

分类号: O242.1
U.D.C.: 519.6

密级: 公开
编号: 82817

中国工程物理研究院

学位论文

基于环流的大偏差理论和涨落定理的研究

姜瑜浩

指导教师姓名 贾晨 助理教授

申请学位级别 硕士 专业名称 应用数学

论文提交日期 2022 年 3 月 论文答辩日期 2022 年 6 月

授予学位单位和日期 中国工程物理研究

答辩委员会主席

评阅人

2022 年 3 月 19 日

Classified Index: O242.1
U.D.C.: 519.6
Secret State: Public
Number: 82817

China Academy of Engineering Physics

Dissertation for the Master Degree in Engineering

Large deviations and fluctuation theorems for cycle currents

Yuhao Jiang

Supervisor: Prof. Chen Jia

Academic Degree Applied for: Master of Science

Specialty: Applied Mathematics

Date of Submitting: April, 2022 Date of Defence: June, 2022

Degree-Confering-Institution: China Academy of Engineering Physics

Chairman of defence committee

Paper Reviewers

19th 3, 2022

独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得中国工程物理研究院或其他教育机构的学位或证书使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签名：

签字日期： 年 月 日

学位论文版权使用授权书

本学位论文作者完全了解并接受中国工程物理研究院研究生部有关保存、使用学位论文的规定，允许论文被查阅、借阅和送交国家有关部门或机构，同时授权中国工程物理研究院研究生部可以将学位论文全部或部分内容编入有关数据库进行检索，可以影印、缩印或扫描等复制手段保存、汇编学位论文。

学位论文作者签名：

导师签名：

签字日期： 年 月 日

签字日期： 年 月 日

摘 要

环流理论一直是随机热力学理论的关键内容。马氏系统中的环流和净环流可以由两种方式定义，即环擦除和生成树。该研究针对两种研究下的大偏差理论和涨落定理进行比较研究。首先研究环拓扑结构系统，并计算出环擦除定义下大偏差速率函数的隐式表达式，然后研究生成树定义下的大偏差速率函数，并阐明两者之间的关系。进一步，检验两个定义下的结论和涨落定理的相容性。经过严格对照，生成树定义下的环流不满足涨落定理，然而生成树的结果满足涨落定理的弱形式。

关键词：环流；大偏差；涨落定理

Abstract

The cycle current is a crucial quantity in stochastic thermodynamics. The absolute and net cycle currents of a Markovian system can be defined in the loop-erased (LE) or the spanning tree (ST) manner. Here we make a comparative study between the large deviations and fluctuation theorems for the LE and ST currents, i.e. cycle currents defined in the LE and ST manners. First, we give the explicit expression of the large deviation rate functions for the LE currents of a system with a cyclic topology and for the ST currents of a general system. The relationship between the rate functions for the LE and ST currents are clarified. Furthermore, we examine various types of fluctuation theorems satisfied by the LE and ST currents. We show that both the absolute and net LE currents satisfy the strong form of all types of fluctuation theorems. In contrast, the absolute ST currents do not satisfy fluctuation theorems, while the net ST currents only satisfy the weak form of fluctuation theorems.

Keywords: cycle currents, large deviations, fluctuation theorems

目录

第一章 绪论

1.1 问题的背景和研究现状

本文中研究的椭圆方程具有如下的一般形式。在 d 维空间中的求解区域 $\Omega \subset \mathbb{R}^d$ 上, 椭圆方程表示为

$$\begin{cases} -\nabla(a(x)\nabla u(x)) + c(x)u(x) = f(x) & \text{for } x \in \Omega \\ u(x) = u_0(x) & \text{for } x \in \partial\Omega \end{cases} \quad (1-1)$$

其中 $a(x) > a_0 > 0, c(x) \geq 0, f(x), u_0(x)$ 是已知函数, $u(x)$ 是方程的解。

椭圆特征值问题表示为

$$\begin{cases} -\nabla(a(x)\nabla u(x)) + c(x)u(x) = \lambda u(x) & \text{for } x \in \Omega \\ u(x) = 0 & \text{for } x \in \partial\Omega \end{cases} \quad (1-2)$$

其中, $u(x)$ 是待求解的特征函数, λ 是对应的特征值。

椭圆方程是应用很广泛的一类方程, 例如: 描述溶液中蛋白质电势分布的 Poisson-Boltzmann 方程 [?], 描述量子效应的 Schrodinger 方程等, 都可以看作是椭圆方程。目前已经有很多传统方法 (如: 有限差分法, 有限元方法, 有限体积方法, 谱方法等) 可以高效且高精度地求解椭圆方程。在本文中, 我们主要关注用深度学习方法求解椭圆方程。

深度学习方法是当前研究的热点问题。在许多领域内 (如: 图像处理, 语音识别等) 深度学习方法已经被广泛应用, 并且取得了超越传统方法的效果。在科学计算问题中, 深度学习方法也已经取得了一定的进展。由于神经网络的逼近性质和维数的关系不大, 深度学习方法能够一定程度上缓解维数灾难, 在求解高维方程时有很好的效果。此外, 由于深度学习方法不需要像传统的有限元或有限差分法那样使用结构化网格, 因此它可以轻易处理复杂的求解区域。文献 [? ? ?] 中讨论了深度学习方法在求解微分方程中的应用。文献 [? ?] 中讨论了如何通过深度学习方法模拟物理模型。文献 [? ?] 中讨论了深度神经网络和传统有限元方法的联系, 从数学角度解释了深度神经网络的逼近性质。

对于深度学习方法在求解椭圆方程上的具体应用, 文献 [? ?] 中提出了 Deep-Ritz 方法, 通过极小化能量泛函的方式求解方程。同时, 文献 [?] 中提出了 PINN 方法, 通过极小化残量的方式求解方程。

然而, 要将常用的深度神经网络应用于计算科学和工程问题, 我们面临着一些挑战。其中最突出的问题是, 深度学习方法通常只适合于处理低频数据。文献 [? ? ?] 中提出了频率准则。频率准则指出, 许多深度神经网络学习低频数据的速度很快, 效果很好, 但当处理高频数据时, 它们就失去了这些优势。文献 [? ? ? ?] 在理论上严格地论证了频率准则。因此, 对于解具有高震荡的问题, 我们需要采取特殊的处理。

目前用深度学习方法求解偏微分方程的工作已经有许多，但从频域的角度探索深度学习方法的研究还不完善。文献 [?] 中提出了 Phase-DNN。它通过在频域上平移的方式求解带有高震荡的方程，取得了很好的效果。但美中不足的是，Phase-DNN 在求解过程中需要事先知道精确解的频率，因此无法求解频率未知的方程。

神经网络在不同频率下的表现不同，利用这种差异设计神经网络结构可以有利于提高网络的性能。在计算机视觉领域，已经有一系列的工作，如：图像恢复 [?]、超分辨率 [?]、分类问题 [?] 等，通过利用不同图像频率的学习差异，提高了学习性能（包括泛化性能和训练速度）。在多尺度变换方面，文献 [?] 和文献 [?] 中分别基于快速多极算法和小波变换的思想设计了多尺度网络，使网络性能得到了很大提升。

1.2 本文主要研究内容

本文主要研究内容可以分为以下三个方面：

首先，在实际问题中，我们通常希望模型和输入数据的尺度无关。例如在图像处理中，我们希望输入的灰度值无论是在 $[0, 255]$ 还是 $[0, 1]$ 之间，得到的结果应该类似。很多机器学习方法（如：简单的线性模型，SVM 等）都有这种性质，但是并没有理论保证深度学习具有类似的性质。本文论述了尺度变换（在输入数据或者输出数据上）对网络的影响，得到了一些初步的结论。

其次，深度学习是基于优化算法的方法，它把方程求解的问题转化成一个优化问题。但是由于神经网络本身的复杂性，这样的优化问题通常是非凸且病态的。想要高效求解这样的优化问题并不容易，一般来说，我们只能找到问题中靠近初始点的局部极小值。因此，在这样的问題中，初值的选取十分关键，好的初值可以提高网络的精度（使训练过程收敛到误差更小的局部最优值），也可以加速训练过程（使初始值离局部最优值更近）。本文基于尺度变换，提出了一种有效的初值选取方式，同时基于这种初值选取方式提出了多尺度网络结构。

最后，本文实验了多尺度网络在求解椭圆方程和分子计算模拟问题中的效果，同时指出了深度学习方法在求解椭圆特征值问题中的潜力。经过各方面的实验，我们可以验证，相比于传统的全连接网络，多尺度网络求解椭圆方程的速度更快，误差也更小。这些实验清楚地表明，多尺度网络是一种高效且易于实现的无网格椭圆方程求解器。

第二章 模型介绍

在此考虑基于离散时间马氏链 $\xi = (\xi_n)_{n \geq 0}$ 建模的随机热力学系统，该模型的状态空间是 $S = \{1, \dots, N\}$ 转移概率矩阵是 $P = (p_{ij})_{i,j \in S}$ ，其中 p_{ij} 表示从状态 i 到状态 j 的转移概率。该马氏链的转移图表示为 $G = (S, E)$ ，其中顶点集合 S 是状态空间， E 是转移概率有向连接边的集合。在这篇论文里，用 $\langle i, j \rangle$ 表示状态 i 到状态 j 的有向边，所以有 $E = \{\langle i, j \rangle \in S \times S : p_{ij} > 0\}$ ，并且令 $M = |E|$ ，其中 $|E|$ 表示集合 E 中元素的数量。本文中假设马氏链是不可约的，也就是有向图 G 是连通的。因此对某个状态，图 G 不仅包含其他状态到其的边，还包括它到其自身的边，也就是自循环。

本文首先研究只有一个多于两状态的环的图 G ，如图 1(a) 所示，并称具有这种图结构的马氏链为单环。也就是说，如果马氏链的转移概率矩阵满足， $p_{ij} = 0$ 且 $|\text{mod}(i-j, N)| \geq 2$ ($\text{mod}(*, N)$ 表示模 N 同余)。生化反应中的很多物理过程也都用单环马氏链建模，比如酶的构象变化，磷酸化-脱磷酸化环，染色体重塑引起的启动子激活等。本文会过多关注单环系统，很多结论也会延展到一般系统中。

2.1 环擦除定义下的环流

本文研究马氏链中两种类型的环流，该章节回顾环擦除定义下的环流。马氏链中的回路是用路径定义的，比如路径 $i_1 \rightarrow i_2 \rightarrow \dots \rightarrow i_s \rightarrow i_1$ (i_1, i_2, \dots, i_s 是顶点集合 S 中不同点) 的环流为 $p_{i_1 i_2} p_{i_2 i_3} \dots p_{i_s i_1} > 0$ 。令 $j_1 \rightarrow j_2 \rightarrow \dots \rightarrow j_r \rightarrow j_1$ 为另一个环，若上述两个环满足存在一个整数 k 使得

$$j_1 = i_{k+1}, j_2 = i_{k+2}, \dots, j_n = i_{k+s},$$

且 $r = s$ 则称两个回路是等价的，其中指标 $k+1, k+2, \dots, k+s$ 被视为模 n 同余的。环流 $i_1 \rightarrow i_2 \rightarrow \dots \rightarrow i_s \rightarrow i_1$ 在上述等价关系下所属的等价类被表示为 (i_1, i_2, \dots, i_s) 。例如， $(1, 2, 3)$, $(2, 3, 1)$ 和 $(3, 1, 2)$ 表示相同的环。环 (i_1, i_2, \dots, i_s) 的反环被定义为 (i_1, i_s, \dots, i_2) 。通常称马氏链中所用环的集合为环空间 C 。

马氏链的一条轨迹会形成各种环。直观看，如果我们抛弃马氏链 ξ 中环，并且在该过程中，始终关注轨迹中剩余状态形成的轨道，那么称剩余的轨迹为导出链 $\tilde{\xi} = (\tilde{\xi}_n)_{n \geq 0}$ 。例如，如果马氏链 ξ 的轨迹为 $\{1, 2, 3, 3, 2, 3, 4, 1, 4, \dots\}$ ，那么相应的导出链 $\tilde{\xi}$ 的轨迹和环形成表 1 所示

特别地，导出链的状态用 S 的中状态组成的有限序列表示，即 $[i_1, i_2, \dots, i_s]$ 。假设 $\tilde{\xi}_{n-1} = [i_1, i_2, \dots, i_s]$ 且 $\xi_n = i_{s+1}$ 。若 i_{s+1} 不同于 i_1, i_2, \dots, i_s ，那么 $\tilde{\xi}_n$ 被定义为 $\tilde{\xi}_n = [i_1, i_2, \dots, i_s, i_{s+1}]$ 。其次，若 $i_{n+1} = i_r$ ，那么 $\tilde{\xi}_n$ 被定义为 $\tilde{\xi}_n = [i_1, i_2, \dots, i_r]$ 。对于这种情况，称马氏链在时间 n 形成环 $(i_r, i_{r+1}, \dots, i_s)$ 。令 N_n^c 为环 c 在时间 n 时形成的次数。那

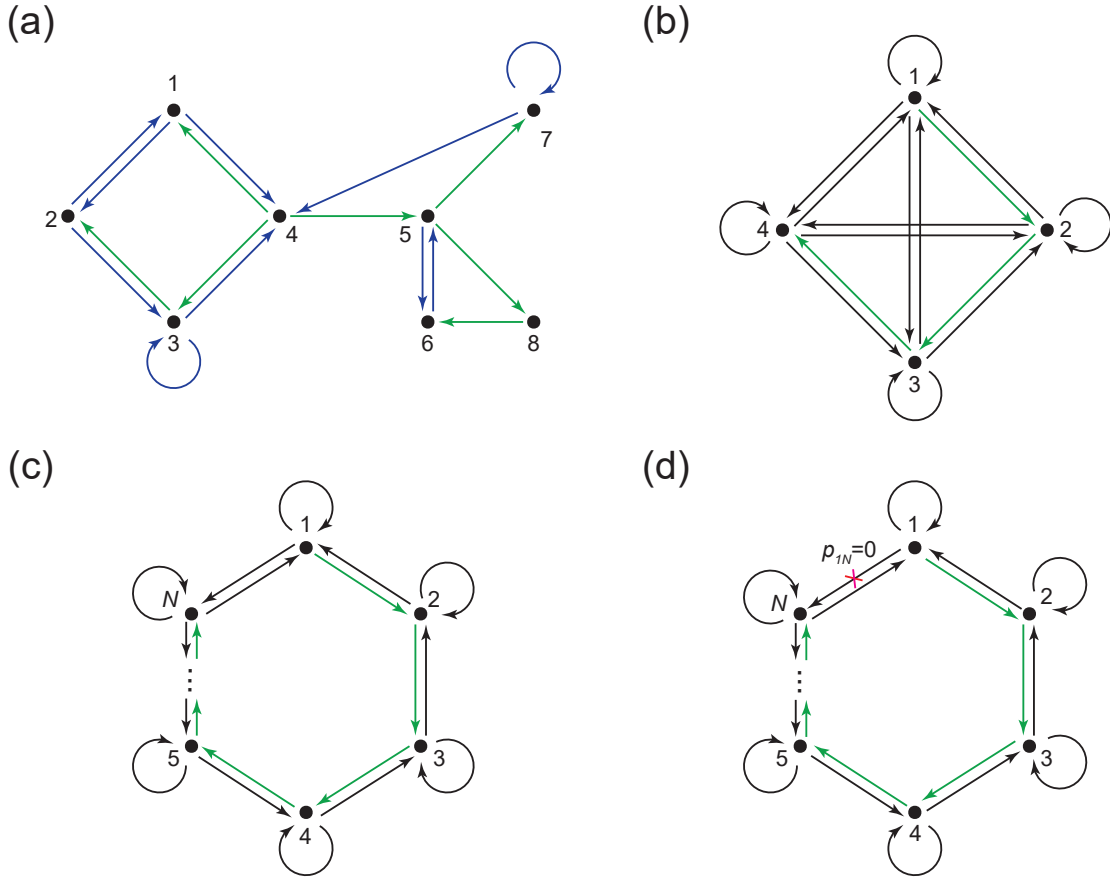


图 2.1 不同马氏链的转移图 (a) 一般马氏链的转移图，绿色线表示生成树 T ，并且蓝色线表示 T 的弦. (b) 转移概率矩阵满足 $p_{ij} > 0, \forall i, j, 1 \leq i, j \leq 4$ 的马氏链。绿色线表示生成树 $T = 1 \rightarrow 2 \rightarrow 3 \rightarrow 4$. (c) N 状态单环马氏链。绿线表示生成树 $T = 1 \rightarrow 2 \rightarrow \dots \rightarrow N$. (d) 转移概率矩阵有 $p_{1N} = 0$ 的单环马氏链，绿色线表示生成树 $T = 1 \rightarrow 2 \rightarrow \dots \rightarrow N$.

n	0	1	2	3	4	5	6	7	8
ξ_n	1	2	3	3	2	3	4	1	4
$\tilde{\xi}_n$	[1]	[1,2]	[1,2,3]	[1,2,3]	[1,2]	[1,2,3]	[1,2,3,4]	[1]	[1,4]
环				(3)	(2,3)			(1,2,3,4)	

表 2.1 导出链和形成的环

么在时间 n 时，环 c 的经验环流可以被定义为：

$$J_n^c = \frac{1}{n} N_n^c,$$

并且在时间 n 时，经验净环流可以定义为：

$$\tilde{J}_n^c = J_n^c - J_n^{c-}.$$

用更直观的表述， J_n^c 表示环 c 平均每个单位时间形成的数量， \tilde{J}^c 表示环 c 平均每个单

位时间形成的净数量。若令 $n \rightarrow \infty$ ，则有经验环流 J_n^c 和经验净环流 \tilde{J}^c 分别以概率为 1 趋近于 J^c 和 \tilde{J}^c 。极限 J^c 和 \tilde{J}^c 分别被视为环 c 的环流和净环流。关于 J_n^c 和 \tilde{J}^c 更为细致的描述，可以参考文献 [1]。著名的环流分解定理可以用上述定义表示为：

$$\pi_i p_{ij} = \sum_{c \ni \langle i, j \rangle} J^c, \quad (2-1)$$

其中 $c \ni \langle i, j \rangle$ 表示环 c 中有边 $\langle i, j \rangle$ 。

2.2 生成树定义下的环流

环的环流可以用生成树的方式定义。令 T 为 G 的一个有向子图，也就是说 T 的所有边也是 G 的边，再令 \hat{T} 表示 \bar{T} 表示与 T 有关的无向图。满足下列三个条件的 T 被称为图 G 的生成树：

- T 是 G 的覆盖子图，也就是说 T 包含 G 的所有顶点。
- \bar{T} 是连通的。
- \bar{T} 没有回路，其中无向图的回路是顶点到自身的无向路径。

易知生成树 T 包含 G 的所有顶点，并且 $|T| = N - 1$ 。接下来，会通过 T 表示生成树和它的边集合。图的生成树并不唯一，一个图可以有很多完全不同的生成树。

若有向边 $l \notin T$ ，则被称为 T 的弦。(图 1 a)。因为 $|T| = N - 1$ ，生成树 T 有 $M - N + 1$ 个弦，也就是说明图 G 中有 $M - N + 1$ 条边不会在 T 中出现。同时 \bar{T} 是连通的，并且没有环的，如果添加一根弦 l 到 T ，则导出的子图 $\overline{T \cup \{l\}}$ 会恰巧有一个回路，表示其为 C_l 。记 c_l 是从回路 C_l 中获得的环，并且和 l 保持同样的指向。由弦生成的环的集合 $\mathcal{L} = \{c_l : l \in E \setminus T\}$ 被称为基本集。很明显，弦和基本集之间没有一一对应的关系，可以用 c_l 形成的次数定义通过弦 l 的次数。那么经验时间 n 时刻， c_l 的经验环流可以被定义为：

$$Q_n^{c_l} = \frac{1}{n} \sum_{i=1}^n 1_{\{\langle \xi_{i-1}, \xi_i \rangle = l\}}.$$

$Q_n^{c_l}$ 表示单位时间通过弦 l 的次数。生成树方式和环擦除方式有很大的不同，环擦除可以很便捷的定义环流，生成树只能定义基本集的环流。

类似的，也可以用生成树定义净环流。最终，假设转移概率满足 $p_{ij} > 0$ ，当且仅当 $p_{ji} > 0$ ，这个条件可以保障熵增量是有限的。【1】考虑 l ，在时间步 n ， c_l 的净环流定义为：

$$\tilde{Q}_n^{c_l} = Q_n^{c_l} - Q_n^{c_l^{-}}.$$

如果 c_l 中只有一个或两个状态，那么易知 $c_l = c_{l^{-}}$ ，因此 $\tilde{Q}_n^{c_l} = 0$ 。对于弦 $l = \langle i, j \rangle$ ，如果 c_l 中有三个以上的状态，那么 l^{-} 也是一个弦，并且 $c_{l^{-}}$ 是由弦 l^{-} 生成的环。在文献 [3-6]，净环流的定义只是考虑有三个及以上状态的环，本文参考文献【2】中的定义，

使得净环流的定义也可以包含只有一个或两个状态的环。

若 $n \rightarrow \infty$, 则经验环流 $Q_n^{c_l}$ 和经验净环流 $\tilde{Q}_n^{c_l}$ 将会分别以概率为 1 趋于 Q^{c_l} 和 \tilde{Q}^{c_l} 。极限 Q^{c_l} 和 \tilde{Q}^{c_l} 分别被当做环 c 的环流与净环流。根据马氏链的遍历性, 有 $Q^{c_l} = \pi_i p_{ij}$ 。

2.3 两种类型环流的比较

下面将简述两种类型环流的差异。为叙述简便, 由环擦除方式定义的环流称为 LE 环流, 由生成树方式定义的环流称为 ST 环流。 LE 环流是对环空间 C 中所有环定义的, 然而 ST 环流仅是针对基本集 \mathcal{L} 定义的。因此, 对于马氏链的环动态性, LE 环流相较于 ST 环流给出了更完整的描述。而且, 由于生成树不具有唯一性, 不同的生成树选择会对应不同的 ST 环流。对比之下, LE 环流并不依赖生成树的选择。

经过上述的比对, 自然会问到基本集 \mathcal{L} 和环空间 C 的差距会有多大。因为每根弦对应集合 \mathcal{L} 唯一一个元素, 即 $|\mathcal{L}| = |E \setminus T| = M - N + 1$, 所以很难对 $|C|$ 给出一般性的表述。下面将主要考虑几个重要的特例。首先考虑转移图是全连接的马氏链 (图 1b), 也就是 $p_{ij} > 0, \forall i, j \in S$ 。有 k 个状态的环的数量是 $\frac{N(N-1)\cdots(N-K+2)(N-k+1)}{k}$, 因此

$$|C| = \sum_{k=1}^N \frac{N \cdots (N-k+1)}{k}.$$

特别地, 若 $N = 4$, 可得 $|C| = 24$, 环空间为:

$$\begin{aligned} C = \{ & (1), (2), (3), (4), (1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4), \\ & (1, 2, 3), (1, 2, 4), (1, 3, 2), (1, 3, 4), (1, 4, 2), (1, 4, 3), (2, 3, 4), (2, 4, 3), \\ & (1, 2, 3, 4), (1, 2, 4, 3), (1, 3, 2, 4), (1, 3, 4, 2), (1, 4, 2, 3), (1, 4, 3, 2) \}. \end{aligned}$$

若选择 $T = 1 \rightarrow 2 \rightarrow 3 \rightarrow 4$ 作为生成树, 那么有基本集 $|\mathcal{L}| = 13$, 且可以表示为:

$$\begin{aligned} \mathcal{L} = \{ & (1), (2), (3), (4), (1, 2), (2, 3), (3, 4) \\ & (1, 2, 3), (1, 3, 2), (2, 3, 4), (2, 4, 3), (1, 2, 3, 4), (1, 4, 3, 2) \}. \end{aligned}$$

这也说明了 LE 环流的值可能会和 ST 的值差很多。

接下来考虑单环马氏链【图 1 c】。为了叙述方便, 假设任意一对相邻状态 i 和 j 满足 $p_{ii} > 0$ 和 $p_{ij} > 0$ 。对于这种情况, $|C| = 2N + 2$, 并且环空间为:

其中前 N 个环是只包含一个状态的环, 也就是自循环的。中间 N 个环是两状态的环, 最后两个环是 N 个状态的环。如果选择 $T = 1 \rightarrow 2 \rightarrow \cdots \rightarrow N$ 作为生成树, 那么 $|\mathcal{L}| = 2N + 1$, 并且基本集是:

$$\mathcal{L} = \{ (1), \cdots, (N), (1, 2), \cdots, (N-1, N), (1, 2, \cdots, N), (1, N, \cdots, 2) \}.$$

对于这种情况, 只有一个环 $(N, 1)$ 在环空间 C 中出现, 而没有在基本集 \mathcal{L} 中出现。

最后考虑单环马氏链 [图 1 d]。从状态 1 到状态 N 的转移概率为 0。为了叙述方便, 假设 $p_{1N} = 0$, 且对于其他相邻状态 i, j , 马氏链满足 $p_{ii} > 0$ 并且 $p_{ij} > 0$ 。易知 $|C| = 2N$,

环空间为:

$$C = \{(1), \dots, (N), (1, 2), \dots, (N-1, N), (1, \dots, N)\}.$$

若令 $T = 1 \rightarrow 2 \rightarrow \dots \rightarrow N$ 为生成树, 那么有 $\mathcal{L} = C$, 也就是说环空间和基本集是一致的。而且, 很容易验证两种类型的环流也是一致的, 也就是:

$$Q_n^{c_l} = J_n^{c_l}, \quad c_l \in \mathcal{L}. \quad (2-2)$$

为了进一步理解经验 LE 环流 J_n^c 和经验 ST 环流 Q_n^c 的关系。下面使用周期边界条件假设, 也就是 $\xi_0 = \xi_1$, 这是文献 [7] 中标准的假设条件。基于这个假设, 对任意弦 l , 易得:

$$Q_n^{c_l} = \sum_{c \in C} J_n^c 1_{\{l \in c\}}. \quad (2-3)$$

注意到上述方程两边都表示弦 l 形成的速度。这表、说明经验 ST 环流可以表示为经验 LE 环流的线性组合。

第三章 环流的大偏差原理

在概率论的研究中，大偏差是与稀有事件发生概率有关的物理量。本文在此考察了经验环流的大偏差。注意到单环马氏链的经验 LE 环流 $(J_n^c)_{c \in C}$ 定义在空间：

$$\mathcal{V} = \left\{ (v^c)_{c \in C} : v^c \geq 0, \sum_{c \in C} |c| v^c = 1 \right\},$$

其中 $|c|$ 表示环 c 的长度，也就是环 c 包含的状态数量。若满足下列三个条件，则称经验环流 $(J_n^c)_{c \in C}$ 满足速率为 n 的大偏差原则：

- 对 $\forall \alpha \geq 0$ ，水平集 $\{x \in \mathcal{V} : I_J(x) \leq \alpha\}$ 是紧的。
- 对任意开集 $U \subset \mathcal{V}$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}((J_n^c)_{c \in C} \in U) \geq - \inf_{x \in U} I_J(x). \quad (3-1)$$

- 对任意闭集 $F \subset \mathcal{V}$,

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}((J_n^c)_{c \in C} \in F) \leq - \inf_{x \in F} I_J(x). \quad (3-2)$$

同样地，也可以定义经验 ST 环流的大偏差，和经验 LE/ST 净环流的大偏差。从上述条件中可以看出，(ii) 和 (iii) 表明了上述定义对 $\forall (v^c)_{c \in C} \in \mathcal{V}$ ，满足：

$$\mathbb{P}(J_n^c = v^c, \forall c \in C) \propto e^{-nI_J(v)}, \quad n \rightarrow \infty, \quad (3-3)$$

在此考虑的是所有环的大偏差，而不是单个环的大偏差。

3.1 单环马氏链 LE 环流的大偏差

一般的马氏链中，经验环流 LE 的速率函数很难求出。本文将证明经验 LE 环流的大偏差原理，并给出【图 1c】单环系统速率函数的表达式。单环系统中所有可能形成的环都在 (2) 中列出。

为简化叙述推导过程，假设系统从状态 1 出发，这并不会降低命题的一般性。为了计算经验环流的速率函数，下面将证明 (7) 对某些离散值 v 成立。若环 c 形成的速度是 k^c ，那么环 c 的经验环流是 $v^c = k^c/n$ 。为书写方便，记 k^i 替代 k^c 表示包含一个状态的环 $c = (i)$ 的形成速度。类似的，记 $k^{i,i+1}$ 表示环 $c = (i, i+1)$ 形成的速度，记 k^+ 表示环 $c = (1, 2, \dots, N)$ 形成的速度， k^- 表示环 $c = (1, N, \dots, 2)$ 形成的速度（图 1c）。类似地，本文用 $v^i, v^{i,i+1}, v^+, v^-$ 表示相应的经验环流，用 $J^i, J^{i,i+1}, J^+, J^-$ 表示相应的经验净环流。因为上述假设周期边界条件，对于某些 $k = (k^c)_{c \in C} \in \mathbb{N}^{2N+2}$ ，容易得到：

$$\begin{aligned} \mathbb{P}(J_n^c = v^c, \forall c \in C) &= \mathbb{P}(N_n^c = k^c, \forall c \in C) \\ &= |G_n(k)| \prod_{c \in C} (\gamma^c)^{k^c}, \end{aligned} \quad (3-4)$$

其中 $\gamma^c = p_{i_1 i_2} p_{i_2 i_3} \cdots p_{i_s i_1}$ 是沿着环 c 产生的所有转移概率的乘积。 $G_n(k)$ 是 n 时刻可能形成的轨迹的集合，该集合中的轨迹满足 $c \in C$ 出现恰好 k^c 次，并称这样的轨迹为容许轨迹。例如，若 $N = 3, n = 8, k^3 = k^{12} = k^{23} = k^- = 1, k^1 = k^2 = k^{13} = k^+ = 0$ ，那么有 8 个容许轨迹，在表 (2) 中已经列出。从这个例子中，可看到容许轨迹可以通过重拍环的顺序互相转化。

m	0	1	2	3	4	5	6	7	8
ξ_m	1	3	3	2	3	2	1	2	1
ξ_m	1	3	2	3	3	2	1	2	1
ξ_m	1	3	3	2	1	2	3	2	1
ξ_m	1	3	2	1	2	3	3	2	1
ξ_m	1	2	3	3	2	1	3	2	1
ξ_m	1	2	3	2	1	3	3	2	1
ξ_m	1	2	1	3	3	2	3	2	1
ξ_m	1	2	1	3	2	3	3	2	1

表 3.1 给定每种环形成的次数 k ，所有可能的轨迹

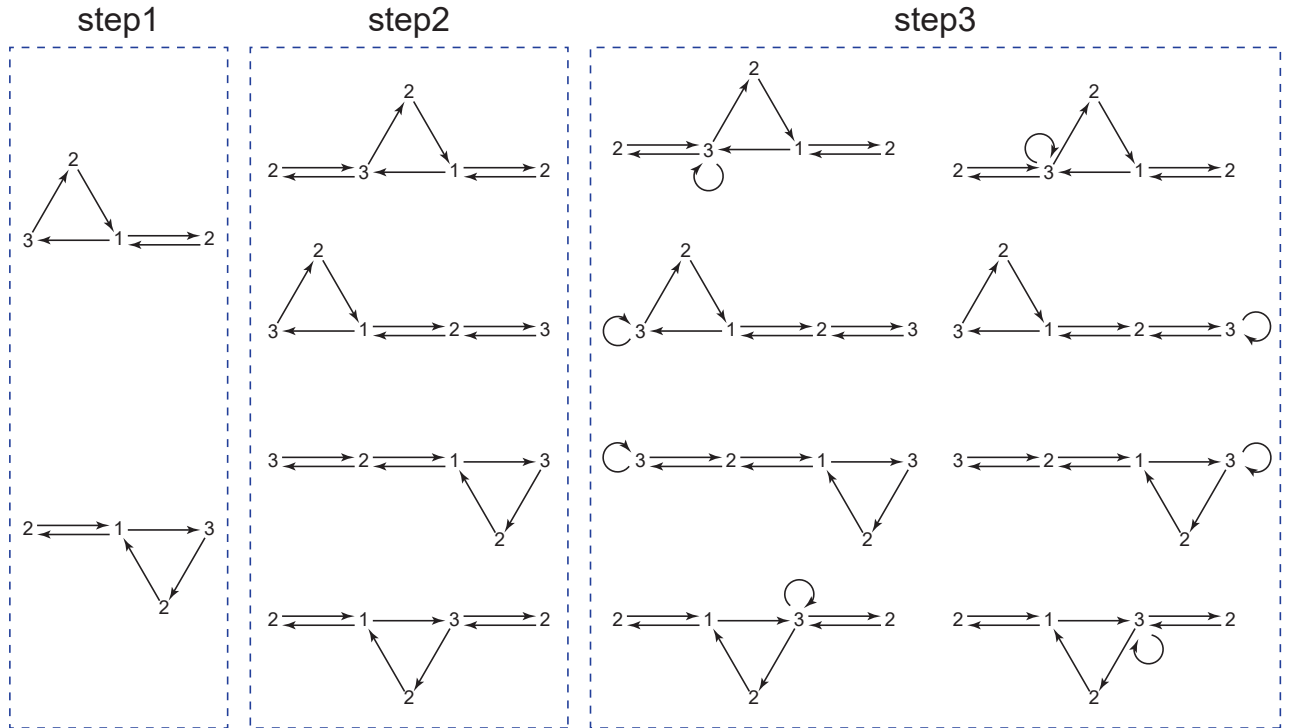


图 3.1 给定每种环形成的次数 k ，所有可能的轨迹 ($k^3 = k^{12} = k^{23} = k^- = 1, k^1 = k^2 = k^{13} = k^+ = 0$)。

所有轨迹都是从状态 1 出发。在第一步之后，顺时针向左移动

接下来，计算容许轨迹的数量 $G_n(k)$ 。计算过程分为三个步骤：

1) 已假设系统从状态 1 出发, 从环集合 C 选出通过状态 1 的环, 也就是 $(1), (1, 2), (1, N), (1, 2, \dots, N), (1, N, \dots, 2)$, 并且插入到初始状态 1 中。这些环的排列数为:

$$A_1 = \binom{k^1 + k^{12} + k^{1N} + k^+ + k^-}{k^1, k^{12}, k^{1N}, k^+, k^-} := \frac{(k^1 + k^{12} + k^{1N} + k^+ + k^-)!}{k^1! k^{12}! k^{1N}! k^+! k^-!}.$$

2) 在轨迹中插入剩余的两状态环。仔细观察系统形成环 $(i, i+1)$ 的情况, 可能在状态 i , 也可能是 $i+1$ 。例如, 轨迹 $\{1, 3, 2, 3, \dots\}$, 导出链是 1, 3 时, 形成环 $(2, 3)$ 。对此, 称这个环是在状态 3 处形成。作对比, 轨迹 $\{1, 2, 3, 2, \dots\}$ 在导出链是 1, 2 时, 形成环 $(2, 3)$, 因此称环是在状态 2 处形成的。

考虑两状态环 $i, i+1, 2 \leq i \leq N-1$, 令 l^i 和 m^i 分别表示在状态 i 和 $i+1$ 处形成环 $(i, i+1)$ 的数量, 显然 $l^i + m^i = k^{i, i+1}$ 。固定 l^i 和 m^i , 容许轨迹的数量可以计算得到。首先, 在状态 2 处插入 l^2 个环 $(2, 3)$, 总共有 $k^+ + k^{12}$ 个可能的位置可以插入, 环 $(1, 2, \dots, N)$ 和环 $(1, 2)$ 已经在步骤 1 中考虑。还要注意到, 这些位置没有包含环 $(1, N, \dots, 2)$ 中的状态 2, 因此如果插入 $(2, 3)$, 这个环将会在状态 3 处形成, 而不是状态 2 (图 2)。因此可能的插入方式数量是:

$$\binom{k^+ + k^{12} + l^2 - 1}{l^2}.$$

接下来, 把每个状态 $3 \leq i \leq N$ 对应的 l^i 个环 $(i, i+1)$ 插入轨迹中, 且每个状态 i 总共有 $k^+ + l^{i-1}$ 个可能的位置插入, 这对应了环 $(1, 2, \dots, N)$ 和环 $(i-1, i)$ 。因此可能的插入方式数量是:

$$\binom{k^+ + l^{i-1} + l^i - 1}{l^i}, \quad 3 \leq i \leq N-1.$$

最终, 把每个状态 $2 \leq i \leq N-1$ 对应的 m^i 个环 $(i, i+1)$ 插入轨迹, 所有两状态的环就已经完全被插入。步骤 2) 中所有可能的插入方式数量为:

$$A_2 = \sum_{l^2 + m^2 = k^{23}} \dots \sum_{l^{N-1} + m^{N-1} = k^{N-1, N}} \prod_{i=2}^{N-1} \binom{l^i + l^{i-1} + k^+ - 1}{l^i} \prod_{i=2}^{N-1} \binom{m^i + m^{i+1} + k^- - 1}{m^i},$$

其中 $l^1 := k^{12}$ 且 $m^N = k^{1N}$ 。

3) 最终把剩下的所有一状态的环插入轨迹中。对每个环 $(i), 2 \leq i \leq N$, 总共有 $\sum_{c \ni i} k^c - k^i$ 个可选择的位置插入。因此步骤 3) 总共的可能的插入方式数量为:

$$A_3 = \prod_{i=2}^N \binom{\sum_{c \ni i} k^c - 1}{k^i}.$$

结合上述三个步骤, 最终可以得到容许轨迹的数量为:

$$|G_n(k)| = A_1 A_2 A_3. \quad (3-5)$$

为了得到更明确的速率函数表达式, 先回顾 *Stirling* 公式:

$$\log n! = n \log n - n + O(\log n) = h(n) - n + O(\log n),$$

其中 $h(x) = x \log x, x \geq 0$ 。记 $k_i = \sum_{c \ni i} k^c$ 且 $v_i = \sum_{c \ni i} v^c$, 那么有:

$$\begin{aligned}
\log A_1 &= \log \frac{(k^1 + k^{12} + k^{1N} + k^+ + k^-)!}{k^1! k^{12}! k^{1N}! k^+! k^-!} \\
&= h(k_1) - h(k^1) - h(k^{12}) - h(k^{1N}) - h(k^+) - h(k^-) + O(\log n) \\
&= n \left[h(v_1) - h(v^1) - h(v^{12}) - h(v^{1N}) - h(v^+) - h(v^-) \right] + O(\log n).
\end{aligned} \tag{3-6}$$

最后, 估计 $\log A_2$, 令:

$$D = \{(l^i, m^i)_{2 \leq i \leq N-1} : l^i, m^i \in \mathbb{N}, l^i + m^i = k^{i,i+1}\}.$$

表示 l^i 和 m^i 可能的组合形成的集合。记 $L = (l^i, m^i) \in D$, 令

$$B_L = \prod_{i=2}^{N-1} \binom{l^i + l^{i-1} + k^+ - 1}{l^i} \binom{m^i + m^{i+1} + k^- - 1}{m^i}.$$

表示在给定 l^i 和 m^i 时, 步骤 2) 中插入方式的数量。易知 $|D| \leq (n+1)^{N-2}$, 因此, 可以得到:

$$\max_{L \in D} B_L \leq A_2 \leq (n+1)^{N-2} \max_{L \in D} B_L, \tag{3-7}$$

该式中还运用了 $A_2 = \sum_{L \in D} B_L$ 。类似于 (11) 式, 有:

$$\begin{aligned}
\log B_L &= \sum_{i=2}^{N-1} [h(l^i + l^{i-1} + k^+) - h(l^i) - h(l^{i-1} + k^+)] \\
&\quad + \sum_{i=2}^{N-1} [h(m^i + m^{i+1} + k^-) - h(m^i) - h(m^{i+1} + k^-)] + O(\log n) \\
&= \sum_{i=2}^{N-1} n[h(x^i + x^{i-1} + v^+) - h(x^i) - h(x^{i-1} + v^+)] \\
&\quad + \sum_{i=2}^{N-1} n[h(y^i + y^{i+1} + v^-) - h(y^i) - h(y^{i+1} + v^-)] + O(\log n),
\end{aligned} \tag{3-8}$$

其中 $x^i = l^i/n$ 和 $y^i = m^i/n$ 。对于给定的 $v \in \mathcal{V}$, 考虑其到处的空间:

$$V(v) = \{(x^i, y^i)_{2 \leq i \leq N-1} : x^i, y^i \geq 0, x^i + y^i = v^{i,i+1}\},$$

且记 $X = (x^i, y^i) \in V(v)$ 。定义函数:

$$\begin{aligned}
F_v(X) &= \sum_{i=2}^{N-1} [h(x^{i-1} + v^+) + h(x^i) - h(x^{i-1} + x^i + v^+)] \\
&\quad + \sum_{i=2}^{N-1} [h(y^i) + h(y^{i+1} + v^-) - h(y^i + y^{i+1} + v^-)].
\end{aligned} \tag{3-9}$$

那么联系 (12) 式的结果, 可知:

$$\log A_2 = \max_{L \in D} \log B_L + O(\log n) = n \sup_{X \in V(v)} F_v(X) + O(\log n). \tag{3-10}$$

结合(??), (??) 和 (??)可得:

$$\begin{aligned}
I_J(\nu) &= -\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(J_n^c = \nu^c, \forall c \in C) \\
&= -\lim_{n \rightarrow \infty} \frac{1}{n} \left[\log A_1 + \log A_2 + \log A_3 + \sum_{c \in C} k^c \log \gamma^c \right].
\end{aligned}$$

再联系(??), (??), 和 (??)式, 可得:

$$\begin{aligned}
I_J(\nu) &= \left[h(\nu^{12}) + h(\nu^{1N}) + h(\nu^+) + h(\nu^-) - h(\nu^{12} + \nu^{1N} + \nu^+ + \nu^-) \right] \\
&\quad + \inf_{X \in V(\nu)} F_\nu(X) + \sum_{i \in S} \left[h(\nu_i - \nu^i) + h(\nu^i) - h(\nu_i) \right] - \sum_{c \in C} \nu^c \log \gamma^c,
\end{aligned} \tag{3-11}$$

其中 $h(x) = x \log x$ 且 $\nu_i = \sum_{c \ni i} \nu^c$ 。这就是经验 LE 完整的环流速率函数表达式。该式中的 $\inf_{X \in V(\nu)} F_\nu(X)$ 难以直接计算, 不过可以通过拉格朗日乘子法得到。在附录 A 中, 证明了

$$\inf_{X \in V(\nu)} F_\nu(X) = F_\nu(x^i, y^i),$$

其中 $x^1 = \nu^{12}$, $x^N = 0$, $y^1 = 0$, 并且 $y^N = \nu^{1N}$ 。

上述论证, 都是基于系统的初始状态是 1, 自然会联想到, 基于其他初始分布, 速率函数的表达式是否与其一致。附录 B 给出了速率函数不依赖初始状态的证明, 由 (??) 的形式, 可以看出这是一个极不平凡的结论。

为建立经验 LE 环流大偏差原理, 依然要证明速率函数的水平集是紧的, 也就是验证条件 (??) 和 (??)。大偏差原理严格的证明在附录 ?? 中给出。总结上面的论述, 可以得到下面的定理

定理 3.1 单环马氏链的经验环流 $(J_n^c)_{c \in C}$ 满足大偏差原理, 并且相应的速率函数 $I_J: \mathcal{V} \rightarrow [0, \infty]$ 满足上式 (??)。速率函数 I_J 满足有界性, 连续性和凸性。并且, 速率函数 I_J 并不依赖初始分布的选择。

一般单环马氏链的速率函数表达式 (??) 十分复杂。不过, 如果令状态 1 到状态 N 的转移概率为 0, 或者只考虑三状态的马氏链。若 $N = 3$, 则速率函数可以化简为: (证明细节见 ??)

$$I_J(\nu) = \sum_{i \in S} \left[\nu^i \log \left(\frac{\nu^i / \nu_i}{J^i / J_i} \right) + (\nu_i - \nu^i) \log \left(\frac{(\nu_i - \nu^i) / \nu_i}{(J_i - J^i) / J_i} \right) \right] + \sum_{c \in C, |c| \neq 1} \nu^c \log \left(\frac{\nu^c / \tilde{\nu}}{J^c / \tilde{J}} \right),$$

其中

$$\begin{aligned}
\tilde{\nu} &= \sum_{c \in C, |c| \neq 1} \nu^c = \nu^{12} + \nu^{13} + \nu^{23} + \nu^+ + \nu^-, \\
\tilde{J} &= \sum_{c \in C, |c| \neq 1} J^c = J^{12} + J^{13} + J^{23} + J^+ + J^-.
\end{aligned}$$

若 $p_{1N} = 0$, 速率函数可以化简为 (证明细节见 ??):

$$I_J(\nu) = \sum_{i \in S} \left[\nu^i \log \left(\frac{\nu^i / \nu_i}{J^i / J_i} \right) + \nu^{i,i+1} \log \left(\frac{\nu^{i,i+1} / \nu_i}{J^{i,i+1} / J_i} \right) + (\nu^{i-1,i} + \nu^+) \log \left(\frac{(\nu^{i-1,i} + \nu^+) / \nu_i}{(J^{i-1,i} + J^+) / J_i} \right) \right]. \quad (3-12)$$

最好, 考虑经验经验净 LE 环流 $(\tilde{J}_n^c)_{c \in C}$ 的大偏差原理。因为 $\tilde{J}_n^c = 0, |S| = 1, 2$ 且 $\tilde{J}_n^+ = -J_n^-$, 所以只需要考虑环 $(1, 2, \dots, N)$ 的经验净环流。那么由收缩原理可以得到:

$$\begin{aligned} \mathbb{P}(\tilde{J}_n^+ = x) &= \mathbb{P}(J_n^+ - J_n^- = x) \\ &= \sum_{\nu^+ - \nu^- = x} \mathbb{P}(J_n^c = \nu^c, \forall c \in C) \\ &\propto \sum_{\nu^+ - \nu^- = x} e^{-nI_J(\nu)}. \end{aligned} \quad (3-13)$$

由此说明 \tilde{J}_n^+ 满足大偏差原则, 相应的速率函数 $I_J: \mathbb{R} \rightarrow [0, \infty]$ 为:

$$I_J(x) = \inf_{\{\nu \in \mathcal{V}: \nu^+ - \nu^- = x\}} I_J(\nu).$$

从上述表达式, 也可以看出 I_J 与初始分布的选择无关。

3.1.1 一般马氏链的 ST 环流的大偏差

ST 经验净环流的大偏差, 以及速率函数的对称性已经在文献 [?] 中有过相关介绍。本文将着重研究经验 ST 环流的大偏差原理, 涉及对经验测度部分可参考 [?]。

n 时刻的对经验测度 $R_n: E \rightarrow [0, 1]$, 可以定义为

$$R_n(i, j) = \frac{1}{n} \sum_{m=1}^n 1_{\{\xi_{m-1}=i, \xi_m=j\}},$$

其中假设周期边界条件 $X_0 = X_n$ 。显然 $R_n(i, j)$ 表示边 $\langle i, j \rangle$ 形成的速度。注意到, 对经验测度 R_n 处于空间

$$\mathcal{M} = \left\{ R: E \rightarrow [0, 1]: \sum_{i, j \in S} R(i, j) = 1, \sum_{j \in S} R(i, j) = \sum_{j \in S} R(j, i) \right\}.$$

众所周知, 对经验测度满足下面的大偏差原则:

$$\mathbb{P}(R_n(i, j) = R(i, j), \forall \langle i, j \rangle \in E) \propto e^{-nI_{\text{pair}}(R)}, \quad n \rightarrow \infty,$$

上式中的速率函数 $I_{\text{pair}}: \mathcal{M} \rightarrow [0, \infty]$ 的表达式为

$$I_{\text{pair}}(R) = \sum_{\langle i, j \rangle \in E} R(i, j) \log \frac{R(i, j)}{R(i)p_{ij}}$$

其中 $R(i) = \sum_{j \in S} R(i, j)$, 可以看到, 对经验测度的速率函数可以表示为有着明显的相对熵形式。给定生成树 T , 定义在空间 E 上的函数 H^{c_l} 为:

$$H^{c_l}(i, j) = \begin{cases} 1, & \text{if } \langle i, j \rangle \in c_l \text{ and } \langle i, j \rangle \in T, \text{ or } \langle i, j \rangle = l, \\ -1, & \text{if } \langle i, j \rangle \notin c_l, \langle j, i \rangle \in c_l, \text{ and } \langle i, j \rangle \in T, \\ 0, & \text{otherwise.} \end{cases}$$

对经验测度可以被分解为下面 ST 环流的加权和：

$$R_n(i, j) = \sum_{c_l \in \mathcal{L}} H^{c_l}(i, j) Q_n^{c_l},$$

且由文献 [?] 可知，该分解是唯一的。那么有：

$$\mathbb{P}(Q_n^{c_l} = \mu^{c_l}, \forall c_l \in \mathcal{L}) = \mathbb{P}\left(R_n(i, j) = \sum_{c_l \in \mathcal{L}} \mu^{c_l} H^{c_l}(i, j), \forall \langle i, j \rangle \in E\right) \propto e^{-n I_{\text{pair}}(\sum_{c_l \in \mathcal{L}} \mu^{c_l} H^{c_l})}.$$

这表明 ST 经验环流 $(Q_n^{c_l})_{c_l \in \mathcal{L}}$ 满足大偏差原则，相应的速率函数 $I_Q : \mathcal{M} \rightarrow [0, \infty]$ 为：

$$I_Q(\mu) = I_{\text{pair}}\left(\sum_{c_l \in \mathcal{L}} \mu^{c_l} H^{c_l}\right). \quad (3-14)$$

最终，考虑 ST 经验净环流 $(\tilde{Q}_n^{c_l})_{c_l \in \mathcal{L}}$ 的大偏差原理。考虑基本集中三状态或者更多状态的环，令 $l_1, l_1-, l_2, l_2-, \dots, l_s, l_s-$ 为对应的弦。类似于 LE 经验净环流的论证，只考虑环 $c_{l_1}, c_{l_2}, \dots, c_{l_s}$ 的经验净环流。则 $(\tilde{Q}_n^{c_{l_i}})_{1 \leq i \leq s}$ 满足大偏差原则，且相应的速率函数 $I_{\tilde{Q}} : \mathbb{R}^s \rightarrow [0, \infty]$ 为：

$$I_{\tilde{Q}}(x) = \inf_{\{\mu \in \mathcal{M}; \mu^{c_{l_i}} - \mu^{c_{l_i-}} = x_i, \forall 1 \leq i \leq s\}} I_Q(\mu). \quad (3-15)$$

前面论述已得出单环马氏链 LE 经验环流的速率函数，和一般马氏链的 ST 经验环流的速率函数，下面将探讨两者之间的关系。前面也说过 ST 环流可以通过 LE 表示为 $Q_n^{c_l} = \sum_{c \in \mathcal{C}} J_n^c 1_{\{l \in c\}}$ 。从收缩原理中，可以得到：

$$\begin{aligned} \mathbb{P}(Q_n^{c_l} = \mu^{c_l}, \forall l \in \mathcal{L}) &= \mathbb{P}\left(\sum_{c \in \mathcal{C}} J_n^c 1_{\{l \in c\}} = \mu^{c_l}, \forall l \in \mathcal{L}\right) \\ &= \sum_{\{v \in \mathcal{V}; \sum_c v^c 1_{\{l \in c\}} = \mu^{c_l}\}} \mathbb{P}(J_n^c = v^c, \forall c \in \mathcal{C}) \\ &\propto \sum_{\{v \in \mathcal{V}; \sum_c v^c 1_{\{l \in c\}} = \mu^{c_l}\}} e^{-n I_J(v)}. \end{aligned}$$

这表明 LE 和 ST 经验环流的内部关系为：

$$I_Q(\mu) = \inf_{\{v \in \mathcal{V}; \sum_c v^c 1_{\{l \in c\}} = \mu^{c_l}\}} I_J(v).$$

易知上式中的 I_Q 和式 (??) 中的一致，故该式的证明在此省略。

由于 LE 环流的定义相比于 ST 环流更为精确，所以多数情况下， LE 经验环流的速率函数有别于 ST 环流。然而，对于图 ??(d) 所示的单环系统，基本集 \mathcal{L} 与环空间 \mathcal{C} 完全一致。那么结合 (??) 式，可得 LE 经验环流 $(J_n^{c_l})_{c_l \in \mathcal{L}}$ 和 ST 经验环流 $(Q_n^{c_l})_{c_l \in \mathcal{L}}$ 恰好相等，并且

$$I_J(v) = I_Q(v),$$

其中 $I_J(v)$ 已经在 (??) 中给出。

第四章 环流的涨落定理

4.1 单环马氏链 LE 环流的涨落定理

本章节将验证前面讨论的经验环流是否满足各类涨落定理。LE 经验环流的暂态涨落定理已经在文献 [?] 中证明，不过该文中的证明还有些许不完整，下面先简述该文中关于单环马氏链的证明。记单环系统中两个 N 状态环为 $C^+ = (1, 2, \dots, N)$ 和 $C^- = (1, N, \dots, 2)$ ，令 N_n^+ 和 N_n^- 分别表示 n 时刻环 C^+ 和 C^- 分别形成的数量。由文献 (??) 可以得到：

$$\begin{aligned} \mathbb{P}(N_n^+ = k^+, N_n^- = k^-, N_n^c = k^c, \forall c \neq C^+, C^-) \\ = (\gamma^+)^{k^+} (\gamma^-)^{k^-} \prod_{c \neq C^+, C^-} (\gamma^c)^{k^c} |G_n(k^+, k^-, (k^c)_{c \neq C^+, C^-})|, \end{aligned}$$

其中 $\gamma^+ = p_{12}p_{23} \cdots p_{N1}$ 和 $\gamma^- = p_{1N}p_{N,N-1} \cdots p_{21}$ 分别是环 C^+ 和环 C^- 中的转移概率的乘积， $G_n(k^+, k^-, (k^c)_{c \neq C^+, C^-})$ 表示所有容许轨迹的集合，这些轨迹满足 C^+ 形成 k^+ 次， C^- 形成 k^- 次，环 $c \neq C^+, C^-$ 形成 k^c 次。上述方程可以重写为：

$$\mathbb{P}(N_n^+ = k^+, N_n^- = k^-, \dots) = (\gamma^+)^{k^+} (\gamma^-)^{k^-} \prod_{c \neq C^+, C^-} (\gamma^c)^{k^c} |G_n(k^+, k^-, \dots)|. \quad (4-1)$$

类似地，如果交换上式中的 k^+ 和 k^- ，可以得到

$$\mathbb{P}(N_n^+ = k^-, N_n^- = k^+, \dots) = (\gamma^+)^{k^-} (\gamma^-)^{k^+} \prod_{c \neq C^+, C^-} (\gamma^c)^{k^c} |G_n(k^-, k^+, \dots)|. \quad (4-2)$$

文献 [?] 中的证明是基于 $G_n(k^+, k^-, \dots)$ 与 $G_n(k^-, k^+, \dots)$ 之间的一一对应关系的，即对 $G_n(k^+, k^-, \dots)$ 的任意轨迹 $(\xi_0, \xi_1, \dots, \xi_n)$ 都可以在 $G_n(k^-, k^+, \dots)$ 中找到对应的逆轨迹 $(\xi_n, \xi_{n-1}, \dots, \xi_0)$ 。也就是说轨迹中环 C^+ 形成 k^+ 次，环 C^- 形成 k^- 次，相应的逆轨迹中环 C^+ 形成 k^- 次，环 C^- 形成 k^+ 次。同时也可得到两个集合中的元素数量相同，即：

$$|G_n(k^+, k^-, \dots)| = |G_n(k^-, k^+, \dots)|. \quad (4-3)$$

结合 (??)，(??)，和(??)，可以得到 LE 经验环流的暂态涨落定理：

$$\mathbb{P}(N_n^+ = k^+, N_n^- = k^-, \dots) = \mathbb{P}(N_n^+ = k^-, N_n^- = k^+, \dots) \left(\frac{\gamma^+}{\gamma^-} \right)^{k^+ - k^-}. \quad (4-4)$$

然而，上面也指出直观的想法是基于周期边界条件。例如，表?? 中的四状态单环系统的轨迹，其中环 C^+ 形成一次，然而相应的反环并没有形成 C^+ 和 C^- 。这说明逆轨迹并没有 $G_n(k^+, k^-, \dots)$ 和 $G_n(k^-, k^+, \dots)$ 之间的一一对应关系。通过上述论证，确实可以通过假设周期边界条件简化，简化论证。只是逆轨迹无法做到两个集合之间的一一映射，但这并不能否认暂态涨落定理 (??) 是错误的。在附录 E 中，给出了暂态涨落定理的严格证明。该证明的出发点是式 (??) 中 $|G_n(k)|$ 的非平凡对称性。因为证明过于复杂，所以放在附录部分。这表明 LE 经验环流的联合分布满足一种非平凡对称性。

n	0	1	2	3	4	5	6	7
trajectory	1	2	3	4	4	1	4	3
derived chain	[1]	[1,2]	[1,2,3]	[1,2,3,4]	[1,2,3,4]	[1]	[1,4]	[1,4,3]
cycles formed					(4)	(1,2,3,4)		

n	0	1	2	3	4	5	6	7
reversed trajectory	3	4	1	4	4	3	2	1
derived chain	[3]	[3,4]	[3,4,1]	[3,4]	[3,4]	[3]	[3,2]	[3,2,1]
cycles formed				(1,4)	(4)	(3,4)		

表 4.1 环擦除方式形成环的例子

通过暂态涨落定理，可以进一步得到其他两种涨落定理。首先，回顾 LE 经验环流的矩母函数：

$$g_n(\lambda^+, \lambda^-, \dots) = \mathbb{E} \left[e^{\lambda^+ N_n^+ + \lambda^- N_n^- + \sum_{c \in C^+, C^-} \lambda^c N_n^c} \right].$$

可以得出 Kurchan-Lebowitz-Spohn 类型的涨落定理成立：

$$\begin{aligned}
g_n(\lambda^+, \lambda^-, \dots) &= \sum_k e^{\sum_{c \in C} \lambda^c k^c} \mathbb{P}(\dots, N^+ = k^+, N^- = k^-) \\
&= \sum_k e^{\sum_{c \in C} \lambda^c k^c} \mathbb{P}(\dots, N^+ = k^-, N^- = k^+) \left(\frac{\gamma^+}{\gamma^-} \right)^{k^+ - k^-} \\
&= \sum_k e^{\dots + (\lambda^+ - \log \frac{\gamma^+}{\gamma^-}) k^+ + (\lambda^- - \log \frac{\gamma^-}{\gamma^+}) k^-} \mathbb{P}(\dots, N^+ = k^-, N^- = k^+) \\
&= \mathbb{E} \left[e^{\dots + (\lambda^- - \log \frac{\gamma^+}{\gamma^-}) N_n^+ + (\lambda^+ + \log \frac{\gamma^+}{\gamma^-}) N_n^-} \right] \\
&= g_n \left(\lambda^- - \log \frac{\gamma^+}{\gamma^-}, \lambda^+ + \log \frac{\gamma^+}{\gamma^-}, \dots \right),
\end{aligned}$$

其中 $\log(\gamma^+/\gamma^-)$ 是环 C^+ 的匹配度。接下来，考虑 n 趋于无穷是时，单环系统的极限情况，易知：

$$\begin{aligned}
e^{-nI_J(\dots, \nu^+, \nu^-)} &\propto \mathbb{P}(\dots, J_n^+ \approx \nu^+, J_n^- \approx \nu^-) \\
&= \mathbb{P}(\dots, J_n^+ \approx \nu^-, J_n^- \approx \nu^+) \left(\frac{\gamma^+}{\gamma^-} \right)^{n(\nu^+ - \nu^-)} \\
&\propto e^{-n \left[I_J(\dots, \nu^-, \nu^+) - \left(\log \frac{\gamma^+}{\gamma^-} \right) (\nu^+ - \nu^-) \right]}.
\end{aligned}$$

这蕴含着下面 Gallavotti-Cohen 形式的涨落定理成立：

$$I_J(\dots, \nu^+, \nu^-) = I_J(\dots, \nu^-, \nu^+) - \left(\log \frac{\gamma^+}{\gamma^-} \right) (\nu^+ - \nu^-). \quad (4-5)$$

类似地，可以得到 LE 经验净环流的涨落定理。类似于章节 ?? 中讨论 $(\tilde{J}_n^c)_{c \in C}$ ，只需关注环 C^+ 的经验净环流。令 $\tilde{g}_n(\lambda) = \mathbb{E}[e^{\lambda n \tilde{J}_n^+}]$ 为 \tilde{J}_n^+ 的矩母函数， $I_J(x)$ 为相应的速率函数。下面各种有关的涨落皆可得出：1) 暂态涨落定理：

$$\frac{\mathbb{P}(\tilde{J}_n^+ = x)}{\mathbb{P}(\tilde{J}_n^+ = -x)} = \left(\frac{\gamma^+}{\gamma^-}\right)^{nx}.$$

2) Kurchan-Lebowitz-Spohn 类型的涨落定理：

$$\tilde{g}_n(\lambda) = \tilde{g}_n\left(-\left(\lambda + \log \frac{\gamma^+}{\gamma^-}\right)\right).$$

3) 积分涨落定理：取 $\lambda = -\log \gamma^+/\gamma^-$ 带入上式 2) 可得

$$\mathbb{E}[e^{\lambda n \tilde{J}_n^+}] = 1.$$

4) Gallavotti-Cohen 类型的涨落定理：

$$\tilde{I}_J(x) = \tilde{I}_J(-x) - \left(\log \frac{\gamma^+}{\gamma^-}\right)x.$$

最后考虑熵产生涨落和环流涨落之间的关系。回顾熵产生的定义 [?]：

$$W_n = \frac{1}{2} \sum_{c \in C} \tilde{J}_n^c \log \frac{\gamma^c}{\gamma^{c-}} = \tilde{J}_n^+ \log \frac{\gamma^+}{\gamma^-}.$$

熵产生的暂态涨落定理可以整理为：

$$\frac{\mathbb{P}(W_n = x)}{\mathbb{P}(W_n = -x)} = \frac{\mathbb{P}(\tilde{J}_n^+ = x/(\log \frac{\gamma^+}{\gamma^-}))}{\mathbb{P}(\tilde{J}_n^+ = -x/(\log \frac{\gamma^+}{\gamma^-}))} = e^{nx}.$$

其他形式的涨落定理也有类似形式，在此省略。

4.2 单环马氏链 ST 环流的涨落定理

有关 LE 经验环流和 LE 经验净环流的涨落定理已经很完整了。自然会想到是否 ST 经验环流和 ST 经验净环流是否也满足相似的涨落定理。下面将通过一个例子说明，ST 经验环流不具有 Gallavotti-Cohen 类型的涨落定理，即使只考虑单环系统。

考虑三状态马氏链，并且令 $T = 1 \rightarrow 2 \rightarrow 3$ 是一个生成树，那么相应基本集为：

$$\mathcal{L} = \{(1), (2), (3), (1, 2), (2, 3), (1, 2, 3), (1, 3, 2)\}.$$

回顾式 (??) 中的 ST 经验环流的速率函数为：

$$I_Q(\mu) = \sum_{\langle i, j \rangle \in E} R^\mu(i, j) \log \frac{R^\mu(i, j)}{R^\mu(i) p_{ij}},$$

其中 $R^\mu(i, j) = \sum_{c_l \in \mathcal{L}} \mu^{c_l} H^{c_l}(i, j)$ 且 $R^\mu(i) = \sum_{j \in S} R^\mu(i, j)$ 。再记 $\mu^+ = \mu^{(1,2,3)}$ ， $\mu^- = \mu^{(1,3,2)}$ ，易知（参考图 ?? (a)）

$$I_Q(\cdots, \mu^+, \mu^-) \neq I_Q(\cdots, \mu^-, \mu^+) - \left(\log \frac{\gamma^+}{\gamma^-}\right)(\mu^+ - \mu^-),$$

这说明了 Gallavotti-Cohen 类型涨落定理对于 ST 环流不成立。且由于 Gallavotti-Cohen

类型的涨落定理成立条件最弱，所以其他类型的涨落定理也不成立。ST 经验环流无法

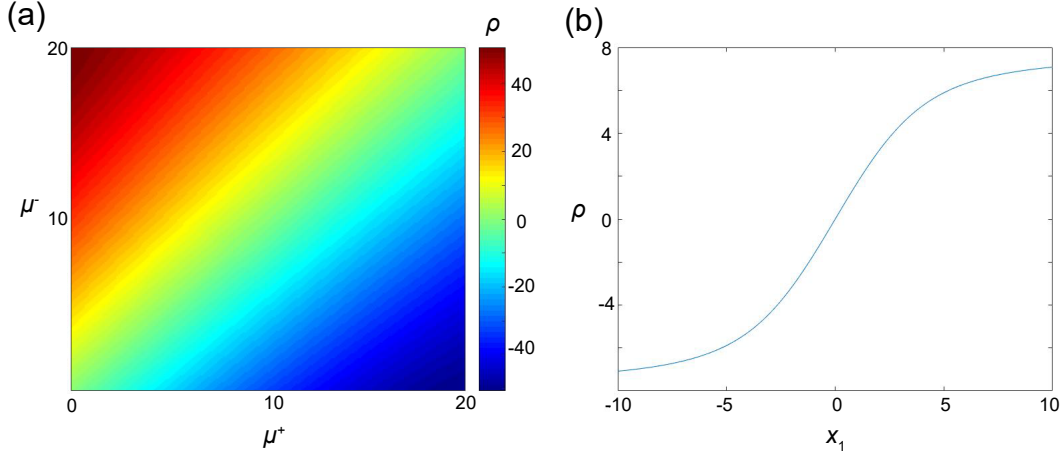


图 4.1 ST 经验环流和经验净环流的速率函数图像 (a)

$$\rho = I_Q(\cdots, \mu^+, \mu^-) - I_Q(\cdots, \mu^-, \mu^+) + (\log \frac{\gamma^+}{\gamma^-})(\mu^+ - \mu^-). \quad (b)$$

$$\rho = I_{\tilde{Q}}(\tilde{\mu}^{c_1}, \tilde{\mu}^{c_2}, \tilde{\mu}^{c_3}) - I_{\tilde{Q}}(-\tilde{\mu}^{c_1}, \tilde{\mu}^{c_2}, \tilde{\mu}^{c_3}) + (\log \frac{\gamma^{c_1}}{\gamma^{c_1-}})\tilde{\mu}^{c_1}$$

满足各种涨落定理，ST 经验净环流却是在很多情况下成立（参考 [? ?]）。对于单环系统，只需考虑环 C^+ 的净环流。从 (??) 式中易知环 C^+ 的 ST 经验净环流 \tilde{Q}_n^+ 等于该环的 LE 经验净环流 \tilde{J}_n^+ 。记弦 l^+ 和 l^- 分别对应于 C^+ 和 C^- ，由于 $l^+ = l^-$ ，则：

$$\begin{aligned} \tilde{Q}_n^+ &= \sum_{c \in C} J_n^c 1_{\{l^+ \in c\}} - \sum_{c \in C} J_n^c 1_{\{l^- \in c\}} \\ &= \sum_{c \in C} J_n^c 1_{\{l^+ \in c\}} - \sum_{c \in C} J_n^{c^-} 1_{\{l^+ \in c\}} = \tilde{J}_n^+. \end{aligned}$$

因此 ST 净环流的涨落定理自然等同于 LE 环流的涨落定理。然而，(??) 式只在周期边界条件下成立，也就意味着，只能得到 ST 经验净环流满足 Gallavotti-Cohen 类型的涨落定理。并且，容易验证 \tilde{Q}_n^+ 对其他三种类型的涨落定理不成立。

4.3 一般马氏链 LE 环流的涨落定理

目前，已经验证了单环马氏链环流相关的涨落定理，自然会问到环流涨落定理是否会适用于一般的马氏链。Thus far, we have examined the fluctuation theorems of cycle currents for a monocyclic Markov chain. It is interesting to ask whether the fluctuation theorems can be generalize to a general Markov chains. Before stating the results, we first recall the following definition [?]. Let $c_1 = (i_1, i_2, \cdots, i_s)$ and $c_2 = (j_1, j_2, \cdots, j_r)$ be two cycles. Then c_1 and c_2 are called similar if $s = r$ and

$$\{i_1, i_2, \cdots, i_s\} = \{j_1, j_2, \cdots, j_r\}.$$

In other words, two cycles are similar if they pass through the same set of states. For example, the six cycles, $(1, 2, 3, 4)$, $(1, 2, 4, 3)$, $(1, 3, 2, 4)$, $(1, 3, 4, 2)$, $(1, 4, 2, 3)$, and $(1, 4, 3, 2)$,

are similar. Note that any cycle c and its reversed cycle c^- must be similar.

We first focus on the empirical LE currents $(J_n^c)_{c \in \mathcal{C}}$. Let c_1, c_2, \dots, c_r be an arbitrary family of cycles; for example, they can be chosen as all cycles in the cycle space. Suppose that cycles c_s and c_t ($1 \leq s, t \leq r$) are similar, then the following transient fluctuation theorem holds:

$$\frac{(N_n^{c_s} = k^{c_s}, N_n^{c_t} = k^{c_t}, N_n^{c_m} = k^{c_m}, \forall m \neq s, t)}{(N_n^{c_s} = k^{c_t}, N_n^{c_t} = k^{c_s}, N_n^{c_m} = k^{c_m}, \forall m \neq s, t)} = \left(\frac{\gamma^{c_s}}{\gamma^{c_t}} \right)^{k^{c_s} - k^{c_t}}. \quad (4-6)$$

This shows that if the cycles c_s and c_t are similar, then the joint distribution of empirical LE currents satisfies a nontrivial symmetric relation under the exchange of k^{c_s} and k^{c_t} . If we choose c_s and c_t be some cycle C^+ and its reversed cycle C^- , then the above equation reduces to

$$\frac{(N_n^+ = k^+, N_n^- = k^-, N_n^c = k^c, \forall c \neq C^+, C^-)}{(N_n^+ = k^-, N_n^- = k^+, N_n^c = k^c, \forall c \neq C^+, C^-)} = \left(\frac{\gamma^+}{\gamma^-} \right)^{k^+ - k^-},$$

which has the same form as the monocyclic case. Actually, the proof of (??) has been given in [?] under the restrictions that the cycles c_1, c_2, \dots, c_r considered pass through a common state $i \in S$ and the Markov chain also starts from i . Fortunately, this technical assumption can be removed, which means that (??) actually holds for an arbitrary family of cycles c_1, c_2, \dots, c_r and an arbitrary initial distribution (manuscript in preparation). Moreover, for any family of cycles c_1, c_2, \dots, c_r , it follows from (??) that for any $1 \leq s \leq r$

$$\begin{aligned} & (\tilde{J}_n^{c_s} = x^{c_s}/n, \tilde{J}_n^{c_m} = x^{c_m}/n, \forall m \neq s) \\ &= (N_n^{c_s} - N_n^{c_s^-} = x^{c_s}, N_n^{c_m} - N_n^{c_m^-} = x^{c_m}, \forall m \neq s) \\ &= \sum_{k^{c_m} - k^{c_m^-} = x^{c_m}, \forall m} (N_n^{c_s} = k^{c_s}, N_n^{c_s^-} = k^{c_s^-}, N_n^{c_m} = k^{c_m}, N_n^{c_m^-} = k^{c_m^-}, \forall m \neq s) \\ &= \sum_{k^{c_m} - k^{c_m^-} = x^{c_m}, \forall m} (N_n^{c_s} = k^{c_s^-}, N_n^{c_s^-} = k^{c_s}, N_n^{c_m} = k^{c_m}, N_n^{c_m^-} = k^{c_m^-}, \forall m \neq s) \left(\frac{\gamma^{c_s}}{\gamma^{c_s^-}} \right)^{k^{c_s} - k^{c_s^-}} \\ &= \sum_{k^{c_m} - k^{c_m^-} = x^{c_m}, \forall m} (N_n^{c_s} = k^{c_s^-}, N_n^{c_s^-} = k^{c_s}, N_n^{c_m} = k^{c_m}, N_n^{c_m^-} = k^{c_m^-}, \forall m \neq s) \left(\frac{\gamma^{c_s}}{\gamma^{c_s^-}} \right)^{x^{c_s}} \\ &= (N_n^{c_s} - N_n^{c_s^-} = -x^{c_s}, N_n^{c_m} - N_n^{c_m^-} = x^{c_m}, \forall m \neq s) \left(\frac{\gamma^{c_s}}{\gamma^{c_s^-}} \right)^{x^{c_s}} \\ &= (\tilde{J}_n^{c_s} = -x^{c_s}/n, \tilde{J}_n^{c_m} = x^{c_m}/n, \forall m \neq s) \left(\frac{\gamma^{c_s}}{\gamma^{c_s^-}} \right)^{x^{c_s}}. \end{aligned}$$

Thus we finally obtained the following transient fluctuation theorems of empirical LE currents:

$$(\tilde{J}_n^{c_s} = x_s, \tilde{J}_n^{c_m} = x_m, \forall m \neq s) = (\tilde{J}_n^{c_s} = -x_s, \tilde{J}_n^{c_m} = x_m, \forall m \neq s) e^{n x_s \log \frac{\gamma^{c_s}}{\gamma^{c_s^-}}}. \quad (4-7)$$

This shows that for any cycle c_s , the joint distribution of empirical net LE currents satisfies a symmetric relation when x_s is changed to $-x_s$. Note that $\tilde{J}_n^c = 0$ for any

one-state or two-state cycle c and $\tilde{J}_n^c = -\tilde{J}_n^{c-}$ for any three or more state cycle c . Let $c_1, c_1-, c_2, c_2-, \dots, c_{r*}, c_{r*}-$ be all three or more state cycles. If we change x_s to $-x_s$ one by one for $1 \leq s \leq r^*$ in the above equation, then we obtain

$$\begin{aligned} & \left(\tilde{J}_n^{c_1} = x_1, \tilde{J}_n^{c_2} = x_2, \dots, \tilde{J}_n^{c_{r*}} = x_{r*} \right) \\ &= \left(\tilde{J}_n^{c_1} = -x_1, \tilde{J}_n^{c_2} = -x_2, \dots, \tilde{J}_n^{c_{r*}} = -x_{r*} \right) e^{n \sum_{i=1}^{r*} x_i \log \frac{\gamma^{c_i}}{\gamma^{c_i-}}}. \end{aligned} \quad (4-8)$$

Note that (??) is much weaker than (??). As a result, in what follows, we term (??) the strong form of the transient fluctuation theorem and term (??) the weak form of the transient fluctuation theorem.

Other types of fluctuation theorems for empirical LE and net LE currents can be easily derived from the transient fluctuation theorems and are summarized as follows (here we only show the strong form). For an arbitrary family of cycles c_1, c_2, \dots, c_r , let $g_n(\lambda) = \mathbb{E}[e^{n \sum_{i=1}^r \lambda_i J_n^{c_i}}]$, $\tilde{g}_n(\lambda) = \mathbb{E}[e^{n \sum_{i=1}^r \lambda_i \tilde{J}_n^{c_i}}]$ be the moment generating function of $(J_n^{c_i})_{1 \leq i \leq r}$, $(\tilde{J}_n^{c_i})_{1 \leq i \leq r}$ respectively and let $I_J(x)$, $I_{\tilde{J}}(x)$ be the rate function of $(J_n^{c_i})_{1 \leq i \leq r}$, $(\tilde{J}_n^{c_i})_{1 \leq i \leq r}$ respectively.

Kurchan-Lebowitz-Spohn-type fluctuation theorem: For any cycles c_s and c_t ($1 \leq s, t \leq r$) are similar

$$\begin{aligned} g_n(\dots, \lambda_s, \dots, \lambda_t, \dots) &= g_n\left(\dots, \lambda_t - \log \frac{\gamma^{c_s}}{\gamma^{c_t}}, \dots, \lambda_s + \log \frac{\gamma^{c_s}}{\gamma^{c_t}}, \dots\right). \\ \tilde{g}_n(\dots, \lambda_s, \dots) &= \tilde{g}_n\left(\dots, -\left(\lambda_s + \log \frac{\gamma^{c_s}}{\gamma^{c_s-}}\right), \dots\right). \end{aligned}$$

Integral fluctuation theorem:

$$\mathbb{E}\left[e^{-n \sum_{i=1}^r (\log \gamma^{c_i} / \gamma^{c_i-}) \tilde{J}_n^{c_i}}\right] = 1.$$

Gallavotti-Cohen-type fluctuation theorem:

$$\begin{aligned} I_J(\dots, x_s, \dots, x_t, \dots) &= I_J(\dots, x_t, \dots, x_s, \dots) - \left(\log \frac{\gamma^{c_s}}{\gamma^{c_t}}\right)(x_s - x_t). \\ I_{\tilde{J}}(\dots, x_s, \dots) &= I_{\tilde{J}}(\dots, -x_s, \dots) - \left(\log \frac{\gamma^{c_s}}{\gamma^{c_s-}}\right)x_s. \end{aligned}$$

4.3.1 Fluctuation theorems of ST currents for general Markov chains

We have seen that for monocyclic Markov chains, the empirical ST currents do not satisfy various types of fluctuation theorems, while the empirical net ST currents satisfy the Gallavotti-Cohen-type fluctuation theorems. Next we focus on general Markov chains. In fact, Andrieux and Gaspard [?] have proved the weak form of Gallavotti-Cohen-type fluctuation theorem for empirical net ST currents: for all three or more state cycles $c_1, c_1-, c_2, c_2-, \dots, c_r, c_r-$ in the fundamental set \mathcal{L} , we have

$$I_{\tilde{Q}}(x_1, x_2, \dots, x_r) = I_{\tilde{Q}}(-x_1, -x_2, \dots, -x_r) - \sum_{i=1}^r x_i \log \frac{\gamma^{c_i}}{\gamma^{c_i^-}}.$$

This shows that the joint distribution of empirical net ST currents satisfies a symmetric relation when all variables x_i are changed to $-x_i$.

Interestingly, in contrast to empirical net LE currents, we find that empirical net ST currents do not satisfy the strong form of Gallavotti-Cohen-type fluctuation theorems. To give a counterexample, we consider a four-state Markov chain with a fully connected transition diagram illustrated in Fig. ??(b). Let the spanning tree be $T = 1 \rightarrow 2 \rightarrow 3 \rightarrow 4$. Since $\tilde{Q}_n^{c_l} = 0$ for all one-state or two-state cycles c_l , we only need to consider the rest of the cycles

$$c_1 = (1, 2, 3), \quad c_2 = (2, 3, 4), \quad c_3 = (1, 2, 3, 4), \quad c_4 = (1, 3, 2), \quad c_5 = (2, 4, 3), \quad c_6 = (1, 4, 3, 2).$$

Note that $\tilde{Q}_n^{c_l} = -\tilde{Q}_n^{c_l^-}$ for all three or more state cycles and $c_1 = c_4^-$, $c_2 = c_5^-$, $c_3 = c_6^-$. Recall that the rate function of empirical ST currents (??), we have the rate function for $(\tilde{Q}_n^{c_1}, \tilde{Q}_n^{c_2}, \tilde{Q}_n^{c_3})$ is given by

$$I_{\tilde{Q}}(x) = \inf_{\{\mu \in \mathcal{M}; \mu^{c_i} - \mu^{c_i^-} = x_i, \forall 1 \leq i \leq 3\}} I_{\tilde{Q}}(\mu).$$

For cycle c_1 , if we exchange $\tilde{\mu}^{c_1}$ to $-\tilde{\mu}^{c_1}$, then (see fig. ?? (b))

$$I_{\tilde{Q}}(\tilde{\mu}^{c_1}, \tilde{\mu}^{c_2}, \tilde{\mu}^{c_3}) \neq I_{\tilde{Q}}(-\tilde{\mu}^{c_1}, \tilde{\mu}^{c_2}, \tilde{\mu}^{c_3}) - \left(\log \frac{\gamma^{c_1}}{\gamma^{c_1^-}} \right) \tilde{\mu}^{c_1},$$

which shows that the strong form of the Gallavotti-Cohen-type fluctuation theorem fails for empirical net ST currents.

第五章 多尺度网络在求解椭圆方程中的应用

这里我们应用带有激活函数为 $\text{Phi}(x)$ 的多尺度网络来求解复杂的椭圆型方程。具体实验包括求解具有宽频率范围、变系数、环形区域上和多孔的立方区域上的椭圆方程。通过这些实验，我们令人信服地证明了多尺度网络是一种有效且易于实现的求解复杂椭圆方程的无网格方法。

5.1 宽频域振荡的椭圆方程

求解椭圆方程 $\Delta u = f(x)$ ，其中参数选取为 $a(x) = 1$, $c(x) = 0$ ，求解区域为 $\Omega = [-1, 1]^d$ ，右端项表达式为

$$f(x) = \sum_{i=1}^d 4\mu^2 x_i^2 \sin(\mu x_i^2) - 2\mu \cos(\mu x_i^2) \quad (5-1)$$

方程的精确解为

$$u(x) = \sum_{i=1}^d \sin(\mu x_i^2) \quad (5-2)$$

对应的边界条件由精确解生成。

每步中我们随机在区域 $[-1, 1]^d$ 内选取 5000 个均匀分布的积分点，在边界上生成 4000 个均匀分布的积分点。这个问题的特殊之处在于，精确解虽然是高频振荡的，但是它没有固定的频率，而是有一个广泛的频率范围。对于 $\mu = 15$ ，二维情况下问题的精确解和两种网络结构下的近似解如图??所示。图中指出，多尺度网络得到的解很好地捕捉到不同尺度的振荡。例如，在红圈所示区域，普通的全连接网络没有捕捉到解的振荡，而多尺度网络的解很好地展示出了振荡的形态。在四个角处的振荡也出现了类似的行为。

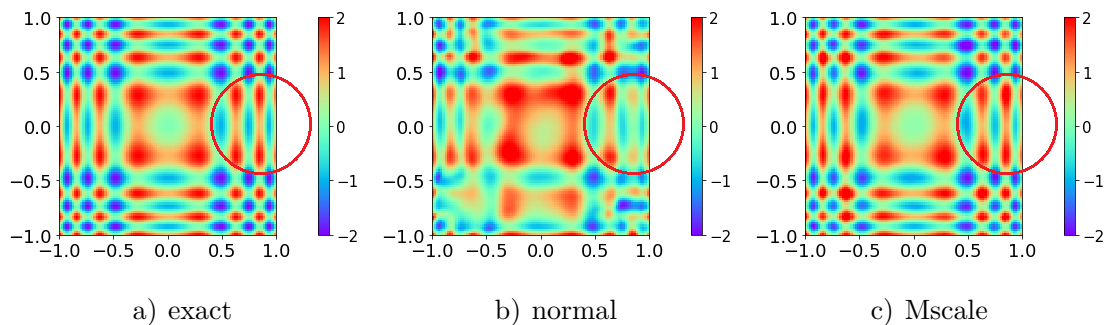


图 5.1 二维情况下椭圆方程的精确解和数值解图像

求解结果如图??，图中画出了求解方程的误差随迭代步数下降的图像。在求解二维和三维的方程中，多尺度网络都表现出了明显的优势。

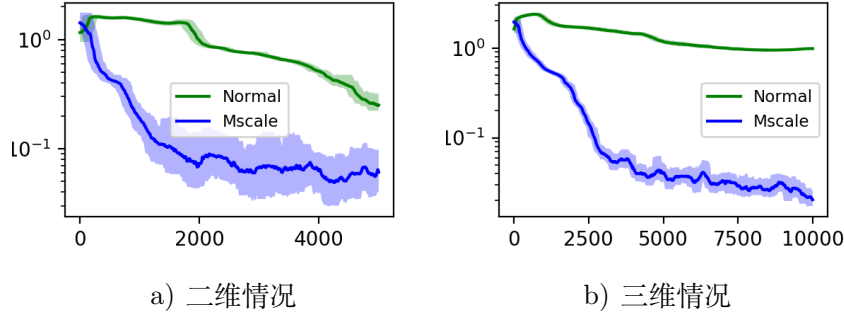


图 5.2 二维和三维情况下两种网络结构的表现

5.2 环形区域上的椭圆方程

求解椭圆方程^{??}，其中参数选取为 $a(x) = 1$ ， $c(x) = 0$ ，求解区域为一个以原点为中心的圆环，内径为 1，外径为 3。方程的右端项表达式为

$$f(x) = \mu^2 J_0(\mu|x - x_0|) \quad (5-3)$$

其中 J_0 是 Bessel 函数，方程的精确解为

$$u(x) = J_0(\mu|x - x_0|) \quad (5-4)$$

对应的边界条件由精确解生成。位移选取为 $x_0 = (0.5, 0)$ ，参数选取为 $\mu = 5$ 和 $\mu = 10$ 分别实验。

每步中我们随机在区域内选取 5000 个均匀分布的积分点，在边界上生成 4000 个均匀分布的积分点。问题的精确解和两种网络结构下的近似解如图^{??}和图^{??}所示。图中用黑色圆圈标记的区域表示解在此处有最大的振幅，正常的全连接网络完全不能刻画振荡，而多尺度网络在这两种情况下都可以得到和精确解类似的效果。

同样，如图^{??}所示，图中画出了求解方程的误差随迭代步数下降的图像。在迭代同样的步数后，多尺度网络的误差更小。

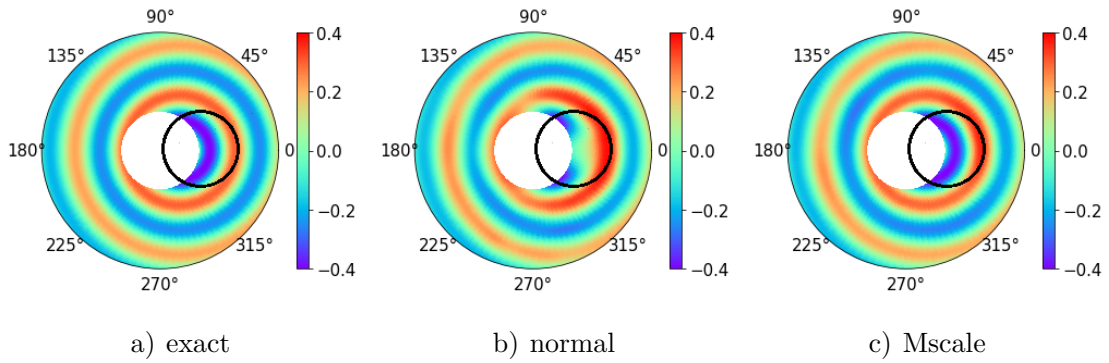


图 5.3 $\mu = 5$ 情况下椭圆方程的精确解和数值解图像

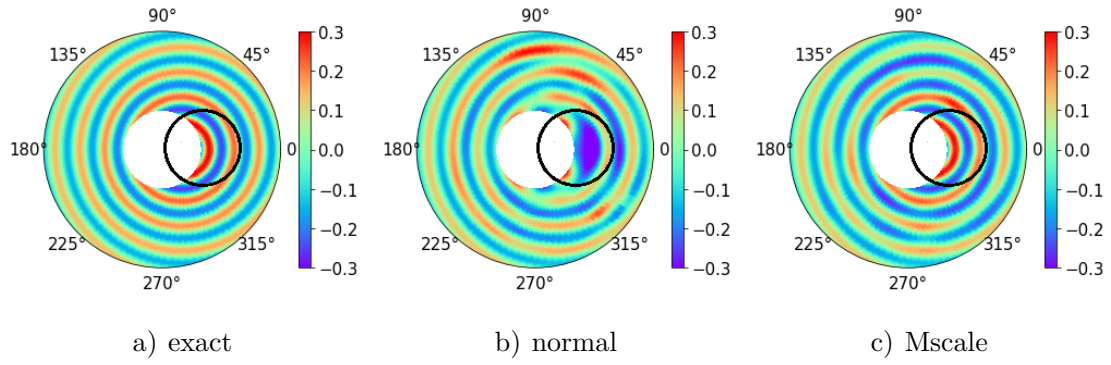


图 5.4 $\mu = 10$ 情况下椭圆方程的精确解和数值解图像

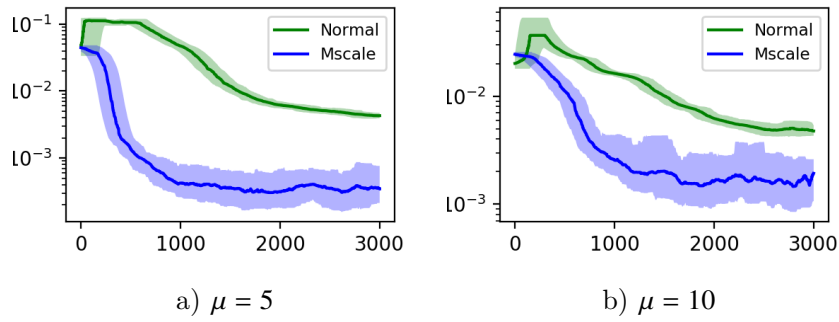


图 5.5 不同参数下两种网络结构的表现

5.3 有孔区域上的椭圆方程

这里我们考虑如下两个区域：

区域 1: 方形区域 $[-1, 1]^2$ 内有三个圆形的孔。三个圆孔的中心分别为 $(-0.5, -0.5)$ 、 $(0.5, 0.5)$ 和 $(0.5, -0.5)$ ，半径分别为 0.1、0.2 和 0.2。在训练过程中，每步我们随机在区域内选取 5000 个均匀分布的积分点，在方形区域的边界上生成 3000 个积分点，每个大圆孔的边界上生成 800 个积分点，每个小圆孔的边界上生成 400 个积分点。

区域 2 方形区域 $[-1, 1]^2$ 内有四个孔。三个圆形孔的中心分别为 $(-0.6, -0.6)$ 、 $(0.3, -0.3)$ 和 $(0.6, 0.6)$ ，半径分别为 0.3、0.6 和 0.3。椭圆形孔由方程 $16(x + 0.5)^2 + 64(y - 0.5)^2 = 1$ 表示。每步我们随机在区域内选取 5000 个均匀分布的积分点，外部边界、大圆孔边界、小圆孔边界和椭圆孔边界生成的积分点数量分别为 2400、1100、550 和 400。

求解椭圆方程 $??$ ，其中参数选取为 $a(x) = 1$ ， $c(x) = 0$ 。方程的右端项表达式为

$$f(x) = 2\mu^2 \sin \mu x_1 \sin \mu x_2 \quad (5-5)$$

方程的精确解为

$$u(x) = \sin \mu x_1 \sin \mu x_2 \quad (5-6)$$

对应的边界条件由精确解生成。其中的参数选取为 $\mu = 7\pi$ 。

问题的精确解和两种网络结构下的近似解如图??和图??所示。显然，多尺度网络能更好地刻画精确解中的振荡。

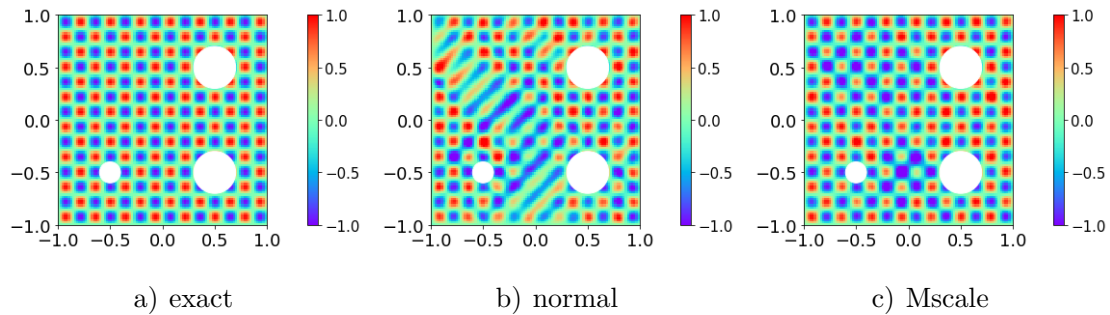


图 5.6 区域 1 上椭圆方程的精确解和数值解图像

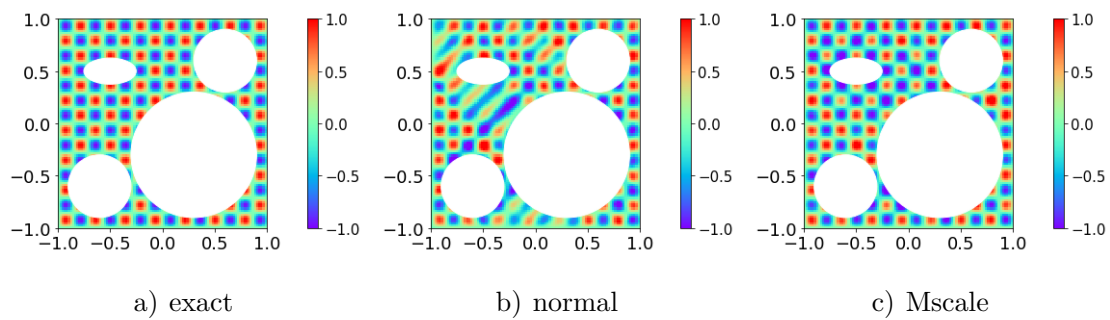


图 5.7 区域 2 上椭圆方程的精确解和数值解图像

同样，如图??所示，图中画出了求解方程的误差随迭代步数下降的图像，多尺度网络的误差下降速度更快，精度也更高。

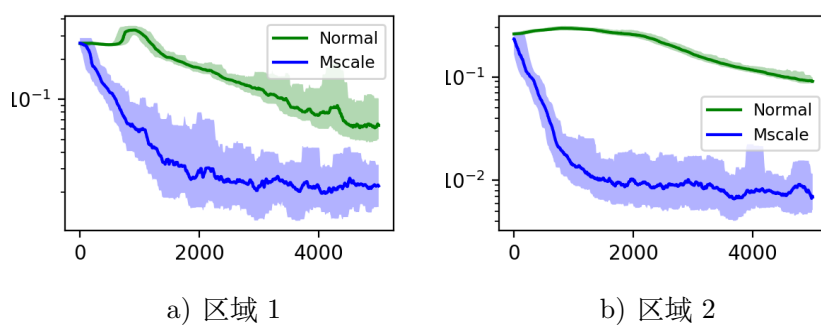


图 5.8 两种区域下不同网络结构的表现

5.4 多孔区域上的椭圆方程

为了验证多尺度网络求解复杂区域上椭圆方程的能力，我们考虑了一个三维立方体 $[-1, 1]^3$ ，内部移除了 125 个孔，如图??，这些球形孔的球心位于正方体内部的均匀网格 $\{-0.8, -0.4, 0, 0.4, 0.8\}^3$ 上。这些孔的半径是从范围在 $[0, 0.15]$ 中的均匀分布随机抽样

产生，这样可以保证各个孔之间没有相交。在训练过程中，每步我们随机在区域内选取 5000 个均匀分布的积分点，外边界生成 2500 个积分点，内部的孔上 1500 个（每个孔 12 个）积分点。

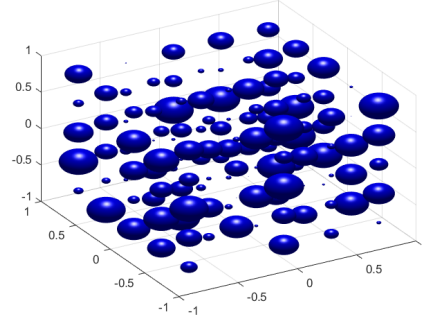


图 5.9 多孔区域示意图，蓝色表示在区域内挖去的部分

求解椭圆方程 $\Delta u = 0$ ，其中参数选取为 $a(x) = 1$, $c(x) = 0$, $\mu = 7\pi$ 。我们测试如下三组示例，它们对应的精确解分别为

1. 示例 1: $u(x) = \sin \mu x_1 \sin \mu x_2 \sin \mu x_3$

2. 示例 2: $u(x) = e^{\sin \mu x_1 + \sin \mu x_2 + \sin \mu x_3}$

3. 示例 3: $u(x) = e^{\sin \mu x_1 \sin \mu x_2 \sin \mu x_3}$

对应的边界条件和右端项由精确解生成。

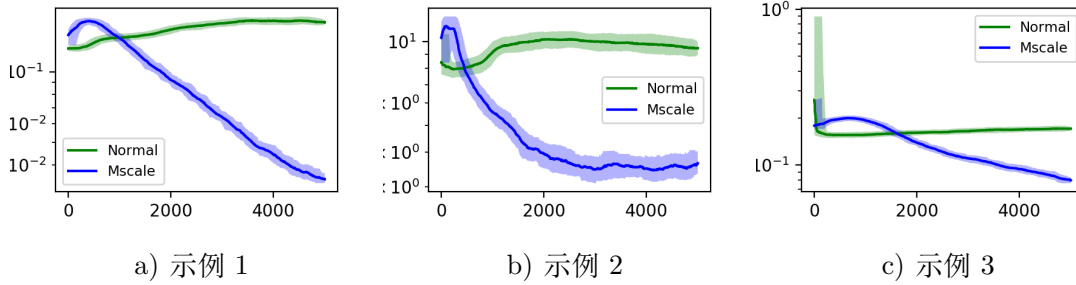


图 5.10 三个示例中不同网络结构的表现

计算结果如图??所示，在三维区域上，我们难以画出精确解和近似解的图像，这里只给出了误差随迭代步数变化的图像。对于这三种情况，正常的全连通网络结构对于此类复杂问题根本不收敛，在迭代的过程中误差几乎不下降，而多尺度网络虽然在计算开始时误差上升，但是在一定的步数之后就可以明显达到更小的误差。

5.5 变系数椭圆方程

求解椭圆方程 $\Delta u = f(x)$ ，其中参数选取为 $a(x) = 1$ ，方程的右端项表达式为

$$f(x) = (\mu_1^2 + \mu_2^2 + \mu_3^2 + x_1^2 + 2x_2^2 + 3x_3^2) \sin(\mu_1 x_1) \sin(\mu_2 x_2) \sin(\mu_3 x_3) \quad (5-7)$$

系数表达式为

$$c(x) = (x_1^2 + 2x_2^2 + 3x_3^2) \quad (5-8)$$

方程的精确解为

$$u(x) = \sin(\mu_1 x_1) \sin(\mu_2 x_2) \sin(\mu_3 x_3) \quad (5-9)$$

对应的边界条件由精确解生成。我们选取各向异性的参数，其中 $\mu_1 = 15, \mu_2 = 20, \mu_3 = 25$ 。

每步中我们随机在求解区域 $[-1, 1]^d$ 内选取 5000 个均匀分布的积分点，在边界上生成 4000 个均匀分布的积分点。

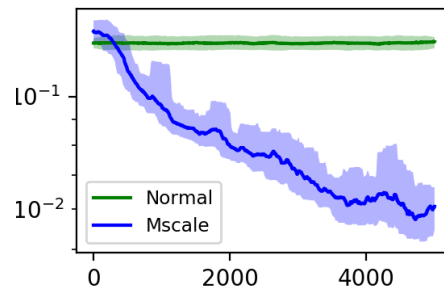


图 5.11 两种网络结构的表现

如图??所示，在训练过程中，多尺度网络的误差显著衰减，而正常的全连接网络误差几乎保持不变。多尺度网络以更高的精度更快地解决了问题。

第 六 章 多尺度网络在分子计算中的应用

这里我们展示多尺度网络在求解 Poisson-Boltzmann 方程 (PB 方程) 中的应用。PB 方程用于描述生物分子在溶液中的电势分布，在分子模拟中有重要的应用。

6.1 PB 方程简介

考虑一个有界开区域 $\Omega_1 \subset \mathbb{R}^3$ ，它将全空间 \mathbb{R}^3 分成两个不相交的子区域，曲面 $\Gamma = \partial\Omega_1$ 。在物理模型中， Ω_1 表示生物分子，而 $\Omega_2 = \mathbb{R}^3 \setminus \Omega_1$ 是溶剂区域。它们之间的边界 Γ 一般不规则。

在整个区域上，PB 方程表示为

$$-\nabla(\epsilon(x)\nabla u(x)) + \kappa(x)u(x) = f(x) \quad x \in \mathbb{R}^3 \quad (6-1)$$

其中 $\epsilon(x)$ 是离子溶剂的介电常数， $\kappa(x)$ 是离子溶剂的 Debye-Huckel 长度的倒数。在实际问题中，介电常数在界面 Γ 处一般不连续。

解在界面上的边界条件满足连续性

$$[u](x) = 0 \quad x \in \Gamma \quad (6-2)$$

和流连续性

$$[\epsilon \frac{\partial u}{\partial n}](x) = 0 \quad x \in \Gamma \quad (6-3)$$

其中 $[\cdot]$ 表示在边界上的跳跃。取无穷远点为电势零点，即

$$\lim_{|x| \rightarrow \infty} u_2(x) = 0 \quad (6-4)$$

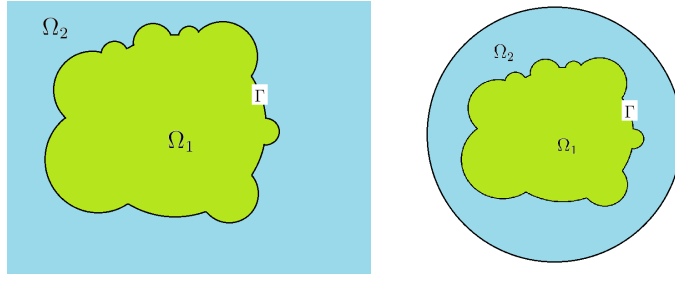
为了处理无界区域，我们将求解区域截断成一个大的球或立方体，用 Ω 表示，它满足 $\Omega_1 \subset \Omega$ 。这里我们重新定义 $\Omega_2 = \Omega \setminus \Omega_1$ ，并在 Ω 的边界上设置一个近似边界条件 $u = 0$ 。对于区域的截断，高阶的人工边界条件已经得到了广泛的研究。但是由于我们更感兴趣的是界面 Γ 附近神经网络方法的性能，这里对区域的截断比较粗糙。图??展示出了对区域的截断。

6.2 求解 PB 方程的有限差分法

这里我们介绍一种简单的有限差分方法，用于求解 PB 方程并与深度学习方法进行对比。

在 PB 方程中，我们可以定义通量为 $F(x) = \epsilon(x)\nabla u(x)$ 。尽管 $\epsilon(x)$, $\kappa(x)$ 和 $u'(x)$ 可能是不连续的，根据边界条件我们可以得到， $u(x)$ 和 $F(x)$ 连续且几乎处处可导。

考虑三维均匀网格上的数值解 $u_{i,j,k}$ ，网格步长为 h ，我们有



a) 截断前 b) 截断后

图 6.1 无界区域截断示意图

$$\frac{\partial u}{\partial x}(x_{i+\frac{1}{2}}, y_j, z_k) \approx \delta_x u_{i+\frac{1}{2}, j, k}$$

其中 $x_{i+\frac{1}{2}} = \frac{x_i + x_{i+1}}{2}$ 和 $\delta_x u_{i+\frac{1}{2}, j, k} = \frac{u_{i+1, j, k} - u_{i, j, k}}{h}$ 。这里我们可以定义数值通量为

$$(F_1)_{i+\frac{1}{2}, j, k} = \epsilon(x_{i+\frac{1}{2}}, y_j, z_k) \delta_x u_{i+\frac{1}{2}, j, k}$$

因此可以得到

$$\frac{\partial F_1}{\partial x}(x_{i+\frac{1}{2}}, y_j, z_k) \approx \delta_x (F_1)_{i, j, k}$$

这就是说

$$\frac{\partial}{\partial x}(\epsilon \frac{\partial u}{\partial x})(x_i, y_j, z_k) \approx \delta_x (\epsilon \delta_x u)_{i, j, k}$$

其中

$$\delta_x (\epsilon \delta_x u)_{i, j, k} = \epsilon(x_{i+\frac{1}{2}}, y_j, z_k) u_{i+1, j, k} - \epsilon(x_{i-\frac{1}{2}}, y_j, z_k) u_{i-1, j, k} + (\epsilon(x_{i+\frac{1}{2}}, y_j, z_k) + \epsilon(x_{i-\frac{1}{2}}, y_j, z_k)) u_{i, j, k}$$

在对其它方向进行类似的处理之后，我们可以得到最终的有限差分格式为

$$-(\delta_x (\epsilon \delta_x u)_{i, j, k} + \delta_y (\epsilon \delta_y u)_{i, j, k} + \delta_z (\epsilon \delta_z u)_{i, j, k}) + \kappa(x_i, y_j, z_k) u_{i, j, k} = f(x_i, y_j, z_k)$$

6.3 具有几何奇异性的 PB 方程

考虑到分子计算中的问题通常具有几何奇异性，我们构造具有几何奇异性的区域如下。首先选择一个中心在位于原点，半径为 0.5 的大球。在大球表面随机抽取 20 个点，以这些点为中心生成许多小球。小球的半径从 [0.1, 0.2] 中随机抽取。选取 Ω_1 是这些小球和大球的并集。 Ω_1 的形状如图??所示。各个球之间的相交会产生几何奇异性，这对传统的有限元法和边界元法的网格生成以及精确求解提出了很大的挑战。

我们在此区域上，首先考虑如下两个示例：

示例 1： 方程的精确解表示为：

$$u(x) = \frac{e^{\sin \mu x_1 + \sin \mu x_2 + \sin \mu x_3}}{|x|^2 + 1} (|x|^2 - 1) \quad (6-5)$$

方程的系数选取为

$$\mu = 15, \epsilon(x) = 1, \kappa(x) = 1 \text{ for } x \in \Omega_1, \epsilon(x) = 1, \kappa(x) = 5 \text{ for } x \in \Omega_2. \quad (6-6)$$

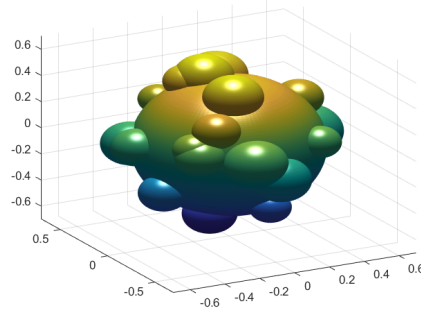


图 6.2 具有几何奇异性的区域

方程的右端项由精确解生成。

求解区域被一个球心位于原点，半径为 1 的大球截断。在球面上，精确解满足对应的边界条件。

示例 2： 方程中，右端项选取为

$$f(x) = \frac{e^{\sin \mu x_1 + \sin \mu x_2 + \sin \mu x_3}}{|x|^2 + 1} (|x|^2 - 1) \quad (6-7)$$

方程的系数选取为

$$\mu = 20, \epsilon(x) = 1 \text{ for } x \in \Omega_1, \epsilon(x) = 80 \text{ for } x \in \Omega_2, \kappa(x) = 1. \quad (6-8)$$

此时方程的精确解无法写出。在这种情况下，求解区域被立方体 $[-1, 1]^3$ 截断，参考解通过有限差分法（FDM）计算。在差分法中，我们选取了具有足够的精细的网格以确保精度。

在示例 1 中，在每个训练步，我们在区域 Ω 内采样 5000 个点，在边界 $\partial\Omega$ 上采样 4000 个点。在示例 2 中，我们对区域 Ω 内的 6000 个点和边界 $\partial\Omega$ 上的 3000 个点进行采样。需要注意的是，由于我们使用了 Ritz 方法，我们无需在界面 Γ 上采样，所有的跳跃边界条件都是自动满足的。

如图??所示。多尺度网络与普通的全连接网络相比，误差下降得更快。在两个示例中，多尺度网络都显示了很大的优势。

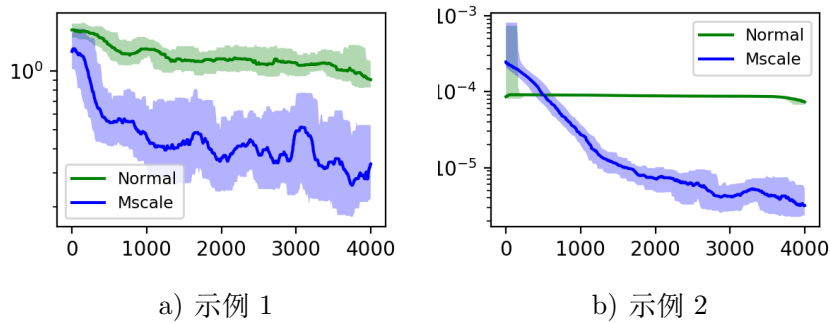


图 6.3 不同网络结构在求解 PB 方程中的表现

我们在图??展示了示例 2 中，不同方法得到的近似解在横截面 $x_1 = x_3 = 0$ 上的取值，包括 FDM ($h = 0.02$)、normal(训练 4000 步) 和 Mscale(训练 4000 步)。普通的全连通网络在区域的内部给出了错误的解，而多尺度网络的结果和参考解相近。这些实例表明，相比于普通的全连接网络，多尺度网络求解复杂区域上的 PB 方程的速度更快，精度更高。

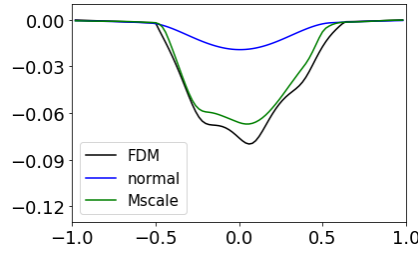


图 6.4 示例 2 中不同方法得到的近似解在截面 $x_1 = x_3 = 0$ 上的图像

6.4 具有源项奇异性的 PB 方程

在实际的生物分子中，电荷往往以点电荷的形式分布在分子内部，这就相当于方程中的右端项变为

$$f(x) = \sum_{k=1}^K q_k \delta(x - s_k) \quad (6-9)$$

其中 $\delta(x)$ 是 Dirac 函数， q_k 和 s_k 分别表示分子中点电荷的带电量 and 位置。我们假设核与分子界面之间的距离大于常量 R_0 。也就是说我们可以选取 $\Omega_0 = \{x : \exists k, |x - s_k| < R_0\} \subset \Omega_1$ 。在图??中，蓝色部分表示溶剂域 Ω_2 ，绿色部分表示生物分子区域 $\Omega_1 \setminus \Omega_0$ ，粉红色部分表示 Ω_0 。这里 Ω_0 是一些球的并集，其中已经包含了所有的点电荷。

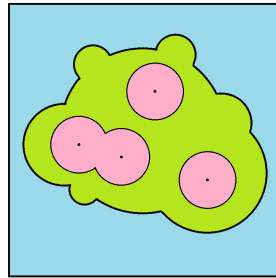


图 6.5 点电荷分布示意图

这里我们给出一种克服源项奇异性的方法。我们首先定义

$$\bar{u}(x) = \sum_{k=1}^K q_k G(x - s_k) m(x - s_k) \quad (6-10)$$

其中格林函数

$$G(x) = \frac{1}{4\pi\epsilon_1} \frac{e^{-\frac{\kappa_1}{\sqrt{\epsilon_1}}|x|}}{|x|} \quad (6-11)$$

截断函数

$$m(x) = \begin{cases} 1 - \left(\frac{|x|}{R_0}\right)^3 \left(4 - 3\frac{|x|}{R_0}\right) & |x| < R_0 \\ 0 & |x| > R_0 \end{cases} \quad (6-12)$$

通过以上定义，我们可以验证 $\bar{u}(x)$ 满足

$$\begin{aligned} -\Delta \bar{u}(x) + \kappa^2 \bar{u}(x) &= \sum_{k=1}^K q_k \delta(x - s_k) + \sum_{k=1}^K q_k F(|x - s_k|) \quad x \in \Omega_0 \\ \bar{u} = \frac{\partial \bar{u}}{\partial n} &= 0 \quad x \in \partial\Omega_0 \\ \bar{u}(x) &= 0 \quad x \in \Omega_0^c \end{aligned} \quad (6-13)$$

我们定义消除奇异性之后的解为

$$w(x) = u(x) - \bar{u}(x)\chi_{\Omega_0}(x) \quad (6-14)$$

奇异性之后的解满足这样一个没有奇异性的方程

$$-\epsilon(x)\Delta w(x) + \kappa^2(x)w(x) = f(x)\chi_{\Omega_0}(x) \quad (6-15)$$

其中

$$f(x) = -\sum_{k=1}^K q_k F(|x - s_k|) \quad (6-16)$$

和

$$F(r) = \begin{cases} \frac{3e^{-\frac{\kappa_1}{\sqrt{\epsilon_1}}r}}{\pi R_0^4} (2R_0 - 3r + 2\frac{\kappa_1}{\sqrt{\epsilon_1}}r^2 - 2R_0\frac{\kappa_1}{\sqrt{\epsilon_1}}r) & r < R_0 \\ 0 & r > R_0 \end{cases} \quad (6-17)$$

这里我们只要求解消去奇异性之后的方程，就可以得到原方程的解了。

下面我们选取示例，验证多尺度网络对消去奇异性之后方程的求解能力。

示例 1: 这里我们选取大区域 $\Omega = [-1, 1]^3$, Ω_1 是一个中心位于原点，半径为 $R = 0.7$ 的球。其它相关参数选择为

$$s = (0, 0, 0), \quad q = 1, \quad R_0 = 0.5, \quad \epsilon(x) = 1 \text{ for } x \in \Omega_1, \quad \epsilon(x) = 80 \text{ for } x \in \Omega_2, \quad \kappa(x) = 0.$$

由于此时的区域不再具有几何奇异性，我们可以算出方程的精确解为

$$u(x) = \begin{cases} \frac{1}{4\pi|x|\epsilon_1} - \left(\frac{1}{\epsilon_1} - \frac{1}{\epsilon_2}\right) \frac{1}{4\pi R} & x \in \Omega_1 \\ \frac{1}{4\pi|x|\epsilon_2} & x \in \Omega_2 \end{cases} \quad (6-18)$$

相应地，去掉奇异性的解为

$$w(x) = \begin{cases} \frac{1}{4\pi\epsilon_1|x|} \left(\frac{|x|}{R_0}\right)^3 \left(4 - 3\frac{|x|}{R_0}\right) - \left(\frac{1}{\epsilon_1} - \frac{1}{\epsilon_2}\right) \frac{1}{4\pi R} & |x| < R_0 \\ \frac{1}{4\pi|x|\epsilon_1} - \left(\frac{1}{\epsilon_1} - \frac{1}{\epsilon_2}\right) \frac{1}{4\pi R} & R_0 < |x| < R \\ \frac{1}{4\pi|x|\epsilon_2} & |x| > R \end{cases} \quad (6-19)$$

示例 2: 在第二个示例中，我们选择大区域 $\Omega = [-1, 1]^3$ 。区域 Ω_1 构造如下。我们选择一个中心为原点，半径 0.7 的大球。在大球表面随机抽取 20 个点作为小球的中心。小球的半径从 $[0.1, 0.3]$ 范围内的均匀分布中随机抽取。 Ω_1 是这些球的并集。带有奇异性的源项构造为。每个电荷的位置在一个球心位于原点，半径 0.5 的球中随机选择，电荷量从 $[-0.5, 0.5]$ 范围内的均匀分布中随机抽样。我们选择 $R_0 = 0.2$ 。如图??，图中蓝色区域表示 Ω_1 ，在蓝色区域内的红色部分表示 Ω_0 。

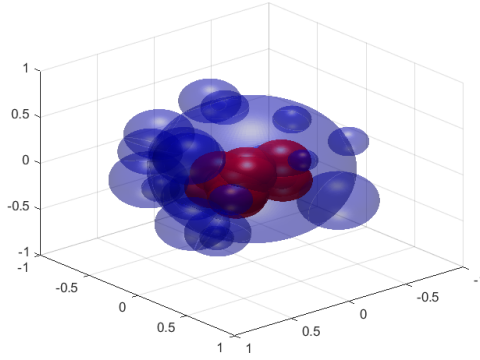


图 6.6 求解区域图像

其它参数选择如下

$$\epsilon(x) = 1 \text{ for } x \in \Omega_1 \quad \epsilon(x) = 80 \text{ for } x \in \Omega_2 \quad \kappa(x) = 0$$

需要注意的是，示例 2 中同时具有几何奇异性和源项的奇异性，这给普通的数值方法带来了很大的挑战。这里我们无法得到精确解，参考解同样由 FDM 给出。

每步中我们随机在区域 $[-1, 1]^d$ 内选取 5000 个均匀分布的积分点，在边界上生成 4000 个均匀分布的积分点。

在示例 1 中，方程具有源项奇异性，我们测试的网络结构如下：

1. 规模为 1-1000-1000-500-1 的全连接网络。(Normal)
2. 5 个规模为 1-200-200-100-1 的多尺度网络，系数为 $\{1, 2, 4, 8, 16\}$ 。(Mscale)

在示例 2 中，方程同时具有几何奇异性和源项奇异性，因此我们需要更多的神经元，使网络具有更强的逼近能力。这里测试的网络结构如下：

1. 规模为 1-1500-1000-1000-500-1 的全连接网络。(Normal)
2. 5 个规模为 1-300-200-200-100-1 的多尺度网络，系数为 $\{1, 2, 4, 8, 16\}$ 。(Mscale)

通过三种方法：FDM ($h = 0.01$)、normal (训练 10000 步) 和 Mscale (训练 10000

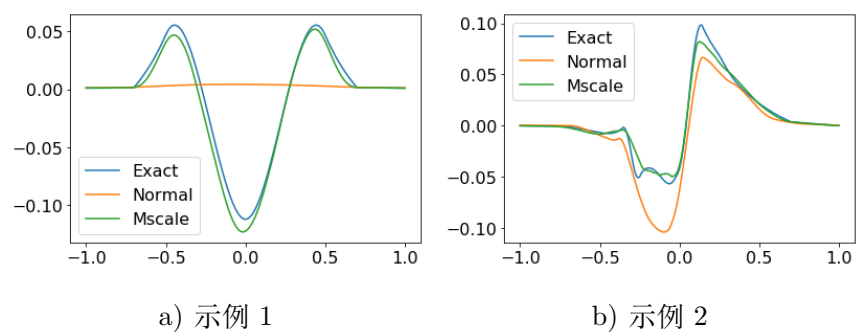


图 6.7 不同方法得到的近似解在截面 $x_2 = x_3 = 0$ 上的图像

步) 得到的截面上的数值解如图??所示。正常全连通网络的输出不能很好地捕捉精确解中的峰值，而多尺度网络得到了更加精确的结果。

第七章 深度学习方法在特征值问题中的应用

这里我们应用深度学习方法来求解椭圆特征值问题。由于特征值问题中，我们要求解的特征函数一般不具有高频震荡的性质，此外，特征值问题的结果随区间变换而改变。这里我们用普通的没有尺度变换的全连接网络求解特征值问题，目的主要是展示深度学习方法在求解特征值问题中的能力。

7.1 高维方形区域上的特征值问题

求解椭圆方程特征值问题^{??}，其中参数选取为 $a(x) = 1$ ， $c(x) = 0$ ，求解区域为 $\Omega = [0, 1]^d$ 。

问题的最小特征值为 $\lambda = d\pi^2$ ，对应的特征函数

$$u(x) = \prod_{i=1}^d \sqrt{2} \sin(\pi x_i) \quad (7-1)$$

我们用 Ritz 方法求解问题，测试规模为 1-1000-1000-1000-1000-1 的全连接网络。激活函数是 ReLU。每次训练区域内部生成 5000 个点，区域边界生成 4000 个点。选取边界惩罚系数为 1000，归一化惩罚系数为 10000。分别在 $d = 3$ ， $d = 5$ 和 $d = 10$ 的情况下求解特征值问题。实验结果如图^{??}，图^{??}，图^{??}。

在这些图中，图 (a) 画出了特征值的近似值和真实值随迭代步数变化的图像，图 (b) 画出了特征函数的近似解和真实解之间的误差。由于高维情况下的特征函数难以画图表示，我们选取两个截面作图。图中，“特征函数 1”画出的是在 $x_1 = x_2 = \dots = x_{d-1} = 0.5$ 截面上的数值解图像，“特征函数 2”画出的是在 $x_1 = x_2 = \dots = x_d$ 对角线上的数值解图像。

实验中展示了深度学习方法求解高维问题的能力，对于维数很高的问题，传统方法很难应用，但深度学习方法可以给出一些相对可以接受的近似解。但是值得注意的是，随着维度的增高，需要求解得到的目标函数也越来越复杂，这就需要更宽和更深的神经网络来逼近，这同样对计算能力提出了一定的挑战。

7.2 变系数的特征值问题

求解椭圆方程特征值问题^{??}，其中参数选取为 $a(x) = 1$ ，求解区域为 $\Omega = [0, 1]^d$ 。这里我们把 $c(x)$ 选为分片常数，测试深度学习方法对变系数特征值问题的效果。

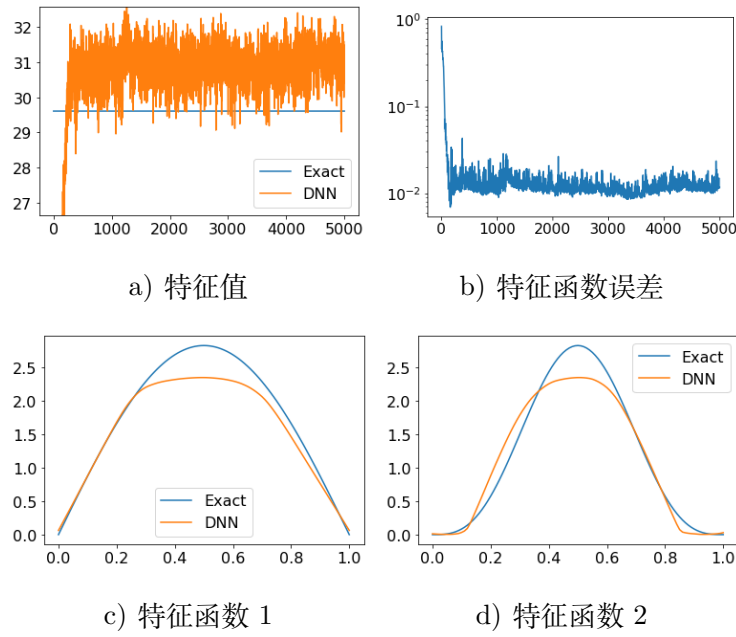


图 7.1 高维区域上的实验结果 ($d=3$)

7.2.1 一维数值实验

首先我们测试一维的情况，测试规模为 1-200-200-200-1 的全连接网络。每次训练区域内部生成 1000 个点，区域边界生成 200 个点。边界惩罚系数 β_2 为 1000，归一化惩罚系数 β_1 为 1000。其它参数相同。

计算结果如图??。图 (a) 中画出了势函数的选取，图 (b) 表示特征值的近似值和真实值随迭代步数变化的图像。图 (c) 中画出了精确解和深度学习方法给出的近似解的图像。可以看出，在一维的变系数问题上，虽然网络没有刻画出解精确的细节，但是峰值的位置已经可以确定了。

7.2.2 二维数值实验

在二维的情况下，我们测试规模为 1-800-800-800-1 的全连接网络。参数选取为每次训练区域内部生成 4000 个点，区域边界生成 1000 个点。边界惩罚系数 β_2 为 1000，归一化惩罚系数 β_1 为 10000。其它参数相同。

计算结果如图??。图 (a) 中画出了势函数的选取，图 (b) 表示特征值的近似值和真实值随迭代步数变化的图像。图 (c) 和 (d) 中画出了精确解和深度学习方法给出的近似解的图像。可以看出，二维情况下误差比一维的更大了，但还是在可以接受的范围内。

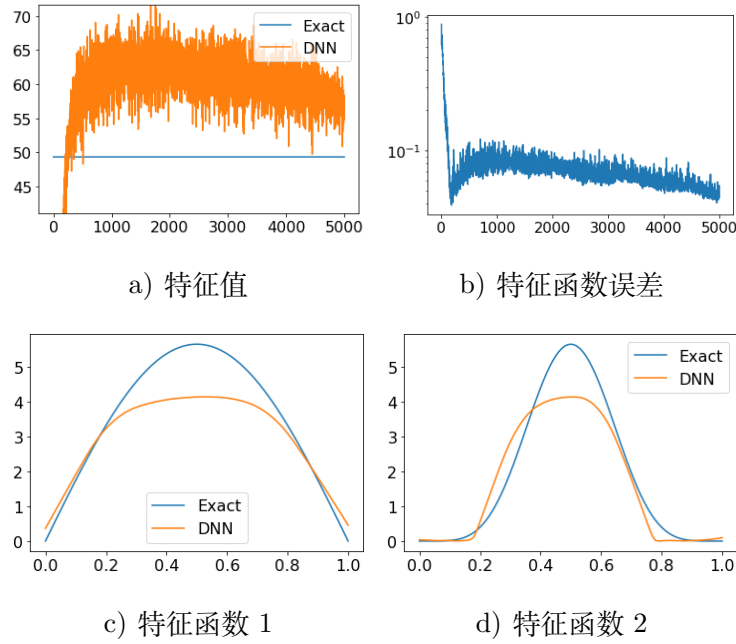


图 7.2 高维区域上的实验结果 ($d=5$)

7.3 具有几何奇异性的特征值问题

在椭圆方程特征值问题??中，参数选取为 $a(x) = 1$ ， $c(x) = 0$ ，我们选取不同的求解区域，验证几何奇异性对深度学习方法求解特征值问题的影响。

7.3.1 复杂单连通区域

这里的求解区域为一个多边形区域，多边形的顶点依次为 $(-1, 1) \rightarrow (-1, -1) \rightarrow (1, -1) \rightarrow (1, -0.6) \rightarrow (-0.3, -0.6) \rightarrow (-0.3, 0) \rightarrow (0, 0) \rightarrow (0, -0.5) \rightarrow (1, -0.5) \rightarrow (1, 1) \rightarrow (0.5, 1) \rightarrow (0.5, 0.3) \rightarrow (-1, 1)$ 。

如图??为通过 matlab 的 PDE 工具箱生成的网格和求解得到的特征函数参考解。问题的最小特征值参考解为 $\lambda = 19.2725$ 。

在深度学习方法中，我们测试规模为 1-800-800-800-1 的全连接网络。每次训练区域内部生成 8000 个点，区域边界生成 2500 个点。边界惩罚系数 β_2 为 1000，归一化惩罚系数 β_1 为 10000。计算结果如图??。图 (a) 画出了特征值的近似值和真实值随迭代步数变化的图像，图 (b) 画出了特征函数的近似解和真实解之间的误差。图 (c) 画出了迭代 5000 步后近似解的图像。可以看出，通过深度学习得到的特征函数近似解和参考解相差不大。在这样的区域上，深度学习方法效果很好。

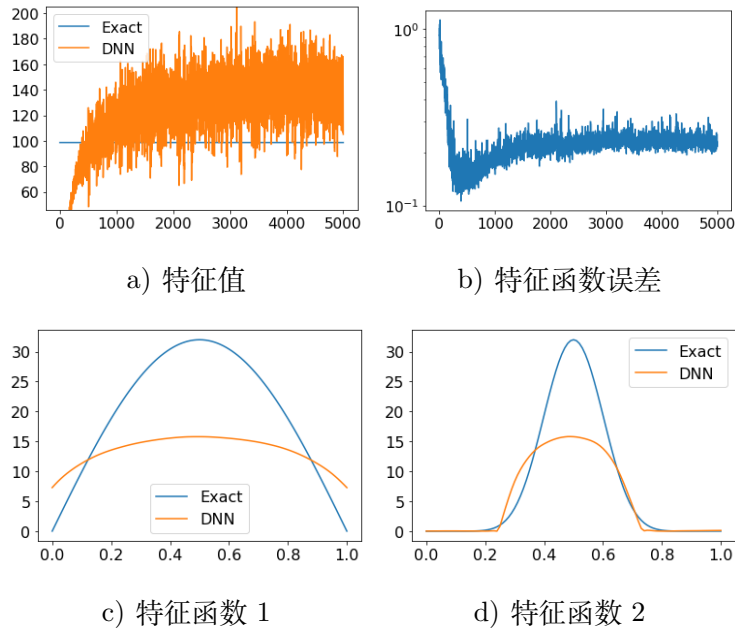


图 7.3 高维区域上的实验结果 (d=10)

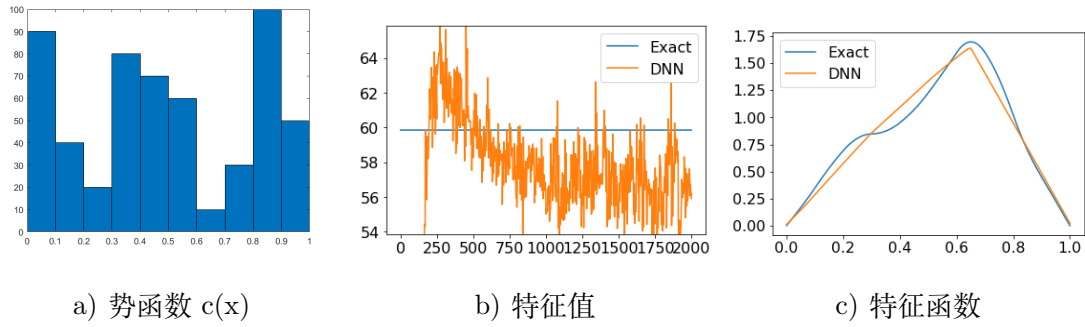


图 7.4 一维变系数方程中的实验结果

7.3.2 有孔区域

下面我们求解有孔区域上的特征值问题。求解区域为正方形 $[-1, 1]^2$ 挖去三个圆形的洞。洞的圆心和半径依次为 $c_1 = (-0.3, 0.3), r_1 = 0.5, c_2 = (0.6, 0), r_2 = 0.3, c_3 = (-0.1, -0.7), r_3 = 0.2$ 。

如图??为通过 matlab 的 PDE 工具箱生成的网格和求解得到的特征函数参考组。问题的最小特征值为 $\lambda = 28.8807$ 。

在同样的参数下用深度学习方法求解，实验结果如图??。可以看出，深度学习方法与 PDE 工具箱中给出的参考解有一定差距，但是特征函数的大致形状和峰的位置都已经被刻画出来了。

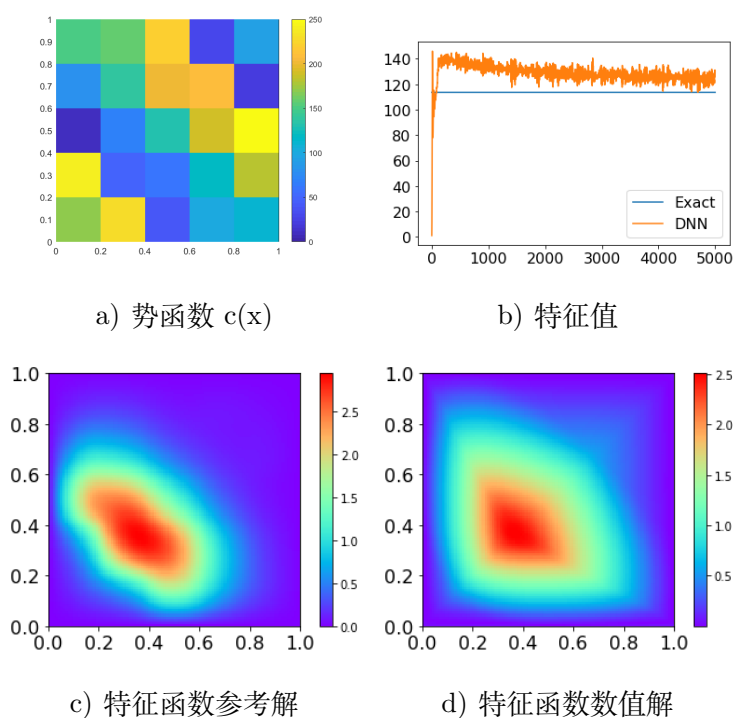


图 7.5 二维变系数方程中的实验结果

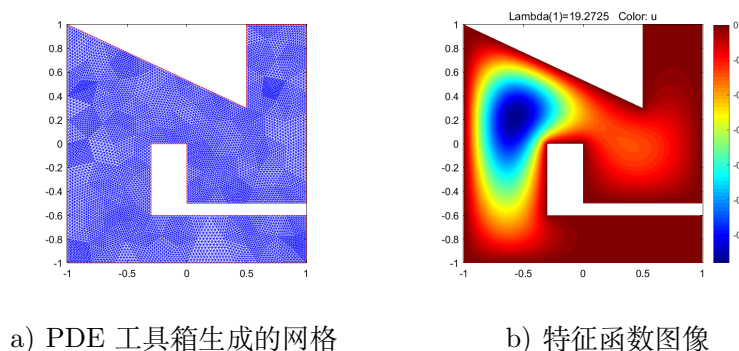


图 7.6 参考解图像

7.3.3 多孔区域

我们选取求解区域为正方形 $[-1, 1]^2$ 挖去 25 个圆形的洞。洞的圆心位于区域内的均匀网格上，半径依次在 0.05, 0.15 之间均匀划分。

如图??为通过 matlab 的 PDE 工具箱求解得到的特征函数和网格。最小特征值为 $\lambda = 68.9198$ 。

在同样的参数下用深度学习方法求解，实验结果如图??。对于这样复杂的区域，深度学习方法的效果明显变差，几乎完全没有得到正确的结果。这说明深度学习方法在处理一些复杂的特征值问题时，还有很大的改进空间。

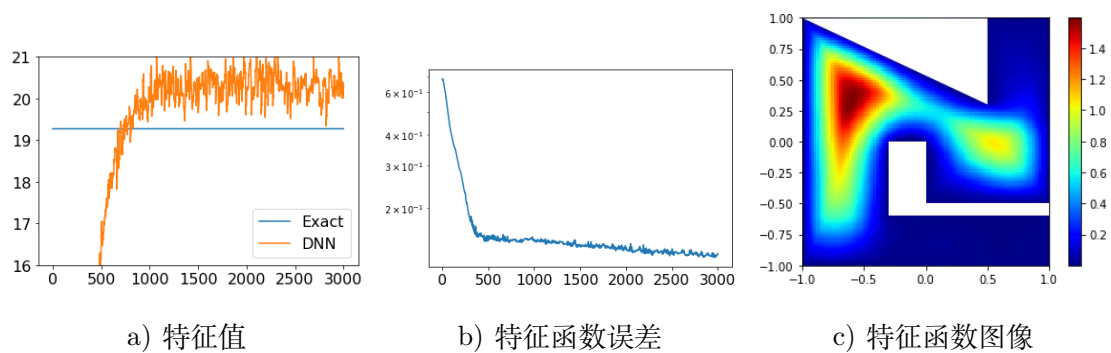


图 7.7 深度学习结果

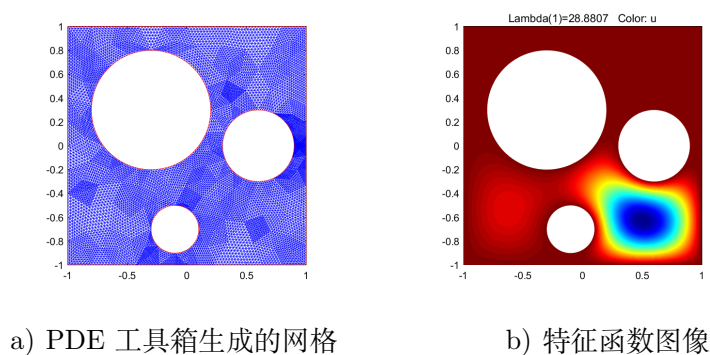


图 7.8 参考解图像

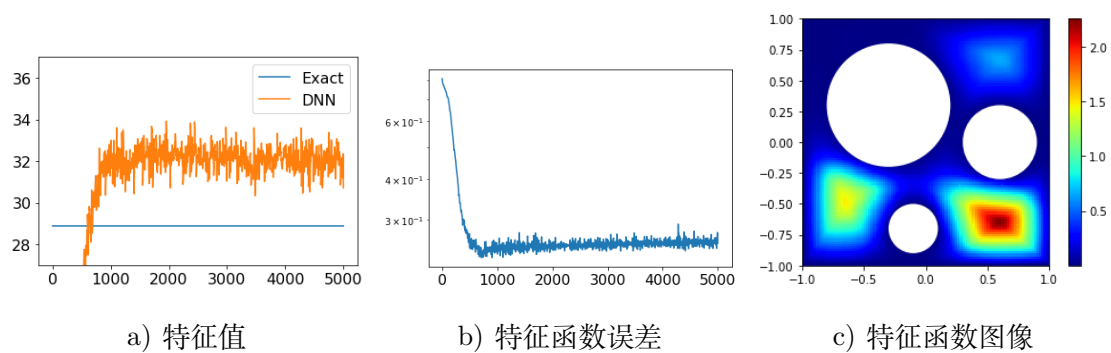


图 7.9 深度学习结果

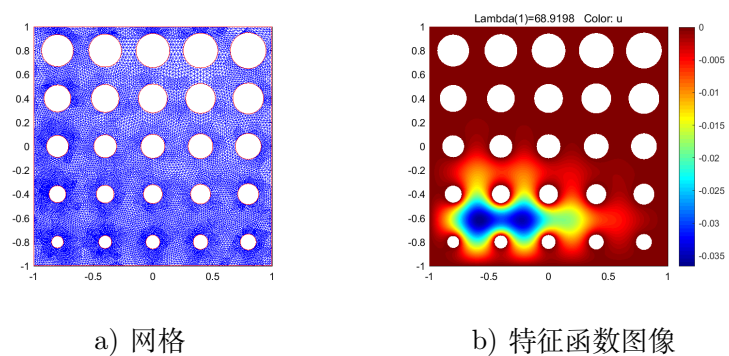
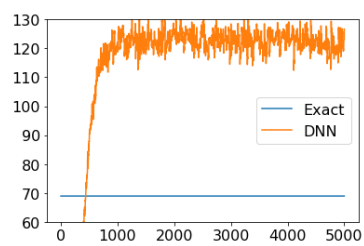
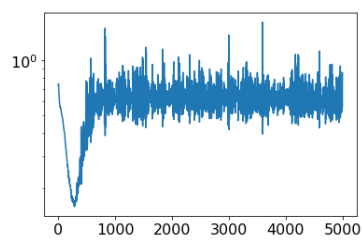


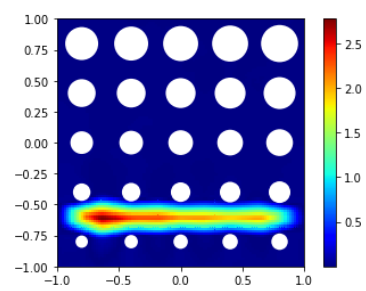
图 7.10 参考解图像



a) 特征值



b) 特征函数误差



c) 特征函数图像

图 7.11 深度学习结果

第八章 总结与展望

本文基于深度学习方法，从初始化的角度出发，介绍了一种新的多尺度网络结构。我们利用尺度变换的技术和紧支集的激活函数，构造了一种可以求解高频震荡椭圆方程的多尺度网络。本文的主要创新在于：在通常的全连接网络下，求解高频震荡的椭圆方程比较困难，而多尺度网络克服了这一缺点。大量的实验证明，多尺度网络是一种无网格、易于实现的椭圆方程求解方法。

相比于传统的有限元和有限差分方法，深度学习方法具有如下优势：首先，在复杂的区域上，由于需要生成网格和求解大型线性方程组，传统的有限元和有限差分法求解方法可能代价高昂。而深度学习方法只需要在边界和内部生成随机点，无需生成网格。此外，深度学习方法是本质非线性的，虽然本文中只讨论了线性方程的情形，这些方法都可以毫无难度地推广到非线性方程的情况。最后，深度学习方法采用了类似蒙特卡洛的抽样，因此可以克服维度灾难，更加快速地求解高维问题。

本文的末尾通过深度学习方法求解了椭圆特征值问题，指出深度学习方法在求解特征值问题中还有很大的改进空间。未来我们可以针对特征值问题设计对应的神经网络，希望可以提高求解效率。在进一步的工作中，我们还将探索在多尺度网络中，充分利用小波激活函数的思想，进一步提高多尺度网络的频率局部化和分离能力。此外，我们将尝试将多尺度网络应用于大规模计算工程问题，特别是与有限元法和有限差分法的比较。更进一步，我们将研究将多尺度网络应用于高维，更复杂的方程，如多体量子系统的薛定谔方程。在这些问题中，方程往往是非常高维的，用深度学习方法求解会为我们带来很大的便利。

致 谢

三年前，不知什么原因让我执着地来到了这里，可能是科研兴趣，可能是当时家庭变故，也可能只是想为慌乱的心灵找一个可以安静地方。这么说，并不是过分夸张，我本科的朋友不理解我为什么来到此处做这个方向；研究生的同学和老师对我的评价是，你在这条路好像可以走下去，但并不是很适合；找工作面试我的 HR 也问我，你为什么去那个地方读研。好像每次解释这件事的时候，我说的都不是很一致三年即将结束，三年时间很漫长，甚至有些不愿承认这只是三年，他不是很精彩，却依然丰富。

不敢谈英雄主义，三年让我认清了些生活的真相，也不减对生活的热爱。

附录 9 攻读硕士学位期间发表的论文及其他成果

(一) 发表的学术论文

- [1] Liu Z, Cai W, Xu Z. Multi-scale Deep Neural Network (MscaleDNN) for Solving Poisson-Boltzmann Equation in Complex Domains[J]. 2020.

(二) 参加学术活动情况

- [1] 2019.06 深度学习在计算机视觉、求解微分方程和计算建模中的应用
- [2] 2019.07 第三届“偏微分方程数值方法与理论暑期学校”
- [3] 2020.07 2020 年度北京大学“应用数学专题讲习班”
- [4] 2020.07 机器学习和大数据在复杂性科学中的应用暑期培训班
- [5] 2020.12 【百旺科学论坛】认识我们的宇宙