бесконечномерный случай:

***Теорема* 30** (*Шаудера*)**.** *Пусть X – локально-выпуклое топологическое линейное пространство (например, банахово или гильбертово пространство) и $B \subset X$ – выпуклый компакт. Тогда любое непрерывное отображение $f : B \to B$ имеет неподвижную точку.*

Теорема Брауэра носит чисто топологический характер – она верна для любого множества, гомеоморфного шару. Однако (как мы видели в доказательстве), выпуклость шара играет большое значение. Еще одно обобщение теоремы Брауэра – *теорема Какутани* – верна в выпуклой геометрии, но уже не переносится на произвольный топологический случай.

***Теорема* 31** (Какутани)**.** *Пусть $B \subset \mathbb{R}^d$ – выпуклый компакт и $\Phi : B \to 2^B$ – многозначное отображение, ставящее в соответствие каждой точке $x \in B$ непустое выпуклое компактное подмножество $\Phi(x) \subset B$. Тогда если его график*

[8]Так как шар B – строго выпуклое, компактное множество. Существует обобщение теоремы Брауэра на бесконечномерный случай – теорема Шаудера, – где B уже не может быть шаром (так как шар в этом случае некомпактен), но может быть любым компактным выпуклым множеством.

$$\operatorname{graph}\Phi = \{(x,y) : x \in B, y \in \Phi(x)\} \subset B \times B$$
замкнут, то найдется неподвижная точка $\hat{x}$, т.е. $\hat{x} \in \Phi(\hat{x})$.

Отметим, что если множество $\Phi(x)$ является одноточечным при любом x, то замкнутость графика Φ равносильна непрерывности Φ как обычного отображения, и мы получаем просто теорему Брауэра.

Доказательство теоремы 31. Разберем сначала главный случай, когда множество B является d-мерным симплексом. Разобьем B на маленькие симплексы размера не больше $1/n$ и построим отображение $\varphi_n : B \to B$ с помощью этого разбиения. Для каждой вершины x из разбиения выберем любой элемент $y \in \Phi(x)$ и положим $\varphi_n(x) \stackrel{\text{def}}{=} y$. Таким образом, мы задали φ_n на вершинах разбиения. Продолжим отображение φ_n внутрь любого симплекса $\Delta = \operatorname{conv}\{x_1, \dots, x_{d+1}\}$ построенного разбиения по линейности: $\varphi_n(\sum_{i=1}^{d+1} \lambda_i x_i) = \sum_{i=1}^{d+1} \lambda_i \varphi_n(x_i)$, $\lambda_i \geqslant 0$, $\sum \lambda_i = 1$. Поскольку на вершинах $\varphi_n(x_i) \in B$, то $\varphi_n(x) \in B$ для любого x в силу выпуклости B. Отметим, что построенное отображение φ_n корректно определено на гранях разбиения и непрерывно. Поэтому по теореме Брауэра у него найдется неподвижная точка $\hat{x}^n = \varphi_n(\hat{x}^n)$. Обозначим $x_1^n, \dots, x_{d+1}^n$ – вершины симплекса разбиения, в котором лежит $\hat{x}^n$ (т.е. $\hat{x}_n = \sum_{i=1}^{d+1} \lambda_i^n x_i^n$), и обозначим $y_i^n = \varphi_n(x_i^n) \in \Phi(x_i^n)$. Ввиду линейности φ_n на симплексе, получаем $\hat{x}^n = \sum_{i=1}^{d+1} \lambda_i^n y_i^n$.

Таким образом, мы построили при $n \to \infty$ такие последовательности точек $\hat{x}^n$, x_i^n, y_i^n и чисел λ_i^n, $i = 1 \dots, d+1$, что

$$\operatorname{dist}(x_i^n, x_j^n) \to 0, \quad y_i^n \in \Phi(x_i^n) \quad \text{и} \quad \hat{x}^n = \sum_{i=1}^{d+1} \lambda_i^n x_i^n = \sum_{i=1}^{d+1} \lambda_i^n y_i^n.$$

Поскольку все построенные последовательности лежат в компактных множествах ($\hat{x}^n, x_i^n, y_i^n \in B$ и $\lambda^n \in \{\lambda : \lambda_i \geqslant 0, \sum_{i=1}^{d+1} \lambda_i = 1\}$), то, выбирая подпоследовательности, будем считать, что все они имеют пределы при $n \to \infty$:

$$\hat{x}^n \to \hat{x}, \quad x_i^n \to x_i, \quad y_i^n \to y_i, \quad \lambda_i^n \to \lambda_i.$$

При этом $\operatorname{dist}(x_i^n, x_j^n) \to 0$, поэтому $x_i = \hat{x}$ при всех i. Значит, $y_i \in \Phi(\hat{x})$, так как $(x_i^n, y_i^n) \to (\hat{x}, y_i)$, а график Φ замкнут по условию теоремы. Осталось воспользоваться выпуклостью $\Phi(\hat{x})$: поскольку $y_i \in \Phi(\hat{x})$, то точка $\hat{x} = \sum_{i=1}^{d+1} \lambda_i y_i$ тоже лежит в $\Phi(\hat{x})$, что и требовалось.

Доказательство для случая произвольного выпуклого компактного множества B получается очень просто. Поместим B внутрь произвольного симплекса Δ и продолжим отображение Φ с B на Δ. Например, это можно сделать так: для любой точки $x \in \Delta$ выбрать (единственную в силу выпуклости B) ближайшую к ней точку $\psi(x)$ в B и доопределить $\Phi(x) = \Phi(\psi(x))$ для $x \in \Delta$ (заметим, $\psi(x) = x$ для $x \in B$). Поскольку отображение ψ непрерывно (в силу выпуклости B), то Φ на Δ удовлетворяет условиям теоремы, и его неподвижная точка $\hat{x} \in \Delta$ должна лежать в B, так как $\Phi(x) \subset B$ при всех $x \in \Delta$. □

Теоремы Радона и Хелли.

Теорема 32 (*Радона*)**.** *В произвольном множестве из $d+2$ или более точек $\mathbb{R}^d$ можно выбрать два непересекающихся подмножества, чьи выпуклые оболочки пересекаются.*

Доказательство. Сначала докажем теорему для подмножеств, состоящих ровно из $d+2$ точек. Пусть $x_1, \dots, x_{d+2} \in \mathbb{R}^d$. Будем действовать аналогично теореме Каратеодори. Рассмотрим векторы $x_i - x_{d+2}$, $i = 1, \dots, d+1$. Поскольку их на 1 больше, чем размерность пространства, то они должны быть линейно зависимыми, т.е. найдутся такие не равные одновременно 0 числа a_i, $i = 1, \dots, d+1$, что $\sum_{i=1}^{d+1} a_i(x_i - x_{d+2}) = 0$. Обозначая $a_{d+2} = -\sum_{i=1}^{d+1} a_i$, получаем

$$\sum_{i=1}^{d+2} a_i x_i = 0 \quad \text{и} \quad \sum_{i=1}^{d+2} a_i = 0.$$

Поскольку сумма всех коэффициентов a_i равна 0, то часть из них должна быть положительной, часть отрицательной и, возможно, часть равна 0. Обозначим все положительные коэффициенты a_i через $\beta_1, \dots, \beta_k$, а отвечающие им точки x_i через $y_1, \dots, y_k$. Аналогично, обозначим модули всех отрицательных коэффициентов a_i через $\gamma_1, \dots, \gamma_m$, а отвечающие им точки x_i через $z_1, \dots, z_m$. Тогда $k + m \leqslant d+2$ и

$$\sum_{i=1}^{k} \beta_i y_i = \sum_{i=1}^{m} \gamma_i z_i \overset{\text{def}}{=} \hat{x} \quad \text{и} \quad \sum_{i=1}^{k} \beta_i = \sum_{i=1}^{m} \gamma_i \overset{\text{def}}{=} \hat{a}.$$

Поскольку $\hat{a} > 0$, точка $\hat{x}/\hat{a}$ представляется и в виде выпуклой комбинации точек y_i, и в виде выпуклой комбинации точек z_i. Следовательно,

$$\hat{x}/\hat{a} \in \operatorname{conv}\{y_1, \dots, y_k\} \cap \operatorname{conv}\{z_1, \dots, z_m\} \neq \emptyset,$$

что и требовалось.

Для бо́льших подмножеств теорема получается элементарно – достаточно выбрать любое подмножество ровно из $d+2$ точек, и далее выбрать в нем два непересекающихся подмножества по уже доказанному. □

Теорему Радона часто формулируют немного другим (но эквивалентным) способом:

***Следствие* 39.** *Произвольное множество из $d+2$ или более точек $\mathbb{R}^d$ можно разбить на два непересекающихся подмножества, чьи выпуклые оболочки пересекаются.*

Доказательство. Если $A \subset \mathbb{R}^d$ и $\#A \geqslant d+2$, то по уже доказанной теореме Радона найдутся такие непересекающиеся подмножества $A_1, A_2 \subset A$, что $\operatorname{conv} A_1 \cap \operatorname{conv} A_2 \neq \emptyset$. Обозначим $A_3 = A\ (A_1 \cup A_2)$. Тогда в качестве нужного разбиения можно взять $A = A_1 \sqcup (A_2 \cup A_3)$ или $A = (A_1 \cup A_3) \sqcup A_2$. □

***Теорема* 33** (*Хелли*)**.** *Пусть C_α, $\alpha \in I$, – выпуклые подмножества $\mathbb{R}^d$. Предположим, что любые $d+1$ из них имеют общую точку. Тогда любые $n \in \mathbb{N}$ из них имеют общую точку. Если дополнительно все множества C_α замкнуты и хотя бы одно из них компактно, то все они имеют общую точку.*

Доказательство. Сначала докажем первую часть теоремы индукцией по n. База индукции при $n \leqslant d+1$ уже есть. Сделаем шаг. Пусть $C_{\alpha_1}, \dots, C_{\alpha_{n+1}}$ – произвольные множества из набора, $n \geqslant d+1$. По индуктивному предположению, любой набор из n множеств имеет общую точку. Обозначим через x_1 общую точку множеств $C_{\alpha_2}, \dots, C_{\alpha_{n+1}}$. Аналогично, обозначим через x_j общую точку всех множеств, исключая j-ое множество:

$$x_j \in C_{\alpha_1} \cap \dots \cap C_{\alpha_{j-1}} \cap C_{\alpha_{j+1}} \cap \dots \cap C_{\alpha_{n+1}}.$$

Тогда множество $\{x_1, \dots, x_{n+1}\}$ содержит не менее $d+2$ точек и, следовательно, по теореме Радона может быть разбито на два непересекающихся подмножества, выпуклые оболочки которых пересекаются. Без ограничения общности $\operatorname{conv}\{x_1, \dots, x_m\} \cap \operatorname{conv}\{x_{m+1}, \dots, x_n\} \neq \emptyset$. Возьмем точку $\hat{x}$ из этого пересечения. Мы утверждаем, что она лежит в каждом множестве C_{a_i}. Действительно, если $i \leqslant m$, то $\hat{x} \in C_{a_i}$ по следующей причине: точки $x_{m+1}, \dots, x_n$ лежат в C_{a_i} по построению, а, значит, $\operatorname{conv}\{x_{m+1}, \dots, x_n\} \in C_{a_i}$ (так как множество C_{a_i} выпукло), т.е. $\hat{x} \in C_{a_i}$. Если $i \geqslant m+1$, то все получается аналогично, так как в C_{a_i} теперь лежат точки $x_1, \dots, x_m$.

Докажем теперь вторую часть теоремы. Обозначим $B_a = \mathbb{R}^d\ C_a$. Тогда B_a – открытые множества. Доказательство проведем от противного. Если $\bigcap_a C_a = \emptyset$, то $\bigcup_a B_a = \mathbb{R}^d$. Пусть C_{a_0} – компактное выпуклое множество из набора C_a. Тогда открытые множества B_a покрывают компакт C_{a_0}. Выберем конечное подпокрытие: $C_{a_0} \subset \bigcup_{i=1}^n B_{a_i}$. Следовательно, $\bigcap_{i=0}^n C_{a_i} = \emptyset$, что противоречит первой части теоремы. □

Пример 14. Отказаться от требования компактности хотя бы одного из множеств нельзя – в этом случае пересечение $\bigcap_a C_a$ может оказаться пустым. Простейший пример такой: $C_a = [a; +\infty) \subset \mathbb{R}$, $a \subset \mathbb{R}$. Все множества замкнуты, но ни одного компактного среди них нет. Любой конечный набор имеет общую точку, а вот общей точки у всех множеств нет.

Крайние точки и Теорема Минковского-Штейнера. Линейное подпространство удобно описывать как линейную оболочку набора некоторых его векторов. Аналогично, выпуклое множество удобно описывать как выпуклую оболочку набора некоторых его точек. И в первом, и во втором случаях встает естественный вопрос: как выбрать этот набор поменьше? В первом случае мы знаем ответ: минимальный такой набор – это базис линейного подпространства. Для того, чтобы получить ответ на второй вопрос, поймем, какие точки выпуклого множества невозможно представить в виде выпуклой оболочки других точек из этого множества. Такие точки называют *крайними точками* выпуклого множества.

Определение 25. Точка $x \in C$ называется *крайней точкой* выпуклого множества $C \subset \mathbb{R}^d$, если она не является внутренней точкой никакого отрезка из C. Множество крайних точек C обозначают через $\operatorname{extr} C$.

Ясно, что если мы хотим описать выпуклое множество как выпуклую оболочку, то без крайних точек обойтись не получится. Оказывается, что если выпуклое множество компактно, то крайние точки полностью описывают множество:

***Теорема* 34** (*Минковского-Штейнера*)**.** *Выпуклый компакт в B есть замыкание выпуклой оболочки своих крайних точек:*

$$B = \operatorname{conv} \operatorname{extr} B.$$

Доказательство. Без ограничения общности $\dim B = d$. Включение $\operatorname{conv} \operatorname{extr} B \subset B$ очевидно. Доказательство включения $B \subset \operatorname{conv} \operatorname{extr} B$ проще всего провести индукцией по d. Если $d = 0, 1$, то B – это точка или отрезок и включение, очевидно, выполняется. Далее, пусть $d \geqslant 2$. Сначала докажем, что $B = \operatorname{conv} \partial B$ – действительно, если $x \in \operatorname{int} B$, то проведем через x любую прямую, и она пересечет ∂B в двух точках. Теперь докажем, что $\partial B \subset \operatorname{conv} \operatorname{extr} B$.

Выберем точку $x \in \partial B$ и проведем опорную гиперплоскость H к B в x (она существует по первой теореме отделимости). Множество $B \cap H$ выпукло и имеет размерность строго меньше d. Поскольку $x \in B \cap H$, то по индуктивному предположению $x \in \operatorname{conv}\operatorname{extr}(B \cap H)$. Осталось заметить, что, поскольку H – опорная гиперплоскость, то любая крайняя точка $B \cap H$ является крайней точкой B. Таким образом, $B = \operatorname{conv}\partial B \subset \operatorname{conv}\operatorname{conv}\operatorname{extr} B = \operatorname{conv}\operatorname{extr} B$, что и требовалось. □

Теорема Минковского-Штейнера верна и в бесконечномерном случае в локально-выпуклых топологических пространствах и носит название *теоремы Крейна-Мильмана*. Только в бесконечномерном случае в формулу надо еще добавить замыкание: $B = \operatorname{cl}\operatorname{conv}\operatorname{extr} B$. Доказательство почти такое же, только необходимо использовать вместо обычной индукции трансфинитную (лемму Цорна).

Выпуклые многогранники. Выпуклый многогранник в $\mathbb{R}^d$ имеет два эквивалентных определения:

Определение 26. *выпуклым многогранником* в $\mathbb{R}^d$ называется выпуклая оболочка конечного числа точек.

Определение 27. Пересечение конечного числа замкнутых полупространств в $\mathbb{R}^d$ называется *Выпуклым многогранником*, если оно компактно.

Априори не ясно, почему эти определения совпадают. Другими словами, почему выпуклая оболочка конечного числа точек может быть представлена как пересечение конечного числа полупространств, и наоборот. Тем не менее, докажем, что это так.

***Теорема* 35** (об эквивалентности двух определений выпуклого многогранника)**.** *Данные определения выпуклого многогранника эквивалентны.*

Пока теорема не доказана, будем для определенности называть эти два определения так: выпуклый многогранник в смысле прямого описания и в смысле двойственного.

Сначала докажем удобную для нас лемму:

***Лемма* 4.** *Пусть $\Omega \subset \mathbb{R}^d$ – выпуклое компактное множество с непустой внутренностью и $0 \in \operatorname{int}\Omega$. Тогда если Ω является многогранником в смысле прямого описания, то поляра Ω° является многогранником в смысле двойственного.*

Доказательство. Итак, пусть $\Omega = \operatorname{conv}\{x_1, \dots, x_n\}$. Поскольку 0 лежит в $\operatorname{int}\Omega$, то множество $\Omega^\circ = \{p : \sup_{x \in \Omega}\langle p, x\rangle \leqslant 1\}$ ограничено. Покажем, что оно есть пересечение конечного числа полупространств. Если $p \in \Omega^\circ$, то, очевидно, $\langle p, x_i\rangle \leqslant 1$ для всех $i = 1, \dots, n$. Более того, если $\langle p, x_i\rangle \leqslant 1$ для всех i, то $p \in \Omega^\circ$, так как

$$\langle p, \sum_{i=1}^{n} \lambda_i x_i\rangle = \sum_{i=1}^{n} \lambda_i \langle p, x_i\rangle \leqslant 1$$

при $\lambda_i \geqslant 0$ и $\sum_i \lambda_i = 1$. Поэтому

$$\Omega^\circ = \bigcap_{i=1}^{n} \{p : \langle p, x_i\rangle \leqslant 1\},$$

т.е. Ω° есть пересечение конечного числа полупространств $\{p : \langle p, x_i\rangle \leqslant 1\}$, что и требовалось. $\square$

Доказательство теоремы 35. Сначала докажем, что многогранник в смысле двойственного описания имеет конечное количество крайних точек – этого будет достаточно по теореме Минковского-Штейнера.

Итак, пусть $\Omega = \bigcap_{i=1}^{n}\{x : \langle p_i, x\rangle \leqslant b_i\}$. Обозначим для $x \in \Omega$ через $I(x)$ множество таких индексов i, что $\langle p_i, x\rangle = b_i$. Докажем, что если $\hat{x} \in \Omega$ – крайняя точка, то система линейных уравнений $\langle p_i, x\rangle = b_i$, $i \in I(\hat{x})$ имеет единственное решение. Действительно, система имеет по крайней мере одно решение $\hat{x}$. Следовательно, решения системы образуют некоторое аффинное подпространство A размерности $k \geqslant 0$. Если $k \geqslant 1$, то пересечение A с открытыми полупространствами $\{x : \langle p_i, x\rangle \leqslant 1\} < b_i$, $i \notin I(\hat{x})$, является относительно открытым выпуклым множеством размерности $k \geqslant 1$, лежит в Ω и содержит $\hat{x}$, что невозможно, так как $\hat{x}$ – крайняя точка Ω.

Итак, рассмотрим все возможные наборы индексов $I \subset \{1, \dots, n\}$. Их конечное количество. Оставим только те из них, для которых система $\langle p_i, x\rangle = b_i$, $i \in I(\hat{x})$ имеет единственное решение x_I. Таким образом, все крайние точки Ω содержатся в конечном множестве $\{x_I\}$, что и требовалось.

Остается показать, что многогранник Ω в смысле прямого описания является многогранником в смысле двойственного. Здесь нам и потребуется поляра. Пусть сначала Ω имеет непустую внутренность, и в этом случае будем без ограничения общности считать, что $0 \in \operatorname{int}\Omega$. Тогда множество Ω° является многогранником в смысле двойственного описания по лемме 4. Значит, согласно только что доказанному, Ω° является многогранником в смысле прямого описания, а биполяра $\Omega^{\circ\circ} = \Omega$ является многогранником в смысле двойственного описания опять по лемме 4, что и требовалось.

Разберем теперь общий случай: $\operatorname{int}\Omega = \emptyset$. Тогда аффинное подпространство $\operatorname{aff}\Omega$ можно задать с помощью конечного числа равенств $\langle p_i, x\rangle = b_i$ (или пар неравенств $\langle p_i, x\rangle \gtrless b_i$). Внутри $\operatorname{aff}\Omega$ выпуклое множество Ω имеет непустую внутренность, и, значит, по уже доказанному, Ω можно задать внутри $\operatorname{aff}\Omega$ как пересечение конечного числа полупространств. $\square$

***Следствие* 40.** *Если $\Omega \subset \mathbb{R}^d$ – выпуклый многогранник и $0 \in \operatorname{int}\Omega$, то его поляра Ω° тоже является выпуклым многогранником.*

***Следствие* 41.** *Компактное выпуклое множество является выпуклым многогранником, если и только если множество его крайних точек конечно.*

Доказательство. Прямо вытекает из теоремы Минковского-Штейнера. $\square$

Тот факт, что многогранник имеет конечное количество крайних точек (называемых вершинами многогранника) очень удобен при нахождении максимума выпуклой функции. Напомним, что минимум выпуклой функции обычно искать намного проще, чем ее максимум (вспомним хотя бы теорему Каруша-Куна-Таккера). Также, минимум выпуклой замкнутой функции на компакте должен достигаться (в силу ее полунепрерывности снизу). Если же мы ищем максимум выпуклой (и даже замкнутой) функции на компакте, то он может не достигаться. А даже если он и достигается, то условия типа Лагранжа нет. Тем не менее, в случае многогранника все сильно упрощается:

***Теорема* 36.** *Пусть $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ – выпуклая собственная функция и Ω – выпуклый многогранник. Тогда f достигает максимума на Ω и найдется вершина, в которой принимается максимальное значение.*

Доказательство. Пусть $M = \sup_{x\in\Omega} f(x)$ и x_i, $i = 1, \dots, n$ – вершины Ω. Если $f(x_i) \leqslant M - \varepsilon$, $\varepsilon > 0$, для всех i, то по неравенству Йенсена для любой точки $x \in \Omega = \mathrm{conv}\{x_1, \dots, x_n\}$ получаем

$$f(x) = f(\sum_{i=1}^{n} \lambda_i x_i) \leqslant \sum_{i=1}^{n} \lambda_i f(x_i) < \sum_{i=1}^{n} \lambda_i (M - \varepsilon) = M - \varepsilon,$$

так как $\lambda_i \geqslant 0$ и $\sum_i \lambda_i = 1$. □

Таким образом, для того, чтобы найти максимум выпуклой функции (например, линейной) на многограннике, достаточно перебрать его вершины.

Замечание 4. В доказательстве теоремы 35 мы получили грубую оценку на число вершин многогранника: если он задается системой из n неравенств, то число вершин не превосходит 2^n.

Вершин конечное количество, однако, при больших размерностях их может быть очень много, поэтому множество исследований посвящено вопросам быстрого перебора вершин многогранника. Наиболее известный результат в этом направлении – симплексный метод.

Упражнения:

12.1 Докажите, что любое открытое звездное множество $A \subset \mathbb{R}^d$ гомеоморфно открытому шару в $\mathbb{R}^d$.

12.2 Пусть $A \subset \mathbb{R}^d$ – выпуклое множество и P – проективное преобразование на $\mathbb{R}^d$. Докажите, что если $P(x)$ не уходит на бесконечность ни при каком $x \in A$, то $P(A)$ – выпуклое подмножество $\mathbb{R}^d$.

12.3 Верно ли, что множество крайних точек компактного выпуклого множества компактно?

12.4 Легко понять, что выпуклая замкнутая функция f не обязана достигать своего максимума на выпуклом компактном множестве C. Верно ли, что если она его все-таки достигает, то найдется точка из $\mathrm{extr}\, C$ с максимальным значением?