

# Одномерная оптимизация. Матрично-векторное дифференцирование

МЕТОДЫ ВЫПУКЛОЙ ОПТИМИЗАЦИИ

НЕДЕЛЯ 2

Даня Меркулов



# **Матрично-векторное дифференцирование. Линейный поиск**

## **Семинар**

**Оптимизация для всех! ЦУ**



# Напоминание с лекции

# Вспоминаем теорию. Дифференциал

- Дифференциал  $df(x)[\cdot] : U \rightarrow V$  в точке  $x \in U$  для  $f(\cdot) : U \rightarrow V$ :

$$f(x + h) - f(x) = \underbrace{df(x)[h]}_{\text{дифференциал}} + \bar{o}(\|h\|)$$

| $U \rightarrow V$         | $\mathbb{R}$           | $\mathbb{R}^n$  | $\mathbb{R}^{n \times m}$ |
|---------------------------|------------------------|-----------------|---------------------------|
| $\mathbb{R}$              | $f'(x)dx$              | $\nabla f(x)dx$ | $\nabla f(x)dx$           |
| $\mathbb{R}^n$            | $\nabla f(x)^T dx$     | $J(x)dx$        | —                         |
| $\mathbb{R}^{n \times m}$ | $tr(\nabla f(X)^T dX)$ | —               | —                         |

# Вспоминаем теорию. Дифференциал

- Дифференциал  $df(x)[\cdot] : U \rightarrow V$  в точке  $x \in U$  для  $f(\cdot) : U \rightarrow V$ :

$$f(x + h) - f(x) = \underbrace{df(x)[h]}_{\text{дифференциал}} + \bar{o}(\|h\|)$$

- Каноническая форма дифференциала:

| $U \rightarrow V$         | $\mathbb{R}$           | $\mathbb{R}^n$  | $\mathbb{R}^{n \times m}$ |
|---------------------------|------------------------|-----------------|---------------------------|
| $\mathbb{R}$              | $f'(x)dx$              | $\nabla f(x)dx$ | $\nabla f(x)dx$           |
| $\mathbb{R}^n$            | $\nabla f(x)^T dx$     | $J(x)dx$        | —                         |
| $\mathbb{R}^{n \times m}$ | $tr(\nabla f(X)^T dX)$ | —               | —                         |

# Вспоминаем теорию. Правила дифференцирования

- Полезные правила дифференцирования и стандартные производные:

| Правила дифференцирования   | Производные стандартных функций                             |
|---|---|
| $dA = 0$  | $d(\langle A, X \rangle) = \langle A, dX \rangle$           |
| $d(\alpha X) = \alpha(dX)$  | $d(\langle Ax, x \rangle) = \langle (A + A^T)x, dx \rangle$ |
| $d(AXB) = A(dX)B$   | $d(Det(X)) = Det(X)\langle X^{-T}, dX \rangle$              |
| $d(X + Y) = dX + dY$  | $d(X^{-1}) = -X^{-1}(dX)X^{-1}$                             |
| $d(X^T) = (dX)^T$   |   |
| $d(XY) = (dX)Y + X(dY)$   |   |
| $d(\langle X, Y \rangle) = \langle dX, Y \rangle + \langle X, dY \rangle$ |   |
| $d\left(\frac{X}{\phi}\right) = \frac{\phi dX - (d\phi)X}{\phi^2}$        |   |

# Вспоминаем теорию. Дифференциал и градиент / гессиан

Градиент можно найти по следующей формуле:

$$df(x) = \langle \nabla f(x), dx \rangle$$

# Вспоминаем теорию. Дифференциал и градиент / гессиан

Градиент можно найти по следующей формуле:

$$df(x) = \langle \nabla f(x), dx \rangle$$

Тогда, если у нас есть дифференциал в форме выше и мы хотим вычислить вторую производную матричной/векторной функции, мы рассматриваем "старый"  $dx$  как константу  $dx_1$ , затем вычисляем  $d(df) = d^2 f(x)$

# Вспоминаем теорию. Дифференциал и градиент / гессиан

Градиент можно найти по следующей формуле:

$$df(x) = \langle \nabla f(x), dx \rangle$$

Тогда, если у нас есть дифференциал в форме выше и мы хотим вычислить вторую производную матричной/векторной функции, мы рассматриваем "старый"  $dx$  как константу  $dx_1$ , затем вычисляем  $d(df) = d^2f(x)$

$$d^2f(x) = \langle \nabla^2 f(x)dx_1, dx \rangle = \langle H_f(x)dx_1, dx \rangle$$



# Вспоминаем теорию. Линейный поиск

- Методы локализации решения:

# Вспоминаем теорию. Линейный поиск

- Методы локализации решения:
  - Метод дихотомии

# Вспоминаем теорию. Линейный поиск

- Методы локализации решения:
  - Метод дихотомии
  - Метод золотого сечения

# Вспоминаем теорию. Линейный поиск

- Методы локализации решения:
  - Метод дихотомии
  - Метод золотого сечения
- Неточный линейный поиск:

# Вспоминаем теорию. Линейный поиск

- Методы локализации решения:
  - Метод дихотомии
  - Метод золотого сечения
- Неточный линейный поиск:
  - Условие достаточного убывания

# Вспоминаем теорию. Линейный поиск

- Методы локализации решения:
  - Метод дихотомии
  - Метод золотого сечения
- Неточный линейный поиск:
  - Условие достаточного убывания
  - Условия Гольдштейна

# Вспоминаем теорию. Линейный поиск

- Методы локализации решения:
  - Метод дихотомии
  - Метод золотого сечения
- Неточный линейный поиск:
  - Условие достаточного убывания
  - Условия Гольдштейна
  - Условие ограничения на кривизну

# Вспоминаем теорию. Линейный поиск

- Методы локализации решения:
  - Метод дихотомии
  - Метод золотого сечения
- Неточный линейный поиск:
  - Условие достаточного убывания
  - Условия Гольдштейна
  - Условие ограничения на кривизну
  - Идея заключается в использовании бэктрекинга для нахождения шага, удовлетворяющего условию Армихо.

# **Задачи на матрично-векторное дифференцирование**

# Матрично-векторное дифференцирование. Задача 1



## Example

Найдите  $\nabla f(x)$ , если  $f(x) = \frac{1}{2}x^T Ax + b^T x + c$ .

# Матрично-векторное дифференцирование. Задача 2



## Example

Найдите  $\nabla f(X)$ , если  $f(X) = \text{tr}(AX^{-1}B)$

# Матрично-векторное дифференцирование. Задача 3



## Example

Найдите градиент  $\nabla f(x)$  и гессиан  $\nabla^2 f(x)$ , если  $f(x) = \frac{1}{3} \|x\|_2^3$

# Примеры линейного поиска

# Линейный поиск. Пример 1: Сравнение методов (Colab)

clubs



$$f_1(x) = x(x - 2)(x + 2)^2 + 10$$

$$[a, b] = [-3, 2]$$

Случайный поиск: 72 вызова функции. 36 итераций.  $f_1^* = 0.09$

Метод дихотомии: 23 вызова функции. 13 итераций.  $f_1^* = 10.00$

Золотое сечение: 19 вызова функции. 18 итераций.  $f_1^* = 10.00$

Параболический поиск: 20 вызова функции. 17 итераций.  $f_1^* = 10.00$



Рисунок 1. Сравнение различных методов линейного поиска с  $f_1$

# Линейный поиск. Пример 2: Сравнение методов (Colab)

clubs



$$f_2(x) = -\sqrt{\frac{2}{\pi}} \frac{x^2 e^{-\frac{x^2}{8}}}{8}$$
$$[a, b] = [0, 6]$$

Случайный поиск: 68 вызова функции. 34 итераций.  $f_2^* = 0.71$

Метод дихотомии: 23 вызова функции. 13 итераций.  $f_2^* = 0.71$

Золотое сечение: 20 вызова функции. 19 итераций.  $f_2^* = 0.71$

Параболический поиск: 17 вызова функции. 14 итераций.  $f_2^* = 0.71$



Рисунок 2. Сравнение различных методов линейного поиска с  $f_2$

# Линейный поиск. Пример 3: Сравнение методов (Colab)

clubs



$$f_3(x) = \sin \left( \sin \left( \sin \left( \sqrt{\frac{x}{2}} \right) \right) \right)$$
$$[a, b] = [5, 70]$$

Random search: 66 function calls. 33 iterations.  $f_3^* = 0.25$   
Метод дихотомии: 32 вызова функции. 17 итераций.  $f_3^* = 0.25$   
Золотое сечение: 25 вызова функции. 24 итераций.  $f_3^* = 0.25$   
Параболический поиск: 103 вызова функции. 100 итераций.  $f_3^* = 0.25$

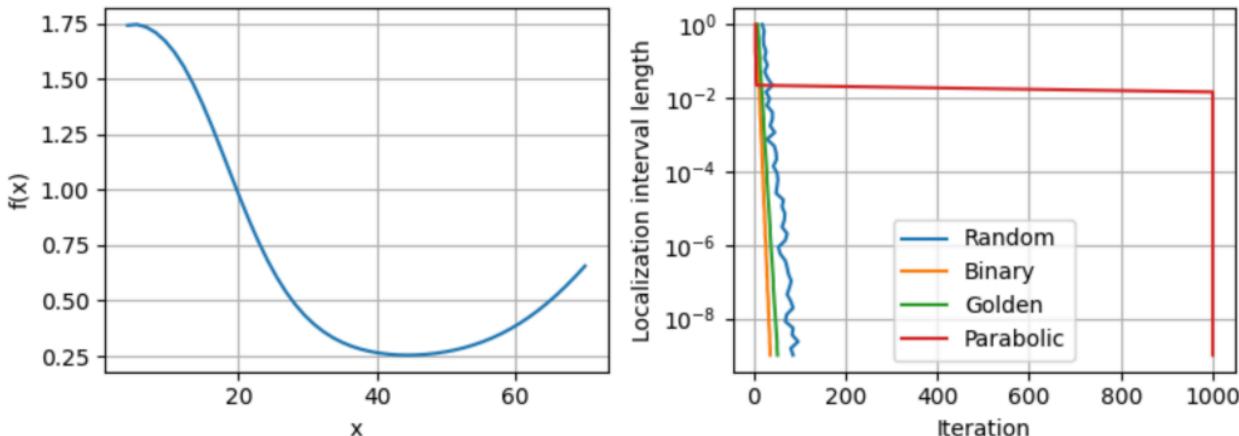


Рисунок 3. Сравнение различных методов линейного поиска с  $f_3$

# Линейный поиск. Пример 4: Метод Брента

- Параболическая интерполяция + Золотое сечение = Метод Брента



Рисунок 4. Идея метода Брента

# Линейный поиск. Пример 4: Метод Брента

- Параболическая интерполяция + Золотое сечение = Метод Брента
- Основная идея метода заключается в отслеживании значения оптимизируемой скалярной функции в шести точках  $a, b, x, w, v, u$



Рисунок 4. Идея метода Брента

# Линейный поиск. Пример 4: Метод Брента

- Параболическая интерполяция + Золотое сечение = Метод Брента
- Основная идея метода заключается в отслеживании значения оптимизируемой скалярной функции в шести точках  $a, b, x, w, v, u$
- $[a, b]$  — интервал локализации в текущей итерации



Рисунок 4. Идея метода Брента

# Линейный поиск. Пример 4: Метод Брента

- Параболическая интерполяция + Золотое сечение = Метод Брента
- Основная идея метода заключается в отслеживании значений оптимизируемой скалярной функции в шести точках  $a, b, x, w, v, u$
- $[a, b]$  — интервал локализации в текущей итерации
- Точки  $x, w$  и  $v$  такие, что выполняется неравенство  $f(x) \leq f(w) \leq f(v)$



Рисунок 4. Идея метода Брента

# Линейный поиск. Пример 4: Метод Брента

- Параболическая интерполяция + Золотое сечение = Метод Брента
- Основная идея метода заключается в отслеживании значения оптимизируемой скалярной функции в шести точках  $a, b, x, w, v, u$
- $[a, b]$  — интервал локализации в текущей итерации
- Точки  $x, w$  и  $v$  такие, что выполняется неравенство  $f(x) \leq f(w) \leq f(v)$
- $u$  — минимум параболы, построенной на точках  $x, w$  и  $v$ , или точка золотого сечения наибольшего из отрезков  $[a, x]$  и  $[x, b]$ .



Рисунок 4. Идея метода Брента

# Линейный поиск. Пример 5: Метод Брента

Парабола строится только если точки  $x$ ,  $w$  и  $v$  различны, и ее вершина  $u^*$  берется как точка  $u$  только если

- $u^* \in [a, b]$



Рисунок 5. Пример работы метода Брента

# Линейный поиск. Пример 5: Метод Брента

Парабола строится только если точки  $x$ ,  $w$  и  $v$  различны, и ее вершина  $u^*$  берется как точка  $u$  только если

- $u^* \in [a, b]$
- $u^*$  не более половины длины шага, предшествующего предыдущему, от точки  $x$



Рисунок 5. Пример работы метода Брента

# Линейный поиск. Пример 5: Метод Брента

Парабола строится только если точки  $x$ ,  $w$  и  $v$  различны, и ее вершина  $u^*$  берется как точка  $u$  только если

- $u^* \in [a, b]$
- $u^*$  не более половины длины шага, предшествующего предыдущему, от точки  $x$
- Если условия выше не выполняются, то точка  $u$  находится из золотого сечения



Рисунок 5. Пример работы метода Брента

# Линейный поиск. Пример 5: Метод Брента

Парабола строится только если точки  $x$ ,  $w$  и  $v$  различны, и ее вершина  $u^*$  берется как точка  $u$  только если

- $u^* \in [a, b]$
- $u^*$  не более половины длины шага, предшествующего предыдущему, от точки  $x$
- Если условия выше не выполняются, то точка  $u$  находится из золотого сечения
- Пример в Colab ♣

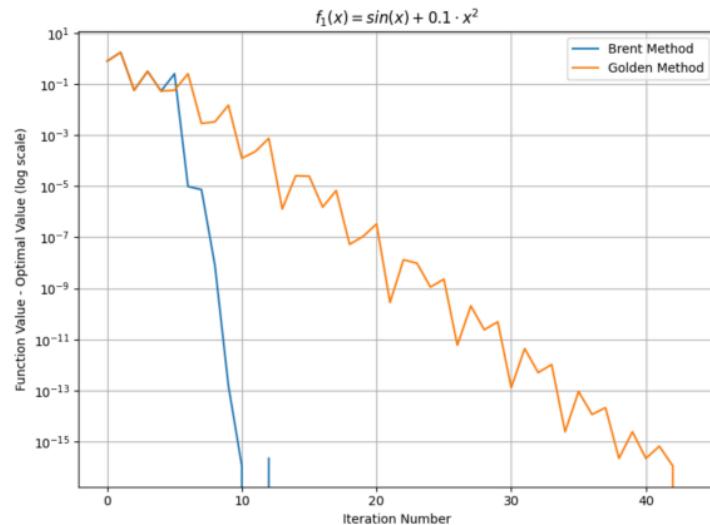


Рисунок 5. Пример работы метода Брента