

Modified Moving Least Square Method

Xi Zhang, Renzhong Feng

School of Mathematics and System Science, Beijing University of Aeronautics and Astronautics, Beijing
Email: daxuekanshijie@sina.cn, fengrz@buaa.edu.cn

Received: Nov. 2nd, 2016; accepted: Nov. 19th, 2016; published: Nov. 24th, 2016

Copyright © 2016 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In this paper, based on the MLH (Least-Hardy Moving) method, we present a modified MLS (Moving Least-Squares) method by introducing the correction weight in MLS method. The method is the same as the MLH method, which can effectively approximate the scattered data with outliers. At the same time, this method avoids the problem of MLS method to approximate this kind of scattered data, and its computing efficiency is significantly better than that of MLH method. In order to improve the accuracy of approximation, we also introduce the natural neighbor points, which replace the role of MLS weight function.

Keywords

Scattered Data, Outlier, Moving Least Square (MLS), Moving Least-Hardy Approximation (MLH), Modified Moving Least Square (MMLS)

修正的移动最小二乘方法

张 熙, 冯仁忠

北京航空航天大学, 数学与系统科学学院, 北京
Email: daxuekanshijie@sina.cn, fengrz@buaa.edu.cn

收稿日期: 2016年11月2日; 录用日期: 2016年11月19日; 发布日期: 2016年11月24日

摘 要

本文在MLH (Moving Least-Hardy)逼近方法的基础上, 通过在MLS (Moving Least-Squares)方法中引入

修正权给出了一种修正的MLS方法。该方法与MLH方法一样, 能够对带有奇异值的散乱数据进行有效的逼近, 避免了MLS方法对这类数据逼近的不理想问题, 并且其计算效率明显高于MLS方法。为了提高逼近精度, 该方法还引入了自然邻点, 以自然邻点替代目标函数中权函数的作用。

关键词

散乱数据, 奇异值, 移动最小二乘(MLS), 移动最小Hardy逼近(MLH), 修正移动最小二乘(MMLS)

1. 引言

散乱数据的函数逼近是函数逼近论中的一个重要问题, 可以用于数据拟合、几何建模、CAD、医学成像、微分方程数值解等许多领域。构造逼近函数的常用方法有插值和拟合两类方法。McLain [1] 在 1974 年提出了一种移动最小二乘逼近方法(Moving Least-Squares 简记为 MLS)。这种方法可以用于任意维的散乱数据逼近且已用于计算机图形学、微分方程数值解等领域, 见文献[2] [3] [4]。Zuppa, Armentano 等在 [5] [6] 给出了 MLS 方法关于函数值、一阶和二阶导函数值的逼近误差分析。MLS 方法在含有噪音数据的情况下, 其逼近能力也有一定的保证。但是对于存在奇异值(采样值远远偏离真实值)的散乱数据, MLS 方法的逼近效果就变得非常不理想, 见[7]。针对采样数据中存在奇异值的情况, Levin [7] 在 MLS 的方法基础上提出了一种叫 MLH (Moving Least-Hardy)的逼近方法。该方法对带奇异值的散乱数据逼近有着理想的逼近效果。

MLH 方法是引进了 Hardy's multiquadric 函数[8]并结合了 MLS 和 MLl_1 (移动最小 l_1 逼近)两种逼近方法的思想。当数据中存在奇异值时, MLH 方法中发挥主要作用的是 MLl_1 部分, 而 MLl_1 具有对奇异值不敏感性。与 MLS 不同的是 MLH 方法不是仅通过求解一次法方程就能获得解而是经过有限次迭代后的一个近似解, 属于一种效率低的逼近方法。在本文中我们提出了一种修正的 MLS (Moving Least-Squares)方法, 它不仅具有 MLH 方法处理带奇异值散乱数据的能力, 而且计算效率显著高于 MLH。本文余下部分安排如下: 第二节简要介绍 MLH 方法的工作原理; 第三节介绍修正的 MLS 方法的构造; 第四节为数值实验; 第五节提出结论。

2. MLH 方法概述

给定一组散乱数据 $\{(x_i, f_i)\}_{i=1}^n$, $x_i \in \mathbb{R}^d, f_i \in \mathbb{R}$ 其中的某些 f_i 的采样存在奇异值 f_ξ , 即 $|f_i - \bar{f}_i| \gg 1$ 。MLH 方法构造逼近这类散乱数据的方法如下: 设 $\{p_r(x)\}_{r=1}^K \subseteq \Pi_d^m$ 是一组基(Π_d^m 表示所有次数不超过 m 次的 d 元多项式空间), 取误差函数

$$E_{H_d, \theta, x}(p) = \sum_{i=1}^n H_d \left(f_i - \sum_{r=1}^K c_r p_r(x_i) \right) \theta(\|x_i - x\|) \quad (2.1)$$

其中 $H_d \left(f_i - \sum_{r=1}^K c_r p_r(x_i) \right) = \sqrt{\left(f_i - \sum_{r=1}^K c_r p_r(x_i) \right)^2 + d^2}$, $\theta(\|x_i - x\|)$ 是权函数, d 是一个小参数。

$H_d(t) = \sqrt{t^2 + d^2}$ 称为 *multiquadric* 函数, 对于很小的 t , $H_d(t) \sim d + 1/(2d)t^2 + 1/d^3 O(t^4), t \rightarrow 0$; 对于很大的 t , $H_d(t) \sim |t| \left(1 + O(d^2/t^2) \right), t \rightarrow \infty$ 。

MLH 方法是要确定一组值(解) $\{c_k\}_{k=1}^K$ 使误差函数(2.1)极小化, 不难知道这组解是满足下列方程组

$$\frac{\partial}{\partial c_k} E_{H_d, \theta, x}(p) = \sum_{i=1}^n \frac{p_k(x_i) \left(\sum_{r=1}^K c_r p_r(x_i) - f_i \right)}{H_d \left(f_i - \sum_{r=1}^K c_r p_r(x_i) \right)} \theta(\|x_i - x\|) = 0, k = 1, \dots, K. \quad (2.2)$$

奇异值点误差 $e_\xi = \left| f_\xi - \sum_{r=1}^K c_r p_r(x_\xi) \right| \gg 1$, 该项变为 $\left| p_k(x_\xi) \theta(\|x_\xi - x\|) \right| / \left(1 + O(d^2/|e_\xi|^2) \right)$, 是关于 d 的常数项, 从而对奇异值不敏感。

因方程组(2.2)是一个关于 $\{c_k\}_{k=1}^K$ 的非线性方程组, 所以 MLH 方法定义了一个迭代过程来求解。它以 $c_k^{(0)} = 0, k = 1, \dots, K$ 为初始值, 通过下列迭代过程由第 s 步的值得到第 $s+1$ 步的值

$$\sum_{i=1}^n \frac{p_k(x_i) \left(\sum_{r=1}^K c_r^{(s+1)} p_r(x_i) - f_i \right)}{H_d \left(f_i - \sum_{r=1}^K c_r^{(s)} p_r(x_i) \right)} \theta(\|x_i - x\|) = 0, k = 1, \dots, K, s = 0, 1, \dots \quad (2.3)$$

记 $\mathbf{c} = [c_1, \dots, c_K]^T$, 上面的迭代过程可以写为

$$\mathbf{c}^{(s+1)} = \left(A(\mathbf{c}^{(s)}) \right)^{-1} \mathbf{b}(\mathbf{c}^{(s)}), \quad (2.4)$$

其中(2.4)右端的矩阵 $A(\mathbf{c}^{(s)})_{K \times K}$ 和向量 $\mathbf{b}(\mathbf{c}^{(s)})_{K \times 1}$ 具体形式为

$$\begin{aligned} A(\mathbf{c}^{(s)})_{k,j} &= \sum_{i=1}^n p_k(x_i) p_j(x_i) \frac{\theta(\|x_i - x\|)}{H_d \left(f_i - \sum_{r=1}^K c_r^{(s)} p_r(x_i) \right)}, \\ \mathbf{b}(\mathbf{c}^{(s)})_k &= \sum_{i=1}^n f_i p_k(x_i) \frac{\theta(\|x_i - x\|)}{H_d \left(f_i - \sum_{r=1}^K c_r^{(s)} p_r(x_i) \right)}, \quad k, j = 1, \dots, K. \end{aligned} \quad (2.5)$$

对于迭代格式(2.4), 文献[7]给出了以下两点结论:

- 1) 该方法经有限次迭代收敛;
- 2) 对带有奇异值的散乱数据具有较好的逼近性。

MLH 方法具体描述详见[7]。该方法虽经有限次迭代达到收敛, 但是每次迭代相当于求解一次移动最小二乘问题, 就是需要计算(2.4)中 $A(\mathbf{c}^{(s)})^{-1}, \mathbf{b}(\mathbf{c}^{(s)})$, 导致其计算效率低下。我们针对这个问题提出了一种修正的 MLS 方法(简记为 MMLS 方法), 它只需求一次移动最小二乘问题, 但对带有奇异值的散乱数据也有类似 MLH 一样的逼近性质。

3. 修正的 MLS 方法(MMLS)

MLS 的目标是欲在 Π_d^m 中找到一个多项式 $\sum_{r=1}^K c_r p_r(x)$, 使下列误差达到极小:

$$E_{p, \theta, x} = \sum_{i=1}^n \left(f_i - \sum_{r=1}^K c_r p_r(x_i) \right)^2 \theta(\|x - x_i\|). \quad (3.1)$$

优化问题(3.1)的解向量 $\mathbf{c} = [c_1 \ c_2 \ \dots \ c_K]^T$ 是下列法方程

$$\mathbf{A}\mathbf{c} = \mathbf{b}, \quad (3.2)$$

的解, 其中 \mathbf{A} , \mathbf{b} 的表达式如下:

$$\mathbf{A} = \begin{pmatrix} \sum_{i=1}^n p_1(x_i) p_1(x_i) \theta(\|x-x_i\|) & \cdots & \sum_{i=1}^n p_1(x_i) p_K(x_i) \theta(\|x-x_i\|) \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n p_K(x_i) p_1(x_i) \theta(\|x-x_i\|) & \cdots & \sum_{i=1}^n p_K(x_i) p_K(x_i) \theta(\|x-x_i\|) \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} \sum_{i=1}^n p_1(x_i) f_i \theta(\|x-x_i\|) \\ \vdots \\ \sum_{i=1}^n p_K(x_i) f_i \theta(\|x-x_i\|) \end{pmatrix}. \quad (3.3)$$

如果从给定的散乱数据集 $\{(x_i, f_i)\}_{i=1}^n$ 中去掉数据点 (x_1, f_1) , 此时相应的 MLS 逼近问题的解是下列方程组的解

$$\mathbf{A}' \mathbf{c}' = \mathbf{b}', \quad (3.4)$$

其中

$$\mathbf{A}' = \mathbf{A} - \mathbf{A}_1, \quad \mathbf{A}_1 = \begin{pmatrix} p_1(x_1) p_1(x_1) \theta(\|x-x_1\|) & \cdots & p_1(x_1) p_K(x_1) \theta(\|x-x_1\|) \\ \vdots & \ddots & \vdots \\ p_K(x_1) p_1(x_1) \theta(\|x-x_1\|) & \cdots & p_K(x_1) p_K(x_1) \theta(\|x-x_1\|) \end{pmatrix}, \quad (3.5)$$

$$\mathbf{A}' = \begin{pmatrix} \sum_{i=2}^n p_1(x_i) p_1(x_i) \theta(\|x-x_i\|) & \cdots & \sum_{i=2}^n p_1(x_i) p_K(x_i) \theta(\|x-x_i\|) \\ \vdots & \ddots & \vdots \\ \sum_{i=2}^n p_K(x_i) p_1(x_i) \theta(\|x-x_i\|) & \cdots & \sum_{i=2}^n p_K(x_i) p_K(x_i) \theta(\|x-x_i\|) \end{pmatrix}, \quad (3.6)$$

$$\mathbf{b}' = \mathbf{b} - \mathbf{b}_1, \quad \mathbf{b}_1 = \begin{pmatrix} p_1(x_1) f_1 \theta(\|x-x_1\|) \\ \vdots \\ p_K(x_1) f_1 \theta(\|x-x_1\|) \end{pmatrix}, \quad \mathbf{b}' = \begin{pmatrix} \sum_{i=2}^n p_1(x_i) f_i \theta(\|x-x_i\|) \\ \vdots \\ \sum_{i=2}^n p_K(x_i) f_i \theta(\|x-x_i\|) \end{pmatrix}. \quad (3.7)$$

由(3.2)~(3.4)得

$$\mathbf{A}'(\mathbf{c} - \mathbf{c}') = \mathbf{b}_1 - \mathbf{A}_1 \mathbf{c}. \quad (3.8)$$

把(3.5)和(3.7)代入(3.8)式的右端, 展开得到

$$\mathbf{A}'(\mathbf{c} - \mathbf{c}') = \begin{pmatrix} p_1(x_1) \left(f_1 - \sum_{r=1}^K c_r p_r(x_1) \right) \theta(\|x-x_1\|) \\ \vdots \\ p_K(x_1) \left(f_1 - \sum_{r=1}^K c_r p_r(x_1) \right) \theta(\|x-x_1\|) \end{pmatrix}. \quad (3.9)$$

文献[2] [3] [4] [7]给出: 如果真实采样函数 $f_{\text{really}} \in C^{k+1}$, 那么 MLS 的逼近误差

$$|f_{\text{really}} - f_{\text{MLS}}| < O(h^{k+1}), \quad (3.10)$$

这里 f_{MLS} 通过 MLS 方法得到的逼近函数值, h 是网格范数。那么通过(3.9)和(3.10)得到

$$\mathbf{A}'(\mathbf{c} - \mathbf{c}') \leq \mathbf{M}_p \theta(\|x-x_1\|) O(h^{k+1}), \quad \mathbf{M}_p = [p_1(x_1) | \cdots | p_K(x_1)]^T.$$

由于 \mathbf{A}' 是对称正定的, 有 $\lim_{h \rightarrow 0} \mathbf{c} = \mathbf{c}'$ 。关于这个极限我们可以理解为当采样点取得非常密集时, 去掉

某一个散乱数据点对我们逼近的结果影响可以忽略。

假设在散乱点集 $\{(x_i, f_i)\}_{i=1}^n$ 中, 某些点处有 $|f_e - \tilde{f}_e| \gg 1$ (f_e 是真实的值, \tilde{f}_e 是采样得到的值)。不失一般性我们设 (x_1, \tilde{f}_1) 为奇异点 ($|f_1 - \tilde{f}_1| \gg 1$)。那么, 此时 MLS 的法方程是 $\mathbf{A}c' = \tilde{\mathbf{b}}$, 其中

$$\tilde{\mathbf{b}} = \begin{pmatrix} p_1(x_1)\tilde{f}_1\theta(\|x-x_1\|) + \sum_{i=2}^n p_1(x_i)f_i\theta(\|x-x_i\|) \\ \vdots \\ p_K(x_1)\tilde{f}_1\theta(\|x-x_1\|) + \sum_{i=2}^n p_K(x_i)f_i\theta(\|x-x_i\|) \end{pmatrix}. \quad (3.11)$$

可以发现, 奇异值的存在主要是将列向量 \mathbf{b} 变成了 $\tilde{\mathbf{b}}$, 进而影响了解向量 \mathbf{c} 。根据前述的记号

$$(\mathbf{A}' + \mathbf{A}_1)c' = \tilde{\mathbf{b}} = \mathbf{b}' + (\tilde{\mathbf{b}} - \mathbf{b}'). \quad (3.12)$$

为了使得逼近多项式对应的权系数 $\|\mathbf{c} - \mathbf{c}'\|_\infty < \varepsilon$, ε 是任意正常数。根据线性方程组解的稳定性理论, 如果 $\|\mathbf{A}_1\| < \eta_1$, $\|\tilde{\mathbf{b}} - \mathbf{b}'\| < \eta_2$, η_1, η_2 是很小的正常数, 那么 \mathbf{c}' 可以满足 $\|\mathbf{c} - \mathbf{c}'\|_\infty < \varepsilon$, 这里 \mathbf{c} 是无奇异值采样时最小二乘(3.2)的解。因此我们希望在带有奇异值的散乱数据的 MLS 逼近问题得到的法方程形式为

$$\left(\mathbf{A}' + \frac{\eta}{w}\mathbf{A}_1\right)c' = \mathbf{b}' + \frac{\eta}{w}(\tilde{\mathbf{b}} - \mathbf{b}'), \quad (3.13)$$

其中要求 $w \gg \eta > 0$ 。这样, 奇异值就被削弱成了一个较小的扰动项加在了不含奇异值的节点组法方程中。

基于以上两种假设分析, 我们可以构造一个修正的 MLS 方法来处理带奇异值的散乱数据逼近问题, 其思想是: 在我们修正的 MLS 方法中使用下列误差函数

$$E_{w,p,\theta,x} = \sum_{i=1}^n w(f_i)\theta(\|x-x_i\|) \left(f_i - \sum_{r=1}^K c_r p_r(x_i) \right)^2, \quad (3.14)$$

其中 $\{p_r(x)\}_{r=1}^K \subseteq \Pi_d^n$, $\theta(\|x-x_i\|)$ 是标准 MLS 方法中的权函数, $w(f_i)$ 是对采样函数值 f_i 的一个修正表达式。加入这个修正就是为了针对存在奇异点值时, 满足前述 MLS 逼近中去除奇异值点或者弱化奇异值点对逼近问题的影响。假设采样点 $\{(x_i, f_i)\}_{i=1}^n$ 来自连续函数。

按照求解 MLSS 问题的方法对误差函数(3.14)极小化得到

$$\frac{\partial E_{w,p,\theta,x}}{\partial c_s} = 2 \sum_{i=1}^n w(f_i)\theta(\|x-x_i\|) \left(f_i - \sum_{r=1}^K c_r p_r(x_i) \right) (-p_s(x_i)) = 0, s = 1, \dots, K \quad (3.15)$$

整合得到

$$\mathbf{A}_M \mathbf{c}_M = \mathbf{b}_M, \quad (3.16)$$

其中, \mathbf{A}_M 是法方程矩阵, \mathbf{b}_M 是一个列向量, 它们的具体表达式如下:

$$\mathbf{A}_M = \begin{pmatrix} \sum_{i=1}^n p_1(x_i)p_1(x_i)w(f_i)\theta(\|x-x_i\|) & \cdots & \sum_{i=1}^n p_1(x_i)p_K(x_i)w(f_i)\theta(\|x-x_i\|) \\ \vdots & \vdots & \vdots \\ \sum_{i=1}^n p_K(x_i)p_1(x_i)w(f_i)\theta(\|x-x_i\|) & \cdots & \sum_{i=1}^n p_K(x_i)p_K(x_i)w(f_i)\theta(\|x-x_i\|) \end{pmatrix}$$

$$\mathbf{b}_M = \begin{pmatrix} \sum_{i=1}^n p_1(x_i) f_i w(f_i) \theta(\|x - x_i\|) \\ \vdots \\ \sum_{i=1}^n p_K(x_i) f_i w(f_i) \theta(\|x - x_i\|) \end{pmatrix}. \quad (3.17)$$

关于修正的 MLS 问题(3.14)~(3.17)中的修正 $\{w(f_i)\}_{i=1}^n$, 我们希望其能根据数据 $\{(x_i, f_i)\}_{i=1}^n$ 的不同情况满足相应的条件:

(i) 若数据中不存在奇异值时, 我们希望当网格范数 $h \rightarrow 0$ 时, $\{w(f_i)\}_{i=1}^n$ 满足

$w(f_i) = w(f_j), i, j = 1, \dots, n, i \neq j$, 这样 $\mathbf{A}_M, \mathbf{b}_M$ 退化成(3.3)中 \mathbf{A}, \mathbf{b} , MMLS 逼近趋向于 MLS 逼近。

(ii) 若存在奇异值时, 不妨假设存在一个奇异点 $(x_1, f_{e_1}), |f_{e_1} - f_1| \gg 1$, f_{e_1} 是采样奇异值, f_1 是真实值, 可以将 $\mathbf{A}_M, \mathbf{b}_M$ 转化为 $\mathbf{A}_M = \mathbf{A}_e + (\mathbf{A}_M - \mathbf{A}_e), \mathbf{b}_M = \mathbf{b}_e + (\mathbf{b}_M - \mathbf{b}_e)$, 其中

$$\mathbf{A}_e = w(f_{e_1}) \begin{pmatrix} p_1(x_1) p_1(x_1) \theta(\|x - x_1\|) & \cdots & p_1(x_1) p_K(x_1) \theta(\|x - x_1\|) \\ \vdots & \ddots & \vdots \\ p_K(x_1) p_1(x_1) \theta(\|x - x_1\|) & \cdots & p_K(x_1) p_K(x_1) \theta(\|x - x_1\|) \end{pmatrix}, \quad \mathbf{b}_e = w(f_{e_1}) \begin{pmatrix} p_1(x_1) f_{e_1} \theta(\|x - x_1\|) \\ \vdots \\ p_K(x_1) f_{e_1} \theta(\|x - x_1\|) \end{pmatrix}. \quad (3.18)$$

我们希望当网格范数 $h \rightarrow 0$ 时, $\{w(f_i)\}_{i=1}^n$ 满足 $w(f_p) = w(f_q), q, p = 2, \dots, n; p \neq q$ 并且 $w(f_p) \gg w(f_{e_1}) \geq 0$ 。当 $w(f_{e_1}) = 0$ 时, 带有奇异点 (x_1, f_{e_1}) 的逼近问题退化为关于散乱数据 $\{(x_i, f_i)\}_{i=2}^n$ 的 MLS 逼近问题, 法方程形式变为(3.4); 当 $w(f_{e_1}) \neq 0$ 时, 由于 $w(f_p) \gg w(f_{e_1})$, 可以认为在法方程(3.4)的矩阵 $\mathbf{A}_M - \mathbf{A}_e$ 加一个小的扰动矩阵 \mathbf{A}_e , 向量 $\mathbf{b}_M - \mathbf{b}_e$ 加一个小的扰动向量 \mathbf{b}_e , 法方程形式变为(3.13)。因此, 有 $\|\mathbf{c}_M - \mathbf{c}\|_\infty < \varepsilon$, 对 $\|\mathbf{A}_e\|_\infty < \eta_1, \|\mathbf{b}_e\|_\infty < \eta_2$ 。

注释 1: 修正的 MLS 方法是考虑了关于目标点处的函数值信息对逼近效果的影响, $w(f)$ 是关于采样值的修正, f_i 的值与真实值的偏离越大其对逼近的影响越小。

注释 2: 修正的 MLS 方法同样也可以逼近存在 s ($s > 1, s \in N_+$) 个奇异值的散乱数据。

根据 $w(f_i)$ 所需要满足的条件(i)和(ii), 我们给出下列两个 $w(f_i)$ 模型。

模型一:

$$w(f_i) = \begin{cases} 0, \frac{|f_i| + d^2}{\sum_{i=1}^n (|f_i| + d^2)} \geq \frac{2}{n} \\ 1, \frac{|f_i| + d^2}{\sum_{i=1}^n (|f_i| + d^2)} < \frac{2}{n} \end{cases} \quad (\text{其中 } d \text{ 是一个小参数}).$$

这个模型利用符号函数的性质排除奇异点。如果存在奇异值将退化为不含有奇异值的逼近问题(3.4)。且当 $h \rightarrow 0$ 时, $w(f_i)$ 满足条件(i)、(ii)。

模型二:

$$w(f_i) = \frac{1}{\sqrt{\frac{(f_i - \bar{f})^2}{\sum_{j=1}^n (f_j - \bar{f})^2} + d^2}}, d > 0.$$

d 是一个小参数, $\bar{f} = \sum_{i=1}^n f_i/n$ 。 $w(f_i)$ 可以看作 f_i 对 $\{f_1, \dots, f_n\}$ 的方差贡献率的倒数, 或者 f_i 对集合 $\{f_1, \dots, f_n\}$ 的偏离程度。 偏离程度越高 $w(f_i)$ 越小, 反之越大。 d 根据 \bar{f} 而定, d 正比于 \bar{f} 。 因为我们的模型 $w(f_i)$ 不能严格满足(i)、(ii)中的设定, 在 $h \rightarrow 0$ 的条件下, $w(f_q) \approx w(f_p) \gg w(f_{e_1}) \geq 0$ 。 但是模型的作用同前述分析。

为了测试这两个模型的效果, 我们在第四节的数值试验中进行了验证。

4. 数值实验

4.1. 自然邻点

MLS 中非负权函数 $\theta(\|x-x_i\|)$ 一般被选取成当 x_i 距离 x 越大时, θ 值快速下降接近零, 甚至 θ 被选成具有紧支撑性, 即 $\|x-x_i\| > D, \theta(\|x-x_i\|) = 0$ 。 所以移动最小二乘中权函数的作用是选取目标点 x 附近的参与逼近的数据节点。 自然邻点是“最靠近”目标点的邻域节点。 在本节我们使用自然邻点来替代权函数的这种选取性质, 即修正 MLS 逼近中通过使用自然邻点作为(3.14)中的逼近节点, 而不是通过权函数来选取逼近节点集。 这样做既提高了 $MMLS$ 的计算效率同时又保证逼近精度。 图 1 是二维平面上的自然邻点的一个示例。 自然邻点的详细内容和选取方法见[9] [10]。

在目标函数(3.1)和(3.14)中, 如果 $n \leq K$, 根据 MLS 逼近理论知道逼近函数将会是一个插值于这 n 个散乱节点的插值多项式。 因此, 在使用自然邻点时, 如果得到的自然邻点个数小于 K , 我们拟将自然邻点的自然邻点也加入其中。

图 2 给出了一组二维散乱节点, 红色的点代表目标点的自然邻点, 绿色的点代表红色点自然邻点, 黑色点代表绿色点的自然邻点, 我们称为三层自然邻点模板。 在后面的数值试验中, 将统一采用这种三层自然邻点模板进行逼近。

4.2. $MMLS$ 方法与 MLH 方法的逼近精度比较

我们选取 Franke 函数作为被采样函数, 其表达式为

$$f(x, y) = 0.75 \exp \left[-\frac{(9x-2)^2 + (9y-2)^2}{4} \right] + 0.75 \exp \left[-\frac{(9x+1)^2}{49} - \frac{(9y+1)^2}{10} \right] - 0.2 \exp \left[-(9x-4)^2 - (9y-7)^2 \right] + 0.5 \exp \left[-\frac{(9x-7)^2 + (9y-3)^2}{4} \right], 0 \leq x, y \leq 1$$

选择(2.1)和(3.14)中 $\{p_r(x, y)\}_{r=1}^{10} = \{1, x, y, x^2, xy, y^2, x^3, x^2y, xy^2, y^3\}$, 目标逼近散乱点取 511 个, 如图 3 中红色点表示, 为了消除边界影响加入蓝色散乱点, 这样散乱点总个数为 1000 个。 我们加入的奇异点为 $\tilde{f}(0.3742, 0.2033) = 5$, $\tilde{f}(0.7078, 0.8428) = -5$, 其中真实值为 $f(0.3742, 0.2033) = 0.7077$, $f(0.7078, 0.8428) = 0.0020$ 。 模型中 $d = 0.01$ 。

本文使用下列两种误差表达式

$$RMS_f = \sqrt{\frac{\sum_{i=1}^n (f_i - \tilde{f}_i)^2}{n}},$$

f_i 表示精确函数值, \tilde{f}_i 表示逼近的函数值

$$MaxError_f = \max_{i=1, \dots, n} |f_i - \tilde{f}_i|,$$

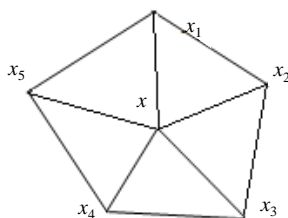


Figure 1. Target point x and its surrounding natural neighbor points $\{x_1, \dots, x_5\}$

图 1. 目标点 x 及其周围的自然邻点 $\{x_1, \dots, x_5\}$

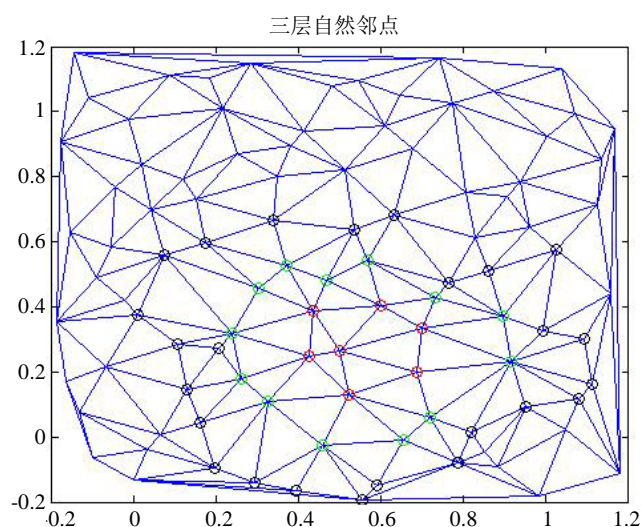


Figure 2. Extended three layer natural neighbor points

图 2. 扩展的三层自然邻点

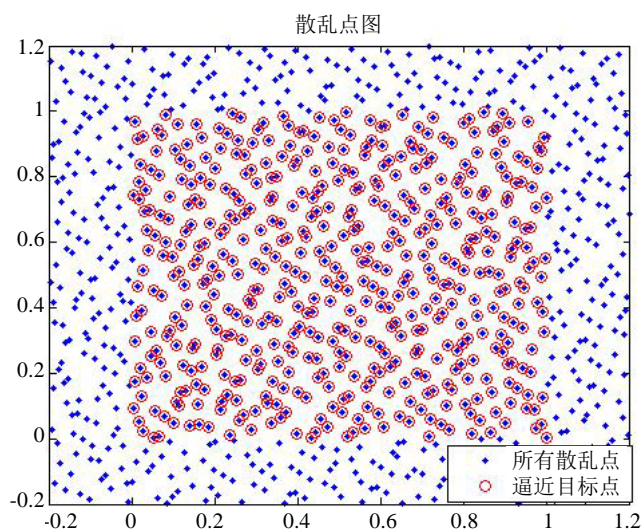


Figure 3. Two dimensional scattered nodes (red dots represent the point of approximation, the blue dots around the red dots are the scattered nodes to eliminate the influence of the boundary)

图 3. 二维散乱节点(红色点代表待逼近点, 围绕红色点的蓝色点是为了消除边界影响加入的散乱节点)

表 1 说明了, MMLS 模型一是将含有奇异值的散乱节点逼近问题转化为排除散乱节点后的 MLS 逼近问题, MMLS 模型二是将奇异值的影响转化为了求解过程中法方程和右端向量的小扰动的影响。在这里模型一优于模型二。

从表 2 看出 MMLS 模型二的精度稍低于 MLH 方法, 但计算效率却优于 MLH 方法, 这和 MLH 方法需要求解多遍移动最小二乘问题, 而 MMLS 只需求解一次该问题的结论一致。

表 3 中的噪音是按照 $f(1+s\%)$, 如果是 5% 的噪音, s 是 $[-5,5]$ 之间的随机数。从表中的结果看出, 在相同条件下, MLH 方法的逼近效果稍优于 MMLS 方法, 但是从计算效率看, MMLS 方法明显优于 MLH 方法。这也说明了 MMLS 方法对含有奇异值和噪音的数据集逼近都能保证一定的精度, 但是计算效率优于 MLH 方法。

5. 结论

文中提出的 MMLS 方法同 MLH 方法一样, 能够解决带有奇异点和小噪声的散乱数据集的逼近问题, 但本文中的方法较 MLH 方法在计算效率上有着显著的提高。可是本文中的 $w(f)$ 函数还可以进一步的改进和提高, 因为针对带有奇异点的散乱数据集的逼近中, 更加合理的选择函数 $w(f)$ 将会直接影响我们对于逼近问题的逼近精度。

Table 1. Error comparison of MMLS method with two different correction methods

表 1. 两种不同修正权的 MMLS 方法误差对比

| 方法 | RMS_f | $MaxError_f$ | 执行时间(秒) |
|----------|---------|--------------|---------|
| MMLS 模型一 | 0.0062 | 0.0559 | 2.8790 |
| MMLS 模型二 | 0.0068 | 0.0577 | 3.2060 |

Table 2. Error comparison between MMLS method and MLH method

表 2. MMLS 方法与 MLH 方法误差对比

| 方法 | RMS_f | $MaxError_f$ | 执行时间(秒) |
|----------|---------|--------------|---------|
| MLH | 0.0053 | 0.0316 | 4.7860 |
| MMLS 模型二 | 0.0066 | 0.0577 | 2.5380 |

Table 3. Scattered data approximation error comparison with outliers and noise

表 3. 带有奇异值和噪音的散乱数据逼近对比

| 方法 | RMS_f | $MaxError_f$ | 执行时间(秒) |
|----------|---------|--------------|---------|
| 噪声比率 2% | | | |
| MLH | 0.0054 | 0.0320 | 4.8750 |
| MMLS 模型二 | 0.0068 | 0.0575 | 3.2490 |
| 噪声比率 3% | | | |
| MLH | 0.0058 | 0.0312 | 5.0150 |
| MMLS 模型二 | 0.0072 | 0.0577 | 3.2660 |
| 噪声比率 5% | | | |
| MLH | 0.0066 | 0.0339 | 5.1760 |
| MMLS 模型二 | 0.0080 | 0.0580 | 3.2380 |

基金项目

该工作受到国家自然科学基金项目(No.11271041)的支持。

参考文献 (References)

- [1] McLain, D.H. (1974) Drawing Contours from Arbitrary Data Points. *Computer Journal*, **17**, 318-324. <https://doi.org/10.1093/comjnl/17.4.318>
- [2] Fasshauer, G.E. and Zhang, J.G. (2007) Scattered Data Approximation of Noisy Data via Iterated Moving Least Squares. In: Lyche, T., Merrien, J.L. and Schumaker, L.L., Eds., *Proceedings of Curve and Surface Fitting: Avignon 2006*, Nashboro Press, 150-159.
- [3] Ding, H., Shu, C. and Tang, D.B. (2005) Error Estimates of Local Multiquadric-Based Differential Quadrature (LMQDQ) Method through Numerical Experiments. *International Journal for Numerical Methods in Engineering*, **63**, 1513-1529. <https://doi.org/10.1002/nme.1318>
- [4] Wendland, H. (2004) Scattered Data Approximation. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511617539>
- [5] Zuppa, C. (2001) Error Estimates for Moving Least Square Approximations. *Applied Numerical Mathematics*, **37**, 397-416. [https://doi.org/10.1016/S0168-9274\(00\)00054-4](https://doi.org/10.1016/S0168-9274(00)00054-4)
- [6] Armentano, M.G. (2001) Error Estimates in Sobolev Spaces for Moving Least Square Approximations. *Society for Industrial & Applied Mathematics*, **39**, 38-51. <https://doi.org/10.1137/s0036142999361608>
- [7] Levin, D. (2014) Between Moving Least-Squares and Moving Least-11. *BIT Numerical Mathematics*, **55**, 781-796. <https://doi.org/10.1007/s10543-014-0522-0>
- [8] Hardy, R.L. (1971) Multiquadric Equations of Topography and Other Irregular Surfaces. *Journal of Geophysical Research Atmospheres*, **76**, 1905-1915. <https://doi.org/10.1029/JB076i008p01905>
- [9] Sibson, R. (1981) A Brief Description of Natural Neighbour Interpolation. *Interpreting Multivariate Data*, **21**, 21-36.
- [10] Watson, D.F. (1985) Natural Neighbour Sorting. *Australian Computer Journal*, **17**, 189-193.

期刊投稿者将享受如下服务:

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: aam@hanspub.org