Chapter 7 One-dimensional Search Methods

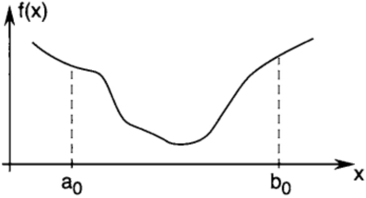
1. Golden Section Search
2. Fibonacci Method
3. Bisection Method
4. Newton's Method
5. Secant Method
6. Bracketing
7. Line Search in Multidimensional Optimization

optimization problem: $\min\{f(x) \mid x \in \Omega\}$

- The function $f : \mathbb{R} \to \mathbb{R}$ is one-dimensional objective/cost function.
- The scalar $x \in \mathbb{R}$ is one-dimensional decision variables.
- The set $\Omega = [a_0, b_0] \subset \mathbb{R}$ is an interval.

★ with an initial point $x^{(0)}$ and produce a sequence of iterates $\{x^{(k)}\}_{k=1}^{\infty}$.

★ Premises: the objective function $f : [a_0, b_0] \to \mathbb{R}$ is unimodal.



Optimization Problem

golden section search

Assume that $f : \Omega \to \mathbb{R}$ is unimodal and $\Omega = [0, 1]$.

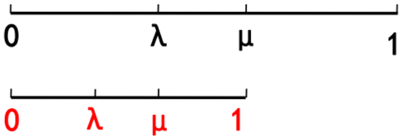
1. insert intermediate points λ and μ in interval $[0, 1]$;
2. marginal intervals are symmetric, i.e., $\lambda = 1 - \mu$;
3. λ (resp. μ) can be used as the counterpart μ (resp. λ) in new interval.

Optimization Problem

golden section search

Assume that $f : \Omega \to \mathbb{R}$ is unimodal and $\Omega = [0, 1]$.

1. insert intermediate points λ and μ in interval $[0, 1]$;
2. marginal intervals are symmetric, i.e., $\lambda = 1 - \mu$;
3. λ (resp. μ) can be used as the counterpart μ (resp. λ) in new interval.

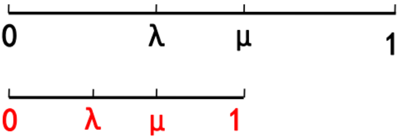


Optimization Problem

golden section search

Assume that $f : \Omega \to \mathbb{R}$ is unimodal and $\Omega = [0, 1]$.

1. insert intermediate points λ and μ in interval $[0, 1]$;
2. marginal intervals are symmetric, i.e., $\lambda = 1 - \mu$;
3. λ (resp. μ) can be used as the counterpart μ (resp. λ) in new interval.



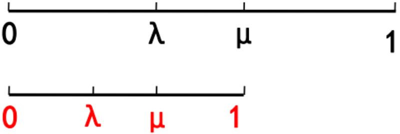
How to derive λ and μ?

Optimization Problem

golden section search

Assume that $f:\Omega\to\mathbb{R}$ is unimodal and $\Omega=[0,1]$.

1. insert intermediate points λ and μ in interval $[0,1]$;
2. marginal intervals are symmetric, i.e., $\lambda=1-\mu$;
3. λ (resp. μ) can be used as the counterpart μ (resp. λ) in new interval.



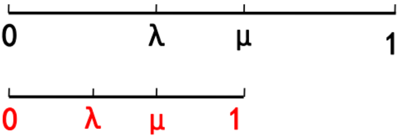
How to derive λ and μ?

$$\begin{cases}\lambda=1-\mu\\ \frac{\mu}{1}=\frac{\lambda}{\mu}\end{cases}$$

golden section search

Assume that $f:\Omega \to \mathbb{R}$ is unimodal and $\Omega = [0,1]$.

1. insert intermediate points λ and μ in interval $[0,1]$;
2. marginal intervals are symmetric, i.e., $\lambda = 1-\mu$;
3. λ (resp. μ) can be used as the counterpart μ (resp. λ) in new interval.

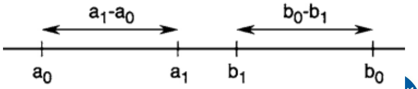


How to derive λ and μ?

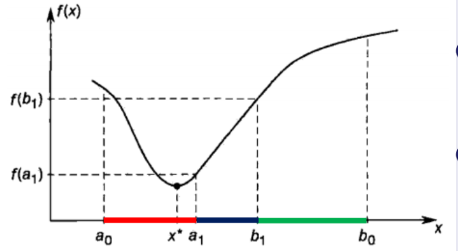
$$\begin{cases} \lambda = 1-\mu \\ \frac{\mu}{1} = \frac{\lambda}{\mu} \end{cases} \Rightarrow \begin{cases} \lambda = \frac{3-\sqrt{5}}{2} \approx 0.382 \\ \mu = \frac{\sqrt{5}-1}{2} \approx 0.618 \end{cases}$$

Question:

1. Why not average the interval $[0,1]$?
2. How to insert point for $[a_0, b_0]$?



How to remove marginal intervals?



1. Case 1: if $f(a_1) < f(b_1)$, then $x^* \in [a_0, b_1]$. Thus, remove the right interval $[b_1, b_0]$.
2. Case 1: if $f(a_1) \geq f(b_1)$, then $x^* \in [a_1, b_0]$. Thus, remove the left interval $[a_0, a_1]$.

Properties of golden section method:

1. At every stage, the function value of f need only be evaluated at one new point.
2. The interval is reduced by the ratio $\mu \approx 0.618$ at every stage, i.e., after N steps, the ratio of interval reduces by the factor μ^N.

Golden Section Search

Example (objective function $f(x) = x^4 - 14x^3 + 60x^2 - 70x$)

Find the minimizer of f in $[0, 2]$ by golden section search. How many iterations need to guarantee the minimizer to locate within a range of 0.3?

Ans: $\because\ 2\mu^N < 0.3 \Longrightarrow N \geq \log_\mu(\frac{0.3}{2}) \approx 4.$

$\therefore$ 4 iterations can guarantee the minimizer to locate within a range of 0.3.

S1: $\begin{cases} a_1 = a_0 + \lambda(b_0 - a_0) = 0.763 \\ b_1 = a_0 + \mu(b_0 - a_0) = 1.236 \end{cases} \Rightarrow \begin{cases} f(a_1) = -24.36 \\ f(b_1) = -18.96 \end{cases}$

$\therefore$ the new interval is $[a_0, b_1] = [0, 1.236]$.

S2: let $b_2 = a_1$, $a_1 = a_0$ and compute

$\begin{cases} a_2 = a_1 + \lambda(b_1 - a_1) = 0.472 \\ f(a_2) = -21.10 \end{cases} \Rightarrow \begin{cases} f(a_2) = -21.10 \\ f(b_2) = -24.36 \end{cases}$

$\therefore$ the new interval is $[a_2, b_1] = [0.472, 1.236]$.

S3: implement by C/C++/matlab/$\cdots$ programs

the minimizer of f over [0,2] is $x^* \approx 0.780884$.

Fibonacci method

Motivation: an improvement of golden section method by replacing fixed (λ, μ) by dynamic (λ_k, μ_k), where $\{\lambda_k\}, \{\mu_k\}$ are sequences satisfying

$$\lim_{k\to\infty} \lambda_k = \lambda \text{ and } \lim_{k\to\infty} \mu_k = \mu.$$

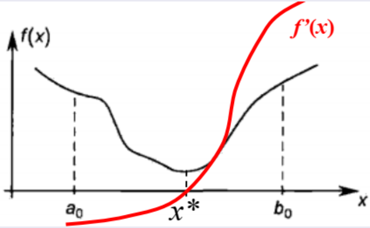
★ E.g., $\lambda_k = 1 - \frac{F_k}{F_{k+1}}$ with F_k as Fibonacci sequence.

Bisection method

Let $f : \Omega \to \mathbb{R}$ is unimodal with $\Omega = [a_0, b_0]$. Moreover, $f \in \mathcal{C}^1(\Omega)$.
Motivation: find $x^* = \arg\min_{x\in\Omega} f$ by solving the 1st-order necessary condition, i.e., $f'(x^*) = 0$.

Bisection Method

pseudo codes of bisection method



1. compute the midpoint $x^{(0)} = \frac{a_0+b_0}{2}$.
2. evaluate $f'(x^0)$, and check

$$x^* \in \begin{cases} [a_0, x^{(0)}], & \text{if } f'(x^{(0)}) > 0, \\ [x^{(0)}, b_0], & \text{if } f'(x^{(0)}) < 0, \\ \text{is minimizer}, & \text{if } f'(x^{(0)}) = 0. \end{cases}$$

different between bisection method and golden section method

- instead of using values of f, the bisection methods uses values of f'.
- at each iteration, the length of the uncertainty interval is reduced by a factor of $\frac{1}{2}$, which is smaller than μ for the golden section method.

Example (objective function $f(x) = x^4 - 14x^3 + 60x^2 - 70x$)

Find the minimizer of f in $[0, 2]$ by bisection metod. How many iterations need to guarantee the minimizer to locate within a range of 0.3?

Ans: $\because\ 2(\frac{1}{2})^N < 0.3 \Longrightarrow N \geq \log_2(\frac{0.3}{2}) \approx 3.$

$\therefore$ 3 iterations need to guarantee the minimizer to locate within a range of 0.3.

S1: $x^{(0)} = \frac{a_0+b_0}{2} = 1 \Longrightarrow f'(x^{(0)}) = 12$
$\therefore$ the new interval is $[a_0, x^{(0)}] = [0, 1]$.

S2: let $a_1 = a_0$, $b_1 = x^{(0)}$ and compute
$x^{(1)} = \frac{a_1+b_1}{2} = 0.5 \Longrightarrow f'(x^{(1)}) = -20$
$\therefore$ the new interval is $[x^{(1)}, b_1] = [0.5, 1]$.

S3: let $a_2 = x^{(1)}$, $b_2 = b_1$ and compute
$x^{(2)} = \frac{a_2+b_2}{2} = 0.75 \Longrightarrow f'(x^{(2)}) = -1.937$
$\therefore$ the new interval is $[x^{(2)}, b_2] = [0.75, 1]$.

The minimizer of f over [0,2] is $x^* \approx 0.780884$.

Newton's Method

Newton's method

Let $f : \Omega \to \mathbb{R}$ is unimodal with $\Omega = [a_0, b_0]$. Moreover, $f \in \mathcal{C}^2(\Omega)$.
Motivation: find $x^* = \arg\min_{x \in \Omega} f$ by a quadratic approximation of f at some point.

Given iterate $x^{(k)}$, let $q(x)$ be the quadratic approximation of f at $x^{(k)}$,

$$f(x) \approx q(x) = f(x^{(k)}) + f'(x^{(k)})(x - x^{(k)}) + \frac{f''(x^{(k)})}{2}(x - x^{(k)})^2,$$

The minimizer of $q(x)$ is defined as the new iterate $x^{(k+1)}$.
By 1st-order necessary condition, we have

$$0 = q'(x^{(k+1)}) = f'(x^{(k)}) + f''(x^{(k)})(x^{(k+1)} - x^{(k)})^2.$$

Thus, the iterative scheme of Newton method is

$$x^{(k+1)} = x^{(k)} - \frac{f'(x^{(k)})}{f''(x^{(k)})}, \quad k = 0, 1, \ldots$$

Newton's Method

Example (objective function $f(x) = \frac{1}{2}x^2 - \sin x$)

Find the minimizer of f by Newton's method with $x^{(0)} = 0.5$, and the accuracy $|x^{(k+1)} - x^{(k)}| < \varepsilon = 10^{-5}$.

Ans: $\because f'(x) = x - \cos x$ and $f''(x) = 1 + \sin x$. With initial point $x^{(0)} = 0.5$

S1: $x^{(1)} = x^{(0)} - \frac{f'(x^{(0)})}{f''(x^{(0)})} = 0.5 - \frac{0.5 - \cos 0.5}{1 + \sin 0.5} = 0.7552.$

S2: $x^{(2)} = x^{(1)} - \frac{f'(x^{(1)})}{f''(x^{(1)})} = 0.7391.$

S3: $x^{(3)} = x^{(2)} - \frac{f'(x^{(2)})}{f''(x^{(2)})} = 0.7390.$

S4: $x^{(4)} = x^{(3)} - \frac{f'(x^{(3)})}{f''(x^{(3)})} = 0.7390.$

$\because\ |x^{(4)} - x^{(3)}| < \varepsilon = 10^{-5}$, we stop the iteration.

Moreover, $f'(x^{(4)}) = -8.6 \times 10^{-6} \approx 0$ and $f''(x^{(4)}) = 1.673 > 0$,

$\therefore\ x^{(4)}$ is an "good" approximation of strict minimizer.

★ Newton's method works well if $f''(x) > 0$ everywhere. If $f''(x) < 0$ for some x, Newton's method may diverge.

application of Newton's method to nonlinear equation $g(x) = 0$

Given iterate $x^{(k)}$, let $l(x)$ be the linear approximation of g at $x^{(k)}$,

$$g(x) \approx l(x) = g(x^{(k)}) + g'(x^{(k)})(x - x^{(k)}),$$

The zero-point of $l(x)$ is defined as the new iterate $x^{(k+1)}$. Thus,

$$x^{(k+1)} = x^{(k)} - \frac{g(x^{(k)})}{g'(x^{(k)})}, \quad k = 0, 1, \ldots$$

Definition (bracket)

An interval $[a, b]$ in which the minimizer lie in.

finding a bracket $[a, b]$ of unimodal function f

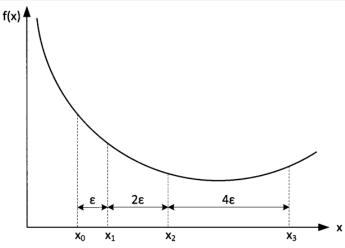
Aim: find three points $a < c < b$ such that $f(c) < f(a)$ and $f(c) < f(b)$.

S0: pick three arbitrary points $x_0 < x_1 < x_2$.

S1.1 If $f(x_1) < f(x_0)$ and $f(x_1) < f(x_2)$, then $[a, b] = [x_0, x_2]$.

S1.2 If not, e.g., $f(x_0) > f(x_1) > f(x_2)$, then pick a point $x_3 > x_2$.

- S1.2.1 if $f(x_2) < f(x_3)$, then $[a, b] = [x_1, x_3]$.
- S1.2.2 Otherwise, continue this process until the function increases.

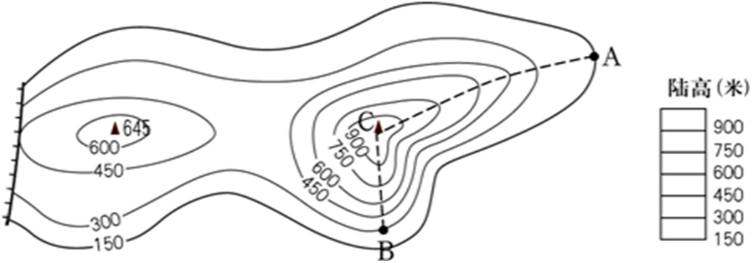


★ new point expands distance between successive test points.

Unconstrained Optimization

unconstrained optimization: $\min\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in \mathbb{R}^n\}$

- The function $f : \mathbb{R}^n \to \mathbb{R}$ is n-dimensional objective/cost function.
- The vector $\boldsymbol{x} \in \mathbb{R}^n$ is n-dimensional decision variables.



★ Premises: the objective function $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differential.

optimality condition

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

If f is convex around optima, then it is also sufficient.

Line Search

iterative scheme of line search

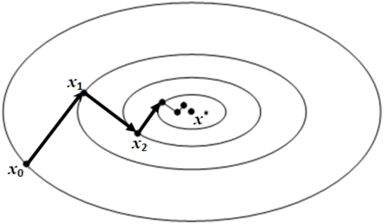
Generate an iterative sequence $\{\boldsymbol{x}_k\}_{k=1}^{\infty} \subset \mathbb{R}^n$ via
$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k$,

- α_k — step size or step length,
- $\boldsymbol{p}_k$ — searching direction.

If $\boldsymbol{p}_k$ and α_k are "properly" selected or designed, it can be guaranteed that $\lim_{k\to\infty} \|\boldsymbol{x}_k - \boldsymbol{x}^*\| = 0$, i.e., the limit of the sequence is a minimizer.

Basic procedures:

1. determine a searching direction $\boldsymbol{p}_k$;
2. find the optimal step size α_k such that the objective function is minimum along the given searching direction $\boldsymbol{p}_k$.



Choice of Searching Direction $\boldsymbol{p}_k$

descent property

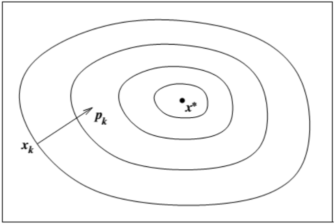
If the angle between $\boldsymbol{p}_k$ and $\nabla f(\boldsymbol{x}_k)$ in $(\frac{\pi}{2}, \pi)$, or equivalently $\boldsymbol{p}_k^\top \nabla f(\boldsymbol{x}_k) < 0$, then $\boldsymbol{p}_k$ is a descent direction of $f(\boldsymbol{x})$ at $\boldsymbol{x}_k$.

proof. because f is differentiable, by Taylor expansion, we have

$$f(\boldsymbol{x}_k + \alpha \boldsymbol{p}_k) = f(\boldsymbol{x}_k) + \alpha \boldsymbol{p}_k^\top \nabla f(\boldsymbol{x}_k) + o(\|\alpha \boldsymbol{p}_k\|)$$

some state-of-the-art descent direction

- gradient: $\boldsymbol{p}_k = -\nabla f(\boldsymbol{x}_k)$;
- conjugate gradient: $\boldsymbol{p}_k = -\nabla f(\boldsymbol{x}_k) + \beta_k \boldsymbol{p}_{k-1}$;
- Newton direction: $\boldsymbol{p}_k = -(\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k)$;
- quasi-Newton direction: $\boldsymbol{p}_k = -(\boldsymbol{B}_k)^{-1} \nabla f(\boldsymbol{x}_k)$, where $\boldsymbol{B}_k \approx \nabla^2 f(\boldsymbol{x}_k)$;
- $\cdots\cdots$



choice of stepsize

- exact stepsize:

  $\alpha_k = \arg\min_{\alpha>0} \; f(\boldsymbol{x}_k + \alpha \boldsymbol{p}_k)$,

  where $\phi(\alpha) := f(\boldsymbol{x}_k + \alpha \boldsymbol{p}_k)$.

exact stepsize when minimizing quadratic function

$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c$, where $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is symmetric and $\boldsymbol{A} \succ 0$.

Given the iterate $\boldsymbol{x}_k$ and descent direction $\boldsymbol{p}_k$, what is the exact stepsize α_k?

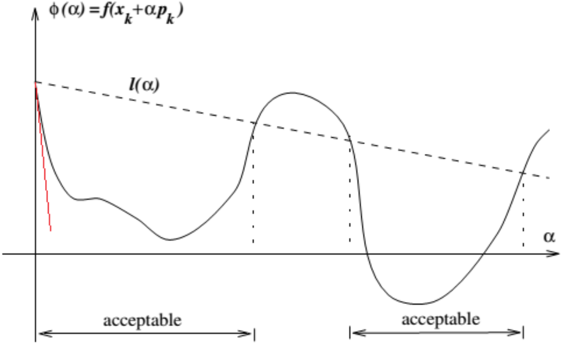
Ans: $\because \; \nabla f(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ and $\phi(\alpha) := f(\boldsymbol{x}_k + \alpha \boldsymbol{p}_k)$,

$$\therefore \; \phi'(\alpha) = \boldsymbol{p}_k^\top \nabla f(\boldsymbol{x}_k + \alpha \boldsymbol{p}_k) = \boldsymbol{p}_k^\top \nabla f(\boldsymbol{x}_k) + \alpha \boldsymbol{p}_k^\top \boldsymbol{A}\boldsymbol{p}_k = 0,$$

$$\therefore \; \alpha = -\frac{\boldsymbol{p}_k^\top \nabla f(\boldsymbol{x}_k)}{\boldsymbol{p}_k^\top \boldsymbol{A}\boldsymbol{p}_k}.$$

Choice of Stepsize α_k

- exact stepsize α_k: $\alpha_k = \arg\min_{\alpha>0} \phi(\alpha) := f(\boldsymbol{x}_k + \alpha \boldsymbol{p}_k)$
- inexact stepsize α_k:

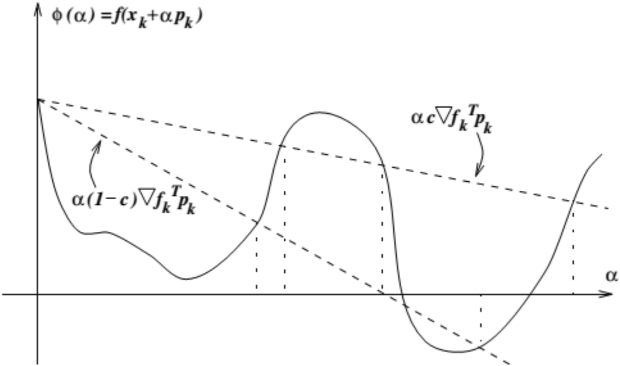


Armijo criterion

$$f(\boldsymbol{x}_k + \alpha \boldsymbol{p}_k) \leq f(\boldsymbol{x}_k) + c\alpha \boldsymbol{p}_k^\top \nabla f(\boldsymbol{x}_k), \ c \in (0,1).$$

Choice of Stepsize α_k

- exact stepsize α_k: $\alpha_k = \arg\min_{\alpha>0} \ \phi(\alpha) := f(\boldsymbol{x}_k + \alpha \boldsymbol{p}_k)$
- inexact stepsize α_k:

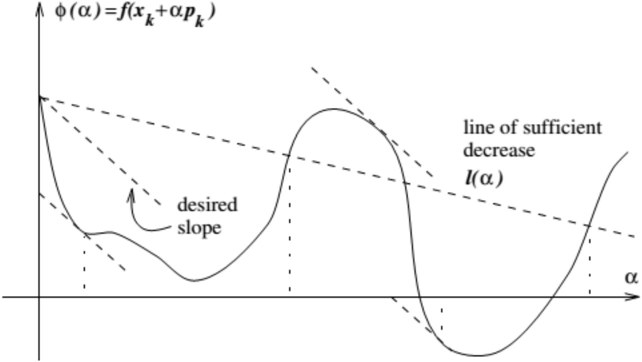


Goldstein criterion

1. $f(x_k + \alpha p_k) \le f(x_k) + c\alpha p_k^\top \nabla f(x_k),$
2. $f(x_k + \alpha p_k) \ge f(x_k) + (1-c)\alpha p_k^\top \nabla f(x_k), \ c \in (0, 0.5).$

Choice of Stepsize α_k

- exact stepsize α_k: $\alpha_k = \arg\min_{\alpha>0} \ \phi(\alpha) := f(\boldsymbol{x}_k + \alpha \boldsymbol{p}_k)$
- inexact stepsize α_k:



Wolfe criterion

1. $f(\boldsymbol{x}_k + \alpha \boldsymbol{p}_k) \leq f(\boldsymbol{x}_k) + c_1 \alpha \boldsymbol{p}_k^\top \nabla f(\boldsymbol{x}_k)$,
   $\boldsymbol{p}_k^\top \nabla f(\boldsymbol{x}_{k+1}) \geq c_2 \boldsymbol{p}_k^\top \nabla f(\boldsymbol{x}_k)$, $0 < c_1 < c_2 < 1$.

backtracing for α_k

Require: Initial stepsize $\bar{\alpha} > 0$, $\rho \in (0,1)$, $c \in (0,1)$. Set $\alpha = \bar{\alpha}$.

1: **while** $f(\boldsymbol{x}_k + \alpha \boldsymbol{p}_k) > f(\boldsymbol{x}_k) + c\alpha \boldsymbol{p}_k^\top \nabla f(\boldsymbol{x}_k)$ **do**

2: $\quad \alpha \leftarrow \rho\alpha$;

3: **end while**

4: $\alpha_k \leftarrow \alpha$

Theorem (convergence analysis)

Suppose that the objection function $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable with Lipschitz continuous gradient, for the iterative scheme $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k$, if $\boldsymbol{p}_k$ is a descent direction and α_k satisfies that Wolfe criterion, then $\sum_{k=1}^{\infty} \cos^2\theta_k \|\nabla f(\boldsymbol{x}_k)\|^2 < +\infty$, where θ_k is the angle between $\boldsymbol{p}_k$ and $\nabla f(\boldsymbol{x}_k)$.

Homework

Exercise in textbook: 7.10, 7.12