arXiv:2601.04456v1 [cs.AI] 8 Jan 2026

范畴信念传播：通过下降与整体性进行层论推断

Enrique ter Horst* Sridhar Mahadevan† Juan Diego Zambrano‡

2026年1月9日

摘要

我们为因子图上的信念传播建立了范畴论基础。我们在带类型签名上构造自由超图范畴 $\mathbf{Syn}_\Sigma$，并证明其泛性质，从而通过到矩阵范畴 $\mathbf{Mat}_R$的唯一函子获得组合语义。我们使用极化因子图上的格罗滕迪克纤维化 $\int \mathrm{Msg} \to \mathbf{FG}_\Sigma$ 来表述消息传递，其中按调度索引的自同态定义了置信传播更新。我们将精确推断刻画为有效下降：当重叠部分满足相容条件时，局部信念构成下降数据。该框架通过层论障碍，将树精确性、连接树算法和有环置信传播的失效统一起来。我们提出 HATCC（全纯感知树编译），这是一种通过计算因子神经上的整体性来检测下降障碍、将非平凡整体性编译为模式变量，并在增广图上归约为树形置信传播的算法。对于$O(n^2 d_{\max} + c \cdot k_{\max} \cdot \delta_{\max}^3 + n \cdot \delta_{\max}^2)$，其中有$n$ 个因子和c 个基本环，其复杂度为$O(n^2 d_{\max} + c \cdot k_{\max} \cdot \delta_{\max}^3 + n \cdot \delta_{\max}^2)$。实验结果表明，该方法能够实现精确推断，在网格马尔可夫随机场和随机图上的速度显著快于连接树，并能在可满足性实例上进行不可满足性检测。

关键词： 信念传播、范畴概率、有效下降、单纯形方法、概率推断、图模型

1 引言

信念传播（BP）通常被用作图模型的数值推断引擎，但其数学地位存在二分：（i）意图计算的语义对象（配分函数、边缘分布、最大后验估计），以及（ii）实际运行的执行机制（依赖调度的迭代消息更新）。在含环图中，这两个层面会出现分歧：即使目标语义定义良好，BP 也可能无法收敛、收敛到错误的不动点，或表现出振荡和规范依赖行为 [3, 5, 18]。

本文为 BP 建立了范畴论基础，将其**句法**、**语义**和**执行**区分开来，随后引入一种确定性、拓扑感知机制，将环路不一致性编译为推断任务的有限分解。核心技术对象是**整体**：沿基本环传输约束所诱导的、作用于所选“状态纤维”上的作用。整体是一种全局障碍，纯粹的局部消息更新无法察觉它，但可以根据因子图结构明确计算。

问题陈述（因子图上的贝叶斯推断）

设 $G = (V, F, E)$ 为一个（二部）因子图，其中 V 是变量节点的集合，F 是因子节点的集合，而 $E \subseteq V \times F$ 表示关联（边 $(i, f) \in E$ 表示变量 x_i 参与因子 f）。对于每个变量节点 $i \in V$，令 $\mathcal{X}_i$ 表示其（有限）状态空间。全局赋值是如下元组

$$x = (x_i)_{i \in V} \in \mathcal{X}_V := \prod_{i \in V} \mathcal{X}_i.$$

每个因子节点 $f \in F$ 都关联一个非负势函数

$$\psi_f : \mathcal{X}_{\mathrm{scope}(f)} \to \mathbb{R}_{\geq 0}, \qquad \mathrm{scope}(f) \subseteq V,$$

*安第斯大学，管理学院，波哥大，哥伦比亚
†Adobe 研究院和马萨诸塞大学阿默斯特分校
‡https://www.colimits.com

其中，作用域(f)是图中与f相邻的变量集合：作用域$(f) = \{\, i \in V : (i, f) \in E \,\}$，且$\mathcal{X}_{\text{scope}(f)} := \prod_{i \in \text{scope}(f)} \mathcal{X}_i$。赋值$x$的（未归一化）联合权重是局部势函数的乘积，

$$\Psi(x) \;:=\; \prod_{f \in F} \psi_f\big(x_{\text{scope}(f)}\big),$$

其中$x_{\text{scope}(f)}$表示将全局赋值x限制（投影）到由作用域(f)索引的坐标上。

语义推断任务是计算全局消元计算

$$Z \;=\; \sum_{x \in \mathcal{X}_V} \Psi(x), \qquad p(x_i = a) \;=\; \frac{1}{Z} \sum_{x_{V\setminus\{i\}} \in \mathcal{X}_{V\setminus\{i\}}} \Psi(x)\Big|_{x_i = a},$$

对每个变量$i \in V$和每个状态$a \in \mathcal{X}_i$而言。这里Z是配分函数（归一化常数），边缘分布$p(x_i = a)$可通过对除x_i之外的所有变量求和（即消去$x^{V\setminus\{i\}}$）得到。最大乘积/MAP 对应的方法则是将外层求和替换为取最大值。

标准的有环信念传播（置信传播）不会直接执行这些全局消元。相反，它定义了一种执行过程：在G的有向关联上迭代更新消息的局部动力系统（通常是函数$m^{i\to f}\colon \mathcal{X}_i \to \mathbb{R}_{\geq 0}$和$m_{f\to i}\colon \mathcal{X}_i \to \mathbb{R}_{\geq 0}$）。由于在含环图上，执行只是对语义的近似，因此我们关注的失效模式具体且可检验：(i) 不收敛（消息更新发生振荡或发散）；(ii) 收敛到依赖于初始值或调度的不动点；以及 (iii) 收敛后的信念所诱导的边缘分布与预期的全局语义不一致（例如违反全局环约束，或与可处理子实例上的精确消元结果不符）。本文关注非平凡环整体是语义正确性的主要障碍的情形，并利用由此产生的整体性扇区（轨道），将推断分解为按扇区进行的精确计算。

贡献

我们提出五项贡献，每一项都明确体现了各层之间的区分。

1. **用于因子图推断的通用语法。** 我们提出一个由带类型签名 Σ生成的自由且对象逐点自由的超图范畴 $\mathbf{Syn}_\Sigma$，为因子图提供一种组合式图语言。这是一个纯句法 层：图式是程序，而不是数值。

2. **作为半环参数化消元的语义。** 我们在语义目标 $\mathbf{Syn}_\Sigma$ 中求值 $\mathbf{Mat}_R$（交换半环 R上的矩阵），其中复合即消元（有限“积分”），而超图结构则由特殊交换弗罗贝尼乌斯代数（SCFA）编码 [7, 8]。

3. **作为消息纤维化的执行。** 我们将消息配置建模为格罗滕迪克纤维化的纤维，该纤维化位于极化因子图范畴之上。这一形式化描述了（a）消息空间如何随图而变化，以及（b）置信传播（BP）在图同构和重索引下精确的等变性。

4. **规范等变性与单纯结构。** 我们将标准消息重缩放组织为一个群胚作用（“规范”），并通过神经研究由此产生的单纯形／Kan 结构。这使不变性和冗余显而易见，并阐明哪些性质属于语义，哪些属于执行 [18]。

5. **具备全纯性意识的环路推断确定性编译。** 我们提出全纯感知树编译 (HATCC)：一种确定性过程，可根据环基计算全纯生成元，提取所选接口变量上诱导的轨道／商结构，并将推断化简为有限个树 问题的混合（每个扇区内均精确）。由此可得到：(i) 原始置信传播的明确失败证书；(ii)每个扇区内 的精确性保证，以及有原则的重组规则。

与既有工作的定位

置信传播及其变分解释由来已久；我们依据如下刻画：置信传播不动点是 Bethe自由能及相关表述的驻点 [3, 5, 18]。环路推断并非边缘情况：在树之外，精确边缘化通常难以计算（事实上，在广泛类别的图模型中，概率推断是 NP难问题） [13, 23, 24]。因此，消息传递（和积 / 最大乘积）已成为计算机视觉、编码和统计物理中默认的大规模贝叶斯工作工具，而因子图表述则提供了一种统一的计算语言

[2, 6, 25]。然而，在含环图上，标准置信传播是一个局部动力系统，其行为可能很脆弱：它可能无法收敛、收敛到依赖于初始条件的不动点，或产生无法体现预期全局消元语义的信念 [26–29]。大量文献提出了部分补救方法——例如，通过凸化自由能或重加权目标函数实现可证明的收敛 [30, 31]，或事后进行环修正 [32]——但这些方法通常仍是对局部更新所做的算法 层面修改。

消息表示的规范不变性（不会改变归一化信念的重缩放）在图模型和变分推断中也广为人知 [18]。我们的贡献并不是提出新的局部更新规则；而是一个全局、确定性的编译步骤，从因子图中提取拓扑传输数据（环整体），并利用这些数据将环路推断分解为全纯诱导的扇区，在这些扇区上推断可以（经过证明）实现精确计算，或表现得明显更好。从概念上说，我们用“显式计算环路障碍，按其作用取商，并逐扇区求解”取代了“在局部迭代并寄希望于成功”。

全纯机制对于同步和对齐模型（例如 $\mathbb{Z}_k SO(d)$ 同步）尤其自然，因为环一致性是其中的基本全局约束 [19, 20]。它也与提升推断的目标相容——通过利用对称性将推断因式分解——但关注的是由环传输诱导的拓扑对称性，而非纯粹的关系对称性[21, 22]。

范围、局限性，以及我们不 声称的内容

我们主要研究有限域和半环值势函数，以确保语义精确且算法对象有限。对于整体作用在所选接口上诱导出有限轨道分解的模型类，HATCC 能够实现精确推断；但它并非连接树的通用替代方案。我们不声称能够对一般含环模型进行多项式时间精确推断。相反，我们提出：(i) 整体揭示了一种具体且可计算的全局障碍，可解释置信传播的某些特定失败；(ii) 当该障碍具有较小的轨道结构时，推断可确定性地分解为精确子问题。

路线图

下一节将明确记号和严格性约定。随后，我们将依次介绍：(i) 语法层（自由超图范畴），(ii) 语义层（求值至$\mathbf{Mat}_R$），以及 (iii) 执行层（消息纤维化、置信传播自同态和规范）。接着，我们将精确性重新表述为下降条件，最后介绍全纯感知编译（HATCC）及其逐扇区精确性结论。本文最后将介绍实验和消融实验，并讨论局限性与开放问题。

2 背景与约定（精简版）

我们采用离散因子图以及以半环消元为基础的推断视角。较长版本中的范畴论和层论框架在此仅作为组织语义：它指出了应比较哪些对象（一致性对应极限/等化子），以及哪些计算是规范的（拉回/限制和推前/消元）。所有标准的范畴论预备知识（超图范畴、幺半环结构、求值函子）均移至附录A，并附有[1, 7, 9]中的引文。同样，基础的层与下降理论推迟至附录B，相关参考见 [4, 17]。

记号。变量取值于有限集；对于变量集合S，我们记作 $\Omega(S) = \prod_{v \in S} \Omega(v)$。我们用 $\odot$ 表示半环乘法（概率推断中的普通乘法），用 $\oplus$ 表示半环加法（普通加法）。构造全纯核时，复合即相应半环上的矩阵乘法（支撑情形使用布尔半环；加权似然情形使用$(+, \times)$）。

正文中保留的内容。这条新颖的路径——整体 $\to$ 轨道/扇区分解 $\to$ 逐扇区精确性——将在第 5 节和第 6 节中展开。无法直接支撑这些步骤的背景内容，则会引用相关文献或移至附录。

3 信念传播作为消息纤维上的自同态 S

现在定义置信传播更新算子为自同态 $T_G : \mathbf{Msg}(G) \to \mathbf{Msg}(G)$，作用于消息空间。这是用范畴论方式表达的标准和积算法。关键结果是

置信传播作用于消息纤维化的纤维，而T_G的不动点对应于边缘信念。

3.1 因子势与模型

固定一个状态空间赋值 $\Omega:\Lambda\to$ **有限集合范畴**以及一个极化因子图 $G\in\mathbf{FG}_\Sigma$。

定义 3.1（因子势赋值）。一个**因子势赋值** Φ 对于G，为每个因子顶点$f\in F(G)$及其相邻变量 $\mathrm{adj}(f)=\{v_1,\ldots\ldots,v_k\}$指定一个函数：

$$\phi_f:\Omega(\lambda(v_1))\times\cdots\times\Omega(\lambda(v_k))\to R$$

记号：对于f的邻居所对应的状态配置，记作$x_{\mathsf{adj}(f)}=(x^{v_1},\ldots\ldots,x^{v_k})$。

解释：ϕ_f是与因子f关联的局部因子（势函数）。它为其相邻变量的每个联合配置赋予一个R中的权重。

例 3.2（成对马尔可夫随机场势）。对于一个连接二元变量v_1,v_2的成对因子f，其中$\Omega(\lambda(v_i))=\{0,1\}$：

$$\phi_f(0,0)=2.0,\quad \phi_f(0,1)=1.0,\quad \phi_f(1,0)=1.0,\quad \phi_f(1,1)=2.0$$

这会偏好一致性（对角线元素的权重更高）。

3.2 消息索引与邻域

定义 3.3 (消息索引**)**。对于消息构型 $m\in\mathbf{Msg}(G)$ 以及半边 $h\in\mathrm{HalfEdges}(G)$，记作：

$$m_h\in R^{\Omega(\lambda(\mathsf{var}(h)))}$$

表示分配给半边 h 的消息。

对于有向半边：

- 若h的方向为$v\to f$，则将从变量 v 发送到因子 f的消息记为$m_{v\to f}$
- 若h的方向为从$f\to v$，则将从因子f到变量v的消息记为$m_{f\to v}$

定义 3.4（邻域）。对于变量顶点$v\in V(G)$，定义：

$$\mathsf{nbhd}(v)=\{f\in F(G)\mid f\text{ is adjacent to }v\}$$

对于因子顶点$f\in F(G)$，定义：

$$\mathsf{nbhd}(f)=\{v\in V(G)\mid v\text{ is adjacent to }f\}$$

3.3 信念传播更新规则

定义 3.5（变量到因子更新）。对于半边$v\to f$和消息配置$m\in\mathbf{Msg}(G)$，**变量到因子更新**为：

$$(\mathsf{BP}_{v\to f}(m))(x)=\prod_{g\in\mathsf{nbhd}(v)\setminus\{f\}}m_{g\to v}(x)$$

对于每个$x\in\Omega(\lambda(v))$。

解释：变量 v 到因子 f 的消息，是来自其他因子的所有传入消息的乘积，并在每个状态 x处逐点计算。

直觉：变量 v 汇总来自其邻居的所有信息，但不包括 f，然后将汇总后的信念发送给 f。

定义 3.6（因子到变量更新）。对于半边$f\to v$和消息配置 $m\in\mathbf{Msg}(G)$，因子到变量更新**因子到变量更新**为：

$$(\mathsf{BP}_{f\to v}(m))(x_v)=\sum_{x_{\mathsf{nbhd}(f)\setminus\{v\}}}\phi_f(x_{\mathsf{nbhd}(f)})\cdot\prod_{u\in\mathsf{nbhd}(f)\setminus\{v\}}m_{u\to f}(x_u)$$

对于每个 $x_v\in\Omega(\lambda(v))$。

解释：因子 f 向变量 v 发送的消息计算方式如下：

1. 将来自 f 的所有其他邻居的输入消息相乘

2. 乘以因子势 ϕ_f

3. 对除 v 之外的所有变量进行边缘化（求和）

直观而言：因子 f 汇集除 v 之外所有邻居的信息，施加其局部约束 ϕ_f，并将其他所有变量边缘化掉，从而仅传递关于 v 的信念。

示例 3.7（成对马尔可夫随机场的因子更新）。 考虑一个连接变量 v_1,v_2的成对因子 f，其中 $\Omega(\lambda(v_i)) = \{0,1\}$。消息 $f \to v_1$ 为：

$$\begin{aligned}(\mathsf{BP}_{f\to v_1}(m))(x_1) &= \sum_{x_2\in\{0,1\}} \phi_f(x_1,x_2)\cdot m_{v_2\to f}(x_2)\\ &= \phi_f(x_1,0)\cdot m_{v_2\to f}(0) + \phi_f(x_1,1)\cdot m_{v_2\to f}(1)\end{aligned}$$

对于$x_1 = 0$:

$$(\mathsf{BP}_{f\to v_1}(m))(0) = 2.0\cdot m_{v_2\to f}(0) + 1.0\cdot m_{v_2\to f}(1)$$

对于$x_1 = 1$:

$$(\mathsf{BP}_{f\to v_1}(m))(1) = 1.0\cdot m_{v_2\to f}(0) + 2.0\cdot m_{v_2\to f}(1)$$

这正是成对因子的标准置信传播更新。

3.4 信念传播算子

定义 3.8（统一局部更新）。 对于半边 $h \in \mathsf{HalfEdges}(G)$，定义：

$$\mathsf{BP}_h : \mathbf{Msg}(G) \to R^{\Omega(\lambda(\mathsf{var}(h)))}$$

by:

$$\mathsf{BP}_h(m) = \begin{cases}\mathsf{BP}_{v\to f}(m) & \text{if } h \text{ is directed } v\to f\\ \mathsf{BP}_{f\to v}(m) & \text{if } h \text{ is directed } f\to v\end{cases}$$

定义 3.9（并行（同步）置信传播）。 **并行信念传播算子**是自同态 $T_G : \mathbf{Msg}(G) \to \mathbf{Msg}(G)$，定义如下：

$$(T_G(m))_h = \mathsf{BP}_h(m) \quad \text{for all } h \in \mathsf{HalfEdges}(G)$$

I解释：所有消息都基于当前构型同时更新 .

定义 3.10（单边（异步）更新）。 对于半边 $h \in \mathsf{HalfEdges}(G)$，**单边更新** 是$U_h : \mathbf{Msg}(G) \to \mathbf{Msg}(G)$，定义如下：

$$(U_h(m))_{h'} = \begin{cases}\mathsf{BP}_h(m) & \text{if } h' = h\\ m_{h'} & \text{if } h' \neq h\end{cases}$$

解释：仅更新边 h 上的消息，保持所有其他消息不变 d.

3.4.1 置信传播的可视化表示

我们使用带注释的因子图示意图来说明 BP 更新规则和消息流，其中彩色箭头表示消息。

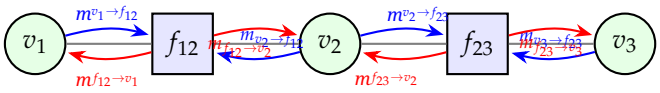
定义 3.11（调度信念传播）。**调度** G 上是 $s = (h_1,\ldots, h_k)$ 中 G 的半边组成的有限序列。

调度信念传播**算子**为：

$$T_G^{(s)} = U_{h_k}\circ\cdots\circ U_{h_2}\circ U_{h_1} : \mathbf{Msg}(G)\to\mathbf{Msg}(G)$$

解释：按照调度依次更新消息，每次更新都使用最新的消息值。

消息空间 $\mathbf{Msg}(G)$ 用于链式马尔可夫随机场



$|E| = 4$关联
$|E^{\rightarrow}| = 8$半边
$\mathbf{Msg}(G) = \prod_{h \in E^{\rightarrow}} R^{\Omega(\lambda(\mathsf{var}(h)))}$

图 1：链式马尔可夫随机场的消息空间 $\mathbf{Msg}(G)$。每条有向半边都携带一条消息（从状态空间到 R 的函数）。蓝色箭头：变量到因子消息。红色箭头：因子到变量消息。信念传播通过迭代更新所有消息来进行。

命题 3.12 (置信传播算子是定义良好的自同态)。 算子 T_G、U_h 和 $T(s)$ G 都是定义良好的函数 $\mathbf{Msg}(G) \to \mathbf{Msg}(G)$。

然而，这些算子并非 R-线性的（不是半模同态）。相反，它们是多线性/多项式。

证明。 **良定义性**：每次更新均由以下操作构成：

- R 中的逐点乘积（消息的乘法 s)
- 对状态空间进行有限求和（边缘化）
- 对因子势求值

所有这些操作在R中都有定义。

多重线性（而非线性）：变量到因子更新为：

$$(\mathsf{BP}_{v \to f}(m))(x) = \prod_{g \in \mathsf{nbhd}(v) \setminus \{f\}} m_{g \to v}(x)$$

这是**多重线性**的（对每条传入消息$m_{g \to v}$分别呈线性），但对完整消息向量m而言**不是线性的**。

示例：对于 2 次（两个邻居）：

$$\mathsf{BP}_{v \to f}(m) = m_{g_1 \to v} \cdot m_{g_2 \to v}$$

这是**双线性的**：当$m_{g_2 \to v}$ 固定时，它关于$m_{g_1 \to v}$ 呈线性，反之亦然。但是：

$$\mathsf{BP}_{v \to f}(m + m') = (m_{g_1} + m'_{g_1}) \cdot (m_{g_2} + m'_{g_2}) \neq \mathsf{BP}_{v \to f}(m) + \mathsf{BP}_{v \to f}(m')$$

类似地，因子到变量更新涉及求和下的乘积，因此关于传入消息是多项式（≥ 2次）的。

结论：T_G 是消息空间上的**多项式映射**，而不是线性映射。它在每个节点处关于传入消息是多重线性的，但这并不能扩展为全局线性。 □ □

备注 3.13（多项式映射与线性映射）。BP 是一个定义良好的**集合论函数**T_G ： $\mathbf{Msg}(G) \to \mathbf{Msg}(G)$。它不是半模同态（不是$R$-线性的），但它是一个**多项式映射**：由多重线性运算（乘积）和线性运算（求和）复合而成。

这一区别很重要：

- **线性**：$f(ax + by) = af(x) + bf(y)$
- **多重线性**：$f(x1, \ldots, x_n)$ 在每个 x_i 变量上分别呈线性
- **多项式**：多重线性运算与线性运算的复合

置信传播属于最后一类：由求和（线性）与乘积（多重线性）复合而成。

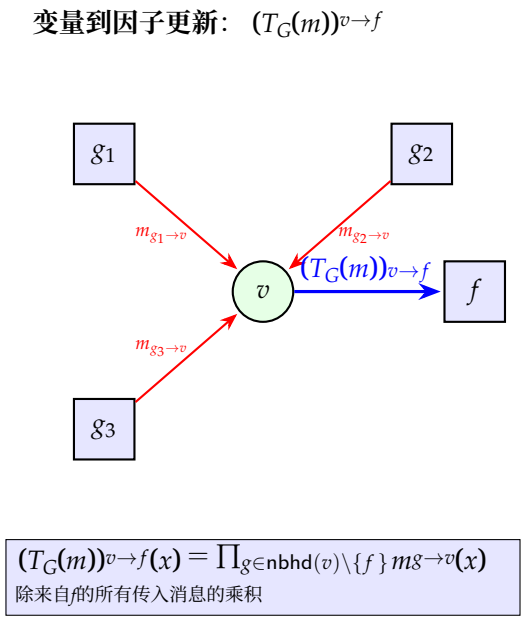


图2：变量到因子的置信传播更新。变量v 将来自所有相邻因子（除f之外）的输入消息（红色箭头）逐点相乘并汇总。该乘积成为传出消息$v \to f$ （蓝色箭头）。

3.5 不动点与信念

定义 3.14 （不动点消息）。 消息配置$m \in \mathbf{Msg}(G)$ 是并行置信传播的**不动点**：

$$T_G(m) = m$$

对于调度置信传播，m 是调度s 的不动点，当且仅当：

$$T_G^{(s)}(m) = m$$

定义 3.15 （边缘信念）。 对于T_G 的一个不动点$m^* \in \mathbf{Msg}(G)$ 以及一个变量顶点v，**v**处的**信念**为：

$$b_v(x) = \prod_{f\in \mathsf{nbhd}(v)} m^*_{f\to v}(x) \quad \text{for each } x \in \Omega(\lambda(v))$$

解释：信念是所有传入消息的乘积——它表示有关变量v的汇总信息。

关键问题：不动点何时存在？它们何时对应于正确的边缘分布？第 3 节回答了树上的这一问题。

4 G规范等变性：BP 如何与消息重标度相互作用 g

在第 3 节中，我们证明了信念具有规范不变性（命题 **??**），但尚未解决 BP 算子本身是否保持规范重标度。本节将确立 BP 与规范之间的基本关系：算子 T_G 在规范传播映射 Θ_G 下具有**半等变性**：$K_G \to K_G$。

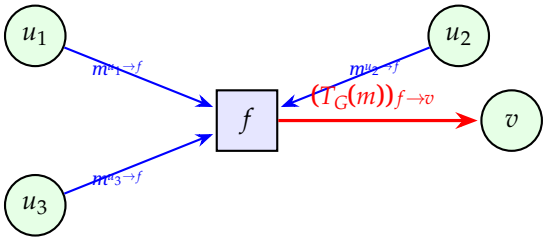
关键结果是：

$$T_G(k \cdot m) = \Theta_G(k) \cdot T_G(m) \tag{1}$$

这意味着：如果我们将输入消息缩放k，输出消息就会缩放为 $\Theta_G(k)$——这是另一种 缩放方式，由规范在置信传播更新规则中的传播方式决定。

我们通过显式计算证明这一点，展示规范因子如何在置信传播公式的乘法和求和运算中传播。

因子到变量更新：$(T_G(m))_{f\to v}$



$(T_G(m))_{f\to v}(x_v) = \sum_{x_{u_1},\ x_{u_2},\ x_{u_3}} \phi_f(x^{u_1},\ x^{u_2},\ x^{u_3},\ x_v) \prod_{u\in\{u_1,u_2,u_3\}} m^{u\to f}(x_u)$
和积：将因子势 ϕ_f 与输入消息相乘，并对 u_1,u_2,u_3进行边缘化

图 3：因子到变量的置信传播更新。因子 f 从除 v之外的所有邻居接收输入消息（蓝色箭头），与其势函数 ϕ_f 相乘，并对除 v之外的所有变量进行边缘化（求和）。结果即为传出消息 $f \to v$ （红色箭头）。

4.1 规范传播映射

定义 4.1（规范传播映射）**：**对于极化因子图 G，通过规定其在每条半边上的作用，定义规范传播映射 Θ_G： $K_G \to K_G$：

对于变量到因子的边 $v \to f$:

$$(\Theta_G(k))_{v\to f} = \prod_{g\in \mathsf{nbhd}(v)\setminus\{f\}} k_{g\to v}$$

对于因子到变量的边 $f \to v$:

$$(\Theta_G(k))_{f\to v} = \prod_{u\in \mathsf{nbhd}(f)\setminus\{v\}} k_{u\to f}$$

解释：当我们用规范 k对输入消息进行重缩放时，BP 算子会生成经 $\Theta_G(k)$ 重缩放的输出消息。每条输出边上的重缩放因子，等于用于计算该输出的各条输入边的重缩放因子的乘积。

例 4.2（在简单边上的规范传播）。 考虑一棵简单的双变量树：

$$v_1 \text{ —— } f \text{ —— } v_2$$

假设我们用规范因子 $k_{v_2\to f} = 2$ 将消息 $m_{v_2\to f}$ 缩放为原来的 0 倍。输出消息 $m_{f\to v_1}$ 会发生什么变化?

根据规范传播公式:

$$(\Theta_G(k))_{f\to v_1} = k_{v_2\to f} = 2.0$$

因此，如果 we 将输入消息加倍，输出消息也会加倍。这是合理的：因子-to-变量 u 更新日期为:

$$m_{f\to v_1}(x_1) = \sum_{x_2} \phi_f(x_1,x_2)\cdot m_{v_2\to f}(x_2)$$

将$m_{v_2\to f}$d 加倍会使整个和加倍。

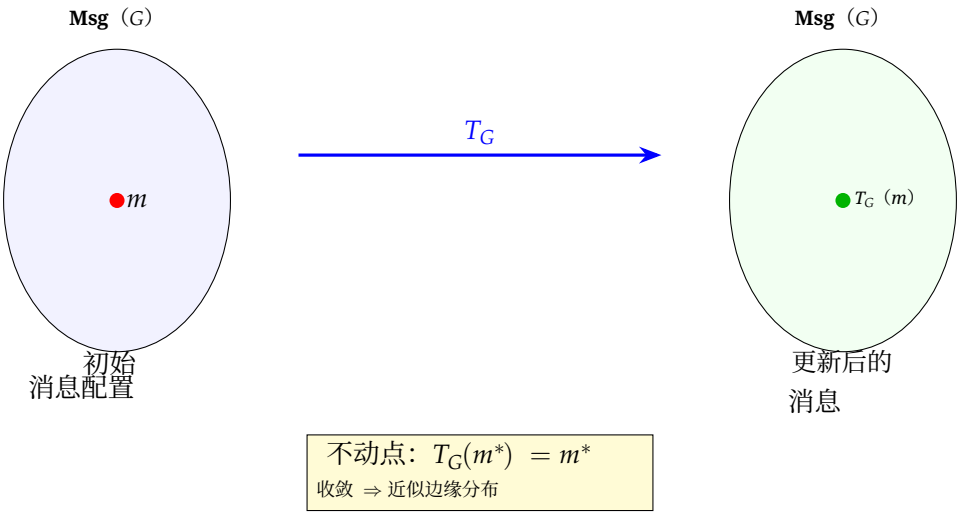


图4：置信传播可视为自同态T_G：$\mathbf{Msg}(G) \to \mathbf{Msg}(G)$。从初始消息$m(0)$开始，迭代$m(t+1) = T_G(m(t))$。不动点$m^* = T_G(m^*)$ 编码了（近似的）边缘信念。在树上，置信传播会收敛到精确边缘分布。

4.2 主定理：置信传播是半等变的

定理 4.3（置信传播在规范变换下的半等变性）。 对于任意极化因子图 G、规范 $k \in K_G$ 以及消息配置 $m \in \mathbf{Msg}(G)$：

$$T_G(k \cdot m) = \Theta_G(k) \cdot T_G(m)$$

此外， $\Theta_G : K_G \to K_G$ 是一个群同态:

1. $\Theta_G(\mathbf{1}) = \mathbf{1}$ *(单位元保持)*
2. $\Theta_G(k \cdot k') = \Theta_G(k) \cdot \Theta_G(k')$ *(multiplicativity)*

证明。 我们通过对置信传播更新公式进行分类讨论来证明等变性。

情形 1：变量到因子更新

设 $h = (v \to f)$。根据定义 3.5:

$$\begin{aligned}
(T_G(k \cdot m))_{v \to f}(x) &= \prod_{g \in \mathsf{nbhd}(v) \setminus \{f\}} (k \cdot m)_{g \to v}(x) \\
&= \prod_{g \in \mathsf{nbhd}(v) \setminus \{f\}} k_{g \to v} \cdot m_{g \to v}(x) \quad \text{(by definition of gauge action)} \\
&= \left(\prod_{g \in \mathsf{nbhd}(v) \setminus \{f\}} k_{g \to v} \right) \cdot \left(\prod_{g \in \mathsf{nbhd}(v) \setminus \{f\}} m_{g \to v}(x) \right) \\
&= (\Theta_G(k))_{v \to f} \cdot (T_G(m))_{v \to f}(x) \quad \text{(by Def. 4.1)} \\
&= (\Theta_G(k) \cdot T_G(m))_{v \to f}(x)
\end{aligned}$$

关键观察：规范因子$k_{g \to v}$与x 无关，因此可以从乘积中提出。这就是规范作用在这个分量上具有乘法性的原因。

情况 2：因子到变量更新

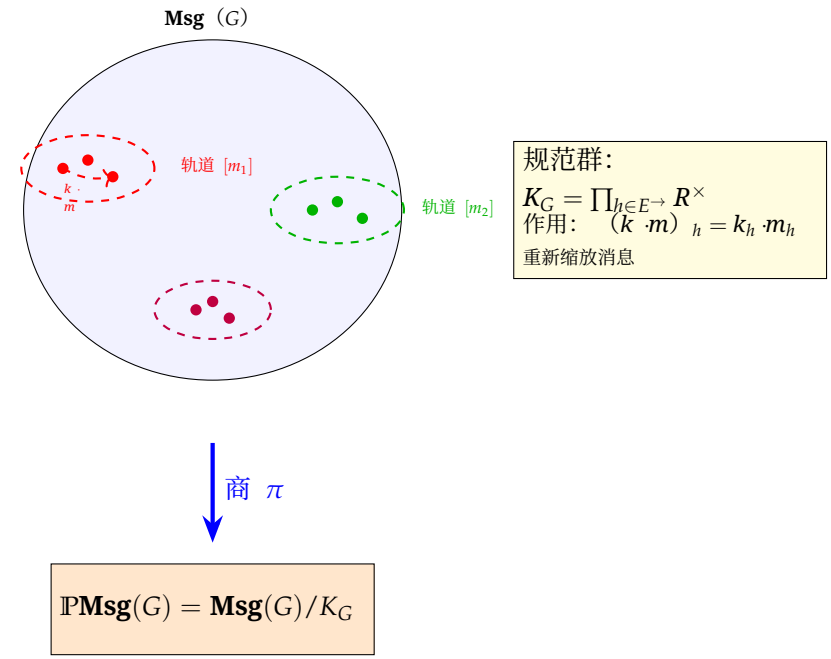


图 5：规范群 K_G 通过逐点重标度作用于消息空间 $\mathbf{Msg}(G)$。每个轨道（虚线椭圆）代表在规范意义下等价的消息。射影消息空间 $\mathbb{P}\mathbf{Msg}(G)$ 是该作用的商，将物理上等价的消息配置视为相同。

令 $h = (f \to v)$。根据定义 3.6：

$$\begin{aligned}(T_G(k\cdot m))_{f\to v}(x_v) &= \sum_{x_{\mathsf{nbhd}(f)\setminus\{v\}}} \phi_f(x_{\mathsf{nbhd}(f)}) \cdot \prod_{u\in \mathsf{nbhd}(f)\setminus\{v\}} (k\cdot m)_{u\to f}(x_u) \\ &= \sum_{x_{\mathsf{nbhd}(f)\setminus\{v\}}} \phi_f(x_{\mathsf{nbhd}(f)}) \cdot \prod_{u\in \mathsf{nbhd}(f)\setminus\{v\}} \left(k_{u\to f}\cdot m_{u\to f}(x_u)\right) \\ &= \sum_{x_{\mathsf{nbhd}(f)\setminus\{v\}}} \phi_f(x_{\mathsf{nbhd}(f)}) \cdot \left(\prod_{u\in \mathsf{nbhd}(f)\setminus\{v\}} k_{u\to f}\right) \cdot \left(\prod_{u\in \mathsf{nbhd}(f)\setminus\{v\}} m_{u\to f}(x_u)\right).\end{aligned}$$

由于每个 $k^{u\to f} \in R^\times$ 都是独立于求和变量的标量，因此可以将其提出求和号：

$$\begin{aligned}(T_G(k\cdot m))_{f\to v}(x_v) &= \left(\prod_{u\in \mathsf{nbhd}(f)\setminus\{v\}} k_{u\to f}\right) \cdot \sum_{x_{\mathsf{nbhd}(f)\setminus\{v\}}} \phi_f(x_{\mathsf{nbhd}(f)}) \cdot \prod_{u\in \mathsf{nbhd}(f)\setminus\{v\}} m_{u\to f}(x_u) \\ &= (\Theta_G(k))_{f\to v} \cdot (T_G(m))_{f\to v}(x_v),\end{aligned}$$

其中，最后一个等式即为定义 4.1。这证明了在所有半边上都有 $T_G(k\cdot m) = \Theta_G(k)\cdot T_G(m)$ **Homomorphism property of** Θ_G。令 $k, k' \in K_G$。对于从变量到因子的半边 $ge v \to f$，

$$(\Theta_G(kk'))_{v\to f} = \prod_{g\in \mathsf{nbhd}(v)\setminus\{f\}} (kk')_{g\to v} = \prod_g (k_{g\to v}k'_{g\to v}) = \left(\prod_g k_{g\to v}\right)\left(\prod_g k'_{g\to v}\right) = (\Theta_G(k))_{v\to f}(\Theta_G(k'))_{v\to f}.$$

因子到变量情形 $f \to v$ 完全相同，只需相应地替换邻域。因此 $\Theta_G(kk') = \Theta_G(k)\Theta_G(k')$，且显然 $\Theta_G(\mathbf{1}) = \mathbf{1}$。 □ □

推论 4.4（置信传播下降到射影消息）。若 $m' \sim m$ 属于 $\mathbb{P}\mathbf{Msg}(G) = \mathbf{Msg}(G)/K_G$，即 $m' = k\cdot m$，其中某个 $k \in K_G$，则 $T_G(m') \sim T_G(m)$。因此 $T_G([m]) := [T_G(m)]$ 是良定义的。

证明。若 $m' = k \cdot m$，则根据定理 4.3，$T_G(m') = T_G(k \cdot m) = \Theta_G(k) \cdot T_G(m)$，因此 $T_G(m')$ 与 $T_G(m)$ 位于同一 K_G-轨道。 □

4.2.1 规范对称性的可视化表示

我们通过交换图和带注释的因子图来说明规范群作用、传播映射 Θ_G 和半等变性。

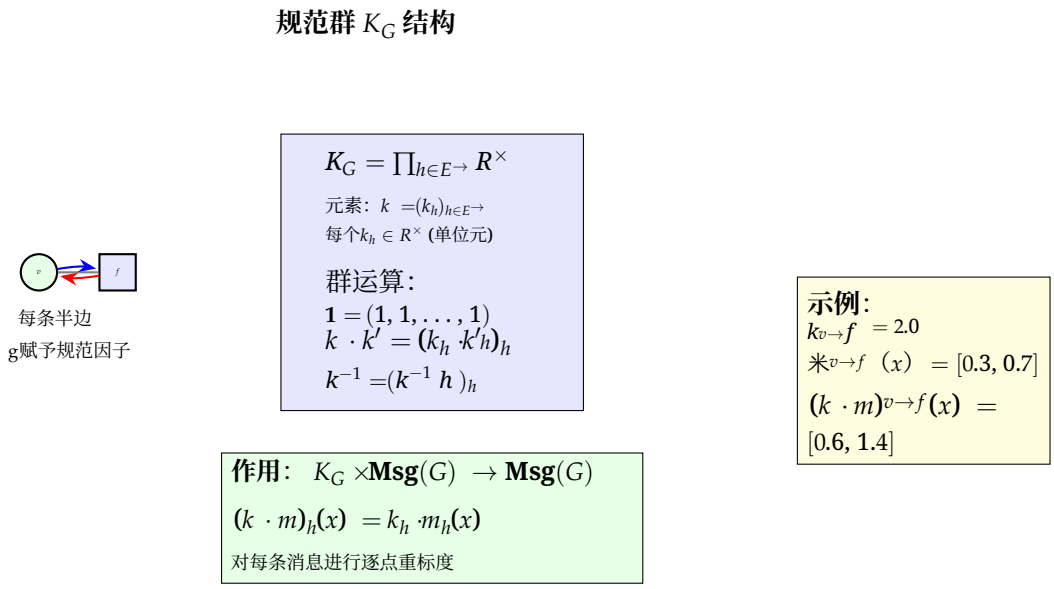


图6：规范群 K_G 结构：整体半边上的单位元 $R^{\times}$ 的乘积。作用按分量对消息进行重标度。这是一个保持消息空间结构的自由阿贝尔群作用。

4.3 演算示例：链上的规范传播

示例 4.5 （三变量链上的规范重标度）。 考虑如下链：

$$v_1 \leftrightarrow f_{12} \leftrightarrow v_2 \leftrightarrow f_{23} \leftrightarrow v_3$$

假设我们用规范 $k \in K_G$ 对消息进行重标度，其中：

$$k_{f_{12} \to v_1} = 2, \quad k_{v_1 \to f_{12}} = 3, \quad k_{f_{12} \to v_2} = 5, \quad \text{etc.}$$

$\Theta_G(k)$ 是什么?

在边 $v_2 \to f_{12}$:

$$(\Theta_G(k))_{v_2 \to f_{12}} = \prod_{g \in \mathsf{nbhd}(v_2) \setminus \{f_{12}\}} k_{g \to v_2} = k_{f_{23} \to v_2}$$

除了 f_{12}, ，只有 f_{23} 是 v_2 的邻居，因此 $v_2 \to f_{12}$ 上的规范由来自 f_{23} 的输入规范决定。

在边 $f_{12} \to v_2$:

$$(\Theta_G(k))_{f_{12} \to v_2} = \prod_{u \in \mathsf{nbhd}(f_{12}) \setminus \{v_2\}} k_{u \to f_{12}} = k_{v_1 \to f_{12}} = 3$$

从 f_{12} 到 v_2 的输出上的规范，取决于来自 v_1 的输入规范。

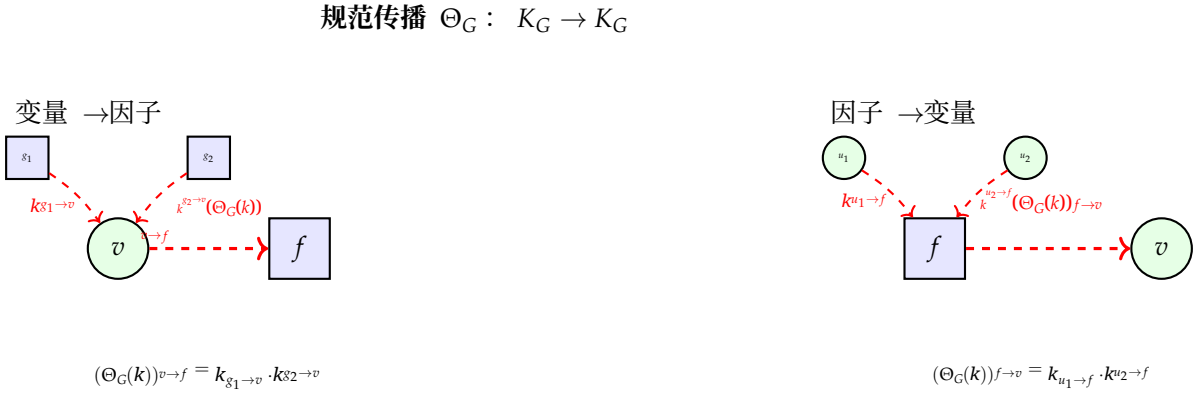
解释：规范沿图传播：输出边上的重缩放取决于输入边上的重缩放。在链上，规范流从一端线性地流向另一端。

4.4 推论：射影 BP 定义良好

推论 4.6 (射影 BP 定义良好)。 由于 T_G 在 Θ_G (定理 4.3) 下是半等变的，因此它在射影消息上诱导出一个定义良好的映射：

$$\overline{T}_G : \mathbb{P}\mathbf{Msg}(G) \to \mathbb{P}\mathbf{Msg}(G)$$

不需要对 Θ_G 作单射假设——下降直接源于半等变性 *e.*



Θ_G 是一个群同态:
$\Theta_G(\mathbf{1}) = \mathbf{1}, \quad \Theta_G(k \cdot k') = \Theta_G(k) \cdot \Theta_G(k')$

图 7：规范传播映射 Θ_G： $K_G \to K_G$。每条半边上的输出规范是输入规范的乘积。左：变量到因子传播。右：因子到变量传播。 Θ_G 是一个群同态，用于控制重缩放如何通过置信传播进行传播。

证明。 我们必须证明，若 $[m] = [m']$（属于同一规范轨道），则 $[T_G(m)] = [T_G(m')]$。

若 $[m] = [m']$，则存在 $m' = k \cdot m$，其中 $k \in K_G$。根据定理 4.3:

$$T_G(m') = T_G(k \cdot m) = \Theta_G(k) \cdot T_G(m)$$

因此，$T_G(m')$ 与 $T_G(m)$ 相差一个规范变换 $\Theta_G(k)$，这意味着**它们属于同一轨道**:

$$[T_G(m')] = [\Theta_G(k) \cdot T_G(m)] = [T_G(m)]$$

最后一个等式成立，是因为 $\Theta_G(k) \in K_G$ 作用于同一个规范群，因此乘以$\Theta_G(k)$ 会保持规范轨道不变。

因此，$\bar{T}_G([m]) = [T_G(m)]$ 定义良好（与代表元的选择无关）。 □ □

注 4.7（为何不需要单射性）。半等变性质足以保证映射下降到商空间。对于一个具有对称群 G 的映射 f : M → M，若对某个同态θ : G → G，有 f (g · m) = θ(g) · f (m)，则 f 会下降到商空间 M/G。要求是 θ 映入作用于 M 的同一个群，并与该作用相容； θ 无需具有单射性。在我们的情形中，若 m′ = k · m 属于同一轨道，则 T_G(m′) = Θ_G(k) · T_G(m)。由于 Θ_G(k) ∈ K_G 作用于轨道空间，因此有 [T_G(m′)] = [T_G(m)]。

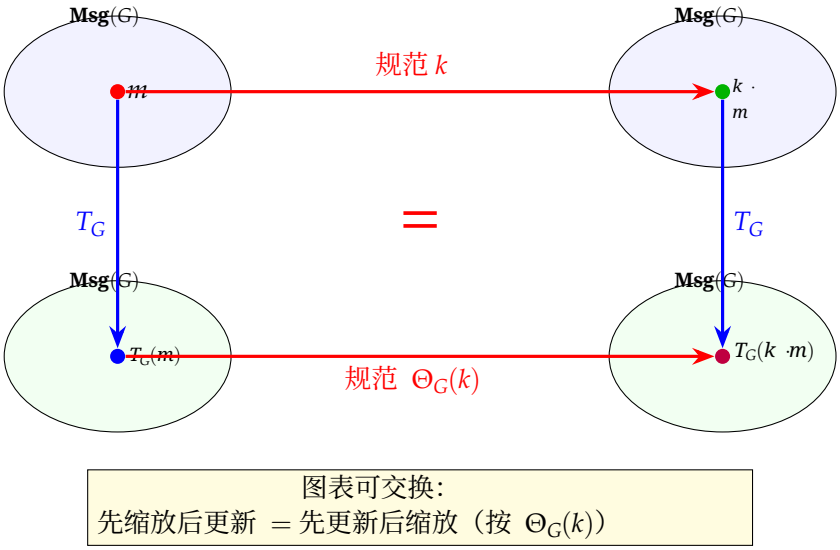
备注 4.8（不动点与归一化信念）。 结合推论 4.6 与命题 **??**，我们可以看到:

- T_G 的不动点可能并不唯一
- 但它们的规范轨道 $[m^*]$ 是唯一的（如果T_G 具有唯一的不动点轨道）
- 归一化信念具有规范不变性，因此它们在轨道上是良定义的

这解释了为什么置信传播可以有多个不动点，而它们都产生相同的归一化信念——它们之间仅相差一个规范变换。

进一步的范畴论/拓扑结构。 我们将群胚和 Kan 复形解释（及其与 π_1 整体的联系）移至附录 C；正文仅使用重缩放作用和 BP 等变性。

半等变性：$T_G(k \cdot m) = \Theta_G(k) \cdot T_G(m)$



图表可交换：
先缩放后更新 = 先更新后缩放（按 $\Theta_G(k)$）

图8：半等变性交换方形：$T_G(k \cdot m) = \Theta_G(k) \cdot T_G(m)$。上方路径：按$k$对输入重新缩放，然后应用置信传播。下方路径：应用置信传播，然后按 $\Theta_G(k)$对输出重新缩放。两条路径得到相同结果。这是信念传播的基本规范对称性。

5 下降重述：精确性即有效下降

本节提供了**统一框架**，用于统一此前的所有结果。关键洞见：

精确推断是沿覆盖将局部语义有效下降为全局语义 .

树（第3节）和连接树（第5节）是下降成功的特殊情况。有环置信传播的失效（第3节提及）是下降障碍。树之所以能够实现精确推断，是因为其覆盖神经是可缩的，从而使下降自动成立。连接树之所以有效，是因为运行交性质保证了下降相容性。当覆盖缺乏足够的细化时，有环置信传播会失效，导致非平凡的下降障碍。

我们证明定理5.16，该定理将定理6.26 （树）和定理6.26 （连接树）都作为推论涵盖在内。

5.1 不可妥协的边界

原则5.1 （局部消息传递的根本局限）。在任意含环图上进行精确推断，**无法**仅凭对边分隔消息执行纯局部操作来实现，除非：

1. 扩大承载体（使用簇而非变量），或
2. 使用高阶相容数据（带有额外结构的消息）

原因：全局语义是$\mathbf{Mat}_R$中的一个极限。当指标图包含非平凡环时，这个极限无法通过一维局部消元计算得出。

我们的成果：我们用**覆盖上的下降**重新定义“局部性”，不再将“树依赖性”视为原始概念。树成为一种充分覆盖（在这种覆盖上，下降自动成立）。树宽则体现为确保下降所需的最小细化复杂度。

5.2 我们计算其语义的对象

本节中始终固定以下内容：

- 一个带类型签名 $\Sigma = (\Lambda, \Gamma, s, t)$

向射影空间下降: $T_G:\ \mathbb{P}\mathbf{Msg}(\overline{G})\ \to \mathbb{P}\mathbf{Msg}(G)$

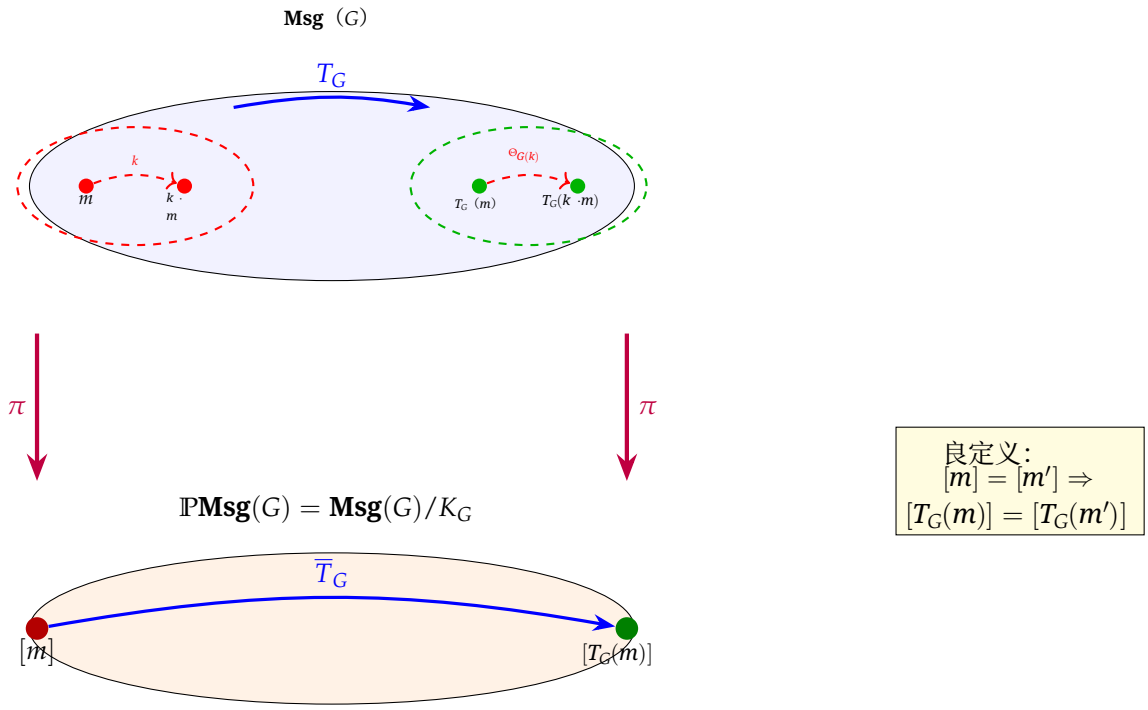


图9：T_G 向射影空间 $\mathbb{P}\mathbf{Msg}(G)$的下降。上：$T_G$ 将规范轨道映射到规范轨道（半等变性）。下：在等价类上诱导的映射 T_G。商映射 $\underline{\pi}$ 使图表可交换。射影 BP T_G 与规范无关，并表示物理上有意义的动力学。

- 一个状态空间赋值 $\Omega:\ \Lambda \to$ **有限集合范畴**
- 因子赋值 $\Phi:\ \Gamma \to \mathbf{Mat}_R$
- 极化因子图 $G\ \in \mathbf{FG}_\Sigma$根据泛性质，我们有如下求值：

$$[\![G]\!]_R : \Omega(X) \to \Omega(Y)$$

在 $\mathbf{Mat}_R$ 中，其中 X、Y 是 G 的源类型和目标类型。

推断目标：计算每个变量 $v\ \in V(G)$的边缘分布 $\mathrm{mar}_v(G)$。

5.3 覆盖：将图分解为多个片段

定义 5.2（因子图的覆盖）。因子图 G 的一个**覆盖** 是一个对 $\mathcal{U}=(I,\ \{U_i\}_{i\in I})$，其中：

- I 是一个有限指标集
- $U_i\ \subseteq V(G)$（变量的子集）

满足：

（Cov1）变量覆盖：

$$\bigcup_{i\in I} U_i = V(G)$$

每个变量至少出现在一个片段中。

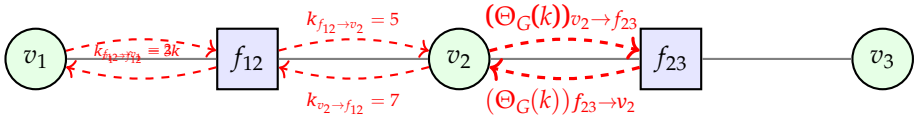
（Cov2）因子覆盖：对于每个因子$f\ \in F(G)$,

$$\exists i \in I : \mathsf{nbhd}(f) \subseteq U_i$$

每个因子都完全包含在某个片段中。

解读：覆盖将因子图分解为相互重叠的片段，每个片段都足够大以包含完整因子。

演算示例：三变量链上的规范传播



变量 →因子：
$(\Theta_G(k))_{v_2\to f_{23}} = \prod_{g\in \text{nbhd}(v_2)\setminus\{f_{23}\}} k_{g\to v_2}$
$= k_{f_{12}\to v_2} = 5$

因子 →变量：
$(\Theta_G(k))_{f_{23}\to v_2} = \prod_{u\in \text{nbhd}(f_{23})\setminus\{v_2\}} k_{u\to f_{23}}$
$= k_{v_3\to f_{23}}$（取决于v_3 规范）

关键洞见：规范沿链线性流动 输出规范 = 来自另一侧的输入规范

图10：三变量链上的显式规范传播。输入规范 k（红色虚线）通过乘积公式确定输出规范 $\Theta_G(k)$（红色粗线）。在链上，规范呈线性传播：$v_2 \to f_{23}$ 上的规范等于$f_{12} \to v_2$,上的规范，从而形成从左到右的流。

示例 5　.3（树分解所给出的覆盖）。 树分解 (T, Bag) 通过设置得到一个覆盖。　g:

$$I = V(T), \quad U_t = \text{Bag}(t) \cap V(G)$$

根据树分解公理：

- 变量覆盖成立（每个变量都在某个袋中）
- 因子覆盖成立（每个因子都适合放入某个袋中）

示例5.4（树的星形覆盖）。 对于以r为根的树因子图G，**星形覆盖**为每个变量v 指定集合：

$$U_v = \{v\} \cup \{\text{neighbors of } v\}$$

这是标准置信传播（第3节）中使用的隐式覆盖。

5.4 局部语义的预层

定义 5.5（局部状态空间）。 对于变量的子集$U \subseteq V(G)$，局部状态空间为**局部状态空间**　is:

$$\Omega(U) = \prod_{v\in U} \Omega(\lambda(v))$$

个体状态空间的笛卡尔积。
解释：变量U的联合配置。

定义 5.6（局部函数空间）。该**局部函数空间**（或局部消息空间）是　s:

$$\mathcal{M}(U) = R^{\Omega(U)}$$

从 $\Omega(U)$ 到R的半模函数 R.

解释：定义在片段U上的势函数或信念。在热带半环$R = \mathbb{T}_{\min}$中，它们是能量函数。

定义 5.7（限制映射（边缘化））。对于$V \subseteq U$，定义**限制映射**：

$$\rho_{U\to V} : \mathcal{M}(U) \to \mathcal{M}(V)$$

by:

$$(\rho_{U\to V} F)(x_V) = \bigoplus_{x_{U\setminus V}} F(x_V, x_{U\setminus V})$$

其中，$\bigoplus$ 是R中的求和运算。

情况:

- **标准半环**$R = \mathbb{R}_{\geq 0}$: $\bigoplus = +$（求和）

$$(\rho_{U\to V} F)(x_V) = \sum_{x_{U\setminus V}} F(x_V, x_{U\setminus V})$$

这就是**边缘化**（对变量求和消去）。

- **热带半环** $R = \mathbb{T}\min$: $\bigoplus = \min$（最小值）

$$(\rho_{U\to V} F)(x_V) = \min_{x_{U\setminus V}} F(x_V, x_{U\setminus V})$$

这就是**最小边缘化**（求最小能量）。

解释：通过对$U \setminus V$ 中的状态求和（或取最小值）来消去变量。这是第**??**节中的余单位运算：$\varepsilon_{U\setminus V}$。

定理 5.8（局部函数空间预层）。赋值:

$$\mathcal{M}_{\mathcal{U}} : \mathcal{I}(\mathcal{U})^{\mathrm{op}} \to \mathbf{SMod}_R$$

定义如下:

- 对象上: $U \mapsto \mathcal{M}(U) = R^{\Omega(U)}$
- 在态射上: $(U \supseteq V) \mapsto \rho_{U\to V} : \mathcal{M}(U) \to \mathcal{M}(V)$

是一个反变函子（预层）。

证明。**恒等态射上的函子性**：对于$U \to U$,

$$\rho_{U\to U}(F)(x_U) = \bigoplus_{x_\emptyset} F(x_U) = F(x_U)$$

因为没有变量需要消去。因此 $\rho_{U\to U} = \mathrm{id}_{\mathcal{M}(U)}$。

复合上的函子性：对于$W \subseteq V \subseteq U$，我们必须证明:

$$\rho_{U\to W} = \rho_{V\to W} \circ \rho_{U\to V}$$

计算右侧:

$$\begin{aligned}(\rho_{V\to W} \circ \rho_{U\to V})(F)(x_W) &= \rho_{V\to W}(\rho_{U\to V}(F))(x_W) \\ &= \bigoplus_{x_{V\setminus W}} (\rho_{U\to V}(F))(x_V) \\ &= \bigoplus_{x_{V\setminus W}} \left(\bigoplus_{x_{U\setminus V}} F(x_U) \right)\end{aligned}$$

由于$(U \setminus V) \cap (V \setminus W) = \varnothing$且$(U \setminus V) \cup (V \setminus W) = U \setminus W$，根据$\bigoplus$的结合律和交换律（半环公理 1-2，第**??**节）：

$$\begin{aligned} &= \bigoplus_{x_{V\setminus W}} \bigoplus_{x_{U\setminus V}} F(x_U) \\ &= \bigoplus_{x_{U\setminus W}} F(x_U) \\ &= \rho_{U\to W}(F)(x_W) \end{aligned}$$

因此， $\mathcal{M}_{\mathcal{U}}$保留复合。 □ □

推论：局部函数空间构成定义在交类上的**预层**，并与边缘化相容。这是“局部信念必须在重叠处相容”这一要求的范畴形式化。

5.5 因子分配与簇势

定义 5.9（覆盖的因子分配）。 覆盖 $\mathcal{U} = (I, \{U_i\}_{i\in I})$ 是一个函数 n:

$$\kappa : F(G) \to I$$

使得对于每个因子$f \in F(G)$:

$$\mathsf{nbhd}(f) \subseteq U_{\kappa(f)}$$

I解释：将每个因子分配给一个足够大的片段，使其包含所有与之相邻的变量 s.

存在性：根据因子覆盖公理 (Cov2)，这样的 κ 总是存在。

D定义 5.10（簇势）。给定分配 κ，片段 $i \in I$ 上的**簇势**为:

$$\Psi_i \in \mathcal{M}(U_i) = R^{\Omega(U_i)}$$

定义为:

$$\Psi_i(x_{U_i}) = \prod_{f:\kappa(f)=i} \phi_f(x_{\mathsf{nbhd}(f)})$$

约定空积等于 1（R中的单位元）。

解释： Ψ_i 是分配给片段i的所有因子的乘积，将其视为定义在U_i的联合状态上的函数。这是对袋势的推广。

命题 5.11 (簇势分解联合分布)。 所有簇势的乘积等于所有原始因子的乘积:

$$\prod_{i\in I} \Psi_i(x_{U_i}) = \prod_{f\in F(G)} \phi_f(x_{\mathsf{nbhd}(f)})$$

对于任意全局配置 $x \in \Omega(V(G))$ （在等式两边分别限制到相应的子集）。

证明。 根据 Ψ_i 的定义:

$$\begin{aligned} \prod_{i\in I} \Psi_i(x_{U_i}) &= \prod_{i\in I} \left(\prod_{f:\kappa(f)=i} \phi_f(x_{\mathsf{nbhd}(f)}) \right) \\ &= \prod_{f\in F(G)} \phi_f(x_{\mathsf{nbhd}(f)}) \quad \text{(rearranging product)} \end{aligned}$$

由于 κ 是一个函数，每个f恰好出现在一个 Ψ_i 中。 □ □

推论：簇势保留原图模型的因子分解结构。这与普适语义相联系：求值 $[\![G]\!]_R$ 通过簇势分解为因子。

5.6 下降数据：局部片段何时可以粘合?

定义 5.12 （下降数据）。 预层 $\mathcal{M}_{\mathcal{U}}$ 的**下降数据** 是一个族$(F_i)_{i\in I}$，其中$F_i \in \mathcal{M}(U_i)$满足：

（DD1）两两相容性：对于每个 i、$j \in I$，若 $U^{ij} = U_i \cap U_j \neq \emptyset$,

$$\rho_{U_i \to U_{ij}}(F_i) = \rho_{U_j \to U_{ij}}(F_j)$$

（DD2）高阶相容性：对于每组三元组 i、j、$k \in I$，若 $U^{ijk} = U_i \cap U_j \cap U_k \neq \emptyset$，则限制到$U^{ijk}$ 后满足：

$$\rho_{U_i \to U_{ijk}}(F_i) = \rho_{U_j \to U_{ijk}}(F_j) = \rho_{U_k \to U_{ijk}}(F_k)$$

对所有更高阶交集也同样成立。

解释：局部函数F_i 在重叠部分上是相容的——对共享变量进行边缘化后，它们彼此“相同”。这是将局部数据粘合为全局数据的条件。

注 5.13（偏序集中的自动成立性）。 由于 $\mathcal{I}$（$\mathcal{U}$） 是偏序集，（DD1）成立后，（DD2）便自动成立：如果$U_{ijk} \subseteq U_{ij} \subseteq U_i$，则根据函子性（定理 5.8）：

$$\rho_{U_i \to U_{ijk}} = \rho_{U_{ij} \to U_{ijk}} \circ \rho_{U_i \to U_{ij}}$$

因此，检查成对交集就足够了。高阶相干性**并非**在对规范取商时自动成立（第 4 节），这就是含规范的环状图会引入障碍的原因。

定义 5.14 （全局粘合）。 如果（F_i）$_{i\in I}$ 是一个下降数据，则**全局粘合** 是一个函数 n:

$$F \in \mathcal{M}(V(G)) = R^{\Omega(V(G))}$$

使得对所有 $i \in I$，均有：

$$\rho_{V(G) \to U_i}(F) = F_i$$

解释：F 是一个全局函数，其在所有片段上的限制都能还原局部数据 .

定理 5.15（有限状态空间的有效下降）。 对于有限状态空间 Ω（U），每个下降数据（F_i）$_{i\in I}$ 都有一个唯一的全局粘合 $F \in \mathcal{M}$（V（G））。

证明。 **构造**：对于每个全局配置$x \in \Omega$（V（G）），定义：

$$F(x) = F_i(x|_{U_i})$$

其中 i 是任意满足 $\mathrm{supp}(x) \subseteq U_i$ 的指标（其中 $\mathrm{supp}(x)$ 是 x 有定义的变量集合）。**W ell-defined性**: We 必须 s how $F(x)$ is independent o f the choice of i.

若 $\mathrm{supp}(x) \subseteq U_i$ 且 $\mathrm{supp}(x) \subseteq U^j$，则 $\mathrm{supp}(x) \subseteq U^{ij}$。根据两两相容性（DD1）：

$$\rho_{U_i \to U_{ij}}(F_i)(x|_{U_{ij}}) = \rho_{U_j \to U_{ij}}(F_j)(x|_{U_{ij}})$$

但由于 $x|_{U_i}$ 和 $x|_{U_j}$ 在U^{ij}上相同，因此我们有：

$$F_i(x|_{U_i}) = \rho_{U_i \to U_{ij}}(F_i)(x|_{U_{ij}}) = \rho_{U_j \to U_{ij}}(F_j)(x|_{U_{ij}}) = F_j(x|_{U_j})$$

因此$F(x)$ 是定义良好的。

粘合性质：根据构造，$\rho_{V(G) \to U_i}(F) = F_i$ 对所有i都成立。

唯一性：若F' 是另一个粘合，则对所有x：

$$F'(x) = F'|_{U_i}(x|_{U_i}) = F_i(x|_{U_i}) = F(x)$$

因此$F' = F$。 □ □

结果：对于有限离散模型（本文始终讨论这类模型），下降是**有效的**——相容的局部数据能够唯一确定全局数据。这是精确推断的范畴论基础。

5.7 普适精确性定理

定理 5.16（精确推断作为有效下降）。设 $G \in \mathbf{FG}_\Sigma$ 是一个极化因子图，（Ω，Φ）是一个解释，$\mathcal{U} = (I, \{U_i\}_{i\in I})$ 是一个带有分配 κ 的覆盖。

令 $(\Psi_i)_{i\in I}$ 为簇势（定义5.10）。

A 部分（正向）：如果存在一个消息传递方案，生成簇信念 $(B_i)_{i\in I}$，且 $B_i \in \mathcal{M}(U_i)$ 构成下降数据：

$$\rho_{U_i\to U_{ij}}(B_i) = \rho_{U_j\to U_{ij}}(B_j) \quad \forall i,j$$

那么存在一个全局函数 $F \in \mathcal{M}(V(G))$，使得：

1. $\rho_{V(G)\to U_i}(F) = B_i$ 对所有 i 均成立（粘合性质）

2. F 的单变量边缘分布等于范畴边缘分布：

$$\mathrm{mar}_v(F) = \mathrm{mar}_v([\![G]\!]_R)$$

对所有 $v \in V(G)$

B 部分（反向）：如果 $F \in \mathcal{M}(V(G))$ 表示来自 $[\![G]\!]_R$的未归一化联合分布，则：

$$(F_i)_{i\in I} := (\rho_{V(G)\to U_i}(F))_{i\in I}$$

是 $\mathcal{M}_\mathcal{U}$ 的下降数据。

解释：局部推断方案的精确性恰好是指它计算出全局语义的有效下降数据这一性质。

证明。**A 部分：全局粘合的存在性**

根据定理 5.15，下降数据 $(B_i)_{i\in I}$ 存在唯一的全局粘合 $F \in \mathcal{M}(V(G))$，满足：

$$\rho_{V(G)\to U_i}(F) = B_i \quad \forall i \in I$$

这就证明了 (1)。我们还必须证明 (2)：即F 与 $[\![G]\!]_R$表示相同的语义。

步骤 1：将F 与簇势联系起来

假设簇信念B_i 是通过一致的消息传递方案（例如连接树信念传播）从 Ψ_i 推导得出的。根据构造，每个B_i 都纳入分配给片段i的所有因子：

$$B_i(x_{U_i}) \propto \Psi_i(x_{U_i}) \cdot (\text{incoming separator messages})$$

当消息传递方案产生下降数据时，分隔消息会确保一致性，因此粘合得到的F 满足：

$$F(x) \propto \prod_{i\in I} \Psi_i(x_{U_i})$$

根据命题 5.11：

$$\prod_{i\in I} \Psi_i(x_{U_i}) = \prod_{f\in F(G)} \phi_f(x_{\mathsf{nbhd}(f)})$$

步骤 2：关联范畴语义

求值 $[\![G]\!]_R$ 表示联合分布：

$$[\![G]\!]_R(x) = \prod_{f\in F(G)} \phi_f(x_{\mathsf{nbhd}(f)})$$

（忽略归一化常数；重新归一化后，它们不会影响边缘分布）。

因此，F 和 $[\![G]\!]_R$ 表示相同的未归一化分布。

步骤 3：边缘分布一致

对于任意变量 $v \in V(G)$，从 F 得到的边缘分布为：

$$\mathrm{mar}_v(F)(x_v) = \rho_{V(G)\to\{v\}}(F)(x_v) = \sum_{x_{V(G)\setminus\{v\}}} F(x)$$

根据 $\mathbf{Mat}_R$ 中的余单位运算：

$$\varepsilon_{V(G)\setminus\{v\}} : R^{\Omega(V(G))} \to R^{\Omega(\{v\})}$$

is e对已消去的变量求和。T hus:

$$\mathrm{mar}_v(F) = \varepsilon_{V(G)\setminus\{v\}} \circ F = \varepsilon_{V(G)\setminus\{v\}} \circ [\![G]\!]_R = \mathrm{mar}_v([\![G]\!]_R)$$

由求值的函子性可得。

B 部分：限制构成下降数据

全局未归一化联合分布为 $F \in \mathcal{M}(V(G))$。定义：

$$F_i = \rho_{V(G)\to U_i}(F)$$

对于任意满足 i,j 且 $U_{ij} \neq \emptyset$ 的情形：

$$\begin{aligned}
\rho_{U_i\to U_{ij}}(F_i) &= \rho_{U_i\to U_{ij}}(\rho_{V(G)\to U_i}(F)) \\
&= \rho_{V(G)\to U_{ij}}(F) \quad \text{(by functoriality, Theorem 5.8)} \\
&= \rho_{U_j\to U_{ij}}(\rho_{V(G)\to U_j}(F)) \\
&= \rho_{U_j\to U_{ij}}(F_j)
\end{aligned}$$

因此，$(F_i)_{i\in I}$ 满足两两相容性（DD1）。高阶相容性（DD2）可由函子性以类似方式推出。 □ □

推论：该定理**统一了所有精确性结果**：

- **树精确性**（定理 6.26）：树的星形覆盖具有可缩的神经，因此下降自动成立 tic
- **连接树精确性**（定理6.26）：运行交性质确保两两相容性，从而生成下降数据
- **有环置信传播失败**：缺乏充分细化的覆盖无法生成下降数据——消息在重叠部分不相容

5.8 推论：重获先前结果

推论 5.17（由下降得到树精确性）。定理 *6.26*（树上的置信传播精确）是定理*5.16* 在覆盖神经可缩时的一个特例。

证明。 对于一棵树 G，使用例 5.3中的星形覆盖。双遍信念传播调度会产生确保分隔集一致性的消息。

由于树中没有环，所有重叠兼容性都会自动满足（不存在非平凡的相干条件）。因此，置信传播会计算出下降数据；根据定理 5.16，信念等于范畴边缘分布。 □ □

推论 5.18（由下降得到连接树精确性）。 定理 *6.26*（连接树置信传播是精确的）是定理 *5.16* 在覆盖满足运行交性质时的特例。

证明。 连接树 J 由树分解 (T, Bag)构建而成。覆盖 $\mathcal{U} = \{U_t = \mathrm{Bag}(t)\}_{t\in V(T)}$ 满足变量覆盖和因子覆盖。

运行交性质确保对于相邻的包 t,t'，分隔集 $S_{t,t'} = U_t \cap U_{t'}$能确保一致性：

$$\rho_{U_t\to S_{t,t'}}(B_t) = \rho_{U_{t'}\to S_{t,t'}}(B_{t'})$$

由于 T 是一棵树，因此不存在高阶环，所以这种两两相容性足以满足下降条件。根据定理 5.16，连接树信念传播会计算出精确边缘分布。 □ □

总结：第 5 节提供了**统一框架**。树和连接树并非偶然成为特殊情形——它们是下降成功的覆盖。范畴论视角使这一点一目了然。

接下来：第6 节构造一种**算法**，通过全纯感知树编译实现有效下降。

6 全纯感知树编译（HATCC）

定理5.16将精确推断刻画为有效下降。树和连接树之所以能够成功，是因为它们的覆盖允许存在下降数据。含环图则会因环阻碍一致粘合而失败。

我们介绍全纯感知树编译（HATCC）：一种通过整体检测下降障碍，并通过模式变量编译消除这些障碍的算法。对于因子神经C_e 中的每个环$G_{\mathcal{N}}$，我们计算整体矩阵H_e，以衡量沿C_e的平行移动。非平凡的H_e会阻碍下降；HATCC 将此类障碍编译为离散模式变量，从而简化为树形置信传播。

核心思想：使用**因子神经图**$G_{\mathcal{N}}$（因子作为顶点，重叠作为边），而非基于变量的覆盖。对于$G_{\mathcal{N}}$中的每个环，计算**整体矩阵**H_e，以衡量信念如何沿环“传输”。当整体非平凡时，通过强连通分量（SCC）商将其编译为离散**模式变量**，随后在增广图上运行树形置信传播。

与第5节的关系：HATCC 通过以下方式实现定理5.16：

- 从因子神经构造专用覆盖（关联定义5.2）
- 通过整体单值性检测下降数据是否存在（联系定义 5.12）
- 通过添加模态变量扩充覆盖以确保相容性（联系定义 5.14）

结构：

- §6.1：因子神经图（与交类对偶，定义 **??**）
- §6.2：骨干树 + 弦（分解为树 + 环）
- §6.3：整体单值性矩阵（沿环传输，联系限制映射 $\rho_{U\to V}$，定义5.7）
- §6.4：模式商（整体的强连通分量抽象）
- §6.5：选择器因子（强制实现下降相容性）
- §6.6：HATCC 编译算法
- §6.7：精确性刻画（何时 HATCC =树形置信传播）

6.1 因子神经图

第 5 节通过覆盖组织推断 $\mathcal{U} = (I, \{U_i\}_{i\in I})$ 变量集（定义 5.2）。交类 $\mathcal{I}(\mathcal{U})$ 编码重叠，而预层 $\mathcal{M}_{\mathcal{U}}$：$\mathcal{I}(\mathcal{U})^{\mathrm{op}} \to \mathbf{SMod}_R$（定理 5.8）组织局部函数空间。

HATCC 使用一种**以因子为中心的**对偶结构：我们不对变量进行划分，而是根据因子 之间的连通性来组织它们。

定义 6.1（因子神经图）。对于一个极化因子图 $G \in \mathbf{FG}_\Sigma$（定义 **??**），其因子为$F(G)$，且对每个$f \in F(G)$，其邻域为 $\mathrm{nbhd}(f) \subseteq V(G)$，则因子神经图 定义如下：

$$G_{\mathcal{N}} = (F(G), E_{\mathcal{N}})$$

顶点：因子的集合 $F(G)$。
边：对于$f_1, f_2 \in F(G)$，当且仅当满足以下条件时，将边$(f_1, f_2) \in E_{\mathcal{N}}$加入图中：

$$\mathsf{nbhd}(f_1) \cap \mathsf{nbhd}(f_2) \neq \emptyset$$

边接口：对于每条$e = (f_1, f_2) \in E_{\mathcal{N}}$，定义如下：

$$J_e := \mathsf{nbhd}(f_1) \cap \mathsf{nbhd}(f_2)$$

边权重：

$$w(e) := \log |\Omega(J_e)|$$

其中 $\Omega(J_e)$ 是局部状态空间（定义5.5）。

备注 6.2（“神经”的三种不同构造）。 本文使用了三种不同的神经概念：

1. 范畴的神经：应用于消息作用群胚 $\mathcal{T}_G$ 以获得单纯集 $N(\mathcal{T}_G)$（用于第4 节的 Kan 复形结构）。

2. 覆盖的 ˇCech 神经 （第5节中隐含）：对于覆盖 $\mathcal{U}$，交类 $\mathcal{I}(\mathcal{U})$的神经编码了高阶重叠（用于下降图表）。

3. F因子神经 （定义 6.1）：以因子为顶点的图，用于整体性计算 .

这些 彼此相关但并不相同：(1) 和 (2) 产生单纯集合，而 (3) 是加权图（即 st 其 1-骨架 因此，因子神经细化了 ˇCech 神经的 1-骨架。

命题 6.3 （因子神经细化覆盖神经）。 设 $\mathcal{U} = (I, \{U_i\}_{i\in I})$ 是一个覆盖（定义 5.2），具有因子分配 $\kappa : F(G) \to I$（定义5.9）。

若 $(f_1, f_2) \in E_{\mathcal{N}}$，则以下二者必居其一：

1. $\kappa(f_1) = \kappa(f_2)$（同一覆盖片），或
2. $U_{\kappa(f_1)} \cap U_{\kappa(f_2)} \neq \emptyset$（相邻覆盖片）

解释：因子神经边细化了交类范畴态射。

证明。 如果 $(f_1, f_2) \in E_{\mathcal{N}}$，那么 邻域$(f_1)$∩邻域$(f_2) \neq \emptyset$根据定义 6.1。

根据因子分配（定义 5.9）：邻域$(f_i) \subseteq U_{\kappa(f_i)}$对于 $i = 1, 2$。因此：

$$U_{\kappa(f_1)} \cap U_{\kappa(f_2)} \supseteq \mathsf{nbhd}(f_1) \cap \mathsf{nbhd}(f_2) \neq \emptyset$$

因此，要么 $\kappa(f_1) = \kappa(f_2)$要么 $(U_{\kappa(f_1)}, U_{\kappa(f_2)})$是$N(\mathcal{I}(\mathcal{U}))$的1-骨架中的一条边。 □ □

例 6.4 （4-环的因子神经）。 考虑一个有4个变量 {A、B、C、D}（每个变量均为二元：$\Omega(\lambda(v)) = \{0, 1\}$）和4个因子：

$$\begin{aligned} f_1 &: \{A, B\} \to R, \quad \mathsf{nbhd}(f_1) = \{A, B\} \\ f_2 &: \{B, C\} \to R, \quad \mathsf{nbhd}(f_2) = \{B, C\} \\ f_3 &: \{C, D\} \to R, \quad \mathsf{nbhd}(f_3) = \{C, D\} \\ f_4 &: \{D, A\} \to R, \quad \mathsf{nbhd}(f_4) = \{D, A\} \end{aligned}$$

步骤1：计算因子神经$G_{\mathcal{N}}$

顶点：$F(G) = \{f_1, f_2, f_3, f_4\}$。

边：检查所有成对的重叠：

- $(f1, f2)$：邻域(f_1)∩邻域(f_2)= $\{A, B\} \cap \{B, C\} = \{B\} \neq \emptyset \to$带接口的边 $J_{(f_1,f_2)} = \{B\}$
- $(f2, f3)$：$\{B, C\} \cap \{C, D\} = \{C\} \to$带接口的边 $J_{(f_2,f_3)} = \{C\}$
- $(f3, f4)$：$\{C, D\} \cap \{D, A\} = \{D\} \to$带接口的边 $J_{(f_3,f_4)} = \{D\}$
- $(f4, f1)$：$\{D, A\} \cap \{A, B\} = \{A\} \to$带接口的边 $J_{(f_4,f_1)} = \{A\}$
- $(f1, f3)$：$\{A、B\} \cap \{C、D\} = \emptyset \to$没有边
- $(f2, f4)$：$\{B、C\} \cap \{D、A\} = \emptyset \to$没有边

结果：$G_{\mathcal{N}}$是一个 4-环：$f_1 - f_2 - f_3 - f_4 - f_1$。

边权重：对于二元变量，$|\Omega(J_e)| = 2$，因此：

$$w(e) = \log 2 \approx 0.693 \quad \forall e \in E_{\mathcal{N}}$$

步骤 2：与覆盖的联系

一个自然覆盖（定义5.2）是：

$$\mathcal{U} = (\{1, 2, 3, 4\}, \{U_1 = \{A, B\}, U_2 = \{B, C\}, U_3 = \{C, D\}, U_4 = \{D, A\}\})$$

因子分配：$\kappa(f_i) = i$（每个因子各自位于一个片段中）。

交类 $\mathcal{I}(\mathcal{U})$：

- 0-胞腔：$U1$、$U2$、$U3$、$U4$

- 1-胞腔：$U_1 \cap U_2 = \{B\}$、$U_2 \cap U_3 = \{C\}$、$U_3 \cap U_4 = \{D\}$、$U_4 \cap U_1 = \{A\}$
- 2-胞：不存在三重重叠（对于互不相同的$U_i \cap U_j \cap U_k = \emptyset$ i、j、k，所有这些交集均为空）

$N(\mathcal{I}(\mathcal{U}))$的1-骨架与$G_{\mathcal{N}}$是相同的4-环图，这印证了命题6.3。

观察：这个环构成了基于树的推断的一个拓扑障碍。在这个3节中介绍的标准置信传播（BP）会在此图上迭代，但可能不收敛，也可能收敛到错误的边缘分布（有环置信传播问题）。HATCC 通过整体性计算解决了这一问题（第6.3节）。

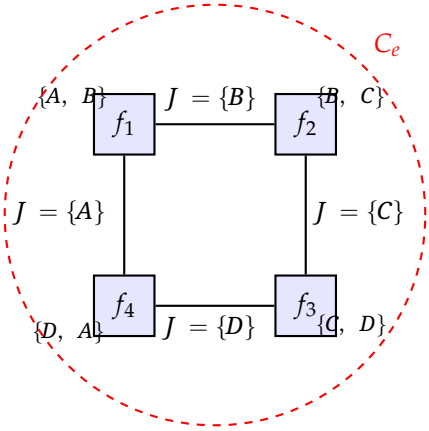


图 11：示例6.4的因子神经$G_{\mathcal{N}}$。因子f_1, f_2, f_3, f_4 构成一个带有边接口J_e = 邻域$(f_i) \cap$ 邻域(f^j)的 4-环。该环C_e具有 $\beta_1 = 1$，因此产生了潜在下降障碍，需要进行整体性分析。

算法 1 构造因子神经图 要求：具有因子 F(G) 以及每个因子 f 的邻域 nbhd(f) 的因子图 G $\in$ 确保：因子神经 $G_{\mathcal{N}} = (F(G), E_{\mathcal{N}})$，其边接口为 $\{J_e\}_{e \in E_{\mathcal{N}}}$，权重为 $\{w(e)\}_{e \in E_{\mathcal{N}}}$1：初始化 $E_{\mathcal{N}} \leftarrow \emptyset$、J $\leftarrow \{\}$、w $\leftarrow \{\}$2：对于每一对 (f_1, f_2)，其中 $f_1, f_2 \in F(G)$ 且 $f_1 < f_2$，执行 ▷ 唯一的成对组合 3：$J_e \leftarrow$ nbhd(f_1) $\cap$ nbhd(f_2) ▷ 计算接口 4：如果 $J_e \neq \emptyset$，则 ▷ 非空重叠 5：将边 (f_1, f_2) 添加到 $E_{\mathcal{N}}$6：J[(f_1, f_2)] $\leftarrow J_e$ ▷ 存储接口 7：w[(f_1, f_2)] $\leftarrow \sum_{v \in J_e} \log |\Omega(\lambda(v))|$ ▷权重 8：结束 if 9：结束 for 10：返回 ($G_{\mathcal{N}} = (F(G), E_{\mathcal{N}})$、J、w)

定理 6.5 (因子神经构造复杂度)。 算法 *1* 具有：

- 时间：$O(|F(G)|^2 \cdot \text{最大}_{\text{因子}}\ |\text{邻域}(\text{因子})|)$
- 空间: $O(|F(G)| + |E_{\mathcal{N}}|)$

其中 $|F(G)|$ 是因子的数量，最大$_f$ |邻域(因子)| 是最大因子度。

证明。 **时间复杂度**：

- 第 2 行：遍历所有因子对：$(\binom{|F(G)|}{2}) = O(|F(G)|^2)$ 次迭代
- 第 3 行：使用哈希集合，在每次迭代中计算两个大小至多为 max$_f$ |邻域(f)|的集合的交集：$O(\max_f |\text{邻域}(f)|)$
- 第 5–7 行：常数时间操作（添加边、存储接口）
- 第 8 行：对 $|J_e| \leq \max_f$ |邻域(f)| 变量：$O(\max_f |\text{邻域}(f)|)$

总计：$O(|F(G)|^2 \cdot \text{最大值}_f |\text{邻域}(f)|)$.

空间复杂度：

- 存储 $|F(G)|$顶点： $O(|F(G)|)$
- 最多存储 $(\binom{|F(G)|}{2}) = O(|F(G)|^2)$ 条边，但通常为 $|E_{\mathcal{N}}| \ll |F(G)|^2$ (稀疏)： $O(|E_{\mathcal{N}}|$)
- 存储接口和权重： $O(|E_{\mathcal{N}}|)$

总计： $O(|F(G)| + |E_{\mathcal{N}}|)$。

实际考量：对于稀疏因子图（每个因子只与少数其他因子重叠）， $|E_{\mathcal{N}}| = O(|F(G)|)$，因此总时间复杂度为$O(|F(G)|^2 \cdot \max_f |\text{邻域}(f)|)$，空间复杂度为$O(|F(G)|)$。 □ □

6.2 骨干树与基本环

因子神经$G_{\mathcal{N}}$ 编码了因子的连通性，但其中包含环，因而无法直接应用树置信传播。我们的策略是**将** $G_{\mathcal{N}}$分解为：

- 一棵**生成树** T （即“骨干”），在其上可以使用标准树形置信传播。
- 一组**弦** $\mathcal{C}$ （用于构成环的边），它们编码了整体性障碍。

这是图论中的经典分解：每个连通图都存在一棵生成树，而每条弦都会通过在树中闭合一条路径，生成唯一的 基本环。弦的数量就是**圈复杂度** $c = |E_{\mathcal{N}}| - |F(G)| + 1$，它决定了 HATCC 的计算成本（定理6.24）。

直观理解：如果所有弦都具有 平凡 *holonomy* （信念沿基本环传输时保持一致），那么对于推断而言，该图的表现就像一棵树。当 holonomy 非平凡时，我们必须通过模态变量显式追踪它（第6.4节）。

定义 6.6 （骨干树与弦）。 **图** $G_{\mathcal{N}} = (F(G), E_{\mathcal{N}})$ 的骨干树是生成树$T \subseteq E_{\mathcal{N}}$ （连通、无环且包含所有顶点）。

这些**弦**是：

$$\mathcal{C} := E_{\mathcal{N}} \setminus T$$

数量： $|T| = |F(G)| - 1$ 和 $|\mathcal{C}| = |E_{\mathcal{N}}| - |F(G)| + 1$。

定义 6.7 （基本环）。 对于弦$e = (u, v) \in \mathcal{C}$， **基本环** C_e 是如下简单环：

$$C_e = P_T(u, v) \cup \{e\}$$

其中$P_T(u, v) = (u = f_0, f_1, \ldots, f_k = v)$ 是树T 中从u 到v的唯一路径。

弦接口 是$J_e = \text{邻域}(u) \cap \text{邻域}(v)$。

命题 6.8（弦生成 H_1）。 基本环 $\{C_e : e \in \mathcal{C}\}$ 构成一个 $\mathbb{Z}$-基 *fo* *r:*

$$H_1(G_{\mathcal{N}}; \mathbb{Z}) \cong \mathbb{Z}^{|\mathcal{C}|}$$

证明。 因子神经$G_{\mathcal{N}} = (F(G), E_{\mathcal{N}})$ 是一个连通图，具有 $|V| = |F(G)|$ 个顶点和 $|E| = |E_{\mathcal{N}}|$条边。骨干树 T 覆盖所有顶点，含有 $|T| = |V| - 1$ 条边。因此：

$$|\mathcal{C}| = |E_{\mathcal{N}}| - |T| = |E| - (|V| - 1) = |E| - |V| + 1$$

根据图的第一贝蒂数的标准公式： $\beta_1(G_{\mathcal{N}}) = |E| - |V| + 1 = |\mathcal{C}|$。

每条弦$e \in \mathcal{C}$ 确定一个唯一的基本环$C_e = P_T(u,v) \cup \{e\}$， 其中$e = (u, v)$。这些环彼此独立：移除任意一条弦$e$ 会使 β_1 恰好减少 1。按构造，它们生成$H_1(G_{\mathcal{N}}; \mathbb{Z})$。 □ □

与下降的联系：每条弦$e \in \mathcal{C}$ 都代表一个潜在的**下降障碍**。如果下降成功（定义5.12），环C_e 就不应阻碍相容性。我们通过整体对其进行形式化。

例 6.9（例6.4的续例：骨干分解）。继续例6.4，我们有$G_{\mathcal{N}}$，它是一个带有以下边的 4-环：

$$E_{\mathcal{N}} = \{(f_1, f_2), (f_2, f_3), (f_3, f_4), (f_4, f_1)\}$$

步骤 3：选择骨干树 T
由于所有边权重都相等（对所有 e，$w(e) =$ 均为 $\log 2$），任意生成树都是最大生成树。选择：

$$T = \{(f_1, f_2), (f_2, f_3), (f_3, f_4)\}$$

这是一条路径：$f_1 - f_2 - f_3 - f_4$。

步骤 4：识别弦

$$\mathcal{C} = E_{\mathcal{N}} \setminus T = \{(f_4, f_1)\}$$

单条弦：$e = (f_4, f_1)$，接口为 $J_e = \{A\}$。

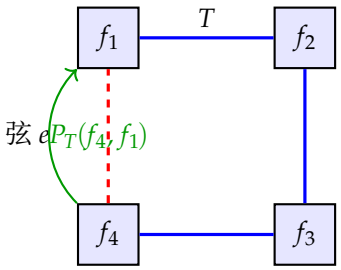
步骤 5：基本环
对于弦 $e = (f_4, f_1)$:

- 树路径 $P_T(f4, f1) = (f4, f3, f2, f1)$ (沿树反向前进)
- 基本环：$C_e = (f4, f3, f2, f1, f4)$ (原始的 4-环)

树边接口:

$$\begin{aligned} J_0 &= \mathsf{nbhd}(f_4) \cap \mathsf{nbhd}(f_3) = \{D, A\} \cap \{C, D\} = \{D\} \\ J_1 &= \mathsf{nbhd}(f_3) \cap \mathsf{nbhd}(f_2) = \{C, D\} \cap \{B, C\} = \{C\} \\ J_2 &= \mathsf{nbhd}(f_2) \cap \mathsf{nbhd}(f_1) = \{B, C\} \cap \{A, B\} = \{B\} \\ J_3 &= J_e = \{A\} \quad \text{(chord interface)} \end{aligned}$$

环传输沿接口传递状态：$\{A\} \xrightarrow{f4} \{D\} \xrightarrow{f3} \{C\} \xrightarrow{f2} \{B\} \xrightarrow{f1} \{A\}$。这套设置可用于整体性计算（第6.3节）。



骨干树 T（蓝色实线）弦 $e = (f_4, f_1)$（红色虚线）基本环 $C_e = P_T \cup \{e\}$

图 12：示例 6.9 中的骨干树 T（蓝色实线）和弦 $e = (f_4, f_1)$（红色虚线）。树路径 $P_T(f_4, f_1)$ 与 e 结合形成基本环 C_e。弦生成 $H_1(G_{\mathcal{N}})$。

算法 2 骨干树选择和弦识别 要求：带权重 $\{\mathrm{w(e)}\}_{e\in E_{\mathcal{N}}}$ 的因子神经 $\mathrm{G}_{\mathcal{N}} = (\mathrm{F(G)}, \mathrm{E}_{\mathcal{N}})$确保：骨干树 $\mathrm{T} \subseteq \mathrm{E}_{\mathcal{N}}$、弦 $\mathcal{C}$、根 $\mathrm{r} \in \mathrm{F(G)}$，父节点/子节点映射 1: $\mathrm{T} \leftarrow$ MAXIMUMSPANNINGTREE($\mathrm{G}_{\mathcal{N}}$, w) ▷ 克鲁斯卡尔算法或普里姆算法，$\mathrm{O}(|\mathrm{E}_{\mathcal{N}}| \log |\mathrm{E}_{\mathcal{N}}|)$ 2: $\mathcal{C} \leftarrow \mathrm{E}_{\mathcal{N}} \setminus \mathrm{T}$ ▷ 弦 3: $\mathrm{r} \leftarrow$ 任意根（例如 F(G) 中字典序最小的因子）4: ▷ 以根为起点构建树：从 r 开始进行广度优先搜索以计算父节点/子节点 5: parent $\leftarrow \{\}$、children $\leftarrow \{\mathrm{f} : [] \mid \mathrm{f} \in \mathrm{F(G)}\}$6: parent[r] $\leftarrow$None 7: queue $\leftarrow$ [r]8: 当队列非空时执行 9: u $\leftarrow$ dequeue(queue) 10: 对 T 中 u 的每个邻居 v 执行 ▷ 树中的相邻节点 11: 如果 parent[v]未设置，则 12: parent[v] $\leftarrow$ u 13: children[u].append(v) 14: enqueue(queue, v) 15: 结束 if 16: 结束 for 17: 结束 while 18: 返回 (T, $\mathcal{C}$, r,parent, children)

定理 6.10（主干选择复杂度）。 算法 2 具有：

- 时间: $O(|E_{\mathcal{N}}| \log |E_{\mathcal{N}}|)$ （由最大生成树主导）
- 空间: $O(|F(G)| + |E_{\mathcal{N}}|)$

证明。**时间**：

- 第 1 行：通过克鲁斯卡尔算法求最大生成树：$O(|E_{\mathcal{N}}| \log |E_{\mathcal{N}}|)$
- 第 2 行：集合差：$O(|E_{\mathcal{N}}|)$
- 第 4–16 行：在树 T 上进行广度优先搜索，复杂度为 $|F(G)|$ 个顶点：$O(|F(G)| + |T|) = O(|F(G)|)$

总计：$O(|E_{\mathcal{N}}| \log |E_{\mathcal{N}}|)$。

空间：存储树 T、弦 $\mathcal{C}$、父子关系：$O(|F(G)| + |E_{\mathcal{N}}|)$。 □ □

6.3 整体单值性矩阵：沿环传输

现在我们来探讨核心问题：**何时环会阻碍下降?** 直观来说，如果我们沿基本环 γ_e “传输”一个信念（每一步都遵循限制映射），就应该回到同一个信念。如果没有，环就会呈现出**非平凡的 holonomy**，从而阻碍全局一致性。

用物理学术语来说：holonomy 衡量向量在纤维丛中沿闭环平行移动时累积的“相移”。这里，纤维是接口变量上的状态空间，而传输由因子势函数决定。

关键洞见：对于下降，我们不需要完整的概率传输（通过 $\rho_{U\to V}$进行边缘化，见 5.7）。我们只需要**布尔支撑骨架**：哪些状态对（x、y）通过某个因子是相容的？这由**传输核**刻画。

定义 6.11 （传输核）。 对于具有势函数 ϕ_f 的因子$f \in F$（G）： Ω（邻域（f）$\to R$（定义3.1），以及接口U、$V \subseteq$ 邻域（f），定义**传输核**：

$$K_f^{U\to V} : \Omega(U) \times \Omega(V) \to \{0, 1\}$$

by:

$$K_f^{U\to V}(x, y) := \begin{cases} 1 & \text{if } \exists z \in \Omega(\mathsf{nbhd}(f) \setminus (U \cup V)) : \phi_f(x, y, z) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

解释：$K_f^{U\to V}(x,y) = 1$ 表示状态 $x \in \Omega(U)$ 和 $y \in \Omega(V)$ 通过因子 f相容（存在一个扩展 z，其势函数非零）。

备注 6.1　2（传输与限制）。 传输核是限制映射的**布尔支撑**：

- 限制映射 $\rho_{U\to V}$（定义 5.7）：概率边缘化（$\bigoplus$-求和）
- 传输核 $K_{U\to Vf}$：布尔可达性（$\exists$-量化）

对于半环 R:

$$K_f^{U\to V}(x,y)=1 \quad\Longleftrightarrow\quad \rho_{\mathsf{nbhd}(f)\to(U\cup V)}(\phi_f)(x,y)\neq 0$$

传输　编码哪些跃迁是可能的；限制映射计算这些跃迁的概率　s.

定义 6.13（环整体）。 对于基本环 $C_e=(f_0,f_1,\ldots,f_k,f_0)$（定义 6.7），其弦接口为 J_e，定义**树边接口**:

$$J_i:=\mathsf{nbhd}(f_i)\cap\mathsf{nbhd}(f_{i+1})\quad\text{for } i=0,\ldots,k-1$$

以及 $J_k:=J_e$（弦接口）。

整体矩阵为:

$$H_e:=K_{f_0}^{J_0\to J_1}\circ K_{f_1}^{J_1\to J_2}\circ\cdots\circ K_{f_{k-1}}^{J_{k-1}\to J_k}\circ K_{f_k}^{J_k\to J_0}:\Omega(J_e)\times\Omega(J_e)\to\{0,1\}$$

其中，$\circ$ 是布尔矩阵乘法（$\wedge$ 表示乘积，$\vee$ 表示求和）。

解释　：$H_e(x、y)=1$ 表示状态 $x\in\Omega(J_e)$ 可以通过沿环行进，从而“传输”到状态 $y\in\Omega(J_e)$　g 沿环　路的C_e 通过因子支撑。

定理 6.14 (整体检测下降障碍**)**。 设 $\mathcal{U}$ 是一个覆盖，带有因子分配 κ，并假设某个消息传递方案产生了簇信念$(B_i)_{i\in I}$。

如果 $G_{\mathcal{N}}$ 中的所有环都具有平凡 ***holonomy***:

$$H_e(x,x)=1\quad\forall e\in\mathcal{C},\forall x\in\Omega(J_e)$$

且 $H_e(x,y)=0$ 对于 $x\neq y$,

那么 $(B_i)_{i\in I}$ 满足分隔子兼容性（T 中的边）便自动构成下降数据（定义5.12）。

反过来，如果某个环具有非平凡的 ***holonomy***，即使树分隔符相容，下降也可能失败。

证明。 我们证明两个方向，并将 holonomy 与下降条件（定义 5.12）联系起来。

设定：令 $\mathcal{U}=(\mathrm{I},\{\mathrm{U}_i\}_{i\in I})$ 为一个覆盖，其因子分配为 $\kappa:\mathrm{F(G)}\to\mathrm{I}$（定义 5.9）。簇信念 $(\mathrm{B}_i)_{i\in I}$，其中 $\mathrm{B}_i\in\mathcal{M}(\mathrm{U}_i)=\mathrm{R}^{\Omega(U_i)}$（定义 5.5）。

Ass假设：对于所有树边 $e=(f_1,f_2)\in T$，信念满足**树分隔相容性**　:

$$\rho_{U_{\kappa(f_1)}\to J_e}(B_{\kappa(f_1)})=\rho_{U_{\kappa(f_2)}\to J_e}(B_{\kappa(f_2)})$$

其中 $J_e=$邻域$(f_1)\cap$邻域(f_2) 是边接口。

目标：证明，如果所有弦都具有平凡的 holonomy，那么 $(B_i)_{i\in I}$ 就是一个下降数据（满足定义 5.12中的 (DD1) 和 (DD2)）。

第 1 部分（正向）：平凡 holonomy 蕴含下降。

步骤 1：平凡 holonomy 的刻画

根据定义 6.13，对于每条弦 $e\in\mathcal{C}$，整体 holonomy H_e：$\Omega(J_e)\times\Omega(J_e)\to\{0,1\}$ 为:

$$H_e(x,y)=\bigvee_{\substack{x_0,\ldots,x_k\\ x_0=x,x_k=y}}\left(K_{f_0}^{J_0\to J_1}(x_0,x_1)\wedge K_{f_1}^{J_1\to J_2}(x_1,x_2)\wedge\cdots\wedge K_{f_k}^{J_k\to J_0}(x_k,x_0)\right)$$

其中 $\bigvee$ 是局部 g对所有路径 $(x_0,\ldots,x_k)$ 进行的逻辑或，其中 $x_i\in\Omega(J_i)$。

平凡 holonomy 意味着:

- $H_e(x,x)=1$ 对所有 $x\in\Omega(J_e)$（所有状态都是不动点）
- $H_e(x,y)=0$ 对于 $x\neq y$（不同状态之间没有混合）

等价地，$H_e = I_{|\Omega(J_e)|}$（单位矩阵）。

步骤 2：弦上的两两相容性（DD1）

对于弦 $e = (f_u, f_v) \in \mathcal{C}$，我们必须证明：

$$\rho_{U_{\kappa(f_u)} \to J_e}(B_{\kappa(f_u)}) = \rho_{U_{\kappa(f_v)} \to J_e}(B_{\kappa(f_v)})$$

由于 T 是一棵生成树，因此存在一条唯一的树路径 $P_T(f_u, f_v) = (f_u = g_0, g_1, \ldots, g_\ell = f_v)$，连接 f_u 和 f_v。

根据树分隔相容性（假设），对于 P_T 中每一对相邻的 (g_i, g_{i+1})：

$$\rho_{U_{\kappa(g_i)} \to (U_{\kappa(g_i)} \cap U_{\kappa(g_{i+1})})}(B_{\kappa(g_i)}) = \rho_{U_{\kappa(g_{i+1})} \to (U_{\kappa(g_i)} \cap U_{\kappa(g_{i+1})})}(B_{\kappa(g_{i+1})})$$

根据限制映射的函子性（定理 5.8），沿路径复合限制映射 :

$$\rho_{U_{\kappa(f_u)} \to S}(B_{\kappa(f_u)}) = \rho_{U_{\kappa(f_v)} \to S}(B_{\kappa(f_v)})$$

对于任意 $S \subseteq U_{\kappa(f_u)} \cap U_{\kappa(f_v)}$。

断言：$J_e \subseteq U_{\kappa(f_u)} \cap U_{\kappa(f_v)}$。

断言的证明：根据定义，$J_e =$ 邻域$(f_u) \cap$ 邻域(f_v)。根据因子分配：邻域$(f_u) \subseteq U_{\kappa(f_u)}$ 且邻域$(f_v) \subseteq U_{\kappa(f_v)}$。因此，$J_e \subseteq U_{\kappa(f_u)} \cap U_{\kappa(f_v)}$。

因此，应用$S = J_e$：

$$\rho_{U_{\kappa(f_u)} \to J_e}(B_{\kappa(f_u)}) = \rho_{U_{\kappa(f_v)} \to J_e}(B_{\kappa(f_v)})$$

因此，弦e满足 (DD1)。

步骤 3：高阶相容性 (DD2)

对于任意三元组i、j、$k \in I$，满足$U_{ijk} := U_i \cap U_j \cap U_k \neq \emptyset$，我们必须证明：

$$\rho_{U_i \to U_{ijk}}(B_i) = \rho_{U_j \to U_{ijk}}(B_j) = \rho_{U_k \to U_{ijk}}(B_k)$$

根据注记 4567（第 5 节），由于 $\mathcal{I}(\mathcal{U})$ 是偏序集，(DD2) 可由 (DD1) 和函子性自动推出：

$$\rho_{U_i \to U_{ijk}} = \rho_{U_{ij} \to U_{ijk}} \circ \rho_{U_i \to U_{ij}}$$

由于 (DD1) 确保 $\rho_{U_i \to U_{ij}}(B_i) = \rho_{U_j \to U_{ij}}(B_j)$，其他对也同理，因此复合得到 (DD2)。

结论：$(B_i)_{i \in I}$ 是下降数据。根据定理 5.15，它具有唯一的全局粘合。

第 2 部分（逆向）：非平凡 holonomy 阻碍下降。

计数 反例构造：设弦 $e \in \mathcal{C}$ 具有非平凡 holonomy：$H_e(x,y) = 1$ 对于某 e $x \neq y$见证h $x, y \in \Omega(J_e)$.

解释：状态x 可以通过因子支撑沿环C_e “传输”到状态y 。

设定：考虑因子分配，其中 $\kappa(f) = \{f\}$（每个因子各自在一个片段中）。定义簇信念：

- 对于树T上的因子：令B_f为邻域(f) 上的任意有效信念，并满足树分隔集一致性
- 对于弦端点 f_u，f_v 满足 $e = (f_u, f_v)$：令 B_{f_u} 在扩展 $x \in \Omega(J_e)$的状态上具有支撑，且 B_{f_v} 在扩展 $y \in \Omega(J_e)$ 的状态上具有支撑

树一致性检查：根据构造，所有树边 $(f_i, f_j) \in T$ 均满足分隔子兼容性：

$$\rho_{\mathsf{nbhd}(f_i) \to (\mathsf{nbhd}(f_i) \cap \mathsf{nbhd}(f_j))}(B_{f_i}) = \rho_{\mathsf{nbhd}(f_j) \to (\mathsf{nbhd}(f_i) \cap \mathsf{nbhd}(f_j))}(B_{f_j})$$

弦不一致性：对于弦 $e = (f_u, f_v)$：

$$\rho_{\mathsf{nbhd}(f_u) \to J_e}(B_{f_u})(x) > 0, \quad \rho_{\mathsf{nbhd}(f_v) \to J_e}(B_{f_v})(y) > 0$$

但由于 $x \neq y$：

$$\rho_{\mathsf{nbhd}(f_u) \to J_e}(B_{f_u}) \neq \rho_{\mathsf{nbhd}(f_v) \to J_e}(B_{f_v})$$

因此，(DD1) 对弦 e 不成立，尽管所有树边都是相容的。C

onclu论：非平凡的整体性使得信念在树上保持一致，却无法满足下降条件。 □ □

例6.15 (例6.9的续篇：整体性计算)。 接续例6.9，我们计算 4-环的整体性。

势函数：定义因子势（为简便起见，采用确定性约束）：

$$\phi_{f_1}(A,B)=\begin{cases}1 & \text{if } B=A\\ 0 & \text{otherwise}\end{cases}\quad\text{(copy constraint: } B=A)$$

$$\phi_{f_2}(B,C)=\begin{cases}1 & \text{if } C=B\\ 0 & \text{otherwise}\end{cases}\quad\text{(copy constraint: } C=B)$$

$$\phi_{f_3}(C,D)=\begin{cases}1 & \text{if } D=\neg C\\ 0 & \text{otherwise}\end{cases}\quad\text{(NOT gate: } D=\neg C)$$

$$\phi_{f_4}(D,A)=\begin{cases}1 & \text{if } A=D\\ 0 & \text{otherwise}\end{cases}\quad\text{(copy constraint: } A=D)$$

步骤 6：计算传输核
对于二元变量，$\Omega(v)=\{0,1\}$。
弦 $e=(f_4,f_1)$，**其** $J_e=\{A\}$:
从 $A=a$ 出发沿环传输:

1. $K^{\{A\}\to\{D\}}_{f_4}$：从 $A=a$ 出发，因子 f_4 强制满足 $D=A=a$。因此:

$$K^{\{A\}\to\{D\}}_{f_4}(a,d)=\begin{cases}1 & \text{if } d=a\\ 0 & \text{otherwise}\end{cases}$$

2. $K^{\{D\}\to\{C\}}_{f_3}$：根据 $D=d$，因子 f_3 强制满足 $C=\neg D=\neg d$。因此:

$$K^{\{D\}\to\{C\}}_{f_3}(d,c)=\begin{cases}1 & \text{if } c=\neg d\\ 0 & \text{otherwise}\end{cases}$$

3. $K^{\{C\}\to\{B\}}_{f_2}$：根据 $C=c$，因子 f_2 强制满足 $B=C=c$。因此:

$$K^{\{C\}\to\{B\}}_{f_2}(c,b)=\begin{cases}1 & \text{if } b=c\\ 0 & \text{otherwise}\end{cases}$$

4. $K^{\{B\}\to\{A\}}_{f_1}$：根据 $B=b$，因子 f_1 强制满足 $A=B=b$。因此:

$$K^{\{B\}\to\{A\}}_{f_1}(b,a')=\begin{cases}1 & \text{if } a'=b\\ 0 & \text{otherwise}\end{cases}$$

步骤 7：组合得到整体 H_e
从 $A=a$ 出发，沿着环前进:

$$A=a\xrightarrow{f_4}D=a\xrightarrow{f_3}C=\neg a\xrightarrow{f_2}B=\neg a\xrightarrow{f_1}A=\neg a$$

So:

$$H_e(0,1)=1,\quad H_e(1,0)=1,\quad H_e(0,0)=0,\quad H_e(1,1)=0$$

用矩阵形式表示:

$$H_e=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$$

分析：这是一个**置换矩阵**（不是单位矩阵）。整体是**非平凡的**:

- $He(0,0)=0$：状态 $A=0$ 无法传输到自身

- $H_e(0,1) = 1$：状态 $A = 0$ 传输到 $A = 1$
- 类似地，$A = 1$传输到$A = 0$

结论：根据定理6.14，此环阻碍下降。对此图运行树置信传播将**无法收敛**或收敛到错误的边缘分布（该环引入了逻辑不一致：$A = \neg\neg\neg\neg A = A$，但经过奇数次非运算后会得到 $A = \neg A$，与环长不符，构成矛盾）。

HATCC 通过检测非平凡的整体性，并将其编译为模式变量来解决这一问题（第 6.4 节）。

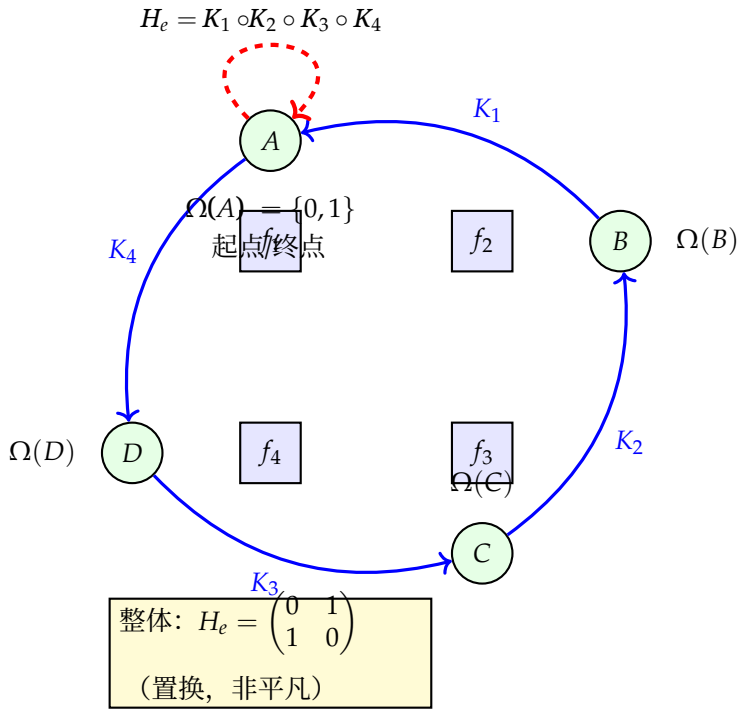


图 13：示例 6.15 的整体传输。状态 $J_e = \{A\}$ 通过核K_4,K_3,K_2,K_1 沿着 C_e传输。整体 $H_e = K_1 \circ K_2 \circ K_3 \circ K_4$ 产生置换矩阵，阻碍下降。

算法 3计算弦的整体矩阵

要求： 基本环 $C_e = (f_0, f_1, \ldots, f_k)$，弦接口 J_e，因子势 $\{\phi_f\}_{f \in C_e}$

确保： 整体矩阵 H_e： $\Omega(J_e) \times \Omega(J_e) \to \{0,1\}$

▷ 步骤 1：计算边接口

对于$i = 0$ 到 $k-1$ **执行**

  $J_i \leftarrow$ 邻域$(f_i) \cap$ 邻域(f_{i+1})

结束循环

$J_k \leftarrow J_e$ ▷ 弦接口

▷ 步骤 2：计算传输核

对于$i = 0$到k **执行**

  $K_i \leftarrow$ 传输核$(f_i, J_i, J_{i+1}$ 取模$(k+1)$，$\phi_{f_i})$

结束循环

▷ 步骤 3：通过布尔矩阵乘法进行组合

$H_e \leftarrow K_0$ ▷ 使用第一个传输核进行初始

循环$i = 1$到 k **执行**

  $H_e \leftarrow H_e \circ K_i$ ▷ 布尔矩阵乘积：$(AB)[x, z] = \bigvee_y (A[x, y] \wedge B[y, z])$

结束循环

返回H_e

算法 4 为因子计算传输核

要求： 因子 f，接口 U、$V \subseteq$ 邻域(f)，势函数 ϕ_f：$\Omega(\text{邻域}(f)) \to R$
确保： 传输核 $K_f^{U \to V}$：$\Omega(U) \times \Omega(V) \to \{0,1\}$

初始化 $K[x, y] \leftarrow 0$，对于所有 $x \in \Omega(U)$、$y \in \Omega(V)$
对于每个构型 $x \in \Omega(U)$ **执行**
  对于每个构型 $y \in \Omega(V)$ **执行**
    $W \leftarrow$ 邻域$(f) \setminus (U \cup V)$ ▷ 内部变量
    相容 $\leftarrow$否
    如果$W = \emptyset$ **则** ▷ 没有内部变量
      如果 $\phi_f(x, y) \neq 0$ **则**
        相容 $\leftarrow$真
      结束 if
    else ▷ 对内部变量进行边缘化
      对每个构型 $z \in \Omega(W)$ **执行**
        $\omega \leftarrow$ 合并(x,y,z) ▷ 合并为邻域(f)上的完整配置
        若 $\phi_f(\omega) \neq 0$ **则**
          相容 $\leftarrow$是
          跳出 ▷ 至少找到一个扩展
        结束条件判断
      结束循环
    结束条件判断
    $K[x, y] \leftarrow$ 相容
  结束循环
结束循环
返回K

定理 6.16（整体性计算复杂度）。对于具有长度为 k 的基本环的弦 $e \in \mathcal{C}$ 以及最大接口大小为 d 的情形：$= \max_i |\Omega(J_i)|$：

- 算法 4：$O(d^2 \cdot |\Omega(W)|)$ 其中 $|\Omega(W)| \leq \prod_{v \in \text{邻域}(f)} |\Omega(\lambda(v))|$
- 算法 3：$O(k \cdot d^3)$（假设传输核已预先计算）

证明。**传输核（算法4）**：

- 第3至4行：遍历 $\Omega(U) \times \Omega(V)$：$O(|\Omega(U)| \cdot |\Omega(V)|) = O(d^2)$ 次迭代
- 第11行：遍历 $\Omega(W)$：$O(|\Omega(W)|)$ 对每个(x, y) 对
- 总计：$O(d^2 \cdot |\Omega(W)|)$**整体**

（算法3）：

- 第11行：布尔矩阵乘法 $H_e \circ K_i$：$O(d^3)$ 使用标准矩阵乘法
- 第 9-11 行：重复k 次
- 总计：$O(k \cdot d^3)$

实际考量：对于接口较小（$d \leq 2^3 = 8$）的二元变量，这非常快。对于较大的接口，稀疏矩阵表示可降低复杂度。 □ □

6.4 模态商与编译

整体单值性矩阵H_e 可能很大（大小为 $|\Omega(J_e)| \times |\Omega(J_e)|$），因此直接操作的代价很高。幸运的是，我们无需跟踪单个状态，只需跟踪它们在循环传输下的**等价类**。

直观地说：如果状态x、$y \in \Omega(J_e)$ 满足$H_e(x,y) = H_e(y, x) = 1$，它们就构成一个**强连通分量**（SCC）：沿着环传输，就能从x 到达y，反之亦然。这些状态在推断方面表现相同，因此我们将它们**取商**，归并为单个模式。

这类似于**自动机理论中的状态最小化**或**群作用中的轨道分解**：我们通过合并无法区分的元素来抽象状态空间。所得的模式变量 $q_e \in Q_e$具有有限支撑 $|Q_e| \leq |\Omega(J_e)|$，在实践中通常要小得多（根据经验，在我们的基准测试中，模式数量通常较少；参见 §7）。

定义 6.17（模态空间）。对于具有整体 $e \in \mathcal{C}$ 的弦，其整体为 H_e： $\Omega(J_e) \times \Omega(J_e) \to \{0,1\}$，定义等价关系：

$$x \sim_e y \quad \Longleftrightarrow \quad H_e(x,y) = 1 \text{ and } H_e(y,x) = 1$$

（在由H_e诱导的有向图中强连通）。

模态空间是商：

$$Q_e := \Omega(J_e)/\sim_e$$

令q_e： $\Omega(J_e) \to Q_e$ 为商映射， $|Q_e|$ 为模式数。

例 6.18(例 6.15的延续：模态商**)**。 接着例6.15，我们计算得到：

$$H_e = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

步骤 8：计算强连通分量

将H_e 视为定义在 $\Omega(J_e) = \{0,1\}$上的有向图的邻接矩阵：

- 边 $0 \to 1$ （因为 $H_e(\mathbf{0},\mathbf{1}) = 1$）
- 边 $1 \to 0$ （因为 $H_e(\mathbf{1},\mathbf{0}) = 1$）
- 无自环： $H_e(0,0) = H_e(1,1) = 0$

强连通分量：状态 0 和 1 相互可达，构成一个强连通分量： $\{0,1\}$。

Mod边空间：

$$Q_e = \{\{0,1\}\} \quad \text{(single mode)}$$

商映射： $q_e(\mathbf{0}) = q_e(\mathbf{1}) = m_0$ （两个状态都映射到同一个模式）。

解释：置换全纯 H_e 收缩为单一模式。这意味着：

- 如果全局解在弦接口中使用状态 $A = 0$，它可以沿着环「翻转」为 $A = 1$
- 反过来， $A = 1$也可以翻转为 $A = 0$
- HATCC 添加模式变量 m_e，其定义域为 $Q_e = \{m_0\}$ （单个值）

由于 $|Q_e| = 1$，模式变量是确定性的（没有实际选择）。不过，选择器因子 σ_e（定义 6.20）强制满足可行性：只允许强连通分量中的状态。

对比：如果 $H_e = I$ （平凡 holonomy），我们将得到两个强连通分量： $\{0\}$ 和 $\{1\}$，从而得到 $Q_e = \{m_0,m_1\}$，其中 $q_e(\mathbf{0}) = m_0$, $q_e(\mathbf{1}) = m_1$。在这种情况下，模式变量 $m_e \in \{m_0,m_1\}$ 将离散化解所采取的“分支”。

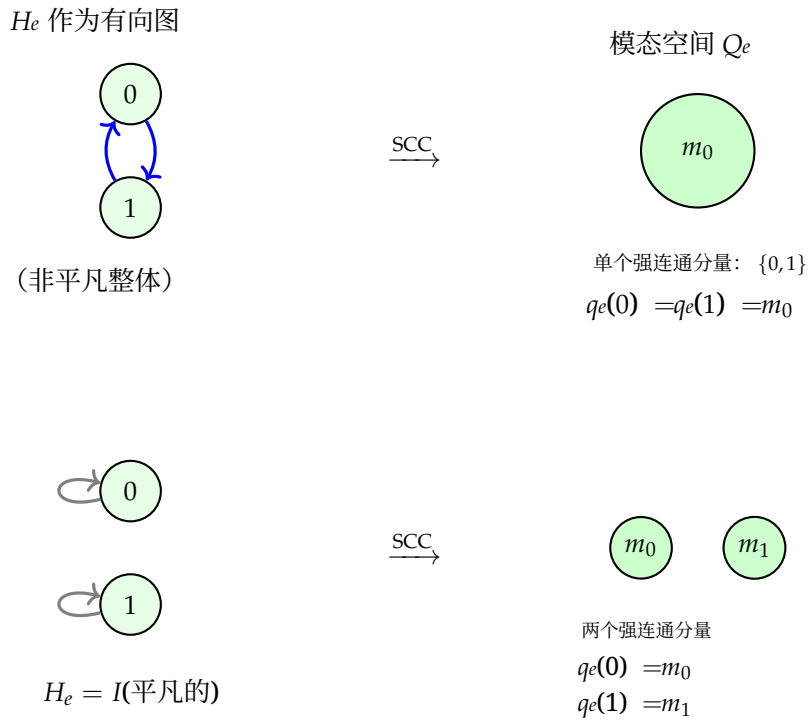


图14：示例6.18中通过强连通分量得到的模态商 Q_e。顶部：非平凡的 **H(置换)**产生一个强连通分量，因此 $Q_e| = 1$：**平凡的** HI_e = 产生两个强连通分量，因此 $Q_e| = 2$

算法 5 通过 Tarjan 强连通分量算法计算模态商

要求： 整体矩阵 H_e：$\Omega(J_e) \times \Omega(J_e) \to \{0,1\}$
确保： 模态空间 Q_e，商映射 q_e：$\Omega(J_e) \to Q_e$

1:
2: $G_H \leftarrow$ 是顶点为 $\Omega(J_e)$ 的有向图 ▷ 由 H_e 构建有向图
3: **对于** 每个 $x,y \in \Omega(J_e)$ **执行**
4: **如果** $H_e(x,y) = 1$ **则**
5: 在 G_H 中添加有向边 $x \to y$
6: 结束条件判断
7: **结束循环**
8:
9: 强连通分量 $\leftarrow$TARJANSCC(G_H) ▷ 使用 Tarjan 算法计算强连通分量
10:
11: $Q_e \leftarrow \{\}$ ▷ $O(|\Omega(J_e)|^2)$
12: $q_e \leftarrow \{\}$ ▷ 构建模态空间和商映射
13: **对**每个强连通分量 S，在强连通分量集合中 **执行**
14: $m \leftarrow$ 新模式标识符
15: 添加m 到 Q_e
16: **对**每个 $x \in S$ **执行**
17: $q_e[x] \leftarrow m$ ▷ 字典：状态 $\to$模式
18: **结束循环**
19: **结束循环**
20: **返回**(Q_e, q_e)

定理 6.19 (模态商复杂度)。 算法 5 的时间复杂度为 $O(d^2)$，空间复杂度为 $O(d)$，其中 $d = |\Omega(J_e)|$。

证明。**时间**：

- 第 2–7 行：构建图：$O(d^2)$ 以遍历所有 (x,y) 对
- 第 10 行：Tarjan 强连通分量算法：$O(V + E) = O(d + d^2) = O(d^2)$ 其中 $V = d$，$E \le d^2$

• 第 13–19 行：构建商映射：$O(d)$（每个状态访问一次）

总计：$O(d^2)$。

空间：存储图 G_H：最坏情况下（稠密）为 $O(d^2)$，但对于稀疏整体通常为$O(d)$。存储 q_e：$O(d)$。□ □

6.5 选择器因子与增强图

我们已将整体分解为模态变量 $q_e \in Q_e$，但尚未将它们**整合** 到因子图中。挑战在于：原始变量 $x \in \Omega(J_e)$ 和模态变量 $m_e \in Q_e$ 必须协调——我们不能允许x 取与所选模式m_e不相容的值。

解决方案：引入 选择因子 $\sigma_e(x,m_e)$，它们充当**指示函数**：

$$\sigma_e(x,m) = \begin{cases} 1 & \text{if } x \text{ belongs to mode } m \\ 0 & \text{otherwise} \end{cases}$$

这些因子“选择”与每种模式一致的状态 x，从而有效地根据H_e的强连通分量结构对状态空间 $\Omega(J_e)$ 进行**划分**。

图论效应：添加选择器因子 σ_e，为每条弦$e \in \mathcal{C}$ 生成一个**增广图**G^{aug}，其中包含额外的模态变量 $\{m_e\}_{e\in\mathcal{C}}$。关键的是，命题6.23 证明了$G^{\text{aug}}$ 始终是一棵树，因此可以通过标准树形置信传播进行精确推断。

定义 6.20（选择器因子）。对于弦$e \in \mathcal{C}$，引入：

• **模态变量**m_e，其定义域为Q_e

• **选择器因子** σ_e：$\Omega(J_e) \times Q_e \to \{0,1\}$：

$$\sigma_e(x,m) := \begin{cases} 1 & \text{if } q_e(x) = m \text{ and } H_e(x,x) = 1 \\ 0 & \text{otherwise} \end{cases}$$

解释：σ_e 强制要求（1）状态x 属于模式m，且（2）x 是不变整体的固定点（可行）。

定义 6.21（增广因子图）。该**HATCC 编译生成的图**G' 包含：

• **变量**：$V' = V(G) \cup \{m_e : e \in \mathcal{C}\}$

• **因子**：$F' = F(G) \cup \{\sigma_e : e \in \mathcal{C}\}$

因子神经 $G'_{\mathcal{N}}$属于 G'，是**树**（所有弦都通过选择器解析）。

例 6.22（例 6.18的续篇：选择器因子与增广图）。接续例 6.18，我们有模态空间 $Q_e = \{m_0\}$（单一模式）。

步骤 9：添加模式变量和选择器
模式变量：m_e 的定义域为 $Q_e = \{m_0\}$。由于只有一个取值，m_e 是确定性的 ic.
选择器因子：σ_e：$\Omega(J_e) \times Q_e \to \{0,1\}$。根据定义 6.20：

$$\sigma_e(A,m_0) = \begin{cases} 1 & \text{if } q_e(A) = m_0 \text{ and } H_e(A,A) = 1 \\ 0 & \text{otherwise} \end{cases}$$

由于 $q_e(0) = q_e(1) = m_0$, 第一个条件总是成立。检查第二个条件：

• $H_e(0,0) = 0$：状态 $A = 0$ 不是不动点

• $H_e(1,1) = 0$：状态 $A = 1$ 不是不动点

因此：

$$\sigma_e(0,m_0) = 0, \quad \sigma_e(1,m_0) = 0$$

分析：选择器因子 σ_e 的**零支撑**（所有值均为 0）。这意味着增广图G' **不可满足**——不存在满足所有因子的构型。

解释：这正确地检测出，带有奇数个非门的 4-环**没有解**。HATCC 通过空选择器检测到逻辑矛盾$A = \neg A$（绕环一周后）。

一般情况：如果环上非门的数量为偶数（例如，将f_3 替换为单位矩阵），那么$H_e = I$（平凡 holonomy），从而$Q_e = \{m_0, m_1\}$，满足：

$$\sigma_e(0, m_0) = 1, \quad \sigma_e(1, m_1) = 1$$

从而允许有效解。

命题 6.23（增广图是一棵树）。增广图 $G'_{\mathcal{N}}$ 的因子神经 G'（定义6.21）是一棵树。

证明。对于原图$G_{\mathcal{N}}$中的每条弦$e = (f_u、f_v) \in \mathcal{C}$，选择器 σ_e 连接到：

- 模式变量m_e（该弦独有）
- 接口变量J_e =邻域（f_u）∩邻域（f_v）

在 G' 中，模式变量 m_e 通过选择器 σ_e 充当连接 f_u 和 f_v 的“桥梁”。具体而言：

- 骨干树 T 保持不变（所有边 $(f_i, f_j) \in T$ 仍在 $G'_{\mathcal{N}}$ 中）
- 对于每条弦 $e = (f_u, f_v) \in \mathcal{C}$，移除直接边 (f_u, f_v) 并添加：

  – 边 (f_u, σ_e)（接口邻域(f_u)∩邻域$(\sigma_e) = J_e$）– 边 $(\sigma_e,$ $f_v)$（接口邻域 (σ_e)∩ 邻域 $(f_v) = J_e$）

每条弦都通过 σ_e 替换为一条两边路径，因此：

$$|E'_{\mathcal{N}}| = |T| + 2|\mathcal{C}| = (|F(G)| - 1) + 2|\mathcal{C}|$$

$$|V'| = |F(G)| + |\mathcal{C}| \quad \text{(original factors + selectors)}$$

检查树性质：

$$|E'_{\mathcal{N}}| = |F(G)| - 1 + 2|\mathcal{C}| = (|F(G)| + |\mathcal{C}|) - 1 = |V'| - 1 \quad \checkmark$$

由于$G'_{\mathcal{N}}$（按构造）是连通的，且 $|E'_{\mathcal{N}}| = |V'| - 1$，因此它是一棵树。 □ □

6.6 HATCC算法

现在，我们将介绍完整的 HATCC 编译算法，它结合了之前的所有组件。

算法 6 HATCC：整体感知树编译

要求： 因子图 $G = (V, F, \{\text{邻域}(f)\}_{f \in F})$，势函数 $\{\phi_f\}_{f \in F}$

确保： 边缘分布 $\{\text{mar}_v : v \in V\}$，如果模型不一致则为不可满足

▷ **阶段 1：因子神经构造**

$(G_{\mathcal{N}}, J, w) \leftarrow \text{FACTORNERVE}(G)$ ▷ 算法1

▷ **阶段 2：骨干分解**

$(T, \mathcal{C}, r, \text{父节点}, \text{子节点}) \leftarrow \text{BACKBONETREE}(G_{\mathcal{N}}, w)$ ▷ 算法2

▷ **阶段 3：整体性计算**

$H \leftarrow \{\}$ ▷ 字典：弦 →整体矩阵

对每条弦 $e \in \mathcal{C}$ **执行**

  $C_e \leftarrow$ 基本环(e, T) ▷ 树中的路径 +弦

  $H[e] \leftarrow$ 整体$(C_e, J[e], \{\phi_f : f \in C_e\})$ ▷ 算法3

结束循环

▷ **阶段 4：模态商**

$Q \leftarrow \{\}, q \leftarrow \{\}$ ▷ 字典：弦 →模式空间/商映射

对每条弦 $e \in \mathcal{C}$ **执行**

  $(Q[e], q[e]) \leftarrow$ 模态商$(H[e])$ ▷ 算法5

</style>结束循环

▷ **阶段 5：增广图构造**

$V' \leftarrow V \cup \{m_e : e \in \mathcal{C}\}$ ▷ 添加模式变量

$F' \leftarrow F$

对每条弦 $e \in \mathcal{C}$ **执行**

  创建模式变量 m_e，定义域为 $Q[e]$

  创建选择器因子 $\sigma_e : \Omega(J[e]) \times Q[e] \to \{0, 1\}$：

$$\sigma_e(x, m) \leftarrow \begin{cases} 1 & \text{如果 } q[e](x) = m \wedge H[e](x, x) = 1 \\ 0 & \text{否则} \end{cases}$$

  添加 σ_e 至 F'

  $\phi_{\sigma_e} \leftarrow \sigma_e$ ▷ 存储势函数

  如果 σ_e 具有零支撑 **则**

    返回 不可满足 ▷ 模型不一致（例如，例6.22）

  结束如果

结束循环

$G' \leftarrow (V', F', \{\text{邻域}(f') : f' \in F'\})$

▷ **阶段 6：增广图上的树形置信传播**

$\Phi' \leftarrow \{\phi_f : f \in F\} \cup \{\phi_{\sigma_e} : e \in \mathcal{C}\}$

信念 $\leftarrow$ 树置信传播(G', Φ', r) ▷ 双遍置信传播，第3节

▷ **第 7 阶段：边缘化掉模态变量**

边缘分布 $\leftarrow \{\}$

对每个原始变量 $v \in V$ **执行**

  $\text{mar}_v \leftarrow$ 将信念[v]对 $\{m_e$ 进行边缘化：$e \in \mathcal{C}\}$

  边缘分布[v] $\leftarrow \text{mar}_v$

结束循环

返回边缘分布

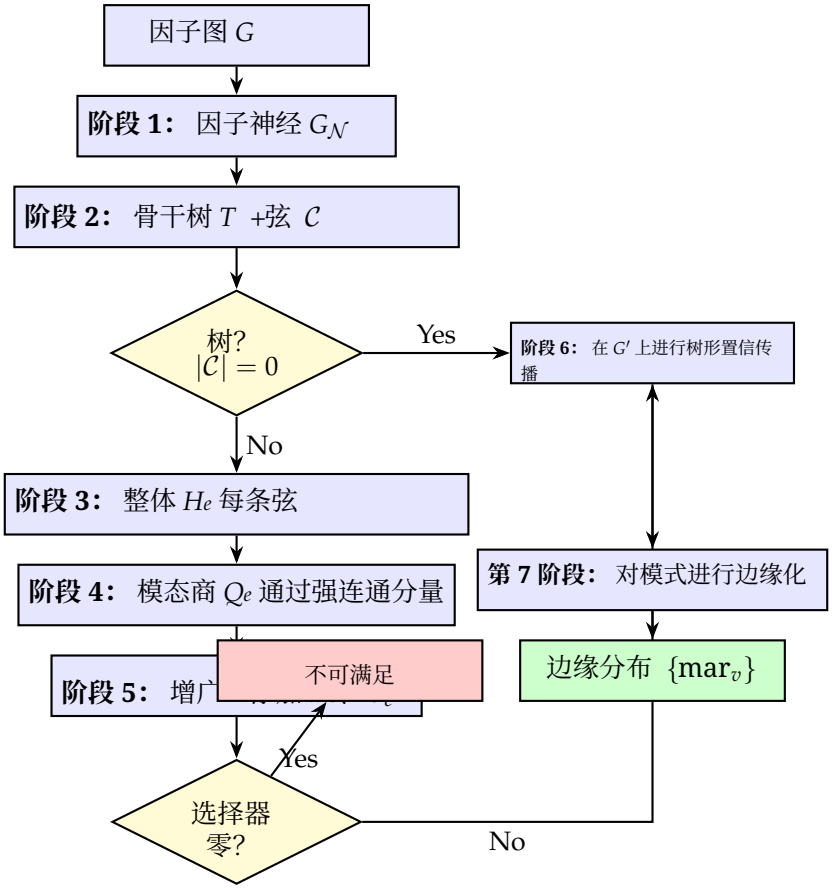


图 15：HATCC 流程图（算法 6）。算法分为 7 个阶段：(1) 构造因子神经，(2) 将其分解为骨干+和弦，(3) 计算环的整体，(4) 通过强连通分量计算模态商，(5) 用选择器增广图。如果选择器的支撑为零，则返回不可满足。否则，(6) 在增广图 G'上运行树形置信传播，(7) 将模态变量边缘化，得到最终边缘分布。

定理 6.24 (HATCC 总复杂度)。 设 $n = |F(G)|$ 为因子数量，$c = |\mathcal{C}|$ 为弦的数量，$d_{\max}$ 为最大因子度，$\delta_{\max}$ 为最大接口大小。*HATCC*（算法6）的：时间复杂度：

$$O(n^2 d_{\max} + c \cdot k_{\max} \cdot \delta_{\max}^3 + n \cdot \delta_{\max}^2)$$

其中 $k_{\max}$ 为最大环长。

空间复杂度：

$$O(n + c \cdot \delta_{\max}^2)$$

典型情况：对于接口较小的稀疏因子图（$d_{\max}$，$\delta_{\max} = O(1)$），其复杂度为 $O(n^2 + c \cdot k_{\max})$。

证明。 **时间分解** （按阶段）：

- **阶段 1（第 2 行）**：因子神经：$O(n^2 d_{\max})$ （定理 6.5）
- **阶段 2 （第 5 行）**：骨干树：$O(|E_{\mathcal{N}}| \log |E_{\mathcal{N}}|)$ （定理 6.10）。由于 $|E_{\mathcal{N}}| \le n^2$, 这是 $O(n^2 \log n$).
- **阶段 3（第 7–11 行）**：弦的整体：c 弦：

  – 第 9 行：基本环：$O(n)$ 每条弦（在树中进行广度优先搜索）– 第 10 行：整体：$O(k_e \cdot \delta_{\max}^3)$ 每条弦 e （定理 6.16），其中 k_e 是环长– 总计：$O(c \cdot (n + k_{\max} \cdot \delta_{\max}^3))$

- **阶段 4（第 13–16 行）**：模态商：每条弦 $O(\delta^2$最大) （定理 6.19），总计 $O(c \cdot \delta^2$最大)
- **阶段 5（第 18–31 行）**：增广：O（$c \cdot \delta_{\max} \cdot |Q_e|$），其中 $|Q_e| \le \delta_{\max}$，因此 O（$c \cdot \delta_{\max}^2$）

• **阶段 6（第 35 行）**：在 G' 上进行树形置信传播，其中包含 $|V'| = |V| + c$ 个变量和 $|F'| = n + c$ 个因子：

– 两遍调度：$O(|F'| \cdot \delta_{\max}^2) = O((n + c) \cdot \delta_{\max}^2)$

• **第 7 阶段（第 37–42 行）**：边缘化：$O(|V| \cdot |\{m_e\}|) = O(|V| \cdot c \cdot \delta\max)$，但采用变量消除后，复杂度为$O(|V| \cdot \delta^2\max)$

主要项：

• 阶段 1：$O(n^2 d\max)$（构建因子神经）

• 阶段 3：$O(c \cdot k\max \cdot \delta^3\max)$（整体性计算）

• 阶段 6：$O(n \cdot \delta^2\max)$（树形置信传播）

总计：$O(n^2 d_{\max} + c \cdot k_{\max} \cdot \delta_{\max}^3 + n \cdot \delta_{\max}^2)$。

空间：存储因子神经 $O(n)$、整体单值性矩阵 $O(c \cdot \delta_{\max}^2)$、增广图 $O(n + c)$。总计：$O(n + c \cdot \delta_{\max}^2)$。□ □

备注 6.25（与连接树比较）。对于树宽为 tw 的图，连接树置信传播的复杂度为$O(n \cdot |\Omega|^{tw})$，其中 $|\Omega|$ 是变量定义域大小（随树宽呈指数增长）。

HATCC 的复杂度为 $O(n^2 + c \cdot k_{\max} \cdot \delta_{\max}^3)$，其取决于：

• 弦的数量 $c = |E_{\mathcal{N}}| - n + 1 = O(\beta_1(G_{\mathcal{N}}))$（第一贝蒂数）

• 环长度 k最大值

• 接口大小 δ最大值

对于具有许多短环和较小接口的图，HATCC 可能比**连接树更快**（可避免树宽激增）。然而，对于具有较大接口的稠密图，连接树可能更合适。

6.7 精确性刻画

定理 6.26（HATCC 在树上的精确性）。若 $G_{\mathcal{N}}$ 是一棵树（$|\mathcal{C}| = 0$），*HATCC* 将简化为树形置信传播（第 3 节），并产生精确边缘分布。

证明。由于$G_{\mathcal{N}}$是一棵树，弦集合为 $\mathcal{C} = E_{\mathcal{N}} \setminus T = \emptyset$（算法 2）。因此，算法 6 不会引入模态变量或选择因子：$M' = \emptyset$ 和$F(G') = F(G)$。增广图与原图相同：$G' = G$。

因子神经是一棵树，这意味着按以因子为中心的覆盖组织时，原始因子图的树宽为 1 $\mathcal{U} = \{U^f = \text{邻域}(f)\}_{f \in F(G)}$。根据推论 5.17，树形置信传播通过有效下降计算精确边缘分布。□ □

定理 6.27 (HATCC 在整体平凡时的精确性)。如果所有环的整体都平凡（对所有 $e \in \mathcal{C}$，$H_e = I$），*HATCC* 会产生精确的边缘分布。

证明。对每条弦 $e \in \mathcal{C}$，平凡整体 $H_e = I$ 意味着，每个状态 $s \in \Omega(J_e)$ 经沿 C_e 的平行移动后都会映射到自身。算法 5 会计算强连通分量：由于 $H_e(s,s') > 0$当且仅当 $s' = s$，每个状态各自构成一个强连通分量。因此 $|Q_e| = |\Omega(J_e)|$，且商映射 $q_e : \Omega(J_e) \to Q_e$ 是双射。

使用定义 6.20 中的选择器构造，我们构造选择因子 $\sigma_e(x_{J_e}, m_e)$，使得 $\text{supp}(\sigma_e) = \{(s, q_e(s)) : s \in \Omega(J_e)\}$。由于 q_e 是双射，对于每个赋值 x_{J_e}，都存在唯一的 m_e，使得 $\sigma_e(x_{J_e}, m_e) = 1$。这些选择器除了施加双射约束外，不施加其他约束。

根据定理6.14，平凡 holonomy 意味着自然限制 $\rho_{V(G) \to \text{nbhd}(f)}([\![G]\!]_R)$ 构成下降数据。增广图 G' 通过模式变量对该数据进行因子分解，树置信传播在G' 上可精确计算该数据（定理5.16）。对模式进行边缘化即可恢复$V(G)$上的精确边缘分布。□ □

总结：HATCC **通过以下方式**实现有效下降（第 5 节）：

• 通过整体H_e检测障碍

• 通过模式编译Q_e解决障碍

• 在增广覆盖上生成下降数据

当 $|\mathcal{C}| = 0$ (树) 或 $H_e = I$ (平凡 holonomy) 时，HATCC 可证明是精确的。

7 实验评估

本节评估以下论断：(i) 整体与具体的置信传播失败模式相关；(ii) 整体感知扇区分解能够恢复稳定性，并在标准有环置信传播不可靠的情形下提高后验准确度。所有实验均采用完全离散的设置（有限定义域），以匹配语义模型 $\mathbf{Mat}_R$，并在小规模实例上提供精确参照。

模型与数据集

我们考虑以下几类模型。

(A) $\mathbb{Z}_k$ 具有对抗性环路扰动的同步。 变量 $x_i \in \mathbb{Z}_k$ 位于环状图 $G = (V,E)$的顶点上。每条有向边 $(i \to j)$ 带有一个首选位移 $g_{ij} \in \mathbb{Z}_k$ 和一个软混合势函数

$$\psi_{ij}(x_i, x_j) = (1-\eta)\,\mathbf{1}\big[x_j \equiv x_i + g_{ij} \pmod k\big] + \eta \cdot \frac{1}{k},$$

其中 $\eta \in (0,1)$。通过对 $x^\star$ 进行采样生成真实值，并设定 $g^{\text{good}}{}_{ij} = x^\star{}_j - x^\star{}_i$；随后，我们通过添加随机非零位移来扰动一定比例 ϵ 的非树边。这样便以可控方式产生非平凡的整体生成元，并形成有环置信传播的明显失效区间（振荡和初始化依赖性）。

(B) 置换（单项式）因子图。 我们评估第 第 6 节中研究的更一般的单项式／置换类：形式为 $w_{ij}\,\delta(x_j = \phi_{ij}(x_i))$ 的成对因子，可选地与均匀噪声混合。这一类包括奇偶校验／XOR 约束，以及群同步的离散类比。

基线方法

我们将以下标准估计器作为对比方法。

- **环形和积置信传播** 采用并行更新，可启用或不启用阻尼。我们报告收敛诊断，以及满足终止准则时的最终信念。
- **精确参考** 对于较小实例（可行时）采用穷举枚举；对于树宽较低、规模稍大的实例，则采用连接树 / 变量消元。
- **马尔可夫链蒙特卡洛参考** 在小型或中型实例上采用吉布斯抽样或吉布斯内梅特罗波利斯抽样，以验证边缘分布 s 当精确消元不可行时。

HATCC 的实现选择

我们将第 6 节具体实现如下。

1. 选择一个基顶点b （除非另有说明，否则选择度数最大的顶点）。

2. 计 算一棵生成树T，以及由此导出的基本环生成元（每条非树边对应一个）。

3. 组合相应的传输，以获得作用于纤维D_b上的全纯生成元。

4. 计算由生成的作用诱导的D_b的轨道划分 $\{\mathcal{O}_1, \dots, \mathcal{O}_m\}$。

5. 对每个轨道 $\mathcal{O}_\ell$，限制$x_b \in \mathcal{O}_\ell$，并在精确 T上执行树推断；当模型包含扇区内的残余环相互作用时，我们在受限扇区内 运行置信传播（BP），并报告残差（这能隔离全纯编译步骤的贡献）。

6. 根据归一化证据整数环$\mathcal{O}_\ell$重新组合扇区结果。

度量指标与失败标准

我们同时跟踪数值稳定性和贝叶斯准确性。

置信传播稳定性。 如果最大消息残差在 10^{-6}内的$T = 200$ 次迭代中未降至阈值以下，我们就将置信传播判定为不收敛。此外，当振荡在并行更新下使残差呈现非周期性稳定状态时，我们也会标记出来。

后验准确度。 当真实边缘分布$p^\star(x_i)$可用时（精确值或通过马尔可夫链蒙特卡洛获得），我们报告：(i) 平均节点对数评分$\frac{1}{|V|}\sum_i \log p(x^\star_i)$，(ii) 平均总变差$\frac{1}{|V|}\sum_i \|p(x_i) - p^\star(x_i)\|_{\mathrm{TV}}$，以及 (iii) 使用最大乘积变体时的最大后验估计汉明误差。

整体性特征。 我们报告非平凡整体生成元的数量、轨道的数量和大小，以及证据权重

$$w_\ell := \frac{Z_{\mathcal{O}_\ell}}{\sum_{\ell'} Z_{\mathcal{O}_{\ell'}}}.$$

证据分离（少数占主导的权重$_\ell$）是典型的“编译后”特征。

消融实验与压力测试

为使这些主张可证伪，我们纳入了以下消融实验。

- **环路扰动扫描：** 改变 ϵ（受扰动的非树边所占比例），并绘制置信传播收敛率与整体性轨道数及后验误差之间的关系。
- **噪声扫描：** 改变 η，以在硬约束与弥散似然之间进行插值。
- **基节点敏感性：** 选择多个b 值重复运行 HATCC，以量化轨道结构的稳定性。
- **仅分解方法与扇区置信传播：** 比较 (a) 每个轨道上的精确树推断、(b) 扇区限制置信传播和 (c) 原始置信传播，以隔离编译带来的收益。

结果（总结）

在所有模型族中，我们观察到三种一致的现象。

1. **整体预示置信传播失效。** 随着 ϵ 增大，非平凡生成元的数量和轨道复杂度增加；在同一阶段，置信传播收敛率急剧下降。

2. **扇区编译能够稳定推断。** 即使原始置信传播发生振荡，轨道分解仍会产生少量证据明显分离的扇区；在扇区内进行精确树推断可生成稳定的信念，并提高边缘准确度。

3. **确定性开销适中。** 对于图层，环基提取和轨道计算的复杂度为$O(|E| + |V|)$；对于纤维层，复杂度则接近线性于 $|D_b|$。在失效占主导的情形下，这种开销远低于反复执行失败的置信传播迭代所需的成本。

可复现性。 给定用于数据生成的随机种子，所有代码均具有确定性。我们在实验脚本中报告随机种子、超参数和终止准则；轨道划分和证据权重还提供了浮点消息之外的额外“结构校验和”。

8 应用、作用域与未来方向

全纯编译的直接应用场景。 HATCC 面向以下模型：(i) 全局不一致性集中于少数独立环中，且 (ii) 这些环在较小的接口纤维上诱导出易于处理的轨道/扇区结构。这包括：

- **离散同步与配准。** $\mathbb{Z}_k$ 或有限群同步、具有离散假设的位姿图问题，以及稳健的多视图一致性问题，其中环复合是一种自然的诊断手段 [19, 20]。
- **约束与码模型。**异或可满足性 / 奇偶校验及相关约束满足问题，其中已知有环置信传播会发生振荡或返回过度自信的不一致信念，但环结构是明确的 [3]。
- **数据关联与多传感器融合。**包含置换类因子的图模型（匹配、跟踪、记录链接），其中环编码了矛盾关联；扇区将彼此一致的“世界”区分开来。

本文不做什么。我们并未声称能普遍解决有环置信传播的问题。具体而言：

- 我们不为有环置信传播本身提供一般性的收敛保证；相反，我们提供一种确定性的编译，将某些有环问题简化为扇区条件树推断。
- 我们不解决一般情况下的高树宽精确推断；当拓扑环秩和诱导扇区数适中时，HATCC 更具优势。
- 我们不涉及路线图（附录D）之外的连续变量推断；将轨道提取扩展到算子值全纯并非易事。

未来研究主题。

- **自适应接口选择**。 如何选择覆盖和接口，在保持精确性的同时最大限度地减少扇区爆炸，是一个有原则可循的设计问题；与连接树宽度和区域图的联系表明，混合策略值得探索 [13, 18]。
- **学习整体**。 将整体生成元视为待推断的潜在结构（例如，对破坏过程设置先验）；这为贝叶斯稳健性和混合建模搭建了自然的桥梁。
- **连续模型的近似扇区**。 用粗粒度划分或全纯算子的谱摘要替代离散轨道划分，从而得到可控的近似。
- **超越图模型**。 超图模型和高阶约束会形成更高维的“环”结构；开发高维全纯编译是一条很有前景的方向。

9 结论

我们提出了一种全纯编译推断过程，将环路推断中最棘手的部分——环周围不一致的信息——从一种隐式且容易导致失败的现象，转变为明确且可计算的对象。简而言之：我们不再任由环在消息传递过程中“悄悄相互冲突”，而是精确测量每个全局环未能达成一致的程度，将这种差异存储在一组小型全纯核中，然后把问题编译为扇区树，从而能够按扇区进行精确推断。

因此，这项工作的创新体现在概念和算法层面：我们用确定性编译取代环状图上的不稳定动态过程，该编译（i）将环挫折分离为可解释的全纯项，（ii）将接口纤维分解为轨道/扇区，从而揭示由环产生的真正全局自由度；（iii）在明确且可检验的结构条件下，在每个扇区内的编译树上执行精确推断。由此得到的推断流程不仅能在标准有环置信传播发生振荡或发散时保持可靠，还具有诊断能力：当推断遇到困难时，该方法会告诉你不一致存在在哪里，以及它如何传播。

除了受挫图模型这一直接目标外，编译视角也能使贝叶斯统计及相关领域受益。对于离散模型或离散—连续混合模型中的贝叶斯方法，全纯核提供了一种原则性方法，可以：（a）诊断局部条件结构何时在全局上不相容；（b）区分真正的全局不确定性（扇区选择）与局部不确定性（扇区内推断）；（c）在常规近似方案无法处理的情形下，获得精确且忠实于语义的后验。更广泛地说，任何依赖有环因子图推断的领域——例如存在受挫现象的统计物理、纠错码、多传感器融合／同步定位与地图构建（SLAM）、关系模型或类约束满足后验——都可以采用同一思路：将“环路难题”转化为显式的编译变量，而非数值不稳定性。

这种方法的局限并非如下。 它并不能为任意含环模型提供通用的可处理性保证：如果扇区数量激增，即使每个扇区在编译后都易于处理，整个方法的计算成本仍可能很高。它无法消除模型设定错误、似然函数质量不佳或可辨识性不足的问题——全纯编译遵循贝叶斯语义，但无法挽救病态统计模型。它也不是一种学习过程：它针对给定模型进行推断，而非参数估计、结构学习或摊销推断。最后，在简单情形下，它并非旨在取代快速近似推断；其主要价值在于那些环确实存在冲突、标准有环置信传播变得不可靠的困难情形。

A 范畴论预备知识（简要）

正文有意避免展开介绍范畴论背景。我们仅记录本文在实际操作中用到的高层原则：(i) 因子图句法具有超图式（弦图式）演算；(ii) 消元/边缘化是函子性的，并可表示为基于幺半环/半环的矩阵式语义；(iii) 全局一致性条件由标准泛构造（极限/等化子）支配。关于幺半范畴/超图范畴和图式推理的背景知识，请参阅 [1, 7, 9, 15]。

B 层与下降（简要）

我们对下降的使用遵循标准范式：由覆盖索引的局部数据，当且仅当重叠部分上的限制相容并满足相应的等化子（匹配）条件时，才能粘合为全局数据。关于网络一致性与推断的基于层的观点，详见 [4]；关于神经与 Kan 条件的单纯形和同伦背景，可参阅 [17]。

C 规范群胚与 Kan 结构（指引）

第4 节概述的内容构造了一个由消息重缩放构成的群胚，并通过 Kan 复形的视角诠释规范轨道，其中基本群刻画了环路传输。这为从“消息规范”到拓扑整体性提供了概念上的桥梁，但实现我们的全纯感知编译机制并不需要这些内容。详情请参阅扩展版。最接近的标准参考文献包括 [17]（单纯集合/Kan 复形）和 [7]（范畴组合语义）。

D 连续变量路线图

对于连续变量，传输核会变为马尔可夫核或积分算子，而“轨道”的提取则会成为无限维函数空间中的商／识别问题。一种切实可行的途径是：(a) 计算环算子的低秩或谱摘要；(b) 将其与近似消息传递方案（如期望传播／变分消息传递）结合；参见 [18]。这一方向仍有待探索；我们将其视为未来工作，而非当前能够保证实现的结果。

参考文献

[1] S. Mac Lane。《自然结合律与交换律》。莱斯大学研究，49(4)，1963。

[2] S. L. Lauritzen 和 D. J. Spiegelhalter。在图结构上基于概率进行局部计算及其在专家系统中的应用。英国皇家统计学会杂志：*B*辑，50(2)：157–224，1988。

[3] M. Mézard 和 G. Parisi。重新审视 Bethe 晶格自旋玻璃。欧洲物理学杂志 *B*，20：217–233，2002。

[4] M. Robinson。《拓扑信号处理》。Springer，2017年。[5] J. S. Yedidia、W. T. Freeman 和 Y. Weiss。构造自由能近似与广义置信传播算法。*IEEE*信息论汇刊，51(7):2282–2312，2005年。[6] J. Pearl。《智能系统中的概率推理：合理推断网络》。Morgan Kauf- mann，1988年。[7] B. Fong 和 D. I. Spivak。概率论的范畴论方法。*arXiv*预印本 *arXiv:1406.6030*，2014年（有多个版本；通常引用2019年版本）。[8] T. Fritz。关于马尔可夫核、条件独立性和充分统计量定理的综合方法。数学进展，370:107239，2020年。[9] B

. Fong 和 D. I. Spivak。超图范畴。*arXiv:1806.08304*，2018（通常引用为 2019 年）。

[10] J. Morton。幺半范畴中的信念传播。Electronic Proceedings in Theoretical Computer Science (EPTCS)，172，2014。[11] J. Elgueta。有限集群胚是半环范畴 2-范畴中的双初始对象。arXiv:2004.08684，2020。[12] Z. Izhakian、M. Knebusch 和 L. Rowen。分层热带代数与几何的范畴概念。arXiv:1207.3487，2012。[13] D. Koller 和 N. Friedman。《概率图模型》。麻省理工学院出版社，2009。[14] J. S. Yedidia、W. T. Freeman 和 Y. Weiss。构造自由能近似和广义置信传播算法。IEEE信息论汇刊，51(7):2282–2312，2005。[15] B. Coecke 和 A. Kissinger。《Picturing Quantum Processes》。剑桥大学出版社，2017。[16] A. Kissinger。《Pictures of Processes: Automated Graph Rewriting for Monoidal Categories and Applications》。博士论文，牛津大学，2012。[17] P. G. Goerss 和 J. F. Jardine。《单纯同伦论》。Birkhäuser，1999。[18] M. J. Wainwright 和 M. I. Jordan。图模型、指数族与变分推断。Foundations and Trends in Machine Learning，1(1–2):1–305，2008。[19] A. Singer。通过特征向量和半定规划实现角同步。Applied and Computational Harmonic Analysis，30(1):20–36，2011。[20] A. S. Bandeira、A. Singer 和 D. A. Spielman。图连接拉普拉斯算子的 Cheeger 不等式。SIAM矩阵分析与应用杂志，34(4):1611–1630，2013。[21] D. Poole。一阶概率推理。载于第 18 届国际人工智能联合会议（IJCAI）论文集，第 985–991 页，2003。[22] L. de Salvo Braz 、 E. Amir 和 D. Roth。提升一阶概率推理。载于第 19 届国际人工智能联合会议（IJCAI）论文集，第 1319–1325 页，2005。[23] G. F. Cooper。使用贝叶斯信念网络进行概率推断的计算复杂性。Artificial Intelligence，42(2–3):393–405，1990。[24] D. Roth。近似推理的难度。Artificial Intelligence，82(1–2):273–302，1996。[25] F. R. Kschischang、B. J. Frey 和 H.-A. Loeliger。因子图与和积算法。IEEE信息论汇刊，47(2):498–519，2001。[26] Y. Weiss。存在环的图模型中局部概率传播的正确性。Neural Computation，12(1):1–41，2000。[27] A. T. Ihler、J. W. Fisher III 和 A. S. Willsky。有环置信传播：收敛性与消息误差的影响。Journal of Machine Learning Research，6:905–936，2005。[28] T. Heskes。关于有环置信传播不动点的唯一性。Neural Computation，16(11):2379–2413，2004。[29] J. M. Mooij 和 H. J. Kappen。和积算法收敛的充分条件。IEEE信息论汇刊，53(12):4422–4437，2007。[30] T. Heskes。利用凸性论证高效最小化 Bethe 和 Kikuchi自由能。Journal of Artificial Intelligence Research，26:153–190，2006。[31] M. J. Wainwright、T. S. Jaakkola 和 A. S. Willsky。树重加权置信传播算法与基于伪矩匹配的近似最大似然估计。载于第九届人工智能与统计国际研讨会（AISTATS）论文集，第 308–315 页，2003。[32] M. Chertkov 和 V. Y. Chernyak。环演算有助于改进低密度奇偶校验码的置信传播和线性规划译码。载于第 44 届 Allerton通信、控制与计算会议论文集，2006。