

某些极小化问题的一个直接算法*

费浦生

(武汉大学)

A DIRECT ALGORITHM FOR SOME MINIMIZATION PROBLEMS

Fei Pu-sheng

(Wuhan University)

Abstract

In this paper an algorithm for minimizing a nonlinear function is obtained by using approximate gradient from the simplex method. It is economical for problems with less variables and a complicated objective function.

一、引言

在一些物理过程或化学反应过程中,随时间变化的物理量 $y = (y_1, \dots, y_m)^T$ 满足含有与时间无关的参数 $x = (x_1, \dots, x_n)^T$ 的常微分方程初值问题

$$\frac{dy}{dt} = f(t, y, x), \quad (1)$$

$$y(a) = \varphi(x). \quad (2)$$

当 x_1, \dots, x_n 确定后,从(1)、(2)可以用数值方法(例如 Runge-Kutta 方法等^[1])近似解出 y . 但是,实际上往往会出现如下情况: 这些参数是未知的,不能用实验方法直接测出,然而可以用实验测出某些 y_p 在一些时刻 t_q 上的值. 例如,已知 $y_1, \dots, y_l (1 \leq l \leq m)$ 在时刻 $t_1, \dots, t_k (a \leq t_1 < \dots < t_k \leq b)$ 处的测量值是 $y_{pq} (p = 1, \dots, l; q = 1, \dots, k)$. $l = 1$ 是常见的情况,特别是(1)由单个高阶方程化成的一阶方程组的情况. 因此我们提出一个反问题: 如何选择参数 $x_i (i = 1, \dots, n)$, 用(1)、(2)的解曲线对测量值进行拟合. 这是一种利用实验测量结果间接确定参数的方法. 如果采用最小二乘误差进行拟合,就得到极小化问题. 求 $x \in \mathbb{R}^n$, 让(1)、(2)的解 y 使目标函数

$$I(x) = \sum_{p=1}^l \sum_{q=1}^k [y_p(t_q) - y_{pq}]^2 \quad (3)$$

* 1984年2月23日收到.

达到极小.

在实际问题中, 一般来说 n 是一个较小的正整数, 主要困难是方程 (1)、(2) 的解不能写成关于 t , x 的显式表达式, 更不知道 $I(x)$ 关于 x_i 的偏导数的表达式. 因此, 用带导数的极小化算法是困难的. 在采用不用导数的直接算法时, 也要尽量避免计算过多的点 x 上的 $I(x)$ 值. 因为对每一点 x , 要算出 y 在 t_1, \dots, t_k 上的值就要数值求解 (1)、(2), 而且要有足够的精确度.

为了避免使用导数, 在求 (3) 的极小时可以选择单纯形算法^[2]. 但单纯形算法在求重心、反映点、扩展点或收缩点以及整个单纯形尺度缩小时用到的新点都要计算 $I(x)$ 的值, 因此就是过多地求解 (1)、(2). 我们宁可增加其他一些辅助计算工作来换取少用新点. 我们结合利用近似梯度, 对单纯形法进行了修改, 提出另一个算法.

二、算 法

我们考虑 \mathbf{R}^n 中的极小化问题

$$\min I(x)$$

并假设变量个数 n 不很大, 但计算目标函数 $I(x)$ 的值相对来说工作量是很大的.

设已经有了 \mathbf{R}^n 中的 $n+1$ 个点 $x^{(0)}, x^{(1)}, \dots, x^{(n)}$, 且 $x^{(i)} - x^{(0)} (i = 1, \dots, n)$ 线性无关, 并由此 $n+1$ 个点为顶点构成一个单纯形. 设

$$I(x^{(0)}) \leq I(x^{(1)}) \leq \dots \leq I(x^{(n)}).$$

假如 $I(x)$ 关于 $x_j (j = 1, \dots, n)$ 的偏导数都存在且连续, 近似地有

$$I(x^{(i)}) - I(x^{(0)}) = \sum_{j=1}^n (x_j^{(i)} - x_j^{(0)}) \frac{\partial I(x^{(0)})}{\partial x_j} \quad (i = 1, \dots, n). \quad (4)$$

从 (4) 中可求出负梯度

$$-\text{grad } I(x^{(0)}) = \left(-\frac{\partial I(x^{(0)})}{\partial x_1}, \dots, -\frac{\partial I(x^{(0)})}{\partial x_n} \right)$$

的近似值 $u = (u_1, \dots, u_n)^T$, 并从 $x^{(0)}$ 出发沿 u 方向寻找新点. u 满足线性代数方程组

$$Au = d, \quad (5)$$

其中

$$A = \begin{pmatrix} x_1^{(1)} - x_1^{(0)} & \dots & x_n^{(1)} - x_n^{(0)} \\ \dots & \dots & \dots \\ x_1^{(n)} - x_1^{(0)} & \dots & x_n^{(n)} - x_n^{(0)} \end{pmatrix}, \quad (6)$$

$$d = (d_1, \dots, d_n)^T,$$

$$d_i = I(x^{(0)}) - I(x^{(i)}) \quad (i = 1, \dots, n). \quad (7)$$

由于 $x^{(i)} - x^{(0)} (i = 1, \dots, n)$ 的无关性, A 是满秩的.

为了不增加过多的新点, 我们避免在 u 方向采取一维寻查求极小的方法, 而只是取一个点来比较. 例如取

$$x^* = x^{(0)} + \frac{\alpha}{|u|^2} [I(x^{(n)}) - I(x^{(0)})]u, \quad (8)$$

其中 α 是选定的常数, 一般取 $\frac{1}{2} \leq \alpha \leq 2$.

若 $I(x^*) < I(x^{(n)})$, 我们就用 x^* 替换 $x^{(n)}$, 否则用 $x^{(0)} + \beta(x^{(n)} - x^{(0)})$ 替换 $x^{(n)}$, 其中 $0 < \beta < 1$ 是选定的收缩系数. 得到新的 $n+1$ 个点之后再重复上述过程.

我们需要防止在若干次迭代后出现“降维”, 即出现 $x^{(i)} - x^{(0)} (i = 1, \dots, n)$ 近似线性相关的情况, 即 A 的行列式近似为零, 这一点可以在求解方程组(5)时加以检验.

设上一次的单纯形是 n 维的, 则经过用 x^* 代替 $x^{(n)}$ 后至少也是 $n-1$ 维的, 即 A 的秩 $r(A) \geq n-1$. 在求解方程组(5)时我们选用满矩阵(不包括右端项)选主元的消去法. 由于 $r(A) \geq n-1$, 故前 $n-1$ 个主元都不会为零. 设第一个主元的绝对值为 h_1 , 第 n 个主元的绝对值为 h_n , 且最后这个主元出现在 A 的第 s 行. 如果

$$\frac{h_n}{h_1} < \varepsilon_1,$$

则认为出现了降维, 这里 ε_1 是预先给定的很小的正数. 如果出现了降维, 这时 $x^{(i)} - x^{(0)} (i = 1, \dots, n; i \neq s)$ 是线性无关的. 下面我们求(5)中去掉第 s 个方程后所得的方程组相应的齐次方程组的一个非零解 v , 则 v 与诸 $x^{(i)} - x^{(0)} (i \neq s)$ 均正交, 因而用

$$x^{(0)} + \frac{\zeta h_1}{\max_{1 \leq i \leq n} |v_i|} v$$

代替 $x^{(s)}$ 后得到的 $n+1$ 个点为顶点的单纯形, 此单纯形是 n 维的, 可以保证迭代继续进行, 其中 ζ 是选定的常数, 常取 $0 < \zeta < 1$.

过程的迭代终止条件, 可参照单纯形法来选用. 例如用

$$I(x^{(n)}) - I(x^{(0)}) < \varepsilon$$

或

$$\frac{1}{n+1} \sum_{i=1}^n [I(x^{(i)}) - I(x^{(0)})]^2 + \frac{1}{n+1} [I(x^*) - I(x^{(0)})]^2 < \varepsilon$$

等等. $\varepsilon > 0$ 为选定的误差判断常数.

具体计算步骤如下:

1) 选取初始点 $x^{(0)}$, 令

$$x^{(i)} = x^{(0)} + \sigma_i e^{(i)} \quad (i = 1, \dots, n),$$

其中 $\sigma_1, \dots, \sigma_n$ 是给定的非零数. $e^{(i)}$ 为第 i 坐标方向的单位向量.

计算 $I(x^{(i)}) (i = 0, 1, \dots, n)$.

2) 将 $(x^{(0)}, I(x^{(0)})), (x^{(1)}, I(x^{(1)})), \dots, (x^{(n)}, I(x^{(n)}))$ 按 $I(x^{(i)})$ 的值从小到大重新排队.

3) 计算 A, d . 对方程组(5)用主元消去法消元. 若 $\frac{h_n}{h_1} < \varepsilon_1$ 转 10); 否则转 4).

4) 解出 u ,

$$x^* := x^{(0)} + \frac{\alpha}{|u|^2} [I(x^{(n)}) - I(x^{(0)})]u.$$

- 5) 计算 $I(x^*)$. 若 $I(x^*) < I(x^{(0)})$ 转 8); 否则转 6).
- 6) 若 $I(x^*) \geq I(x^{(n)})$ 转 9); 否则转 7).
- 7) 若 $I(x^{(n)}) - I(x^{(0)}) < \varepsilon$ 则算法终止; 否则转 8).
- 8) $(x^{(n)}, I(x^{(n)})) := (x^*, I(x^*))$, 转 2).
- 9) $x^{(n)} := x^{(0)} + \beta(x^{(n)} - x^{(0)})$, 计算 $I(x^{(n)})$, 转 2)
- 10) 解出 v ,

$$x^{(s)} := x^{(0)} + \frac{\zeta h_1}{\max_{1 \leq i \leq n} |v_i|} v,$$

计算 $I(x^{(s)})$, 转 2).

这个过程由步骤 7) 产生终止后得到的 $x^{(0)}$ 就作为近似最优解.

三、说 明

1. 我们记每次迭代的单纯形中各顶点目标函数值之和为 $\tilde{I}(x^{(0)}, x^{(1)}, \dots, x^{(n)})$.

由算法可知, 当经过步骤 8) 用 x^* 代替 $x^{(n)}$ 时, 必有 $I(x^*) < I(x^{(n)})$, 新单纯形的 $\tilde{I}(x^{(0)}, x^{(1)}, \dots, x^{(n)})$ 将减小. 若经过步骤 9) 用 $x^{(0)} + \beta(x^{(n)} - x^{(0)})$ 代替 $x^{(n)}$, 如果 $I(x)$ 是严格拟凸的^[4], 则 $I(x^{(0)} + \beta(x^{(n)} - x^{(0)})) < I(x^{(n)})$, $\tilde{I}(x^{(0)}, x^{(1)}, \dots, x^{(n)})$ 也将减小. 因此, 本算法关于函数 $\tilde{I}(x^{(0)}, x^{(1)}, \dots, x^{(n)})$ 是一个下降算法.

2. 本算法曾用于计算沸石分子筛晶体生长动力学模型. 该模型是形如 (1)、(2) 的常微分方程初值问题. 微分方程个数 $m = 6$, 待定参数个数为 $n = 4$. 针对不同沸石分子筛在不同反应条件下的晶体生长, 共计算 60 多个例子, 都取得了满意的效果^[5,6]. 采用单纯形方法对同样这些问题进行了计算, 在相同的初始单纯形及相同的误差要求之下, 得到的结果是很接近的, 但所需计算目标函数值的次数, 单纯形方法为本方法的 4 倍至 11 倍.

3. 在所算各例中, 大部分始终未出现降维情况. 在概率很小的几次降维出现后, 经过处理后均能继续进行, 直到满足要求而终止.

4. 如果目标函数 $I(x)$ 依各变量 x_1, \dots, x_n 的变化速度有较大的数量级差别, 则无论用单纯形法还是本算法都不太理想, 出现降维的可能性增大. 最好对这种问题先作一个自变量变换

$$\tilde{x}_i = \theta_i x_i \quad (i = 1, \dots, n),$$

其中选择正数 θ_i 使得目标函数关于新变量 $\tilde{x}_1, \dots, \tilde{x}_n$ 的变化速度稍均匀一些, 然后再使用本算法求极小.

5. 在实际问题中, 有时物理过程不是由常微分方程 (1) 来描述, 而是满足某个微分代数方程组^[3]或者其他形式的数学物理问题. 只要这些问题在参数固定时有可靠的数值解法, 并且满足适当的条件以保证根据给出的实验值来确定这些参数是合理的, 本算法原则上可以使用.

6. 这个算法仅仅对于变量个数不多、目标函数计算复杂而又不便于用导数的极小化

问题是有效的。对于一般极小化问题,本算法由于解线代数方程组增加了计算量,因而就不合算了。

参 考 文 献

- [1] 武汉大学,山东大学编,计算方法,人民教育出版社,1979.
- [2] 席少霖,赵凤治,最优化计算方法,上海科学技术出版社,1983.
- [3] C. W. Gear, H. H. Hsu, L. Petzold, *Differential-Algebraic Equations Revisited*, Proc. Numerical Methods for Solving Stiff Initial Value Problems, Oberwolfach, West Germany, 1981.
- [4] R. T. Rockafellar, *Convex Analysis*.
- [5] 徐如人,李守贵,韩淑芸,曹惠,费浦生,高等学校化学学报,6: 9(1985), 765—770.
- [6] 冯守华,李守贵,徐如人,费浦生,高等学校化学学报,6: 10(1985), 855—860.