

单位代码	10602
学 号	2020012189
分 类 号	N949
密 级	公开



广西师范大学
GUANGXI NORMAL UNIVERSITY

硕士学位论文

平行相移深度神经网络
模型及算法研究

Research on Parallel Phase Shift Deep Neural
Network Model and Algorithm

学院：物理科学与技术学院

学科专业：系统科学

研究方向：复杂网络理论与应用

年级：2020 级

研究生：邢康

指导教师：孙小军 教授

完成日期：2023 年 5 月

平行相移深度神经网络 模型及算法研究

专业名称：系统科学

申 请 人：邢康

指导教师：孙小军 教授

论文答辩委员会

主席：_____

委员：_____

平行相移深度神经网络 模型及算法研究

研究生: 邢康 指导教师: 孙小军 教授
专业: 系统科学 方向: 复杂网络理论与应用 年级: 2020 级

摘要

随着实验核物理的发展,核数据的类别及数量越来越多,机器学习方法在核物理领域中的应用也越来越广,但在中子共振截面领域至今仍无人涉及。中子诱发燃料核 ^{235}U 裂变反应的共振截面,存在低频、高频和不可分辨共振区,这些共振裂变截面大于快中子能区裂变截面几个数量级,这在新一代核能技术、国防战略武器、基础核物理研究等领域中有着非常重要的实际应用和学术价值。

本文以 $^{235}\text{U}(n, f)$ 裂变共振截面为研究对象,主要的工作有以下三个方面。首先通过理论分析和算法验证了传统深度神经网络方法(DNN)在低频振荡数据情况下的适用性,但在高频振荡数据情况下具有很大的局限性。其次,借鉴平行相移理论(PPS),在DNN基础上发展能同时适用于低频和高频变化的共振裂变截面的平行相移深度神经网络(PPSDNN)模型。该模型主要包含以下几方面的内容:(1)傅立叶变换和尼奎斯特采样定理,傅立叶变换可以将目标函数从时域空间变换到频域空间,尼奎斯特采样定理揭示了在不失真的情况下需要的最小采样频率;(2)单位分解定理,它可以将目标函数分解成若干部分;(3)B-样条函数,它可以用来自构造满足单位分解的函数;(4)平行相移在频率空间中的实现,这是PPSDNN算法核心,可以将目标函数从高频平移到低频。最后,为应用该模型,本文通过构建算法,使用python语言研发出程序代码。算法主要思路:第一步把目标函数通过单位分解分解成若干部分,对低频部分直接使用DNN训练并给出最佳结果;对高频部分通过傅立叶变换将目标函数在频域空间上使用平行相移将高频平移到低频,第二步使用DNN训练这些低频函数并将训练结果再还原到高频振荡数据;第三步将训练得到的若干目标函数叠加得到最终对目标函数的训练结果。通过 $^{235}\text{U}(n, f)$ 裂变共振截面评价库数据的学习,发现PPSDNN模型可以得到较好的效果,为进一步针对实验数据和获得工程所需的中子共振参数奠定了坚实的基础,具有很大的应用前景。

关键词: $^{235}\text{U}(n, f)$; 中子共振截面; 深度神经网络方法; 平行相移深度神经网络模型; 算法

Research on Parallel Phase Shift Deep Neural Network Model and Algorithm

Postgraduate Student: kang Xing

Advisor: Prof. Xiaojun Sun Grade: 2020

Major: Systems Science Direction: Theory and Applications of Complex Networks

Abstract

As experimental nuclear physics has developed, the categories and quantities of nuclear data have become increasingly diverse, and the application of machine learning methods in the field of nuclear physics has become more and more widespread. However, no one has yet ventured into the field of neutron resonance cross sections. The resonance cross section of the neutron-induced fission reaction of the fuel nucleus ^{235}U has low-frequency, high-frequency, and indistinguishable resonance regions, and these resonance fission cross sections are several orders of magnitude greater than the fission cross sections in the fast neutron energy range. This has very important practical applications and academic value in fields such as new generation nuclear energy technology, national defense strategic weapons, and basic nuclear physics research.

This paper focuses on the study of the resonance cross section of $^{235}\text{U}(n, f)$ fission, with the following three main objectives. Firstly, the applicability of the traditional deep neural network (DNN) method is theoretically analyzed and algorithmically verified for low-frequency oscillation data, but its performance is limited in high-frequency oscillation data. Secondly, the parallel phase shift deep neural network (PPSDNN) model is developed based on the DNN, drawing on the parallel phase shift theory (PPS), which is capable of simultaneously handling resonant fission cross sections with low and high frequency changes. This model mainly includes the following aspects: (1) Fourier transform and Nyquist sampling theorem, where the Fourier transform can transform the target function from the time domain space to the frequency domain space, and the Nyquist sampling theorem reveals the minimum sampling frequency required to avoid distortion; (2) unit decomposition theorem, which can decompose the target function into several parts; (3) B-spline function, which can be used to construct functions that satisfy the unit decomposition; (4) the implementation

of parallel phase shift in the frequency domain space, which is the core of the PPSDNN algorithm, and can shift the target function from high frequency to low frequency. Finally, in order to apply the model, this paper develops a program code using the Python language by constructing an algorithm. The algorithm's main idea is to first decompose the target function into several parts using unit decomposition, and then directly use DNN to train the low-frequency part and obtain the best result. For the high-frequency part, the target function is shifted from high frequency to low frequency using parallel phase shift in the frequency domain space with the Fourier transform. In the second step, the DNN is used to train these low-frequency functions and the training results are restored to high-frequency oscillation data. In the third step, the training results of the various target functions are superimposed to obtain the final training result for the target function. Through the evaluation of the $^{235}\text{U}(n, f)$ fission resonance cross-section library data, the PPSDNN model is found to achieve good results and lay a solid foundation for further targeting experimental data and obtaining the neutron resonance parameters needed for engineering, with great potential for application.

Keywords: $^{235}\text{U}(n, f)$; Neutron resonance cross section; Deep Neural Network method; Parallel Phase Shift Deep Neural Network model; Algorithm

目录

第 1 章 引言	1
1.1 机器学习简介	1
1.2 中子共振数据简介	1
1.3 机器学习方法在核物理中的应用	4
第 2 章 平行相移深度神经网络模型	5
2.1 深度神经网络模型	5
2.1.1 深度神经网络算法简介	5
2.1.2 频率原则	6
2.2 平行相移深度神经网络模型	12
2.2.1 模型及算法简介	12
2.2.2 傅里叶变换与尼奎斯特采样定理	13
2.2.3 单位分解定理	14
2.2.4 B-样条函数	15
2.2.5 相移在频域空间中的实现	18
2.3 算法流程	21
第 3 章 初步结果	23
3.1 toy model 测试结果	23
3.2 $^{235}\text{U}(n, f)$ 评价库数据测试结果	27
第 4 章 总结与展望	33
附录 A 计算双曲正切函数的傅立叶变换	34
参考文献	37
致谢	
论文独创性声明	
论文使用授权声明	

插图

1-1 $^{235}\text{U}(n, f)$ 反应截面数据	2
2-1 图 (a): 使用 DNN 训练有噪声的正弦函数。图 (b): 损失函数分别在训练集 测试集上的表现。图 (c): 训练集测试集在频率空间上的表现。	9
2-2 DNN 学习函数 $\sin(x) + \sin(5x)$, 学习结果随 epoch 的变化, 并与 $\sin(x)$ 进 行比较	10
2-3 使用梯度下降学习 11 阶插值函数	10
2-4 使用 tanh 和 relu 激活函数分别对 ENDF/B-VIII.0 中 $^{235}\text{U}(n, f)$ 反应截面 进行学习	11
2-5 不同激活函数下 DNN 的损失函数	11
2-6 2 阶 B 样条函数	16
2-7 3 阶 B 样条函数	17
2-8 PPSDNN 算法流程图	22
3-1 toy model 目标函数 3-1 式的图形	23
3-2 使用 PPSDNN 训练目标函数 3-1 式	25
3-3 toy model 在区间 $[0, 1]$ 上的训练细节	25
3-4 toy model 在区间 $[1, 2]$ 上的训练细节	26
3-5 toy model 在区间 $[2, 3.14]$ 上的训练细节	26
3-6 使用 frequency sweep 训练高频函数	27
3-7 PPSDNN 方法对 ENDF/B-VIII.0 数据库中 $^{235}\text{U}(n, f)$ 反应截面的学习 . . .	28
3-8 PPSDNN 方法在能量区间为 $[10^{-6}, 10^{-5}]\text{MeV}$ 的学习细节	29
3-9 PPSDNN 方法在能量区间为 $[10^{-5}, 10^{-4}]\text{MeV}$ 的学习细节	29
3-10 PPSDNN 方法在能量区间为 $[10^{-4}, 10^{-3}]\text{MeV}$ 的学习细节	30
3-11 PPSDNN 方法在能量区间为 $[10^{-3}, 10^{-2}]\text{MeV}$ 的学习细节	30
3-12 PPSDNN 方法在能量区间为 $[10^{-2}, 10^{-1}]\text{MeV}$ 的学习细节	31
3-13 PPSDNN 方法在能量区间为 $[0.1, 30]\text{MeV}$ 的学习细节	31
1-1 半圆围道	35

第1章 引言

1.1 机器学习简介

机器学习^[1]是一种基于数据和统计学的人工智能方法，旨在自动从数据中提取规律和模式，并基于这些规律和模式进行预测和决策。与传统的编程方式不同，机器学习的程序不需要明确规定所有的规则和逻辑，而是通过学习和建模来自动提取规律和模式，从而处理大量复杂的数据，甚至能够发现人类无法感知的模式和规律^[2]。

机器学习包括多种算法和技术，例如监督学习、无监督学习和强化学习。在监督学习中，机器学习算法从标记好的数据中学习，以预测未来的输出结果。在无监督学习中，机器学习算法自主从数据中发现结构和模式，而不需要任何标记好的数据。在强化学习中，机器学习算法通过与环境交互，学习最优策略，以达成某种目标。

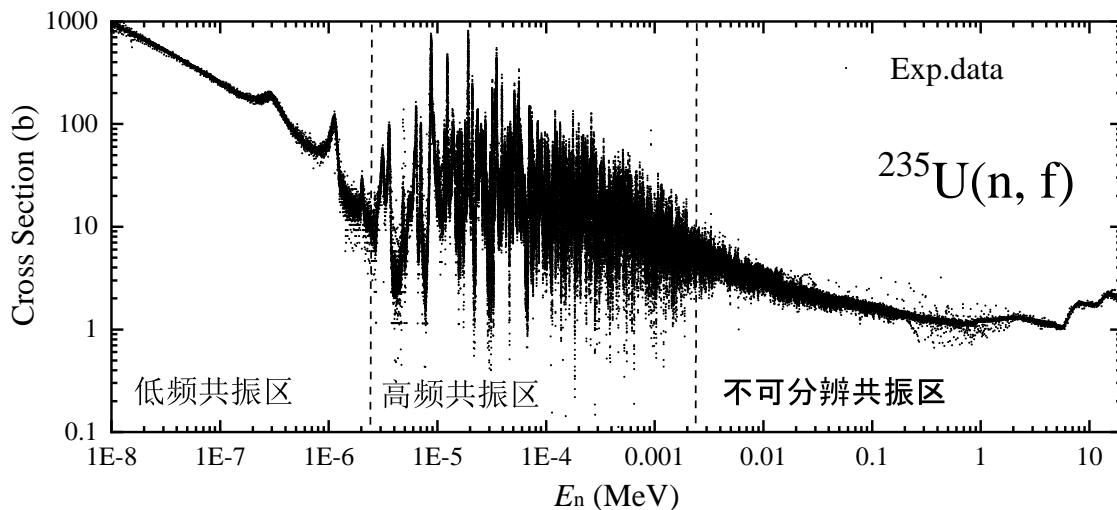
机器学习在各个领域都得到了广泛应用，例如在自然语言处理^[3]领域，机器学习可以通过学习大量文本数据，自动分析文本中的语法和语义，从而实现文本分类和命名实体识别^[4]等任务；在图像识别领域，机器学习可以通过训练模型识别出图片中的物体、人脸等信息；在信用评估中可以通过分析用户历史数据预测其信用风险；在金融风险管理中可以对市场数据进行分析预测投资风险；在医疗诊断中可以通过模型学习预测疾病风险或判断疾病种类。在这些领域中，机器学习可以通过从大量数据中学习和发现模式，提高预测准确性和效率，甚至超过人类的表现。

因此，机器学习是一种基于数据和统计学的人工智能方法，通过不断学习和优化，让计算机可以自主完成某些任务和决策，具有广泛的应用前景。

1.2 中子共振数据简介

中子核反应的基本过程包括中子的入射、原子核的吸收和复合核的形成，以及复合核的衰变等多个阶段^[5]。中子与原子核相互作用的方式取决于中子的能量和原子核的结构。当中子的能量较低时，会与原子核中的核子发生相互作用，导致原子核激发或熔合。而当中子的能量足够高时，原子核会发生裂变等其他核反应。在核物理学中，截面（Cross section）是衡量原子核反应中发生特定相互作用的可能性大小的物理量，截面越大，发生某特定核反应的概率也就越大。截面通常会随着中子能量的变化而变化，当用低能中子去轰击靶核并测量其激发曲线时，会出现曲线的形状出现许多宽窄不等、高低不齐的共振峰，共振峰对应的截面称为共振截面。以 $^{235}\text{U}(n, f)$ 反应为例，此反应的实验数据数据如图1-1所示。从图中分析得到，在高频共振区有 900 多个共振峰。

实验表明：除氕、氘、氚以外的所有核素均存在中子共振现象，但不同核素的共振谱

图 1-1 $^{235}\text{U}(n, f)$ 反应截面数据

各不相同，每种核素都有其独一无二的中子共振谱。中子共振截面具有几大鲜明的特征^[6]：1) 特定核素的峰值出现在特定的能量下，这一性质为同位素识别、同位素定量测量以及它们衍生的各类核探测技术提供了强有力的支撑。2) 中子共振截面非常大，常常高于快中子能区几个数量级，这在实际应用中非常重要并产生极大的影响。3) 中子共振截面不受压力、温度、湿度、形状、大小等物质形态的影响。中子共振截面因具备上述特性，可应用于多种极端条件，如炸药爆轰或冲击波等动态系统中最难测量的瞬时温度；无损检测极其珍贵稀缺文物、涡轮叶片、航空发动机等特定核素的含量和分布；核反应堆中铀富集度检测或二氧化钚分布检测；发生融堆核事故后核燃料形态和新核素形成检测等。

每个峰的最高点对应的激发能被称为共振能量或共振能级，每个峰对应的半宽度是共振宽度。共振截面、共振能量和共振宽度是描述中子共振现象的基本物理量，这些用于描述中子与原子核间相互作用中的共振现象的物理量又称为共振参数，共振参数主要包含以下几个物理量：

- (1) 共振截面。共振截面是指粒子在通过物质时，发生特定频率的共振吸收的概率。在核物理领域，共振截面通常指中子与原子核相互作用时，发生裂变反应的截面。中子与原子核相互作用会导致中子在原子核内部不断反弹，当中子的能量和原子核的固有能量匹配时，就会发生共振吸收并引起裂变反应。这个匹配能量对应的中子能量就是共振能，对应的共振截面就是中子能量等于共振能时的截面。
- (2) 共振能量。中子共振能量是指中子与原子核之间相互作用发生共振时的能量^[7]。中子与原子核相互作用时，复合核会形成一些离散的能级，这些能级与原子核结构有关。中子共振能量的测量可以通过中子散射实验或共振吸收实验来获得。中子散射实验通常是将一束中子入射到样品上，然后测量散射角度和能量分布等参数，通过分析实验数

据得到中子共振能量。共振吸收实验是通过测量材料对中子束的吸收率来获得中子共振能量。

- (3) 共振宽度。中子共振宽度描述了共振态能量附近反应截面的变化范围。共振宽度反映了共振态的寿命，即共振态存在的时间长短。根据反应道的类型，共振宽度可以分为总宽度、光子辐射宽度，中子宽度，质子宽度等几种类型。总宽度是指所有粒子通道的总和，包括弹性和非弹性散射以及各种反应产物。光子辐射宽度表示发射 γ 光子的能量宽度，中子宽度表示发射中子的宽度，质子宽度表示发射质子的宽度。
- (4) 自旋宇称。中子共振自旋是指中子与原子核相互作用形成的共振态所具有的自旋角动量^[8]。核子是费米子，费米子的自旋通常用半整数表示，例如 $1/2$ 等。中子共振自旋对于核反应的选择规律和谱线结构有着重要影响。中子共振中宇称是指中子与原子核相互作用形成的共振态所具有的宇称量子数。宇称通常用正负号表示，分别表示偶宇称和奇宇称。中子共振宇称对于核反应的对称性和选择规律有着重要影响。

中子共振参数在核反应相关领域有着广泛的应用。中子共振参数可以用于预测核反应截面^[9]，这对于新一代核能技术，基础核物理学，国防战略武器研究^[10]中起着重要作用。总之，中子共振参数是研究核反应机制和核结构的重要物理参数，具有广泛的应用前景。通过不断深入研究和应用，可以更好地掌握核反应的规律性和特征，为核能领域的应用提供更好的理论支持和技术保障。

要得到精准的共振核数据，一个关键性的前提条件就是要能拟合好这些高频振荡的截面数据，这个问题在我国一直没有得到很好的解决。目前，能拟合这些高频振荡的截面数据并得到共振参数的方法有 R 矩阵方法^[11]，核反应 R 矩阵理论是研究中重核核反应中共振能区核反应的重要理论方法。在国际上有很多程序都可以用 R 矩阵方法来计算中子共振，美国橡树岭国家实验室早在上世纪七十年代就组建团队，并于 1980 年研制出大型 SAMMY 程序^[12]用于中子共振核数据的评价工作，其后不断改进和更新多种版本。SAMMY 程序在美国各重要单位（如 ORNL^[13]、KAPL、LANL^[14]、TUNL^[15]等）及很多国家（如比利时、日本、法国、保加利亚等）得到广泛应用，但该程序的最新版本在我国并未得到授权使用。在国内，清华大学陈振鹏教授等人曾先后耗时三十余年发展了 RAC 模型^[16]、中国核数据中心及南开大学蔡崇海教授组成联合团队在最近 10 年发展了 FDRR 模型^[17]。两团队各自取得显著进展，能给出一定的结构信息，并成功应用于中子诱发 ${}^6\text{Li}$ 、 ${}^7\text{Li}$ 和 ${}^9\text{Be}$ 等几个轻核反应^[18]，只是这些轻核反应的中子共振峰仅有 1 个或很少几个。受理论模型及计算方法局限，如要推广到几十至几百个共振峰、入射能区达几个数量级、从轻核到重核较大范围靶核体系，RAC 和 FDRR 模型与 SAMMY-8 仍有较大差距。

1.3 机器学习方法在核物理中的应用

随着实验核物理与核理论的发展，中子共振数据的种类和数量越来越多。在国内，有CSNS反角白光中子源^[19]获得的最新各类中子共振截面^[20]。而机器学习方法恰好依赖这些大数据。目前机器学习方法在核物理领域的各个领域均得到了广泛应用^[21-22]，例如核理论、核实验、核数据和加速器等领域。

在核理论领域，可使用深度神经网络、贝叶斯神经网络^[23]和基于决策树的模型来预测原子核的各种性质，如核质量^[24]、电荷半径^[25]、 α 衰变半衰期^[26]、 β 衰变半衰期^[27-28]、核裂变碎片质量分布^[29]、核聚变截面^[30]等。

在低能核物理领域，可使用机器学习表示原子核的变分波函数，解决“从头计算”方法中的维数灾难问题。还可使用贝叶斯神经网络确定核能量密度泛函中的参数及其不确定度，评估快中子的捕获过程、致密核物质状态方程以及中子星质量半径关系中的不确定性。

在高能核物理领域，可使用贝叶斯分析从高能重离子碰撞数据中提取夸克胶子等离子体QGP的性质和相图信息^[31]。还可使用生成对抗网络(GAN)和流模型(flow-based model)产生格点量子色动力学Lattice QCD的组态(configurations)，并使用深度神经网络表示QGP中重味夸克反夸克之间的相互作用势能V，可数值求解薛定谔方程。

在实验核物理、加速器设计和核数据领域，可使用聚类做反常探测，自动标注出现问题的谐振腔。还可使用强化学习^[32]优化加速器中多个组件的配合。

中子共振作为核物理学中的一个重要研究领域。使用机器学习算法，可以加速计算和提高计算精度，帮助研究人员更深入地了解中子共振现象和物理机制。

总之，机器学习为中子共振研究提供新的思路和方法，加速计算、提高计算精度、进行数据分析和拟合物理模型。随着机器学习算法的不断发展和优化，它将在中子共振研究领域中发挥越来越重要的作用。

本文将基于机器学习方法对核反应的共振现象展开研究。本文安排如下：在引言中，介绍了什么是机器学习和中子共振数据，以及机器学习在核物理中的应用。在第二章中，先介绍机器学习中最常用的深度神经网络(DNN)算法，然后引出DNN模型及频率原则(f-principle)。接着详细介绍所采用的机器学习算法和模型——PPSDNN模型，并给出算法流程。在第三章中，首先给出PPSDNN模型在玩具模型(toy model)中的测试结果。接着介绍PPSDNN在学习中子共振数据中的应用。最后，在第四章中，先对以上内容作出总结，接着介绍机器学习用于中子共振研究的潜在应用及其和R-Matrix模型相结合的可能性，给出展望。

第 2 章 平行相移深度神经网络模型

2.1 深度神经网络模型

2.1.1 深度神经网络算法简介

深度神经网络 (Deep Neural Network, DNN) 模型是由多个神经网络层组成的模型^[33]。在全连接神经网络中，每层由多个神经元组成，且每个神经元与上一层的所有神经元相连。DNN 可以自主从训练数据中学习和发现模式，并执行高级特征提取和分类任务。因此，在图像识别、自然语言处理和语音识别等领域广泛使用。

DNN 的每一层都是一种特征提取器，可以自动提取数据的不同层次的特征，例如，在图像识别中，低层神经网络可以提取边缘和纹理特征，中层神经网络可以提取形状和结构特征，而高层神经网络可以提取语义信息，实现更高级别的分类任务。

训练 DNN 通常采用反向传播算法^[34]，该算法使用损失函数的梯度更新每一层神经网络的权重和偏置。在 DNN 的训练过程中，使用一些技巧和方法，如批量归一化^[35]、残差连接和 Dropout 等^[36]，以提高模型的性能和稳健性。

DNN 是由多个线性函数和非线性激活函数复合形成的复合函数，于是可以把这种函数关系看成是 n 维空间到 m 维空间的非线性映射 ($m, n > 1$)。其中，线性映射关系为 $\Theta(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ，这种线性映射关系的具体表达形式为 $\Theta(x) = Wx + b$ ，其中， $W = (w_{ij}) \in \mathbb{R}^{n \times m}$ ， $b \in \mathbb{R}^m$ ，它们是神经网络的参数。非线性激活函数是一个一维空间到一维空间的映射，可表示为 $\sigma(u) = \mathbb{R} \rightarrow \mathbb{R}$ 。在 DNN 中，可以将其拓展到 n 维空间到 n 维空间的映射，即 $\sigma(u) = \mathbb{R}^n \rightarrow \mathbb{R}^n$ 。于是，一个 $L + 1$ 层的神经网络可以表示为：

$$\begin{aligned} T(\mathbf{x}) &= T^L(\mathbf{x}) \\ T^l(\mathbf{x}) &= [\Theta^l \circ \sigma] (T^{l-1}(\mathbf{x})), \quad l = 1, 2, \dots, L \end{aligned} \tag{2-1}$$

且 $T^0(x) = \Theta^0(x)$ 。通过这种递推方式，DNN 也可以更简洁地表示为

$$T(\mathbf{x}) = \Theta^L \circ \sigma \circ \Theta^{L-1} \circ \sigma \cdots \circ \Theta^1 \circ \sigma \circ \Theta^0(\mathbf{x}) \tag{2-2}$$

其中， $\Theta^l(x) = W^l x + b^l : \mathbb{R}^{n^{(l)}} \rightarrow \mathbb{R}^{n^{(l+1)}}$ 。这样的 DNN 模型一共有 L 个隐藏层，且第 l 个隐藏层有 $n^{(l)}$ 个神经元。

本文中只考虑单隐藏层和单输出层的 DNN。使用上文中的数学符号可以表示为， $\Theta^0(x) : \mathbb{R} \rightarrow \mathbb{R}^{n^{(1)}}$ ， $\Theta^L(\mathbf{x}) : \mathbb{R}^{n^{(L-1)}} \rightarrow \mathbb{R}$ 。也就是说， $T(x) : \mathbb{R} \rightarrow \mathbb{R}$ 。由于核数据的学习和拟合要保证数据的光滑性，一般选取激活函数为 $\sigma(u) = \tanh(u)$ 。

为了使 DNN 逼近目标函数，定义损失函数：

$$L(\mathbf{W}^0, \mathbf{b}^0, \mathbf{W}^1, \mathbf{b}^1, \dots, \mathbf{W}^L, \mathbf{b}^L) = \|f(\mathbf{x}) - T(\mathbf{x})\|^2 = \int_{-\infty}^{+\infty} |f(\mathbf{x}) - T(\mathbf{x})|^2 dx \quad (2-3)$$

并使用梯度下降法最小化该损失函数。同时，根据帕萨瓦尔等式^[37-38]，在频域空间上的损失函数等于其在时域空间上的损失函数。定义目标函数 $f(x)$ 的傅里叶变换及其逆变换：

$$\mathcal{F}[f](k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x) e^{-ikx} dx, \quad \mathcal{F}^{-1}[\hat{f}](x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \hat{f}(k) e^{ikx} dk \quad (2-4)$$

令 $D(k) = \mathcal{F}[T - f](k) = A(k) e^{i\varphi(k)}$, $L(k) = |D(k)|^2$, 根据帕萨瓦尔等式得：

$$L(\mathbf{W}^0, \mathbf{b}^0, \mathbf{W}^1, \mathbf{b}^1, \dots, \mathbf{W}^L, \mathbf{b}^L) = \int_{-\infty}^{+\infty} L(k) dk \quad (2-5)$$

在 DNN 的参数空间可以表示为

$$\theta = (\mathbf{W}_1^0, \mathbf{W}_2^0, \dots, \mathbf{W}_{n^1}^0, \mathbf{b}^0, \mathbf{W}_{11}^1, \dots, \mathbf{W}_{n^1 n^2}^1, \mathbf{b}_1^1, \dots, \mathbf{b}_{n^2}^1, \dots) \in \mathbb{R}^p \quad (2-6)$$

其中，DNN 参数个数为 $p = 2n^2 + (n^1 + 1) \times n^2 + (n^2 + 1) \times n^3 + \dots (n^L + 1)$

DNN 的训练过程可以看成函数的数值逼近过程，根据函数的万能逼近定理，一个带激活函数的单隐藏层 DNN 可以逼近任意函数^[39]。

总之，DNN 是一种由多个神经网络层组成的神经网络模型，可以自主从训练数据中学习和发现模式，以及进行高级的特征提取和分类任务。DNN 在图像识别、自然语言处理、语音识别等领域中得到广泛应用，是人工智能领域中非常重要的技术之一。

2.1.2 频率原则

频率原则 (f-principle)^[40] 是一个用傅里叶分析来研究 DNN 训练过程的方法。它表明 DNN 通常倾向于从低频到高频拟合目标函数，这与传统的数值迭代方法相反。这个原则揭示了 DNN 的一种隐式偏置，即它们更容易学习低频的特征，而对高频的特征不太敏感。这种对高频低频数据的敏感程度，可以用频域空间下的损失函数随参数的变化率 $\frac{\partial L(k)}{\partial \theta_j}$ 表示。

为了证明频率原则，考虑一个单隐藏层神经网络，激活函数是双曲正切。并且 DNN 是一输入量，一输出量。于是，DNN 可以写成

$$h(x) = \sum_{j=1}^m a_j \sigma(w_j x + b_j), \quad a_j, w_j, b_j \in \mathbb{R}, \quad (2-7)$$

其中, m 是神经网络的层数。其中激活函数为:

$$\sigma(x) = \tanh(x). \quad (2-8)$$

对该激活函数做傅里叶变换得 $\hat{\sigma}(k) = -\frac{i\pi}{\sinh(\pi k/2)}$ ^[41], 计算过程大致为

$$\begin{aligned} & \int_{-\infty}^{\infty} \tanh(x) e^{-ikx} dx = \\ &= \int_{-\infty}^{\infty} \frac{e^x - e^{-x}}{e^x + e^{-x}} e^{-ikx} dx \\ &= \int_{-\infty}^{\infty} e^{-ikx} dx - \int_{-\infty}^{\infty} \frac{2e^{-ikx}}{e^{2x} + 1} dx \\ &= -\frac{i\pi}{\sinh(\pi k/2)} \end{aligned} \quad (2-9)$$

计算详细过程见附录A。计算结果与文献结果一致。于是, 根据相移特性和尺度变换特性^[42]

$$q(\omega t + b) \xrightarrow{\mathcal{F}} \frac{1}{|\omega|} e^{i \frac{b}{\omega} \omega} Q\left(\frac{k}{\omega}\right) \quad (2-10)$$

其中, $Q(k)$ 是函数 $q(t)$ 的傅立叶变换。对式2-9使用变换特性2-10, 可以得到 $\sigma(w_j x + b_j)$ 的傅里叶变换形式为

$$\mathcal{F}[\sigma](k) := \widehat{\sigma(w_j \cdot + b_j)}(k) = \frac{2\pi i}{|w_j|} \exp\left(\frac{ib_j k}{w_j}\right) \frac{1}{\exp\left(-\frac{\pi k}{2w_j}\right) - \exp\left(\frac{\pi k}{2w_j}\right)}. \quad (2-11)$$

不难看出, 它随 k 的值按指数衰减, 这是由激活函数 \tanh 的光滑性导致的。设当 $\frac{\pi k}{2w_j} > 0$ 时, $\exp\left(\frac{\pi k}{2w_j}\right) \gg \exp\left(-\frac{\pi k}{2w_j}\right)$, DNN 模型的傅里叶变换可以写成

$$\begin{aligned} \hat{h}(k) &= \sum_{j=1}^m \frac{2\pi a_j i}{|w_j|} \exp\left(\frac{ib_j k}{w_j}\right) \frac{1}{\exp\left(-\frac{\pi k}{2w_j}\right) - \exp\left(\frac{\pi k}{2w_j}\right)} \\ &\approx -\sum_{j=1}^m \frac{2\pi a_j i}{|w_j|} \exp\left(\frac{ib_j k}{w_j}\right) \exp\left(-\frac{\pi k}{2w_j}\right) \end{aligned} \quad (2-12)$$

由于目标函数为 $f(x)$, 于是 DNN 在频域空间上的损失函数可以定义成 L , L 由如下公式给出

$$\begin{aligned} D(k) &\triangleq \hat{h}(k) - \hat{f}(k) \\ L(k) &= \frac{1}{2} |D(k)|^2 \\ L &= \int_{-\infty}^{+\infty} \frac{1}{2} (h(x) - f(x))^2 dx \end{aligned} \quad (2-13)$$

其中，可以令 $D(k) = A(k)e^{i\phi(k)}$ 。为了探究 L 在进行梯度下降时对频率的依赖关系，把 $L(k)$ 对其参数求偏导，为了半定量地对其中一个参数 a_j 求偏导得：

$$\begin{aligned}\frac{\partial L(k)}{\partial a_j} &= D(k) \frac{\partial D(k)}{\partial a_j} \\ &= D(k) \exp\left(-\frac{\pi k}{2\omega_j}\right) G(\theta)\end{aligned}\quad (2-14)$$

其中， θ 是神经网络的参数，即 $\{\omega_j, b_j, a_j\}$ 。通过严格的计算可以得到

$$\begin{aligned}\frac{\partial L(k)}{\partial a_j} &= \frac{2\pi}{w_j} \sin\left(\frac{b_j k}{w_j} - \phi(k)\right) E_0 \\ \frac{\partial L(k)}{\partial w_j} &= \left[\sin\left(\frac{b_j k}{w_j} - \phi(k)\right) \left(\frac{\pi^2 a_j k}{w_j^3} E_1 - \frac{2\pi a_j}{w_j^2} \right) \right. \\ &\quad \left. - \frac{2\pi a_j b_j k}{w_j^3} \cos\left(\frac{b_j k}{w_j} - \phi(k)\right) \right] E_0 \\ \frac{\partial L(k)}{\partial b_j} &= \frac{2\pi a_j b_j k}{w_j^2} \cos\left(\frac{b_j k}{w_j} - \phi(k)\right) E_0\end{aligned}\quad (2-15)$$

其中

$$\begin{aligned}E_0 &= \frac{\operatorname{sgn}(w_j) A(k)}{\exp\left(\frac{\pi k}{2w_j}\right) - \exp\left(-\frac{\pi k}{2w_j}\right)}, \\ E_1 &= \frac{\exp\left(\frac{\pi k}{2w_j}\right) + \exp\left(-\frac{\pi k}{2w_j}\right)}{\exp\left(\frac{\pi k}{2w_j}\right) - \exp\left(-\frac{\pi k}{2w_j}\right)}. \\ \frac{\partial L}{\partial \theta_{lj}} &= \int_{-\infty}^{+\infty} \frac{\partial L(k)}{\partial \theta_{lj}} dk.\end{aligned}\quad (2-16)$$

其中，将 $L(k)$ 对参数 θ_{lj} 求偏导取绝对值

$$\left| \frac{\partial L(k)}{\partial \theta_{lj}} \right| \approx A(k) \exp(-|\pi k/2w_j|) F_{lj}(\theta_j, k), \quad (2-17)$$

其中， $\theta_j \triangleq \{w_j, b_j, a_j\}$, $\theta_{lj} \in \theta_j$, $F_{lj}(\theta_j, k)$ 是和 θ_j, k 有关的函数, $A(k)$ 与 $|\hat{h}(k) - \hat{f}(k)|$ 有关。这和前面推导出来的2-14从式很类似。从式中可以看出，当 $A(k) > 0$ 时，如果 w_i 是个小量，那么 $\exp(-|\pi k/2w_j|)$ 将占主导地位，即低频占主导地位。如果 $A(k) \approx 0$ ，那么频率空间的参数对梯度的贡献就很少。

这个特例针对的是 \tanh 激活函数，当我们选取不同的激活函数时，这些激活函数的光滑性和正则性根据 f-principle 也会影响梯度下降。通过以上对一个特例的分析，可以总结

出如下定理, 定义

$$W = (w_1, w_2, \dots, w_m)^T \in \mathbb{R}^m \quad (2-18)$$

有

定理 1 给定一个函数 $f(\vec{x}) \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$ (为了使函数存在傅里叶变换且损失函数可以解析表达) 和一个DNN模型 $T(\vec{x})$, 通过公式 $\Theta^0(\vec{x}) = \epsilon \vec{W}^0 \vec{x} + \vec{b}$ 改变 $T(\vec{x})$, 其中 $\epsilon > 0$ 是一个无穷小的参数. 假设对所有的 $0 \leq l \leq L-1$, $1 \leq r \leq n^l$ 有 $\vec{W}_{rs}^l \neq 0$ 且 $1 \leq s \leq n^{l+1}$. 对任意频率 k_1 和 k_2 使得 $|\mathcal{F}[f](k_1)| > 0$, $|\mathcal{F}[f](k_2)| > 0$ 且 $|k_2| > |k_1| > 0$, 存在 $\delta > 0$ 使得当 $\epsilon < \delta$ 时, 有

$$\frac{\mu\left(\{\theta \mid \left|\frac{\partial L(k_1)}{\partial \theta_j}\right| > \left|\frac{\partial L(k_2)}{\partial \theta_j}\right|\} \cap B_\delta\right)}{\mu(B_\delta)} > 1 - C \exp(-c/\delta)$$

对所有 j 都成立, 其中 $B_\delta = \{\vec{x} \in \mathbb{R}^p \mid |\vec{x}| < \delta\}$, $\mu(\cdot)$ 是 Lebesgue 测度且常数 $c > 0$, $C > 0$.

此即 PPSDNN 的基础, 当 $\delta \rightarrow 0$ 时, 在基于梯度训练的 DNN 中, 对于损失函数 $L = \int_{-\infty}^{\infty} L(k) dk$, 低频的梯度会大于高频的梯度, 即低频数据的收敛速度要比高频数据的收敛速度快。

下面构造一个有噪声的正弦函数, 将其生成的数组作为 DNN 的训练集, 训练结果如图2-1所示。图 (a) 中的训练数据是正弦函数加上噪声的数据, 使用 DNN 进行训练, 并在测试集损失函数变大时 (如图 (b) 所示) 提前停止训练, 发现 DNN 可以很好的学习到正弦函数的大致趋势, 但因为噪声属于高频数据, 所以无法很好的学习噪声。

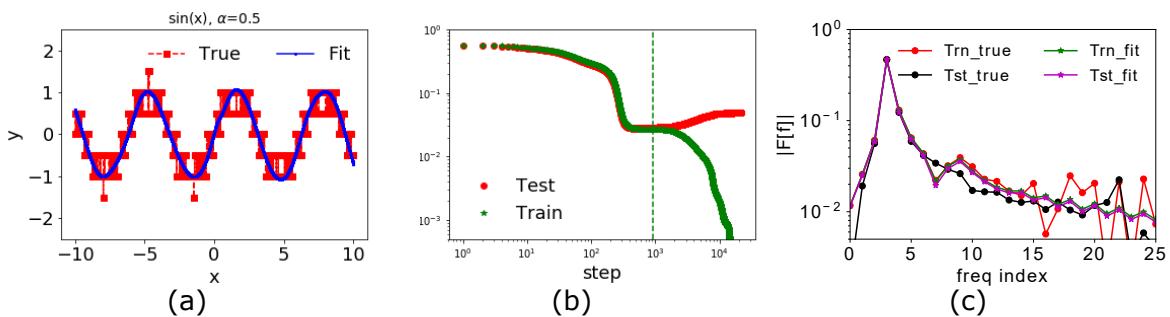


图 2-1 图 (a): 使用 DNN 训练有噪声的正弦函数。图 (b): 损失函数分别在训练集测试集上的表现。图 (c): 训练集测试集在频率空间上的表现。

下面再举个更定量的例子, 用来说明当正弦函数与其他高频函数叠加时, DNN 会优先学习到更低频的部分, 即正弦函数部分。从图2-2可以看出, 当 DNN 学习由函数 $\sin(x) + \sin(5x)$ 生成的数据时, 当 epoch 等于 18000, 神经网络可以学习处这个函数的大致趋势, 并能较好的学习到 $\sin(x)$ 的部分。只有当 epoch 继续增加, DNN 才能学好这个函数各种高频的细节。

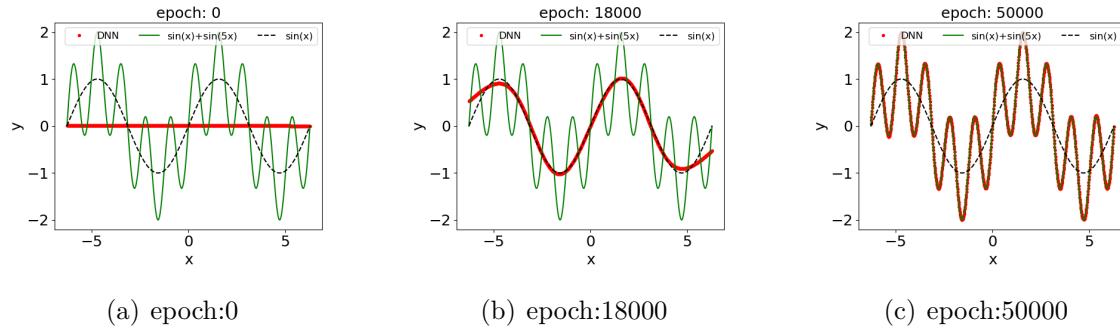


图 2-2 DNN 学习函数 $\sin(x) + \sin(5x)$, 学习结果随 epoch 的变化, 并与 $\sin(x)$ 进行比较

从上面两个例子可以看出 DNN 确实可以先学习到低频部分，再学习到高频部分。那么，在学习中子共振的数据时，是不是只要我的 epoch 足够大，就一定能学好呢？答案显然不是，否则也没有发明 PPSDNN 算法的必要了。下面使用一个具体例子说明当频率稍微大一点时，传统 DNN 要学习好他们就必须付出极高的代价。在数值分析中，当插值函数的次数比较大时，会出现龙格现象。当使用 DNN 训练这类数据时，当参数数量和插值次数相仿时，就需要花费无穷大的 epoch 去训练他们，如图2-3所示。有了 f-principle 的指导，在 DNN 中就不用再担龙格现象，并可以通过训练提前停止来避免过拟合。

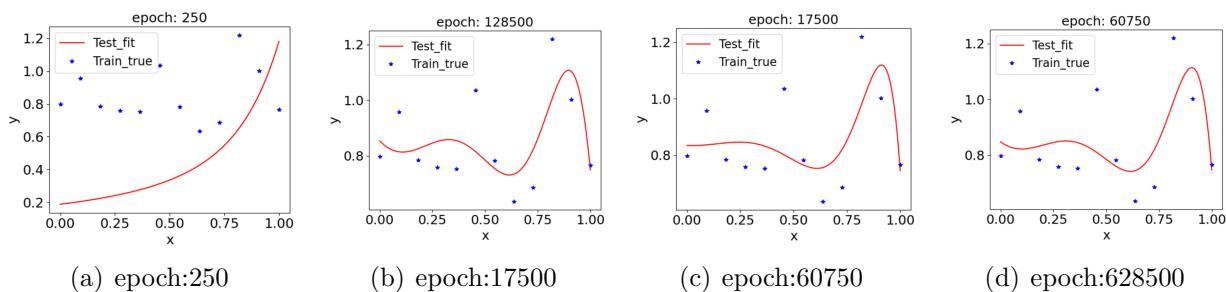


图 2-3 使用梯度下降学习 11 阶插值函数

f -principle 可以帮助我们理解 DNN 在不同问题上的优势和局限性，以及如何利用它们来设计更有效的算法。例如，在偏微分方程的数值求解中^[43]，可以结合 DNN 和传统迭代方法，先利用 DNN 快速收敛低频部分，再用传统迭代方法细化高频部分。

在中子共振核反应中，由于共振峰的存在，并根据第一章中的分析，在有限的高频共振能区中含有 900 多个共振峰，DNN 并不能直接很好地应用于中子共振核反应的研究中。

使用 DNN 训练 $^{235}\text{U}(n, f)$ 评价库数据, 如图2-4所示, 发现 DNN 只能较好的学习 1/v 能区和不可分辨能区的数据, 对共振区的数据只能学习大致趋势, 无法学到共振细节。为了验证 DNN 的训练效果不会随着 epoch 显著增加, 可以画出损失函数的变化趋势, 如图2-5所示, 可以看出 loss 值在一定 epoch 之后不会再发生显著下降, 即神经网络的学习效果不会再变好了。

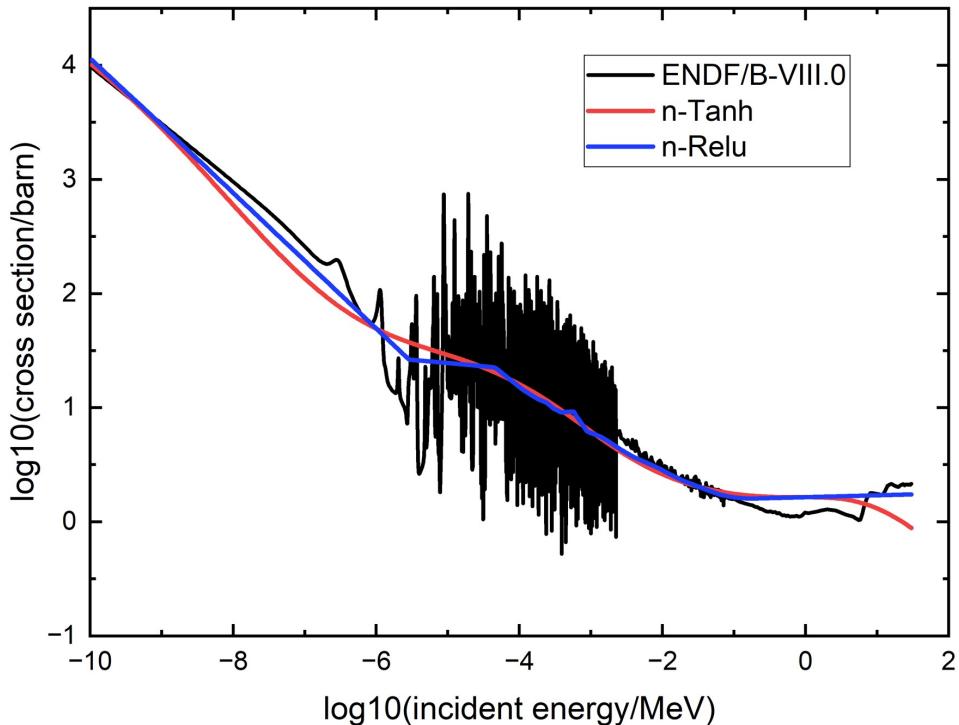


图 2-4 使用 tanh 和 relu 激活函数分别对 ENDF/B-VIII.0 中 $^{235}\text{U}(n, f)$ 反应截面进行学习

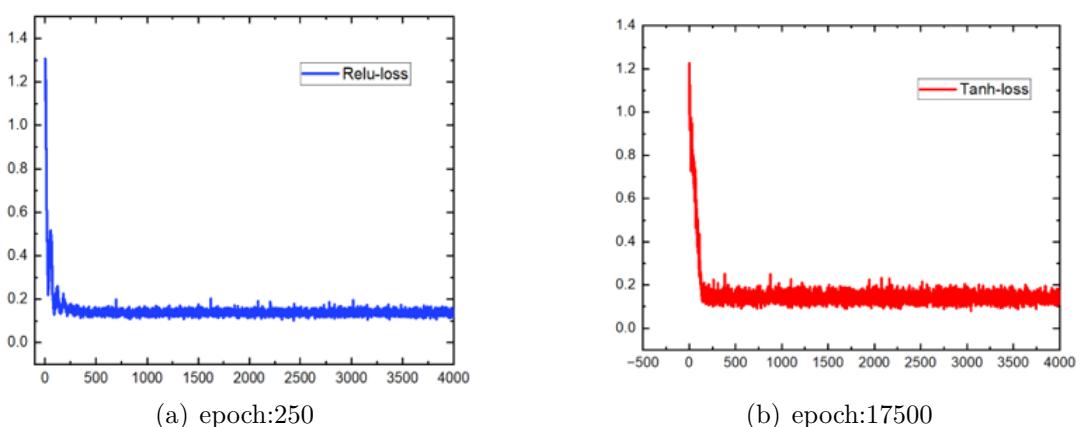


图 2-5 不同激活函数下 DNN 的损失函数

为解决 DNN 难以学习高频数据这个问题。本文中采用平行相移深度神经网络模型^[44] (Parallel Phase Shift Deep Neural Network,PPSDNN)。

2.2 平行相移深度神经网络模型

2.2.1 模型及算法简介

PPSDNN 可以在训练过程中实现对函数的收敛，即同时逼近目标函数的低频和高频成分。PPSDNN 利用了一个事实，即常见的 DNN 在训练时先收敛于低频范围，因此，它构造了一系列大小适中的 DNN，并通过在频域上对训练数据进行相移操作，使得每个 DNN 能够逼近目标函数在特定范围内的高频内容。由于相移的作用，每个 DNN 都能够像在低频范围一样快速收敛。因此，PPSDNN 能够将高频学习转化为低频学习，从而实现对函数的全部频域空间均匀学习和自适应训练。^[45] 算法的大致思想是利用神经网络在训练过程中对低频成分的快速收敛，通过相移技术将目标函数的高频成分转换为低频成分，然后用一系列并行的神经网络来近似不同频率范围的函数，最后将这些神经网络的输出相加，从而得到目标函数的近似。

PPSDNN 算法的基本思路如下：首先，将 $f(x)$ 分解成各个频段函数的叠加。

$$f(x) = \sum_{j=0}^m f_j(x), \quad (2-19)$$

其中，

$$f_j(x) = \mathcal{F}^{-1}[\hat{f}_j](x).$$

谱区间为 $[\omega_{j-1}, \omega_{j+1}]$ 设对于低频项，频域 $k \in [-K_0, k_0]$ ，普通 DNN 方法可以给出很好的近似。对于高频项，通过相移 (phase shift) $r_j^{\text{shift}}(x_i) = e^{i\omega_j x_i} r_j(x_i)$ ，把高频数据转换成 $[-K_0, k_0]$ 上的低频数据，进而使用 DNN 训练。将训练得到的结果再通过 phase shift 还原到高频部分。最后将训练的到的所有结果加起来就是总的训练结果。

PPSDNN 算法可按如下几个模块逐步实现：

(1) 傅里叶变换和与尼奎斯特采样定理^[46]

(2) 单位分解定理 (POU)

(3) B-样条函数

(4) 相移在频域空间中的实现

2.2.2 傅里叶变换与尼奎斯特采样定理

傅里叶变换是把一个函数或一些离散数据从时域空间转换到频域空间，傅里叶逆变换则是从频域空间转换到时域空间^[47]。先考虑连续傅里叶变换（CFT）和连续傅里叶逆变换（ICFT），其公式为

$$\mathcal{F}[g(x)](\xi) = \int_{-\infty}^{\infty} g(x) e^{-2\pi i \xi x} dx, \quad \mathcal{F}^{-1}[\hat{g}(\xi)](x) = \int_{-\infty}^{\infty} \hat{g}(\xi) e^{2\pi i \xi x} d\xi \quad (2-20)$$

再考虑离散时域空间上的傅里叶变换（DTFT）及其逆变换

$$\begin{aligned} \mathcal{F}_{DTFT,\Delta}[g(x)](\xi) &= \hat{g}_{DTFT}(\xi) = \sum_{j=-\infty}^{\infty} g(j\Delta) e^{-2\pi i \xi j \Delta} \\ \mathcal{F}_{DTFT,\Delta}^{-1}[\hat{g}(\xi)](j) &= \Delta \cdot \int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} \hat{g}_{DTFT}(\xi) e^{2\pi i \xi j \Delta} d\xi \end{aligned} \quad (2-21)$$

由于客观世界的数据往往是连续分布的，而我们认识的世界往往是通过离散的采样得到的数据。于是我们要尽可能在最大限度保留真实世界信息的情况下把 CFT 向 DTFT 转换。下面是转换公式的推导

$$\begin{aligned} \mathcal{F}_{DTFT,\Delta}[g(x)](\xi) &= \sum_{j=-\infty}^{\infty} g(j\Delta) e^{-2\pi i \xi j \Delta} \\ &= \sum_{j=-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{g}(\xi') e^{2\pi i \xi' j \Delta} d\xi' e^{-2\pi i \xi j \Delta} \\ &= \sum_{j=-\infty}^{\infty} \sum_{M=-\infty}^{\infty} \int_{\frac{M}{\Delta} - \frac{1}{2\Delta}}^{\frac{M}{\Delta} + \frac{1}{2\Delta}} \hat{g}(\xi') e^{2\pi i \xi' j \Delta} d\xi' e^{-2\pi i \xi j \Delta} \\ &= \sum_{j=-\infty}^{\infty} \sum_{M=-\infty}^{\infty} \int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} \hat{g}\left(\xi' + \frac{M}{\Delta}\right) e^{2\pi i \xi' j \Delta} d\xi' e^{-2\pi i \xi j \Delta} \end{aligned} \quad (2-22)$$

定义 $\hat{h}(\xi') = \hat{g}\left(\xi' + \frac{M}{\Delta}\right)$ ，得

$$\mathcal{F}_{DTFT,\Delta}[g(x)](\xi) = \frac{1}{\Delta} \sum_{M=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \left(\Delta \int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} \hat{h}(\xi') e^{2\pi i \xi' j \Delta} d\xi' \right) e^{-2\pi i \xi j \Delta} \quad (2-23)$$

通过逆 DTFT 得

$$\mathcal{F}_{DTFT,\Delta}[g(x)](\xi) = \frac{1}{\Delta} \sum_{M=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} h(j\Delta) e^{-2\pi i \xi j \Delta}. \quad (2-24)$$

再通过 DTFT 得

$$\begin{aligned}\mathcal{F}_{DTFT,\Delta}[g(x)](\xi) &= \frac{1}{\Delta} \sum_{M=-\infty}^{\infty} \bar{h}(\xi) \\ &= \frac{1}{\Delta} \sum_{M=-\infty}^{\infty} \hat{g}\left(\xi + \frac{M}{\Delta}\right).\end{aligned}\quad (2-25)$$

在上式中，左边是离散时间中的采样，右边代表真是信号。其中，左边的 $\xi \in [-\frac{1}{2\Delta}, \frac{1}{2\Delta}]$ 。于是，可以得到尼奎斯特采样定理^[48]。

定理 2 一个有限带宽的连续信号可以被采样并且可以从它的采样中完美重构的条件为：采样频率大于真实频率的两倍。

有了尼奎斯特采样定理，我们就能知道至少需要多少实验数据点才能学好中子共振截面的数据。

2.2.3 单位分解定理

在数学上，任意流形都能够表示成足够大的 n 维空间 \mathbb{R}^n 中某个曲面形式，而这可以用单位分解定理 (Partition of Unity, POU)^[49] 阐明。

定理 3 设 M 是 $C^{(k)}$ 光滑类的流形， X 是 M 的子集。如果函数组 $E = \{e_\alpha, \alpha \in A\}$ 由 $e_\alpha = C^{(k)}(M, \mathbb{R})$ 组成，且满足：

在 X 上 $\sum_{e_\alpha \in E} e_\alpha(x) \equiv 1$ 。

那么，我们就说该函数组是集合 X 上的 k 阶光滑单位分解。

在 PPSDNN 中，我们首先要将一个含有高频部分的目标函数 $f(x)$ 通过傅里叶变换高频谱函数 $\hat{f}(k)$ ，再通过 phase shift 将频谱转换到 $[-K_0, K_0]$ 的范围。下面考虑 $\hat{f}(k)$ 上的支撑集，因为 $\text{supp } \hat{f}(k) = \{k \in \mathbb{R} | \hat{f}(k) \neq 0\}$ ，所以对给定的 Δk ，可以设 $\hat{f}(k)$ 上的支撑集为 $\text{supp } \hat{f}(k) \subset [-m\Delta k, m\Delta k]$ 。再将区间 $[-m\Delta k, m\Delta k]$ 格点化，令 $\omega_j = j\Delta k$ ，其中， $j = -m, \dots, -1, 0, 1, \dots, m$ 。下面考虑在区间 $[-m\Delta k, m\Delta k]$ 上进行单位分解

$$1 = \sum_{j=-m}^{m-1} \phi_j(k), \quad k \in [-m\Delta k, m\Delta k]. \quad (2-26)$$

下面举几个满足单位分解定理的例子。满足单位分解最简单的函数为 $\phi_j(k)$ ：

$$\phi_j(k) = \phi\left(\frac{k - \omega_j}{\Delta k}\right) = \phi\left(\frac{k}{\Delta k} - j\right), \quad (2-27)$$

其中, $\phi(k)$ 是区间 $[-1, 1]$ 上的特征函数, 特征函数的形式为

$$\phi(k) = \chi_{[0,1]}(k) = \begin{cases} 1, & k \in [0, 1) \\ 0, & \text{otherwise} \end{cases} \quad (2-28)$$

则对 $\phi_j(k)$ 而言, $0 \leq \frac{k-\omega_j}{\Delta k} < 1$ 时, 即 $j\Delta k \leq k < (j+1)\Delta k$ 时为 1, 其余区间为 0, 即

$$\phi_j(k) = \begin{cases} 1, & k \in [j\Delta k, (j+1)\Delta k) \\ 0, & \text{otherwise} \end{cases} \quad (2-29)$$

这样的 $2m$ 个函数 $\phi_{-m}(k), \phi_{-m+1}(k), \dots, \phi_0(k), \dots, \phi_{m-2}(k), \phi_{m-1}(k)$, 对应的值为 1 的非零区间为 $[-m\Delta k, (-m+1)\Delta k], [-(m+1)\Delta k, (-m+2)\Delta k], \dots, [(m-2)\Delta k, (m-1)\Delta k], [(m-2)\Delta k, (m-1)\Delta k]$ 。因此, 当 $-m\Delta k \leq k \leq m\Delta k$ 时, 必落在某个区间, 使某个 $\phi_j(k) = 1$, 其余的 $\phi_j(k) = 0$, 即 $\sum_{j=-m}^{m-1} \phi_j(k) = 1$ 。因此 $\phi_j(k)$ 满足单位分解定理 (POU)。下面计算 $\phi(k) = \chi_{[0,1]}(k)$ 的逆傅里叶变换, 我们用符号 \vee 表示。

$$\phi^\vee(x) = \frac{1}{\sqrt{2\pi}} \int_0^1 dk e^{ikx} = \frac{e^{\frac{ix}{2}(e^{\frac{ix}{2}} - e^{-\frac{ix}{2}})}}{ix\sqrt{2\pi}} = \frac{1}{\sqrt{2\pi}} e^{i\frac{x}{2}} \frac{\sin x/2}{x/2}. \quad (2-30)$$

虽然特征函数可以构造出满足单位分解的函数, 但特征函数是分段函数, 且不够光滑。下面我们使用 B-样条函数构造满足单位分解的函数, 这样的函数满足连续且光滑的条件。

2.2.4 B-样条函数

B-样条函数^[50]是基于特征函数迭代递推得到的, 它和特征函数相比具有更连续更光滑的性质。采用如下卷积迭代的方式可以定义 B-样条函数

$$\begin{cases} B_1(k) = \chi_{[0,1]}(k), \\ B_m(k) = B_{m-1}(k) * \chi_{[0,1]}(k), m = 2, \dots . \end{cases} \quad (2-31)$$

这种递推关系式也可以写成

$$B_m(k) = \underbrace{B_1(k) * B_1(k) * \dots * B_1(k)}_{m \uparrow} \quad (2-32)$$

其中 * 代表卷积, 即 $(f * g)(t) \stackrel{\text{def}}{=} \int_{\mathbb{R}^n} f(\tau)g(t-\tau)d\tau$ 可以用归纳法证明

$$\text{supp } B_m(k) = [0, m] \quad (2-33)$$

下面，分别举 $m=2$ 和 $m=3$ 的例子来说明 B-样条函数的支撑集以及它的连续性和光滑性。

当 $m=2$ 时

$$B_2(k) = B_1(k) \star B_1(k) = \int_{-\infty}^{\infty} B_1(k - \lambda) B_1(\lambda) d\lambda = \int_0^1 B_1(k - \lambda) d\lambda$$

其中

$$\begin{cases} 0 \leq \lambda < 1, \\ 0 \leq k - \lambda < 1 \end{cases} \Rightarrow 0 \leq k < 2$$

于是不难看出

$$supp B_2(k) = [0, 2] \quad (2-34)$$

通过积分，不难得出

$$B_2(k) = \begin{cases} k, & 0 \leq k < 1 \\ 2 - k, & 1 \leq k < 2 \\ 0, & otherwise \end{cases} \quad (2-35)$$

可以看出， $B_2(k)$ 虽然连续，但不光滑。如图2-6

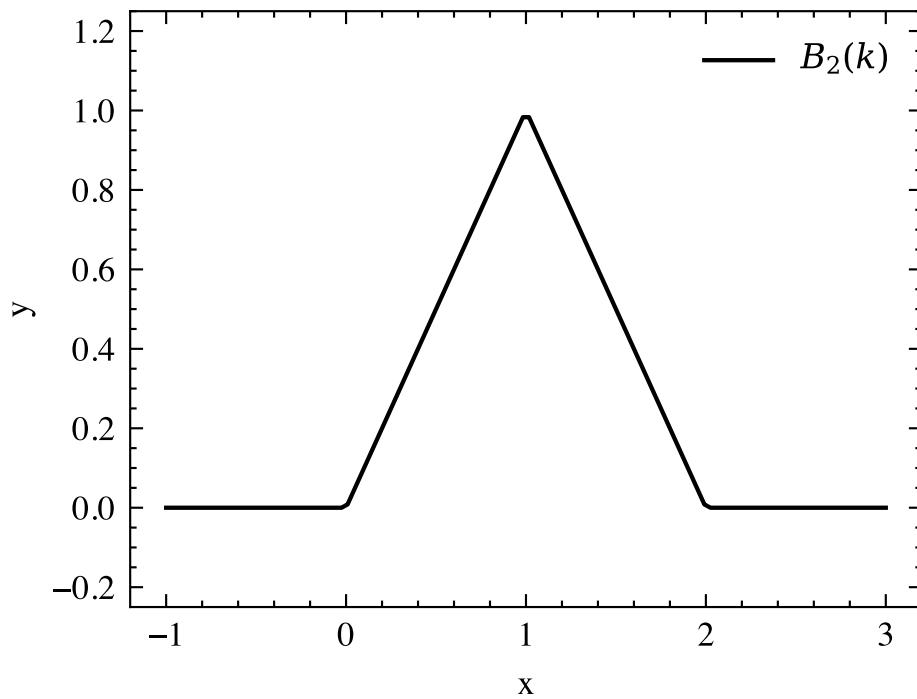


图 2-6 2 阶 B 样条函数

当 $m=3$ 时，同理可得

$$B_3(k) = B_2(k) \star B_1(k) = \int_{-\infty}^{\infty} B_2(k - \lambda) B_1(\lambda) d\lambda = \int_0^1 B_2(k - \lambda) d\lambda$$

且不难看出

$$\text{supp}B_3(k) = [0, 3] \quad (2-36)$$

通过积分，不难得出

$$B_3(k) = \begin{cases} \frac{1}{2}k^2, & 0 \leq k < 1 \\ \frac{1}{2} + (k-1) - (k-1)^2, & 1 \leq k < 2 \\ \frac{1}{2} - (k-2) + \frac{1}{2}(k-2)^2, & 2 \leq k < 3 \\ 0, & \text{otherwise} \end{cases} \quad (2-37)$$

此时，在 $k=1$ 和 $k=2$ 处， $B_3(k)$ 不仅连续，而且光滑（一阶导数连续）。如图2-7

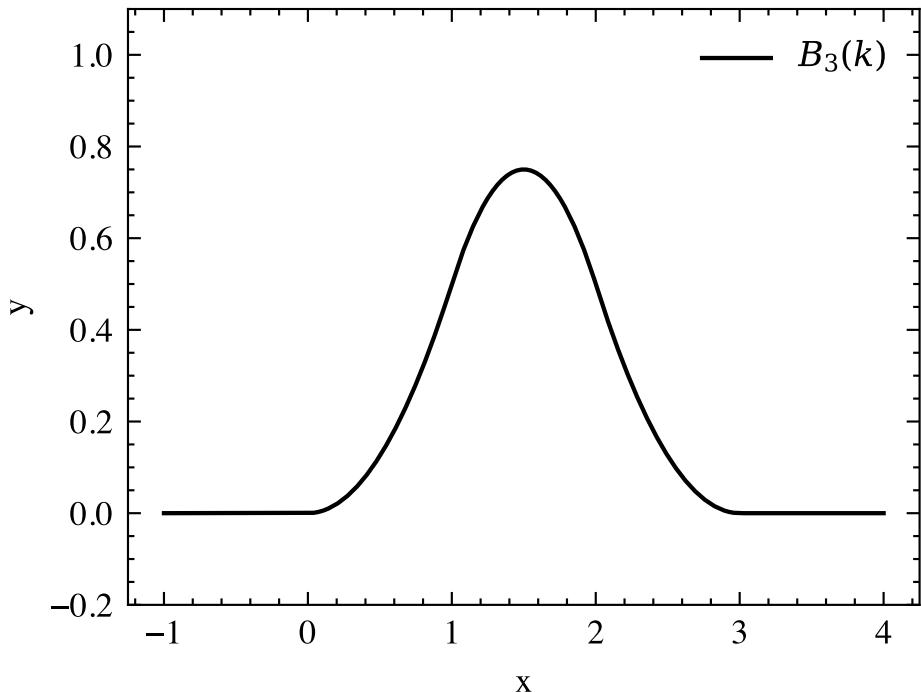


图 2-7 3 阶 B 样条函数

于是，可以整理出 B-样条函数的诸多性质，如

$$(1) \text{ supp}B_m(k) = [0, m]$$

证明：采用归纳法。

在上文中，我们通过计算 $m=1, m=2, m=3$ 给出了具体几个常用特例的证明。下面，采用数学归纳法给出普式的 m 的证明。

设 $\text{supp}B_m(k) = [0, m]$ ，则

$$B_{m+1}(k) = B_m(k) \star B_1(k) = \int_{-\infty}^{\infty} B_m(k-\lambda) B_1(\lambda) d\lambda = \int_0^1 B_m(k-\lambda) d\lambda$$

其中

$$\begin{cases} 0 \leq \lambda < 1, \\ 0 \leq k - \lambda < m \end{cases} \Rightarrow 0 \leq k < m + 1$$

于是, $\text{supp } B_m(k) = [0, m]$, 证毕。

(2) $B_m(k)$ 的逆傅里叶变换

通过卷积定理:

$$\begin{cases} \mathcal{F}[f(x) * g(x)] = \hat{f}(k)\hat{g}(k) \\ \mathcal{F}[f(x)g(x)] = \hat{f}(k) * \hat{g}(k) \Rightarrow f(x)g(x) = \mathcal{F}^{-1}[\hat{f}(k) * \hat{g}(k)] \end{cases} \quad (2-38)$$

其中, 为了符号书写方便, 把傅里叶逆变换定义为 $\mathcal{F}[\cdot]$ 。

这样, $B_m(k)$ 的逆傅里叶变换, 记为 $B_m^\vee(k)$, 有

$$\begin{aligned} B_m^\vee(x) &= (B_1^\vee(x))^m = \left(\frac{1}{\sqrt{2\pi}} e^{i\frac{x}{2}} \frac{\sin \frac{x}{2}}{\frac{x}{2}} \right)^m \\ &= \frac{e^{i\frac{mx}{2}}}{(2\pi)^{m/2}} \left(\frac{\sin \frac{x}{2}}{\frac{x}{2}} \right)^m. \end{aligned} \quad (2-39)$$

下面将在“相移在频域空间中的实现”部分介绍如何使用 B-样条函数构造满足单位分解的函数。

2.2.5 相移在频域空间中的实现

我们以四次 B-样条函数为例, 构造满足单位分解的函数。首先, 令 $\phi(k) = B_4(k+2)$, 接着构造

$$\phi_j(k) = \phi(k/\Delta k - j) = B_4\left(\frac{k}{\Delta k} + 2 - j\right) \quad (2-40)$$

可以证明, 该函数满足单位分解。将 $\phi_j(k)$ 做傅里叶逆变换, 得到频率选择函数 (frequency selection kernel), 为了得到 $\phi_j^\vee(x)$, 我们先求 $\phi^\vee(x)$, 得

$$\begin{aligned} \phi^\vee(x) &= \frac{1}{\sqrt{2\pi}} \int_{\mathcal{R}} B_4(k+2) e^{ikx} dk \\ &\stackrel{k+2=k'}{=} \frac{1}{\sqrt{2\pi}} \int_{\mathcal{R}} dk' B_4(k') e^{ik'x} e^{-2ix} \\ &= B_4^\vee(x) e^{-2ix} = \frac{1}{(2\pi)^2} \left(\frac{\sin x/2}{x/2} \right)^4 \end{aligned} \quad (2-41)$$

下面计算 $\phi_j(k)$ 的傅里叶逆变换，的

$$\begin{aligned}\phi_j^\vee(x) &= \frac{1}{\sqrt{2\pi}} \int_{\mathcal{R}} dk \phi\left(\frac{k}{\Delta k} - j\right) e^{ikx} \\ &\stackrel{k'=k/\Delta k}{=} \frac{1}{\sqrt{2\pi}} \int_{\mathcal{R}} (\Delta k) dk' \phi(k' - j) e^{ik'\Delta kx} \\ &\stackrel{k''=k'-j}{=} \frac{1}{\sqrt{2\pi}} \int_{\mathcal{R}} (\Delta k) dk'' \phi(k'') e^{i(k''+j)\Delta kx} \\ &= \Delta k e^{-ij\Delta kx} \phi^\vee(\Delta kx) = K_0 e^{-ijK_0 x} \phi^\vee(K_0 x),\end{aligned}\quad (2-42)$$

下面求 $\phi_j(k)$ 的支撑集。由前面的定义和结论，显然，有 $\text{supp} \phi_j(k) = \text{supp} \mathcal{F}[\phi_j^\vee(x)]$ 。于是，根据 $0 \leq \frac{k}{\Delta k} + 2 - j < 4$ 得， $\omega_{j-2} \leq k < \omega_{j+2}$ ，即

$$\text{supp} \mathcal{F}[\phi_j^\vee(x)] \subset [\omega_{j-2}, \omega_{j+2}]. \quad (2-43)$$

同样，可以证明 B-样条函数满足单位分解性质。前面假设了低频空间为 $k \in [-m\Delta k, \Delta k]$ ，通过移项得 $-m \leq \frac{k}{\Delta k} \leq m$ 。又 $0 \leq \frac{k}{\Delta k} + 2 - j < 4$ ，移项得 $-2 + \frac{k}{\Delta k} < j \leq 2 + \frac{k}{\Delta k}$ 。不难得出

$$-m - 2 < j \leq m + 2$$

因此， j 实际的范围是 $j = -m - 1, -m, \dots, 0, \dots, m + 1, m + 2$ 。因此，可以证明 $\phi_j(k)$ 满足单位分解。

$$\sum_{j=-m-1}^{m+2} \phi_j(k) = 1, \quad \forall k \in [-m\Delta k, m\Delta k]. \quad (2-44)$$

在实际我们选择 POU 函数时，对于 $\sum_j \phi_j(k) = 1$ ， j 的具体范围取决于：

1) k 的范围；2) 样条函数的选取；3) 样条函数的平移量。

然后，在 POU 的基础上，我们对想要分解的函数 $f(x)$ 做傅里叶展开（傅里叶逆变换），即

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathcal{R}} dk \hat{f}(k) e^{ikx} \quad (2-45)$$

其中， $\hat{f}(k)$ 可以利用 POU 的性质分解为

$$\hat{f}(k) = \sum_{j=-m}^{m-1} \phi_j(k) \hat{f}_j(k) \triangleq \sum_{j=-m}^{m-1} \hat{f}_j(k). \quad (2-46)$$

相应的，把2-46式代入2-45式，可以在时域空间里将 $f(x)$ 做分解

$$f(x) = \sum_{j=0}^m f_j(x), \quad (2-47)$$

其中，

$$f_j(x) = \mathcal{F}^{-1}[\hat{f}_j](x).$$

根据 POU，

$$f_j(x) = \frac{1}{\sqrt{2\pi}} \int_{\omega_j}^{\omega_{j+1}} dk \hat{f}_j(k) e^{ikx} = \frac{1}{\sqrt{2\pi}} \int_{\omega_j}^{\omega_{j+1}} dk \phi_j(k) \hat{f}(k) e^{ikx} = \mathcal{F}^{-1}[\phi_j \hat{f}]$$

由卷积定理， $\mathcal{F}[f(x) * g(x)] = \hat{f}(k)\hat{g}(k)$ ，即 $f(x) * g(x) = \mathcal{F}^{-1}[\hat{f}(k)\hat{g}(k)]$ 。所以，

$$f_j(x) = \phi_j^\vee(x) * f(x) = \int_{-\infty}^{\infty} \phi_j^\vee(x-s) f(s) ds \quad (2-48)$$

从上式可以看出， $\phi_j^\vee(x)$ 通过卷积作用到 $f(x)$ 上得到频率空间为 $[\omega_j, \omega_{j+1}] = [j\Delta k, (j+1)\Delta] = [jK_0, (j+1)K_0]$ 的分量。 $f_j(x)$ 在此称为频率选择函数 (frequency selection kernel)。

基于频率选择函数，就可以实现平行相移。设当函数 $f(x)$ 通过傅里叶变换转换到频域空间，且 $|k| < K_0$ 时，函数处于低频，DNN 可以很好的训练 $f(x)$ 的低频量。对 $f(x)$ 的高频量 $f_j(x)$ ，我们可以通过乘一个相移因子，将其转换到 $|k| < K_0$ 的区域，具体形式为：

$$f_j^{\text{shift}}(x) = e^{i\omega_j x} f_j(x) \quad (2-49)$$

下面分析 $f_j^{\text{shift}}(x)$ 的频率范围，这里的推导需要用到 δ 函数的性质。

$$f_j^{\text{shift}}(k) = \hat{f}_j(k - \omega_j) \quad (2-50)$$

如果 $f_j(k)$ 的支撑集为 $[\omega_j, \omega_{j+1}]$ ，则 $\text{supp}[f_j^{\text{shift}}(k)] \subset [0, K_0]$ 。于是，我们通过 phase shift 将 $f(x)$ 的高频部分转换到了低频部分。于是可以将低频部分使用 DNN 去训练。下面介绍 PPSDNN 的算法流程。

2.3 算法流程

首先，给定一个要用来训练的目标函数 $f(x)$ 。我们使用 $\text{DNNT}_*(x, \theta^{(n_0)}), x \in R$, 去逼近它。

$$T_*(x, \theta^{(n_0)}) = f_{\text{DNN}}(x) \approx f(x), \quad (2-51)$$

$f_{\text{DNN}}(x)$ 和 $f(x)$ 各自对应的低频部分可以拟合的很好，但高频部分仍然会有差距。于是定义余项 $r(x) = f(x) - f_{\text{DNN}}(x)$, 并规定低频区域

$$|\mathcal{F}[r](k)| \ll 1 \text{ for } |k| < K_0, \quad (2-52)$$

这样，才能确保

$$\mathcal{F}[f](k) \approx \mathcal{F}[f_{\text{DNN}}(x)](k), \quad k \in (-K_0, K_0). \quad (2-53)$$

如果精确知道 $r(x)$, 则 $f(x) = r(x) + f_{\text{DNN}}(x)$ 。为尽量精确求得 $r(x)$, 首先将 $r(x)$ 分解,

$$r(x) = \sum_{j=0}^m r_j(x), \quad (2-54)$$

其中

$$r_j(x) = \mathcal{F}^{-1}[\hat{r}_j(k)] = \mathcal{F}^{-1}[\phi_j(k)\hat{r}(k)](x). \quad (2-55)$$

训练数据 $\{r_j(x_i)\}_{i=1}^N$ 可以由2-55式计算得到，在时域空间上，这些离散数据点可以通过下面的卷积公式得到

$$\begin{aligned} r_j(x_i) &= \phi_j^\vee * r(x_i) = \int_{-\infty}^{\infty} \phi_j^\vee(x_i - s)r(s)ds \\ &\approx \frac{2\delta}{N_s} \sum_{x_s \in (x_i - \delta, x_i + \delta)} \phi_j^\vee(x_i - x_s)r(x_s), \end{aligned} \quad (2-56)$$

取相应的频率选择核 $r_j(x)$, 可以将 $\phi_j(x)$ 的频率范围，即支撑集调整到 $[\omega_{j-1}, \omega_{j+1}]$ 。

接下来对 $r_j(x)$ 作 phase shift。

$$r_j^{\text{shift}}(x) = e^{i\omega_j x} r_j(x) = \mathcal{F}^{-1}[\hat{r}_j(k - \omega_j)](x), \quad (2-57)$$

则 $r_j^{\text{shift}}(x)$ 的支撑集在 $[-K_0, K_0]$ 之间。所以 $r_j^{\text{shift}}(x)$ 属于低频部分，所以它可以由 DNN 相对精确且快速地得到。即

$$T_j(x, \theta^{(n_0)}) \approx r_j^{\text{shift}}(x), 0 \leq j \leq m, \quad (2-58)$$

相应的 $r_j(x)$ 为

$$r_j(x) \approx e^{-i\omega_j x} T_j(x, \theta^{(n_0)}). \quad (2-59)$$

于是

$$r(x) \approx \sum_{j=0}^m e^{-i\omega_j x} T_j(x, \theta^{(n_0)}). \quad (2-60)$$

最终，我们得到

$$f(x) \approx f_{\text{DNN}}(x) + \sum_{j=0}^m e^{-i\omega_j x} T_j(x, \theta^{(n_0)}). \quad (2-61)$$

以上便是 PPSDNN 的全部算法。本算法可以使用 python 语言实现，用 pytorch^[51]编写，算法大致流程如图2-8：

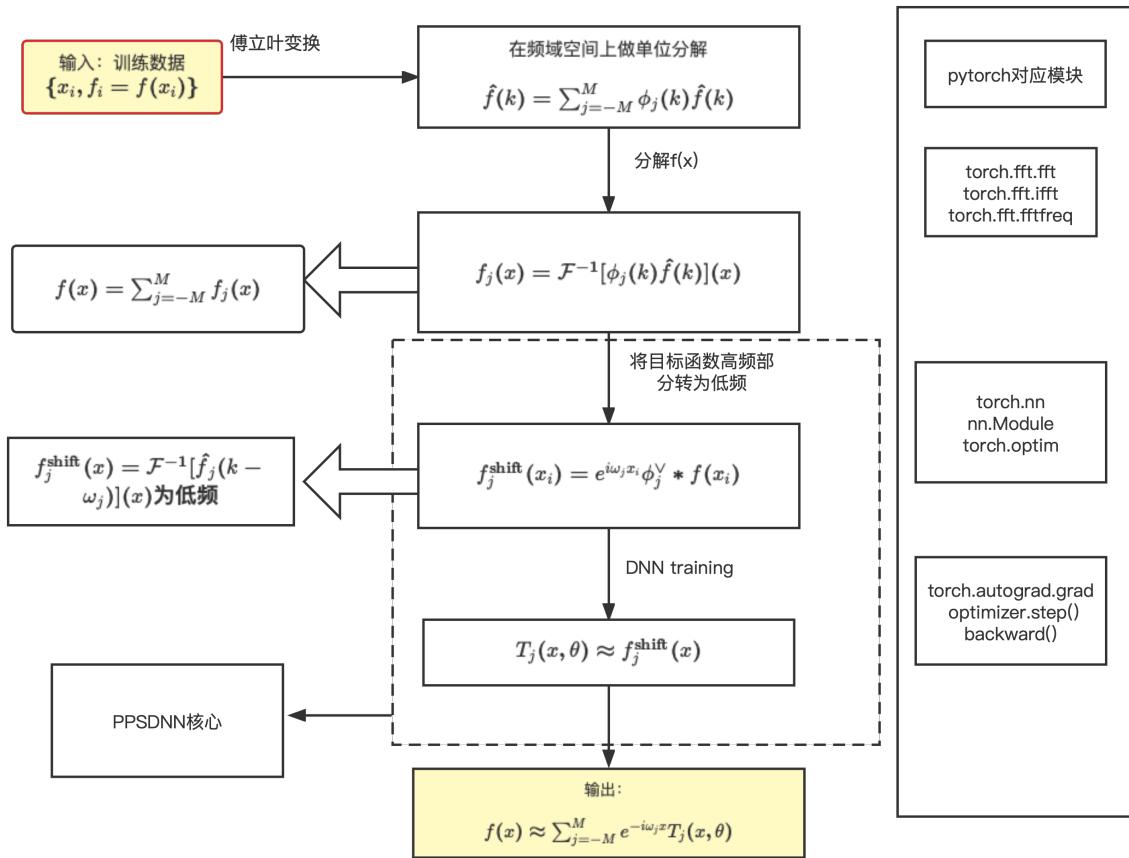


图 2-8 PPSDNN 算法流程图

第3章 初步结果

3.1 toy model 测试结果

下面通过两个数值例子介绍 PPSDNN 的算法实现以及一些更多的算法细节。第一个数值的例子是训练一个有确定函数形式生成的数据，且我们可以根据它的函数形式大致确定其频域空间的范围，在第一个例子中将引入和 PPSDNN 等价的另一种实现 phase shift 的方法。第二个数值例子是训练一个不光滑的函数生成的数据，它的特点是根据函数形式无法大致确定其频域空间的范围，因此在第二个例子中引入频率扫荡方法 (frequency sweep)。此外，PPSDNN 还可以用来拟合多维的振荡数据，解高频的波动方程，亥姆霍兹方程等，但这些应用不在本文的范围之内。

首先介绍第一个数值例子。在第一个例子中，我们选取目标函数 $f(x)$ 的定义域为 $[-\pi, \pi]$ ，函数形式为

$$f(x) = \begin{cases} 10(\sin x + \sin 3x), & \text{如果 } x \in [-\pi, 0] \\ 10(\sin 23x + \sin 137x + \sin 203x), & \text{如果 } x \in [0, \pi]. \end{cases} \quad (3-1)$$

在定义域上画出函数图形，如图3-1所示：

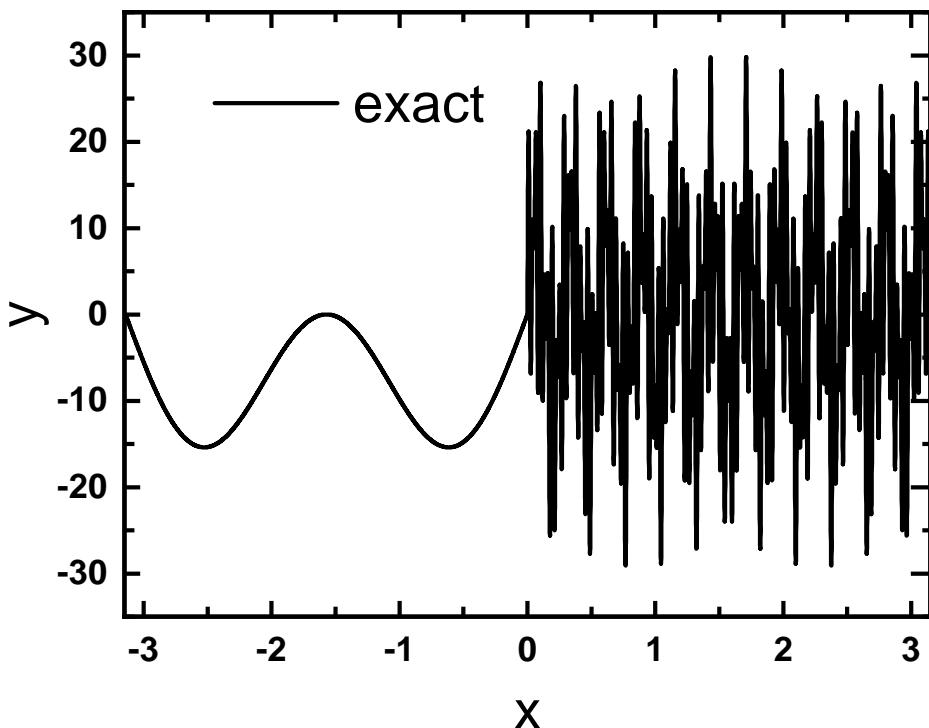


图 3-1 toy model 目标函数3-1式的图形

因为该函数的频率空间可以从它的函数形式或通过傅里叶变换得到，因此我们取 $\Delta k =$

5, 用以下的特征函数对目标函数进行单位分解。

$$\begin{aligned}
 \phi_1(k) &= \chi_{[-205, -200]}(k) & \phi_2(k) &= \chi_{[-140, -135]}(k) \\
 \phi_3(k) &= \chi_{[-25, -20]}(k) & \phi_4(k) &= \chi_{[-5, 0]}(k) \\
 \phi_5(k) &= \chi_{[0, 5]}(k) & \phi_6(k) &= \chi_{[20, 25]}(k) \\
 \phi_7(k) &= \chi_{[135, 140]}(k) & \phi_8(k) &= \chi_{[200, 205]}(k)
 \end{aligned} \tag{3-2}$$

则 $f_j(x) = \mathcal{F}^{-1}[\widehat{f}\phi_j](x)$ 。训练 DNN 的参数为：

- DNN 结构: 1-40-40-40-40-1
- 训练集: 函数定义域上均匀分布的 10,000 个数据点
- 测试集: 函数定义域上均匀分布的 1000 个数据点
- epochs: 1000
- 优化器: Adam
- 学习率: 0.002
- batchsize: 2000

训练结果如图3-2, 训练细节如图3-3,3-4,3-5:

在 PPSDNN 中, 由于频率选择核要与训练数据做卷积运算, 因此, 在高维的卷积运算中, 它的空间复杂度为 $O(N \times N)$, 说明平行相移的方法存在维数灾难, 为了避免这个问题, 人们提出了 Coupled Weighted phase shift DNN^[52] (CPSDNN):

$$T(x) = \sum_{m=1}^M e^{i\omega_m x} T_m(x), \tag{3-3}$$

其中 $\{\omega_m\}_{m=1}^M$ 是和目标函数有关的频率空间。将指数项写成实数的形式为

$$T(x) = \sum_{m=1}^M A_m \cos(\omega_m x) + B_m \sin(\omega_m x) \tag{3-4}$$

于是, 我们可以定义并最小化损失函数:

$$L(\theta) = \int_{-\infty}^{+\infty} |f(x) - T(x)|^2 dx, \tag{3-5}$$

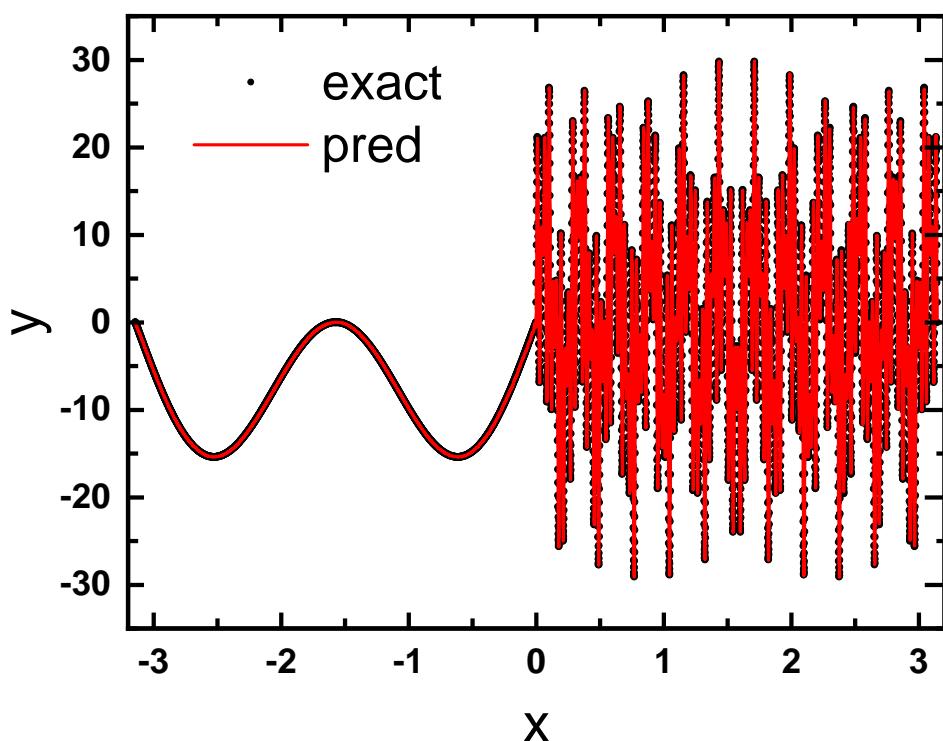
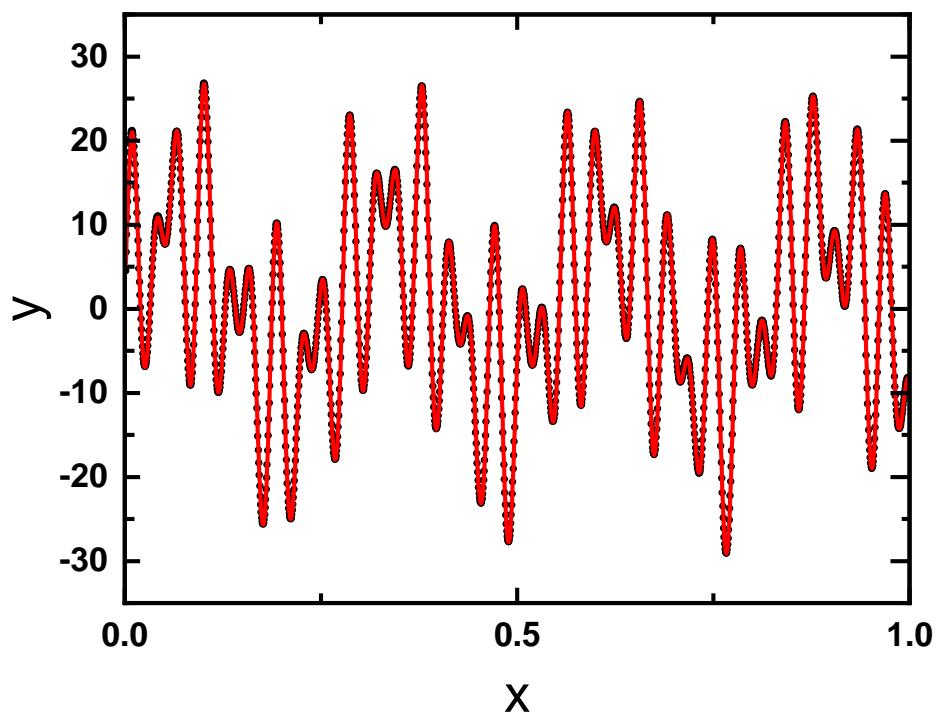


图 3-2 使用 PPSDNN 训练目标函数3-1式

图 3-3 toy model 在区间 $[0, 1]$ 上的训练细节

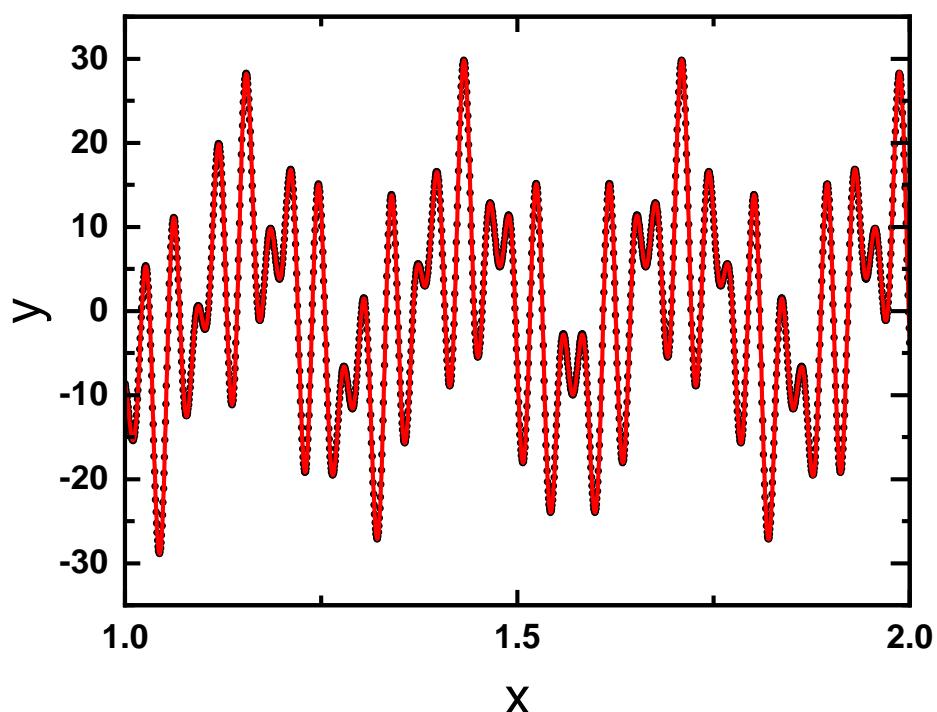


图 3-4 toy model 在区间 $[1, 2]$ 上的训练细节

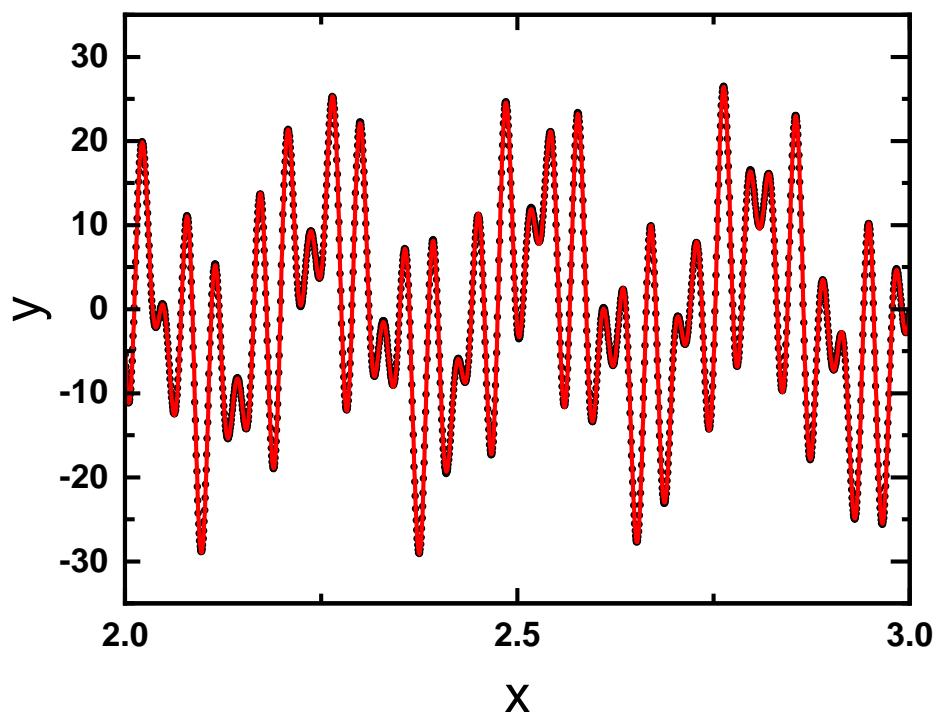


图 3-5 toy model 在区间 $[2, 3.14]$ 上的训练细节

在离散情况下可以写成：

$$L_N(\theta) = \sum_{i=1}^N |f(x_i) - T(x_i)|^2 = \sum_{i=1}^N \left| f(x_i) - \sum_{m=1}^M e^{i\omega_m x_i} T_m(x_i) \right|^2. \quad (3-6)$$

在实际应用中，我们往往引入正则项，于是损失函数可以改写为

$$L_N^R(\theta) = \sum_{i=1}^N |f(x_i) - T(x_i)|^2 + \beta \sum_{m,l} \left\| \vec{W}^{m,l} \right\|_F^2. \quad (3-7)$$

其中， β 是正则化参数， $\vec{W}^{m,l}$ 是 T_m 的权重矩阵。可以证明，PPSDNN 方法和 CPSDNN 方法训练效果是等价的，而 CPSDNN 来源于 PPSDNN 算法的启发。

下面介绍第二个数值例子。在实际工作中，我们学习的核数据一般都是离散的，不像第一个例子生成的函数是连续且光滑的。这时候我们就要用到频率扫荡 (frequency sweep) 方法。前面取的 ω_m 都是几个离散的值，在 frequency sweep 中，我们可以取形如 $\omega_m = -1600 : 10 : 1600$ 这样连续的值。我们将 3-1 的函数稍作修改，将 x 小于零的部分改成阶跃函数，使它的频率空间不再那么确定，接着使用 frequency sweep 进行训练。

训练效果如图 3-6

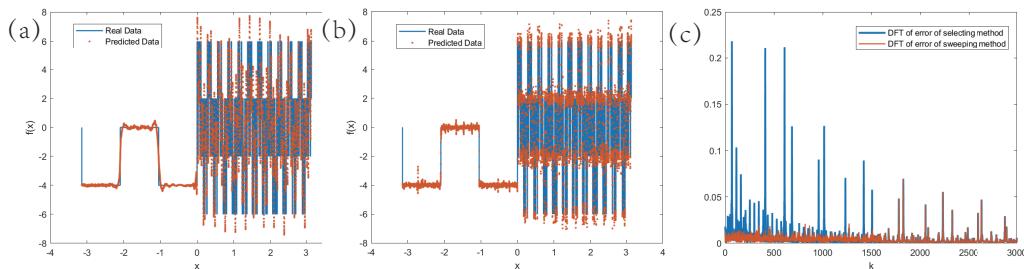


图 3-6 使用 frequency sweep 训练高频函数

3.2 $^{235}\text{U}(n, f)$ 评价库数据测试结果

根据核反应的复合核理论，当中子与靶核形成复合核时，复合核的激发能为

$$E^* = \frac{m_A}{m_n + m_A} E_n + B_{nA} \quad (3-8)$$

其中， m_A 为靶核的质量， m_n 为中子的质量， E_n 为入射中子的相对动能， B_{nA} 为中子与靶核的结合能。当形成的复合核能级等于 E^* 时，入射粒子被强烈吸收，形成共振峰。

近几十年以来， ^{235}U 中子共振区的实验测量和理论研究一直是生机勃勃的，累计的实

验截面数据估计在 10 万以上。由于测量技术，参数分析方法的不同，共振参数间存在着分歧。同时，一家测量分析的结果往往不能覆盖全部能区和给出全部共振参数。因此，综合多家实验数据进行分析评价是十分必要的^[53]。而目前最核心的技术难题就是难以拟合好高频振荡的数据。在 $n + ^{235}\text{U}$ 反应中，存在多种截面数据，包括裂变截面，俘获截面，全截面等。本工作研究的是裂变截面。

ENDF/B-VIII.0 是美国核反应数据库 ENDF 的最新版本，于 2018 年 2 月 2 日发布^[54]。它完全采用了新的中子数据标准，改进了热中子散射数据，并使用了国际评价库组织 (CIELO) 试点项目的新评价数据，涵盖了氢、氧、铁、铀和钚等元素的中子反应^[55]。由于评价库中的核数据与实验核数据相比，其截面数据的连续性与光滑性更好，因此更适合使用神经网络方法进行学习。下面将分别使用普通 DNN 方法和 PPSDNN 方法学习这些实验数据以及评价库中的数据。

在此部分，首先介绍普通 DNN 方法在研究 $^{235}\text{U}(n, f)$ 反应中的应用。根据第二章的分析，由于 f-principle，DNN 方法更偏向于先学习低频数据，再学习高频数据，因此，在中子共振区域，在有限运算资源的情况下，DNN 方法的学习效果并不会很好。

由于实验数据每家测量的分歧较大，因此为了快速从实践上验证 PPSDNN 在中子共振中应用的可行性，本文采用 ENDF/B-VIII.0 的数据作为学习数据集。下面是使用 PPSDNN 方法对 $^{235}\text{U}(n, f)$ 反应的评价库数据进行学习的结果。从图中可以看出，PPSDNN 可以

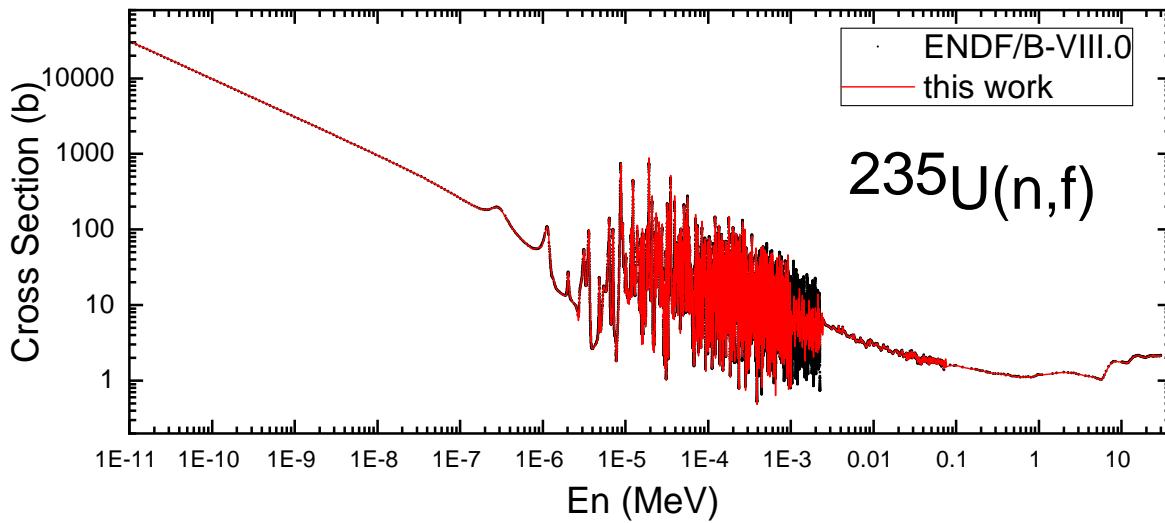


图 3-7 PPSDNN 方法对 ENDF/B-VIII.0 数据库中 $^{235}\text{U}(n, f)$ 反应截面的学习

较好地学习到入射能为 $10^{-6} \sim 0.001\text{MeV}$ 共振区的评价库数据。当入射能超过 0.001MeV 时，可以用来学习的数据量大大减少，且接近不可分辨共振区，因此会在一小段能区无法重现评价库数据。当能量大于 0.003MeV 时，PPSDNN 可以学习到不可分辨共振区的数据。

由于评价库的核数据和实验数据相比更具有光滑性，而我们的目的是根据实验数据给出一套自己的核数据评价方法。因此后续将尝试使用 PPSDNN 方法直接学习 $^{235}\text{U}(n, f)$ 反

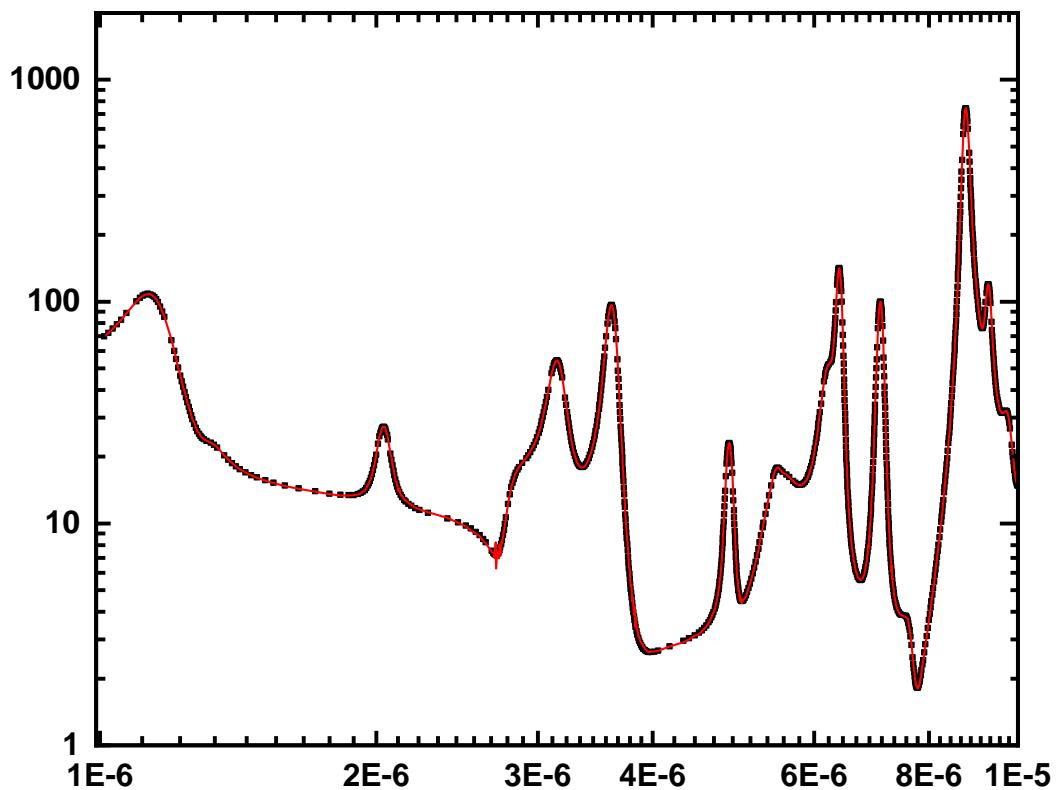


图 3-8 PPSDNN 方法在能量区间为 $[10^{-6}, 10^{-5}]$ MeV 的学习细节

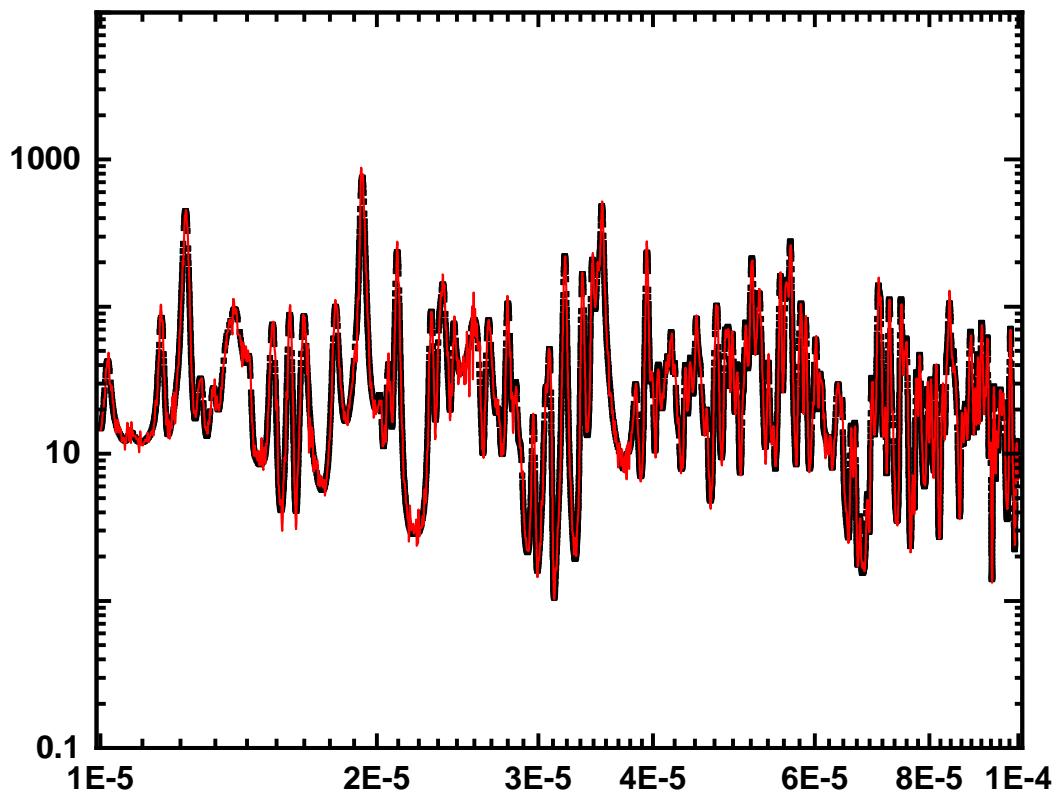


图 3-9 PPSDNN 方法在能量区间为 $[10^{-5}, 10^{-4}]$ MeV 的学习细节

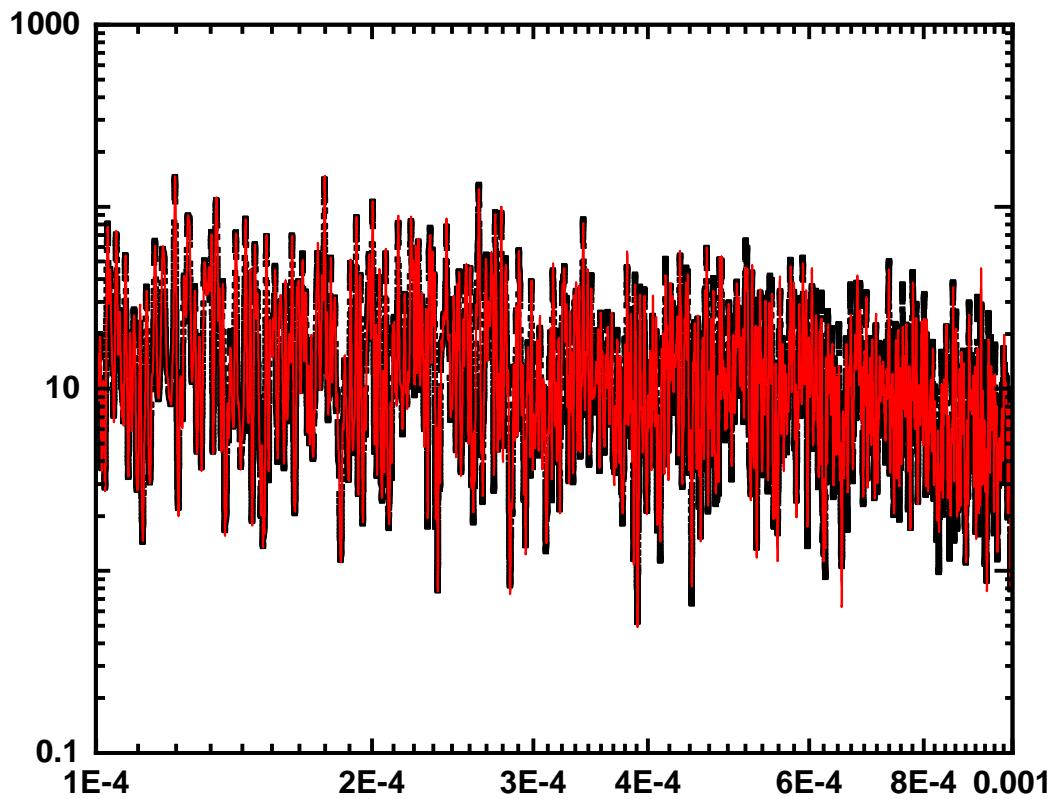


图 3-10 PPSDNN 方法在能量区间为 $[10^{-4}, 10^{-3}] \text{MeV}$ 的学习细节

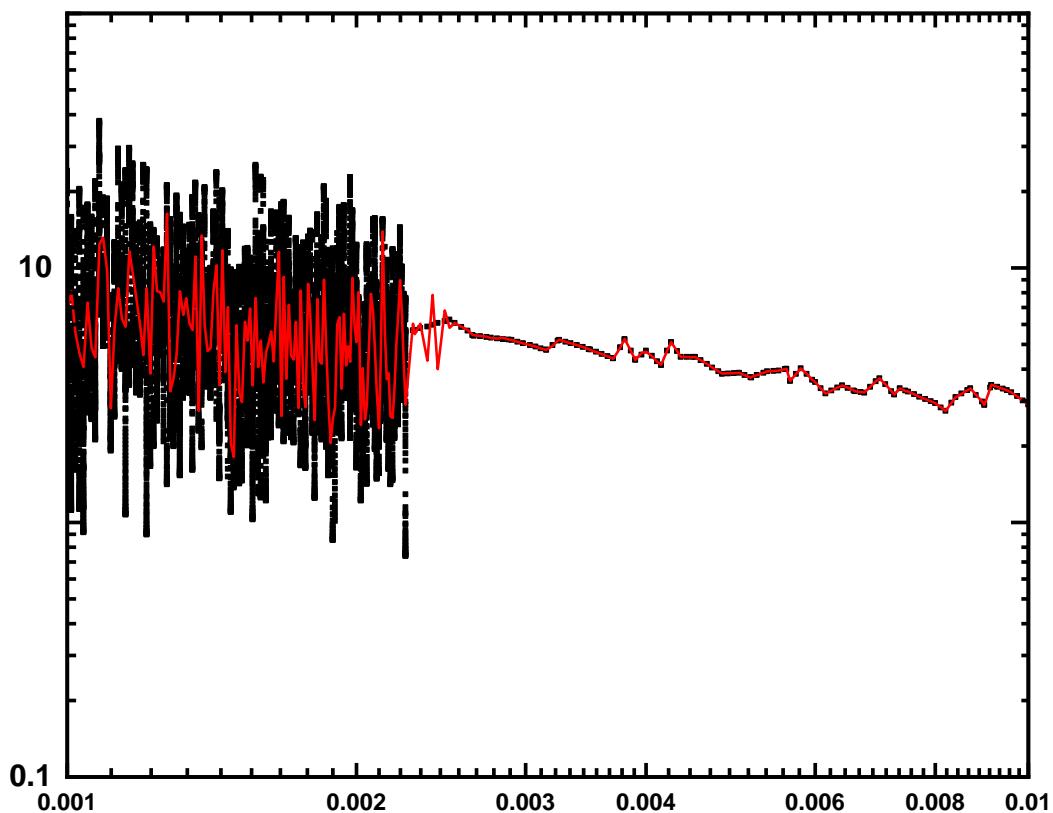
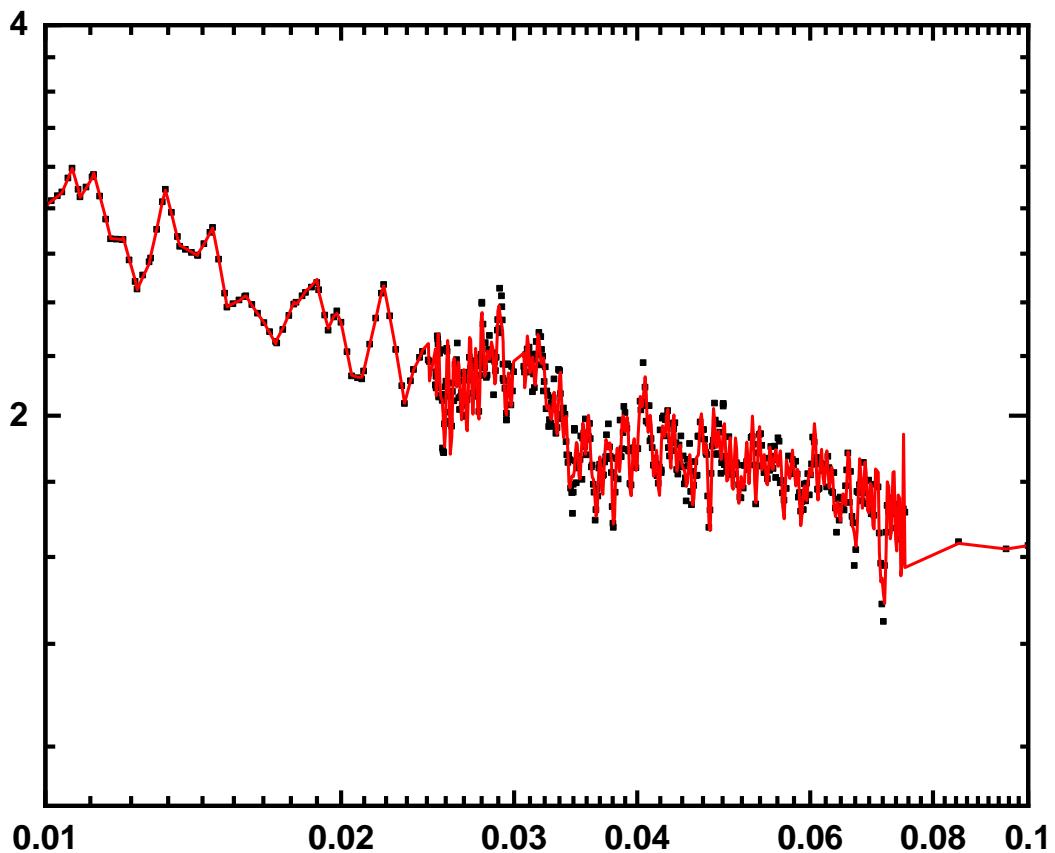
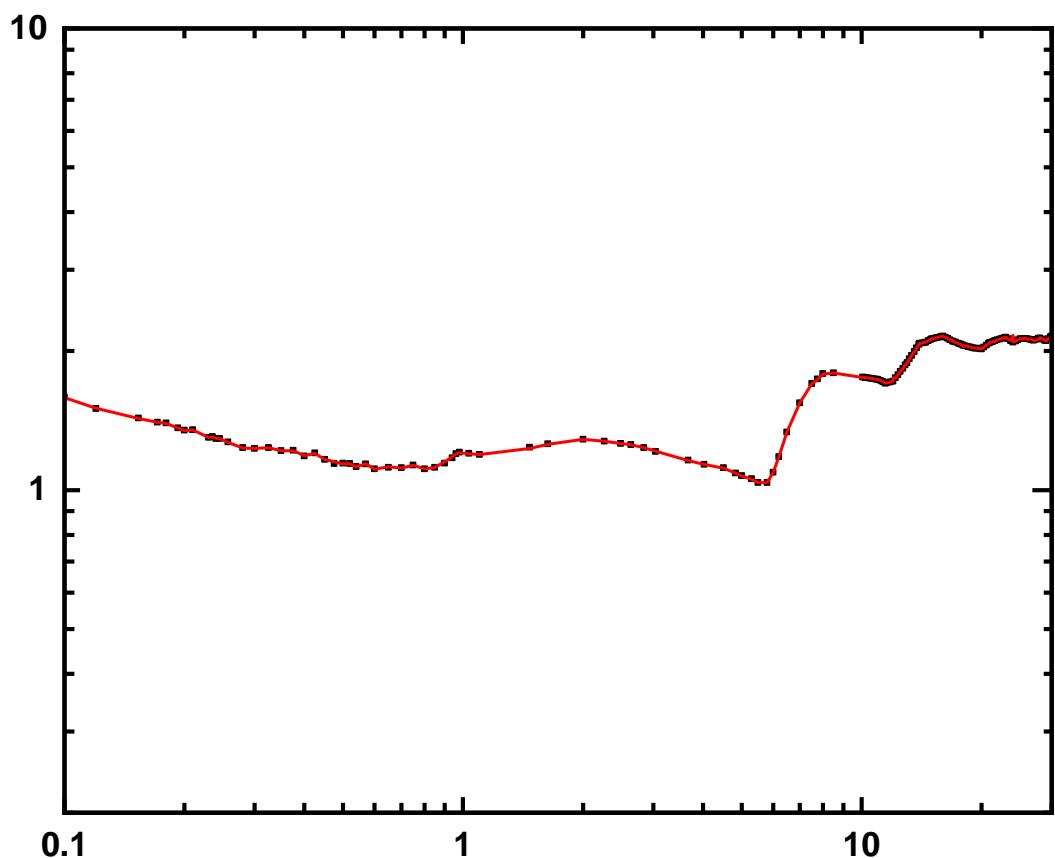


图 3-11 PPSDNN 方法在能量区间为 $[10^{-3}, 10^{-2}] \text{MeV}$ 的学习细节

图 3-12 PPSDNN 方法在能量区间为 $[10^{-2}, 10^{-1}] \text{ MeV}$ 的学习细节图 3-13 PPSDNN 方法在能量区间为 $[0.1, 30] \text{ MeV}$ 的学习细节

应截面的实验数据。

第 4 章 总结与展望

本文详细推导并复现了 PPSDNN 算法，实现了 PPSDNN 算法对 toy model 的学习，在 toy model 中，目标函数是连续可导的，其傅立叶变换也是解析的，根据其傅立叶变换得到的采样频率，通过平行相移，就可以很好的学习 toy model。在实际应用中，训练数据是离散的，比如 $^{235}\text{U}(n, f)$ 反应的评价数据，通过 frequency sweep，也可以做到较好的学习这类离散数据。这表明该算法确实具有较高的预测精度和效率和较好高频数据学习能力。这为今后从事中子共振截面的评价工作奠定了坚实的基础。

机器学习用于中子共振研究还有以下潜在应用：

- (1) 加速计算^[56]。中子共振计算中需要进行大量的矩阵运算和复杂的数值计算，通常非常耗时。使用机器学习算法，通过对之前计算过的数据进行训练，从而加速计算过程并保持相同的计算精度。
- (2) 提高计算精度。中子共振计算中存在着一些难以处理的误差和不确定性。使用机器学习算法通过对大量数据进行训练，从而减小误差和降低不确定性，提高计算精度。
- (3) 数据分析。中子共振实验通常产生大量数据，如能谱和截面等。使用机器学习算法，可以对这些数据进行分析和分类，更好地理解中子共振现象和物理机制。
- (4) 拟合物理模型参数。中子共振研究中需要对物理模型进行拟合，以更好地描述中子共振现象和物理机制。使用机器学习算法，可以对大量的实验数据进行训练，以获得更好的拟合效果和更准确的模型参数。随着中子共振核反应数据量的增加，机器学习方法也逐渐体现出其独有的优势。

在今后的工作中，将继续发挥机器学习的优势，并将 PPSDNN 方法用于学习以 $^{235}\text{U}(n, f)$ 反应为代表的中子共振实验数据中。并使用学习和预言得到的激发曲线给出核反应的共振参数。

目前得到共振参数的主流方法是使用 R 矩阵拟合实验数据。尽管 R 矩阵方法已经被广泛应用，但仍然存在一些困难和限制。例如，该方法只能唯象地给出道半径 R，无法清晰地解释反应机制，而且 R 矩阵程序的编写也较为困难。为此，在今后的科研中计划将 PPSDNN 方法和 R 矩阵相结合，以改进共振参数的计算方法。这一方法将利用 PPSDNN 算法的高预测精度和高效率，以及 R 矩阵的核反应共振的原理，提高共振参数的准确性和高效性。

附录 A 计算双曲正切函数的傅立叶变换

$$\begin{aligned}
 & \int_{-\infty}^{\infty} \tanh(x) e^{-ikx} dx = \\
 &= \int_{-\infty}^{\infty} \frac{e^x - e^{-x}}{e^x + e^{-x}} e^{-ikx} dx \\
 &= \int_{-\infty}^{\infty} e^{-ikx} dx - \int_{-\infty}^{\infty} \frac{2e^{-ikx}}{e^{2x} + 1} dx
 \end{aligned} \tag{A-1}$$

先计算积分第一项：

$$\begin{aligned}
 \int_{-\infty}^{\infty} e^{-ikx} dx &= \lim_{\alpha \rightarrow \infty} \int_{-\alpha}^{\alpha} e^{-ikx} dx \\
 &= \lim_{\alpha \rightarrow \infty} \left. \frac{-e^{-ikx}}{ik} \right|_{-\alpha}^{\alpha} \\
 &= \lim_{\alpha \rightarrow \infty} \frac{1}{ik} (e^{i\alpha k} - e^{-i\alpha k}) \\
 &= \lim_{\alpha \rightarrow \infty} \frac{2i}{ik} \sin \alpha k = 2\pi \delta(k).
 \end{aligned} \tag{A-2}$$

下面证明 $\lim_{\alpha \rightarrow \infty} \frac{\sin \alpha k}{\pi k} = \delta(k)$ 。首先证明它的积分等于 1。注意到：

$$\begin{aligned}
 I &= \int_0^{\infty} e^{-\gamma x} \cos \beta x dx = e^{-\gamma x} \frac{\sin \beta x}{\beta} \Big|_0^{\infty} + \frac{\gamma}{\beta} \int_0^{\infty} e^{-\gamma x} \sin \beta x dx \\
 &= \frac{\gamma}{\beta} \int_0^{\infty} e^{-\gamma x} \sin \beta x dx = -\frac{\gamma}{\beta^2} e^{-\gamma x} \cos \beta x \Big|_0^{\infty} - \frac{\gamma^2}{\beta^2} \int_0^{\infty} e^{-\gamma x} \cos \beta x dx \\
 &= \frac{\gamma}{\beta^2} - \frac{\gamma^2}{\beta^2} I
 \end{aligned} \tag{A-3}$$

解得：

$$I = \int_0^{\infty} e^{-\gamma x} \cos \beta x dx = \frac{\gamma}{\beta^2 + \gamma^2}, \quad \gamma > 0 \tag{A-4}$$

将式A-4的等式两边对 β 从 0 到 α 进行积分，得

$$\begin{aligned}
 & \int_0^{\alpha} d\beta \left(\int_0^{\infty} e^{-\gamma x} \cos \beta x dx \right) = \int_0^{\infty} dx e^{-\gamma x} \left(\int_0^{\alpha} d\beta \cos \beta x \right) \\
 &= \int_0^{\infty} dx e^{-\gamma x} \frac{\sin \alpha x}{x} = \int_0^{\alpha} d\beta \frac{\gamma}{\beta^2 + \gamma^2} = \arctan \frac{\alpha}{\gamma}.
 \end{aligned} \tag{A-5}$$

于是有

$$\lim_{\gamma \rightarrow 0} \int_0^{\infty} dx e^{-\gamma x} \frac{\sin \alpha x}{x} = \int_0^{\infty} dx \frac{\sin \alpha x}{x} = \lim_{\gamma \rightarrow 0} \arctan \frac{\alpha}{\gamma} = \arctan \infty = \frac{\pi}{2} \tag{A-6}$$

再通过等式变化，就可以完成证明。可以看出，第一项积分可以化为 δ 函数，当 $k > 0$ 时

为零；下面计算第二项积分，计算第二项积分需要使用留数定理，由于被积函数中的 $\frac{1}{e^{2x}+1}$ 存在无穷多个奇点，为了避免计算无穷多个奇点，令 $t = e^x$ ，在上半圆形，逆时针的围道上只需要计算一个奇点 $t = i$ 的留数即可。下面是计算第二项积分的主要步骤：

$$\begin{aligned} & \int_{-\infty}^{\infty} \frac{e^{-ikx}}{e^{2x}+1} dx \\ & \stackrel{t=e^x}{=} \int_0^{\infty} \frac{t^{-ik}}{t^2+1} \frac{dt}{t}. \end{aligned} \quad (\text{A-7})$$

为了使用留数定理，给出一个逆时针半圆围道：

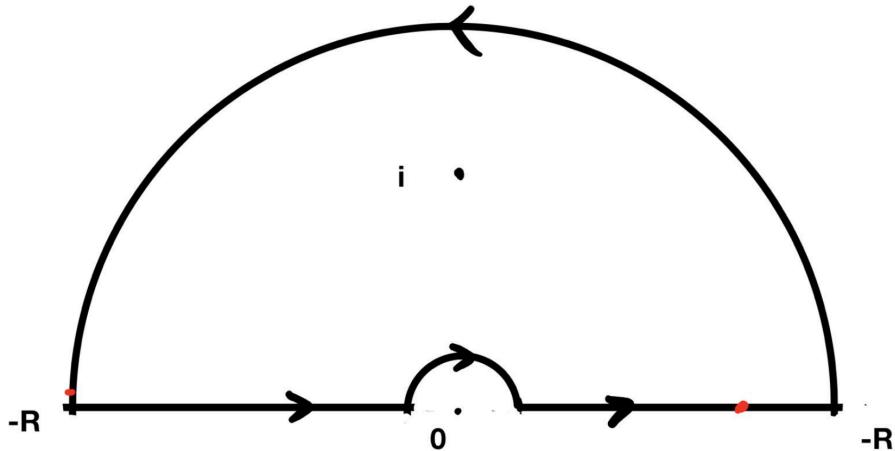


图 1-1 半圆围道

在半圆围道上，存在：

$$\begin{aligned} & \lim_{R \rightarrow \infty} \left(\int_0^R \frac{z^{-ik}}{z^2+1} \frac{dz}{z} + \int_{-R}^0 \frac{z^{-ik}}{z^2+1} \frac{dz}{z} \right) \\ &= \lim_{R \rightarrow \infty} \left(\int_0^R \frac{z^{-ik}}{z^2+1} \frac{dz}{z} - \int_0^{-R} \frac{z^{-ik}}{z^2+1} \frac{dz}{z} \right) \\ &= \lim_{R \rightarrow \infty} \left(\int_0^R \frac{z^{-ik}}{z^2+1} \frac{dz}{z} - \int_0^R \frac{(ze^{i\pi})^{-ik}}{(ze^{i\pi})^2+1} \frac{d(-z)}{ze^{i\pi}} \right) \\ &= \lim_{R \rightarrow \infty} \left(\int_0^R \frac{z^{-ik}}{z^2+1} \frac{dz}{z} - e^{k\pi} \int_0^R \frac{z^{-ik}}{z^2+1} \frac{dz}{z} \right). \end{aligned} \quad (\text{A-8})$$

在该半圆形围道上使用留数定理：

$$\begin{aligned} & \int_C \frac{z^{-ik}}{z^2+1} \frac{dz}{z} = 2\pi i \cdot \text{Res } f(z=i) + \pi i \cdot \text{Res } f(z=0) \\ &= 2\pi i \left(\lim_{z \rightarrow i} \frac{z^{-ik}}{3z^2+1} + \lim_{z \rightarrow 0} \frac{z^{-ik}}{3z^2+1} \right) = -\pi i e^{\frac{k\pi}{2}} + 0 \\ &= \lim_{R \rightarrow \infty} \left(\int_0^R \frac{z^{-ik}}{z^2+1} \frac{dz}{z} + \int_{-R}^0 \frac{z^{-ik}}{z^2+1} \frac{dz}{z} \right) + \int_{\text{arc}} \frac{z^{-ik}}{z^2+1} \frac{dz}{z} \end{aligned} \quad (\text{A-9})$$

其中, $\int_{\text{arc}} \frac{z^{-ik}}{z^2+1} dz$ 是半圆弧线上的积分, 可以估算:

$$\left| \int_{\text{arc}} \frac{z^{-ik}}{z^2+1} dz \right| \leq \pi R \cdot \sup_{\text{arc}} \left| \frac{z^{-ik}}{z^2+1} \right| \leq \pi R \cdot \sup_{\text{arc}} \frac{1}{|z^3+z|} \leq \frac{\pi R}{R^3 + R - 1} \quad (\text{A-10})$$

且:

$$\lim_{R \rightarrow \infty} \frac{\pi R}{R^3 + R - 1} = 0 \quad (\text{A-11})$$

所以:

$$\int_{\text{arc}} \frac{z^{-ik}}{z^2+1} dz = 0 \quad (\text{A-12})$$

将式A-8,A-12代入式A-9可得:

$$\begin{aligned} & \lim_{R \rightarrow \infty} \int_0^R \frac{z^{-ik}}{z^2+1} dz = \frac{-\pi i e^{\frac{k\pi}{2}}}{1 - e^{k\pi}} \\ &= \frac{\pi i}{2 \sinh \frac{k\pi}{2}} \end{aligned} \quad (\text{A-13})$$

于是第二项积分为:

$$\int_{-\infty}^{\infty} \frac{2e^{-ikx}}{e^{2x}+1} dx = \frac{\pi i}{\sinh \frac{k\pi}{2}} \quad (\text{A-14})$$

于是, 双曲正切函数的傅立叶变换为:

$$\int_{-\infty}^{\infty} \tanh(x) e^{-ikx} dx = -\frac{\pi i}{\sinh \frac{k\pi}{2}} \quad (\text{A-15})$$

参考文献

- [1] Mitchell, Michael T. Machine learning: Vol. 1[M]. McGraw-hill New York, 2007.
- [2] 李航. 统计学习方法[M]. Qing hua da xue chu ban she, 2012.
- [3] Chowdhary K, Chowdhary K. Natural language processing[J]. Fundamentals of artificial intelligence, 2020: 603-649.
- [4] 刘浏, 王东波. 命名实体识别研究综述[J]. 情报学报, 2018, 37(3): 329-340.
- [5] 卢希庭. 原子核物理[M]. 原子能出版社, 2000.
- [6] 庚先国, 刘福乐, 王琦标, 等. 中重核中子共振研究及应用进展[J]. 核技术, 2020, 43(10): 7-21.
- [7] Cierjacks S, Hinterberger F, Schmalz G, et al. High precision time-of-flight measurements of neutron resonance energies in carbon and oxygen between 3 and 30 mev[J]. Nuclear Instruments and Methods, 1980, 169(1): 185-198.
- [8] Hore P J. Nuclear magnetic resonance[M]. Oxford University Press, USA, 2015.
- [9] Seidl F. An interpretation of the resonance scattering of neutrons by cobalt in terms of the breit-wigner equation[J]. Physical Review, 1949, 75(10): 1508.
- [10] 郭之虞, 王宇钢, 包尚联. 核技术及其应用的发展[J]. 北京大学学报: 自然科学版, 2003 (z1): 82-91.
- [11] Descouvemont P, Baye D. The r-matrix theory[J]. Reports on progress in physics, 2010, 73(3): 036301.
- [12] Larson N M. Updated users' guide for sammy multilevel r-matrix fits to neutron data using bayes' equation[R]. Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States), 1998.
- [13] Bell M. Origen: the ornl isotope generation and depletion code[R]. Oak Ridge National Lab., Tenn.(USA), 1973.
- [14] Ren X, Zelenay P, Thomas S, et al. Recent advances in direct methanol fuel cells at los alamos national laboratory[J]. Journal of Power Sources, 2000, 86(1-2): 111-116.

- [15] Kelley J, Tilley D, Purcell J, et al. Tunl nuclear data evaluation project[Z]. 2004.
- [16] Chen Z, Qi H. R-matrix analysis code (rac)[R]. China Nuclear Information Centre, 1990.
- [17] Ge Z, Xu R, Wu H, et al. Cendl-3.2: The new version of chinese general purpose evaluated nuclear data library[C]//EPJ Web of Conferences: Vol. 239. EDP Sciences, 2020: 09001.
- [18] 陶曦, 蔡崇海, 申庆彪. 20 MeV 以下 $n + {}^6 Li$ 反应的理论计算[J]. 原子核物理评论, 2015, 32(3): 368-374.
- [19] 唐靖宇. 白光中子源及其多学科应用[J]. 中国基础科学, 2020, 22(4): 22-31.
- [20] 张奇玮, 栾广源, 任杰, 等. 基于 CSNS 反角白光中子源的中子俘获反应截面测量技术研究[J]. 物理学报, 2021, 70(22): 222801.
- [21] Boehnlein A, Diefenthaler M, Sato N, et al. Colloquium: Machine learning in nuclear physics[J]. Reviews of Modern Physics, 2022, 94(3): 031003.
- [22] 李庆峰, 马余刚. 机器学习在原子核物理中的应用专题 · 编者按[J]. 中国科学: 物理学 力学 天文学, 2022.
- [23] 樊春玲, 孙四通, 金志华. 贝叶斯神经网络建模预测方法及其应用[J]. 中国惯性技术学报, 2009, 17(1): 85-88.
- [24] Niu Z, Liang H. Nuclear mass predictions based on bayesian neural network approach with pairing and shell effects[J]. Physics Letters B, 2018, 778: 48-53.
- [25] Shang T S, Li J, Niu Z M. Prediction of nuclear charge density distribution with feedback neural network[J]. Nuclear Science and Techniques, 2022, 33(12): 153.
- [26] Saxena G, Sharma P, Saxena P. Modified empirical formulas and machine learning for α -decay systematics[J]. Journal of Physics G: Nuclear and Particle Physics, 2021, 48(5): 055103.
- [27] Minato F, Niu Z, Liang H, et al. Calculation of β -decay half-lives within a skyrme-hartree-fock-bogoliubov energy density functional with the proton-neutron quasiparticle random-phase approximation and isoscalar pairing strengths optimized by a bayesian method[J]. Physical Review C, 2022, 106(2): 024306.

- [28] Niu Z, Fang J, Niu Y, et al. Comparative study of radial basis function and bayesian neural network approaches in nuclear mass predictions[J]. Physical Review C, 2019, 100(5): 054311.
- [29] Wang Z A, Pei J. Optimizing multilayer bayesian neural networks for evaluation of fission yields[J]. Physical Review C, 2021, 104(6): 064608.
- [30] Akkoyun S. Estimation of fusion reaction cross-sections by artificial neural networks[J]. Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, 2020, 462: 51-54.
- [31] 庞龙刚, 周凯, 王新年. 深度学习在核物理中的应用[J]. 原子核物理评论, 2020, 37(3).
- [32] Wiering M A, Van Otterlo M. Reinforcement learning[J]. Adaptation, learning, and optimization, 2012, 12(3): 729.
- [33] 王珏, 周志华, 周傲英. 机器学习及其应用: 第 4 卷[M]. 清华大学出版社有限公司, 2006.
- [34] Rojas R, Rojas R. The backpropagation algorithm[J]. Neural networks: a systematic introduction, 1996: 149-182.
- [35] Deng L, Yu D, et al. Deep learning: methods and applications[J]. Foundations and trends® in signal processing, 2014, 7(3-4): 197-387.
- [36] 尹宝才, 王文通, 王立春, 等. 深度学习研究综述[J]. 北京工业大学学报, 2015, 41(1): 48-59.
- [37] 姚端正, 梁家宝. 数学物理方法[M]. 武汉大学出版社, 1997.
- [38] Hughes H K. The physical meaning of parseval's theorem[J]. American Journal of Physics, 1965, 33(2): 99-101.
- [39] Cybenko G. Approximation by superpositions of a sigmoidal function[J]. Mathematics of control, signals and systems, 1989, 2(4): 303-314.
- [40] Xu Z Q J, Zhang Y, Luo T, et al. Frequency principle: Fourier analysis sheds light on deep neural networks[A]. 2019.
- [41] Bracewell R N, Bracewell R N. The fourier transform and its applications: Vol. 31999 [M]. McGraw-Hill New York, 1986.

- [42] 刘树棠. 信号与系统[M]. 西安交通大学出版社, 2010.
- [43] Cai W, Wang J. Adaptive multiresolution collocation methods for initial-boundary value problems of nonlinear pdes[J]. SIAM Journal on Numerical Analysis, 1996, 33(3): 937-970.
- [44] Cai W, Li X, Liu L. Phasednn-a parallel phase shift deep neural network for adaptive wideband learning[A]. 2019.
- [45] Daubechies I. Ten lectures on wavelets[M]. SIAM, 1992.
- [46] Vaidyanathan P. Generalizations of the sampling theorem: Seven decades after nyquist [J]. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, 2001, 48(9): 1094-1109.
- [47] Xu Z J. Understanding training and generalization in deep learning by fourier analysis [A]. 2018.
- [48] Benedetto J J, Ferreira P J. Modern sampling theory: mathematics and applications [M]. Springer Science & Business Media, 2001.
- [49] 卓里奇. 数学分析: 第二卷[M]. 高等教育出版社, 2019.
- [50] Prochazkova J. Derivative of b-spline function[C]//Proceedings of the 25th Conference on Geometry and Computer Graphics, Prague, Czech Republic. 2005: 12-16.
- [51] Ketkar N, Moolayil J, Ketkar N, et al. Introduction to pytorch[J]. Deep Learning with Python: Learn Best Practices of Deep Learning Models with PyTorch, 2021: 27-91.
- [52] Cai W, Li X, Liu L. A phase shift deep neural network for high frequency approximation and wave problems[J]. SIAM Journal on Scientific Computing, 2020, 42(5): A3285-A3312.
- [53] 何锦昌, 李冕丰. 中子共振参数的联合拟合分析[J]. 原子能科学技术, 1990, 24(2): 21-28.
- [54] Brown D, Chadwick M, Capote R, et al. ENDF/b-VIII.0: The 8 th major release of the nuclear reaction data library with CIELO-project cross sections, new standards and thermal scattering data[J]. Nuclear Data Sheets, 2018, 148: 1-142.
- [55] Sobes V, Holcomb A, Marshall B, et al. Augmented ENDF/b-VIII.0 covariance library for SCALE 6.3[J]. Annals of Nuclear Energy, 2021, 160: 108353.

- [56] Verbraeken J, Wolting M, Katzy J, et al. A survey on distributed machine learning[J].
Acm computing surveys (csur), 2020, 53(2): 1-33.

致 谢

我在此向所有在我硕士论文完成过程中提供支持和帮助的人们致以诚挚的感谢和敬意。

首先，我要向我的导师孙小军教授表达最真挚的感谢之情。导师不仅在研究过程中给予我专业而精准的指导和建议，更是以身作则，教会我如何做一个有追求、有担当的科研工作者。导师在我人生道路上的指引和支持，将成为我前行的力量和动力，使我更加坚定迈向未来的信心和决心。

其次，我要感谢罗秋莲等计算物理组的师兄师姐们，感谢他们在我研一时的帮助和支持。我还要感谢我的同门彭海圆师兄，梁艳师姐，张凯师兄，蓝秀丽师姐，邹方磊，唐诺程，崔嫣然，曹寒梅。另外，我还要感谢姚红师兄，万梦婷，杨贊，张国峰，陈立欣，陈叶若溪，李文琳，莫腾欢等核组的同学，我研究生大部分时间都是和他们一起度过的。以上这些人为我提供了最基础、最实用的技术支持，让我可以在学术中尽情发挥。他们严谨的治学态度、乐观向上的精神，都在潜移默化中影响和鼓励着我。

此外，我要感谢我的家人和朋友们，感谢他们在我完成学业的道路上给予我物质上的鼓励和支持。正是有了你们的陪伴和支持，我才有了追求知识和实现梦想的勇气和动力。谢谢你们的默默付出，是你们的陪伴，让我更坚定地走向自己的梦想之路。

另外，我要感谢一只特别的小生命——薛咪。她是我在研究生期间被收养的一只流浪猫。她的出现给了我无穷的欢乐和慰藉。

最后，感谢当年选择物理专业的那个自己。一路走来，世界很美好。

论文独创性声明

本人郑重声明：所提交的学位论文是本人在导师的指导下进行的研究工作及取得的成果。除文中已经注明引用的内容外，本论文不含其他个人或其他机构已经发表或撰写过的研究成果。对本文的研究作出重要贡献的个人和集体，均已在文中以明确方式标明。本人承担本声明的法律责任。

研究生签名: _____ 日 期: _____

论文使用授权声明

本人完全了解广西师范大学有关保留、使用学位论文的规定。本人授权广西师范大学拥有学位论文的部分使用权，即：学校有权保留本人所送交学位论文的复印件和电子文档，可以采用影印、缩印或其他复制手段保存论文；学校有权向国家有关部门或机构送交学位论文的复印件和电子版。本人电子文档的内容和纸质论文的内容相一致。除在保密期内的保密论文外，允许论文被查阅和借阅，可以公布（包括刊登）论文的全部或部分内容。论文的公布（包括刊登）授权广西师范大学学位办办理。

本学位论文属于：

- 保密，在_____年解密后适用授权。
- 不保密。

论文作者签名: _____ 日 期: _____

指导教师签名: _____ 日 期: _____

作者联系电话: _____ 电子邮箱: _____