arXiv:2602.14722v1 [cs.FL] 16 Feb 2026

亮点

通过内段二分法对上下文无关交集进行几何刻画

Jorge Miguel Silva

- 下推自动机弧之间的交叉间隙并不能决定上下文无关语言交集是否为上下文无关语言。
- 内部片段度量——交错交叉端点之间较长内段的长度——是起决定作用的量。
- 有界内部片段可实现缓冲下推自动机构造；不断增长的内段若伴有抽运敏感链接，则意味着该语言非上下文无关。
- 对于块计数上下文无关语言，这一二分结论是完备的：当且仅当弧是联合良嵌套的，它才是上下文无关语言。
- 一般的上下文无关语言交集问题可归约为一个关于泵敏感分解的特定问题。

通过内段二分定理对上下文无关交集进行几何刻画

Jorge Miguel Silva*

IEETA/DETI，阿威罗大学，圣地亚哥大学城校区，阿威罗，*3810-193*，葡萄牙

摘要

两个上下文无关语言的交集通常并非上下文无关语言，但尚无几何判据能够刻画何时交集仍是上下文无关语言。交叉间隙（两条交叉的压入弹出弧之间的 $\max(i' - i,\ \ j' - j)$）是一个自然的候选指标。我们否定了这一猜想：我们给出了上下文无关语言，其交集虽存在无界间隙交叉，仍是上下文无关语言。

起决定作用的是内部片段度量：对于导出分解 $w = P_1P_2P_3P_4$ 的交叉弧，该度量为 $\max(|P_2|,\ |P_3|)$，即交错交叉端点之间较长内段的长度。我们证明了该度量满足二分定理：有界内段可通过有限缓冲区构造保证上下文无关性；若内段不断增长且链接对抽运敏感，则可推出非上下文无关性。内段这一概念适用于所有上下文无关语言交集；由此得到的刻画是否严格，取决于语言类。

对于块计数上下文无关语言（要求指定的成对区块长度相等的语言），二分结论已完整确立：当且仅当组合弧联合良嵌套时，交集才是上下文无关语言。对于一般上下文无关语言，上下文无关语言方向无条件成立；非上下文无关方向则要求存在泵敏感的链接，而这些链接是否必然存在是主要的开放问题。这使一般上下文无关语言交集问题归结为对泵敏感分解某一特定性质的研究。

关键词：上下文无关语言交集、下推自动机、内部片段度量、良嵌套性、交叉依存关系、泵敏感

*通讯作者电子邮箱：jorge.miguel.ferreira.silva@ua.pt (Jorge Miguel Silva)

链接，线性上下文无关重写系统*2000*
年 *MSC*： 68Q45、68Q42、03D05

---

1. 引言

上下文无关语言对于交集运算不封闭。下推自动机的力量源于其栈，而栈的局限在于它的刚性：它提供后进先出（LIFO）访问，因此，埋在栈顶之下的信息无法访问，除非先移除其上方的所有内容。

当两个上下文无关语言取交集时，问题在于：两个彼此独立的栈协议（每个协议都跟踪自身的入栈—出栈依赖关系）能否交错合并到同一个栈中。如果两组依赖弧能够良好嵌套，一个栈就可以同时处理它们；如果不能，交集可能完全超出上下文无关语言类。由 $\{a^m b^* c^m d^*\} \cap \{a^* b^n c^* d^n\}$产生的语言 $\{a^m b^n c^m d^n\}$就是经典例子：其 a到c 以及 b到d 的依赖关系交叉，因此单个栈无法同时处理二者 Shieber（1985年）；Joshi（1990年）。Bodirsky、Kuhlmann 和 Möhl 等人 Bodirsky et al.（2005年）在线性上下文无关重写系统层级中通过良嵌套性 对这一现象进行了形式化。

然而，并非所有交叉都会破坏上下文无关性。如果来自两个下推自动机的弧只在相邻位置交叉，就像定理 2.1中的交错回文语言一样，这种交叉是局部的，而成对字母表编码可以将其吸收。这引出了一个几何学问题：交叉的依赖弧具有什么性质，决定了两个栈协议能否共用一个栈？这一问题最初源于序列分析和数据压缩中的结构分解问题，但它适用于所有上下文无关语言，并且在自动机理论文献中一直悬而未决。

自然的猜想及其反驳。两个交叉弧的交叉间隙，形式化定义为 max($i' - i$, $j' - j$)（即入栈端点距离与弹出端点距离中的较大值；定理 5.1），是一个自然的候选指标：如果间隙随输入规模增长，栈就应该会失效；间隙有界的交叉应该无害。

这个猜想是错误的。在定理 6.5 中，我们给出两个上下文无关语言，它们的交集仍然是上下文无关的，尽管存在交叉间隙为 $\Theta(n)$ 的交叉弧。该构造是显式的：一个下推自动机执行一次跨越两个位置的入栈-出栈对，而另一个下推自动机的匹配弧在其周围跨越$\Theta(n)$个位置。两个内段的长度都恰好为 1，因此这种“中断”被有限缓冲区所吸收，但交叉间隙却不断增长

而无上界。直观地说，一条跨越许多位置、但不包围任何其他机器操作的弧，会形成很大的间隙，却不会给栈带来任何负担。

核心度量：内段。合适的量是内部片段度量 $\max(|P_2|, |P_3|)$。当两条弧 (i, j) 和 (i', j') 在 $i < i' < j < j'$ 的条件下交叉时，字符串可分解为 $w = P_1P_2P_3P_4$，而此度量反映了交错端点之间较长内段的长度。有界中断是可由缓冲区处理的：有限状态缓冲区吸收较短的交叉弧，而单个下推自动机处理其余部分。无界中断则会不可挽回地破坏后进先出规则。

我们的主要结果将这一直觉形式化如下：

1. **有界内部片段 $\Rightarrow$ 上下文无关语言**（定理 6.7）：若所有交叉对均满足 $\max(|P_2|, |P_3|) = O(1)$，则我们可以构造一个识别该交集的下推自动机。缓冲区大小相对于界限呈指数级增长，但与输入长度无关。2. **增长的内段 $\Rightarrow$ 非上下文无关语言**（定理 6.1）：若某个交叉对满足 $\max(|P_2|, |P_3|) = \omega(1)$，且具有抽运敏感链接（即依赖关系：若修改一个链接片段而不修改其配对片段，就会导致字符串不再属于该语言），则该交集不是上下文无关语言。3.**块计数上下文无关语言的完整刻画**（定理 4.4）：当且仅当组合弧联合良嵌套时，该交集才是上下文无关语言。4. **否定间隙猜想**（定理 6.5）：交叉间隙无界并不意味着语言是非上下文无关语言；决定性度量是内部片段度量。

结果 (1) 与 (2) 合并为**内段二分定理**（定理 6.8），汇总于下表：

内部片段度量	交叉间隙	交集	参考
0（无交叉）	–	CFL	（乘积构造）
any	$O(1)$	CFL	定理 5.4
$O(1)$	$\omega(1)$	CFL	定理 6.7
$\omega(1)^\dagger$	$\omega(1)$	非上下文无关语言	定理 6.1

†需要使用对抽引敏感的连接；对块计数上下文无关语言自动成立，对一般上下文无关语言仍为开放问题（定理 6.9）。注：有界间隙（第 2 行）总是得到上下文无关语言，因此仅当间隙同样无界时，链接条件才可满足。

范围。 第 2 节确定符号约定。第 3 节至第 6 节阐述形式化结果。第 7 节综述相关工作，第 8 节将这些结果置于乔姆斯基层级中，第 9 节列出开放问题。

2. 预备知识

我们使用格雷巴赫范式（GNF）下的非确定性下推自动机：每次转移恰好读取一个输入符号，并至多执行一次栈操作，即一次入栈或一次出栈。二元产生式（$A \to aBC$）要求压入两个符号；我们通过一个辅助的 ϵ-移动，将每个这样的步骤分解为两次转移。因此，每台机器的每个输入位置至多对应两次压入操作。这种分解的详细内容见第 5 节。

对于接受字符串 w 的下推自动机 M，压栈-出栈匹配 $\mu_M(w)$ 是由弧 $(i,\ j)$ 构成的集合，这些弧将压入操作发生的位置 i 与对应出栈操作发生的位置 j 配对。由于 M 具有非确定性，不同的接受计算可能产生不同的匹配。我们的结果只要求每台机器某个 接受计算产生相关的交叉结构；一旦选定具体计算，便在全文中始终使用该计算。

两条弧 $(i,\ j) \in \mu_{M_1}(w)$ 和 $(i',\ j') \in \mu_{M_2}(w)$ 交叉，当且仅当 $i < i' < j < j'$或 $i' < i < j' < j$。（通过重新标记机器，任何交叉对都可以写成 $i < i' < j < j'$；我们在全文中采用这一约定。）这种交叉会在交错端点之间产生内段；形式化分解见第 6.3 节。

示例 2.1。 交错回文语言

$$L = \{w \in \{0,1\}^{2n} : w_1 w_3 \cdots w_{2n-1} \in \mathrm{Pal},\ w_2 w_4 \cdots w_{2n} \in \mathrm{Pal}\}$$

是上下文无关语言：定义成对符号 $X_i = (w_{2i-1},\ w_{2i})$，两个回文条件都要求 $X_i = X_{n-i+1}$，因此 L 是$\{00, 01, 10, 11\}$上的回文语言。两种正则性共享相同的匹配结构（位置 i 链接到 $n+1-i$），交叉间隙为 1，即相邻位置。成对字母表编码吸收了交叉。

3. 交叉泵敏感联结蕴含非上下文无关语言

3.1. 抽运敏感链接

我们的结果中的非上下文无关方向要求字符串片段满足耦合条件。通俗地说，如果两个字符串片段紧密耦合，以至于只修改其中一个而不修改另一个就会破坏语言成员关系，那么这两个片段就是抽运敏感链接的。不同于标准泵引理（它保证某种因式分解可以被抽运，且抽运后仍属于该语言），抽运敏感链接要求相反：每一种孤立抽运（任何触及一个链接片段但避开另一个的因式分解）都会使字符串不属于该语言。

定义 3.1. 设 $L \subseteq \Sigma^*$。在字符串 $w = P_1 P_2 \cdots P_k$ 上，L 中的一个抽运敏感链接 (P_i, P_j)（其中 $P_1, \ldots, P_k$ 是连续片段，且 $i < j$）具有以下性质：对于任意因式分解 $w = uvxyz$，若 $|vy| \geq 1$，且 vxy 与 P_i 相交但与 P_j 不相交，或者与 P_j 相交但与 P_i 不相交，则有 $uv^2xy^2z \notin L$。

通俗地说：在一个链接片段中的任意位置进行抽运，而不同时修改另一个片段，会使该字符串不再属于 L。一个熟悉的例子是，在语言 $\{a^n b^n\}$ 中，片段 a^n 和 b^n 是泵敏感链接的：任何只抽运 a 区域内字符的因式分解（而不触及 b 区域）都会破坏等式 $\#_a = \#_b$，反之亦然。该定义将这种紧密耦合从简单的计数约束推广到了任意语言成员条件。请注意，链接是一种输入条件，它根据语言的成员条件进行验证；这类似于在 Ogden 引理中选择标记位置，而不是根据已知该语言为非上下文无关语言而推导出的结论。对于块计数上下文无关语言，这直接源于相等约束（第 6.6 节）；对于一般上下文无关语言，能否始终仅凭弧几何推导出这种链接，是主要的开放问题（第 9 节，问题 1）。

备注 3.2（全称量化是有意为之的）。 定义要求，凡是只涉及某个关联片段的因式分解，都 无法泵，而不仅仅是存在这样一种因式分解。这个要求严格强于泵引理所提供的结论（泵引理只保证存在某种 可泵的因式分解）。这种强度是必要的：定理 3.3 通过证明泵引理所保证的因式分解必定属于七种情况之一来成立，而每种情况都恰好只涉及某个关联对中的一个成员。为了

使论证成立，所有 这类情况都必须失败；因此需要全称量化。

对于特定语言，要验证这一条件，就需要证明这些片段与其对应部分“紧密耦合”：无论局部修改位于何处，对一侧的任何局部修改都会破坏结构约束。对于带有交叉弧的按块计数上下文无关语言，这一条件在交集 $L_1 \cap L_2$ 上成立：任何只改变一个块的计数、却不改变其受约束对应部分计数的抽运，都会违反该交集的至少一个相等约束（见定理 3.6）。对于一般上下文无关语言，验证这一条件并非易事，也是一般二分结论具有条件性的根源。

3.2. 交叉连接定理

定理 3.3 (交叉连接的抽运不相容性)。 设 $L \subseteq \Sigma^*$ 是一种语言，对于无穷多个 n，其中包含一个字符串 w_n，该字符串具有一个满足以下条件的因式分解 $w_n = P_1P_2P_3P_4$：

(i) $|P_i| \geq n$ 对于每个 $i \in \{1, 2, 3, 4\}$。

(ii) L 具有一个泵敏感链接(P_1, P_3)。(iii) L 具有一个泵敏感链接(P_2, P_4)。

那么 L不是上下文无关语言。

图 1 展示了情况分析：有界长度的子串 vxy 无论位于何处，都恰好触及链接对中的一个成员，从而触发两个连接条件之一。

证明。反设 L 是上下文无关语言，其泵引长度为 p。选取 $n > p$，并考虑满足 $|P_i| \geq n > p$ 的 $w_n = P_1P_2P_3P_4$。

根据泵引理，存在一种因式分解 $w_n = uvxyz$，满足$|vxy| \leq p$、$|vy| \geq 1$，且对所有 $i \geq 0$ 都有 $uv^ixy^iz \in L$。由于 $|vxy| \leq p < n \leq |P_i|$，子串 vxy 至多与两个相邻片段相交。我们枚举所有可能的位置：

情况 *1*： $vxy \subseteq P_1$。于是 vxy 与 P_3 不相交。由 (ii)， $uv^2xy^2z \notin L$。矛盾。

情况 *2*： vxy 跨越 P_1P_2 边界。于是 vxy 与 P_3 和 P_4均不相交。由于 vxy 与 P_1 相交，却不与 P_3 相交，条件 (ii) 导致矛盾。

情况 *3*： $vxy \subseteq P_2$。 与 P_4 不相交。根据 (iii)，矛盾。

情况 *4*： vxy 跨越 P_2P_3 边界。 由于 $|P_2| > p \geq |vxy|$， vxy 的起点严格位于 P_2内部，因此 vxy 与 P_1不相交。由于 $|P_3| >$

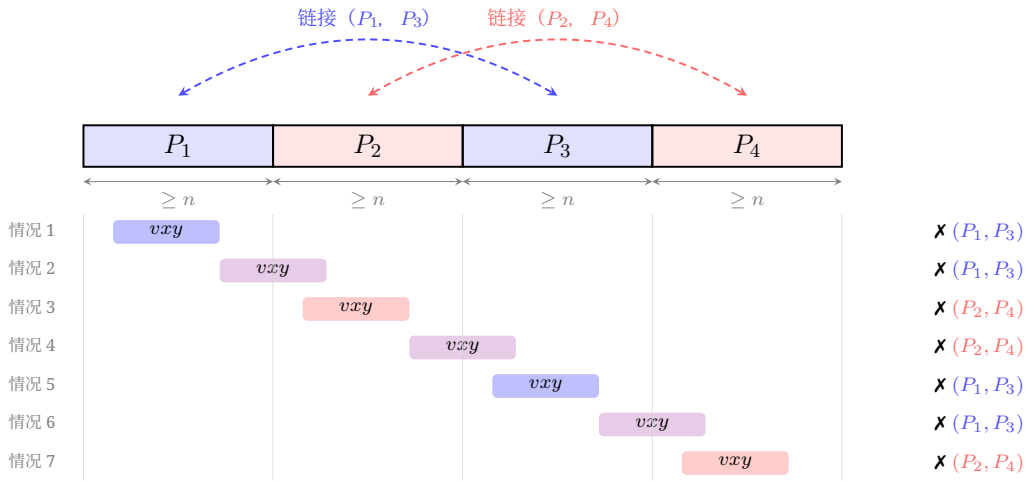


图 1：交叉连接的抽引障碍（定理 3.3）。栈顶：字符串$w_n = P_1P_2P_3P_4$ 具有交叉连接（P_1, P_3）和（P_2, P_4）。底部：由于 $|vxy| \leq p < n \leq |P_i|$，可抽引子串（每种情况均已示出）至多与两个相邻片段相交。每种放置都会触发连接矛盾（）：它恰好触及一对链接中的一个成员，而无法到达另一个。

$p \geq |vxy|$，vxy 的终点严格位于 P_3内部，因此 vxy 与 P_4不相交。因此 vxy 与 P_2 相交，但不与 P_4相交；由 (iii)，矛盾。

情形 *5–7* ($vxy \subseteq P_3$，跨越 P_3P_4，$vxy \subseteq P_4$)是对称的。所有情形都会导出矛盾。 □

3.3. 上下文无关语言交集的应用

推论3.4. 设 L_1、L_2 是上下文无关语言。假设对于无穷多个 n，存在一个字符串 $w_n \in L_1 \cap L_2$ 及其因式分解 $w_n = P_1P_2P_3P_4$，满足 $|P_i| \geq n$，并且 L_1 具有抽运敏感链接(P_1, P_3)，而 L_2 具有抽运敏感链接(P_2, P_4)。那么 $L_1 \cap L_2$ 不是上下文无关语言。

证明。由于 $L_1 \cap L_2 \subseteq L_1$，$L_1$ 的抽运敏感链接可由$L_1 \cap L_2$继承。L_2同理。应用定理 3.3。 □

3.4. 对已知示例的验证

示例 3.5 ($\{a^nb^nc^nd^n\}$)。设 $L_1 = \{w \in a^*b^*c^*d^* : \#_a(w) = \#_c(w)\}$，$L_2 = \{w \in a^*b^*c^*d^* : \#_b(w) = \#_d(w)\}$，二者都是上下文无关语言。令 $L = L_1 \cap L_2 = \{a^nb^nc^nd^n\}$。将 $a^nb^nc^nd^n$ 因式分解为 $P_1 = a^n$, $P_2 = b^n$, $P_3 = c^n$, $P_4 = d^n$，它们的大小均为 n。

我们验证 L 本身上的抽运敏感链接（而非分别在 L_1 或 L_2 上；见下文定理 3.6）。 L上的链接 $(P_1,\ P_3)$：任意满足 vxy 与 P_1 相交、但与 P_3 不相交的因式分解$w = uvxyz$，都有 vxy 包含在 a-b 区域内。由于 $|vy| \geq 1$，抽运会改变 $\#_a$ 或 $\#_b$ 中至少一个：如果 $\#_a$ 改变，那么 $\#_a \neq \#_c = n$ 违反了 L_1约束；如果 $\#_a$ 不变，那么某些 b 被抽运，从而得到 $\#_b \neq n = \#_d$，违反L_2约束。无论哪种情况，$uv^2xy^2z \notin L$。 L上的链接 $(P_2,\ P_4)$：对称地。由定理 3.3，L 不是上下文无关语言。

注记 3.6。上述链接在 $L = L_1 \cap L_2$ 上成立，而不只是单独在 L_1 上成立。在 $L_1 = \{w :\ \#_a = \#_c\}$ 上，链接 (P_1, P_3) 不成立：当因子分解跨越边界，且 $v = \epsilon$、$x = a \cdot b^s$、$y = b^t$ 满足条件时，vxy 与 P_1 相交，却与 P_3 不相交；然而抽引只会增加 b，因此 $\#_a$ =$\#_c$ 保持不变。交集的额外约束（$\#_b$ =$\#_d$）可防止这种规避。

这一细微之处意味着，要求环境语言 L_1、L_2 各自具有链接的定理 3.4 并不适用于此例。相反，我们直接对 L 应用定理 3.3。对于块计数上下文无关语言，交集上的链接总是成立：任何只改变一个块的计数而不改变其配对块计数的抽引，都会违反交集的一项约束。

例 3.7 ($\{a^nb^nc^n\}$)。取 $L_1 = \{a^nb^nw :\ w \in c^*\}$ 和 $L_2 = \{wb^nc^n :\ w \in a^*\}$，二者都是上下文无关语言。块计数弧分别是来自 L_1 的 $(1, 2)$ 和来自 L_2 的 (2,3)，它们共用端点 2；因此，$L_1 \cap L_2 = \{a^nb^nc^n\}$ 强制满足三方相等关系 $|B_1| = |B_2| = |B_3|$，而根据 Hopcroft 等人（2007）的标准抽引论证，该语言不是上下文无关语言。（这说明的是定理 4.3(iv) 中共用端点的机制，而非交叉弧；完整刻画见定理 4.4。）

3.5. 交叉定理涵盖的内容

定理 3.3 适用于任意一对上下文无关语言，只要它们的交集包含具有四个线性增长片段和交叉泵敏感联结的字符串，其中包括所有带有交叉弧的块计数上下文无关语言。定理 6.1（见第 6 节）将要求从四个片段都增长放宽为只需一个内侧片段无限增长。

定理 3.3 未涵盖有界间隙交叉。条件 $|P_i| \geq$$n$确保所有四个片段都随 n 增长。定理 5.4 证明，有界间隙交叉（$O(1)$）总是可以消解的。

4. 块计数上下文无关语言的完整刻画

4.1. 块计数上下文无关语言

定义 4.1。 定义在互不相交的子字母表 $\Sigma_1, \ldots, \Sigma_k$上的块计数上下文无关语言 是由满足 $B_i \in \Sigma_i^*$ 的字符串 $w = B_1 B_2 \cdots B_k$ 构成的语言，并受制于关于区块长度的相等约束 $C \subseteq \binom{[k]}{2}$：对每个 $(i, j) \in C$， $|B_i| = |B_j|$。每个约束 $(i, j) \in C$ 称为一条弧，从块 i 指向块 j。

4.2. 联合良好嵌套性

定义 4.2. 如果弧集 C在 $\{1, \ldots, k\}$ 上满足以下条件，则称其为良嵌套：其中任意两条弧都不 C交叉；也就是说，不存在满足 $i < i' < j < j'$ 的 (i, j),$(i', j') \in C$。

当 C 良嵌套时，弧集为 C 的块计数语言是上下文无关语言：对于每条弧，下推自动机在左端点块中压入，在右端点块中弹出，并由良嵌套性确保遵循后进先出规则。

定义 4.3. 如果两个弧集 C_1, C_2在 $\{1, \ldots, k\}$ 上满足以下条件，则称它们联合良嵌套：(i) C_1 良嵌套；(ii) C_2 良嵌套；(iii) C_1 中的弧不会与 C_2中的弧交叉：不存在满足 $i < i' < j < j'$ 或$i' < i < j' < j$的 $(i, j) \in C_1$ 和 $(i', j') \in C_2$；并且 (iv) 任意两条不同的弧 $(i, j) \in C_1$ 和 $(i', j') \in C_2$，若满足 $(i, j) \neq (i', j')$，则它们不共享端点（即 $\{i, j\} \cap \{i', j'\} = \emptyset$）。同时出现在 C_1 和 C_2 中的相同弧不违反联合良嵌套性。

4.3. 刻画定理

定理 4.4 （区块计数上下文无关语言交集的刻画）。 设L_1和 L_2 是具有相同区块结构 （Σ_1、.....、 Σ_k）且分别具有良嵌套弧集 C_1和C_2 的块计数上下文无关语言。则

$$L_1 \cap L_2 \text{ is context-free} \quad \Longleftrightarrow \quad C_1 \text{ and } C_2 \text{ are jointly well-nested.}$$

证明。（$\Leftarrow$: 良嵌套蕴含上下文无关语言。）假设 C_1 和 C_2 是联合良好嵌套的。构造一个识别 $L_1 \cap L_2$的下推自动机 M。栈字母表为$\Gamma = \{c_e : e \in C_1 \cup C_2\} \cup \{\$\}$。下推自动机通过互不相交的子字母表跟踪当前块索引。对于每个块 $B_\ell (\ell \in \{1, \ldots, k\})$：如果 ℓ是弧 $e = (\ell, r) \in C_1 \cup C_2$的左端点，则对读入的每个符号压入 c_e；如果 ℓ 是弧 $e = (s, \ell)$的右端点，则出栈并验证 c_e；如果 ℓ 是自由的（不是任何弧的端点），则直接读取，不执行栈操作。仅当栈中只剩下 $ 时接受。

联合良好嵌套性确保后进先出兼容性：到达弧 $e = (s,\ \ell)$的闭合块 B_ℓ 时，在 e 之后开启的所有弧要么嵌套其中（在 B_ℓ之前闭合），要么彼此不相交且位于其后（尚未开启）。没有弧与 e交叉或共享其端点。因此，需要时 c_e 标记位于栈顶。

（⇒：非良嵌套性蕴含非上下文无关语言。）如果联合良好嵌套性不成立，则要么（a）存在交叉弧，要么（b）不同的弧共享一个端点。

对于（a）：设（$i,\ j$）$\in C_1$与（$i',\ j'$）$\in C_2$交叉；不失一般性（如有必要，重新标记 $C_1,\ C_2$），$i < i' < j < j'$。反设 $L = L_1 \cap L_2$ 是上下文无关语言，其泵引长度为 p。取 $n > p$。如果一个块索引 ℓ 与交叉相连，则称其可通过 $C_1 \cup C_2$中的弧从 $\{i, i', j, j'\}$ 到达（将弧视为块索引上的无向边）。令 $w \in L$ 对每个与交叉相连的块 ℓ 具有 $|B_\ell| = n$，并令 $w \in L$ 对所有其他块具有 $|B_\ell| = 0$ 。所有约束均得到满足：两个与交叉相连的块之间的约束具有 $n = n$；两个不与交叉相连的块之间的约束具有 $0 = 0$；并且根据可达性的定义，没有弧连接与交叉相连的块和不与交叉相连的块。泵引理给出一个因式分解 $uvxyz$，其中 $|vxy| \le p$ 且 $|vy| \ge 1$。

由于交叉连通块的大小为 $n > p$，而所有其他块均为空，vxy 最多只会落在两个相邻的交叉连通块中。每个这样的块都通过一串弧与某个交叉端点 $\alpha \in \{i, i', j, j'\}$相连，其交叉配对端点 β满足 $|\alpha - \beta| \ge 2$（因为 $i < i' < j < j'$，其中 i' 位于 i 和 j之间，而 j 位于 i' 和 j'之间）；因此，β 与 α 之间至少隔着一个大小为 n的块，使二者在字符串中的距离为 $> p$。由于 $|vxy| \le p$，任何涉及交叉连通块的因式分解都无法到达该块的交叉配对端点。如果 vxy 完全位于单个块 B_ℓ内，那么泵送会改变 $|B_\ell|$，而配对块无法到达，这就违反了一项约束。如果 vxy 横跨两个相邻块 B_ℓ 和 $B_{\ell+1}$，泵送会改变至少一个块的长度；沿着该块的弧链追踪到某个交叉端点后，会找到一个距离很远的配对端点，其长度仍为 n，这同样违反了一项约束。因此，L是非上下文无关语言。

对于 (b)：来自 C_1 和 C_2 的不同弧共享一个端点：设为 $(i,\ j) \in C_1$ 和 $(i,\ j') \in C_2$，其中 $j \ne j'$。那么 $L_1 \cap L_2$ 强制满足 $|B_i| = |B_j|$ 和 $|B_i| = |B_{j'}|$，即 $|B_i| = |B_j| = |B_{j'}|$，这是不相交字母表上的三方相等关系，根据标准抽引论证可知它不是上下文无关语言（参见 Hopcroft 等人（2007））。（如果 C_1 和 C_2 共享同一条弧 $(i,\ j)$，则不会添加新的约束，而且共享弧不会妨碍其成为上下文无关语言。）

□

4.4. 讨论

定理 4.4 对块计数上下文无关语言给出了完整刻画。定理 6.1（见第 6 节）将这一刻画作为基于内段的分析的推论重新得出，因为在块计数情形中，所有交叉都具有增长的内段，并在 $L_1 \cap L_2$ 上带有抽运敏感链接（参见定理 3.6）。推广到一般上下文无关语言需要两个要素：一个与文法无关的“匹配弧”定义，以及一项证明，即非上下文无关交集总会在某个交叉对上产生抽运敏感链接。内段二分定理的上下文无关方向（定理 6.7）是无条件的；非上下文无关方向（定理 6.1）则以连结假设为条件。与线性上下文无关重写系统层级中的良嵌套性的联系将在第 7 节讨论。

5. 有界间隙交叉始终可消解

5.1. 定义

下推自动机范式约定。 在本节和第 6 节中，我们假设下推自动机处于一种标准形式：每次转换恰好读取一个输入符号，并且至多执行一次栈操作（压入单个符号或出栈）。任何下推自动机都可以转换为这种形式：首先将上下文无关文法转换为格雷巴赫范式（GNF），然后构造标准的格雷巴赫范式到下推自动机的转换。所得下推自动机每一步最多可压入两个符号（来自二元产生式）；我们通过辅助状态将每个此类步骤分解为一系列单次压入转换，这些辅助状态执行 ϵ-移动。

辅助状态会使 $|Q|$状态数增加一个依赖于文法的常数倍，但不会读取输入符号，因此每个输入位置 至多对应于两个单符号压入操作和一个出栈操作。因此，根据弧跨度（以输入位置衡量）确定的缓冲区界限，可以正确限制同时处于待处理状态的操作数量，其上限只需乘以 2 的常数因子。

回想一下，在第 2 节中， $\mu_M(w)$ 表示下推自动机 M 在字符串 w上的压栈-出栈匹配。

定义 5.1（交叉间隙）。 对于下推自动机 M_1, M_2 以及字符串 $w \in L(M_1) \cap L(M_2)$，交叉对 由来自不同机器的弧 $(i,\ j)$ 和 $(i',\ j')$ 组成（即一条来自 $\mu_{M_1}(w)$，另一条来自 $\mu_{M_2}(w)$），且满足 $i < i' < j < j'$。该交叉对的交叉间隙 为 $\max(i' - i,\ \ j' - j)$。

定义 5.2（有界交叉间隙）。下推自动机 M_1，M_2 具有 k-有界交叉间隙，如果对于每个 $w \in L(M_1) \cap L(M_2)$，每个交叉对的交叉间隙都至多为 k。

请注意，对于此构造，间隙的两侧都必须有界。若一对弧的左端点彼此接近（$i' - i = O(1)$），且内部重叠较小（$j - i' = O(1)$），但右端点相距很远（$j' - j = \omega(1)$），则其交叉间隙无界，无法仅凭定理 5.4 解决；不过，内部片段度量 $\max(i' - i,\ \ j - i')$ 仍然有界，而定理 6.7 表明此类交叉可以解决。

5.2. 不交叉弧集的良嵌套性

下面的位移缓冲区构造会将所有条目直接压入栈中，并在出栈时，暂时移开来自另一台机器的至多 $2k$ 个中间条目。下面这个初等引理（稍后在定理 6.7 中也会用到）说明，来自两台机器的非交叉弧可以共用一个栈。

引理 5.3（非交叉弧的后进先出兼容性）。设 S_1 和 S_2 是良嵌套弧集（同一集合内没有两条弧交叉）。如果 S_1 中没有弧与 S_2 中的任何弧交叉，那么 $S_1 \cup S_2$ 就是良嵌套的。

证明。反设 $(a, b) \in S_1 \cup S_2$ 与 $(c, d) \in S_1 \cup S_2$ 交叉，即 $a < c < b < d$。若二者都属于 S_1（或都属于 S_2），则与内部良嵌套性矛盾。若一个属于 S_1 而另一个属于 S_2，则与无交叉假设矛盾。因此 $S_1 \cup S_2$ 是良嵌套的。

□

由于任意良嵌套弧集都容许对其压入/弹出操作采用后进先出顺序，定理 5.3 保证来自两台机器的所有非交叉弧均可由单个栈处理而无需重排。

5.3. 有界间隙消解定理

定理 5.4（有界间隙消解）。若下推自动机 M_1、M_2 具有 k-有界交叉间隙，则 $L(M_1) \cap L(M_2)$ 是上下文无关的。

证明。我们构造一个非确定性下推自动机 M，用于识别 $L(M_1) \cap L(M_2)$。核心思路是将所有符号直接连同所有者标签一起压入栈中，并在出栈时使用有界的位移缓冲区，以跳过来自另一台机器的至多 $2k$ 个条目。

构造。 M 的有限控制由 (q_1, q_2, β) 组成，其中 $q_j \in Q_j$ 是 M_j 的当前状态， β 是一个最多容纳 $2k$ 个带标签符号的位移缓冲区。栈字母表由带标签的对 (j, γ) 组成，其中 $j \in \{1, 2\}$ 和 $\gamma \in \Gamma_1 \cup \Gamma_2$。

在每个输入位置， M 都会分别在各自的状态分量上模拟 M_1 和 M_2：

- 压入操作： 当 M_j 将 γ 入栈时，直接将 (j, γ) 压入 M 的栈中。
- 弹出操作： 当 M_j 想要弹出预期符号 γ 时：如果栈顶是 (j, γ)，就将其出栈。否则，将栈中的条目弹出到移位缓冲区 β 中（最多 $2k$ 个条目，且这些条目的所有者标签都不正确），直到找到 (j, γ)。弹出 (j, γ)，然后按照原始顺序将 β 中的所有条目压回栈中。如果在 $2k$ 个条目内未找到 (j, γ)，则此计算路径拒绝。

图2展示了位移机制。

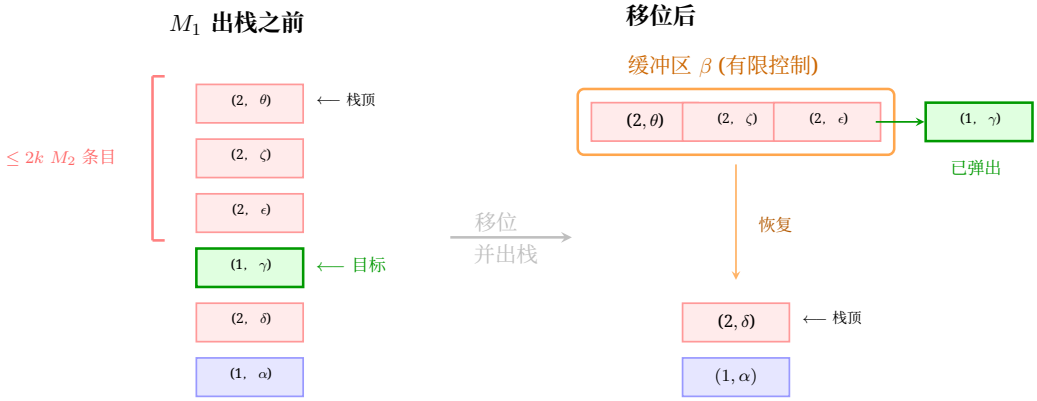


图2：位移缓冲区机制（定理 5.4）。左：组合栈交错存放来自 M_1 （蓝色）和 M_2 （红色）的条目。当 M_1 执行出栈操作时，其目标 $(1, \gamma)$（绿色）上方至多有 $2k$ 个交叉的 M_2 条目。右：M_2 个条目被移位到缓冲区 β （有限控制）中，目标被弹出，然后恢复被移位的条目。

位移缓冲区界。固定 $w \in L(M_1) \cap L(M_2)$，并考虑该接受计算。我们声称，当机器 M_ℓ （$\ell \in \{1, 2\}$）弹出弧 (i, j_{pop}) 时，来自另一台机器 $M_{3-\ell}$ 的至多 $2k$ 个条目位于栈上 M_ℓ 的目标条目之上。

在时刻 j_{pop}，来自 $M_{3-\ell}$ 的条目位于 M_ℓ 从位置 i 入栈的条目之上，当且仅当它是在位置 i 之后入栈且尚未出栈。这样的 $M_{3-\ell}$

弧 $(i',\ j')$ 满足 $i < i'$ 和 $j' > j_{\mathrm{pop}}$，这意味着 $i < i' < j_{\mathrm{pop}} < j'$，也就是说，它与 $(i,\ j_{\mathrm{pop}})$交叉 。根据 k-有界交叉间隙假设，$i' - i \le k$。由于每个位置在每台机器上至多贡献两次入栈操作（标准形式），因此此类条目至多有 $2k$ 个。

（来自嵌套在 i, j_{pop} 内的 $M_{3-\ell}$ 弧的条目，在 $i < i' < j' < j_{\mathrm{pop}}$ 的情况下，已在时间 j_{pop} 之前弹出操作。来自在 i 之前压入操作的 $M_{3-\ell}$ 弧的条目位于 M_ℓ 的目标下方。来自尚未压入操作的弧的条目不在栈上。）

栈顺序保持。 位移操作会弹出 $\le 2k$ 个条目，移除目标，并按原始顺序将移位的条目重新压入栈中。这保持了所有剩余栈条目的相对顺序。

完备性。 对于 $w \in L(M_1) \cap L(M_2)$，位移缓冲区中的条目数不会超过 $2k$ ，因此没有计算路径会因缓冲区溢出而拒绝。两个模拟机器都能找到预期符号并到达接受状态。

可靠性。 如果 M 在任何非确定性路径上接受 w，那么两个模拟机器都已接受 w （每次出栈操作都验证了正确的符号），因此 $w \in L(M_1) \cap L(M_2)$。

因此，$L(M) = L(M_1) \cap L(M_2)$ 是上下文无关的，具有 $|Q_1| \cdot |Q_2| \cdot (|\Gamma_1|+|\Gamma_2|+1)^{2k}$ 个状态。 □

5.4. 讨论

定理 5.4 解决了交叉弧刻画中的有界间隙情形：所有入栈操作都直接进入栈中，而在出栈时，来自另一台机器的至多 $2k$ 个条目会被移位并恢复。结合定理 3.3，两个已解决的情形分别是 $O(1)$ 间隙（上下文无关语言）和具有交叉连接的 $\Theta(n)$ 间隙（非上下文无关语言）。在下一节中，我们将引入内段度量$\max(|P_2|,\ |P_3|)$，并证明该量决定上下文无关性，从而缩小未解决范围。

交错回文（定理 2.1）的交叉间隙为 1。定理 5.4 与 $k = 1$ 构造了一种位移缓冲区大小为 2 的下推自动机：每次出栈时，来自另一台机器的条目最多有两个必须移位。定理 2.1 的成对字母表构造利用了交错回文语言的特定结构，是一种更直接的方法。

6. 内段二分定理

定理 3.3 证明，在大小为 Ω(n) 的四个片段上存在交叉泵敏感链接，意味着该语言不是上下文无关语言。定理 5.4 证明，交叉间隙为 $O(1)$ 时，总能进行下推自动机构造。现在，我们通过证明一个强化的交叉定理来缩小中间区域：该定理将“四个片段都很大”的条件替换为只对两个内段提出要求，并且只要求两个内段中有一个 足够大。

6.1. 强化交叉定理

定理 6.1（强化交叉链接定理）。 设 $L \subseteq \Sigma^*$ 对无穷多个 n 包含满足以下条件的字符串 $w_n = P_1P_2P_3P_4$：

（i） L 具有一个泵敏感链接（P_1， P_3）；

（ii） L 具有一个泵敏感链接（P_2， P_4）；（iii） $|P_2| \geq 1$、 $|P_3| \geq 1$，以及max（$|P_2|$， $|P_3|$） $\geq n$。

则 L 不是上下文无关的。

请注意这与定理 3.3 的区别：条件（iii）只要求两个内段中较大的那个 P_2 和 P_3 增长，而 P_1、 P_4 以及较小的内段可以是任意大小，包括 $O(1)$（但不能为零：要求 $|P_2| \geq 1$、 $|P_3| \geq 1$ 确保两个链接都非平凡；空片段上的链接总是可以平凡地满足，因此无法提供任何约束）。在交叉弧应用中， $|P_2| = i' - i \geq 1$ 和 $|P_3| = j - i' \geq 1$ 由 $i < i' < j$ 推出。

证明。 反设 L 是上下文无关的，其泵引长度为 p。取 $n > p$，并考虑满足 max($|P_2|$， $|P_3|$) $\geq n > p$ 的 $w_n = P_1P_2P_3P_4$。

根据泵引理，存在一种因式分解 $w_n = uvxyz$，使得$|vxy| \leq p$、 $|vy| \geq 1$，并且对所有 $i \geq 0$，都有 $uv^ixy^iz \in L$。我们分两种情况讨论。

情况 *A*： $|P_2| \geq n > p$。我们枚举 vxy 所有可能的位置：

情况 *1*： $vxy \subseteq P_1$。与 P_3 不相交（两者之间隔着长度为 $> p$ 的 P_2）。由 (i)，矛盾。

情况 *2*： vxy 跨越 P_1P_2 边界。由于 $|vxy| \leq p < |P_2|$， vxy无法越过 P_2，因此它与 P_3 和 P_4 不相交。它与 P_1 相交，但不与 P_3 相交；由 (i)，矛盾。

情况 *3:* $vxy \subseteq P_2$。与 P_4 不相交（由 P_3 隔开）。根据 (ii)，矛盾。

情况 *4:* vxy 跨越 P_2P_3 边界。由于 $|vxy| \leq p < |P_2|$，vxy 的起点严格位于 P_2内部，因此 vxy 与 P_1不相交。分两种子情况：如果$|P_3| \geq p$，那么 vxy 的终点严格位于 P_3内部，因此 vxy 与 P_4不相交；它与 P_2 相交，但不与 P_4相交，而 (ii) 导致矛盾。如果 $|P_3| < p$，那么vxy 可能穿过 P_3 延伸至 P_4；但它仍与 P_3 相交，同时与 P_1 不相交（因为 $|P_2| > p$），所以 (i) 导致矛盾。

情况 *5*：$vxy \subseteq P_3$。与 P_1 不相交（由长度为 $> p$ 的 P_2 隔开）。根据 (i)，矛盾。

情况 *6*：vxy 横跨 P_3P_4 边界。与 P_1 不相交（因为 $|P_2| > p$将 P_1 与 P_3隔开）。与 P_3 相交，但不与 P_1相交；根据 (i)，矛盾。

情况 *7*：$vxy \subseteq P_4$。与 P_2 不相交（由 P_3隔开）。根据 (ii)，矛盾。

情况 *B*：$|P_3| \geq n > p$（且 $|P_2|$ 可能很短）。情况 3 和 7 可从情况 A 沿用，其中 P_3（长度为 $> p$）将 P_2与 P_4隔开。情况 5 和 6 也可沿用，因为 vxy 从 P_3内部开始，因此无论 $|P_2|$如何，都与 P_1 不相交；根据 (i)，矛盾。不同的是情况 1、2 和 4，此时较短的 P_2 可能使 vxy 横跨多个片段：

情况 *1*：$vxy \subseteq P_1$。完全位于 P_1之内，因此与 P_3不相交。由(i)得出矛盾。

情况 *2*：vxy 跨越 P_1P_2 边界。若 $|P_2| \geq p$，则情况 A 的论证适用。若 $|P_2| < p$，则 vxy 可能延伸穿过整个 P_2 而进入 P_3。由于 $|P_3| > p \geq |vxy|$，vxy 无法到达 P_4。因此 vxy 与 P_2 相交，但与 P_4不相交；由(ii)得出矛盾。

情况 *4*：vxy 跨越 P_2P_3 边界。由于 $|P_3| > p \geq |vxy|$，vxy 与 P_4不相交。若 vxy 也与 P_1不相交，则它与 P_2 相交但不与 P_4相交；由(ii)得出矛盾。若 $|P_2| < p$ 且 vxy 向回延伸进入 P_1，则 vxy跨越 $P_1P_2P_3$，但仍与 P_4不相交；它与 P_2 相交但不与 P_4相交，因此(ii)再次导致矛盾。

在这两种情况下，vxy 的所有位置都会导致矛盾，因此 L不是上下文无关的。□

注记 6.2（跨越三个线段与链接的普遍性）。当一个内段（例如 P_2）较小时，会出现一个微妙的问题：泵引理的因式分解 vxy 可能从 P_1 跨越 P_2，延伸至 P_3。在这种配置下，vxy 同时与两者 P_1 和 P_3相交，因此链接 (i) 不适用。然而，vxy 也与 P_2 相交，同时与 P_4 不相交（因为

另一个内段 P_3 较大），因此链接 (ii) 可以解决这种情况。这就是为什么定理 3.1 中的全称量化至关重要：对于与某个链接片段相交的所有 因式分解，链接都必须成立，包括那些也触及其他片段的跨片段因式分解。像 $\{a^n bc^n d\}$这样的语言，其中泵 $a^k bc^k$ 后仍保持成员资格，按照这一全称定义不 满足链接 (P_2, P_4)，尽管单独泵单元素集合 b 确实会破坏成员资格。因此，这类语言不满足该定理的假设。

6.2. 上下文无关语言交集的应用

推论 6.3. 设 L_1, L_2 是上下文无关语言。假设对于无穷多个 n，存在字符串 $w_n \in L_1 \cap L_2$ 及其因式分解 $w_n = P_1P_2P_3P_4$，满足 $|P_2| \geq 1$, $|P_3| \geq 1$, 且 $\max(|P_2|, |P_3|) \geq n$，使得 L_1 具有抽运敏感链接 (P_1, P_3)，而 L_2 具有抽运敏感链接 (P_2, P_4)。那么 $L_1 \cap L_2$ 不是上下文无关语言。

证明。 抽运敏感链接会由 $L_1 \cap L_2 \subseteq L_i$ 继承。应用定理 6.1。

□

6.3. 从交叉几何角度解读

在交叉间隙设置中，弧 $(i, j) \in \mu_{M_1}(w)$ 和 $(i', j') \in \mu_{M_2}(w)$ 满足 $i < i' < j < j'$，由此在交叉端点之间定义出三个片段：左内片段 $w[i{+}1..i']$（长度为 $i' - i$）、右内段$w[i'{+}1..j]$（长度为 $j - i'$）以及右间隙 $w[j{+}1..j']$（长度为 $j' - j$）。该字符串还有一个前缀 $w[1..i]$ 和一个后缀 $w[j'{+}1..|w|]$。应用抽象因式分解定理（定理 3.3 和 6.1；若链接成立于环境语言上，则应用推论 6.3）时，我们令 $P_1 = w[1..i]$（前缀）、$P_2 = w[i{+}1..i']$（左内片段）、$P_3 = w[i'{+}1..j]$（右内段）以及 $P_4 = w[j{+}1..|w|]$（右间隙与后缀的组合），从而得到完整因式分解 $w = P_1P_2P_3P_4$。决定二分情形的内段 P_2 和 P_3，恰好对应于交错交叉端点之间的几何片段。

交叉端点上有两个自然的度量：

- 交叉间隙：$\max(i' - i,\ j' - j)$，即推端点距离与弹出端点距离中的较大者。
- 内部片段度量：$\max(|P_2|,\ |P_3|) = \max(i' - i,\ \ j - i')$，即内部线段长度的最大值。

这些度量不同，是因为右间隙 $j' - j$ 独立于内段。内部片段度量决定弧的跨度（$i,\ j$），我们称之为左起始弧（其压入端点 i先于另一条弧的压入端点 i'）。它的跨度为 $j - i = (i' - i) + (j - i') = |P_2| + |P_3|$。当 $\max(|P_2|,\ |P_3|) \le D$ 时，这条弧的跨度至多为 $2D$，因此它是一条“短弧”。

定理 6.1 表明：如果对于某个具有抽运敏感链接的交叉对族，$\max(|P_2|,\ |P_3|) \to \infty$，那么 $L_1 \cap L_2$ 不是上下文无关语言。此前的定理 3.3 要求所有四个片段都满足 $|P_i| \ge n$；而只要求一个内段增长，这一加强条件严格地更为一般。

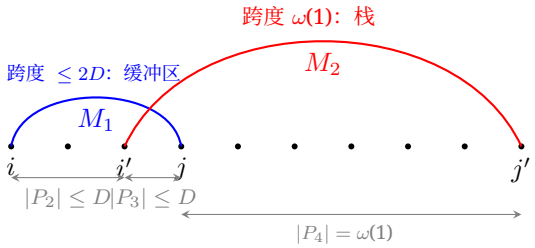
6.4. 内部线段二分法

定义 6.4（内部片段度量）。对于来自不同机器的一对弧的交叉对，记作 $i < i' < j < j'$，其内部片段度量为 $\max(|P_2|,\ |P_3|) = \max(i' - i,\ \ j - i')$。

当内部片段度量 $\max(i' - i,\ \ j - i') = O(1)$，而交叉间隙 $\max(i' - i,\ \ j' - j) = \omega(1)$ 时，我们非正式地称该交叉为近切线：两个内端点 i' 和 j 都距离左起始弧的端点不远，但右起始弧在 j' 处的出栈位置却很远。无界间隙完全源于右间隙 $j' - j$，而两个内段都保持有界。这正是定理 6.7 所解决的内部线段有界、间隙无界的情形。

图 3 展示了内段二分定理的两种情况。

有界内段：上下文无关语言



增长的内段：非上下文无关语言

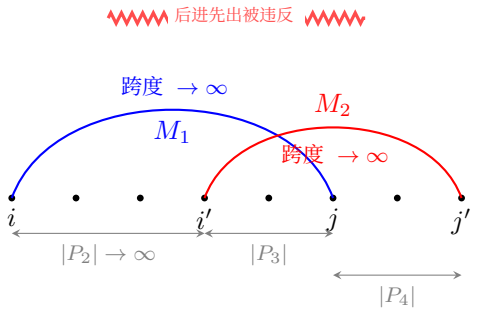


图 3：内段二分定理。上：有界内部片段（$\max(|P_2|, |P_3|) \leq D$）；左起始交叉弧的跨度为 $\leq 2D$，并被纳入有限缓冲区（定理 6.7）。下：不断增长的内段（$\max(|P_2|, |P_3|) \to \infty$）；左起始弧无限增长，泵引理中的子串 vxy 无法跨越两个关联片段，因此该语言不是上下文无关语言（定理 6.1）。

完整分类如下：

交叉类型	条件	交集	参考
无交叉	–	CFL	（乘积构造）
有界间隙	最大值$(i'-i,\ j'-j) = O(1)$	CFL	定理 5.4
有界内部片段	$\max(\lvert P_2\rvert,\ \lvert P_3\rvert) = O(1)$，间隙 $\omega(1)$	CFL	定理 6.7
增长的内部分段†	$\max(\lvert P_2\rvert,\ \lvert P_3\rvert) \to \infty$，间隙 $\omega(1)$	非上下文无关语言	定理 6.1

†要求交叉分段上的抽运敏感链接。对于块计数上下文无关语言，这些链接在 $L_1 \cap L_2$ 上总是成立。对于一般上下文无关语言，请参见定理 6.9。间隙限制 $\omega(1)$ 可由有界间隙行推出：有界间隙总能得到上下文无关语言，因此只有当间隙也无界时，连接条件才可能满足。

该表揭示了一个此前未被注意到的情形：内段有界，但交叉间隙无界，而交集仍然是上下文无关语言（定理 6.7）。剩下的开放问题涉及非上下文无关方向：内段不断增长却没有抽运敏感链接，这种情况是否可能出现？对于块计数上下文无关语言，链接是自动存在的；对于一般上下文无关语言，这是主要的开放问题（定理 6.9）。

6.5. 尖锐间隙猜想的反驳

有人可能会猜想，仅交叉间隙 $\omega(1)$ 就意味着非上下文无关。内段二分定理将其归结为一个几何问题：交叉间隙 $\omega(1)$ 是否必然意味着存在某个交叉对，使得 $\max(|P_2|, |P_3|) = \omega(1)$？我们证明答案是否。

命题 6.5 (对尖锐间隙猜想的反驳)。存在上下文无关语言 L_1, L_2，其下推自动机为 M_1, M_2，满足：

(i) $L(M_1) \cap L(M_2)$ 是上下文无关的；(ii) 对于无穷多个 n，存在交叉间隙超过 n 的交叉弧；(iii) 对于所有交叉对，内部片段度量满足 $\max(|P_2|, |P_3|) = O(1)$。

证明。令 $\Sigma = \{a, b, d, e, f, g, h\}$。定义：

$$L_1 = \{a\, b\, a\, w_1\, f\, g^k h^k : w_1 \in \{d, e\}^*,\ k \geq 0\},$$
$$L_2 = \{a\, b\, a\, d^n e^n f\, w_2 : w_2 \in \{g, h\}^*,\ n \geq 0\}.$$

两者都是上下文无关的。M_1 在位置 1 入栈（读取 a），在位置 3 出栈（读取第二个 a），自由读取 d、e、f，然后压入 g 并弹出 h。M_2 在位置 2 入栈（读取 b），压入 d，弹出 e（嵌套在b入栈操作之内），在位置 $3{+}2n{+}1$ 出栈（读取 f），然后自由读取 g 和 h。

交集。 $L_1 \cap L_2 = \{a\, b\, a\, d^n e^n f\, g^k h^k :\ n, k \geq 0\}$，它是上下文无关的：一个下推自动机以有限状态读取前缀 aba，压入 d 并弹出 e，读取 f，然后压入 g 并弹出 h。

交叉结构。M_1的弧（1,3）与 M_2的弧（2, $3{+}2n{+}1$）满足 $1 < 2 < 3 < 3{+}2n{+}1$，得到一个交叉对，其内段为 $|P_2| = i' - i = 2 - 1 = 1$ 和 $|P_3| = j - i' = 3 - 2 = 1$，右间隙为 $j' - j = 2n{+}1$。内部片段度量为$\max(|P_2|, |P_3|) = 1$：M_1的入栈-出栈对对M_2的弧的“中断”在每个方向上恰好持续一个位置。

同时，交叉间隙$\max(i'-i,\ j'-j) = \max(1, 2n{+}1) \to \infty$。不存在其他交叉对： M_1 的长弧（位于 g-h区域）与 M_2的弧（位于位置 $\le 3{+}2n{+}1$）占据互不相交的位置范围。

因此， $L_1 \cap L_2$ 是上下文无关语言，交叉间隙为 $\omega(1)$，且所有交叉对的内部片段度量均为 $O(1)$。 □

这是因为 M_1的交叉弧具有有界跨度（$j - i = 2$）：它只是对 M_2的长弧的一次短暂中断。这类短弧可以吸收到有限缓冲区中；它们表现得像有限状态条件。两台机器的长非交叉弧都是良嵌套的，并且可以无冲突地共享栈。我们将在下面的定理 6.7 中将这一设想形式化。

注 6.6。 该反证区分了两个看似等价的度量。交叉间隙 $\max(i'-i,\ j'-j)$ 无界增长，而内部片段度量 $\max(|P_2|, |P_3|)$ 保持有界：尽管一条弧与一条“长”弧（无界跨度）交叉，但它本身是“短”弧（有界跨度）。上下文无关性取决于内部片段度量，而不是交叉间隙。

6.6. 块计数交叉总是具有增长的内段

对于具有交叉弧的块计数上下文无关语言，右内段始终为 $\Theta(n)$：每条弧连接块 B_i 与 B_j，并带有 $|B_i| = |B_j| = \Theta(n)$，且任意交叉对在内端点之间至少有一个完整块。因此，块计数设定中的所有交叉均有$\max(|P_2|, |P_3|) = \Theta(n)$。此外，对抽引敏感的连接在交集 $L_1 \cap L_2$上成立：在交叉弧几何中，每个片段 P_i 都位于受某个等式约束的块之内，因此任何泵若修改某一链接片段中的字符而不修改其配对片段，就会改变受约束块计数，同时使其配对片段的计数保持不变，从而违反交集的某一约束（参见定理 3.6）。于是，定理 6.1解决了整个块计数情形，并将定理 4.4作为基于内段的分析的推论加以恢复。

6.7. 有界内段消解定理

当所有交叉对都满足 $\max(|P_2|, |P_3|) = O(1)$ 时，每对中的左起始弧跨度为 $O(1)$。其几何意义如下：内部片段度量刻画了两条交叉弧的交错端点之间较长内段的长度。左起始弧（即 i 是四个端点中最小的端点）(i, j) 的入栈位置 i 与出栈位置 j 相距至多 $2D$ 个位置；在此期间，另一台机器可以入栈和出栈，但这种“中断”是有界的。这类短弧只施加

局部约束，类似于有限状态条件，可由有限控制中的有界缓冲区吸收，而所有长弧都直接进入栈（不同机器的长弧不会交叉，因为任何交叉都会迫使其中一条弧变短）。缓冲能力取决于中断持续时间，而不是两台机器的弧之间的整体距离。

定理 6.7 (有界内部片段分辨)。设 M_1, M_2 为下推自动机，其状态集分别为 Q_1, Q_2，栈字母表分别为 Γ_1, Γ_2。若存在常数 D，使得对每个 $w \in L(M_1) \cap L(M_2)$ 以及每个满足 $i < i' < j < j'$ 的交叉对，内部片段度量均满足 $\max(i' - i,\ j - i') \le D$，那么 $L(M_1) \cap L(M_2)$ 可由一台至多具有以下数量状态的非确定性下推自动机识别

$$|Q_1| \cdot |Q_2| \cdot \big(1 + (|\Gamma_1| + |\Gamma_2|) \cdot 2D\big)^{8D}$$

。特别地，$L(M_1) \cap L(M_2)$ 是上下文无关的，其状态复杂度关于 D 呈指数增长，但与输入长度无关。

证明。我们构造一台非确定性下推自动机 M，用于识别 $L(M_1) \cap L(M_2)$。证明分为两个阶段：首先，我们定义一种预言机策略，利用对接受计算的完整了解，将每条弧分类为“短”或“长”；然后，我们证明非确定性下推自动机可以猜测这种分类，并正确模拟两台机器。

阶段 *1*：预言机分类。固定 $w \in L(M_1) \cap L(M_2)$，并固定 M_1 和 M_2 在 w 上的接受计算，从而确定压入-弹出匹配 $\mu_{M_1}(w)$ 和 $\mu_{M_2}(w)$。若弧 $(i, j) \in \mu_{M_k}(w)$ 满足 $j - i \le 2D$，则称其为短弧；否则称其为长弧。（这种分类仅取决于弧自身的端点。阶段 2 将证明，非确定性下推自动机可以猜测这种分类；预言机仅用于确定存在正确的猜测。）

断言 *A*：每一对交叉对中，至少有一条弧是短弧。假设来自不同机器的弧 $(i,\ j)$ 和 $(i',\ j')$ 在 $i < i' < j < j'$ 处交叉。根据假设，$\max(i' - i,\ j - i') \le D$，因此 $j - i = (i' - i) + (j - i') \le 2D$。也就是说，弧 $(i,\ j)$ 是短弧。（另一条弧 $(i',\ j')$ 的跨度可能无界；$j' - i'$ 不受 D 的约束。）

断言 *B*：长弧是联合良嵌套的。每台机器的弧各自都是良嵌套的。我们证明，没有任何来自 M_1 的长弧会与来自 M_2 的长弧交叉。若弧 $(a, b) \in \mu_{M_1}(w)$ 和 $(c, d) \in \mu_{M_2}(w)$ 交叉，则断言 A 保证其中一条的跨度为 $\le 2D$，也就是说，其中一条是短弧。反过来说，如果两条都是长弧（跨度为 $> 2D$），它们就不会交叉。根据定理 5.3，来自两台机器的长弧构成一个联合良嵌套的集合。

阶段 2：下推自动机构造。 M 的有限控制由状态对 $(q_1, q_2) \in Q_1 \times Q_2$ 和一个缓冲区 B 组成；该缓冲区最多可容纳 $8D$ 个待处理条目，每个条目记录来自 $\Gamma_1 \cup \Gamma_2$ 的一个栈符号以及位于 $\{1, \ldots, 2D\}$ 的一个倒计时计时器。缓冲区为状态数贡献一个有限因子：至多为 $(1 + (|\Gamma_1| + |\Gamma_2|) \cdot 2D)^{8D}$，因为 $8D$ 个槽位中的每一个要么为空，要么存放一个符号—计时器对。

M 的栈字母表由带标签的对（γ，所有者）组成，其中 $\gamma \in \Gamma_1 \cup \Gamma_2$ 和所有者 $\in \{1, 2\}$。

在每个输入位置， M 分别在各自的状态分量上模拟 M_1 和 M_2。当某台被模拟的机器执行入栈操作时， M 非确定性地猜测该弧是短弧还是长弧：

- 长弧猜测：将符号压入 M 的栈中，并标记其所有者。
- 短弧猜测：将符号放入缓冲区，并设置倒计时 $2D$。

每一步中，所有倒计时计时器都会递减。若缓冲区条目的计时器降至 0 时该条目仍未出栈，则此计算路径将被拒绝。

当某台被模拟的机器执行出栈操作时：若缓冲区中有预期符号，则在那里解决；若栈顶具有正确标签和符号，则将其出栈；否则，拒绝此路径。

完备性（正确的猜测会被接受）。给定 $w \in L(M_1) \cap L(M_2)$，第 1 阶段的预言机分类会确定一个特定的猜测序列。在这条路径上：

- 每条短弧（跨度为 $\leq 2D$）都会被缓冲，并在计时器到期前得到处理。
- 每条长弧都会被压入栈中。根据断言 B，来自不同机器的长弧是联合良嵌套的，因此后进先出规则能够正确处理它们：嵌套弧在包围它的弧之后入栈、之前出栈，从而保持栈顺序。所有者标签可确定每个符号属于哪台机器。
- 每台机器同时处于待处理状态的短弧最多有 $4D$ 条（每条跨度为 $\leq 2D$ 个位置，每台机器每个位置最多有两次压入操作），因此缓冲区不会超过 $8D$ 个条目。

两个模拟机器都到达接受状态，且缓冲区为空，因此M 接受。

可靠性。如果 M 在任意非确定性路径上接受了 w，那么两个模拟机器都接受了 w，因此 $w \in L(M_1) \cap L(M_2)$。

因此，$L(M_1) \cap L(M_2) = L(M)$ 是上下文无关的。 □

结合定理 6.1 和定理 6.7：

推论 6.8 (内段二分定理)。设 M_1，M_2 为下推自动机。设 $L = L(M_1) \cap L(M_2)$。

(a) 如果存在常数 D，使得每个 $w \in L$上的每个交叉对的内段度量均小于 $\max(|P_2|, |P_3|) \leq D$，那么 L 是上下文无关的(定理 6.7)。(b) 如果对于无穷多个 n，某个 $w_n \in L$ 具有一对带有泵敏感链接的交叉对，且$\max(|P_2|, |P_3|) \geq n$，那么 L 不是上下文无关的(定理 6.1)。

我们称之为“二分定理”，因为这两个方向指出了起决定作用的量（内段度量，而非交叉间隙）。对于块计数上下文无关语言，这一术语完全准确，因为条件(a) 和(b) 穷尽了所有可能。对于一般上下文无关语言，这一说法则是有条件的：(a) 与(b)之间的差距，涉及不断增长的内段是否总是带有泵敏感链接（见定理 *6.9*）。

备注 6.9（穷尽性）。对于块计数上下文无关语言，条件 (a) 和 (b) 是穷尽的：刚性的区块结构保证，每个内段不断增长的交叉对都带有抽运敏感链接（第 6.6 节），因此二分情形是完整的，并由此重新得到定理 4.4。

对于一般上下文无关语言，穷尽性尚无定论。缺口在于这样一种情形：内段不断增长，但没有交叉对带有抽运敏感链接；L_1 或 L_2 中的其他接受计算可能吸收原本会离开 L的泵。方向 (a) 无条件成立，且不需要对链接作任何假设。方向 (b) 则要求由 M_1，M_2的匹配所诱导的特定片段分解，能在 $L = L_1 \cap L_2$上产生链接。非上下文无关交集是否总能得到这样的分解，等价地说，是否存在某些下推自动机，其交叉结构能通过抽运敏感链接见证其非上下文无关性，这仍是一个开放问题（第 9 节）。

我们注意到，目前尚无已知的非上下文无关交集例子缺乏抽运敏感链接。文献中的每个非上下文无关交集（块计数语言、Shieber 的瑞士德语构造，

位置匹配语言）都在其交叉分段上呈现抽运敏感链接。反例需要两个上下文无关语言，它们的交集是非上下文无关语言，但对于每个内部分段不断增长的交叉对，都存在一种只涉及一个链接分段的因式分解，仍能在 $L_1 \cap L_2$ 中抽运。这意味着交叉分段有足够的“余量”，可以吸收局部修改，尽管整体语言仍是非上下文无关语言。构造这样的例子似乎很困难：大多数自然的上下文无关语言交集构造都利用相等约束，形成紧密的分段耦合。

6.8. 与有界间隙构造的比较

有界间隙构造（定理 5.4）会将所有符号直接压入栈中，并在出栈时移位至多 $2k$ 个条目。当交叉间隙有界时，移位数量也保持有界。当交叉间隙无界时，移位数量可能无限增长，超过任何有限缓冲区。定理 6.7 采用了另一种策略：它不在出栈时移位，而是在压入时将每条弧分类为短弧（跨度为 $\leq 2D$）或长弧。短弧（根据断言 A，每个交叉对中的左起始弧都属于短弧）完全缓存在有限控制中；长弧则压入栈中。断言 B 保证，压入栈中的长弧联合良嵌套，因为任何交叉对都至少包含一条短弧。非确定性下推自动机无需预先观察后续输入，即可猜测短弧／长弧的分类。

7. 相关工作

7.1. 自动机交集的状态复杂度

Yu、Zhuang 和 Salomaa Yu 等（1994）证明，状态数分别为 m 和 n 的两个确定性有限自动机的交集至多需要 mn 个状态，且在最坏情况下这一上界是紧的。对于一元字母表，当m 和 n 互素时，状态复杂度为 lcm(m, n)，这与中国剩余定理有关。Dutta 和 Saikia Dutta 和 Saikia（2022）也将中国剩余定理与有限自动机乘积构造联系起来。我们的定理 6.7 将这项工作拓展到了下推情形，在该情形下，交集可能完全不属于上下文无关语言类。

7.2. 下推自动机的交集

两个上下文无关语言的交集非空问题是不可判定的（Hopcroft 等人（2007））。Swernofsky 和 Wehar（2015）为判定涉及正则语言和上下文无关语言的交集是否非空建立了紧确的复杂度界

，并将这些问题与图论困难性联系起来。我们的结果通过刻画两个上下文无关语言交集的语言类，补充了这些可判定性和复杂度界：我们找出了交叉弧上的几何条件，这些条件决定了交集是否仍为上下文无关语言。

7.3. 温和上下文敏感语言与线性上下文无关重写系统

Vijay-Shanker、Weir 和 Joshi Vijay-Shanker 等（1987）提出了线性上下文无关重写系统（LCFRS），它通过允许非终结符推导字符串元组来推广上下文无关文法。扇出为 k的 LCFRS 恰好生成 k-多重上下文无关语言（k-MCFL）；扇出为 1 时得到上下文无关语言。Seki、Matsumura、Fujii 和 Kasami Seki 等（1991）独立提出了多重上下文无关文法（MCFG），其与 LCFRS 等价。

许多两个上下文无关语言的非上下文无关语言交集落在 LCFRS(2) 中。例如，$\{a^n b^n c^n d^n\}$ 属于 LCFRS(2)，但不是上下文无关语言；它作为两个带有交叉弧的上下文无关语言的交集而产生。具有有界数量独立交叉族的两个上下文无关语言的交集是否总是落在有界 k 的 LCFRS(k) 中，仍是一个未决问题（第 9 节，问题7）。

我们的内部片段度量与线性上下文无关重写系统扇出度有着自然的联系。当 $\max(|P_2|, |P_3|) \leq D$ 时，每个交叉对中的左起始弧都有有界跨度（$\leq 2D$），而缓冲区构造（定理 6.7）表明，该交集是上下文无关语言（扇出度为 1）。随着内段增长，每个交叉对都会增加一条依存关系“线索”，这表明非上下文无关交集的扇出度为 2。

7.4. 交叉依存关系与良嵌套性

Shieber Shieber（1985年）通过将瑞士德语中的交叉依存关系归约为 $\{a^m b^n c^m d^n\}$，证明瑞士德语不是上下文无关语言。Joshi Joshi（1990年）指出，交叉依存关系需要温和上下文有关形式体系。Bodirsky、Kuhlmann 和 Möhl Bodirsky 等（2005年）将良嵌套性形式化，而 Kuhlmann Kuhlmann（2013年）则在LCFRS框架中建立了一个层级。我们的定理 4.4 可视为一种交集理论类比：他们的研究刻画了哪些单个文法仍属于上下文无关语言类，而我们的结果则刻画了哪些成对的上下文无关语言约束在取交集后仍属于上下文无关语言类。

8. 讨论

8.1. 在乔姆斯基层级中的位置

内段二分定理将上下文无关语言交集置于一个分级层级中。对于有限自动机，两个正则语言的交集始终是正则语言，最坏情况下需要 mn 个乘积状态 Yu 等（1994年）。对于下推自动机，我们的结果表明，交集可能完全超出上下文无关语言类。结合LCFRS框架 Vijay-Shanker 等（1987年）；Seki 等（1991年）：

扇出	类别	交行为	参考资料
–	正则	始终封闭；mn乘积状态（一元情形：最小公倍数）	Yu 等（1994）
1	CFL	在有界间隙、有界内部片段或联合良嵌套的情况下封闭‡	定理 4.4、5.4 和 6.7
2	LCFRS(2)	带有交叉弧的非上下文无关语言（推测的上界）	定理 6.1；问题 7

‡对于块计数上下文无关语言，双向条件成立（定理 4.4）；对于一般上下文无关语言，仅“若”方向成立（非上下文无关方向需要对抽引敏感的连接；见定理 6.9）。

对于 $k \geq 3$ 上下文无关语言，情况在性质上截然不同。任意有限个上下文无关语言的交集仍然是上下文有关语言（因此可判定），但每个递归可枚举语言都是两个线性上下文无关语言之交集的同态像 Baker 和 Book（1974）：引入完整图灵计算能力的是同态，而非交集本身。

内段二分定理（定理 6.8）刻画了跨越上下文无关语言／线性上下文无关重写系统（2）边界的一个充分条件：不断增长的内段（带有抽运敏感链接）会使交集进入 LCFRS(2) 或更高层级；在该层级中，单个栈已不足以满足需要，必须使用多条依赖链。

8.2. 几何与复杂度

非上下文无关语言的相关结果涉及的是内存访问规则，而非内存量。对于我们示例中处于核心地位的块计数上下文无关语言，各个语言及其非上下文无关交集（例如 $\{a^n b^n c^n d^n\}$）都可在图灵机上以 $O(\log n)$ 空间通过计数并比较区块长度来验证。区别在于，单个上下文无关语言的正则性可放入栈中（后进先出访问），而联合正则性则需要非后进先出访问。内部片段度量恰好刻画了这一点：当某个内段无界增长时，两台机器的栈操作会以任何单个栈都无法模拟的方式交错。

9. 开放问题

1.**对抽引敏感的连接的必要性**。 内段二分法（定理 6.8）对块计数上下文无关语言是完备的，但对一般上下文无关语言是有条件的：方向 (b) 要求交叉分段上存在对抽引敏感的连接。若 $L_1 \cap L_2$ 不是上下文无关的，且具有内段不断增长的交叉弧，是否必定存在某个针对 (L_1, L_2) 的下推自动机对 (M_1, M_2)，使其产生在 $L_1 \cap L_2$ 上携带对抽引敏感的连接的交叉分段？肯定的回答将使该二分法无条件完备。否定的回答则会展示一种非上下文无关交集，其中任何由机器诱导的分解都不产生该链接条件，从而需要一种根本不同的证明技术（例如 Ogden 引理或交替引理）来处理非上下文无关方向。

2. **与文法无关的匹配弧**。匹配弧 $\mu_M(w)$ 取决于具体的下推自动机 M 及其接受计算，而不仅仅取决于语言。对于同一个上下文无关语言，不同的下推自动机可能产生不同的匹配和不同的交叉结构。可见下推语言（VPL）通过固定划分 $\Sigma = \Sigma_c \cup \Sigma_r \cup \Sigma_\ell$ 提供了一种自然的、与文法无关的匹配弧概念。不同划分下的可见下推语言（Σ_c，Σ_r，Σ_ℓ）与（Σ'_c，Σ'_r，Σ'_ℓ）的交理论是什么？这或许是实现完全与文法无关的刻画最可行的路径，也能解决可见下推语言情形下的问题 1（因为可见下推语言的匹配由输入字符串决定）。

3. **下推自动机转换下交叉度量的不变性**。定理 6.7 中的内段界 D 取决于具体的下推自动机M_1，M_2 及其乔姆斯基范式，而不仅仅取决于 L_1 和 L_2。性质“存在下推自动机 M_1 用于 L_1，以及下推自动机 M_2 用于 L_2，使得所有交叉对的 $\max(|P_2|, |P_3|)$ 均为 $\leq D$”是否等价于“$L_1 \cap L_2$ 是上下文无关语言”？更一般地，对于（L_1，L_2），每一对下推自动机是否都会得出相同的内段有界／无界分类，还是说不同的下推自动机选择会定性地改变交叉结构？

4. **缓冲区构造的状态复杂度**。定理 6.7 构造了一个具有 $(1+(|\Gamma_1|+|\Gamma_2|)\cdot 2D)^{8D} \cdot |Q_1| \cdot |Q_2|$ 个状态的下推自动机，其状态数关于 D 呈指数增长。这种状态数膨胀是否不可避免，还是可以找到多项式状态下推自动机？更具体地说，对于某些下推自动机族而言，这个界是否是紧的？底数或指数能否降低？

5. **上下文无关语言交集的描述复杂度**。是否存在这样的情况 s

两个上下文无关语言的交集也是上下文无关语言，但其文法所需的规模（非终结符或规则更多）严格大于任一组成语言的文法？这将为上下文无关语言类内部提供一种定量的复杂度度量。

6. **索引语言**。某些下推自动机交集问题可能会产生索引语言层级上的语言 Aho（1968年），而非上下文无关语言或线性上下文无关重写系统层级上的语言。哪些交叉结构会产生索引语言、哪些会产生线性上下文无关重写系统语言，仍是一个开放问题。

7. **上下文无关语言交集的线性上下文无关重写系统扇出度**。当 $L_1 \cap L_2$ 不是上下文无关语言时，通常可以用扇出度为 2 的线性上下文无关重写系统识别 Seki 等（1991年）；Vijay-Shanker 等（1987年）。对于两个上下文无关语言的交集，情况是否总是如此？更准确地说，匹配图中独立交叉族的数量是否决定线性上下文无关重写系统的扇出度？对于恰好两个上下文无关语言的交集，该扇出度是否总是有界？

8. **有界内部片段的可判定性**。给定两个上下文无关文法 G_1, G_2，是否可以判定是否存在下推自动机 M_1, M_2，使得对于$L(G_1)$ $L(G_2)$，所有交叉对的内部片段度量都有界？由于上下文无关语言交集空性通常不可判定，答案很可能是否定的；不过，证明或归约将有助于明确内段二分定理的计算边界。

致谢

本研究部分经费来自国家资金，通过科学技术基金会（FCT）提供，资助项目为 UIDB/00127/2020（https://doi.org/10.54499/UIDB/00127/2020）。

生成式人工智能使用声明

在准备本研究期间，作者使用 Claude（Anthropic）协助起草文本并完善数学论证的结构。作者已手动核查所有证明，并对出版内容承担全部责任。

作者贡献声明

Jorge Miguel Silva： 概念化、方法论、形式分析、撰写——初稿、撰写——审阅与编辑。

利益冲突声明

作者声明不存在利益冲突。

数据可用性

本文所述研究未使用任何数据。所有结果均为理论性质。

参考文献

Aho, A.V., 1968年。索引文法——上下文无关文法的一种扩展。《ACM学报》15，647–671。doi:10.1145/321479.321488。

Baker, B.S., Book, R.V., 1974年。反转有界多栈下推自动机。《计算机与系统科学杂志》8，315–332。doi:10.1016/ S0022-0000(74)80027-9。

Bodirsky, M.、Kuhlmann, M.、Möhl, M.，2005年。良嵌套图作为句法结构的模型，载于：第十届形式语法会议与第九届语言数学会议（FG-MoL）论文集，英国苏格兰爱丁堡。第88–106页。

Dutta, M.、Saikia, H.K.，2022年。中国剩余定理中的自动机。《数学与应用通讯》13，183–197。doi:10. 26713/cma.v13i1.1704。

Hopcroft, J.E., Motwani, R., Ullman, J.D., 2007年。《自动机理论、语言与计算导论》。第3版，艾迪生-韦斯利。Joshi, A.K., 1990年。处理交叉和嵌套依存关系：从自动机视角看心理语言学研究结果。《语言与认知过程》5，1–27。doi:10.1080/01690969008402095。Kuhlmann, M., 2013年。轻度非投射依存语法。《计算语言学》39，355–387。doi:10.1162/COLI_a_00125。Seki, H., Matsumura, T., Fujii, M., Kasami, T., 1991年。论多重上下文无关文法。《理论计算机科学》88，191–229。doi:10.1016/ 0304-3975(91)90374-B。

Shieber, S.M., 1985年。反对自然语言具有上下文无关性的证据。《语言学与哲学》8，333–343。doi:10.1007/BF00630917。Swernofsky, J., Wehar, M., 2015年。论正则语言、上下文无关语言和树语言交集的复杂度，载于：《自动机、语言与编程》（ICALP 2015），施普林格。第414–426页。doi:10.1007/ 978-3-662-47666-6_33。Vijay-Shanker, K., Weir, D.J., Joshi, A.K., 1987年。刻画各种语法形式体系所产生的结构描述，载于：第25届计算语言学协会年会论文集，第104–111页。doi:10.3115/981175.981190。Yu, S., Zhuang, Q., Salomaa, K., 1994年。一些正则语言基本运算的状态复杂度。《理论计算机科学》125，315–328。doi:10.1016/0304-3975(92)00011-F。