

有限内存 BFGS 法的对角预条件子

曹蕙语 周群艳 薛瑾

江苏理工学院数理学院 江苏常州 213001

摘要: 基于修正弱拟牛顿方程, 构造了一种有限内存 BFGS 法对角预条件子, 实验表明算法有效.

关键词: 预条件; 对角校正; 修正弱拟牛顿方程

1 引言

问题 $\min_{x \in R^n} f(x)$ 的线搜索法迭代格式是

$$x_{k+1} = x_k - \alpha_k H_k g_k, k = 0, 1, 2, \dots, (1)$$

α_k 是步长, g_k 是 x_k 处的梯度, H_k 是 Hesse 阵逆的近似, 满足拟牛顿方程

$$H_k y_{k-1} = s_{k-1}, (2)$$

$s_{k-1} = x_k - x_{k-1}$, $y_{k-1} = g_k - g_{k-1}$. H_k 用 BFGS 等拟牛顿公式校正得到. 但 n 很大时, 有限内存 BFGS(L-BFGS) 法更受青睐.

$$\text{BFGS 法中, } H_{k+1} = V_k^T H_k V_k + \rho_k s_k s_k^T, \rho_k = \frac{1}{y_k^T s_k}$$

, $V_k = I - \rho_k y_k s_k^T$. L-BFGS 法由 BFGS 法改进得到. 若已

有 $(s_i, y_i) (i = k-m, \dots, k-1)$, 取 H_{k+1} 的初始近似 $H_{k+1}^{(0)}$,

用 BFGS 公式校正 m 次, 可得

$$H_{k+1} = (V_k^T \dots V_{k-m}^T) H_{k+1}^{(0)} (V_{k-m}^T \dots V_k^T) + \rho_{k-m} (V_k^T \dots V_{k-m+1}^T) s_{k-m} s_{k-m}^T (V_{k-m+1}^T \dots V_k^T) + \dots + \rho_{k-1} s_{k-1} s_{k-1}^T,$$

$\hat{m} = \min(m, k)$. L-BFGS 只需存储第 k 步的向量 $H_k g_k$, 用文 [1] 提出的双循环算法得到.

尽管 L-BFGS 法及其改进很有效 [2], 但遇坏条件问题可能极慢, 可对 $H_{k+1}^{(0)}$ 使用预条件子提高效率, 如取文 [1] 中的

$$H_{k+1}^{(0)} = \frac{s_k^T y_k}{y_k^T y_k} I. (3)$$

(3) 是数量矩阵, 对有些问题仍可能很慢. 文 [3] 提出一

种对角校正, D_k 为正定阵, D_{k+1} 由 D_k 校正得到, 满足弱拟牛顿方程, 且和 D_k 的偏离最小. 基于此, 文 [4] 令

$H_{k+1}^{(0)} = \gamma I + \Delta$, Δ 为对角阵, $\gamma \geq 0$, 满足经典弱拟牛顿方程

$$y_k^T H_{k+1}^{(0)} y_k = y_k^T s_k.$$

通过求解

$$\min \frac{1}{2} \|\Delta\|_F^2$$

$$s.t. y_k^T (\gamma I + \Delta) y_k = y_k^T s_k,$$

得

$$H_{k+1}^{(0)} = \gamma I + \frac{y_k^T s_k - \gamma y_k^T y_k}{tr(Y_k^2)} Y_k.$$

$Y_k = \text{diag}(y_{k,1}^2, y_{k,2}^2, \dots, y_{k,n}^2)$, γ 自由选取.

2 算法

文 [5] 提出修正拟牛顿方程

$$H_{k+1} y_k = \alpha_k s_k,$$

$$\alpha_k = \frac{s_k^T y_k}{2(f_k - f_{k+1} + g_{k+1}^T s_k)}. \text{ 为保证收敛性, 将 } \alpha_k \text{ 截}$$

取在 $[0.01, 100]$ 内. 本文要求 $H_{k+1}^{(0)} = \theta I + \Delta$ ($\theta \geq 0$) 满足

修正弱拟牛顿方程

$$y_k^T H_{k+1} y_k = \alpha_k y_k^T s_k.$$

定理 2.1 假设 $y_k \neq 0$, 则

$$\min \frac{1}{2} \|\Delta\|_F^2$$

$$s.t. y_k^T (\theta I + \Delta) y_k = \alpha_k y_k^T s_k, (4)$$

的最优解为

$$\Delta = \frac{\alpha_k y_k^T s_k - \theta y_k^T y_k}{tr(Y_k^2)} Y_k. (5)$$

证 目标函数和可行集都是凸的, 故有唯一解, Lagrange 函数为

$$L = \frac{1}{2} tr(\Delta^T \Delta) + \mu (y_k^T \Delta y_k + \theta y_k^T y_k - \alpha_k y_k^T s_k),$$

μ 为乘子. 将 L 对 Δ 的每个分量求偏导数并令为 0 得

$$\Delta = -\mu Y_k. (6)$$

将 (6) 左乘和右乘 y_k , 结合约束 $y_k^T \Delta y_k = \alpha_k y_k^T s_k - \theta y_k^T y_k$ 得

$$\mu = \frac{\theta y_k^T y_k - \alpha_k y_k^T s_k}{y_k^T Y_k y_k}. (7)$$

由 (6), (7) 及 $y_k^T Y_k y_k = tr(Y_k^2)$ 得 (5).

(1) 令 $\theta = \min(1, \alpha_k) \frac{s_k^T y_k}{y_k^T y_k}$, 则

$$H_{k+1}^{(0)} = \begin{cases} \alpha_k \frac{s_k^T y_k}{y_k^T y_k} I, \alpha_k < 1, \\ \frac{s_k^T y_k}{y_k^T y_k} I + \frac{(\alpha_k - 1) s_k^T y_k}{tr(Y_k^2)} Y_k, \alpha_k \geq 1. \end{cases} (8)$$

(2) 令 $\theta = \frac{\alpha_k s_k^T y_k}{y_k^T y_k}$, 则

$$H_{k+1}^{(0)} = \frac{\alpha_k s_k^T y_k}{y_k^T y_k} I. (9)$$

算法 2.1:

步 0. $x_0 \in R^n$, $0 < \beta_1 < \frac{1}{2} < \beta_2 < 1$, $\varepsilon > 0$, $H_0 = I$, $k = 0$.

步 1. 若 $\|g_k\| \leq \varepsilon$, 则停算.

步 2. 计算 $d_k = -H_k g_k$ 及 $x_{k+1} = x_k + \lambda_k d_k$, λ_k 满足

$$f(x_k + \lambda_k d_k) \leq f(x_k) + \beta_1 \lambda_k g_k^T d_k$$

, $g(x_k + \lambda_k d_k)^T d_k \geq \beta_2 g_k^T d_k$.

步 3. 计算 α_k .

步 4. 计算 $H_{k+1}^{(0)}$. 置 $k = k + 1$, 转步 1.

3 收敛性

假设 3.1 (1) $f(x)$ 二次连续可微, 在 $L(x_0) = \{x \in R^n | f(x) \leq f(x_0)\}$ 为凸的.

(2) 存在常数 M_1, M_2 使对 $\forall d \in R^n$ 及 $\forall x \in L(x_0)$ 有

$$M_1 \|d\|^2 \leq d^T H(x) d \leq M_2 \|d\|^2. (10)$$

引理 3.1 在假设 3.1 下, $\theta \in (0, c]$, $c = \frac{\alpha_k s_k^T y_k}{y_k^T y_k}$, 对

$\forall k > 0$, $\{H_{k+1}^{(0)}\}$ 有界.

证 令 $\bar{H}_k = \int_0^1 H(x_k + \tau s_k) d\tau$, 则

$$y_k = \bar{H}_k s_k. (11)$$

(10) 及 (11) 表明

$$\frac{s_k^T y_k}{\|y_k\|^2} = \frac{s_k^T \bar{H}_k s_k}{s_k^T \bar{H}_k^2 s_k} \leq \frac{M_2}{M_1^2}, \quad s_k^T y_k \leq \frac{M_2}{M_1^2} \|y_k\|^2.$$

$H_{k+1}^{(0)}$ 对角元满足

$$\theta \leq H_{k+1,i}^{(0)} \leq \theta + \frac{\alpha_k s_k^T y_k - \theta y_k^T y_k}{tr(Y_k^2)} y_{k,M}^2 \leq \theta + \frac{\alpha_k s_k^T y_k}{tr(Y_k^2)} y_{k,M}^2$$

$$\leq \theta + \frac{\alpha_{\max} M_2 \|y_k\|^2}{M_1^2 \sum_{i=1}^n (y_{k,i}^4) y_{k,M}^2} y_{k,M}^4 \leq \theta + \frac{n \alpha_{\max} M_2 y_{k,M}^2}{M_1^2 y_{k,M}^2} = \theta + \frac{n \alpha_{\max} M_2}{M_1^2}, \forall i,$$

$y_{k,M}^2$ 为各 $y_{k,i}^2$ 中最大者, 得证.

定理 3.2 若 $\{H_{k+1}^{(0)}\}$ 有界且正定, 则算法 2.1 产生的 $\{x_k\}$ 收敛到唯一极小值点 x^* , 且存在常数 $r(0 \leq r \leq 1)$, 使得

$$f(x_k) - f(x^*) \leq r^k [f(x_0) - f(x^*)],$$

即 $\{x_k\}$ r -线性收敛到 x^* .

证 见文 [1].

4 实验

选取 48 个无约束问题测试算法 2.1. 分别运行算法: L-BFGS(带预条件子 (3)), L-BFGS-M1(带预条件子 (8)), L-BFGS-M2(带预条件子 (9)).

用剖面分析图比较 (图 1), L-BFGS-M1 和 L-BFGS-M2 可解出所有问题, L-BFGS 仅解出 95% 的问题, 且需更多函数估值和迭代次数, 新对角预条件子比文 [1] 中的效果更好.

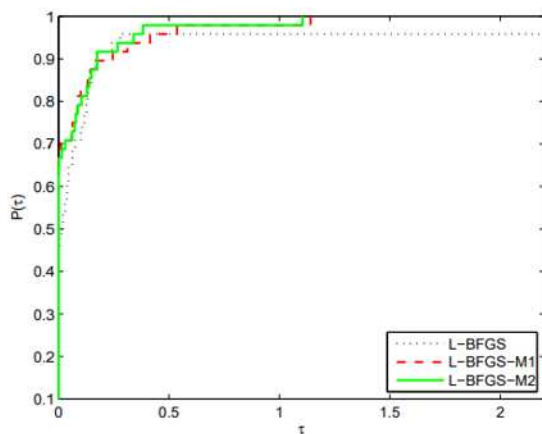


图 1 函数 / 梯度估值次数

参考文献:

- [1]Liu D C, Nocedal J. On the limited memory BFGS for large scale optimization, Math. Program. [J], 1989, 45: 503–528.
- [2] 杨博等. 一类求解非凸无约束优化问题的改进 L-BFGS 方法 [J]. 火箭军工程大学学报, 2024, 6(38):67–72.
- [3]Dennis J E, Wolkowicz H. Sizing and least-change secant methods, SIAM J. Numer. Anal.[J], 1993, 30: 1291–1314.
- [4]Leong W J, Chen C Y. A class of diagonal preconditioners for limited memory BFGS method, Optim. Method Softw.[J], 2013, 28: 379–392.
- [5]Yuan Y. A modified BFGS algorithm for unconstrained optimization, IMA J. Numer. Anal.[J], 1991, 11: 325–332.

作者简介:

曹蕙语 (2003.10–), 女, 汉, 江苏盐城人, 大学本科, 研究方向: 统计学

周群艳 (1980.04–), 女, 汉, 江苏常州人, 研究生, 副教授, 研究方向: 最优化算法

薛瑾 (2003.11), 女, 汉, 江苏盐城人, 大学本科, 研究方向: 数学与应用数学

基金项目:

2024 江苏理工学院大学生创新创业训练计划项目: “有限内存 BFGS 算法的对角预条件子研究” (编号: KYX24054)》