

# An Adaptive Method for Choosing Collocation Points of RBF Interpolation

Yu Liu<sup>1</sup>, Guanglei Liu<sup>2</sup>, Ziwu Jiang<sup>1</sup>, Dianxuan Gong<sup>3</sup>

<sup>1</sup>School of Sciences, Linyi University, Linyi Shandong

<sup>2</sup>School of Informatics, Linyi University, Linyi Shandong

<sup>3</sup>College of Sciences, Hebei Polytechnic University, Tangshan Hebei

Email: 413491191@qq.com, zwjiang@lyu.edu.cn

Received: Jan. 29<sup>th</sup>, 2016; accepted: Feb. 19<sup>th</sup>, 2016; published: Feb. 22<sup>nd</sup>, 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

Radial basis function (RBF) is one of effective meshfree methods for interpolation on high dimensional scattered data. Since the approximation quality and stability seriously depend on the distribution of the collocation points, it is urgent to find algorithm of choosing optimal point sets for the reconstruction process. In this paper, we give a short overview of existing algorithms including thinning algorithm, greedy algorithm, and so on. A new adaptive data-dependent method is provided at the end with a numerical example to show its efficiency.

## Keywords

Adaptive Method, Collocation Points, Radial Basis Function, Greedy Algorithm

---

# 径向基函数插值配置点的自适应选取算法

刘雨<sup>1</sup>, 刘广磊<sup>2</sup>, 姜自武<sup>1</sup>, 龚佃选<sup>3</sup>

<sup>1</sup>临沂大学理学院, 山东 临沂

<sup>2</sup>临沂大学信息学院, 山东 临沂

<sup>3</sup>河北联合大学理学院, 河北 唐山

Email: 413491191@qq.com, zwjiang@lyu.edu.cn

收稿日期: 2016年1月29日; 录用日期: 2016年2月19日; 发布日期: 2016年2月22日

## 摘要

径向基函数是一种处理高维散乱数据插值的有效方法。由于逼近精度和稳定性都严重依赖于配置点的分布，因此在重建过程中如何设计配置点的优化选取算法成为一个迫切需要解决的问题。在本文中，我们将简要介绍已有的选取算法，例如：细化算法、贪婪算法等。文章的最后，我们给出一种新的自适应选取算法，并通过数值算例验证该方法的高效性。

## 关键词

自适应算法，配置点，径向基函数，贪婪算法

## 1. 引言

依据大量的散乱数据通过插值方法重建一个光滑曲线或曲面，在反向工程、计算机图形学、计算机视觉等很多领域中有着重要的应用[1]。由于原始数据的无规则性，在选取插值方法的时候，尽量使得逼近函数的产生不依赖于数据位置，即选择所谓的无网格方法。径向基函数插值方法作为一种常用的无网格方法，其实质是用一元函数来描述多元函数。由于在通常的方法中所有的原始数据均选取为径向基函数配置中心，这样导致了所求插值函数的不稳定性[2] [3]。因此众多的学者关注于如何从大量的原始数据中选择较少的配置点以使得所求的插值函数近似精度高并且稳定性好[4] [5]。在本文中，我们将给出一种新的一元自适应配置点优化选取方法。本文内容安排如下：在第二部分我们简要介绍径向基函数相关概念性质。第三部分概述诸如：细化算法、贪婪算法等已有的几类选取算法。在第四部分，介绍我们所设计出一种新选取算法，最后通过数值算例演示新算法的有效性。

## 2. 径向基函数

考虑一个定义在  $\Omega$  上的未知函数  $f(x)$ ，其定义在离散样本集合  $X = \{x_k\}_{k=1}^N \subset \Omega \subseteq \mathbb{R}^d$  上的函数值集合为  $\{f(x_k)\}_{k=1}^N$ 。给定一个一元函数  $\varphi: \mathbb{R}^+ \rightarrow \mathbb{R}$ ，令  $\Phi(x) = \varphi(\|x\|)$ ，其中  $\|\cdot\|$  为欧几里得范数，则函数  $f(x)$  可以近似表示成如下线性组合形式

$$\hat{f}_N(x) = \sum_{k=1}^N \alpha_k \varphi(r_k), \quad (1)$$

其中：  $r_k = \|x - x_k\|$ ，  $\{\alpha_k\}_{k=1}^N$  为待定系数，其值由如下的插值条件决定

$$\hat{f}_N(x_k) = f(x_k), \quad 1 \leq k \leq N.$$

称  $X = \{x_k\}_{k=1}^N$  为径向基函数插值配置点集，由径向基函数  $\varphi$  与配置点集  $X$  决定的矩阵记为

$$A_\varphi = \begin{pmatrix} \varphi(r_{11}) & \cdots & \varphi(r_{1N}) \\ \vdots & \ddots & \vdots \\ \varphi(r_{N1}) & \cdots & \varphi(r_{NN}) \end{pmatrix}.$$

如果  $A_\varphi$  非奇异，则线性方程组

$$\sum_{k=1}^N \alpha_k \varphi(r_{ik}) = f(x_i), \quad 1 \leq i \leq N,$$

的解存在并唯一，其中  $r_{ik} = \|x_i - x_k\|$ 。

下面为常用的径向基函数：

$$\text{多二次函数(MQ): } \varphi(r_j) = \sqrt{r^2 + c^2}$$

$$\text{逆多二次函数(IMQ): } \varphi(r_j) = 1/\sqrt{r^2 + c^2}$$

$$\text{薄板样条函数(TPS): } \varphi(r_j) = r_j^2 \log(r)$$

$$\text{高斯函数(Gauss): } \varphi(r_j) = e^{-c^2 r^2}$$

大多数径向基函数不能保证插值矩阵  $A_\varphi$  的非奇异性，此时应对插值函数进行处理，即为了保证插值问题解的存在唯一性，在式(1)中等号的后边添加多项式项  $P(x)$ ：

$$\hat{f}_N(x) = \sum_{k=1}^N \alpha_k \varphi(r_k) + \sum_{s=1}^M b_s P_s(x),$$

且满足条件

$$\sum_{k=1}^N \alpha_k P_s(x_k) = 0, \quad 1 \leq s \leq M, \quad (2)$$

$\{P_s\}_{s=N+1}^M$  是  $P_{m-1}$  的一组基， $M = \binom{m+1}{2}$  为  $P_{m-1}$  的维数。令  $P^T = \begin{pmatrix} p_1(x_1) & \cdots & p_1(x_N) \\ \vdots & \ddots & \vdots \\ p_M(x_1) & \cdots & p_M(x_N) \end{pmatrix}$ ，结合式(2)，

则插值条件可以写成如下形式：

$$\begin{pmatrix} A_\varphi & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} f \\ 0 \end{pmatrix},$$

其中  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)^T$ ， $\beta = (b_1, b_2, \dots, b_M)^T$ ， $f = (f_1, f_2, \dots, f_N)^T$ 。

任给  $\mathbb{R}^d$  上的一个径向基函数  $\Phi(x) = \varphi(\|x\|)$ ，对于任意选取的配置点集  $X$ ，若矩阵  $A_\varphi$  均为正定矩阵，此时我们可以定义一个由  $\Phi$  诱导的内积：

$$(\Phi(x-\cdot), \Phi(y-\cdot))_\Phi = \Phi(x-y), \quad x, y \in \mathbb{R}^d.$$

利用此内积可以定义如下的线性空间：

$$F_\Phi := \left\{ \sum_{i=1}^n \lambda_i \Phi(x_i - \cdot) \mid \lambda_i \in \mathbb{R}, n \in \mathbb{N}, x_i \in \mathbb{R}^d \right\}$$

即  $F_\Phi$  是所有有限个  $\Phi(x_i - \cdot)$  的线性组合构成的空间。取  $F_\Phi$  的闭包为  $\Gamma_\Phi$ ，则  $\Gamma_\Phi$  是一个 Hilbert 空间，其上的内积仍然是  $(\cdot, \cdot)_\Phi$ 。并且任取空间  $\Gamma_\Phi$  中的一个元素  $f(x)$ ，都有

$$f(x) = (f, \Phi(x-\cdot))_\Phi, \quad x \in \mathbb{R}^d, f \in \Gamma_\Phi.$$

称  $\Gamma_\Phi$  是  $\Phi$  的本性空间[6]。为了给出径向基插值的误差估计，定义配置点分布的填充距离如下：

$$h_{X,\Omega} = \sup_{x \in \Omega} \min_{x_j \in X} \|x - x_j\|.$$

则对于任意的本性空间中的函数  $f \in N_\Phi(\Omega)$ ，有如下的误差估计

$$|\hat{f}_N - f| \leq C \cdot h_{X,\Omega} \cdot \|f\|_\Phi, \quad (3)$$

其中  $C$  为常数,  $\|f\|_{\Phi}$  为函数在本性空间中的范数。一方面, 从误差估计式中可以看出要使得近似精度高, 应该尽可能多选取配置点。但另一方面, 过多的选点将导致矩阵  $A_{\Phi}$  的条件数过大, 从而导致求解插值函数的不稳定性[7]。若定义配置点分布的分散距离为  $q_X := \frac{1}{2} \min_{1 \leq i < j \leq N} \|x_i - x_j\|$ , 则配置点的选取应使得  $h_{X, \Omega}$  尽量小, 并且  $q_X$  尽量大。

### 3. 配置中心的选取方法

通过前面的介绍可知, 利用径向基函数插值方法对散乱数据进行处理, 如果数据点非常多, 为了降低插值矩阵条件数, 节省运算成本, 我们希望能够用尽量少的配置点并获得尽量高的近似精度。这个问题可描述为: 给定了大量的原始散乱数据, 要从中选取  $n$  个配置点进行插值, 如何选择才能使  $h_{X, \Omega}$  尽量小, 并且  $q_X$  尽量大。M. S. Floater [8], A. Iske [9]和 S. D. Marchi [10]及其合作者在此方面做了较多工作。到目前为止, 这方面的研究成果主要可以分为两类: 依赖数据的和不依赖数据的。

给定插值区域  $\Omega$  和大量散乱数据, 要在从中选取的一个包含  $n$  个配置点的集合  $X$  上插值并使得效果最优, 即应该使得所选结点的分散距离  $q_X$  尽量大的同时填充距离  $h_X$  尽量小。令  $\rho_X = \frac{q_X}{h_X}$ , Iske 指出对  $d$  维有界区域做正单纯性剖分得到的顶点集满足  $\rho_X$  最大, 且有

$$\rho_X = \sqrt{\frac{2(d+1)}{d}}。$$

基于同样的思想, 当给定了插值节点集  $X$  时, Floater 和 Iske 通过添加点、细化点和交换点等方法构造分层点集序列  $X_1 \subset X_2 \subset \dots \subset X_N = X$ , 这里  $\#(X_i) = i$ , 满足相应的  $(\rho_{X_1}, \rho_{X_2}, \dots, \rho_{X_N})$  在某种意义上最大[8]。Iske 将同样的想法也应用到了分片算法中。

径向基函数插值可写成 Lagrange 形式

$$f^*(x) = \sum_{j=1}^n f_j l_j(x),$$

其中  $l_j(x_i) = \delta_{ij}$ , 定义能量函数如下

$$P_{\Phi, X}^2(x) = \Phi(x, x) - 2 \sum_{j=1}^N l_j(x) \Phi(x, x_j) + \sum_{j, k=1}^N l_j(x) \Phi(x_j, x_k) l_k(x)$$

则  $P_{\Phi, X}^2(x)$  的大小直接反应了插值的误差, 并且当  $X \subset Y$  时, 有  $P_{\Phi, X}^2(x) \leq P_{\Phi, Y}^2(x)$ 。

借助于能量函数, Marchi 给出了下面的贪婪算法[10]:

- 1) 初始步骤: 有界区域  $\Omega \subset \mathbb{R}^d$ , 在  $\Omega$  的边界上取一点  $x_1$ , 令  $X_1 = \{x_1\}$ ;
  - 2) 迭代步骤: 对于  $j \geq 2$ , 在  $\Omega \setminus X_{j-1}$  上选取点  $x_j$  满足  $x_j$  到  $X_{j-1}$  的距离最远, 令  $X_j = X_{j-1} \cup \{x_j\}$ 。
- 设  $E \subset \mathbb{R}^2$  是一个有界区域, 令  $x_0 \in E$ , 对  $j = 1, 2, \dots$ , 定义  $x_j$  满足:

$$\min_{0 \leq k \leq j-1} \|x_j - x_k\| = \max_{x \in E \setminus \{x_0, x_1, \dots, x_{j-1}\}} \min_{0 \leq k \leq j-1} \|x - x_k\|,$$

称这样的点列  $\{x_j\}$  为  $E$  上的 Leja-Bos 点列, 显然通过贪婪算法选取的点列是一个 Leja-Bos 点列。

另外, 在径向基神经网络中常用  $K$ -均值聚类算法选取中心,  $K$ -均值聚类算法不仅简单而且性能良好, 在模式识别中应用十分广泛。这种方法的选点过程与被插值数据无关, 也是一种不依赖数据的选点方法[11]。

### 4. 我们的选取方法

从前面的分析可以看到，径向基插值精度与  $h_x$  有关，插值矩阵的条件数与  $q_x$  关系密切，而  $h_x$  和  $q_x$  只与配置点的几何分布有关，而与要插值的函数或者数据没有关系。所以依据  $h_x$  和  $q_x$  所得到的取点方法都是不依赖数据的，都是要使所选的点集在插值区域上的分布尽量均匀。这类算法所得到的配置点集都可以说是一种 Leja-Bos 点列。观察式(3)右端的误差函数，它不仅与分布密度有关，还与被插值函数  $f$  有关。被插值函数  $f$  的性质由给定的散乱数据的性质反映，所以在选择插值中心的时候应该同时考虑散乱数据的值及其变化性质。基于此，我们希望找到一个依赖数据的自适应的选点方法，能够根据具体的数据选择最佳配置点。

为了能够在选择配置点时同时考虑点集分布和具体数据的变化，我们定义广义的 Leja-Bos 点列：设  $E$  是给定原数据点集， $\omega$  是定义在  $E$  上的一个正定函数，令  $x_0 \in E$  满足

$$\omega(x_0)\|x_0\| = \max_{x_i \in E} \omega(x_i) \oplus \|x_i\|,$$

其中  $\oplus$  表示某种运算，并且当  $j=1,2,\dots$  时，令  $x_j \in E$  满足：

$$\min_{0 \leq k \leq j-1} \omega(x_j) \oplus \|x_j - x_k\| = \max_{x_l \in E \setminus \{x_0, x_1, \dots, x_{j-1}\}} \min_{0 \leq k \leq j-1} \omega(x_l) \oplus \|x_l - x_k\|,$$

其中， $\|\cdot\|$  表示欧几里得范数。称  $\omega$  是  $E$  上的广义权函数，称如此选取的点列  $\{x_j\}$  为  $E$  上的一个广义 Leja-Bos 点列。

权函数  $\omega$  需要根据具体的数据点集来确定，不同的权函数对插值误差的影响比较明显，一般我将函数值的变化剧烈程度与广义权函数联系起来。在本文中，我们取广义权函数为最近 3 个点之间的二阶差商平方平均值

$$\omega(x_i) = \sqrt{\sum_{l=i-1}^{i+1} \left( \frac{1}{x_i - x_{l-1}} \left( \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} - \frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}} \right) \right)^2} / \sqrt{3},$$

$\oplus$  运算取两者的加权平均即  $\omega(x) \oplus \|x\| = \lambda\omega(x) + (1-\lambda)\|x\|, 0 \leq \lambda \leq 1$ ，然后根据找出广义 Leja-Bos 点列即为所求的配置点集。

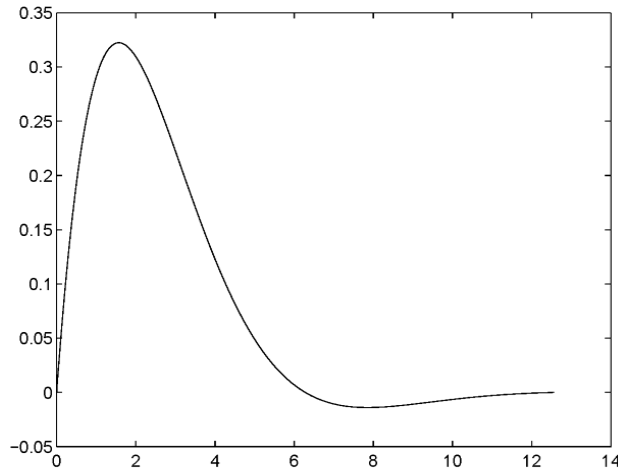


Figure 1. The profile of the  $f(x)$

图 1. 曲线  $f(x)$  的形状

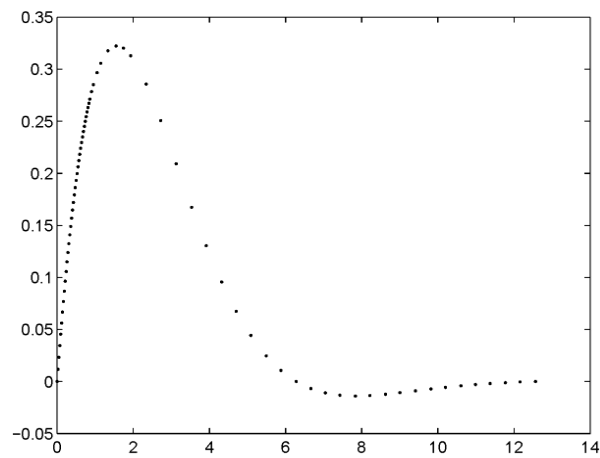


Figure 2. The distribution of chosen collocation points  
图 2. 配置点的分布情况

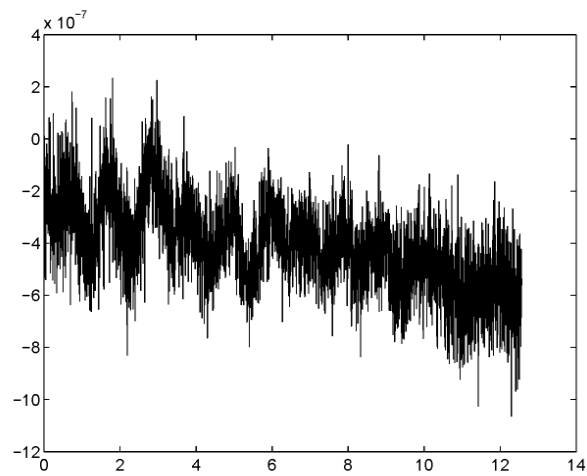


Figure 3. Error function corresponding to 527 equidistant nodes  
图 3. 均匀分布选取 527 个配置点的误差图形

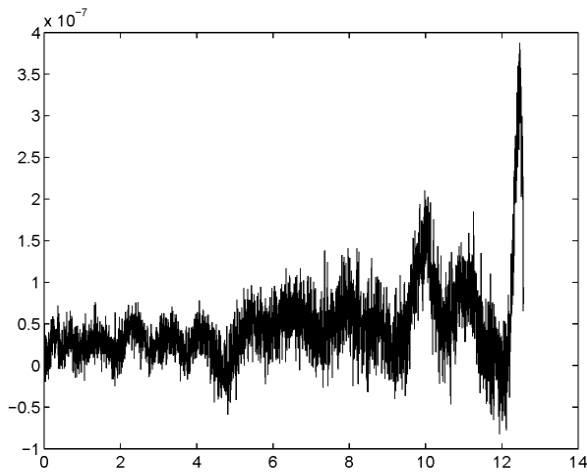


Figure 4. Error function corresponding to generalized-Leja-Bos sequence of 72 points  
图 4. 选取 72 个广义 Leja-Bos 配置点的误差图形

## 5. 数值算例

本数值算例中, 我们选择函数  $f(x) = e^{-\frac{x}{2}} \sin\left(\frac{x}{2}\right)$ , 给定原始数据为区间  $[0, 4\pi]$  的 527 等分点(包括端点)及其函数值。径向基函数选择为 MQ 函数, 其中的形状参数  $r = 5$ 。  $x_i$  点处的广义权值取该点与其最近 3 个点的平方平均值, 权系数  $\lambda = 0.5$ , 配置点根据算法自适应取到 72 个。原函数图像如图 1 所示, 从函数图像可以看出曲线兼具变化较为剧烈和平缓的部分。利用本文的方法所选取的配置点如图 2 所示。插值等分点的误差函数如图 3 所示, 利用本文方法所做插值方法的误差图像如图 4 所示。从误差图像容易看出, 得到同样近似精度, 本文所采用的方法选取的配置点更少因而效率更高。

## 6. 结论

本文主要探讨了如何利用大量的原始数据, 自适应选取较少的径向基函数插值配置点。我们利用配置点分布的分散距离并结合数据点处的二阶插商信息自适应选取了一系列广义 Leja-Bos 点列, 将此点列作为径向基函数插值配置点, 从数值算例可以看出, 利用本文给出的算法所得出的插值函数效率更高。

## 基金项目

大学生创新创业训练计划项目(编号: 201410452002)、国家自然科学基金(项目编号: 11301252)。

## 参考文献 (References)

- [1] 吴宗敏. 散乱数据拟合的模型、方法和理论[M]. 北京: 科学出版社, 2007.
- [2] Fasshauer, G.E. (2007) Meshfree Approximation Methods with Matlab. World Scientific, Singapore. <http://dx.doi.org/10.1142/6437>
- [3] Franke, R. (1982) Scattered Data Interpolation, Test of Some Methods. *Mathematics of Computation*, **38**, 181-200.
- [4] Dyn, N., Floater, M.S. and Iske, A. (2002) Adaptive Thinning for Bivariate Scattered Data. *Journal of Computational and Applied Mathematics*, **145**, 505-517. [http://dx.doi.org/10.1016/S0377-0427\(02\)00352-7](http://dx.doi.org/10.1016/S0377-0427(02)00352-7)
- [5] Iske, A. (2003) Progressive Scattered Data Filtering. *Journal of Computational and Applied Mathematics*, **158**, 297-316. [http://dx.doi.org/10.1016/S0377-0427\(03\)00449-7](http://dx.doi.org/10.1016/S0377-0427(03)00449-7)
- [6] Wendland, H. (2005) Scattered Data Approximation (Cambridge Monographs on Applied and Computational Mathematics; 17). Cambridge University Press, Cambridge.
- [7] Schaback, R. (1995) Error Estimates and Condition Numbers for Radial Basis Function Interpolation. *Advances in Computational Mathematics*, **3**, 251-264. <http://dx.doi.org/10.1007/BF02432002>
- [8] Floater, M.S. and Iske, A. (1998) Thinning Algorithms for Scattered Data Interpolation. *BIT Numerical Mathematics*, **38**, 705-720. <http://dx.doi.org/10.1007/BF02510410>
- [9] Behrens, J. and Iske, A. (2002) Grid-Free Adaptive Semi-Lagrangian Advection Using Radial Basis Functions. *Computers and Mathematics with Applications*, **43**, 319-327. [http://dx.doi.org/10.1016/S0898-1221\(01\)00289-9](http://dx.doi.org/10.1016/S0898-1221(01)00289-9)
- [10] Marchi, S.D. (2003) On Optimal Center Locations for Radial Basisfunction Interpolation: Computational Aspects. *Rendiconti del Seminario Matematico Università e Politecnico di Torino (Splines Radial Basis Functions and Applications)*, **61**, 343-358.
- [11] Xu, B.Z., Zhang, B.L. and Wei, G. (1994) Neural Network Theory and Its Application. South China University of Technology Press, Guangzhou.