Operations Research Letters 49 (2021) 855–861

Contents lists available at ScienceDirect

Operations Research Letters

www.elsevier.com/locate/orl

Performance guarantees of forward and reverse greedy algorithms for minimizing nonsupermodular nonsubmodular functions on a matroid

Orcun Karaca [a,*], Daniel Tihanyi [a], Maryam Kamgarpour [b]

[a] Automatic Control Laboratory, D-ITET, ETH Zürich, ETL K12, Physikstrasse 3, 8092, Zürich, Switzerland
[b] Electrical and Computer Engineering Department, University of British Columbia, Vancouver, Canada

ARTICLE INFO

Article history:
Received 26 June 2021
Received in revised form 27 September 2021
Accepted 29 September 2021
Available online 5 October 2021

Keywords:
Combinatorial optimization
Greedy algorithm
Matroid theory

ABSTRACT

This letter studies the problem of minimizing increasing set functions, equivalently, maximizing decreasing set functions, over the base matroid. This setting has received great interest, since it generalizes several applied problems including actuator and sensor placement problems in control, task allocation problems, video summarization, and many others. We study two greedy heuristics, namely, the forward and reverse greedy. We provide two novel performance guarantees for the approximate solutions obtained by them depending on both submodularity ratio and curvature.

1. Introduction

Set function optimization is an active field of research which is used in a broad range of applications including video summarization in machine learning [2], splice site detection in computational biology [3], actuator and sensor placement problems in control [5], task allocation problems in robotics and many other domains [11]. In this letter, we study the following instance: the problem of minimizing an increasing nonsubmodular and nonsupermodular set function (equivalently, maximizing a decreasing function) over the base of a matroid. This is general enough to model many of these applications for which NP-hardness results are readily available in the literature [25]. Thus, it is desirable to obtain scalable algorithms with provable suboptimality bounds.

Many studies have adopted greedy heuristics, because of their polynomial-time complexity and the performance bounds they are equipped with [6,9,20,22]. For the forward greedy algorithm applied to our setting, [22, Thm 7] provides a performance guarantee based on the notion of *strong* curvature, describing how modular the objective is (we use the term strong curvature to differentiate between the definition of curvature from [22] and the definition utilized in our letter and also in [2]). [12, Props 4 and 5] show that for this setting one cannot derive any performance guarantee unless both submodular and supermodular-like properties are present. As an alternative to [22, Thm 7], [12, Thm 2] provides a guarantee based on the notions of submodularity ratio (describing how submodular the objective is) and curvature (describing how supermodular the objective is), which together form a weaker condition than strong curvature. However, this guarantee scales with the cardinality of the ground set. To the best of our knowledge, there exists no forward greedy guarantee applicable to our setting utilizing both submodularity ratio and curvature, simultaneously, which is also problem-size independent. This will be the first goal of this letter.

An inherent drawback of the forward greedy algorithm is that any performance guarantee has to involve the objective function evaluated at the empty set as the reference value. This reference is known to have an undesirable effect in several applications [4]. An alternative is to adopt the reverse greedy, which excludes the least desirable elements iteratively starting from the full set. In this case, any potential performance guarantee would instead involve the objective function evaluated at the full set, which could be a more preferable reference point. For the reverse greedy applied to our setting, [22, Thm 6] provides a performance guarantee again based on the notion of strong curvature. When only cardinality constraints are present, [2, Thm 1] is applicable and it provides a guarantee based on the weaker notions of submodularity ratio and curvature. However, to the best of our knowledge, there exists no reverse greedy guarantee applicable to our setting involving matroid constraints utilizing both curvature and submodularity ratio, simultaneously.

Our contributions are as follows. We obtain a performance guarantee for the forward greedy algorithm applied to minimizing increasing nonsubmodular and nonsupermodular set functions,

* Corresponding author.
E-mail address: okaraca@ethz.ch (O. Karaca).

https://doi.org/10.1016/j.orl.2021.09.006
0167-6377/

characterized by submodularity ratio and curvature, over the base of a matroid. In contrast to [12, Thm 2], this new guarantee is problem-size independent. This result is presented in Theorem 1. For the same setting, we then obtain a performance guarantee for the reverse greedy algorithm, see Theorem 2. This guarantee is novel, and no previous work obtained a reverse greedy performance guarantee for this setting involving both curvature and submodularity ratio. We remark that the derivations of the two guarantees are distinct from each other. This is because even though both algorithms are essentially greedy heuristics, their iterations will be shown to exhibit opposite behavior (greedy minimization vs. greedy maximization). Each guarantee utilizes a variant of the ordering property of matroids in its derivation steps, see Lemmas 1 and 2. These properties are adaptations of [14, Lem 1], which was originally inspired by its continuous polymatroid counterpart from [6, Thm 6.1]. For both guarantees, we demonstrate more efficient greedy notions of the curvature and the submodularity ratio, since the original definitions are computationally intractable. After each theoretical result, we provide a detailed comparison of the guarantees with those found in the literature. Finally, we compare the two performance guarantees for different values of submodularity ratio and curvature.

2. Preliminaries

Let V be a finite ground set and $f : 2^V \to \mathbb{R}$. Function is called normalized whenever $f(\emptyset) = 0$, however, we assume this is not necessarily the case. For simplicity, we use j and $\{j\}$ interchangeably for singleton sets.

Definition 1. A function f is *increasing* if $f(S) \leq f(R)$, for all $S \subseteq R \subseteq V$. Function $-f$ is then *decreasing*. If the inequality is strict whenever $S \subsetneq R$, then f is *strictly increasing* and $-f$ is *strictly decreasing*.

Definition 2. For $S \subseteq V$, $j \in V$, the *discrete derivative* (or marginal gain) of f at S with respect to j is given by $\rho_j(S) = f(S \cup j) - f(S)$.

If $j \in S$, we have $\rho_j(S) = 0$. For $R \subseteq V$, we generalize the definition above to $\rho_R(S) = f(S \cup R) - f(S)$.

Definition 3. A function f is *submodular* if $\rho_j(R) \leq \rho_j(S)$, for all $S \subseteq R \subseteq V$, for all $j \in V \setminus R$.

In several practical problems, the discrete derivative diminishes as S expands yielding the submodularity property, see the examples in [1]. Unfortunately, functions used in many problems, including the ones we consider, do not have this property. Instead, these problems involve increasing set functions, allowing the use of *submodularity ratio* describing how far a nonsubmodular set function is from being submodular. This property was first introduced by [19].

Definition 4. The *submodularity ratio* of an increasing function f is the largest scalar $\gamma \in \mathbb{R}_+$ such that $\gamma \rho_j(R) \leq \rho_j(S)$, for all $S \subseteq R \subseteq V$, for all $j \in V \setminus R$. Function with submodularity ratio γ is called γ-submodular.

Observe that the definition above is not well-posed unless f is increasing, in other words, both $\rho_j(R)$ and $\rho_j(S)$ should be nonnegative. It can easily be verified that, for an increasing function f, we have $\gamma \in [0, 1]$ and submodularity is attained if and only if $\gamma = 1$. We briefly review an alternative but nonequivalent submodularity ratio notion from [2,7]. The *cumulative submodularity ratio* is the largest scalar $\gamma' \in \mathbb{R}_+$ such that $\gamma' \rho_R(S) \leq \sum_{j \in R \setminus S} \rho_j(S)$, for all $S, R \subseteq V$. The ratio γ of Definition 4 satisfies the inequalities for the definition of γ', but the reverse argument does not necessarily hold. Hence, $\gamma \leq \gamma'$, see [12, App. B]. This notion γ' is generally restricted to the case when deriving guarantees for greedy heuristics with only cardinality constraints, because only then it allows to derive bounds of the form of [2, Lem 1] to be used for the linear programming proof from [6]. Utilizing the submodularity notion as per Definition 4 is needed for the guarantees we will derive.

Definition 5. A function f is *supermodular* if $\rho_j(R) \geq \rho_j(S)$, for all $S \subseteq R \subseteq V$, for all $j \in V \setminus R$.

Other than submodularity, another widely-used notion is supermodularity we defined above, that is, the increasing discrete derivatives property. A function which is both supermodular and submodular is modular/additive. Similar to the case with submodularity, objective functions we consider do not exhibit supermodularity as well. By introducing the *curvature*, that is, how far a nonsupermodular increasing function is from being supermodular, we obtain a more precise description on how the discrete derivatives change.

Definition 6. The *curvature* of an increasing function f is the smallest scalar $\alpha \in \mathbb{R}_+$ such that $\rho_j(R) \geq (1 - \alpha)\rho_j(S)$, for all $S \subseteq R \subseteq V$, for all $j \in V \setminus R$. Function f with curvature α is called α-supermodular.

It can easily be verified that, for an increasing function f, we have $\alpha \in [0, 1]$ and supermodularity is attained if and only if $\alpha = 0$. A cumulative definition is also applicable, however, we leave it out. As a remark, the computation of both the curvature and the submodularity ratio requires an exhaustive enumeration of inequalities, which is typically addressed by either approximations or ex-ante bounds. For each theoretical guarantee, we will show that some of these inequalities in the definitions are not needed. This way, we can improve the tractability of these two notions when evaluating the guarantees.

Next, we provide two propositions regarding these ratios. The first is an observation relevant for the applications from the literature we consider. The second will be useful when adopting the reverse greedy algorithm.

Proposition 1. *Suppose f is strictly increasing and $0 < \underline{b} \leq \rho_j(S) \leq \overline{b}$, for all $S \subsetneq V$, for all $j \notin S$. Then, we have $\gamma \geq \underline{b}/\overline{b}$ and $\alpha \leq 1 - \underline{b}/\overline{b}$.*

Proof. Since $\rho_j(R) > 0$, by reorganizing Definition 4, we obtain $\gamma = \min_{S \subseteq R \subseteq V,\ j \in V \setminus R}(\rho_j(S)/\rho_j(R))$. Clearly, the term on the right is lower bounded by $\underline{b}/\overline{b}$. Similarly, we can reorganize Definition 6 as follows $(1/(1-\alpha)) = \max_{S \subseteq R \subseteq V,\ j \in V \setminus R}(\rho_j(S)/\rho_j(R))$. The term on the right is upper bounded by $\overline{b}/\underline{b}$. A simple manipulation of the inequality gives us the desired result. □

Condition above also implies that our function can be upper and lower bounded by two (sub)modular functions. In literature, this property is called differential submodularity, which found its applications in deriving guarantees for parallel adaptive sampling algorithms, see [21].

Proposition 2. *Let $\hat{f}(S) = -f(V \setminus S)$, for all $S \subseteq V$. Let $\hat{\gamma}$ and $\hat{\alpha}$ be submodularity ratio and curvature of $\hat{f}$, respectively. Then, $\hat{\gamma} = 1 - \alpha$ and $\hat{\alpha} = 1 - \gamma$.*

Proof. Observe that function $\hat{f}$ is increasing, hence submodularity ratio and curvature are well-defined. Let $\hat{\rho}_j(S) = \hat{f}(S \cup j) - \hat{f}(S) = f(V \setminus S) - f(\{V \setminus S\} \setminus j\})$, for all S, j. The submodularity ratio of function $\hat{f}$ is the largest scalar $\hat{\gamma}$ such that $\hat{\gamma}\hat{\rho}_j(R) \le \hat{\rho}_j(S)$, for all $S \subseteq R \subseteq V$, for all $j \in V \setminus R$. Using the definition of $\hat{\rho}_j$, this ratio is also the largest scalar $\hat{\gamma}$ such that $\hat{\gamma}\rho_j(S) \le \rho_j(R)$, for all $S \subseteq R \subseteq V$, for all $j \in V \setminus R$. Thus, $\hat{\gamma} = 1 - \alpha$ by Definition 6. Using a similar reasoning, one would also obtain $\hat{\alpha} = 1 - \gamma$. □

Many combinatorial optimization problems from the literature are subject to constraints that are more complex than simple cardinality constraints, see the examples in [18]. Among those, we introduce matroids. They will capture problems of interest stated in Section 2.1. Moreover, they are known to allow performance guarantees for greedy heuristics thanks to their specific properties outlined below [8].

Definition 7. A *matroid* $\mathcal{M}$ is an ordered pair $(V, \mathcal{F})$ consisting of a ground set V and a nonempty collection $\mathcal{F}$ of subsets of V which satisfies *(i)* $\emptyset \in \mathcal{F}$, *(ii)* if $S \in \mathcal{F}$ and $R \subsetneq S$, then $R \in \mathcal{F}$, *(iii)* if $S_1, S_2 \in \mathcal{F}$ and $|S_1| < |S_2|$, there exists $j \in S_2 \setminus S_1$ such that $j \cup S_1 \in \mathcal{F}$. Every set in $\mathcal{F}$ is called *independent*. Maximum independent sets refer to those with the largest cardinality, and they are called the bases of a matroid. All bases have the same cardinality, and the cardinality of the bases is called the rank of the matroid.

The last property (iii) of a matroid is considered as the generalization of the linear independence relation from linear algebra. Intuitively, this property will later let us to keep track of the elements that the greedy algorithm is missing from the optimal solution. To give an example, two well-studied matroids are the uniform matroid (i.e., $\{S \subseteq V \mid |S| \le Q\}$, where $Q \in \mathbb{Z}_+$, $Q \le |V|$) and the partition matroid (i.e., $\{S \subseteq V \mid |S \cap Q_i| \le q_i, \text{ for } i = 1, \dots, \ell\}$, where $\{Q_i\}_{i=1}^{\ell}$ is a family of disjoint sets that from a partition of V, and $\{q_i\}_{i=1}^{\ell}$ are some given positive integers).

To adopt the reverse greedy algorithm, an additional concept will be required, that is, the dual of a matroid.

Definition 8. Given a matroid $\mathcal{M} = (V, \mathcal{F})$, let $\hat{\mathcal{F}} = \{U \mid \exists$ a base $M \in \mathcal{F}$ such that $U \subseteq V \setminus M\}$. The pair $\hat{\mathcal{M}} = (V, \hat{\mathcal{F}})$ is called the *dual* of the matroid $(V, \mathcal{F})$.

The pair $\hat{\mathcal{M}} = (V, \hat{\mathcal{F}})$ satisfies all the axioms of a matroid. Suppose $\{M_i\}_{i=1}^{q}$ is the collection of all bases of matroid $(V, \mathcal{F})$. Then, $\{V \setminus M_i\}_{i=1}^{q}$ defines the collection of all bases for the dual $(V, \hat{\mathcal{F}})$, see [24, Ch. 2].

2.1. Problem formulation

Our goal is to solve

$$\begin{aligned} &\min_{S \subseteq V} f(S), \text{ increasing, } \gamma\text{-submodular, } \alpha\text{-supermodular} \\ &\text{s.t. } S \in \mathcal{F},\ \mathcal{M} = (V, \mathcal{F}) \text{ is a matroid}, |S| = N, \end{aligned} \tag{1}$$

where the rank of $\mathcal{M}$ is given by $N \in \mathbb{Z}_+$. The guarantees we derive will be applicable as long as the rank of $\mathcal{M}$ is larger than or equal to N. Note that if not, the problem would be infeasible. We can reformulate problems with a matroid rank larger than N as (1), since the intersection of a uniform matroid and any matroid results in another matroid. Finally, let S^* denote the optimal solution of (1). Due to page limitations, we relegated a detailed discussion on three of the well-studied relevant applications the problem above can model to an online appendix in [17]. In each of these applications (task allocation, actuator placement and video summarization), the objective functions are known to be nonsubmodular and nonsupermodular, and bounds on submodularity ratio and curvature can be obtained via Proposition 1 above. The constraints range from partition matroids (in task allocation and summarization) to transversal matroids (in structural controllability), see also the examples in [23].

Algorithm 1 Forward Greedy Algorithm.

Input: Set function f, ground set V, matroid $(V, \mathcal{F})$, cardinality constraint N
Output: Forward greedy solution S^f

```
1: function ForwardGreedy(f, V, F, N)
2:     S^0 = ∅, U^0 = ∅, t = 1
3:     while |S^{t-1}| < N do
4:         j*(t) = argmin_{j ∈ V\U^{t-1}} ρ_j(S^{t-1})
5:         if S^{t-1} ∪ j*(t) ∉ F then U^{t-1} ← U^{t-1} ∪ j*(t)
6:         else
7:             ρ_t ← ρ_{j*(t)}(S^{t-1}) and s_t = j*(t)
8:             S^t ← S^{t-1} ∪ j*(t) and U^t ← U^{t-1} ∪ j*(t)
9:             t ← t + 1
10:        end if
11:    end while
12:    S^f ← S^N
13: end function
```

3. Greedy algorithms

3.1. Forward greedy and the performance guarantee

For Algorithm 1, the following definitions and explanations are in order. Let S^t denote the set at the end of iteration t. At iteration t, the forward greedy algorithm chooses $s_t := S^t \setminus S^{t-1}$ with the corresponding discrete derivative $\rho_t := f(S^t) - f(S^{t-1})$. In Line 5, we have a matroid feasibility check. We assume that this can be done in polynomial-time, which is the case for all the applications we consider, e.g., [12, §5.B&C] proves polynomial-time complexity for the structural controllability matroid. Thanks to the properties of a matroid, we do not need to reconsider an element that has already been rejected by the feasibility check. To this end, the set U^t denotes the set of elements having been considered by the matroid feasibility check before choosing s_{t+1}. The final forward greedy solution is $S^f := S^N$, and it is a base of $(V, \mathcal{F})$, since it lies in $\mathcal{F}$ and has cardinality N by the properties of a matroid.

Main result is shown in the following theorem.

Theorem 1. *If Algorithm 1 is applied to* (1)*, then*

$$\frac{f(S^f) - f(\emptyset)}{f(S^*) - f(\emptyset)} \le \frac{1}{\gamma(1-\alpha)}. \tag{2}$$

We first need the following lemma.

Lemma 1. *For any base $M \in \mathcal{F}$, the elements of $M = \{m_1, \dots, m_N\}$ can be ordered so that $\rho_{m_t}(S^{t-1}) \ge \rho_t = \rho_{s_t}(S^{t-1})$, holds for $t = 1, \dots, N$. Moreover, whenever $s_t \in M$, we have that $m_t = s_t$.*

Proof. We prove by induction. Assume the elements $\{m_{t+1}, \dots, m_N\}$ are already identified. Let $\tilde{M}_t = M \setminus \{m_{t+1}, \dots, m_N\}$. If $s_t \in \tilde{M}_t$, we let $m_t = s_t$, and the inequality condition is satisfied, and the second statement of the lemma holds. If $s_t \notin \tilde{M}_t$, by property *(iii)* of Definition 7 and $\tilde{M}_t \in \mathcal{F}$, we know that there exists $j \in \tilde{M}_t \setminus S^{t-1}$ such that $j \cup S^{t-1} \in \mathcal{F}$. Moreover, $\rho_j(S^{t-1}) \ge \rho_{s_t}(S^{t-1})$, since j is not the element chosen by the greedy algorithm. Hence, we pick $m_t = j$. The base case for our induction proof is as follows. The existence of m_N follows from property *(iii)* of Definition 7: there exists $m_N \in M \setminus S^{N-1}$ such that $m_N \cup S^{N-1} \in \mathcal{F}$. This concludes the proof. □

For this proof, we extended a method from a similar ordering property which is derived when applying greedy heuristics on dependence systems [14, Lem 1] (specifically, on comatroids [13,15,16]). [14, Lem 1] was originally inspired by a study on greedy heuristics over integral polymatroids (that is, a continuous extension of matroids), see [6, Thm 6.1]. This lemma above plays a significant role in obtaining suboptimality bounds, which becomes clear once we pick M to be the optimal solution S^* in the following proof of Theorem 1.

Proof of Theorem 1. Let $S^* = \{s_1^*, \ldots, s_N^*\}$, where the elements s_t^* are ordered according to Lemma 1. Let $S_t^* = \{s_1^*, \ldots, s_t^*\}$ for $t = 1, \ldots, N$, and $S_0^* = \emptyset$. Using this definition, we obtain

$$f(S^*) - f(\emptyset) = \sum_{t=1}^{N} \rho_{s_t^*}(S_{t-1}^*) \geq (1-\alpha) \sum_{t=1}^{N} \rho_{s_t^*}(\emptyset). \quad (3)$$

The equality follows from a telescoping sum. The inequality follows from Definition 6. A similar observation can also be made for the forward greedy solution:

$$f(S^{\mathrm{f}}) - f(\emptyset) = \sum_{t=1}^{N} \rho_t = \sum_{t:s_t \in S^*} \rho_{s_t}(S^{t-1}) + \sum_{t:s_t \notin S^*} \rho_{s_t}(S^{t-1}),$$

where the last equality decomposes the greedy steps into those that coincide with the optimal solution and those that do not. Invoking Lemma 1 for the term on the right-hand side, we obtain

$$f(S^{\mathrm{f}}) - f(\emptyset) \leq \sum_{t:s_t \in S^*} \rho_{s_t}(S^{t-1}) + \sum_{t:s_t^* \notin S^{\mathrm{f}}} \rho_{s_t^*}(S^{t-1}), \quad (4)$$

where the term on the right also utilizes $\{t : s_t \notin S^*\} = \{t : s_t^* \notin S^{\mathrm{f}}\}$, which is a direct result of the last statement of Lemma 1. Now, notice that for any $s_t^* \notin S^{\mathrm{f}}$ we have $s_t^* \notin S^{t-1}$. Hence using Definition 4, we obtain

$$f(S^{\mathrm{f}}) - f(\emptyset) \leq \sum_{t:s_t \in S^*} \rho_{s_t}(S^{t-1}) + \frac{1}{\gamma} \sum_{t:s_t^* \notin S^{\mathrm{f}}} \rho_{s_t^*}(\emptyset).$$

Next, the term on the left can also be upper bounded by utilizing Definition 4, giving us

$$f(S^{\mathrm{f}}) - f(\emptyset) \leq \frac{1}{\gamma} \sum_{t:s_t \in S^*} \rho_{s_t}(\emptyset) + \frac{1}{\gamma} \sum_{t:s_t^* \notin S^{\mathrm{f}}} \rho_{s_t^*}(\emptyset)$$

$$= \frac{1}{\gamma} \sum_{t:s_t \in S^*} \rho_{s_t}(\emptyset) + \frac{1}{\gamma} \sum_{t:s_t \notin S^*} \rho_{s_t^*}(\emptyset) = \frac{1}{\gamma} \sum_{t=1}^{N} \rho_{s_t^*}(\emptyset). \quad (5)$$

The first equality reapplies the set reformulation $\{t : s_t \notin S^*\} = \{t : s_t^* \notin S^{\mathrm{f}}\}$ (previously found in (4)), and the second equality combines the two summations. Finally, we can combine (3) and (5), to obtain $\frac{f(S^{\mathrm{f}}) - f(\emptyset)}{f(S^*) - f(\emptyset)} \leq \frac{1}{\gamma(1-\alpha)}$. □

[12, Thm 2] offers the only guarantee in the literature for this setting involving both submodularity ratio and the curvature. This is given by $\frac{\gamma}{1-\gamma}\left((2N+1)^{\frac{1-\gamma}{\gamma(1-\alpha)}} - 1\right)$. In contrast, the guarantee in Theorem 1 above is independent of the problem size, and also tighter for any of the (α, γ, N) pairs. Moreover, [12, Props 4 and 5] prove that there is no performance guarantee for the forward greedy algorithm unless both submodular-like and supermodular-like properties are present in the objective function. Observe that this is confirmed by Theorem 1, since it grows infinitely large.

As an alternative, [22, Thm 7] offers another guarantee, which is given by $1/(1-c)$, where the strong curvature c quantifies how far a function is from being modular: the smallest parameter $c \in [0, 1]$ such that $\rho_j(R) \geq (1-c)\rho_j(S)$, for all $S, R \subseteq V \setminus j$. This novel notion is a significantly stronger requirement than having both the submodularity ratio and the curvature, simultaneously [12]. Hence, it is not possible to compare it with our guarantee other than the case of a modular objective, that is, $c = 0$, $\gamma = 1$, $\alpha = 0$. For both guarantees, modularity confirms the optimality of the forward greedy algorithm, as it is well-established by the Rado-Edmonds theorem. Note that computing the strong curvature notion requires an exhaustive enumeration of all inequalities in its definition. The proof method of [22, Thm 7] does not allow any greedy strong curvature computation as we present below for the submodularity ratio and the curvature we use.

Corollary 1. *Let γ^{fg} be the largest $\tilde{\gamma}$ that satisfies $\tilde{\gamma}\rho_s(S) \leq \rho_s(\emptyset)$ for all $S \in \mathcal{F}$, $|S| \leq N-1$ and $S \cup s \in \mathcal{F}$. Then, γ^{fg} is called the forward greedy submodularity ratio with $\gamma^{\mathrm{fg}} \geq \gamma$. Let α^{fg} be the smallest $\tilde{\alpha}$ that satisfies $\rho_s(S) \geq (1-\tilde{\alpha})\rho_s(\emptyset)$ for all $S \in \mathcal{F}$, $|S| \leq N-1$ and $S \cup s \in \mathcal{F}$. Then, α^{fg} is called the forward greedy curvature with $\alpha^{\mathrm{fg}} \leq \alpha$.*

The performance guarantee can then be written as

$$\frac{f(S^{\mathrm{f}}) - f(\emptyset)}{f(S^*) - f(\emptyset)} \leq \frac{1}{\gamma^{\mathrm{fg}}(1-\alpha^{\mathrm{fg}})}, \text{ or equivalently,}$$

$$f(S^{\mathrm{f}}) \leq \frac{1}{\gamma^{\mathrm{fg}}(1-\alpha^{\mathrm{fg}})} f(S^*) + \left(1 - \frac{1}{\gamma^{\mathrm{fg}}(1-\alpha^{\mathrm{fg}})}\right) f(\emptyset). \quad (6)$$

The forward greedy submodularity ratio and the forward greedy curvature can be obtained after analyzing $O(\binom{|V|}{N})$ inequalities, which could still be large. However, they are significantly more tractable than the original definitions. Since $\gamma^{\mathrm{fg}} \geq \gamma$ and $\alpha^{\mathrm{fg}} \leq \alpha$, the performance guarantee in Corollary 1 can essentially be better than the one in Theorem 1. Notice that $(\gamma^{\mathrm{fg}}, \alpha^{\mathrm{fg}})$ changes with the constraint set of the problem since the inequalities defining $(\gamma^{\mathrm{fg}}, \alpha^{\mathrm{fg}})$ would then be different. In contrast, submodularity ratio and curvature depend only on the objective function.

The performance guarantee in Corollary 1 can still be loose, because of the reference $f(\emptyset)$. For instance, in task allocation to robots, $f(\emptyset)$ may correspond to minus the safety probability of a plan with no tasks assigned. In such applications, the values $f(\emptyset) \approx -1$ and $(1 - 1/[\gamma^{\mathrm{fg}}(1-\alpha^{\mathrm{fg}})]) < 0$ can make the bound in (6) large.

Next, we consider a greedy variant that comes along with a guarantee that does not depend on $f(\emptyset)$.

3.2. Reverse greedy and the performance guarantee

For Algorithm 2, the following definitions and explanations are in order. For compactness, we define a *shifted* discrete derivative $\delta_j(S) := \rho_j(S \setminus j) = f(S) - f(S \setminus j)$, for all $S \subseteq V$, $j \in S$. Let X^t denote the set at the end of iteration t. At iteration t, the reverse greedy algorithm chooses $r_t := X^{t-1} \setminus X^t$ to remove, with the maximal reduction $\delta_t := f(X^{t-1}) - f(X^t)$. In Line 5, we have a matroid feasibility check. In contrast to Algorithm 1, this matroid feasibility check requires that our intermediate solutions remain always as supersets of a base of the matroid $\mathcal{M}$. This follows since the algorithm starts with a large set, while our final goal is to obtain a base of the matroid. The set Y^t denotes the set of elements having been considered by the matroid feasibility check before choosing x_{t+1}. The final reverse greedy solution is $S^{\mathrm{r}} := X^{|V|-N}$, and it is a base of $(V, \mathcal{F})$, since it lies in $\mathcal{F}$ and has cardinality N by the properties of a matroid.

Main result is shown in the following theorem.

Algorithm 2 Reverse Greedy Algorithm.

Input: Set function f, ground set V, matroid $(V,\mathcal{F})$, cardinality constraint N
Output: Reverse greedy solution S^r
function ReverseGreedy$(f,V,\mathcal{F},N)$
  $X^0=V$, $Y^0=\emptyset$, $t=1$
  while $|X^{t-1}|>N$ **do**
    $k^*(t)=\arg\max_{j\in V\setminus Y^{t-1}}\delta_j(X^{t-1})$
    if $\exists M\in\mathcal{F}$ such that $M\subseteq\{X^{t-1}\setminus k^*(t)\}$ and $|M|=N$ **then** $Y^{t-1}\leftarrow Y^{t-1}\cup k^*(t)$
    else
      $\delta_t\leftarrow\delta_{k^*(t)}(X^{t-1})$ and $r_t=k^*(t)$
      $X^t\leftarrow X^{t-1}\setminus k^*(t)$ and $Y^t\leftarrow Y^{t-1}\cup k^*(t)$
      $t\leftarrow t+1$
    end if
  end while
  $S^r\leftarrow X^{|V|-N}$
end function

Algorithm 3 Forward Greedy Reformulation of Reverse Greedy Algorithm.

Input: Set function $\hat{f}$, ground set V, dual matroid $(V,\hat{\mathcal{F}})$, cardinality constraint $\hat{N}$
Output: Reverse greedy solution S^r
function ReverseGreedyReformulated$(f,V,\hat{\mathcal{F}},\hat{N})$
  $R^0=\emptyset$, $Y^0=\emptyset$, $t=1$
  while $|R^{t-1}|<\hat{N}$ **do**
    $k^*(t)=\arg\max_{j\in V\setminus Y^{t-1}}\hat{\rho}_j(R^{t-1})$
    if $R^{t-1}\cup k^*(t)\notin\hat{\mathcal{F}}$ **then** $Y^{t-1}\leftarrow Y^{t-1}\cup k^*(t)$
    else
      $\hat{\rho}_t\leftarrow\hat{\rho}_{k^*(t)}(R^{t-1})$ and $r_t=k^*(t)$
      $R^t\leftarrow R^{t-1}\cup k^*(t)$ and $Y^t\leftarrow Y^{t-1}\cup k^*(t)$
      $t\leftarrow t+1$
    end if
  end while
  $S^r\leftarrow V\setminus R^{\hat{N}}$
end function

Theorem 2. *If Algorithm 2 is applied to (1), then*

$$\frac{f(V)-f(S^r)}{f(V)-f(S^*)}\geq\frac{1-\alpha}{1+(1-\gamma)(1-\alpha)}.$$

For the sake of clarity of the notation, we now define $\hat{f}(R):=-f(V\setminus R)$, for all $R\subseteq V$. Our proof will utilize the following reformulation of (1):

$$\begin{aligned}&\max_{R\subseteq V}\hat{f}(R),\ \text{increas.},\ (1-\alpha)\text{-submod.},\ (1-\gamma)\text{-supermod.}\\&\text{s.t. } R\in\hat{\mathcal{F}},\ \hat{\mathcal{M}}=(V,\mathcal{F})\text{ is a matroid}, |R|=|V|-N=\hat{N},\end{aligned}\tag{7}$$

where the cardinality of any base of $\hat{\mathcal{M}}$ is given by $|V|-N=\hat{N}\in\mathbb{Z}_+$. The equivalence of (1) and (7) follows directly from Proposition 2 and Definition 8 (that is, the definition of the dual matroid). Denote its optimal solution by R^*. Clearly, we have $R^*=V\setminus S^*$.

Forward greedy algorithm applied to (7) is presented in Algorithm 3, where we define $\hat{\rho}_j(R)=\hat{f}(R\cup j)-\hat{f}(R)$, for all $R\subseteq V$ and $j\in V$. Assuming uniqueness, iterations of this algorithm coincide with those of the reverse greedy algorithm applied to (1), which was presented earlier in Algorithm 2. Denote the forward greedy iterates by R^t. At any iteration, we have $X^t=V\setminus R^t$. At iteration t, the forward greedy algorithm chooses $r_t:=R^t\setminus R^{t-1}$ with the corresponding discrete derivative $\hat{\rho}_t:=\hat{f}(S^t)-\hat{f}(S^{t-1})$. With the observations above, we now bring the ordering lemma.

Lemma 2. *For any base $M\in\hat{\mathcal{F}}$, the elements of $M=\{m_1,\dots,m_{\hat{N}}\}$ can be ordered so that $\hat{\rho}_{m_t}(R^{t-1})\leq\hat{\rho}_t=\hat{\rho}_{r_t}(R^{t-1})$, holds for $t=1,\dots,\hat{N}$. Moreover, whenever $r_t\in M$, we have that $m_t=r_t$.*

The proof of the lemma above involves steps similar to those found in Lemma 1, and hence, it is relegated to our online appendix in [17]. Note that in contrast to the similarities found in the proofs of Lemmas 1 and 2, the proof below for our theorem is completely different from the proof of Theorem 1. This follows from greedy minimization and maximization distinction in the steps of Algorithms 1 and 3, respectively. We show below that this distinction requires us to find different lower and upper bound terms. The terms below involve slightly more complicated steps by considering the objective function evaluated at the union: $R^{\hat{N}}\cup R^*$.

Proof of Theorem 2. Let $R^*=\{r_1^*,\dots,r_{\hat{N}}^*\}$, where the elements r_t^* are ordered according to Lemma 2. Let $R_t^*=\{r_1^*,\dots,r_t^*\}$ for $t=1,\dots,\hat{N}$, and $R_0^*=\emptyset$. Using this definition, we obtain

$$\begin{aligned}\hat{f}(R^{\hat{N}}\cup R^*)-\hat{f}(\emptyset)&=\hat{f}(R^{\hat{N}})-\hat{f}(\emptyset)+\sum_{t=1}^{\hat{N}}\hat{\rho}_{r_t^*}(R^{\hat{N}}\cup R_{t-1}^*)\\&=\hat{f}(R^{\hat{N}})-\hat{f}(\emptyset)+\sum_{t:r_t^*\notin R^{\hat{N}}}\hat{\rho}_{r_t^*}(R^{\hat{N}}\cup R_{t-1}^*)\\&\leq\hat{f}(R^{\hat{N}})-\hat{f}(\emptyset)+\frac{1}{1-\alpha}\sum_{t:r_t^*\notin R^{\hat{N}}}\hat{\rho}_{r_t^*}(R^{t-1})\\&\leq\hat{f}(R^{\hat{N}})-\hat{f}(\emptyset)+\frac{1}{1-\alpha}\sum_{t:r_t^*\notin R^{\hat{N}}}\hat{\rho}_{r_t}(R^{t-1}).\end{aligned}\tag{8}$$

The first equality follows from a telescoping sum, whereas the second equality removes the zero-valued terms. The first inequality follows from the submodularity ratio of $\hat{f}$ (which is $(1-\alpha)$), whereas the second inequality follows from Lemma 2.

A similar observation can be made to obtain a lower bound to the term $[\hat{f}(R^{\hat{N}}\cup R^*)-\hat{f}(\emptyset)]$ above as follows

$$\begin{aligned}&\hat{f}(R^{\hat{N}}\cup R^*)-\hat{f}(\emptyset)=\hat{f}(R^*)-\hat{f}(\emptyset)+\sum_{t=1}^{\hat{N}}\hat{\rho}_{r_t}(R^{t-1}\cup R^*)\\&\geq\hat{f}(R^*)-\hat{f}(\emptyset)+\gamma\sum_{t=1}^{\hat{N}}\hat{\rho}_{r_t}(R^{t-1})-\gamma\sum_{t:r_t\in R^*}\hat{\rho}_{r_t}(R^{t-1})\\&=\hat{f}(R^*)-\hat{f}(\emptyset)+\gamma\left[\hat{f}(R^{\hat{N}})-\hat{f}(\emptyset)\right]-\gamma\sum_{t:r_t\in R^*}\hat{\rho}_{r_t}(R^{t-1})\\&=\hat{f}(R^*)-\hat{f}(\emptyset)+\gamma\left[\hat{f}(R^{\hat{N}})-\hat{f}(\emptyset)\right]-\gamma\sum_{t:r_t^*\in R^{\hat{N}}}\hat{\rho}_{r_t}(R^{t-1}).\end{aligned}\tag{9}$$

The first equality follows from a telescoping sum. The inequality follows from the application of curvature of $\hat{f}$ (which is $(1-\gamma)$) together with the fact that some of the terms in the sum $\sum_{t=1}^{\hat{N}}\hat{\rho}_{r_t}(R^{t-1}\cup R^*)$ are zero whenever $r_t\in R^*$. The second equality sums up the terms in the telescoping sum, whereas the third applies $\{t:r_t\in R^*\}=\{t:r_t^*\in R^{\hat{N}}\}$, invoking the last part of Lemma 2.

Now, combining (9) and (8), we obtain

$$\begin{aligned}&(1-\gamma)\left[\hat{f}(R^{\hat{N}})-\hat{f}(\emptyset)\right]\\&\geq\hat{f}(R^*)-\hat{f}(\emptyset)-\frac{1}{1-\alpha}\sum_{t:r_t^*\notin R^{\hat{N}}}\hat{\rho}_{r_t}(R^{t-1})-\gamma\sum_{t:r_t^*\in R^{\hat{N}}}\hat{\rho}_{r_t}(R^{t-1})\\&\geq\hat{f}(R^*)-\hat{f}(\emptyset)-\frac{1}{1-\alpha}\sum_{t=1}^{\hat{N}}\hat{\rho}_{r_t}(R^{t-1})\\&=\hat{f}(R^*)-\hat{f}(\emptyset)-\frac{1}{1-\alpha}\left[\hat{f}(R^{\hat{N}})-\hat{f}(\emptyset)\right].\end{aligned}$$

(10)

The first inequality presents only the combination of (9) and (8). Observing that $(1/1-\alpha) \geq \gamma$ for any (α, γ) pair, the second inequality combines the two sums: $\{1, \ldots, \hat{N}\} = \{t : r_t^* \notin R^{\hat{N}}\} \cup \{t : r_t^* \in R^{\hat{N}}\}$. The last step sums all the terms involved in the telescoping sum.

By reorganizing (10), we get

$$\frac{\hat{f}(R^{\hat{N}}) - \hat{f}(\emptyset)}{\hat{f}(R^*) - \hat{f}(\emptyset)} \geq \frac{1-\alpha}{1+(1-\gamma)(1-\alpha)}.$$

From the equivalence of the two problems and Algorithms 2 and 3, we complete the proof:

$$\frac{f(V) - f(S^{\mathrm{r}})}{f(V) - f(S^*)} = \frac{\hat{f}(R^{\hat{N}}) - \hat{f}(\emptyset)}{\hat{f}(R^*) - \hat{f}(\emptyset)} \geq \frac{1-\alpha}{1+(1-\gamma)(1-\alpha)}. \qquad \square$$

[22, Thm 6] offers a guarantee for this setting. This is given by $1-c$, where c is again the strong curvature as in [22, Thm 7] discussed in Section 3.1. Here, arguments similar to the case of the forward greedy can be made both on the strength of this requirement and its computational aspect. Observe that for the case of a modular objective, both guarantees confirm the optimality of the reverse greedy on the base of a matroid.

As an alternative, [2, Thm 1] offers another guarantee for this setting if the constraint is a uniform matroid, that is, only a cardinality constraint. This is given by $(1/(1-\gamma))\left(1 - e^{-(1-\alpha)(1-\gamma)}\right)$. This guarantee is tighter than the one in Theorem 2, since it treats a specialized case. However, when we have exact submodularity $\gamma = 1$, both guarantees still coincide $\lim_{\gamma \to 1}(1/(1-\gamma))\left(1 - e^{-(1-\alpha)(1-\gamma)}\right) = 1-\alpha$. Moreover, both guarantees tend to 0 as $\alpha \to 1$ independent of γ, in other words, when supermodular-like properties are not present at all. We highlight that the proof method in [2] is not applicable to general matroids.

When $\gamma = 0$, $\alpha = 0$, we recover the classical 1/2 guarantee of [9], since setting $\gamma = 0$ can be considered to be the case when the submodularity property of the objective is completely unknown. For the same setting without the submodularity ratio knowledge, the guarantee from [10, Thm 6] is also applicable, and it is given by $0.4(1-\alpha)^2/(1+(1-\alpha)^{(\hat{N}/2)})$. In contrast, our guarantee is $(1-\alpha)/(2-\alpha)$, which is tighter for all α.

Corollary 2. *Let γ^{rg} be the largest $\tilde{\gamma}$ that satisfies $\tilde{\gamma}\hat{\rho}_{r_t}(R^{t-1}) \leq \hat{\rho}_{r_t}(R^{t-1} \cup R)$ for all t, for all $R \subset V \setminus r_t$ and $|R| = \hat{N}$. (We remind the reader that $\hat{\rho}_j(R) = \hat{f}(R \cup j) - \hat{f}(R) = f(V \setminus R) - f(\{V \setminus R\} \setminus j)$, for all $R \subseteq V$ and $j \in V$.) Then, γ^{rg} is called the reverse greedy submodularity ratio with $\gamma^{\mathrm{rg}} \geq \gamma$. Let α^{rg} be the smallest $\tilde{\alpha}$ that satisfies $\hat{\rho}_r(R^{t-1}) \geq (1-\tilde{\alpha})\hat{\rho}_r(R^{\hat{N}} \cup R)$ for all t, for all $R \subset V$ and $|R| = t-1$, for all $r \notin R^{\hat{N}} \cup R$. Then, α^{rg} is called the reverse greedy curvature with $\alpha^{\mathrm{rg}} \leq \alpha$.*

The performance guarantee can then be written as

$$\frac{f(V) - f(S^{\mathrm{r}})}{f(V) - f(S^*)} \geq \frac{1-\alpha^{\mathrm{rg}}}{1+(1-\gamma^{\mathrm{rg}})(1-\alpha^{\mathrm{rg}})},$$

$$f(S^{\mathrm{r}}) \leq \frac{1-\alpha^{\mathrm{rg}}}{1+(1-\gamma^{\mathrm{rg}})(1-\alpha^{\mathrm{rg}})} f(S^*) + \left(1 - \frac{1-\alpha^{\mathrm{rg}}}{1+(1-\gamma^{\mathrm{rg}})(1-\alpha^{\mathrm{rg}})}\right) f(V).$$

The reverse greedy submodularity ratio and the reverse greedy curvature can be obtained in an ex-post manner after analyzing $O(\hat{N}\binom{|V|-1}{\hat{N}})$ and $O(\hat{N}\binom{|V|}{\hat{N}})$ inequalities, respectively. Moreover, since $\gamma^{\mathrm{rg}} \geq \gamma$ and $\alpha^{\mathrm{rg}} \leq \alpha$, the performance guarantee in Corollary 2 can essentially be better than the one in Theorem 2. Finally, note that a large value for $f(V)$ can affect the tightness of this guarantee. For instance, in the actuator removal application discussed in [12], when we pick the full set of actuators V to remove from the networked system, the control energy metric $f(V)$ could be infinitely large.

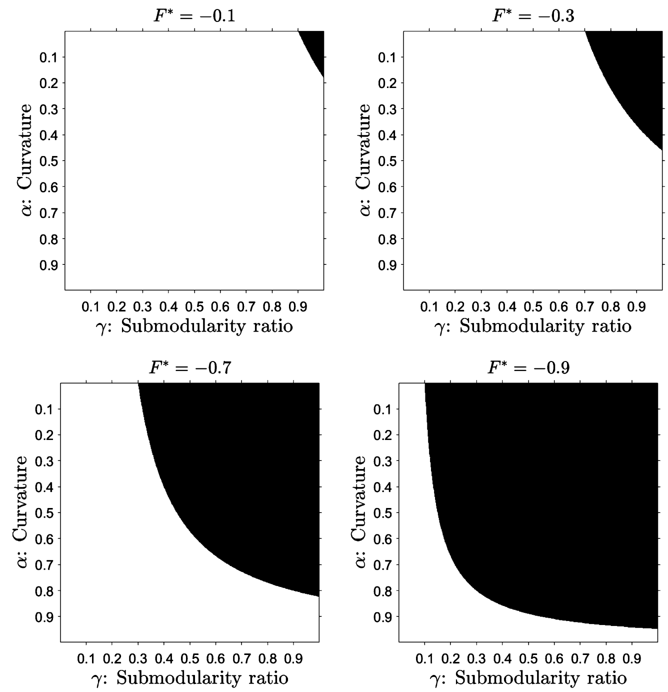

Fig. 1. For each F^* value, the shaded regions represent the (α, γ) pairs for which the forward greedy algorithm outperforms the reverse greedy algorithm. For $F^* \geq 0$, the reverse greedy algorithm outperforms the forward greedy algorithm for any (α, γ) pair.

4. Comparison of the performance guarantees

For the sake of visualization, we let $f(\emptyset) = -1$, $f(V) = 1$, $f(S^*) = F^* \in [-1, 1]$. Fig. 1 shades the area where the guarantee for the forward greedy is better than the one for the reverse greedy. By decreasing the value of F^* from 0 to -1, one can observe that the area where the forward greedy is better, expands. When F^* is small, and the function is close to being both submodular and supermodular, the forward greedy guarantee is more desirable. In fact, [12, Props 4 and 5] prove that there is no performance guarantee for the forward greedy unless both submodularity ratio and curvature are utilized, simultaneously. When F^* is large enough, $F^* \geq 0$, the effect of $f(\emptyset)$ on the forward greedy guarantee is more dominant, thus, the reverse greedy outperforms the forward greedy for all (α, γ) pairs. One can scale the comparison for different $f(\emptyset)$ and $f(V)$ values. In practice, it could be useful to implement both greedy algorithms (which can be done efficiently with polynomial time complexity) and choose the best out of the two. Our future work is focused on obtaining the problem instances for which these guarantees are potentially tight.

Acknowledgements

The authors would like to thank Prof. Victor Il'ev for his helpful feedback on existing bounds for greedy algorithms, Dr. Yatao An Bian for discussions on the literature for properties of set functions.

References

[1] F. Bach, Learning with submodular functions: a convex optimization perspective, Found. Trends Mach. Learn. 6 (2–3) (2013) 145–373.

[2] A.A. Bian, J.M. Buhmann, A. Krause, S. Tschiatschek, Guarantees for greedy maximization of non-submodular functions with applications, in: 34th ICML, 2017, pp. 498–507.

[3] L. Chen, M. Feldman, A. Karbasi, Weakly submodular maximization beyond cardinality constraints: does randomization help greedy?, in: 35th ICML, 2018, pp. 804–813.

[4] M. Chrobak, C. Kenyon, N. Young, The reverse greedy algorithm for the metric k-median problem, Inf. Process. Lett. 97 (2) (2006) 68–72.

[5] A. Clark, B. Alomair, L. Bushnell, R. Poovendran, Submodularity in input node selection for networked linear systems: efficient algorithms for performance and controllability, IEEE Control Syst. Mag. 37 (6) (2017) 52–74.

[6] M. Conforti, G. Cornuéjols, Submodular set functions, matroids and the greedy algorithm: tight worst-case bounds and some generalizations of the Rado-Edmonds theorem, Discrete Appl. Math. 7 (3) (1984) 251–274.

[7] A. Das, D. Kempe, Submodular meets spectral: greedy algorithms for subset selection, sparse approximation and dictionary selection, in: 28th ICML, 2011, pp. 1057–1064.

[8] J. Edmonds, Matroids and the greedy algorithm, Math. Program. 1 (1) (1971) 127–136.

[9] M.L. Fisher, G.L. Nemhauser, L.A. Wolsey, An analysis of approximations for maximizing submodular set functions—ii, in: Polyh. Comb., Springer, 1978, pp. 73–87.

[10] K. Gatmiry, M. Gomez-Rodriguez, Non-submodular function maximization subject to a matroid constraint, with applications, arXiv:1811.07863, 2018.

[11] B.P. Gerkey, M.J. Matarić, A formal analysis and taxonomy of task allocation in multi-robot systems, Int. J. Robot. Res. 23 (9) (2004) 939–954.

[12] B. Guo, O. Karaca, T. Summers, M. Kamgarpour, Actuator placement under structural controllability using forward and reverse greedy algorithms, IEEE Trans. Autom. Control (2021).

[13] V. Il'ev, Hereditary systems and greedy-type algorithms, Discrete Appl. Math. 132 (1–3) (2003) 137–148.

[14] V. Il'ev, S. Il'eva, On minimizing supermodular functions on hereditary systems, in: Opt. Prob. and Their Appl., ISBN 978-3-319-93800-4, 2018, pp. 3–15.

[15] V. Il'ev, N. Linker, Performance guarantees of a greedy algorithm for minimizing a supermodular set function on comatroid, Eur. J. Oper. Res. 171 (2) (2006) 648–660.

[16] O. Karaca, B. Guo, M. Kamgarpour, A comment on performance guarantees of a greedy algorithm for minimizing a supermodular set function on comatroid, Eur. J. Oper. Res. 290 (1) (2020) 401–403.

[17] O. Karaca, D. Tihanyi, M. Kamgarpour, Performance guarantees of forward and reverse greedy algorithms for minimizing nonsupermodular nonsubmodular functions on a matroid, Online Appendix, arXiv:2103.01135, 2021.

[18] A. Krause, D. Golovin, Submodular function maximization, Tractability 3 (01 2011) 71–104, https://doi.org/10.1017/CBO9781139177801.004.

[19] B. Lehmann, D. Lehmann, N. Nisan, Combinatorial auctions with decreasing marginal utilities, Games Econ. Behav. 55 (01 2001) 270–296, https://doi.org/10.1016/j.geb.2005.02.006.

[20] G.L. Nemhauser, L.A. Wolsey, M.L. Fisher, An analysis of approximations for maximizing submodular set functions—i, Math. Program. 14 (1) (1978) 265–294.

[21] S. Qian, Y. Singer, Fast parallel algorithms for statistical subset selection problems, in: Pro. of the 33rd NeurIPS, 2019, pp. 5072–5081.

[22] M. Sviridenko, J. Vondrák, J. Ward, Optimal approximation for submodular and supermodular optimization with bounded curvature, Math. Oper. Res. 42 (4) (2017) 1197–1218.

[23] V. Tzoumas, A. Jadbabaie, G.J. Pappas, Resilient non-submodular maximization over matroid constraints, arXiv preprint arXiv:1804.01013, 2018.

[24] D. Welsh, Matroid theory, in: Dover Books on Mathematics, Dover Publications, ISBN 9780486474397, 2010, https://books.google.ch/books?id=QL2iYMBLpFwC.

[25] L.A. Wolsey, G.L. Nemhauser, Integer and Combinatorial Optimization, vol. 55, John Wiley & Sons, 1988.