

東南大學  
博士学位论文

基于传递熵的非线性时间序列因果推断及其在  
路面时序建模的应用<sup>1</sup>

专业名称: 数学

研究生姓名: 张金壬

导师姓名: 曹进德 教授



**NONLINEAR TIME SERIES  
CASUAL DETECTION BASED  
ON TRANSFER ENTROPY AND  
ITS APPLICATION ON  
PAVEMENT TIME SERIES  
MODELING**

A Dissertation Submitted to  
Southeast University  
For the Academic Degree of Doctor of Science

By  
ZHANG Jinren

Supervised by  
Professor CAO Jinde

School of Mathematics  
Southeast University  
October 2023



## 东南大学 学位论文独创性声明

本人声明所呈交的学位论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，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，也不包含为获得东南大学或其它教育机构的学位或证书而使用过的材料。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了谢意。

研究生签名: \_\_\_\_\_ 日期: \_\_\_\_\_

## 东南大学 学位论文使用授权声明

研究生签名: \_\_\_\_\_ 导师签名: \_\_\_\_\_ 日期: \_\_\_\_\_



# 基于脉冲控制的时滞复杂网络的动力学分析

## 摘要

复杂网络是具有大量节点和多种连接拓扑的网络模型，广泛存在于现实生活中，例如社交网络、万维网、电网以及交通网络等。在信息资源和能量有限的情况下，复杂网络内部节点在执行信息交互以及反馈控制过程中，不可避免地会产生时滞。通常，时滞的存在会导致复杂网络产生振荡、发散和不稳定等不良性能，因此，时滞复杂网络一直以来都是一个重要的研究模型。在复杂网络的动态发展过程中，系统的高度复杂性与时滞的相互影响可以产生丰富多样的动力学行为。因此，时滞复杂网络的动力学行为（如同步、一致性等）分析具有重要的理论价值和实际意义。本文主要讨论在不同类型脉冲控制下时滞复杂网络动力学特性，主要创新工作如下：

1、分别研究了在混杂脉冲控制下耦合多稳态时滞忆阻神经网络的动态多同步和静态多同步问题。其中，混杂脉冲控制包括时滞脉冲控制和连续时间状态反馈控制。首先，基于状态空间划分方法和激励函数的几何特性，给出了系统具有多个局部指数稳定的周期轨道或平衡点的充分条件。其次，利用新型的 Halanay 微分不等式和脉冲控制理论，分别建立了基于线性矩阵不等式 (LMIs) 的动态多同步和静态多同步的充分性判据。结果表明，在具有可允许的时变时滞上界和合适的脉冲间隔的混杂脉冲控制下，动态多同步和静态多同步都可以被实现。

2、研究了事件触发脉冲控制下具有耦合时滞的复杂网络同步问题。首先，设计了一种新型的基于二次 Lyapunov 函数的事件触发时滞脉冲控制机制，此机制将时滞脉冲控制考虑到事件触发机制中。通过构造合适的辅助函数，并利用递推法、脉冲控制理论和 Lyapunov–Razumikhin 技术，研究了在事件触发时滞脉冲控制下具有耦合常时滞的复杂网络全局指数同步问题，并建立了基于 LMIs 的保守性更小的同步判据。其次，进一步将牵制控制考虑到事件触发脉冲控制机制中，设计出事件触发牵制脉冲控制，探讨了具有耦合比例时滞的基因振荡器网络集群同步问题，并建立起相应的集群同步准则。值得注意的是触发时刻即为脉冲时刻，其由所设计的事件触发牵制脉冲控制机制产生。同时，每个集群中只有一小部分节点依据所设计的算法在触发时刻被施加脉冲控制，进一步节约了网络资源。

3、研究了分布式饱和脉冲控制下时滞复杂网络的局部动力学行为。首先，利用反证法、脉冲系统的比较原理和平均脉冲区间的方法研究了分布式饱和脉冲控制下具有耦合时滞的鲁里叶网络的局部指数同步问题，并给出了不依赖于时滞的基于双线性矩阵不等式 (BMIs) 的充分性判据。为了降低保守性，选取最新的具有更多松弛变量的改进凸包表示法来处理分布式饱和脉冲项，并且开发了一种新的估计吸引域方法，该方法与传统的借助于收缩不变集来估计吸引域的方法完全不同。其次，通过构造依赖于脉冲时刻的复合型 Lyapunov 函数进一步降低保守性，讨论了分布式饱和脉冲控制下具有切换拓扑的非线性时滞多智能体系统的局部一致性，并建立起一致性准则。为了估计出最大的吸引域，通过适当的矩阵变换，建立起基于 LMIs 的优化问题，并通过 Matlab 软件中的 Yalmip 工具箱求解相应的最大吸引域的数值解。

**关键词：** 时滞复杂网络；同步；一致性；混杂脉冲控制；事件触发脉冲控制；分布式饱和脉冲控制。

# Dynamic Analysis of Time-delay Complex Network Based on Impulse Control

## Abstract

Complex networks is a network model with a large number of nodes and a variety of connection topologies, which widely exists in real life, such as social networks, World Wide Web, power grids, transportation networks, and so on. In the case of limited information resources and energy, time delay is inevitable in the process of information interaction and feedback control of nodes in complex networks. Usually, the existence of time delay can cause poor performance of complex networks, such as oscillation, divergence and instability. Therefore, the time-delay complex networks has always been an important research model. In the dynamic development of complex networks, the interaction between the high complexity of the system and time delay can produce various dynamic behaviors. Therefore, it has great theoretical value and practical significance to analyze the dynamic behaviors of time-delay complex networks, such as synchronization, consensus and so on. This thesis mainly discusses the dynamic characteristics of time-delay complex networks under different types of impulsive control, the main innovations are as follows:

1. The problems of dynamic multisynchronization (DMS) and static multisynchronization (SMS) of coupled multistable time-delay memristive neural networks under hybrid impulsive control are investigated separately. Among them, the hybrid impulsive control includes delayed impulsive control and continuous-time state feedback control. Firstly, based on the state space partition method and the geometrical properties of the activation function, several sufficient conditions are given such that system has multiple local exponential stable periodic orbits or equilibrium points. Secondly, by employing a new Halanay differential inequality and the impulsive control theory, some DMS and SMS sufficient conditions based on linear matrix inequalities (LMIs) are established, respectively. It is shown that DMS and SMS can be realized under hybrid impulsive control with allowable upper bound of time-varying delay and suitable impulsive interval.

2. The synchronization problem of complex networks with coupling time delay under event-triggered impulsive control (ETIC) is studied. Firstly, a novel event-triggered delayed impulsive control (ETDIC) mechanism combining delayed impulsive control with the event-triggering mechanism is designed based on the quadratic Lyapunov function. By constructing a suitable auxiliary function and using recurrence method, impulsive control theory and Lyapunov–Razumikhin technique, the global exponential synchronization problem of complex networks with coupling constant delay under ETDIC is studied, and some synchronization criteria with less conservatism are established in terms of LMIs. Secondly, the pinning control is further considered in the ETIC mechanism, and the event-triggered pinning impulsive control (ETPIC) is designed. Based on this, the cluster synchronization of genetic oscillator networks with coupling proportional delay is discussed and the corresponding cluster synchronization criteria are established. Note that the triggering instant is the impulsive instant, which is generated by the designed ETPIC mechanism. Meanwhile, only a fraction of nodes of each cluster are controlled at each triggering instant according to the designed algorithm, which further saves network resources.

3. The dynamic behavior of time-delay complex networks under distributed saturated impulsive control (DSIC) is researched. Firstly, by utilizing proof by contradiction, comparison principle of impulsive system, and average impulsive interval method, the local exponential synchronization problem of Lur'e networks with coupling time-varying delay under the DSIC is studied, and some delay-independent sufficient criteria are presented in the form of bilinear matrix inequalities (BMIs). To reduce conservatism, a new improved convex hull representation with more slack variables be chosen to deal with the DSIC terms, and a new method is developed to estimate the domain of attraction, which is quite distinct from the traditional method of estimating the domain of attraction by means of contractive invariant set. Secondly, a novel composite impulsive-instant-dependent Lyapunov function is constructed to further reduce conservatism. And then, the local consensus problem of nonlinear time-delay multi-agent systems with switching topologies via DSIC is discussed and the some local consensus criteria are derived. In order to estimate the maximum domain of attraction, some LMIs-based optimization problems are formulated through appropriate matrix transformation, and the corresponding

numerical solutions of the maximum domain of attraction are solved through the Yalmip toolbox in Matlab software.

**Keywords:** Time-delay complex networks; synchronization; consensus; hybrid impulsive control; event-triggered impulsive control; distributed saturated impulsive control.



# 符 号 说 明

$\mathbb{N}$	自然数域;
$\mathbb{Z}_+$	正整数域;
$\mathbb{R}$	实数域;
$\mathbb{R}_+$	正实数域;
$\mathbb{R}^n$	$n$ 维的 Euclidean 空间;
$\mathbb{R}^{m \times n}$	$\mathbb{R}$ 上的 $m \times n$ 维矩阵;
$\mathbf{c}_n$	所有元素均为数字 $c$ 的 $n$ 维列向量;
$ x _\infty$	向量 $x \in \mathbb{R}^n$ 的 $\infty$ 范数, $ x  = \max_{1 \leq i \leq n}  x_i $ ;
$ x $	向量 $x \in \mathbb{R}^n$ 的 Euclidean 范数, $ x  = \sqrt{\sum_{i=1}^n x_i^2}$ ;
$P > 0$ ( $P \geq 0$ )	矩阵 $P$ 是正定 (半正定) 矩阵;
$P < 0$ ( $P \leq 0$ )	矩阵 $P$ 是负定 (半负定) 矩阵;
$\lambda_{\max}(P)$ ( $\lambda_{\min}(P)$ )	矩阵 $P$ 的最大 (小) 特征值;
$a \vee b$ ( $a \wedge b$ )	$a$ 和 $b$ 的最大 (小) 值;
$P^T$	矩阵 $P$ 的转置;
$P^{-1}$	矩阵 $P$ 的逆矩阵;
$\mathbb{O}_{m \times n}$	$m \times n$ 维零矩阵;
$I_n$	$n$ 阶单位矩阵;
$\text{diag}(p_1, p_2, \dots, p_n)$	对角线元素为 $p_1, p_2, \dots, p_n$ , 非对角线元素为零的 $n \times n$ 矩阵;
$\mathcal{C}(D, F)$	定义域为 $D$ , 值域为 $F$ 的连续函数全体;
$\mathcal{C}^1(D, F)$	定义域为 $D$ , 值域为 $F$ 的连续可微函数全体;
$\mathcal{PC}(D, F)$	定义域为 $D$ , 值域为 $F$ , 除在有限个点 $t$ 之外, 在其它任何时刻都是连续的函数 $\psi$ 全体, 其中 $\psi(t^+)$ 和 $\psi(t^-)$ 都存在, 并且 $\psi(t^+) = \psi(t)$ ;
$v_0$	表示一个函数类, 其中函数 $V(t, x) \in \mathcal{PC}(-\tau, +\infty)$ 是正定的, 且关于 $x$ 是局部 Lipschitz 的, 其中 $\tau$ 为正常数;

$I[m, n]$	整数集 $\{m, m + 1, \dots, n\}$ ;
$:=$	定义符号;
$\lfloor \cdot \rfloor$	下取整函数;
$co(\cdot)$	一组向量组成的凸包;
$\mathcal{E}^c$	集合 $\mathcal{E}$ 的补集;
$D^+$	右上 <i>Dini</i> 导数;
$\otimes$	Kronecker 积;
$\star$	对称矩阵中的对称块。

# 目录

<b>第一章 绪论</b>	<b>1</b>
1.1 时滞复杂网络概述 . . . . .	1
1.2 时滞复杂网络动力学的研究背景及现状 . . . . .	2
1.2.1 同步的研究背景及现状 . . . . .	2
1.2.2 一致性的研究背景及现状 . . . . .	4
1.3 脉冲控制的研究背景及现状 . . . . .	6
1.3.1 脉冲控制的研究背景 . . . . .	6
1.3.2 脉冲控制的研究现状 . . . . .	7
1.4 预备知识 . . . . .	10
1.4.1 代数图理论 . . . . .	10
1.4.2 矩阵理论 . . . . .	12
1.5 本文的主要内容和创新点 . . . . .	13
1.5.1 主要研究内容 . . . . .	13
1.5.2 主要创新点 . . . . .	16
<b>第二章 两种基于rényi熵的分数阶累计剩余时间序列度量</b>	<b>21</b>
2.1 引言 . . . . .	21
2.2 Rényi 累积剩余KL散度及其性质 . . . . .	24
2.3 Rényi累积剩余传递熵及其性质 . . . . .	28
2.4 仿真数据实例 . . . . .	33
2.4.1 VAR 模型 . . . . .	33
2.4.2 Hénon映射 . . . . .	33
2.4.3 关于 RCRKL 的仿真实验结果 . . . . .	34
2.4.4 关于 RCRTE 的仿真实验结果 . . . . .	35
2.5 实例应用 . . . . .	37
2.5.1 基于 RCRKL 的车辙时间序列聚类结果及讨论 . . . . .	37
2.5.2 基于 RCRTE 的车辙时间序列因果检测结果及讨论 . . . . .	40
2.6 小结 . . . . .	42

<b>第三章 基于自适应小波分解和多元核密度估计的多尺度因果推断</b>	<b>43</b>
3.1 引言 . . . . .	43
3.2 模型描述 . . . . .	45
3.2.1 预备知识 . . . . .	45
3.2.2 基于自编码器的自适应离散小波分解 . . . . .	48
3.2.3 基于核密度估计的rényi 传递熵 . . . . .	49
3.3 实验结果及讨论 . . . . .	55
3.3.1 数据简介 . . . . .	55
3.3.2 基准数据集CauseEffectPairs . . . . .	56
3.3.3 RIOHTrack数据集 . . . . .	57
3.3.4 结果及分析 . . . . .	57
3.4 小结 . . . . .	68
<b>第四章 基于多级小波分解和传递熵的车辙预测框架</b>	<b>71</b>
4.1 引言 . . . . .	71
4.2 相关工作 . . . . .	72
4.2.1 车辙预测模型 . . . . .	72
4.2.2 路面材料和结构的非稳态时间序列预测 . . . . .	73
4.3 模型描述 . . . . .	74
4.3.1 时序预测模型框架 . . . . .	76
4.3.2 高斯随机过程 . . . . .	77
4.3.3 基于融合注意力机制的自编码器 . . . . .	78
4.4 实例结果及讨论 . . . . .	81
4.4.1 小波基函数的选择 . . . . .	81
4.4.2 基于传递熵的因果关系结果验证 . . . . .	82
4.4.3 预测模型的性能分析 . . . . .	83
4.5 小结 . . . . .	88
<b>第五章 总结与展望</b>	<b>89</b>
5.1 总结 . . . . .	89
5.2 展望 . . . . .	90

附录 A 博士期间撰写和发表的论文 113

附录 B 博士期间主持和参加的科研项目、学术会议和获得的荣誉 115

# 第一章 绪论

## 1.1 时滞复杂网络概述

从生态系统到社会系统，从人际关系网到万维网，广泛存在于自然界和人类社会中的复杂系统都可以由相应的复杂网络模型刻画。复杂网络的数学理论来源于经典的图论 [1]，其发展经历了一个漫长的探索过程。最早的研究始于 1736 年欧拉提出的柯尼斯堡七桥问题，但在之后相当长的一段时间都没有实质性进展。直到 20 世纪 60 年代，随机图理论 [2] 的建立为网络构造提供了一种新的方法，正式开启了复杂网络的理论研究。随后，在 20 世纪 90 年代，小世界网络 [3] 和无标度网络 [4] 的提出打破了随机图理论的框架，开启了复杂网络研究的新纪元。在过去的几十年里，复杂网络已经成为众多学科争相关注的焦点，例如生物科学、计算机科学、物理科学、图论以及社会学等 [5–7]。在实际的复杂网络中，由于信息传输和采样速度有限、网络带宽受限以及物理元件老化等原因，时滞往往是不可避免的，这导致系统当前的发展趋势不仅依赖于当前的状态，还依赖于过去的状态 [8–10]。时滞的存在通常会导致复杂网络产生振荡、发散等不良性能。因此，时滞复杂网络的研究是非常有必要和有意义的。

在时滞复杂网络中，时滞可以分为耦合时滞（即节点之间通讯产生的时滞）[11–17] 和状态时滞（即节点自身信息反馈产生的时滞）[18,19]；也可以分为常时滞和时变时滞。为了研究不同类型的时滞，不同的研究方法相继被提出。最早的方法为频域法（直接法）[20]，通过分析系统的特征根在复平面上的分布来研究系统的动态性能。虽然这种方法更加直观形象且易于理解，但是对于具有大规模节点的复杂网络来说，其特征方程为超越方程，计算出所有的特征根是不切实际的。随后，时域法（间接法）[21, 22] 被提出，它被广泛用于处理时变时滞系统、参数不确定的时滞系统和时滞非线性系统，是目前最主要的研究方法。时域法主要有 Lyapunov-Krasovskii 泛函法（充分利用时滞信息）和 Lyapunov-Razumikhin 函数法（不借助于时滞信息）。通常，Krasovskii 法被用于处理具有较小或变化缓慢的时滞的系统，提供依赖于时滞的条件；而 Razumikhin 法则被用于处理具有较大或变化比较快的时滞的系统，提供不依赖于时滞的条件。

网络拓扑结构是时滞复杂网络的重要组成部分，可以清楚地反映网络节点之间的连接关系，即节点之间的信息交流 [23]。起初，时滞复杂网络的网络拓扑主要设定为节点间的拓扑关系不随时间发生改变且网络中的所有边都是无方向的，此时网络具有

无向的静态(固定)拓扑结构[24]。而现实世界中大多数网络,例如电网、食物链网络和引文网络等都是有向的[25–27]。此外,随着空间和时间的推移,网络的节点数可能会不断地增加或减少,并且节点之间的权重和连接方式都可能是动态变化的,这导致网络拓扑是不断发生变化的。例如无线传感器网络中节点可能出现移动、故障等情况,万维网的网页随时可能出现或断开。因此,具有有向动态网络的时滞复杂网络吸引了大量研究者的关注,其中切换拓扑作为动态网络结构中比较重要的类型已有大量相关结果被提出并发表[28–31]。

## 1.2 时滞复杂网络动力学的研究背景及现状

由于时滞复杂网络中不同的网络拓扑、每个子系统的非线性以及时滞的存在,使时滞复杂网络在演化过程中能够产生丰富的动力学行为。

### 1.2.1 同步的研究背景及现状

同步作为一种重要的集群行为,广泛存在于自然界、工程和社会生活中,例如钟摆的同步摆动,成千上万的萤火虫以共同的频率发出光脉冲,人体大脑神经网络和心脏肌肉细胞的共同振动,管弦乐队中小提琴手们齐声演奏等[32–34]。在过去的十几年里,时滞复杂网络的同步问题吸引了众多国内外学者的关注且收获颇丰。其中,完全同步是形式最简单也是研究最多的同步类型,是指在不同的初始条件下两个或多个动力系统,通过动力系统间的相互作用,使得各个动力系统状态逐渐趋同,最终达到完全相同,其数学定义如下:

**定义 1.2.1** [35] 记时滞复杂网络中节点  $i$  的状态变量为  $x_i(t)$ ,  $i \in I[1, N]$ , 如果

$$\lim_{t \rightarrow \infty} |x_j(t) - x_i(t)| = 0$$

对所有  $i, j \in I[1, N]$  成立,那么,称时滞复杂网络可以实现完全同步。

称  $\mathcal{S} = \{x = (x_1^T, x_2^T, \dots, x_N^T)^T : x_1(t) = x_2(t) = \dots = x_N(t) = s(t)\}$  为时滞复杂网络状态控制中的同步流形,其中  $s(t)$  是时滞复杂网络中孤立节点的解。那么,上述定义可以重新描述为:

**定义 1.2.2** 记时滞复杂网络中节点  $i$  的状态变量为  $x_i(t)$ ,  $i \in I[1, N]$ , 如果

$$\lim_{t \rightarrow \infty} |x_i(t) - s(t)| = 0$$

对所有  $i \in I[1, N]$  成立。那么，称时滞复杂网络可以实现完全同步。

近十几年来，众多学者从不同的方面、不同角度对时滞复杂网络的完全同步进行了研究 [36–47]。从时滞的角度来看，借助 Lyapunov 方法和随机分析技术，song 等人研究了具有状态和耦合常时滞的复杂网络的完全同步问题 [37]，得到了依赖于时滞的同步判据。在此基础上，Shi 等人通过设计合适的控制协议进一步讨论了具有时变时滞的复杂网络完全同步问题 [38]。最近，Tang 等人将有界时滞的结果推广到无穷时滞，研究了自适应策略下具有无穷时滞的复杂网络的完全同步问题 [42]。从网络拓扑的角度来看，前期的研究结果主要集中于具有固定拓扑的时滞复杂网络。然而，在网络的实际演化过程中，由于节点的增多或减少以及网络问题等原因，网络拓扑结构总是不断变化的，并且时变的网络拓扑可以传输更多的网络信息。因此，时变的网络拓扑越来越受到人们的关注，例如，具有时变网络拓扑的时滞复杂网络的完全同步问题在文献 [40, 46] 中进行了研究，为网络拓扑时变时的完全同步提供了新的见解。此外，作为时变网络拓扑的最为常见的类型，具有切换拓扑的时滞复杂网络的完全同步也被充分讨论 [41, 47]。研究表明，网络拓扑结构对时滞复杂网络同步性能有着显著的影响。从不同类型的耦合项来看，大部分的完全同步结果都集中于具有线性耦合项的时滞复杂网络。而在现实的网络中，节点的邻居信息在绝大多数情况下无法实时获得，那么存在邻居节点未知但是满足非线性函数的情况，因此具有非线性耦合项的时滞复杂网络得到了关注，相应的完全同步准则也被提出 [43–45]。

时滞复杂网络的完全同步实际上是比较理想的情况，而现实的时滞复杂网络的动力学可能是非常复杂的，完全同步可能无法实现，尤其是节点存在不同的动力学特征。因此，其他形式的有序行为，例如相位同步、集群同步、投影同步等广义同步被相继提出，并且越来越受到关注 [48–54]。例如，大规模时滞复杂网络的相位同步在文献 [48, 53] 中被讨论，从能量的角度观察和分析网络的相位同步而振幅可以相关很差或没有关联。大多数情况下，投影同步都是将系统的状态同步到给定的常数比例因子，而文献 [49, 54] 改进了这一结果，研究了时滞复杂网络的函数投影同步，其中比例因子是函数比例因子。随后，为揭示耦合基因振荡器网络不同集群规律行为的产生机制，Guan 等人分析了时滞耦合基因调控网络的集群同步 [50]，描绘了在同一集群中的节点可以实现完全同步，在不同集群中的节点无法实现同步的动力学行为。最近，通过设计基于离散时间状态观测的非周期间歇控制协议，Chen 等人研究了模糊异构的复杂网络的拟同步，描绘了具有不同动力学的节点之间存在有界误差的动力学行为 [52]。为了描述更多丰富的有序行为，时滞复杂网络的广义同步还值得进一步探索与研究。

### 1.2.2 一致性的研究背景及现状

一致性是复杂网络除同步外又一重要而特殊的动力学行为，主要用于描述为满足工程需求而提出的结构更加简单的分布式多智能体系统。所谓一致性是指多智能体系统中的个体在局部协作和相互通信下，随时间的演化最终使得所有智能体在某个目标量上趋于一致，描述了智能体之间相互作用、传递信息的过程。在过去的十几年，多智能体系统的一致性问题已被广泛应用在无人机编队控制、交通车辆控制、网络的资源分配等众多领域，成为当前工程领域研究的热点 [55–57]。一致性的数学定义有以下两种：

**定义 1.2.3** [58](领导者-跟随一致性) 假设多智能体系统中有  $N + 1$  个智能体，记第  $i$  个智能体的状态变量为  $x_i(t)$ ,  $i \in I[0, N]$ , 如果

$$\lim_{t \rightarrow \infty} |x_i(t) - x_0(t)| = 0$$

对所有  $i \in I[0, N]$  成立，则称多智能体系统达到状态一致。其中，智能体 0 被称作是领导者，其他的智能体  $i$ ,  $i \in I[1, N]$  被称作是跟随者。

**定义 1.2.4** [58](无领导者一致性) 假设多智能体系统中有  $N$  个智能体，记第  $i$  个智能体的状态变量为  $x_i(t)$ ,  $i \in I[1, N]$ , 如果

$$\lim_{t \rightarrow \infty} |x_j(t) - x_i(t)| = 0, \forall i \neq j$$

对所有  $i, j \in I[1, N]$  成立，则称多智能体系统达到一致性。

现实的网络环境是复杂多变的，多智能体系统在实际运行中普遍具有非线性特性和时滞。因此，非线性时滞多智能体系统的一致性问题受到越来越多的关注 [59–62]。相比时滞复杂网络的同步，非线性时滞多智能体系统的一致性则更强调智能体之间的局部信息的交流方式。近几年来，国内外的大量学者沿着不同的思路和方法对非线性时滞多智能体的一致性从多个方面进行了深入研究，不同的一致性算法被开发。

**领导者-跟随一致性** 在带有领导者的非线性时滞多智能体系统中，领导者代表着跟随者共同跟踪的目标或者整个系统的共同利益，通过对跟随者设计合适的控制协议，使得领导者和跟随者的最终状态达到一致。目前，带有领导者的一致性主要包括单个领导者的一致性和多个领导者的一致性 [63–70]。对于具有单个领导者的非线性时滞多智能体系统，通过设计连续的反馈控制器，非线性时滞多智能体系统的领导者-跟随一

致性问题在文献 [64–66, 70] 中被充分讨论。而通过设计不连续的分布式脉冲控制策略, 文献 [67, 70] 提出了不连续的分布式脉冲牵制控制策略, 分别研究了具有时变时滞和分布式时滞的非线性多智能体系统的领导者-跟随一致性。当领导者具有高维的动态且无法进行测量时, Yue 等人讨论了非线性时滞多智能体系统一致性跟踪问题 [68]。当具有多个领导者时, Li 等人利用分布式自适应输出反馈控制器解决了不确定的非线性时滞多智能体系统包围控制问题 [69], 此时跟随者的状态最终渐近收敛到领导者所形成的凸包内。

**切换拓扑下一致性** 在非线性时滞多智能体系统的运行过程中, 由于移动的智能体超出彼此的有效检测范围, 或者移动的智能体存在未知障碍物等原因, 智能体之间的通信链路可能发生变化 (中断或增加), 这意味着网络拓扑是动态变化的 [71–75]。其中, 切换拓扑是最重要也是研究最多的动态拓扑, Jiang 等人讨论了在采样控制下具有切换拓扑非线性时滞多智能体系统的非脆弱  $H_\infty$  一致性跟踪问题 [71]。Li 等人通过设计分布式数据驱动的一致性协议分析了具有切换拓扑的非线性时滞多智能体系统的输出一致性 [72]。基于脉冲控制, Yaghoubi 等人研究了具有切换拓扑的分数阶非线性时滞多智能体系统的集群一致性 [73]。特别地, 当网络拓扑是随机切换时, 具有马尔可夫切换拓扑的非线性时滞多智能体系统的一致性问题在文献 [74, 75] 中被讨论, 其中网络拓扑之间服从马尔可夫过程, 马尔可夫过程的状态空间相当于所有可能的拓扑。

**具有通信时滞一致性** 由于智能体之间的传输速度受限以及通道带宽有限等原因, 多智能体系统不可避免地存在通讯时滞。通常, 通信时滞的存在会降低多智能体系统的系统性能甚至导致系统的不稳定。因此, 研究具有通信时滞的多智能体的一致性是非常有必要的。近年来, 具有通信时滞的非线性多智能体系统的一致性得到了研究者们极大的关注, 一些有趣的结果被相继提出 [76–80]。例如, Sharifi 等人研究了在混合连续控制下具有通信常时滞的非线性多智能体系统的有限时间一致性问题 [76]。通过设计可靠控制器, Subramanian 等人分析了具有时变通信时滞的一阶非线性多智能体系统领导者-跟随一致性问题 [77]。Jiang 等人通过引入非周期性数据采样控制机制, 研究了具有通信时滞的高阶非线性多智能体系统一致性追踪问题 [78]。当网络拓扑图是具有拮抗作用的符号图时, Lu 等人研究了具有不同的通信时滞的非线性多智能体系统的二部一致性问题 [79, 80]。

**局部一致性** 目前, 许多研究人员已经对非线性时滞多智能体系统进行了大量研究, 但不难发现, 大多数一致性的理论结果都是建立在理想通信条件下且要求智能体的非线性项满足全局 Lipschitz 条件, 因此所得都是全局一致性结果, 这些全局性的假

设具有一定的局限性和保守性。因此，近几年非线性时滞多智能体系统的局部一致性得到关注 [81, 82]。为了避免全局 Lipschitz 假设带来的保守性，Qian 等人通过定义智能体的加权平均状态并应用局部线性化研究了非线性时滞多智能体的局部一致性 [81]。此外，当执行器达到饱和时，一致性无法对多智能体系统的所有初值都实现，只能在有界初值下才能实现。Liu 等人研究了非线性时滞多智能体系统的局部一致性并设计优化算法估计出吸引域 [82]。目前，对非线性时滞多智能体的局部一致性分析还较少，值得进一步深入研究。

## 1.3 脉冲控制的研究背景及现状

### 1.3.1 脉冲控制的研究背景

20 世纪 90 年代，随着脉冲微分方程理论在控制领域中应用的不断深入以及工业上越来越多的元件以脉冲微分方程建模，例如捕食者-食饵系统、混沌扩频系统和纳米电子器件等，脉冲控制系统引起了数学家、物理学家以及工程师的极大兴趣，并获得了突飞猛进的发展 [83]。随后，脉冲控制的基本框架被提出 [84]，不同于连续的控制方法，脉冲控制仅通过瞬时改变系统的状态就能达到控制效果，为控制领域提供一种全新的控制策略，其定义如下：

**定义 1.3.1** [84] 给定一个状态变量为  $z \in \mathbb{R}^n$  的非脉冲的机构  $\mathcal{P}$ ，控制时间序列  $T = \{t_k\}$ ,  $t_k \in \mathbb{R}$ ,  $t_k < t_{k+1}$ ,  $k \in \mathbb{Z}_+$  和控制律  $\mathcal{I}(k, x) \in \mathbb{R}^n$ , 如果在每个脉冲时刻  $t_k$ , 状态变量  $z$  被瞬时改变为  $z(t_k^-) + \mathcal{I}(k, z)$ , 使得当  $t \rightarrow \infty$  时, 系统的输出  $y = g(t, z)$ ,  $g : \mathbb{R}_+ \times \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $y \in \mathbb{R}^m$  逼近目标  $y^* \in \mathbb{R}^m$ , 该控制方法被称为脉冲控制。

其中，机构  $\mathcal{P}$  中至少有一个状态变量被瞬时改变为控制律给定的任意值，这意味着并不是机构  $\mathcal{P}$  中所有的物理系统都能被施加脉冲控制，也可能同时受到非脉冲控制。另一方面，脉冲控制的时间序列  $\{t_k\}$ ,  $k \in \mathbb{Z}_+$  可以是周期的，也可以非周期的。

脉冲控制的数学基础是脉冲微分方程，根据脉冲发生的不同特征，主要有以下三种类型：

(1) **脉冲控制发生在固定时刻** 脉冲控制发生在时间序列  $\{t_k\}$ ,  $k \in \mathbb{Z}_+$  上，即所谓的依赖于时间的脉冲控制，其脉冲微分方程可以写成

$$\begin{cases} \dot{z}(t) = f(t, z(t)), & t \neq t_k, k \in \mathbb{Z}_+, \\ \Delta z(t_k) = \mathcal{I}_k(z(t_k^-)), \end{cases}$$

其中  $\Delta z(t_k) = z_i(t_k) - z_i(t_k^-)$ ,  $z_i(t_k^-) = \lim_{t \rightarrow t_k^-} z_i(t)$ ,  $z_i(t_k) = z_i(t_k^+) = \lim_{t \rightarrow t_k^+} z_i(t)$ 。

**(2) 脉冲控制发生在可变时刻** [85] 当系统的轨迹撞击增广相空间的超曲面  $\mathcal{H}(t, z) = 0$  时, 产生脉冲控制, 即所谓的依赖于状态的脉冲控制, 其脉冲微分方程可以写成

$$\begin{cases} \dot{z} = f(t, z), & \mathcal{H}(t, z) \neq 0, \\ \Delta z = \mathcal{I}(z(\tau_k(x))), & \mathcal{H}(t, z) = 0. \end{cases}$$

在不同初始条件下, 该类系统的解具有不同的脉冲点。在这种情况下, 系统的一个解可能会撞击同一个超曲面很多次, 导致脉动现象。另一方面, 系统的不同的解也可能在一段时间后重合, 并在此后表现为单一的解, 从而导致解的合并。因此, 该类脉冲控制的研究比发生在固定时刻的脉冲控制更加困难。

**(3) 自治的脉冲控制** [86] 这类系统的状态不依赖于时间  $t$ , 脉冲的发生依赖于系统的状态, 其脉冲微分方程可以写成

$$\begin{cases} \dot{z} = f(z), & z \notin \mathcal{S}, \\ \Delta z = \mathcal{I}(z), & z \in \mathcal{S}. \end{cases}$$

脉冲控制作为一种强大的不连续控制方法, 受到了大多数研究人员的青睐, 因为它仅在某些离散时刻改变系统的状态就能达到控制效果 [87, 88]。与状态反馈控制等连续控制方法相比, 瞬时的脉冲跳变增强了对外部干扰的鲁棒性, 并能有效减少信息传输量。因此, 脉冲控制具有成本更低、效率更高、鲁棒性更强、保密性更好等优点。

### 1.3.2 脉冲控制的研究现状

在过去的几十年里, 脉冲控制已经被广泛应用于各种领域, 例如市场中货币供用控制、深空航天器轨道修正、房室模型中静脉给药治疗以及通信保密等 [86, 89]。因而, 脉冲控制得到国内外学者的广泛重视, 大量的研究成果被报道 [90–95]。其中, Yang 等人通过脉冲控制研究了非线性动态系统的渐近同步并提出一种新的混沌通信方案 [90], 并给出了脉冲间隔的上界的估计。随后, Lu 等人提出了平均脉冲区间的概念, 消除了对脉冲间隔上下界的限制, 在脉冲控制下解决了复杂网络的同步问题 [91]。对于脉冲控制下的时滞复杂网路的同步问题, 借助于脉冲时滞系统的比较原理, 时滞复杂网络指数同步问题在文献 [93, 94] 中被讨论, 但是它们对于脉冲间隔的上下界都有严格的限制。Feng 等人则取消了这一限制, 利用平均脉冲区间的方法, 讨论了脉冲控制下具有状态时滞和耦合时滞的复杂网络指数同步问题 [95], 允许非均匀分布的脉冲控制信号的存在。

近几年来，随着对脉冲控制研究的逐渐深入，许多专家不再满足传统的脉冲控制，不同类型的脉冲控制相继被提出，并被推广到时滞复杂网络的动力学分析中，主要包括下面几种类型：

**时滞脉冲** 在实际操作过程中，由于频繁的数据采样导致的数据包冲突会造成丢包，这意味着脉冲控制中的瞬时跳变不仅取决于系统的当前状态，还取决于过去的状态，即时滞脉冲控制。在现实控制系统中，时滞脉冲控制广泛应用于股市市场、病虫害的防治、疫苗的脉冲免疫接种等，越来越多的学者们加入到对时滞脉冲控制的研究中来，许多重要的成果被报道 [96–102]。通常情况下，时滞的存在会降低脉冲控制器的控制性能。利用拓展的 Halanay 不等式和新型的脉冲微分方程，Zhang 等人讨论了时滞脉冲控制下基于忆阻器的耦合时滞递归神经网络的同步问题 [98]。基于分布式时滞脉冲控制，非线性多智能体系统的领导者-跟随一致性问题在文献 [100] 得到解决。另一方面，脉冲控制中时滞信号的存在可以促进系统的稳定性甚至镇定不稳定的系统。例如，Li 等人利用脉冲时滞不等式研究了时滞系统的指数镇定性 [97]，该研究成果充分考虑了脉冲中时滞的有益影响。受此启发，基于 Lyapunov 方法和平均脉冲区间的方法，He 等人讨论了时滞脉冲控制下复杂网络的全局指数同步问题 [102]，其中脉冲控制中的时滞的镇定作用被充分考虑。因此，脉冲控制中的时滞对复杂网络的动态性能具有两面性。

**牵制脉冲** 传统的脉冲控制需要在脉冲时刻对网络中所有的节点施加控制，但是当网络规模比较大的时候，不可能控制网络中的每一个节点，这是非常昂贵且不切实际的。为了解决这个问题，牵制脉冲控制策略被提出，其在脉冲控制的基础上进一步考虑了牵制控制，本质上是仅对大规模网络中的部分关键节点施加脉冲控制，通过节点之间的相互耦合影响网络中的其它节点，从而达到控制整个网络的目的。由于可以节约更多的网络资源，牵制脉冲控制更适合节点较多的复杂网络，被广泛用于解决时滞复杂网络的同步问题 [103–107]。通过牵制脉冲控制，Yang 等人讨论了具有时变时滞的耦合的反应-扩散神经网络的同步问题 [104]。随后，Wang 等人提出了具有记忆功能的牵制脉冲控制方法，分析了具有参数不匹配和时变时滞的复杂网络的集群同步问题 [105]。特别地，Zhou 等人考虑仅对网络中的单个节点施加脉冲控制，设计了牵制的单个脉冲控制策略，验证了只要正确选择网络的耦合强度和耦合时滞，牵制的单个脉冲控制仍然可以保证时滞复杂网络的同步 [106]。

**混杂脉冲控制** 由于传统的脉冲控制仅在离散时刻发挥作用，而在脉冲间隔内没有任何控制行为，为了获得更好的收敛性能，在脉冲控制间隔内进一步施加连续的控制，

即混杂脉冲控制。这种控制方法同时拥有脉冲控制和连续控制的优点。尤其适用于其中一种控制效果不佳甚至失效，另一种控制仍然会使系统达到所期望的控制效果。因此，混杂脉冲控制比单个控制方法具有控制性能更高、鲁棒性更强的优点。近几年，混杂脉冲控制被广泛用于时滞复杂网络的同步的研究 [108–110]。例如，通过设计包含自适应控制和脉冲控制的混杂脉冲控制策略，文献 [108] 研究了具有不确定扰动的时滞复杂网络同步问题。Yang 等人文献同时考虑了脉冲控制和执行器故障反馈控制，并在此混杂脉冲控制器下分析了具有切换拓扑的时滞复杂网络的几乎必然全局指数同步问题 [109]。Li 等人利用包括连续时间状态反馈控制和脉冲控制的混杂脉冲控制策略，讨论了具有耦合时滞的分数阶复杂网络的全局同步控制问题 [110]。

**事件触发脉冲控制** 为了降低与连续控制的偏差、确保控制性能，脉冲控制都需要频繁的采样频率。这意味着不管系统性能有无明显变化，脉冲控制都需要以固定的速度（基于时间触发）进行信息传输和控制更新，故不再适应于大规模的复杂网络。最近，资源利用率更高的控制方法——事件触发脉冲控制被提出，主要思想是将脉冲控制考虑到事件触发机制中。不同于传统的脉冲控制和事件触发控制，事件触发脉冲控制仅在系统的性能指标超出一定阈值时才更新脉冲控制信号，并且在触发的时间间隔内不执行任何控制指令，这极大地减少了发送的数据包数量，节约了通信资源，减少了网络问题对系统性能和稳定性的负面影响。近年来，一些有关事件触发脉冲控制的结果被提出 [111–114]。例如，Zhu 等人借助事件触发脉冲控制研究了连续时间动态系统的指数稳定并将结果推广到忆阻神经网络的同步 [112]。基于分布式事件触发的脉冲控制，Tan 等人研究了多智能体系统的一致性问题 [111]。最近，Hai 等人进一步将牵制控制和事件触发的脉冲控制相结合，提出了事件触发的牵制脉冲控制策略 [115]，讨论了分数阶时滞复杂网络的同步问题。尽管事件触发脉冲控制优势明显，但也会导致时滞复杂系统呈现更复杂的不连续行为，因而分析与设计更具挑战性。

**饱和脉冲控制** 在许多实际控制系统中，执行器饱和是不断发生的，这源于系统的物理特性的限制和安全运行的需求 [116–118]。当执行器达到饱和，如果输入较大的控制信号，执行器只能传输具有幅度或速率限制的信号，这就导致控制器状态被错误更新从而降低系统的控制性能。因此，在设计脉冲控制方法时充分考虑执行器饱和是非常有必要的。实际上，饱和脉冲系统在真实网络中广泛存在，例如工业污水处理系统、汽车操纵系统等。近年来，具有执行饱和的脉冲控制得到了关注，但是由于脉冲饱和项难处理以及吸引域难于估计，关于饱和脉冲控制下复杂网络的研究工作甚少，尚还处于起步阶段 [82, 119–123]。例如，通过设计饱和脉冲控制器，Li 等人文献研究了耦

合的时滞神经网络的同步问题 [123]，并将所得结果应用到图像加密中。从现有的饱和脉冲的结果中不难发现，收缩的不变集被普遍应用于吸引域的估计，因为它具有所有从集合出发的状态轨迹始终保持在其内部并最终收敛到原点的性质。也就是说系统的 Lyapunov 能量函数是单调递减的，这意味着利用收缩的不变集来估计吸引域对具有饱和脉冲控制的系统来说具有保守性。目前，迫切需要探索一种新的估计吸引域的方法。

## 1.4 预备知识

### 1.4.1 代数图理论

$\mathcal{G} = (\mathcal{V}, \mathcal{E})$  表示  $N$  阶有向图，其中  $\mathcal{V} = \{v_i, i \in I[1, N]\}$  为顶点集， $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$  为边集。有向图  $\mathcal{G}$  中的边  $\mathcal{E}_{ij}$  可由有序对  $(v_i, v_j)$ ， $i, j \in I[1, N]$  表示，其中  $v_i$  为起点， $v_j$  为终点，这意味着节点  $v_j$  可以接收来自节点  $v_i$  的信息， $v_i$  被称为父节点， $v_j$  被称为子节点。节点  $v_i$  的邻居可记为  $\mathcal{N}_i = \{v_j \in \mathcal{V} | (v_j, v_i) \in \mathcal{E}\}$ 。图  $\mathcal{G}$  中从节点  $v_i$  到节点  $v_j$  的有向路径是有向边序列  $(v_{i_1}, v_{i_2}), (v_{i_2}, v_{i_3}), \dots, (v_{i_{m-1}}, v_{i_m})$ ，其中  $v_i = v_{i_1}$ ， $v_j = v_{i_m}$ 。如果存在一个节点  $v_r$ ，使得对于其他任意节点  $v_s \in \mathcal{V}$ ，都存在一条从  $v_r$  到  $v_s$  的有向路径，那么就称  $\mathcal{G}$  包含有向生成树。如果任意两个节点之间都存在有向路径，则此有向图  $\mathcal{G}$  是强连通的。

有向图  $\mathcal{G}$  的邻接矩阵  $\mathcal{A} = [a_{ij}]_{N \times N}$  定义为  $a_{ij} > 0$  当且仅当  $(v_j, v_i) \in \mathcal{E}$ ，且对所有的  $i \in I[1, N]$  有  $a_{ii} = 0$ 。拉普拉斯矩阵  $\mathcal{L} = [l_{ij}]_{N \times N}$  定义为  $\mathcal{L} = \mathcal{D} - \mathcal{A}$ ，其中  $\mathcal{D} = \text{diag}\{d_1, d_2, \dots, d_N\}$  是度矩阵，且元素  $d_i = \sum_{j \in \mathcal{N}_i} a_{ij}$  为节点  $v_i$  的入度。对于由  $N$  个跟随者和一个领导者组成的多智能体系统，网络拓扑由有向图  $\bar{\mathcal{G}}$  表示，其由有向图  $\mathcal{G}$  和节点  $v_0$  构成。定义  $\mathcal{H} = \mathcal{L} + \mathcal{B}$ ，其中  $\mathcal{B} = \text{diag}\{b_1, b_2, \dots, b_N\}$ ，如果领导者  $v_0$  是跟随者  $v_i$  的邻居，则  $b_i > 0$ ，否则  $b_i = 0$ 。

**定义 1.4.1** [55] 设矩阵  $A \in \mathbb{R}^{N \times N}$ ， $N \geq 2$ ，若存在置换矩阵  $P \in \mathbb{R}^{N \times N}$ ，使得

$$P^T A P = \begin{bmatrix} A_{11} & \mathbb{O}_{k \times (N-k)} \\ A_{21} & A_{22} \end{bmatrix},$$

其中  $A_{11} \in \mathbb{R}^{k \times k}$ ， $A_{21} \in \mathbb{R}^{(N-k) \times k}$ ， $A_{22} \in \mathbb{R}^{(N-k) \times (N-k)}$  ( $1 \leq k \leq N-1$ )，则称矩阵  $A$  是可约的；否则，称矩阵  $A$  是不可约的。

**引理 1.4.1** [124] 假设  $A = [\tilde{a}_{ij}]_{i,j=1}^N \in \mathbb{R}^{N \times N}$ ，如果

$$(1) \tilde{a}_{ij} \geq 0, i \neq j, \tilde{a}_{ii} = -\sum_{j=1, j \neq i}^N \tilde{a}_{ij}, i \in I[1, N];$$

(2)  $A$  是不可约的,

那么下式成立:

(a)  $\text{rank}(A) = N - 1$ , 即 0 是  $A$  代数几何重数为 1 的特征值, 且  $A$  的所有非零特征值均有负实部;

(b) 假设  $\xi = (\xi_1, \xi_2, \dots, \xi_N)^T \in \mathbb{R}^N$  (不失一般性, 假设  $\sum_{i=1}^N \xi_i = 1$ ) 是矩阵  $A$  特征值等于 0 所对应的左特征向量。那么  $\xi_i > 0, \forall i \in I[1, N]$ 。令  $\xi_{\max} = \max\{\xi_i\}$ ,  $\xi_{\min} = \min\{\xi_i\}$ ;

(c) 令  $\Xi = \text{diag}\{\xi_1, \xi_2, \dots, \xi_N\}$ , 那么  $\tilde{A} = \Xi A + A^T \Xi$  是一个对称矩阵, 使得所有特征值都是实数并且满足  $0 = \lambda_1(\tilde{A}) > \lambda_2(\tilde{A}) \geq \dots \geq \lambda_N(\tilde{A})$ 。

**引理 1.4.2** [26] 如果有向图  $G$  是强连通的, 其拉普拉斯矩阵  $\mathcal{L}$  特征值等于 0 所对应的左特征向量为  $\tilde{\xi} = (\tilde{\xi}_1, \tilde{\xi}_2, \dots, \tilde{\xi}_N)^T$ , 不失一般性, 假设  $\sum_{i=1}^N \tilde{\xi}_i = 1$ , 那么存在正定对角矩阵  $\Xi = \text{diag}\{\tilde{\xi}_1, \tilde{\xi}_2, \dots, \tilde{\xi}_N\}$ , 使得  $\check{\mathcal{L}} = 1/2(\Xi \mathcal{L} + \mathcal{L}^T \Xi)$  是对称矩阵, 且  $\sum_{j=1}^N \check{l}_{ij} = \sum_{i=1}^N \check{l}_{ij} = 0, \forall i, j \in I[1, N]$ 。

**定义 1.4.2** [125] 考虑强连通图  $G$  的拉普拉斯矩阵  $\mathcal{L}$ , 代数连通度被定义为

$$\lambda_\mu(\mathcal{L}) = \min_{x^T \xi = 0, x \neq 0} \frac{x^T \check{\mathcal{L}} \chi}{x^T \Xi x}.$$

**引理 1.4.3** [63] 有向图  $\bar{G}$  包含有向生成树, 其根节点为领导节点  $v_0$ , 则  $\mathcal{H}$  的非对角项是非正的, 并且其所有特征值都具有正实部。

## 1.4.2 矩阵理论

**定义 1.4.3** 设  $A = (a_{ij})_{m \times n}$ ,  $B = (b_{ij})_{p \times q}$ , 则  $A$  和  $B$  的 Kronecker 积是一个  $mp \times nq$  的分块矩阵:

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}.$$

**引理 1.4.4** [126] 假设矩阵  $A$ 、 $B$ 、 $C$  和  $D$  具有适合维数, 则有下列性质:

- (1) 设  $k$  为常数, 则有  $k(A \otimes B) = (kA) \otimes B = A \otimes (kB)$ ;
- (2) 如果  $A$  与  $B$  为同型矩阵, 则有  $(A + B) \otimes C = A \otimes C + B \otimes C$ ,  $C \otimes (A + B) = C \otimes A + C \otimes B$ ;
- (3)  $(A \otimes B)^T = A^T \otimes B^T$ ;
- (4)  $(A \otimes B) \otimes C = A \otimes (B \otimes C)$ ;
- (5) 如果矩阵乘积  $AC$  和  $BD$  存在, 那么  $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$ ;
- (6) 如果  $A$  和  $B$  可逆, 则  $A \otimes B$  也是可逆的, 且  $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ 。

**引理 1.4.5** [27] 考虑向量  $y = [(I_N - \mathbf{1}_N \tilde{\xi}^T) \otimes I_n]x$ , 其中  $x, y \in \mathbb{R}^{Nn}$ , 则下列等式成立:

$$\begin{aligned} y^T(\Xi \otimes U)(\mathbf{1}_N \tilde{\xi}^T \otimes I_m) &= \mathbb{O}_{1 \times Nm}, \quad \forall U \in \mathbb{R}^{n \times m}, \\ y^T(\Xi \otimes U)(\mathbf{1}_N \otimes z) &= 0, \quad \forall z \in \mathbb{R}^m. \end{aligned}$$

**引理 1.4.6 (Schur 补引理)** [126] 假设对称矩阵  $M$  可以分成  $M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$ , 其中  $M_{11}$ 、 $M_{22}$  是对称方阵。下列三个条件是等价的:

- (i)  $M < 0$ ;
- (ii)  $M_{11} < 0$ ,  $M_{22} - M_{21}M_{11}^{-1}M_{12} < 0$ ;
- (iii)  $M_{22} < 0$ ,  $M_{11} - M_{12}M_{22}^{-1}M_{21} < 0$ 。

**引理 1.4.7** [127] 对于任意的向量  $x, y$ , 适合维数的正定矩阵  $\mathcal{Q}$ , 和任意的常数  $\varepsilon > 0$ , 有下列的不等式成立:

$$2x^T y \leq \varepsilon x^T \mathcal{Q} x + \varepsilon^{-1} y^T \mathcal{Q}^{-1} y.$$

**引理 1.4.8** [96] 对于任意的向量  $x, y, z$ , 和任意的常数  $\varepsilon_1 > 0, \varepsilon_2 > 0$ , 有下列的不等式成立:

$$(x + y + z)^T(x + y + z) \leq (1 + \varepsilon_1)(1 + \varepsilon_2)x^T x + (1 + \varepsilon_1^{-1})(1 + \varepsilon_2)y^T y + (1 + \varepsilon_2^T)z^T z.$$

## 1.5 本文的主要内容和创新点

### 1.5.1 主要研究内容

在 Lyapunov 稳定性理论框架下，本文分析了不同类型脉冲控制下时滞复杂网络的动力学。首先，设计了包含时滞脉冲控制和连续时间状态反馈控制的混杂脉冲控制策略，分别研究了耦合多稳态时滞忆阻神经网络的动态多同步和静态多同步问题，其中对脉冲控制中时变时滞的上界不再有任何限制，进一步降低了相关结论的保守性。随后，讨论了在事件触发脉冲控制下时滞复杂网络的同步问题，设计了新型的依赖于 Lyapunov 函数的事件触发脉冲控制机制，得到了不依赖于时滞的同步判据，减少了不必要的网络资源浪费。最后，通过设计分布式饱和脉冲控制分析了时滞复杂网络的局部动力学，开发了一种全新的估计吸引域的方法，建立了相关的优化问题，并给出了最接近真实吸引域的估计值。

全文共包括五章，其结构框架如图 1.1 所示，主要研究内容如下：

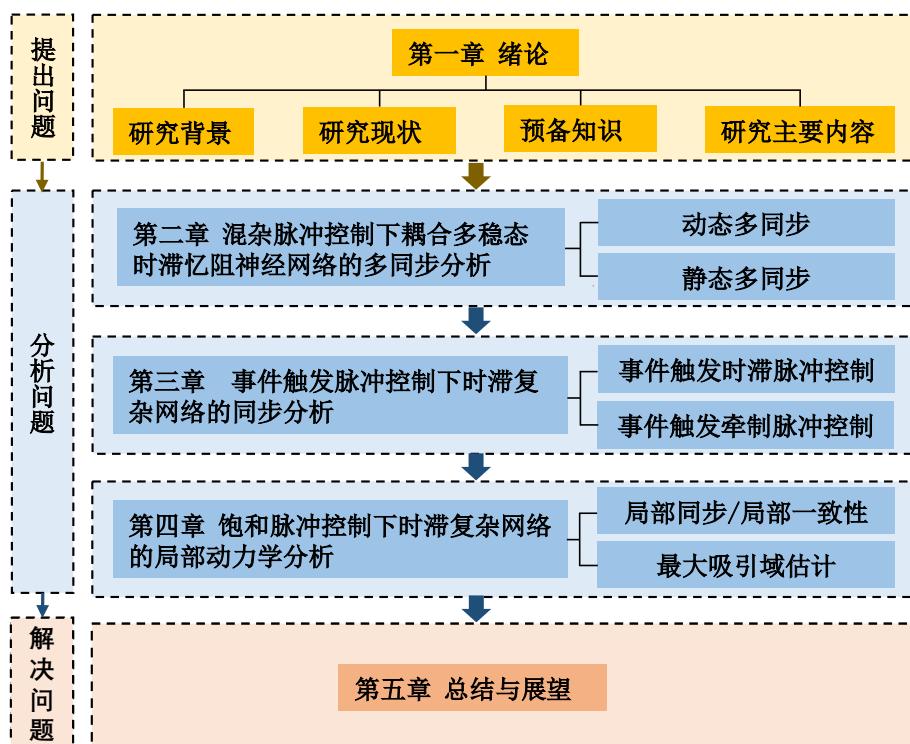


图 1.1: 论文的结构框架

第一章概述了时滞复杂网络以及脉冲控制理论的研究背景和研究现状，给出了所需的代数图论和矩阵理论的相关定义与引理，并总结出主要内容和创新点。

第二章主要研究了在混杂脉冲控制下耦合多稳态时滞忆阻神经网络的动力多同步以及静态多同步问题，其数学模型为

$$\begin{aligned}\dot{x}_i(t) \in & -(\check{C} + co[-\hat{C}, \hat{C}])x_i(t) + (\check{B} + co[-\hat{B}, \hat{B}])f(x_i(t)) \\ & + (\check{D} + co[-\hat{D}, \hat{D}])f(x_i(t - \eta(t))) + \alpha \sum_{j=1}^N a_{ij} \Gamma x_j(t) + I(t) + u_i(t),\end{aligned}$$

和

$$\begin{aligned}\dot{S}(t) \in & -(\check{C} + co[-\hat{C}, \hat{C}])S(t) + (\check{B} + co[-\hat{B}, \hat{B}])f(S(t)) \\ & + (\check{D} + co[-\hat{D}, \hat{D}])f(S(t - \eta(t))) + I(t).\end{aligned}$$

针对此网络，设计包括时滞脉冲控制和连续时间状态反馈控制的混杂脉冲控制器：

$$u_i(t) = -K \text{sign}(e_i(t)) - \mu \sum_{k=1}^{\infty} e_i(t - \tau(t)) \delta(t - t_k), \quad k \in \mathbb{Z}_+.$$

基于状态空间划分方法和激励函数的几何特性，给出了系统具有多个局部指数稳定的周期轨道或平衡点的充分条件。然后，利用新型的 Halanay 微分不等式和脉冲控制理论，分别获得基于 LMIs 的动态多同步以及静态多同步充分性判据。最后，给出了一个数值例子来验证所设计控制器的可行性和有效性。

第三章研究了事件触发脉冲控制下具有耦合时滞的复杂网络的全局同步问题。首先，分析了如下具有耦合常时滞的复杂网络的全局指数同步：

$$\dot{x}_i(t) = Bx_i(t) + Cf(x_i(t)) + c \sum_{j=1}^N a_{ij} \Gamma x_j(t - \tau_0), \quad i \in I[1, N],$$

和

$$\dot{\delta}z(t) = B\delta z(t) + Cf(\delta z(t)).$$

设计事件触发时滞脉冲控制机制：

$$t_k = \inf\{t > t_{k-1} : \Phi(t, z(t)) \leq 0\}, \quad k \in \mathbb{Z}_+,$$

其中  $\Phi(t, z(t)) = e^{\alpha_k - \mu t} \{ [e^{\mu t_{k-1}} z^T(t_{k-1})(I_N \otimes P)z(t_{k-1})] \vee (e^{\mu t_0} z_{t_0}^T(I_N \otimes P)z_{t_0}) - z^T(t)(I_N \otimes P)z(t) \}$ ，此机制是将时滞脉冲控制考虑到事件触发机制中。通过构造合适的辅助函数，并利用递推法、脉冲控制理论和 Lyapunov–Razumikhin 技术，建立了基于 LMIs 的保守性更小的同步判据。

其次，进一步将牵制控制考虑到事件触发脉冲控制机制中，探讨了如下具有耦合比例时滞的基因振荡器网络的集群同步问题：

$$\dot{z}_i(t) = A_v z_i(t) + B_{1v} f(z_i(t)) - B_{2v} g(z_i(t)) + B_2 I_n + c \sum_{r=1}^{\kappa} \sum_{j \in \mathcal{S}_r} a_{ij} \Upsilon z_j(\mu t) + u_i(t),$$

和

$$\dot{w}_v(t) = A_v w_v(t) + B_{1v} f(w_v(t)) - B_{2v} g(w_v(t)) + B_2 I_n.$$

针对此网络，设计事件触发牵制脉冲控制器：

$$u_i(t) = \begin{cases} \sum_{l=1}^{\infty} K_l e_i(t_l) \delta(t - t_l), & i \in \mathcal{D}_v(t_l), \\ 0, & i \notin \mathcal{D}_v(t_l), \end{cases}$$

其中，事件触发机制为

$$t_k = \inf\{t > t_{l-1} : \Lambda(t) \geq 0\},$$

这里  $\Lambda(t) = (1+\lambda(t-t_0)) \sum_{v=1}^{\kappa} \sum_{i \in \mathcal{S}_v} e_i^T(t) P e_i(t) - p_l [(1+\lambda(t_{l-1}-t_0)) \sum_{v=1}^{\kappa} \sum_{i \in \mathcal{S}_v} e_i^T(t_{l-1}) P e_i(t_{l-1}) \vee \lambda_{\max}(P) \sum_{v=1}^{\kappa} \sum_{i \in \mathcal{S}_v} \|\varphi_i\|_{\mu}^2]$ 。值得注意的是触发时刻即为脉冲时刻，其由所设计的事件触发牵制脉冲控制机制产生。并且每个集群中只有一小部分节点依据所设计的算法在触发时刻被施加脉冲控制，进一步节约了网络资源。最后，提供了数值实例以证明所设计的事件触发机制的有效性和优越性。

第四章分析了分布式饱和脉冲控制下时滞复杂网络的局部动力学行为。首先，讨论了如下具有耦合时变时滞的鲁里叶网络的局部指数同步：

$$\dot{x}_i(t) = Ax_i(t) + Bf(\tilde{D}x_i(t)) + c \sum_{j=1}^N \omega_{ij} \Gamma x_j(t - q(t)) + \mathcal{U}_i(t), \quad i \in I[1, N].$$

针对此网络，设计如下具有执行器饱和的分布式脉冲控制输入：

$$\mathcal{U}_i(t) = \sum_{l=1}^{\infty} \text{sat}(u_i(t)) \delta(t - t_l), \quad l \in \mathbb{Z}_+,$$

其中  $u_i(t) = K \sum_{j \in \mathcal{N}_i} a_{ij} (x_j(t) - x_i(t))$ 。利用反证法、脉冲系统的比较原理和平均脉冲区间的方法，得到了不依赖于时滞的基于 BMIs 的局部同步判据。为了降低保守性，选取最新的具有更多松弛变量的改进凸包表示法来处理分布式饱和的脉冲项，并且开发

了一种与收缩不变集完全不同的估计吸引域方法。

其次，研究了分布式饱和脉冲控制下如下具有切换拓扑的非线性时滞多智能体系统的局部一致性：

$$\begin{cases} \dot{x}_0(t) = Ax_0(t) + Bf(x_0(t - \tau(t))), \\ \dot{x}_i(t) = Ax_i(t) + Bf(x_i(t - \tau(t))) + u_i(t). \end{cases}$$

设计分布式饱和脉冲控制器：

$$u_i(t) = \sum_{k=1}^{\infty} \text{sat}(w_i(t))\delta(t - t_k), \quad k \in \mathbb{Z}_+,$$

其中  $\omega_i(t) = F\left(\sum_{j \in \mathcal{N}_i} a_{ij}^{\sigma(t)}(x_j(t) - x_i(t)) + b_i^{\sigma(t)}(x_0(t) - x_i(t))\right)$ 。利用反证法、Lyapunov-Razumikhin 技术，建立起相应的局部一致性判据。通过构造依赖于脉冲时刻的复合的 Lyapunov 函数，进一步降低了受控系统的保守性。为了估计出最大的吸引域，通过适当的矩阵变换，建立起基于 LMIs 的优化问题，并通过 Matlab 软件中的 Yalmip 工具箱求解相应的最大吸引域的数值解。最后，通过数值实例验证了所提理论结果的有效性和优越性。

第五章总结了全文的主要内容，并对未来的研究工作进行展望。

### 1.5.2 主要创新点

本文通过设计不同类型的脉冲控制，对时滞复杂网络的多种动力学行为进行了全面分析：通过设计混杂脉冲控制器，实现了耦合多稳态时滞忆阻神经网络的动态多同步以及静态多同步，脉冲控制中时变时滞的上界小于脉冲间隔的限制被取消，降低了结果的保守性。此外，设计了两种依赖于 Lyapunov 函数的事件触发脉冲控制协议，分别获得了时滞复杂网络不依赖于时滞的全局指数同步和集群同步判据，降低了不必要的网络资源的浪费。最后，通过设计分布式的饱和脉冲控制，对时滞复杂网络的局部动力学进行分析，并开发了一种全新的估计吸引域的方法，通过建立相关的优化问题估计出最大的吸引域。

文中结果是对现有一些结果的丰富和扩展，主要创新点具体如下：

- 探索了耦合多稳态时滞忆阻神经网络的多同步问题，设计了由时滞脉冲控制和连续时间状态反馈控制构成的混杂脉冲控制器，分别给出了动态多同步和静态多同步的充分性判据。文中讨论了脉冲控制中时变时滞存在的情况，取消了脉冲控制中时变时滞的上界小于任意脉冲区间的限制，进一步降低了结果的保守性。

在已有的耦合时滞忆阻神经网络中，通常存在一个唯一的平衡态（平衡点或周期轨道或混沌吸引子）。为了提高神经网络的储存能力，基于状态空间划分方法和激励函数多分段的几何性质，给出了系统具有多个局部指数稳定的周期轨道或平衡点的充分条件。其次，在现有的多同步文章中，脉冲控制器没有考虑时滞的影响。为此，设计了一种由时滞脉冲控制和连续时间状态反馈控制组成的新型混杂脉冲控制器，再利用改进的 Halanay 不等式技术、微分包含以及脉冲控制理论，分别获得了耦合多稳态时滞忆阻神经网络的动态多同步和静态多同步的充分性判据。值得注意的是，系统的状态时变时滞可微的限制和脉冲控制中的时变时滞的上界小于任意脉冲区间的要求都被取消，即允许时滞区间中可能存在多个脉冲。

**2.** 讨论了具有耦合常时滞的复杂网络的全局指数同步问题，设计了基于 Lyapunov 函数的事件触发时滞脉冲控制协议，给出了全局指数同步判据。其中，取消了对 Lyapunov 函数单调性的限制，进一步降低了结果的保守性。

已有的脉冲控制大多都是基于时间触发来进行信息传输和控制更新，为了减少不必要的网络资源的浪费和提高控制效率，将时滞脉冲控制与事件触发机制相结合，设计了新型的基于 Lyapunov 函数的事件触发时滞脉冲控制协议，极大地减少控制成本，增加系统的鲁棒性，避免了控制冗余。通过构造合适的辅助函数并利用 Lyapunov-Razumikhin 技术和递推法，得到了不依赖于时滞的时滞复杂网络的全局指数同步判据。一般来说，Lyapunov 函数要求在触发的时间间隔内严格单调递减，以确保触发时间序列是无穷的。文中放弃了这一限制，允许 Lyapunov 能量函数在下一个触发瞬间之前是递增，并能保证在有限的触发次数下实现同步。最后，验证时滞复杂网络在所设计的事件触发时滞脉冲控制下能够有效地避免芝诺现象。

**3.** 针对具有耦合比例时滞的基因振荡器网络集群同步问题，建立起事件触发牵制脉冲控制机制，给出了集群同步判据。通过设计合适的算法，每个集群中只有一小部分节点在触发时刻被施加脉冲控制，进一步降低了控制成本。

进一步将事件触发脉冲控制机制推广到具有比例时滞的脉冲泛函微分方程，提出了新型的依赖于 Lyapunov 函数的事件触发脉冲控制机制。其中，触发时刻由所设计的事件触发机制产生，触发时刻即为脉冲时刻。目前已经存在的事件触发脉冲控制都是对大规模网络中所有的节点施加控制，从而导致网络具有较高的能耗。为了解决这一问题，进一步将牵制控制考虑到所建立的事件触发脉冲控制机制中，设计了事件触发牵制脉冲控制协议，解决了具有耦合比例时滞的基因振荡器网络集群同步问题，并给出了基于 LMIs 的集群同步判据。通过设计合适的算法，每个集群中只有一小部分

节点在触发时刻被施加脉冲控制。与已有的事件触发牵制控制不同，所设计的事件触发牵制脉冲控制仅在触发时刻瞬时改变网络中部分节点的状态，而在非触发的脉冲时刻不再施加任何控制，并且消除了芝诺行为。

4. 针对具有耦合时变时滞的有向鲁里叶网络，研究了分布式饱和脉冲控制下的局部指数同步和吸引域估计问题。开发了全新的吸引域的估计方法并建立优化问题分别给出吸引域和平均脉冲区间的最大估计。

目前，由于脉冲饱和项难处理以及吸引域难于估计，关于饱和脉冲控制下时滞复杂网络的研究工作甚少，尚处于起步阶段。通过采用最新改进的饱和函数凸包表示来处理饱和脉冲约束项，并利用反证法、脉冲系统的比较原理以及平均脉冲区间方法，给出了具有耦合时变时滞的鲁里叶网络的局部指数同步判据并建立优化问题估计出最大的吸引域。目前，已有的关于饱和脉冲的结果基本上都是借助于收缩不变集来估计吸引域，这要求 Lyapunov 函数在此集合中的任意非零点上是严格单调递减的。本文取消了这一限制，开发了一种保守性更小的估计吸引域的方法。其次，由于耦合的时变时滞的上界与局部同步判据无关，仅与收敛指数有关，因此，时滞的上界小于某个常数的限制被取消，使得所得结果保守性更小。最后，只要平均脉冲区间满足一定条件，不再对脉冲间隔的上下界有任何限制，这意味着允许非均匀分布脉冲信号的存在。

5. 通过设计分布式饱和脉冲控制协议，讨论了具有切换拓扑的非线性时滞多智能体的局部一致性。构造了一种全新的依赖于脉冲时刻的复合型 Lyapunov 函数，以获得最大的吸引域估计。

已有一致性结果主要集中在全局一致性上，本文讨论了具有执行器饱和切换拓扑的非线性时滞多智能体系统的局部一致性问题，利用 Lyapunov-Razumikhin 技术和反证法，得到了不依赖于时滞的局部一致性判据并给出了最大吸引域的估计。由于结构简单，目前存在的结果大多数都是基于一般的二次 Lyapunov 函数。然而，利用二次 Lyapunov 函数估计出的吸引域与真实的吸引域之间存在较大差距。为了降低由二次 Lyapunov 函数带来的保守性，通过构造了一个具有更多辅助矩阵的复合的依赖于脉冲时刻的 Lyapunov 函数，并采用了一种具有更多可用的松弛变量的改进的凸包表示法，在最大程度上降低了保守性，使得所估计的吸引域更接近真实吸引域。

本文部分结果发表在 Neural Networks, IEEE Transactions on Cybernetics, IEEE Transactions on Neural Networks and Learning Systems 等国际刊物上，部分结果已投稿至 IEEE Transactions on Automatic Control, IEEE Transactions on Systems, Man, and Cybernetics: Systems 等期刊。具体详见作者发表论文的清单。

## 第二章 两种基于rényi熵的分数阶累计 剩余时间序列度量

### 2.1 引言

在信息论中，香农熵（或称信息熵）被提出来衡量随机变量的信息量（cite1）。给定一个随机变量  $X$ ，其概率密度函数为  $F(x)$ ，则  $X$  的香农熵 (SE) 定义为：

$$SE(X) = - \int_0^\infty F(x) \log F(x) dx, \quad (2.1.1)$$

其中， $\log$  是以 2 为底的对数。香农熵的离散形式是基于离散概率分布  $P(X)$ ，给定一个随机变量  $X_t = \{x_1, x_2, \dots, x_t\}$ ，相应的概率为  $P(X) = P(x = x_i) \forall i = 1, 2, \dots, t$ ，离散香农熵 (DSE) 可定义为：

$$DSE(X) = - \sum_{i=1}^t P(X = x_i) \log(P(X = x_i)) dx. \quad (2.1.2)$$

SE的计算公式表明，概率越高，信息的不确定性就越小，而所需信息的预期数量则以比特为单位，用对数来表示。由于其稳健的数学基础和固有的灵活性，SE 在信息检索 [?]、数据处理 [?] 和机器学习 [?] 等不同领域得到了广泛应用。然而，一些特性限制了 SE 在时间序列分析中的应用。例如，单独计算每个数据点的概率会导致对异常值的敏感性。此外，香农熵只关注每个事件的概率，而忽略了每个事件的值，这使得信息不确定性的计算不可靠。为了解决上述问题，Rao 等人提出了基于生存函数和累积分布函数的累积残差熵 [?]，其定义为：

$$CRE(X_t) = - \int_0^\infty \hat{F}(x) \log \hat{F}(x) dx, \quad (2.1.3)$$

其中  $\hat{F}(x) = 1 - F(x) = P(X \geq x_i)$  是生存函数。CRE 是 SE 的扩展，考虑了概率分布的累积残差。CRE 包含了之前观测的残差信息，而 SE 只考虑了当前观测的熵。换句话说，CRE 考虑了过去观测值的累积信息，使其成为时间序列中整体不确定性或无序性的度量，同时考虑了历史残差 [?]。与此类似，CRE 也可以离散形式存在，给定一

一个非负离散随机过程  $X$ , 离散累积残差熵 (DCRE) 可以定义为:

$$DCRE(X_t) = - \sum_{i=1}^t P(X \geq x_i) \log P(X \geq x_i)(x_i - x_{i-1}), \quad (2.1.4)$$

其中,  $P(X \geq x_i)$  是  $X$  的累积分布函数 (CDF), 也可用于计算  $P(a < X \leq b)$  形式的事件概率, 方法是求出在  $a$  和  $b$  处评估的 CDF 之差:  $P(X \geq x_i)$ 。

在 SE 和 CRE 的基础上, 人们提出了一些度量方法来量化时间序列的各种特征或属性。其中, Kullback-Leibler (KL) 散度又称相对熵, 是对两个随机变量概率分布之间相似度的一种度量。给定两个随机变量  $X$  和  $Y$  的概率分布分别为  $F(x)$  和  $G(x)$ , 则  $X$  和  $Y$  之间的 KL 散度  $KL(X, Y)$  定义为:

$$KL(X, Y) = - \int_0^\infty F(x) \log \frac{F(x)}{G(x)} dx, \quad (2.1.5)$$

KL 散度量化了使用为  $G(x)$  设计的最佳编码来表示从  $F(x)$  采样的数据所需的平均额外比特数。它衡量了用一种分布逼近另一种分布时损失或获得的信息。KL 散度总是非负的, 而且只有当  $F(X) = G(Y)$  时,  $KL(X, Y) = 0$ 。基于 CRE 的概念, KL 散度可以扩展到累积残差 KL 散度 (CEKL) [?]:

$$CRKL(X, Y) = - \int_0^\infty \hat{F}(x) \log \frac{\hat{F}(x)}{\hat{G}(x)} dx, \quad (2.1.6)$$

另一种时间序列度量是传递熵 (TE)。传递熵的形式与 KL 散度相似, 也可以写成概率分布乘以对数项的形式。与 KL 散度相比, TE 是两个随机过程之间信息定向流动的度量, 因此, TE 通常用于识别因果关系中因果之间的时间阶梯。从数学表达式上来看 TE 本质上是与变量历史数据影响的条件互信息 (CMI)。给定两个连续的随机过程  $X$  和  $Y$ , 从  $X$  到  $Y$  的  $rmTE$  可以表示为 [?]:

$$TE_{Y \rightarrow X} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x_{t+1}, x_t^{(k)}, y_t^{(l)}) \log_2 \frac{p(x_{t+1}|x_t^{(k)}, y_t^{(l)})}{p(x_{t+1}|x_t^{(k)})} dx dy \quad (2.1.7)$$

其中  $k$  和  $l$  是马尔可夫阶数, 它控制变量  $X$  和  $Y$  历史信息的滞后时间。一些研究者在  $rmTE$  中用 CDF 代替了 PDF, 以使用 CRE 的特性 (cite10,11,12)。然而, 这些工作没有给出累积残余传递熵 (CRTE) 的正式定义, 这促使我们进一步将 TE 扩展到具有累计剩余度量性质的 CRTE。

总之, KL 散度和 TE 都是信息传递或流动的度量, 但它们在度量这种流动的方

法上有所不同。KL 散度测量的是各系统概率分布之间的差异，因此，KL 常被用作距离度量 [?, ?, ?, ?]。TE 衡量的是系统中变量间信息的定向流动，更适合研究具有时间依赖性的复杂系统，可用于识别系统中因果互动的方向和强度 [?, ?, ?, ?]。

此外，作为SE的分数阶形式扩展，Rényi引入了Rényi熵（RE）来度量时间序列的复杂性 [?]。令  $\alpha \geq 0$ ，随机变量  $X$  的  $\alpha$  阶 RE 定义为：

$$H_\alpha(X) = \frac{1}{1-\alpha} \log_2 \left( \sum_{x \in \mathcal{X}} p(x)^\alpha \right) \quad (2.1.8)$$

其中， $X$  是分布为  $p(x)$  的离散型随机变量， $\mathcal{X}$  是  $X$  可能结果的集合。RE 是一个关于  $\alpha$  的非负的、单调递减的函数 [?]。根据 L'Hospital 规则，当  $\alpha = 1$  时，RE 收敛到 SE。当  $\alpha > 1$  时，RE 更偏向于代表变量  $X$  分布中的高频事件。当  $\alpha = 2$  时，RE 转换为 collision entropy；当  $\alpha = +\infty$  时，RE 转换为 minimum entropy。此外，当  $\alpha \leq 1$  时，RE 中强调的是  $X$  的低频事件，而当  $\alpha > 1$  时，强调的是  $X$  中的高频事件 [?]。与传统的 SE 相比，RE 能够根据需要自适应地测量复杂分布。由于 Rényi 熵在复杂分布上的有效性，一些研究者将 KL 散度和 TE 扩展到了分数阶 [?, ?]，即所谓的 Rényi 散度 (RKL) 和 Rényi 传递熵 (RTE)。这些工作也启发我们在 RE 的基础上将 CRKL 和 CRTE 扩展到分数阶，从而更全面地衡量时间序列之间的关系。

综上，本章引入了两种分数阶累积剩余时间序列度量，分别称为 rényi cumulative residual kullback-leibluer (RCRKL) 和 rényi cumulative residual transfer entropy (RCRTE)。这两种度量方法都基于累积残差熵和 rényi 熵，其中 RCRKL 主要用于比较时间序列之间的相似性，而 RCRTE 则用于分析时间序列数据中的信息流和变量之间的因果关系。同时，我们证明了 RCRKL 作为距离度量的一些性质，并研究了 RCRTE 的零点和边界。为了验证本章所提出度量的有效性，我们分别在仿真数据和真实数据上进行了数值验证。在仿真数据的验证部分，我们研究了参数变换对两种度量的影响。在真实数据验证部分，我们采用了来自 RIOHTrack 的路面车辙数据，基于 RCRKL 作为距离度量，设计了一种新型的聚类算法，对不同路面的车辙进行聚类，并基于 RCRTE 对每个簇里的车辙进行因果关系检测。结果显示，基于 RCRKL 的聚类算法具有良好的聚类性能，而 RCRTE 也能有效检测出显著的因果关系。

## 2.2 Rényi 累积剩余KL散度及其性质

**定义 1.** 给定两个随机变量  $X$  和  $Y$ 。Rényi cumulative residual kullback-leibluer divergence (RCRKL) 的定义是：

$$RCRKL(X, Y) = \frac{1}{\alpha - 1} \log \left( \int_{-\infty}^{\infty} \hat{F}(x)^{\alpha} \hat{G}(y)^{1-\alpha} dx \right), \quad (2.2.1)$$

在这个公式中， $\hat{F}(x)$  表示随机变量  $X$  的生存函数， $\hat{G}(y)$  表示随机变量  $Y$  的生存函数。

在本节接下来的内容中，将提供 CRKL 的一些一般性质和相应的证明。

**命题 1.** 令  $X$  和  $Y$  是两个随机变量，并且  $\alpha \geq 1$ ，那么

$$RCRKL(X, Y) \geq 0,$$

**证明**

首先，回顾RCRKL公式为：

$$RCRKL(X, Y) = \frac{1}{\alpha - 1} \log \left( \int_{-\infty}^{\infty} \hat{F}(x)^{\alpha} \hat{G}(y)^{1-\alpha} dx \right),$$

由于  $\alpha \geq 1$ ，分母  $\alpha - 1$  是非负数。因此，只证明 RCRKL 中的对数项大于 1，即：

$$\int_{-\infty}^{\infty} \hat{F}(x)^{\alpha} \hat{G}(y)^{1-\alpha} dx \geq 1.$$

根据定义 [?], 生存函数  $\hat{F}(x)$  和  $\hat{G}(y)$  是非负值，其值介于 0 和 1 之间。因此，通过应用广义的 Hölder 不等式，我们可以得到下面的不等式：

$$\int_{-\infty}^{\infty} \hat{F}(x)^{\alpha} \hat{G}(y)^{1-\alpha} dx \geq \left( \int_{-\infty}^{\infty} \hat{F}(x) dx \right)^{\alpha} \left( \int_{-\infty}^{\infty} \hat{G}(y) dy \right)^{1-\alpha}.$$

同时，生存函数  $\hat{F}(x)$  和  $\hat{G}(y)$  的此积分为 1，即：

$$\int_{-\infty}^{\infty} \hat{F}(x) dx = \int_{-\infty}^{\infty} \hat{G}(y) dy = 1,$$

将这些值代入上一不等式，我们得出：

$$\begin{aligned} \int_{-\infty}^{\infty} \hat{F}(x)^{\alpha} \hat{G}(y)^{1-\alpha} dx &\geq 1^{\alpha} \cdot 1^{1-\alpha} \\ &= 1. \end{aligned}$$

因此，我们得出结论： $RCRKL(X, Y) \geq 0$  for  $\alpha \geq 1$ 。

证明完毕。

**命题 2.** 令  $X$  和  $Y$  为两个随机变量， $a, b$ , 和  $c$  为常数，那么

$$RCRKL(aX + b, cY + b) = a \cdot RCRKL(X, \frac{c}{a}Y),$$

其中  $\hat{F}(x)$  表示随机变量  $X$  的生存函数， $\hat{G}(x)$  表示随机变量  $Y$  的生存函数。

**证明**

首先，考虑待证等式：

$$RCRKL(aX + b, cY + b) = a \cdot RCRKL(X, \frac{c}{a}Y),$$

将 RCRKL 公式代入 LHS，可得：

$$\begin{aligned} RCRKL(aX + b, cY + b) &= \frac{1}{\alpha - 1} \log \left( \int_{-\infty}^{\infty} \hat{F}^{\alpha}(aX + b) \hat{G}^{1-\alpha}(cY + b) dx \right), \\ &= \frac{1}{\alpha - 1} \log \left( \int_{-\infty}^{\infty} \hat{F}_{aX+b}(x)^{\alpha} \hat{G}_{cY+b}(x)^{1-\alpha} dx \right), \end{aligned}$$

令  $u = aX + b$  and  $v = cY + b$ ，则上式可写为：

$$RCRKL(aX + b, cY + b) = \frac{1}{\alpha - 1} \log \left( \int_{-\infty}^{\infty} \hat{F}_{\frac{1}{a}(u-b)}(x)^{\alpha} \hat{G}_{\frac{1}{c}(v-b)}(y)^{1-\alpha} dx \right)$$

再将带证等式 RHS 中变量  $X$  和  $Y$  替换为  $u$  和  $v$ ，可得：

$$\begin{aligned} a \cdot RCRKL(X, \frac{c}{a}Y) &= a \cdot \frac{1}{\alpha - 1} \log \left( \int_{-\infty}^{\infty} \hat{F}_X(x)^{\alpha} \hat{G}_{\frac{c}{a}Y}(x)^{1-\alpha} dx \right) \\ &= \frac{1}{\alpha - 1} \log \left( \int_{-\infty}^{\infty} \hat{F}_{\frac{1}{a}(u-b)}(x)^{\alpha} \hat{G}_{\frac{1}{a}(v-b)}(y)^{1-\alpha} dx \right) \end{aligned}$$

显然，LHS = RHS。因此，我们有：

$$RCRKL(aX + b, cY + b) = a \cdot RCRKL(X, \frac{c}{a}Y),$$

证明完毕。

**命题3.** 令  $X$  和  $Y$  为两个随机变量,  $0 \leq \lambda \leq 1$  且  $\alpha \geq 0$ , 则:

$$RCRKL(\lambda X_1 + (1-\lambda)X_2, \lambda Y_1 + (1-\lambda)Y_2) \leq \lambda RCRKL(X_1, Y_1) + (1-\lambda) RCRKL(X_2, Y_2).$$

**证明**

令  $Z_1 = \lambda X_1 + (1 - \lambda)X_2$  及  $W_1 = \lambda Y_1 + (1 - \lambda)Y_2$ , 其中  $0 \leq \lambda \leq 1$ .

那么,  $Z_1$  和  $W_1$  之间的 RCRKL 可以表示为:

$$RCRKL(Z, W) = \frac{1}{\alpha - 1} \log_2 \int_{-\infty}^{\infty} \hat{F}_{Z_1}(x)^{\alpha} \hat{G}_{W_1}(x)^{1-\alpha} \lambda dx,$$

利用生成函数的性质, 我们可以得到:

$$\begin{aligned} \hat{F}_{Z_1}(x) &= P(Z_1 \geq x) \\ &= P(\lambda X_1 + (1 - \lambda)X_2 \geq x) \\ &\leq \lambda P(X_1 \geq x) + (1 - \lambda)P(X_2 \geq x) \\ &= \lambda \hat{F}_{X_1}(x) + (1 - \lambda) \hat{F}_{X_2}(x) \end{aligned}$$

类似的, 我们有:

$$\hat{G}_{W_1}(x) \leq \lambda \hat{F}_{Y_1}(x) + (1 - \lambda) \hat{F}_{Y_2}(x),$$

将这些不等式代入 RCRKL 公式, 即可得出:

$$RCRKL(Z_1, W_1) \leq \frac{1}{\alpha - 1} \log_2 \left( \int_{-\infty}^{\infty} (\lambda \hat{F}_{X_1}(x) + (1 - \alpha) \hat{F}_{X_2}(x))^{\alpha} (\lambda \hat{G}_{Y_1}(x) + (1 - \alpha) \hat{G}_{Y_2}(x))^{1-\alpha} dx \right),$$

利用对数性质, 我们可以简化上式:

$$RCRKL(Z_1, W_1) \leq \lambda \cdot RCRKL(X_1, Y_1) + (1 - \lambda) \cdot RCRKL(X_2, Y_2).$$

因此, 可以证明到:

$$RCRKL(\lambda X_1 + (1 - \lambda)X_2, \lambda Y_1 + (1 - \lambda)Y_2) \leq \lambda \cdot RCRKL(X_1, Y_1) + (1 - \lambda) \cdot RCRKL(X_2, Y_2),$$

对于  $0 \leq \lambda \leq 1$  范围内的任何  $\lambda$  值和任何  $\alpha \geq 0$  来说都成立。

证明完毕。

**命题 4.** 当 $\alpha \geq 1$ 时, RCRKL满足三角不等式, 即:

$$RCRKL(X, Y) + RCRKL(Y, Z) \geq RCRKL(X, Z)$$

### 证明

利用 RCRKL 散度的定义, 我们可以把 LHS 写成:

$$\begin{aligned} RCRKL(X, Y) + RCRKL(Y, Z) &= \frac{1}{\alpha-1} \log \left( \int_{-\infty}^{\infty} \hat{F}_X^{\alpha}(x) \int_{-\infty}^{\infty} \hat{G}_Y^{1-\alpha}(x) dx \right) \\ &\quad + \frac{1}{\alpha-1} \log \left( \int_{-\infty}^{\infty} \hat{F}_Y^{\alpha}(y) \int_{-\infty}^{\infty} \hat{G}_Z^{1-\alpha}(y) dy \right), \end{aligned}$$

而 RHS 可以写成:

$$RCRKL(X, Z) = \frac{1}{\alpha-1} \log \left( \int_{-\infty}^{\infty} \hat{F}_X^{\alpha}(x) \int_{-\infty}^{\infty} \hat{G}_Z^{1-\alpha}(x) dx \right).$$

我们首先考虑RCRKL散度内部的对数项, 根据 Cauchy-Schwarz 不等式, 我们可以得到:

$$\sqrt{\int_{-\infty}^{\infty} \hat{F}_X^{\alpha}(x) \int_{-\infty}^{\infty} \hat{G}_Y^{1-\alpha}(x) dx} \cdot \sqrt{\int_{-\infty}^{\infty} \hat{F}_Y^{\alpha}(y) \int_{-\infty}^{\infty} \hat{G}_Z^{1-\alpha}(y) dy} \geq \int_{-\infty}^{\infty} \hat{F}_X^{\alpha}(x) \int_{-\infty}^{\infty} \hat{G}_Z^{1-\alpha}(x) dx.$$

将不等式两边平方, 就得到:

$$\int_{-\infty}^{\infty} \hat{F}_X^{\alpha}(x) \int_{-\infty}^{\infty} \hat{G}_Y^{1-\alpha}(x) dx \cdot \int_{-\infty}^{\infty} \hat{F}_Y^{\alpha}(y) \int_{-\infty}^{\infty} \hat{G}_Z^{1-\alpha}(y) dy \geq \left( \int_{-\infty}^{\infty} \hat{F}_X^{\alpha}(x) \int_{-\infty}^{\infty} \hat{G}_Z^{1-\alpha}(x) dx \right)^2.$$

对两边取对数, 我们有:

$$\log \left( \int_{-\infty}^{\infty} \hat{F}_X^{\alpha}(x) \int_{-\infty}^{\infty} \hat{G}_Y^{1-\alpha}(x) dx \right) + \log \left( \int_{-\infty}^{\infty} \hat{F}_Y^{\alpha}(y) \int_{-\infty}^{\infty} \hat{G}_Z^{1-\alpha}(y) dy \right) \geq 2 \log \left( \int_{-\infty}^{\infty} \hat{F}_X^{\alpha}(x) \int_{-\infty}^{\infty} \hat{G}_Z^{1-\alpha}(x) dx \right)$$

当  $alpha \geq 1$  时, 我们可以得到:

$$\begin{aligned} \frac{1}{\alpha-1} \log \left( \int_{-\infty}^{\infty} \hat{F}_X^{\alpha}(x) \int_{-\infty}^{\infty} \hat{G}_Y^{1-\alpha}(x) dx \right) + \frac{1}{\alpha-1} \log \left( \int_{-\infty}^{\infty} \hat{F}_Y^{\alpha}(y) \int_{-\infty}^{\infty} \hat{G}_Z^{1-\alpha}(y) dy \right) \geq \\ \frac{2}{\alpha-1} \log \left( \int_{-\infty}^{\infty} \hat{F}_X^{\alpha}(x) \int_{-\infty}^{\infty} \hat{G}_Z^{1-\alpha}(x) dx \right). \end{aligned}$$

因此，我们证明了

$$RCRKL(X, Y) + RCRKL(Y, Z) \geq 2RCRKL(X, Z).$$

又因为当  $\alpha \geq 1$  时，  $RCRKL(X, Y) \geq 0$ ，因此我们有：

$$RCRKL(X, Y) + RCRKL(Y, Z) \geq RCRKL(X, Z).$$

证明完毕。

## 2.3 Rényi 累积剩余传递熵及其性质

首先，我们先介绍累积残余传递熵（Cumulative Residual Transfer Entropy, CRTE）的定义，CRTE 是传递熵（Transfer Entropy, TE）的扩展，它将累积剩余熵（cumulative residual entropy, CRE）的概念纳入到两个离散马尔可夫过程之间的信息传递或因果关系的计算中。CRTE量化了在考虑了两个过程的历史依赖性的影响之后，根据另一个过程的过去历史预测一个过程的未来状态的累积剩余熵。**定义 2.** 给定两个随机变量  $X$  和  $Y$ ，从  $Y$  到  $X$  的 CRTE 被表示为：

$$CRTE(Y \rightarrow X) = - \int_0^\infty \int_0^\infty \hat{F}(x_{t+1}|x_t^{(k)}, y_t^{(l)}) \log_2 \frac{\hat{F}(x_{t+1}|x_t^{(k)}, y_t^{(l)})}{\hat{F}(x_{t+1}|x_t^{(k)})} dx dy, \quad (2.3.1)$$

基于CRTE的形式，我们再给出 Rényi 累积残余传递熵（Rényi cumulative residual transfer entropy, RCRTE）的定义：

**定义 3.** 将3.2.5中的  $H_{alpha}$  代入 2.3.3，CRTE可以扩展到分数阶，其定义为 Rényi 累积残差传递熵，RCRTE：

$$RCRETE(Y \rightarrow X) = \frac{1}{1-\alpha} \log_2 \left[ \frac{\int \rho_\alpha(x_t^l) \hat{F}^\alpha(x_{t+1}|x_t^l) dx}{\int \int \rho_\alpha(x_t^l, y_t^k) \hat{F}^\alpha(x_{t+1}|x_t^l, y_t^k) dx dy} \right] \quad (2.3.2)$$

其中， $\rho_\alpha(x) = \frac{\hat{F}^\alpha(x)}{\int \hat{F}^\alpha(x) dx}$  为分配分布。

同时，与传统的传递熵类似， $rmCRTE$  也是一个非对称度量。因此，为了表示 CRTE 的方向性，方程2.3.2也可以改写成条件熵差的形式：

$$RCRTE(Y \rightarrow X) = RCRE(x_{t+1}|x_t^{(l)}) - RCRE(x_{t+1}|x_t^{(l)}, y_t^k) \quad (2.3.3)$$

$$= RCRE(x_{t+1}, x_t^{(l)}) - RCRE(x_t^{(l)}) + RCRE(x_{t+1}, x_t^{(l)}, y_t^{(k)}) - RCRE(x_t^{(l)}, y_t^{(k)}) \quad (2.3.4)$$

其中,  $CRE(X|Y)$  是条件累积残差熵, 它表示了与  $Y$  相同的概率质量函数;  $RCRE(X, Y)$  是联合累积残差熵, 与  $p(y_t)$  相关联的结果对应于以  $Y = y_i$  为条件的概率累积残差熵, 即  $CRE(X|Y = y_i)$ 。

此外, 离散的 Rényi 累积残差传递熵 (discrete rényi cumulative residual transfer entropy, DRCRTE) 可以定义为:

$$DRCRTE_{Y \rightarrow X}^{\alpha} = \frac{1}{1-\alpha} \log_2 \left[ \frac{\sum_{x \in X} \rho_{\alpha}(x_t^{(l)}) \cdot F^{\alpha}(x_{t+1}|x_t^{(l)})}{\sum_{y \in Y} \sum_{x \in X} \rho_{\alpha}(x_t^{(l)}, y_t^{(k)}) \cdot \hat{F}^{\alpha}(x_{t+1}|x_t^{(k)}, y_t^{(l)})} \right]. \quad (2.3.5)$$

在本节的其余部分, 我们将给出 RCRTE 的一些性质和证明:

**命题 5.** RCRTE 随着  $\alpha$  的增加而单调递减。

首先, 观察 RCRTE 的定义可知, 为了证明 RCRTE 随  $\alpha$  的增加而单调递减, 我们需要证明 RCRTE 中对数项的随  $\alpha$  的增加而递减。

RCRTE 的定义公式为:

$$RCRTE(Y \rightarrow X) = \frac{1}{1-\alpha} \log \left( \frac{\int_0^{\infty} \hat{F}^{\alpha}(x_{t+1}, x_t^{(l)}) dx \int_0^{\infty} \int_0^{\infty} \hat{F}^{\alpha}(x_{t+1}, x_t^{(l)}, y_t^{(k)}) dxdy}{\int_0^{\infty} \hat{F}^{\alpha}(x_t^{(l)}) dx \int_0^{\infty} \int_0^{\infty} \hat{F}^{\alpha}(x_t^{(l)}, y_t^{(k)}) dxdy} \right)$$

我们将把对数中的分子和分母分别记为  $N(\alpha)$  和  $D(\alpha)$ , 则有:

$$\begin{aligned} N(\alpha) &= \int_0^{\infty} \hat{F}^{\alpha}(x_{t+1}, x_t^{(l)}) dx \int_0^{\infty} \int_0^{\infty} \hat{F}^{\alpha}(x_{t+1}, x_t^{(l)}, y_t^{(k)}) dxdy \\ D(\alpha) &= \int_0^{\infty} \hat{F}^{\alpha}(x_t^{(l)}) dx \int_0^{\infty} \int_0^{\infty} \hat{F}^{\alpha}(x_t^{(l)}, y_t^{(k)}) dxdy \end{aligned}$$

利用对数求和不等式, 我们可以得出: 对于任何正数  $a$  和  $b$ :

$$\log(a) - \log(b) \leq \frac{a-b}{b}$$

将这一不等式应用于  $N(\alpha)$  和  $D(\alpha)$ , 我们得到:

$$\log \left( \frac{N(\alpha_2)}{D(\alpha_2)} \right) - \log \left( \frac{N(\alpha_1)}{D(\alpha_1)} \right) \leq \frac{\frac{N(\alpha_2)}{D(\alpha_2)} - \frac{N(\alpha_1)}{D(\alpha_1)}}{\frac{N(\alpha_1)}{D(\alpha_1)}}$$

将不等式右侧简化，我们有：

$$\begin{aligned} \log\left(\frac{N(\alpha_2)}{D(\alpha_2)}\right) - \log\left(\frac{N(\alpha_1)}{D(\alpha_1)}\right) &\leq \frac{\frac{N(\alpha_2)}{D(\alpha_2)} - \frac{N(\alpha_1)}{D(\alpha_1)}}{\frac{N(\alpha_1)}{D(\alpha_1)}} \\ &\leq \frac{N(\alpha_2)D(\alpha_1) - N(\alpha_1)D(\alpha_2)}{D(\alpha_1)N(\alpha_1)} \\ &= \frac{N(\alpha_2)}{N(\alpha_1)} - \frac{D(\alpha_2)}{D(\alpha_1)} \end{aligned}$$

由于  $\alpha_1 < \alpha_2$ , 则  $\frac{N(\alpha_2)}{N(\alpha_1)} \leq 1$  and  $\frac{D(\alpha_2)}{D(\alpha_1)} \geq 1$ .

因此，我们可以得出：

$$\log\left(\frac{N(\alpha_2)}{D(\alpha_2)}\right) - \log\left(\frac{N(\alpha_1)}{D(\alpha_1)}\right) \leq \frac{N(\alpha_2)}{N(\alpha_1)} - \frac{D(\alpha_2)}{D(\alpha_1)} \leq 0$$

由于对数是单调递增函数，上述不等式意味着  $\frac{N(\alpha_2)}{D(\alpha_2)} \leq \frac{N(\alpha_1)}{D(\alpha_1)}$ 。因此，我们可以得出结论，RCRTE随着  $\alpha$  的增加而单调递减。

证明完毕。

**命题 6.** 令  $X$  和  $Y$  为两个随机变量，那么当且仅当随机变量  $X$  和  $Y$  的分布处处相同时， $\text{RCRTE}(Y \rightarrow X) = 0$ 。

要证明  $\text{RCRTE}(Y \rightarrow X) = 0$  当且仅当  $X$  和  $Y$  的分布处处相同时，我们需要证明该命题的充要性。

充分性： $\text{RCRTE}(Y \rightarrow X) = 0$  意味着  $X$  和  $Y$  的分布处处相同。

证明

如果  $\text{RCRTE}(Y \rightarrow X) = 0$ , 那么根据 RCRTE 的定义，我们有：

$$\frac{1}{1-\alpha} \log\left(\frac{\int_0^\infty \hat{F}^\alpha(x_{t+1}, x_t^{(l)}) dx \int_0^\infty \int_0^\infty \hat{F}^\alpha(x_{t+1}, x_t^{(l)}, y_t^{(k)}) dxdy}{\int_0^\infty \hat{F}^\alpha(x_t^l) dx \int_0^\infty \int_0^\infty \hat{F}^\alpha(x_t^{(l)}, y_t^{(k)}) dxdy}\right) = 0$$

这意味着该分数项的分子等于分母，即：

$$\int_0^\infty \hat{F}^\alpha(x_{t+1}, x_t^{(l)}) dx \int_0^\infty \int_0^\infty \hat{F}^\alpha(x_{t+1}, x_t^{(l)}, y_t^{(k)}) dxdy = \int_0^\infty \hat{F}^\alpha(x_t^l) dx \int_0^\infty \int_0^\infty \hat{F}^\alpha(x_t^{(l)}, y_t^{(k)}) dxdy$$

由于累积分布函数 (CDF)  $\hat{F}$  表示随机变量大于或等于某个值的概率，因此积分相等意味着联合 CDF  $\hat{F}(x_t^{(l)}, y_t^{(k)})$  等于边际 CDF  $\hat{F}(x_t^l)$  和  $\hat{F}(x_t^{(l)})$  的乘积。

因此，我们可以得出结论：如果  $\text{RCRTE}(Y \rightarrow X) = 0$ , 那么  $X$  和  $Y$  的分布处处相

同。

必要性：X 和 Y 的分布处处相同意味着  $\text{RCRTE}(Y \rightarrow X) = 0$ 。

### 证明

如果 X 和 Y 在所有地方都有相同的分布，这意味着它们的联合 CDF  $\hat{F}(x_t^{(l)}, y_t^{(k)})$  等于它们的边际 CDF  $\hat{F}(x_t^{(l)})$  和  $\hat{F}(y_t^{(k)})$  的乘积。

因此，我们有：

$$\hat{F}(x_t^{(l)}, y_t^{(k)}) = \hat{F}(x_t^{(l)}) \cdot \hat{F}(y_t^{(k)})$$

考虑到 RCRTE 中的每一项，我们有：

$$\begin{aligned} \int_0^\infty \hat{F}^\alpha(x_{t+1}, x_t^{(l)}) dx &= \int_0^\infty (\hat{F}(x_{t+1}, x_t^{(l)}))^\alpha dx, \\ \int_0^\infty \int_0^\infty \hat{F}^\alpha(x_{t+1}, x_t^{(l)}, y_t^{(k)}) dxdy &= \int_0^\infty \int_0^\infty (\hat{F}(x_{t+1}, x_t^{(l)}, y_t^{(k)}))^\alpha dxdy, \\ \int_0^\infty \hat{F}^\alpha(x_t^{(l)}) dx &= \int_0^\infty (\hat{F}(x_t^{(l)}))^\alpha dx, \\ \int_0^\infty \int_0^\infty \hat{F}^\alpha(x_t^{(l)}, y_t^{(k)}) dxdy &= \int_0^\infty \int_0^\infty (\hat{F}(x_t^{(l)}, y_t^{(k)}))^\alpha dxdy. \end{aligned}$$

现在，把边际 CDF 的表达式代入联合 CDF，我们就得到了：

$$(\hat{F}(x_{t+1}, x_t^{(l)}))^\alpha = (\hat{F}(x_{t+1}, x_t^{(l)}))^\alpha \cdot (\hat{F}(y_t^{(k)}))^\alpha,$$

以及

$$(\hat{F}(x_t^{(l)}))^\alpha = (\hat{F}(x_t^{(l)}))^\alpha \cdot (\hat{F}(y_t^{(k)}))^\alpha$$

由于指数  $\alpha$  不为零，我们可以将方程两边分别除以公因子  $(\hat{F}(x_{t+1}, x_t^{(l)}))^\alpha$  和  $(\hat{F}(x_t^{(l)}))^\alpha$ ，从而得到：

$$(\hat{F}(y_t^{(k)}))^\alpha = 1$$

由于指数  $\alpha$  不为零，这些等式成立的唯一条件是  $(\hat{F}(y_t^{(k)}))^{\alpha} = 1$ ，这意味着  $\hat{F}(y_t^{(k)}) = 1$ 。这意味着随机变量 Y 是一个概率为 1 的常数。因此，如果 X 和 Y 的分布处处相同，则有  $\text{RCRTE}(Y \rightarrow X) = 0$ 。

证明完毕。

**命题 7.** 令  $X$  和  $Y$  是两个随机变量, 且  $l = k = t$ , 则有:

$$\begin{aligned} 1 - \left( \frac{\mathbb{E}(x_{t+1}, x_t^{(l)}) \cdot \mathbb{E}(x_{t+1}, x_t^{(l)}, y_t^{(k)})}{\mathbb{E}(x_t^{(l)}) \mathbb{E}(x_t^{(l)}, y_t^{(k)})} \right)^{\frac{1-\alpha}{\alpha}} &\leq RCRTE(Y \rightarrow X) \\ &\leq \left( \frac{\mathbb{E}(x_{t+1}, x_t^{(l)}) \cdot \mathbb{E}(x_{t+1}, x_t^{(l)}, y_t^{(k)})}{\mathbb{E}(x_t^{(l)}) \mathbb{E}(x_t^{(l)}, y_t^{(k)})} \right)^{\frac{\alpha}{1-\alpha}} - 1 \end{aligned}$$

证明.

根据定义,  $RCRTE(Y \rightarrow X)$  的公式可改写为:

$$\begin{aligned} RCRTE(Y \rightarrow X) &= RCRE(x_{t+1}, x_t^{(l)}) - RCRE(x_t^{(l)}) + RCRE(x_{t+1}, x_t^{(l)}, y_t^{(k)}) - RCRE(x_t^{(l)}, y_t^{(k)}) \\ &= \frac{1}{1-\alpha} \log \frac{\int_0^\infty \hat{F}^\alpha(x_{t+1}, x_t^{(l)}) dx \int_0^\infty \int_0^\infty \hat{F}^\alpha(x_{t+1}, x_t^{(l)}, y_t^{(k)}) dxdy}{\int_0^\infty \hat{F}^\alpha(x_t^{(l)}) dx \int_0^\infty \int_0^\infty \hat{F}^\alpha(x_t^{(l)}, y_t^{(k)}) dxdy} \\ &= \log \frac{(\mathbb{E}^\alpha(x_{t+1}, x_t^{(l)}) \cdot \mathbb{E}^\alpha(x_{t+1}, x_t^{(l)}, y_t^{(k)}))^{1-\alpha}}{(\mathbb{E}^\alpha(x_t^{(l)}) \mathbb{E}^\alpha(x_t^{(l)}, y_t^{(k)}))^{1-\alpha}} \\ &= \log \frac{\mathbb{E}^{\frac{\alpha}{1-\alpha}}(x_{t+1}, x_t^{(l)}) \cdot \mathbb{E}^{\frac{\alpha}{1-\alpha}}(x_{t+1}, x_t^{(l)}, y_t^{(k)})}{\mathbb{E}^{\frac{\alpha}{1-\alpha}}(x_t^{(l)}) \mathbb{E}^{\frac{\alpha}{1-\alpha}}(x_t^{(l)}, y_t^{(k)})} \end{aligned}$$

由于当  $\alpha \geq 0, \alpha \neq 1$ , and  $x \geq 0$  时, 有  $\log^\alpha(x) \leq x - 1$ , 那么我们有:

$$\begin{aligned} RCRTE(Y \rightarrow X) &= \log \frac{\mathbb{E}^{\frac{\alpha}{1-\alpha}}(x_{t+1}, x_t^{(l)}) \cdot \mathbb{E}^{\frac{\alpha}{1-\alpha}}(x_{t+1}, x_t^{(l)}, y_t^{(k)})}{\mathbb{E}^{\frac{\alpha}{1-\alpha}}(x_t^{(l)}) \mathbb{E}^{\frac{\alpha}{1-\alpha}}(x_t^{(l)}, y_t^{(k)})} \\ &\leq \frac{\mathbb{E}^{\frac{\alpha}{1-\alpha}}(x_{t+1}, x_t) \cdot \mathbb{E}^{\frac{\alpha}{1-\alpha}}(x_{t+1}, x_t, y_t) - \mathbb{E}^{\frac{\alpha}{1-\alpha}}(x_t) \cdot \mathbb{E}^{\frac{\alpha}{1-\alpha}}(x_t, y_t)}{\mathbb{E}^{\frac{\alpha}{1-\alpha}}(x_t) \cdot \mathbb{E}^{\frac{\alpha}{1-\alpha}}(x_t, y_t)} \\ &= \left( \frac{\mathbb{E}(x_{t+1}, x_t^{(l)}) \cdot \mathbb{E}(x_{t+1}, x_t^{(l)}, y_t^{(k)})}{\mathbb{E}(x_t^{(l)}) \mathbb{E}(x_t^{(l)}, y_t^{(k)})} \right)^{\frac{\alpha}{1-\alpha}} - 1 \end{aligned}$$

不等式右侧得证。

又由于当  $\alpha \geq 0, \alpha \neq 1$ , 且  $x \geq 0$  时, 有  $\log_\alpha(x) \geq 1 - \frac{1}{x}$ , 那么我们有:

$$\begin{aligned} RCRTE(Y \rightarrow X) &= \log \frac{\mathbb{E}^{\frac{\alpha}{1-\alpha}}(x_{t+1}, x_t^{(l)}) \cdot \mathbb{E}^{\frac{\alpha}{1-\alpha}}(x_{t+1}, x_t^{(l)}, y_t^{(k)})}{\mathbb{E}^{\frac{\alpha}{1-\alpha}}(x_t^{(l)}) \mathbb{E}^{\frac{\alpha}{1-\alpha}}(x_t^{(l)}, y_t^{(k)})} \\ &\leq 1 - \left( \frac{\mathbb{E}(x_{t+1}, x_t^{(l)}) \cdot \mathbb{E}(x_{t+1}, x_t^{(l)}, y_t^{(k)})}{\mathbb{E}(x_t^{(l)}) \mathbb{E}(x_t^{(l)}, y_t^{(k)})} \right)^{\frac{1-\alpha}{\alpha}} \end{aligned}$$

不等式左侧得证。

证明完毕。

## 2.4 仿真数据实例

为了在合成数据上验证本文提出的两种时间序列度量方法的有效性，我们选择了两个系统进行实验模拟，即多元自回归（Vector Autoregressive, VAR）模型和Hénon映射。我们首先对这两个仿真系统进行简单介绍。

### 2.4.1 VAR 模型

自回归模型（Autoregressive, AR）是一种时间序列模型，根据变量以前的值来预测其当前值。VAR模型是一种用于分析多个时间序列变量之间关系的统计模型。它将AR模型的概念从单一变量扩展到变量系统。它为宏观经济预测、金融分析和动态过程建模等复杂系统的分析提供了一个灵活的框架。多变量 VAR 模型的一般形式可写成：

$$\begin{cases} X_t = C + \sum(A_i * X_{t-i}) + \sum(B_i * Y_{t-i}) + E_t \\ Y_t = D + \sum(C_i * X_{t-i}) + \sum(D_i * Y_{t-i}) + F_t, \end{cases}$$

其中， $X_t$  和  $Y_t$  是两个变量在时间  $t$  的值， $C$  和  $D$  是常数项， $A_i$ 、 $B_i$ 、 $C_i$  和  $D_i$  是捕捉变量滞后值之间关系的系数， $X_{t-i}$  和  $Y_{t-i}$  是变量在滞后期  $i$  的滞后值， $E_t$  和  $F_t$  是时间  $t$  的残差项。在本章设计的实验中，使用了两个二阶 VAR 模型，分别称为  $\text{VAR}_1$  和  $\text{VAR}_2$ 。具体参数如下：在  $\text{VAR}_1$  中，系数矩阵为  $A_1, B_1, A_2, B_2 = [0.8, 0.2, 0.4, 0.6]$ ， $C_1, D_1, C_2, D_2 = [0.3, 0.7, 0.1, 0.9]$ ； $x_0 = 0.1, y_0 = 0.1$ 。在  $\text{VAR}_2$  中，系数矩阵为： $A_1, B_1, A_2, B_2 = [1.5, 2.3, 3.7, 1.1]$ ， $C_1, D_1, C_2, D_2 = [2.6, 0.3, 2.5, 0.3]$ ； $x_0 = 0.5$ ， $y_0 = 0.5$ 。

### 2.4.2 Hénon映射

Hénon 映射是一个表现出混沌行为的多维离散动力系统。它是分析混沌系统动力学的一种有效模型，在物理学、生物学和计算机科学等多个领域都有应用。以二维 Hénon 映射为例，其定义如下式所示：

$$\begin{cases} x_{n+1} = 1 - a \cdot x_n^2 + y_n \\ y_{n+1} = b \cdot x_n, \end{cases}$$

其中  $x_t$  和  $t_t$  是时间  $t$  的状态变量,  $a$  和  $b$  是控制映射行为的参数。在下面的实验中, 我们同样设计了两组 Hénon 映射, 即  $\text{map}_1$  和  $\text{map}_2$ 。在  $\text{map}_1$  中, 参数设置为  $a = 1.4, b = 0.3$ , 初始值为  $x_0 = 0.1, y_0 = 0.1$ 。在  $\text{map}_2$  中, 参数设置为:  $a = 0.9, b = 0.6$ , 初始值为:  $x_0 = 0.5, y_0 = 0.5$ 。

### 2.4.3 关于 RCRKL 的仿真实验结果

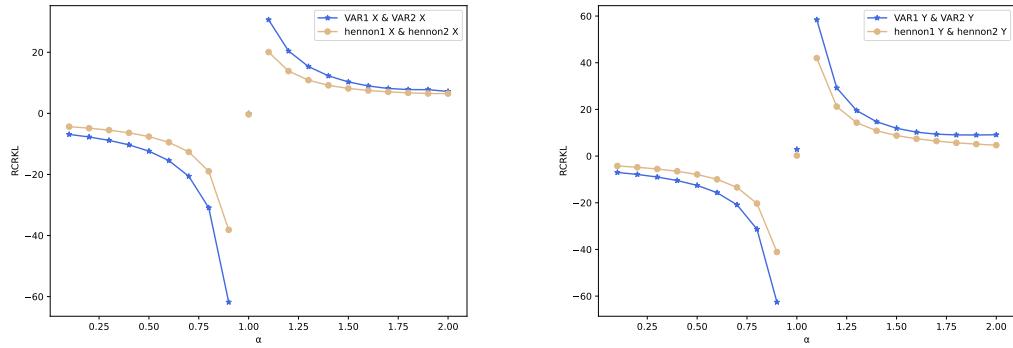


图 2.1: 参数  $\alpha$  对 RCRKL 的影响, 图(a)为序列X之间的 RCRKL, 图 (b) 为序列Y之间的 RCRKL

为了验证 RCRKL 在测量两个时间序列差异方面的有效性。我们设计了一些合成数据对比实验。第一个实验是验证 RCRKL 在测量由不同系数的系统产生的序列之间的差异时的有效性。结果如图2.1所示。图2.1.a的曲线为 $\text{RCRKL}(X_{\text{VAR1}}, X_{\text{VAR2}})$ 和 $\text{RCRKL}(X_{\text{map1}}, X_{\text{map2}})$ 。图2.1.b的曲线为 $\text{RCRKL}(Y_{\text{VAR1}}, Y_{\text{VAR2}})$ 和 $\text{RCRKL}(Y_{\text{map1}}, Y_{\text{map2}})$ 。正如我们在命题1中所证明的, 当  $\alpha \geq 1$  时, 无论时序列X还是Y, 各系统之间的RCRKL都是正值, 相反, 当  $0 \leq \alpha \leq 1$  时, 各系统之间的RCRKL则都是负值。此外, 根据KL散度的理论, 当 Rényi KL 散度为负时, 意味着被比较的两个概率分布有重叠的支持, 在结构上表现出相似性; 当 Rényi KL 散度为正时, 表明被比较的两个概率分布是不同的 [?]. 而根据 Rényi 熵理论 [?]. 当  $\alpha > 1$  时, Rényi 熵更关注序列中的低频事件, 而  $\alpha \in [0, 1]$  时, Rényi 熵更关注高频事件。因此, 我们可以得出结论, 对于 VAR 模型的趋势项 (低频事件) 来说, VAR1 和 VAR2 是相似的。H' enon 映射也同样存在这个现象。然而, 在两个不同系数的 VAR 模型序列 (或两个 H' enon 映射) 之间, 波动项 (高频事件) 的差异更大。相反, 当  $\alpha = 1$  时, RCRKL 退化为传统的累积残差散度。从图中可以看出, 不管是时间序列  $X$  还是序列  $Y$ , 两个系统的 CRKL 差异都不大。而且正负性也不确定。同时, 我们注意到 RCRKL 是随着  $\alpha$  的增加而单调递减的, 特别是

当其取值在  $(0.7, 1)$  和  $(1, 1.3)$  时, RCRKL 的变化是显著的。因此, 当使用 RCRKL 的任务对事件的相似性或差异性更为敏感时,  $\alpha$  的取值范围应为  $(0.7, 1)$  或  $(1, 1.3)$ 。

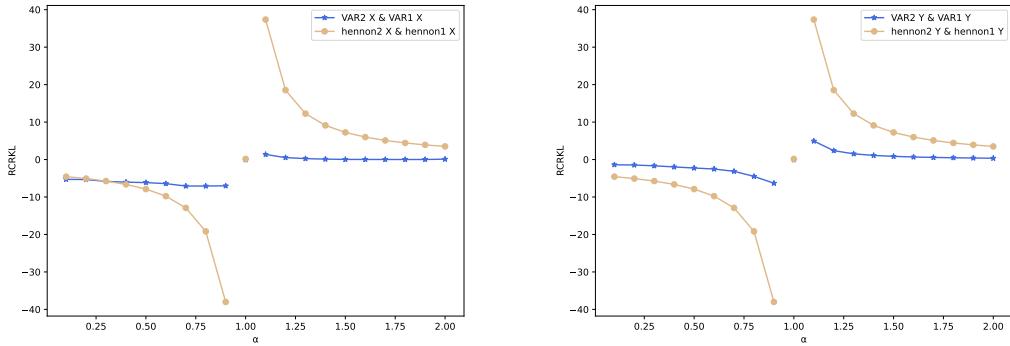


图 2.2: 参数  $\alpha$  对 RCRKL 的影响, 图(a)为从 VAR2 到 VAR1 的 RCRKL, 图 (b) 为从 Hénon map2 到 Hénon map1 的 RCRKL

由于 CRKL 是不对称的, 这意味着  $RCRKL(x, y) \neq RCRKL(y, x)$ 。为了进一步说明 CRKL 的有效性, 我们还比较了从 VAR2 到 VAR1 以及从 textH'enon2 到 textH'enon1 的 RCRKL。结果如图4.3 所示, 其中序列 X 的 RCRKL 如图4.3.a 所示, 系列 Y 的 RCRKL 如图4.3.b 所示。与图 1 中的 RCRKL 结果相似, 无论时序 X 还是 Y, 从 VAR2 到 VAR1 以及从 textH'enon2 到 textH'enon1 的 RCRKL 都在  $\alpha \geq 1$  时为负值, 并且随着  $\alpha$  的增大单调减小。同时, 我们注意到在这种情况下, 当  $0 < \alpha < 1$  时, RCRKL 的正负性是不确定的。从图 4.3.b 中 VAR2 到 VAR1 的 RCRKL 可以看出, 当  $\alpha = 0.1$  时, RCRKL 的值为正, 当  $\alpha \geq 0.1$  时, RCRKL 的值变为负, 这时, RCRKL 是与序列的分布有关的。

#### 2.4.4 关于 RCRTE 的仿真实验结果

RCRTE 是在 Rényi 熵和累积残差熵的基础上对传递熵的扩展。它测量系统中变量序列之间的因果关系。为了验证本章提出的 RCRTE 的有效性, 本节在合成数据上也进行了一些实验, 不同于上一节有关 RCRKL 上的实验, 本节的实验均作用在系统内部, 即序列 X 和 Y 之间。在第一个实验中, 我们验证  $\alpha$  对 RCRTE 的影响, 结果如图2.5所示。在 VAR1 模型中, 从系列 X 到 Y 的 RCRTE 如图2.5.a 所示; 在 VAR2 模型中, 从系列 X 到 Y 的 RCRTE 如图 2.5.b 所示。结果表明, 当  $\alpha \in (\infty, 1)$  和  $(1, \infty)$  时, RCRTE 分别随  $\alpha$  的增加而单调递减。同时, 正如命题 6 所证明的那样,  $RCRKL \neq 0$ , 因为

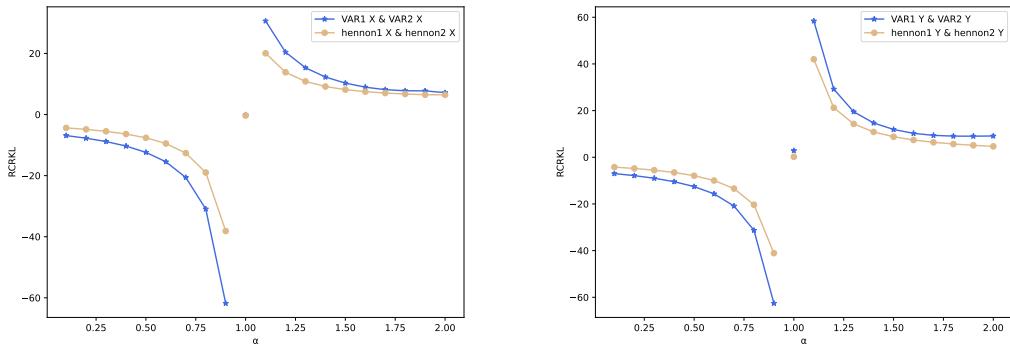


图 2.3: 参数  $\alpha$  对 VAR 模型和Hénon映射的 RCRTE 值的影响。

$X$  和  $Y$  的分布并非处处相等。这一特性可以帮助我们在因果检测中剔除没有时间依赖性或信息流的变量。

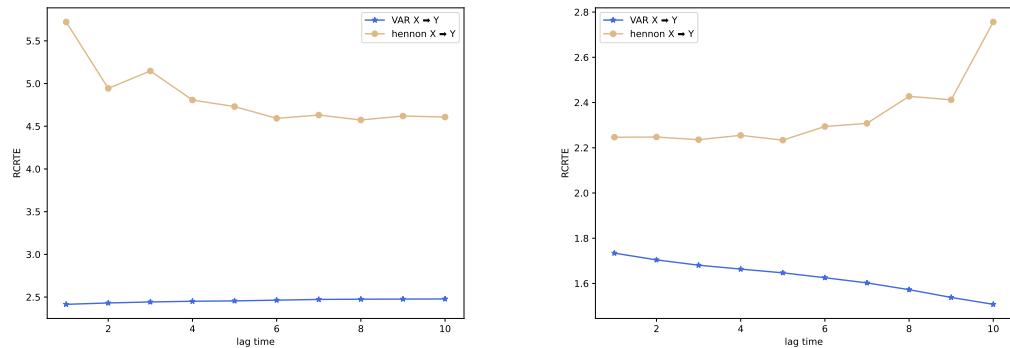


图 2.4: 滞后时间  $l$  和  $k$  对 VAR 模型和Hénon映射的 RCRTE 值的影响。

马尔可夫阶数  $l$  和  $k$  是另一个会影响 RCRTE 值的参数。它决定了用于计算  $t$  时刻 RCRTE 的  $X$  序列和  $Y$  序列历史信息的长度。图 2.4直观地展示了马尔可夫阶数对RCRTE值的影响。为方便起见, 将  $l$  和  $k$  设置为相同值, 范围为:  $l = k \in [1, 10]$ 。总的来说, 无论  $\alpha$  是否大于 1, 两个系统的  $\text{RCRTE}(X \rightarrow Y)$  都是正值。这表明, 无论是长尾事件还是普通事件,  $X$ 都是 $Y$ 的因果源。此外, 马尔可夫阶数对 VAR 模型和 hénnon映射的影响也不尽相同。具体来说, VAR 模型的 RCRTE 值在两个方向上都是稳定的, 而 Hénon 映射的 RCRTE 值在而当  $l \in [3, 10]$  时两个方向上都是相对稳定的, 变化幅度较小, 而当  $l \in [1, 2]$  时, RCRTE( $X \rightarrow Y$ ) 值有快速下降的趋势, 当  $l > 9$  时, RCRTE( $Y \rightarrow X$ ) 值快速上升。不同的结果意味着, 在处理不同系统时, 发生统计意义上的因果相关性的历史序列长度是不同的, 而且需要谨慎选择马尔可夫

阶。

## 2.5 实例应用

本章的实例数据来自RIOHTrack数据集，有关该数据集的详细介绍在第一章中已经给出，本章不在赘述。这里介绍一下本章所使用数据，车辙及其影响因素被选用来验证 RCRKL 和 RCRTE 的有效性，影响因素包括温度、荷载量和中心点弯沉。路面分为19组结构，每组结构144个数据点。

RIOHTrack 车辙实验分为两部分：1. 基于 RCRKL 的车辙时间序列聚类；2. 基于 RCRTE 的车辙因果分析。实验的流程设计如下：在第一部分中，基于 RCRKL 对 19 个路面车辙的时间序列进行分层聚类，将其聚成相似的簇。对于每个聚类，在第二部分中通过 RCRTE 检测车辙与其影响因素之间的因果关系。

### 2.5.1 基于 RCRKL 的车辙时间序列聚类结果及讨论

本节提出了一种基于 RCRKL 的新型 k-means 聚类方法，其中用 RCRKL 代替了欧式距离度量。在第二节中，我们证明了当  $\alpha > 1$  时，RCRKL 是非负的，并且三角形不等式在 RCRKL 上成立，然而，由于其不对称性，RCRKL 不能直接被用作距离度量，即  $RCRKL(X, Y) \neq RCRKL(Y, X)$ 。为了解决这个问题，本文设计了一种基于 RCRKL 的对称距离度量：

$$D_{RCRKL}(X, Y) = \frac{RCRKL(X, Y) + RCRKL(Y, X)}{2}.$$

显然， $D_{RCRKL}(X, Y)$  是对称的，即  $D_{RCRKL}(X, Y) = D_{RCRKL}(Y, X)$ 。因此，基于 RCRKL 的距离度量已被完整定义，在此基础上，我们可以使用已有的聚类算法对时间序列进行聚类。作为一种典型的无监督聚类算法，AP 聚类在本节中被使用，其设计算法为 1：

算法1的聚类可视化结果如图??所示。聚类结果如下：STR1-STR5 的车辙在第 1 簇，STR6-STR9 的车辙在第 2 簇，STR10-STR11 的车辙在第 3 簇，STR12-STR14 的车辙在第 4 簇，其余车辙在第 5 簇。由于 AP 算法是一种无监督算法，因此不存在聚类结果的绝对正确结果。为了便于比较，我们参考了一些工程经验分类结果。根据 RIOHTrack 之前的研究，路面结构设计可以根据沥青层的组合和厚度分为 5 组，如表4.1所示。与基于结构的分类相比，基于 RCRKL 的 AP 算法的结果在聚类数量上是

---

**算法 1 基于RCRKL的AP聚类**

---

```
1: 输入:
2: - 多元时间序列  $X_n$ 
3: - 最大迭代次数  $max\_iterations$ 
4: 初始化:
5: - 将可用矩阵  $A$  初始化为零矩阵
6: - 将责任矩阵  $R$  初始化为零矩阵
7: 信息传递:
8: for each iteration from 1 to  $max\_iterations$  do
9:   for each time series  $i$  from  $X_1$  to  $X_n$  do
10:    for each time series  $j$  from  $X_1$  to  $X_n$  except  $i$  do
11:      通过下列公式计算责任矩阵  $R(i, j)$ :
12:      
$$R(i, j) = (RCRKL(x_i, x_j) + RCRKL(x_j, x_i)) / 2 - \max(A(i, :)) + (RCRKL(x_i, x_j) + RCRKL(x_j, x_i)) / 2 \text{ except for } R(i, j)$$

13:    end for
14:  end for
15:  for each time series  $i$  from 1 to  $X_n$  do
16:    for each time series  $j$  from 1 to  $X_n$  do
17:      通过下列公式计算可用矩阵  $A(i, j)$  :
18:      
$$A(i, j) = \min(0, R(j, j) + \sum(\max(0, R(k, j))) \text{ except for } i \text{ and } j))$$

19:    end for
20:  end for
21: end for
22: 范例点和簇的确定:
23: - 通过查找具有正净责任的数据点来确定范例点: ( $\sum R(i, i)$  and  $A(i, i)$ )
24: - 将每个数据点分配到与其最近的示例相关联的簇中
25: 输出:
26: - 范例点及其相应的簇
```

---

相同的，都是 5 个。但是，基于所提聚类算法的每个聚类中的元素与基于经验的路面结构类别并不相同。主要区别在于，半刚性基层结构中的路面被分为 STR1-STR3 和 STR6-STR9 两组，而倒置结构中的 STR10 和 STR12 则被分为不同的簇。针对这一现象，一些基于复杂网络分析的研究表明，作为粘弹性半刚性体，路面的变形与沥青的厚度有关。这一解释反映在聚类结果中，即具有相似沥青层厚度的路面更有可能聚类在同一个聚类中 [?].

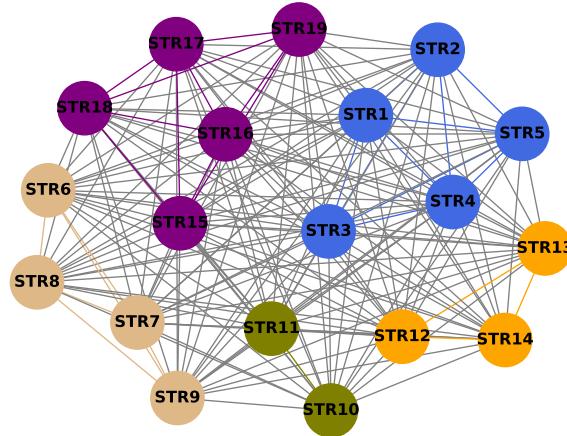


图 2.5: 基于 RCRKL 的 AP 聚类算法得出的 RIOHTrack 车辙时间序列聚类结果

表 2.1: 基于经验设计的 RIOHTrack 路面结构分类。

簇编号	路面类别	路面编号
1	semi-rigid base structure	STR1-STR3, STR6-STR9
2	rigid base structure	STR4, STR5
3	inverted structure	STR10, STR12
4	flexible base structure	STR11, STR13-STR17
5	full thickness structure	STR18, STR19

此外，为了验证基于 RCRKL 的聚类算法的性能，本节还对基于不同距离度量的聚类算法进行了比较。我们将重点限制在应用于时间序列的方法上，即皮尔逊相关性（Pearson Correlation, PC）[?]、动态时间扭曲（Dynamic Time Warping, DTW）[?]和马哈拉诺比斯距离（Mahalanobis Distance, MD）[?]。性能评估指标是一些内在的衡量标准：剪影系数（Silhouette Coefficient, SC）、戴维斯-博尔丁指数（Davis Boulding Index, DBI）和邓恩指数（Dunn Index, DI）。SC 计算的是所有时间序列点的平均剪影系数，其中点的剪影系数被定义为与自身聚类中点的平均距离和与最近相邻聚类中点的平均距离之差。SC 越高，表示聚类越清晰。DB 指数评估聚类之间的紧

凑性和分离度。它被定义为每个聚类与其最相似聚类之间的平均相似度。DBI 指数越低，表示聚类越清晰、分离度越高。DI 衡量聚类的紧凑程度（聚类内距离）和聚类间的分离程度（聚类间距离）。DI 值越高，表示聚类质量越好。本次验证中使用的所有距离指标均为 RCRKL。

总体而言，基于 RCRKL 的聚类算法的性能优于其他方法。特别是在 SC 测量中，本文提出的聚类算法表现较好， $SC(RCRKL) = 0.43$ ，表明聚类结果合理，聚类之间的分离度适中。在 DB 的测量中，只有  $DB(RCRKL)=0.37$  和  $DB(DTW)=0.27$  小于 0.5。结果表明，聚类结果相对较好，聚类分离度较高。最后，基于 RCRKL 的聚类算法的 DI 值为 0.5，这表明聚类结果具有良好的分离性。

表 2.2: 基于不同距离度量的 AP 算法的平均内在测量值比较。

距离度量	SC	DBI	DI
PC	-0.03	1.12	0.18
DTW	0.27	0.49	0.35
MD	0.15	0.63	0.22
RCRKL	0.43	0.37	0.50

### 2.5.2 基于 RCRTE 的车辙时间序列因果检测结果及讨论

最后，我们分析了 5 个群组内车辙影响因素的 RCRTE 值。对于参数  $\alpha$ ，我们选择了两个值，分别为  $\alpha = 0.5$  和  $\alpha = 1.5$ ，分别对应于时间序列中的低概率事件和高概率事件。滞后时间  $l = k = 5$  的结果如图2.6和图2.7所示。很明显，从挠度盆深度到车辙的RCRTE在两个 $\alpha$ 值上均为正，这说明路面刚度是车辙的重要成因。同时，大多数聚类上从路面温度到车辙的 RCRTE 在  $\alpha \leq 1$  时为负值，而在聚类 1 上，当  $\alpha \geq 1$  时为正值。这表明对于长尾事件，如车辙的快速加深，温度变化是其因果来源，而对于普通事件，如车辙的平稳发展，温度不应被视为因果来源。具体来说，对于属于第 1 组和第 2 组的半刚性基层结构路面和刚性基层结构路面，在长尾事件中，温度是车辙产生的原因，而在普通事件中，交通荷载是车辙产生的原因。对于属于第 3 组和第 4 组的厚沥青混凝土基层结构路面，交通荷载也是长尾事件车辙产生的原因。而对于属于第 5 组的全深度沥青混凝土结构路面，温度是长尾事件和普通事件车辙产生的原因。同时，我们还注意到，在某些群组中，从温度和交通荷载到挠度盆深度的 RCRTE 也是正值，这意味着这两个因素在某些情况下会影响路面刚度。

与经典 TE 相比，分数 RCRTE 可以在分数阶  $\alpha$  的范围内更全面地检测影响因素与车辙之间的因果关系。

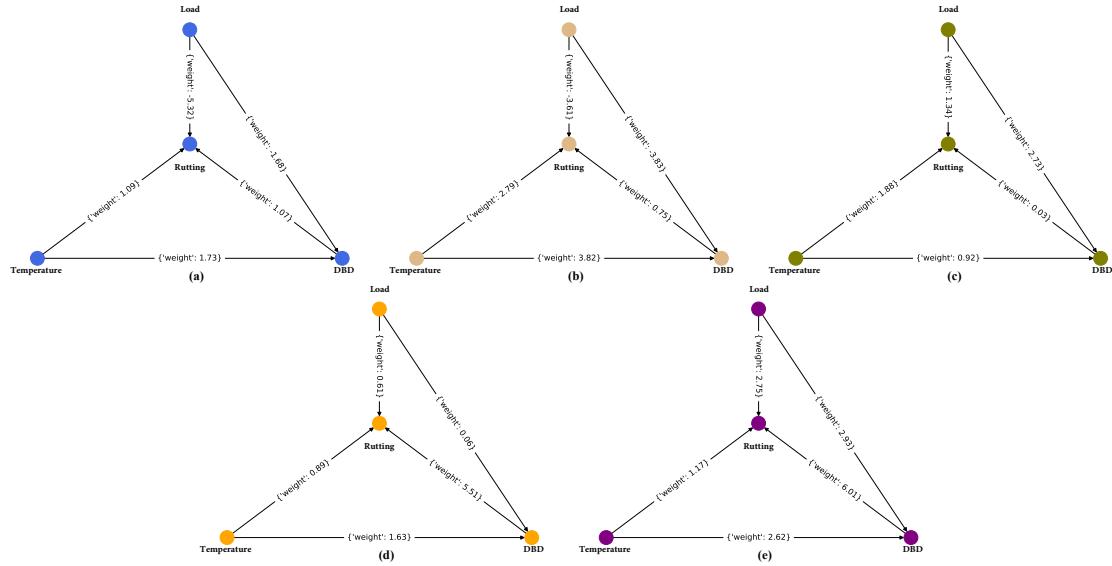


图 2.6: 5 个聚类上从影响因素到车辙的 RCRTE 值图表可视化, 其中参数为  $\alpha = 0.5$ ,  $l = k = 2$ 。

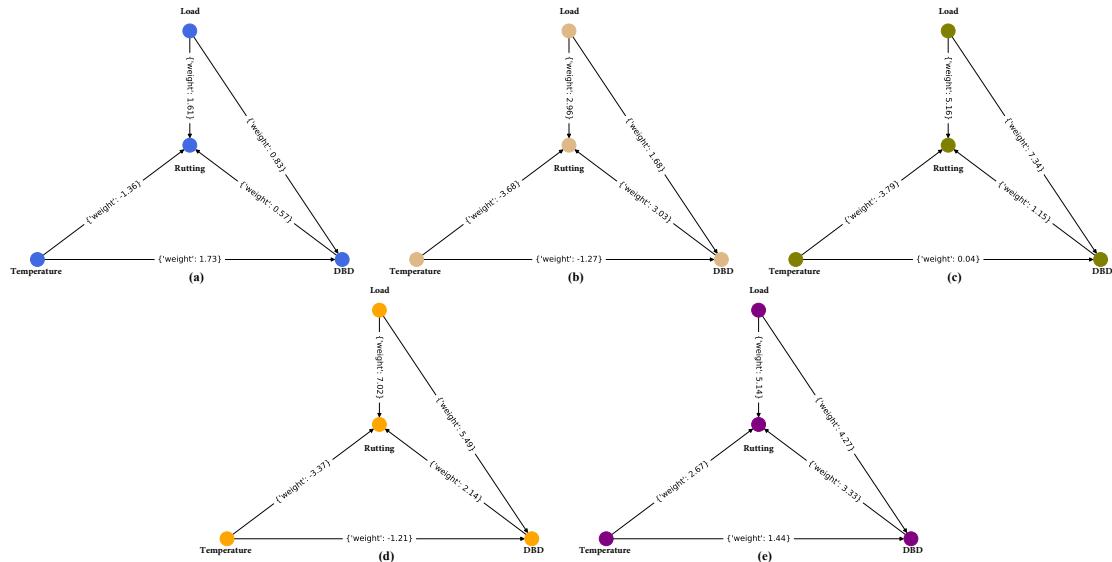


图 2.7: 5 个聚类上从影响因素到车辙的 RCRTE 值图表可视化, 其中参数为  $\alpha = 1.5$ , 滞后时间为  $l = k = 2$ 。

## 2.6 小结

本文引入了两种新的分数阶时间序列度量—Rényi累积残差KL散度（Rényi cumulative residual Kullback-Leibler, RCRKL）和Rényi累积残差转移熵（Rényi cumulative residual transfer entropy, RCRTE）。研究并证明了 RCRKL 和 RCRTE 的一些数学性质。我们证明了 RCRKL 作为距离度量的一些性质。并研究了 RCRTE 的零点和边界。然后，我们提供了一些仿真数据的例子来研究参数变化对度量值的影响。最后，将两个分数阶时间序列度量应用于路面车辙时间序列，设计了基于 RCRKL 的AP聚类，并通过 RCETE 检测了影响因素与车辙之间的因果关系。总的来说，本章所提出的 RCRKL 和 RCRTE 兼具了 CRE 和 RE 的优点，同时对时间序列的度量提供了更加全面的见解和分析。在未来的工作中，我们将进一步把这两个度量应用到其他领域的时间序列数据中以验证它们的泛化性，同时，基于诸如Tsallis熵这样的分数阶熵，将这两个度量进一步扩展到更多形式的分数阶度量以增加它们的全面性。本章结果已发表在国际刊物 Neural Networks 上。具体详见作者发表论文的清单。

# 第三章 基于自适应小波分解和多元核密度估计的多尺度因果推断

## 3.1 引言

因果检测是时间序列分析的一项基本任务，旨在识别变量之间的顺序关系，强调变量之间的方向性影响 [128]。与时间序列相关分析中的对称关系不同，因果关系被定义为不对称关系 [129]。传递熵(Transfer entropy, TE)是一种显著的基于信息熵的时间序列因果检测方法，用于度量两个时间序列之间的因果关系 [130]。一般来说，从序列 $X$ 到序列 $Y$ 的TE表示在掌握了 $X$ 和 $Y$  [131]的历史信息后，下一个时间步预测未来值 $Y$ 的不确定性降低。Rényi传递熵 (Rényi RTE) [132]作为TE的扩展，通过参数 $\alpha$ 既能考虑罕见事件和频繁事件，又能考虑序列中的特殊分布部分。近年来，基于RTE的多种因果检测方法在过去几十年中被提出，因为它具有分析具有不对称、时滞、间接相关甚至非线性系统的机制分析的时间序列的优势 [133]。例如，Mi等人结合振动模态分解和RTE，基于脑电图(EEG)信号探索脑电极间多层次连接状态 [134]。基于RTE和有效传递熵，He等。提出了一种有效的rényi传递熵来分析股票时间序列之间的信息流，可以减弱噪声对股票时间序列的影响 [135]。Armah等人提出了一种基于rte的方法来衡量发达国家房地产与政策不确定性之间的因果关系 [136]。

尽管有这些成功的先例，但在将RTE引入特定领域时仍存在一些挑战。首先，作为一种基于信息熵的模型，在计算RTE时不可避免地要计算概率密度分布，在实际中需要对其进行估计。概率密度分布估计的准确性将直接决定TE在因果检测中的有效性。一类方法是直接利用现有的观测数据，基于不同类型的估计器估计联合概率密度分布和边际概率密度分布 [137]。举几个例子，Kraskov等 [138]通过引入 $k$ -最近邻来解决概率分布的局部近似，提出了互信息的估计。Faes等 [139]提出了一种基于修正条件熵估计序列间信息流的非线性格兰杰因果关系方法。上述方法虽然可以有效地估计概率密度函数，但往往需要预先定义一些超参数，并假设整个原始数据的概率密度分布。因此，RTE构造的另一个分支是基于内核密度估计(KDE) [140]。例如，Giraldo等 [141]提出了一种基于无限可分核定义的算子在再现核希尔伯特空间中估计概率的非参数方法。在Giraldo等人 [141]的基础上，De La Pava等人 [142]提出了一种基于RTE估计的脑电信号因果检测的数据驱动测量方法，该方法提供了RTE的线性组合

表达式。随后, Zhou等人通过显著性分析 [143]将KDE方法扩展到基于矩阵的RTE估计的多变量情况,既可以检测间接因果关系,也可以检测共同因果关系。因此,基于kde的方法可以通过估计RTE值而不是估计概率分布来获得更好的因果检测结果。然而,上述文献并未提供一维核方法与多维核方法之间的关系,这促使我们在单变量和多变量情况下寻求基于kde的RTE估计方法的测度统一准则。此外,现实世界中的复杂系统通常是一个多层次系统,传统的RTE无法捕获多尺度情况下的因果关系。为了解决这一问题,一些学者将小波变换(WT)作为分析非平稳信号的经典时间序列分解方法引入到因果检测中。Materassi等人基于TE和连续小波变换(CWT) [144]研究了地球磁层尖端磁场动力学各变量之间不同频带的因果关系。Storhas等 [145]基于CWT和符号传递熵检测了原油和成品油返回动态的多尺度因果关系。此外,离散小波变换(DWT)也被用于多尺度因果检测。例如,提出了一种基于DWT和TE的因果检测方法来测量耦合系统 [146]中变量之间的信息传递。Guo等 [147]利用二进平稳小波变换将脑电图和肌电图信号分解为5个频段,应用TE检测不同时间尺度下的皮质-肌肉耦合。这些工作证明了将基于小波变换的时间序列分解引入因果检测的可行性和必要性。然而,对于离散小波,最优滤波器是非常理想的,这也会影响最终因果检测的性能。近年来,一些学者基于机器学习求解最优小波系数。Saadatmorad等人( [148])引入卷积神经网络设计最优小波系数。Lu等人 [149]提出了一种基于小波神经网络和切换粒子群优化的混合方法,将小波系数作为待训练模型的参数。同样, Zhao等人提出了一种基于深度残差网络的方法来融合不同频段的多个小波系数,每个小波系数可以通过模型训练自适应调整 [150]。需要指出的是,上述基于机器学习的方法都需要预先定义一些超参数。因此,基于机器学习的方法在克服预定义超参数问题上还有一定的改进空间,这激发了我们去填补这一空白。

针对上述问题,本文提出了一种基于核密度估计的复杂系统因果检测新方法。该方法对各变量在不同频带上的RTE进行估计,分为两部分:基于自适应离散小波变换(Adaptive Discrete Wavelet Transform, ADWT) 的时间序列分解部分和基于多元核密度估计(Multivariate Kernel Density Estimation, MKDE) 的RTE因果关系检测部分。ADWT模块是一个自编码器,其中编码器和解码器由同一对正交小波组成,分别对应于小波分解和重构过程。在基于adwt的时间序列分解中,系统的多变量时间序列在自编码器框架下,通过最小化原始时间序列与重构时间序列之间的均方误差(MSE)产生最优逼近系数和详细系数,在多层次中将系统的多变量时间序列分解为高频分量和低频分量。每层的分量构成分量变量邻接矩阵,节点表示变量,边表示因果关系。这些

变量邻接矩阵被输入到因果网络生成部分来计算因果关系。在基于MKDE的因果网络生成中，每个变量之间的因果关系用RTE值表示，RTE值由MKDE模块估计。KDE的度量是基于最小化KDE中每个变量的概率密度函数与核函数之间的渐近平均积分误差(AMISE)。本文给出了单变量和多变量情况下基于AMISE的KDE统一评价准则。此外，我们还选择了单变量和多变量核情况下的最优带宽。为了验证该方法的有效性，在合成数据和实际数据上进行了实验。结果表明，本文提出的方法基于ADWT模块和MKED模块，可以对不同频段的传输信息进行量化。与其他估计方法的比较结果表明，所提出的方法可以更准确、更显著地估计变量之间的因果关系。

## 3.2 模型描述

新型RTE的估计框架如图3.1所示，它由两个模块组成：基于自适应离散小波变换（ADWT）的时间序列分解和基于多变量核密度估计（MKDE）的RTE因果关系网络生成。输入的 $X \in \mathbb{R}^{m \times t}$ 是 $m$ 组时间序列， $t$ 是每个时间序列的长度。 $X$ 首先被送入ADWT模块以计算近似系数 $H_D \in \mathbb{R}^{n \times \frac{t}{2}}$ 和详细系数 $G_D$ 中。通过最小化输入数据 $X$ 和重建数据 $\hat{X}$ 之间的均方误差（MSE），在 $\mathbb{R}^{n \times \frac{t}{2}}$ 中获得 $H_D$ 和 $G_D$ 。在基于ADWT的时间序列分解模块中，原始时间序列在最优小波系数的作用下被分解为不同层的高频成分 $X^h(n)$ 和低频成分 $X^l(n)$ 。在基于MKDE的因果关系网络模块中，系统内各变量的因果关系由变量相邻矩阵表示。每个变量的分解分量依次形成邻接矩阵，其中最后一层的低频分量作为趋势项，形成低频分量变量邻接矩阵 $G^l$ ，各层的高频分量作为波动项，形成高频分量变量邻接矩阵 $G^h$ 。这些邻接矩阵作为输入进入基于MKDE的因果网络部分的生成部分。邻接矩阵的值表示变量之间的因果关系，初始值为0。变量之间的因果关系将由RTE值呈现，在每一层中，MKDE模块通过最小化概率分布和最优带宽的核函数之间的渐近平均积分误差（AMISE）来估计。最后，每层的变量邻接矩阵 $G_n = (V, E)$ 是整个模型的输出。在详细介绍所提出的方法中的两个模块之前，我们先对离散小波变换和RTE进行简单的初步介绍。

### 3.2.1 预备知识

#### 3.2.1.1 离散小波分解

在现实的复杂系统中，观察结果总是以离散的形式存在，因此，通常采用DWT来捕捉时间和频率的定位。小波分解的完整框架是一个分层结构，每一层都有两个部分：时间序列分解和时间序列重构。如图中的时间序列分解部分。图3.2，DWT通过高通

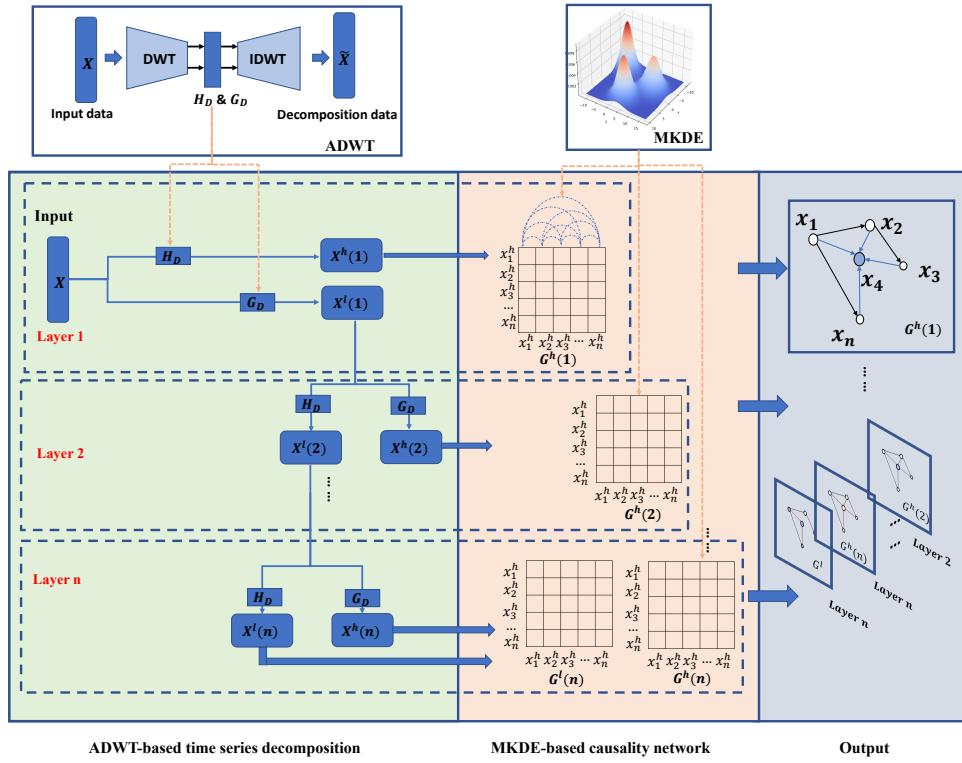


图 3.1: 新型RTE的估计框架。

滤波器 $h_i$ 和低通滤波器 $g_i$ 在第*i*层将时间序列 $x_i^l$ 分解为高频分量和低频分量。其中，时间序列分解过程可以通过以下方式给出：

$$\begin{cases} y_t^l(i+1) = H_D(M \cdot t) = \sum_{t=1}^t x_i^l(M \cdot T - t) \cdot h_i, \\ y_t^h(i+1) = G_D(M \cdot t) = \sum_{t=1}^t x_i^l(M \cdot T - t) \cdot g_i, \end{cases} \quad (3.2.1)$$

其中， $y_t^l(i+1)$ 和 $y_t^h(i+1)$ 是*i*+1层原始序列的*t*时间步长的低频分量和高频分量，对应于所谓的近似序列和详细序列 [151]，它们代表原始序列的趋势项和波动项。此外， $x_i^l(M \cdot T - t)$ 是第*i*层的子序列，是对上层输出的下采样。*M*为比例因子，*T*为滤波器的分解长度。 $H_D(t)/inh_i$ 是第*i*层的近似系数， $G_D(t)/ing_i$ 是第*i*层的详细系数。在时间序列重建部分，分解后的序列可以通过反离散小波变换进行重建：

$$y_i'^l(t) = \sum_{t=1}^t x_{i+1}^l(M \cdot T - t) \cdot h'_{i+1} + \sum_{t=1}^t x_{i+1}^h(M \cdot T - t) \cdot g'_{i+1}, \quad (3.2.2)$$

其中重建滤波器 $h'$ 和 $g'$ 是分解滤波器的镜像函数。因此，这对重建滤波器也是正交镜像滤波器。 $x(M \cdot T - t)$ 的作用是，在与相应的滤波器卷积之前，要重建的序列被上采样（每个元素之间插入零）。单级DWT的图示摘要见图3.2。该图说明了高通滤波器 $h$ 和低

通滤波器 $g$ 对于DWT是必要的，包括过滤、下采样和上采样的因素。同时，DWT的中间结果，低频分量和高频分量可以用来表示原始序列中不同频段的特征。

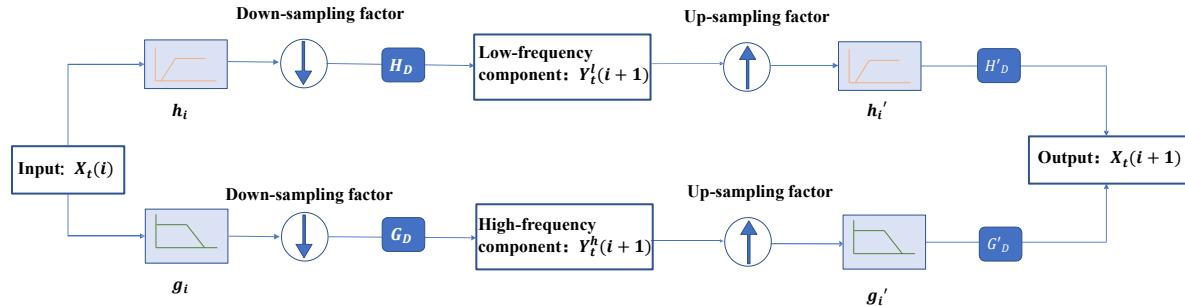


图 3.2: 基于单层DWT的时间序列分解和重建的图示

### 3.2.1.2 Rényi 传递熵

Rényi传递熵 (Rényi transfer entropy, RTE) 是对香农传递熵 (Shannon transfer entropy, STE) 在复杂分布下的扩展和泛化。因此，在介绍RTE之前，先简单介绍一下STE。STE是一种基于信息论和概率论的非参数测量，用于量化系统间的传递信息和因果关系。考虑到一个复杂的系统 $S$ ，它产生了两个离散的Markov过程 $X_t = x_1, x_2, \dots, x_t$ 和 $Y_t = y_1, y_2, \dots, y_t$ ，顺序为 $k$ 和 $l$ ，其中每个变量的概率分布为 $p_{x_t}$ 和 $p_{y_t}$ 。时间步数是离散的，单步为 $tau$ ， $t_n$ 为未来时间步数。从 $Y$ 到 $X$ 的STE,即  $T_{Y \rightarrow X}(k, l)$  的定义是：

$$\begin{aligned} T_{Y \rightarrow X}(k, l) &\equiv I(X_{t+1} : Y_t^{(l)} | X_t^{(k)}) \\ &= H(X_{t+1} : X_t^{(k)}) - H(X_{t+1} : (Y_t^{(l)}, X_t^{(k)})) \\ &= \sum_{x_t^{(k)} \in X_t^{(k)}, y_t^{(l)} \in Y_t^{(l)}} p(x_{t+1}, x_t^{(k)}, y_t^{(l)}) \log_2 \left( \frac{p(x_{t+1} | x_t^{(k)}, y_t^{(l)})}{p(x_{t+1} | x_t^{(k)})} \right), \end{aligned} \quad (3.2.3)$$

TE在理论上等同于从 $Y_t^l$ 到 $X_{t+1}$ 的条件互信息，以历史序列 $X_t^k$ 为条件，其中 $X_t^k = X_t, X_{t-1}, \dots, X_{t-k+1}$ 和 $Y_t^l = Y_t, Y_{t-1}, \dots, Y_{t-l+1}$ 为联合过程集合。香农传递熵(STE) (4.3.3) 是定量测量  $X_{n_{t+1}}$ 通过观察 $Y_t$ 到时间 $t_n$ 会减少多少不确定性的基本方法。考虑到香农熵形式 $H_s = -\sum p(x) \log(p(x))$ ，方程(4.3.3)也可以表示为熵形式：

$$T_{Y \rightarrow X}(k, l) = H_s(X_t^k, Y_t^l) - H_s(x_t, Y_t^l, X_t^k) + H_s(x_t, X_t^l) - H_s(x_t^l), \quad (3.2.4)$$

其中 $H_s(\cdot)$ 和 $H_s(\cdot, \cdot)$ 分别为边际熵和联合熵。不同于SE, Rényi熵 (RE) 是一个指数加权的不确定性度量的平均值, 在形式上为SE的单参数泛化。假设 $\alpha > 0$ , 具有概率分布 $\mathcal{P}$ 的变量 $X$ 的RE被定义为:

$$H_\alpha[\mathcal{P}] = \frac{1}{1-\alpha} \log_2 \sum_{x \in X} p^\alpha(x), \quad (3.2.5)$$

特别是, 当 $\alpha = 1$ 时, RE指的是SE。有关RE及其参数 $\alpha$ 的性质在上一章中有过介绍, 在本章中不再赘述。基于RE的引入, 以及上述STE的定义 (4.3.3), RTE可以被表达为:

$$T_{\alpha, Y \rightarrow X}^R(k, l) = \frac{1}{1-\alpha} \log_2 \frac{\sum \rho_\alpha(x_n^{(k)}) p^\alpha(x_{n+1}|x_n^{(k)})}{\sum \rho_\alpha(x_n^{(k)}, y_n^{(l)}) p^\alpha(x_{n+1}|x_n^{(k)}, y_n^{(l)})}, \quad (3.2.6)$$

其中 $\rho_\alpha(x_n^k) \equiv \frac{p^\alpha(x_n^k)}{\sum p^\alpha(x_n^k)}$ 是 $x_n^k$ 的护送分布(escort distribution),  $\rho_\alpha(x_n^k, y_n^l)$ 是 $x_n^k, y_n^l$ 的联合护送分布(joint escort distribution)。RTE可以通过调整 $\alpha$ 的值来关注概率密度函数的不同部分。特别是, 对于 $\alpha = 1$ , 可以证明RE指的是SE。对于 $0 < \alpha < 1$ , 像长尾事件这样的边际事件会被强调出来。对于 $\alpha > 1$ , 普遍事件将被强调。类似地, 将方程 (3.2.5) 代入方程 (3.2.6), RTE也可以被改写为熵的形式:

$$T_{\alpha, Y \rightarrow X}^R(k, l) = H_\alpha(X_t^k, Y_t^l) - H_\alpha(x_t, Y_t^l, X_t^k) + H_\alpha(x_t, X_t^l) - H_\alpha(x_t^l). \quad (3.2.7)$$

### 3.2.2 基于自编码器的自适应离散小波分解

在预备知识中我们介绍到, DWT中的高频和低频成分是原始序列与小波系数的乘积。因此, 基于DWT的时间序列分解的一个关键步骤是找到合适的高通滤波系数 $H_D$ 和低通滤波系数 $G_D$ 来生成子序列。本章提出了一种基于编码器-解码器的自适应离散小波变换 (ADWT) 来求解最优系数, 它可以在没有任何超参数的情况下求解最优小波系数。本章所提出的自适应离散小波分解 (Adaptive discrete wavelet decomposition, ADWT) 本质上是一个自编码器, 其中的编码器和解码器, 分别相当于离散小波分解 (3.2.1) 和重构过程 (3.2.2), 分别由高通滤波器 $h(t)$ 和低通滤波器 $g(t)$ 组成, 编码-解码器的最佳高通滤波器系数 $H_D$ 和低通滤波器系数 $G_D$ 被用于基于ADWT的时间序列分解中。采样系数 $M = 2$ , 在这种情况下,  $h(t)$ 和 $g(t)$ 都是半带滤波器, 这意味着分解后的序列长度是原始序列的一半。编码器-解码器模型致力于寻找最优正交的小波滤波器 $h(t)$ 和 $g(t)$ , 这里, 最优是指在对序列进行小波变换时, 这些滤波器应该足够好, 能够保留与原始序列尽可能多相同的信息。

同时，根据小波分解理论，待优化的高通滤波器和低通滤波器也必须满足以下条件：

$$\sum_t h[t]^2 = 1, \quad \sum_t h[t] = \sqrt{2}, \quad \sum_t g[t]^2 = 1, \quad \sum_t g[t] = \sqrt{2}, \quad g(t) = (-1)^t h(1-t).$$

因此，求解高通滤波器和低通滤波器的损失函数，即最优滤波器  $L_w(h, g)$  被设计为：

$$L_w(h, g) = (\sum_t h[t]^2 - 1)^2 + (\sum_t h[t] - \sqrt{2})^2 + (\sum_t g[t]^2 - 1)^2 + (\sum_t g[t] - \sqrt{2})^2. \quad (3.2.8)$$

每个项都被设定为平方项，因此损失函数是可微的，以便通过梯度下降学习参数。最终，结合重构序列与原序列之间的均方根误差，本章所提出的自适应离散小波分解模块的损失函数被设计为：

$$\min_{(h,g)} L_{loss} = W_1 \text{MSE}(X, X') + W_2 L_w(h, g) + W_3 \|H_D\|_1 + W_4 \|G_D\|_1. \quad (3.2.9)$$

损失函数是三个部分的加权和：基于均方根误差（MSE）的重建序列和原始序列之间的估计，基于高通和低通滤波器系数的惩罚项目，以及基于滤波器的损失函数  $L_w(h, g)$  的约束。其中，MSE是对原始序列  $X$  和通过ADWT重建的序列  $X'$  之间差异的一种度量：

$$\text{MSE}(X, X') = \frac{1}{t} \sum_{i=1}^t (\hat{x}_i - x_i)^2. \quad (3.2.10)$$

在上述工作的基础上，用于求解最优滤波器的自动编码器的步骤已经完成。然后，得到的最优滤波器系数被用来生成不同层次的系统变量的高频和低频部分，后续章节所讨论的因果关系，都是在这些层次中独立存在的。

### 3.2.3 基于核密度估计的rényi 传递熵

在预备知识中，我们提到对于公式 (3.2.7)，边际熵和联合熵的估计是RTE计算的关键步骤。在本章中，我们通过核密度估计 (kernel density estimation, KDE)，研究多变量系统中的RTE的计算。首先，我们先研究较为简单的单变量核密度估计 (Univariate kernel density estimation, UKDE)。给定一个离散随机变量  $X_t = x_1, x_2, \dots, x_t$ ，

对 $X$ 进行KDE的标准过程可以被定义为 $\tilde{f}(x)$ :

$$\tilde{f}(x) = \frac{1}{t} \sum_{i=1}^t K_h(x - x_i) = \frac{1}{th} K\left(\frac{x - x_i}{h}\right), \quad (3.2.11)$$

其中 $t$ 是变量的长度,  $K(\cdot)$ 是核函数,  $h > 0$ , 作为公式的平滑参数, 是带宽。 $K_h(\cdot)$ 是以 $h$ 为比例因子的比例核函数。一般来说, 核函数决定了概率密度函数的形状, 它应该满足以下特性。

1. 非负性:  $K(x) > 0, \forall x \in D_K$ ;
2. 对称性:  $K(x) = K(-x), \forall x \in D_K$ ;
3. 实值性:  $K(x) \in \mathbb{R}, \forall x \in D_K$ ;
4. 归一性:  $\int K(x)dx = 1, \forall x \in D_K$ ;
5. 有界性:  $\int x^p K(x)dx = 0, \int x^{p+1} K(x)dx = k < \infty$ .

KDE的本质是将输入的数据和带宽作为核函数的参数, 从而得到几个核函数, 例如、高斯核函数:  $k(x, x') = e^{-\frac{\|x-x'\|^2}{2\sigma^2}}$ , 其中 $x'$ 是核函数的中心,  $\|x-x'\|^2$ 是向量 $x$ 和 $x'$ 之间的欧氏距离(L2范数), 而 $\sigma$ 控制高斯核函数的范围,  $\sigma$ 值越大, 高斯函数的局部范围越大。高斯核函数值随着两个向量之间的距离增加而单调地减少。此外, 内核密度估计函数是基于内核函数的线性求和和归一化得到的。理论上, 所有满足函数在定义域上的积分为1的平滑函数都可以作为KDE的核函数。由于高斯核较为典型的数学特性, 因此在过往的研究中都使用高斯核对KDE的数学性质进行讨论。因此在本章中, 我们选择高斯核进行后续讨论。

带宽 $h$ 是KDE性能的另一个重要因素。带宽反映了整个KDE曲线的平坦程度, 也量化了观察到的数据点在KDE曲线形成过程中的比例。带宽越大, 观察到的数据点在最终曲线形状中的比例越小, 整体KDE曲线越平坦; 带宽越小, 观察到的数据点在最终曲线形状中的比例越大, KDE曲线越陡峭。在公式(3.2.11)中, 对于输入数据 $x_i$ , 给定带宽 $h$ , KDE函数在 $x_i$ 处形成的曲线函数为 $\frac{1}{th} K\left(\frac{x-x_i}{h}\right)$ , 其中内核函数的内部 $h$ 用于调整KDE曲线的宽度, 而内核函数的外部 $h$ 则用于确保KDE函数积分为1。容易看出, 对于 $n \rightarrow \infty$ , 带宽 $h \rightarrow 0$ 。

接下来是关于为给定的系列样本量选择最佳带宽的讨论。我们首先关注MSE, 以

衡量概率密度函数 $f(x)$ 和其估计值 $\hat{f}(x)$ 之间的均方根误差MSE( $\hat{f}(x), f(x)$ ):

$$\begin{aligned} \text{MSE}(\hat{f}(x), f(x)) &= E(\hat{f}(x) - f(x))^2 \\ &= E(\hat{f}(x) - E\hat{f}(x) + E\hat{f}(x) - f(x))^2 \\ &= E(\hat{f}(x) - E\hat{f}(x))^2 + E(\hat{f}(x) - f(x))^2 \\ &= \text{var}(\hat{f}(x)) + \text{bias}^2(\hat{f}(x)), \end{aligned} \quad (3.2.12)$$

从上式可以看出, KDE的均方根误差MSE( $\hat{f}(x), f(x)$ )是关于 $x$ 的函数, 它由方差项(标准误差)和偏差组成。在此基础上, 我们将其扩展至对所有由 $\hat{f}(x)$ 组成的核函数的整体准确性的测量, 即平均分级平方误差(mean intergraded squared error, MISE):

$$\begin{aligned} \text{MISE}(\hat{f}(x), f(x)) &= E \int_{-\infty}^{+\infty} (\hat{f}(x) - f(x))^2 \\ &= \int_{-\infty}^{+\infty} \text{MSE}(\hat{f}(x), f(x)) \\ &= \int_{-\infty}^{+\infty} \text{var}(\hat{f}(x)) + \int_{-\infty}^{+\infty} \text{bias}^2(\hat{f}(x)). \end{aligned} \quad (3.2.13)$$

将KDE公式((3.2.11))代入上述两项, 并假设概率密度函数 $f(x)$ 至少是二阶可推导的, 对其进行二阶泰勒展开, 就可以得到:

$$\begin{aligned} \text{var}(\hat{f}(x)) &= \frac{1}{th} \int_{-\infty}^{+\infty} K^2 dx - \frac{1}{t} \int_{-\infty}^{+\infty} f^2(x) dx + o(t^{-1} + h^4), \\ \text{bias}(\hat{f}(x)) &= \frac{h^2}{2} \left( \int_{-\infty}^{+\infty} x^2 K \right) \int_{-\infty}^{+\infty} (f(x)'' dx) + o(h^2). \end{aligned}$$

因此, 对于 $X$ 整体分布进行估计所使用的 $\hat{f}(x)$ 的整体准确性度量MISE( $\hat{f}(x), f(x)$ )可以表示为:

$$\text{MISE}(\hat{f}(x), f(x)) = \frac{1}{th} \int_{-\infty}^{+\infty} K^2 dx - \frac{1}{t} \int_{-\infty}^{+\infty} f^2(x) dx + \frac{h^4}{4} \left( \int_{-\infty}^{+\infty} x^2 K \right)^2 \int_{-\infty}^{+\infty} (f(x)''^2 dx) + o(t^{-1} + h^4), \quad (3.2.14)$$

并且从上式显然得, 当 $t \rightarrow \infty$ 时, 且 $h \rightarrow 0$ 时,  $\text{MISE}(\hat{f}(x), f(x))=0$ , 并且 $th \rightarrow \infty$ 。

由此，我们可以求解最佳带宽（optimal bandwidth）。进一步，我们定义渐进平均综合平方误差（asymptotic mean integrated squared error, AMISE）为：

$$\text{AMISE}(\hat{f}(x), f(x)) = \frac{1}{th} \int_{-\infty}^{+\infty} K^2 dx - \frac{1}{t} \int_{-\infty}^{+\infty} f^2(x) dx + \frac{h^4}{4} \left( \int_{-\infty}^{+\infty} x^2 K dx \right)^2 \int_{-\infty}^{+\infty} (f(x)'' )^2 dx. \quad (3.2.15)$$

假设核函数K的属性1-5成立，它表明  $\text{MISE}(\hat{f}(x), f(x)) = \text{AMISE}(\hat{f}(x), f(x)) + o(t^{-1} + h^4)$ 。那么，通过微分公式 (3.2.15)，寻找使  $\text{MISE}(\hat{f}(x), f(x))$  最小化的最佳带宽  $h_{opti}$  的问题被转化为寻找  $\text{AMISE}(\hat{f}(x), f(x))$  的极值点：

$$\frac{\partial \text{AMISE}}{\partial h} = -\frac{1}{(th)^2} \int_{-\infty}^{+\infty} K^2 dx + h^3 \left( \int_{-\infty}^{+\infty} x^2 K dx \right)^2 \int_{-\infty}^{+\infty} (f(x)'' )^2 dx, \quad (3.2.16)$$

使上面的方程等于0，然后求解，可得：

$$h_{opti} = \left( \frac{1}{t} \frac{\int_{-\infty}^{+\infty} K^2 dx}{\int_{-\infty}^{+\infty} x^2 K dx \int_{-\infty}^{+\infty} (f(x)'' )^2 dx} \right)^{\frac{1}{5}}, \quad (3.2.17)$$

将高斯核  $K$  代入方程 3.2.17 中，最优带宽为

$$h_{opti} = \left( \frac{4}{3t} \right)^{\frac{1}{5}} \sigma \approx \frac{1.06\sigma}{t^{\frac{1}{5}}}, \quad (3.2.18)$$

其中  $\sigma$  是样本的方差，在方程 (3.2.18) 提供的最佳带宽选择下，我们容易得到，当  $t \rightarrow \infty$  时，则  $h \rightarrow 0$ ，且当  $\text{AMISE} \rightarrow 0$ ， $th \rightarrow \infty$ 。

值得注意的是，在上述讨论中，只有带宽  $h$  被认为是衡量估计结果的一个指标。然而，核函数的选择对估计结果也有影响，但一般认为，带宽对估计效果的影响更大，而核函数的选择一般不影响估计结果的误差范围。因此，在 AMISE 的讨论中我们只给出了关于带宽参数的讨论。

### 3.2.3.1 多变量核密度估计

在上一节中，对于具有一维核函数的单变量 KDE，我们得到了最佳带宽  $h$ 。然而，有一些多变量的联合概率项，例如： $p(x_{t+1}|x_t^{(k)}, y_t^l) = \frac{p(x_{t+1}, x_t^{(k)}, y_t^l)}{p(x_t^{(k)}, y_t^l)}$ 。在 RTE 中，其边际

概率与核估计器一致:

$$\begin{cases} \tilde{f}(x_{t+1}, x_t, y_t) = \frac{1}{th_u} \sum_{i=1}^t \mathcal{K}\left(\frac{u-u_i}{h_u}\right) = \frac{1}{th_x^2 h_y} \sum_{i=1}^t K_x^2\left(\frac{x-x_i}{h_x}\right)^2 K_y\left(\frac{y-y_i}{h_y}\right), \\ \tilde{f}(x_t, y_t) = \frac{1}{th_u} \sum_{i=1}^t \mathcal{K}\left(\frac{u-u_i}{h_u}\right) = \frac{1}{th_x h_y} \sum_{i=1}^t K_x\left(\frac{x-x_i}{h_x}\right) K_y\left(\frac{y-y_i}{h_y}\right), \end{cases} \quad (3.2.19)$$

其中核函数 $\mathcal{K}(u)$ 是乘法核:  $\mathcal{K}(u) = K(u_1) \times K(u_2) \times \cdots \times K(u_d)$ 。在将核函数从单变量扩展到多变量的情况下, 会出现一些重要的差异, 因此, 在这里我们选择使用乘法核。在等式左边的部分, 我们主要关注MKDE的公式和最佳带宽选择。首先, 类似于一维空间中的核估计器, 我们定义高维空间中概率密度函数 $f_H(x)$ 的一般核估计器为:

$$\tilde{f}_H(x) = \frac{1}{t} \sum_{i=1}^t K_H(x - X_i) = \frac{1}{tH} K\left(\frac{x - X_i}{H}\right), \quad (3.2.20)$$

其中带宽矩阵 $H \in \mathbb{H}^{d \times d}$ , 是一个 $d$ 阶对称正定矩阵,  $K$ 是一个 $d$ 变量球形对称密度函数,  $K_H(x) = |H|^{-\frac{1}{2}} K(|H|^{-\frac{1}{2}} x)$ 是带宽比例核函数。可以注意到, 带宽矩阵控制了多变量KDE中核函数的带宽和方向, 这与单变量情况下的一维核不同。在本文中, RTE的估计只与两个变量 $X$ 和 $Y$ 有关, 因此我们将注意力转向双变量的情况( $d = 2$ )。综上, 多变量核函数估计的MISE可以写成:

$$\text{MISE}(\hat{f}_H(x), f_H(x)) = E \int_{-\infty}^{+\infty} (\hat{f}_H(x) - f_H(x))^2 = \int_{-\infty}^{+\infty} \text{var}(\hat{f}_H(x)) + \int_{-\infty}^{+\infty} \text{bias}^2(\hat{f}_H(x)). \quad (3.2.21)$$

与单变量的情况一样, 多变量核函数估计的MISE也可以分解为变化项和偏置项: $\text{var}(\hat{f}_H(x))$ 和 $\text{bias}(\hat{f}_H(x))$ 。对这两项进行二阶泰勒展开, 我们得到:

$$\begin{cases} \text{var}(\hat{f}_H(x)) = \frac{1}{t \det(\mathbf{H})} \|\mathcal{K}\|_2^2 f_H(x) + o(t^{-1} H^{-\frac{1}{2}}), \\ \text{bias}(\hat{f}_H(x)) = \frac{1}{2} \mu_2(\mathcal{K}) \int_{-\infty}^{+\infty} [\text{tr}\{\mathbf{H}^\top \mathcal{H}_f(x) \mathbf{H}\}] dx + o(\|H\|^2), \end{cases} \quad (3.2.22)$$

其中 $\|\mathcal{K}\|_2^2$ 是 $\mathcal{K}$ 的 $d$ -维的 $L_2$ 范数,  $\mu_2(\mathcal{K}) I_d = \int_{-\infty}^{+\infty} x^\top x \mathcal{K}(x) dx$ , 而 $I_d$ 是 $d$ -维单位矩阵,  $\mathcal{H}_f(x)$ 是函数 $f_H(x)$ 的Hessian矩阵。 $\text{tr}\{\cdot\}$ 是该矩阵的秩。同样的, 为了求解最优带宽, 我们需要令多变量核函数估计的AMISE为0。在这里AMISE( $\hat{f}_H(x), f_H(x)$ )被定义为:

$$\text{AMISE}(\hat{f}_H(x), f_H(x)) = \frac{1}{t \det(\mathbf{H})} \|\mathcal{K}\|_2^2 f(x) + \frac{1}{4} \mu_2^2(\mathcal{K}) \int_{-\infty}^{+\infty} [\text{tr}\{\mathbf{H}^\top \mathcal{H}_f(x) \mathbf{H}\}]^2 dx. \quad (3.2.23)$$

多变量情况下的MISE和AMISE的关系可以写成:

$$\text{MISE}(\hat{f}_H(x), f_H(x)) = \text{AMISE}(\hat{f}_H(x), f_H(x)) + o(t^{-1}H^{-\frac{1}{2}} + \|H\|^2)$$

。并且从中可以得出, 当  $t \rightarrow \infty$  时, 那么  $\frac{1}{t \det(\mathbf{H})} \|\mathcal{K}\|_2^2 f(x) \rightarrow 0$ 。至此, 本文完成了基于AMISE的单变量和多变量情况下的KDE的统一度量, 现在我们转向寻求MKDE的最佳带宽矩阵。

与单变量的情况不同, 最佳带宽矩阵的计算与矩阵  $H$  的形式有关。根据矩阵的形式, 带宽矩阵可以分为三类:

I. Positive scalars times the identity matrix:  $\mathbf{H} = \begin{bmatrix} h^2 & 0 \\ 0 & h^2 \end{bmatrix}$

D. Siagonal positive definite matrix:  $\mathbf{H} = \begin{bmatrix} h_1^2 & 0 \\ 0 & h_2^2 \end{bmatrix}$

F. Symmetric positive definite matrix:  $\mathbf{H} = \begin{bmatrix} h_1^2 & h_1 h_2 \\ h_1 h_2 & h_2^2 \end{bmatrix}$

在第一类矩阵**I**中, 带宽矩阵是最简单的情况, 即, 核函数在坐标方向上的传播具有相同的平移量, 但这在实际场景中是一个过于严格的条件。在第二种情况下**D**, 带宽矩阵是一个对角线带宽矩阵, 它对应于具有坐标方向和平移量的核函数。在第三种情况**H**中, 带宽矩阵**F**是最一般的情况, 它允许核函数在任意方向和平移量上移动, [152]。

对于第一类矩阵**I**, 其最优带宽矩阵是一个对角阵, 对角线元素即为每个单元最优带宽。由于多元核密度估计的AMISE与未知的  $f_H(x)$  有关, 所以**D**和**F**类的最佳带宽不能直接计算。我们引入了经验法则的带宽选择方法, 它首先用于**I**类矩阵, 计算**F**类矩阵的最优带宽。在经验法则的带宽选择中, 多变量高斯分布  $N(\mu, \Sigma)$  的概率密度函数通常被用来作为多变量情况下的参考分布候选。不失一般性, 考虑将高斯分布  $N(0, I)$  作为多变量情况的参考分布, 其概率密度函数  $f_{\mathcal{K}}$ , 其中核函数  $\mathcal{K}$  是多变量高斯核:  $\mathcal{K} = K_{Gaussian}(x_1) \times K_{Gaussian}(x_2) \times \cdots \times K_{Gaussian}(x_n)$ 。在这种情况下, 我们有  $\mu_2(\mathcal{K}) = 1, \|\mathcal{K}\|_2^2 = 2^{-d}\pi^{\frac{-d}{2}}$ 。那么方程(3.2.23)中的第二个项可以写成:

$$\int_{-\infty}^{+\infty} [\text{tr}\{\mathbf{H}^\top \mathcal{H}_f(x) \mathbf{H}\}]^2 dx = \frac{1}{2^d \pi^{\frac{d}{2}} \det(\Sigma)^{\frac{1}{2}}} [2\text{tr}(\mathbf{H}^\top \mathcal{H}_f(x) \mathbf{H})^2 + \{\text{tr}\mathbf{H}^\top \mathcal{H}_f(x) \mathbf{H}\}^2],$$

其中  $\Sigma$  是方差矩阵。这个结果表明, AMISE的经验法则公式与  $\mathbf{H}$  和  $\Sigma$  的估计有关。考虑到最简单的情况, 即**I**类矩阵, 其中  $\mathbf{H}$  和  $\Sigma$  是对角矩阵:  $\mathbf{H} = \text{diag}(h_1, \dots, h_i, \dots, h_d)$  和  $\Sigma =$

$\text{diag}(\sigma_1, \dots, \sigma_i, \dots, \sigma_d)$ 。与单变量情况下的结果(3.2.17)类似, 对角线矩阵的最佳带宽为  $\tilde{h}_i = (\frac{4}{d+2}t)^{\frac{1}{d+4}}\sigma_i$ 。根据 [153] 变换, 第一个因子总是在 0.923 和 1.059 之间, 本文将其取为 1。对于  $\mathbf{D}$  和  $\mathbf{F}$  类矩阵, 我们通过将  $\mathbf{I}$  类中的  $\Sigma$  替换为其估计值  $\tilde{S}$ , 即样本方差矩阵, 可以得到 MKDE 的最优带宽如下:

$$\tilde{\mathbf{H}} = t^{-\frac{1}{d+4}} \tilde{S}. \quad (3.2.24)$$

值得注意的是, 方程式(3.2.24)相当于对数据应用了 Mahalanobis 转换, 将估计的协方差矩阵转换为一个单位方阵。基于最佳带宽矩阵, 可以计算出来自样本随机变量的概率密度分布的估计。至此, 我们就可以确定核函数并对 RTE 进行有效估计。

### 3.3 实验结果及讨论

#### 3.3.1 数据简介

为了确定所提出的方法的有效性, 我们使用合成和真实的数据集进行了全面的评估。合成数据集包括 Henon 映射模型和多元回归模型, 而真实数据集的来源是公开可用的因果推断开放数据库, 特别是混凝土抗压强度数据集和电力负荷数据集。此外, 我们还采用了一个私有的领域数据集, 即沥青路面结构的 RIOHTrack 数据集, 可以通过指定网站访问: [https://www.roadsdata.cn/pages/zhuantifuwu/zuchihuandao\\_1/kexueshuju/index.html](https://www.roadsdata.cn/pages/zhuantifuwu/zuchihuandao_1/kexueshuju/index.html)。这些数据的细节将在本节的其余部分提供。

##### 3.3.1.1 多元 Hénon 映射

Hénon 映射是一种经典的离散时间动态系统, 可以产生混沌现象。本文使用的多变量 Hénon 映射公式如下:

$$\begin{cases} x_{t+1} = 1 - ax_t^2 + y_t, \\ y_{t+1} = bx_t + cy_t, \\ z_{t+1} = dx_t + ey_t, \end{cases}$$

其中  $x$ ,  $y$  和  $z$  为状态变量, 参数设置为  $a = 1.4$ ,  $b = c = 0.3$ 、 $d = 0.5$  和  $e = 0.5$ , 初始值设置为  $x_0 = 0.1$ 、 $y_0 = 0.1$  和  $z_0 = 0.1$ 。

### 3.3.1.2 多变量自回归交互移动平均模型

多变量自回归交互移动平均（Multivariate Autoregressive Interacted Moving Average, ARIMA）模型常用于非平稳时间序列数据建模。本文使用的多变量ARIMA模型有以下三个变量：

$$\begin{cases} X_t = [Wx_t + (1 - W)y_t] + \varepsilon_{X_t}, \\ Y_t = [(1 - W)x_t + Wy_t] + \varepsilon_{Y_t}, \\ Z_t = W_x X_t + W_y Y_t + \varepsilon_{Z_t}, \end{cases}$$

其中， $\varepsilon \sim N(0, 1)$ 是白噪声， $W \in [0, 1]$ 是耦合系数，而关于 $x$ 和 $y$ 的自相关方程可以被写作：

$$\begin{cases} x_{t+1} = \sum_{n=1}^{\infty} a_n(d_x) x_{t-n}, \\ y_{t+1} = \sum_{n=1}^{\infty} a_n(d_x) y_{t-n}, \end{cases} \quad (3.3.1)$$

其中， $n$ 为滑动窗口的大小， $d \in (0, 0.5)$ 为与Hurst系数相关的记忆参数， $a_n(d)$ 为权重系数，定义为 $a_n(d) = d \frac{\Gamma(n-d)}{\Gamma(1-d)\Gamma(n+1)}$ ， $\Gamma$ 为伽马函数。耦合系数 $W$ 设为0.5，初始值设为 $x_0 = 0.1$ 和 $y_0 = 0.1$ 。

### 3.3.2 基准数据集CauseEffectPairs

CauseEffectPairs数据集由从不同领域（如气象学、生物学、医学、工程学、经济学等）的37个数据集中选取的100个不同因果对的数据组成，旨在为评估因果关系发现的可信度。我们选取了其中两组多元数据，分别是混凝土抗压强度（concrete compressive strength）和电力负荷（electricity load）。混凝土抗压强度数据集是CauseEffectPairs（[154]）中的一组开放数据集，可在UCI机器学习资源库中获取。该数据集来自于混凝土抗压强度及其多元影响因素建模研究（[155]）。除混凝土抗压强度外，变量还包括水泥、高炉矿渣、粉煤灰、水、超塑化剂、粗骨料、细骨料和混凝土龄期。每个变量有1030个数据点，我们选取前800个数据点进行实验，并将其分为4组。

电力负荷是另一组来自CauseEffectPairs数据集的数据，用于研究城市电力负荷变换的多元因果关系（[154]）。该数据集是基准的第94-96对，其中三个变量是小时数、摄氏度温度和每小时兆瓦的用电量。每个变量有9504个数据点。本文选取前8000个数据点，分成10组，每组800个数据点。

### 3.3.3 RIOHTrack数据集

有关RIOHTrack数据集的背景及具体信息在之前的章节已经介绍过，这里不再过多赘述。这里主要探究本章所提出的自适应多层次rényi传递熵在路面车辙及其影响因素之间的因果检测性能。车辙数据来自RIOHTrack数据集。车辙是指在温度、湿度等环境因素的影响下，沥青路面面层在车辆荷载作用下出现的路面沟槽现象。除了车辙，RIOHTrack还提供了其他影响因素的观测：环境温度、交通和中心点挠度(CPD)。车辙和三个变量定期采集，采集周期定义为荷载车辆累计行驶里程20000km。其中，环境温度是RIOHTrack气象站在一个时期内的平均温度。由于沥青材料在高温下最易发生变形，因此采用日最高温度代表日温度。在RIOHTrack上施加的交通荷载是从2016年12月开始的挂车荷载，交通荷载的观测值是一个加载周期内的累计荷载。此外，RIOHTrack还提供了路面弯沉盆的中心点挠度，即由落锤式挠度仪(FWD)产生的沥青路面整体结构的垂直位移。作为表征路面结构刚度的重要指标，中心点挠度被认为是路面车辙性能的影响因素。迄今为止，RIOHTrack已经收集了144组有效数据。为了用更精细的观测数据计算因果关系，我们对原始数据进行了上采样，通过线性插值将数据点增加到288个。插入值基于各维度邻域基本真实值的线性插值。上采样数据按年份分为两组，选取前144个数据点，即2017年至2018年的观测数据作为第一阶段的数据，选取其余数据点，即2019年至2021年的观测数据作为第二阶段的数据。在每个阶段，数据点被分为4组，每组有36个数据点。

### 3.3.4 结果及分析

为了验证本章所提出的方法在发现复杂系统中各变量时间序列之间因果关系方面的有效性，我们在上述合成数据和真实数据上进行了一些实验，实验结果将在本节的其余部分提供。

#### 3.3.4.1 仿真数据实验

仿真数据实验包括两个部分：对基于AWDT的时间序列分解模块的性能验证和对基于MKDE的因果关系检测性能验证。在第一个部分中，AWDT模块的度量指标是原始序列 $X$ 与AWDT重建序列 $\tilde{X}$ 之间的均方误差MSE，定义为： $MSE(X, \tilde{X}) = \frac{1}{n} \sum_{i=1}^n (X_i - \tilde{X}_i)^2$ . 70%的数据用于AWDT模块中自编码器的训练，其余20%的数据用于模型测试。在第二个部分中，为了验证因果关系检测的有效性，从 $X$ 到 $Y$ 的RTE值在噪声和时延的影响下并不是唯一的。因此，需要判断RTE值在统计上是否显著大于0。根据[156]方

法的建议，本章采用马尔可夫块引导法（Markov block bootstrap）进行显著性检验，显著水平设置为5%四分位数 ( $p < 0.05$ )，每个block大小为20。此外，根据 [157] 的建议，方程(4.3.3)中的Maekov阶数设置为  $l = k = 5$ 。

**结果 1：多元hénon映射** 在基于AWDT的时间序列分解模块中，一个影响因素是分解层数。为了避免过多分解造成的计算成本，实验中将分解层设置为2。本实验以不同的初始值重复多次。以  $x_0, y_0, z_0$  为例，两层分解结果如图3.3所示。图中从上到下依次为：原始数据、第二层产生的低频分量、第一层产生的高频分量、第二层产生的高频分量。对于每个变量，最后一层产生的低频分量表示序列的趋势，高频分量表示序列波动变化的细节，第二层的高频分量保留了更多的细节信息。可以明显看出，两级DWT可以从 hénon 图中得到各频段的序列特征。作为对比，我们选择了另一种基于深度残差网络（DRN）的最优小波系数方法作为对比基准模型。表4.1总结了不同初始小波函数下基于不同方法重构时间序列的平均MSE。

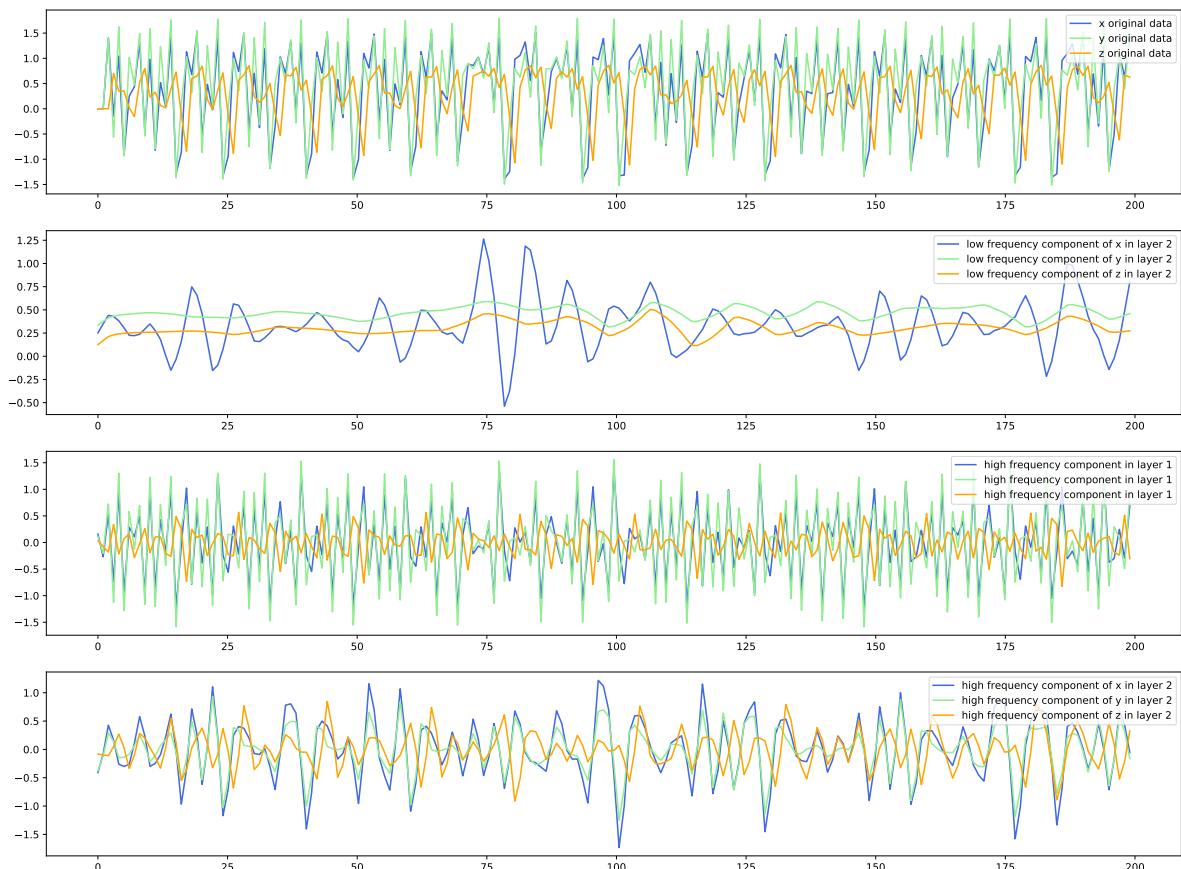


图 3.3: 利用ADWT模块对hénon映射中的序列进行db20小波分解。

候选小波包括Haar小波、Daubechies (dbN) 小波和Coiflets小波。所有候选小波

表 3.1: Average MSE of the time series reconstruction for Henon map. The values in parentheses are the standard deviation of performance.

Wavelet function	original wavelet coefficient	DRN based wavelet coefficient	ADWT based wavelet coefficient
Haar	0.4139 (0.1164)	0.3482 (0.0446)	0.2731 (0.0311)
db4	0.3579 (0.2416)	0.3095 (0.1601)	0.2125 (0.0191)
db20	0.2582 (0.0735)	0.2246 (0.0275)	0.1315 (0.0305)
Symlets4	0.3263 (0.1537)	0.2554 (0.0150)	0.1752 (0.0251)
Symlets20	0.2471 (0.0437)	0.1722 (0.0341)	0.1539 (0.0217)

应满足正交性、规则性和长度限制。一般来说，基于ADWT模块生成的小波系数和基于DRN的小波系数在时间序列重构上的表现优于原始小波系数，基于编码器-解码器框架的ADWT模块可以在hénon映射的序列上找到离散小波分解的最优系数，并且对初始值没有约束。特别是，相较于原始小波系数基于Symlets4的ADWT模块性能提升最大；而重构后的序列在Symlets20小波上的性能最好，即重构序列于原始序列之间差异最小。同时，分解结果与消失矩和对称性（即非线性相位）有关，具体来说，db小波的特点是消失矩的阶数随阶数 $N$ 的增大而增大，消失矩越大，平滑度越好，频域定位能力越强。Haar小波是db小波族的一个特例，小波函数在支持集上是离散的矩形波，支持集为 $N = 1$ 。symeltN是由dbN修改而来的近似对称小波。在高消失矩的小波函数的性能方面，基于ADWT的时间序列重构的性能略优于基于DRN基准方法。在消失矩较小的小波函数中，基于ADWT方法的性能提升更为明显，这说明本文方法能够以较小的计算代价获得较好的小波系数。总结来说，上述结果说明了ADWT模块在hénon映射的序列分解及重构上的有效性和鲁棒性。

基于ADWT模块的验证结果，我们继续验证基于MKDE的因果检测模块在不同频率层的有效性。从公式3.2.6可知，参数 $\alpha$ 是对RTE估计值的另一个影响因素。在这里， $\alpha$ 分别取0.5、1、1.5和2。基于MKDE的因果关系结果于参数 $\alpha$ 的关系如3.4所示。左图是在已知变量Z的情况下，从变量X到变量Y的RTE估计值。右图是已知变量Z的情况下，从变量Y到变量X的RTE估计值。x轴代表参数 $\alpha$ 值，y轴代表分解层数，z轴代表RTE值。一般来说， $X$ 和 $Y$ 之间存在双向的显著因果关系，这表明变量之间的因果关系取决于这些变量是否在一个动态系统中，而不是取决于序列之间的相似性。从参数 $\alpha$ 的角度来看，RTE值随着 $\alpha$ 的增大而减小，说明因果关系逐渐趋近于高概率随机事件。从分解层的角度看，信息在 $X$ 和 $Y$ 之间同时传递。从分解层的角度看，信息在 $X$ 和 $Y$ 之间同时传递，对应于低频分量的L2层的信息量从 $X$ 传递到 $Y$ ，再从 $Y$ 传递到 $X$ 是最大的，而H1层的RTE值在两个方向上都是最小的。这一结果与RTE的结论相吻合，即低频分量的因果关系受长尾事件的影响更大，而高频分量的因果关系受普遍

事件的影响更大。至此，本章所提出方法的两个模块在hénon映射上的性能验证，即对基于AWDT的时间序列分解模块的性能验证和对基于MKDE的因果关系检测性能验证，已全部完成。

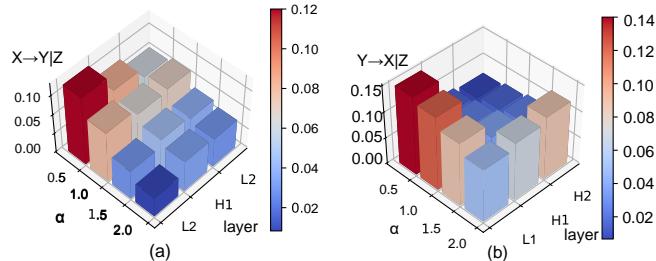


图 3.4: 基于MKDE的RTE在hénon映射上结果随参数 $\alpha$ 的和分解层数变化的可视化。

**结果 2：多元自回归模型** 在第二个实验中，我们将使用mVAR模型生成数据继续验证所提出方法的有效性。同样，实验的目的是验证ADWT模块的性能和MKDE模块的性能。表4.2总结了基于不同小波系数生成方法的mVAR时间序列重构的性能。

表 3.2: Average MSE of the ADWT module for mVAR model. The values in parentheses are the standard deviation of performance.

Wavelet function	original wavelet coefficient	DRN based wavelet coefficient	ADWT based wavelet coefficient
Haar	0.3751 (0.1817)	0.3370 (0.0681)	0.3091 (0.0010)
db4	0.3528 (0.1258)	0.3193 (0.0730)	0.3094 (0.0541)
db20	0.2736 (0.0288)	0.2433 (0.0052)	0.2018 (0.1355)
Symlets4	0.3193 (0.1306)	0.2791 (0.0680)	0.2373 (0.0446)
Symlets20	0.2122 (0.0857)	0.1736 (0.0529)	0.1757 (0.0836)

从表4.2可知，与原始小波系数相比，使用ADWT模块重构的时间序列的mse在所有重构精度上都有所提高，结果更加稳健。以Symlets20与ADWT的分解结果为例，初始值为 $X(0)$ 、 $Y(0)$ 和 $Z(0)$ ，如图3.5所示。从各变量的低频成分来看，本文采用的mVAR模型中的序列呈现出明显的上升趋势。mVAR模型的性能表明ADWT模块在面对非平稳序列时具有良好的鲁棒性。特别是，由于初始小波的重构性能已经足够好，高阶消失短小波的ADWT性能改善不大。因此我们可以得出结论，ADWT模块的性能提升并不适用于所有情况，有必要先使用初始小波进行对比来避免计算成本的浪费。

对于基于MKDE的因果检测模块，RTE值与其参数 $\alpha$ 之间及分解层数的关系如图3.6所示，左图是已知 $Z$ 的情况下，基于MKDE的RTE( $X \rightarrow Y$ )随参数变化的可视化，右图是已知 $Z$ 的情况下，基于MKDE的RTE( $Y \rightarrow X$ )随参数变化的可视化,其中参数 $\alpha$ 也设置为0.5、1、1.5和2。在3.6 (a)中，结果显示当 $\alpha = 0.5$ 时，低频段中，在已知 $Z$ 的情况下

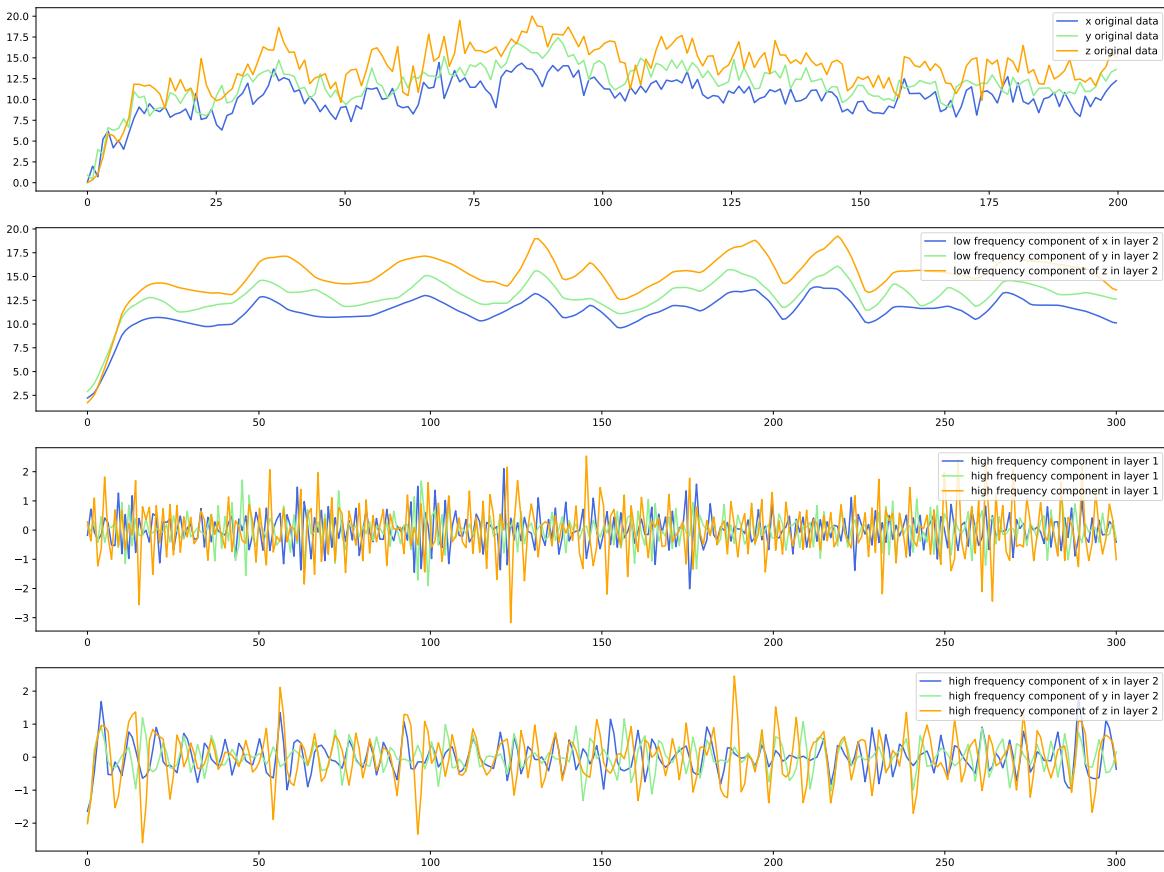


图 3.5: 利用ADWT模块对mVAR模型中的序列进行Symlets4小波分解。

下, 从 $X$ 到 $Y$ 的因果关系最大; 而当 $\alpha = 1.5$ 时, 在高频段的因果关系最大。与hénon映射的结果类似, mVAR模型中低频分量的RTE值随着 $\alpha$ 的增大而减小。另一个不可忽略的现象是, 低频分量中RTE值最统计显著时的参数是 $\alpha = 0.5$ 。这意味着在计算非平稳序列的RTE值时, 应重视低概率事件, 如mVAR模型中具有增长趋势的低频分量。基于hénon map和mVAR模型的RTE值与参数 $\alpha$ 的关系的实验, 本文在后面的实验中将L2层的参数设置为 $\alpha = 0.5$ , 高频分量设置为 $\alpha = 2$ 。

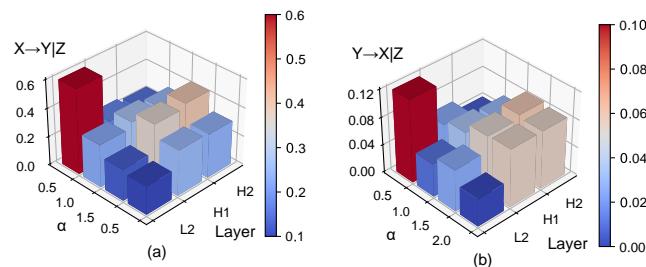


图 3.6: 基于MKDE的RTE在VAR模型上结果随参数 $\alpha$ 的和分解层数变化的可视化。

### 3.3.4.2 真实数据实验

基于合成数据的结果，我们得到了RTE相关参数的先验知识，现在我们转向真实数据的实验验证。真实数据的实验也包括两部分：ADWT模块性能验证和基于 MKDE 的 RTE 效果验证。真实数据被分成两组，ADWT模块的数据分割与仿真数据实验相同。在MKDE模块的实验中，显著性检验同样采用马尔可夫块引导法（Markov block bootstrap），显著水平设置为5%四分位数 ( $p < 0.05$ )，其中block大小也为20。方程(4.3.3) 中的时间延迟长度设为 $l = k = 5$ 。实验和比较结果将在本节的其余部分给出。

**实验 1：基准数据集CauseEffectPairs** 与仿照数据实验类型，首先，我们进行了对ADWT模块的性能实验，其中分解层设置为2，不同小波函数的重构性能如表3.3所示。从表3.3中两个基准数据集的MSE结果可以看出，与原始小波相比，基于ADWT模块的所有小波函数的分解重构性能都得到了普遍提高。同时，消失矩是影响重建性能的另一个因素，高消失矩的symlets系列小波具有较好的对称性，在对两组数据进行分析和重建时可以减少相位失真。基于Symlets20的ADWT模块是所有小波函数中性能最好的。然而，在计算成本更低的情况下，Symlets4在混凝土抗压强度数据集上的重建性能并不明显劣于Symlets20，表3.3中Symlets4的方差显示了其更好的鲁棒性。因此，我们采用Symlets4对混凝土抗压强度序列进行分解，采用Symlets20对电力负荷数据进行分解。

表 3.3: ADWT模块在混凝土抗压强度 (CCS) 和电力负荷 (EL) 上的平均MSE。括号中的值为性能标准偏差。

Wavelet function	original		ADWT	
	CCS	EL	CCS	EL
Haar	0.3527 (0.1597)	0.4136 (0.0353)	0.1137 (0.3919)	0.1754 (0.3309)
db4	0.3139 (0.0739)	0.4443 (0.1597)	0.1347 (0.1184)	0.1847 (0.5512)
db20	0.3035 (0.0510)	0.2759 (0.3431)	0.0577 (0.1666)	0.2047 (0.3217)
Symlets4	0.3570 (0.1971)	0.3841 (0.1471)	0.0534 (0.8710)	0.1573 (0.0227)
Symlets20	0.3135 (0.1730)	0.3158 (0.2828)	0.0510 (0.0293)	0.1129 (0.0583)

对于因果关系检测模块的性能评估，基准数据集CauseEffectPairs提供了每个变量之间的真实因果关系，在CCS数据集中，影响因素包括：水泥含量 (Cement)、高炉矿渣含量 (Blast furnace slag)、粉煤灰含量 (Fly ash)、水含量 (Water)、超塑化剂含量 (Superplasticizer)、粗集料含量 (Coarse aggregate)、细集料含量 (Fine aggregate) 以及使用年份 (Age)。所有的影响因素变量都是抗压强度的真实因果源。RTE模块生成的不同频段的因果关系如表3.4所示。按照仿真数据实验中的结果，RTE的参数 $\alpha$ 低

频分量和高分量分别设置为0.5和1.5。 $p$ 值小于0.05的RTE被认为具有统计意义的因果关系，并在表中加重表示。表中结果表示，在低频段中，水泥、水、使用年份对抗压强度的RTE高于显著性临界值；在高频段中，高炉矿渣、超塑化剂、粗骨料、细骨料对抗压强度的RTE高于显著性临界值。其中，从使用年份到抗压强度的信息传递 RTE ( $\text{Age} \rightarrow \text{Compressive strength}$ ) = 0.3709，从含水量到抗压强度的信息传递 RTE ( $\text{Water} \rightarrow \text{Compressive strength}$ ) = 0.3110 是低频段最大的两个值，说明混凝土的龄期和水胶比是影响抗压强度变化趋势的主要因素。在高频段，含高炉矿渣量对抗压强度的RTE值：RTE ( $\text{Blast furnace slag} \rightarrow \text{Compressive strength}$ ) = 0.2059 0.1919 在高频段上 ( $H_1, H_2$ ) 上是最大值，这表明高炉矿渣占粘结剂总量的百分比是抗压强度波动项的最大影响因素。同时，我们还注意到龄期和水分对抗压强度在低频段和高频段都有影响。同时，除了验证MKDE模块在不同层中的性能外，我们还将其与其变体模型（不含ADWT模块）进行了比较，结果显示，变体只检测高炉矿渣和龄期对抗压强度的因果关系，这点也同样说明了ADWT模块的不可替代性。

表 3.4: 不同频段混凝土抗压强度与其影响因素之间的RTE值，其中括号内为 $p$ 值， $\rightarrow$ 为因果关系方向。加重字段表示统计显著的因果关系。

	L2	H1	H2	variant
Cement $\rightarrow$ Compressive strength	<b>0.0714 (0.0187)</b>	0.2324 (0.3610)	0.2024 (0.0910)	0.2024 (0.0910)
Blast furnace slag $\rightarrow$ Compressive strength	<b>0.0371 (0.0027)</b>	0.2059 (0.0371)	0.1918 (0.0273)	<b>0.0314 (0.0305)</b>
Fly ash $\rightarrow$ Compressive strength	<b>0.2714 (0.0187)</b>	<b>0.0024 (0.0173)</b>	<b>0.0013 (0.0210)</b>	0.2715 (0.8486)
Water $\rightarrow$ Compressive strength	<b>0.3110 (0.0372)</b>	0.0215 (0.0835)	<b>0.0311 (0.0254)</b>	0.3062 (0.4136)
Super plasticizer $\rightarrow$ Compressive strength	<b>0.0412 (0.0158)</b>	0.0514 (0.0179)	0.0704 (0.0122)	0.2581 (0.3193)
Coarse aggregate $\rightarrow$ Compressive strength	<b>0.0813 (0.0369)</b>	<b>0.0144 (0.0014)</b>	<b>0.0637 (0.0023)</b>	0.0404 (0.7473)
Fine aggregate $\rightarrow$ Compressive strength	<b>0.1073 (0.0032)</b>	<b>0.2324 (0.0010)</b>	<b>0.3014 (0.0110)</b>	0.0079 (0.6751)
Age $\rightarrow$ Compressive strength	<b>0.1209 (0.0087)</b>	<b>0.1692 (0.0610)</b>	<b>0.2097 (0.0031)</b>	<b>0.3105 (0.0457)</b>

同样的MKDE模块性能实验也适用于电力负荷数据集。真实的因果关系也在数据集中提供，分别是：用电时刻  $\rightarrow$  温度，用电时刻  $\rightarrow$  用电量，温度  $\rightarrow$  用电量。基于MKDE模块的RTE估计结果如表3.5所示。

表 3.5: 不同频段电力负荷与其影响因素之间的RTE值，其中括号内为 $p$ 值， $\rightarrow$ 为因果关系方向。加重字段表示统计意义上的因果关系。

	L2	H1	H2	variants
Hour $\rightarrow$ Temperature	<b>0.0035 (0.0404)</b>	/	/	0.3318 (0.1736)
Hour $\rightarrow$ Electricity Load	<b>0.3136 (0.0349)</b>	/	/	0.0055 (0.3910)
Temperature $\rightarrow$ Electricity Load	<b>0.2075 (0.0045)</b>	<b>0.2763 (0.0060)</b>	0.1818 (0.0749)	<b>0.0024 (0.0091)</b>

总的来说，基于 MKDE模块的 RTE 估计值在真实三种因果关系上都是统计显著

的，这一意味着三种因果关系都被检测到。由于小时是单调递增序列，因此在ADWT模块中不会对其进行分解。在趋势项中，用电时刻和温度是电力负荷的重要影响因素，小时的变化也会影响温度的变化。特别是， $\text{RTE}(\text{Hour} \rightarrow \text{Electricity Load}) = 0.3136 > \text{RTE}(\text{Temperature} \rightarrow \text{Electricity Load}) = 0.2075$ ，这表明电负荷趋势受时间变化的影响更大。这一结果背后的事实是，人类活动由一天中的时间决定，而这些活动所需的电力进一步决定了电力负荷的变化趋势。同时，温度变化也是电力负荷波动的原因，但不会影响高频段的电力负荷波动。此外，对于不带ADWT模块的变型，虽然可以检测出温度与电负荷的因果关系，但从RTE值来看，用电时刻与温度、电负荷的因果关系并不显著。

在接下来的实验中，我们还比较了基于MKDE模块和其他不同估计器计算RTE的性能，本文使用了线性估计器（[146]）和二值估计器（[158]）。首先是混凝土抗压强度数据集，以低频分量为例，混凝土抗压强度数据集的计算结果如图3.7(a)-(c)所示，列和行的数字1-9依次表示水泥、高炉矿渣、粉煤灰、水、超塑化剂、粗骨料、细骨料和抗压强度。纵轴元素表示源变量，横轴元素表示目标变量。矩阵元素值表示从源变量到目标变量的RTE值。矩阵元素颜色越深，表示从源变量到目标变量的RTE值越高。根据原文中的观点 [155]，所有变量都会引起混凝土抗压强度的变化。所有估计值的结果都表明变量与混凝土抗压强度之间存在显著的因果关系。然而，只有基于MKDE的RTE计算的结果发现了其他变量与粉煤灰含量之间的最大因果关系，而其他两个估计器所得到的RTE忽略了这一点。我们的方法计算的结果还表明，混凝土的使用年份和含水量水是对混凝土抗压强度影响最大的两个因素。相反，基于二值估计的RTE识别出了从混凝土抗压强度到含水量、超塑化剂含量、粗骨料含量和细骨料含量的显著因果关系，这与事实有明显的冲突。此外，基于线性估计的RTE会遗漏非线性偶然性，例如，基于该估计器的RTE值检测水与其他变量的非线性因果关系，这也是违反常理的，是一个假性因果关系。

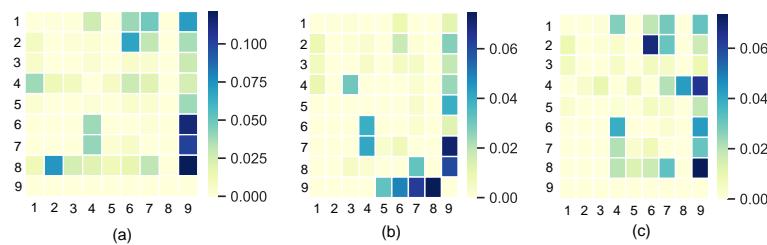


图 3.7：不同估计器的RTE值在混凝土抗压强度数据集上的对比。（a）基于MKDE的RTE。（b）基于线性估计的 RTE。（c）基于二值估计的RTE。

接下来是电力负荷数据集上，不同估计器的RTE的比较，结果如图3.8 (a)-(c)所示，列和行的数字1-3依次表示时刻、温度和用电负荷。纵轴元素表示源变量，水平轴元素表示目标变量。。数据集中真实的因果关系是，时刻和温度是互为因果关系的，它们都是电力负荷的原因。这可以从图3.8 (a) MKDE模块结果可以有效地检测出共同因果关系。时刻和温度的因果关系显著高于图3.8 (c)的结果，即bin估计器生成的RTE。同时，bin估计器给出了从电力负荷到时刻的显著RTE值，以及时刻与温度之间的显著RTE值，这些都与事实相反。这个结果说明了bin估计器在处理高耦合系统时估计假因果关系。因此，基于混凝土抗压强度和电荷载的结果，我们可以得出结论，MKDE模块可以更准确地捕捉多变量之间的因果关系。

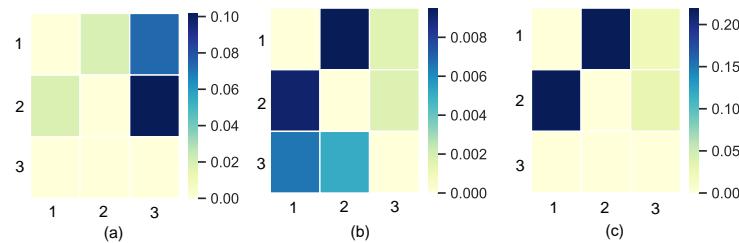


图 3.8: 不同估计器的RTE值在电力负荷数据集上的对比。(a)基于MKDE模块的RTE。  
(b)基于RTE的线性估计。(c)基于bin估计的RTE。

## 实验 2: RIOHTrack 数据集

最后，我们验证ADWT模块的性能和基于MKDE的RTE模块对来自RIOHTrack的车辙数据及其影响因素的有效性。在ADWT模块的验证中，初始设置分解层数为4，并选择Symlets20小波函数。车辙分解结果如图3.9所示，图中从上到下分别为：原始数据、第4层产生的低频分量、第1层、第2层、第3层和第4层产生的高频分量。

在每个分解层中，低频滤波器和高频滤波器对应的频段分别为  $(0, \frac{f_l}{2^{l+1}})$  和  $(\frac{f_l}{2^{l+1}}, \frac{f_l}{2^l})$ ，其中  $f_l$  为采样频率， $l$  为层数。因此，原始序列的分量被分解为5个频段： $L4 : (0Hz - 4Hz)$ ,  $H4 : (4Hz - 8Hz)$ ,  $H3 : (8Hz - 16Hz)$ ,  $H2 : (16Hz - 32Hz)$ ,  $H1 : (32Hz - 64Hz)$ ，其中  $L4$  代表低频段， $H1 - H4$  代表高频段。ADWT模块在RIOHTrack数据上的性能如表3.6所示。与仿真数据的结果一样，在ADWT模块中，消失矩阶越高的小波的mse值越高，其中Symlets20与ADWT模块的mse值最佳，为0.1039。此外，我们还比较了上采样到原始序列长度的高频分量，度量指标为两个分量之间的mse。结果表3.7中所示，低频分量的变化率从第三层开始变小，到第四层高频分量几乎没有变化，这说明在更深的一层没有更多的信息被分解出来。RIOHTrack数据集中STR1路面结构在不同层的

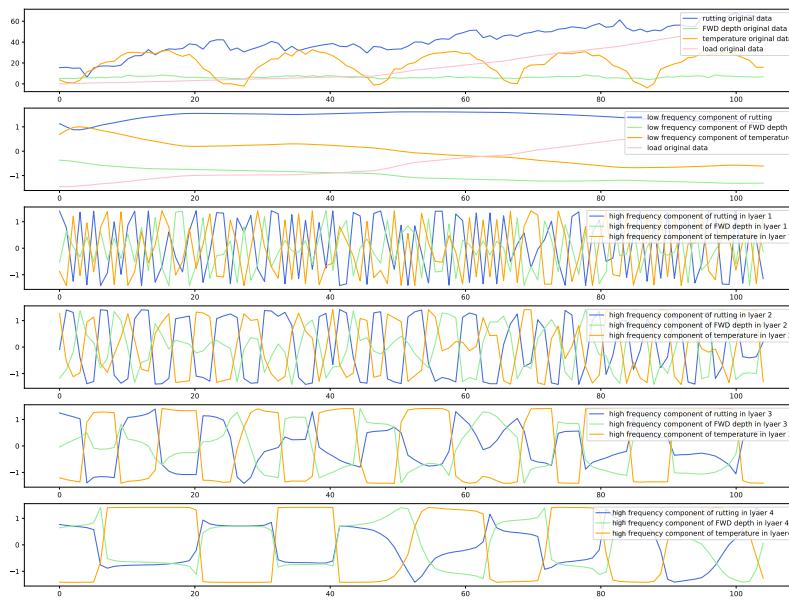


图 3.9: 利用ADWT模块对车辙及其影响因素进行Symlets20小波分解。

分解结果如图3.8所示。各变量的分解序列被归一化，映射为[0,1]，以消除量纲的影响。有趣的是，车辙的高频分量、中心点挠度和温度在第4层( $4Hz - 8Hz$ )前40个采样周期的变化周期基本相同，在第3层( $8Hz - 16Hz$ )后30个采样周期的变化周期也基本相同，这预示着这些变量在这些频段可能存在因果关系。因此，在接下来的对MKDE模块的验证部分中，我们将继续关注这两个频段来验证我们的猜想。

表 3.6: ADWT模块在RIOHTrack数据集上的平均重构性能，括号中的值为性能的标准偏差。

Wavelet	original	ADWT
Haar	0.4943 (0.3710)	0.2751 (0.0935)
db4	0.5335 (0.6233)	0.2159 (0.1481)
db20	0.3573 (0.3710)	0.1365 (0.1179)
Symlets4	0.5228 (0.8710)	0.2021 (0.1510)
Symlets20	0.2920 (0.1531)	0.1039 (0.0813)

在RIOHTrack的MKDE模块实验部分，我们设置RTE参数 $\alpha = 0.5$ 为低频分量， $\alpha = 1.5$ 为高频分量。根据沥青路面弹性层系理论，路面车辙与路面结构、层厚、路面力学响应等内部因素和温度、交通荷载轴线、荷载轴重等外部因素有关。影响因素对车辙产生的RTE值见表4.6和3.9。我们还将结果与框架变量（无ADWT）进行了比较。总的来说，本文提出的方法所确定的具有统计意义的因果关系与实际情况相吻合。特别地，三个影响因素与车辙的因果关系在低频品牌中都是显著的，这意味着它们是车辙

表 3.7: ADWT模块在RIOHTrack数据集上每一层平均重构性能，括号中的值为性能的标准偏差。

Wavelet	$L_1 \sim L_2$	$L_2 \sim L_3$	$L_3 \sim L_4$
Haar	0.4738 (0.2609)	0.1381 (0.1110)	0.0013 (0.0105)
db4	0.5141 (0.6130)	0.2173 (0.3537)	0.0079 (0.0310)
db20	0.3299 (0.8710)	0.1818 (0.0100)	0.0016 (0.0007)
Symlets4	0.4107 (0.2310)	0.1539 (0.1710)	0.0031 (0.0010)
Symlets20	0.3866 (0.1919)	0.1581 (0.2391)	0.0013 (0.0126)

加深趋势项的原因。并且RTE(Temperature → Rutting) = 0.1307为最高值，表明温度是影响轨道车辙演变趋势的最大影响因素。此外，基于MKDE的RTE值从温度和中心点挠度高频分量到车辙高频分量在第3层和第4层也是显著的，这与上一小节的推测不谋而合。有趣的是，车辙与中心点挠度在高频段具有明显的因果关系，也就是说车辙与中心点挠度在路面系统中是一对互为因果的耦合变量。相反，不含ADWT的框架仅检测到荷载与车辙之间存在显著的因果关系，没有证据表明其他变量对车辙有影响，而且中心点挠度似乎不受车辙的影响。因此，上述结果表明ADWT可以帮助框架检测RIOHTrack数据的多尺度因果关系。

表 3.8: RIOHTrack数据集在不同频段 (L4-H2层) 车辙与影响因素之间RTE值，括号内为p值，→为因果关系方向。加重字段表示统计意义上的因果关系。

	L4	H1	H2
Temperature → Rutting	<b>0.1307 (0.0320)</b>	<b>0.0683 (0.0041)</b>	0.0803 (0.3021)
Load → Rutting	<b>0.1094 (0.0017)</b>	/	/
CPD → Rutting	<b>0.1135 (0.0217)</b>	<b>0.2324 (0.0037)</b>	<b>0.2024 (0.0261)</b>
Rutting → CPD	0.0311 (0.0031)	0.0164 (0.3610)	<b>0.0424 (0.0210)</b>

表 3.9: RIOHTrack数据集在不同频段 (H3-H4层) 车辙与影响因素之间RTE值，括号内为p值，→为因果关系方向。加重字段表示统计意义上的因果关系。

	L4	H3	H4	variant
Temperature → Rutting	0.0803 (0.3021)	0.0402 (0.0734)	<b>0.0392 (0.0210)</b>	0.0065 (0.4910)
Load → Rutting	/	/	/	<b>0.1027 (0.0010)</b>
CPD → Rutting	<b>0.2024 (0.0261)</b>	<b>0.2024 (0.0391)</b>	0.0052 (0.0910)	0.0030 (0.3907)
Rutting → CPD	<b>0.0424 (0.0210)</b>	0.1024 (0.0535)	0.0076 (0.0769)	0.0013 (0.1703)

在最后的实验中，我们将基于MKDE的RTE模块的性能与上述其他两种方法，即线性估计器和bin估计进行了比较。以第4层低频分量的全部19种路面为例，基于RTE的2016-2018年和2019-2021年的因果网络分别如图3.10和3.11所示。如上所述，所有影响因素都是车辙产生的原因，车辙也是中心点挠度产生的原因。总之，图3.10 (a) 和图3.11 (a) 中的两个结果都表明MKDE模块能够有效地检测出各变量之间的显著因果关系。对于基于线性估计的RTE，如图3.10(b)所示，车辙深度与中心点挠度之间的因果关系未被检测到，两个周期内温度与中心点挠度之间的因果关系均被忽略。图3.10(c)和

图3.11(c)中的结果表明，bin估计器识别出了荷载与温度之间的强因果关系，这显然是一個错误的结论。此外，两组结果为RIOHTrack的因果检测提供了重要启示，后期中心点挠度与车辙波动项的因果关系显著低于前期，而线性发展的交通荷载与车辙的因果关系变得更强。这些结论与车辙演变三阶段理论相一致，结果证明了本章提出的ADWTRTE的有效性，也为进一步表征车辙演变动力学模型提供了理论依据。至此，对于本章提出的ADWTRTE所有的实验完成，结果验证了本章所提出的方法的有效性与健壮性。

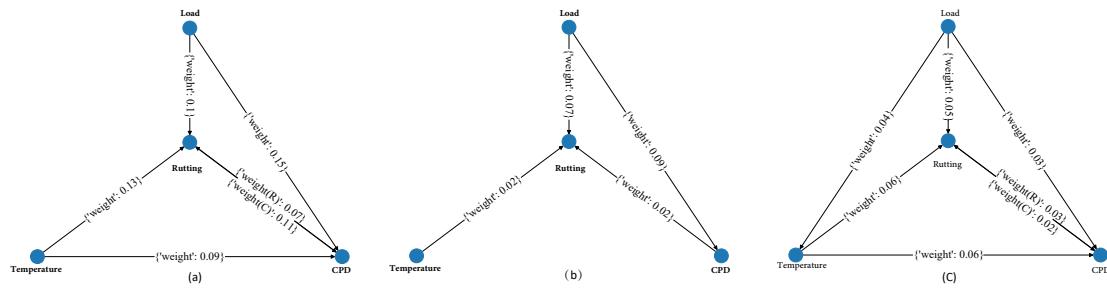


图 3.10: 2016-2018年RIOHTrack数据集上车辙及其影响因素间，不同估计器生成的RTE值比较。(a) 基于MKDE模块的RTE。(b) 基于线性估计的RTE。(c) 基于二进制估计的RTE。

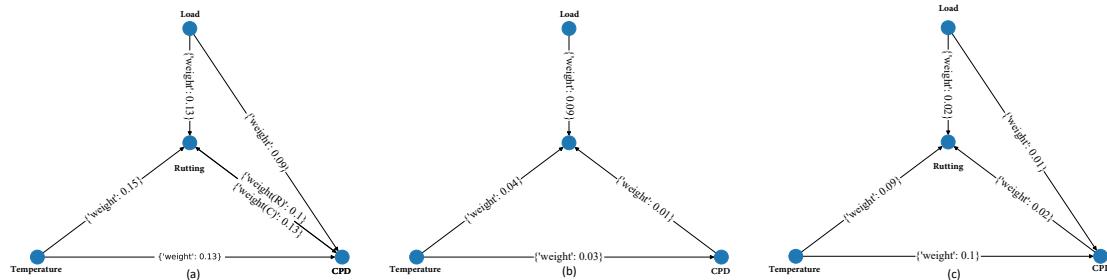


图 3.11: 2018-2021年RIOHTrack数据集上车辙及其影响因素间，不同估计器生成的RTE值比较。(a) 基于MKDE模块的RTE。(b) 基于线性估计的RTE。(c) 基于二进制估计的RTE。

### 3.4 小结

在基于rényi传递熵的因果检测模型研究中，如何更准确地估计rényi传递熵以及如何度量多层系统的因果关系是许多学者关注的问题。针对这两个问题，本文提出了一种基于核密度估计的新型多尺度RTE方法。该方法包括两个部分：基于ADWT的时间序列分解和基于MKDE的因果网络生成。该方法的流程是将复杂系统的所有变量作为

输入，通过ADWT模块计算出最优小波系数。这些系数用于将原始时间序列分解为多层高频和低频分量，分别对应波动项和趋势项。本文用变量邻接矩阵表示系统内变量的因果关系。变量邻接矩阵由最后一层的低频分量和每一层的高频分量组成。矩阵中的节点表示系统的变量，边表示变量之间的因果关系，由MKDE模块确定。每一层的因果网络就是该方法的输出结果。具体来说，在基于ADWT的时间序列分解中，设计了一个最优小波系数的自动编码器框架，以减少分解过程中的信息损伤。自动编码器框架由小波函数组成，通过最小化原始时间和重构时间序列之间的均方误差，在编码器-解码器中训练最优小波系数。在基于MKDE的因果网络生成中，RTE值由KDE估计，KDE的核函数由最优带宽选择决定，KDE通过最小化概率密度分布与核函数之间的AMISE来测量。为了验证本研究中提出的方法的有效性，设计了一系列实验，在合成数据和真实数据上测试该模型。进行这些实验的目的是评估模型在各种情况和不同条件下的性能。合成数据使用多种参数生成，而真实数据则从多个不同来源收集。在使用合成数据进行验证时，我们进行了实验来研究ADWT模块在不同小波上重建序列的性能。我们的研究结果表明，ADWT模块能够有效降低重建数据与原始数据之间的误差。随后，我们探讨了参数 $\alpha$ 的选择对基于RTE的因果关系检测的影响。在真实数据的验证中，我们对ADWT的性能进行了评估，并确定了在该数据上性能最好的小波基。最后，我们利用本文提出的方法构建了基于RTE值的因果网络，并利用不同的估计方法比较了因果网络构建的效果。研究结果表明，该方法在处理合成数据和真实数据时都非常有效，非常适合因果检测领域的一系列应用。本章结果已发表在国际刊物 Neural Networks 上。具体详见作者发表论文的清单。



# 第四章 基于多级小波分解和传递熵的车辙预测框架

## 4.1 引言

在前两章中，我们对以传递熵为基础的因果关系模型进行了理论和方法上的创新。从本章开始，我们将展示基于传递熵的因果网络在复杂系统中建模的应用。以沥青路面车辙预测模型为例，本章将展示因果网络作为模型中特征选取模块的应用。车辙作为沥青路面的一种主要破坏形式，不仅会导致路面结构性能衰减，还会加速路面损坏。此外，车辙还会威胁行车安全，对正常交通极为不利 [?, ?]。同时，由于车辙变形既发生在沥青路面的面层，也发生在下层，因此车辙的养护工作十分困难 [?]。车辙问题的解决不仅有赖于抗车辙材料的研发和沥青路面抗车辙性能的提高，还可以在设计阶段根据路面的影响因素预测车辙变形的发展，从源头上保证路面的抗车辙性能 [?]。因此，高精度车辙预测模型的研究对于评价车辙病害、提高路面使用寿命具有重要意义。近年来，车辙预测研究取得了长足的发展。然而，以下两个突出问题依然存在，并严重制约了车辙预测性能及其应用。第一个问题是缺乏有效的、通用的数据处理、特征选择和建模框架。车辙预测模型的研究需要分析车辙变形的影响因素，车辙变形是沥青路面在交通荷载因素、气候环境因素和路面材料因素的耦合作用下产生的永久变形。常见的经验模型是基于多元回归分析，这种方法会忽略历史数据的依赖性。因此，时间序列模型通常用于研究复杂系统并了解其随时间变化的行为方式 (cite5)。从统计学角度来看，时间序列模型的一般定义可表述为：

$$\hat{y}_{i,t+n} = f(Y_{i,t-k}, X_{j,t-k}, s),$$

其中，模型输出  $\hat{y}_{i,t+n}$  为序列， $n$  为预测步数； $Y_{t-k} = \{y_t, y_{t-1}, \dots, y_{t-k}\}$  是  $\hat{y}$  的历史数据； $i$  是前提条件维度，在本任务中等于 1； $X_{j,t-k} = \{x_{j,t}, x_{j,t-1}, \dots, x_{j,t-k}\}$  是输入变量序列， $j$  代表输入维度； $s$  是与实体相关的静态信息。上述三项构成了模型的整个输入，而  $f : R^{i \times k} \rightarrow R^m$  是待确定的预测模型。随着统计学和大数据的发展，基于机器学习和深度学习的车辙预测模型已被引入，有助于实现长期时间序列建模 [?]。特别地，作为复杂系统数据处理的常用技术，时间序列分解通常用于识别原始时间序列的趋势分量和波动分量 [?]。此外，基于因果关系的特征值对于建立可解释且稳健的预测模型也有

潜在的益处 [?].

第二个问题是缺乏可信的沥青路面数据。沥青路面的数据来源主要包括：实验室试验、加速路面试验（APT）、路面长期性能跟踪（LTPP）等。与实验室试验和 LTPP 研究相比，加速路面试验不仅能更真实地反映路面的使用状态，而且能在更短的时间内获得路面的使用状态数据。因此，全尺寸轨迹作为一种收集多种路面结构和材料时空性能数据的典型 APT 方法，已被许多国家的研究机构和研究人员广泛采用。交通运输部公路科学研究院轨道（RIOHTrack）是我国第一条全尺寸轨道，于2017年在北京建成 [?]. 为克服上述问题，本文提出了一种基于小波分解和多元转移熵的新型车辙预测框架，旨在以可解释的分析方法捕捉车辙及其影响因素之间的因果关系，准确预测车辙演变曲线。本文提出的框架包括三个模块：基于多级离散小波分解（MDWD）的数据处理模块、基于多元传递熵（MTE）的特征选择模块和基于机器学习模型的预测模块。在车辙预测模块中，使用了两个预测模型进行车辙预测，分别是用于预测高频分量的多元高斯过程回归模型（GPR）和用于预测低频分量的双向门控循环单元（bi-GRU），并进一步提供了基于这两个模块的一些因果关系分析。最后，预测结果通过反小波变换进行重构，将各频段的预测结果融合为框架的输出结果。性能比较结果表明，所提出的框架比一些最先进的基线时间序列预测模型更有效。此外，本章还设置了消融实验，以评估每个模块对框架性能的影响。

## 4.2 相关工作

### 4.2.1 车辙预测模型

沥青路面车辙评价机理可分为三个阶段 [?]. 第一阶段是初始沥青混合料的压实过程。该阶段的车辙是人工为提高路面承载能力而产生的。第二阶段为稳定阶段，车辙是沥青混合料在高温或重载作用下抗剪能力不足而产生的流动变形。两侧的隆起量逐渐与凹陷量相等。这一阶段车辙演变的速度相对较慢。第三阶段为快速发展阶段，沥青混合料骨架结构的重新排列过程和沥青混合料的剪切破坏是车辙产生的主要原因。在这一阶段，车辙加深的速度明显加快。从以上分析可以看出，沥青路面车辙主要是沥青层在外部环境因素和内部结构因素的双重作用下产生的流动变形 [?, ?]. 近年来，人们在车辙预测模型方面做了大量工作。车辙预测模型可分为三类：力学模型、经验模型和力学-经验模型。力学模型以弹性或粘弹性-塑性分层体系理论为基础，通过分析材料变形与应力、应变之间的关系来预测沥青路面的车辙深度。 [?]根据实验室蠕变试

验的结果提出了修正的麦克斯韦模型。[?]提出了带有时间和温度因子的臼化 Burgers 模型。cite15利用统计学原理分析了车辙与其他影响因素之间的关系，并在此基础上建立了预测车辙的经验模型。[?]提出了车辙预测模型，并分析了多变量影响因素。引入了一些深度学习模型以提高车辙预测模型的准确性。[?]提出了基于多层感知（MLP）的车辙演变模型。力学-经验模型基于力学理论建立车辙预测模型，并通过实词数据计算参数。其代表作是 AASTO2002 模型，该模型提供了一种称为 MEPDG 的模型设计方法，用于建立各层的车辙预测模型。M-E模型的参数训练可以采用多元线性回归（MLR）和随机预测的方法，这样可以提高50%的精度 [?].

#### 4.2.2 路面材料和结构的非稳态时间序列预测

实践中的复杂系统往往表现出多变量非稳态演化行为，所获得的系统信息通常是不完整和不确定的。一般来说，对于沥青路面车辙演变预测这样的典型复杂任务，很难建立精确的分析数学模型。因此，通常通过观测或实验获取时间序列来分析复杂系统。非平稳时间序列预测旨在分析多变量之间的关系，了解纵向观测数据的依赖性，历来是学术界研究的热门领域，应用广泛。从广义上讲，非平稳时间序列的预测模型可分为参数模型和非参数模型，这取决于模型是采取预定义的形式还是纯粹由数据驱动的构造 [?].

参数预测模型通常提供一个具有有限数量参数  $\theta$  的显式函数来描述输入  $X$  和输出  $Y$  之间的关系，其中  $\theta$  是通过时间序列变现来优化的 [?]. 作为静态时间序列的经典自回归模型，ARIMA 通过差分来降低非静态性。[?]使用基于 ARMIA 的模型，通过建立结构时间序列来预测交通基础设施。基于神经网络和深度学习的方法因其强大的拟合能力和无需静态假设而被广泛应用于非静态时间序列预测任务中。本文提出了一种基于特征融合 LSTM-BPNN 的路面性能预测模型，并引入了注意力机制 [?]. 基于支持向量机（SVM）的模型是另一种经典的参数模型，例如，[?]提出了一种基于粒子群优化（particle swarm optimization, PSO）和支持向量回归（support vector regression, SVR）的沥青路面性能预测模型。

与参数模型相比，非参数模型是在大量观测数据的基础上表示非平稳的动态变化，不需要结构假设和复杂的建模。[?]和 cite25在路面裂缝问题上测试并比较了多种非参数机器学习方法，包括随机森林、朴素贝叶斯和 K-近邻算法。[?] 将经验模态分解（empirical mode decomposition, EMD）引入路面裂缝分析。[?]使用小波分解处理车辙深度时间序列数据，并用ARIMA预测车辙，该方法可以捕捉时间序列的频率特性进

行车辙预测。

### 4.3 模型描述

本章提出的多元时间序列分析方法可分为三个模块：（1）基于多级小波分解（MDWD）的数据处理；（2）基于多元传递熵（MTE）的因果推断；（3）混合时间序列预测模型。该方法的架构如图4.1所示，其中 MDWD 模块和 MTE 模块在框架的多尺度因果特征检测中发挥了重要作用。基于 MDWD 模块，原始时间序列被分解为低频分量和高频分量，分别代表原始时间序列的趋势项和波动项。然后，MTE 模块选择与车辙具有显著静态偶然关系的影响因素作为各频段时间序列预测模型的特征。最后，对各频段的预测序列进行重构，设计基于融合注意力机制的预测模型，生成车辙预测结果。本节接下来将详细介绍这三个模块。

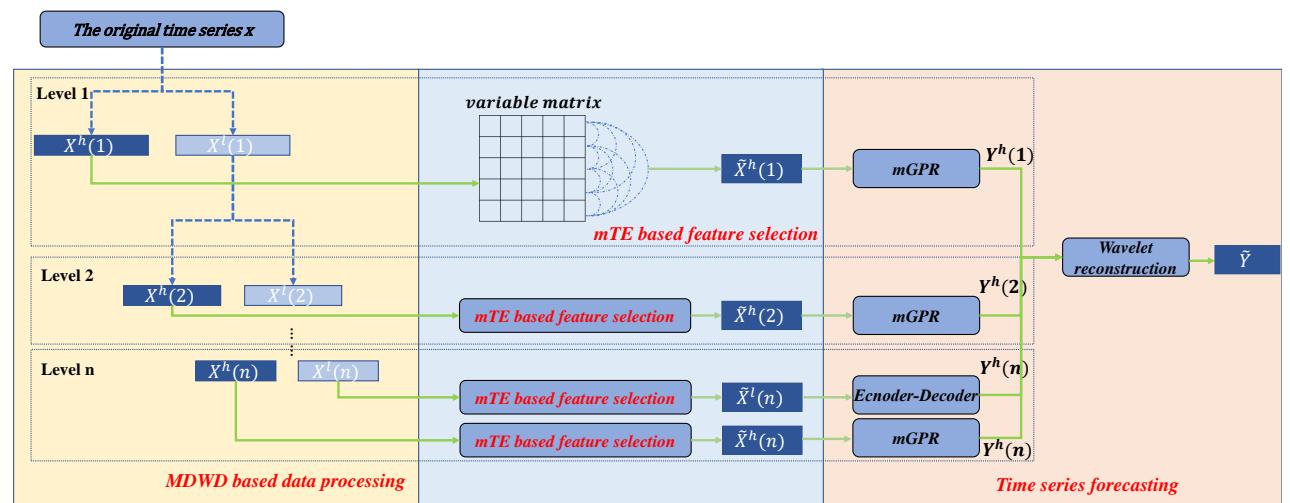


图 4.1: 基于多级小波分解和多元传递熵的时间序列分析框架示意图

所提模型的结构包括三个模块：基于 MDWD 的数据处理模块、基于 MTE 的特征选择模块和时间序列预测模块。模型的输入  $X \in \mathbb{R}^{n \times T}$  由  $n - 1$  个影响因素和长度为  $T$  的车辙时间序列组成。输出  $\tilde{Y} \in \mathbb{R}^{1 \times \tau}$  为沥青路面车辙预测值，其中  $\tau$  为预测步长。基于 MDWD 的数据处理模块将输入分解成不同的层次。每一级中的高频成分和最后一级中的低频成分作为每一级中相应的基于 MTE 的特征选择模块的输入。最后，在时间序列预测模块中，高频成分预测模型采用多元高斯过程回归模型，低频成分预测模型采用基于双向长短记忆网络（LSTM）和融合注意力机制的预测模型。

**基于 MDWD 的数据处理模块** 有关多级小波分解（MDWD）概念介绍及推导

已在第二章中给出，本章不再赘述，这里只给出本章所提出框架中使用的内容。其中MDWD模块分为时间序列分解部分和重构部分，对于分解部分，给定输入时间序列  $X_t = x_1 \text{fix}_2 \text{fi} \dots \text{fix}_t$ ，MDWD 在  $(i+1)$ -th 层的分解输出可定义为：

$$\begin{aligned} y_t^l(i+1) &= \sum_{k=1}^K x_{t+k-1}^l(i) \cdot l_k, \\ y_t^h(i+1) &= \sum_{k=1}^K x_{t+k-1}^h(i) \cdot h_k, \end{aligned} \quad (4.3.1)$$

其中，右侧的输入为第  $i$  层的低频序列分量  $X^l$  和高频序列分量  $X^h$ ，它们是上一层分解的子序列  $Y^l$  和  $Y_h$  的下采样，输出  $y_t^l(i+1)$  和  $y_t^h(i+1)$  分别是  $t$  时间步长的低频序列分量和  $i$  级别的高频序列分量。函数  $l_k$  和  $h_k$  是低通和高通滤波器的脉冲响应。从每一层的低频分量  $Y^l$  和高频分量  $Y_h$  可以重建级别  $i-1$  的时间序列  $X$ ：

$$x_t^l(i-1) = \sum_{k=1}^K y_{t+k-1}^l(i) \cdot l_k + \sum_{k=1}^K y_{t+k-1}^h(i) \cdot h_k. \quad (4.3.2)$$

这一过程也被称为逆离散小波变换（inverse discrete wavelet transform, IDWT）。它的过程包括向上采样和重建。变换函数  $l_k$  和  $h_k$  是上一级函数的缩放和移位。

**基于 MTE 的特征抽取模块** 在上一章中，我们已经介绍了一种新型的基于核函数估计的方法来计算MTE，尽管该方法具有准确性高和计算代价小的优点。但由于其本身对于MTE中参数马尔可夫阶数 $\tau$ 的选取并没有针对性的讨论，导致在选取 $\tau$ 值时只能根据经验选择，会出现虚假因果关系，甚至无法发现显著因果关系。为了能够选择出统计显著的传递熵（effective transfer entropy, ETE），本章将继续介绍一种基于贪心算法的MTE计算方法，进一步确定公式中的Markov阶数，首先，我们回顾TE的定义，从  $Y$  到  $X$  的 TE 是以历史数据  $X_t^k$  为条件，从源过程  $Y_t$  到  $X$  的实现值  $x_\tau$  的条件互信息：

$$T_{Y \rightarrow X} = I(Y_t^k; x_\tau | X_t^k) = - \sum_{i=1}^{t-k} \frac{\log(p(x_\tau | x_i, y_i))}{p(p(x_\tau | x_i))}, \quad (4.3.3)$$

其中  $Y_t^k$  和  $X_t^k$  为长度为  $k$  的历史序列， $x_\tau$  为  $\tau$  时刻的随机变量。 $T_{Y \rightarrow X}$  大于 0 意味着  $Y$  是  $X$  的原因。否则， $Y$  就是  $X$  的结果。值得注意的是，TE 取决于历史序列长度  $k$ 。 $T_{Y \rightarrow X}$  可以表示整个概念信息传递，而  $k \rightarrow \infty$ 。但是， $\tau$  值的选取会影响 TE 的显著性，甚至会带来冗余信息。为了克服这些问题，[?] 提出了集体传递熵(CTE)，它表示因果源的增量条件互信息之和：

$$T_{Y \rightarrow X} = \sum_G I(z_n, t; y_{t+\tau} | y_t^k, z^{<n}), \quad (4.3.4)$$

其中  $z^{<n} \in Z$  中的  $z^{<n}$  是额外的源变量集，用于检测来自  $X$  的冗余变量： $z^{<n} =$

$\{Z_c | \forall c : 1 \leq c \leq n\}$ 。受 [?] 的启发，我们提供了一种贪婪算法，根据 CMI 对候选变量  $C$  的贡献从  $X$  中选择  $Z$ 。算法如算法 1 所示：

---

**算法 2** 基于贪心算法的有效传递熵 (ETE) 的计算

---

**输入：**多元源时间序列  $X$ , 目标时间序列  $Y$ , 相关源序列集  $Z$  和变量历史数据候选集  $C_X$  and  $C_Y$ .

**输出：**多元因果关系集  $E$ .

```

1: Initialise  $Z = \emptyset$ ,  $E = \emptyset$ 
2: for  $c \in C_X$  do
3:    $T_{c \rightarrow y_{t+\tau}} = I(c, y_t | Z)$ ;
4:   if  $T_{c \rightarrow y_{t+\tau}}$  meets Maximum Statistics then
5:     add  $c$  to  $Z$  and remove it from  $C_Y$ ;
6:   else
7:     break;
8:   end if
9: end for
10: for  $c' \in C_Y$  do
11:    $T_{c' \rightarrow y_{t+\tau}} = I(c', y_t | Z)$ ;
12:   if  $T_{c' \rightarrow y_{t+\tau}}$  meets Maximum Statistics then
13:     add  $c'$  to  $Z$  and remove it from  $C_X$ ;
14:   else
15:     break;
16:   end if
17: end for
18: for  $z \in Z$  do
19:   add  $z$  to  $Z'$ ;
20:    $T_{z \rightarrow y_{t+\tau}} = I(Z_X, y_t | Z_Y)$ ;
21:   if  $T_{z \rightarrow y_{t+\tau}}$  meets Minimum Statistics then
22:     remove  $z$  from  $Z$ 
23:   end if
24: end for
25:  $T_{Z_X \rightarrow y_{t+\tau}} = I(Z_X, y_t | Z_Y)$ ;
26: if  $T_{Z_X \rightarrow y_{t+\tau}}$  meets omnibus test then
27:   add  $Z_X$  to  $E$ 
28: end if
29: return  $E$ 

```

---

### 4.3.1 时序预测模型框架

在小波分解模块中，车辙及其影响因素的时间序列被分解为 3 层，其中最后一层的低频分量 (0-20Hz) 表示原始时间序列的趋势项，所有 3 层的高频分量 (20-30Hz、30-50Hz 和 50-100Hz) 表示原始时间序列的波动项。作为递归神经网络 (RNN) 的扩

展, LSTM能够捕捉时间序列的长期依赖关系, 而注意力机制的概念在近几年的深度学习, 特别是神经网络领域的研究中得到了普及, 它能有效捕捉输入数据不同部分之间的依赖关系, 这在许多任务中都至关重要。( [?]), 因此也被用于时间序列预测。因此, 我们设计了一种基于融合注意力机制和LSTM的自编码器为趋势项的预测模型。此外, 为了克服 RNN 模型在面对时间序列高频波动时的不敏感性, 本文将基于统计理论的复杂时间序列高维非线性机器学习模型 GPR 引入到所提出的框架中, 以预测波动项。本节的其余部分将介绍这两个模型的机制。

### 4.3.2 高斯随机过程

高斯过程是随机变量的集合, 其中任意有限个变量都具有联合高斯分布。给定一组点  $X = \{x_1, x_2, \dots, x_t\}$ , 其值  $f(X)$  完全由其均值函数  $m(X)$  和协方差函数  $k(X, X')$  指定, 它们分别定义为:

$$\begin{aligned} m(X) &= \mathbb{E}[f(X)], \\ k(X, X') &= \mathbb{E}[(f(X) - m(X))(f(X') - m(X'))], \end{aligned} \quad (4.3.5)$$

其中,  $f(X)$  是高斯过程, 可写成:

$$f(X) \sim \mathcal{GP}(m(X), k(X, X')). \quad (4.3.6)$$

用  $y$  表示观测值, 它与真实值  $f(X)$  之间存在加性噪声。我们假设噪声服从独立、同分布的高斯分布, 均值为零, 方差为  $\sigma^2$ :

$$\varepsilon \sim \mathcal{N}(0, \sigma^2).$$

根据  $f(X)$  和先验分布, 可以推断出在一组点  $X$  上产生的真值的期望值和方差。具体来说, 要根据已知观测值和先验值估计真值  $f(X)$ , 我们需要联合分布:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}' \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu}' \end{bmatrix}, \begin{bmatrix} \mathbf{K}(X, X) + \sigma_n^2 I & \mathbf{K}(X', X) \\ \mathbf{K}(X, X') & \mathbf{K}(X', X') \end{bmatrix} \right), \quad (4.3.7)$$

以及观测数据的联合高斯先验分布

$$P(f' | y, X, X') = \mathcal{N}(m, C). \quad (4.3.8)$$

有一些常用的协方差矩阵函数  $K$ , 通过对数似然框架中的优化, 可以自适应地获得最佳超参数。函数值  $f'$  (对应于测试输入  $X'$ ) 可以通过评估点  $X$  的估计均值  $m$  和

估计方差  $C$  从联合后验分布中采样:

$$\begin{aligned} \mathbf{m} &= \boldsymbol{\mu}' + \mathbf{K}(X, X') (\mathbf{K}(X, X) + \sigma_n^2 I)^{-1} (\mathbf{y} - \boldsymbol{\mu}) \\ \mathbf{C} &= \mathbf{K}(X', X') - \mathbf{K}(X, X') (\mathbf{K}(X, X) + \sigma_n^2 I)^{-1} \mathbf{K}(X', X) \end{aligned} \quad (4.3.9)$$

### 4.3.3 基于融合注意力机制的自编码器

本节所提出的时间序列预测模型利用了长短期记忆和融合注意力机制，是一个端到端的深度学习框架，它将编码上下文向量和注意力向量相互融合，从而进行时间序列表示学习。模型架构见图4.2。模型的流程如下：在经过特征提取后的序列  $X_n = \{x_1, x_2, \dots, x_n\}$  被当模型输入送入编码器模块，首先进入LSTM模型，生成隐向量  $h_t$ ，作为注意力层的输入。注意力层包括两种注意力机制：自注意力（self-attention）和时间注意力（temporal attention），图中  $h'_t$  表示自注意力层在时间步长为  $t$  时的输出向量，其维度为  $l$ ，滑动窗口大小为  $t$ 。在时间注意力层中有  $t$  个 1 维 CNN 滤波器  $C_i$ , ( $i \in l$ )，其大小为  $t \times 1$ 。通过滤波器对  $h'_t$  进行卷积，产生一组行向量  $H^C$ ，其大小为  $k \times l$ 。接下来，关注度权重  $\alpha_i$ , ( $i \in l$ )，即行向量中元素相关性，由评分函数  $f : \mathbb{R}^{k \times l} \times \mathbb{R}^l \mapsto \mathbb{R}$ 。最后，编码器的输出向量  $\tilde{h}_t$  是由  $h'_t$  和原始向量  $h_t$  连接生成的。在本节的剩下部分，我们将详细介绍模型内所涉及到的模块。

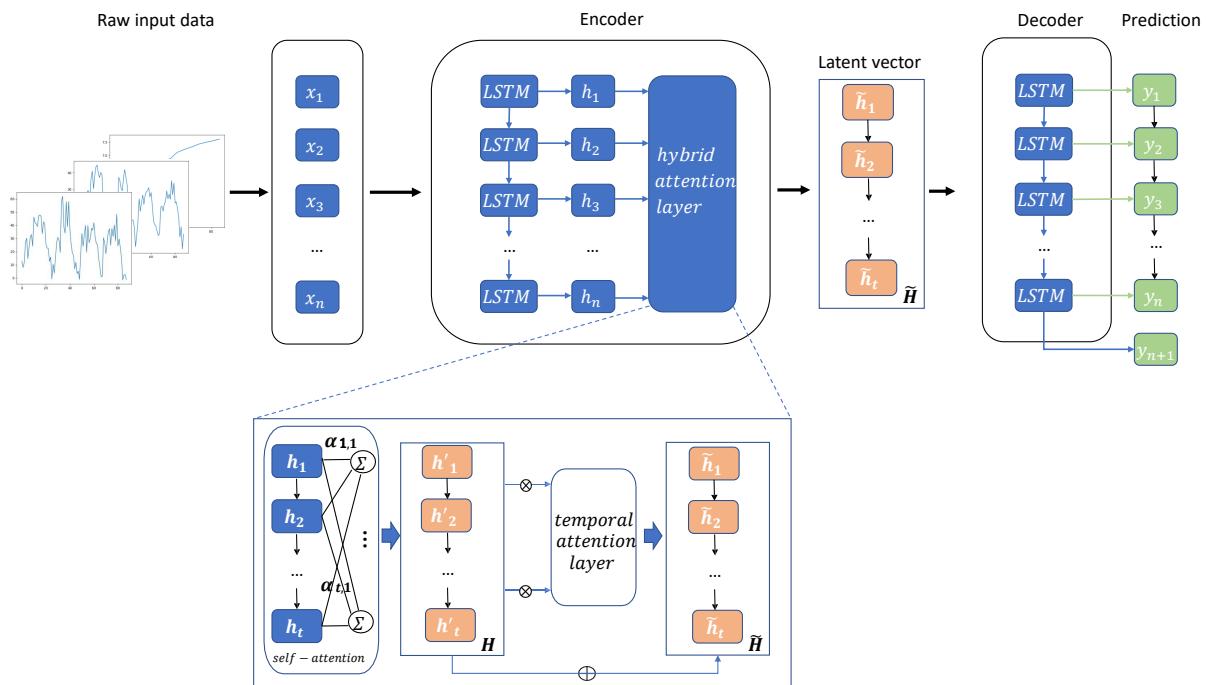


图 4.2: 基于融合注意力机制的自编码器

**长短记忆网络** 长短记忆网络 (long short term memory, LSTM) 是 Hochreiter 等人提出的 RNN 变体 [?]. LSTM 继承了循环神经网络 (recurrent neural network, RNN) 对时间序列数据的上下文信息进行动态建模的能力, 还能处理不同长度和维度的变量序列。通过引入门机制, LSTM 克服了 RNN 在长期序列依赖问题和梯度消失问题。为了解决这些问题, LSTM 使用单元状态  $c_t$  中存储长期信息, 并在隐藏状态  $h_t$  来处理当前信息:

$$h_t = o_t \odot \tanh(c_t), \quad (4.3.10)$$

$$o_t = \delta(W_{x_o}x_t + W_{h_o}h_{t-1}), \quad (4.3.11)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_{x_g}x_t + W_{h_g}h_{t-1}), \quad (4.3.12)$$

$$i_t = \delta(W_{x_i}x_t + W_{h_i}h_{t-1}), \quad (4.3.13)$$

$$f_t = \delta(W_{x_f}x_t + W_{h_f}h_{t-1}), \quad (4.3.14)$$

其中,  $i_t$ ,  $f_t$ ,  $o_t$  分别是输入门、遗忘门和输出门,  $W_{x_i}, W_{x_f}, W_{x_o}, W_{x_g} \in \mathbb{R}^{m \times i}$  是每个门控的可学习参数矩阵。编码器中的 LSTM 层用于将多变量输入编码为隐藏状态  $H_t = \{h_1, h_2, \dots, h_t\}$ , 这是一个固定大小的向量, 作为混合注意力层的输入。而输出的大小  $Y_t = \{y_1, y_2, \dots, y_t\}$  由解码器中的 LSTM 生成, 长度可以按需指定。

**注意力机制** 注意力机制模拟了认知心理学中引入的人脑注意力。注意力机制的发展也引领了长期时间序列依赖性学习。本节引入了两种注意力机制, 首先介绍自注意力机制, 自我注意力的提出是为了克服序列生成的变换器架构中的一个问题, 即固定大小的嵌入引起的信息损失 [?]. 注意力层将每个时间步的时间信息与动态生成的注意力值聚合在一起, 这样模型就能锁定重要的时间步, 并消除长期序列中的无用信息。从概念上讲, 注意力是一种基于给定查询的键值查找机制, 可表述为:

$$h_t = \sum_{\tau=0}^T \alpha(q_\tau, k_t) v_{t-\tau}, \quad (4.3.15)$$

$$\alpha(q_\tau, k_t) = \exp\left(\frac{q_\tau \cdot k_t^T}{\sqrt{d_k}}\right) / \sum_{\tau=0}^T \exp\left(\frac{q_\tau \cdot k_t^T}{\sqrt{d_k}}\right), \quad (4.3.16)$$

其中,  $H_t$  是自注意力机制的输出;  $Q_\tau$ 、 $K_t$  和  $v_{t-\tau}$  是输入线性变换的中间矩阵,  $Q \Delta K$  和  $V$  在不同时间步的特征向量表示;  $\alpha$  是可学习的自注意系数。自注意力机制的第一步是对输入特征  $x_\tau$  和输出状态  $h_t$  进行相似度测量。相似度越高, 输入的注意力就越

强。上式中的相似度运算是向量点积，也可以用余弦相似度或输入  $h_t$  和输出  $\alpha_t$  的多层感知器（MLP）来代替。第二步是加权平均算子，用于消除量纲影响 [?].

自注意力主要集中在描述每个连续隐藏向量元素在时间维度的重要性上，这种设计忽略了行元素对输出包含信息的影响，尤其是在多变量任务中，它无法检测到有用的时态模式。而隐向量的行元素之间的关系则可以时间注意力来捕捉，时间注意力的具体架构如图4.3 所示。给定回视窗口  $t$  中的 RNN 单元隐藏状态矩阵  $H_{t-1} \in \mathbb{R}^{m \times t}$ ，时态模式关注机制关注用  $K$  个 1 维的 CNN 过滤器  $C_i \in \mathbb{R}^{t \times 1}$  操作每个向量  $h$  中的行元素，所有向量得到的结果组成矩阵  $H^C \in \mathbb{R}^{l \times k}$ 。最后，使用评分函数  $f : \mathbb{R}^{l \times k} \mapsto \mathbb{R}$  来计算矩阵  $H^C$  和隐藏状态  $h_t$  在时间步长  $t$  中的相关性，因此时间注意力可以表述为：

$$\hat{h}_t = W_h h'_t + W_h h_t, \quad (4.3.17)$$

$$h'_t = \sum_{i=1}^n \alpha_i \tilde{H}_i^C, \quad (4.3.18)$$

$$\alpha_i = \frac{\exp(f(\tilde{H}_i^C, h_t))}{\sum_{i=1}^n \exp(f(\tilde{H}_i^C, h_t))}, \quad (4.3.19)$$

$$f(\tilde{H}_i^C, h_t) = (\tilde{H}_i^C)^\top W_a h_t, \quad (4.3.20)$$

$$\tilde{H}_{i,j}^C = \sum_{j=1}^K r_j \times C_i, \quad (4.3.21)$$

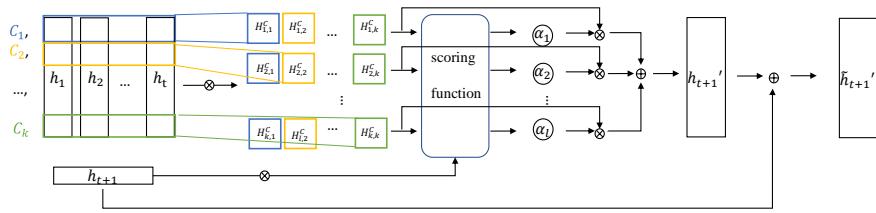


图 4.3: 时间模式注意力原理图。

其中， $h'_t$  是时间注意力机制的输出矩阵， $h_t$  是上一步LSTM的输出矩阵， $W_h', W_h$  分别是对应的参数矩阵； $\alpha_i$  是  $H_t$  中  $i$  行元素的注意力权重，由式 (4.3.19) 中的评分函数  $f$  计算得出；式 (4.3.20) 中的  $\tilde{H}^C$  表示卷积运算的权重； $r_j$  是 RNN 单元的输出矩阵， $W_h', W_h$  分别是参数矩阵。评分矩阵中每行的权重可以代表行元素与输出  $h'_t$  之间的相关性。

表 4.1: 小波基函数的选择

小波基函数	公式	MSE	GPV	
			$R^{h1}$	$R^{h2}$
db4	369.8	5.3	3.8	14
db20	369.7	5.6	3.3	10
Symlets4	370.6	5.5	3.3	11
Symlets20	369.8	5.1	3.7	9
Coiflets21	369.6	5.3	3.6	10
Coiflets22	370.0	5.3	3.8	9

## 4.4 实例结果及讨论

为了验证本章所提出的时间序列预测框架的有效性，采用RIOHTrack数据集作为实例并设计验证实验。实例结果包括三个部分：时间序列分解模块中小波基函数的选择；特征选择模块中有效传递熵的结果验证；以及预测模型的性能分析及对比。

### 4.4.1 小波基函数的选择

车辙序列进行小波变换的目的是提高车辙数据的稳定性。与上一章选择小波基函数的原则相同，主要是根据小波变换后数据的平滑度，具体可量化为以下几个方面：正交性、支持集、规则性和消失矩的阶数。根据上述原则，我们继续从Haar小波、Daubechies小波、Symlets小波，和Coiflets小波中进行选择，每个小波的最优滤波系数由第二章所提出的自适应算法求得。我们给出两个指标来衡量小波函数的分解性能：1.时间序列的小波重构误差，即重构序列 $X'$ 与原序列 $X$ 的均方误差（mean square error, MSE）： $MSE(X, X')$ ；2.高斯过程验证（Gaussian process validation, GPV），即验证采样时间点中不符合高斯随机过程的点数。此外，对于分解层数，我们继续按照第三章中的经验，设置分解层数为2层。验证结果如表4.1所示：

对于小波基函数筛选的两个指标， $MSE(X, X')$ 的值越小，证明小波分解后的信息损失的越少，小波基函数的性能越好；GPV的值越小，证明分解后的序列更符合高斯分布，越有助于提升GPR预测的性能。同时，考虑到消失矩越大，所需要的计算代价越高。从表4.1中的结果可知，消失矩越高的的小波基函数，其对应的MSE越小，其中，又以Symlet20小波和Coiflets20小波的MSE最小；同时，无论是第一层高频率列的GPV的Symlets20的GPV。综合以上因素，我们选择Symlets20作为时间序列分解模块的小波基。

#### 4.4.2 基于传递熵的因果关系结果验证

我们还比较并显示了每个变量在不同频率下的情况，如图 7 所示。然而，变量之间的相关性并不能直观地发现。为了定量测量不同频段车辙及其影响因素之间的偶然性，本文采用了贪心算法的显著TE计算方法，其中源变量和目标变量候选集的长度均为10。在对所有19种路面数据集进行影响因素与车辙之间的偶然性分析后，结果如表4.2, 所示，括号内表示转移熵的P值。高亮显示的数据表示该数据具有统计意义。根据公式，MTE 的核心思想是：如果一个变量引起另一个变量，那么源变量的历史数据的分布应该为目标变量的未来值提供额外的转移信息，而不是仅仅从目标变量的过去值就能预测到的。因此，从源变量到目标变量的转移熵值越高，因果关系就越强。表4.2中的结果表明，轴载与车辙的因果关系在低频段（L1）统计显著，这表明轴载主要影响车辙的变化趋势。同时，从温度到车辙的传递熵在统计意义上是显著的。这也表明，在趋势项中，全轮驱动挠度盆面积与车辙的因果关系最强，而在波动项中，温度（H1 层）和全轮驱动中心点挠度与车辙的因果关系最强。} 特别地，在所有频段中，我们都可以观察到车辙与全轮驱动深度和全轮驱动面积之间存在着非常显著的双向信息传递。因此，低频段预测模型选择轴载、全断面深度和全断面面积作为特征，高频段预测模型选择温度、全断面深度和全断面面积作为特征。此外，虽然不同路面的因果关系值不同，但所有路面的偶然关系是相同的。这表明多元转移熵可以在车辙及其影响因素之间建立稳定的因果关系。此外，本章所提出的基于贪心算法的显著TE计算方法，进一步的明确了源变量和目标变量的有效Markov阶数，确定了用于计算的历史长度，可以看到，由本章所提出的方法所计算出的统计显著的因果关系，除了TE(load → rutting)，目标变量和源变量的Markov阶数不是相同的。具体的，以温度和荷载为源变量的因果关系，其源变量的Markov阶数较小，说明其数值变化对车辙的影响较快。以弯沉深度和弯沉面积为源变量的因果关系，其源变量

表 4.2: 基于转移熵的车辙及其影响因素之间的因果关系，括号中为  $p$  值， $\rightarrow$  左侧为源变量及其Markov阶数 $l$ ， $\rightarrow$  左侧为目标变量及其Markov阶数 $k$ 。

因果关系	第一次低频分量	第一层高频分量	第二层高频分量
温度(3) → 车辙(1)	0.0013 (0.8710)	<b>0.0792 (0.0040)</b>	<b>0.0812 (0.0021)</b>
荷载(1) → 车辙(1)	<b>0.1190 (0.0003)</b>	0.0036 (0.9310)	0.0049 (0.6711)
弯沉深度(5) → 车辙(3)	<b>0.0943 (0.0000)</b>	<b>0.0596 (0.0035)</b>	<b>0.0889 (0.0014)</b>
弯沉面积(5) → 车辙(2)	<b>0.1251 (0.0002)</b>	<b>0.0341 (0.0079)</b>	<b>0.0666 (0.0061)</b>
车辙(3) → 弯沉深度(1)	0.0091 (0.9130)	0.0013 (0.9005)	0.0018 (0.9116)
车辙(3) → 弯沉面积(1)	<b>0.0875 (0.0010)</b>	0.0037 (0.9375)	0.0017 (0.3671)

的Markov阶数较大，说明其数值变化对车辙的影响更久。在之后的时间序列预测模块中，这些Markov阶数将作为模型自变量的历史滑动窗口尺寸进行使用。

#### 4.4.3 预测模型的性能分析

为了验证所提出的框架在利用 RIOHTrack 数据集进行车辙预测方面的性能，将其与以下基线时间序列预测模型进行了比较：统计回归模型，包括 GPR 模型和 VARIMA；两个基于深度学习的时间序列预测模型，这两个模型为自编码器，其中的编码器分别由 transformer 模型和双向 LSTM 构成，其中滑动窗口大小为 5，层数为 2；transformer 模型中的多头注意数为 3，注意力机制为自注意力；解码器中则是一个线性回归模型。所有模型都进行了 1000 次训练。每个路面类别选择一个路面数据作为训练集，其余路面数据作为测试集。在训练集中，80% 用于训练，20% 用于验证。

我们使用三个指标来衡量模型的预测性能：平均绝对百分比误差 (MAE)、均方根误差 (RMSE) 和判定系数  $R^2$ ，它们的定义分别是：

$$MAPE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|, \quad (4.4.1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}, \quad (4.4.2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}, \quad (4.4.3)$$

其中， $y_t$  为时间步长  $t$  时的车辙地面实况， $\hat{y}$  为预测值， $\bar{y}$  为车辙时间序列的真实值。对于 MAPR 和 RMSE，数值越小表示性能越好。 $R^2$  的取值范围为  $[0, 1]$ ，该值越接近 1，模型的预测性能越好。 $R^2$  的值小于 0 意味着预测值的性能比平均值的性能要差。

在性能对比实验中，我们在来自 RIOHTrack 的 19 种路面车辙数据上比较了所提出的框架和其他先进的时间序列预测模型。预测结果的步长分别为 2、4、8 和 16 步。按照第二章中的聚类结果，我们将路面分为五个簇，结果如表 3、表 4 和表 5 所示。其中，每个路面数据的最佳性能以粗体标出。总体而言，在大多数路面类别的车辙数据上，所提出的框架显然比其他基线模型取得了更好、更稳健的性能。在硬基路面 (STR4-5) 上的两步预测任务的预测性能最好，其指标为  $MAE = 0.97$ 、 $RMSE = 1.15$ 、 $R^2 = 0.92$ 。单向 LSTM 在 STR18-19 的两步预测任务中取得了最佳结果。此外，我们还发现传统的统计方法—多元 GPR 在所有指标上都表现最差。这

是因为该模型对预测目标有很强的假设。GPR 和 VARIMA 的 MAE 和 RMSE 值较大，因为这些模型会忽略一些跳跃点。结果中另一个有趣的现象是：基于深度学习的模型（LSTM、GRU、转换器和所提出的框架）普遍优于浅层模型，而所提出的框架在小波分解和偶然性推理的帮助下，对于泛化数据具有更强的鲁棒性。在合同中，基于注意力机制的模型，如转换器，可以学习较长时间序列的趋势。而在小波分解和偶然性推理的帮助下，拟议框架中基于 RNN 的模型也能保证长期预测的性能。从预测步长方面来看，我们发现预测性能随着预测步长的增加而降低，这一现象在基于 RNN 的模型中尤为明显。拟议框架在 16 个水平线上的平均性能为  $MAE = 2.07 \pm 0.15$ 、 $RMSE = 3.09 \pm 1.14$ 、 $R^2 = 0.72 \pm 0.20$ ，这表明所提出的框架在较长的预测步长上具有鲁棒性。从不同的路面数据来看，随着数据的变化，所提框架的性能变化不大。建议框架在所有水平面上的平均性能为  $MAE = 1.79 \pm 1.07$ 、 $RMSE = 2.85 \pm 1.68$ 、 $R^2 = 0.77 \pm 0.15$ ，结果表明所提框架对数据变化的鲁棒性。

表 4.3: 在STR1-9的RIOHTrack数据集上，本章所提框架与其他基线模型的平均性能比较

路面类别		semi-rigid base (STR1-3)				rigid base (STR4-5)				common forms (STR6-9)			
		距离步长 2	距离步长 4	距离步长 8	距离步长 16	距离步长 2	距离步长 4	距离步长 8	距离步长 16	距离步长 2	距离步长 4	距离步长 8	距离步长 16
模型	度量												
	<i>MAE</i>	10.11	13.41	15.36	16.91	9.31	10.33	13.72	15.11	12.87	16.14	18.33	20.26
	<i>RMSE</i>	11.58	14.79	16.26	18.05	9.89	11.02	14.14	16.26	13.97	18.33	20.12	22.37
GPR	<i>R</i> <sup>2</sup>	0.43	0.32	0.31	0.21	0.43	0.39	0.34	0.29	0.35	0.31	0.24	0.12
	<i>MAE</i>	9.75	15.12	18.26	19.97	8.33	16.38	17.99	20.37	10.33	14.62	23.15	25.71
	<i>RMSE</i>	11.31	17.81	21.89	22.31	12.33	18.57	20.12	22.37	17.97	18.33	23.15	27.15
VARIMA	<i>R</i> <sup>2</sup>	0.41	0.32	0.15	0.03	0.64	0.41	0.25	0.10	0.62	0.37	0.19	0.13
	<i>MAE</i>	1.61	1.74	3.53	4.97	1.24	1.63	2.59	3.59	2.33	4.09	5.93	8.39
	<i>RMSE</i>	1.83	2.27	5.59	6.97	1.47	1.94	2.73	5.31	2.74	6.59	6.06	9.16
LSTM	<i>R</i> <sup>2</sup>	0.80	0.74	0.68	0.53	0.79	0.73	0.69	0.50	0.64	0.55	0.42	0.26
	<i>MAE</i>	1.57	1.71	3.97	4.31	1.33	1.52	2.76	3.85	2.73	3.12	6.26	8.15
	<i>RMSE</i>	1.80	2.33	6.22	7.73	1.83	1.92	2.37	6.97	3.33	3.64	7.53	8.93
GRU	<i>R</i> <sup>2</sup>	0.83	0.71	0.66	0.54	0.83	0.80	0.69	0.51	0.70	0.61	0.51	0.36
	<i>MAE</i>	1.59	1.67	1.71	1.93	1.12	1.56	1.97	2.33	2.12	3.26	3.57	4.14
	<i>RMSE</i>	1.83	2.12	2.26	2.57	1.31	1.93	2.54	3.31	3.09	3.74	4.15	4.38
transformer	<i>R</i> <sup>2</sup>	0.82	0.81	0.79	0.71	0.83	0.80	0.76	0.71	0.69	0.65	0.59	0.41
	<i>MAE</i>	1.22	1.23	1.51	1.67	0.97	1.17	1.24	1.46	1.25	1.53	1.74	1.96
	<i>RMSE</i>	1.37	1.67	1.75	1.93	1.15	1.33	1.74	1.97	1.55	1.93	1.94	2.13
proposed framework	<i>R</i> <sup>2</sup>	0.87	0.82	0.79	0.71	0.92	0.87	0.83	0.80	0.86	0.81	0.77	0.72

由于提议的框架由三个模块组成，为了研究每个模型的贡献，我们设计了一组消融实验。消融实验中包含6个变体框架，变体 1、2 和 3 是针对数据处理模块的变体。变体 4 和 5 针对基于转移熵的特征提取模块的变体。变体 6 使用单一预测模型来预测车辙。这些变体模型的具体设置如下：

- 在变体 1 中，删除了基于小波分解的数据处理模块。框架中只使用了基于传递熵的特征提取模块和单一时间序列预测模块。
- 在变体 2 中，结构与本章所提出框架相同，但特征提取模块中的小波函数由 dB20 代替。

表 4.4: 在STR10-17的RIOHTrack数据集上, 本章所提框架与其他基线模型的平均性能比较

pavement	semi-rigid base (STR10& 12)				rigid base (STR11 & 13-14)				common forms (STR15-17)							
	距离步长 度量	2	4	8	16	距离步长 度量	2	4	8	16	距离步长 度量	2	4	8	16	
GPR	MAE	6.34	10.15	13.64	20.51	9.31	13.33	19.44	23.89	10.81	14.14	17.95	21.73			
	RMSE	9.74	12.59	15.05	26.35	9.89	15.91	21.07	26.66	13.98	16.27	19.99	23.17			
	R <sup>2</sup>	0.39	0.31	0.27	0.22	0.35	0.26	0.17	0.09	0.40	0.35	0.23	0.10			
VARIMA	MAE	10.09	12.35	20.03	27.93	6.15	7.34	16.87	22.59	10.33	12.61	17.15	23.33			
	RMSE	11.31	17.81	21.89	22.31	7.71	9.15	20.42	29.13	12.22	15.85	20.15	26.92			
	R <sup>2</sup>	0.41	0.27	0.08	0.02	0.48	0.46	0.29	0.22	0.54	0.33	0.21	0.15			
LSTM	MAE	1.43	1.70	3.53	4.97	1.24	1.63	2.59	3.59	2.67	4.15	6.35	8.88			
	RMSE	1.65	1.92	5.59	6.97	1.47	1.94	2.73	5.31	3.03	5.66	8.10	10.33			
	R <sup>2</sup>	0.90	0.81	0.74	0.62	0.83	0.77	0.61	0.42	0.68	0.60	0.58	0.32			
GRU	MAE	1.61	2.79	4.18	4.98	1.41	1.55	2.63	3.77	2.73	4.45	7.14	9.66			
	RMSE	2.01	2.93	6.21	7.14	1.69	1.70	2.81	5.99	3.33	6.66	10.15	12.15			
	R <sup>2</sup>	0.73	0.70	0.65	0.63	0.81	0.75	0.62	0.45	0.68	0.57	0.52	0.34			
transformer	MAE	1.61	1.66	1.75	2.44	2.13	2.73	2.97	3.28	1.92	2.08	2.61	2.88			
	RMSE	1.64	1.74	2.26	3.05	2.54	1.93	2.54	3.31	2.09	2.74	3.33	4.19			
	R <sup>2</sup>	0.79	0.71	0.66	0.43	0.74	0.70	0.67	0.54	0.74	0.66	0.61	0.50			
proposed framework	MAE	1.21	1.27	1.74	2.01	1.70	1.78	1.99	2.35	1.68	1.73	1.97	2.70			
	RMSE	1.67	1.81	2.50	2.77	2.67	2.31	2.50	4.17	1.72	2.47	3.13	4.60			
	R <sup>2</sup>	0.90	0.85	0.81	0.76	0.80	0.78	0.73	0.67	0.78	0.72	0.70	0.66			

表 4.5: 在STR18-19的RIOHTrack数据集上, 本章所提框架与其他基线模型的平均性能比较

pavement	semi-rigid base (STR18-19)				
	预测步长	1	3	6	10
GPR	MAE	5.96	6.12	16.28	28.12
	RMSE	7.39	7.68	18.29	30.17
	R <sup>2</sup>	0.53	0.37	0.26	0.02
VARIMA	MAE	8.85	10.12	18.58	21.97
	RMSE	9.89	13.31	20.02	28.97
	R <sup>2</sup>	0.41	0.30	0.24	0.19
LSTM	MAE	<b>1.33</b>	1.92	4.16	5.35
	RMSE	<b>1.61</b>	2.37	5.07	7.33
	R <sup>2</sup>	<b>0.88</b>	0.72	0.66	0.49
GRU	MAE	1.51	1.75	2.03	4.21
	RMSE	1.73	2.12	2.25	6.03
	R <sup>2</sup>	0.81	0.73	0.68	0.59
transformer	MAE	1.42	1.63	1.93	2.33
	RMSE	1.77	1.81	2.17	2.47
	R <sup>2</sup>	0.80	0.76	0.70	0.67
proposed framework	MAE	1.35	<b>1.73</b>	<b>2.01</b>	<b>2.31</b>
	RMSE	1.71	<b>2.05</b>	<b>2.67</b>	<b>4.03</b>
	R <sup>2</sup>	0.87	<b>0.81</b>	<b>0.74</b>	<b>0.72</b>

- 在变体 3 中，架构与本章所提出框架相同，但特征提取模块中的小波函数由 Coefit20 替代。
- 在变体 4 中，删除了基于传递熵的特征提取模块。框架中只使用了基于小波分解的数据处理模块和时间序列预测模块。
- 在变体 5 中，其结构与建议的框架相同，其中特征提取模块中传递熵的计算不使用本章所提出的贪心算法。
- 在变体 6 中，其结构与拟议框架相同，但时间序列预测模型只使用基于注意力机制的自编码器的单一预测模型。

我们对所有路面数据进行了消融实验，平均性能对比结果如图4.4和图4.5所示。总体而言，结果表明所提出的框架能比其他变体更精确地模拟车辙曲线，这表明所有模块缺一不可。特别是，从图4.4的结果来看，变体1 的对比结果表明，基于小波分解的数据处理模块的使用大大提高了性能，使各预测步长上的 $R^2$ 指标分别提高了16.23%、18.53%、24.35%和36.66%。同时，变体2 和变体3 的 GPV 验证值分别为 10、12 和 10、10，比基于Symlet20小波基函数的值要差。变体2和变体3的对比也表明，基于Symlet20的时间序列分解模块对车辙数据的定位更准确，对多步骤预测的鲁棒性更高。在图4.5 中，本章所提出框架与变体 4 的性能差异体现了基于传递熵的特征选择模块的必要性。而和变体 5 的性能差异显示了基于贪心算法所计算的统计显著转移熵在特征选择模块的优势。我们还注意到，单预测模型的变体 6在预测步长较小时（horizon=2, 4）性能接近本章所提出框架，但随着预测步长的增加，其性能会迅速下降，这表明变体 3 提取的特征中产生了冗余信息和虚假因果关系。此外，与变体 4 的比较表明，超时间序列预测模型更适合 RIOHTrack 的车辙数据。

此外，在低频部分的基于混合注意力的预测模型中，混合注意力层由两部分组成：自我注意力层和时间注意力层。在这里，我们将进行进一步的实验和分析，以验证这两个部分的有效性。

自我注意层旨在增强模型捕捉长距离依赖关系的能力。如上文不同滑动窗口大小的实验结果所示，带有注意力机制的模型可以获得每个隐藏状态之间的全局相关性。注意分值的计算是自注意机制的核心步骤，为了进一步验证自注意层的有效性，我们对查询向量  $q$  和关键向量  $k$  之间的注意分值的不同计算方法的模型性能进行了研究。表4.6总结了不同技术水平的注意力向量的实验结果，包括余弦相似度（cosine similar-

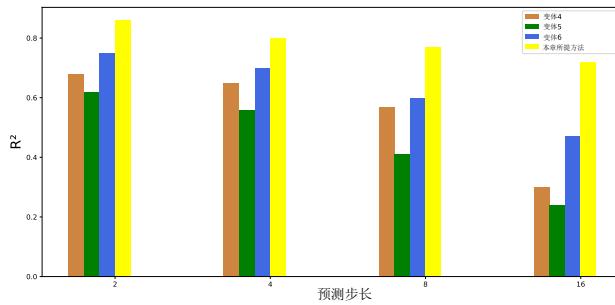


图 4.4: 变体 (1-3) 与本章所提出框架的消融实验性能比较。

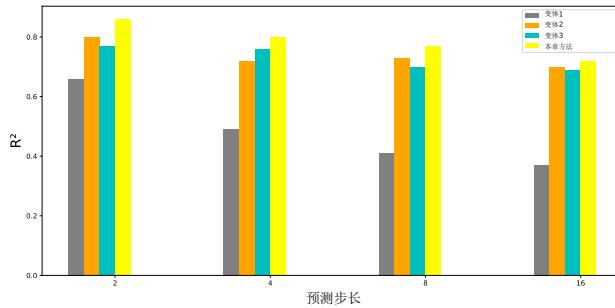


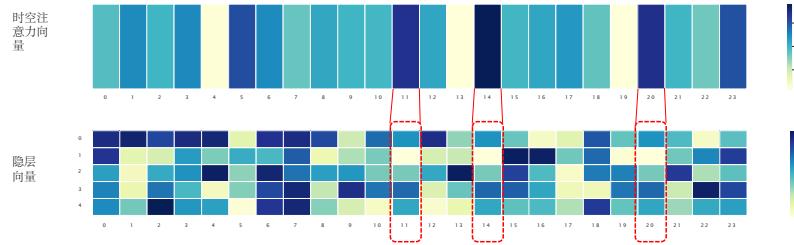
图 4.5: 变体 (4-6) 与本章所提出框架的消融实验性能比较。

ity) [?], MLP [?], 矩阵串联 (matrix concatenation) 和带权重的点积 (dot product with weights)，本文提出的模型采用了带权重的点积 (dot product)。表4.6显示，本文采用的注意力向量以较少的计算量获得了最低的平均 MAE 和 RMSE。

最后，为了更好地理解时态注意层的有效性，我们在 STR1 路面数据集的实验中通过可视化图形显示对时态注意权重向量进行了分析。时间注意力向量的热图如图4.6所示，热图中的每个网格代表注意力值，颜色越深，表示注意力值越高。注意力向量为  $\alpha \in \mathbb{R}^l$ ，其中  $l$  等于前 LSTM 层生成的隐藏状态的维度，例如本实验中为  $l = 24$ 。注意向量的值是每一行元素的相关性度量。参数越接近 0，相应位置特征对预测的贡献就越小。以在 STR1 数据上训练的时间模式注意力值为例，我们选取三个最高值，并追溯到隐藏状态  $h_t \in \mathbb{R}^l$  在时间步长  $t$  的相应位置上，滑动窗口大小为 5。这三个值对应的隐藏状态行元素分别为  $[0.1129, -0.4297, -0.1249, 0.2026, 0.04882]$ 、 $[0.1125, -0.4264, -0.1050, 0.2278, 0.0524]$  和  $[0.1180, -0.4804, -0.1140, 0.2154, 0.0553]$ 。显然，这三个行元素的分布是相似的。也就是说，这些位置上隐藏状态的行元素在与下一个隐藏状态  $h_t$  连接时会做出更多贡献。

表 4.6: 采用不同注意力权重计算方法的自我注意力层的性能

注意力机制	对应公式	MAE	RMSE
dot product	$\alpha = \exp(\frac{q \cdot k^T}{\sqrt{d_k}}) / \sum_{\tau=0}^T \exp(\frac{Q_\tau \cdot K_\tau^T}{\sqrt{d_k}})$	0.55	0.81
cosine similarity	$\alpha = \frac{q \cdot k^T}{\ q\  \cdot \ k\ }$	0.64	1.23
concatenation	$\alpha = W[q; k]$	0.73	1.47
MLP	$\alpha = \delta(W_q \cdot q + W_k \cdot k)$	0.57	0.85

图 4.6: 时间注意力向量  $\alpha \in \mathbb{R}^{24 \times l}$  和 5 个滑动窗口中的隐藏状态  $h \in \mathbb{R}^{5 \times 24}$  可视化。前3个注意力值分别为 0.6396、0.6809 和 0.6421。

## 4.5 小结

本章分析了分布式饱和脉冲控制下时滞复杂网络的局部动力学行为。首先，利用反证法、脉冲系统的比较原理和平均脉冲区间的方法研究了分布式饱和脉冲控制下具有耦合时滞的鲁里叶网络的局部指数同步问题，并给出了不依赖于时滞的基于 BMIs 的充分性判据。为了降低保守性，选取最新的具有更多松弛变量的改进凸包表示法来处理分布式饱和的脉冲项，并且开发了一种与收缩不变集完全不同的估计吸引域的方法。其次，通过构造依赖于脉冲时刻的复合型 Lyapunov 函数进一步降低保守性，讨论了分布式饱和脉冲控制下具有切换拓扑的非线性时滞多智能体系统的局部一致性，并建立了相应的局部一致性判据。为了估计出最大的吸引域，通过适当的矩阵变换，建立了基于 LMIs 的优化问题，并通过 Matlab 软件中的 Yalmip 工具箱求解相应的最大吸引域的数值解。

本章部分结果已发表在国际期刊 IEEE Transactions on Neural Networks and Learning Systems 上，部分结果已投稿至国际期刊 IEEE Transactions on Automatic Control。具体详见作者发表论文的清单。

# 第五章 总结与展望

## 5.1 总结

本文主要在 Lyapunov 稳定性理论框架下，研究了具有脉冲控制的时滞复杂网络的动力学行为。通过设计不同类型的脉冲控制，包括混杂脉冲控制，事件触发脉冲控制，分布式饱和脉冲控制等，利用时滞系统理论、脉冲控制理论以及 LMIs 等方法和工具，分别对具有不同类型时滞的复杂网络的全局动力学以及局部动力学进行了全面分析。所得结果完善了部分已有的结论，补充和丰富了脉冲控制器的设计方案以及时滞复杂网络动力学方面的理论成果，具体工作总结如下：

1. 研究了在混杂脉冲控制下耦合多稳态时滞忆阻神经网络的动态多同步以及静态多同步问题，其中，混杂脉冲控制包括时滞脉冲控制和连续时间状态反馈控制。首先，基于状态空间划分方法和激励函数的几何特性，给出了系统具有多个局部指数稳定的周期轨道或平衡点的充分条件。其次，利用新型的 Halanay 微分不等式和脉冲控制理论，分别得到了不依赖于时滞的动态多同步和静态多同步的相关判据。值得注意的是耦合多稳态时滞忆阻神经网络中状态时变时滞可微的限制和脉冲控制中的时变时滞的上界小于任意脉冲区间的要求都被取消，进一步降低了结果的保守性。

2. 讨论了事件触发时滞脉冲控制下具有耦合常时滞的复杂网络的全局指数同步问题。一种新型的基于二次 Lyapunov 函数的事件触发时滞脉冲控制机制被设计，此机制是将时滞脉冲控制考虑到事件触发机制中。其中，时滞脉冲控制仅在触发时刻被施加，而在非触发的时滞脉冲时刻不再发挥作用，进一步节约了网络资源。通过构造合适的辅助函数，并利用递推法、脉冲控制理论以及 Lyapunov–Razumikhin 技术，得到了基于 LMIs 的不依赖于时滞的同步判据并有效地避免芝诺现象。值得一提的是 Lyapunov 函数在触发的时间间隔内严格单调递减的要求被取消，改进了现有的结果。

3. 探讨了事件触发牵制脉冲控制下具有耦合比例时滞的基因振荡器网络的集群同步问题。首先，针对具有比例时滞的脉冲系统，设计了事件触发脉冲控制协议，保证了脉冲系统的全局渐近稳定性。在此基础上，进一步将牵制控制考虑到所建立的事件触发脉冲控制机制中，建立起事件触发牵制脉冲控制，研究了具有耦合比例时滞的基因振荡器网络的集群同步，得到了相应的集群同步判据并且消除了芝诺行为。不同于已有的事件触发脉冲控制，通过设计合适的算法，每个集群中只有一小部分节点在触

发时刻被施加脉冲控制，进一步节约了网络资源。

4. 研究了分布式饱和脉冲控制下具有耦合时变时滞的鲁里叶网络的局部指数同步问题。利用反证法、脉冲系统的比较原理和平均脉冲区间的方法得到了基于 BMIs 的局部同步判据。为了降低保守性，选取最新的具有更多松弛变量的改进凸包表示法来处理分布式饱和脉冲项。目前，已有的关于饱和脉冲的结果基本上都是借助于收缩不变集来估计吸引域，这种方法要求 Lyapunov 函数在此集合中的任意非零点上是严格单调递减的。我们取消了这一限制，并开发了一种保守性更小的估计吸引域的方法。利用平均脉冲区间的方法处理脉冲信号，对脉冲间隔的上下界没有限制，这意味着允许存在非均匀分布的脉冲信号。

5. 分析了分布式饱和脉冲控制下具有切换拓扑的非线性时滞多智能体系统的局部一致性。利用反证法和 Lyapunov–Razumikhin 方法，得到了局部一致性判据。由于结构简单和计算方便，目前存在的处理饱和脉冲的结果大多数都是基于二次 Lyapunov 函数。然而，利用二次 Lyapunov 函数估计出的吸引域与真实的吸引域之间存在较大差距。为了降低由二次 Lyapunov 函数带来的保守性，构造了一个具有更多辅助矩阵的复合的依赖于脉冲时刻的 Lyapunov 函数，通过引入更多的决策变量来降低保守性，使得所估计的吸引域更接近真实吸引域。

## 5.2 展望

时滞复杂网络具有广泛的应用背景，其动力学特性的研究是近年来的热门课题。通过设计不同类型的脉冲控制器，本文对具有不同类型时滞的复杂网络的全局动力学以及局部动力学进行了分析，但仍有一些问题值得更深入的研究。首先，对于混杂脉冲控制下时滞复杂网络的多同步研究已经取得了一些理论成果，但是其实际应用还需要进一步探索。其次，对于事件触发脉冲控制，已经取得了一些成果，但尚有许多值得进一步探究的课题。此外，目前对于饱和脉冲控制下时滞复杂网络的局部动力学研究相对较少，还存在着大量开放的问题。基于本人的前期研究和本文的讨论，下面几个问题值得考虑：

1. 多同步的实际应用需要进一步探索。由于具有更高的存储能力，耦合多稳态时滞神经网络在处理高维数据时具有更高的效率。目前耦合多稳态时滞神经网络的多同步大多还是在理论研究上，只有较少的文献将耦合多稳态时滞神经网络的多同步应用于图像处理中，今后可以考虑进一步探索其在联想记忆、模式识别、人工智能等相关

领域的应用。此外，本文仅讨论了耦合多稳态时滞忆阻神经网络模型，更多的类型的耦合多稳态神经网络值得进一步探索，以处理不同的实际问题。

2. 事件触发脉冲控制值得进一步改进。到目前为止，关于事件触发脉冲控制的研究成果，大部分都是同步事件触发机制和集中式控制策略。首先，同步事件触发机制需要网络中所有节点在同一时刻的采样信息来判断触发条件是否满足，这可能导致不必要的资源浪费。因此，在设计事件触发脉冲控制时，每个节点都有单独的触发时刻进行信息传输和控制更新的异步事件触发机制值得被考虑。此外，相比于集中式的脉冲控制，分布式的脉冲控制中每个节点可以以协作的方式低代价地完成控制任务，因此，分布式脉冲控制在设计事件触发脉冲控制机制时值得被充分考虑。

3. 关于饱和脉冲控制有待进一步分析。对于执行器饱和而言，传输信号速率约束比幅值大小的约束对系统的影响更为显著，因此，具有执行器速率饱和的脉冲控制具有重要的研究意义。此外，目前关于饱和脉冲控制的研究成果大部分考虑的是执行器饱和的情况。而在实际的工业生产中，由于制造技术以及使用安全等诸多因素限制，传感器会无法识别或者提供幅值过大的信号，因而产生传感器的饱和特性。因此，具有传感器饱和的脉冲控制也值得进一步研究。

4. 多种网络问题值得进一步考虑。本文的结果都是在理想的网络环境中，而由于通信噪声、带宽受限、信息拥塞等原因，实际无线通信网络的丢包普遍存在。网络丢包会降低系统的性能甚至导致不稳定，因此研究丢包环境下脉冲控制是很有必要的。此外，由于通信环境的不稳定，复杂网络会受到很多干扰，例如网络攻击。网络攻击会破坏节点之间的通信链路，从而破坏系统的性能。因此，具有网络攻击的复杂网络的信息安全问题也具有重要的研究价值。

5. 复杂网络的网络拓扑结构值得进一步研究。为了实现所期待的动力学行为，本文考虑的网络拓扑都是连通或强连通的。而在许多实际网络中，网络拓扑可能是切换的，但每个切换的子图不满足连通或强连通条件。目前，已经有一些关于连续控制下具有联合连通或序列联通的复杂网络同步的研究成果，而对不连续的脉冲控制来说，不连通的复杂网络的研究还较少。脉冲控制是否能在联合连通或序列联通的网络拓扑下实现复杂网络的同步，这值得进一步探索。



# 参考文献

- [1] J. A. Bondy and U. S. R. Murty. *Graph Theory with Applications*. Amsterdam: North-Holland, 1976.
- [2] E. N. Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.
- [3] D. J. Watts and S. H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684):440–442, 1998.
- [4] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- [5] M. Dehmer and F. Emmert-Streib. *Analysis of Complex Networks: From Biology to Linguistics*. Hoboken, New Jersey, USA: Wiley-Blackwell, 2009.
- [6] L. Costa, O. Oliveira, G. Travieso, F. Rodrigues, P. Boas, L. Antiqueira, M. Viana, and L. Rocha. Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Advances in Physics*, 60(3):329–412, 2011.
- [7] E. Estrada. *The Structure of Complex Networks: Theory and Applications*. Oxford, U.K.: Oxford University Press, 2012.
- [8] X.-J. Xu, H.-O. Peng, X.-M. Wang, and Y.-H. Wang. Epidemic spreading with time delay in complex networks. *Physica A: Statistical Mechanics and its Applications*, 367:525–530, 2006.
- [9] H. Liu, J.-A. Lu, J. Lü, and D. J. Hill. Structure identification of uncertain general complex dynamical networks with time delay. *Automatica*, 45(8):1799–1807, 2009.
- [10] L. Wang, Z. Wang, T. Huang, and G. Wei. An event-triggered approach to state estimation for a class of complex networks with mixed time delays and nonlinearities. *IEEE Transactions on Cybernetics*, 46(11):2497–2508, 2015.

- [11] J. Lu and D. W. C. Ho. Local and global synchronization in general complex dynamical networks with delay coupling. *Chaos, Solitons and Fractals*, 37(5):1497–1510, 2008.
- [12] Z. Fei, H. Gao, and W. Zheng. New synchronization stability of complex networks with an interval time-varying coupling delay. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 56(6):499–503, 2009.
- [13] B. Wang and Z. H. Guan. Chaos synchronization in general complex dynamical networks with coupling delays. *Nonlinear Analysis: Real World Applications*, 11(3):1925–1932, 2010.
- [14] Z.-G. Wu, P. Shi, H. Su, and J. Chu. Sampled-data exponential synchronization of complex dynamical networks with time-varying coupling delay. *IEEE Transactions on Neural Networks and Learning Systems*, 24(8):1177–1187, 2013.
- [15] M. Park, S.-H. Lee, K. Oh-Min, and S. Alexandre. Closeness-centrality-based synchronization criteria for complex dynamical networks with interval time-varying coupling delays. *IEEE Transactions on Cybernetics*, 48(7):2192–2202, 2018.
- [16] J.-W. Wang, Q. Ma, L. Zeng, and M. S. Abd-Elouahab. Mixed outer synchronization of coupled complex networks with time-varying coupling delay. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 21(1):013121, 2011.
- [17] M. S. Ali, M. Usha, J. Cao, and G. Lu. Synchronisation analysis for stochastic T–S fuzzy complex networks with coupling delay. *International Journal of Systems Science*, 50(3):585–598, 2019.
- [18] Y. Liu, Z. Wang, Y. Yuan, and F. E. Alsaadi. Partial-nodes-based state estimation for complex networks with unbounded distributed delays. *IEEE Transactions on Neural Networks and Learning Systems*, 29(8):3906–3912, 2017.
- [19] L. Zou, Z. Wang, H. Gao, and X. Liu. Event-triggered state estimation for complex networks with mixed time delays via sampled data information: The continuous-time case. *IEEE Transactions on Cybernetics*, 45(12):2804–2815, 2015.

- [20] K. Gu, J. Chen, and V. L. Kharitonov. *Stability of time-delay systems*. Boston: Birkhäuser, 2003.
- [21] E. Fridman. *Introduction to time-delay systems: Analysis and control*. Birkhäuser Basel, 2014.
- [22] 孙健, 陈杰, 刘国平. 时滞系统稳定性分析与应用. 科学出版社, 2012.
- [23] X. F. Wang. Complex networks: topology, dynamics and synchronization. *International Journal of Bifurcation and Chaos*, 12(05):885–916, 2002.
- [24] L. Cheng, Z.-G. Hou, and M. Tan. A mean square consensus protocol for linear multi-agent systems with communication noises and fixed topologies. *IEEE Transactions on Automatic Control*, 59(1):261–267, 2013.
- [25] J. Camacho, R. Guimerà, and L. A. N. Amaral. Robust patterns in food web structure. *Physical Review Letters*, 88(22):228102, 2002.
- [26] W. Yu, G. Chen, M. Cao, and J. Kurths. Second-order consensus for multiagent systems with directed topologies and nonlinear dynamics. *IEEE Transactions on Systems, Man, and Cybernetics: Cybernetics*, 40(3):881–891, 2009.
- [27] H. Gu, P. Liu, J. Lü, and Z. Lin. PID control for synchronization of complex dynamical networks with directed topologies. *IEEE Transactions on Cybernetics*, 51(3):1334–1346, 2021.
- [28] L. Wang and Q. Wang. Synchronization in complex networks with switching topology. *Physics Letters A*, 375(34):3070–3074, 2011.
- [29] Y. Jin and S. Zhong. Function projective synchronization in complex networks with switching topology and stochastic effects. *Applied Mathematics and Computation*, 259:730–740, 2015.
- [30] Y. Chen, W. Yu, S. Tan, and H. Zhu. Synchronizing nonlinear complex networks via switching disconnected topology. *Automatica*, 70:189–194, 2016.

- [31] W. Li, Y. Jia, and J. Du. State estimation for stochastic complex networks with switching topology. *IEEE Transactions on Automatic Control*, 62(12):6377–6384, 2017.
- [32] R. E. Mirollo and S. H. Strogatz. Synchronization of pulse-coupled biological oscillators. *SIAM Journal on Applied Mathematics*, 50(6):1645–1662, 1990.
- [33] P. Fries, J. H. Reynolds, A. E. Rorie, and R. Desimone. Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, 291(5508):1560–1563, 2001.
- [34] P. N. Steinmetz, A. Roy, P. Fitzgerald, S. Hsiao, K. Johnson, and E. Niebur. Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature*, 404(6774):187–190, 2000.
- [35] 卢文联. 时滞复杂系统动力学：从神经网络到复杂网络. 复旦大学出版社有限公司, 2018.
- [36] J. Liang, Z. Wang, and X. Liu. Exponential synchronization of stochastic delayed discrete-time complex networks. *Nonlinear Dynamics*, 53(1):153–165, 2008.
- [37] B. Song, J. H. Park, Z.-G. Wu, and Y. Zhang. Global synchronization of stochastic delayed complex networks. *Nonlinear Dynamics*, 70(4):2389–2399, 2012.
- [38] L. Shi, H. Zhu, S. Zhong, Y. Zeng, and J. Cheng. Synchronization for time-varying complex networks based on control. *Journal of Computational and Applied Mathematics*, 301:178–187, 2016.
- [39] Q. Zhang, J. Lu, J. Lu, and K. T. Chi. Adaptive feedback synchronization of a general complex dynamical network with delayed nodes. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 55(2):183–187, 2008.
- [40] D. J. Stilwell, E. M. Bollt, and D. G. Roberson. Sufficient conditions for fast switching synchronization in time-varying network topologies. *SIAM Journal on Applied Dynamical Systems*, 5(1):140–156, 2006.

- [41] T. Hu, J. H. Park, X. Liu, Z. He, and S. Zhong. Sampled-data-based event-triggered synchronization strategy for fractional and impulsive complex networks with switching topologies and time-varying delay. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021, DOI: 10.1109/TSMC.2021.3071811.
- [42] Z. Tang, J. H. Park, Y. Wang, and J. Feng. Adaptively synchronize the derivative coupled complex networks with proportional delay. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51(8):4969–4979, 2021.
- [43] X. Liu and T. Chen. Exponential synchronization of nonlinear coupled dynamical networks with a delayed coupling. *Physica A: Statistical Mechanics and its Applications*, 381:82–92, 2007.
- [44] S. Yang, Z. Guo, and J. Wang. Robust synchronization of multiple memristive neural networks with uncertain parameters via nonlinear coupling. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 45(7):1077–1086, 2015.
- [45] Y. Wu and L. Liu. Exponential outer synchronization between two uncertain time-varying complex networks with nonlinear coupling. *Entropy*, 17(5):3097–3109, 2015.
- [46] A. Popov and H. Werner. Robust stability of a multi-agent system under arbitrary and time-varying communication topologies and communication delays. *IEEE Transactions on Automatic Control*, 57(9):2343–2347, 2012.
- [47] T. Wang, W. Zhou, and S. Zhao. Robust synchronization for stochastic delayed complex networks with switching topology and unmodeled dynamics via adaptive control approach. *Communications in Nonlinear Science and Numerical Simulation*, 18(8):2097–2106, 2013.
- [48] X. Yu, Q. Ren, J. Hou, and J. Zhao. The chaotic phase synchronization in adaptively coupled-delayed complex networks. *Physics Letters A*, 373(14):1276–1282, 2009.

- [49] H. Du, S. Peng, and L. Ning. Function projective synchronization in complex dynamical networks with time delay via hybrid feedback control. *Nonlinear Analysis Real World Applications*, 14(2):1182–1190, 2013.
- [50] Z.-H. Guan, D. Yue, B. Hu, T. Li, and F. Liu. Cluster synchronization of coupled genetic regulatory networks with delays via aperiodically adaptive intermittent control. *IEEE Transactions on Nanobioscience*, 16(7):585–599, 2017.
- [51] Q. Wang, J.-L. Wang, Y.-L. Huang, and S.-Y. Ren. Generalized lag synchronization of multiple weighted complex networks with and without time delay. *Journal of the Franklin Institute*, 355(14):6597–6616, 2018.
- [52] T. Chen, W. Wang, and Y. Wu. Quasi-synchronization of fuzzy heterogeneous complex networks via intermittent discrete-time state observations control. *IEEE Transactions on Fuzzy Systems*, 2021, DOI: 10.1109/TFUZZ.2021.3103597.
- [53] L. Lu, M. Ge, Y. Xu, and Y. Jia. Phase synchronization and mode transition induced by multiple time delays and noises in coupled fitzhugh–nagumo model. *Physica A: Statistical Mechanics and its Applications*, 535:122419, 2019.
- [54] M. Han, Y. Zhang, and M. Zhang. Modified function projective synchronization between two different complex networks with delayed couplings and delayed nodes of different dimensions. *International Journal of Modern Physics C*, 27(11):1650126, 2016.
- [55] R. Wei and Y. Cao. *Distributed Coordination of Multi-agent Networks: Emergent Problems, Models, and Issues*. Springer London, 2011.
- [56] Y. Cao, W. Yu, W. Ren, and G. Chen. An overview of recent progress in the study of distributed multi-agent coordination. *IEEE Transactions on Industrial Informatics*, 9(1):427–438, 2013.
- [57] K. Sakurama, S. I. Azuma, and T. Sugie. Multi-agent coordination to high-dimensional target subspaces. *IEEE Transactions on Control of Network Systems*, 5(1):345–358, 2018.

- [58] 高利新, 徐晓乐. 多智能体系统基于观察器的一致性控制. 上海交通大学出版社, 2018.
- [59] D. Lei and G. Ge. Sampled-data leader-following consensus for nonlinear multi-agent systems with Markovian switching topologies and communication delay. *Journal of the Franklin Institute*, 352(1):369–383, 2015.
- [60] H. Meng, Z. Chen, L. Zhu, and R. Middleton. Consensus of a class of second-order nonlinear heterogeneous multi-agent systems with uncertainty and communication delay. *International Journal of Robust and Nonlinear Control*, 26:3311–3329, 2016.
- [61] C. L. P. Chen, G.-X. Wen, Y.-J. Liu, and F.-Y. Wang. Adaptive consensus control for a class of nonlinear multiagent time-delay systems using neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 25(6):1217–1226, 2014.
- [62] L. Liu, H. Sun, L. Ma, J. Zhang, and Y. Bo. Quasi-consensus control for a class of time-varying stochastic nonlinear time-delay multiagent systems subject to deception attacks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51(11):6863–6873, 2021.
- [63] H. Li, X. Liao, T. Huang, and W. Zhu. Event-triggering sampling based leader-following consensus in second-order multi-agent systems. *IEEE Transactions on Automatic Control*, 60(7):1998–2003, 2015.
- [64] H. Kim, H. Shim, and J. H. Seo. Output consensus of heterogeneous uncertain linear multi-agent systems. *IEEE Transactions on Automatic Control*, 56(1):200–206, 2011.
- [65] H. Du, G. Wen, G. Chen, J. Cao, and F. E. Alsaadi. A distributed finite-time consensus algorithm for higher-order leaderless and leader-following multiagent systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47(7):1625–1634, 2017.
- [66] H. Li, Q. Liu, G. Feng, and X. Zhang. Leader-follower consensus of nonlinear time-delay multiagent systems: A time-varying gain approach. *Automatica*, 126(11):109444, 2021.

- [67] K.-P. Wang, D. Ding, Z. Tang, and J. Feng. Leader-following consensus of nonlinear multi-agent systems with hybrid delays: Distributed impulsive pinning strategy. *Applied Mathematics and Computation*, 424:127031, 2022.
- [68] D. Yue, J. Cao, Q. Li, and X. Shi. Neuro-adaptive consensus strategy for a class of nonlinear time-delay multi-agent systems with an unmeasurable high-dimensional leader. *IET Control Theory and Applications*, 13(2):230–238, 2019.
- [69] Y. Li, J. H. Park, C. Hua, and G. Liu. Distributed adaptive output feedback containment control for time-delay nonlinear multiagent systems. *Automatica*, 127:109545, 2021.
- [70] G. Wen, X. Zhai, Z. Peng, and A. Rahmani. Fault-tolerant secure consensus tracking of delayed nonlinear multi-agent systems with deception attacks and uncertain parameters via impulsive control. *Communications in nonlinear science and numerical simulation*, 82:105043, 2020.
- [71] X. Jiang, G. Xia, Z. Feng, and T. Li. Non-fragile  $H_\infty$  consensus tracking of nonlinear multi-agent systems with switching topologies and transmission delay via sampled-data control. *Information Sciences*, 509:210–226, 2020.
- [72] C.-J. Li and G.-P. Liu. Data-driven consensus for non-linear networked multi-agent systems with switching topology and time-varying delays. *IET Control Theory and Applications*, 12(12):1773–1779, 2018.
- [73] Z. Yaghoubi and H. A. Talebi. Cluster consensus of fractional-order non-linear multi-agent systems with switching topology and time-delays via impulsive control. *International Journal of Systems Science*, 51(10):1685–1698, 2020.
- [74] X. Jiang, G. Xia, Z. Feng, and Z. Jiang.  $H_\infty$  delayed tracking protocol design of nonlinear singular multi-agent systems under Markovian switching topology. *Information Sciences*, 545:280–297, 2021.
- [75] H. Li, X. Liao, X. Lei, T. Huang, and W. Zhu. Second-order consensus seeking in multi-agent systems with nonlinear dynamics over random switching directed net-

- works. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 60(6):1595–1607, 2013.
- [76] M. Sharifi and M. J. Yazdanpanah. Finite time consensus of nonlinear multi-agent systems in the presence of communication time delays. *European Journal of Control*, 53:10–19, 2020.
- [77] K. Subramanian, P. Muthukumar, and Y. H. Joo. Leader-following consensus of nonlinear multi-agent systems via reliable control with time-varying communication delay. *International Journal of Control, Automation and Systems*, 17(2):298–306, 2019.
- [78] X. Jiang, G. Xia, Z. Feng, and Z. Jiang. Consensus tracking of data-sampled nonlinear multi-agent systems with packet loss and communication delay. *IEEE Transactions on Network Science and Engineering*, 8(1):126–137, 2021.
- [79] J. Lu, X. Guo, T. Huang, and Z. Wang. Consensus of signed networked multi-agent systems with nonlinear coupling and communication delays. *Applied Mathematics and Computation*, 350:153–162, 2019.
- [80] D. Wang, H. Ma, and D. Liu. Distributed control algorithm for bipartite consensus of the nonlinear time-delayed multi-agent systems with neural networks. *Neurocomputing*, 174:928–936, 2016.
- [81] W. Qian, L. Wang, and M. Z. Chen. Local consensus of nonlinear multiagent systems with varying delay coupling. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 48(12):2462–2469, 2018.
- [82] X. Liu, J.-W. Xiao, D. Chen, and Y.-W. Wang. Dynamic consensus of nonlinear time-delay multi-agent systems with input saturation: an impulsive control algorithm. *Nonlinear Dynamics*, 97(2):1699–1710, 2019.
- [83] V. Lakshmikantham, D. D. Bainov, and P. S. Simeonov. *Theory of Impulsive Differential Equations*, volume 6. Singapore: World Scientific, 1989.
- [84] T. Yang. *Impulsive control theory*. New York: Nova Science, 2001.

- [85] S. Yang, C. Li, and T. Huang. Synchronization of coupled memristive chaotic circuits via state-dependent impulsive control. *Nonlinear Dynamics*, 88(1):115–129, 2017.
- [86] R. Goedel, R. G. Sanfelice, and A. R. Teel. *Hybrid Dynamical Systems: Modeling Stability, and Robustness*. Princeton University Press Princeton, New Jersey, 2012.
- [87] Y. Wang, J. Lu, X. Li, and J. Liang. Synchronization of coupled neural networks under mixed impulsive effects: a novel delay inequality approach. *Neural Networks*, 127:38–46, 2020.
- [88] Y. Tang, H. Gao, W. Zhang, and J. Kurths. Leader-following consensus of a class of stochastic delayed multi-agent systems with partial mixed impulses. *Automatica*, 53:346–354, 2015.
- [89] W. M. Haddad, V. Chellaboina, and S. G. Nersesov. *Impulsive and Hybrid Dynamical Systems: Stability, Dissipativity and Control*. Princeton University Press Princeton, New Jersey, 2006.
- [90] T. Yang and L. O. Chua. Impulsive control and synchronization of nonlinear dynamical systems and application to secure communication. *International Journal of Bifurcation and Chaos*, 7(03):645–664, 1997.
- [91] J. Lu, D. W. Ho, and J. Cao. A unified synchronization criterion for impulsive dynamical networks. *Automatica*, 46(7):1215–1221, 2010.
- [92] N. Wang, X. Li, J. Lu, and F. E. Alsaadi. Unified synchronization criteria in an array of coupled neural networks with hybrid impulses. *Neural Networks*, 101:25–32, 2018.
- [93] J. Lu, Z. Wang, J. Cao, D. W. Ho, and J. Kurths. Pinning impulsive stabilization of nonlinear dynamical networks with time-varying delay. *International Journal of Bifurcation and Chaos*, 22(07):1250176, 2012.
- [94] W. Sun, Z. Chen, J. Lü, and S. Chen. Outer synchronization of complex networks with delay via impulse. *Nonlinear Dynamics*, 69(4):1751–1764, 2012.

- [95] J. Feng, F. Yu, and Y. Zhao. Exponential synchronization of nonlinearly coupled complex networks with hybrid time-varying delays via impulsive control. *Nonlinear Dynamics*, 85(1):621–632, 2016.
- [96] X. Liu, K. Zhang, and W.-C. Xie. Consensus of multi-agent systems via hybrid impulsive protocols with time-delay. *Nonlinear Analysis: Hybrid Systems*, 30:134–146, 2018.
- [97] X. Li and S. Song. Stabilization of delay systems: Delay-dependent impulsive control. *IEEE Transactions on Automatic Control*, 62(1):406–411, 2016.
- [98] W. Zhang, C. Li, T. Huang, and X. He. Synchronization of memristor-based coupling recurrent neural networks with time-varying delays and impulses. *IEEE Transactions on Neural Networks and Learning Systems*, 26(12):3308–3313, 2015.
- [99] X. Li and J. Wu. Stability of nonlinear differential systems with state-dependent delayed impulses. *Automatica*, 64:63–69, 2016.
- [100] W. He, G. Chen, Q. L. Han, and F. Qian. Network-based leader-following consensus of nonlinear multi-agent systems via distributed impulsive control. *Information Sciences*, 380:145–158, 2017.
- [101] X. Yang, X. Li, Q. Xi, and P. Duan. Review of stability and stabilization for impulsive delayed systems. *Mathematical Biosciences and Engineering*, 15(6):1495–1515, 2018.
- [102] X. He, D. Peng, and X. Li. Synchronization of complex networks with impulsive control involving stabilizing delay. *Journal of the Franklin Institute*, 357(8):4869–4886, 2020.
- [103] J. Lu, J. Kurths, J. Cao, N. Mahdavi, and C. Huang. Synchronization control for nonlinear stochastic dynamical networks: Pinning impulsive strategy. *IEEE Transactions on Neural Networks and Learning Systems*, 23(2):285–292, 2012.

- [104] X. Yang, J. Cao, and Z. Yang. Synchronization of coupled reaction-diffusion neural networks with time-varying delays via pinning-impulsive controller. *SIAM Journal on Control and Optimization*, 51(5):3486–3510, 2013.
- [105] X. Wang, J. H. Park, H. Yang, and S. Zhong. An improved impulsive control approach for cluster synchronization of complex networks with parameter mismatches. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51(4):2561–2570, 2021.
- [106] J. Zhou, Q. Wu, and L. Xiang. Pinning complex delayed dynamical networks by a single impulsive controller. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 58(12):2882–2893, 2011.
- [107] H. Liu, J. Li, Z. Li, Z. Zeng, and J. Lü. Intralayer synchronization of multiplex dynamical networks via pinning impulsive control. *IEEE Transactions on Cybernetics*, 52(4):2110–2122, 2022.
- [108] X. Yang and J. Cao. Hybrid adaptive and impulsive synchronization of uncertain complex networks with delays and general uncertain perturbations. *Applied Mathematics and Computation*, 227:480–493, 2014.
- [109] X. Yang, X. Li, J. Lu, and Z. Cheng. Synchronization of time-delayed complex networks with switching topology via hybrid actuator fault and impulsive effects control. *IEEE Transactions on Cybernetics*, 50(9):4043–4052, 2020.
- [110] H.-L. Li, J. Cao, C. Hu, L. Zhang, and Z. Wang. Global synchronization between two fractional-order complex networks with non-delayed and delayed coupling via hybrid impulsive control. *Neurocomputing*, 356:31–39, 2019.
- [111] X. Tan, J. Cao, and X. Li. Consensus of leader-following multiagent systems: A distributed event-triggered impulsive control strategy. *IEEE Transactions on Cybernetics*, 49(3):792–801, 2019.
- [112] W. Zhu, D. Wang, L. Liu, and G. Feng. Event-based impulsive control of continuous-time dynamic systems and its application to synchronization of mem-

- ristive neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 29(8):3599–3609, 2018.
- [113] D. Peng and X. Li. Leader-following synchronization of complex dynamic networks via event-triggered impulsive control. *Neurocomputing*, 412:1–10, 2020.
- [114] J. Chen, B. Chen, and Z. Zeng. Synchronization in multiple neural networks with delay and disconnected switching topology via event-triggered impulsive control strategy. *IEEE Transactions on Industrial Electronics*, 68(3):2491–2500, 2021.
- [115] X. Hai, G. Ren, Y. Yu, C. Xu, and Y. Zeng. Pinning synchronization of fractional and impulsive complex networks via event-triggered strategy. *Communications in Nonlinear Science and Numerical Simulation*, 82:105017, 2020.
- [116] O. Brieger, M. Kerr, D. Leißling, I. Postlethwaite, J. Sofrony, and M. Turner. Flight testing of a rate saturation compensation scheme on the ATTAS aircraft. *Aerospace Science and Technology*, 13(2-3):92–104, 2009.
- [117] E. Farantatos, R. Huang, G. J. Cokkinides, and A. Meliopoulos. A predictive generator out-of-step protection and transient stability monitoring scheme enabled by a distributed dynamic state estimator. *IEEE Transactions on Power Delivery*, 31(4):1826–1835, 2016.
- [118] B. Xiao, Q. Hu, and Y. Zhang. Adaptive sliding mode fault tolerant attitude tracking control for flexible spacecraft under actuator saturation. *IEEE Transactions on Control Systems Technology*, 20(6):1605–1612, 2012.
- [119] H. Li, C. Li, D. Ouyang, and S. K. Nguang. Impulsive synchronization of unbounded delayed inertial neural networks with actuator saturation and sampled-data control and its application to image encryption. *IEEE Transactions on Neural Networks and Learning Systems*, 32(4):1460–1473, 2020.
- [120] H. Li, C. Li, D. Ouyang, and S. K. Nguang. Impulsive stabilization of nonlinear time-delay system with input saturation via delay-dependent polytopic approach. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51(11):7087–7098, 2021.

- [121] Z. Shen, C. Li, H. Li, and Z. Cao. Estimation of the domain of attraction for discrete-time linear impulsive control systems with input saturation. *Applied Mathematics and Computation*, 362:124502, 2019.
- [122] Z. He, C. Li, Y. Li, Z. Cao, and X. Zhang. Local synchronization of nonlinear dynamical networks with hybrid impulsive saturation control inputs. *Applied Mathematics and Computation*, 410:126452, 2021.
- [123] D. Ouyang, J. Shao, H. Jiang, S. K. Nguang, and H. T. Shen. Impulsive synchronization of coupled delayed neural networks with actuator saturation and its application to image encryption. *Neural Networks*, 128:158–171, 2020.
- [124] W. Lu and T. Chen. New approach to synchronization analysis of linearly coupled ordinary differential systems. *Physica D: Nonlinear Phenomena*, 213(2):214–230, 2006.
- [125] C. Godsil and G. F. Royle. *Algebraic graph theory*. New York, USA: Springer-Verlag, 2001.
- [126] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge University Press, 1991.
- [127] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*. Philadelphia, PA: SIAM, 1994.
- [128] J. Pearl. Causal inference in statistics: An overview, stat. *Surv.*, 3:96–146, 2009.
- [129] S. Epskamp and E. I. Fried. A tutorial on regularized partial correlation networks, psychol. *Methods*, 23(4):617–634, 2018.
- [130] T. Schreiber. Measuring information transfer. *Phys. Rev. Letters*, 85:461464, 2000.
- [131] A. Sensoy, C. Sobaci, S. Sensoy, and F. Alali. Effective transfer entropy approach to information flow between exchange rates and stock markets. *Chaos Solitons Fractals*, 68:180–185, 2014.

- [132] A. Rényi. On measures of entropy and information. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, pages 547–561, Berkeley, California., 1961. Univeristy of Berkeley.
- [133] M. Xu, P. Shang, and S. Zhang. Multiscale renyi cumulative residual distribution entropy: Reliability analysis of financial time series. *Chaos Solitons Fractals*, 143:110410, 2021.
- [134] Y. Mi and A. Lin. Kernel based multiscale partial renyi transfer entropy and its applications. *Commun. Nonlinear Sci. Numer. Simul.*, 119:107084, 2023.
- [135] J. He and P. Shang. Comparison of transfer entropy methods for financial time series,. *Phys. A*, 482:772–785, 2017.
- [136] A. Bossman, Z. Umar, S. Agyei, and P. Junior. A new iceemdan-based transfer entropy quantifying information flow between real estate and policy uncertainty. *Res. Econ.*, 76(3):189–205, 2022.
- [137] M. Staniek and K. Lehnertz. Symbolic transfer entropy. *Phys. Rev. Lett.*, 100(15):158101, 2008.
- [138] A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. *Phys. Rev. E.*, 69:066138, 2004.
- [139] L. Faes, G. Nollo, and A. Porta. Non-uniform multivariate embedding to assess the information transfer in cardiovascular and cardiorespiratory variability series. *Comput. Biol. Med.*, 42:290–297, 2012.
- [140] J. D. Wren. Extending the mutual information measure to rank inferred literature relationships. *BMC Bioinform.*, 5:1–13, 2004.
- [141] L. G. S. Giraldo, M. Rao, and J. C. Principe. Measures of entropy from data using infinitely divisible kernels, ieee trans. *IEEE Trans. Inf. Theory*, 61:535–548, 2015.
- [142] I. D. L. P. Panche, A. M. Alvarez-Meza, and A. Orozco-Gutierrez. A data-driven measure of effective connectivity based on renyi’s  $\alpha$ -entropy. *Front. Neurosci.*, 13:1277, 2019.

- [143] W. Zhou, S. Yu, and B. Chen. Causality detection with matrix-based transfer entropy. *Inf. Sci.*, 613:357–375, 2022.
- [144] M. Materassi, A. Wernik, and E. Yordanova. Determining the verse of magnetic turbulent cascades in the earth’s magnetospheric cusp via transfer entropy analysis: preliminary results. *Nonlinear Process. Geophys.*, 14:153–161, 2007.
- [145] D. Storhas, L. D. Mello, and A. Singh. Multiscale lead-lag relationships in oil and refined product return dynamics: A symbolic wavelet transfer entropy approach. *Energy Econ.*, 92:104927, 2020.
- [146] M. Lungarella, A. Pitti, and Y. Kuniyoshi. Information transfer at multiple scales. *Phys. Rev. E.*, 76(5):056117, 2007.
- [147] Z. Guo, V. McClelland, O. Simeone, K. Mills, and Z. Cvetkovic. Multiscale wavelet transfer entropy with application to corticomuscular coupling analysis. *IEEE Trans. Biomed. Eng.*, 69:771–782, 2022.
- [148] Y. Lu, N. Zeng, Y. Liu, and N. Zhang. A hybrid wavelet neural network and switching particle swarm optimization algorithm for face direction recognition. *Neurocomputing*, 155:219–224, 2015.
- [149] M. Saadatmorad, R. Jafari-Talookolaei, M. Pashaei, and S. Khatir. Damage detection on rectangular laminated composite plates using wavelet based convolutional neural network technique. *Comp. Struct.*, 278:114656, 2021.
- [150] M. Zhao, M. Kang, B. Tang, and M. Pecht. Multiple wavelet coefficients fusion in deep residual networks for fault diagnosis, ieee trans. *IEEE Trans. Ind. Electron.*, 66(6):4696–4706, 2018.
- [151] L. Bruce, C. Koger, and J. Li. Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction. *IEEE Trans. Geosci. Remote Sens.*, 40:2331–2338, 2002.
- [152] T. Duong and M. Hazelton. Plug-in bandwidth matrices for bivariate kernel density estimation. *J. Nonparametr. Stat.*, 15:17–30, 2003.

- [153] D. W. Scott. *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons Press, 2015.
- [154] J. M. Mooij, J. Peters, D. Janzing, J. Zscheischler, and B. Schölkopf. Distinguishing cause from effect using observational data: methods and benchmarks. *J. Mach. Learn. Res.*, 17:1103–1204, 2016.
- [155] I. C. Yeh. Modeling of strength of high-performance concrete using artificial neural networks. *Cem. Concr. Res.*, 28:1797–1808, 1998.
- [156] T. Dimpfl and F. J. Peter. Using transfer entropy to measure information flows between financial markets. *Stud. Nonlinear Dyn. Econom.*, 17:85–102, 2013.
- [157] K. Bae, G. A. Karolyi, and R. M. Stulz. A new approach to measuring financial contagion. *Rev. Financ. Stud.*, 16:717–763, 2003.
- [158] K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka, and J. Bhattacharya. Causality detection based on information-theoretic approaches in time series analysis. *Phys. Rep.*, 441(1):1–46, 2007.
- [159] M. Yalcin, J. Suykens, and J. Vandewalle. Experimental confirmation of 3-and 5-scroll attractors from a generalized Chua’s circuit. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 47(3):425–429, 2000.
- [160] Y.-W. Wang, W. Yang, J.-W. Xiao, and Z.-G. Zeng. Impulsive multisynchronization of coupled multistable neural networks with time-varying delay. *IEEE Transactions on Neural Networks and Learning Systems*, 28(7):1560–1571, 2016.
- [161] R. Rakkiyappan, G. Velmurugan, J. N. George, and R. Selvamani. Exponential synchronization of Lur’e complex dynamical networks with uncertain inner coupling and pinning impulsive control. *Applied Mathematics and Computation*, 307:217–231, 2017.
- [162] W. Zhang, Y. Tang, X. Wu, and J.-A. Fang. Synchronization of nonlinear dynamical networks with heterogeneous impulses. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 61(4):1220–1228, 2013.

- [163] Z. Yang and D. Xu. Stability analysis and design of impulsive control systems with time delay. *IEEE Transactions on Automatic Control*, 52(8):1448–1454, 2007.
- [164] M. C. Fan, Z. Chen, and H. T. Zhang. Semi-global consensus of nonlinear second-order multi-agent systems with measurement output feedback. *IEEE Transactions on Automatic Control*, 59(8):2222–2227, 2014.
- [165] W. He, B. Zhang, Q. L. Han, Q. Feng, J. Kurths, and J. Cao. Leader-following consensus of nonlinear multiagent systems with stochastic sampling. *IEEE Transactions on Cybernetics*, 47(2):327–338, 2017.
- [166] C. Chen, C. Wen, Z. Liu, K. Xie, Y. Zhang, and C. L. P. Chen. Adaptive consensus of nonlinear multi-agent systems with non-identical partially unknown control directions and bounded modelling errors. *IEEE Transactions on Automatic Control*, 62(9):4654–4659, 2017.
- [167] X. Zhang, X. Lv, and X. Li. Sampled-data-based lag synchronization of chaotic delayed neural networks with impulsive control. *Nonlinear Dynamics*, 90(3):2199–2207, 2017.
- [168] W.-H. Chen, Z. Ruan, and W. Zheng. Stability and  $L_2$ -gain analysis for impulsive delay systems: An impulse-time-dependent discretized Lyapunov functional method. *Automatica*, 86:129–137, 2017.
- [169] B. Zhou. Analysis and design of discrete-time linear systems with nested actuator saturations. *Systems and Control Letters*, 62(10):871–879, 2013.
- [170] Y. Tao, Z. Meng, D. V. Dimarogonas, and K. H. Johansson. Global consensus for discrete-time multi-agent systems with input saturation constraints. *Automatica*, 50(2):499–506, 2014.
- [171] B. Wang, J. Wang, B. Zhang, and X. Li. Global cooperative control framework for multiagent systems subject to actuator saturation with industrial applications. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47(7):1270–1283, 2017.

- 
- [172] H. Chu, B. Yi, G. Zhang, and W. Zhang. Performance improvement of consensus tracking for linear multiagent systems with input saturation: A gain scheduled approach. *IEEE Transactions on Systems Man and Cybernetics Systems*, 50(3):734–746, 2020.
  - [173] H. Arianfar and M. Haeri. Adaptive flocking control of nonlinear multi-agent systems with directed switching topologies and saturation constraints. *Journal of the Franklin Institute*, 350(6):1545–1561, 2013.
  - [174] X. You, C. Hua, and X. Guan. Event-triggered leader-following consensus for nonlinear multiagent systems subject to actuator saturation using dynamic output feedback method. *IEEE Transactions on Automatic Control*, 63(12):4391–4396, 2018.
  - [175] A. U. Rehman, M. Rehan, M. R. Mehar, M. Abid, and N. I. Awan. Consensus tracking of nonlinear multi-agent systems under input saturation with applications: A sector-based approach. *ISA Transactions*, 107:194–205, 2020.
  - [176] H. Yang and D. Ye. Adaptive fuzzy nonsingular fixed-time control for nonstrict-feedback constrained nonlinear multi-agent systems with input saturation. *IEEE Transactions on Fuzzy Systems*, 29(10):3142–3153, 2021.
  - [177] Y. Li and Z. Lin. *Stability and Performance of Control Systems with Actuator Saturation*. Switzerland: Birkhäuser, 2018.



# 附录 A 博士期间撰写和发表的论文

## 一、发表的期刊论文列表：

- [1]. 第一作者. Local synchronization of directed Lur'e networks with coupling delay via distributed impulsive control subject to actuator saturation, *IEEE Transactions on Neural Networks and Learning Systems*, 2022, DOI: 10.1109/TNNLS.2021.3138997. (SCI 收录)
- [2]. 第一作者. Synchronization analysis for complex dynamical networks with coupling delay via event-triggered delayed impulsive control, *IEEE Transactions on Cybernetics*, 51(11):5269–5278, 2021. (SCI 收录)
- [3]. 第一作者. Dynamical and static multisynchronization analysis for coupled multistable memristive neural networks with hybrid control, *Neural Networks*, 143:515-524, 2021. (SCI 收录)
- [4]. 第一作者. Dynamical and static multisynchronization of coupled multistable neural networks via impulsive control, *IEEE Transactions on Neural Networks and Learning Systems*, 29(12):6062–6072, 2018. (SCI 收录)
- [5]. 第四作者. Extended dissipativity performance of high-speed train including actuator faults and probabilistic time-delays under resilient reliable control, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51(6): 3808-3819, 2021. (SCI 收录)
- [6]. 第四作者. Enhanced  $L_2-L_\infty$  state estimation design for delayed neural networks including leakage term via quadratic-type generalized free-matrix-based integral inequality, *Journal of the Franklin Institute*, 356(13):7371-7392, 2019. (SCI 收录)

## 二、在审的期刊论文列表:

- [1]. 第一作者. Distributed saturated impulsive control for local consensus of non-linear time-delay multi-agent systems with switching topologies, *IEEE Transactions on Automatic Control*. Under review.
- [2]. 第一作者. Event-triggered pinning impulsive control for cluster synchronization of coupled genetic oscillator networks with proportional delay, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. Under review.

## 三、发表的会议论文列表:

- [1]. 第一作者. Finite-time average consensus of multi-agent systems with impulsive perturbations. *2020 12th International Conference on Advanced Computational Intelligence*, (ICACI 2020), 558–563, August 14–16, 2020, Dali, China. (EI 收录)
- [2]. 第一作者. Impulsive local leader-following consensus and estimation of domain of attraction of multi-agent systems with actuator saturation. *2021 11th International Conference on Information Science and Technology*, (ICIST 2021), 553–558, May 21–23, 2021, Chengdu, China. (EI 收录)

## 附录 B 博士期间主持和参加的科研项目、学术会议和获得的荣誉

### 一、博士期间主持和参加的科研项目：

1. 不连续控制下多稳态系统的动力学分析，东南大学优秀博士学位论文培育基金，(No. YBPY1967)，主持
2. 基于群体智能的分布式优化理论、方法及应用研究，国家自然科学基金重点项目 (No. 61833005)，参与
3. 多智能体最优合作调控及其在电力系统中应用，国家自然科学基金面上项目 (No. 61573096)，参与

### 二、博士期间参加的学术会议与学术论坛：

1. 2021 年 05 月，第 11 届信息科学与技术国际会议 (ICIST 2021)，成都，四川
2. 2020 年 11 月，第 44 届复杂系统与网络科学研究中心论坛，南京，江苏
3. 2020 年 08 月，第 12 届高级计算智能国际会议 (ICACI 2020)，大理，云南
4. 2019 年 10 月，第 12 届复杂系统与应用国际研讨会，南京，江苏
5. 2019 年 07 月，第 38 届中国控制会议 (CCC 2019)，广州，广东
6. 2018 年 10 月，2018 年国际复杂系统与网络论坛 (IWCSN 2018)，南京，江苏

### 三、博士期间科研工作经历：

1. 2019.10–2019.11 香港城市大学 (CityU)，香港，中国 (助理研究员)。

### 四、博士期间获得的荣誉：

1. 2018 年 12 月，东南大学新生奖学金