Л2 Условия оптимальности. Выпуклость и гладкость

Напомним, что ключевой задачей курса является (Л1.1). Начнем изучение с задачи без ограничений (безусловной задачи оптимизации):

$$\min_{x \in \mathbb{R}^d} f(x). \tag{Л2.1}$$

Формализуем понятия решения данной задачи.

Л2.1 Условия оптимальности

Определение Л2.1. Точка x^* называется *локальным минимумом* функции f на $\mathbb{R}^d$ (локальным решением задачи минимизации f на $\mathbb{R}^d$), если существует $r > 0$ такое, что для любого $y \in B_2^d(r, x^*) = \left\{ y \in \mathbb{R}^d \mid \|y - x^*\|_2 \leq r \right\}$ следует, что $f(x^*) \leq f(y)$.

Определение Л2.2. Точка x^* называется *глобальным минимумом* функции f на $\mathbb{R}^d$ (глобальным решением задачи минимизации f на $\mathbb{R}^d$), если для любого $x \in \mathbb{R}^d$ следует, что $f(x^*) \leq f(x)$.

Понятно, что глобальный минимум является одновременно и локальным. Попробуем понять, какие есть свойства локального минимума. В частности, следующая теорема приводит необходимое условие локального минимума безусловной задачи оптимизации (Л2.1).

Теорема Л2.1 (Теорема 1.2.1. из [29])**.** Пусть x^* — локальный минимум функции f на $\mathbb{R}^d$. Тогда если f дифференцируема, то $\nabla f(x^*) = 0$.

Доказательство. Пойдем от противного и предположим, что x^* — локальный минимум, но $\nabla f(x^*) \neq 0$. Разложим функцию f в ряд в окрестности локального минимума:

$$f(x) = f(x^*) + \langle \nabla f(x^*), x - x^* \rangle + o(\|x - x^*\|_2), \tag{Л2.2}$$

где $\lim_{x \to x^*} \frac{o(\|x - x^*\|_2)}{\|x - x^*\|_2} = 0$.

Рассмотрим $x_\lambda = x^* - \lambda \nabla f(x^*)$. Найдем $\lambda_1 > 0$ такое, что для любого $0 < \lambda \leq \lambda_1$ можно гарантировать, что $\|x_\lambda - x^*\|_2 \leq r$, т.е. x_λ попадает в нужную окрестность из определения локального минимума (Определение Л2.1). Понятно, что такое λ_1 можно найти в силу $r > 0$, а $\nabla f(x^*)$ конечно. Тогда для любого $0 < \lambda \leq \lambda_1$ справедливо

$$f(x_\lambda) \geq f(x^*).$$

При этом разложение в ряд (Л2.2) для точек x_λ имеет вид:

$$\begin{aligned} f(x_\lambda) &= f(x^*) + \langle \nabla f(x^*), x_\lambda - x^* \rangle + o(\|x_\lambda - x^*\|_2) \\ &= f(x^*) - \lambda \|\nabla f(x^*)\|_2^2 + o(\lambda \|\nabla f(x^*)\|_2). \end{aligned}$$

Сделаем еще одно ограничение на «малость» λ. А именно, найдем $\lambda_2 > 0$ такое, что для любого $0 < \lambda \leq \min\{\lambda_1, \lambda_2\}$ выполнено

$$|o(\lambda \|\nabla f(x^*)\|_2)| \leq \frac{\lambda}{2} \|\nabla f(x^*)\|_2^2.$$

Тогда для любого $\lambda > 0$ такого, что $\lambda \le \min\{\lambda_1, \lambda_2\}$, следует

$$f(x_\lambda) \le f(x^*) - \frac{\lambda}{2}\|\nabla f(x^*)\|_2^2.$$

Пришли к противоречию, что x^* — локальный минимум. А значит $\nabla f(x^*) = 0$. ■

Наша цель — находить глобальный минимум, а локальных хотелось бы наоборот избегать (зависит от конкретной задачи, но обычно цель именно такая). Как мы поняли в Параграфе Л1, без дополнительных предположений на задачу (Л1.1) в худшем случае полный равномерный перебор является оптимальным алгоритмом. Поэтому пора ввести новые понятия, которые помогут сузить класс изучаемых задач и построить оптимистичную теорию поиска глобального минимума.

Л2.2 Выпуклость

Первое понятие — это выпуклость целевой функции f в задаче (Л2.1).

Определение Л2.3. Пусть дана непрерывно дифференцируемая на $\mathbb{R}^d$ функция $f : \mathbb{R}^d \to \mathbb{R}$. Будем говорить, что она является *выпуклой*, если для любых $x, y \in \mathbb{R}^d$ выполнено

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

Наряду с выпуклостью также вводят еще одно, более сильное понятие.

Определение Л2.4. Пусть дана непрерывно дифференцируемая на $\mathbb{R}^d$ функция $f : \mathbb{R}^d \to \mathbb{R}$. Будем говорить, что она является *μ-сильно выпуклой* ($\mu > 0$), если для любых $x, y \in \mathbb{R}^d$ выполнено

$$f(y) \ge f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|x - y\|_2^2.$$

Из определений видно, что выпуклость — это сильная выпуклость с $\mu = 0$.
В Параграфе С4 также даны другие определения выпуклости и сильной выпуклости, которые не требуют дифференцируемости. В случае дифференцируемой функции данные выше определения эквивалентны определениям из Параграфа С4.
Физический смысл Определений Л2.3 и Л2.4 проиллюстрирован на Рисунке Л2.1: выпуклая функция в любой точке «подперта» снизу линейно аппроксимацией, а сильно выпуклая — квадратичной функцией.
Введя новые классы функций, попробуем понять, что теперь можно сказать про точки минимума/решения задач оптимизации с такими целевыми функциями.

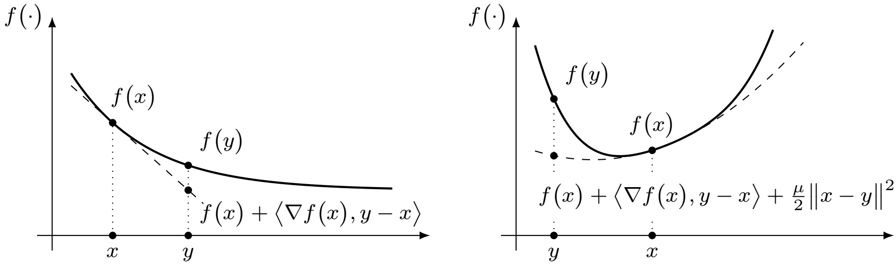
Теорема Л2.2. Пусть дана выпуклая на $\mathbb{R}^d$ функция $f : \mathbb{R}^d \to \mathbb{R}$. Если для некоторой точки $x^* \in \mathbb{R}^d$ верно, что $\nabla f(x^*) = 0$, то x^* — глобальный минимум f на всем $\mathbb{R}^d$.

Доказательство. Достаточно записать определение выпуклости:

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

■

Докажем несколько полезных фактов о минимумах выпуклых безусловных задач оптимизации.



(a) Выпуклость

(b) μ-сильная выпуклость

Рис. Л2.1: Иллюстрация понятий выпуклости и μ-сильной выпуклости.

Теорема Л2.3. Пусть дана выпуклая на $\mathbb{R}^d$ функция f. Тогда

- всякий локальный минимум f на $\mathbb{R}^d$ является и глобальным,
- если дополнительно f сильно выпуклая, то минимум существует и единственен.

Доказательство. Докажем последовательно пункты теоремы.

- Пусть x^* — локальный минимум. Согласно необходимому условию минимума функции (Теорема Л2.1)
$$\nabla f(x^*) = 0.$$
Пусть также x — произвольная точка из $\mathbb{R}^d$. Воспользуемся выпуклостью f:
$$f(x) \geq f(x^*) + \langle \nabla f(x^*), x - x^* \rangle = f(x^*).$$
Получаем, что $f(x^*) \leq f(x)$ для любого $x \in \mathbb{R}^d$. Значит, x^* — глобальный минимум.

- Покажем существование глобального минимума сильно выпуклой функции. Выберем произвольную точку $x^0 \in \mathbb{R}^d$ и положим $\beta = f(x^0)$. Если $\nabla f(x^0) = 0$, то x^0 — глобальный минимум (Теорема Л2.2). Далее рассуждаем в предположении, что $\nabla f(x^0) \neq 0$. Рассмотрим множество подуровня функции f:
$$L_\beta = \left\{ x \in \mathbb{R}^d \;\middle|\; f(x) \leq \beta \right\}.$$
По сильной выпуклости и неравенству Коши — Буняковского — Шварца (0.3) имеем:
$$\begin{aligned} f(x) &\geq f(x^0) + \langle \nabla f(x^0), x - x^0 \rangle + \frac{\mu}{2}\|x - x^0\|_2^2 \\ &\geq \beta - \|\nabla f(x^0)\|_2 \|x - x^0\|_2 + \frac{\mu}{2}\|x - x^0\|_2^2. \end{aligned}$$
Из условия $f(x) \leq \beta$ вытекает:
$$\frac{\mu}{2}\|x - x^0\|_2^2 - \|\nabla f(x^0)\|_2 \|x - x^0\|_2 \leq 0 \implies \|x - x^0\|_2 \leq \frac{2\|\nabla f(x^0)\|_2}{\mu}.$$

Значит

$$L_\beta \subseteq B_R(x^0) = \left\{ x \in \mathbb{R}^d \;\middle|\; \|x - x^0\|_2 \le R \right\}, \text{ где } R = \frac{2\|\nabla f(x^0)\|_2}{\mu}.$$

Из этого вложения получаем, что вне шара $B_R(x^0)$ значения функции f строго больше β:

$$f(x) > \beta, \ \forall x \notin B_R(x^0).$$

При этом, шар $B_R(x^0)$ замкнутый и ограниченный в конечномерном $\mathbb{R}^d$, следовательно, компактен. Тогда по теореме Вейерштрасса (см. Теорему 2 Параграфа 7 Главы 2 в [27]) f достигает на $B_R(x^0)$ своего минимума, который автоматически является глобальным (вне $B_R(x^0)$ значения f больше β).

Теперь покажем единственность. Пусть x^* — глобальный минимум, $x \in \mathbb{R}^d$. Поскольку f является μ-сильно выпуклой:

$$\begin{aligned} f(x) &\ge f^* + \langle \nabla f(x^*), x - x^* \rangle + \frac{\mu}{2}\|x - x^*\|_2^2 \\ &= f^* + \frac{\mu}{2}\|x - x^*\|_2^2. \end{aligned}$$

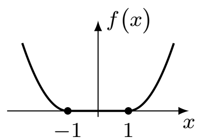
Слагаемое с градиентом занулилось, поскольку в точке оптимума $\nabla f(x^*) = 0$. Из свойств нормы получаем, что $f(x)$ достигает минимума только в точке x^*. Значит, если решение существует, то оно единственно.

■

Получается, что в случае выпуклой функции локальный минимум совпадает с глобальным. А значит $\nabla f(x^*) = 0$ является необходимым и достаточным условием.
В теореме не сказано про существование или единственность минимума выпуклой функции. Приведем два примера выпуклых функций, где эти свойства могут не выполняться.

Пример Л2.1. Покажем, что у выпуклой функции может быть больше одного минимума. Рассмотрим кусочно-заданную функцию f:

$$f(x) = \begin{cases} (x-1)^2, & x \in (1, +\infty) \\ 0, & x \in [-1, 1] \\ (x+1)^2, & x \in (-\infty, -1). \end{cases}$$



Она является выпуклой, однако, все точки отрезка $[-1, 1]$ доставляют минимум f, то есть, у f несчетное число точек минимума.

Пример Л2.2. Теперь приведем пример, когда выпуклая функция не имеет минимума на $\mathbb{R}$. Для этого подойдет линейная функция

$$f(x) = x.$$

Действительно, она выпукла, поскольку в каждой точке совпадает со своей линейной аппроксимацией, при этом функция не ограничена снизу, поэтому ни одна из точек $\mathbb{R}$ не является точкой минимума f.

Л2.3 Гладкость

Введем еще одно свойство, которое также пригодится для того, чтобы строить теорию сходимости оптимизационных методов.

Определение Л2.5. Пусть дана непрерывно дифференцируемая на $\mathbb{R}^d$ функция $f : \mathbb{R}^d \to \mathbb{R}$. Будем говорить, что данная функция имеет *L-Липшицев градиент* (говорить, что она является L-гладкой), если для любых $x, y \in \mathbb{R}^d$ выполнено

$$\big\|\nabla f(x) - \nabla f(y)\big\|_2 \le L\big\|x - y\big\|_2.$$

Определение L–гладкости можно задавать и в не евклидовой норме. Обобщение понятия на произвольную норму мы введем в Параграфе Л9.

Теорема Л2.4 (Лемма 1.2.3. из [29])**.** Пусть дана L–гладкая непрерывно дифференцируемая на $\mathbb{R}^d$ функция $f : \mathbb{R}^d \to \mathbb{R}$. Тогда для любых $x, y \in \mathbb{R}^d$ выполнено

$$\big|f(y) - f(x) - \big\langle \nabla f(x), y - x\big\rangle\big| \le \frac{L}{2}\big\|x - y\big\|_2^2.$$

Доказательство. Используем (см. страницу 84 из [28]) формулу Ньютона-Лейбница для криволинейного интеграла второго рода по кривой, заданной вектор функцией $r(\tau)$:

$$\int\limits_a^b \big\langle \nabla f\big(r(\tau)\big), \mathrm{d}r(\tau)\big\rangle = f\big(r(b)\big) - f\big(r(a)\big).$$

В нашем случае выберем кривую следующим образом $r(\tau) = x + \tau(y - x)$, где $\tau \in [0, 1]$. Тогда

$$\begin{aligned} f(y) - f(x) &= \int\limits_0^1 \big\langle \nabla f\big(x + \tau(y - x)\big), y - x\big\rangle \,\mathrm{d}\tau \\ &= \big\langle \nabla f(x), y - x\big\rangle + \int\limits_0^1 \big\langle \nabla f\big(x + \tau(y - x)\big) - \nabla f(x), y - x\big\rangle \,\mathrm{d}\tau\,. \end{aligned}$$

Переместив скалярное произведение влево и взяв модуль от обоих частей, получим:

$$\begin{aligned} \big|f(y) - f(x) - \big\langle \nabla f(x), y - x\big\rangle\big| &= \left|\int\limits_0^1 \big\langle \nabla f\big(x + \tau(y - x)\big) - \nabla f(x), y - x\big\rangle \,\mathrm{d}\tau\right| \\ &\le \int\limits_0^1 \big|\big\langle \nabla f\big(x + \tau(y - x)\big) - \nabla f(x), y - x\big\rangle\big| \,\mathrm{d}\tau\,. \end{aligned}$$

В последнем переходе мы использовали факт, что модуль суммы не превосходит сумму моделей слагаемых. Далее воспользуемся неравенством Коши — Буняковского — Шварца (0.3), а затем L-гладкостью (Определение Л2.5):

$$\begin{aligned}\left|f(y) - f(x) - \langle \nabla f(x), y - x\rangle\right| &\le \int\limits_0^1 \left\|\nabla f(x + \tau(y-x)) - \nabla f(x)\right\|_2 \left\|y - x\right\|_2 \mathrm{d}\tau \\ &\le L\left\|y - x\right\|_2^2 \int\limits_0^1 \tau \,\mathrm{d}\tau = \frac{L}{2}\left\|x - y\right\|_2^2.\end{aligned}$$

■

Отметим, что Теорема Л2.4 требует только L-гладкости функции f. Посмотрим, что можно получить, если дополнительно предположить еще и выпуклость функции f.

Теорема Л2.5 (Теорема 2.1.5. из [29]). Непрерывно дифференцируемая функция $f : \mathbb{R}^d \to \mathbb{R}$ является выпуклой и L-гладкой тогда и только тогда, когда для любых $x, y \in \mathbb{R}^d$ выполнены следующие неравенства:

$$0 \le f(y) - f(x) - \langle \nabla f(x), y - x\rangle \le \frac{L}{2}\left\|x - y\right\|_2^2, \tag{Л2.3}$$

$$\frac{1}{2L}\left\|\nabla f(x) - \nabla f(y)\right\|_2^2 \le f(y) - f(x) - \langle \nabla f(x), y - x\rangle, \tag{Л2.4}$$

$$\frac{1}{L}\left\|\nabla f(x) - \nabla f(y)\right\|_2^2 \le \langle \nabla f(x) - \nabla f(y), x - y\rangle. \tag{Л2.5}$$

Доказательство. Докажем следующее:

$$\begin{cases}\text{выпуклость}\\ \text{гладкость}\end{cases} \Longrightarrow (\text{Л2.3}) \Longrightarrow (\text{Л2.4}) \Longrightarrow (\text{Л2.5}) \Longrightarrow \begin{cases}\text{выпуклость}\\ \text{гладкость}\end{cases}$$

Выпуклость + гладкость $\Longrightarrow$ (Л2.3). Первое неравенство есть просто определение выпуклости, а второе является следствием из Теоремы Л2.4.
(Л2.3) $\Longrightarrow$ (Л2.4). Рассмотрим $\phi(y) = f(y) - \langle \nabla f(x), y\rangle$ для некоторого фиксированного $x \in \mathbb{R}^d$. Удостоверимся, что $\nabla \phi(y) = \nabla f(y) - \nabla f(x)$ является L-гладкой:

$$\begin{aligned}\left\|\nabla\phi(y_1) - \nabla\phi(y_2)\right\|_2 &= \left\|\nabla f(y_1) - \nabla f(x) - \nabla f(y_2) + \nabla f(x)\right\|_2 \\ &= \left\|\nabla f(y_1) - \nabla f(y_2)\right\|_2 \le L\left\|y_1 - y_2\right\|_2.\end{aligned}$$

Проверим также, что $\phi(y)$ выпуклая (по определению). Так как f выпуклая, для произвольных y_1 и y_2 имеем:

$$\begin{gathered} f(y_1) \ge f(y_2) + \langle \nabla f(y_2), y_1 - y_2\rangle \\ \Updownarrow \\ f(y_1) - \langle \nabla f(x), y_1\rangle \ge f(y_2) - \langle \nabla f(x), y_2\rangle + \langle \nabla f(y_2) - \nabla f(x), y_1 - y_2\rangle \\ \Updownarrow \\ \phi(y_1) \ge \phi(y_2) + \langle \nabla\phi(y_2), y_1 - y_2\rangle. \end{gathered}$$

А это и есть выпуклость $\phi(y)$. Заметим, что $\nabla\phi(x) = 0$, тогда в силу того, что функция ϕ выпуклая, то $y^* = x$ — точка глобального минимума (Теорему Л2.2). Откуда

$$\phi(x) = \phi(y^*) \leq \phi\left(y - \frac{1}{L}\nabla\phi(y)\right). \tag{Л2.6}$$

Теперь применим первый пункт теоремы для $f \to \phi$, $y \to y - \frac{1}{L}\nabla\phi(y)$, $x \to y$:

$$\phi\left(y - \frac{1}{L}\nabla\phi(y)\right) - \phi(y) - \left\langle \nabla\phi(y), -\frac{1}{L}\nabla\phi(y)\right\rangle \leq \frac{1}{2L}\left\|\nabla\phi(y)\right\|_2^2,$$

а значит после небольшой перестановки получим:

$$\phi\left(y - \frac{1}{L}\nabla\phi(y)\right) \leq \phi(y) - \frac{1}{2L}\left\|\nabla\phi(y)\right\|_2^2. \tag{Л2.7}$$

Осталось объединить (Л2.6) и (Л2.7):

$$\phi(x) \leq \phi(y) - \frac{1}{2L}\left\|\nabla\phi(y)\right\|_2^2,$$

и подставить ϕ:

$$f(x) - \langle\nabla f(x), x\rangle \leq f(y) - \langle\nabla f(x), y\rangle - \frac{1}{2L}\left\|\nabla f(x) - \nabla f(y)\right\|_2^2.$$

Из этого легко получить то, что и хотели доказать

$$\frac{1}{2L}\left\|\nabla f(x) - \nabla f(y)\right\|_2^2 \leq f(y) - f(x) - \langle\nabla f(x), y - x\rangle.$$

(Л2.4) $\Longrightarrow$ (Л2.5). Запишем два раза (Л2.4):

$$\begin{aligned}\frac{1}{2L}\left\|\nabla f(x) - \nabla f(y)\right\|_2^2 &\leq f(y) - f(x) - \langle\nabla f(x), y - x\rangle,\\ \frac{1}{2L}\left\|\nabla f(y) - \nabla f(x)\right\|_2^2 &\leq f(x) - f(y) - \langle\nabla f(y), x - y\rangle.\end{aligned}$$

Сложим эти два неравенства:

$$\frac{1}{L}\left\|\nabla f(x) - \nabla f(y)\right\|_2^2 \leq \langle\nabla f(x) - \nabla f(y), x - y\rangle.$$

(Л2.5) $\Longrightarrow$ выпуклость + гладкость. Из (Л2.5) имеем, что:

$$0 \leq \langle\nabla f(x) - \nabla f(y), x - y\rangle.$$

Снова применим (см. страницу 84 из [28]) формулу Ньютона-Лейбница для криволинейного интеграла второго рода по кривой, заданной вектор функцией $r(\tau)$:

$$\int\limits_a^b \left\langle\nabla f\big(r(\tau)\big), \mathrm{d}r(\tau)\right\rangle = f\big(r(b)\big) - f\big(r(a)\big).$$

В нашем случае выберем кривую следующим образом $r(\tau) = x + \tau(y - x)$, где $\tau \in [0, 1]$:

$$
\begin{aligned}
f\big(y\big) - f\big(x\big) &= \int\limits_0^1 \big\langle \nabla f\big(x + \tau(y-x)\big), y - x \big\rangle \,\mathrm{d}\tau \\
&= \big\langle \nabla f\big(x\big), y - x \big\rangle + \int\limits_0^1 \big\langle \nabla f\big(x + \tau(y-x)\big) - \nabla f\big(x\big), y - x \big\rangle \,\mathrm{d}\tau\,.
\end{aligned}
$$

Используя, что $\big\langle \nabla f\big(x\big) - \nabla f\big(x\big), x - y \big\rangle \geq 0$, получим:

$$
\begin{aligned}
f(y) - f(x) &= \big\langle \nabla f(x), y - x \big\rangle + \int\limits_0^1 \frac{1}{\tau} \cdot \big\langle \nabla f(x + \tau(y-x)) - \nabla f(x), \tau(y-x) \big\rangle \,\mathrm{d}\tau \\
&\geq \big\langle \nabla f(x), y - x \big\rangle.
\end{aligned}
$$

А это и есть эквивалентное определение выпуклости для непрерывно дифференцируемой функции. Также (Л2.5) вместе с неравенством Коши — Буняковского — Шварца (0.3) дает:

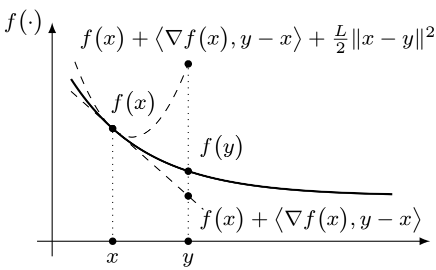
$$
\begin{aligned}
\frac{1}{L}\big\|\nabla f\big(x\big) - \nabla f\big(y\big)\big\|_2^2 &\leq \big\langle \nabla f\big(x\big) - \nabla f\big(y\big), x - y \big\rangle \\
&\leq \big\|\nabla f\big(x\big) - \nabla f\big(y\big)\big\|_2 \cdot \big\|x - y\big\|_2.
\end{aligned}
$$

Откуда

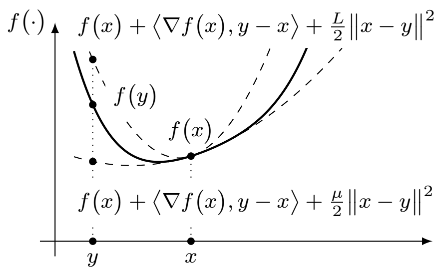
$$
\big\|\nabla f\big(x\big) - \nabla f\big(y\big)\big\|_2 \leq L\big\|x - y\big\|_2,
$$

а это и есть Определение Л2.5. ■

Первое неравенство из Теоремы Л2.5 может быть полезно в понимании физического смысла L-гладкости (Рисунок Л2.2): функция «подперта» снизу линейной аппроксимацией, а сверху квадратичной функцией. Похожая ситуация с L-гладкой и μ-сильно выпуклой функцией. Из Теоремы Л2.5 и Определения Л2.4 легко заметить, что $L \geq \mu$.



(a) Выпуклость и L-гладкость



(b) μ-сильная выпуклость и L-гладкость

Рис. Л2.2: Иллюстрация выпуклости, μ-сильной выпуклости и L-гладкости.

Упражнение 11.1. Рассмотрите квадратичную функцию:

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

где $b, x \in \mathbb{R}^d$, $A \in \mathbb{S}^d_+$. Покажите, что константы сильной выпуклости и гладкости можно оценить:

$$\mu \leq \frac{1}{2}\lambda_{\min}\big(A + A^\top\big), \quad L \geq \frac{1}{2}\lambda_{\max}\big(A + A^\top\big).$$

Упражнение 11.2. Рассмотрите логистическую функцию потерь:

$$f(x) = -\frac{1}{n}\sum_{i=1}^{n} l(g(x, a_i), b_i) + \frac{\lambda}{2}\|x\|_2^2,$$

где $g : \mathbb{R}^d \times \mathbb{R}^d \to \{0, 1\}$ — классификатор:

$$g(x, a_i) = \frac{1}{1 + \exp\big(-\langle x, a_i\rangle\big)},$$

а $l : \mathbb{R}^d \times \mathbb{R}^d \to [0, +\infty]$ — бинарная кросс-энтропия:

$$l(x, y) = y \ln x + (1 - y)\ln(1 - x),$$

$a_i \in \mathbb{R}^d$, $b_i \in \{0, 1\}$, $i = \overline{1, n}$ — данные, $x \in \mathbb{R}^d$ — целевая переменная, $\lambda \in \mathbb{R}_+$ — параметр регуляризации. Покажите, что константы сильной выпуклости и гладкости можно оценить:

$$\mu \leq \frac{1}{n}\lambda_{\min}\big(AA^\top\big) + \lambda, \quad L \geq \frac{1}{n}\lambda_{\max}\big(AA^\top\big) + \lambda.$$