Large Sets Avoiding Prescribed Differences

Paul Raff

February 05, 2009

Aesthetics

Throughout this talk, we will refer to many sets of integers, in all kinds of places. We will use a shorthand notation. Therefore, for example, instead of

$$f_{\{1,4\}}(n, \{2, 9\})$$

we will write

$$f_{1\centerdot 4}(n, 2 \centerdot 9)$$

and instead of

$$f_{\{\{1,2\},\{2,4\}\}}(n, \{\{1, 3, 5\}, \{2, 4\}\})$$

we will write

$$f_{\{1\centerdot 2\ ,\ 2\centerdot 4\}}(n, \{1 \centerdot 3 \centerdot 5\ ,\ 2 \centerdot 4\}).$$

Background - Coding Theory

We are interested in building words over the alphabet $\{x, y\}$ in a special way. For an integer m, let

$$\mathcal{A}_m = \{x^i y x^j \mid i + j + 1 \leq m\}.$$

(Recall that x^i is shorthand - for example, $x^4 = xxxx$.)

Definition. $A \subseteq \mathcal{A}_m$ is a *code* if any word created from the concatenation of elements of A can be decomposed uniquely. Algebraically speaking, A is a code if the free monoid $A^\star$ generated by A exhibits unique factorization.

Examples

For any m, the set

$$D_m = \{x^i y x^{m-i-1} \mid 0 \leq i < m\}$$

is a code.

However, the set $\{xy, y, yx\}$ is *not* a code, for

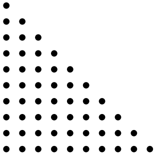
$$yxy = y \cdot xy = yx \cdot y.$$

The Triangle Conjecture

In 1981, D. Perrin and M. P. Schützenberger gave the following conjecture, now called the *Triangle Conjecture*:

Conjecture. If $A \subseteq \mathcal{A}_m$ is a code, then $|A| \leq m$.

Why Triangle Conjecture? Viewed graphically, the elements of $\mathcal{A}_m$ form a triangle:



A Counterexample

The Triangle Conjecture did not last long - less than two years after the conjecture was published, P. Shor provided a counterexample:

y	x^3y	x^8y	$x^{11}y$
yx	x^3yx^2	x^8yx^2	$x^{11}yx$
yx^7	x^3yx^4	x^8yx^4	$x^{11}yx^2$
yx^{13}	x^3yx^6	x^8yx^6	
yx^{14}			

Proof

Suppose a word of length 2 could be decomposed in two unique ways:

$$x^i yx^{j_1} \cdot x^{i_2} yx^j = x^i yx^{j_3} \cdot x^{i_4} yx^j$$

We must then have $j_1 + i_2 = j_3 + i_4$, or $i_2 - i_4 = j_3 - j_1$.

i_2 and i_4 were prefixes, so $i_2, i_4 \in \{0, 3, 8, 11\}$. Additionally, j_1 and j_3 were suffixes of words with the same prefix. Therefore, $j_1, j_3 \in \{0, 1, 7, 13, 14\}$, $j_1, j_3 \in \{0, 2, 4, 6\}$, or $j_1, j_3 \in \{0, 1, 2\}$.

Differences

However, denoting $\Delta(a_1, a_2, \ldots, a_n)$ as the difference set of $\{a_1, a_2, \ldots, a_n\}$, we have

$$\begin{aligned}
\Delta(0,3,8,11) &= \{3,5,8,11\} \\
\Delta(0,1,7,13,14) &= \{1,6,7,12,13,14\} \\
\Delta(0,2,4,6) &= \{2,4,6\} \\
\Delta(0,1,2) &= \{1,2\}
\end{aligned}$$

Since $\Delta(0,3,8,11)$ is disjoint from the other difference sets, our proof is complete.

Consequences

We can define γ as

$$\gamma = \sup_m \left(\frac{\text{size of largest code in } \mathcal{A}_m}{m} \right).$$

The Triangle Conjecture can then be restated as saying $\gamma \leq 1$.

By counting all words created from $\mathcal{A}_m$, G. Hansel showed that $\gamma \leq 1 + \frac{1}{\sqrt{2}}$. Hence, the current state of the Triangle Conjecture is

$$\frac{16}{15} \leq \gamma \leq 1 + \frac{1}{\sqrt{2}}.$$

Finding Large Sets Avoiding Differences

The key to Shor's proof was finding large subsets of [15], [12], [7] and [4] that avoided differences in $\Delta(0, 3, 8, 11) = \{3, 5, 8, 11\}$.

Definition. Given a set Δ, $f_\Delta(n)$ is defined as the size of the largest subset $X \subseteq [n]$ such that X avoids differences in Δ. We can extend this definition to $f_\Delta(I)$, where I is any set of integers.

Rephrased: Words Avoiding Patterns

We can rephrase the problem as a problem of pattern avoidance in words by viewing a subset of $[n]$ as a n-length $0/1$ string.

Example. Avoiding differences in $\{2, 3\}$ is the same as avoiding the pattern $\{1 \bullet 1, 1 \bullet \bullet 1\}$, where $\bullet$ can be either 0 or 1. The set $\{1, 2, 6, 7\}$ avoids the differences in $\{2, 3\}$, and the word 1100011 avoids the patterns in $\{1 \bullet 1, 1 \bullet \bullet 1\}$.

Rephrased: Circulant Graphs

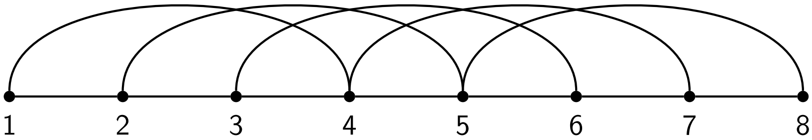
We can also rephrase the problem in terms of circulant graphs, which are very important structures in graph theory.

Definition. Given a set S of positive integers, the *unhooked circulant graph on n vertices* $UC_S(n)$ is the graph with vertex set $[n]$ and

$$i \sim j \iff |i - j| \in S.$$

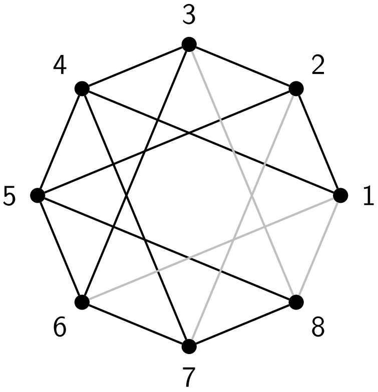
An Example

The following is $UC_{1,3}(8)$:



Another Example

Unhooked circulant graphs are very closely related to standard circulant graphs, $C_S(n)$. Here is $C_{1,3}(8)$:



The Connection

It is clear that finding $f_\Delta(n)$ is the same as finding the independence number of $UC_\Delta(n)$.

However: It is well-known that the problem of finding the clique number in general graphs is NP-complete. In 1998, Codenotti et al. showed that it is still NP-hard when reduced to considering only circulant graphs. As far as I know, a similar result has not been shown explicitly for unhooked circulant graphs, but it is likely that it is also NP-hard.

A Very Useful Recurrence

We introduce another parameter, S, which denotes elements to avoid outright. Therefore,

$$f_{\Delta}(I, S) = f_{\Delta}(I \setminus S).$$

Theorem. If $1 \in S$ then

$$f_{\Delta}(n, S) = f_{\Delta}(n-1, S-1).$$

Otherwise,

$$f_{\Delta}(n, S) = \max\{f_{\Delta}(n-1, S-1), 1 + f_{\Delta}(n-1, \Delta \cup (S-1))\}.$$

Where

$$S - 1 = \{s - 1 \mid s \in S\}.$$

Proof

The proof is based on the following:

Claim. If $1 \notin I$, then the map $X \mapsto X - 1$ is a cardinality-preserving bijection between subsets of I that avoids differences in Δ and elements in S and subsets of $I - 1$ that avoids differences in Δ and elements in $S - 1$.

Furthermore, if $1 \in I$, then the map $X \mapsto X - 1$ is a bijection between subsets of I that avoids differences in Δ and elements in S and subsets of $I - 1$ that avoids differences in Δ and elements in $\Delta \cup (S - 1)$.

Proof - Continued

From the claim, the first part is immediate, for if $1 \in S$, then

$$f_\Delta(n, S) = f_\Delta([2 \ldots n], S) = f_\Delta([1 \ldots n{-}1], S{-}1) = f_\Delta(n{-}1, S{-}1)$$

For the second part of the proof, we note that

$$f_\Delta(n, S) = \max\{\text{sets that don't contain } 1, \text{sets that do contain } 1\}.$$

Using The Recurrence

We can define the Δ-*closure* of a set S to be the smallest family $\mathfrak{S} \ni S$ that satisfies the following:

$$X \in \mathfrak{S}, 1 \notin X \Rightarrow X - 1 \in \mathfrak{S}$$
$$X \in \mathfrak{S}, 1 \in X \Rightarrow X - 1 \in \mathfrak{S}, \Delta \cup (X - 1) \in \mathfrak{S}$$

The closure contains the other parameters S' that are necessary to compute $f_\Delta(n, S)$.
We can graphically view the closure.

Investigating The Sequences

As an example, consider the first few terms of the sequence $f_{3.8.10}(n)$:

$$1, 2, 3, 3, 3, 3, 4, 5, 5, 5, 5, 5, 6, 6, 6, 7, 7, 8, 8, 8, 9, 9, 9, 9, 10,$$
$$11, 11, 11, 12, 12, 12, 12, 12, 13, 13, 14, 14, 15, 15, 16, 16, 16, 16, 17,$$
$$17, 17, 18, 18, 19, 19, 20, 20, 20, 20, 20$$

Any Pattern?

With clever structuring and coloring of the terms, a pattern emerges.

Definition. A sequence of integers is *(eventually) pseudoperiodic* if the sequence of successive differences is (eventually) periodic.

Theorem (Raff). For any Δ and S, the sequence $\{f_\Delta(n, S)\}$ is eventually pseudoperiodic.

Proof and Limitations

The proof is based on a standard finite-automata argument: the "program" to compute the sequence $\{f_\Delta(n, S)\}$ can be expressed as a finite automata, and it is then immediate that the sequence is eventually pseudoperiodic.

However, there is little known about specifics:

- How long is the period?
- How much does the sequence increase over a period?
- How long is the offset?

Consequences

Corollary. For every Δ and S, there is a rational $\alpha = \alpha_{\Delta,S}$ (or α_Δ if $S = \emptyset$) such that

$$\lim_{n \to \infty} \frac{f_\Delta(n, S)}{n} = \alpha.$$

α will be expressed as a potentially unreduced fraction r/s, where s is the period length.

Finding α quickly is probably a hopeless problem, but some special-case results are known, specifically:

Theorem. If $\Delta = [i, i+1, \ldots, i+k]$, then $\alpha_\Delta = \frac{i}{2i+k}$.

Extensions - Part 1

By extending what it means to avoid a difference and avoid elements, we can go further:

Definition. If $D = \{i_i, \ldots, i_k\}$ is a set of integers with $i_1 < i_2 < \cdots < i_k$, then a set X *avoids generalized differences in* D if

$$x \in X \rightarrow \{x, x + i_1, x + i_2, \ldots, x + i_k\} \not\subseteq X.$$

Similarly, if S is a set of integers, then X *avoids* S *generally* if $X \not\subseteq S$.

To achieve a similar recurrence, we need to extend and modify an operator. If $\mathfrak{S}$ is a family of sets, then

$$\mathfrak{S} - 1 = \{S - 1 \mid S \in \mathfrak{S}\}$$
$$(\mathfrak{S} - 1)^\star = \{S - 1 \mid S \in \mathfrak{S}, 1 \notin S\}$$

A New Recurrence

We can then extend the definition of f: for example, $f_{\{1.2,2.4\}}(n)$ is the size of the largest subset of $[n]$ that avoids three-term arithmetic sequences of difference 1 and 2.

Theorem. If $\mathfrak{D}$ and $\mathfrak{S}$ are families of sets:
If $\{1\} \in \mathfrak{S}$, then

$$f_{\mathfrak{D}}(n, \mathfrak{S}) = f_{\mathfrak{D}}(n-1, (\mathfrak{S}-1)^{\star}).$$

If $\{1\} \notin \mathfrak{S}$, then

$$f_{\mathfrak{D}}(n, \mathfrak{S}) = \max\{f_{\mathfrak{D}}(n-1, (\mathfrak{S}-1)^{\star}), 1 + f_{\mathfrak{D}}(n-1, \mathfrak{S}-1)\}.$$

An Application - Experimental Roth's Theorem

We can use the extended recurrence to find the sizes large sets of integers that avoid 3-term arithmetic progressions.

max difference to avoid	α
1,2	$2/3$
3	$4/8$
4,5,6,7,8	$4/9$
9	$4/10$
10	$4/11$
11	$8/24$
12	$56/177$
13,14,15,16,17	$6/19$

How To Be Sure?

The ratios given on the previous page were obtained by analyzing the sequences and looking for the pseudoperiodic pattern. We can obviously only compute a finite number of terms - how can we be certain that we have the actual pattern instead of being part of a larger pattern?

PROVE IT!

A Cyclic Extension

What if we want to avoid differences modulo n? We can define $f_\Delta^c(n)$ to be the size of the largest subset of $[n]$ that avoids differences *modulo* n in Δ.

There is a similar recurrence for the cyclic extension, and everything stated previously about the structure of the sequence $\{f_\Delta^c(n)\}$ holds true for $\{f_\Delta(n)\}$, with the following exception:

$f_\Delta(n+1)$ *may be smaller than* $f_\Delta(n)$.

Conjectures

Since the Triangle Conjecture has been disproved, I offer the following asymptotic version:

Conjecture. If I is a set and X is the difference set of I,

$$\alpha_X \leq \frac{1}{I}.$$

Another conjecture:

Conjecture. For any Δ with $|\Delta| \geq 2$, the period of $\{f_\Delta(n)\}$ is less than or equal to the sum of the elements of Δ.

Future Work

- Find some sort of bounds on the period of $\{f_\Delta(n)\}$ in terms of Δ.
- Find more recurrences - specifically, recurrences that involve changing Δ.
- Investigate connections between $f_\Delta(n)$ and $f_\Delta^c(n)$.

Thanks!

Thanks for listening to the talk. Voltaire said:

> *The more you know, the less sure you are.*

Contact me to learn more: `praff@math.rutgers.edu`.

Check my website (and OEIS) shortly for preprints and results:

`http://math.rutgers.edu/~ praff`