

Submodular Maximization with Nearly-optimal Approximation and Adaptivity in Nearly-linear Time

Alina Ene*

Huy L. Nguyễn†

Abstract

In this paper, we study the tradeoff between the approximation guarantee and adaptivity for the problem of maximizing a monotone submodular function subject to a cardinality constraint. The adaptivity of an algorithm is the number of sequential rounds of queries it makes to the evaluation oracle of the function, where in every round the algorithm is allowed to make polynomially-many parallel queries. Adaptivity is an important consideration in settings where the objective function is estimated using samples and in applications where adaptivity is the main running time bottleneck. Previous algorithms achieving a nearly-optimal $1 - 1/e - \epsilon$ approximation require $\Omega(n)$ rounds of adaptivity. In this work, we give the first algorithm that achieves a $1 - 1/e - \epsilon$ approximation using $O(\ln n/\epsilon^2)$ rounds of adaptivity. The number of function evaluations and additional running time of the algorithm are $O(n \text{ poly}(\log n, 1/\epsilon))$.

1 Introduction

The general problem of maximizing a monotone submodular function subject to a size constraint captures many problems of interest both in theory and in practice, including sensor placement, clustering, and influence maximization in social networks. This problem has received considerable attention over the past few decades. The classical work of Nemhauser, Wolsey, and Fischer [NWF78] showed that a very natural Greedy algorithm achieves a $1 - 1/e$ approximation for the problem, and this approximation is known to be optimal [Fei98, Von09, DV12]. The ensuing decades have led to the development of powerful algorithmic frameworks as well as new applications in areas such as machine learning and data mining.

The Greedy algorithm and its variants play a central role in these developments: Greedy algorithms are natural and simple to use and they achieve the best

known approximation guarantees in many settings of interest. The main drawback of Greedy algorithms is that they are inherently sequential and their decisions are intrinsically adaptive.

A recent line of work has focused on addressing the first drawback of Greedy algorithms, and it has led to the development of distributed algorithms for submodular maximization problems in parallel models of computation such as MapReduce [KMVV13, MKSK13, MZ15, BENW15, MKBK15, BENW16, EMZ17]. The main focus of these works is on parallelizing sequential algorithms such as Greedy and its variants in order to achieve tradeoffs between the approximation guarantee and resources such as the number of rounds of MapReduce computation and the total amount of communication. In particular, Barbosa *et al.* [BENW16] show that it is possible to achieve a nearly-optimal $1 - 1/e - \epsilon$ approximation using $O(1/\epsilon)$ MapReduce rounds. The algorithms developed in these works run a Greedy algorithm on each of the machines, and thus they are just as adaptive as the sequential algorithms.

Very recently, Balkanski and Singer [BS18] initiated the study of the following question: can we design algorithms for submodular maximization that are less adaptive? The *adaptivity* of an algorithm is the number of sequential rounds of queries it makes to the evaluation oracle of the function, where in every round the algorithm is allowed to make polynomially-many parallel queries:

DEFINITION 1. ([BS18]) *Given an oracle f , an algorithm is r -adaptive if every query q to the oracle f occurs at a round $i \in [r]$ such that q is independent of the answers $f(q')$ to all other queries q' at round i .*

Adaptivity is an important consideration in settings where the objective function is estimated using samples and in applications where adaptivity is the main running time bottleneck. In applications of submodular maximization such as influence maximization and experimental design, queries are experiments that take time and can benefit greatly from parallel execution. In the broader context of optimization and computation, a lot of effort has been devoted to studying the trade-

*Department of Computer Science, Boston University, aene@bu.edu. Supported in part by NSF CAREER award 1750333 and NSF CCF award 1718342.

†College of Computer and Information Science, Northeastern University, hlnghuyen@cs.princeton.edu. Supported in part by NSF CAREER award 1750716.

off of adaptivity and other resources. For example, in property testing, adaptivity has been shown to be crucial with huge gaps in query complexity between non-adaptive and adaptive algorithms and more generally algorithms with different number of adaptive rounds [RS06, CG17]. In compressed sensing, adaptive algorithms can have exponentially fewer measurements than non-adaptive ones (see e.g. [IPW11]). We refer the reader to [BS18] for a more detailed discussion of applications of submodular maximization and the importance of adaptivity in their contexts as well as the study of adaptivity in various areas.

Balkanski and Singer give an algorithm that achieves a $1/3 - \epsilon$ approximation using $O(\log n/\epsilon^2)$ rounds of adaptivity, and they show that $\Omega(\log n/\log \log n)$ rounds of adaptivity are needed in order to obtain a $1/\log n$ approximation.

Thus there are now two incomparable algorithms for submodular maximization: the classical Greedy with optimal $1 - 1/e$ approximation but $O(k)$ adaptivity and the algorithm of [BS18] with $O(\log n/\epsilon^2)$ adaptivity but $1/3 - \epsilon$ approximation. One cannot help but ask

Is there an inherent tradeoff between adaptivity and approximation?

In this work, we obtain an algorithm that is the best of both worlds with nearly optimal approximation $1 - 1/e - \epsilon$ and $O(\log n/\epsilon^2)$ adaptivity and $O(n \text{ poly}(\log n, 1/\epsilon))$ total number of queries and additional running time.

THEOREM 1.1. *For the problem of maximizing a monotone submodular function subject to a cardinality constraint and any $\epsilon > 0$, there exists an $O(\log n/\epsilon^2)$ -adaptive randomized algorithm which obtains a $1 - 1/e - \epsilon$ approximation with high probability. The number of function evaluations and additional running time of the algorithm are $O(n \text{ poly}(\log n, 1/\epsilon))$.*

Comparison to [BS18]. Let us briefly highlight some of the differences between our work and that of [BS18] (see also Section 1.1). The algorithm of [BS18] is based on the single threshold Greedy algorithm, whereas our algorithm is based on the standard Greedy algorithm that achieves the optimal approximation guarantee in the sequential setting. The number of rounds of adaptivity that we obtain matches that of [BS18], and it is optimal up to lower order terms. Another important point of departure in our algorithm and its analysis is in the running time and function evaluations in each round of adaptivity: we sample only a *poly-logarithmic* number of random sets, and we evaluate the marginal gains only on these random sets, whereas the algorithm

of [BS18] uses a *polynomial* number of random sets and it evaluates the marginal gains with respect to all of these random sets. This allows us to obtain an overall nearly-linear running time and function evaluations, which matches up to logarithmic factors the best running time that we can achieve in the sequential setting. The low number of queries as well as their structure make it possible to obtain improved running times for applications such as the ones discussed above.

1.1 Our techniques The starting point of our algorithm is the standard Greedy algorithm that achieves the optimal $1 - 1/e$ approximation in the sequential setting. The Greedy algorithm constructs the solution sequentially over k iterations, where each iteration adds the element with maximum marginal gain on top of the current solution. An important observation about the standard analysis of Greedy is that the only property that we use about the element selected in each iteration is that its gain $f(S \cup \{e\}) - f(S)$ is at least $\frac{1}{k}(f(\text{OPT}) - f(S))$. Following [BS18], to achieve a low adaptivity, we want to add not just a single element but a much larger set of elements in each round of adaptivity. In contrast to [BS18], which based their approach on the single threshold Greedy algorithm, we draw inspiration from the standard Greedy algorithm and its analysis. Thus we aim to add a large set R in each iteration whose density (ratio of gain to size) nearly matches the density of the optimal solution, i.e., we have $\frac{f(R \cup S) - f(S)}{|R|} \geq (1 - O(\epsilon)) \frac{f(\text{OPT}) - f(S)}{k}$.

Perhaps surprisingly, we show that we can implement this strategy using $O(\ln n)$ rounds of adaptivity, which matches up to lower order terms the hardness result of [BS18]. The algorithm leverages the following dichotomy inspired by the work of [BS18] and the earlier work on sample and prune of [KMVV13]. Suppose that we choose the set R by sampling suitably many elements uniformly at random. If the expected density of the random set is almost as high as the target density of $\frac{f(\text{OPT}) - f(S)}{k}$, then we can add the random set to our solution and gain as much as Greedy. On the other hand, if the density is low, we are guaranteed that a constant fraction of the elements have low expected marginal gain. Using this insight, we make progress by filtering the elements with low marginal gain. Since each filtering step removes a constant fraction of the elements, we ensure that we have only $O(\ln n)$ rounds of adaptivity. At the same time, we are able to argue that the filtering steps preserve most of the value of OPT, which is essential for obtaining the nearly-optimal approximation guarantee.

We also ensure that the overall running time and function evaluations of our algorithm is nearly-linear.

This requires new insights and a different analysis from that of [BS18]. In each adaptive round, the algorithm needs to estimate the expected gain of a random set and the expected marginal gain of each element by sampling sufficiently many random sets. The analysis of [BS18] relies on having accurate estimates for the marginal gain of each element. Since some of the marginal gains can be very small (even among the elements of the optimal solution, most marginal gains might be as small as $\frac{1}{k} \cdot f(\text{OPT})$), it is necessary to sample a *polynomial* number of random sets to ensure that every gain is estimated accurately enough. We take a very different approach that allows us to sample only a *poly-logarithmic* number of random sets.

A key difficulty is to ensure that the filtering does not remove too much value from OPT. The crucial insight here is the following. Evaluating the marginal gain over a common set (random or otherwise) can significantly decrease the marginal gains of elements, but the *aggregate value* decreases by at most the value of the common set itself. Since we only filter when the random set has small value, even though some elements in OPT appear to have very small marginal gain with respect to the random set and they get filtered, the overall decrease in value can be charged to the random set and therefore it is low.

Independent work. Finally, we note two independent results with the same approximation and number of rounds of adaptivity [BRS18, FMZ18]. The algorithm of [BRS18] is similar to ours. The main difference between the two algorithms is in the number of random sets used in each round of adaptivity. Our algorithm uses a poly-logarithmic number of random sets and has an overall nearly-linear running time. In contrast, the algorithm of [BRS18] follows the approach of [BS18] and uses a polynomial number of random sets and an analysis based on estimating the marginal values. As discussed above, we use a different analysis to handle the much smaller number of random sets. The algorithm of [FMZ18] makes a linear number of queries in expectation.

1.2 Preliminaries and notation Let $f : 2^V \rightarrow \mathbb{R}_{\geq 0}$ be a set function on a ground set V of size $n = |V|$. The function is *submodular* if $f(A) + f(B) \geq f(A \cap B) + f(A \cup B)$ for all subsets A, B . The function is *monotone* if $f(A) \leq f(B)$ for all subsets A, B satisfying $A \subseteq B$.

We consider the problem of maximizing a monotone submodular function subject to a cardinality constraint: find $S^* \in \arg \max_{S \subseteq V: |S| \leq k} f(S)$.

For any two sets A and B , we use the notation $f(A|B)$ to denote the marginal gain of A on top of B , i.e., $f(A|B) = f(A \cup B) - f(B)$. For an element $e \in V$,

we use $f(e|B)$ as a shorthand for $f(\{e\}|B)$.

We will use the following Chernoff inequality, which follows from the standard Chernoff bound (see, e.g., [DP09]).

THEOREM 1.2. ([DP09]) *Let X_1, \dots, X_n be mutually independent and identically distributed random variables with $X_i \in [0, 1]$. Let $X = \frac{1}{n} \sum_{i=1}^n X_i$. Suppose that $\mathbb{E}[X] \leq \mu_H$. Then, for every $0 < \epsilon < 1$, we have*

$$\Pr[X > (1 + \epsilon)\mu_H] \leq \exp\left(-\frac{\epsilon^2}{3}n\mu_H\right).$$

2 Submodular Maximization Algorithm

Algorithm 1 The input is a submodular function $f : 2^V \rightarrow \mathbb{R}_{\geq 0}$ that is monotone and non-negative, a cardinality constraint k , and an error parameter ϵ .

```

1: For a set  $U \subseteq V$  and an integer  $\ell \leq |U|$ , we let
    $\mathcal{U}(U, \ell)$  be the uniform distribution over subsets of
    $U$  of cardinality  $\ell$ 
2:  $M$  is an approximate optimal solution value:  $M \leq$ 
    $f(\text{OPT}) \leq (1 + \epsilon)M$ 
3:  $m = O((\ln n)^2/\epsilon^4)$ ,  $\ell = \epsilon^2 k / (100 \ln n)$ 
4:  $S = \emptyset$ 
5: while  $f(S) \leq (1 - 1/e - O(\epsilon))M$  do
6:    $U_1 = V \setminus S$  ▷ Unfiltered elements
7:    $t = 0$ 
8:    $Old = f(S)$ 
9:   while  $f(S) - Old < \epsilon M / 100$  do
10:     $t \leftarrow t + 1$ 
11:    Let  $R_1, \dots, R_m$  be independent samples from
     $\mathcal{U}(U_t, \ell)$ 
12:    Let  $R_{max} = \arg \max_{R \in \{R_1, \dots, R_m\}} f(R|S)$ 
13:    if  $f(R_{max}|S) \geq (1 - 10\epsilon)\frac{\ell}{k}(M - Old)$  then
14:       $S \leftarrow S \cup R_{max}$ 
15:       $U_{t+1} = U_t \setminus R_{max}$ 
16:    else
17:      Let  $\{R_{i,j} : 1 \leq i \leq \Theta(\ln n/\epsilon), 1 \leq j \leq m\}$ 
      be independent samples from  $\mathcal{U}(U_t, \ell)$ 
18:      Let  $v_{i,e} = \frac{1}{m} \sum_{j=1}^m f(e|S \cup R_{i,j})$ 
19:      Let  $avg_i = \frac{1}{m} \sum_{j=1}^m f(R_{i,j}|S)$ 
20:      Let  $i$  be s.t.  $avg_i \leq (1 - 8\epsilon)\frac{\ell}{k}(M - Old)$ 
      and  $\sum_{e \in U_t} v_{i,e} \leq |U_t|(1 - 8\epsilon)\frac{M - Old}{k}$ 
21:      if there is no such  $i$ , declare failure and
      terminate
22:       $U_t^- = \{e \in U_t : v_{i,e} < (1 - 7\epsilon)\frac{M - Old}{k}\}$ 
23:       $U_{t+1} = U_t \setminus U_t^-$ 
24:    end if
25:  end while
26: end while
27: return  $S$ 

```

The algorithm is given in Algorithm 1. We assume that the algorithm has access to a value M such that $M \leq f(\text{OPT}) \leq (1 + \epsilon)M$. An n -approximation to $f(\text{OPT})$ is $M_0 = \max_{e \in V} f(\{e\})$. Given this value, we can try $2\epsilon^{-1} \ln n$ guesses for M : $M_0, (1 + \epsilon)M_0, (1 + \epsilon)^2 M_0, \dots$ in parallel and return the best solution from all the guesses.

The algorithm builds the solution over $1/\epsilon$ phases, where each phase increases the solution value by $\Omega(\epsilon f(\text{OPT}))$; a phase corresponds to a single iteration of the while loop on line 5. In each iteration of a phase (an iteration of the while loop on line 9), we aim to find a set with density $(1 - O(\epsilon)) \frac{f(\text{OPT}) - \text{Old}}{k}$ and size $\Theta\left(\frac{\epsilon^2}{\ln n}\right) \cdot k$, where Old is the value $f(S)$ of the solution at the beginning of the phase. If the expected density of a random set is at least the target density then, by sampling enough random sets, we can guarantee that with high probability we find a good set to add to our solution (see Lemma 2.2). On the other hand, if the expected density of a random set is below the target, then we can show that an ϵ fraction of the elements have expected marginal gain at most $1 + O(\epsilon)$ times the target density (see Lemma 2.1). Thus, by sampling enough random sets, we can guarantee that with high probability we filter an ϵ fraction of the elements on line 22.

Another key issue is determining how many random sets we need to sample. Since we are filtering based on marginal gains, it is tempting to proceed by ensuring that each marginal gain is estimated to sufficient accuracy. Unfortunately, this requires sampling polynomially many random sets. To obtain a fast running time, we take a different approach that uses only a polylogarithmic number of random sets. Since each random set has many elements, we can show using a Chernoff bound argument that our estimate of the expected value $\mathbb{E}_R[f(R|S)]$ of the random set is correct with high probability (see Lemma 2.2). When the expected value of the random set is small, it holds deterministically that the average of the expected marginal gains of the elements is small (see Claim 1). This fact together with a straightforward application of Markov's inequality and the Chernoff inequality gives us that with high probability the algorithm executes the filtering step on line 22 (see Lemma 2.2). This ensures that, when the expected value of the random set is low, we filter many elements with high probability. We also need to argue that we do not filter too much of the value of OPT. Here we cannot rely on the marginal values being estimated accurately and we need a different analysis. A crucial insight is that we can analyze the loss on aggregate. A key idea is to track the value $f((\text{OPT} \cap U_t) \cup S)$ of the optimal solution $\text{OPT} \cap U_t$ that has survived the filtering steps

so far. Instead of relying on having accurate marginal gains, we use a *deterministic* analysis to bound the loss in the value $f((\text{OPT} \cap U_t) \cup S)$ in a filtering round: since we are evaluating the marginal gains on common sets that have low value, even though some elements of OPT appear to have very small marginal gain with respect to these sets and they get filtered, the overall decrease in value is at most the value of the random sets themselves (see Claim 3).

2.1 The analysis of Algorithm 1 We divide the execution of the algorithm into phases corresponding to the iterations of the outer while loop. We will show that the number of phases is bounded by $O(1/\epsilon)$ and the number of iterations of the inner while loop in each phase is $O(\ln n/\epsilon)$. Therefore, the total number of rounds of adaptivity is $O(\ln n/\epsilon^2)$.

Consider an iteration of the inner while loop (line 9). We show that if $\mathbb{E}_{R \sim \mathcal{U}(U_t, \ell)}[f(R|S)] \geq (1 - 9\epsilon) \frac{\ell}{k} (M - \text{Old})$ then with high probability, the algorithm executes line 14 (see Lemma 2.1). On the other hand, if $\mathbb{E}_{R \sim \mathcal{U}(U_t, \ell)}[f(R|S)] < (1 - 9\epsilon) \frac{\ell}{k} (M - \text{Old})$ then, with high probability, the elements filtered on line 22 account for at least an ϵ fraction of all elements in U (see Lemma 2.2).

LEMMA 2.1. *Consider an iteration t with $\mathbb{E}_{R \sim \mathcal{U}(U_t, \ell)}[f(R|S)] \geq \frac{\ell}{k} (1 - 9\epsilon)(M - \text{Old})$. With probability $1 - 1/n^2$, we have $f(R_{\max}|S) \geq \frac{\ell}{k} (1 - 9\epsilon)(M - \text{Old})$ and the algorithm executes line 14.*

Proof. Using Markov's inequality, we will show that a given random set has a high value with probability at least $\Omega(\epsilon^3/\ln n)$. Since we are independently sampling $m = \Theta(\ln^2 n/\epsilon^4)$ sets, at least one of the sets has a high value with high probability.

Note that, for a random set $R \sim \mathcal{U}(U_t, \ell)$, we have $0 \leq f(R|S) \leq f(\text{OPT}) \leq (1 + \epsilon)M$: the first inequality follows by monotonicity, and the second inequality follows from the fact that $f(R|S) \leq f(R) \leq f(\text{OPT})$, since R is feasible. Since $(1 + \epsilon)M - f(R|S)$ is a non-negative random variable, it follows from Markov's inequality that

$$\begin{aligned} & \Pr_{R \sim \mathcal{U}(U_t, \ell)} \left[(1 + \epsilon)M - f(R|S) > \right. \\ & \quad \left. (1 + \epsilon)M - \frac{\ell}{k} (1 - 10\epsilon)(M - \text{Old}) \right] \\ & \leq \frac{\mathbb{E}_{R \sim \mathcal{U}(U_t, \ell)} [(1 + \epsilon)M - f(R|S)]}{(1 + \epsilon)M - \frac{\ell}{k} (1 - 10\epsilon)(M - \text{Old})} \\ & \leq \frac{(1 + \epsilon)M - \frac{\ell}{k} (1 - 9\epsilon)(M - \text{Old})}{(1 + \epsilon)M - \frac{\ell}{k} (1 - 10\epsilon)(M - \text{Old})} \end{aligned}$$

$$\begin{aligned}
&= 1 - \frac{\frac{\ell}{k}\epsilon(M - Old)}{(1 + \epsilon)M - \frac{\ell}{k}(1 - 10\epsilon)(M - Old)} \\
&= 1 - \Theta\left(\frac{\epsilon^3}{\ln n}\right)
\end{aligned}$$

The second inequality is our assumption, and the last equality follows from the fact that $\ell/k = \Theta(\epsilon^2/\ln n)$ and $(1/e + O(\epsilon))M \leq M - Old \leq M$.

Therefore, with probability at least $\Omega(\epsilon^3/\ln n)$, we have $f(R|S) \geq \frac{\ell}{k}(1 - 10\epsilon)(M - Old)$. Since we independently sample $m = \Theta((\ln n)^2/\epsilon^4)$ sets, with probability at least $1 - 1/n^2$, we find a set R_{max} such that $f(R_{max}|S) \geq \frac{\ell}{k}(1 - 10\epsilon)(M - Old)$.

We now consider the case when the expected value of the random set is below the target and show that, with high probability, the algorithm filters many elements on line 22.

LEMMA 2.2. *Consider an iteration t in which $\mathbb{E}_{R \sim \mathcal{U}(U_t, \ell)}[f(R|S)] \leq (1 - 9\epsilon)\frac{\ell}{k}(M - Old)$ and the algorithm executes line 16. With probability at least $1 - 1/n^3$, the algorithm does not fail on line 21. Additionally, if the algorithm does not fail then $|U_t^-| \geq \epsilon|U_t|$.*

Proof. We first give an overview of the proof. In Claim 1, we show that the expected density $\frac{\mathbb{E}_R[f(R|S)]}{|R|}$ of the random set is at least the average expected marginal gain $\frac{1}{|U_t|} \sum_{e \in U_t} \mathbb{E}_R[f(e|S \cup R)]$ of the elements. Note that this is a claim about expected values and thus it holds deterministically. Using a Chernoff inequality (Theorem 1.2), we show that, with high probability, we correctly determine that the expected value $\mathbb{E}_R[f(R|S)]$ of the random set is below the target. Additionally, we show that, with high probability, the algorithm succeeds to determine that the average marginal gain of the elements is low: by Claim 1 and Markov's inequality, a single random set succeeds with constant probability; since we independently sample poly-logarithmically many random sets, we obtain high probability overall. Thus, with high probability, the algorithm executes the filtering step on line 22 and a straightforward averaging argument shows that it filters an ϵ fraction of the elements.

CLAIM 1. *We have*

$$\mathbb{E}_{R \sim \mathcal{U}(U_t, \ell)}[f(R|S)] \geq \ell \cdot \frac{1}{|U_t|} \sum_{e \in U_t} \mathbb{E}[f(e|S \cup R)].$$

Proof. Consider a random set R . Order the elements of R arbitrarily as e_1, e_2, \dots, e_ℓ and let $R_0 = \emptyset$ and $R_i = \{e_1, \dots, e_i\}$.

$$f(R|S) = f(S \cup R) - f(S)$$

$$\begin{aligned}
&= \sum_{i=1}^{\ell} (f(S \cup R_i) - f(S \cup R_{i-1})) \\
&= \sum_{i=1}^{\ell} f(e_i|S \cup R_{i-1}) \\
&\geq \sum_{i=1}^{\ell} f(e_i|S \cup (R \setminus \{e_i\}))
\end{aligned}$$

where the last inequality follows from submodularity.

Therefore we have

$$\begin{aligned}
&\mathbb{E}_{R \sim \mathcal{U}(U_t, \ell)}[f(R|S)] \\
&\geq \mathbb{E}\left[\sum_{e \in R} f(e|S \cup (R \setminus \{e\}))\right] \\
&= \sum_e \Pr[e \in R] \cdot \mathbb{E}[f(e|S \cup (R \setminus \{e\}))|e \in R] \\
&= \sum_e \frac{\ell}{|U_t|} \cdot \mathbb{E}[f(e|S \cup (R \setminus \{e\}))|e \in R]
\end{aligned}$$

Let us now show that, for every e , we have

$$\begin{aligned}
&\mathbb{E}_{R \sim \mathcal{U}(U_t, \ell)}[f(e|S \cup (R \setminus \{e\}))|e \in R] \\
&\geq \mathbb{E}_{R \sim \mathcal{U}(U_t, \ell)}[f(e|S \cup (R \setminus \{e\}))|e \notin R]
\end{aligned}$$

To this end, note that we may assume that $R \sim \mathcal{U}(U_t, \ell)$ is generated by choosing a permutation π of U_t uniformly at random and letting $R = \{e_{\pi_1}, e_{\pi_2}, \dots, e_{\pi_\ell}\}$ be the first ℓ elements in this permutation. We have

$$\begin{aligned}
&\mathbb{E}_{R \sim \mathcal{U}(U_t, \ell)}[f(e|S \cup (R \setminus \{e\}))|e \in R] \\
&= \mathbb{E}_{R' \sim \mathcal{U}(U_t \setminus \{e\}, \ell - 1)}[f(e|S \cup R')] \\
&= \mathbb{E}_{R \sim \mathcal{U}(U_t, \ell)}[f(e|S \cup (R \setminus \{e, e_{\pi_1}\}))|e \notin R] \\
&\geq \mathbb{E}_{R \sim \mathcal{U}(U_t, \ell)}[f(e|S \cup (R \setminus \{e\}))|e \notin R]
\end{aligned}$$

In the first equality, we have used that, if $R \sim \mathcal{U}(U_t, \ell)$ and $e \in R$, then $R \setminus \{e\}$ has the distribution $\mathcal{U}(U_t \setminus \{e\}, \ell - 1)$. In the second equality, we have used that, if $R \sim \mathcal{U}(U_t, \ell)$ and $e \notin R$, then $R \setminus \{e, e_{\pi_1}\}$ has the distribution $\mathcal{U}(U_t \setminus \{e\}, \ell - 1)$, since e_{π_1} is an element of R . The inequality follows by submodularity.

Therefore

$$\begin{aligned}
&\mathbb{E}_{R \sim \mathcal{U}(U_t, \ell)}[f(e|S \cup (R \setminus \{e\}))] \\
&= \mathbb{E}[f(e|S \cup (R \setminus \{e\}))|e \in R] \Pr[e \in R] \\
&\quad + \mathbb{E}[f(e|S \cup (R \setminus \{e\}))|e \notin R] \Pr[e \notin R] \\
&\leq \mathbb{E}[f(e|S \cup (R \setminus \{e\}))|e \in R] (\Pr[e \in R] + \Pr[e \notin R])
\end{aligned}$$

$$= \mathbb{E}[f(e|S \cup (R \setminus \{e\})|e \in R]$$

By combining the inequalities above and using submodularity, we obtain

$$\begin{aligned} & \mathbb{E}_{R \sim \mathcal{U}(U_t, \ell)}[f(R|S)] \\ & \geq \sum_e \frac{\ell}{|U_t|} \cdot \mathbb{E}[f(e|S \cup (R \setminus \{e\}))] \\ & \geq \sum_e \frac{\ell}{|U_t|} \cdot \mathbb{E}[f(e|S \cup R)] \end{aligned}$$

Let us now show that, with probability $1 - 1/n^3$, there is a batch of random sets $\{R_{i,j} : j \in [m]\}$ with the properties stated on line 20.

Fix a batch i . Using Theorem 1.2, we can upper bound the probability of the event that $avg_i > (1 - 8\epsilon)\frac{\ell}{k}(M - Old)$ as follows. For each $j \in [m]$, let $X_j = \frac{1}{(1+\epsilon)M}f(R_{i,j}|S) \in [0, 1]$. Let $X = \frac{1}{m} \sum_{j=1}^m X_j$. By our assumption, we have

$$\mathbb{E}_{R \sim \mathcal{U}(U_t, \ell)}[X_j] \leq \frac{1}{(1+\epsilon)M}(1 - 9\epsilon)\frac{\ell}{k}(M - Old).$$

Let $\mu_H = \frac{1}{(1+\epsilon)M}(1 - 9\epsilon)\frac{\ell}{k}(M - Old)$. By Theorem 1.2,

$$\Pr[X > (1 + \epsilon)\mu_H] \leq \exp\left(-\frac{\epsilon^2}{3}m\mu_H\right) \leq \frac{1}{n^3},$$

where the second inequality follows by substituting m and μ_H , and using the fact that $(M - Old)/M = \Theta(1)$.

We now upper bound the probability of the event that $\sum_{e \in U_t} v_{i,e} > |U_t|(1 - 8\epsilon)\frac{M - Old}{k}$. By Markov's inequality, with probability at least ϵ , we have

$$\begin{aligned} & \frac{1}{m} \sum_{j=1}^m \left(\sum_e f(e|S \cup R_{i,j}) \right) \\ & \leq \frac{1}{1 - \epsilon} \mathbb{E}_{R \sim \mathcal{U}(U_t, \ell)} \left[\sum_e f(e|S \cup R) \right] \\ & \leq \frac{1}{1 - \epsilon} \frac{|U_t|}{\ell} \mathbb{E}_{R \sim \mathcal{U}(U_t, \ell)} [f(R|S)] \\ & \leq \frac{1}{1 - \epsilon} \frac{|U_t|}{\ell} \frac{\ell}{k} (1 - 9\epsilon)(M - Old) \\ & \leq |U_t|(1 - 8\epsilon) \frac{M - Old}{k} \end{aligned}$$

In the second inequality, we have used Claim 1.

Therefore each batch i satisfies both conditions of line 20 with probability at least $\epsilon - 1/n^3$. Since the batches are independent, it follows that the probability that the algorithm fails on line 21 is at most $(1 - \epsilon + 1/n^3)^{\Theta(\ln n/\epsilon)} \leq 1/n^3$.

Let us now condition on the event that the algorithm does not fail. Consider the set U_t^- filtered on line 22. We have

$$\begin{aligned} & |U_t \setminus U_t^-|(1 - 7\epsilon) \frac{M - Old}{k} \\ & \leq \sum_{e \in U_t} v_{i,e} \\ & \leq |U_t|(1 - 8\epsilon) \frac{M - Old}{k}, \end{aligned}$$

and thus $|U_t^-| \geq \epsilon|U_t|$.

We now show that the number of phases is $O(1/\epsilon)$ and the number of iterations in each phase is $O(\ln n/\epsilon)$. The former simply follows from the fact that each phase increases the value of the solution by $\Omega(\epsilon f(\text{OPT}))$. Most of the work is to show that the filtering steps do not remove too much of the optimal solution. A subtle but crucial choice is to track the value of the optimal solution $\text{OPT} \cap U_t$ that has survived the filtering steps so far. In Claim 3, we analyze how much this value decreases in each filtering iteration and show that this decrease can be charged to the value of the random sets, which have low value. Claim 3 then allows us to show that we cannot keep filtering without eventually finding a good set to add on line 14: since each filtering iteration removes an ϵ fraction of the elements, after $O(\ln n/\epsilon)$ filtering iterations the ground set becomes empty; on the other hand, Claim 3 shows that $O(\ln n/\epsilon)$ filtering iterations is not enough to remove all of the value of OPT , since $f(\text{OPT} \cap U_t|S)$ is strictly positive (see Claim 4).

LEMMA 2.3. *Consider a phase of the algorithm. The phase increases $f(S)$ by $\Omega(\epsilon M)$. Additionally, with probability at least $1 - 1/n^2$, the phase has $O(\ln n/\epsilon)$ iterations.*

Proof. The lower bound on the increase follows from the terminating condition for the phase. Thus it only remains to bound the number of iterations. We refer to each iteration as a gain iteration if line 13 is executed, and as a filtering iteration if line 16 is executed. We show that the number of gain iterations is $O(\ln n/\epsilon)$ with probability 1, and the number of filter iterations is $O(\ln n/\epsilon)$ with probability at least $1 - 1/n^2$.

CLAIM 2. *The number of gain iterations is at most $6 \ln n/\epsilon$.*

Proof. Each gain iteration increases $f(S)$ by $\frac{\ell}{k}(1 - 10\epsilon)(M - Old)$. Since $M - Old \geq M/3$ and the phase ends when $f(S) - Old$ becomes $\epsilon M/100$, the number of gain iterations is at most $6 \ln n/\epsilon$.

Let us now consider the filtering iterations. Let T be the minimum of $2 \ln n / \epsilon + 1$ and the number of filtering iterations of the phase. By Lemma 2.2, the probability that none of the first T filtering iterations fails is at least $1 - T/n^3 \geq 1 - 1/n^2$. In the following, we condition on the event that none of the first T filtering iterations fails.

We can show the following invariant.

CLAIM 3. *Consider an iteration t and suppose that the number of filtering iterations so far is at most T . At the beginning of iteration t , we have*

$$\begin{aligned} & f((\text{OPT} \cap U_t) \cup S) \\ & \geq M - \frac{(t-1)\ell}{k}(M - \text{Old}) \\ & \quad - \frac{|\text{OPT} \setminus (U_t \cup S)|}{k}(1 - 7\epsilon)(M - \text{Old}). \end{aligned}$$

Proof. We will prove the invariant by induction on t . The invariant is true at the beginning of iteration 1 since $f((\text{OPT} \cap U_1) \cup S) \geq f(\text{OPT}) \geq M$.

Consider iteration $t > 1$ and suppose the invariant holds at the beginning of iteration t . We will show that the invariant continues to hold at the beginning of iteration $t + 1$.

Suppose that iteration t is a gain iteration. The algorithm adds the random set R_{max} to S on line 14 and we have

$$f((\text{OPT} \cap (U_t \setminus R_{max})) \cup S \cup R_{max}) \geq f((\text{OPT} \cap U_t) \cup S),$$

since $(\text{OPT} \cap (U_t \setminus R_{max})) \cup S \cup R_{max} \supseteq (\text{OPT} \cap U_t) \cup S$ and f is monotone. Thus the invariant continues to hold at the beginning of iteration $t + 1$.

Therefore we may assume that iteration t is a filtering iteration. Recall that we are conditioning on the event that the algorithm does not fail in the first T filtering iterations, and thus iteration t executes line 22. Let i be the index satisfying the conditions on line 20. We have

$$\begin{aligned} & f(\text{OPT} \cap U_{t+1} | S) \\ & \geq \frac{1}{m} \sum_{j=1}^m f(\text{OPT} \cap U_{t+1} | S \cup R_{i,j}) \\ & \geq \frac{1}{m} \sum_{j=1}^m \left(f(\text{OPT} \cap U_t | S \cup R_{i,j}) \right. \\ & \quad \left. - \sum_{e \in \text{OPT} \cap U_t^-} f(e | S \cup R_{i,j}) \right) \\ & \geq f(\text{OPT} \cap U_t | S) - \frac{1}{m} \sum_{j=1}^m f(R_{i,j} | S) \end{aligned}$$

$$\begin{aligned} & - |\text{OPT} \cap U_t^-| (1 - 7\epsilon) \frac{1}{k} (M - \text{Old}) \\ & \geq f(\text{OPT} \cap U_t | S) - \frac{\ell}{k} (M - \text{Old}) \\ & \quad - |\text{OPT} \cap U_t^-| (1 - 7\epsilon) \frac{1}{k} (M - \text{Old}) \end{aligned}$$

The first and second inequalities follow by submodularity. In the third inequality, we have used monotonicity to bound $f((\text{OPT} \cap U_t) \cup S \cup R_{i,j}) \geq f((\text{OPT} \cap U_t) \cup S)$, and we have used that $v_{i,e} \leq (1 - 7\epsilon) \frac{1}{k} (M - \text{Old})$ for all $e \in U_t^-$. In the fourth inequality, we have used that $\text{avg}_i \leq \frac{\ell}{k} (1 - 8\epsilon) (M - \text{Old})$.

It follows that the invariant continues to hold at the beginning of iteration $t + 1$.

CLAIM 4. *The number of filtering iterations is at most $2 \ln n / \epsilon$.*

Proof. Recall that we are conditioning on the event that the first T filtering iterations do not fail. Suppose for contradiction that the number of filtering iterations reaches $2 \ln n / \epsilon + 1$, and let t be the iteration when this happens. By Lemma 2.2, each filtering iteration removes an ϵ fraction of U and thus $U_t = \emptyset$. By Claim 2, the number of gain iterations is at most $6 \ln n / \epsilon$ and thus $t \leq 8 \ln n / \epsilon + 1$. By Claim 3, at the beginning of iteration t , we have

$$\begin{aligned} & f((\text{OPT} \cap U_t) \cup S) - f(S) \\ & \geq M - \frac{(t-1)\ell}{k}(M - \text{Old}) \\ & \quad - \frac{|\text{OPT} \setminus (U_t \cup S)|}{k}(1 - 7\epsilon)(M - \text{Old}) - f(S) \\ & \geq M - \frac{8\epsilon}{100}(M - \text{Old}) - (1 - 7\epsilon)(M - \text{Old}) - f(S) \\ & = \left(7\epsilon - \frac{8\epsilon}{100} \right) (M - \text{Old}) - (f(S) - \text{Old}) \\ & \geq \left(7\epsilon - \frac{8\epsilon}{100} \right) \cdot \frac{1}{3} M - \frac{\epsilon}{100} M \\ & > 0 \end{aligned}$$

In the second inequality, we used that $|\text{OPT} \setminus (S_t \cup S)| \leq k$. In the third inequality, we used that $M - \text{Old} \geq (1/e + O(\epsilon))M \geq 1/3M$. In the last inequality, we used the fact that $f(S) - \text{Old} < \epsilon M / 100$, since the phase has not ended.

It follows that U_t is non-empty, which is a contradiction. Therefore the number of filtering phases is at most $2 \ln n / \epsilon$.

LEMMA 2.4. *With probability $1 - 1/n$, the algorithm uses $O(\ln n / \epsilon^2)$ rounds of queries.*

Proof. The lemma follows from Lemma 2.3. Each phase increases $f(S)$ by $\Omega(\epsilon M)$ and the algorithm stops when $f(S) \geq (1 - 1/e - O(\epsilon))M$. Thus, the number of phases is $O(1/\epsilon)$. With probability $1 - \frac{1}{\epsilon n^2} \geq 1 - \frac{1}{n}$, every phase uses $O(\ln n/\epsilon)$ rounds of queries. Therefore, with probability $1 - 1/n$, the total number of rounds of queries is $O(\ln n/\epsilon^2)$.

Finally, we argue that the algorithm returns a feasible solution and achieves a $1 - 1/e - O(\epsilon)$ approximation ratio.

LEMMA 2.5. *The algorithm returns a feasible solution and achieves a $1 - 1/e - O(\epsilon)$ approximation.*

Proof. Consider iteration j of a phase where the algorithm executes line 14. We have

$$\begin{aligned} f(S \cup R_{max}) - f(S) &\geq |R_{max}|(1 - 10\epsilon)(M - Old)/k \\ &\geq |R_{max}|(1 - 11\epsilon)(M - f(S))/k \end{aligned}$$

Using the inequality above, we can show by induction that

$$M - f(S) \leq \exp\left(-\frac{(1 - 11\epsilon)|S|}{k}\right)M.$$

Initially, $S = \emptyset$ and the inequality holds. Consider an iteration where the inequality holds at the beginning of the iteration. We have

$$\begin{aligned} M - f(S \cup R_{max}) &\leq (M - f(S))\left(1 - \frac{|R_{max}|(1 - 11\epsilon)}{k}\right) \\ &\leq M \exp\left(-\frac{(1 - 11\epsilon)|S|}{k}\right) \exp\left(-\frac{(1 - 11\epsilon)|R_{max}|}{k}\right) \\ &= M \exp\left(-\frac{(1 - 11\epsilon)|S \cup R_{max}|}{k}\right) \end{aligned}$$

This completes the induction.

By the induction, in the first iteration where $f(S) \geq (1 - \exp(-(1 - 12\epsilon)))M$ we must have $|S| \leq (1 - \epsilon)k + \ell < k$. Thus, the final solution satisfies the constraint $|S| \leq k$ and has value $f(S) \geq (1 - \exp(-(1 - 12\epsilon)))M$.

References

- [BENW15] Rafael D.P. Barbosa, Alina Ene, Huy L. Nguyen, and Justin Ward. The power of randomization: Distributed submodular maximization on massive datasets. In *International Conference on Machine Learning (ICML)*, pages 1236–1244, 2015.
- [BENW16] Rafael da Ponte Barbosa, Alina Ene, Huy L. Nguyen, and Justin Ward. A new framework for distributed submodular maximization. In *IEEE Foundations of Computer Science (FOCS)*, pages 645–654, 2016.
- [BRS18] Eric Balkanski, Aviad Rubinfeld, and Yaron Singer. An exponential speedup in parallel running time for submodular maximization without loss in approximation. *CoRR*, abs/1804.06355, 2018.
- [BS18] Eric Balkanski and Yaron Singer. The adaptive complexity of maximizing a submodular function. In *ACM Symposium on Theory of Computing (STOC)*, pages 1138–1151, 2018.
- [CG17] Clément L. Canonne and Tom Gur. An adaptivity hierarchy theorem for property testing. *CoRR*, abs/1702.05678, 2017.
- [DP09] Devdatt P. Dubhashi and Alessandro Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009.
- [DV12] Shahar Dobzinski and Jan Vondrák. From query complexity to computational complexity. In *ACM Symposium on Theory of Computing (STOC)*, pages 1107–1116, 2012.
- [EMZ17] Alessandro Epasto, Vahab Mirrokni, and Morteza Zadimoghaddam. Bicriteria distributed submodular maximization in a few rounds. In *PACM Symposium on Parallelism in Algorithms and Architectures (SPAA)*, pages 25–33, 2017.
- [Fei98] Uriel Feige. A threshold of $\ln n$ for approximating set cover. *Journal of the ACM*, 45:634–652, 1998.
- [FMZ18] Matthew Fahrbach, Vahab Mirrokni, and Morteza Zadimoghaddam. Submodular maximization with optimal approximation, adaptivity and query complexity. *CoRR*, abs/1807.07889, 2018.
- [IPW11] Piotr Indyk, Eric Price, and David P. Woodruff. On the power of adaptivity in sparse recovery. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 285–294. IEEE, 2011.
- [KMVV13] Ravi Kumar, Benjamin Moseley, Sergei Vassilvitskii, and Andrea Vattani. Fast greedy algorithms in mapreduce and streaming. In *PACM Symposium on Parallelism in Algorithms and Architectures (SPAA)*, pages 1–10, 2013.
- [MKBK15] Baharan Mirzasoleiman, Amin Karbasi, Ashwinkumar Badanidiyuru, and Andreas Krause. Distributed submodular cover: Succinctly summarizing massive data. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2881–2889, 2015.
- [MKS13] Baharan Mirzasoleiman, Amin Karbasi, Rik Sarkar, and Andreas Krause. Distributed submodular maximization: Identifying representative elements in massive data. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2049–2057, 2013.
- [MZ15] Vahab Mirrokni and Morteza Zadimoghaddam. Randomized composable core-sets for distributed submodular maximization. In *ACM Symposium on Theory of Computing (STOC)*, pages 153–162, 2015.

- [NWF78] G L Nemhauser, L A Wolsey, and M L Fisher. An analysis of approximations for maximizing submodular set functions—i. *Mathematical Programming*, 14(1):265–294, 1978.
- [RS06] Sofya Raskhodnikova and Adam D. Smith. A note on adaptivity in testing properties of bounded degree graphs. *Electronic Colloquium on Computational Complexity (ECCC)*, 13(089), 2006.
- [Von09] Jan Vondrák. Symmetry and approximability of submodular maximization problems. In *IEEE Foundations of Computer Science (FOCS)*, pages 651–670, 2009.