基于三次模型的线搜索方法及其收敛性*

A Line Search Method with Cubic Model and It's Convergence Property

陆　莎[1],韦增欣[2],袁功林[2]
LU Sha[1],WEI Zeng-xin[2],YUAN Gong-lin[2]

(1.广西师范学院数学科学学院,广西南宁　530023;2.广西大学数学与信息科学学院,广西南宁　530004)

(1. School of Mathematical Science,Guangxi Teachers Education University,Nanning,Guangxi,530023,China;2. School of Mathematics and Information Science,Guangxi University,Nanning,Guangxi,530004,China)

摘要:针对无约束优化问题,给出一种基于三次模型的线搜索型算法,并在适当的条件下证明算法的全局收敛性.该算法以对称矩阵代替原三次模型中的 Hessian 阵,并且不需要保持正定和 Dennis-Moré 条件,它与一般线搜索法不同,在每次迭代中步长可以在下降方向上由显性公式直接确定,从而可以减少搜索计算.

关键词:三次模型　线搜索　无约束优化　全局收敛
中图法分类号:O221.2　　**文献标识码**:A　　**文章编号**:1005-9164(2010)03-0209-03

Abstract:A line search method with conic model for unconstrained optimization was proposed,and its global convergence result was proved under some suitable conditions. In the cubic model of the algorithm, the Hessian matrix of the objective function is replaced by a symmetric matrix without the positive definite assumption and the Dennis-Moré condition. At each iteration, the step-size is computed by an explicit formula on the descent direction which is different from general line search methods so that the search computation can be simplified.

Key words:cubic model,line search method,unconstrained optimization,global convergence

求解非线性规划的迭代算法,通常是在当前迭代点 x_k 构造搜索方向 d_k,然后采用某一线搜索准则确定步长 t_k,由此得到下一迭代点 $x_{k+1}=x_k+t_kd_k$,并判断是否满足算法终止条件.不同的搜索方向 d_k 和确定步长 t_k 的方法构成了求解非线性规划的不同算法.在此类线搜索型算法中,步长因子 t_k 的确定,对算法的收敛性分析有着重要作用,采用不同的线搜索策略可能导致不同的收敛结果.在常用的非精确线搜索策略(Armijo 线搜索、Armijo-Goldstein 线搜索和 Wolfe-Powell 线搜索)基础上,不少学者采用修改的线搜索策略与各种搜索方向相结合,得到性质更好的算法,见文献[1,2]等.

对于无约束优化问题

$$\min f(x),x\in R^n,\tag{1}$$

其中 $f:R^n\rightarrow R$ 为连续可微函数 Nesterov 和 Polyak[3] 给出求解问题(1)的三次正则化牛顿法.不同于一般牛顿法,他们采用 x_k 上的三次超估模型

$$\zeta_{2k}(y)=f(x_k)+\langle\nabla f(x_k),y-x_k\rangle+\frac{1}{2}\langle\nabla^2 f(x_k)(y-x_k),y-x_k\rangle+\frac{L}{6}\|y-x_k\|^3\tag{2}$$

做为原目标函数 $f(x)$ 的近似,取模型(2)的极小点 $T_{M_k}(x_k)$ 作为下一迭代点 x_{k+1}.他们证明该算法在较弱条件下所产生点列的任一极限点满足二阶优化必要条件,且算法具有局部二次收敛性质和很好的全局复杂性结果.在此基础上,Cartis 等[4] 进一步建立了更一般的三次超估模型并将其应用于信赖域算法.他们的研究结果表明,利用三次模型构造的算法具有良好的收敛性质和更小的算法复杂性边界.对三次模型算法的进一步研究还可参见文献[5,6]等.

在本文中,我们给出基于上述三次模型的新线搜索算法,在每次迭代过程中只要确定了下降方向 d_k,其步长 t_k 可由显性公式一次计算得到.同时我们

收稿日期:2010-06-09
作者简介:陆　莎(1974-),女,讲师,主要从事优化理论及其应用。
*国家自然科学基金项目(10761001),广西自然科学基金项目(0991028),广西大学科研基金项目(X081082)资助。

以对称矩阵 B_k 代替原三次模型(2) 中的 $\nabla^2 f(x)$，由此避免了 Hessian 阵的计算. 最后在适当条件下证明了所给算法的收敛性.

1 基于三次模型的新线搜索算法

设 $f:R^n \mapsto R$ 为二阶连续可微函数. 定义点 x_k 上的三次模型函数：

$$f_k(y) = f(x_k) + g^T(x_k)(y - x_k) + \frac{1}{2}(y - x_k)^T B_k (y - x_k) + \frac{M}{6}\| y - x_k \|^3, \quad (3)$$

其中 $g(x) = \nabla f(x)$，B_k 为一 $n \times n$ 阶对称矩阵，$0 < M \in R$，$\|\cdot\|$ 表示 l_2 范数. 为简便，下文记 $g_k = g(x_k)$.

设 $d_k \in R^n$ 是满足 $g_k^T d_k < 0$ 的下降方向，令

$$\varphi_k(t) = t g_k^T d_k + \frac{1}{2} t^2 d_k^T B_k d_k + \frac{M}{6} t^3 \| d_k \|^3,$$

$$t_k = \underset{t \geqslant 0}{\operatorname{argmin}} \varphi_k(t). \quad (4)$$

由于 $\varphi'_k(t) = g_k^T d_k + t d_k^T B_k d_k + \frac{M}{2} t^2 \| d_k \|^3$，$\varphi''_k(t) = d_k^T B_k d_k + M t^2 \| d_k \|^3$，

由(4) 式知，t_k 应满足

$$g_k^T d_k + t_k d_k^T B_k d_k + \frac{M}{2} t_k^2 \| d_k \|^3 = 0, \quad (5)$$

即

$$t_k = \frac{- d_k^T B_k d_k + \sqrt{(d_k^T B_k d_k)^2 - 2M \| d_k \|^3 g_k^T d_k}}{M \| d_k \|^3} > 0. \quad (6)$$

又由(5) 式并注意到 $g_k^T d_k < 0$，有

$$t_k d_k^T B_k d_k + \frac{M}{2} t_k^2 \| d_k \|^3 = - g_k^T d_k > 0,$$

可推出 $\varphi''_k(t_k) = d_k^T B_k d_k + M t_k^2 \| d_k \|^3 > 0$.

于是由(6) 式可确定满足 $\min \varphi_k(t)$ 的 t_k.

给出基于三次模型(3) 的模式算法：

算法 1

初始步 给定初始点 $x_0 \in R^n$ 和初始对称矩阵 $B_0 \in R^{n \times n}$，常数 $\varepsilon > 0$，令 $k := 0$；

步骤 1 若 $\| g_k \| < \varepsilon$，则算法停止；否则找到一下降方向 d_k 满足 $g_k^T d_k < 0$；

步骤 2 由(6) 式得到步长 t_k；

步骤 3 令 $x_{k+1} = x_k + t_k d_k$，更新 $B_{k+1} \in R^{n \times n}$，$k := k + 1$，转步骤 1.

在上述算法中，下降方向 d_k 和 B_k 的更新是容易得到的，比如分别采用共轭下降方向和 BFGS 公式等[7]. 与一般的下降算法相比，算法 1 在确定步长时不需做多次搜索尝试而直接由显性公式计算得到，从而减少了搜索计算.

2 算法的收敛性分析

在对算法的收敛性分析中采用如下假设：

A_1 水平集 $\Omega = \{x \in R^n : f(x) \leqslant f(x_0)\}$ 有界.

A_2 函数 $f(x)$ 的 Hessian 阵 $\nabla^2 f(x)$ 在水平集 Ω 上 Lipschitz 连续，即存在常数 $L > 0$，使得

$$\| \nabla^2 f(x) - \nabla^2 f(y) \| \leqslant L \| x - y \|, \forall x, y \in \Omega. \quad (7)$$

A_3 存在 $\alpha, \beta \geqslant 0$，使得对任意 $d \in R^n$，有

$$\beta \| d \|^2 \geqslant d^T [B_k - \nabla^2 f(x_k)] d \geqslant \alpha \| d \|^2. \quad (8)$$

引理 1 若 $\nabla^2 f(x)$ 在水平集 Ω 上 Lipschitz 连续，则对任意 $x, y \in \Omega$，有

$$\| \nabla f(y) - \nabla f(x) - \nabla^2 f(x)(y - x) \| \leqslant \frac{1}{2} L \| y - x \|^2, \quad (9)$$

$$| f(y) - f(x) - \langle \nabla f(x), y - x \rangle - \frac{1}{2} \langle \nabla^2 f(x)(y - x), y - x \rangle | \leqslant \frac{1}{6} L \| y - x \|^3. \quad (10)$$

引理 1 给出二阶 Lipschitz 连续函数的性质，其详细证明可见文献[3] 的引理 1.

下一个引理给出算法 1 在每一次迭代中所获得的下降量估计，并证明了算法所产生的点列 $\{x_k\}$ 收敛.

引理 2 设 $A_1 \sim A_3$ 成立，$\{x_k\}$ 由算法 1 产生，则当 $M \geqslant \frac{L}{3}$ 时，有

$$f(x_k) - f(x_k + t_k d_k) \geqslant \frac{\alpha}{2} t_k^2 \| d_k \|^2 + \frac{3M - L}{12} t_k^3 \| d_k \|^3 > 0, \quad (11)$$

并且，

$$\lim_{k \to \infty} t_k \| d_k \| = 0. \quad (12)$$

证明 由(3) 式，有

$$f(x_k) - f_k(x_k + t_k d_k) = - t_k g_k^T d_k - \frac{1}{2} t_k^2 d_k^T B_k d_k - \frac{M}{6} t_k^3 \| d_k \|^3 = - t_k g_k^T d_k - t_k^2 d_k^T B_k d_k - \frac{M}{2} t_k^3 \| d_k \|^3 + \frac{1}{2} t_k^2 d_k^T B_k d_k + \frac{M}{3} t_k^3 \| d_k \|^3.$$

由(5) 式代入上式，并注意到 $g_k^T d_k < 0$，有 $d_k^T B_k d_k > - \frac{M}{2} t_k \| d_k \|^3$，于是推出

$$f(x_k) - f_k(x_k + t_k d_k) = \frac{1}{2} t_k^2 d_k^T B_k d_k + \frac{M}{3} t_k^3 \| d_k \|^3 \geqslant - \frac{M}{4} t_k^3 \| d_k \|^3 + \frac{M}{3} t_k^3 \| d_k \|^3 = \frac{M}{12} t_k^3 \| d_k \|^3. \quad (13)$$

利用(10) 式，$f_k(x)$ 的定义以及假设 A_3，有

$$f_k(x_k + t_k d_k) - f(x_k + t_k d_k) \geqslant f_k(x_k + t_k d_k) - f(x_k) - t_k g_k^T d_k - \frac{1}{2} t_k^2 d_k^T \nabla^2 f(x_k) d_k - \frac{L}{6} t_k^3 \| d_k \|^3 =$$

$\frac{1}{2}t_k^2d_k^T[B_k - \nabla^2 f(x_k)]d_k + \frac{M-L}{6}t_k^3\|d_k\|^3 \geqslant \frac{\alpha}{2}t_k^2\|d_k\|^2 + \frac{M-L}{6}t_k^3\|d_k\|^3$. (14)

由(13)、(14)式即知，当 $M \geqslant \frac{L}{3}$ 时，

$$f(x_k) - f(x_k + t_kd_k) \geqslant \frac{\alpha}{2}t_k^2\|d_k\|^2 + \frac{3M-L}{12}t_k^3\|d_k\|^3 > 0.$$

因为 $x_{k+1} = x_k + t_kd_k$，上述不等式两边对 k 求和，并注意到 $f(x)$ 有下界，可推出 $\lim\limits_{k\to\infty} t_k\|d_k\| = 0$，于是 $\lim\limits_{k\to\infty}\|x_{k+1} - x_k\| = 0$.

为保证算法的全局收敛性，一般要求避免搜索方向 d_k 和负梯度方向 $-g_k$ 几乎直交的情形，即要求 d_k 偏离 g_k 正交方向一定的角度. 否则，当 $d_k^Tg_k \to 0$ 时，由于 d_k 几乎不是下降方向，此时并不能保证 $g_k \to 0$. 为此，对搜索方向 d_k 和 $-g_k$ 之间的夹角 $\theta_k \in [0, \frac{\pi}{2})$ 作如下假设：

A_4 对任意 k，$\exists\ \mu > 0$，使得 $\theta_k \leqslant \frac{\pi}{2} - \mu$，其中 $\cos\theta_k = \frac{-g_k^Td_k}{\|g_k\|\ \|d_k\|}$.

定理 1 若假设 $A_1 \sim A_4$ 成立，$\{x_k\}$ 由算法 1 产生，则要么存在某个 k，使得 $g_k = 0$；要么 $\lim\limits_{k\to\infty}\|g_k\| = 0$.

证明 用反证法. 不妨假设对任意 k，$g_k = \nabla f(x_k) \neq 0$. 若命题结论不成立，则存在常数 $\varepsilon > 0$ 和子列 K，使得对所有 $k \in K$，有

$$\|g_k\| \geqslant \varepsilon. \tag{15}$$

由算法构造 $x_{k+1} = x_k + t_kd_k$，其中 t_k 满足(5)式，由(9)式有

$$\|g_{k+1} - g_k - t_k\nabla^2 f(x_k)d_k\| \leqslant \frac{1}{2}Lt_k^2\|d_k\|^2.$$

于是

$$|g_{k+1}^Td_k - g_k^Td_k - t_kd_k^T\nabla^2 f(x_k)d_k| \leqslant \|g_{k+1} - g_k - t_k\nabla^2 f(x_k)d_k\|\ \|d_k\| \leqslant \frac{1}{2}Lt_k^2\|d_k\|^3. \tag{16}$$

同时由(5)式，

$$g_k^Td_k + t_kd_k^T\nabla^2 f(x_k)d_k = -t_kd_k^T[B_k - \nabla^2 f(x_k)]d_k - \frac{M}{2}t_k^2\|d_k\|^3,$$

可推出

$$|g_k^Td_k + t_kd_k^T\nabla^2 f(x_k)d_k| \leqslant \frac{M}{2}t_k^2\|d_k\|^3 + t_kd_k^T[B_k - \nabla^2 f(x_k)]d_k \leqslant \frac{M}{2}t_k^2\|d_k\|^3 + \beta t_k\|d_k\|^2.$$

结合(16)式即得

$$|g_{k+1}^Td_k| \leqslant |g_{k+1}^Td_k - g_k^Td_k - t_kd_k^T\nabla^2 f(x_k)d_k| + |g_k^Td_k + t_kd_k^Tf(x_k)d_k| \leqslant \frac{L+M}{2}t_k^2\|d_k\|^3 + \beta t_k\|d_k\|^2. \tag{17}$$

再由假设 A_2，$t_k\|d_k\| \to 0$，$\nabla^2 f(x)$ 在水平集 Ω 上 Lipschitz 连续，可知 $g(x)$ 在 Ω 上是一致连续的，即

$$g_{k+1}^Td_k = g_k^Td_k + t_kd_k^T\nabla^2 f(\zeta)d_k = g_k^Td_k + o(\|d_k\|),\ \zeta \in (x_k, x_k + t_kd_k).$$

注意到 $g_k^Td_k < 0$，有

$$|g_{k+1}^Td_k| = |g_k^Td_k + o(\|d_k\|)| \geqslant |g_k^Td_k| - o(\|d_k\|) = -g_k^Td_k - o(\|d_k\|).$$

再由假设 A_4 和(15)式，有

$$-g_k^Td_k = \|g_k\|\ \|d_k\|\cos\theta_k \geqslant \varepsilon\sin\mu\|d_k\|,$$

故得

$$|g_{k+1}^Td_k| \geqslant \varepsilon\sin\mu\|d_k\| - o(\|d_k\|). \tag{18}$$

结合(17)、(18)式，即有

$$\varepsilon\sin\mu\|d_k\| - o(\|d_k\|) \leqslant \frac{L+M}{2}t_k^2\|d_k\|^3 + \beta t_k\|d_k\|^2,$$

对不等式两边同时除以 $\|d_k\|$，得到

$$\varepsilon\sin\mu \leqslant \frac{L+M}{2}t_k^2\|d_k\|^2 + \beta t_k\|d_k\|. \tag{19}$$

令 $k\to\infty$，则由引理 2，上面不等式右边 $\to 0$，但这与常数 $\varepsilon > 0$，$\mu > 0$ 矛盾. 故定理 1 证明完毕.

参考文献：

[1] Wei Z X, Li G, Qi L. Global convergence of the Polak-Ribière-Polyak conjugate gradient method with an Armijo-type inexact line search for nonconvex unconstrained optimization problems[J]. Mathematics of computation, 2008, 77: 2173-3193.

[2] Yuan G L, Lu X W. A new line search method with trust region for unconstrained optimization[J]. Communications on Applied Nonlinear Analysis, 2008, 15: 35-49.

[3] Nesterov Y, Polyak B T. Cubic regularization of Newton method and its global performance[J]. Math Program, 2006, 108: 177-205.

[4] Cartis C, Gould N I M, Toint P L. Adaptive cubic regularization methods for unconstrained optimization [J]. Math Program, 2010, doi: 10. 1007/s10107-009-0337-y(online).

[5] Nestero Y. Accelerating the cubic regularization of Newton's methods on convex problems [J]. Math Program, 2008, 112: 159-181.

[6] Wei Zeng-xin, Li Lve, Lu Sha. Nonmonotone adaptive cubic overestimation methods for unconstrained optimization[J], 广西大学学报：自然科学版, 2009, 34: 115-119.

[7] 袁亚湘，孙文瑜. 最优化理论与方法[M]. 北京：科学出版社，1997.

（责任编辑：尹　闯）