

基于 BFGS 修正的 HSDY 混合共轭梯度

祝子长, 刘丽平

(重庆师范大学数学科学学院, 重庆 401331)

摘要: 本文研究了大规模无约束优化问题, 利用 BFGS 逼近搜索方向, 提出了两种关于 HSDY 方法的自适应共轭梯度算法 (HSDY1 和 HSDY2). 新算法具有充分下降性和全局收敛性. 数值实验表明, 新方法比 HSDY 的计算性能更优.

关键词: 共轭梯度法; BFGS 算法; 充分下降性; 全局收敛性

MR(2010) 主题分类号: 90C30; 65K05 中图分类号: O224

文献标识码: A 文章编号: 0255-7797(2022)03-0246-13

1 引言

本文考虑无约束优化问题:

$$\min_{x \in R^n} f(x), \quad (1.1)$$

其中 $f: R^n \rightarrow R$ 是连续可微函数. 在众多的算法中, 共轭梯度法由于所需储存量小且收敛速度快, 因此被广泛用来求解大规模无约束优化问题.

共轭梯度法的基本迭代格式如下:

$$x_{k+1} = x_k + \alpha_k d_k, \quad (1.2)$$

其中 α_k 是由 Wolfe 线搜索得到的第 k 步步长; d_k 是第 k 步的搜索方向, 其第 $k+1$ 步迭代格式如下:

$$d_{k+1} = \begin{cases} -g_{k+1}, & k = 0, \\ -g_{k+1} + \beta_k d_k, & k \geq 1, \end{cases} \quad (1.3)$$

其中 $g_{k+1} = \nabla f(x_{k+1})$, β_k 是共轭参数.

共轭梯度法常见的 Wolfe 线搜索有: 标准 Wolfe 线搜索、强 Wolfe 线搜索、推广的 Wolfe 线搜索以及近似 Wolfe 线搜索. 本文采用强 Wolfe 线搜索以及近似 Wolfe 线搜索:

(1) 强 Wolfe 线搜索的步长 α_k 满足

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k, \quad (1.4)$$

$$|g(x_k + \alpha_k d_k)^T d_k| \leq -\sigma g_k^T d_k, \quad (1.5)$$

其中 $0 < \delta < \sigma < 1$.

*收稿日期: 2021-08-13 接收日期: 2021-12-14

基金项目: 重庆市教委科学技术研究项目基金资助 (KJQN 201800520).

作者简介: 祝子长 (1995-), 男, 四川旺苍, 硕士, 主要研究方向: 最优化计算方法及理论.

(2) 近似 Wolfe 线搜索 [1] 的步长 α_k 满足

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k, \quad (1.6)$$

$$g(x_k + \alpha_k d_k)^T d_k \leq (2\delta - 1)g_k^T d_k, \quad (1.7)$$

其中 $0 < \delta < \sigma < 1$.

共轭参数 β_k 经典选取有 Hestenes-Stiefel(HS) [2], Fletcher-Reeves(FR) [3], Polak-Ribière-Rolyak(PRP) [4, 5], Conjugate Descent(CD) [6], Liu-Storey(LS) [7] 和 Dai-Yuan(DY) [8] 六种, 其具体表达式如下:

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k}; \quad \beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}; \quad \beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2};$$

$$\beta_k^{CD} = -\frac{\|g_{k+1}\|^2}{g_k^T d_k}; \quad \beta_k^{LS} = -\frac{g_{k+1}^T y_k}{g_k^T d_k}; \quad \beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T y_k}.$$

其中 $\|\cdot\|$ 表示为 Euclidean 范数, $y_k = g_{k+1} - g_k$. 当 f 是一般连续可微函数或者适用非精确线搜索时, HS、PRP、LS 三种方法数值试验效果较好, 但是理论性质较差; FR、CD、DY 三种方法理论性质较好, 但是数值试验效果不理想. 为了获得理论收敛性和数值计算效果都更佳的共轭梯度方法, Touati-Ahmed 和 Storey [9] 首次引入混合共轭梯度法, 将数值表现较好的 PRP 方法与理论收敛性较好的 FR 方法混合, 其参数如: $\beta_k^{TS} = \max\{0, \min\{\beta_k^{PRP}, \beta_k^{FR}\}\}$. 混合共轭梯度法的提出, 使得共轭梯度方法的理论性质和数值试验都表现的更佳. 随后, 许多学者对混合共轭梯度法进一步进行研究, 见文献 [10]、[11]、[12]、[13] 等.

2008 年, Neculai Andrei [14] 为了同时利用 HS 方法数值试验效果好, DY 方法理论性质好这样的优点, 提出了一种修正的混合共轭梯度算法, 其参数 β_k^{HSDY} 是 β_k^{HS} 和 β_k^{DY} 的凸组合, 即:

$$\beta_k^{HSDY} = (1 - \vartheta_k)\beta_k^{HS} + \vartheta_k\beta_k^{DY}, \quad (1.8)$$

其中 $\vartheta_k \in [0, 1]$. 注意到当 $\vartheta_k = 0$ 时, HSDY 方法等价于 HS 方法; 当 $\vartheta_k = 1$ 时, HSDY 方法等价于 DY 方法. 由 (1.3) 式可得:

$$d_{k+1} = -g_{k+1} + (1 - \vartheta_k) \frac{g_{k+1}^T y_k}{d_k^T y_k} d_k + \vartheta_k \frac{\|g_{k+1}\|^2}{d_k^T y_k} d_k, \quad (1.9)$$

将上式 d_{k+1} 记为 d_{k+1}^{HD} . ϑ_k 的计算方式是采用牛顿方向去逼近 d_{k+1}^{HD} , 从而获得

$$\vartheta_k = -\frac{s_k^T g_{k+1}}{g_k^T g_{k+1}}. \quad (1.10)$$

Neculai Andrei 证明了 HSDY 方法的收敛性, 其数值试验的结果比 HS 方法和 DY 方法都更有效.

2015 年 Zhifeng Dai [15] 等对文献 [14] 中定理 2 和定理 4 进行修正, 并提出条件 $0 < c_1 \leq \theta_k < 1$ 可以被删除.

2020 年 Xiuyun Zheng [16] 等对文献 [14] 中定理 2 和定理 4 和文献 [15] 中定理 2.1 和定理 3.1 进行了进一步修正, 并指出条件 $0 < c_1 \leq \theta_k < 1$ 不能被删除.

1977 年 perry [17] 和 1978 年 shanno [18] 提出自调比无记忆 BFGS 方法:

$$d_{k+1} = -H_{k+1}g_{k+1}, \quad (1.11)$$

$$H_{k+1} = \frac{1}{\tau_k} \left(I - \frac{s_k y_k^T + y_k s_k^T}{s_k^T y_k} \right) + \left(1 + \frac{\|y_k\|^2}{\tau_k s_k^T y_k} \right) \frac{s_k s_k^T}{s_k^T y_k}, \quad (1.12)$$

即

$$d_{k+1} = -g_{k+1} + \left[\frac{g_{k+1}^T y_k}{s_k^T y_k} - \left(\tau_k + \frac{\|y_k\|^2}{s_k^T y_k} \right) \frac{g_{k+1}^T d_k}{d_k^T y_k} \right] s_k + \frac{g_{k+1}^T d_k}{d_k^T y_k} y_k. \quad (1.13)$$

若删除 (1.13) 最后一项, 再取 $\tau_k = \frac{\|y_k\|^2}{s_k^T y_k}$, 则上述方法就退化到 2005 年 Hager 和 Zhang [1] 提出的共轭梯度法, 其共轭参数为:

$$\beta_k^N = \frac{g_{k+1}^T y_k}{d_k^T y_k} - \frac{2\|y_k\|^2 g_{k+1}^T d_k}{(d_k^T y_k)^2}. \quad (1.14)$$

2 一种修正的 HSDY 方法

HSDY 方法同时具有 HS 方法和 DY 方法各自的优点, 然而 HSDY 方法是利用牛顿方向去逼近 d_{k+1} 获得 ϑ_k , 在这种情况下要求目标函数是二次可微并且对应的 Hesse 矩阵是正定的, 但是, 这样一来使得计算复杂、占用内存较大. 因此, 受 Perry [17] 和 shanno [18] 思想的启发, 本文提出如下策略计算 ϑ_k :

用自调比无记忆 BFGS 去逼近搜索方向 d_{k+1} , 即

$$d_{k+1} = -g_{k+1} + \left[\frac{g_{k+1}^T y_k}{s_k^T y_k} - \left(\tau_k + \frac{\|y_k\|^2}{s_k^T y_k} \right) \frac{g_{k+1}^T d_k}{d_k^T y_k} \right] s_k + \frac{g_{k+1}^T d_k}{d_k^T y_k} y_k. \quad (2.1)$$

计算化简得

$$\vartheta_k^N = -\tau_k \alpha_k \frac{d_k^T g_{k+1}}{g_k^T g_{k+1}} = -\tau_k \frac{s_k^T g_{k+1}}{g_k^T g_{k+1}}. \quad (2.2)$$

其中, τ_k 与 Dai 和 Kou [19] 文一致, 即

$$\tau_k^1 = \min \left\{ 1, \frac{\|y_k\|^2}{s_k^T y_k} \right\}, \tau_k^2 = \min \left\{ 1, \frac{s_k^T y_k}{\|s_k\|^2} \right\}. \quad (2.3)$$

本文将参数 τ_k^1 和 τ_k^2 对应的方法分别记作 HSDY1 方法和 HSDY2 方法, 即

$$\beta_k^{HSDY1} = (1 + \tau_k^1 \frac{s_k^T y_k}{g_k^T g_{k+1}}) \beta_k^{HS} - \tau_k^1 \frac{s_k^T y_k}{g_k^T g_{k+1}} \beta_k^{DY}, \quad (2.4)$$

$$\beta_k^{HSDY2} = (1 + \tau_k^2 \frac{s_k^T y_k}{g_k^T g_{k+1}}) \beta_k^{HS} - \tau_k^2 \frac{s_k^T y_k}{g_k^T g_{k+1}} \beta_k^{DY}. \quad (2.5)$$

3 下降性和充分下降性

为方便后续讨论, 在此先介绍对一些目标函数的基本假设.

假设 (A) 水平集 $\mathcal{L} = \{x \in R^n \mid f(x) \leq f(x_1)\}$ 是有界的, 其中 $x_1 \in R^n$ 为算法初始点, 即存在常数 $B > 0$, 使得

$$\|x\| \leq B, \quad \forall x \in \mathcal{L}. \quad (3.1)$$

假设 (B) 目标函数 $f(x)$ 在水平集 \mathcal{L} 的某些邻域 \mathcal{N} 内是连续可微的, 且梯度函数 $g(x)$ 是 Lipschitz 连续的, 即存在常数 $L > 0$, 使得

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathcal{N}. \quad (3.2)$$

由假设 (A)-(B) 可知, 存在常数 $\bar{\gamma} \geq 0$, 使得

$$\|g(x)\| \leq \bar{\gamma}, \quad \forall x \in \mathcal{L}. \quad (3.3)$$

引理 3.1 若 $f(x)$ 对于满足假设 (A)-(B). 考虑具有 (1.2)-(1.3) 形式的共轭梯度法, 其中 d_k 是一个下降方向即满足 $g_{k+1}^T d_{k+1} \leq 0$, α_k 满足强 Wolfe 条件 (1.4) 和 (1.5), 则有

$$\sum_{k \geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \quad (3.4)$$

引理 3.1 由 Zoutendijk [20] 和 Wolfe [21, 22] 共同证明, 称 (3.4) 式为 Zoutendijk 条件.

下面的定理 3.1 和定理 3.2 证明了 HSDY1 方法和 HSDY2 方法的下降性.

定理 3.1 假设 (A)-(B) 成立. 考虑 HSDY1 共轭梯度法, 对任意的 $k \geq 0$, 若 $d_k^T y_k \neq 0$, 则有

$$g_{k+1}^T d_{k+1} \leq 0. \quad (3.5)$$

证 由 (1.3) 可知, 当 $k = 0$ 时, 有 $g_1^T d_1 = -\|g_1\|^2$. 当 $k \geq 1$ 时, 我们可以得到

$$\begin{aligned} g_{k+1}^T d_{k+1} &= -\|g_{k+1}\|^2 + \beta_k^{HSDY1} g_{k+1}^T d_k \\ &= -\|g_{k+1}\|^2 + \left(1 + \tau_k \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}}\right) \frac{g_{k+1}^T d_k g_{k+1}^T y_k}{y_k^T d_k} - \tau_k \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \frac{g_{k+1}^T d_k \|g_{k+1}\|^2}{y_k^T d_k}. \end{aligned} \quad (3.6)$$

接下来对 τ^1 进行分类讨论:

当 $\tau^1 = \min \left\{1, \frac{\|y_k\|^2}{s_k^T y_k}\right\} = 1$ 时, 此时与 HSDY 方法的证明过程一致, 详见 [14]、[15]、[16].

当 $\tau^1 = \min \left\{1, \frac{\|y_k\|^2}{s_k^T y_k}\right\} = \frac{\|y_k\|^2}{s_k^T y_k}$ 时, 则有

$$\begin{aligned} g_{k+1}^T d_{k+1} &= -\|g_{k+1}\|^2 + \left(1 + \frac{\|y_k\|^2}{s_k^T y_k} \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}}\right) \frac{g_{k+1}^T d_k g_{k+1}^T y_k}{y_k^T d_k} \\ &\quad - \frac{\|y_k\|^2}{s_k^T y_k} \frac{g_{k+1}^T s_k g_{k+1}^T d_k \|g_{k+1}\|^2}{g_k^T g_{k+1} y_k^T d_k} \\ &= -\|g_{k+1}\|^2 \left(1 + \frac{\|y_k\|^2}{s_k^T y_k} \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \frac{g_{k+1}^T d_k}{y_k^T d_k}\right) + \left(1 + \frac{\|y_k\|^2}{s_k^T y_k} \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}}\right) \|g_{k+1}\|^2 \\ &= -\|g_{k+1}\|^2 \left(\frac{\|y_k\|^2}{s_k^T y_k} \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \frac{g_{k+1}^T d_k}{y_k^T d_k} - \frac{\|y_k\|^2}{s_k^T y_k} \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}}\right) \\ &= -\|g_{k+1}\|^2 \frac{\|y_k\|^2}{s_k^T y_k} \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \left(\frac{g_{k+1}^T d_k}{y_k^T d_k} - 1\right). \end{aligned} \quad (3.7)$$

在 (1.4)-(1.5) 式下有 $s_k^T y_k > 0$ 成立, 所以

$$\frac{\|y_k\|^2}{s_k^T y_k} > 0. \quad (3.8)$$

由 [14] 中证明可得

$$\frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} < 0 \quad (3.9)$$

以及

$$\frac{g_{k+1}^T d_k}{y_k^T d_k} - 1 < 0. \quad (3.10)$$

再由 (3.8)、(3.9) 和 (3.10) 可知

$$-\frac{\|y_k\|^2}{s_k^T y_k} \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \left(\frac{g_{k+1}^T d_k}{y_k^T d_k} - 1 \right) \leq 0.$$

即 $g_{k+1}^T d_{k+1} \leq 0$ 成立, 故 HSDY1 方法是具有下降性的. 证毕.

定理 3.2 假设 (A)-(B) 成立. 考虑 HSDY2 共轭梯度法, 对任意的 $k \geq 0$, 若 $d_k^T y_k \neq 0$, 则有

$$g_{k+1}^T d_{k+1} \leq 0. \quad (3.11)$$

注 3.1 其证明方法与定理 3.1 一致, 需要注意的是在 (1.4)-(1.5) 式下, 仍然可以得到 $\frac{s_k^T y_k}{\|s_k\|^2} > 0$. 从而 HSDY2 方法是具有下降性的.

下面定理 3.3 和定理 3.4 证明了 HSDY1 方法和 HSDY2 方法的充分下降性.

定理 3.3 假设 (A)-(B) 成立. 考虑 HSDY1 共轭梯度法, 若 $d_k^T y_k \neq 0$ 成立, 则存在一自适应参数 $c_1 > 0$ 使

$$g_{k+1}^T d_{k+1} \leq -c_1 \|g_{k+1}\|^2, \quad (3.12)$$

其中

$$c_1 = \begin{cases} 1, & k = 1, \\ \frac{\|y_k\|^2}{s_k^T y_k} \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \left(\frac{g_{k+1}^T d_k}{y_k^T d_k} \right) - 1, & k > 1 \text{ 且 } \tau^1 = \frac{\|y_k\|^2}{s_k^T y_k}, \\ \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \left(\frac{g_{k+1}^T d_k}{y_k^T d_k} \right) - 1, & k > 1 \text{ 且 } \tau^1 = 1. \end{cases}$$

证 由 (1.3) 可知, 当 $k = 0$ 时, 有 $g_1^T d_1 = -\|g_1\|^2$. 此时 $c_1 = 1$.

当 $k \geq 1$ 时, 由 (3.9) 式可知

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 \frac{\|y_k\|^2}{s_k^T y_k} \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \left(\frac{g_{k+1}^T d_k}{y_k^T d_k} - 1 \right). \quad (3.13)$$

由 (3.5)、(3.10) 和 (3.12) 式可知, 存在一自适应参数 $c_1 > 0$ 使得

$$c_1 = \frac{\|y_k\|^2}{s_k^T y_k} \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \left(\frac{g_{k+1}^T d_k}{y_k^T d_k} - 1 \right). \quad (3.14)$$

又或者当 $\tau^1 = \min \left\{ 1, \frac{\|y_k\|^2}{s_k^T y_k} \right\} = 1$ 时, 有

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \left(\frac{g_{k+1}^T d_k}{y_k^T d_k} - 1 \right). \quad (3.15)$$

由 (3.5), (3.10) 和 (3.12) 式可知, 存在一自适应参数 $c_1 > 0$ 使得

$$c_1 = \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \left(\frac{g_{k+1}^T d_k}{y_k^T d_k} - 1 \right). \quad (3.16)$$

所以 HSDY1 方法是具有充分下降性. 证毕.

定理 3.4 假设 (A)-(B) 成立. 考虑 HSDY2 共轭梯度法, 若 $d_k^T y_k \neq 0$ 成立, 则存在一自适应参数 $c_2 > 0$ 使

$$g_{k+1}^T d_{k+1} \leq -c_2 \|g_{k+1}\|^2, \quad (3.17)$$

其中

$$c_2 = \begin{cases} 1, & k = 1, \\ \frac{s_k^T y_k}{\|s_k\|^2} \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \left(\frac{g_{k+1}^T d_k}{y_k^T d_k} - 1 \right), & k > 1 \text{ 且 } \tau^2 = \frac{s_k^T y_k}{\|s_k\|^2}, \\ \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \left(\frac{g_{k+1}^T d_k}{y_k^T d_k} \right) - 1, & k > 1 \text{ 且 } \tau^2 = 1. \end{cases}$$

证 由 (1.3) 式, 当 $k = 0$ 时, 有 $g_1^T d_1 = -\|g_1\|^2$, 此时 $c_2 = 1$.

当 $k \geq 1$ 时, 由 (3.9) 式可知:

$$g_{k+1}^T d_{k+1} = -g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 \frac{s_k^T y_k}{\|s_k\|^2} \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \left(\frac{g_{k+1}^T d_k}{y_k^T d_k} - 1 \right). \quad (3.18)$$

由 (3.10), (3.11) 和 (3.17) 式可知, 存在一自适应参数 $c_2 > 0$ 使得

$$c_2 = \frac{s_k^T y_k}{\|s_k\|^2} \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \left(\frac{g_{k+1}^T d_k}{y_k^T d_k} - 1 \right). \quad (3.19)$$

又或者当 $\tau^2 = \min \left\{ 1, \frac{s_k^T y_k}{\|s_k\|^2} \right\} = 1$ 时, 则有

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \left(\frac{g_{k+1}^T d_k}{y_k^T d_k} - 1 \right). \quad (3.20)$$

由 (3.10), (3.11) 和 (3.17) 式可知, 存在一自适应参数 $c_2 > 0$ 使得

$$c_2 = \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \left(\frac{g_{k+1}^T d_k}{y_k^T d_k} - 1 \right). \quad (3.21)$$

所以 HSDY2 方法是具有充分下降性. 证毕.

4 全局收敛性

定理 3.1–定理 3.4 证明了 HSDY1 和 HSDY2 方法的下降性和充分下降性. 接下来, 我们将证明 HSDY1 和 HSDY2 方法对一致凸函数具有强收敛性, 再证明 HSDY1 和 HSDY2 方法对一般函数具有全局收敛性.

我们先给出一致凸函数的等价定义: 如果存在常数 $\mu > 0$, 使得

$$(\nabla f(x) - \nabla f(y))^T(x - y) \geq \mu \|x - y\|^2, \quad \forall x, y \in \mathcal{L}. \quad (4.1)$$

定理 4.1 假设 (A)-(B) 成立. 考虑 HSDY1 共轭梯度法, 其中步长 α_k 满足强 Wolfe 线搜索 (1.4) 和 (1.5). 若 f 为一致凸函数, 则有

$$\lim_{k \rightarrow \infty} g_k = 0. \quad (4.2)$$

证 由 Lipschitze 条件 (3.2) 式以及 f 为一致凸函数, 有

$$\|y_k\| \leq L \|s_k\|, \quad (4.3)$$

$$d_k^T y_k \geq \mu \|d_k\| \|s_k\|. \quad (4.4)$$

当 $\tau^1 = \min \left\{ 1, \frac{\|y_k\|^2}{s_k^T y_k} \right\} = 1$ 时, 由 (1.3), (4.3) 和 (4.4), 利用三角不等式及柯西不等式有

$$\begin{aligned} |\beta_k^{HSDY1}| &= |(1 - \vartheta_k^N) \beta^{HS} + \vartheta_k^N \beta^{DY}| \\ &= \left| \left(1 + \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \right) \frac{g_{k+1}^T y_k}{y_k^T d_k} - \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \frac{\|g_{k+1}\|^2}{y_k^T d_k} \right| \\ &= \left| \frac{g_{k+1}^T y_k}{y_k^T d_k} + \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \left(\frac{g_{k+1}^T y_k}{y_k^T d_k} - \frac{\|g_{k+1}\|^2}{y_k^T d_k} \right) \right| \\ &= \left| \frac{g_{k+1}^T y_k}{y_k^T d_k} - \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \frac{g_{k+1}^T g_k}{y_k^T d_k} \right| \\ &\leq \left| \frac{g_{k+1}^T y_k}{y_k^T d_k} + \frac{g_{k+1}^T s_k}{y_k^T d_k} \right| \leq \left| \frac{(L+1) \|g_{k+1}\| \|s_k\|}{\mu \|d_k\| \|s_k\|} \right| \leq \frac{(L+1) \|g_{k+1}\|}{\mu \|d_k\|}, \end{aligned} \quad (4.5)$$

则

$$\|d_{k+1}\| \leq \|g_{k+1}\| + |\beta_k| \|d_k\| \leq \|g_{k+1}\| + \frac{(L+1) \|g_{k+1}\|}{\mu \|d_k\|} \|d_k\| \leq \left(1 + \frac{(L+1)}{\mu} \right) \|g_{k+1}\|. \quad (4.6)$$

当 $\tau^1 = \min \left\{ 1, \frac{\|y_k\|^2}{s_k^T y_k} \right\} = \frac{\|y_k\|^2}{s_k^T y_k}$ 时, 由 (1.3), (4.3) 和 (4.4), 利用三角不等式及柯西不等式有

$$\begin{aligned} |\beta_k^{HSDY1}| &= |(1 - \vartheta_k^N) \beta^{HS} + \vartheta_k^N \beta^{DY}| \\ &= \left| \left(1 + \frac{\|y_k\|^2}{s_k^T y_k} \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \right) \frac{g_{k+1}^T y_k}{y_k^T d_k} - \frac{\|y_k\|^2}{s_k^T y_k} \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \frac{\|g_{k+1}\|^2}{y_k^T d_k} \right| \\ &= \left| \frac{g_{k+1}^T y_k}{y_k^T d_k} - \frac{\|y_k\|^2}{s_k^T y_k} \frac{g_{k+1}^T s_k}{y_k^T d_k} \right| \\ &\leq \left| \frac{g_{k+1}^T y_k}{y_k^T d_k} + \frac{\|y_k\|^2}{(y_k^T d_k)^2} \right| \\ &\leq \left| \frac{L \|g_{k+1}\| \|s_k\|}{\mu \|d_k\| \|s_k\|} + \frac{L^2 \|g_{k+1}\| \|s_k\|^2}{\mu^2 \|d_k\| \|s_k\|^2} \right| \\ &= \left(\frac{L}{\mu} + \frac{L^2}{\mu^2} \right) \frac{\|g_{k+1}\|}{\|d_k\|}, \end{aligned} \quad (4.7)$$

则

$$\begin{aligned}
 \|d_{k+1}\| &\leq \|g_{k+1}\| + |\beta_k^{HSDY1}| \|d_k\| \\
 &\leq \|g_{k+1}\| + \left(\frac{L}{\mu} + \frac{L^2}{\mu^2}\right) \frac{\|g_{k+1}\|}{\|d_k\|} \|d_k\| \\
 &\leq \left(1 + \frac{L}{\mu} + \frac{L^2}{\mu^2}\right) \|g_{k+1}\|.
 \end{aligned} \tag{4.8}$$

根据 Zoutentijk 条件 (3.4) 和定理 2.3 可得:

$$\sum_{k \geq 1} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty, \tag{4.9}$$

由 (4.6) 和 (4.9) 式可得:

$$\sum_{k \geq 1} \|g_k\|^2 < \infty. \tag{4.10}$$

所以 (4.2) 式成立. 证毕.

接下来, 我们将证明三个重要的引理.

引理 4.1 假设 (A)-(B) 成立. 考虑 HSDY1 共轭梯度法, 其中步长 α_k 满足强 Wolfe 线搜索 (1.4) 和 (1.5). 若存在一个常数 $\gamma > 0$, 则有

$$\|g_k\| \geq \gamma, \quad \forall k \geq 1, \tag{4.11}$$

则 $d_k \neq 0$ 且

$$\sum_{k \geq 1} \|u_k - u_{k-1}\|^2 < \infty, \tag{4.12}$$

其中 $u_k = \frac{d_k}{\|d_k\|}$.

证明 因为充分下降性条件 (3.12) 成立, 所以 $d_k \neq 0$, 故对于 u_k 的定义是有意义的. 将 (2.4) 式中的 β_k^{HSDY1} 分为如下两部分:

$$\beta_k^1 = (1 - \vartheta_k^N) \beta_k^{HS}, \quad \beta_k^2 = \vartheta_k^N \beta_k^{DY}. \tag{4.13}$$

记

$$w_k := \frac{-g_k + \beta_{k-1}^1 d_{k-1}}{\|d_k\|}, \quad \delta_k := \frac{\beta_{k-1}^2 \|d_{k-1}\|}{\|d_k\|}. \tag{4.14}$$

当 $k \geq 2$ 时, $d_k = -g_k + \beta_{k-1} d_{k-1}$, 得

$$u_k = w_k + \delta_k u_{k-1}. \tag{4.15}$$

由 $\|u_k\| = \|u_{k-1}\| = 1$ 和 (4.15) 得

$$\|w_k\| = \|u_k - \delta_k u_{k-1}\| = \|\delta_k u_k - u_{k-1}\|. \tag{4.16}$$

又因为 $\delta_k \geq 0$, 再根据 (4.16) 我们可以得到

$$\begin{aligned} \|u_k - u_{k-1}\| &\leq \|(1 + \delta_k)u_k - (1 + \delta_k)u_{k-1}\| \\ &\leq \|u_k - \delta_k u_{k-1}\| + \|\delta_k u_k - u_{k-1}\| \\ &= 2\|w_k\|. \end{aligned} \quad (4.17)$$

存在一个自适应参数 $c_3 > 0$, 使得 $c_3 = \frac{\|g_{k+1}\|^2}{|g_k^T g_{k+1}|}$. 由 (4.13) 得

$$\|-g_k + \beta_{k-1}^2 d_{k-1}\| \leq \|g_k\| + |\beta_{k-1}^2| \|d_{k-1}\| \leq (1 + c_3)\|g_k\|, \quad (4.18)$$

另外, 再根据 (4.14), (4.17) 和 (4.18) 可知

$$\|u_k - u_{k-1}\| \leq 2\|w_k\| \leq 2(1 + c_3) \frac{\|g_k\|}{\|d_k\|}. \quad (4.19)$$

最后, 结合 (4.11), 再根据 (3.4) 和 (3.12) 可以得到

$$\sum_{k \geq 1} \frac{g_k^2}{d_k^2} \leq \frac{1}{\gamma^2} \sum_{k \geq 1} \frac{g_k^4}{d_k^2} \leq \frac{1}{\gamma^2 c_1^2} \sum_{k \geq 1} \frac{(g_k^T d_k)^2}{d_k^2} < +\infty. \quad (4.20)$$

由 (4.19) 和 (4.20) 式可知

$$\sum_{k \geq 1} \|u_k - u_{k-1}\|^2 \leq 4(1 + c_3)^2 \sum_{k \geq 1} \frac{\|g_k\|^2}{\|d_k\|^2} < \infty. \quad (4.21)$$

证毕.

文献 [24] 中给出的性质 (*), 对于收敛性结果的证明和分析产生了重要的作用. 下面证明 HSDY1 和 HSDY2 方法满足性质 (*).

引理 4.2 考虑由 (1.2), (1.3) 以及 (2.4) 产生的 HSDY1, 其中步长 α_k 满足强 Wolfe 线搜索 (1.4) 和 (1.5), 且 (4.12) 成立, 则 HSDY1 方法满足性质 (*), 即若存在常数 $b > 1$ 和 $\lambda > 0$, 使得对所有的 k 有 $|\beta_k| \leq b$ 以及 $\|s_{k-1}\| \leq \lambda \Rightarrow |\beta_k| \leq \frac{1}{b}$.

证明 因为水平集 \mathcal{L} 有界, 可知点列 $\{x_k\}$ 有界, 由梯度函数 g 的连续性, 则存在 $\bar{\gamma} > 0$, 使得

$$\|x_k\| \leq \bar{\gamma}, \quad \|g_k\| \leq \bar{\gamma}, \quad \forall k \geq 1, \quad (4.22)$$

根据 (1.6), (3.5) 和 (4.11), 我们可以得到

$$d_k^T y_k \geq -(1 - \sigma)g_k^T d_k \geq c_1(1 - \sigma)g_k^2 \geq c_1(1 - \sigma)\gamma^2 > 0. \quad (4.23)$$

由 (4.23) 最左端的不等式得

$$\sigma g_k^T d_k \geq \frac{\sigma}{\sigma - 1} d_k^T y_k. \quad (4.24)$$

进一步, 由 (4.23) 和 (4.24) 有

$$\frac{g_{k+1}^T d_k}{d_k^T y_k} \geq \frac{\sigma}{\sigma - 1}, \quad (4.25)$$

再根据 (3.5) 和 (1.6) 得

$$\frac{g_{k+1}^T d_k}{d_k^T y_k} = 1 + \frac{g_k^T d_k}{d_k^T y_k} \leq 1, \quad (4.26)$$

将 (4.25) 和 (4.26) 合并得

$$\frac{\sigma}{\sigma - 1} \leq \frac{g_{k+1}^T d_k}{d_k^T y_k} \leq 1, \quad (4.27)$$

进而有

$$\left| \frac{g_{k+1}^T d_k}{d_k^T y_k} \right| \leq M, \quad (4.28)$$

其中 $M = \max \left\{ 1, \frac{\sigma}{1 - \sigma} \right\}$. 由 (4.22) 式可以得到 $\|s_k\| = \|x_{k+1} - x_k\| \leq 2\bar{\gamma}$ 成立.

下面根据 (2.4), (4.22), (4.23), (4.28) 及 $\|y_k\| \leq L\|s_k\|$ 对 $|\beta_k|$ 分类讨论.

对 HSDY1 方法有

$$\begin{aligned} |\beta_k| &= \left| \frac{g_{k+1}^T y_k}{y_k^T d_k} + \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \left(\frac{g_{k+1}^T y_k}{y_k^T d_k} - \frac{\|g_{k+1}\|^2}{y_k^T d_k} \right) \right| \\ &= \left| \frac{g_{k+1}^T y_k}{y_k^T d_k} - \frac{g_{k+1}^T s_k}{g_k^T g_{k+1}} \frac{g_{k+1}^T g_k}{y_k^T d_k} \right| \\ &\leq \left| \frac{g_{k+1}^T y_k}{y_k^T d_k} \right| + \left| \frac{g_{k+1}^T s_k}{y_k^T d_k} \right| \\ &\leq \frac{(L+1)\|g_{k+1}\|\|s_k\|}{\bar{c}(1-\sigma)\gamma^2}. \end{aligned} \quad (4.29)$$

令 $N = \frac{(L+1)\|g_{k+1}\|\|s_k\|}{\bar{c}_1(1-\sigma)\gamma^2}$, 则有 $|\beta_k| \leq N\|s_k\|$, 定义 $b = 2N\bar{\gamma}$, $\lambda = \frac{1}{2N^2\bar{\gamma}}$, 则对所有 $k \geq 1$, 有

$$|\beta_k| \leq b, \quad (4.30)$$

且

$$\|s_k\| \leq \lambda \implies |\beta_k| \leq \frac{1}{b}. \quad (4.31)$$

由 (4.30) 和 (4.31) 式, HSDY1 方法满足性质 (*).

定义 4.1 正整数集为 N^* , 常数 $\lambda > 0$ 和一个正整数 Δ , 记

$$K_{k,\Delta}^\lambda := \{i \in N^* : k \leq i \leq k + \Delta - 1, \|s_{i-1}\| > \lambda\}. \quad (4.32)$$

令 $|K_{k,\Delta}^\lambda|$ 为 $K_{k,\Delta}^\lambda$ 中的元素个数.

引理 4.3 假设 (A)-(B) 成立. 考虑形如 (1.2), (1.3) 和 (2.4) 的 HSDY1 方法, 步长 α_k 通过强 Wolfe 线搜索 (1.4) 和 (1.5) 计算得到. 如果 (4.11) 成立, 则存在常数 $\lambda > 0$, 使得任意 $\Delta \in N^*$ 和任意指标集 k_0 , 存在 $k \geq k_0$, 使得

$$|K_{k,\Delta}^\lambda| > \frac{\Delta}{2}. \quad (4.33)$$

注 4.1 由性质 (*) 中证明到的 HSDY1 方法满足性质 (A) 以及 $\|d_k\|^2$ 至多线性增长, 类似于文献 [24] 中引理 3.5 的证明可得到上述引理的证明.

下面是 HSDY1 方法全局收敛性定理.

定理 4.2 假设 (A)-(B) 成立. 考虑形如 (1.2), (1.3) 和 (2.5) 式的 HSDY1 方法, 步长 α_k 满足强 Wolfe 线搜索 (1.4) 和 (1.5), 则有

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0.$$

由引理 4.1-4.3, HSDY1 方法满足性质 (*) 以及 $\{x_k\}$ 是有界的, 参照文献 [24] 中定理 4.3 证明, 可以证得上述定理.

HSDY2 方法的全局收敛性证明与 HSDY1 的方法类似.

5 数值实验

由于 DK 方法是当前共轭梯度算法研究领域中公认为的数值计算效果最好的方法之一, 在这一小节中, 为了验证 HSDY, HSDY1, HSDY2 方法的有效性, 我们将这三种方法同 DK+ 方法作数值比较, 其中 $\eta = 0.5$. 本文选取文献 [23] 中的 62 个测试问题, 维数 1000-100000. 计算中这四个方法的步长 α_k 都通过近似 Wolfe 线搜索 (1.6) 和 (1.7) 计算获得, 线搜索中的具体参数取 $\delta = 0.1, \sigma = 0.9$. 算法的测试环境为华硕 Windows 10 操作系统 Intel(R) Core(TM) i5-3337U CPU@1.80GHz 1.80GHz, 4GB 内存.

对 HSDY, HSDY1, HSDY2 和 DK 方法的数值进行数据处理结果见表 1. 我们得到下面的比值表 (见表 1) 和性能曲线图 (见图 1-4). 通过表 1 中的比值结果可以看出, HSDY、HSDY1 以及 HSDY2 方法都优于 DK 方法, 且 HSDY2 方法的计算效果最好.

表 1: DK, HSDY, HSDY1 和 HSDY2 方法的计算效率

DK	HSDY	HSDY1	HSDY2
1.0000	0.8224	0.8642	0.8061

本文采用文献 [25] 的技术得到以下性能曲线图. 下图 1-4 分别对应的是 DK, HSDY, HSDY1 以及 HSDY2 四种方法, 在近似 Wolfe 线搜索条件下解决测试问题所用的计算时间, 函数计算次数, 梯度计算次数以及迭代次数的性能曲线图. 从图中也可以看出 HSDY, HSDY1 以及 HSDY2 方法都优于 DK 方法, 且 HSDY1 方法的计算效果最好.

6 结束语

本文通过新的放缩方式, 提出两个参数自适应选取的 HSDY 共轭梯度法; 通过采用 BFGS 方法而不是牛顿法去逼近搜索方向, 在此条件下, 不要求目标函数是二次可微的并且对应的 Hesse 矩阵是正定的. 本文提出的这两类共轭梯度法 HSDY1, HSDY2 均可以证明在强 Wolfe 线搜索条件下具有全局收敛性, 这在理论上是很好的; 在计算上与当前公认的数值计算最好的 DK 共轭梯度法都是可比的.

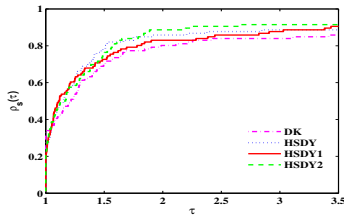


图 1: 计算时间性能曲线

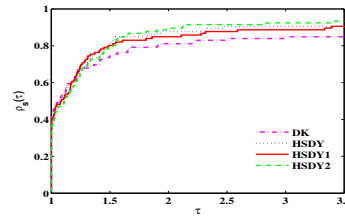


图 2: 函数计算次数性能曲线

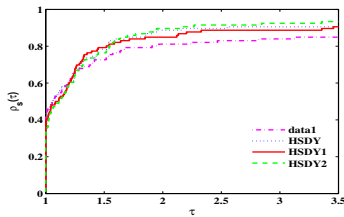


图 3: 梯度计算次数性能曲线

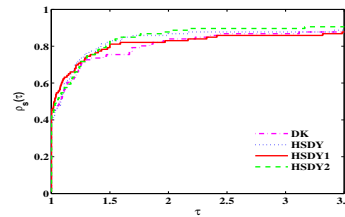


图 4: 迭代次数性能曲线

参 考 文 献

- [1] Hager W W, Zhang H. A new conjugate gradient method with guaranteed descent and an efficient line search[J]. *SIAM Journal on Optimization*, 2005, 16(1): 170–192.
- [2] Hestenes M R, Stiefel E. Methods of conjugate gradients for solving linear systems[J]. *Journal of Research of the National Bureau of Standards*, 1952, 49(6): 409–436.
- [3] Fletcher R, Reeves C M. Function minimization by conjugate gradients[J]. *The Computer Journal*, 1964, 7(2): 149–154.
- [4] Polak E, Ribière G. Note sur la convergence de méthodes de directions conjuguées[J]. *Revue Française d'Informatique et de Recherche Opérationnelle*, 1969, 16(16): 35–43.
- [5] Polyak B T. The conjugate gradient method in extremal problems[J]. *USSR Computational Mathematics and Mathematical Physics*, 1969, 9(4): 94–112.
- [6] Fletcher R. *Practical Methods of Optimization, Vol. 1: Unconstrained Optimization*[M]. New York: John Wiley and Sons, 1987.
- [7] Liu Y, Storey C. Efficient generalized conjugate gradient algorithms, part 1: Theory[J]. *Journal of Optimization Theory and Applications*, 1991, 69(1): 129–137.
- [8] Dai Y H, Yuan Y. A nonlinear conjugate gradient method with a strong global convergence property[J]. *SIAM Journal on Optimization*, 1999, 10(1): 177–182.
- [9] Touati-Ahmed D, Storey C. Global convergent hybrid conjugate gradient method[J]. *Journal of Optimization Theory and Applications*, 1990, 64(2): 379–397.
- [10] Gilbert J C, Nocedal J. Global convergence of the conjugate gradient methods for optimization[J]. *SIAM Journal on Optimization*, 1992, 2(1): 21–42.
- [11] Dai Y H, Yuan Y. An efficient hybrid conjugate method for unconstrained optimization[J]. *Annals of Operations Research*, 2001, 103: 33–47.
- [12] Dai Y H. A nonmonotone conjugate gradient algorithm for unconstrained optimization[J]. *Journal of Systems Science and Complexity*, 2002, 15(2): 139–145.
- [13] Dai Z, Wen F. Another improved Wei-Yao-Liu nonlinear conjugate gradient method with sufficient descent property[J]. *Applied Mathematics and Computation*, 2012, 218(14): 7421–7430.

- [14] Andrei N. Another hybrid conjugate gradient algorithm for unconstrained optimization[J]. Numer. Algor., 2008, 47(2): 143–156.
- [15] Dai Z F, Wen F H. Comments on another hybrid conjugate gradient algorithm for unconstrained optimization by Andrei[J]. Numer. Algor., 2015, 69(2): 337–341.
- [16] Zheng X Y, Dong X L, Shi J R, Yang W. Further comment on another hybrid conjugate gradient algorithm for unconstrained optimization by Andrei[J]. Numer. Algor., 2020, 84(1): 603–608.
- [17] Perry J M. A class of conjugate gradient algorithms with a two-step variable-metric memory[J]. Discussion Paper 269, Center for Mathematical Studies in Economics and Management Sciences, Northwestern University, Evanston, IL., 1977.
- [18] Shanno D F. On the convergence of a new conjugate gradient algorithm[J]. SIAM J Numer Anal., 1978, 15(6): 1247–1257.
- [19] Dai Y H, Kou C X. A nonlinear conjugate gradient algorithm with an optimal property and an improved Wolf line search[J]. SIAM Journal on Optimization, 2013, 23(1): 296–320.
- [20] Zoutendijk G. Nonlinear programming computational methods[J]. Integer and Nonlinear Programming, 1970, 143(1): 37–86.
- [21] Wolfe P. Convergence conditions for ascent methods[J]. SIAM Review., 1969, 11(2): 226–235.
- [22] Wolfe P. Convergence conditions for ascent methods[J]. II: some corrections, SIAM Review., 1971, 13(2): 185–188.
- [23] Gould N I M, Orban D, Toint P L. CUTeR and SifDec: A constained ang unconstrained and testing environment revisited[J]. ACM Trainscations on Mathematical Software, 2003, 29(4): 373–394.
- [24] Gilbert J C, Nocedal J. Global convergence properties of conjugate gradient methods for optimization[J]. SIAM Journal on Optimization, 1992, 2(1): 21–42.
- [25] Dolan E D, Moré J J. Benchmarking optimization software with performance profiles[J]. Mathematical Programming, 2002, 91(2): 201–213.

HSDY MIXED CONJUGATE GRADIENT WAS MODIFIED BASED ON BFGS

ZHU Zi-chang, LIU Li-ping

(*School of Mathematical Sciences, Chongqing Normal University, Chongqing 401331, China*)

Abstract: In this paper, we are concerned with large-scale unconstrained optimization problems. We use BFGS to approximate the search direction, and propose two adaptive conjugate gradient algorithms(HSDY1 and HSDY2) for the HSDY method. We prove that the new algorithms have sufficient descent and global convergence. Numerical experiments are reported to show the computational performance of proposed algorithms are better than HSDY.

Keywords: conjugate gradient methods; BFGS algorithm; sufficient descensibility; global convergence

2010 MR Subject Classification: 90C30; 65K05