Submodular Optimization with Contention Resolution Extensions

Benjamin Moseley

Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA, USA
Relational AI, Berkeley CA, USA
moseleyb@andrew.cmu.edu

Maxim Sviridenko

Yahoo Research, New York, NY, USA
sviri@oath.com

Abstract

This paper considers optimizing a submodular function subject to a set of downward closed constraints. Previous literature on this problem has often constructed solutions by (1) discovering a fractional solution to the multi-linear extension and (2) rounding this solution to an integral solution via a contention resolution scheme. This line of research has improved results by either optimizing (1) or (2).

Diverging from previous work, this paper introduces a principled method called contention resolution extensions of submodular functions. A contention resolution extension combines the contention resolution scheme into a continuous extension of a discrete submodular function. The contention resolution extension can be defined from effectively any contention resolution scheme. In the case where there is a loss in both (1) and (2), by optimizing them together, the losses can be combined resulting in an overall improvement. This paper showcases the concept by demonstrating that for the problem of optimizing a non-monotone submodular subject to the elements forming an independent set in an interval graph, the algorithm gives a .188-approximation. This improves upon the best known $\frac{1}{2e} \simeq .1839$ approximation.

2012 ACM Subject Classification Mathematics of computing → Combinatorial algorithms

Keywords and phrases Submodular, Optimization, Approximation Algorithm, Interval Scheduling

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2019.3

Category APPROX

Funding *Benjamin Moseley*: Supported in part by a Google Research Award, a Yahoo Research Award and NSF Grants CCF-1830711, CCF-1824303, and CCF-1733873.

1 Introduction

Submodular function maximization has numerous applications and there has been a rich theory developed on the topic. See [9] for pointers to relevant work. In this problem, the input consists of a universe of n elements U and a submodular set function $f : 2^U \to \mathbb{R}^+$. A function is submodular if for all sets $A, B \subseteq U$ where $A \subseteq B$ and any element $e \in U \setminus B$ it is the case that $f(A \cup \{e\}) - f(A) \geq f(B \cup \{e\}) + f(B)$.[1] Submodular functions are a general class of functions that capture the concept of diminishing returns. Natural occurrences of submodular functions include the cut function [8] and the coverage function [3]. Due to their generality, submodular functions capture many common objective functions. For example, submodular functions are frequently used in machine learning for problems such as document summarization [18], exemplar clustering [12], influence in social networks [13] and other problems [15].

[1] Equivalently, a function is submodular if for all sets $A, B \subseteq U$ it is the case that $f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$.

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019).
Editors: Dimitris Achlioptas and László A. Végh; Article No. 3; pp. 3:1–3:17
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

The submodular maximization problem is to select a set S maximizing $f(S)$ such that $S \in \mathcal{I}$ where $\mathcal{I}$ is a family of sets of feasible solutions. The set $\mathcal{I}$ is usually assumed to be downward closed.[2] The set of feasible solutions $\mathcal{I}$ is defined based on the constraints of the given problem. Prior work has focused on two cases. In one, the function f is additionally assumed to be *monotone* and in the other the function f is *non-monotone*. A submodular function is monotone if $f(S \cup \{e\}) \geq f(S)$ for all $S \subseteq U$ and $e \in U$. The function f is said to be non-monotone if there is no monotonicity restriction.

Optimizing a submodular function subject to classes of downward closed constraints has been extensively studied [7, 11, 4, 1, 10]. The most widely considered classes of constraints are a cardinality constraint [3], matroid constraints [17], knapsack constraints [16], and interval constraints [9]. Through this line of research, a general algorithmic method has emerged. The method consists of two parts. (1) Find a fractional solution to the multilinear extension, and then (2) use a contention resolution scheme or techniques like pipage rounding [5] to round the fractional solution to a feasible integral solution. The *multilinear extension* is an extension of a discrete submodular set function f to the fractional continuous setting. This algorithmic method is general enough to give strong results for numerous problems, including the best known results for monotone and non-monotone submodular maximization under a single matroid constraint [3, 7, 1].

Several past works have focused on optimizing either steps (1) or (2) to improve state-of-the-art methods. Generally, past work has focused on improving (1), the procedure to construct a fractional solution. This is because [6] gave general methods for converting fractional solutions to the multilinear extension to an integral solution. The algorithm typically used in (1) is the *continuous greedy* algorithm and its variants [2, 4, 19, 11, 7].

The Multi-Linear Extension, Continuous Greedy, and Contention Resolution Schemes. Let F be the multilinear extension of f. The **multilinear extension** F is a continuos function that extends f to the fractional domain $[0,1]^{|U|}$. The input to F is a vector $\mathbf{x}$ where $0 \leq x_i \leq 1$ for all i. Let S contain each element U with probability x_i. The value of $F(\mathbf{x})$ is $\mathbb{E}[f(S)]$. It is important to note that S may not be in $\mathcal{I}$. Past work uses the continuous greedy framework to discover a vector $\mathbf{x}$ such that $F(\mathbf{x})$ is close to the optimal solution. Then, this is rounded to an integral solution using a contention resolution scheme $\mathcal{C}(\mathbf{x})$. The idea is to first construct the set S at random, as is done in the computation of $F(\mathbf{x})$. Then some elements from S are dropped to find a set $S' \subseteq S$ such that $S' \in \mathcal{I}$. Key is showing that $\mathbb{E}[f(S')]$ is close to $F(\mathbf{x})$, and thereby bounding $\mathbb{E}[f(S')]$ by the optimal solution.

The **continuous greedy algorithm** iteratively builds a fractional solution $\mathbf{x}$. The algorithm adds a small fractional amount $\mathbf{x}^*$ of some elements to $\mathbf{x}$ such that it greedily increases $F(\mathbf{x} + \mathbf{x}^*)$. Past work has focused on the optimizing the greedy choice of $\mathbf{x}^*$.

This line of work has mostly focused on optimizing (1). This is due to (1) is being the core part of the algorithm where there is loss in the approximation factor. In many cases though, there is additionally loss when performing (2) as well [6, 9].

Contention Resolution Extensions. As mentioned, past work has focused on optimizing (1) and (2) in isolation. This paper for the first time considers optimizing (1) and (2) together to combine the losses in the two procedures and show overall improved results. Our main results are enabled by a principled algorithmic method called *contention resolution extensions*, going beyond optimizing the multi-linear extension.

---

[2] A set $\mathcal{I}$ is said to be downward closed if $S \in \mathcal{I}$ implies $S' \in \mathcal{I}$ for all $S' \subseteq S$.

The framework takes as input a (randomized) contention resolution scheme $\mathcal{C}(\mathbf{x})$. The contention resolution scheme takes as input a fractional solution and returns a feasible integral solution. Past work constructs $\mathbf{x}$ and then produces the final solution using $\mathcal{C}$ only in the last step. Instead, this paper uses $\mathcal{C}$ to construct $\mathbf{x}$. At each step the new method greedily selects a small fractional amount of each element $\mathbf{x}^*$ to maximize the expected value of $\mathcal{C}(\mathbf{x}+\mathbf{x}^*)$. When the algorithm terminates, it simply returns $\mathcal{C}(\mathbf{x})$ for the final vector $\mathbf{x}$ computed. In this way, the algorithm's greedy choices at each step are closely connected to the final solution that the algorithm will return.

Improved results can be shown using this framework because the loss in step (1) and (2) can be combined in the analysis. Further, the loss in the contention resolution scheme is optimized over in each step, allowing the algorithm to converge to a fractional solution that is chosen directly to optimize the final solution.

1.1 Applications of the Contention Resolution Extension Framework

This paper shows how contention resolution extensions can be used to improve state-of-the-art results for optimizing submodular functions.

The paper considers the problem of optimizing a submodular function over independent sets in an interval graph. In this problem, each element is associated with an interval. The goal is to select a set of intervals that do not intersect to maximize a non-monotone submodular function. The best known previous result is a $\frac{1}{2e} \simeq .1839$-approximation [9].

▶ **Theorem 1.** *For any non-monotone submodular function where $f(\emptyset) = 0$ there is a $.188$-approximation algorithm for maximizing the function subject to an interval constraint.*

Overview of the Improved Analysis. To describe how our analysis improves over previous work, first consider the unified continuous greedy algorithm of [11]. Let $\mathcal{C}$ be a contention resolution scheme and OPT denote the value of the optimal solution. As discussed, the algorithm greedily builds a fractional solution $\mathbf{x}$. At each step, an amount $\mathbf{x}^*$ is added to $\mathbf{x}$ where $\mathbf{x}^*$ contains a small amount of some of the elements. Past analysis of the continuous greedy framework proves that in each step $F(\mathbf{x})$ increases by an amount proportional to $(1 - ||\mathbf{x}||_\infty)\mathrm{OPT}$. That is, the incremental improvement of $F(\mathbf{x})$ at each step is proportional to OPT multiplied by an amount that depends on the most any element is fractionally selected in $\mathbf{x}$. The analysis crucially relies on a bound on $||\mathbf{x}||_\infty$ at each step. The algorithm arrives at the final solution using $\mathcal{C}$ on the vector $\mathbf{x}$ at the end of the continuous greedy procedure. For many contention resolution schemes, the expected value of the solution returned is bounded by $F(\mathbf{x})$ multiplied by the minimum probability an element is not discarded by the contention resolution scheme.

Following the above, notice that improving the bound on $||\mathbf{x}||_\infty$ in each step will improve the overall analysis. Our algorithmic framework will allow us to achieve better bounds on $||\mathbf{x}||_\infty$. In particular, we know that the final solution returned is obtained by running $\mathcal{C}$, which increases the probability that an element is not included in the final solution. If somehow the probability an element is discarded by $\mathcal{C}$ could be incorporated into each step of the algorithm to ensure $||\mathbf{x}||_\infty$ is small, then this would improve the overall analysis.

Our algorithm uses $\mathcal{C}$ at each step in the continuous process of constructing $\mathbf{x}$. In particular, by using $\mathcal{C}$ there is less of a chance an element is selected. For this reason, the analysis effectively gets a tighter bound on $||\mathbf{x}||_\infty$, resulting in an overall improved analysis.

A challenge in this approach is that no prior analysis has considered optimizing $\mathcal{C}(\mathbf{x})$ and have always used $F(\mathbf{x})$. Consequently, our analysis introduces new techniques for optimizing over contention resolutions extensions.

2 Preliminaries

Let f be a non-monotone submodular function. The input to the problem is a universe of n elements S. The goal is to select a set of elements $S' \subseteq \mathcal{I}$ such that $f(S')$ is maximized where $\mathcal{I}$ is a set of feasible solution sets. Let $f_R(S') := f(R \cup S') - f(R)$ be the value of adding elements in the set S' to the set R. In this paper it is assumed that $f(\emptyset) = 0$.

The paper considers a hereditary set system defined by independent sets in interval graphs. In this problem, each element $i \in U$ is an interval $(s_i, d_i]$. A set S' is in $\mathcal{I}$ if no two intervals in S' intersect.

The analysis framework in this paper builds on previous submodular optimization work. The next lemma follows from the contention resolution framework of [6]. It is not proven explicitly, but follows from the proof in the paper. Consider a contention resolution scheme that takes as input a set S' and returns a set $D(S') \subseteq S'$. The scheme is said to be monotonic if the probability an element $i \in D(S'')$ is only greater than the probability $i \in D(S')$ for $S'' \subseteq S'$ and $\{i\} \in S''$.

▶ **Theorem 2** ([6]). *Let S' be a set constructed using a randomized procedure. Consider a deterministic monotonic contention resolution scheme that given a set S' of elements constructs a set $D(S') \subseteq S'$ such that $\mathbf{Pr}[i \in D(S') \mid i \in S'] \geq c$ for all S' and i. Further, there exists an ordering of elements $e_1, e_2, \ldots$ in $D(S')$ such that $f_{e_1,e_2,\ldots e_i}(\{e_{i+1}\}) > 0$ for all $0 \leq i < |D(S')|$. Then it is the case that $c\mathbb{E}[f(S')] \leq \mathbb{E}[f(D(S'))]$.*

The following lemma is implied by a well known relationship between the Lovasz extension and multilinear extension of submodular functions. See [9] and [20]. We prove this here for completeness.

▶ **Theorem 3.** *Let f be a non-negative submodular function with $f(\emptyset) = 0$. Fix any set O. Let R be a set of elements constructed at random where element i is in R with probability p_i. Say that $p_i \leq \alpha$ for all $i \notin O$. It is the case that $\mathbb{E}[f(R \cup O)] \geq (1 - \alpha) f(O)$.*

Proof. Let p_i be the probability that i is in R for $i \notin O$ and let $p_i = 1$ for $i \in O$. Consider ordering all of the intervals so that $p_1 \geq p_2 \geq \ldots \geq p_n$. For notational convienience, assume $p_{n+1} = 0$. Recall that for any sets S' and S'' we set $f_{S'}(S'') = f(S' \cup S'') - f(S')$. In the following $[k]$ is the set $\{1, 2, \ldots, k\}$. Let $R' = R \cup O$ in the following. We see the following.

$$
\begin{aligned}
& \mathbb{E}[f(R')] = f(\emptyset) + \sum_{k=1}^{n} \mathbb{E}[f(R' \cap [k]) - f(R' \cap [k-1])] \\
= \; & \sum_{k=1}^{n} \mathbb{E}[f_{R' \cap [k-1]}(R' \cap \{k\})] \geq \sum_{k=1}^{n} \mathbb{E}[f_{[k-1]}(R' \cap \{k\})] \quad [f(\emptyset) = 0 \text{ and submodularity}] \\
= \; & \sum_{k=1}^{n} p_k f_{[k-1]}(k) = \sum_{k=1}^{n} p_k (f([k]) - f([k-1])) = \sum_{k=1}^{n} (p_k - p_{k+1}) f([k]) \\
\geq \; & (1 - \alpha) f(O) \quad [f \text{ is positive and } p_i \leq (1 - \alpha) \text{ for all } i \notin O \text{ by assumption}]
\end{aligned}
$$
◀

3 Non-Monotone Function Subject to an Interval Constraint

In this section, we consider the problem of optimizing a non-monotone submodular function f subject to an interval scheduling constraint. In this problem, there is a set S of possible intervals $(s_i, d_i]$. We note that the intervals do not contain their starting point. This is simply for notational purposes and is without loss of generality. A set S' of intervals is feasible (in $\mathcal{I}$) if no two intersect and the goal is to maximize $f(S')$. It is said that two intervals intersect if they both include a common point.

The algorithm maintains a vector y of size n. Let y^i denote the ith entry in the vector. Intuitively, one can think of the entry y^i as the probability of selecting interval i. The vector y will be chosen such that the following holds. Fix any point t. It is the case that $\sum_{i:t\in(s_i,d_i]} y^i \leq 1$. That is, the total weight of intervals intersecting point t is at most one.

The Function $F(y)$. The function F is defined as follows. A set R of intervals is selected by choosing each interval i with probability y^i. The function $F(y) = \mathbb{E}[f(R)]$. This function is the multi-linear extension. Notice that R may not be in the set of feasible solutions $\mathcal{I}$.

The Function $G(y)$. The function G is constructed similarly to F, but it removes additional intervals from R to get a set $D(R)$. The value of $G(y)$ is set to $\mathbb{E}[f(D(R))]$. Intervals are removed from R so that $D(R)$ forms a feasible solution. In this way, G acts as a contention resolution scheme. Each interval i in R is added to $D(R)$ if there is no other interval in R that intersects the start point s_i of i. This function is a *contention resolution extension*[3] of the set function $f(S)$. Notice that the set $D(R)$ is a feasible solution.

Formally, each interval i is in R with probability y^i. Given R let $D(R) = \{i \in R \mid \forall j \in R, s_i \notin (s_j, d_j]\}$. Set $G(y) = \mathbb{E}[f(D(R))]$.

The Algorithm. The algorithm works as follows. The algorithm continuously optimizes G. At time t a vector y_t has been constructed. Let δ be very small, $\frac{1}{\mathtt{poly}(n)}$. The algorithm initializes $y_{t+\delta}$ to y_t and then increases some of the entires. Pseudocode can be found in Algorithm 1. In the following description, for any vector v let $v + 1_i$ denote the vector v except that the coordinate of i is fixed to 1.

Separately for each element i, the algorithm finds the value of $\gamma_i = \sum_{S'\subseteq S} \mathbf{Pr}[R = S']f(D(S' \cup \{i\}))$, equivalently the value of $G(y_t + 1_i)$. This can be estimated to high accuracy following sampling techniques used in previous work [6, 9, 7] and for ease of explanation we assume that it can be computed exactly. Let $\beta_i := \delta e^{-y_t^i}(1 - y_t^i)$ and $w_i = \beta_i(\gamma_i - G(y_t)) = \beta_i(G(y_t + 1_i) - G(y_t))$. The value of w_i is precisely the change in $G(y_t)$ if y_t^i is increased to 1 and then scaled by β_i.

The algorithm finds a maximum weight independent set I over all intervals where an interval i is given weight w_i. It is well known that such a solution can be found in polynomial time using dynamic programming [14]. For each interval $i \in I$, $y_{t+\delta}^i$ is increased by an additive β_i.

The procedure can stop at any time t where $0 \leq t \leq 1$.[4] When the procedure stops, the final solution is produced by constructing $D(R)$ as in the description of G. This set is returned as the solution. This is a feasible solution by construction and the expected value of the algorithm's solution will be $G(y_t)$.

[3] We note that this is not the only contention resolution extension and there are other natural contention resolution schemes that could be used.

[4] One could stop at $t > 1$ so long as the contention resolution scheme constructs a feasible solution. This did not result in improvement in our analysis.

Algorithm 1 Computing $y_{t+\delta}$ from y_t.

for $i \in U$ **do**
  $\gamma_i \leftarrow G(y_t + 1_i)$
  $\beta_i \leftarrow \delta e^{-y_t^i}(1 - y_t^i)$
  $w_i \leftarrow \beta_i(\gamma_i - G(y_t))$ $// \ = \beta_i(G(y_t + 1_i) - G(y_t))$
end for
Give each interval i a weight of w_i. Using these weights, find a maximum weight subset of intervals I that do not intersect.
for $i \in U$ **do**
  if $i \in I$ **then**
    $y_{t+\delta}^i = y_t^i + \beta_i$
  else
    $y_{t+\delta}^i = y_t^i$
  end if
end for
Output $y_{t+\delta}$

3.1 Analysis

Let O denote the intervals in a fixed optimal solution. For each interval i, let E_i be the set of intervals at or before s_i that intersect i and let i be in E_i. The analysis begins by showing that any single interval is selected with at most a small probability.

▶ **Lemma 4.** *The maximum value an entry in y_t can have is $\alpha(t) := \frac{100(e^{37t/100}-1)}{100e^{37t/100}-63} + 2\delta \leq 1 - e^{-t} + 2\delta t$ for any $0 \leq t \leq 1$.*

Proof. In each step, an interval i chosen to be in I has its probability of selection increased by the algorithm. This increase is at most $\delta(1 - y_t^i)e^{-y_t^i}$ at time t. In the worst case, y_t^i is increased at each time step t. The proof will assume that this is the case for element i. For all $y_t^i \leq 1$, from convexity of $e^{-y_t^i}$ we derive,

$$e^{-y_t^i} \leq 1 - (1 - e^{-1})y_t^i \leq 1 - 0.63y_t^i.$$

We now define a function $\rho^i(t)$ which is a piecewise linear version of y_t^i over times t. Define the function $\rho^i(t)$ for any integer $j \geq 2$ and $t \in [0, 1]$ as follows: for each $t \in [(j-1)\delta, j\delta]$ let $\rho^i(t) = \delta \sum_{\tau=0}^{j-2}(1 - y_{\tau\delta}^i)(1 - 0.63y_{\tau\delta}^i) + (t - (j-1)\delta)(1 - y_{(j-1)\delta}^i)(1 - 0.63y_{(j-1)\delta}^i)$. Set $\rho^i(0) = 0$. Obviously $y_{\tau\delta}^i \leq \rho^i(t)$ when $t \leq \tau\delta$ and $y_{(\tau+1)\delta}^i \leq y_{\tau\delta}^i + \delta$ for all τ. Moreover,

$$\begin{aligned} \frac{d\rho^i(t)}{dt} &= (1 - y_{(j-1)\delta}^i)(1 - 0.63y_{(j-1)\delta}^i) \leq (1 - y_{j\delta}^i + \delta)(1 - 0.63y_{j\delta}^i + \delta) \\ &\leq (1 - \rho^i(t))(1 - 0.63\rho^i(t)) + 4\delta \end{aligned}$$

Consider setting up a new function $\alpha(t)$ where $\alpha(0) = 0$ and $\frac{d\alpha}{dt} = (1-\alpha(t))(1-0.63\alpha(t)) + 4\delta$. Solving this differential equation gives that $\alpha(t) = \frac{100(e^{37t/100}-1)}{100e^{37t/100}-63} + 4\delta t$. We know that $\rho^i(0) = \alpha(0) = 0$. The function $\alpha(t)$ is continuous and the function $\rho(t)$ is piecewise linear. Further, for any $0 \leq t \leq 1$ whenever $\rho^i(t) = \alpha(t)$ the derivative of $\alpha(t)$ is larger than $\rho^i(t)$. This gives that $\rho^i(t) \leq \alpha(t)$ for all $0 \leq t \leq 1$.

Thus, we have that $y_t^i \leq \rho^i(t) \leq \alpha(t)$ for all $0 \leq t \leq 1$, proving the lemma. ◀

We will begin by relating the functions G and F. To do this, we will use Theorem 2. This theorem requires that we bound the probability an interval in R is in $D(R)$. We do this in the following lemma.

▶ **Lemma 5.** *For any time* $0 \leq t \leq 1$ *it is the case that* $\mathbf{Pr}[i \in D(R) \mid i \in R] = \mathbf{Pr}[R \cap (E_i \setminus \{i\}) = \emptyset] \geq e^{-(t-y_t^i)} \geq e^{-t}$.

Proof. Fix an interval $i = (s_i, d_i]$. If this interval is in R, then the only reason it is not in $D(R)$ is because there is another interval $j \in R$ such that j intersects the start point of i. That is if $j \in E_i \cap R$ and $j \neq i$ then in this case i will not be in $D(R)$; otherwise, if $R \cap (E_i \setminus \{i\}) = \emptyset$ then i is in $D(R)$ when $i \in R$. Thus, it suffices to bound the probability any interval is sampled to be in R which intersects s_i. The probability no interval in $E_i \setminus \{i\}$ is sampled is $\prod_{j \neq i, s_i \in (s_j, d_j]} (1 - y_t^j) \geq e^{-t+y_t^i}$. Where the inequality follows from the fact that $\sum_{j: s_i \in (s_j, d_j]} y_t^j \leq t$ for any step of the algorithm, i.e. any time t where $0 \leq t \leq 1$. ◀

Now we show two key lemmas. The first shows a relationship between G and F.

▶ **Lemma 6.** $G(y) \geq \frac{1}{e^t} F(y)$ *for all vectors* y.

Proof. We utilize Theorem 2. First notice that the procedure to construct $D(R)$ in the definition of G is a monotonic scheme. This is because the probability an interval is in $D(R)$ only decreases if intervals are added to R. Lemma 5 and Theorem 2 give the lemma. ◀

The next lemma is the key technical lemma that bounds the increase in the G at each step of the algorithm.

▶ **Lemma 7.** *It is the case that* $G(y_{t+\delta}) \geq (1-\delta)G(y_t) + \frac{\delta}{e^t} \mathbb{E}\left[\sum_{i \in O} \left(f(R \cup \{i\}) - f(R)\right)\right] - O(n^2\delta^2)f(O)$ *for all* $t \leq \ln 2 - \delta$.

We defer the proof of the lemma and first show how this can be used to construct our result. Using the previous two lemma, we can bound the total increase in the function by the optimal solution.

▶ **Lemma 8.** *It is the case that* $G(y_{t+\delta}) \geq (1-\delta)G(y_t) + \frac{\delta}{e^t}\left((1-\alpha(t))\, f(O) - e^t G(y_t)\right) - O(n^2\delta^2)f(O)$ *for all* $t \leq \ln 2 - \delta$.

Proof. Lemma 7 says that $G(y_{t+\delta}) \geq (1-\delta)G(y_t) + \frac{\delta}{e^t}\mathbb{E}[\sum_{i \in O} f(R \cup \{i\}) - f(R)] - O(n^2\delta^2)f(O)$. By definition, $\mathbb{E}[f(R)] = F(y_t)$ and Lemma 6 states that $F(y_t) \leq e^t G(y_t)$. This gives the following. The first inequality follows from submodularity.

$$
\begin{aligned}
& (1-\delta)G(y_t) + \frac{\delta}{e^t}\mathbb{E}[\sum_{i \in O} f(R \cup \{i\}) - f(R)] - O(n^2\delta^2)f(O) \\
\geq \quad & (1-\delta)G(y_t) + \frac{\delta}{e^t}\mathbb{E}[f(R \cup O) - f(R)] - O(n^2\delta^2)f(O) \\
\geq \quad & (1-\delta)G(y_t) + \frac{\delta}{e^t}\left(\mathbb{E}[f(R \cup O)] - e^t G(y_t)\right) - O(n^2\delta^2)f(O). \qquad (1)
\end{aligned}
$$

Notice that $\mathbb{E}[f(R \cup O)] \geq (1-\alpha(t))f(O)$ by Theorem 3 because Lemma 4 gives that the maximum probability any interval is in R is bounded by $\alpha(t)$. Combining this with equation (1) gives the lemma. ◀

Using the two above lemmas, we can show our main result.

Proof of Theorem 1. Lemma 8 states that $G(y_{t+\delta}) \geq (1-\delta)G(y_t) + \frac{\delta}{e^t}((1-\alpha(t))f(O) - e^t G(y_t)) - O(n^2\delta^2)f(O)$ wherever $t \leq \ln 2 - \delta$. This implies that $G(y_{t+\delta}) - G(y_t) \geq -2\delta G(y_t) + \frac{\delta}{e^t}((1-\alpha(t))f(O)) - O(n^2\delta^2)f(O)$ for $t \leq \ln 2 - \delta$.

By choosing δ to be sufficiently small, $G(y_{t+\delta})$ can be bounded using a differential equation. Consider a function $g(t)$ where $g(0) = 0$ and for any $t \in [(j-1)\delta, j\delta]$ it is the case that

$$\begin{aligned} g(t) \quad &= \delta \sum_{\tau=0}^{j-2} \left(-2G(y_{\tau\delta}) + \frac{f(O)}{e^{\tau\delta}}(1-\alpha(\tau\delta)) \right) \\ &+ (t-(j-1)\delta) \left(-2G(y_{(j-1)\delta}) + \frac{f(O)}{e^{(j-1)\delta}}(1-\alpha((j-1)\delta)) \right). \end{aligned}$$

Inductively, notice that $G(y_t) + O(n^2\delta^2 \cdot \frac{t}{\delta})f(O) \geq g(t)$ for any t divisible by δ and t less than $\ln 2 - \delta$. Further, $\frac{dg}{dt} = -2G(y_{(j-1)\delta}) + \frac{f(O)}{e^{(j-1)\delta}}(1-\alpha((j-1)\delta)) \geq -2g(t) + \frac{f(O)}{e^t}(1-\alpha(t)) - 2\delta f(O)$. Consider a new function $h(t)$ where $h(0) = 0$ and $\frac{dh}{dt} = -2h(t) + \frac{f(O)}{e^t}(1-\alpha(t)) - 2\delta f(O)$. Solving this differential equation results in $h(.54) > .188f(O)$[5]. Note that $.54 \leq \ln 2 - \delta$ for sufficiently small δ.

We know that $h(0) = g(0) = 0$. We also know that $h(t)$ is a continuous function and $g(t)$ is piecewise linear. Further, for any $0 \leq t \leq 1$ whenever $h(t) = g(t)$ the derivative of $g(t)$ is only larger than that of $h(t)$. Thus, we have that $h(t) \leq g(t)$ for all t. Knowing that $g(t) \leq G(y_t) + O(n^2\delta^2\frac{t}{\delta})f(O) \leq G(y_t) + O(n^2\delta)f(O)$ for $t \leq \ln n - \delta$, it is the case that $.188f(O) < h(.54) \leq g(.54) \leq G(y_{.54}) + O(n^2\delta)f(O)$, proving the theorem for $\delta \leq \frac{1}{n^3}$. ◀

It only remains to prove Lemma 7. The proof can be found in Section 4.

4 Proof of Lemma 7

For this section, let y be the current solution computed by our algorithm at some fixed stage t. Throughout the section all lemmas and proofs will assume that $t \leq \ln 2 - \delta$, an assumption in the statement of Lemma 7. Let v be a vector equal to $y_{t+\delta} - y_t$. For simplicity, we drop the index t and throughout this section we only focus on stage t and drop the index t in y_t. We want to bound $G(y+v)$. Throughout this section, let I be the intervals in the support of v. These are the elements the algorithm chooses in the independent set and whose variables get increased. Let O be the intervals in the optimal solution.

Let S be the set of all intervals. Let R be the random set of intervals chosen according to y where every interval is sampled independently. Formally, for each interval i draw a number r_i uniformly at random from $[0,1]$ and let i be in R if $r_i < y^i$. Let $\mathcal{E}_i$ denote the event $y^i < r_i \leq y^i + \beta_i$. Intuitively, $\mathcal{E}_i$ is the event that i would not be in R if y^i is used for the sampling, but would have if y^i was increased by β_i. For $i \in I$ this is the event i was chosen in the computation of $G(y+v)$, but not $G(y)$.

For any set S', let $D(S')$ contain the intervals from S' chosen according to the algorithm that is used in G. That is $D(S')$ is constructed from S' by only adding an interval $j \in S'$ to be in $D(S')$ if there is no other interval in S' with earlier start point that also intersects j.

We would like to bound $G(y+v)$ by quantities involving O and $G(y)$. Let $\mathcal{E}(I')$ denote the event that $\mathcal{E}_i$ occurs for all $i \in I'$ and $\mathcal{E}_i$ does not occur for any $i \in I \setminus I'$ and recall that $\mathbf{Pr}[\mathcal{E}_i] = \beta_i = \delta(1-y^i)e^{y^i}$, the amount the algorithm would increase y^i if $i \in I$. It will

[5] This was verified using a differential equation solving software from Mathematica and independently verified using numerical evaluation.

be useful to first bound the probability that $R = S'$ for some S'. To do this, the following lemmas bound the probability of either an interval being in R or $R = S'$ depending on the events $\mathcal{E}_i$. The claim isn't difficult and the proof is deferred to the appendix.

▷ Claim 9. For any $i \in I$ it is the case that $\mathbf{Pr}[i \in R \mid \overline{\mathcal{E}_i}] = \frac{\mathbf{Pr}[i \in R]}{1-\beta_i} \geq \mathbf{Pr}[i \in R]$ and $\mathbf{Pr}[i \notin R \mid \overline{\mathcal{E}_i}] \geq (1-\beta_i)\mathbf{Pr}[i \notin R]$ when $t \leq \ln 2 - \delta$. Further, for any $i \in I$ and any set $S' \subseteq S$ it is the case that $\mathbf{Pr}[S' = R \mid \mathcal{E}(\{i\})] \geq \mathbf{Pr}[R = S' | \mathcal{E}_i] \prod_{j \in I, j \neq i}(1-\beta_j)$ when $t \leq \ln 2 - \delta$.

Intuitively, the next claim relates the probability R would be the same set if intervals are drawn randomly using y or $y + v$.

▷ Claim 10. Fix any set $S' \subseteq S$. It is the case that $\mathbf{Pr}[R = S' \text{ and } \mathcal{E}(\emptyset)] \geq (1 - \sum_{i \in I \setminus S'} \frac{\beta_i}{1-y^i})\mathbf{Pr}[R = S']$.

Proof. Notice that for any $i \in I$, it is the case that $\mathbf{Pr}[i \in R \text{ and } \mathcal{E}(\emptyset)] = \mathbf{Pr}[i \in R]$ and $\mathbf{Pr}[i \notin R \text{ and } \mathcal{E}(\emptyset)] = \mathbf{Pr}[i \notin R] - \beta_i = \mathbf{Pr}[i \notin R](1 - \frac{\beta_i}{1-y^i})$. The last equality follows from $\mathbf{Pr}[i \notin R] = 1 - y^i$ by definition. Knowing that elements are sampled independently, we have the following. The first equality follows since elements are sampled independently. The three terms break up the cases on if an elements is not in I, is in $I \cap S'$ or is in I and not S'.

$$\begin{aligned}
& \mathbf{Pr}[R = S' \text{ and } \mathcal{E}(\emptyset)] \\
= \; & \mathbf{Pr}[R \setminus I = S' \setminus I] \prod_{i \in I \cap S'} \mathbf{Pr}[i \in R \text{ and } \mathcal{E}(\emptyset)] \prod_{i \in I \setminus S'} \mathbf{Pr}[i \notin R \text{ and } \mathcal{E}(\emptyset)] \\
= \; & \mathbf{Pr}[R = S'] \prod_{i \in I \setminus S'} (1 - \frac{\beta_i}{1-y^i}) \geq (1 - \sum_{i \in I \setminus S'} \frac{\beta_i}{1-y^i})\mathbf{Pr}[R = S'].
\end{aligned}$$

The second equality follows from the observation at the beginning of the proof of the lemma. ◁

The next lemma bounds $G(y+v)$ by $G(y)$. Intuitively, the first term says that if $\mathcal{E}_i$ does not occur for any i then $G(y+v)$ is the same as $G(y)$. The second term captures the case for $\mathcal{E}_i$ occurs for exactly one $i \in O$. Finally, the probability that $\mathcal{E}_i$ occurs for more than one i is very small (proportional to δ^2) so this effect is negligible. The proof is deferred to Section 5.

▶ **Lemma 11.** *It is the case that,* $G(y+v) \geq (1 - \sum_{i \in I} \beta_i)G(y) + \sum_{i \in I} \sum_{S' \subseteq S \setminus \{i\}} \beta_i \mathbf{Pr}[R = S' \mid \mathcal{E}_i] f(D(S' \cup \{i\})) - O(n^2\delta^2 f(O))$.

Next it is observed that the choice of the set I allows us to swap the terms in the expression in the previous lemma by the optimal solution O.

▶ **Lemma 12.** $G(y+v) \geq (1 - \sum_{i \in O} \beta_i)G(y) + \sum_{i \in O} \sum_{S' \subseteq S \setminus \{i\}} \beta_i \mathbf{Pr}[R = S' \mid \mathcal{E}_i] f(D(S' \cup \{i\})) - O(n^2\delta^2 f(O))$

Proof. Consider the value of

$$\sum_{i \in I} \beta_i \left(\sum_{S' \subseteq S \setminus \{i\}} \mathbf{Pr}[R = S' \mid \mathcal{E}_i] f(D(S' \cup \{i\})) - G(y) \right).$$

This equals

$$\sum_{i \in I} \beta_i \left(\sum_{S' \subseteq S \setminus \{i\}} \mathbf{Pr}[R \setminus \{i\} = S'] f(D(S' \cup \{i\})) - G(y) \right).$$

This is equal to the following since elements are sampled independently

$$\sum_{i \in I} \beta_i \left(\sum_{S' \subseteq S} \mathbf{Pr}[R = S'] f(D(S' \cup \{i\})) - G(y) \right) = \sum_{i \in I} w_i.$$

By definition, this is only greater than $\sum_{i \in O} w_i$. Reversing the above steps for O and combining with Lemma 11 gives the lemma. ◀

Our remaining goal is to bound part of the expression from the prior lemma,

$$\sum_{S' \subseteq S \setminus \{i\}} \sum_{i \in O} \beta_i \mathbf{Pr}[R = S' \mid \mathcal{E}_i] f(D(S' \cup \{i\})) + \sum_{i \in O} \beta_i G(y).$$

Recall that E_i is the set of intervals starting earlier than i that intersect i and also the interval i itself. The intervals in $E_i \setminus \{i\}$ are the intervals, which if they are sampled to be in R then i will not be in $D(R)$. Let B_i be the set containing intervals that start during interval i and also i. The following fact will be useful for applying submodularity.

▶ **Lemma 13.** *For any set $S' \subseteq S$ consider $\{S' \setminus B_i\}_{i \in O}$, a collection of subsets of S'. It is the case that every interval in S' appears in exactly $|O| - 1$ sets in this collection. Further, each interval in S appears in exactly one set B_i.*

Proof. To show the lemma, it suffices to show that every interval in S appears in exactly one set B_i for some $i \in O$. Indeed, we may assume that the intervals in O span the entire time horizon (adding dummy intervals as needed). Then, an interval $j \in S$ can only be in B_i if j starts during i. Knowing that O cannot have two intervals that overlap, we have the lemma. ◀

The next lemma is a technical lemma. The purpose is to take an expression $f(D(S') \setminus B_i)$ depending on a set S' and B_i for $i \in O$ and bound it by an expression depending on $f(D(S'))$ without B_i inside the function input. The lemma follows from submodularity and the previous lemma.

▶ **Lemma 14.** *Fix any set $S' \subseteq S$. It is the case that $\delta f(D(S')) \geq \sum_{i \in O} \beta_i \left(f(D(S')) - f(D(S') \setminus B_i)\right)$.*

Proof. Consider the term $\sum_{i \in O} \beta_i (f(D(S')) - f(D(S') \setminus B_i))$. We will remove all negative terms as they only makes the expression smaller. Let O' be all i where $f(D(S')) - f(D(S') \setminus B_i) > 0$. The lemma follows if we prove that $f(D(S')) \geq \sum_{i \in O'} (f(D(S')) - f(D(S') \setminus B_i))$ because this implies $\delta f(D(S')) \geq \delta \sum_{i \in O'} (f(D(S')) - f(D(S') \setminus B_i)) \geq \sum_{i \in O'} \beta_i (f(D(S')) - f(D(S') \setminus B_i))$ knowing that $\beta_i \leq \delta$ and all terms are positive.

Now it is established that $f(D(S')) \geq \sum_{i \in O'} (f(D(S')) - f(D(S') \setminus B_i))$, which follows by submodularity. Indeed, let $A_0 = D(S') \setminus \cup_{i \in O'} B_i$. Arbitrarily order the sets $B_1, B_2, \ldots B_{|O'|}$ and let $A_i = A_{i-1} \cup (B_i \cap D(S'))$ for $1 \leq i \leq |O'|$. By submodularity, $\sum_{i \in O'} (f(D(S')) - f(D(S') \setminus B_i)) \leq \sum_{i \in O'} (f(A_i) - f(A_{i-1})) = f(D(S')) - f(A_0) \leq f(D(S'))$. The equality follows from the function being positive and the inequality from submodularity. ◀

Assuming $\mathcal{E}_i$ occurs, the purpose of the following lemma is to separate the cases where at least one interval in E_i is in R and the other where no interval in E_i is in R. Intuitively, if no interval in E_i is in R then i will be in $D(R)$ otherwise i will not. In either case, when $\mathcal{E}_i$ occurs the interval i ensures no interval in B_i is in $D(R)$ and the lemma bounds the cost of removing B_i by applying Lemma 14. The proof is deferred to Section 6.

▶ **Lemma 15.** *It is the case that,*

$$G(y+v) \geq (1-\delta)G(y) + \sum_{S' \subseteq S} \mathbf{Pr}[R = S' \mid \mathcal{E}_i] \sum_{i \in O, S' \cap E_i = \emptyset} \beta_i (f(D(S') \setminus B_i \cup \{i\}) - f(D(S') \setminus B_i)) - O((n\delta)^2 f(O)).$$

Our goal now is to bound the second term in the previous lemma by showing this following. This shows that the second term is at least $\frac{\delta}{e^t}$ multiplied by the expected value of adding each element of O to R individually.

▶ **Lemma 16.** $\sum_{S' \subseteq S} \mathbf{Pr}[R = S' \mid \mathcal{E}_i] \sum_{i \in O, S' \cap E_i = \emptyset} \beta_i (f(D(S') \setminus B_i \cup \{i\}) - f(D(S') \setminus B_i)) \geq \frac{\delta}{e^t} \sum_{S' \subseteq S} \mathbf{Pr}[R = S'] \sum_{i \in O} f_{S'}(i)$

Before we prove the lemma, we show how this can be used to complete the proof of Lemma 7.

Proof of Lemma 7. By combining lemmas 15 and 16 we have the following.

$$\begin{aligned} G(y+v) &\geq (1-\delta)G(y) + \frac{\delta}{e^t} \sum_{S' \subseteq S} \mathbf{Pr}[R = S'] \sum_{i \in O} f_{S'}(i) - O((n\delta)^2 f(O)) \\ &\geq (1-\delta)G(y) + \frac{\delta}{e^t} \mathbb{E}[\sum_{i \in O} f_R(i)] - O((n\delta)^2 f(O)) \end{aligned}$$

This completes the proof. ◀

It only remains to prove Lemma 16.

Proof of Lemma 16. Consider the term $\sum_{S' \subseteq S} \mathbf{Pr}[R = S' \mid \mathcal{E}_i] \sum_{i \in O, S' \cap E_i = \emptyset} \beta_i (f(D(S') \setminus B_i \cup \{i\}) - f(D(S') \setminus B_i))$. Rearranging the summations and using the definition of $f_{S'}(i)$ this is equal to $\sum_{i \in O} \sum_{S' \subseteq S, S' \cap E_i = \emptyset} \mathbf{Pr}[R = S' \mid \mathcal{E}_i] \beta_i f_{D(S') \setminus B_i}(i)$. We know that for any set $S' \subseteq S$ if $S' \cap E_i = \emptyset$ then $S' \subseteq (S \setminus E_i)$. Using this, the term is equal to the following.

$$\begin{aligned} &\sum_{i \in O} \sum_{S' \subseteq (S \setminus E_i)} \beta_i \mathbf{Pr}[R = S' \mid \mathcal{E}_i] f_{D(S') \setminus B_i}(i) \\ = &\sum_{i \in O} \sum_{S' \subseteq (S \setminus E_i)} \beta_i \mathbf{Pr}[S' = R \setminus E_i \text{ and } R \cap E_i = \emptyset \mid \mathcal{E}_i] f_{D(S') \setminus B_i}(i) \end{aligned}$$

To see why the previous equality holds, notice that $R = S'$ if and only if $S' = R \setminus E_i$ and $R \cap E_i = \emptyset$ for $S' \subseteq (S \setminus E_i)$. Now we continue to lower bound this expression.

$$\begin{aligned} &\sum_{i \in O} \sum_{S' \subseteq (S \setminus E_i)} \beta_i \mathbf{Pr}[S' = R \setminus E_i \mid \mathcal{E}_i] \mathbf{Pr}[R \cap E_i = \emptyset \mid \mathcal{E}_i] f_{D(S') \setminus B_i}(i) \\ &\qquad \text{[Definition of } R \text{ implies independence]} \\ = &\sum_{i \in O} \sum_{S' \subseteq (S \setminus E_i)} \beta_i \mathbf{Pr}[S' = R \setminus E_i] \mathbf{Pr}[R \cap (E_i \setminus \{i\}) = \emptyset] f_{D(S') \setminus B_i}(i) \\ &\qquad \text{[Definition of } \mathcal{E}_i \text{]} \\ \geq &\sum_{i \in O} \sum_{S' \subseteq (S \setminus E_i)} \frac{\beta_i}{e^{t-y^i}} \mathbf{Pr}[S' = R \setminus E_i] f_{D(S') \setminus B_i}(i) \quad \text{[Lemma 5]} \\ = &\frac{\delta}{e^t} \sum_{i \in O} \sum_{S' \subseteq (S \setminus E_i)} (1 - y^i) \mathbf{Pr}[R = S' \setminus E_i] f_{D(S') \setminus B_i}(i) \quad \text{[Definition of } \beta_i \text{]} \end{aligned} \tag{2}$$

Notice that $1 = \sum_{E \subseteq E_i} \mathbf{Pr}[R \cap E_i = E \mid i \notin R]$ because the right hand side captures all the events in a probability distribution. Further, fix an element $i \in O$ and notice that for any set $S' \subseteq (S \setminus E_i)$ and any set $E \subseteq E_i$ it is the case that $\mathbf{Pr}[S' = R \setminus E_i] \cdot \mathbf{Pr}[R \cap E_i = E \mid i \notin R] = \mathbf{Pr}[R = S' \cup E \mid i \notin R]$. This follows for two reasons. One is because elements are sampled independently. The other is because $\mathbf{Pr}[S' = R \setminus E_i] = \mathbf{Pr}[S' = R \setminus E_i \mid i \notin R]$ since $i \in E_i$ and the independence of sampling elements. Using these facts, the following holds.

$$
\begin{aligned}
(2) \quad &= \frac{\delta}{e^t} \sum_{i \in O} \sum_{S' \subseteq (S \setminus E_i)} (1 - y^i) \mathbf{Pr}[S' = R \setminus E_i] f_{D(S') \setminus B_i}(i) \sum_{E \subseteq E_i} \mathbf{Pr}[R \cap E_i = E \mid i \notin R] \\
&= \frac{\delta}{e^t} \sum_{i \in O} \sum_{S' \subseteq (S \setminus E_i)} (1 - y^i) f_{D(S') \setminus B_i}(i) \sum_{E \subseteq E_i} \mathbf{Pr}[R = S' \cup E \mid i \notin R] \\
&\quad [\text{Independence}] \\
&\geq \frac{\delta}{e^t} \sum_{i \in O} \sum_{S' \subseteq (S \setminus E_i)} (1 - y^i) f_{S' \setminus B_i}(i) \sum_{E \subseteq E_i} \mathbf{Pr}[R = S' \cup E \mid i \notin R] \\
&\quad [\text{Submodularity}] \\
&\geq \frac{\delta}{e^t} \sum_{i \in O} \sum_{S' \subseteq (S \setminus E_i)} (1 - y^i) \sum_{E \subseteq E_i} f_{(S' \cup E) \setminus \{i\}}(i) \mathbf{Pr}[R = S' \cup E \mid i \notin R] \\
&\quad [\text{Submodularity and } i \in B_i] \\
&= \frac{\delta}{e^t} \sum_{i \in O} \sum_{S' \subseteq (S \setminus \{i\})} (1 - y^i) f_{S' \setminus \{i\}}(i) \mathbf{Pr}[R = S' \mid i \notin R] \\
&= \frac{\delta}{e^t} \sum_{i \in O} \sum_{S' \subseteq (S \setminus \{i\})} f_{S' \setminus \{i\}}(i) \mathbf{Pr}[R = S'] \\
&\quad [(1 - y^i) = \mathbf{Pr}[i \notin R] \text{ and definition of conditional probability}] \\
&= \frac{\delta}{e^t} \sum_{i \in O} \sum_{S' \subseteq S} f_{S'}(i) \mathbf{Pr}[R = S'] \quad [f_{S'}(i) = 0 \text{ if } i \in S']
\end{aligned}
$$

◀

5 Proof of Lemma 11

This section is devoted to proving Lemma 11.

Consider $G(y + v)$. The value of $G(y + v)$ is equal to $\sum_{S' \subseteq S} \sum_{I' \subseteq I} \mathbf{Pr}[R = S'$ and $\mathcal{E}(I')] f(D(S' \cup I'))$. This is equal to the following by breaking this into cases. This is a partitioning of the event space by definition of $\mathcal{E}(I')$.

$$
\begin{aligned}
&\sum_{S' \subseteq S} \mathbf{Pr}[R = S' \text{ and } \mathcal{E}(\emptyset)] f(D(S')) \\
&+ \sum_{i \in I} \sum_{S' \subseteq S} \mathbf{Pr}[R = S' \text{ and } \mathcal{E}(\{i\})] f(D(S' \cup \{i\})) \\
&+ \sum_{I' \subseteq I, |I'| \geq 2} \sum_{S' \subseteq S} \mathbf{Pr}[R = S' \text{ and } \mathcal{E}(I')] f(D(S' \cup \{i\}))
\end{aligned}
$$

Knowing that f is positive, this is greater than the following.

$$
\sum_{S' \subseteq S} \mathbf{Pr}[R = S' \text{ and } \mathcal{E}(\emptyset)] f(D(S')) \tag{3}
$$

$$
+ \sum_{i \in I} \sum_{S' \subseteq S} \mathbf{Pr}[R = S' \text{ and } \mathcal{E}(\{i\})] f(D(S' \cup \{i\})) \tag{4}
$$

The proof bounds these two terms separately. First consider (3). Using Claim 10 this is greater than $\sum_{S' \subseteq S}(1 - \sum_{i \in I \setminus S'} \frac{\beta_i}{1-y^i})\mathbf{Pr}[R = S']f(D(S')) = G(y) - \sum_{i \in I} \frac{\beta_i}{1-y^i} \sum_{S' \subseteq S \setminus \{i\}} \mathbf{Pr}[R = S']f(D(S'))$. The definition of β_i gives that this is equal to $G(y) - \sum_{i \in I} \beta_i \sum_{S' \subseteq S \setminus \{i\}} \mathbf{Pr}[R = S']f(D(S')) - \sum_{i \in I} y^i e^{-y^i} \delta \sum_{S' \subseteq S \setminus \{i\}} \mathbf{Pr}[R = S']f(D(S'))$. We will establish that this is only greater than $(1 - \sum_{i \in I} \beta_i)G(y)$. Consider the last term.

$$\begin{aligned}
& \sum_{i \in I} y^i e^{-y^i} \delta \sum_{S' \subseteq S \setminus \{i\}} \mathbf{Pr}[R = S']f(D(S')) \\
= & \sum_{i \in I} y^i e^{-y^i} \delta \sum_{S' \subseteq S \setminus \{i\}} \mathbf{Pr}[R \setminus \{i\} = S']\mathbf{Pr}[i \notin R]f(D(S')) \\
= & \sum_{i \in I} y^i e^{-y^i} (1 - y^i) \delta \sum_{S' \subseteq S \setminus \{i\}} \mathbf{Pr}[R \setminus \{i\} = S']f(D(S')) \\
= & \sum_{i \in I} y^i \beta_i \sum_{S' \subseteq S \setminus \{i\}} \mathbf{Pr}[R \setminus \{i\} = S']f(D(S'))
\end{aligned}$$

By definition of the algorithm $\frac{w_i}{\beta_i} = \mathbb{E}[f(D(R \cup \{i\})) - f(D(R))] = \sum_{S' \subseteq S} \mathbf{Pr}[R = S'](f(D(S' \cup \{i\})) - f(D(S'))) = \sum_{S' \subseteq S \setminus \{i\}} \mathbf{Pr}[R = S'](f(D(S' \cup \{i\})) - f(D(S'))) > 0$ for all $i \in I$. The last equality follows since a term is 0 if i is in S'. Since elements are sampled independently, this gives that the previous term is only less than the following.

$$\begin{aligned}
& \sum_{i \in I} y^i \beta_i \sum_{S' \subseteq S \setminus \{i\}} \mathbf{Pr}[R \setminus \{i\} = S']f(D(S')) \\
\leq & \sum_{i \in I} y^i \beta_i \sum_{S' \subseteq S \setminus \{i\}} \mathbf{Pr}[R \setminus \{i\} = S']f(D(S' \cup \{i\})) \\
= & \sum_{i \in I} \beta_i \sum_{S' \subseteq S \setminus \{i\}} \mathbf{Pr}[R = S' \cup \{i\}]f(D(S' \cup \{i\})) \quad [\text{Note that } \mathbf{Pr}[i \in R] = y_i]
\end{aligned}$$

Now we use this to bound (3). (3) is greater than or equal to the following.

$$\begin{aligned}
& G(y) - \sum_{i \in I} \beta_i \sum_{S' \subseteq S \setminus \{i\}} \mathbf{Pr}[R = S']f(D(S')) - \sum_{i \in I} y^i e^{-y^i} \delta \sum_{S' \subseteq S \setminus \{i\}} \mathbf{Pr}[R = S']f(D(S')) \\
\geq & G(y) - \sum_{i \in I} \beta_i \sum_{S' \subseteq S \setminus \{i\}} \mathbf{Pr}[R = S']f(D(S')) - \sum_{i \in I} \beta_i \sum_{S' \subseteq S \setminus \{i\}} \mathbf{Pr}[R = S' \cup \{i\}]f(D(S' \cup \{i\})) \\
= & (1 - \sum_{i \in I} \beta_i)G(y)
\end{aligned}$$

It remains to bound (4). Using conditional probability, this can be bounded as follows.

$$\begin{aligned}
& \sum_{i \in I} \mathbf{Pr}[\mathcal{E}(\{i\})] \sum_{S' \subseteq S} \mathbf{Pr}[R = S' | \mathcal{E}(\{i\})]f(D(S' \cup \{i\})) \\
= & \sum_{i \in I} \mathbf{Pr}[\mathcal{E}_i] \prod_{j \in I, j \neq i} \mathbf{Pr}(\overline{\mathcal{E}}_j) \sum_{S' \subseteq S} \mathbf{Pr}[R = S' | \mathcal{E}(\{i\})]f(D(S' \cup \{i\})) \\
= & \sum_{i \in I} \beta_i \prod_{j \in I, j \neq i} (1 - \beta_j) \sum_{S' \subseteq S} \mathbf{Pr}[R = S' | \mathcal{E}(\{i\})]f(D(S' \cup \{i\})) \\
\geq & \prod_{i \in I} (1 - \beta_i) \sum_{i \in I} \beta_i \sum_{S' \subseteq S} \mathbf{Pr}[R = S' | \mathcal{E}(\{i\})]f(D(S' \cup \{i\})) \\
\geq & (1 - \sum_{i \in I} \beta_i) \sum_{i \in I} \beta_i \sum_{S' \subseteq S} \mathbf{Pr}[R = S' | \mathcal{E}(\{i\})]f(D(S' \cup \{i\}))
\end{aligned}$$

Knowing that $\beta_i \leq \delta$, this is greater than $\sum_{i \in I} \beta_i \sum_{S' \subseteq S} \mathbf{Pr}[R = S' | \mathcal{E}(\{i\})] f(D(S' \cup \{i\})) - O(\delta^2 n^2 f(O))$.

Now we know that,

$$\begin{aligned}
& \sum_{i \in I} \beta_i \sum_{S' \subseteq S} \mathbf{Pr}[R = S' | \mathcal{E}(\{i\})] f(D(S' \cup \{i\})) \\
= & \sum_{i \in I} \beta_i \sum_{S' \subseteq S} \mathbf{Pr}[R = S' | \mathcal{E}_i \text{ and } \overline{\mathcal{E}}_j \text{ for } j \neq i)] f(D(S' \cup \{i\})) \quad \text{[Def. of } \mathcal{E}(\{i\})] \\
\geq & \sum_{i \in I} \beta_i \prod_{j \in I, j \neq i} (1 - \beta_j) \sum_{S' \subseteq S} \mathbf{Pr}[R = S' | \mathcal{E}_i)] f(D(S' \cup \{i\})) \quad \text{[Claim 9 and independence]} \\
\geq & \sum_{i \in I} \beta_i \sum_{S' \subseteq S} \mathbf{Pr}[R = S' | \mathcal{E}_i)] f(D(S' \cup \{i\})) - |I| \delta^2 f(O) \\
\geq & \sum_{i \in I} \beta_i \sum_{S' \subseteq S \setminus \{i\}} \mathbf{Pr}[R = S' | \mathcal{E}_i)] f(D(S' \cup \{i\})) - |I| \delta^2 f(O)
\end{aligned}$$

The last line follows since if $\mathcal{E}_i$ occurs then R does not contain i. Putting this all together gives the lemma.

6 Proof of Lemma 15

This section is devoted to proving Lemma 15.

Consider the following expression. Lemma 12 gives the following.

$$G(y + v) \geq \sum_{S' \subseteq S} \sum_{i \in O} \beta_i \mathbf{Pr}[R = S' \mid \mathcal{E}_i] f(D(S' \cup \{i\})) \tag{5}$$

$$+ (1 - \sum_{i \in O} \beta_i) G(y). \tag{6}$$

We see that (5) equals the following.

$$\sum_{i \in O} \beta_i \Big(\sum_{S' \subseteq S, S' \cap E_i = \emptyset} \mathbf{Pr}[R = S' \mid \mathcal{E}_i] f(D(S' \cup \{i\})) + \sum_{S' \subseteq S, S' \cap E_i \neq \emptyset} \mathbf{Pr}[R = S' \mid \mathcal{E}_i] f(D(S' \cup \{i\})) \Big)$$

By definition of G, for any set $S' \subseteq S$ it is the case that $D(S' \cup \{i\})$ includes i only if S' includes no interval in $E_i \setminus \{i\}$. We also know that for any $j \in S'$ it is the case that $j \in D(S' \cup \{i\})$ if and only if $j \in D(S')$ and $j \notin B_i$. Using these two facts, the previous term is equal to the following.

$$\sum_{i \in O} \beta_i \Big(\sum_{S' \subseteq S, S' \cap E_i = \emptyset} \mathbf{Pr}[R = S' \mid \mathcal{E}_i] f(D(S') \setminus B_i \cup \{i\})) + \sum_{S' \subseteq S, S' \cap E_i \neq \emptyset} \mathbf{Pr}[R = S' \mid \mathcal{E}_i] f(D(S') \setminus B_i) \Big)$$

This is equal to the following.

$$\sum_{i \in O} \beta_i \left(\sum_{S' \subseteq S, S' \cap E_i = \emptyset} \mathbf{Pr}[R = S' \mid \mathcal{E}_i] \Big(f(D(S') \setminus B_i \cup \{i\})) - f(D(S') \setminus B_i) \Big) + \sum_{S' \subseteq S} \mathbf{Pr}[R = S' \mid \mathcal{E}_i] f(D(S') \setminus B_i) \right)$$

We focus on bounding $\sum_{i \in O} \beta_i \sum_{S' \subseteq S} \mathbf{Pr}[R = S' \mid \mathcal{E}_i] f(D(S') \setminus B_i)$ along with (6). The rest of the expression is carried to the end of the proof. First we establish a bound on $\sum_{i \in O} \beta_i \sum_{S' \subseteq S} \mathbf{Pr}[R = S' \mid \mathcal{E}_i] f(D(S') \setminus B_i)$ in the following claim and then it is combined with (6). The purpose of the following claim is to remove the conditioning on $\mathcal{E}_i$.

▷ Claim 17. $\sum_{i \in O} \beta_i \sum_{S' \subseteq S} \mathbf{Pr}[R = S' \mid \mathcal{E}_i] f(D(S') \setminus B_i) = \sum_{i \in O} \beta_i \sum_{S' \subseteq S} \mathbf{Pr}[R = S'] f(D(S') \setminus B_i)$.

Proof. First note that $\mathbf{Pr}[R = S' \mid \mathcal{E}_i] > 0$ if and only if $i \notin S'$. Thus we have that the left hand side is equal to $\sum_{i \in O} \beta_i \sum_{S' \subseteq S \setminus \{i\}} \mathbf{Pr}[R = S' \mid \mathcal{E}_i] f(D(S') \setminus B_i)$. Using the definition of conditional probability and the definition of $\mathcal{E}_i$ this is equal to $\sum_{i \in O} \beta_i \sum_{S' \subseteq S \setminus \{i\}} \frac{\mathbf{Pr}[R \setminus \{i\} = S' \text{ and } \mathcal{E}_i]}{\mathbf{Pr}[\mathcal{E}_i]} f(D(S') \setminus B_i)$. By independence, this equals $\sum_{i \in O} \beta_i \sum_{S' \subseteq S \setminus \{i\}} \mathbf{Pr}[R \setminus \{i\} = S'] f(D(S') \setminus B_i) = \sum_{i \in O} \beta_i \sum_{S' \subseteq S \setminus \{i\}} \mathbf{Pr}[R \setminus \{i\} = S'] f(D(S') \setminus B_i)(\mathbf{Pr}[i \in R] + \mathbf{Pr}[i \notin R]) = \sum_{i \in O} \beta_i \sum_{S' \subseteq S \setminus \{i\}} (\mathbf{Pr}[R = S'] + \mathbf{Pr}(R = S' \cup \{i\})) f(D(S') \setminus B_i)$.

We know that for any $S' \subseteq S \setminus \{i\}$ it is the case that $f(D(S') \setminus B_i) = f(D(S' \cup \{i\}) \setminus B_i)$ because B_i is the set of intervals that are not in the contention resolution scheme if i is input and also i is in B_i. Using this, we have that the previous expression is equal to $\sum_{i \in O} \beta_i \sum_{S' \subseteq S \setminus \{i\}} (\mathbf{Pr}[R = S'] f(D(S') \setminus B_i) + \mathbf{Pr}(R = S \cup \{i\}) f(D(S' \cup \{i\}) \setminus B_i)) = \sum_{i \in O} \beta_i \sum_{S' \subseteq S} \mathbf{Pr}[R = S'] f(D(S') \setminus B_i)$. ◁

Going back to $\sum_{i \in O} \beta_i \sum_{S' \subseteq S} \mathbf{Pr}[R = S' \mid \mathcal{E}_i] f(D(S') \setminus B_i)$ with the expansion of (6) using the definition of G and the previous claim, we have the following.

$$\sum_{i \in O} \beta_i \sum_{S' \subseteq S} \mathbf{Pr}[R = S'] f(D(S') \setminus B_i) + (1 - \sum_{i \in O} \beta_i) \sum_{S' \subseteq S} \mathbf{Pr}[R = S'] f(D(S'))$$

We apply Lemma 14 for each term S' to get that this is greater than the following.

$$(1 - \delta) \sum_{S' \subseteq S} \mathbf{Pr}[R = S'] f(D(S')) = (1 - \delta) G(y)$$

The above gives that $G(y + v) \geq (1 - \delta) G(y) + \sum_{i \in O} \beta_i \sum_{S' \subseteq S, S' \cap E_i = \emptyset} \mathbf{Pr}[R = S' \mid \mathcal{E}_i] \Big(f(D(S') \setminus B_i \cup \{i\}) - f(D(S') \setminus B_i) \Big)$, giving the lemma.

7 Conclusion

This paper introduces the approach of using contention resolution extensions to optimize a submodular function subject to a set of constraints. This algorithmic approach can be used to improve the best known result when the constraints correspond to independent sets in an interval graph. The next direction is to determine if this approach can be used to improve on the best known approximation for other submodular optimization problems.

References

1 Niv Buchbinder and Moran Feldman. Constrained Submodular Maximization via a Non-symmetric Technique. *CoRR*, abs/1611.03253, 2016. arXiv:1611.03253.

2 Gruia Călinescu, Chandra Chekuri, Martin Pál, and Jan Vondrák. Maximizing a Submodular Set Function Subject to a Matroid Constraint (Extended Abstract). In *IPCO*, pages 182–196, 2007.

3 Gruia Călinescu, Chandra Chekuri, Martin Pál, and Jan Vondrák. Maximizing a Monotone Submodular Function Subject to a Matroid Constraint. *SIAM J. Comput.*, 40(6):1740–1766, 2011.

4 Chandra Chekuri, T. S. Jayram, and Jan Vondrák. On Multiplicative Weight Updates for Concave and Submodular Function Maximization. In *ITCS*, pages 201–210, 2015.

5 Chandra Chekuri, Jan Vondrák, and Rico Zenklusen. Dependent Randomized Rounding via Exchange Properties of Combinatorial Structures. In *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010, October 23-26, 2010, Las Vegas, Nevada, USA*, pages 575–584, 2010.

6 Chandra Chekuri, Jan Vondrák, and Rico Zenklusen. Submodular Function Maximization via the Multilinear Relaxation and Contention Resolution Schemes. *SIAM J. Comput.*, 43(6):1831–1879, 2014.

7 Alina Ene and Huy L Nguyen. Constrained Submodular Maximization: Beyond 1/e. In *FOCS*, 2016.

8 Uriel Feige, Vahab S. Mirrokni, and Jan Vondrák. Maximizing Non-monotone Submodular Functions. *SIAM J. Comput.*, 40(4):1133–1153, 2011.

9 Moran Feldman. *Maximization Problems with Submodular Objective Functions*. PhD thesis, Technion - Israel Institute of Technology, 2013.

10 Moran Feldman, Christopher Harshaw, and Amin Karbasi. Greed Is Good: Near-Optimal Submodular Maximization via Greedy Optimization. In *Proceedings of the 30th Conference on Learning Theory, COLT 2017, Amsterdam, The Netherlands, 7-10 July 2017*, pages 758–784, 2017.

11 Moran Feldman, Joseph Naor, and Roy Schwartz. A Unified Continuous Greedy Algorithm for Submodular Maximization. In *FOCS*, pages 570–579, 2011.

12 Ryan Gomes and Andreas Krause. Budgeted Nonparametric Learning from Data Streams. In *ICML*, pages 391–398, 2010.

13 David Kempe, Jon M. Kleinberg, and Éva Tardos. Maximizing the Spread of Influence through a Social Network. *Theory of Computing*, 11:105–147, 2015.

14 Jon Kleinberg and Eva Tardos. *Algorithm Design*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2005.

15 Matt J. Kusner, Wenlin Chen, Quan Zhou, Zhixiang Eddie Xu, Kilian Q. Weinberger, and Yixin Chen. Feature-Cost Sensitive Learning with Submodular Trees of Classifiers. In *AAAI*, pages 1939–1945, 2014.

16 Jon Lee, Vahab S. Mirrokni, Viswanath Nagarajan, and Maxim Sviridenko. Maximizing Nonmonotone Submodular Functions under Matroid or Knapsack Constraints. *SIAM J. Discrete Math.*, 23(4):2053–2078, 2010.

17 Jon Lee, Maxim Sviridenko, and Jan Vondrák. Submodular Maximization over Multiple Matroids via Generalized Exchange Properties. *Math. Oper. Res.*, 35(4):795–806, 2010.

18 Hui Lin and Jeff A. Bilmes. Multi-document Summarization via Budgeted Maximization of Submodular Functions. In *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 2-4, 2010, Los Angeles, California, USA*, pages 912–920, 2010.

19 Jan Vondrák. Optimal approximation for the submodular welfare problem in the value oracle model. In *STOC*, pages 67–74, 2008.

20 Jan Vondrák. Symmetry and Approximability of Submodular Maximization Problems. *SIAM J. Comput.*, 42(1):265–304, 2013.

A Omitted Proofs

Proof of Claim 9. By definition $\mathbf{Pr}[i \in R \mid \overline{\mathcal{E}_i}] = \frac{\mathbf{Pr}[i \in R]}{1-\beta_i}$. To see the other part of the claim, by definition of $\mathcal{E}_i$ it is the case that $\mathbf{Pr}[i \notin R \mid \overline{\mathcal{E}_i}] = \frac{(1-y^i-\beta_i)}{1-\beta_i}$ and $(1-\beta_i)\mathbf{Pr}[i \notin R] = (1-\beta_i)(1-y^i)$. For all $\beta_i \in [0,1)$ it is the case that $\frac{(1-y^i-\beta_i)}{1-\beta_i} \geq (1-\beta_i)(1-y^i)$ if $0 \leq y^i \leq \frac{1-\beta_i}{2}$. Finally, $0 \leq y^i \leq \frac{1-\delta}{2} \leq \frac{1-\beta_i}{2}$ when $t \leq \ln 2 - \delta$. This is because $t \leq \ln 2 - \delta$ by assumption and Lemma 4 states that any entry in y is at most $1 - e^{-t} \leq 1 - e^{-(\ln 2 - \delta)} \leq \frac{1-\delta}{2}$.

Now consider the second part of the lemma. Recall that $\mathcal{E}(\{i\})$ is the event where $\mathcal{E}_i$ occurs as well as $\overline{\mathcal{E}}_j$ for all $j \in I$ where $j \neq i$. We have the following.

$$\mathbf{Pr}[S' = R \mid \mathcal{E}(\{i\})] = \frac{\mathbf{Pr}[S' = R \text{ and } \mathcal{E}(\{i\})]}{\mathbf{Pr}[\mathcal{E}(\{i\})]}$$

By independence this equals the following.

$$\frac{\mathbf{Pr}[S' \setminus I = R \setminus I]}{\mathbf{Pr}[\mathcal{E}(\{i\})]}\mathbf{Pr}[\{i\} \cap S' = \{i\} \cap R \text{ and } \mathcal{E}_i] \prod_{j \in I, j \in S', j \neq i} \mathbf{Pr}[j \in R \text{ and } \overline{\mathcal{E}}_j]$$
$$\cdot \prod_{j \in I, j \notin S', j \neq i} \mathbf{Pr}[j \notin R \text{ and } \overline{\mathcal{E}}_j]$$

By independence we know that $\mathbf{Pr}[\mathcal{E}(\{i\})] = \mathbf{Pr}[\mathcal{E}_i] \prod_{j \in I, j \neq i} \mathbf{Pr}[\overline{\mathcal{E}}_j]$. Using this and conditional probability, the prior term is equal to the following.

$$\frac{\mathbf{Pr}[S' \setminus I = R \setminus I]}{\mathbf{Pr}[\mathcal{E}_i]}\mathbf{Pr}[\{i\} \cap S' = \{i\} \cap R \text{ and } \mathcal{E}_i] \prod_{j \in I, j \in S', j \neq i} \mathbf{Pr}[j \in R \mid \overline{\mathcal{E}}_j]$$
$$\cdot \prod_{j \in I, j \notin S', j \neq i} \mathbf{Pr}[j \notin R \mid \overline{\mathcal{E}}_j]$$

The first argument shown in the lemma gives that this is at least the following. This argument allows us to remove the conditioning on $\overline{\mathcal{E}}_j$.

$$\prod_{j \neq i, j \in I} (1-\beta_j)\frac{\mathbf{Pr}[S' \setminus I = R \setminus I]}{\mathbf{Pr}[\mathcal{E}_i]}\mathbf{Pr}[\{i\} \cap S' = \{i\} \cap R \text{ and } \mathcal{E}_i] \prod_{j \in I, j \in S', j \neq i} \mathbf{Pr}[j \in R]$$
$$\cdot \prod_{j \in I, j \notin S', j \neq i} \mathbf{Pr}[j \notin R]$$

Using independence, this is equal to the following.

$$\prod_{j \neq i, j \in I} (1-\beta_j)\frac{\mathbf{Pr}[S' = R \text{ and } \mathcal{E}_i]}{\mathbf{Pr}[\mathcal{E}_i]}$$

Finally, conditional probability gives the following.

$$\prod_{j \neq i, j \in I} (1-\beta_j)\mathbf{Pr}[S' = R \mid \mathcal{E}_i]$$

◁