基于 Wolfe-Powell 准则的非精确一维步长搜索算法

Xiaoma

2023 年 1 月 6 日

问题描述

给定目标函数 $f(x)$:

- 二维 Rosenbrock 函数：

$$f(x) = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2$$

求解无约束优化问题：

$$\min_x f(x)$$

实验原理

拟牛顿法

牛顿法

牛顿法的基本思想是利用目标函数的二次 Taylor 展开，并将其极小化。

设 $f(x)$ 是可微实函数，$x^{(k)} \in \mathbb{R}^n$，Hesse 矩阵 $\nabla^2 f(x^{(k)})$ 正定，在 $x^{(k)}$ 附近用二次 Taylor 展开近似 f

$$q^{(k)}(s) = f(x^{(k)}) + \nabla f(x^{(k)})^T s + \frac{1}{2} s^T \nabla^2 f(x^{(k)}) s$$

其中 $s=x-x^{(k)}$，$q^{(k)}(s)$ 为 $f(x)$ 的二次近似，将上式右边极小化得

$$x^{(k+1)}=x^{(k)}-[\nabla^2 f(x^{(k)})]^{-1}\nabla f(x^{(k)})$$

在该公式中，步长因子 $a_k=1$，令 $G_k=\nabla^2 f(x^{(k)}), g^{(k)}=\nabla f(x^k)$，则原式可写为

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

显然，牛顿法也可看成在椭球范数 $\|\cdot\|_{G_k}$ 下的最速下降法，对于 $f(x^k+s)\approx f(x^{(k)})+g^{(k)\mathcal{T}}s$，$s^{(k)}$ 是极小化问题

$$\min \frac{g^{(k)\mathcal{T}}}{\|s\|}$$

的解，当采用 l_2 范数时

$$s^{(k)}=-g^{(k)}$$

所得方法是最速下降法，当采用椭球范数 $\|\cdot\|_{G_k}$ 时

$$s^{(k)}=-G_k^{-1}g^{(k)}$$

所得方法是牛顿法。

经典牛顿迭代法的运算步骤为

Algorithm 1: 牛顿法

Input: 初始点 x_0，阈值误差 ε

Output: $x^{(k)}, f(x^{(k)})$

(1) 初始化 k=0

(2) 计算 $g^{(k)}=\nabla f(x^{(k)})$，如果 $\|g^{(k)}\|<\varepsilon$，则停止迭代

(3) 解线性方程组 $s^{(k)}=-G_k^{-1}g^{(k)}$

(4) 更新 $x^{(k+1)}=x^{(k)}+s^{(k)}, k=k+1$，返回步骤 (2)

拟牛顿法

牛顿法虽然收敛速度较快，但计算 Hense 矩阵的成本过大，并且若矩阵不是正定的，则牛顿法失效。

对 $\nabla f(x)$ 在 $x^{(k)}$ 出 Taylor 展开得到如下近似

$$\nabla f(x) = g^{(k)} + H_k(x - x^{(k)})$$

令 $x = x^{(k+1)}$ 即可得到 $g^{(k+1)} - g^{(k)} = H_k(x^{(k+1)} - x^{(k)})$
记 $y^{(k)} = g^{(k+1)} - g^{(k)}, \delta_k = x^{(k+1)} - x^{(k)}$

$$y^{(k)} = H_k\delta_k$$

称为拟牛顿条件

DFP 法

设对称秩二矫正为

$$H_{k+1} = H_k + auu^T + bvv^T$$

令拟牛顿条件满足，则

$$H_k y^{(k)} + auu^T y^{(k)} + bvv^T y^{(k)} = s^{(k)}$$

这里 u 和 v 并不唯一确定，但 u 和 v 的明显选择是

$$u = s^{(k)} \quad v = H_k y^{(k)}$$

确定出

$$a = 1/y^{(k)}u^T = 1/s^{(k)T}y^{(k)}b = -1v^T/y^{(k)} = -1/y^{(k)T}H_k y^{(k)}$$

因此

$$H_{k+1} = H_k + \frac{s^{(k)}s^{(k)T}}{s^{(k)T}y^{(k)}} - \frac{H_k y^{(k)} y^{(k)T} H_k}{y^{(k)T} H_k y^{(k)}}$$

这个公式称为 DFP 公式。DFP 法的运算步骤为

BFGS

类似的，可以得到关于 B_k 的对称秩二矫正公式

$$B_{k+1}^{(BFGS)} = B_k + \frac{\mathbf{y}^{(k)}\mathbf{y}^{(k)^T}}{\mathbf{y}^{(k)^T}\mathbf{s}^{(k)}} - \frac{B_k\mathbf{s}^{(k)}\mathbf{s}^{(k)^T}B_k}{\mathbf{s}^{(k)^T}B_k\mathbf{s}^{(k)}}$$

Algorithm 2: DFP

Input: 初始点 x_0，阈值误差 ε

Output: $x^{(k)}, f(x^{(k)})$

(1) 初始化 $H_0 = I, k = 0$

(2) 计算搜索方向：$d^{(k)} = -H_k \nabla f(x^{(k)})$，如果 $\|g^{(k)}\| < \varepsilon$，停止迭代

(3) 一维搜索确定步长 α_k，令 $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$

(4) 令 $\mathbf{s}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}, \mathbf{y}^{(k)} = \nabla f\left(\mathbf{x}^{(k+1)}\right) - \nabla f\left(\mathbf{x}^{(k)}\right)$, 当 $\mathbf{s}^{(k)^T}\mathbf{y}^{(k)} > 0$, 作更新 $H_{k+1} = H_k + \frac{\mathbf{s}^{(k)}\mathbf{s}^{(k)^T}}{\mathbf{s}^{(k)^T}\mathbf{y}^{(k)}} - \frac{H_k\mathbf{y}^{(k)}\mathbf{y}^{(k)^T}H_k}{\mathbf{y}^{(k)^T}H_k\mathbf{y}^{(k)}}$ 。置 $k = k + 1$, 返回步骤 (2)

H_k 的 BFGS 校正公式为

$$
\begin{aligned}
H_{k+1}^{(BFGS)} =& H_k + \left(1 + \frac{\mathbf{y}^{(k)^T}H_k\mathbf{y}^{(k)}}{\mathbf{s}^{(k)^T}\mathbf{y}^{(k)}}\right)\frac{\mathbf{s}^{(k)}\mathbf{s}^{(k)^T}}{\mathbf{s}^{(k)^T}\mathbf{y}^{(k)}} \\
&- \frac{H_k\mathbf{y}^{(k)}\mathbf{s}^{(k)^T} + \mathbf{s}^{(k)}\mathbf{y}^{(k)^T}H_k}{\mathbf{s}^{(k)^T}\mathbf{y}^{(k)}}.
\end{aligned}
$$

BFGS 法的运算步骤为

Algorithm 3: BFGS

Input: 初始点 x_0，阈值误差 ε

Output: $x^{(k)}, f(x^{(k)})$

(1) 初始化 $H_0 = I, k = 0$

(2) 计算搜索方向：$d^{(k)} = -H_k \nabla f(x^{(k)})$，如果 $\|g^{(k)}\| < \varepsilon$，停止迭代

(3) 一维搜索确定步长 α_k，令 $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$

(4) 令 $\mathbf{s}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}, \mathbf{y}^{(k)} = \nabla f\left(\mathbf{x}^{(k+1)}\right) - \nabla f\left(\mathbf{x}^{(k)}\right)$, 作更新 $H_{k+1}^{(BFGS)} = H_k + \left(1 + \frac{\mathbf{y}^{(k)^T}H_k\mathbf{y}^{(k)}}{\mathbf{s}^{(k)^T}\mathbf{y}^{(k)}}\right)\frac{\mathbf{s}^{(k)}\mathbf{s}^{(k)^T}}{\mathbf{s}^{(k)^T}\mathbf{y}^{(k)}} - \frac{H_k\mathbf{y}^{(k)}\mathbf{s}^{(k)^T}+\mathbf{s}^{(k)}\mathbf{y}^{(k)^T}H_k}{\mathbf{s}^{(k)^T}\mathbf{y}^{(k)}}$ 。置 $k = k + 1$, 返回步骤 (2)

非精确一维步长搜索

非精确一维搜索

找出满足某些适当条件的粗略近似解作为步长，提升算法的整体计算效率

Wolfe-Powell conditions :

$$\varphi(\alpha) \le \varphi(0) + \rho\alpha\varphi'(0)$$
$$\varphi'(\alpha) \ge \sigma\varphi'(0)$$

其中 $\rho \in (0, \frac{1}{2}), \sigma \in (\rho, 1)$ 是固定参数。

设 $\hat{\alpha}_k$ 是使得 $f(x^{(k)} + \alpha d^{(k)}) = f(x^{(k)})$ 的最小正数 α

基于 Wolfe-Powell 准则的非精确一维步长搜索

(1) 给定初始一维搜索区间 $[0, \bar{\alpha}]$, 以及 $\rho \in \left(0, \frac{1}{2}\right), \sigma \in (\rho, 1)$. 计算 $\varphi_0 = \varphi(0) = f\left(x^{(k)}\right), \varphi_0' = \varphi'(0) = \nabla f\left(x^{(k)}\right)^T d^{(k)}$. 并令 $a_1 = 0, a_2 = \bar{\alpha}, \varphi_1 = \varphi_0, \varphi_1' = \varphi_0'$. 选取适当的 $\alpha \in (a_1, a_2)$.

(2) 计算 $\varphi = \varphi(\alpha) = f\left(x^{(k)} + \alpha d^{(k)}\right)$. 若 $\varphi(\alpha) \le \varphi(0) + \rho\alpha\varphi'(0)$, 则转到第 (3) 步。否则，由 $\varphi_1, \varphi_1', \varphi$ 构造二次插值多项式 $p^{(1)}(t)$, 并得其极小点 $\hat{\alpha}$. 令 $a_2 = \alpha, \alpha = \hat{\alpha}$, 重复第 (2) 步.

(3) 计算 $\varphi' = \varphi'(\alpha) = \nabla f\left(x^k + \alpha d^{(k)}\right)^T d^{(k)}$. 若 $\varphi'(\alpha) \ge \sigma\varphi'(0)$, 则输出 $\alpha_k = \alpha$, 并停止搜索. 否则由 $\varphi, \varphi', \varphi_1'$ 构造两点二次插值多项式 $p^{(2)}(t)$，并求得极小点 $\hat{\alpha}$. 令 $a_1 = \alpha, \alpha = \hat{\alpha}$, 返回第 (2) 步.

数据说明

二维 Rosenbrock 函数

程序输入输出说明

输入 1、2 选择 DFP/BFGS，然后输入初始点。

输出值为迭代次数与最优解

程序测试结果

DFP	epoch	x1, x2	f
0 0	17	0.9999999986994 0.9999999981660	6.055373455850762e-17
0.5 0.5	16	1.00000001203758 1.0000000247276	1.8747613102236447e-16
-1 1	1	1.0 1.0	0.0
2 -3	37	0.99999999503545 0.99999999346	1.176336523561564e-15
-100 150		无法收敛	

BFGS	epoch	x1, x2	f
0 0	18	0.99999999786 0.99999999574	4.635405677357143e-18
0.5 0.5	15	0.9999999995 0.9999999990	2.4722121607796934e-19
-1 1	1	1.0 1.0	0.0
2 -3	26	1.00000002922187 1.000000058506	8.543076984615268e-16
-100 150	271	0.99999999993 0.9999999999	4.539167488746724e-19

分析总结

不同的初始点也可能产生截然不同的结果，甚至无法收敛。

经过测试发现，BFGS 法具有比 DFP 法更稳定的数值。