

含参数 Dai - Yuan 共轭梯度法及其收敛性

张劲松 李 红

(九江学院 理学院 江西 九江 332005)

摘要: 通过对 Dai - Yuan 共轭梯度法的分析, 将 β_k^{DY} 推广到更一般的形式. 根据搜索方向的下降性要求, 得出含参数 Dai - Yuan 共轭梯度法. 在 Wolfe 条件下, 证明了方法的收敛性; 在强 Wolfe 条件下, 证明了方法的充分下降性. 含参数 Dai - Yuan 共轭梯度法不仅仅是 Dai - Yuan 共轭梯度法在形式上的推广, 其参数的合理选择有望使 Dai - Yuan 共轭梯度法良好的数值表现得到进一步改善.

关键词: Dai - Yuan 共轭梯度法; Wolfe 条件; 收敛性; 充分下降性

中图分类号: O151

文献标识码: A

1952 年 Hestenes 和 Stiefel 提出了求解正定性方程组的共轭梯度法, 1964 年 Fletcher 和 Reeves 将该方法推广到求解无约束优化问题

$$\min f(x) \quad x \in R^n$$

其中 $f: R^n \rightarrow R^1$ 为连续可微函数.

考虑一般的共轭梯度法

$$x_{k+1} = x_k + \alpha_k d_k \quad (1.1)$$

$$d_k = \begin{cases} -g_k, & \text{若 } k=1 \\ -g_k + \beta_k d_{k-1}, & \text{若 } k \geq 2 \end{cases} \quad (1.2)$$

其中 $g_k = \nabla f(x_k)$, $x_k \in R^n$. 步长因子 α_k 满足 Wolfe 条件:

$$f(x_k) - f(x_k + \alpha_k d_k) \geq -\delta \alpha_k g_k^T d_k \quad (1.3)$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k \quad (1.4)$$

其中 $0 < \delta < \sigma < 1$. 假设第 $k-1$ 步的搜索方向 d_{k-1} 为下降方向, 我们希望 β_k 的定义使得第 k 步的搜索方向 d_k 也为下降方向, 即:

$$d_k^T g_k < 0 \quad (1.5)$$

1995 年 Dai 和 Yuan^[1] 提出:

$$\beta_k^{DY} = \|g_k\|^2 / (d_{k-1}^T (g_k - g_{k-1}))$$

其中 $\|\cdot\|$ 表示 Euclid 范数. 上式和 (1.1)、(1.2) 一起被称为 Dai - Yuan 共轭梯度法. 本文拟给出一

个符合要求的更一般的 β_k .

1 含参数 Dai - Yuan 共轭梯度法

由 (1.2) 知, (1.5) 需要 β_k 满足:

$$-\|g_k\|^2 + \beta_k g_k^T d_{k-1} < 0$$

设 $\beta_k > 0$, 则:

$$\|g_k\|^2 / \beta_k > g_k^T d_{k-1}$$

这时可令

$$\begin{aligned} \|g_k\|^2 / \beta_k &= g_k^T d_{k-1} - \lambda g_{k-1}^T d_{k-1} \\ &= d_{k-1}^T (g_k - \lambda g_{k-1}) \end{aligned}$$

这里 λ 为参数, 且 $\lambda \geq 1$. 由此得:

$$\beta_k = \|g_k\|^2 / d_{k-1}^T (g_k - \lambda g_{k-1}) \quad (2.1)$$

我们把 (2.1) 和 (1.1)、(1.2) 一起称为含参数 Dai - Yuan 共轭梯度法. 显然, $\lambda = 1$ 时, 为 Dai - Yuan 共轭梯度法.

2 收敛性证明

引理 3.1^[2] 设目标函数 $f(x)$ 下方有界, 导数 $\nabla f(x)$ 满足 Lipschitz 条件

收稿日期: 2007-12-17

基金项目: 江西省高校省级教改重点资助项目 (JXJG-07-17-47)

作者简介: 张劲松 (1970-) 男, 江西九江人, 讲师, 硕士, 研究方向: 运筹学.

$$\|\nabla f(x) - \nabla f(\tilde{x})\| \leq L \|x - \tilde{x}\|, \forall x, \tilde{x} \in R^n$$

考虑一般方法 $x_{k+1} = x_k + \alpha_k d_k$, 其中 d_k 满足 $d_k^T g_k < 0$, 步长因子 α_k 满足 Wolfe 条件, 则

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

上式常被称为 Zoutendijk 条件.

引理 3.2 考虑 (1.1) 和 (1.2) β_k 按 (2.1) 计算 α_k 满足 (1.3) 和 (1.4) 则 $g_k = 0$ 对某 k 成立, 或者 $d_k^T g_k < 0, \forall k \geq 1$.

证明 设对所有的 $k \geq 1$, 均有 $\|g_k\| > 0$, 对 (1.2) 两端与 g_k 作内积, 并由 (2.1) 得:

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 + \frac{\|g_k\|^2}{d_{k-1}^T (g_k - \lambda g_{k-1})} g_k^T d_{k-1} \\ &= \frac{\|g_k\|^2}{d_{k-1}^T (g_k - \lambda g_{k-1})} \lambda g_{k-1}^T d_{k-1} \end{aligned} \quad (3.1)$$

由 $d_1 = -g_1$ 知 $k=1$ 时结论成立; 设 $g_{k-1}^T d_{k-1} < 0$ 成立, 由 (1.4) 知

$$g_k^T d_{k-1} \geq \sigma g_{k-1}^T d_{k-1} > g_{k-1}^T d_{k-1} > \lambda g_{k-1}^T d_{k-1}$$

由 (3.1) 知 $g_k^T d_k < 0$, 由数学归纳法即得证. 证毕.

推论 3.3 由 (3.1) 式 (2.1) 式也可写成

$$\beta_k = \frac{g_k^T d_k}{\lambda g_{k-1}^T d_{k-1}} \quad (3.2)$$

定理 3.4 设目标函数 $f(x)$ 下方有界, 导数 $\nabla f(x)$ 满足 Lipschitz 条件

$$\|\nabla f(x) - \nabla f(\tilde{x})\| \leq L \|x - \tilde{x}\|, \forall x, \tilde{x} \in R^n$$

考虑方法 (1.1) 和 (1.2), 其中 β_k 按 (2.1) 计算, 步长 α_k 满足 Wolfe 条件 (1.3) 和 (1.4). 则 $g_k = 0$ 对某 k 成立, 或者 $\liminf_{k \rightarrow \infty} \|g_k\| = 0$.

证明 用反证法. 假设结论不真, 则必存在常数 $c > 0$, 使得 $\|g_k\|^2 \geq c$ 对所有 $k \geq 1$ 成立. 由引理 3.2 知, 每一个搜索方向必为下降方向. 现记 (1.2) 为:

$$d_k + g_k = \beta_k d_{k-1}$$

对两端取模平方并移项得:

$$\|d_k\|^2 = \beta_k^2 \|d_{k-1}\|^2 - 2g_k^T d_k - \|g_k\|^2$$

将上式除以 $(g_k^T d_k)^2$ 且由 (3.2) 得:

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} = \frac{\|d_{k-1}\|^2}{\lambda^2 (g_{k-1}^T d_{k-1})^2} - \frac{2}{g_k^T d_k} - \frac{\|g_k\|^2}{(g_k^T d_k)^2} \quad (3.3)$$

记 $l_{k-1} = \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}}$, 由线搜索条件 (1.4) 有:

$$l_{k-1} \leq \sigma.$$

另一方面 (3.1) 式可写为

$$g_k^T d_k = \frac{\lambda}{l_{k-1} - \lambda} \|g_k\|^2 \quad (3.4)$$

代入 (3.3) 式最后两项, 得:

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} = \frac{\|d_{k-1}\|^2}{\lambda^2 (g_{k-1}^T d_{k-1})^2} + \frac{\lambda^2 - l_{k-1}^2}{\lambda^2 \|g_k\|^2} \quad (3.5)$$

在 (3.5) 式中, 依次令 $i=1, 2, \dots, k$ 然后相加, 并注意到 $\|d_1\|^2 = -g_1^T d_1 = \|g_1\|^2$,

$$\begin{aligned} \frac{\|d_k\|^2}{(g_k^T d_k)^2} &= \left(\frac{1}{\lambda^2}\right)^{k-1} \frac{1}{\lambda^2 \|g_1\|^2} \\ &+ \sum_{i=2}^k \left(\frac{1}{\lambda^2}\right)^{k-i} \frac{\lambda^2 - l_{i-1}^2}{\lambda^2 \|g_i\|^2} = \sum_{i=1}^k \left(\frac{1}{\lambda^2}\right)^{k-i} \frac{\lambda^2 - l_{i-1}^2}{\lambda^2 \|g_i\|^2} \end{aligned}$$

由 $\lambda \geq 1$, 题设 $\|g_k\|^2 \geq c$ 及上式, 得:

$$\frac{(g_k^T d_k)^2}{\|d_k\|^2} \geq \frac{c}{k}$$

由调和级数的发散性, 得:

$$\sum_{k \geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = +\infty$$

与 Zoutendijk 条件矛盾. 表明结论为真. 证毕.

容易看出, 定理 3.4 的已知条件中包含引理 3.

1 的已知条件, Zoutendijk 条件应当成立, 由此想到反证法. (3.2) 式在定理 3.4 的证明中起到了很重要的作用, 它使得我们能够建立起递推式 (3.5).

(3.5) 式中的递推项 $\frac{\|d_k\|^2}{(g_k^T d_k)^2}$ 与 Zoutendijk 条件中的

项 $\frac{(g_k^T d_k)^2}{\|d_k\|^2}$ 互为倒数, 通过分析得出矛盾, 从而证明了定理.

另外, Al-Baali^[3] 在建立 FR 方法的收敛结果, 以及 Gilbert 和 Nocedal^[4] 在建立 PRP 方法的收敛结果时, 证明了 (或需要) 线搜索满足充分下降条件. 若步长因子 α_k 满足强 Wolfe 条件, 即, 将 (1.4) 改为 $|g(x_k + \alpha_k d_k)|^T \leq \sigma |g_k^T d_k|$, 则也有类似结论.

推论 3.5 由 $|l_{k-1}| \leq \sigma$ 及 (3.4) 式易得, 含参数 Dai-Yuan 共轭梯度法对 $c = \frac{\lambda}{\lambda + \sigma}$ 满足充分下降条件, 即: $g_k^T d_k \leq -c \|g_k\|^2$.

3 结束语

本文将 β_k^{DY} 推广到更一般的形式, 得到了含参数 Dai-Yuan 共轭梯度法, 使 Dai-Yuan 共轭梯度法是在 $\lambda=1$ 时的特殊情形. 在 Wolfe 条件下, 证明了其收敛性; 在强 Wolfe 条件下, 证明了其充分下降性. 当然, 其数值表现还有待检验, λ 如何选取也

有待进一步分析.

参考文献:

- [1] Dai Y H, Yuan Y. A Nonlinear Conjugate Gradient with a Strong Global Convergence Property [J]. SIAM Journal of Optimization 2000, (10): 177 – 182
- [2] Zoutendijk G. Nonlinear Programming, Computational Methods [A]. Abadie J, ed. Integer and Nonlinear Programming [M]. Amsterdam: North – Holland, 1970, 37 – 86
- [3] Al – Baali M. Descent property and global convergence of the Fletcher – Reeves method with inexact line search [J]. IMA J Numer. Anal, 1985, 5: 121 ~ 124
- [4] Gilbert J C, Nocedal J. Global convergence properties of conjugate gradient methods for optimization [J]. SIAM. J. Optimization, 1992, 2(1): 21 – 42

Dai – Yuan Conjugate Gradient Method with Parameter and Its Convergence

ZHANG Jin – song, LI Hong

(College of Science, Jiujiang University, Jiujiang 332005, China)

Abstract: Through the analysis of Dai – Yuan conjugate gradient method β_k^{DY} will be generalized to a more common form. According to the request of descent property of searching direction, Dai – Yuan conjugate gradient method with parameter is obtained. In the condition of Wolfe line search, its convergence is proved; in the condition of Strong Wolfe line search, its plenary descent property is proved. Dai – Yuan conjugate gradient method with parameter is generalizing its form, the rational election of its parameter is also expected to improve the better numerical representation of Dai – Yuan conjugate gradient method.

Key words: Dai – Yuan conjugate gradient; wolfe condition; convergence; plenary descent property.

(责任编辑: 周尚超)