Математическое моделирование и методы оптимизации
Часть 2.

Тремба Андрей Александрович
Москва, к.ф.-м.н., с.н.с.
Институт проблем управления РАН

21 июня 2019 г., Иннополис

Вторая часть

Теория: «геометрия», формулы и методы

- Выпуклость
- Градиент и градиентные методы (безусловная минимизация)
- Безградиентные методы
- А также
  - Определение типа задачи
  - Фундаментальные идеи и подходы
  - Выбор подходящего метода решения
  - Немного об эффективности методов
  - Преобразование задач

Выпуклость и минимумы

Выпуклые и невыпуклые

Некоторые задачи существенно неразрешимы*:

$$2\sin x - x \to \min_x$$

Некоторые задачи решаются эффективно

$$x^2 - 2x \to \min_x$$

В чём разница?

Выпуклая оптимизация и глобальная оптимизация.

Выпуклые множества

Q выпукло

$$\Updownarrow$$

Для любых $x, z \in Q$

$$\alpha x + (1 - \alpha) z \in Q, \quad \alpha \in [0, 1]$$

Примеры выпуклых множеств

- шары в любых нормах $\{x : \|x\| \leq 1\}$
- аффинные множества $\{x \in \mathbb{R}^n : x^0 + Az,\ z \in \mathbb{R}^m\}$
- полупространства $\{x : (c, x) \leq c_0\}$
- пересечение выпуклых множеств (не объединение!)
- аффинные преобразования выпуклых множеств $\{Ax + b : x \in Q\}$

Примеры выпуклых множеств (продолжение)

- выпуклые конусы $C : \forall x^a, x^b \in C \to \lambda_1 x^a + \lambda_2 x^b \in C,\ \ \lambda_i \geq 0$
- выпуклая оболочка множества $conv(Q) = \{\lambda x + (1-\lambda)z, \quad x, z \in Q, \lambda \in [0,1]\}$
- выпуклая оболочка точек $conv(x^1, x^2, ..., x^m) = \{\sum_{i=1,m} \lambda_i x^i, \quad \lambda_i \geq 0,\ \sum_i \lambda_i = 1\}$

Свойства выпуклых множеств

- Отделимость:
  Если выпуклые множества $Q_1 \cap Q_2 = \emptyset$, то

$$\exists c, \alpha : (c, x) \leq \alpha, \forall x \in Q_1,\ (c, x) \geq \alpha, \forall x \in Q_2$$

- Существование опорной гиперплоскости к любой точке границы

$$\forall x \in \partial Q, \exists c, \alpha : (c, x) = \alpha, \quad (c, z) \leq \alpha, \quad z \in Q$$

Определение с помощью опорных гиперплоскостей:

$$Q = \bigcap_{c \in \mathbb{R}^n, \|c\|=1} \{x : (c, x) \leq S_Q(c)\}$$

Выпуклые функции

«Нулевое» определение: надграфик функции является выпуклым множеством.

$$\textbf{epi}\, f = \{(x,t) : f(x) \le t\}$$

Эквивалентные определения ($\forall \alpha \in [0,1],\ x, y \in \textbf{dom}\, f$):

1. $f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(y)$
2. $f(\frac{x+y}{2}) \le \frac{f(x)+f(y)}{2}$
3.
4.
5.

Определения минимума

Точка x^* называется локальным минимумом, если $\exists \varepsilon > 0$:

$$f(x) \geq f(x^*), \quad \forall x : \|x - x^*\| \leq \varepsilon$$

Определения минимума

Точка x^* называется локальным минимумом, если $\exists \varepsilon > 0$:

$$f(x) \geq f(x^*), \quad \forall x : \|x - x^*\| \leq \varepsilon$$

x^* называется глобальным минимумом, если

$$f(x) \geq f(x^*), \quad \forall x \in \mathbb{R}^n \quad (\text{или } \forall x \in Q)$$

Определения минимума

Точка x^* называется локальным минимумом, если $\exists \varepsilon > 0$:

$$f(x) \geq f(x^*), \quad \forall x : \|x - x^*\| \leq \varepsilon$$

x^* называется глобальным минимумом, если

$$f(x) \geq f(x^*), \quad \forall x \in \mathbb{R}^n \quad (\text{или } \forall x \in Q)$$

Для выпуклых функций каждый* локальный минимум является глобальным.

Область определения функции

Область определения $\mathbf{dom}\, f$ это множество, где функция «определена»

$$\mathbf{dom}\, f \subseteq \mathbb{R}^n$$

а) как неявные ограничения («естественная» область определения)

$$f(x) = x - \log x : \ \mathbf{dom}\, f = ?$$

б) как часть определения
$f(x) = x^2, x \in [1, 3) \cup [4, +\infty) \subset \mathbb{R}.$

Расширенная числовая прямая

Функция со значениями в $\mathbb{R} \cup \{+\infty\}$,

$$\overline{f} : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$$

$$\overline{f}(x) = \begin{cases} f(x), \; x \in \textbf{dom}\, f, \\ +\infty, \; \text{иначе} \end{cases}$$

$$\min_{x \in \textbf{dom}\, f} f(x) \Leftrightarrow \min \overline{f}(x)$$

$f(x)$ – выпуклая $\Leftrightarrow$ $\overline{f}(x)$ – выпуклая

Внимание, нюанс с постановкой задачи!

$$\mathbf{dom}\,\overline{f} = \,?$$

Внимание, нюанс с постановкой задачи!

$$\mathbf{dom}\ \overline{f} = \mathbb{R}^n$$

$$\min \overline{f}(x) \Leftrightarrow \min_{x \in \mathbf{dom}\ f} f(x)$$

Задача без ограничений может быть «скрыто» ограничена:

$$\min_x f(x), \quad \mathbf{dom}\ f \neq \mathbb{R}^n \Leftrightarrow \min_{x \in \mathbf{dom}\ f} f(x)$$

В безусловной оптимизацией подразумевается

$$\mathbf{dom}\ f = \mathbb{R}^n$$

Операции, сохраняющие выпуклость

- $\sum_i \lambda_i f_i(x)$
- $-f(x)$, если f — вогнутая
- $f(Ax + b)$
- $\max_i f_i(x)$
- $\sup_a f_a(x)$

Правила композиции (скалярные)

$$f(x) = h(g(x))$$

$$h : \mathbb{R} \to \mathbb{R}, \quad g : \mathbb{R}^n \to \mathbb{R}$$

$$\mathbf{dom}\, f = \{x \in \mathbf{dom}\,(g) : g(x) \in \mathbf{dom}\, h\}$$

выпуклая, если:

- h выпуклая, $\overline{h}$ неубывающая, g выпуклая
- h выпуклая, $\overline{h}$ невозрастающая, g вогнутая
- h выпуклая, g_i аффинная
- ...

Правила композиции (векторные)

$$f(x) = h(g(x)) = h(g_1(x), \ldots, g_m(x))$$

$$h : \mathbb{R}^m \to \mathbb{R}, \quad g : \mathbb{R}^n \to \mathbb{R}^m$$

$$\mathbf{dom}\, f = \{x \in \mathbf{dom}\,(g) : g(x) \in \mathbf{dom}\, h\}$$

выпуклая, если для каждой компоненты (i-й):

- h выпуклая, $\overline{h}$ неубывающая по i-й компоненте и g_i выпуклая
- h выпуклая, $\overline{h}$ невозрастающая по i-й компоненте и g_i вогнутая
- h выпуклая, g_i аффинная
- ...

Строгая выпуклость

Для множеств:

если $\widehat{x} \in \operatorname{bnd} Q$, то $\nexists\, x, y \in Q : \widehat{x} = \frac{x+y}{2}$

Для функций:

$$f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha)f(y), \quad \alpha \in (0,1)$$

Theorem

Минимум строго выпуклой функции единственный.

Сильная выпуклость

Константа $\mu > 0$.

Эквивалентные определения

($\forall \alpha \in [0, 1],\ x, z \in \textbf{dom} f$):

1. $f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha) f(y) - \mu \frac{\alpha(1-\alpha)}{2} \|x - y\|^2$
2. $f(\frac{x+y}{2}) \leq \frac{f(x)+f(y)}{2} - \frac{\mu}{8}\|x - y\|^2$
3.
4.
5.

Градации выпуклости

1. Выпуклая функция:
   локальный минимум* = глобальный.
2. Строго выпуклая функция:
   минимум* единственный.
3. Сильно выпуклая функция:
   - Если существует (*), то минимум единственный,
   - Существует и единственный, если $\mathbf{dom}\, f = \mathbb{R}^n$.

Особенности задачи

Даже с выпуклыми функциями:

- нет минимума (неограниченная обл. опр.),
- нет сходимости по x (неограниченная область определения),
- нет сходимости f (неограниченные значения),
- нет минимума (ограниченная область опр.),
- не единственный минимум (для не строго/сильно выпуклых)
- может быть разрывной (на границе).

Визуализация функций: графики и множества Лебега.

Множества Лебега и линии уровня

$$Q(c) = \{x : f(x) \leq c\}$$

- вложенность
- легко находить минимум
- выпуклые для выпуклых функций
- ограниченность

Линии уровня

$$\{x : f(x) = c\}$$

Методы!

Одномерная минимизация,
основная структура
и дифференцируемость.

Одномерная минимизация

$$f : \mathbb{R} \to \mathbb{R}$$

$$x \in [a, b]$$

- Сеточный метод

Одномерная минимизация

$$f : \mathbb{R} \to \mathbb{R}$$

$$x \in [a, b]$$

- Сеточный метод
- Метод золотого сечения
- Метод Фибоначчи

Сложность вычислений

1. Сложность вычисления функции
2. Число вычислений функций в одном шаге алгоритма
3. Число шагов алгоритма:

$$N - \varepsilon \text{ обозначения и } o(\cdot), O(\cdot) \text{ функции}$$

Сложность вычислений

1. Сложность вычисления функции
2. Число вычислений функций в одном шаге алгоритма
3. Число шагов алгоритма:

   $N - \varepsilon$ обозначения и $o(\cdot), O(\cdot)$ функции

   - $N(\varepsilon) : f(x^{N(\varepsilon)}) - f^* \leq \varepsilon$
   - $\varepsilon(N) \geq f(x^N) - f^*$

Скорость сходимости 1

$$\min_{x \in [a,b]} f(x)$$

Сеточный метод (1D) (для $|x^k - x^*| \leq \varepsilon$):

$$N(\varepsilon) = \frac{2(b-a)}{\varepsilon} + 1 = O(\frac{1}{\varepsilon}), \ \varepsilon(N) = \frac{2(b-a)}{N-1} = o(\frac{1}{N})$$

Скорость сходимости 1

$$\min_{x\in[a,b]} f(x)$$

Сеточный метод (1D) (для $|x^k - x^*| \le \varepsilon$):

$$N(\varepsilon)=\frac{2(b-a)}{\varepsilon}+1 = O(\frac{1}{\varepsilon}),\ \varepsilon(N)=\frac{2(b-a)}{N-1} = o(\frac{1}{N})$$

Сеточный метод (n-D) ($x \in \mathbb{R}^n,\ x_i \in [a,b]$):

$$N(\varepsilon) \approx \left(\frac{b-a}{\varepsilon}\right)^n, \quad \varepsilon(N) \approx \frac{b-a}{\sqrt[n]{N}}$$

Скорость сходимости 2

$$\min_{x \in [a,b]} f(x)$$

Метод золотого сечения (1D)
(для $|x^k - x^*| \leq \varepsilon$):

$$\varepsilon(N) = \frac{(b-a)}{2}\phi^{N-1} = o(1{,}618^N)$$

$$N(\varepsilon) = \log_\phi \left(\frac{b-a}{2\varepsilon}\right) + 1 = O(\log \frac{1}{\varepsilon})$$

Разные скорости сходимости

1. Сублинейная

$$f(x^{k+1}) - f^* \leq q_k \cdot (f(x^k) - f^*), \;\; q_k < 1, q_k \to 1$$

2. Линейная

$$f(x^{k+1}) - f^* \leq q \cdot (f(x^k) - f^*), \quad q < 1$$

3. Квадратичная*

$$f(x^{k+1}) - f^* \leq c \cdot (f(x^k) - f^*)^2$$

Общий метод, дифференцируемость.

Безусловная минимизация: общий метод

$$f(x) \to \min_x$$

(предполагается $\mathbf{dom}\, f = \mathbb{R}^n$)

«Выбор направления + выбор длины шага»

$$x^{k+1} = x^k + \gamma_k p^k$$

Координатный спуск

$$x^{k+1} = x^k + \gamma_k p^k$$

Оси как направления: $p^k = e^{i_k}$

- циклический перебор

$$i_k = (k \bmod n) + 1$$

Скалярное произведение

- Симметричность $(\cdot,\cdot) : V \times V \to \mathbb{R}$
$$(y, x) = (x, y)$$
- Линейность по первому* аргументу
$$(ax, y) = a(x, y)$$
$$(x + z, y) = (x, y) + (z, y)$$
- Положительная полуопределённость
$$(x, x) \geq 0$$
$$(x, x) = 0 \Leftrightarrow x = 0$$

Стандартное скалярное произведение,

$x, y \in \mathbb{R}^n$: $$(x, y) = x^T y = \sum_{i=\overline{1,n}} x_i y_i$$

Дифференцируемость

$$f : \mathbb{R}^n \to \mathbb{R}$$

Функция дифференцируема в точке x), если:

$$\exists p : f(x+z) = f(x) + (p, z) + o(\|z\|)$$

Дифференцируемость

$$f : \mathbb{R}^n \to \mathbb{R}$$

Функция дифференцируема в точке x), если:

$$\exists p : f(x+z) = f(x) + (p, z) + o(\|z\|)$$

$p(x)$ называется «производной» ($f'(x)$):

$$n = 1 : \;\; f(x+z) = f(x) + f'(x)z + o(|z|)$$

Дифференцируемость

$$f : \mathbb{R}^n \to \mathbb{R}$$

Функция дифференцируема в точке x), если:

$$\exists p : f(x+z) = f(x) + (p, z) + o(\|z\|)$$

$p(x)$ называется «производной» ($f'(x)$):

$$n = 1 : \; f(x+z) = f(x) + f'(x)z + o(|z|)$$

Для векторных функций $(n > 1)$ называется «градиент», обозначается $\nabla f(x)$.

Дифференцируемость

- Непрерывные функции: класс C^0
- Дифференцируемые функции
- Непрерывно дифференцируемые функции (гладкие функции): класс C^1
- Гладкие функции (бесконечно дифференцируемые): класс C^∞

Градиент и его свойства

- $f(x+z) = f(x) + (\nabla f(x), z) + o(\|z\|)$
- То же, что ряд Тейлора функции многих переменных

$$f(x+\Delta) = f(x) + \sum_{i=1,n} \frac{\partial f(x)}{\partial x_i}\Delta_i + \sum_{i=1,n} o(\Delta_i)$$

Градиент и его свойства

- $f(x+z) = f(x) + (\nabla f(x), z) + o(\|z\|)$
- То же, что ряд Тейлора функции многих переменных

$$f(x+\Delta) = f(x) + \sum_{i=1,n} \frac{\partial f(x)}{\partial x_i}\Delta_i + \sum_{i=1,n} o(\Delta_i)$$

Полезные следствия:

- $\lim_{\|z\|\to 0} \frac{f(x+z)-f(x)}{\|z\|} = (\nabla f(x), \frac{z}{\|z\|})$
- $f(y) = f(x) + (\nabla f(x), y-x) + o(\|y-x\|)$

Формула Ньютона-Лейбница

Если $f(x)$ дифференцируема, то верна точная формула

$$f(x+z) = f(x) + \int_0^1 \big(\nabla f(x + \tau z), z\big) d\tau,$$

и теорема о среднем значении:

$$f(x+z) = f(x) + \big(\nabla f(x + \widehat{\tau} z), z\big)$$

для некоторой $\widehat{\tau} \in (0, 1)$.

Визуализация градиентов

$$f(x+z) = f(x) + (\nabla f(x), z) + o(\|z\|)$$

$\nabla f(x)$ и касательная гиперплоскость

Исчисление градиентов

- «линейность»

$$\nabla_x \Big(af(x) \Big) = a\nabla f(x), \quad a \in \mathbb{R}$$

$$\nabla_x (f(x) + g(x)) = \nabla f(x) + \nabla g(x)$$

- дифференцируемость сложных функций (скалярных): $h : \mathbb{R} \to \mathbb{R}, \quad g : \mathbb{R} \to \mathbb{R},$

$$h(g(x))'_x = h'(g(x))g'(x)$$

и векторных $h : \mathbb{R} \to \mathbb{R}, \quad g : \mathbb{R}^n \to \mathbb{R},$

$$\nabla_x h(g(x)) = h'(g(x))\nabla_x g(x)$$

- $f(x) = g(Ax)$, то $\nabla_x f(x) = A^T \nabla_y g(Ax)$

Выпуклые функции (обновление)

Эквивалентные определения
($\forall \alpha \in [0, 1],\ x, y \in \textbf{dom}\, f$):

1. $f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$
2. $f(\frac{x+y}{2}) \leq \frac{f(x)+f(y)}{2}$
3. $f(y) \geq f(x) + (\nabla f(x), y - x)$
(дифференцируемые)
4. $(\nabla f(x) - \nabla f(y), x - y) \geq 0$
(дифференцируемые)
5.

Сильная выпуклость (обновление)

Константа $\mu > 0$.

Эквивалентные определения
$(\forall \alpha \in [0,1],\ x, z \in \mathbf{dom}\, f)$:

1. $f(\alpha x + (1-\alpha)y) \leq$
   $\alpha f(x) + (1-\alpha)f(y) - \mu\frac{\alpha(1-\alpha)}{2}\|x-y\|^2$
2. $f(\frac{x+y}{2}) \leq \frac{f(x)+f(y)}{2} - \frac{\mu}{8}\|x-y\|^2$
3. $f(y) \geq f(x) + (\nabla f(x), y - x) + \frac{\mu}{2}\|x-y\|^2$
4. $(\nabla f(x) - \nabla f(y), x - y) \geq \mu\|x-y\|^2$

Константа Липшица градиента

1.
2.
3. $f(y) \leq f(x) + (\nabla f(x), y - x) + \frac{L}{2}\|x - y\|^2$
4. $\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$
5.

Необходимое условие оптимальности (1го порядка, без ограничений)

Если x^* — точка экстремума, то

$$\nabla f(x^*) = 0$$

«x^* экстремальная точка» $\Rightarrow$ $\nabla f(x^*) = 0$

$$\Downarrow$$

«x^* точка минимума» $\Rightarrow$ $\nabla f(x^*) = 0$

Необходимые и достаточные условия оптимальности (1го порядка, без ограничений)

Для выпуклой функции без ограничений

$$\nabla f(x^*) = 0 \Leftrightarrow x^* \in \operatorname{Arg\,min} f$$

Безусловная минимизация!

1. Аналитическое решение или упрощение уравнения

$$\nabla f(x) = 0 \in \mathbb{R}^n$$

2. Практический критерий остановки

$$\|\nabla f(x^k)\| \to 0$$

Пример: аппроксимация и метод наименьших квадратов

Дано: наборы точек (x_j, y_j), $\quad j = 1, \ldots, N$

Найти: аппроксимирующую функцию $F(x)$.

Пример: аппроксимация и метод наименьших квадратов

Дано: наборы точек $(x_j, y_j), \quad j = 1, ..., N$
Найти: аппроксимирующую функцию $F(x)$.

Дополнительно: модельные функции
$\phi_i : \mathbb{R} \to \mathbb{R}, \quad i = 1, ..., n$ и $F(x) = \sum_{i=1}^{n} w_i \phi_i(x)$

Пример: аппроксимация и метод наименьших квадратов

Дано: наборы точек (x_j, y_j), $\quad j = 1, ..., N$
Найти: аппроксимирующую функцию $F(x)$.

Дополнительно: модельные функции
$\phi_i : \mathbb{R} \to \mathbb{R}$, $\quad i = 1, ..., n$ и $F(x) = \sum_{i=1}^{n} w_i \phi_i(x)$

Задача: $$\sum_{j=1}^{N} \left\| \sum_{i=1}^{n} w_i \phi_i(x_j) - y_j \right\|^2 \to \min_w$$

$$\min_w \|Aw - b\|^2$$

Возвращаясь к общему методу

Выбор направления (1 часть)

$$x^{k+1} = x^k + \gamma_k p^k, \; \gamma_k \geq 0$$

- Градиентный спуск (скорейшего спуска)

$$p^k = -\nabla f(x^k)$$

Выбор направления (1 часть)

$$x^{k+1} = x^k + \gamma_k p^k, \;\; \gamma_k \geq 0$$

- Градиентный спуск (скорейшего спуска)

$$p^k = -\nabla f(x^k)$$

- Допустимые направления

$$p^k : (p^k, \nabla f(x^k)) < 0$$

Выбор направления (1 часть)

$$x^{k+1} = x^k + \gamma_k p^k, \ \gamma_k \geq 0$$

- Градиентный спуск (скорейшего спуска)

$$p^k = -\nabla f(x^k)$$

- Допустимые направления

$$p^k : (p^k, \nabla f(x^k)) < 0$$

- Покоординатный спуск (?):

$$p^k = e_{i_k}$$

Покоординатный спуск $p^k = e^{i_k} \frac{-\nabla f(x)_{i_k}}{|\nabla f(x)_{i_k}|}$

- цикл $i_k = (k \mod n) + 1$

Покоординатный спуск $p^k = e^{i_k} \frac{-\nabla f(x)_{i_k}}{|\nabla f(x)_{i_k}|}$

- цикл $i_k = (k \mod n) + 1$
- случайный выбор (равномерно): $\mathsf{P}(i_k = j) = \frac{1}{n}$

Покоординатный спуск $p^k = e^{i_k}\frac{-\nabla f(x)_{i_k}}{|\nabla f(x)_{i_k}|}$

- цикл $i_k = (k \mod n) + 1$

- правило Гаусса-Саусвелла
  $i_k = \arg\max_i |\nabla f(x^k)_i|$

Покоординатный спуск $p^k = e^{i_k} \frac{-\nabla f(x)_{i_k}}{|\nabla f(x)_{i_k}|}$

- цикл $i_k = (k \mod n) + 1$

- правило Гаусса-Саусвелла-Липшица

$$i_k = \arg\max_i \frac{|\nabla f(x^k)_i|}{\sqrt{L_i}}$$

Покоординатный спуск $p^k = e^{i_k} \frac{-\nabla f(x)_{i_k}}{|\nabla f(x)_{i_k}|}$

- цикл $i_k = (k \mod n) + 1$

- случайный выбор (правило Нестерова)

$$\mathsf{P}(i_k = j) = \frac{L_j}{\sum_{m=1,n} L_j}$$

Покоординатный спуск $p^k = e^{i_k} \frac{-\nabla f(x)_{i_k}}{|\nabla f(x)_{i_k}|}$

- цикл $i_k = (k \mod n) + 1$
- случайный выбор (равномерно): $\mathsf{P}(i_k = j) = \frac{1}{n}$
- правило Гаусса-Саусвелла $i_k = \arg\max_i |\nabla f(x^k)_i|$
- правило Гаусса-Саусвелла-Липшица
$$i_k = \arg\max_i \frac{|\nabla f(x^k)_i|}{\sqrt{L_i}}$$
- случайный выбор (правило Нестерова)
$$\mathsf{P}(i_k = j) = \frac{L_j}{\sum_{m=1,n} L_j}$$

Выбор длины шага

$$h(\gamma) = f(x + \gamma p), \quad \gamma \in \mathbb{R} \text{ (or } \mathbb{R}_+)$$

- аналитический
- постоянный (learning rate)
- одномерная оптимизация $f(x^k + \alpha p^k)$
- правило Армихо
  $f(x+\gamma p) \le f(x)+\alpha\gamma(\nabla f(x), p), \quad \alpha \in (0, 1)$

Выбор длины шага

$$h(\gamma) = f(x + \gamma p), \quad \gamma \in \mathbb{R} \text{ (or } \mathbb{R}_+)$$

- аналитический
- постоянный (learning rate)
- одномерная оптимизация $f(x^k + \alpha p^k)$
- правило Армихо
  $f(x+\gamma p) \leq f(x)+\alpha\gamma(\nabla f(x), p), \quad \alpha \in (0, 1)$
- + правило Гольштейна
  $f(x + \gamma p) \geq f(x) + (1 - \alpha)\gamma(\nabla f(x), p),$
  $\alpha \in (0, 1/2)$

Выбор длины шага

$$h(\gamma) = f(x + \gamma p), \quad \gamma \in \mathbb{R} \text{ (or } \mathbb{R}_+)$$

- аналитический
- постоянный (learning rate)
- одномерная оптимизация $f(x^k + \alpha p^k)$
- правило Армихо + поиск с возвратом
  $f(x+\gamma p) \leq f(x)+\alpha\gamma(\nabla f(x), p), \quad \alpha \in (0, 1)$
- + правило Гольштейна
  $f(x + \gamma p) \geq f(x) + (1 - \alpha)\gamma(\nabla f(x), p)$,
  $\alpha \in (0, 1/2)$

Постоянный шаг. Теория

$$x^{k+1} = x^k - \gamma \nabla f(x^k)$$

Theorem

Пусть $f(x)$ *дифференцируема на* $\mathbb{R}^n$*, имеет липшицев градиент с константой* L *и ограничена снизу*

$$f(x) \geq f^* > -\infty.$$

Тогда если $0 < \gamma < \frac{2}{L}$*, то*
а) $\lim_{k\to\infty} \nabla f(x^k) = 0$*,*
б) $f(x^{k+1}) < f(x^k), \quad$ *if* $\quad \nabla f(x^k) \neq 0$

Доказательство и свойства

Формула Ньютона-Лейбница

- теорема верна для невыпуклых функций
- для сходимости по $f(x^k)$ либо x^k нужны дополнительные свойства,
- для выпуклых функций с ограниченным множеством Лебега - сходимость по $f(x^k)$,
- для сильно выпуклых функций - $x^k \to x^*$,
- «глобальная» сходимость,
- «медленная» сходимость.

Скорость сходимости градиентного метода

- Для выпуклой функции с липшицевым градиентом (L) - сублинейная:

$$f(x^k) - f^* \sim \frac{c}{k}$$

- Для сильно выпуклой (μ) функции с Липшицевым градиентом (L) - линейная:

$$\|x^{k+1} - x^*\| \leq \left(\frac{L - \mu}{L + \mu}\right) \|x^k - x^*\|$$

Промежуточные итоги

- «Направление + шаг»
- Дифференцируемость
- Условия оптимальности
- Выбор направления: градиентный метод
- Выбор длины шага

Методы нулевого порядка

Безградиентные методы

- Одномерная минимизация
- Аппроксимация градиента (численная, конструктивная)
- Симплексный метод (Нилдера-Мида) - не симплекс-метод!
- Метод имитации отжига

Метод Нилдера-Мида

Набор $n+1$ точек $\{x^i\}, i = 0, \dots, n$

Параметр «отражения» $\alpha = 1$

Параметр «расширения» $\gamma = 2$

Параметр «сжатия вершины» $\beta_1 = 0.5$

Параметр «сжатия симплекса» $\beta_2 = 0.5$

Один шаг метода Нилдера-Мида

1. Вычислить $f(x^i)$ и отсортировать:

$$\mathbf{f}_\ell = f(x^\ell) \le \ldots \le \mathbf{f}_g = f(x^g) \le \mathbf{f}_h = f(x^h),$$

Вычислить «центр масс» грани: $x^c = \frac{1}{n}\sum_{i \ne h} x^i$.

2. Отражение: $x^r = x^c + \alpha(x^c - x^h)$, и
$\mathbf{f}_r = f(x^r)$.

3. Принятие решения:

A. Если $\mathbf{f}_r < \mathbf{f}_\ell$, то выполнить Расширение
$x^e = x^c + \gamma(x^r - x^c)$, $\mathbf{f}_e = f(x^e)$.
Если $\mathbf{f}_e < \mathbf{f}_r$, то установить $x^h \leftarrow x^e$, иначе
$x^h \leftarrow x^r$, и перейти на Шаг 5.

Б. Если $\mathbf{f}_\ell < \mathbf{f}_r < \mathbf{f}_g$, то установить $x^h \leftarrow x^r$ и перейти на Шаг 5.

В. Если $\mathbf{f}_g < \mathbf{f}_r < \mathbf{f}_h$ попробовать Сжатие вершины $x^s = x^c + \beta_1(x^h - x^c),\ \mathbf{f}_s = f(x^s)$. Если $\mathbf{f}_s < \mathbf{f}_h$, установить $x^h \leftarrow x^s$, иначе* перейти на Шаг 5.

D. Если $\mathbf{f}_h < \mathbf{f}_r$, то на следующий шаг.

4. Сжатие симплекса $x^i = x^\ell + \beta_2(x^i - x^\ell),\ i \neq \ell$, и перейти на Шаг 1.

5. Проверить условия останова.

Случайный поиск и имитация отжига

- случайное направление и одномерный поиск
- имитация отжига
  1. из x^k получить $\to x^+$ (случайную близкую точку)
  2. если $f(x^+) < f(x^k)$, then $x^{k+1} = x^+$, иначе
  $$x^{k+1} = \begin{cases} x^k, & \text{с вероятностью } p_k, \\ x^+, & \text{с вероятностью } 1 - p_k, \end{cases}$$
  3. перейти на Шаг 1.

Недифференцируемые функции

Дифференцируемое и недифференцируемое

Для непрерывных функций:

1. Дифференцируемая функция (по Фреше)

$$f(x + \Delta x) = f(x) + (\nabla f(x), \Delta x) + o(\|\Delta x\|)$$

2. Производная по направлению (производная Гато)

$$f(x + \gamma d) = f(x) + \gamma D f(x, d) + o(\gamma), \;\; \gamma > 0$$

Субдифференциал

Для выпуклой дифференцируемой функции

$$f(x + \Delta x) \geq f(x) + \langle \nabla f(x), \Delta x \rangle, \quad \forall \Delta x$$

$$f(z) \geq f(x) + \langle \nabla f(x), z - x \rangle, \quad \forall z$$

Субдифференциал

Для выпуклой дифференцируемой функции

$$f(x + \Delta x) \geq f(x) + (\nabla f(x), \Delta x), \quad \forall \Delta x$$

$$f(z) \geq f(x) + (\nabla f(x), z - x), \quad \forall z$$

Для выпуклой недифференцируемой функции

$$\partial f(x) = \{v : f(z) \geq f(x) + (v, z - x), \forall z\}$$

Субдифференциал

$$\partial f(x) = conv\{Df(x, d) : d \in \mathbb{R}^n\}$$

Элементы субдифференциала - субградиенты (тоже обозначаются как $\partial f(x)$).

Свойства субградиентов

- Если функция дифференцируема в x, то
$$\partial f(x) = \{\nabla f(x)\}$$
- Сумма функций – сумма множеств
$$\partial\left(\sum_i \alpha_i f_i(x)\right) = \sum_i \alpha_i \partial f_i(x), \; \alpha_i \geq 0$$
- Максимум функций – выпуклая оболочка множеств
$$\partial \max_i f_i(x) = conv\left(\partial f_i(x), i \in \operatorname{Arg}\max_i f_i(x)\right)$$

Необходимые и достаточные условия оптимальности

Дифференцируемые функции

$$\nabla f(x) = 0$$

Недифференцируемые функции

$$\partial f(x) \ni 0$$

Методы?

Субградиентные метод

$$x^{k+1} = x^k - \gamma_k \partial f(x^k)$$

Типично

- $\gamma_k \geq 0$
- $\gamma_k \to 0$
- $\sum_k \gamma_k \to \infty$
- + дополнительные требования к задаче

Постоянная длина шага может не сработать!

Три способа выбора шага: первый

Theorem

Для выпуклой функции $f(\cdot)$, $\|\partial f(\cdot)\| \leq c$, $\exists x^*$, *такой, что*

- $\gamma_k \geq 0$,
- $\gamma_k \to 0$,
- $\sum_k \gamma_k \to \infty$.

Тогда алгоритм

$$x^{k+1} = x^k - \gamma_k \frac{\partial f(x^k)}{\|\partial f(x^k)\|}$$

сходится $\underline{\lim}_{k\to\infty} f(x^k) = f^*$

Три способа выбора шага: второй

Theorem

Для выпуклой функции $f(\cdot)$, $\exists x^*$

- $\gamma_k \geq 0$,
- $\gamma_k \to 0$,
- $\sum_k \gamma_k \to \infty$,
- $\sum_k \gamma_k^2 < \infty$.

Тогда алгоритм

$$x^{k+1} = x^k - \gamma_k \frac{\partial f(x^k)}{\|\partial f(x^k)\|}$$

приводит к $f(x^k) \to_{k \to \infty} f^*, x^k \to_{k \to \infty} x^*$

Три способа выбора шага: третий

Theorem (шаг По́ляка)

Для выпуклой $f(\cdot)$, f^* существует и известно, $\exists x^*$. Тогда алгоритм

$$x^{k+1} = x^k - \gamma_k \partial f(x^k)$$

с

$$\gamma_k = \frac{f(x^k) - f^*}{\|\partial f(x^k)\|^2}$$

сходится

$$\underline{\lim}_{k\to\infty} f(x^k) = f^*$$

Приложение к линейным неравенствам

$$Ax \leq b$$

$$(a^i, x) \leq b_i, \quad i = 1, \ldots, m$$

Положить $d_i(x) = [(a^i, x) - b_i]_+$ и минимизировать

$$f(x) = \max_i d_i(x)$$

$\partial f(x) =?$

Приложения к линейным неравенствам (с выпуклым функциям)

Найти x:

$$f_i(x) \leq 0, \quad i = 1, \ldots, m$$

Эквивалентная задача минимизации

$$f(x) = \max_i [f_i(x)]_+ \to \min_x$$

$\partial f(x) = ?$

$$x^{k+1} = x^k - \ldots$$

Преобразование функций

$$\min_x f(x)$$

- Добавление константы
- Мажоранты
- Сдвиг по координатам
- Масштабирование
- Смена координат
- Модификация функции