Van der Waerden's Theorem and Avoidability in Words

Yu Hin (Gary) Au[1] Jeffrey Shallit[2]

[1]Department of Combinatorics & Optimization
University of Waterloo

[2]David R. Cheriton School of Computer Science
University of Waterloo

CanaDAM 2009

Outline

1. Basic results in avoidability in words.
2. The problem of avoiding arithmetic squares and some of its variants.

Avoidability in Words

Definitions

A **square** is a word of the form xx. E.g. "murmur".

Facts

- Any word over $\{0, 1\}$ of length ≥ 4 contains a square.
- There exists an infinite word over $\{0, 1, 2\}$ that avoids squares (Thue, 1906)

Avoidability in Words

Definitions

A **square** is a word of the form *xx*. E.g. “murmur”.

Facts

- Any word over $\{0, 1\}$ of length ≥ 4 contains a square.
- There exists an infinite word over $\{0, 1, 2\}$ that avoids squares (Thue, 1906)

Avoidability in Words

Definition

An **abelian square** is a word of the form xx', with x' being a permutation of the symbols in x. E.g. “reappear”.

Fact

Any word over $\{0, 1, 2\}$ of length ≥ 8 contains an abelian square.

Question (Erdös, 1961)

Is there an infinite word over a finite alphabet that avoids abelian squares?

Avoidability in Words

Definition

An **abelian square** is a word of the form xx', with x' being a permutation of the symbols in x. E.g. “reappear”.

Fact

Any word over $\{0, 1, 2\}$ of length ≥ 8 contains an abelian square.

Question (Erdös, 1961)

Is there an infinite word over a finite alphabet that avoids abelian squares?

Avoidability in Words

Question (Erdös, 1961)

Is there an infinite word over a finite alphabet that avoids abelian squares?

Answer

Yes! Can be done over an alphabet of size

- 25 (Evdokimov,1968)
- 5 (Pleasants,1970)
- 4 (Keränen, 1991)

Avoidability in Words

Definition

An **arithmetic square** is a word over the integers of the form xx', such that

- x, x' have the same length;
- the sum of symbols in x is equal to the sum of symbols in x'.

E.g. "32341434".

Question (Pirillo & Varricchio, 1994)

Is there an infinite word over a finite subset of $\mathbb{Z}$ that avoids arithmetic squares?

Avoidability in Words

Definition

An **arithmetic square** is a word over the integers of the form xx', such that

- x, x' have the same length;
- the sum of symbols in x is equal to the sum of symbols in x'.

E.g. "32341434".

Question (Pirillo & Varricchio, 1994)

Is there an infinite word over a finite subset of $\mathbb{Z}$ that avoids arithmetic squares?

Avoiding Arithmetic Squares

Three Slightly "Stronger" Problems

1. Is there a $p \in \mathbb{N}$ and an infinite word over $\mathbb{Z}_p$ that does not contain two consecutive blocks of the same length and the same sum modulo p?
2. Is there an infinite word over a finite subset of $\mathbb{Z}$ that avoids consecutive blocks of the same sum?
3. Is there an infinite word over a finite alphabet that
   a) avoids abelian squares, and
   b) in any subword, the difference between the number of occurrences of the most and least frequent letter is bounded above by a constant?

Answer

No, no and no.

Avoiding Arithmetic Squares

Three Slightly "Stronger" Problems

1. Is there a $p \in \mathbb{N}$ and an infinite word over $\mathbb{Z}_p$ that does not contain two consecutive blocks of the same length and the same sum modulo p?
2. Is there an infinite word over a finite subset of $\mathbb{Z}$ that avoids consecutive blocks of the same sum?
3. Is there an infinite word over a finite alphabet that
   a) avoids abelian squares, and
   b) in any subword, the difference between the number of occurrences of the most and least frequent letter is bounded above by a constant?

Answer

No, no and no.

Avoiding Arithmetic Squares

Three Slightly "Stronger" Problems

1. Is there a $p \in \mathbb{N}$ and an infinite word over $\mathbb{Z}_p$ that does not contain two consecutive blocks of the same length and the same sum modulo p?
2. Is there an infinite word over a finite subset of $\mathbb{Z}$ that avoids consecutive blocks of the same sum?
3. Is there an infinite word over a finite alphabet that
   a) avoids abelian squares, and
   b) in any subword, the difference between the number of occurrences of the most and least frequent letter is bounded above by a constant?

Answer

No, no and no.

Avoiding Arithmetic Squares

Three Slightly "Stronger" Problems

1. Is there a $p \in \mathbb{N}$ and an infinite word over $\mathbb{Z}_p$ that does not contain two consecutive blocks of the same length and the same sum modulo p?
2. Is there an infinite word over a finite subset of $\mathbb{Z}$ that avoids consecutive blocks of the same sum?
3. Is there an infinite word over a finite alphabet that
   a) avoids abelian squares, and
   b) in any subword, the difference between the number of occurrences of the most and least frequent letter is bounded above by a constant?

Answer

No, no and no.

Avoiding Arithmetic Squares

Theorem (Van der Waerden, 1921)

Suppose we colour $\mathbb{N}$ with finitely many colours. Then for every integer $k \geq 0$, there exist integers a, d such that

$$a, a + d, a + 2d, \ldots, a + kd$$

are all assigned the same colour.

Avoiding Arithmetic Squares

Notation

For any word x:

- $x[i]$ denotes i-th symbol in x;
- $x[i..j]$ denotes the subword $x[i]x[i+1]\cdots x[j]$;
- $|x|_a$ denotes the number of occurrences of the symbol "a" in x.

Avoiding Arithmetic Squares

Problem 1

Is there a $p \in \mathbb{N}$ and an infinite word over $\mathbb{Z}_p$ that does not contain two consecutive blocks of the same length and the same sum modulo p?

Avoiding Arithmetic Squares

Solution

- Suppose $\mathbf{w}$ is such a word.
- For every $i \in \mathbb{N}$, define $f(i) := \sum_{j=1}^{i} \mathbf{w}[j] \pmod{p}$.
- VDW: $\exists a, d \in \mathbb{N}$ such that $f(a) = f(a+d) = f(a+2d)$.
- $f(a) = f(a+d)$ implies

$$\sum_{j=1}^{a} \mathbf{w}[j] \equiv \sum_{j=1}^{a+d} \mathbf{w}[j] \pmod{p} \Rightarrow \sum_{j=a+1}^{a+d} \mathbf{w}[j] \equiv 0 \pmod{p}$$

- Similarly,

$$f(a+d) = f(a+2d) \Rightarrow \sum_{j=a+d+1}^{a+2d} \mathbf{w}[j] \equiv 0 \pmod{p}. \quad \square$$

Avoiding Arithmetic Squares

Problem 2

Is there an infinite word over a finite subset of $\mathbb{Z}$ that avoids consecutive blocks of the same sum?

Avoiding Arithmetic Squares

Solution (Halbeisen & Hungerbühler, 2000)

- Suppose **w** is an infinite word over a finite subset of $\mathbb{N}$. E.g. $\mathbf{w} = 5414541\ldots$;
- Construct sequence

$$\mathbf{u} := 543214321143215432143211\ldots$$

and apply VDW.

- Can be slightly tweaked to take care of sequences over $\mathbb{Z}$.

$\square$

Avoiding Arithmetic Squares

Solution (Halbeisen & Hungerbühler, 2000)

- Suppose **w** is an infinite word over a finite subset of $\mathbb{N}$. E.g. $\mathbf{w} = 5414541\ldots$;
- Construct sequence

$$\mathbf{u} := 54321432114321543214321 1\ldots$$

and apply VDW.

- Can be slightly tweaked to take care of sequences over $\mathbb{Z}$.

□

Avoiding Arithmetic Squares

Solution (Halbeisen & Hungerbühler, 2000)

- Suppose **w** is an infinite word over a finite subset of $\mathbb{N}$. E.g. $\mathbf{w} = 5414541\ldots$;
- Construct sequence

$$\mathbf{u} := 543214321143215432143211\ldots$$

and apply VDW.

- Can be slightly tweaked to take care of sequences over $\mathbb{Z}$.

$\square$

Avoiding Arithmetic Squares

Solution (Halbeisen & Hungerbühler, 2000)

- Suppose **w** is an infinite word over a finite subset of $\mathbb{N}$. E.g. $\mathbf{w} = 5414541\ldots$;
- Construct sequence

$$\mathbf{u} := 5432143211432154321432 11\ldots$$

and apply VDW.

- Can be slightly tweaked to take care of sequences over $\mathbb{Z}$.

□

Avoiding Arithmetic Squares

Solution (Halbeisen & Hungerbühler, 2000)

- Suppose $\mathbf{w}$ is an infinite word over a finite subset of $\mathbb{N}$. E.g. $\mathbf{w} = 5414541\ldots$;
- Construct sequence

$$\mathbf{u} := 54321432114321543214321 1\ldots$$

and apply VDW.

- Can be slightly tweaked to take care of sequences over $\mathbb{Z}$.

□

Avoiding Arithmetic Squares

Solution (Halbeisen & Hungerbühler, 2000)

- Suppose **w** is an infinite word over a finite subset of $\mathbb{N}$. E.g. $\mathbf{w} = 5414541\ldots$;
- Construct sequence

$$\mathbf{u} := 54321432114321543214321 1\ldots$$

and apply VDW.

- Can be slightly tweaked to take care of sequences over $\mathbb{Z}$.

□

Avoiding Arithmetic Squares

Solution (Halbeisen & Hungerbühler, 2000)

- Suppose **w** is an infinite word over a finite subset of $\mathbb{N}$. E.g. $\mathbf{w} = 5414541\ldots$;
- Construct sequence

$$\mathbf{u} := 543214321143215432143211\ldots$$

and apply VDW.

- Can be slightly tweaked to take care of sequences over $\mathbb{Z}$.

$\square$

Avoiding Arithmetic Squares

Solution (Halbeisen & Hungerbühler, 2000)

- Suppose **w** is an infinite word over a finite subset of $\mathbb{N}$. E.g. $\mathbf{w} = 5414541\ldots$;
- Construct sequence

$$\mathbf{u} := 54321432114321543214321 1\ldots$$

and apply VDW.

- Can be slightly tweaked to take care of sequences over $\mathbb{Z}$.

□

Avoiding Arithmetic Squares

Solution (Halbeisen & Hungerbühler, 2000)

- Suppose **w** is an infinite word over a finite subset of $\mathbb{N}$. E.g. $\mathbf{w} = 5414541\ldots$;
- Construct sequence

$$\mathbf{u} := 54321432114321543214321 1\ldots$$

  and apply VDW.

- Can be slightly tweaked to take care of sequences over $\mathbb{Z}$.

  $\square$

Avoiding Arithmetic Squares

Solution (Halbeisen & Hungerbühler, 2000)

- Suppose **w** is an infinite word over a finite subset of $\mathbb{N}$. E.g. $\mathbf{w} = 5414541\ldots$;
- Construct sequence

$$\mathbf{u} := 54\underline{3}2143211 4\underline{3}2154321 4\underline{3}211\ldots$$

  and apply VDW.
- Can be slightly tweaked to take care of sequences over $\mathbb{Z}$.

  □

Avoiding Arithmetic Squares

Solution (Halbeisen & Hungerbühler, 2000)

- Suppose **w** is an infinite word over a finite subset of $\mathbb{N}$. E.g. $\mathbf{w} = 5414541$
- Construct sequence

$$\mathbf{u} := 543214321143215432143211\ldots$$

and apply VDW.

- Can be slightly tweaked to take care of sequences over $\mathbb{Z}$.

$\square$

Avoiding Arithmetic Squares

Solution (Halbeisen & Hungerbühler, 2000)

- Suppose **w** is an infinite word over a finite subset of $\mathbb{N}$. E.g. $\mathbf{w} = 5414541$
- Construct sequence

$$\mathbf{u} := 54\underline{3}21\underline{4}321\underline{14}\underline{3}21\underline{5}4321\underline{4}\underline{3}211\ldots$$

and apply VDW.

- Can be slightly tweaked to take care of sequences over $\mathbb{Z}$. □

Avoiding Arithmetic Squares

Problem 3

Is there an infinite word over a finite alphabet that

a) avoids abelian squares, and

b) in any subword, the difference between the number of occurrences of the most and least frequent letter is bounded above by a constant?

Avoiding Arithmetic Squares

Solution

- Suppose we have an infinite word **w** over the alphabet $\Sigma := \{a_1, \ldots, a_k\}$ and $M \in \mathbb{N}$ such that

$$|\mathbf{w}[1..l]_{a_i} - \mathbf{w}[1..l]_{a_j}| \leq M$$

  for all $l \in \mathbb{N}, i, j \in \{1, \ldots, k\}$.
- Define f that maps $i \in \mathbb{N}$ to the $(k-1)$-tuple $(|\mathbf{w}[1..i]|_{a_1} - |\mathbf{w}[1..i]|_{a_2}|, |\mathbf{w}[1..i]|_{a_1} - |\mathbf{w}[1..i]|_{a_3}|, \ldots, |\mathbf{w}[1..i]|_{a_1} - |\mathbf{w}[1..i]|_{a_k}|)$.
- $f(i) \in [-M, M]^{k-1}$ and is integral $\forall i \in \mathbb{N}$.
- VDW: $\exists a, d$ such that $f(a) = f(a+d) = f(a+2d)$.

Avoiding Arithmetic Squares

Solution

- Suppose we have an infinite word **w** over the alphabet $\Sigma := \{a_1, \ldots, a_k\}$ and $M \in \mathbb{N}$ such that

$$|\mathbf{w}[1..l]_{a_i} - \mathbf{w}[1..l]_{a_j}| \leq M$$

  for all $l \in \mathbb{N}, i, j \in \{1, \ldots, k\}$.
- Define f that maps $i \in \mathbb{N}$ to the $(k-1)$-tuple $(|\mathbf{w}[1..i]|_{a_1} - |\mathbf{w}[1..i]|_{a_2}|, |\mathbf{w}[1..i]|_{a_1} - |\mathbf{w}[1..i]|_{a_3}|, \ldots, |\mathbf{w}[1..i]|_{a_1} - |\mathbf{w}[1..i]|_{a_k}|)$.
- $f(i) \in [-M, M]^{k-1}$ and is integral $\forall i \in \mathbb{N}$.
- VDW: $\exists a, d$ such that $f(a) = f(a+d) = f(a+2d)$.

Avoiding Arithmetic Squares

Solution

- Suppose we have an infinite word **w** over the alphabet $\Sigma := \{a_1, \ldots, a_k\}$ and $M \in \mathbb{N}$ such that

$$|\mathbf{w}[1..l]_{a_i} - \mathbf{w}[1..l]_{a_j}| \leq M$$

  for all $l \in \mathbb{N}, i, j \in \{1, \ldots, k\}$.
- Define f that maps $i \in \mathbb{N}$ to the $(k-1)$-tuple $(|\mathbf{w}[1..i]|_{a_1} - |\mathbf{w}[1..i]|_{a_2}|, |\mathbf{w}[1..i]|_{a_1} - |\mathbf{w}[1..i]|_{a_3}|, \ldots, |\mathbf{w}[1..i]|_{a_1} - |\mathbf{w}[1..i]|_{a_k}|)$.
- $f(i) \in [-M, M]^{k-1}$ and is integral $\forall i \in \mathbb{N}$.
- VDW: $\exists a, d$ such that $f(a) = f(a+d) = f(a+2d)$.

Avoiding Arithmetic Squares

Solution

- Suppose we have an infinite word $\mathbf{w}$ over the alphabet $\Sigma := \{a_1, \ldots, a_k\}$ and $M \in \mathbb{N}$ such that

$$|\mathbf{w}[1..l]_{a_i} - \mathbf{w}[1..l]_{a_j}| \leq M$$

  for all $l \in \mathbb{N}, i, j \in \{1, \ldots, k\}$.
- Define f that maps $i \in \mathbb{N}$ to the $(k-1)$-tuple $(|\mathbf{w}[1..i]|_{a_1} - |\mathbf{w}[1..i]|_{a_2}|, |\mathbf{w}[1..i]|_{a_1} - |\mathbf{w}[1..i]|_{a_3}|, \ldots,$ $|\mathbf{w}[1..i]|_{a_1} - |\mathbf{w}[1..i]|_{a_k}|)$.
- $f(i) \in [-M, M]^{k-1}$ and is integral $\forall i \in \mathbb{N}$.
- VDW: $\exists a, d$ such that $f(a) = f(a+d) = f(a+2d)$.

Avoiding Arithmetic Squares

Solution (continued)

- $f(a) = f(a+d)$ implies

$$\mathbf{w}[1..a]_{a_1} - \mathbf{w}[1..a]_{a_2} = \mathbf{w}[1..a+d]_{a_1} - \mathbf{w}[1..a+d]_{a_2}.$$

- Observe that

$$\mathbf{w}[1..a+d]_{a_1} = \mathbf{w}[1..a]_{a_1} + \mathbf{w}[a+1..a+d]_{a_1}.$$

$$\Rightarrow \mathbf{w}[a+1..a+d]_{a_1} - \mathbf{w}[a+1..a+d]_{a_2} = 0;$$

Avoiding Arithmetic Squares

Solution (continued)

- $f(a) = f(a + d)$ implies

$$\mathbf{w}[1..a]_{a_1} - \mathbf{w}[1..a]_{a_2} = \mathbf{w}[1..a+d]_{a_1} - \mathbf{w}[1..a+d]_{a_2}.$$

- Observe that

$$\mathbf{w}[1..a+d]_{a_1} = \mathbf{w}[1..a]_{a_1} + \mathbf{w}[a+1..a+d]_{a_1}.$$

$$\Rightarrow \mathbf{w}[a+1..a+d]_{a_1} - \mathbf{w}[a+1..a+d]_{a_2} = 0;$$

Avoiding Arithmetic Squares

Solution (continued)

- Similarly, we have

$$\mathbf{w}[a+1..a+d]_{a_1} = \mathbf{w}[a+1..a+d]_{a_i} \ \forall i \in \{3, \dots, k\}.$$

- Since $\sum_{i=1}^{k} \mathbf{w}[a+1..a+d]_{a_i} = d$,

$$\mathbf{w}[a+1..a+d]_{a_i} = \frac{d}{k} \ \forall i \in \{1, \dots, k\}.$$

- Analogously, $f(a+d) = f(a+2d) \Rightarrow$ $\mathbf{w}[a+d+1..a+2d]_{a_i} = \frac{d}{k} \ \forall i \in \{1, \dots, k\}$.
- $\mathbf{w}[a+1..a+2d]$ is an abelian square. $\square$

Avoiding Arithmetic Squares

Solution (continued)

- Similarly, we have

$$\mathbf{w}[a+1..a+d]_{a_1} = \mathbf{w}[a+1..a+d]_{a_i} \ \forall i \in \{3,\ldots,k\}\,.$$

- Since $\sum_{i=1}^{k} \mathbf{w}[a+1..a+d]_{a_i} = d$,

$$\mathbf{w}[a+1..a+d]_{a_i} = \frac{d}{k} \ \forall i \in \{1,\ldots,k\}\,.$$

- Analogously, $f(a+d) = f(a+2d) \Rightarrow \mathbf{w}[a+d+1..a+2d]_{a_i} = \frac{d}{k} \ \forall i \in \{1,\ldots,k\}$.
- $\mathbf{w}[a+1..a+2d]$ is an abelian square. $\square$

Avoiding Arithmetic Squares

Solution (continued)

- Similarly, we have

$$\mathbf{w}[a+1..a+d]_{a_1} = \mathbf{w}[a+1..a+d]_{a_i} \ \forall i \in \{3,\ldots,k\}\,.$$

- Since $\sum_{i=1}^{k} \mathbf{w}[a+1..a+d]_{a_i} = d$,

$$\mathbf{w}[a+1..a+d]_{a_i} = \frac{d}{k} \ \forall i \in \{1,\ldots,k\}\,.$$

- Analogously, $f(a+d) = f(a+2d) \Rightarrow$ $\mathbf{w}[a+d+1..a+2d]_{a_i} = \frac{d}{k} \ \forall i \in \{1,\ldots,k\}$.
- $\mathbf{w}[a+1..a+2d]$ is an abelian square. $\square$

Avoiding Arithmetic Squares

Solution (continued)

- Similarly, we have

$$\mathbf{w}[a+1..a+d]_{a_1} = \mathbf{w}[a+1..a+d]_{a_i} \ \forall i \in \{3, \dots, k\} .$$

- Since $\sum_{i=1}^{k} \mathbf{w}[a+1..a+d]_{a_i} = d$,

$$\mathbf{w}[a+1..a+d]_{a_i} = \frac{d}{k} \ \forall i \in \{1, \dots, k\} .$$

- Analogously, $f(a+d) = f(a+2d) \Rightarrow$ $\mathbf{w}[a+d+1..a+2d]_{a_i} = \frac{d}{k} \ \forall i \in \{1, \dots, k\}$.
- $\mathbf{w}[a+1..a+2d]$ is an abelian square. $\square$

Avoiding Arithmetic Squares

Proposition 1

Suppose

1. **w** is an infinite word over a finite alphabet, and
2. in any prefix of **w**, the difference of the number of occurrences of the most frequent letter and that of the least frequent letter is bounded by a constant.

Then **w** contains arbitrarily many consecutive blocks of the same length that are permutations of each other.

Avoiding Arithmetic Squares

Definition

For any finite word x on an alphabet $\Sigma := \{a_1, \ldots, a_k\}$, the **Parikh Map** of x is

$$\Psi(x) := (|x|_{a_1}, |x|_{a_2}, \ldots, |x|_{a_k}).$$

Avoiding Arithmetic Squares

Proposition 2

For any infinite word **w** over a finite alphabet of size k, if there exists $\vec{v} \in \mathbb{Q}^k$ and $M \in \mathbb{N}$ such that

$$\{\Psi(\mathbf{w}[1..l]) - l\vec{v} : l \in \mathbb{N}\} \subseteq [-M, M]^k,$$

then for any $p \geq 2$, there exist a, d such that

$$\Psi(\mathbf{w}[a + (j-1)d + 1..a + jd]) = d\vec{v} \quad \forall j \in \{1, \ldots, p\}.$$

Remark

Proposition 1 is the special case when $\vec{v} = (\frac{1}{k}, \frac{1}{k}, \ldots, \frac{1}{k})$.

Avoiding Arithmetic Squares

Other Relevant Facts

- Not known : whether there is an infinite sequence over a finite subset of $\mathbb{Z}$ that avoids 3 consecutive blocks of the same length and the same sum.
- There is a binary sequence that avoids abelian 4th powers (Dekking, 1979).

Avoiding Arithmetic Squares

Other Relevant Facts

- Not known : whether there is an infinite sequence over a finite subset of $\mathbb{Z}$ that avoids 3 consecutive blocks of the same length and the same sum.
- There is a binary sequence that avoids abelian 4th powers (Dekking, 1979).

Future Work

Settle the arithmetic squares problem!

- Positively – construct a sequence that is the fixed point of a morphism and verify that it has such properties?
- Negatively – another application of Van der Waerden's theorem?

Future Work

Settle the arithmetic squares problem!

- Positively – construct a sequence that is the fixed point of a morphism and verify that it has such properties?
- Negatively – another application of Van der Waerden's theorem?

Future Work

Settle the arithmetic squares problem!

- Positively – construct a sequence that is the fixed point of a morphism and verify that it has such properties?
- Negatively – another application of Van der Waerden's theorem?