优化理论与算法

第七章 迭代算法

郭加熠|助理教授

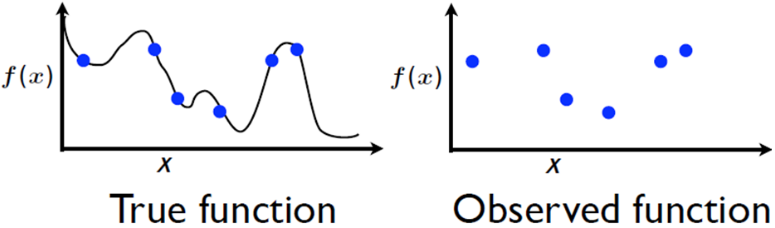
目录

讲　　员

郭加熠，江波，刘慧康

如何寻找函数的最小值

- **精确方法**: 适用于简单问题
- **数值方法**: 适用于大多数大规模复杂工程问题。通过计算机实现优化算法



我们通常对优化函数知之甚少，只能计算少量点处的函数值和梯度，且每次计算成本较高。

迭代方法

在计算数学中，迭代方法是通过初始值生成一系列**逐步改进的近似解**的数学过程，其中第 $(k+1)$ 次近似解由第 k 次解导出。

Algorithm 1 Iterative method

initial point x^0
for $k = 0, 1, 2, \ldots$ **do**
  update $x^{k+1} = \Phi(x^k)$
  if termination criteria is met **then**
    stop
  end if
end for

三个关键要素:

1. 初始点 x^0
2. 更新策略 $\Phi(\cdot)$
3. 终止准则

示例：不动点迭代

求解二次方程：

$$x^2 - x - 1 = 0$$

等价于寻找满足 $x = 1 + \frac{1}{x}$ 的解，定义迭代格式：

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

当 $x^{k+1} \approx x^k$ 时停止。取 $x^0 = 2$，迭代过程如下：

$$\begin{aligned} x^1 &= 1 + \frac{1}{2} = 1.5 & x^2 &= 1 + \frac{1}{1.5} = 1.666 \\ x^3 &= 1 + \frac{1}{1.666} = 1.6 & x^4 &= 1 + \frac{1}{1.6} = 1.625 \\ x^5 &= 1 + \frac{1}{1.625} = 1.615 & x^6 &= 1 + \frac{1}{1.615} = 1.619 \end{aligned}$$

序列 $\{x^k\}$ 收敛至解 $x^\star = 1.618$。

示例：不动点迭代

求解方程实根：

$$\cos x - 3x + 1 = 0$$

等价于寻找满足 $x = \frac{1}{3}(\cos x + 1)$ 的解，定义迭代格式：

$$x^{k+1} = \Phi(x^k) = \frac{1}{3}(\cos x^k + 1)$$

当 $x^{k+1} \approx x^k$ 时停止。取 $x^0 = 0$，迭代过程如下：

$$\begin{aligned}
x^1 &= \frac{1}{3}(\cos 0 + 1) = 0.6667 & x^2 &= \frac{1}{3}(\cos 0.6667 + 1) = 0.5953 \\
x^3 &= \frac{1}{3}(\cos 0.5953 + 1) = 0.6093 & x^4 &= \frac{1}{3}(\cos 0.6093 + 1) = 0.6067 \\
x^5 &= \frac{1}{3}(\cos 0.6067 + 1) = 0.6072 & x^6 &= \frac{1}{3}(\cos 0.6072 + 1) = 0.6071
\end{aligned}$$

求得解 $x^\star \approx 0.607$。

示例：单纯形法

考虑标准线性规划问题：

$$
\begin{aligned}
\min \quad & c^T x \\
\text{s.t.} \quad & Ax = b,\ x \geq 0
\end{aligned}
$$

单纯形法每次迭代的单纯形表：

$0 \quad c_N^T - c_B^T A_B^{-1} A_N$		$-c_B^T A_B^{-1} b$
I	$A_B^{-1} A_N$	$A_B^{-1} b$

- x^0：对应初始基的基本可行解
- $\Phi(\cdot)$：转轴运算（从基本可行解 x^k 移动到相邻基本可行解 x^{k+1}）
- 终止条件：$c_N^T - c_B^T A_B^{-1} N \geq 0$

示例：幂方法

求对称矩阵 $A \in \mathbb{R}^{n\times n}$ 的最大特征值，转化为优化问题：

$$\max_{x\in\mathbb{R}^n} \; x^T A x \quad \text{s.t.} \; \|x\|_2 = 1$$

幂方法（Power Method）迭代格式：

$$x^{k+1} = \frac{Ax^k}{\|Ax^k\|_2}$$

- x^0：随机生成
- $\Phi(x) = \frac{Ax}{\|Ax\|_2}$
- 终止条件：$x^{k+1} \approx x^k$

示例：黄金分割法（零阶方法）

最小化一维凸函数 $f(x), x \in [a, b]$

0) 初始化：设 $x_l = a$， $x_r = b$

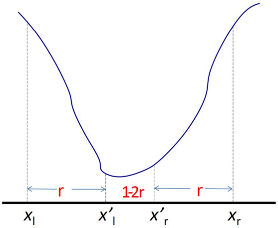
1) 选取两点：

$$x_l' = x_l + r(x_r - x_l)$$
$$x_r' = x_l + (1 - r)(x_r - x_l)$$
$$r = 0.382$$

2) 比较 $f(x_l')$ 和 $f(x_r')$：

- 若 $f(x_l') < f(x_r')$，更新 $x_r = x_r'$
- 若 $f(x_l') \geq f(x_r')$，更新 $x_l = x_l'$

重复步骤 1) 直至 x_l' 与 x_r' 足够接近



为何选择 $r = 0.382$?

示例：下降方法

考虑无约束优化问题:

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

下降方法迭代格式:

$$x^{k+1} = x^k + t_k p_k, \quad k = 0, 1, 2, \ldots$$

- $t_k > 0$：自定义步长
- p_k：基于 $f(x)$ 计算的下降方向
- 通过合理选择 t_k 和 p_k 保证

$$f(x^{k+1}) < f(x^k)$$

目录

讲　　　　员

郭加熠，江波，刘慧康

终止准则

一阶条件 $\|\nabla f(x)\| = 0$ 不实用

实用终止准则:

- 梯度足够小: $\|\nabla f(x^k)\| \le \epsilon$
- 目标函数变化量: $|f(x^{k+1}) - f(x^k)| \le \epsilon$ 或

$$\frac{|f(x^{k+1}) - f(x^k)|}{|f(x^k)|} \le \epsilon$$

- 迭代点变化量: $\|x^{k+1} - x^k\| \le \epsilon$ 或

$$\frac{\|x^{k+1} - x^k\|}{\|x^k\|} \le \epsilon$$

- 为避免除以极小值，分母可替换为 $\max\{1, |f(x^k)|\}$ 和 $\max\{1, \|x^k\|\}$

误差度量与收敛性

- 问题误差度量：
  - $e_k = \left\| x^k - x^* \right\|$
  - $e_k = f(x^k) - f(x^*)$
  - $e_k = \left\| \nabla f(x^k) \right\|$
- 收敛性

$$\lim_{k\to\infty} x^k = x^* \text{ 或 } \lim_{k\to\infty} f(x^k) = f(x^*)$$

- 迭代复杂度（全局）：

$$e_k \leq O\left(\frac{1}{k}\right) \left(O\left(\frac{1}{k^2}\right), O\left(\frac{1}{\sqrt{k}}\right), O\left(\frac{1}{2^k}\right), O\left(\frac{1}{2^{2^k}}\right)\right)$$

收敛速率（局部）

设序列 $\{x^k\}$ 收敛于 x^*。

Q-收敛定义：$\lim_{k\to\infty} \frac{e_{k+1}}{e_k} = \mu$

- 若 $\mu = 1$，称为**次线性收敛**
- 若 $\mu \in (0,1)$，称为**线性收敛**
- 若 $\mu = 0$，称为**超线性收敛**
- 为区分超线性收敛速率，检验

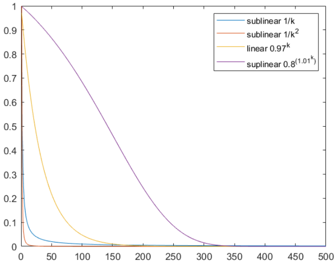
$$\lim_{k\to\infty} \frac{e_{k+1}}{(e_k)^q} = \mu > 0$$

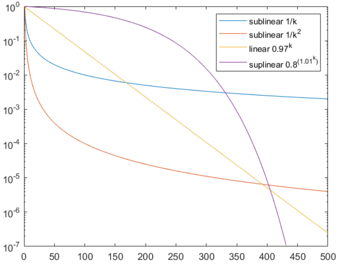
  当 $q = 2$ 时，称为**二次收敛**

- 若对任意 k 有 $|e_k| \le \epsilon_k$ 且 $\{\epsilon_k\}$ Q-收敛于 0，则称 $\{x^k\}$ 为 **R-收敛**

收敛速率示例

- 次线性收敛：$\frac{1}{k}$
- 次线性收敛：$\frac{1}{k^2}$
- 线性收敛：0.97^k
- 超线性收敛：$0.8^{1.01^k}$





- R-收敛：$0.5^{\lceil k/2 \rceil}$
- 二次收敛：2^{2^k}

次线性 vs 线性收敛

- 次线性收敛速度**较慢**

  例如误差按 $O(1/k)$ 或 $O(1/k^2)$ 衰减（k 为迭代次数）。要达到精度 ϵ，分别需要 $O(1/\epsilon)$ 和 $O(\sqrt{1/\epsilon})$ 次迭代

- 线性收敛渐近**更快**，达到精度 ϵ 需要 $O(\log(1/\epsilon))$ 次迭代

- 二次收敛在实践中通常只需**常数**次迭代