# Discrete Newton's Algorithm for Parametric Submodular Function Minimization

Michel X. Goemans, Swati Gupta, and Patrick Jaillet

Massachusetts Institute of Technology, USA,
goemans@math.mit.edu, swatig@mit.edu, jaillet@mit.edu

**Abstract.** We consider the line search problem in a submodular polytope $P(f) \subseteq \mathbb{R}^n$: Given an arbitrary $a \in \mathbb{R}^n$ and $x_0 \in P(f)$, compute $\max\{\delta : x_0 + \delta a \in P(f)\}$. The use of the discrete Newton's algorithm for this line search problem is very natural, but no strongly polynomial bound on its number of iterations was known [Iwata, 2008]. We solve this open problem by providing a quadratic bound of $n^2 + O(n \log^2 n)$ on its number of iterations. Our result considerably improves upon the only other known strongly polynomial time algorithm, which is based on Megiddo's parametric search framework and which requires $\tilde{O}(n^8)$ submodular function minimizations [Nagano, 2007]. As a by-product of our study, we prove (tight) bounds on the length of chains of ring families and geometrically increasing sequences of sets, which might be of independent interest.

**Keywords:** Discrete Newton's algorithm, submodular functions, line search, ring families, geometrically increasing sequence of sets, fractional combinatorial optimization

## 1 Introduction

Let $f$ be a submodular function on $V$, where $|V| = n$. We often assume that $V = [n] := \{1, 2, \cdots, n\}$. Let $P(f) = \{x \in \mathbb{R}^n \mid x(S) \leq f(S) \text{ for all } S \subseteq V\}$. The only assumption we make on $f$ is that $f(\emptyset) \geq 0$ (otherwise $P(f)$ is empty). Given $x_0 \in P(f)$ (this condition can be verified by performing a single submodular function minimization) and $a \in \mathbb{R}^n$, we would like to find the largest $\delta$ such that $x_0 + \delta a \in P(f)$. For any vector $b \in \mathbb{R}^n$ and any set $S \subseteq V$, it is convenient to use the notation $b(S) := \sum_{e \in S} b_e$. By considering the submodular function $f'$ taking the value $f'(S) = f(S) - x_0(S)$ for any set $S$, we can equivalently find the largest $\delta$ such that $\delta a \in P(f')$. Since $x_0 \in P(f)$ we know that $0 \in P(f')$ and thus $f'$ is nonnegative. Thus, without loss of generality, we consider the problem

$$\delta^* = \max\left\{\delta : \min_{S \subseteq V} f(S) - \delta a(S) \geq 0\right\}, \tag{1}$$

for a nonnegative submodular function $f$.

Since $x_0 = 0 \in P(f)$ we know that $\delta^* \geq 0$ and that the minimum could be taken only over the sets $S$ with $a(S) > 0$, although we will not be using this fact. To make this problem nontrivial, we assume that there exists some $i$ with $a_i > 0$. Geometrically, the problem of finding $\delta^*$ is a line search problem. As we go along the line segment $\ell(\delta) = x_0 + \delta a$ (or just $\delta a$ if we assume $x_0 = 0$), when do we exit the submodular polyhedron $P(f)$? This is a basic subproblem needed in many algorithmic applications. For example, for the algorithmic version of Carathéodory's theorem (over any polytope), one typically performs a line search from a vertex of the face being considered in a direction within the same face. This is, for example, also the case for variants of the Frank-Wolfe algorithm (see for instance [Freund et al., 2015]).

A natural way to solve this line search problem is to use a cutting plane approach. Start with any upper bound $\delta_1 \geq \delta^*$ and define the point $x^{(1)} = \delta_1 a$. One can then generate a most violated inequality for $x^{(1)}$, where most violated means the one minimizing $f(S) - \delta_1 a(S)$ over all sets $S$. The hyperplane corresponding to a minimizing set $S_1$ intersects the line in $x^{(2)} = \delta_2 a$. Proceeding analogously, we obtain a sequence of points and eventually will reach the optimum $\delta$.

This cutting-plane approach is equivalent to Dinkelbach's algorithm or the discrete Newton's algorithm for solving (1). At the risk of repeating ourselves, we let $\delta_1 \geq \delta^*$. For example we could set $\delta_1 =$

$\min_{i:a_i>0} f(\{i\})/a_i$. At iteration $i \geq 1$ of Newton's algorithm, we consider the submodular function $k_i(S) = f(S) - \delta_i a(S)$, and compute

$$h_i = \min_S k_i(S),$$

and define $S_i$ to be *any* minimizer of $k_i(S)$. Now, let $f_i = f(S_i)$ and $g_i = a(S_i)$. As long as $h_i < 0$, we proceed and set

$$\delta_{i+1} = \frac{f_i}{g_i}.$$

As soon as $h_i = 0$, Newton's algorithm terminates and we have that $\delta^* = \delta_i$. We give the full description of the discrete Newton's algorithm in Algorithm 1.

**Algorithm 1:** DISCRETE NEWTON'S ALGORITHM

**input** : submodular $f : 2^V \to \mathbb{R}$, $f$ nonnegative, $a \in \mathbb{R}^n$
**output**: $\delta^* = \max \{\delta : \min_S f(S) - \delta a(S) \geq 0\}$
$i = 0, \delta_1 = \min_{i \in V, a(\{i\})>0} f(\{i\})/a(\{i\})$;
**repeat**
    $i = i + 1$;
    $h_i = \min_{S \subseteq V} f(S) - \delta_i a(S)$;
    $S_i \in \arg\min_{S \subseteq V} f(S) - \delta_i a(S)$;
    $\delta_{i+1} = \frac{f(S_i)}{a(S_i)}$;
**until** $h_i = 0$;
Return $\delta^* = \delta_i$.

When $a \geq 0$, it is known that Newton's algorithm terminates in at most $n$ iterations (for e.g. [Topkis, 1978]). Even more, the function $g(\delta) := \min_S f(S) - \delta a(S)$ is a concave, piecewise affine function with at most $n$ breakpoints (and $n + 1$ affine segments) since for any set $\{\delta_i\}_{i \in I}$ of $\delta$ values, the submodular functions $f(S) - \delta_i a(S)$ for $i \in I$ form a sequence of strong quotients (ordered by the $\delta_i$'s), and therefore the minimizers form a chain of sets. See [Iwata et al., 1997] for definitions of strong quotients and details.

When $a$ is arbitrary (not necessarily nonnegative), little is known about the number of iterations of the discrete Newton's algorithm. The number of iterations can easily be bounded by the number of possible distinct positive values of $a(S)$, but this is usually very weak (unless, for example, the support of $a$ is small as is the case in the calculation of exchange capacities). A weakly polynomial bound involving the sizes of the submodular function values is easy to obtain, but no strongly polynomial bound was known, as mentioned as an open question in [Nagano, 2007], [Iwata, 2008]. In this paper, we show that the number of iterations is quadratic. This is the first strongly polynomial bound in the case of an arbitrary $a$.

**Theorem 1.** *For any submodular function $f : 2^{[n]} \to \mathbb{R}_+$ and an arbitrary direction $a$, the discrete Newton's algorithm takes at most $n^2 + O(n \log^2(n))$ iterations.*

Previously, the only strongly polynomial algorithm to solve the line search problem in the case of an arbitrary $a \in \mathbb{R}^n$ was an algorithm of Nagano et al. [Nagano, 2007] relying on Megiddo's parametric search framework. This requires $\tilde{O}(n^8)$ submodular function minimizations, where $\tilde{O}(n^8)$ corresponds to the current best running time known for *fully combinatorial* submodular function minimization [Iwata and Orlin, 2009]. On the other hand, our main result in Theorem 1 shows that the discrete Newton's algorithm takes $O(n^2)$ iterations, i.e. $O(n^2)$ submodular function minimizations, and we can use any submodular function minimization algorithm. Each submodular function minimization can be computed, for example, in $\tilde{O}(n^4 + \gamma n^3)$ time using a result of [Lee et al., 2015], where $\gamma$ is the time for an evaluation of the submodular function.

Radzik [Radzik, 1998] provides an analysis of the discrete Newton's algorithm for the related problem of $\max \delta : \min_{S \in \mathcal{S}} b(S) - \delta a(S) \geq 0$ where both $a$ and $b$ are modular functions and $\mathcal{S}$ is an arbitrary collection of sets. He shows that the number of iterations of the discrete Newton's algorithm is at most $O(n^2 \log^2(n))$.

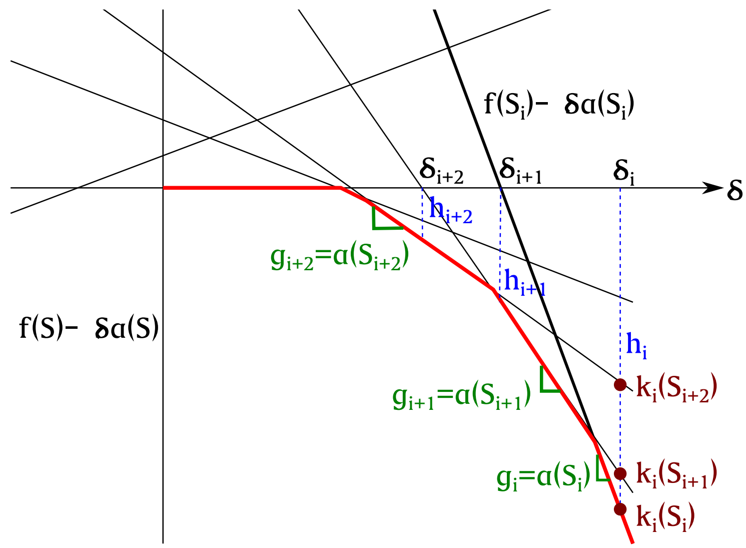


**Fig. 1.** Illustration of Newton's iterations and notation in Lemma 1.

Our analysis does not handle an arbitrary collection of sets, but generalizes his setting as it applies to the more general case of submodular functions $f$. Note that considering submodular functions (as opposed to modular functions) makes the problem considerably harder since the number of input parameters for modular functions is only $2n$, whereas in the case of submodular functions the input is exponential (we assume oracle access for function evaluation).

Apart from the main result of bounding the number of iterations of the discrete Newton's algorithm for solving $\max \delta : \min_S f(S) - \delta a(S) \geq 0$ in Section 3, we prove results on ring families (set families closed under taking intersections and unions) and geometrically increasing sequences of sets, which may be of independent interest. As part of the proof of Theorem 1, we first show a tight (quadratic) bound on the length of a sequence $T_1, \cdots, T_k$ of sets such that no set in the sequence belongs to the smallest ring family generated by the previous sets (Section 2). Further, one of the key ideas in the proof of Theorem 1 is to consider a sequence of sets (each set corresponds to an iteration in the discrete Newton's algorithm) such that the value of a submodular function on these sets increases geometrically. We show a quadratic bound on the length of such sequences for any submodular function and construct two (related) examples to show that this bound is tight, in Section 4. Interestingly, one of these examples is a construction of intervals and the other example is a weighted directed graph where the cut function already gives such a sequence of sets.

## 2 Ring families

A *ring family* $\mathcal{R} \subset 2^V$ is a family of sets closed under taking unions and intersections. From Birkhoff's representation theorem, we can associate to a ring family a directed graph $D = (V, E)$ in the following way. Let $A = \bigcap_{R \in \mathcal{R}} R$ and $B = \bigcup_{R \in \mathcal{R}} R$. Let $E = \{(i, j) \mid R \in \mathcal{R}, i \in R \Rightarrow j \in R\}$. Then for any $R \in \mathcal{R}$, we have that (i) $A \subseteq R$, (ii) $R \subseteq B$ and (iii) $\delta^+(R) = \{(i, j) \in E \mid i \in R, j \notin R\} = \emptyset$. But, conversely, any set $R$ satisfying (i), (ii) and (iii) must be in $\mathcal{R}$. Indeed, for any $i \neq j$ with $(i, j) \notin E$, there must be a set $U_{ij} \in \mathcal{R}$ with $i \in U_{ij}$ and $j \notin U_{ij}$. To show that a set $R$ satisfying (i), (ii) and (iii) is in $\mathcal{R}$, it suffices to observe that

$$R = \bigcup_{i \in R} \bigcap_{j \notin R} U_{ij}, \tag{2}$$

and therefore $R$ belongs to the ring family.

Given a collection of sets $\mathcal{T} \subseteq 2^V$, we define $\mathcal{R}(\mathcal{T})$ to be the smallest ring family containing $\mathcal{T}$. The directed graph representation of this ring family can be obtained by defining $A$, $B$ and $E$ directly from $\mathcal{T}$ rather than from the larger $\mathcal{R}(\mathcal{T})$, i.e. $A = \bigcap_{R \in \mathcal{T}} R = \bigcap_{R \in \mathcal{R}(\mathcal{T})} R$, $B = \bigcup_{R \in \mathcal{T}} R = \bigcup_{R \in \mathcal{R}(\mathcal{T})} R$, and $E = \{(i,j) \mid R \in \mathcal{T}, i \in R \Rightarrow j \in R\}$. Further, in the expression (2) of any set $R \in \mathcal{R}(\mathcal{T})$, we can use sets $U_{ij} \in \mathcal{T}$.

Given a sequence of subsets $T_1, \cdots, T_k$ of $V$, define $\mathcal{L}_i := \mathcal{R}(\{T_1, \cdots, T_i\})$ for $1 \leq i \leq k$. Assume that for each $i > 1$, we have that $T_i \notin \mathcal{L}_{i-1}$. We should emphasize that this condition depends on the *ordering* of the sets, and not just on this collection of sets. For instance, $\{1\}, \{1,2\}, \{2\}$ is a valid ordering whereas $\{1\}, \{2\}, \{1,2\}$ is not. We have thus a chain of ring families: $\mathcal{L}_1 \subset \mathcal{L}_2 \subset \cdots \subset \mathcal{L}_k$ where all the containments are proper. The question is how large can $k$ be, and the next theorem shows that it can be at most quadratic in $n$.

**Theorem 2.** *Consider a chain of ring families, $\mathcal{L}_0 = \emptyset \neq \mathcal{L}_1 \subsetneq \mathcal{L}_2 \subsetneq \cdots \subsetneq \mathcal{L}_k$ within $2^V$ with $n = |V|$. Then*

$$k \leq \binom{n+1}{2} + 1.$$

Before proving this theorem, we show that the bound on the number of sets is tight.

*Example 1.* Let $V = \{1, \cdots, n\}$. For each $1 \leq i \leq j \leq n$, consider intervals $[i,j] = \{k \in V \mid i \leq k \leq j\}$. Add also the empty set $\emptyset$ as the trivial interval $[0,0]$ (as $0 \notin V$). We have just defined $k = \binom{n+1}{2} + 1$ sets. Define a complete order on these intervals in the following way: $(i,j) \prec (s,t)$ if $j < t$ or ($j = t$ and $i < s$). We claim that if we consider these intervals in the order given by $\prec$, we satisfy the main assumption of the theorem that $[s,t] \notin \mathcal{R}(\mathcal{T}_{st})$ where $\mathcal{T}_{st} = \{[i,j] \mid (i,j) \prec (s,t)\}$. Indeed, for $s = 1$ and any $t$, we have that $[1,t] \notin \mathcal{R}(\mathcal{T}_{1t})$ since $\bigcup_{I \in \mathcal{T}_{1t}} I = [1, t-1] \not\supset [1,t]$. On the other hand, for $s > 1$ and any $t$, we have that $[s,t] \notin \mathcal{R}(\mathcal{T}_{st})$ since for all $I \in \mathcal{T}_{st}$ we have ($t \in I \Rightarrow s - 1 \in I$) while this is not the case for $[s,t]$.

*Proof.* For each $1 \leq i \leq k$, let $T_i \in \mathcal{L}_i \setminus \mathcal{L}_{i-1}$. We can assume that $\mathcal{L}_i = \mathcal{R}(\{T_1, \cdots T_i\})$ (otherwise a longer chain of ring families can be constructed). If none of the $T_i$'s is the empty set, we can increase the length of the chain by considering (the ring families generated by) the sequence $\emptyset, T_1, T_2, \cdots, T_k$. Similarly if $V$ is not among the $T_i$'s, we can add $V$ either in first or second position in the sequence. So we can assume that the sequence has $T_1 = \emptyset$ and $T_2 = V$, i.e. $\mathcal{L}_1 = \{\emptyset\}$ and $\mathcal{L}_2 = \{\emptyset, V\}$.

When considering $\mathcal{L}_2$, its digraph representation has $A = \emptyset$, $B = V$ and the directed graph $D = (V, E)$ is the bidirected complete graph on $V$. To show a weaker bound of $k \leq 2 + n(n-1)$ is easy: every $T_i$ we consider in the sequence will remove at least one arc of this digraph and no arc will get added.

To show the stronger bound in the statement of the theorem, consider the digraph $D'$ obtained from $D$ by contracting every strongly connected component of $D$ and discarding all but one copy of (possibly) multiple arcs between two vertices of $D'$. We keep track of two parameters of $D'$: $s$ is its number of vertices and $a$ is its the number of arcs. Initially, when considering $\mathcal{L}_2$, we have $s = 1$ strongly connected component and $D'$ has no arc: $a = 0$. Every $T_i$ we consider will either keep the same strongly connected components in $D$ (i.e. same vertices in $D'$) and remove (at least) one arc from $D'$, or will break up at least one strongly connected component in $D$ (i.e. increases vertices in $D'$). In the latter case, we can assume that only one strongly connected component is broken up into two strongly connected components and the number of arcs added is at most $s$ since this newly formed connected component may have a single arc to every other strongly connected component. Thus, in the worst case, we move either from a digraph $D'$ with parameters $(s,a)$ to one with $(s, a-1)$ or from $(s,a)$ to $(s+1, a+s)$. By induction, we claim that if the original one has parameters $(s,a)$ then the number of steps before reaching the digraph on $V$ with no arcs with parameters $(n, 0)$ is at most

$$a + \binom{n+1}{2} - \binom{s+1}{2}.$$

Indeed, this trivially holds by induction for any step $(s,a) \to (s,a-1)$ and it also holds for any step $(s,a) \to (s+1,a+s)$ since:

$$(a+s) + \binom{n+1}{2} - \binom{s+2}{2} + 1 = a + \binom{n+1}{2} - \binom{s+1}{2}.$$

As the digraph corresponding to $\mathcal{L}_2$ has parameters $(1,0)$, we obtain that $k \leq 2 + \binom{n+1}{2} - 1 = \binom{n+1}{2} + 1$. □

## 3 Analysis of the discrete Newton's algorithm

To prove Theorem 1, we start by recalling Radzik's analysis of Newton's algorithm for the case of modular functions ([Radzik, 1998]). First of all, the discrete Newton's algorithm, as stated in Algorithm 1 for solving $\max \delta : \min_{S\subseteq V} f(S) - \delta a(S) \geq 0$ terminates. Recall that $h_i = \min_{S\subseteq V} f(S) - \delta_i a(S)$, $S_i \in \arg\min_S f(S) - \delta_i a(S)$, $g_i = a(S_i)$ and $\delta_{i+1} = \frac{f(S_i)}{a(S_i)}$. Let $f_i = f(S_i)$ and $g_i = a(S_i)$. Figure 1 illustrates the discrete Newton's algorithm and the notation.

**Lemma 1.** *Newton's algorithm as described in Algorithm 1 terminates in a finite number of steps t and generate sequences:*

*(i)* $h_1 < h_2 < \cdots < h_{t-1} < h_t = 0$,
*(ii)* $\delta_1 > \delta_2 > \cdots > \delta_{t-1} > \delta_t = \delta^* \geq 0$,
*(iii)* $g_1 > g_2 > \cdots > g_{t-1} > g_t \geq 0$.

*Furthermore, if $g_t > 0$ then $\delta^* = 0$.*

The first proof of the above lemma is often attributed to [McCormick and Ervolina, 1994] and is omitted for conciseness. As in Radzik's analysis, we use the following lemma, illustrated in Figure 2, and we reproduce here its proof.

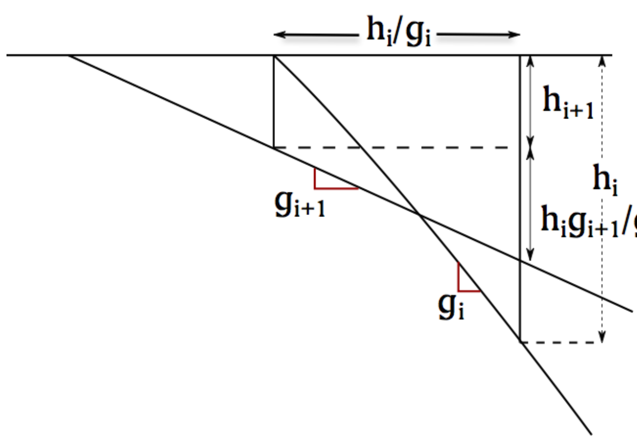


**Fig. 2.** Illustration for showing that $h_{i+1} + h_i \frac{g_{i+1}}{g_i} \leq h_i$, as in Lemma 2.

**Lemma 2.** *For any $i < t$, we have $\frac{h_{i+1}}{h_i} + \frac{g_{i+1}}{g_i} \leq 1$.*

*Proof.* By definition of $S_i$, we have that

$$\begin{aligned} h_i = f(S_i) - \delta_i a(S_i) = f_i - \delta_i g_i &\leq f(S_{i+1}) - \delta_i a(S_{i+1}) = f_{i+1} - \delta_i g_{i+1} \\ &= h_{i+1} + \frac{f_i}{g_i} g_{i+1} - \frac{f_i - h_i}{g_i} g_{i+1} = h_{i+1} + h_i \frac{g_{i+1}}{g_i}. \end{aligned}$$

Since $h_i < 0$, dividing by $h_i$ gives the statement. □

Thus, in every iteration, either $g_i$ or $h_i$ decreases by a constant factor smaller than 1. We can thus partition the iterations into two types, for example as

$$J_g = \left\{ i \mid \frac{g_{i+1}}{g_i} \leq \frac{2}{3} \right\}$$

and $J_h = \{i \notin J_g\}$. Observe that $i \in J_h$ implies $\frac{h_{i+1}}{h_i} < \frac{1}{3}$. We first bound $|J_g|$ as was done in [Radzik, 1998].

**Lemma 3.** $|J_g| = O(n \log n)$.

*Proof sketch.* Let $J_g = \{i_1, i_2, \cdots, i_k\}$ and let $T_j = S_{i_j}$. From the monotonicity of $g$, these sets $T_j$ are such that $a(T_{j+1}) \leq \frac{2}{3} a(T_j)$. These can be viewed as linear inequalities with small coefficients involving the $a_i$'s, and by normalizing and taking an extreme point of this polytope, Goemans (see [Radzik, 1998]) has shown that the number $k$ of such sets is $O(n \log n)$.

Although we do not need this for the analysis, the bound of $O(n \log n)$ on the number of geometrically decreasing sets defined on $n$ numbers is tight, as was shown by Mikael Goldmann in 1993 by an intricate construction based on a Fourier-analytic approach of Håstad [Håstad, 1994]. As this was never published, we include (a variant of) this construction here. The reader is welcome to skip directly to Section 3.1 without break in continuity.

**Theorem 3.** *Let $n$ be a power of 2. Then there exists $a \in \mathbb{R}^n$ and a sequence of sets $\{S_i\}_{i \in [k]}$ with $a(S_1) > 0$ and $a(S_i) \geq 2a(S_{i-1})$ for $i > 1$ where $k = \frac{1}{2} n \log_2 n - O(n \log \log n)$.*

*Proof.* Let $m$ be such that $n = 2^m$. Consider all $2^m$ subsets of $[m]$ and order them as $\alpha_1, \alpha_2, \cdots, \alpha_n$ such that $|\alpha_i| \leq |\alpha_j|$ for $i < j$. Thus, $\alpha_1 = \emptyset$ and $\alpha_n = [m]$. We say $\alpha_i \prec \alpha_j$ if $i < j$. Consider the $n \times n$ Hadamard matrix $Q$ in which the $i$th row and column are indexed by subset $\alpha_i$ of $[m]$ and

$$q_{ij} = (-1)^{|\alpha_i \cap \alpha_j|}.$$

$Q$ is invertible and $Q^{-1} = \frac{1}{n} Q$. Set $b_1 = 0$ and $b_i = 2^{mi}$ for $i > 1$. Now, let $a \in \mathbb{R}^n$ be the solution to $Qa = b$. We claim that there is a sequence of sets of length $\frac{1}{2} nm + O(n \log \log n) = \frac{1}{2} n \log n + O(n \log \log n)$ whose $a(\cdot)$ values increase geometrically by a factor of 2.

First, observe that $q_{1j} = 1$ for all $j$ and thus $a([n]) = 0$. This means that if we have a $r \in \{-1, 1\}^n$ such that $\langle r, a \rangle = p$ then $a(S) = \frac{p}{2}$ where $S = \{i | r_i = 1\} \subseteq [n]$. Thus we focus on constructing a sequence of vectors $r \in \{-1, 1\}^n$ whose inner product with $a$ increases geometrically. We already have $n-1$ such vectors, namely the rows $q_i \in \{-1, 1\}^n$ of $Q$ for $i > 1$: $\langle q_i, a \rangle = 2^{mi}$.

Now, for each $i > 1$, we show how to construct $\pm 1$ vectors $v$ such that $2^{m(i-1)} < \langle v, a \rangle < 2^{mi}$ and whose $a$ values increase geometrically. We will be able to construct one such set for almost all values between 1 and $|\alpha_i|$. Fix $i > 1$ and let $k = |\alpha_i|$. For any $\ell$ with $1 \leq \ell \leq k - 2$, consider a set $\alpha_{h_\ell} \subset \alpha_i$ of cardinality $\ell$. Define the vector

$$w^{(\ell)} = \sum_{u : \alpha_{h_\ell} \subseteq \alpha_u \subseteq \alpha_i} q_u.$$

Its $j$th component is:

$$\begin{aligned} w_j^{(\ell)} &= \sum_{u : \alpha_{h_\ell} \subseteq \alpha_u \subseteq \alpha_i} q_{uj} = \sum_{u : \alpha_{h_\ell} \subseteq \alpha_u \subseteq \alpha_i} (-1)^{|\alpha_u \cap \alpha_j|} = (-1)^{|\alpha_{h_\ell} \cap \alpha_j|} \sum_{u : \alpha_{h_\ell} \subseteq \alpha_u \subseteq \alpha_i} (-1)^{|(\alpha_u \setminus \alpha_{h_\ell}) \cap \alpha_j|} \\ &= \begin{cases} 0 & \text{if } (\alpha_i \setminus \alpha_{h_\ell}) \cap \alpha_j = \emptyset \\ 2^{k-\ell} (-1)^{|\alpha_{h_\ell} \cap \alpha_j|} & \text{otherwise} \end{cases} \end{aligned}$$

Now consider $v^{(\ell)} = 2^{1-(k-\ell)} w^{(\ell)} - q_h$. We claim this is a $\pm 1$ vector. Its $j$th component is equal to $-q_{hj} \in \{-1, 1\}$ if $(\alpha_i \setminus \alpha_{h_\ell}) \cap \alpha_j = \emptyset$ and, otherwise, is equal to

$$2(-1)^{|\alpha_{h_\ell} \cap \alpha_j|} - (-1)^{|\alpha_{h_\ell} \cap \alpha_j|} = (-1)^{|\alpha_{h_\ell} \cap \alpha_j|} \in \{-1, 1\}.$$

Now for this vector $v^{(\ell)}$ (corresponding to a given pair $\alpha_{h_\ell} \subset \alpha_i$), we have:

$$\langle v^{(\ell)}, a\rangle = 2^{1-(k-\ell)}\langle w^{(\ell)}, a\rangle - \langle q_{h_\ell}, a\rangle = 2^{1-(k-\ell)}\left(\sum_{u:\alpha_{h_\ell}\subseteq\alpha_u\subseteq\alpha_i} b_u\right) - b_{h_\ell}. \quad (3)$$

Now the $b_j$'s increase geometrically with $j$. In the summation (with $2^{k-\ell}$ terms), the dominant one will be $b_i = 2^{mi}$, and as a first approximation, we have that $\langle v^{(\ell)}, a\rangle$ is roughly $2^{mi+1-k+\ell}$, and therefore they appear to be between $b_{i-1}$ and $b_i$, and increase appropriately by a factor 2. Unfortunately, lower terms matter and, therefore, we need to select carefully the indices $h_\ell$'s.

A simple construction of these sets $\{\alpha_{h_\ell}\}$ is as follows. Let $f = \lceil \log_2 k\rceil$. For any $f \le \ell \le k-f$, let $\alpha_{h_\ell}$ be such that (i) $\alpha_{h_\ell} \cap [f]$ has as characteristic vector the $f$-bit representation of $k-f-\ell$ and (ii) the elements in $\alpha_{h_\ell} \cap ([k] \setminus [f])$ are chosen arbitrarily so that $|\alpha_{h_\ell}| = \ell$. Observe that (i) is possible for all $f \le \ell \le k-f$ since $k-f-\ell \le k-2f \le 2^f - 1$ and therefore $k-f-\ell$ can be represented by $f$ bits. And (ii) is feasible as well by our choice of $\ell$. We have just constructed $k-2f+1 \ge k - 2\log_2(k) - 1$ sets.

One can show (proof omitted for space considerations) that, for such a choice of $\{\alpha_{h_\ell}\}$, we have for $f \le \ell < k-f$: $\langle v^{(\ell+1)}, a\rangle \ge 2\langle v^{(\ell)}, a\rangle$.

The number of vectors/sets we have constructed this way is therefore at least:

$$\sum_{k=0}^{m}\binom{m}{k}(k - 2\log_2(k) - 1) \ge \frac{m+1}{2}2^m - 2\log_2(m)2^m - 2^m = \frac{1}{2}n\log_2(n) - O(n\log\log(n)),$$

and this completes the proof. □

### 3.1 Weaker upper bound

Before deriving the bound of $O(n^2)$ on $|J_g| + |J_h|$ for Theorem 1, we show how to derive a weaker bound of $O(n^3 \log n)$. For showing the $O(n^3\log n)$ bound, first consider a block of *consecutive* iterations $[u, v] := \{u, u+1, \cdots, v\}$ within $J_h$.

**Theorem 4.** *Let $[u, v] \subseteq J_h$. Then $|[u, v]| \le n^2 + n + 1$.*

The strategy of the proof is to show (i) that, for the submodular function $k_v(S) = f(S) - \delta_v a(S)$, the values of $k_v(S_i)$ for $i \in [u, v-1]$ form a geometrically decreasing series (Lemma 4), (ii) that each $S_i$ cannot be in the ring family generated by $S_{i+1}, \ldots, S_{v-1}$ (Lemma 5 and Theorem 5), and (iii) then conclude using our Theorem 2 on the length of a chain of ring families.

**Lemma 4.** *Let $[u, v] \subseteq J_h$. Then for $k_v(S) = f(S) - \delta_v a(S)$, we have (i) $k_v(S_v) = \min_S k_v(S) = h_v$, (ii) $k_v(S_{v-1}) = 0$, (iii) $k_v(S_{v-2}) > 2|h_v|$ and (iv) $k_v(S_{i-1}) > 2k_v(S_i)$ for $i \in [u+1, v-1]$.*

*Proof.* Since $\frac{g_{i+1}}{g_i} > \frac{2}{3}$ for all $i \in [u, v]$, Lemma 2 implies that $\frac{h_{i+1}}{h_i} \le \frac{1}{3}$, and thus

$$\frac{|h_{i+1}|}{g_{i+1}} \le \frac{1}{2}\frac{|h_i|}{g_i}.$$

Since $\delta_{i+1} - \delta_i = \frac{f_i}{g_i} - \frac{f_i - h_i}{g_i} = \frac{h_i}{g_i}$. We deduce that

$$\delta_{i+1} - \delta_{i+2} = -\frac{h_{i+1}}{g_{i+1}} \le \frac{1}{2}(\delta_i - \delta_{i+1}), \quad (4)$$

for all $i \in [u, v]$. Now, observe that for any $i \in [u, v-2]$, we have

$$\delta_{i+1} - \delta_v = \sum_{k=i+1}^{v-1} \delta_k - \delta_{k+1} \le \frac{1}{2}\sum_{k=i+1}^{v-1}(\delta_{k-1} - \delta_k) = \frac{1}{2}(\delta_i - \delta_{v-1}) < \frac{1}{2}(\delta_i - \delta_v).$$

Thus

$$\delta_{i+1} - \delta_v < \frac{1}{2}\left(\delta_i - \delta_v\right), \tag{5}$$

and we can even extend the range of validity to $i \in [u, v]$ since for $i = v - 1$ or $i = v$, this follows from Lemma 1.

Consider the submodular function $k_v(S) = f(S) - \delta_v a(S)$. We have denoted its minimum value by $h_v < 0$ and $S_v$ is one of its minimizers. For each $i \in [u, v-1]$ we have

$$k_v(S_i) = f_i - \delta_v g_i = g_i(\delta_{i+1} - \delta_v),$$

and therefore $k_v(S_{v-1}) = 0$ while $k_v(S_i) > 0$ for $i \in [u, v-2]$. Furthermore, (5) implies that

$$k_v(S_i) = g_i(\delta_{i+1} - \delta_v) < \frac{1}{2}\frac{g_i}{g_{i-1}} g_{i-1}(\delta_i - \delta_v) < \frac{1}{2} g_{i-1}(\delta_i - \delta_v) = \frac{1}{2} k_v(S_{i-1}),$$

and this is valid for $i \in [u, v-1]$. Thus the $k_v(S_i)$'s decrease geometrically with increasing $i$. In addition, we have $k_v(S_{v-2}) = g_{v-2}(\delta_{v-1} - \delta_v)$ while (by (4) and Lemma 1)

$$-k_v(S_v) = |h_v| = -h_v = g_v(\delta_v - \delta_{v+1}) < \frac{1}{2} g_{v-2}(\delta_{v-1} - \delta_v) = \frac{1}{2} k_v(S_{v-2}).$$

Summarizing, we have $k_v(S_v) = \min_S k_v(S) = h_v$, $k_v(S_{v-1}) = 0$, $k_v(S_{v-2}) > 2|h_v|$ and $k_v(S_{i-1}) > 2k_v(S_i)$ for $i \in [u, v-1]$. □

We now show that for any submodular function and any ring family on the same ground set, the values attained by the submodular function cannot increase much when the ring family is increased to the smallest ring family including a single additional set. This lemma follows from the submodularity of $f$ and Birkhoff's representation theorem for subsets contained in a ring family.

**Lemma 5.** *Let $f : 2^V \to \mathbb{R}$ be a submodular function with $f_{min} = \min_{S \subseteq V} f(S) \leq 0$. Let $\mathcal{L}$ be any ring family over $V$ and $T \notin \mathcal{L}$. Define $\mathcal{L}' := \mathcal{R}(\mathcal{L} \cup \{T\})$, $m = \max_{S \in \mathcal{L}} f(S)$ and $m' = \max_{S \in \mathcal{L}'} f(S)$. Then*

$$m' \leq 2(m - f_{min}) + f(T).$$

*Proof.* Consider $S \in \mathcal{L}'$. Using (2), we can express $S$ as $S = \bigcup_{i \in S} S_i$ where $S_i$ can be either (i) $T$, or (ii) $R$ for some $R \in \mathcal{L}$, or (iii) $R \cap T$ for some $R \in \mathcal{L}$. Taking the union of the sets $R$ of type (ii), resp. (iii), into $P$, resp. $Q$, we can express $S$ as $S = P \cup T$ or as $S = P \cup (Q \cap T)$ where $P, Q \in \mathcal{L}$ (since the existence of any case (i) annihilates the need for case (iii)).

Now using submodularity, we obtain that

$$f(P \cup T) \leq f(P) + f(T) - f(P \cap T) \leq m + f(T) - f_{min},$$

in the first case and

$$\begin{aligned} f(P \cup (Q \cap T)) &\leq f(P) + f(Q \cap T) - f(P \cap Q \cap T) \\ &\leq f(P) + f(Q) + f(T) - f(Q \cup T) - f(P \cap Q \cap T) \\ &\leq 2m + f(T) - 2f_{min}. \end{aligned}$$

In either case, we get the desired bound on $f(S)$ for any $S \in \mathcal{R}'$. □

We will now use the bound in Lemma 5 to show that if a sequence of sets increases in their submodular function value by a factor of 4, then any set in the sequence is not contained in the ring family generated by the previous sets.

**Theorem 5.** *Let $f : 2^V \to \mathbb{R}$ be a submodular function with $f_{min} = \min_{S \subseteq V} f(S) \leq 0$. Consider a sequence of distinct sets $T_1, T_2, \cdots, T_q$ such that $f(T_1) = f_{min}$, $f(T_2) > -2f_{min}$, and $f(T_i) \geq 4f(T_{i-1})$ for $3 \leq i \leq q$. Then $T_i \notin \mathcal{R}(\{T_1, \cdots, T_{i-1}\})$ for all $1 < i \leq q$.*

*Proof.* This is certainly true for $i = 2$. For any $i \geq 1$, define $\mathcal{L}_i = \mathcal{R}(\{T_1, \cdots, T_i\})$ and $m_i = \max_{S \in \mathcal{L}_i} f(S)$. We know that $m_1 = f_{min} \leq 0$ and $m_2 = f(T_2)$ since $T_1 \cap T_2$ and $T_1 \cup T_2$ cannot have larger $f$ values than $T_2$ by submodularity of $f$ and minimality of $T_1$.

We claim by induction that $m_k \leq 2f(T_k) + 2f_{min}$ for any $k \geq 2$. This is true for $k = 2$ since $m_2 = f(T_2) \leq 2f(T_2) + 2f_{min}$. Assume the induction claim to be true for $k - 1$.

We get that $m_{k-1} \leq 2f(T_{k-1}) + 2f_{min} < 4f(T_{k-1})$. Since $f(T_k) > m_{k-1}$, $T_k \notin \mathcal{L}_{k-1} = \mathcal{R}(T_1, \cdots, T_{k-1})$. Using Lemma 5, we get that

$$\begin{aligned} m_k &\leq 2(m_{k-1} - f_{min}) + f(T_k) \\ &\leq 2(2f(T_{k-1}) + 2f_{min} - f_{min}) + f(T_k) \\ &\leq 2f(T_k) + 2f_{min}. \end{aligned}$$

Thus proving the induction step for $k$, and hence the statement of the theorem. $\square$

We are now ready to prove Theorem 4.

*Proof.* (of Theorem 4) Apply Theorem 5 to the submodular function $k_v$ given in Lemma 4. Let $T_1 = S_v$ and skip every other set to define $T_i = S_{v-2(i-1)}$ for $v - 2(i-1) \geq u$ i.e. $i \leq q := 1 + (v-u)/2$. Then the conditions of Theorem 5 are satisfied (thanks to Lemma 4), and we obtain a sequence of sets $T_1, \cdots, T_q$ such that $T_i \notin \mathcal{R}(T_1, \cdots, T_{i-1})$. Therefore, Theorem 2 on the length of a chain of ring families implies that $q \leq \binom{n+1}{2} + 1$, or $v - u \leq (n+1)n$. This means $|[u, v]| \leq n^2 + n + 1$. $\square$

Since Lemma 3 shows that $|J_g| = O(n \log n)$ and we know from Theorem 4 that the intervals between two indices of $J_g$ have length $O(n^2)$, this implies that $|J_g| + |J_h| = O(n \log n) \cdot O(n^2) = O(n^3 \log n)$.

### 3.2 Main Result of Theorem 1

The analysis of Theorem 4 can be improved by showing that we can extract a chain of ring families not just from one interval of $J_h$ but from all of $J_h$. Instead of discarding every other set in $J_h$, we also need to discard the first $O(\log n)$ sets in every interval of $J_h$. This helps prove the main result of the paper that bounds the number of iterations in the discrete Newton's algorithm by at most $n^2 + O(n \log^2 n)$.

**Theorem 6.** *We have $|J_h| = n^2 + O(n \log^2 n)$.*

Before proving this, we need a variant of Lemma 5. The proof of the lemma again follows from the submodularity of $f$ and Birkhoff's representation theorem for subsets contained in a ring family.

**Lemma 6.** *Let $\mathcal{T} \subseteq 2^V$ and assume that $f(S) \leq M$ for all $S \in \mathcal{T}$. Then for all $S \in \mathcal{R}(\mathcal{T})$*

$$f(S) \leq \frac{n^2}{4}(M - f_{min}).$$

*Proof.* Consider any $S \in \mathcal{R}(\mathcal{T})$. We know that $S = \bigcup_{i \in S} \bigcap_{j \notin S} U_{ij}$, for some $U_{ij} \in \mathcal{T}$. Define $S_i = \bigcap_{j \notin S} U_{ij}$; thus $S = \bigcup_{i \in S} S_i$.

We first claim that, for any $k$ sets $T_1, T_2, \cdots, T_k \in \mathcal{T}$, we have that

$$f(\bigcap_{i=1}^{k} T_i) \leq kM - (k-1)f_{min}.$$

This is proved by induction on $k$, the base case of $k = 1$ being true by our assumption on $f$. Indeed, applying submodularity to $P = \bigcap_{i=1}^{k-1} T_i$ and $T_k$ (and the inductive hypothesis), we get

$$f(\bigcap_{i=1}^{k} T_i) = f(P \cap T_k) \leq f(P) + f(T_k) - f(P \cup T_k) \leq (k-1)M - (k-2)f_{min} + M - f_{min} = kM - (k-1)f_{min}.$$

Using this claim, we get that for any $i \in S$, we have

$$f(S_i) = f(\bigcap_{j \notin S} U_{ij}) \leq |V \setminus S|M - (|V \setminus S| - 1)f_{min} \leq |V \setminus S|(M - f_{min}).$$

By a similar argument on the union of the $S_i$'s, we derive that

$$\begin{aligned} f(S) &\leq |S|\left(|V \setminus S|M - (|V \setminus S| - 1)f_{min}\right) - (|S| - 1)f_{min} \\ &\leq |S||V \setminus S|M - (|S||V \setminus S| - 1)f_{min} \\ &\leq \frac{n^2}{4}(M - f_{min}). \end{aligned}$$

□

We are now ready to prove Theorem 6.

*Proof.* (of Theorem 6.) Let $J_h = \bigcup_{i=1}^{\ell}[u_i, v_i]$ where $u_{i-1} > v_i + 1$ for $1 < i \leq \ell$. Notice that these intervals are ordered in a *reverse* order (compared to the natural ordering). We construct a sequence of sets $T_1, \cdots$ such that each set in the sequence is not in the ring closure of the previous ones. The first sets are just every other set $S_i$ from $[u_1, v_1]$ obtained as before by using Theorem 5 and Lemma 4 with the submodular function $k_{v_1}$. Let $\mathcal{T}_1$ denote this sequence of sets.

Suppose now we have already considered the intervals $[u_j, v_j]$ for $j < i$ and have extracted a (long) sequence of sets $\mathcal{T}_{i-1}$ such that each set in the sequence is not in the ring closure of the previous ones. Consider now the submodular function $f := k_{v_i}$, and let $f_{min} \leq 0$ be its minimum value. Notice that from the order of iterations in the discrete Newton's algorithm we have that $f(T) < 0$ for $T \in \mathcal{T}_{i-1}$. Therefore by Lemma 6 with $M = 0$ we have that $f(S) \leq -\frac{n^2}{4} f_{min}$ for all $S \in \mathcal{R}(\mathcal{T}_{i-1})$. Using Lemma 4 with $f = k_{v_i}$, we have that only sets $S_k$ with $k > v_i - \log(n^2/4)$ could possibly be in $\mathcal{R}(\mathcal{T}_{i-1})$, and therefore we can safely add to $\mathcal{T}_{i-1}$ every other set in $[u_i, v_{i-O(\log n)}]$ while maintaining the property that every set is not in the ring closure of the previous ones. Over all $i$, we have thus constructed a chain of ring families of length $\frac{1}{2}|J_h| - O(\log n)\ell = \frac{1}{2}|J_h| - O(\log n)|J_g|$. The theorem now follows from Lemma 3 and Theorem 2. □

Finally, combining Theorem 6 and Lemma 3 proves Theorem 1.

*Proof.* (of Theorem 1.) In every iteration of discrete Newton's algorithm, either $g_i$ or $h_i$ decreases by a constant factor smaller than 1. Thus, the iterations can be partitioned into two types $J_g = \left\{ i \mid \frac{g_{i+1}}{g_i} \leq \frac{2}{3} \right\}$ and $J_h = \{i \notin J_g\}$. Lemma 3 shows that $|J_g| = O(n \log n)$ and Theorem 6 shows that $|J_h| = n^2 + O(n \log^2 n)$. Thus, the total number of iterations is $n^2 + O(n \log^2 n)$. □

## 4 Geometrically Increasing Sequences

In the proof for Theorem 1, we considered a sequence of sets $S_1, \cdots, S_k$ such that $f(S_i) \geq 4f(S_{i-1})$ for all $i \leq k$ for a submodular function $f$. In the special case when $f$ is modular, we know that the maximum length of such a sequence is at most $O(n \log n)$ (Lemma 3). When $f$ is submodular, we show that the maximum length is at most $\binom{n+1}{2} + 1$ by applying Theorem 2 to Theorem 5. In this section, we show that the bound for the submodular case is tight by constructing two related examples: one that uses interval sets of the ground set $\{1, \cdots, n\}$, and the other that assigns weights to arcs in a directed graph such that the cut function already gives such a sequence of quadratic (in the number of vertices) number of sets.

### 4.1 Interval Submodular Functions

In this section, we show that the bound for the submodular case is tight by constructing a sequence of $\binom{n+1}{2} + 1$ sets $\emptyset, S_1, \cdots, S_{\binom{n+1}{2}}$ for a nonnegative submodular function $f$, such that $f(S_i) = 4f(S_{i-1})$ for all $i \leq \binom{n+1}{2}$.

For each $1 \leq i \leq j \leq n$, consider intervals $[i,j] = \{k \mid i \leq k \leq j\}$ and let the set of all intervals be $\mathcal{I} = \bigcup_{i,j}\{[i,j]\}$. Let $[i,j] = \emptyset$ whenever $i > j$. Consider a set function $f : \mathcal{I} \to \mathbb{R}_+$ such that $f(\emptyset) = 0$. We say $f$ is *submodular on intervals* if for any $S, T \in \mathcal{I}$ such that $S \cup T \in \mathcal{I}$ and $S \cap T \in \mathcal{I}$, we have

$$f(S) + f(T) \geq f(S \cup T) + f(S \cap T).$$

**Lemma 7.** *Let $\tau$ and $\kappa$ be monotonically increasing, nonnegative functions on the set $[n]$, then $f$ defined by $f([i,j]) = \tau(i)\kappa(j)$ is submodular on intervals.*

*Proof.* Consider two intervals $S$ and $T$. The statement follows trivially if $S \subseteq T$, so consider this is not the case. Let $S = [s_i, s_j]$ and $T = [t_i, t_j]$ and assume w.l.o.g that $s_j \geq t_j$.

i. Case $S \cap T \neq \emptyset$. This implies $t_i < s_i$ and $s_i \leq t_j \leq s_j$. In this case, $f(S) + f(T) - f(S \cap T) - f(S \cup T) = \tau(s_i)\kappa(s_j) + \tau(t_i)\kappa(t_j) - \tau(s_i)\kappa(t_j) - \tau(t_i)\kappa(s_j) = (\tau(s_i) - \tau(t_i))(\kappa(s_j) - \kappa(t_j)) \geq 0$.
ii. Case $S \cap T = \emptyset, S \cup T = [t_i, s_j]$. In this case, $f(S) + f(T) - f(S \cup T) = \tau(s_i)\kappa(s_j) + \tau(t_i)\kappa(t_j) - \tau(t_i)\kappa(s_j) \geq \kappa(s_j)(\tau(s_i) - \tau(t_i)) \geq 0$. □

We show that one can *extend* any function that is submodular on intervals to a submodular function (defined over the ground set). This construction is general, and might be of independent interest. For any set $S \subseteq V$, define $\mathcal{I}(S)$ to be the set of maximum intervals contained in $S$. For example, for $S = \{1, 2, 3, 6, 9, 10\}$, $\mathcal{I}(S) = \{[1,3], [6,6], [9,10]\}$.

**Lemma 8.** *Consider a set function $f$ defined over intervals such that (i) $f(\emptyset) = 0$, (ii) $f([i,j]) \geq 0$ for interval $[i,j]$, (iii) for any $S, T \in \mathcal{I}$ such that $S \cap T, S \cup T \in \mathcal{I}$, $f(S) + f(T) \geq f(S \cup T) + f(S \cap T)$. Then, $g(S) = \sum_{I \in \mathcal{I}(S)} f(I)$ is submodular over the ground set $\{1, \ldots, n\}$.*

*Proof.* We will show that $g$ is submodular by proving that for any $T \subseteq S$ and any $k \notin S$, $g(S \cup \{k\}) - g(S) \leq g(T \cup \{k\}) - g(T)$. Let the marginal gain obtained by adding $k$ to S be $g_k(S) = g(S \cup \{k\}) - g(S)$.

Note that $\mathcal{I}(S \cup k) \setminus \mathcal{I}(S)$ can either contain (i) $[s,k]$, for some $s \leq k$, or (ii) $[k,u]$, for some $u > k$, or (iii) $[s,u]$ for $s \leq k \leq u$. In case (i), $g_k(S) = f([s,k]) - f([s,k-1])$; in case (ii), $g_k(S) = f([k,u]) - f([k+1,u])$; and in case (iii), $g_k(S) = f([s,u]) - f([s,k-1]) - f([k+1,u])$. Thus, when comparing the values of $g_k(S)$ and $g_k(T)$, we are only concerned with intervals that are modified due to the addition of $k$.

Let $S \cup \{k\}$ contain the interval $[s,k-1] \cup \{k\} \cup [k+1,u]$ and $T \cup \{k\}$ contain the interval $[t,k-1] \cup \{k\} \cup [k+1,v]$ where $s \leq t$, $v \leq u$ (as $T \subseteq S$) and $s \leq k \leq u$ ($s = k$ implies $[s,k-1] = \emptyset$ and $u = k$ implies that $[k+1,u] = \emptyset$) and $t \leq k \leq v$ ($t = k$ implies $[t,k-1] = \emptyset$ and $v = k$ implies that $[k+1,v] = \emptyset$).

$$\begin{aligned}
&g(S \cup \{k\}) - g(S) - g(T \cup \{k\}) + g(T) \\
&= f([s,u]) - f([s,k-1]) - f([k+1,u]) - (f([t,v]) - f([t,k-1]) - f([k+1,v])) \\
&= f([s,u]) - f([s,k-1]) - f([k+1,u]) - f([t,v]) + f([t,k-1]) + f([k+1,v]) \\
&\leq f([s,u]) - f([s,k-1]) - f([k+1,v]) - f([t,u]) + f([t,k-1]) + f([k+1,v]) \qquad (6) \\
&= f([s,u]) - f([s,k-1]) - f([t,u]) + f([t,k-1]) \leq 0. \qquad (7)
\end{aligned}$$

where (6) follows from submodularity of $f$ on intervals $[k+1,u]$ and $[t,v]$, i.e., $f([k+1,u]) + f([t,v]) \geq f([t,u]) + f([k+1,v])$, and (7) follows from submodularity of $f$ on intervals $[s,k-1]$ and $[t,u]$. □

**Construction.** Consider the function $f([i,j]) = 4^{\frac{j(j-1)}{2}} 4^i$ for $[i,j] \in \mathcal{I}$, obtained by setting $\tau(i) = 4^i$ and $\kappa(j) = 4^{\frac{j(j-1)}{2}}$. This is submodular on intervals from Lemma 7. This function defined on intervals can be extended to a submodular function $g$ by Lemma 8. Consider the total order $\prec$ defined on intervals $[i,j]$ specified in example 1 (Section 2). By our choice of $\tau$ and $\kappa$ we have that $S \prec T$ implies $4g(S) \leq g(T)$. The submodular function $g$ thus contains a sequence of length $\binom{n+1}{2} + 1$ of sets that increase geometrically in their function values.

### 4.2 Cut functions

The example from the previous section and Birkhoff's representation theorem motivates a construction of a complete directed graph $G = (V, A)$ ($|V| = n$) and a weight vector $w \in \mathbb{R}_+^{|A|}$ such that there exists a sequence of $m = \binom{n}{2}$ sets $\emptyset, S_1, \cdots, S_m \subseteq V$ that has $w(\delta^+(S_k)) \geq 4w(\delta^+(S_{k-1}))$ for all $k \geq 2$.

**Construction.** The sets $S_i$ are all intervals of $[n-1]$, and are ordered by the complete order $\prec$ as defined previously. One can verify that the $k$th set $S_k$ in the sequence is $S_k = [i, j]$ where $k = i + j(j-1)/2$.

Note that, if $i > 1$, for each interval $[i, j]$, arc $e_{i,j} := (j, i-1) \in \delta^+([i, j])$ and $(j, i-1) \notin \delta^+([s, t])$ for any $(s, t) \prec (i, j)$. For any interval $[1, j]$, arc $e_{1,j} := (j, j+1) \in \delta^+([1, j])$ and $(j, j+1) \notin \delta^+([s, t])$ for any $(s, t) \prec (1, j)$. Define arc weights $w$ by $w(e_{i,j}) = 5^{i+j(j-1)/2}$. Thus, the arcs $e_{i,j}$ corresponding to the intervals $[i, j]$ increase in weight by a factor of 5. We claim that $w(\delta^+(S_k)) \geq 4w(\delta^+(S_{k-1}))$. This is true because $4 \sum_{e_{s,t}:(s,t)\prec(i,j)} w(e_{s,t}) \leq w(e_{i,j})$.

## 5 Open Question

In this paper, we showed an $O(n^2)$ bound on the number of iterations of the discrete Newton's algorithm for the problem of finding $\max \delta : \min_S f(S) - \delta a(S) \geq 0$ for an arbitrary direction $a \in \mathbb{R}^n$. Even though we showed that certain parts of our analysis were tight, we do not know whether this bound is tight. More fundamentally, we know little about the number of breakpoints of the piecewise linear function $g(\delta) = \min_S f(S) - \delta a(S)$ in the case of an arbitrary direction $a$. Our results do not imply anything on this number of breakpoints, and this number could still be quadratic, exponential or even linear. In the simpler, nonnegative setting $a \in \mathbb{R}_+^n$, it is not just that the discrete Newton's algorithm takes at most $n$ iterations, but it is also the case that the number of breakpoints of the lower envelope is at most $n$ (by the property of strong quotients). On the other hand, there exist instances of parametric minimum $s-t$ cut problems where the minimum cut value has an exponential number of breakpoints [Mulmuley, 1999]. However, this corresponds to the more general problem $\min_S f(S) - \delta a(S)$ where $f(\cdot)$ is submodular but the function $a(\cdot)$ is not modular (and not even supermodular or submodular as the slopes of the parametric capacities can be positive or negative).

## References


[Freund et al., 2015] Freund, R. M., Grigas, P., and Mazumder, R. (2015). An extended Frank-Wolfe method with "In-Face" directions, and its application to low-rank matrix completion. *arXiv preprint arXiv:1511.02204*.

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

[Iwata, 2008] Iwata, S. (2008). Submodular function minimization. *Mathematical Programming*, 112(1):45–64.

[Iwata et al., 1997] Iwata, S., Murota, K., and Shigeno, M. (1997). A fast parametric submodular intersection algorithm for strong map sequences. *Mathematics of Operations Research*, 22(4):803–813.

[Iwata and Orlin, 2009] Iwata, S. and Orlin, J. B. (2009). A simple combinatorial algorithm for submodular function minimization. In *Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1230–1237. Society for Industrial and Applied Mathematics.

[Lee et al., 2015] Lee, Y. T., Sidford, A., and Wong, S. C. (2015). A faster cutting plane method and its implications for combinatorial and convex optimization. In *Foundations of Computer Science (FOCS)*, pages 1049–1065. IEEE.

[McCormick and Ervolina, 1994] McCormick, S. T. and Ervolina, T. R. (1994). Computing maximum mean cuts. *Discrete Applied Mathematics*, 52(1):53–70.

[Mulmuley, 1999] Mulmuley, K. (1999). Lower bounds in a parallel model without bit operations. *SIAM Journal on Computing*, 28(4):1460–1509.

[Nagano, 2007] Nagano, K. (2007). A strongly polynomial algorithm for line search in submodular polyhedra. *Discrete Optimization*, 4(3):349–359.

[Radzik, 1998] Radzik, T. (1998). Fractional combinatorial optimization. In *Handbook of Combinatorial Optimization*, pages 429–478. Springer.

[Topkis, 1978] Topkis, D. M. (1978). Minimizing a submodular function on a lattice. *Operations Research*, 26(2):305–321.