arXiv:2603.07816v1 [math.CO] 8 Mar 2026

因子复杂度极低的无限词：词组合学导论

Mélodie Andrieu[1,2]

[1]约瑟夫·刘维尔纯粹与应用数学实验室，滨海奥帕尔大学，法国。[2]数学建模中心，智利大学及 *IRL-CNRS 2807*，智利。melodie.andrieu@univ-littoral.fr

2025年12月

Lecture Notes

摘要

本讲义介绍词组合学及其与动力系统、代数和算术的相互作用。

本文的核心主题是无限词的低因子复杂度这一概念。我们探讨以下引导性问题：二元、三元或更一般的有限字母表上的非平凡无限词，其最小复杂度是多少？应如何形式化“非平凡性”？哪些词能达到这一最小复杂度？这样的词多吗？它们有趣吗？

在探讨这些问题时，我们介绍词组合学中的经典对象与工具，例如斯图姆词和 *Rauzy* 图，以及鲜为人知的结果和新成果。特别是，第三章专门讨论 R. Tijdeman 于 1999 年提出的一个定理。该定理推广了 M. Morse 和 G. Hedlund 于 1938 年得到的一项开创性结果；我们还给出了由 J. Cassaigne 和作者于 2022年提出的一种概念上全新的代数证明。

第一章不要求读者预先掌握词组合学和符号动力系统知识，适合研究生阅读。第二章和第三章则面向研究生以及熟悉斯图姆词及其推广的研究人员。

本文根据*2024*年*12*月在智利瓦尔帕莱索举行的第七届国际会议 *Dyadisc* 上的一场讲座整理而成。文中包含的新成果自*2023*年起也已在多个研讨会上发表。

目录

概述 **3**

1 Morse 和 Hedlund 的一个开创性定理（1938年） **5**

1.1 预备知识：无限词、复杂度.5 1.2 Morse 和 Hedlund 的奠基性定理.6 1.3 斯图姆词类.7 1.3.1 基本性质.7 1.3.2 插曲：连分数速成.8 1.3.3 斯图姆词与连分数之间的联系.9 1.3.4 斯图姆词的高效构造.13 1.3.5 斯图姆词的等价动力学描述与几何描述 16 1.4 评注式参考文献.21

2 迈向将 Morse–Hedlund 定理推广至 d元字母表：Tijdeman 的一个定理（1999）23

2.1 动机.23 2.2 斯图姆词的 d元推广实例.24 2.3 关于“最小复杂度”对 d元词可能含义的讨论 . . . 24 2.4 字母频率的有理独立性条件.26 2.5 Tijdeman 的一个定理.27 2.6 迈向对 Tijdeman 意义下具有最小复杂度的 d元无限词的刻画.28 2.7 Tijdeman 定理的原始版本与强化版本 29

3 Tijdeman 定理的另一种代数证明及其推论（2022年）31

3.1 引言.31 3.2 预备知识：Rauzy 图与流矩阵的定义 . . . . . 32 3.2.1 Rauzy 图. 32 3.2.2 流矩阵.33 3.3 定理 3.1 的证明.34 3.3.1 定理 3.1 证明的总体结构.34 3.3.2 引理 3.9 的证明.35 3.3.3 引理 3.10 的证明.36 3.3.4 引理 3.11 的证明.37 3.3.5 附录：线性代数简要回顾.38 3.4 与 Tijdeman 原始证明的比较.38 3.5 本证明的一个副产品：具有 Tijdeman 最小复杂度的词是树状的.39

概述

设 $d \geq 1$ 为一个整数。无限 d元词是一个序列 $w \in \{1, \ldots\ldots, d\}^{\mathbb{N}}$。 w 的复杂度是一个函数，它对每个整数 n，统计 w 中出现的长度为 n的不同子词的数量。本课程旨在回答以下两个问题：

1）什么是有趣的 d元词的最小复杂度？2）哪些词能达到这一最小复杂度？当然，要回答这些问题，就需要对最小和有趣给出严谨且恰当的定义。

对于二元词，这一问题自20世纪40年代以来便已得到透彻理解：斯图姆词恰好是具有最小复杂度的有趣的词（在此情形下，“有趣”意为非最终周期）。第1章将全面介绍 Sturmian 词。我们从纯粹的组合视角引入它们，并说明它们与连分数以及某些动力系统的关系。特别是，我们着重阐述字母频率的有理独立性为何是 Sturmian 词的一项核心性质。第2章将探讨如何把 M. Morse 和 G. Hedlund 于1938年提出的开创性定理以及 Sturmian 词推广到 d 元字母表，其中$d \geq 3$。我们将看到，在这一背景下，“有趣”不再能意味着“非最终周期”，而应改为“具有有理无关字母频率”。本章最后将讨论 R. Tijdeman 于1999年提出的一条定理（J. Cassaigne 与作者于2022年独立重新发现），该定理给出了具有有理无关字母频率的 d 元词的最小复杂度。最后，第3章将给出 Tijdeman 定理一种概念上全新的代数证明，并探讨其推论：具有有理无关字母频率且复杂度最小的 d 元词的组合结构。

这三章可以独立阅读。

致谢 作者对 Léo Vivion 在准备这些讲义时给予的无价帮助深表感激。

第1章

Morse 和 Hedlund（1938年）的一项开创性定理

1.1 预备知识：无限词、复杂度

设 u 是一个有限词，w 是一个无限词（根据定义：一个$\mathcal{A}^{\mathbb{N}}$ 中的序列，其中$\mathcal{A}$ 是一个称为字母表的有限集合）。我们称 u 是一个长度为 n 的 w的因子，记作 $u \in \mathcal{L}_n(w)$，如果 u 在 w 中作为由 n 个连续字母组成的块出现。例如，art 是法语词 montmartre 的一个长度为 3 的因子（art $\in \mathcal{L}_3$(montmartre)）。

定义 1.1. 无限词 w 的复杂度是一个函数，对于每个整数 n，计算 w中长度为 n的因子数量：

$$\begin{array}{rccc} p_w: & \mathbb{N} & \longrightarrow & \mathbb{N} \\ & n & \longmapsto & \mathrm{card}(\mathcal{L}_n(w)). \end{array}$$

（注意，按照约定，我们总是有 $p_w(0) = 1$：每个词恰好有一个长度为 0 的因子，即“空词”，记为 ϵ。）

以下例子留给读者验证。

例 1.2. (i) 若 $w_1 = 1^\omega = 111111...$ 是常一元词，则对所有 $n \in \mathbb{N}$，$p_{w_1}(n) = 1$。事实上，w_1 唯一的长度为 n 的因子是 1...1（即字母 1 重复n 次）。注意，w_1 具有绝对最小复杂度。

(ii) 若 $w_2 =$(12)$^\omega = 12121212...$ 是周期为 12 的周期词，则对所有 $n \in \mathbb{N}_{\geq 1}$，$p_{w_2}(n) = 2$。事实上，$w_2$ 恰有两个长度为 n的因子：一个以字母 1 开头，另一个以字母 2 开头。

(iii) 若 $w_3 =$ (1233)$^\omega = 1233 \cdot 1233 \cdot 1233...$ 是周期为 1233 的周期词，则对所有 $n \in \mathbb{N}_{\geq 2}$，$p_{w_3}(1) = 3$ 且 $p_{w_3}(n) = 4$。事实上，w_3 恰由 3 个字母组成，且每个长度为 $n \geq 2$ 的因子完全由其前两个字母决定：12、23、33 或 31。

（iv）十进制 *Champernowne* 词，通过连接所有整数的十进制展开得到：

$$w_C = 01234567891011121314151617181920 21...$$

其复杂度为 $p_{w_C}(n) = 10^n, \forall n \in \mathbb{N}$。事实上，每个有限的十元词都出现在 w_C 中。（练习：因子 02 首次出现在哪里？）对于无限十元词而言，这个复杂度是最大的。

（v）人们认为，尽管远未得到证明，黄金比例的十进制展开

也具有最大复杂度。这一说法是一个范围大得多的猜想的特例。该猜想由 E. Borel 于1950年提出，断言每个无理代数数都“在每个进制下正规” [Bor50]。

顾名思义，复杂度量化了一个无限词有多么复杂。这一组合函数应被理解为动力熵概念的细化，它衡量对轨迹一小部分的了解能在多大程度上提供关于其可能长期行为的信息。下一截面证明，在这一尺度上，周期词是最简单的词。

1.2 莫尔斯与Hedlund的奠基性定理

下一则定理出现在M. Morse与G. Hedlund的第一篇文章 [MH38]中；它是词组合学的奠基性成果之一。

定理1.3 （M. Morse与G. Hedlund，1938年）。 以下断言是等价的：

(i) 无限词 w 最终周期；(ii) 其复杂度 p_w 有界；(iii) 存在 $n \in \mathbb{N}$ 使得 $p_w(n) \leq n$。

证明。 (i) $\Rightarrow$ (ii) 一个最终周期词 $w = u \cdot p^\omega$ 在其前周期 u 或周期 p的每个位置上，至多产生一个长度为 n （无论 n 为何）的新因子。因此，其复杂度 p_w 不超过其前周期与周期长度之和： $|u| + |p|$。

(ii) $\Rightarrow$ (iii) 这是显然的：只需取 $n = M$，其中 $M \in \mathbb{N}$ 是复杂度函数的上界。

(iii) $\Rightarrow$ (i) 由于 $p_w(0) = 1$ （这是我们的约定）， $p_w(n) \leq n$，并且由于 $n \mapsto p_w(n)$ 单调不减（事实上， w 的每个长度为 n 的因子总能延伸为一个长度为 $n+1$的因子），因此必定存在 $n_0 \in \{1, \ldots, n-1\}$ 使得 $p_w(n_0) = p_w(n_0+1)$。这个等式意味着，若 u是一个长度为 n_0的因子，那么它在 w 中的所有出现之后总是跟着同一个字母。（否则， u 会产生两个长度为 n_0+1的因子，从而意味着 $p_w(n_0+1) > p_w(n_0)$。）此外，在 w中出现的 $p_w(n_0)$ 个长度为 n_0的因子中，至少有一个（仍记作 u）会无限次出现。因此，存在 $v \in \mathrm{L}(w)$ 使得 $uvu \in \mathrm{L}(w)$（我们用 $\mathrm{L}(w)$ 表示 w的所有因子组成的集合，不论其长度）。由于 w 的每个长度为 n_0 的因子之后总是跟着同一个字母，我们可以推得，因子 u 之后总是跟着因子 vu。这表明存在 $p \in \mathrm{L}(w)$ 使得 $w = pu(vu)^\omega$，即 w 是最终周期词。□

注1.4. 由该证明可知，词 w 是最终周期的，当且仅当其复杂度函数是局部常值的：存在一个长度 m，使得 $p_w(m+1) = p_w(m)$；当且仅当其复杂度是最终常值的：对所有 $n \geq m$， $p_w(n) = p_w(m)$。

(iii) $\Rightarrow$ (i) 的逆否命题指出：若 w 不是最终周期的，则其复杂度对所有 $n \in \mathbb{N}$满足 $p_w(n) \geq n+1$ 。这引出了如下自然的问题。

问题 1.5. 是否存在一个无限词 w ，其复杂度对于所有 $n \in \mathbb{N}$ 都恰好为 $p_w(n) = n+1$？

正如我们将看到的，这个问题的答案是“是”。这个肯定的答案将引出三个新问题：

1. 有多少个无限词满足这一复杂度条件？2. 它们是什么样的？我们能否有效地构造它们？3. 它们有趣吗？

本章余下部分将致力于回答这些问题。先透露一下，答案将分别是：“不可数多个，基本上每个无理数对应一个”，“是”，“是”，以及“是的，当然是！”

定义 1.6。 具有复杂度 w $p_w(n) = n+1$ （对所有 $n \in \mathbb{N}$）的无限词称为Sturmian词。

例 1.7。 虽然我们尚未证明Sturmian词的存在性（这将在稍后完成），但这里给出第一个例子，称为斐波那契词。

$$w_{fibo} = 1211212112112121121121121121121121121121121121121121121121...$$

在接下来的几节中，我们将了解这个词是如何构造的，以及它为什么被称为“斐波那契”。

练习 1.8. 证明：如果一个无限词对于无限多个 $n \in \mathbb{N}$ 满足 $p_w(n) = n+1$，那么它就是Sturmian词。

1.3 斯图姆词类

1940年，M. Morse 和 G. Hedlund 在他们的第二篇论文 [MH40] 中开启了对斯图姆词的研究。此后，人们一直在广泛研究斯图姆词。这一节旨在对斯图姆词作一个自成一体的介绍。不过，内容远非面面俱到（例如，有意略去了阿贝尔性质）。有兴趣的读者可参阅 [Lot02，第 2] 章，以了解补充内容。

本节提纲。 接下来的两个小节中，我们假设斯图姆词确实存在，并研究它们的性质。我们将看到，如果它们存在，就必定与连分数密切相关。在第三个小节中，我们利用连分数设计一种有效且高效的构造方法——从而最终证明它们确实 存在。在第四个也是最后一个小节中，我们将说明如何从动力学和几何角度直观呈现斯图姆词。

1.3.1 基本性质

下一个命题汇集了后续内容中需要用到的斯图姆词的基本性质。

命题 1.9. 设 w 是一个*Sturmian*词。那么：

(i) w 是一个二元词（下文将字母表设为 $A = \{1,2\}$）。(ii) w 是“递归的”（定义：w 的每个因子都出现无穷多次）。(iii) </style}如果 表示“移位映射”（定义：删除无限词首字母的映射；例如，(314159...) $= 14159...$），那么 $S(w)$ 也是一个 *Sturmian*词。(iv) </style}对于每个，都存在唯一的因子 L(w)，使得其两个延伸$u1$ 和 $u2$ 都是 w 的因子（定义：我们称 u 为“右特殊”）。

证明。 (i) 显然：因为 $p_w(1) = 2$, w 恰好使用两个字母。

(ii) 反证法：假设 $u \in \mathcal{L}_n(w)$ 并非出现无穷多次；例如，假设它在位置 $N \in \mathbb{N}$ 之后不再出现。那么，若用 $w' := S^N(w)$ 表示 w 在位置 N 之后的副本，我们有 $u \notin \mathcal{L}_n(w')$，因此 $p_{w'}(n) < p_w(n) = n+1$。这与定理1.3矛盾，因为 w' 仍然是非最终周期的。

(iii) 显然， $S(w)$ 的每个因子都是 w 的因子。由于 w 是递归的，反过来也成立： w 的每个因子都是 $S(w)$ 的因子。因此，$S(w)$ 与 w 的复杂度函数相等，且 $S(w)$ 是一个 Sturmian词。

(iv) 由于 w 是一个二元词，量 $p_w(n+1) - p_w(n)$ 恰好等于 w 中长度为 n 的右特殊因子的数量。对于Sturmian词 w，这个差值等于 1，因此对每个长度 n， w 恰好有一个长度为 n 的右特殊因子。

□

习题 1.10. 设 w 为一个 Sturmian词。证明：对每个 $n \in \mathbb{N}$，存在唯一的 $v \in \mathcal{L}_n(w)$，使得两个左扩展 $1v$ 与 $2v$ 都是 w 的因子（定义：我们称 v 为“左特殊”）。

1.3.2 插曲：连分数速成课

在继续之前，让我们回顾一下什么是连分数，以及它为何是一个极其有趣的对象。（在下一小节中，我们将说明它如何与斯图姆词相关联。）

连分数是一种记数系统，也就是用有限或无限的整数序列来表示实数的一种方式。正实数 x 的连分数展开 $(a_n)_n$ 通常记为

$$x = [a_0; a_1, a_2, a_3, \cdots].$$

例如， $17/6$ 的连分数展开为 $[2;\ 1, 5]$, 黄金比例 $\varphi := (1+\sqrt{5})/2$ 的连分数展开为 $[1;\ \overline{1,\ 1,\ 1,\ 1,\ 1,\ 1}...]$（恒等于 1 的无限序列），而 π 的连分数展开是一个以 $[3;\ 7,\ 15,\ 1,\ 292,\ 1,\ 1,\ 1,\ 2,\ ...]$开头的无限序列。与进制展开不同，出现在连分数中的整数 a_n （我们称之为“部分商”）未必有界。

问题：如何计算部分商？

a_n是 $x \in \mathbb{R}_{>0}$ 的部分商，它们恰好是对这对实数（x,1）应用“欧几里得算法”时依次得到的整数 商。例如， a_0 的第一个部分商是唯一满足以下条件的整数 q：

$$x = q \times 1 + r,$$

其中 $0 \le r < 1$。（换句话说， $a_0 = \lfloor x \rfloor$ 和 $r = \{x\}$ 分别是 x的整数部分和小数部分。）接着， a_1 的第二个部分商是用 1 除以 r 时“欧几里得除法”所得的整数商（等价地，它也是用 1 除以 $1/r$ 时欧几里得除法所得的整数商），依此类推。用算法表示，我们可以写成：

```
def CF(x):
    print⌊x⌋
    if{x} ≠ 0, return CF(1/{x})
```

注意，该算法并不总是终止。

下一命题说明了连分数为何是一种极为有趣的代数与分析工具：它能无歧义地表示实数，提供高质量的有理数逼近，并刻画有理数与二次[1] 数。

命题 1.11. 设 $x = [a_0;\, a_1, a_2, a_3, \cdots]$ 是一个正实数。

(1) x 的连分数展开是有限的，当且仅当 $x \in \mathbb{Q}$。(2) [狄利克雷定理] 有理数序列

$$\left(\frac{p_n}{q_n}\right)_n := \left([a_0; a_1, a_2, \cdots, a_n]\right)_n = \left(a_0 + \cfrac{1}{a_1 + \cfrac{1}{\ddots + \frac{1}{a_n}}}\right)_n$$

能够有效地收敛到 x；此外，这些有理数是 x 的“最佳逼近”，其含义如下：若不愿付出分母增大的代价，就无法找到比 p_n/q_n更接近 x 的有理数。

(3) 以下映射是一个双射：

$$\begin{array}{ccc} \mathbb{R}_{>0} \setminus \mathbb{Q} & \longrightarrow & \mathbb{N} \times (\mathbb{N}_{\geq 1})^{\mathbb{N}} \\ x & \longmapsto & (a_n)_{n \in \mathbb{N}} \end{array}$$

(4) [拉格朗日定理] 无理数 x 是二次无理数，当且仅当其连分数展开是最终周期的。

将拉格朗日定理推广到三次数，或更一般地推广到代数数，是一个长期存在的未解问题（该问题最初由C. Hermite于1850年提出）。这一问题是C. G. J. Jacobi引入“多维连分数”（我们稍后将回到这一点）的开创性动机之一。

1.3.3 斯图姆词与连分数之间的联系

斯图姆词与连分数通过我们现在引入的重整化过程相联系。

设 w 为一个Sturmian词。由于 w 不是最终周期的，它必定同时包含因子12和21。由于 $p_w(2) = 3$， w 具有第三个长度为2的因子，该因子可以是**11异或**（互斥或）22。若 w 包含11，则称其为“类型1”，否则称其为“类型2”。

重整化过程。令 $i \in \{1, 2\}$，并用 j 表示另一个字母。设 w 是类型为 i 的Sturmian词。那么，在 w 中，每次出现字母 j （当 w 以 j 开头时的首次出现除外）之前，至少会出现一次 i。令 $R_i(w)$ 表示通过删去 w 中每个紧邻 j 之前的 i 后得到的无限词。于是，我们将重整化过程定义为

$$R(w) = \begin{cases} R_1(w) \text{ if } w \text{ is of type } 1, \\ R_2(w) \text{ otherwise.} \end{cases}$$

[1]定义：若一个实数是一个具有整数系数的二次多项式（分别为三次多项式或任意次数多项式）的根，则称它为二次（分别为三次或代数数。

例 1.12. 通过一个例子来说明重整化过程很有用。为此，我们再次考虑斐波那契词，暂且假设它是斯图姆词。

$$w_{fibo} = 121121211211212112121121112121121112121121121121121121121121211212...$$

斐波那契词属于类型1，因此

$$R(w_{fibo}) = R_1(w_{fibo}) = 212212122122121221212212212212212212...$$

我们断言，这个新的无限词也是 Sturmian词。

定理 1.13。 如果 w 是 *Sturmian*词，那么 $R(w)$ 也是 *Sturmian*词。

因此，我们可以无限次地迭代 R。我们将 Sturmian词 w 的"游程长度"序列（b_n）$_n$ 定义为记录操作R_1 （分别为 R_2）在下一次应用 R_2 （分别为 R_1）之前重复次数的序列。具体而言，如果（R_{i_n}）$_{n\in\mathbb{N}}$ 表示算子 R 作用于 w并进行迭代时所取得的连续值，则有

$$i_0 i_1 i_2 i_3 i_4 i_5 i_6 ... = 1^{b_0} 2^{b_1} 1^{b_2} 2^{b_3} ...$$

（注意，按照惯例， b_0 表示首次应用 R_2之前连续应用 R_1 的次数。）

由于（R_{i_n}）$_n$ 中算子 R_1 和 R_2 都出现了无穷多次（见练习 1.17），其游程长度（b_n）是一个无限的整数序列。更准确地说，对于每个 $k \geq 1$，都必须有$b_k > 0$；当且仅当所考察的 Sturmian词 w 是类型2时， $b_0 = 0$。

例 1.14. 算子 R 作用于斐波那契词时的前几次迭代如下：

w	$= w_{fibo} = 121121211211212112121121112121121112121121121121121121121121211212...$	类型1
w'	$= R(w) = R_1(w) = 212212122122121221212212...$	type 2
w''	$= R^2(w) = R_2(w') = 121121211211212...$	type 1
w'''	$= R^3(w) = R_1(w'') = 212212122...$等等。	type 2

事实上，我们可以证明 $w'' = w_{fibo}$ （这是斐波那契词的一个特殊性质），并由此推得w_{fibo} 的游程长度是恒等于 1 的常数序列。

注记 1.15. 根据定义，"替换"是一个从字母表到有限词集合的函数。它可扩展为有限词和无限词上的态射，并且保持拼接。（例如，有限词 $aaba$ 在替换s 下的像为： $a \mapsto aab$, $b \mapsto ba$ 是 $s(aaba) = s(a)s(a)s(b)s(a) = aabaabbaaab$。）现在，若 σ_1 和 σ_2表示特定的替换

$$\sigma_1 : \begin{array}{l} 1 \mapsto 1 \\ 2 \mapsto 12 \end{array} \qquad \sigma_2 : \begin{array}{l} 1 \mapsto 21 \\ 2 \mapsto 2, \end{array}$$

那么请注意，在例 1.14 中，可以通过施加σ_1从像 w' 重构出 w。类似地，无限词 w' 和 w'' 可以分别通过 $w' = \sigma_2(w'')$和 $w'' = \sigma_1(w''')$从它们各自的像 w'' 和 w''' 重构出来。接下来的内容中，这两个替换 σ_1 和 σ_2将发挥核心作用。

练习 1.16. 设 w 是类型为 i的 Sturmian词。
1) 证明等式 $w = \sigma_i(R_i(w))$ 并非总是成立。（提示：考虑 $S(w_{fibo})$。）2) 不过，证明如果 u 是 $R_i(w)$ 的因子，那么 $\sigma_i(u)$ 就是 w 的因子。

定理 *1.13* 的证明。设 w 是类型为 i的 Sturmian词，另一个字母仍记作 j。我们要证明，$w' := R(w) = R_i(w)$ 也是一个 Sturmian词。首先，注意到 w' 最终周期，当且仅当 w 最终周期——而它并非如此。因此，根据 Morse–Hedlund 定理 1.3，对于每个$n \in \mathbb{N}$，必有 $p_{w'}(n) \geq n+1$。

对于反向不等式，我们将证明：对每个 $n \in \mathbb{N}$， w' 至多有一个长度为 n的右特殊因子。（因此，再次利用命题 1.9 断言 (iv) 证明中的论证，对于所有 $n \in \mathbb{N}$，我们有 $p(n+1) - p(n) \leq 1$，从而$p(n) \leq n+1$。）我们用反证法。假设 w' 含有两个不同且长度相同的右特殊因子， u 和 v。那么，四种扩展 ui、 uj、 vi和 vj 都是 w'的因子。根据练习 1.16，它们在替换 σ_i 下的像都是原始 Sturmian词 w的因子： $\sigma_i(ui) = \sigma_i(u)i$、$\sigma_i(uj) = \sigma_i(u)ij$、 $\sigma_i(vi) = \sigma_i(v)i$，以及$\sigma_i(vj) = \sigma_i(v)ij$。我们可以更精确地指出，两个词 $\sigma_i(u)ii$和 $\sigma_i(v)ii$ 都是 w的因子。事实上，由于 ui 是 w'的因子， uii 或 uij （也可能两者都是）也是 w'的因子，由此可知， $\sigma_1(u)ii$ 或 $\sigma_1(u)iij$ 是 w的因子；无论哪种情况，$\sigma_1(u)ii$ 都是 w的因子——对 $\sigma_1(v)ii$的论证也类似。

至此，我们已经证明：两个有限词 $\sigma_i(u)i$ 与 $\sigma_i(v)i$（因 $u \neq v$ 而互不相同）构成 Sturmian词 w 的两个右特殊因子。这尚不构成矛盾，因为这两个因子的长度未必相同。然而，由于 u 与 v 是等长且互不相同的词，它们总可以分解为（至多交换 u与 v）

$$\begin{cases} u = pis, \\ v = rjs, \end{cases}$$

其中 p, r 与 s 是可能为空的有限词。因此，它们在 σ_i 作用下的像的后缀，再以字母 i 延拓，即 $i\sigma_i(s)i$ 与 $j\sigma_i(s)i$，是 Sturmian词 w 的两个互不相同、等长的右特殊因子。这与命题 1.9 矛盾。

□

练习 1.17. 设 w 为一个 Sturmian词。
1）证明算子 R_1 与 R_2 在 w 的连续重整化中均出现无穷多次。

2）证明 w 是一致回归的。（定义：若对 w 的每一个因子 u，都存在整数 N，使得 u 在 w 中任意两次相继出现之间至多隔着 N 个字母，则称 w 为“一致回归”。）

下一个定理揭示了斯图姆词、我们的重整化过程与连分数之间的联系。

定理1.18. 设 w 是一个*Sturmian*词。记其在重整化过程 R的迭代下的游程长度为 $(b_n)_{n\in\mathbb{N}}$ 。那么

(i) w 存在字母频率（以下记作 f_1, f_2）；(ii) $[b_0;\ b_1, b_2, b_3, \cdots]$ 是比值 $\frac{f_1}{f_2}$的连分数展开；(iii) 比值 $\frac{f_1}{f_2}$ 是一个无理数。 —

—

根据定义，无限词 w 中“字母 a的频率”是指：如果极限存在，则为 a 在 w不断增长的前缀中所占比例的极限。具体来说，它是以下极限（同样，若其存在）

$$f_a = \lim_{n\to+\infty} \frac{|\mathrm{pref}_n(w)|_a}{n},$$

其中， $|\mathrm{pref}_n(w)|_a$ 表示 w的前 n 个字母中 a 的数量。

例 1.19. 由于斐波那契词的游程长度是恒等于 1 的无限序列，其字母频率 f_1/f_2 之比就是黄金比例。

定理*1.18*的证明，断言 *(i)*。设 w 是一个斯图姆词。由于 w 是一致回归的（如练习1.17所证），并且满足

$$\liminf_{n\to+\infty} \frac{p_w(n)}{n} = 1,$$

根据 Boshernitzan 于 1984 年的一条定理，可知 w 存在字母频率 [Bos84]（另见 [FM10, 第 7.3] 节）。 □

注 1.20。经典的做法是利用一种称为“平衡性”的阿贝尔性质，推导出斯图姆词中字母频率的存在性；本讲座不打算介绍这一性质。Boshernitzan定理使我们能够直接从复杂度出发，证明同样的结果。

定理*1.18*的证明，断言 *(ii)* 和 *(iii)*。设 w 是一个斯图姆词。用 f_1 和 f_2 表示它的字母频率。

（ii）根据 w的游程长度序列（b_n）$_n$ 的定义，首项 b_0 是算子 R 最初以 R_1 作用、随后切换为 R_2之前所作用的次数。步骤 1： R_1的初始迭代。假设 $b_0 \neq 0$ （否则，我们跳过此步骤，直接处理 b_1）。令 $w' = R（w）= R_1（w）$。由于 w 和 w' 是斯图姆词（定理 1.13），它们分别具有字母频率，记为 f_1、 f_2、 f_1'和 f_{r2}。我们声称这些字母频率满足关系

$$\frac{f_1'}{f_2'} = \frac{f_1}{f_2} - 1. \tag{1.1}$$

事实上，只需通过算子 R_1统计长度为 n、$n \in \mathbb{N}$的前缀的像中各字母的数量：

$$\begin{cases} |R_1(\mathrm{pref}_n(w))|_2 = |\mathrm{pref}_n(w)|_2 \\ |R_1(\mathrm{pref}_n(w))|_1 = |\mathrm{pref}_n(w)|_1 - |\mathrm{pref}_n(w)|_2 \text{ or } |\mathrm{pref}_n(w)|_1 - |\mathrm{pref}_n(w)|_2 + 1 \end{cases}$$

（字母 2 的出现次数不变；在每次出现字母 2 之前，我们都删去一个 1，但可能位于开头的 2 除外。）由于（R_i(pref$_n$(w))）$_n$ 是 w'的一列不断增长的前缀，因此有

$$\frac{|R_1(\mathrm{pref}(w))|_1}{|R_1(\mathrm{pref}(w))|} \longrightarrow_{n\to\infty} f_1'$$

以及关于 f_2' 的对称表达式（回顾一下， $|u|$ 表示一个有限词的长度u）。因此

$$\frac{f_1'}{f_2'} = \lim_n \frac{|R_1(\mathrm{pref}(w))|_1}{|R_1(\mathrm{pref}(w))|_2} = \lim_n \frac{|\mathrm{pref}(w)|_1 - |\mathrm{pref}(w)|_2}{|\mathrm{pref}(w)|_2} = \frac{f_1}{f_2} - 1$$

由此得到（1.1）。现在，在切换到 R_2 之前，操作 R_1 恰好重复 b_0 次，因此 $w'' := R^{b_0}$(w) =(R_1)^{b_0}(w) 的字母频率必须满足

$$\begin{cases} f_1''/f_2'' = f_1/f_2 - b_0 \\ f''1/f''2 < 1 \end{cases} \tag{1.2}$$

（在第二个表达式中，较大的不等式是因为 w'' 属于类型2，而等号成立的情形则被下文的练习 1.21 排除。）表达式（1.2）等价于

$$\begin{cases} b_0 = \lfloor f_1/f_2 \rfloor \\ f_1''/f_2'' = \{f_1/f_2\} \end{cases}$$

（这已经很像连分数了……）

步骤 2：迭代 R_2。现在我们对新的 Sturmian词 w''应用算子 R_2，恰好 b_1 次。我们按照步骤 1 进行，只是交换字母 1 和 2 的角色，因为 w'' 是类型2。因此，我们不考虑商 $\frac{f_1}{f_2}$，而是考虑商

$$\frac{f_2''}{f_1''} = \frac{1}{\{\frac{f_1}{f_2}\}}.$$

现在很清楚，我们正在对输入 $x = f_1/f_2$执行算法 `CF`（见第 1.3.2 节），它展示了 x的连分数。

（iii）由于 w 是 Sturmian词，其重整化序列（R_n）$_n$ 中，算子 R_1 和 R_2 各自都出现无穷多次（练习 1.17）。这意味着它的游程长度（b_n）——我们已证明它是 f_1/f_2的连分数展开——是无限的。由经典命题 1.11 可知，比值 f_1/f_2 是一个无理数。 □

练习 1.21。 证明：不使用定理1.18的断言（ii）和（iii）（否则该定理的证明会陷入循环），任何斯图姆词都没有字母频率 $f_1 = f_2 = 1/2$。提示：若 w 是类型为 i的斯图姆词，则将 w 划分为长度为N的连续块，其中整数 N 的选取应确保所有长度为 N 的因子都至少包含两个词 ii。你需要用到一致回归这一性质。

1.3.4 斯图姆词的一种高效构造

在本小节中，我们证明斯图姆词确实存在，而且有不可数多个：每个正无理数本质上对应一个斯图姆词。这些结果依赖于利用连分数高效构造斯图姆词的方法。

下一个定理给出了这种构造。它用到了替换 σ_1 和 σ_2，这两个替换曾在第1.3.3节中作为重整化过程 R的逆过程出现。它们定义如下：

$$\sigma_1 : \begin{array}{l} 1 \mapsto 1 \\ 2 \mapsto 12 \end{array} \qquad \sigma_2 : \begin{array}{l} 1 \mapsto 21 \\ 2 \mapsto 2. \end{array}$$

定理 1.22. 设$(s_n)_n \in \{\sigma_1, \sigma_2\}^{\mathbb{N}}$ 是一个序列，其中两种替换 σ_1 和σ_2 都无限次出现。那么

(i) 有限词的两个序列$(s_1 \circ s_2 \circ \cdots \circ s_n(1))_n$和$(s_1 \circ s_2 \circ \cdots \circ s_n(2))_n$ 当 $n \to \infty$时收敛到同一个无限词 w_0，该词是 *Sturmian*词；
(ii)替换序列$(s_n)_n$ 的游程长度（定义同前，即 σ_1 重复的次数，直到 σ_2首次出现；接着是 σ_2 重复的次数，直到 σ_1再次出现，依此类推）恰好是比值$\frac{f_1}{f_2}$的连分数展开，其中 f_1 和 f_2 表示 *Sturmian*词 w_0中的字母频率。

（我们使用积集 $\{1,2\}^{\mathbb{N}}$上的蒂霍诺夫拓扑：有限词或无限词的序列 $(u_n)_n$ 收敛到无限词 w，是指这些词 u_n 与 w共享的前缀越来越长。）

例1.23。 设 $(s_n)_n = \sigma_1, \sigma_2, \sigma_1, \sigma_2, \sigma_1, \sigma_2, \ldots$ 为周期为 (σ_1, σ_2) 的周期序列。那么，有限词序列 $u_n := (s_1 \circ s_2 \circ \cdots \circ s_n(1))_n$，其开头为

$$u_0 = u_1 = 1 u_2 = u_3 = 121 u_4 = u_5 = 12112$$
$$u_5 = u_6 = 1211212112112$$
$$u_7 = u_8 = 12112121121121211212112112121121121\ldots$$

收敛于一个Sturmian词 w_0。这个Sturmian词 w_0 被称为“斐波那契词”——它正是我们在本讲中预先提到的Sturmian词例子。正如我们在例1.14和例1.19中所述，其字母频率之比是黄金比例。事实上，替换序列$(s_n)_n$ 的游程长度是常数序列，等于1；因此，$f_1/f_2 = [1;\ 1,1,1,1,1,1,1,\ldots] = \varphi$。下面的练习说明了它为何得名“斐波那契”。

练习 1.24. 令 $(v_n)_n$ 为由下式定义的有限词序列

$$\begin{cases} v_0 = 2 \\ v_1 = 1 \\ v_{n+2} = v_{n+1} + v_n \text{ for every } n \geq 0, \end{cases}$$

其中符号 $+$表示拼接运算。证明 $(v_n)_n$ 收敛到斐波那契词。

定义 1.25. 定理 *1.22* 中得到的斯图姆词称为标准斯图姆词。由其构造标准斯图姆词的替换序列$(s_n)_n \in \{\sigma_1, \sigma_2\}^{\mathbb{N}}$ 称为该词的指令序列。

一个自然的问题是：每个斯图姆词都是标准的吗？也就是说，每个斯图姆词都是由替换 σ_1 和 σ_2作用所得的迭代像序列的极限吗？答案是否定的。不过，每个斯图姆词都与某个标准斯图姆词相关。

练习 1.26. 1）证明存在非标准斯图姆词。2）不过，证明对于每个斯图姆词 w，都存在唯一的标准斯图姆词 w_0，使得 w 属于 w_0的“子移位”。

$$w \in \overline{\{S^n(w_0) \mid n \in \mathbb{N}\}},$$

或者等价地，使得 w 与 w_0 具有相同的因子集合。

下一个结果是定理 1.22 的推论。它表明存在不可数多个斯图姆词，本质上每个正无理数对应一个。

推论 1.27. 对每个正无理数 x，存在唯一的标准*Sturmian*词 w_0，其字母频率满足 $\frac{f_1}{f_2} = x$。

证明。 根据命题 1.11，连分数在正无理数集合与整数无穷序列集合 $\mathbb{N} \times (\mathbb{N}_{\geq 1})^{\mathbb{N}}$之间构成双射；后者又通过“游程长度”的概念，与代换序列集合$\{\sigma_1,\ \sigma_2\}^{\mathbb{N}}$ 构成双射，其中 σ_1 和 σ_2 出现无穷多次。因此，根据定理 1.22，对于每个正无理数 x，都存在唯一一个标准Sturmian词，其字母频率满足 $f_1/f_2 = x$。

□

我们以一条注记和定理 1.22 的证明结束本小节。

注 1.28. 本节介绍的斯图姆词构造已在通用数学软件系统 `Sagemath` 中实现。为了说明该算法的高效性，我们可以这样说：已知 f_1/f_2 的连分数的前 n 项，就能确定斯图姆词 w_0 的前 $\simeq \exp(n)$ 个字母。

定理 *1.22* 的证明。 设 $(s_n)_n \in \{\sigma_1,\ \sigma_2\}^{\mathbb{N}}$ 是一个序列，其中每种替换 σ_1 和 σ_2 都出现无穷多次。对于 $n \in \mathbb{N}$，记 $u_n := s_1 \circ \cdots \circ s_n(1)$，以及 $v_n := s_1 \circ \cdots \circ s_n(2)$。回顾一下，$|u|$ 表示有限词 u 的长度。我们将证明:

1）“整数序列 $(|u_n|)_n$ 和 $(|v_n|)_n$ 单调不减，并且在无穷多个 n 处递增。”因此，若 $(u_n)_n$ 和 $(v_n)_n$ 收敛，它们必定收敛到无限词。2）“序列 $(u_n)_n$ 和 $(v_n)_n$ 收敛到同一个极限。”这个极限记作 w_0。3）“无限词 w_0 对每个 $n \in \mathbb{N}$ 都恰好有一个长度为 n 的右特殊因子。”因此，它的复杂度为 $n \mapsto n+1$，且 w_0 是一个 Sturmian词。

4）“Sturmian词 w_0 的连续重整化序列 $(Q_n)_n \in \{R_1, R_2\}^{\mathbb{N}}$ 由下式给出

$$Q_n = R_1 \iff s_n = \sigma_1.''$$

因此，$(Q_n)_n$ 和 $(s_n)_n$ 的游程长度相同，并且（根据定理 1.18）等于 f_1/f_2的连分数展开，即 w_0中字母出现频率之比。至此，这一步完成，定理也就完全得证。

1）由于任意字母经 σ_1 和 σ_2 映射后的结果都至少包含一个字母，因此总有 $|u_{n+1}| \geq |u_n|$。此外，由于序列 $(s_n)_n$ 中 σ_1 和 σ_2都出现无穷多次，对于无穷多个 $n \in \mathbb{N}$，有 $u_{n+2} = \sigma_1 \circ \sigma_2(u_n)$。由$\sigma_1 \circ \sigma_2(1) = 121$ 和 $\sigma_1 \circ \sigma_2(2) = 12$ 可知，对于这些整数 n，有 $|u_{n+2}| > |u_n|$。因此，序列 $(|u_n|)_n$ 单调不减，并且会无穷多次递增。对于序列 $(v_n)_n$，也可以用类似的论证。

2) 设 $n \in \mathbb{N}$。注意，若 $s_{n+1} = \sigma_1$，则

$$u_{n+1} = s_1 \circ \cdots \circ s_{n+1}(1) = s_1 \circ \cdots \circ s_n(1) = u_n,$$
$$v_{n+1} = s_1 \circ \cdots \circ s_{n+1}(2) = s_1 \circ \cdots \circ s_n(12) = u_n v_n.$$

- 类似地，若 $s_{n+1} = \sigma_2$，则 $u_{n+1} = v_n u_n$ 且 $v_{n+1} = v_n$。
因此，若对每个 $n \in \mathbb{N}$ 我们令

$$p_{n+1} := \begin{cases} u_n & \text{if } s_{n+1} = \sigma_1, \\ v_n & \text{otherwise,} \end{cases}$$

那么，p_{n+1} 是 u_{n+1} 和 v_{n+1}的公共前缀，因此也是 p_{n+2}的公共前缀。由此可知，由于 $|p_n| \to \infty$，三个有限词序列（u_n）$_n$、（v_n）$_n$ 和（p_n）$_n$ 收敛到同一个无限词 $w_0 \in \{1,\ 2\}^{\mathbb{N}}$。

3）我们先作一个初步观察：有限词或无限词经σ_1 映射后的像不包含因子22；对称地，有限词或无限词经 σ_2 映射后的像不包含因子11。

首先，我们证明，对于每个 $n \in \mathbb{N}$， w_0 至多包含一个长度为 n 的右特殊因子。我们采用反证法。设 $u \neq v$ 是 w_0 中两个长度相等的右特殊因子。由于 u_n 收敛于 w_0，存在 n，使得 $u1$、$u2$、$v1$ 和 $v2$ 出现在 $u_n = s_1 \circ \cdots \circ s_n(1)$ 中。我们将通过 s_1 对 $u1$、$u2$、$v1$ 和 $v2$ 进行“去替换”，从而得到 $s_2 \circ \cdots \circ s_n(1)$ 的两个更短且长度相等的右特殊因子 u'' 和 v''。

为严谨起见，我们引入以下定义：如果 u 是 $s(u')$ 的一个因子，并且 u' 对于此性质是极小的（所采用的序关系为：$u < v$ 当且仅当 u 是 v 的因子），则称 u' 是有限词 u 关于替换 s 的一个原像。例如，212 是 12112 关于 σ_1 的唯一原像，而 211 和 212 是 1211 的两个原像。

我们继续证明。不失一般性，假设 $s_1 = \sigma_1$。在这种情况下，右特殊因子 u 和 v 的末字母必定是 1（否则，词 22 就会出现在 w_0中）。于是，当它们后面跟着字母 2 时（如 $u2$和 $v2$），这些末尾的 1 必定源自 $s_2 \circ \cdots \circ s_n(1)$ 中的字母 2。相反，当它们后面跟着另一个 1 时（如 $u1$和 $v1$），这些末尾的 1 必定源自 $s_2 \circ \cdots \circ s_n(1)$ 中的字母 1。现在，从倒数第二个字母开始反向读取 u 和 v，可知它们在 σ_1 下的其余原像是唯一确定的。我们将它们记作 u' 和 v'。因此，词 u' 和v' 是 $s_2 \circ \cdots \circ s_n(1)$ 的两个不同的右特殊因子。此外，它们的长度满足 $|u'|, |v'| < |u| = |v|$，并且它们互不为对方的前缀（否则，$|u| \neq |v|$）。因此，如有必要，我们可以从其中一个词的开头删去几个字母，从而确保 u'和 v'是 $s_2 \circ \cdots \circ s_n(1)$ 中两个不同且等长的右特殊因子，且 $|u'| = |v'| < |u| = |v|$。现在，有限次重复这一“去替换”过程，我们将得到一个整数 k，以及 $u'' \neq v''$ 中两个不同的长度为 1 的右特殊因子，它们属于 $s_k \circ \cdots \circ s_n(1)$。这种情况不可能发生：根据我们先前的观察，11 和 22 不可能同时出现在 $s_k \circ \cdots \circ s_n(1)$ 中。矛盾。

现在我们要证明，对于每个 $n \in \mathbb{N}$， w_0 至少存在一个长度为 n 的右特殊因子。由于右特殊因子的每个后缀也都是右特殊因子，因此只需对无穷多个 n 证明这一性质。注意，形如 $\sigma_1 \circ \sigma_2^k \circ \sigma_1(2)$ 的所有有限词中都出现了因子 11 和 12，其中 $k \geq 1$。由于模式 $\sigma_1 \circ \sigma_2^k \circ \sigma_1$ 必须在 $(s_n)_n$ 中出现无穷多次，我们推知，对于无穷多个 n，有限词 $s_1 \circ \cdots \circ s_n(11)$ 和 $s_1 \circ \cdots \circ s_n(12)$ 都是 w_0 的因子。这两个词互不相同。因此，它们的最长公共前缀（以 $u_n = s_1 \circ \cdots \circ s_n(1)$ 开始）是一个长度大于 $|u_n|$ 的右特殊因子。根据第一步，我们得出 w_0 存在任意长的右特殊因子。第三步的证明至此完成：对于每个长度 $n \in \mathbb{N}$，w_0 恰好存在一个右特殊因子。

4）只需注意，若 $s_1 = \sigma_1$，则 $w_0 = \lim_{k\to\infty} s_1 \circ \cdots \circ s_k(1)$ 属于类型1，而$R(w_0) = R_1(w_0)$ 恰好是 Sturmian词 $w_1 := \lim_{k\to\infty} s_2 \circ \ldots \circ s_k(1)$。对称地，若 $s_1 = \sigma_2$，则 w_0 属于类型2，且 $R(w_0) = R_2(w_0) = \lim_{k\to\infty} s_2 \circ \ldots \circ s_k(1)$。由归纳法即可得出结论。 □

1.3.5 Sturmian词的等价动力学与几何描述

在最后这一小节中，我们将介绍（不作证明）Sturmian词的其他动力学与几何构造。具体而言，我们将看到，Sturmian词编码了最小轨迹

- 正方形台面上的一颗台球，

- 在二维环面上线性流作用下的一个点，
- 圆上的旋转 ，

并且它们将具有无理斜率的 $\mathbb{R}^2$ 半直线“数字化”。

（我们回顾一下，若每个点都有一条稠密轨道，则称拓扑动力系统是极小的 。下一章中，我们将讨论极小性和无理性假设的重要性。）

我们首先定义上文提到的动力系统与几何系统，并说明如何用无限词对它们进行编码。

台球词。令 $x \in [0,1)^2$ 和 $\theta \in (\mathbb{R}_{>0})^2$。台球词 $w = w_{\text{billiard}}(x,\theta)$ 是对台球轨迹进行编码的无限二元词。台球最初位于位置 x，并以动量 θ 在正方形台面 $[0,1]^2$ 上出发。其第 n 个字母定义如下：

$$w[n] = \begin{cases} 1 \text{ if the } n\text{-th edge hit by the ball is vertical,} \\ 2 \text{ if the } n\text{-th edge hit by the ball is horizontal.} \end{cases}$$

（由于本讲中不会用到，我们选择不对撞到角点的轨迹进行编码。）例如，在图 1.1 中，台球词以 $w_{\text{billiard}}(x,\theta) = 12112\ldots$ 开头。

直线数字化。同样，令 $x \in [0,1)^2$ 和 $\theta \in (\mathbb{R}_{>0})^2$。用 $\mathcal{G} = \mathbb{Z} \times \mathbb{R} \cup \mathbb{R} \times \mathbb{Z}$ 表示 $\mathbb{R}^2$ 的整数格。切割序列 $w = w_{line}(x,\theta)$ 是对由 $x + t\theta,\ t \in \mathbb{R}_{>0}$ 参数化的半直线 $\mathcal{D}$ 进行数字化得到的无限二元词。其第n个字母定义为

$$w[n] = \begin{cases} 1 \text{ 若直线 } \mathcal{D} \text{ 与格点网格 } \mathcal{G} \text{ 的第 } n \text{ 个交点位于垂直分支上；否则。} \\ 2 \end{cases}$$

（我们不对经过格点交点的直线进行数字化。）例如：图 21.1 中的半直线可由一个以 $w_{line}(x,\theta) = 12112\ldots$ 开头的无限词进行数字化。解释：如果将字母 1 解释为向右的单位步，将字母 2 解释为向上的单位步，那么词 w 就是最能表示直线 $\mathcal{D}$ 的“阶梯线”。

环面上的线性流编码。令 $x \in [0,1)^2$ 和 $\theta \in (\mathbb{R}_{>0})^2$。“流词” $w = w_{flow}(x,\theta)$ 是一个无限二元词，它编码点 x 在环面 $\mathbb{R}^2/\mathbb{Z}^2 = [0,1)^2$ 上沿方向 θ 的线性流作用下的轨迹。若 P 表示庞加莱截面 $P = \{0\} \times [0,1) \cup [0,1) \times \{0\}$，则编码定义为

$$w[n] = \begin{cases} 1 \text{ 如果第 } n \text{ 次穿过 } P\text{，则是经由其垂直分支穿过；否则，} \\ 2 \end{cases}$$

（我们不对经过环面角点 (0,0) 的轨迹进行编码。）例如：图 1.1 中 x 的流轨迹由无限词 $w_{flow}(x,\theta)$ =12112... 编码。

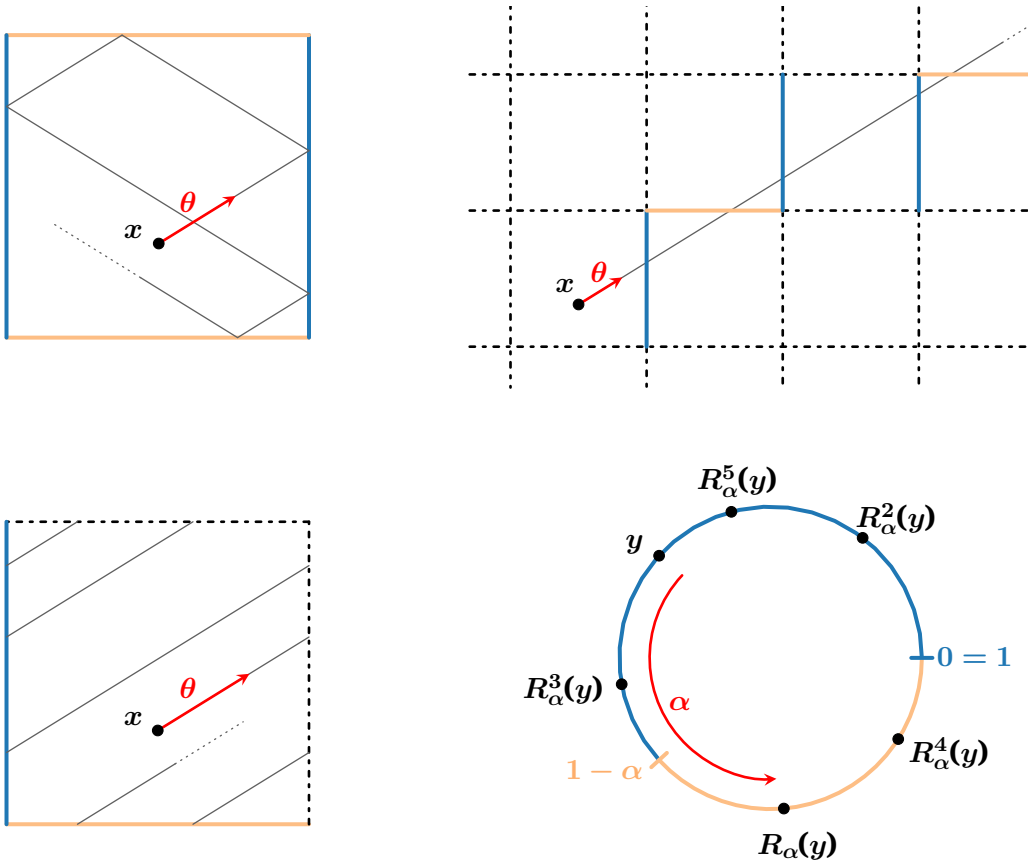


图 1.1：左上：台球轨迹。右上：直线的数字化。左下：环面上的线性流。右下：圆周旋转。

旋转编码。 角度为 $\alpha \in \mathbb{R}$ 的圆 $\mathbb{R}/\mathbb{Z} = [0,1)$ 上的旋转是映射 $R_\alpha : y \mapsto y + \alpha \bmod 1$。点 $y \in [0,1)$ 的轨迹由无限二元词 $w = w_{rotation}(y,\alpha)$ 编码，其第 n 个字母定义为

$$w[n] = \begin{cases} 1 \text{ if } R_\alpha^{n-1}(y) \in [0, 1-\alpha) \\ 2 \text{ otherwise.} \end{cases}$$

示例：在图 1.1 中，y 的轨迹由无限词 $w_{rotation}(y,\ \alpha)$ =12112..... 编码。

下一个定理确立了这些动力系统和几何系统与Sturmian词之间的联系。

定理 1.29。 设 w 是一个标准*Sturmian*词，其字母频率为 $(f_1、\ f_2)$。则

$$\begin{aligned} w &= w_{\text{billiard}}(0,(f_1,f_2)) \\ &= w_{\text{line}}(0,(f_1,f_2)) \\ &= w_{\text{flow}}(0,(f_1,f_2)) \\ &= w_{\text{rotation}}(f_2,f_2). \end{aligned}$$

这个定理为斯图姆词的存在性提供了另一种证明。它还有助于培养对其组合性质的直觉，有时也能让我们证明这些性质，正如下一个练习所示。

练习 1.30. 设 w 是一个Sturmian词。
1) 证明：若 u 是 w 的因子，则 $\texttt{mirror}(u)$（即将词 u 倒序书写）也是 w 的因子。

2) [挑战] 仅使用组合论论证证明相同的结果。

本章最后，我们将作两点说明，并给出定理 1.29 的证明。

说明 1.31。 根据定理1.18，我们知道 Sturmian词的字母频率满足 $f_1/f_2 \notin \mathbb{Q}$。这一条件是方形台球具有极小性以及环面上的线性流具有极小性的充要条件。它也等价于要求 $f_2 \notin \mathbb{Q}$（这是由于等式 $f_1/f_2 = (1-f_2)/f_2) = 1/f_2 - 1$），因此也等价于要求圆周旋转具有极小性。

注释 1.32. 我们几乎得到了集合相等式

$$\{\text{Sturmian words}\} = \{w_{\text{rotation}}(y,\alpha) \mid y \in [0,1), \alpha \in [0,1)\backslash\mathbb{Q}\}.$$

事实上，反向包含成立，而正向包含几乎成立：对于每一对可能的字母频率（f_1, f_2），只有可数多个斯图姆词缺失（见本章末的练习 1.34）。

定理 *1.29* 的证明。 我们来证明第二、第三和第四个等式。第一个等式将在下面的练习 1.33 中证明。

显然，我们总有 $w_{\text{billiard}}(x,\ \theta) = w_{\text{line}}(x,\ \theta)$。事实上，我们可以不让球反弹，而是等价地将台球桌镜像展开，让球沿同一条直线继续运动。（镜像展开台球桌的想法是平移曲面理论中的经典方法。）同样显然，对于每一对参数（x, θ），都有 $w_{\text{flow}}(x,\ \theta) = w_{\text{line}}(x,\ \theta)$：只需对平面、整数格和半直线按格 $\mathbb{Z}^2$ 取商即可。

为了确立第四个等式，我们将方向为 $\theta = (\theta_1,\ \theta_2) \in \mathbb{R}^2_{>0}$ 的线性流投影到与流方向正交的一条直线上，该流定义在环面 $\mathbb{R}^2/\mathbb{Z}^2 = [0,1)^2$ 上，并证明所得的一维动力系统是两个区间交换，或者等价地，是圆周旋转。图 1.2 展示了示意图。

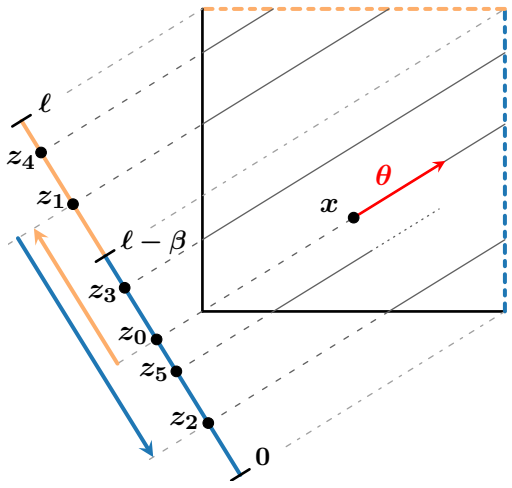


图 1.2：从二维环面上的线性流到圆上的旋转。

注意，投影后，- 单位正方形 $[0,1)^2$ 的像是一条斜线段（其长度记为 l）；- 正方形的上边是线段上方的橙色/浅灰色区间（其长度记为 β）；- 正方形的右边是线段下方的蓝色/深灰色区间（其长度为 $l-\beta$）；- 点 $x\in[0,1)^2$ 在环面上的流作用下的轨迹，是线段上的一列点，记为 $(z_n)_{n\in\mathbb{N}_{>0}}$。

我们将斜线段与 $[0,l)$ 视为同一对象，其中 0 位于底端点。我们断言：1) 序列 $(z_n)_n$ 是 z_1 在长度为 l 的圆 $\mathbb{R}/l\mathbb{Z}=[0,l)$ 上按角度 β 旋转所得的轨迹。

2) 流词 $w_{\text{flow}}(0,\theta)$ 等于旋转词 $w_{\text{rotation}}(y,\alpha)$，其中 $y=\alpha=\theta_2\theta_1+\theta_2$。

第一个断言可由以下观察得到：环面中的点 x 下一次穿过的截面是竖直的，当且仅当其投影 $z=:z_k$ 属于线段的蓝色／深灰色区间（即 $z_k\in[0,1-\beta)$）。在这种情况下，$z_{k+1}=z_k+\beta$。另一种情况下（即下一次穿过的截面是水平的，或者等价地，当 $z_k\in[1-\beta,1$ 时），我们有 $z_{k+1}=z_k+\beta-l$。两种情况下，$z_{k+1}=z_k+\beta$ 模 l，第一个断言得证。

接下来，我们将线段按因子 $1/l$ 进行缩放，以便得到单位圆上的旋转。记 $\alpha=\beta/l$ 和 $y_k=z_k/l$。于是，对于所有 $k\in\mathbb{N}_{>0}$

$$w_{\text{flow}}(x,\theta)[k]=\begin{cases}1 \text{ if } z_k\in[0,1-\alpha)\\ 2 \text{ otherwise,}\end{cases}$$

由此可得 $w_{\text{flow}}(x,\theta)=w_{\text{rotation}}(z_0,\alpha)$。最后，我们通过计算得出

$$\alpha=\frac{\beta}{l}=\frac{\theta_2}{\theta_1+\theta_2},$$

并且，点 $(0,0)$ 在 $[0,1)^2$ 上的投影为 α。 □

练习 1.33。设 $\alpha\in[0,1]\backslash\mathbb{Q}$。本练习旨在证明，$w_{\text{rotation}}(\alpha,\alpha)$ 是字母频率为 $(f_1,f_2)=(1-\alpha,\alpha)$ 的标准Sturmian词。这个等式完成了定理 1.29 的证明。

1）设 $y\in[0,1)$。证明旋转词 $w_{\text{rotation}}(y,\alpha)$ 的字母频率存在，且等于 $(1-\alpha,\alpha)$。

2）设 $y\in[0,1)$。证明 $w_{\text{rotation}}(y,\alpha)$ 是一个 Sturmian词。
3）证明：当且仅当一个 Sturmian词的所有有限前缀都是左特殊时，该词是标准的。

4）证明旋转词 $w_0=w_{\text{rotation}}(\alpha,\alpha)$ 的所有有限前缀都是左特殊的。5）得出结论。

练习 1.34。本练习旨在证明注记 1.32。设 $\alpha\in\mathbb{R}\backslash\mathbb{Q}$。
1）设 $x,y\in[0,1)$。证明：如果 $x\neq y$，那么 $w_{\text{rotation}}(x,\alpha)\neq w_{\text{rotation}}(y,\alpha)$。
2) 记 w_0 为字母频率为 $(1-\alpha,\alpha)$ 的标准Sturmian词。证明 $w=2\cdot w_0$ 是一个Sturmian词，但根据我们的定义，它不是旋转词。提示：我们在练习 1.33 中证明了 $w_0=w_{\text{rotation}}(\alpha,\alpha)$。

1.4 注释参考文献

有兴趣进一步了解斯图姆词的读者，不妨查阅：

- M. Morse 和 G. Hedlund 在20世纪40年代发表的历史性论文 [MH38, MH40],
- C. Coven 和 G. Hedlund 在20世纪70年代发表的历史性论文 [CH73a, CH73b],
- 21世纪初的两部专著中的相关章节：[Lot02，第 2] 章，以及 [PF02，第 6]章。

我们在本讲中对斯图姆词的介绍，与上述两本书章节中的介绍有所不同。事实上，我们直接从其作为复杂度为 $n \mapsto n+1$ 的词这一界定出发，确立了它们与连分数以及几何系统和动力系统之间的联系。我们没有使用词上的阿贝尔组合学（尤其没有使用平衡性）。

第2章

将 Morse–Hedlund 定理推广到 d 元字母表：Tijdeman 的一个定理（1999）

2.1 动机

本课程的第一章专门研究了斯图姆词，它们构成了一类引人注目的二元无限词。简要回顾如下：

- 无限词 $w \in$ 复杂度 $\mathbb{N}$ 是一个函数 $p_w : \mathbb{N} \to \mathbb{N}$，它对每个整数 n，统计 w中长度为 n 的因子（即子词）的数量。
- M. Morse 和 G. Hedlund 于 1938年提出的一个著名定理指出，一个无限词是最终周期的，当且仅当存在一个整数 n，使得 $p_w(n) \leq n$。因此，每个非最终周期词都必须满足：对所有 n，$p_w(n) \geq n+1$。
- 存在复杂度恰好为 $p_w(n) = n+1$ 且对所有 n都成立的无限词；它们称为斯图姆词。
- 斯图姆词由二元字母表构成（事实上，它们满足 $p_w(1) = 2$）。它们与算术、动力系统和几何之间存在着显著的联系。具体而言，它们与连分数密切相关，也可以等价地定义为无理旋转的符号编码、方形台球桌上的无理轨迹、最小二区间交换，以及具有无理斜率的直线的数字化。

因此，很自然会想知道，这些令人满意的结果在多大程度上可以推广到更大的字母表。

问题 2.1. 设 $d \geq 3$。是否存在 *Morse–Hedlund* 定理的 d元类比，也就是说，是否存在一个定理保证：低于某个给定的复杂度界时，无限d元词都是“平凡的”？

问题 2.2. 如果是这样，我们能否在 $d \geq 3$ 时，动态地并通过算术刻画具有最小复杂度的非平凡 d 元无限词集合？

第二章旨在使读者相信，蒂耶德曼定理为问题 2.1 提供了一个令人十分满意的答案，而问题 2.2 在本文写作时基本上仍未解决，值得进一步研究。

2.2 d元斯图姆词的推广示例 s

斯图姆词可以通过多种方式推广到 d元字母表：

1）从组合学角度出发，对特殊因子的集合施加条件，或将其视为在 d 字母字母表上通过特定替换复合而成的极限；

2）从算术角度出发，将其视为由 $d-1$ 维连分数算法生成的词（请注意，连分数有几种同样合理的定义，因此产生了几类不同的 d 元词）；3）从几何或动力系统角度出发，将其视为多边形台球或超立方台上的台球轨迹的编码，或环面 $\mathbb{R}^{d-1}/\mathbb{Z}^{d-1}$ 上旋转的编码，或 d 区间交换的编码，等等。

这些构造产生了各种经过深入研究的 d元无限词族，例如*Arnoux–Rauzy*词 [AR91], *Episturmian* 词 [DJP01, JP02, GJ09], 与 *Cassaigne–Selmer* 连分数算法相关的词 [CLL22],树状词[BDFD+15]、*S-adic*词 [PF02, Chapter 12],[BD14], 区间交换变换的自然编码 [Kea75, Yoc05, FZ08, Del16], 超立方台球词[AMST94, Bar95, Béd09], 等。它们都被视为 Sturmian 词的推广。

有趣的是，这些 d元 Sturmian词推广的复杂度函数并不都相同。例如，在三元字母表上，Arnoux–Rauzy词和 Cassaigne–Selmer词的复杂度为 $n \mapsto 2n+1$，非 Arnoux–Rauzy 的 Episturmian 词的复杂度严格低于 $n \mapsto 2n+1$，而几乎所有立方台球词的复杂度为 $n \mapsto n^2+n+1$。

2.3 d元词的最小复杂度是什么？（讨论“最小”应如何定义。）

从现在起，在第 2 章和第 3 章中，当我们说无限词 w 是“d元无限词”时，我们指的是其中实际出现了 d 个不同的字母。具体而言，我们要求 $p_w(1) = d$。

下一个定理断言，常函数 $p: n \mapsto d$ 是 d元无限词的绝对最小复杂度。

定理 2.3。 设 $d \geq 1$。设 $w \in \{1, \ldots, d\}^{\mathbb{N}}$ 是一个满足 $p_w(1) = d$的无限词。那么，对于每个 $n \in \mathbb{N}_{>0}$，都有$p_w(n) \geq d$ 。此外，存在一个 d元无限词 w ，使得对所有 $n \in \mathbb{N}_{>0}$，都有$p_w(n) = d$ 。

证明。 第一个断言由等式 $p_w(1) = d$以及复杂度p_w 是一个非递减函数这一事实得出。对于第二个断言，只需验证周期词 $w = (12\ldots d)^\omega$ 是一个具有所需复杂度的 d 元无限词。□

然而，定理 2.3 的意义有限：取 $d = 2$ 时，我们无法得到希望推广的 Morse–Hedlund 定理：“非最终周期二进制词的最小复杂度函数是 $n \mapsto n + 1$”。为了推广这一定理，我们至少需要继续将研究范围限定在非最终周期 d 元词上。正如我们将在接下来的段落中看到的，这一限制甚至还不够。

下面的定理断言，对于一个非最终周期的 d 元词，线性函数 $n \mapsto n + d - 1$ 是其最小复杂度。

定理2.4。 设 $d \geq 2$。令 $w \in \{1, \ldots, d\}^{\mathbb{N}}$ 为一个非最终周期词，且满足$p_w(1) = d$。则对每个 $n \in \mathbb{N}_{>0}$，都有 $p_w(n) \geq n + d - 1$ 。此外，存在一个非最终周期的 d 元词 w，使得对所有 $n \in \mathbb{N}_{>0}$，都有 $p_w(n) = n - d - 1$ 。

证明。 由于 w 是非最终周期的，它必定对每个长度都有至少一个右特殊因子；因此，其复杂度函数严格递增。由于 $p_w(1) = d$，直接归纳可得，对每个 $n \geq 1$，都有 $p_w(n) \geq n + d - 1$ 。

对于第二个断言，设 w_0 是二元字母表$\{d-1, d\}$上的Sturmian词，并考虑由拼接 $w := 12 \ldots (d-2)w_0$得到的 d元无限词。由于 w_0 是Sturmian词， w 是非最终周期的，并且具有所需的复杂度函数$n \mapsto n + d - 1$。事实上，对每个 $n \geq 1$，词 w 分别含有 $d - 2$ 个以字母$1, 2, \ldots, d-2$开头的因子，以及 $n+1$ 个源自Sturmian词w_0、仅由字母 $d - 1$ 和 d 写成的因子。 □

读者可能会对上述证明中的构造感到失望：我们只是选取了一个Sturmian词，并在其前面加上一些额外字母，以使其符合“d元词”的定义。这些额外字母甚至不会在 w中再次出现。我们距离斯图姆词的令人满意的推广还很远。

为了排除这些繁琐的例子，我们可以要求 w 中每个字母都出现无穷多次。我们甚至可以采取额外的预防措施，要求 w 中每个因子都出现无穷多次，且间隔有界。我们回顾一下，这一性质称为一致回归，而斯图姆词是一致回归的（练习 1.17 的目标就是证明这一性质，见第1章）。

问题 2.5. 是否存在一个非最终周期的一致回归 d 元词w，其复杂度为 $n \mapsto n + d - 1$?

答案是“是”，下面的练习便是证明。

练习2.6. 设 $w_0 \in \{1,2\}^{\mathbb{N}}$ 是一个 Sturmian词。记 w 为通过替换 σ 对 w_0 作出的映像： $1 \mapsto 31, 2 \mapsto 32$.1）证明 w 是一个非最终周期的一致回归三元词。2）证明 w 的复杂度为 p_w： $n \mapsto n+2$。（提示：研究 w 的右特殊因子，证明 $p_w(n+1) - p_w(n) = 1$。）3）将这一构造推广到 d 字母表，其中 $d \geq 4$。

再者，练习2.6中的构造令人沮丧：词 w 实在过于贫乏，难以视为斯图姆词的一种推广。

问题 2.7。 是否存在一个复杂度为 $n \mapsto n + d - 1$ 的非最终周期一致回归词，而且它并非对斯图姆词进行人为扭曲的产物?

答案是否定的。下一个定理断言，所有非最终周期的、一致回归的、复杂度为 $n \mapsto n + d - 1$ 的 d元词，都可以构造为斯图姆词在类似于练习2.6的代换下的像。

定理 2.8（S. Ferenczi与C. Mauduit，1997；见 [FM97], 引理4）。设 $d \geq 3$，且w 为一个无限词。词 w 是递归的，且复杂度为 $n \in \mathbb{N}_{>0} \mapsto n + d - 1$，当且仅当存在：

- 一个整数 $m \in \mathbb{N}$，- 一个 *Sturmian*词 $w_0 \in \{a, b\}^{\mathbb{N}}$，- 字母表 $\{1, \ldots, d\}$ 的一个划分 $= A \sqcup B \sqcup C$，其中 $A = \{a_1, \ldots, a_{N_A}\}$，$B = \{b_1, \ldots, b_{N_B}\}$，$C = \{c_1, \ldots, c_{N_C}\}$，且 $C \neq \emptyset$和 $A \sqcup B \neq \emptyset$，

使得 $w = S^m \sigma(w_0)$，其中 S 表示移位算子，σ 是由下式定义的替换

$$a \mapsto c_1 \ldots c_{N_C} a_1 \ldots a_{N_A}, \quad b \mapsto c_1 \ldots c_{N_C} b_1 \ldots b_{N_B}.$$

此外，在这种情况下，词 w 是一致回归的。

（该定理似乎此前已由 P. Alessandri 针对三元字母表证明，见 [Ale95],另见 [KT07]。）

因此，在文献中，复杂度为 $n \mapsto n + d - 1$ 的无限词，以及更一般地说，复杂度函数最终等于某个常数 k 的 $n + k$，被称为“拟 Sturmian 词” [Cas98]。

2.4 字母频率有理独立的条件。

在上一节中，我们看到，非最终周期性这一条件对于排除二元字母表上的平凡词非常有效，但对于较大字母表上的平凡词却不足以将其排除。本节旨在使读者相信，对于 d字母的字母表，无论 $d \geq 1$ 如何，适宜采用的条件是字母频率的 有理独立性。

回顾一下，实数 $x_1, \ldots, x_d$ 称为“有理独立”，是指其中任何一个都不能表示为其余实数的有理系数线性组合；等价地说，它们在域 $\mathbb{Q}$上张成的线性空间的维数为 d。在处理大小任意的字母表时，字母频率有理独立的条件应取代非最终周期性，这一观点由以下四个事实所支持。

事实 1. Morse–Hedlund 定理中有趣的部分（即：具有低复杂度的无限词是最终周期的）可以用字母频率的有理相关性来表述。具体而言，我们有：

命题 2.9. 设 $w \in \{1,2\}^{\mathbb{N}}$。若存在整数 $n \geq 1$ 使得 $p_w(n) \leq n$，则 w 具有有理相关的字母频率。

证明。在这一假设下，Morse–Hedlund 定理（第1章，定理1.3）保证词 w 是最终周期的；我们将它写作 $w = u \cdot p^\omega$。容易验证，w 中字母 1 和 2 的频率与其周期后缀 p^ω中的频率相同，也就是说，$f_1 = |p|_1/|p|$ 和 $f_2 = |p|_2/|p|$。显然，这些（有理）数是有理相关的。 □

事实 2. 斯图姆词的字母频率有理无关，这是它们的核心性质之一。 我们在第1章中回顾过，这一性质最初是通过它们与连分数的联系显现出来的。具体而言，定理1.18指出，任意 Sturmian词的字母频率 f_1/f_2 之比是一个无理数，这等价于断言 f_1 和 f_2 有理无关。这一性质也体现在斯图姆词的所有几何和动力学刻画中，正如注记 1.31 所强调的：斯图姆词编码圆的极小 旋转、极小 方形台球轨迹、二维环面上的极小 线性流，以及斜率为无理数 的平面半直线。

最后值得注意的是，根据定义以及命题 2.9，在具有有理无关字母频率的二元词集合中，Sturmian词恰好是具有最小复杂度的词。

事实 3。通常被视为 Sturmian词的良好组合或动力学推广的词类，也具有有理无关字母频率。例如，对编码超立方体台球桌中最小轨道的词（参见例如 [AV24]）来说，情况就是如此；Arnoux–Rauzy词 [And21b,DHS22], 严格拟Sturmian词 [And21a，第 4], 章以及 Cassaigne–Selmer词 [CLL17, CLL22]也同样如此。

事实 4。不存在 d 元拟Sturmian词，其 $d \geq 3$具有有理无关字母频率。这将是下一节所介绍的 Tijdeman定理的一个推论。对于具有复杂度$n \in \mathbb{N}_{>0} \mapsto n + d - 1$的递归的 d 元词，这一事实也可以由定理 2.8 推导出来，正如下方练习所示。

练习 2.10. 设 $d \geq 3$。设 w 是一个复杂度为 $n \in \mathbb{N}_{>0} \mapsto n + d - 1$ 的递归 d 元词。证明 w 存在字母频率，且这些字母频率张成一个维数为 2 的有理线性空间。

现在，对于任意大小的字母表，字母频率的有理独立性似乎可以合理地替代非最终周期性这一条件，于是一个新的问题随之而来。

问题 2.11. 设 $d \geq 3$。具有有理独立字母频率的 d 元词，其最小复杂度是多少？

R. Tijdeman 于 1999 年完整地回答了这个问题。

2.5 Tijdeman 的一个定理

遗憾的是，Tijdeman 的答案大多未受关注（例如，在 [BB13] 中就没有提及），直到2022年，J. Cassaigne 和本讲义作者重新发现了它。

Tijdeman 定理指出，字母频率有理无关的无限 d 元词，其最小复杂度为线性函数 $n \mapsto (d-1)n+1$

定理2.12 (Tijdeman, 1999, [Tij99])。 设 $d \geq 1$。设 $w \in \{1, \ldots, d\}^{\mathbb{N}}$ 是一个无限词，其 $p_w(1) = d$ 且字母频率有理无关。那么，对于每个 $n \in \mathbb{N}_{>0}$，都有 $p_w(n) \geq (d-1)n+1$ 。此外，存在一个 d 元无限词 w，其字母频率有理无关，并且对于所有 $n \in \mathbb{N}_{>0}$，都有 $p_w(n) = (d-1)n+1$ 。

当然，当 $d = 2$ 时，我们重新得到促使我们开展这项研究的命题 2.9。

2.6 试论 Tijdeman 意义下具有最小复杂度的 d 元无限词的刻画。

本小节旨在开始研究达到 Tijdeman 最小复杂度的无限词。首先，我们断言它们都是递归的。

命题 2.13. 设 $d \geq 1$，且 w 是一个复杂度为 $p_w(n) = (d-1)n+1$、字母频率有理无关的无限词。那么 w 是递归的。

证明. 设 w 是一个复杂度为 $p_w(n) = (d-1)n+1$、字母频率 $f_1, \ldots, f_d$ 有理无关的无限词。我们用反证法。假设 w 不是递归的，并记 u 为在 w 中仅出现有限次的一个因子。设 w' 是 w 的一个后缀，其中不含因子 u。注意，w' 仍然是一个无限 d 元词（事实上，如果其中缺少某个字母，那么该字母在 w' 中的频率为 0，这不可能），且 w' 的字母频率与 w 完全相同（事实上，字母频率不受有限前缀影响）。此外，对于 $n = |u|$，有 $p_{w'}(n) \leq p_w(n) - 1 < (d-1)n+1$。因此，$w'$ 是一个字母频率有理无关的无限 d 元词，其复杂度低于 Tijdeman 定理 2.12 给出的最小下界。由此得到矛盾。

□

自然会想知道，哪些经典无限词类适用于 Tijdeman 定理

问题 2.14. 设 $d \geq 3$。哪些 d元无限词具有有理无关字母频率，且复杂度为 $p_w : n \in \mathbb{N}_{>0} \mapsto (d-1)n+1$?

我们回顾一下，当 $d = 1$ 和 $d = 2$时，这些词类已得到完全刻画 .

- 对于 $d = 1$，它就是单词 $\{w = 111111111111...\}$。事实上，它是唯一的 1 元词，并且显然具有有理无关字母频率和所需的复杂度 $n \mapsto 1$。

- 对于 $d = 2$，它们恰好是斯图姆词。事实上，根据定义，斯图姆词就是复杂度为 $n \mapsto n+1$ 的无限词；此外，由定理1.18可知，它们的字母频率有理无关。

相比之下，迄今为止，当 $d \geq 3$时，我们只知道部分答案。

- 对于 $d = 3$，我们知道，具有 Tijdeman 最小复杂度的词类至少包含所有 Arnoux–Rauzy词、所有 Cassaigne–Selmer词，以及几乎所有三段区间交换变换的编码：事实上，这些都是复杂度为 $n \mapsto 2n+1$、且字母频率有理无关的三元词。

- 更一般地，对于任意字母表大小 d，具有 Tijdeman 最小复杂度的词类至少包含所有严格episturmian词，以及几乎所有 d 段区间交换变换的编码。

有趣的是，以上提到的所有词类（即 Sturmian 词、Arnoux– Rauzy 词、Cassaigne–Selmer 词和严格episturmian词，以及正则区间交换变换的编码）都属于一类称为树状词的无限词（将在下一章定义）。事实上，我们对 Tijdeman 定理的另一种证明将顺带说明，在 Tijdeman 意义下具有最小复杂度的所有无限词都是树状的。

定理 2.15 (Andrieu, Cassaigne, 2022年，未发表)。 设 $d \geq 1$。所有具有有理无关字母频率且复杂度为 $n \mapsto (d-1)n+1$ 的无限 d 元词都是树状的。

由于每个 d 元树状词的复杂度为 $n \mapsto (d-1)n+1$ （见 [BDFD+15] 或第3章中的命题3.19），我们可得出以下刻画。

推论2.16。 设 $d \geq 1$。具有有理无关字母频率且复杂度为 $n \mapsto (d-1)n+1$ 的无限 d 元词集合，与具有有理无关字母频率的 d 元树状词集合相同。

遗憾的是，并非所有树状词都具有有理无关字母频率。值得注意的是，某些正则 d 区间交换变换的编码并不满足这一点：例如参见[Via06，例3.3]。因此，截至 $d \geq 3$ 本文撰写之时，如何刻画具有 Tijdeman 最小复杂度的无限 d 元词，仍然是一个未解问题；这方面可借鉴对斯图姆词所作的优美组合、几何和动力学刻画。

2.7 Tijdeman 定理的原始版本与强化版本

Tijdeman 的原始定理实际上比我们所陈述的定理2.12更强。它形式化并量化 了如下思想：若无限词 w 的复杂度增长得不够快，则 w 是代数贫乏的。在这最后一截面中，我们给出 Tijdeman 定理的完整原始版本，并引入它的一个新的强化版本。

为了陈述 Tijdeman 定理的原始形式——其中甚至不要求字母频率的存在——我们首先需要引入一些记号。给定一个 d元无限词 w，我们考虑序列

$$\Big(\frac{1}{n}(|\mathrm{pref}_n(w)|_1, \ldots, |\mathrm{pref}_n(w)|_d)\Big)_n$$

它位于紧集 $[0,1]^d$中。若该序列收敛，则依定义，其极限即为 w中字母的频率。否则，我们用 F_w 表示其所有子序列极限的集合：

$$F_w := \Big\{(f_1,\ldots,f_d) \in [0,1]^d \;\Big|\; \begin{array}{c}\text{there exists an increasing sequence } (n_l)_l \subseteq \mathbb{N} \text{ s.t.} \\ \forall\, i \in \{1,\ldots,d\},\ |\mathrm{pref}_{n_l}(w)|_i/n_l \underset{l\to\infty}{\longrightarrow} f_i\end{array}\Big\}. \quad (2.1)$$

现在，我们将整数 $\delta_w \in \{1,\ldots,d\}$ 定义为 w 中各字母的最小无理度：

$$\delta_w := \min\big\{\dim \mathrm{Span}_{\mathbb{Q}}(f_1,\ldots,f_d) \,|\, (f_1,\ldots,f_d) \in F_w\big\}.$$

定理2.17 （Tijdeman 1999年，原始版本）。 设 w 是一个无穷 d 元无限词。记 δ_w 为 w 中各字母的最小无理度。则对于每个$n \geq 1$,

$$p_w(n) \geq (\delta_w - 1)(n-1) + d.$$

事实上，可以通过将字母无理性的最小次数 δ_w 替换为字母无理性的最大次数 来加强 Tijdeman 的原始定理：

$$\Delta_w := \max\big\{\dim \mathrm{Span}_{\mathbb{Q}}(f_1,\ldots,f_d) \,|\, (f_1,\ldots,f_d) \in F_w\big\}. \quad (2.2)$$

在下面的习题2.21中，我们给出一个无限词 w ，其中字母无理性的最小次数与最大次数不同。

定理2.18 (Andrieu, Cassaigne, 2022年,未发表)**.** 设 w 为一个无限 d元词。用 Δ_w 表示 w 中字母的最大无理度。则对每个$n \geq 1$,

$$p_w(n) \geq (\Delta_w - 1)(n-1) + d.$$

定理2.18将在第3章中证明。

我们以两条注记和一道习题结束本章。

注记2.19. 若 w 具有字母频率，并且这些频率有理无关，则定理2.17与定理2.18都归结为定理2.12。事实上，在此情形下，我们有 $\Delta_w = \delta_w = d$，由此立即可得

$$(\Delta_w - 1)(n-1) + d = (d-1)n + 1.$$

注2.20。 从其证明可知，定理2.18也可等价地表述如下：“若存在整数 $k \geq 1$ 和长度 $m \geq 0$ ，使得 $p_w(m+1) - p_w(m) \leq k-1$，则 $\Delta_w \leq k$”，或者等价地，“$\Delta_w \leq \min_{m \geq 0}\{ p_w(m+1) - p_w(m) + 1\}$”。这种表述应理解为“一个词最终周期，当且仅当其复杂度函数局部常值”这一命题的 d元对应版本。正如我们在第1章注1.4中讨论的，该命题等价于 Morse–Hedlund 定理。

习题2.21 （构造一个满足 $\delta_w < \Delta_w$的无限二元词）。 令(u_n)为按如下方式构造的有限词序列。首先， $u_0 = 1$。然后，对每个偶整数 n，我们在 u_n末尾连接足够多个0，使得 u_{n+1} 中0和1的数量相同（这总是可行的，因为对每个偶整数 n， u_n 中1的数量都多于0的数量）。对于每个奇整数 n，我们在 u_n末尾连接 $n \times |u_n|$ 个1，使得 u_{n+1} 的长度是 u_n的$n+1$ 倍。由此得到 $u_0 = 1$、 $u_1 = 10$、 $u_2 = 1011$、 $u_3 = 101100$、 $u_4 = 10110011111111111111111111$等。根据构造，有限词序列$(u_n)$收敛于一个无限词，我们将其记为 w:

$$w = 10110011111111111111111100000000000000000...$$

1）证明 w 不具有字母频率，并且（0，1）和（$1/2$，$1/2$）属于 F_w （如（2.1）中所定义）。

2）证明实际上 $F_w = \{ (t,\ 1-t) :\ 0 \leq t \leq 1/2\}$。推导出 $\delta_w = 1$ 和 $\Delta_w = 2$。

现在，我们逐一将 w 中的每个1替换为斐波那契词 $w_{Fibo} = 12112121121121211212112112121121121...$ 中依次出现的字母（我们回顾一下，斐波那契词在例1.23中定义过，并且我们已经证明它的字母频率存在且满足 $f_1/f_2 = \varphi$，其中 φ 表示黄金比例。）由此我们得到一个由字母0、1和2组成的无限词:

$$w' = 10210012121121121211212100000000000000000...$$

3) 证明 $F_{w'} = \{(t,(1-t)f_1,\ (1-t)f_2) :\ 0 \leq t \leq 1/2\}$。由此推出 $\delta_{w'} = 2$ 与$\Delta_{w'} = 3$。验证定理2.17仅保证 $p_{w'}(n) \geq n+2$，而定理2.18保证 $p_{w'}(n) \geq 2n+1$。

第3章

Tijdeman 定理及其推论的另一种代数证明（2022年）

3.1 引言

在这第三也是最后一章中，我们给出1999年 Tijdeman 的原始定理的另一种证明。该证明由 J. Cassaigne 与作者于2022年得到，但尚未发表。它实际上建立了一个比 Tijdeman 的原始定理略强的结果，即：

定理 3.1（Andrieu, Cassaigne；在第2章中亦编号为定理2.18）。 设 w为一个无限的 d 元词。用 Δ_w 表示 w中字母的最大无理度。则对每一个 $n \geq 1$,

$$p_w(n) \geq (\Delta_w - 1)(n-1) + d.$$

我们回顾，在上述定理中，并不要求 w 中字母的频率存在；相反，我们使用它们的“伪频率”集合 F_w ，即序列 $\left(\frac{1}{n}(|\mathrm{pref}_n(w)|_1, \ldots\ldots, |\mathrm{pref}_n(w)|_d)\right)_n$ 的所有子序列极限所构成的集合。于是，“字母的最大无理度”按定义则为

$$\Delta_w := \max\left\{ \dim \mathrm{Span}_{\mathbb{Q}}(f_1, \ldots, f_d) \,|\, (f_1, \ldots, f_d) \in F_w \right\}.$$

我们的证明依赖于线性代数（尤其是秩-零化度定理以及对某些特征子空间维数的研究），并将其应用于一个特定的矩形矩阵，我们称之为“流矩阵”。这类矩阵与词组合学中两个经典且相互关联的对象密切相关：*Rauzy* 图 和扩展图。相比之下，Tijdeman 的原始证明主要是组合性的。

此外，这一替代证明中所发展的技术使我们能够建立如下新结果，这标志着在刻画 Tijdeman 意义下具有最小复杂度的 d 元词方面向前迈进了一步。

定理3.2。（Andrieu, Cassaigne；在第2章中也编号为定理 2.15）。 令 $d \geq 1$。每个具有有理无关字母频率且复杂度为$n \mapsto$（$d-1$）$n+1$ 的无限 d 元词都是树状的。

本章分为四个小节。在第3.2节中，我们介绍经典的 Rauzy 图概念以及新的流矩阵概念。在第3.3节中，我们证明定理 3.1。在第3.4节中，我们将我们的证明与 Tijdeman 的证明进行比较。最后，在第3.5节中，我们回顾扩展图和树状性的经典概念，并运用定理 3.1 的证明来证明定理3.2。

3.2 预备知识：Rauzy 图与流矩阵的定义

3.2.1 Rauzy 图

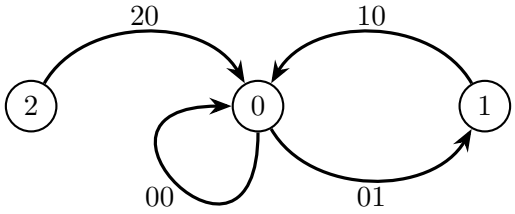
Rauzy 图是 Rauzy 在 [Rau83] 中提出的诸多出色想法之一。它们是词组合学中经过充分研究且十分有用的工具。根据定义，无限词 w 在给定长度 $n \in \mathbb{N}$ 下的“Rauzy 图”是有向图 $\mathcal{G} = (V, E)$，其：

- 顶点是 w 的长度为 n 的因子（$V := \mathcal{L}_n(w)$）；
- 边是 w 的长度为 $n+1$ 的因子（$E := \mathcal{L}_{n+1}(w)$）。具体而言，由 $v \in \mathcal{L}_{n+1}$ 标记的边将 v 的前 n 个字母所构成的词连接到由其最后 n个字母所构成的词。

例3.3. 最终周期词的 Rauzy 图

$$w = 2(010)^\omega = 20100100100100 10...$$

对于长度 $n = 1$ 为：



事实上，不难验证$\mathcal{L}_1(w) = \{0, 1, 2\}$ 且$\mathcal{L}_2(w) = \{00, 01, 10, 20\}$。

注 3.4. 根据定义，若对每一对顶点 u 与 u'，都存在从 u到 u' 的路径或从 u' 到 u的路径，则有向图是“半连通”的。每个无限词的 Rauzy 图都是半连通的。事实上，当沿 w从左向右滑动大小为 n的窗口时，我们沿着其长度为n 的 Rauzy 图上的一条路径行进，该路径访问每一个顶点。

我们建议不熟悉 Rauzy 图的读者思考以下两个练习。

练习 3.5. 根据定义，如果对于任意一对顶点 u 和 u'，都存在一条从 u 到 u'的路径，那么有向图就是“强连通”的。
1）所有无限词的 Rauzy 图总是强连通的吗?
2）证明：给定无限词 w 的所有 Rauzy 图强连通，当且仅当
且仅当 w 是递归的。
3) 如何根据无限词 w 的 Rauzy 图来解释其一致回归性?

习题3.6。
1) 画出斐波那契词在长度 4 时的 Rauzy 图。（我们记得，斐波那契词是在第1章习题1.23中定义的。）
2）如何在这幅 Rauzy 图上识别长度为 4 的右特殊因子和左特殊因子？3）设 $n \in \mathbb{N}$ 和 w 是任意一个 Sturmian 词。描述 w 的长度为 n 时 Rauzy 图的形状。

3.2.2 流矩阵

设 w 是一个无限词， n 是一个非负整数。根据定义，w 的长度为 n 时的“流矩阵”是矩形矩阵 M，其指标分别取自 $u \in \mathcal{L}_n(w)$ 和 $v \in \mathcal{L}_{n+1}(w)$，定义如下：

$$M_{u,v} := \begin{cases} 1 & \text{若 } v \text{ 以 } u\text{开头但不以它结尾，} -1 \quad v \qquad\qquad u \text{ 若以它结尾但不以它开头，} \\ 0 & v \qquad\qquad u \text{ 若既不以它开头也不以它结尾，或者以 } u\text{开头且结尾。} \end{cases}$$

其大小为 $p_w(n) \times p_w(n+1)$，按Rauzy图的记号，它也等于 $|V| \times |E|$。为了显式表示它，我们在集合$\mathcal{L}_n(w)$与$\mathcal{L}_{n+1}(w)$上固定一个任意次序。在本课程中，我们始终选择字典序。

例3.7. 承接例3.3，最终周期词 $w = 2(010)^\omega = 20100100100100 10\ldots$ 在长度 $n = 1$下的流矩阵为：

$$M = \begin{pmatrix} 0 & 1 & -1 & -1 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

由同一参数的Rauzy图很容易导出流矩阵：在由顶点 $u \in \mathcal{L}_n(w)$索引的行上，每条入边贡献 -1，每条出边贡献 $+1$。

“流矩阵 ”这一名称源自水力和电路中的基尔霍夫节点定律。流矩阵不同于图论中的经典邻接矩阵 ，后者始终是方阵。

注 3.8. 流矩阵可以理解为 w在长度为 n时的“右扩展矩阵和左扩展矩阵”之差： $M = R - L$, 其中

$$R_{u,v} := \begin{cases} 1 & \text{if } v \text{ starts with } u, \\ 0 & \text{otherwise,} \end{cases}$$

是表示 w 的长度为 n 的因子的右扩展的矩阵（事实上， $u \in \mathcal{L}_n(w)$ 的右扩展恰好是满足 $R_{u,v} = 1$) 的因子 $v \in \mathcal{L}_{n+1}(w)$）， 并且

$$L_{u,v} := \begin{cases} 1 & \text{if } v \text{ ends with } u, \\ 0 & \text{otherwise,} \end{cases}$$

类似地，是表示 w 的长度为 n 的因子的左扩展的矩阵。

特别地，根据它们的定义可得：
- R 和 L 的每一列都恰好有一个等于 1 的元素。

- R 的每一行至少有一个等于 1 的元素。如果 w 是递归的，那么L 的每一行也至少有一个等于 1 的元素。相反，如果 w 不是递归的，那么当 n充分大时， L 恰好有一行全为零（如例3.7所示）。

- 当且仅当矩阵 R （分别是 L）中以 u为索引的行包含多个等于 1 的项时，因子 $u \in \mathcal{L}(w)$ 是右特殊的（分别是左特殊的）。

现在我们已准备好证明定理 3.1。

3.3 定理 3.1 的证明

3.3.1 证明的总体结构

设 $d \geq 1$。设 w 是一个无限的 d 元词（回顾一下，这意味着 $p_w(1) = d$）。我们用反证法并结合极小性来证明。反证法地，假设存在一个长度 $n \geq 1$ ，使得

$$p_w(n) < (\Delta_w - 1)(n-1) + d, \tag{3.1}$$

其中 Δ_w 表示 w 中字母的最大无理度。（注意，要使不等式成立，必须有 $n \geq 2$。）令 $m+1$ 为使不等式 (3.1) 成立的最小长度 n。由 $m+1$ 的极小性，我们得到以下两重估计

$$\begin{cases} p_w(m) \geq (\Delta_w - 1)(m-1) + d, \\ p_w(m+1) < (\Delta_w - 1)m + d \end{cases} \tag{3.2}$$

由此我们推得

$$p_w(m+1) - p_w(m) \leq \Delta_w - 2. \tag{3.3}$$

根据 Δ_w 的定义，存在一个实数元组 $(f_1, \ldots, f_d) \in [0,1]^d$（称为字母的伪频率），使得 $\dim \mathrm{Span}_{\mathbb{Q}}(f_1, \ldots, f_d) = \Delta_w$，并且

$$\Big(\frac{1}{n_l}(|\mathrm{pref}_{n_l}(w)|_1, \ldots, |\mathrm{pref}_{n_l}(w)|_d)\Big)_l \longrightarrow_{l\to\infty} (f_1, \ldots, f_d), \tag{3.4}$$

对于某个递增整数序列 $(n_l)_l$。从序列 $(n_l)_l$ 中抽取一个子序列后，我们还可以假设，每个长度为 $m+1$ 的因子 v（估计式 (3.2) 中的参数为 m ）都具有一个伪频率 f_v，也就是说

$$\Big(\frac{|\mathrm{pref}_{n_l}(w)|_v}{n_l}\Big)_l \longrightarrow_{l\to\infty} f_v. \tag{3.5}$$

证明的目标是表明长度为$m+1$的因子的伪频率张成一个维数至多为 $\Delta_w - 1$的有理线性空间，并由此推出字母（$f_1, \ldots, f_d$）的伪频率同样张成一个维数至多为$\Delta_w - 1$的有理线性空间。这将与（3.4）中对伪频率族（$f_1, \ldots, f_d$）所作的选择相矛盾。

将 w 的长度为 m 的流矩阵记为 M。证明的核心在于研究 M 的核：首先将其视为一个具有有理系数的矩阵，然后视为一个具有实系数的矩阵。证明分为以下三个引理，并将在接下来的三个小节中予以证明。

引理 3.9。 设 M 为 w 的长度为 m 的流矩阵。那么，长度为-$m+1$ 的因子的伪频率向量满足

$$(f_v)_{v\in\mathcal{L}_{m+1}(w)} \in \ker_{\mathbb{R}}(M).$$

引理 3.10。 设 M 为长度为 m 的无限词 w 的流矩阵。那么 M 的核的维数满足

$$\dim \ker_{\mathbb{Q}}(M) \leq \Delta_w - 1.$$

回顾一下，$\ker_{\mathbb{R}}(M)$ 和 $\ker_{\mathbb{Q}}(M)$ 分别表示所有满足 $Mx = 0$ 的向量 $x \in \mathbb{R}^{\mathcal{L}_{m+1}(w)}$ 和 $x \in \mathbb{Q}^{\mathcal{L}_{m+1}(w)}$ 的集合。它们分别构成 $\mathbb{R}^{\mathcal{L}_{m+1}(w)}$ 和 $\mathbb{Q}^{\mathcal{L}_{m+1}(w)}$ 的线性子空间。

引理 3.11。 由字母伪频率张成的有理线性空间满足

$$\dim \mathrm{Span}_{\mathbb{Q}}(f_1, \ldots, f_d) \leq \Delta_w - 1.$$

若引理 3.11 成立，则定理 3.1 随之成立。事实上，我们同时有 $\dim \mathrm{Span}_{\mathbb{Q}}(f_1, \ldots, f_d) = \Delta_w$（根据字母伪频率（$f_1, \ldots, f_d$）的构造），以及 $\dim \mathrm{Span}_{\mathbb{Q}}(f_1, \ldots, f_d) < \Delta_w$。矛盾。

3.3.2 引理 3.9 的证明

设 $u \in \mathcal{L}_m(w)$ 和 $l \in \mathbb{N}$。由于无限词 $w \in \{1, \ldots, d\}^{\mathbb{N}}$ 中每次出现的 u，除非 w 以 u 开头时的第一次出现，否则前面都有一个字母，从而构成一个长度为 $m+1$ 的因子 v，因此我们得到一般估计：

$$|\mathrm{pref}_{n_l}(w)|_u = \sum_{\substack{v\in\mathcal{L}_{m+1}(w)\\ \text{s.t. } v \text{ ends with } u}} |\mathrm{pref}_{n_l}(w)|_v \quad + \quad \begin{cases} 1 & \text{if } w \text{ starts with } u, \\ 0 & \text{otherwise.} \end{cases} \tag{3.6}$$

用类似的论证，我们得到第二个估计：

$$|\mathrm{pref}_{n_l}(w)|_u = \sum_{\substack{v\in\mathcal{L}_{m+1}(w)\\ \text{s.t. } v \text{ starts with } u}} |\mathrm{pref}_{n_l}(w)|_v \quad + \quad \begin{cases} 1 & \text{if } \mathrm{pref}_{n_l}(w) \text{ ends with } u, \\ 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

由这些估计以及 w 的长度为 $m+1$ 的因子的伪频率定义（3.5），可推得每个长度为 m 的因子 u 也有一个伪频率 f_u：

$$\Big(\frac{|\mathrm{pref}_{n_l}(w)|_u}{n_l}\Big)_l \longrightarrow_{l\to\infty} f_u,$$

它满足以下守恒定律：

$$f_u = \sum_{v\in\mathcal{L}_{m+1}(w) \text{ s.t. } v \text{ starts with } u} f_v = \sum_{v\in\mathcal{L}_{m+1}(w) \text{ s.t. } v \text{ ends with } u} f_v. \tag{3.8}$$

注 3.12。 在一个形状为 w 长度为 m 的 Rauzy 图的电路中，若标记为 $v \in \mathcal{L}_{m+1}(w)$ 的支路上的电流强度为 f_v，则这一守恒定律恰好就是基尔霍夫节点定律。

(3.8) 中的第二个等式显然等价于

$$\sum_{v\in\mathcal{L}_{m+1}(w) \text{ s.t. } v \text{ starts with } u} f_v - \sum_{v\in\mathcal{L}_{m+1}(w) \text{ s.t. } v \text{ ends with } u} f_v = 0$$

并且对每个 $u \in \mathcal{L}_m(w)$ 都成立。因此，根据流矩阵 M 的定义，我们有

$$M \cdot (f_v)_{v\in\mathcal{L}_{m+1}(w)} = 0.$$

由于向量 $(f_v)_{v\in\mathcal{L}_{m+1}(w)}$ 先验上具有实数项，我们得出

$$(f_v)_{v\in\mathcal{L}_{m+1}(w)} \in \ker_{\mathbb{R}}(M).$$

引理 3.9 得证。

3.3.3 引理 3.10 的证明

令 M 为无限词 w 在长度 m下的流矩阵。回忆一下，流矩阵的元素属于$\{-1, 0, 1\} \subseteq \mathbb{Q}$，其大小为 $p_w(m) \times p_w(m+1)$。

对 M 及其转置矩阵应用秩-零化度定理，可得

$$\begin{cases} \dim \ker_{\mathbb{Q}}(M) = p_{m+1}(w) - \mathrm{rk}(M), \\ \mathrm{rk}(M) = \mathrm{rk}(M^t) = p_m(w) - \dim \ker_{\mathbb{Q}}(M^t); \end{cases}$$

由此，利用估计式 (3.3)，我们得到

$$\begin{array}{rcl} \dim \ker_{\mathbb{Q}}(M) & = & p_w(m+1) - p_w(m) + \dim \ker_{\mathbb{Q}}(M^t) \\ & \leq & \Delta_w - 2 + \dim \ker_{\mathbb{Q}}(M^t). \end{array} \tag{3.9}$$

现在我们将证明 $\dim \ker_{\mathbb{Q}}(M^t) = 1$，更确切地说，证明 $\ker_{\mathbb{Q}}(M^t) = \mathrm{Span}_{\mathbb{Q}}((1, \ldots, 1)^t)$。证明这一点后，上述不等式便成为

$$\dim \ker_{\mathbb{Q}}(M) \leq \Delta_w - 1,$$

从而证明引理 3.10。

首先，根据构造，M 每一列的元素之和都等于 0（这也可由注 3.8 直接得出）；因此，$(1, \ldots, 1)^t \in \ker_{\mathbb{Q}}(M^t)$。反过来，我们要证明，若 $x \in \ker_{\mathbb{Q}}(M^t)$，则 $x = (x_u)_{u\in\mathcal{L}_m(w)}$ 的所有坐标都相等。（换言之，我们要证明 $\ker_{\mathbb{Q}}(M^t) \subseteq \mathrm{Span}_{\mathbb{Q}}\{(1, \ldots, 1)^t\}$。）令 $u \neq u' \in \mathrm{L}_m(w)$。由于无限词 w 的长度为 m 时的 Rauzy 图 G 是半连通的（所有无限词的 Rauzy 图都是如此，见注 3.4），因此 G 中存在一条从 u 到 u'的路径，或者一条从 u' 到 u的路径。不妨设路径从 u 到 u'。将从 u 到 u'的连续边依次记为 $v_1, \ldots, v_l$，将依次经过的顶点（包括端点）记为 $u_0, \ldots, u_l$。这条路径可表示如下：

$$u = u_0 \underset{v_1}{\to} u_1 \underset{v_2}{\to} u_2 \underset{v_3}{\to} \cdots \to u_{l-1} \underset{v_l}{\to} u_l = u'. \tag{3.10}$$

由于 $u' \neq u$，经过适当缩短路径后，我们不妨假设顶点 $u_0, \ldots\ldots, u_l$ 两两不同。

现在，根据 Rauzy 图的定义，对于每个 $i \in \{1, \ldots, l\}$，因子 v_i 以u_{i-1} 开头，并以 u_i结尾。因此，在由因子 v_i索引的 M^t 行中，恰好有两个非零元素：位置 u_i 处的元素“$+1$”，以及位置 u_{i-1} 处的元素 “-1”，

它们彼此不同。最后，由于 $x \in \ker_{\mathbb{Q}}(M^t)$，我们有 $x_{u_i} - x_{u_{i-1}} = 0$。等式$x_{u_{i-1}} = x_{u_i}$对每个 $i \in \{1,\ldots,l\}$ 都成立，因此它沿着路径 (3.10) 传递，由此得到所需的等式 $x_u = x_{u'}$。由此我们证明了$x \in \ker_{\mathbb{Q}}(M^t)$ 的所有坐标都相等，进而可以推知 $\ker_{\mathbb{Q}}(M^t)$ 包含于 $\mathrm{Span}_{\mathbb{Q}}((1,\ldots,1)^t)$，事实上两者相等。因此，$\dim \ker_{\mathbb{Q}}(M^t) = 1$，等式 (3.9) 变为

$$\dim \ker_{\mathbb{Q}}(M) \leq \Delta_w - 1.$$

引理 3.10 已得到证明。

3.3.4 引理 3.11 的证明

对于此证明，我们建议读者注意每个对象所属的类别（标量、具有有理数项或实数项的向量等）。

记 $\delta = \dim \ker_{\mathbb{Q}}(M)$（根据引理 3.10，我们已经知道 $\delta \leq \Delta_w - 1$）。令 $b_1,\ldots,b_\delta \in \mathbb{Q}^{\mathcal{L}_{m+1}(w)}$ 为 $\ker_{\mathbb{Q}}(M)$ 的一个基，即，将 M 视为具有有理数项的矩阵时，其核的一个基。证明的关键是线性代数中的以下一般性结果：由于 $\mathbb{Q} \subseteq \mathbb{R}$，向量族 $b_1,\ldots,b_\delta$ 也构成 $\ker_{\mathbb{R}}(M)$ 的一个基，即，将 $\ker(M)$ 视为具有实数项的矩阵。（为完整起见，我们将在下一小节回顾这一一般性结果。）因此，由于根据引理 3.9，伪频率向量 $f := (f_v)_{v\in\mathcal{L}_{m+1}(w)}$ 属于 $\ker_{\mathbb{R}}(M)$，所以存在实数 $\lambda_1,\ldots,\lambda_\delta \in \mathbb{R}$，使得

$$f = \sum_{i=1}^{\delta} \lambda_i b_i.$$

这个向量等式等价于以下 $p_{m+1}(w)$ 个标量等式：对每个$v \in \mathcal{L}_{m+1}(w)$，都有

$$f_v = \sum_{i=1}^{\delta} \lambda_i b_{i,v}$$

其中 $b_{i,v}$ 表示向量 b_i 中由 v索引的坐标。在这个等式中，我们有$f_v \in \mathbb{R}$, $\lambda_i \in \mathbb{R}$ 和 $b_{i,v} \in \mathbb{Q}$。因此，实数 f_v 属于有理线性空间 $\mathrm{Span}_{\mathbb{Q}}(\lambda_1,\ldots,\lambda_\delta)$；并且由于 λ_i 不依赖于 $v \in \mathcal{L}_{m+1}(w)$ 的选取，可得

$$\mathrm{Span}_{\mathbb{Q}}((f_v)_{v\in\mathcal{L}_{m+1}(w)}) \subseteq \mathrm{Span}_{\mathbb{Q}}(\lambda_1,\ldots,\lambda_\delta).$$

现在回到 w中字母 的伪频率 $f_1,\ldots,f_d$。通过与 (3.6) 类似的计数论证，很容易证明

$$f_j = \sum_{v\in\mathcal{L}_{m+1}(w) \text{ s.t. } v \text{ starts with } j} f_v \ ,$$

对每个 $j \in \{1,\ldots,d\}$，由此可推得

$$f_j \in \mathrm{Span}_{\mathbb{Q}}((f_v)_{v\in\mathcal{L}_{m+1}(w)})$$

（仍然对每个 $j \in \{1,\ldots,d\}$），因此

$$\mathrm{Span}_{\mathbb{Q}}(f_1,\ldots,f_d) \subseteq \mathrm{Span}_{\mathbb{Q}}((f_v)_{v\in\mathcal{L}_{m+1}(w)}) \subseteq \mathrm{Span}_{\mathbb{Q}}(\lambda_1,\ldots,\lambda_\delta).$$

上述线性空间的包含关系意味着不等式

$$\dim \mathrm{Span}_{\mathbb{Q}}(f_1,\ldots,f_d) \leq \dim \mathrm{Span}_{\mathbb{Q}}(\lambda_1,\ldots,\lambda_\delta) \leq \delta.$$

由于 $\delta = \dim \ker_{\mathbb{Q}}(M) \leq \Delta_w - 1$ 根据引理 3.10，引理 3.11 得证。定理 3.1 的证明至此完成。

3.3.5 附录：线性代数简要回顾

为方便读者，我们证明如下一般引理，它在引理 3.11 的证明中起了关键作用。

引理 3.13。 (i) 如果向量 $e_1, \dots, e_d \in \mathbb{Q}^n$ 在 $\mathbb{Q}$ 上线性无关，那么它们在 $\mathbb{R}$ 上也线性无关。

(ii) 设 $M \in \boldsymbol{M}_{m,n}(\mathbb{Q})$。如果向量 $e_1, \dots, e_d \in \mathbb{Q}^n$ 构成 核$_{\mathbb{Q}}(M)$ 的一组基，那么它们也构成 核$_{\mathbb{R}}(M)$ 的一组基。

证明。 **(i)** 首先，设 $d = n$。在这种情况下， $\mathbb{Q}$ 线性无关的向量 $e_1, \dots, e_n$ 构成 $\mathbb{Q}^n$ 的一组基，因此通过有理数线性组合生成 $\mathbb{Q}^n$ 的标准基。由于 $\mathbb{Q}^n$ 的标准基与 $\mathbb{R}^n$ 的标准基相同，因此 $\mathbb{R}^n$ 中的每个向量都可以表示为向量 $e_1, \dots, e_n$ 的实系数线性组合。因此，$e_1, \dots, e_n$ 构成 $\mathbb{R}^n$ 的一组基。

现在考虑一般情形 $d \leq n$，我们将这组 $\mathbb{Q}$-线性无关向量 $\{e_1, \dots, e_d\}$ 扩充为 $\mathbb{Q}^n$ 的一组基 $\{e_1, \dots, e_n\}$。根据上述论证，这个扩充后的集合构成 $\mathbb{R}^n$ 的一组基，由此可知，向量 $e_1, \dots, e_d$ 是 $\mathbb{R}$-线性无关的。

(ii) 根据前一点，向量组 $(e_1, \dots, e_d)$ 是 $\mathbb{R}$-线性无关的。要证明它构成 $\ker_{\mathbb{R}}(M)$ 的一组基，只需验证 $\dim \ker_{\mathbb{R}}(M) = d$。应用秩-零化度定理可得

$$\dim \ker_{\mathbb{R}}(M) = n - \mathrm{rk}(M);$$

并且，无论将 M 视为定义在 $\mathbb{Q}$ 上还是 $\mathbb{R}$ 上，其秩都相同。（观察可知，高斯消元不受底域影响。）因此，在 $\mathbb{Q}$ 上应用秩-零化度定理可得

$$\dim \ker_{\mathbb{Q}}(M) = n - \mathrm{rk}(M),$$

因此 $\dim \ker_{\mathbb{R}}(M) = \dim \ker_{\mathbb{Q}}(M) = d$。这证明了 $(e_1, \dots, e_d)$ 构成 $\ker_{\mathbb{R}}(M)$ 的一组基，从而完成了引理 3.13 的证明。 □

3.4 与 Tijdeman 原始证明的比较

Tijdeman 的原始证明和我们的另一种证明都采用反证法，并依赖于对 w 的 Rauzy 图进行深入研究，其中 m 表示复杂度函数的一阶差分小于预期值时对应的最小整数。然而：

- 我们的证明完全是代数性的，且关键在于流矩阵这一概念，以及以下新结果。

定理 3.14 (Andrieu, Cassaigne)。 设 $d \geq 1$，且 w 是一个无限 d 元词。设 $n \in \mathbb{N}$，并用 M 表示 w 中长度为 n 的流矩阵。令 K 为 $\mathbb{Q}$ 或 $\mathbb{R}$。则有

1. $\ker_K(M^t) = \mathrm{Span}((1, \dots, 1)^t)$;
2. $\dim \ker_K(M) = p_w(n+1) - p_w(n) + 1$.

此外，如果 $(f_v)_{v\in\mathcal{L}_{n+1}(w)}$ 表示 w 中长度为 $n+1$ 的因子的频率，或其任意子序列极限

$$\Big(\Big(\frac{|\mathrm{pref}_l(w)|_v}{l}\Big)_{v\in\mathcal{L}_{n+1}(w)}\Big)_l,$$

则 $(f_v)_{v\in\mathcal{L}_{n+1}(w)} \in$ 核$_K(M)$。

相比之下，Tijdeman 最初的证明并未使用线性代数，而是依赖组合论论证，以及有向图中“p-通过数”这一概念。

- Tijdeman 的原始结果比我们的定理 3.1 稍弱：他对无限词 w 的复杂度函数增长所作的估计，依据的是其中字母的最小有理次数 w；而我们的估计依据的是字母的最大有理次数（参见第2章第2.7节的讨论）。不过，我们认为可以重新组织 Tijdeman 的原始证明，以确立我们更强的结果。

3.5 我们证明的一个副产品：具有 Tijdeman 最小复杂度的词是树状的

本节旨在运用我们对 Tijdeman 定理的另一种证明中的要素，证明以下新结果。

定理3.2。 (Andrieu, Cassaigne；在第2章中也编号为定理 2.15)。 设 $d \geq 1$。每个字母频率有理无关且复杂度为$n \mapsto (d-1)n+1$ 的无限 d 元词都是树状的。

这个定理推动了我们对字母频率有理无关且具有最小复杂度的无限词之组合结构的理解；完整讨论和未解问题见第2章第2.6节。

我们先回顾树状词的定义。

如果一个无限词的每个因子的扩展图都是一棵树，那么它就是树状的 。我们回顾一下，无向图在连通的 （按定义：任意两个顶点之间都存在一条路径）且无环 （按定义：不包含任何圈）时称为树 。因子 $u \in\mathcal{L}_n(w)$ 的扩展图 是无向二分图 Ext$(u) = (L \sqcup R, B)$, 定义如下：

- 我们用 L 表示所有满足 $au \in\mathcal{L}_{n+1}(w)$ 的字母 a 的集合。对称地，我们用 R 表示所有满足 $ub \in\mathcal{L}_{n+1}(w)$ 的字母 b 的集合；

  当且仅当 $aub \in\mathcal{L}_{n+2}(w)$ 时，两个顶点 $a \in L$ 和 $b \in R$ 之间有一条边。所有边的集合记为 B。

树状词的概念由 [BDFD$^+$15] 引入，而延拓图的概念可追溯至 [Cas97]。

注 3.15. 扩展图与 Rauzy 图密切相关。更确切地说，要构造常返词 w 的长度为 $n+1$ 的 Rauzy 图 $\mathcal{G}(n+1)$，只需知道它的长度为 n 的 Rauzy 图 $\mathcal{G}(n)$，以及所有长度为 n 、同时为左特殊和右特殊的因子的扩展图。（注意，这些因子在 $\mathcal{G}(n)$ 中由具有多条入边和多条出边的顶点表示。）

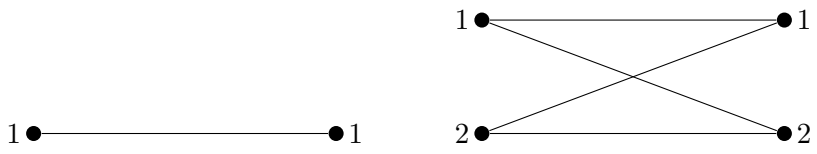


图3.1： $w_1 = 11111...$ 中每个因子的扩展图（左），以及 $w_2 = 112211221122...$ 中空词的扩展图（右）。

例 3.16. 无限词 $w_1 = 11111111...$ 是树状的。事实上，它的各个因子的扩展图都等于图 3.1 左侧的二分图。相比之下，周期无限词 $w_2 = 112211221122...$ 不是树状的。事实上，空词 $\varepsilon \in$ 的扩展图 $\mathcal{L}(w_2)$ 是图 3.1 右侧的非无环图。

练习 3.17 （树状性与回归性）。
1）证明无限词 $w_1 = 21111111...$ 不是树状的。2）证明每个树状无限词都是递归的。3）若一个词的字母由 $\mathbb{Z}$ 而非 $\mathbb{N}$索引，则称该词为双向无限的 。作为对比，证明非递归的双向无限词 $w_2 = ...1112111...$ 是树状的。

练习 3.18. 证明每个Sturmian词都是树状的。提示：利用第1章的命题 1.9。

众所周知，所有 d 元树状词的复杂度为 $n \mapsto (d-1)n+1$[BDFD+15]。有趣的是，只要其扩展图是连通的，树状词甚至可以仅由其复杂度来刻画。

命题 3.19。 无限 d 元词 w 是树状的，当且仅当其复杂度为$n \mapsto (d-1)n+1$，且其所有因子的扩展图都是连通的。

证明。 设 G是一个有 v 个顶点和 e 条边的无向图。我们回顾图论中的两个经典结论：

(1) 若 G 连通，则有 $e \geq v-1$;
(2) G 是非空树，当且仅当 G 连通且 $e = v-1$。

设 w 是一个无限词。记 $\mathrm{Ext}(u) = (L(u) \sqcup R(u\ B(u))$ 为其任意因子的扩展图 u。首先注意到， w 的复杂度函数的离散导数为

$$p(n+1) - p(n) = \sum_{u \in \mathcal{L}_n(w)} (|R(u)| - 1),$$

对所有 $n \in N$。类似地，我们有

$$p(n+2) - p(n+1) = \sum_{v \in \mathcal{L}_{n+1}(w)} (|R(v)| - 1) = \sum_{u \in \mathcal{L}_n(w)} \sum_{a \in L(u)} (|R(au)| - 1).$$

结合这两个恒等式，我们得到复杂度$\Delta(n) := (p(n+2) - p(n+1)) - (p(n+1) - p(n))$ 的二阶导数等于[1]

$$\begin{aligned} \Delta(n) &= \textstyle\sum_{u \in \mathcal{L}_n(w)} \sum_{a \in L(u)} (|R(au)| - 1) - \sum_{u \in \mathcal{L}_n(w)} (|R(u)| - 1) \\ &= \textstyle\sum_{u \in \mathcal{L}_n(w)} (|B(u)| - |L(u)| - |R(u)| + 1) \end{aligned} \tag{3.11}$$

[1]有趣的是（尽管对本讲并无用处），恒等式 (3.11) 对每一个因子语言 $\mathcal{L}$都成立，而无需对其可扩展性作任何假设。

对每个 $n \in \mathbb{N}$ 都成立。

现在，若 w 是一个 d元树状词，则由(2)可知，和式(3.11)中的每一项贡献都等于0，因此复杂度的二阶导数为0。于是复杂度是一个仿射函数。由于 $p(0) = 1$ 且 $p(1) = d$，我们得出对所有$n \in \mathbb{N}$都有 $p(n) = (d-1)n + 1$ ，从而命题 3.19 中的正向蕴含得证。

我们证明反向蕴含。一方面，由于 w 的复杂度为$p(n) = (d-1)n+1$，其二阶导数为0。另一方面，由(1)，由于 w 的所有因子的扩展图都是连通的，和式(3.11)中的每一项贡献 $|B(u)| - |L(u)| - |R(u)| + 1$ 都是非负的。因此，对每个 $u \in \mathcal{L}_n(w)$以及每个 $n \in \mathbb{N}$，我们有

$$|B(u)| - |L(u)| - |R(u)| + 1 = 0.$$

由（2），我们得出 w 是树状的。命题 3.19 的证明至此完成。 □

习题 3.20。 与命题 3.19 类似，证明一个无限 d元词 w 是树状的，当且仅当其复杂度为 $n \mapsto (d-1)n + 1$，且其所有因子的扩展图都是无环的。

有了命题 3.19 的刻画，我们便可以证明定理3.2。

定理3.2的证明。设 $d \geq 1$。设 w 是一个无限 d元词，其字母频率有理无关，且复杂度为 $n \mapsto (d-1)n + 1$。由命题 3.19，为证明 w 是树状的，只需证明其所有因子的扩展图都是连通的。我们采用反证法，并假设存在 w的一个因子 u，其扩展图 $\mathrm{Ext}(u) = (R \sqcup L, B)$ 不是连通的。于是 $\mathrm{Ext}(u)$ 的顶点可以划分为两个非空集合 V_1、 V_2 ，使得 V_1 中的顶点与 V_2中的顶点之间不存在边。用 L_1、 L_2、 R_1、 R_2 分别表示子划分 $L_1 := L \cap V_1$、$L_2 := L \cap V_2$、 $R_1 := R \cap V_1$ 和 $R_2 := R \cap V_2$。由V_1、 V_2的非空性可知，子集 L_1、 L_2、 R_1、 R_2 中至少有三个是非空的。（尽管这对证明并不重要，但请注意， w的回归性——已在第2章命题 2.13 中证明——蕴含四个子集 L_1、 L_2、 R_1、 R_2 都是非空的。）

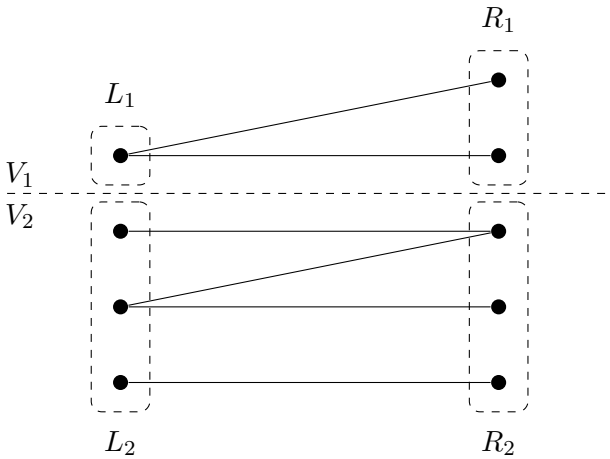


图 3.2：扩展图划分的示例。

记 $\mathcal{G}$ 为 w 关于长度 $n = |u|$ 的 Rauzy 图。我们将把 $\mathcal{G}$ 修改为一个新图 $\mathcal{G}'$ ，方法是将顶点 u 分裂为两个副本 u_1 和 u_2，并分配其

入边和出边于 u_1 与 u_2 之间，依据其扩展图 Ext(u) 的划分。更精确地说，我们如下定义图 $\mathcal{G}' = (V', E')$。

- 其顶点为 $V' = \mathcal{L}_n(w) \backslash \{u\} \cup \{u_1, u_2\}$。

- $\mathcal{G}$ 中所有与顶点 u 不相邻的边都保留在$\mathcal{G}'$中。在$\mathcal{G}'$ 中，进入 u_1 （分别是 u_2）的入边，是$\mathcal{G}$ 中进入 u且同时属于 $L_1 \cdot u = \{au, a \in L_1\}$（分别是 $L_2 \cdot u$）的入边。类似地，在$\mathcal{G}'$ 中，从u_1 （分别是 u_2）出发的出边，是$\mathcal{G}$ 中从 u出发且同时属于$u \cdot R_1$ （分别是 $u \cdot R_2$）的出边。

如此定义的图$\mathcal{G}'$ 有 $|V'| = p_w(n)$ 个顶点和 $|E'| = |E| = p_w(n+1)$ 条边。尽管$\mathcal{G}'$ 不再是 Rauzy 图，但它继承了其中许多性质。值得注意的是，$\mathcal{G}'$ 仍然是半连通的。

引理 3.21。 图 G' 是半连通的。

证明。 证明只需对引理 3.10 的证明稍作修改。由于$\mathcal{G}'$的边恰好是无限词 w中长度为 $n+1$ 的因子，因此$\mathcal{G}'$中总存在一条路径，将任意两条边 v_1 和 $v_2 \in \mathcal{L}_{n+1}(w)$连接起来。（例如，可以通过在 w中，从 v_1 的一次出现滑动窗口至 v_2 的一次出现来得到这样一条路径。）现在只需注意到$\mathcal{G}'$ 中的每个顶点都至少与一条边相连。对于 $V' \backslash \{u_1, u_2\}$ 中的所有顶点，这显然成立，因为它们都有出边。对于 u_1 和 u_2，这也成立，因为四个子集 L_1, L_2, R_1, R_2 中至少有三个非空。引理 3.21 得证。

□

此外，$\mathcal{G}'$ 存在一个大小为 $|V'| \times |E'|$ 的“流矩阵” M'，定义如下：位置 $(u, v) \in V' \times E'$ 处的元素，在有向边 v 从 u 出发且不抵达 u时为 1；在 v 抵达 u 但不从 u出发时为 -1；其他情况下为 0。我们声称，证明 Tijdeman 定理的核心引理（即引理 3.9 和 3.10）对 M'仍然成立，且其证明只需稍作调整。

引理 3.22。 采用上述记号，有

1. 维数 $\ker_{\mathbb{Q}}(M') = |E'| - |V'| + 1 = d - 1$，
2. w 的长度为-$n+1$ 因子的伪频率向量属于 M'的核，即，

$$M' \cdot (f_v)_{v \in \mathcal{L}_{n+1}(w)} = 0.$$

引理 *3.22* 的证明。 第一条断言的证明是对引理 3.10 证明的直接改编。我们回顾一下，该证明依赖于$\mathcal{G}'$ 的连通性（已在引理 3.21 中证明），由此我们可以证明 $\ker_{\mathbb{Q}}(M'^t) = \mathrm{Span}((1, \dots, 1)^t)$。随后，两次应用秩-零化度定理即可得到所需的等式。下面证明第二条断言，它仍可理解为$\mathcal{G}'$上的基尔霍夫节点定律。记 $r \in V'$所索引的 M' 的行为 $M'[r]$。我们的目标是证明，对于每个 $r \in V'$，

$$M'[r] \cdot (f_v)_{v \in \mathcal{L}_{n+1}(w)} = 0. \tag{3.12}$$

对于每个 $r \neq u_1, u_2$，这个等式显然成立，因为 $M'[r]$ 也是原始 Rauzy 图 $\mathcal{G}$ 的流矩阵（长度为 n）中的一行。因此，只需再次应用引理 3.9。对于 $r = u_1$ 和 $r = u_2$，等式 (3.12) 可由以下等式得出：对于每个 $a \in \{1, \dots, d\}$，我们有

$$f_{au} = \sum_{b \in \{1, \dots, d\}} f_{aub};$$

类似地，对于每个 $b \in \{1,\ldots,d\}$，我们有

$$f_{ub} = \sum_{a\in\{1,\ldots,d\}} f_{aub}.$$

由于不存在连接 V_1 与 V_2的边，我们有 $f_{aub} = 0$ 若（a, b） $\in L_1 \times R_2 \cup L_2 \times R_1$。因此

$$\sum_{a\in L_1} f_{au} = \sum_{a\in L_1}\sum_{b\in R_1} f_{aub} = \sum_{b\in R_1} f_{ub}, \tag{3.13}$$

and

$$\sum_{a\in L_2} f_{au} = \sum_{a\in L_2}\sum_{b\in R_2} f_{aub} = \sum_{b\in R_2} f_{ub}. \tag{3.14}$$

方程(3.13)与(3.14)分别等价于 $M'[u_1]\cdot(f_v)_{v\in\mathcal{L}_{n+1}(w)} = 0$与 $M'[u_2]\cdot(f_v)_{v\in\mathcal{L}_{n+1}(w)} = 0$。引理 3.22 得证。 □

由引理 3.22，采用与引理 3.11 完全相同的计算，我们证明 w 的字母频率张成一个维数至多为 $d-1$的有理线性空间。我们得到了所需的矛盾。定理3.2的证明完成。

□

参考文献

[Ale95] Pascal Alessandri. 复杂度为 n+2 的词的分类与表示. 技术报告，艾克斯-马赛第二大学，1995年.[AMST94] Pierre Arnoux、Christian Mauduit、Iekata Shiokawa 和 Jun-Ichi Tamura. 由立方体台球定义的序列的复杂度 . 法国数学学会公报，122(1):1–12，1994年.[And21a] Mélodie Andrieu. 多维连分数算法符号动力系统中的特殊轨道. 博士论文，法国艾克斯-马赛大学，2021年3月. 可见于 https://www-lmpa.univ-littoral.fr/~mandrieu/research/andrieu_manuscript.pdf.[And21b] Mélodie Andrieu. 在平面所有方向上均无界的 Rauzy 分形. 数学报告，359(4):399–407，2021年.[AR91] Pierre Arnoux 和 Gérard Rauzy. 复杂度为 2n+1 的序列的几何表示. 法国数学学会公报，119:199–215，1991年.[AV24] Mélodie Andrieu 和 Léo Vivion. 超立方台球词中的不平衡性（II）. 载于第19届蒙斯理论计算机科学日，法国尼斯，2024年.[Bar95] Yuliv Baryshnikov. 矩形台球轨迹的复杂度. 数学物理通讯，174(1):43–56，1995年.[BB13] Nicolas Bédaride 和 Jean-François Bertazzon. 环面上的平移所关联复杂度函数的下界. 数学月刊，171:291–304，2013 年 .[BD14] Valérie Berthé 和 Vincent Delecroix. 超越替代动力系统：S-adic 展开. 载于《Lecture notes Kokyuroku Bessatu》，第81–123页，2014年.[BDFD+15] Valerie Berthé、Clelia De Felice、Francesco Dolce、Julien Leroy、Dominique Perrin、Christophe Reutenauer 和 Rindone Giuseppina. 无环、连通和树集. 数学月刊，176(4):521–550，2015年.[Béd09] Nicolas Bédaride. 超立方体台球的方向复杂度. 离散数学，309(8):2053–2066，2009年.[Bor50] Émile Borel. 关于 $\sqrt{2}$ 的小数数字及若干链式概率问题. 巴黎科学院报告，230:591–593，1950年.

–

[Bos84] Michael Boshernitzan. 具有线性分块增长的极小符号流的唯一遍历性. 《数学分析杂志》，44(1):77–96，1984年.[Cas97] Julien Cassaigne. 复杂度与特殊因子. 《比利时数学学会公报——Simon Stevin》，4(1):67–88，1997年.[Cas98] Julien Cassaigne. 具有分组因子的序列. 载于 DLT'97，《语言理论发展 III》，塞萨洛尼基，第211–222页，1998年.[CH73a] Ethan M. Coven 和 Gustav A. Hedlund. 具有最小分块增长的序列 I. 数学系统理论，7:138–153，1973年.[CH73b] Ethan M. Coven 和 Gustav A. Hedlund. 具有最小分块增长的序列 II. 数学系统理论，8:376–382，1973年.[CLL17] Julien Cassaigne、Sébastien Labbé 和 Julien Leroy. 一类复杂度为 $2n+1$ 的序列. 载于《WORDS 2017 论文集》，第144–156页. Springer，2017年.[CLL22] Julien Cassaigne、Sébastien Labbé 和 Julien Leroy. 几乎处处平衡且复杂度为 $2n+1$ 的序列. 《莫斯科组合学与数论杂志》，11(4):287–333，2022年.[Del16] Vincent Delecroix. 区间交换变换. 讲义，2016年. 可见于 `https://www.labri.fr/perso/vdelecro/Salta2015-notes.pdf`.[DHS22] Ivan Dynnikov、Pascal Hubert 和 Alexandra Skripchenko. Rauzy 胎面周围的动力系统及其遍历性质. 《国际数学研究通报》，2023(8):6461–6503，2022年.[DJP01] Xavier Droubay、Jacques Justin 和 Giuseppe Pirillo. Episturmian 词以及 de Luca 和 Rauzy 的一些构造. 《理论计算机科学》，255(1):539–553，2001年.[FM97] Sébastien Ferenczi 和 Christian Mauduit. 具有低复杂度展开的数的超越性. 《数论杂志》，67(2):146–161，1997年.[FM10] Sébastien Ferenczi 和 Thierry Monteil. 具有一致频率的无限词与不变测度. 载于 V. Berthé 和 M. Rigo 编，《组合学、自动机与数论》，《数学及其应用百科全书》，第373–409页. 剑桥大学出版社，2010年.[FZ08] Sébastien Ferenczi 和 Luca Q. Zamboni. k-区间交换变换的语言. 《伦敦数学学会公报》，40(4):705–714，2008年.[GJ09] Amy Glen 和 Jacques Justin. Episturmian 词：综述. RAIRO - 理论信息学与应用，43(3):403–442，2009年.[JP02] Jacques Justin 和 Giuseppe Pirillo. Episturmian 词与 Episturmian 态射. 《理论计算机科学》，276(1):281–313，2002年.[Kea75] Michael Keane. 区间交换变换. 《数学杂志》，141:25–32，1975年.

[KT07] Idrissa Kaboré 与 Théodore Tapsoba. 复杂度为 n+2 的重复词组合学. *RAIRO* - 理论信息学与应用, 41(4):425–446, 2007年.[Lot02] M. Lothaire. 词的代数组合学. 数学及其应用百科全书. 剑桥大学出版社, 2002年.[MH38] Marston Morse 与 Gustav A. Hedlund. 符号动力系统. 《美国数学杂志》, 60(4):815–866, 1938年.[MH40] Marston Morse 与 Gustav A. Hedlund. 符号动力系统 II. Sturmian 轨道. 《美国数学杂志》, 62:1–42, 1940年.[PF02] N. Pytheas Fogg. 动力系统、算术与组合学中的替换. 数学讲义. 施普林格·柏林·海德堡, 2002年.[Rau83] Gérard Rauzy. 有限字母表上的序列. 波尔多数论研讨班, 12:1–16, 1982-1983.[Tij99] Robert Tijdeman. 关于无限词的最小复杂度. 《数学探究》, 10(1):123–129, 1999.[Via06] Marcelo Viana. 区间交换映射的遍历理论. 《康普顿斯数学杂志》, 19(1):7–100, 2006年.[Yoc05] Jean-Christophe Yoccoz. 区间交换. 法兰西公学院课程, 2005年.