Comenius University, Bratislava
Faculty of Mathematics, Physics and Informatics

Finite Difference Sets

Bachelor's Thesis

2019
Erik Szalay

Comenius University, Bratislava
Faculty of Mathematics, Physics and Informatics

Finite Difference Sets

Bachelor's Thesis

Study Programme:	Applied Computer Science
Field of Study:	9.2.9. Applied Informatics
Školiace pracovisko:	Department of Applied Informatics
Supervisor:	doc. RNDr. Tatiana Jajcayova, PhD.

Bratislava, 2019
Erik Szalay

Acknowledgements: ...

Abstract

Keywords:

Abstrakt

Kľúčové slová:

Contents

List of Figures

List of Tables

Introduction

Chapter 1

Preliminaries

1.1 Definitions and Terminology

Difference sets are interesting topic combining discrete mathematics, combinatorics and group theory. It has applications in the communication and cryptography. The original study in difference sets focused on symmetrical design of classical difference sets, where each difference is represented exactly the same number of times [7].

Definition 1. *A non-empty subgroup D is* ***a (v, k, λ)-difference set*** *of an abelian group G, if the order of G is v, the size of D is k and each non-identity element of G can be expressed in exactly λ ways as a difference $d_1 - d_2$, where $d_1, d_2 \in D$.*

Another way how to describe the difference sets is to consider a multiset of differences $\Lambda = \{d1 - d2 | d1, d2 \in D, d1 < d2\}$. Then D is a difference set if every nonzero element of G appears the same number of times in Λ.

While properties of classical difference sets are interesting for lot of purposes, there is also an interesting generalization of this concept, where we no longer require each element of D to be present exactly λ-times.

Definition 2. *A group S as a subset of N is a* ***generalized difference set (g.d.s.)*** *of type $\Lambda(S) = (\lambda_i)_{i=1}^{max(D)}$, if for every $i \in N$ the number i can be expressed as a difference $s_1 - s_2$ in exactly λ_i ways, where $s_1, s_2 \in S$*

To be able to better describe properties of generalized difference sets, we can further define mulitset of differences and frequency sequence:

Definition 3. *A $D(S)$ is* ***a multiset of differences*** *of generalized difference set S, where S is a subset of N and $D(S)$ contains differences $s1 - s2$ of all pairs $s1, s2 \in S$, where $s1 > s2$.*

Definition 4. *A* $\Lambda(S) = (\lambda_i)_{i=1}^{max(D)}$ *is* ***a frequency sequence*** *of generalized difference set* S*, where* S *is a subset of* N *and each positive integer* i *appears as a difference* $s1 - s2$ *of elements from* S *exactly* λ_i *times.*

The properties of generalized difference set, multiset of differences and frequency sequence can be best described on an example.

Example Let us take a set $S = \{1, 2, 4, 7\}$. The multiset of differences will contain differences of all the pairs of elements of S, like $2 - 1$, $4 - 1$, $7 - 1$ etc. The whole multiset of differences will be $D(S) = \{1, 2, 3, 3, 5, 6\}$, which can be also described by the frequency sequence $\Lambda(S) = \{1, 1, 2, 0, 1, 1\}$, where the first element represents the frequency of number 1 in $D(S)$, the second element of number 2 etc. The example is represented in the Figure 1.2.

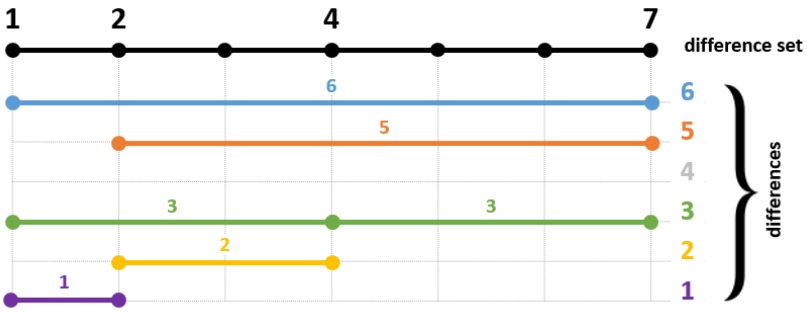

Figure 1.1: Example of generalized difference set with construction of multiset of differencies.

1.2 Construction of generalized difference sets

It is a trivial task to construct a frequency sequence form a generalized difference sets, since it just requires to take each pair of elements form g.d.s. and calculate the difference. But the problem arise when we want to construct a generalized difference set from a frequency sequence.

First of all, not all frequencies represent a frequency sequence of a difference set. For example, for the frequency sequence $\{1, 10\}$, no generalized difference set can be constructed.

Furthermore, two generalized difference sets can have the same frequency sequence. For example, g.d.s. $S = \{1, 2, 3\}$ and $S' = \{2, 3, 4\}$ have the same frequency sequence $\{2, 1\}$. By generalization of this example, if P is a g.d.s. of a group G and $g \in G$ then $P + g = p + g : p \in P$ is also a g.d.s. Therefore, if

there is one g.d.s for a frequency sequence, there are infinite number of them. This kind of transformation is called a **translate** of P [10].

There can be more than one g.d.s. for a frequency sequence which does not represent translate. The g.d.s. $S = \{1, 2, 4\}$ and $S' = \{1, 3, 4\}$ also have the same frequency sequence $\{1, 1, 1\}$, but sequence S does not translate to S, but they were constructed by **reversing of the order** of first level differences in the S.

The aim of this thesis is to recognize those frequencies where a generalized difference set can be constructed by eliminating frequencies, which does not fulfill necessary condition for a frequency sequence.

There are two kinds of generalized difference sets: finite and infinite. Each is defined by corresponding finite or infinite frequency sequence. The properties of finite and infinite g.d.s. differs significantly, especially, when it comes to construction of difference set.

1.2.1 Infinite frequency sequences

Although it seems counter-intuitive, most of infinite frequency sequences allow existence of generalized difference set. Grosek and Jajcay [2] showed, that any infinite frequency sequence where $\lambda_i \geq 2$ allows for g.d.s.

Theorem 1. *(Theorem 3 [2]). Let $\Lambda = \{\lambda_i\}_{i=1}^{\infty}$ be a sequence of positive integers such that $\lambda_i \geq 2$ for all but finitely many $i \in N$. Then there exists a generalized difference set S of type Λ.*

The idea that allows construction of such a g.d.s. relies on the possibility to push pairs of elements for an unfit difference further into positive numbers. E.g. when we come to a difference λ_k in a sequence, that does not fit within existing elements, we can always find very large numbers that differ by this number. The exact construction method described by Grosek and Jajcay [2] follows:

Construction 1. *Let $\Lambda = \{\lambda_i\}_{i=1}^{\infty}$ be a sequence of positive integers, let $\{M_n\}_{i=1}^{\infty}$ be a sequence of subsets of N defined recursively as follows:*

1. *$M_1 = \{m_0, m_0 + 1\}$, where m_0 is an arbitrary element of N;*
2. *the set M_{n+1} is defined from the set M_n by setting*

$$M_{n+1} = M_n \cup \{2(k+1), 2(k+1) + j\}$$

where k is the maximal element of Mn and j is the smallest positive integer which appears in D(Mn) fewer than λ_j times.

Then, let S_Λ *denote the union* $\cup\{M_n | n \in N\}$

Jajcayova and Jajcay [4] further defined conditions for other types of infinite frequency sequences to determine, if they allow for generalized frequency sets. Proofs and construction methods are described in their work [4].

1. Let $\Lambda = \{\lambda_i\}_{i=1}^{\infty}$ be a sequence consisting entirely of 1's and 2's, $\lambda_i \in \{1, 2\}$, for all $i \in N$, such that $\lambda_i = 2$ for infinitely many i's. Then there exists a generalized difference set $S\Lambda$ of type Λ.

2. Let $N = N_1 \cup N_2$ be a partition of the set of natural numbers into two infinite sets with the second set satisfying the property $n + n_p rime \notin N_2$, for all $n, n' \in N_2$. Let $\Lambda = \{\lambda_i\}_{i=1}^{\infty}$ be any sequence of positive integers such that $\lambda_i = 1$ for all $i \in N_1$ and $\lambda_j \geq 3$ for all $j \in N_2$. Then Λ does not allow the existence of a g.d.s.

3. Let $N = N_1 \cup N_2$ be a partition of the set of natural numbers into two non-empty sets with the first set satisfying the property $n + n_p rime \notin N_1$, for all $n, n' \in N_1$ Let $\Lambda = \{\lambda_i\}_{i=1}^{\infty}$ be any sequence of positive integers such that $\lambda_j \geq 3$ for all $i \in N_1$ and $\lambda_j = 1$ for all $j \in N2.4$Then there exists a g.d.s. S of type Λ.

Kopparty [6] described other conditions of infinite frequency sets:

1. If generalized difference set is in the form $S = \{1, \alpha, \alpha^2, \alpha^3, ...\}$, where $\alpha \geq 2$, its frequency sequences consist only of 0's or 1's.

2. If $\Lambda = \{\lambda_i\}_{i=1}^{\infty}$ such that $\lambda_i \leq 1$ and $\lambda_i = 1$ for all but finitely many i, then Λ is a frequency sequence.

3. Any finite sequence of nonnegative numbers is the initial segment of some infinite frequency sequence.

4. The sequence $\Lambda = (x_1, 0, x_2, 0, x_3, 0...)$, where $x_i \geq 1$ for all i, is not the frequency sequence of any set of natural numbers.

5. Let $N = N_1 \cup N_2$be a partition of the set of natural numbers with N_2 satisfying the property that $n + n_0$ is not in N_2 for all n and n_0 in N_2. Let $\Lambda = (\lambda_k)_{k=1}^{\infty}$ be any sequence of positive integers such that $\lambda_i = 0$ for all i in N_1 and $\lambda_i = 1$ for all i in N_2. Then Λ is not a frequency sequence.

1.2.2 Finite frequency sequences

While constructing generalized difference sets form finite frequency sequences, the conditions and process differs from infinite ones. Because of the "limited space" there are more restrictions and the process of reconstruction of the original difference set is more complicated. Kopparty [6] described several necessary conditions for a finite frequency sequence to represent a generalized difference set.

Theorem 2. *(Theorem 6 [6]). For any finite frequency sequence, the sum of the elements of the sequence is of the form* $\binom{n}{2}$ *for some n.*

Proof. Let $S = \{s_1, s_2, s_3, ..., s_n\}$ be a subset of N and $\Lambda = (\lambda_1, \lambda_2, \lambda_3, ..., \lambda_k)$ be its frequency sequence. The total number of differences in the set S is given by $(\lambda_1 + \lambda_2 + \lambda_3 + ... + \lambda_k)$. A difference is between any two numbers, so $\binom{n}{2}$ is the total number of differences in S, and $\binom{n}{2} = (\lambda_1 + \lambda_2 + \lambda_3 + ... + \lambda_k)$. □

Theorem 3. *(Theorem 7 [6]). For any finite sequence* $\Lambda = (\lambda_i)_{i=1}^n$ $\lambda_i \leq n - i + 1, \forall i \neq n$ *and* $\lambda_n = 1$.

Proof. Let $S = \{s_1, s_2, s_3, ..., s_n\}$ be a subset of N and $\Lambda = \{\lambda_1, \lambda_2, \lambda_3, ..., \lambda_k\}$ be its frequency sequence, where λ_i is the number of times i appears as a difference of elements in S. Since k is the largest difference, λ_k must be 1 because there is only one largest difference (which is $s_n - s_1$). Similarly, λ_{k-1} can be atmost 2 second largest differences $s_{n-1} - s_1$ and $s_n - s_2$. Hence $k-1 \leq 2$. Similar arguments give the results for $\lambda_{k-3}, \lambda_{k-4}, ..., \lambda_1$. □

Theorem 4. *(Theorem 8 [6]). A finite sequence* $\{\lambda_i\}_{i=1}^k$*, where* $\lambda_i = 1$ *and* $k \geq 10$ *does not allow the existence of a generalized frequency set.*

Proof. Let us try to construct a g.d.s. S with a frequency sequence $\lambda_i = 1$, where $k \geq 10$. Let us then take frequency $\lambda_k = 1$. Based on that we know that 1 and $k + 1$ must be in S. Then the next two elements of S could be either 2 or $k - 1$. When we take 2 as the next element of S, we will need a difference $k - 2$, so for the next step we will have following possibilities:
(1) $1 + (k - 2) = k - 1 \in S$
(2) $2 + (k - 2) = k \in S$
(3) $k + 1 - (k - 2) = 3 \in S$
The third options is impossible: 3 cannot be an element in S, because that would require 1 as a difference of $(2 - 1)$ and $(3 - 2)$ to be present at least twice in Λ. Also the second option would require k to be present twice. Only the first option is possible. So we have $S \supset \{1, 2, k - 1, k + 1\}$ and $D(S) \supset$

$\{1, 2, k-3, k-2, k-1, k\}$. The next difference we could add could be $k-4$, which gives us following options:
(1) $1 + (k-4) = k-3 \in S$
(2) $2 + (k-4) = k-2 \in S$
(3) $k-1 + (k-4) = 3 \in S$
(4) $k+1 - (k-4) = 5 \in S$
This time the first three options would lead to multiple differences and thus to violation of the definition of our g.d.s., so only last option is plausible. So $S \supset \{1, 2, 5, k-1, k+1\}$ and $D(S) \supset \{1, 2, 3.4, k-6, k-4, k-3, k-2, k-1, k\}$. So we are missing a difference $k-5$, but by trying to add this difference we violate the conditions of our g.d.s., because we would found duplicate differences:
(1) $1 + (k-5) = k-4 \in S$
(2) $2 + (k-5) = k-3 \in S$
(3) $5 + (k-5) = k \in S$
(4) $k-1 - (k-5) = 4 \in S$
(5) $k+1 - (k-5) = 6 \in S$
Thus, we cannot move on to constuct a five-element set which has a frequency sequence consisting entirely of ones. The set $S = \{1, 2, 5, 7\}$ has the frequency sequence $\Lambda = (1, 1, 1, 1, 1, 1)$. No set consisting of more than four elements can have a frequency sequence consisting entirely of ones. □

1.3 Algorithms for construction of finite g.d.s

For construction of finite g.d.s. from frequency sequence, we could try to find the g.d.s. by different approaches. Basic algorithm could take an integer to be the lowest element in g.d.s. and then choose the first non-zero element of frequency sequence and by adding it to the first element of g.d.s. generate the next g.d.s. element. Each new element must satisfy condition of having differences to any previous elements within unused multiset of differences. If a new element does not satisfy this condition, other element is chosen by backtracking. This process would be very time consuming, since the input sizes grow quadratically. Stefanak [9] described two other approaches that could be used in finding finite generalized difference sets. For the introduction of the algorithms, it is useful to define specific set of differences: the base differences are the differences between neighboring elements in ordered generalized difference set, and the slope differences are the differences between the largest element and any other element. The names are derived from pyramid representation of visualization of differences in [9].

Imagine the elements of B as base bricks of the pyramid. We shall put other elements of D on top of them in such a fashion that the element $dt = b_i + ... + b_j$ would be on top of the elements $dm = b_i + ... + b_{j-1}$ and $dn = b_{i+1} + ... + b_j$ where $b_i, b_j \in B = \{b_1, ..., b_n\}; i < j \leq |B|$.

The representation can be best shown on an example in Figure X. Given a generalized difference set $S = \{1, 2, 4, 7, 14\}$, we align the base differences, e.g. differences between neighboring elements of g.d.s., as the base of a pyramid. In our example, the base differences are $B = \{1, 2, 3, 7\}$. Next level of the pyramid would be constructed by differences between the first and the third element, the second and the fourth, etc. When we build the pyramid all the way up, we would get all the differences of g.d.s. and the bricks furthest right at each level of the pyramid define the slope differences, e.g. the differences between the highest element, 14 in our example, and each other element in g.d.s., consisting the set of slope differences $\{7, 10, 12, 13\}$.

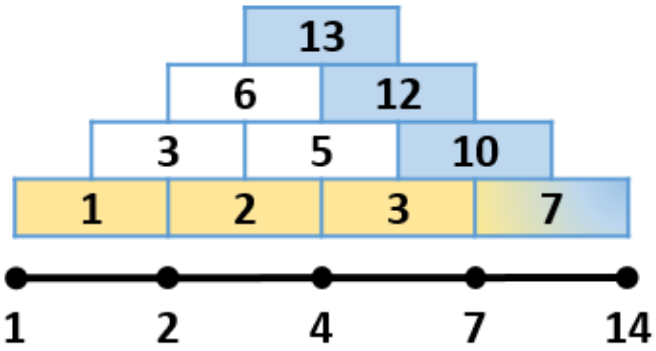

Figure 1.2: Example of pyramid visualization of g.d.s. and its difference set. The set of base differences is in yellow $B = \{1, 2, 3, 7\}$ and the set of slope differences is in blue $\{7, 10, 12, 13\}$.

Base on this representation of g.d.s. and difference set, [9] introduced two algorithms to optimize construction of g.d.s.: the base finding algorithm and the slope finding algorithm.

The base finding algorithm, as the name suggests, is trying to construct the g.d.s. by identifying the set of base differences first. Knowing the base set as an ordered list, we can choose a starting integer and exactly construct the whole difference set, either by adding or subtracting base differences starting with the first element. [9] therefore defines several conditions that a set of base differences must fulfil.

Extrapolating from theorem 6, we can calculate the number of elements in the set of base differences.

Theorem 5. *(Theorem 4 [9]). The size of the set of base differences |B| is defined as |D| = (|B|+1)*|B| / 2, where |D| is the size of its generalized difference set.*

Proof. From Theorem 6, the number of all differences is $\binom{n}{2}$, where n is number of elements in g.d.s. and the number of elements in set of base differences is n-1 as the number of gaps between n elements of g.d.s. Therefore

$$|D| = n = \binom{n}{2} = \frac{n(n-1)}{2} = \frac{|B|+1)|B|}{2}$$

□

Then we can define necessary condition for each element of the set of all differences D derived from the associative property of addition.

Theorem 6. *(Theorem 5 [9]). Every element of the set of all differences D can be described by the set of base differences B as* $D = \{d|d = \sum_{k=i}^{j} b_k\}$

Stefanak [9] further describes three other conditions for reducing the number of candidates for set of base differences:

Theorem 7. *(Theorem 6 [9]). For every element* $d = \sum_{k=i}^{j} b_k$*, there are at least* $j - i$ *different pairs of elements* d'_1, d'_2 *such that* $d'_1 + d'_2 = d$*.*

Theorem 8. *(Theorem 7 [9]).* $\forall b \in B$*, there are at least |B| -1 elements* $p \in D$ *such that* $b + p \in D$*.*

Theorem 9. *(Theorem 8 [9]). For any set of differences D, the first element of B (and last of reverse(B)) will be equal to the difference of the largest and the second largest element of D.*

Proofs for these theorems can be found in [9].
The base finding algorithm consists of the following steps:

1. Create a subset M of D, consisting of those elements of D that cannot be expressed as a sum of other elements of D.

2. Find every subset B^0 of D, that contains every element from M and where $\sum B^0 = max(D)$ and put them at the beginning of a stack (depth-first search).

3. Get the first element of a stack. If any permutation of B^0 creates a difference bag, return this permutation. Else, get the next candidate from the stack until the stack is empty.

The slope finding algorithm in [9] on the other hand, is trying to construct the g.d.s. by identifying the slope. It starts with the peak element and tries to divide it between base element and a slope element. Then follows with further dividing the slope element, until it gets enough elements for the base set. If at any point the algorithm comes to an non-existing or non-plausible element, it backtracks to previous step. Unlike the base finding algorithm, the slope finding algorithm populates an ordered list of base differences

Variables: a stack of partial results, d_{next} as element to be divided to possible slope and base members, B^0 as possible base.

1. Insert a partial result consisting of empty possible base and a peak element as the next divided element d_{next} into a stack.

2. If a stack is not empty, take the first Partial result, else inputed sequence is not a frequency sequence.

3. If size of the possible base + 1 is equal to the size of a base list then jump to 4, else jump to 5.

4. Add the divided element to the possible base and check if it constructs a set of differences equal to a bag of differences. If yes, return it as result, else go to 2.

5. For every two elements $e_1, e_2 \in D$such that $e_1 + e_2 = d_{next}$, add 2 new partial results where $d_{next} = e_1$, $B^0 = B^0 + e_2$ and $d_{next} = e_2$, $B^0 = B^0 + e_1$ to the head of the stack. Then go to 2.

Since every element of the bag of differences can be expressed as $d = \sum_j^{k=i} b_k$ where $i \leq j, b_k \in B$, there exists at least $j - i - 1$ different pairs of elements with a sum equal to d. In this case, algorithm is looking for the elements e_1 and e_2 such that e_1 is from the base and e_2 lies on the side of "pyramid". But since we can't predict which pair that is, backtracking is needed. If we were to divide peak element to the number of elements equal to base size, we will end up with the base finding algorithm (2.3).

When comparing the base and slope finding algorithms, Stefanak [9] found that the slope finding algorithm is far superior to the base finding algorithm for any input. The reason for the superiority is the fact, that slope finding algorithm identifies ordered list of base differences.

Chapter 2

Implementation

Chapter 3

Results

Conclusion

Bibliography

[1] A.Pott, P.V. Kumar, T. Helleseth, and D. Jungnickel. *Difference Sets, Sequences and their Correlation Properties.* Springer-Science+Business Media, B.v., 1999.

[2] Otakar Grosek and Robert Jajcay. Generalized difference sets on an infinite cyclic semigroup. *J. Combin. Math. Combin. Comput*, 13:167–174, 1993.

[3] Tatiana B. Jajcayová. Generalized difference sets. 2015.

[4] Tatiana B. Jajcayová and Robert Jajcay. Notes on subtractive properties of natural numbers. *Bulletin of the ICA*, 25:29–40, 2008.

[5] Ivana Kellyérová. Generalized difference sets. Bachelor's thesis, Comenius University, 2014.

[6] Swara Kopparty. Results on frequency sequences. Thesis, Terre Haute South Vigo High School, 2008.

[7] Emily H. Moore and Harriet S. Pollatsek. *Difference Sets: Connecting Algebra, Combinatorics, and Geometry.* American Mathematical Society, 2013.

[8] Bc. Marián Opial. Generalized difference sets – an algorithmic approach. Master's thesis, Comenius University, 2017.

[9] Marek Štefaňák. Frequency sequences of finite difference sets. Bachelor's thesis, Comenius University, 2014.

[10] J. H. van Lint and R. M. Wilson. *A Course in Combinatorics, 2nd edition.* Cambridge University Press, 2001.