Approximating Sumset Size

Shivam Nadimpalli
(Columbia)

Joint work with

Anindya De
(Penn)

Rocco Servedio
(Columbia)
Approximating **Sumset** Size

Shivam Nadimpalli  
(Columbia)

Joint work with

Anindya De  
(Penn)

Rocco Servedio  
(Columbia)
Definition: Given an abelian group \((G, +)\) and a subset \(A \subseteq G\), we define the sumset \(A + A\) as

\[
A + A := \{a + b : a, b \in A\}.
\]

- Note \(A + A \neq 2A := \{a + a : a, a \in A\}\).
- Fundamental object of study in additive combinatorics.
Easy Example

\[ A = \{1, 3, \ldots, 99\} \]

\[ A + A = \{0, 2, \ldots, 98\} \]

- Note that \( |A| = |A + A| \).
- \( A \) is a coset of the subgroup of even residues modulo 100.
Why Sumset Size?

**Easy Exercise:** For $A \subseteq G$, if $|A| = |A + A|$, then $A = x + H$ for some subgroup $H \leq G$ and $x \in G$.

Robustifications & Variants

Freiman–Rusza, Plünneke–Rusza, Balog–Szemerédi–Gowers, etc.
A Natural Question

Question: Given query access to \( A \subseteq G \), what is \(|A+A|/|G|\) up to an error of \( \pm \epsilon \)?

This work: \( \mathbb{F}_2^n \)
A Natural Question over $\mathbb{F}_2^n$

Question: Given query access to $A \subseteq \mathbb{F}_2^n$ and writing

$$\text{Vol}(A) := \frac{|A|}{2^n},$$

what is $\text{Vol}(A + A)$ up to an error of $\pm \epsilon$?

- Cost measure: number of queries (as a function of $n$ and $\epsilon$).
- At first glance: To confirm $z \notin A + A$, have to check that at least one of $x, y \notin A$ for the $2^n$ pairs $(x, y)$ satisfying $x + y = z$. 
No Query-Efficient Algorithm over $\mathbb{F}_2^n$

$A = \emptyset$

$\text{Vol}(A + A) = 0$

$A$ is a random set of size $2^{0.51n}$

$\text{Vol}(A + A) \geq 1 - \exp(-n)$ w.h.p.

Need $\Omega(2^{0.49n})$ queries to distinguish $A$ from $A$. 
Refining The Original Question

Original Question: Given query access to \( A \subseteq \mathbb{F}_2^n \) and writing

\[
\text{Vol}(A) := \frac{|A|}{2^n},
\]

what is \( \text{Vol}(A + A) \) up to an error of \( \pm \varepsilon \)?

- Adding a small (random) collection \( R \subseteq \mathbb{F}_2^n \) of \( 2^{0.51n} \) elements to \( A \) can blow up \( \text{Vol}(A + A) \) to almost 1.

- Natural relaxation: Output \( \text{Vol}(A' + A') \) for set \( A' \subseteq A \) that is close to \( A \).
An Analogous Situation: Approximating Surface Area

“Given a nice convex set such as a sphere, one can add a very thin tentacle to it with negligible volume but arbitrarily large surface area.”

– Kothari, Nayyeri, O’Donnell, Wu (2014)
Refining The Original Question

Original Question: Given query access to \( A \subseteq \mathbb{F}_2^n \) and writing

\[
\text{Vol}(A) := \frac{|A|}{2^n},
\]

what is \( \text{Vol}(A + A) \) up to an error of \( \pm \varepsilon \)?

- Adding a small (random) collection \( R \subseteq \mathbb{F}_2^n \) of \( 2^{0.51n} \) elements to \( A \) can blow up \( \text{Vol}(A + A) \) to almost 1.

- Natural relaxation: Output \( \text{Vol}(A' + A') \) for set \( A' \subseteq A \) that is close to \( A \).
**New Question**: Given query access to $A \subseteq \mathbb{F}_2^n$ and writing

$$\text{Vol}(A) := \frac{|A|}{2^n},$$

what is $\text{Vol}(A' + A')$ up to an error of $\pm \varepsilon$ for some $A' \subseteq A$ such that

$$\text{Vol}(A \setminus A') \leq \varepsilon?$$
The Question We Consider

New Goal: Output $\text{Vol}(A' + A')$ instead of $\text{Vol}(A + A)$. 

$\text{Vol}(\Box) \leq \varepsilon$
Revisiting Our Earlier Example

$A = \emptyset$

$\text{Vol}(A + A) = 0$

$A$ is a random set of size $2^{0.51n}$

$\text{Vol}(A + A) \geq 1 - \exp(-n)$ w.h.p.

For $\varepsilon \geq 2^{-0.49n}$, simply output $\text{Vol}(A' + A') = 0$. 
New Question: Given query access to \( A \subseteq \mathbb{F}_2^n \) and writing

\[
\text{Vol}(A) := \frac{|A|}{2^n},
\]

what is \( \text{Vol}(A' + A') \) up to an error of \( \pm \varepsilon \) for some \( A' \subseteq A \) such that

\[
\text{Vol}(A \setminus A') \leq \varepsilon?
\]

**Main Theorem:** Can be done using \( O_{\varepsilon}(1) \) queries to \( A \).

(Bonus: Outputs an exact oracle to \( A' \) and an approximate oracle to \( A' + A' \).)
Proof Sketch

Almost all of \( \mathbb{F}_2^n \)

**Ingredient 1**: “Non-tiny” random-like sets have “large” sumsets.

**Ingredient 2**: Green’s Regularity Lemma.

Need an algorithmic version
Green’s Regularity Lemma

- Decomposes $\mathbb{F}_2^n$ into translates of $H \leq \mathbb{F}_2^n$ such that:
  - $H \cong \mathbb{F}_2^{n-k}$ where $k$ is does not depend on $n$.
  - $A \cap (x + H)$ is “random-like,” i.e. has small Fourier coefficients.

- Made algorithmic by closely following the original proof and using the Goldreich–Levin algorithm.
Defining $A'$: Iterate through $2^k$ cosets of $H$:

- If $|A \cap (x + H)| \leq \varepsilon \cdot 2^{n-k}$, then set $A' \cap (x + H) = \emptyset$.
- Else set $A' \cap (x + H) = A$.

Approximately Defining $A' + A'$: If $A'$ intersects with cosets $x + H, y + H$,

$$(A' + A') \cap (x + y + H) \approx x + y + H.$$
Obtaining $O_\varepsilon(1)$ Query Complexity

• Explicitly obtaining a description of the subspace $H$ necessarily requires a number of queries that scales at least linearly in $n$.
• Require implicit versions of aforementioned algorithms.
  – For Goldreich–Levin: Equivalent to being a local list corrector for the Hadamard code.
Conclusion & Future Directions

• Our approach extends to estimate $\text{Vol}(A + B)$ and $\text{Vol}(A + \ldots + A)$ for $A, B \subseteq \mathbb{F}_2^n$.

• Generalizing to groups other than $\mathbb{F}_2^n$?
  • Green’s Regularity Lemma does hold for arbitrary abelian groups.
  • Implicitly finding significant Fourier coefficients?
Thanks for listening! Questions?