A black and white Corgi dog and a yellow rubber duck are sliding down a curved, white ramp. The dog is on the left, lying on its back with its paws up. The duck is on the right, also lying on its back. The ramp is curved and has some motion lines around it, suggesting speed. The background is a light gray.

**Классификация и обозначения в задачах  
оптимизации. Скорость сходимости.  
Линейный поиск. Неточная одномерная  
оптимизация.**

**Даниил Меркулов**

Методы оптимизации. МФТИ

## Классификации скорости сходимости

## Классификации скорости сходимости

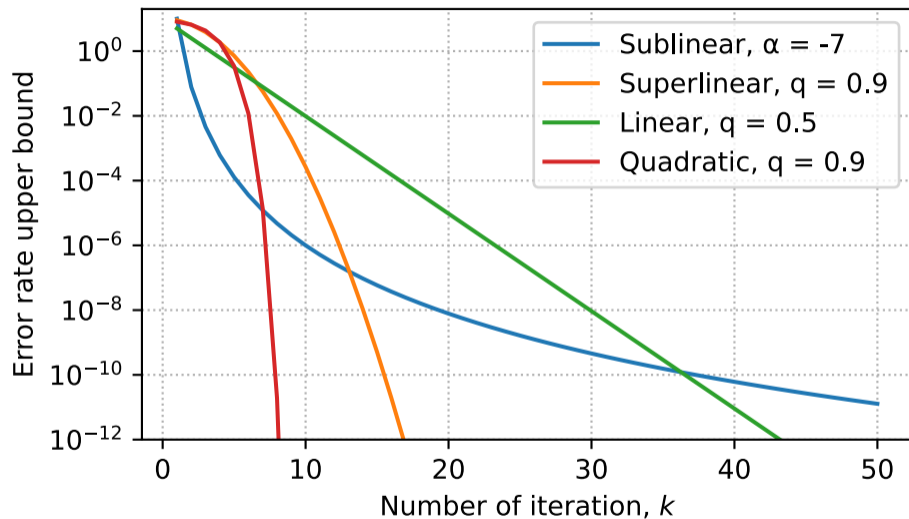


Рис. 1: Разные типы скорости сходимости

# Линейная сходимость

Чтобы сравнить производительность алгоритмов, нам нужно определить терминологию для различных типов сходимости.

Пусть  $r_k = \{\|x_k - x^*\|_2\}$  — это последовательность в  $\mathbb{R}^n$ , сходящаяся к нулю.

# Линейная сходимость

Чтобы сравнить производительность алгоритмов, нам нужно определить терминологию для различных типов сходимости.

Пусть  $r_k = \{\|x_k - x^*\|_2\}$  — это последовательность в  $\mathbb{R}^n$ , сходящаяся к нулю.

Мы можем определить *линейную* сходимость в двух формах:

$$\|x_{k+1} - x^*\|_2 \leq Cq^k \quad \text{или} \quad \|x_{k+1} - x^*\|_2 \leq q\|x_k - x^*\|_2,$$

# Линейная сходимость

Чтобы сравнить производительность алгоритмов, нам нужно определить терминологию для различных типов сходимости.

Пусть  $r_k = \{\|x_k - x^*\|_2\}$  — это последовательность в  $\mathbb{R}^n$ , сходящаяся к нулю.

Мы можем определить *линейную* сходимость в двух формах:

$$\|x_{k+1} - x^*\|_2 \leq Cq^k \quad \text{или} \quad \|x_{k+1} - x^*\|_2 \leq q\|x_k - x^*\|_2,$$

для всех достаточно больших  $k$ . Здесь  $q \in (0, 1)$  и  $0 < C < \infty$ . Это означает, что расстояние до решения  $x^*$  уменьшается на каждой итерации как минимум на некоторый постоянный множитель. Обратите внимание, что иногда такой тип сходимости также называют *экспоненциальной* или *геометрической*. Величина  $q$  называется скоростью сходимости.

# Линейная сходимость

Чтобы сравнить производительность алгоритмов, нам нужно определить терминологию для различных типов сходимости.

Пусть  $r_k = \{\|x_k - x^*\|_2\}$  — это последовательность в  $\mathbb{R}^n$ , сходящаяся к нулю.

Мы можем определить *линейную* сходимость в двух формах:

$$\|x_{k+1} - x^*\|_2 \leq Cq^k \quad \text{или} \quad \|x_{k+1} - x^*\|_2 \leq q\|x_k - x^*\|_2,$$

для всех достаточно больших  $k$ . Здесь  $q \in (0, 1)$  и  $0 < C < \infty$ . Это означает, что расстояние до решения  $x^*$  уменьшается на каждой итерации как минимум на некоторый постоянный множитель. Обратите внимание, что иногда такой тип сходимости также называют *экспоненциальной* или *геометрической*. Величина  $q$  называется скоростью сходимости.

## i Question

Предположим, у вас есть две последовательности с линейными скоростями сходимости  $q_1 = 0.1$  и  $q_2 = 0.7$ . Какая из них сходится быстрее?

# Линейная сходимость

## i Example

Рассмотрим следующую последовательность:

$$r_k = \frac{1}{2^k}$$

Нетрудно заметить, что здесь наблюдается линейная сходимость с параметрами  $q = \frac{1}{2}$  и  $C = 1$ .

# Линейная сходимость

## i Example

Рассмотрим следующую последовательность:

$$r_k = \frac{1}{2^k}$$

Нетрудно заметить, что здесь наблюдается линейная сходимость с параметрами  $q = \frac{1}{2}$  и  $C = 1$ .

## i Question

Определите сходимость следующей последовательности:

$$r_k = \frac{3}{2^k}$$

# Суб- и сверхлинейная сходимость

## Сублинейная сходимость

Если последовательность  $r_k$  сходится к нулю, но не обладает линейной сходимостью, то такая сходимость называется сублинейной. Иногда можно рассматривать следующий класс сублинейной сходимости:

$$\|x_{k+1} - x^*\|_2 \leq Ck^q,$$

где  $q < 0$  и  $0 < C < \infty$ . Заметьте, что сублинейная сходимость означает, что последовательность сходится медленнее, чем любая геометрическая прогрессия.

## Сверхлинейная сходимость

Сходимость называется *сверхлинейной*, если она происходит быстрее, чем любая линейно сходящаяся последовательность.

## Тест корней

### i Theorem

Пусть  $(r_k)_{k=m}^{\infty}$  — последовательность неотрицательных чисел, стремящаяся к нулю, и положим  $\alpha := \limsup_{k \rightarrow \infty} r_k^{1/k}$ . (Заметим, что  $\alpha \geq 0$ .)

(a) Если  $0 \leq \alpha < 1$ , то  $(r_k)_{k=m}^{\infty}$  сходится линейно с константой  $\alpha$ .

**Доказательство.**

## Тест корней

### i Theorem

Пусть  $(r_k)_{k=m}^{\infty}$  — последовательность неотрицательных чисел, стремящаяся к нулю, и положим  $\alpha := \limsup_{k \rightarrow \infty} r_k^{1/k}$ . (Заметим, что  $\alpha \geq 0$ .)

- (a) Если  $0 \leq \alpha < 1$ , то  $(r_k)_{k=m}^{\infty}$  сходится линейно с константой  $\alpha$ .
- (b) В частности, если  $\alpha = 0$ , то  $(r_k)_{k=m}^{\infty}$  сходится сверхлинейно.

**Доказательство.**

## Тест корней

### i Theorem

Пусть  $(r_k)_{k=m}^{\infty}$  — последовательность неотрицательных чисел, стремящаяся к нулю, и положим  $\alpha := \limsup_{k \rightarrow \infty} r_k^{1/k}$ . (Заметим, что  $\alpha \geq 0$ .)

- (a) Если  $0 \leq \alpha < 1$ , то  $(r_k)_{k=m}^{\infty}$  сходится линейно с константой  $\alpha$ .
- (b) В частности, если  $\alpha = 0$ , то  $(r_k)_{k=m}^{\infty}$  сходится сверхлинейно.
- (c) Если  $\alpha = 1$ , то  $(r_k)_{k=m}^{\infty}$  сходится сублинейно.

**Доказательство.**

## Тест корней

### i Theorem

Пусть  $(r_k)_{k=m}^{\infty}$  — последовательность неотрицательных чисел, стремящаяся к нулю, и положим  $\alpha := \limsup_{k \rightarrow \infty} r_k^{1/k}$ . (Заметим, что  $\alpha \geq 0$ .)

- (a) Если  $0 \leq \alpha < 1$ , то  $(r_k)_{k=m}^{\infty}$  сходится линейно с константой  $\alpha$ .
- (b) В частности, если  $\alpha = 0$ , то  $(r_k)_{k=m}^{\infty}$  сходится сверхлинейно.
- (c) Если  $\alpha = 1$ , то  $(r_k)_{k=m}^{\infty}$  сходится сублинейно.
- (d) Случай  $\alpha > 1$  невозможен.

**Доказательство.**

## Тест корней

### i Theorem

Пусть  $(r_k)_{k=m}^{\infty}$  — последовательность неотрицательных чисел, стремящаяся к нулю, и положим  $\alpha := \limsup_{k \rightarrow \infty} r_k^{1/k}$ . (Заметим, что  $\alpha \geq 0$ .)

- (a) Если  $0 \leq \alpha < 1$ , то  $(r_k)_{k=m}^{\infty}$  сходится линейно с константой  $\alpha$ .
- (b) В частности, если  $\alpha = 0$ , то  $(r_k)_{k=m}^{\infty}$  сходится сверхлинейно.
- (c) Если  $\alpha = 1$ , то  $(r_k)_{k=m}^{\infty}$  сходится сублинейно.
- (d) Случай  $\alpha > 1$  невозможен.

### Доказательство.

1. Покажем, что если  $(r_k)_{k=m}^{\infty}$  сходится линейно с константой  $0 \leq \beta < 1$ , то обязательно  $\alpha \leq \beta$ .

Действительно, по определению константы линейной сходимости, для любого  $\varepsilon > 0$ , такого что  $\beta + \varepsilon < 1$ , существует  $C > 0$  такое, что  $r_k \leq C(\beta + \varepsilon)^k$  для всех  $k \geq m$ .

Отсюда  $r_k^{1/k} \leq C^{1/k}(\beta + \varepsilon)$  для всех  $k \geq m$ . Пройдя к пределу при  $k \rightarrow \infty$  и учитывая, что  $C^{1/k} \rightarrow 1$ , получаем  $\alpha \leq \beta + \varepsilon$ . Учитывая произвольность  $\varepsilon$ , заключаем, что  $\alpha \leq \beta$ .

## Тест корней

### i Theorem

Пусть  $(r_k)_{k=m}^{\infty}$  — последовательность неотрицательных чисел, стремящаяся к нулю, и положим  $\alpha := \limsup_{k \rightarrow \infty} r_k^{1/k}$ . (Заметим, что  $\alpha \geq 0$ .)

- (a) Если  $0 \leq \alpha < 1$ , то  $(r_k)_{k=m}^{\infty}$  сходится линейно с константой  $\alpha$ .
- (b) В частности, если  $\alpha = 0$ , то  $(r_k)_{k=m}^{\infty}$  сходится сверхлинейно.
- (c) Если  $\alpha = 1$ , то  $(r_k)_{k=m}^{\infty}$  сходится сублинейно.
- (d) Случай  $\alpha > 1$  невозможен.

### Доказательство.

1. Покажем, что если  $(r_k)_{k=m}^{\infty}$  сходится линейно с константой  $0 \leq \beta < 1$ , то обязательно  $\alpha \leq \beta$ .

Действительно, по определению константы линейной сходимости, для любого  $\varepsilon > 0$ , такого что  $\beta + \varepsilon < 1$ , существует  $C > 0$  такое, что  $r_k \leq C(\beta + \varepsilon)^k$  для всех  $k \geq m$ .

Отсюда  $r_k^{1/k} \leq C^{1/k}(\beta + \varepsilon)$  для всех  $k \geq m$ . Пройдя к пределу при  $k \rightarrow \infty$  и учитывая, что  $C^{1/k} \rightarrow 1$ , получаем  $\alpha \leq \beta + \varepsilon$ . Учитывая произвольность  $\varepsilon$ , заключаем, что  $\alpha \leq \beta$ .

2. Следовательно, в случае  $\alpha = 1$  последовательность  $(r_k)_{k=m}^{\infty}$  не может обладать линейной сходимостью по приведённому выше результату (доказанному от противного). Поскольку, тем не менее,  $(r_k)_{k=m}^{\infty}$  стремится к нулю, она должна сходиться сублинейно.

## Тест корней

3. Рассмотрим теперь случай  $0 \leq \alpha < 1$ . Пусть  $\varepsilon > 0$  — произвольное число такое, что  $\alpha + \varepsilon < 1$ .

## Тест корней

3. Рассмотрим теперь случай  $0 \leq \alpha < 1$ . Пусть  $\varepsilon > 0$  — произвольное число такое, что  $\alpha + \varepsilon < 1$ .

## Тест корней

3. Рассмотрим теперь случай  $0 \leq \alpha < 1$ . Пусть  $\varepsilon > 0$  — произвольное число такое, что  $\alpha + \varepsilon < 1$ .

В соответствии со свойствами верхнего предела (limsup), существует  $N \geq m$  такое, что  $r_k^{1/k} \leq \alpha + \varepsilon$  для всех  $k \geq N$ .

## Тест корней

3. Рассмотрим теперь случай  $0 \leq \alpha < 1$ . Пусть  $\varepsilon > 0$  — произвольное число такое, что  $\alpha + \varepsilon < 1$ .

В соответствии со свойствами верхнего предела (limsup), существует  $N \geq m$  такое, что  $r_k^{1/k} \leq \alpha + \varepsilon$  для всех  $k \geq N$ .

Отсюда  $r_k \leq (\alpha + \varepsilon)^k$  для всех  $k \geq N$ . Следовательно, последовательность  $(r_k)_{k=m}^{\infty}$  сходится линейно с параметром  $\alpha + \varepsilon$  (не имеет значения, что неравенство верно только начиная с номера  $N$ ).

## Тест корней

3. Рассмотрим теперь случай  $0 \leq \alpha < 1$ . Пусть  $\varepsilon > 0$  — произвольное число такое, что  $\alpha + \varepsilon < 1$ .

В соответствии со свойствами верхнего предела (limsup), существует  $N \geq m$  такое, что  $r_k^{1/k} \leq \alpha + \varepsilon$  для всех  $k \geq N$ .

Отсюда  $r_k \leq (\alpha + \varepsilon)^k$  для всех  $k \geq N$ . Следовательно, последовательность  $(r_k)_{k=m}^{\infty}$  сходится линейно с параметром  $\alpha + \varepsilon$  (не имеет значения, что неравенство верно только начиная с номера  $N$ ).

Ввиду произвольности  $\varepsilon$  это означает, что константа линейной сходимости последовательности  $(r_k)_{k=m}^{\infty}$  не превосходит  $\alpha$ .

## Тест корней

3. Рассмотрим теперь случай  $0 \leq \alpha < 1$ . Пусть  $\varepsilon > 0$  — произвольное число такое, что  $\alpha + \varepsilon < 1$ .

В соответствии со свойствами верхнего предела (limsup), существует  $N \geq m$  такое, что  $r_k^{1/k} \leq \alpha + \varepsilon$  для всех  $k \geq N$ .

Отсюда  $r_k \leq (\alpha + \varepsilon)^k$  для всех  $k \geq N$ . Следовательно, последовательность  $(r_k)_{k=m}^{\infty}$  сходится линейно с параметром  $\alpha + \varepsilon$  (не имеет значения, что неравенство верно только начиная с номера  $N$ ).

Ввиду произвольности  $\varepsilon$  это означает, что константа линейной сходимости последовательности  $(r_k)_{k=m}^{\infty}$  не превосходит  $\alpha$ .

Поскольку, как было показано выше, константа линейной сходимости не может быть меньше  $\alpha$ , следует, что константа линейной сходимости последовательности  $(r_k)_{k=m}^{\infty}$  равна ровно  $\alpha$ .

1. Наконец, покажем, что случай  $\alpha > 1$  невозможен.

## Тест корней

3. Рассмотрим теперь случай  $0 \leq \alpha < 1$ . Пусть  $\varepsilon > 0$  — произвольное число такое, что  $\alpha + \varepsilon < 1$ .

В соответствии со свойствами верхнего предела (limsup), существует  $N \geq m$  такое, что  $r_k^{1/k} \leq \alpha + \varepsilon$  для всех  $k \geq N$ .

Отсюда  $r_k \leq (\alpha + \varepsilon)^k$  для всех  $k \geq N$ . Следовательно, последовательность  $(r_k)_{k=m}^{\infty}$  сходится линейно с параметром  $\alpha + \varepsilon$  (не имеет значения, что неравенство верно только начиная с номера  $N$ ).

Ввиду произвольности  $\varepsilon$  это означает, что константа линейной сходимости последовательности  $(r_k)_{k=m}^{\infty}$  не превосходит  $\alpha$ .

Поскольку, как было показано выше, константа линейной сходимости не может быть меньше  $\alpha$ , следует, что константа линейной сходимости последовательности  $(r_k)_{k=m}^{\infty}$  равна ровно  $\alpha$ .

1. Наконец, покажем, что случай  $\alpha > 1$  невозможен.

## Тест корней

3. Рассмотрим теперь случай  $0 \leq \alpha < 1$ . Пусть  $\varepsilon > 0$  — произвольное число такое, что  $\alpha + \varepsilon < 1$ .

В соответствии со свойствами верхнего предела ( $\limsup$ ), существует  $N \geq m$  такое, что  $r_k^{1/k} \leq \alpha + \varepsilon$  для всех  $k \geq N$ .

Отсюда  $r_k \leq (\alpha + \varepsilon)^k$  для всех  $k \geq N$ . Следовательно, последовательность  $(r_k)_{k=m}^{\infty}$  сходится линейно с параметром  $\alpha + \varepsilon$  (не имеет значения, что неравенство верно только начиная с номера  $N$ ).

Ввиду произвольности  $\varepsilon$  это означает, что константа линейной сходимости последовательности  $(r_k)_{k=m}^{\infty}$  не превосходит  $\alpha$ .

Поскольку, как было показано выше, константа линейной сходимости не может быть меньше  $\alpha$ , следует, что константа линейной сходимости последовательности  $(r_k)_{k=m}^{\infty}$  равна ровно  $\alpha$ .

1. Наконец, покажем, что случай  $\alpha > 1$  невозможен.

Действительно, пусть  $\alpha > 1$ . Тогда из определения верхнего предела ( $\limsup$ ) следует, что для любого  $N \geq m$  найдётся  $k \geq N$  такое, что  $r_k^{1/k} \geq 1$ , и, в частности,  $r_k \geq 1$ .

## Тест корней

3. Рассмотрим теперь случай  $0 \leq \alpha < 1$ . Пусть  $\varepsilon > 0$  — произвольное число такое, что  $\alpha + \varepsilon < 1$ .

В соответствии со свойствами верхнего предела ( $\limsup$ ), существует  $N \geq m$  такое, что  $r_k^{1/k} \leq \alpha + \varepsilon$  для всех  $k \geq N$ .

Отсюда  $r_k \leq (\alpha + \varepsilon)^k$  для всех  $k \geq N$ . Следовательно, последовательность  $(r_k)_{k=m}^{\infty}$  сходится линейно с параметром  $\alpha + \varepsilon$  (не имеет значения, что неравенство верно только начиная с номера  $N$ ).

Ввиду произвольности  $\varepsilon$  это означает, что константа линейной сходимости последовательности  $(r_k)_{k=m}^{\infty}$  не превосходит  $\alpha$ .

Поскольку, как было показано выше, константа линейной сходимости не может быть меньше  $\alpha$ , следует, что константа линейной сходимости последовательности  $(r_k)_{k=m}^{\infty}$  равна ровно  $\alpha$ .

1. Наконец, покажем, что случай  $\alpha > 1$  невозможен.

Действительно, пусть  $\alpha > 1$ . Тогда из определения верхнего предела ( $\limsup$ ) следует, что для любого  $N \geq m$  найдётся  $k \geq N$  такое, что  $r_k^{1/k} \geq 1$ , и, в частности,  $r_k \geq 1$ .

Но это означает, что у  $r_k$  есть подпоследовательность, отстоящая от нуля на положительную величину. Следовательно,  $(r_k)_{k=m}^{\infty}$  не может стремиться к нулю, что противоречит условию.

## Тест отношений

Пусть  $\{r_k\}_{k=m}^{\infty}$  — последовательность строго положительных чисел, стремящаяся к нулю. Пусть

$$q = \lim_{k \rightarrow \infty} \frac{r_{k+1}}{r_k}$$

- Если существует  $q$  и  $0 \leq q < 1$ , то  $\{r_k\}_{k=m}^{\infty}$  сходится линейно с константой  $q$ .

## Тест отношений

Пусть  $\{r_k\}_{k=m}^{\infty}$  — последовательность строго положительных чисел, стремящаяся к нулю. Пусть

$$q = \lim_{k \rightarrow \infty} \frac{r_{k+1}}{r_k}$$

- Если существует  $q$  и  $0 \leq q < 1$ , то  $\{r_k\}_{k=m}^{\infty}$  сходится линейно с константой  $q$ .
- В частности, если  $q = 0$ , то  $\{r_k\}_{k=m}^{\infty}$  сходится сверхлинейно.

## Тест отношений

Пусть  $\{r_k\}_{k=m}^{\infty}$  — последовательность строго положительных чисел, стремящаяся к нулю. Пусть

$$q = \lim_{k \rightarrow \infty} \frac{r_{k+1}}{r_k}$$

- Если существует  $q$  и  $0 \leq q < 1$ , то  $\{r_k\}_{k=m}^{\infty}$  сходится линейно с константой  $q$ .
- В частности, если  $q = 0$ , то  $\{r_k\}_{k=m}^{\infty}$  сходится сверхлинейно.
- Если  $q$  не существует, но  $q = \lim_{k \rightarrow \infty} \sup_k \frac{r_{k+1}}{r_k} < 1$ , то  $\{r_k\}_{k=m}^{\infty}$  сходится линейно с константой, не превышающей  $q$ .

## Тест отношений

Пусть  $\{r_k\}_{k=m}^{\infty}$  — последовательность строго положительных чисел, стремящаяся к нулю. Пусть

$$q = \lim_{k \rightarrow \infty} \frac{r_{k+1}}{r_k}$$

- Если существует  $q$  и  $0 \leq q < 1$ , то  $\{r_k\}_{k=m}^{\infty}$  сходится линейно с константой  $q$ .
- В частности, если  $q = 0$ , то  $\{r_k\}_{k=m}^{\infty}$  сходится сверхлинейно.
- Если  $q$  не существует, но  $q = \lim_{k \rightarrow \infty} \sup_k \frac{r_{k+1}}{r_k} < 1$ , то  $\{r_k\}_{k=m}^{\infty}$  сходится линейно с константой, не превышающей  $q$ .
- Если  $\lim_{k \rightarrow \infty} \inf_k \frac{r_{k+1}}{r_k} = 1$ , то  $\{r_k\}_{k=m}^{\infty}$  сходится сублинейно.

## Тест отношений

Пусть  $\{r_k\}_{k=m}^{\infty}$  — последовательность строго положительных чисел, стремящаяся к нулю. Пусть

$$q = \lim_{k \rightarrow \infty} \frac{r_{k+1}}{r_k}$$

- Если существует  $q$  и  $0 \leq q < 1$ , то  $\{r_k\}_{k=m}^{\infty}$  сходится линейно с константой  $q$ .
- В частности, если  $q = 0$ , то  $\{r_k\}_{k=m}^{\infty}$  сходится сверхлинейно.
- Если  $q$  не существует, но  $q = \lim_{k \rightarrow \infty} \sup_k \frac{r_{k+1}}{r_k} < 1$ , то  $\{r_k\}_{k=m}^{\infty}$  сходится линейно с константой, не превышающей  $q$ .
- Если  $\lim_{k \rightarrow \infty} \inf_k \frac{r_{k+1}}{r_k} = 1$ , то  $\{r_k\}_{k=m}^{\infty}$  сходится сублинейно.
- Случай  $\lim_{k \rightarrow \infty} \inf_k \frac{r_{k+1}}{r_k} > 1$  невозможен.

## Тест отношений

Пусть  $\{r_k\}_{k=m}^{\infty}$  — последовательность строго положительных чисел, стремящаяся к нулю. Пусть

$$q = \lim_{k \rightarrow \infty} \frac{r_{k+1}}{r_k}$$

- Если существует  $q$  и  $0 \leq q < 1$ , то  $\{r_k\}_{k=m}^{\infty}$  сходится линейно с константой  $q$ .
- В частности, если  $q = 0$ , то  $\{r_k\}_{k=m}^{\infty}$  сходится сверхлинейно.
- Если  $q$  не существует, но  $q = \lim_{k \rightarrow \infty} \sup_k \frac{r_{k+1}}{r_k} < 1$ , то  $\{r_k\}_{k=m}^{\infty}$  сходится линейно с константой, не превышающей  $q$ .
- Если  $\lim_{k \rightarrow \infty} \inf_k \frac{r_{k+1}}{r_k} = 1$ , то  $\{r_k\}_{k=m}^{\infty}$  сходится сублинейно.
- Случай  $\lim_{k \rightarrow \infty} \inf_k \frac{r_{k+1}}{r_k} > 1$  невозможен.
- Во всех остальных случаях (то есть когда  $\lim_{k \rightarrow \infty} \inf_k \frac{r_{k+1}}{r_k} < 1 \leq \lim_{k \rightarrow \infty} \sup_k \frac{r_{k+1}}{r_k}$ ) мы не можем утверждать ничего конкретного про скорость сходимости  $\{r_k\}_{k=m}^{\infty}$ .

## Тест отношений (лемма)

### i Theorem

Пусть  $(r_k)_{k=m}^{\infty}$  — последовательность строго положительных чисел. (Строгое положительное значение необходимо, чтобы отношения  $\frac{r_{k+1}}{r_k}$ , которые встречаются ниже, были корректно определены.) Тогда

$$\liminf_{k \rightarrow \infty} \frac{r_{k+1}}{r_k} \leq \liminf_{k \rightarrow \infty} r_k^{1/k} \leq \limsup_{k \rightarrow \infty} r_k^{1/k} \leq \limsup_{k \rightarrow \infty} \frac{r_{k+1}}{r_k}.$$

## Тест отношений (лемма)

### i Theorem

Пусть  $(r_k)_{k=m}^{\infty}$  — последовательность строго положительных чисел. (Строгое положительное значение необходимо, чтобы отношения  $\frac{r_{k+1}}{r_k}$ , которые встречаются ниже, были корректно определены.) Тогда

$$\liminf_{k \rightarrow \infty} \frac{r_{k+1}}{r_k} \leq \liminf_{k \rightarrow \infty} r_k^{1/k} \leq \limsup_{k \rightarrow \infty} r_k^{1/k} \leq \limsup_{k \rightarrow \infty} \frac{r_{k+1}}{r_k}.$$

### Доказательство.

1. Среднее неравенство следует из того, что  $\liminf$  любой последовательности всегда не больше её  $\limsup$ . Докажем последнее неравенство; первое доказывается аналогично.

## Тест отношений (лемма)

### i Theorem

Пусть  $(r_k)_{k=m}^{\infty}$  — последовательность строго положительных чисел. (Строгое положительное значение необходимо, чтобы отношения  $\frac{r_{k+1}}{r_k}$ , которые встречаются ниже, были корректно определены.) Тогда

$$\liminf_{k \rightarrow \infty} \frac{r_{k+1}}{r_k} \leq \liminf_{k \rightarrow \infty} r_k^{1/k} \leq \limsup_{k \rightarrow \infty} r_k^{1/k} \leq \limsup_{k \rightarrow \infty} \frac{r_{k+1}}{r_k}.$$

### Доказательство.

1. Среднее неравенство следует из того, что  $\liminf$  любой последовательности всегда не больше её  $\limsup$ . Докажем последнее неравенство; первое доказывается аналогично.
2. Обозначим  $L := \limsup_{k \rightarrow \infty} \frac{r_{k+1}}{r_k}$ . Если  $L = +\infty$ , то неравенство очевидно, поэтому предположим, что  $L$  конечно. Заметим, что  $L \geq 0$ , поскольку отношение  $\frac{r_{k+1}}{r_k}$  положительно для всех  $k \geq m$ . Пусть  $\varepsilon > 0$  — произвольное число. По свойствам верхнего предела существует  $N \geq m$  такое, что  $\frac{r_{k+1}}{r_k} \leq L + \varepsilon$  для всех  $k \geq N$ . Отсюда  $r_{k+1} \leq (L + \varepsilon)r_k$  для всех  $k \geq N$ . Применяя индукцию, получаем  $r_k \leq (L + \varepsilon)^{k-N} r_N$  для всех  $k \geq N$ . Положим  $C := (L + \varepsilon)^{-N} r_N$ . Тогда  $r_k \leq C(L + \varepsilon)^k$  для всех  $k \geq N$ , откуда  $r_k^{1/k} \leq C^{1/k}(L + \varepsilon)$ . Взяв  $\limsup$  при  $k \rightarrow \infty$  и учитывая, что  $C^{1/k} \rightarrow 1$ , получаем  $\limsup_{k \rightarrow \infty} r_k^{1/k} \leq L + \varepsilon$ . Учитывая произвольность  $\varepsilon$ , заключаем, что  $\limsup_{k \rightarrow \infty} r_k^{1/k} \leq L$ .

## Линейный поиск

## Задача

Предположим, у нас есть задача минимизации функции  $f(x) : \mathbb{R} \rightarrow \mathbb{R}$  от вещественной переменной:

$$f(x) \rightarrow \min_{x \in \mathbb{R}}$$

## Задача

Предположим, у нас есть задача минимизации функции  $f(x) : \mathbb{R} \rightarrow \mathbb{R}$  от вещественной переменной:

$$f(x) \rightarrow \min_{x \in \mathbb{R}}$$

Иногда мы рассматриваем похожую задачу поиска минимума на отрезке  $[a, b]$ :  $f(x) \rightarrow \min_{x \in [a, b]}$

## Задача

Предположим, у нас есть задача минимизации функции  $f(x) : \mathbb{R} \rightarrow \mathbb{R}$  от вещественной переменной:

$$f(x) \rightarrow \min_{x \in \mathbb{R}}$$

Иногда мы рассматриваем похожую задачу поиска минимума на отрезке  $[a, b]$ :  $f(x) \rightarrow \min_{x \in [a, b]}$

### Example

Типичный пример задачи одномерного поиска — выбор подходящего шага для алгоритма градиентного спуска:

$$\begin{aligned} x_{k+1} &= x_k - \alpha \nabla f(x_k) \\ \alpha &= \operatorname{argmin} f(x_{k+1}) \end{aligned}$$

Одномерный поиск — важная задача оптимизации, часто играющая роль при решении более сложных задач. Чтобы упростить постановку, предположим, что функция  $f(x)$  *унимодальна*, то есть убывает до какой-то точки (не включительно), а затем возрастает или наоборот.

# Унимодальная функция

## i Definition

Функция  $f(x)$  называется **унимодальной** на  $[a, b]$ , если существует  $x_* \in [a, b]$  такое, что  $f(x_1) > f(x_2) \quad \forall a \leq x_1 < x_2 < x_*$  и  $f(x_1) < f(x_2) \quad \forall x_* < x_1 < x_2 \leq b$ .

# Унимодальная функция

## i Definition

Функция  $f(x)$  называется **унимодальной** на  $[a, b]$ , если существует  $x_* \in [a, b]$  такое, что  $f(x_1) > f(x_2) \quad \forall a \leq x_1 < x_2 < x_*$  и  $f(x_1) < f(x_2) \quad \forall x_* < x_1 < x_2 \leq b$ .

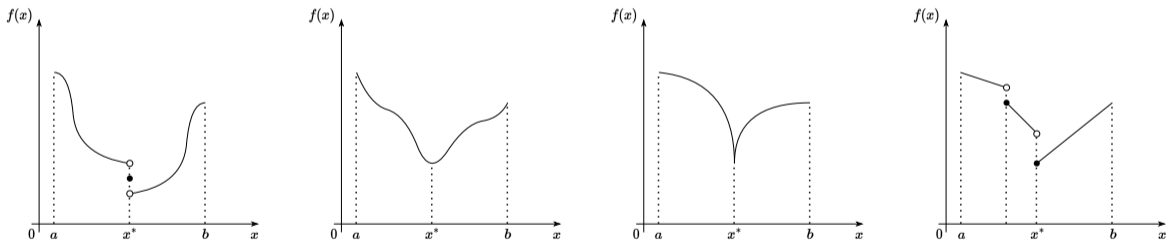


Рис. 2: Примеры унимодальных функций

## Ключевое свойство унимодальных функций

Пусть  $f(x)$  — унимодальная функция на  $[a, b]$ . Тогда если  $x_1 < x_2 \in [a, b]$ , то:

- если  $f(x_1) \leq f(x_2)$ , то  $x_* \in [a, x_2]$ ;

## Ключевое свойство унимодальных функций

Пусть  $f(x)$  — унимодальная функция на  $[a, b]$ . Тогда если  $x_1 < x_2 \in [a, b]$ , то:

- если  $f(x_1) \leq f(x_2)$ , то  $x_* \in [a, x_2]$ ;
- если  $f(x_1) \geq f(x_2)$ , то  $x_* \in [x_1, b]$ .

## Ключевое свойство унимодальных функций

Пусть  $f(x)$  — унимодальная функция на  $[a, b]$ . Тогда если  $x_1 < x_2 \in [a, b]$ , то:

- если  $f(x_1) \leq f(x_2)$ , то  $x_* \in [a, x_2]$ ;
- если  $f(x_1) \geq f(x_2)$ , то  $x_* \in [x_1, b]$ .

## Ключевое свойство унимодальных функций

Пусть  $f(x)$  — унимодальная функция на  $[a, b]$ . Тогда если  $x_1 < x_2 \in [a, b]$ , то:

- если  $f(x_1) \leq f(x_2)$ , то  $x_* \in [a, x_2]$ ;
- если  $f(x_1) \geq f(x_2)$ , то  $x_* \in [x_1, b]$ .

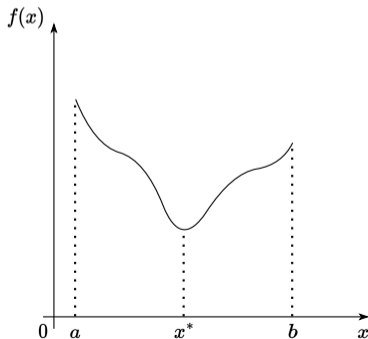
**Доказательство.** Докажем первое утверждение. Допустим противное: пусть  $f(x_1) \leq f(x_2)$ , но  $x^* > x_2$ . Тогда обязательно  $x_1 < x_2 < x^*$ , и по унимодальности функции  $f(x)$  должно выполняться неравенство  $f(x_1) > f(x_2)$ . Получаем противоречие.

## Ключевое свойство унимодальных функций

Пусть  $f(x)$  — унимодальная функция на  $[a, b]$ . Тогда если  $x_1 < x_2 \in [a, b]$ , то:

- если  $f(x_1) \leq f(x_2)$ , то  $x_* \in [a, x_2]$ ;
- если  $f(x_1) \geq f(x_2)$ , то  $x_* \in [x_1, b]$ .

**Доказательство.** Докажем первое утверждение. Допустим противное: пусть  $f(x_1) \leq f(x_2)$ , но  $x^* > x_2$ . Тогда обязательно  $x_1 < x_2 < x^*$ , и по унимодальности функции  $f(x)$  должно выполняться неравенство  $f(x_1) > f(x_2)$ . Получаем противоречие.

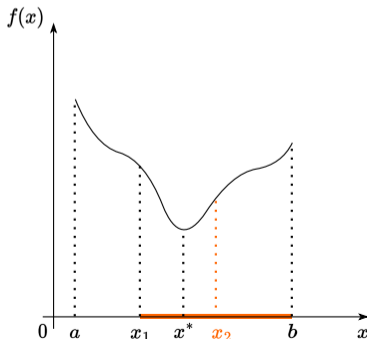
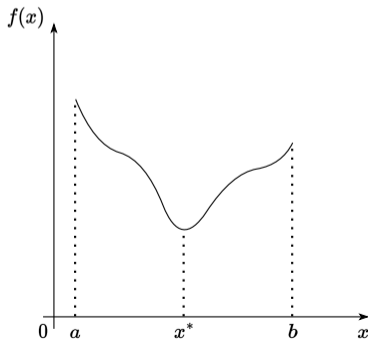


## Ключевое свойство унимодальных функций

Пусть  $f(x)$  — унимодальная функция на  $[a, b]$ . Тогда если  $x_1 < x_2 \in [a, b]$ , то:

- если  $f(x_1) \leq f(x_2)$ , то  $x_* \in [a, x_2]$ ;
- если  $f(x_1) \geq f(x_2)$ , то  $x_* \in [x_1, b]$ .

**Доказательство.** Докажем первое утверждение. Допустим противное: пусть  $f(x_1) \leq f(x_2)$ , но  $x^* > x_2$ . Тогда обязательно  $x_1 < x_2 < x^*$ , и по унимодальности функции  $f(x)$  должно выполняться неравенство  $f(x_1) > f(x_2)$ . Получаем противоречие.

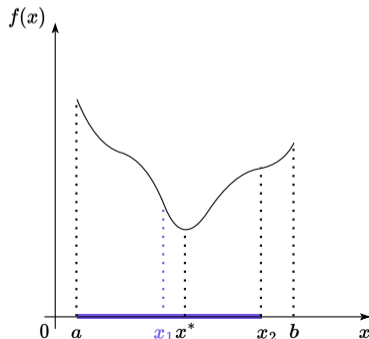
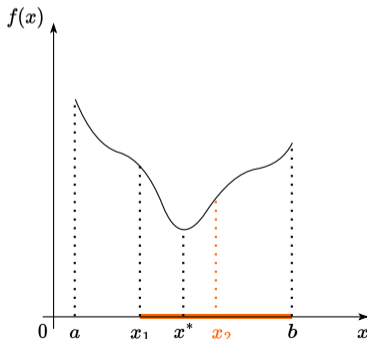
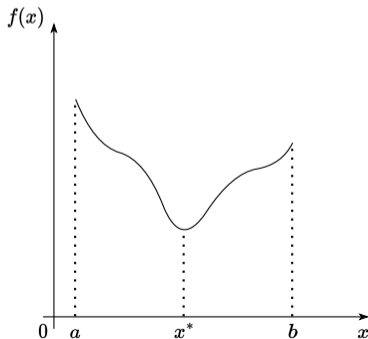


## Ключевое свойство унимодальных функций

Пусть  $f(x)$  — унимодальная функция на  $[a, b]$ . Тогда если  $x_1 < x_2 \in [a, b]$ , то:

- если  $f(x_1) \leq f(x_2)$ , то  $x_* \in [a, x_2]$ ;
- если  $f(x_1) \geq f(x_2)$ , то  $x_* \in [x_1, b]$ .

**Доказательство.** Докажем первое утверждение. Допустим противное: пусть  $f(x_1) \leq f(x_2)$ , но  $x^* > x_2$ . Тогда обязательно  $x_1 < x_2 < x^*$ , и по унимодальности функции  $f(x)$  должно выполняться неравенство  $f(x_1) > f(x_2)$ . Получаем противоречие.



## Метод дихотомии

Мы стремимся решить следующую задачу:

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

Разделяем отрезок на две равные части и выбираем ту, которая содержит решение задачи, используя значения функции и опираясь на ключевое свойство, описанное выше. Наша цель после одной итерации метода — сократить область, содержащую решение, вдвое.

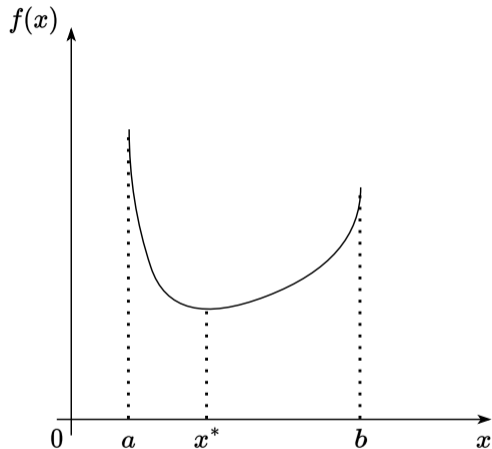


Рис. 3: Метод дихотомии для унимодальной функции

## Метод дихотомии

Мы измеряем значение функции в середине отрезка.

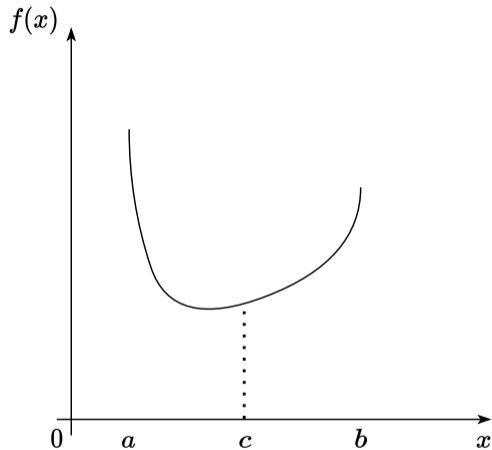


Рис. 4: Метод дихотомии для унимодальной функции

## Метод дихотомии

Чтобы применить ключевое свойство, мы выполняем ещё одно измерение.

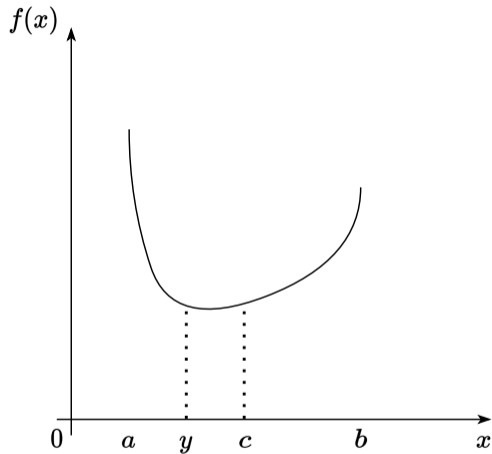


Рис. 5: Метод дихотомии для унимодальной функции

## Метод дихотомии

Выбираем целевой отрезок. В данном случае нам повезло — область, содержащая решение, уже сократилась вдвое. Но так бывает не всегда.

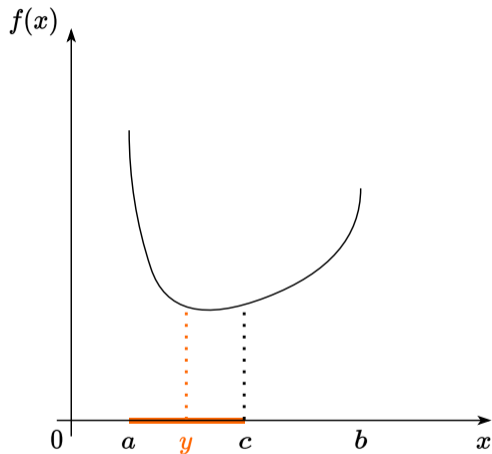


Рис. 6: Метод дихотомии для унимодальной функции

## Метод дихотомии

Рассмотрим другую унимодальную функцию.

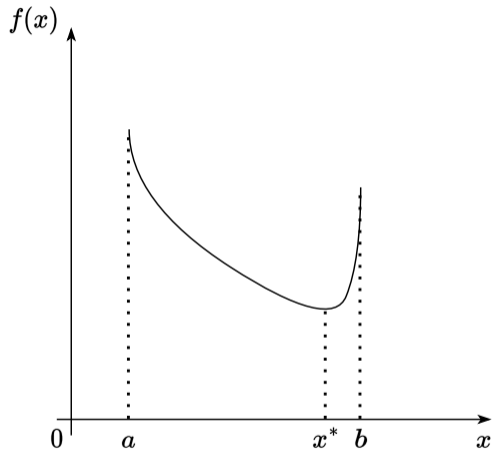


Рис. 7: Метод дихотомии для унимодальной функции

## Метод дихотомии

Измерим значение функции в середине отрезка.

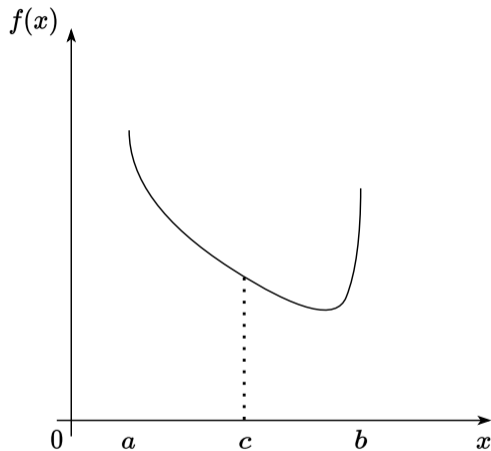


Рис. 8: Метод дихотомии для унимодальной функции

## Метод дихотомии

Проведём ещё одно измерение.

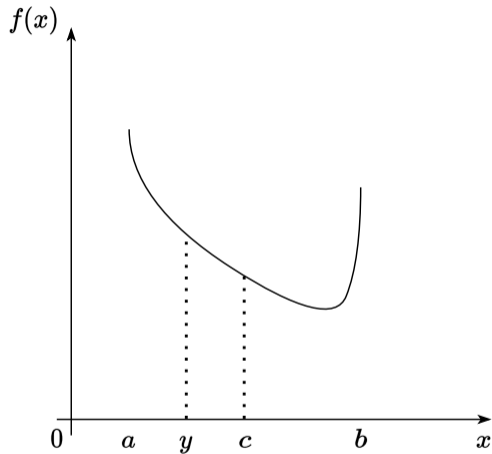


Рис. 9: Метод дихотомии для унимодальной функции

## Метод дихотомии

Выбираем целевой отрезок. Очевидно, что получившийся отрезок не равен половине исходного — он равен  $\frac{3}{4}(b - a)$ . Чтобы исправить это, нужен ещё один шаг алгоритма.

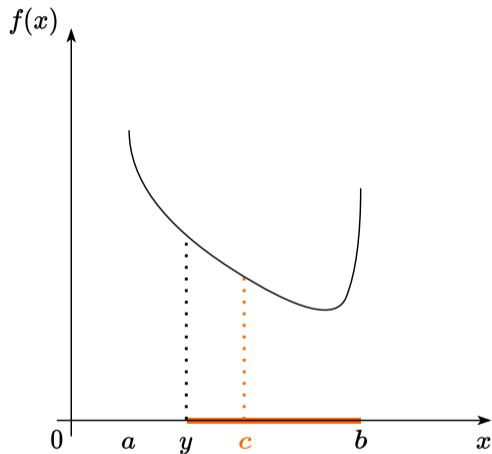


Рис. 10: Метод дихотомии для унимодальной функции

## Метод дихотомии

После ещё одного дополнительного измерения мы получим  $\frac{2}{3} \frac{3}{4}(b-a) = \frac{1}{2}(b-a)$ .

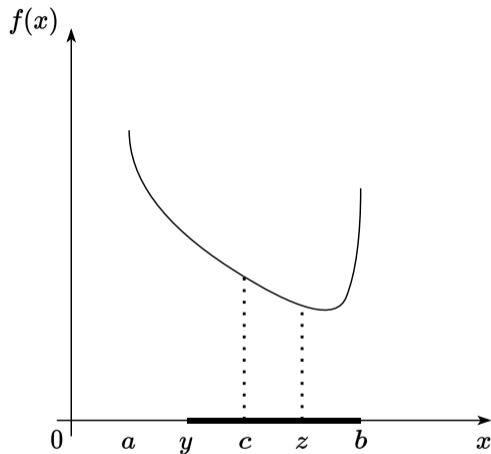


Рис. 11: Метод дихотомии для унимодальной функции

## Метод дихотомии

В итоге: каждая последующая итерация потребует не более двух измерений значения функции.

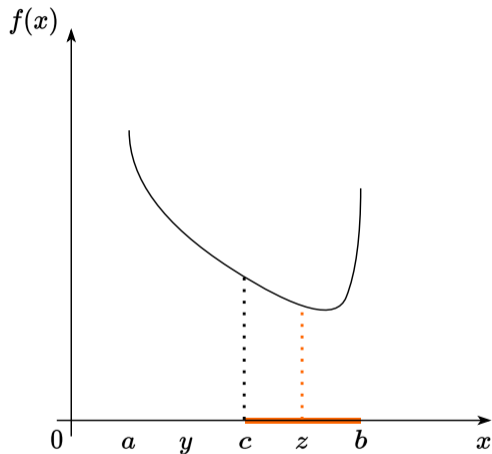


Рис. 12: Метод дихотомии для унимодальной функции

## Метод дихотомии: код алгоритма

```
def binary_search(f, a, b, epsilon):  
    c = (a + b) / 2  
    while abs(b - a) > epsilon:  
        y = (a + c) / 2.0  
        if f(y) <= f(c):  
            b = c  
            c = y  
        else:  
            z = (b + c) / 2.0  
            if f(c) <= f(z):  
                a = y  
                b = z  
            else:  
                a = c  
                c = z  
    return c
```

## Метод дихотомии. Оценки

Длина отрезка на  $(k + 1)$ -й итерации:

$$\Delta_{k+1} = b_{k+1} - a_{k+1} = \frac{1}{2^k}(b - a)$$

## Метод дихотомии. Оценки

Длина отрезка на  $(k + 1)$ -й итерации:

$$\Delta_{k+1} = b_{k+1} - a_{k+1} = \frac{1}{2^k}(b - a)$$

Для унимодальных функций это справедливо, если в качестве выхода итерации выбирать середину отрезка  $x_{k+1}$ :

$$|x_{k+1} - x_*| \leq \frac{\Delta_{k+1}}{2} \leq \frac{1}{2^{k+1}}(b - a) \leq (0.5)^{k+1} \cdot (b - a)$$

## Метод дихотомии. Оценки

Длина отрезка на  $(k + 1)$ -й итерации:

$$\Delta_{k+1} = b_{k+1} - a_{k+1} = \frac{1}{2^k}(b - a)$$

Для унимодальных функций это справедливо, если в качестве выхода итерации выбирать середину отрезка  $x_{k+1}$ :

$$|x_{k+1} - x_*| \leq \frac{\Delta_{k+1}}{2} \leq \frac{1}{2^{k+1}}(b - a) \leq (0.5)^{k+1} \cdot (b - a)$$

Заметим, что на каждой итерации мы обращаемся к оракулу не более двух раз, поэтому число оценок функции равно  $N = 2 \cdot k$ , что влечёт за собой:

$$|x_{k+1} - x_*| \leq (0.5)^{\frac{N}{2}+1} \cdot (b - a) \leq (0.707)^N \frac{b - a}{2}$$

## Метод дихотомии. Оценки

Длина отрезка на  $(k + 1)$ -й итерации:

$$\Delta_{k+1} = b_{k+1} - a_{k+1} = \frac{1}{2^k}(b - a)$$

Для унимодальных функций это справедливо, если в качестве выхода итерации выбирать середину отрезка  $x_{k+1}$ :

$$|x_{k+1} - x_*| \leq \frac{\Delta_{k+1}}{2} \leq \frac{1}{2^{k+1}}(b - a) \leq (0.5)^{k+1} \cdot (b - a)$$

Заметим, что на каждой итерации мы обращаемся к оракулу не более двух раз, поэтому число оценок функции равно  $N = 2 \cdot k$ , что влечёт за собой:

$$|x_{k+1} - x_*| \leq (0.5)^{\frac{N}{2}+1} \cdot (b - a) \leq (0.707)^N \frac{b - a}{2}$$

Положив правую часть последнего неравенства равной  $\varepsilon$ , получим число итераций метода, необходимое для достижения точности  $\varepsilon$ :

$$K = \left\lceil \log_2 \frac{b - a}{\varepsilon} - 1 \right\rceil$$

## Метод золотого сечения

Идея во многом похожа на метод дихотомии. На отрезке выбираются две точки золотого сечения (левая и правая), и проницательная мысль состоит в том, что на следующей итерации одна из этих точек останется «золотой».

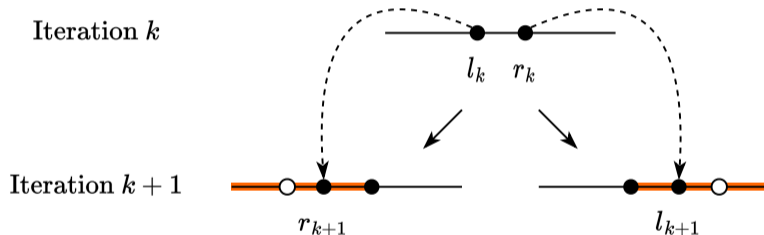


Рис. 13: Ключевая идея, позволяющая уменьшить число обращений к функции

## Метод золотого сечения: код алгоритма

```
def golden_search(f, a, b, epsilon):  
    tau = (sqrt(5) + 1) / 2  
    y = a + (b - a) / tau**2  
    z = a + (b - a) / tau  
    while b - a > epsilon:  
        if f(y) <= f(z):  
            b = z  
            z = y  
            y = a + (b - a) / tau**2  
        else:  
            a = y  
            y = z  
            z = a + (b - a) / tau  
    return (a + b) / 2
```

## Метод золотого сечения. Оценки

$$|x_{k+1} - x_*| \leq b_{k+1} - a_{k+1} = \left(\frac{1}{\tau}\right)^{N-1} (b - a) \approx 0.618^k (b - a),$$

где  $\tau = \frac{\sqrt{5}+1}{2}$ .

- Константа геометрической прогрессии **больше**, чем в методе дихотомии — 0.618 против 0.5 (хуже по скоростному сокращению интервала).

## Метод золотого сечения. Оценки

$$|x_{k+1} - x_*| \leq b_{k+1} - a_{k+1} = \left(\frac{1}{\tau}\right)^{N-1} (b - a) \approx 0.618^k (b - a),$$

где  $\tau = \frac{\sqrt{5}+1}{2}$ .

- Константа геометрической прогрессии **больше**, чем в методе дихотомии — 0.618 против 0.5 (хуже по скоростному сокращению интервала).
- Число вызовов функции **меньше**, чем в методе дихотомии — (показатель 0.707 хуже, чем 0.618) — для каждой итерации метода дихотомии, за исключением первой, функция вычисляется не более 2 раз, а для метода золотого сечения — не более одного раза.

## Последовательная параболическая интерполяция

Выборка трёх точек функции однозначно определяет параболу. Используя эту информацию, мы перейдём непосредственно к её минимуму. Пусть у нас есть три точки  $x_1 < x_2 < x_3$  такие, что отрезок  $[x_1, x_3]$  содержит минимум функции  $f(x)$ . Тогда нужно решить следующую систему уравнений:

## Последовательная параболическая интерполяция

Выборка трёх точек функции однозначно определяет параболу. Используя эту информацию, мы перейдём непосредственно к её минимуму. Пусть у нас есть три точки  $x_1 < x_2 < x_3$  такие, что отрезок  $[x_1, x_3]$  содержит минимум функции  $f(x)$ . Тогда нужно решить следующую систему уравнений:

$$ax_i^2 + bx_i + c = f_i = f(x_i), \quad i = 1, 2, 3$$

Заметим, что эта система линейна по неизвестным  $a, b, c$ . Минимум этой параболы вычисляется как:

## Последовательная параболическая интерполяция

Выборка трёх точек функции однозначно определяет параболу. Используя эту информацию, мы перейдём непосредственно к её минимуму. Пусть у нас есть три точки  $x_1 < x_2 < x_3$  такие, что отрезок  $[x_1, x_3]$  содержит минимум функции  $f(x)$ . Тогда нужно решить следующую систему уравнений:

$$ax_i^2 + bx_i + c = f_i = f(x_i), \quad i = 1, 2, 3$$

Заметим, что эта система линейна по неизвестным  $a, b, c$ . Минимум этой параболы вычисляется как:

$$u = -\frac{b}{2a} = x_2 - \frac{(x_2 - x_1)^2(f_2 - f_3) - (x_2 - x_3)^2(f_2 - f_1)}{2[(x_2 - x_1)(f_2 - f_3) - (x_2 - x_3)(f_2 - f_1)]}$$

Заметим, что если  $f_2 < f_1$  и  $f_2 < f_3$ , то  $u$  лежит в  $[x_1, x_3]$ .

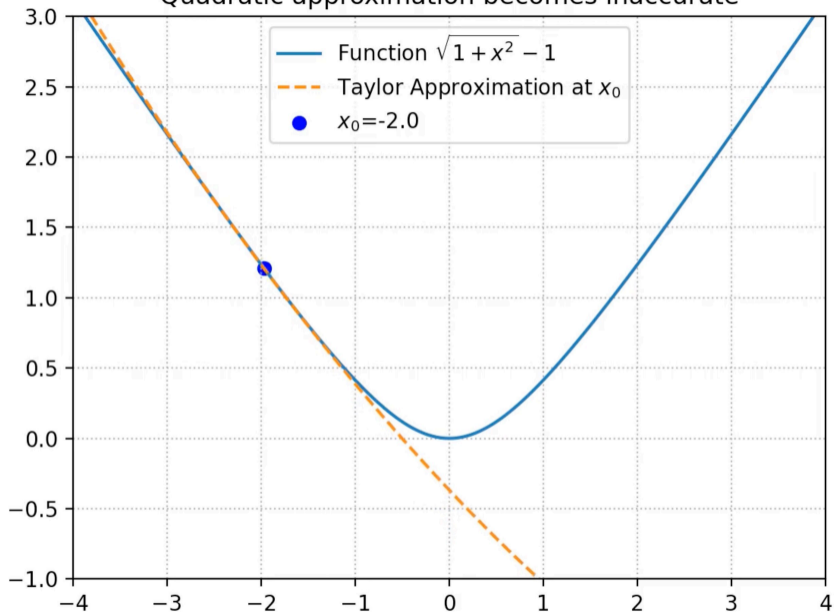
# Последовательная параболическая интерполяция: код алгоритма <sup>1</sup>

```
def parabola_search(f, x1, x2, x3, epsilon):
    f1, f2, f3 = f(x1), f(x2), f(x3)
    while x3 - x1 > epsilon:
        u = x2 - ((x2 - x1)**2*(f2 - f3) - (x2 - x3)**2*(f2 - f1))/(2*((x2 - x1)*(f2 - f3) - (x2 - x3)*(f2 - f1)))
        fu = f(u)

        if x2 <= u:
            if f2 <= fu:
                x1, x2, x3 = x1, x2, u
                f1, f2, f3 = f1, f2, fu
            else:
                x1, x2, x3 = x2, u, x3
                f1, f2, f3 = f2, fu, f3
        else:
            if fu <= f2:
                x1, x2, x3 = x1, u, x2
                f1, f2, f3 = f1, fu, f2
            else:
                x1, x2, x3 = u, x2, x3
                f1, f2, f3 = fu, f2, f3
    return (x1 + x3) / 2
```

<sup>1</sup>Сходимость этого метода сверхлинейная, но локальная, что означает, что воспользоваться этим методом можно лишь в некоторой окрестности оптимума. Здесь (ссылка) приведено доказательство сверхлинейной сходимости.

## Quadratic approximation becomes inaccurate



## Неточный линейный поиск

Иногда достаточно найти решение, которое приблизительно решает нашу задачу. Это типичный сценарий для упомянутой задачи выбора длины шага

$$x_{k+1} = x_k - \alpha \nabla f(x_k)$$

$$\alpha = \operatorname{argmin} f(x_{k+1})$$

## Неточный линейный поиск

Иногда достаточно найти решение, которое приблизительно решает нашу задачу. Это типичный сценарий для упомянутой задачи выбора длины шага

$$x_{k+1} = x_k - \alpha \nabla f(x_k)$$

$$\alpha = \operatorname{argmin} f(x_{k+1})$$

Рассмотрим скалярную функцию  $\phi(\alpha)$  в точке  $x_k$ :

$$\phi(\alpha) = f(x_k - \alpha \nabla f(x_k)), \quad \alpha \geq 0$$

## Неточный линейный поиск

Иногда достаточно найти решение, которое приблизительно решает нашу задачу. Это типичный сценарий для упомянутой задачи выбора длины шага

$$\begin{aligned}x_{k+1} &= x_k - \alpha \nabla f(x_k) \\ \alpha &= \operatorname{argmin} f(x_{k+1})\end{aligned}$$

Рассмотрим скалярную функцию  $\phi(\alpha)$  в точке  $x_k$ :

$$\phi(\alpha) = f(x_k - \alpha \nabla f(x_k)), \quad \alpha \geq 0$$

Аппроксимация первого порядка функции  $\phi(\alpha)$  в окрестности  $\alpha = 0$  имеет вид:

$$\phi(\alpha) \approx f(x_k) - \alpha \nabla f(x_k)^T \nabla f(x_k)$$

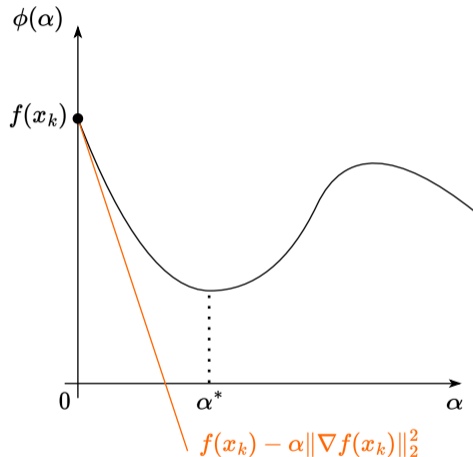


Рис. 14: Иллюстрация Тейлоровской аппроксимации  $\phi_0^I(\alpha)$

## Неточный линейный поиск. Условие достаточного убывания

Условие неточного одномерного поиска, известное как условие Армихо, требует, чтобы  $\alpha$  обеспечивал достаточное уменьшение функции  $f$ , а именно:

$$f(x_k - \alpha \nabla f(x_k)) \leq f(x_k) - c_1 \cdot \alpha \nabla f(x_k)^T \nabla f(x_k)$$

## Неточный линейный поиск. Условие достаточного убывания

Условие неточного одномерного поиска, известное как условие Армихо, требует, чтобы  $\alpha$  обеспечивал достаточное уменьшение функции  $f$ , а именно:

$$f(x_k - \alpha \nabla f(x_k)) \leq f(x_k) - c_1 \cdot \alpha \nabla f(x_k)^T \nabla f(x_k)$$

для некоторой константы  $c_1 \in (0, 1)$ . Обратите внимание, что выбор  $c_1 = 1$  соответствует приближению первого порядка для  $\phi(\alpha)$ . Однако такое условие допускает очень малые значения  $\alpha$ , что может замедлить процесс поиска решения. На практике обычно берут  $c_1 \approx 10^{-4}$ .

## Неточный линейный поиск. Условие достаточного убывания

Условие неточного одномерного поиска, известное как условие Армихо, требует, чтобы  $\alpha$  обеспечивал достаточное уменьшение функции  $f$ , а именно:

$$f(x_k - \alpha \nabla f(x_k)) \leq f(x_k) - c_1 \cdot \alpha \nabla f(x_k)^T \nabla f(x_k)$$

для некоторой константы  $c_1 \in (0, 1)$ . Обратите внимание, что выбор  $c_1 = 1$  соответствует приближению первого порядка для  $\phi(\alpha)$ . Однако такое условие допускает очень малые значения  $\alpha$ , что может замедлить процесс поиска решения. На практике обычно берут  $c_1 \approx 10^{-4}$ .

### i Example

Если  $f(x)$  представляет собой функцию потерь в задаче оптимизации, то выбор подходящего  $c_1$  критичен. Например, при обучении модели машинного обучения неудачно подобранный  $c_1$  может привести либо к медленной сходимости, либо к тому, что минимум будет пропущен.

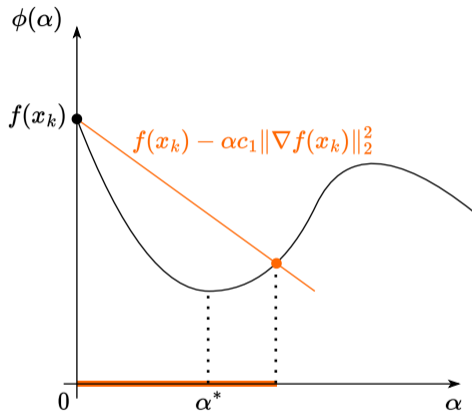


Рис. 15: Иллюстрация условия Армихо с коэффициентом  $c_1$

## Неточный линейный поиск. Условия Голдштейна

Рассмотрим две линейные скалярные функции  $\phi_1(\alpha)$  и  $\phi_2(\alpha)$ :

$$\phi_1(\alpha) = f(x_k) - c_1 \alpha \|\nabla f(x_k)\|^2$$

$$\phi_2(\alpha) = f(x_k) - c_2 \alpha \|\nabla f(x_k)\|^2$$

## Неточный линейный поиск. Условия Голдштейна

Рассмотрим две линейные скалярные функции  $\phi_1(\alpha)$  и  $\phi_2(\alpha)$ :

$$\phi_1(\alpha) = f(x_k) - c_1 \alpha \|\nabla f(x_k)\|^2$$

$$\phi_2(\alpha) = f(x_k) - c_2 \alpha \|\nabla f(x_k)\|^2$$

Условия Голдштейна–Армихо помещают функцию  $\phi(\alpha)$  между  $\phi_1(\alpha)$  и  $\phi_2(\alpha)$ . Как правило, берут  $c_1 = \rho$  и  $c_2 = 1 - \rho$ , где  $\rho \in (0, 0.5)$ .

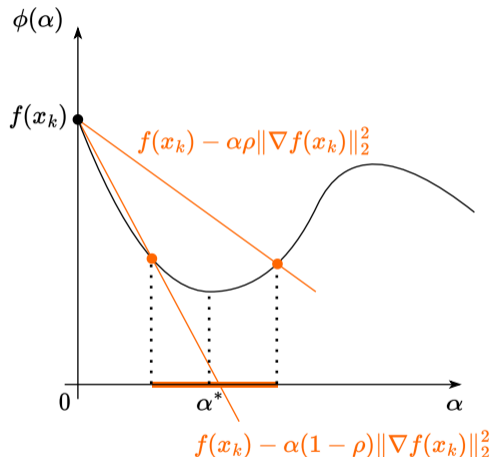


Рис. 16: Иллюстрация условий Голдштейна

## Неточный линейный поиск. Условие кривизны

Чтобы избежать чрезмерно коротких шагов, вводят второе критериальное требование:

$$-\nabla f(x_k - \alpha \nabla f(x_k))^T \nabla f(x_k) \geq c_2 \nabla f(x_k)^T (-\nabla f(x_k))$$

## Неточный линейный поиск. Условие кривизны

Чтобы избежать чрезмерно коротких шагов, вводят второе критериальное требование:

$$-\nabla f(x_k - \alpha \nabla f(x_k))^T \nabla f(x_k) \geq c_2 \nabla f(x_k)^T (-\nabla f(x_k))$$

для некоторого  $c_2 \in (c_1, 1)$ . Здесь  $c_1$  — константа из условия Армихо.

Левая часть есть производная по параметру  $\nabla_{\alpha}\phi(\alpha)$ , что гарантирует, что наклон функции  $\phi(\alpha)$  в целевой точке не меньше, чем  $c_2$  от начального наклона  $\nabla_{\alpha}\phi(\alpha)(0)$ .

Обычно для методов Ньютона или квазиньютоновских методов берут  $c_2 \approx 0.9$ . Вместе условие достаточного уменьшения и условие кривизны образуют условия Вульфа.

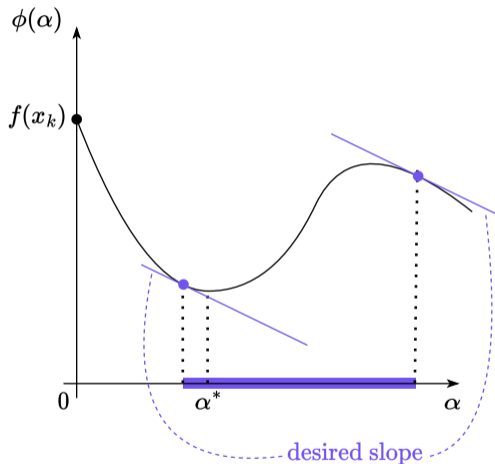


Рис. 17: Иллюстрация условия кривизны

## Неточный линейный поиск. Условие Вульфа

$$-\nabla f(x_k - \alpha \nabla f(x_k))^T \nabla f(x_k) \geq c_2 \nabla f(x_k)^T (-\nabla f(x_k))$$

Вместе условие достаточного уменьшения и условие кривизны образуют условия Вульфа.

### Theorem

Пусть  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  — непрерывно дифференцируемая функция, и пусть  $\phi(\alpha) = f(x_k - \alpha \nabla f(x_k))$ . Предположим, что  $\nabla f(x_k)^T p_k < 0$ , где  $p_k = -\nabla f(x_k)$ , то есть  $p_k$  является направлением спуска. Также предположим, что  $f$  ограничена снизу вдоль луча  $\{x_k + \alpha p_k \mid \alpha > 0\}$ . Цель — показать, что при  $0 < c_1 < c_2 < 1$  существуют интервалы длин шагов, удовлетворяющие условиям Вульфа.

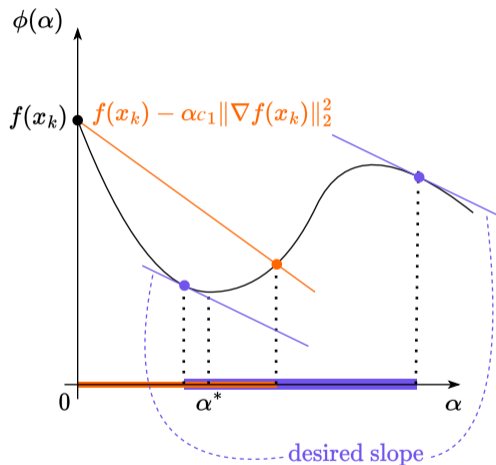


Рис. 18: Иллюстрация условия Вульфа

## Неточный линейный поиск. Условие Вульфа. Доказательство

1. Поскольку  $\phi(\alpha) = f(x_k + \alpha p_k)$  ограничена снизу, а  $l(\alpha) = f(x_k) + \alpha c_1 \nabla f(x_k)^T p_k$  неограниченна снизу (так как  $\nabla f(x_k)^T p_k < 0$ ), график  $l(\alpha)$  обязан пересечь график  $\phi(\alpha)$  по крайней мере один раз. Пусть  $\alpha' > 0$  — наименьшее такое значение, при котором выполняется:

$$f(x_k + \alpha' p_k) \leq f(x_k) + \alpha' c_1 \nabla f(x_k)^T p_k. \quad (1)$$

Это гарантирует выполнение **условия достаточного убывания**.

## Неточный линейный поиск. Условие Вульфа. Доказательство

1. Поскольку  $\phi(\alpha) = f(x_k + \alpha p_k)$  ограничена снизу, а  $l(\alpha) = f(x_k) + \alpha c_1 \nabla f(x_k)^T p_k$  неограниченна снизу (так как  $\nabla f(x_k)^T p_k < 0$ ), график  $l(\alpha)$  обязан пересечь график  $\phi(\alpha)$  по крайней мере один раз. Пусть  $\alpha' > 0$  — наименьшее такое значение, при котором выполняется:

$$f(x_k + \alpha' p_k) \leq f(x_k) + \alpha' c_1 \nabla f(x_k)^T p_k. \quad (1)$$

Это гарантирует выполнение **условия достаточного убывания**.

2. По теореме о среднем значении существует  $\alpha'' \in (0, \alpha')$  такое, что:

$$f(x_k + \alpha' p_k) - f(x_k) = \alpha' \nabla f(x_k + \alpha'' p_k)^T p_k. \quad (2)$$

Подставляя  $f(x_k + \alpha' p_k)$  из (1) в (2), получаем:

$$\alpha' \nabla f(x_k + \alpha'' p_k)^T p_k \leq \alpha' c_1 \nabla f(x_k)^T p_k.$$

Разделив на  $\alpha' > 0$ , получаем:

$$\nabla f(x_k + \alpha'' p_k)^T p_k \leq c_1 \nabla f(x_k)^T p_k. \quad (3)$$

## Неточный линейный поиск. Условие Вульфа. Доказательство

1. Поскольку  $\phi(\alpha) = f(x_k + \alpha p_k)$  ограничена снизу, а  $l(\alpha) = f(x_k) + \alpha c_1 \nabla f(x_k)^T p_k$  неограниченна снизу (так как  $\nabla f(x_k)^T p_k < 0$ ), график  $l(\alpha)$  обязан пересечь график  $\phi(\alpha)$  по крайней мере один раз. Пусть  $\alpha' > 0$  — наименьшее такое значение, при котором выполняется:
3. Поскольку  $c_1 < c_2$  и  $\nabla f(x_k)^T p_k < 0$ , имеем  $c_1 \nabla f(x_k)^T p_k < c_2 \nabla f(x_k)^T p_k$ . Отсюда следует, что существует  $\alpha''$  такое, что:

$$\nabla f(x_k + \alpha'' p_k)^T p_k \leq c_2 \nabla f(x_k)^T p_k. \quad (4)$$

Неравенства (3) и (4) вместе обеспечивают выполнение условий Вульфа.

$$f(x_k + \alpha' p_k) \leq f(x_k) + \alpha' c_1 \nabla f(x_k)^T p_k. \quad (1)$$

Это гарантирует выполнение **условия достаточного убывания**.

2. По теореме о среднем значении существует  $\alpha'' \in (0, \alpha')$  такое, что:

$$f(x_k + \alpha' p_k) - f(x_k) = \alpha' \nabla f(x_k + \alpha'' p_k)^T p_k. \quad (2)$$

Подставляя  $f(x_k + \alpha' p_k)$  из (1) в (2), получаем:

$$\alpha' \nabla f(x_k + \alpha'' p_k)^T p_k \leq \alpha' c_1 \nabla f(x_k)^T p_k.$$

Разделив на  $\alpha' > 0$ , получаем:

$$\nabla f(x_k + \alpha'' p_k)^T p_k \leq c_1 \nabla f(x_k)^T p_k. \quad (3)$$

## Неточный линейный поиск. Условие Вульфа. Доказательство

1. Поскольку  $\phi(\alpha) = f(x_k + \alpha p_k)$  ограничена снизу, а  $l(\alpha) = f(x_k) + \alpha c_1 \nabla f(x_k)^T p_k$  неограниченна снизу (так как  $\nabla f(x_k)^T p_k < 0$ ), график  $l(\alpha)$  обязан пересечь график  $\phi(\alpha)$  по крайней мере один раз. Пусть  $\alpha' > 0$  — наименьшее такое значение, при котором выполняется:

$$f(x_k + \alpha' p_k) \leq f(x_k) + \alpha' c_1 \nabla f(x_k)^T p_k. \quad (1)$$

Это гарантирует выполнение **условия достаточного убывания**.

2. По теореме о среднем значении существует  $\alpha'' \in (0, \alpha')$  такое, что:

$$f(x_k + \alpha' p_k) - f(x_k) = \alpha' \nabla f(x_k + \alpha'' p_k)^T p_k. \quad (2)$$

Подставляя  $f(x_k + \alpha' p_k)$  из (1) в (2), получаем:

$$\alpha' \nabla f(x_k + \alpha'' p_k)^T p_k \leq \alpha' c_1 \nabla f(x_k)^T p_k.$$

Разделив на  $\alpha' > 0$ , получаем:

$$\nabla f(x_k + \alpha'' p_k)^T p_k \leq c_1 \nabla f(x_k)^T p_k. \quad (3)$$

3. Поскольку  $c_1 < c_2$  и  $\nabla f(x_k)^T p_k < 0$ , имеем  $c_1 \nabla f(x_k)^T p_k < c_2 \nabla f(x_k)^T p_k$ . Отсюда следует, что существует  $\alpha''$  такое, что:

$$\nabla f(x_k + \alpha'' p_k)^T p_k \leq c_2 \nabla f(x_k)^T p_k. \quad (4)$$

Неравенства (3) и (4) вместе обеспечивают выполнение условий Вульфа.

4. Для условий Вульфа условие кривизны

$$|\nabla f(x_k + \alpha p_k)^T p_k| \leq c_2 |\nabla f(x_k)^T p_k| \quad (5)$$

выполняется потому, что  $\nabla f(x_k + \alpha p_k)^T p_k$  отрицательно и снизу ограничено значением  $c_2 \nabla f(x_k)^T p_k$ .

## Неточный линейный поиск. Условие Вульфа. Доказательство

1. Поскольку  $\phi(\alpha) = f(x_k + \alpha p_k)$  ограничена снизу, а  $l(\alpha) = f(x_k) + \alpha c_1 \nabla f(x_k)^T p_k$  неограниченна снизу (так как  $\nabla f(x_k)^T p_k < 0$ ), график  $l(\alpha)$  обязан пересечь график  $\phi(\alpha)$  по крайней мере один раз. Пусть  $\alpha' > 0$  — наименьшее такое значение, при котором выполняется:

$$f(x_k + \alpha' p_k) \leq f(x_k) + \alpha' c_1 \nabla f(x_k)^T p_k. \quad (1)$$

Это гарантирует выполнение **условия достаточного убывания**.

2. По теореме о среднем значении существует  $\alpha'' \in (0, \alpha')$  такое, что:

$$f(x_k + \alpha' p_k) - f(x_k) = \alpha' \nabla f(x_k + \alpha'' p_k)^T p_k. \quad (2)$$

Подставляя  $f(x_k + \alpha' p_k)$  из (1) в (2), получаем:

$$\alpha' \nabla f(x_k + \alpha'' p_k)^T p_k \leq \alpha' c_1 \nabla f(x_k)^T p_k.$$

Разделив на  $\alpha' > 0$ , получаем:

$$\nabla f(x_k + \alpha'' p_k)^T p_k \leq c_1 \nabla f(x_k)^T p_k. \quad (3)$$

3. Поскольку  $c_1 < c_2$  и  $\nabla f(x_k)^T p_k < 0$ , имеем  $c_1 \nabla f(x_k)^T p_k < c_2 \nabla f(x_k)^T p_k$ . Отсюда следует, что существует  $\alpha''$  такое, что:

$$\nabla f(x_k + \alpha'' p_k)^T p_k \leq c_2 \nabla f(x_k)^T p_k. \quad (4)$$

Неравенства (3) и (4) вместе обеспечивают выполнение условий Вульфа.

4. Для условий Вульфа условие кривизны

$$|\nabla f(x_k + \alpha p_k)^T p_k| \leq c_2 |\nabla f(x_k)^T p_k| \quad (5)$$

выполняется потому, что  $\nabla f(x_k + \alpha p_k)^T p_k$  отрицательно и снизу ограничено значением  $c_2 \nabla f(x_k)^T p_k$ .

5. В силу гладкости  $f$  существует окрестность вокруг  $\alpha''$ , на которой выполняются условия Вульфа (а значит и сильные условия Вульфа). Следовательно, доказательство завершено.

## Метод обратного шага (Backtracking Line Search)

Метод backtracking line search — приём для нахождения длины шага, удовлетворяющей условию Армихо, условиям Голдштейна или другим критериям неточного одномерного поиска. Начинают с относительно большого шага и итеративно уменьшают его, пока выбранное условие не выполнится.

## Метод обратного шага (Backtracking Line Search)

Метод backtracking line search — приём для нахождения длины шага, удовлетворяющей условию Армихо, условиям Голдштейна или другим критериям неточного одномерного поиска. Начинают с относительно большого шага и итеративно уменьшают его, пока выбранное условие не выполнится.

### Алгоритм:

1. Выбрать начальную длину шага  $\alpha_0$  и параметры  $\beta \in (0, 1)$  и  $c_1 \in (0, 1)$ .

## Метод обратного шага (Backtracking Line Search)

Метод backtracking line search — приём для нахождения длины шага, удовлетворяющей условию Армихо, условиям Голдштейна или другим критериям неточного одномерного поиска. Начинают с относительно большого шага и итеративно уменьшают его, пока выбранное условие не выполнится.

### Алгоритм:

1. Выбрать начальную длину шага  $\alpha_0$  и параметры  $\beta \in (0, 1)$  и  $c_1 \in (0, 1)$ .
2. Проверить, выполняется ли выбранное условие (например, условие Армихо) для текущего шага.

## Метод обратного шага (Backtracking Line Search)

Метод backtracking line search — приём для нахождения длины шага, удовлетворяющей условию Армихо, условиям Голдштейна или другим критериям неточного одномерного поиска. Начинают с относительно большого шага и итеративно уменьшают его, пока выбранное условие не выполнится.

### Алгоритм:

1. Выбрать начальную длину шага  $\alpha_0$  и параметры  $\beta \in (0, 1)$  и  $c_1 \in (0, 1)$ .
2. Проверить, выполняется ли выбранное условие (например, условие Армихо) для текущего шага.
3. Если условие выполнено — остановиться; иначе положить  $\alpha := \beta\alpha$  и вернуться к шагу 2.

## Метод обратного шага (Backtracking Line Search)

Метод backtracking line search — приём для нахождения длины шага, удовлетворяющей условию Армихо, условиям Голдштейна или другим критериям неточного одномерного поиска. Начинают с относительно большого шага и итеративно уменьшают его, пока выбранное условие не выполнится.

### Алгоритм:

1. Выбрать начальную длину шага  $\alpha_0$  и параметры  $\beta \in (0, 1)$  и  $c_1 \in (0, 1)$ .
2. Проверить, выполняется ли выбранное условие (например, условие Армихо) для текущего шага.
3. Если условие выполнено — остановиться; иначе положить  $\alpha := \beta\alpha$  и вернуться к шагу 2.

## Метод обратного шага (Backtracking Line Search)

Метод backtracking line search — приём для нахождения длины шага, удовлетворяющей условию Армихо, условиям Голдштейна или другим критериям неточного одномерного поиска. Начинают с относительно большого шага и итеративно уменьшают его, пока выбранное условие не выполнится.

### Алгоритм:

1. Выбрать начальную длину шага  $\alpha_0$  и параметры  $\beta \in (0, 1)$  и  $c_1 \in (0, 1)$ .
2. Проверить, выполняется ли выбранное условие (например, условие Армихо) для текущего шага.
3. Если условие выполнено — остановиться; иначе положить  $\alpha := \beta\alpha$  и вернуться к шагу 2.

Длина шага обновляется по правилу

$$\alpha_{k+1} := \beta\alpha_k$$

на каждой итерации до тех пор, пока выбранное условие не выполнится.

### Example

При обучении моделей машинного обучения метод backtracking line search можно применять для адаптации скорости обучения. Если значение функции потерь уменьшается недостаточно, скорость обучения умножается на множитель меньше единицы до тех пор, пока не будет выполнено условие Армихо.

## Численные эксперименты

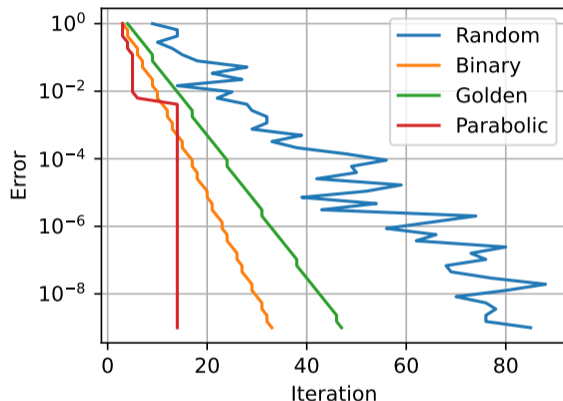
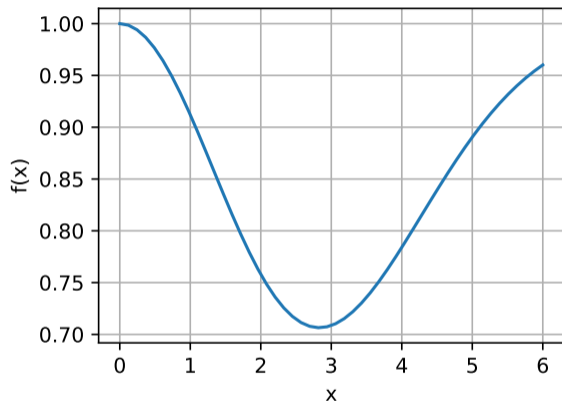


Рис. 19: Сравнение разных алгоритмов линейного поиска

Open In Colab ♣