arXiv:2602.05033v1 [cs.LG] 4 Feb 2026

一般几何下因果表示与随机建模

Jiaxu Ren，[1,3] Yixin Wang，[2] 和 Biwei Huang[3]

[1]CDS，库朗数学研究所[2]密歇根大学统计学系[3]
哈利乔格鲁数据科学研究所，加州大学圣迭戈分校

2026年2月6日

摘要

从观测数据中学习有意义的因果表示，已成为一项关键任务，有助于推动机器学习应用，并促进气候科学、生物学和物理学等领域的科学发现。这一过程涉及从低层次观测数据中解缠高层次潜变量及其因果关系。该领域以往能够实现可识别性的研究，通常聚焦于观测数据独立同分布或遵循潜在离散时间过程的情形。然而，许多现实场景需要识别作为连续时间随机过程的潜变量（例如，多元点过程）。为此，我们针对连续时间潜在随机点过程，开发了可识别的因果表征学习方法。我们通过分析参数空间的几何结构来研究其可识别性。此外，我们开发了 MUTATE，这是一种可识别的变分自编码器框架，配备时间自适应转移模块，用于推断随机动力学。在模拟研究和实证研究中，我们发现 MUTATE 能够有效回答科学问题，例如基因组学中的突变积累，以及时变动力学如何驱动神经元尖峰触发。

1 引言

从观测数据中推断变量之间的因果关系，能够发挥机器学习推动科学发现的潜力，因为这有助于揭示仅凭观测分布无法识别的底层机制 [Pearl, 2009]。然而，我们往往无法获取因果变量，只能获得高维感知数据；因果变量及其结构未知，因此需要通过学习来确定。但这些潜在因果变量通常不可识别 [Hyv¨arinen and Pajunen, 1999, Sorrenson 等人, 2020]。近年来，越来越多关于潜在因果表征解耦的研究提出了可识别性保证，并提出了估计潜在因果变量的方法。其中的开创性工作利用多源数据所产生的潜在分布的充分变异性，确立了可识别性 [Yao et al., 2022a, Song 等人, 2023]，辅助变量 [Hyv¨arinen and Pajunen, 1999, Hyvarinen and Morioka, 2016, 2017, Hyvarinen 等人, 2019]，，或对潜在因果图进行干预 [Ahuja 等人, 2023, Squires 等人, 2023, Jiang and Aragam, 2023, Bing 等人, 2024, Buchholz 等人, 2023]。

上述大多数近期研究旨在恢复遵循离散时间过程的潜在因果变量 [Yao 等人, 2022a, Song 等人, 2023] 并且这些变量由可逆函数混合。然而，在实践中，许多感兴趣的潜在因果变量都是连续时间过程；而由随机过程或随机微分方程系统驱动的潜在连续时间因果变量研究鲜受关注，尤其是在混合函数不可逆且更加一般的情况下[1]。例如

[1]按照代数几何中的标准用法，“ X 的一个一般点具有性质 P”意味着存在一个稠密开子集 $U \subseteq X$，使得 U 中的每一点都具有性质 P [参见 Eisenbud 和 Harris, 2000, 第 I 章]

例如，在视频监控系统中，摄像机经过策略性布置，用于检测和遏制犯罪、防范可能危及公众的威胁，并在自然灾害和人为灾难发生时管理应急响应 [Lima 等人，2020年，Bacry 和 Muzy，2014b，Bacry 等人，2012]。在生物学中，癌症等致命疾病主要由克隆扩增过程中驱动基因内多次累积突变所致。在神经科学中，潜在事件动力学会触发生物信号 [Reynaud- Bouret 和 Schbath，2010年，Lorch 等人， 2024]近几十年来，通过表征识别癌症相关突变基因并追踪其行为已受到越来越多的重视 [Torkamani 和 Schork，2009年，Bailey 等人，2018年，Nourbakhsh 等人， 2024]。鉴于这些应用领域蕴含的实际潜力，我们研究连续时间潜在随机点过程及其因果结构何时具有可识别性，并开发用于从高维数据中学习这些潜在动力学的算法。

贡献。

- 我们建立了首个必要且充分的条件，以保证在一般的非可逆混合下，潜在点过程的完全可识别性。
- 我们提出了 MUTATE，这是一种新颖的、具有可识别性的变分自编码方法，用于学习随机点过程的因果表征。

2 相关工作

因果解耦与时间序列学习。 尽管时间序列的估计和预测一直是传统统计学与现代机器学习中的经典问题，但表征学习开辟了新的途径，可利用潜在信息更好地刻画时间序列数据 [Wu 等人，2021年，Liu 等人，2024]。近年来，学习时间序列中的因果表征已成为推动科学发现的一种基础方法。这一研究方向主要通过利用非平稳数据来确立因果潜变量的可识别性 [Yao 等人，2022b，a] 以及模块化分布变化 [Song 等人，2023年，Cai 等人， 2024] 并施加稀疏性约束 [Song 等人，2024年，Zhang 等人， 2024]研究潜在转移。这些工作利用充分变异性解决了潜在因果模型的可辨识性问题，而这种变异性可来自适当的干预或被动分布变化。另一条研究路线则侧重于学习潜变量之间潜在的因果图。

学习随机过程中的因果影响。 尽管学习因果关系仍然比因果发现或表征学习更具挑战性，但已有多项工作尝试弥合这些领域之间的鸿沟。本文回顾了将因果学习与随机建模联系起来的现有方法。我们的范围并不局限于基于随机过程的因果表征学习，而是扩展到与任一领域密切相关的更广泛问题集合。

动力系统因果学习的一个代表性方向是研究格兰杰因果关系——与结构严格的因果模型相比，这是一种更宽泛且较为宽松的概念 [Achab 等人， 2018]。人们普遍认为，在此类系统中实现完整的因果恢复是不可能的。因此，即使是关于随机过程的最新研究，也只能确定一个点过程 a 相对于另一个过程是否具有格兰杰因果关系 b，通常通过 局部独立性 和 δ-分离规则 [Didelez， 2008]。另一个活跃的研究方向是动力系统中的可识别性 [Lippe 等人，2023]。然而，据我们所知，这些模型中没有一个能够为高度动态的系统（如自激发过程或更一般的随机过程）提供可证明的保证。

因果表示与动力系统之间的联系也通过常微分方程（ODEs） [Yao 等人、 2024]进行了探索。从技术上说，这些方法只能恢复一组难以在潜在空间中解释为因果关系的参数，或者至多允许观测变量呈现随机动力学。最近，人们提出了因果扩散模型 [Karimi Mamaghan 等人、2024年、Lorch 等人、 2024], ，但它们在很大程度上仍将扩散视为标准的去噪过程，因此无法实现结构良好的随机潜在因果表示。

另一个重要的研究方向是研究对随机过程的干预及其相应的干预后分布，这些分布构成了因果推断的基础 [Sokol，2013年，Bongers 等人，2018年，2022，Boeken 和 Mooij，2024年，Lorch 等人， 2024]。首次尝试将因果解释引入随机微分方程（SDEs）是由 [Sokol 和 Hansen， 2014],

其中，干预被定义为移除随机微分方程（SDEs）中的单个变量。他们表明，随机微分方程（SDEs）中的因果原理可以形式化为干预，而由此产生的干预后分布则可通过无穷小生成元加以识别。然而，这类干预过于受限，无法刻画更复杂的动力学情景。在这一研究方向的早期工作之后，[Bongers 等人，2018] 进一步发展了通过最小化扩散平稳性偏差来估计平稳因果模型的方法。不过，他们仅考虑了观测到的扩散过程，这些过程对来自改变漂移项的软干预所产生的因果效应进行建模。

3 问题设定：随机点过程下的因果表示

3.1 预备知识与符号说明

设 $O_t \in \mathbb{R}^n$ 为可观测数据，$Z_t \in \mathbb{R}^p$ 为具有独立噪声 $\epsilon_t \in \mathbb{R}^p$的潜在因果过程。$O_t$ 由潜在点过程 Z_t 经由未知的任意混合函数 f生成。多维数组 $A^{\otimes d}$ 表示张量／克罗内克积。在时间过程中，$\Phi \in \mathbb{R}^{p\times p}$ 表示转移算子（例如自回归系数矩阵或连续核矩阵），符号 $\star$表示带核效应的卷积算子。我们假设存在概率空间 $(S, \mathcal{B}(S)\ \mathbb{P})$，其中 S 是波兰空间（即完备可分度量空间），$\mathcal{B}(S)$是 Borel σ-代数，$\mathbb{P}$ 是概率测度，μ 是一般测度（例如噪声或强度的测度）。$\mathcal{F}_t$ 是过程截至时间 t 的自然滤过。令 K 表示特征零的代数闭域 [2] 。在整个理论证明中，除非另有说明，我们都在该域 K上进行。

3.2 随机点过程的生成模型

本文始终考虑一类非齐次随机过程（Hawkes过程），其动力学由如下定义的条件强度支配。

定义 1（条件强度，非正式定义 [Bacry 和 Muzy，2014a]）。假设有一组随时间随机演化并表现出自激动力学的潜在过程。具体而言，令 $Z_t := N_t$ 表示截至时间 t 的累积计数过程。我们用 $i \leftarrow j$ 表示过程 j 对 i 施加影响。据此，过程 i 在时间 t 处的条件强度为

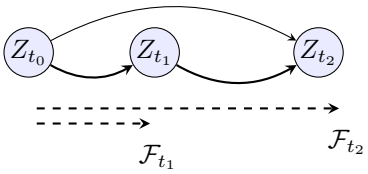


图3.1：递增滤过中的信息损失可视化。

$$\lambda_t^i = \mu_i + \sum_j \int_0^t \phi_{i\leftarrow j}(t-s)\, N_s(\Delta)^j,$$

其中，$\mu_i \in U$ 是基线速率，$\phi_{i\leftarrow j} \in \Phi$刻画从过程 j 到 i的激发核。计数过程 N_t^i 和条件强度 λ_t^i 满足：$\mathbb{E}[dN_t^i \mid \mathcal{F}_t] = \lambda_t^i dt$，即表示截至 t 时所有可用信息的 σ代数（参见 *A.2.1*）。$\mathcal{F}_t$ 称为该随机过程的滤过，$\mathcal{F}_{t+\Delta}$ 则表示

要使这样的点过程定义良好，需要满足一些非平凡约束，其中之一是平稳条件。这一假设在大多数随机过程文献中被广泛采用，以确保该过程的唯一性。

假设 1. *1.*（平稳增量）过程 $N_t^{(\Delta)}$ 是广义平稳的，即其一阶矩和二阶矩存在且不随时间变化。特别地，强度过程 $\mathbb{E}[\Lambda_t]$ 一致有界，且 $dN_t^{(\Delta)}$ 具有平稳增量。

[2]根据定义，域 k 是 *algebraically closed*，如果每个非恒定多项式 $f(x) \in k[x]$ 在 k中都有根；或者等价地，如果 k 不存在真代数扩张 [参见 Lang，2002，Atiyah 和 Macdonald，1969年，第 v 章]。

2. (K 核可积性）卷积因果核 $\Phi_t \in \mathbb{R}^{p\times p}$ 平方可积，即

$$\int_0^\infty \|\Phi_t\|_F^2 \, dt < \infty,$$

其中 $\|\cdot\|_F$ 表示 *Frobenius* 范数。

现在，我们正式提出随机点过程生成模型的识别问题。我们考虑一组无结构的低层观测数据 $O = (O_t)_{t\le T}$，它们由潜在过程N_t 经任意混合函数 f生成。简而言之，将核矩阵 Φ 和计数过程 N_t 并入标准卷积算子后，生成模型可写为

$$O_t = f(N_t(\Delta)), \qquad \lambda_t = U + \Phi_t \star N_t(\Delta). \tag{1}$$

因此，核心目标是根据样本或观测数据的完整分布 O_t，恢复参数空间 $\Theta := (f,\ N_t,\ \lambda_t,\ \Phi,\ U)$。为确保理论严谨性，我们采用这样一种设定：混合函数的形式、潜在因果过程的数量及其因果结构均完全未知。

4 可辨识性理论

本节将确立潜在因果随机点过程的可识别性。我们首先介绍一族一般的等价类，这是一种能够从给定数据中最大程度识别的模型。随后，我们考察该等价类参数空间的几何性质，以确保能够完整恢复混合映射和核函数，以及参数代数结构中所编码的因果结构。所有详细证明均见附录 B 和 C，关于可识别性推广的讨论见附录 D。

4.1 最大可辨识等价类

我们首先介绍可从离散时间观测中识别出的最大等价类。假设我们观测到一个离散时间观测序列 O_{t_0}，$O_{t_0+\Delta}$，$O_{t_0+2\Delta}$，……，$O_{t_0+k\Delta}$，观测时刻为t_0，$t_0+\Delta$，$t_0+2\Delta$，……，$t_0+k\Delta$。给定线性 Hawkes 型强度，我们得到一个在子序列 Δ 下的离散化潜在过程 $Z_t^{(\Delta)}$，其对应的强度为：$\lambda_t^{(\Delta)} = u + \phi(\Delta) \cdot \Delta dN_t^{(\Delta)}$。不同实现之间的差异源于连续时间动力学与其离散近似项之间的不匹配，即 $\lambda_t^{(\Delta\to 0)} \neq \lambda_t^{(\Delta)}$；这意味着，只有在与观测分辨率相同的离散化尺度 Δ 下生成的潜在过程，才能从 $O_t^{(\Delta)}$ 中恢复。因此，底层潜在动力学的可辨识性受限于由观测分辨率决定的离散时间等价类。为刻画分布层面的变化与动力学，我们认为，对于大多数科学任务而言，恢复潜在过程的分布行为已足够，并且可用于在任何其他 Δ尺度下生成潜在过程。因此，我们只能识别一个等价类，定义如下。

定义 2 (弱收敛等价类)。 设(dN_t, λ_t) 表示真实潜在点过程及其对应的连续时间强度函数。若一对$(Z^{(\Delta)}, \lambda^{(\Delta)})$ 满足以下弱收敛条件，则称其属于 (dN_t, λ_t) 的弱收敛等价类：

$$(Z^{(\Delta)}, \lambda^{(\Delta)}) \xrightarrow[\Delta\to 0]{d} (dN_t, \lambda_t),$$

也就是说，随着分辨率参数 $\Delta \to 0$，估计的潜在过程及其离散时间强度在分布上收敛到真实连续时间过程。

感谢 定义 2，只需找到一个属于该等价类的模型，并证明其可识别性。遵循Kirchner [2016], ，我们通过收敛极限重新审视阶数为l 整数值自回归过程（INAR(l)）与多变量随机点过程之间的密切联系，其中 $l \to \infty$。INAR(∞) 过程定义为变量取整数的无限阶自回归模型

Z_t。具体而言，其形式为：

$$Z_t = \sum_{\tau>0}^{\infty} a_\tau Z_{t-\tau} + \epsilon_t, \tag{2}$$

其中， a_τ 是一个常系数， ϵ 是时间上相互独立的噪声。我们的直觉是，用时不变核 ϕ_t 替代常系数，仍能保证弱收敛到随机点过程 N_t。主要目标是证明，这种替代属于所定义的弱收敛类，正如下列引理所确立的。

引理 1（变分近似中的点过程界定）。 设 $N_t \in \mathbb{R}^p$ 是一个多元点过程，其条件强度函数 λ_t 由式（1）中描述的 ϵ_t 卷积结构所支配，且为均值为零、相互独立的噪声。那么，该强度模型满足以下弱收敛：

$$Z_k^\Delta := \lambda_k^\Delta + \epsilon_k^\Delta \xrightarrow{w} N_k^\Delta, \tag{3}$$

其中，下标 k 表示任意子序列过程， λ_k^Δ 是对应子序列下的强度。

引理 2（收敛到潜在等价类）。 假设 引理 *1* 中的弱收敛条件成立。那么，存在一个潜在过程 Z_t^Δ，使得过程 N_t 及其变分近似在分布上收敛到同一个潜在因果类。形式上，

$$\lim_{\Delta\downarrow 0} N_k^\Delta := Z^\Delta \xrightarrow{d} N_t. \tag{4}$$

这意味着，在无穷小分辨率 Δ下，所估计的过程具有由相同卷积动力学支配的潜在表示。

引理 1 和 2共同确立了连续随机过程在相应弱拓扑下的弱收敛。粗略地说，对于任意紧时间区间 $[a, b]$，原过程在该区间上的一个子序列过程收敛于一个 连续时间 因果点过程 N_t。这种收敛确保 Z^Δ 有效地表示 N_t 并保持所有因果结构。因此，在不失一般性的情况下，我们可以直接研究 Z^Δ，并通过分析相关参数空间的几何来研究其可识别性。

4.2 几何刻画模型可识别性

作为迈向主要结果的 stepping stone，所提出潜在模型的几何刻画了生成模型参数的唯一性。Carreno 等人 [2024] 的研究表明，当且仅当由可用数据定义的系统的解集是零维时，参数空间才能被完全恢复，即该解集由有限多个点组成，并且这些点的数量与潜变量或参数的数量相一致。在几何上，给定有限维观测分布 $P(O_t)$（即相应分布截至任意有限阶的矩和统计量），我们考虑理想[3]

$$\mathcal{I} = \langle P(O_t) - P_\Theta \rangle, \tag{5}$$

而参数的可识别性 Θ 要求 $\mathcal{I}$ 的维数为零。直观而言，生成模型的参数空间$\Theta := (f, Z^\Delta)$（即混合映射 f 与 Z^Δ的完整参数）对应于 $\mathcal{I}$的零点集；在这种情况下，它不能位于任何更高维的超曲面上，如图4.1所示。实际上，完整分布无法获取，因此我们旨在仅利用部分分布信息来确立可识别性。遵循Wang 和 Seigal [2024], 的做法，我们选取观测分布的累积量作为中介，以研究参数空间的几何 Θ。无穷阶累积量是一种重要的代数几何特征，因为它能够精确编码整个分布，包括变量间分量内依赖关系和时间依赖 [Achab 等人，2018年， Jovanovi´c

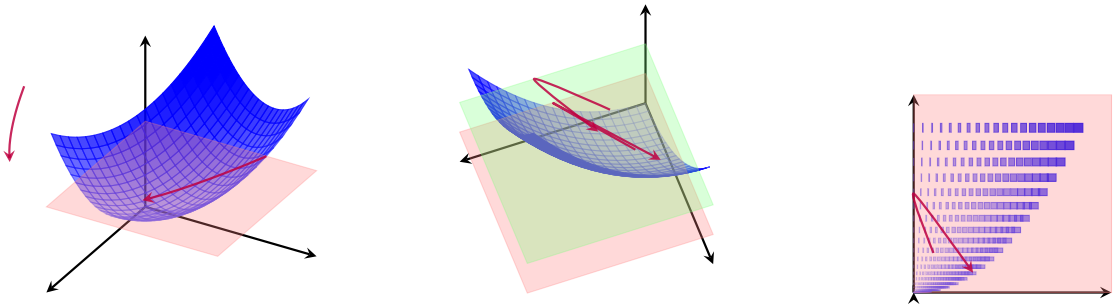
[3]下文中，我们将“理想”一词泛用于描述有限阶分布信息。严格来说，理想应当是在某个环上有限生成的。

等人，2015年，Landsberg， 2011]。因此，高阶累积量能够以相同的阶数捕捉分布的因果结构，从而支持对干预效应进行精细的数学分析，而不仅局限于传统的均值和方差变化。

在一般的（非高斯）条件下，Carreno 等人 [2024] 证明了，如果观测变量满足如下形式的完全线性因果模型 $X = FZ$ 其中 $Z = AZ + \epsilon$，那么 d阶累积量 X，记作 $\kappa_d(X)$，可以唯一分解为

$$\kappa_d(X) = \sum_{j=1}^{p} \kappa_d(\epsilon) \cdot K_j \otimes \cdots \otimes K_j, \quad K = F(\mathbb{I} - A)^{-1}.$$

与公式（5）相比，上述方程导出了一个简化的理想 $\mathcal{I} := \langle K（\mathbb{I} - A） - F \rangle$，其中 A 和F 被视为一般不定元。因此，编码于 K 中的参数空间在缩放和置换意义下是可识别的。这些参数的线性混合的可识别性，等价于证明由 κ_d（X）定义的代数簇是非 p-缺陷的 [Chiantini 等人， 2017]。参数几何与可识别性之间的这一联系，使我们能够阐明 INAR 等价类在线性情形下的可识别性，我们将在下一节中对此进行展开。



(a) (b) (c)

图4.1：具有一般超平面的曲面（平滑梯度网格，三维效果）。(a) 含一个超平面的曲面；(b) 含两个超平面的 Veronese；(c) 含有限点的曲面

4.3 识别线性混合

我们首先给出线性情形下的可识别性结果。具体而言，我们证明可以从线性混合 $\Theta := (f, \Phi, U)$中恢复完整生成模型 $O_t = FZ_t^{\triangle}$，其中 $Z_t^{\triangle}$表示第4.1节中引入的弱等价类。为清晰起见，只要上下文明确，我们就省略子序列记号。根据引理 2，该潜在过程满足 $Z_t = \lambda_t + \epsilon$，其中 $\lambda_t = U + \Phi \star Z_t$。

一般而言，当从完整分布 $P(O_t)$中识别因果变量时，需要潜在分布具有充分变异性Yao 等人，2022b，Song 等人，2023年，Zhang 等人， 2024]。与此不同，在这种情形下，我们只能观测到 Z_t 的线性混合，潜在分布的任何变异性都只能源自参数 $\Theta := (F, \Phi, U)$。这一观察促使我们利用累积量的代数结构，它能够自然地捕捉分布变异性，并对可识别性作出清晰的解释。

如果 O_t的 d阶累积量（记为 $\kappa_d(O_t)$）也具有唯一分解，那么约化理想 $\mathcal{I}' := \langle K(\mathbb{I} - \Phi) - F \rangle$（其中 $K = F(\mathbb{I} - \Phi)^{-1}$）的次数至多为一，并且可分解为有限个素理想的乘积。解空间的维数与相关理想 $\mathcal{I}'$ 的维数一致，从而确定完整生成模型的可识别性。为此，我们需要控制相关参数空间的几何。

假设 2。

1. F 几乎必然具有一般的可能最大秩。

2. 存在一个集合 $D := \{d \in \mathbb{Z} \mid \kappa_{d+1}(O_t) = 0\}$，使得集合 $\{\kappa_d(O_t)\}_{d\in D}$ 至少包含p 个非零元素。

3. 理想 $\mathcal{I} := \langle F - K^{(k)}(\mathbb{I}_p - \Phi^{(k)}) \quad k = 1, 2, \ldots, p\rangle$ 满足空间 Θ 是零维的。

这三个条件既非彼此独立，也非彼此平行，而是按顺序相互关联：每个条件都建立在前一个条件的基础上，从而逐步实现生成模型的完全可识别性。也就是说，每个观测分量 $\kappa_d(O_t)$ 都具有有限深度 d，且截至此阶的非零信息足够丰富，能够通过张量分解确保可识别性：所提出的秩条件 (1) 是经典条件，可使每个 d 阶张量 $\kappa_d(O_t)$ 几乎处处具有唯一分解，从而唯一恢复分量 $K_j := (F(\mathbb{I} - \Phi)^{-1})^{(:,j)}_{t-s}$；对所有 $j \in [p]$ 和所有 $s \in S := \{s : s < t\}$，结果均在置换和缩放意义下成立。对于 Kruskal 秩条件 krank(V) [Kruskal, 1977年，Lovitz 和 Petrov， 2023], 该结论成立；这一条件要求每个秩一分量 $(K_j)^{\otimes d}$ 在环境空间中都没有共线列[4][Wang 和 Seigal，2024年，Wang 等人， 2025]。条件 (2) 保证 $(K_j)^{\otimes d}$ 的外部空间的线性张成中至少存在 p 个这样的非零点，因此 F 和 Φ 构成一个零维参数空间。由条件 (1)、(2) 和 (3) 的综合结果，我们实现了模型的完全可识别性。

定理 1 （等价类的线性可识别性）。 在假设 *1* 与*2*下，潜在点过程的弱收敛等价类和因果结构在逐分量缩放与置换意义下是可识别的。

我们进一步指出，一旦生成模型被识别，因果变量便可从模型 $\Theta := (F, \Phi, U)$ 中采样。

4.4 识别一般非线性等价类

在本节中，我们放宽线性假设，并证明观测流形的代数结构同样能够在任意非线性变换下保证可识别性（参见 一般的 F 于 假设 2）。具体而言，我们现在假设 f 是一般的（可能为单射），从而映射 $f : Z_t \mapsto O_t$ 是良定义的，并保持潜在表示的可区分性。我们将此形式化于 假设 3。

假设 3。

1. 设 f 为一个一般的 C^d 映射。

2. 在混合累积量流形上，该系统存在一种线性退化，使得假设 *2*对某个阶数 d 和 p 成立。

该 C^d 正则性假设严格弱于要求 f 为微分同胚的条件；诸多先前研究均要求满足这一条件，例如 [Song 等人的研究 2023]。即使潜在空间中存在方向坍缩，所诱导的代数结构仍可能忠实地将本质的依赖关系传递到观测域。不出所料，当混合函数 f的雅可比矩阵 J_f 中的元素是足够自由的函数（例如多项式或C^d 函数）时，其一般性自然成立。事实上，使 J_f 不满秩的函数集合对应于函数空间中的一个真代数子簇，因此，几乎所有 f 的选取都会得到满秩雅可比矩阵。因此，与雅可比矩阵元素为常数相比，若其元素为函数，则 f 在代数几何意义上更有可能是一般的。

定理 2 (等价类的完全非线性可识别性)。 在 假设 *1* 和 假设 *3*下，潜在点过程的弱收敛等价类及其因果结构在逐分量变换和置换意义下是可识别的。

[4]该 *ambient space* 是给定簇或概形所嵌入的更高维空间，通常为 $\mathbb{A}^n$ 或 $\mathbb{P}^n$ （代数几何中）。详见附录C.1

定理的直观解释。 累积量通过非线性变换传递因果结构，使得我们能够从混合流形 O_t上的部分代数几何信息中恢复潜在依赖。我们区分两种情况：能够获取完整的观测分布 $P(O_t$，或只能获取从受限分布 $P^R(O_t$ 中抽取的样本。在这两种情况下，都可以检验累积量的几何性质：允许分布信息在某一阶内有所损失，累积量仍存在唯一的线性退化。该退化以典范方式确定 Veronese 簇的射影嵌入，且该嵌入在逐分量缩放和置换意义下唯一。此外，在条件（2）下，给定多个环境（干预或变异性）后，可推出完整生成模型不可能位于任何正维超曲面内。

值得强调的是，假设 2 中提出的条件并非只是技术性假设，而是共同构成了一组可识别性的充要条件。在给定数据下，不存在其他阶数 $d_0 < d$ 能够使更简单的累积量流形仍足以保证可识别性，除非引入额外数据或干预。具体形式化如下：

定理 3. 定理 *1*中陈述的可识别性结果，*2* 当且仅当满足假设 *2* 中的条件时成立。

4.5 从可识别性到可检验条件

我们的理论证明表明，通过从外围参数空间中排除一个次线性张成，可以显著提高可识别性，从而控制代数参数簇的预期维数。所有条件最初都是在4 的代数几何框架下提出的；这一框架保证了数学严谨性，但在一定程度上妨碍了经验证据的呈现。为了将我们的结果与实践联系起来，我们打算给出若干示例，用以检验零维理想假设何时成立、何时不成立。由于按定义，我们的结果是在完全代数条件下对充分可变性条件的推广，接下来我们将重点讨论干预的包含关系。

命题4.1 （几乎必然有界 dim($I(V)$）。 若对生成核矩阵Φ中的每个元素都施加硬干预，则与簇$\dim(I(V))$ 相关联的理想的维数有界为零。

命题4.2 （有界 dim($I(V)$）的正概率。 若对生成核矩阵 Φ中的多个元素施加软干预，则与簇$\dim(I(V))$ 相关联的理想的维数有界为零。

我们对参数簇最大维数的界表明，即使缺乏足够的内部变异性，也可以通过设计适当的干预方案恢复可识别性。例如，可以通过在任意层级调节核算子的通常行为来实现干预。假设干预会扰动每个核函数的生成参数，我们证明，系统性的干预就足以奏效，而无论干预效应持续的时间长度或时间窗口如何 [Lippe 等人, 2023]。

5 MUTATE：估计等价随机因果过程

受可辨识性理论的启发，我们旨在开发一种实用算法，从观测数据中恢复其潜在的生成因果机制。我们提出 MUTATE（多时间自适应转移编码器），这是一种用于估计潜在多变量随机点过程的新型变分自编码（VAE）框架。我们的目标是获得生成分布及其参数的最优表示，并使其与理论结果保持一致。

与变分自编码器中的常见做法一致，我们使用一对编码器和解码器来参数化 O_t 和 Z_t的边缘分布。借助引理 1 和引理 2，我们可以通过自卷积因果机制 $s_\phi(R_t, \Phi) := s_\phi(\epsilon_t, U, \Phi)$生成潜变量 Z_t 。因此，编码器 $q_\phi(R_t|O_t)$ 将观测 O_t 编码为独立噪声 ϵ 和恒定基线强度 U ，随后将其交由 s_ϕ 进行参数化。最后，解码器 $p_\theta(O_t|s)$根据 s_ϕ重建虚拟观测样本 $\hat{} O_t$ 。有多种方法可以处理 s_ϕ 中的卷积，以对动态因果机制进行建模。

例如，如果假设生成式建模等同于神经网络，则可以使用线性卷积网络 [Kohn等人， 2021, 2023]。而我们利用时频联系来解决这一问题 [Lima 等人，2020年，Bacry 和 Muzy，2014b]。给定编码器输出的噪声嵌入序列 $\epsilon_{t_i}, \cdots, \epsilon_{t_k}$，傅里叶变换会将其映射为 $\epsilon_{s_i}, \cdots, \epsilon_{s_k}$，从而使 $Z_t = (I - \Phi) \star R_t$ 到$Z_s = (I - \Phi)^{-1} R_s$，其中 R_s 结合了独立噪声 ϵ_s 的影响，以及常数 U所对应的附加影响 $2\pi U\delta(k)$。该变换仅在频域中对潜变量 Z_s 进行参数化，无法直接映射回时域分布中的先验构造，而这正是我们优化观测样本 O_t证据所需要的。

因此，我们引入一个概念验证模块，利用神经网络学习核编码因果结构，并通过简单求和评估时域先验分布。神经功率谱密度兼具神经网络的能力，以及我们在可辨识性证明中主要依赖的代数结构：在假设噪声服从独立噪声分布的前提下，对 s_ϕ 进行参数化等价于寻找一个群作用 G，使其作用于噪声的方差矩阵，并满足：

$$S_{Z_s} = G^{-1} \Sigma_{R_s} G^{-H}, \tag{6}$$

其中 Σ_{R_s} 表示驱动过程的谱密度 R_s，它结合了随机噪声 ϵ_t和编码器推断出的基线强度 U。具体而言，基线项仅对谱的零频分量有贡献。这里， A^H 表示 A的厄米共轭（共轭转置）。求解（6）既可解释为针对 Φ_s参数化的神经网络优化问题，也可等价地解释为频域中厄米二次型的代数分解问题。

遵循后一种表述，我们设计了神经功率谱密度模块，通过一个高阶簇可辨识性问题，支持其他变分自编码器组件的联合优化。神经功率谱密度首先在每个频率 s 处，将一个随机选取的元素 $g \in G$（即前述作用于频域的群）作用于嵌入向量 ϵ_s 的功率谱密度矩阵，从而构造二阶谱簇。在附录中，我们证明，这个二阶簇存在正则多元分解，其中潜嵌入方向 G_i 在一般情况下可识别至最小相位。得到 G 后，二阶谱由时域中的总能量分量计算得出；根据维纳–辛钦定理，这些分量可参数化为另一个高斯分布。最后，变换后的高斯分布可根据先验分布 $p(Z_t)$ 进行评估。我们通过最小化观测样本 O_t 的证据（记为证据下界）来优化训练目标：

$$\mathcal{L}_s = \mathbb{E}_{q_\phi(R_t)} \log p_\theta(O_t \mid R_t) - D_{\mathrm{KL}}\big(q_\phi(Z_t) \,\|\, p(G, R_t)\big). \tag{7}$$

训练期间样本批次的总损失为：

$$\mathbb{E}_O^\lambda\{\mathcal{L}_s\} + \lambda\|G\|$$

Neural PSD表述具有若干理想性质。首先，潜变量 Z_t 由 s_ϕ 参数化，并同时由重构误差更新，该误差在训练过程中动态变化。其次，群作用 G 定义了一个在我们的标准假设下完全可识别的射影簇。最后，Neural PSD将Wilson 分解 [Wilson， 1978] 推广到方差矩阵不是单位矩阵的情形，且对于这一更广的可识别性类别不需要额外的正则化损失。

与以往主要依赖时间戳条件独立性来约束潜在结构的框架不同，我们的方法考虑了渐进自适应随机过程的性质。在这类系统中，能够刻画过程内在历史的滤过 $\mathcal{F}_T$定义为$\sigma\left(\bigcup_{0<t<T} \sigma(Z_t^{\Delta})\right)$，并随时间严格增长。如图3.1所示，这种动态扩展的信息结构为可识别性和表征学习带来了独特挑战，而 MUTATE 正是专为应对这些挑战而设计的。此外，为了利用相互独立噪声，神经功率谱密度会自动确保噪声的全局白化。

6 模拟研究

我们模拟多元点过程及其收敛等价类 $\mathcal{Z}_t$，后者在我们的可辨识性理论中得到了广泛研究。我们使用泊松叠加法对所有点过程进行采样（从条件强度 [Cinlarand Agnew，1968，Albin，1982]的上界进行拒绝采样），以模拟条件强度的高度动态变化，并捕捉随机过程中更密集的信息。随后，我们构造相应的收敛类，作为概念验证：针对五种核函数——指数型、幂律、矩形型、简单非线性和灵活混合——在两种噪声机制（异质噪声和高斯混合噪声）下，各采样 20,000 条潜在轨迹。为了展示无结构数据背后的潜在事件，我们还使用 SERGIO 模拟生物数据的随机动力学 [Dibaeinia and Sinha，2020], 这是一种基因调控网络引导的基因表达模拟器，也用于 Lorch 等人的Lorch 等人 [2024] 因果建模。所有观测数据 O_t 均由潜变量通过多层感知机和带泄漏的修正线性单元非线性混合得到。详细的模拟流程见E.5。

为验证我们的可识别性结果，我们将其与几种具有代表性的基线方法进行比较，包括 TDRL [Yao 等人,], BetaVAE [Higgins 等人，2017], SlowVAE [Klindt 等人，2021], 以及 PCL [Hyvarinen 和 Morioka，2017]。其中，PCL 和 TDRL 利用历史信息，并明确约束潜变量之间的条件独立性，以纳入时间依赖关系并恢复潜在动态。相比之下，BetaVAE 和 SlowVAE 假设潜在成分相互独立，并忽略任何时滞机制。

表 1：五种核函数的 MCC 分数及标准差

方法	平均值（↑越大越好）指数型	幂律	矩形型非线性	非参数
TDRL	0.599	0.593±0.028	0.609±0.043	0.618±0.056 0.556±0.016 **0.616**±0.043
BetaVAE	0.141	0.153±0.863 0.128± 0.077	0.128±0.078	0.146±0.108 0.149±0.096
SlowVAE	0.115	0.108±0.075	0.104±0.073	0.104±0.073 0.126±0.074 0.131±0.076
PCL	0.375	0.395±0.034	0.330±0.029	0.330±0.029 0.414±0.028 0.404±0.028
MUTATE（我们的方法）		**0.853**±0.218		98±0.013

评估指标 我们在合成数据集和真实世界数据集上验证了我们的方法。对于选定的基线模型，我们对非线性独立成分分析中常用的因子化推断模块进行改造，使其支持更深层次的内在滤过组合，从而能够对复杂的真实世界动态进行建模。在合成数据集上，我们使用平均相关系数（MCC）评估可识别性。MCC 是一种标准指标，用于量化潜变量的恢复准确度。具体而言，MCC 通过求解线性分配问题来处理置换不确定性，然后对真实值与推断出的潜在成分之间的绝对相关系数取平均得到。

结果 所有基线方法和我们模型的性能如表1 所示，扩展结果见表2。训练过程中，BetaVAE 和 SlowVAE 往往会过早收敛，通常在第一个 epoch 内就达到局部最优并触发提前停止。这一现象凸显了它们在建模时间结构方面的局限，而时间结构对于识别潜在事件驱动过程至关重要。当滞后模块设置得较长时（实验中我们使用 $L = 9$），TDRL 表现尚可，因为它能够利用较短期的上下文信息。我们发现，只需向合成数据集添加域索引，或调节会改变潜在空间中边对的分布偏移，就能将我们的可识别性结果直接应用于先前的框架。不过，我们也认识到，完全非参数设定难以解释，因为我们的可识别性理论并未涵盖这一点。

7 结语

本研究将因果表征学习框架扩展到随机因果动态（即多变量 Hawkes 过程），这一主题尚未在当前的因果表征学习文献中得到探讨。我们提出了一种新视角，认为

随机过程的一个分支可通过 INAR(∞) 表示及其在弱拓扑上的弱收敛，视为相应的等价类。我们证明，在足够一般的条件下，生成模型与潜在因果模型均可被完全识别，且仅存在逐分量变换和置换的不确定性。我们的理论结果弥合了随机建模与因果表示之间的差距。我们还提出了一种名为时间自适应转移编码的新框架，以忠实地估计潜在过程。然而，我们的工作避开了完全非参数核这一最棘手、最复杂的情形；在实证实践中，可以用更简单的核替代它。未来的研究方向可能包括解决这一条件问题，以及针对能够引出丰富科学问题的随机微分过程开展因果表征学习。

参考文献

Massil Achab、Emmanuel Bacry、St´ephane Ga¨ıffas、Iacopo Mastromatteo 和 Jean-Fran¸cois Muzy。《从多元 Hawkes 积分累积量中揭示因果关系》。2018年。Kartik Ahuja、Divyat Mahajan、Yixin Wang 和 Yoshua Bengio。《干预式因果表示学习》。载于《国际机器学习会议》，第 372–407 页。PMLR，2023年7月。URL https://proceedings.mlr.press/v202/ahuja23a.html。ISSN：2640-3498。Susan L Albin。《排队系统中叠加到达过程的泊松近似》。《管理科学》，28(2):126–137，1982年。M. F. Atiyah 和 I. G. Macdonald。《交换代数导论》。Addison-Wesley，1969年。ISBN 978-0201407518。E. Bacry、K. Dayri 和 J. F. Muzy。《对称 Hawkes 过程的非参数核估计及其在高频金融数据中的应用》。《欧洲物理学杂志B》，85(5):157，2012年5月。ISSN 1434-6028，1434-6036。doi：10.1140/epjb/e2012-21005-8。URL http://arxiv.org/abs/1112.1838。arXiv:1112.1838 [物理学，q-fin]。Emmanuel Bacry 和 Jean-Francois Muzy。《Hawkes 过程的二阶统计特征刻画与非参数估计》，2014年1月。URL https://arxiv.org/abs/1401.0903v2。Emmanuel Bacry 和 Jean-Fran¸cois Muzy。《用于价格和交易高频动态的 Hawkes 模型》。《定量金融》，14(7):1147–1166，2014b。ISSN 1469-7688，1469-7696。doi：10.1080/14697688. 2014.897000。URL http://www.tandfonline.com/doi/abs/10.1080/14697688.2014.897000。Emmanuel Bacry、Iacopo Mastromatteo 和 Jean-Fran¸cois Muzy。《金融中的 Hawkes 过程》。《市场微观结构与流动性》，01(01):1550005，2015年6月。ISSN 2382-6266，2424-8037。doi：10.1142/ S2382626615500057。URL https://www.worldscientific.com/doi/abs/10.1142/S2382626615500057。Matthew H. Bailey、Collin Tokheim、Eduard Porta-Pardo、Sohini Sengupta 等，Bertrand。《癌症驱动基因与突变的全面特征分析》。《细胞》，173(2):371–385.e18，2018年4月。ISSN 00928674。doi：10.1016/j.cell.2018.02.060。URL https://linkinghub.elsevier.com/retrieve/pii/S009286741830237X。Patrick Billingsley。《概率测度的收敛》。Wiley 概率与统计丛书·概率与统计分册。Wiley，纽约、魏恩海姆，第2版，1999年。ISBN 978-0-471-19745-4 978-0-470-31780-8。Simon Bing、Urmi Ninad、Jonas Wahl 和 Jakob Runge。《从多节点干预中识别线性混合因果表示》。载于《因果学习与推理》，第 843–867 页。PMLR，2024 年 3 月。URL https://proceedings.mlr.press/v236/bing24a.html。ISSN：2640-3498。Philip Boeken 和 Joris M. Mooij。《动态结构因果模型》，2024年6月。URL https://arxiv.org/abs/2406.01161v2。

Stephan Bongers、Tineke Blom 和 Joris M. Mooij。《动力系统的因果建模》，2018年3月。URL https://arxiv.org/abs/1803.08784v4。

Stephan Bongers、Tineke Blom和Joris M. Mooij。动力系统的因果建模，2022年3月。网址 http://arxiv.org/abs/1803.08784。arXiv:1803.08784 [cs]。

Pierre Br´emaud和Laurent Massouli´e。非线性霍克斯过程的稳定性。《概率年鉴》，24(3)，1996年7月。ISSN 0091-1798。doi: 10.1214/aop/1065725193。网址 https://projecteuclid.org/journals/annals-of-probability/volume-24/issue-3/Stability-of-nonlinear-Hawkes-processes/10.1214/aop/1065725193.full。

Simon Buchholz、Goutham Rajendran、Elan Rosenfeld、Bryon Aragam、Bernhard Sch¨olkopf 和 Pradeep Ravikumar。从干预中学习线性因果表征：一般非线性混合。神经信息处理系统进展，36：45419–45462，2023年12月。URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/8e5de4cb639ef718f44060dc257cb04f-Abstract-Conference.html。Ruichu Cai、Zhifang Jiang、Zijian Li、Weilin Chen、Xuexin Chen、Zhifeng Hao、Yifan Shen、Guangyi Chen 和 Kun Zhang。从正交性到依赖关系：为多模态时间序列传感信号学习解耦表征，2024年5月。URL https://arxiv.org/abs/2405.16083v1。Paula Leyes Carreno、Chiara Meroni 和 Anna Seigal。通过高阶累积量实现线性因果解耦。arXiv 预印本 arXiv:2407.04605，2024年。Luca Chiantini 和 Ciro Ciliberto。弱缺陷簇。《美国数学会汇刊》，354(1)：151–178，2002。Luca Chiantini、Giorgio Ottaviani 和 Nick Vannieuwenhoven。关于次一般秩对称张量的一般可识别性。《美国数学会汇刊》，369(6)：4021–4042，2017年。Erhan Cinlar 和 RA Agnew。关于点过程的叠加。《皇家统计学会会刊：B辑（方法学）》，30(3)：576–581，1968年。Daryl J. Daley 和 David Vere-Jones。点过程理论导论。1：初等理论与方法。施普林格，纽约，第二版，第二次修订印刷版，2005年。ISBN 978-0-387-95541-4。Payam Dibaeinia 和 Saurabh Sinha。SERGIO：由基因调控网络引导的单细胞表达模拟器。《细胞系统》，11(3)：252–271，2020年。Vanessa Didelez。基于局部独立性的标记点过程图模型。《皇家统计学会会刊B辑：统计方法学》，70(1)：245–264，2008年2月。ISSN 1369-7412，1467-9868。doi：10.1111/j.1467-9868.2007.00634.x。URL https://academic.oup.com/jrsssb/article/ 70/1/245/7109395。David Eisenbud 和 Joe Harris。概形几何。施普林格，2000年。Robin Hartshorne。代数几何，数学研究生教材第52卷。施普林格，1977年。ISBN 978-0387902441。Alan G Hawkes。一些自激发与相互激发点过程的谱。1971。doi：10.1093/ biomet/58.1.83。Irina Higgins、Loic Matthey、Arka Pal、Christopher Burgess、Xavier Glorot、Matthew Botvinick、Shakir Mohamed 和 Alexander Lerchner。beta-VAE：利用受约束变分框架学习基本视觉概念。载于国际表征学习会议，2017年。Aapo Hyvarinen 和 Hiroshi Morioka。通过时间对比学习和非线性独立成分分析进行无监督特征提取。神经信息处理系统进展，29，2016年。URL https://proceedings.neurips.cc/paper_files/paper/2016/hash/ d305281faf947ca7acade9ad5c8c818c-Abstract.html。

Aapo Hyvarinen 和 Hiroshi Morioka。时间依赖的平稳源的非线性独立成分分析。载于人工智能与统计，页码460–469。PMLR，2017年4月。URL https://proceedings. mlr.press/v54/hyvarinen17a.html。ISSN：2640-3498。Aapo Hyvarinen、Hiroaki Sasaki 和 Richard Turner。利用辅助变量和广义对比学习进行非线性独立成分分析。载于第22届人工智能与统计国际会议，页码859–868。PMLR，2019年4月。URL https://proceedings.mlr.press/v89/hyvarinen19a.html。ISSN：2640-3498。Aapo Hyv¨arinen 和 Petteri Pajunen。非线性独立成分分析：存在性与唯一性结果。《神经网络》，12(3)：429–439，1999年4月。ISSN 08936080。doi：10.1016/S0893-6080(98)00140-3。URL https://linkinghub.elsevier.com/retrieve/pii/S0893608098001403。Yibo Jiang 和 Bryon Aragam。利用未知干预学习非参数潜在因果图。神经信息处理系统进展，36：60468–60513，2023年12月。URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/bdeab378efe6eb289714e2a5abc6ed42-Abstract-Conference.html。Jikai Jin 和 Vasilis Syrgkanis。从一般环境中学习因果表征：可识别性与内在歧义，2023年11月。URL https://arxiv.org/abs/2311.12267v2。Stojan Jovanovi´c、John Hertz 和 Stefan Rotter。霍克斯点过程的累积量。《物理评论E》，91(4)：042802，2015年4月。ISSN 1539-3755，1550-2376。doi：10.1103/PhysRevE.91.042802。URL https: //link.aps.org/doi/10.1103/PhysRevE.91.042802。Amir Mohammad Karimi Mamaghan、Andrea Dittadi、Stefan Bauer、Karl Henrik Johansson 和 Francesco Quinzan。基于扩散的因果表征学习。26(7)：556，2024年。ISSN 1099-4300。doi：10.3390/e26070556。URL https://www.mdpi.com/1099-4300/26/7/556。Matthias Kirchner。Hawkes 和 INAR(∞) 过程。《随机过程及其应用》，126 (8)：2494–2525，2016 年 8 月。ISSN 03044149。doi：10.1016/j.spa.2016.02.008。URL http://arxiv.org/ abs/1509.02007。arXiv：1509.02007 [math]。David Klindt、Lukas Schott、Yash Sharma、Ivan Ustyuzhaninov、Wieland Brendel、Matthias Bethge 和 Dylan Paiton。利用时间稀疏编码，在自然数据中迈向非线性解缠，2021年3月。URL http://arxiv.org/abs/2007.10930。arXiv：2007.10930 [stat]。Kathl´en Kohn、Thomas Merkh、Guido Mont´ufar 和 Matthew Trager。线性卷积网络的几何，2021年8月。URL https://arxiv.org/abs/2108.01538v2。Kathl´en Kohn、Guido Mont´ufar、Vahid Shahverdi 和 Matthew Trager。线性卷积网络的函数空间与临界点，2023年4月。URL https://arxiv.org/abs/2304.05752v2。Joseph B. Kruskal。三维数组：秩与三线性分解的唯一性，以及其在算术复杂度和统计学中的应用。《线性代数及其应用》，18(2)：95–138，1977年。Joseph M Landsberg。张量：几何与应用：几何与应用，第128卷。美国数学学会，2011年。Serge Lang。代数，数学研究生教材第211卷。施普林格，第三版，2002年。ISBN 978-0387953854。Vinicius Lima、Fernanda Jaiara Dellajustina、Renan O. Shimoura、Mauricio Girardi-Schappo、Nilton L. Kamiji、Rodrigo F. O. Pena 和 Antonio C. Roque。频域中的格兰杰因果关系：推导与应用。Revista Brasileira de Ensino de F´ısica，42：e20200007，2020年。ISSN 1806-9126，1806-1117。doi：10.1590/1806-9126-RBEF-2020-0007。URL http://arxiv.org/abs/2106.03990。arXiv：2106.03990[physics，q-bio，stat]。

Phillip Lippe、Sara Magliacane、Sindy L¨owe、Yuki M. Asano、Taco Cohen 和 Efstratios Gavves。《交互系统中瞬时与时间效应的因果表征学习》。2023年。网址http://arxiv.org/abs/2206.06169。

Yong Liu、Tengge Hu、Haoran Zhang、Haixu Wu、Shiyu Wang、Lintao Ma 和 Mingsheng Long。《iTransformer：倒置Transformer对于时间序列预测十分有效》，2024年3月。网址 http://arxiv.org/abs/2310.06625。arXiv:2310.06625 [cs]。

Lars Lorch、Andreas Krause 和 Bernhard Sch¨olkopf。《平稳扩散的因果建模》。载于 Sanjoy Dasgupta、Stephan Mandt 和 Yingzhen Li（编），《第*27*届人工智能与统计国际会议论文集》，第238卷，收录于《机器学习研究论文集》，第1927–1935页。PMLR，2024年5月2–4日。网址https://proceedings.mlr.press/v238/lorch24a.html。

Benjamin Lovitz 和 FedorPetrov。张量分解中 Kruskal 定理的推广。载于*Forumof Mathematics, Sigma*，第 11 卷，文章 e27。剑桥大学出版社，2023年。

Mona Nourbakhsh、Kristine Degn、Astrid Saksager、Matteo Tiberti 和 Elena Papaleo。癌症驱动基因与突变的预测：整合计算框架的潜力。生物信息学简报，25(2):bbad519，2024年1月。ISSN 1467-5463，1477-4054。doi: 10.1093/bib/bbad519。URL https://academic.oup.com/bib/article/doi/10.1093/bib/bbad519/7584784。

Judea Pearl。因果关系。剑桥大学出版社，2009年。

Patricia Reynaud-Bouret 和 Sophie Schbath。Hawkes 过程的自适应估计；在基因组分析中的应用。统计学年刊，38(5)，2010年10月。ISSN 0090-5364。doi: 10.1214/10-AOS806。URL https://projecteuclid.org/journals/annals-of-statistics/volume-38/issue-5/Adaptive-estimation-for-Hawkes-processes-application-to-genome-analysis/10.1214/10-AOS806.full。

Alexander Sokol。干预 Ornstein-Uhlenbeck随机微分方程，2013年8月。网址https://arxiv.org/abs/1308.2152v1。

Alexander Sokol 和 Niels Richard Hansen。随机微分方程的因果解释。《电子概率期刊》，19（无），2014年1月。ISSN 1083-6489。doi：10.1214/EJP.v19-2891。网址http://arxiv.org/abs/1304.0217。arXiv:1304.0217 [数学]。

Xiangchen Song、Weiran Yao、Yewen Fan、Xinshuai Dong、Guangyi Chen、Juan Carlos Niebles、Eric Xing 和 Kun Zhang。未知非平稳性下的时间解耦表示学习。神经信息处理系统进展，36：8092–8113，2023年12月。网址https://proceedings.neurips.cc/paper_files/paper/2023/hash/19a567abaec3990cb40d7a013556fecd-Abstract-Conference.html。

Xiangchen Song、Zijian Li、Guangyi Chen、Yujia Zheng、Yewen Fan、Xinshuai Dong 和 Kun Zhang。基于非平稳稀疏转移的因果时间表示学习。载于 A. Globerson、L. Mackey、D. Belgrave、A. Fan、U. Paquet、J. Tomczak 和 C. Zhang 编，神经信息处理系统进展，第 37 卷，77098–77131 页。Curran Associates, Inc.，2024年。网址https://proceedings.neurips.cc/paper_files/paper/2024/file/8cef4e4bcb85f7d4a1005a9db018d6b6-Paper-Conference.pdf。Peter Sorrenson、Carsten Rother 和 Ullrich K¨othe。通过广义不可压缩流网络（GIN）实现非线性独立成分分析解耦，2020年1月。网址https://arxiv.org/abs/2001.04872v1。Chandler Squires、Anna Seigal、Salil S Bhate 和 Caroline Uhler。通过干预实现线性因果解耦。载于 Andreas Krause、E mma Brunskill、Kyunghyun Cho、Barbara Engelhardt、Sivan Sabato 和 Jonathan Scarlett 编，第*40*届国际机器学习会议论文集，第 202 卷，载于机器学习研究论文集，32540–32560 页。PMLR，2023年7月23–29日。网址https://proceedings.mlr.press/v202/squires23a.html。

Ali Torkamani 和 Nicholas J. Schork。通过网络重构识别罕见癌症驱动突变。《基因组研究》，19(9):1570–1578，2009年9月。ISSN 1088-9051。doi: 10.1101/gr.092833.109。网址 http://genome.cshlp.org/lookup/doi/10.1101/gr.092833.109。

Kexin Wang 和 Anna Seigal。《超完备独立成分分析的可识别性》，2024年1月。链接 https://arxiv.org/abs/2401.14709v1。

Kexin Wang、Aida Maraj 和 Anna Seigal。《对比独立成分分析》，2025年5月。链接 http://arxiv.org/abs/2407.02357。arXiv:2407。02357 [数学]。

Christian H. Weiß。用于计数时间序列建模的稀疏化操作——综述。《*AStA*统计分析进展》，92(3):319–341，2008年8月。ISSN 1863-8171, 1863-818X。doi: 10.1007/ s10182-008-0072-3。链接 http://link.springer.com/10.1007/s10182-008-0072-3。

诺伯特·维纳。《平稳时间序列的外推、插值与平滑：工程应用》。麻省理工学院出版社，马萨诸塞州剑桥，1949年。

G. Tunnicliffe Wilson。谱因子分解的收敛定理。《多元分析杂志》，8 (2):222–232，1978年。ISSN 0047-259X。doi: https://doi.org/10.1016/0047-259X(78)90073-8。URLhttps://www.sciencedirect.com/science/article/pii/0047259X78900738。

Haixu Wu、Jiehui Xu、Jianmin Wang 和 Mingsheng Long。Autoformer：用于长期序列预测的自相关分解变换器。神经信息处理系统进展，34:22419–22430，2021年12月。URL https://proceedings.neurips.cc/paper/2021/hash/bcc0d400288793e8bdcd7c19a8ac0c2b-Abstract.html。

Dingling Yao、Caroline Muller 和 Francesco Locatello。将因果表征学习与动力系统结合以推进科学。2024年。URLhttp://arxiv.org/abs/2405.13888。

Weiran Yao、Guangyi Chen 和 Kun Zhang。《时间解耦表示学习》。神经信息处理系统进展，35:26492–26503，2022 年 12 月。网址https://proceedings.neurips.cc/paper_files/paper/2022/hash/a938292feb86b94ebe3e6200ff7786ef-Abstract-Conference.html。

Weiran Yao、Yuewen Sun、Alex Ho、Changyin Sun 和 Kun Zhang。从一般时间数据中学习时间因果潜在过程，2022年2月。网址 http://arxiv.org/abs/2110.05428。arXiv:2110.05428 [cs, stat]。Jiaqi Zhang、Kristjan Greenewald、Chandler Squires、Akash Srivastava、Karthikeyan Shanmugam 和 Caroline Uhler。软干预下因果解耦的可识别性保证。载于 A. Oh 、 T. Naumann 、 A. Globerson 、 K. Saenko、M. Hardt 和 S. Levine 编，神经信息处理系统进展，第36卷，第50254–50292页。Curran Associates, Inc.，2023年。网址https://proceedings.neurips.cc/paper_files/paper/2023/file/9d3a4cdf6f70559e8c6fe02170fba568-Paper-Conference.pdf。Kun Zhang、Shaoan Xie、Ignavier Ng 和 Yujia Zheng。从多个分布中进行因果表征学习：一般设定，2024年2月。网址 https://arxiv.org/abs/2402.05052v1。

补充材料

“因果表示与随机建模”附录组织：

A 有用的引理 **17**
A.1 初步引理 .17A.2 随机点过程讨论.17A.2.1 关于滤过的备注 .18A.3 累积量与张量 .19

B 支持性结果的证明 性结果的证明 **19**
B.1 引理 1 的证明 .20B.2 引理 2 的证明 .21

C 主识别理论的证明 识别理论 **2** **2**
C.1 符号说明 .22C.2 定理 1 的证明.22C.2.1 代数量的分解 .22C.2.2 有用引理 .23C.2.3 定理 1 的证明 .26C.2.4 识别因果结构 .30C.3 定理 2 的证明.30C.3.1 定理 2 的预备内容 .30C.3.2 定理 2 的证明 .32C.4 定理 3 的证明.33C.5 变化的相关噪声 .33C.6 存在瞬时影响时的识别 .34

D 扩展标识符 性结果 **3** **5**
D.1 层级最小性的存在性 .35D.2 命题 D.1 的证明 .36D.3 命题 D.2 的证明 .38

E MUTA 详述 TE 构型 **3** **9**
E.1 神经功率谱密度 .39E.2 时间自适应模块的先验分解 .40E.3 卷积先验的显式控制 .42E.4 与学习潜在因果变量的方法比较 .42E.5 模拟方案 .42

F 详细实验结果 **43**

G 扩展讨论 **43**G.1 与三个潜在动态过程的联系 .43G.2 不同层级识别 .44G.3 关于非参数强度的讨论 .45

一些有用的引理

A.1 预备引理

定理 2 中所述的可识别性 定理 2 建立在 引理 1 和 引理 2 之上，这两条引理讨论了空间度量和拓扑空间下广泛的收敛条件。为便于理解，我们通过示例介绍这些基础但关键的概念。

引理 A.1 (弱收敛 Billingsley [1999]). 设 $(S, \mathcal{S})$ 是一个配备其 *Borel*σ-代数的波兰空间，并设 $\{Z_n\}_{n\in\mathbb{N}}$ 和 Z 是定义在同一概率空间上的 S-值随机元素。那么序列 $\{Z_n\}$ 依分布（即弱收敛）收敛到 Z，记作 $Z_n \Rightarrow Z$，当且仅当

$$\lim_{n\to\infty} \mathbb{E}[f(Z_n)] = \mathbb{E}[f(Z)]$$

对所有有界连续函数 $f\colon\ S \to \mathbb{R}$ 均成立。

这个引理的证明和论证是基础概率论中的经典内容，因此这里略去。在大多数情况下，弱收敛对应于过程或变量的有限维分布收敛。

引理 A.2 (测度的紧性 $\mathbb{P}_{Z^{(\Delta)}}$). 设 $\{Z_n^{(\Delta)}\}_{n\in\mathbb{N}}$ 是一列以 S 为值的随机元素（例如随机过程或路径取值），由 Δ 索引并定义在带有 *Borel*σ-代数的波兰空间 S 上。那么相应的概率测度 $\{\mathbb{P}_{Z_n^{(\Delta)}}\}$ 序列是紧的。特别地，任意子序列都存在一个进一步的弱收敛子序列。

概率测度序列的紧性确保存在性质良好的子序列：每个子序列都包含一个进一步的弱收敛子序列。这一性质在波兰空间中特别有用，因为在弱拓扑下，紧性等价于相对紧性（预紧性）。不过，值得注意的是，预紧性并不意味着完全紧致；一般而言，紧序列未必收敛，除非另有唯一性或极限识别论证。因此，紧性为子序列的行为提供了必要的控制，但不能保证整个序列完全收敛。

引理 A.3 (高阶矩界蕴含低阶界)。 设 $\{Z_n\}_{n\in\mathbb{N}}$ 是定义在同一概率空间上的实值随机变量序列。固定一个整数 $d > 0$。假设存在一个常数 $C > 0$，使得

$$\sup_{n\in\mathbb{N}} \mathbb{E}[|Z_n|^d] \le C.$$

那么，对于任意 $0 < p < d$，存在一个常数 $C_p > 0$，使得

$$\sup_{n\in\mathbb{N}} \mathbb{E}[|Z_n|^p] \le C_p.$$

A.2 随机点过程的讨论

首先，我们回顾良好定义随机点过程所需的关键假设。

假设 4 （稳定性和平稳增量， [Bacry 等人的 2015]命题 1）。 过程 N_t具有渐近平稳增量，且强度 λ_t 在核满足以下假设时渐近平稳：

$$\rho_{\Phi(t)} = \|\Phi(t)\| = \int_0^t |\Phi(t)|\ dt \text{ has spectral radium smaller than } 1 \tag{A.1}$$

假设 4 给出了一个必要条件，使得点过程的强度具有稳定的平稳增量。特别地，这意味着整个过程趋于稳定，其条件强度的期望未知但固定 $\mathbb{E}[\lambda_t^i] = \Lambda^i$。在平稳增量假设的约束下，相应过程的存在性由引理 3得到保证。为说明这些条件，我们在例 1中展示一个更简单版本的核函数。

例 1。 考虑一个点过程，其核函数以指数衰减的方式将因果影响传递给其他过程。因此，生成过程可相应地表示为

$$\lambda_t^i = u^i + \sum_{j=1}^{p} \int_0^t \alpha^{ij} e^{-\beta(t-t')} \, dN_{t'}^j$$

这表明指数型核会引发持续但随时间推移而衰减的影响。从技术上说，尽管诱发的因果影响会在系统动力学内部衰减，但只要所有 j 的因果强度 $\alpha = 0$，它们就不会消失。

引理 3 （Kirchner [2016]中的命题 6）。如果假设 4 中的所有条件和结论几乎处处成立，那么就仅存在一个确定的过程，其动力学与关于 Λ^i 的观测数据相符。

卷积核与强度。我们评估连续时间中的随机积分，其中由核诱发的因果影响会随时间平滑衰减。设 $t < t'$ 表示早于 t 的任意时刻。那么，针对 p 随机过程，在 t' 处求得的核向量为:

$$\phi_i(t-t') = (\phi_{i,1}(t-t'), \phi_{i,2}(t-t'), \phi_{i,3}(t-t'), \cdots, \phi_{i,p}(t-t'))$$

这将得到如下形式的积分

$$I_i(t) = \int_0^t \sum_{j=1}^{2} \phi_{ij} \, dN_t^j = \int_0^t \sum_{j=1}^{2} \phi_{ij}(t-t') \, d \begin{bmatrix} N_1(t') \\ N_2(t') \\ \vdots \\ N_p(t') \end{bmatrix} = \int_0^t \phi_i(t-t') \, dN_{t'} \tag{A.2}$$

汇总 p 积分

$$\begin{bmatrix} I_1(t) \\ I_2(t) \\ \vdots \\ I_p(t) \end{bmatrix} = \int_0^t \begin{bmatrix} \phi_{11}(t-t') & \cdots \\ \phi_{21}(t-t') & \cdots \\ \vdots & \ddots \\ \phi_{p1}(t-t') & \phi_{pp}(t-t') \end{bmatrix} d \begin{bmatrix} N_1(t') \\ N_2(t') \\ \vdots \\ N_p(t') \end{bmatrix} = \int_0^t \Phi(t-t') dN_{t'} = \int_0^t \Phi(t') dN_{t-t'}$$

由此得出 $I(t)$ 是核矩阵 Φ与随机跳跃向量 dN 截至时间 t 的卷积

$$I(t) = \Phi_t \star dN_t \tag{A.3}$$

A.2.1 关于滤过的备注

在概率论中，滤过 $\mathcal{F}_t$ 定义为使强度过程 λ_t 成为 $\mathcal{F}_t$-适应且可测的最小 σ代数。该滤过通过对过去事件进行集合运算（例如并、交）并取其最小闭包来构造，从而确保 λ_t 随着可观测历史 [一致演化（Hawkes，1971；Daley 和 Vere-Jones， 2005]）。因此，对于任何以其内部历史作为滤过的情形，我们有 $\mathcal{F}_s \subseteq \mathcal{F}_t$，其中 $s \leq t$。请注意，滤过 $\mathcal{F}_t$ 在理论上可能不同于内在历史 $\mathcal{H}_t$，这给点过程的评估和建模带来了额外挑战。关于 $\mathcal{F}_t$ 和 $\mathcal{H}_t$ 定义不同的情形，读者可参阅 [Daley 和 Vere-Jones， 2005]，以了解全面讨论。我们有时会将随机微积分中的积分元素 dN_t^i用作重载记号，以便将其与确定性对应项区分开来。尽管记号可能相似，但二者本质上不同：标准微积分考察固定网格宽度上的无穷小增量（例如，当 $\Delta t \to 0$ 时， $dg(x)$），而增量 dN_t^i 是由随机过程支配的随机变量。具体而言，每个无穷小区间上的实现值均取自强度为 λ_t^i的伯努利过程，使得 $\mathbb{P}(dN_t^i > 0 \mid \mathcal{F}_t) = \lambda_t^i dt$。与确定性微分不同， dN_t^i涵盖了每个区间内事件发生的不确定性。核矩阵 Φ_t 由随时间衰减的核函数组成，用于传递过去事件对各过程的影响。它同时捕捉

时间延迟依赖和因果依赖，并在自激或互激动态的建模中发挥核心作用。

A.3 累积量与张量

累积量张量记法。随机向量 $X \in \mathbb{R}^p$ 的 d 阶累积量张量记作 $\kappa_d(X) \in \mathbb{R}^{p\times\cdots\times p}$，并且在所有模态上都是对称的。在独立成分分析和因果表示学习中，独立成分的累积量通常具有 CP 形式：

$$\kappa_d(X) = \sum_{r=1}^{R} \lambda_r \cdot v_r^{\otimes d},$$

其中 $v_r \in \mathbb{R}^p$ 和 $\lambda_r \in \mathbb{R}$。这种结构使得能够从累积量信息中识别潜在源。

张量记法与运算。我们将 d 阶张量记作 $\mathcal{T} \in \mathbb{R}^{I_1\times I_2\times\cdots\times I_d}$。外积 $u^{(1)} \otimes \cdots \otimes u^{(d)} \in \mathbb{R}^{I_1\times\cdots\times I_d}$ 生成一个秩一张量，其元素为：

$$\mathcal{T}_{i_1,\ldots,i_d} = u^{(1)}_{i_1} \cdots u^{(d)}_{i_d}.$$

给定张量 $\mathcal{T} \in \mathbb{R}^{I_1\times\cdots\times I_N}$ 和矩阵 $U \in \mathbb{R}^{J\times I_n}$，模态-$n$ 积 $\mathcal{T}\times_n U \in \mathbb{R}^{I_1\times\cdots\times I_{n-1}\times J\times I_{n+1}\times\cdots\times I_N}$ 定义为：

$$(\mathcal{T} \times_n U)_{i_1,\ldots,i_{n-1},j,i_{n+1},\ldots,i_N} = \sum_{i_n=1}^{I_n} \mathcal{T}_{i_1,\ldots,i_N} \cdot U_{j,i_n}.$$

B 支持性结果的证明

点过程与计数过程 N_t 相关联，后者源于其在可测空间 S 上的随机测度。令网格宽度趋于零，可得条件强度过程：

$$\lambda_t = u + \Phi_t \star dN_t. \tag{B.1}$$

即使是单变量点过程，也不满足自回归性质，因为强度 λ_t 本身受内部动态随机驱动。这使得该过程具有自激性，并适应于滤过 $\mathcal{F}_t$。这种非自回归结构给如何为一组随机强度变量 $\{\lambda_t^i\}_{t\in T,\ i\in[p]}$构建因果模型带来了挑战。为了对随机跳跃 dN_t 与其条件强度 λ_t 之间的关系建模，我们旨在寻找一种形式为 $dN_t = f(\lambda_t)$ 的表示，使其与潜在因果模型相容，这正是我们的弱收敛类所追求的目标。

注B.1. dN_t 与 λ_t 之间的关系可以写得更为简洁。根据定义，条件强度满足：

$$\lambda_t = \frac{\mathbb{E}[N_{t+dt} - N_t \mid \mathcal{F}_t]}{dt} = \frac{\mathbb{E}[dN_t \mid \mathcal{F}_t]}{dt}.$$

这自然导出如下分解：

$$dN_t = \lambda_t dt + dM_t,$$

其中，dM_t 是一个局部鞅，用来刻画相对于条件期望的随机偏差。这种分解类似于标准回归形式 $Y = \mathbb{E}[Y \mid X] + \epsilon$，其中滤过 $\mathcal{F}_t$ 起到协变量的作用，而 dM_t 则表示随机误差项。

在因果充分系统的假设下，残差噪声向量 $R_t = (r_t^1, r_t^2, \ldots, r_t^p)$ 的各个分量相互独立。这种点过程的特定表示有助于运用算子理论工具进行进一步分析。条件强度 λ_t 可解释为对过程 i 在时刻 t的预期事件数的短期估计。这个观点在 事实1中进行了形式化，其证明见附录 B。

事实1。 设 N_t 是一个计数过程，其条件强度由式 (1)定义。则离散时间过程 N_t^{Δ}与近似项 $\lambda_t dt + dM_t$ 之间的残差以高概率一致有界。具体而言，以至少 $1-\epsilon$的概率，以下结论成立：

$$\left|N_t^{\Delta} - \lambda_t dt - dM_t\right| = o\left(\log\left(\frac{f(\Lambda)}{T^2}\right)\right).$$

B.1 证明引理1

该引理发挥着核心作用，并引出了我们的可识别性结果。我们重新陈述原始结论，以提供有关点过程以及弱拓扑和弱收敛理论的更多细节与背景。

引理B.1 （强度中的点过程界，构造性）。 对于一个可测映射 $N^{\Delta}:(\Omega\ \mathcal{F}) \to (M_p, \mathcal{M})$，使得 $\omega \mapsto N(\omega)$ 在尺度 Δ下是一个点过程。令 Δ 为其点过程任意子序列的控制算子。考虑由拓扑 $\mathcal{M}_p := \mathcal{B}(M_p)$生成的 $A \in \mathcal{B}$。 λ 和 ϵ 定义在该指标空间上。如果 λ 满足平稳增量条件，则我们可以建立所构造等价类的弱收敛：

$$\sum_{k:k\Delta\in A} \lambda_k^{(\Delta)} + \epsilon_k^{(\Delta)} \overset{w}{\Rightarrow} N(A) \quad for\ \lambda_k^{\Delta} = \lim_{\Delta\to 0}\lim_{\delta\to 0}\frac{\mathbb{E}[dN^{\Delta}|\mathcal{F}]}{\delta}$$

作为一个示例性过程，一个平凡情形可以很容易地得到证明， $\Delta = 1$这与离散时间自回归模型一致。对于更一般的情形 $\Delta \in (0,1)$，为保持条件强度的良好性质，分子 $\mathbb{E}[N(\Delta)\ |\ \mathcal{F}]$ 和分母 Δ 应当同时且成比例地变化。关于构造性引理，请参阅附录A。

证明。 我们先从平凡情形开始证明。如果 $\delta = \Delta_1 = 1$，这些条件总是显然成立。在这种情况下，我们只需通过简单地使用塔式法则来证明 $N^{(\Delta_1)}{}_t = \lambda^{(\Delta_1)}{}_t + R_t$。因此，我们的证明将更多关注 $\delta \neq 1$的非平凡情形。

情形 *2*： $\delta \in (0, \Delta_1)$

如果用于生成子序列的时间步长算子按比例缩小到足够小的单位 $(0, \Delta_1)$，那么这一情形的推理会变得更加复杂。我们重写逼近序列N ，以利用该空间的可度量性。由于我们在波兰空间中工作，Borel δ代数是可数生成的，并且该空间可分且可度量。给定一个可测集合 $A \in \mathcal{B}$ 和一个指标ρ，将开 δ邻域定义为：

$$\mathcal{A} = A^{\delta} := \{x \in \mathbb{R}^d : \rho(x, A) < \delta\}$$

根据波兰空间上 Borel 概率测度的外正则性，对每个 $\epsilon > 0$，都存在一个开集的可数集族 $\{A_i\}_{i\in\mathbb{N}}$，使得 $\bigcup_i A_i \supset \mathcal{A}$ 且 $\sum_i \mu(A_i \setminus A) < \epsilon$。这使我们能够用开集从外部逼近任意紧子集，并使其多出的质量任意小，同时确保逼近序列定义在一个非递减的基上。我们将收敛改述为

$$\sum_{k:k\Delta\in A} \lim_{i\to\infty}\frac{\mathbb{E}[N^{\Delta}(A_i)|\mathcal{F}]}{|A_i|} + \epsilon_k^{(\Delta)} \overset{w}{\Rightarrow} N(A) \quad \text{for } \Delta \to 0 \tag{B.2}$$

上述方程改编自点过程的连续时间强度。然而，这要求我们针对 A_i 考察两个极限条件： $1/k$ 闭球的测度趋于零，以及子序列算子 Δ 趋于 0。一种常见的方法是确保极限的支配收敛和一致收敛。为了利用强度信息，使我们的收敛推广到更一般的过程，我们首先只使用算子 Δ来诱导强度函数的相同时间尺度。因此，我们得到等价条件

$$\sum_{k:k\Delta\in A} \frac{\mathbb{E}[\sum_{k=1} Z_k^{\Delta} - \sum_{k=1} Z_{k-1}^{\Delta}|\mathcal{F}]}{|A_i|} + \epsilon_k^{(\Delta)} = \sum_{k:k\Delta\in A}\frac{\mathbb{E}[Z^{\Delta}(\Delta)|\mathcal{F}]}{\Delta} + \epsilon_k^{(\Delta)} \overset{w}{\Rightarrow} N(A) \quad \text{for } \Delta \to 0 \tag{B.3}$$

我们去掉极限条件，因为显然，当 $\Delta = 0$时， $|A_i|$ 是零测集；这确保了我们的拓扑性质与合理性保持对齐，从而在下文中只分析子序列。根据 [Kirchner 的引理 2， 2016], 对于任意紧区间 $[a,b]^{(\delta)}$ 及分箱数 $[b-a]/\delta$， $\mathbb{E}[N^{(\delta)}([a,b])] < (b-a+2)(I - G^{(\delta)}(a,b))^{-1}\Lambda$ 其中 $G(a,b) = \int_a^b \Phi(s)ds$ 是随机微分方程系统的一个解

$$\mathbb{E}[\lambda([a,b])] = \mathbb{E}[u + G(a,b)\Lambda], \text{ for } \mathbb{E}[\lambda(a,b)] = \Lambda$$

注意，重新应用塔式法则后，式（B.2）意味着：

$$\lim_{\delta\to 0} \mathbb{E}[N_A^{(\delta\in(0,1))}] \to \lim_{\delta\to 0} \mathbb{E}[\frac{\mathbb{E}[N_A^{(\delta)}|\mathcal{F}_t]}{\delta}]$$

接下来，我们证明相应概率测度 $\mathbb{P}^\Delta$ 的紧性对于式（B.3）左侧实现所需的收敛是必要的。不失一般性地，我们考虑一个非参数强度函数

$$\lambda_t = \psi\ u + \int \phi\ t - s\ Z^\Delta\ s\ \ ds \qquad \mathbb{E}[\lambda_t] = \Lambda \qquad \mathbb{E}[\lambda_t] = \mathbb{E}[\psi\ u + \int \phi\ t - (( ) ( ))。$$

$s\ Z^\Delta\ s\ \ ds$ α 因此，和 $(( ) ( ))]$。假设 ψ 是 -Lipschitz 的，并且 $\alpha\|\phi\|_1 < 1$ [Bremaud and Massoulie， 1996], 因此映射$F(\Lambda) = \psi(u + \|\phi\|_1\Lambda)$ 是 $\mathbb{R}_+$上的压缩映射。根据巴拿赫不动点定理，存在唯一解 $\Lambda^\star$ 满足方程：

$$\Lambda^\star = \psi(u + \|\phi\|_1\Lambda^\star)$$

形式上，可以将其重新排列为：

$$\psi^{-1}(\Lambda^\star) - \|\phi\|_1\Lambda^\star = u \quad \Longrightarrow \quad \Lambda^\star = \left(\mathrm{id} - \|\phi\|_1 \cdot \psi^{-1}\right)^{-1}(-u)$$

前提是 $\mathrm{id} - \|\phi\|_1 \cdot \psi^{-1}$ 在 ψ的像上可逆。

为了控制尾概率，我们应用马尔可夫不等式：

$$\mathbb{P}\left(\sum_{k:k\Delta\in A} \frac{\mathbb{E}[Z^\Delta(\Delta)|\mathcal{F}]}{\Delta} + \epsilon_k^{(\Delta)} > M_\varepsilon\right) \le \frac{\mathbb{E}[\sum_k \Lambda^\Delta]}{M_\varepsilon} \le \frac{(b-a+2\delta)\cdot\Lambda^\star}{M_\varepsilon}$$

这里，我们定义：

$$M_\varepsilon := \frac{(b-a+2\delta)\cdot\Lambda^\star}{\varepsilon} \quad \text{where } \Lambda^\star = \psi(u + \|\phi\|_1\Lambda^\star)$$

这种选择确保上界始终处于规定的 ε 级别内，对所有 $\Delta \in (0,\Delta_1)$ 均成立。由于我们所需的仅是相对紧性，因此不会建立更紧的上界。引理 A.2中给出的测度紧性表明，我们总能找到一个子序列 $\lambda_{k_n}^\Delta + \epsilon_{k_n}^\Delta$ ，使其在 $\lambda_k^\Delta + \epsilon_k^\Delta$中弱收敛到一个序列 $\lambda^\star + \epsilon^\star$。然而，这种子序列的弱收敛无法保证每个序列的极限唯一性。因此，我们还应通过一致收敛进一步控制每个序列的极限行为；这里的一致收敛针对由逼近过程和目标过程定义的特征泛函，这正是引理 2的核心思想。

□

B.2 引理 2 的证明

引理 B.2 （收敛到等价类，并针对有限维分布给出构造性证明）。 根据引理 *1* 及其构造性版本，每个定义在测度上的子序列 $N_k^\Delta(A_i)$ $\mathbb{P}^\Delta$ 弱收敛到极限过程 N_t ，且该极限存在且唯一。

证明。这个证明虽然繁琐，但较为直接，改编自对任意紧支撑连续函数 f的有限维矩母函数（MGF）的一致收敛。证明过程安排如下：引理 1 中的子序列收敛选取一个任意序列，其特征函数 $\Psi(N_k^\Delta$也收敛到 $\Psi(N^{\Delta\to 0}$。根据引理 A.3，若过程的二次变差有界，则该子序列与原过程 N具有相同的极限。特征函数之间的等价性表明，它们的行为是一致收敛的。

□

C 主可识别性理论的证明

C.1 符号说明

因果表示的射影空间 在本证明的其余部分，我们研究所提出的潜在因果模型的代数结构。我们在复射影空间 $\mathbb{P}^n$（也写作 $\mathbb{CP}^n$）中进行研究；该空间形式化了通常的可识别性约定，即矩阵在相差一个非零标量倍数时视为等价。具体而言，对于向量 $v \in \mathbb{C}^{n+1} \setminus \{0\}$，其在 $\mathbb{P}^n$ 中的等价类为 $[v] = \{\lambda v \mid \lambda \in \mathbb{C} \setminus \{0\}\}$。例如，给定一个代数对象 $V := V(f(x, y)) \subseteq \mathbb{P}^1$，其中 V是齐次多项式 $f(x, y)$的零点集，则每个点 $[x : y] \in \mathbb{P}^1$ 对应于 $\mathbb{C}^2$中一条经过原点的直线。在通常的识别 $\mathbb{P}^1(\mathbb{C}) \simeq \hat{\mathbb{C}}$ （黎曼球面）下，每条直线与单位球面 $S^2 \subset \mathbb{R}^3$ 相交于两个对径点。因此，从拓扑角度看，我们有 $\mathbb{P}^1 \cong S^2$。除非另有说明，本文所考虑的所有环均假定为交换的、诺特的（有限生成的），并具有乘法单位元。特别地，我们关注诸如 $K(x, y, z)$ 及其子环，例如 $f(x, y) \subseteq K(x, y, z)$ 这样的环，并始终认为它们满足这些性质。若需要额外假设，例如为整环或域，将明确说明。

一般性与满秩。 在本文中，我们将矩阵

$$F \in \mathbb{C}^{n \times p} \quad \text{as points} \quad P = (p_0, p_1, \ldots, p_{np-1}) \in \mathbb{P}^{np-1} \text{ or } \mathbb{CP}^{np-1},$$

其中， F 的各项与 P的齐次坐标相对应。若不存在非零多项式 $f \in K[x_0, \ldots, x_{mn-1}]$ 使得 $f(p_0, \ldots, p_{mn-1}) = 0$，则称点 P 为一般的。等价地，一般点的坐标在基域 K上代数独立。一般性意味着相应矩阵 F 几乎必然满秩，因为任一子式为零都对应于一个非零多项式的零点集，而一般点不可能落在该零点集上。然而，反过来通常不成立：矩阵可以满秩，但其各项并不代数独立。因此，一般点涵盖的矩阵范围比仅限于单射或可逆矩阵更广。

低维嵌入 为了嵌入由我们的模型产生的代数对象，我们引入一个更高维的射影空间 $\mathbb{P}^N$ ，其中包含 $N \geq n$，称为环境空间； $\mathbb{P}^n$ 自然嵌入其中 [参见 Hartshorne，1977年，第 I 章]。我们称射影簇 $V \subset \mathbb{P}^n \subset \mathbb{P}^N$ 在 $\mathbb{P}^N$中的余维为 $N - n$，即 $\dim \mathbb{P}^N - \dim V = N - n$，其中 $\dim V$表示 V的射影维数。按照这些约定，与潜在因果模型相关的所有代数对象均按射影方式理解，因此缩放等价性已内置于该框架中。

C.2 定理 1的证明

我们将定理 1 的证明加以分解，逐步考察与整个生成模型的完全识别日益相关的代数几何观点。首先，我们通过多线性变换得到无限阶 INAR 模型的代数累积量，从而保证可以恢复 $K = F(\mathbb{I}_p - \Phi)^{-1}$，其中因子 $[\![K^{(1)}, K^{(2)}, \ldots, K^{(p)}]\!]$ 位于一个 Veronese 簇上，前提是引理 C.1 成立。由此可知，拓扑排序表示 $\mathcal{G}$ 可通过将任意 $\Phi(\tau)$ 嵌入到更大的环境空间 $\mathbb{R}^{2p \times 2p}$而得到。最后，我们证明，在我们的假设下，位于相关环境空间上的相关代数簇的维数是零维的。

C.2.1 代数量的分解

为了有可能识别动力学中的任何潜在成分，我们必须引入超出当前设定的张量代数，如下列重要结果所示。

推论 C.1 (CP 分解). 设 $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ 是一个 N 阶张量。若存在分量向量 $a_r^{(n)} \in \mathbb{R}^{I_n}$ ，分别对应每个，则称 $\mathcal{T}$存在一个精确的秩-R 典范多元分解（CP）分解

$r = 1$、.....、 R、 $n = 1$、.....、 N，使得：

$$\mathcal{T} = \sum_{r=1}^{R} a_r^{(1)} \otimes a_r^{(2)} \otimes \cdots \otimes a_r^{(N)} = [\![A^{(1)}, A^{(2)}, \ldots, A^{(N)}]\!],$$

其中， $A^{(n)} = [a_1^{(n)}\ a_2^{(n)}\ \cdots\ a_R^{(n)}] \in \mathbb{R}^{I_n \times R}$ 为因子矩阵。

推论 C.2。 设 $X^{(1)}$、$X^{(2)}$、...、$X^{(n)} \in \mathbb{R}^p$ 为具有非零 d阶累积量的独立随机向量，且每个向量均可表示为

$$\kappa_d(X^{(i)}) = \lambda_i \cdot v_i^{\otimes d}, \quad for\ i = 1, \ldots, n,$$

其中 $v_i \in \mathbb{R}^p$ 和 $\lambda_i \in \mathbb{R} \setminus \{0\}$。设 $\mathcal{T}: = \kappa_d\ (X^{(1)} + \cdots + X^{(n)}) \quad \in \mathbb{R}^{p \times \cdots \times p}$ 为其和的 d阶累积量张量。

假设矩阵 $V = [v_1\ v_2\ \cdots\ v_n] \in \mathbb{R}^{p \times n}$ 满足

$$\operatorname{krank}(V) \geq \left\lceil \frac{2n + (d-1)}{d} \right\rceil.$$

则 *CP* 分解

$$\mathcal{T} = \sum_{i=1}^{n} \lambda_i \cdot v_i^{\otimes d}$$

在缩放和置换意义下是唯一的。

C.2.2 有用引理

回忆一下， F 足够一般， Φ 是谱半径为 $\rho < 1$的核矩阵。令 K $=F(\mathbb{I}_p - \Phi)^{-1}$，并用 $\mathcal{F}[\cdot]$ 表示正则傅里叶变换。我们用 $[\![K_1$、 K_2、.....、 $K_p]\!]_{\mathcal{F}}$表示傅里叶变换后 K的各列。对于每个$(K_j)_{\mathcal{F}}$，其坐标为 $[k_1$： k_2： $\cdots$： $k_n]_{\mathcal{F}}$，它们表示向量中的自由不定元。

接下来，我们将扩展 [Wang 和 Seigal的命题 4.6 2024]，以证明一个重要的预可识别性结果，即混合参数空间 $F(\mathbb{I}_p - \Phi)^{-1}$。

引理 C. 1 ψ $\mathbb{P}_{\mathcal{F}}^{n-1} \mapsto \nu_2\ \mathbb{P}_{\mathcal{F}}^{n-1}$ $\subset$（与一般线性子空间的有限交）。定义 $\mathbb{P}_{\mathcal{F}}^{m-1}$ 有理映射: ( ) 如下：

$$\psi : [k_1 : k_2 : \cdots : k_n]_{\mathcal{F}} \mapsto [k_1^d : k_1^{d-1} k_2 : \cdots : k_n^d]_{\mathcal{F}} \tag{C.1}$$

其中， $m = \binom{n+d-1}{d}$ 表示维数为 $m - 1$的射影空间中的坐标。 ν_d 表示所有d 阶张量的d次 *Veronese* 嵌入。设 $\mathcal{V}$ 为由所有秩为 *1* 的矩阵构成的投影空间。考虑由 $\{K_1^{\otimes d}, K_2^{\otimes d}, \cdots, K_p^{\otimes d}\}$张成的一个充分一般的线性子空间，记为 $\mathcal{W}$。由此可知，簇 $\mathcal{W} \bigcap \mathcal{V}$的维数为零，且 $\mathcal{W}$与 $\mathcal{V}$相交于 d^{n-1}个不同点。因此，当且仅当 $\mathcal{V}(f) = \{K_1^{\otimes d}, K_2^{\otimes d}, \cdots, K_p^{\otimes d}\}$时，$\{K_1^{\otimes d}, K_2^{\otimes d}, \cdots, K_p^{\otimes d}\}$ 是可识别的。

证明。 在经典线性源分解（LSD）设定中， X 的 d 阶累积量具有如下张量分解： $\kappa_d(X) = \sum_{i=1}^{p} \kappa_d(s_i) \cdot (A_i)^{\otimes d}$ 假设 ϵ 的分量均为非高斯，且其 d 阶累积量非零，并且可以进行多次干预。对于固定的 $t = t_i$，每个 Z_t 的充分 d 阶累积量为

$$\kappa_d(Z_t) = \kappa_d[(I - \Phi)^{-1} \star \epsilon_t] = \kappa_d[\sum_{k=1}^{t} H_{t-s} \epsilon_s]$$

对于每个 s， 线性变换 H_{t-s} 会使其累积量产生多线性变换 s

$$\kappa_d(H_{t-s}\epsilon_k) = (H_{t-s})^{\otimes d}\mathcal{C}^d_{\epsilon_s} = \sum_{i=1}^{p} \kappa_d(\epsilon_s^i)(H_{t-s})_j^{\otimes d}$$

完整的 d 阶累积量为

$$\begin{aligned}
\kappa_d(O_t) &= \kappa_d(\underbrace{FZ_t, FZ_t, \ldots, FZ_t}_{d \text{ times}}) \\
&= F^{\otimes d} \cdot \kappa_d(\underbrace{Z_t, Z_t, \ldots, Z_t}_{d \text{ times}}) \qquad \text{(C.2)} \\
&= \underbrace{F \otimes F \otimes \cdots \otimes F}_{d \text{ times}} \cdot \kappa_d \left(\underbrace{\sum_{s_1=1}^{t} H_{t-s_1}\epsilon_{s_1}, \sum_{s_2=1}^{t} H_{t-s_2}\epsilon_{s_2}, \ldots, \sum_{s_d=1}^{t} H_{t-s_d}\epsilon_{s_d}}_{d \text{ times}} \right) \\
&= \underbrace{F \otimes F \otimes \cdots \otimes F}_{d \text{ times}} \cdot \sum_{s_1=1}^{t} \cdots \sum_{s_d=1}^{t} \kappa_d \left(H_{t-s_1}\epsilon_{s_1}, H_{t-s_2}\epsilon_{s_2}, \ldots, H_{t-s_d}\epsilon_{s_d} \right) \\
&= \underbrace{F \otimes F \otimes \cdots \otimes F}_{d \text{ times}} \cdot \sum_{s_1=1}^{t} \cdots \sum_{s=1}^{t} \sum_{j=1}^{p} \kappa_d \left(H_{t-s}^{(:,j)}\epsilon_s^{(j)}, H_{t-s}^{(:,j)}\epsilon_s^{(j)}, \ldots, H_{t-s}^{(:,j)}\epsilon_s^{(j)} \right) \\
&= \underbrace{F \otimes F \otimes \cdots \otimes F}_{d \text{ times}} \cdot \left(\sum_{s=1}^{t} \sum_{j=1}^{p} \kappa_d^{(j)}(\epsilon) \cdot \underbrace{H_{t-s}^{(:,j)} \otimes H_{t-s}^{(:,j)} \otimes \cdots \otimes H_{t-s}^{(:,j)}}_{d \text{ times}} \right) \\
&= \left(\sum_{s=1}^{t} \sum_{j=1}^{p} \kappa_d^{(j)}(\epsilon) \cdot \underbrace{F \otimes F \otimes \cdots \otimes F}_{d \text{ times}} \cdot \left(H_{t-s}^{(:,j)} \right)^{\otimes d} \right) \\
&= \sum_{s=1}^{t} \sum_{j=1}^{p} \kappa_d^{(j)}(\epsilon) \cdot \left(F H_{t-s}^{(:,j)} \right)^{\otimes d} \qquad \text{(C.3)}
\end{aligned}$$

由于噪声在时间上不相关（白噪声），我们省略 $\kappa(\epsilon)$ 中的时间索引，即

$$\mathrm{Cov}(\epsilon_s, \epsilon_t) = \sigma^2 \delta_{s,t},$$

与无时间过程不同，时间过程的联合累积量的阶数为 d，并且与时间滞后的数量相关:

$$\begin{aligned}
\kappa_d(O_{t_1}, \ldots, O_{t_d}) &= \sum_{s=1}^{\min(t_1,\ldots,t_d)-1} \sum_{j=1}^{p} \kappa_d^{(j)}(\epsilon) \cdot \bigotimes_{\ell=1}^{d} \left(F H_{t_\ell - s}^{(:,j)} \right) \\
&= \sum_{j=1}^{p} \sum_{s=1}^{\min(t_1,\ldots,t_d)-1} \kappa_d^{(j)}(\epsilon) \cdot \bigotimes_{\ell=1}^{d} \left(F H_{t_\ell - s}^{(:,j)} \right) \qquad \text{(C.4)}
\end{aligned}$$

我们将 $x(t)$关于时间 t的傅里叶变换记为 $\mathcal{F}[x](\omega)$。利用卷积定理和傅里叶变换的线性性质，我们得到:

$$\mathcal{F}[\kappa_d(O_t)](\omega) = \mathcal{F}\left[\sum_{s\geq 0}^{t} \sum_{j=1}^{p} \kappa_d^{(j)}(\epsilon) \cdot \left(F H_{t-s}^{(:,j)} \right)^{\otimes d} \right]$$

$$
\begin{aligned}
&= \mathcal{F}\left[\int_0^t \sum_{j=1}^{p} \kappa_d^{(j)}(\epsilon) \cdot \left(FH(t-s)^{(:,j)}\right)^{\otimes d} ds\right] \\
&= \left(\sum_{j=1}^{p} \kappa_d^{(j)}(\epsilon) \cdot \mathcal{F}\left[\left(FH^{(:,j)}\right)^{\otimes d} \star \mathcal{U}(t-s)\right]\right) \\
&= \left(\sum_{j=1}^{p} \kappa_d^{(j)}(\epsilon) \cdot \mathcal{F}\left[\left(FH^{(:,j)}\right)^{\otimes d}\right](\omega) \cdot \left(\pi\delta(\omega) + \frac{1}{i\omega}\right)\right)
\end{aligned} \tag{C.5}
$$

此后，我们应用 $\mathcal{U}$ 来构造一个广义卷积积分。由于 $\delta(\omega)$ 除了在 $\omega = 0$处以外处处为零，将其与 ω 相乘结果为零，因此式（C.5）得到

$$
\begin{aligned}
i\omega\mathcal{F}[\kappa_d(O_t)](\omega) &= \left(i\omega \sum_{j=1}^{p} \kappa_d^{(j)}(\epsilon) \cdot \mathcal{F}\left[\left(FH^{(:,j)}\right)^{\otimes d}\right](\omega) \cdot \left(\pi\delta(\omega) + \frac{1}{i\omega}\right)\right) \\
&= \left(\sum_{j=1}^{p} \kappa_d^{(j)}(\epsilon) \cdot \mathcal{F}\left[\left(FH^{(:,j)}\right)^{\otimes d}\right](\omega) \cdot i\omega \cdot \left(\pi\delta(\omega) + \frac{1}{i\omega}\right)\right).
\end{aligned}
$$

即

$$
\left(\sum_{j=1}^{p} \kappa_d^{(j)}(\epsilon) \cdot \mathcal{F}\left[\left(FH^{(:,j)}\right)^{\otimes d}\right](\omega) \cdot i\omega \cdot \left(\pi\delta(\omega) + \frac{1}{i\omega}\right)\right) = \left(\sum_{j=1}^{p} \kappa_d^{(j)}(\epsilon) \cdot \mathcal{F}\left[\left(FH^{(:,j)}\right)^{\otimes d}\right](\omega) \cdot 1\right) \tag{C.6}
$$

按约定将 FH 用 K 表示，可得:

$$
\kappa_{O_t,\mathcal{F}} = \left(\sum_{j=1}^{p} \kappa_d^{(j)}(\epsilon) \cdot (k_j)_{\mathcal{F}}^{\otimes d}\right)
$$

$(k_j)_{\mathcal{F}}^{\otimes d}$是一个阶为 d、秩为1的张量，因而可以表示为 $V \otimes \cdots \otimes V$的空间。在射影空间 $\mathbb{P}^n$中，每个点表示一条过原点的直线，并且所有点在相差倍数的意义下失去一个维数。因此，簇 $\mathcal{V}$ 的射影维数为 $n-1$，嵌入于射影维数为 $m-1$的环境空间中。线性子空间 $\mathcal{W}$ 是格拉斯曼流形$\mathrm{Gr}(p-1, n-1)$中的一个元素，并且是充分一般的。由 [Chiantini and Ciliberto的命题 2.6, 2002], 对于 $\mathbb{P}^{m-1}$中一个充分一般的、既约且不可约的簇 $\mathcal{V}$，若存在一个整数使得 $k + (n-1) < m-1$，则 $\mathcal{V}$与一般线性子空间 $\mathcal{W}$的交是一般点 $\langle P_0$、P_1、$\cdots$、$P_k\rangle$的并。我们可以通过选取 $k = p$来证明这一点。由于 $m = \binom{n+d-1}{d}$的组合性质，若我们有更高阶的 $d \geq 2$，该条件会更容易得到满足。

因此，假设 ϵ_t中具有非高斯性，对于每个 j，式（C.6）得出 $i\omega\mathcal{F}[\kappa_d(O_t)](\omega)$ 可唯一分解为秩一张量之和（若 $d = 2$，则为矩阵之和）。因此，若满足 $\mathcal{F}\left[\left(FH^{(:,j)}\right)\right]$ 的所有假设，则次线性混合传递矩阵π 的每一列在理论上都可以恢复至缩放和置换的不确定性 Φ。这表明，无论是否需要显式计算该分解，这种唯一性都为进一步解耦奠定了基础。 □

此后，$\mathcal{F}\left[\left(FH^{(:,j)}\right)\right] DP$已可用；因此，唯一的不确定性可直接由引理C.2得出。

引理 C.2 (通过线性群的置换和相位不确定性)。 设 $\mathbb{F}$ 为一个代数闭域。^ ^考虑与时域因子 H_τ 相关联的核混合矩阵 FH_τ，并设 FH_τ 表示通过频域功率谱密度堆叠和 *Wilson* 分解得到的估计因子。那么，存在一个对角相位矩阵 D_τ 和一个置换矩阵 P_τ，使得

$$
FH_\tau = \hat{F}\hat{H}_\tau D_\tau P_\tau,
$$

其中 $(D_\tau, P_\tau) \in \mathrm{GL}_1(\mathbb{F})^p \times S_p$ 对应于从频域分解中继承的不确定性。

˜证明。 考虑由 Wilson 分解得到的频域因式分解 H_ω ， 它满足

$$\tilde{H}_\omega = H_\omega D_\omega P.$$

˜根据构造， H_ω 只能确定到列方向上的可逆缩放 $D_\omega \in \mathbf{GL}_1(\mathbb{F})^p$ 和置换 $P_\omega \in S_p$。这种不确定性恰好对应于可逆线性变换群 GL(V)，它作用于核混合矩阵的每一列，其中 V 是定义在 $\mathbb{F}$ 上的 p 维向量空间。应用傅里叶逆变换（频域 $\to$ 时域）会保留这种线性不确定性：频域中的相位旋转 D_ω 映射为时域中的对角相位矩阵 D_τ ，而列置换保持不变。因此，时域核矩阵满足

$$F H_\tau = \hat{F}\hat{H}_\tau \, D_\tau P_\tau,$$

这表明，频域因式分解中的全部置换和相位不确定性都会在时域中保留下来。

□

利用 引理 C.2，具有一般射影线性张成的泛型点 $(k_j)_{\mathcal{F}}^{\otimes d}(w)$ 表明泛型点 $(k_j)^{\otimes d}(\tau)$。因此，对于任意 FH_τ，原始核混合矩阵均可在相同的置换和缩放意义下恢复。接下来要证明的是因果结构及其完整参数空间的可恢复性。我们的证明聚焦于多重线性约束生成的多项式系统及其关联理想 $\mathcal{I}$。

C.2.3 定理1的证明

剩下需要证明的是，在定理 1 中，条件（2）确保条件（3），从而保证模型的完全可识别性 $\Theta :=$（F， Φ， U）。为研究潜在模型参数空间的几何结构，我们为 Φ赋予一个拓扑序。

引理 C.3。 给定所提出的 *INAR(∞)* 结构的一个二部图，它存在一个核有向无环图，记为 $\mathcal{G}$，对应于一个矩阵 $\mathcal{M}_\mathcal{G} \in \mathbb{R}^{2p\times 2p}$ ，使得 $\mathbb{I}_{2p} - \mathcal{M}_\mathcal{G}$ 可逆。因此，其逆矩阵可以表示为关于 $\mathcal{M}_\mathcal{G}$ 的有限阶指数展开，

$$(\mathbb{I}_{2p} - \mathcal{M}_\mathcal{G})^{-1} = \sum_{i=0}^{k} \mathcal{M}_\mathcal{G}^i, \quad k \le p, \quad \mathcal{M}_\mathcal{G} = \begin{bmatrix} \mathcal{M}_\mathcal{G}[U] & \Phi \\ 0 & \mathcal{M}_\mathcal{G}[V] \end{bmatrix}$$

其中 k 表示该有向无环图中最长路径的长度，且 $\mathcal{M}_\mathcal{G}[U] = \mathcal{M}_\mathcal{G}[V] = \mathbf{0}_{p\times p}$.

证明。 令 $\mathcal{M}_\mathcal{G} \in \mathbb{R}^{2p\times 2p}$ 为与该二部图相关联的核矩阵，其中变量被划分为两个子集 U 和 V。考虑一种拓扑排序，使 U 中的所有节点都排在 V 中的节点之前。由于二部图结构禁止从 V 到 U 的边， $\mathcal{M}$ 的左下角分块中没有元素可以非零。此外， U 内的边仅按拓扑序排列，因此与 U 对应的非对角块和右上角分块的对角线元素均为零。 V 内的边构成一个有向无环图，在对 V 进行拓扑排序后， $\mathcal{M}$ 中对应的分块为严格上三角。因此，整体而言， $\mathcal{M}$ 是严格上三角矩阵，这意味着它表示一个有向无环图。

由于 $\mathcal{M}_\mathcal{G}$ 是严格上三角矩阵，因此它是幂零矩阵。令 k 表示该有向无环图中最长有向路径的长度。那么 $\mathcal{M}_\mathcal{G}^{k+1} = 0$，且 $\mathbb{I} - \mathcal{M}_{\mathcal{G}_K}$ 的逆矩阵可表示为 $\mathcal{M}$ 的有限幂和：

$$(\mathbb{I} - \mathcal{M}_\mathcal{G})^{-1} = \sum_{i=0}^{k} \mathcal{M}_\mathcal{G}^i.$$

每一项 $\mathcal{M}_\mathcal{G}^i$ 都对应于有向无环图中长度为 i 的路径所产生的贡献。这表明，逆矩阵完全由最长路径长度 k以内的路径乘积确定，从而完成证明。 □



图 C.1：图 (a) 是与 $\mathcal{M}_{\mathcal{G}}$兼容的核延迟有向无环图 $\mathcal{G}$；系统 $\mathcal{M}_{\mathcal{G}}$ 是严格上三角矩阵。图 (b) 通过额外的边 $\{u_2 \to u_1, v_3 \to v_2\}$违反了拓扑序。

例 1. 考虑一个允许存在内部箭头的核矩阵 $\Phi \in \mathbb{R}^{2\times 2}$：

$$\Psi_U = \begin{bmatrix} 0 & \psi_{12} \\ 0 & 0 \end{bmatrix}, \quad \Phi = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}, \Psi_V = \begin{bmatrix} 0 & \psi'_{34} \\ 0 & 0 \end{bmatrix}$$

其中，Ψ_U、Ψ_V编码了子图 $\mathcal{G}_U$和 $\mathcal{G}_V$的内部箭头；Φ 编码了从 s到 t的时间延迟核效应。

相应的扩展核矩阵 $\mathcal{M} \in \mathbb{R}^{4\times 4}$ 为

$$\mathcal{M} = \begin{bmatrix} 0 & \psi_{12} & \phi_{11} & \phi_{12} \\ 0 & 0 & \phi_{21} & \phi_{22} \\ 0 & 0 & 0 & \psi'_{34} \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

条件 (3) 中的多项式系统具有几何等价表示。也就是说，条件 (3) 立即表明了推论 C.3。

推论 C.3。 理想 $I:\ \langle K_{\mathcal{G}}(\mathbb{I}_{2p} - M_{\mathcal{G}}) - F_{\mathcal{G}}\rangle$ 满足空间 $\Theta_{\mathcal{G}}:=(F_{\mathcal{G}}、\ M_{\mathcal{G}})$ 是零维的。

根据推论 C.3，识别 F 和 Φ 中所有参数的空间等价于求解以下理想 $\mathcal{I}:\ \langle F_{\mathcal{G}} - F_{\mathcal{G}}(\mathbb{I}_{2p} - \mathcal{M}_{\mathcal{G}})^{-1}(\mathbb{I}_{2p} - \mathcal{M}_{\mathcal{G}})\rangle$。在潜在因果模型中，$F_{\mathcal{G}}$、$\mathcal{M}_{\mathcal{G}}$被填充为需要恢复的不定元，其中 $F_{\mathcal{G}}$ 是通过将 $F \in \mathbb{R}^{n\times p}$ 填入大小为 $2n \times 2p$的更大分块对角矩阵中而得到的扩展线性混合，记为 $F_{\mathcal{G}}$。

请记住，由于分解在缩放和置换意义下唯一，因此我们有 $F(\mathbb{I} - \Phi(\tau))^{-1}$。我们记作 $H_{\mathcal{G}} = (\mathbb{I}_{2p} - \mathcal{M}_{\mathcal{G}})^{-1}$。在 INAR 下，$H_{\mathcal{G}}$ 的对角块是大小为 $p \times p$ 的单位矩阵。然后

$$\begin{aligned} F_{\mathcal{G}}(\mathbb{I}_{2p} - \mathcal{M}_{\mathcal{G}_K})^{-1} &= \begin{bmatrix} A & A' \\ F & F' \end{bmatrix} \begin{bmatrix} \mathbb{I}_{2p} & H \\ \mathbf{0} & \mathbb{I}_{2p} \end{bmatrix} \\ &= \begin{bmatrix} A + A'\mathbf{0} & AH + A' \\ A + F'\mathbf{0} & AH + F' \end{bmatrix} \\ &= K_{\mathcal{G}}. \end{aligned}$$

因此，我们得到理想 $\mathcal{I}:\ \langle F_{\mathcal{G}} - K_{\mathcal{G}}(\mathbb{I}_{2p} - \mathcal{M}_{\mathcal{G}})\rangle$，其中 $K_{\mathcal{G}}$是已知的，因为 FH 已知，根据 引理 C.1，因此它们不再被视为不定量，也就不对$\mathcal{I}$的维数作出贡献。显然，在被动观测设置下，无法恢复完整模型，因为当前的$\mathcal{I}$ 必为正维，导致伴随簇 $\mathcal{V}$中没有定义不动点。因此，核心目标是确定能够表明存在充分变异性或干预设置的情境数量，从而恢复参数空间。为此，我们首先需要讨论 $F_{\mathcal{G}}$的重要性质。

引理 C.4。 $F \in \mathbb{R}^{n\times p}$ 是一个一般满秩矩阵。那么 $F_{\mathcal{G}}$ 是满秩的，且 秩$(F_{\mathcal{G}}) = 2\cdot$秩$(F) = 2\min(n,p)$，并且它在 $\mathbb{R}^{2n\times 2p}$ 的开稠密子集中并非一般的，这是由于分块对角结构施加了额外的线性约束。因此， $F_{\mathcal{G}}$ 属于 $\mathbb{R}^{2n\times 2p}$ 的一个真线性子簇，该子簇由以下条件定义：

$$\mathcal{V}_{I_\star} := \left\{ B \in \mathbb{R}^{2n\times 2p} : B = \begin{pmatrix} F' & 0 \\ 0 & F' \end{pmatrix},\ F' \in \mathbb{R}^{n\times p} \right\}.$$

因此，对于任意方阵 A^{2p}，若 秩$(A) \le r$，则 秩$(F_{\mathcal{G}} A) \le r$

证明。 利用线性代数即可轻易证明。 □

我们首先证明，识别 $\mathcal{M}$ 等价于构造一个簇，该簇落在一个零维射影空间$\mathbb{P}^d$ 中。然而，对于当前的理想，其维数可能要大得多，如下述引理所示。

引理 C.5。 与理想 $\mathcal{I}^\star$ 相关联的子簇在 $\mathcal{M}_{\mathcal{G}}$ 下的维数至少为 $3p^2 - 2p$。

证明。 考虑环境射影空间中的矩阵 $\mathcal{M}_{\mathcal{G}}$。由于左下块 $\mathcal{M}_{\mathcal{G}}$ 被强制为零，因此有 $\mathcal{M}_{\mathcal{G}} \in \mathbb{P}^{4p^2-1-p^2}$。 $\mathcal{M}_{\mathcal{G}}$ 是幂零的，即 $\mathcal{M}_{\mathcal{G}}^k = 0$。这种幂零性施加了额外的代数约束，将 $\mathcal{M}_{\mathcal{G}}$限制在子簇 $V^\star = \langle$Tr$(\mathcal{M}_{\mathcal{G}})$，det$(\mathcal{M}_{\mathcal{G}})\rangle$中。该簇的维数为

$$\dim V^\star = (2p)^2 - 1 - (2p) = 4p^2 - 2p - 1.$$

现在，两个代 维度分别为 d_1 和 d_2 的两个代数簇在维度为 n 的环境空间中满足。 s

$$\dim(V_1 \cap V_2) \ge d_1 + d_2 - n.$$

将此应用于我们的情形，可得

$$\dim(\text{variety satisfying all constraints}) \ge (4p^2 - 1 - p^2) + (4p^2 - 2p - 1) - 4p^2 = 3p^2 - 2p.$$

因此，与 $\mathcal{I}^\star$ 相关联的子簇的维数至少为 $3p^2 - 2p$。 □

如引理 C.5所示，我们至少需要额外的 $3p^2 - 2p - 1$ 个线性无关多项式来排除 $\mathcal{V}$，从而识别出所有 $2p \times 2p$ 增广核矩阵。

注意，我们有 p 个过程；通常假设我们从至少K 个环境中获得不同的上下文信息（即 $K = p$），这在因果表征学习中是一个温和的条件。每个子理想$\mathcal{I}_k$： $\langle F_{\mathcal{G}} - K_{\mathcal{G}}\ (\mathbb{I}_{2p} - \mathcal{M}_{\mathcal{G}})\ \rangle$构成一个多项式系统，记为$\mathcal{S}_{\mathbb{P}}^k$，使得：

$$F_{\mathcal{G}} + K_{\mathcal{G}}^{(k)} \mathcal{M}_{\mathcal{G}}^{(k)} = K_{\mathcal{G}}^{(k)}. \tag{C.7}$$

$2n \times 2p$不定元用于 $F_{\mathcal{G}}$和 $2|e(\mathcal{G})|$，用于 $\mathcal{M}$，因为每个环境 k都会引入一个新的 $\mathcal{M}_{k,\ j}$。考虑所有 K 理想（$\mathcal{I}_0$，$\mathcal{I}_2$，$\cdots$，$\mathcal{I}_K$），我们得到所有 K 代数簇的并集：

$$V(\mathcal{I}) = \left\{ (F_{\mathcal{G}}, \mathcal{M}_{\mathcal{G}}^{(0)}, \ldots, \mathcal{M}_{\mathcal{G}}^{(K)}) \;\middle|\; F_{\mathcal{G}} - K_{\mathcal{G}}^{(k)} \left(\mathbb{I}_{2p} - \mathcal{M}_{\mathcal{G}}^{(k)}\right) = 0,\ \forall k \in K \right\}. \tag{C.8}$$

通过将 0 对应的式子从 C.7 对应的式子中相减来添加多项式约束，得到 k：

$$V(\mathcal{I}^\star) = \left\{ V(\mathcal{I}) \;\middle|\; K_{\mathcal{G}}^{(k)} \mathcal{M}_{\mathcal{G}}^{(k)} - K_{\mathcal{G}}^{(0)} \mathcal{M}_{\mathcal{G}}^{(0)} - \left(K_{\mathcal{G}}^{(k)} - K_{\mathcal{G}}^{(0)}\right) = 0,\ \forall k \in K \right\}. \tag{C.9}$$

它在多项式环 $R = k[F_{\mathcal{G}_{i,j}}, \mathcal{M}_{i,j}^k]$ 中诱导出坐标环 $R/\mathcal{I}$。坐标环的阶数是原始代数簇的维数。我们用 $\begin{pmatrix} & \end{pmatrix}$ 表示分块系统，以免读者将其与矩阵括号混淆。对于每个 $m \in [n], j \in [p]$ 和 $i \in$ ch$_{\mathcal{G}}(j)$，我们有 $|\text{ch}_{\mathcal{G}}(j)|$ 个属于 $\mathcal{M}_{\mathcal{G}}$ 的列。 $F_v, \mathcal{M}_v^{(k)}$将条目 $f_{i,j}, \mathcal{M}_{m,i}^k$ 写成向量，从而将多项式约束描述为

一个线性系统：

$$\left(\begin{array}{c|c} \mathbb{I}_{4np} & \star \\ \hline \mathbf{0} & \star \end{array}\right)\begin{pmatrix} F_v \\ \mathcal{M}_v^{(0)} \\ \mathcal{M}_v^{(k)} \end{pmatrix} = \left(\begin{array}{c} K_{\mathscr{G}}^{(k)} \\ \hline K_{\mathscr{G}}^{(k)} - K_{\mathscr{G}}^{(0)} \end{array}\right). \tag{C.10}$$

在 INAR (∞) 中，每个 $\mathcal{M} := \mathcal{M}_{t-s}$ 保留所有路径 $\{(\varphi(j) \to \varphi(i)|Z^{\Delta}_{\varphi(j),\ t-s} \to Z^{\Delta}_{\varphi(i),\ t},\ \mathcal{M}_{i,\ j} \neq 0\}$ 从 Φ_{t-s} 经由同构 φ得到。因此，元素 $(\mathbb{I}_{2p} - \mathcal{M})^{-1}_{i,\ j}$ 是路径 $j \to m \to n \to i$对应的 $\mathcal{M}_{n,\ m}$ 的乘积。上下文清楚时，我们省略时间索引和图标签。这样的 $(\mathbb{I}_{2p} - \mathcal{M})^{-1}_{i,\ j}$ 可表示为

$$(\mathbb{I}_{2p} - \mathcal{M})^{-1} = \mathbb{I} + \mathcal{M}.$$

采用了 [Carreno 等人的结果， 2024] 以得到秩$(\mathbb{I}_{2p} - \mathcal{M}^{(k)})^{-1} - (\mathbb{I}_{2p} - \mathcal{M}^{0})^{-1} \leq 1$。利用 引理 C.4，我们得到秩$(F_{\mathscr{G}}(\mathbb{I}_{2p} - \mathcal{M}^{(k)})^{-1} - F_{\mathscr{G}}(\mathbb{I}_{2p} - \mathcal{M}^{0})^{-1}) \leq 1$。因此，式（C.10）的左侧部分的列彼此成比例：

$$(K_{\mathscr{G}}^{(k)} - K_{\mathscr{G}}^{(0)})_{l,j} = (K_{\mathscr{G}}^{(0)})_{l,k}\Delta_{k,i}, \quad \Delta := (I - \mathcal{M}^{(k)})^{-1} - (I - \mathcal{M}^{(0)})^{-1} \tag{C.11}$$

我们考察式（C.10）中右下块 $\star$的非零子块，其大小为 $|\mathsf{de}(j) \setminus \mathsf{ch}(j)| \times |\mathsf{ch}(j)|$。按照惯例，我们将子块表示为 $M[j]$，并选择与 $M[j]$ 的大小相对应的较小块 $[K_{\mathscr{G}}^{(k)} - K_{\mathscr{G}}^{(0)}]$，记作 $b[j]$。簇 $V(\mathcal{I}^{\star})$ 的维数是满足以下条件的点 $(\mathcal{M}^{0}_{i,j})$，$i \in \mathsf{ch}(j)$ 的维数：

$$M[j](\mathcal{M}^{0}_{i,j}) = b[j] \tag{C.12}$$

当上述约束无解时，该簇为空集。因此，我们要求 $\mathsf{rank}(M[j])$ $=\mathsf{rank}(M[j]|b[j])$。$[M[j] \mid b[j]]$ 是用于检验多项式方程组稳定性的共同增广矩阵。我们将在下一个引理中正式陈述关于 $\mathcal{V}(\mathcal{I}^{\star})$维数的结论，以此结束证明。

引理 C.6。 对于由累积量分解得到的一般的 F 和 FH，当且仅当簇 $\mathcal{V}(\mathcal{I}^{\star})$ 的维数为零时，完整的生成模型是可辨识的，也就是说，

$$\dim(\mathcal{V}(\mathcal{I}^{\star})) = \sum_{j=1}^{q} \mathrm{ch}(j) - \mathrm{rank}(M[j]) = 0.$$

证明。对于左侧子集 U_s，每个节点 j 的出边仅指向右侧子集 V_t 中的节点 i，而这些节点都是 j 的直接子节点。根据构造，U_s 内部和 V_t 内部均不存在边。因此，对于每个 j，都有 $M[j] = \emptyset$，因为 $\mathsf{de}(j) = \mathsf{ch}(j)$。 □

当 $\mathcal{M}_{t-s}$ 被完全恢复后，可得到完整矩阵 $\mathcal{F}_{\mathscr{G}}$，如下所示

$$\mathcal{F}_{\mathscr{G}} = K_{\mathscr{G}}(\mathbb{I}_{2p} - \mathcal{M}_{\mathscr{G}}).$$

若 F_v 是单射，则由于将 F^{-1}（或伪逆 $F^{\dagger}$）直接与 K相乘，识别 F_v（从而识别 $\mathcal{F}_{\mathscr{G}}$和 F）等价于分别识别每个 $\mathcal{M}$。然而，F的一般性阻碍了对完整生成模型的识别。即使 $K_{\mathscr{G}}$ 在通常的不确定性下唯一，恢复其他核矩阵 $\mathcal{M}_{t-s'}$ 仍需要针对各个核矩阵的类似可辨识性条件。因此，模型的可辨识性取决于其嵌入的几何复杂性。具体而言，环境射影空间 $\mathbb{P}^n$ 中至少需要 p 个一般点。如果这些一般点仅通过改变参数 ϕ得到，那么所得点就位于同一个射影轨道

$$\mathcal{O}_{\phi} = \{[\Phi(\phi + \delta)] \mid \delta \in \mathbb{R}^k\}$$

在……之下 参数偏移。因此，可辨识性仅在这种等价关系下成立，也就是说 is,

$$\phi_1 \sim \phi_2 \quad \Longleftrightarrow \quad \Theta(\Phi_{i:}(\phi_1)) = \Theta(\Phi_{i:}(\phi_2)),$$

因此，只要同一过程的不同参数化对应着相同的参数行对 $\Theta(\Phi_i)$，它们就会变得无法区分。因此，在连续时间下，只要改变 $\Theta(\Phi_i)$，该过程就会发生自然的平移。由此，在 $k \in 1$、2、..... K 个不同的情境下——每个情境都会在分布中引入充分变异性，或确保每个滞后 k 至少受到一次干预，以改变下游机制——完整的潜在结构在相同的不确定性范围内是可识别的。

基线 U 的恢复。 一旦恢复了完整的因果结构（除一个缩放与置换矩阵外），就可以直接计算 $\mathbb{E}[F(\mathbb{I}_p - \Phi)^{-1} \star (U + \epsilon_t)]$ 的低阶矩，以找出在相同不确定性范围内的 U。

关于条件的说明。 白噪声确保过程 ϵ_t 的所有分量相互独立，因此，按照 $\epsilon(\Delta)$ 中的傅里叶变换，积分除 $\Delta = 0$ 外恒为零，并使代数累积量 $\kappa(O_t)$ 呈现出简化且易于解析的形式，从而允许唯一分解。我们着重指出这样一种情况：若选择随机维纳过程 dW_t 代替白噪声扰动，则所提出的条件会受到严重违反。因此，识别完全随机的随机微分过程仍然具有挑战性，因为应用伊藤积分后，增量独立扰动仍会形成时变噪声。为完整起见，我们在 C.5 中展示如何将相关噪声化为独立增量噪声，同时保持可识别性。

C.2.4 因果结构的识别

我们的可识别性涵盖了任意软干预的情形，因此只需证明对于一般的分布偏移，识别仍然成立。对于通过软干预得到的元组 $(d,\ K,\ \{\kappa_d(O_t)\}_{0:K})$，若 $\Phi(w)$中条目 $(i \to j)$ 的任何固定值都不会导致依赖关系消除，则可将真实过程的传递闭包 $\mathcal{G}$ 及其因果结构恢复到前述的平凡变换为止。没有瞬时影响的时间过程必须满足 $TC(\mathcal{G}) = \mathcal{G}$，因此我们可以在缩放和置换 π下恢复原过程及其因果结构。

不失一般性，我们可以放心地选择阶数为 p 的核矩阵，因为弱收敛成立。因此，我们将连续核矩阵离散化至 $t - s = \tau$，每个 $\tau \in (\tau_1,\ \cdots,\ \tau_p)$。对上述方程应用傅里叶变换，取阶数为 p，得到：

$$\mathcal{F}[F(\mathbb{I}_{2p} - \mathcal{M}_{\mathcal{G}_K})^{-1}] = \mathcal{F}\left[F\sum_{p=0}^{n}\mathcal{M}_{\mathcal{G}_K}^{\star p}[\tau]\right] = F\sum_{p=0}^{n}(\mathcal{M}_{\mathcal{G}_K}(\omega))^p \tag{C.13}$$

在满足秩 条件的情况下，我们构造差分矩阵 $\Delta = H[\tau]^0 - H[\tau]^k$，然后

$$\begin{aligned}\Delta &= F(\mathbb{I}_{2p} - \mathcal{M}_{\mathcal{G}_K})^{-1}(w)^{(0)} - F(\mathbb{I}_{2p} - \mathcal{M}_{\mathcal{G}_K})^{-1}(w)^{(k)} \\ &= F(\mathcal{M}_{\mathcal{G}_K}^{(0)}(w) + \mathcal{M}_{\mathcal{G}_K}^{2(0)}(w) + \cdots + \mathcal{M}_{\mathcal{G}_K}^{p(0)}(w) - \mathcal{M}_{\mathcal{G}_K}^{(k)}(w) + \mathcal{M}_{\mathcal{G}_K}^{1(k)}(w) + \cdots + \Phi^{p(k)}(w))\end{aligned}$$

C.3 证明定理 2

C.3.1 定理2的预备知识定理2

我们通过证明元组 (f, Φ, U) 在逐分量缩放和置换下可识别，来证明主定理。我们的证明基于定义多项式环 K^n 中特殊超曲面的伴随簇的维数。我们研究累积量结构的非线性传播，以确定 INAR(∞) 过程的可辨识性条件。给定一个一般的非线性 f，其精确累积量

$\kappa_d(O_t)$ 遵循具有贝尔多项式系数的阶数 d 展开。因此，我们将假设 3重述如下。

假设3（重述）设观测数据由未知混合函数生成 f，该混合函数作用于由线性强度驱动的潜在随机过程 $Z_t^{(\Delta)}$ λ_t。假设：

1. f 是一个一般的 C_d 映射，其雅可比矩阵 J_f 几乎必然具有满秩。

2. 对于 $d \in D := \{d \mid \bar{\kappa}_{d+1}(\Delta O_t) = 0\}$，至少存在 p 个非零张量 $\bar{\kappa}_d(\Delta O_t)$，其中 $\bar{\kappa}_d(\Delta O_t)$ 是由 $O_t := f(Z_t)$ 的一阶泰勒展开计算得到的差分累积量。

3. 理想 $\mathcal{I}^\star$： $\langle J_f - K_{\mathscr{G}}^{(k)}(\mathbb{I}_{2p} - \mathcal{M}^{(k)})\ k = 1、2、\cdots、p\rangle$ 具有零维伴随簇 $\mathcal{V}(\mathcal{I}^\star)$

引理 C.7。设 $\mathbb{Q}$ 为基域。假设 f 是一般的，其雅可比矩阵为 J_f。以下是对 J_f的一般性的等价刻画：

1. J_f 的各项是关于 $\mathbb{Q}$代数独立的可交换不定元；等价地，它们生成了 $\mathbb{Q}$的纯超越扩张。*2.* 点 $\big((J_f)_{i,j}\big)$ 不属于多项式环$\mathbb{Q}[x_{ij}]$中任何非零多项式的零点集。*3.* 因此， J_f 在一个扎里斯基开稠密子集上满秩；特别地，它几乎必然满秩。

通常，与可辨识性分析中的大多数假设一样，一般性假设在实践中几乎无法验证。令人惊讶的是，从概率意义上说，该假设仍然几乎必然成立。回顾一个简单的例子，考虑一般矩阵 F。

当我们说一个矩阵是一般的时，并不是说它是通过给矩阵各项任意赋予数值而得到的。相反，矩阵的每一项都被视为一个纯形式符号——一个代数变量——并未被赋予任何具体值。等价地，我们是在这些符号构成的有理函数域上进行运算，因此矩阵的各项是代数独立的不定元。这确保了该矩阵不会满足任何退化代数关系，除非位于某个真代数子簇上（一个在欧几里得意义下测度为零的集合）。因此，尽管这一假设无法通过数值检验，但在随机选取下它几乎必然成立；通过将矩阵各项视为代数独立符号，这一假设也得到了严格的形式化。

接下来，我们通过例2说明，假设混合函数具有一般性是多么自然。

例2（矩形雅可比矩阵： f： $\mathbb{R}^2 \to \mathbb{R}3$，混合超越元）。令 $x = (x_1,\ x_2) \in \mathbb{R}^2$ 并定义

$$
\begin{aligned}
f_1(x) &= \sin(\alpha_1 x_1) + e^{\beta_1 x_2}, \\
f_2(x) &= \cos(\alpha_2 x_1 + \alpha_3 x_2) + x_1^2, \\
f_3(x) &= x_1 x_2 + e^{\beta_2 x_1 + \beta_3 x_2},
\end{aligned}
$$

其中所有系数 α_i, β_i均视为代数独立符号。雅可比矩阵 $J_f(x) \in \mathbb{R}^{3\times 2}$ 为

$$
J_f(x) = \begin{pmatrix}
\alpha_1 \cos(\alpha_1 x_1) & \beta_1 e^{\beta_1 x_2} \\
-\alpha_2 \sin(\alpha_2 x_1 + \alpha_3 x_2) & -\alpha_3 \sin(\alpha_2 x_1 + \alpha_3 x_2) + 2x_1 \\
x_2 + \beta_2 e^{\beta_2 x_1 + \beta_3 x_2} & x_1 + \beta_3 e^{\beta_2 x_1 + \beta_3 x_2}
\end{pmatrix}.
$$

各元素的代数独立性。 $J_f(x)$ 的每个元素要么包含一个唯一的符号因子（α_i, β_i），要么依赖于 x_1, x_2的一个不同的线性组合。通过将系数视为代数独立符号，可以看出各元素之间在 $\mathbb{Q}$上不存在非平凡的多项式关系。因此， $J_f(x)$ 的各元素在基域 $\mathbb{Q}(\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3)$上是代数独立的，从而$J_f(x)$ 在可辨识性理论的意义下是一般的。

一般混合 f 由其雅可比矩阵 J_f所保持；因此，辨识 J_f 等价于在至多相差一个常数的意义下恢复 f 。现在，我们已准备好证明主定理。不失一般性，我们将 J_f 写成 F ，因为它们在代数闭域中的行为相同。

C.3.2 定理2的证明

证明。 我们研究累积量结构的非线性传播，以寻找 INAR过程的可辨识性条件。给定一个一般的非线性 f，其精确累积量 $\kappa_d(O_t)$ 遵循带有贝尔多项式系数的 d 阶指数展开。对于光滑映射 $f: Z_t \to f(Z_t)$，我们可以使用泰勒展开构造 $O_\cdot = f(Z_{t+\Delta t})$:

$$f(Z_{t+\Delta t}) = f(Z_t) + \frac{\partial}{\partial Z_t} f(Z_t)\Delta Z_t + \frac{1}{2}\frac{\partial^2}{\partial Z_t^2} f(Z_t)(\Delta Z_t)^2 + o(\Delta Z_t^3)$$

此时，展开表现出相当异常的行为，因为阶数可能大得难以承受。然而，正如下文所解释的，一阶截断展开具有理论上的吸引力。令高阶分量˜为 $\mathcal{R}(1)$，并回顾 $Z_t = H \star \epsilon_t$，我们得到截断微分过程 $\Delta f(Z_t)$，记为:

$$\Delta f(Z_t) - \mathcal{R}(1) = J_f \sum_{k=1}^{t} H_{t-s}\epsilon_s \tag{C.14}$$

我们将式（C.14）中出现的所有量视为多项式环中的不定元

$$R = k\big[\{\Delta f(Z_t)\}_t,\ \mathcal{R}(1),\ J_f,\ \{H_{t-s}\}_s,\ \{\epsilon_s\}_s\big],$$

其中， k 是一个基域，例如 $\mathbb{R}$ 或 $\mathbb{C}$。对于每个时间索引 t，定义多项式为 $g_t := \Delta f(Z_t)$ _ $\mathcal{R}$ 1 ( ) $- J_f \sum_{s=1}^{t} H_{t-s}\epsilon_s \in R$. 该多项式生成主理想 $\mathcal{I}_t = \langle g_t \rangle \subset R$，并考虑所有时间索引 $t = 1$、 2、.....、 T，我们得到全局理想 $\mathcal{I} = \langle g_1$、 g_2、.....、 $g_T \rangle \subset R$。相应的簇为

$$V(\mathcal{I}) = \Big\{(Z_t, \Delta f(Z_t), J_f, H_{t-s}, \epsilon_s, \mathcal{R}(1)) \in k^N \;\Big|\; g_t = 0 \text{ for all } t\Big\}.$$

显然， $V(\mathcal{I})$ 是正维的，因为其定义关系并未指定有限个点。为获得更多结构，我们考察高阶统计量。特别地，变换后增量的 d 阶累积量张量具有如下形式。

$$\kappa_d\Big(\Delta\tilde{f}(Z_t)\Big) = \sum_{s=1}^{t}\sum_{j=1}^{p} \kappa_d^{(j)}(\epsilon) \cdot \Big(J_f H_{t-s}^{(:,j)}\Big)^{\otimes d}.$$

这一表达 式表明，累积量自然地在射影张量空间中定义了一个点 e.

$$\mathbb{P}(V^n \otimes V^n \otimes \cdots \otimes V^n),$$

其中张量因子的数量等于 p。因此，尽管仿射簇 $V(\mathcal{I})$ 过于庞大，无法实现可识别性，但累积量张量将问题提升到了射影几何框架中；在此框架下，与 Veronese嵌入的割线簇之间的联系为研究唯一性与分解提供了自然的途径。 □

此后，一般的 f 具有由退化线性截断累积量所形成的代数结构。这种退化形式确保 Veronese 嵌入定义在更大的环境空间中。注意，$\mathcal{R}(1)$完全由 f 决定，因此可以通过一个优化问题求得：选择一个初始 $\mathcal{R}(1)$，使得 $\kappa_d\Big(\Delta\tilde{f}(Z_t)\Big)$ 具有稳定的唯一分解解。我们指出，只要解存在且唯一，这一优化便表明整个生成模型的恢复能够得到保证。也可以采用其他任何技术来估计生成模型。正如正文所示，为了利用生成模型的计算能力，我们采用变分方法来建模因果动力学和混合映射。

C.4 定理 3 的证明

最后，关于我们的可辨识性理论，我们证明，所提出的条件对于识别随机过程的完整生成模型既是充分的，也是必要的。注意，在讨论随机过程时，它随时间包含无限多个变量（分别地时间滞后时间序列）；因此，我们并不打算恢复所谓的因果“变量”，而是聚焦于具有完整参数的过程生成模型。

⇒ 假设带来可识别性：

对于充分性，按照我们对定理 2的证明即可轻松得证。

⇐ 可识别性意味着假设：

假设即使违反假设 3中的某些条件，完整的生成模型仍然是可识别的。那么，存在参数空间的一个线性混合 $F\Phi$，它只能在逐分量变换 g' 和置换 π' 下唯一确定，且 $\kappa(\epsilon)$ 和 π不同。

因此，重构出的观测数据 O_t 满足

$$\kappa(O_t) = \sum_{j=1}^{p} g'(F\hat{H}_j)^{\otimes d},$$

它具有唯一分解。

然而，这与假设违反假设 3中的条件相矛盾，因为只有在所有这些条件都成立时，才应存在唯一分解。因此，完整生成模型的可识别性意味着假设 3 中的所有条件都必须成立。

C.5 变化且相关的噪声

我们构造了一个带有驱动噪声的模型，其中驱动噪声并非时间独立的，而是一个连续随机过程 R_t。为了利用上述证明，我们对观测混合流形O_t 应用伊藤引理。将带有卷积核的因果过程代入：

$$\begin{aligned}
\Delta Z_t &= (Z_{t+\Delta t} - Z_t) \\
&= \left[(u_0 + \int_0^{t+\Delta t} \Phi(t+\Delta t - s) Z_s \; ds + R_{t+\Delta t}) - (u_0 + \int_0^{t} \Phi(t - s) Z_s \; ds + R_{t}) \right] \\
&= \left[(\int_0^{t+\Delta t} \Phi(t+\Delta t - s) Z_s \; ds) - (\int_0^{t} \Phi(t - s) Z_s \; ds) \right] + \Delta R_t
\end{aligned}$$

对于第一个积分，我们在 $t-s$ 处应用泰勒展开：

$$\begin{aligned}
\Delta Z_t &= Z_{t+\Delta t} - Z_t \\
&= \left\{ \int_0^{t+\Delta t} \left[\Phi(t-s) + \frac{\partial}{\partial \Delta t}\Phi(t-s)\Delta t + \frac{\partial^2}{\partial \Delta t^2}\Phi(t-s)(\Delta t)^2 \right] Z_s \, ds \right. \qquad \text{(C.15)} \\
&\quad \left. - \int_0^{t} \Phi(t-s) Z_s \, ds \right\} + \Delta R_t \\
&= \int_t^{t+\Delta t} \Phi(t-s) Z_s \, ds + \int_0^{t+\Delta t} \left[\frac{\partial}{\partial \Delta t}\Phi(t-s)\Delta t + \frac{\partial^2}{\partial \Delta t^2}\Phi(t-s)(\Delta t)^2 \right] Z_s \, ds + \Delta R_t \\
&= \Phi^* Z_t \Delta t + \Delta t \int_0^{t+\Delta t} \frac{\partial}{\partial \Delta t}\Phi(t-s) Z_s \, ds + (\Delta t)^2 \int_0^{t+\Delta t} \frac{\partial^2}{\partial \Delta t^2}\Phi(t-s) Z_s \, ds + \Delta R_t \\
&= \Phi^* Z_t \Delta t + \Delta t \, (G^{(1)} \star Z_t) + (\Delta t)^2 \, (G^{(2)} \star Z_t) + \Delta R_t \qquad \text{(C.16)}
\end{aligned}$$

进一步说明 在不失一般性的卷积情况下，我们要求 $s < t$，第一次和第二次展开的部分可简洁地写成由新核函数定义的退化卷积 $G^{(1)} := \Phi'(t-s)$ 和 $G^{(2)}$。

于是，$(\Delta Z_t)^2$ 需要将其所有二次项分别按 1、2 和 4 阶展开：

$$\begin{aligned}(\Delta Z_t)^2 &= \left(\Phi^* Z_t \Delta t + \Delta t (G^{(1)} \star Z_t) + (\Delta t)^2 (G^{(2)} \star Z_t) + \Delta R_t\right)^2 \\ &= \left(\Phi^{*2} Z_t^2 \Delta^2 t + \Delta^2 t (G^{(1)} \star Z_t)^2 + \Delta^4 t (G^{(2)} \star Z_t)^2 + \Delta^2 R_t\right) \\ &\quad + 2\left(AB + AC + AD + BC + BD + CD\right) \\ &= \left(\Phi^{*2} Z_t^2 \Delta^2 t + \Delta^2 t (G^{(1)} \star Z_t)^2 + o(\Delta^4 t) + \Delta^2 R_t\right) \qquad \text{(C.17)} \\ &\quad + 2\left(AB + o(\Delta^3 t) + AD + o(\Delta^3 t) + BD + CD\right) \qquad \text{(C.18)}\end{aligned}$$

我们将论证，$(\Delta Z_t)^2$ 中的所有项都是通过对核 $\Phi(t-s)$ 和 $f(Z_{t+\Delta t})$ 进行泰勒展开得到的。因此，从卷积核推导伊藤引理，需要仔细研究每个无穷小量的阶，以及它们相对于随机噪声增量阶数的极限分布行为。增量噪声的阶数与潜在因果随机过程能否实现完全可识别，以及实现这一目标的难度，密切相关。

C.6 瞬时影响下的识别

本节简要讨论模型存在瞬时影响的情形，并着重说明恢复整个生成模型的复杂性；这种复杂性与求解一个二次方程组的复杂性相当。

在模型不存在瞬时影响的情况下，我们的主定理给出了可识别性结果。在这种情况下，$\mathcal{M}$ 不仅是一个上三角矩阵，其元素是不定元，并且属于由特征多项式定义的射影簇 $\mathbb{P}^{4p^2-2p-1}$，而且其大小为 $p \times p$ 的右上角分块中的元素也都严格非零。考虑如下形式的增广矩阵 $F_{\mathscr{G}}$：

$$\left[\begin{array}{ccc|ccc} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ f_1 & \cdots & f_q & \mathrm{e}_{q+1} & \cdots & \mathrm{e}_{2q} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathrm{e}_1 & \cdots & \mathrm{e}_q & f_1 & \cdots & f_q \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{array}\right]$$

我们认为这个增广矩阵是一般的，因为在射影空间 $\mathbb{P}^d$ 中，几乎每个点都对应一种构型，其中任意两个方向都不会重合；也就是说，相关直线仅在无穷远处相交，因此不会出现线性退化。该增广矩阵一般，是指它属于 $\mathbb{P}^d$ 的一个扎里斯基开稠密集，因此用其增广形式替换 $F_{\mathscr{G}}$ 不会改变 $\dim(V)$。然而，存在瞬时影响时，定义方程会施加代数约束，从而使这个开集退化。在这种情况下，唯一可行的增广方式是强制其他所有块均为零，因此无法构造非平凡的一般矩阵。假设 INAR 模型包含由矩阵 B 表示的瞬时影响。该模型如下所示，

$$BX_t = A_1 X_{t-1} + A_2 X_{t-2} + \cdots + A_p X_{t-p} + u_t, \qquad \text{(C.19)}$$

其中：

B 是一个非对角矩阵，用于捕捉变量之间的同期（瞬时）影响。

u_t 表示结构冲击，通常假设其满足 $\mathrm{Cov}(u_t) = I$。

由此可得，无穷阶模型为

$$O_t = FB^{-1}(\mathbb{I}_p - \Phi)^{-1} \star \epsilon_t \tag{C.20}$$

显然，按照我们的推理，仍然可以恢复混合参数空间 $K' = FB^{-1}(\mathbb{I}_p - \Phi)^{-1}$ ，但结果仅确定到缩放和置换。然而，得到的矩阵构成一个三次多项式系统，需要更多约束才能确定各个参数空间。

断言 C.1。 对于具有瞬时影响的 *INAR* 模型，定理 *2* 永远不足以恢复整个生成模型

D 扩展可识别性结果

D.1 层级最小性的存在性

我们已证明，INAR（∞）下的因果解耦可由定理 1和定理 2保证。接下来，这个问题可归结为寻找最小累积量层级结构（复杂度），并搜索 d 以尽可能简洁地达到这一复杂度。作为启发，我们还表明，代数几何性质唯一确定了累积量复杂度的最小性。因此，可以从几何角度控制潜在因果结构的可识别性。随后，我们展示了若干示例，说明如何在任意数据流形下寻找最小复杂度。

注释 1。 在本节中，我们暂且忽略弱收敛类，考虑不再要求弱收敛的所有广义情形。作为一个简单示例，我们将由任意噪声族驱动的潜在因果时间过程进行线性满秩多项式混合，作为对定理 *1*的一种变体。

我们首先讨论不同噪声过程在路径不可微过程或半鞅情况下的一系列变化。我们回顾两种基本识别理论，并通过发现不同的层次结构最小性，阐明我们的定理严格涵盖了这两种理论。

命题 D.1 （特殊噪声下的进一步可识别性）。 给定定理 *2*中定义的一列随机过程，潜在因果表示 $\mathcal{G}$ 当且仅当噪声服从拉普拉斯分布时，除符号翻转和置换外可识别。

命题 D.2 （线性混合下可识别性降低，改编自 F，取自 Carreno 等人 [2024]）。 给定定理 *2*中定义的一列随机过程，潜在因果结构 $\mathcal{G}$ 除缩放和置换外可识别。

现在，我们关注一种更复杂但更有用的情形，在该情形中，噪声过程的行为是任意的。

定理 4 （有限噪声支撑上的识别）。

(*1.*) 在定理 *1*下，通过寻找 d_0 以满足假设 *2*中提出的最小累积量层级关系，可保证可识别性。

(*2.*) d_0 和 T 是可处理的。 (*3.*) 假设 *2*仍然是充分且必要的。

根据标准张量代数，线性变换 F 通过多重线性变换（当 F 是线性映射时）或多项式映射（当 F 是满秩多项式函数时），将累积量信息从观测空间传递到潜在空间。在后一种情况下，由此产生的多项式方程组复杂度呈组合式增长，反映了多项式结构与潜变量高阶累积量之间的相互作用。值得注意的是，所有此类方程都受到底层噪声过程相同的尺度和分布结构支配。

D.2 证明命题 D.1

我们采用**频域变换**来刻画连续时间因果影响和标准因果转变。重要的是，该变换仅在潜在空间中应用，而可观测量在时域中仍为实值，这与真实世界的数据相符。

备注 在频域中，变量 f 作为连续索引，用于表征信号的谱行为。然而，该域中的分布偏移——尤其是涉及复值结构的偏移——往往体现在细粒度层面，标准的基于似然的密度建模无法捕捉这些特征。相比之下，功率谱密度 *(PSD)* 及其高阶扩展（例如双谱、三谱）等统计表示，能够更忠实地刻画分布在不同频率间的结构依赖关系。这些量与信号的底层累积量直接相关，因此提供了一种自然的多尺度视角，用于检测和解释干预引起的偏移。与充分变异性等传统准则不同，后者通常较为粗略，而高阶累积量和谱则能够更细致地识别各统计阶次上的结构变化。

设 $S_X(f)$ 表示随机时域信号 X_t 的功率谱密度，并将自相关函数定义为

$$R_X(\Delta) := \mathbb{E}[X_t X_{t+\Delta}].$$

根据维纳–辛钦定理， $S_X(f)$ 是 $R_X(\Delta)$ 的傅里叶变换 .

我们将这一功率谱密度分析应用于定义为卷积的变换后潜变量:

$$Z_t^{\Delta} = K_t \star R_t,$$

并在频域中分析其结构。

$$\begin{aligned} S_Z(f) = S_{K\star R}(f) &= ||K(f)||_F^2 S_R(f) \qquad (D.1)\\ &= ||K(f)||_F^2 S_{\epsilon+u}(f)\\ &= ||K(f)||_F^2 (S_\epsilon(f) + S_u(f))\\ &= ||K(f)||_F^2 \frac{1}{2\pi}\int_G R_R(\Delta) e^{-i\Delta f}\, d\Delta \qquad (D.2)\end{aligned}$$

第一个方程描述了原始信号与经过滤波器核处理的复合信号之间的关系。随后两个方程是通过对域 G 应用傅里叶变换得到的，该域分为两种情况： $\Delta = 0$ ，以及其他情况。上面的结果称为功率谱密度矩阵，其中每个元素 $[S_Z(f)]_{ij}$ 描述了所有时间滞后 Δ 的完整密度信息。 $\delta(f)$ 是狄拉克δ函数， A^H 是共轭转置。我们的可识别性建立在潜在空间功率谱密度以及 $Z_t^{(\Delta)}$ 与 $\hat{}Z_t^{(\Delta)}$ 之间的重构等价性之上。

我们从式（D.5）所示的过度简化情形开始。请注意，功率谱密度矩阵表示不同频率下总功率的分布，因此包含能够捕捉不同潜在随机过程之间相互作用的互谱分量。令$S_Z(f)$ 为一个对角矩阵，使得频域中的所有互谱消失，即

$$[S_Z]_{ij} = \sum_{j=1}^{p}\sum_{i=1}^{p} K_{ik} S_{R,kl} \overline{K_{lj}} = 0,\ i \neq j \qquad (D.3)$$

我们首先重新审视变量替换变换这一技术分析的基础方法，该方法常用于因果表征学习框架中。在此，我们推导其在频域中的对应形式。

事实 2。 设 Z_t 为实值潜在过程，并设 $h := \hat{}f \circ f^{-1}$ 为应用于 Z_t 的重构映射。若噪声服从拉普拉斯分布，则变换后的微分过程 $dh^{-1}(Z_t) := d\hat{}Z_t$ 的功率谱密度（*PSD*）满足:

$$S_{dh(Z)}(f) = J_{h^{-1}} S_{dZ}(f) J_{h^{-1}}^T, \qquad (D.4)$$

其中， J_h 是 h 的雅可比矩阵。

证明。 根据功率谱密度的定义，我们有：

$$S_Z(f) = \frac{1}{2\pi}\int_0^\infty R_Z(\Delta)e^{-i\Delta f}\, d\Delta,$$

其中 $R_Z(\Delta) := \mathbb{E}[Z_t Z_{t+\Delta}^T]$ 表示自相关函数。对于变换过程 $h(Z_t)$，其自相关函数为：

$$R_{h(Z)}(\tau) = \mathbb{E}[h(Z_t)h(Z_{t+\tau})^T].$$

将变量替换公式应用于微分 $dh(Z_t)$，我们需要：

$$\mathbb{E}[dh(Z_t)dh(Z_{t+\tau})^T] = J_h\,\mathbb{E}[dZ_t dZ_{t+\tau}^T]\,J_h^T.$$

当映射 h 为仿射映射时，这一条件成立，即 h^{-1} 必须是线性映射。此结果改编自 Klindt 等人 [2021]。因此，变换后变量的自相关函数通过雅可比矩阵与 Z_t 的自相关函数呈线性关系，功率谱密度也通过傅里叶变换呈线性关系。由此证明了等式（D.4）中的恒等式。 □

根据推导出的功率谱密度等价公式，若学习器编码器使估计变量与真实变量之间的分布相匹配，则满足与常规时域概率空间中类似的条件：

$$S_{\hat{Z}:=h(Z)}(f) = S_Z(f) \tag{D.5}$$

代入此前推导出的所有项：

$$\mathcal{F}\{(I-\Phi_t)^{-1}\}(S_\epsilon + uu^T\delta(f))\mathcal{F}\{(I-\Phi_t)^{-1}\}^H = J_h S_Z J_h^T \tag{D.6}$$

求出该关系中的每一项：

$$\begin{bmatrix} K_{f,11} & \dots & \dots \\ \vdots & \ddots & \vdots \\ K_{f,1p} & \dots & K_{f,pp} \end{bmatrix} \begin{bmatrix} S_\epsilon(f)_{11} + u_1^2\delta(f) & \dots & \dots \\ \vdots & \ddots & \vdots \\ S_\epsilon(f)_{p1} + u_p u_1\delta(f) & \dots & S_\epsilon(f)_{pp} + u_p u_p\delta(f) \end{bmatrix} \begin{bmatrix} \overline{K_{f,11}} & \dots & \dots \\ \vdots & \ddots & \vdots \\ \overline{K_{f,1p}} & \dots & \overline{K_{f,pp}} \end{bmatrix}^T \tag{D.7}$$

$$= J_h S_Z J_h^T \tag{D.8}$$

其中， $X(f)$ 是关于 f 的谱的复共轭，而 ϕ_{ij} 是滤波器核函数。接下来，我们关注真实随机过程 Z_t 的可识别性。

如果移民参数 $u = 0$ 和噪声过程都是白噪声，这意味着它们的方差不随时间变化。那么， S_R 是方程（D.2）中的一个对角矩阵，该方程可写成更明确的形式：

$$\begin{aligned} &||K||_F^2 \frac{1}{2\pi}\int_0^\infty \mathbb{E}[R_t R_{t+\Delta}]e^{-i\Delta f}\, d\Delta \\ &= ||K||_F^2 \frac{1}{2\pi}\int_G \mathbb{E}[(\epsilon_t)\epsilon_{t+\Delta})]\delta(\Delta)e^{-i\Delta f}\, d\Delta \qquad \text{(D.9)} \\ &= ||K||_F^2 \frac{1}{2\pi}\{\int_{G=0} \mathbb{E}[(R_t)(R_{t+\Delta})]\delta(t)e^{-i\Delta f}\, d\Delta + \int_{G\neq 0} \mathbb{E}[(R_t)(R_{t+\Delta})]\delta(t)e^{-i\Delta f}\, d\Delta\} \qquad \text{(D.10)} \\ &= \mathcal{F}\{(I-\Phi_t)^{-1}\}\Sigma_{\epsilon_t}\mathcal{F}\{(I-\Phi_t)^{-1}\}^H \qquad \text{(D.11)} \end{aligned}$$

因此，为了使 t 条件成立时，唯一的解是使 K （因此 K^H）成为对角矩阵 ，

导致 K 和 K^T 应为置换缩放矩阵。也就是说，我们有：

$$P_\pi \begin{bmatrix} K_{f,11} & \dots & \dots \\ \vdots & \ddots & \vdots \\ 0 & \dots & K_{f,pp} \end{bmatrix} \Sigma \begin{bmatrix} K_{f,11} & \dots & \dots \\ \vdots & \ddots & \vdots \\ 0 & \dots & K_{f,pp} \end{bmatrix}^H P_\sigma = J_h S_Z J_h^T \tag{D.12}$$

K 的逆矩阵仍然是一个置换缩放矩阵，因此这意味着 $\mathcal{F}\{(I-\Phi_t)^{-1}\}$ 的每一列和每一行中也都只有一个非零值。随机延迟过程至少应当与自身相关，因此仅当其（i，j）项为 0 时，$[(I-\Phi_t)_f^{-1}]_{ii} \neq 0$，并使 $I-\Phi_f$ 成为一个置换缩放矩阵。因此，$I-\Phi_f$ 及其逆矩阵 K 都是对角矩阵，并且

$$\phi_{f,ij} = 0, i \neq j \tag{D.13}$$

这等价于 $\int_G \phi(t)_{ij} e^{2\pi i t} dt = 0$，并意味着 $\phi(t)_{ij} = 0$。根据构造，右侧因此应为对角矩阵，$S_{\hat{Z}}$也是如此。该条件可化简为：

$$\Lambda_Z = J_h \Lambda_Z J_h^T \tag{D.14}$$

我们总可以左乘 J_h^{-1} 并右乘 J_h^{-T} 以得到 $J_h^{-1}\Lambda_Z J_h^{-T} = \Lambda_Z$。由于 Λ_Z 是一个对角矩阵，因此 J_h 至多只能有一个非零元素，这表明 J_h 可以写成$P_\sigma$diag(k_1，k_2，k_3，.....，k_p)，从而得到置换矩阵与缩放矩阵的混合形式。现在我们说明可识别性还可以进一步提高。由于我们已经确立了 $K_f S_R K_f^H = J_h K_f S_R K_f^H J_h^T$，并将两侧固定为对角矩阵。若首先考虑一个二阶实置换缩放矩阵，则有：

$$P_\pi D(K_{11}, K_{22}, K_{33}) \Lambda(\sigma_{11}, \sigma_{22}, \sigma_{33}) D(K_{11}, K_{22}, K_{33}) P_\sigma = Q \Sigma Q^T \tag{D.15}$$

由于 $P_\sigma = P_\pi^T$，我们还得到 $PD\Lambda DP^T = (PD)\Lambda(DP^T) = (DP)\Lambda(P^T D) = D(P\Lambda P^T)D$。所得矩阵是一个相似变换，它按 K_{ii} 的值缩放原始噪声谱密度矩阵中的每个元素。在我们的设定中，考虑复值谱密度矩阵时，这一关系仍然成立，因为只需将 a 替换为 $a+bi$，使得 Λ 的每个元素都按 $(a+pi)(a+pi) = a^2+p^2 = A$，$(b+mi)(b+mi) = b^2+m^2 = B$ 进行缩放，同理 C：

$$\begin{bmatrix} a+pi & 0 & 0 \\ 0 & 0 & c+qi \\ 0 & b+mi & 0 \end{bmatrix} \begin{bmatrix} \sigma_{11}^2 & 0 & 0 \\ 0 & \sigma_{22}^2 & 0 \\ 0 & 0 & \sigma_{33}^2 \end{bmatrix} \begin{bmatrix} a-pi & 0 & 0 \\ 0 & 0 & b-mi \\ 0 & c-qi & 0 \end{bmatrix} \tag{D.16}$$

$$= \begin{bmatrix} A\sigma_{11}^2 & 0 & 0 \\ 0 & C\sigma_{33}^2 & 0 \\ 0 & 0 & B\sigma_{22}^2 \end{bmatrix} \tag{D.17}$$

我们知道，J_h 是一个置换缩放矩阵，满足 $\Lambda = J_h \Lambda J_h^T$。这意味着雅可比矩阵无法改变主对角线上的值。利用置换缩放矩阵会放大主对角线元素并重新排列各元素。设置换缩放雅可比矩阵的元素为 A，B，且 C。由此可得 $A^2 = B^2 = C^2 = 1$。这就实现了在置换和符号翻转意义下的分量可辨识性。置换源于对每个过程的随机标记，例如，将 K_{11} 设为 K_{22}；这不会改变矩阵的对角形式，只会对过程重新编号。

D.3 证明命题 D.2

作为一般映射下可识别性研究中最有趣的结果之一，f我们证明，我们的定理将自回归模型涵盖为一个特例，但对f的条件要求较宽松。我们采用了 [Yao 等人，2022a] 中的符号说明，并回顾 f、g 以及潜变量$z_t = \{z_t^i\}_{i \in p}$的一些重要符号：

$$x_t = g(z_t); z_t^i = f_i(\{z_{j,t-\tau} | z_{j,t-\tau} \in \mathbf{Pa}(z_{it})\}, \epsilon_t^i, \theta_r^c, \theta_r^o) \tag{D.18}$$

$_p$其中，f 是一般映射，$g = \sum B_\tau Z_{t-\tau} + E_t$ 是一个 τ 阶自回归因果模型（即，$_{\tau=1}$对于每个 $t-\tau$ 时间点均满足条件独立性）。由于它是一个 VAR（p）模型，识别潜在模型无需收敛。

为了将我们的工作与足够多关于因果表征学习和时滞因果模型的先前研究联系起来，我们提出了一个与标准自回归时滞因果模型相对应的类比模型：时变时间序列和滤波系统。我们将连续时间序列 X_t 视为输入滤波器 $A(L)$ 的源信号，其中滤波器是实值矩阵或矩阵值函数，用以生成新的信号序列：

$$Y_t = A(L)X_t + B_t \tag{D.19}$$

$_T$它可以更明确地展开为 $Y_t = \sum A_k X_{t-k} + B_t$。$A_k$ 是一个取值可能随$_{k=1}$时间变化的矩阵，而 B_t 是另一个随机噪声过程。在许多应用中，B_t 表示布朗运动或平稳的 $I(0)$ 泊松过程。这些过程常见于随机微分方程系统中，用以描述物理对象（如粒子或分子）的随机运动。另一个更特殊的情形是取 $B_t :=$ 伯努利分布(p, q)，从而将 Y_t 转变为伴有稀疏化算子的整数值自回归过程（INAR）$A \circ Y$ [Weiß, 2008]。滤波矩阵与时间相互独立时，可得到自回归模型的简化形式。

我们的理论支持具有任意固定时间滞后的向量自回归模型：首先识别向量自回归(∞)，然后令$\tau_{\geq p} = 0$。由于证明的第二步导出了一个 GL 表示，因此核中的 0 总会映射到原始的 0，从而完成证明。

E 详细 MUTATE 配置

E.1 神经功率谱密度

从本节开始，我们假设向量嵌入已由编码器得到。回顾一下，神经功率谱密度需要独立噪声嵌入来应用群作用 G。因此，我们现在关注非高斯性在整个时间过程中不成立的情况。当噪声服从高斯分布时，$\kappa_d(R_t) = 0$ 对所有 $d \geq 3$都成立，这会使任意簇 $\kappa_d(G, R_t)$ 映射为零张量 $\mathbf{0} \in \mathbb{R}^{p \times d}$。分解的解有无穷多个，因此 G 无法在列缩放和置换意义下恢复，潜在转移图 $\mathcal{G}$也无法恢复。我们认为，仅保留阶数为 d 的 $d \leq 2$ 阶累积量，是实现识别的最小构件。$\kappa_2(R_t)$ 是一个二阶谱矩阵，这正是神经功率谱密度模块所采用的方法，如 推论 4.1所述。

推论 4.1 (通过威尔逊谱因子分解实现的可辨识性)。令 $\mathcal{S}^{(3)} \in \mathbb{C}^d \otimes \mathbb{C}^d \otimes \mathbb{C}^T$ 表示沿频率索引堆叠二阶功率谱密度（*PSD*）矩阵 $\{S_{\epsilon_s}\}_{s=1}^T$ 所得到的三阶张量。假设对于每个频率 s，*PSD* 切片均存在 *Wilson* 谱因子分解

$$\mathcal{S}^{(3)}_{:,:,s} = \tilde{G}_s \tilde{G}_s^H,$$

其中 $\tilde{G}_s = G\Sigma_R^{1/2}$，且矩阵值函数 $G(\omega)$ 在频域中是解析的外函数（最小相位）。

则谱因子 G 在酉常数因子意义下唯一可识别，因此其各列在置换和缩放意义下可识别。

证明。 我们通过矩阵值函数群的作用来表述可辨识性问题，而非通过代数理想来表述。

对于每个频率索引 s，二阶谱切片

$$\mathcal{S}^{(3)}_{:,:,s} = S_{\epsilon_s}$$

存在因式分解

$$S_{\epsilon_s} = H_s \Sigma H_s^H,$$

这并非唯一　　由于在右乘全通矩阵值函数下具有不变性

$$H_s \mapsto H_s U_s, \qquad U_s^H U_s = I.$$

因此，仅凭二阶谱只能确定 H_s，且只确定到由该酉规范群诱导的一个等价类。

Wilson 谱因子分解通过限制 $H(\omega)$ 为解析且外函数（最小相位），解决了这种不可辨识性。在此约束下，谱因子唯一，但相差一个与 ω 无关的常值酉矩阵。因此，歧义从频率依赖函数群缩减为有限维的常值酉群。

沿频率维度堆叠功率谱密度矩阵，

$$\mathcal{S}^{(3)} = \mathrm{con}(S_{\epsilon_1}, \ldots, S_{\epsilon_T}),$$

不会在各切片之间引入新的代数关系，而是确保同一谱因子 $H(\omega)$ 在所有频率上保持一致。因此，残余歧义必须在所有切片之间共享，因而不能依赖于 s。

这种共享的酉歧义通过对因子 $G = H\Sigma^{1/2}$ 进行列置换和缩放而发挥作用。因此，G 的列可识别到置换和缩放。

最后，由于 Wilson 因式分解仅取决于二阶谱结构，因此该构造不受噪声分布 ϵ_t 的影响。一旦 FH 在置换和缩放意义下固定，诱导的高阶累积量作用便使相关等价类保持不变。

□

E.2 时间自适应模块的先验分解

不失一般性，我们考虑潜在随机生成过程的非有限步情形，如引理 2中所述，其中 $\Delta t \to 0$。这导出一种等价关系：内在历史——滤过$\mathcal{F}_t := \sigma\left(\bigcup_{0<t<T} \sigma(Z_t^{\Delta})\right)$——确保过程 $Z_t^{(\Delta)}$ 是 $\mathcal{F}_t$-自适应且可测的。

我们将 ELBO 目标函数分解如下：

$$\begin{aligned}
\mathrm{ELBO} &= \log p(O) - D_{\mathrm{KL}}(q_\phi(Z|O) \| p(Z)) \\
&= \mathbb{E}_{z\sim q(Z_t|O_t)}[\log p(O_t|Z_t)] + \mathbb{E}_{z\sim q(Z_t|O_t)}\left[\log \frac{q(Z_t|O_t)}{p(Z_t)}\right] \\
&= \mathbb{E}_{z\sim q(Z_t|O_t)}[\log p(O_t|Z_t)] - \mathbb{E}_{z\sim q(Z_t|O_t)}\left[\log q(Z_t|O_t) - \log p(Z_t)\right] \\
&= \mathbb{E}_{z\sim q(Z_t|O_t)}\left[\log p(O_t|Z_t) - \log q(Z_t|O_t)\right] + \mathbb{E}_{z\sim q(Z_t|O_t)}\left[\log p(Z_t)\right] \\
&= \mathbb{E}_{z\sim q(Z_t|O_t)}\left[\log p(O_t|Z_t) - \log q(Z_t|O_t)\right] + \mathbb{E}_{z\sim q(Z_t|O_t)}\left[\sum_{\mathcal{F}_0^+}^{\mathcal{F}_T} \log p(Z_t^{(\Delta)}|\mathcal{F}_t)\right]
\end{aligned}$$

我们能够对最后一项中的递增滤过进行分段，原因在于 $\mathcal{F}_t$-可测序列具有良好的性质。我们可以证明，$Z_t|Z_s, R_t$ 和 $Z_t|R_s$ 的滤过相等，因为众所周知，任何具有平稳增量的 p 阶 INAR 序列都存在移动平均（MA）表示。˜进一步构造它们的滤过 $\mathcal{F}_t$（分别为 $.R_{s<t}$）以及 $\mathcal{F}_t$（分别为 $.Z_{s<t}, R_t$）可以证明

$$\tilde{\mathcal{F}}_t = \mathcal{F}_t$$

˜我们将在下文证明该结果。对于 $\mathcal{F}_t$, Z_t 是关于 $s < t$ 的可测函数。根据卷积核 $\Psi = (I - \Phi)^{-1}$ 的因果性，满足 $\Psi_\tau = 0$ （对于 $\tau < 0$），这表明 $Z_t \in \sigma(R_s\colon s < t)$。然后，

˜ ˜ ˜我们构造另一个滤过 $\mathcal{F}_s\colon \sigma(R_u\colon u < s)$。由适应性可知，$\mathcal{F}_s\colon \sigma(R_u\colon u < s) \subseteq \mathcal{F}_t\colon \sigma(R_u\colon u < t)$。因此，$Z_s$ 也是 $\sigma(R_u : u < t)$-可测的。由于原始 $\mathcal{F}_t$ 可测函数的最小 σ-代数必包含于其 σ-代数中，因此我们有 $\sigma(Z_s) \subseteq \sigma(R_u\colon u < t)$，且 $\sigma(\bigcup \sigma(Z_s)) \subseteq \sigma(R_u\colon u < t)$。$s\le t$ 对于 $\mathcal{F}_t$, $R_t = Z_t - \Psi \star Z_t$，因此 R_t 是 $\sigma\left(Z_s\colon s \le t\right)$ 可测的。因此，通过类似的构造，显然˜ ˜ ˜，$\sigma\left(\bigcup \sigma\left(R_s\right)\right) \subseteq \sigma\left(Z_s\colon s \le t\right)$。因此，由于 $\mathcal{F}_t \subseteq \mathcal{F}_t$ 和 $\mathcal{F}_t \subseteq \mathcal{F}_t$，必定有 $\mathcal{F}_t = \mathcal{F}_t$。$s<t$

按照这一设定，先验变为：

$$Z_t \mid \mathcal{F}_t \sim \mathcal{N}\left(\begin{bmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ u_p(t) \end{bmatrix} \sum_{t'<t} (I-\Phi)^{-1},\ \sum_{t'<t} (I-\Phi)^{-1} \Sigma_{t'} (I-\Phi)^{-T}\right)$$

潜变量由 $Z_t = (I-\Phi) \star R_t$生成，其中 R_t 被建模为均值为U、方差为 Σ的各向同性高斯噪声。注意，对于任意 $t-t' \neq 0$，方差矩阵 Σ_{Z_t} 均为零。根据维纳－辛钦$_{N-1}$定理 [维纳， 1949], 我们得到协方差矩阵 $C_{Z_t}(0) = \frac{1}{N} \sum S_z(w_k)$。当上下文明确时，我们省略下标 $Z_t{}_{k=0}$。现在可以将卷积先验分解推导为

$$\begin{aligned}
&\mathbb{E}_{z\sim q(Z_t|O_t)} \left\{ \sum_{\mathcal{F}_0^+, Z_t}^{\mathcal{F}_T} \log p(Z_t^{(\Delta)} | \mathcal{F}_t) \right\} \\
&= \mathbb{E}_{z\sim q(Z_t|O_t)} \left\{ \sum_{\mathcal{F}_0^+, Z_t}^{\mathcal{F}_T} \log p \left[(I-\Phi_t) \star \hat{R}_t^{(\Delta)} \right] \right\} = \mathbb{E}_{z\sim q(Z_t|O_t)} \left\{ \sum_{\mathcal{F}_0^+, Z_t}^{\mathcal{F}_T} \log p \left[\int_0^t (I-\Phi_{t-t'}) \hat{R}_{t'}^{(\Delta)}\ dt \right] \right\} \\
&= \mathbb{E}_{z\sim q(Z_t|O_t)} \left\{ \sum_{\mathcal{F}_0^+, Z_t}^{\mathcal{F}_T} \log p \left[\mathcal{N}(\hat{U}_R, \sum H_t^{-1} \Sigma_{\hat{R}_t'} H_t^{-T}) \right] \right\} \\
&= \mathbb{E}_{z\sim q(Z_t|O_t)} \left\{ \sum_{\mathcal{F}_0^+, Z_t}^{\mathcal{F}_T} \log p \left[\mathcal{N}(\hat{U}_R, \underbrace{\sum H_t^{-1} (PSD_{\hat{Z}_t}) \Sigma_{\hat{R}_t'} H_t^{-T} (PSD_{\hat{Z}_t})}_{C_{p(Z_t)}(0)}) \right] \right\} \\
&= \mathbb{E}_{z\sim q(Z_t|O_t)} \left\{ \sum_{\mathcal{F}_0^+, Z_t}^{\mathcal{F}_T} \log p \left[\mathcal{N}(\hat{U} \sum_{t'<t} \underbrace{(I-\Phi(t-t'))^{-1}}_{(1-\Phi)(w_k)^{-1} \Sigma (1-\Phi)(w_k)^{-H} = S_Z(w_k)}, \frac{1}{N} \sum_{k=0}^{N-1} S_{Z_t}(w_k)) \right] \right\} \qquad \text{(E.1)} \\
&= \mathbb{E}_{z\sim q(Z_t|O_t)} \left\{ \sum_{\mathcal{F}_0^+, Z_t}^{\mathcal{F}_T} \log p \left[\mathcal{N}(\hat{U} \underbrace{\sum_{\tau>0, w=0} (I-\Phi(\tau))^{-1} e^{-jw\tau}}_{\text{the inverse Fourier at } \quad w=0}, \frac{1}{N} \sum_{k=0}^{N-1} S_{Z_t}(w_k)) \right] \right\} \\
&= \mathbb{E}_{z\sim q(Z_t|O_t)} \left\{ \sum_{\mathcal{F}_0^+, Z_t, N\in(N_0,T)}^{\mathcal{F}_T} \log p \left[\mathcal{N}(\hat{U} \text{PSD}_{Z_t}(H(0)), \frac{1}{N} \sum_{k=0}^{N-1} S_{Z_t}(w_k)) \right] \right\} \qquad \text{(E.2)}
\end{aligned}$$

E.3 卷积先验的显式控制

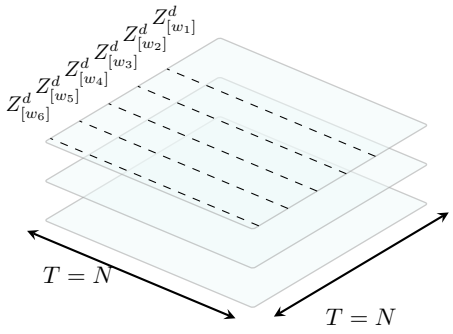


图 E.1：可视化时间自适应功率谱密度计算

备注。 只要时间离散化足够密集，核函数乘积与积分噪声变量的求和就能保证在 $\mathcal{F}_t$ 下收敛到真实的时间自适应过程。潜变量 Z_t^{Δ}从编码器分布 q_ϕ 中采样，并传入功率谱密度分解模块，以计算完整核矩阵 $F_w[1-\Phi_t]$ 的频域表示和功率谱密度 $S_{\hat{R}_t}$。我们进一步指出，关键步骤——谱分解——是在整个编码轨迹 $Z_{t_0:\,T}$ 上完成的，而先验结构则通过滤过分段来确保。这与先前工作的主要区别在于，后者会为每个潜变量递归地构造一个等长滑动窗口。滤过分段可以使用因果掩码，表达能力更强的编码器能够利用这一点。请注意，对于较短的序列，即 $T < 100$，Transformer 模块并非必需组件。然而，对于常规随机点过程类别中的极长序列，可以使用 Transformer 替代标准的多层感知机编码器，通过利用对任意长过去事件的滤过注意力，学习表达能力更强的潜嵌入。

总体训练损失 为鼓励传递核函数的稀疏性，我们遵循广泛使用的惩罚项以联合优化：

$$\mathcal{L}_{Total} = \mathcal{L} - \lambda|G| - \omega\mathcal{L}_{\text{PSD}} \tag{E.3}$$

该训练目标确保所学潜在过程由一族广义白噪声过程驱动。在编码器-功率谱密度流程中，分解由预先规定的各向同性噪声强制实现，从而省略了任何判别器模块，如 Yao 等人 [2022b] 中所使用的那样。稀疏性损失和功率谱密度精度的系数被视为可调超参数。

E.4 与学习潜在因果变量的方法的比较

在本文中，可识别性保证对任何一般映射 f 均成立。具体而言，我们的框架允许多个潜在过程 $Z^{i,}_{\ j\,k\ ,}$ 映射到较少数量的观测变量$O^{s,}_{\ t}{}^{m}$，也允许单个潜在过程 Z_t^i 映射到多个观测数据 O_t。我们旨在将我们的模型与那些专注于恢复潜在因果变量的方法进行比较。因此，我们在仿真中也使用可逆混合。

E.5 模拟方案

我们展示了 INAR 等价类的生成过程。为了与主要针对逐步条件独立性的基线方法进行公平比较，我们将核宽度设为非常短的值 $\tau \in (0.001, 0.01)$，对应于 $t - t'$，从而生成时间步和更密集动态的数据。我们

根据三种基本核响应函数生成随机点过程：

$$\phi_{\text{exponential}}(t) = \alpha e^{-\beta t'}, \quad \alpha \sim \text{均匀分布}(0.1, 0.5) \text{ 和 } \beta \sim \text{均匀分布}[0.5, 2) \phi_{\text{powerlaw}}(t) = \frac{\alpha}{(t+c)^{\beta}} \cdot \mathbf{1}, \quad \alpha \sim \text{均匀分布}(0.5, 1.2) \ \beta \sim \text{均匀分布}[0.1, 0.8) \text{ 和 } \gamma \in \text{均匀分布}(1, 3, 1.8)$$
$$\phi_{\text{rectangular}}(t) = \frac{1}{T - T'} \cdot \mathbf{1}_{\{t' \le T\}}$$

基线强度 u_0 从均匀分布(0, 1, 0.2)中采样。基本核的所有参数均通过均匀采样确定，确保 $\alpha < \beta$ 采用指数响应， $\alpha < \gamma$ 采用幂律响应，从而满足平稳增量条件，使得 $|\phi| < 1$。在模拟中，我们还考虑了简单非线性强度和非参数强度的两种极端情况。我们通过线性变换混合潜在特征，再经过非线性激活函数，构造条件强度函数。具体而言，我们首先使用以下表达式计算对数线性强度

$$\lambda_t = \log(1 + \exp(z_\ell - r_\ell[:, \Delta, :]))$$

该函数通过平滑的 ReLU（即 Softplus）确保输出为正，并控制输出的尺度。在另一种设置（核 == "np"）中，我们使用一个小型神经网络（多层感知机）学习强度函数：一个带 ReLU 激活的两层感知机，后接 Softplus 激活，以确保输出为正。这种设置能够灵活地、以数据驱动的方式对强度动态进行建模，而不局限于纯加性或线性形式。我们使用一个带 ReLU 和 Softplus 激活的两层前馈神经网络来定义混合强度函数。具体而言，其架构如下：

$$\lambda_t = \sigma_+ \left(W_2 \cdot \text{ReLU}(W_1 \lambda_t(l) + b_1) + b_2 \right), \tag{E.4}$$

其中

- $\lambda_t(l) \in \mathbb{R}^d$ 是时间 t 的输入线性基础强度，
- $W_1 \in \mathbb{R}^{64 \times d}$， $b_1 \in \mathbb{R}^{64}$ 是第一层的权重和偏置，
- $W_2 \in \mathbb{R}^{d \times 64}$， $b_2 \in \mathbb{R}^d$ 是第二层的权重和偏置，
- $\sigma_+(x) := \log(1 + e^x)$ 表示 Softplus 激活函数。

这种设计确保输出 λ_t 始终严格为正，并能在保持数值稳定性的同时对潜在动力学中的复杂依赖关系进行建模。

我们通过一个多层混合网络对潜变量 $Z_t \in \mathbb{R}^d$ 到观测空间的变换进行建模。具体而言，对于每一层 $l = 1, \ldots, L-1$，变换表示为 $Z_t^{(l)} = \mathbf{A}^{(l)} \cdot \sigma_{\text{leaky}}(Z_t^{(l-1)})$，其中 $\mathbf{A}^{(l)} \in \mathbb{R}^{d \times d}$ 是一个正交混合矩阵，σ_{leaky} 表示斜率为 $\alpha = 0.2$的带泄漏的ReLU激活函数。初始输入为 $Z_t^{(0)} = Z_t$，最终输出 $Z_t^{(L-1)}$ 表示观测空间信号。

F 详细实验结果

G 扩展讨论

G.1 与三个潜在动态过程的联系

可识别性保证建立在有限维分布的适当弱收敛之上，这反映了每对过程之间的双向路径。我们通过G.1所示的耦合与退化图将它们联系起来。

表2：报告各基线的最佳性能

方法	指标	核平均	Exp	幂	整流	非线性	非参数
TDRL	MCC	0.657	0.629	0.653	**0.773**	0.584	**0.644**
	$\mathcal{L}_{vae}$	**0.449**	**0.308**	**0.302**	0.302	0.871	**0.461**
BetaVAE	MCC	0.419	0.395	0.414	0.420	0.433	0.433
	$\mathcal{L}_{vae}$	9.480	8.538	7.533	8.424	11.683	11.220
SlowVAE	MCC	0.410	0.384	0.405	0.420	0.425	0.412
	$\mathcal{L}_{vae}$	362.890	395.107	448.105	452.472	238.520	280.247
PCL	MCC	0.440	0.469	0.379	0.430	**0.474**	0.449
	$\mathcal{L}_{vae}$(训练)	0.693	0.693	0.694	0.693	**0.693**	0.693
MUTATE	MCC	**0.811**	**0.922**	**0.784**	**0.964**	**0.885**	0.501
	$\mathcal{L}_{vae}$	0.670	0.448	0.508	**0.253**	0.942	1.201

G.2 不同层次的可识别性

尽管因果表征学习常被认为已通过可识别性保证得到解决，但当前文献在很大程度上忽视了对可识别性本身的深入理解。本文强调不同可识别性目标之间的区别，它们各自面临独特的挑战，并指出可将可识别性归纳为三条主要研究路线。

识别 $\mathcal{G}$。从观测数据中恢复因果结构，通常被认为是因果表征学习最根本的目标——这一任务正是该领域得名的由来。识别 $\mathcal{G}$ 至关重要，因为了解因果结构往往足以揭示潜在机制，尤其有助于预测干预后分布。给定一组因果变量的表示，如果相关的条件独立约束能够被唯一确定，那么因果结构就得以完全恢复。从建模角度看，编码器输出的是随机变量 Z的分布，而它可能与真实因果变量 Z并不完全一致。尽管如此，这种可能存在"未对齐"的表示仍会诱导出有效的因果结构，从而为分解观测分布提供一种有原则的方法。这一研究方向通常称为潜在因果结构学习 [Jiangand Aragam，2023年，Jinand Syrgkanis，2023年，Zhang 等人， 2023]。

识别潜在 Z。这一目标比单纯恢复潜在结构 $\mathcal{G}$更具挑战性，因为它要求每个因果变量都能实现精确的分量级对应。一般而言，识别完整潜变量所需的假设往往更强，也不太现实。为恢复 Z，我们假设未知混合函数 f 无噪声且为微分同胚。利用整个分布的信息，我们能够保证估计的因果变量与真实因果变量精确对齐，从而确保构造映射的雅可比矩阵 $f^{-1} \circ \hat{} f$ 具有稀疏的、类似置换的形式。大多数先前研究都遵循这一思路；我们仅列举其中几项代表性工作 [Yao 等人，2022b，Zhang 等人，2024年，Song 等人，2023年，Hyvarinenand Morioka， 2016]

识别完整参数和混合 f。最后，我们通过将参数层级识别（这是最具挑战性的一项目标）与上述两个目标进行比较，结束本节。请注意，在某些条件有所放宽的情况下，仅识别$\mathcal{G}$ 或 Z可能与完整参数识别重叠，因为如果识别出所有因果参数，就必须得到 $\mathcal{G}$ 。如果混合是可逆的或单射的，只需从观测中恢复混合函数 f，便可轻松恢复潜变量。然而，当潜在因果模型和混合函数都不具有任何线性结构时，恢复参数就意味着恢复整个生成模型；这是一项要求极高的任务，因为参数空间处于任意大的环境空间中，却没有任何闭式形式。

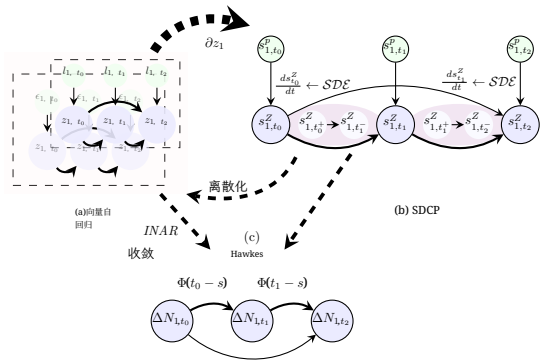


图 G.1：连接三种因果过程：因果过程的演变与退化。(a) 允许存在时间延迟因果影响的经典自回归模型。(b) 以随机微分动力学为特征的因果过程。(c) Hawkes过程，一种特殊的自激过程。NE

.

G.3 非参数强度讨论

为完善可识别性理论，我们通过讨论一种具有非线性、灵活条件强度的随机过程，补充定理2和定理3。请注意，对于实际应用，定理 2 和 命题 D.2 已足够。按照我们的推理，我们可以将非线性强度过程重新表述如下：

$$O_t = f(Z_t^{\Delta}) \tag{G.1}$$

$$Z_t^{\Delta} = \psi(\lambda_t) + \epsilon_t \tag{G.2}$$

有人可能会认为，这与 Song 等人[2023], Yao 等人 [2022b,a] , 中主要讨论的非线性时间序列过程相似，这样的模型 $z_t = f(\text{Pa}(z_t)^{z_{t-1,j}}) + \epsilon_t$。一种猜想是，线性强度的非线性混合可能会也可能不会盖过截至当前序列急剧增加的影响 $\mathcal{F}_t$ 。此外，强度可以定义在任意稠密的区间上，这反映了核效应。因此，对于较长的 $\mathcal{F}_t$ - 预测过程，我们无法保证其严格可识别性。使用谱方法或许是未来研究的一个有益方向。