arXiv:2601.04515v1 [math.DS] 8 Jan 2026

Eugene Tan、David Walker、Michael S
mall、Braden Thorne

动力学、Complexity 与时间序列分析

理论与应用入门

2026年1月9日

施普林格

使用模板 dedic.tex，并配合施普林格的文档类 *SVMono*（用于专著类书籍）或 *SVMult*（用于编著论文集），即可在书籍开篇处设置引语或献辞

前言

使用模板*foreword.tex*，并搭配文档类 SVMono（专著类书籍）或 SVMult（编著书籍），为您的前言设置格式。

前言是由本书作者或编辑以外的人士撰写的介绍性文字，置于书籍正文之前（如适用）。前言位于由本书作者或编辑撰写的序言之前。

地点，月份 年份　　名字 姓氏

前言

本书改编自同一作者为西澳大利亚大学本科生动力系统课程编写的一系列讲义。虽然大部分内容保持不变，但为适应文本形式并提升清晰度，我们对内容作了若干调整和扩充。此书还会定期更新，以适应后续年份专题内容的变化，既丰富内容，也反映数学领域当前最前沿的发现与方法。

本书旨在以通俗易懂的语言，全面介绍动力系统及其应用的多种主题。具备动力系统方面的基础知识会有所帮助，但并非必需；读者只需熟悉微积分、线性代数和基础统计学等本科数学基础课程。最后四章收录了该领域近期发表的成果和研究活动中的多个拓展主题，有兴趣的读者可借此进一步阅读。本书还收录了一系列理论和计算练习与问题。这些问题是在多年使用原课程材料的过程中逐步整理而成的。

版本历史记录

1. v1.0 - 第一版（2024年4月29日）E. Tan 2. v1.1 - 新增关于储备池计算机和复现图的章节。（2024年5月29日）B. Thorne 3. v1.2 - 新增关于动力网络的章节（2025年5月12日）E. Tan 4. v1.3 - 新增关于序数时间序列分析的章节（2025年5月17日）E. Tan 5. v1.3.1 - 更新前言（2025年5月17日）E. Tan 6. v1.4 - 将手绘图更新为数字图（2025年6月25日）E. Tan 7. v1.5 - 第一轮编辑和整理（2025年7月1日）E. Tan 8. v1.6 - 在末尾添加练习题（2025年7月2日）E. Tan 9. v1.7 - 编纂成书（2025年7月2日）E. Tan

地点，
月份 年份

名字 姓氏
名字 姓氏

致谢

如果您希望将致谢部分单独设为一章，而不是将其作为前言的最后一部分，请将模板 *acknow.tex* 与文档类 SVMono（专著）或 SVMult（编著书籍）配合使用。

目录

1 动力系统理论. 1

1.1 动力系统.1 1.2 流与线性化.3 1.3 中心流形.8 1.4 中心流形计算.11 1.5 映射.13 1.5.1 庞加莱映射.13 1.6 周期性与混沌.14 1.6.1 周期行为与符号动力学.14 1.6.2 倍周期与混沌.16

2 数论与递推 .19

2.1 随机系统与确定性系统.19 2.2 克罗内克系统：分数数的动力学 .. . . . . 19 2.3 复现 .22

3 嵌入理论与重构 .25

3.1 嵌入定理.26 3.2 嵌入方法.29 3.2.1 时间延迟嵌入.29 3.2.2 导数嵌入.30 3.2.3 积分-微分嵌入.30 3.2.4 全局主值嵌入.31 3.3 均匀与非均匀延迟嵌入.31

4 嵌入实践.35

4.1 嵌入参数是否会影响参数？.35 4.2 测量嵌入质量.37 4.3 均匀延迟嵌入——同时选择 τ 和 m . . . . 38 4.3.1 Gao & Zheng——特征长度.38 4.3.2 摆动乘积.39

4.4 均匀延迟嵌入——分离选择.41 4.4.1 选择 τ.41 4.4.2 选择 m——全局伪近邻（GFNN） . . . . 46 4.5 非均匀延迟嵌入.47 4.5.1 Garcia 和 Almeida.48 4.5.2 连续性统计量.50 4.5.3 PECUZAL.51 4.5.4 其他方法.52

5 不变量 .53

5.1 遍历性与自然测度.54 5.2 李雅普诺夫指数.56 5.2.1 李雅普诺夫指数（繁琐的方法）.57 5.2.2 李雅普诺夫指数（简便的方法）.61 5.2.3 Wolf 算法.62 5.2.4 伪李雅普诺夫指数.66 5.3 分形维数与标度.66 5.3.1 盒计数维数.67 5.3.2 相关性与广义维数.68 5.3.3 Grassberger-Procaccia算法.69 5.3.4 Judd 算法.71 5.3.5 高斯核算法.72 5.3.6 信息维数.73 5.4 熵与信息论.73

6 预测与函数逼近 .77

6.1 使用动力系统建模.77 6.2 自回归过程.78 6.3 最近邻（类比法）.79 6.4 神经网络（谁会在乎参数的数量呢？）80 6.4.1 感知器.80 6.4.2 梯度下降算法.82 6.4.3 神经网络变体.83 6.5 符号回归（即：把神经网络交给代数学家／泛函分析学家）.84

7 模型选择 .87

7.1 赤池信息准则（AIC）.88 7.2 贝叶斯信息准则（BIC）.89 7.3 模型选择与编码理论.91 7.3.1 重新探讨消息传递问题.91 7.3.2 编码理论.92 7.3.3 编码树.93 7.3.4 最优编码.94 7.3.5 霍夫曼编码.97

7.3.6 自定界解码器.99 7.4 最小模型描述长度（MDL）.101 7.4.1 基于MDL的模型选择.107

8 Surro闸门 . 111
8.1 替代算法层级.112 8.1.1 算法0：随机置换替代数据.112 8.1.2 算法1：傅里叶变换替代序列.113 8.1.3 算法2：幅值调整傅里叶替代数据 . . . . . 114 8.2 其他替代数据算法 .115 8.2.1 截断傅里叶变换替代数据 . 115 8.2.2 小幅洗牌替代数据（SSS）.115 8.2.3 伪周期替代数据（PPS）. 116 8.2.4 吸引子网络替代数据.117 8.3 约束替代数据.117

9 储备池计算.119
9.1 回声状态性质.120 9.2 储备池设计.122 9.2.1 回声状态网络（ESN）. 123 9.2.2 其他储备池计算机.125 9.3 储备池计算机训练.126 9.3.1 时间序列预测.126 9.3.2 系统仿真.127 9.3.3 时间序列分类与储备池时间序列分析127

10 递归图 .131
10.1 构造注意事项.132 10.2 复现量化分析.133 10.2.1 点密度.133 10.2.2 对角线.134 10.2.3 垂直线.135 10.2.4 复现时间.136

11 动力网络 .139
11.1 耦合与同步.139 11.1.1 耦合函数.139 11.1.2 二体系统与同步.140 11.2 网络上的动力学.142 11.2.1 振子网络与奇美拉态.142 11.2.2 FitzHugh-Nagumo 奇美拉模型.144 11.3 意见动力学.145 11.3.1 社会物理学的动力网络方法.145 11.3.2 异步与同步模拟.147 11.3.3 DeGroot 模型与 Friedkin-Johnsen 模型.148 11.3.4 投票者模型.148

11.3.5 边界置信度模型 (BCM).149

12 序数时间序列分析 .151
12.1 符号动力学.151 12.2 序数方法.152 12.2.1 时间序列的序数表示.
152 12.2.2 序数划分.152 12.3 网络方法.154 12.3.1 序数划分转换网络.
154 12.4 复杂度度量.155 12.4.1 置换熵.156 12.4.2 条件排列熵.157
12.5 序数 Poincar´e 截面.157 12.5.1 Poincar´e 截面与首次返回映射.
157 12.5.2 Poincar´e 截面的序数方法.158

13 练习与计算实验.161
参考文献.171

第 1 章
动力系统理论

动力学指随时间发生的各种变化，无论大小。由于这一术语涵盖范围广泛，因此与物理世界中观察到的诸多现象都密切相关，且能对其进行描述。从经典牛顿物理学的基本原理，到混沌、复杂网络和奇美拉态，动力学无处不在，也一直是人们持续探究的主题。若想理解观测到的动力学所遵循的规则，那么求知者不可避免地会接触到动力系统理论，希望借此建立一定程度的逻辑结构。

动力系统理论的研究旨在理解并编纂支配动力学的数学规则。这样做的目的是为更好地理解人们在工作应用中遇到的观测到的动力学提供指导和工具。随着我们开始讨论这一主题，毫无疑问，我们必须首先对研究对象作出定义。

1.1 动力系统

定义 1.1. – 动力系统
动力系统是一个三元组（T， M， Φ），其中：

- T：时间指标
- M：状态空间
- Φ：演化算子

使得

$$\Phi : U \subset (T \times M) \to M$$

换言之，动力系统是由位于状态空间中的某个状态构成的数学对象，该状态的位置根据演化算子 Φ 给出的规则随时间（或任何有序变量）演变。给定一个初始条件

$x(0) = x_0$，演化算子表示为

$$\Phi_t(x(0)) = x(t) \tag{1.1a}$$

$$\Phi_{t+s}(x) = \Phi_t \circ \Phi_s(x) = \Phi_s \circ \Phi_t(x) \tag{1.1b}$$

这里，$x(t) \in M$ 表示动力系统在相空间中的轨迹（轨道），是时间的函数。动力系统的另一种表示形式是常微分方程（ODE）：

$$\dot{x} = F(x, t) \tag{1.2}$$

如果 F 与时间无关（即 $F(x,t) = F(x)$），则称动力系统为自治的。因此，轨迹的实现仅取决于当前状态和演化算子。上述常微分方程初值问题的解给出了轨迹 $x(t)$。动力系统主要有两种可视化方式：

1. 作为向量场 $\dot{x} = F(x)$ 2. 作为相空间中的轨迹 $x(t) \in M$（或者 $(x,t) \in M \times T$）

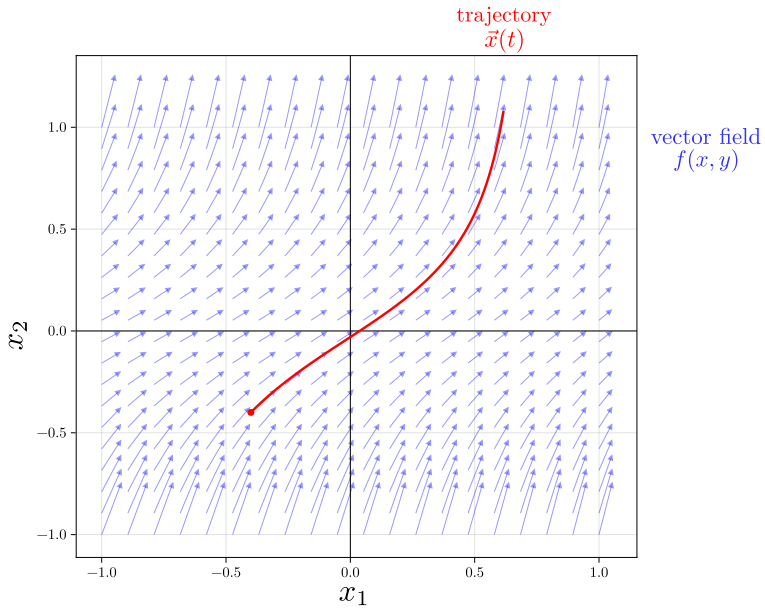


图 1.1 动力系统的向量场和轨迹

1.2 流与线性化

在大多数情况下，人们通常关注微分方程典型解的长期行为。通过分析微分方程及其迭代的不动点，可以获得有关这种行为的信息。具体而言，可以通过考察不动点的稳定性来确定长期行为。

定义 1.2——不动点
给定定义在空间 $\mathbb{R}^n$ 中的一般常微分方程 $\dot{x} = f(x)$。任何满足 $f(x^*) = 0$ 的点 $x^* \in \mathbb{R}^n$ 都是不动点。

在不动点 x^* 处， $f(x^*) = 0$ 这一性质意味着轨迹的速度为零，因此在所有时间内都无法离开该不动点 t。请注意，对于连续且可微的（C^1）系统，这同样适用于过去的历史 t 。上述定义使我们能够轻松计算不动点的位置。然而，它并未提供有关不动点周围邻域内稳定性和行为的任何信息。

考虑一维流最简单的情形（即 $x(t) \in \mathbb{R}$），并设相应的常微分方程为

$$\dot{x} = f'(x) = x(x+1)(x-1). \tag{1.3}$$

绘制该系统的相图，可以发现三个不动点位于 $x^* = -1$，0、1，其行为如下（见图 1.2）

- 在 $x = -1$，从不动点左侧出发的轨迹会加速向左运动。同样，从不动点右侧出发的轨迹会加速向右运动。 →**不稳定**
- 在 $x = 0$，从不动点左侧出发的轨迹会向右运动。同样，从不动点右侧出发的轨迹会向左运动。 →**稳定**
- 在 $x = 1$，行为与 $x = -1$ →**不稳定** 时相同

一般而言，对于可微的常微分方程，可以通过在不动点处计算导数来确定不动点的稳定性。

- $f\prime(x^*) > 0 \implies$ **不稳定**
- $f\prime(x^*) < 0 \implies$ **稳定**

对于方程 1.3， f 是一个简单函数，其导数易于求值。请注意， $f\prime(x)$ 也是非线性的（二次函数）。通过围绕每个不动点 x^* 对 f 作泰勒级数展开，可以简化利用导数推断稳定性的方法，

$$\dot{x} = f'(x^*)(x - x^*) + R_1(x, x^*). \tag{1.4}$$

如果 f 是 C^1 （即一次可微函数），高阶项的大小$||R_1$（x， x^*）$|| \to 0$ 比 $||x - x^*||$ 更快地趋于零，当 $x \to x^*$时。换句话说，动力学

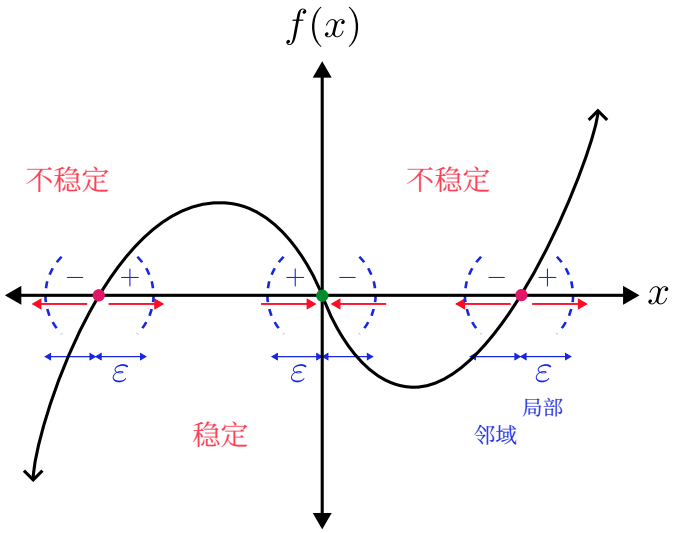


图 1.2 $\dot{x} = x(x+1)$ 的相空间图（$x-1$）。不动点附近的局部区域会出现有趣的行为。

在不动点邻域附近主要由一阶导数f'决定。因此，可以通过坐标变换 $u = x - x^*$ 将非线性常微分方程线性化，从而得到动力学的局部线性近似，

$$\dot{u} = Df(x^*)u, \tag{1.5}$$

其中 $Df(x) = f'(x)$。令 $Df(x^*) = A$，则具有初始条件 $u(0) = u_0$ 的轨迹所对应的近似线性系统的解可精确表示为：

$$u(t) = u_0 e^{tA}. \tag{1.6}$$

在方程 1.6 中，不动点的稳定性完全由 A 的符号决定。对于 $x \in \mathbb{R}^n$，这一结果通常也成立；最终的线性常微分方程用矩阵 $Df(x^*)$ 的特征向量和特征值表示。

$$\vec{u}(t) = \vec{v}_1 e^{t\lambda_1} + \vec{v}_2 e^{t\lambda_2} \ldots + \vec{v}_1 e^{t\lambda_n}. \tag{1.7}$$

定理 1.1 – 线性化定理
设 f：$\mathbb{R}^n \to \mathbb{R}^n$ 是一个C向量场，且 x^* 是一个不动点。如果 $A = Df(x^*)$ 的所有特征值实部均非零，那么存在 $\exists \epsilon > 0$ 使得 $\forall x$、$||x - x^*|| < \epsilon$、$\dot{x} = f(x)$ 与 $\dot{y} = Ay$ 拓扑等价。

你可以试着自行推导这个表达式，这会是一个有益的练习。注意，只有当 $\lambda_i \neq 0$ 对所有 i均成立时，这种表述才有用。如果确实如此，相应的不动点就称为**双曲型**。

定义 1.3. – 稳定与不稳定流形
考虑 $\dot{x} = f(x)$ 和 $x \in \mathbb{R}^n$，其中 x^* 是一个以原点为中心的双曲不动点，而 $A = Df(x^*)$ 是 f 的雅可比矩阵。

令 $\lambda_1, \ldots \lambda_n$ 和 $\vec{v}_1, \ldots, \vec{v}_n$ 分别为A 的有序特征值和特征向量，使得

$$\lambda_1 < \lambda_2 < \ldots < \lambda_i < 0 < \lambda_j, \ldots \lambda_n.$$

Then

- $E_s = \text{span}(\vec{1}, \ldots, \vec{v}_i)$ is the stable manifold
- $E_u = \text{span}(\vec{v}_j, \ldots, \vec{v}_n)$ is the unstable manifold

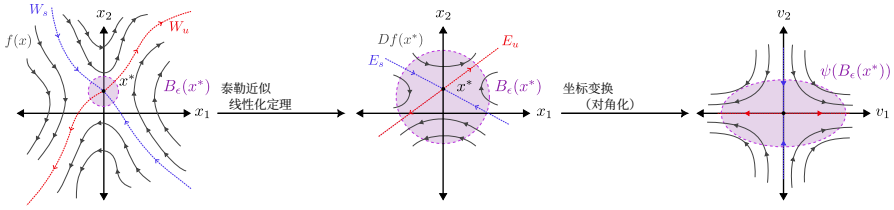


图 1.3 关于不动点的线性化

稳定与不稳定流形定义了轨迹趋近不动点时的切向方向。双曲不动点大致可分为以下几种情况：

- 稳定结点 $\lambda_1 < \ldots < \lambda_n < 0$
- 不稳定结点 $0 < \lambda_1 < \ldots < \lambda_n$
- 鞍点 $\lambda_1 < \ldots < \lambda_p < 0 < \lambda_q, \ldots \lambda_n$
- 稳定焦点 $\text{Re}(\lambda_1) < \ldots < \text{Re}(\lambda_n) < 0$ 和 $\text{Im}(\lambda_i) \neq 0$
- 不稳定焦点 $\text{Re}(\lambda_1) > \ldots > \text{Re}(\lambda_n) > 0$ 和 $\text{Im}(\lambda_i) \neq 0$

上述分类还伴随着关于常微分方程动力学存在性与唯一性的一些理论保证。

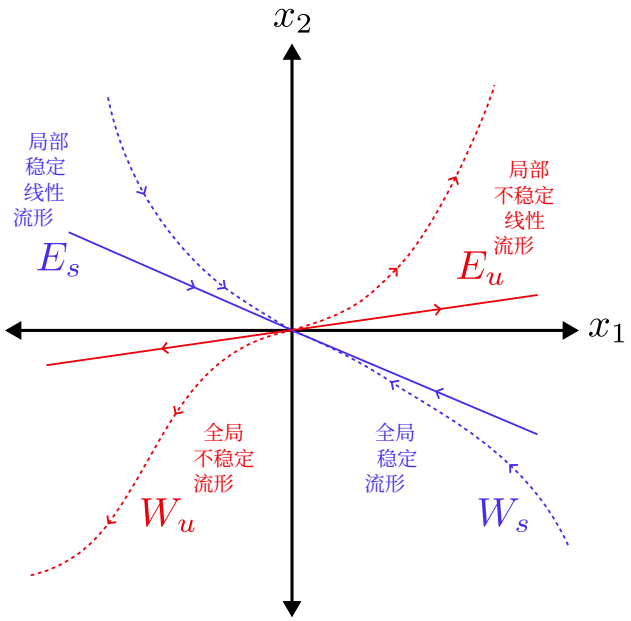


图 1.4 全局与局部流形

定理 1.2. – 流的存在唯一性
如果 f： $\mathbb{R}^n \to \mathbb{R}^n$ 是 C类的向量场，那么对于所有满足 $\dot{x} = f(x)$ 具有唯一解 $\psi(t, x_0)$ 的 $x_0 \in \mathbb{R}^n$、 $\exists T > 0$，其中$\psi(0, x_0) = x_0$。因此：

1. 两个不同的状态不能在同一时间演化到同一个状态 *2.* 如果两个不同的状态确实演化到同一个状态，那么它们位于同一条轨道上 *3*

· 轨迹永不相交 ，相交，也不会合并

该定理保证，只要 f 性质良好（一次可微），就总存在唯一的轨迹解在局部经过每个点。对于 f 不是 C^1的情况，这一条件并不成立。为说明这一点，我们考虑一个违反解的唯一性条件的病态例子。

考虑以下初值问题，

$$\dot{x} = \sqrt{|x|}, \quad x(t_0) = x_0. \tag{1.8}$$

显然， $f(x)$ 在点 $x = 0$ 处不可微，因此我们预期存在性与唯一性的保证不再适用。为说明这一点，我们考虑一条从不动点 $x_0 = 0$ 出发的轨迹。该轨迹的显式解

可以通过如下分离变量法求得,

$$\frac{1}{\sqrt{x}}\,dx = dt$$
$$x(t) = \left(\frac{t+C}{2}\right)^2.$$

代入初始条件 $x(t_0) = 0$，求出积分常数 C，得到以下轨迹解

$$x(t) = \frac{(t-t_0)^2}{4}. \tag{1.9}$$

上述结果表明出现了矛盾：相同的初始条件 $x(t_0) = 0$演化成了两条不同的轨迹。在第一种情况下，$\dot{x} = 0$ 是不动点，因此对于所有 $t > t_0$都应保持为 $x = 0$。然而，显式轨迹解表明，当 $t > t_0$ 时，轨迹 $x(t)$ 会自发离开不动点 $x = 0$，并沿着二次轨迹前进。解的不唯一性直接源于向量场在不动点处不可微。

$$\dot{x} = \frac{1}{x^2} \tag{1.10}$$

定理 1.3. – 解的连续性

若 $f\colon\ \mathbb{R}^n \to \mathbb{R}^n$ 是 C^1类向量场，则对于所有满足以下条件的 $x_0 \in \mathbb{R}^n$、$\exists \epsilon > 0$和 $T > 0\colon\ \dot{x} = f(x)$ 在 $-T < t < T$和$||x - x_0|| < \epsilon$上有解 $\phi(t,x)$，其中 $\phi(0,x) = x$，且 $\psi(t,x)$ 关于 t 和 x 连续。

定理 1.3 保证在局部邻域内，状态会平滑演变并连续变形（即位置不会发生跳跃）。需要强调的是，这些保证仅适用于空间和时间上的局部范围，并不一定能保证解对所有 $T > 0$ 或 $\epsilon > 0$都存在。以下系统示例可以说明这一点

$$\dot{x} = x^2, \quad x(t_0) = x_0. \tag{1.11}$$

由于该常微分方程是 C^1，我们可以保证轨迹对所有初始条件 x_0都存在且唯一。这些解轨迹可以通过分离变量显式求解。

$$x(t) = \frac{1}{\frac{1}{x_0} + (t_0 - t)}. \tag{1.12}$$

我们可以看到，$t^* = \frac{1}{x_0} + t_0$ 处存在一个奇点，且轨迹会在有限时间内趋于无穷大（有限时爆破）。因此，$t < t^*$ 的解在 $t > t^*$处不连续。

定义 1.4. – 拓扑共轭（等价）
令 t here b^{e} 两个 dynamical 系统

- $\dot{x} = f(x)$， $x \in \mathbb{R}^n$ 其解为 $q(t,\ x)$
- $\dot{y} = g(y)$， $y \in \mathbb{R}^n$ 其解为 $r(t,\ y)$

如果存在坐标变换 $\Psi(x)$，它是一个可微的一一映射 $\Psi:\ U \to \mathbb{R}^n$，使得对所有 x 和 t， $r(t,\ \Psi(x)) = \Psi(q(t,\ x))$，且 $q(t,\ x)$ 在 U 中存在，那么这两个动力系统在定义域 $U \subseteq \mathbb{R}^n$ 上**等价（拓扑共轭）**。

定理 1.4. – 流的基本定理
若 $f:\ \mathbb{R}^n \to \mathbb{R}^n$ 是 C类向量场， $x_0 \in \mathbb{R}^n, \epsilon > 0, f\ (x) \neq 0$ 对于 $||x - x_0|| < \epsilon, v \in \mathbb{R}^n, v \neq 0$成立，则存在一个可微坐标变换 $y = \Psi\ (x)$，适用于 $||x - x_0|| < \epsilon$，使得 $\dot{x} = f\ (x)$ 与$\dot{y} = v$等价。因此：

1. 远离不动点的轨迹看起来都像以恒定速度平行流动（即直线）。
2. 有趣的行为只会出现在不动点附近。

1.3 中心流形

用于确定稳定性的线性化方法仅适用于双曲不动点（即特征值非零的不动点）。要确定非双曲不动点的稳定性，我们需要采用稍微复杂一些的方法。

定义 1.5 – 中心流形（特征子空间）
设 $\dot{x} = f(x)$ 是 $\mathbb{R}^{s+u+c}$ 中的一个非线性动力系统，其对应的线性化流为 $\dot{y} = Ay$。设 $\Lambda_- = \{\lambda_1^-,\ \ldots\ldots,\ \lambda_s^-\}$、 $\Lambda_+ = \{\lambda_1^+,\ \ldots\ldots,\ \lambda_u^+\}$、$\Lambda_0 = \{\lambda_1^0,\ \ldots\ldots,\ \lambda_c^0\}$ 分别为实部为正、负和零的特征值的有序集合。因此，设 $V_- = \{\vec{v}_1^-,\ \ldots\ldots,\ \vec{v}_s^-\}$、 $V_+ = \{\vec{v}_1^+,\ \ldots\ldots,\ \vec{v}_u^+\}$ 和 $V_0 = \{\vec{v}_1^0,\ \ldots\ldots,\ \vec{v}_c^0\}$ 为相应的特征向量。由此将稳定流形定义为

- 稳定流形 $E^s =$ 张成(V_-)
- 不稳定流形 $E^u =$ 张成(V_+)
- 中心流形 $E^c =$ 张成(V_0)

根据线性化理论，我们知道，双曲不动点的稳定性可以通过稳定与不稳定流形来确定。然而，对于非双曲不动点，线性化这种简化方法过于粗略，无法得出任何有用的或

准确的结论。与线性化试图将动力学约化为在不动点上不变类似，中心流形理论试图将动力学局部化于某个经过不动点的不变非线性**中心流形** 上。

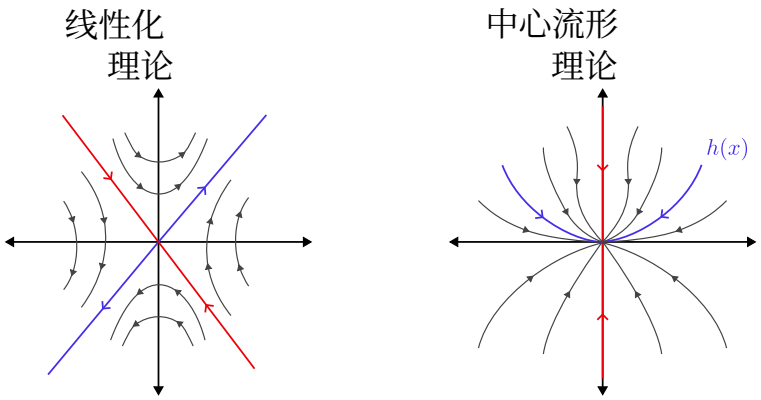


图 1.5 线性化与中心流形理论

考虑由以下向量场定义的一般动力系统

$$\dot{x} = Ax + f(x, y), \tag{1.13a}$$
$$\dot{y} = By + g(x, y) \tag{1.13b}$$

其中 $(x,\ y) \in \mathbb{R}^c \times \mathbb{R}^s$ 满足以下边界条件：

$$f(0,0) = 0, \qquad Df(0,0) = 0, \tag{1.14a}$$
$$g(0,0) = 0. \qquad Dg(0,0) = 0. \tag{1.14b}$$

这个条件使我们关注位于原点、且一阶导数为零的不动点（注意，通过应用线性坐标变换，动力系统中的任何不动点都可以写成方程 1.13 和 1.14 的形式。在上述系统中，与矩阵 A 对应的前 c 个分量描述中心流形（即实部为零的特征值），矩阵 B 描述特征值实部为负的分量。一般而言，也可以将结果推广到所有非零特征值分量的情形。该结果对应一个鞍点，这在当前讨论中并无特别意义。

定义 1.6. – 中心流形
对于上述系统，如果一个不变流形可以在局部表示如下，则它是中心流形:

$$W^c(0) = \{(x, y) \in \mathbb{R}^c \times \mathbb{R}^s | y = h(x), |x| < \delta, h(0) = 0, Dh(0) = 0\}$$

当 δ 足够小时。见图 1.5

为了确定不动点的稳定性，我们需要确定中心流形的稳定性。值得庆幸的是，[1] 提供了几个有用的定理来解决这一问题。

定理 1.5. – 中心流形的存在性
对于由方程 *1.13* 和 *1.14* 给出的系统，存在一个 C^r 中心流形。当 u足够小时，系统在中心流形上的动力学由以下 c维向量场给出，

$$\dot{u} = Au + f(u, h(u)), \quad u \in \mathbb{R}^c.$$

定理 1.6– 中心流形的稳定性
对于限制在中心流形上的动力系统。

1. 假设零解（不动点）是稳定的（同理，对于渐近稳定/不稳定的情形），那么原动力系统的零解也稳定（渐近稳定/不稳定）。*2.* 假设受限系统的零解是稳定的。那么，如果$(x(t),\ y(t))$ 是原系统的一个解，且$(x(0),\ y(0))$ 足够小，则受限系统存在一个解 $u(t)$，使得当 $t \to \infty$

$$x(t) = u(t) + O(e^{-\gamma t}),$$
$$y(t) = h(u(t)) + O(e^{-\gamma t})$$

其中 $\gamma > 0$ 是一个常数。

定理的第一个结论指出，可以通过确定中心流形的稳定性来推断原系统的稳定性。也就是说，中心流形近似足以表示动力学。第二个结论指出，只要将分析限制在不动点的一个足够小的邻域内，这一结果就成立。在这种情况下，中心流形能够很好地表示我们的动力学，误差会按指数函数衰减，如 $t \to \infty$所示。

1.4 中心流形计算

现在我们已做好充分准备，可以开始计算中心流形$h(x)$。完成后，我们便可以利用这些结果进一步推断非双曲型不动点的稳定性。

假设我们有一个不变中心流形

$$W^c(0) = \{(x, y) \in \mathbb{R}^c \times \mathbb{R}^s | y = h(x), |x| < \delta, h(0) = 0, Dh(0) = 0\} \tag{1.15}$$

1. 任何位于 $W^c(0)$ 上的点，其（x, y）坐标必须满足

$$y = h(x)$$

2. 对时间求导意味着，任何位于$W^c(0)$ 上的点，其（$\dot{x}, \dot{y}$）坐标必须满足

$$\dot{y} = Dh(x)\dot{x}$$

3. 对于原动力系统,

$$\begin{aligned} \dot{x} &= Ax + f(x, h(x)), \\ \dot{y} &= Bh(x) + g(x, h(x)) \end{aligned}$$

代入时间导数 $\dot{y}$，得到以下条件

$$\mathcal{N}(h(x)) \equiv Dh(x)[Ax + f(x, h(x))] - Bh(x) - g(x, h(x)) = 0. \tag{1.16}$$

假设 $h(x)$[2] 可表示为幂级数展开，即可求解该表达式 N。

作为上述中心流形固定点稳定性计算的示例，我们考虑 Wiggins 等人给出的以下系统。 [2],

$$\dot{x} = \frac{x}{2} + y + x^2 y, \tag{1.17a}$$

$$\dot{y} = x + 2y + y^2. \tag{1.17b}$$

在本例中，我们尝试确定位于原点 $(x, y) = (0,0)$ 的不动点的稳定性。首先尝试通过在 $(0,0)$ 处进行线性化来确定稳定性，可得到以下雅可比矩阵,

$$Df(x, y) = \begin{bmatrix} \frac{1}{2} + 2xy & 1 + x^2 \\ 1 & 2 + 2y \end{bmatrix}, \tag{1.18}$$

$$Df(0, 0) = A = \begin{bmatrix} \frac{1}{2} & 1 \\ 1 & 2 \end{bmatrix}. \tag{1.19}$$

矩阵 A 的特征值和特征向量分别为 $(\lambda_1, \lambda_2) = (0, 5/2)$ 和 $v_1 = (-2, 1)$, $v_2 = (1,2)$。由此可知，该不动点确实是非双曲的，并且沿 v_2 方向存在不稳定流形。然而，线性近似无法揭示 v_1 方向上的稳定性。为确定这一点，我们首先使用计算得到的特征向量进行坐标变换，

$$x = -2u + v, \tag{1.20a}$$
$$y = u + 2v, \tag{1.20b}$$

其中， u 和 v 分别对应中心流形方向和不稳定方向。因此，式 1.17 中的原始常微分方程可以用 u 和v表示，

$$\dot{u} = \frac{2}{5}(-2u + v)^2(u + 2v) + \frac{1}{5}(u + 2v)^2 = g_1(u, v), \tag{1.21a}$$
$$\dot{v} = \frac{5}{2}v + \frac{1}{5}(-2u + v)^2(u + 2v) + \frac{2}{5}(u + 2v)^2 = g_2(u, v). \tag{1.21b}$$

我们选择一个多项式近似来表示中心流形，

$$h(u) = au^2 + bu^3 + O(u^4), \tag{1.22}$$

其中，$O(u^4)$表示 4th 阶或更高阶项，当 u 较小时可忽略不计（即接近不动点时）。将其代入式 1.16 给出的条件，可得以下方程，

$$(2au + bu^2 + ...)[0 + \frac{2}{5}(-2u + au^2 + bu^3 + ...)^2(u + 2au. . . . . 2 + 2bu^3 + ...) + \frac{1}{5}(u + 2au^2 + 2bu^3 + ...)^2] = 0. \tag{1.23}$$

去掉所有高阶项 $O(u^4)$，并令 u^2 项和 u^3项的系数相等，可得 $a = -\frac{4}{25}$和 $b = -\frac{116}{625}$的值，因此中心流形近似为

$$h(u) = -\frac{4}{25}u^2 - \frac{116}{625}u^3 + O(u^4). \tag{1.24}$$

将$h(u)$代入 $g_1(u,h(u))$，以便将分析限制在中心流形上的动力学，可得以下简化方程，

$$\dot{u} = \frac{1}{5}u^2 - \frac{8}{5}u^3 - \frac{12}{5}au^4 + O(u^5), \tag{1.25}$$

当 $u < 0$ 时稳定，而当 $u > 0$时不稳定。因此，我们可以得出结论：原点处的不动点是不稳定的。

1.5 映射

如前所述，流可以理解为与某个常微分方程（ODE）相关联的一族连续解轨迹，

$$\dot{x} = f(x), \quad x \in M \subseteq \mathbb{R}^n. \tag{1.26}$$

类似地，映射是流的一种离散类比，可理解为与某个差分方程（DE）相关联的离散解序列族，

$$x_{n+1} = f(x_n). \tag{1.27}$$

流的许多概念也适用于映射，相关定义如下。

定义 1.7——映射的不动点与稳定性
设有一个映射 $x_{n+1} = f(x_n)$，其中 $x_n \in M$。不动点 x^* 是满足下式的任一解，

$$x^* = f(x^*)$$

不动点的稳定性由 $f\prime(x^*)$ 的幅值决定，

- $|f'(x^*| < 1 \implies$ 稳定
- $|f'(x^*| > 1 \implies$ 不稳定

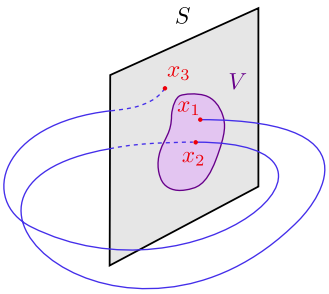
1.5.1 庞加莱映射

通过定义庞加莱曲面截面 S，可以将连续流与离散映射的动力学联系起来。具体做法如下：

1. 定义一条轨迹 $x(t) \in \mathbb{R}^n$ 以及一个开截面 $S \in \mathbb{R}^n$，其中 S 的维数为$n-1$。
2. 令 $t_1 < t_2 < ...$ 表示 $x(t)$与 S 相交时的时间（即 $x(t_i) \in S$）。3. 定义一个新序列，其各项取值为

$$y_n = x(t_n)$$

从本质上说，庞加莱截面使我们能够用离散映射来表示连续流的动力学。正如我们将在后续章节中看到的，这是一个有用的工具，可以轻松地将映射的研究结果推广到连续流的研究中。



Poincaré Map

$P : V \to S,\ V \subseteq S$

$x_{n+1} = P(x_n)$

图 1.6 Poincaré 映射的构造

1.6 周期性与混沌

1.6.1 周期行为与符号动力学

考虑逻辑斯蒂映射的情形:

$$x_{n+1} = r x_n (1 - x_n) \tag{1.28}$$

对于参数 $r \in [0, 4]$。简单计算表明，在 $x_1 = 0$ 和 $x_2 = \frac{r-1}{r}$ 处存在两个不动点。这些解的稳定性由 $f_r(x, r)$ 的幅值决定，而幅值又取决于 r,

$$|f'(x, r)| = |r(1 - 2x)|. \tag{1.29}$$

练习： 确定在 r 的不同取值下这些不动点的存在性和稳定性。

根据上述内容，我们对周期点的存在性可以得出什么结论?

定义 1.8. – 周期点
对于由函数 $x_{n+1} = f(x_n)$ 给出的映射，点 a 的周期为 $n \in \mathbb{N}$，如果,

$$a = f^{(n)}(a), \quad a \neq f^{(i)}(a) \quad \forall i \neq kn, k \in \mathbb{Z}$$

类似地，对于演化算子为 Φ_t 的流，点 a 的周期为 $T \in \mathbb{R}$，$T > 0$，如果,

$$a = \Phi_T(a), \quad a \neq \Phi_{t(a)} \quad \forall t \neq kT, k \in \mathbb{Z}$$

请注意， n-周期点与 $f^{(n)}$ 的不动点相对应。所有观测值的轨迹（序列）都位于连续区间 $[0, 1]$ 的某处。另一种观察动力学的方法是查看给定值是否位于

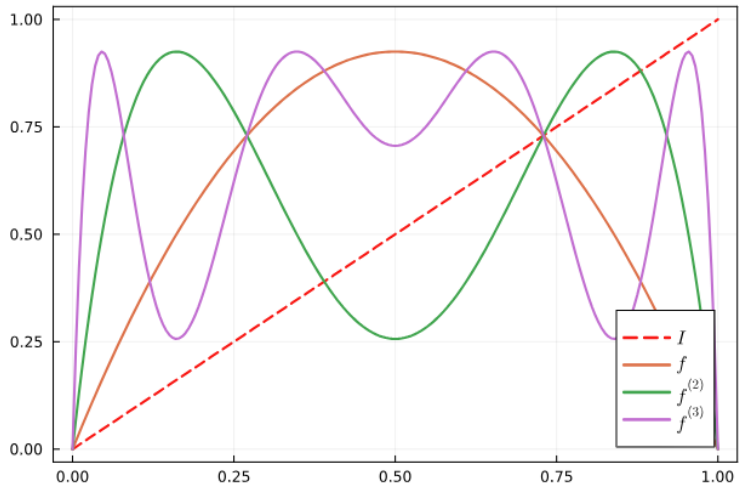


图 1.7 逻辑斯蒂函数的迭代

序列位于区间 $x_n < 0.5$ 的左半部分，还是右半部分 $x_n \geq 0.5$。通过以这种方式对序列进行划分，我们可以构造另一个由符号组成的序列 y_n（例如 L 和 R），

$$y_n = \begin{cases} L, & x_n < 0.5 \\ R, & x_n \geq 0.5 \end{cases} \tag{1.30}$$

事实证明，这种由离散符号（而非连续值）组成的轨迹表示保留了逻辑斯蒂映射的全部动力学性质。x_n 中的周期动力学对应于 y_n中的重复模式。将状态空间划分以生成符号动力学的方法，是一种在保留系统动力学行为的同时降低轨迹复杂性的有用工具。符号动力学也构成了解释混沌行为出现的基础，正如我们稍后将要展示的那样。然而，划分状态空间的方法选择并不总是明确的，这将在第12章中进一步讨论。

回到最初的问题，对于 r的不同取值，关于周期点的存在性我们能说些什么？为此，我们可以绘制逻辑斯蒂映射的分岔图。

从分岔图可以看出，周期点仅在 $r > 3$时存在。随着 r 的增大，逐渐形成周期更长的周期点（每次周期都加倍），直到大约 $r = 3。7$，此时轨迹似乎覆盖了状态空间的一个区间，表明出现了混沌行为。

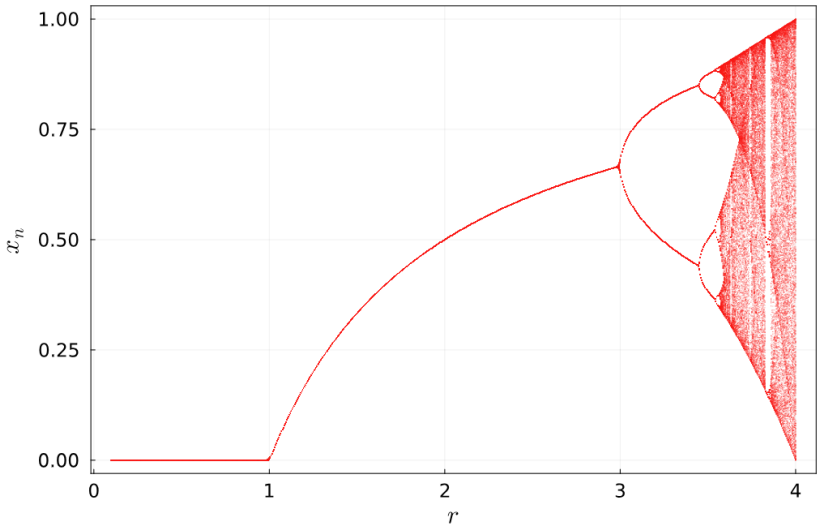


图 1.8 逻辑斯蒂映射的分岔图

1.6.2 倍周期与混沌

到目前为止，我们只是简要提及了**混沌**这一主题，却没有给出明确的定义。在动力系统文献中，混沌有许多定义和性质。以下列出一些常见的定义和性质：

- 对初始条件的敏感依赖
- 确定性的，但不可预测
- 正李雅普诺夫指数
- “蝴蝶效应”
- 有界、确定性且非周期
- 拓扑混合
- 分形不变集的存在性

最值得注意的是，对初始条件的敏感依赖（SDIC）的定义及其更为通俗的名称蝴蝶效应很有用，因为它描述了初始条件接近时轨迹的指数发散行为。也就是说，初始条件存在极微小差异的轨迹，随着演化会变成指数发散的轨迹。对于流，混沌行为要求状态空间至少具有 3 维，才能实现充分的拓扑混合和非周期性，同时避免轨迹交叉。然而，对于离散映射，只需 1 维即可产生混沌。

为满足好奇的读者，我们提供一些关于混沌流和离散映射的示例：

例 *1.1*。 **– 洛伦兹（1963） [3]**

$$\dot{x} = \sigma(y - x) \tag{1.31a}$$
$$\dot{y} = -xz + rx - y \tag{1.31b}$$
$$\dot{z} = xy - bz \tag{1.31c}$$
$$\tag{1.31d}$$

对于参数 $\sigma = 16,\quad b = \frac{8}{3}, -\ r = 28$

例 *1.2*。 **– R¨ossler（1976） [4]**

$$\dot{x} = -y - z \tag{1.32a}$$
$$\dot{y} = x + ay \tag{1.32b}$$
$$\dot{z} = b + z(x - c) \tag{1.32c}$$
$$\tag{1.32d}$$

对于参数 $a = 0。1,\quad b = 0。1,\quad c \in [4, 20]$

混沌映射示例：

例 *1.3.* **– 海农映射 [5]**

$$x_{n+1} = 1 + y_n - ax_n^2 \tag{1.33a}$$
$$y_n = bx_n \tag{1.33b}$$

当参数为 $a = 1。4,\quad b = 0。3$ 时

例 *1.4.* – **Ikeda 映射 [6]**

$$z_n = x_n + iy_n \tag{1.34a}$$

$$z_{n+1} = p + Bz_n \exp\{i\kappa - i\frac{\alpha}{1+|z_n|^2}\} \tag{1.34b}$$

当参数为 $p = 1.0$， $B = 0.9$， $\kappa = 0.4$， $\alpha = 6.0$ 时

混沌行为的主要特征之一是存在有界但非周期的轨迹。从逻辑斯蒂映射（以及一般而言，整个单峰映射族）可以看出，混沌行为是由依次发生的多次倍周期分岔引起的。随着 r 增大，周期行为按照整数的沙尔科夫斯基序，从长度为 2， 4，8，……，3 依次增加，直到达到某个临界分岔值 r^*；此后，系统进入完全混沌状态。这一视角也导出了一个联系周期性与混沌的关键定理。

定义 1.9. – 混沌函数
若函数映射 f 对任意 n 都存在一个周期为 n 的轨道，则称其为混沌的。

定理 1.7. – 周期三蕴含混沌
设由函数 f 给出一个保测映射。若 f 中包含一个周期为 3 的点 a，则 f 中包含所有周期 $n \in \mathbb{N}$ 的点。因此， f 是混沌的。

回想一下，庞加莱截面可以将任意流表示为离散映射。因此，如果连续系统的离散庞加莱映射呈现倍周期分岔，并最终出现周期3，这意味着原连续系统中存在所有可能周期长度的轨道。此外，这些周期轨道在状态空间中稠密（即，对于状态空间中的任意非周期点 M，它都可以与某个周期轨道任意接近），从而产生非常复杂的行为。

第2章
数论与递推

2.1 随机系统与确定性系统

动力系统的三元组表述（T， M， Φ）涵盖范围很广，并未对演化算子 Φ 的性质给出明确规则。如果 Φ包含随机过程（例如，从均匀分布中随机生成后续观测值序列），它仍然是一个有效的动力系统（尽管是随机的）。显然，随机过程会对动力系统许多便捷保证所依赖的 C^1 假设造成问题。另一方面，我们也可以定义一个完全确定性的系统（例如一个均匀递增的数列），它具备连续性、解的存在性与唯一性等所有有用特征。这样的系统如果足够简单，就会产生非常可预测的行为，不具备人们可能希望研究的那些更有趣、更复杂的动力学。

在上述上下文中，混沌系统可以理解为处于确定性与随机性边界上的情形。它们可以通过解析方法完全描述（具有确定性），但同时又表现出拓扑混合和轨迹发散，因此时间上相距甚远的观测值高度不相关，显得几乎随机。更重要的是，这种表面上的“随机性”源自确定性规则，而非初始条件或演化算子中内在的随机性。本章将通过一个取自数论的例子来探讨这种区别。

2.2 克罗内克系统：分数数的动力学

设有两个函数 frac(α) 和 int(α)，它们以正实数α 为输入，分别返回小数部分（即 $\alpha \bmod 1$）和整数部分（即 $\lfloor \alpha \rfloor$）。据此，考虑如下移位映射：

$$\tau_\alpha(x) = \text{frac}(x + \alpha) \tag{2.1a}$$
$$x_{n+1} = \tau_\alpha(x_n) \tag{2.1b}$$

其中 α 是预先给定的常数。这个动力系统称为克罗内克系统，可表示为由三元组 $(\mathbb{Z},\ [0,1],\ \tau_\alpha)$ 定义的动力系统。

引理2.1. – 克罗内克系统
设有一个离散映射，定义如下：

$$x_{n+1} = \tau_\alpha(x_n),$$

其中，τ_α 是关于 $\alpha \in [0,1]$的移位映射。

1. τ_α（x） 是非周期的 $\forall\alpha \in [0,1] \setminus \mathbb{Q}$*2.* τ_α（x） 是最终周期的 $\forall\alpha \in [0,1] \cap \mathbb{Q}$

克罗内克系统的行为完全取决于 α的选择。移位映射 τ_α 实质上生成一个分数序列，该序列按步骤截去 α 小数点后的前 n 位，并返回余数（初始条件 x_0 仅仅是增加了一个移位）。请注意，无理数的小数（n进制）小数部分永不重复。因此，α 的无理数值将产生一个永不重复的序列 x_n。反过来也成立：如果 α 是有理数，其小数（n进制）表示中包含一个周期序列，因此 x_n 也将是周期性的，除去最初的暂态阶段。

克罗内克系统十分有趣，并引出了几个问题：

问题1。x_n 可能取哪些值？**问题2**。我们能如何描述 x_n 在定义域 $[0,1]$上的分布？**问题3**。克罗内克系统是否被认为是混沌的？

第一个问题的答案由克罗内克定理给出。

定理2.1. – 克罗内克定理
设 $\alpha \in \mathbb{R} \setminus \mathbb{Q}$。那么，$\forall U \subseteq [0,1], U \neq \emptyset, \exists m \in \mathbb{N}$ 使得 $frac(m\alpha) \in U$

简单来说，克罗内克定理指出，对于 $[0,1]$ 中的任意子区间以及任意无理数 α，都存在 α 的某个倍数，其小数部分落在该子区间内。因此，我们可以在 $[0,1]$ 上定义宽度为 ϵ 的子区间，并证明克罗内克系统必然会访问距离 ϵ 以内的任意数值。证明克罗内克定理有多种方法。一种直观的方法如下：

1. 注意，当 $\alpha < 0.5$ 时，移位映射会使一个区间在数轴上向右移动 α，即移动 I_n。同样地， $\alpha > 0.5$ 会使 x_n 在数轴上向左移动 $1 - \alpha$。2. 不失一般性，令 α 较小。对于移位映射的少数几次迭代， I_n 线性增加，而 frac() 不会改变轨迹。3. 令 k 为 I_n 的一部分首次超过边界值 1 的时刻。因此，移位映射会使 I_n 的某些部分跳到数轴左侧，其取值由对 α 进行小数移位和截断得到。4. 如果 α 是无理数，那么截断和连续相加应当包含小数的每一种可能的子序列。因此，通过对区间 I_n 重复这一过程，根据鸽巢原理，整个单位区间最终将被未来区间 I_n 与 ´ 相对的像完全覆盖。5. 对 $1 - \alpha$ 的情况重复上述过程。

我们已经确定，含有无理数 α 的克罗内克系统最终会访问 $[0, 1]$ 中的每一个数。那么，对于 **Q2** –点的分布，我们能说些什么呢?

定理 2.2。 – 韦尔定理
设 $\alpha \in \mathbb{R} \setminus \mathbb{Q}$。则序列 $X = \{frac(n\alpha)\}_{n=1}^{N}$ 在 $[0, 1)$ 中分布，使得 $\forall J \subseteq [0, 1)$,

$$|X \cap J| \to |J|$$

或者，前 n 项中属于 J 的项所占的比例在 $n \to \infty$ 时存在极限，并且等于 $\mu(J)$，其中 μ 是长度测度。参见 *Nillsen*，以获取出色的证明和讨论 *[7]*。

本质上，韦尔定理意味着，对于克罗内克系统，点的序列在单位区间上均匀分布（即，每个宽度为 ϵ 的唯一子区间都被访问同样多次）。

上述讨论引导我们探讨**Q3**：克罗内克系统是混沌的吗？解决这一问题的一种方法是从信息损失的角度出发。取$x_0 = 0$和较小的无理数值α（例如$\frac{\pi}{10}$）。对于移位映射连续迭代的较短时期，下位置 x_{n+1} 可以由 x_n“预测”。它在每一步以α 线性增长。或者，只要我们处于序列线性增长的区间内，已知位置 x_n 就能提供关于先前值 x_{n-1}， x_{n-2}.....的完整信息。从这个意义上说，克罗内克系统确实包含类似混沌的性质（即拓扑混合），我们定义如下：

定义 2.1. – 拓扑传递性
设 S 为一个区间，且 $f: \ S \to S$。如果 $\forall x, \ y \in S$ 以及 $\epsilon > 0$、 $\exists z \in S$ 和 $n \in \mathbb{N}$ 使得$S, \ f$，则系统$S, \ f$是拓扑传递的，

$$|z - x| < \epsilon, \quad |f^{(n)}(z) - y| < \epsilon$$

反之，如果一个系统不具有传递性，那么存在 $\exists S_1, S_2 \subseteq S$ 使得

- $S = S_1 \cup S_2, S_1 \cap S_2 = \emptyset$
- $f^{(n)}(S_1) \subseteq S_1, f^{(n)}(S_2) \subseteq S_2 \ \forall n$

拓扑传递性（拓扑混合）这一性质表明，对于状态空间中的任意两个点的邻域，总能找到连接这两个邻域的轨迹，且准确性可达到任意程度 $S \ \epsilon$。不具有拓扑传递性的系统包含一些子集，它们之间不存在相互连接的轨迹。

然而，混沌系统的一个关键特征是对初始条件的敏感性（SDIC）。为了最终确定克罗内克系统是否为混沌系统，我们需要给出 SDIC 的定义。

定义 2.2. – 对初始条件的敏感依赖（SDIC）
如果 $\exists \delta > 0$ 使得$\forall x \in S$ 且 $\epsilon > 0$、 $\exists z \in S$ 和 $n \in \mathbb{N}$，其中一个系统（$S, \ f$）就对初始条件敏感（SDIC）。

$$|x - z| < \epsilon, \quad |f^{(n)}(x) - f^{(n)}(z)| > \delta$$

克罗内克系统是否基于 SDIC 而具有混沌性？（是）

2.3 复现

动力系统的另一个特征是复现。回想一下，通过将轨迹与一个定义好的庞加莱截面的连续交点识别为同一点，任何连续流都可以化为离散映射。由此，可以提出几个问题：

1. 如何定义复现？2. 关于动力系统中状态的复现，可以得出什么结论？

我们首先通过给出动力系统中状态复现的数学定义来回答第一个问题。

定义 2.3. – 回归点与区间
设 $(S,\ f)$ 是一个动力系统。如果存在 $\exists n \in \mathbb{N}$ 使得，则点 $x \in S$ 在界限ϵ 内是回归的

$$|x - f^{(n)}(x)| < \epsilon.$$

类似地，若对于 $J \subset S$，其中 $\mu(J) > 0$，$\exists x \in J$ 且存在 $n \in \mathbb{N}$使得

$$f^{(n)}(x) \in J.$$

或者，存在 $\exists n \in \mathbb{N}$使得

$$f^{(n)}(J) \cap J \neq \emptyset.$$

请注意，上述定义并未区分回归点的数量或 n 是否有限。可以定义这样一种动力系统，其中复现只发生有限次。

关于动力系统中复现的第二个问题，通常由庞加莱复现定理给出。

定理 2.3. – 庞加莱复现定理
设 S 为一个有界区间，U 为 S 的一个正长度基本子集，且 $f:\ S \to S$ 为一个保长度变换。那么，对于几乎所有 $x \in U$，存在 $n \in \mathbb{N}$ 使得 $f_{(n)}(x) \in U$。

与许多数学证明一样，庞加莱复现定理需要厘清几个关键词的含义。

定义 2.4. – 基本集合
基本集合 S 是一个可以定义为有限个区间之并的集合。

$$S = S_1 \cup S_2 \cup \ldots \cup S_n$$

由此可知，μ 的测度（长度）S 是有限个两两不相交且并集为 S的区间长度之和，

$$\mu(S) = \mu(S_1) + \mu(S_2) + \ldots + \mu(S_n), \quad S_i \cap S_j = \emptyset$$

定义 2.5. – 保长度函数
定义在定义域 f 上的函数 S，如果 $\forall J \subseteq S$，则称其关于测度 μ 保持长度，其中 J 是基本的，

$$\mu(f^{-1}(J)) = \mu(J)$$

并且 $f^{-1}(J)$ 也是基本的。

定义 2.6. – 几乎所有
对于空间 S 和测度 μ，$Z \subset S$ 是一个零测度集，如果 $\forall \epsilon > 0$，存在一族连通子集 $A_n \subset S$，使得

$$Z \subseteq \bigcup_{n=1}^{\infty} A_{n,}, \quad \sum_{n=1}^{\infty} \mu(A_n) < \epsilon.$$

如果某个陈述对 S 中几乎所有点都成立，那么存在某个 Z，使得该陈述对 $S \backslash Z$ 中的所有点都成立。

庞加莱复现定理指出，任何由保长度函数 f 所定义的有界动力系统都是回归的。此外，点会无限多次回归。这一结果为混沌动力系统提出了几个关键结论：

1. 不稳定周期轨道（即零测度集）是稠密的。2. 会发生拓扑混合（在不稳定周期轨道之间来回跳动）。3. 状态会无限多次返回到任意接近的位置

此外，对于混沌系统，短期动力学由 f 支配，并因对初始条件的敏感依赖和拓扑混合而发散；但长期行为取决于相空间的性质，而非初始条件（即最终趋于某个吸引子流形。）

第 3 章
嵌入理论与重构

回顾一下，任何一般动力系统都可以表示为三元组 $(T,\ M,\ \Phi)$，由时间排序 T、状态空间流形 M 和演化算子／动力学 Φ 组成。如果 $\mathbf{s}(t) \in M$ 是系统状态，那么自治动力系统可以表示为：

$$\dot{\mathbf{s}} = f(\mathbf{s}) \tag{3.1}$$

这与现实世界的问题有何关联？几乎总是无法观测到完整状态 $\mathbf{s}$。即使能够做到，受测量和计算限制，观测值在时间上也会离散化。我们可以用测量函数 h 来描述这一限制：$M \to \mathbb{R}$,

$$x(t) = h(\mathbf{s}(t)). \tag{3.2}$$

其中，$x(t)$ 是对动力学的低维、不完整表征。通常，采样时（使用 h），观测是在离散时刻进行的。因此，我们也可以将观测值写成时间序列 $\{x_t\}_{t=1}^{T}$。

在上述上下文中，我们可以利用观测时间序列做哪些有趣且有用的事情？一般来说，我们可以进行两类任务

1. **系统表征：** 底层系统的动力学和运行机制是什么？虽然这是一个定性问题，但可以通过李雅普诺夫指数、关联维数等描述性测度进行量化。2. **状态预测：** 给定一段状态历史记录 $\mathbf{s}(t_0)$，……，$\mathbf{s}(t)$，未来某个时间 τ 的预测状态 $\mathbf{s}(t+\tau)$ 是什么？

对于状态预测，我们通常不知道或无法知道 $\mathbf{s}$，因此只能退而求其次，回答一个替代性问题：给定一段观测历史 $x(t_0)$，……$x(t)$ 其中 $x(t) = h(\mathbf{s}(t))$，未来某个 τ 时刻的观测值 $x(t+\tau)$ 的最佳估计是什么？

这两个问题都面临若干有待解决的挑战

- 预测需要对动力学进行特征描述 f（或演化算子）。

- 如果已知 f，那么只有在我们掌握完整状态$\mathbf{s}(t)$的情况下，这才有用。
- 我们无法直接访问完整的状态空间 $\mathbf{s}(t)$

可以将 f视为对$\mathbf{s}$ 的各个分量如何相互作用并彼此反馈的描述。因此，可以合理地认为，经过足够长的时间后，一个分量中的信息必然会传播到其他分量。由此，单个时间步上的高维状态与跨越多个时间步的低维状态之间存在等价性。这为我们如何解决上述问题提供了线索。幸运的是，嵌入理论为我们应对这一有趣的挑战提供了一种方法。

3.1 嵌入定理

定义 3.1。- 嵌入
考虑一个状态$\mathbf{s}(t) \in \mathrm{S} \subseteq \mathbb{R}^m$ 的动力系统，其演化算子为 f_T，满足：

$$\mathbf{s}(t+T) = f_T(\mathbf{s}(t)),$$

以及相应的观测函数h：$\mathbb{R}^m \to \mathbb{R}$，它会生成一个时间序列，

$$x(t) = h(\mathbf{s}(t)).$$

嵌入是一种变换 Ψ：$\mathbb{R} \to X \subseteq \mathbb{R}^d$，满足存在一个微分同胚 Φ：$\mathrm{S} \to X$，其中，

$$f_T = \Phi^{-1} \circ F_T \circ \Phi,$$

使得，

$$x(t+T) = \Psi^{-1} \circ F_T \circ \Psi(x(t)).$$

我们从流形及其基本嵌入性质开始讨论。具体而言，我们将流形定义为 $M \subseteq \mathbb{R}^n$ 中的任意曲面，它在某个小邻域内局部呈欧几里得空间。（关于如何将邻域（坐标图）映射以构成一个图册，还有更严谨的描述；这些内容我们将留待其他更具权威性的文献讨论。）

首先，我们应该思考研究流形嵌入的动机是什么？请记住，在许多情况下，混沌系统（周期系统也一样）通常不会占据整个环境状态空间，而是会稳定在某个子空间流形上。考虑到我们的观测值也受到维数较低的限制，研究流形的嵌入定理及其维数可能会有所助益。

定理 3.1. – 惠特尼嵌入定理
任何将一个 d 维流形映射到 m 维流形（$d, m \in \mathbb{Z}^+$）的连续函数，在满足 $m \geq 2d+1$ 的条件下，都可以用光滑嵌入来近似。示例见图 *3.1*。

示例：
- 克莱因瓶：$d = 2 \implies \mathbb{R}^4$
- 带环的圆：$d = 1 \implies \mathbb{R}^3$

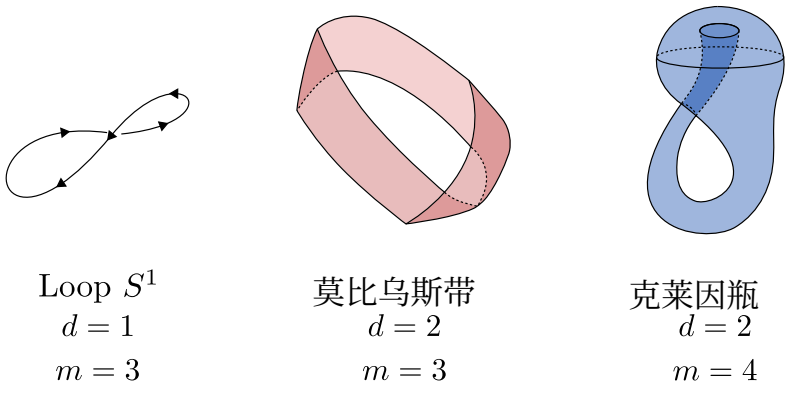


图 3.1 惠特尼嵌入定理的示例。请注意，$m = 2d + 1$ 是保证嵌入所需的最小值。较低维度的嵌入有时也是可行的。

惠特尼嵌入定理为所需的嵌入维数提供了一个有用的上界。然而，Sauer 等人在其开创性著作**"Embedology"**中进一步改进了这一界限，利用更一般化的量——**关联维数**——给出了一个更紧的上界。

推论 3.1. – 惠特尼嵌入定理（***Sauer*** 等人，***1991***）每个 d 维流形（$d \in \mathbb{R}^+$）都可以嵌入到 $\lceil 2d \rceil$ 中，其中 d 是关联维数。

回到状态空间估计的问题。假设我们有一条轨迹$\psi(t,\ \mathbf{s}_0) \subset M \subset \mathbb{R}$ 以及测量函数 $h(\psi(k\Delta t,\ \mathbf{s}_0)) = x_k$，其中 $n \in \mathbb{Z}$。对于$n = 1$，我们可以定义如下内容：

$$x_{k+1} = \int_0^{\Delta t} f(\mathbf{s})\, dt = \phi_{\Delta t}(\mathbf{s}). \tag{3.3}$$

对于 $n > 1$，我们需要估计多个导数（f）来计算 x_{k+1}。反过来，我们能否使用（x_k，x_{k+1}，x_{k+2}，……，$x_{k+(n-1)}$）个连续测量来表示这些导数？（即，我们能否将高维状态空间表示为连续观测组成的向量？）

定理 3.2. – *Takens* 嵌入定理 *[8]*
设 $\mathcal{S}$ 是一个维数为 d 的紧流形。对于满足 $\phi \in Diff^2(M)$ 的数对 (ϕ, h)，其中 $h \in C^2(\mathcal{S}, \mathbb{R})$，映射 $\Psi_{(\phi,h)} : \mathcal{S} \to \mathbb{R}^{2d+1}$ 是一个泛性质，其定义如下

$$\Psi_{(\phi,h)}(s) = (h(s), h(\phi(s)), ..., h(\phi^{2d}(s))),$$

是一个有效嵌入。

这里，ϕ 等价于演化算子。Takens 定理基本上指出，对于几乎所有观测函数 h 和可微演化动力学 f_T 的选取，构造向量 $\Psi_{(\phi,h)}(s)$ 会得到一个有效嵌入。这一结果主要依赖于流形之间 C^k 的映射 C^k的横截性和线性化思想。更直观地说，该定理保证，对于给定的动力系统，由时间滞后标量观测构成的延迟向量

$$\mathbf{x}(t) = (x(t), x(t-\tau), ..., x(t-2d\tau)), \tag{3.4}$$

是一个有效嵌入。因此，对于任何维数已知为m的动力系统，

$$\dot{\mathbf{s}} = f(\mathbf{s}), \tag{3.5}$$

可以利用标量观测值的延迟向量重构一个拓扑共轭动力系统，其中

$$\dot{\mathbf{x}} = F(\mathbf{x}), \quad \mathbf{x} \in X \subseteq \mathbf{R}^m, \tag{3.6}$$

只要 $m \geq 2d + 1$。因此，Takens 定理还强调了自治动力系统的演化与一般标量自回归过程演化之间的密切联系。此后，人们提出了众多嵌入方法，但这些方法无一例外都利用了有效嵌入与延迟观测之间的复杂联系。然而，由于延迟嵌入方法简单，且对噪声具有相对稳健性，它一直是时间序列分析应用中最常用的方法。请参阅 [9]，其中给出了 Takens 定理的证明概要。

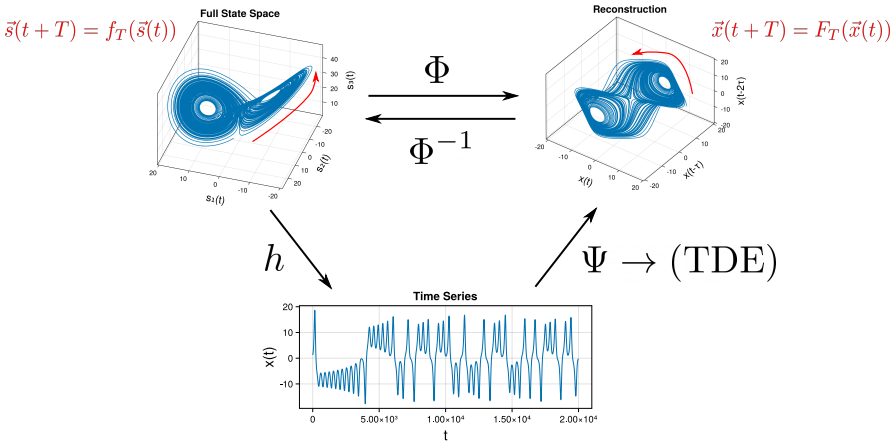


图 3.2 嵌入过程及其各分量之间关系的示意图。

3.2 嵌入方法

到目前为止，我们的讨论为利用嵌入方法重构相空间的有效性奠定了理论基础。然而，关于实践中如何实现这一点的细节仍然十分有限。本节将介绍几种可用于构造嵌入的计算方法。

3.2.1 时间延迟嵌入

时间延迟嵌入最早由 [10], 提出，它通过构造延迟向量 $\vec{x}$（t），将标量时间序列 x（t）扩展到更高维度，具体形式如下：

$$\vec{x}(t) = (x(t), x(t-\tau), ..., x(t-(m-1)\tau)), \tag{3.7}$$

其中需要选定的嵌入参数为延迟滞后 τ 和嵌入维数 m。根据 Takens 定理的保证，只要 m 足够大且测量值具有无限精度，τ 的任意取值都能得到有效嵌入。然而，实际中无法达到这一条件，而且延迟滞后和嵌入维数的不同选择会产生不同结果。

请注意，选择理想的延迟滞后 τ 和嵌入维数 m 并非时间延迟嵌入所独有。选择 τ 和 m 的取值，也是置换熵 [11] 和序数分区网络 [12] 等其他时间序列分析方法中的关键决策。在这两种情况下，都会构造一个延迟向量，并根据各分量的大小顺序进行编码表示。时间序列

随后可以被视为不同编码状态之间的转换，并用于进一步分析。

3.2.2 导数嵌入

导数嵌入方法利用观测时间序列中逐阶递增的时间导数来重构嵌入向量，

$$\vec{x}(t) = \left(x(t), \frac{dx(t)}{dt}, ..., \frac{d^m x(t)}{dt^m}\right) \tag{3.8}$$

导数通过数值近似求得。只要能够准确计算所需导数，当 m 足够大时，导数嵌入方法就是有效嵌入。

3.2.3 积分-微分嵌入

导数嵌入方法的一个缺点是需要根据数据计算数值导数。虽然这对于一阶导数而言或许可以接受，但随着信噪比往往受到负面影响，连续高阶导数的近似通常并不准确。即使拥有非常干净的数据集，情况也是如此。

导数嵌入的一种替代方法是积分-微分嵌入 [13]。这种方法通过用积分代替二阶导数，避免了计算连续的高阶导数。由此得到以下嵌入构造：

$$\vec{x}(t) = \left(\int_{-\infty}^{t} x(t) - \langle x(t)\rangle_t \, dt, x(t), \frac{dx(t)}{dt}\right), \tag{3.9}$$

其中，首先将第一个分量设为零均值，然后进行积分。一阶积分和数值微分的使用，仅使第一个和第三个嵌入分量的信噪比分别降低一个数量级。相比之下，导数嵌入方法中的每一次后续数值微分，都会使信噪比按递增的数量级下降。然而，对于更高维的嵌入，积分-微分嵌入方法会受到与纯导数方法相同的噪声影响。这限制了它在那些被认为具有高维系统动力学的系统中的适用性。

3.2.4 全局主值嵌入

Broomhead 和 King 提出了主值嵌入方法[14]，将其作为一种经过改进的时间延迟嵌入替代方法，并运用了 Takens 的定理。该方法借鉴主成分分析的思想，对具有足够高维数的时间延迟嵌入进行理想旋转（见图 3.3）。给定一个长度为 N_T 的时间序列 $x(t)$和一个长度为 M的滑动窗口，我们可以构造一组 $N = N_T - (M - 1)$个延迟向量，

$$\mathbf{X} = N^{-1/2} \begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vdots \\ \vec{x}_N \end{bmatrix}, \tag{3.10}$$

其中， $\vec{x}_i$ 是利用时间序列中以 i^{th} 值为第一分量构造的延迟向量，

$$\vec{x}_i = (x(t_i), x(t_{i-1}), ..., x(t_{i-(M-1)})). \tag{3.11}$$

可根据$\mathbf{X}$计算 $M \times M$ 协方差矩阵 C。该矩阵的元素 C_{ij} 可简要表示为：

$$C_{ij} = \langle x(t)\, x(t + (i - j)) \rangle_t \tag{3.12}$$

其中，$...$$_t$ 表示时间平均。随后，通过计算...的相应特征值和特征向量，求得其主成分。取与所需嵌入维数相对应的前 m 个主成分后，可利用特征向量矩阵计算$\mathbf{X}$ 的投影，从而得到最终的嵌入坐标。感兴趣的读者可参阅 [14] 和 [15]。

主成分值嵌入旨在将高维延迟嵌入（通常通过取大量时滞分量获得）提炼并简化到一个低维子空间。其余子空间被认为对应于动力学变化和重要性都较小的分量方向。该方法的一项应用是尝试简化最优嵌入维数的选择：理想的嵌入维数m 对应于那些显著高于某个“噪声底”的奇异值的数量。

3.3 均匀与非均匀延迟嵌入

延迟嵌入有两种形式。第一种是均匀延迟嵌入，选择单个时滞 τ 和嵌入维数 m，以构造由观测值组成的延迟向量，这些观测值在时间上等间隔分布，

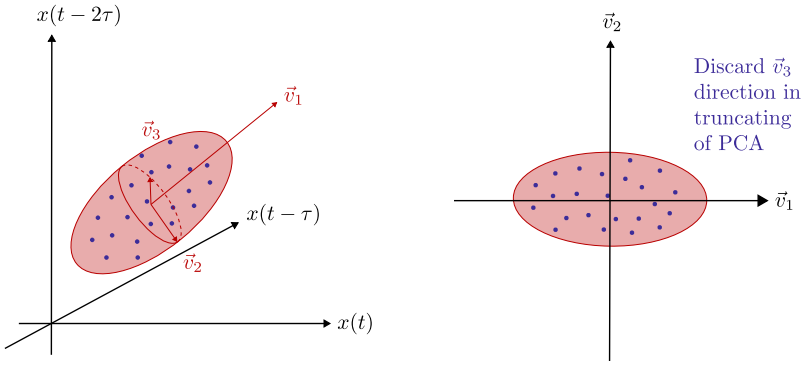


图 3.3 通过降维将 3 维降至 2 维的主成分分析嵌入（即 2 个主成分 v_1 和 v_2）

$$\mathbf{x}(t) = (x(t), x(t-\tau), ..., x(t-(m-1)\tau))). \tag{3.13}$$

由于只需选择两个参数，这种方法是常用的经典延迟嵌入形式。

尽管均匀方法对于呈现单一主导周期性的系统效果良好，且易于实现，但这种便利也付出了代价：灵活性降低，并且在分析具有更复杂多尺度动力学的系统时存在局限 [16]。

首先，选择使用单一时滞会限制重构凸显多个彼此差异显著的时间尺度上的特征的能力 [17]。例如，对于特征时间尺度为 τ_1 和 τ_2 且 $\tau_1/\tau_2 \gg 1$ 的快慢系统，选择 τ_1（即慢动力学）作为嵌入滞后，可能会限制重构充分展开对应于快动力学的吸引子拓扑的能力。时间尺度为 τ_2 的动力学（即快动力学）将在重构状态空间中表现为噪声波动。

其次，充分的均匀延迟嵌入所得到的重构未必是最优的。在此，我们必须说明，最优的定义预设了某种评价标准或质量概念。 [15] 指出，嵌入的质量（定义为重构在预测含噪数据时的稳健性）会因吸引子的不同区域而在局部发生变化。 [18] 在扩展 Casdagli 的噪声放大与失真方法时，也强调了这一现象。此外，我们还应考虑到，李雅普诺夫指数等不变测度也会在局部发生变化。因此，选择单一的嵌入滞后意味着这些变化都可能被平均化。

另一种提出的延迟嵌入形式是非均匀延迟嵌入。

$$\mathbf{x}(t) = (x(t), x(t-\tau_1), ..., x(t-\tau_{m-1})) \tag{3.14}$$

这是对均匀嵌入的自然扩展，旨在通过将多个时间尺度编码到延迟向量中，解决后者的一些局限性。

第4章
嵌入实践

在进行时间序列分析时，嵌入方法似乎是应对实际操作挑战和局限性的一种可行方式。假设我们采用某种形式的延迟嵌入（因为它简单易用），仍然需要回答两个问题：

问题1。均匀还是非均匀？这取决于具体应用，也仍有争议。详见[19]，其中有详细讨论。**问题2**。嵌入参数（维数和时滞）应取什么值？

4.1 嵌入参数是否重要？

回想一下，Takens 定理并未规定延迟嵌入要有效，Δt（或 τ）必须取什么值。它只要求维数足够高。然而，正如我们将看到的，这并不意味着所有嵌入都一样，也不意味着它们在时间序列分析中的表现同样出色。此外，这还没有考虑到我们尚未讨论“足够大的维数”是什么意思，以及怎样才算“足够大”。

在 Takens 定理的原始表述中，所引述的演化算子ϕ：$x(t) \to x(t+\tau)$具有一般性，因此应适用于任意的 τ选择。然而，这仅在采样分辨率和精度无限高且不存在噪声的情况下成立。实际上，这些条件即便并非不可能实现，也极少能够满足。

为了说明噪声和精度的影响，考虑时间序列 $s(t) = \cos(t)$，它也可以表示为线性常微分方程的一个解：

$$\ddot{s} + s = 0, \tag{4.1}$$

其中，底层吸引子由一个环表示。该系统也可以写成符合以下形式的一阶二维动力系统：

方程

$$\dot{x} = y \tag{4.2a}$$
$$\dot{y} = x \tag{4.2b}$$

由$\mathbf{x}(t) = (x(t), x(t-\tau))$构成的延迟向量，在所有 $\tau > 0$ 取值下都会形成一个椭圆（见图 4.1）。增大 τ 的值会降低椭圆的离心率；当 τ 等于周期的四分之一时，离心率最小。因此，位置$\mathbf{x}(t)$与振荡器的相位 ϕ之间存在明确的一一映射。根据观测值 $x(t)$，可以准确确定底层系统的状态。

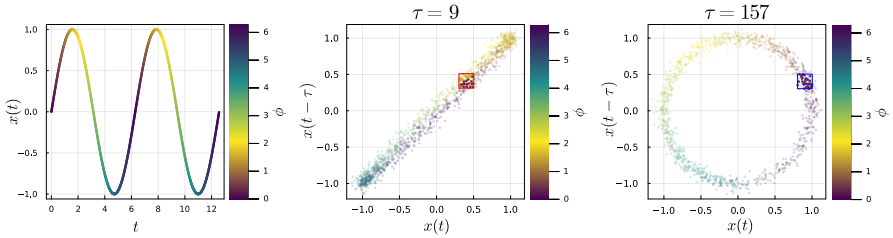


图 4.1 正弦时间序列（左）。对受高斯噪声污染的时间序列，采用不同时间滞后构造的延迟嵌入 $\xi \sim \mathrm{N}(0, 0.05^2)$。时滞选择不佳、过小时，可能导致相位识别错误（中）。取四分之一周期的时滞 $\tau = 157$ 可更好地分离轨迹（右）。边界框对应于潜在的精度误差 $\epsilon = 0.05$

观测噪声的存在会给真实状态的估计带来不确定性。在有限精度 ϵ的情况下，不确定性会在每个观测点周围形成一个宽度为 ϵ的正方形区域。当对应不同相位的多个观测值所产生的不确定区域相互重叠，导致真实状态难以判定时，这种影响最为显著 $s(t)$。对于周期时间序列，当 τ 等于周期的四分之一时，这种影响最小，此时椭圆的展开程度最大。尽管噪声和精度对周期信号的影响似乎相对较小，但对于包含稠密轨道的混沌系统，不确定性的影响不容忽视，而且可能使真实底层状态难以估计。此外，对于动力学跨越多个空间和时间尺度的系统，并不能保证存在理想的 τ。

设 Ψ_θ 是关于嵌入参数 θ 选择的嵌入映射。应选择什么样的参数 θ 才能得到最佳嵌入？

对于均匀延迟嵌入，θ 由嵌入滞后 τ和嵌入维数 m组成。更一般地，非均匀延迟嵌入则需要

选择一组时滞 $\mathrm{T} = \{\tau_1, \ldots\ldots, \tau_{m-1}\}$ 来构造延迟向量，

$$\mathbf{x}(t) = (x(t), x(t-\tau_1), ..., x(t-\tau_{m-1})). \tag{4.3}$$

然而，由于嵌入质量这一概念尚未定义，嵌入问题本身是不适定的。因此，要回答嵌入问题，同样需要先解决另一个问题：“如何量化嵌入的优劣？”

4.2 测量嵌入质量

嵌入的主要用途在于保留原始（部分观测到的）系统的动力学。这也意味着真实状态$\mathbf{s} \in M$与重构状态$\mathbf{x} \in \mathbb{R}^m$之间存在唯一的一一对应。如前所示，当数据中存在噪声且分辨率有限时，这种对应关系无法得到保证；因此，某些嵌入比其他嵌入更能保留这种一一对应关系。所以，我们希望嵌入（参数为 θ 的延迟嵌入）尽可能减少对真实状态的模糊性。但从拓扑学角度看，这意味着什么？

1. **尽量减少自交次数**
   根据存在性与唯一性定理以及流的基本定理，轨迹不会相交。因此，我们的轨迹和流形也不应与自身相交。<0x0A>这表明嵌入维数可能不够大。

   从状态估计的角度来看，自交会在进行状态推断时造成歧义。（当你试图将两个不同频率的正弦波之和嵌入二维空间时，会发生什么？）

2. **尽量减少层状流形**
   与自交类似，但可以说问题没那么严重，流形中的层流结构虽然并未直接违反存在性与唯一性，却会因与自交类似的原因导致流形病态。除非观测值和训练好的预测模型完全准确，否则层状曲面上的区域便是状态估计的不确定区域。

选择延迟坐标（即 τ 和 m）时，可以从两个概念来概括这些考量：（1）无关性和（2）冗余。我们不必分别考虑 τ 和 m，而可以考虑嵌入窗口 $\tau_w = m\tau$ 的大小（用于均匀嵌入）。理想情况下，嵌入窗口 τ_w不应太短（以免产生冗余），因为分量具有较高的时间相关性，这会沿主对角线形成层流性。同样，τ_w 也不应太长（以免产生无关性），以至于坐标完全不再相关，嵌入流形变得过于折叠，并出现大量自交。

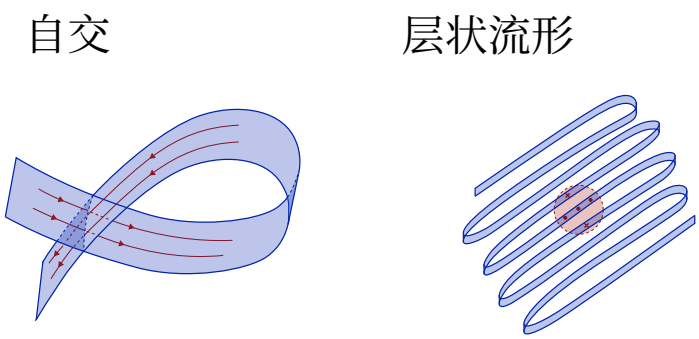


图 4.2 自交与层状流形。层状流形上的开球揭示了真正的最近邻和伪最近邻。

通常，所有嵌入参数选择方法都旨在根据某种质量概念进行优化，而这种质量概念与冗余和无关性的思想有关。

4.3 均匀延迟嵌入——同时选择 τ 和 m

4.3.1 Gao & Zheng - 特征长度

Gao 和 Zheng 提出的嵌入方法 [20, 21]基于伪最近邻的出现情况。伪最近邻可能源于冗余（展开不足）或无关性（吸引子中的虚假交叉）。Gao 和 Zheng 提出的方法基于这样一种观点：在理想的嵌入中，应尽可能减小分离距离和伪最近邻所占的比例。

考虑嵌入空间中的一对点 $\vec{x}_i$， $\vec{x}_j$ 以及它们向未来演化 k 步后的情况 $\vec{x}_{i+k}$， $\vec{x}_{j+k}$。随着吸引子在时间延迟嵌入中展开，伪最近邻倾向于比真正的邻居分离得更快。因此，对于伪最近邻点对，其距离之比 $|\vec{x}_{i+k} - \vec{x}_{j+k}|/|\vec{x}_i - \vec{x}_j|$ 会更大；对于真正的邻居，该比值则约等于 1。随后，Gao 和 Zheng 提出以下测度 Λ来优化嵌入参数，

$$\Lambda(k, m, \tau) = \frac{1}{N_{ref}} \sum_{i,j} \ln \frac{|\vec{x}_{i+k} - \vec{x}_{j+k}|}{|\vec{x}_i - \vec{x}_j|}, \tag{4.4}$$

其中， N_{ref} 是随机抽取的点对数量，在这些点对上对距离比值取平均。对于点对的选择，还有几项额外限制 $\vec{x}_i$， $\vec{x}_j$。首先，这些点的初始间距应满足 $|\vec{x}_i - \vec{x}_j| \leq r$

其中 r 是一个选定的较小阈值，即点的初始间距应足够小，从而使增长间距的计算具有意义。其次，所选点对不应具有相交的 Theiler 窗口 $|i-j| > l_{\text{Theiler}}$，其中 $l_{\text{Theiler}} \in \mathbb{N}^{+}$。这样做是为了避免同一局部轨迹上的点之间产生不需要的相关性 [22]。最后，常数 k 不应过大，并且应根据系统动力学的自然时间尺度进行选择。

为确定合适的嵌入参数，需要针对不断增大的嵌入维数 m计算 $\Lambda(\tau)$的剖面。选择剖面 $\Lambda(\tau)$中降幅最大的 m 所对应的值作为嵌入维数。然后，将嵌入滞后 τ 选为 $\Lambda(\tau)$的第一个极小值。

这种方法的问题在于，它需要选择一个演化时间k。与其任意选择 k，不如计算特征长度 $J(m, \tau)$，用以描述系统吸引子的自然空间尺度，

$$J(m,\tau) = \langle |\vec{x}_i - \vec{x}_j| \rangle, \tag{4.5}$$

其中 ... 表示对吸引子上采样点对取平均。随后，利用特征长度计算分离时间 $T_J(\vec{x}_i, \vec{x}_j)$，其定义为最近邻点对发散至特征长度 $J(m, \tau)$某一比例所需的时间。对于真正的邻居，随着嵌入维数 m增加，T_J将收敛至与系统李雅普诺夫指数相关的值；而伪最近邻会得到较小的值 T_J，因为轨迹会迅速分离。用于确定嵌入参数的新测度定义如下，

$$C(m,\tau) = \frac{1}{N_{ref}} \sum_{i,j} T_J(\vec{x}_i, \vec{x}_j), \tag{4.6}$$

其中 N_{ref}是采样的近邻对数量。选取使 $C(m,\tau)$ 最大化的 m和 τ 值，分别作为嵌入维数和时滞。

4.3.2 摆动乘积

摆动乘积 [23] 与 Gao 和 Zheng 的方法以及特征长度相似，因为它们都基于最近邻的概念。作者提出，良好的嵌入应保留给定参考点附近邻居的顺序对应关系（即，应保留按照与某个参考点 $\vec{x}_i$ 的距离排序后的邻居顺序）。具体做法是比较点 $\vec{x}_i$ 的 N 个邻近邻居在给定嵌入 Φ_k中的顺序，其中有序序列邻居表示为，

$$X_{\Phi_k} = \{\vec{x}_{i,1}, ..., \vec{x}_{i,N}\}_{\Phi_k}, \tag{4.7}$$

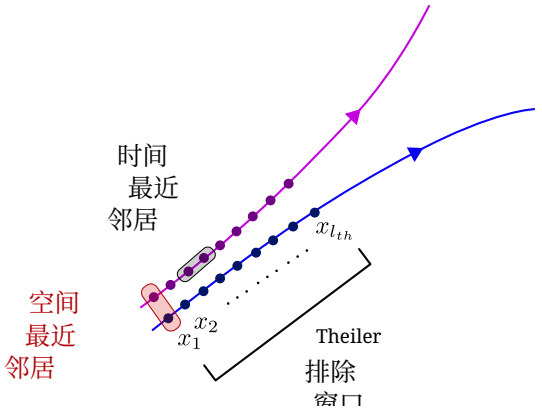


图 4.3 用于排除近邻的 Theiler 窗口：这些近邻之所以接近，是因为它们在时间上相距很近。这样做是为了避免统计量产生偏差。

以及其在下一个阶数嵌入 Φ_{k+1} 上的投影（通过增加 m 或 τ）所对应的序列，表示为，

$$Z_{\Phi_{k+1}} = \{\vec{z}_{i,1}, ..., \vec{z}_{i,N}\}_{\Phi_{k+1}}. \tag{4.8}$$

这里，$\vec{x}_{i,\ n}$ 对应于 i^{th} 参考点 $\vec{x}_i$的 n^{th} 最近邻。投影 $\vec{z}_{i,\ n}$ 对应于同一个邻居数据点 $\vec{x}_{i,\ n}$，其位置根据下一阶嵌入 Φ_{k+1}重新计算。

类似的比较也可以投影到阶数更低的嵌入中（通过减小 m 或 τ），从而得到一组新的有序点，

$$V_{\Phi_{k-1}} = \{\vec{v}_{i,1}, ..., \vec{v}_{i,N}\}_{\Phi_{k-1}}. \tag{4.9}$$

理想情况下，良好的嵌入应在这些有序序列中保持一一对应关系。这将使以下比值等于 1，

$$\frac{|\vec{x}_i - \vec{z}_{i,n}|}{|\vec{x}_i - \vec{x}_{i,n}|}, \text{ and } \frac{|\vec{x}_i - \vec{x}_{i,n}|}{|\vec{x}_i - \vec{v}_{i,n}|}. \tag{4.10}$$

Liebert 等人提出的方法将以下测度定义为上述两个比值的乘积，

$$W_i(m,\tau) = \prod_{n=1}^{N} \left\{ \left(\frac{|\vec{x}_i - \vec{z}_{i,n}|}{|\vec{x}_i - \vec{x}_{i,n}|} \right) \cdot \left(\frac{|\vec{x}_i - \vec{x}_{i,n}|}{|\vec{x}_i - \vec{v}_{i,n}|} \right) \right\}. \tag{4.11}$$

待优化的测度由对随机抽样的 N_{ref} 个参考点取平均得到，

$$W(m,\tau) = \ln\left(\frac{1}{N_{ref}}\sum_{i=1}^{N_{ref}} W_i(m,\tau)\right) \tag{4.12}$$

其中选择 m 作为实现W极限行为的维数，而 τ 对应于所得剖面的第一个极小值。

4.4 均匀延迟嵌入 - 分离选择

在实践中，执行均匀延迟嵌入时，很少使用同时参数选择方法。相反，通常的做法是分别为 m 和 τ 选择合适的值。在一种参数的选择依赖于先前对另一种参数的选择的方法中，可以采用迭代方法（例如选择 $\tau_1 \to m_1|\tau_1 \to \tau_2|m_1...$）。

4.4.1 选择 τ

有多种方法可用于指导嵌入时滞 τ 的选择，其复杂程度从启发式方法到更具理论性的方法不等。下面简要介绍其中几种方法。

四分之一周期
这种方法可以将系统动力学的自然时间尺度编码到嵌入过程中。这种启发式方法源于将正弦波嵌入二维 $x(t) = \sin(\omega t)$ 的问题。在这种情况下，$\tau = \frac{2\pi}{4\omega}$ 会产生一个最接近圆形的二维延迟嵌入。然而，这种启发式方法无法直接应用于信号呈非周期性的混沌系统。相反，需要估算某种形式的伪周期，例如连续极大值之间的平均时间，

自相关
这种方法旨在选择滞后，使延迟向量中的分量之间的线性相关性最小，从而尽量减少嵌入中的冗余，

$$R(\tau) = \frac{\mathbb{E}[(x(t)-\bar{x})(x(t+\tau)-\bar{x})]}{\sigma^2}. \tag{4.13}$$

延迟滞后 τ 可以选择为 $R(\tau) < \frac{1}{e}$ 的第一个零点、第一个极小值或第一次交叉。各项判据的选择具有任意性，通常根据哪一种在目标应用中表现最佳来决定。

最小互信息
自相关的一个缺点是无法考虑时间序列中的非平稳性

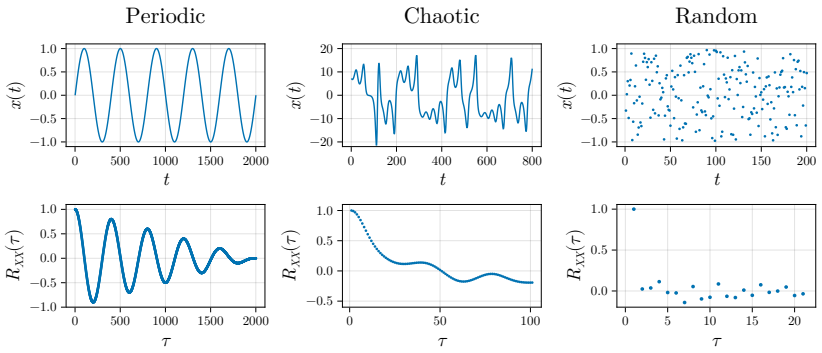


图 4.4 周期、混沌和随机信号的自相关曲线

（例如相位、频率和幅值的漂移）。此外，它的应用只测量线性相关（Abarbanel，1993年），这对于非线性动力系统而言可能会造成问题。除了最简单的情况外，动力系统都表现出一定程度的非平稳和非线性行为。[24]提出，应使用系统的自互信息取代自相关。在他们的原始论文中，Fraser和Swinney首先为互信息的理论论证提供了几何解释。具体来说，考虑一组点，其某个分量的取值 x 落在一个固定窗口内。从这组点出发，追踪它们在未来 τ 步后的位置，并计算同一组点在同一分量上的取值 $p_\tau(x)$ 的分布。若某个 τ 对应的分布 $p_\tau(x)$ 更宽，则它应当是一个合适的时滞，同时也对应于较小的互信息值。

互信息可解释为利用信息论中的思想计算得到的自相关函数的非线性类比。给定两个随机变量 X，Y 它们之间的互信息为：

$$I(X,Y) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{X,Y}(x,y)\log_2\left(\frac{P_{X,Y}(x,y)}{P_{X(x)}P_Y(y)}\right) \tag{4.14}$$

对于连续随机变量，则为：

$$I(X,Y) = \int_{\mathcal{X}}\int_{\mathcal{Y}} P_{X,Y}(x,y)\log_2\left(\frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)}\right)dydx. \tag{4.15}$$

简单来说，量 $I(X,Y)$ 以比特（以 2 为底时）或纳特（以 e 为底时）表示，衡量的是 X 的测量值中所包含、同时也包含于 Y中的“信息”量。或者说，观测 X 能让你对尚未观测到的 Y了解多少？将这一类比应用于时间序列，我们可以对 τ 提出类似的问题：一次观测 $x(t)$ 与

在 $x(t+\tau)$? 如果目标是尽量减少冗余，我们需要选择 τ，使互信息 $I(\tau)$ 最小，其中，

$$I(\tau) = \int P(x(t), x(t+\tau)) \log_2 \left(\frac{P(x(t), x(t+\tau))}{P(x(t))P(x(t)+\tau)} \right) dt \tag{4.16}$$

$$I(\tau) = \sum_{n=1}^{N} P(x(n), x(n+\tau)) \log_2 \left(\frac{P(x(n), x(n+\tau))}{P(x(n))P(x(n+\tau))} \right). \tag{4.17}$$

其中，$P(x(t))$ 是在任意给定时间观测到状态 $x(t)$ 的概率，而 $P(x(t), x(t+\tau))$ 是以类似方式定义的时间 t 和未来时间 $t+\tau$ 的联合概率。根据信息论，互信息 $I(\tau)$旨在量化在时间 t的观测中包含的有关时间 $t+\tau$ 未来状态的信息量。给定时滞 τ 下较高的互信息意味着状态之间具有较强的相关性，并会导致延迟重构中的冗余度更高。

最小互信息和自相关的优势在于，它们能够以相对简单、快速的计算为时滞提供合理估计。然而，无法保证给定的互信息剖面 $I(\tau)$ [25]一定存在明确的最小值。此外，计算互信息需要对概率密度函数 $P(x(t))$ 和 $P(x(t), x(t+\tau))$ 进行数值估计，因此需要考虑最优直方图箱宽和数据长度要求 [26, 27]。

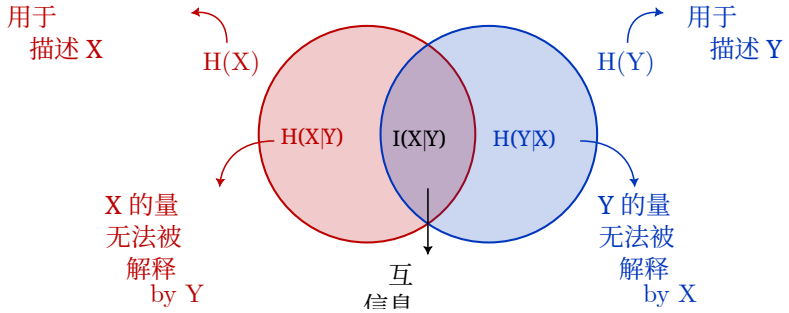


图 4.5 互信息与香农熵关系示意图

填充因子

由 [28] 首次提出的填充因子方法是一种完全基于几何的方式，用于计算给定嵌入的质量。该方法假设，理想的嵌入应能够展开吸引子，并最大限度地增大轨迹之间的间隔。作者认为，这种嵌入能够最优地利用环境空间，并减少系统真实状态在重构状态空间中不同

点上的歧义。

填充因子的计算方法是：先从数据的 m维延迟嵌入中随机采样 $m+1$ 个点。然后从这组点中选取一个参考点 $\vec{x}_r$，并计算相应的相对距离向量，

$$\vec{d}_i(\tau) = \begin{bmatrix} x_i(t) - x_r(t) \\ x_i(t-\tau) - x_r(t-\tau) \\ \vdots \\ x_i(t-(m-1)\tau) - x_r(t-(m-1)\tau) \end{bmatrix}. \tag{4.18}$$

相应的 $m \times m$ 矩阵可表示为

$$M(\tau) = (\vec{d}_1, \vec{d}_2, ..., \vec{d}_m), \tag{4.19}$$

所得平行六面体的体积可通过计算 M的行列式得到，

$$V(\tau) = \det(M(\tau)). \tag{4.20}$$

填充因子的最终表达式为：计算随机采样的一组平行六面体 V_i（τ）的平均体积，再除以采样数据点的范围，

$$f = \log\left(\frac{\frac{1}{N}\sum_{i=1}^{N} V_i(\tau)}{(\max_k x(t_k) - \min_k x(t_k))^m}\right). \tag{4.21}$$

作者建议选择能使填充因子 f 在区间 $\tau \in (0,\ T_c/2)$上达到最大值的 τ，其中 T_c 为特征复现时间。T_c 的值为，

$$T_c = \frac{1}{\omega_c}, \tag{4.22}$$

其中，ω_c 是时间序列功率谱中占主导地位的频率。

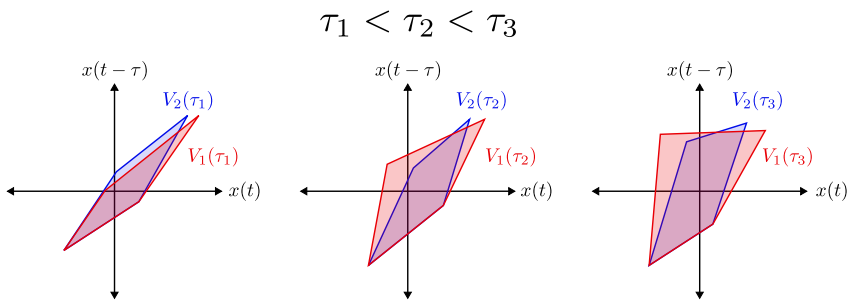


图 4.6 评估 V 的填充因子法

噪声放大

噪声放大是 Casdagli [15] 提出的一种测度，旨在量化嵌入的质量。这一概念基于这样的观点：良好的嵌入应有助于进行预测。此外，即使存在噪声，良好的嵌入也应能保持相对出色的表现。给定嵌入 $\vec{x}(t)$ 的噪声放大，是根据存在噪声时系统未来 T 步的可预测性来定义的。通常，其定义如下：

$$\sigma(T,\vec{x}) = \lim_{\epsilon\to 0} \sigma_\epsilon(T,\vec{x}), \tag{4.23}$$

其中

$$\sigma_\epsilon(T,\vec{x}) = \frac{1}{\epsilon}\sqrt{\mathrm{Var}[x(T)|B_\epsilon(\vec{x})]}. \tag{4.24}$$

这里，$\mathrm{Var}[x(T)|B_\epsilon(\vec{x})]$ 对应于在嵌入空间 $\mathbb{R}^m$中、从受附加小幅观测噪声 ϵ污染的初始条件 $\vec{x}$ 出发，对未来 T 步进行预测的条件方差。此处假设预测不存在模型误差。通过在嵌入 $\mathbb{R}^m$ 中选择相邻邻居，作为含噪初始条件 [18] 的替代，可以满足这一条件。

最后，对沿时间序列抽取的一组参考点计算噪声放大量 $\sigma(T,\ \vec{x})$ 的平均值，以得到噪声放大值 σ。高噪声放大的嵌入意味着嵌入空间 $\mathbb{R}^m$ 中的相邻邻居，其未来轨迹往往会迅速分离，因为它们并非真实流形 M 状态空间中的真实邻居。因此，噪声的影响会被大幅放大：重构空间 $\mathbb{R}^m$ 中的微小扰动会导致系统真实状态出现较大的不确定性。

L统计量

噪声放大测度的一个缺点是，它要求定义 T，即计算噪声放大时所采用的预测时域。Uzal等人通过将噪声放大的定义修改如下，解决了这一问题：

$$\sigma_\epsilon^2(\vec{x}) = \frac{1}{T_M}\int_0^{T_M} \sigma_\epsilon^2(T,\vec{x})\, dT. \tag{4.25}$$

该定义针对一系列预测时域计算噪声放大，预测时域的最大值为 T_M；研究发现，当 T_M足够大时，该定义相对稳健。

用于计算 σ 的算法依赖于将来自参考点 $\vec{x}_i$ 的 k 最近邻作为代理。根据点的分布，这可能导致每个点的有效噪声水平 ϵ 大小不同。因此，Uzal 提出了一个归一化常数 α_k，以适应这种变化，其定义如下：

$$\alpha_k^2 = \left[\sum_i \epsilon_k^{-2}(\vec{x}_i)\right]^{-1} \tag{4.26}$$

结合这两个思路，作者提出，噪声放大 σ 衡量了某种冗余性，而 α_k 衡量了某种无关性。随后，L统计量被描述为一个代价函数，用于同时最小化这两个值，

$$L = \log(\sigma\alpha_k) = \log(\sigma) + \log(\alpha_k). \tag{4.27}$$

4.4.2 选择 m ——全局伪近邻（*GFNN*）

Takens 定理的保证要求 $m > 2d+1$，其中 d 通常是未知的。可以采用 Sauer、York和Casdagli的改进 [29]，使得$m > \lceil c_d \rceil$ 即可满足要求，其中 c_d。但同样，估计 c_d 并非易事，而且还要求数据采用状态空间的形式（即需要预先进行嵌入）。一种常见方法是全局伪近邻（GFNN），其思想类似于 Gao & Zheng/特征长度/摆动乘积方法。

1. 选择时滞 τ 和起始嵌入维数 m （例如 $m = 2$），以生成重构时间序列

$$y(t) = [x(t), x(t+\tau), ..., x(t+(m-1)\tau)]$$

2. 对于重构吸引子中的每个点 y（t^*），找到其空间最近邻，同时使用 Theiler 窗口排除时间邻居，

$$y^{NN}(t^*) = [x(t^*_{NN}), x(t^*_{NN}+\tau), ..., x(t^*_{NN}+(m-1)\tau)]$$

为了根据长度为l_{Th}的 Theiler 窗口排除邻居，我们确保$|t^* - t^*_{NN}| > l_{Th}$。这样可以确保 $y^{NN}(t)$ 是 $y(t)$ 的近邻，这是由动力学和流形结构造成的，而非时间相关性。更详细的讨论请见第 5 章。

3. 使用维数为 $m+1$ 的数据进行另一次重构，并检验在维数增加后，$y(t^*)$ 和 $y^{NN}(t^*)$ 是否仍然是近邻（保持相同的时间指标）。如果维数增加 1 后，点对不再是最近邻，这意味着它们原本彼此接近是因为吸引子尚未充分展开。
4. 计算 $\mathbb{R}^{m+1}$ 中最近邻之间的距离

$$
\begin{aligned}
R_{m+1}(t^*)^2 &= \sum_{d=1}^{m+1}[x(t^* + (d-1)\tau) - x(t^*_{NN} + (d-1)\tau)]^2 \\
&= \left(\sum_{d=1}^{m}[x(t^* + (d-1)\tau) - x(t^*_{NN} + (d-1)\tau)]^2\right) \\
&+ [x(t^* + m\tau) - x(t^*_{NN} + m\tau)]^2 \\
&= R_m(t^*)^2 + [x(t^* + m\tau) - x(t^*_{NN} + m\tau)]
\end{aligned}
$$

5. 在考虑维数标度的情况下，计算最近邻之间的距离比 $R(t^*)$

$$
\frac{|x(t^* + m\tau) - x(t^*_{NN} + m\tau)|}{R_m(t^*)} = \sqrt{\frac{R_{m+1}(t^*)^2 - R_m(t^*)^2}{R_m(t^*)^2}}
$$

6. 如果 $R(t^*) > 15$ （或某个选定的阈值），那么时间指标 t^* 处的最近邻对就是伪邻居。此外，在维数 $m+1$ 中的距离远大于吸引子平均规模 的最近邻，也会被归类为伪邻居：

$$
\frac{|x(t^* + m\tau) - x(t^*_{NN} + m\tau)|}{R_A} > 2
$$

其中， R_A 是围绕数据均值居中后计算的均方根。

7. 重复上述步骤，并计算每次增加 m时，属于伪最近邻的点所占的百分比。当 $\%FNN$接近 0 时，选择嵌入维数 m 。

4.5 非均匀延迟嵌入

均匀嵌入是嵌入时间序列最常用的方法。这主要是因为它易于实现和优化。在直接应用中，均匀延迟只需选取两个参数： τ 和 m。然而，这种方法虽然方便，却以降低灵活性并带来局限性为代价，尤其是在分析动力学发生于多个相异时间尺度上的系统时。

首先，选择使用单一时滞会限制重构呈现跨越多个迥异时间尺度的特征的能力。例如，对于一个特征时间尺度为 τ_1 和 τ_2 、且 $\tau_1/\tau_2 \gg 1$ 的快慢系统，选择 τ_1 （即慢动力学）作为嵌入滞后，可能会限制重构充分展开与快动力学相对应的吸引子拓扑的能力。时间尺度为 τ_2 的动力学（即快动力学）会表现为噪声波动

，出现在重构状态空间中。

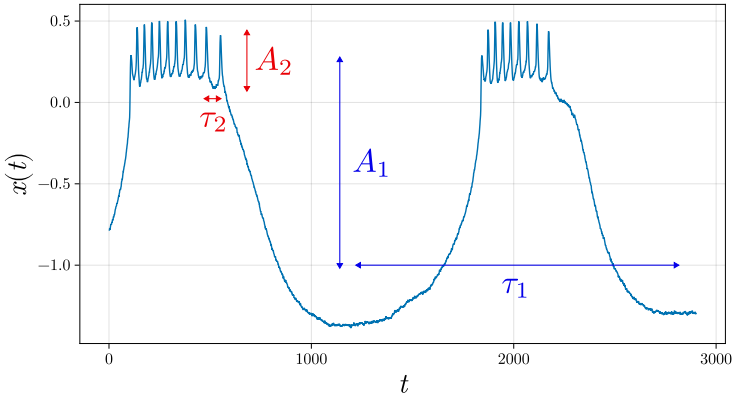


图 4.7 具有两种不同时间尺度和幅值的时间序列示例。数据取自实验测量的龙虾侧幽门（LP）神经元，承蒙 Abarbanel 等人提供 [30]。

其次，由充分的均匀延迟嵌入得到的重构不一定是最优的。这里必须明确，最优的定义预设了某种标准或质量概念。Casdagli 指出，嵌入的质量，即重构对用于预测的含噪数据的稳健性，可能在吸引子的不同区域局部变化。此外，我们还应考虑到，李雅普诺夫指数等不变测度也会发生局部变化。因此，选择单一嵌入滞后意味着所有这些变化都可能被平均化。

解决这些问题的一种显而易见的方法是在构造延迟向量时纳入多个非均匀时间滞后：

$$\vec{x}(t) = (x(t), x(t-\tau_1), ..., x(t-\tau_{m-1})). \tag{4.28}$$

延迟滞后的选择代表一个组合难题，其复杂度会随着嵌入维数的增加而增长。用于构造非均匀延迟嵌入的方法通常涉及迭代选择时间滞后，逐步构造延迟向量，直到达到所需的嵌入维数。

4.5.1 Garcia 和 *Almeida*

最早用于选择非均匀时滞的方法之一由 [31]提出。他们提出了一种全局伪近邻方法的变体，并将其应用于

时间延迟选择问题。该方法还利用一种提出的 N-统计量，在多个嵌入周期中递归地选择滞后。每个周期结束时，都会使用假最近邻算法评估新构建的嵌入质量。重复这一过程，直到假最近邻统计量 F 降至临界阈值以下。

一般算法如下：

1. 选择第一个时间滞后 τ_1 时，首先针对一个待测试的候选时间滞后 τ^*，进行二维延迟嵌入 $\vec{x}(t)$,

$$\vec{x}(t) = (x(t), x(t - \tau^*)).$$

2. 在重构嵌入中，考虑 Theiler 窗口，为每个点 $\vec{x}(t_i)$ 找出最近邻 $\vec{x}(t_j)$。

3. 计算任意两点之间的欧几里得距离 $d_{1,\ \tau^*}(t_i)$，$d_{2,\ \tau^*}(t_i)$,

$$\begin{aligned} d_{1,\tau^*}(t_i) &= ||\vec{x}(t_i) - \vec{x}(t_j)|| \\ d_{2,\tau^*}(t_i) &= ||\vec{x}(t_i + \delta t) - \vec{x}(t_j + \delta t)||, \end{aligned} \tag{4.29}$$

其中 δt 是数据的采样时间。简单来说，$d_{1,\ \tau^*}$ 是重构状态空间中最近邻点对之间的空间间隔，而 $d_{2,\ \tau^*}$ 是向前一个时间步后得到的间隔。

4. 计算 F统计量，即距离比$d_{2,\ \tau^*}/d_{2,\ \tau^*} > 10$的点所占的比例,

$$F(\tau^*) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\left(\frac{d_{2,\tau^*}(t_i)}{d_{1,\tau^*}(t_i)} > 10 \right), \tag{4.30}$$

其中，N 是时间序列的长度，$\mathbb{1}$ 是示性函数。作者根据 Kennel 等人[32]的数值计算，启发式地选取阈值 10。$N(\tau^*)$中第一个极小值对应的时间滞后被视为嵌入滞后。

5. 将维数增加一，并纳入新选出的时间滞后，重复步骤 1-4。因此，在非均匀嵌入过程中选择 m^{th} 嵌入滞后时，需要使用由已选定的 $m-1$ 个滞后和新的候选滞后 τ^* 构成的嵌入来计算邻居和距离$d_{2,\ \tau^*}$，$d_{2,\ \tau^*}$

$$\vec{x}(t) = (x(t), x(t - \tau_1), ..., x(t - \tau_{m-1}), x(t - \tau^*)).$$

4.5.2 连续性统计量

连续性统计量最初由 [17]提出，作为一种基于向量坐标之间函数独立性这一理念、以程序化方式构造非均匀延迟向量的方法。Takens 和 Sauer 都讨论了嵌入重构所要求的条件，即向量的各个坐标必须相互独立。Pecora 等人提出通过检验延迟向量各分量之间的函数依赖来评估嵌入的质量。向量坐标之间存在函数依赖意味着，

$$x(t-\tau_m) = F(x(t), x(t-\tau_1), ..., x(t-\tau_{m-1})), \tag{4.31}$$

其中， F 是某个任意函数。构造非均匀延迟嵌入，需要通过迭代构建一个时间滞后集合 $\vec{\tau} = \{\tau_1, ...\tau_{m-1}\}$，以尽可能降低各分量之间存在函数依赖的可能性。在每次迭代中，都会检验一个候选滞后 τ_i 与对应于 $\vec{\tau}$的现有滞后分量之间是否存在函数依赖。如果不存在显著的函数依赖，那么τ_i 就可以加入滞后集合。为检验这一点，作者假设 F 是光滑的，并利用连续性这一性质来量化函数依赖。

1. 考虑一个已有的 $m+1$ 维嵌入 $\vec{x}_m(t) \in \mathbb{R}^m$，它由时滞 $\tau = \{\tau_1, ...\tau_m\}$构建，并有一个待检验的潜在新嵌入时滞 τ_{m+1}。要检验新时滞的函数依赖关系，请在嵌入空间中选择一个参考点 $\vec{x}_m(t_0)$。如果存在平滑的函数依赖关系，那么连续性条件表明，在重构空间 $\mathbb{R}^m$ 中，靠近参考点（$||\vec{x}_{m,i} - \vec{x}_m(t_0)|| < \delta$）的点 $\vec{x}_{m,i}$，其时滞 $m+1^{th}$ 分量也应彼此接近（$|x_i(t-\tau_{m+1}) - x(t_0-\tau_{m+1})| < \epsilon$）。2. 计算满足以下条件的点 $\vec{x}_{m,i}$ 所占的比例 p： 其时滞分量与参考点的时滞分量之间的距离小于 ϵ。3. 将 p 与零假设进行比较；也就是说，假设这两组之间的对应关系纯属偶然。 p 的值较大，表明m维重构与新的 τ_{m+1} 时滞分量之间存在较强的关系。Pecora 等人建议使用临界值为 $p^* = 0.5$ 的二项分布来判断相对于某个选定的ϵ 是否存在函数依赖关系，因为这种方法简单且对噪声具有稳健性。4. 逐渐减小 ϵ，重复连续性检验，直到无法拒绝零假设。拒绝零假设所需的最小可能值记为 ϵ^*。对从数据中抽取的一组参考点计算该值的平均数，以得到连续性统计量 $\epsilon^*(\tau_{m+1})$。5. 将新时滞 τ_{m+1} 设为连续性统计量曲线上相对极大值所对应的时滞。重复此过程，直至达到所需的嵌入维数（根据惠特尼定理确定）。Pecora 等人还提出了一种欠采样统计量，可用作迭代选择时间延迟时的终止准则。更多详情请参阅原论文。

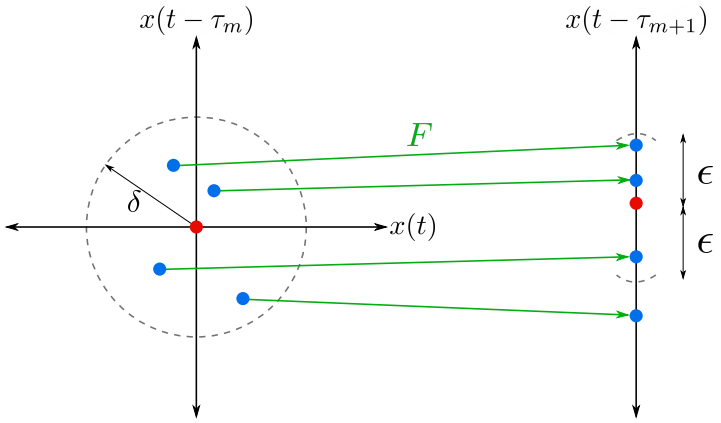


图 4.8 连续性统计量的计算

4.5.3 PECUZAL

连续性统计量方法的一个批评是，在每次嵌入迭代中选择最优时滞 τ 存在歧义。在 Pecora 等人的原论文中，相对极大值 的定义并不明确，也没有客观准则从多个可能的局部极大值中选出最佳时滞。此外，该方法也没有考虑选择不同距离 δ 来定义重构中的邻近点所产生的影响。最后，最初提出作为嵌入算法终止条件的欠采样统计量计算量很大，而且无法指示应当选择哪一个候选时滞。

Kr¨amer 等人提出，可将连续性统计量方法与 L统计量方法结合，构建一种完全自动化的非均匀嵌入延迟方法 [33]。

1. 按照 Pecora 等人的方法计算连续性统计量曲线。2. 找出与 $\epsilon^*(\tau_m)$ 的所有局部极小值相对应的滞后集合 T $= \{\tau_1^*, \ldots\ldots\}$。3. 计算当前嵌入配置的 L-统计量（初始值为$m = 1$）。（即 L_m）4. 计算将各个候选滞后纳入 T 后， L-统计量的改善幅度。5. 选择对应于 L-统计量最大降幅的滞后 τ_i^*（即找出 ΔL_i 的最大值），并将此滞后添加到滞后列表中。6. 重复步骤 1–5，直到 $|L_{m+1} - L_m| > 0$

$$\Delta L_i = L_m - (L_{m+1}|\tau_i^*)$$

4.5.4 其他方法

对于感兴趣的读者，我们列举了一些其他方法，这些方法也被提出用于解决非均匀嵌入的参数选择问题。

- 投影上的最大导数（Nichkawde，2013）——根据某种推断得到的函数依赖 F的导数，对重构吸引子的展开程度进行量化。
- 降阶自回归模型（Judd & Mees，1998）——逐步选择时滞，使其用于伪线性自回归模型时能获得良好的预测性能。使用模型描述长度测度来确定何时终止。
- 搜索优化算法——使用各种优化算法，在由用户选择的目标函数指导下搜索状态空间。示例包括蚁群优化（ACO） [34, 35]和蒙特卡洛决策树搜索（MCDTS） [36]。
- 持久链上的显著时刻（SToPS）——一种持久同调方法，用于检测并量化时间序列中伪周期结构的重要性，以指导并排序嵌入滞后选择 [19]。

第 5 章 不变量

根据嵌入理论，所给出的变换

$$y_t = (x_t, x_{t-\tau}, ..., x_{t-(m-1)\tau}) \tag{5.1}$$

可得到一个在微分同胚意义下保持系统动力学的嵌入。在最简单的均匀嵌入情形中，需要选定嵌入维数 m 和嵌入滞后 τ。对于状态预测，下一步是构建一个模型函数 F，使得，

$$y_{t+1} = F(y_t). \tag{5.2}$$

构建 F 的过程并非易事，后续章节将对此进行讨论。但仅通过考察观测数据（即系统表征），我们能了解到关于底层动力系统的什么信息？由于 y_t 和 x_t 通过一个（复杂的）变换映射相关联，我们自然会寻求找出相对于该映射保持不变的测度、量和特征。

如果一个量 $s(\{y_t\}_t)$ 不依赖于测量方式，则称其为不变量。在理想情况下，我们希望测量函数$h: \mathbb{R} \to \mathbb{R}$ 足够光滑，因为这会带来若干有用的解析性质。

定义 5.1. – 不变量
设 $s: T \times X \to \mathbb{R}$ 是一个测量某种量的函数。如果该量 s 相对于测量函数 $h: S \to \mathbb{R}, S \subset X$ 不变，则

$$s(\{x_t\}_t) = s(\{h(x_t\}_t).$$

用更数学化的术语来说，我们可以进一步细化，从而给出不变测度的定义：

定义 5.2——不变测度
设 X 是一个紧致可度量空间，B 是一个博雷尔 σ 代数（即包含 X 的所有 + 开集且在补集、可数并和可数交运算下封闭的最小集合。换句话说，这只是赋予它良好的性质）。设 $T\colon\ X \to X$ 是一个可测映射（即 T 的原像和像都是可测的 $\implies T^{-1}\mathrm{B} \subset \mathrm{B}$）。若概率测度 $\mu\colon\ \mathrm{B} \to \mathbb{R}^{+}$ 在 T 下满足 $\mu(T^{-1}B) = \mu(B),\quad \forall B \in \mathrm{B}$，则称其为不变测度。

不变测度很有用，因为它们不依赖于测量事物的方式。在实践中，现实世界的测量值会因数字化而不够平滑。但只要分辨率足够高，它们就足够平滑。此外，许多不变量所测量的是与系统动力学有关的某种“相关”量。我们可以通过计算时间平均和/或空间平均，从时间序列中估算不变量的值（有时还能给出置信区间）。有大量先进的技术工作和理论论证了取这些平均值的有效性。不过，深入探讨这些理论超出了本文的范围和目的。尽管如此，本文仍作简要讨论，旨在介绍相关概念，并为感兴趣的读者提供入门指引。

5.1 遍历性与自然测度

研究需要计算空间平均和时间平均的不变量时，考虑遍历性这一性质很有帮助。简而言之，如果相空间中几乎每个点的轨迹最终都会任意接近其他所有点，那么该系统就是遍历的。回想一下，在有界区域上，保测映射具有这一性质，这是由庞加莱复现定理保证的。

定理 5.1. - 遍历轨道
集合 S 中的时间演化是遍历的，当且仅当所有遍历分量 $R \in S$ 要么具有 $\mu(R) = 0$，要么具有 $\mu(R) = \mu(S)$。或者说，若不变测度 $\mu \in \mathcal{M}(\mathcal{X}, \mathcal{T})$ 满足：对于某个 $B \subset \mathcal{B}$，只要 $T^{-1}B = B$，则必有 $\mu(B) = 0$ 或 $\mu(B) = 1$，那么该不变测度就是遍历的。

事实证明，在尝试估计不变量时，遍历轨道的存在能带来一些非常有用的性质。遍历性保证轨迹最终会遍历系统的整个吸引子流形（不变集）。因此，时间平均的计算可以等同于空间平均的计算。本质上，遍历性为我们提供了两种不同的方式来进行计算不变测度所需的平均：

1. 对时间进行平均（即跨越多次迭代，或一条长期轨迹）

2. 对不变密度 $\mu(x)$ 进行加权积分（即任意给定时间观测状态的长期概率分布。）

不变密度，

$$d\mu = \mu(x)dx,$$

也称为吸引子上的**自然测度**，是典型轨迹在无限小的相空间元素 dx中所花费的平均时间。当所有轨迹最终都会遍历各处时，若此测度独立于初始条件，则可保证遍历性。由此，我们可以用不变密度来描述某个量 O 的长期平均，

$$\langle O \rangle_t = \int O(x)\mu(x)\,dx. \tag{5.3}$$

如果给定一条很长的轨迹，并根据观测到的位置 x 在状态空间中构建直方图，就可以数值计算自然测度 $\mu(x)$：

$$\mu(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \delta^d(x - y_k), \quad y_k \in \mathbb{R}^d \tag{5.4}$$

其中， δ^d 是狄拉克δ函数。自然测度 $\mu(x)$是不变的，因为在正向映射 F下它保持不变。将长度为 N 的长轨迹向前迭代一步，相当于向轨迹 x_{N+1} 追加一个新值，并移除 x_1。

定理 5.2. - 自然测度的不变性
设 $\mu(x)$ 为状态空间 $S \subset \mathbb{R}^d$中一个由正向映射 F （即 $y_{k+1} = F(y_k)$）刻画的动力系统的自然测度。设 $f\colon\ S \to \mathbb{R}$ 为一个可测量量。 f在 S 上的空间平均为

$$\begin{aligned}\bar{f} &= \int_S \mu(x) f(x)\, dx^d \qquad \text{– (spatial average)} \\ &\approx \frac{1}{N}\sum_{i=1}^{N} f(y_k) \\ &= \frac{1}{N}\sum_{i=1}^{N} f(F^{k-1}(y_{1)}) \qquad \text{– (time average)}\end{aligned}$$

为证明在前向映射 F 下的不变性，

$$\begin{aligned}\int_S \mu(x) f(F(x))\, dx^d &= \bar{f} + \frac{1}{N}\left[f(y_{N+1} - f(y_1))\right] \\ &\approx \bar{f} \quad \textit{for } N \to \infty\end{aligned}$$

因此，对于无限长轨迹，自然测度在前向映射下是不变的。

鉴于自然测度具有便于使用的性质，现在计算有趣的不变测度取决于选择 f 的 “ 有趣 / 有用 ” 函数。

5.2 李雅普诺夫指数

回顾一下，混沌动力系统的一个重要特征是对初始条件的敏感依赖（SDIC）。也就是说，初始条件非常接近的轨迹会以指数速度迅速分离。简单来说，这可以部分地用以下描述来说明，

$$|\delta(t)| \approx \delta_0 e^{\lambda t}, \tag{5.5}$$

其中， δ 是两条轨迹之间的分离距离， λ 是一个称为最大李雅普诺夫指数的量。混沌行为（对初始条件的敏感依赖，SDIC）的一个指标是最大李雅普诺夫指数为正 λ。事实上，正如我们将要看到的，李雅普诺夫指数对于遍历系统是不变的。我们将介绍两种研究李雅普诺夫指数的不同方法：首先在第 5.2.1 节采用更偏理论的方法，然后在第 5.2.2 节采用计算方法。

5.2.1 李雅普诺夫指数（艰难的方法）

考虑一个 C^r $(r \geq 1)$向量场，

$$\dot{x} = f(x), \quad \in \mathbb{R}^n. \tag{5.6}$$

设 $x(t, \ x_0)$ 是该系统的一条解轨迹，初始条件为 $x(0, \ x_0) = x_0$。李雅普诺夫指数关注初始时彼此接近的轨道之间的散度，因此我们想了解与 $x(t, \ x_0)$ 附近及相对于它的轨道吸引和/或排斥相关的几何特征。我们可以将其与不动点的稳定性分析进行类比。研究不动点时，我们会在不动点附近进行线性化，并观察从不动点某个邻域内出发的状态的行为（即吸引/排斥）。分析李雅普诺夫指数，类似于将不动点替换为整条轨道。

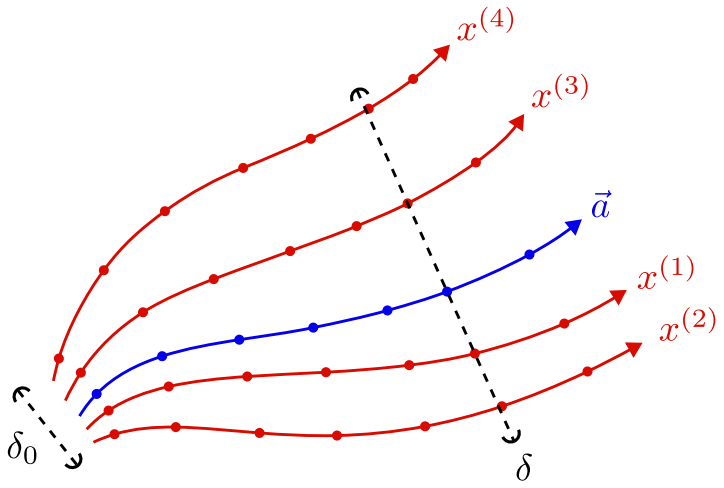


图 5.1 相对于一条轨迹的线性化 a

如何围绕一条轨道对上述系统进行线性化？为此，我们考虑用一系列 N 个点来替代轨道 $x(t, \ x_0)$。

$$\begin{aligned} \vec{a} &= [a_0, a_1, ..., a_N] \\ &= [x(0, x_0), x(t_1, x_0), ..., x(t_{N-1}, x_0)] \end{aligned} \tag{5.7}$$

然后，我们继续对轨道上每个点附近的系统进行线性化。为此，设存在坐标变换 $\xi_i = x - a_i$，并将 a_i 视为不动点。根据线性化定理，我们可以推导出如下结果，

$$x_i = \xi_i + a_i \tag{5.8a}$$
$$\dot{x}_i = \dot{\xi} \tag{5.8b}$$
$$\begin{aligned} f(x_i) &= f(a_i) + f'(a_i)(x - a_i) + \dots \\ &\approx 0 + f'(a_i)\xi_i \end{aligned} \tag{5.8c}$$

因此，对点 a_i 附近进行线性化后，我们可以写出如下线性等价的动力系统

$$\dot{\xi}_i = f'(a_i)\xi_i. \tag{5.9}$$

对轨道上每个点 a_i 重复上述步骤，我们正在围绕这些点进行线性化，可以写出一个描述轨道 $x(t,\ x_0)$ 附近线性化动力学的常微分方程系统，

$$\begin{aligned} \dot{\xi}_0 &= Df(x(0,x_0))\xi_0 \\ \dot{\xi}_1 &= Df(x(t_1,x_0))\xi_1 \\ &\vdots \\ \dot{\xi}_{N-1} &= Df(x(t_{N-1},x_0))\xi_{N-1}, \end{aligned} \tag{5.10}$$

(5.11)

其中每个 ξ_i 实际上都是定义在 $\mathbb{R}^n$ 中的向量。我们可以将此记法简化，写成下面的矩阵常微分方程，

$$\dot{\boldsymbol{\xi}} = Df(x(t,x_0))\boldsymbol{\xi} \tag{5.12}$$

这个系统是表示 $N \to \infty$ 的恰当形式。由于这是一个矩阵常微分方程，其解也是 $n\times N$ 矩阵。令 $X(t;\ x(t,\ x_0))$ 为轨道 $x(t,\ x_0)$ 附近线性化系统的基本解矩阵。假设我们要测量 $x(t,\ x_0)$ 周围邻域沿 $\vec{v}$ 方向的膨胀或收缩程度。我们可以定义沿经过 x_0 的轨迹、沿方向 $v \in \mathbb{R}^n$ 的膨胀系数，

$$\chi_t(x_0, v) = \frac{||X(t;x(t,x_0))v||}{||v||}. \tag{5.13}$$

注意，$\chi_t(x_0,\ v)$ 是一个与时间无关、但也取决于特定轨道的量。假设存在指数发散的形式（即 χ_t 按指数规律变化），我们可以定义沿经过 x_0 的轨迹、沿方向 v 的李雅普诺夫指数，

$$\lambda(X(t;x(t,x_0)),x_0,v) = \sup \lim_{t\to\infty} \log \chi_t(x_0,v). \tag{5.14}$$

一些重要说明

1. λ 是一个渐近量，因此假定对于所有 $t > 0,\ x(t,\ x_0)$都存在。这要求相空间是紧致无边流形，或者 x_0 位于

一个正向不变区域内（关于负向不变性也有一些有趣的结果）。

2. $\lambda(X(t;\, x(t,x_0)),\ \ x_0,\vec{0}) = -\infty$
3. 由于基本解 X 与轨迹有关，因此严格来说，结果也取决于轨迹。然而，如果系统是遍历的，这并不重要，因为只要 $t \to \infty$足够大，轨迹总会遍历相空间的所有部分。

4. $\lambda(X(t;\, x(t,x_0)), x_0, v) = \lambda(X(t;\, x(t,x_0)), x_1, v)$ if $x_1 = x(t_1, x_0)$ for $t_1 > 0$. 因此，我们可以简化记法

$$\lambda(X(t; x(t,x_0)), x_0, v) = \lambda(v)$$

定理 5.3. – 李雅普诺夫指数的几何性质
对于任意向量 f、$g \in \mathbb{R}^n$ 以及非零常数 $c \in \mathbb{R}$,

$$\lambda(f+g) \le \max\{\lambda(f), \lambda(g)\} \tag{5.15}$$
$$\lambda(cg) = \lambda(g) \tag{5.16}$$

定理 5.4。 – 李雅普诺夫方向与子空间
对于任意 $r \in \mathbb{R}$,

$$\{g \in \mathbb{R}^n | \lambda(g) \le r\}$$

是 $\mathbb{R}^n$ 的向量子空间。

该结果表明，对于维数为 n 的状态空间，与该轨迹相关联的李雅普诺夫指数至多有 n 个。更准确地说，我们可以按照特定方式排列这些李雅普诺夫指数，从而形成一个李雅普诺夫谱。

命题 5.1。 – 李雅普诺夫谱与嵌套子空间
数的集合

$$\{\lambda(g)\}_{g\in\mathbb{R}^n, g\neq 0}$$

至多取 $n = dim(\mathbb{R}^n)$ 个值 ν_i，这些值可以按如下方式排序

$$\nu_1 > \nu_2 > \ldots > \nu_s, \quad 1 \le s \le n.$$

此外，令 $L_i = \{g \in \mathbb{R}^n | \lambda(g) \le \nu_i\}$。那么

$$\{0\} = K_n \subset \ldots \subset L_{s+1} \subset L_s \subset \ldots \subset L_1 = \mathbb{R}^n,$$

使得 $L_{i+1} \neq L_i$，且 $\lambda(g) = \nu_i$ 当且仅当 $g \in L_i \setminus L_{i+1}$, $1 \le i \le s$。李雅普诺夫指数 ν_i 的重数由 k_i 给出，

$$k_i = dim(L_i) - dim(L_{i+1})$$

上述关于李雅普诺夫谱的命题表明，只有当对应的子空间L_i 缩小时，李雅普诺夫指数才能严格递减排列。因此，李雅普诺夫谱对应的方向构成一个基，逐步构建出状态空间 $\mathbb{R}^n$。

与分析流和映射中的不动点时需要确立轨迹的存在性与唯一性类似。接下来我们要问：在什么条件下，上确界极限（因此，相关轨迹的李雅普诺夫指数）存在？为此，我们还需要给出另外两个定义。

定义 5.3. – 正规基
若 {v1，……，v6} 是 $\mathbb{R}^n$ 的一个基，且满足

$$\sum_{i=1}^{n} \lambda(e_i) \leq \sum_{i=1}^{n} \lambda(f_i),$$

其中 { {v1，……， v4} 是 $\mathbb{R}^n$ 的任意其他基。换言之，正规基对应的李雅普诺夫谱之和最小。

定义 5.4. – 正则族
基本解矩阵 $X(t;\, x(t,\ x_0))$ 称为正规，因为 $t \to \infty$ 如果
1. $\lim_{t\to\infty} \frac{1}{t} \log\, |\det\, X(t;\, x(t,\ x_0))|$ 存在且为有限值，
并且 2. 对于每个正规基 $\{e_1,\ \ldots,\ e_n\}$

$$\sum_{i=1}^{n} \lambda(e_i) = \lim_{t\to\infty} \frac{1}{t} \log |\det X(t; x(t, x_0))|$$

定理 5.5。 – 李雅普诺夫指数的存在性
如果 X（t；x（t、x_0）） 是正规的，因为 $t \to \infty$，则

$$\lambda(e) = \lim_{t\to\infty} \frac{1}{t} \log \chi_t(x_0, e)$$

对于任意向量 $e \in \mathbb{R}^n$，其存在且为有限值

只要存在一个正则的基本解矩阵 $X(t,\ x(t,\ x_0))$，李雅普诺夫指数就存在且定义良好。剩下要做的就是检验这一性质。**乘法遍历定理**解决了这个问题。该定理表明，对于某个不变测度，几乎所有轨迹（即除一个零测度集之外的所有轨迹）都会产生正则的基本解矩阵。

5.2.2 李雅普诺夫指数（简单方法）

在了解了李雅普诺夫指数的一般理论和基本思想之后，我们接下来讨论一种更实用的方法，以理解并最终计算李雅普诺夫指数和谱。如前所述，混沌系统具有**对初始条件的敏感依赖（SDIC）**这一特性，这意味着空间中无限接近的向量会产生以指数速度发散的轨迹。这种散度可以用线性化动力学和李雅普诺夫指数的工具来完整描述。

考虑由映射 F 描述的动力系统中的两条不同轨迹 $\mathbf{x}_n$ 和 $\mathbf{y}_n$，它们在 m 维状态空间中起初彼此相近。我们可以利用这种表述写出一个以轨迹 x_n 为中心的线性映射，用来描述这对轨迹之间的距离，

$$\begin{aligned}\mathbf{y}_{n+1} - \mathbf{x}_{n+1} &= F(\mathbf{y}_n) - F(\mathbf{x}_n) \\ &= J_n(\mathbf{y}_n - \mathbf{x}_n) + O(||\mathbf{y}_n - \mathbf{x}_n||^2),\end{aligned} \tag{5.17}$$

其中，$J_n = J(x_n)$ 是 F 在 x 处的雅可比矩阵。为简单起见，我们将扰动表示为 $\boldsymbol{\delta}_n = \mathbf{y}_n - \mathbf{x}_n$，并写出等价的系统

$$\boldsymbol{\delta}_{n+1} \approx J_n \boldsymbol{\delta}_n. \tag{5.18}$$

聚焦于时间步 n，令 $\mathbf{e}_i$ 为 J 的一个特征向量，其对应的特征值为 Λ_i。因此，我们可以用 J_n 的特征向量及基系数 β_i 来表示 $\boldsymbol{\delta}_n$，其中，

$$\boldsymbol{\delta}_n = \sum_{i=1}^{m} \beta_i \mathbf{e}_i. \tag{5.19}$$

由此可得，

$$\boldsymbol{\delta}_{n+1} = \sum_{i=1}^{m} \beta_i \Lambda_i \mathbf{e}_i. \tag{5.20}$$

简单来说，我们将初始扰动表示为各个 m 线性特征子空间的组合，并估算其在每个特征向量方向上的收缩或扩张。

到目前为止，由于我们是针对特定点 $\mathbf{x}_n$进行线性化，因此分析在时间和空间上都是局部的。显然，随着 n 变化且轨迹移动到状态空间中的另一个点，J_n,$\mathbf{e}_i$ 和 Λ_i 的值也会变化。为了刻画整个系统，我们需要定义全局对象和量。我们可以通过对不同的拉伸因子和方向取平均值来实现这一点。为此，我们注意到，沿长度为$N-1$ 的轨迹进行向前迭代，等价于将 N 线性化矩阵 J_n 相乘。依照这一思路，我们可以针对方向向量 $\mathbf{u}_i^{(N)}$ 写出以下特征向量方程，

$$\left(\prod_{n=1}^{N} J_n\right)\mathbf{u}_i^{(N)} = \Lambda_i^{(N)}\mathbf{u}_i^{(N)}. \tag{5.21}$$

接下来，我们可以将 i^{th} 全局李雅普诺夫指数 λ_i 定义为沿轨迹取平均值，

$$\lambda_i = \lim_{N\to\infty}\frac{1}{N}\ln|\Lambda_i^{(N)}|. \tag{5.22}$$

与其理论表述中的结论类似，上述极限的存在性与唯一性由**乘法遍历定理**给出。这是一个高度非平凡的结果，尤其是因为矩阵的乘法不可交换，并且对数运算不能与特征值的形成交换。对于一维映射，雅可比矩阵是一个实数，上述定义可化为：

$$\lambda = \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\ln|f'(x_n)|. \tag{5.23}$$

对于 $\mathbb{R}^m$中的状态空间， m 个不同指数的集合称为李雅普诺夫谱。除去不导致自然不变测度的零测度初始条件集合，对于所有初始条件，该谱都是相同的。此外，李雅普诺夫指数在状态空间的光滑变换下保持不变。如果由 F 给出的动力学是可逆的，则 F 的每个不变测度也在时间反演动力学 F^{-1}下保持不变。计算这些值将得到一个李雅普诺夫谱，其幅值相等但符号相反。与李雅普诺夫指数相对应的特征向量对应于轨迹的拉伸与收缩方向。这些方向与全局不变流形相切（即稳定与不稳定流形）。

5.2.3 Wolf 算法

在实践中，我们希望根据数据计算李雅普诺夫指数。该如何进行呢？一种方法是采用 Wolf 算法 [37, 38],，它用于计算完整的李雅普诺夫谱，具体如下：

1. 给定时间序列 $x(t)$，使用由延迟向量构成的时间延迟嵌入进行相空间重构：

$$\mathbf{y}(t) = (x(t), x(t-\tau), x(t-(d-1)\tau)).$$

2. 取一个初始点 $\mathbf{y}(t_0)$，并找到它的最近邻 $\mathbf{y}^{NN}(t_0) = \mathbf{y}(t_0^{NN})$，同时确保排除时间最近邻（即 Theiler 窗口）3. 计算这对点之间的距离，

$$\begin{aligned} L(t_0) &= |\mathbf{y}(t_0) - \mathbf{y}^{NN}(t_0)| \\ &= |\mathbf{y}(t_0) - \mathbf{y}(t_0^{NN})| \end{aligned}$$

4. 在稍后的时间 $t_1 = t_0 + n$，其中 n 是步数，计算新的距离，

$$L'(t_1) = |\mathbf{y}(t_0 + n) - \mathbf{y}(t_0^{NN} + n)|$$

请注意，n 的选择应确保 $L(t_0^*)$ 不超过吸引子的尺度（例如，避免吸引子的折叠效应）。

5. 令 $t_1 = t_0 + n$，并重复步骤 2–4。

6. 对所需的所有基准点 M（t_M）重复上述过程。然后用它来计算第一个（最大）李雅普诺夫指数，

$$\lambda_1 = \frac{1}{t_M - t_0} \sum_{k=1}^{M} \log \frac{L'(t_k)}{L(t_{k-1})}$$

7. 对于下一个最大的正李雅普诺夫指数（如果存在）λ_2，以一条基准轨迹为参照重复上述过程。但这次不选择单个最近邻，而是选择两个最近邻，并计算这些点形成的面积 $A(t_0)$，而不是长度 $L(t_0)$。同样，当 $A(t_0^*)$ 超过吸引子的尺度时，在基准轨迹附近重新选取另外两个点，使所得三角形平面的方向尽可能接近 $A(t_0^*)$。按照与步骤 8 相同的公式，可得 $\lambda_1 + \lambda_2$ 的值，随后可用它来计算 λ_2。

8. 根据需要，对正李雅普诺夫指数重复上述步骤。

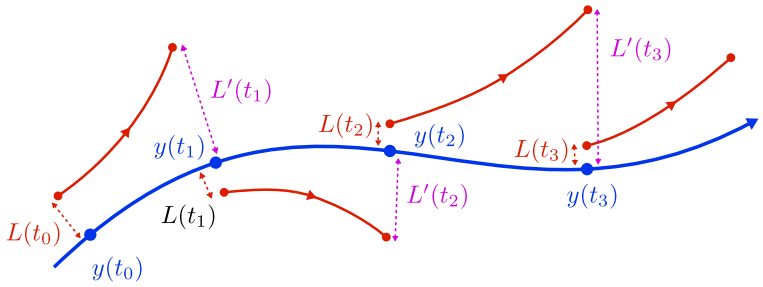


图 5.2 用于计算最大李雅普诺夫指数的 Wolf 算法 λ_1

在实践中，该算法通常仅用于获取第一个（最大）李雅普诺夫指数。这种截断算法也被称为 Rosenstein 算法 [38]。这些算法相对容易理解和实现。然而，它们存在一些缺陷：

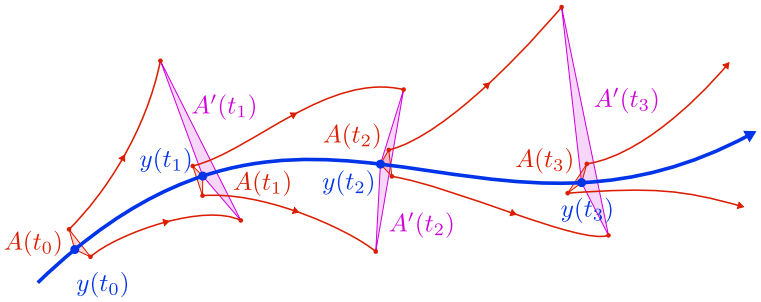


图 5.3 用于计算前两个李雅普诺夫指数的 Wolf 算法 $\lambda_1 + \lambda_2$

- 通过替换进行持续重置并不能克服与混沌轨迹相关的不准确性
- 只跟踪一条轨迹。

有几种方法可以解决这一问题：

- 计算参考点邻域内一组距离的平均值（而不仅仅是最近邻的距离）。
- 选择一个时间跨度
- 选择许多初始点 n_0，而不是单条轨迹。

在这些解中，最后一种方法带来了最显著的改进。为此，改为跟踪最近邻球体的平均点间距离，以表示其轨迹的散度，

$$S(n) = \frac{1}{N}\sum_{n_0=1}^{N} \ln\Big(\frac{1}{|\mathcal{U}(B_{n_0})|}\sum_{y_s \in \mathcal{U}(B_{n_0})} |y_{n_0+n} - y_{s+n}|\Big). \tag{5.24}$$

这里，$\mathrm{U}(B_{n_0})$ 是一个近邻球体，其点间距离均在 ε 以内。$S(n)$ 是针对固定的 ε 和嵌入计算得到的剖面。必须在一系列 ε 和 m—嵌入维数范围内绘制许多曲线。这样才能在其中识别出线性标度区间，并以估计斜率作为最大李雅普诺夫指数的估计值。

考虑 $\Delta_0 = S_n - S_{n'}$和 $\Delta_l = S_{n+l} - S_{n'+l}$。如果 $|\Delta_l \approx |\Delta_0|e^{\lambda l}$，那么最大李雅普诺夫指数 $\lambda_1 \approx \frac{1}{l}\ln\frac{|\Delta_l|}{|\Delta_0|}$。我们对不同嵌入维数 $m \geq m_0$下的轨迹 $\{s_n\}$ 取平均，从而

$$S(\varepsilon, m, t) = \langle \ln \frac{1}{|\mathcal{U}_n|}\sum_{s_{n'} \in \mathcal{U}_n} |s_{n+t} - s_{n'+t}|\rangle_n. \tag{5.25}$$

确定许多 U_n 和 $t = 1, 2、\ldots\ldots、T$，但不要延伸到太遥远的未来。将 S（以对数刻度表示）相对于 t 作图，可以揭示一个标度区域（直线部分），其斜率给出 λ_1（见图5.4）。

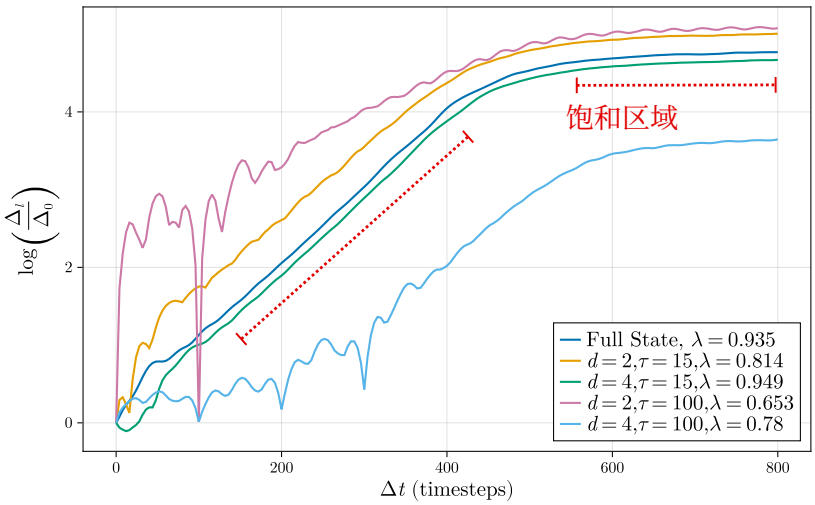


图 5.4 将李雅普诺夫指数作为线性标度区间提取。不同曲线展示了嵌入维数过小或延迟滞后过大的影响。

如果我们能够可靠地估计李雅普诺夫谱的值，那么李雅普诺夫指数与其他动力学不变量之间就存在有趣且实用的关系。

佩辛恒等式 通过以下关系将柯尔莫哥洛夫-辛奈熵（请自行查阅）与正李雅普诺夫指数联系起来：

$$h_{KS} = \sum_{i:\lambda_i>0} \lambda_i. \tag{5.26}$$

卡普兰-约克猜想 将信息维数与李雅普诺夫指数联系起来。具体而言：

$$D_{KY} = k + \frac{\sum_{i=1}^{k} \lambda_i}{|\lambda_{k+1}|} \tag{5.27}$$

其中 $\sum_{i=1}^{k} \lambda_i \geq 0$ 和 $\sum_{i=1}^{k+1} \lambda_i < 0$。该维数也称为李雅普诺夫维数，并被推测等于信息维数。

5.2.4 伪李雅普诺夫指数

考虑状态空间 $\mathbb{R}^d$中的动力系统。回顾一下，只要满足 $m \geq 2d + 1$，Takens 嵌入定理便能提供一个状态空间重构，其动力学和流形与原始动力学微分同胚。根据上述关于李雅普诺夫指数的细节， $\mathbb{R}^m$ 中的动力系统在其谱中至多有 m个唯一的李雅普诺夫指数。

现在，在大多数情况下，我们不可避免地需要嵌入观测时间序列数据，并且所有分析都将在 $\mathbb{R}^m$ 中进行。因此，直接进行数值计算会得到最多包含 m 个李雅普诺夫指数的谱。然而，我们从中得知，这样的系统最多只有 $d < m$ 个李雅普诺夫指数，因为原始动力学是在 $\mathbb{R}^d$ 中定义的。这就带来了一个问题：我们该如何处理额外的李雅普诺夫指数，它们是否重要?

我们可以得出一个结论：原始系统的指数不会超过 d 个。因此，存在 $d-m$ 个伪指数，它们与动力学无关。那么，如何识别哪些指数是伪指数呢? Parlitz [39], 提议，可以考察李雅普诺夫指数的时间可逆性质。如果反转轨迹的时间排序并重新计算李雅普诺夫谱，理论上只有真实的李雅普诺夫指数会保留下来，但符号会变为负号。其他指数都将被归类为伪指数。然而，Kantz 和 Schreiber [40] 指出，这种方法在实践中效果并不理想，因此大多数计算只关注最大指数 λ_1。

5.3 分形维数与标度

首先以有理数的密度分布为例。虽然还有更正式的定义，但在数轴上，这意味着对于任意一对有理数 i 和 j，都可以找到另一个有理数 k，使得 $i < k < j$。现在，给定一个较小的 $\epsilon \in \mathbb{R}$，需要多少个半径为 ϵ 的开球才能覆盖所有有理数? 将这个数量记作 $N(\epsilon)$。直观而言，这个问题关注的是某种几何性质（例如点密度）的标度性质。有理数的情况相当简单，因为它们均匀稠密，因此我们预期 $N(\epsilon) \propto \frac{1}{\epsilon}$。开球的数量随粒度呈线性增长。但这对于所有数学对象都成立吗?

作为对均匀分布且稠密的有理数的替代，我们考虑一个来自地理学的示例。挪威的主要陆地块体由崎岖地形构成，周围有众多峡湾和形状不规则的海岸线边界。挪威的海岸线有多长? 正如人们所预期的那样，这个问题的答案取决于用于计算海岸线边的分辨率。较粗的分辨率会忽略峡湾的细节边界，并提供平滑的边界。随着分辨率逐渐提高，计算得到的海岸线长度也会相应增加。鉴于这一边界具有（近似的）自相似性，

人们可以问：海岸线长度如何随分辨率变化？在研究分形结构时，如何将这一点推广？

5.3.1 盒计数维数

盒计数维数（分形维数）为我们提供了一种解答这一标度问题的方法。考虑 $S \in \mathbb{R}^n$ 中的一个曲面，以及边长为ϵ的网格盒。令 $N(\epsilon)$表示覆盖 S所需的盒子数量。在最简单的情况下，长度为 L 的线段将按以下方式缩放

$$N(\epsilon) \approx \frac{L}{\epsilon^1}. \tag{5.28}$$

类似地，对于面积，

$$N(\epsilon) \approx \frac{A}{\epsilon^2}, \tag{5.29}$$

对于体积，

$$N(\epsilon) \approx \frac{V}{\epsilon^3}. \tag{5.30}$$

按照这一逻辑，对于某种维数概念为d的流形，一般有

$$N(\epsilon) \approx \frac{C}{\epsilon^d}, \tag{5.31}$$

其中 C 是一个常数，用来度量流形。我们可以利用这个定义推导出 d的表达式

$$\begin{aligned} \log N(\epsilon) &\approx \log C - d \log \epsilon \\ d &= \frac{\log N(\epsilon) - \log C}{-\log \epsilon}. \end{aligned} \tag{5.32}$$

据此，我们可以定义盒计数维数

定义 5.5. – 盒计数维数
流形的盒计数（分形）维数 d 定义为

$$d = \lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{-\log \varepsilon}$$

对于分形等简单数学对象，可以精确计算这个量。作为示例，我们来考察康托尔集。回顾一下，康托尔集通过以下算法构造

1. 令 C_0 为单位区间 $[0, 1]$。

2. 从 C_0 中移除中间三分之一的线段。令所得的两个不相交区间的并集为 C_1。3. 对 C_n 中的每个不相交区间，移除其中间的三分之一，并令所有所得不相交区间的并集为 C_{n+1}4. 康托尔集定义为

$$C = \lim_{n\to\infty} C_n$$

C 的一个显著性质是，它是测度为零的集合的一个例子。考虑康托尔集的第 1st 次迭代（C_∞）。这个集合用直径为 $\epsilon_1 = 1$ 的单个球即可最有效地覆盖。类似地，对于 C_ϵ，我们需要两个直径各为 $\epsilon_2 = \frac{1}{3}$ 的开球。一般而言，第 n^{th} 次迭代可用 2^n 个球覆盖，每个球的直径为 $\frac{1}{3^n}$。因此，当 $n \to \infty$ 时，康托尔集的盒计数维数计算如下：

$$\begin{aligned} d &= \lim_{\epsilon\to 0} \frac{\log 2^n}{-\log \frac{1}{3^n}} \\ &= \frac{\log 2}{\log 3} \\ &= 0.6309 \end{aligned} \tag{5.33}$$

盒计数维数可以理解为豪斯多夫维数的一个粗略上界（豪斯多夫维数也衡量平滑度）。这两个量通常只有在某些构造出来的例子中才会有所不同 [40]。

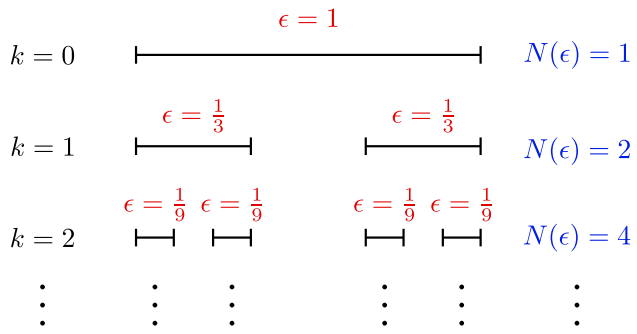


图5.5 康托尔集的构造

5.3.2 相关性与广义维数

回想一下，吸引子和有界动力系统在相空间中具有某种自然测度，用以描述点密度分布。

此外，自然测度通常并非均匀分布。从洛伦兹吸引子两翼上的轨道分布可以看出这一点：轨迹沿着分形轨道带聚集。因此，仅凭吸引子的支撑集维数（即其所在的集合或流形），来确定吸引子的标度维数并不十分准确。理想情况下，我们还需要考虑点密度，并对吸引子的不同区域进行适当加权。

为了考虑系统的自然测度，令 $p_\epsilon(x) = \int_{\mathcal{U}_\epsilon} d\mu(x)$ 表示在球 $\mathrm{U}_\epsilon(x)$ 中找到一条典型轨迹的概率。因此，我们可以定义以下量，称为关联积分，它描述了某种形式的自相似性，

$$C_q(\epsilon) = \int_x (p(x)_\epsilon)^{q-1} d\mu(x) = \langle (p_\epsilon)^{q-1} \rangle, \tag{5.34}$$

其中 q 是一个整数参数。与盒计数维数一样，我们可以定义一种指数标度关系，用以描述某种形式的维数，

$$C_q \propto \epsilon^{(q-1)D_q}, \quad \epsilon \to 0 \tag{5.35}$$

取等号并求极限，可得到广义维数的表达式

定义 5.6——广义维数
阶数为 q 的广义维数定义为，

$$D_q = \lim_{\epsilon \to 0} \frac{1}{q-1} \frac{\ln C_q(\epsilon)}{\ln \epsilon}$$

其中 $C_q(\epsilon)$ 是关联积分，

$$C_q(\epsilon) = \int_x (p(x)_\epsilon)^{q-1} d\mu(x) = \langle (p_\epsilon)^{q-1} \rangle$$

广义维数的前几阶有专门的名称：

- $q = 0$：容量维数／盒计数维数
- $q = 1$：信息维数
- $q = 2$：关联维数（共同出现的测度）

通常，如果 $q_1 > q_2$，那么 $D_{q_1} < D_{q_2}$。

*5.3.3 Grassberger-Procaccia*算法

关联维数包含有关系统吸引子的分形特征和标度性质的有用信息，因此我们希望找到一种计算它的方法。常用的计算方法是 Grassberger-Procaccia算法 [41]。为此，我们首先嵌入到 $d_E > 2m + 1$ 中，其中 m 是真实维数。由于

我们不知道 m，所以要对多个 d_E 进行计算。考虑相关和，

$$C(\varepsilon, N) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \Theta(\varepsilon - \|x_i - x_j\|), \tag{5.36}$$

其中，$\Theta(x)$ 是赫维赛德阶跃函数，而 $C(\varepsilon, N)$ 则针对彼此距离不超过 ε 的点对进行计算。

定义 5.7. – 关联维数
关联维数定义为

$$D_2 = \lim_{\varepsilon \to 0} \lim_{N \to \infty} \frac{\log C(\varepsilon, N)}{\log \varepsilon}.$$

也就是说，假设相关和按幂律标度，即对于较小的 ε 和足够大的 d_E，$C(\varepsilon) \approx \varepsilon^{D_2}$。若我们取 $N \to \infty$ 的极限，相关和便趋近于相关积分，

$$C \to C(\epsilon, N) := \int \int \Theta(\|x - y\| < \varepsilon) d\mu(x) d\mu(y). \tag{5.37}$$

Grassberger-Procaccia算法直接计算相关和，并绘制 $\log C(\varepsilon, N)$ 随 $\log \varepsilon$ 变化的图像。随后找到一个线性标度区间，并近似估算其斜率，从而得到 D_2。与大多数用于估算不变测度的算法一样，最佳实践要求生成具有不同初始条件的多条轨迹，并在识别近邻时使用 Theiler 窗口。

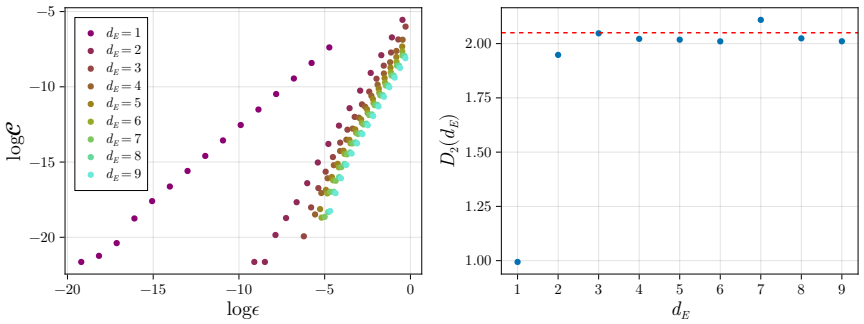


图 5.6 使用相关和的线性标度区间，根据包含 $\tau = 15$ 个时间步的数据推断洛伦兹数据的关联维数，并改变嵌入维数 d_E。右图中的红色虚线对应真实的相关维数值 2.05

请注意，拟合误差估计并不是关联维数估计的置信区间。噪声会填满可用空间，因此在存在噪声或高

含噪信号中，应预计 D_2 为嵌入维数。务必查看（并展示）关联和图，然后根据数据决定提取 D_2 是否有用，即展示标度图以及用于估计 D_2 的区域。

Grassberger-Proccacia 算法存在的问题之一，是它假设标度区域存在，然后再对其位置进行最佳猜测。Kantz 和 Schreiber [40] 讨论了克服这一问题的各种补救措施。顺便指出，其他维数也已有相关描述——其中最常见的是盒计数维数 d_0、豪斯多夫维数和信息维数。这些维数与关联维数的差异，在于关联积分中用于测量接近程度的度量选择不同。

下文将介绍适用于有限数据集的、比关联维数更为实用的替代方法。我们首先列出普通 Grassberger-Procaccio 方案面临的若干挑战。

- 在 G-P 实现中，首先假定标度区域存在，然后再寻找它。
- 在大尺度下，关联积分中的标度受到吸引子大小的限制。这一点可以从剖面的饱和现象中看出。
- 在较小的长度尺度下，即使数据没有噪声，量化效应也会产生伪影，改变 log(C) 的估计值（这是另一个问题）；量化会造成混乱。
- 在较小的长度尺度下，点间距离的分布会产生偏差，因为这些点（来自一条轨迹）并非相互独立。
- 较小尺度下的点较少，而较大尺度下的点很多；由于用于估计关联积分的样本量会随尺度变化，因此会引入统计相关性。
- 由于上述所有因素，标度区域始终覆盖有限的长度尺度范围——这反映的是吸引子的确定性结构特性，而非其渐近特性。

关联维数估计值的方差（误差）尚无现成的估计方法。

5.3.4 Judd 算法

Judd [42]提出了对Grassberger-Procaccia算法的一项改进，将关联积分 C 的公式表示为概率分布的形式。这为关联维数提供了以下替代公式

$$d_c(V,\mu) = \limsup_{\epsilon \to 0} \frac{\log \Pr(||X-Y|| < \epsilon | X, Y \in V)}{\log \epsilon}, \tag{5.38}$$

其中 μ 是自然测度， V 是相空间中的集合（例如吸引子）。在这里，相关和 $C(\epsilon)$可视为.....的估计量

$$\rho_V(\epsilon) = \Pr(||X - Y|| < \epsilon | X, Y \in V). \tag{5.39}$$

Judd 算法的第二次改进是将吸引子建模为两个组成部分的组合 $V = U \times Z$，其中 Z 是光滑流形的一个连通子集，在拓扑上等价于欧几里得空间的一个闭子集，而U 是类似“康托尔集”的分形部分。此方法用于证明以下结论

$$\rho_V \approx \epsilon^{dC} p(\epsilon) \tag{5.40}$$

其中，ϵ^{dC} 与由 U导致的分形自相似性有关，而 $p(\epsilon)$对应于 Z的光滑曲率。此外，$p(\epsilon)$被建模为 k次多项式，通常 $k = 1,2,3$。这种方法的动机在于，集合 V 上随机点的间距主要由 U 的自相似结构（指数函数项）以及光滑流形的弯曲和扭转（多项式项）决定 Z。多项式项取决于我们试图估计的底层集合，因此也需要根据数据进行估计。感兴趣的读者可查阅原论文了解详情。

5.3.5 高斯核算法

Grassberger-Procaccia 和 Judd 算法都没有考虑噪声的影响。对此予以考虑非常重要，因为当分辨率等于噪声幅值时，噪声可能会有效地破坏自相似分形结构，从而导致高估关联维数。一种替代方法是假设所有观测值 X 和 Y 都包含观测高斯噪声。为此，Diks 等人 [43] 和 Yu 等人 [44] 通过用高斯核替换 Heaviside 函数来修改原始关联积分。

定义 5.8。– 高斯核关联和
假设观测值受到高斯噪声污染，相关和表示为:

$$T(h,m) = \int \int w\left(\frac{||x-y|| - \epsilon}{h}\right) d\mu(x) d\mu(y)$$

其中 h 是带宽参数，m 是时间延迟嵌入的嵌入维数，并且

$$w(x) = e^{-\frac{x^2}{4}}$$

是高斯核。

令 ρ_m 为与 m维时间延迟嵌入相关的自然测度，并假设自相似性关系

$$T(h,m) \propto h^{d_C}. \tag{5.41}$$

高斯核关联和的离散形式为

$$T(h,m) = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j \neq i}^{N} e^{-||x_i - y_j||^2/4h^2}. \tag{5.42}$$

此处，所有观测值都必须归一化为均值为零、标准差为一。随后，按照Grassberger-Procaccia算法，在使用线性标度区间的基础上估计 d_C 。

5.3.6 信息维数

请注意，对于 q 的不同取值，广义维数 D_q 均存在。D_1 的情况通常称为**信息维数**，可由原始广义维数定义推导得出。

定义 5.9. – 信息维数

$$D_1 = \lim_{\epsilon \to 0} \frac{\langle \ln p_\epsilon \rangle_\mu}{\ln \epsilon}$$

信息维数得名于以下事实：$p_\epsilon \mu$ 是以准确性 ϵ 指定点 x 所需的平均香农信息量。信息维数描述了所需信息量如何随分辨率 ϵ 变化。

5.4 熵与信息论

动力系统的熵是另一种在拓扑共轭下保持不变的测度，并且有多种不同的形式。它与热力学中的概念有若干类比，也是**信息论**的一部分。这一理论自20世纪40年代以来不断发展，其主要贡献者包括Shannon、Renyi、Kolomogorov和Sinai。信息论试图以数学方式描述动力过程中的信息测量与流动，从而为时间序列分析提供一种重要方法。

我们首先考虑一个离散随机变量的静态分布X，其可能结果有 M 种。令 p_X 表示 $X = M$的概率。由此，我们定义信息和香农熵。

定义 5.10. – Shannon 信息与熵
随机过程的一次观测或结果 $X = i$ 所包含的信息为

$$I_i = -\log(p_X(i)).$$

其中， p_X 是随机变量 X 的概率质量函数。离散概率分布 p_X 的 Shannon 熵定义为该分布所包含信息的平均值

$$H(X) = -\sum_i p_X(i)\log(p_X(i)).$$

类似地，对于连续随机变量 X，

$$H(X) = -\int p_X(x)\log p_X(x)dx$$

与关联积分类似，我们可以定义熵的一般化 q 阶形式，称为 q 阶 Renyi 熵

定义 5.11. – R´enyi 熵
设观测值 x 定义在由边长为 ϵ 的互不相交的盒子组成的集合 P_j 上。令 $p_j = \int_{\mathcal{P}_j} d\mu(x)$ 表示该盒子中包含的测度所占的比例。分辨率为 ϵ 的 q 阶 R´enyi 熵定义为

$$\tilde{H}_q(\epsilon) = \sup_{\mathcal{P}_\epsilon} \frac{1}{1-q}\ln\left(\sum_j p_j^q\right)$$

在这种形式下，香农熵是 R´enyi 熵在$q = 1$时的一个特例。对于任意给定的概率分布，当 p_X 趋于均匀分布时，熵会增加。

另一个值得关注的熵是柯尔莫哥洛夫-辛奈熵。这个量试图衡量动力系统的可预测性以及信息损失速率。

定义 5.12. – Kolomogorov-Sinai 熵
设 P_ϵ 是某个观测量的相空间划分为 m 个划分I_1，I_2，..... I_m，其尺度为 ϵ。考虑联合概率$p_{i_1,\ i_2,\ i_m}$即在任意时间 n 该观测量落入划分I_{i_1}，在时间 $n+1$落入区间 I_{i_2} 等。考虑以下熵：

$$K_n = -\sum_{i_1,i_2,\ldots,i_m} p_{i_1,i_2,\ldots,i_n} \log p_{i_1,i_2,\ldots,i_n}.$$

柯尔莫哥洛夫-辛奈熵定义为

$$h_{KS} = \lim_{T\to 0} \lim_{\epsilon\to 0^+} \lim_{N\to\infty} \frac{1}{NT} \sum_{n=0}^{N-1} K_{n+1} - K_n$$

与香农熵描述概率分布的扩散程度类似，柯尔莫哥洛夫-辛奈熵在动力系统的上下文中发挥着类似的作用。直观地说，h_{KS}描述了某条轨迹向前演化时信息丢失的速率。具体做法是用离散马尔可夫链取代连续动力学。另一种理解h_{KS}的方式是：已知完整的历史轨迹后，在历史信息变得无关紧要之前，我们还能在多大程度上可靠地预测未来?

第6章
预测与函数逼近

6.1 使用动力系统建模

George E.P. Box 是一位统计学家，他在 1979 年的一份报告中创造了这句话：“所有模型都是错误的，但有些是有用的——如果现实世界中存在的任何系统都能由某个简单模型精确表示，那将是非常了不起的”[45]。一般而言，建模任务正处于应用数学家通常所承担工作范围的核心位置。很多时候，我们难以全面、精确地分析和描述现实世界中观察到的现象。为此，我们需要进行简化，使后续分析变得可行且易于处理。对应用数学家来说，这意味着要谨慎选择恰当的假设。建立模型后，便可将其用于各种任务，例如预测和特征描述。

一般而言，我们应致力于构建足以支持所需分析和结论的模型。这意味着要考虑数据的数量与质量，以及待执行任务的需求。不过，在动力系统的上下文中，考虑模型具备以下优良特性会比较恰当：

Q1. 模型的预测效果好吗？ - 如果需要，预测可以有多种形式。这些形式包括未来预测（接下来的 n 个时间步）、预测某个现象的起始、检测变点、分类等。每种情况都有各自独特的性能测度。例如：预测时域、F1召回率、AUROC、马修斯相关系数。

Q2. 模型是否再现了观测到的动力学？ - 良好的预测性能并不一定意味着模型准确描述了动力学。例如，生成混沌时间序列的短期预测并不困难。相比之下，长期预测很困难，因为模型中的任何微小误差都会导致预测误差随时间呈指数增长。如果我们接受这一自然限制，那么评估模型的一种替代方法是：模型至少应保持观测到的系统的整体动力学。能够生成不正确的长期预测、但

其行为与系统相似的模型（例如再现不变测度、产生相似的吸引子）比无法做到这一点的模型（例如偏离感兴趣的定义域）更具信息量

Q3. 模型是否简单且具有良好的泛化能力? - 不要将模型简单性与泛化能力混为一谈，这一点很重要。简单模型需要调整的参数可能较少，模型的结构也可能相对简单。这通常是一个有利特征，因为大量参数可能导致模型设定错误，进而影响模型性能。泛化能力是指模型在训练/拟合数据以外的任务上表现良好的能力。泛化能力有两种类型需要考虑：定义域内的泛化（插值）和定义域外的泛化（外推）。后者要实现得困难得多，但也更有用。总的来说，模型简单并不意味着具有良好的泛化能力（反之亦然），但确实有理由认为这两种性质往往存在相关性。神经网络就是一个例外，目前有研究致力于为此类情形下的模型性能建立泛化理论。

这些讨论的范围相当广泛，可应用于数学中任何涉及建模和预测的领域。动力系统方法，尤其是时间序列分析，通过假设任何给定系统通常都可表示为任意自治动力系统，对此进行了进一步细化：

$$\dot{x} = f(x), \quad x(t + \delta t) = \phi_{\delta t}(x(t)). \tag{6.1}$$

我们也可以考虑它的离散表示，

$$x_{n+1} = f(x_n). \tag{6.2}$$

在许多情况下，目标是近似或建模函数 f （或演化算子 ϕ）。如果能准确地做到这一点，就可以进行预测和系统表征。不过，重要的是要注意，这并非在所有情况下都可行。我们很可能不得不接受其他较差的结果。

6.2 自回归过程

时间序列预测最常见、最简单的方法之一，是根据观测数据拟合一个自回归模型。自回归模型的一般形式为 $AR(p)$，可定义如下：

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \epsilon_t. \tag{6.3}$$

还可以在模型中考虑更大且间隔不均匀的滞后项。对于随机时间序列，也可以纳入移动平均项和噪声项，

$$X_t = \epsilon_t + \sum_{i=1}^{p} \varphi_i x_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2). \tag{6.4}$$

一般来说，拟合自回归模型 $AR(p)$ 等价于根据 $p-1$ 维延迟嵌入拟合线性模型。选择 φ_i 时必须谨慎，以确保模型的迭代应用不会导致数值无限增长。为此，我们可以用后移算子 $BX_t = X_{t-1}$ 来表示自回归过程：

$$\Phi(B) = 1 - \sum_{j=1}^{p} \varphi_j B^j = 1 - \varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_p B^p \tag{6.5a}$$

$$\Phi(B) X_t = \epsilon_t \tag{6.5b}$$

.

定理 6.1. – 自回归过程的稳定性
自回归过程 $AR(p)$ </style}平稳，当且仅当特征多项式 $(B) = 0$ 的所有根 β_i 都大于一。

6.3 最近邻（类比法）

预测下一时间步的一种简单方法是采用类比法，也称为最近邻预测，该方法基于一个简单的天气预测模型。假设我们要预测明天（$n+1$）的天气。一种直接的方法是回顾历史，找出另一天（$m < n$），其天气与今天（n）最为相似。由此我们可以推断，明天的天气（$n+1$）将与第（$m+1$）天相同，误差在假定的范围内。

对于时间序列（嵌入）{$x(t)$}数据，最近邻预测算法的步骤如下：

1. 令当前状态为 $x(t)$。2. 找到最近邻 $z^{NN}(t) = x(k)$，其中 $t - k > l_{Th}$，且 l_{Th} 是 Theiler 窗口。3. 预测值为

$$x(t+1) = x(k+1)$$

显然，这种方法局限性很大，也存在许多潜在问题。例如，观测点可能没有足够近的近邻

。在尝试预测不变区域边界附近的状态时（例如分界轨附近），也很可能会遇到问题。我们可以采用更好的方法：对最近的 N 个邻居取局部平均：

1. 令当前状态为 $x(t)$。2. 在某个 Theiler 窗口内找到 N 最近邻 $z_i^{NN}(t) = x(k_i)$。3. 预测值为

$$x(t+1) = \frac{1}{N}\sum_{i=1} x(k_i + 1)$$

N 最近邻方法对噪声更具鲁棒性。然而，当尝试预测吸引子上密度较低的点，或基于稀疏数据进行预测时，该方法容易产生误差，因为最近邻可能距离很远。对此，可以根据距离的某个上界 ϵ，选择参考点附近的邻居（即找到所有 $||z_i^{NN}(t)x^{NN}(t)|| < \epsilon$）。此外，还可以使用核函数，根据每个邻居到参考点的距离 $x(t)$对其预测贡献进行加权。例如，我们可以使用大小参数为 S的指数核函数，

$$w(d) = e^{-Sd/\bar{d}}, \tag{6.6}$$

¯其中 $d = ||z(t) - x(t)||$， d 是这些距离的平均值。因此，我们可以将预测写成

$$x(t+1) = \frac{\sum x^{NN}(k_i + 1)w(d_i)}{\sum w(d_i)} \tag{6.7}$$

6.4 神经网络（反正谁在乎参数的数量？）

6.4.1 感知器

神经网络是一大类日益受到关注的数学模型。然而，它们并非新生事物，在文献中已有 70 多年的历史，其灵感源自人们对生物神经元如何通信的早期认识。神经元序列彼此传递消息，信号逐渐累积，超过阈值后便会“触发”。这也引出了赫布描述：“共同触发的神经元会连接在一起”。我们首先探讨通用逼近定理 [46]，以此开始讨论神经网络。

定理 6.2. – 通用逼近定理
令 $C(X, \mathbb{R}^m)$ 表示从子集 $X \subset \mathbb{R}^n$ 到 的连续函数集合。
令 $\sigma \in C(\mathbb{R}, \mathbb{R})$。请注意，$(\sigma \circ x)_i = \sigma(x_i)$ 是逐元素应用的。
则有以下结论：

σ 是非多项式的，当且仅当对于每个 n、$m \in \mathbb{N}$、紧集 $K \subset \mathbb{R}^n$，$f \in C(K, \mathbb{R}^m)$、$\epsilon > 0$，存在 $k \in \mathbb{N}$、$A \in \mathbb{R}^{k \times n}$、$b \in \mathbb{R}^k$、$C \in \mathbb{R}^{m \times k}$ 使得

$$\sup_{x \in K} ||f(x) - g(x)|| < \epsilon,$$

其中 $g(x) = C \cdot (\sigma \circ (A \cdot x + b))$

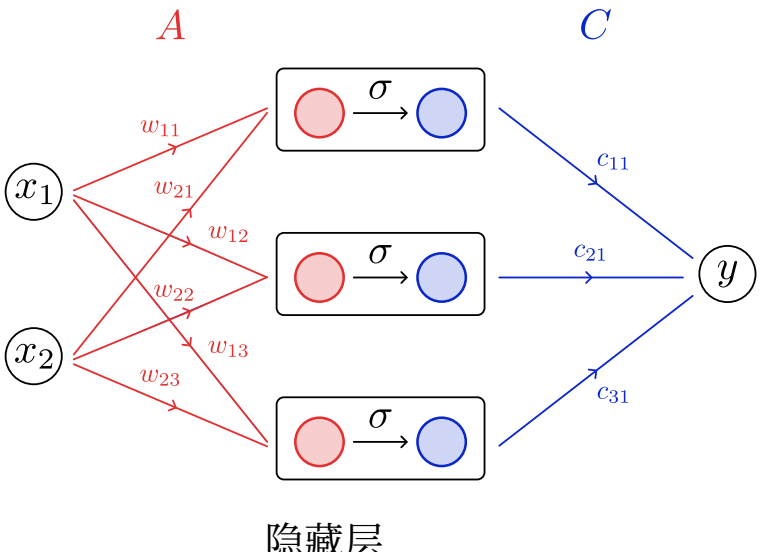


图 6.1 带有一个隐藏层的前馈神经网络（感知器）。

考虑基本神经网络（又称感知器）的模型结构时，通用逼近定理具有重要意义。本质上，该定理指出，只要 σ 是非多项式的非线性函数（例如指数函数、对数函数、S形函数），那么只要隐藏层足够宽，单隐藏层神经网络就能逼近任意连续函数。这种极强的灵活性意味着，人们可以轻松运用神经网络架构来训练用于时间序列预测的演化算子，而这些算子往往复杂且无法通过解析方法求得：

$$\begin{aligned} x(t+1) &= f(x(t)) \\ &= \phi^{NN}(x(t)) \\ &= C \cdot (\sigma \circ (A \cdot x + b)) \end{aligned} \tag{6.8}$$

其中， C 和 A 分别是感知器的输出向量和输入向量， b 是偏置向量。非线性函数 σ （如双曲正切函数、指数线性单元、整流线性单元等）有多种选择。不同的选择会影响模型的性能。

6.4.2 梯度下降算法

拟合感知器模型需要选择（训练） C, A 和 b 中的常数元素（权重），使函数（预测）误差最小化，该误差也称为损失 L。更一般地，损失可表示为

$$\mathcal{L}(x) = d(f(x), \phi^{NN}(x)), \tag{6.9}$$

其中， d 是距离度量/核。具体过程如下，使用梯度下降算法:

1. 设初始模型权重集合为 $\lambda_0 = \{\alpha_0,1,\alpha_0,2,...\}$，使用这些权重对神经网络进行评估的结果为 $\phi^{NN}_{\lambda_0}$。2. 定义输入数据 $\{x_n\}$ 和输出数据 $\{y_n\}$，我们希望在它们之间学习一个映射。3. 将 $\{x_n\}$ 输入网络，计算预测输出数据:

$$\hat{y}_n = \phi^{NN}_{\lambda_0}(x_n)$$

4. 计算给定的损失

$$\mathcal{L}(x_n) = d(y_n, \hat{y}_n)$$

5. 计算损失相对于模型权重的梯度向量

$$\vec{\nabla} = \frac{\partial \mathcal{L}(x_n)}{\partial \lambda_0} = \left(\frac{\partial \mathcal{L}(x_n)}{\partial \alpha_{0,1}}, \frac{\partial \mathcal{L}(x_n)}{\partial \alpha_{0,2}}, \ldots \right)$$

6. 更新模型权重以降低损失

$$\lambda_{n+1} = \lambda_n - \beta \cdot \frac{\vec{\nabla}}{||\vec{\nabla}||},$$

其中， $\beta > 0$ 是一个较小的值，称为学习率。

梯度下降算法无法保证找到全局最优解，而且往往会停留在局部最优解上。此外，由于优化景观的 Complexity 和维数，通常不可能找到全局极小值。不过，某些局部极小值仍可能足以用于预测。我们可以通过几项改进来优化梯度下降算法，并部分解决这些缺陷：

- **批量梯度下降 -**在计算梯度并更新权重之前，先计算所有输入数据的平均损失。这样得到的损失曲面更加平滑，也更容易沿着曲面优化。
- **随机梯度下降 -** 它通过随机选择输入点，而不是对整个批次求平均，来解决与批量梯度下降相同的问题，从而最大限度地减少连续权重更新之间输入数据时间相关性的影响。
- **小批量梯度下降 -** 将批量梯度下降和随机梯度下降结合起来，在每次更新时随机选取较小的输入子样本进行平滑处理。
- **动量 -** 在计算 ∇时纳入某种动量概念

$$\vec{\nabla}_{n+1} = \gamma\vec{\nabla}_n + (1-\gamma)\vec{\nabla}^*,$$

其中，$\vec{\nabla}^*$ 是最新的梯度估计值。这种实现方式可以避免梯度方向突然变化，并使算法越过局部最优解，从而有助于更充分地探索损失景观。

- 其他优化算法（例如 ADAM、AdaGrad、AdaDelta）

6.4.3 神经网络变体

前馈神经网络（FNN）- 这是对经典感知器的扩展，包含多个隐藏层。

循环神经网络（RNN）和长短期记忆网络（LSTM）-这种架构专为序列到序列预测而设计。通常只有一个隐藏层，但也可以有更多。每个隐藏层节点都具有一个状态值，该状态值根据方程演变

$$\vec{s}(t+1) = \sigma \circ (C_{in}\vec{x}(t) + C_{rec}\vec{s}(t) + \vec{b}), \tag{6.10}$$

而下一步预测则表示为

$$\vec{x}(t+1) = C_{out}\vec{s}(t+1). \tag{6.11}$$

这种结构在模型中内置了复现机制，使网络本质上能够成为一种通用的自回归过程。然而，训练循环神经网络需要使用经过修改的梯度下降算法，以处理误差随时间传播的问题。这通过时间反向传播（BPTT）算法实现

该算法试图将具有 T 时间步复现的循环网络展开为一个以前馈网络形式、包含 T 隐藏层的网络。然而，由于连续乘以非常大或非常小的数，这种算法会受到梯度爆炸与梯度消失问题的影响。长短期记忆（LSTM）网络是循环神经网络的一种改进，通过在节点中编码记忆缓冲区来缓解这一问题 [47]。

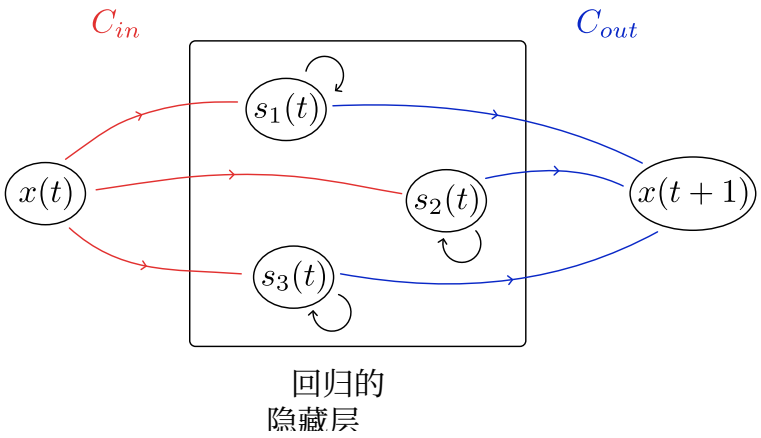


图 6.2 含有三个隐藏节点的单隐藏层循环神经网络结构。

径向基网络 - 其形式与感知器相似，但使用了不同的非线性函数。具体而言，这类模型使用径向基函数

$$\phi_j^{RBN}(x(t)) = C \cdot (\sigma \circ (\frac{||x(t) - c_j||}{r_j})) \tag{6.12}$$

储备池计算机 - 这是一种循环神经网络的变体，它避免直接优化隐藏节点权重，转而只训练读出权重，从而将训练算法简化为一个正则化最小二乘回归问题。它更适用于时间序列分析：将时间序列序列输入循环神经网络，并允许其在隐藏层中回响，以生成基函数，随后用于未来预测。它本质上相当于内置的无限维延迟嵌入。储备池计算机本质上相当于无限维延迟嵌入（见第 9 章）

6.5 符号回归（即让代数学家/泛函分析学家使用神经网络）

神经网络是通用函数逼近器。这是通过用预先定义的非线性函数嵌套集合来逼近所需函数实现的

结构。函数复现的准确性取决于常数的优化。更一般地，任何前馈神经网络本质上都是一个具有如下数学形式的函数

$$\phi^{NN}(x) = \sigma(A...\sigma(\sigma(Ax + b) + b)... + b) \tag{6.13}$$

尽管这种形式对数值分析有用，但它存在若干问题

- **可解释性 -** 这种函数表示形式无法直观地描述其底层行为。可以想想用泰勒展开替代指数函数的情形。
- **复杂度高 -** 需要训练大量模型参数，导致模型复杂度非常高。其中大多数参数并未被使用，而且模型也不够简约。

神经网络的一种替代方案是符号回归。这种方法试图在不预先设定数学形式的情况下，以解析方式重构原始函数 f，目的是希望得到数学上更具可解释性且更简约的解。非线性动力学稀疏识别（SINDy）算法是目前旨在直接从数据中实现这一目标的一种算法 [48]。该算法如下：

1. 设有一个 n 维多元时间序列，由 $\mathbf{x}(t) = (x_1(t), x_2(t), ..., x_n(t))$ 表示，并在 $m + 1$ 个时间步上进行测量。构造两个矩阵 $\mathbf{X}, \dot{\mathbf{X}} \in \mathbb{R}^{m \times n}$，其元素分别包含观测状态和数值导数。

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}^T(t_1) \\ \mathbf{x}^T(t_2) \\ \vdots \\ \mathbf{x}^T(t_m) \end{bmatrix} = \begin{bmatrix} x_1(t_1) & x_2(t_1) & \dots & x_n(t_1) \\ x_1(t_2) & x_2(t_2) & \dots & x_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(t_m) & x_2(t_m) & \dots & x_n(t_m) \end{bmatrix} \tag{6.14a}$$

$$\dot{\mathbf{X}} = \begin{bmatrix} \dot{\mathbf{x}}^T(t_1) \\ \dot{\mathbf{x}}^T(t_2) \\ \vdots \\ \dot{\mathbf{x}}^T(t_m) \end{bmatrix} = \begin{bmatrix} \dot{x}_1(t_1) & \dot{x}_2(t_1) & \dots & \dot{x}_n(t_1) \\ \dot{x}_1(t_2) & \dot{x}_2(t_2) & \dots & \dot{x}_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \dot{x}_1(t_m) & \dot{x}_2(t_m) & \dots & \dot{x}_n(t_m) \end{bmatrix} \tag{6.14b}$$

2. 设 $F = \{f_1, f_2, ..., f_K\}$ 是候选非线性函数库（例如 $1, x, x^2, \sin x$ ...）。构造矩阵 $\Theta(\mathbf{X}) \in \mathbb{R}^{m \times nK}$，其元素由基函数 F 在$\mathbf{X}$取值处求值得到。3. 设 $\Xi = (\xi_1, \xi_2, ..., \xi_n)$ 是一组系数稀疏向量，用于确定哪些非线性函数处于激活状态：

$$\dot{\mathbf{X}} = \Theta(\mathbf{X})\Xi$$

对上述方程的每一列进行稀疏优化，以确定所需的系数。结果是一个近似的解析表达式，用于描述向量场方程。

SINDy 的一种较新替代方法是方程学习网络（EQL）[49]，它对基本的前馈神经网络进行了修改，使其由基本运算组成，并尝试 f 从头开始构造所需函数。不过，总体而言，符号回归领域正在迅速发展，感兴趣的读者可以进一步研究这一领域。

对于感兴趣的读者而言，近期发展起来的物理信息机器学习方法可用于求解偏微分方程。该方法通常是指将结构化物理模型与神经网络相结合，以求解预先定义的偏微分方程问题。物理信息神经网络（PINN）算法[50]由两部分架构组成：用于表示未知函数的前馈神经网络 u，以及通过数值方式进行评估的偏导数偏微分方程组件。这两部分结合使用，用于计算损失，随后可通过反向传播更新神经网络权重。

第7章 模型选择

与时间序列分析中的其他任务相比，模型相对容易构建和优化。然而，模型选择并不那么直接，因而构成了更大的挑战。为说明这一点，可以考虑一个用于优化传输给定长度文本消息的模型 N。我们可以考虑两种发送消息的方法：

1. 直接逐字编码消息并发送。这种方法精度和准确性都很高，但计算成本高昂，且难以扩展。2. 对消息的一般语义（即简单模型）进行编码，并附上足够的细节（即误差分布），以便接收方重构消息。这种方法精度较低，容易出错，但发送成本低得多，也更简约。

上述示例说明了模型选择的一个关键方面：在 Complexity 与准确性之间取得平衡。高 Complexity 的模型（尽管直接传递消息几乎称不上模型）功能强大且灵活，但也可能容易出错、难以解释、计算成本高昂，而且泛化能力可能较差。另一方面，低 Complexity 的模型可能具有更好的泛化能力（得益于其简单性），校准和运行成本也较低，但准确性或灵活性可能不足。那么，该如何选择模型呢?

模型选择可以借助若干信息准则来完成。这些准则类似于代价函数，用于平衡模型复杂度与准确性的权重。这样一来，模型选择任务就转化为一个优化问题，即需要将信息准则得分最小化。在此，我们考察三种选择严格程度逐渐提高的信息准则，分别是赤池信息准则（AIC）、贝叶斯（施瓦茨）信息准则（BIC）和模型描述长度（MDL）。

7.1 赤池信息准则（AIC）

假设我们有一组观测数据点，它们由底层真实概率分布 $g(y)$ 生成。在建模问题中，我们希望构建一个参数模型 $f(y|\theta)$，其模型参数为 θ，以逼近 $g(y)$。具体而言，我们假设已经选定了模型类，其中每个模型具有 k 个参数，其分布形式如下

$$\mathcal{F}(k) = \{f(y|\theta_k, \theta_k \in \Theta(k)\}, \tag{7.1}$$

其中， $\Theta(k)$ 是所有可能的 k 维参数向量的集合。

直观地说，好的模型 $f(y|\theta_k)$ 应与真实分布更加接近。因此，我们希望评估模型的拟合优度。我们的目标是在所有拟合模型中进行搜索，找出对 $g(y)$ 近似效果最好的模型。为此，我们首先需要定义一种量化相似性的方法。

散度 是一种常用于描述两个分布之间相似性的函数。这里的“散度”一词不要与其另一种定义 $\nabla \cdot \mathbf{F}$ 混淆。散度与度量非常相似，但条件更宽松。最值得注意的是，散度不必具有对称性（即 $\mathrm{D}(f, g) = \mathrm{D}(g, f)$）。

就我们的目的而言，我们考虑一种源自信息论的散度——Kullback-Leibler（KL）散度

定理 7.1. – 库尔贝克-莱布勒散度

$$\mathcal{D}_{KL}(g||f) = \mathbb{E}\left[\log\left(\frac{g(y)}{f(y|\theta_k)}\right)\right]$$

此外，

$$\mathcal{D}_{KL}(g||f) = 0 \iff f = g$$

KL散度 $\mathrm{D}_{KL}(g||f)$ 衡量使用针对编码 f 优化的码来编码 $g(y)$ 所需的信息量。尽管 D_{KL} 并非形式上的度量，但其值会随着 f 与 g 之间差异的增大而增大。因此，$\mathrm{D}_{KL}(g||f)$ 衡量模型分布 f 对真实分布 g 的近似程度。接下来，我们定义

$$d(\theta_k) = \mathbb{E}[-2\log f(y|\theta_k)]. \tag{7.2}$$

由此可得

$$
\begin{aligned}
2\mathcal{D}_{KL}(g||f) &= \mathbb{E}\left[-2\log\left(\frac{g(y)}{f(y|\theta_k)}\right)\right] \\
&= \mathbb{E}[-2\log g(y) + 2\log f(y|\theta_k)] \\
&= d(\theta_k) - \mathbb{E}[-2\log g(y)].
\end{aligned} \tag{7.3}
$$

由于我们无法控制 $g(y)$，因此我们旨在针对 θ_k 进行所有优化。根据 D_{KL} 对模型进行排名，等同于仅根据 $d(\theta_k)$进行排名。因此，我们应该可以通过评估以下值，按照参数为 $\hat{\theta}_k$的模型进行排名：

$$
d(\theta_k) = \mathbb{E}[-2\log f(y|\hat{\theta}_k)]. \tag{7.4}
$$

遗憾的是，无法直接估计 $d(\hat{\theta}_k/0$，因为它取决于观测值 y，而观测值又取决于原始的 $g(y)$。(赤池，1974) 提出了一种有偏估计量，方法是使用项 $-2\log f(y|\hat{\theta}_k)$，其中 y 是迄今为止的观测数据。在这种表述中，$-2\log f(y|\hat{\theta}_k)$ 是一个负偏估计量。可以通过使用偏差调整项 $\delta > 0$ 来对此进行修正，其定义为

$$
\begin{aligned}
\delta &= \mathbb{E}[2d(\hat{\theta}_k)] - \mathbb{E}[-2\log f(y|\hat{\theta}_k] \\
&\approx k.
\end{aligned} \tag{7.5}
$$

将上述分量合并，可得赤池信息准则（AIC）的表达式

定义 7.1。– 赤池信息准则（AIC）

$$
\begin{aligned}
\mathrm{AIC} &= -2\log f(y|\hat{\theta}_k) + 2k \\
&= 2k - 2\ln\hat{\mathcal{L}}
\end{aligned}
$$

其中，$\hat{L} = \max(\mathrm{L}(\hat{\theta}_k|y))$ 是似然函数的最大化值。详细推导见 [51]。

简单来说，AIC 分数会根据模型所需参数的数量，对模型的拟合优度进行惩罚。$2k$ 项可确保模型不会因参数过多而过拟合。

7.2 贝叶斯信息准则（BIC）

BIC 是施瓦茨提出的一种贝叶斯方法，用于解决模型选择问题 [52]。同样，假设存在一个真实的底层模型分布 $g(y)$，我们的目标是从一组参数类中选择一个模型 M_k

分布。换言之，给定一个模型类 M_K，存在一个 k 维的参数化集合，其中包含概率密度函数

$$\mathcal{F}(k) = \{f(y|\theta_k|\theta_k \in \Theta(k)\}, \tag{7.6}$$

一旦我们选定模型类 M_k，就可以根据观测数据 y 计算似然函数 $\mathrm{L}(\theta_k|y)$。令 $\hat{}\ \theta_k$ 表示通过对 $\Theta(k)$ 中的 M_k 进行最大化而得到的参数估计值向量。

假设我们可以构建一组模型 M_{k_1}，M_{k_2}、……、M_{k_L}（例如形式、统计量、协方差结构等各不相同的模型）。我们要问：哪个 M_k 最适合近似 $g(y)$?

与贝叶斯方法中常见的做法一样，我们首先从陈述贝叶斯定理开始。

定义 7.2. – 贝叶斯定理
设 A 和 B 是两个随机事件，$P(A)$ 和 $P(B)$ 分别是各事件发生的概率。那么，

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$P(A)$称为 A的先验分布，而 $P(B)$是 B的边际分布。

设模型类 $M_{k_1}, M_{k_2}, \ldots, M_{k_L}$的离散先验分布为 $\pi(k)$，$g(\theta_k|k)$表示给定模型 M_k时 θ_k 的先验分布。应用贝叶斯定理，可得:

$$h((k,\theta_k)|y) = \frac{\mathcal{L}(\theta_k(y)g(\theta_k|k)\pi(k))}{m(y)} \tag{7.7}$$

其中，$m(y)$是观测值 y的边际分布。模型选择的贝叶斯表述试图使后验似然 $h((k,\theta_k)|y)$相对于 k最大化。我们也可以将其写成后验概率，

$$P(k|y) = \frac{1}{m(y)}\pi(k)\int_{\Theta_k}\mathcal{L}[\theta_k|y]g(\theta_k|k)d\theta_k \tag{7.8}$$

或者，这等价于最小化对数后验，

$$-2\ln P(k|y) = 2\ln m(y) - 2\ln\pi(k) - 2\ln\left[\int \mathcal{L}[\theta_k|y]g(\theta_k|k)d\theta_k\right] \tag{7.9}$$

由于我们只关心模型选择，因此可以舍弃第一项，将上述表达式改写得更简单。

$$-2\ln P(k|y) = -2\ln\pi(k) - 2\ln\left[\int \mathcal{L}[\theta_k|y]g(\theta_k|k)d\theta_k\right] \tag{7.10}$$

还可以进一步简化该表达式并舍弃一些项，从而得到以下简化后的 BIC 表达式。

定义 7.3——贝叶斯（施瓦茨）信息准则（BIC）

$$\begin{aligned}\mathrm{BIC} &= k\ln(n) - 2\ln\mathcal{L}[\theta_k|y] \\ &= k\ln(n) - 2\hat{\mathcal{L}}\end{aligned}$$

其中，$\hat{L} = \max(\mathrm{L}(\hat{\theta}_k|y))$ 是似然函数的最大值。详细推导见 [53]。

注意，BIC 与 AIC 之间存在相似之处。这两个评分仅在惩罚项上有所不同。BIC 根据观测值的数量对参数数量施加惩罚。总体而言，BIC 对模型选择施加了更严格的考量。在贝叶斯框架下，人们预期，随着观测数据的增加，应当能够构建出更好的模型。因此，BIC 在进行模型选择时，将观测数据的质量和数量作为正则化项。

7.3 模型选择与编码理论

7.3.1 重新审视消息传递问题

假设我们有一个时间序列 $x_t, t = 1, 2, 3, \ldots\ldots N(x = (x_1, \ldots\ldots, x_N))$。根据 Takens 定理，当 d 足够大时，我们可以通过时间延迟嵌入重构底层动力学，

$$v_t = (x_t, x_{t-1}, \ldots x_{t-d+1}) \tag{7.11}$$

其中 $v_{t+1} = \phi(v_t)$ 的动态演化等价于底层动力系统的演化算子。我们希望估计 ϕ。我们可以通过如下形式的模型 $f\colon \mathbb{R}^d \to \mathbb{R}$ 来实现

$$x_{t+1} \approx f(x_t, x_{t-1}, \ldots x_{t-d+1}; \lambda) \tag{7.12}$$

其中 $\lambda \in \mathbb{R}^k$ 对可能的模型集合进行参数化。但我们该如何选择最佳模型？例如，如果将模型限制在某个模型类中，比如径向基模型，那么仍然需要确定应包含多少个中心，以及精度应达到多少。我们需要一个客观原则来从多个备选模型中进行选择。我们将运用信息论的思想来探讨这个问题。

考虑一个消息传递问题。我们希望在双方之间传输一个数据流。例如，数据流 1，1，2，3，5，8，13，21，34，55，……。实现这一目标有两种方式：

1. 直接逐字发送原始数据 2. 发送数据的模型、初始条件以及模型的预测误差，使接收方能够重构数据流。

上例中的数据流可以识别为斐波那契数列的前几项。要重现该数据流，我们可以使用一个**完美模型**，即 x_1和 x_2，并发送规则 $x_t = x_{t-1} + x_{t-2}$，$t \geq 3$。我们还应发送项数，以便接收方知道何时停止。

有时可能无法获得**完美模型** ——模型未知、转录模型时出错，等等——但发送一个带有误差的**不完美模型**，而不是直接发送原始数据流，仍可能有用。例如，我们不发送斐波那契数列的完美模型，而是发送规则$x_t = 2x_{t-1}$，$t \geq 2$，其中 $x_1 = 1$，并同时发送误差 e_t，这样数据流就是 $\hat{x}_t = x_t + e_t$。如果模型“足够好”（模型简洁且误差较小），那么发送模型加误差的成本仍可能低于直接发送原始数据。若限制对接收方的预测范围（例如，限于前100 项），这一点尤其成立。

我们将最佳 定义为：在某个有限的指定精度下，能够对数据进行最紧凑描述的模型。也就是说，如果要通过容量有限的通信信道传输原始数据，哪种模型能够最紧凑地描述这些数据？因此，我们有两种选择：

1. **完美模型 -** 模型能够精确捕捉动力学。求解 λ，然后发送λ 和初始条件 x_0
2. **不完美模型 -** 估计 λ，并优化权衡

   - 模型效果不好：传输数据 x更快（成本更低）
   - 模型效果好：发送模型参数 λ，再加上 x_0 和模型预测误差，效果更好。

因此，我们应致力于构建满足以下条件的模型

$$L(model) + L(errors) < L(data), \tag{7.13}$$

其中 L 是发送数字的成本。如何量化 L? 为此，我们借鉴**编码理论**中的思想。

7.3.2 编码理论

令 A 表示由符号组成的**有限字母表**。在**数字**信道上通信，需要用一组独立的符号 B 来表示该字母表。令 B 表示信道上允许使用的码符号集合，并用 B^* 表示该字母表上的词集。由 B 中的 l 个符号组成的一个词（$b_1, b_2, \ldots, b_\ell$）是

B 中元素的一个有序序列。 B^* 包含所有此类允许的序列。因此，我们有一个映射 L： $A \to B^*$，其中 $a \mapsto (b_1, b_2, \ldots, b_\ell)$ 满足 $b_i \in B$。

定义码长 L： $A \to \mathbb{Z}^+$ 使得 $a \mapsto l$，即 $L(a) = l$ 是用字母表 B对 A 进行编码的长度。理想情况下，我们希望码既无歧义又高效。

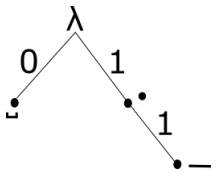
例 *7.1.* – **摩尔斯电码**
在摩尔斯电码中，$|B| = 3$（点、划和空格）。摩尔斯电码没有歧义——空格分隔由点和划组成的码字，并且 B^* 与 A之间存在一一对应关系。摩尔斯电码也很高效——在 A 中出现频率较高的符号，在 B^*中对应的码字较短。也就是说，如果 $L(a_1) < L(a_2)$，那么 a_1 比 a_2更常见。考虑摩尔斯电码的一种二进制编码，即令 $B = \{0, 1\}$，且 $A =$ {点,划, 空格}。考虑表 7.1 中四种可能的编码。

空格 点 划 （） （•） （–） 1. 0 1 11 有歧义 2. 00 01 11 无歧义（但不完整）3. 0 10 11 无歧义且效率更高（完整）4. 00 1 01 无歧义且效率更高（按频率而言）

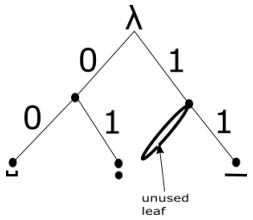
表 7.1 摩尔斯电码的编码

7.3.3 编码树

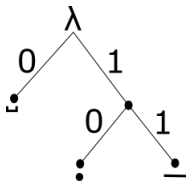
编码树是判断一种编码是否存在歧义的有用工具。对于无歧义编码，所用码字是树的“叶子”。也就是说，任何码字都不是其他码字的前缀子集。这也称为前缀无歧义编码。完备码没有未使用的叶子。考虑表 7.1 中给出的摩尔斯电码符号{点,划, 空格}的各种二进制编码 $= A$，其中 $B =$ {0,1}。

1. 

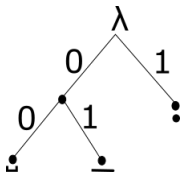
空格 点 划
0 1 11
这种编码的问题在于存在歧义。如果传输消息 11，就无法确定它指的是“● ●”还是“–”的编码。

2. 

空格 点 划
00 01 11
这种编码没有歧义，但不完整，因为码 10 对应的叶节点未被使用。

3. 

空格 点 划
0 10 11
这种编码没有歧义且完整（没有未使用的叶节点），但并非完全高效，因为它没有考虑每个符号的相对出现频率。理想情况下，出现频率较高的符号应使用更短的编码。

4. 

空格 点 短划线
00 1 01
这种编码无歧义、完备，而且效率也更高。

7.3.4 最优编码

我们简要使用过这个术语，但什么是**高效编码**呢？我们的目标是构造一种码，使出现频率更高（概率更大）的符号使用较短的码字。码字长度应根据预期出现频率来确定，从而使预期消息长度最小。设

$$p(a) = \text{frequency/probability of encountering symbol } a \in A. \tag{7.14}$$

现在可以用平均（期望）码字长度来定义编码的**效率**

$$\text{Efficiency} = \sum_{a \in A} p(a)L(a). \tag{7.15}$$

我们有如下结果，称为**克拉夫特不等式**。

推论 7.1. – 克拉夫特不等式
对于前缀无关码，$\sum_{a \in A} 2^{-\ L(a)} \leq 1$，当 $|B| = 2$ 时。此外，如果该码是完备的，则 $\sum_{a \in A} 2^{-L(a)} = 1$。这个结果可以利用树形图证明，留作读者练习。

假设 $A = \{a_1, a_2, \ldots, a_n\}$，且 p_i 是 a_i ($\sum p_i = 1$) 的频率。定义 $l_i = \lceil -\log_2 p_i \rceil$（其中 $\lceil x \rceil$ 是不小于 x 的最小整数）。那么，

$$\begin{aligned} \sum_i 2^{-\ell_i} &= \sum_i 2^{-\lceil -\log_2 p_i \rceil} \\ &\leq \sum_i 2^{\log_2 p_i} \\ &= \sum_i p_i = 1 \end{aligned} \tag{7.16}$$

下面的结果使我们可以考虑最优编码的定义。

定理 7.2。 – 最优编码
设 $p = (p_1, p_2, \ldots, p_n)$ 和 $q = (q_1, \ldots q_n)$，其中 $0 \leq p_i \leq 1, 0 \leq q_i \leq 1$ 和 $\sum_{i=1}^n p_i = \sum_{i=1}^n q_i = 1$。那么

$$-\sum_{i=1}^{n} p_i \log p_i \leq -\sum_{i=1}^{n} p_i \log q_i$$

等号成立当且仅当 $p_i = q_i, \forall i$。

量 $H(p) = -\sum_{i=1}^n p_i \log_2 p_i$ 称为信息源的**熵**。熵 $H(p)$ 是平均码字长度的下界：

$$\sum_{a \in A} p(a)L(a) \geq H(p) \tag{7.17}$$

并且，如果 p_i 都是 2 的幂，那么 $\lceil -\log_2 p_i \rceil$ 都是整数，我们可以达到 $H(p)$。因此，给定字母表 A 和频率 $p = (p_1, p_2, \ldots)$，则存在一个基于二进制字母表 $B = \{0, 1\}$ 的前缀无关码，使得

$$\sum_{a \in A} p(a)L(a) - H(p) \leq 1 \tag{7.18}$$

也就是说，$H(p) \leq \sum_{a \in A} p(a)L(a) \leq H(p) + 1$。

回顾一下克拉夫特不等式：如果一种码可唯一解码，那么其码长必须满足

$$\sum_i 2^{-l_i} \leq 1. \tag{7.19}$$

因此，对于任何满足克拉夫特不等式的码长，都存在一种具有这些码长的前缀无关码。

7.3.4.1 系综的码长

系综的最优源码长度等于香农信息量

$$l_i = \log_2 \frac{1}{p_i}, \tag{7.20}$$

反过来，任意选择一组码长都会定义隐含的概率，

$$q_i = \frac{2^{-l_i}}{Z} \tag{7.21}$$

其中，如果该码是完备的，则 Z 归一化为 1。

7.3.4.2 符号码的信源编码定理

对于系综 X，存在一种前缀无关码，其期望长度满足

$$H(X) \leq L(C, X) < H(X) + 1, \tag{7.22}$$

其中

$$H(X) = \sum_i p_i \log \frac{1}{p_i} \text{ – entropy} \tag{7.23}$$

and

$$L(C, X) = \sum_i p_i l_i \text{ – expected length of a code} \tag{7.24}$$

推论 7.2。 – 数据码长的上界
设数据为 $D = \{\alpha_1, \alpha_2, \ldots \alpha_N\}$，其中 $\alpha_i \in A$，并假设符号 D 出现的概率为 p。令 $L(D)$ 表示发送这些数据所需的比特总数。那么，

$$\begin{aligned} L(D) &= \sum_{i=1}^{N} L(\alpha_i) \\ &= \sum_{i=1}^{N} \lceil -\log_2 p(\alpha_i) \rceil \\ &\approx N \sum_{a \in A} p(a) \lceil -\log_2 p(a) \rceil \\ &\leq N(H(p) + 1) \end{aligned} \tag{7.25}$$

7.3.5 霍夫曼编码

那么，我们如何构造最优码呢？**霍夫曼编码**可以保证实现最优编码。假设我们用 (w, p) 表示符号 $a \in A$ 对应的（码字，频率）对。接下来的算法可以得到最优码，也就是说，平均码字长度最小。

1. 找出概率（频率）最小的两对 (w, p) 和 (w', p')——若有并列，任意选择。
2. 将这两对替换为一对 $(\hat{w}, \hat{p})$，其中 $\hat{p} = p + p'$, $w = \hat{w}_0$ and$w' = \hat{w}_1$。
3. 重复以上步骤，直到只剩下一对 $(w, p) = (\lambda, 1)$。 4. 回代。

码字不会被唯一确定——我们会任意打破平局——但它们的长度是确定的。

例 *7.2.* – **霍夫曼编码**
为以下码字/频率配对集合计算霍夫曼编码：$\{(a, 0.1), (b, 0.2), (c, 0.4), (d, 0.3)\}$。

符号	频率/概率
a	0.1
b	0.2
c	0.4
d	0.3
A	0.3
B	0.6

表 7.2 霍夫曼编码（非唯一）

例 *7.3.* – **霍夫曼编码 2**
为字符串“万维网”计算霍夫曼编码 ␣ ␣

包括空格字符在内，共有 14 个符号。表 7.3 列出了这些符号及其频率、熵、霍夫曼码和码长。图 7.1 展示了一种特定的编码树。

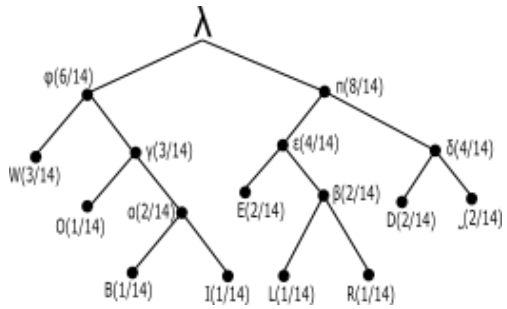


图 7.1 WWW 霍夫曼编码的编码树。

霍夫曼编码的缺点

- 取决于系综（“文本”）
- 忽略上下文（q 后面跟着字母 u）
- 可能产生额外比特开销

上述方案提供了一种对给定的（单个）数字进行最优编码的方法——该数字来自 $\mathbb{N}$、 $\mathbb{Z}$或 $\mathbb{Q}$；如何将该方案准确地扩展到这些情形中的每一种，留作读者练习。尽管霍夫曼码是最优符号码，但在实际应用中，符号码未必总是理想的。这是因为这种编码没有包含指示码序列结束的方法。我们可以引入一个新符号（“逗号”）来分隔码字（类似于摩尔斯电码中的“空格”）；或者设计一种自定界编码。也就是说，我们构建一种方法来指示一个码字何时

符号 a_i	p_i	h_i 码	$C(a_i)$	l_i
W	3/14 2.2	00		2
O	1/14 3.8	010		3
R	1/14 3.8	1011		4
L	1/14 3.8	1010		4
D	2/14 2.8	110		3
␣	2/14 2.8	111		3
I	1/14 3.8	0111		4
E	2/14 2.8	100		3
B	1/14 3.8	0110		4
α	2/14			
β	2/14			
γ	3/14			
δ	4/14			
ε	4/14			
ϕ	6/14			
π	8/14			

表 7.3 万维网的霍夫曼编码。

结束、另一个码字何时开始。本质上，我们先传达整数l_b（n）的长度，然后再传达原始整数。我们还必须传达该整数长度的长度，依此类推。

7.3.6 自定界解码器

考虑一个二进制位序列。

1. 如果第一个数字是 0，则编码整数为 1。停止。2. 否则，将最开始的三个数字（其中第一个为 1）视为当前代码块。如果第三个数字为 0，则前两个数字编码所需的整数（即 $10 = 2$、 $11 = 3$），并停止。3. 否则，前两个数字编码待读取的下一个代码块的长度——如果为 10，则令 $w = 3$；如果为 11，则令 $w = 4$。4. 丢弃当前代码块，并读取接下来的 w 比特作为下一个（新的）代码块。5. 如果第 $w + 1$ 个比特（即当前代码块之后的下一个比特）为 0，则当前的 w 比特代码块以标准二进制编码所需的数字——停止。6. 否则，当前的 w 比特代码块编码待读取的下一个代码块的长度 u。丢弃当前代码块及前 w 比特，读取接下来的 u 比特作为（新的）当前代码块，令 $w = u$，然后返回步骤 5。

例 *7.4.* – **整数自定界码**
某些整数的二进制自定界码示例如表 7.4 所示。

整数	码字
1	0
2	10
0	3
11	0
4	10
100	0
7	10
111	0
14	11
11100	15
11	11110
16	10
101	100000

表 7.4 整数二进制自定界码的实现。

例 *7.5.* – **整数自定界码 2**
使用二进制自定界码对去年 2017 年进行编码。

$$
\begin{aligned}
2017_2 &= && \text{11111100001—11 比特}\\
11_2 &= && \text{1011—4 比特}\\
4_2 &= && \text{100—3 比特}\\
3_2 &= && \text{11—2 比特}
\end{aligned}
$$

因此，发送 10 100 1011 11111100001 0。回顾第 3 步。

如何解码： 检查前三个数字。如果第三个数字是 1，则前两个数字表示接下来要读取的数字位数，记作 r_n。检查第$r_n + 1$ 个数字。如果它是 0，则 r_n 个数字就是该数字；否则，它表示接下来要读取的数字位数。

Elias 编码器
操作编码器由**Elias 编码器**给出（摘自 [54]）

写入 ’0’ 循环 { 如果 $\lfloor \log n \rfloor = 0$ 则
停止。将 $C_2(n)$前置到已写入的字符
串$n := \lfloor \log n \rfloor$}

可以计算自定界码对$n \in \mathbb{N}$进行编码所需的比特数，如下：

$$L^*(n) = \lceil \log_2 n \rceil + \lceil \log_2 \lceil \log_2 n \rceil \rceil + \lceil \log_2 \lceil \log_2 \lceil \log_2 n \rceil \rceil \rceil + \ldots + 1 \tag{7.26}$$

对于整数 $n \in \mathbb{Z}$ 的编码，我们定义 $n' = 2n,\ n \geq 0$ 和 $n = -(2n-1),\ n < 0$然后对于 $n \in \mathbb{Z}, L^*(2n)$。也就是说，$n \in \mathbb{Z}$ 如果 $n > 0$ 则 $n \mapsto 2n$，否则 $n \mapsto -2n+1$，因此

$$L^*(n) = \lceil \log_2 2n \rceil + \lceil \log_2 \lceil \log_2 2n \rceil \rceil + \lceil \log_2 \lceil \log_2 \lceil \log_2 2n \rceil \rceil \rceil + \ldots + 1 \tag{7.27}$$

为了将编码扩展到有理浮点数 $\lambda \in \mathbb{Q}$，我们首先考虑以下表示形式，

$$\lambda = a \times 2^b \tag{7.28}$$

其中 $a \in [\frac{1}{2}, 1)$ 和 $b \in \mathbb{Z}$。由于计算内存有限，λ 的截断近似可以写为，

$$\overline{\lambda} = \pm 0.a_1 a_2 a_3 a_4 \ldots a_n \times 2^{\pm m} \tag{7.29}$$

其中我们始终只关注有限位表示（$n < \infty$），并且还需要额外的两比特来编码尾数和指数的符号。令 δ表示有限表示 λ 的精度 λ

$$\delta > |\lambda - \overline{\lambda}| \tag{7.30}$$

因此，在 n 位处进行截断（舍入）得到 $\delta = 2^{-n}$ （分别为 $\delta = 2^{-(n+1)}$）。

因此，若要对一个实数进行编码，可以改为¯编码一个截断的 λ，通过 n 位浮点数实现，所需码长为

$$\begin{aligned} L^*(\lambda) &= L^*(\pm a_1 a_2 a_3 a_4 \ldots a_n) + L^*(2m) \\ &= L^*(2 \times 2^{(n-1)}) + L^*(2m) \\ &= L^*(\lceil \delta^{-1} \rceil) + L^*(2m) \end{aligned} \tag{7.31}$$

7.4 最小模型描述长度（MDL）

我们希望通过计算使用模型来描述（编码）数据的代价，评估模型的性能（“优劣”）。

1. 对大量具有**已知** 分布（取决于系综）的数据进行霍夫曼编码——我们可以将其用于模型预测误差。2. 对少量**没有特定分布**的数据使用自定界码，$\mathbb{Z}^+$，$\mathbb{Z}$ 或者 $\mathbb{R}$ 对具有指定精度的数据进行编码——我们可以将其用于模型参数。

回想一下，我们有有限精度的数据 x_1，x_2，.....，$x_n \in \mathbb{R}$ 以及一个模型

$$x_{t+1} = f(x_t, x_{t-1}, \ldots, x_{t-d}; \lambda) + e_{t+1}, \quad \lambda \in \mathbb{R}^k \tag{7.32}$$

其误差 e_t 具有已知（或假定的）分布 P_e。

注意： e_t 也是有限精度的，而且希望其编码长度平均小于 x_t。我们还需要初始条件 $x_{-(d-1)}$, $x_{-(d-2)}$, ..., x_0，并且不失一般性地将它们包括在 λ中。因此，总而言之，我们有一个参数向量 θ，其中包括 λ、初始条件**和**精度 δ。e 和 x 都是有限精度向量；我们希望 $L(e) \ll L(x)$。因此，我们需要两段式编码：先编码参数，再编码 e_t。这些参数定义了 P_e和 f，并允许对 e_t进行解码。

因此，令 θ 表示完整的参数向量（λ、初始条件及所有其他内容——从现在开始，我们将 $|\theta| = k$ 以及 θ 的 j^{th} 分量记作 λ_j，并称其精度为 δ_j）。然后，我们可以通过发送两段式编码来评估模型：先发送 θ，再发送 e，总码长为

$$\begin{aligned} L(x, \theta) &= L(x|\theta) + L(\theta) \\ &= L(e) + L(\theta) \\ &\leq n(H(P_e) + 1) + L(\theta) \\ &\approx n(H(P_e) + 1) + \sum_{j=1}^{k} \left(L^*(\lceil \delta_j^{-1} \rceil) + L^*(2m_j) \right) \end{aligned}$$

其中，m_j是相应 λ_j的指数。但在实践中，我们的做法是

$$L(x, \theta) = L(\theta) + L(x|\theta) \tag{7.33}$$

$$\approx L(\theta) + L(x|\hat{\lambda}) + \frac{1}{2}\delta^T Q \delta \tag{7.34}$$

$$\approx \sum_{j=1}^{|\theta|} (L^*(2m_j) + L^*(\lceil \delta_j^{-1} \rceil)) + L(x|\hat{\lambda}) + \frac{1}{2}\delta^T Q \delta \tag{7.35}$$

因此，对于选定的某个模型类（例如，径向基、神经网络、自回归过程.....），用于模型选择的一般模型描述长度方法如下：

1. 找出使 $H(P_e)$ 最小的 $\hat{\lambda}$，但实际上应最小化假定误差分布的对数似然。通常，我们假设模型误差服从加性高斯分布。 2. 给定 $\hat{\lambda}$，选择 δ 的最优值。3. 使用总码长选择最佳模型（即总码长最低的模型）

路线图：
首先作出一些假设和近似会比较方便。

- 将 λ_j 的指数限制在某个常数以内，例如 γ（= 32）。

- 用首项近似 L^*，这样使用 γ 后，参数的码长变为

$$L(\theta) \approx \sum_{j=1}^{k} \log(\frac{\gamma}{\delta_j})$$

- 找到使 $L(e)$ 最小的 $\hat{}\lambda$，其中 $-\log(P_e)$

路线图：续

- 给定 $\hat{}\lambda$，选择最优的 δ
- 计算 $(L(e) + L(\theta))$，并用它从一般形式（7.12）中选出最佳模型（即最优的 k）。

- **注意：** P_e 取决于 σ^2，我们还必须求出其解，并连同其精度一起给出，因此，

$$L(\theta) \approx \sum_{j=0}^{k} \log(\frac{\gamma}{\delta_j})$$

且 $\delta_0 \hat{\sigma}^2$ 的精度。
- 我们也将使用纳特而非比特（即使用 log 而不是 log2）。

考虑如下形式的广义线性模型

$$f(x_t, x_{t-1}, \ldots, x_{t-d+1}; \lambda) = \sum_{j=1}^{k} \lambda_j \phi_j(x_t, x_{t-1}, \ldots, x_{t-d+1}) \tag{7.36}$$

其中 k非零，λ_j精度为 δ_j。对于线性模型

$$f(x_t, x_{t-1}, \ldots x_{t-d+1}; \lambda) = \sum_{j=1}^{d} \lambda_j x_{t-j+1}. \tag{7.37}$$

$L(x$ 的典型剖面 $,\theta)$ 关于模型规模 k 的示意图见图7 2
. . 近似地，$L(\theta) \approx \sum_{j=1}^{k} L^*(\lceil \delta^{-1} \rceil)$ + 常数。我们将尾数 m_j界定为 γ，这需要为其单独设置精度。为了求出各项精度，即 δ's，考虑

$$L(x) \approx L(x|\hat{\lambda}) + \frac{1}{2}\delta^T Q \delta + O(\delta^3) + \sum_{j=1}^{k} L^*(\lceil \delta_j^{-1} \rceil) + \sum_{j=1}^{k} \log \gamma. \tag{7.38}$$

我们使用纳特（自然对数比特）作为单位，实际得到的是 $-\sum_{j=0}^{k} \log \frac{\gamma}{\delta_j}$，因为 $j = 0$对应于 σ^2，其精度为 η，见下文。我们希望关于 δ_i 最小化 $L(x)$。因此，

$$\frac{\partial}{\partial \delta_i} L(x) = 0 \tag{7.39}$$

这意味着

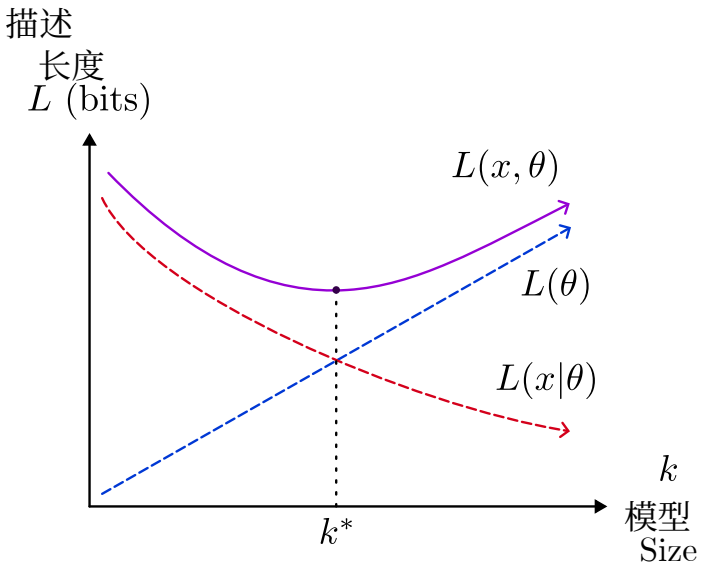


图 7.2 广义线性模型总码长随模型规模变化的典型剖面 k。模型预测误差的代价与模型参数的代价之间存在权衡。

$$\frac{1}{\delta_i} = [Q\delta]_i \tag{7.40}$$

而 $[\cdot]_i$ 表示该向量的第 ith 个分量。我们对 δ 求解，因此需要求出 Q。为此，我们考虑一个特定的模型。具体而言，对于广义线性模型，指定基函数。令 $x^{(t)} = x_{t-1}$, ... x_{t-d}，则:

- $\phi_j = \phi(a_j x^{(t)} - b_j)$ yields a feedforward neural network;
- $\phi_j = \phi(\frac{\|x^{(t)} - c_j\|}{r_j})$ is a radial basis network.

令 V 为（$n-d$）$\times k$ 矩阵，其列和行分别表示每个基函数在各个数据点上的取值。则该模型可用矩阵表示法表示为

$$e = x - V\lambda \tag{7.41}$$

假设（对于霍夫曼编码），误差 e 服从高斯分布 $e \sim N(0, \sigma^2)$，因此

$$P_{e_i} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{e_i^2}{2\sigma^2}} \tag{7.42}$$

So,

$$\begin{aligned} L(x,\theta) &= L(x|\theta) + L(\theta) \\ &= -\sum_{i=1}^{n} \log P_{e_i} + L(\theta) \\ &\approx \frac{n}{2}\log 2\pi\sigma^2 + \frac{1}{2\sigma^2}(x - V\lambda)^T(x - V\lambda) + L(\theta) \end{aligned} \tag{7.43}$$

我们要寻找 $\lambda = \hat{\lambda}$ 和 $\sigma^2 = \hat{\sigma}^2$的最优值。因此，同时求解 $D_\lambda L(x,\theta) = 0$ 和$D_{\sigma^2} L(x,\theta) = 0$。

对于 $D_\lambda L = 0$:

$$\left[-\frac{1}{\sigma^2} V^T (x - V\lambda) \right] = 0 \tag{7.44}$$

意味着 $\hat{\lambda} = (V^T V)^{-1} V^T x$，即通常所说的“最小二乘法”解。

对于 $D_{\sigma^2} L = 0$:

$$\left[\frac{n}{2}\frac{1}{\sigma^2} - \frac{1}{2(\sigma^2)^2}(x - V\lambda)^T(x - V\lambda) \right] = 0 \tag{7.45}$$

意味着 $\hat{\sigma}^2 = \left[(x - V\lambda)^T (x - V\lambda) \frac{1}{n} \right] = \frac{e^T e}{n}$。现在，

$$Q = \begin{bmatrix} D_{\lambda\lambda} L(x,\theta)|_{\hat{\lambda},\hat{\sigma}^2} & D_{\lambda\sigma^2} L(x,\theta)|_{\hat{\lambda},\hat{\sigma}^2} \\ D_{\sigma^2\lambda} L(x,\theta)|_{\hat{\lambda},\hat{\sigma}^2} & D_{\sigma^2\sigma^2} L(x,\theta)|_{\hat{\lambda},\hat{\sigma}^2} \end{bmatrix} \tag{7.46}$$

我们求得 $D_{\lambda\lambda} L|_{\hat{\lambda},\ \hat{\sigma}^2} = \frac{1}{\hat{\sigma}^2} V^T V$, $D_{\lambda\sigma^2} L|_{\hat{\lambda},\ \hat{\sigma}^2} = 0$和

$$\begin{aligned} D_{\hat{\sigma}^2\hat{\sigma}^2} L|_{\hat{\lambda},\hat{\sigma}^2} &= -\frac{n}{2}\frac{1}{(\sigma^2)^2} + \frac{2}{2(\sigma^2)^3} e^T e|_{\hat{\lambda},\hat{\sigma}^2} \\ &= -\frac{n}{2}\frac{1}{(\hat{\sigma}^2)^2} + \frac{n}{(\hat{\sigma}^2)^2} \\ &= \frac{n}{2}\frac{1}{(\hat{\sigma}^2)^2} \end{aligned} \tag{7.47}$$

因此,

$$Q = \begin{bmatrix} \frac{1}{\hat{\sigma}^2} V^T V & 0 \\ 0 & \frac{n}{2}\frac{1}{(\hat{\sigma}^2)^2} \end{bmatrix} \tag{7.48}$$

因此，精度 δ 和 η 可以由以下公式得出

$$\left[\frac{1}{\hat{\sigma}^2} V^T V \delta \right]_j = \frac{1}{\delta_j} \tag{7.49}$$

and

$$\frac{n}{2}\frac{1}{(\hat{\sigma}^2)^2}\eta = \frac{1}{\eta}, \tag{7.50}$$

其解为 $\eta = \sqrt{\frac{2}{n}}\hat{\sigma}^2$。据此可得模型描述长度（MDL）的以下表达式：

$$
\begin{aligned}
L(x,\theta) &\approx \frac{n}{2}\log(2\pi\hat{\sigma}^2) + \frac{1}{2\hat{\sigma}^2}e^T e + \frac{1}{2}\delta^T Q\delta + \sum_{j=0}^{k}\log(\frac{\gamma}{\delta_j}) \\
&\approx \frac{n}{2}\log(2\pi\hat{\sigma}^2) + \frac{1}{2\hat{\sigma}^2}e^T e + \frac{1}{2}(k+1) + (k+1)\log\gamma \\
&- \sum_{j=1}^{k}\log\hat{\delta}_j - \log\eta \\
&= \frac{n}{2}\log 2\pi + \frac{n}{2}\log\frac{e^T e}{n} + \frac{n}{2} + (k+1)(\frac{1}{2} + \log\gamma) \\
&- \sum_{j=1}^{k}\log\hat{\delta}_j - \log\sqrt{\frac{2}{n}}\hat{\sigma}^2 \\
&= (\frac{n}{2} - 1)\log\frac{e^T e}{n} + (k+1)(\frac{1}{2} + \log\gamma) - \sum_{j=1}^{k}\log\hat{\delta}_j \\
&+ \frac{n}{2}(1 + \log 2\pi) + \frac{1}{2}\log\frac{n}{2}
\end{aligned}
\tag{7.51}
$$

定义 7.4. – 模型描述长度（最小描述长度）
对于误差服从高斯分布的模型，其模型描述长度定义为

$$DL(k) = (\frac{n}{2}-1)\log\frac{e^T e}{n} + (k+1)(\frac{1}{2}+\log\gamma) - \sum_{j=1}^{k}\log\hat{\delta}_j + \frac{n}{2}(1+\log 2\pi) + \frac{1}{2}\log\frac{n}{2}$$

注意：对于线性模型，BIC（SIC）是描述长度 [42, 55] 的渐近近似。

回顾一下其原理。我们有一个时间序列 y_t，并假设它来自对某个动力系统状态的测量，该系统在某个低维吸引子上演化。我们使用时间延迟嵌入方法进行嵌入，以重构一个等价的状态空间，其中状态为 x_t；从某种意义上说，其动力学等价于未知的“真实”动力学。因此，通过坐标变换保持不变的量，无论是直接从（嵌入后的）数据计算，还是通过演化模型间接计算，都有助于刻画原始系统。

如果我们构建一个全局模型，如何选择一个有用的模型，使得

$$\begin{aligned} y_{t+1} &= f(x_t) \\ &= \sum_{i=1}^{k} \lambda_i \phi_i(x_t) + e_t \end{aligned} \tag{7.52}$$

其中 ϕ_i 是基函数，使得 $e = y - V\lambda$? 请注意，该模型关于（未知）参数 λ 是线性的。我们会发送估计出的参数、精度和预测误差。为了比较模型，我们使用模型描述长度（最小描述长度）

$$DL_k(\theta) = (\frac{1}{2}n-1)\ln\frac{e^Te}{n} + (k+1)(\frac{1}{2}+\ln\gamma) - \sum_{j=1}^{k}\ln\hat{\delta}_j + \frac{1}{2}n(1+\ln 2\pi) + \frac{1}{2}\ln(\frac{1}{2}n) \tag{7.53}$$

其中， n 是数据的数量， γ 是常数， $\hat{}\delta$ 求解 $[Q\delta]_j = \frac{1}{\delta_j}$ ，而 θ 表示数据、参数、精度等。

7.4.1 基于*MDL*的模型选择

总体方案

选择描述长度最小的模型作为“最佳”或最有用的模型。因此，我们得到以下模型选择路线图。

1. 选择模型类（即 $y_t = \sum_{i=1}^{k}\lambda_i\phi_i(x_t)$中的基函数，使得$y = V\lambda$）。2. 求出使 $e = y - V\lambda$最小的$\hat{}\lambda$。3. 给定$\hat{}\lambda$，选择最优的 δ （即求解$[Q\delta]_i = \frac{1}{\delta_i}$）。4. 计算 DL_k ，并据此选择最佳模型（即使DL_k 达到最小值的最优 k ）。5. 或者，也可以额外计算 $AIC(k)$或 $SIC(k)$，并选择最佳模型（即使 $AIC(k)$或 $SIC(k)$达到最小值的模型）。

子集选择

模型选择本质上归结为求解以下优化问题

$$\text{minimize } |y - V\lambda| \text{ subject to } N(\lambda) = k.$$

其中， N 是参数的数量。这很困难。为了解决这一问题，我们首先从一个固定的 k 入手，并假设误差服从高斯独立同分布噪声，然后求解相应的优化问题，

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}e^Te \\ \text{over} & e, \lambda \\ \text{subject to} & V\lambda + e = y, N(\lambda) = k. \end{array}$$

在这种形式下， $N(\lambda) = k$ 是复杂因素。其思路是从一个规模为 k 的模型开始，然后从一组候选基函数中确定要添加到模型中的“最佳”

函数，使其规模增至 $k+1$。然后确定从模型中剔除哪一个 $k+1$ 基函数是“最佳”的，使规模恢复为 k 个基函数。通过迭代方式完成这一过程，以找到“最优”模型，其中最优意味着该模型具有最小描述长度。

扩展模型

若我们将 B 表示为基本（模型项）指标的集合，即

$$B = \{j : \lambda_j \neq 0\}, \tag{7.54}$$

于是 $N(\lambda) = |B|$，因此我们可以通过灵敏度分析，研究改变 B 大小所产生的影响。为此，使用辅助变量并应用拉格朗日优化理论会很方便。我们可以用非基本变量表示约束 $N(\lambda) = k$，即

$$\lambda_j = u_j, \quad j \notin B, \tag{7.55}$$

其中参数 u 均为零。该优化问题的拉格朗日函数为

$$L = \frac{1}{2} e^T e + \nu^T (y - V\lambda - e) + \mu^T (u - \lambda), \tag{7.56}$$

其中 μ 和 ν 为对偶变量。对 e 和 λ 进行无约束优化（$\frac{\partial L}{\partial e} = 0 = \frac{\partial L}{\partial \lambda}$）可得

$$\nu = e \tag{7.57}$$

and

$$\mu = -V^T e. \tag{7.58}$$

现在，μ 是对应于非基约束的对偶变量，因此它表示在最优状态下对 u 变化的敏感度。因此，应将 μ 中绝对值最大的元素加入模型，即将其设为基变量，以获得最大的边际收益。

精简模型

我们希望找出应从模型中移除的基变量，使均方误差受到的影响最小。将 λ 和 e 的最优值代入拉格朗日函数，考虑对偶优化问题，即

$$\begin{aligned} &\text{maximize } -\tfrac{1}{2}\nu^T\nu + \nu^T y + \mu^T u \\ &\text{over} \qquad \mu, \nu \\ &\text{subject to } (V^T\nu)_j = w_j, \text{ for } j \in B. \end{aligned}$$

与之前类似，$w = 0$ 保留为参数，但我们立即设置 $u = 0$。注意，涉及 w 的约束会影响基变量，因此对 w 的敏感性反映了已在基中的变量。由于问题是凸的，原始解与对偶解相同。因此，关于 μ 的对偶变量为λ，并且

$$\lambda_j = \frac{\partial \psi}{\partial w_j}, \tag{7.59}$$

其中 ψ 是最优解。因此，选择 λ_j 中绝对值最小的元素作为非基变量，即从模型中剔除它，以尽可能减少对收益的边际损害。

由此可得出以下算法，用于确定并评估一个“良好”模型。令 V_B 为由 V 中索引属于 B的各列组成的 n × k 矩阵。令 λ_B 为 $y = V_B\lambda$ 的最小二乘解，并令 $e_B = y - V_B\lambda_B$。

1. 令 $S_0 = (\frac{1}{2}n - 1)\ln(\frac{y^Ty}{n}) + \frac{1}{2} + \ln\gamma$。2. 令 $B = \{j\}$，其中 V_j是 V中使 $|V_j^Ty|$达到最大值的列。3. 令 $\mu = V^Te_B$和 i 为 μ 中绝对值最大的分量的索引。（索引 i 是进入基的索引。）令 $B' = B \cup \{i\}$。4. 计算 $\lambda_{B'}$。令 o 为 B' 中对应于 $\lambda_{B'}$绝对值最小分量的索引。（索引 o 是离开基的索引。）5. 如果 $i \neq o$，则令 $B = B' \setminus \{o\}$，并转到第 3 步。6. 定义 $B_k = B$，其中 $k = |B|$。求 δ，使得对每个 $j = \{1, 2, \ldots, k\}$，$(V_B^TV_B\delta)_j = \frac{1}{\delta_j}$，并计算 $S_k = (\frac{1}{2}n - 1)\ln(\hat{e}^T\hat{e}/n) + (k+1)(\frac{1}{2} + \ln\gamma) - \sum_{j=1}^{k}\ln\hat{}\delta_j$。7. 如果 $S_k < S_{k-1}$，则转到第 3 步。

8. 选择使 S_k 最小的基 B_k 作为最优模型。

第 8 章 替代数据

在研究时间序列和动力系统时，我们通常会求解相关的正问题。也就是说，给定一组方程（常微分方程），求解这些常微分方程以生成相关的时间序列。在现实中，逆问题更为常见，也更难解决。给定一个观测到的时间序列，我们能从中了解到有关动力过程的什么信息？解决逆问题时会遇到许多挑战：

- 学习全局模型可能有助于预测，但不一定能提供有用的信息
- 噪声和非线性会影响线性自相关的表现。
- 估计不变测度（例如，相关维数、李雅普诺夫指数）容易出错并导致误解，除非有大量高质量数据可用。

我们还可以考虑以下相关问题：

Q1. 时间序列中变异性（动力学）的成因是什么？**Q2.** 我们如何找到简单的方法来检验并排除“简单”的解释？

为什么这很重要？了解数据生成过程的潜在形式，有助于选择适当的分析方法。例如，如果动力学是非线性的，我们会认为自相关没有用处。如果过程是随机的，那么李雅普诺夫指数就无法提供有用信息。基于这些认识，我们旨在找出最适合分析的模型。一种解决方法是构造替代数据来进行假设检验。具体做法是回答两个更简单的问题：

Q3. 该过程是非线性的吗？**Q4.** 如果是，那么应该采用哪一类非线性模型，才能得到与数据一致的结果？

统计学中的假设检验，是指根据观测数据计算某个统计量，将其与某个理论分布进行比较，并

得出结论：要么拒绝某个潜在的零假设，要么无法拒绝该假设。替代数据假设检验遵循以下一般流程：

1. 定义关于数据类型的零假设 H_0。2. 定义一个与 H_0 一致的数据生成过程（即替代过程），同时保留与观测数据相符的其他特征（例如均值、傅里叶谱等）。3. 生成一个替代数据集合 s_i（t）。4. 定义一个检验统计量 d（李雅普诺夫指数、关联维数），并计算 d（s_i（t））~D 的分布。5. 计算 $d_o = d$（x（t）），即原始观测数据的检验统计量。6. d_0 是否符合分布 D?

- **是** $\implies$ H_0 无法被拒绝。数据要么由替代过程生成，要么是因为 $d()$ 并非区分这两个过程的恰当选择
- **不** $\implies$ 拒绝 H_0，并得出结论：x（t）是更“复杂”过程的结果。

替代数据假设检验程序需要满足两项条件：

1. 一个零假设，以及与之对应的算法，用于生成替代时间序列。2. 用于通过 p-值检验进行决策的判别统计量。

8.1 替代算法层级

基本的替代数据检验涉及三种主要假设，每种假设都有各自的替代数据生成算法。为简便起见，我们将这些算法称为算法0、算法1和算法2。如果一个时间序列拒绝了所有算法的零假设，我们可以得出结论：这些观测值是更复杂过程的结果。

8.1.1 算法*0*：随机置换替代数据

假设：观测数据是一个独立同分布噪声过程的产物。

为检验这一假设，生成替代数据以保留数据的概率分布，同时破坏任何时间结构。这可以通过随机打乱数据的时间顺序来实现。算法0的步骤如下：

1. 令 $\{x_t\}_{t=1}^N$ 为原始观测时间序列 2. 令 $\{s_t\}_{t=1}^N$ 为新生成的时间序列 3. s_t 通过打乱 x_t 的时间排序计算得出（无放回抽样

a. u = randn(N,
1) b. idu =sort(u)
c. $s_t = x\,[idu]$

随机打乱会消除数据中的任何自相关，因此 $s_t \perp s_{t-\tau} \forall\, \tau$

8.1.2 算法*1*：傅里叶变换替代序列

假设：观测数据本质上是经过线性滤波的高斯噪声（等价于自回归过程）。

给定一个标量时间序列 x_t，线性滤波器是一种变换，它会生成另一个具有以下形式的时间序列 y_t：

$$y_t = \sum_{j=1}^{N} \beta_j x_{t-j}, \tag{8.1}$$

其中 β_j 为常数。也可以在频域中表示线性滤波器：

$$Y_\omega = \beta_\omega X_\omega \tag{8.2}$$

这 β （即 β_ω的集合）称为传递函数。

算法1尝试生成一个部分随机的时间序列 [56]。底层数据值是随机生成的，但线性平滑过程会使时间上相近的值彼此相关。因此，数据中所有有趣的特征完全源于自相关函数。通常，应用线性滤波器只会改变功率谱。

在实践中，我们旨在构造一个替代时间序列 s_t，它保留相同的自相关函数（即相同的功率谱），但除此之外是随机的。为此，我们对数据的相位进行随机置换，同时保留振幅结构：

1. 令$\{x_t\}_{t=1}^{N}$ 为观测到的原始时间序列。2. 应用离散傅里叶变换
$X_n = \mathrm{F}(x_t) = \{R_n e^{-i\pi\theta_n + \phi_n}\}$ 3. 将复数相位随机化：

a. 令 $\psi_n \sim \mathrm{U}(0, 2\pi)$ b. 将每个振幅乘以 $e^{i\psi}$ 以生成 S_n。确保 $\psi(f) = -\psi(-f)$，这样逆傅里叶变换的结果就是实数

4. 应用逆傅里叶变换 $\{s_t\}_{t=1}^{N} = \mathrm{F}^{-1}(S_n)$

对于具有区分性的统计量而言，线性自相关是不够的，因为该算法会有意保留这一量。因此，采用非线性形式的自相关（如平均互信息）作为统计量是合适的。算法0存在几个相关问题：

- 数值傅里叶变换并不完美，因为它无法完全保留自相关结构。
- 处理非周期数据或首尾不匹配的数据时，傅里叶变换也可能存在问题。

8.1.3 算法*2*：幅值调整傅里叶替代数据

假设： 观测数据由经过线性滤波的噪声描述，但随后又通过一个平滑的观测函数h进行变换，其中h： $\mathbb{R} \to \mathbb{R}$ 是一个单调的非线性函数（即$h(a) > h(b) \Longleftrightarrow a > b$）。

算法2的零假设认为，所有观测到的非线性都源于观测函数，而不是内在动力学。当数据显然不服从高斯分布，但人们怀疑其根本原因是线性随机过程 [57]时，这一假设很有用。过程如下：

1. 令 $\{x_t\}_{t=1}^{N}$ 为观测数据。生成一组 N 高斯随机数 $\{z_t\} \sim \mathrm{N}(0,1)$，并按照与 x_t 相同的秩分布对 z_t 重新排序，以复制功率谱。2. 对 $\{z_t\}$ 应用傅里叶变换算法1，生成序列 $\{\hat{s}_t\}$，使其具有相同的功率谱（自相关结构）。3. 按照 $\{\hat{s}_t\}$ 的秩排序对 $\{x_t\}$ 重新排序，生成替代数据 $\{s_t\}$，使其具有与原始观测数据相同的谱、自相关、均值和标准差。

在实践中，幅值调整傅里叶替代数据（算法2）无法完美保留自相关和功率谱。这是因为秩排序（重新缩放）过程会导致谱发生改变。为了更好地保留功率谱，可以对上述过程进行迭代，这种方法恰当地称为迭代幅值调整傅里叶变换 [58],

1. 计算由 $\{S_k^2\}$ 给出的傅里叶变换 $\{x_t^{(0)}\}$ 的振幅平方。2. 对观测数据 $\{x_t\}$ 应用随机打乱算法0，生成 $\{\hat{s}_t^{(i)}\}$。3. 对 $\{\hat{s}_t^{(i)}\}$ 进行傅里叶变换，并将其振幅替换为 $\{S_k^2\}$，以保留所需的功率谱。4. 进行逆傅里叶变换，同时保留序列的相位。这样可以确保谱正确，但标量分布不正确。5. 按照未经变换的 $\{\hat{s}_t^{(0)}\}$ 对 $\{x_t\}$ 重新排序，生成替代序列 $\{\hat{s}_t^{(i+1)} \quad\}$

6. 检查功率谱；如果不够接近，则重复操作。

适用于这些替代数据的判别统计量是关联维数。

8.2 其他替代数据算法

算法0、1和2的标准替代数据组易于实现，可用于判断是否有必要进行非线性分析。然而，还有其他算法是对这三种检验的改进，旨在检验更一般（或更具体）的零假设。由于这类算法数量众多且各有不同，我们仅简要讨论其中少数几种。

8.2.1 截断傅里叶变换替代数据

TFTS算法是对标准傅里叶变换替代数据（算法1）的一种变体。该算法旨在检验某一类非平稳过程，即具有长期趋势（周期性）但除此之外还包含不规则波动的数据。具体来说，我们想探究这些不规则波动中是否存在非线性，还是它们由随机过程所致。其思路是“破坏”不规则波动中的非线性，同时保留“全局”行为（即趋势和周期性）。

为实现这一点，对数据进行傅里叶变换并检查功率谱。不要打乱所有相位，而只随机打乱高频相位，即落在某个频带内的那些频率 f_z（见图 8.1）。这个参数需要调节。替代数据是对截断后的置乱傅里叶变换进行逆傅里叶变换得到的。由于只打乱高频分量，长期趋势（低频周期）得以保留，只有不规则波动被随机化。短期波动中的任何非线性都应被破坏。对于这些替代数据，可采用跨多个时滞计算的平均互信息作为判别统计量。

8.2.2 小幅洗牌替代数据（*SSS*）

TFTS 方法是对算法1替代数据生成方法的改进，用于考察具有长期趋势的数据中的不规则波动是否存在非线性的证据。它通过在频域的高频受限频带内打乱相位来实现。在时域中运行、可视为算法0替代数据方法的类似改进的一种替代数据生成方法是小幅洗牌替代数据（SSS） [59]。该检验的零假设是，不规则波

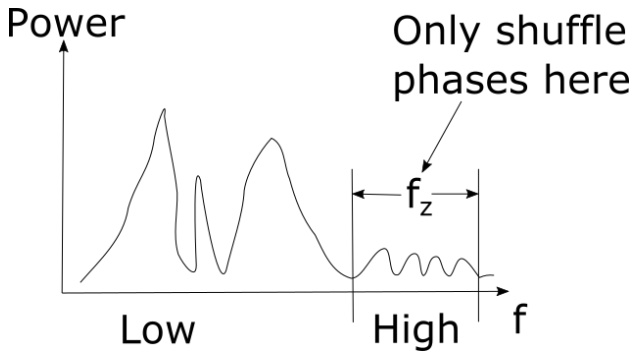


图 8.1 TFTS 背后的思想。

动是独立同分布的随机变量，即不存在短期动力学。正如前文所述，该方法本质上是尺度更小的算法0。

1. 从时间序列 $x(t)$ 开始，令 $i(t)$ 为 $x(t)$ 的索引，使得 $i(t) = t$。2. 令 $g(t)$ 为一个高斯随机数，则对于 $A > 0$，$i'(t) = i(t) + Ag(t)$。3. 对 $i'(t)$ 按秩排序，使得 $i'(t)$ 的索引为 $\hat{i}(t)$。4. 设定 $s(t) = x(\hat{i}(t))$。

8.2.3 伪周期替代数据（*PPS*）

这些方法适用于具有强周期性的数据。零假设检验数据是否与非周期轨道加不相关噪声相一致，因此采用伪周期替代数据（PPS） [60]。取时间序列 $\{x_t\}$，使用时滞为 τ、维数为 d 的时间延迟嵌入进行嵌入，得到重构状态 $\{z_t\}$。

1. 随机选择 $s_1 \in \{z_t\}(i = 1)$。2. 以概率 $P(s_j = z_t) \approx \exp(\frac{-\|z_t - s_j\|}{\rho})$ 随机选择一个邻居 s_j。3. 设置 $s_{i+1} = s_{j+1}$，递增 i。4. 重复，直到 $i > N$。

该方法结合了最近邻预测和 S-map 预测器。现在，将标量时间序列 y_t（替代数据）选为向量 s_t 的第一个分量。注意，s_t 和 z_t 大致是同一个“吸引子”，但 $\{y_t\}$ 的嵌入只是对该吸引子的近似。替代数据（近似地）遵循与原始数据相同的向量场，但细尺度动力学已被抹去。参数 ρ 需要调节：当 ρ 较大时，我们实际上得到的是算法0替代数据；当 ρ 较小时，不会有变化（我们会持续选择 s_i，因此遵循原始数据）。在中间某处存在一个 ρ，此时周期间动力学被破坏，而周期内动力学得以保留。（关于如何选择

一个合适的 ρ，请参见论文）。与周期打乱替代数据进行比较。

8.2.4 吸引子网络替代数据

PPS 算法通过构造在相邻轨道间随机切换的轨迹来生成替代数据。这会破坏轨迹的分形结构，但能近似保留向量场。也可以考虑动力学的离散形式。与符号动力学类似，我们可以将连续相空间中的流简化为相空间中互不重叠（不一定大小相同）区域之间的离散跃迁。在本单元前面讨论的传统符号动力学中，任何一组动力学都可以表示为一个 n^{th} 阶马尔可夫链，适用于无限的 n。（例如，对于洛伦兹系统，我们可以将 L 和 R 翼视为符号。）可以通过增加离散区域的数量 N 来进一步发展这一思想；随着 $N \to \infty$ 增大，动力学可以由 1st 阶马尔可夫链任意逼近（即无记忆性）。

吸引子网络替代数据（ANS）利用上述原理，构建系统动力学的离散马尔可夫链表示，其中各区域的边界由一组训练数据 [61] 确定。构建完成后，可以沿马尔可夫链进行随机游走，生成替代时间序列。与 PPS 类似，ANS 大体上保留了目标系统的向量场（在这种情况下，即相空间转移概率），因此能够复现观测数据的遍历测度（李雅普诺夫指数和关联维数），但它是随机的，而非确定性的。这为评估分类算法提供了更严格的替代数据检验。

8.3 约束替代数据

在一般的算法 0、1 和 2 中，目标是保留数据的某项给定性质，例如随机性或线性性质。在这种情况下，我们可以将生成的替代数据称为“受约束实现”，因为替代数据保留了原始数据的某些性质。不过，也可以使用**无约束实现**。进一步讨论前，有必要先对两个概念作出恰当的定义。

定义 8.1. – 受约束替代数据
设 H_0 是给定替代数据检验的零假设，P 是与 H_0 一致的所有数据生成过程的集合。假设 P 中的所有过程元素均可由一组参数 α 参数化。因此，任何与 H_0 一致的 $f \in \mathrm{P}$ 中的数据生成过程，都可以通过其对应参数 $\alpha(f)$ 唯一确定。设 x_t 为观测数据，s_t 为一组生成的替代数据。如果参数 $\alpha(x_t)$ 和 $\alpha(s_t)$ 的估计值服从相同的分布（但不一定相等），那么该替代数据生成算法生成的就是受约束实现。也就是说，这不依赖于拟合模型。相比之下，由 $\alpha(x_t)$ 生成的是无约束替代数据。这可以通过拟合某个模型过程（例如使用最大似然法）来实现；然后利用所得模型生成新数据。

定义 8.2. – 枢轴统计量
如果从 $f \in \mathrm{P}$ 获得的统计量值的分布独立人士于 f 的选择，则该统计量是枢轴统计量。换言之，由所有 f 与零假设 H_0 一致的过程生成的数据集的统计量遵循相同的分布。

枢轴统计量意味着，在进行假设检验时，无须确保替代数据生成方案受到约束；否则就必须确保这一点。当待检验的假设是与关联维数相符的动力学一般类别（即迄今讨论过的所有非线性动力学和混沌）时，关联维数的无偏估计可提供枢轴统计量。因此，借助关联维数（以及其他动力学不变量），可以采用基于模型的方法（该方法会生成受约束替代数据），检验某个系统是否属于广泛的动力系统类别。

第 9 章 储备池计算

回顾一下第 6.4.3 节中对循环神经网络的定义，以及演化函数（对激活状态的时间指标稍作调整后）：

$$\vec{s}(t) = \sigma \circ (C_{in}\vec{x}(t) + C_{rec}\vec{s}(t-1) + \vec{b}), \tag{9.1}$$

与下一步预测

$$\vec{x}(t+1) = C_{out}\vec{s}(t)\,. \tag{9.2}$$

在典型的循环神经网络设置中，$C_{in} \in \mathbb{R}^{n\times k}$、$C_{rec} \in \mathbb{R}^{k\times k}$、$\vec{b} \in \mathbb{R}^{k}$ 和 $C_{out} \in \mathbb{R}^{m\times k}$ 中包含的每个参数都需要调优，即使网络规模适中（k），计算量也会很快变得十分庞大。相反，如果我们能利用对动力系统的理解，预先以某种精巧的方式设计网络，从而降低后续的计算需求，会怎么样呢?

Herbert Jaeger 在 2001 [62] 中的研究聚焦于此。Jaeger 观察到，在某些参数设置下，循环神经网络的状态可以看作是输入历史的“回声”。在这些条件下会出现一些有趣的性质，使我们能够简化储备池计算机所需的训练过程。虽然储备池计算的发展通常归功于 Jaeger 及其回声状态网络（ESN），但 Mass 等人 [63] 于 2002 年独立提出了这一构想，并将其应用于液态状态机（LSM）。

9.1 回声状态性质

定义 9.1. – 回声状态
考虑一个左无限输入序列 $\vec{x}(-\infty)$、……、$\vec{x}(T-1)$、$\vec{x}(T)$，以及由此产生的网络状态 $\vec{s}(-\infty)$、……、$\vec{s}(T-1)$、$\vec{s}(T)$ 和 $\vec{s}'(-\infty)$、……、$\vec{s}'(T-1)$、$\vec{s}'(T)$，其中 $\vec{s}(t) = f_{RC}(\vec{s}(t-1), \vec{x}(t))$，且 $\vec{s}'(t) = f_{RC}(\vec{s}'(t-1), \vec{x}(t))$。如果网络状态由输入序列 $\vec{x}$ 唯一确定（即满足 $\vec{s}(T) = \vec{s}'(T)$），则称该网络具有回声状态。

储备池计算机的能力取决于回声状态性质，不过近来一致性性质一词也开始被使用 [64]。其理念是，在具备回声状态的网络中，网络的初始化并不重要，每个状态都仅由输入决定。实际应用中，检验回声状态需要评估底层系统的三个特征。

定理 9.1. – 实用回声状态性质
若网络具备回声状态，则称该网络：

- 是一致状态收缩，
- 具有状态遗忘性，并且
- 具有渐 *fading* 记忆。

定义 9.2。– 一致状态收缩
如果存在一个零序列 $\delta(h)$（即趋于 0 的序列），使得对于所有右无限输入序列 $\vec{x}(T)$，$\vec{x}(T+1)$，$\ldots$，$\vec{x}(\infty)$及其状态 $\vec{s}$，$\vec{s}$，，均满足$||f_{RC}(\vec{s}, \vec{x}(h) - f_{RC}(\vec{s}, \vec{x}(h))||_2 < \delta(h)$，且对所有输入序列前缀$\vec{x}(h) = \vec{x}(T), \vec{x}(T+1), \ldots, \vec{x}(h)$，其中$h \geq 0$，则称该网络为一致状态收缩。

定义 9.3。– 状态遗忘
如果存在一个零序列 $\delta(\mathrm{h})$，使得对于所有具有状态 $\vec{s}, \vec{s}'$ 的左无穷输入序列 $\vec{x}(-\infty), \ldots, \vec{x}(T-1), \vec{x}(T)$，均满足 $||f_{RC}(\vec{s}, \vec{x}(\mathrm{h}) - f_{RC}(\vec{s}', \vec{x}(\mathrm{h}))||_2 < \delta(\mathrm{h})$；对于所有输入序列后缀 $\vec{x}(\mathrm{h}) = \vec{x}(T-\mathrm{h}), \ldots, \vec{x}(T-1), \vec{x}(T)$（其中 $\mathrm{h} \geq 0$），网络便称为具有状态遗忘性质。

前两项性质与网络状态随时间的收缩有关。这些性质确保，对于相同的输入序列，具有不同初始条件的回声状态网络中的状态会随时间逐渐收敛；只要持续输入相同的输入序列，这些状态就不会发散。

定义 9.4。——渐 fading 记忆
当近期过去的输入彼此接近时，与之对应的输出也彼此接近，即使这些输入在遥远的过去可能大不相同，此时网络就被称为具有渐 *fading* 记忆 。

形式化地说，如果对于所有左无穷输入序列$\vec{x}$({v2)、.....、$\vec{x}(T-1)$、$\vec{x}(T)$，都存在一个零序列 δ(h)，使得对于所有输入序列后缀 $\vec{x}$(h) = $\vec{x}(T-\mathrm{h})$、.....、$\vec{x}(T-1)$、$\vec{s}(T)$（其中 h ≥ 0），对于所有左无穷输入序列 $\vec{w}\cdot\vec{x}$(h) 和 $\vec{w}'\cdot\vec{x}$(h)（即其后缀为 $\vec{x}$(h) 的左无穷序列），以及所有与 $\vec{w}\cdot\vec{x}$(h) 末端相容的状态 $\vec{s}$ 和与 $\vec{w}'\cdot\vec{x}$(h) 末端相容的状态 $\vec{s}'$，均有 $||\vec{s}-\vec{s}'||_2 < \delta$(h)，则称网络具有渐 fading 记忆。

渐忘记忆属性指出，过去输入对回声状态网络当前状态的影响，会随着这些过去输入在时间上距离当前状态越来越远而减弱。实际而言，这意味着网络的每个状态都仅是网络有限个先前输入的函数，使得

$$\begin{aligned}\vec{s}(t) &= \mathcal{E}_\infty\Big(\vec{s}(t-1),\vec{s}(t-2),\dots\Big)\\ &\approx \mathcal{E}\Big(\vec{x}(t-1),\vec{x}(t-2),\dots,\vec{x}(t-\tau_{MC})\Big).\end{aligned} \tag{9.3}$$

其中 τ_{MC} 是衡量储备池记忆时长的指标，并且在某种程度上与记忆容量（MC）有关。特定储备池计算机的记忆容量可以通过多种方式量化，不过，Jaeger [62]提出的原始方法最为常用

$$MC = \sum_{\hat{\tau}=0}^{\hat{\tau}_{max}} \text{correlation}\Big(\{x(t)\}_{t=1}^{T-\hat{\tau}}, C_{out}\{\vec{s}(t)\}_{t=\hat{\tau}+1}^{T}\Big) \tag{9.4}$$

其中 x 是标准高斯噪声信号， C_{out} 是用于预测该信号第 $\hat{\tau}$个先前值的学习得到的参数。对于足够大的$\hat{\tau}_{max}$，记忆容量会收敛，图 9.1 和图 9.2 展示了这一点。

此时，我们可以注意到输入信号 $\vec{x}$ 的储备池状态表示与嵌入生成的表示之间的相似之处。事实上，最早提及储备池计算机可充当嵌入机器的，是 Jaeger 的原始研究 [62]。然而，真正的突破来自 Hart 等人 [65] 的研究，他们提出了证明，表明回声状态网络确实以正概率生成底层信号的嵌入。Hart [66]后来进一步证明，连续时间线性储备池计算机几乎必然也能生成嵌入。

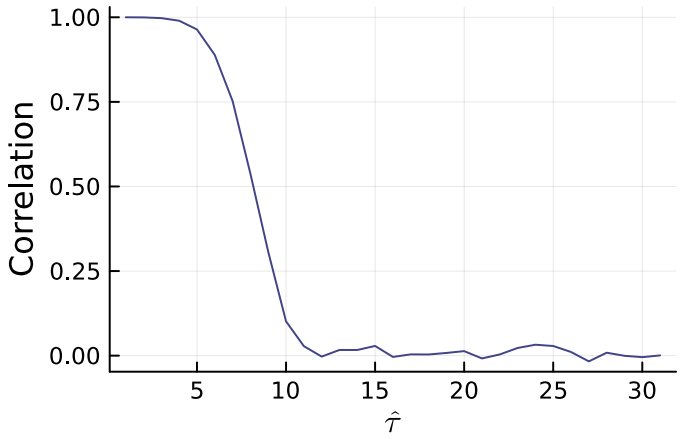


图 9.1 通过回忆性能演示渐忘记忆属性。

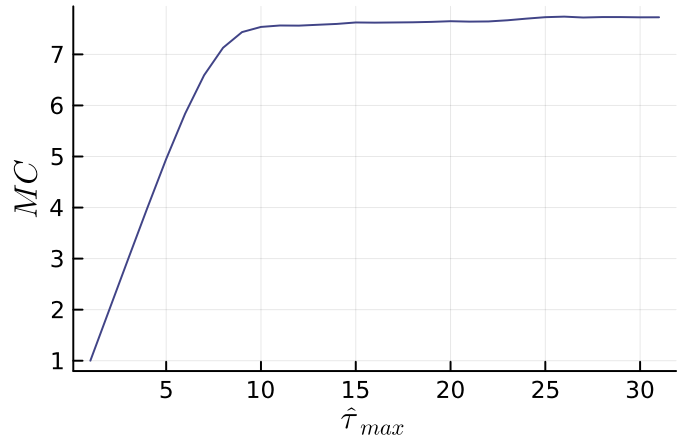


图 9.2 记忆容量的收敛。

9.2 储备池设计

虽然到目前为止我们讨论的是循环神经网络，但最简单的储备池计算机是一种输入输出自动机，它利用一个动力系统（称为储备池）作为信息处理的核心单元，并具有回声状态。时间序列 $\vec{x}(t)$作为一种驱动信号输入动力系统，系统随后按照以下方式自然演化

$$\frac{d\vec{s}}{dt}(t) = f_{RC}(\vec{s}(t), \vec{x}(t), V) \tag{9.5}$$

其中， f 描述动力学， V 则收集储备池的固定 参数。随后，可以记录该系统的状态 $\vec{s}(t)$（称为激活状态），并将其用于训练。

使用固定动力系统可得出两个推论：

1. 唯一需要进行的训练，是从 $\vec{s}(t)$ 到我们期望的输出 $\vec{y}(t)$。2. 构建良好模型的任务可简化为选择一个有用的动力系统。

乍看之下，第二点似乎令人望而生畏，但实际上，有许多系统都可以考虑。

9.2.1 回声状态网络（*ESN*）

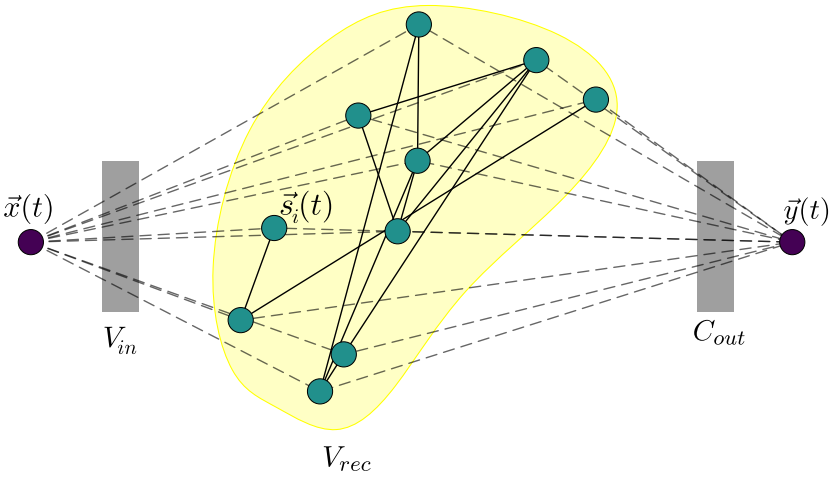


图 9.3 回声状态网络的可视化。

研究最深入的储备池计算机是回声状态网络（ESN），我们可以将其视为固定的循环神经网络（RNN）。在这种情况下，我们现在使用离散时间，因此可以将演化函数改写为

$$\vec{s}(t+1) = \tanh(V_{in}\vec{x}(t) + V_{rec}\vec{s}(t) + V_{bias}) \tag{9.6}$$

注意，这里选择了 Sigmoid 激活函数，这在回声状态网络中很常见。

在回声状态网络中，固定参数 V 是随机生成的：

- 输入矩阵 $V_{in} \in \mathbb{R}^{n \times k}$ 通常从适当的分布中随机抽取元素（通常为均匀分布或高斯分布）。
- 储备池邻接矩阵 $V_{rec} \in \mathbb{R}^{k \times k}$ 最初生成为具有平均度 d 的 Erd˝os-R´enyi 随机图的邻接矩阵，随后进行缩放，使其最大绝对特征值（或谱半径）为 ρ。
- 偏置矩阵 $V_{bias} \in \mathbb{R}^{k}$ 以类似于 V_{in}的方式生成。

因此，对动力学的控制及其对性能产生的影响取决于我们对超参数 k、d 和 ρ的选择。由于我们并不寻求训练这些超参数，因此依靠一系列启发式方法来选择它们。以下是对这些启发式方法的总结；Luko˘sevi˘cius [67]的工作对如何实现回声状态网络提供了更为详尽的介绍。

储备池规模 k
关于储备池规模的一般直觉是，越大越好。不同于传统的循环神经网络，在传统循环神经网络中使用大量节点会导致训练负担显著增加，而固定的回声状态网络特性可以缓解这一问题。因此，通常会将 k 设置为在所使用硬件条件下合理可行的最大值。储备池规模为 k 、达到数百或数千的回声状态网络十分常见。

平均度 d
在相关的超参数中， d 的重要性最低。储备池计算领域普遍认为， d 应足够小，以使网络保持稀疏。典型值约为网络规模 k 的 5%，很少超过 10%。更一般地说，储备池计算领域的研究通常表明，边的选择相对而言并不重要。其他网络结构（例如格点或环形网络）已被证明与随机网络具有类似的功能，因此人们通常不会过多考虑这些结构问题，而是更加关注关键的超参数 ρ。

谱半径 ρ
谱半径也称为内部增益，它是我们控制储备池动力学并确保回声状态存在的机制。通常可以将 ρ 视为控制储备池的 记忆 ，因此可以考虑以下极端情况：

- $\rho = 0$ 会移除回声状态网络中的所有记忆（每个状态仅是输入的函数）
- $\rho \gg 1$ 回声状态网络的记忆占据主导，状态因此变得没有信息量。
- $\rho \approx 1$ 输入信号的驱动与回声状态网络的记忆之间达到了一种平衡。

随着我们增大 $\rho > 1$ ，会逐渐接近一个临界点；越过该点后，回声状态网络的记忆将占据主导。由于观察到的动力学发生了变化，这个临界点通常被称为稳定性边缘或混沌边缘。因此，一种常见的启发式方法是将 $\rho = 1$设为……，以确保记忆能力，同时避免破坏性动力学，不过

研究表明，将 ρ 设得更接近混沌边缘往往有助于提升性能。选择 ρ 的方法仍是一个活跃的研究领域。

其他参数
在回声状态网络演化方程中，加入更多超参数的变体并不少见，其形式如下：

$$\vec{s}(t+1) = (1-\alpha)\vec{s}(t) + \alpha \tanh(\eta V_{in}\vec{x}(t) + V_{rec}\vec{s}(t) + V_{bias}) . \tag{9.7}$$

这引入了以下额外的超参数：

- α 是一种额外机制，可用于人为增加储备池中的记忆量。对于时间尺度较大的信号，这可以使储备池更好地提取动力学特征，不过也可以通过其他超参数实现类似效果。
- η 是一个用于控制输入幅值的参数。由于采用了双曲正切函数 tanh(·) 作为 sigmoid 激活函数，超出范围 $[-1, 1]$ 的信号值会被压缩，从而导致信息丢失。通过在输入前对输入信号 $\vec{x}$ 进行归一化，可以获得与引入超参数 η 类似的效果。

9.2.2 其他储备池计算机

液态状态机（LSM）
另一类基于循环神经网络的储备池计算机是 LSM，它采用脉冲神经网络，而非结构更简单的回声状态网络。就结构设计而言，许多适用于回声状态网络的相关直觉同样适用于 LSM：随机生成矩阵、高维网络，以及通过控制超参数来确保动力学。

单延迟节点
Appeltant 等人 [68] 没有采用时空回声状态网络，而是提出了一种纯时间方法：通过时间复用输入，并将其馈入带有延迟的单个节点。常见的延迟系统是 Mackey-Glass 振荡器，因此

$$\frac{ds}{dt}(t) = -s(t) + \frac{\eta\,(s(t-\tau) + \gamma x(t))}{1 + (s(t-\tau) + \gamma x(t))^p} \tag{9.8}$$

其中 η， τ， γ 和 p 均为超参数。请注意，由于采用了时间复用，输入和输出都是标量序列，因此高维状态是通过在不同延迟下读取 $s(t)$ 得到的。

基于智能体的储备池计算机
我们可以不采用节点，而是考虑大量代理人，并测量它们对外部驱动的行为。这正是 Lymburn 等人研究的重点。 [69]其中

采用了一个基于智能体的模型，将受外部捕食者影响的群体作为储备池计算机。这类模型相对而言探索得较少，但此前的荣誉项目也曾研究过一个领地性捕食者模型，其中单个猎物充当驱动信号。

物理储备池计算
储备池计算（RC）的吸引力之一在于，我们可以使用物理动力系统，而非数字系统，大幅提升状态生成的速度。为此，已出现许多物理模型：

1. 实际的“液体”状态机 2. 光子
电路 3. 自旋电子学 4. 软体机
器人

这里不详细介绍物理储备池计算，但 Nakajima 所著的书籍 [70] 涵盖了许多当前的研究领域。

9.3 储备池计算机训练

由于动力系统是固定的，唯一需要调节的参数是连接激活状态与期望输出 (C_{out}) 的参数。

9.3.1 时间序列预测

储备池计算机最常见的学习任务是时间序列预测，即尝试将输入信号 $\vec{x}(t)$ 映射到某个期望输出信号$\vec{y}(t)$。由于 C_{out} 与储备池的动力学之间没有反馈，因此可以通过线性回归来处理训练过程，使得

$$\hat{y}(t) = C_{out}\vec{s}(t) \tag{9.9}$$

其中，$\hat{y}(t)$ 是对目标信号 $\vec{y}(t)$ 得出的近似结果。通常， C_{out} 的确定通过岭回归（或Tikhonov正则化）来处理，岭回归通过求解最小二乘问题来确定相关参数值

$$\min_{C_{out}} \left(\sum_t \left(\vec{z}(t) - C_{out}\vec{s}(t) \right)^2 + \beta \sum_i \left(C_{out,i} \right)^2 \right) \tag{9.10}$$

其中， $\vec{z}(t)$ 是我们的训练目标，用于区分训练中使用的信号与测试中使用的信号 $\vec{y}(t)$。参数 β 称为正则化参数，用于防止训练步骤中过拟合。该问题的闭式解

上述最小化问题可以写为

$$C_{out} = (\vec{s}\vec{s}^{\top} + \beta I)^{-1}\vec{s}\vec{z}\,. \tag{9.11}$$

9.3.2 系统仿真

或者，我们可能希望构建一个能够模拟并预测原始系统的模型。这项任务只是时间序列预测任务的延伸：首先求解方程 9.10 中的回归问题，其目标信号对应于输入信号的单步预测（即$\vec{z}(t) = \vec{x}(t+1)$），使得

$$\vec{x}(t+1) = C_{out}\vec{s}(t)$$

与传统循环神经网络的情况相同。随后，可以将这一近似结果反馈到储备池计算机中，以替代输入信号，从而实现对系统的自由运行预测。对于回声状态网络，新的演化方程为

$$\vec{s}(t) = \tanh(V_{in}C_{out}\vec{s}(t-1) + V_{rec}\vec{s}(t-1) + V_{bias})\,. \tag{9.12}$$

这里仍有一个问题：自由运行预测时，我们应如何选择储备池的初始状态。理想情况下，在自由运行开始之前应有一段足够长的序列，使储备池状态能够收敛。若缺少这样的序列，我们就会遇到无法落入吸引子的问题。不过，已有方法可解决这一问题，例如训练另一个储备池计算机，将输入空间映射到输出空间（即 $\vec{z}(t) = \vec{x}(t)$）。

9.3.3 时间序列分类与储备池时间序列分析

时间序列分类是储备池计算得到广泛应用的另一项任务。分类并非关注时间点之间的关系，而是试图将整个时间序列与单个分类或标签联系起来。这里的主要区别在于，不再是根据激活状态训练输出序列，而是将激活状态汇总为某种特征 H。随后可以针对不同时间序列生成大量特征，并据此训练分类器。典型的分类器包括支持向量机和随机森林。

人们考察了许多不同的特征，其中一些最常见的特征包括：

$$H_{RC}^{last} = \vec{s}(T) \tag{9.13}$$

$$H_{RC}^{mean} = \frac{1}{T}\sum_{t}\vec{s}(t) \tag{9.14}$$

$$H_{RC}^{R} = C_{out} \text{ with } \vec{z}(t) = \vec{x}(t+1) \tag{9.15}$$

这些特征本质上都旨在量化底层系统的某些方面，从而帮助我们区分来自不同系统的信号。

虽然这一方法是在监督式机器学习的背景下发展起来的，但通过某种特征来量化动力系统的某些方面，这一想法对你来说并不陌生。这正是不变测度背后的核心思想；事实上，储备池计算机已用于近似某些此类测度，以完成时间序列分类任务。虽然分类 侧重于监督学习任务，储备池时间序列分析则是一个更广泛的领域，关注如何从储备池计算机生成的嵌入（即储备池空间嵌入）中提取信息 [71]。

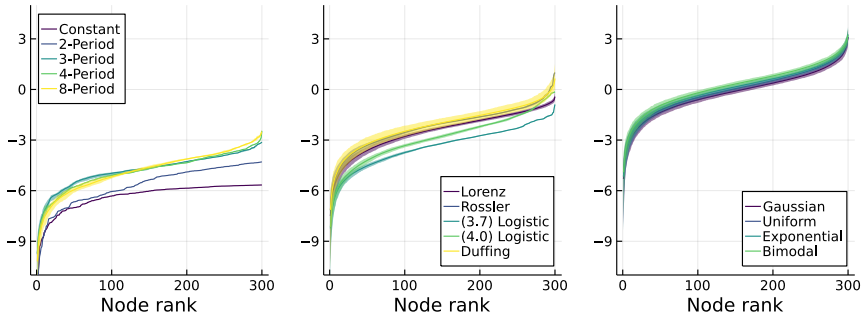


图 9.4 由一步预测生成的周期、混沌和随机信号（从左到右）的读出幅值曲线对比。

储备池时间序列分析仍是一个新兴领域，目前许多方法与分类方法相互重叠，不过已有一些工作致力于开发专门的不变测度。一个例子是储备池计算复杂度（RCC） [72]。RCC 背后的想法是，与简单系统相比，储备池在预测复杂系统时会使用更大比例的节点，而这一比例的增加可通过正则化条件下读出向量的幅值 C_{out}来量化。图 9.4 展示了这一现象。曲线的行为可以通过近似计算曲线下面积来概括

$$H_{RC}^{cplx} = \log\left(\sum_{i=1}^{k-1}\frac{\hat{C}_{out,i} + \hat{C}_{out,i+1}}{2}\right) \tag{9.16}$$

9.3 训练储备池计算机，其中 C_{out} 是 C_{out} 的按秩排序的变化。

第10章 递归图

回顾一下，我们可以用回归行为来描述动力学。原则上，分析复现时间和状态所编码的信息，应当与分析连续流中的信息等价。复现（定义某种邻域后）比完整的状态动力学更易于分析，并且有助于快速获得对系统的基本理解。那么，如何构造复现的直观表示呢?

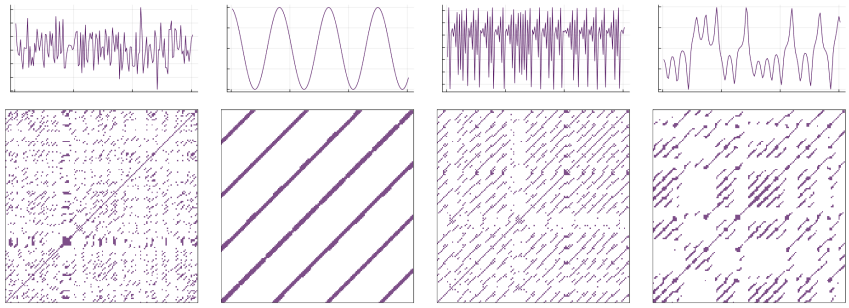
一种可视化和量化复现的方法是使用递归图 $RP \in \mathbb{R}^{T\times T}$，使得

$$R_{ij} = \begin{cases} 1, & d(\vec{x}(i), \vec{x}(j)) < \epsilon \\ 0, & \text{otherwise} \end{cases}$$

其中，$d(\cdot)$ 是某种（不一定是欧几里得）距离度量，ϵ 是复现阈值。递归图（理论上）包含时间序列动力学的所有“有趣”细节。因此，RP 中的某些模式和结构可以与动力学性质联系起来:

- 对角线编码了递归状态保持一致的持续时间，其中线段长度等于时间长度。线段长度的分布还有助于刻画系统的混合行为。值得注意的是，这些对角线的长度与李雅普诺夫时间（$\frac{1}{\lambda_{max}}$）成正比。
- 垂直线 可以突出显示静止/稳态，或层流性很强的状态。
- 垂直空隙或复现时间 提供有关状态何时复现的信息，规律性则表明存在周期动力学。

通过绘制 RP（例如热图），可以对上述结构进行定性分析，从而了解底层系统，图10.1 提供了一个示例。



(a) Random (b) Periodic (c) Chaotic Flow (d) Chaotic System

图10.1 复杂度各异的信号 *RP*示例。信号分别对应随机噪声、周期正弦波，以及混沌洛伦兹系统和混沌逻辑斯蒂映射的观测值（子图(a)至(d)依次对应）。

10.1 构造注意事项

此时显然，在生成递归图之前需要做出多项选择，而做出这些决定通常是此类分析中最困难的部分。

距离度量 $||\cdot||$
处理数值时间序列数据时，最常用的距离度量是曼哈顿距离

$$d_1(\vec{a},\vec{b}) = \sum_i |a_i - b_i|\,, \tag{10.1}$$

欧几里得距离

$$d_2(\vec{a},\vec{b}) = \sqrt{\sum_i (a_i - b_i)^2} \tag{10.2}$$

以及上确界（或最大）距离

$$d_\infty(\vec{a},\vec{b}) = \max_i(|a_i - b_i|)\,. \tag{10.3}$$

在这些方法中，使用上确界范数在生成 *RP*s 时能够显著降低计算成本，同时还可解析地推导出某些结果，例如关联维数。

然而，这种对范数选择的自由，使复现的思想得以应用于结构各异的序列。例如，可以将克罗内克δ函数作为距离度量，在符号序列上构造 *RP*s [73]。近年来，编辑距离也被用于在事件数据序列上定义 *RP*s，这是一种

持续得到应用和研究的方法 [74]。

复现阈值 ϵ
通常会根据对预期距离的某种启发式估计来选择并固定阈值 ϵ：阈值要足够大，使复现能够发生，但又不能大到让所有点都发生复现。我们也可以不采用固定距离，而是根据固定的复现率来选择 ϵ，即选择一个预期会发生复现的点所占的百分比。还存在其他启发式和数值方法，可用于估算这个固定阈值；这些方法在分辨率和抗噪性方面各有取舍。

嵌入 $\vec{x}$
理想情况下，我们会在完整的状态空间轨迹上构建 RP；然而，正如本单元所讨论的，我们通常并没有掌握所有这些信息。更常见的做法是构建嵌入，并使用嵌入空间来确定哪些状态会复现。有趣的是，关于这一步的重要性，目前仍存在争论，因为研究表明，基于观测值生成的 RP在没有嵌入的情况下，与经过嵌入后生成的结果在定性上相似；许多重要结果（例如推导关联维数）也可以在不进行嵌入的情况下得出。尽管如此，学界普遍认为，如果能够在分析 RP 之前进行嵌入，就应该这么做！

10.2 复现量化分析

递归图对系统行为进行定性评估，而复现量化分析（RQA）则将递归图中的几何属性形式化并量化。这些信息可用于变点检测、确定复杂度度量或检测周期—混沌转变等任务。总体而言，RQA 方法遵循相似的思路：概括 RP 中某种定性特征的经验分布，然后根据该分布计算某种定量测度。考察对角 H^{diag} 线、垂直 H_{vert} 线和复现时间 H_{recc} 的分布，是最常见的定性特征。

10.2.1 点密度

这里首先考察的几个特征仅涉及 RP 中点的分布，而非线结构。对于给定的距离阈值，第一个特征是复现率

$$RR = \frac{1}{T^2} \sum_{i=1}^{T} \sum_{j=1}^{T} RP_{i,j}, \tag{10.4}$$

它用于测量递归点的密度。若排除主对角线，这就对应于我们对相关和的定义。一个类似的值

$$N_{avg} = \frac{1}{T}\sum_{i=1}^{T}\sum_{j=1}^{T} RP_{i,j}, \tag{10.5}$$

也可以计算相空间中各点的平均递归数（或邻居数）。

给定固定的复现率后，我们可以反向确定固定的距离阈值

$$\epsilon_{RR} = \max_{i,j}\left(d(\vec{x}(i),\vec{x}(j))\middle| RP_{i,j}=1\right). \tag{10.6}$$

10.2.2 对角线

对角线直方图定义为

$$\mathcal{H}_{diag}(l) = \sum_{i,j=1}^{T}(1-RP_{i-1,j-1})(1-RP_{i+l,j+l})\prod_{k=0}^{l-1} RP_{i+k,j+k}, \tag{10.7}$$

其对应的实验分布为

$$\mathcal{P}_{diag}(l) = \frac{\mathcal{H}_{diag}(l)}{\sum_{\hat{l}=l_{\min}}^{T}\mathcal{H}_{diag}(\hat{l})}, \tag{10.8}$$

请注意， l_{min} 应与针对相应特征所考虑线段的最小长度一致。对角线特征可用作动力系统中混沌—有序转变的指标。以下详细介绍从 H^{diag} 得出的一些常见定量特征。

确定性
不相关、弱相关、随机或具有混沌行为的过程不会产生对角线，或只产生较短的对角线。确定性过程会产生较长的对角线，且孤立的递归点较少。因此，确定性（或可预测性）的测度是构成长对角线（$l > l_{min}$）的递归点占所有递归点的比例（不包括切向运动和主对角线）。

$$DET = \frac{\sum_{l=l_{min}}^{T} l\mathcal{H}_{diag}(l)}{\sum_{l=1}^{T} l\mathcal{H}_{diag}(l)}. \tag{10.9}$$

平均对角线长度
平均对角线长度与轨迹发散有关，因此也是一种

平均预测时间的测度,

$$L_{avg} = \frac{\sum_{l=l_{min}}^{T} l\mathcal{H}_{diag}(l)}{\sum_{l=l_{min}}^{T} \mathcal{H}_{diag}(l)}. \tag{10.10}$$

最长对角线

最长对角线与相空间轨迹的指数发散有关，因此与系统的李雅普诺夫时间成正比。一个相关的测度是散度的倒数，因此它与最大李雅普诺夫指数成正比。轨迹片段发散得越快，对角线就越短，散度也越大,

$$L_{max} = \max(l_i|\mathcal{H}_{diag}(l_i) > 0) \propto \frac{1}{\lambda_{max}}, \tag{10.11}$$

$$DIV = \frac{1}{L_{max}}. \tag{10.12}$$

递归熵

对角线分布的 Shannon 熵,

$$ENTR = -\sum_{l=l_{\min}}^{T} \mathcal{P}_{diag}(l) \log_2 \left(\mathcal{P}_{diag}(l)\right), \tag{10.13}$$

反映了递归图在对角线方面的复杂度。例如，不相关噪声和周期信号的 $ENTR$ 值往往都较低。

10.2.3 垂直线

垂直线直方图如下:

$$\mathcal{H}_{vert}(l) = \sum_{i,j=1}^{T} (1 - RP_{i,j-1})(1 - RP_{i,j+l}) \prod_{k=0}^{l-1} RP_{i,j+k}, \tag{10.14}$$

相应的实验分布 $\mathrm{P}^{diag}(l)$ 的计算方式与对角线情形相同。垂直线主要表示轨迹在某个其他时间经过该邻域附近（即切向运动）。然而，研究还表明，垂直线也可指示层流状态的存在；相应的测度不仅能够检测混沌-有序转变，还能检测混沌-混沌转变。下面详细介绍由 H_{vert} 得出的一些常见定量特征。

层流性
层流性可按照与 DET 类似的方式计算，表示为:

$$LAM = \frac{\sum_{l=l_{min}}^{T} l\mathcal{H}_{vert}(l)}{\sum_{l=1}^{T} l\mathcal{H}_{vert}(l)}, \tag{10.15}$$

表示某些状态的出现，这些状态在一段时间内保持不变或变化缓慢，称为层流状态。这些状态是间歇性的典型特征。

滞留时间Th
e 的平均值 l垂直结构的长度或滞留时间，

$$TT = \frac{\sum_{l=l_{min}}^{T} l\mathcal{H}_{vert}(l)}{\sum_{l=l_{min}}^{T} \mathcal{H}_{vert}(l)}, \tag{10.16}$$

估算系统处于特定状态的平均时长，可用于检测混沌-混沌转变，因此也可用于检测间歇性。

最长垂直线与熵
最大垂直线 V_{max} 和垂直线结构熵 $VENTR$ 也有类似的测度，不过使用频率相对较低。其解释与对角线的情况类似，其中 V_{max} 反映系统被困在特定状态中的最长时间，而 $VENTR$ 则再次表示复杂性的一种测度。

10.2.4 复现时间

复现时间的分布由回归结构之间垂直距离的分布给出

$$\mathcal{H}_{recc}(l) = \sum_{i,j=1}^{T} RP_{i,j-1} RP_{i,j+l} \prod_{k=0}^{l-1} (1 - RP_{i,j+k}), \tag{10.17}$$

同样对应一个 P_{recc}。复现时间与点态维数有关，点态维数是系统的不变量。具体而言，点态维数D_p 与平均回归时间有关，

$$r_{avg} = \frac{\sum_{l=l_{min}}^{T} l\mathcal{H}_{recc}(l)}{\sum_{l=l_{min}}^{T} \mathcal{H}_{recc}(l)}, \tag{10.18}$$

由关系 $r_{avg} \sim \epsilon^{D_p}$ 给出。

对于复现时间，也可以像对角线和垂直线那样计算长度比、最大长度和熵等类似特征。总体而言，研究表明，复现时间度量能够检测动力系统中的多种周期性和混沌转变，包括涉及奇异非混沌吸引子的转变。

第11章
动力网络

到目前为止，我们已经广泛研究了处于规范形式中的孤立动力系统，

$$\dot{x} = f(x), x \in \mathbb{R}.$$

然而，如果允许两个或更多个体动力系统相互作用，会发生什么？在极端情况下，可以考虑一个相互作用的动力系统网络（不一定是全连接的）。这种系统通常被称为**动力网络**，因为它指的是沿某种拓扑结构（即网络）发生的动力学（时间行为），而后者编码了个体动力系统之间某种形式的信息流。

11.1 耦合与同步

11.1.1 耦合函数

考虑以下形式的两个动力系统

$$\begin{aligned}\dot{x} &= f_1(x),\\ \dot{y} &= f_2(y) + G(x, y),\end{aligned}$$

其中 $x,y \in \mathbb{R}^m$ 和 $G\colon \mathbb{R}\times\mathbb{R} \to \mathbb{R}$。这两个系统都符合标准形式（$\dot{x} = f(x)$），主要区别在于第二个系统中增加了一个相互作用项。该项 G 称为**耦合函数**，用于描述一个系统影响另一个系统的机制。上述情况对应单向耦合，其中 x 影响 y，但反之则不然（有时分别称为驱动系统和受驱系统）。当耦合施加于直接状态变量 x, y 时，这种情况称为振幅耦合。如果将动力系统的研究范围限定为相位可以定义的振荡器，那么以这些变量实现的耦合同样称为相位耦合（见 11.1.2）

耦合函数可分为两部分来描述：（1）强度，以及（2）形式，

$$G(x, y) = Kq(x, y)$$

前者通常通过常数乘数 K 介导，其中较大的值表示更强的耦合。耦合函数 $q(x, y)$ 的形式描述了相互作用的性质，并且会以不可预测的方式显著影响耦合系统的动力学 [75]。最常见的两种耦合函数形式是直接耦合 $q(x, y) = q(y)$，表示单向影响；以及扩散耦合 $q(x, y) = q(y - x)$。

Kuramoto 在研究相位耦合振荡器时推广的扩散耦合情形是一个特别有趣的案例 [76]。扩散耦合的数学形式表示这样一种情况：相互作用的强度取决于两个状态之间的差异。通常会将其设置为差异越大，对应的耦合越强。$q'(0) < 0$ 的情况也称为耗散耦合，指的是推动系统趋于同一状态的耦合力。相反，$q'(0) > 0$ 的情况则称为排斥耦合，原因相同。 [75, 76]

11.1.2 二体系统与同步

两个系统之间存在足够强的扩散耦合，可能使系统状态收敛，这种现象称为同步。这样一来，两个系统便具有相同的状态，并随时间共同演化。为说明这一点，我们考虑以下双向耦合的系统对，

$$\begin{aligned} \dot{x} &= f(x) + Kg(y - x), \\ \dot{y} &= f(y) + Kg(x - y). \end{aligned}$$

为了更好地分析两个系统之间的同步性程度，我们将分析限制在辅助变量 $z = x - y$ 上，它描述了两个系统之间的偏差。因此，我们可以写出描述 z 演化的以下方程，

$$\begin{aligned} \dot{z} &= \dot{x} - \dot{y} \\ &= f(x) - f(y) + K[g(-z) - g(z)]. \end{aligned}$$

如果假设 $g(z) > 0$ for $z > 0$,and $g(-z) = -g(z)$ (or at least $g(-z) \propto -g(z)$，那么状态差异的演化可简化为，

$$\dot{z} = f(x) - f(y) - 2Kg(z).$$

分析这个约化系统，我们可以观察到在 $z^* = 0$ 处存在一个不动点（即 $x = y$）。此外，尽管我们无法保证 $f(x) - f(y)$ 的符号对于

所有时间都成立，但显然，当 K 足够大且存在耗散耦合（$g'(0) < 0$）时，该不动点 z^* 是稳定的。

为说明同步的起始，考虑以下一对耗散耦合的 R¨ossler 振荡器 [77]，

$$\begin{aligned} \dot{x}_1 &= -y_1 - z_1 + K(x_2 - x_1) & \dot{x}_2 &= -y_2 - z_2 + K(x_1 - x_2) \\ \dot{y}_1 &= x_1 + a y_1 & \dot{y}_2 &= x_2 + a y_2 \\ \dot{z}_1 &= 0.4 + z_1(x_1 - 8.5) & \dot{z}_2 &= 0.4 + z_2(x_2 - 8.5) \end{aligned}$$

常规的 R¨ossler 吸引子由围绕原点处一个不稳定不动点旋转的轨道组成，同时流形在 z 方向的一侧发生折叠（见图 11.1）。这种折叠会导致混沌行为。当 $a = 0.165$ 时，系统动力学处于相位相干状态。“相位相干”一词指的是，轨迹的旋转频率在吸引子的所有部分大致保持恒定。因此，任意两条轨迹之间的差异应仅在于它们的相位 ϕ 和半径（相对于 xy 平面）。为简单起见，我们可以将轨迹上任意一点的相位估算为 $\phi = atan(\frac{y}{x})$。

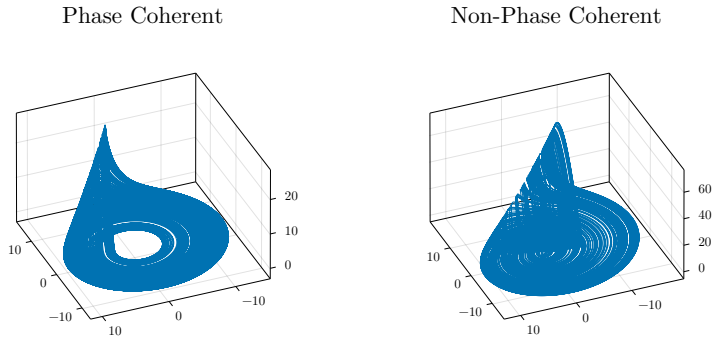


图 11.1 相位相干 $a = 0.165$（左）和非相位相干 $a = 0.265$ 状态下的 R¨ossler 吸引子。

当 $K = 0$ 时，两个振荡器完全没有相互作用，其相位根据各自的初始条件而有所不同。然而，随着 K 逐渐增大，振荡器的耦合动力学会经历几个不同的变化阶段（见图 11.2 和 11.3）。在低至中等耦合强度下，会出现相位同步：振荡器的相位开始收敛，但振幅仍然不同。稍微增大 K 后，系统开始呈现间歇同步（不要与间歇性混淆），此时两个振荡器会相互收敛，并不时短暂同步。这表现为耦合 x 分量的偏差出现阶段性塌缩。最后，当 K 足够大时，便会出现完全同步，两条轨迹永久收敛。

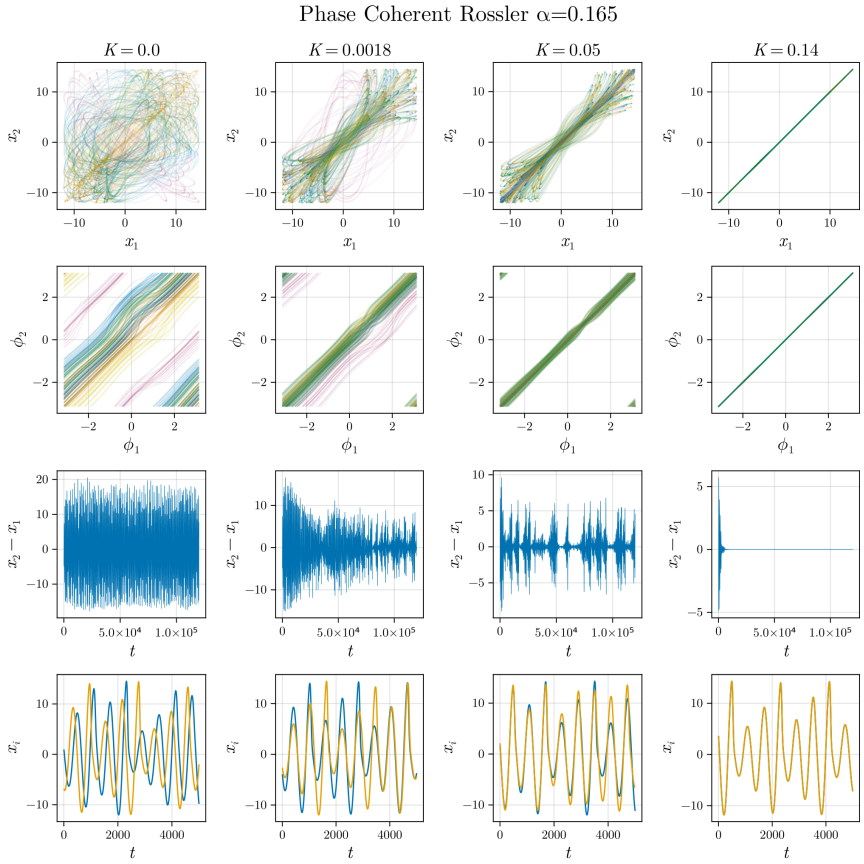


图 11.2 两个耦合的相位相干（$\alpha = 0.165$）R¨ossler 振荡器，耦合强度 K 从左到右逐渐增大：异步（无耦合）、相位同步、间歇同步，随后在 $K = 0.14$ 时达到完全同步。注意 $\phi_1 - \phi_2$ 图中显示的对角斜率，这是相位相干的特征，此时所有动力学都会自然地遵循单一的主导频率。

11.2 网络上的动力学

11.2.1 振子网络与奇美拉态

振荡器之间的耦合可以轻松扩展到包含任意 N 个（通常相同的）振荡器。耦合的具体连接方式也可以直接调整，不必采用全连接形式。通常可以用以下方程表示为网络形式：

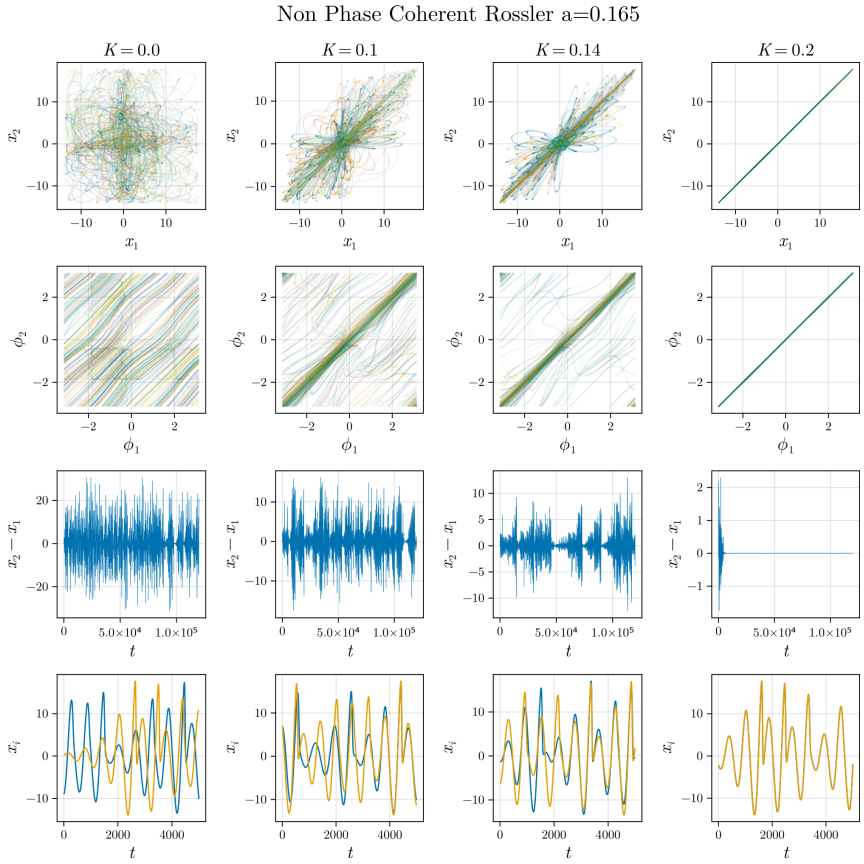


图 11.3 非相位相干状态下的两个耦合 R¨ossler 振荡器（$\alpha = 0.265$）。缺乏相位相干意味着，要达到与相位相干状态下类似的同步行为，需要更强的耦合 K。

$$\dot{x}_i = f(x_i) + K \sum_{j=1}^{N} A_{ij} G(x_i, x_j),$$

其中， A 是底层网络的邻接矩阵，网络中的节点表示各个振荡器，边表示一对振荡器之间的耦合。如果A 是对称的，耦合就是双向的。

此前，人们认为耦合的同一相位振荡器的行为相对简单，只要耦合足够强，就会发生相位同步。人们推测，这些结果也适用于振子网络：耦合足够强时，整个网络会完全同步。然而，Kuramoto 等人 [76]证明，对于某些网络配置和初始条件，可能会出现有序簇与无序簇共存的情况——

因此，网络中的一部分振荡器会相位锁定，而其余振荡器仍保持异步。这两种不协调状态共存的现象被称为“奇美拉态”，名称源自神话中的希腊神兽。

全局耦合振荡器网络中奇美拉态的存在性和稳定性已从多个角度进行了研究。聚焦于由两个簇组成的典型 Kuramoto 振荡器网络 σ， σ'，

$$\frac{d\theta_i^\sigma}{dt} = \omega - \mu\langle\sin(\theta_i^\sigma - \theta_j^\sigma + \alpha)\rangle_{j\in\sigma} - \nu\langle\sin(\theta_i^\sigma - \theta_j^\sigma + \alpha)\rangle_{j\in\sigma'},$$

其中 μ 和 ν 分别表示簇内和簇间耦合强度，Ott 和 Antonsen 证明，无限大网络的宏观动力学可以约化为一组有限维常微分方程 [78]。对于由两个簇构成的系统，还可以进一步化为 2 对方程（下方展示的是单个簇的情形）：

$$\begin{aligned}
0 &= \dot{\rho}_X + \frac{\rho_X^2 - 1}{2}[\mu\rho_X\cos\alpha + \nu\rho_Y\cos(\phi_Y - \phi_X - \alpha)],\\
0 &= -\rho_X\dot{\phi}_X + \rho_X\omega - \frac{1+\rho_X^2}{2}[\mu\rho_X\sin\alpha + \nu\rho_Y\sin(\phi_X - \phi_Y + \alpha)],
\end{aligned}$$

其中 X， Y 表示两个簇， ρ_X， ρ_Y 是各簇的序参量（同步程度）， ϕ_X， ϕ_Y 是它们的相位 [79]。通过将上述方程限制在流形 $\rho_X = 1$ 上（即簇 X完全同步），并应用坐标变换 $r = \rho_Y, \psi = \phi_X - \phi_Y$，可以研究奇美拉态的动力学，

$$\begin{aligned}
\dot{r} &= \frac{1-r^2}{2}[\mu r\cos\alpha + \nu\cos(\psi - \alpha)],\\
\dot{\psi} &= \frac{1+r^2}{2r}[\mu r\sin\alpha - \nu\sin(\psi - \alpha)] - \mu\sin\alpha - \nu r\sin(\psi + \alpha).
\end{aligned}$$

这组约化后的方程在 $r = 1$ 处有一个平凡不动点（完全同步态），并且在 $r < 1$ 处还有一个非平凡不动点（奇美拉态）。感兴趣的读者可参阅 [80]，了解更详细的讨论。

11.2.2 FitzHugh-Nagumo 奇美拉模型

为了通过数值方法展示一个呈现奇美拉态的系统，我们参考了 [81, 82] 提供的 FitzHugh-Nagumo 振荡器示例。简而言之，FitzHugh-Nagumo 振荡器是广为人知的 Hodgkin-Huxley 模型的一种简化形式，用于通过以下方程描述神经元尖峰动力学，

$$\dot{u} = u - \frac{u^3}{3} - v + I,$$
$$\epsilon \dot{v} = u + a - bv,$$

其中 I, a 和 b 为常数，状态变量 u 和 v 描述神经元中的激活和抑制机制。对于奇美拉态，我们考虑一个由 N 个非局部耦合的 FitzHugh-Nagumo（FHN）振荡器组成的网络，

$$\epsilon \dot{u_k} = u_k - \frac{u_k^3}{3} - v_k + \sigma \sum j = 1^N A_{kj}[b_{uu}(u_j - u_k) + b_{uv}(v_j - u_k)],$$
$$\dot{v_k} = u_k + a + \sigma \sum_{j=1}^{N} A_{kj}[b_{vu}(u_j - v_k) + b_{vv}(v_j - v_k)],$$

其中，振荡器由 k 编号，$\epsilon > 0$ 是一个较小的参数，用于提供时间尺度分离，而 σ 是耦合强度。此外，振荡器的耦合既包含直接的 $u - u$ 和 $v - v$，也包含交叉耦合 $u - v$。

非局部耦合是指振荡器之间的相互作用，而这些振荡器并非直接的空间邻居。在最简单的情况下，我们考虑按照一维环形拓扑排列的振荡器，其耦合项具有以下形式：

$$\sigma \sum_{j=1}^{N} A_{kj} = \frac{\sigma}{2R} \sum_{j=k-R}^{k+R} 1.$$

其中，R 是每个节点所连接的邻居数量。一个便于追踪的量是耦合 $r = \frac{R}{N}$。$r = \frac{1}{N}$ 的值对应于最近邻（局部）耦合，而 $r = 0.5$ 则对应于全连接耦合/全局耦合（即完全连接的网络）。r 的中间值对应于非局部耦合。

对该模型进行数值模拟得到图 11.4 所示的结果。值得注意的是，在精心选择参数值的情况下，具有非局部耦合的 FHN 振子环形网络会产生一个由完全同步的振荡器组成的连续簇，同时与另一组异步（低阶）振荡器共存。

11.3 意见动力学

11.3.1 社会物理学的动力网络方法

动力网络模型的另一个例子虽然有所不同，但同样具有说明性，来自社会物理学和意见动力学。意见动力学与共识形成的研究是一个活跃的研究领域，旨在对群体行为中人类互动的驱动力和机制进行建模与理解。一般而言，意见动力学的网络形式如下所示。设一个意见网络由一组 N 智能体组成，其个体

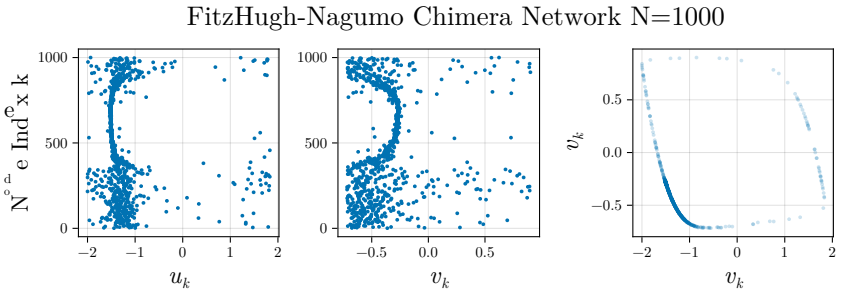


图 11.4 包含 1000 个节点的 FHN 奇美拉网络仿真。参数取自 [81]: $\epsilon = 0.05$, $a = 0.05$, $r = 0.35$, $\sigma = 0.1$, $\phi = \pi/2 - 0.1$。初始条件在圆 $u_k^2 + v_k^2 = 4$ 上均匀随机分布。节点 $500 \sim 700$ 呈现为一个同步簇，而其他节点处于异步状态。

意见在给定时间 t 由一个 k维向量$\mathbf{x}_i(t) \in \mathbb{R}^k$ 表示，并通过一个演化算子 ϕ 将当前意见映射为未来状态。因此，社交网络在时间 t 的意见状态可以写成矩阵 $X_t \in \mathbb{R}^{N\times k}$，其中，

$$X_{t+1} = \phi(X_t). \tag{11.1}$$

基于上述表述，经典的意见动力学问题旨在研究当系统趋于稳态时，ϕ 对网络中意见形成的影响。上述表述的普适性主要体现在 ϕ 的定义上，这使得所产生的动力学范围具有极大的灵活性。经典的 ϕ 表述通常假设全连接结构（即节点与所有其他节点交互）。然而，这些思想也可以推广到非全连接的随机网络，但往往会牺牲关于收敛和共识形成的解析保证。

研究意见形成时，有几种值得关注的特征。常见的研究特征之一是稳态中的簇数量，可分为三种不同结果：(1) 共识——所有节点都持有相同的意见；(2) 极化——形成两个占主导地位的意见簇；(3) 分裂——形成两个以上的意见簇。其他值得关注的特征还包括收敛时间、簇大小，以及最终状态下的意见分布。

人们提出了许多基于不同机制的意见动力学模型。一些经典模型包括用于离散动力学的投票者模型 [83–85], ，以及在考察意见概率分布时使用的 DeGroot [86] 和 Friedkin-Johnsen [87] 模型。另一类考虑连续取值意见的模型，是 Hegselman-Krause [88] 和 Deffuant [89, 90]提出的有限信任模型。这些模型纳入了选择性接触 [91, 92], 这一概念，即智能体不太可能与意见和自己相差悬殊的邻居互动。为此，模型设置了一个阈值

（半径/置信度界限） c，用于判定在任一给定时间，一对发生互动的邻居是否会在其持有的意见上作出一定程度的妥协 μ。一般而言，即使网络中只有一些不大的复杂性，例如随机图（无标度、小世界等）、异质的主体阈值 c 以及不对称的置信度阈值，分析结果也往往难以求得。因此，对这些情况的分析通常依赖蒙特卡洛数值模拟来获得结果。不过，Meng等人近期的研究已就各种网络拓扑及其产生的意见动力学 [93]的影响提供了大量实证结果。Di Mare等人 [94]也提出了通过博弈论方法理解智能体互动的思路。

11.3.2 异步与同步模拟

与振子网络类似，由于在数学上难以处理，研究意见动力学的许多方法都需要使用数值模拟。然而，与振荡器不同，意见动力学模型中的相互作用发生在离散时间步（差分方程）上，而非连续进行（常微分方程）。模拟此类网络通常采用基于主体的方法。意见更新具有这种间歇性，因此模拟可以通过两种不同的方式进行。

同步更新是一种计算简单且成本低廉的意见动力学模拟方式。在这种方法中，所有智能体都根据前一时间步的信息同时更新其状态。对于 DeGroot（DG）和 Friedkin-Johnsen（FJ）模型等线性模型，同步方法可以大幅简化实现所需的代码（即迭代一个矩阵方程）。由于每个时间步都必定会更新状态，同步更新程序往往表现出更快的收敛速度。然而，这种方法受到的一项批评是它缺乏现实性，因为个体通常不会协调一致地在离散时间步更新意见。

异步更新是一种更“连续”的方法，用于模拟基于智能体的模型或意见模型，尤其适用于研究发生在非规则网络拓扑上的意见动力学（即不属于环、格点、全连接网络，或无法用连续空间近似描述的任何网络）。其实现方式是随机选择一个智能体（或一条边），对该智能体进行局部状态更新，同时保持其他智能体的状态不变。重复这一过程，直到系统收敛。不出所料，这种方法的计算成本远高于同步更新，且通常收敛更慢，因为时间步对应的时间单位更小，并且与智能体数量成反比。即使模型类别相同，并且对时间步长的差异进行校正后，同步模拟与异步模拟中观测到的动力学仍可能大不相同。

11.3.3 DeGroot 模型与 Friedkin-Johnsen 模型

最早的意见形成模型之一由 DeGroot 等人提出。[86]该模型是一个线性矩阵模型，形式如下：

$$\mathbf{o}^{(t+1)} = \mathbf{A}\mathbf{o}^{(t)} = \cdots = \mathbf{A}^{(t+1)}\mathbf{o}^{(0)},$$

其中，$\mathbf{o}^{(t)} \in \mathbb{R}^n$ 是一个实值意见向量，$\mathbf{A} \in \mathbb{R}^{n\times n}$ 是一个双随机矩阵，对应于任意代理人/个体从他人那里获得的影响程度。因此，代理人会将当前意见与相连邻居的意见加权求和，以此更新自己的意见。该模型相对简单，具有解析可处理性，其动力学由$\mathbf{A}$刻画。

DeGroot 模型的扩展是 Friedkin-Johnsen（FJ）模型。FJ 模型通过加入描述固执代理人行为的项来修改演化方程，

$$\mathbf{o}^{(t+1)} = \mathbf{D}\mathbf{A}\mathbf{o}^{(t)} + (\mathbf{I} - \mathbf{D})\mathbf{o}^{(0)},$$

其中，$\mathbf{D}$是一个对角矩阵，其非零元素对应于个体改变意见的易感性。第二项的加入描述了个体对其初始意见的依附，因此体现了代理人的固执性。在 DG 和 FJ 模型中，更新均以同步方式进行，因此个体只能了解最近观察到的邻居意见状态。

11.3.4 投票者模型

与确定性的 DG 和 FJ 模型不同，投票者模型是一种随机意见模型，可采用同步或异步方案进行模拟。对于给定的、包含 N 个节点的网络，每个节点都可能持有一种二元意见$o_i^{(n)} \in \{0, 1\}$。在经典表述中，代理人遵循以下三条规则：

1. 若采用异步方案，随机选取一名投票者 2. 投票者采用随机邻居的状态 3. 重复以上步骤，直至达成共识

更一般地，投票者模型可用于描述这样一类智能体：它们的二元状态会根据某个转移函数 $c(x, \beta)$发生翻转，其中 x 表示网络中的位置/节点，而 β 表示网络当前的整体状态/配置（即，通过查找 $\beta(x)$，即可得到 x 处的状态）。

11.3.5 边界置信度模型（*BCM*）

有限信任模型描述了这样一种现象：智能体可能更倾向于与意见更相近的邻居互动，并受其影响 [95]。该模型通过定义一个阈值（半径/置信度界限） c 来实现这一点；该阈值决定了在任意给定时间，一对互动中的邻居是否会在其现有意见的基础上妥协一定程度 μ。目前广泛研究的有限信任模型主要有两类。最初提出的方法称为 Deffuant-Weisbuch（DW）模型 [89], ，采用异步更新规则，每个时间步只更新一对智能体的意见。具体而言，在每个时间步，随机选取一条对应于一对相邻节点的边 i， j，并根据以下方程改变它们的意见，

$$x_i(t+1) = f(x_i(t), x_j(t)), \tag{11.2}$$

$$f(x_i(t), x_j(t)) = \begin{cases} x_i(t) + \mu(x_j(t) - x_i(t)), & |x_i(t) - x_j(t)| < c \\ x_i(t), & \text{otherwise} \end{cases}. \tag{11.3}$$

另一种表述是 Hegselman-Krause（HK）模型，其中所有节点的意见都会在每个时间步根据前一时间步的意见同步更新 [88]。在这种表述中，每个代理人都会考虑其所有邻居的加权平均意见 $\bar{x}_i$（t），并据此更新自己的意见，

$$f(x_i(t), \bar{x}_j(t)) = \begin{cases} x_i(t) + \mu(\bar{x}_i(t) - x_i(t)), & |x_i(t) - \bar{x}_i(t)| < c \\ x_i(t), & \text{otherwise} \end{cases}. \tag{11.4}$$

这种替代表述采用了更具包容性的相互作用策略：只要 $\bar{x}_i$ 位于置信阈值 [95] 之内，意见更极端的代理人也可能仍会被部分纳入考量。还可以调整 BCM，通过允许每个代理人采用不同的阈值，来考虑异质个体。

除了最简单的网络结构（例如格点、全连接网络和环）之外，其他网络结构通常很难建立解析模型来描述 BCM 的行为，或者根本无法建立；而在最简单的网络结构中，可能可以采用简化方法和平均场方法。在大型社交网络中，这类简单的网络结构并不常见，因为个体不太可能与所有其他成员都直接相连。相反，网络中可能存在富人俱乐部、团和枢纽等其他非平凡结构。这些特征可能会扭曲个体的影响力，从而改变最终的收敛行为。因此，大多数实验都依赖计算机模拟来提供见解。

BCM 的函数形式内在地导致智能体形成共识或分裂，因为任意两个相互作用的主体要么保持静止，要么彼此收敛。通常，较小的阈值 c 和较大的妥协率 μ 往往会带来更快的收敛。在中间情况下，网络拓扑会影响

收敛时间和最终簇数等特征[93]。

第12章
序数时间序列分析

12.1 符号动力学

许多分析工具和相空间分析方法都适用于描述定义明确的动力系统的行为。然而，随着系统动力学变得愈加复杂，这些方法也会越来越难以应用。在现实应用中，往往只有数值观测数据（例如时间序列），而依赖数学结构存在这一前提的分析方法几乎难以适用。

解决这一难题的传统方法是依托嵌入定理的原理，希望通过延迟嵌入方法重构状态空间中的轨迹，使该状态空间与真实状态空间（若存在）微分同胚。如果希望学习动力学以用于预测等任务，可以采用各种模型构建方法来近似演化算子。然而，延迟嵌入方法会将数据扩充到 $\mathbb{R}^n$ 的空间中；如果观测数据含有噪声或分辨率有限，就会带来若干数值困难。

应对嘈杂且不完美的数据，一种变通方法是转而分析动力学的简化近似（即符号动力学）。回顾第 1 章，符号动力学是指将连续取值上的动力学（例如 $\mathbb{R}^n$）约化为有限个取值上的动力学，同时近似保留大部分有趣的动力学特征（例如周期轨道、同宿轨道/异宿轨道、稳定性）。也就是说，原系统中的周期序列应当对应于符号动力学中的周期序列。例如，可以将定义在单位区间上的帐篷映射和伯努利映射的动力学约化为由两个符号构成的序列（L，R）（留作读者练习），并证明周期轨道得以保留。

通过符号动力学实现的降维能够显著提高计算效率，并增强对噪声的鲁棒性。然而，将时间序列转换为符号动力学所需的变换尚不明确。通常，可以通过采用适当的状态空间划分来实现这种向符号动力学的转换，其中每个符号对应一个唯一的划分。然——

而，并非所有划分都相同；不同划分对符号动力学保留真实动力学的程度也可能不同。因此，定义划分时必须谨慎。

定义 12.1. – 划分
设 S 为一个集合。划分 P (*X*) 定义为由互不相交的集合 $\{X_i\}_i$ 组成的集合，使得

$$\mathcal{S} = \bigcup_i X_i$$

构造状态空间划分的一些示例如下：

1. 体素划分 [96] （即将状态空间分割成微小的方框）2. 最近邻划分 [61]3. 超平面（例如洛伦兹系统中 x 分量的符号、序数模式）

12.2 序数方法

12.2.1 时间序列的序数表示

序数方法旨在利用时间序列中观测值的秩排序，构造离散表示。为此，我们考察时间序列中由时滞 τ 分隔的 m 个点构成的序列。随后，将这组点映射为一个整数向量，用以表示序数符号，记为 π_i；该符号描述了这 m 个点的幅值排序（见图 12.1）。

定义序数符号有两种常用且功能等价的方法。振幅排序构造一种排序，使符号中的 m^{th} 个整数对应于观测值序列中第 m^{th} 个元素的幅值排序。时间顺序排序则利用序列的时间排序来定义符号，其中序数符号的 m^{th} 是序列中幅值最大的第 m^{th} 个元素的时间指标。在这两种方法中，对长度为 τ（m − 1）的滑动窗口依次应用该变换，可得到序数符号序列 π（n），称为序数时间序列，其中 π（n） ∈ X，且 X 是所有可能序数符号的集合。值得注意的是，X 是有限集，且 |X| = m!

12.2.2 序数划分

细心的读者会注意到，序数符号的构造与均匀延迟嵌入中的延迟向量有许多相似之处。简单来说，一个 m阶序数符号集对应于将延迟重构的相

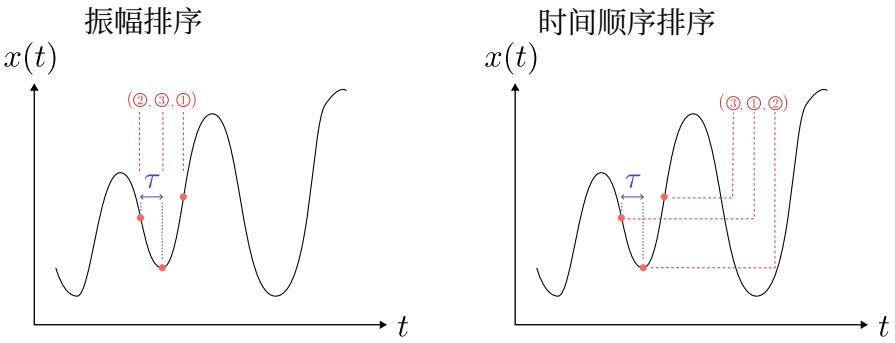


图 12.1 从数值序列定义序数符号的两种不同方法：振幅排序与时间顺序排序

空间划分为 $m!$ 个区域，这些区域由一组均通过原点的超平面构成。作为特例，阶数为 $m = 2$ 的序数符号编码了时间序列的给定子序列是递增还是递减的信息。此外，还可以证明，只要系统地打破排序相同的情况，序数符号的构造就能产生一个适当的完备划分

为说明这一点，设 $x(t)$ 为一个时间序列，$\vec{x}(t)$ 为按常规方式构造的 m 维延迟向量，

$$\vec{x}(t) = (x(t), x(t-\tau), \ldots, x(t-(m-1)\tau)) = (x_1(t), x_1(2), \ldots, x_m(t)),$$

其中 $m \in \mathbb{N}$，$m > 2$ 且 $\vec{x}(t) \in \mathrm{M} \subseteq \mathbb{R}^m$。我们将 m 记为序数构造的阶数，M为重构吸引子流形。此外，设 $\mathrm{X} = \{\pi_i\}_{i=1}^{m!}$ 为序数划分的集合。根据定义，每个序数符号 π_i 表示延迟向量元素的一种特定秩排序，这些元素满足以下一组嵌套不等式，

$$x_{i_1} \geq x_{i_2} \geq \cdots \geq x_{i_m}, \tag{12.1}$$

其中，$\{i_1, \ldots\ldots, i_m\} \in \{1, \ldots\ldots, m\}$ 是延迟向量中按秩排序的元素的索引。于是，我们可以将对应于 π_i 的划分定义为：

$$W_i = \{\vec{x} \in \mathbb{R}^m \,|\, x_{i_1} \geq x_{i_2} \geq \cdots \geq x_{i_m}\}.$$

由于 12.1 由 $m-1$ 个独立的线性不等式构成，因此 W_i 是一个楔形区域，其边界由 $m-1$ 个超平面定义，且这些超平面都经过原点。此外，

$$W_i \cap W_j \neq \emptyset \quad \Longleftrightarrow \quad \exists\, i, j \ \text{ s.t. } \ x_i = x_j,$$

且 $\mathrm{P} = \bigcup_i W_i = \mathbb{R}^m$。也就是说，序数符号描述了 $\mathbb{R}^m$ 的一种划分，直至每个划分的边界；这些边界对应于延迟向量分量中取值相等的元素。为了构造一个适当的划分，使得

$W_i \cap W_j = \emptyset$ 对所有 $i \neq j$ 均成立，我们可以按照元素索引系统地打破平局：若 $x_i = x_j$，其中 $i > j$，则将 $\vec{x}$ 分配到最符合条件 $x_i > x_j$ 的划分。

由于动力学发生所在的流形 M 位于 $\mathbb{R}^m$ 的一个子集上，因此无法保证轨迹会遍历所有可能的划分。实际上，并非所有 $m!$ 个序数符号都是完整描述动力学所必需的。这些未使用的符号也称为禁止符号。因此，对于无噪声系统^，只需考虑较小的序数符号集合 $\mathrm{P} \subset \mathrm{P}$，其划分覆盖 M 即可。引入噪声后，观测到的轨迹可能看似进入了先前被禁止、且不属于 M 的划分，从而增加所使用的序数符号数量。

12.3 网络方法

12.3.1 序数划分转换网络

与庞加莱映射的思想类似，原系统中的动力学被简化为一个映射，并附加了定义域为离散有限域的约束，

$$\pi(n+1) = f_m(\pi(n)), \qquad \pi(n) \in \mathcal{X}_{(m)},$$

其中 $\mathrm{X}^{(m)}$ 是由 m阶划分生成的所有可能序数符号的集合。对于离散时间观测（即大多数现实世界的数据），当 $m \to \infty$ 时，映射 f_m 的动力学近似于真实动力学。这对应于重构状态空间中具有无限精细度的楔形划分。对于具有离散观测的有限m，f_m 既不能保证重现动力学，也不能保证是一个函数。两条不同的轨迹可能从同一个划分出发，却映射到不同的划分（见图 12.2）。接受这一局限后，我们可以放弃对确定性描述的所有期望，转而采用更具随机性的方法来理解动力学。为此，我们可以考虑使用符号之间的转移概率，

$$p(\pi(n+1) = \pi_i \,|\, \pi(n) = \pi_{i_n}, \pi(n-1) = \pi_{i_{n-1}} \dots).$$

由于下一步的概率取决于符号的历史观测轨迹，转移行为不一定具有马尔可夫性质（即，下一状态的转移概率完全取决于当前状态）。不过，如果底层动力学是遍历的，则可以假设马尔可夫性质成立。因此，我们可以仅用一步转移概率来完整描述动力学的随机近似，

$$p_{ij} = p(\pi(n+1) = \pi_j \,|\, \pi(n) = \pi_i).$$

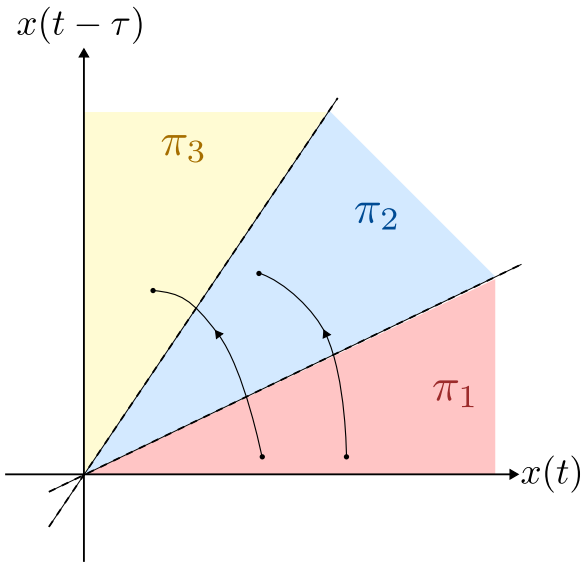


图 12.2 有限个划分的前向映射或序数符号不一定是唯一的。图中展示了一个示例轨迹。为便于说明，未显示 $x(t-\tau)$ 的第三个坐标轴。

在实践中，我们可以通过统计任意两个序数符号 π_i 和 π_j 之间观测到的转移频率，来数值近似 p_{ij} 的值。计算完成后，这些概率可以表示为一个序数转移网络，其连通性由加权矩阵 $A \in \mathbb{R}^{|\mathcal{X}|\times|\mathcal{X}|}$ 表示，矩阵中的元素为 $A_{ij} = p_{ij}$。其中，每个节点代表一个独特的序数符号，有向边则表示符号对之间的转移概率。

与禁止符号类似，对于给定的动力系统，特定符号对之间的某些转移也可能永远不会发生。因此，这些转移称为禁用转移，可用作衡量给定时间序列动力学复杂度的初步指标。周期轨道等较简单的动力学只会用到一小部分可能的序数符号，可能的转移数量也只占一小部分。对于混沌动力学等更“复杂”的系统，这些比例会增加；而在纯随机噪声的情况下，所有可能的符号和转移都可能出现，比例则会达到饱和。

12.4 复杂度度量

假设我们需要一种方法来区分周期性数据、混沌数据和随机数据。如果将这三种不同状态视为位于同一谱系上，就可以将它们描述为动力复杂度逐渐增加的情形。周期信号的复杂度最低，

而随机独立同分布噪声的复杂度最高。时间序列的复杂度也与其可预测性有关。

顾名思义，复杂性度量是量化和比较不同时间序列复杂程度的方法。此前讨论过的一个简单例子是李雅普诺夫谱，其中正值表示存在对初始条件的敏感依赖（SDIC）。更复杂的例子包括各种递归图度量，以及储层计算复杂性度量。

除了比较两个不同的系统，复杂性度量的一个主要实际用途是状态分类和变点检测。处于不同动力学状态的系统，其时间序列观测值应呈现出特征各异的复杂度值（例如，分岔参数在混沌状态与周期状态之间变化，或心电图（ECG）中从窦性心律转变为室性心动过速）。复杂性度量通过提取并概括时间序列的关键动力学特征，将时间序列转换为一种更适用于传统分类和变点检测方法的形式。实际应用中，通常会结合使用一组复杂性度量，以分类并检测动力学变化。

12.4.1 置换熵

我们可以利用序数符号的数量及其频率分布来计算给定时间序列的复杂度。更规则、更具周期性的时间序列只会使用可能的 $m!$ 序数符号集合中的一小部分。相比之下，复杂度更高的时间序列可能会涉及更多种类的划分，因此会使用更多序数符号。Bandt & Pompe [11] 提出了置换熵的概念，利用序数符号的频率分布来量化复杂度。置换熵可按如下方式计算，

$$h^{PE} = -\sum_{i=1}^{m!} p_i \log_2 p_i,$$

其中 p_i 是时间序列中观测到 i^{th} 序数符号的频率。该表达式是将香农熵计算应用于序数符号概率分布的结果。置换熵值较高，表示序数符号的种类更加多样，因此意味着时间序列更加复杂。置换熵的一个显著性质是，在一般情形 $m \to \infty$ 下，它有下界 Kolmogorov-Sinai 熵 h_{KS}。因此，置换熵也可以被视为衡量时间序列可预测性的代理指标。

12.4.2 条件排列熵

Unakafov [97] 提出了一种置换熵的扩展形式，即条件排列熵。条件排列熵不只是计算序数符号的频率分布，而是关注成对序数符号之间的转移概率，

$$h_{CPE} = -\sum_{i=1}^{m!} p_i \sum_{j=1}^{m!} p_{ij} \log_2(p_{ij}),$$

其中，p_{ij} 是从符号 π_i 转移到符号 π_j 的概率。这一熵测度旨在解决标准置换熵 h_{PE} 在有限 $m!$ 情况下估计柯尔莫哥洛夫-辛奈熵 h_{KS} 时存在的不足。尽管其计算复杂度与置换熵相同，h_{CPE} 却是 h_{KS} 的更佳估计量；对于周期动力系统（$h_{KS} = 0$），当 $m!$ 足够大时，等式成立。

12.5 序数庞加莱截面

12.5.1 庞加莱截面与首次返回映射

如前文简要讨论过的，首次返回映射（庞加莱映射）为分析连续动力系统提供了一种便捷的方法，可将维数降低一维。然而，构造首次返回映射所需的庞加莱截面的定义并非总是容易，而且不同截面选择的效果也未必相同。一种常用的构造截面方法，是将时间序列中连续出现的极大值作为映射序列中的值。洛伦兹于1963年首次采用这种方法，定义了一个穿过 Lorenz吸引子各叶瓣的庞加莱截面（见图12.3）。这种技术也可用于为其他系统（如 Rössler 系统）构造首次返回映射，以表明其动力学属于单峰映射族。

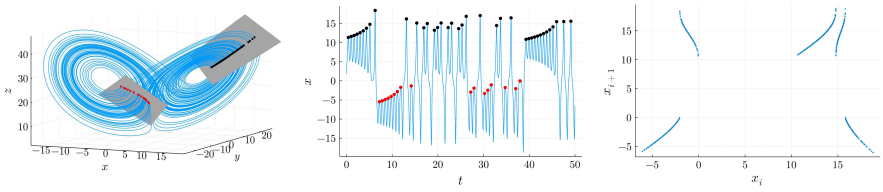


图 12.3 取自 [98] 的洛伦兹庞加莱映射。从左到右：(1) 由连续极大值法得到的庞加莱截面；(2) 时间序列的 x 分量上的极大值；(3) 连续极大值首次返回映射

12.5.2 庞加莱截面的序数方法

Shahriari 等人 [98] 提出了一种利用序数时间序列的思想定义庞加莱截面的替代方法。序数庞加莱截面方法依靠序数划分的超平面边界来定义一大类庞加莱截面。算法步骤如下：

1. 设 $x(t)$ 为一个时间序列。对于给定的阶数 m 和延迟滞后 τ，计算相应的序数划分、序数符号集合 π_1、π_2、……、π_n 以及出现频率 K_1、K_2、……、K_n。2. 根据各自对应的序数序列对完整时间序列 $x(t)$ 进行分离，构造 n 个不同的分量时间序列 $\hat{x}_1(t)$、$\hat{x}_2(t)$、……、$\hat{x}_n(t)$。3. 对每个分量时间序列 $\hat{x}_i(t)$，使用序数符号 $\hat{\pi}_1$、$\hat{\pi}_2$、……、$\hat{\pi}_m$ 以及出现频率 $\hat{}p_{i_1}$、$\hat{}p_{i_2}$、……、$\hat{}p_{i_m}$ 构造另一个序数划分。4. 对分量时间序列 $\hat{x}_i(t)$ 计算加权排列熵 $\mathrm{h}w_i$，其权重由 K_i 在原始完整时间序列 $x(t)$ 中的出现频率给出。

$$h_{W_i} = -\sum_{j=1}^{m} K_i p_{i_j} \log_2(K_i p_{i_j})$$

5. 还可以使用基于进入给定序数划分的首次进入频率计算的权重 $\hat{}\hat{}\hat{}K_1$、K_2、……、K_n 来计算入口加权排列熵（也就是说，如果一段时间序列观测值具有相同的序数符号 π_i，则只计数一次）。因此，给定序数时间序列分量 $x_i(t)$ 的入口加权排列熵计算如下：

$$h_{EW_i} = -\sum_{j=1}^{m} \hat{K}_i p_{i_j} \log_2(\hat{K}_i p_{i_j})$$

6. 按照加权排列熵（或入口加权排列熵）对每个序数符号 π_i 进行排序。对于每个符号 π_i，找出对应于进入π_i序数划分的首次进入的时间索引序列 $\{t_1^i$、t_2^i、……$\}$。对应于序数符号 π_i 的首次返回映射为 $(x(t_n^i$、$x(t_{n+1}^i))$

序数庞加莱截面方法会产生一整族候选庞加莱截面，可用于构造首次返回映射。加权熵和入口加权熵被用作筛选标准，优先选择与熵值最高的序数符号对应的映射。这样得到的庞加莱截面能够以平滑且定义明确的方式横切吸引子。与传统的连续极大值首次返回映射相比，利用序数庞加莱截面构造的首次返回映射对噪声也更加稳健（见图 12.5）。

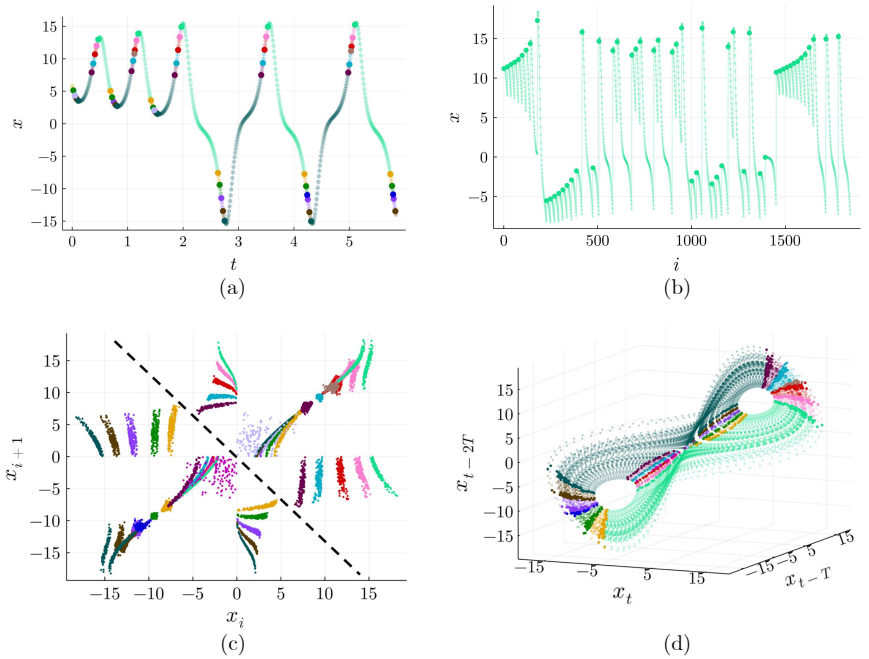


图 12.4 序数庞加莱截面的构造：(a) 根据序数符号着色的时间序列，并将入口点加粗显示。(b) 与单调递减序数序列 $\pi_1 = (4, 3, 2, 1)$ 对应的提取分量时间序列。(c) 根据序数符号着色的族或首次返回映射。(d) 带有加粗入口点、并根据序数符号着色的洛伦兹延迟吸引子。

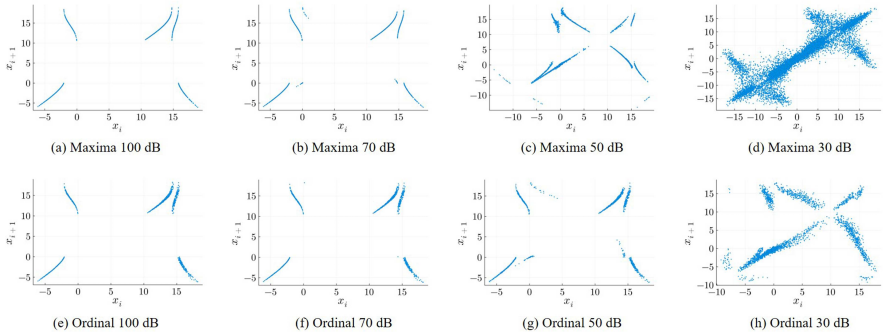


图 12.5 在不同信噪比下为 Lorenz 系统构造的首次返回映射。所施加的噪声为加性高斯噪声。顶行 (a-d) 是使用连续极大值方法构造的最大值，底行 (e-g) 是使用序数庞加莱截面构造的最大值。

第13章
练习与计算实验

1.（线性化）考虑以下含有实常数参数 δ， μ 的二维动力系统，其中 $\delta \neq 0$。不动点原点的稳定性如何随 δ 和 μ变化?

$$\dot{x} = -\delta x - \mu y + xy$$
$$\dot{y} = \mu x - \delta y + \frac{1}{2}(x^2 - y^2)$$

2.（中心流形）回顾一下，流的中心流形的特征方程为

$$N(h(x)) = Dh(x)[Ax + f(x, h(x))] - Bh(x) - g(x, h(x)) = 0$$

推导一个类似的特征方程，用于计算离散映射上的中心流形。

3.（稳定性）设 $\bar{x}(t)$ 是一个自治常微分方程组的解，其中 $\bar{x}(0) = x_0$。经过 x_0、对应于 $t > t_0$ 的正向轨道定义如下:

$$O^+(x_0, t_0) = \{x \in \mathbb{R}^n | x = \bar{x}(t),\ t \geq t_0,\ \bar{x}(t_0) = x_0\}$$

接下来，考虑以下稳定性定义:

轨道稳定性– $\bar{x}(t)$ 若给定 $\epsilon > 0, \exists \delta > 0$ ，对于任意其他解 $y(t)$，其中 $|x(t_0) - y(t_0)| < \delta$，都有 $d(y(t), O^+(x_0, t_0)) < \epsilon$对于 $t > t_0$，则称其轨道稳定

渐近轨道稳定性 – $\bar{x}(t)$ 是渐近轨道稳定的，当且仅当它具有轨道稳定性，并且 $\exists b > 0$ 满足：如果 $|x(t_0) - y(t_0)| < b$，则

$$\lim_{t \to \infty} d(y(t), O^+(x_0, t_0)) = 0.$$

证明以下向量场的所有轨迹都是渐近轨道稳定的：

$$\dot{\theta} = -\sin^2\theta + (1 - r^2)$$
$$\dot{r} = r(1 - r),$$

其中（θ, r） $\in S^1$, $\mathbb{R}$。

4.（中心流形）考虑以下含分岔参数 ϵ 的动力系统，其中 ϵ 很小，

$$\dot{x} = 2x + 2y$$
$$\dot{y} = x + y + x^4 + \epsilon y^2$$

a. 证明当 $\epsilon \geq 0$ 时，该系统有一个不动点；当 $\epsilon < 0$ 时，有三个不动点
b. 计算不动点（0，0）处的线性化动力学，并证明它是非双曲的，且线性化不足。因此，得出结论：它必定是不稳定不动点或鞍点。

c. 利用（b）部分中求得的特征值和特征向量，对（u, v）进行线性坐标变换。证明所得系统为

$$\begin{bmatrix}\dot{u}\\ \dot{v}\end{bmatrix} = \begin{bmatrix}0 & 0\\ 0 & 3\end{bmatrix}\begin{bmatrix}u\\ v\end{bmatrix} + \frac{1}{3}\begin{bmatrix}-2[(u+2v)^4 + \epsilon(-u+v)^2]\\ (u+2v)^4 + \epsilon(-u+v)^2\end{bmatrix}$$

wh这里u一个d v 是 th中心t是一个d and unstabl分量 t分别 tivel
y. d. 求中心流形的 4th 阶近似。

$$h(u) = a_1u^2 + a_2u\epsilon + a_3\epsilon^2 + a_4u^3 + a_5u^2\epsilon + a_6u\epsilon^2 + a_7\epsilon^3 + a_8u^4 + a_9u^3\epsilon + a_{10}u^2\epsilon^2 + a_{11}u\epsilon^3 + a_{12}\epsilon^4$$

注意：对于参数 ϵ 为常数的系统， n^{th} 阶近似包含所有 u 和 ϵ 的幂次之和小于或等于 n 的项。为减少项数，建议每当出现高于 4 阶的项时，就主动将其删去。

e. 利用中心流形描述原点处不动点的稳定性。它如何随 ϵ 的变化而改变？ f. 在（u, v）坐标中画出该动力系统的全局流。其行为如何随 ϵ 改变?

5.（复现）证明三倍映射 $f(x_n) = 3x_n \bmod 1$ 是单位区间上的保长度函数。验证这并不意味着 $\mu(U) = \mu(f(U))$，其中 μ 是勒贝格测度（即长度），而 U 是 $[0, 1]$ 的一个子区间。

6.（非自治系统）考虑以下非自治动力系统

$$\dot{x} = x(x+a)(x-a) + q(t)$$

a. 设 $q(t) = 0$。求出所有不动点。对于稳定的不动点，给出其吸引域。b. 求出使某个稳定吸引域存在所需的 $q(t)$ 条件。c. 假设 $q(t)$ 满足 (b) 部分中的条件。令 $\Phi_\infty(x_0|q(0) = q_0)$ 为初始条件为 $x(t) = x_0$ 、初始扰动为 $q(t) = p_0$ 的点的长期演化。对于所有 $q_0 \neq p_0$， $\Phi_\infty(x_0|q(0) = q_0) \implies \Phi_\infty(x_0|q(0) = p_0)$ 吗？d. 令 $q(t) = \frac{-2a^3}{3\sqrt{3}}\cos(\omega t + \phi)$。针对多个初始条件下的不同 ω 和 phi 值，生成时间序列和 ω极限集。结果是否符合你对 (c) 部分的预期?

7.（嵌入）使用标准参数值（$\sigma = 10$、 $\beta = \frac{8}{3}$、 $\rho = 28$）对洛伦兹方程（见下文）进行数值积分。

$$\begin{aligned}\dot{x} &= \sigma(y-x)\\ \dot{y} &= x(\rho - z) - y\\ \dot{z} &= xy - \beta z\end{aligned}$$

证明仅根据 x 和 y 分量进行时间延迟重构（使用“恰当选择的”嵌入参数）可以得到一个与原始吸引子拓扑等价的对象。然而，说明仅根据洛伦兹方程的 z 分量无法得到有效嵌入，并解释原因。此外，提出一种将洛伦兹时间序列约化为符号动力学的方法。

8.（嵌入）考虑混沌洛伦兹系统的 x 分量（见问题 7）。

a. 实现一种可生成全局主值嵌入的算法。b. 使用 Julia DynamicalSystems 软件包或自行实现的方法，根据自互信息的第一个极小值计算理想嵌入时滞。比较通过主成分分析和均匀延迟嵌入得到的重构吸引子。c. 考虑一个最大时滞为τ_{max}的 m维主成分分析嵌入。其分量是否等价于时滞为$\tau = \frac{\tau_{max}}{m}$的均匀 m维嵌入?

9.（对初始条件的敏感依赖）考虑伯努利映射，

$$x_{n+1} = \begin{cases} 2x_n, & x_n < 0.5 \\ 2x_n - 1, & x_n \geq 0.5 \end{cases}$$

a. 使用各种有理数和无理数初始条件模拟伯努利映射。绘制轨迹，并验证有理数初始条件总会产生周期行为。b. 令 $a_0 = \frac{\pi}{20}$ 和 b_0 等于 a_0，但四舍五入到小数点后 10 位。证明分离量 a_n 和 b_n 近似按以 2 为底的指数函数增长。（你可能需要使用对数刻度绘图。）

10.（信息论、代数）证明互信息可以用联合概率和边缘概率表示为以下恒等式

$$\begin{aligned} I(x,y) &= \sum_{i,j} p_{i,j}^{(xy)} \ln p_{i,j}^{(xy)} - \sum_i p_i^{(x)} \ln p_i^{(x)} - \sum_j p_j^{(y)} \ln p_j^{(y)} \\ &= \sum_{i,j} p_{i,j}^{(xy)} \ln \frac{p_{i,j}^{(xy)}}{p_i^{(x)} p_j^{(y)}} \end{aligned}$$

11.（模型选择、代数）设有观测数据 x_y 和 y_t，以及一个模型类 ϕ_θ，其中 θ 是一组 k 模型参数。假设已经找到最优参数集 $\hat{}\theta$，使得

$$y_t = \phi_{\hat{\theta}}(x_t) + \epsilon_t,\ \epsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma^2).$$

证明该模型的 AIC（除去一个常数后）可表示为

$$AIC(\phi_{\hat{\theta}}) = 2k + n \log\left(\frac{SSE}{n}\right)$$

12.（不变量）计算门格海绵的分形维数（搜索相关资料会有所帮助。）

13.（不变量）证明，有界集 $B \in \mathbb{R}^n$ 的任意仿射变换（对于 $n \in \mathbb{Z}^+$）都不会改变该集合的关联维数。

14.（关联积分）证明，阶数为 $q = 1$ 的广义维数（即信息维数）由下列表达式给出。

$$D_1 = \lim_{\epsilon \to 0} \frac{\langle \ln p_\epsilon \rangle_\mu}{\ln \epsilon}$$

15.（熵）证明 $H(p)$ 满足以下性质：对于概率分布 $p = (p_1、\ldots\ldots、p_n)$ 和 $q = (q_1、\ldots\ldots、q_n)$，若 $p \neq q$，则 $H(p) < -\sum_{i=1}^n p_1 \log q_i$。

16.（神经网络）考虑一个将 $\mathbb{R} \to \mathbb{R}$ 映射的一般前馈神经网络。令 n_l 和 n_h 分别表示隐藏层的数量和每个隐藏层的节点数。假设训练后的模型误差服从高斯分布，且所有参数的精度均为常数，未经优化。请给出用 n_l 和 n_h 表示的模型描述长度表达式。假设均值误差随隐藏节点数呈指数函数增长。这会如何改变最终表达式？

17.（信息密度）英语和印度尼西亚语是三种书面语言结构迥异的语言。例如，印度尼西亚语是一种马来语族语言，常通过添加前缀和后缀来构造新词和表达新的语义。

请看《世界人权宣言》第一条的以下译文：

英语
人人生而自由，在尊严和权利上一律平等。他们赋有理性和良心，并应以兄弟情谊的精神彼此相待。

印度尼西亚语
人人生而自由，在尊严和权利上一律平等。

人人赋有理性和良知，应该本着兄弟情谊的精神彼此相待。

编写一个程序，根据一段文本和一组码字频率自动构造霍夫曼编码。计算上述每段文字的平均码字长度（以比特为单位）和编码长度。比较各段文字的长度。前缀和后缀的频繁使用会如何影响码字数量及其相对频率？这又会如何影响给定长度文本的信息密度？

18.（预测）使用随机初始条件对洛伦兹方程进行积分，得到一条轨迹。利用由此产生的时间序列，编写代码进行最近邻预测，并将其应用于积分得到的时间序列，分别采用（a）无嵌入、（b）二维延迟嵌入和（c）三维延迟嵌入，并尝试不同的嵌入滞后。请评论你的发现。减少训练数据量和／或添加观测高斯噪声时会发生什么？

19.（心电图，实验）使用你的最近邻预测器，尝试对 ECG.csv 数据集进行单步和多步自由运行预测。效果如何？预测器在什么情况下会失效，是否存在某些规律？造成最近邻预测器失效的潜在原因有哪些？

如果你有时间，可以通过哪些方法构建更好的预测器?

20.（预测）使用随机初始条件对洛伦兹方程进行积分，得到一条轨迹。利用生成的时间序列编写代码，实现最近邻预测，并将其应用于积分得到的时间序列，分别进行：（a）不使用嵌入；（b）使用二维延迟嵌入；（c）使用三维延迟嵌入，并尝试不同的嵌入滞后。评论你的发现。预测是否容易在某些特定区域失效？原因是什么？减少训练数据量和/或添加观测高斯噪声会产生什么影响？请提出一些可通过改进算法来提升预测效果的建议。

21.（挑战——李雅普诺夫指数）在不借助预构建动力系统软件包的情况下，实现 Wolf 算法，并通过数值方法估算你选择的 2 个连续混沌系统的最大李雅普诺夫指数。构建该算法的两个版本：（a）采用轨迹对的标准方法；（b）基于邻域的方法。分别在以下两种情况下进行估算：（i）使用完整状态向量的数据；（ii）使用延迟重构。将结果与文献中已确立的李雅普诺夫指数（或使用现有预构建工具得到的结果）进行比较，并讨论你的发现。该算法面临哪些挑战?

22.（心电图，实验）使用你的最近邻预测器，尝试对 ECG.csv 数据集进行单步和多步自由运行预测。它的表现如何？预测器在什么情况下会失效，是否存在任何规律？最近邻预测器失效的潜在原因有哪些？如果你有时间，可以通过哪些方法构建更好的预测器?

23.（序数时间序列分析）回想一下下面给出的洛伦兹动力系统:

$$\begin{aligned}\dot{x} &= \sigma(y - x)\\ \dot{y} &= x(\rho - z) - y\\ \dot{z} &= xy - \beta z\end{aligned}$$

a. 使用不同的时间步长$\Delta t = (0.001，0.05、0.01)$ 对方程进行数值积分。b. 使用时间序列中的单个分量绘制首次返回映射（FRM）。c. 使用你在 (a) 部分生成的时间序列，应用维数为 $m = 4$ 的序数划分。每个序数划分的出现概率是多少？d. 选择序数划分中权重最高的一项，并为每个时间序列绘制首次返回映射。不同时间序列及其分量的首次返回映射看起来如何?

e. 选择合适的时滞，将 x 分量时间序列延迟嵌入到 3 维空间中。如何根据嵌入吸引子上属于这一特定序数划分的对应点的位置，解释首次返回映射之间的差异?

24.（替代数据）考虑处于相位给定状态下运行的 R¨ossler 混沌动力系统

$$\begin{aligned}\dot{x} &= -y - z \\ \dot{y} &= x + ay \\ \dot{z} &= b + z(x - c)\end{aligned}$$

a. R¨ossler 系统呈现出两种独特的状态：相位相干（a，b，c）=（0.165，0.4，8.5）和非相位相干（0.265，0.4，8.5）。使用这些值进行积分，并描述从时间序列及其重构吸引子中可以观察到的几何和动力学差异。这两个系统的信号的傅里叶谱有何不同？你可以查阅并阅读一些文章，以进一步了解相关内容并讨论你的发现。b. 编写一个算法，从非相位相干系统的 x 分量构造幅值调整傅里叶替代数据和伪周期替代数据。描述这两种替代时间序列之间的差异。

25.（储备池计算）实现一个可运行的储备池计算机，并用它模拟 Lorenz 系统。

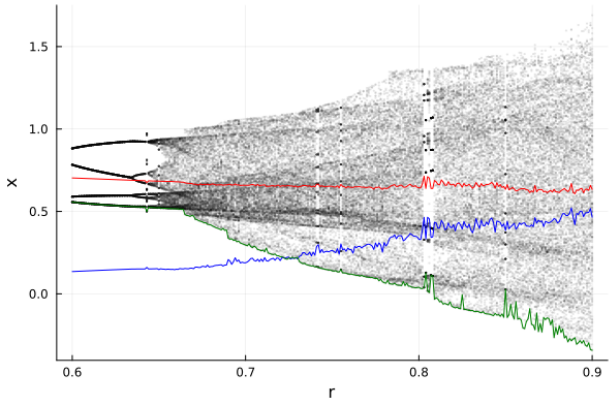
a. 编写一个 create ESN 函数，该函数接收超参数，并返回 ESN 所需的相关矩阵。假设 ESN 的输入为标量。b. 编写一个 run ESN 函数，该函数接收标量输入信号 x 及相关矩阵，以生成 ESN 的激活状态。c. 训练你的 ESN，使其接收 Lorenz 系统中 x 分量的轨迹，并输出未来一个时间步的同一分量。生成 Lorenz 轨迹时，时间步长设为 $\Delta t = 0.05$。你必须选择并明确说明合理的超参数，但无需对其进行优化。在独立测试序列上绘制预测结果，并以相关性报告误差。d. 重现第 9 章中的图**31** 和**32**，并据此说明你的储备池计算机的记忆容量。e. 改变储备池计算机的谱半径，并描述其对记忆容量和预测任务误差的影响。f. 使用你的储备池计算机，对 Lorenz 系统的 x 分量进行自由运行预测。

26.（递归图）下方是由 Ikeda 映射的 x 分量生成的分岔图，

$$x_{n+1} = 1 + r(x_n \cos(t_n) - y_n \sin(t_n))$$
$$y_{n+1} = r(x_n \sin(t_n) + y_n \cos(t_n))$$

带有 $t_n = 0.4 - \frac{6}{1+x_n^2+y_n^2}$。

参数范围内的均值、标准差和最小值轨迹分别以红色、蓝色和绿色显示。



a. 对于参数值 $r \in \{0.60, 0.65, 0.70\}$，从 Ikeda 映射的 x 分量生成递归图，不进行嵌入，并评论你的观测结果。使用嵌入重复此过程。
b. 重现该系统的分岔图，并绘制标量对角线、垂直线和递归时间特征的轨迹。评论这些特征检测状态变化的能力。注意：最终图应为包含三条轨迹的分岔图，但可能需要对这些特征进行标准化，以统一不同的尺度。
c. 使用其他非线性时间序列分析领域中的三个标量特征重复上述过程（例如储备池时间序列分析、符号分析、持久同调）。这三个特征可以全部来自同一分析领域，也可以分别来自三个不同的领域。

注意：为了让绘图更易于处理，绘制分岔图时只使用 *200* 个点，但计算特征时使用 *1000* 个点（或更多）。别忘了从轨迹中移除瞬态。

27.（最小描述长度）假设我们有一个参数为 $\theta = (\theta_1, \theta_2, ...\theta_k)$ 的线性模型。模型输出为 $X\theta$，期望的模型输出为 Y，其中

$$Y = (y(t_1), y(t_2), ...y(t_n))^T, X = \begin{bmatrix} x_1(t_1) & x_2(t_1) & \dots & x_k(t_1) \\ x_1(t_2) & x_2(t_2) & \dots & x_k(t_2) \\ \dots & \dots & \dots & \dots \\ x_1(t_n) & x_2(t_n) & \dots & x_k(t_n) \end{bmatrix}.$$

为了求出该模型的描述长度，我们必须估算其误差（$E = Y - X\theta$）和参数（θ）的描述长度。我们使用误差概率分布的微分熵来估算误差的长度。假设误差服从均值为零、方差为 σ^2 的正态分布，则 n 个误差点的描述长度为

$$L(E|\theta) \approx \frac{n}{2}\log(2\pi e\sigma^2) \text{ where } \sigma^2 = \frac{E^T E}{n} .$$

如果参数以精度为 γ 的浮点数编码（即 $\theta_i = 0。a_1a_2...a_{\log(\gamma)} \times e^{2m}$），并且忽略它们的符号、任何自定界前缀以及编码 $2m$ 的成本，那么参数的长度（$L(\theta)$）就是$-k\log(\gamma)$。通过添加精确的扰动** $\delta \in \mathbb{R}^k$，我们可以将 θ_i 写成 $0。a_1a_2...a_{(\log(\delta_i)-2m)} \times e^{2m_i}$，并使 $L(\theta)$ 减少 $\sum_i(\log(\gamma) - \log(|\delta_i|) + 2m_i)$，即

$$L(\theta + \delta) = L(\theta) + k\log(\gamma) - \sum_i(\log(|\delta_i|) - 2m_i)$$

最优舍入扰动使 $L(\theta+\delta)$ 与 $L(E|\theta+\delta)$ 的和最小，即：

$$\frac{\partial}{\partial\delta_i}\left(L(E|\theta) + \delta^T\frac{\partial L(E|\theta)}{\partial\delta} + \frac{1}{2}\delta^T\frac{\partial^2 L(E|\theta)}{\partial\delta^2}\delta\right)$$
$$+\frac{\partial}{\partial\delta_i}\left(L(\theta) + k\log(\gamma) - \sum_i(\log(|\delta_i|) - 2m_i)\right) = 0$$

模型的最小总描述长度（当 δ 满足方程 3 时取得）为

$$\begin{aligned} L_{tot} &= L(E|\theta + \delta_{opt}) + L(\theta + \delta_{opt}) \\ &= \frac{n}{2}\log(2\pi e\sigma^2) + \delta^T\frac{\partial L(E|\theta)}{\partial\delta} + \frac{1}{2}\delta^T\frac{\partial^2 L(E|\theta)}{\partial\delta^2}\delta \\ &\quad - \partial\delta^2\delta - \sum_{i=1}^{k}(\log(|\delta_i|) \text{ (ignoring the 2m)} . \end{aligned}$$

如果使用线性回归来求参数 θ，方程 3 可以简化，总描述长度将与单元读本中的表达式相似。考虑模型改用岭回归训练且惩罚项为 α 时的描述长度。在岭回归的情况下，θ 满足

$$\frac{\partial}{\partial\theta}\left(E^T E + \alpha\theta^T\theta\right) = 0 .$$

**只有当 δ_i 恰好与真实参数值及其低精度舍入值之间的差完全一致时，情况才确实如此。不过，假设对于所有 $\delta \in \mathbb{R}^k$（包括 $|\delta_i| > |\theta_i|$ 时）这一结论都成立。

a. θ 的岭回归方程（方程 4）的解是什么？ b. $\partial L_{(E)}/\partial\theta$ 和 $\partial^2 L(E)/\partial\theta^2$ 分别是什么？ ~~（你需要用到方程~~ 1 和 4） c. 化简方程 3，求出 δ 的表达式（提示：方程 3 中的大多数项都不依赖于 δ ） d. 求出 L_{tot} 的表达式，并尽可能化简。 e. 当 $\alpha \to 0$ 和 $n \to \infty$时，会发生什么？

28.（动力网络）考虑一个用于连续意见动力学的动力网络模型。回顾一下，2种意见的BCM模型如下所示：

$$x_i(t+1) = \begin{cases} x_i(t) + \alpha(x_j(t) - x_i(t)), & |x_i(t) - x_j(t)| < \rho \\ x_i(t), & |x_i(t) - x_j(t)| \geq \rho \end{cases}$$

a. 提出一种扩展上述意见更新模型的方法，以考虑3个不同的意见方向。对于n 种意见的情况，应如何将其推广？

b. 考虑一个用于表示选举中选民情绪的3种意见动力网络，分别对应3个政治团体：工党、LNP和独立人士。使用你在第（a）节中提出的模型，实现BCM模型的基于主体的仿真，并采用两种不同的配置：（i）同步更新；（ii）异步更新。针对不同的网络拓扑和置信度阈值 α，展示仿真结果。

c. 假设所有个体此后都在每个时间步参加选举投票，且所有选票价值相等。采用简单多数制与排序投票时，选举结果有何不同？对于不同的 α 值，结果又有何差异？（你需要选择一种合理的方法，根据个体的意见确定其投票偏好。随着不同意见的数量 n 增加，这些结果会如何变化？

参考文献

[1]J. Carr，《中心流形理论的应用》（Springer-Verlag，1983）。[2]S. Wiggins，《应用非线性动力系统与混沌导论》（Springer-Verlag，2003）。[3]E. Lorenz，“确定性非周期流”，《大气科学杂志》20（1963）。[4]O. E. R¨ossler，“一个关于连续混沌的方程”，《物理快报 A》57，397–398（1976）。[5]M. Henon，“具有奇异吸引子的二维映射”，Communications in Mathematical Physics 50，69–77（1976）。[6]K. Ikeda，“环形腔系统透射光的多值稳态及其不稳定性”，《光学通讯》30，257–261（1979）。[7]R. Nillsen，《动力系统中的随机性与复现：实分析方法》，第31卷（美国数学会，2010）。[8]F. Takens，“动力系统与湍流”，华威，1980，366–381（1981）。[9]L. Noakes，“Takens 嵌入定理”，International Journal of Bifurcation and Chaos 1，867–872（1991）。[10]N. H. Packard、J. P. Crutchfield、J. D. Farmer 和 R. S. Shaw，“从时间序列中提取几何结构”，《物理评论快报》45，712（1980）。[11]C. Bandt 和 B. Pompe，“置换熵：一种自然的时间序列复杂性度量”，《物理评论快报》88，174102（2002）。[12]M. McCullough、M. Small、T. Stemler 和 H. H.-C. Iu，“用于捕捉连续动力系统动力学的时间滞后序数分区网络”，Chaos 25（2015）。[13]R. Gilmore，“混沌动力系统的拓扑分析”，Reviews of Modern Physics 70，1455（1998）。[14]D. S. Broomhead 和 G. P. King，“从实验数据中提取定性动力学”，Physica D: Nonlinear Phenomena 20，217–236（1986）。[15]M. Casdagli、S. Eubank、J. D. Farmer 和 J. Gibson，“噪声环境下的状态空间重构”，Physica D: Nonlinear Phenomena 51，52–98（1991）。[16]K. Judd 和 A. Mees，“作为建模问题的嵌入”，Physica D: Nonlinear Phenomena 120，273–286（1998）。[17]L. M. Pecora、L. Moniz、J. Nichols 和 T. L. Carroll，“吸引子重构的统一方法”，Chaos 17（2007）。[18]L. C. Uzal、G. L. Grinblat 和 P. F. Verdes，“动力系统的最优重构：一种噪声放大方法”，Physical Review E 84，016223（2011）。[19]E. Tan、S. Algar、D. Corrˆea、M. Small、T. Stemler 和 D. Walker，“选择嵌入时延：嵌入技术概述及一种利用持久同调的新方法”，Chaos 33（2023）。[20]J. Gao 和 Z. Zheng，“混沌时间序列的局部指数发散图与最优嵌入”，《物理快报 A》181，153–158（1993）。[21]J. Gao 和 Z. Zheng，“确定性混沌的直接动力学检验”，《欧洲物理学快报》25，485（1994）。

[22]J. Theiler, 《应用于有限时间序列数据的相关算法所产生的伪维数》, 《物理评论A》**34**, 2427（1986）。[23]W. Liebert、K. Pawelzik 和 H. G. Schuster, 《基于拓扑学考虑的混沌吸引子最优嵌入》, 《欧洲物理学快报》**14**, 521（1991）。[24]A. M. Fraser 和 H. L. Swinney, 《基于互信息的奇异吸引子独立坐标》, 《物理评论A》**33**, 1134（1986）。[25]S. Wallot 和 D. Mønster, 《在 MATLAB 中计算平均互信息（ami）和伪近邻（fnn）, 以估计多维时间序列的嵌入参数》, 《心理学前沿》**9**, 1679（2018）。[26] A. Kraskov、H. St¨ogbauer 和 P. Grassberger, 《互信息估计》, 《物理评论E》**69**, 066138（2004）。[27]A. Papana 和 D. Kugiumtzis, 《时间序列互信息估计器的评估》, International Journal of Bifurcation and Chaos **19**, 4197–4215（2009）。[28]T. Buzug 和 G. Pfister, 《计算延迟时间坐标最优嵌入参数的算法比较》, Physica D: Nonlinear Phenomena **58**, 127–137（1992）。[29]T. Sauer、J. A. Yorke 和 M. Casdagli, 《Embedology》, 《统计物理学杂志》**65**, 579–616（1991）。[30]H. D. Abarbanel、R. Huerta、M. I. Rabinovich、N. F. Rulkov、P. F. Rowat 和 A. I. Selverston, 《突触耦合混沌模型神经元的同步活动》, Neural Computation **8**, 1567–1602（1996）。[31]S. P. Garcia 和 J. S. Almeida, 《通过具有不同时间延迟的最近邻嵌入进行多变量相空间重构》, 《物理评论E》**72**, 027205（2005）。[32]M. B. Kennel、R. Brown 和 H. D. Abarbanel, 《利用几何构造确定相空间重构的嵌入维数》, 《物理评论A》**45**, 3403（1992）。[33]K.-H. Kr¨amer、G. Datseris、J. Kurths、I. Z. Kiss、J. L. Ocampo-Espindola 和 N. Marwan, 《吸引子重构的统一自动化方法》, 《新物理学杂志》**23**, 033017（2021）。[34]M. Dorigo 和 L. M. Gambardella, 《蚁群系统：一种解决旅行商问题的协同学习方法》, 《IEEE进化计算汇刊》**1**, 53–66（1997）。[35]M. Shen、W.-N. Chen、J. Zhang、H. S.-H. Chung 和 O. Kaynak, 《利用蚁群优化为混沌时间序列的非均匀嵌入选择最优参数》, 《IEEE控制论汇刊》**43**, 790–802（2013）。[36]K. H. Kr¨aemer、M. Gelbrecht、I. Pavithran、R. Sujith 和 N. Marwan, 《通过蒙特卡洛决策树搜索实现最优状态空间重构》, 《非线性动力学》**108**, 1525–1545（2022）。[37]A. Wolf、J. B. Swift、H. L. Swinney 和 J. A. Vastano, 《从时间序列确定李雅普诺夫指数》, Physica D **16**, 285–317（1985）。[38]M. T. Rosenstein、J. J. Collins 和 C. J. De Luca, 《从小型数据集计算最大李雅普诺夫指数的实用方法》, Physica D **65**, 117–134（1993）。

[39]U. Parlitz，《从时间序列中识别真实和伪李雅普诺夫指数》，International Journal of Bifurcation and Chaos **2**，155–165（1992）。[40]H. Kantz 和 T. Schreiber，非线性时间序列分析（剑桥大学出版社，2003年）。[41]P. Grassberger 和 I. Procaccia，《测量奇异吸引子的奇异性》，Physica D **9**，189–208（1983）。[42]K. Judd，《一种改进的维数估计器及关于提供置信区间的若干评论》，Physica D **56**，216–228（1992）。[43]C. Diks，《具有随机波动率的收益率的关联维数》，Quanti- tative Finance **4**，45（2003）。[44]D. Yu、M. Small、R. G. Harrison 和 C. Diks，《高斯核算法在从含噪时间序列数据估计不变量和噪声水平中的高效实现》，《物理评论E》**61**，3750（2000）。[45]G. E. Box，《科学建模策略中的稳健性》，载于统计学中的稳健性（爱思唯尔，1979年），第201–236页。[46]K. Hornik、M. Stinchcombe 和 H. White，《多层前馈网络是通用逼近器》，Neural Networks **2**，359–366（1989）。[47]S. Hochreiter 和 J. Schmidhuber，《长短期记忆》，Neural Computa- tion**9**，1735–1780（1997）。[48]S. L. Brunton、J. L. Proctor 和 J. N. Kutz，《通过非线性动力系统的稀疏辨识从数据中发现控制方程》，美国国家科学院院刊**113**，3932–3937（2016）。[49]G. Martius 和 C. H. Lampert，《外推与学习方程》，arXiv（2016）。[50]M. Raissi、P. Perdikaris 和 G. E. Karniadakis，《物理信息神经网络：用于求解涉及非线性偏微分方程的正问题与逆问题的深度学习框架》，Journal of Computational Physics**378**，686–707（2019）。[51]J. E. Cavanaugh 和 A. A. Neath，《赤池信息准则：背景、推导、性质、应用、解释与改进》，Wiley Inter- disciplinary Reviews: Computational Statistics **11**，e1460（2019）。[52]G. Schwarz，《模型维数的估计》，The Annals of Statistics，461–464（1978）。[53]A. A. Neath andJ.E.Cavanaugh，《贝叶斯信息准则：背景、推导与应用》，Wiley Interdisciplinary Reviews: Computational Statistics**4**，199–203（2012）。[54]D. J. MacKay，信息论、推断与学习算法（剑桥大学出版社，2003年）。[55]J. Rissanen，统计研究中的随机复杂度，第15卷（世界科学出版社，1998年）。[56]J. Theiler、S. Eubank、A. Longtin、B. Galdrikian 和 J. D. Farmer，《时间序列中的非线性检验：替代数据方法》，Physica D **58**，77–94（1992）。[57]J. Theiler 和 D. Prichard，《用于假设检验的约束实现蒙特卡洛方法》，Physica D: Nonlinear Phenomena**94**，221–235（1996）。

[58]T. Schreiber 和 A. Schmitz, 《用于非线性检验的改进替代数据》, Physical Review Letters 77, 635 (1996)。[59]T. Nakamura 和 M. Small, 《小幅打乱替代数据：检验带趋势的波动数据中的动力学》, Physical Review E 72, 056216 (2005)。[60]M. Small 和 C. K. Tse, 《将替代数据方法应用于循环时间序列》, Physica D 164, 187–201 (2002)。[61]E. Tan、S. D. Algar、D. Corrˆea、T. Stemler 和 M. Small, 《用于变点检测的吸引子网络表示》, Communications Physics 6, 340 (2023)。[62]H. Jaeger, 《分析和训练循环神经网络的"回声状态"方法——附勘误说明》, 德国波恩：德国国家信息技术研究中心 GMD 技术报告 148, 13 (2001)。[63]W. Maass、T. Natschl¨ager 和 H. Markram, 《没有稳定状态的实时计算：一种基于扰动的神经计算新框架》, Neural Computation 14, 2531–2560 (2002)。[64]T. Lymburn、A. Khor、T. Stemler、D. C. Corrˆea、M. Small 和 T. J¨ungling, 《回声状态网络的一致性》, Chaos 29 (2019)。[65]A. Hart、J. Hook 和 J. Dawes, 《回声状态网络的嵌入与逼近定理》, Neural Networks 128, 234–247 (2020)。[66]A. G. Hart, 《连续时间储备池计算机的广义同步、嵌入与逼近》, Physica D: Nonlinear Phenomena 458, 133956 (2024)。[67]M. Lukoˇseviˇcius, 《应用回声状态网络实用指南》, 载于《神经网络：行业诀窍：第二版》（施普林格，2012），第 659–686 页。[68]L. Appeltant、M. C. Soriano、G. Van der Sande、J. Danckaert、S. Massar、J. Dambre、B. Schrauwen、C. R. Mirasso 和 I. Fischer, 《利用单个动力学节点作为复杂系统进行信息处理》, Nature Communications 2, 468 (2011)。[69]T. Lymburn、S. D. Algar、M. Small 和 T. J¨ungling, 《利用群体进行储备池计算》, Chaos 31 (2021)。[70]K. Nakajima 和 I. Fischer, 《储备池计算》（施普林格，2021）。[71]B. Thorne、T. J¨ungling、M. Small、D. Corrˆea 和 A. Zaitouny, 《储备池时间序列分析：将复杂动力系统的响应用作通用变化指标》, Chaos 32 (2022)。[72]B. Thorne、T. J¨ungling、M. Small、D. Corrˆea 和 A. Zaitouny, 《通过储备池计算研究时间序列复杂性的新方法》, 载于《澳大拉西亚人工智能联合会议》（施普林格，2022），第 442–455 页。[73]P. Faure 和 A. Lesne, 《符号序列的递归图》, International Journal of Bifurcation and Chaos 20, 1731–1749 (2010)。[74] N. Marwan, 《事件时间序列递归分析的挑战与展望》, Frontiers in Applied Mathematics and Statistics 9, 1129105 (2023)。[75]T. Stankovski、T. Pereira、P. V. McClintock 和 A. Stefanovska, 《耦合函数：对动力学相互作用机制的普适洞见》, Reviews of Modern Physics 89, 045001 (2017)。[76] Y. Kuramoto, 《化学湍流》（施普林格，1984）。[77]Y. Zou、R. V. Donner 和 J. Kurths, 《关于相位相干与非相干混沌的几何和动力学视角》, Chaos 22 (2012)。

[78]E. Ott 和 T. M. Antonsen, 《全局耦合振子大型系统的低维行为》, Chaos **18** (2008)。[79]D. M. Abrams、R. Mirollo、S. H. Strogatz 和 D. A. Wiley, 《耦合振子奇美拉态的可解模型》, Physical review letters **101**, 084103 (2008)。[80]M. J. Panaggio、D. M. Abrams、P. Ashwin 和 C. R. Laing, 《相位振子网络中的奇美拉态：两个小型群体的情形》, Physical Review E**93**, 012218 (2016)。[81]E. Sch¨oll, 《复杂网络中的同步模式与奇美拉态：拓扑结构与动力学的相互作用》, 《欧洲物理学杂志专题》**225**, 891–919 (2016)。[82]I. Omelchenko、A. Provata、J. Hizanidis、E. Sch¨oll 和 P. H¨ovel, 《耦合 FitzHugh-Nagumo 振子的奇美拉态鲁棒性》, Physical Review E**91**, 022917 (2015)。[83]P. Holme 和 M. E. Newman, 《网络与意见协同演化中的非平衡相变》, Physical Review E **74**, 056108 (2006)。[84]L. Horstmeyer 和 C. Kuehn, 《单纯复形上的自适应投票者模型》, Physical Review E**101**, 022305 (2020)。[85]V. Sood、T. Antal 和 S. Redner, 《异质网络上的投票者模型》, Physical Review E**77**, 041121 (2008)。[86]M. H. DeGroot, 《达成共识》, Journal of the American Statistical Association **69**, 118–121 (1974)。[87]N. E. Friedkin 和 E. C. Johnsen, 《社会影响与意见》, Journal of Mathematical Sociology **15**, 193–206 (1990)。[88]H. Rainer 和 U. Krause, 《意见动力学与有限信任：模型、分析与模拟》, Journal of Artificial Societies and Social Simulation **5**(2002)。[89]G. Deffuant、D. Neau、F. Amblard 和 G. Weisbuch, 《相互作用的主体之间的信念交流》, 《复杂系统进展》**3**, 87–98 (2000)。[90]G. Weisbuch, 《有限信任与社交网络》, 《欧洲物理学杂志B辑》**38**, 339–343 (2004)。[91]A. Hickok、Y. Kureh、H. Z. Brooks、M. Feng 和 M. A. Porter, 《超图上意见动力学的有限信任模型》, 《SIAM应用动力系统杂志》**21**, 1–32 (2022)。[92]D. O. Sears 和 J. L. Freedman, 《对信息的选择性接触：批判性综述》, 《公众舆论季刊》**31**, 194–213 (1967)。[93]X. F. Meng、R. A. Van Gorder 和 M. A. Porter, 《各种网络上的有限信任模型中的意见形成与分布》, Physical Review E**97**, 022312 (2018)。[94]A. Di Mare 和 V. Latora, 《基于博弈论的意见形成模型》, 《现代物理学国际杂志C辑》**18**, 1377–1395 (2007)。[95]H. Noorazar、K. R. Vixie、A. Talebanpour 和 Y. Hu, 《从经典到现代的意见动力学》, 《现代物理学国际杂志C辑》 **31**, 2050101 (2020)。

[96]S. St Luce 和 H. Sayama, 《使用基于网络的方法分析和可视化高维动力系统的相空间》, Complexity **2022**, 3937475 (2022).[97]A. M. Unakafov 和 K. Keller, 《序数模式的条件熵》, Physica D: Nonlinear Phenomena **269**, 94–102 (2014).[98]Z. Shahriari、S. D. Algar、D. M.Walker 和 M.Small, 《序数庞加莱截面：从时间序列的序数分割重建首次返回映射》, Chaos **33** (2023).