Методы оптимизации, ВМК, осень 2018

Семинар 10: Субдифференциалы

1 Определение и базовые свойства

Для дифференцируемой функции f ее градиент ∇f позволяет построить *линеаризацию* $x \mapsto f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle$, достаточно точно приближающую функцию f в окрестности точки x_0. Если функция f выпуклая, то эта линеаризация обладает дополнительным свойством: она является *глобальной аффинной нижней оценкой* на функцию f, совпадающую с ней в точке x_0. Оказывается, что для минимизации функций достаточно иметь лишь *одностороннюю* аппроксимацию указанного вида. Поэтому вместо градиента можно рассматривать понятие *субградиента*.

Определение 1.1 (Субградиент и субдифференциал)**.** Пусть $f : E \to \mathbb{R}$ — функция, определенная на множестве E в евклидовом пространстве V, и пусть $x_0 \in E$. Вектор $s \in V$ называется *субградиентом функции f в точке x_0*, если

$$f(x) \geq f(x_0) + \langle s, x - x_0 \rangle$$

для всех $x \in E$. Множество всевозможных субградиентов функции f в точке x_0 называется *субдифференциалом функции f в точке x_0* и обозначается $\partial f(x_0)$. Функция $\partial f : E \to 2^V$, ставящая каждой точке $x \in E$ в соответствие множество $\partial f(x)$ в пространстве V, называется *субдифференциалом функции f*. Если $\partial f(x_0) \neq \emptyset$, то функция f называется *субдифференцируемой в точке x_0*; если f субдифференцируема в любой точке области определения, то f называется просто *субдифференцируемой*.

Замечание 1.2. Как следует из определения, субградиент/субдифференциал (точно так же, как и обычный градиент для дифференцируемой функции) зависят от введенного в пространстве V скалярного произведения $\langle , \rangle$. Таким образом, при работе с субградиентами/субдифференциалами необходимо четко указывать, какое конкретно используется скалярное произведение. Всюду в дальнейшем, если не оговорено иное, будем считать, что во всех пространствах, с которыми мы работаем, введено стандартное скалярное произведение, и субградиенты/субдифференциалы определяются относительно него. Отметим также, что существует более общий взгляд на рассматриваемую теорию, в котором субградиент определяется как элемент *двойственного пространства* V^* всевозможных непрерывных линейных функционалов на V; в этом случае $s \in V^*$ является линейным функционалом, а операция $\langle s, x \rangle$ интерпретируется как вычисление линейного функционала s на аргументе x. При таком определении никаких проблем вышеуказанного характера не возникает (и при этом даже можно работать с пространствами, на которых скалярное произведение ввести невозможно). Тем не менее, для простоты и понятности мы не будем рассматривать это обобщение.

Пример 1.3. Пусть $f : \mathbb{R} \to \mathbb{R}$ — функция $f(x) := |x|$. Тогда $\partial f(0) = [-1, 1]$.

Доказательство. Пусть $s \in [-1, 1]$. Тогда $sx \leq |s||x| \leq |x|$ для всех $x \in \mathbb{R}$, т. е., по определению, $s \in \partial f(0)$. В силу произвольности s, это означает, что $[-1, 1] \subseteq \partial f(0)$. Остается показать, что вложение, на самом деле, является равенством.

Предположим, чтобы прийти к противоречию, что найдется $s \in \partial f(0)$, такое, что $|s| > 1$. Тогда, по определению субградиента, $|x| \geq sx$ для всех $x \in \mathbb{R}$. Полагая $x = s$, получаем $|s| \geq |s|^2$, что невозможно. $\square$

Из определения сразу же следует, что в любой точке $x_0 \in E$ субдифференциал $\partial f(x_0)$ является выпуклым и замкнутым множеством как пересечение $\cap_{x \in E \setminus \{x_0\}} H_x$ замкнутых полупространств $H_x := \{s \in V : \langle s, x - x_0 \rangle \leq f(x) - f(x_0)\}$.

Покажем, что для дифференцируемых функций субградиент является обобщением градиента.

Утверждение 1.4 (Субдифференциал дифференцируемой функции)**.** *Пусть $f : E \to \mathbb{R}$ — функция, определенная на множестве E в евклидовом пространстве V. Пусть $x_0 \in \mathrm{int}(E)$, и пусть f дифференцируема в точке x_0. Тогда либо $\partial f(x_0) = \emptyset$, либо $\partial f(x_0) = \{\nabla f(x_0)\}$. Если при этом функция f выпуклая, то первый вариант невозможен.*

Доказательство. Предположим, чтобы прийти к противоречию, что $s \in \partial f(x_0)$ для некоторого $s \in V$, отличного от $\nabla f(x_0)$. Пусть $v \in V$ — единичный вектор. В силу того, что x_0 — внутренняя точка E, найдется $\delta > 0$, такое, что $x_0 + tv \in E$ для всех $0 < t < \delta$. Из определения субградиента, получаем

$$f(x_0 + tv) \geq f(x_0) + t\langle s, v\rangle$$

для всех $0 < t < \delta$, откуда

$$\frac{f(x_0 + tv) - f(x_0)}{t} \geq \langle s, v\rangle$$

для всех $0 < t < \delta$. Переходя к пределу и используя определение градиента, получаем

$$\langle \nabla f(x_0), v\rangle = \lim_{t\to 0; 0<t<\delta} \frac{f(x_0 + tv) - f(x_0)}{t} \geq \langle s, v\rangle.$$

Отсюда $\langle s - \nabla f(x_0), v\rangle \geq 0$. В силу произвольности, можно положить $v = -\frac{s - \nabla f(x_0)}{\|s - \nabla f(x_0)\|}$, что дает $s = \nabla f(x_0)$.

Если же функция f выпуклая, то, согласно дифференциальному условию выпуклости (см. конспект по выпуклым функциям), $f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0\rangle$ для всех $x \in E$. Но это, по определению, означает $\nabla f(x_0) \in \partial f(x_0)$. $\square$

Пример 1.5. Пусть $f : \mathbb{R} \to \mathbb{R}$ — выпуклая функция $f(x) := |x|$. Согласно примеру 1.3, $\partial f(0) = [-1, 1]$. Пусть теперь $x_0 \neq 0$. Тогда функция f дифференцируема в точке x_0 с производной (градиентом) $\operatorname{sign}(x_0)$. Значит, $\partial f(x_0) = \{\operatorname{sign}(x_0)\}$. Таким образом,

$$\partial f(x_0) = \begin{cases} [-1, 1], & \text{если } x_0 = 0, \\ \{\operatorname{sign}(x_0)\}, & \text{если } x_0 \neq 0. \end{cases}$$

Оказывается, что для выпуклой функции верно и обратное: если субдифференциал $\partial f(x_0)$ состоит из одного элемента, то функция f является дифференцируемой в точке x_0, и при этом этот единственный элемент субдифференциала является градиентом $\nabla f(x_0)$. Таким образом, утверждение 1.4 можно усилить:

Теорема 1.6 (Критерий дифференцируемости)**.** *Пусть $f : E \to \mathbb{R}$ — выпуклая функция, определенная на множестве E в конечномерном евклидовом пространстве V, и пусть $x_0 \in \operatorname{int}(E)$. Тогда f дифференцируема в точке x_0, если и только если субдифференциал $\partial f(x_0)$ содержит ровно один элемент. В этом случае $\partial f(x_0) = \{\nabla f(x_0)\}$.*

Приведенный критерий оказывается чрезвычайно полезным для установления дифференцируемости функций, полученных с помощью выпуклых недифференцируемых операций, таких как максимизация. Точно так же, как и при доказательстве существования предела иногда оказывается полезно отдельно вычислить нижний и верхние пределы (которые существуют всегда) и показать, что они равны друг другу, так же и при доказательстве дифференцируемости выпуклой функции иногда оказывается проще напрямую вычислить субдифференциал и показать, что он состоит ровно из одного элемента (особенно, если такое вычисление удается провести с помощью правил субдифференциального исчисления).

Несмотря на то, что формально субградиент может быть определен для любой функции, а не только выпуклой, в действительности, это впечатление обманчиво. Далее будет видно, что практически все наиболее интересные результаты, связанные с субдифференциалами, требуют выпуклости. Во многом ответ на вопрос о том, почему субградиенты и выпуклость тесно связаны, дает следующее простое утверждение:

Утверждение 1.7 (Из субдифференцируемости следует выпуклость)**.** *Пусть $f : E \to \mathbb{R}$ — функция, определенная на выпуклом множестве E в евклидовом пространстве. Если f субдифференцируемая, то f выпуклая.*

Доказательство. Пусть $x, y \in E$, $0 < \alpha < 1$. В силу выпуклости множества, $\alpha x + (1-\alpha)y \in E$. Пусть $s \in \partial f(\alpha x + (1-\alpha)y)$ (в силу субдифференцируемости, такое s существует). Тогда, из определения субградиента, получаем

$$f(x) \geq f(\alpha x + (1-\alpha)y) + (1-\alpha)\langle s, x-y\rangle,$$
$$f(y) \geq f(\alpha x + (1-\alpha)y) - \alpha\langle s, x-y\rangle.$$

Умножая первое неравенство на α, второе на $(1-\alpha)$ и складывая, получаем $\alpha f(x) + (1-\alpha)f(y) \geq f(\alpha x + (1-\alpha)y)$. $\square$

Одно из наиболее полезных применений субдифференциалов заключается в формулировании условий оптимальности. Хорошо известно, что если дифференцируемая функция $f : E \to \mathbb{R}$ имеет минимум в точке $x_0 \in \mathrm{int}(E)$, то необходимо $\nabla f(x_0) = 0$; если при этом функция f выпуклая, то условие равенства градиента нулю является не только необходимым, но и достаточным. Аналогичное утверждение можно сформулировать и на языке субдифференциалов:

Упражнение 1.8 (Критерий минимума)**.** Пусть $f : E \to \mathbb{R}$ — функция, определенная на множестве E в евклидовом пространстве, и пусть $x_0 \in E$. Покажите, что точка x_0 является минимумом функции f, если и только если $0 \in \partial f(x_0)$.

Таким образом, любая функция f в точке минимума x_0 гарантированно имеет хотя бы один субградиент. Обратите внимание, что критерий минимума говорит не о том, что $\partial f(x_0) = \{0\}$, т. е. субдифференциал состоит только из одного элемента, равного 0, а о том, что субдифференциал $\partial f(x_0)$ *содержит* элемент 0; при этом $\partial f(x_0)$ может содержать и другие ненулевые субградиенты.

Пример 1.9. Пусть $f : \mathbb{R} \to \mathbb{R}$ — функция $f(x) := |x|$. Поскольку $0 \in \partial f(0) = [-1, 1]$, то 0 является минимумом f. Любая другая точка $x_0 \neq 0$ не является минимумом f, поскольку $0 \notin \partial f(x_0) = \{\mathrm{sign}\{x_0\}\}$.

Согласно утверждению 1.7, из субдифференцируемости следует выпуклость. Естественно задаться вопросом, а верно ли обратное: любая ли выпуклая функция является субдифференцируемой? Оказывается, что, вообще говоря, ответ на этот вопрос отрицательный.

Пример 1.10. Пусть $f : [0, +\infty) \to \mathbb{R}$ — функция $f(x) := -\sqrt{x}$. Тогда $\partial f(0) = \emptyset$.

Доказательство. Предположим, чтобы прийти к противоречию, что $s \in \partial f(0)$ для некоторого $s \in \mathbb{R}$. Тогда, по определению, $sx \leq -\sqrt{x}$ для всех $x \geq 0$. Отсюда $s \leq -\frac{1}{\sqrt{x}}$ для всех $x > 0$. Переходя к пределу по $x \to 0; x > 0$, получаем $s \leq -\infty$, что невозможно. $\square$

Итак, не всякая выпуклая функция является субдифференцируемой. Тем не менее, оказывается, что отсутствие субградиентов у выпуклой функции — это довольно экзотическая ситуация, которая может произойти только в граничных точках области определения. Один из основных результатов теории субдифференциалов состоит в том, что в любой *внутренней* точке области определения выпуклая функция гарантированно имеет субградиенты, причем субдифференциал является не просто выпуклым и замкнутым множеством, но также *ограниченным* (и, тем самым, компактным, согласно теореме Гейне–Бореля).

Теорема 1.11 (Субдифференциал во внутренних точках области определения)**.** *Пусть $f : E \to \mathbb{R}$ — выпуклая функция, определенная на множестве E в конечномерном евклидовом пространстве V, и пусть $x_0 \in \mathrm{int}(E)$. Тогда множество $\partial f(x_0)$ не пусто и является выпуклым компактом.*

В заключение этого раздела получим важную формулу для субдифференциала нормы и обсудим связь субдифференциала с двойственностью.

Пример 1.12 (Субдифференциал нормы)**.** Пусть V — конечномерное евклидово пространство, $x_0 \in V$. Пусть $\|\|$ — произвольная норма в V (не обязательно индуцированная скалярным произведением), и пусть $\|\|_*$ — соответствующая сопряженная норма. Тогда

$$\partial\|\|(x_0) = \{s \in V : \|s\|_* \leq 1;\ \langle s, x_0\rangle = \|x_0\|\} = \begin{cases} \overline{B}_{\|\|_*}(0,1), & \text{если } x_0 = 0, \\ \{s \in V : \|s\|_* = 1;\ \langle s, x_0\rangle = \|x_0\|\}, & \text{иначе,} \end{cases} \tag{1.1}$$

где $\overline{B}_{\|\|_*}(0,1) := \{s \in V : \|s\|_* \le 1\}$ — замкнутый единичный шар с центром в нуле относительно сопряженной нормы. Другими словами, вектор $s \in V$, $\|s\|_* = 1$, является субградиентом нормы $\|\|$ в точке $x_0 \neq 0$, если и только если *неравенство Гельдера* $\langle s, x_0 \rangle \le \|x_0\|$ переходит в равенство.

Доказательство. Пусть $s \in V$. По определению, $s \in \partial\|\|(x_0)$, если и только если

$$\langle s, x \rangle - \|x\| \le \langle s, x_0 \rangle - \|x_0\|,$$

для всех $x \in V$, или, эквивалентно

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

По определению супремума, последнее равносильно (почему?)

$$\sup_{x \in V}\{\langle s, x \rangle - \|x\|\} = \langle s, x_0 \rangle - \|x_0\|. \tag{1.2}$$

Заметим, что выражение в левой части есть супремум из определения сопряженной функции Фенхеля для нормы, который, как известно (см. конспект по двойственности), равен

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

Таким образом, (1.2) эквивалентно

$$\|s\|_* \le 1 \qquad \text{и} \qquad \langle s, x_0 \rangle = \|x_0\|.$$

В итоге,

$$\partial\|\|(x_0) = \{s \in V : \|s\|_* \le 1;\ \langle s, x_0 \rangle = \|x_0\|\}.$$

Остается заметить, что для $x_0 \neq 0$ неравенство $\|s\|_* \le 1$ обязано переходить в равенство, поскольку при $\|s\|_* < 1$, из неравенства Гельдера следует $\langle s, x_0 \rangle \le \|s\|_* \|x_0\| < \|x_0\|$. □

Сопряженная норма в примере 1.12 возникла не случайно. Оказывается, что полностью аналогичным образом для произвольной функции f (а не только для нормы) ее субдифференциал можно описать в терминах двойственного объекта — сопряженной функции Фенхеля:

Упражнение 1.13 (Характеризация субдифференицала через сопряженную функцию)**.** Пусть $f : E \to \mathbb{R}$ — функция, определенная на множестве E в евклидовом пространстве. Пусть $x_0 \in E$, и пусть $f^* : E_* \to \mathbb{R}$ — сопряженная функция. Покажите, что

$$\partial f(x_0) = \{s \in E_* : \langle s, x_0 \rangle = f^*(s) + f(x_0)\}, \tag{1.3}$$

Другими словами, вектор $s \in E_*$ является субградиентом функции f в точке x_0, если и только если *неравенство Фенхеля–Юнга* $\langle s, x_0 \rangle \le f^*(s) + f(x_0)$ переходит в равенство.

В случае $f = \|\|$, получаем $f^* = \delta_{\overline{B}_{\|\|_*}(0,1)}$, т. е. сопряженная функция равна индикаторной функции шара $\overline{B}_{\|\|_*}(0,1)$ (см. конспект по двойственности), и формула (1.3) переходит в формулу (1.1).

Следствие 1.14 (Критерии равенства в неравенстве Фенхеля–Юнга)**.** *Пусть $f : E \to \mathbb{R}$ — выпуклая замкнутая функция, $f^* : E_* \to \mathbb{R}$ — сопряженная функция, и пусть $x \in E$, $s \in E_*$. Тогда следующие утверждение эквивалентны:*

(a) $\langle s, x \rangle = f^*(s) + f(x)$.

(b) $s \in \partial f(x)$.

(c) $x \in \partial f^*(s)$.

Доказательство. Согласно упражнению 1.13, условие $\langle s, x \rangle = f^*(s) + f(x)$ равносильно вложению $s \in \partial f(x)$. С другой стороны, поскольку f выпуклая и замкнутая, то, по теореме Фенхеля–Моро (см. конспект по двойственности), $f^{**} = f$. Применяя упражнение 1.13 к функции f^*, получаем, что равенство $\langle s, x \rangle = f^*(s) + f(x)$ эквивалентно вложению $x \in \partial f^*(s)$. □

2 Субдифференциальное исчисление

Неотъемлемой частью классического дифференциального исчисления являются правила, позволяющие вычислить производную суммы, произведения, композиции и пр. двух функций через производные операндов. При отсутствии такого рода правил полезность любой теории обобщенного дифференцирования представляется сомнительной. Во многом рассматриваемая теория субдифференциалов приобрела широкое распространение благодаря тому, что для важного (хоть и частного) класса *выпуклых функций* каждую стандартную операцию, сохраняющую выпуклость (сумма, композиция, взятие максимума и др.), удается снабдить соответствующим правилом пересчета субдифференциалов. (Для невыпуклых функций и операций существуют разнообразные обобщения классического понятия субдифференциала. Наиболее известным из таких обобщений является *субдифференциал Кларка*. Однако исчисление для него существенно более слабое, чем классическое выпуклое субдифференциальное исчисление и выходит за рамки данного текста.)

Отметим, что доказательство правил субдифференциального исчисления (даже таких простых, как сумма или композиция) является существенно менее тривиальным, чем доказательство аналогичных правил классического дифференциального исчисления. Причиной этому является то, что в обычном дифференциальном исчислении градиент является вектором, в то время как субдифференциал является целым множеством, и операции над субдифференциалами соответствуют операциям над множествами.

Начнем с элементарного правила, которое сразу же следует из определения.

Упражнение 2.1 (Умножение на положительный скаляр)**.** Пусть $f : E \to \mathbb{R}$ — функция, определенная на множестве E в евклидовом пространстве, $c > 0$, и пусть $x_0 \in E$. Покажите, что

$$\partial(cf)(x_0) = c\partial f(x_0).$$

Одним из самых первых результатов, появившихся в теории субдифференциального исчисления, является *правило суммы*, которое в литературе известно под названием теоремы *Моро–Рокафеллара.*

Теорема 2.2 (Теорема Моро–Рокафеллара)**.** *Пусть $f : E \to \mathbb{R}$ и $g : G \to \mathbb{R}$ — функции, определенные на множествах E и G в одном и том же евклидовом пространстве, и пусть $x_0 \in E \cap G$. Тогда*

$$\partial(f + g)(x_0) \supseteq \partial f(x_0) + \partial g(x_0). \tag{2.1}$$

Если дополнительно функции f и g выпуклые, и при этом $E \cap \operatorname{int}(G) \neq \emptyset$, то

$$\partial(f + g)(x_0) = \partial f(x_0) + \partial g(x_0).$$

Еще раз отметим, что, несмотря на свою естественность, правило суммы является крайне нетривиальным: используя только лишь определение субдифференциала через неравенства, удается показать лишь вложение (2.1); для доказательства обратного вложения приходится задействовать теорему об отделимости выпуклых множеств.

Следующее упражнение показывает, что условие $E \cap \operatorname{int}(G) \neq \emptyset$, вообще говоря, убрать нельзя:

Упражнение 2.3. Пусть $f : [0, +\infty) \to \mathbb{R}$ — функция $f(x) := -\sqrt{x}$, и пусть $g : (-\infty, 0] \to \mathbb{R}$ — функция $g(x) := -\sqrt{-x}$. Покажите, что $\partial(f + g)(0) = \mathbb{R}$, в то время как $\partial f(0) = \partial g(0) = \emptyset$.

Используя индукцию и свойство о том, что внутренность конечного пересечения множеств равна пересечению внутренностей, получаем обобщение теоремы Моро–Рокафеллара для конечным сумм:

Следствие 2.4 (Теорема Моро–Рокафеллара II)**.** *Пусть $m \geq 2$ — целое, $f_1 : E_1 \to \mathbb{R}, \ldots, f_m : E_m \to \mathbb{R}$ — функции, определенные на множествах $E_1, \ldots, E_m$ в одном и том же евклидовом пространстве, и пусть $x_0 \in \cap_{i=1}^m E_i$. Тогда*

$$\partial \left(\sum_{i=1}^m f_i \right)(x_0) \supseteq \sum_{i=1}^m \partial f_i(x_0).$$

Если дополнительно функции $f_1, \dots, f_m$ выпуклые, и при этом $\cap_{i=1}^{m-1} \operatorname{int}(E_i) \cap E_m \neq \emptyset$, то

$$\partial \left(\sum_{i=1}^{m} f_i \right) (x_0) = \sum_{i=1}^{m} \partial f_i(x_0).$$

Еще одним естественным обобщением соответствующего правила из классического дифференциального исчисления является правило композиции с аффинным преобразованием. Его формулировка и доказательство полностью аналогичны теореме Моро–Рокафеллара.

Теорема 2.5 (Композиция с аффинным преобразованием). *Пусть V и W — евклидовы пространства, $T : V \to W$ — аффинное преобразование $T(x) := Ax + b$, где $A : V \to W$ — линейное преобразование, $b \in W$. Пусть $g : G \to \mathbb{R}$ — функция, определенная на множестве G в пространстве W, и пусть $x_0 \in T^{-1}(G)$. Тогда*

$$\partial(g \circ T)(x_0) \supseteq A^* \partial g(T(x_0)),$$

где $A^* : W \to V$ — сопряженный оператор для A. Если дополнительно функция g выпуклая, и при этом $T(V) \cap \operatorname{int}(G) \neq \emptyset$, то

$$\partial(g \circ T)(x_0) = A^* \partial g(T(x_0)).$$

Пример 2.6. Пусть $f : \mathbb{R}^n \to \mathbb{R}$ — функция $f(x) := |\langle a, x \rangle|$, где $a \in \mathbb{R}^n$. Пусть $A : \mathbb{R}^n \to \mathbb{R}$ — линейное преобразование $A(x) := \langle a, x \rangle$. В этом случае $A^* t = ta$ для всех $t \in \mathbb{R}$ (почему?). Поскольку $A(\mathbb{R}^n) \cap \operatorname{int}(\mathbb{R}^n) \neq \emptyset$ (например, это множество содержит ноль), то

$$\partial f(x) = \partial |{|}(\langle a, x \rangle) a = \begin{cases} [-1, 1] a, & \text{если } \langle a, x \rangle = 0, \\ \{\operatorname{sign}(\langle a, x \rangle) a\}, & \text{иначе} \end{cases} = \begin{cases} [-a, a], & \text{если } \langle a, x \rangle = 0, \\ \{\operatorname{sign}(\langle a, x \rangle) a\}, & \text{иначе.} \end{cases}$$

(Второе равенство следует из формулы для субдифференциала модуля, полученной в примере 1.5.)

Следующие два правила, позволяющие вычислять субдифференциал максимума, являются жемчужинами субдифференциального исчисления и не имеют аналогов в классическом дифференциальном исчислении. Напомним, что для множества C *выпуклой оболочкой* $\operatorname{Conv}(C)$ множества C называется пересечение всевозможных выпуклых множеств, содержащих C (или, эквивалентно, множество, образованное всевозможными выпуклыми комбинациями произвольного конечного числа точек из C).

Теорема 2.7 (Максимум конечного числа функций). *Пусть $f_1 : E_1 \to \mathbb{R}, \dots, f_m : E_m \to \mathbb{R}$ — функции, определенные на множествах $E_1, \dots, E_m$ в одном и том же конечномерном евклидовом пространстве, пусть $f : \cap_{i=1}^{m} E_i \to \mathbb{R}$ — функция*

$$f(x) := \max\{f_1(x), \dots, f_m(x)\},$$

и пусть $x_0 \in \cap_{i=1}^{m} E_i$. Тогда

$$\partial f(x_0) \supseteq \overline{\operatorname{Conv}} \left(\cup_{i \in I(x_0)} \partial f_i(x_0) \right),$$

где $I(x_0)$ — множество индексов, на которых достигается максимум в точке x_0 (гарантированно непустое):

$$I(x_0) := \{1 \leq i \leq m : f_i(x_0) = f(x_0)\}.$$

Если дополнительно функции $f_1, \dots, f_m$ выпуклые, и при этом $x_0 \in \operatorname{int}(\cap_{i=1}^{m} E_i)$ (или, эквивалентно, $x_0 \in \cap_{i=1}^{m} \operatorname{int}(E_i)$), то

$$\partial f(x_0) = \operatorname{Conv} \left(\cup_{i \in I(x_0)} \partial f_i(x_0) \right).$$

Замечание 2.8. В некоторых источниках теорема 2.7 называется *теоремой Дубовицкого–Милютина.*

Пример 2.9. Пусть $f : \mathbb{R} \to \mathbb{R}$ — функция $f(x) := |x| = \max\{-x, x\}$, и пусть $x_0 \in \mathbb{R}$. Применяя теорему 2.7 и утверждение 1.4, получаем

$$\partial f(x_0) = \begin{cases} \operatorname{Conv}(\{-1, 1\}), & \text{если } x_0 = 0, \\ \operatorname{Conv}(\{\operatorname{sign}(x_0)\}), & \text{если } x_0 \neq 0 \end{cases} = \begin{cases} [-1, 1], & \text{если } x_0 = 0, \\ \{\operatorname{sign}(x_0)\}, & \text{если } x_0 \neq 0, \end{cases}$$

что совпадает с результатом примера 1.5.

Для того, чтобы распространить предыдущее правило на бесконечное семейство функций, необходимы дополнительные топологические условия: компактность и полунепрерывность.

Теорема 2.10 (Максимум произвольного числа функций)**.** *Пусть V — конечномерное евклидово пространство, I — произвольное множество (не обязательно конечное и не обязательно счетное), и пусть для каждого $i \in I$ задано множество E_i в V и функция $f_i : E_i \to \mathbb{R}$. Пусть $f : E \to \mathbb{R}$ — функция*

$$f(x) := \sup_{i \in I} f_i(x),$$

определенная на множестве $E := \{x \in \cap_{i \in I} E_i : \sup_{i \in I} f_i(x) < +\infty\}$, и пусть $x_0 \in E$. Тогда

$$\partial f(x_0) \supseteq \overline{\operatorname{Conv}} \left(\cup_{i \in I(x_0)} \partial f_i(x_0) \right),$$

где $I(x_0)$ — множество индексов, на которых достигается супремум в точке x_0 (возможно, пустое):

$$I(x_0) := \{i \in I : f_i(x) = f(x)\}.$$

Если дополнительно функция f_i выпукла для каждого $i \in I$, множество I является компактом (в некотором топологическом пространстве), и при этом для каждого $x \in \cap_{i \in I} E_i$ функция $i \mapsto f_i(x)$ является полунепрерывной сверху на I, то $E = \cap_{i \in I} E_i$, множество $I(x_0)$ не пусто, и для $x_0 \in \operatorname{int}(E)$ имеет место

$$\partial f(x_0) = \operatorname{Conv} \left(\cup_{i \in I(x_0)} \partial f_i(x_0) \right).$$

Замечание 2.11. В некоторых источниках теорема 2.10 (или ее аналоги) называется *теоремой Данскина.*

Пример 2.12. Получим еще одно доказательство формулы субдифференциала нормы (см. пример 1.12) с помощью теоремы 2.10. Пусть V — конечномерное евклидово пространство, $\|\|$ — произвольная норма в V, и пусть $x_0 \in V$. Используя сопряженную норму $\|\|_*$ (и тот факт, что сопряжение является инволюцией), можно записать следующее вариационное представление:

$$\|x\| = \max_{\|s\|_* = 1} \langle s, x \rangle,$$

справедливое для любого $x \in V$. В этом случае индексное множество $I := \{s \in V : \|s\|_* = 1\}$ представляет собой единичную сферу и является компактом (в пространстве V). Поскольку $x_0 \in \operatorname{int}(V)$, и для каждого $x \in V$ функция $s \mapsto \langle s, x \rangle$ является линейной (а, значит, непрерывной, и, тем более, полунепрерывной сверху), то, согласно теореме 2.10, получаем

$$I(x_0) = \{s \in V : \|s\|_* = 1;\ \langle s, x_0 \rangle = \|x_0\|\},$$

и

$$\partial \|\|(x_0) = \operatorname{Conv}(\{s \in V : \|s\|_* = 1;\ \langle s, x_0 \rangle = \|x_0\|\}).$$

Здесь была использована формула $\partial \langle s, \rangle (x_0) = \{s\}$ (см. утверждение 1.4).

Если $x_0 = 0$, то условие $\langle s, x_0 \rangle = \|x_0\|$ выполняется бессодержательно, и

$$\partial \|\|(0) = \operatorname{Conv}(\{s \in V : \|s\|_* = 1\}) = \overline{B}_{\|\|_*}(0, 1),$$

(обоснуйте последнее равенство). Если же $x_0 \neq 0$, то множество

$$\{s \in V : \|s\|_* = 1;\ \langle s, x_0 \rangle = \|x_0\|\}$$

является выпуклым (почему?), и, значит, операцию взятия выпуклой оболочки можно убрать.

Таким образом,

$$\partial\|\|(x_0) = \begin{cases} \overline{B}_{\|\|_*}(0,1), & \text{если } x_0 = 0, \\ \{s \in V : \|s\|_* = 1;\ \langle s, x_0 \rangle = \|x_0\|\}, & \text{иначе.} \end{cases}$$