文章编号：0253-374X(2014)05-0804-03 DOI:10.3969/j.issn.0253-374x.2014.05.023

一类多项式全局优化的差分算法

朱经浩，何似菡
(同济大学 数学系，上海 200092)

摘要：引入一类 n 元多项式的倒向微分流以求解全局优化问题. 沿着倒向微分流，建立一个差分-牛顿混合算法，并证明了由算法所得迭代点的绝对误差受到差分步长的一致界囿. 应用所建立的算法，给出了一个数值计算的例子.

关键词：多项式全局优化；倒向微分方程；差分-牛顿混合算法
中图分类号：O224 **文献标志码**：A

A Difference Algorithm to Find Global Minimizers of a Polynomial

ZHU Jinghao, *HE Sihan*
(Department of Mathematics, Tongji University, Shanghai 200092, China)

Abstract: This paper presents a differential flow concerning the global optimization with polynomials. Along the flow which leads to a global minimizer there is posed a mixed difference algorithm with Newton method. An error analysis of the algorithm is given with a numerical example to demonstrate the mixed difference method.

Key words: global optimization with polynomial; backward differential flow; mixed difference algorithm with Newton method

考虑如下多项式全局优化问题(P)：

$$\min P(\boldsymbol{x})$$
$$\text{s.t. } \boldsymbol{x} \in \mathbf{R}^n$$

其中对于 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}\in\mathbf{R}^n$

$$P(\boldsymbol{x}) = \sum_{i=1}^{n} x_i^{2m} + \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{f}^{\mathrm{T}}\boldsymbol{x} \qquad (1)$$

这里 $m\geqslant 2$，$\boldsymbol{A}=(a_{i,j})_{i,j=1,2,\cdots,n}$ 是半正定矩阵，而 $\boldsymbol{f}=(f_1,\cdots,f_n)^{\mathrm{T}}$ 是一个 n 维非零向量. 多项式 $P(\boldsymbol{x})$ 的梯度为

$$\nabla P(\boldsymbol{x}) = (2mx_1^{2m-1}, 2mx_2^{2m-1}, \cdots, 2mx_n^{2m-1})^{\mathrm{T}} + \boldsymbol{A}\boldsymbol{x} - \boldsymbol{f}$$

且其 Hesse 矩阵为

$$\nabla^2 P(\boldsymbol{x}) = \mathrm{diag}(2m(2m-1)x_1^{2m-2}, 2m(2m-1)x_2^{2m-2}, \cdots, 2m(2m-1)x_n^{2m-2}) + \boldsymbol{A}$$

此矩阵中除了对角线上为非负项外其他处均为 0. 可见，对于 $\forall\, \boldsymbol{x}\in\mathbf{R}^n$，有 $\nabla^2 P(\boldsymbol{x})\geqslant\mathbf{0}$.

这类多项式具有一定的实际意义，在文献[1-2]阐述了下述问题：

$$\min_{\boldsymbol{x}\in\mathbf{R}^n}\left\{P(\boldsymbol{x}) = W(\boldsymbol{x}) + \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{f}^{\mathrm{T}}\boldsymbol{x}\right\} \qquad (2)$$

其中

$$W(\boldsymbol{x}) = \sum_{k=1}^{p} \frac{1}{2}\alpha_k\left(\frac{1}{2}\|\boldsymbol{B}_k\boldsymbol{x}\|^2 - \mu_k\right)^2$$

这个问题经常出现在许多复杂系统中，涵盖相变的非凸分析、网络设计及通信的离散优化等方面. 如果 $A\geqslant 0$，令式(2)中 $\mu_k=0$，$\frac{1}{2}\alpha_k=4$，$p=n$，$\boldsymbol{B}_k=\boldsymbol{E}_{kk}$（$\boldsymbol{E}_{kk}$ 为只有在第 k 行与第 k 列交叉位置的元素为 1，其余元素都为 0 的 n 阶方阵)，于是得到 $W(\boldsymbol{x}) = \sum_{k=1}^{n} x_k^4$，这样问题(2)所涉及的目标函数即为本文所关注的一类多项式.

1 全局最优化

1.1 一个球约束优化问题

文献[3-4]证明了多项式全局优化问题(P)的所有最优点都含于一个以原点为中心的闭球内部，以下记此闭球为

$$D := \{\boldsymbol{x} \mid \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} \leqslant d^2\}$$

这样无约束多项式全局优化问题(P)就等价地转化

收稿日期：2013-06-24
基金项目：国家自然科学基金(10671145)
第一作者：朱经浩(1949—)，男，教授，博士生导师，主要研究方向为最优控制理论与应用. E-mail: jinghaok@online.sh.cn
通讯作者：何似菡(1990—)，女，硕士生，主要研究方向为最优控制理论与应用. E-mail: hesihan0804@foxmail.cn

为以下球约束优化问题：

$$\min P(\boldsymbol{x}) = \sum_{i=1}^{n} x_i^{2m} + \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{f}^{\mathrm{T}}\boldsymbol{x},$$
$$\text{s. t. } \|\boldsymbol{x}\| \leqslant d. \tag{3}$$

显然这里的关键问题是如何得到闭球的半径 d，按照文献[4]给出的计算公式可得到

$$d = \max\left\{1, \frac{(2m)!\left(n + \frac{n^2}{2!} + \cdots + \frac{n^{2m-1}}{(2m-1)!}\right)M}{J_{2m}^{*}}\right\} \tag{4}$$

其中

$$M = \max\{|a_{ij}|, |f_i|\}_{1\leqslant i\leqslant n, 1\leqslant j\leqslant n} \tag{5}$$

而 J_{2m}^{*} 是最优化问题 $\min J_{2m}(s) = (2m)!\sum_{i=1}^{n} s_i^{2m}$，s. t. $\|\boldsymbol{s}\|^2 = 1$ 的最优值.

1.2　全局最优点

本文将引用文献[5]中的结果，求解球约束下的全局优化问题(3).

由于在 $\mathbf{R}^n$ 上，$\nabla^2 P(\boldsymbol{x}) \geqslant 0$，从而对任意 $\rho \in (0, +\infty)$，有 $\nabla^2 P(\boldsymbol{x}) + \rho\boldsymbol{I} > 0$. 注意到 $\nabla P(\boldsymbol{0}) = -\boldsymbol{f} \neq \boldsymbol{0}$，由文献[6]中不动点定理可知：对于适当大的 $\rho^* > 0$，存在一个非零向量 $\boldsymbol{x}^* \in D = \{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} \leqslant d^2\}$ 满足以下方程：

$$\nabla P(\boldsymbol{x}^*) + \rho^* \boldsymbol{x}^* = 0$$

由于对 $\boldsymbol{x} \in \mathbf{R}^n$，$\rho > 0$，恒有 $\nabla^2 P(\boldsymbol{x}) + \rho\boldsymbol{I} > 0$，根据经典常微分方程理论，由以下微分方程：

$$\frac{d\hat{\boldsymbol{x}}}{d\rho} + [\nabla^2 P(\hat{\boldsymbol{x}}) + \rho\boldsymbol{I}]^{-1}\hat{\boldsymbol{x}} = 0 \tag{6}$$

$$\hat{\boldsymbol{x}}(\rho^*) = \boldsymbol{x}^* \tag{7}$$

可以惟一得到定义在 $(0, +\infty)$ 内的微分流 $\hat{\boldsymbol{x}}(\rho)$，满足

$$\nabla P(\hat{\boldsymbol{x}}(\rho)) + \rho\hat{\boldsymbol{x}}(\rho) = 0$$

而且，对给定的 $\rho > 0$，$\hat{\boldsymbol{x}}(\rho)$ 是惟一满足方程 $\nabla P(\boldsymbol{x}) + \rho\boldsymbol{x} = 0$ 的点，当然也是 $L(\boldsymbol{x}, \rho) = P(\boldsymbol{x}) + \frac{\rho}{2}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} - d^2)$ 在 $\mathbf{R}^n$ 上的惟一全局最小点. 关于微分流 $\hat{\boldsymbol{x}}(\rho)$，定义如下对偶函数[5]：对于 $\rho > 0$，$P_d(\rho) = P(\hat{\boldsymbol{x}}(\rho)) + \frac{\rho}{2}\hat{\boldsymbol{x}}^{\mathrm{T}}(\rho)\hat{\boldsymbol{x}}(\rho) - \frac{\rho d^2}{2}$.

引理 1[5]

$$\frac{\mathrm{d}P_d(\rho)}{\mathrm{d}\rho} = \frac{1}{2}\hat{\boldsymbol{x}}^{\mathrm{T}}(\rho)\hat{\boldsymbol{x}}(\rho) - \frac{d^2}{2}$$

$$\frac{\mathrm{d}^2 P_d(\rho)}{\mathrm{d}\rho^2} = -\left(\frac{\mathrm{d}\hat{\boldsymbol{x}}(\rho)}{\mathrm{d}\rho}\right)^{\mathrm{T}}[\nabla^2 P(\hat{\boldsymbol{x}}(\rho)) + \rho\boldsymbol{I}]\frac{\mathrm{d}\hat{\boldsymbol{x}}(\rho)}{\mathrm{d}\rho}$$

引理 2[5]　$\frac{\mathrm{d}P_d(\rho)}{\mathrm{d}\rho}$ 关于 $\rho > 0$ 单调递减，而 $P_d(\rho)$ 是关于 $\rho > \rho^*$ 单调递减的.

定理 1[5]　设 $\hat{\boldsymbol{x}}(\rho)$，$\rho \in (0, +\infty)$ 为式(6)，(7)定义的微分流，记

$$\hat{\boldsymbol{x}}(0^+) = \lim_{\rho \to 0^+}\hat{\boldsymbol{x}}(\rho)$$

若 $\hat{\boldsymbol{x}}^{\mathrm{T}}(0^+)\hat{\boldsymbol{x}}(0^+) \leqslant d^2$，则 $\hat{\boldsymbol{x}}(0^+)$ 是 $P(x)$ 在球 D 上的全局最小点.

以下说明应用定理 1 可求解球约束下的全局优化问题(3). 由引理 1 和引理 2 可知，$\|\hat{\boldsymbol{x}}(\rho)\|$ 关于 $\rho > 0$ 单调递减. 若存在正实数 $\hat{\rho}$ 使得 $\|\hat{\boldsymbol{x}}(\hat{\rho})\| = d$，则对任意 $\boldsymbol{x} \in D$ 有

$$\begin{aligned} P(\boldsymbol{x}) &\geqslant P(\boldsymbol{x}) + \frac{\hat{\rho}}{2}[\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} - d^2] \geqslant \\ &\inf_{D}\left\{P(\boldsymbol{x}) + \frac{\hat{\rho}}{2}[\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} - d^2]\right\} = \\ &P(\hat{\boldsymbol{x}}(\hat{\rho})) + \frac{\hat{\rho}}{2}(\hat{\boldsymbol{x}}^{\mathrm{T}}(\hat{\rho})\hat{\boldsymbol{x}}(\hat{\rho}) - d^2) = \\ &P(\hat{\boldsymbol{x}}(\hat{\rho})) \end{aligned}$$

于是 $\hat{\boldsymbol{x}}(\hat{\rho})$ 是 $P(\boldsymbol{x})$ 在球 D 上的全局最小点，但是 $\|\hat{\boldsymbol{x}}(\hat{\rho})\| = d$，这与全局优化问题($P$)的所有最优点都含于球 D 的内部这一事实相矛盾，所以 $[\hat{\boldsymbol{x}}(0^+)]^{\mathrm{T}} \cdot \hat{\boldsymbol{x}}(0^+) < d^2$. 再由定理 1 推知，$\hat{\boldsymbol{x}}(0^+)$ 是 $P(\boldsymbol{x})$ 在球 D 上的全局最小点.

注 1：由于当 $\rho > 0$，$\nabla^2 P(\hat{\boldsymbol{x}}(\rho)) + \rho\boldsymbol{I} > 0$，且 $\|\hat{\boldsymbol{x}}(\rho)\|$ 关于 $\rho > 0$ 单调递减，以及 $\hat{\boldsymbol{x}}^{\mathrm{T}}(0^+)\hat{\boldsymbol{x}}(0^+) < d^2$，考察倒向微分方程(6)可知，存在正数 σ 使得在 $[0, \rho^*]$ 上一致有 $\nabla^2 P(\hat{\boldsymbol{x}}(\rho)) + \hat{\rho}\boldsymbol{I} > \sigma\boldsymbol{I}$.

2　一个差分-牛顿混合算法

为了在数值计算上实现以上求解球约束优化问题(3)的数学过程，本节建立一个差分-牛顿混合算法. 算法的基本依据是，对给定的 $\rho > 0$，微分流上的点 $\hat{\boldsymbol{x}}(\rho)$ 是 $L(\boldsymbol{x}, \rho) = P(\boldsymbol{x}) + \frac{\rho}{2}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} - d^2)$ 在 $\mathbf{R}^n$ 上的惟一全局最小点，而 $\hat{\boldsymbol{x}}(0^+)$ 是 $P(\boldsymbol{x})$ 在球 D 上的全局最小点. 具体做法是，利用欧拉差分方法对倒向微分方程进行迭代求解，而对每次由差分格式得到的点进行一次牛顿法迭代，即对由差分格式得到的预估点进行一次牛顿法矫正.

算法 1　取误差限 $\varepsilon > 0$ 和迭代次数 N.

步骤 1　给定 $\rho_0 > 0$，对 $[0, \rho_0]$ 作 N 等分，$\rho_k = \frac{(N-k)}{N}\rho_0$，$k = 0, 1, 2, \cdots, N$，则步长为 $h = \frac{1}{N}\rho_0$. 取 $\hat{\boldsymbol{x}}_0 = \arg\{\nabla P(\boldsymbol{x}) + \rho_0\boldsymbol{x} = 0\}$（即 $\nabla P(\boldsymbol{x}) + \rho_0\boldsymbol{x} = 0$ 的解）. $k = 0$.

步骤2 $k=k+1$,计算差分迭代

$$\tilde{\boldsymbol{x}}_k=\hat{\boldsymbol{x}}_{k-1}+h[\nabla^2P(\hat{\boldsymbol{x}}_{k-1})+\rho_{k-1}\boldsymbol{I}]^{-1}\hat{\boldsymbol{x}}_{k-1}$$

步骤3 进行一步牛顿法迭代

$$\hat{\boldsymbol{x}}_k=\tilde{\boldsymbol{x}}_k-[\nabla^2P(\tilde{\boldsymbol{x}}_k)+\rho_k\boldsymbol{I}]^{-1}[\nabla P(\tilde{\boldsymbol{x}}_k)+\rho_k\tilde{\boldsymbol{x}}_k]$$

当 $k<N$,转到步骤2,否则转到步骤4.

步骤4 若 $\|\nabla P(\hat{\boldsymbol{x}}_N)\|<\varepsilon$,取优化问题(3)的最优点 $\boldsymbol{x}^*\approx\hat{\boldsymbol{x}}_N$. 否则,$N=2N$,转到步骤1. 证毕.

注2:由于 $L(\boldsymbol{x},\rho)=P(\boldsymbol{x})+\frac{\rho}{2}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}-d^2)=P(\boldsymbol{x})+\frac{\rho}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}-\frac{\rho d^2}{2}$,所以对于给定的 $\rho(>0)$,$\min\limits_{x\in\mathbf{R}^n}L(\boldsymbol{x},\rho)\Leftrightarrow\min\limits_{x\in\mathbf{R}^n}\left(P(\boldsymbol{x})+\frac{\rho}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\right)$. 这表明算法1并不明显依赖于 d.

以下是关于上述差分-牛顿混合算法的误差估计,推导中每次出现的正常数 C 有所不同,仅依赖于多项式 $P(\boldsymbol{x})$ 本身,同时也注意到上一节的注1. 由于算法是沿倒向微分流进行的欧拉格式,所以有以下3个基本估计:对于 $k=1,2,\cdots,N$,有

微分流初始条件

$$\|\hat{\boldsymbol{x}}_0-\hat{\boldsymbol{x}}(\rho_0)\|=0 \tag{8}$$

微分流 Lipschitz 性质

$$\|\hat{\boldsymbol{x}}(\rho_{k-1})-\hat{\boldsymbol{x}}(\rho_k)\|\leqslant Ch \tag{9}$$

Euler 差分公式误差

$$\|\tilde{\boldsymbol{x}}_k-\hat{\boldsymbol{x}}_{k-1}\|\leqslant Ch \tag{10}$$

定理2 对于 $k=1,2,\cdots,N$,有 $\|\hat{\boldsymbol{x}}_k-\hat{\boldsymbol{x}}(\rho_k)\|\leqslant Ch$.

证明 基于式(8),首先施行归纳法,假设

$$\|\hat{\boldsymbol{x}}_{k-1}-\hat{\boldsymbol{x}}(\rho_{k-1})\|\leqslant Ch \tag{11}$$

由算法1中的步骤3得到

$$\begin{aligned}&\|\hat{\boldsymbol{x}}_k-\tilde{\boldsymbol{x}}_k\|=\\&\|[\nabla^2P(\hat{\boldsymbol{x}}_k)+\rho_k\boldsymbol{I}]^{-1}[\nabla P(\hat{\boldsymbol{x}}_k)+\rho_k\hat{\boldsymbol{x}}_k]\|\leqslant\\&\|[\nabla^2P(\hat{\boldsymbol{x}}_k)+\rho_k\boldsymbol{I}]^{-1}\|\|\nabla P(\hat{\boldsymbol{x}}_k)+\rho_k\hat{\boldsymbol{x}}_k\|\leqslant\\&[\|[\nabla^2P(\hat{\boldsymbol{x}}_k)+\rho_k\boldsymbol{I}]^{-1}-[\nabla^2P(\hat{\boldsymbol{x}}(\rho_k))+\rho_k\boldsymbol{I}]^{-1}\|+\\&\|[\nabla^2P(\hat{\boldsymbol{x}}(\rho_k))+\rho_k\boldsymbol{I}]^{-1}\|][\|(\nabla P(\hat{\boldsymbol{x}}_k)+\rho_k\hat{\boldsymbol{x}}_k)-\\&(\nabla P(\hat{\boldsymbol{x}}(\rho_k))+\rho_k\hat{\boldsymbol{x}}(\rho_k))\|]\leqslant\\&(C\|\hat{\boldsymbol{x}}_k-\hat{\boldsymbol{x}}(\rho_k)\|+C)(C\|\hat{\boldsymbol{x}}_k-\hat{\boldsymbol{x}}(\rho_k)\|)\end{aligned} \tag{12}$$

这里注意到,在倒向微分流上 $\nabla P(\hat{\boldsymbol{x}}(\rho))+\rho\hat{\boldsymbol{x}}(\rho)=0$,又注意到(注1),存在正数 σ 使得在 $[0,\rho^*]$ 上一致有 $\nabla^2P(\hat{\boldsymbol{x}}(\rho))+\rho\boldsymbol{I}>\sigma\boldsymbol{I}$. 进一步,利用归纳法假设式(11)和基本估计式(9)和(10),有以下估计:

$$\begin{aligned}\|\tilde{\boldsymbol{x}}_k-\hat{\boldsymbol{x}}(\rho_k)\|\leqslant&\|\tilde{\boldsymbol{x}}_k-\hat{\boldsymbol{x}}_{k-1}\|+\|\hat{\boldsymbol{x}}_{k-1}-\hat{\boldsymbol{x}}(\rho_{k-1})\|+\\&\|\hat{\boldsymbol{x}}(\rho_{k-1})-\hat{\boldsymbol{x}}(\rho_k)\|\leqslant Ch\end{aligned}$$

于是由式(12)得到

$$\|\hat{\boldsymbol{x}}_k-\tilde{\boldsymbol{x}}_k\|\leqslant Ch \tag{13}$$

利用式(13),归纳法假设式(11)以及基本估计式(9)和(10),最后得到

$$\begin{aligned}\|\hat{\boldsymbol{x}}_k-\hat{\boldsymbol{x}}(\rho_k)\|\leqslant&\|\hat{\boldsymbol{x}}_k-\tilde{\boldsymbol{x}}_k\|+\|\tilde{\boldsymbol{x}}_k-\hat{\boldsymbol{x}}_{k-1}\|+\\&\|\hat{\boldsymbol{x}}_{k-1}-\hat{\boldsymbol{x}}(\rho_{k-1})\|+\|\hat{\boldsymbol{x}}(\rho_{k-1})-\hat{\boldsymbol{x}}(\rho_k)\|\leqslant Ch\end{aligned}$$

证毕

3 数值计算的实例

例1 考虑如下全局优化问题:

$$\min P(\boldsymbol{x})=x_1^4+x_2^4+2x_1^2+2x_2^2+4x_1x_2+2x_1+2x_2 \tag{14}$$

首先,计算多项式 $P(\boldsymbol{x})$ 的梯度和 Hesse 矩阵,得到

$$\nabla P(\boldsymbol{x})=\begin{pmatrix}4x_1^3+4x_1+4x_2+2\\4x_2^3+4x_2+4x_1+2\end{pmatrix}$$

$$\nabla^2P(\boldsymbol{x})=\begin{pmatrix}12x_1^2+4&4\\4&12x_2^2+4\end{pmatrix}=\begin{pmatrix}12x_1^2&0\\0&12x_2^2\end{pmatrix}+\begin{pmatrix}4&4\\4&4\end{pmatrix}\geqslant0$$

应用算法1,取误差限 $\varepsilon=10^{-4}$ 和迭代次数 $N=300$. 取 $\rho_0=0.1$,则步长可以表示为 $h=\frac{1}{N}\rho_0=\frac{1}{10N}>0$. 求解非线性方程组

$$\begin{pmatrix}4x_{11}^3+4x_{11}+4x_{12}+2\\4x_{12}^3+4x_{12}+4x_{11}+2\end{pmatrix}+\begin{pmatrix}0.1x_{11}\\0.1x_{12}\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$$

得到 $\hat{\boldsymbol{x}}_0=\begin{pmatrix}-0.2401\\-0.2401\end{pmatrix}$. 计算差分迭代

$$\begin{pmatrix}(\tilde{x}_k)_1\\(\tilde{x}_k)_2\end{pmatrix}=\begin{pmatrix}(\hat{x}_{k-1})_1\\(\hat{x}_{k-1})_2\end{pmatrix}+h\begin{pmatrix}12(\hat{x}_{k-1})_1^2+\rho_{k-1}+4&4\\4&12(\hat{x}_{k-1})_2^2+\rho_{k-1}+4\end{pmatrix}^{-1}\times\begin{pmatrix}(\hat{x}_{k-1})_1\\(\hat{x}_{k-1})_2\end{pmatrix}$$

得到 $\hat{\boldsymbol{x}}_k$ 的预估值 $\tilde{\boldsymbol{x}}_k$. 再进一步应用牛顿法迭代,矫正预估值得

$$\begin{pmatrix}(\hat{x}_k)_1\\(\hat{x}_k)_2\end{pmatrix}=\begin{pmatrix}(\tilde{x}_k)_1\\(\tilde{x}_k)_2\end{pmatrix}-\begin{pmatrix}12(\tilde{x}_k)_1^2+\rho_k+4&4\\4&12(\tilde{x}_k)_2^2+\rho_k+4\end{pmatrix}^{-1}\times\begin{pmatrix}4(\tilde{x}_k)_1^3+(4+\rho_k)(\tilde{x}_k)_1+4(\tilde{x}_k)_2+2\\4(\tilde{x}_k)_2^3+4(\tilde{x}_k)_1+(4+\rho_k)(\tilde{x}_k)_2+2\end{pmatrix}$$

(下转第822页)