Asymptotic growth of codes and related combinatorial problems

Stefano Della Fiore
Supervisor: Prof. Marco Dalai

Università degli studi di Brescia

27 June, 2023

Overview

1. Asymptotic growth of codes
   - (b, k)-hash codes
   - Codes for multimedia fingerprinting
2. Related combinatorial problems
   - Erdős Sum-Distinct problem
   - Sequenceability of abelian groups

q-ary codes

Definition

A q-ary **code** C of size M and length n is a subset of $\{0, 1, \ldots, q-1\}^n$. The elements of $C = \{x_1, x_2, \ldots, x_M\}$ are called the **codewords** of C.

Example (4-ary code)

$$\begin{array}{ll}
x_1 & \cdots 0\ 2\ 0\ 2\ 3\ 1\ 0\ 2\ 0\ 2 \cdots \\
x_2 & \cdots 2\ 3\ 1\ 0\ 1\ 1\ 2\ 3\ 1\ 0 \cdots \\
x_3 & \cdots 2\ 3\ 3\ 2\ 1\ 2\ 2\ 3\ 3\ 2 \cdots \\
 & \vdots \\
x_M & \cdots 1\ 0\ 3\ 2\ 3\ 0\ 0\ 2\ 1\ 1 \cdots
\end{array}$$

Codes having some combinatorial properties

We are going to see codes (or related structures) where groups of codewords have some combinatorial properties.

Codes where the symbols in at least one coordinate have some properties.

Example (trifferent code)

Codewords

$$
\begin{array}{ll}
x_1 & \cdots \boxed{0}\,1\,0\,1\,0\,1\,0\,1\,0\,1 \cdots \\
x_2 & \cdots \boxed{1}\,1\,1\,1\,0\,1\,1\,1\,1\,1 \cdots \\
x_3 & \cdots \boxed{2}\,1\,0\,1\,2\,1\,1\,1\,0\,1 \cdots \\
x_4 & \cdots 0\,1\,2\,1\,1\,1\,2\,1\,0\,1 \cdots \\
 & \vdots
\end{array}
$$

Codes having some combinatorial properties

We are going to see codes (or related structures) where groups of codewords have some combinatorial properties.

Codes where the sums between pairs of codewords have some properties.

Example (binary $\overline{2}$-separable code)

Codewords	Sum
$x_1 \cdots 0\,1\,0\,0\,1\,1\,1\,0\,0\,1 \cdots$	Sum $x_1 + x_2$: $\cdots 1\,1\,0\,1\,2\,2\,2\,0\,0\,2 \cdots$
$x_2 \cdots 1\,0\,0\,1\,1\,1\,1\,0\,0\,1 \cdots$	Sum $x_1 + x_3$:
$x_3 \cdots 0\,0\,0\,0\,0\,0\,1\,0\,1\,0 \cdots$	$\cdots 0\,1\,0\,0\,1\,1\,2\,0\,1\,1 \cdots$
$\vdots$	

The sums are performed over the integers.

Codes having some combinatorial properties

We are going to see codes (or related structures) where groups of codewords have some combinatorial properties.

Sequences (codes of length 1) where all the subset sums have some properties.

Example (sum-distinct sequence)

$$(x_1, x_2, x_3) = (1, 2, 4)\,,$$

where

$$x_1 = 1, \qquad x_2 = 2, \qquad x_3 = 4,$$

$$x_1 + x_2 = 3, \qquad x_1 + x_3 = 5, \qquad x_2 + x_3 = 6$$

$$x_1 + x_2 + x_3 = 7$$

The sums are performed over the integers.

Codes having some combinatorial properties

We are going to see codes (or related structures) where groups of codewords have some combinatorial properties.

Sequences (codes of length 1) where the partial sums over a finite abelian group have some properties.

Example (sequence with distinct partial sums over $\mathbb{Z}_5$)

$$(x_1, x_2, x_3) = (1, 3, 4)\,,$$

where

$$x_1 = 1, \qquad x_1 + x_2 = 1 + 3 = 4, \qquad x_1 + x_2 + x_3 = 1 + 3 + 4 = 3$$

The sums are performed over $\mathbb{Z}_5$.

Outline

1. Asymptotic growth of codes
   - (b, k)-hash codes
   - Codes for multimedia fingerprinting
2. Related combinatorial problems
   - Erdős Sum-Distinct problem
   - Sequenceability of abelian groups

(b,k)-hash codes

Definition ((b,k)-hash code)

A b-ary code C is a (b,k)-hash code if for every k codewords there exists a coordinate in which all the symbols differ.

If $b = k = 3$ they are known as trifferent codes.

Example ($(3,3)$-hash code / trifferent code)

$$\begin{array}{ll} & \text{Codewords} \\ x_1 & \cdots 0\,1\,0\,1\,0\,1\,0\,1\,0\,1 \cdots \\ x_2 & \cdots 1\,1\,1\,1\,0\,1\,1\,1\,1\,1 \cdots \\ x_3 & \cdots 2\,1\,0\,1\,2\,1\,1\,1\,0\,1 \cdots \\ x_4 & \cdots 0\,1\,2\,1\,1\,1\,2\,1\,0\,1 \cdots \\ & \vdots \end{array}$$

(b, k)-hash codes

Definition ((b, k)-hash code)

A b-ary code C is a (b, k)-hash code if for every k codewords there exists a coordinate in which all the symbols differ.

If $b = k = 3$ they are known as trifferent codes.

Example ($(3, 3)$-hash code / trifferent code)

$$\begin{array}{cl} & \text{Codewords} \\ x_1 & \cdots \boxed{0}\,1\,0\,1\,0\,1\,0\,1\,0\,1 \cdots \\ x_2 & \cdots \boxed{1}\,1\,1\,1\,0\,1\,1\,1\,1\,1 \cdots \\ x_3 & \cdots \boxed{2}\,1\,0\,1\,2\,1\,1\,1\,0\,1 \cdots \\ x_4 & \cdots 0\,1\,2\,1\,1\,1\,2\,1\,0\,1 \cdots \\ & \vdots \end{array}$$

(b, k)-hash codes

Definition ((b, k)-hash code)

A b-ary code C is a (b, k)-hash code if for every k codewords there exists a coordinate in which all the symbols differ.

If $b = k = 3$ they are known as trifferent codes.

Example ($(3, 3)$-hash code / trifferent code)

$$\begin{array}{ll} & \text{Codewords} \\ x_1 & \cdots 0\,1\,0\,1\,0\,1\,0\,1\,0\,1 \cdots \\ x_2 & \cdots 1\,1\,\boxed{1}\,1\,0\,1\,1\,1\,1\,1 \cdots \\ x_3 & \cdots 2\,1\,\boxed{0}\,1\,2\,1\,1\,1\,0\,1 \cdots \\ x_4 & \cdots 0\,1\,\boxed{2}\,1\,1\,1\,2\,1\,0\,1 \cdots \\ & \vdots \end{array}$$

(b, k)-hash codes

Definition ((b, k)-hash code)

A b-ary code C is a (b, k)-hash code if for every k codewords there exists a coordinate in which all the symbols differ.

If $b = k = 3$ they are known as trifferent codes.

Example ($(3, 3)$-hash code / trifferent code)

$$
\begin{array}{ll}
 & \text{Codewords} \\
x_1 & \cdots 0\,1\,0\,1\,\boxed{0}\,1\,0\,1\,0\,1 \cdots \\
x_2 & \cdots 1\,1\,1\,1\,0\,1\,1\,1\,1\,1 \cdots \\
x_3 & \cdots 2\,1\,0\,1\,\boxed{2}\,1\,1\,1\,0\,1 \cdots \\
x_4 & \cdots 0\,1\,2\,1\,\boxed{1}\,1\,2\,1\,0\,1 \cdots \\
 & \vdots
\end{array}
$$

(b,k)-hash codes

Definition ((b,k)-hash code)

A b-ary code C is a (b,k)-hash code if for every k codewords there exists a coordinate in which all the symbols differ.

If $b = k = 3$ they are known as trifferent codes.

Example ($(3,3)$-hash code / trifferent code)

$$\begin{array}{ll} & \text{Codewords} \\ x_1 & \cdots 0\,1\,0\,1\,0\,1\,\boxed{0}\,1\,0\,1 \cdots \\ x_2 & \cdots 1\,1\,1\,1\,0\,1\,\boxed{1}\,1\,1\,1 \cdots \\ x_3 & \cdots 2\,1\,0\,1\,2\,1\,1\,1\,0\,1 \cdots \\ x_4 & \cdots 0\,1\,2\,1\,1\,1\,\boxed{2}\,1\,0\,1 \cdots \\ & \vdots \end{array}$$

A very challenging problem

Fredman and Komlós in 1985 posed the following question.

(b,k)-hashing problem

What is the asymptotic behaviour of the size of the largest (b,k)-hash code with length n as n goes to infinity?

Definition (Rate of a code)

Given a code C of length n

$$R = \frac{\log_2 |C|}{n}$$

We are interested in the asymptotic rate of (b,k)-hash codes of maximum cardinality, i.e.

$$R_{(b,k)} = \limsup R\,,$$

where the limsup is over all (b,k)-hash codes as n goes to infinity.

A very challenging problem

Fredman and Komlós in 1985 posed the following question.

(b,k)-hashing problem

What is the asymptotic behaviour of the size of the largest (b,k)-hash code with length n as n goes to infinity?

Definition (Rate of a code)

Given a code C of length n

$$R = \frac{\log_2 |C|}{n}$$

We are interested in the asymptotic rate of (b,k)-hash codes of maximum cardinality, i.e.

$$R_{(b,k)} = \limsup R\,,$$

where the limsup is over all (b,k)-hash codes as n goes to infinity.

A very challenging problem

Fredman and Komlós in 1985 posed the following question.

(b, k)-hashing problem

What is the asymptotic behaviour of the size of the largest (b, k)-hash code with length n as n goes to infinity?

Definition (Rate of a code)

Given a code C of length n

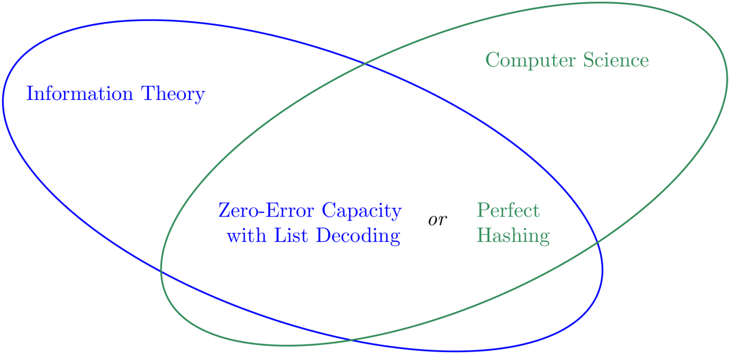
$$R = \frac{\log_2 |C|}{n}$$

We are interested in the asymptotic rate of (b, k)-hash codes of maximum cardinality, i.e.

$$R_{(b,k)} = \limsup R \,,$$

where the limsup is over all (b, k)-hash codes as n goes to infinity.

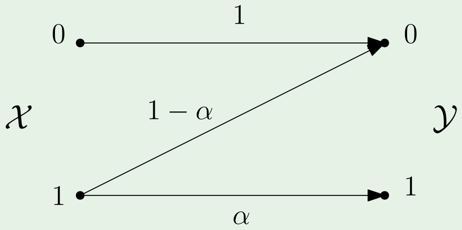
Information Theory and Computer Science interpretation



Discrete channels

A discrete channel is typically characterized by a bipartite graph $W = (\mathcal{X}, \mathcal{Y}, E)$ where $\mathcal{X}$ are the channel inputs, $\mathcal{Y}$ are the channel outputs and E is a subset of paris $(x, y) \in \mathcal{X} \times \mathcal{Y}$ that represents the channel links.

Example (Z-channel)

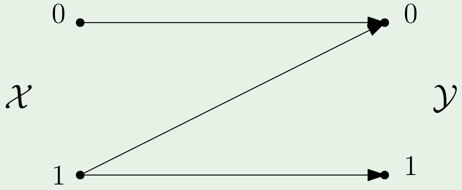


We note that $(x, y) \in E$ if and only if y can be received at the channel output when x is transmitted over the channel.

Discrete channels

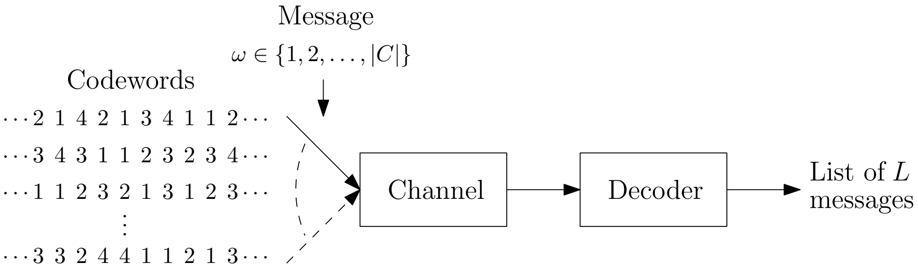
A discrete channel is typically characterized by a bipartite graph $W = (\mathcal{X}, \mathcal{Y}, E)$ where $\mathcal{X}$ are the channel inputs, $\mathcal{Y}$ are the channel outputs and E is a subset of paris $(x, y) \in \mathcal{X} \times \mathcal{Y}$ that represents the channel links.

Example (Z-channel)



We note that $(x, y) \in E$ if and only if y can be received at the channel output when x is transmitted over the channel.

Zero-Error Codes under List Decoding

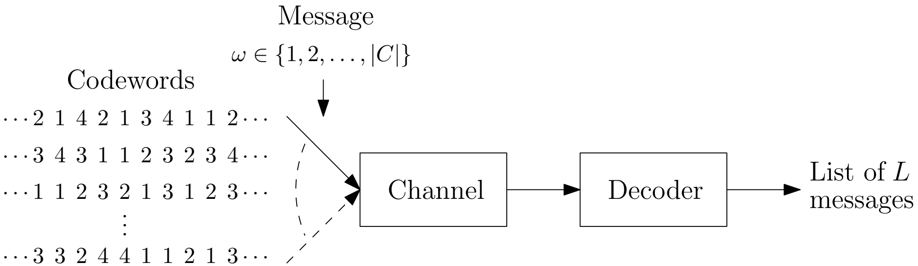


1. The decoder outputs a list of L messages
2. There is an **error** if the original message is not in the list
3. **Zero-error** code: the correct message is always in the list $\iff$ No $L+1$ codewords are compatible with any output sequence

Definition (Zero-error capacity)

The largest asymptotic rate that zero-error codes with list L can achieve for a specific channel is known as the zero-error capacity with list of size L.

Zero-Error Codes under List Decoding



1. The decoder outputs a list of L messages
2. There is an **error** if the original message is not in the list
3. **Zero-error** code: the correct message is always in the list $\iff$ No $L + 1$ codewords are compatible with any output sequence

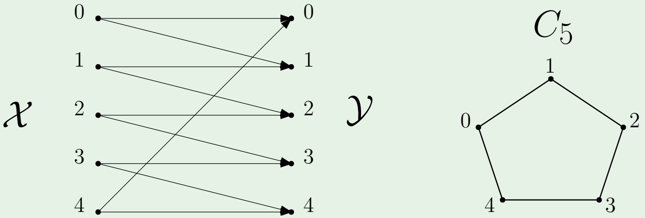
Definition (Zero-error capacity)

The largest asymptotic rate that zero-error codes with list L can achieve for a specific channel is known as the zero-error capacity with list of size L.

Zero-Error Capacity for $L = 1$

Shannon introduced this concept in 1956. Given a discrete channel $W = (\mathcal{X}, \mathcal{Y}, E)$. We can associate to W a confusability graph G.

Example ($5/2$-channel)



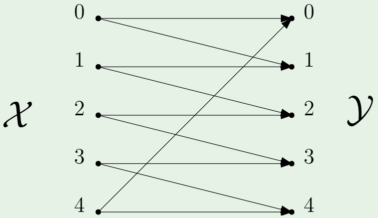
The zero-error capacity $C(G)$ with $L = 1$ only depends on G.

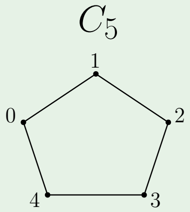
Shannon in 1956 proved that $C(C_5) \geq \log_2 \sqrt{5}$. Then Lovász in 1979 showed that $C(C_5) \leq \log_2 \sqrt{5}$. For C_7, the value $C(C_7)$ is still unknown.

Zero-Error Capacity for $L = 1$

Shannon introduced this concept in 1956. Given a discrete channel $W = (\mathcal{X}, \mathcal{Y}, E)$. We can associate to W a confusability graph G.

Example ($5/2$-channel)





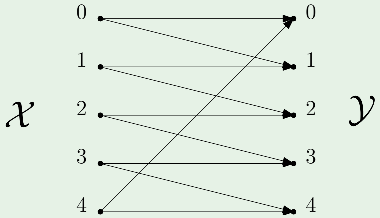
The zero-error capacity $C(G)$ with $L = 1$ only depends on G.

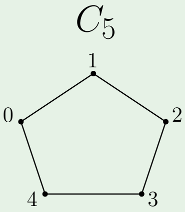
Shannon in 1956 proved that $C(C_5) \geq \log_2 \sqrt{5}$. Then Lovász in 1979 showed that $C(C_5) \leq \log_2 \sqrt{5}$. For C_7, the value $C(C_7)$ is still unknown.

Zero-Error Capacity for $L = 1$

Shannon introduced this concept in 1956. Given a discrete channel $W = (\mathcal{X}, \mathcal{Y}, E)$. We can associate to W a confusability graph G.

Example ($5/2$-channel)





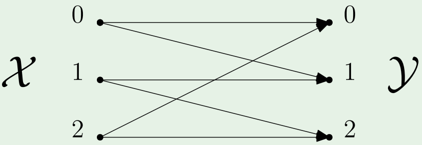
The zero-error capacity $C(G)$ with $L = 1$ only depends on G.

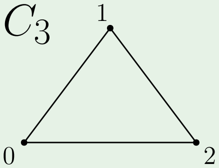
Shannon in 1956 proved that $C(C_5) \geq \log_2 \sqrt{5}$. Then Lovász in 1979 showed that $C(C_5) \leq \log_2 \sqrt{5}$. For C_7, the value $C(C_7)$ is still unknown.

Zero-Error Capacity for $L > 1$

Elias introduced this concept in 1988. Given the following channel W:

Example ($3/2$ channel)





It can be seen that $C(C_3) = 0$. A code that achieves zero-error with list of size 2 for this channel is known as 3-hash code or trifferent code.

$$\frac{1}{4}\log_2\frac{9}{5} \leq C_0(2) \leq \log_2\frac{3}{2}\,,$$

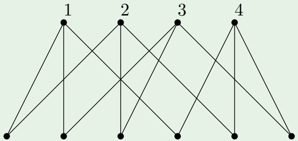
where $C_0(2)$ is the zero-error capacity with list of size 2 of W.

$b/(k-1)$ Channels

Definition

A $b/(k-1)$ channel is a channel where any $k-1$ of the b inputs share one output but no k inputs do.

Example ($4/2$-channel)



A $(4,3)$-hash code achieves zero-error with $L=2$ for the $4/2$-channel.

$$
\begin{array}{ll}
x & \cdots 0\ 2\ 0\ 2\ 3\ 1 \cdots \\
y & \cdots 2\ 3\ 1\ 0\ 1\ 1 \cdots \\
z & \cdots 2\ 3\ 3\ 2\ 1\ 2 \cdots
\end{array}
$$

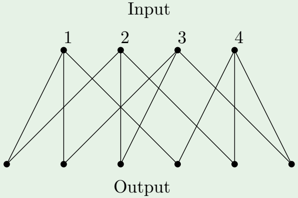
Zero-error capacity for $L < k-1$ is 0 while for $L = k-1$ is positive.

$b/(k-1)$ Channels

Definition

A $b/(k-1)$ channel is a channel where any $k-1$ of the b inputs share one output but no k inputs do.

Example ($4/2$-channel)



A $(4,3)$-hash code achieves zero-error with $L = 2$ for the $4/2$-channel.

$$
\begin{array}{ll}
x & \cdots 0\ 2\ \color{red}{0}\ 2\ 3\ 1 \cdots \\
y & \cdots 2\ 3\ \color{red}{1}\ 0\ 1\ 1 \cdots \\
z & \cdots 2\ 3\ \color{red}{3}\ 2\ 1\ 2 \cdots
\end{array}
$$

Zero-error capacity for $L < k-1$ is 0 while for $L = k-1$ is positive.

Known upper bounds from Literature

The quantity $R_{(b,k)}$ represents the zero-error capacity with list of size $k-1$ of the $b/(k-1)$ channel.

Using a graph theoretical lemma (Hansel's lemma) and a probabilistic argument.

Theorem (Fredman-Komlós (1985))

$$R_{(b,k)} \leq \frac{b^{\underline{k-1}}}{b^{k-1}} \log_2(b-k+2)$$

Generalizing the procedure of F-K (Hansel's for hypergraphs)

Theorem (Körner-Marton (1988))

$$R_{(b,k)} \leq \min_{0 \leq j \leq k-2} \frac{b^{\underline{j+1}}}{b^{j+1}} \log_2 \frac{b-j}{k-j-1}$$

Known upper bounds from Literature

The quantity $R_{(b,k)}$ represents the zero-error capacity with list of size $k-1$ of the $b/(k-1)$ channel.

Using a coding theoretic argument

Theorem (Arikan (1994))

$$R_{(4,4)} \leq 0.3512$$

Mixing the ideas of Arikan and F-K

Theorem (Dalai, Guruswami, Radhakrishnan (2017))

$$R_{(4,4)} \leq 6/19 \approx 0.3158$$

Known upper bounds from Literature

The quantity $R_{(b,k)}$ represents the zero-error capacity with list of size $k-1$ of the $b/(k-1)$ channel.

Theorem (Guruswami, Riazanov (2018))

The Fredman-Komlós bound is not tight for every b and k.

Theorem (Costa, Dalai (2020))

$$R_{(5,5)} \leq 0.1697, \qquad R_{(6,6)} \leq 0.0875$$

Our method

Following the work of Costa and Dalai (2020). We obtained the following upper bound on $R_{(b,k)}$

$$R_{(b,k)} \leq (1+o(1))\frac{1}{2}\log_2(b-k+2)\sum_{i}\sum_{\omega,\mu\in\Omega}\lambda_\omega\lambda_\mu\Psi(f_{i|\omega}, f_{i|\mu}),$$

where Ω is a family of subcodes, $\sum_{\omega\in\Omega}\lambda_\omega = 1$ and $\lambda_\omega \geq 0\ \forall\omega\in\Omega$.

Definition (Ψ function)

Given two probability vectors $p = (p_1, p_2, \dots, p_b)$ and $q = (q_1, q_2, \dots, q_b)$

$$\Psi(p;q) = \frac{1}{(b-k+1)!}\sum_{\sigma\in S_b} p_{\sigma(1)}p_{\sigma(2)}\cdots p_{\sigma(k-2)}q_{\sigma(k-1)} + q_{\sigma(1)}q_{\sigma(2)}\cdots q_{\sigma(k-2)}p_{\sigma(k-1)}$$

New upper bounds for different (b,k)-cases

Analyzing carefully the quadratic form we obtain the following bounds on $R_{(b,k)}$

Theorem (Della Fiore, Costa, Dalai (2022))

(b,k)	*Ours*	*(1)*	*(2)*	*(3)*	*(4)*
$(5,5)$	**0.16894**	*0.16964*	*0.25050*	*0.23560*	*0.19079*
$(6,5)$	**0.34512**	*0.34597*	*0.45728*	*0.44149*	*0.43207*
$(6,6)$	**0.08475**	*0.08760*	*0.21170*	*0.15484*	*0.09228*
$(7,7)$	**0.04090**	*0.04379*	*0.18417*	*0.09747*	*0.04279*
$(8,8)$	**0.01889**	*0.02077*	*0.16323*	*0.05769*	*0.01922*
$(9,8)$	**0.05616**	*0.05686*	*0.30348*	*0.12874*	*0.06001*
$(10,9)$	**0.02773**	*0.02889*	*0.27417*	*0.07668*	*0.02874*
$(11,10)$	**0.01321**	*0.01407*	*0.25018*	*0.04289*	*0.01342*

Table: Upper bounds on $R_{(b,k)}$. All numbers are rounded upwards.

(1,2) $\rightarrow$ *S. Costa, M. Dalai, 2020; M. Dalai, V. Guruswami, and J. Radhakrishnan, 2017;*

(3,4) $\rightarrow$ *E. Arikan, 1994; V. Guruswami, A. Riazanov, 2019.*

All the bounds have been computed symbolically with Mathematica, $R_{(6,6)} \leq 5/59$.

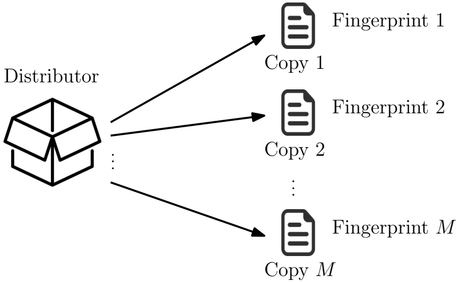
S. Della Fiore, S. Costa and M. Dalai, Improved Bounds for (b,k)-hashing, IEEE Transactions on Information Theory 68 (2022)

Outline

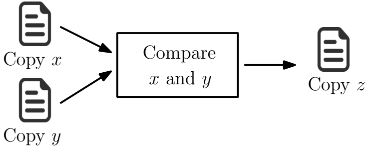
1. Asymptotic growth of codes
   - (b, k)-hash codes
   - Codes for multimedia fingerprinting
2. Related combinatorial problems
   - Erdős Sum-Distinct problem
   - Sequenceability of abelian groups

Multimedia fingerprinting

A distributor wants to sell M copies of a digital product. Each copy has its own fingerprint.



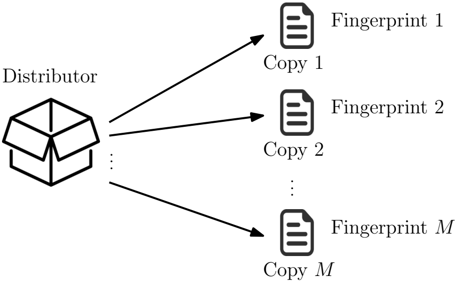
A coalition of malicious users (x and y) can compare their copies



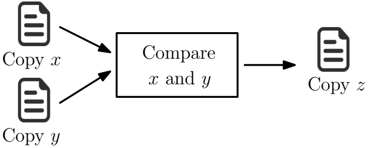
to produce a new feasible copy z (x, y, z are all distinct).

Multimedia fingerprinting

A distributor wants to sell M copies of a digital product. Each copy has its own fingerprint.



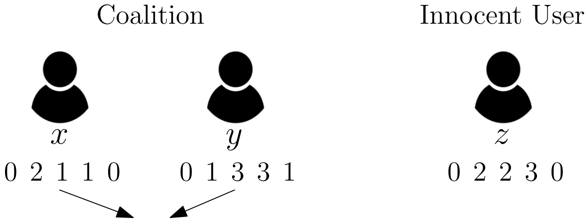
A coalition of malicious users (x and y) can compare their copies



to produce a new feasible copy z (x, y, z are all distinct).

Frameproof codes were introduced due to their applications of protecting innocent authorized users against collusion attacks in digital fingerprinting.

Suppose that C is a 4-ary code of length 5 and $x, y, z \in C$ are distinct codewords.

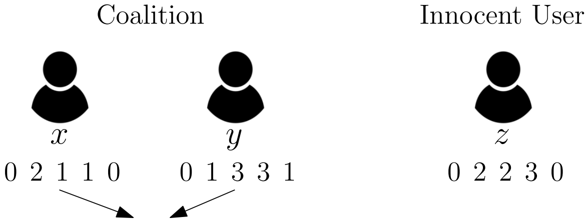


$$A = \{0\} \times \{2,1\} \times \{1,3\} \times \{1,3\} \times \{0,1\}$$

If $z \notin A$ then C is a 4-ary 2-frameproof code. This property has to hold for any distinct $x, y, z \in C$.

Frameproof codes were introduced due to their applications of protecting innocent authorized users against collusion attacks in digital fingerprinting.

Suppose that C is a 4-ary code of length 5 and $x, y, z \in C$ are distinct codewords.



$$A = \{0\} \times \{2,1\} \times \{1,3\} \times \{1,3\} \times \{0,1\}$$

If $z \notin A$ then C is a 4-ary 2-frameproof code. This property has to hold for any distinct $x, y, z \in C$.

Frameproof and Separable codes

Frameproof codes were introduced due to their applications of protecting innocent authorized users against collusion attacks in digital fingerprinting.

Suppose that C is a 4-ary code of length 5 and $x, y, z \in C$ are distinct codewords.

Coalition   Innocent User

x: $0\ 2\ 1\ 1\ 0$ y: $0\ 1\ 3\ 3\ 1$ z: $0\ 2\ 2\ 3\ 0$

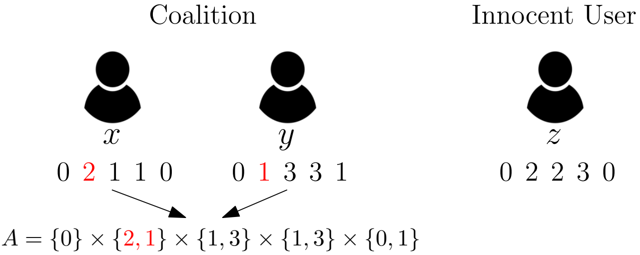
$$A = \{0\} \times \{2,1\} \times \{1,3\} \times \{1,3\} \times \{0,1\}$$

If $z \notin A$ then C is a 4-ary 2-frameproof code. This property has to hold for any distinct $x, y, z \in C$.

Frameproof and Separable codes

Frameproof codes were introduced due to their applications of protecting innocent authorized users against collusion attacks in digital fingerprinting.

Suppose that C is a 4-ary code of length 5 and $x, y, z \in C$ are distinct codewords.



If $z \notin A$ then C is a 4-ary 2-frameproof code. This property has to hold for any distinct $x, y, z \in C$.

New randomized algorithms for frameproof codes

Question

We ask for efficient algorithms to construct frameproof codes of fixed size M with length n as small as possible.

Theorem (Dalai, Della Fiore, Rescigno, Vaccaro (2023))

There exists a randomized algorithm to construct frameproof codes of a fixed size M and length n of complexity $O(nM^2)$ where $n = O(\log M)$.

It can be shown that the length n in the Theorem in near the theoretical optimal length of frameproof codes.

M. Dalai, S. Della Fiore, A. A. Rescigno and U. Vaccaro, Bounds and Algorithms for Frameproof Codes and Related Combinatorial Structures, IEEE ITW (2023)

New randomized algorithms for frameproof codes

Question

We ask for efficient algorithms to construct frameproof codes of fixed size M with length n as small as possible.

Theorem (Dalai, Della Fiore, Rescigno, Vaccaro (2023))

There exists a randomized algorithm to construct frameproof codes of a fixed size M and length n of complexity $O(nM^2)$ where $n = O(\log M)$.

It can be shown that the length n in the Theorem in near the theoretical optimal length of frameproof codes.

M. Dalai, S. Della Fiore, A. A. Rescigno and U. Vaccaro, Bounds and Algorithms for Frameproof Codes and Related Combinatorial Structures, IEEE ITW (2023)

Binary $\overline{2}$-separable codes / B_2 codes

Definition (binary $\overline{2}$-separable code / B_2 code)

We say that a binary code is $\overline{2}$-separable if all sums $x_i + x_j$ over $\mathbb{Z}$, where x_i and x_j are two codewords, are different.

Example (A binary $\overline{2}$-separable code)

$$
\begin{array}{cc}
 & \text{Codewords} \\
x_1 & \cdots 0\,1\,0\,0\,1\,1\,1\,0\,0\,1 \cdots \\
x_2 & \cdots 1\,0\,0\,1\,1\,1\,1\,0\,0\,1 \cdots \\
x_3 & \cdots 0\,0\,0\,0\,0\,0\,1\,0\,1\,0 \cdots \\
 & \vdots
\end{array}
$$

Binary $\overline{2}$-separable codes / B_2 codes

Definition (binary $\overline{2}$-separable code / B_2 code)

We say that a binary code is $\overline{2}$-separable if all sums $x_i + x_j$ over $\mathbb{Z}$, where x_i and x_j are two codewords, are different.

Example (A binary $\overline{2}$-separable code)

$$\begin{array}{ll} \text{Codewords} & \text{Sum } \; x_1 + x_2 \\ x_1 \cdots 0\,1\,0\,0\,1\,1\,1\,0\,0\,1 \cdots & \cdots 1\,1\,0\,1\,2\,2\,2\,0\,0\,2 \cdots \\ x_2 \cdots 1\,0\,0\,1\,1\,1\,1\,0\,0\,1 \cdots & \\ x_3 \cdots 0\,0\,0\,0\,0\,0\,1\,0\,1\,0 \cdots & \\ \vdots & \end{array}$$

Binary $\overline{2}$-separable codes / B_2 codes

Definition (binary $\overline{2}$-separable code / B_2 code)

We say that a binary code is $\overline{2}$-separable if all sums $x_i + x_j$ over $\mathbb{Z}$, where x_i and x_j are two codewords, are different.

Example (A binary $\overline{2}$-separable code)

Codewords	Sum $x_1 + x_2$
$x_1 \cdots 0\,1\,0\,0\,1\,1\,1\,0\,0\,1 \cdots$	$\cdots 1\,1\,0\,1\,2\,2\,2\,0\,0\,2 \cdots$
$x_2 \cdots 1\,0\,0\,1\,1\,1\,1\,0\,0\,1 \cdots$	Sum $x_1 + x_3$
$x_3 \cdots 0\,0\,0\,0\,0\,0\,1\,0\,1\,0 \cdots$	$\cdots 0\,1\,0\,0\,1\,1\,2\,0\,1\,1 \cdots$
$\vdots$	

Binary $\overline{2}$-separable codes / B_2 codes

Definition (binary $\overline{2}$-separable code / B_2 code)

We say that a binary code is $\overline{2}$-separable if all sums $x_i + x_j$ over $\mathbb{Z}$, where x_i and x_j are two codewords, are different.

Example (A binary $\overline{2}$-separable code)

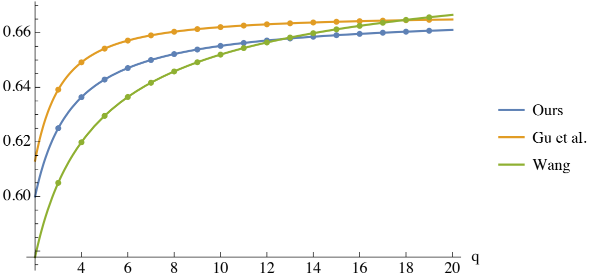
Codewords	
	Sum $x_1 + x_2$
	$\cdots 1\,1\,0\,1\,2\,2\,2\,0\,0\,2 \cdots$
$x_1 \cdots 0\,1\,0\,0\,1\,1\,1\,0\,0\,1 \cdots$	Sum $x_1 + x_3$
$x_2 \cdots 1\,0\,0\,1\,1\,1\,1\,0\,0\,1 \cdots$	$\cdots 0\,1\,0\,0\,1\,1\,2\,0\,1\,1 \cdots$
$x_3 \cdots 0\,0\,0\,0\,0\,0\,1\,0\,1\,0 \cdots$	Sum $x_2 + x_3$
$\vdots$	$\cdots 1\,0\,0\,1\,1\,1\,2\,0\,1\,1 \cdots$

New upper bounds for q-ary $\overline{2}$-separable codes

Theorem (Della Fiore, Dalai (2022))

Let C be a q-ary $\overline{2}$-separable code of length n for $q \geq 2$. Then

$$|C| \leq q^{\frac{2q-1}{3q-1}n(1+o(1))}$$



Improves the best known bounds for every $q \geq 13$.

S. Della Fiore and M. Dalai, A note on $\overline{2}$-separable codes and B_2 codes, Discrete Mathematics 345 (2022)

Outline

1. Asymptotic growth of codes
   - (b, k)-hash codes
   - Codes for multimedia fingerprinting
2. Related combinatorial problems
   - Erdős Sum-Distinct problem
   - Sequenceability of abelian groups

Erdős Sum Distinct Problem

Let $\{a_1, ..., a_n\}$ be a set of positive integers with $a_1 < \ldots < a_n$ such that all 2^n subset sums are distinct.

Conjecture

A famous conjecture by Erdős states that $a_n > c \cdot 2^n$ for some constant c.

The best results known to date are of the form $a_n > c \cdot 2^n/\sqrt{n}$ for some constant c.

Improving the factor $\sqrt{n}$ is a very hard task and so only the constant c has been improved in the past 65 years.

Variations on the original problem

First variation.
The distinct-sums condition is weakened by only requiring that the sums of up to λn elements of the set be distinct with $0 < \lambda < 1$.

Second variation.
The elements $a_i \in \mathbb{Z}^k$ for some $k \geq 1$.

Question

If $a_i \in [0, M]^k \; \forall i$. What is the minimum M for the existence of a sequence $(a_1, \ldots, a_n)$ where all the sums of up to λn elements are distinct?

Our results.
We proved upper and lower bounds on M using probabilistic and polynomial arguments.

S. Costa, M. Dalai and S. Della Fiore, Variations on the Erdős distinct-sums problem, Discrete Applied Mathematics 325 (2023)

Variations on the original problem

First variation.
The distinct-sums condition is weakened by only requiring that the sums of up to λn elements of the set be distinct with $0 < \lambda < 1$.

Second variation.
The elements $a_i \in \mathbb{Z}^k$ for some $k \geq 1$.

Question

If $a_i \in [0, M]^k$ $\forall i$. What is the minimum M for the existence of a sequence $(a_1, \ldots, a_n)$ where all the sums of up to λn elements are distinct?

Our results.
We proved upper and lower bounds on M using probabilistic and polynomial arguments.

S. Costa, M. Dalai and S. Della Fiore, Variations on the Erdős distinct-sums problem, Discrete Applied Mathematics 325 (2023)

Variations on the original problem

First variation.
The distinct-sums condition is weakened by only requiring that the sums of up to λn elements of the set be distinct with $0 < \lambda < 1$.

Second variation.
The elements $a_i \in \mathbb{Z}^k$ for some $k \geq 1$.

Question

If $a_i \in [0, M]^k \ \forall i$. What is the minimum M for the existence of a sequence $(a_1, \dots, a_n)$ where all the sums of up to λn elements are distinct?

Our results.
We proved upper and lower bounds on M using probabilistic and polynomial arguments.

S. Costa, M. Dalai and S. Della Fiore, Variations on the Erdős distinct-sums problem, Discrete Applied Mathematics 325 (2023)

Outline

1. Asymptotic growth of codes
   - (b, k)-hash codes
   - Codes for multimedia fingerprinting
2. Related combinatorial problems
   - Erdős Sum-Distinct problem
   - Sequenceability of abelian groups

Sequenceability of abelian groups

Let S be a subset of a finite abelian group G.

Definition

We say that S is sequenceable if there exists an ordering of its elements such that the partial sums are distinct and not-null (exc. $\sum S = 0$).

Example

Let $S = \{1, 4, 2\} \subset \mathbb{Z}_5$ and $\sigma = (1, 2, 4)$ be an ordering of S. Then the partials sums $(1, 1+2, 1+2+4) = (1, 3, 2)$ are all distinct and not-null.

Conjecture

Every subset $S \subseteq G \setminus \{0\}$ is sequenceable.

Sequenceability of abelian groups

Let S be a subset of a finite abelian group G.

Definition

We say that S is sequenceable if there exists an ordering of its elements such that the partial sums are distinct and not-null (exc. $\sum S = 0$).

Example

Let $S = \{1, 4, 2\} \subset \mathbb{Z}_5$ and $\sigma = (1, 2, 4)$ be an ordering of S. Then the partials sums $(1, 1+2, 1+2+4) = (1, 3, 2)$ are all distinct and not-null.

Conjecture

Every subset $S \subseteq G \setminus \{0\}$ is sequenceable.

Sequenceability of abelian groups

Let S be a subset of a finite abelian group G.

Definition

We say that S is sequenceable if there exists an ordering of its elements such that the partial sums are distinct and not-null (exc. $\sum S = 0$).

Example

Let $S = \{1, 4, 2\} \subset \mathbb{Z}_5$ and $\sigma = (1, 2, 4)$ be an ordering of S. Then the partials sums $(1, 1+2, 1+2+4) = (1, 3, 2)$ are all distinct and not-null.

Conjecture

Every subset $S \subseteq G \setminus \{0\}$ *is sequenceable.*

Sequenceability of abelian groups - Our results

Let S be a subset of a finite abelian group G. Then using the polynomial method we proved the following theorems.

Theorem (Costa, Della Fiore, Ollis and R-Frydman (2022))

For $G = \mathbb{Z}_p$ with p an odd prime and $|S| = 11, 12$, S is sequenceable.

Theorem (Costa, Della Fiore, Ollis and R-Frydman (2022))

Let $p > 3$ be a prime and let $G = \mathbb{Z}_p \times \mathbb{Z}_t \cong \mathbb{Z}_{pt}$, $S \subseteq G \setminus \{(0, 0)\}$, $|S| = 11, 12$ and $t = 2, 3, 4$. Then S is sequenceable.

S. Costa, S. Della Fiore, M. A. Ollis and S. Z. Rovner-Frydman, On Sequences in Cyclic Groups with Distinct Partial Sums, The E. J. of Combinatorics 3 (2022)

References - Some publications

- S. Della Fiore, S. Costa and M. Dalai, Improved Bounds for (b, k)-hashing, IEEE Transactions on Information Theory 68 (2022)
- S. Della Fiore and M. Dalai, A note on $\overline{2}$-separable codes and B_2 codes, Disc. Math. 345 (2022)
- S. Costa, M. Dalai and S. Della Fiore, Variations on the Erdős distinct-sums problem, Discrete Applied Mathematics 325 (2023)
- S. Della Fiore, A. Gnutti and S. Polak, The maximum cardinality of trifferent codes with lengths 5 and 6, Examples and Counterexamples 2 (2022)
- S. Costa, S. Della Fiore, M. A. Ollis and S. Z. Rovner-Frydman, On Sequences in Cyclic Groups with Distinct Partial Sums, The E. J. of Combinatorics 3 (2022)
- S. Della Fiore, M. Dalai and U. Vaccaro, Achievable Rates and Algorithms for Group Testing with Runlength Constraints, IEEE ITW (2022)
- S. Costa and S. Della Fiore, Weak Sequenceability in Cyclic Groups, J. of Comb. Designs (2022)
- M. Dalai, S. Della Fiore, A. A. Rescigno and U. Vaccaro, Bounds and Algorithms for Frameproof Codes and Related Combinatorial Structures, IEEE ITW (2023)