

无约束优化问题1

翟卫欣 副教授

zhaiweixin@cau.edu.cn

中国农业大学信息与电气工程学院

目录

无约束最优化问题

一阶必要条件

二阶必要条件

二阶充分条件

凸充分性定理

线性搜索问题

非精确线性搜索方法

下降算法的全局收敛性与收敛速率

目录

无约束最优化问题

一阶必要条件

二阶必要条件

二阶充分条件

凸充分性定理

线性搜索问题

非精确线性搜索方法

下降算法的全局收敛性与收敛速率

无约束最优化问题

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

无约束最优化问题

局部最优解

- 设 $x^* \in \mathbf{R}^n$, 若存在 x^* 的 $\delta (\delta > 0)$ 邻域 $N_\delta(x^*) = \{x | \|x - x^*\| \leq \delta\}$,
使得 $f(x) \geq f(x^*), \forall x \in N_\delta(x^*)$,
则称 x^* 是 $f(x)$ 的局部最优解 (局部解);
- 若 $f(x) > f(x^*), \forall x \in N_\delta(x^*)$,
则称 x^* 是 $f(x)$ 的严格局部最优解 (严格局部解)。

全局最优解

- 若对任意 $x \in \mathbf{R}^n$, 有 $f(x) \geq f(x^*)$,
则称 x^* 是 $f(x)$ 的全局最优解 (全局解);
 - 若 $f(x) > f(x^*)$,
则称 x^* 是 $f(x)$ 的严格全局最优解 (严格全局解)。
- 求全局解较复杂, 无约束最优化方法通常只求局部解。

目录

无约束最优化问题

一阶必要条件

二阶必要条件

二阶充分条件

凸充分性定理

线性搜索问题

非精确线性搜索方法

下降算法的全局收敛性与收敛速率

一阶必要条件

- 设 $f(x)$ 一阶连续可微，若 x^* 是无约束问题的局部解，则 $\nabla f(x^*) = \mathbf{0}$
- 证明：对任意的 $d \neq \mathbf{0}, d \in R^n$ ，构造单变量函数 $\varphi(\alpha) = f(x^* + \alpha d)$ ，因为 x 是约束问题的局部解，所以 $\alpha = 0$ 是一元函数 $\varphi(\alpha)$ 的局部极小点，由一元函数极小点的必要条件，有：

$$0 = \varphi'(0) = \langle \nabla f(x^*), d \rangle = d^T \nabla f(x^*), \forall d \neq \mathbf{0}, d \in R^n$$

因为 d 的任意性，所以 $\nabla f(x^*) = \mathbf{0}$ ，即在 x^* 位置上的任何方向上的导数均为0。此处 $\langle \nabla f(x^*), d \rangle$ 代表内积。

一阶必要条件举例



一阶必要条件不是充分条件，例如：

$$\min\{f(\mathbf{x}) = x_1^2 + x_2^3\}, \mathbf{x} \in \mathbf{R}^2$$

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 2x_1 \\ 3x_2^2 \end{bmatrix}$$

目录

无约束最优化问题

一阶必要条件

二阶必要条件

二阶充分条件

凸充分性定理

线性搜索问题

非精确线性搜索方法

下降算法的全局收敛性与收敛速率

二阶必要条件

- 设 $f(\mathbf{x})$ 二阶连续可微，若 \mathbf{x}^* 是无约束问题的局部解，则 $\nabla f(\mathbf{x}^*) = \mathbf{0}$, $\nabla^2 f(\mathbf{x}^*)$ 为半正定。
- 证明：对任意的 $\mathbf{d} \neq \mathbf{0}, \mathbf{d} \in \mathbf{R}^n$ ，由于 \mathbf{x} 是约束问题的局部解，所以存在 $\varepsilon > 0$ ，使得 $0 < \alpha < \varepsilon$ 时，有

$f(\mathbf{x}^* + \alpha\mathbf{d}) \geq f(\mathbf{x}^*)$ ，另一方面，二阶Taylor展开式 $f(\mathbf{x}^* + \alpha\mathbf{d}) = f(\mathbf{x}^*) + \alpha\mathbf{d}^T \nabla f(\mathbf{x}^*) + \frac{1}{2}\alpha^2 \mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} + o(\alpha^2)$

由于 $\nabla f(\mathbf{x}^*) = \mathbf{0}$ ，所以 $\frac{1}{2}\alpha^2 \mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} + o(\alpha^2) \geq 0$ ，

令 $\alpha \rightarrow 0^+$ ，得到 $\mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} + \frac{o(\alpha^2)}{\alpha^2} \geq 0, \forall \mathbf{d} \in \mathbf{R}^n$ ，

即 $\nabla^2 f(\mathbf{x}^*)$ 为半正定。

二阶必要条件举例

- $\min\{f(\mathbf{x}) = \frac{1}{2}x_1^2 + x_1x_2 + 2x_2^2 - 4x_1 - 4x_2 - x_2^3\}, \mathbf{x} \in \mathbb{R}^2$

$$\nabla f(\mathbf{x}) = \begin{bmatrix} x_1 + x_2 - 4 \\ x_1 + 4x_2 - 4 - 3x_2^2 \end{bmatrix}$$

- Hesse矩阵为

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} 1 & 1 \\ 1 & 4 - 6x_2 \end{bmatrix}$$

- 令 $\nabla f(\mathbf{x}) = \mathbf{0}$ 可得 $\mathbf{x} = \begin{bmatrix} 4 \\ 0 \end{bmatrix}$ 或者 $\mathbf{x} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ 。

- 但是当 $\mathbf{x} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ 时, $\nabla^2 f(\mathbf{x}) = \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix}$

不是半正定矩阵, 仅有 $\mathbf{x} = \begin{bmatrix} 4 \\ 0 \end{bmatrix}$ 可能为解。

二阶必要条件举例

- 满足 $\nabla f(\mathbf{x}^*) = \mathbf{0}$ 的点 \mathbf{x}^* 称为 f 的平稳点或驻点，该 \mathbf{x}^* 可能是极大点、极小值点或既不极大又不极小的点（鞍点）。

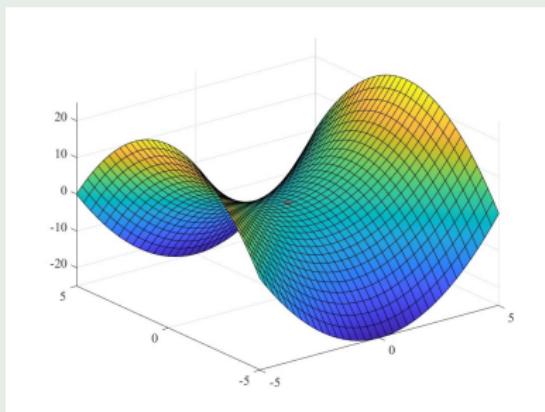


图 1: 鞍点示例

目录

无约束最优化问题

一阶必要条件

二阶必要条件

二阶充分条件

凸充分性定理

线性搜索问题

非精确线性搜索方法

下降算法的全局收敛性与收敛速率

二阶充分条件

- 设 $f(\mathbf{x})$ 二阶连续可微，且 $\nabla f(\mathbf{x}^*) = \mathbf{0}$, $\nabla^2 f(\mathbf{x}^*)$ 正定，则 \mathbf{x}^* 是无约束问题的严格局部解。
- 证明：令 \mathbf{d} 为任意方向，其中 $\|\mathbf{d}\| = 1$ ，再令 $\mathbf{x} = \mathbf{x}^* + \lambda\mathbf{d}$, $f(\mathbf{x})$ 在 \mathbf{x}^* 处的泰勒展开为：

$$f(\mathbf{x}) = f(\mathbf{x}^* + \lambda\mathbf{d}) = f(\mathbf{x}^*) + \lambda\mathbf{d}^T \nabla f(\mathbf{x}^*) + \frac{1}{2}\lambda^2 \mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} + o(\lambda^2),$$

其中 $\lambda > 0$

因为 $\nabla f(\mathbf{x}^*) = \mathbf{0}$, 所以

$$f(\mathbf{x}) - f(\mathbf{x}^*) = \frac{1}{2}\lambda^2 \mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} + o(\lambda^2)$$

由于 $\nabla^2 f(\mathbf{x}^*)$ 正定，所以 $\frac{f(\mathbf{x}) - f(\mathbf{x}^*)}{\lambda^2} = \frac{1}{2}\mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} + \frac{o(\lambda^2)}{\lambda^2} > 0$

所以 $f(\mathbf{x})$ 在 \mathbf{x}^* 可取到局部最小值 $f(\mathbf{x}^*)$ 。

二阶充分条件举例

对于如下问题 $\min\{f(\mathbf{x}) = \frac{1}{2}x_1^2 + x_1x_2 + 2x_2^2 - 4x_1 - 4x_2 - x_2^3\}, \mathbf{x} \in \mathbb{R}^2$

在 $\mathbf{x} = \begin{bmatrix} 4 \\ 0 \end{bmatrix}$ 时, Hesse矩阵为

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix} \text{ 正定}$$

该点是局部最优解。

目录

无约束最优化问题

一阶必要条件

二阶必要条件

二阶充分条件

凸充分性定理

线性搜索问题

非精确线性搜索方法

下降算法的全局收敛性与收敛速率

凸充分性定理

- 设 $f(x)$ 是一阶连续可微的凸函数，则 $\nabla f(x^*) = \mathbf{0}$ 是 x^* 为全局解的充分必要条件。

证明： $f(x) \geq f(x^*) + \nabla f(x^*)^T(x - x^*) = f(x^*)$

(参考lecture note 2的“凸函数的判别定理2”章节)

无约束最优化问题的求解

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

- 常用方法是构造序列 $\{x^{(k)}\}$, 使其满足:
$$\lim_{k \rightarrow \infty} f(x^{(k)}) = f(x^*) = \min_{x \in \mathbb{R}^n} f(x), \quad \lim_{k \rightarrow \infty} x^{(k)} = x^*$$
称 x^* 是问题的解,
 $\{x^{(k)}\}$ 是极小化点列。极小化点列的构造方法通常采取逐步构造法, 即取 $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}, k = 1, 2, \dots$
- 其中, $d^{(k)}$ 为 $x^{(k)}$ 处的搜索方向, α_k 为沿 $d^{(k)}$ 方向的步长。采用不同的方法构造 $d^{(k)}$ 和确定 α_k 对应不同的算法。

目录

无约束最优化问题

一阶必要条件

二阶必要条件

二阶充分条件

凸充分性定理

线性搜索问题

非精确线性搜索方法

下降算法的全局收敛性与收敛速率

线性搜索问题

- 假定在 $x^{(k)}$ 处的方向 $d^{(k)}$ 已确定，如何沿着 $d^{(k)}$ 选取合适的步长 α_k 以确定 $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}, k = 1, 2, \dots$ 且满足 $f(x^{(k+1)}) < f(x^{(k)})$ 的问题称为**线性搜索问题**。
- 线性搜索问题常分为**精确线性搜索、非精确线性搜索问题**。
- 以一维函数为例介绍**二分法、黄金分割法、斐波那契数列法、牛顿法、割线法**等精确线性搜索方法以及**Armijo方法、Goldstein方法、Wolfe-Powell方法**等非精确线性搜索方法。

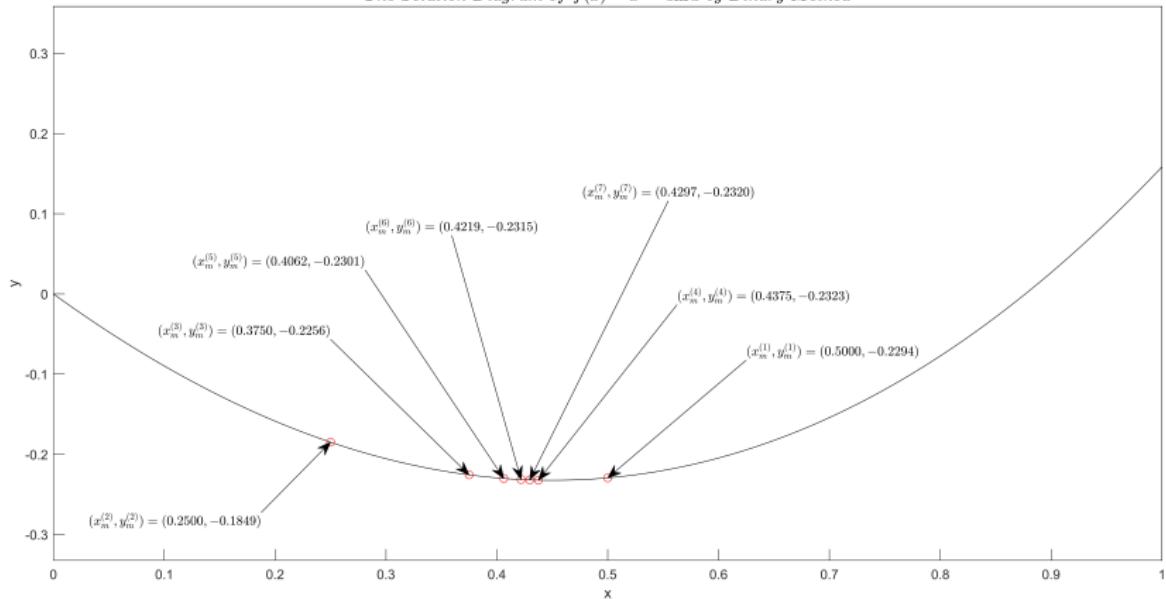
二分法

- 针对一元单值函数 $f : \mathbb{R} \rightarrow \mathbb{R}$, 求 $[a_0, b_0]$ 上的极小点。
- 前提: f 在 $[a_0, b_0]$ 上是单峰的, 并且是连续可微的, 一阶导数为 f'
- 步骤 (令 x^* 为极小值):

取 $x^{(k)} = \frac{a_k + b_k}{2}$, f 在 $x^{(k)}$ 处的导数为 $f'(x^{(k)})$,

$$\begin{cases} f'(x^{(k)}) > 0 \Rightarrow x^* \in [a_k, x^{(k)}] \Rightarrow [a_{k+1}, b_{k+1}] = [a_k, x^{(k)}] \\ f'(x^{(k)}) < 0 \Rightarrow x^* \in [x^{(k)}, b_k] \Rightarrow [a_{k+1}, b_{k+1}] = [x^{(k)}, b_k] \\ f'(x^{(k)}) = 0 \Rightarrow x^* = x^{(k)} \end{cases}$$

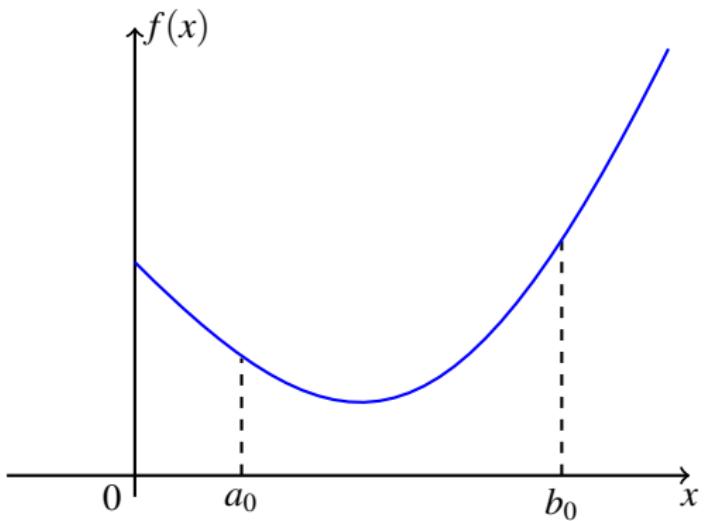
二分法举例

The Solution Diagram of $f(x) = x^2 - \sin x$ by Binary Method

黄金分割法

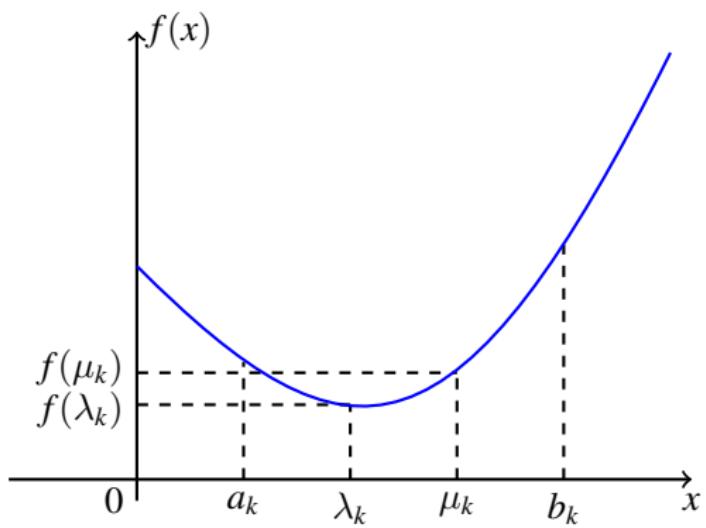
针对一元单值函数 $f : \mathbb{R} \rightarrow \mathbb{R}$, 求 $[a_0, b_0]$ 上的极小点。

前提: f 在 $[a_0, b_0]$ 上是单峰的。



可否尽可能减少计算次数?

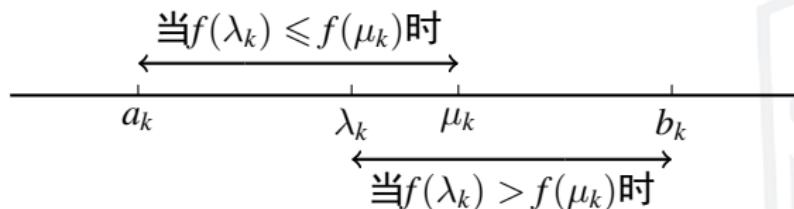
黄金分割法



$$|\lambda_k - a_k| = |b_k - \mu_k| = \rho |b_k - a_k| \text{ 其中 } 0 < \rho < \frac{1}{2}.$$

- 由于函数值先单调下降，后单调上升，所以如果 $f(\lambda_k) \leq f(\mu_k)$ ，则极小值在 $[a_k, \mu_k]$ 可取到；否则，极小值在 $[\lambda_k, b_k]$ 可取到， $k = 0, 1, 2, \dots$

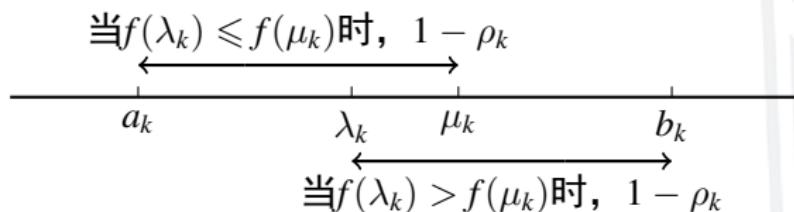
黄金分割法



- 当 $f(\lambda_k) \leq f(\mu_k)$ 时, $[a_{k+1}, b_{k+1}] = [a_k, \mu_k]$;
- 当 $f(\lambda_k) > f(\mu_k)$ 时, $[a_{k+1}, b_{k+1}] = [\lambda_k, b_k]$ 。
- 在迭代过程中, 我们希望能尽量少地减少对于 f 函数的计算, 所以希望在进行第 $k + 1$ 次分割时, 第 k 次分割点的数值也能用上。
- 当 $[a_{k+1}, b_{k+1}] = [a_k, \mu_k]$ 时, 我们希望 $\mu_{k+1} = \lambda_k$;
- 当 $[a_{k+1}, b_{k+1}] = [\lambda_k, b_k]$ 时, 我们希望 $\lambda_{k+1} = \mu_k$ 。
- 令 $\overline{a_k, b_k} = 1$, 第 k 次分割时, $\overline{\lambda_k, \mu_k} = 1 - 2\rho$; 第 $k + 1$ 次分割时, $\overline{\lambda_k, \mu_k} = \rho(1 - \rho)$ 。解得 $\rho = \frac{3-\sqrt{5}}{2}$, 为黄金比例分割。

斐波那契数列法

利用黄金分割法进行区间压缩的过程中， ρ 保持不变。如果不限制 ρ 保持不变，对于第 k 次迭代需确定一个参数 ρ_k 。



$|\lambda_k - a_k| = |b_k - \mu_k| = \rho_k |b_k - a_k|$ 其中 $0 < \rho_k < \frac{1}{2}$ 。

- 在迭代过程中，我们希望能尽量减少对于 f 函数的计算，所以希望在进行第 $k+1$ 次分割时，第 k 次分割点的数值也能用上。
- 当 $[a_{k+1}, b_{k+1}] = [a_k, \mu_k]$ 时，我们希望 $\mu_{k+1} = \lambda_k$ ；
- 当 $[a_{k+1}, b_{k+1}] = [\lambda_k, b_k]$ 时，我们希望 $\lambda_{k+1} = \mu_k$ 。
- 令 $\overline{a_k, b_k} = 1$ ，第 k 次分割时， $\overline{\lambda_k, \mu_k} = 1 - 2\rho_k$ ；第 $k+1$ 次分割时， $\overline{\lambda_k, \mu_k} = \rho_{k+1}(1 - \rho_k)$ 。

斐波那契数列法

$$\rho_{k+1} = 1 - \frac{\rho_k}{1 - \rho_k}$$

- 存在多个序列能满足该条件，例如 $\rho_k = \frac{3-\sqrt{5}}{2}$ ，再例如 $\rho_k = 1 - \frac{F_{N-k+1}}{F_{N-k+2}}$ ，其中 F_i 是斐波那契数列， N 是总迭代次数。

- 斐波那契数列： $F_{i+1} = F_i + F_{i-1}, i \geq 0$ 。

$F_{-1} = 0, F_0 = 1, F_1 = 1, F_2 = 2, F_3 = 3, F_4 = 5, F_5 = 8, F_6 = 13, F_7 = 21, F_8 = 34, \dots$

- 证明斐波那契数列满足该条件：

$$\begin{aligned}\rho_{k+1} &= 1 - \frac{F_{N-k}}{F_{N-k+1}} = 1 - \left(1 - \frac{F_{N-k+1}}{F_{N-k+2}}\right) / \left(1 - \left(1 - \frac{F_{N-k+1}}{F_{N-k+2}}\right)\right) \\ &\Leftrightarrow F_{N-k+2} = F_{N-k} + F_{N-k+1} , \text{ 其中 } F_i > 0, i = 1, 2, 3, \dots.\end{aligned}$$

斐波那契数列法

- 斐波那契数列法 N 次压缩的总压缩比为：

$$(1 - \rho_1)(1 - \rho_2)(1 - \rho_3) \cdots (1 - \rho_N) = \frac{F_N}{F_{N+1}} \frac{F_{N-1}}{F_N} \cdots \frac{F_1}{F_2} = \frac{1}{F_{N+1}}$$

- 但是由于 $\rho_N = \frac{1}{2}$, 两个中间点重合, 无法压缩, 所以常令 $\rho_N = \frac{1}{2} - \varepsilon$, 其中 ε 是很小的正实数。则 $1 - \rho_N = \frac{1+2\varepsilon}{2}$, 总压缩比为 $\frac{1+2\varepsilon}{F_{N+1}}$ 。

黄金分割法和斐波那契数列法举例

- 求目标函数在 $[0, 2]$ 上的极小点，要求极小点所在区间的长度压缩到0.3以内

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

黄金分割法举例

经过 N 次压缩可以在 $[0, 2]$ 上取得极小点 $(\frac{\sqrt{5}-1}{2})^N \leq 0.3/2 \Rightarrow N \geq 4$

第1次迭代：

- $a_1 = a_0 + \rho(b_0 - a_0) = 0.7639;$
 $b_1 = a_0 + (1 - \rho)(b_0 - a_0) = 1.236$
 $\Rightarrow f(a_1) = -24.36 < f(b_1) = -18.96$

- 因此，区间被压缩为 $[a_0, b_1] = [0, 1.236]$

第2次迭代：

- $a_2 = a_0 + \rho(b_1 - a_0) = 0.4721$
 $\Rightarrow f(a_2) = -21.10 > f(b_2) = f(a_1) = -24.36$

- 因此，区间被压缩为 $[a_2, b_1] = [0.4721, 1.236]$

黄金分割法举例

第3次迭代：

- $b_3 = a_2 + (1 - \rho)(b_1 - a_2) = 0.9443;$
 $\Rightarrow f(a_3) = f(b_2) = -24.36 < f(b_3) = -23.59$
- 因此，区间被压缩为 $[a_2, b_3] = [0.4721, 0.9443]$

第4次迭代：

- $a_4 = a_2 + \rho(b_3 - a_2) = 0.6525$
 $\Rightarrow f(a_4) = -23.84 > f(b_4) = f(a_3) = -24.36$
- 因此，区间被压缩为 $[a_4, b_3] = [0.6525, 0.9443]$

斐波那契数列法举例

经过 N 次压缩可以在 $[0, 2]$ 上取得极小点 $\frac{1+2\varepsilon}{F_{N+1}} \leq 0.3/2 \Rightarrow N \geq 4$

第1次迭代：

- $\rho_1 = 1 - \frac{F_4}{F_5} = \frac{3}{8}$
 $a_1 = a_0 + \rho(b_0 - a_0) = \frac{3}{4}$
 $b_1 = a_0 + (1 - \rho)(b_0 - a_0) = \frac{5}{4}$
 $\Rightarrow f(a_1) = -24.34 < f(b_1) = -18.65$
- 因此，区间被压缩为 $[a_0, b_1] = [0, \frac{5}{4}]$

第2次迭代：

- $\rho_2 = 1 - \frac{F_3}{F_4} = \frac{2}{5}$
 $a_2 = a_0 + \rho_2(b_1 - a_0) = \frac{1}{2}$
 $b_2 = a_1 = \frac{3}{4}$
 $\Rightarrow f(a_2) = -21.69 > f(b_2) = f(a_1) = -24.34$
- 因此，区间被压缩为 $[a_2, b_1] = [\frac{1}{2}, \frac{5}{4}]$

斐波那契数列法举例

第3次迭代：

- $\rho_3 = 1 - \frac{F_2}{F_3} = \frac{1}{3}$
 $a_3 = b_2 = \frac{3}{4}$
 $b_3 = a_2 + (1 - \rho_3)(b_1 - a_2) = 1$
 $\Rightarrow f(a_3) = f(b_2) = -24.34 < f(b_3) = -23$
- 因此，区间被压缩为 $[a_2, b_3] = [\frac{1}{2}, 1]$

第4次迭代：

- 令 $\varepsilon = 0.05$
 $\rho_4 = 1 - \frac{F_1}{F_2} = \frac{1}{2}$
 $a_4 = a_2 + (\rho_4 - \varepsilon)(b_3 - a_2) = 0.725$
 $b_4 = a_3 = \frac{3}{4}$

$$\Rightarrow f(a_4) = -24.27 > f(b_4) = f(a_3) = -24.34$$

- 因此，区间被压缩为 $[a_4, b_3] = [0.725, 1]$



牛顿法

- 针对一元单值函数 $f : \mathbb{R} \rightarrow \mathbb{R}$, 求 $[a_0, b_0]$ 上的极小点。
- 前提: f 在 $[a_0, b_0]$ 上是单峰的, 并且是连续二阶可微的, 一阶导数为 f' , 二阶导数为 f'' 。
- 步骤: 构造经过 $(x^{(k)}, f(x^{(k)}))$ 的函数

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

此时满足 $q(x^{(k)}) = f(x^{(k)})$, $q'(x^{(k)}) = f'(x^{(k)})$, $q''(x^{(k)}) = f''(x^{(k)})$ 。
 $q(x)$ 可认为是 $f(x)$ 的近似, 所以求 $f(x)$ 的极小值点可近似于求解 $q(x)$ 的极小值点。

- 求极小点应满足一阶必要条件, 即 $0 = q'(x) = f'(x^{(k)}) + f''(x^{(k)})(x - x^{(k)})$, 解得 $x^{(k+1)} = x^{(k)} - f'(x^{(k)})/f''(x^{(k)})$ 。(注意: 牛顿法未必收敛, 需注意初值条件)



牛顿法举例

求目标函数 $f(x)$ 的极小点，初始值为 $x^{(0)} = 0.5$ ，要求精度为 $\varepsilon = 10^{-5}$

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

求得其一阶导数和二阶导数为

$$f'(x) = x - \cos x; f''(x) = 1 + \sin x$$

$$\begin{aligned}x^{(1)} &= 0.5 - \frac{0.5 - \cos 0.5}{1 + \sin 0.5} \\&= 0.5 - \frac{-0.3775}{1.479} \\&= 0.7552\end{aligned}$$

牛顿法举例

$$x^{(2)} = x^{(1)} - \frac{x^{(1)} - \cos x^{(1)}}{1 + \sin x^{(1)}} = 0.7391$$

$$x^{(3)} = x^{(2)} - \frac{x^{(2)} - \cos x^{(2)}}{1 + \sin x^{(2)}} = 0.7390$$

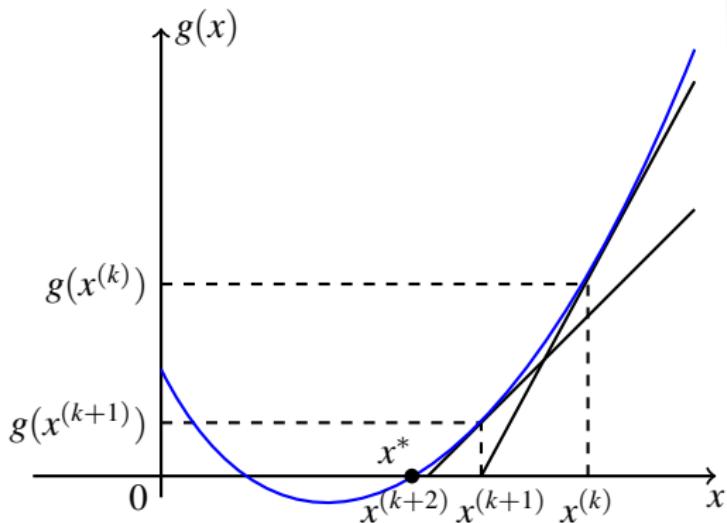
$$x^{(4)} = x^{(3)} - \frac{x^{(3)} - \cos x^{(3)}}{1 + \sin x^{(3)}} = 0.7390$$

$|x^{(3)} - x^{(4)}| < \varepsilon$ 此时, $f'(x^{(4)}) \approx 0; f''(x^{(4)}) = 1.673 > 0$
所以, $x^{(4)} = 0.7390$ 是极小点。

牛顿法举例

用牛顿法求解时常令 $g(x) = f'(x)$, 可以得到一个迭代公式, 求解方程 $g(x) = 0$

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

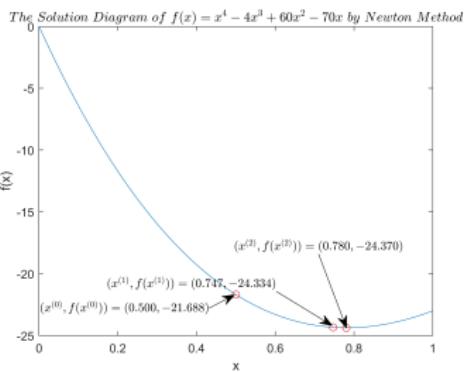
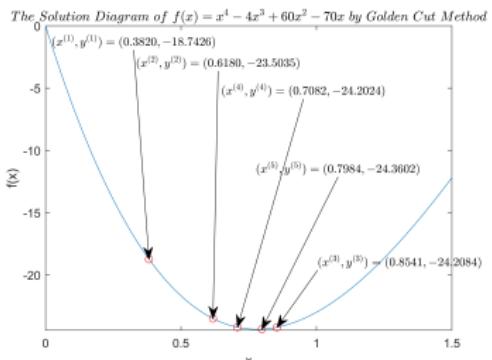
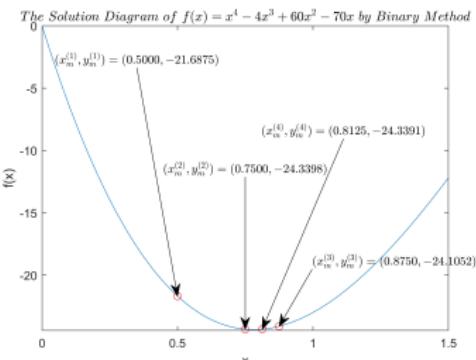


三种方法举例

问题

分别使用二分法，黄金分割法，牛顿法求解函数 $f(x) = x^4 - 14x^3 + 60x^2 - 70x$ 在[0,1]上的最小值，要求精度小于0.1，给出每一步的详细计算结果

三种方法举例



割线法

- 已知牛顿法要求 $x^{(k+1)} = x^{(k)} - f'(x^{(k)})/f''(x^{(k)})$ 。如果二阶导数不存在，则可以采用不同点处一阶导数对其近似得到。

- 例如：

$$f''(x^{(k)}) = \frac{f'(x^{(k)}) - f'(x^{(k-1)})}{x^{(k)} - x^{(k-1)}}$$

代入牛顿公式中，得到新的迭代公式：

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

- 整理后即为：

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

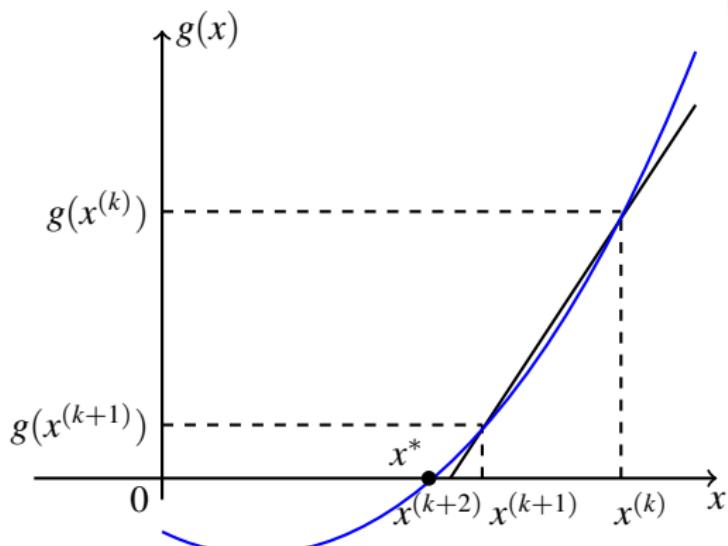
- 该方法需要两个起始点 $x^{(-1)}$ 、 $x^{(0)}$ 。

割线法

令 $g(x) = f'(x)$, 可以得到一个迭代公式, 求解方程 $g(x) = 0$

$$x^{(k+1)} = x^{(k)} - g(x^{(k)}) \frac{x^{(k)} - x^{(k-1)}}{g(x^{(k)}) - g(x^{(k-1)})}$$

或 $x^{(k+1)} = \frac{g(x^{(k)})x^{(k-1)} - g(x^{(k-1)})x^{(k)}}{g(x^{(k)}) - g(x^{(k-1)})}$



目录

无约束最优化问题

一阶必要条件

二阶必要条件

二阶充分条件

凸充分性定理

线性搜索问题

非精确线性搜索方法

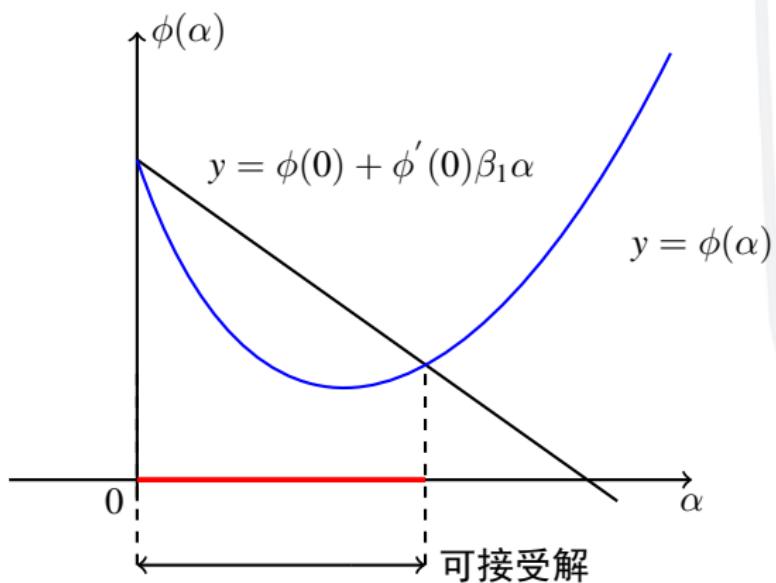
下降算法的全局收敛性与收敛速率

非精确线性搜索方法

- 用精确线性搜索方法求得的 $\phi(\alpha) = f(\mathbf{x}^{(k)} + \alpha\mathbf{d}^{(k)})$ 是精确（或近似）极小点，迭代过程中尽管每次下降比较多，但计算量相对较大。
- 许多时候在求解 $f(\mathbf{x})$ 的最优解时，无需十分精确，特别是在迭代的最初阶段。过分追求精确度会降低整个方法的运算速度。因此，我们可以适当放宽对 α_k 的要求，只需目标函数每次迭代都有充分下降即可，这样就可减少每次迭代的时间，使整体效果更好。
- 常见的方法包括Armijo方法、Goldstein方法、Wolfe-Powell方法。

Armijo方法

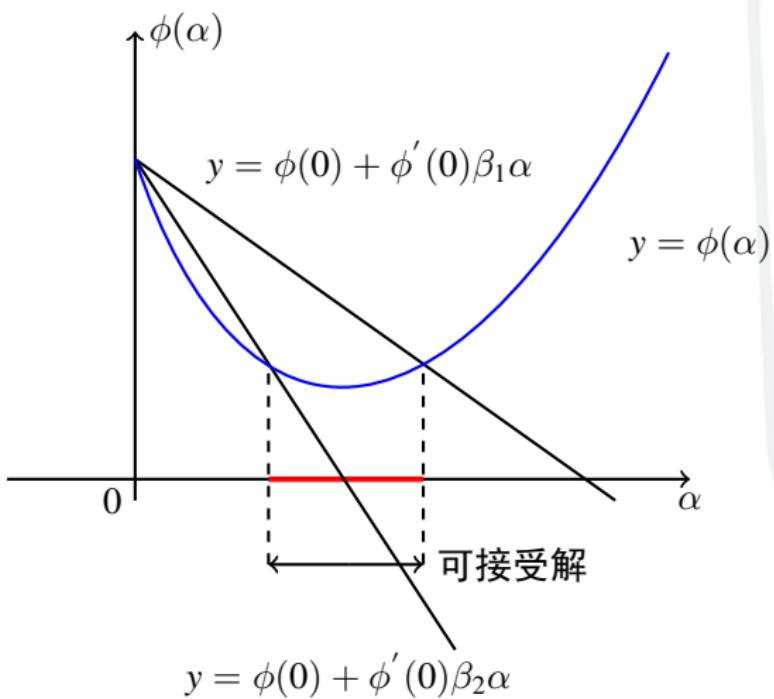
- 预先指定一个参数 β_1 , 满足 $0 < \beta_1 < 1$, 需要满足 $0 < \phi(\alpha) \leq \phi(0) + \phi'(0)\beta_1\alpha$;



Goldstein方法

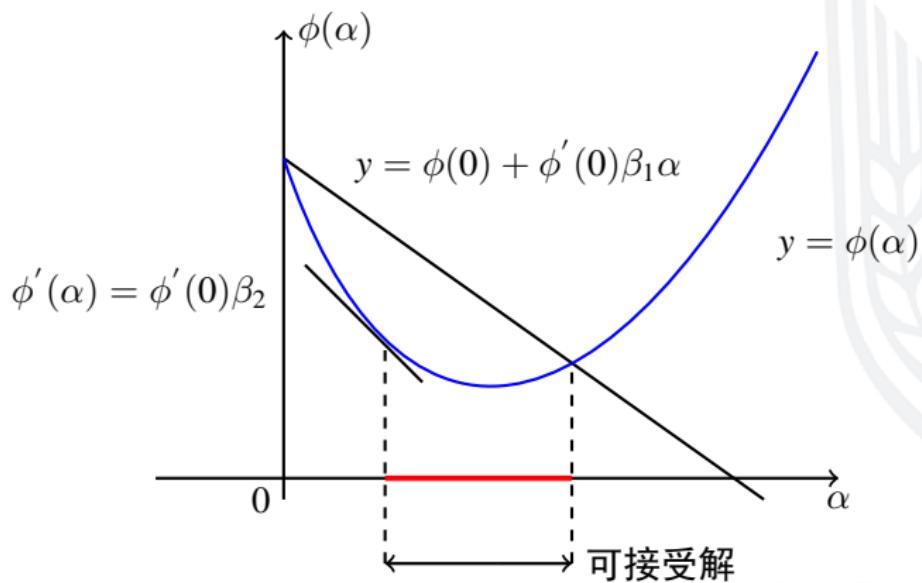
- 预先指定两个参数 β_1 和 β_2 , 满足 $0 < \beta_1 < \beta_2 < 1$, α 需要满足 $\phi(0) + \phi'(0)\beta_2\alpha \leq \phi(\alpha) \leq \phi(0) + \phi'(0)\beta_1\alpha$;
- 两个系数 β_1 和 β_2 控制 α 既不过大又不过小。

Goldstein方法



Wolfe-Powell方法

- 预先指定两个参数 β_1 和 β_2 , 满足 $0 < \beta_1 < \beta_2 < 1$, α 需要满足 $\phi(\alpha) \leq \phi(0) + \phi'(0)\beta_1\alpha$ 且 $\phi'(0)\beta_2 \leq \phi'(\alpha)$;
- 两个系数 β_1 和 β_2 控制 α 既不过大又不过小。



目录

无约束最优化问题

一阶必要条件

二阶必要条件

二阶充分条件

凸充分性定理

线性搜索问题

非精确线性搜索方法

下降算法的全局收敛性与收敛速率

下降算法的全局收敛性与收敛速率

下降算法的一般迭代格式如下：

- ① 给定初始点 $x^{(1)}$, $k = 1$;
- ② 确定局部下降方向 $d^{(k)}$, 使得 $\nabla f(x^{(k)})^T d^{(k)} < 0$
- ③ 确定步长 $\alpha_k > 0$ 使得 $f(x^{(k)} + \alpha_k d^{(k)}) < f(x^{(k)})$
- ④ 令 $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$
- ⑤ 若 $x^{(k+1)}$ 满足终止准则, 则停; 否则, $k = k + 1$, 转至步骤(1)

下降算法的全局收敛性与收敛速率

三个问题：

- ① 如何确定某点处的搜索方向？
- ② 如何进行一维搜索以确定步长？
- ③ 如何确定当前点的终止准则？

解决方法：

- 对第(1)个问题，今后讨论；
- 对第(2)个问题，本节之前已讨论；
- 终止准则一般与算法相关，常定为 $\nabla f(\mathbf{x}^{(k)}) = \mathbf{0}$

下降算法的全局收敛性与收敛速率

- 设 $\nabla f(\mathbf{x})$ 在水平集 $L(\mathbf{x}^0) = \{\mathbf{x} | f(\mathbf{x}) \leq f(\mathbf{x}^0)\}$ 上存在且一直连续，下降算法的搜索方向 $d^{(k)}$ 与 $-\nabla f(\mathbf{x}^{(k)})$ 之间的夹角为 $\theta_k = \arccos\left(\frac{-\nabla f(\mathbf{x}^{(k)})d^{(k)}}{\|\nabla f(\mathbf{x}^{(k)})\|\|d^{(k)}\|}\right)$ ，满足条件：
存在 $\bar{u} > 0$ ，使得 $\theta_k \leq \frac{\pi}{2} - \bar{u}, \forall k$
- 步长 α_k 由以下3种方法：
(1)精确线性搜索，(2)Goldstein方法，(3)Wolfe-Powell方法之一确定，则
对某个 k ，有 $\nabla f(\mathbf{x}^{(k)}) = \mathbf{0}$ ，或者有 $f(\mathbf{x}^{(k)}) \rightarrow -\infty (k \rightarrow \infty)$ ，或者
有 $\nabla f(\mathbf{x}^{(k)}) \rightarrow \mathbf{0} (k \rightarrow \infty)$

下降算法的全局收敛性与收敛速率

- 设序列 $\{x^{(k)}\}$ 收敛到 x^* , 若极限

$$\lim_{k \rightarrow \infty} \frac{\|x^{(k+1)} - x^*\|}{\|x^{(k)} - x^*\|} = \beta$$

存在, 则当 $0 < \beta < 1$ 时, 则称 $\{x^{(k)}\}$ 为线性收敛; 当 $\beta = 0$ 时, 则称 $\{x^{(k)}\}$ 为超线性收敛; 当 $\beta = 1$ 时, 则称 $\{x^{(k)}\}$ 为次线性收敛, 因为次线性收敛的收敛速度太慢, 一般不予以考虑。

- 若存在某个 $p \geq 1$, 有

$$\lim_{k \rightarrow \infty} \frac{\|x^{(k+1)} - x^*\|}{\|x^{(k)} - x^*\|^p} = \beta < +\infty$$

则称 $\{x^{(k)}\}$ 为 p 阶收敛。当 $p > 1$ 时, p 阶收敛必为超线性收敛, 反之不一定成立。

Thank you for your
attention!

翟卫欣 副教授

zhaiweixin@cau.edu.cn

中国农业大学信息与电气工程学院