I. Безусловная оптимизация

20 февраля 2019 г. 14:09

Tags used:

Homework

Example

Definitions used

$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ - вектор ($\boldsymbol{x}$ в методичке)

В примерах accent обычно опускается, т.к. надо быстро печатать

$x^{(0)}, x^{(1)}, \ldots$ — результаты итерации, если используются x_n как компоненты.

В противном случае, результаты итерации - $x_0, x_1, \ldots$

Безусловная оптимизация

Преподаватель - Балтрашевич В.Э.

$X = \mathbb{R}^n$

Если f — дифференцируема, то в точке минимума $\nabla f(x) = 0$

$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right)$

Релаксационные методы - методы построения последовательности $\{x_i\}, x_i \in X$:

- $f(x_0) > f(x_1) > \cdots$
- $x_i \to x^* = \arg\min\{f(x), x \in X\}$

Порядок метода - порядок старшей производной f, используемой при работе метода.

Общая схема

$\boxed{\overrightarrow{x_{k+1}} = \overrightarrow{x_k} + t_k \overrightarrow{S_k}}$

$\overrightarrow{S_k}$ — некоторый вектор, определяющий направление изменения $\overrightarrow{x_k}$

t_k — длина шага

Методы могут быть

- **Одношаговые** $\overrightarrow{S_k} = \varphi(x_k)$
- **Двухшаговые** $\overrightarrow{S_k} = \varphi(x_k, x_{k-1})$
- …

Сходимость релаксационного метода

$d = const$ — наибольшее из чисел, при котором выполняется условие

$\boxed{\exists c : \forall k \ \left\|\overrightarrow{x_{k+1}} - \overrightarrow{x^*}\right\| \le c \cdot \|x_k - x^*\|^d}$

Тогда d — скорость сходимости.

$\Delta_k := \left\|\overrightarrow{x_k} - \overrightarrow{x^*}\right\|; \Delta_k \xrightarrow[k\to\infty]{} 0 \Rightarrow d = \overline{\lim_{k\to\infty}} \frac{\ln \Delta_{k+1}}{\ln \Delta_k}$

$\Delta_{k+1} \approx q\Delta_k, 0 < q < q \Rightarrow d = 1 -$ **геометрическия / линейная** скорость
$\Delta'_{k+1} \approx const \cdot \Delta_k^2 \Rightarrow d = 2 -$ **квадратичная** скорость

Методы первого порядка (градиентные методы)

См. ФОИТ :: Метод градиентного спуска

Градиентный метод с постоянным шагом

t_k не зависит k, т.е. $t_k = const$
Тогда итерационная формула:

$$\boxed{\overrightarrow{x_{k+1}} = \overrightarrow{x_k} - t\nabla f(\overrightarrow{x_k})}$$

Теорема о сходимости метода с постоянным шагом

Если:

- $f -$ дифференцируема на $\mathbb{R}^n$
- $\nabla f(x)$ удовлетворяет **глобальному условию Липшица**
  $\exists \mathrm{L} > 0: \forall \vec{x}, \vec{y}: \|\nabla f(\vec{x}) - \nabla f(\vec{y})\| \le L \cdot \|\vec{x} - \vec{y}\|$
- $f -$ ограничена снизу, т.е.
  $\exists f^*, \text{т.ч.} \forall \vec{x}: f(\vec{x}) \ge f^* > -\infty$
- $0 < t < \dfrac{2}{L}$

Тогда для последовательности $\{\overrightarrow{x_k}\}$, получаемого по формуле $\vec{x}_{i+1} = \overrightarrow{x_k} - t\nabla f(x_k)$

- $\forall k$: $f(x_{k+1}) \le f(x_k) - t\left(1 - t\frac{L}{2}\right) \cdot \|\nabla f(\overrightarrow{x_k})\|^2$
  $\{f(\overrightarrow{x_k})\}$ монотонно убывает: $f(\overrightarrow{x_{k+1}}) < f(\overrightarrow{x_k})$ при $\nabla f(\overrightarrow{x_k}) \ne 0$
- $\nabla f(\overrightarrow{x_k}) \xrightarrow[k\to\infty]{} 0$

Доказательство

$$\frac{df(\vec{x} + \tau\vec{y})}{d\tau} = (\nabla f(\vec{x} + \tau\vec{y}), y) \Rightarrow$$

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

Пусть $\vec{x} = \overrightarrow{x_k}; y = -t\nabla f(\overrightarrow{x_k})$
Тогда

$$f(\overrightarrow{x_{k+1}}) = f(\overrightarrow{x_k}) - t \cdot \|\nabla f(x_k)\|^2 + \int_0^1 (\nabla f(\overrightarrow{x_k} + \tau t\nabla f(\overrightarrow{x_k})) - \nabla f(\vec{x}_k), -t\nabla f(\overrightarrow{x_k}))d\tau$$

С учётом $(x, y) \le \|x\| \cdot \|y\|$ и неравенства Липшица:

$$\left(\nabla f\left(\overrightarrow{x_k} - t\tau\nabla f(\overrightarrow{x_k})\right) - \nabla f(\overrightarrow{x_k}), -t\nabla f(\overrightarrow{x_k})\right) \le$$
$$\le \left\|\nabla f\left(\overrightarrow{x_k} - t\tau\nabla f(\overrightarrow{x_k})\right) - \nabla f(\overrightarrow{x_k})\right\| \cdot \|-t\nabla f(\overrightarrow{x_k})\| \le$$
$$\le L\|-t\tau\nabla f(\overrightarrow{x_k})\| \cdot \|-t\nabla f(\overrightarrow{x_k})\| = Lt^2\tau\|\nabla f(\overrightarrow{x_k})\|^2$$

Тогда

$$f(\overrightarrow{x_{k+1}}) \le f(\overrightarrow{x_k}) - t\|\nabla f(\overrightarrow{x_k})\|^2 + Lt^2\|\nabla f(\overrightarrow{x_k})\|^2 \int_0^1 \tau d\tau = f(\overrightarrow{x_k}) - \tau\left(1 - \frac{Lt}{2}\right)\|\nabla f(\overrightarrow{x_k})\|^2$$

$a := t\left(1 - \frac{Lt}{2}\right); 0 < \mathrm{t} < \frac{2}{\mathrm{L}} \Rightarrow \mathrm{a} > 0 \Rightarrow \forall k: f(\overrightarrow{x_{k+1}}) \le f(\overrightarrow{x_k}) - a\|\nabla f(\overrightarrow{x_k})\|^2$

Сложим полученные неравенства при $k = 0,1, \dots, s$:

$$f(\overrightarrow{x_{s+1}}) \le f(\overrightarrow{x_0}) - a\sum_{k=0}^{s}\|\nabla f(\overrightarrow{x_k})\|^2$$

$$a > 0 \Rightarrow \sum_{k=0}^{s}\|\nabla f(\overrightarrow{x_k})\|^2 \le a^{-1}\big(f(\overrightarrow{x_0}) - f(\overrightarrow{x_{s+1}})\big)$$

$$f(\vec{x}) \ge f^* > -\infty \Rightarrow \forall s\colon \sum_{k=0}^{s}\|\nabla f(\overrightarrow{x_k})\|^2 \le a^{-1}(f(\overrightarrow{x_0}) - f^*) \Rightarrow$$

$$\Rightarrow \sum_{k=0}^{s}\|\nabla f(\overrightarrow{x_k})\|^2 < \infty;\ \|\nabla f(\overrightarrow{x_k})\| \xrightarrow[k\to\infty]{} 0 \blacksquare$$

<u>Покоординатный спуск</u>

На каждом этапе выполняется шаг только по одной координате

<u>Пример [ДЗ 27.02.2019]</u>

$f(x_1, x_2) = x_1^2 + (x_2 - 1)^2$

Пусть метод градиентного спуска начинается из точки $(0,0)$

Первое направление - $\begin{pmatrix}1\\0\end{pmatrix}$, второе - $\begin{pmatrix}0\\1\end{pmatrix}$

$x^{(0)} = \begin{pmatrix}0\\0\end{pmatrix}$

$x^{(1)} = \begin{pmatrix}0\\0\end{pmatrix} - \gamma\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}-\gamma\\0\end{pmatrix}$

$f\big(x^{(1)}\big) = \gamma^2$

$f' = 2\gamma$

$f' = 0 \Rightarrow \gamma = 0 \Rightarrow x^{(1)} = \begin{pmatrix}0\\0\end{pmatrix}$

$x^{(2)} = \begin{pmatrix}0\\0\end{pmatrix} - \gamma\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}0\\-\gamma\end{pmatrix}$

$f\big(x^{(2)}\big) = (-\gamma - 1)^2 = \gamma^2 + 2\gamma + 1 = \psi(\gamma)$

$f' = 2\gamma + 2$

$f' = 0 \Rightarrow \gamma = -1 \Rightarrow x^{(2)} = \begin{pmatrix}1\\0\end{pmatrix}$

<u>Выпуклые функции и множества</u>

$X \subseteq \mathbb{R}^n -$ **выпуклое**, если $\forall \overrightarrow{x_1}, \overrightarrow{x_2} \in X, \lambda \in [0,1]\colon\ \lambda\vec{x}_1 + (1-\lambda)\overrightarrow{x_2} \in X$

$\vec{z} -$ **выпуклая комбинация** $\overrightarrow{x_1}, \overrightarrow{x_2}, \dots, \overrightarrow{x_m}$, если $\vec{z} = \sum_{i=1}^{m} a_i\overrightarrow{x_i}\,, a_i \ge 0, \sum a_i = 1$

Теорема о выпуклом множестве

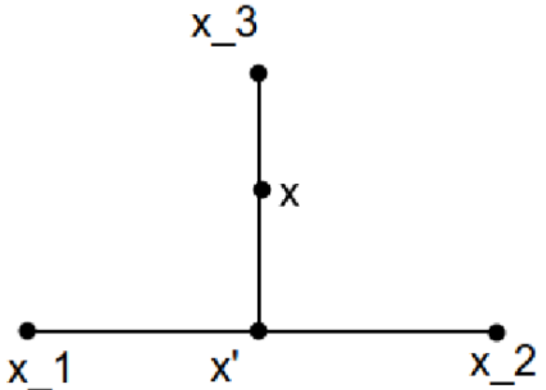
Выпуклое множество X содержит все выпуклые комбинации своих точек

План доказательства

Доказательство методом индукции.

Для двух точек очевидно

$\forall x\ \exists\lambda\colon\ x_1 \cdot \lambda + x_2 \cdot (1-\lambda) = x$
Добавим ещё одну точку:



$x' = \lambda_1 x_1 + (1-\lambda_1)x_2$
$x = \lambda_2 x' + (1-\lambda_2)x_3 = \lambda_2\lambda_1 x_1 + \lambda_2(1-\lambda_1)x_2 + (1-\lambda_2)x_3$
$\alpha_1 = \lambda_1\lambda_2$
$\alpha_2 = \lambda_2 - \lambda_1\lambda_2$
$\alpha_3 = 1 - \lambda_2$

$$\sum \alpha_i = \lambda_1\lambda_2 + \lambda_2 - \lambda_1\lambda_2 + 1 - \lambda_2 = 1$$

Далее по индукции

Теорема. Пересечение выпуклых множеств выпукло.
Доказательство
Пусть X, Y — выпуклые, т.е.
$Z \coloneqq X \cap Y$ — пересечение
По определению $Z \subseteq X, Z \subseteq Y$. По теореме о выпуклом множестве Z — выпукло ■

Функция f называется **выпуклой**, если:
- $D(f)$ — область определения - выпукла
- $\forall \overrightarrow{x_1}, \overrightarrow{x_2} \in D(f)\colon f(\lambda\overrightarrow{x_1} + (1-\lambda)\overrightarrow{x_2})) \le \lambda f(\overrightarrow{x_1}) + (1-\lambda)f(\overrightarrow{x_2}), \lambda \in (0,1)$

Функция f называется **вогнутой**, если $-f$ выпукла.

Функция f называется **строго выпуклой**, если
$\forall \vec{x} \ne y, 0 < \lambda < 1\colon f(\lambda\vec{x} + (1-\lambda)\vec{y}) < \lambda f(x) + (1-\lambda)f(y)$

Функция f называется **сильновыпуклой** с константой $l > 0$, если
$\forall \lambda \in [0,1]\colon f(\lambda\vec{x} + (1-\lambda)\vec{y}) \le \lambda f(x) + (1-\lambda)f(y) - l\lambda(1-\lambda)\|\vec{x} - \vec{y}\|$

Свойства выпуклых функций
1. Любая точка локального минимума выпуклой функции является точкой её глобального минимума
   Доказательство

Пусть $\vec{x}^*$ — локальный, но не глобальный минимум. Значит
$\exists \vec{y} \in X: f(\vec{y}) < f(\vec{x}^*)$
Пусть $\vec{x} = \lambda\vec{y} + (1-\lambda)\vec{x}^*, \lambda \in (0,1)$
Т.к. X — выпукло, $\vec{x} \in$ X.
Т.к. f — выпукла, $f(\vec{x}) = f(\lambda\vec{y} + (1-\lambda)\vec{x}^*) \leq \lambda f(y) + (1-\lambda)f(\vec{x}^*) + (1-\lambda)f(\vec{x}^*) = f(\vec{x}^*)$
Выходит, что $\forall \lambda \in (0,1): f(\vec{x}) < f(\vec{x}^*)$. Значит, $\vec{x}^*$ — не локальный минимум. ■

2. **Критерий выпуклости дифференцируемых функций**
   Пусть $\exists f''$ — непрерывна.
   f — выпукла $\Leftrightarrow f'' > 0$
   f — сильновыпукла $\Leftrightarrow \exists l, \forall x: \nabla^2 f(\vec{x}) \geq l\vec{I}$
3. **Неравенство Йенсена**
   f — выпукла
   $$\Leftrightarrow \forall m \geq 2, \overrightarrow{x_1}, \overrightarrow{x_2}, \dots, \overrightarrow{x_m} \in X, \lambda_i \geq 0, \sum_{i=1}^{m} \lambda_i = 1:$$
   $$f\left(\sum_{i=1}^{m} \lambda_i \overrightarrow{x_i}\right) \leq \sum_{i=1}^{m} \lambda_i f(\overrightarrow{x_i})$$
   [док-во см. 9-10 мет. опт.]

4. f — выпукла, $X = D(f)$ — выпукло $\Rightarrow f$ — непрерывна в каждой внутренней точке X и имеет в ней производную по любому направлению
5. f — дифференцируемая функция на выпуклом множестве X.
   > Выпуклость эквивалентна
   $\forall \vec{x}, \vec{y} \in X: f(\vec{x} + \vec{y}) \geq f(\vec{x}) + (\nabla f(\vec{x}), \vec{y})$
   > Строгая выпуклость эквивалентна
   $\forall \vec{y} \neq 0, \vec{x} \in X: f(\vec{x} + \vec{y}) \geq f(\vec{x}) + (\nabla f(\vec{x}), \vec{y})$
   > Сильная выпуклость эквивалентна
   $$\forall \vec{x}, \vec{y} \in X, l > 0: f(\vec{x} + \vec{y}) \geq f(\vec{x}) + (\nabla f(\vec{x}), \vec{y}) + \frac{l\|\vec{y}\|^2}{2}$$
6. Для сильновыпуклых функций:
   > $f(\vec{x}) \geq f(\vec{x}^*) + \dfrac{l\|\vec{x} - \vec{x}^*\|^2}{2}$
   > $(\nabla f(\vec{x}), x - \vec{x}^*) \geq l\|\vec{x} - \vec{x}^*\|^2$
   > $\|\nabla f(\vec{x})\| \geq l\|\vec{x} - \vec{x}^*\|$

Теорема. Пусть $\exists f''$. Если для $l, L > 0\ \forall x$ верно:
$l\vec{I} \leq \nabla^2 f(x) \leq L\vec{I}$
То $\left\|\overrightarrow{x_k} - \overrightarrow{x^*}\right\| \leq \|x_0 - x^*\| \cdot q^k$
$q = \max\{|1-\gamma l|, |1-\gamma L|\}$
$$\min q = \frac{L-l}{L+l} < 1$$
[доказательство см. 11 мет. опт.]

Теорема. Если $\varphi(x)$ — выпукла на X, то $f(x) = \max\{\varphi(x), 0\}$ — тоже выпукла [ДЗ

Доказательство.
По определению выпуклой функции
$\lambda \in (0,1), \forall \overrightarrow{x_1}, \overrightarrow{x_2} \in X$:
$\varphi(\lambda\overrightarrow{x_1} + (1-\lambda)\overrightarrow{x_2})) \leq \lambda\varphi(\overrightarrow{x_1}) + (1-\lambda)\varphi(\overrightarrow{x_2})$

Докажем, что: $\max\{a+b, 0\} \leq \max\{a, 0\} + \max\{b, 0\}$
- $a \geq 0, b \geq 0 \Rightarrow \max\{a+b, 0\} = a + b = \max\{a, 0\} + \max\{b, 0\}$
- $a < 0, b \geq 0, a+b > 0 \Rightarrow a + b < b \Rightarrow$
  $\max\{a+b, 0\} = a + b < b = \max\{a, 0\} + \max\{b, 0\}$
- $a < 0, b \geq 0, a+b < 0 \Rightarrow$
  $\max\{a+b, 0\} = 0 < b = \max\{a, 0\} + \max\{b, 0\}$

Очевидно: $\max\{ac, 0\} = c \cdot \max\{a, 0\}$, если $c > 0$
Очевидно: $max\{a, 0\} \leq ma\,x\{b, 0\}$, если $a \leq b$

$f(\lambda\overrightarrow{x_1} + (1-\lambda)\overrightarrow{x_2}) = \max\{\varphi(\lambda\overrightarrow{x_1} + (1-\lambda)\overrightarrow{x_2}), 0\} \leq \max\{\lambda\varphi(\overrightarrow{x_1}) + (1-\lambda)\varphi(\overrightarrow{x_2}), 0\} \leq$
$\leq \max\{\lambda\varphi(\overrightarrow{x_1}), 0\} + \max\{(1-\lambda)\varphi(\overrightarrow{x_2}), 0\} = \lambda \max\{\varphi(\overrightarrow{x_1}, 0\} + (1-\lambda)\max\{\varphi(\overrightarrow{x_2}), 0\} =$
$= \lambda f(\overrightarrow{x_1}) + (1-\lambda) f(\overrightarrow{x_2})$ ■

Метод с дроблением шага, метод наискорейшего спуска
См. ФОИТ :: Метод градиентного спуска

Пример метода наискорейшего спуска

$$f(x_1, x_2) = x_1^2 + (x_2 - 1)^2$$

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 2x_1 \\ 2x_2 - 2 \end{pmatrix}$$

$$x^{(0)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$x^{(1)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \gamma \begin{pmatrix} 0 \\ -2 \end{pmatrix} = \begin{pmatrix} 0 \\ 2\gamma \end{pmatrix}$$

$$f\Big|_{x^{(1)}} = (2\gamma - 1)^2 = 4\gamma^2 - 4\gamma + 1$$

$$f' = 8\gamma - 4 = 0 \Rightarrow \gamma = \frac{1}{2} \Rightarrow x^{(1)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Чтобы проверить, что это точка минимума, нужно найти вторую производную

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}$$

$$\nabla^2 f\Big|_{x^{(1)}} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

Функция положительно определена, значит $x^{(1)}$ - точка минимума

Масштабирование
Заключается в замене переменных вида
$x_i = \mu_i y_i$
Это позволяет уменьшить вытянутость функции вдоль некоторых осей.

Пример. ДЗ 06.03.2019
$f(x) \coloneqq x_1^2 + 16x_2^2$
Найти минимум с масштабированием и без масштабирования.
$x^{(0)} = \begin{pmatrix} 5 \\ 5 \end{pmatrix}$

Без масштабирования
$$\nabla f(x) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 2x_1 \\ 32x_2 \end{pmatrix}$$
$$x^{(0)} = \begin{pmatrix} 5 \\ 5 \end{pmatrix}$$
$$x^{(1)} = \begin{pmatrix} 5 \\ 5 \end{pmatrix} - \gamma \begin{pmatrix} 10 \\ 160 \end{pmatrix} = \begin{pmatrix} 5 - 10\gamma \\ 5 - 160\gamma \end{pmatrix}$$
$$f\Big|_{x^{(1)}} = (5 - 10\gamma)^2 + 16(5 - 160\gamma)^2$$
$$= 25 - 100\gamma + 100\gamma^2 + 400 - 25600\gamma + 409600\gamma^2$$
$$= 409700\gamma^2 - 25700\gamma + 425$$
$$f'\Big|_{x^{(1)}} = 819400\gamma - 25700$$
$$f'\Big|_{x^{(1)}} = 0 \Rightarrow 891400\gamma = 25700 \Rightarrow \gamma = \frac{257}{8194}$$
$$x^{(1)} = \begin{pmatrix} \dfrac{5 \cdot 8194 - 10 \cdot 257}{8194} \\ \dfrac{5 \cdot 8194 - 160 \cdot 257}{8194} \end{pmatrix} = \begin{pmatrix} \dfrac{38400}{8194} \\ \dfrac{-150}{8194} \end{pmatrix} = \begin{pmatrix} \dfrac{19200}{4097} \\ \dfrac{-75}{4097} \end{pmatrix}$$
$$x^{(2)} = x^{(1)} - \gamma \nabla f\left(x^{(1)}\right) = \begin{pmatrix} x_1^{(1)} \\ x_2^{(1)} \end{pmatrix} - \gamma \begin{pmatrix} 2x_1^{(1)} \\ 32x_2^{(1)} \end{pmatrix} = \begin{pmatrix} x_1^{(1)} - 2\gamma x_1^{(1)} \\ x_2^{(1)} - 32\gamma x_2^{(1)} \end{pmatrix} = \begin{pmatrix} x_1^{(1)}(1 - 2\gamma) \\ x_2^{(1)}(1 - 32\gamma) \end{pmatrix}$$
$$= \begin{pmatrix} \dfrac{38400}{8194}\left(1 - 2 \cdot \dfrac{257}{8194} \cdot \gamma\right) \\ -\dfrac{150}{8194}\left(1 - 32 \cdot \dfrac{257}{8194} \cdot \gamma\right) \end{pmatrix} = \begin{pmatrix} 38400\,\dfrac{8194 - 514\gamma}{8194} \\ -150\,\dfrac{8194 - 16864\gamma}{8194} \end{pmatrix}$$
$$= \begin{pmatrix} \dfrac{314649600 - 19737600\gamma}{8194} \\ \dfrac{-1229100 + 2529600\gamma}{8194} \end{pmatrix}$$
$$f\Big|_{x^{(2)}} = \left(\frac{314649600 - 19737600\gamma}{8194}\right)^2 + 16\left(\frac{1229100 - 2529600\gamma}{8194}\right)^2$$
Не получается

С масштабированием

$$f(x) = x_1^2 + 16x_2^2$$

$$\mu_1 = \frac{1}{\sqrt{\frac{\partial^2 f}{\partial x_1^2}(x^{(0)})}} = \frac{1}{\sqrt{2}}$$

$$\mu_2 = \frac{1}{\sqrt{\frac{\partial^2 f}{\partial x_2^2}(x^{(0)})}} = \frac{1}{\sqrt{32}} = \frac{1}{4\sqrt{2}}$$

$$f(y) = \frac{y_1^2}{2} + \frac{y_2^2}{2}$$

$$\nabla f(y) = \binom{y_1}{y_2}$$

$$x^{(0)} = \begin{pmatrix} \frac{5}{\sqrt{2}} \\ \frac{5}{4\sqrt{2}} \end{pmatrix}$$

$$x^{(1)} = \begin{pmatrix} \frac{5}{\sqrt{2}} \\ \frac{5}{4\sqrt{2}} \end{pmatrix} - \gamma \begin{pmatrix} \frac{5}{\sqrt{2}} \\ \frac{5}{4\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{5}{\sqrt{2}}(1-\gamma) \\ \frac{5}{4\sqrt{2}}(1-\gamma) \end{pmatrix}$$

$$f\Big|_{x^{(1)}} = \frac{\left(\frac{5}{\sqrt{2}}(1-\gamma)\right)^2}{2} + \frac{\left(\frac{5}{4\sqrt{2}}(1-\gamma)\right)^2}{2} = \frac{(1-\gamma)^2}{2}\left(\frac{25}{2} + \frac{25}{32}\right)$$

$$= \frac{425}{64} - \frac{425}{32}\gamma + \frac{425}{64}\gamma^2$$

$$f'\Big|_{x^{(1)}} = \frac{425}{32}\gamma - \frac{425}{32} = 0 \Rightarrow \gamma = 1$$

$$x^{(1)} = \begin{pmatrix} \frac{5}{\sqrt{2}}(1-1) \\ \frac{5}{4\sqrt{2}}(1-1) \end{pmatrix} = \binom{0}{0}$$

<u>Метод Ньютона</u>

Ряд Тейлора:

$$f(\vec{x}) = f(\vec{x}_k) + (\nabla f(\vec{x}_k), \vec{x} - \vec{x}_k) + \frac{1}{2}(\nabla^2 f(\vec{x}_k)(\vec{x} - \vec{x}_k), \vec{x} - \vec{x}_k) + o(\|\vec{x} - \vec{x}_k\|)$$

Пусть $f_2(\vec{x})$ — квадратичная аппроксимация $f(\vec{x})$

$f_2(\vec{x})$ имеет единственную точку минимума, которая является корнем $\nabla f_2(\vec{x}) = 0$

В данном случае

$$\nabla f_2(\vec{x}) = 0 = \nabla f(\vec{x}_k) + \nabla^2 f(\vec{x}_k)(\vec{x} - \vec{x}_k)$$

$\vec{x}_{k+1} = \vec{x}_k - [\nabla^2 f(\vec{x}_k)]^{-1} \nabla f(\vec{x}_k)$

Пример

$f(\vec{x}) = \frac{1}{2}(A\vec{x}, \vec{x}) - \left(\vec{b}, \vec{x}\right)$

$\nabla f(\vec{x}) = A\vec{x} - \vec{b}$

$\nabla f(\vec{x}) = 0 \Rightarrow \vec{x} = A^{-1}\vec{b}$

$\nabla^2 f(\vec{x}_k) = A$

$\overrightarrow{x_1} = \overrightarrow{x_0} - A^{-1}\left(A\overrightarrow{x_0} - \vec{b}\right) = A^{-1}b$

Достоинства и недостатки

Градиентный метод	Метод Ньютона
✓ Слабые требования к исходным данным. x_0 может быть далека от x^* ✓ Используется только градиент функции f ✓ Относительная простота вычислений – Медленная скорость сходимости	✓ Быстрая скорость сходимости (квадратичная) – Только локальная сходимость. Начальное приближение должно быть достаточно близким – Жесткие требования к функции (должна быть дважды непрерывно дифференцируема) – Большой объем вычислений, связанный с необходимостью вычисления матрицы вторых производных и её обращений

Метод Пауэлла

- Выбираются две точки и направление
- Через две точки рисуются векторы
- На векторах ищется минимальное значение
- Через точки, в которых на векторе минимальное значение рисуется следующий вектор

Пример

$f(x_1, x_2) = x_1^2 + (x_2 - 1)^2$

$x^{(-1)} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}; x^{(0)} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$; Направление - $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$

$x^{(1)} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} - \gamma \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 - \gamma \\ 2 \end{pmatrix}$

$f\left(x^{(1)}\right) = (2 - \gamma)^2 + 1$

$f'(x^{(1)}) = 2\gamma - 4 = 0 \Rightarrow \gamma = 2$

$x^{(1)} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$

$x^{(2)} = \begin{pmatrix} 2 \\ 3 \end{pmatrix} - \gamma \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 - \gamma \\ 3 \end{pmatrix}$

$f\left(x^{(2)}\right) = (2 - \gamma)^2 + 4$

$f'\left(x^{(2)}\right) = 2\gamma + 4 = 0 \Rightarrow \gamma = 2$

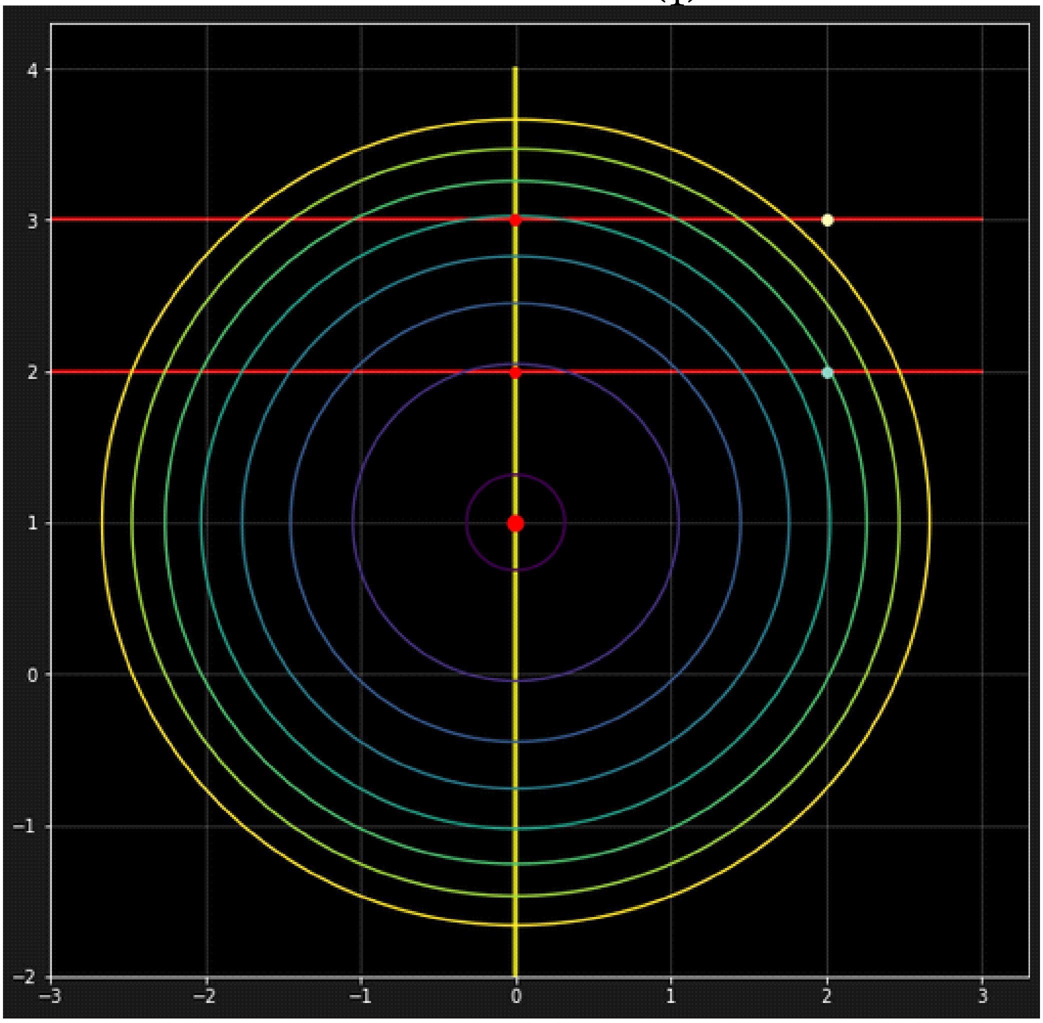
$x^{(2)} = \begin{pmatrix} 0 \\ 3 \end{pmatrix}$

Новое направление - $x^{(2)} - x^{(1)} = \binom{0}{1}$

$x^{(3)} = \binom{0}{2} - \gamma\binom{0}{1} = \binom{0}{2-\gamma}$

$f\left(x^{(3)}\right) = (1-\gamma)^2$

$f'\left(x^{(3)}\right) = 2\gamma - 2 = 0 \Rightarrow \gamma = 1 \Rightarrow x^{(3)} = \binom{0}{1}$



Многошаговые методы

Метод тяжелого шарика

Заключается в учёте "инерции движения"

$\vec{x}_{k+1} = \overrightarrow{x_k} - \alpha\nabla f(\overrightarrow{x_k}) + \beta(\overrightarrow{x_k} - \overrightarrow{x_{k-1}})$

Теорема о скорости сходимости метода тяжелого шарика

Если:

- $0 < l \leq \nabla^2 f(\vec{x}) \leq L$ (сильная выпуклость функции)
- $0 \leq \beta \leq 1; 0 \leq \alpha \leq \dfrac{2(1+\beta)}{L}$

То:

- $\exists c, q, \forall k: \|\vec{x}_k - \vec{x}^*\| \leq cq^k$
- $q_{min} = \dfrac{\sqrt{L} - \sqrt{l}}{\sqrt{L} + \sqrt{l}}$

Получается, что этот метод сходится не быстрее геометрической прогрессии. Поэтому при плохой обусловленности предпочтительно применение одношагового градиентного метода.

Метод сопряженных градиентов
$\vec{x}_{k+1} = \vec{x}_k - \alpha_k \nabla f(\vec{x}_k) + \beta(\vec{x}_k - \vec{x}_{k-1})$
В отличие от метода тяжелого шарика, α, β вычисляются следующим образом:
$(\alpha_k, \beta_k) = \arg\min\{f\big(\vec{x}_k - \alpha\nabla f(\vec{x}_k) + \beta(\vec{x}_k - \vec{x}_{k-1})\big): \alpha > 0, \beta > 0\}$

Для квадратичной функции $\frac{1}{2}(A\vec{x}, \vec{x}) - \left(\vec{b}, \vec{x}\right), A > 0$

- Метод сходится за конечное число шагов, не превосходящее размерности пространства состояний
- Градиенты в точках $\vec{x}_k$ попарно ортогональны
  $\forall i \neq k: \big(\nabla f(\vec{x}_i), \nabla f(\vec{x}_k)\big) = 0$
- $\vec{p}_k \coloneqq \vec{x}_k - \vec{x}_{k-1}$
  $\forall i \neq j: \left(A\vec{p}_i, \vec{p}_j\right) = 0$

Для некоторой положительно определенной матрица A векторы $\vec{p}_i$, связанные соотношением $\left(\mathrm{A}\vec{p}_i, \vec{p}_j\right) = 0$, называются **сопряженными** или $\boldsymbol{A}$ **− ортогональными**

Модификация Полака-Ривьера
$\vec{x}_{k+1} = \vec{x}_k + \alpha_k \vec{p}_k$
$\alpha_k = \arg\min f(\vec{x}_k + \alpha\vec{p}_k), \alpha > 0$
$\vec{p}_k = -\nabla f(\vec{x}_k) + \beta\vec{p}_{k-1}$
$$\beta_k = \frac{\Big(\big(\nabla f(\vec{x}) - \nabla f(\vec{x}_{k-1})\big), \nabla f(\vec{x}_k)\Big)}{\|\nabla f(\vec{x}_{k-1})\|}, \beta_0 = 0$$

Квазиньютоновские методы

Итерационная схема метода имеет вид:
$\vec{x}_{k+1} = \vec{x}_k - \gamma_k H_k \nabla f(\vec{x}_k)$
Если $H_k = 1$, это градиентный метод
Если $H_k = \left(\nabla^2 f(\vec{x}_k)\right)^{-1}, \gamma_k = 1$ – это метод Ньютона

Достоинство такого подхода в том, что не нужно вычислять обратную матрицу вторых производных.

Обозначим:
$p_k \coloneqq -H\nabla f(\vec{x}_k)$
$y_k \coloneqq \nabla f(\vec{x}_{k-1}) - \nabla f(\vec{x}_k)$
$f(\vec{x}) \coloneqq \frac{A\vec{x}, \vec{x}}{2} + \left(\vec{b}, \vec{x}\right), A > 0$
Тогда для
$\vec{y}_k = A(\vec{x}_{k+1} - \vec{x}_k) = \gamma_k A\vec{p}_k, \gamma_k\vec{p}_k = A^{-1}\vec{\gamma}_k$
Квазиньютоновское условие: $H_{k+1}\vec{y}_k = \gamma_k \vec{p}_k$

Метод Давидона-Флетчера-Пауэлла

$$H_{k+1} = H_k - \frac{H_k \vec{y}_k \left(H_k \vec{y}_j\right)^T}{(H_k \vec{y}_k, \vec{y}_k)} + \gamma_k \frac{\vec{p}_k \vec{p}_k^T}{(\vec{p}_k, \vec{y}_k)}; H_0 > 0$$

На каждом шаге, имея H_k, делается шаг в направлении $\vec{p}_k$. γ_k можно получить, например, по методу наискорейшего спуска.
После чего получается x_{k+1}, вычисляется y_k и пересчитывается H_{k+1}

Метод Бройдена-Флетчера-Шенно

$$H_{k+1} = H_k - \frac{\rho_k \vec{p}_k (\vec{p}_k)^T - \vec{p}_k (\vec{y}_k)^T H_k - H_k \vec{y}_k (\vec{p}_k)}{(\vec{y}_k, \vec{p}_k)}$$

$$\rho_k = \gamma_k + \frac{(H_k \vec{y}_k, \vec{y}_k)}{(\vec{y}_k, \vec{p}_k)}$$

Методы нулевого порядка

Методы аппроксимации

Пусть e_j — орт оси j

$$\varphi\left(\vec{x} + \gamma \vec{e}_j\right) \approx f(\vec{x}) + \frac{\partial f}{\partial \vec{x}} \gamma + o(\gamma)$$

$$\frac{\partial f}{\partial \vec{x}_j} \approx \frac{f\left(\vec{x} + \gamma \vec{e}_j\right) - f(\vec{x})}{\gamma} \approx \frac{f\left(\vec{x} + \gamma \vec{e}_j\right) - f\left(\vec{x} - \gamma \vec{e}_j\right)}{2\gamma}$$

Метод покоординатного спуска

Метод симплексов

Задача - найти экстремум функции

$c_1 x_1 + c_2 x_2 + \cdots + c_n x_n \to min$

При ограничениях:

$a_{11} x_1 + a_{12} x_2 + \cdots + a_{1n} x_n \leq b_1$

$a_{21} x_1 + a_{22} x_2 + \cdots + a_{2n} x_n \leq b_2$

…

$a_{m1} x_1 + a_{m2} x_2 + \cdots + a_{mn} x_n \leq b_m$

Симплекс-таблица

	x_1	x_2	…	x_n	
y_1	a_{11}	a_{12}	…	a_{1n}	b_1
y_2	a_{21}	a_{22}	…	a_{2n}	b_2
…	…	…	…	…	…
y_m	a_{m1}	a_{m2}	…	a_{mn}	b_m
	c_1	c_2	…	c_n	0

Алгоритм

1. Выбрать разрешающий элемент
   Для поиска крайней точки:

- Разрешающая строка j – любой, где $b_j < 0$
- Разрешающий столбец i – такой, чтобы $a_{ij} > 0$, и $\frac{b_j}{a_{ij}}$ было максимальным

Для оптимальной точки:

- Разрешающий столбец i – такой, где $c_i < 0$
- Разрешающая строка j – такая, чтобы $a_{ij} < 0$, и $\frac{b_j}{a_{ij}}$ было максимальным

  Оптимальная точка ищется после крайней, поэтому все b_j будут положительными

2. Строится новая симплекс-таблица.
   - x_i меняется местами с y_j
   - Элемент (i, j) возводится в -1 степень
   - Все остальные элементы разрешающего столбца делятся на (i, j)
   - Все остальные элементы разрешающей строки делятся на (i, j) и умножаются на -1.
   - Все остальное вычисляется следующему правилу:
     $a_{qv} \coloneqq \frac{a_{ij} \cdot a_{qv} - a_{iv} \cdot a_{qv}}{a_{ij}}$
3. Крайняя точка не существует, если в таблице существует строка, в которой все элементы ≤ 0, а последний - < 0
   Оптимальная точка не существует, если в таблице есть столбец, в котором $c_j < 0$; но все $a_{ij} > 0$
4. Если в столбце b есть отрицательные элементы, обратно на шаг 1. Иначе решение найдено
   Оптимальная точка ищется следующим образом:
   - Если x_j находится на $i - \text{м}$ месте левого столбца, то его значение равно b_i
   - Если x_i находится на $j -$ м месте верхней строки, то его значение равно 0

Пример

Нужно найти минимум функции:

$f(x) = x_1 + x_2$

Ограничения:

$$\begin{cases} 3x_1 + 2x_2 - 6 > 0 \\ x_1 + 4x_2 - 4 > 0 \end{cases}$$

Симплекс-таблица:

	x_1	x_2	b
y_1	3	2	-6
y_2	1	4	-4
	1	1	0

1. Разрешающая строка - 1
   $-\frac{6}{3} > -\frac{6}{2} \Rightarrow a_{11}$ – разрешающий элемент

	y_1	x_2	b

x_1	3^{-1}	$-\frac{2}{3}$	$-\frac{-6}{3}$
y_2	$\frac{1}{3}$	$\frac{3\cdot 4-2\cdot 1}{3}$	$\frac{-4\cdot 3-(-6)\cdot 1}{3}$
	$\frac{1}{3}$	$\frac{3\cdot 1-2\cdot 1}{3}$	$\frac{3\cdot 0-(-6)\cdot 1}{3}$

	y_1	x_2	b
x_1	$\frac{1}{3}$	$-\frac{2}{3}$	2
y_2	$\frac{1}{3}$	$\frac{10}{3}$	-2
	$\frac{1}{3}$	$\frac{1}{3}$	2

2. Разрешающая строка - 2.
   $-\frac{2\cdot 3}{1} < -\frac{2\cdot 3}{10} \Rightarrow a_{22}$ — разрешающий элемент

	y_1	y_2	b
x_1	$\frac{1}{3}$	$-\frac{2}{3}$	2
x_2	$\frac{1}{3}$	$\frac{10}{3}$	-2
	$\frac{1}{3}$	$\frac{1}{3}$	2

	y_1	y_2	b
x_1	$\frac{\frac{1}{3}\cdot\frac{10}{3}+\frac{2}{3}\cdot\frac{1}{3}}{\left(\frac{10}{3}\right)}$	$-\frac{2}{3}\cdot\frac{3}{10}$	$\frac{2\cdot\frac{10}{3}-2\cdot\frac{2}{3}}{\left(\frac{10}{3}\right)}$
x_2	$-\frac{1}{3}\cdot\frac{3}{10}$	$\frac{3}{10}$	$-\left(-2\cdot\frac{3}{10}\right)$
	$\frac{\frac{1}{3}\cdot\frac{10}{3}-\frac{1}{3}\cdot\frac{1}{3}}{\left(\frac{10}{3}\right)}$	$\frac{1}{3}\cdot\frac{3}{10}$	$\frac{2\cdot\frac{10}{3}+2\cdot\frac{1}{3}}{\left(\frac{10}{3}\right)}$

	y_1	y_2	b
x_1	0.4	-0.2	1.6
x_2	-0.1	0.3	0.6

	0.3	0.1	2.2

Поиск ответа
Если x находится вверху, то это свободная переменная и он равен 0. Если x находится слева, он равен b.
Ответ: $x_1 = 1.6; x_2 = 0.6$

Методы прямого поиска

f — унимодальная функция.
Общая схема метода: $\boxed{x^{(k+1)} = x^{(k)} + t_k}$
Нужно определить t_k. Для этого ищется минимум функции $y(t) = f(x^{(k)} + t)$

Метод квадратичной интерполяции

Пусть f задана на прямой. Даны точки $a < b < c$: $f(a) \geq f(b); f(b) \leq f(c)$. На отрезке $[a, c]$ ищется минимум.
Через эти три точки проводится парабола:
$g(t) = g_0 + g_1 t + g_2 t^2$
Коэффициенты находятся из системы:

$$\begin{cases} g(a) = f(a) \\ g(b) = f(b) \\ g(c) = f(c) \end{cases}$$

Решение.

$$x_1 \coloneqq a; x_2 \coloneqq b; x_3 \coloneqq c; y_1 \coloneqq f(x_1); y_2 \coloneqq f(x_2); y_3 \coloneqq f(x_3)$$

$$g_2 = \frac{y_3 - \dfrac{x_3(y_2 - y_1) + x_2 y_1 - x_1 y_2}{x_2 - x_1}}{x_3(x_3 - x_1 - x_2) + x_1 x_2}$$

$$g_1 = \frac{y_2 - y_1}{x_2 - x_1} - g_2(x_1 + x_2)$$

$$g_0 = \frac{x_2 y_1 - x_1 y_2}{x_2 - x_1} + g_2 x_1 x_2$$

После этого находится значение минимума:

$$t^* = \arg\min g(t) \Rightarrow -\frac{g_1}{2g_2}$$

После этого:

$$a \leq t^* \leq b \Rightarrow c \coloneqq b; b \coloneqq t^*$$

$$b \leq t^* \leq c \Rightarrow a \coloneqq b; b \coloneqq t^*$$

Метод дихотомии (половинного деления)

См. ВМ :: метод бисекции

Метод "золотого" сечения

$[a, b]$ — интервал. $t', t'' \in [a, b]$
Пусть минимум функции расположен на интервале.
Тогда выберем точки так:

$\begin{cases} t' = a + F_1(b-a) \\ t'' = a + F_2(b-a) \end{cases}; F_1, F_2 \in (0,1)$

Как и прежде, идея состоит в замене $[a, b]$ на $[a, t'']$ или $[t', b]$

Выбор коэффициентов

$\frac{b-t'}{b-a} = \frac{t''-a}{b-a} = F_2 \Rightarrow F_1 + F_2 = 1$, т.к.

$\frac{b-a-F_1(b-a)}{b-a} = 1 - F_1 = F_2$ - после замены отрезок уменьшится в $\frac{1}{F_2} = \tau$

По правилу "золотого" сечения

$\frac{t'-a}{t''-a} = F_2 \Rightarrow F_1 = F_2^2$, т.к. $\frac{F_1(b-a)}{F_2(b-a)} = \frac{F_1}{F_2} = F_2$

$$\begin{cases} F_1 = F_2^2 \\ F_1 + F_2 = 1 \end{cases} \Rightarrow F_2^2 + F_1 = 1 \Rightarrow \begin{cases} F_1 = \dfrac{3-\sqrt{5}}{2} \approx 0.382 \\ F_2 = \dfrac{\sqrt{5}-1}{2} \approx 0.618 \end{cases}$$

Алгоритм

$f(t') \geq f(t'') \Rightarrow a \coloneqq t'; t' = t''; t'' = a + F_2(b-a)$

$f(t') < f(t'') \Rightarrow b \coloneqq t''; t'' = t'; t' = a + F_1(b-a)$