arXiv:2602.03746v1 [math.CO] 3 Feb 2026

因子平衡性、线性递归与因子复杂度

´BASTIAN ESPINOZA 和 MANON STIPULANTI

数学系，滑铁卢大学，*Li`ège，All´ée de la D´écouverte 12，4000 Li`ège*，比利时

PIERRE POPOLI

*Sc*滑铁卢大学计算机科学学院，加拿大安大略省滑铁卢市，邮编 *N2L 3G1*

Abstract. 在无限词研究中，各种平衡性概念为字母或因子出现的规律性提供了定量度量，并应用于数学和理论计算机科学的多个领域。本文研究因子平衡性和一致因子平衡性，并作出两项主要贡献。首先，我们为无限词成为（具有一致）因子平衡词建立了一般性的充分条件，尤其适用于任意给定的线性递归词。这些条件以 $\mathcal{S}$-进表示的形式表述，并推广了 Adamczewski 关于本原替代词的结果；该结果表明，长度为 2 的因子的平衡性已经足以推出一致因子平衡性。作为这些判据的应用，我们刻画了具有一致因子平衡性的 Sturmian 词和三元 Arnoux–Rauzy 词：它们恰好是弱部分商有界的那些词。我们的第二项主要贡献是研究因子平衡性与因子复杂度之间的关系。特别是，我们分析了非本原替代情形，并构造了一个具有指数因子复杂度的因子平衡词，从而推动了对 Arnoux、Berth´é、Minervino、Steiner 和 Thuswaldner 于 2025 年提出的关于平衡性与离散谱之间关系的问题的研究。

关键词：词组合学、平衡性、因子平衡性、均匀因子平衡性、偏差、Sturmian 词、Arnoux–Rauzy词、线性递归词、替换、决定性、因子复杂度、Toeplitz 词。

2020年数学主题分类：68R15。

致谢。我们感谢 Boris Adamczewski、Shigeki Akiyama、Pierre Arnoux、Val´erie Berth´e 和 Jeffrey Shallit 在整个项目过程中给予的真知灼见。

E-mail addresses：baespinoza@uliege.be，m.stipulanti@uliege.be，pierre.popoli@uwaterloo.ca。*Date*：2026年2月4日。

Basti´an Espinoza 获得列日大学特别研究基金 IPD-STEMA 项目的资助。Pierre Popoli 获得加拿大自然科学与工程研究委员会 2024-03725 号基金的资助。Manon Stipulanti 是比利时国家科学研究基金研究副研究员，获得研究基金 1.C.104.24F 的资助。

Contents

1. 引言 2 2. 开放问题 6 3. 预备知识 6 4. 因子平衡词概述 12 5. 线性递归词中的因子平衡性 16 6. Arnoux–Rauzy词中的一致因子平衡性 23 7. 因子平衡性与因子复杂度 27 参考文献 36 附录 39

1. Introduction

对平衡无限词的研究——也称为序列、字符串或数据流——始于 Morse 和 Hedlund 在 20 世纪 30 年代末的开创性工作 [MH38, MH40] 。他们的研究引入了平衡性这一基本性质，将所谓$Sturmian$ 词的组合学与看似相距甚远的领域联系起来，例如正方形中的台球动力学和连分数算术。这一视角为后来成为词组合学反复出现的主题奠定了基础（例如，参见 [DP22, Fog02, Lot02, Thu20]），并且也影响了其他多个领域。在非周期序理论中，平衡性是 Delone 集、铺砌和准晶体的一项重要规则性性质，可定量衡量点或铺砌单元分布得有多均匀；例如，参见 [BG13]。它也是一致分布理论中反复出现的概念 [KN74]。在康托尔集上的拓扑动力系统和遍历理论中，Gottschalk–Hedlund 定理将既开又闭集的平衡性与上同调性质及拓扑特征值的存在联系起来 [DP22, GH55]。从计算角度看，平衡性与某些优化和运筹学问题相关；例如，参见 [AGH00, Tij80]。在算术和数论中，它以有界余差集之名出现，并与连分数密切相关（例如，参见 [Fog02]），也与 Fraenkel 猜想等未解问题相关 [Fra73]。这一猜想最初提出于数论领域 [Fra73],，可以改写为如下组合学命题：对于每个 $k \geq 3$，在一个含有 k 个字母的字母表上，除字母置换外，至多只有一个各字母频率不同的 1-平衡无限词。该猜想已在较小的 k 取值时得到验证，参见 [AGH00, Tij00a, Tij00b] ；对于 $k \in \{3, 4, 5, 6\}$，以及某些特定词族（例如 episturmian 词），也有相关结果，参见 [PV07]。最后，在台球动力学的背景下，完全不平衡性与拓扑弱混合的概念相关。

在词组合学和符号动力系统中，平衡性是衡量因子在给定无限词中出现得有多均匀的一个定量指标。Morse 和 Hedlund [MH38, MH40] 最初将无限词$\mathbf{x}$ 定义为如今通常所称的[1] 1-平衡，即对于任意字母 a 以及, v 中任意两个长度相同、属于$\mathbf{x}$ 的因子，我们都有 $\big||u|_a - |v|_a\big| \leq 1$（其中，记号 $|w|_a$ 表示字母 a 在词 w中的出现次数）。随后，将常数 1 替换为任意整数 $C \geq 1$，便将 1-平衡性推广为 C-字母平衡性。我们注意到，这等价于有界偏差条件（对应的常数可能不同），参见 [Ada03]或下文的引理 3.4；关于偏差理论的更一般参考资料，读者可参见 [KN74]。正如 [RSZ11] 所指出的，它也等价于有界阿贝尔因子复杂度。Rauzy 的开创性论文 [Rau82]将字母平衡性作为关键性质，用以构造与*Tribonacci* 替换相关联的分形集。这项工作将算术几何与词组合学之间的联系拓展到了二维情形；这种联系在 Sturmian 情形中早已出现。Rauzy 的论文开启了大量后续研究。尤其是，一系列研究致力于将这种相互作用推广到更广泛的无限词族。这条研究脉络包括引入 Arnoux–Rauzy 词 [AR91], 、推翻了关于它们在所有情形下均为字母平衡的猜想 [CFZ00], 、研究与这些系统相关的李雅普诺夫指数 [AD19]，以及证明它们几乎必然平衡并具有几何表示的结果 [BST22]。此后，研究还拓展到其他词族，例如 Brun 字 [BST22]。最近，Berth´e、Cecchi–Bernales 和 Espinoza 根据关联矩阵和所谓的扩展图，给出了本原代换生成词具有字母平衡性的充要条件，参见 [BCBE, 命题 6.1]。

除了考察字母，我们还可以转而考察因子。给定一个整数 $C \geq 1$ 和一个无限词 $\mathbf{x}$，若有限词 w 满足：对于 $\mathbf{x}$ 中任意两个长度相同的因子 u, v，它们各自包含 w 和 u, v 的出现次数之差不超过 C，则称 w 是$\mathbf{x}$ 中的-平衡词。若$\mathbf{x}$ 的每个因子都在$\mathbf{x}$ 中是 C-平衡的，其中 C取某个值，则称$\mathbf{x}$是因子平衡的。当界 C 是一致的，即与$\mathbf{x}$中所选的因子无关时，我们称之为一致因子平衡性。本文主要研究因子平衡性及其一致版本。

与字母平衡性不同，字母平衡性只取决于词的阿贝尔化，而因子平衡性取决于更精细的组合结构，因此通常不能归结为对字母的研究。然而，Adamcwzeski [Ada03]证明，本原代换词的因子平衡性等价于长度为 2 的因子的平衡性。关于著名的词族，Fagnot 和 Vuillon [FV02] 表明，Sturmian 词是因子平衡的，并给出了它们成为一致因子平衡的条件。此外，许多研究致力于考察字母平衡性与不同长度下因子平衡性之间的联系。例如，Berth´e 和 Cecchi–Bernales [BCB19] 证明，在树状词中，字母平衡性蕴含因子平衡性。近年来，这类词日益受到关注

[1]提醒读者，这种性质有时也直接称为 *balancedness*。本文中，我们使用术语 *letter-balanced* 和 *factor-balanced*，分别表示对某个常数 C而言的 C-字母平衡和 C-因子平衡；这些概念将在后文介绍。

，并将 Arnoux–Rauzy 词和区间交换变换的编码推广开来；例如参见 [BDFD+15a、BDFD+15b、BDFD+15c、GGOLS25、GL25]。此后，Berth´e 和 Cecchi–Bernales 的结果由 Berth´e 等人 [BCBD+21推广至本原幺模适当的 $\mathcal{S}$-adic 子移位。最近，Arnoux 等人 [ABM+25] 将因子平衡性与 Pisot猜想联系起来；该猜想断言，任何不可约Pisot代换都会生成具有纯离散谱的符号动力系统；关于这一猜想的综述，参见 [ABB+15] 。在此背景下，Arnoux 等人 [ABM+25在问题 18.2] 中提出：哪些因子平衡无限词会生成具有离散谱的符号系统? 关于字母平衡性这一原始概念与其向因子平衡性的扩展之间的关系，我们还提及 Poirier 和 Steiner [PS24]以及 Qu´effelec [Que10]的研究。

虽然一致因子平衡性强于因子平衡性，但 Adamczewski 的结果 [Ada03, Ada 04] 表明，若某替换的不动点所对应的长度2滑动块码是本原 $Pisot$ 替换（即其模最大的次大特征值严格小于 1），则该不动点满足这一性质。因此，许多经典无限词都属于这一类，例如特里波纳契词；更一般地，任意代换生成的 Arnoux-Rauzy 词也属于这一类 [AI01]。据我们所知，这一性质直到二十年后才作为一项计算假设明确出现于 [CDO+25]；该假设的提出受到定理证明器 Walnut [Sha23], 的启发，而 Walnut 建立在一阶逻辑与自动机理论之间的联系之上。基于这一假设，Couvreur 等人 [CDO+25] 证明，自动一致因子平衡词的二维广义阿贝尔复杂度是正则的，这里的“正则”采用 Allouche 和 Shallit 的定义 [AS92]。此外， [CDO+25]的作者还证明了以下关于特里波纳契词的显式结果，将 [RSZ10] 关于字母平衡性的结果加以推广。

定理 1.1 ([CDO+25])。 *Tribonacci* 词具有一致因子平衡性，常数为2。

然而，在相关文献中，Adamczewski 的一般定理似乎在因子平衡性方面未得到充分利用。当 [CDO+25] 的作者向词组合学、动力系统和数制系统领域的听众介绍其工作时，这一点变得尤为明显，因为一致因子平衡性（尤其是定理 1.1）常常令他们感到惊讶。这一观察构成了本文工作的主要动机。我们的目标是通过重新审视已知结果、阐明其适用范围和重要性，并拓展当前对这一性质的理解，同时给出一些关于因子平衡性的新的一般性结果，使一致因子平衡性重新成为关注焦点。

1.1.**我们的贡献**。 本文以两种不同的方式，为因子平衡性和一致因子平衡性的理论作出贡献。

首先，在第5节中，我们给出线性递归词成为（统一）因子平衡词的条件。我们在定理 1.2 中陈述这一结果，但必要的定义和符号将在第5节中介绍，读者可参阅该节。需要注意的是，对于给定的线性递归词，下文所述的相容序列总是存在。

定理 1.2. 设$\mathbf{x}$ 是一个线性递归词，且$(\tau_n, a_n)_{n\geq 0}$ 是生成$\mathbf{x}$ 并满足条件$(\mathbf{P})$、$(\mathbf{F})$和$(\mathbf{D})$的相容序列。令$\mathbf{x}^{(\mathbf{n})}$ 表示以下对象的极限：

$(\tau_{n,\ m}(a_m))_{m>n}$。假设存在一个序列$(B_n)_{n\geq 0}$ ，使得词$\mathbf{x}^{(\mathbf{n})}$是 B_n-字母平衡的。那么词 $\mathbf{x}$ 是因子平衡的。此外，若序列$(B_n)_{n\geq 0}$ 有界，则词$\mathbf{x}^{(\mathbf{n})}$ 是一致因子平衡的。

我们证明定理 1.2 的方法基于线性递归词的丰富的 $\mathcal{S}$-进表示理论，并在第 5.1 节中加以说明。我们方法的另一个关键要素是第 5.2 节中的决定性概念。这使我们能够在第 5.3 节中将研究因子平衡性的问题，化归为研究原词的适当逆替换所得词中字母平衡性的问题。

定理 1.2 可应用于任意本原代换生成的词，因为这类词都是线性递归的 [DHS99]。在这种情况下，我们在第 5.3 节中得到一致因子平衡性与长度为 2 的因子的平衡性之间的等价关系。

命题 1.3。 设 $\mathbf{x}$ 是一个本原替换的不动点。当且仅当词$\mathbf{x}$ 一致因子平衡时，$\mathbf{x}$的每个长度为 *2* 的因子在$\mathbf{x}$ 中都是平衡的。

Adamczewski 的研究中已出现这一现象，这可直接由 [Ada03, 定理 22] 得出。

更重要的是，定理 1.2 的另一项应用完整刻画了具有一致因子平衡性的二元和三元 Arnoux–Rauzy 词，我们将在第 6 节讨论这一结果。所有必要的定义都在第 3.6 节中给出。另请注意，二元情形，即关于 Sturmian 词的情形，已在 [FV02] 中讨论过。

定理 1.4. 字母表至多含3个字母的 *Arnoux–Rauzy* 词是一致因子平衡的，当且仅当它具有有界的弱部分系数。

其次，在第 7 节中，我们研究因子平衡性与因子复杂度之间的联系。我们首先在命题 7.3 中证明，替代因子平衡词具有线性因子复杂度；随后在更专门的命题 7.3 中，我们描述了这些词的一种特殊代换结构。此外，我们通过定理 1.5 给出一个因子平衡词的例子，其因子复杂度为指数级。这一论证基于 Toeplitz 词，其构造将在第 7.2 节中回顾。

定理 1.5. 存在一个因子平衡且具有指数增长因子复杂度的无限词。特别地，由$\mathbf{x}$生成的符号动力系统具有正的拓扑熵，因而不能具有纯离散谱。

这推进了 Arnoux 等人在 [ABM^{+}25中提出的问题 18。2] 通过表明并非所有因子平衡的符号系统都具有离散谱。

最后，读者将在第3节找到本文所需的背景知识；特别是，其中若干小节专门给出关于词、替换、偏差与平衡性之间的密切联系、 $\mathcal{S}$-进表示、Sturmian 词以及 Arnoux–Rauzy词的定义。第4节包含关于（一致）因子平衡性的一般性内容，尤其是它在可逆高阶块码下的稳定性，以及它与无限词中一致回归的关系。我们没有把一些未决问题留到论文结尾，而是直接在下文第2节中提出。

2. Open problems

在本节中，我们汇集了撰写本文过程中提出的一些未决问题，这些问题促使人们进一步研究（一致）因子平衡性这一主题。

在第5节中，定理 1.2 给出了使线性递归词成为（均匀）因子平衡词的充分条件。由此自然引出一个问题：是否存在使这一性质成立的必要条件?

问题 2.1. 使线性递归词成为（均匀）因子平衡词的必要条件是什么?

要回答这个问题，似乎需要某种可识别性假设。在第6节中，我们完整刻画了均匀因子平衡的 Sturmian 词和三元 Arnoux–Rauzy 词。自然会想到，这一性质是否能推广到任意大小的字母表。

问题 2.2. 是否可能刻画在任意 d 元字母表上哪些 Arnoux–Rauzy 词是均匀因子平衡的，其中 $d \geq 2$?

请注意，三个字母已经是 [BCS13], 中的一个限制，因此一般情形可能更加复杂，解决这个问题可能需要新的方法。

在第 7.3 节中，我们明确给出了一个因子平衡词的例子，其因子复杂度呈指数增长。人们可能会进一步提出以下问题。

问题 2.3. 每一种指数增长的因子复杂度，都能由因子平衡无限词实现吗?

若不作重大修改，我们的 Toeplitz构造无法得到因子复杂度呈指数增长的一致因子平衡例子。此外，目前所有已知的一致因子平衡词的因子复杂度都是线性的。因此，自然会提出以下问题。

问题 2.4. 能否构造一个因子复杂度超线性增长的一致因子平衡无限词?

根据命题 7.3，因子复杂度超线性的（一致）因子平衡词不可能是代换生成的。因此，必须采用不同的词，正如我们在利用 Toeplitz 词证明定理 1.5 时所做的那样。

3. Preliminaries

3.1. 词的一般性质。 字母表是由有限个元素组成的集合，这些元素称为字母。字母表上的长度为 $n \in \mathbb{N}$ 的有限词 w 是一个元素 $w = w_1 w_2 \cdots w_n$，其中每个 w_i 都属于 $\mathcal{A}$。我们用 $\mathcal{A}^n$ 表示字母表 $\mathcal{A}$ 上长度为 n 的词的集合。我们用 $\mathcal{A}^*$ 表示 $\mathcal{A}$ 上的有限词集合，即 $\mathcal{A}^* = \cup_{n \geq 0} \mathcal{A}^n$；用 ε 表示空词。无限词 $\mathbf{x}$ 是 $\mathcal{A}^{\mathbb{N}}$ 的一个元素（除非另有说明，我们从 0 开始为它们编索引）。为区分有限词和无限词，我们将后者用粗体表示。

有限词 $v = v_1 \cdots v_k$ 是另一个有限词 $u = u_1 \cdots u_\ell$ 的因子，如果存在 $i \geq 1$ 使得 $v_1 \cdots v_k = u_i \cdots x_{i+k-1}$。若 $i = 1$，我们称 v 是 u 的前缀，

若 $i = \ell - k + 1$，我们称 v 是 u的后缀。若 v 是 u的因子，那么 v 在 u 中的一次出现是指满足 $v_1 \cdots v_k = u_i \cdots x_{i+k-1}$的任意整数 i。我们用 $|u|_v$ 表示 v 在 u 中的出现次数（允许重叠）。当 u 是无限词时，我们使用相同的术语。给定一个无限词 $\mathbf{x}$，我们还用 $\mathcal{L}_n(\mathbf{x})$表示其长度为 n 的因子集合，并用 $\mathcal{L}(\mathbf{x}) = \bigcup_{n\geq 0} \mathcal{L}_n(\mathbf{x})$表示其因子集合。我们用 $p_{\mathbf{x}}$ 表示 $\mathbf{x}$ 的因子复杂度，即映射 $p_{\mathbf{x}}\colon \mathbb{N} \to \mathbb{N}, n \mapsto \#\mathcal{L}_n(\mathbf{x})$，它统计 $\mathbf{x}$ 中每种长度的因子数量。若 $\mathbf{x}$ 可写为 $uvvv\cdots$，其中 u, v 是有限词，则称 $\mathbf{x}$ 是周期最终或最终周期的，周期为 $|v|$；若 $u = \varepsilon$，则称 $\mathbf{x}$ 是（完全）周期的。若 $\mathbf{x}$ 不是周期的，则称其为非周期词。若存在 $v \in \mathcal{A}^*$ 和 $k \geq 2$ 满足 $u = v^k$，则词 $u \in \mathcal{A}^*$ 是（另一个词的）幂；否则称 u 是本原的。对于任意整数 P，若一个词的任何因子都不具有 v^P 的形式，其中 v是非空词，则称该词是 P-幂自由的。

考虑一个无限词 $\mathbf{x} = (x_n)_n \in \mathcal{A}^{\mathbb{N}}$ 以及一个有限词 $u \in \mathcal{A}^*$。若存在，频率$\mu_{\mathbf{x}}(u)$ *of* u *in* $\mathbf{x}$定义为以下极限

$$\lim_{n\to+\infty} \frac{|x_0 \cdots x_{n-1}|_u}{n}.$$

当该极限存在时，我们称 $\mathbf{x}$ 具有词 u 的频率。如果 u 在 $x_k \cdots x_{k+n-1}$ 中的出现次数除以 n 在 k 趋于无穷时存在一致极限，即存在一个实数 μ_u 使得，我们称 $\mathbf{x}$具有词 u 的一致频率。

$$\forall \varepsilon > 0, \exists n_0 \geq 0, \forall n \geq n_0, \forall k, \left| \frac{1}{n} |x_k \cdots x_{n+k-1}|_u - \mu_u \right| \leq \varepsilon.$$

最后，如果 $\mathbf{x}$的所有因子都具有一致频率，我们称 $\mathbf{x}$具有一致频率。若 $\mathbf{x} \in \mathcal{A}^{\mathbb{N}}$ 中的每个因子都在$\mathbf{x}$中出现无穷多次，则称无限词 $\mathbf{x}$是复现的。若存在一个整数 k，使得每个因子 $w \in \mathcal{L}(\mathbf{x})$ 都出现在$\mathbf{x}$的每个长度为k 的因子中，则称词$\mathbf{x}$ 是一致复现的，或称极小的。对于任意无限词 $\mathbf{x}$，都存在一个一致复现词 $\mathbf{y}$，使得 $\mathcal{L}(\mathbf{y}) \subseteq \mathcal{L}(\mathbf{x})$（该结果源于一个众所周知的事实：每个拓扑动力系统都包含一个极小子系统；例如参见 [Wal82, 定理 5.2]）。

设 $\mathbf{x} \in \mathcal{A}^{\mathbb{N}}$ 为一个无限词，且 $u \in \mathcal{L}(\mathbf{x}) \setminus \{\varepsilon\}$。$\mathbf{x}$ 中到 u 的回返词是一个词 w，使得 $wu \in \mathcal{L}(\mathbf{x})$，且 u 在 wu 中恰好出现两次（一次作为前缀，一次作为后缀）。我们用 $\mathcal{R}_{\mathbf{x}}(u)$ 表示 $\mathbf{x}$ 中到 u 的回返词集合。于是，若对每个 $u \in \mathcal{L}(\mathbf{x}) \setminus \{\varepsilon\}$，集合 $\mathcal{R}_{\mathbf{x}}(u)$ 非空且有限，则 $\mathbf{x}$ 是一致复现的。此外，若对所有 $u \in \mathcal{L}(\mathbf{x}) \setminus \{\varepsilon\}$ 以及所有 $w \in \mathcal{R}_{\mathbf{x}}(u)$，我们有 $|w| \leq L|u|$，则称 $\mathbf{x}$ 是线性复现的（常数为 $L \geq 0$）。特别地，若 $\mathbf{x}$ 的每个长度为 $(L+1)n$ 的因子都包含 $\mathcal{L}_n(\mathbf{x})$ 的每个元素，则 $\mathbf{x}$ 以常数L 线性复现。我们回顾如下结果。

引理3.1 ([DHS99，定理 24])。设 $\mathbf{x} \in \mathcal{A}^{\mathbb{N}}$ 为一个以常数 L线性复现的无限非周期字。

(1) 对任意非空词 $u \in \mathcal{A}^*$，u^{L+1} 都不是$\mathbf{x}$的因子。*(2)* 对任意非空 $u \in \mathcal{L}(\mathbf{x})$ 以及任意 $w \in \mathcal{R}_{\mathbf{x}}(u)$，都有 $\frac{|w|}{L} \leq |u| \leq L|x|$。*(3)* 对任意非空 $u \in \mathcal{L}(\mathbf{x})$，都有 $\mathrm{Card}\big(\mathcal{R}_{\mathbf{x}}(u)\big) \leq L(L+1)^2$。

考虑一个无限词$\mathbf{x} \in \mathcal{A}^{\mathbb{N}}$ 和一个有限词 $w \in \mathcal{A}^*$。固定一个常数 $C > 0$。如果对于 $\mathbf{x}$ 的所有等长因子 u, v ，都有$||u|_w - |v|_w| \leq C$，我们就称 w 在$\mathbf{x}$ 中是-平衡的。如果任意字母 a在$\mathbf{x}$中都是 C-平衡的，我们就称$\mathbf{x}$ 是 C-字母平衡的，或称其为字母平衡的。如果任意词 w 在$\mathbf{x}$中对于某个常数 $C_w > 0$都是 C_w-平衡的，我们就称$\mathbf{x}$ 是因子平衡的。请注意，我们特别指出，平衡常数可能取决于所考虑的词 w。最后，如果存在一个常数 $C > 0$ ，使得每个有限词 w 在$\mathbf{x}$中都是 C-平衡的，我们就称$\mathbf{x}$ 是一致因子平衡的。显然，一致因子平衡性蕴含因子平衡性，而因子平衡性又蕴含字母平衡性。关于字母平衡性与一致字母频率之间的联系，有如下经典结果。

命题 3.2 ([Ada03，命题 7])。 无限词 $\mathbf{x} \in \mathcal{A}^{\mathbb{N}}$ 是字母平衡的，当且仅当它具有一致字母频率，并且存在一个常数 B ，使得对于 $\mathbf{x}$ 的任意因子 u ，以及任意字母 $a \in \mathcal{A}$，都有 $||u|_a - \mu_a|u|| \leq B$ ，其中 μ_a 是 a 的频率。

通过类似的证明，可以得到以下关于一致因子频率和（均匀）因子平衡性的结果。

命题 3.3。 设 $\mathbf{x} \in \mathcal{A}^{\mathbb{N}}$ 是一个无限词。

(1) 词 $\mathbf{x}$ 是因子平衡的，当且仅当它具有一致因子频率，并且对于 $\mathbf{x}$ 的任意因子 u ，存在一个常数 B_u ，使得对于任意字母 $a \in \mathcal{A}$，都有 $||u|_a - \mu_a|u|| \leq B_u$ ，其中 μ_a 是 a 的频率。*(2)* 词 $\mathbf{x}$ 是一致因子平衡的，当且仅当它具有一致因子频率，并且存在一个常数 B ，使得对于 $\mathbf{x}$ 的任意因子 u ，以及任意字母 $a \in \mathcal{A}$，都有 $||u|_a - \mu_a|u|| \leq B$ ，其中 μ_a 是 a 的频率。

3.2. **替换**。 给定字母表 $\mathcal{A}$，替换 是从 $\mathcal{A}$到 $\mathcal{A}^*$ 且与拼接相容的映射 τ，通常其从 $\mathcal{A}^*$ 到 $\mathcal{A}^*$ 的扩张也称为τ。若存在字母 a，使得 $\tau(a)$的首字母为 a，$\lim_{n\to+\infty} |\tau^n(a)| = +\infty$，并且对每个字母 b都有 $|\tau(b)| \neq 0$，则极限 $\tau^\omega(a)$: $=\lim_{n\to+\infty} \tau^n(a)$存在，并称为的不动点 τ 以 a开头。此时，我们称所生成的无限词是代换生成的。

替换 τ: $\mathcal{A}^* \to \mathcal{A}^*$ 是本原 的，如果存在整数 k ，使得对所有字母 a， $b \in \mathcal{A}$， a 都是 $\tau^k(b)$ 的因子。若$\mathbf{x}$ 是代换生成的，则容易验证：只要所涉及的替换是本原的，$\mathbf{x}$便是一致复现的。

如果对于任意 u ，只要它是 $v \in \mathcal{L}$ 的因子，就有 $u \in \mathcal{L}$，那么词集合 $\mathcal{L}$ （由 $\mathcal{A}$ 上的词构成）就是因子封闭的。如果 $u \in \mathcal{L}$ 蕴含存在字母 $a, b \in \mathcal{A}$，使得 $aub \in \mathcal{L}$ ，那么它就是双向可扩展的。对于替换 τ: $\mathcal{A}^* \to \mathcal{A}^*$，考虑由所有出现在某个 $\tau^n(a)$ 中的词 $u \in \mathcal{A}^*$ 构成的集合 L ，其中 $n \geq 1$ 且 $a \in \mathcal{A}$。于是， L 含有唯一一个包含关系下极大的、因子封闭且双向可扩展的词集合，我们称之为由 τ 生成的语言，并记作 $\mathcal{L}(\tau)$。如果对于每个满足 $\mathcal{L}(x) \subseteq \mathcal{L}(\tau)$ 的 $\mathbf{x} \in \mathcal{A}^{\mathbb{N}}$，$\mathbf{x}$ 也满足某个给定性质（例如，复现或 P-幂自由），那么我们称 $\mathcal{L}(\tau)$ 满足该性质。

字母 $a \in \mathcal{A}$ 对于 τ 是有界的，当且仅当 $\left(|\tau^n(a)|\right)_{n\geq 0}$ 有界。我们用 B_τ 表示 τ的有界字母集合，并定义 $\mathcal{C}_\tau = \mathcal{A} \setminus B_\tau$为 τ的增长字母集合。若 $\mathcal{C}_\tau = \mathcal{A}$，则称替换 τ 为增长替换，即每个字母都是增长字母。若存在 $n \geq 1$，使得对于所有 $a, b \in \mathcal{C}_\tau$， b都出现在 $\tau^n(a)$ 中，则称 τ 为拟本原。这

等价于 l-本原性这一概念，见 [MR16, Shi19]。注意，对于增长替换， τ 是本原的，当且仅当 τ 是拟本原的。若存在 $C > 0$ 使得

$$|\tau^n(a)| \leq C\,|\tau^n(b)| \text{ for all } n \geq 1 \text{ and } a, b \in \mathcal{A}.$$

最后， τ 当且仅当 $\mathcal{L}(\tau) \cap \mathcal{B}_\tau^*$ 为有限时，才称其为温和的。正如 [MR16, 定理 2.9] 和 [Shi19, 命题 3.17]所示，该定义等价于温和性的原始概念。

3.3. **偏差与平衡性**。 对于无限词$\mathbf{x} \in \mathcal{A}^{\mathbb{N}}$ 和 $u \in \mathcal{L}(\mathbf{x})$，若其在$\mathbf{x}$ 中的频率 $\mu_{\mathbf{x}}(u)$ 存在，则将其偏差定义为 $D_{\mathbf{x}}(u) = \sup_{n\geq 0} D_{\mathbf{x}}(u,n)$，其中

$$D_{\mathbf{x}}(u,n) = \sup\big\{\big||w|_u - |w|\mu_{\mathbf{x}}(u)\big| : w \in \mathcal{L}_n(\mathbf{x})\big\} \in \mathbb{R}.$$

定义 u 在$\mathbf{x}$ 中的平衡函数为 $B_{\mathbf{x}}(u) = \sup_{n\geq 0} B_{\mathbf{x}}(u,n)$，其中

$$B_{\mathbf{x}}(u,n) = \sup\big\{\big||t|_u - |s|_u\big| : t, s \in \mathcal{L}_n(\mathbf{x})\big\}.$$

偏差与平衡函数之间的基本关系由以下引理给出。该引理将 [Ada03 的命题 7] 从字母情形（形如 $|u|_a - |v|_a$ 且满足 $a \in \mathcal{A}$的表达式）推广到了词情形。

引理 3.4。 设$\mathbf{x}$ 是一个无限词，且有 $u \in \mathcal{L}(\mathbf{x})$。

(1) 如果 u 在$\mathbf{x}$中是 C-平衡的，则 u 在$\mathbf{x}$ 中的频率 $\mu_{\mathbf{x}}(u)$ 一致存在，且偏差 $D_{\mathbf{x}}(u)$ 由 C有界。*(2)* 如果偏差 $D_{\mathbf{x}}(u)$ 由 C有界，则 u 在$\mathbf{x}$中是 $2C$-平衡的。

证明。 我们证明该陈述的第一部分。假设 u在$\mathbf{x}$中是 C-平衡的。对于长度为 $n+|u|-1$ 的$z \in \mathcal{L}(\mathbf{x})$，我们定义 $g(z) = |z|_u/n$，并令 m_n（分别为 M_n）表示长度为 $n + |u| - 1$ 的 $z \in \mathcal{L}(\mathbf{x})$上 $g(z)$的最小值（分别为最大值）。首先，我们来证明

$$\limsup_{\ell\to\infty} M_\ell \leq M_n \tag{1}$$

对每个 $n \geq 0$均成立。对于长度为 $\ell + |u| - 1$ 且满足 $\ell \geq n$的 $w \in \mathcal{L}(\mathbf{x})$，我们令 $k = \lfloor \ell/n \rfloor$。对于每个 $i \in \{0, \ldots, k-1\}$，令 z_i表示 w的长度为 $n + |u| - 1$、出现在位置 in 处的因子，并令 z_k 表示从位置 kn开始的 w 的后缀。由于$|w|_u = \sum_{i=0}^{k} |z_i|_u$，我们得到

$$g(w) = \frac{|w|_u}{\ell} = \sum_{i=0}^{k} \frac{|z_i|_u}{\ell} = \frac{n}{\ell}\sum_{i=0}^{k-1} g(z_i) + \frac{|z_k|_u}{\ell}.$$

由于 $|z_k| \leq n$，我们得到 $g(w) \leq \frac{nk}{\ell} M_n + \frac{n}{\ell}$，由此可得 (1)。类似地，我们可以证明

$$m_n \leq \liminf_{\ell\to\infty} m_\ell \tag{2}$$

对于每个 $n \geq 0$。现在，由于 u 在 $\mathbf{x}$ 中是 C-平衡的，我们有 $M_n - m_n \leq \frac{C}{n}$ 对于每个$n \geq 0$。由不等式 (1) 和 (2)，我们推出 u 在$\mathbf{x}$ 中的频率 $\mu_{\mathbf{x}}(u)$ 存在，并且是序列 $(m_n)_{n\geq 0}$ 与 $(M_n)_{n\geq 0}$ 两者的极限。由不等式 (1)

以及 (2) 再一次，我们有 $m_n \le \mu_{\mathbf{x}}(u) \le M_n$ 对所有 $n \ge 0$，这意味着$D_{\mathbf{x}}(u) \le C$ 利用界 $M_n - m_n \le \frac{C}{n}$。

我们转而证明陈述中的第二项。假设 $D_{\mathbf{x}}(u)$ 有界于 C。那么，对于长度相同的 $w,\ w' \in \mathcal{L}_{\mathbf{x}}(u)$ ℓ，我们有

$$\big||w|_u - |w'|_u\big| \le \big||w|_u - \mu_{\mathbf{x}}(u)\ell\big| + \big||w'|_u - \mu_{\mathbf{x}}(u)\ell\big| \le 2C,$$

这表明 u 在 $\mathbf{x}$ 中是 $2C$-平衡的。 ■

3.4. $\mathcal{S}$-进表示。表示低复杂度词的一种高效方法是取替换的复合的极限。这里回顾基本定义，这些定义将在下文中发挥核心作用。

设 $(\mathcal{A}_n)_{n\ge 0}$ 是一个字母表序列，并且对每个 $n \ge 0$，考虑一个替换 $\tau_n\colon \mathcal{A}_{n+1}^* \to \mathcal{A}_n^*$。假设存在一个字母序列 $(b_n)_{n\ge 0}$，使得对每个 $n \ge 0$，每个像 $\tau_n(b_{n+1})$ 都以 b_n 开头。沿用 Berth´e、Karimov 和 Vahanwala [BKV25], 的术语，我们称 $\boldsymbol{\tau} = (\tau_n,\ b_n)_{n\ge 0}$ 为一个 相容序列。对于 $n > m \ge 0$，我们简写为

$$\tau_{m,n} = \tau_m \circ \tau_{m+1} \circ \tau_{m+2} \circ \cdots \circ \tau_{n-1}.$$

由于 $\tau_n(b_{n+1})$ 以 b_n 开头，因此 $\tau_{m,\ n}(b_n)$ 是 $\tau_{m,\ n+1}(b_{n+1})$ 的前缀。因此，对每个 $m \ge 0$，极限

$$\mathbf{x}^{(\mathbf{m})} = \lim_{n\to\infty} \tau_{m,n}(b_n)$$

存在于 $\mathcal{A}_m^{\mathbb{N}} \cup \mathcal{A}_m^*$ 中。此外，容易验证

$$\mathbf{x}^{(\mathbf{m})} = \tau_{m,n}(\mathbf{x}^{(\mathbf{n})}) \quad \text{for all } n > m \ge 0. \tag{3}$$

我们研究的主要对象是 $\mathbf{x} \coloneqq \mathbf{x}^{(\mathbf{0})}$，我们称之为由 $\boldsymbol{\tau}$ 生成的词。在这种情况下，$\boldsymbol{\tau}$ 称为 $\mathbf{x}$ 的 $\mathcal{S}$-进制展开。对于 $m \ge 1$，辅助无限词 $\mathbf{x}^{(\mathbf{m})}$ 在分析 $\mathbf{x}$ 时发挥着关键作用。

3.5. Sturmian 词。Sturmian 词的 $\mathcal{S}$进制展开的典型例子，是 Morse 和 Hedlund [MH40], 得到的展开，现将对此加以介绍。我们首先介绍 Sturmian 词，遵循 Berth´e、Holton 和 Zam- boni [BHZ06] 的表述。

令 α 为 (0，1) 中的一个无理数。考虑以下两个二区间交换变换$\widetilde{\ }$： $R_\alpha\colon [-\alpha, 1-\alpha) \to [-\alpha, 1-\alpha)$ 和 $R_\alpha\colon (-\alpha,\ 1-\alpha] \to (-\alpha,\ 1-\alpha]$，它们分别定义为

$$R_\alpha(z) = \begin{cases} z+\alpha, & \text{if } z \in [-\alpha, 1-2\alpha); \\ z+\alpha-1, & \text{if } z \in [1-2\alpha, 1-\alpha); \end{cases}$$

and

$$\widetilde{R}_\alpha(z) = \begin{cases} z+\alpha, & \text{if } z \in (-\alpha, 1-2\alpha]; \\ z+\alpha-1, & \text{if } z \in (1-2\alpha, 1-\alpha]. \end{cases}$$

两者都可以视为角度为 $2\pi\alpha$ 的旋转，因为识别点 $-\alpha$ 和 $1-\alpha$ 后，它们与圆周旋转共轭。

斜率为 α 的 *Sturmian*词 $\mathbf{x} = (x_n)_{n\geq 0} \in \{0,1\}^{\mathbb{N}}$ 是某个点 $\rho \in [-\alpha, 1-\alpha]$ 在这些变换之一作用下的正向轨道编码（相对于自然划分），该点称为 截距 $\mathbf{x}$，即:

(1) 对所有 $n \in \mathbb{N}$, $x_n = 0$，当且仅当 $R_\alpha^n(x) \in [-\alpha, 1-2\alpha)$,

(2) 或对所有 $n \in \mathbb{N}$, $x_n = 0$，当且仅当 $R_\alpha^n(x) \in [-\alpha, 1-2\alpha)$。

If $z = 0$，那么上述两种定义都会生成同一个词 $\mathbf{x}$，称为，在 th

是这种情况

斜率为 α 的特征 *Sturmian* 字。

现在我们描述特征 Sturmian 词的 $\mathcal{S}$-进展开。为此，令θ_0，θ_1: $\{0,1\}^* \to \{0,1\}^*$ 为如下定义的替换

$$\theta_0 : \begin{cases} 0 & \mapsto 0, \\ 1 & \mapsto 01; \end{cases} \quad \text{and} \quad \theta_1 : \begin{cases} 0 & \mapsto 10, \\ 1 & \mapsto 1. \end{cases}$$

我们进一步定义：对于 $n \geq 2$，若 n 为偶数，则为 $\theta_n = \theta_0$ ；若 n 为奇数，则为 $\theta_n = \theta_1$ 。以下结果是经典结论，最早由 Morse 和 Hedlund [MH40]证明。此处引用的版本沿用 Berth´e、Holton 和 Zamboni [BHZ06], 的表述，并采用上一小节引入的语言。

命题 3.5 ([MH40]). 设 α 是 $(0,1)$ 中的一个无理数，其连分数展开为 $[0; a_1, a_2, \ldots]$。对于 $n \geq 0$，若 n 为偶数，则定义 $b_n = 0$；否则定义 $b_n = 1$。那么， $\boldsymbol{\tau} = (\theta_n^{a_{n+1}}, b_n : n \geq 0)$ 是一个生成斜率为 α 的特征 *Sturmian* 词 $\mathbf{x} \in \{0,1\}^{\mathbb{N}}$ 的相宜指示序列，即，

$$\mathbf{x} = \lim_{n\to\infty} \theta_0^{a_1} \circ \theta_1^{a_2} \circ \cdots \circ \theta_n^{a_{n+1}}(b_{n+1}).$$

3.6. **Arnoux–Rauzy 词**。本节中，设 $\mathcal{A}$ 是一个大小为 $d \geq 2$ 的有限字母表。字母表 $\mathcal{A}$ 上的无限词 $\mathbf{x}$ 若为复现词，且其因子复杂度满足$\rho_{\mathbf{x}}(n) = (d-1)n + 1$ （对所有 $n \in \mathbb{N}$），并且对于所有 $n \in \mathbb{N}$，恰有一个长度为 n 的左特殊因子和一个右特殊因子，则称为 $Arnoux–Rauzy$ 词。本文采用其 $\mathcal{S}$-进展开的描述（例如，参见 [BCS13]），这是 Sturmian 词的 $\mathcal{S}$-进展开的推广。对于每个字母 $a \in \mathcal{A}$，我们定义 *Arnoux–Rauzy*替换σ_a: $\mathcal{A} \to \mathcal{A}^*$ 将 a 映射为 a，并将其他任意字母 $b \in \mathcal{A} \setminus \{a\}$ 映射为 ba。于是，我们得到 Arnoux–Rauzy 词关于 Arnoux–Rauzy 替换的如下描述，这与 Sturmian 词的命题 3.5 类似。

定理 3.6 ([AR91])。 无限词 $\mathbf{x}$ 在 $\mathcal{A}$ 上是 *Arnoux–Rauzy* 词，当且仅当它的因子集合与形式为下列形式的无限词 $\mathbf{y}$ 的因子集合相同

$$\mathbf{y} = \lim_{n\to\infty} \sigma_{a_0} \circ \sigma_{a_1} \circ \cdots \circ \sigma_{a_{n+1}}(a),$$

其中 a 和 a_n 是任意 $n \in \mathbb{N}$ 的字母，并且使得 $\mathcal{A}$ 中的每个字母都在序列 $(a_n)_{n\geq 0}$中无限次出现。

对于由 $\mathcal{A}$的字母构成的序列$(a_n)_{n\geq 0}$，记 $a_0a_1a_2\cdots = b_0^{k_0}b_1^{k_1}b_2^{k_2}\cdots$，使得$b_n \in \mathcal{A}$、$k_n \geq 1$，且对所有 $n \geq 0$ 有 $b_n \neq b_{n+1}$ （换言之，我们将$(a_n)_{n\geq 0}$中相等的连续字母分组）。则$(k_n)_{n\geq 0}$ 称为弱部分商序列。设$(K_n)_{n\geq 0}$ 为递增的整数序列，满足 $K_0 = 0$ 且

$$\{a_{K_n}, a_{K_n+1}, \ldots, a_{K_{n+1}}\} = \mathcal{A} \text{ and } \{a_{K_n}, a_{K_n+1}, \ldots, a_{K_{n+1}-1}\} \neq \mathcal{A}$$

对所有 $n \geq 0$。一阶差分序列$(K_{n+1} - K_n)_{n\geq 0}$ 称为强部分商序列。注意，这一术语来自相应的所谓*Arnoux–Rauzy*连分数算法 （另见 [AR91]）。

4. Generalities on factor-balanced words

我们定义了子移位之间高阶块码概念的一个自然类比，使其适用于单侧无限词的语境。一个高阶 k-块码是一个映射 $\pi\colon\ \mathcal{A}^k \to \mathcal{B}$，其中 $k \geq 1$和 $\mathcal{A}$，$\mathcal{B}$是字母表。我们令 π作用于长度为 $\ell + k$的词$w = a_0a_1\cdots a_{\ell+k-1}$，其定义为

$$\pi(w) = \pi(a_0a_1\cdots a_{k-1})\,\pi(a_1a_2\cdots a_k)\cdots\pi(a_\ell a_{\ell+1}\cdots a_{\ell+k-1}).$$

设$\mathbf{x} = (x_n)_{n\geq 0} \in \mathcal{A}^{\mathbb{N}}$ 是一个无限词。$\mathbf{x}$ 在 π 下的像是无限词$\pi(\mathbf{x}) \in \mathcal{B}^{\mathbb{N}}$，定义为

$$\pi(\mathbf{x}) = \pi(x_0x_1\cdots x_{k-1})\,\pi(x_1x_2\cdots x_k)\,\pi(x_2x_3\cdots x_{k+1})\cdots$$

我们称 π 在$\mathbf{x}$上**可逆**，如果存在一个高阶 ℓ-块码 $\rho\colon\ \mathcal{B}^\ell \to \mathcal{A}$，使得 $\rho(\pi(\mathbf{x}))$等于$\mathbf{x}$的第（$k + \ell$）次移位，即 $\rho(\pi(\mathbf{x})) = x_{k+\ell}\,x_{k+\ell+1}\cdots$。

高阶 k 块码的一个著名例子是所谓的长度为 k 滑动块码，它将不同的长度为 k 的因子映射为不同的字母。下面我们用一个例子来说明。此外，每个长度为 k 的滑动块码都是可逆的；粗略地说，其逆映射将每个长度为 k 的因子映射到它的第一个字母。

例 4.1。我们考虑随处可见的 Thue–Morse 词$\mathbf{t} = 01101001\cdots$，它是二元替换 $0 \mapsto 01$, $1 \mapsto 10$ 的不动点，参见 [Sea, A010060]。图 1 展示了它的~~长度为 2~~ 的滑动块码 $\pi(\mathbf{t})$。注意，由于$\mathbf{t}$ 有四个长度为 2 的因子， $\pi(\mathbf{t})$ 是一个由四个字母构成的无限词。更具体地说，因子 01、11、10 和 00 分别编码为 0,1, 2, 3，由此定义了 π。接着， π 在$\mathbf{t}$上可逆，其逆映射为 $\rho\colon 0, 3 \mapsto 0;\ 1, 2 \mapsto 1$。

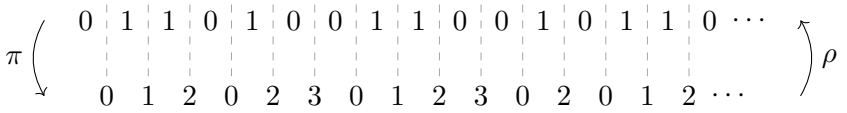


Figure 1. Thue–Morse 词$\mathbf{t} = 01101001\cdots$及其长度2滑动块码 $\pi(\mathbf{t}) = 012023\cdots$。

我们还给出了一个不可逆块码的例子。

例4.2. 我们再次考虑 Thue–Morse 词$\mathbf{t} = 01101001\cdots$。令 π为由 $\pi(ab) = a + b$ 模 2定义的二分块码，其中任意 $a, b \in \{0, 1\}$均适用，见图2。令$\mathbf{p} = \pi(\mathbf{t}) = (p_n)_{n\geq 0}$。注意，$\mathbf{p}$ 被称为倍周期，见 [Sea, A035263]，，且它也是2-自动的（见 [AS03]）。

我们证明， π 在$\mathbf{t}$上不可逆。容易验证，$\mathbf{p}$ 也满足$p_0 = 1$、 $p_1 = 0$以及以下递推关系

$$p_{4n+2} = p_{4n} = p_{2n}, p_{4n+3} = p_n, p_{8n+5} = p_{8n+1} = p_{4n+1}$$

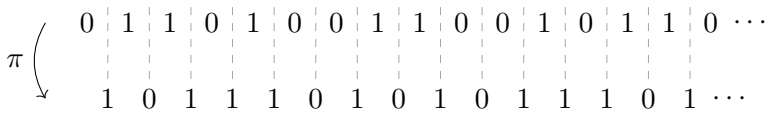


FIGURE 2. Thue–Morse 词 $\mathbf{t} = 01101001\cdots$ 及其长度2和块码。

对所有 $n \geq 0$均成立。假设 π 在$\mathbf{t}$ 上可逆，逆映射为 ρ：固定 $\{0,1\}^{\ell} \to \{0,1\}$ 和 $\ell \geq 1$，于是 $\rho(p_i p_{i+1} \cdots p_{i+\ell-1}) = t_{i+\ell+2}$。对所有 $k \geq 0$，$\mathbf{p}$ 的递推关系给出

$$p_{2^\ell k} p_{2^\ell k+1} \cdots p_{2^\ell k+\ell-1} = p_0 p_1 \cdots p_{\ell-1}.$$

因此，这意味着序列 $(t_{2^\ell k+\ell+2})_{k\geq 0}$ 是常数，但事实并非如此。

这个概念将子移位之间的高阶块码推广到了无限词的情境。事实上，存在一个块码 $\tilde{\pi}$，将传递子移位 X 映射到 Y，当且仅当存在一个传递点 $\mathbf{x} \in X$ 和一个块码 π，使得$\pi(\mathbf{x}) \in Y$；关于符号动力系统的一般背景，见 [LM95]。下面这个简单引理留给读者证明。

引理 4.3。设 $\mathbf{x}$、$\mathbf{y}$ 是满足 $\mathcal{L}(\mathbf{x}) = \mathcal{L}(\mathbf{y})$的无限词。那么，$\mathbf{x}$ 一致因子平衡，当且仅当$\mathbf{y}$ 一致因子平衡。

现在我们关注因子平衡性如何通过高阶块码得以保持。

命题 4.4。设 $k \in \mathbb{N}$ 为正的整数，设 $\mathbf{x} \in \mathcal{A}^{\mathbb{N}}$ 为无限词，并设 $\pi\colon\ \mathcal{A}^k \to \mathcal{B}$ 为高阶 k-块码。

(1) 如果$\mathbf{x}$ 是因子平衡的，那么 $\pi(\mathbf{x})$ 也是因子平衡的。*(2)* 如果 $\mathbf{x}$ 是一致因子平衡的，且 π 在 $\mathbf{x}$ 中可逆，那么 $\pi(\mathbf{x})$ 是一致因子平衡的。

证明。我们证明命题的第一部分。我们证明 $\pi(\mathbf{x})$ 中的每个因子 w 都是对某个常数 C 而言的 C-平衡的。设 w 是 $\pi(\mathbf{x})$ 的一个因子。由于$\mathbf{x}$ 是因子平衡的，对于$\mathbf{x}$ 的每个因子 w'，存在一个常数 $C_{w'}$，使得 w' 在 $\mathbf{x}$ 中是 $C_{w'}$-平衡的，并定义 $W = \{w' \in \mathcal{L}(\mathbf{x}) \cap \mathcal{A}^{|w|+k}\colon\ \pi(w') = w\}$。我们证明存在一个常数 $K > 0$，使得对于所有 $u,\ v \in \mathcal{L}(\pi(\mathbf{x}))$ 且 $|u| = |v|$，都有

$$\big||u|_w - |v|_w\big| \leq K \cdot \mathrm{Card}(W), \tag{4}$$

也就是说，w在 $\pi(\mathbf{x})$ 中是 $(K\,\mathrm{Card}(W))$-平衡的。令 $q = |u| = |v|$。由于 $u,\ v \in \mathcal{L}(\pi(\mathbf{x}))$，存在长度为 $q+k$ 的 $u',\ v' \in \mathcal{L}(\mathbf{x})$，使得 $\pi(u') = u$ 且 $\pi(v') = v$。于是，对所有 $w' \in W$，都有 $\big||u'|_{w'} - |v'|_{w'}\big| \leq C_{w'}$。对于每个 $t \in \{u, v\}$，根据 W 的定义，并且由于 $t = \pi(t')$，我们得到 $|t|_w = \sum_{w' \in W} |t'|_{w'}$。因此，

$$\big||u|_w - |v|_w\big| \leq \sum_{w' \in W} \big||u'|_{w'} - |v'|_{w'}\big| \leq \mathrm{Card}(W) \max_{w' \in W} C_{w'}.$$

这就对 $K := \max_{w' \in W} C_{w'}$ 证明了 (4)，由此可得 $\pi(\mathbf{x})$ 是因子平衡的。接下来，我们证明该陈述的第二部分；为此，只需证明出现在 (4) 中的常数 $\mathrm{Card}(W)$ 和 K 独立于 w 有界。如果

x 进一步是一致 C-因子平衡的，则 $K := \max_{w' \in W} C_{w'}$ 以 C 为上界。因此，只需证明 Card(W) 独立于w有界。设 π 具有逆映射 ρ，即 ρ 是一个 ℓ-块码，使得 $\rho(\pi(\mathbf{x})) = x_{k+\ell}x_{k+\ell+1}\cdots$。令 $S = \{s \in \mathcal{L}(\pi(\mathbf{x})) \cap \mathcal{B}^{|w|+k+\ell} : \ \rho(s) \in W\} \cap \mathcal{L}(\pi(\mathbf{w}))$。于是，对任意 $s \in S$，$\pi(\rho(s)) = w$ 是 s 的一个后缀。这表明 $S \subseteq \{t\,w : \ t \in \mathcal{B}^{k+\ell}\}$，由此可得 Card($S$) $\le$ Card($\mathcal{B}^{k+\ell}$)。由于 $\rho(S) = W$，我们得到 Card(W) $\le$ Card($\mathcal{B}^{k+\ell}$)。我们得出 K 独立于 w有界，如所欲证。■

注 4.5。 注意，若我们仅考虑字母平衡性，则前一命题的第一项不成立。例如，Thue–Morse 词是字母平衡的，而倍周期词则不是。事实上，将 τ定义为 $\tau(1) = 10$ 且 $\tau(0) = 11$，则$\mathbf{x} = \tau^{\omega}(1)$ 是倍周期词。对 $n \ge 0$ 定义 $u_0 = 1$ 与 $u_{n+1} = \tau^2(u_n)1$，即

$$u_n = \tau^{2n}(1)\,\tau^{2(n-1)}(1)\cdots\tau^2(1)\,1.$$

我们来证明 $u_n \in \mathcal{L}(\mathbf{x})$。对 $n = 0$ 而言这是显然的；若 $u_n \in \mathcal{L}(\mathbf{x})$，则对某个 $a \in \{0,1\}$ 有 $u_n a \in \mathcal{L}(\mathbf{x})$，因此 $\tau^2(u_n a) = \tau^2(u_n)\tau^2(a)$ 也属于 $\mathcal{L}(\mathbf{x})$。由于 $\tau^2(0)$ 与 $\tau^2(1)$ 都以 1 开头，$u_{n+1} = \tau^2(u_n)1$ 是 $\tau^2(u_n)\tau^2(a)$ 的前缀，因而它属于 $\mathcal{L}(\mathbf{x})$。

现在，我们通过估计字母 1 的偏差，证明字母 1 在 $\mathbf{x}$ 中不是平衡的。字母 1 在 $\mathbf{x}$ 中的频率 μ_1 由 τ 的关联矩阵的归一化右佩龙特征向量给出，即 $\begin{pmatrix} 0 & 2 \\ 1 & 1 \end{pmatrix}$；在此情形下，$\mu_1 = 2/3$。$\tau^{2n}$ 的关联矩阵为

$$\frac{1}{3}\begin{pmatrix} 2+4^n & 2(4^n-1) \\ 4^n-1 & 2\cdot 4^n+1 \end{pmatrix}.$$

因此，

$$|u_n|_1 - \mu_1|u_n| = \sum_{0\le i\le n}\big(|\tau^{2i}(1)|_1 - \mu_1|\tau^{2i}(1)|\big) = \frac{1}{3}\sum_{0\le i\le n}\big(2\cdot 4^i + 1 - 2\cdot 4^i\big) = \frac{n+1}{3},$$

也就是说，字母 1 的偏差无界。

现在我们证明，因子平衡性迫使无限词中的复现具有一致性。

命题 4.6。 设 $\mathbf{x}$ 是一个因子平衡的无限词，$\mathbf{y}$ 是一个复现的无限词，且 $\mathcal{L}(\mathbf{y}) \subseteq \mathcal{L}(\mathbf{x})$。

(1) 词 $\mathbf{y}$ 是一致复现的。*(2)* 对每个复现的无限词 $\mathbf{y}'$，若满足 $\mathcal{L}(\mathbf{y}') \subseteq \mathcal{L}(\mathbf{x})$，则有 $\mathcal{L}(\mathbf{y}) = \mathcal{L}(\mathbf{y}')$.

特别地，每个因子平衡的回归无穷词都是一致复现的。

证明。 设$\mathbf{y}$,$\mathbf{y}'$ 是回归无穷词，且 $\mathcal{L}(\mathbf{y})\ \mathcal{L}(\mathbf{y}') \subseteq \mathcal{L}(\mathbf{x})$，并任取一个因子 $t \in \mathcal{L}(\mathbf{y})$。我们断言

$$\text{there exists } \ell \ge 1 \text{ such that every length-}\ell \text{ factor of } \mathbf{y}' \text{ contains } t. \tag{5}$$

我们先证明断言 (5) 蕴含该陈述。根据断言 (5)，$t \in \mathcal{L}(\mathbf{y}')$，此外，$t$在$\mathbf{y}'$ 中出现，且间隙有界，界为 ℓ。由于 t 是任意的，我们得到$\mathcal{L}(\mathbf{y}) \subseteq \mathcal{L}(\mathbf{y}')$。由对称性（交换$\mathbf{y}$ 和$\mathbf{y}'$），可得 $\mathcal{L}(\mathbf{y}) = \mathcal{L}(\mathbf{y}')$，并且每个

这个公共集合中的词都以有界间隔同时出现在$\mathbf{y}$ 和$\mathbf{y}'$中。特别地，取$\mathbf{y}' =\mathbf{y}$ 即可得到条目 (1)。

剩下需要证明断言 (5)。令 $t \in \mathcal{L}(\mathbf{x})$，设 C_t 为 t在$\mathbf{x}$中的平衡常数，也就是说，对于任意两个长度相等的因子 $u, v \in \mathcal{L}(\mathbf{x})$，都有 $\big||u|_t - |v|_t\big| \leq C_t$。由于$\mathbf{y}$ 是复现的，且 $t \in \mathcal{L}(\mathbf{y})$，因此存在一个$\mathbf{y}$ 的因子 p ，其中至少包含 t的$C_t + 1$ 次出现。令 $\ell := |p|$。现在，取$\mathbf{y}'$ 的任意因子 w ，满足 $|w| = \ell$。那么$p, w \in \mathcal{L}(\mathbf{x})$，因此由$\mathbf{x}$的因子平衡性，

$$|w|_t \geq |p|_t - \big||w|_t - |p|_t\big| \geq |p|_t - C_t.$$

由于 $|p|_t \geq C_t + 1$，我们得到 $|w|_t \geq 1$。因此， t 出现在$\mathbf{y}'$的每个长度为 ℓ 的因子中，这就证明了断言 (5)。■

现在我们给出一个一致复现但不平衡的词的例子 .

例 4.7. 考虑所谓的*Chacon*词$\mathbf{c} = 0012001212012\cdots$，它是本原替换 $0 \mapsto 0012, 1 \mapsto 12, 2 \mapsto 012$ 的不动点；参见 [Sea, A049321]。已知$\mathbf{c}$ 是非周期、一致复现的，且具有线性因子复杂度，参见 [Fer95], ，但对于任意词 $w \in \{0, 1, 2\}^*$， w 在$\mathbf{c}$中都不是平衡的。事实上，已知$\mathbf{c}$ 生成一个弱混合拓扑动力系统 [Cha69], ，而这类系统已知不存在平衡因子 [BCB19]。

在词组合学中，若非空词 $w \in \mathcal{A}^*$ 与自身重叠，我们称之为如此，当且仅当ww 包含超过两次 w的出现。否则，我们称 w 不与自身重叠；注意，此时 ww 恰好包含两次 w 的出现，一次作为前缀，一次作为后缀。

引理 4.8 ([Lot02, Chapter 1])。 *</style}*一个词是本原的，当且仅当它不与自身重叠。

引理 4.9。 设$\mathbf{x}$ 是一个无限词，且 $u \in \mathcal{L}(\mathbf{x})$ 非空。如果对于某个 $C > 0$ 和 $u^{2C+3} \in \mathcal{L}(\mathbf{x})$， u 在$\mathbf{x}$ 中是 C-平衡的，那么$\mathbf{x}$ 是最终周期的，且其周期整除$|u|$。

证明。不失一般性，可假设 u 是本原的。我们用反证法，假设$\mathbf{x}$ 不是最终周期的且周期整除 $|u|$。那么，由于 $u^{2C+3} \in \mathcal{L}(\mathbf{x})$，存在一个词 $w \in \mathcal{L}(\mathbf{x})$，其长度为 $|u^{2C+3}|$ ，以 u^{C+1}为前缀，但不是 u^{C+2}。我们将 $w = u^{C+1}ts$分解，其中 $|t| = |u|$ 和 $|s| = (C+1)|u|$; 注意 $t \neq u$。现在， u^{2C+3} 至少包含 $C + 2$ 次 u^{C+2}的出现，因此，由于 u在$\mathbf{x}$中是 C-平衡的，$\mathbf{x}$ 的每个长度为 $|u^{2C+3}|$ 的因子至少有两次u^{C+2}的出现。这特别适用于 w；设 $i < j$ 是 u^{C+2}在 w中出现的两个不同位置。

另一方面，我们注意到 $i \notin \{0, |u|, \ldots, C|u|\}$ 因为否则， u^{C+2} 将是 w的前缀，这与 w 的选取相矛盾。这意味着，如果 $i \leq C|u|$，那么 u 会与自身重叠；根据引理 4.8，这是不可能的，因为 u 是本原的。因此， $i > C|u|$。另一方面，显然有 $j \leq |w|-|u^{C+2}| = |s| = (C+1)|u|$。由此可得 $C|u| < i < j \leq (C + 1)|u|$。因此， u^{C+2} 在 w 中出现时所对应的区间 $[i, i + |u^{C+2}|)$ 和 $[j, j + |u^{C+2}|)$ 至少在大小为 $|u|$的窗口内重叠，从而产生了 u 与自身的重叠。根据引理 4.8，这就构成了矛盾。■

因此，根据前一个引理，我们证明以下两个关于幂避免的结果。

推论 4.10. 以常数 C 一致因子平衡的非周期递归词是 $(2C+3)$-幂自由的。

证明. 若 $\mathbf{x}$ 不是 $(2C+3)$-幂自由的，则由引理 4.9 可知它是最终周期的。由于$\mathbf{x}$ 是复现的，这意味着它是周期的，与假设矛盾。■

引理 4.11. 设 $\mathbf{x} \in \mathcal{A}^{\mathbb{N}}$ 对某个整数 $P \geq 1$是 P-幂自由的。则对任意 u， $|v|_u \leq \frac{P|v|}{|u|}$ $v \in \mathcal{L}(\mathbf{x})$。

证明. 反设 $|v|_u > \frac{P|v|}{|u|}$ 对某个非空的 $u, v \in \mathcal{L}(\mathbf{x})$。由抽屉原理， u 存在两次出现，分别始于 v的位置 i 与$i+\ell$，且满足 $\ell \leq |u|/P$。这两次出现发生重叠，从而迫使 u 的长度为 ℓ的前缀 t 至少重复 $P+1$ 次。换言之， t^{P+1} 出现在v的位置 i，这与 $\mathbf{x}$ 是 P-幂自由的相矛盾。■

5. Factor-balancedness in linearly recurrent words

本节旨在刻画哪些线性复现词是均匀因子平衡的。我们的方法依赖于 Durand [DHS99] 提出并在第 5.1 节介绍的这些词的 $\mathcal{S}$-进制描述。线性复现词特有的代换结构，使我们能够将因子平衡性的研究问题，化约为该结构不同层级上的字母平衡性问题。所得刻画见定理 1.2，其证明将在第 5.3 节给出。作为直接推论，我们重新得到 Adamczewski 关于代换词的类似结果 [Ada03, Ada04], ，即命题 1.3。

5.1. 线性复现词的代换描述。提醒读者，对于相容序列 $\boldsymbol{\tau} = (\tau_n\colon \mathcal{A}_{n+1}^* \to \mathcal{A}_n^*,$ $a_n \in \mathcal{A}_n)_{n\geq 0}$，记号 $\tau_{m,\,n}$，其中 $n > m \geq 0$，表示复合 $\tau_m \circ \tau_{m+1} \circ \cdots \circ \tau_{n-1}$。此外，若对于所有 $a \in \mathcal{A}_n$ 和 $b \in \mathcal{A}_m$， b 都出现在 $\tau_{m,\,n}(a)$ 中，我们就称 $\tau_{m,\,n}$ 为正的 。

下面的经典定理给出了线性复现无限词的一种具体刻画，借助的是代换结构。回顾一下，替换$\tau\colon$ $\mathcal{A}^* \to \mathcal{A}^*$ 是左适当的，如果存在一个字母 a，使得每个 τ（b）都以 a开头。

定理 5.1 （[DHS99]）。 无限词$\mathbf{x}$ 是线性复现的，当且仅当存在一个生成$\mathbf{x}$ 的相容序列 $\boldsymbol{\tau} =$ （τ_n, a_n）$_{n\geq 0}$ ，满足：

（**P**）存在 $k \in \mathbb{N}$ 使得$\tau_{n,\,n+k}$ 对所有 $n \geq 0$都是正的；（**F**）集合 $\{\tau_n :$ $n \geq 0\}$ 是有限的；并且（ **L** ）替换 τ_n 对所有 $n \geq 0$ 都是左适当的。

定理 5.1 中的相容序列 $\boldsymbol{\tau}$ 的原型例子对应于代换词。更准确地说，设 $\tau\colon$ $\mathcal{A}^* \to \mathcal{A}^*$ 是一个左适当的本原替换，且 $a \in \mathcal{A}$。那么，常值序列 $\boldsymbol{\tau} := (\tau,\ a)$ $_{n\geq 0}$ 是相容的，并满足定理 5.1 的三个条件。这表明，左适当本原替换的任何不动点都是线性复现的。事实上，任意本原替换的任何不动点（不一定是适当的）都是线性复现的 [DHS99]。因此，

线性复现词推广了本原代替词。这一推广十分广泛，因为代换词只有可数多个，而线性复现词却有不可数多个。例如，在 Sturmian 词中，代换生成的词对应于斜率为二次代数数的情形，而线性复现词对应于其连分数系数有界的斜率（这是一个勒贝格测度为零的不可数集合），参见定理 6.1。

现在，我们汇集定理 5.1 中关于线性递归词及其代换生成结构的一些事实。回想一下，对于替换 $\tau\colon \mathcal{A}^* \to \mathcal{B}^*$，我们使用记号$|\tau| = \max_{a\in\mathcal{A}} |\tau(a)|$。

命题 5.2 ([Dur98,DHS99])。 设$\mathbf{x}$ 是由相容序列 $\boldsymbol{\tau} = (\tau_n, a_n)_{n\geq 0}$ 生成并满足定理 *5.1* 的条件$(\mathbf{P})$ 和$(\mathbf{F})$ 的线性递归词。那么，存在 $K > 0$ 使得：

(1) 对每个 $n \geq 0$，都有 $|\tau_n| \leq K$ 和 $|\mathcal{A}_n| \leq K$；并且 *(2)* 对所有 a、$b \in \mathcal{A}_n$ 和 $n \geq 1$，都有$|\tau_{0,\,n}(a)| \leq K \cdot |\tau_{0,\,n}(b)|$。

证明。 根据假设，集合 $\{\tau_n : n \geq 0\}$ 是有限的，因此存在满足第（1）项的 $K_1 > 0$。此外，[DHS00] 中的引理 8 给出了一个常数 K_2，使第（2）项成立。因此，$K = \max(K_1, K_2)$ 满足命题的陈述。■

引理 5.3。 设$\mathbf{x}$ 是一个常数为 L的非周期线性递归词。那么：

(1) 词$\mathbf{x}$ 是$(L+1)$ 幂自由的。*(2)* 对任意 u，$v \in \mathcal{L}(\mathbf{x})$，$|v|_u \leq (L+1)\,\frac{|v|}{|u|}$。*(3)* 每个$(\mathbf{x})$ 的频率都存在，且满足 $\mu_{\mathbf{x}}(u) \leq \frac{L}{|u|}$。

证明。 第（1）项见 [DHS99], 中的定理 24，第（3）项见 [Dur00]中的命题 13。第（2）项由第（1）项和引理 4.11 得出。■

5.2. **决定性**。 现在我们引入一个概念，该概念在定理 1.2 的证明中发挥着关键作用，并介绍其证明中将使用的性质。

定义 5.4。 设$\mathbf{x} \in \mathcal{A}^{\mathbb{N}}$ 是一个无限词，且有 $k \in \mathbb{N}$。替换 $\tau\colon \mathcal{A}^* \to \mathcal{B}^*$ 是 k-决定性的，关于 $\mathbf{x}$，如果存在一个映射 $r\colon \mathcal{A} \to \mathcal{B}^k$，使得对于所有 a，$b \in \mathcal{A}$（满足 $ab \in \mathcal{L}(\mathbf{x})$），$\tau(b)$ 均以 $r(a)$ 开头。对于长度为 $|u| \leq k+1$ 的 $u \in \mathcal{B}^*$ 和 $a \in \mathcal{A}$，我们定义记号 $q_{\tau,\,\mathbf{x}}(u,\,a)$ 如下。对于 $a \in \mathcal{A}$，令 $r_{|u|}(a)$ 表示 $r(a)$ 的长度为 $|u|-1$ 的前缀，然后令

$$q_{\tau,\mathbf{x}}(u,a) = |\tau(a)\, r_{|u|}(a)|_u. \tag{6}$$

我们可以如下改述决定性：如果字母 a 出现在$\mathbf{x}$ 中的某个位置 i，那么这不仅确定了块 $\tau(a)$ 会在 $\tau(\mathbf{x})$ 中的预期位置 $j := |\tau(\mathbf{x}^{[0,i)})|$ 出现，还确定了扩展块 $\tau(a)\,r(a)$ 会在位置 j 出现。

决定性可以看作适当性的推广，具体如下所述。回忆一下，若存在某个字母 a，使得每个 $\tau(b)$、$b \in \mathcal{A}$ 都以 a 开头，则称替换 $\tau\colon \mathcal{A}^* \to \mathcal{A}^*$ 是左适当的。注意，在这种情况下，τ 有唯一的不动点，即 $\tau^\omega(a)$。

命题 5.5. 设 $\tau\colon \mathcal{A}^* \to \mathcal{B}^*$ 是一个左真替换，在 $\mathbf{x}$ 中为其不动点。那么 τ 在1-决定性方面具有 $\mathbf{x}$。

证明。 令 $r(b) = a$ 为 $b \in \mathcal{A}$，这表明左真替换在其唯一不动点中总是 1-决定性的。■

更一般地，对于任意替换 $\tau\colon \mathcal{A}^* \to \mathcal{A}^*$，其长度2滑动块码在其任意一个不动点中都是$k$-决定性的，其中 $k = \min\{|\tau(a)|\colon a \in \mathcal{A}\} - 1$。此外，如果某个替换 τ 在某个 $\mathbf{x}$ 中是1-决定性的，那么 τ^n 在 $\mathbf{x}$ 中是 k_n-决定性的，其中 $k_n = \min\{|\sigma^n(a)|\colon a \in \mathcal{A}\}$。

决定性替换的一个重要性质是，它使我们能够仅利用字母在 $\mathbf{x}$ 中的出现次数，以及一些取决于 u、但与 u 出现的位置无关的数据，来计算给定短词 u 在 $\tau(\mathbf{x})$ 中的出现次数。下一个引理将对此作出精确说明。

命题 5.6。 设 $\tau\colon \mathcal{A}^* \to \mathcal{B}^*$ 是 $\mathbf{x} \in \mathcal{A}^{\mathbb{N}}$ 中的一个 k-决定性替换。假设$w \in \mathcal{A}^*$ 在 $\mathbf{x}$ 中的位置 $i \in \mathbb{N}$ 处出现，并令 $p = |\tau(\mathbf{x}^{[0,i+|w|)})|$ 为 $\tau(w)$ 在 $\tau(\mathbf{x})$ 中结束处的自然位置。那么，对于长度为 $|u| \le k+1$ 的每个 $u \in \mathcal{L}(\tau(\mathbf{x}))$，都有

$$|\tau(w)\,\tau(\mathbf{x})_{[p,p+|u|)}|_u = \sum_{a \in \mathcal{A}} |w|_a \, q_{\tau,\mathbf{x}}(u,a). \tag{7}$$

证明。 我们需要统计起始位置落在区间$[p-|\tau(w)|, p+|u|)$内的 u 的出现次数，其中 u 是 $\tau(\mathbf{x})$的子词。对每个 $j \in \{i, i+1, \dots, i+|w|-1\}$，考虑 $\tau(x_j x_{j+1}) = \tau(x_j)\tau(x_{j+1})$。由于 τ是 k-决定性的，且 $|u| \le k+1$，所以 $\tau(x_{j+1})$的前 $|u|-1$ 个字母是固定的词$r_{|u|}(x_j)$，与x_{j+1}无关。因此，起始位置在 $\tau(x_j)$内（可能延伸至 $\tau(x_{j+1})$中）的 u的出现次数等于

$$\big|\tau(x_j)\, r_{|u|}(x_j)\big|_u = q_{\tau,\mathbf{x}}(u, x_j).$$

当 j在 $\{i, i+1, \dots, i+|w|-1\}$ 中变化时，这些起始位置集合两两不相交，且恰好覆盖区间 $[p-|\tau(w)|, p+|u|)$内 u 的所有起始位置。因此，

$$|\tau(w)\,\tau(\mathbf{x})_{[p,p+|u|)}|_u = \sum_{i \le j < i+|w|} q_{\tau,\mathbf{x}}(u, x_j) = \sum_{a \in \mathcal{A}} |w|_a \, q_{\tau,\mathbf{x}}(u,a)$$

因为 $w = x_i x_{i+1} \cdots x_{i+|w|-1}$。 ■

我们在此记录以下简单的观察，稍后将会用到。

引理 5.7。 *:* 设 $\tau\colon \mathcal{A}^* \to \mathcal{B}^*$ 和 $\sigma\colon \mathcal{B}^* \to \mathcal{C}^*$ 是替换。假设 τ 在词 $\mathbf{x} \in \mathcal{A}^{\mathbb{N}}$中是 k-决定性的。那么， $\sigma\circ\tau$ 在 $\mathbf{x}$中是 ℓ-决定性的，其中 $\ell = k \cdot \min\{|\sigma(b)|\colon b \in \mathcal{A}\}$。

回想一下，对于替换 $\tau\colon \mathcal{A}^* \to \mathcal{B}^*$，我们使用记号 $\|\tau\| = \sum_{a \in \mathcal{A}} |\tau(a)|$。

命题 5.8。 设 $\mathbf{x} \in \mathcal{A}^{\mathbb{N}}$ 是 B-字母平衡的，其字母频率为 ν_a，对应于 $a \in \mathcal{A}$。假设 $\tau\colon \mathcal{A}^* \to \mathcal{B}^*$在 $\mathbf{x}$中是 k-决定性的，且 $\tau(\mathbf{x})$是 P-幂自由的。固定长度为 $|u| \le k+1$ 的 $u \in \mathcal{L}(\tau(\mathbf{x}))$，并使用 (6) 中定义的记号，定义

$$\mu_u = \frac{\sum_{a \in \mathcal{A}} \nu_a \, q_{\tau,\mathbf{x}}(u,a)}{\sum_{a \in \mathcal{A}} \nu_a \, |\tau(a)|}. \tag{8}$$

那么，对于任意的 $w \in \mathcal{L}(\mathbf{x})$，

$$\big||\tau(w)|_u - \mu_u |\tau(w)|\big| \le \frac{2PB\|\tau\|}{|u|} + 2P. \tag{9}$$

特别地， u 在 $\tau(\mathbf{x})$ 中的频率存在，且等于 μ_u。

证明。令 $q_{\tau,\mathbf{x}}(u,w) := \sum_{a\in\mathcal{A}} |w|_a\, q_{\tau,\mathbf{x}}(u,a)$ 为式 (7) 中出现的量。根据命题 5.6，$q_{\tau,\mathbf{x}}(u,w) = |\tau(w)v|_u$，其中某个词 v 的长度为 $|u|-1$。利用引理 4.11，我们可以给 $|q_{\tau,\mathbf{x}}(u,w) - |\tau(w)|_u| \le 2P$ 一个上界。因此，只需证明 $\big|q_{\tau,\mathbf{x}}(u,w) - \mu_u|\tau(w)|\big|$ 也以上界 $2PB\|\tau\|/|u|$ 为界。

对于 $a\in\mathcal{A}$，令 $\varepsilon_a = |w|_a - \nu_a|w|$，并注意到根据引理 3.4 有 $|\varepsilon_a| \le B$。我们可以计算

$$q_{\tau,\mathbf{x}}(u,w) = \sum_{a\in\mathcal{A}} |w|_a\, q_{\tau,\mathbf{x}}(u,a) = |w| \sum_{a\in\mathcal{A}} \nu_a\, q_{\tau,\mathbf{x}}(u,a) + \sum_{a\in\mathcal{A}} \varepsilon_a\, q_{\tau,\mathbf{x}}(u,a). \tag{10}$$

右侧的第二项很小：由引理 4.11 可得 $q_{\tau,\,\mathbf{x}}(u,\ a) \le P|\tau(a)|/|u|$，因此，由于 $|\varepsilon_a| \le B$，

$$\sum_{a\in\mathcal{A}} \varepsilon_a\, q_{\tau,\mathbf{x}}(u,a) \le B\cdot\frac{P\|\tau\|}{|u|}. \tag{11}$$

接下来，我们证明（10）右侧的第一项接近 $\mu_u\,|\tau(w)|$。利用 ε_a 的定义，我们可以写成

$$|\tau(w)| = \sum_{a\in\mathcal{A}} |w|_a\,|\tau(a)| = |w| \sum_{a\in\mathcal{A}} \nu_a\,|\tau(a)| + \sum_{a\in\mathcal{A}} \varepsilon_a\,|\tau(a)|.$$

由此以及（8）中 μ_u 的定义，我们得到

$$\Big||w| \sum_{a\in\mathcal{A}} \nu_a\, q_{\tau,\mathbf{x}}(u,a) - \mu_u|\tau(w)|\Big| = \mu_u \Big|\sum_{a\in\mathcal{A}} \varepsilon_a\,|\tau(a)|\Big|.$$

注意，根据引理 4.11，$q_{\tau,\,\mathbf{x}}(u,\ a) \le P|\tau(a)|/|u|$，因此 $\mu_u \le P/|u|$。由此以及 $|\varepsilon_a| \le B$ 可得估计式：

$$\Big||w| \sum_{a\in\mathcal{A}} \nu_a\, q_{\tau,\mathbf{x}}(u,a) - \mu_u|\tau(w)|\Big| \le \frac{P}{|u|}\cdot B\cdot\|\tau\|. \tag{12}$$

最后，我们可以先利用 (10) 替换 $\big|q_{\tau,\mathbf{x}}(u,w) - \mu_u|\tau(w)|\big|$ 中的 $q_{\tau,\mathbf{x}}(u,w)$，再用 (11) 和 (12) 估计剩余各项，从而得到其上界，即

$$\big|q_{\tau,\mathbf{x}}(u,w) - \mu_u|\tau(w)|\big| = \Big||w| \sum_{a\in\mathcal{A}} \nu_a q_{\tau,\mathbf{x}}(u,a) - \mu_u|\tau(w)| + \sum_{a\in\mathcal{A}} \varepsilon_a q_{\tau,\mathbf{x}}(u,a)\Big| \le \frac{2BP\|\tau\|}{|u|},$$

如所需。 ■

5.3. 线性递归词的一致因子平衡性。 本节的主要结果是定理 1.2。为清楚起见，我们再次陈述该定理，并给出线性递归词具有（统一）因子平衡性的条件。判据基于生成该词且满足与定理 5.1 中条件相同条件的相容序列。不过，考虑条件 **(L)** 的以下弱化形式会更方便。这对于稍后处理替代情形至关重要。设 $\boldsymbol{\tau} = (\tau_n,\, a_n)_{n\ge 0}$为一个相容序列，并令 $\mathbf{x}^{(\mathbf{n})}$ 表示 $(\tau_{n,m}(a_m))_{m>n}$ 的极限。那么，$\boldsymbol{\tau}$ 满足 **(D)**，当且仅当

（**D**）替换 τ_n 在 $\mathbf{x}^{(\mathbf{n+1})}$ 中对每个 $n\ge 0$ 都是决定性的。

定理 *1.2.* 设 x 是一个线性递归词，且 $(\tau_n, a_n)_{n\geq 0}$ 是生成 x 并满足条件 *(P)*、*(F)* 和 *(D)* 的相容序列。令 $x^{(\mathbf{n})}$ 表示 $(\tau_{n,m}(a_m))_{m>n}$的极限。假设存在一个序列 $(B_n)_{n\geq 0}$ ，使得词 $x^{(\mathbf{n})}$是 B_n-字母平衡的。那么词 x 是因子平衡的。此外，如果序列 $(B_n)_{n\geq 0}$ 是有界的，那么词 $x^{(\mathbf{n})}$ 是一致因子平衡的。

在给出定理 1.2 的证明之前，我们指出，该结果的一个特殊情形正是 Adamczewski 的结果，即第 1.1 节中所述的命题 1.3。

命题 *1.3* 的证明。 由于必要条件是显然的，我们只证明充分条件。假设 $\mathbf{x}$ 的每个长度为 2 的因子都是字母平衡的。令$\tau\colon\ \mathcal{A}^*\to\mathcal{A}^*$ 为一个本原替换，使得 $\mathbf{x}=(x_n)_{n\geq 0}$ 是其不动点之一。考虑 τ的长度2滑动块码 $\tau^{[2]}$，令 $\bar{a}=[x_0x_1]\in\mathcal{A}^{[2]}$，并令$\pi\colon\ \mathcal{A}^{[2]}\to\mathcal{A}$ 为将二分块码解码回 $\mathcal{A}$中字母的映射。于是， $\tau^{[2]}$在 $\bar{a}$上是可延长的，且所生成的不动点 $\bar{\mathbf{x}}$ 满足 $\pi(\bar{\mathbf{x}})=\mathbf{x}$。每个长度为 2 的因子都平衡这一假设意味着，对于某个C， $\bar{\mathbf{x}}$ 是 C-字母平衡的。

我们现在对所有 $n\geq 0$定义 $\tau_n=\tau^{[2]}$ 与 $a_n=\bar{a}$ ，并令 $\boldsymbol{\tau}=(\tau_n,\, a_n)_{n\geq 0}$。于是， $\boldsymbol{\tau}$ 是适宜的并生成 $\bar{\mathbf{x}}$。我们来验证 $\boldsymbol{\tau}$ 满足定理 1.2的假设。首先，定理 5.1的条件（$\mathbf{P}$）与（$\mathbf{F}$）均成立，因为 $\boldsymbol{\tau}$ 是常数（每个 τ_n都等于本原替换 $\tau^{[2]}$）。至于（$\mathbf{D}$），注意由于 $\boldsymbol{\tau}$ 是常数，$\boldsymbol{\tau}$ 的所有极限点 $\bar{\mathbf{x}}^{(\mathbf{n})}$ 都与 $\bar{\mathbf{x}}$重合。因此，由于滑动分块码是决定性的，每个$\tau_n=\tau^{[2]}$ 在 $\bar{\mathbf{x}}^{(\mathbf{n+1})}=\bar{\mathbf{x}}$中都是决定性的。从而，由于对所有 $n\geq 0$都有 $\bar{\mathbf{x}}^{(\mathbf{n})}=\bar{\mathbf{x}}$ ，每个 $\bar{\mathbf{x}}^{(\mathbf{n})}$ 因而都是C-字母平衡的。

如上一段所示，定理 1.2 的假设均得到满足，因此 $\bar{\mathbf{x}}$ 是一致因子平衡的。最后，由命题 4.4 以及 π 是可逆分块码这一事实，我们得出$\mathbf{x}$也是一致因子平衡的。■

现在我们给出定理 1.2 的证明。我们先介绍一个技术引理，它指出，对于定理 1.2 中这类代换结构，任意词 u 都能找到一个合适的结构层级 n，其尺度$|\tau_{0,\ n}|$ 与该词的长度u相当。这是有效控制命题 5.8 所给界的关键步骤。

引理 5.9. 设 $\boldsymbol{\tau}=(\tau_n\colon\ \mathcal{A}_{n+1}^*\to\mathcal{A}_n,\, a_n\in\mathcal{A}_n)_{n\geq 0}$ 是一个相容序列，满足命题 *5.2* 的条件（*1*）和（*2*），常数为 $K>0$，并且满足条件$(\mathbf{D})$。令$\mathbf{x}^{(\mathbf{m})}$ 表示$(\tau_{m,n},\, a_n)_{n>m}$的极限。那么，对于每个 $u\in\mathcal{L}(\mathbf{x}^{(\mathbf{0})})$，存在 $n\geq 1$使得

(i) 替换$\tau_{0,n}$ 在 $\mathbf{x}^{(\mathbf{n})}$中是 $|u|$-决定性的；并且 *(ii)*

对所有 $a\in\mathcal{A}_n$，我们有 $|\tau_{0,n}(a)|\leq K^3|u|$ 。

特别地，对所有 $a\in\mathcal{A}_n$，$|u|\leq|\tau_{0,n}(a)|\leq K^3|u|$ 。

证明。设 $u\in\mathcal{L}(\mathbf{x}^{(\mathbf{0})})$。由于当 $n\to\infty$ 时，$|\tau_{0,n}(a_n)|\to\infty$ ，故存在最大的 $n\geq 1$，使得对所有 $a\in\mathcal{A}_n$ 有 $|\tau_{0,n}(a)|\leq K^3|u|$ 。于是， n 满足第 (ii) 项。

我们证明 n 也满足条件 (i)。根据条件 (**D**)， τ_{m-1} 对每个 $m\geq 1$ 在 $\mathbf{x}^{(m)}$ 中都是决定性的这一事实，由引理 5.7 可推出 $\tau_{0,n}=\tau_{0,n-1}\circ\tau_{n-1}$ 在 $\mathbf{x}^{(\mathbf{n})}$ 中是 ℓ_n-决定性的，其中 $\ell_n=\min\{|\tau_{0,n-1}(a)|\ :\ a\in\mathcal{A}_{n-1}\}$。因此，要得到条件 (i)，只需

证明 $|u| \le \ell_n$。利用 n 的极大性，我们可以找到 $b \in \mathcal{A}_{n+1}$ 使得$K^3|u| < |\tau_{0,n+1}(b)|$，因此可以估计

$$K^3|u| < |\tau_{0,n+1}(b)| \le |\tau_{n-1} \circ \tau_n| \cdot |\tau_{0,n-1}|.$$

命题 5.2 的第 (1) 项给出 $|\tau_{n-1} \circ \tau_n| \le |\tau_{n-1}| \cdot |\tau_n| \le K^2$，而命题 5.2 的第 (2) 项给出 $|\tau_{0,\,n-1}| \le K\ell_n$。因此我们得出 $K^3|u| < K^3\ell_n$，从而 $|u| \le \ell_{n+1}$，正如所需。■

定理 *1.2* 的证明。我们固定一个常数为 L 的线性复现词 $\mathbf{x}$，并令$\boldsymbol{\tau} = (\tau_n\colon\ \mathcal{A}_{n+1}^* \to \mathcal{A}_n^*,\ \ a_n \in \mathcal{A}_n)_{n\ge 0}$ 为定理 5.1 给出的生成 $\mathbf{x}$ 的相容序列。我们进一步定义 K 为命题 5.2 中出现的常数。

假设对于所有 $n \ge 0$，都存在 $B_n > 0$，使得 $\mathbf{x}^{(\mathbf{n})}$ 是 B_n-字母平衡的。此外，假设 $\mathbf{x}$ 是非周期的，因为否则结论显然成立。我们证明，$\mathbf{x}$ 的任意给定因子 u 的偏差均由 $2(L+1)(B_nK^4+2K^3+1)$ 有界，其中 n 仅取决于 u。由引理 3.4，这意味着 $\mathbf{x}$ 是因子平衡的，正如所需。

令 $u, w \in \mathcal{L}(\mathbf{x})$。我们的目标是估计量 $\big||w|_u - \mu_{\mathbf{x}}(u)|w|\big|$。我们先引入必要的对象。根据假设，可以对 $\boldsymbol{\tau}$和 u应用引理 5.9，从而得到一个整数 $n \ge 1$，使得

(i) 替换$\tau_{0,n}$ 在 $\mathbf{x}^{(\mathbf{n})}$ 中是 $|u|$-决定性的；

(ii) 对所有 $a \in \mathcal{A}_n$，都有$|\tau_{0,n}(a)| \le K^3|u|$。

按照 (8)，我们定义

$$\mu_u = \frac{\sum_{a\in\mathcal{A}_n} \nu_a\, q_{\tau_{0,n},\mathbf{x}^{(\mathbf{n})}}(u,a)}{\sum_{a\in\mathcal{A}_n} \nu_a\, |\tau_{0,n}(a)|},$$

其中，ν_a 是 a在 $\mathbf{x}^{(\mathbf{n})}$中的频率；由命题 3.2 可知该频率存在，因为 $\mathbf{x}^{(\mathbf{n})}$ 是字母平衡的。由于 w 出现在 $\tau_{0,n}(\mathbf{x}^{(\mathbf{n})})$ 中，因此存在 $w' \in \mathcal{L}(\mathbf{x}^{(\mathbf{n})})$，使得 w 出现在$\tau_{0,n}(w')$ 中。我们取这样的最短词。

现在，我们通过$\tau_{0,n}(w')$中字母的偏差，来估计 u 在 w 中的偏差。更准确地说，我们将其分解为

$$|w|_u - \mu_u|w| = \underbrace{|w|_u - |\tau_{0,n}(w')|_u}_{:=T_1} + \underbrace{|\tau_{0,n}(w')|_u - \mu_u|\tau_{0,n}(w')|}_{:=T_2} + \underbrace{\mu_u|\tau_{0,n}(w')| - \mu_u|w|}_{:=T_3}, \quad (13)$$

并分别对 $|T_1|$、$|T_2|$和 $|T_3|$给出界。

我们先给出一些一般性观察。由于 w 出现在$\tau_{0,\,n}(w')$ 中，因此对于某个 $p,\ s \in \mathcal{A}^*$，我们可以写成$p\,w\,s = \tau_{0,\,n}(w')$。此外，$w'$ 的极小性确保：若 α和 β 分别是 w'的首字母和末字母，则 p 是 $\tau_{0,\,n}(\alpha)$ 的前缀，而 s 是$\tau_{0,\,n}(\beta)$ 的后缀。因此，根据上文第 (ii) 项，我们有

$$|p| \le K^3|u| \ \text{ and } \ |s| \le K^3|u|. \quad (14)$$

此外，由于 $\mathbf{x}$ 是非周期且线性复现的，复现常数为 L，由引理 5.3 可知

$$\mu_u \le \frac{L}{|u|} \ \text{ and } \ |t|_u \le \frac{(L+1)|t|}{|u|} \ \text{ for every } t \in \mathcal{L}(\mathbf{x}). \quad (15)$$

现在我们估计各项 T_1，T_2，T_3。由于 $\tau_{0,\,n}(w') = pws$，注意到 $|T_1|$ 对应于 u 在$\tau_{0,\,n}(w')$ 中的出现次数，这些出现要么始于前缀 p，要么结束于

后缀 s 内。由于 $|p|,\ |s| \le K^3|u|$ 根据（14），我们可以将（15）与 $t \in \{p, s\}$ 结合使用，从而得到界

$$|T_1| \le 2(L+1)K^3. \tag{16}$$

对于 T_3，我们观察到 $|\tau_{0,\,n}(w')| - |w| = |s| + |p|$，因此使用 (14) 和 (15) 可得

$$|T_3| = \mu_u||\tau_{0,n}(w')| - |w|| = \mu_u\,(|p| + |s|) \le 2(L+1)K^3. \tag{17}$$

最后，为了控制 $|T_2|$，注意到由于 $\mathbf{x^{(n)}}$ 是 B_n-字母平衡的，并且 $\tau_{0,\,n}(\mathbf{x^{(n)}}) = \mathbf{x}$ 根据引理 5.3 是 $(L+1)$-幂自由的，我们可以使用命题 5.8 得到

$$|T_2| \le \frac{2(L+1)B_n\|\tau_{0,n}\|}{|u|} + 2(L+1).$$

由于 $\|\tau_{0,\,n}\| \le |\mathcal{A}_n| \cdot |\tau_{0,\,n}|$，命题 5.2 的第 (1) 项给出的界 $|\mathcal{A}_n| \le K$ 以及上述第 (ii) 项给出的界 $|\tau_{0,\,n}| \le K^3|u|$ 最终得到

$$|T_2| \le 2(L+1)B_nK^4 + 2(L+1) = 2(L+1)(B_nK^4 + 1). \tag{18}$$

结合 (16)、(17) 和 (18)，我们得到

$$\begin{aligned}\big||w|_u - \mu_u|w|\big| &\le |T_1| + |T_2| + |T_3| \\ &\le 2(L+1)K^3 + 2(L+1)(B_nK^4+1) + 2(L+1)K^3 \\ &\le 2(L+1)(B_nK^4 + 2K^3 + 1).\end{aligned} \tag{19}$$

如所需。

为证明该陈述的最后一部分，如果序列 $(B_n)_{n\ge0}$ 有界，则（19）证明词 $\mathbf{x}$ 是 C-一致因子有界的，其中 $C = 2(L+1)(\sup_{n\ge0} B_n\,K^4 + 2K^3 + 1)$。■

人们可能会问（19）中所得界的最优性。在下一个例子中，我们将更明确地研究 Tribonacci 词的情形。

例 5.10。定义 Tribonacci 替换 $\tau\colon\ \mathcal{A}^* \to \mathcal{A}^*$ 为 $a \mapsto ab$、$b \mapsto ac$，以及 $c \mapsto a$，其唯一不动点为 $\mathbf{t}$，称为 ***Tribonacci*** 词。于是，相容序列 $\boldsymbol{\tau} = (\tau,\ a)_{n\ge0}$ 生成 $\mathbf{t}$。此外，由于 $\boldsymbol{\tau}$ 为常数，所有极限点均为由 $\boldsymbol{\tau}$ 生成的 $\mathbf{t^{(n)}}$，且都等于 $\mathbf{t}$。我们确定以下量：

- 由于已知 $\mathbf{t}$ 是 2 字母平衡的，参见 [RSZ10]，以及 $\mathbf{t^{(n)}} =$ 对所有 $n \ge 0$ 均成立，因此我们得到 $B_n = 2$ 对所有 $n \ge 0$ 均成立。
- 命题 5.2 中的所有量 $|\mathcal{A}|$、$|\tau|$ 以及 $\sup\{\frac{|\tau^n|}{|\tau^n\ \begin{smallmatrix}a\\b\end{smallmatrix}|}()\colon\ a、b \in \mathcal{A}、\ n \ge 1\}$ 均由 3 有界，因此我们可以选择 $K = 3$。
- 线性递归常数满足 $L = 2\psi^2 + \psi + 1 \approx 9.605$，其中 ψ 是 Tribonacci 根。有关如何计算该常数的更多细节（使用定理证明器 Walnut），请参阅附录 A。

因此，(19) 中的界等于 $2(2\psi^2 + \psi + 2)(2\cdot 3^4 + 2\cdot 3^3 + 1) \approx 4602.6734$，也就是说，平衡常数至多为 4602。注意，Tribonacci 词的最优常数为 2，已在 [CDO+25，定理 22] 中证明（另见定理 1.1）。

注 5.11。 在定理 1.2 中，可以将线性复现的假设替换为幂自由且满足命题 5.2 的条件 (1) 和 (2)。事实上，应用引理 5.3 和命题 5.2 时才用到了这一假设。然而，根据引理 4.11 以及平衡字母的频率总是存在这一事实，在幂自由性条件下，引理 5.3 的相同结论仍然成立，详见引理 3.4。因此，可以将我们的定理重新表述如下。

定理 *5.12.* 设 x 是一个幂自由词，且 $(\tau_n, a_n)_{n\geq 0}$ 是生成 x、满足条件 *(P)*、*(F)* 和 *(D)* 的相容序列。假设命题 *5.2* 的条件 *(1)* 和 *(2)* 均成立。令 $x^{(\mathbf{n})}$ 表示 $(\tau_{n,m}(a_m))_{m>n}$ 的极限。假设存在一个序列 $(B_n)_{n\geq 0}$，使得词 $x^{(\mathbf{n})}$ 是 B_n-字母平衡的。那么词 x 是因子平衡的。此外，若序列 $(B_n)_{n\geq 0}$ 有界，则词 $x^{(\mathbf{n})}$ 是一致因子平衡的。

我们将在下一节的一个具体例子中应用这一小幅推广；其中所考察的词是一致因子平衡的，但不是复现的。

6. Uniform factor-balancedness in Arnoux–Rauzy words

本节的目标是证明下面回顾的定理 1.4。

定理 1.4。 字母表中至多有3 个字母的 *Arnoux–Rauzy* 词一致因子平衡，当且仅当其弱部分系数有界。

我们先讨论字母表含两个字母的情形，即 Sturmian 词，然后讨论三元 Arnoux-Rauzy 词的情形。

6.1. Sturmian 词。

下一条结果的原始来源将其表述为子移位，而非无限词；但由于 Sturmian 词一致复现 [MH40]，两种表述等价。

定理 6.1 ([Dur00，命题 10])。 设 $\mathbf{x}$ 是斜率为 $\alpha \in (0,1)$的 *Sturmian* 词。那么，$\mathbf{x}$ 线性复现，当且仅当其连分数的系数有界。

二元情形下定理 *1.4* 的证明。 充分条件已由 Fagnot 和 Vuillon [FV02，推论 13], 证明，不过他们使用了不同的术语。

还需证明必要条件。由于斜率为无理数α 、截距为 ρ 的 Sturmian词与斜率为 α 的特征 Sturmian 字具有相同的因子集合（参见 [Lot02，命题 2.1.18]），根据引理 4.3，我们只需考虑 $\mathbf{x}$ 是这个特征 Sturmian 字的情形。于是，根据命题 3.5，$\mathbf{x}$ 具有$\mathcal{S}$-进展开

$$\mathbf{x} = \lim_{n\to\infty} \theta_0^{a_1} \circ \theta_1^{a_2} \circ \cdots \circ \theta_1^{a_{2n}}(0),$$

其中 $\alpha = [0;\ a_1,\ a_2,\ \ldots]$ 是 α 的连分数。对所有整数 n，定义词 $u_n = \theta_0^{a_1} \circ \theta_1^{a_2} \circ \cdots \circ \theta_{\bar{n}}^{a_n}(0)$ 和 $v_n = \theta_0^{a_1} \circ \theta_1^{a_2} \circ \cdots \circ \theta_{\bar{n}}^{a_n}(1)$，其中 $\bar{n} = n$ 模 2。于是，当 n 为偶数时，$u_{n+1} = u_n^{a_n} v_n$；当 n 为奇数时，$v_{n+1} = v_n^{a_n} u_n$。由于 $u_{n+1},\ v_{n+1} \in \mathcal{L}(\mathbf{x})$ 和 $\mathbf{x}$是一个非周期的 C-一致因子平衡词，我们可以应用推论 4.10，得到对所有 n，$a_n \leq 2C+3$。因此，α 的连分数系数均由 $2C+3$ 有界。最后，注意到 $(a_n)_{n\geq 1}$ 构成弱部分商序列，证明完毕。■

注 6.2。 我们提到，也可以使用与下一小节类似的技巧，借助定理 5.12 证明前一个结果的充分条件。

定理 6.1 立即推出以下推论。

推论 6.3。 *Sturmian*词 $\mathbf{x}$ 一致因子平衡，当且仅当$\mathbf{x}$ 是线性复现的。

6.2. 三**字母表上的 Arnoux–Rauzy 词**。 已知，有界的强部分商刻画了线性复现的Arnoux–Rauzy词。

定理 6.4 ([BST19，命题 8。4])。 任意字母表上的 *Arnoux-Rauzy* 词是线性复现的，当且仅当它具有有界的强部分商；也就是说，每个 *Arnoux–Rauzy*替换都以有界间隔出现在其指示序列中。

请注意，在Sturmian词的情形下，强部分商与弱部分商的概念相同。另一方面，关于弱部分商有界的条件，只能给出 Arnoux–Rauzy词字母平衡性的充分条件。事实上，Berth´e 等人 [BCS13]证明了以下结果。

定理 6.5 ([BCS13，定理 7])。 设$\mathbf{x}$ 是定义在含有3个字母的字母表上的 *Arnoux-Rauzy* 词，其 $\mathcal{S}$指示序列为$(i_m)_{m\geq 0}$。若弱部分商有界于C，则$\mathbf{x}$ 是$(2C+1)$-字母平衡的。

然而，前述结果的逆命题并不成立，因为 [BCS13]的作者给出了一个弱部分商无界的二字母平衡 Arnoux–Rauzy 词的例子。

为了证明三元情形下的定理 1.4，我们将利用定理 5.12。注意，我们也可以用这种方法证明二元情形。在此之前，我们先证明下面两个引理。请注意，它们对任意大小的字母表都成立。

引理 6.6。 设 $\boldsymbol{\tau}=(\tau_n)_{n\geq 0}$ 是一列定义在 $\mathcal{A}$ 上的 *Arnoux–Rauzy*替换，其弱部分系数有界。对于所有 a、 $b\in\mathcal{A}$，序列$\left(|\tau_{0,\ n}(a)|/|\tau_{0,\ n}(b)|\right)_{n\geq 0}$一致有界。

证明。 给定一个作用在 $\mathcal{A}$ 上的替换 η，令 $T(\eta)=\max_{a,\ b\in\mathcal{A}}|\eta(a)|/|\eta(b)|$。注意，对于任意 Arnoux–Rauzy 替换 α，我们有

$$T(\eta\alpha)\leq T(\eta)+1. \tag{20}$$

此外，直接计算字母计数可得，对于任意不同的 Arnoux–Rauzy 替换 α，β 以及任意替换 η

$$T(\eta\alpha\beta)\leq \frac{1}{2}T(\eta)+2. \tag{21}$$

我们首先证明这个不等式在二元情形下成立。设 $\alpha=\sigma_a,\beta=\sigma_b$with $a\neq b$。注意到 $\sigma_a\circ\sigma_b$ 将 $a\mapsto aba, b\mapsto ba$映射到。假设 $\eta(a)\geq\eta(b)$，则 $T(\eta)=\eta_{(a)\,\eta(b)}$，于是

$$T(\eta\sigma_a\sigma_b)=\frac{|\eta(aba)|}{|\eta(ba)|}=1+\frac{|\eta(a)|}{|\eta(ba)|}\leq 1+\frac{|\eta(a)|}{2|\eta(b)|}\leq 1+\frac{T(\eta)}{2}.$$

当 $\eta(b) > \eta(a)$ 时，证明方法类似。现在，假设 $|\mathcal{A}| \geq 3$，并设 $\alpha = \sigma_a, \beta = \sigma_b$ with $a \neq b$。注意到 $\sigma_a \circ \sigma_b$ 将 $a \mapsto aba, b \mapsto ba, c \mapsto caba$ 映射到，其中 $c \in \mathcal{A} \setminus \{a, b\}$。令 $c_m \in \mathcal{A} \setminus \{a, b\}$（分别地，$c_M$）满足：$|\eta(c_m)|$ 是 $\{\eta(c) : c \in \mathcal{A} \setminus \{a, b\}\}$ 中最短（分别地，最长）的词。因此，$\eta\alpha\beta$ 的最短可能像是 $\eta(ba)$，最长的像是 $\eta(c_M aba)$。于是，我们有

$$T(\eta\sigma_a\sigma_b) = \frac{|\eta(c_M aba)|}{|\eta(ba)|} = 1 + \frac{|\eta(c_M a)|}{|\eta(ba)|} \leq 2 + \frac{|\eta(c_M)|}{|\eta(ba)|} \leq 2 + \frac{|\eta(c_M)|}{2|\eta(c_m)|} \leq 2 + \frac{1}{2}T(\eta).$$

因此，对于 $\mathcal{A}$的所有长度，(21)均已得到证明。

现在，由于 $\boldsymbol{\tau}$ 具有有界的弱部分系数，存在一个递增的正整数序列$(n_j)_{j\geq 0}$，使得对于所有 $j \geq 0$，$\tau_{n_j} \neq \tau_{n_j+1}$，并且对于某个常数 $C \geq 0$，有 $2 \leq n_{j+1} - n_j \leq C$。由(21)，我们得到

$$T(\tau_{0,n_j+2}) = T(\tau_{0,n_j} \circ \tau_{n_j} \circ \tau_{n_j+1}) \leq \frac{1}{2}T(\tau_{0,n_j}) + 2$$

对于每个 $j \geq 0$均成立。此外，对于所有 $j \geq 1$，我们有 $0 \leq n_j - (n_{j-1} + 2) \leq C - 2$，因此通过归纳地应用(20)，前一个不等式变为

$$\begin{aligned} T(\tau_{0,n_j+2}) \leq \frac{1}{2}T(\tau_{0,n_j}) + 2 &\leq \frac{1}{2}T(\tau_{0,n_{j-1}+2}) + \frac{C-2}{2} + 2, \\ &\leq \frac{1}{2}T(\tau_{0,n_{j-1}+2}) + \frac{C}{2} + 1 \end{aligned}$$

对于每个 $j \geq 1$均成立。对于所有 $j \geq 1$，反复应用前一个不等式可得如下界

$$T(\tau_{0,n_j+2}) \leq \frac{1}{2^j}T(\tau_{0,n_0+2}) + (C+2) \tag{22}$$

当 $\sum_{i=0}^{j-1} \frac{1}{2^i}(\frac{C}{2}+1) \leq C+2$时。在指标 $n_{j+1}+2$与 n_j+2之间有 $n_{j+1}-n_j \leq C$ 次替换，因此至多应用 C 次(20)，得到对于所有 $m \in [n_j + 2, n_{j+1} + 2)$，$T(\tau_{0,m}) \leq T(\tau_{0,n_j+2}) + C$。结合这一结果与(22)，我们得出 $T(\tau_{0,m}) \leq \frac{1}{2^j}T(\tau_{0,n_0+2}) + 2C + 2$，从而得到一个一致的界。■

引理 6.7. 若某个 *Arnoux–Rauzy* 词在字母表 $\mathcal{A}$ 上具有有界的弱部分系数，则它是 $P \geq 0$-幂无子的。

证明。设 x 是字母表 $\mathcal{A}$ 上的 Arnoux–Rauzy 词，由指示序列 $\boldsymbol{\tau} = (\tau_n)_{n\geq 0}$ 给出，且具有有界的弱部分系数。设 C_1 是弱部分系数的一个上界。由引理 6.6，存在一个常数 $C_2 \geq 0$，使得$|\tau_{0,n}(a)|/|\tau_{0,n}(b)| \leq C_2$ 对所有 $a, b \in \mathcal{A}$ 和 $n \geq 0$ 均成立。令 $P = 16C_2^2(C_1 + 5)$。我们用反证法，假设 x 不是 P-幂无子的。设 $u \in \mathcal{L}(\mathrm{x})$ 非空，且满足 $u^P \in \mathcal{L}(\mathrm{x})$。不妨设 u 是本原的，即 u 不是某个严格更短的词的幂。证明分为三个步骤：第一步是找到一个整数 n 和一个足够长的词 $v \in \mathcal{L}(\mathrm{x}^{(n+1)})$，使得 $\tau_{0,n+1}(v)$ 出现在 u^P 中，且其周期 $|u|$ 远小于$|\tau_{0,n}(a)|$ 对所有 $a \in \mathcal{A}$ 均成立。接着，第二步证明，这种强周期性迫使 v 实际上具有 $v = v_1\lambda^{|v|-1}$ 的形式，其中 $v_1, \lambda \in \mathcal{A}$ 是某些字母。最后，我们利用弱部分商有界这一假设，证明这样的 v 不存在。

步骤 1：首先，令 $n \geq 1$ 为满足以下条件的最小整数

$$|\tau_{0,n}(a)| \geq 4|u| \tag{23}$$

对于所有 $a \in \mathcal{A}$。注意，对于所有 a，$a' \in \mathcal{A}$ 和 $m \geq 1$，根据 Arnoux–Rauzy 替换的定义以及 C_2 的定义，可得

$$|\tau_{0,m+1}(a)| \leq 2 \max_{b \in \mathcal{A}} |\tau_{0,m}(b)| \leq 2C_2 |\tau_{0,m}(a')|.$$

将此不等式应用两次（当 $m = 0$ 时，$\tau_{0,m}$ 为恒等映射）可得$|\tau_{0,n+1}(a)| \leq 2C_2|\tau_{0,n}(a')| \leq 4C_2^2|\tau_{0,n-1}(a')|$ 对所有 $a, a' \in \mathcal{A}$ 均成立。现在，由于 n 的极小性，存在 $a' \in \mathcal{A}$ 使得 $|\tau_{0,n-1}(a')| < 4|u|$。因此，

$$|\tau_{0,n+1}(a)| \leq 4C_2^2 |\tau_{0,n-1}(a')| < 16C_2^2|u| \tag{24}$$

对所有 $a \in \mathcal{A}$ 均成立。令 $\mathbf{x}^{(n+1)}$ 为由 $(\tau_m)_{m>n+1}$ 生成的 Arnoux–Rauzy 词。由于 u^P 出现在 $\mathbf{x}$ 中，存在 $v \in \mathcal{L}(\mathbf{x}^{(n+1)})$，使得 $\tau_{0,n+1}(v)$ 出现在 u^P 中。我们取长度最大的 v。由 v 的极大性可知，$|u^P| \leq |\tau_{0,n+1}(v)| + 2 \max_{b \in \mathcal{A}} |\tau_{0,n+1}(b)|$，进而得到

$$P|u| = |u^P| \leq |\tau_{0,n+1}(v)| + 2 \max_{b \in \mathcal{A}} |\tau_{0,n+1}(b)| \leq (|v| + 2) \cdot 16C_2^2|u|$$

利用（24）。因此，我们有

$$|v| \geq \frac{P}{16C_2^2} - 2 \geq C_1 + 3.$$

步骤 2：我们证明，v 以某个字母 $\lambda \in \mathcal{A}$ 结尾。注意到，Arnoux– Rauzy 替换是右适当的，即每个像 $\tau_n(a)$ 都以同一个固定字母结尾；设该字母为 a_n。因此，$\tau_{0,\,n+1}(a)$ 对所有 $a \in \mathcal{A}$ 都以 w_n 结尾：$= \tau_{0,\,n}(a_n)$。于是，根据（23），$|w_n| \geq 4|u|$。此外，对于 v 的每个非空前缀 pa，若 $p \in \mathcal{A}^*$ 且 $a \in \mathcal{A}$，则在 $\tau_{0,\,n+1}(v)$ 中存在 $\tau_{0,\,n+1}(a)$ 的一次出现，其结束位置为 $\ell(pa)$：$= |\tau_{0,\,n+1}(pa)|$。令 $v = v_1 v_2 \cdots v_{|v|}$，其中对所有 $i \in \{1, \ldots, |v|\}$ 均有 $v_i \in \mathcal{A}$。我们断言

$$|\tau_{0,n+1}(v_i)| \equiv 0 \pmod{|u|} \tag{25}$$

对所有 $i \in \{2, 3, \ldots, |v|\}$。事实上，对每一个这样的 i，$\tau_{0,n+1}(v_i)$ 在 $\tau_{0,n+1}(v)$ 中有一次出现，结束于 $\ell(v_1 \cdots v_i)$。由于 $\tau_{0,n+1}(v_i)$ 以 w_n 结尾，w_n 在 $\tau_{0,n+1}(v)$ 中有一次出现，结束于同一位置 $\ell(v_1 \cdots v_i)$。因此，若对两个不同的 i, i' 有 $\ell(v_1 \cdots v_i) \neq \ell(v_1 \cdots v_{i'})$（模 $|u|$），则由于 $\tau_{0,n+1}(v)$ 出现在 u^P 中，w_n 在无限词 u^∞ 中出现于模 $|u|$ 互不相同的位置。由于 $|w_n| \geq 4|u|$，我们可以应用 Fine–Wilf 定理（例如参见 [Lot83，命题 1.3.5]），由此可得 u 不是本原词，这与 u 的定义相矛盾。因此，v 的所有前缀 $v_1 \cdots v_i$ 都产生相同的余数 $\ell(v_1 \cdots v_i)$（模 $|u|$）。这意味着对所有 $2 \leq i \leq |v|$，有 $|\tau_{0,n+1}(v_i)| = \ell(v_1 \cdots v_i) - \ell(v_1 \cdots v_{i-1}) = 0$（模 $|u|$），从而如所期望地证明了 (25)。

接下来，由于 $\tau_{0,\,n+1}(v)$ 出现在 u^P 中，并且对所有 $2 \leq i \leq k$ 使用（25）中的同余关系，每个 $\tau_{0,\,n+1}(v_i)$ 都以同一个字母开头，记为 λ。现在，根据 Arnoux– Rauzy 替换的定义，对于所有 $a \in \mathcal{A}$ 和 $m \geq 0$，$\tau_m(a)$ 都以 a 开头，因此，对于所有 i，$\tau_{0,\,n+1}(v_i)$ 也以 v_i 开头。因此，对所有 $2 \leq i \leq |v|$，都有 $v_i = \lambda$。这意味着$v = v_1 \lambda^{|v|-1}$。

步骤 3：因此，我们得到 $\lambda^{C_1+2} \in \mathcal{L}(\mathbf{x}^{(n+1)})$，因为 $|v| \geq C_1 + 3$。现在我们证明这与弱部分系数的有界性假设矛盾。注意，Arnoux–Rauzy 词 $\mathbf{x}^{(n+1)}$ 的弱部分系数也由 C_1 有界。因此，对于某些 Arnoux–Rauzy 替换 σ_a，我们有 $\mathbf{x}^{(n+1)} = \sigma_a^r(\sigma_b(\mathbf{x}^{(n+r+2)}))$

以及 σ_b、 $a \neq b$ 和某个 $1 \leq r \leq C_1$，其中 $\mathbf{x}^{(n+r+2)}$ 是由 $(\tau_m)_{m>n+r+2}$ 生成的 Arnoux–Rauzy 词。我们有

$$\sigma_a^r \sigma_b : a \mapsto aba^r,\ b \mapsto ba^r,\ c \mapsto ca^r ba^r$$

对 $c \notin \{a, b\}$。由于 $\mathbf{x}^{(n+1)}$ 是由形如 $\sigma_a^r(\sigma_b(c))$ 的词无限拼接而成的，且$c \in \mathcal{A}$，我们推得，对于所有 $c \in \mathcal{A}$，都有 $c^{r+2} \notin \mathcal{L}(\mathbf{x}^{(n+1)})$。因此，由于 $\lambda^{C_1+2} \in \mathcal{L}(\mathbf{x}^{(n+1)})$，必有 $C_1 + 2 < r + 2$，这与 $r \leq C_1$ 矛盾。由此可知，词 $\mathbf{x}$不含 P 次幂。■

现在我们可以在三元情形下证明定理 1.4。请注意，必要条件对每个字母表都成立。

三元情形下定理 *1.4* 的证明。 我们先证明必要条件。此时 $\mathbf{x}$ 有如下 $\mathcal{S}$ 进制展开

$$\mathbf{x} = \lim_{n\to\infty} \sigma_{b_0}^{k_0} \circ \sigma_{b_1}^{k_1} \circ \cdots \circ \sigma_{b_n}^{k_n}(a),$$

其中 $a \in \mathcal{A}$、 $b_i \in \mathcal{A}$，且 $(k_n)_{n\geq 0}$ 是弱部分商序列。对于$c \in \mathcal{A}$，定义 $x_n^{(c)} = \sigma_{b_0}^{k_0} \circ \sigma_{b_1}^{k_1} \circ \cdots \circ \sigma_{b_n}^{k_n}(c)$。对每个 $n \geq 0$ 和字母 $c \neq b_{n+1}$，有

$$x_{n+1}^{(c)} = \sigma_{b_0}^{k_0} \circ \sigma_{b_1}^{k_1} \circ \cdots \circ \sigma_{b_n}^{k_n} \circ \sigma_{b_{n+1}}^{k_{n+1}}(c) = x_n^{(c)}(x_n^{(b_{n+1})})^{k_{n+1}}.$$

由于 $x_{n+1}^{(c)} \in \mathcal{L}(\mathbf{x})$ 且 $\mathbf{x}$ 是一个非周期均匀因子平衡词，推论 4.10 得出 $(k_n)_{n\geq 0}$ 有界。

现在我们证明充分条件。正如之前所述，我们将用到定理 5.12。由于以下各项均成立，所有必要的假设都得到验证：

(i) 命题 5.2 的第 (1) 和第 (2) 项成立，由引理 6.6 得证。(ii) 幂自由性成立，由引理 6.7 得证。(iii) 每个层级上的字母平衡性成立，由定理 6.5 得证。

因此，由于字母平衡性的常数与所考虑的层级无关，该词是一致因子平衡的。■

7. Factor-balancedness and factor complexity

本节旨在展示（均匀）因子平衡性与因子复杂度之间的某些联系。众所周知，线性递归词具有线性因子复杂度 [DHS99], ，因此探究因子平衡性方面的这一问题也显得顺理成章。值得一提的是，Berth´e 和 Cecchi Bernales 的早期研究 [BCB19]也聚焦于低复杂度的无限词，因为树状词的复杂度至多为线性。本文证明，替代因子平衡词具有线性因子复杂度（见命题 7.3），即使其不动点不是一致复现的。请回想 Pansiot 的一个结果 [Pan84]：代换词的因子复杂度可以达到二次。最后，在第 7.3 节中，我们还利用托普利茨词给出一个具有指数因子复杂度的因子平衡词示例；其构造将在第 7.2 节中回顾并加以说明。

7.1. **非本原替换词**。接下来我们需要以下结果：第一项的证明可由 [MR16, 引理 2.8] 和 [Shi19, 引理 3.8]推出；第二项可由 [Shi19, 定理 A]推出；最后一项可由 [Pan84, 定理 1，第 (i) 项]推出。

命题 7.1 ([MR16,Pan84, Shi19])。 设 τ 是一个替换。

(1) 若 $\mathcal{L}(\tau)$ 是 P-幂自由的，其中 $P \in \mathbb{N}$满足某个条件，则 τ是温和的。*(2)* 若 τ 是温和且拟本原的，则 $\mathcal{L}(\tau)$ 是一致复现的。*(3)* 若 τ 是增长且拟一致的，则 τ 的任意不动点 $\mathbf{x}$ 都具有线性增长因子复杂度。

在证明命题 7.3 之前，我们先确立 Pansiot 结果的一个变体（命题 7.1 的第（3）项），其中的假设仅基于增长字母。

引理 7.2。 设 $\mathbf{x}$ 是替换 $\tau\colon\ \mathcal{A}^* \to \mathcal{A}^*$的一个不动点。假设存在 $C > 0$，使得对于所有增长字母 a、$b \in \mathcal{C}_\tau$,

$$|\tau^n(a)| \le C\,|\tau^n(b)| \ \textit{for every } n \ge 1. \tag{26}$$

若 τ 是温和的，则 $\mathbf{x}$ 具有线性因子复杂度。

证明。 策略如下。我们构造一个新的替换 σ，使其具有不动点 $\mathbf{y}$，且对于某个替换 π, 有 $\mathbf{x} = \pi(\mathbf{y})$, 并使得 σ 是增长的。接着我们证明，假设（26）蕴含 σ 是拟一致的, 从而可以使用命题 7.1 的第（3）项；因此，$\mathbf{y}$ 具有线性增长因子复杂度。由于 $\mathbf{x} = \pi(\mathbf{y})$, 线性增长因子复杂度也由 $\mathbf{x}$ 继承。

令 $\mathcal{R}$ 为 $\mathbf{x}$ 中以增长字母 $\mathcal{C}_\tau$ 开头的回返词集合，也就是说，当且仅当 $u \in \mathcal{R}$ 在 $\mathbf{x}$ 的位置 $i \ge 0$ 出现，且满足 x_i、$x_{i+|u|} \in \mathcal{C}_\tau$，并且对所有 $i < j < i+|u|$ 都有 $x_j \notin \mathcal{C}_\tau$ 时， u。由于 τ 是温和的（即 $\mathcal{B}_\tau^* \cap \mathcal{L}(\tau)$ 是有限的），集合 $\mathcal{R}$ 是有限的；此外，$\mathbf{x}$ 可以分解为 $\mathbf{x} = w_0 w_1 w_2 w_3 \cdots$，其中对所有 $j \ge 0$ 都有 $w_j \in \mathcal{R}$。（注意，我们用到了 $x_0 \in \mathcal{C}_\tau$。）令 $\pi\colon$ $\mathcal{B} \to \mathcal{R}$ 为有限字母表 $\mathcal{B}$ 与 $\mathcal{R}$ 之间的任意双射，并将 π 扩展为替换 $\pi\colon \mathcal{B}^* \to \mathcal{A}^*$。现在，由于 τ 是温和的，对于任意 $a \in \mathcal{C}_\tau$， $\tau(a)$ 都以一个增长字母开头。因此，对于每个 $w \in \mathcal{R}$， $\tau(w)$ 都可分解为 $\mathcal{R}$ 中若干词的拼接。这诱导出一个替换 $\sigma\colon \mathcal{B}^* \to \mathcal{B}^*$，使得对所有 $b \in \mathcal{B}$，都有 $\pi(\sigma(b)) = \tau(\pi(b))$，如下方交换图所示：

$$\begin{array}{ccc} \mathcal{B}^* & \xrightarrow{\ \sigma\ } & \mathcal{B}^* \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ \mathcal{A}^* & \xrightarrow{\ \tau\ } & \mathcal{A}^* \end{array}$$

随后通过归纳论证可得

$$\pi(\sigma^n(b)) = \tau^n(\pi(b)) \ \text{ for all } n \ge 1 \text{ and } b \in \mathcal{B}, \tag{27}$$

如下面的交换图所示：

$$\begin{array}{ccccccc} \mathcal{B}^* & \xrightarrow{\sigma} & \mathcal{B}^* & \xrightarrow{\sigma} & \cdots & \xrightarrow{\sigma} & \mathcal{B}^* \\ \downarrow{\scriptstyle\pi} & & \downarrow{\scriptstyle\pi} & & \downarrow{\scriptstyle\pi} & & \downarrow{\scriptstyle\pi} \\ \mathcal{A}^* & \xrightarrow{\tau} & \mathcal{A}^* & \xrightarrow{\tau} & \cdots & \xrightarrow{\tau} & \mathcal{A}^* \end{array}$$

设 $b_0 \in \mathcal{B}$ 是与 $w_0 \in R$ 对应的字母，即 $\pi(b_0) = w_0$。那么， σ 在 b_0 上可延长（因为 τ 在 x_0 上可延长），所以 $\mathbf{y} = \sigma^\omega(b_0)$ 存在。此外，由（ 27 ）可得 $\mathbf{x} = \pi(\mathbf{y})$。

我们来证明 σ 是增长的且拟一致。对于替换 η，回顾一下， $|\eta|$ 表示最长的像 $|\eta(a)|$，其中 a 是一个字母。再令 ℓ 表示 $\mathcal{R}$中最长词的长度。那么，对于 $b \in \mathcal{B}$ 和 $n \geq 1$，利用（27）可得上界：

$$|\sigma^n(b)| \leq |\pi(\sigma^n(b))| = |\tau^n(\pi(b))| \leq \ell|\tau^n|.$$

我们也可以如下证明一个下界。令 c为 $\pi(b)$的首字母，它按 $\pi(b) \in \mathcal{R}$的方式增长。于是再次利用（27），

$$|\sigma^n(b)| \geq \frac{1}{|\pi|}|\pi(\sigma^n(b))| = \frac{1}{|\pi|}|\tau^n(\pi(b))| \geq \frac{1}{|\pi|}|\tau^n(c)|.$$

根据假设（26），有 $|\tau^n(c)| \geq \frac{1}{C}|\tau^n|$，进而得到

$$\frac{1}{C|\pi|}|\tau^n| \leq |\sigma^n(b)| \leq \ell|\tau^n|$$

对任意 $b \in \mathcal{B}$均成立。这证明了 σ 是拟一致的。注意，由于 σ 含有一个增长字母（即 b_0），这意味着 σ 是增长的。因此我们可以使用命题 7.1 的第（3）项，得到$\mathbf{y}$以及$\mathbf{x} = \pi(\mathbf{y})$具有线性增长因子复杂度。■

下面的结果描述了因子平衡的代换词的结构。特别是，它表明任何此类词都具有线性增长复杂度。我们将使用以下定义。设$\mathbf{x} = (x_n)_{n\geq 0}$ 是替换τ的一个不动点：由一个可延长的字母生成 $\mathcal{A}^* \to \mathcal{A}^*$ 。我们将$\mathbf{x}$的重复字母集合 $R \subseteq \mathcal{A}$ 定义为在$\mathbf{x}$中出现无限多次的字母。由于 $\mathcal{A}$ 是有限的，我们可以找到 $i \geq 1$ ，使得对所有 $j \geq i$，都有 $x_j \in R$ 。因此，对于在位置$j \geq i$出现的任意 $a \in R$ ， $\tau(x_j)$会在位置 $k \geq j \geq i$出现，这意味着 $\tau(x_j)$只包含 R中的字母。这表明 $\tau(R) \subseteq R^*$。此外，记$\mathbf{x}$ $u =$ （其中 i 取上述值），并且对某个非空词 v，有 $\tau(u) = uv$ 。接下来我们定义以下替换

$$\sigma\colon \begin{cases} a & \mapsto \tau(a), \text{ for } a \in R \\ * & \mapsto *v \end{cases} \quad \text{and} \quad \pi\colon \begin{cases} a & \mapsto a, \text{ for } a \in R \\ * & \mapsto u. \end{cases} \tag{28}$$

命题 7.3. 设 $\mathbf{x}$ 是由一个可延长字母生成的替换 $\tau\colon \mathcal{A}^* \to \mathcal{A}^*$ 的不动点。考虑上文定义的集合 R 以及 (28)中定义的替换 σ, π 。若$\mathbf{x}$是因子平衡的，则限制 $\tau|_R\colon R^* \to R^*$ 满足： $\mathcal{L}(\tau|_R)$一致复现且因子平衡。此外，我们有$\mathbf{x}$ $= \pi(\sigma^\omega(*))$。特别地，$\mathbf{x}$ 具有线性增长因子复杂度。

证明。不妨设$\mathbf{x}$ 不是最终周期的，否则结论显然成立。此外，不失一般性，假设 $\mathcal{A}$中的每个字母都在某个 $\tau^n(x_0)$ 中出现，其中 $n \geq 1$ 可能依字母而异。这意味着$\mathcal{L}(\mathbf{x}) \supseteq \mathcal{L}(\tau)$。

令 $\tau|_R\colon R^* \to R^*$ 表示 τ到 R^*的限制。证明关于 $\tau|_R$ 的性质时，策略是使用命题 7.1。按照命题 4.6 中所用的相同论证，对于 R中的每个字母 a ，我们可以利用 a 是平衡且复现的这一点，

$\mathbf{x}$ 由此推得它在 $\mathbf{x}$中以有界间隔出现。因此，对于每个增长字母 $a \in \mathcal{C}_\tau \cap R$，每个足够大的幂 $\tau^n(a)$都包含 R中的每个字母。由此可知，限制 $\tau|_R$ 是拟本原的。接下来证明它也是温和的。根据我们在第一段中的证明，有 $\mathcal{L}(\mathbf{x}) \supseteq \mathcal{L}(\tau) \supseteq \mathcal{L}(\tau|_R)$。由于推论 4.10 表明，对于某个 $P \in \mathbb{N}$，$\mathbf{x}$是 P-幂自由的，我们可以推得 $\mathcal{L}(\tau|_R)$是 P-幂自由的。因此，根据命题 7.1 的第 (1) 项，$\tau|_R$是温和的。由于 $\tau|_R$温和且拟本原，根据命题 7.1 的第 (2) 项，$\mathcal{L}(\tau|_R)$是一致复现的。此外，由于 $\mathcal{L}(\mathbf{x}) \supseteq \mathcal{L}(\tau|_R)$，所以 $\mathcal{L}(\tau|_R)$是因子平衡的。

接下来，我们构造 σ 和 π。回顾一下，$i \geq 1$ 的定义使得对所有 $j \geq i$都有 $x_j \in R$。按照陈述，令 $u = \mathbf{x}^{[0,i)}$。由于$\mathbf{x}$ 是 τ的不动点，u 是 $\tau(u)$的前缀，因此存在 v，使得 $\tau(u) = uv$ 且 $v \in R^* \setminus \{\varepsilon\}$。令 σ 和 π 如（28）中所述，并注意到对于每个 $a \in R \cup \{*\}$，都有 $\pi(\sigma(a)) = \tau(\pi(a))$。通过归纳论证，我们得到

$$\pi(\sigma^n(a)) = \tau^n(\pi(a)) \text{ for all } n \geq 1 \text{ and } a \in R \cup \{*\}. \tag{29}$$

显然，σ 在 $*$上是可延长的，因此$\mathbf{y} = \sigma^\omega(*)$是定义良好的无限词。由于$\pi(*) = u = \mathbf{x}^{[0,i)}$，$\pi(*)$以 x_0开头，结合（29）可将其写为

$$\mathbf{x} = \tau^\omega(\pi(*)) = \pi(\sigma^\omega(*)) = \pi(\mathbf{y}).$$

还需证明 $\mathbf{x}$ 具有线性增长复杂度。根据上一个恒等式，只需证明 $\mathbf{y}$ 具有线性增长复杂度。为此，我们检验引理 7.2 的假设。先证明 σ 是温和的。由推论 4.10 可知，对于某个 $P > 0$，$\mathbf{x}$ 是P-幂自由的。由于 $\mathbf{x} = \pi(\mathbf{y})$，$\mathbf{y}$ 也具有同样的性质。又由于 $\mathcal{L}(\mathbf{x}) \supseteq \mathcal{L}(\tau)$，根据 (29)，有 $\mathcal{L}(\mathbf{y}) \supseteq \mathcal{L}(\sigma)$。由此可得，$\mathcal{L}(\sigma)$ 是 P-幂自由的，因此根据命题 7.1 的第 (1) 项，σ 是温和的。接下来证明 σ 是拟一致的。根据佩龙-弗罗贝尼乌斯理论中的经典结果，只需找到一个增长字母 $\bar{a} \in \mathcal{C}_\sigma$，使得对于任意 $b \in \mathcal{C}_\sigma$，存在 $\sigma^n(b)$ 的某个幂，其中出现 $\bar{a}$；例如参见 CLR16，引理 11]。固定 $\bar{a} \in \mathcal{C}_\sigma \cap R$，并任取一个 $b \in \mathcal{C}_\sigma$。由于 $\tau|_R$ 是拟本原的，且 $\sigma|_R = \tau|_R$，因此 $\sigma|_R$ 也是拟本原的。于是，对于每个 $b \in R \cap \mathcal{C}_\sigma$，存在 $\sigma^n(b)$ 的某个幂，其中出现 $\bar{a}$。此外，$\mathbf{y} = \sigma^\omega(*)$，因此存在 $\sigma^n(*)$ 的某个幂，其中出现 $\bar{a}$。这就证明了所述结论；再应用引理 7.2，即完成证明。■

下一个例子给出一个因子平衡、由代换生成的词，但它不是复现的。因此，在命题 7.3 中，辅助符号 $*$ 对表示可能存在的瞬态部分是必不可少的。

例 7.4。令 $\tau\colon \{0,1\}^* \to \{0,1\}^*$ 为斐波那契代换，即 $\tau(0) = 01$ 且 $\tau(1) = 0$。定义 $\sigma\colon \{0,1,*\}^* \to \{0,1,*\}^*$，使得对 $a \in \{0,1\}$ 有 $\sigma(a) = \tau(a)$，并且 $\sigma(*) = *11$。注意，11 不 属于 $\mathcal{L}(\tau)$。在 B节中，我们证明不动点

$$\mathbf{x} = \sigma^\omega(*) = *11\tau(11)\tau^2(11)\cdots = *11000101010010010010100\cdots$$

具有因子平衡性，且其任何移位都不复现。为此，我们首先证明$\mathbf{x}$ 是字母平衡的。然后，我们证明$\mathbf{x}$ 是斐波那契自动的，也就是说，它可以由确定性有限自动机输出生成，其输入采用 Zeckendorf 计数系统表示。最后，利用这一性质，我们使用自动定理证明器 Walnut [Sha23] 证明$\mathbf{x}$ 是 5 次幂自由的，且$\mathbf{x}$ 的任何移位都不复现。因此，

根据定理 5.12，$\mathbf{x}$ 是一致因子平衡的。此外，根据命题 7.3，它具有线性复杂度。

作为这个例子的最后一点，注意，$\mathbf{x}' = 11000\cdots$ 的第一次移位 $\mathbf{x}$ 的因子复杂度为 $2n$，也就是说，它属于*Rote* 词这一类。此外，这个词似乎是 2 字母平衡的，这是这类词可能达到的最小常数。与词 $h(\psi^\omega(0))$ 一起，其中 $\psi\colon 0 \mapsto 0121, 1 \mapsto 0, 2 \mapsto 01$ 以及 $h\colon 0 \mapsto 01, 1 \mapsto 0, 2 \mapsto 011$，它们是目前仅知的具有这一性质的 Rote 词[2]。

7.2.**托普利茨词**。 托普利茨词 是一个一致复现的无限词 $\mathbf{x} = (x_n)_{n\geq 0} \in \mathcal{A}^{\mathbb{N}}$，满足对每个 $n \in \mathbb{N}$，都存在一个周期 $p \geq 1$，使得对每个 $k \in \mathbb{N}$，$x_{n+kp} = x_n$。我们介绍一种按照 [DP22],构造托普利茨词的方法，该方法基于 [GJ00]。我们还提醒读者，这一定义比例如 [CK97]中给出的“托普利茨词”定义更为一般。最后，关于更多细节，请参阅 Downarowicz 的综述 [Dow05]。

固定一个字母表 $\mathcal{A}$。该构造取决于两个参数。首先，我们需要一个递增的正整数序列 $(p_n)_{n\geq 0}$，使得

($\mathbf{Tp_1}$) 对所有 $n \geq 0$，整数 p_n 都整除 p_{n+1}。

对于 $n \geq 1$，定义整数 $q_n = p_n/p_{n-1}$。第二个参数是一个词集合序列 $(W_n)_{n\geq 0}$，其中 $W_n \subseteq \mathcal{A}^{p_n}$，并满足以下条件：对每个 $n \geq 0$，

($\mathbf{Tp_2}$) 每个 $w \in W_{n+1}$ 都是 W_n 中 q_n 个词的拼接 $w = \phi_n(w, 1)\phi_n(w, 2)\cdots\phi_n(w, q_n)$；

并且 **($\mathbf{Tp_3}$)** 存在一个 $w_n \in W_n$，它是 W_{n+1} 中每个词的前缀。

注意，最后一个性质保证 w_n 始终是 w_{n+1} 的前缀。由于 $|w_n| = p_n$ 递增，我们得到一个无限词

$$\mathbf{x} = \lim_{n\to\infty} w_n \in \mathcal{A}^{\mathbb{N}}.$$

此外，若我们假设这些参数对每个 $n \geq 1$ 都满足，

($\mathbf{Tp_4}$) 集合 $\{\phi_n(w,\ i)\colon\ 1 \leq i \leq q_n\}$ 对每个 $w \in W_{n+1}$ 都等于 W_n，

那么$\mathbf{x}$ 是一致复现的，并且$\mathbf{x}$是一个托普利茨词。

例 7.5。 我们通过证明倍周期替换的不动点是一个托普利茨词，来举例说明这些定义。令 $\mathcal{A} = \{0, 1\}$ 和 $p_n = 4^n$ 对于 $n \geq 0$，它们满足 **($\mathbf{Tp_1}$)**。我们归纳地定义词的集合 $W_n = \{u_0(n)\ u_1(n)\} \subseteq \mathcal{A}^{p_n}$ 如下。首先，$u_0(0) = 0$ 且 $u_1(0) = 1$，然后令

$$\begin{cases} u_{n+1}(0) &= u_n(0)\, u_n(1)\, u_n(0)\, u_n(0), \\ u_{n+1}(1) &= u_n(0)\, u_n(1)\, u_n(0)\, u_n(1), \end{cases} \tag{30}$$

对于 $n \geq 0$。条件 **($\mathbf{Tp_2}$)** 和 **($\mathbf{Tp_4}$)** 由构造本身满足，而 **($\mathbf{Tp_3}$)** 也成立，其中 $w_n = u_n(0)$。因此，词序列 $(u_n(0))_{n\geq 0}$ 收敛到一个一致复现词 $\mathbf{x} \in \mathcal{A}^{\mathbb{N}}$，它是一个托普利茨词。

可以验证，$\mathbf{x}$恰好是倍周期替换τ的 不动点：$\mathcal{A}^* \to \mathcal{A}^*$，其定义为 $\tau(0) = 01$和 $\tau(1) = 00$。两种构造方法等价，是因为式（30）中的递推关系模仿了 τ^2，其定义为

[2]这是根据 L. Dvoˇr´akov´a、E. Pelantov´a 与 J. Shallit 的个人通信得出的。

$\tau^2(0) = 0100$和 $\tau^2(1) = 0101$。注意，在式（30）中需要使用 τ^2而不是 τ，以确保该构造满足（$\mathbf{Tp_4}$）。

请注意，(,) 中的词 $\phi_n(w\,i)$ 本质上定义了替换。现在我们将这一想法形式化，因为使用替换语言来描述 Toeplitz构造 将会非常方便。

固定一个字母表序列 $(\mathcal{A}_n)_{n\geq 0}$，使得 $\mathcal{A}_0 = \mathcal{A}$，并且对于每个 $n \geq 0$，$\mathcal{A}_n$ 与 W_n 存在双射。于是，这些双射将 $(\mathbf{Tp_2})$ 中的分解 $w = \phi_n(w,1)\phi_n(w,2)\cdots\phi_n(w,q_n)$ 转化为映射 $\tau_n\colon \mathcal{A}_{n+1} \to \mathcal{A}_n^{q_n}$。该映射 τ_n 唯一延拓为替换 $\tau_n\colon \mathcal{A}_{n+1}^* \to \mathcal{A}_n^*$，使得 $(\mathbf{Tp_1'})$ 我们有 $|\tau_n(a)| = \frac{p_n}{p_{n-1}}$，对于每个 $n \geq 1$ 和每个 $a \in \mathcal{A}_{n+1}$ 均成立。

条件（$\mathbf{Tp_3}$）与（$\mathbf{Tp_4}$）于是等价于：对每一个 $n \geq 0$,

（$\mathbf{Tp_3'}$）存在 $a_n \in \mathcal{A}_n$，使得对每一个 $a \in \mathcal{A}_{n+1}$，$\tau_n(a)$ 都以 a_n 开头。

（$\mathbf{Tp_4'}$）对每一个 $a \in \mathcal{A}_{n+1}$，每一个 $b \in \mathcal{A}_n$ 都出现在 $\tau_n(a)$ 中。

条件（$\mathbf{Tp_3'}$）通常称为左恰当性，而（$\mathbf{Tp_4'}$）强本原性 .

因此，托普利茨词 $\mathbf{x}$ 完全由一个满足（$\mathbf{Tp_1'}$）、（$\mathbf{Tp_3'}$）和（$\mathbf{Tp_4'}$）的替换序列 $(\tau_n\colon \mathcal{A}_{n+1} \to \mathcal{A}_n^{q_n})_{n\geq 0}$ 通过如下公式描述

$$\mathbf{x} = \lim_{n\to\infty} \tau_0 \circ \tau_1 \circ \cdots \circ \tau_n(a_{n+1}).$$

注意，对每一个 $m \in \mathbb{N}$，移位序列 $(\tau_n)_{n\geq m}$ 同样满足托普利茨条件，因此它定义了一个托普利茨词

$$\mathbf{x^{(m)}} = \lim_{n\to\infty} \tau_m \circ \tau_{m+1} \circ \cdots \circ \tau_n(a_{n+1}) \text{ for each } m \geq 0. \tag{31}$$

中间词 $(\mathbf{x^{(m)}})_{m\geq 0}$ 在理解$\mathbf{x}$方面起着重要作用。它们满足以下关系

$$\mathbf{x^{(m)}} = \tau_m \circ \tau_{m+1} \circ \cdots \circ \tau_{n-1}(\mathbf{x^{(n)}}), \text{ for all } n > m \geq 0. \tag{32}$$

这尤其表明，$\mathbf{x^{(m)}}$ 是词 $\tau_m \circ \tau_{m+1} \circ \cdots \circ \tau_{n-1}(a)$ 的无限拼接，其中 $a \in \mathcal{A}_n$。

注 7.6. 我们郑重提醒读者，在 Toeplitz 框架下，记号$\mathbf{x^{(m)}}$ 并非指本文前几节中诱导元组的记号。由于我们不会同时使用这两种对象，因此结合上下文应当可以清楚理解记号 $\mathbf{x^{(m)}}$ 的含义。

7.3. 指数因子复杂度。 我们的目标是证明定理 1.5。因此，我们构造一个因子平衡的托普利茨词$\mathbf{x}$，其复杂度函数 $p_{\mathbf{x}}$ 满足

$$\lim_{n\to\infty} \frac{1}{n} \log p_{\mathbf{x}}(n) > 0,$$

也就是说，$\mathbf{x}$ 具有正熵。

设 $\mathcal{A}_0$ 为一个字母表。考虑两个正整数序列 $(k_n)_{n\geq 0}$ 和 $(r_n)_{n\geq 0}$，它们满足 $r_n \leq 2k_n$，且将在之后固定。我们归纳构造 $\tau_n\colon \mathcal{A}_{n+1}^* \to \mathcal{A}_n^*$ 如下。

首先，我们只需令 $\mathcal{A}_1 = \mathcal{A}_0$ 和 $\tau_0\colon \mathcal{A}_1 \to \mathcal{A}_0$ 为恒等映射，即对所有 $a \in \mathcal{A}_1$，$\tau_0(a) = a$，并通过拼接将其扩展为一个替换。假设 τ_{n-1} 已定义。固定一个满足 $|u_n| = |v_n| = r_n$ 的本原词 $v_nu_n \in \mathcal{A}_n^*$。

令 W_n 为所有满足以下条件的词 $w \in \mathcal{A}_n^*$ 的集合：

(1) 词 w 以 u_n 开头，以 v_n结尾；(2) 每个 $a \in \mathcal{A}_n$ 在 w中恰好出现 k_n 次，即 $|w|_a = k_n$；

注意，第二个条件意味着对所有 $w \in W_n$，都有 $|w| = k_n|\mathcal{A}_n|$。令 $\tau_n\colon \mathcal{A}_{n+1} \to W_n$ 为字母表 $\mathcal{A}_{n+1}$ 与 W_n 之间的任意双射，然后将 τ_n 扩展为替换 $\tau_n\colon \mathcal{A}_{n+1}^* \to \mathcal{A}_n^*$。

例 7.7。让我们通过一个简单的例子来说明第一步如何进行，其中$r_1 = 2$、$k_1 = 4$、$\mathcal{A}_1 = \{0,1\}$、$u_1 = 01$和 $v_1 = 10$。那么，如果 $w = 01w'10$，$|w'|_0 = |w'|_1 = 2$，则 $w \in W_1$。因此，w'有六种选择：0011，0101，0110，1001，1010，1100。最后，我们定义 $\mathcal{A}_2 = \{0,1,\dots,5\}$ 和 τ_1：$0 \mapsto 0011$，$1 \mapsto 0101$，..... $5 \mapsto 1100$。

不难看出，上文定义的替换序列 $(\tau_n)_{n\geq 0}$ 满足托普利茨条件，因此它们按 (31) 定义了 Toeplitz 词$\mathbf{x}^{(\mathbf{m})}$。我们令 $\mathbf{x} =\mathbf{x}^{(\mathbf{0})}$。

定理 7.8. 词 $\mathbf{x}$是因子平衡的。此外，存在序列 (r_n, k_n) 的一种选取，使得词 $\mathbf{x}$ 具有指数增长因子复杂度。

注意，定理 1.5 直接由前述命题推出。我们借助一系列中间结果证明定理 7.8。引理 7.11 证明该定理的第一个命题，引理 7.14 证明第二个命题。

引理 7.9. 对每个 $m \geq 0$，$\mathbf{x}^{(\mathbf{m})}$ 是 $(4k_m)$-字母平衡的。

证明. 固定 $m \geq 0$，设 $u \in \mathcal{L}(\mathbf{x}^{(\mathbf{m})})$ 与 $a \in \mathcal{A}_m$。只需证明

$$\left| |u|_a - k_m \frac{|u|}{|\mathcal{A}_m|} \right| \leq 2k_m, \tag{33}$$

因为这样一来，对任何其他长度为 $|v| = |u|$的 $v \in \mathcal{L}(\mathbf{x}^{(\mathbf{m})})$，前述不等式都给出

$$\big||u|_a - |v|_a\big| \leq \left| |u|_a - k_m \frac{|u|}{|\mathcal{A}_m|} \right| + \left| k_m \frac{|v|}{|\mathcal{A}_m|} - |v|_a \right| \leq 4k_m.$$

让我们证明 (33)。注意，根据构造，$\mathbf{x}^{(\mathbf{m})}$ 是由形如 $\tau_m(b)$ 的词无限拼接而成的，其中 $b \in \mathcal{A}_{m+1}$。特别地，由于 u 出现在 $\mathbf{x}^{(\mathbf{m})}$中，它必定出现在某个词 $b_1 \cdots b_\ell \in \mathcal{A}_{m+1}^*$的 $\tau_m(b_1 \cdots b_\ell)$ 中，其中 $\ell \geq 1$。我们取长度最小的这样一个词 ℓ。根据构造，对于任意 $b \in \mathcal{A}_{m+1}$，$\tau_m(b)$ 满足

$$\begin{cases} |\tau_m(b)| = k_m|\mathcal{A}_m|, \\ |\tau_m(b)|_a = k_m. \end{cases}$$

因此，由于 u出现在 $\tau_m(b_1 \cdots b_\ell)$ 中，我们得到以下上界

$$\begin{cases} |u| \leq \ell k_m|\mathcal{A}_m|, \\ |u|_a \leq \ell k_m. \end{cases}$$

现在，根据 ℓ的极小性，$\ell \leq 2$ 或 $\tau_m(b_2 b_3 \cdots b_{\ell-1})$ 必有一个出现在 u 中。在这两种情况下，我们都得到下界

$$\begin{cases} |u| \geq (\ell-2)k_m|\mathcal{A}_m|, \\ |u|_a \geq (\ell-2)k_m. \end{cases}$$

由前面的不等式可推出

$$\begin{cases} (\ell-2)k_m \le |u|_a \le \ell k_m, \\ \frac{|u|}{k_m|\mathcal{A}_m|} \le \ell \le \frac{|u|}{k_m|\mathcal{A}_m|} + 2. \end{cases}$$

因此，将第二组不等式代入第一组不等式，我们得到

$$\frac{|u|}{|\mathcal{A}_m|} - 2k_m \le |u|_a \le \frac{|u|}{|\mathcal{A}_m|} + 2k_m.$$

这证明了（33），从而完成了引理的证明。 ■

为证明定理 7.8 的第一部分，我们现在使用 Poirier 和 Steiner [PS24] 的以下结果。在他们的论文中，该结果是用符号动力系统的术语表述的；我们给出一个针对无限词的等价版本。

命题 7.10 ([PS24，定理 4。1])。 设 $(\tau_n\colon\ \mathcal{A}_{n+1}^* \to \mathcal{A}_n^*)_{n\ge 0}$ 是一列左适当且本原的替换。设 $\mathbf{x}^{(\mathbf{m})} \in \mathcal{A}_m^{\mathbb{N}}$ 是移位序列 $(\tau_n)_{n\ge m}$ 的极限词，其中对每个 $m \ge 0$ 都成立。若每个 $\mathbf{x}^{(\mathbf{m})}$ 都是字母平衡的，则 $\mathbf{x}^{(\mathbf{0})}$ 是因子平衡的。

引理 7.11。 词$\mathbf{x}$ 是因子平衡的。

证明。 根据（$\mathbf{Tp'_3}$），这些替换 τ_n 是左适当的；根据（$\mathbf{Tp'_4}$），它们是本原的。这两个条件与引理 7.9 一起，恰好构成命题 7.10 的假设；该命题的结论是$\mathbf{x}$ 是因子平衡的。■

现在我们转而估计$\mathbf{x}$的因子复杂度。第一步是证明一个可识别性结果。为此，我们需要构造中得到的词 $v_n u_n$ 。

引理 7.12。 对于每个 $m > n \ge 0$，映射$\tau_{n,m}\colon\ \mathcal{A}_m \to \mathcal{A}_n$ 是单射。

证明。固定某个 $n \ge 0$，我们用归纳法对 m证明该结论。当 $m = n+1$时，有τ_n，$_m\ =\ \tau_n$，因此基础情形直接由 τ_n的构造得出；构造 τ_n时，要求 τ_n 将 $\mathcal{A}_n$ 双射到 W_n。假设 $mn+1$ 成立，并取$a, a' \in \mathcal{A}_m$ ，使得$\tau_{n,m}(a) = \tau_{n,m}(a')$；我们来证明 $a = a'$。因此，我们有$\tau_{n,m-1}(\tau_{m-1}(a))= \tau_{n,m-1}(\tau_{m-1}(a'))$。根据归纳假设，$\tau_{n,m-1}$ 是单射，于是 $\tau_{m-1}(a) = \tau_{m-1}(a')$；又因为 τ_{m-1} 被选为双射，故结论成立。■

我们需要下一个估计结果。

引理 7.13。 对于 $k, d \in \mathbb{N}$，多项式系数$\binom{dk}{k,\dots,k} := (dk)!\ (k!)^d$ 满足

$$\binom{dk}{k,\dots,k} = \big(1+o(1)\big)\frac{\sqrt{d}}{(2k\pi)^{(d-1)/2}}d^{dk}.$$

证明。 这是$\binom{dk}{k,\dots,k}$ 的经典估计，使用斯特林近似可得

$$\log\left(\binom{dk}{k,\dots,k}\right) = dk\log(d) + \frac{1}{2}\log(d) - \frac{d-1}{2}\log(2\pi k) + O\left(\frac{1}{k}\right), \tag{34}$$

结论由此得出。 ■

定理 7.8 的第二部分现在即可得出。

引理 7.14。 存在一组序列 $(r_n, k_n)_{n\geq 0}$ ，使得词 $\mathbf{x}$ 具有指数增长因子复杂度。

证明。 设 $\ell_n = |\mathcal{A}_n|$、 $u_n = 01\cdots(\ell_n - 1)$ 和 $v_n = (\ell_n - 1)(\ell_n - 2)\cdots 0$。因此，对所有 $n \geq 0$，都有 $r_n = \ell_n$。根据构造，我们有 $p_{n+1} = p_n k_n \ell_n$。于是，由引理 7.12，$p_{\mathbf{x}}(p_{n+1}) \geq \ell_{n+1}$。因此，我们的目标是证明，对于所有 $n \geq 0$，以及某个常数 $c > 0$，都有 $\frac{1}{p_n}\log(\ell_n) > c$。

请记住， ℓ_{n+1} 仅取决于 k_n 的选择，而 k_{n+1} 的选择独立于 k_n。假设所有 k_i，即 $i \leq n-1$，都已选定。如果 k_n 足够大， ℓ_{n+1} 就对应于满足以下条件的词 $w \in \mathcal{A}_n^*$

(i) 词以 w 开头，以 u_n 结尾。 (ii) 对所有 $a \in \mathcal{A}_n$，我们都有 $|w|_a = k_n$。特别地，我们有 $|w| = k_n|\mathcal{A}_n|$。

由于对所有 $a \in \mathcal{A}_n$ 都有 $|u_n|_a = |v_n|_a = 1$，我们知道，对于某个 w'，有 $w = u_n w' v_n$，使得对所有 $a \in \mathcal{A}_n$ 都有 $|w'|_a = (k_n - 2)$。因此，这类词的数量恰好是多项式系数 $\binom{\ell_n}{(k_n-2),\ k_n-2,\ \ldots,\ k_n-2}$。于是，我们得到

$$\ell_{n+1} = (1 + o(1))\left(\ell_n^{\ell_n(k_n-2)}\frac{\sqrt{\ell_n}}{(2(k_n-2)\pi)^{(\ell_n-1)/2}}\right).$$

于是，当 k_n 足够大时，我们有

$$\ell_{n+1} > \ell_n^{\ell_n k_n}\frac{1}{(2\pi\ell_n^2 k_n)^{\ell_n}}. \tag{35}$$

利用（35）中的界，我们有

$$\begin{aligned}
\frac{1}{p_{n+1}}\log(\ell_{n+1}) &\geq \frac{1}{p_n k_n \ell_n}\log\left(\ell_n^{\ell_n k_n}\frac{1}{(2\pi\ell_n^2 k_n)^{\ell_n}}\right),\\
&\geq \frac{\log(\ell_n)}{p_n} - \frac{1}{p_n k_n}\log(2\pi\ell_n^2 k_n),\\
&\geq \frac{\log(\ell_n)}{p_n} - \frac{2}{p_n k_n}\log(\ell_n) - \frac{1}{p_n k_n}\log(2\pi k_n),\\
&\geq \frac{\log(\ell_n)}{p_n}\left(1 - \frac{2}{k_n}\right) - \frac{3}{p_n}.
\end{aligned}$$

因此，通过迭代前面的不等式，我们对所有 $n \geq 2$ 得到

$$\frac{1}{p_n}\log(\ell_n) \geq \frac{1}{p_2}\log(\ell_2)\prod_{2\leq i\leq n-1}\left(1 - \frac{2}{k_i}\right) - S_n,$$

其中

$$S_n := \sum_{2\leq i\leq n-1}\frac{3}{p_i}\prod_{i+1\leq j\leq n-1}\left(1 - \frac{2}{k_j}\right).$$

由于 $p_i = p_{i-1}\ell_{i-1}k_{i-1} \geq 2p_{i-1}$，对于所有 $i \geq 2$，我们有 $p_i \geq 2^i$。又因为 $S_n \leq \sum_{2\leq i\leq n-1}\frac{3}{p_i}$，所以总和 $S = \lim_{n\to+\infty} S_n$ 是有限的。现在，我们选择 $(k_n)_{n\geq 2}$，使得

$\prod_{2\le i\le n}\left(1-\frac{2}{k_i}\right)$ 是有限且非零的，例如，对于所有 $n\ge 2$，有 $k_n\ge n^2$。于是我们有

$$\frac{1}{p_n}\log(\ell_n)\ge\frac{1}{p_2}\log(\ell_2)\prod_{i\ge 2}\left(1-\frac{2}{k_i}\right)-S,$$

因为 $\prod_{2\le i\le n-1}\left(1-\frac{2}{k_i}\right)\ge\prod_{i\ge 2}\left(1-\frac{2}{k_i}\right)$ 和 $S_n\le S$。最后，我们按如下方式选取 $|\mathcal{A}|=\ell_1=\ell_0$和 k_1。回想一下，$p_1=1$、$p_2=p_1\ell_1k_1=|\mathcal{A}|k_1$，并由 (34) 得到

$$\frac{1}{p_2}\log\ell_2=\frac{1}{|\mathcal{A}|k_1}\Bigg(|\mathcal{A}|(k_1-2)\log(|\mathcal{A}|)+\frac{1}{2}\log(|\mathcal{A}|)$$
$$-\frac{|\mathcal{A}|-1}{2}\log(2\pi(k_1-2))+O\left(\frac{1}{k_1}\right)\Bigg)$$

因此，首先选择 $\mathcal{A}$，使得 $\log|\mathcal{A}|\prod_{i\ge 2}\left(1-\frac{2}{k_i}\right)>S$。然后，由于 $k_1\to\infty$，我们得到

$$\frac{1}{p_2}\log\ell_2=\log|\mathcal{A}|\prod_{i\ge 2}\left(1-\frac{2}{k_i}\right)+O\left(\frac{1}{k_1}\right).$$

最后，我们可以选择 k_1，使得

$$\frac{1}{p_n}\log p_{\mathbf{x}}(p_n)\ge\frac{1}{p_n}\log|\mathcal{A}_n|\ge\frac{1}{p_2}\log(\ell_2)\prod_{i\ge 2}\left(1-\frac{2}{k_i}\right)-S>0.$$

因此，序列 $\left(\frac{1}{p_n}\log p_{\mathbf{x}}(p_n)\right)_{n\ge 1}$ 始终与零保持正距离，这意味着 $\mathbf{x}$ 具有指数增长因子复杂度，正如所需。■

References

[ABB+15] Shigeki Akiyama、Marcy Barge、Val´erie Berth´e、Jeong-Yup Lee 和 Anne Siegel。关于 Pisot 替换猜想。载于 *Mathematics of aperiodic order* 309 *Progr.*，第 卷，页码 –。 *Math.* 33 72 ¨ 2015

Birkhauser/Springer，巴塞尔，。[ABM+25] Pierre Arnoux、Val´erie Berth´e、Milton Minervino、Wolfgang Steiner 和 J¨org Thuswald-ner。非平稳马尔可夫分割与多维连分数算法，2025年。预印本见 https://arxiv.org/abs/2508.16441。[AD19] Artur Avila 和 Vincent Delecroix。Pisot 矩阵的一些幺半群。载于 *New trends in one dimensional dynamics* 285 *Springer Proc. Math. Stat.* 21 30 ，第 卷，页码 –。施普林格，沙姆，2019年。[Ada03] Boris Adamczewski。本原替换不动点的平衡性。

Theoret. Comput.，( ): –。[Ada04] Boris Adamczewski。符号偏差与自相似动力系统。 *Sci.* 307 1 47 75, 2003

Ann. Inst. Fourier，( ): –。[AGH00] Eitan Altman、Bruno Gaujal 和 Arie Hordijk。平衡序列与最优路由。*J. ACM*，47(4):752–775，2000年。[AI01] Pierre Arnoux 和 Shunji Ito。Pisot 替换与 Rauzy 分形。 *(Grenoble)* 54 7 2201 2234, 2004

Bull. Belg. Math.，( ): –。Journees ´ ´ Montoises d'Informatique Theorique（Marne-la-Val´ee，2000年）。[AR91] Pierre Arnoux 和 G´erard Rauzy。序列的几何表示，复杂度为 $2n+1$。*Bulletin de la Société Mathématique de France*，119(2):199–215，1991年。[AS92] Jean-Paul Allouche 和 Jeffrey Shallit。 *Sci.* 98 2 163 197, 1992

k *Theor. Comput.*-正则序列环。，( ): –。

[AS03] Jean-Paul Allouche 和 Jeffrey Shallit。 *Automatic sequences. Theory、 applications、 generalizations* 2003 。剑桥：剑桥大学出版社，。

[BCB19] Val´erie Berth´e 和 Paulina Cecchi Bernales。符号系统中的平衡性与余边界。 *Theoretical Computer Science*，777:93–110，2019年。[BCBD+21] Val´erie Berth´e、Paulina Cecchi Bernales、Fabien Durand、Julien Leroy、Dominique Perrin 和 Samuel Petite。幺模 S-adic 子移位的维数群。 *Monatsh.* ，( ): –。[BCBE] Val´erie Berth´e、Paulina Cecchi-Bernales 和 Basti´an Espinoza。来自 S-adic 表示的特征值。准备中，2026年。[BCS13] Val´erie Berth´e、Julien Cassaigne 和 Wolfgang Steiner。Arnoux-Rauzy 词的平衡性质。 *Internat. J. Algebra Comput.*，23(4):689–703，2013年。[BDFD+15a] Val´erie Berth´e、Clelia De Felice、Francesco Dolce、Julien Leroy、Dominique Perrin、Christophe Reutenauer 和 Giuseppina Rindone。无环、连通树集。*Monatsh. Math.*，176(4):521–550，2015年。[BDFD+15b] Val´erie Berth´e、Clelia De Felice、Francesco Dolce、Julien Leroy、Dominique Perrin、Christophe Reutenauer 和 Giuseppina Rindone。有限指标基性质。 *J. Pure* ，( ): –。[BDFD+15c] Val´erie Berth´e、Clelia De Felice、Francesco Dolce、Julien Leroy、Dominique Perrin、Christophe Reutenauer 和 Giuseppina Rindone。极大双因子解码。 *Discrete Math.*，338 5 725 742，2015 ( ): –。[BG13] Michael Baake 和 Uwe Grimm。 *Aperiodic order. Vol. 1*，第 149 卷。剑桥大学出版社，剑桥，。[BHZ06] Val´erie Berth´e、Charles Holton 和 Luca Q. Zamboni。Sturmian 序列的初始幂。 *Acta Arithmetica*，122(4):315–347，2006年。[BKV25] Val´erie Berth´e、Toghrul Karimov 和 Mihir Vahanwala。 *s*-adic 词上的自动机，2025年。预印本见 https://arxiv.org/abs/2506.17460。[BST19] Val´erie Berth´e、Wolfgang Steiner 和 J¨org M. Thuswaldner。 *S*-adic 移位的几何、动力学与算术。 *Ann. Inst. Fourier (Grenoble)*，69(3):1347–1409，2019年。[BST22] Val´erie Berth´e、Wolfgang Steiner 和 J¨org Thuswaldner。多维连分数与环面平移的符号编码。 *Journal of the European Mathematical Society* 25 12 4997 5057, 2022 ，( ): –。[CDO+25] Jean-Michel Couvreur、Martin Delacourt、Nicolas Ollinger、Pierre Popoli、Jeffrey Shallit 和 Manon Stipulanti。Pisot 型替代序列的广义阿贝尔复杂度的有效计算，2025年。预印本见 https://arxiv.org/abs/2504.13584。[CFZ00] Julien Cassaigne、S´ebastien Ferenczi 和 Luca Q. Zamboni。Arnoux-Rauzy 序列中的不平衡性。 *Ann. Inst. Fourier (Grenoble)*，50(4):1265–1276，2000年。[Cha69] Rafael V. Chacon。非强混合的弱混合变换。 *Proceedings of the American Mathematical Society* 22 3 559 562, 1969 ，( ): –。[CK97] Julien Cassaigne 和 Juhani Karhum¨aki。Toeplitz 词、广义周期性与周期迭代态射。 *European J. Combin.*，18(5):497–510，1997年。[CLR16] ´Emilie Charlier、Julien Leroy 和 Michel Rigo。自由幺半群态射的渐近性质。 *Linear Algebra and its Applications*，500:119–148，2016年。[DHS99] Fabien Durand、Bernard Host 和 Christian Skau。替代动力系统、Bratteli 图与维数群。 *Ergodic Theory Dyn. Syst.*，19(4):953–993，1999年。[Dow05] Tomasz Downarowicz。里程表系统与 Toeplitz 流综述。载于 *Algebraic and topological dynamics* 385 *Contemp. Math.* 7 37 ，第 卷，页码 –。美国数学会，普罗维登斯，RI，2005年。[DP22] Fabien Durand 和 Dominique Perrin。 *Dimension Groups and Dynamical Systems: Substitutions Bratteli Diagrams and Cantor Systems* 196 *Cambridge Studies in Advanced Mathematics* ，，第 卷。剑桥大学出版社，。[Dur98] Fabien Durand。利用回返词刻画代替序列。 *Discrete Mathematics* 179 1 3 89 101, 1998 ，(–): –。

[Dur00] Fabien Durand。线性复现子移位只有有限个非周期子移位因子。
Ergodic Theory and Dynamical Systems，20(4):1061–1078，2000。[Fer95] S´ebastien Ferenczi。夏孔变换：组合学、几何结构及其与复杂度的联系´eom´etrique，lien avec les syst`emes de complexit´e $2n+1$。
Bulletin de la Société Mathématique de France，123 2 271 292，1995 ( )：–。[Fog02] N. Pytheas Fogg。
Substitutions in dynamics，*arithmetics and combinatorics*，第1794卷，*Lecture Notes in Mathematics*。施普林格出版社，柏林，2002。[Fra73] Aviezri S. Fraenkel。互补与精确覆盖序列。
J. Combinatorial ，：–。[FV02]
Theory Ser. A 14 8 20, 1973
Isabelle Fagnot 和 Laurent Vuillon。Sturmian 词中的广义平衡性。 *Discrete*
Appl. Math. 121 1 3 83 101, 2002 ，(-)：–。[
GGOLS25] France Gheeraert、Herman Goulet-Ouellet、Julien Leroy 和 Pierre Stas。树状性的代数刻画。
The Electronic Journal of Combinatorics，32(1)，2025年3月。[GH55] Walter Helbig Gottschalk 和 Gustav Arnold Hedlund。 *Topological dynamics*，第36卷，
American Mathematical Society Colloquium Publications。美国数学会，普罗维登斯，罗得岛州，1955。[GJ00] Richard Gjerde 和 Ørjan Johansen。Cantor 极小系统的 Bratteli-Vershik 模型：Toeplitz 流的应用。
Ergodic Theory Dynam. Systems，20(6):1687–1710，2000。[GL25] France Gheeraert 和 Julien Leroy。 $\mathcal{S}$-进极小树状移位的刻画。*Discrete Mathematics and Theoretical Computer Science*，27:2（组合学），2025。[KN74] Lauwerens Kuipers 和 Harald Niederreiter。 *Uniform distribution of sequences*。纯粹与应用数学。Wiley-Interscience [John Wiley & Sons], 纽约-伦敦-悉尼，1974。[LM95] Douglas Lind 和 Brian Marcus。
An Introduction to Symbolic Dynamics and Coding。剑桥大学出版社，1995。[Lot83] M. Lothaire。
Combinatorics on words，剑桥大学出版社 *Encycl. Math. Appl.* 第17卷，1983。[Lot02] M. Lothaire。
Algebraic Combinatorics on Words，剑桥大学出版社 *Encycl. Math. Appl.* 第90卷，2002。[MH38] Marston Morse 和 Gustav A. Hedlund。符号动力学。 *Amer. J. Math.*，60(4):815–866，1938。[MH40] Marston Morse 和 Gustav A. Hedlund。符号动力学 II：Sturmian 轨道。*American Journal of Mathematics*，62(1/4)，1940。[Mou16] Hamoon Mousavi。Walnut 中的自动定理证明，2016。预印本见 https://arxiv.org/abs/1603.06017。[MR16] Gregory R. Maloney 和 Dan Rust。一维替换子移位与铺砌空间：超越本原性。 *Ergodic Theory and Dynamical Systems*，38(3):1086–1117，2016。[Pan84] Jean-Jacques
Pansiot。*itérés* 380 389 *Complexité des facteurs des mots infinis engendrés par morphismes* 1984 ，第–页。
Springer Berlin Heidelberg，。[PS24] L
´ *Theoretical Computer* eo Poirier 和
Science 998 114535, 2024
Wolfgang Steiner。因子平衡的 S-adic 语言。，：。[PV07] Genevi`eve Paquin 和 Laurent Vuillon。平衡 episturmian 序列的刻画。 *Electron. J. Combin.*，14(1):研究论文33，12，2007。[Que10] Martine Queff
´ *Substitution dynamical systems—spectral analysis* 1294 *Lec-* elec。，[卷，82] 。
ture Notes in Mathematics 2010
施普林格出版社，柏林，第二版，。´Rau´G *Bull. Soc. Math. France* erard Rauzy。代数数与替换。[，110(2):147–178，1982。00] Rig *Theoret. Comput. Sci.* Michel Rigo。正则语言数制系统上自动序列的推广。[，244(1-2):271–281，2000。02] RM
J. Autom. Lang. Michel Rigo 和
Comb. 7 3 351 376, 2002
Arnaud Maes。关于广义自动序列的进一步研究。[，( ): –。10] RSZ´Gw¨ena *Adv. in Appl. Math.* el Richomme、Kalle Saari 和 Luca Q. Zamboni。Tribonacci 词的平衡性与 Abelian 复杂度。，45(2):212–231，2010。

[RSZ11] Gw´ena¨el Richomme、Kalle Saari 和 Luca Q. Zamboni。极小子移位的 Abelian 复杂度。 *J. Lond. Math. Soc. (2)*, 83(1):79–95, 2011。[Sea] Neil J. A. Sloane 等人。在线整数序列百科全书。在线网址：https://oeis.org。[Sha23] Jeffrey Shallit。 *The logical approach to automatic sequences. Exploring combinatorics on words with Walnut* 482 *Lond. Math. Soc. Lect. Note Ser.* 出版社，2023。[Shi19] Takashi Shimomura。极小替换子移位的一种简单方法。 *Topology and its Applications* 260 203 214, 2019 ，：–。[Thu20] J¨org Thuswaldner。 *S*-进序列：连接动力系统、算术与几何的桥梁。载于 *Substitution and tiling dynamics: introduction to self-inducing structures*，第2273卷， *Lecture Notes in Math.*，第97–191页。施普林格，Cham，2020。[Tij80] Robert Tijdeman。主席分配问题。 *Discrete Math.*，32(3):323–330，1980。[Tij00a] Robert Tijdeman。平衡序列的精确覆盖与 Fraenkel 猜想。 *raic number theory and Diophantine analysis (Graz, 1998)* 467 483 载于 ，第–页。de Gruyter，柏林，2000。[Tij00b] Robert Tijdeman。六个序列的 Fraenkel 猜想。 *Discrete Math.*， 222(1-3):223–234，2000。[Wal82] Peter Walters。 *An Introduction to Ergodic Theory* 79 *Graduate Texts in Mathematics* 1982 ，卷。施普林格出版社，纽约，。

Appendix A. Tribonacci linearly recurrent constant in Example 5.10

在本节中，我们使用自由软件 Walnut 来确定例 5.10 中给出的线性递归常数。Mousavi [Mou16] 设计了 Walnut，使其能够自动判定关于大量词的一个广泛族中多种性质的断言是否成立。我们假定读者熟悉该软件的形式体系；若不熟悉，可参阅 [Sha23]，其中包含完整的介绍，以及对可检验的组合性质的综述。

回顾例 5.10 中的特里波纳契替换 $\tau\colon\ \mathcal{A}^* \to \mathcal{A}^*$ 由 $a \mapsto ab$、$b \mapsto ac$ 和 $c \mapsto a$ 给出，其唯一不动点$\mathbf{t} = (t_i)_{i\geq 0}$ 称为特里波纳契词。*Tribonacci*计数系统 是基于特里波纳契数序列的位置计数系统，也就是说，它是由整数序列 $(T_n)_{n\geq 0}$ 定义的，其中$T_0 = 1$、 $T_1 = 2$、 $T_2 = 4$，且对所有 $n \geq 0$ 均有 $T_{n+3} = T_{n+2} + T_{n+1} + T_n$。随后，特里波纳契词$\mathbf{t}$ 是特里波纳契自动的，也就是说，它可以由一个输入采用特里波纳契计数系统读取的确定性有限自动机输出生成。关于广义自动序列的更多内容，例如可参见 [Rig00, RM02]。在 Walnut 中，特里波纳契计数系统和特里波纳契词$\mathbf{t}$分别以 trib 和 TR 为名存储。

回顾一下，若对所有$u \in \mathcal{L}(\mathbf{x}) \setminus \{\varepsilon\}$ 以及所有 $w \in \mathcal{R}_{\mathbf{x}}(u)$，都有 $|w| \leq L|u|$，则称词 $\mathbf{x}$ 是线性复现的，其常数为 $L \geq 0$。特别地，若 $\mathbf{x}$ 的每个长度为 $(L+1)$ 的 n 因子都包含$\mathcal{L}_n(\mathbf{x})$中的每个元素，则称 $\mathbf{x}$ 是线性复现的，其常数为 L。我们更倾向于使用第二种描述，通过 Walnut 计算线性递归常数。定义 $R_{\mathbf{x}}(w)$为满足以下条件的最小整数 k：$\mathbf{x}$ 的每个长度为k 的因子都包含 w；如果不存在这样的 k，则 ∞。我们还将递归函数定义为$R_{\mathbf{x}}(n) = \max_{w\in\mathcal{L}_n(\mathbf{x})} R_{\mathbf{x}}(w)$。因此，若 $\mathbf{x}$ 是线性复现的，其常数为 $L \geq 0$，则有

$$L + 1 = \lim_{n\to\infty} \frac{R_{\mathbf{x}}(n)}{|n|}. \tag{36}$$

为了确定例 5.10 中的线性递归常数，我们需要一个公式来检验**t**中的因子是否相等，即对所有整数 i, j, n，由 def tribeqfac 定义的 tribeqfac(i,j,n) "?msd_trib Au,v (u>=i & u<i+n & u+j=v+i) => TR[u]=TR[v]": 当且仅当 $t_i \cdots t_{i+n-1} = t_j \cdots t_{j+n-1}$时返回 TRUE。它生成一个包含 26 个状态的自动机。由以下内容定义的函数 hasall(j,m,n)

def hasall "?msd_trib Ai Ek $tribeqfac(i,k,n) & k>=j & k+n<=j+m":
当且仅当 $t_j \cdots t_{j+m-1}$ 包含所有长度为n 的因子时返回 TRUE。它生成一个具有 710 个状态的自动机。由下式定义的命令 recur(m,n)
def recur "?msd_trib Aj $hasall(j,m,n)":
当且仅当**t** 的每个长度为m 的因子都包含所有长度为 n的因子时返回 TRUE。它生成一个具有 34 个状态的自动机。递归函数 $R_{\mathbf{t}}(n)$ 随后由下式给出：def rf "?msd_trib $recur(z,n) & ~$ recur(z-1,n)":

引理A.1。 特里波纳契词**t** 的线性递归常数为 $[9, 10]$.

证明。 使用上面写出的命令，我们可以在 Walnut 中运行以下内容：

def check1 "?msd_trib An,z (n>=1 & $rf(n,z)) => z<=11*n": def
check2 "?msd_trib An,z (n>=1 & $rf(n,z)) => z<=10*n":

第一个返回 TRUE，第二个返回 FALSE，由此证明了该引理。 ■

要得到线性递归常数的精确值，还需要进一步计算。设 $(U_n)_{n\geq 0} = 1$，2，4，8，15，28，52，.....为满足递推关系 $U_n = 2U_{n-1} - U_{n-4}$（对所有 $n \geq 4$）且初始条件为 $(U_n)_{0\leq n\leq 3} = 1$，2，4，8 的整数序列。注意，这些整数在 Tribonacci计数系统中的表示，是以 1，10，100，1001，10010，100100，1001001，.....开头的二进制词序列。以下命令
reg link msd_trib msd_trib "[0,0]*[1,1]([0,0][0,0][0,1]) *(()
|[0,0]|[0,0][0,0])": reg adjtrib msd_trib msd_trib "[0,0]*
[0,1][1,0][0,0]*": def tn4 "?msd_trib Ej,k,l $adjtrib(n,j) & $adjtrib(j,k) &
$adjtrib(k
,l) & $adjtrib(l ,z)":
def adju "?msd_trib Eu,t $link(u,x) & $adjtrib(x,t) & $link(t,y)":
分别提供一个自动机，并行接受数对 (T_n, U_n)、(T_n, T_{n+1})、(T_n, T_{n+4}) 和 (U_n, U_{n+1}) 的特里波那契表示（可能带有前导 0）。

引理 A.2. 设 $R_{\mathbf{t}}$ 为 *Tribonacci* 词 **t**的递归函数。若 $U_i < n \leq U_{i+1}$ 对整数 i, n成立，则 $R_{\mathbf{t}}(n) = T_{i+4} + n - 1$。

证明。 以下命令

eval recurformula "?msd_trib Ax,y,z,v,n (n>=1 & $adju(x,y) & n>x & n<=y &$rf(n,z)&$
tn4(x,v)) => z+1=v+n":

检查语句中写出的递归函数 $R_{\mathbf{t}}(n)$的公式，并返回 TRUE。■

回顾一下，对于函数 $f, g\colon \mathbb{N} \to \mathbb{R}$，若对所有 $n \geq 0$ 都有 $f(n)\ g(n) > 0$，我们记作$f \sim g$ if $\lim_{n\to\infty} f(n)/g(n) = 1$。

命题 A.3. *Tribonacci* 词t 的线性递归常数为 $2\psi^2 + \psi + 1$，其中 $\psi \approx 1.839$ 是多项式 $X^3 - X^2 - X - 1$的主根。

证明。由于 $i \to \infty$，众所周知，我们有 $T_i \sim c_T\psi^{i+2}$，其中 $c_T = \frac{1}{-\psi^2+4\psi-1}$；类似地，可以验证 $U_i \sim c_U\psi^{i+6}$，其中 $c_U = \frac{1}{4\psi^2+6\psi+4}$。利用前一个引理和 (36)，线性递归常数为

$$\lim_{i\to+\infty} \frac{T_{i+4}}{U_i + 1} = \frac{c_U}{c_T} = 2\psi^2 + \psi + 1 \approx 9.605,$$

这就足够了。 ■

Appendix B. Details on Example 7.4

回顾一下例 7.4 中有用的定义。定义 $\tau\colon \{0,1\}^* \to \{0,1\}^*\colon 0 \mapsto 01, 1 \mapsto 0$ 以及 $\sigma\colon \{0,1,*\}^* \to \{0,1,*\}^*$，使其满足 $\sigma(a) = \tau(a)$，其中$a \in \{0,1\}$，以及 $\sigma(*) = *11$。令 $\mathbf{x} = \sigma^\omega(*) = *11000101010010010010100\cdots$，并考虑 $\mathbf{x}' = 11000\cdots$，即$\mathbf{x}$的第一次移位。我们证明这些词的若干性质。

引理 B.1. 词 $\mathbf{x}$ 与 $\mathbf{x}'$ 均为字母平衡的。

证明. 我们来证明$\mathbf{x}$的移位$\mathbf{x}'$ 是字母平衡的，这显然也意味着$\mathbf{x}$ 是字母平衡的。显然，词 $\mathbf{x}'$ 存在字母频率，这些频率分别为 $\mu_0 = 1/\varphi$和 $\mu_1 = 1/\varphi^2$，其中 $\varphi \equiv (1+\sqrt{5})/2$是黄金比例（注意，这些频率与斐波那契词中的频率相同）。接下来，我们只需证明任意因子中 0 的偏差是有限的，因为关于 1 的偏差的证明是类似的。设 w 是$\mathbf{x}'$的一个因子，我们来证明 $||w|_0 - \mu_0|w||$是有界的。对某个整数序列$(\ell_i)_{1\leq i\leq k}$以及斐波那契词的因子 p_w, s_w，记 $w = p_w\tau^{\ell_1}(1)\tau^{\ell_2}(1)\cdots\tau^{\ell_k}(1)s_w$。根据经典的佩龙-弗罗贝尼乌斯理论，众所周知我们有 $|\tau^n(1)|_0 - \mu_0|\tau^n(1)| = O(\bar{\varphi}^n)$，其中 $\bar{\varphi} = (1-\sqrt{5})/2$。

由于斐波那契词是字母平衡的，我们有 $|p_w|_0 - \mu_0|p_w| = O(1)$ 和 $|s_w|_0 - \mu_0|s_w| = O(1)$。因此，我们有

$$\begin{aligned} ||w|_0 - \mu_0|w|| &= \Big||p_w|_0 + \sum_{1\leq i\leq k} |\tau^{\ell_i}(1)|_0 + |s_w|_0 - \mu_0\Big(|p_w| + \sum_{1\leq i\leq k} |\tau^{\ell_i}(1)| + |s_w|\Big)\Big| \\ &= \Big|\sum_{1\leq i\leq k} O(\bar{\varphi}^{\ell_i}) + O(1)\Big|, \end{aligned}$$

并且和式 $\sum_{1\leq i\leq k} O(\bar{\varphi}^{\ell_i})$ 在 $\ell_k \to +\infty$ 时收敛，因为 $|\bar{\varphi}| < 1$。 ■

与上一附录中的特里波那契情形类似，*Zeckendorf* 计数系统是一种基于斐波那契数序列的整数位置计数系统。在该计数系统中，如果一个序列可以由确定性有限自动机输出生成，且其输入按 Zeckendorf 计数系统读取，则称该序列为 斐波那契自动 。

引理 B.2. 词 $\mathbf{x}$ 是斐波那契自动的，并且由 *Zeckendorf* 计数系统中图 *3* 的确定性有限自动机输出生成。

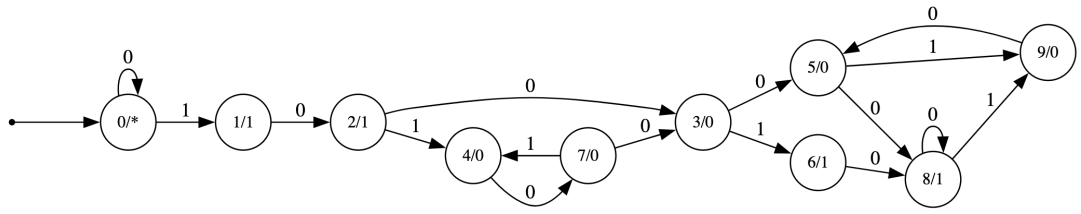


FIGURE 3. 一个在 Zeckendorf 计数系统中生成 x 的确定性有限自动机输出。

证明。 只需证明图 3 中的自动机生成 x。设 $\mathcal{A} = \{*, 0, 1\}$为替换 σ的字母表，并设 $\mathcal{C} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ 为图 3 中自动机的状态集合。由该自动机定义替换 ψ: $\mathcal{C}^* \to \mathcal{C}^*$和 π： $\mathcal{C}^* \to \mathcal{A}^*$，即

ψ:和$\mapsto 01,\ 1 \mapsto 2,\ 2 \mapsto 34,\ 3 \mapsto 56,\ 4 \mapsto 7,\ 5 \mapsto 89,\ 6 \mapsto 8,\ 7 \mapsto 34,\ 8 \mapsto 89,\ 9 \mapsto 5$

$$\pi\colon 0 \mapsto *,\ 1 \mapsto 1,\ 2 \mapsto 1,\ 3 \mapsto 0,\ 4 \mapsto 0,\ 5 \mapsto 0,\ 6 \mapsto 1,\ 7 \mapsto 0,\ 8 \mapsto 1,\ 9 \mapsto 0.$$

我们断言，由该自动机生成的无限词是 x，即 $\pi(\psi^\omega(0)) = \mathbf{x}$ $=\sigma^\omega(*)$。定义替换 ϕ, δ: $\mathcal{C}^* \to \mathcal{A}^*$ 由 $\phi = \pi \circ \psi$和

$$\delta\colon 0 \mapsto \varepsilon,\ 1 \mapsto 0,\ 2 \mapsto \varepsilon,\ 3 \mapsto 1,\ 4 \mapsto 0,\ 5 \mapsto 1,\ 6 \mapsto 1,\ 7 \mapsto \varepsilon,\ 8 \mapsto 1,\ 9 \mapsto 1.$$

注意到

$$\pi(\psi^2(a))\,\delta(b) = \delta(a)\,\sigma(\phi(a))$$

对所有 $ab \in \mathcal{L}(\psi^\omega(0)) \cap \mathcal{C}^2 = \{01, 12, 23, 34, 45, 48, 56, 58, 67, 78, 83, 89, 95, 98\}$ 都成立。因此，通过归纳论证，对所有 $\ell \geq 1$

$$\pi(\psi^2(a_0 \cdots a_{\ell-1}))\,\delta(a_\ell) = \delta(a_0)\,\sigma(\phi(a_0 \cdots a_{\ell-1})) \tag{37}$$

对所有 $a_0 \cdots a_\ell \in \mathcal{L}(\psi^\omega(0)) \cap \mathcal{C}^{\ell+1}$ 都成立。

令 $k \geq 1$ 为整数，设 $a_0 \cdots a_{\ell-1} = \psi^k(0)$，并令 $a_\ell \in \mathcal{C}$ 满足 $\psi^k(0)a_\ell \in \mathcal{L}(\psi^\omega(0)) \cap \mathcal{C}^{\ell+1}$。由于词 $\psi^k(0)$ 以 0 开头，且 $\delta(0) = \varepsilon$，等式 (37) 推出 $\pi(\psi^{k+2}(0))\delta(a_\ell) = \sigma(\phi(\psi^k(0)))$，因此 $\pi(\psi^{k+2}(0))$ 是 $\sigma(\phi(\psi^k(0)))$ 的前缀。令 $k \to +\infty$ 趋于极限，并设 $\mathbf{y} := \psi^\omega(0)$，得到 $\pi(\mathbf{y}) = \sigma(\phi(\mathbf{y}))$。由于 $\phi = \pi \circ \psi$且 $\psi(\mathbf{y}) = \mathbf{y}$，可知 $\mathbf{z} := \pi(\mathbf{y})$ 满足 $\mathbf{z} = \sigma(\mathbf{z})$。因此，对所有$m \in \mathbb{N}$，$\mathbf{z} = \sigma^m(\mathbf{z})$。注意，z 以 $*$开头，因为 y 以 0 开头且 $\pi(0) = *$。因此，令 $m \to +\infty$ 趋于极限，我们得出 $\mathbf{z} = \sigma^\omega(*) = \mathbf{x}$，证毕。■

注 B.3. 要注意，x 是斐波那契自动的，这一点绝非显而易见。事实上，替换 σ 的特征多项式是 $(X-1)(X^2-X-1)$，而这一般并不能保证它是斐波那契自动的。

我们使用自由软件 Walnut 来证明下一个命题。

命题 B.4. 词x 是5-幂自由的，且x 的任何移位都不是复现的。

证明。 为证明该陈述的第一部分，以下两个命令分别用于检验词x是否包含 4 次幂或 5 次幂:

eval pow4 "?msd_fib Ei,n n>0 & At (t>=i & t<i+3*n)=> X[t]=X[t+n]": eval pow5 "?msd_fib Ei,n n>0 & At (t>=i & t<i+4*n)=> X[t]=X[t+n]":
第一个返回 True，第二个返回 False，这表明$\mathbf{x}$ 包含 4 次幂，但它是 5次幂自由的。

在 Walnut 中，我们将因子相等定义如下：谓词 eqfac(i, j, n) 当且仅当$\mathbf{x}[i..i+n] = \mathbf{x}[j..j+n]$ 时返回 True；该谓词通过以下命令定义：

def xeqfac "?msd_fib Au,v (u>=i & u<i+n & u+j=v+i) => X[u]=X[v]":
这会生成一个包含 199 个状态的自动机。为了证明该命题的第二部分，我们使用以下命令：

def norec "?msd_fib An Ei,m i>n & m>n & Aj (j!=i)=> ~$xeqfac(i,j,m)":
这将返回 True，表明 $\mathbf{x}$ 的任何移位都不是复现的。 ■