arXiv:2602.12711v2 [cs.DM] 1 Mar 2026

降低不同平方数上界的另一种方法

Eitatsu Tomita 九州工业大学,
日本
Tomohiro I
九州工业大学，日本

摘要

平方词是指形如 xx 的词，其中 x 是非空词。Brlek 和 Li [在 Comb. Theory, 2025] 证明，词 w 中不同平方子串的数量最多为 $n-\sigma$，其中 σ 是 w 中所用字母的数量。两位作者在 [WORDS, 2023] 中将证明推广，将上界降低至 $n-\Theta(\log n)$。本文提出另一种证明，得到相同的上界 $n-\Theta(\log n)$。

2012年ACM学科分类 计算数学 → 词组合学

关键词和短语 不同平方数的数量；重复；词组合学；Rauzy图

数字对象唯一标识符 10.4230/LIPIcs...

经费资助 *Tomohiro I*: 科研费 24K02899，日本科学技术振兴机构 AIP加速研究 JPMJCR24U4

1 引言

若非空词 x，则形如 xx 的词称为一个平方词 。在重复形式体系中，平方词一直是词组合学的核心研究对象之一 [16, 17, 18]，而 Thue 对无平方词的开创性研究便是其中的代表 [24]。

Fraenkel 和 Simpson [9] 关注了有限词 w （长度为 n）中不同平方子串集合SQ_w 的大小。在猜想 $|SQ_w| < n$的同时，他们通过证明SQ_w 中至多有两个平方词的最右出现位置相同，证明了 $|SQ_w| < 2n$。此后，他们的证明技巧不断改进，从而得到了更紧的上界：Ilie [10]给出的为$2n-\Theta(\log n)$，Lam [12]给出的为 $\frac{95}{48}n$，Deza等人 [7], 给出的为 $\frac{11}{6}n$，而 Thierry [23]给出的为 $1.5n$。

除了这条研究路线外，人们还提出了该猜想的一些变体：Deza等人 [6] 将 w 中所用不同字母的数量 σ 作为参数，并猜想 $|SQ_w| \le w-\sigma$；这一猜想在 [5] 被修订为一个略强的猜想 $|SQ_w| \le n-\sigma-\Theta(\log(n-\sigma))$。[1] 此外，Jonoska等人 [11] 猜想，对于二元词 w，有 $|SQ_w| \le \frac{2k-1}{2k+2}n$，其中 $k \le n/2$ 是 w中较少出现的那个字母的出现次数。

最近，Brlek 和 Li [4] 证明了 $|\text{平方词}_w| \le n-\sigma$，他们采用基于 Rauzy 图 [21]的新方法，突破了原有瓶颈。$\Gamma_w(\ell)$ 的 ℓ 阶 Rauzy 图是一个有向图，其中每个顶点对应 w 的一个长度为 ℓ 的因子，而每条弧对应 w 的一个长度为 $\ell+1$ 的因子，并连接其长度为 ℓ 的前缀和后缀。Brlek 和 Li [4] 证明，存在一个从平方词w 到图 $\bigcup_{\ell=1}^{n}\Gamma_w(\ell)$ 的独立圈集的单射；该独立圈集由阶数从 1 到 n 的 Rauzy 图构成。结果 $|\text{平方词}_w| \le n-\sigma$ 成立，是因为 $\bigcup_{\ell=1}^{n}\Gamma_w(\ell)$ 的独立圈集最大规模为 $n-\sigma$。此外，研究者还扩展了这些技巧，证明了 $|\text{平方词}_w| \le n-\Theta(\log n)$ 在 [3], 中成立，并在 [14]肯定地证明了 [11] 的猜想。

[1] 在 [6, 5], 中，只考虑原始根平方。

© Eitatsu Tomita 和 Tomohiro I；依据知识共享署名 4.0 国际许可协议授权。信息学领域莱布尼茨国际会议论文集，Schloss Dagstuhl – Leibniz信息学中心，Dagstuhl 出版社，德国
LIPICS

XX:2 降低不同平方子串上界的另一种方法

关于长度为 n的所有词 w 中最大 $|\text{平方词}_w|$ 的下界，Fraenkel 和 Simpson [9] 展示了如何构造包含 $n-\Theta(\sqrt{n})$ 个平方词的二元词。Jonoska等人 [11] 证明，通过结构更简单的二元词也能得到相同的渐近下界。由于最大$|\text{平方词}_w|$ 的上下界的首项在 n 处相等，不同平方子串的最大密度 $\frac{|\text{平方词}_n{}^w|}{n}$ 会随着 n 增大而趋近于 1。Manea 和 Seki 的结果 [20] 意味着，这一收敛性对任意 $\sigma \geq 2$ 都成立，且极限永远无法达到。

尽管上界中的首项 n 一般而言是最优的，但人们预计当前最佳上界 $n-\Theta(\log n)$ 中的低阶项 [3] 仍可进一步降低。基于 $|SQ_w|$ 关于二进制词的计算结果（在在线整数数列百科全书中列为 A248958 [22]），Brlek 和 Li [3]提出了以下猜想：

$$|SQ_w| \leq \lceil n+1-\sqrt{n}-\log_2(\sqrt{n})\rceil.$$

为推进这一研究，我们旨在为 $|SQ_w| \leq n-\Theta(\log n)$ 给出另一种证明。

我们还从略有不同于先前研究 [3, 13, 15, 4]的视角，描述环与平方词之间的联系：我们不再通过比较有趣的环的长度与 Rauzy 图的阶数 ℓ 来识别它们，而是根据 Lyndon词对这些环进行分类。我们从这一视角重述并证明已有引理，使本文内容自成一体。

2 预备知识

对于两个整数 b和 e，若 $b \leq e$，则令 $[b..e]$ 表示整数集合$\{b, b+1, \ldots, e\}$；否则，它表示空集。我们还用 $[b..]$ 表示所有大于或等于 b的整数组成的集合$\{b, b+1, \ldots, \}$。

有限词 $w = w[1]w[2]\cdots w[|w|]$ 是由字母组成的序列，其中 $|w| \in [0..]$ 表示 w 的长度；对于 $i \in [1..|w|]$，$w[i]$ 表示 w的第 i 个字母。空词 ε 是长度为 0 的词。对于整数 $1 \leq b \leq e \leq |w|$，w 中起始于 b 、结束于e 的因子 记作 $w[b..e] = w[b]w[b+1]\cdots w[e]$。为方便起见，令 $w[b..e]$，其中 $b > e$ 为空词。w 的前缀（分别为后缀）是指与某个 $e \in [0..|w|]$（分别为 $b \in [1..|w|+1]$）相对应的因子。两个词 x 和 y 的串接 xy 是一种结合二元运算，满足 $(xy)[1..|x|] = x$ 和 $(xy)[|x|+1..|xy|] = y$。

令 $\preceq$ 表示字母上的任意既定全序，并将其扩展为词上的字典序：对于任意两个词 x 和 y，$x \preceq y$ 当且仅当 x 是 y 的前缀，或存在整数 $\ell \geq 0$ 使得 $x[1..\ell] = y[1..\ell]$ 且 $x[\ell+1] \prec y[\ell+1]$。

对于任意词 x 和整数 $m \geq 0$，令 $x^{m/|x|}$ 表示长度为 m 的词，使得$x^{m/|x|}[i] = x[j]$ 对任意满足 $j \equiv i$ 模 m的 $i \in [1..m]$ 和 $j \in [1..|x|]$ 均成立。若 $w = x^{|w|/|x|}$ 且$|x| \leq |w|$，则称整数 $|x|$ 为 w的周期。若词 w 不能写成 $w = x^k$ 的形式，其中 x 为某个词且 $k \in [2..]$，则称其为本原词。非空词若具有 x^k 的形式，其中 x 为某个词且 $k \in [2..]$（此处 x 不一定是本原词），则称其为 k次幂。特别地，2 次幂称为平方词。

对于任意词 x 和 y，如果两个词 xy 和 yx 满足以下条件，则称它们是共轭。对于任意词 z，定义共轭类 $[z]$如下：

$$[z] = \{z[i..|z|]z[1..i-1] \mid i \in [1..|z|]\}$$

对于任意整数 $m \geq 0$，我们也定义

$$[z]_m = \{x^{m/|x|} \mid x \in [z]\}.$$

请注意，我们允许 m 小于 $|z|$。直观地说， $[z]_m$ 是无限词 z^∞ 中长度为m 的因子集合，其中 z^∞ 重复 z。

▶ 例1. 对于 $z = \mathsf{aab}$, $[z]_0 = \{\varepsilon\}$、 $[z]_1 = \{\mathsf{a}, \mathsf{b}\}$、 $[z]_2 = \{\mathsf{aa}, \mathsf{ab}, \mathsf{ba}\}$、 $[z]_3 = [z] = \{\mathsf{aab}, \mathsf{aba}, \mathsf{baa}\}$、 $[z]_4 = \{\mathsf{aaba}, \mathsf{abaa}, \mathsf{baab}\}$ 以及 $[z]_5 = \{\mathsf{aabaa}, \mathsf{abaab}, \mathsf{baaba}\}$。

我们将主要考虑本原词的 $[z]$ 和 $[z]_m$ ，并选择 *Lyndon*词 作为本原词的共轭类中的一个重要代表。

▶ **定义2** ([19, 16])。 **</style}本原词 称为 Lyndon词** ，如果 z 是 $[z]$ 中按字典序最小的词，或者等价地，对于任意 $i \in [2..|z|]$， $z \prec z[i..|z|]$ 均成立。

▶**引理 3**。 *Lyndon*词 z 的最小周期为 $|z|$。

证明。 如果 z的周期为 p且小于 $|z|$，那么其长度为 $z[p+1..|z|]$ 的后缀 $|z|-p>0$ 按字典序小于 z，因为 $z[p+1..|z|] = z[1..|z|-p] \prec z$，这就产生了矛盾。◀

由于任何 k-次幂（包括 $k=2$ 的平方词）都可以写成 x^{kr}，其中 x 是本原词， $r \geq 1$ 是整数，因此我们将其唯一地关联到满足 $x \in [z]$ 的 Lyndon词 z。我们称这样的 z 为该 k-次幂的 *Lyndon* 根。

对于任意词 w，我们定义以下词集：$\mathrm{Fac}_w = \{x \mid w = uxv\}$： w 的因子集合。$\mathrm{Fac}_w(\ell)$ $= \{x \in \mathrm{Fac}_w \mid |x| = \ell\}$： w 中长度为 $\ell \geq 0$ 的因子集合。$\mathrm{Lyn}_w = \{z \in \mathrm{Fac}_w \mid z$ 是 Lyndon词$\}$： w 的 Lyndon因子集合。$\mathrm{SQ}_w = \{x \in \mathrm{Fac}_w \mid x$ 是平方词$\}$： w 中平方词的集合。$\mathrm{SQ}_w(z) = \{x^{2r} \in \mathrm{SQ}_w \mid x \in [z], r \in [1..]\}$： w 中 Lyndon 根为 $z \in \mathrm{Lyn}_w$ 的平方词集合。

▶ 注4. 对于任意词 w， $SQ_w = \bigcup_{z \in Lyn_w} SQ_w(z)$ 和 $|SQ_w| = \sum_{z \in Lyn_w} |SQ_w(z)|$。

有向图由两个互不相交的顶点集和弧集组成，每条弧连接两个顶点，方向从其初始顶点 指向其终止顶点。[2] 如果弧序列 $(u_1, u_2, \ldots, u_m)$ 满足对任意 $i \in [1..m-1]$， u_i 的终止顶点与 u_{i+1} 的初始顶点重合，则称其为长度为 m 的路径。如果 u_m 的终止顶点与 u_1 的初始顶点重合，则称该路径为回路。在路径和回路中，我们沿着序列中弧的方向，从 u_1 的初始顶点移动到 u_m 的终止顶点。如果允许沿相反方向遍历弧，则该图视为无向图。在无向情形下，路径和回路对应的概念分别称为链和环（更正式的定义见 [2, 4] ）。请注意，按照此处的术语，路径（分别为回路）总是链（分别为环）。如果任意一对不同顶点之间都存在一条连接它们的链，则称该图为弱连通图。极大的弱连通子图称为连通分量。

▶ **定义 5.** 对于弧集为 $\{u_1, u_2, \ldots, u_{n_a}\}$ 的图，环 C 的圈向量 是向量 $\mu(C) = (c_1, c_2, \ldots, c_{n_a})$，它位于 n_a 维空间中，并满足 $c_i = r_i - s_i$；其中 r_i 和 s_i 分别表示在该环中沿一个方向和相反方向经过 u_i 的次数。

[2] 在我们的工作中，需要使用多重图来处理0阶 Rauzy 图，因此允许不同的弧具有相同的初始顶点和终止顶点。我们还允许存在环，即初始顶点与终止顶点重合的弧。

XX：4 降低不同平方子串下界的另一种方法

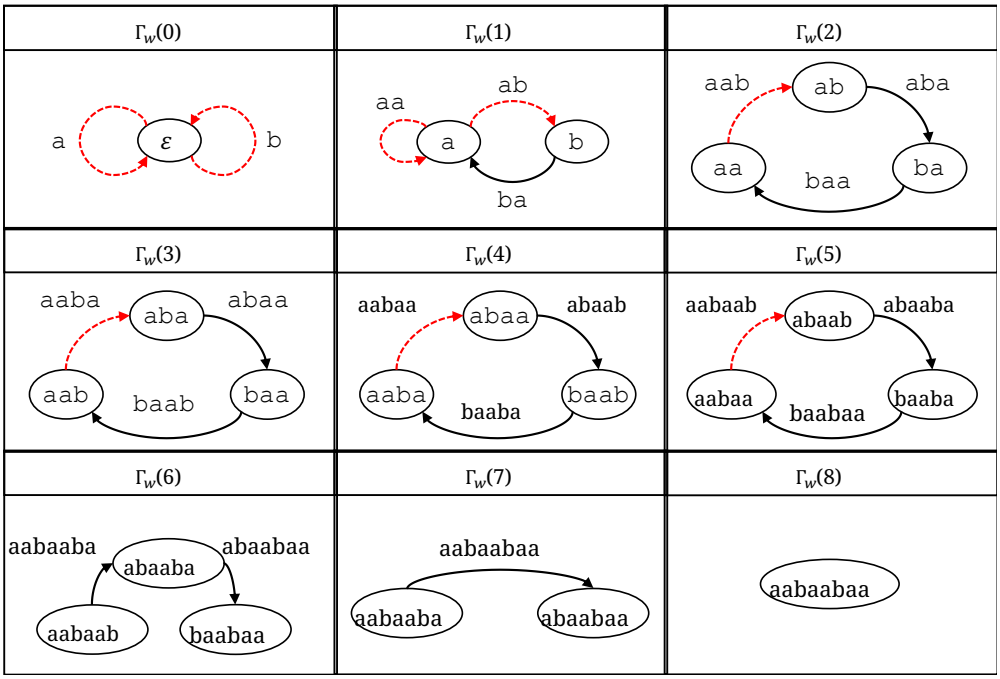


图 1 展示了长度为 8 的词 $w = \texttt{aabaabaa}$ 的所有阶数 $\ell \in [0..8]$ 的 Rauzy 图 $\Gamma_w(\ell)$。每条虚线弧都是某个 $z \in \mathrm{Lyn}_w$（定义见定义 13）对应的 $\mathrm{CS}_w(z)$ 回路中的最小弧，其中 $\mathrm{CS}_w(\texttt{a}) = \{(\texttt{a}), (\texttt{aa})\}$，$\mathrm{CS}_w(\texttt{b}) = \{(\texttt{b})\}$，$\mathrm{CS}_w(\texttt{ab}) = \{(\texttt{ab}, \texttt{ba})\}$，而 $\mathrm{CS}_w(\texttt{aab}) = \{(\texttt{aab}, \texttt{aba}, \texttt{baa}), (\texttt{aaba}, \texttt{abaa}, \texttt{baab}), (\texttt{aabaa}, \texttt{abaab}, \texttt{baaba}), (\texttt{aabaab}, \texttt{abaaba}, \texttt{baabaa})\}$。在第 3 节中，我们将把 $\mathrm{SQ}_w(\texttt{a}) = \{\texttt{aa}\}$ 指派给 $\mathrm{CS}_w(\texttt{a})$，并把 $\mathrm{SQ}_w(\texttt{aab}) = \{\texttt{aabaab}, \texttt{abaaba}, \texttt{baabaa}\}$ 指派给 $\mathrm{CS}_w(\texttt{aab})$。

▶**定义 6.** 若一组环 $\{C_1, C_2, \ldots, C_k\}$ 的圈向量线性无关，则称这组环线性无关，即 $\alpha_1\mu(C_1) + \alpha_2\mu(C_2) + \cdots + \alpha_k\mu(C_k) = \mathbf{0}$ 意味着 $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$。

▶ **定理 7**（定理 2，第 4 章，见 [2]）。对于一个具有 n_v 个顶点、n_a 条弧和 n_c 个连通分量的多重图，其圈秩定义为 $n_\mathsf{a} - n_\mathsf{v} + n_\mathsf{c}$，等于独立圈集中元素数量的最大值。

▶**定义 8**（Rauzy 图 [21]）。词 w 的 $\ell \in [0..|w|]$ 阶Rauzy 图 $\Gamma_w(\ell)$ 定义在顶点集 $\mathit{Fac}_w(\ell)$ 和弧集 $\mathit{Fac}_w(\ell+1)$ 上，其中每条弧 $u \in \mathit{Fac}_w(\ell+1)$ 始于顶点 $u[1..\ell] \in \mathit{Fac}_w(\ell)$，终于顶点 $u[2..|u|] \in \mathit{Fac}_w(\ell)$。

令 $\Gamma_w = \bigcup_{\ell=0}^{|w|} \Gamma_w(\ell)$ 为一个图，它将所有阶的 Rauzy 图 $\ell \in [0..|w|]$ 汇集于同一个图中，其顶点集为 Fac_w，弧集为 $Fac_w \setminus \{\varepsilon\}$，并且有 $|w|+1$ 个连通分量。注意，每个词 $x \in Fac_w \setminus \{\varepsilon\}$ 在 Γ_w中既是顶点也是弧，但两者彼此区分。参见图 1。

▶**引理 9.** 对于任意词 w，Γ_w 的独立圈集的最大大小为 $|w|$.

证明。由于 Γ_w 有 $|Fac_w|$ 个顶点，$|Fac_w \setminus \{\varepsilon\}|$ 条弧和 $|w|+1$ 个连通分量，其圈秩为 $|Fac_w \setminus \{\varepsilon\}| - |Fac_w| + |w| + 1 = |w|$，根据定理 7，这就是 Γ_w 的独立圈集的最大大小。◀

3 关于 $|SQ_w| \leq |w| - \Theta(\log |w|)$ 的另一种证明

Brlek 和 Li [4] 证明了，存在一个从 SQ_w 到 $\bigcup_{\ell=1}^{|w|} \Gamma_w(\ell)$ 中一个独立圈集的单射。更确切地说，Lyndon词 z 对应的 $SQ_w(z)$ 中的平方词被映射到$|SQ_w(z)|$ 回路，其中每个回路都存在于 $\Gamma_w(\ell)$ 中，适用于 $\ell \in [|z|..|z| + |SQ_w(z)| - 1]$。在 [3, 15], 中，独立圈集通过添加另一个回路（文中称为本原回路）得以扩展，该回路存在于 $\Gamma_w(|z| - 1)$ 中。尽管 $|SQ_w| \leq |w| - \Theta(\log |w|)$ 在 [3] 中的证明基于另一种扩展，但我们证明，借助本原回路也能达到相同的上界。

我们先给出两个简单的引理（引理10和引理11），它们将用于识别一些回路，这些回路的弧集 $[z]_m$ 分布于不同阶数的 Rauzy 图中。

▶ **引理10.** 对于任意词 w 和 $z \in Lyn_w$，若对某个整数 $m > 0$有 $[z]_m \subseteq Fac_w$，则有 $[z]_{m-1} \subseteq Fac_w$。

证明. 对于任意词 $u \in [z]_{m-1}$，存在一个词 $x \in [z]$ 使得 $u = x^{(m-1)/|x|}$。由于 u 是 $x^{m/|x|} \in [z]_m \subseteq Fac_w$的前缀，$u$ 必定是 w的因子。◀

▶ **引理11.** 对于任意词 w 和 $z \in Lyn_w$，若对某个整数 $m > 0$有 $[z]_m \subseteq Fac_w$，则 $\Gamma_w(m-1)$ 包含一个弧集为 $[z]_m$ 的回路。

证明。令 $x_i = z[i..|z|]z[1..i-1]$ 对任意 $i \in [1..|z|]$。于是，$[z]_m = \{x_1^{m/|z|}, x_2^{m/|z|}, \ldots, x_{|z|}^{m/|z|}\}$。我们证明，$\Gamma_w(m-1)$ 中的弧序列 $(x_1^{m/|z|}, x_2^{m/|z|}, \ldots, x_{|z|}^{m/|z|})$ 是一个回路：对于任意 $i \in [1..|z| - 1]$，弧 $x_i^{m/|z|}$ 的终止顶点是 $x_i^{m/|z|}[2..m] = x_{i+1}^{(m-1)/|z|} = x_{i+1}^{m/|z|}[1..m-1]$，这与下一条弧 $x_{i+1}^{m/|z|}$ 的初始顶点重合。类似地，弧 $x_{|z|}^{m/|z|}$ 的终止顶点是 $x_{|z|}^{m/|z|}[2..m] = x_1^{(m-1)/|z|} = x_1^{m/|z|}[1..m-1]$，这与 $x_1^{m/|z|}$ 的初始顶点重合。◀

▶ 注 12. 如果 $m < |z|$，引理11的证明中考虑的长度为 $|z|$ 的回路可能会两次经过某个顶点和/或两次使用某条弧。另一方面，对于$m \geq |z|$的情形，可以利用 Fine 和 Wilf 周期引理 [8], 证明该回路恰好经过每个顶点一次，但由于本文不使用这一性质，我们省略证明。

▶ **定义 13。** 对于任意词 w 和 $z \in Lyn_w$，令 $CS_w(z)$ 表示 Γ_w中的回路集合，使得 $CS_w(z)$ 中某个回路的弧集对于某个整数 $m \geq |z|$而言是 $[z]_m$。

我们的目标是将$SQ_w(z)$专门分配给 $CS_w(z)$。[3] 下一个引理表明，对于任意一对 Lyndon因子，$CS_w(z)$的集合彼此不相交，并且它们的并集构成 Γ_w的一个线性无关圈集：

▶**引理 14。** 对于任意词 w，集合 $\bigcup_{z \in Lyn_w} CS_w(z)$ 包含 $\sum_{z \in Lyn_w} |CS_w(z)|$ 个回路，并且是 Γ_w 的独立圈集。

证明。 回想一下，$CS_w(z)$ 中任意回路的弧集都具有以下形式：$[z]_m$ 其中$m \geq |z|$为某个整数。我们声称，可以将该回路唯一地指派给 $z^{m/|z|}$ 中的 $[z]_m$字典序最小的弧。对于两个不同的 Lyndon 因子 z_1 和 $z_2 \in Lyn_w$，若 $|z_1| \leq |z_2|$ 且 $m \geq |z_2|$为整数，则必有 $z_1^{m/|z_1|} \neq z_2^{m/|z_2|}$，因为：若 $|z_1| = |z_2|$，结论显然成立

[3] 注意，即使对于满足 $|SQ_w(z)| = 0$ 的 $z \in Lyn_w$，$CS_w(z)$ 也有明确定义。

XX：6 降低不同平方子串下界的另一种方法

因为 $z_1^{m/|z_1|} \neq z_2^{m/|z_2|}$ 如 $z_1 \neq z_2$。否则， $z_1^{m/|z_1|} = z_2^{m/|z_2|}$ 意味着 z_2 具有周期$|z_1| < |z_2|$，这与引理 3 矛盾。因此， $CS_w(z_1)$ 与 $CS_w(z_2)$ 不相交，由此得到$|\bigcup_{z\in Lyn_w} CS_w(z)| = \sum_{z\in Lyn_w} |CS_w(z)|$。

令 C_1、C_2、……、C_k 表示$\bigcup_{z\in Lyn_w} CS_w(z)$ 中的回路，并按照其字典序最小的弧以严格递增顺序排列，即 $k = \sum_{z\in Lyn_w} |CS_w(z)$。之所以能够如此排列，是因为如上所见，这些弧两两不同。注意，对于任意 $1 \le i < j \le k$， C_i 的字典序最小弧不包含于 C_j 中。现在我们证明

$$\alpha_1\mu(C_1) + \alpha_2\mu(C_2) + \cdots + \alpha_k\mu(C_k) = \mathbf{0} \Longrightarrow \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0.$$

为了使等式 $\sum_{i=1}^{k}(\alpha_i\mu(C_i)) = \mathbf{0}$成立，我们需要 $\alpha_1 = 0$，因为否则，对应于 C_1 的字典序最小弧的向量分量不可能为零。对剩余等式 $\sum_{i=2}^{k}(\alpha_i\mu(C_i)) = \mathbf{0}$ 递归地应用相同的论证，我们可以看到所有系数 α_1、α_2、……、α_k 都必须为零。◀

以下结论可由引理 9 和引理 14 直接得出：

▶ **引理 15.** 对于任意词 w, $|\bigcup_{z\in Lyn_w} CS_w(z)| \le |w|$。

下一个引理给出了$SQ_w(z)$ 和$CS_w(z)$ 的大小。

▶ **引理16.** 对于任意词 w 和 $z \in Lyn_w$，若 $|SQ_w(z)| \ge 1$,

$$|SQ_w(z)| = |z|(r-1) + s,$$
$$|CS_w(z)| \ge 2|z|(r-1) + s + 1,$$

其中

$$r = \max\{i \in [1..] \mid \exists x^{2i} \in SQ_w(z)\},$$
$$s = \left|\{x \in [z] \mid x^{2r} \in SQ_w(z)\}\right|.$$

证明. 设 $x_i = z[i..|z|]z[1..i-1]$ 对任意 $i \in [1..|z|]$，并设 $\{k_j\}_{j=1}^{s}$ 为满足以下条件的整数集合： $0 < k_1 < k_2 < \cdots < k_s < |z|$ 使得 $x_{k_j}^{2r} \in SQ_w$ 对任意 $j \in [1..s]$。

对于每一对 $i \in [1..|z|]$ 和 $d \in [1..r-1]$， x_i^{2d}都会构成一个不同的平方词；由于 x_i^{2d} 是 $x_{k_1}^{2r}$的一个因子，该平方词存在于$SQ_w(z)$中。由 s的定义可知，显然 $|SQ_w(z)| = |z|(r-1)$。

为了证明 $|CS_w(z)| \ge 2|z|(r-1) + s + 1$，令 $g = \max\{k_{j+1} - k_j \mid j \in [1..s]\}$，其中 $k_{s+1} = k_1 + |z|$。容易看出 $g \le |z| - s + 1$，因为当$k_{j+1} - k_j = 1$ 时， g 取最大值；对每个 $j \in [1..s]$ 而言，只有使 g达到最大值的那个例外。现在令$M = 2|z|r - g + 1$，并断言 $[z]_M \subseteq Fac_w$。对于任意满足 $i \in [k_j..k_{j+1}-1]$ 的 $i \in [1..|z|]$，若存在某个 $j \in [1..s]$使得该条件成立，那么 $x_i^{M/|z|}$是 $x_{k_j}^{2r}$的一个因子，因为 $i - i_j + 1 \le g$，并且从 $x_{k_j}^{2r}$中 $i - k_j + 1$开始的 x_i的重复序列会持续 $2|z|r - (i - k_j + 1)$ 个字母，足以覆盖 $x_i^{M/|z|}$。同理，对于任意 $i \in [1..k_1 - 1]$, $x_i^{M/|z|}$ 是 $x_{k_s}^{2r}$的一个因子。因此，$[z]_M \subseteq Fac_w$。图 2 展示了一个示意图。

由引理10可得，对任意 $m \in [1..M]$， $[z]_m \subseteq Fac_w$，根据引理11，这会生成一个弧集为 $[z]_m$ 的回路。在这些回路中， $CS_w(z)$ 包含$M - |z| + 1 = 2|z|r - |z| - g + 2 \ge 2|z|(r-1) + s + 1$ 个回路，分布在阶数从 $|z| - 1$ 到 $M - 1$的 Rauzy 图中。◀

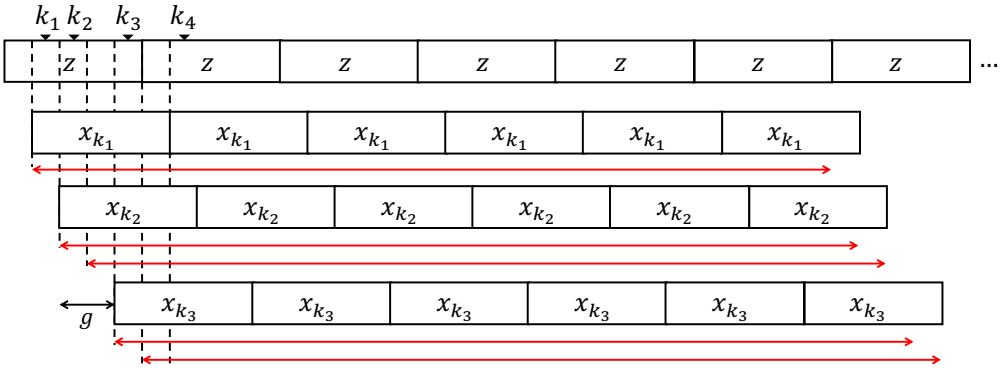


图 2 引理16的证明示意图。这里考虑 $|z|=5$、 $r=3$、 $s=3$ $k_1=2$ $k_2=3$ $k_3=5$ $k_2-k_1=1$ $k_3-k_2=2$ $k_4-k_3=k_1+|z|-k_3=2$ 、以及 的情况。由于,,, 我们有 $g=2$。$SQ_w(z)$包含 s 个形如 $x_{k_j}^{2r}$ for $j\in[1..s]$的正方形，它们在 z^∞ 下对齐。我们看到，z^∞ 中长度为 $2|z|r-g+1$ 的每个因子（以红色双向箭头表示）都必定被某个 $x_{k_j}^{2r}$ 覆盖，因此也是 w 的因子。

▶ **定义17.** 对于任意词 w 和 $z\in Lyn_w$，满足 $|CS_w(z)|\geq 1$，我们定义

$$avg_w(z)=\frac{|SQ_w(z)|}{|CS_w(z)|},$$

并且注意到，$CS_w(z)$ 中的每个回路都装载了 $avg_w(z)$ 个平方词。

根据引理16中以变量 r 和 s得到的 $|CS_w(z)|$ 的下界，可以预期$avg_w(z)$在 $r=1$ 和 $s=|z|$时能达到其最大值 $\frac{|z|}{|z|+1}$，这一点将在下一引理中正式证明。

▶ **引理18.** 对于任意词 w 和 $z\in Lyn_w$，且 $|CS_w(z)|\geq 1$，$avg_w(z)\leq\frac{|z|}{|z|+1}$。 ——

证明。 若 $|SQ_w(z)|=0$，则易见$avg_w(z)=0\leq\frac{|z|}{|z|+1}$。对于其余情形，我们证明[1] $avg_w(z)\geq\frac{|z|+1}{|z|}$。利用引理16中的变量 r 和 s，我们有

$$\frac{1}{avg_w(z)}=\frac{|CS_w(z)|}{|SQ_w(z)|}\geq\frac{2|z|(r-1)+s+1}{|z|(r-1)+s}=1+\frac{|z|(r-1)+1}{|z|(r-1)+s}.$$

由于 $1\leq s\leq|z|$，当 $s=|z|$时，最后一个表达式取最小值，即

$$\frac{1}{avg_w(z)}\geq 1+\frac{|z|(r-1)+1}{|z|(r-1)+|z|}=1+\frac{|z|(r-1)+1}{|z|r}.$$

由于 $|z|\geq 1$，我们有 $|z|(r-1)+1\geq r$，因此，

$$\frac{1}{avg_w(z)}\geq 1+\frac{r}{|z|r}=1+\frac{1}{|z|}=\frac{|z|+1}{|z|}.$$

因此，$avg_w(z)\leq\frac{|z|}{|z|+1}$ 成立。 ◀

我们还有以下性质：

▶ **引理19。** 对于任意词 w 和 $\ell\in[0..|w|]$，$\Gamma_w(\ell)$ 中的一个回路至多装载$\ell_{\ell+2}^{\ell+1}$ 个平方词。 ——

XX：8 另一种降低不同平方子串界的方法

证明。 (ℓ) 中的每个回路都装载了 $SQ_w(z)$ 中的平方词，且 $|z| \le \ell + 1$，因此根据引理18，其装载值至多为 $avg_w(z) \le \frac{|z|}{|z|+1} \le \frac{\ell+1}{\ell+2}$。◀

为了证明定理22，我们根据 w 中最长平方词是否足够短（引理20）将问题分为两种情况，若不够短则应用引理21。

▶ **引理 20。** 对于任意长度为 n 的词 w，如果不包含长度为 $\ge 2\sqrt{n}$ 的平方词，则有 $|SQ_w| \le n - \sqrt{n}$。

证明。 回顾一下，$SQ_w = \sum_{z \in Lyn_w} SQ_w(z)$。对于固定的 Lyndon 因子 $z \in Lyn_w$，满足 $|SQ_w(z)| \ge 1$，关于 w 的假设意味着 $|z| \le \sqrt{n} - 1$。因此，由引理18可知，Γ_w 中的每个回路至多可装载 $avg_w(z) \le \frac{|z|}{|z|+1} \le \frac{\sqrt{n}-1}{\sqrt{n}}$ 个平方词。根据引理15，待装载的回路至多有 n 个，因此我们有

$$|SQ_w| \le \frac{n(\sqrt{n}-1)}{\sqrt{n}} = n - \sqrt{n}.$$

证明至此结束。◀

▶ **引理21。** 对于任意长度为 w n 的词，只要它包含一个长度为 $\ge 2\sqrt{n}$的平方词，就有 $|SQ_w| \le n - \Theta(\log n)$。

证明。 设 $M \ge \sqrt{n}$ 为一个整数，使得存在一个长度为 $2M$ 的 w平方词。由于该平方词包含所有长度为 M的共轭词，因此有 $[z]_M \subseteq Fac_w$ 及其 Lyndon 根z。由引理 10 和引理 11 可知，$\Gamma_w(m-1)$ 对于任意$m \in [1..M]$至少包含一个回路。由于引理19表明，$\Gamma_w(m-1)$ 中的每个回路最多可装载 $\frac{m}{m+1}$ 个平方词，因此存在 M 个回路，其装载的平方词总数达到

$$\sum_{m=1}^{M} \frac{m}{m+1} = \sum_{m=1}^{M} \left(1 - \frac{1}{m+1}\right) = M - \Theta(\log M).$$

根据引理 15，至多还剩下 $n - M$ 个线性无关的回路，它们总共装载至多 $n - M$ 个平方词。综合以上结果，可得

$$|SQ_w| \le (n - M) + (M - \Theta(\log M)) = n - \Theta(\log M) \le n - \Theta(\log n).$$

证明至此结束。◀

定理22由引理20和引理21得出。

▶**定理22。** 对于任意长度为 n 的词 w，w 中不同平方子串的数量上界为 $n - \Theta(\log n)$。

4 结论与未来工作

我们为长度为 n的词中不同平方子串的数量给出了上界 $n - \Theta(\log n)$的另一种证明。我们希望，本文获得的新见解能够得到进一步拓展，从而继续降低这一上界。此外，如果我们的技术能够用于改进相关上界，例如词 [15],中不同 k次幂的上界以及循环词 [1]中不同平方子串的上界，也将很有意义。

参考文献

1 Mika Amit 和 Pawel Gawrychowski。循环词中的不同平方子串。载于第24届字符串处理与信息检索国际研讨会（SPIRE）2017论文集，27–37页，2017年。doi:10.1007/978-3-319-67428-5_3。2 Claude Berge。《图论及其应用》。Greenwood Press，1982年。3 Srecko Brlek 和 Shuo Li。有限序列中不同平方子串的数量：一些新旧结果。载于 Anna E. Frid 和 Robert Mercas（编），第14届词组合学国际会议（WORDS）2023论文集，《计算机科学讲义》第13899卷，35–44页。Springer，2023年。 doi:10.1007/978-3-031-33180-0_3。4 Srecko Brlek 和 Shuo Li。有限词中的平方数量。Comb. Theory，5(1)，2025年。URL: https://doi.org/10.5070/c65165014， doi:10.5070/C65165014。5 Antoine Deza 和 Frantisek Franek。关于字符串中不同平方子串和游程最大数量的d步方法。离散应用数学，163，第3部分(0)：268–274，2014年。doi:http://dx.doi.org/10.1016/j.dam.2013.10.021。6 Antoine Deza、Frantisek Franek 和 Mei Jiang。字符串中不同平方子串的d步方法。载于第22届组合模式匹配年度研讨会（CPM）2011论文集，77–89页，2011年。doi:10.1007/978-3-642-21458-5_9。7 Antoine Deza、Frantisek Franek 和 Adrien Thierry。一个字符串可以包含多少个双平方？离散应用数学，180：52–69，2015年。doi:10.1016/j.dam.2014.08.016。8 Nathan Jacob Fine 和 Herbert Saul Wilf。周期函数的唯一性定理。美国数学会，16(1)：109–114，1965年。9 Aviezri S. Fraenkel 和 Jamie Simpson。一个字符串可以包含多少个平方词？J. Comb. Theory A，82(1)：112–120，1998年。URL: https://doi.org/10.1006/jcta.1997.2843，doi:10.1006/JCTA.1997.2843。10 Lucian Ilie。关于词中平方数量的注记。Theor. Comput. Sci.，380(3)：373–376，2007年。doi:10.1016/j.tcs.2007.03.025。11 Natasa Jonoska、Florin Manea 和 Shinnosuke Seki。关于二元词的更强平方猜想。载于第40届计算机科学理论与实践当前趋势国际会议（SOFSEM）2014论文集，339–350页，2014年。 doi:10.1007/978-3-319-04298-5_30。12 Nguyen Huong Lam。关于字符串中的平方数量。技术报告 AdvOL-Report 2013/2，麦克马斯特大学，2013年。13 Shuo Li。关于lie复杂度及其拓展的注记。Theor. Comput. Sci.，986：114343，2024年。doi:10.1016/J.TCS.2023.114343。14 Shuo Li。二进制词中不同幂数量的一个上界。Discret. Math.，347(4)：113902，2024年。 doi:10.1016/J.DISC.2024.113902。15 Shuo Li、Jakub Pachocki 和 Jakub Radoszewski。关于有限词中k次幂最大数量的注记。Electron. J. Comb.，31(3)，2024年。 doi:10.37236/11270。16 M. Lothaire。《词组合学》。剑桥数学文库。剑桥大学出版社，第2版，1997年。17 M. Lothaire。《词的代数组合学》。剑桥大学出版社，2001年。18 M. Lothaire。《词的应用组合学》。剑桥大学出版社，2005年。19 Roger Conant Lyndon。关于Burnside问题。美国数学会会刊，77(2)：202–215，1954年。20 Florin Manea 和 Shinnosuke Seki。平方密度递增映射。载于第10届词组合学国际会议（WORDS）2015论文集，160–169页，2015年。doi:10.1007/978-3-319-23660-5_14。21 Gérard Rauzy。《有限字母表上的序列》。波尔多数论研讨班，12：1–16，1982—1983年。URL： http://eudml.org/doc/182163。22 OEIS基金会（2026年）。长度为 n的二进制字符串中不同非空平方的最大数量，《在线整数数列百科全书》中的条目A248958。访问日期：2026年2月1日。URL： https://oeis.org/A248958。

XX:10 关于不同平方子串界的另一种降低方法

23 Adrien Thierry. 关于长度为 n 的词含有少于 1.5n 个不同平方的证明，2020年。arXiv:2001.02996。24 Axel Thue. 论无穷字符序列。克里斯蒂安尼亚科学学会文刊，数学与自然科学类，7:1–22，1906。