On Longest Geometrically Increasing Sequences

Deyan Hadzhi-Manich

Under the direction of

Yuchong Pan
Department of Mathematics
Massachusetts Institute of Technology

Research Science Institute
August 1, 2023

Abstract

We consider a problem that comes up in the analysis of the *discrete Newton's algorithm for submodular line optimization*. It has to deal with finding *longest geometrically increasing sequences*, that is finding the longest sequence of partial sums of the elements of a real vector $\mathbf{a}$ with n entries, such that each subsequent partial sum is at least twice as large as the previous one. It is known that if we restrict $\mathbf{a}$ to contain nonnegative components only, then such a sequence can contain at most n subsets, and that the maximum length is $\frac{1}{2}n\log_2 n \pm O(n\log\log n)$ in the general case. This result leads us to consider how the number of negative components in $\mathbf{a}$ affects the longest geometrically increasing sequences. We define $G(n,k)$ to be the largest $m \in \mathbb{N}$ such that there exist $\mathbf{a} \in \mathbb{R}^n$ with exactly k negative components and $T_1,\ldots,T_m \subseteq [n]$ for which $\mathbf{a}(T_1),\ldots,\mathbf{a}(T_m)$ form a geometrically increasing sequence. In this paper, we prove several bounds on $G(n,1)$, $G(n,n-1)$ and $G(n,n/2)$, as well as several inequalities between different values of $G(n,k)$. In addition, we report results from a computational study in which the vector $\mathbf{a}$ is drawn from a normal distribution, implying the conjecture that the maximum length in this case grows logarithmically.

Summary

Imagine that we have a list of numbers, and we want to create a sequence of positive numbers in which each number is a sum of some of the numbers from the list, and at least twice as big as the previous one. Although seemingly simple, solving this problem turns out to be crucial in analyzing the running time of an algorithm called the *discrete Newton's algorithm* that is used in many optimization problems. It is known that if all numbers from the initial list are nonnegative, then any sequence we can construct has at most the number of elements from the initial list. However, if we allow negative numbers, things become more interesting. In this paper, we study how the number of negative elements affects the longest sequence we can construct in our problem. We prove several lower and upper bounds on the maximum length such a sequence can have, with some connections between different parameters. We have also conducted computational experiments which shed some light on the nature of the problem.

1 Introduction

Given some n-dimensional vector $\mathbf{a}$, for a sequence of sets $T_1, \ldots, T_m \in [n]$ to be *geometrically increasing*, we need the sequence $\mathbf{a}(T_i)$, where we denote $\mathbf{a}(S) = \sum_{j \in S} a_j$, to be positive and increasing with at least a factor of 2. We consider the following problem of finding the longest geometrically increasing sequence.

Problem 1.1. For a fixed n, across all $\mathbf{a} = (a_1, \ldots a_n) \in \mathbb{R}^n$, what is the largest number m of sets $T_1, \ldots, T_m \subseteq [n]$ with $\mathbf{a}(T_1) > 0$ and $\mathbf{a}(T_{i+1}) \geq 2\mathbf{a}(T_i)$ for all $i \in [m-1]$?

Given n, we denote with $G(n)$ the largest length, i.e., the answer to Problem 1.1.

Example 1.1. Let $\mathbf{a} = (1, 2, 3)$ and $T_1 = \{1\}$, $T_2 = \{2\}$, and $T_3 = \{1, 2, 3\}$. Then

$$\begin{aligned}
\mathbf{a}(T_1) &= a_1 = 1, \\
\mathbf{a}(T_2) &= a_2 = 2, \\
\mathbf{a}(T_3) &= a_1 + a_2 + a_3 = 6.
\end{aligned}$$

The sequence $T_1, \ldots, T_m$ is geometrically increasing since $2 \geq 2 \times 1$ and $6 \geq 2 \times 2$.

Example 1.2. Now let $\mathbf{a} = (4, 5, -3)$ and $T_1 = \{1, 3\}$, $T_2 = \{2, 3\}$, $T_3 = \{1\}$, $T_4 = \{1, 2\}$. Then

$$\begin{aligned}
\mathbf{a}(T_1) &= 4 - 3 = 1, \\
\mathbf{a}(T_2) &= 5 - 3 = 2, \\
\mathbf{a}(T_3) &= 4, \\
\mathbf{a}(T_4) &= 4 + 5 = 9.
\end{aligned}$$

In fact, this is the longest a geometrically increasing sequence can be when $n = 3$, i.e., $G(3) = 4$.

If $\mathbf{a}$ is nonnegative (i.e., each component of $\mathbf{a}$ is nonnegative), then it is commonly known, and there is an elementary proof, that there exist at most n subsets of $[n]$ which form a geometrically increasing sequence, and this upper bound is tight in the sense that given $n \in \mathbb{N}$, one can construct some nonnegative $\mathbf{a} \in \mathbb{R}^n$ and n subsets of $[n]$ that form a geometrically increasing sequence with respect to $\mathbf{a}$.

For general vectors $\mathbf{a}$, there exist "almost tight" lower and upper bounds on $G(n)$. In 1992, using polyhedral theory, Goemans (see [1]) proved that

$$G(n) \leq O(n \log n).$$

However, with a more careful analysis using Hadamard's inequality, their technique indeed gives

$$G(n) \leq \frac{1}{2} n \log_2 n + O(n). \tag{1}$$

This bound was mentioned in Goemans' talk at Tutte's 100th Distinguished Lecture Series at the University of Waterloo [2], but the proof is not found in the literature. In Section

2, we adapt the proof in [1] to show this bound. An almost matching lower bound for n being a power of 2 was given by Goldmann in 1993 (see [3]) using Hadamard matrices and a Fourier-analytic construction from Håstad [4]:

$$G(n) \geq \frac{1}{2} n \log_2 n - O(n \log\log n). \tag{2}$$

As mentioned in [4], this sharp bound is not known to exist when n is not a power of 2, and it is an open problem whether there exists a Hadamard matrix of order $4k$ for every $k \in \mathbb{N}$. However, (2) can be used to give a fairly good lower bound as follows:

$$G(n) \geq G\left(2^k\right) \geq \frac{1}{2} 2^k k + o\left(2^k k\right) \geq \frac{1}{4} n \log_2 n + o(n \log n).$$

The linear bound for nonnegative vectors $\mathbf{a}$ and the bounds (1) and (2) lead us to consider how the number of negative components in $\mathbf{a}$ affects the longest geometrically increasing sequences. This specific question has not been studied in the literature. Concretely, we study the following problem.

Problem 1.2. Given $n \in \mathbb{N}$ and $k \in [n]$, we define $G(n,k)$ to be the largest number $m \in \mathbb{N}$ where there exists $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$ with exactly k negative components and $T_1, \ldots, T_m \subseteq [n]$ such that $\mathbf{a}(T_i) > 0$ and $\mathbf{a}(T_{i+1}) \geq 2\mathbf{a}(T_i)$ for all $i \in [m-1]$. What does the function $G(n,k)$ look like? What values of k achieve the maximum in Problem 1.1?

This problem originated from the analysis of an algorithm called the discrete Newton's algorithm, which is used for solving certain optimization problems. More formally, it solves the *sumbodular line search problem*: Given a submodular function $f : 2^V \to \mathbb{R}$ on some finite ground set V with its associated polymatroid $P(f)$, a vector $\mathbf{x}_0 \in P(f)$ and a vector $\mathbf{a} \in \mathbb{R}^V$, find the largest δ such that $\mathbf{x}_0 + \delta\mathbf{a} \in P(f)$.[1] Problem 1.1 is a crucial subproblem in analyzing the running time of the discrete Newton's algorithm.

In Section 2, we provide a more careful analysis of Goemans' upper bound on $G(n)$. Then, in Section 3, we present results for some exact values of $G(n,k)$ for small values of n, obtained by a brute-force algorithm. We show a dependency between $G(n,1)$ and $G(n,n-1)$, as well as some bounds regarding them in Section 4. In Section 5, we provide a general inequality for different values of $G(n,k)$. Further, we show a good lower bound for $G(n, \frac{n}{2})$ in Section 6, derived using Goldmann's lower bound construction. In Section 7 we consider the behavior of the longest geometrically increasing sequences when $\mathbf{a}$ is chosen from a normal distribution, which might help us gain some insight into the discrete Newton's algorithm.

2 A more careful analysis of Goemans' upper bound on $G(n)$

For completeness, we present a more careful analysis of Goemans' upper bound on $G(n)$. Radzik [1] cited Goemans' upper bound $G(n) \leq O(n \log n)$ as personal communication, and

[1]We say that a set function $f : 2^V \to \mathbb{R}$ on some ground set V is *submodular* if for all $A, B \subseteq V$, we have $f(A) + f(B) \geq f(A \cap B) + f(A \cup B)$. The *extended polymatroid* associated to a submodular set function $f : 2^V \to \mathbb{R}$ is defined to be the polyhedron $P(f) = \left\{\mathbf{x} \in \mathbb{R}^V : x(S) \leq f(S)\ \forall S \subseteq V\right\}$.

included its proof. However, in Goemans' talk at Tutte's 100th Distinguished Lecture Series at the University of Waterloo, this bound was mentioned as $G(n) \leq \frac{1}{2}n\log_2 n + o(n\log n)$ but no proof of this more careful analysis of Goemans' bound was found in the literature.

In this section, we give this more careful analysis using Goemans' polyhedral technique combined with a direct application of Hadamard's inequality. We start with the definition of a polyhedron.

Definition 2.1. A *polyhedron* is a subset of $\mathbb{R}^n$ that can be represented by a finite system of linear inequalities. More formally, a polyhedron in $\mathbb{R}^n$ is defined as

$$P = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} \geq \mathbf{b}\},$$

where A is an $m \times n$ real-valued matrix, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$, and $\geq$ denotes element-wise inequality.

A useful well-known property of polyhedra that we need for our proof is the following.

Lemma 2.1 (Rephrasing of Theorem 5.7 in [5])**.** *Consider a polyhedron*

$$P = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} \geq \mathbf{b}\}.$$

Extreme points of P are vectors in P which satisfy

$$A'\mathbf{x} = \mathbf{b}'$$

for some nonsingular $n \times n$ submatrix A' of A and an n-dimensional subvector $\mathbf{b}'$ of $\mathbf{b}$. If P does not contain a line, then it has at least one extreme point.

By embedding subsets in the Euclidean space as characteristic vectors and by scaling, we can represent subsets as a matrix, formulating Problem 1.1 in the following equivalent version.

Problem 2.2. For a given $n \in \mathbb{N}$, what is the largest $m \in \mathbb{N}$ for which we can choose a vector $\mathbf{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$ and a matrix $A \in \{0,1\}^{m\times n}$ with $\mathbf{r}_i$ being the vector of the ith row of A, such that

- $\mathbf{r}_1^\top \mathbf{a} = 1$,
- $\mathbf{r}_i^\top \mathbf{a} \geq 2\mathbf{r}_{i-1}^\top \mathbf{a}$ for all $i \in \{2, \ldots, m\}$?

In addition, we need the following version of Hadamard's inequality.

Lemma 2.3 (Hadamard's inequality, see [6])**.** *If an $n \times n$ matrix M with entries (m_{ij}) is such that $|m_{ij}| \leq B$ for all i and j, then*

$$|\det M| \leq B^n n^{n/2}.$$

Now we are ready to give a more careful analysis of the upper bound. As in [1], we give an upper bound in a slightly more general problem in which we can add or subtract any component of $\mathbf{a}$ in a subset.

Theorem 2.4. *Let $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ be an n-dimensional vector with nonnegative components, and let $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m$ be vectors from $\{-1, 0, 1\}^n$. If for all $i = 1, 2, \ldots, m-1$,*

$$0 < 2\mathbf{y}_{i+1}^\top \mathbf{a} \le \mathbf{y}_i^\top \mathbf{a},$$

then $m \le \frac{1}{2} n \log_2 n + O(n)$.

Proof. Consider the following polyhedron:

$$\begin{aligned} P = \{\mathbf{c} = (c_1, c_2, \ldots, c_n) \in \mathbb{R}^n :& \\ (\mathbf{y}_i - 2\mathbf{y}_{i-1})^\top \mathbf{c} \ge 0 &\quad \text{for all } i \in \{2, 3 \ldots, m\}, \\ \mathbf{y}_1^\top \mathbf{c} = 1,& \\ c_i \ge 0 &\quad \text{for all } i \in [m]\} \end{aligned} \tag{3}$$

Let A and $\mathbf{b}$ denote the coefficient matrix and the right-hand side vector of the system defining polyhedron P, respectively. Note that P is nonempty because it contains the vector $\mathbf{a}/(\mathbf{y}_1^\top \mathbf{a})$.

From polyhedral theory (Lemma 2.1), we know that there exists a vector $\mathbf{c}^* = (c_1^*, \ldots, c_n^*)$ such that there is an $n \times n$ nonsingular submatrix A' of A and a subvector $\mathbf{b}'$ of $\mathbf{b}$ such that $A'\mathbf{c}^* = \mathbf{b}'$. Then, from Cramer's rule,

$$c_i^* = \frac{\det A_i'}{\det A'},$$

where matrix A_i' is obtained from A' replacing the ith column with $\mathbf{b}'$. Due to the constraints of our task $|a_{ij}| \le 3$ for every entry a_{ij} in the matrix A. Therefore, by Lemma 2.3,

$$|\det A| \le 3^n n^{n/2}$$

(in [1] a weaker bound on the determinant is used). Hence.

$$c_i^* = \frac{\det A_i'}{\det A'} \le 3^n n^{n/2}.$$

Therefore, we have for all $j \in [m]$

$$\mathbf{y_j}^\top \mathbf{c}^* \le n 3^n n^{n/2},$$

and

$$1 = \mathbf{y}_1^\top \mathbf{c}^* \le \frac{1}{2^{m-1}} \mathbf{y}_m^\top \mathbf{c}^* \le \frac{1}{2^{m-1}} n 3^n n^{n/2}.$$

Hence,

$$m \le 1 + \log_2(n 3^n n^{n/2}) = \frac{1}{2} n \log_2 n + (\log_2 3) n + \log_2 n + 1 = \frac{1}{2} n \log_2 n + O(n).$$

$\square$

Theorem 2.4 implies the following more constrained version for Problem 2.2.

Corollary 2.4.1. *Let* $\mathbf{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$ *and let* $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m$ *be vectors from* $\{0,1\}^n$*. If for all* $i = 1, 2, \ldots, m-1$,

$$0 < 2\mathbf{y}_{i+1}^\top \mathbf{a} \leq \mathbf{y}_i^\top \mathbf{a}$$

then $m \leq \frac{1}{2} n \log_2 n + O(n)$.

In Problem 2.2, using the fact that we cannot have both -1 and 1 as coordinates in the same component of two vectors, we can improve the factor before the term n, making the bound $G(n) \leq \frac{1}{2} n \log_2 n + n + o(n)$.

Corollary 2.4.2. *For all* n*, it holds that* $G(n) \leq \frac{1}{2} n \log_2 n + n + o(n)$*.*

When the vector $\mathbf{a}$ is nonnegative, it is commonly known, that the maximal length of a geometrically increasing sequence is n, or $G(n, 0) = n$. We include the proof of this for completeness.

Theorem 2.5 (folklore)**.** *For all* n*, it holds that* $G(n, 0) = n$*.*

Proof. For any $i > 1$,

$$\mathbf{a}(T_i) \geq 2\mathbf{a}(T_{i-1}) \geq \mathbf{a}(T_{i-1}) + 2\mathbf{a}(T_{i-2}) \geq \cdots \geq \sum_{j=1}^{i-1} \mathbf{a}(T_j) + \mathbf{a}(T_1) > \sum_{j=1}^{i-1} \mathbf{a}(T_j).$$

This means that T_i contains an element that is not contained in $T_1, T_2, \ldots, T_{i-1}$. Therefore, there can be no more than n sets. An easy construction that makes the bound tight is $a_i = 2^i$ and $T_i = \{i\}$ for all $i \in [n]$.

□

3 Exact values of $G(n, k)$ for small values of n

To gain some insight into the problem, we used a brute-force algorithm to compute some values for $G(n, k)$ for small values of n. As $\mathbf{a}$ is not bounded, we are not able to iterate over all possible values of $\mathbf{a}$. In our brute-force algorithm, we fix n and m and iterate over all possible sequences of subsets of $[n]$. Then we construct the polyhedron as in (3) and check if it is nonempty in polynomial time. The results of the execution are presented in Table 1.

n \ k	0	1	2	3	4	5
1	1	0				
2	2	2	0			
3	3	4	3	0		
4	4	5	5	5	0	
5	5	≥ 7	≥ 7	≥ 7	≥ 6	0

Table 1: Values of $G(n, k)$ achieved by the brute-force algorithm

From Table 1, we observe that $G(n, 0) = n$ for small values of n, and that in each column the values increase as n increases. Furthermore, we observe that, in each row, the values increase up to some point and decrease as k increases.

4 The cases of $G(n, 1)$ and $G(n, n-1)$

We observe an interesting connection between $G(n, 1)$ and $G(n, n-1)$, stated in the following theorem.

Theorem 4.1. *For all n, it holds that $G(n, n-1) \leq \max(G(n, 0), G(n, 1))$.*

Proof. Let $n \in \mathbb{N}$. Let $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$ and $T_1, \ldots, T_m \subseteq [n]$ such that $\mathbf{a}$ has exactly $n-1$ negative components and $T_1, \ldots, T_m$ form a geometrically increasing sequence with respect to $\mathbf{a}$. Without loss of generality, $a_1 > 0$ and $a_2, \ldots, a_n < 0$. Since all $\mathbf{a}(T_i) > 0$ for all $i \subseteq [n]$, then $1 \in T_1, \ldots, T_m$. Let $S_i = T_i \setminus \{1\}$ and $F_i = \{2, \ldots, n\} \setminus S_i$. Then

$$\mathbf{a}(T_i) = a_1 + \mathbf{a}(S_i).$$

Let

$$p = a_2 + a_3 + \cdots + a_n.$$

Then,

$$\mathbf{a}(S_i) = \mathbf{a}(\{2, \ldots, n\}) - \mathbf{a}(\{2, \ldots, n\} \setminus S_i) = p - \mathbf{a}(F_i).$$

Hence, $\mathbf{a}(T_i) \geq 2\mathbf{a}(T_{i-1})$ is equivalent to

$$a_1 + p - \mathbf{a}(F_i) \geq 2(a_1 + p - \mathbf{a}(F_{i-1})).$$

Let $x = a_1 + p$. We consider two cases.

Case 4.1.1 Suppose $x \geq 0$. Then for $i \in \{2, \ldots, m\}$,

$$-\mathbf{a}(F_i) \geq x - 2\mathbf{a}(F_{i-1})$$

or equivalently,

$$-\mathbf{a}(F_i) + 2\mathbf{a}(F_{i-1}) \geq x.$$

Since $x \geq 0$, then $F_1, \ldots, F_m \subseteq \{2, \ldots, n\}$ is a geometrically increasing sequence of sets with respect to $-\mathbf{a}$. Since $-a_2, \ldots, -a_n > 0$, this means that $m \leq G(n-1, 0) = n-1 \leq G(n, 0)$.

Case 4.1.2 Suppose $x < 0$. Then we have for $i \in \{2, \ldots, m\}$

$$x - \mathbf{a}(F_i) \geq 2(x - 2\mathbf{a}(F_{i-1})) \quad \text{for } i \in \{2, \ldots, m\}.$$

This shows that, with respect to the vector $\mathbf{b} = \{x, -a_2, \ldots, -a_n\}$, the sequence of sets $Q_1, Q_2, \ldots, Q_m$, where $Q_i = F_i \cup \{1\}$ for all $i \in [m]$, is geometrically increasing. Therefore, from our original construction with $n-1$ negative components in $\mathbf{a}$, we give a new construction with one negative component in the vector and a geometrically increasing sequence of sets having the same length. This implies $m \leq G(n, 1)$.

This completes the proof. $\square$

Next, we give lower bounds on $G(n,1)$ and $G(n,n-1)$ that are greater than n, showing that even one negative component can give us some flexibility in constructing geometrically increasing sequences.

Theorem 4.2. *For all n, it holds that $G(n,1) \geq n + \lfloor \log_2(n-1) \rfloor$ and $G(n,n-1) \geq n + \lfloor \log_2(n-2) \rfloor$.*

Proof. We prove the theorem by constructing a suitable vector $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$.

For $k = 1$, let $a_1 = -2^{n-1} + 1$ and $a_i = 2^{n-1} + 2^{i-2} - 1$ for $i \in \{2, \ldots, n\}$. Let

$$T_i = \{1, i+1\} \quad \text{for } i \in [n-1],$$
$$T_i = \{2, 3, \ldots, 2^{i-n} + 1\} \quad \text{for } i \in \{n, \ldots, n + \lfloor \log_2(n-1) \rfloor\}.$$

Then for $i \in [n-1]$,

$$\mathbf{a}(T_i) = a_1 + a_{i+1} = -2^{n-1} + 1 + 2^{n-1} + 2^{i-1} - 1 = 2^{i-1}.$$

Furthermore,

$$\mathbf{a}(T_n) = a_2 = 2^{n-1} + 2^0 - 1 = 2^{n-1}$$

and for $i \in \{n, \ldots, n + \lfloor \log_2(n-1) \rfloor\}$ the sums $\mathbf{a}(T_i)$ are increasing as $a_2, \ldots, a_n$ is increasing and every sum has twice as many elements as the preceding sum.

For $k = n-1$, there is an analogous construction. Namely, let $\mathbf{a}' = (\mathbf{a}([n]), -a_2, \ldots, -a_n)$ and let

$$T_i' = ([n] \setminus T_i) \cup \{1\} \quad \text{for } i \in [n-1]$$
$$T_i' = [n] \setminus \{2, \ldots, 2^{i-n} + 2\} \quad \text{for } i \in \{n, \ldots, n + \lfloor \log_2(n-2) \rfloor\}$$

Note that $\mathbf{a}(T_i) = \mathbf{a}'(T_i')$ for $i \in [n-1]$ and the remaining sets are also geometrically increasing. Then, T_i' is a geometrically increasing sequence with respect to $\mathbf{a}'$.

The intuition behind the construction was to fix the sets first so that the first n can be computed as in a system of equations.

□

Theorem 4.2 implies that $G(n,1) > n = G(n,0)$. Therefore, in Theorem 4.1, we can replace $\max(G(n,1), G(n,0))$ with $G(n,1)$, getting the following corollary.

Corollary 4.2.1. *For any n, it holds that $G(n,n-1) \leq G(n,1)$.*

Since the lower bound for $G(n,1)$ and $G(n,n-1)$ seems to be tight, at least for small n, a natural step is to try to get an upper bound that is $O(n)$. We provide a proof of the bound $G(n,1) \leq O(n) + log_2(A)$, where A is the maximal absolute value of the components in $\mathbf{a}$.

Theorem 4.3. *For any n, it holds that $G(n,1) \leq O(n) + \log_2 A$, where A is the maximal absolute value of the components in $\mathbf{a}$ when we scale them to integers.*

Proof. Let $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$. Without loss of generality, let $a_1 < 0$ and $a_2, \ldots, a_n > 0$. Then, let $S \subseteq \{2, \ldots, n\}$ be the set of indices of components of $\mathbf{a}$ which are larger than $-a_1$. If we add to $\mathbf{a}$ the elements $a_j + a_1$ for $j \in S$, we can swap every sum containing a_1 and some a_j with $j \in S$, with a sum containing the element $a_j + a_1$. Therefore, this is a sum of only positive numbers and we can bound the number of these sets by n. This means that to prove a bound of at least $O(n)$, we can, without loss of generality, assume $a_2, \ldots, a_n < -a_1$. Therefore, the maximal absolute value is $A = |a_1|$. For $T_i \in [n]$, we can bound

$$1 \leq \mathbf{a}(T_i) \leq -na_1.$$

Therefore, we can have at most $\log_2(-na_1) = \log_2 n + \log_2 A$ sets. In total, we have $G(n, 1) \leq O(n) + \log_2 n + \log_2 A \leq O(n) + \log_2 A$.

□

5 An inequality for $G(n, k)$

Furthermore, we establish a connection between the value of $G(n, k)$ and those of $G(n + 1, n - k)$ and $G(n + 1, n - k + 1)$. Similar bounds can help us gain insight into the behavior of the function. Our result is the following.

Theorem 5.1. *For any given n and k, it holds that $G(n, k) \leq \max(G(n + 1, n - k), G(n + 1, n - k + 1))$.*

Proof. The proof resembles that of Theorem 4.1. We aim to demonstrate that for any vector $\mathbf{a} \in \mathbb{R}^n$ with k negative components, we can construct a vector in $\mathbb{R}^{n+1}$ with $n-k$ or $n-k+1$ negative components, along with a geometrically increasing sequence with the same length as the one constructed from $\mathbf{a}$.

Without loss of generality, let $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$ be such that $a_1, \ldots, a_k < 0$ and $a_{k+1}, \ldots, a_n > 0$, and let the sets $T_1, \ldots, T_m \in [n]$ form a geometrically increasing sequence with respect to $\mathbf{a}$.

Let $S_i = T_i \cap \{1, \ldots, k\}$ and let $F_i = T_i \cap \{k+1, \ldots, n\}$ for all $i \in [m]$. Then $T_i = S_i \cup F_i$ and $F_i \cap S_i = \varnothing$ for all $i \in [m]$. Therefore,

$$\mathbf{a}(S_1) + \mathbf{a}(F_1), \ldots, \mathbf{a}(S_m) + \mathbf{a}(F_m)$$

form a geometrically increasing sequence. Let $x = a_1 + \cdots + a_k$ and $y = a_{k+1} + \cdots + a_n$. Define $S'_i = \{1, \ldots, k\} \setminus S_i$ and $F'_i = \{1, \ldots, k\} \setminus F_i$ for every $i \in [m]$. Then

$$\begin{aligned}\mathbf{a}(S_i) + \mathbf{a}(F_i) &= x - \mathbf{a}(\{1, \ldots, k\} \setminus S_i) + y - \mathbf{a}(\{k + 1, \ldots, n\} \setminus F_i) \\ &= (x + y) - \mathbf{a}(S'_i) - \mathbf{a}(F'_i).\end{aligned}$$

Let $\mathbf{b} = (-a_1, \ldots, -a_n, x + y)$. Then we have

$$(x + y) - \mathbf{a}(S'_i) - \mathbf{a}(F'_i) = b_{n+1} + \mathbf{b}(S'_i) + \mathbf{b}(F'_i).$$

Define $T'_i = S'_i \cup F'_i \cup \{n+1\}$ for all $i \in [m]$. Then $T'_1, \ldots, T'_m$ from a geometrically increasing sequence with respect to $\mathbf{b}$. Since $\mathbf{b} \in \mathbb{R}^{n+1}$ and $\mathbf{b}$ has $n-k$ or $n-k+1$ negative components, depending on the sign of $x + y$, then either

$$G(n, k) \leq G(n + 1, n - k),$$

or

$$G(n, k) \leq G(n+1, n-k+1).$$

This completes the proof. $\square$

6 A lower bound on $G(n, \frac{n}{2})$ for n a power of 2

In this section, we analyze the Fourier-analytic approach construction of Goldmann (1993, see [3]) using Hadamard matrices, giving a lower bound on $G(n, \frac{n}{2})$ when n is a power of 2.

Theorem 6.1. *Let $n = 2^p$ for some $p \in \mathbb{N}$. Define Q as in the proof of the lower bound by Goldmann (1993, see [3]), to be the Hadamard matrix of order n. Let the vector $\mathbf{b} = (b_1, \ldots, b_n) \in \mathbb{R}^n$ be such that $b_{i+1} \geq 2^p b_i > 0$ for $i = 2, \ldots, n-1$ and $b_1 > 0$. Then, the vector $\mathbf{a} \in \mathbb{R}^n$, such that $Q\mathbf{a} = \mathbf{b}$ has exactly $\frac{n}{2}$ negative values.*

Proof. We will prove the statement by induction on p.

For $p = 1$, i.e., $n = 2$, the Hadamard matrix of order 2 is

$$Q = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Therefore, $a_1 + a_2 = b_1$ and $a_1 - a_2 = b_2$. The solutions to this are $a_1 = (b_1 + b_2)/2$ and $a_2 = (b_1 - b_2)/2$. Observe that $b_2 \geq 2b_1$ and $b_2 \neq 0$, therefore a_1 is positive and a_2 is negative.

Assume we have proved the statement for every $p \in [t]$ for some $t \in \mathbb{N}$. Consider the case for $t+1$. We can use Sylvester's construction [7] to derive the Hadamard matrix Q of order n using the Hadamard matrix Q' of order $\frac{n}{2}$

$$Q = \begin{bmatrix} Q' & Q' \\ Q' & -Q' \end{bmatrix}.$$

Since

$$Q\mathbf{a} = \mathbf{b},$$

using row subtractions and additions we get

$$\begin{bmatrix} Q' & 0 \\ 0 & 0 \end{bmatrix} \mathbf{a} = \mathbf{c} = \begin{bmatrix} b_1 + b_{n/2+1} \\ \vdots \\ b_{n/2} + b_n \end{bmatrix},$$

and

$$\begin{bmatrix} 0 & 0 \\ 0 & Q' \end{bmatrix} (-\mathbf{a}) = \mathbf{d} = \begin{bmatrix} -b_1 + b_{n/2+1} \\ \vdots \\ -b_{n/2} + b_n \end{bmatrix}.$$

Let $\mathbf{x} = (a_1, \ldots, a_{n/2})$ and $\mathbf{y} = (-a_{n/2+1}, \ldots, -a_n)$, we have

$$Q'\mathbf{x} = \mathbf{c}$$

and

$$Q'\mathbf{y} = \mathbf{d}.$$

Note that for all $i \in \{2, \ldots, m\}$

$$c_i = b_i + b_{n/2+i} \geq 2^{t+1} b_{i-1} + 2^{t+1} b_{n/2+i-1} \geq 2^t (b_{n/2+i-1} + b_{i-1}) = 2^t c_{i-1},$$

and

$$d_i = -b_i + b_{n/2+i} \geq (2^{t+1} - 1) b_{n/2+i-1} \geq 2^t (b_{n/2+i-1} - b_{i-1}) = 2^t d_{i-1}.$$

By the inductive assumption, both $\mathbf{x}$ and $\mathbf{y}$ have $\frac{n}{4}$ negative components, therefore $\mathbf{a}$ has $\frac{n}{2}$ negative components.

$\square$

The vector $\mathbf{a}$, used in the lower bound construction of Goldmann [3] is the same as in Theorem 6.1. This implies the following corollary.

Corollary 6.1.1. *For n being a power of 2, it holds that $G(n, \frac{n}{2}) \geq \frac{1}{2} n \log_2 n - O(n \log \log n)$.*

7 The case when a is drawn from a normal distribution

Randomized inputs sometimes lead to nice mathematical behavior. Therefore, it is interesting to study the case when $\mathbf{a}$ is chosen from a normal distribution, which might shed some light on a better analysis of the discrete Newton's algorithm when the input is normally distributed. Concretely,

Problem 7.1. For a given $n \in \mathbb{N}$, we define $G_{\mathcal{N}}(n)$ to be the expected value of the largest number $m \in \mathbb{N}$ where for $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$ chosen from a normal distribution $\mathcal{N} = \mathcal{N}(0, \sigma^2 I)$ there exist $T_1, \ldots, T_m \subseteq [n]$ such that $\mathbf{a}(T_i) > 0$ for all i and $\mathbf{a}(T_{i+1}) \geq 2\mathbf{a}(T_i)$ for all $i \in [m-1]$. What does the function $G_{\mathcal{N}}(n)$ look like?

To gain insight into the problem, we conducted a computational experiment using a computer program to analyze the asymptotic behavior of $G_{\mathcal{N}}(n)$. For the sake of efficiency, the algorithm rounds the values chosen for the components of a, which allows the usage of the knapsack algorithm. The obtained results are plotted in Figure 1. From Figure 1, we observe that $G_{\mathcal{N}}(n)$ grows at a similar rate as $\log_2 n$. Therefore, we conjecture that it grows logarithmically.

Conjecture 7.2. *For all n, it holds that $G_{\mathcal{N}}(n) = O(\log n)$.*

8 Conclusion and discussion

In this work, we examined the problem of determining the length of the longest geometrically increasing sequence. In particular, we analyzed what happens in the case of fixing the number of negative values we use. We obtained some exact values of $G(n, k)$ for small values of n ($n \leq 4$) using a computer program. We then presented a connection between $G(n, 1)$ and $G(n, n-1)$, as well as a lower bound for both of them. Further, we showed a

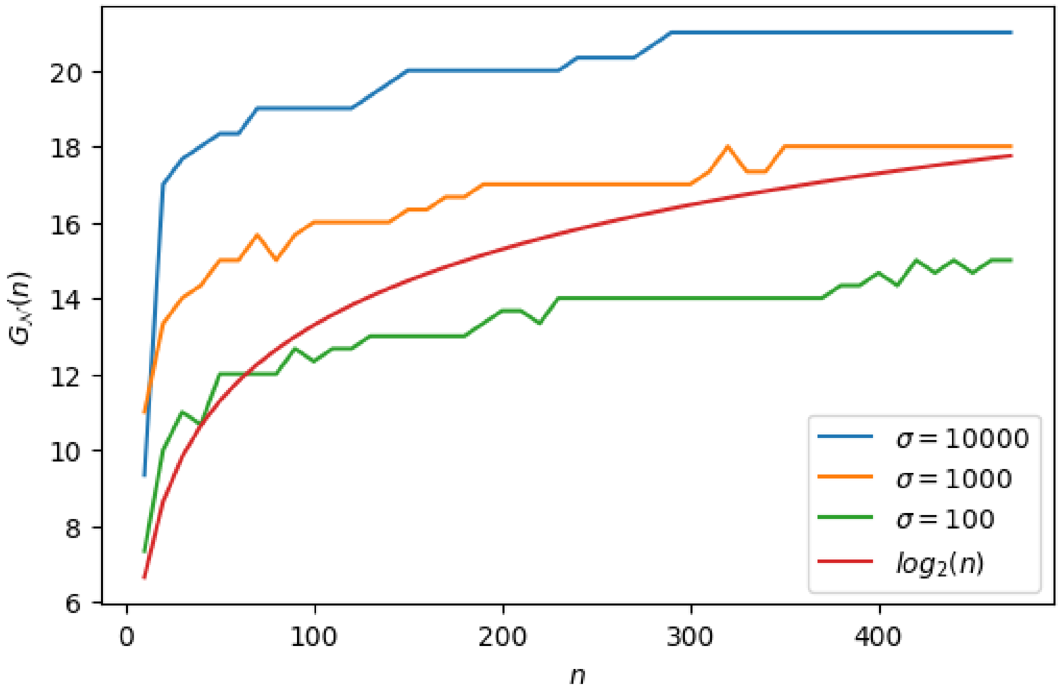

Figure 1: Results from the numerical experiment using a computer program with $\sigma \in \{10000, 1000, 100\}$.

connection between different values of $G(n, k)$, as well as a lower bound for $G(n, \frac{n}{2})$. Further, we conducted a computational experiment, analyzing the problem with the condition that the numbers are drawn from a normal distribution. It led us to conjecture that the function grows logarithmically in this case. For future work, we can try to prove a general lower bound on $G(n, k)$ that is generalized by the lower bound on $G(n, 1)$ and $G(n, n-1)$. Furthermore, it seems from the computational results that the lower bound for $G(n, 1)$ and $G(n, n-1)$ is tight. Therefore, we conjecture that $G(n, 1)$ is asymptotically linear.

Conjecture 8.1. *For all n, it holds that $G(n, 1) = O(n)$.*

For a fixed n, it seems that with k increasing, $G(n, k)$ is also increasing up to some point ($G(n, 0) = n$, $G(n, 1) > n$) and then is decreasing ($G(n, n-1) > n$, $G(n, n) = 0$). Therefore, we may state this as a conjecture.

Conjecture 8.2. *For all n, there exists t, such that*

$$G(n, k) \geq G(n, k-1) \quad \textit{for all } k \in \{2, \ldots, t\}$$

and

$$G(n, k) \geq G(n, k+1) \quad \textit{for all } k \in \{t, \ldots, n-1\}$$

It would be interesting to analyze this behavior and potentially find a good bound for some interval of k's. Therefore, we give this conjecture.

Conjecture 8.3. *There exist numbers* $0 < c_1 < c_2 < 1$, $c_3 \in \mathbb{R}_{>0}$, *such that for every* n, *for every* $k \in [c_1 n, c_2 n] \cap \mathbb{N}$,

$$G(n,k) \geq c_3 n \log_2 n.$$

Another direction worth considering is analyzing ratios different from 2. In particular, the values between 1 and 2 might be interesting, because as the factor tends to 1, the length of the longest geometrically increasing sequence tends to $2^n - 1$, while when the factor tends to 2, the length is $O(n \log n)$.

Lastly, similarly to considering the case when $\mathbf{a}$ is chosen from a normal distribution, it will be interesting to consider the case when it is arbitrarily chosen.

9 Acknowledgements

I would like to thank my mentor Yuchong Pan for suggesting the project's topic and guiding me through the research. I also want to acknowledge the help of Dr. Tanya Khovanova for the organization and presentation of the project. Likewise, I thank my tutor Petar Gaydarov, who helped me with my writing and my presentation. Further, I want to thank RSI's Teaching assistants: Allen Lin, Sally Zhu, AnAn Desimone, and Rich Wang, for providing their valuable help on the presentation of this project. I want to thank Eleonora Pavlova (principal of the High School of Mathematics "Dr. Petar Beron", where I study) for creating a beneficial atmosphere for high school research, and for introducing me to it, together with Momchil Peychev. Then, I want to acknowledge Dr. Konstantin Delchev for helping me prepare to present my research. I also acknowledge the financial support of Veselin Kulev (RSI '05), who made RSI possible for me. Likewise, I want to thank Acad. Petar St. Kenderov for his support. I want to thank Miroslav Marinov and Radoslav Dimitrov for being my research mentors before RSI. I am also grateful for the help of Dr. Jenny Sendova. I am also grateful for the appreciation of Boyka Palanchova. Other sponsors and institutions I want to thank are Cyril and St. Methodius International Foundation, the Union of Bulgarian Mathematicians, the Institute of Mathematics and Informatics at the Bulgarian Academy of Sciences, "MB Consulting" EOOD, and Prosveta-Sofia Foundation. The development of the project would not have been possible without the facilitation of the Research Science Institute, the Center for Excellence in Education, and the Massachusetts Institute of Technology.

References

[1] T. Radzik. *Fractional Combinatorial Optimization*, pp. 429–478. Springer US, Boston, MA, 1998.

[2] M. Goemans. Tutte 100th: May 12, 2017: Professor Michel Goemans. URL `https://www.youtube.com/watch?v=tkPokWOlJi8&ab_channel=Combinatorics%26OptimizationUniversityofWaterloo`.

[3] M. Goemans, S. Gupta, and P. Jaillet. Discrete newton's algorithm for parametric submodular function minimization. 2017.

[4] J. Håstad. On the size of weights for threshold gates. *SIAM Journal on Discrete Mathematics*, 7(3):pp. 484–492, 1994.

[5] A. Schrijver. *Combinatorial Optimization: Polyhedra and Efficiency*. Algorithms and Combinatorics. Springer Berlin Heidelberg, 2003.

[6] J. Hadamard. Resolution d'une question relative aux determinants. *Bull. des Sciences Math.*, 2:pp. 240–246, 1893.

[7] J. Sylvester. Lx. thoughts on inverse orthogonal matrices, simultaneous signsuccessions, and tessellated pavements in two or more colours, with applications to newton's rule, ornamental tile-work, and the theory of numbers. *Philosophical Magazine Series 1*, 34:pp. 461–475, 1867.