

基于泰勒公式和遗传算法的函数拟合研究

刘国伟, 王佳琪*

华北电力大学计算机系, 河北 保定

收稿日期: 2025年4月21日; 录用日期: 2025年5月13日; 发布日期: 2025年5月23日

摘要

基于函数表达式绘制函数图像容易实现, 但是将数据表述为函数表达式较为困难。在统计学和机器学习中, 该问题通常表现为回归建模、插值和参数优化等形式, 虽然计算高效, 但在非线性复杂数据环境下表现受限。近年来, 深度神经网络、进化计算等智能优化方法的引入, 为函数拟合提供了更强大的建模能力和优化策略。该研究基于泰勒公式和遗传算法, 实现了较好的函数拟合效果。具体来说, 泰勒公式可以用函数多项式的形式近似表达函数值, 很大程度上简化了函数拟合的难度。遗传算法可以迭代探索函数拟合的最优结果, 实现了在较短的时间内获得较优的函数拟合表达式。该研究的方法在线性函数、非线性函数的实验中均取得了有竞争性的优化增益结果。

关键词

泰勒公式, 遗传算法, 函数拟合, 进化计算

Study on Function Fitting Based on Taylor Formula and Genetic Algorithm

Guowei Liu, Jiaqi Wang*

Department of Computer, North China Electric Power University, Baoding Hebei

Received: Apr. 21st, 2025; accepted: May 13th, 2025; published: May 23rd, 2025

Abstract

Drawing function image based on function expression is easy to realize, but it is difficult to express data as function expression. In statistics and machine learning, this problem is usually expressed in the form of regression modeling, interpolation and parameter optimization, which are computationally efficient but limited in nonlinear and complex data environments. In recent years, the introduction of intelligent optimization methods such as deep neural networks and evolutionary computation has

*通讯作者。

文章引用: 刘国伟, 王佳琪. 基于泰勒公式和遗传算法的函数拟合研究[J]. 应用数学进展, 2025, 14(5): 291-298.
DOI: 10.12677/aam.2025.145257

provided more powerful modeling capabilities and optimization strategies for function fitting. The study achieves a good function fitting effect based on Taylor formula and genetic algorithm. Specifically, Taylor formula can be used to approximately express function values in the form of function polynomials, which greatly simplifies the difficulty of function fitting. Genetic algorithm can iteratively explore the optimal result of function fitting, and achieve a better function fitting expression in a shorter time. The method of this study achieves competitive optimization gains in both linear and nonlinear functions.

Keywords

Taylor Formula, Genetic Algorithm, Function Fitting, Evolutionary Computation

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

得益于科学技术的发展, 函数拟合深入到科学研究、科创制造的各个领域之中。许多科学研究都离不开函数拟合的支持, 例如: 分子扩散分析[1]、药物敏感性分析[2]和飞船轨迹预测[3]。

给定一个函数表达式, 可以适当选取自变量的值, 代入表达式后计算因变量的值, 将自变量和因变量的值组合后在坐标系中描点, 再依次连线, 即可获取函数图像。在计算机技术高度发达的今天, 以上过程实现起来显得更加容易。但给出一个复杂的函数图像, 没有任何其他辅助提示, 直接用数学表达式描述这个函数图像是很困难的[4]。首先, 函数表达式的组合非常复杂, 难以有统一的表达式形式描述所有的函数图像。其次, 函数图像到函数表达式的过程是一个一对多的过程, 函数图像在数学表达中, 可对应多种形式的表达式。与通过函数描点连线所得到的具有唯一性的结果不同, 函数图像的表达式呈现出非唯一性的特征。

总之, 将函数表达式进行描点、连线, 绘制出函数图像较为容易, 而对给定的函数图像表述为函数表达式较为困难。现有的研究主要有参数化方法、非参数化方法、机器学习方法、启发式方法等, 但是仍未能有效地解决潜在的问题。

针对难以有统一的表达式形式描述所有函数的问题, 泰勒公式[5]给出了可行的解决方案。泰勒公式可以将一些复杂的函数近似地表示为简单的多项式函数, 从而实现了函数表达式形式的统一。

针对难以用函数表达式表述函数图像的问题, 遗传算法[6]给出了可行的解决方案。遗传算法通过模拟生物进化过程, 可以在迭代过程中搜索最优解, 非常适用于将函数图像表述为函数表达式的研究。实验结果表明, 遗传算法在全局搜索和内在启发式随机搜索方面具有优势, 并且可以显著改善自适应模糊神经网络的学习效果[7], 因此将遗传算法应用于函数拟合具有较大的可行性。

2. 研究现状

近些年, 研究函数拟合主要有参数化方法、非参数化方法、机器学习方法、启发式方法等。

参数化方法假设函数具有特定的形式, 通过调整参数来最小化误差。调整参数最经典的方案是最小二乘法, 而特定的函数形式包括但不限于多项式函数、指数函数、三角函数。Smyth [8]在早些年就提出了基于最小二乘法的固定形式函数拟合的相关方案, 给后续的研究提供了很多思路。

非参数化方法不提前预设具体的函数形式, 利用数据的局部或全局特性进行拟合。Nadaraya [9]和

Watson [10]提出了通过核函数对数据点进行加权平均, 以平滑估计回归函数, 从而避免对函数形式的假设。这样的方案被称为Nadaraya-Watson核回归, 能够灵活拟合非线性关系, 使回归曲线更加平滑, 适用于噪声数据的回归分析。

机器学习方法通过学习数据特征来进行拟合, 很适用于复杂或高维的数据。代表方法有多层感知机、支持向量机。Bishop等人[11]系统性地介绍了模式识别和机器学习的基本原理, 包括贝叶斯方法、神经网络、支持向量机等模型, 并结合概率图模型进行深入分析, 为后续的研究提供了一套系统的机器学习理论框架。Vapnik [12]提出了统计学习理论并奠定了支持向量机的数学基础, 强调了结构风险最小化原则, 为后续机器学习理论提供了严密的统计基础, 使得支持向量方法在高维、小样本问题上取得了良好的泛化能力。

启发式方法通过构建优化目标与迭代过程来探索最优解, 适用于复杂、非线性、无显示表达式的情况, 代表方法有粒子群优化、遗传算法。Kennedy等人[13]提出了粒子群优化算法, 受鸟群和鱼群群体行为的启发, 通过个体与群体信息的交互搜索最优解, 为后续的研究提供了一种高效的全局优化方法, 广泛应用于机器学习、工程优化和控制领域。Lambora等人[14]详细描述了遗传算法的渊源、基本方案, 为后续启发式算法的研究奠定了坚实的基础。

3. 理论与方法

3.1. 泰勒公式和遗传算法的简述

泰勒公式是高等数学中的一项重要内容, 可以将一些复杂的函数近似地表示为简单的多项式函数, 实现化繁为简的效果。中学阶段可以接触到三角函数、幂函数、指数函数、对数函数等的泰勒公式, 最终都表述为了多项式函数的形式。在计算机科学中, 许多数学函数的计算实际上是通过截取泰勒级数的前几项来实现的, 这种方法广泛应用于科学计算和工程技术中。

遗传算法基于生物繁殖和进化理论。自然选择是一种普遍接受的理论, 它阐明了优胜劣汰的理论。生物体通过选择、基因重组和基因突变进行迭代繁殖[15], 其中更适合环境的基因更有可能被保留下来。在遗传算法中, 编程中的复杂问题被抽象为自然选择的过程[16], 适应度是衡量种群适应环境的能力的指标。遗传算法的整体流程如图1所示。

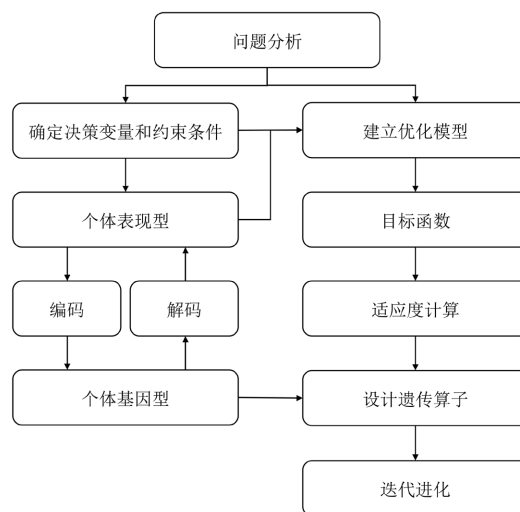


Figure 1. Overall flow of solving practical problems with genetic algorithm

图 1. 用遗传算法解决实际问题的整体流程

在遗传算法的设计中, 个体的表现型决定了基因的编码和解码方式。该研究将初等函数的泰勒展开式的多项式参数视为遗传信息。与经典遗传算法相比, 本研究在设计遗传算子时考虑了全局最优解和局部最优解的差异, 充分注意了保留最优个体的优先性。

3.2. 基于泰勒公式和遗传算法的函数拟合设计

确定决策变量和约束条件方面, 个体表现型的约束条件为有意义的函数表达式, 决策变量为表达式组合, 包含操作数和运算符的组合。将初等函数的泰勒展开式[17]的多项式的参数作为遗传信息, 用于实现遗传信息编码与解码的过程。泰勒展开式的多项式的每一项都为加项, 所以无需考虑运算符的编码。

该研究对传统遗传算法的算子进行了改进, 以便缓解自然选择的不确定性对函数拟合收敛的影响。传统遗传算法的选择算子采取了简单的随机联赛法, 该研究兼顾了随机联赛法和最优部分个体的保留。这样的改进有利于生物进化整体上向着好的方向进行, 从而得到理想的适应度衰减曲线。

适应度评估决定了遗传算法的优化方向。函数拟合的目的是找到一个尽可能地表述图像轨迹的表达式, 因此该研究将目标表达式的图像与待拟合的图像之间的面积作为适应度。通过高等数学中定积分的思想, 将这个面积转化为有限个曲面梯形的面积累加计算, 其中, 累加值越小, 说明目标表达式越理想[18]。

初始群体的基因序列由随机数生成。函数表达式的复杂度决定了随机生成的个体的基因序列的长度。越复杂的函数表达式, 随机生成的个体的基因序列越长。选择算子、交叉算子和变异算子的设计对应着自然选择的迭代流程, 是遗传算法的核心内容。

该研究用随机联赛法设计了选择算子。每代中全部的 80 个个体被平均分为 8 组, 每组中适应度最小的一半个体是优胜者。为了实现全局最优解的保留, 每代中 40 个优胜者进行自我复制用于保留到下一代。40 个优胜者的原始个体将执行后续的交叉算子和变异算子。该研究设计这样的选择算子缓解了遗传算法不确定性大的问题。真实的自然选择过程中有海量的个体, 即便不进行最优个体的备份也很有可能保留优秀基因。而遗传算法中的群体大小是受限的, 通过这样额外的设计可以实现较好的迭代效果。

该研究用单点交叉法设计了交叉算子。选择算子中随机联赛法生成了 40 个优胜者, 这 40 个优胜者随机配对后繁衍产生下一代。繁衍过程中, 基因重组会以一定的概率发生。单点交叉法指的是随机生成一个位置点, 该位置点之前的位置点不进行基因重组, 该位置点与其之后的位置点进行基因重组, 如图 2 所示。单点交叉法有利于保留局部基因片段的结构, 适用于函数拟合、路径优化等具有局部依赖关系的优化问题。

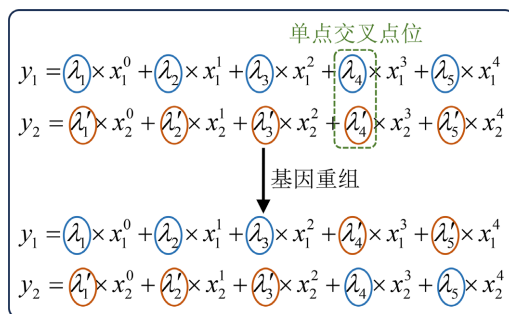


Figure 2. Example of gene recombination by single-point crossover method

图 2. 单点交叉法基因重组示例

该研究用基本位变异法设计变异算子。在生物发育时基因会以一定的概率发生突变。基本位变异法

指的是随机生成首、尾两个位置点, 这两个位置点及其之间的位置点进行基因突变的操作, 其他位置点保留原始基因, 如图 3 所示。基本位变异法可以控制突变发生在局部的片段, 在提高算法收敛效率的同时, 避免了过度破坏个体的基因结构。

$$y_1 = \lambda_1 \times x_1^0 + \lambda_2 \times x_1^1 + \lambda_3 \times x_1^2 + \lambda_4 \times x_1^3 + \lambda_5 \times x_1^4$$

首尾突变位

$$y_1 = \lambda_1 \times x_1^0 + \lambda_2 \times x_1^1 + \lambda_3' \times x_1^2 + \lambda_4' \times x_1^3 + \lambda_5 \times x_1^4$$

基因突变

Figure 3. Example of gene mutation by basic bit mutation method

图 3. 基本位变异法基因突变示例

该研究对超参数进行了特定的取值。个体编码串的长度设置为 5, 在函数拟合问题中体现为函数表达式的复杂度。群体规模设置为 80, 在函数拟合问题中体现为函数表达式的数量。发生基因重组的概率设置为 0.5, 发生基因突变的概率设置为 0.1, 迭代次数设置为 200。

遗传算法作为一种优化搜索算法, 在函数拟合分析中具有重要的应用价值。不过, 该算法存在计算过程较长的固有特性, 且不同用户所使用的机器在硬件配置上存在显著差异。硬件性能的不同会直接影响计算速度, 因此难以确保每台机器都能在瞬间得出计算结果。鉴于此, 当用户使用本系统中对给定函数图像进行拟合分析的功能时, 需要耐心等待遗传算法完成计算。此外, 遗传算法本质上是基于概率的搜索算法, 其搜索过程具有较大的不确定性。这就导致在实际应用中, 无法保证一定能求得效果理想的拟合函数。为了在保证拟合效果与计算效率之间达成平衡, 本函数拟合系统将群体个体数设定为 80, 迭代数设定为 200。这样的参数设置, 既能够在一定程度上有效地开展函数拟合工作, 又能使系统进行遗传算法计算的时间尽可能缩短。

4. 实验与讨论

水平线性拟合、一般线性拟合、类抛物线拟合、不规则图形拟合的函数表达式结果、拟合效果如图 4 所示。可以发现, 该研究提出的基于泰勒公式和遗传算法的函数拟合方法既尽可能地拟合了目标图形, 又避免了过度拟合, 这样的效果对任何难以预料的函数图像具有很强的鲁棒性。

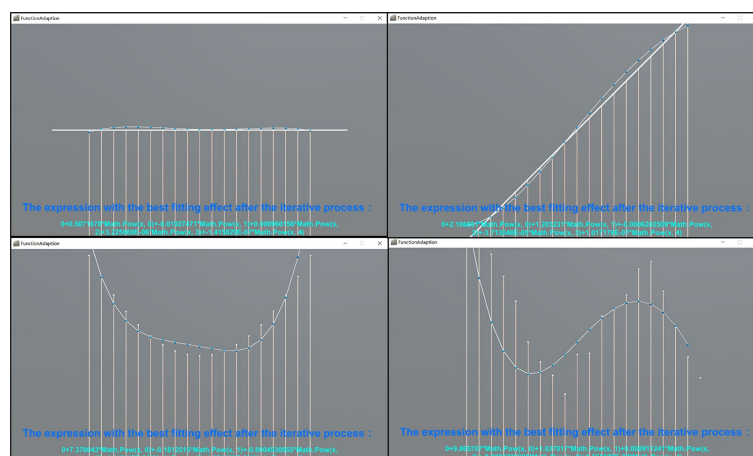


Figure 4. Demonstration of the effects of various fitting examples

图 4. 各种拟合示例的效果展示

水平线性拟合、一般线性拟合、类抛物线拟合、不规则图形拟合的适应度衰减曲线如图 5 所示。可以发现, 该研究提出的基于泰勒公式和遗传算法的函数拟合方法可以通过迭代在较短的时间内取得很小的适应度, 说明该研究的方案具备很高的时效性。

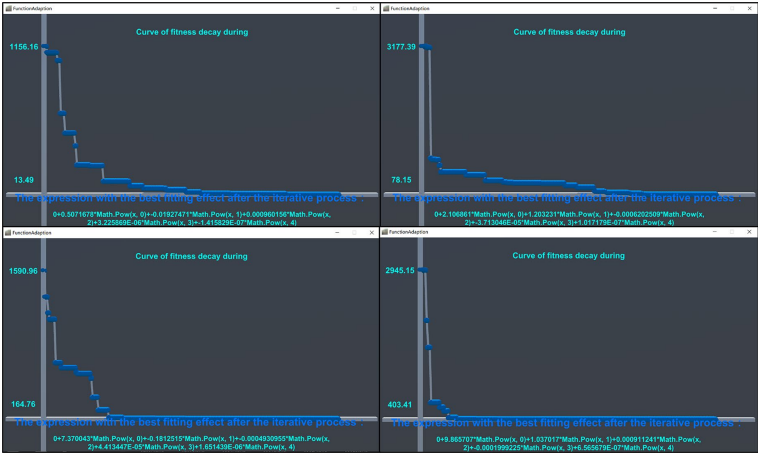


Figure 5. Fitness decay curves for various fitting examples
图 5. 各种拟合示例的适应度衰减曲线

不同的图形, 遗传算法的迭代计算时间也会略有差异。遗传算法对难以拟合的图形需要的计算量更大, 因而计算时间会更长。在 CPU (i7-9750H)、GPU (GTX1660Ti) 的硬件支撑下, 实现对给定函数图像进行拟合分析的功能时, 完成 80 个个体迭代 200 次的函数拟合任务的平均运行时间约为 2 分钟 38 秒。

依据软件工程学的等价类划分、边界值分析的理论[19], 使用多组测试用例进行了对给定函数图像进行拟合分析功能的测试, 运行过程均流畅。将适应度下降的比例称为遗传算法的优化增益, 计算公式为:

$$IR = \frac{F_{\text{initial}} - F_{\text{final}}}{F_{\text{initial}}} \quad (1)$$

其中, F_{initial} 表示初代群体的最佳适应度, F_{final} 表示末代群体的最佳适应度。与以往的研究进行了比较, 以往研究简记为 GA [20], 该研究记为 TGA, 实验结果如表 1 所示。

Table 1. Performance comparison results of function fitting
表 1. 函数拟合的性能对比结果

方法	指标	水平线性拟合	一般线性拟合	类抛物线拟合	不规则图形拟合
GA	F_{final}	10.36	2.93	283.18	None
GA	F_{initial}	54.13	871.54	571.71	None
GA	IR	80.86%	99.66%	50.47%	None
TGA	F_{final}	13.49	78.15	164.76	403.41
TGA	F_{initial}	1156.16	3177.39	1590.96	2945.15
TGA	IR	98.83%	97.54%	89.64%	86.30%

优化增益在 85%以上的标记为绿色, 70%~85%的标记为橙色, 70%以下的标记为红色。通过表 1 可以发现, 该研究对各种测试样例均有较好的优化增益, 说明该研究通过泰勒公式和遗传算法的综合设计实现了较为稳定的进化效果, 可以获得较为理想的函数拟合效果。

我们使用均方误差(MSE)对拟合效果进行了进一步分析, 实验结果如表 2 所示。

Table 2. Comparison results of the mean squared error of function fitting

表 2. 函数拟合的均方误差对比结果

方法	水平线性拟合	一般线性拟合	类抛物线拟合	不规则图形拟合
GA	1.55	2.79	759.95	None
TGA	2.21	26.42	176.29	785.68

从表中可以看出, 相比于 GA, 该研究的方法在类抛物线函数拟合、不规则图形拟合上取得了令人较为满意的 MSE 值。尽管 TGA 在水平线性拟合、一般线性拟合中的 MSE 值略高于 GA, 但仍是可接受的值, 间接反映了 TGA 防止过拟合的能力。综合来看, 该研究的方法具备更强的鲁棒性, 能够更好地适应复杂形状的函数。

5. 局限性与未来展望

目前, 本文采用目标表达式图像与待拟合图像之间的面积差作为适应度, 该方法对整体形状差异具有良好的度量能力, 但对图像平移较为敏感。例如, 两个形状相似但位置偏移较大的图像有较大的面积差, 然而实际上可能是拟合效果较好的图像。未来的工作中将改进适应度评价机制, 综合衡量拟合的形状、图像位置偏差等内容。

此外, 由于无需预设函数形式实验的独特设置, 本文仅与有限的研究现状进行了实验性能对比。未来团队将查阅更多参考文献, 同时自行设计支持向量回归、神经网络等经典方法进行性能对比。

6. 结论

该研究提出了一种基于泰勒公式和遗传算法的函数拟合方法, 在没有任何预设的函数形式的情况下较为高效地实现了对函数图像拟合出函数表达式。具体来说, 该研究使用泰勒公式将一些复杂的函数近似地表示为简单的多项式函数, 从而实现了函数表达式形式的统一; 使用遗传算法在迭代过程中搜索最优解, 解决了用函数表达式表述函数图像的难题。实验结果表明, 该研究提出的方案具备较高的遗传算法优化增益, 函数拟合效果较优。

参考文献

- [1] Ignacio, M. and Slater, G.W. (2021) Using Fitting Functions to Estimate the Diffusion Coefficient of Drug Molecules in Diffusion-Controlled Release Systems. *Physica A: Statistical Mechanics and Its Applications*, **567**, Article ID: 125681. <https://doi.org/10.1016/j.physa.2020.125681>
- [2] Szabó, B., Pintér, P., Antal, Á., Szlávecz, Á., Chase, J.G. and Benyó, B. (2024) Multi-Dimensional Quantile Regression Using Polynomial Function Fitting for Insulin Sensitivity Forecasting. *IFAC-PapersOnLine*, **58**, 327-331. <https://doi.org/10.1016/j.ifacol.2024.11.058>
- [3] Chen, J., Chen, H., Zhao, Y. and Li, X. (2022) Fb-BiGRU: A Deep Learning Model for Ais-Based Vessel Trajectory Curve Fitting and Analysis. *Ocean Engineering*, **266**, Article ID: 112898. <https://doi.org/10.1016/j.oceaneng.2022.112898>
- [4] Ziegel, E.R. (2003) The Elements of Statistical Learning. *Technometrics*, **45**, 267-268. <https://doi.org/10.1198/tech.2003.s770>
- [5] Tom, A. and Apostol, C. (1967) One-Variable Calculus, with an Introduction to Linear Algebra. Blaisdell Publishing Co.
- [6] Sampson, J.R. (1976) Adaptation in Natural and Artificial Systems (John H. Holland). *SIAM Review*, **18**, 529-530. <https://doi.org/10.1137/1018105>
- [7] Liang, J., Cao, H., Lu, Y. and Su, M. (2024) Architecture Search of Accurate and Lightweight CNNs Using Genetic Algorithm. *Genetic Programming and Evolvable Machines*, **25**, Article No. 13. <https://doi.org/10.1007/s10710-024-09484-4>

-
- [8] Smyth, G.K. (2002) Nonlinear Regression. In: *Encyclopedia of Environmetrics*, Vol. 3, Wiley, 1405-1411.
 - [9] Nadaraya, E.A. (1964) On Estimating Regression. *Theory of Probability & Its Applications*, **9**, 141-142.
<https://doi.org/10.1137/1109020>
 - [10] Watson, G.S. (1964) Smooth Regression Analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, **26**, 359-372.
 - [11] Bishop, C.M. and Nasrabadi, N.M. (2006) Pattern Recognition and Machine Learning. Springer.
 - [12] Vapnik, V. (2013) The Nature of Statistical Learning Theory. Springer Science & Business Media.
 - [13] Kennedy, J. and Eberhart, R. (1995) Particle Swarm Optimization. *Proceedings of ICNN'95-International Conference on Neural Networks*, Vol. 4, 1942-1948. <https://doi.org/10.1109/icnn.1995.488968>
 - [14] Lambora, A., Gupta, K. and Chopra, K. (2019) Genetic Algorithm—A Literature Review. 2019 *International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, Faridabad, 14-16 February 2019, 380-384.
<https://doi.org/10.1109/comitcon.2019.8862255>
 - [15] 周明, 孙树栋. 遗传算法原理及应用[M]. 北京: 国防工业出版社, 1999: 4-6.
 - [16] Rooker, T. (1991) Review of Genetic Algorithms in Search, Optimization, and Machine Learning. *AI Magazine*, **12**, 102-103. <https://doi.org/10.1609/aimag.v12i1.889>
 - [17] 濮家程, 段晓慧, 叶阳天, 等. 多点泰勒展开的渐进构造和显式计算及应用[J]. 杭州电子科技大学学报: 自然科学版, 2014, 34(5): 78-81.
 - [18] 同济大学数学系. 高等数学[M]. 第 6 版. 北京: 高等教育出版社, 2007: 202-217.
 - [19] 张海藩. 软件工程导论[M]. 第四版. 北京: 清华大学出版社, 2003: 100-130.
 - [20] Sun, Q., Wang, J. and Zhou, X. (2024) A Function Fitting System Based on Genetic Algorithm. In: Jin, H., *et al.*, Eds., *Green, Pervasive, and Cloud Computing*, Springer, 156-168. https://doi.org/10.1007/978-981-99-9893-7_12