1 Основы гладкой оптимизации

1.1 Обзорная лекция по методам оптимизации.

- Предмет оптимизации.
- Типы задачи оптимизации.
- Решение задачи оптимизации.
- Обзор градиентных методов.
- Обзор методов математической оптимизации для задач с ограничениями.
- Эвристические методы.
- Метод Нелдера-Мида (Nelder-Mead).
- Литература.

Задачи.

1. Доказать, что при рассмотрении задачи оптимизации для квадратичной функции без ограничения общности матрицу A можно считать симметричной.
2. Доказать, что минимум квадратичной функции $f(x) = \frac{1}{2}x^T Ax - b^T x + c$ существует тогда и только тогда, когда матрица A положительно определена.
3. Найти оценку скорости сходимости для квадратичной функции $f(x) = \frac{1}{2}x^T Ax - b^T x + c$.
4. Вычислить $\partial f(x)$ для $f(x) = |x|$. Показать, что для $f(x) = |x|$ при постоянном шаге субградиентный метод не сходится.
5. Вычислить производную по направлению $f'(0; y)$ для $f(x) = \|x\|$.
6. Найти $\nabla f(x)$ для $f(x) = \|x\|$ при $x \neq 0$.
7. Найти $\nabla f(x)$ для $f(x) = \|x_+\|^2$.
8. Найти $\nabla \|(Ax - b)_+\|$.

1.2 Общая постановка задачи оптимизации. Аналитическая и арифметическая сложность. Оценка сложности перебора. [N, §1.1]

- Общая постановка задачи оптимизации.
- Локальное и глобальное решение.
- Классификация задач оптимизации (линейная, гладкая, уловная, выпуклая...).
- Строго допустимая задача - уловие Слейтера.
- Общие примеры задач оптимизации.
- Понятие Оракула. Типы оракула: 0-го порядка, 1-го порядка, 2-го порядка.
- Аналитическая сложность и аримметическая сложность.
- Постановка задачи оптимизации в $B_n \subset \mathbb{R}^n$ для l_∞-липшицевой функции $f(x)$.
- Метод перебора (по p-сетке). Верхние оценки точности и сложности.

$$f(\overline{x}) - f^* \leq \frac{L}{2p}$$

- Понятие сопротивляющегося Оракула. Ничние оценки сложности.
- Асимптоическая оптимальность метода перебора на классе рассматриваемых задач.
- Неразрешимость общей задачи нелинейной оптимизации: оценка времени работы на примере.

Задачи.

1. ◦ Пусть $f(x) : B_n \to \mathbb{R}$ удовлетворяет уловию Липшица (l_∞) с константой L. Пусть f^* — оптимальное значение функции $f(x)$ в посталвенной задаче. Пусть $\overline{x} = \mathrm{argmin}\, f(x)$ на p-сетке. Доказать, что $f(\overline{x}) - f^* \leq \frac{L}{2p}$.
2. ◦ Доказать, что аналитическая сложность задачи 1. для метда перебора по p-сетке не превосходит $\left(\left[\frac{L}{2\varepsilon}\right] + 2\right)^n$.
3. ◦ Пусть $\varepsilon < L/2$. Тогда аналитическая сложность задачи 1. не менее $\left(\left[\frac{L}{2\varepsilon}\right]\right)^n$.
4. ◦ Оценить время работы метода перебора (количество лет) при следующих значениях параметров: $L = 2, n = 10, \varepsilon = 0,01$, производительность компьютера 10^9 арифм. опер. в сек.

1.3 Необходимые условия экстремума 1-го порядка. Необходимые условия экстремума 2-го порядка. Достаточные условия экстремума 2-го порядка. [N, §1.2.1]

- Идея релаксации.
- Постановка задачи безусловной гладкой минимизации.
- Линейные аппроксимации функции.
- Множество уровней $\mathcal{L}_f(a)$ функции $f(x)$.
- Локальное поведение функции $f(x)$ вдоль направления $s \in \mathbb{R}^n$, $||s|| = 1$:

$$\Delta(s) = \lim_{t \to 0} \frac{f(\overline{x} + ts) - f(\overline{x})}{t} = (f'(\overline{x}), s).$$

- Необходимые условия экстремума 1-го порядка.
- Понятие стационарной точки.
- Квадратичные аппроксимации функции. Гессиан.
- Положительно определенная матрица $A > 0$.
- Необходимые условия экстремума 2-го порядка.

$$\nabla f(x^*) = 0, \quad \nabla^2 f(x^*) \succeq 0.$$

- Достаточные условия экстремума 2-го порядка.

$$\nabla f(x^*) = 0, \quad \nabla^2 f(x^*) \succ 0.$$

Задачи.

1. Доказать, что функция $f : Q \to \mathbb{R}$ имеет глобальное решение

$$Q = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 + x_2 > 0\},$$
$$f(x_1, x_2) = x_1^2 + x_2 + \frac{1}{x_1 + x_2}.$$

2. С помощью условий первого и второго порядка найти экстремальные точки функции $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x_1, x_2) = x_1^3 + x_2^3 - 3x_1x_2.$$

3. Найти глобальные решения

$$Q = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1 > 0,\ x_2 > 0,\ x_3 > 0\},$$
$$f(x_1, x_2, x_3) = x_1 + \frac{x_2^2}{4x_1} + \frac{x_3^2}{x_2} + \frac{2}{x_3} \ \to\ \min.$$

4. Построить пример функции $f \in C^1(\mathbb{R}^n)$, имеющей ровно одну стационарную точку, являющуюся ее локальным, но не глобальным экстремумом. Существует ли такой пример при $n = 1$?

5. Привести пример функции $f \in C[a, b]$, не удовлетворяющей условию Липшица (0-го порядка) ни для какого L.

6. Доказать, что при любом значении параметра $\sigma > 1$ система уравнений относительно $x \in \mathbb{R}^2$ имеет ненулевое решение
$$\begin{cases} \sigma \cos x_1 \sin x_2 + x_1 e^{x_1^2 + x_2^2} = 0, \\ \sigma \sin x_1 \cos x_2 + x_2 e^{x_1^2 + x_2^2} = 0. \end{cases}$$

1.4 Классы дифференцируемых функций. [N, §1.2.2]

- Класс функций $C_L^{k,p}(\mathbb{R}^n)$.
- Неравенства для функций из $C_L^{2,1}(\mathbb{R}^n)$ и $C_L^{1,1}(\mathbb{R}^n)$.

$$f(x+y) = f(x) + \big(\nabla f(x), y\big) + \int_0^1 \big(\nabla f(x+\tau y) - \nabla f(x), y\big)\ \mathrm{d}\tau,$$

$$f(x+y) = f(x) + \big(\nabla f(x), y\big) + \int_0^1 \int_0^t \big(\nabla^2 f(x+\tau y)y, y\big)\ \mathrm{d}\tau\ \mathrm{d}t,$$

$$f(x) \in C_L^{k,p}(\mathbb{R}^n) \quad <=> \quad \|\nabla^p f(x) - \nabla^p f(y)\| \le L\|x-y\|,$$

$$f(x) \in C_L^{2,1}(\mathbb{R}^n) \quad <=> \quad \|\nabla^2 f(x)\| \le L,$$

$$f(x) \in C_L^{1,1}(\mathbb{R}^n) \quad => \quad |f(y) - f(x) - \big(\nabla f(x), y-x\big)| \le \frac{L}{2}\|y-x\|^2,$$

$$f(x) \in C_L^{2,2}(\mathbb{R}^n) \quad => \quad \|\nabla f(y) - \nabla f(x) - \nabla^2 f(x)(y-x)\| \le \frac{M}{2}\|y-x\|^2,$$

$$f(x) \in C_L^{2,2}(\mathbb{R}^n) \quad => \quad |f(y) - f(x) - \big(\nabla f(x), y-x\big) - \big(\nabla^2 f(x)(y-x), y-x\big)\| \le \frac{M}{6}\|y-x\|^3,$$

$$\nabla^2 f(x) - MrI_n \preceq \nabla^2 f(y) \preceq \nabla^2 f(x) + MrI_n.$$

Задачи.

1. Пусть $\phi : \mathbb{R}^n \to \mathbb{R}^m$, $f : \mathbb{R}^m \to \mathbb{R}$ дифференцируемые функции и $h(x) = f(\phi(x))$. Выписать $\mathrm{d}h(x)$ и $\mathrm{d}^2 h(x)$ через $\mathrm{d}\phi_j$ и $\mathrm{d}x$.

2. Пусть $f \in C^2(\mathbb{R}^n)$. Доказать, что

   а)
   $$f(y) = f(x) + \int_0^1 \big(f'(x+t(y-x)),\ (y-x)\big)\mathrm{d}t;$$

   б)
   $$f'(y) = f'(x) + \int_0^1 f''(x+t(y-x))(y-x)\mathrm{d}t;$$

3. Найти $f'(x)$ и $f''(x)$ для $f(x) = \alpha + (a,x) + \frac{1}{2}(Ax,x)$.

4. Пусть $f \in C^1(\mathbb{R}^n)$, $\phi(x) = f(Ax+b)$, $b \in \mathbb{R}^m$, матрица $A : \mathbb{R}^n \to \mathbb{R}^m$. Доказать, что для $x, y \in \mathbb{R}^n$:

   а) $(x, Ay) = (A^T x, y)$;

   б) $\phi'(x) = A^T f'(Ax+b)$.

5. Пусть $f \in C_M^{2,2}(\mathbb{R}^n)$. Тогда для любых $x, y \in \mathbb{R}^n$ имеет место неравенства

$$|f(y) - f(x) - \big(f'(x), y-x\big) - \frac{1}{2}\big(f''(x)(y-x), y-x\big)| \le \frac{M}{6}||y-x||^3.$$

2 Методы гладкой оптимизации

2.1 Градиентные методы. [N, §1.2.3-1.2.4]

- Градиентный метод $x_{k+1} = x_k - \alpha_k \nabla f(x_k)$.
- Варианты выбора длины шага h_k.

$$\alpha_k = h, \quad \alpha_k = \frac{h}{\sqrt{k+1}},$$
$$\alpha_k = \arg\min_{h \geq 0} f\big(x_k - h\nabla f(x_k)\big),$$
$$\begin{cases} a\big(\nabla f(x_k), x_k - x_{k+1}\big) \leq f(x_k) - f(x_{k+1}), \\ b\big(\nabla f(x_k), x_k - x_{k+1}\big) \geq f(x_k) - f(x_{k+1}). \end{cases}$$

- Оценки убывания целевой функции.

$$f(x) \in C_L^{1,1}(\mathbb{R}^n) \quad => \quad f(y) \leq f(x) - \frac{1}{2L}\|\nabla f(x)\|^2.$$

- Верхние оценка скорости сходимости для $f \in C_L^{1,1}(\mathbb{R}^n)$.

$$\min_{1 \leq k \leq N} \|\nabla f(x_k)\| \leq \frac{1}{N+1}\left(\frac{L}{\omega}\big(f(x_0) - f^*\big)\right)^{1/2} \leq \varepsilon, \quad => \quad N+1 \geq \frac{L}{\omega\varepsilon^2}\big(f(x_0) - f^*\big).$$

- Пример, когда градиентный метод не сходится к точкам локального минимума.
- Верхняя оценка скорости сходимости для $f \in C_M^{2,2}(\mathbb{R}^n)$: линейная скорость.

$$lI_n \preceq \nabla^2 f(x) \preceq LI_n \quad => \quad \|x_k - x^*\| \leq \frac{rr_0}{r - r_0}\left(1 - \frac{2l}{L+3l}\right).$$

- Метод Ньютона $x_{k+1} = x_k - [f''(x_k)]^{-1} f'(x_k)$.
- Вывод через квадратичную аппроксимацию.
- Пример сходимости и расходимости метода Ньютона.
- Демпфированный метод Ньютона с длиной шага h_k.
- Квадратичная сходимость метода Ньютона.

$$f \in C_M^{2,2}(\mathbb{R}^n),\ lI_n \preceq \nabla^2 f(x) \quad => \quad r_{k+1} \leq \frac{Mr_k^2}{2(l - Mr_k)}, \quad r_k = x_k - x^*.$$

- Классификация скорости сходимости метода.

Задачи.

1. Из начального приближения $x^0 = (-2, 1)$ сделать два шага метода скорейшего спуска для задачи безусловной оптимизации
$$f(x_1, x_2) = x_1^2 + x_1 x_2 + 2x_2^2 + x_1 \ \rightarrow \ \min .$$

2. Из начального приближения $x^0 = (1, -1)$ сделать шаг градиентного метода по правилу Голдштейна-Армийо (подобрать параметры α, β) для задачи безусловной оптимизации
$$f(x_1, x_2) = 2x_1^2 + x_1 x_2 + 3x_2^2 \ \rightarrow \ \min .$$

3. Из начального приближения $x^0 = 1$ сделать два шага метода Ньютона для уравнения $x^p = 0$, где $p \geq 2$ — целочисленный параметр. Проанализировать скорость сходимости, объяснить наблюдаемый эффект. Модифицировать метод Ньютона, введя параметр длины шага, равный p. Проанализировать скорость сходимости.

4. Из начального приближения $x^0 = (1, 1)$ сделать два шага метода Ньютона для задачи безусловной оптимизации
$$f(x_1, x_2) = x_1^2 + e^{x_2^2} \ \rightarrow \ \min .$$
Проанализировать сходимость. Показать, что в данном случае метод Ньютона сходится глобально.

5. Установить верхние границы аналитической сложности (количество итераций для ε-решения) для сублинейной, линейной и квадратичной скорости сходимости.