第15卷第1期 数学研究与评论 Vol. 15 No. 1
1995年2月 JOURNAL OF MATHEMATICAL RESEARCH AND EXPOSITION Feb. 1995

无约束最优化中的一阶曲线法*

漆 涛 游兆永
(大连理工大学数学科学研究所,116024) (西安交通大学,710049)

摘 要 本文对无约束最优化问题提出一阶曲线法,给出其全局收敛结果.对由微分方程定义的一阶曲线,提出渐近收敛指数概念,并计算出几个常用曲线模型的渐近收敛指数.

关键词 无约束最优化,曲线法,全局收敛性.

分类号 AMS(1991) 90C30/ CCL O221.2

§1 引 言

设 $f:R^n\to R$ 为目标函数,R^n 中的无约束优化问题是求 R^n 中一点 x^* 满足 $\min\{f(x):x\in R^n\}=f(x^*)$.现有的一大类算法均可以被称为直线法.而本文将要讨论的是另一类方法,被称为曲线法.令 $p:R^n\times R^n\to R^n$ 满足$(\forall\ x\in R^n)p(x,0)=x$.如果将 x 看成一个固定的参数,则 $p(x,\cdot)$ 为 R^n 中以 x 为初始点的一条曲线.而一般的曲线法可描述为:

算法 1(曲线法):

步 1:选取初始点 x_0,置 $k=0$;

步 2:当 $\nabla f(x_k)\neq 0$ 时做:

2.1:选取 $p(x_k,t)$;

2.2:做曲线搜索 $\min\{f(p(x_k,t)):t\geqslant 0\}=f(p(x_k,t^*))$;

2.3:置 $x_{k+1}=p(x_k,t^*)$;

2.4:置 $k=k+1$;

2.5:转回步 2;

步 3:置 $x^*=x_k$.

算法 1 中的曲线 $p(x_k,t)$ 通常是由微分定义的,或者其导数满足某个不等式.在这里,我们作个不严格的分类.在定义 $p(x,t)$ 的过程中,如果只用到其对 t 的一阶导数,而与其它高阶导数无明显关系,则称曲线 $p(x,t)$ 为一阶曲线.否则就称之为高阶曲线.本文重点讨论一阶曲线.

为了便于讨论,我们假定 $f(x)$ 是二次连续可微的,虽然本文中的大部分结论不需要这么高的要求.但是有了二次连续可微性要求后,定理的陈述及其证明会变得简单清楚些.另外,我们还假定$(\exists\ \alpha>0)$ $L_\alpha(f)$ 为紧.其中 $L_\alpha(f)=\{x\in R^n:f(x)\leqslant\alpha\}$.我们将选取一个紧的 $L_\alpha(f)$,记之为 T.以后讨论的 x 的取值范围将限制在 T 中.即我们讨论下面的无约束化问题:

* 1993年6月28日收到.

(M) $\min\{f(x); x \in T\}$

§2 直线法与曲线法

无约束最优化中广泛使用的方法仍为直线法，但似乎曲线法的思想比拟牛顿算法出现的还要早. Brown and Von Newmann (见[9])1950 年将曲线法应用于对策论中，用分析的方法给出了两人零和对策解的存在性的一个构造性证明. Arrow and Hurwicz 1951 年在[10]中对极大极小问题提出曲线法，其文中称之为梯度法，并在随后的几篇文章中对梯度法进行了进一步的研究. 在这些早期文献中，曲线法的含义没有本文曲线法含义这么广. 本文中的曲线法的定义是由 Botsaris and Jacobson 于 1976 年提出的，见[4]. 在那篇文章中，作者考虑模型

$$\begin{cases} \dot{x}(t) = -\nabla f(x), \\ x(0) = x_k. \end{cases} \tag{1}$$

利用一次近似得到

$$\begin{cases} \dot{x}(t) = -\nabla f(x_k) - H_k(x - x_k), \\ x(0) = x_k, \end{cases} \tag{2}$$

其中 H_k 为 f 在 x_k 处的二阶导数矩阵，而上式的解容易显式给出. 而在文[2]中，*Botsaris* 考虑了更广的一类模型

$$\begin{cases} \dot{x}(t) = -Q^{-1}(x)\nabla f(x) \\ x(0) = x_k \end{cases} \tag{3}$$

与传统的拟牛顿方法比较起来，曲线法似乎有许多缺点. 现在流行一种看法[12],[13]. 认为曲线法作为数值计算的一种方法需要浪费大量的计算来跟踪一个人为的曲线，而事实上所需要的仅是曲线的极限点. 直接应用常微分方程数值解法不会导致有效的算法. 曲线法提供的是一个有效的理论和一个备用的数值技术，作为对标准方法因奇异等原因而失效时的补救措施.

Brown and Bartholomew-Biggs 在[5]中对上面的看法提出异议. 此文认为，从一个初始点出发通过一个光滑曲线的轨迹向解 x^* 逼近这一想法似乎更有吸引力. 对高度非线性优化问题，曲线法比传统方法占有较大的优势. 在传统的直线法中需要沿直线前进一个有限步长. 而当目标函数为非二次函数且有较大的三阶导数时，要找出适当的步长是困难的. 该文还提供了详细的数值试验结果. 这些结果表明，数值积分的方法与步长的选取对算法的效率有巨大影响. 适当地选择这些方法，曲线法的效果是令人满意的. 据此，该文认为曲线法应在数值优化发展的主流中占有一席之地.

另外，在曲线法的具体实现中，可以相对严格地跟踪一条曲线以达到此曲线的极限点. 但是更多的，也是更有吸引力的，是对曲线或者定义此曲线的微分方程本身进行近似. 有各种各样的近似方法，有不同的近似程度，这就给曲线法的研究提供了广阔的研究领域. 用一次近似，如(2)式，则其解为 $x(t) = x_k - (I - e^{-H_k t})H_k^{-1}\nabla f(x_k)$. 它是一个指数曲线. 而最简单的是用零次近似. 那么 $x(t)$ 就是一射线，这就与直线法类似. 这时，直线法与曲线法之间的界限似乎变的模糊了. 但它们之间还是有区别的. 从某种意义上讲，从曲线变为直线是一个过程. 将曲线分成很多小段，每一小段上可以看成为一个直线，即曲线可以看成为无数条直线段所组成的折线.

传统的直线法得到的点列$\{x_k\}k=1,2,\ldots$，$x_k\to x^*$. 顺次连接$x_1,x_2,x_3,\ldots$，得到一折线. 由于$\|x_k-x_{k+1}\|\to 0$，即每一段小折线$[x_k,x_{k+1}]$的长度趋于零. 这样顺次连接$x_1,x_2,x_3,\ldots$，得到的折线可以看作为一曲线. 这条曲线的极限点就是所求的解x^*，但我们却不把它看作是曲线法. 另一面直线法与曲线法的零次近似所循的思想不一样. 拟牛顿法是对目标函数进行近似，将目标函数用二次函数近似代替. 而曲线法的零次近似是对曲线或定义曲线的微分方程进行近似处理. 区别在于两者近似的对象与处理的方法不同.

尽管[5]中的数值试验结果没有得到公认，但[5]中却表明曲线法的潜力是不容忽视的. 这是因为曲线法在理论上还不成熟，在数值试验上还缺乏更有说服力的结果. 本文试图在理论上对曲线法加以完善，对现有的曲线模型进行统一处理，给出他们的全局收敛性分析. 对一类广泛的一阶曲线，给出它们的局部收敛性分析. 对由微分方程定义的一阶曲线，提出收敛指数与渐近收敛指数的概念. 特别是渐近收敛指数是反映曲线收敛速度的一个基本指标. 本文还具体计算出了几个常用曲线模型的渐近收敛指数. 对一阶曲线的一次近似也作了讨论，给出了其全局收敛性结果.

§3 全局收敛性

定义 称$S\stackrel{\triangle}{=}\{x\in T:\nabla f(x)=0\}$为问题$(M)$的解集. $\forall\, x\in S$均被称为(M)的解.

定理 1 假设：

1. $(\forall\, t>0)(\forall\, x\in T)\quad \lim\limits_{y\to x}\quad p(y,t)=p(x,t)$；

2. $(\forall\, x:x\in T-S)(\exists\, t^*>0)\qquad f(p(x,t^*))<f(p(x,0))$.

则由曲线法(算法 1)产生的点列$\{x_k\}$要么为有限个，其最后一个元素就为解；要么为无穷列，其任一聚点为解.

定理 1 中假设条件 2 是基本的. 条件 1 也是合理的实用的. 如果$p(x,t)$由微分方程定义，则由方程解对初值的连续依赖性很容易得出条件 1.

定理 2 假设：

1. $(\forall\, x\in T-S)\qquad f'(x)\frac{d}{dt}p(x,0)<0$，$f'(x)\frac{d}{dt}p(x,0)$在$x$处连续；

2. $(\forall\, x\in T-S)(\exists\, \varepsilon>0,m>0)(\forall\, x':\|x-x'\|<\varepsilon)(\forall\, t>0)$

$$\left\|\frac{d}{dt}p(x',t)\right\|\leqslant m,\ \left\|\frac{d^2}{dt^2}p(x',t)\right\|\leqslant m.$$

则由算法 1 产生的序列$\{x_k\}$要么是有限个，其最后一个元素就是解. 要么为无穷个，其任一聚点为解.

定理 2 中假设 1 是说在$t=0$处曲线$p(x,\cdot)$的切向量应是f在x处的下降方向. 利用T的紧性，假设 2 相当于$(\exists\, M>0)(\forall\, x\in T)(\forall\, t>0)\ \left\|\frac{d}{dt}p(x,t)\right\|\leqslant M$，$\left\|\frac{d^2}{dt^2}p(x,t)\right\|\leqslant M$.

§4 一阶曲线法及渐近收敛指数

1. 下降曲线的基本性质 　考虑满足下式的曲线$x(t)$

$$\begin{cases} f'(x(t))\dot{x}(t)\leqslant 0 \\ x(0)=x_0\in \mathrm{int}(T) \end{cases}$$

由于 $f(x(t))$ 对 t 是单调下降的，称这样的 $x(t)$ 为关于 $f(x)$ 的下降曲线.

定义 如果连续函数 $\alpha:[0,+\infty)\to[0,+\infty)$ 满足 $\alpha(0)=0$，而且 $(\forall\ s>0)\alpha(s)>0$；$(\forall\ t>0)$ $f'(x(t))\dot{x}(t)\leqslant-\alpha(\|f'(x(t))\|)$. 则称 α 为 f 关于 $x(t)$ 的下降控制函数.

定理 3 假设对于 f 与 $x(t)$ 存在下降控制函数 α. 而 $\|\dot{x}(t)\|$，$\|\ddot{x}(t)\|$ 为有界的. 则有：

$$\lim_{t\to+\infty} f(x(t))\text{存在；}\quad \lim_{t\to+\infty}\nabla f(x(t))=0;\quad x(t)\text{的极限集为非空连通紧集，}\tag{4}$$

其中 $x(t)$ 的极限集的定义为：$\Lambda\overset{\triangle}{=}\Lambda(x(t))\overset{\triangle}{=}\{y\in R^n:\exists\ \{t_k\},t_k\to+\infty \text{使 } y=\lim_{k\to\infty}x(t_k)\}$.

由定理 3 知，$\Lambda\subset S$. 一般而言，不能判定 Λ 就是单点集. 即不能判定 $\lim_{t\to\infty}x(t)$ 的存在性. 下面给出 Λ 为单点集的一个充分条件.

定理 4 如果 S 的势小于实数集的势，则 Λ 为单点集.

2. **由微分方程定义的曲线** 设 $\beta:R^n\times R^n\to R^n$ 为连续可微，且 $(\forall\ x\in R^n)\beta(x,0)=0$. 而且 $(\forall\ x,y\in R^n)(\beta(x,y),y)\geqslant\alpha(\|y\|)$. 其中 $\alpha:[0,+\infty)\to[0,+\infty)$ 为一下降控制函数. 考虑由下式定义的曲线 $x(t)$

$$\begin{cases} \dot{x}(t)=-\beta(x,\nabla f(x)), \\ x(0)=x_0\in\mathrm{int}(T). \end{cases}\tag{5}$$

定理 5 由上式定义的曲线 $x(t)$ 仍满足性质(4).

令 $G(x)=\dfrac{d}{dx}\beta(x,\nabla f(x))=\dfrac{\partial\beta}{\partial x}(x,\nabla f(x))+\dfrac{\partial\beta}{\partial y}(x,\nabla f(x))f''(x)$. 则 $G(x)$ 为 R^n 到 R^n 的一个线性算子. 其在基 $\{e_1,\ldots,e_n\}$ 对应的矩阵仍记为 $G(x)$.

定理 6 假定：

1. $G(x)$ 在 $\mathrm{conv}(T)$ 上是一致正定的，即

$$m=\inf\{z^TG(x)z:x\in R^n,z\in\mathrm{conv}(T),\|z\|=1\}>0;$$

2. $x(t)$ 收敛于 x^*，其中 $x(t)$ 满足(5).

则有：

1. $\|x(t)-x^*\|\leqslant e^{-mt}\|x_0-x^*\|$；
2. $|f(x(t)-f(x^*)|<e^{-2m}M\|x_o-x^*\|^2$;
3. $\|\nabla f(x(t))\|\leqslant Me^{-mt}\|x_0-x^*\|$，

其中 $M=\max\{\|f''(x)\|:x\in\mathrm{conv}(T)\}$.

定理中的 m 称为 $x(t)$ 的收敛指数，它是衡量 $x(t)$ 收敛速度大小的一个度量. 但是 $m>0$ 是一个严格的要求，而当 $x(t)$ 趋于 x^* 时，$x(t)$ 的收敛速度应只于 x^* 邻域内的 f 与 β 的性态有关. 下面的定理 7 将改进定理 6 的结果.

定理 7 假设 $x(t)\to x^*(t\to+\infty)$ 且 $m^*=\inf\{y^TG(x^*)y:y\in R^n,\|y\|=1\}>0$.

则 $(\forall\ \varepsilon:0<\varepsilon<m^*)(\exists\ T>0)(\forall\ t\geqslant T)$ 有

1. $\|x(t)-x^*\|\leqslant e^{-(m^*-\varepsilon)(t-T)}\|x(T)-x^*\|$；
2. $|f(x(t))-f(x^*)|\leqslant Me^{-2(m^*-\varepsilon)(t-T)}\|x(T)-x^*\|^2$
3. $\|\nabla f(x(t))\|\leqslant Me^{-(m^*-\varepsilon)(t-T)}\|x(T)-x^*\|$.

定理 7 中的 m^* 被称为 $x(t)$ 的渐近收敛指数. 它仅于 x^* 处 $\beta(x,\nabla f(x))$，及 $\nabla f(x)$ 的性态有

关.渐近收敛指数 m^* 可以用来作为判别模型优劣的一个标准.常见的几个模型(见[2],[3],[4],[5])分别为:

1. $\beta(x,y)=y$ (最降曲线);2. $\beta(x,y)=D(x)y$ ($D(x)$为正定矩阵);

3. $\beta(x,y)=f''(x)y$; 4. $\beta(x,y)=f''(x)^{-1}y$.

它们对应的微分方程分别是:

1. $\dot{x}=-\nabla f(x)$ 2. $\dot{x}(t)=-D(x)\nabla f(x)$

3. $\dot{x}=-f''(x)\nabla f(x)$ 4. $\dot{x}=-f''(x)^{-1}\nabla f(x)$

这几个模型的 $G(x^*)$分别是(注意 $\nabla f(x^*)=0$):

1. $G(x^*)=f''(x^*)$; 2. $G(x^*)=D(x^*)f''(x)$; 3. $G(x^*)=f''(x^*)^2$; 4. $G(x^*)=I$.

它们的渐近收敛指数分别是:

1. $m^*=f''(x^*)$的最小特征值.

2. $m^*=[D(x^*)f''(x^*)+f''(x^*)D(x^*)]/2$ 的最小特征值.

3. $m^*=f''(x^*)^2$的最小特征值.

4. $m^*=1$.

§5 一次近似及其全局收敛性

对于方程

$$\begin{cases}\dot{x}(t)=-\beta(x,\nabla f(x))\\ x(0)=x_k\end{cases} \tag{6}$$

将 $\beta(x,\nabla f(x))$在 x_k 处展开,取其一次近似,有

$$\begin{cases}\dot{x}(t)=-\beta_k-G_k(x-x_k)\\ x(0)=x_k\end{cases} \tag{7}$$

其中 $\beta_k=\beta(x_k,\nabla f(x_k))$, $G_k=G(x_k)=\frac{d}{dx}\beta(x,\nabla f(x))|_{x=x_k}$. 称(7)为(6)的一次近似方程.(7)的解为: $x(t)=x_k-(I-e^{-G_kt})G_k^{-1}\beta_k$.这样得到如下算法:

算法 2(一次近似曲线法):

步 1:取 $x_0\in \mathrm{int}(T)$,置 $k=0$;

步 2:如果 $\nabla f(x_k)\neq 0$ 做:

2.1:作一维优化: $\min\{f(x-(I-e^{-G_kt})G_k^{-1}\beta_k): t\geqslant 0\}$,其极小点记为 t^*;

2.2:置 $x_{k+1}=x_k-(I-e^{-G_kt^*})G_k^{-1}\beta_k$;

2.3:置 $k=k+1$;

2.4:转回步 2;

步 3:置 $x^*=x_k$.

关于一次近似的全局收敛性有下面的结果.

定理 8 假设存在一个下降控制函数 α 满足$(\forall\, x,y\in R^n)(\beta(x,y),y)\geqslant\alpha(\|y\|)$.且 $G(x)=\frac{d}{dx}\beta(x,\nabla f(x))$在 R^n 上连续.则由算法 2 产生的点列$\{x_k\}$,要么为有限个,其最后一个元素就是解;要么为无穷个,其每一个聚点均为解.

参 考 文 献

[1] E. Polak, *Computational Methods in Optimization: a Unified Approach*, Academic Press, 1971.

[2] C. A. Botsaris, *Differential gradient methods*, *JMAA*, 1978, 63(1): 177 — 188.

[3] ——, *A Curvilinear optimization based upon iterative estimation of the eigensystem of the Hessian matrix*, JMAA, 1978, 63(2): 367 — 411.

[4] C. A. Botsaris & D. H. Jacobson, *A Newton-type Curvilinear search methods of optimization*, JMAA, 1976, 54(1): 217 — 229.

[5] A. A. Brown and M. C. Bartholomew-Biggs, *Some effective methods for unconstrained optmization based on the solution of systems of ordinary differential equations*, JOTA, 1989, 62(2): 211 — 224.

[6] 彼得罗夫斯基著，黄克欧译，常微分方程讲义，人民教育出版社，1959 年.

[7] 漆涛，无约束最优化中的曲线方法，西安交通大学博士论文，1992 年.

[8] F. Zirilli, *The use of O. D. E. for the solution of nonlinear systems of equations, in "Nonlinear optimization 1981" ed. by M. J. D. Powell*, Academic Press, 1982, 39 — 46.

[9] G. W. Brown & J. Von Newmann, *Solutions of games by differential equations*, in "Contributions to the theory of games" edited by H. W. Kuhn and A. W. Tuckes, Princeton University Press, 1950.

[10] K. J. Arrow, L. Hurwicz and H. Uzawa, *Studies in Linear and Nonlinear Programming*, Stanford University Press, 1958.

[11] C. A. Botsaris, *A class of methods for unconstrained minimization based on stable numerical integration technique*, JMAA, 1978, 63(3): 722 — 749.

[12] M. J. D. Powell, (editor) *Nonlinear optimization* 1981, Academic Press, 1982.

[13] M. J. D. Powell, *A view of unconstrained optimization*, in "Optimization in action" ed. by L. C. W. Dixon, Academic Press, 1976.

The First Order Curvilinear Methods in Unconstrained Optimization

Qi Tao (Dalian University of Technology)

You Zhaoyong (Xi'an Jiaotong University)

Abstract

The first order curvilinear methods is presented and some relevant global convergence results are given. The asymptotic convergence index for the first order curves defined by some ordinary differential equations is given, which may be used as a criterion for the convergence speed of the first order curvilinear methods.

Keywords unconstrained optimization, curvilinear methods, global convergence.