План лекции

- Выпуклость и гладкость
- Субдифференцируемость
- Условия оптимальности

Гладкие задачи оптимизации

Определение 1

Дифференцируемая функция $f(x)$ заданная на множестве $Q \subseteq \mathbb{R}^d$ называется L-гладкой, если для любых двух точек $x, y \in Q$ выполняется следующее неравенство:

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x - y\|_2. \tag{1}$$

Теорема 1

Пусть функция $f : Q \to \mathbb{R}$ имеет Липшицев градиент с константой L относительно нормы $\|\cdot\|_2$, а множество Q является выпуклым. Тогда

$$|f(y) - f(x) - \langle \nabla f(x), y - x\rangle| \leq \frac{L}{2}\|y - x\|_2^2 \quad \forall x, y \in Q. \tag{2}$$

Доказательство

Рассмотрим две произвольные точки $x, y \in Q$. Так как множество Q является выпкулым, то для любого $\tau \in [0, 1]$ точка $x + \tau(y - x)$ принадлежит множеству Q (иными словами, множество Q вместе с любыми двумя точками содержит и отрезок, их соединяющий).
Рассмотрим функцию $\varphi(\tau) = f(x + \tau(y - x))$. Функция $\varphi(\tau)$ дифференцируема по τ на $[0, 1]$ и $\varphi'(\tau) = \langle \nabla f(x + \tau(y - x)), y - x\rangle$.
Следовательно,

$$f(y) - f(x) \quad = \quad \varphi(1) - \varphi(0) = \int_0^1 \langle \nabla f(x + \tau(y - x)), y - x\rangle d\tau$$

Доказательство

$$\int\limits_0^1 \langle \nabla f(x + \tau(y-x)), y - x\rangle d\tau$$

$$= \int\limits_0^1 \Big(\langle \nabla f(x + \tau(y-x)), y - x\rangle - \underbrace{\langle \nabla f(x), y - x\rangle + \langle \nabla f(x), y - x\rangle}_{\text{не зависит от } \tau} \Big) d\tau$$

$$= \langle \nabla f(x), y - x\rangle + \int\limits_0^1 \langle \nabla f(x + \tau(y-x)) - \nabla f(x), y - x\rangle d\tau$$

Доказательство

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

$$\overset{\text{К.-Б.}}{\leq} \int\limits_0^1 \|\nabla f(x + \tau(y-x)) - \nabla f(x)\|_2 \|y - x\|_2 d\tau$$

$$\overset{\text{Липш.}}{\leq} \int\limits_0^1 L\|x + \tau(y-x) - x\|_2 \|y - x\|_2 d\tau$$

$$= \int\limits_0^1 L\|y - x\|_2^2 \tau d\tau = \frac{L}{2}\|y - x\|_2^2$$

Гладкие задачи оптимизации

Доказанная теорема даёт следующую геометрическую интерпретацию функций с Липшицевым градиентом: это такие функции, которые в каждой точке можно оценить сверху некоторым парабалоидом (если рассматривать график функции f как множество в $\mathbb{R}^{n+1}$), причём оценить на всём множестве Q. Другие интересные свойства функций с Липшицевым градиентом (и не только) можно прочитать в книге Ю. Е. Нестерова "Введение в выпуклую оптимизацию" (глава 2, §2.1.1).

Выпуклые гладкие задачи оптимизации

Теорема 2 (см. Теоремы 2.1.5 и 2.1.6 из книги Нестерова 2010 года)

Все приведённые ниже условия, выполняющиеся для всех $x, y \in \mathbb{R}^d$, эквиваленты тому, что функция f является выпуклой и L-гладкой на $\mathbb{R}^d$:

- $0 \le f(y) - f(x) - \langle \nabla f(x), y - x \rangle \le \frac{L}{2}\|x - y\|_2^2$
- $f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2 \le f(y)$
- $\frac{1}{L}\|\nabla f(x) - \nabla f(y)\|_2^2 \le \langle \nabla f(x) - \nabla f(y), x - y \rangle$
- $0 \le \langle \nabla f(x) - \nabla f(y), x - y \rangle \le L\|x - y\|_2^2.$

Если дополнительно f является дважды непрерывно дифференцируемой, то условия выше эквивалентны условию

$$0 \preceq \nabla^2 f(x) \preceq L\mathrm{I}_d \quad \forall\, x \in Q.$$

Доказанная теорема даёт следующую геометрическую интерпретацию функций с Липшицевым градиентом: это такие функции, которые в каждой точке можно оценить сверху некоторым парабалоидом (если рассматривать график функции f как множество в $\mathbb{R}^{n+1}$), причём оценить на всём множестве Q. Другие интересные свойства функций с Липшицевым градиентом (и не только) можно прочитать в книге Ю. Е. Нестерова "Введение в выпуклую оптимизацию" (глава 2, §2.1.1).

Сильно выпуклые функции

Определение 2

Функция $f(x)$ заданная на выпуклом множестве $Q \subseteq \mathbb{R}^d$ называется μ-сильно выпуклой, если для любых двух точек $x, y \in Q$ и для любого числа $\alpha \in [0, 1]$ выполняется следующее неравенство:

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) - \alpha(1-\alpha)\frac{\mu}{2}\|x-y\|_2^2. \quad (3)$$

Для сильно выпуклой функции график лежит выше некоторого параболоида.

Для дифференцируемых функций существует эквивалентное определение, через нижнюю квадратичную границу:

Определение 3

Пусть функция $f : Q \to \mathbb{R}$ дифференцируема на Q, а множество Q является выпуклым. Будем называть f сильно выпуклой с константой μ если

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|y - x\|_2^2 \quad \forall x, y \in Q. \tag{4}$$

Теорема 3

Пусть функция f дифференцируема на выпуклом множестве Q. Тогда следующие 2 условия эквивалентны:

1. $$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|y - x\|_2^2 \quad \forall x, y \in Q$$

2. $$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) - \frac{\mu}{2}\alpha(1-\alpha)\|x - y\|_2^2$$

$$\forall x, y \in Q, \alpha \in [0, 1].$$

Доказательство

Пусть верно (1). Покажем, что выполнено (2). Рассмотрим две произвольных точки $x, y \in Q$ и произвольное число $\alpha \in [0, 1]$. Покажем, что выполнено неравенство

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) - \frac{\mu}{2}\alpha(1-\alpha)\|x - y\|_2^2.$$

Доказательство

Пусть $x_\alpha = \alpha x + (1-\alpha)y$. Запишем условие

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

сначала для пары точек $y = x, x = x_\alpha$, а затем для пары точек $y = y, x = x_\alpha$:

$$\begin{aligned} f(x) &\geq f(x_\alpha) + \langle \nabla f(x_\alpha), (1-\alpha)(x-y)\rangle + \frac{\mu}{2}\|(1-\alpha)(x-y)\|_2^2 \\ f(y) &\geq f(x_\alpha) + \langle \nabla f(x_\alpha), \alpha(y-x)\rangle + \frac{\mu}{2}\|\alpha(x-y)\|_2^2. \end{aligned}$$

Домножим первое неравенство на α, второе — на $1-\alpha$ и сложим полученные неравенства.

Доказательство

В итоге слагаемые со скалярным произведением сократятся, т.к. они отличаются только знаком, и получим неравенство

$$\alpha f(x) + (1-\alpha) f(y) \geq$$

$$\underbrace{(\alpha + 1 - \alpha)}_{=1} f(\alpha x + (1-\alpha)y) + \frac{\mu}{2}\underbrace{\left(\alpha(1-\alpha)^2 + \alpha^2(1-\alpha)\right)}_{=\alpha(1-\alpha)}\|x - y\|_2^2,$$

что и требовалось доказать.

Доказательство

Теперь предположим, что выполнено (2), и покажем, что выполняется (1). Действительно,

$$
\begin{aligned}
f(y) &\geq \frac{1}{1-\alpha}\left(f(x_\alpha) - \alpha f(x)\right) + \frac{\mu}{2}\alpha\|x-y\|_2^2 \\
&= f(x) + \frac{1}{1-\alpha}\left(f(x_\alpha) - f(x)\right) + \frac{\mu}{2}\alpha\|x-y\|_2^2 \\
&= f(x) + \frac{f(x+(1-\alpha)(y-x)) - f(x)}{1-\alpha} + \frac{\mu}{2}\alpha\|x-y\|_2^2.
\end{aligned}
$$

Переходя к пределу при $\alpha \to 1-0$ (слева), получаем

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

Теорема 4

Если функция $f : \mathbb{R}^d \to \mathbb{R}$ удовлетворяет условию

$$
f(y) \geq f(x) + \langle \nabla f(x), y-x\rangle + \frac{\mu}{2}\|y-x\|_2^2 \quad \forall x, y \in \mathbb{R}^d,
$$

то

$$
f(y) \leq f(x) + \langle \nabla f(x), y-x\rangle + \frac{1}{2\mu}\|\nabla f(x) - \nabla f(y)\|_2^2 \quad \forall x, y \in \mathbb{R}^d. \quad (5)
$$

Доказательство

Рассмотрим произвольные точки $x, y \in \mathbb{R}^d$ и рассмотрим функцию $\varphi(y) = f(y) - \langle \nabla f(x), y \rangle$ для заданного фиксированного x. Нетрудно проверить, что данная функция удовлетворяет условию

$$\varphi(y) \geq \varphi(x) + \langle \nabla \varphi(x), y - x \rangle + \frac{\mu}{2}\|y - x\|_2^2 \quad \forall x, y \in \mathbb{R}^d. \tag{6}$$

Кроме того, $\nabla \varphi(x) = \nabla f(x) - \nabla f(x) = 0$, а значит, из неравенства (6) для функции φ получаем, что для всех точек $y \in \mathbb{R}^n$

$$\varphi(y) \geq \varphi(x) + \frac{\mu}{2}\|y - x\|_2^2,$$

откуда следует, что x — точка минимума функции $\varphi(y)$ (причём единственная).

Доказательство

Тогда, пользуясь (6) для функции φ, получим

$$\varphi(x) = \min_{v \in \mathbb{R}^d} \varphi(v) \geq \min_{v \in \mathbb{R}^d} \left\{ \varphi(y) + \langle \nabla \varphi(y), v - y \rangle + \frac{\mu}{2}\|v - y\|_2^2 \right\}.$$

Теперь распишем правую часть чуть более подробно:

$$\begin{aligned}
&\min_{v \in \mathbb{R}^d} \left\{ \varphi(y) + \langle \nabla \varphi(y), v - y \rangle + \frac{\mu}{2}\|v - y\|_2^2 \right\} \\
&= \varphi(y) + \min_{v \in \mathbb{R}^d} \left\{ \langle \nabla \varphi(y), v - y \rangle + \frac{\mu}{2}\|v - y\|_2^2 \right\} + \frac{1}{2\mu}\|\nabla \varphi(y)\|_2^2 - \frac{1}{2\mu}\|\nabla \varphi(y)\|_2^2 \\
&= \varphi(y) + \min_{v \in \mathbb{R}^d} \left\{ \underbrace{\frac{1}{2\mu}\|\nabla \varphi(y)\|_2^2 + \langle \nabla \varphi(y), v - y \rangle + \frac{\mu}{2}\|v - y\|_2^2}_{\text{полный квадрат}} \right\} - \frac{1}{2\mu}\|\nabla \varphi(y)\|_2^2 =
\end{aligned}$$

Доказательство

$$= \varphi(y) + \underbrace{\min_{v\in\mathbb{R}^d}\left\{\left\|\sqrt{\frac{1}{2\mu}}\nabla\varphi(y) + \sqrt{\frac{\mu}{2}}(v-y)\right\|_2^2\right\}}_{=0,\ \text{при } v=y-\frac{1}{\mu}\nabla\varphi(y)} - \frac{1}{2\mu}\|\nabla\varphi(y)\|_2^2$$

$$= \varphi(y) - \frac{1}{2\mu}\|\nabla\varphi(y)\|_2^2,$$

откуда получаем, что $\varphi(x) \geq \varphi(y) - \frac{1}{2\mu}\|\nabla\varphi(y)\|_2^2$.
Если теперь подставить $\varphi(y) = f(y) - \langle\nabla f(x), y\rangle$, то получится нер-во

$$f(y) \leq f(x) + \langle\nabla f(x), y-x\rangle + \frac{1}{2\mu}\|\nabla f(x) - \nabla f(y)\|_2^2 \quad \forall x, y \in \mathbb{R}^d.$$

Условие Поляка-Лоясевича

В условиях Теоремы 4 выполняется следующее утверждение: если x^* — точка минимума функции f, то

$$\|\nabla f(x)\|_2^2 \geq 2\mu(f(x) - f(x^*)) \quad \forall x \in \mathbb{R}^n.$$

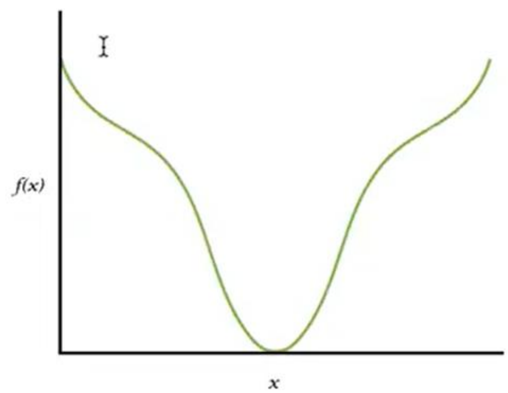
Доказательство

Чтобы доказать это неравенство, достаточно подставить в неравенство (5) точки $x = x^*$ и $y = x$ и учесть, что градиент в решении равен нулю.

На самом деле условие Поляка-Лоясевича может выполняться даже для невыпуклых функций.

Пример

$$x^2 + 3\sin^2(x)$$



Невыпуклая функция, удовлетворяющая условию Поляка-Лоясевича с константной $\mu = \frac{1}{32}$.

Доказательство выпуклости функции

- По определению.
- С помощью эпиграфа.
- Операции сохраняющие выпуклость.
- Дифференциальные критерии.

Выпуклые дифференцируемые функции

Для краткости будем считать, что выпуклая функция – это сильно выпуклая функция с $\mu = 0$.

Теорема 5 (см. Теоремы 2.1.2 и 2.1.9 из книги Нестерова 2010 года)

Пусть $f(x)$ дифференцируемая функция на выпуклом множестве $\mathrm{X} \subseteq \mathbb{R}^d$. Тогда $f(x)$ сильно выпукла на X с константой $\mu \geq 0$ тогда и только тогда, когда

$$f(x) - f(y) \geq \langle \nabla f(y), x - y \rangle + \frac{\mu}{2}\|x - y\|^2 \qquad \forall x, y \in \mathrm{X}$$

(если знак $>$, $\mu = 0$ и $\forall x, y \in \mathrm{X}$, $x \neq y$, тогда $f(x)$ строго выпукла)

Теорема 6 (см. Теоремы 2.1.3 и 2.1.9 из книги Нестерова 2010 года)

Пусть $f(x)$ непрерывно дифференцируемая функция на выпуклом множестве $\mathrm{X} \subseteq \mathbb{R}^d$. Тогда $f(x)$ сильно выпукла на X с константой $\mu \geq 0$ на X тогда и только тогда, когда

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|^2 \qquad \forall x, y \in \mathrm{X}$$

(если знак $>$, $\mu = 0$ и $\forall x, y \in \mathrm{X},\ x \neq y$, тогда $f(x)$ строго выпукла)

Теорема 7 (см. Теоремы 2.1.4 и 2.1.11 из книги Нестерова 2010 года)

Пусть $\mathrm{X} \subseteq \mathbb{R}^d$ - выпуклое множество, причем $\mathrm{int}\mathrm{X} \neq \varnothing$, $f(x)$ – дважды непрерывно дифференцируемая функция на X. Тогда $f(x)$ сильно выпукла на X с константой $\mu \geq 0$ тогда и только тогда, когда

$$h^\top \nabla^2 f(x) h \geq \mu \|h\|^2 \qquad \forall x \in \mathrm{X},\ h \in \mathbb{R}^d$$

$$\left(\nabla^2 f(x) \succeq \mu \mathrm{I} \quad \forall x \in \mathrm{X}\right)$$