Box Progressions and Abelian Power Free Words

Nihan Tanısalı
(joint work with Sadık Eyidoğan and Haydar Göral)

6 November 2026

Number Theory and Combinatorics 3, 2026

Given balls and boxes both enumerated with positive integers, we consider a sequential allocation of the balls into the boxes.

We fix an $\ell \geq 2$. Proceeding in increasing order of box labels, choose an arbitrary integer r between 1 and ℓ, and assign to each box the next r smallest unassigned balls.

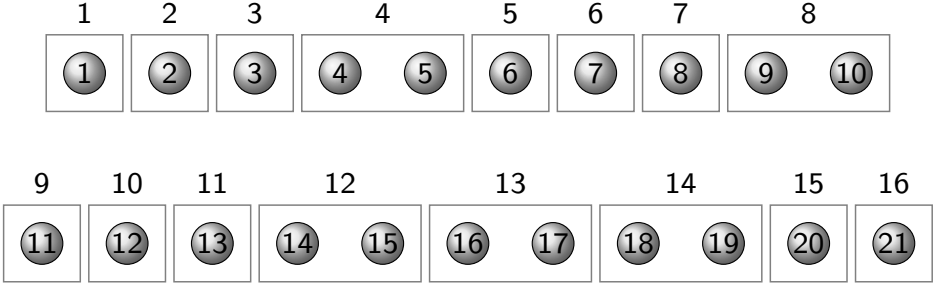
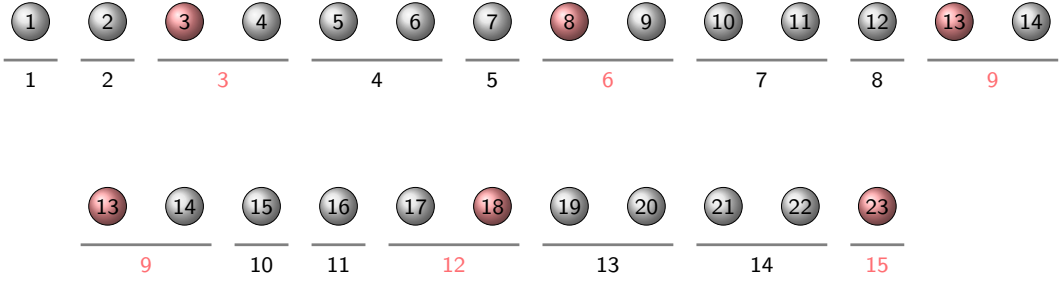


Figure 1: An example of a 2-ball-box distribution

- We focus on ball and box labels that simultaneously form a k-term arithmetic progression.
- We call these balls and boxes k ball box progressions (k-BPs).



Balls: $3, 8, 13, 18, 23$ Boxes: $3, 6, 9, 12, 15$

Question

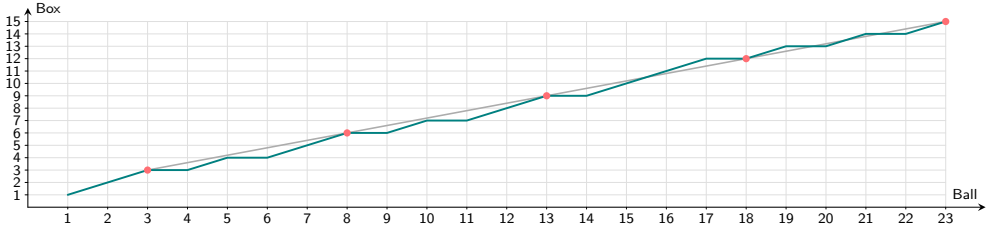
- We investigate the minimal number of balls needed to guarantee the existence of ball-box progressions
- Is there such a minimal number for every k?
- Or are there infinitely long ball-box distributions with no k ball-box progressions?

- $\ell = 2$. First non-trivial case
- The problem seems related to Ramsey theory at first.

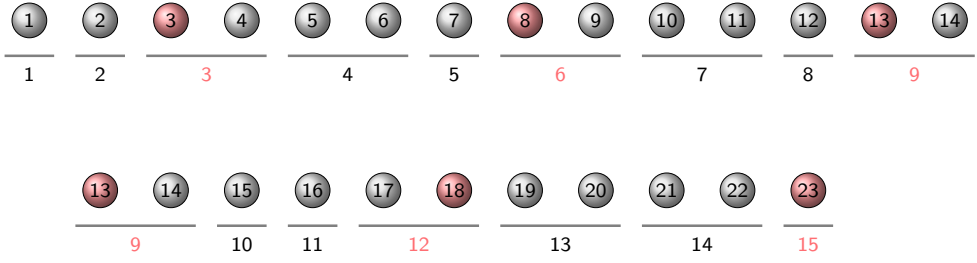
Plan of the Talk

1. Introduction
2. Formalizing the Question
3. Combinatorics on Words
4. Template Method
5. Improvements for Parent Search Algorithm
6. Other Applications of Sieving Idea

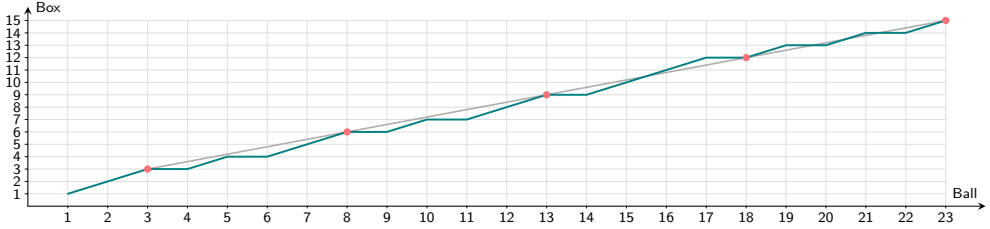
Equivalent Formalization- I



$(3,3),\ (8,6),\ (13,9),\ (18,12),\ (23,15)$



Balls: $3, 8, 13, 18, 23$ Boxes: $3, 6, 9, 12, 15$



$(3,3),\ (8,6),\ (13,9),\ (18,12),\ (23,15)$

- Indeed, we are looking for arithmetic progressions in $\mathbb{N} \times \mathbb{N}$ in the graph above.
- The analogy with Szemerédi is suggestive.

Theorem (Multidimensional Szemerédi Theorem on $\mathbb{N}^2$)

Let $A \subset \mathbb{N}^2$ be a subset with positive upper density:

$$\overline{d}(A) = \limsup_{N\to\infty} \frac{|A \cap [1,N]^2|}{N^2} > 0.$$

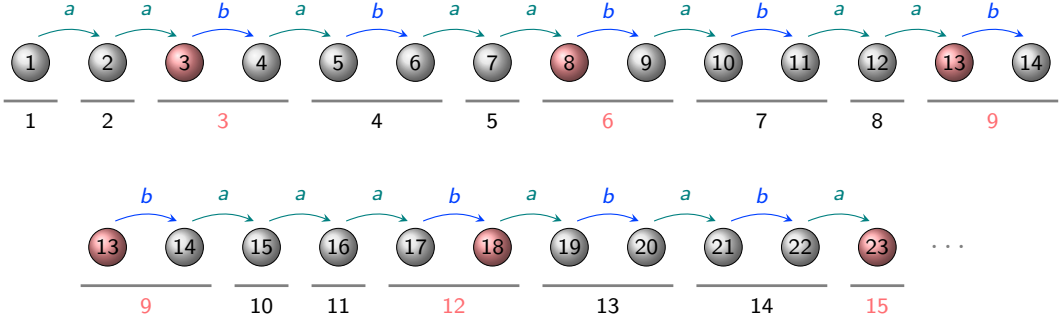
Then, for every $k \in \mathbb{N}$, there exist $x \in \mathbb{N}^2$ and nonzero $v \in \mathbb{N}^2$ such that

$$\{x, x+v, x+2v, \ldots, x+(k-1)v\} \subset A.$$

- The density of integer points on the graph is $O\left(\frac{1}{N}\right)$. This is not a dense subset.
- Multidimensional Szemerédi is not applicable, another formalization?

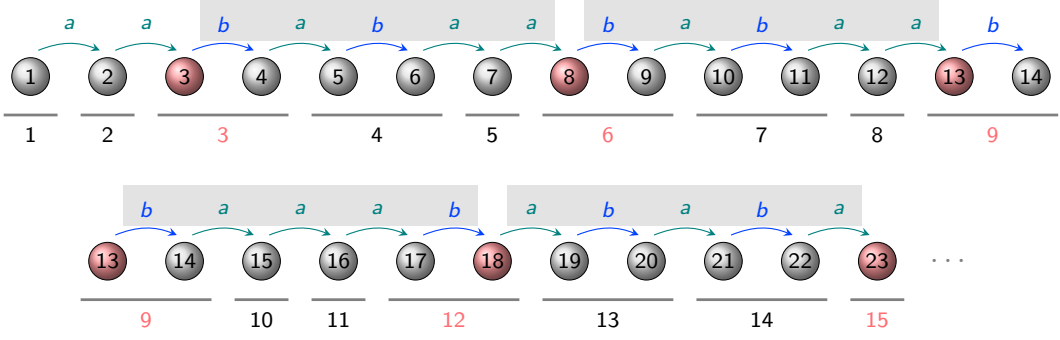
We can encode this sequence into a word over $\Sigma := \{a, b\}$ via the following rules:

- an a if the box changes
- a b if the same box



We can encode this sequence into a word over $\Sigma := \{a, b\}$ via the following rules:

- an a if the box changes
- a b if the same box



Repetition

A k repetition is composed of k consecutive equal factors. Examples:

- $aaaa$
- $\underline{aaabb}\ \underline{aaabb}\ \underline{aaabb}\ \underline{aaabb}$

Abelian Power

A k abelian power is composed of k consecutive factors that are anagrams of each other. Examples:

- $aaaa$
- $\underline{aaabb}\ \underline{aaabb}\ \underline{aaabb}\ \underline{aaabb}$
- angle, angel
- $\underline{babaa}\ \underline{babaa}\ \underline{baaab}\ \underline{ababa}$

Proposition

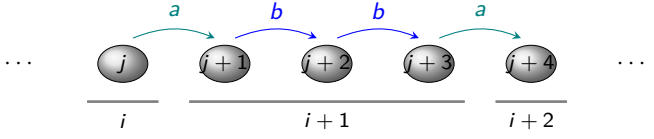
Let $k \geq 2$ be a positive integer. Let D a ball-box distribution, the corresponding word w_D is abelian $(k-1)$-power free if and only if there is no k ball-box progression in D.

being k ball-box progression free $\longleftrightarrow$ being $k-1$ abelian power free

Some previous work on abelian power-free words:

- F. Dekking, Strongly non-repetitive sequences and progression-free sets, Journal of Combinatorial Theory 27 (1979)
- A. Carpi, On abelian power-free morphisms, Internat. J. Algebra Comput. 3 (1993) 151–167.

at most 2 balls in a box $\longleftrightarrow$ no bb as a factor



1 Introduction

2 Formalizing the Question

3 Combinatorics on Words

4 Template Method

5 Improvements for Parent Search Algorithm

6 Other Applications of Sieving Idea

- Let $\Sigma = \{a, b\}$ be an alphabet. An element of Σ^* is called a word.
- The empty word is denoted by ϵ.
- Given a word
$$v = uw,$$
u is called a prefix.
- We set
$$\mathrm{Pref}(v) = \{u \in \Sigma^* : u \text{ is a prefix of } v\}.$$
- The Parikh vector of v denoted by
$$\psi(v) = (v_1, v_2) \in \mathbb{N}^{|\Sigma|}$$
where v_1 and v_2 is the number of appearances of a and b respectively.

- A morphism is a map
$$\begin{aligned} h: \quad & \Sigma^* \to \Sigma^* \\ & h(vw) \mapsto h(v)h(w). \end{aligned}$$
- The frequency matrix of a morphism is
$$\begin{pmatrix} \psi(h(a)) \\ \psi(h(b)) \end{pmatrix}.$$
- Finally, we define
$$\mathrm{Pref}(h) := \bigcup_{\sigma \in \Sigma} \mathrm{Pref}(h(\sigma)).$$
- h is called abelian k-power-free if, for every abelian k-power free word $w \in \Sigma^*$, the image $h(w)$ is also ableian k-power free.

- Consider the morphism $h : \Sigma^* \to \Sigma^*$ given by

$$h(a) = aaaba$$
$$h(b) = bab.$$

- The frequency matrix of h is

$$M = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}$$

  The determinant of the matrix is 7.

- $$\mathrm{Pref}(h) = \{\epsilon, a, aa, aaa, aaab, b, ba\}$$

Question

- Is h a k abelian power free morphism for some k?

- Define the group generated by $\psi(h(a)) = (4,1)$ and $\psi(h(b)) = (1,2)$.

$$G_h = \langle \psi(h(\sigma)) : \sigma \in \Sigma \rangle \subseteq \mathbb{Z}^{\Sigma}.$$

$$\mathsf{Pref}(h) = \{\epsilon, a, aa, aaa, aaab, b, ba\}$$

$$\mathcal{P} := \psi(\mathsf{Pref}(h)) = \{(0,0), (1,0), (2,0), (3,0), (3,1), (0,1), (1,1)\}.$$

- We observe that $(0,1)$ and $(3,0)$ are in the same coset in $\mathbb{Z}^{\Sigma}/G_h$:

$$(3,0) - (0,1) = (4,1) - (1,2).$$

$$\overline{\mathcal{P}} = \{\overline{(0,0)}, \overline{(1,0)}, \overline{(2,0)}, \overline{(3,0)}, \overline{(1,1)}, \overline{(3,1)}\}.$$

Arithmetic Rank (a-rk)

The arithmetic rank of a set A is defined as

$$\text{a-rk}(A) := \text{length of the longest arithmetic progression in } A$$

- Example: $\text{a-rk}(\psi(\mathsf{Pref}(h))) = 6$

Theorem (Eyidoğan, Göral, T.)

Given an integer $n \geq 4$, an alphabet Σ of size at least 2 and a morphism

$$h : \Sigma^* \to \Sigma^*.$$

Let

$$G_h = \langle \psi(h(\sigma)) : \sigma \in \Sigma \rangle \subseteq \mathbb{Z}^\Sigma.$$

Finally, we set

$$\mathcal{P} = \psi(\mathrm{Pref}(h)) \quad \text{and} \quad \overline{\mathcal{P}} = \{\overline{v} : v \in \mathcal{P}\} \subseteq \mathbb{Z}^{|\Sigma|}/G_h.$$

Suppose the following conditions are satisfied:

1. The frequency matrix of h is non-singular,
2. a-rk$(\overline{\mathcal{P}}) \leq n$,
3. For $v \in \mathcal{P} \setminus (\psi(h(\Sigma)) \cup \{0\})$, the class $[v]_{\mathcal{P}} = \{w \in \mathcal{P} : w - v \in G_h\}$ contains at most 2 elements. No nonempty proper prefix of h has its Parikh vector in G_h. Moreover, if $[v]_{\mathcal{P}} = \{v, w\}$ with $v \neq w$, then
$$v - w = \psi(h(p)) - \psi(h(q))$$
for unique $p, q \in \Sigma$ with $v \in \psi(\mathrm{Pref}(h(p)))$ and $w \in \psi(\mathrm{Pref}(h(q)))$.

Then, h is an abelian n-power free morphism.

Corollary

There exists an infinite word Ω over $\Sigma = \{a, b\}$ such that

- bb is not a factor of Ω,
- Ω is abelian 6-power free.

- We consider the morphism

$$\begin{aligned} h(a) &= aaaba \\ h(b) &= bab. \end{aligned} \qquad \text{with non-singular frequency matrix} \quad \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}.$$

- a-rk$(\overline{\mathcal{P}}) = 6$.
- No nonempty proper prefix has its Parikh vector in G_h.
- The only distinct vectors in $\mathcal{P}$ sharing a nonzero class are $(3, 0)$ and $(0, 1)$, and

$$(3, 0) - (0, 1) = (4, 1) - (1, 2) = \psi(h(a)) - \psi(h(b)).$$

- Thus, h is abelian 6-power free. Its fixed point $\Omega = h^\omega(a)$ is therefore abelian 6-power free and also avoids bb.

A First Answer

- Ω admits no 6 abelian powers
- Among the ball-box distributions where we allow at most 2 balls in a box,
  there is an infinitely long ball-box distribution corresponding to Ω which admits no 7 ball box progressions.

A First Answer

- Ω admits no 6 abelian powers
- Among the ball-box distributions where we allow at most 2 balls in a box, there is an infinitely long ball-box distribution corresponding to Ω which admits no 7 ball box progressions.

Question

- Is 7 optimal?

- We constructed an infinite word over $\{a, b\}$ that has no bb as a factor and no abelian 6-powers.

Is 6 Optimal?

- **Question:** Can we construct a word that avoids bb and avoids an abelian 5-power.

- When we constructed the abelian 6 power free word Ω the strategy was to use a morphism

$$\begin{aligned} h : \Sigma &\to \Sigma \\ u &\to h(u) \end{aligned}$$

  such that under h, abelian 6-power free elements have abelian 6-power free images.

- Then, we consider $h^\omega(a)$.

- What about a morphism h that is abelian 5-power free?

- We couldn't find any such h.

The New Strategy

- Find a morphism h that is not necessarily abelian 5-power free, but its fixed point $h^\omega(a)$ is abelian 5-power free.

- When we constructed the abelian 6 power free word Ω the strategy was to use a morphism

$$h : \Sigma \to \Sigma$$
$$u \to h(u)$$

  such that under h, abelian 6-power free elements have abelian 6-power free images.

- Then, we consider $h^\omega(a)$.

- What about a morphism h that is abelian 5-power free?

- We couldn't find any such h.

The New Strategy

- Find a morphism h that is not necessarily abelian 5-power free, but its fixed point $h^\omega(a)$ is abelian 5-power free.
- Spoiler:

$$h(a) = aaaab \qquad h(b) = ababab.$$

1 Introduction

2 Formalizing the Question

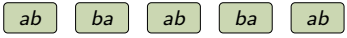
3 Combinatorics on Words

4 Template Method

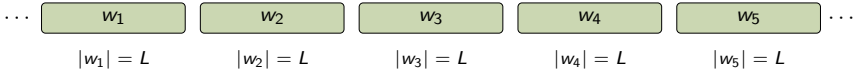
5 Improvements for Parent Search Algorithm

6 Other Applications of Sieving Idea

An abelian 5-power of length $5 \times 2 = 10$:



An abelian 5-power of length $5L$:



- There are always **five blocks**, but the length can be arbitrarily large.
- An occurrence may start arbitrarily far into the infinite word.

How can a finite computation rule out all such occurrences?

Currie–Rampersad, 2012

Under suitable hypotheses on a morphism h, the template method decides whether its fixed point $h^\omega(a)$ is abelian k-power free.

It reduces a question about an **infinite word** to a **finite computation**.

J. D. Currie and N. Rampersad, *Fixed points avoiding abelian k-powers*, Journal of Combinatorial Theory, Series A 119 (2012), 942–948.

- Any **long** abelian power in $h^\omega(a)$ is obtained by applying h repeatedly to a **short** "seed word".
- Template method asks how can this seed word look like.
- **Answer:** The short words have to fit in a certain set of templates.
- After computing the **set of templates** and **short** words in $h^\omega(a)$, we check if any short word fits into any template.

What Is Short? What Is the "Template Set"?

- We will investigate how to compute the template set.
- We will take this as a black box: given a template set there is a natural number n such that the factors w of $h^\omega(a)$ such that $|w| \leq n$ are small.

- A template for abelian 5-powers is a tuple

$$[a_0, a_1, a_2, a_3, a_4, a_5, \vec{d}_1, \vec{d}_2, \vec{d}_3, \vec{d}_4],$$

  where $a_i \in \{a, b, \epsilon\}$ and $\vec{d}_i \in \mathbb{Z}^{|\{a,b\}|}$.
- A word w is an instance of this template if

$$w = a_0 w_0 a_1 w_1 a_2 w_2 a_3 w_3 a_4 w_4 a_5$$

  and

$$\psi(w_i) - \psi(w_{i-1}) = \vec{d}_i \qquad (i = 1, 2, 3, 4).$$

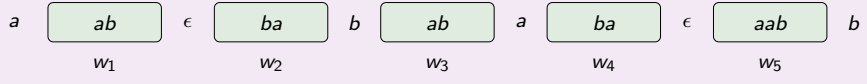
Example

The word

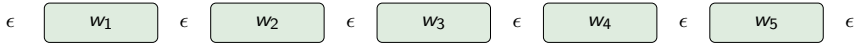
$$w = aabbababababaaabb$$

is an instance of

$$T_5 = [a, \epsilon, b, a, \epsilon, b,\ (0,0), (0,0), (0,0), (1,0)].$$



- An instance of
$$T_5 = [\epsilon, \epsilon, \epsilon, \epsilon, \epsilon, \vec{0}, \vec{0}, \vec{0}, \vec{0}]$$
with nonempty w_i is an abelian 5-power.



- We are looking for instances of T_5 in $h^\omega(a)$ for some h.

- Let $h^r(a)$ be the first iterate containing an instance of T_5, and suppose this instance is **long**.
- Can we trace it back through h to a shorter word in $h^{r-1}(a)$?
- The shorter word **does not** realize T_5: by our choice of r, no instance of T_5 occurs in $h^{r-1}(a)$.

- This sufficiently long instance of T_5 comes from a shorter instance of another template: a **parent** template of T.
- The parent relation depends on the morphism h.

To find a **parent** of T, we need to track how h changes the letter counts in each block.

$$h(a) = aaaab, \qquad h(b) = ababab.$$

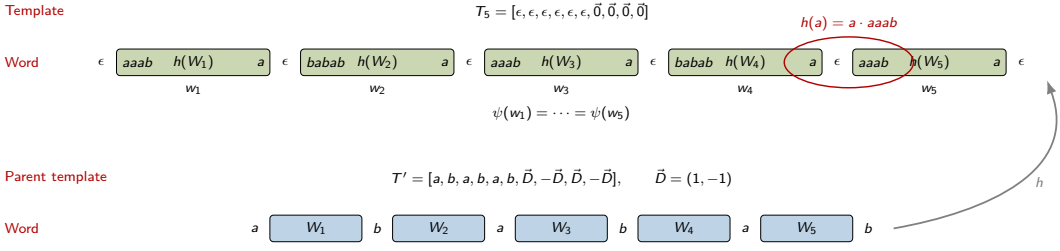
- We will show: $\psi(h(u)) = \psi(u)M$.

If $\psi(u) = (x, y)$, then:

Letters in u	a's in $h(u)$	b's in $h(u)$
x copies of a	$4x$	x
y copies of b	$3y$	$3y$

$$\psi(h(u)) = (4x + 3y,\ x + 3y) = \underbrace{(x, y)}_{\psi(u)} \underbrace{\begin{pmatrix} 4 & 1 \\ 3 & 3 \end{pmatrix}}_{M}.$$

$$h: \quad a \mapsto aaaab, \qquad b \mapsto ababab.$$



- Since $\psi(w_1) = \psi(w_2)$,

$$\underbrace{\psi(W_1)M}_{\psi(h(W_1))} + (4,1) = \underbrace{\psi(W_2)M}_{\psi(h(W_2))} + (3,3).$$

Continuing in this fashion, we compute all difference vectors.

Let

$$T_5 = [\epsilon, \ldots, \epsilon, \vec{0}, \ldots, \vec{0}], \qquad T' = [A_0, \ldots, A_5, \vec{D}_1, \ldots, \vec{D}_4].$$

Let M be the frequency matrix of h, with $\psi(h(u)) = \psi(u)M$.

Definition

We call T' a *parent* of T_5 under h if there exist $p_i, s_i \in \Sigma^*$ such that

$$h(A_i) = p_i s_i \qquad (i = 0, \ldots, 5),$$

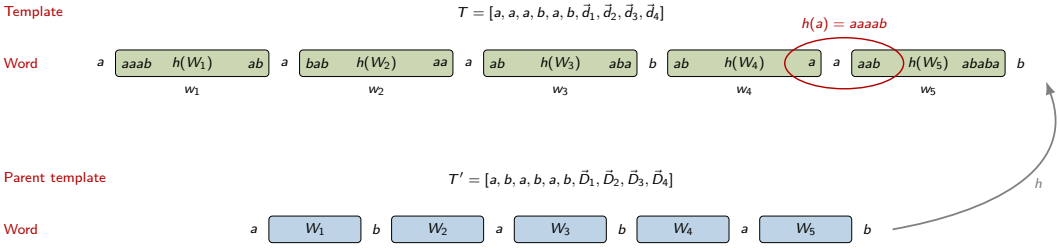
and

$$\vec{D}_i M = \psi(s_{i-1} p_i) - \psi(s_i p_{i+1}) \qquad (i = 1, \ldots, 4).$$

- A long instance of T_5 leads to a shorter instance of a parent template T'.
- But this shorter instance may still be very long.
- If it is sufficiently long, we can trace it back again: this time to an instance of a **parent of** T'.

- We therefore need the parent relation for **general templates**, not only for T_5.

$$h: \quad a \mapsto aaaab, \qquad b \mapsto ababab.$$



- Since $\psi(w_2) - \psi(w_1) = \vec{d}_1$,

$$\underbrace{\psi(W_1)M}_{\psi(h(W_1))} + (4,2) + \vec{d}_1 = \underbrace{\psi(W_2)M}_{\psi(h(W_2))} + (3,2).$$

- Thus, $\vec{D}_1 = \psi(W_2) - \psi(W_1)$ satisfies

$$\vec{D}_1 M = \vec{d}_1 + (1,0).$$

Let

$$T = [a_0, \ldots, a_5, \vec{d}_1, \ldots, \vec{d}_4], \qquad T' = [A_0, \ldots, A_5, \vec{D}_1, \ldots, \vec{D}_4].$$

Let M be the frequency matrix of h, with $\psi(h(u)) = \psi(u)M$.

Definition

We call T' a *parent* of T under h if there exist $p_i, s_i \in \Sigma^*$ such that

$$h(A_i) = p_i a_i s_i \qquad (i = 0, \ldots, 5),$$

and

$$\vec{d}_i = \vec{D}_i M + \psi(s_i p_{i+1}) - \psi(s_{i-1} p_i) \qquad (i = 1, \ldots, 4).$$

- Every sufficiently long instance of T_5 comes from a shorter instance of one of its parent templates.
- Thus, instead of searching for T_5 in a long word, we can search for its parents in a shorter preimage word.
- Repeating this process leads us to parents of parents, and hence to **ancestors**.
- How do we compute these parent templates?

Splits

For $c \in \Sigma \cup \{\epsilon\}$, define

$$S(c) = \{(A, p, s) : A \in \Sigma \cup \{\epsilon\},\ p, s \in \Sigma^*,\ h(A) = pcs\}.$$

For example, since $h(b) = abababab$,

$$h(b) = \underbrace{ab}_{p}\ \underbrace{a}_{c}\ \underbrace{bab}_{s}, \qquad (b, ab, bab) \in S(a).$$

If $c = \epsilon$, a split is simply a cut $h(A) = ps$.

To compute the parents of

$$T = [a_0, \ldots, a_5, \vec{d}_1, \ldots, \vec{d}_4],$$

we do the following

1. Choose $(A_i, p_i, s_i) \in S(a_i)$ for each $i = 0, \ldots, 5$.
2. Compute,
$$\vec{D}_i = \left(\vec{d}_i + \psi(s_{i-1}p_i) - \psi(s_i p_{i+1})\right) M^{-1}, \qquad i = 1, \ldots, 4.$$
3. If every $\vec{D}_i$ is integral, retain the parent
$$T' = [A_0, \ldots, A_5, \vec{D}_1, \ldots, \vec{D}_4].$$

- There are 6 for loops $S(a_0), \ldots, S(a_5)$. Note that $|S(a_i)| \leq \max_{\sigma \in \Sigma} |h(\sigma)|$

Definition

A template T' is an *ancestor* of T if there exists a chain

$$T = T_0,\ T_1, \ldots, T_r = T',$$

where T_{i+1} is a parent of T_i for every $i = 0, \ldots, r-1$. We denote the set of all ancestors of T, including T itself, by

$$\mathrm{Anc}_h(T).$$

Starting from the abelian 5-power template

$$T_5 = [\epsilon, \epsilon, \epsilon, \epsilon, \epsilon, \epsilon, \vec{0}, \vec{0}, \vec{0}, \vec{0}],$$

we compute its ancestor set recursively:

1. Start with T_5.
2. Compute the parents of each newly found template.
3. Add unseen templates and repeat until no new template appears.

We look for the ancestor set of

$$T_5 = [\epsilon, \epsilon, \epsilon, \epsilon, \epsilon, \vec{0}, \vec{0}, \vec{0}, \vec{0}].$$

Assume that $\|M^{-1}\| < 1$.

- There are only finitely many integer vectors within this bound and finitely many splits (p_i, a_i, s_i).
- If we bound $\|\vec{D}_i\|$'s then we are done.

We look for the ancestor set of

$$T_5 = [\epsilon, \epsilon, \epsilon, \epsilon, \epsilon, \vec{0}, \vec{0}, \vec{0}, \vec{0}].$$

Assume that $\|M^{-1}\| < 1$.

- There are only finitely many integer vectors within this bound and finitely many splits (p_i, a_i, s_i).
- If we bound $\|\vec{D}_i\|$'s then we are done.
- The split contributions are bounded by $\max_{\sigma \in \Sigma} |h(\sigma)|$. Therefore, each parent difference vector D_i satisfies

$$\|\vec{D}_i\| \leq \|M^{-1}\| \cdot \|\vec{d}_i\| + C_h.$$

- Starting from the zero difference vectors of T_5, iteration gives

$$\|\vec{D}_i\| \leq \frac{C}{1 - \|M^{-1}\|}.$$

- Under the condition $\|M^{-1}\| < 1$, we obtain a finite set of templates.
- How short must the instances we check be?

Let $\Sigma = \{\sigma_1, \ldots, \sigma_m\}$ and $h : \Sigma^* \to \Sigma^*$ satisfy

- $h(\sigma_1) = \sigma_1 x$ for some nonempty word x;
- $|h(\sigma)| > 1$ for every $\sigma \in \Sigma$.

Set $N = \max_{\sigma \in \Sigma} |h(\sigma)|$.

Inverse Parent Lemma (Currie–Rampersad)

Let I be a factor of $h^\omega(\sigma_1)$ realizing the k-template

$$t = [a_1, \ldots, a_{k+1}, d_1, \ldots, d_{k-1}],$$

and put

$$\Delta = \left\lfloor \max_{1 \le i \le k-1} \|d_i\|_2 \right\rfloor .$$

If

$$|I| > \quad N + k - 1 + (k-1)(N - 2 + mk\Delta)$$

then $h^\omega(\sigma_1)$ contains a factor J realizing a parent of t, with $|J| < |I|$.

Let $\Sigma = \{\sigma_1, \ldots, \sigma_m\}$ and $h : \Sigma^* \to \Sigma^*$ satisfy

- $h(\sigma_1) = \sigma_1 x$ for some nonempty word x;
- $|h(\sigma)| > 1$ for every $\sigma \in \Sigma$.

Set $N = \max_{\sigma \in \Sigma} |h(\sigma)|$.

Inverse Parent Lemma (Currie–Rampersad)

Let I be a factor of $h^\omega(\sigma_1)$ realizing the k-template

$$t = [a_1, \ldots, a_{k+1}, d_1, \ldots, d_{k-1}],$$

and put

$$\Delta = \left\lfloor \max_{1 \le i \le k-1} \|d_i\|_2 \right\rfloor.$$

If

$$|I| > \quad N + k - 1 + (k-1)(N - 2 + mk\Delta)$$

Shortness bound

then $h^\omega(\sigma_1)$ contains a factor J realizing a parent of t, with $|J| < |I|$.

We can now identify the two ingredients from our initial outline.

What Is the Template Set?

It is the ancestor set of the abelian 5-power template:

$$\mathrm{Anc}_h(T_5).$$

These templates describe the possible "seed words".

What Does Short Mean?

The **inverse parent lemma** gives a bound n such that $h^\omega(a)$ contains an abelian 5-power if and only if one of its factors of length at most n realizes a template in $\mathrm{Anc}_h(T_5)$.

Thus, once the ancestor set is finite and computed:

1. Compute the bound $n = N + k - 1 + (k-1)(N - 2 + mk\Delta)$.
2. List the factors of $h^\omega(a)$ of length at most n.
3. Check whether any of them realizes a template in $\mathrm{Anc}(T_5)$.

Theorem (Currie–Rampersad, 2012)

Let $h : \Sigma^* \to \Sigma^*$ be a morphism with frequency matrix M, and let $k \geq 2$.

Assume that:

- $h(a) = ax$ for some $a \in \Sigma$ and nonempty word x;
- $|h(\sigma)| > 1$ for every $\sigma \in \Sigma$;
- M is non-singular and $\|M^{-1}\| < 1$.

Then there is an algorithm deciding whether the fixed point $h^\omega(a)$ is abelian k-power free.

Th morphism

$$h: \quad a \mapsto aaaab, \qquad b \mapsto ababab.$$

Theorem (Eyidoğan, Göral, T.)

The fixed point $h^\omega(a)$ is abelian 5-power free and avoids bb. However, h is not an abelian k-power-free morphism for any $k \geq 2$.

Indeed, for every $k \geq 2$, the word b^{k-1} is abelian k-power free, whereas

$$h(b^{k-1}) = (ab)^{3(k-1)}$$

contains the abelian k-power $(ab)^k$.

Why Is the Fixed Point Abelian 5-Power Free?

- T_5 has 16 163 parents and no further ancestors. A factor w is **short** if $|w| \leq 34$.
- Every factor of length at most 34 occurs in $h^4(a)$.
- A computer check confirms that $h^4(a)$ contains no factor realizing 16 163 parents.

Assume that each box contains one or two balls.

- **Brute-force algorithm:** Every sufficiently long binary word avoiding bb contains an abelian 4-power. Hence every sufficiently long ball–box distribution contains a 5-term ball–box progression.

- The fixed point of
$$h(a) = aaaab, \qquad h(b) = ababab$$
avoids both bb and abelian 5-powers. It therefore gives an infinite ball–box distribution with no 6-term ball–box progression.

The Answer

The smallest k for which an infinite ball–box distribution can avoid k-term ball–box progressions is $k = 6$.

- We are happy with the positive answer, but not with the computation that is "expensive".

1. Generating the Ancestor Set

- For each template, choose a split at each of the $k+1$ boundaries and test whether the parent equations have integer solutions.
- This can require examining up to $|S_h|^{k+1}$ split tuples.
- Every newly retained parent triggers another parent search: the cost accumulates across generations.

2. Testing Realizability

- For every retained template, test the factors of $h^{\omega}(a)$ up to the realization bound.
- A large template set, a large k, and a large realization bound make this stage expensive.

Our goal:

- reduce the search during ancestor generation
- reduce the length of the factors to be examined

1 Introduction

2 Formalizing the Question

3 Combinatorics on Words

4 Template Method

5 Improvements for Parent Search Algorithm

6 Other Applications of Sieving Idea

Definition

Let G be an abelian group. A **scoring morphism** is a map $f : \Sigma^* \longrightarrow G$ satisfying

$$f(\epsilon) = 0, \qquad f(uv) = f(u) + f(v).$$

Thus, the score of a word is the sum of its letter scores.

For the sieve, we choose f so that

$$f(h(c)) = 0 \qquad \text{for every } c \in \Sigma.$$

Consequently, $f(h(W)) = 0$ for every word W.

Example

For $h(a) = aaaab$ and $h(b) = ababab$, take

$$f : \{a, b\}^* \longrightarrow \mathbb{Z}/9\mathbb{Z}, \qquad f(a) = 1, \quad f(b) = 5.$$

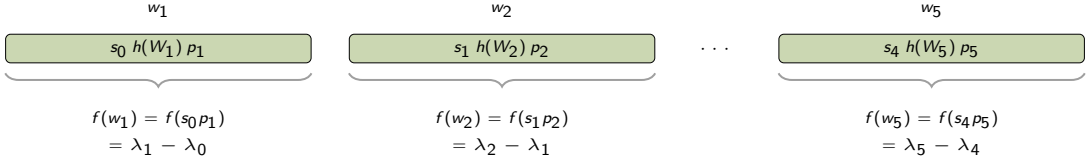
Then $f(u) = |u|_a + 5|u|_b$, and in $\mathbb{Z}/9\mathbb{Z}$,

$$f(h(a)) = 4 + 5 = 0, \qquad f(h(b)) = 3 + 3 \cdot 5 = 0.$$

$$h(a) = aaaab, \qquad h(b) = ababab$$

$$f(a) = 1 \mod 9 \qquad f(b) = 5 \mod 9$$

Choose splits $h(A_i) = p_i s_i$ and put $\lambda_i = f(p_i)$. Then $f(s_i) = -\lambda_i$.



An Abelian 5-Power Has Equal Block Scores

$$\lambda_1 - \lambda_0 = \cdots = \lambda_5 - \lambda_4.$$

Thus, $\lambda_0, \ldots, \lambda_5$ must form an arithmetic progression in the set of available prefix scores:

$$P = \{0, 1, 2, 3, 4, 6, 7\} \subseteq \mathbb{Z}/9\mathbb{Z}.$$

For example, starting with $0, 1$ forces $(0, 1, 2, 3, 4, 5)$, but $5 \notin P$. Discard these choices before computing the $\vec{D}_i$.

Under the same hypotheses as the classical inverse parent lemma, set $N = \max_{\sigma\in\Sigma} |h(\sigma)|$.

Improved Inverse Parent Lemma (Eyidoğan, Göral, T.)

Let I be a factor of $h^\omega(\sigma_1)$ realizing the k-template

$$t = [a_1, \ldots, a_{k+1}, d_1, \ldots, d_{k-1}].$$

If

$$|I| > N + k - 1 + (k-1)\left(N - 2 + \max_{1\le i<j\le k} \left|\sum_{r=i}^{j-1} d_r \cdot (1,\ldots,1)\right|\right),$$

where $d_r = \psi(X_{r+1}) - \psi(X_r)$, then $h^\omega(\sigma_1)$ contains a factor J realizing a parent of t, with $|J| < |I|$.

- Classical bound:
$$N + k - 1 + (k-1)(N - 2 + mk\Delta).$$
- Our bound:
$$N + k - 1 + (k-1)\left(N - 2 + \max_{1\le i<j\le k} \left|\sum_{r=i}^{j-1} d_r \cdot (1,\ldots,1)\right|\right).$$

Consider the morphism

$$h(a) = aaaab, \qquad h(b) = ababab.$$

Proposition (Eyidoğan, Göral, T.)

The sieve selects a set τ of 120 parent templates of T_5 such that the following are equivalent:

- $h^\omega(a)$ contains an abelian 5-power. **(Sieving idea)**
- A factor of $h^\omega(a)$ of length at most 34 realizes a template in τ. **(Improved inverse lemma)**

Thus, only 120 of the 16 163 parent templates need to be checked.

	Classical template method	**With our improvements**
Parent templates to check	16 163	120
Maximum factor length to check	106	34

1 Introduction

2 Formalizing the Question

3 Combinatorics on Words

4 Template Method

5 Improvements for Parent Search Algorithm

6 Other Applications of Sieving Idea

We now apply our reductions beyond the ball–box problem, to study abelian powers in fixed points of other morphisms.

Goal: Decide whether a fixed point contains an abelian power by checking fewer templates and shorter factors.

We illustrate two applications:

- **Detecting a power:** finding an occurrence even when a long initial prefix contains none.
- **Proving avoidance:** establishing that no occurrence exists anywhere in the fixed point.

Consider the morphism

$$h: \quad a \mapsto abaaaba, \qquad b \mapsto babab.$$

Use the scoring morphism

$$f : \{a, b\}^* \longrightarrow \mathbb{Z}/11\mathbb{Z}, \qquad f(a) = 1, \quad f(b) = 3.$$

The sieve selects a set τ of 8 parents of T_8.

	Classical template method	**With our improvements**
Initial parent templates	1 953 162	8
Ancestor templates to check	$\approx$ 2 000 000	224
Maximum factor length to check	273	56

The bound 56 uses the improved inverse parent lemma on the ancestors of τ.

Proposition (Eyidoğan, Göral, T.)

The fixed point $h^\omega(a)$ contains an abelian 8-power. An occurrence is found in $h^6(a)$.

Consider the morphism

$$h: \quad a \mapsto abaababababa, \qquad b \mapsto bababababab.$$

Use the scoring morphism

$$f : \{a, b\}^* \longrightarrow \mathbb{Z}/19\mathbb{Z}, \qquad f(a) = 1, \quad f(b) = 3.$$

The sieve selects a set τ of 32 parents of T_{14}.

	Classical template method	**With our improvements**
Initial parent templates	$> 600\,000\,000$	32
Ancestor templates to check	$> 600\,000\,000$	1 244
Maximum factor length to check	869	154

The selected set τ has 268 parents and 1 244 ancestors.

Theorem (Eyidoğan, Göral, T.)

The morphism h is abelian 16-power free, but not abelian 15-power free, while its fixed point $h^\omega(a)$ is abelian 14-power free.

- We translated ball–box progressions into abelian powers.
- For one or two balls per box, 5-term progressions are unavoidable, while 6-term progressions can be avoided.
- Sieving reduces the template search and applies to other abelian-power problems.

