Sumset Problem

Ari Fiorino, Eric Li, Katrina Liu, Justin Sun

Carnegie Mellon University

August 26, 2020

Introduction

Sumset:

A set A is a sumset if there exists a set S such that $S + S = \{a + b | a, b \in S\} = A$

Sumset Problem:

If we have a set $A \subseteq \mathbb{Z}$, is A a sumset?

Goal:

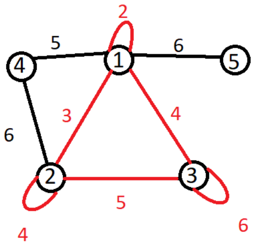
Determine the complexity of the sumset problem.

Complexity Theoretic Approach: Cayley Sum Graphs

Definition:

A Cayley sum graph is a graph where all the vertices are assigned integers and there is a set A where two vertices are connected if and only if their sum is in A.

A={2,3,4,5,6}
S={1,2,3}



Hardness of Cayley-Clique

Cayley Clique Problem:

If we have a Cayley Graph G, what is the maximum clique of the graph?

Theorem (Codenetti, Gerace, Vigna, 1998):

The Cayley Clique Problem is NP-Hard.

Maximum Sumset:

Given a set A. Find the maximum size set S where $S + S \subset A$.

Conclusion

Maximum Sumset is NP-Hard.

Our Approach

Labeled Clique Problem:

Given an edge labeled graph G, is there a clique of the graph that contains all of the labels?

Distinct Labeled Clique Problem:

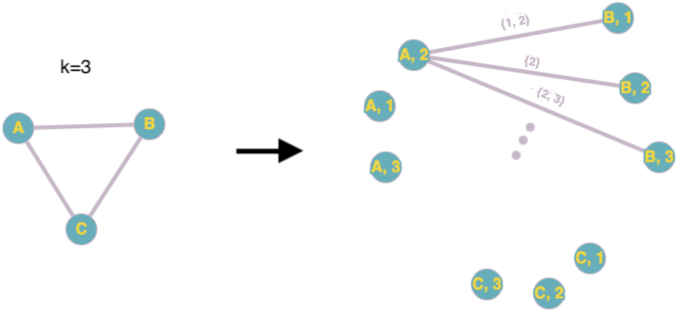
Given a labeled graph G where for every vertex in G, the edges connected to it has distinct labels, what is the maximum clique of the graph that contains all of the labels?

Three reductions:

- Max Clique $\leq$ Labeled Clique
- Labeled Clique $\leq$ Distinct Labeled Clique
- Distinct Labeled Clique $\leq$ Cayley Labelled Clique

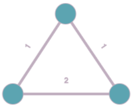
Claim: Max Clique $\leq$ Labeled Clique

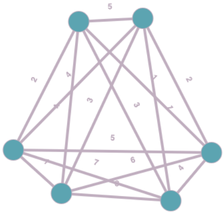
Say we start with a graph G and a number k. The construction is to blow up each vertex in G into k vertices in H. We connect two vertices in H iff their supervertices are connected in G. We connect the vertices with the edge label for the set containing the numbers for each vertex.



Conjecture: Labeled Clique $\leq$ Distinct Labeled Clique

Let G be an instance of the labeled clique problem and is $d-$regular.
Each vertex $v \in G$ gets blown up into a d-clique. For each edge, we connect each vertex in each clique to each other. We can can organize our edge labelling in a latin square or a series of SAT statements.





Other Complexity Related Approaches

- If we have a clique of size $\approx \sqrt{d} \log d$, then it will have all the labels with high probability.
- Cayley Clique $\leq$ Cayley Labelled Clique
- Adding more constraints to the problem

A Naive Approximation Algorithm

Suppose $A = S + S$ where $S = \{s_1, s_2, \ldots, s_n\}$ with $s_1 < \ldots < s_n$. Then we must have $s_1 + s_1 = \min(A)$ and $s_n + s_n = \max(A)$ so $\frac{\min(A)}{2}, \frac{\max(A)}{2} \in S$. If $x \in S$, then $x + \frac{\min(A)}{2}, x + \frac{\max(A)}{2} \in A$, so we take $S' = (A - \frac{\min(A)}{2}) \cap (A - \frac{\max(A)}{2})$ as an approximation of S.

This also ensures $S' \supseteq S$ and subsequently $S' + S' \supseteq A$.

Numerical Results

To verify the strength of this algorithm, we implemented it in C++ and analyzed the error ratio $\frac{|S'+S'|-|A|}{|A|}$ for $S \subseteq \{0, 1, 2, \ldots, 16\}$ and $A = S + S$.

As the range of the potential set increased, this error term grows larger. For example, when $S \subseteq \{0, 1, \ldots, 8\}$ the term never exceeds around 0.22. However when $S \subseteq \{0, 1, \ldots, 16\}$ the error term goes up to 0.5. For wider ranges this factor grows even further, so the algorithm becomes less practical.

Another Approach: Polynomials

Given $A = \{a_1, a_2, ..., a_n\} \subseteq \mathbb{Z}$, we want to find a set $S = \{s_1, s_2, ..., s_m\}$ and positive integers $\alpha_1, ..., \alpha_n$ such that

$$(\sum_{i=1}^{m} x^{s_i})^2 = \sum_{i=1}^{m}\sum_{j=1}^{m} x^{s_i+s_j} = \sum_{k=1}^{n} \alpha_k x^{a_k}$$

Then, $A = S + S$ only if the subspace generated by $\{x^{a_1}, x^{a_2}, \ldots, x^{a_n}\}$ in the vector space of integer polynomials $\mathbb{Z}[x]$ contains a perfect square element with positive coefficients.

Another algorithmic approach

From this, we can conclude that if $A = S + S$, then $\exists \alpha_1, \ldots, \alpha_n > 0$ s.t.

$$(\sum_{i=1}^{m} b^{s_i})^2 = \sum_{i=1}^{m} \sum_{j=1}^{m} b^{s_i+s_j} = \sum_{k=1}^{n} \alpha_k b^{a_k}$$

for all integers b.

Plugging in $b = 1$ yields:

$$\sum_{i=1}^{n} \alpha_i = m^2$$

This suggests other numerical methods such as integer programming.

Approaches to continue

- Show that Distinct Labeled Clique $\leq$ Sumset.
- Improve bounds on the approximation algorithm.
- Find more useful results on perfect square polynomials, or develop integer programming algorithms.