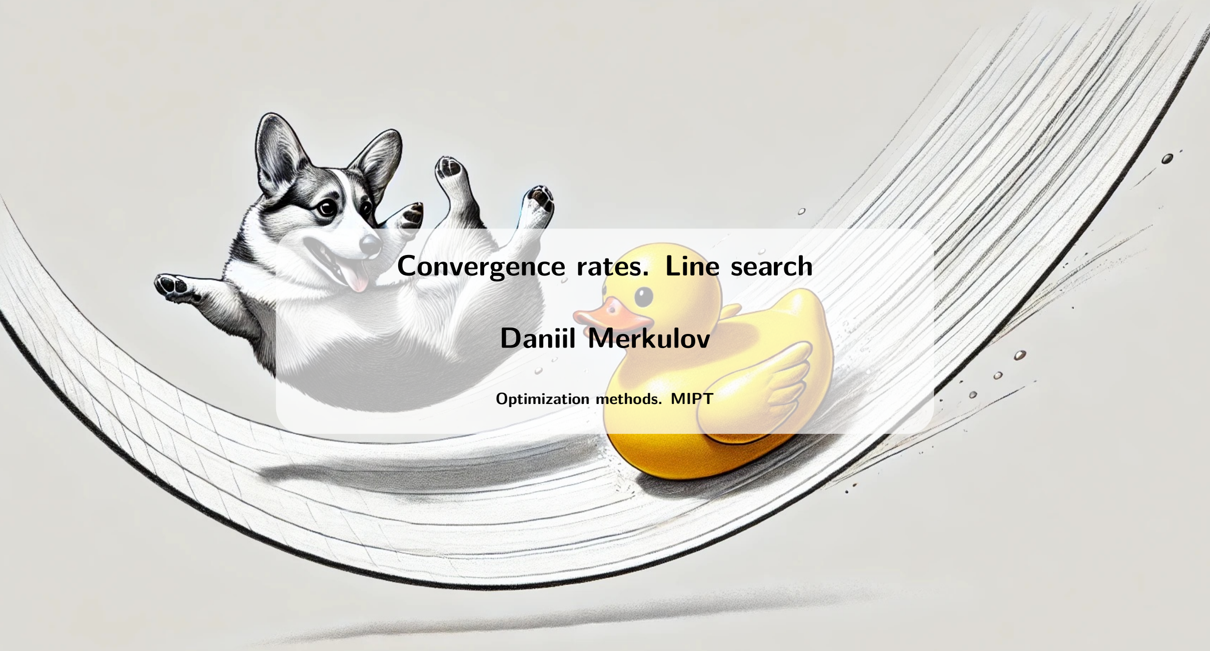


Convergence rates. Line search

Daniil Merkulov

Optimization methods. MIPT

Convergence rates

Convergence rates

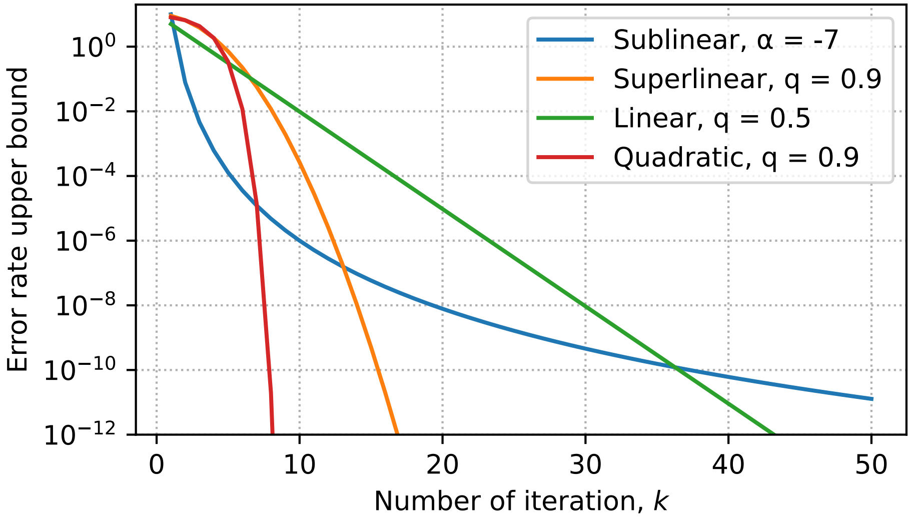


Figure 1: Difference between the convergence speed

Linear convergence

In order to compare perfomance of algorithms we need to define a terminology for different types of convergence. Let $r_k = \{\|x_k - x^*\|_2\}$ be a sequence in $\mathbb{R}^n$ that converges to zero.

Linear convergence

In order to compare perfomance of algorithms we need to define a terminology for different types of convergence. Let $r_k = \{\|x_k - x^*\|_2\}$ be a sequence in $\mathbb{R}^n$ that converges to zero.

We can define the *linear* convergence in a two different forms:

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

Linear convergence

In order to compare perfomance of algorithms we need to define a terminology for different types of convergence. Let $r_k = \{\|x_k - x^*\|_2\}$ be a sequence in $\mathbb{R}^n$ that converges to zero.

We can define the *linear* convergence in a two different forms:

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

for all sufficiently large k. Here $q \in (0, 1)$ and $0 < C < \infty$. This means that the distance to the solution x^* decreases at each iteration by at least a constant factor bounded away from 1. Note, that sometimes this type of convergence is also called *exponential* or *geometric*. The q is called the convergence rate.

Linear convergence

In order to compare perfomance of algorithms we need to define a terminology for different types of convergence. Let $r_k = \{\|x_k - x^*\|_2\}$ be a sequence in $\mathbb{R}^n$ that converges to zero.

We can define the *linear* convergence in a two different forms:

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

for all sufficiently large k. Here $q \in (0, 1)$ and $0 < C < \infty$. This means that the distance to the solution x^* decreases at each iteration by at least a constant factor bounded away from 1. Note, that sometimes this type of convergence is also called *exponential* or *geometric*. The q is called the convergence rate.

Question

Suppose, you have two sequences with linear convergence rates $q_1 = 0.1$ and $q_2 = 0.7$, which one is faster?

Linear convergence

Example

Let us have the following sequence:

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

One can immediately conclude, that we have a linear convergence with parameters $q = \frac{1}{2}$ and $C = 1$.

Linear convergence

Example

Let us have the following sequence:

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

One can immediately conclude, that we have a linear convergence with parameters $q = \dfrac{1}{2}$ and $C = 1$.

Question

Determine the convergence of the following sequence

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

Sub and super

Sublinear convergence

If the sequence r_k converges to zero, but does not have linear convergence, the convergence is said to be sublinear. Sometimes we can consider the following class of sublinear convergence:

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

where $q < 0$ and $0 < C < \infty$. Note, that sublinear convergence means, that the sequence is converging slower, than any geometric progression.

Superlinear convergence

The convergence is said to be *superlinear* if it converges to zero faster, than any linearly convergent sequence.

Root test

Theorem

Let $(r_k)_{k=m}^{\infty}$ be a sequence of non-negative numbers converging to zero, and let $\alpha := \limsup_{k \to \infty} r_k^{1/k}$. (Note that $\alpha \geq 0$.)

(a) If $0 \leq \alpha < 1$, then $(r_k)_{k=m}^{\infty}$ converges linearly with constant α.

Proof.

Root test

Theorem

Let $(r_k)_{k=m}^{\infty}$ be a sequence of non-negative numbers converging to zero, and let $\alpha := \limsup_{k\to\infty} r_k^{1/k}$. (Note that $\alpha \geq 0$.)

(a) If $0 \leq \alpha < 1$, then $(r_k)_{k=m}^{\infty}$ converges linearly with constant α.

(b) In particular, if $\alpha = 0$, then $(r_k)_{k=m}^{\infty}$ converges superlinearly.

Proof.

Root test

Theorem

Let $(r_k)_{k=m}^{\infty}$ be a sequence of non-negative numbers converging to zero, and let $\alpha := \limsup_{k\to\infty} r_k^{1/k}$. (Note that $\alpha \geq 0$.)

(a) If $0 \leq \alpha < 1$, then $(r_k)_{k=m}^{\infty}$ converges linearly with constant α.
(b) In particular, if $\alpha = 0$, then $(r_k)_{k=m}^{\infty}$ converges superlinearly.
(c) If $\alpha = 1$, then $(r_k)_{k=m}^{\infty}$ converges sublinearly.

Proof.

Root test

ℹ Theorem

Let $(r_k)_{k=m}^\infty$ be a sequence of non-negative numbers converging to zero, and let $\alpha := \limsup_{k\to\infty} r_k^{1/k}$. (Note that $\alpha \geq 0$.)

(a) If $0 \leq \alpha < 1$, then $(r_k)_{k=m}^\infty$ converges linearly with constant α.
(b) In particular, if $\alpha = 0$, then $(r_k)_{k=m}^\infty$ converges superlinearly.
(c) If $\alpha = 1$, then $(r_k)_{k=m}^\infty$ converges sublinearly.
(d) The case $\alpha > 1$ is impossible.

Proof.

Root test

Theorem

Let $(r_k)_{k=m}^{\infty}$ be a sequence of non-negative numbers converging to zero, and let $\alpha := \limsup_{k \to \infty} r_k^{1/k}$. (Note that $\alpha \geq 0$.)

(a) If $0 \leq \alpha < 1$, then $(r_k)_{k=m}^{\infty}$ converges linearly with constant α.
(b) In particular, if $\alpha = 0$, then $(r_k)_{k=m}^{\infty}$ converges superlinearly.
(c) If $\alpha = 1$, then $(r_k)_{k=m}^{\infty}$ converges sublinearly.
(d) The case $\alpha > 1$ is impossible.

Proof.

1. let us show that if $(r_k)_{k=m}^{\infty}$ converges linearly with constant $0 \leq \beta < 1$, then necessarily $\alpha \leq \beta$.

   Indeed, by the definition of the constant of linear convergence, for any $\varepsilon > 0$ satisfying $\beta + \varepsilon < 1$, there exists $C > 0$ such that $r_k \leq C(\beta + \varepsilon)^k$ for all $k \geq m$.

   From this, $r_k^{1/k} \leq C^{1/k}(\beta + \varepsilon)$ for all $k \geq m$. Passing to the limit as $k \to \infty$ and using $C^{1/k} \to 1$, we obtain $\alpha \leq \beta + \varepsilon$. Given the arbitrariness of ε, it follows that $\alpha \leq \beta$.

Root test

Theorem

Let $(r_k)_{k=m}^\infty$ be a sequence of non-negative numbers converging to zero, and let $\alpha := \limsup_{k\to\infty} r_k^{1/k}$. (Note that $\alpha \geq 0$.)

(a) If $0 \leq \alpha < 1$, then $(r_k)_{k=m}^\infty$ converges linearly with constant α.
(b) In particular, if $\alpha = 0$, then $(r_k)_{k=m}^\infty$ converges superlinearly.
(c) If $\alpha = 1$, then $(r_k)_{k=m}^\infty$ converges sublinearly.
(d) The case $\alpha > 1$ is impossible.

Proof.

1. let us show that if $(r_k)_{k=m}^\infty$ converges linearly with constant $0 \leq \beta < 1$, then necessarily $\alpha \leq \beta$.

   Indeed, by the definition of the constant of linear convergence, for any $\varepsilon > 0$ satisfying $\beta + \varepsilon < 1$, there exists $C > 0$ such that $r_k \leq C(\beta + \varepsilon)^k$ for all $k \geq m$.

   From this, $r_k^{1/k} \leq C^{1/k}(\beta + \varepsilon)$ for all $k \geq m$. Passing to the limit as $k \to \infty$ and using $C^{1/k} \to 1$, we obtain $\alpha \leq \beta + \varepsilon$. Given the arbitrariness of ε, it follows that $\alpha \leq \beta$.

2. Thus, in the case $\alpha = 1$, the sequence $(r_k)_{k=m}^\infty$ cannot have linear convergence according to the above result (proven by contradiction). Since, nevertheless, $(r_k)_{k=m}^\infty$ converges to zero, it must converge sublinearly.

Root test

3. Now consider the case $0 \leq \alpha < 1$. Let $\varepsilon > 0$ be an arbitrary number such that $\alpha + \varepsilon < 1$.

Root test

3. Now consider the case $0 \leq \alpha < 1$. Let $\varepsilon > 0$ be an arbitrary number such that $\alpha + \varepsilon < 1$.

Root test

3. Now consider the case $0 \leq \alpha < 1$. Let $\varepsilon > 0$ be an arbitrary number such that $\alpha + \varepsilon < 1$.

   According to the properties of the limsup, there exists $N \geq m$ such that $r_k^{1/k} \leq \alpha + \varepsilon$ for all $k \geq N$.

Root test

3. Now consider the case $0 \leq \alpha < 1$. Let $\varepsilon > 0$ be an arbitrary number such that $\alpha + \varepsilon < 1$.

   According to the properties of the limsup, there exists $N \geq m$ such that $r_k^{1/k} \leq \alpha + \varepsilon$ for all $k \geq N$.

   Hence, $r_k \leq (\alpha + \varepsilon)^k$ for all $k \geq N$. Therefore, $(r_k)_{k=m}^{\infty}$ converges linearly with parameter $\alpha + \varepsilon$ (it does not matter that the inequality is only valid from the number N).

Root test

3. Now consider the case $0 \leq \alpha < 1$. Let $\varepsilon > 0$ be an arbitrary number such that $\alpha + \varepsilon < 1$.

   According to the properties of the limsup, there exists $N \geq m$ such that $r_k^{1/k} \leq \alpha + \varepsilon$ for all $k \geq N$.

   Hence, $r_k \leq (\alpha + \varepsilon)^k$ for all $k \geq N$. Therefore, $(r_k)_{k=m}^{\infty}$ converges linearly with parameter $\alpha + \varepsilon$ (it does not matter that the inequality is only valid from the number N).

   Due to the arbitrariness of ε, this means that the constant of linear convergence of $(r_k)_{k=m}^{\infty}$ does not exceed α.

Root test

3. Now consider the case $0 \leq \alpha < 1$. Let $\varepsilon > 0$ be an arbitrary number such that $\alpha + \varepsilon < 1$.

   According to the properties of the limsup, there exists $N \geq m$ such that $r_k^{1/k} \leq \alpha + \varepsilon$ for all $k \geq N$.

   Hence, $r_k \leq (\alpha + \varepsilon)^k$ for all $k \geq N$. Therefore, $(r_k)_{k=m}^{\infty}$ converges linearly with parameter $\alpha + \varepsilon$ (it does not matter that the inequality is only valid from the number N).

   Due to the arbitrariness of ε, this means that the constant of linear convergence of $(r_k)_{k=m}^{\infty}$ does not exceed α.

   Since, as shown above, the constant of linear convergence cannot be less than α, this means that the constant of linear convergence of $(r_k)_{k=m}^{\infty}$ is exactly α.

4. Finally, let's show that the case $\alpha > 1$ is impossible.

Root test

3. Now consider the case $0 \leq \alpha < 1$. Let $\varepsilon > 0$ be an arbitrary number such that $\alpha + \varepsilon < 1$.

   According to the properties of the limsup, there exists $N \geq m$ such that $r_k^{1/k} \leq \alpha + \varepsilon$ for all $k \geq N$.

   Hence, $r_k \leq (\alpha + \varepsilon)^k$ for all $k \geq N$. Therefore, $(r_k)_{k=m}^{\infty}$ converges linearly with parameter $\alpha + \varepsilon$ (it does not matter that the inequality is only valid from the number N).

   Due to the arbitrariness of ε, this means that the constant of linear convergence of $(r_k)_{k=m}^{\infty}$ does not exceed α.

   Since, as shown above, the constant of linear convergence cannot be less than α, this means that the constant of linear convergence of $(r_k)_{k=m}^{\infty}$ is exactly α.

4. Finally, let's show that the case $\alpha > 1$ is impossible.

Root test

3. Now consider the case $0 \leq \alpha < 1$. Let $\varepsilon > 0$ be an arbitrary number such that $\alpha + \varepsilon < 1$.

   According to the properties of the limsup, there exists $N \geq m$ such that $r_k^{1/k} \leq \alpha + \varepsilon$ for all $k \geq N$.

   Hence, $r_k \leq (\alpha + \varepsilon)^k$ for all $k \geq N$. Therefore, $(r_k)_{k=m}^{\infty}$ converges linearly with parameter $\alpha + \varepsilon$ (it does not matter that the inequality is only valid from the number N).

   Due to the arbitrariness of ε, this means that the constant of linear convergence of $(r_k)_{k=m}^{\infty}$ does not exceed α.

   Since, as shown above, the constant of linear convergence cannot be less than α, this means that the constant of linear convergence of $(r_k)_{k=m}^{\infty}$ is exactly α.

4. Finally, let's show that the case $\alpha > 1$ is impossible.

   Indeed, suppose $\alpha > 1$. Then from the definition of limsup, it follows that for any $N \geq m$, there exists $k \geq N$ such that $r_k^{1/k} \geq 1$, and, in particular, $r_k \geq 1$.

Root test

3. Now consider the case $0 \leq \alpha < 1$. Let $\varepsilon > 0$ be an arbitrary number such that $\alpha + \varepsilon < 1$.

   According to the properties of the limsup, there exists $N \geq m$ such that $r_k^{1/k} \leq \alpha + \varepsilon$ for all $k \geq N$.

   Hence, $r_k \leq (\alpha + \varepsilon)^k$ for all $k \geq N$. Therefore, $(r_k)_{k=m}^\infty$ converges linearly with parameter $\alpha + \varepsilon$ (it does not matter that the inequality is only valid from the number N).

   Due to the arbitrariness of ε, this means that the constant of linear convergence of $(r_k)_{k=m}^\infty$ does not exceed α.

   Since, as shown above, the constant of linear convergence cannot be less than α, this means that the constant of linear convergence of $(r_k)_{k=m}^\infty$ is exactly α.

4. Finally, let's show that the case $\alpha > 1$ is impossible.

   Indeed, suppose $\alpha > 1$. Then from the definition of limsup, it follows that for any $N \geq m$, there exists $k \geq N$ such that $r_k^{1/k} \geq 1$, and, in particular, $r_k \geq 1$.

   But this means that r_k has a subsequence that is bounded away from zero. Hence, $(r_k)_{k=m}^\infty$ cannot converge to zero, which contradicts the condition.

Ratio test

Let $\{r_k\}_{k=m}^{\infty}$ be a sequence of strictly positive numbers converging to zero. Let

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

- If there exists q and $0 \leq q < 1$, then $\{r_k\}_{k=m}^{\infty}$ has linear convergence with constant q.

Ratio test

Let $\{r_k\}_{k=m}^{\infty}$ be a sequence of strictly positive numbers converging to zero. Let

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

- If there exists q and $0 \leq q < 1$, then $\{r_k\}_{k=m}^{\infty}$ has linear convergence with constant q.
- In particular, if $q = 0$, then $\{r_k\}_{k=m}^{\infty}$ has superlinear convergence.

Ratio test

Let $\{r_k\}_{k=m}^{\infty}$ be a sequence of strictly positive numbers converging to zero. Let

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

- If there exists q and $0 \leq q < 1$, then $\{r_k\}_{k=m}^{\infty}$ has linear convergence with constant q.
- In particular, if $q = 0$, then $\{r_k\}_{k=m}^{\infty}$ has superlinear convergence.
- If q does not exist, but $q = \lim_{k \to \infty} \sup_k \frac{r_{k+1}}{r_k} < 1$, then $\{r_k\}_{k=m}^{\infty}$ has linear convergence with a constant not exceeding q.

Ratio test

Let $\{r_k\}_{k=m}^\infty$ be a sequence of strictly positive numbers converging to zero. Let

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

- If there exists q and $0 \leq q < 1$, then $\{r_k\}_{k=m}^\infty$ has linear convergence with constant q.
- In particular, if $q = 0$, then $\{r_k\}_{k=m}^\infty$ has superlinear convergence.
- If q does not exist, but $q = \lim\limits_{k\to\infty} \sup_k \frac{r_{k+1}}{r_k} < 1$, then $\{r_k\}_{k=m}^\infty$ has linear convergence with a constant not exceeding q.
- If $\lim\limits_{k\to\infty} \inf_k \frac{r_{k+1}}{r_k} = 1$, then $\{r_k\}_{k=m}^\infty$ has sublinear convergence.

Ratio test

Let $\{r_k\}_{k=m}^{\infty}$ be a sequence of strictly positive numbers converging to zero. Let

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

- If there exists q and $0 \leq q < 1$, then $\{r_k\}_{k=m}^{\infty}$ has linear convergence with constant q.
- In particular, if $q = 0$, then $\{r_k\}_{k=m}^{\infty}$ has superlinear convergence.
- If q does not exist, but $q = \lim\limits_{k \to \infty} \sup_k \frac{r_{k+1}}{r_k} < 1$, then $\{r_k\}_{k=m}^{\infty}$ has linear convergence with a constant not exceeding q.
- If $\lim\limits_{k \to \infty} \inf_k \frac{r_{k+1}}{r_k} = 1$, then $\{r_k\}_{k=m}^{\infty}$ has sublinear convergence.
- The case $\lim\limits_{k \to \infty} \inf_k \frac{r_{k+1}}{r_k} > 1$ is impossible.

Ratio test

Let $\{r_k\}_{k=m}^{\infty}$ be a sequence of strictly positive numbers converging to zero. Let

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

- If there exists q and $0 \leq q < 1$, then $\{r_k\}_{k=m}^{\infty}$ has linear convergence with constant q.
- In particular, if $q = 0$, then $\{r_k\}_{k=m}^{\infty}$ has superlinear convergence.
- If q does not exist, but $q = \lim\limits_{k \to \infty} \sup_k \frac{r_{k+1}}{r_k} < 1$, then $\{r_k\}_{k=m}^{\infty}$ has linear convergence with a constant not exceeding q.
- If $\lim\limits_{k \to \infty} \inf_k \frac{r_{k+1}}{r_k} = 1$, then $\{r_k\}_{k=m}^{\infty}$ has sublinear convergence.
- The case $\lim\limits_{k \to \infty} \inf_k \frac{r_{k+1}}{r_k} > 1$ is impossible.
- In all other cases (i.e., when $\lim\limits_{k \to \infty} \inf_k \frac{r_{k+1}}{r_k} < 1 \leq \lim\limits_{k \to \infty} \sup_k \frac{r_{k+1}}{r_k}$) we cannot claim anything concrete about the convergence rate $\{r_k\}_{k=m}^{\infty}$.

Ratio test lemma

Theorem

Let $(r_k)_{k=m}^{\infty}$ be a sequence of strictly positive numbers. (The strict positivity is necessary to ensure that the ratios $\frac{r_{k+1}}{r_k}$, which appear below, are well-defined.) Then

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

Ratio test lemma

Theorem

Let $(r_k)_{k=m}^{\infty}$ be a sequence of strictly positive numbers. (The strict positivity is necessary to ensure that the ratios $\frac{r_{k+1}}{r_k}$, which appear below, are well-defined.) Then

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

Proof.

1. The middle inequality follows from the fact that the liminf of any sequence is always less than or equal to its limsup. Let's prove the last inequality; the first one is proved analogously.

Ratio test lemma

Theorem

Let $(r_k)_{k=m}^{\infty}$ be a sequence of strictly positive numbers. (The strict positivity is necessary to ensure that the ratios $\frac{r_{k+1}}{r_k}$, which appear below, are well-defined.) Then

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

Proof.

1. The middle inequality follows from the fact that the liminf of any sequence is always less than or equal to its limsup. Let's prove the last inequality; the first one is proved analogously.
2. Denote $L := \limsup_{k\to\infty} \frac{r_{k+1}}{r_k}$. If $L = +\infty$, then the inequality is obviously true, so let's assume L is finite. Note that $L \ge 0$, since the ratio $\frac{r_{k+1}}{r_k}$ is positive for all $k \ge m$. Let $\varepsilon > 0$ be an arbitrary number. According to the properties of limsup, there exists $N \ge m$ such that $\frac{r_{k+1}}{r_k} \le L + \varepsilon$ for all $k \ge N$. From here, $r_{k+1} \le (L+\varepsilon) r_k$ for all $k \ge N$. Applying induction, we get $r_k \le (L+\varepsilon)^{k-N} r_N$ for all $k \ge N$. Let $C := (L+\varepsilon)^{-N} r_N$. Then $r_k \le C(L+\varepsilon)^k$ for all $k \ge N$, from which $r_k^{1/k} \le C^{1/k}(L+\varepsilon)$. Taking the limsup as $k \to \infty$ and using $C^{1/k} \to 1$, we get $\limsup_{k\to\infty} r_k^{1/k} \le L + \varepsilon$. Given the arbitrariness of ε, it follows that $\limsup_{k\to\infty} r_k^{1/k} \le L$.

Line search

Problem

Suppose, we have a problem of minimization of a function $f(x) : \mathbb{R} \to \mathbb{R}$ of scalar variable:

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

Problem

Suppose, we have a problem of minimization of a function $f(x) : \mathbb{R} \to \mathbb{R}$ of scalar variable:

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

Sometimes, we refer to the similar problem of finding minimum on the line segment $[a, b]$:

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

Problem

Suppose, we have a problem of minimization of a function $f(x) : \mathbb{R} \to \mathbb{R}$ of scalar variable:

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

Sometimes, we refer to the similar problem of finding minimum on the line segment $[a, b]$:

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

Example

Typical example of line search problem is selecting appropriate stepsize for gradient descent algorithm:

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

Problem

Suppose, we have a problem of minimization of a function $f(x) : \mathbb{R} \to \mathbb{R}$ of scalar variable:

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

Sometimes, we refer to the similar problem of finding minimum on the line segment $[a, b]$:

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

Example

Typical example of line search problem is selecting appropriate stepsize for gradient descent algorithm:

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

The line search is a fundamental optimization problem that plays a crucial role in solving complex tasks. To simplify the problem, let's assume that the function, $f(x)$, is *unimodal*, meaning it has a single peak or valley.

Unimodal function

Definition

Function $f(x)$ is called **unimodal** on $[a, b]$, if there is $x_* \in [a, b]$, that $f(x_1) > f(x_2) \quad \forall a \leq x_1 < x_2 < x_*$ and $f(x_1) < f(x_2) \quad \forall x_* < x_1 < x_2 \leq b$

Unimodal function

Definition

Function $f(x)$ is called **unimodal** on $[a, b]$, if there is $x_* \in [a, b]$, that $f(x_1) > f(x_2) \quad \forall a \le x_1 < x_2 < x_*$ and $f(x_1) < f(x_2) \quad \forall x_* < x_1 < x_2 \le b$

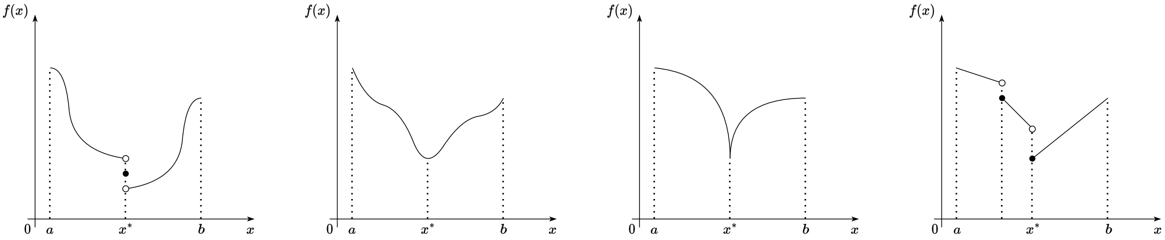


Figure 2: Examples of unimodal functions

Key property of unimodal functions

Let $f(x)$ be unimodal function on $[a, b]$. Than if $x_1 < x_2 \in [a, b]$, then:

* if $f(x_1) \leq f(x_2) \to x_* \in [a, x_2]$

Key property of unimodal functions

Let $f(x)$ be unimodal function on $[a, b]$. Than if $x_1 < x_2 \in [a, b]$, then:

- if $f(x_1) \leq f(x_2) \to x_* \in [a, x_2]$
- if $f(x_1) \geq f(x_2) \to x_* \in [x_1, b]$

Key property of unimodal functions

Let $f(x)$ be unimodal function on $[a, b]$. Than if $x_1 < x_2 \in [a, b]$, then:

- if $f(x_1) \leq f(x_2) \to x_* \in [a, x_2]$
- if $f(x_1) \geq f(x_2) \to x_* \in [x_1, b]$

Key property of unimodal functions

Let $f(x)$ be unimodal function on $[a, b]$. Than if $x_1 < x_2 \in [a, b]$, then:

- if $f(x_1) \leq f(x_2) \to x_* \in [a, x_2]$
- if $f(x_1) \geq f(x_2) \to x_* \in [x_1, b]$

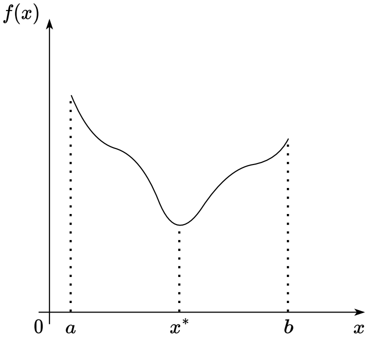
Proof Let's prove the first statement. On the contrary, suppose that $f(x_1) \leq f(x_2)$, but $x^* > x_2$. Then necessarily $x_1 < x_2 < x^*$ and by the unimodality of the function $f(x)$ the inequality: $f(x_1) > f(x_2)$ must be satisfied. We have obtained a contradiction.

Key property of unimodal functions

Let $f(x)$ be unimodal function on $[a, b]$. Than if $x_1 < x_2 \in [a, b]$, then:

- if $f(x_1) \leq f(x_2) \to x_* \in [a, x_2]$
- if $f(x_1) \geq f(x_2) \to x_* \in [x_1, b]$

Proof Let's prove the first statement. On the contrary, suppose that $f(x_1) \leq f(x_2)$, but $x^* > x_2$. Then necessarily $x_1 < x_2 < x^*$ and by the unimodality of the function $f(x)$ the inequality: $f(x_1) > f(x_2)$ must be satisfied. We have obtained a contradiction.

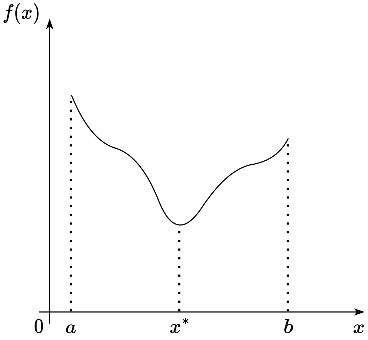


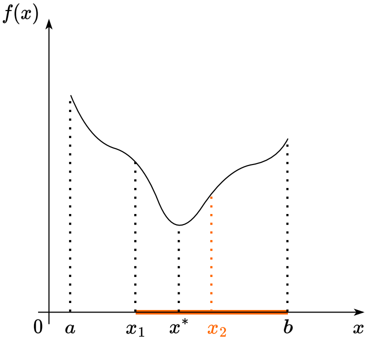
Key property of unimodal functions

Let $f(x)$ be unimodal function on $[a, b]$. Than if $x_1 < x_2 \in [a, b]$, then:

- if $f(x_1) \leq f(x_2) \to x_* \in [a, x_2]$
- if $f(x_1) \geq f(x_2) \to x_* \in [x_1, b]$

Proof Let's prove the first statement. On the contrary, suppose that $f(x_1) \leq f(x_2)$, but $x^* > x_2$. Then necessarily $x_1 < x_2 < x^*$ and by the unimodality of the function $f(x)$ the inequality: $f(x_1) > f(x_2)$ must be satisfied. We have obtained a contradiction.



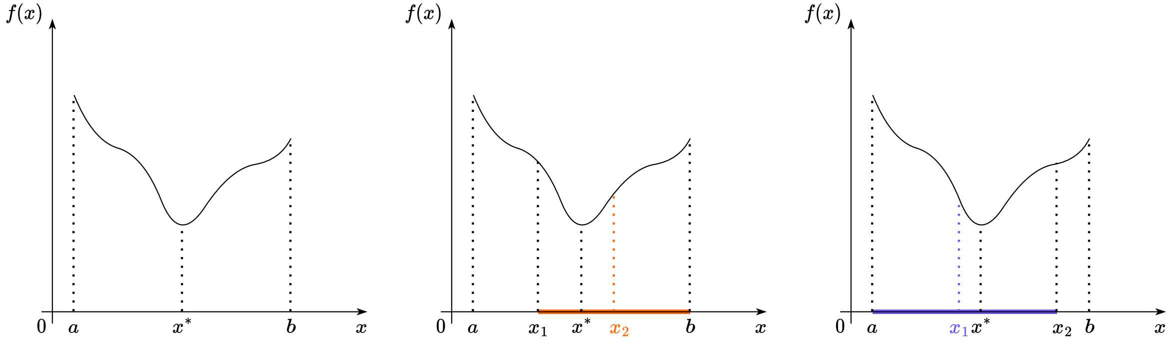


Key property of unimodal functions

Let $f(x)$ be unimodal function on $[a, b]$. Than if $x_1 < x_2 \in [a, b]$, then:

- if $f(x_1) \leq f(x_2) \to x_* \in [a, x_2]$
- if $f(x_1) \geq f(x_2) \to x_* \in [x_1, b]$

Proof Let's prove the first statement. On the contrary, suppose that $f(x_1) \leq f(x_2)$, but $x^* > x_2$. Then necessarily $x_1 < x_2 < x^*$ and by the unimodality of the function $f(x)$ the inequality: $f(x_1) > f(x_2)$ must be satisfied. We have obtained a contradiction.



Dichotomy method

We aim to solve the following problem:

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

We divide a segment into two equal parts and choose the one that contains the solution of the problem using the values of functions, based on the key property described above. Our goal after one iteration of the method is to halve the solution region.

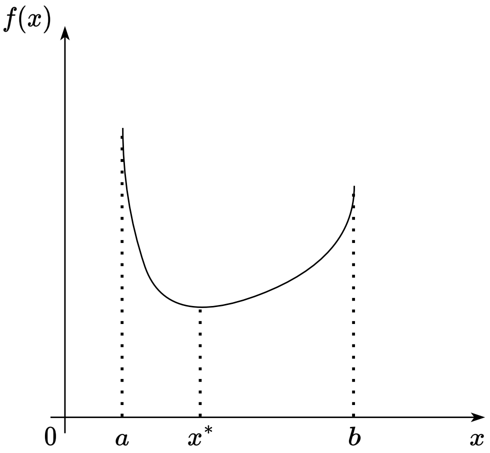


Figure 3: Dichotomy method for unimodal function

Dichotomy method

We measure the function value at the middle of the line segment

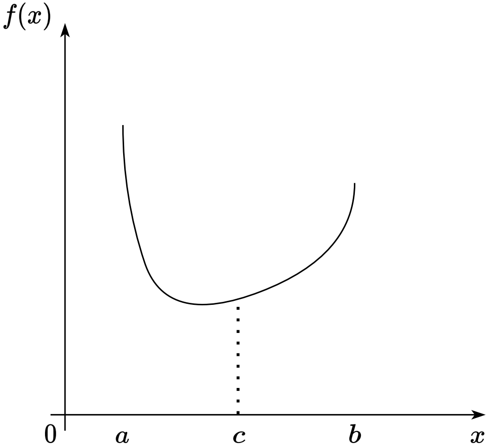


Figure 4: Dichotomy method for unimodal function

Dichotomy method

In order to apply the key property we perform another measurement.

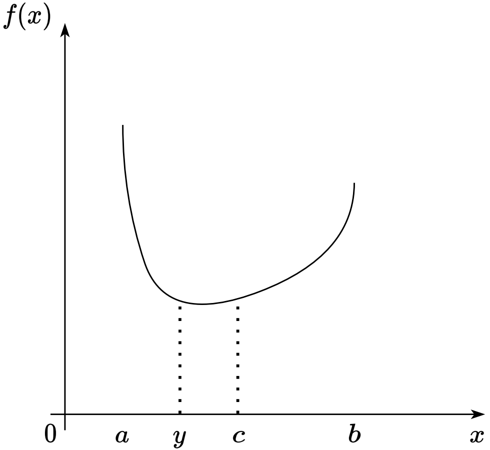


Figure 5: Dichotomy method for unimodal function

Dichotomy method

We select the target line segment. And in this case we are lucky since we already halved the solution region. But that is not always the case.

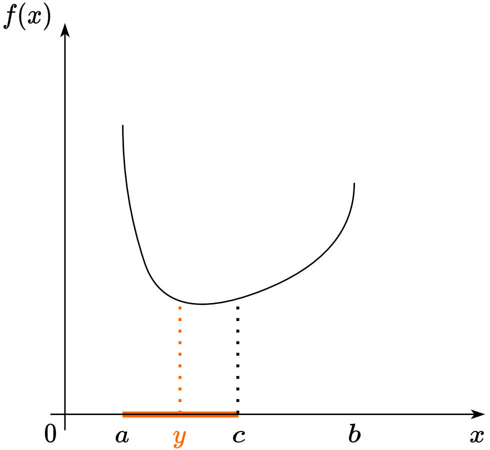


Figure 6: Dichotomy method for unimodal function

Dichotomy method

Let's consider another unimodal function.

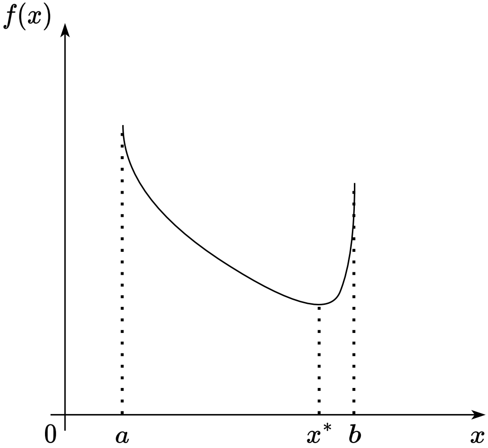


Figure 7: Dichotomy method for unimodal function

Dichotomy method

Measure the middle of the line segment.

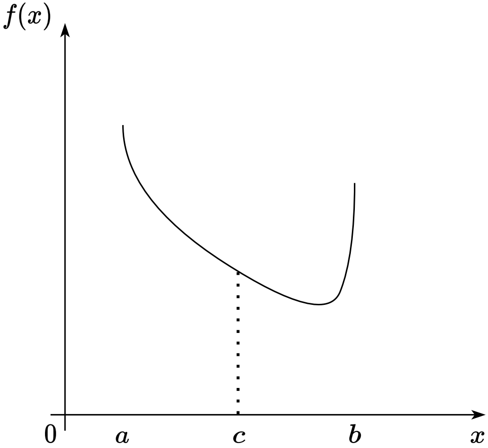


Figure 8: Dichotomy method for unimodal function

Dichotomy method

Get another measurement.

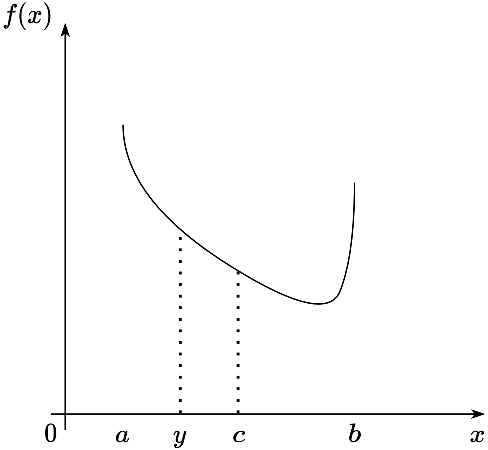


Figure 9: Dichotomy method for unimodal function

Dichotomy method

Select the target line segment. You can clearly see, that the obtained line segment is not the half of the initial one. It is $\frac{3}{4}(b-a)$. So to fix it we need another step of the algorithm.

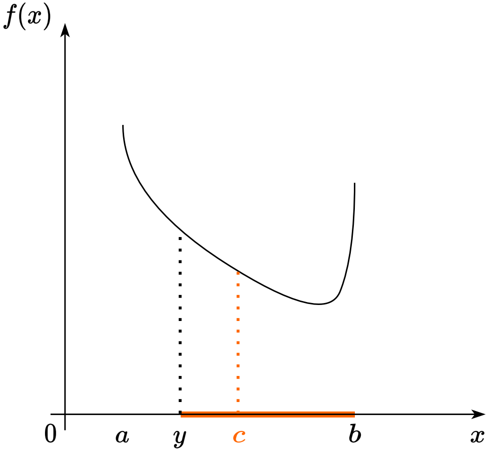


Figure 10: Dichotomy method for unimodal function

Dichotomy method

After another additional measurement, we will surely get $\frac{2}{3}\frac{3}{4}(b-a) = \frac{1}{2}(b-a)$

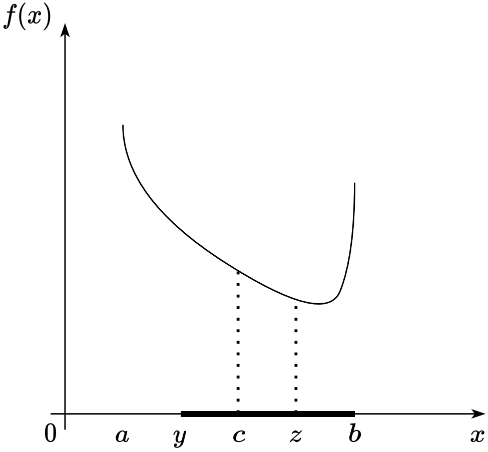


Figure 11: Dichotomy method for unimodal function

Dichotomy method

To sum it up, each subsequent iteration will require at most two function value measurements.

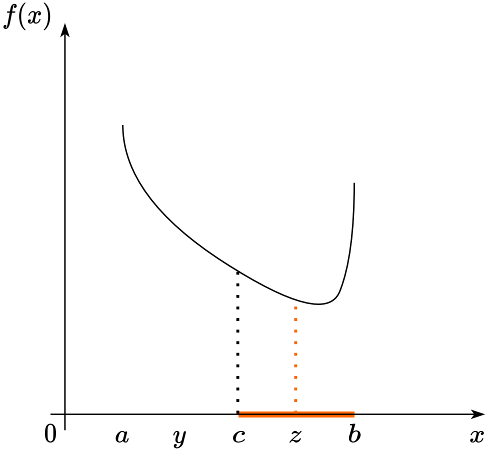


Figure 12: Dichotomy method for unimodal function

Dichotomy method. Algorithm

```python
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
```

Dichotomy method. Bounds

The length of the line segment on $k+1$-th iteration:

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

Dichotomy method. Bounds

The length of the line segment on $k+1$-th iteration:

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

For unimodal functions, this holds if we select the middle of a segment as an output of the iteration x_{k+1}:

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

Dichotomy method. Bounds

The length of the line segment on $k+1$-th iteration:

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

For unimodal functions, this holds if we select the middle of a segment as an output of the iteration x_{k+1}:

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

Note, that at each iteration we ask oracle no more, than 2 times, so the number of function evaluations is $N = 2 \cdot k$, which implies:

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

Dichotomy method. Bounds

The length of the line segment on $k+1$-th iteration:

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

For unimodal functions, this holds if we select the middle of a segment as an output of the iteration x_{k+1}:

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

Note, that at each iteration we ask oracle no more, than 2 times, so the number of function evaluations is $N = 2 \cdot k$, which implies:

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

By marking the right side of the last inequality for ε, we get the number of method iterations needed to achieve ε accuracy:

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

Golden selection

The idea is quite similar to the dichotomy method. There are two golden points on the line segment (left and right) and the insightful idea is, that on the next iteration one of the points will remain the golden point.

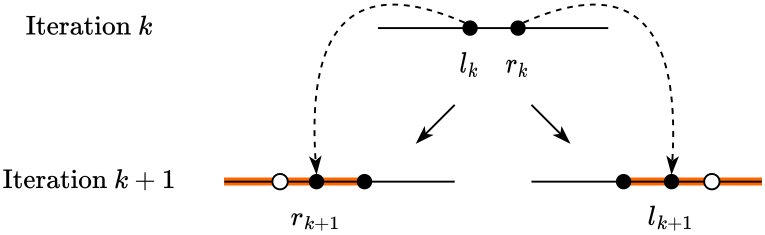


Figure 13: Key idea, that allows us to decrease function evaluations

Golden section. Algorithm

```python
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
```

Golden section. Bounds

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

where $\tau = \frac{\sqrt{5}+1}{2}$.

- The geometric progression constant **more** than the dichotomy method - 0.618 worse than 0.5

Golden section. Bounds

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

where $\tau = \frac{\sqrt{5}+1}{2}$.

- The geometric progression constant **more** than the dichotomy method - 0.618 worse than 0.5
- The number of function calls **is less** than for the dichotomy method - 0.707 worse than 0.618 - (for each iteration of the dichotomy method, except for the first one, the function is calculated no more than 2 times, and for the gold method - no more than one)

Successive parabolic interpolation

Sampling 3 points of a function determines unique parabola. Using this information we will go directly to its minimum. Suppose, we have 3 points $x_1 < x_2 < x_3$ such that line segment $[x_1, x_3]$ contains minimum of a function $f(x)$. Then, we need to solve the following system of equations:

Successive parabolic interpolation

Sampling 3 points of a function determines unique parabola. Using this information we will go directly to its minimum. Suppose, we have 3 points $x_1 < x_2 < x_3$ such that line segment $[x_1, x_3]$ contains minimum of a function $f(x)$. Then, we need to solve the following system of equations:

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

Note, that this system is linear, since we need to solve it on a, b, c. Minimum of this parabola will be calculated as:

Successive parabolic interpolation

Sampling 3 points of a function determines unique parabola. Using this information we will go directly to its minimum. Suppose, we have 3 points $x_1 < x_2 < x_3$ such that line segment $[x_1, x_3]$ contains minimum of a function $f(x)$. Then, we need to solve the following system of equations:

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

Note, that this system is linear, since we need to solve it on a, b, c. Minimum of this parabola will be calculated as:

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

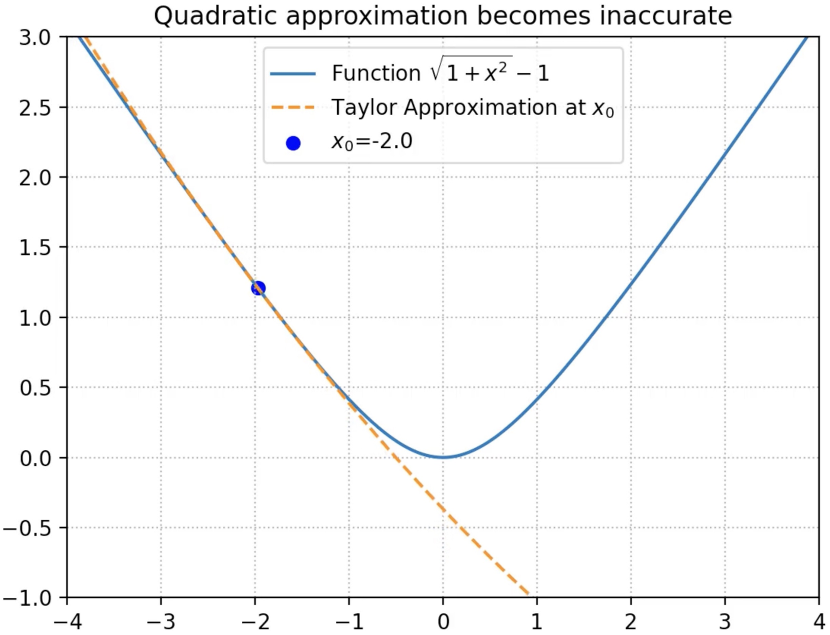
Note, that if $f_2 < f_1, f_2 < f_3$, than u will lie in $[x_1, x_3]$

Successive parabolic interpolation. Algorithm [1]

```
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
```

[1] The convergence of this method is superlinear, but local, which means, that you can take profit from using this method only near some neighbour of optimum. *Here* is the proof of superlinear convergence of order 1.32.



Quadratic approximation becomes inaccurate
Function $\sqrt{1+x^2}-1$
Taylor Approximation at x_0
x_0=-2.0
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−4
−3
−2
−1
0
1
2
3
4

Inexact line search

Sometimes it is enough to find a solution, which will approximately solve out problem. This is very typical scenario for mentioned stepsize selection problem

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

Inexact line search

Sometimes it is enough to find a solution, which will approximately solve out problem. This is very typical scenario for mentioned stepsize selection problem

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

Consider a scalar function $\phi(\alpha)$ at a point x_k:

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

Inexact line search

Sometimes it is enough to find a solution, which will approximately solve out problem. This is very typical scenario for mentioned stepsize selection problem

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

Consider a scalar function $\phi(\alpha)$ at a point x_k:

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

The first-order approximation of $\phi(\alpha)$ near $\alpha = 0$ is:

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

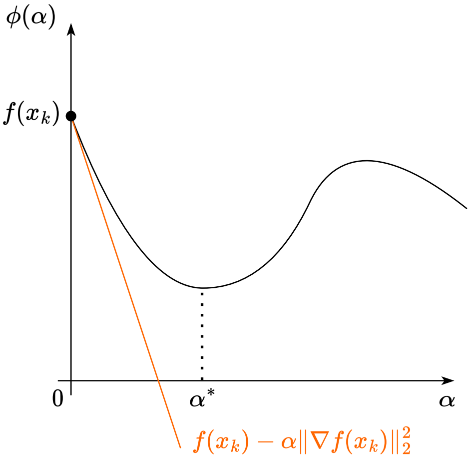


Figure 14: Illustration of Taylor approximation of $\phi_0^I(\alpha)$

Inexact line search. Sufficient Decrease

The inexact line search condition, known as the Armijo condition, states that α should provide sufficient decrease in the function f, satisfying:

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

Inexact line search. Sufficient Decrease

The inexact line search condition, known as the Armijo condition, states that α should provide sufficient decrease in the function f, satisfying:

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

for some constant $c_1 \in (0, 1)$. Note that setting $c_1 = 1$ corresponds to the first-order Taylor approximation of $\phi(\alpha)$. However, this condition can accept very small values of α, potentially slowing down the solution process. Typically, $c_1 \approx 10^{-4}$ is used in practice.

Inexact line search. Sufficient Decrease

The inexact line search condition, known as the Armijo condition, states that α should provide sufficient decrease in the function f, satisfying:

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

for some constant $c_1 \in (0, 1)$. Note that setting $c_1 = 1$ corresponds to the first-order Taylor approximation of $\phi(\alpha)$. However, this condition can accept very small values of α, potentially slowing down the solution process. Typically, $c_1 \approx 10^{-4}$ is used in practice.

Example

If $f(x)$ represents a cost function in an optimization problem, choosing an appropriate c_1 value is crucial. For instance, in a machine learning model training scenario, an improper c_1 might lead to either very slow convergence or missing the minimum.

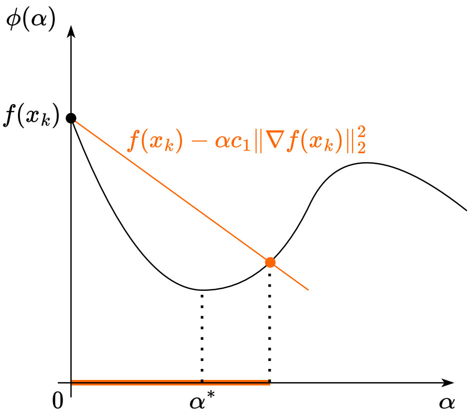


Figure 15: Illustration of sufficient decrease condition with coefficient c_1

Inexact line search. Goldstein Conditions

Consider two linear scalar functions $\phi_1(\alpha)$ and $\phi_2(\alpha)$:

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

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

Inexact line search. Goldstein Conditions

Consider two linear scalar functions $\phi_1(\alpha)$ and $\phi_2(\alpha)$:

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

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

The Goldstein-Armijo conditions locate the function $\phi(\alpha)$ between $\phi_1(\alpha)$ and $\phi_2(\alpha)$. Typically, $c_1 = \rho$ and $c_2 = 1-\rho$, with $\rho \in (0, 0.5)$.

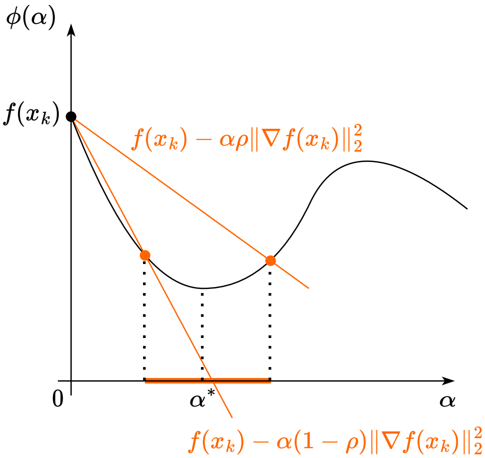


Figure 16: Illustration of Goldstein conditions

Inexact line search. Curvature Condition

To avoid excessively short steps, we introduce a second criterion:

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

Inexact line search. Curvature Condition

To avoid excessively short steps, we introduce a second criterion:

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

for some $c_2 \in (c_1, 1)$. Here, c_1 is from the Armijo condition.
The left-hand side is the derivative $\nabla_\alpha \phi(\alpha)$, ensuring that the slope of $\phi(\alpha)$ at the target point is at least c_2 times the initial slope $\nabla_\alpha \phi(\alpha)(0)$.
Commonly, $c_2 \approx 0.9$ is used for Newton or quasi-Newton methods. Together, the sufficient decrease and curvature conditions form the Wolfe conditions.

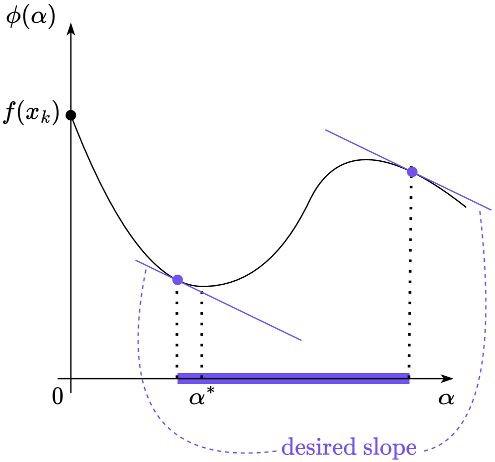


Figure 17: Illustration of curvature condition

Inexact line search. Wolfe Condition

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

Together, the sufficient decrease and curvature conditions form the Wolfe conditions.

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable, and let $\phi(\alpha) = f(x_k - \alpha \nabla f(x_k))$. Assume $\nabla f(x_k)^T p_k < 0$, where $p_k = -\nabla f(x_k)$, making p_k a descent direction. Also, assume f is bounded below along the ray $\{x_k + \alpha p_k \mid \alpha > 0\}$. We aim to show that for $0 < c_1 < c_2 < 1$, there exist intervals of step lengths satisfying the Wolfe conditions.

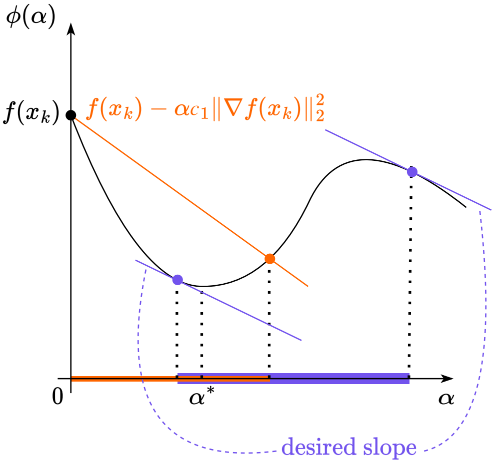


Figure 18: Illustration of Wolfe condition

Inexact line search. Wolfe Condition. Proof

1. Since $\phi(\alpha) = f(x_k + \alpha p_k)$ is bounded below and $l(\alpha) = f(x_k) + \alpha c_1 \nabla f(x_k)^T p_k$ is unbounded below (as $\nabla f(x_k)^T p_k < 0$), the graph of $l(\alpha)$ must intersect the graph of $\phi(\alpha)$ at least once. Let $\alpha' > 0$ be the smallest such value satisfying:

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

This ensures the **sufficient decrease condition** is satisfied.

Inexact line search. Wolfe Condition. Proof

1. Since $\phi(\alpha) = f(x_k + \alpha p_k)$ is bounded below and $l(\alpha) = f(x_k) + \alpha c_1 \nabla f(x_k)^T p_k$ is unbounded below (as $\nabla f(x_k)^T p_k < 0$), the graph of $l(\alpha)$ must intersect the graph of $\phi(\alpha)$ at least once. Let $\alpha' > 0$ be the smallest such value satisfying:

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

   This ensures the **sufficient decrease condition** is satisfied.

2. By the Mean Value Theorem, there exists $\alpha'' \in (0, \alpha')$ such that:

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

   Substituting $f(x_k + \alpha' p_k)$ from (1) into (2), we have:

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

   Dividing through by $\alpha' > 0$, this simplifies to:

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

Inexact line search. Wolfe Condition. Proof

1. Since $\phi(\alpha) = f(x_k + \alpha p_k)$ is bounded below and $l(\alpha) = f(x_k) + \alpha c_1 \nabla f(x_k)^T p_k$ is unbounded below (as $\nabla f(x_k)^T p_k < 0$), the graph of $l(\alpha)$ must intersect the graph of $\phi(\alpha)$ at least once. Let $\alpha' > 0$ be the smallest such value satisfying:

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

   This ensures the **sufficient decrease condition** is satisfied.
2. By the Mean Value Theorem, there exists $\alpha'' \in (0, \alpha')$ such that:

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

   Substituting $f(x_k + \alpha' p_k)$ from (1) into (2), we have:

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

   Dividing through by $\alpha' > 0$, this simplifies to:

   $$\nabla f(x_k + \alpha'' p_k)^T p_k \leq c_1 \nabla f(x_k)^T p_k. \qquad (3)$$
3. Since $c_1 < c_2$ and $\nabla f(x_k)^T p_k < 0$, the inequality $c_1 \nabla f(x_k)^T p_k < c_2 \nabla f(x_k)^T p_k$ holds. This implies there exists α'' such that:

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

   Inequalities (3) and (4) together ensure the Wolfe conditions are satisfied.

Inexact line search. Wolfe Condition. Proof

1. Since $\phi(\alpha) = f(x_k + \alpha p_k)$ is bounded below and $l(\alpha) = f(x_k) + \alpha c_1 \nabla f(x_k)^T p_k$ is unbounded below (as $\nabla f(x_k)^T p_k < 0$), the graph of $l(\alpha)$ must intersect the graph of $\phi(\alpha)$ at least once. Let $\alpha' > 0$ be the smallest such value satisfying:

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

This ensures the **sufficient decrease condition** is satisfied.

2. By the Mean Value Theorem, there exists $\alpha'' \in (0, \alpha')$ such that:

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

Substituting $f(x_k + \alpha' p_k)$ from (1) into (2), we have:

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

Dividing through by $\alpha' > 0$, this simplifies to:

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

3. Since $c_1 < c_2$ and $\nabla f(x_k)^T p_k < 0$, the inequality $c_1 \nabla f(x_k)^T p_k < c_2 \nabla f(x_k)^T p_k$ holds. This implies there exists α'' such that:

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

Inequalities (3) and (4) together ensure the Wolfe conditions are satisfied.

4. For the strong Wolfe conditions, the curvature condition:

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

is met because $\nabla f(x_k + \alpha p_k)^T p_k$ is negative and bounded below by $c_2 \nabla f(x_k)^T p_k$.

Inexact line search. Wolfe Condition. Proof

1. Since $\phi(\alpha) = f(x_k + \alpha p_k)$ is bounded below and $l(\alpha) = f(x_k) + \alpha c_1 \nabla f(x_k)^T p_k$ is unbounded below (as $\nabla f(x_k)^T p_k < 0$), the graph of $l(\alpha)$ must intersect the graph of $\phi(\alpha)$ at least once. Let $\alpha' > 0$ be the smallest such value satisfying:

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

   This ensures the **sufficient decrease condition** is satisfied.

2. By the Mean Value Theorem, there exists $\alpha'' \in (0, \alpha')$ such that:

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

   Substituting $f(x_k + \alpha' p_k)$ from (1) into (2), we have:

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

   Dividing through by $\alpha' > 0$, this simplifies to:

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

3. Since $c_1 < c_2$ and $\nabla f(x_k)^T p_k < 0$, the inequality $c_1 \nabla f(x_k)^T p_k < c_2 \nabla f(x_k)^T p_k$ holds. This implies there exists α'' such that:

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

   Inequalities (3) and (4) together ensure the Wolfe conditions are satisfied.

4. For the strong Wolfe conditions, the curvature condition:

$$\left|\nabla f(x_k + \alpha p_k)^T p_k\right| \leq c_2 \left|\nabla f(x_k)^T p_k\right| \tag{5}$$

   is met because $\nabla f(x_k + \alpha p_k)^T p_k$ is negative and bounded below by $c_2 \nabla f(x_k)^T p_k$.

5. Due to the smoothness of f, there exists an interval around α'' where the Wolfe conditions (and thus the strong Wolfe conditions) hold. Hence, the proof is complete.

Backtracking Line Search

Backtracking line search is a technique to find a step size that satisfies the Armijo condition, Goldstein conditions, or other criteria of inexact line search. It begins with a relatively large step size and iteratively scales it down until a condition is met.

Backtracking Line Search

Backtracking line search is a technique to find a step size that satisfies the Armijo condition, Goldstein conditions, or other criteria of inexact line search. It begins with a relatively large step size and iteratively scales it down until a condition is met.

Algorithm:

1. Choose an initial step size, α_0, and parameters $\beta \in (0, 1)$ and $c_1 \in (0, 1)$.

Backtracking Line Search

Backtracking line search is a technique to find a step size that satisfies the Armijo condition, Goldstein conditions, or other criteria of inexact line search. It begins with a relatively large step size and iteratively scales it down until a condition is met.

Algorithm:

1. Choose an initial step size, α_0, and parameters $\beta \in (0, 1)$ and $c_1 \in (0, 1)$.
2. Check if the chosen step size satisfies the chosen condition (e.g., Armijo condition).

Backtracking Line Search

Backtracking line search is a technique to find a step size that satisfies the Armijo condition, Goldstein conditions, or other criteria of inexact line search. It begins with a relatively large step size and iteratively scales it down until a condition is met.

Algorithm:

1. Choose an initial step size, α_0, and parameters $\beta \in (0, 1)$ and $c_1 \in (0, 1)$.
2. Check if the chosen step size satisfies the chosen condition (e.g., Armijo condition).
3. If the condition is satisfied, stop; else, set $\alpha := \beta\alpha$ and repeat step 2.

Backtracking Line Search

Backtracking line search is a technique to find a step size that satisfies the Armijo condition, Goldstein conditions, or other criteria of inexact line search. It begins with a relatively large step size and iteratively scales it down until a condition is met.

Algorithm:

1. Choose an initial step size, α_0, and parameters $\beta \in (0, 1)$ and $c_1 \in (0, 1)$.
2. Check if the chosen step size satisfies the chosen condition (e.g., Armijo condition).
3. If the condition is satisfied, stop; else, set $\alpha := \beta\alpha$ and repeat step 2.

Backtracking Line Search

Backtracking line search is a technique to find a step size that satisfies the Armijo condition, Goldstein conditions, or other criteria of inexact line search. It begins with a relatively large step size and iteratively scales it down until a condition is met.

Algorithm:

1. Choose an initial step size, α_0, and parameters $\beta \in (0, 1)$ and $c_1 \in (0, 1)$.
2. Check if the chosen step size satisfies the chosen condition (e.g., Armijo condition).
3. If the condition is satisfied, stop; else, set $\alpha := \beta\alpha$ and repeat step 2.

The step size α is updated as

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

in each iteration until the chosen condition is satisfied.

Example

In machine learning model training, the backtracking line search can be used to adjust the learning rate. If the loss doesn't decrease sufficiently, the learning rate is reduced multiplicatively until the Armijo condition is met.

Numerical illustration

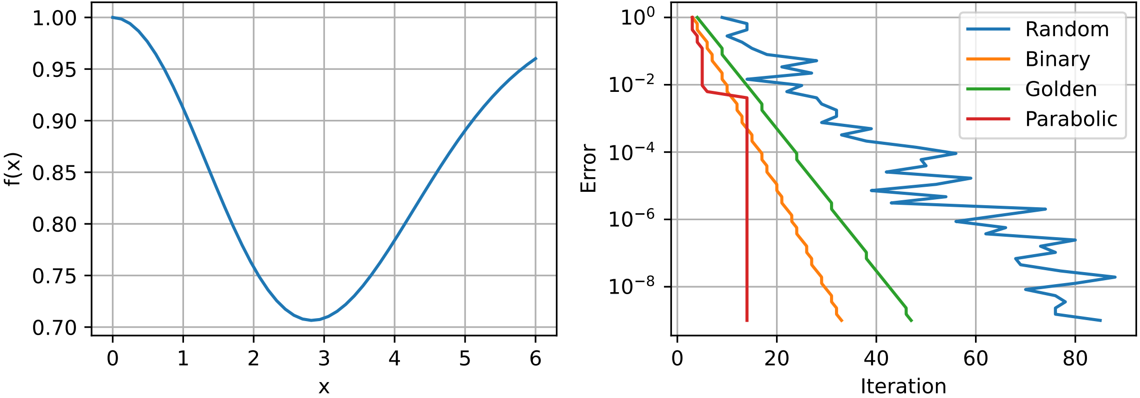


Figure 19: Comparison of different line search algorithms

Open In Colab ♣