Greedy Maximization of Functions with Bounded Curvature under Partition Matroid Constraints

Tobias Friedrich[1], Andreas Göbel[1], Frank Neumann[2], Francesco Quinzan[1] and Ralf Rothenberger[1]
[1]Chair of Algorithm Engineering, Hasso Plattner Institute, Potsdam, Germany
[2]Optimisation and Logistics, School of Computer Science, The University of Adelaide, Adelaide, Australia

Abstract

We investigate the performance of a deterministic GREEDY algorithm for the problem of maximizing functions under a partition matroid constraint. We consider non-monotone submodular functions and monotone subadditive functions. Even though constrained maximization problems of monotone submodular functions have been extensively studied, little is known about greedy maximization of non-monotone submodular functions or monotone subadditive functions.

We give approximation guarantees for GREEDY on these problems, in terms of the *curvature*. We find that this simple heuristic yields a strong approximation guarantee on a broad class of functions.

We discuss the applicability of our results to three real-world problems: Maximizing the determinant function of a positive semidefinite matrix, and related problems such as the maximum entropy sampling problem, the constrained maximum cut problem on directed graphs, and combinatorial auction games.

We conclude that GREEDY is well-suited to approach these problems. Overall, we present evidence to support the idea that, when dealing with constrained maximization problems with bounded curvature, one needs not search for (approximate) monotonicity to get good approximate solutions.

Introduction

Submodular functions capture the notion of diminishing returns, i.e. the more you acquire the less your marginal gain will be. This notion occurs frequently in the real world, thus, the problem of maximizing a submodular function finds applicability in a plethora of scenarios. Examples of such scenarios include: maximum cut problems (Goemans and Williamson 1995), combinatorial auctions (Maehara et al. 2017), facility location (Cornuejols, Fisher, and Nemhauser 1977), problems in machine learning (Elenberg et al. 2017), coverage functions (Krause, Singh, and Guestrin 2008), online shopping (Tschiatschek, Singla, and Krause 2017). As such, the literature on submodular functions contains a vast number of results spanning over three decades.

Formally, a set function $f\colon 2^V \to \mathbb{R}$ is *submodular* if for all $U, W \subseteq V$, $f(U) + f(W) \geq f(U \cup W) + f(U \cap W)$. As these functions come from a variety of applications, in this work we will assume that, given a set $U \subseteq V$, the value $f(U)$ is returned from an oracle. This is a reasonable assumption as in most applications $f(U)$ can be computed efficiently. Often in these applications, a realistic solution is subject to some constraints. Among the most common constraints include Matroid and Knapsack constraints —see (Lee et al. 2009). From these families of constraints the most natural and common type of constraints are uniform matroid constraints also known as cardinality constraints. Optimizing a submodular function given k as a cardinality constraint is to finding a set U, with $|U| \leq k$, that maximizes $f(U)$. In this paper we consider submodular maximization under partition matroid constraints. These constraints are in the intersection of matroid and knapsack constaints and generalize uniform matroid constraints. In partition matroid constraints we are given a collection $B_1, \dots, B_k$ of disjoint subsets of V, integers $d_1 \dots d_k$. Every feasible solution to our problem must then include at most d_i elements from each set B_i. Submodular maximization under partition matroid constraints are considered in various applications, e.g. see (Lin and Bilmes 2010; Jegelka and Bilmes 2011).

As the literature in submodular optimization is immense, we will only review the results on submodular maximization under matroid and knapsack constraints. The classical result of (Cornuejols, Fisher, and Nemhauser 1977) shows that a greedy algorithm achieves a $1/2$ approximation ration when maximizing monotone submodular functions under partition matroid constraints. (Nemhauser and Wolsey 1978) showed that no-polynomial time algorithm can achieve a better approximation ratio than $(1 - 1/e)$. Many years later (Călinescu et al. 2011) where able to achieve this upper bound using a randomized algorithm. Recently (Buchbinder, Feldman, and Garg 2018) achieved a deterministic 0.5008-approximation ratio by derandomizing search heuristics.

The previous approximation ratios can be further improved when assuming that the rate of change of the marginal values of f is bounded. This is expressed by the curvature α of a function as in Definition 1. The results of (Conforti and Cornuéjols 1984; Vondrák 2010) show that a continuous greedy algorithm gives a $\frac{1}{\alpha}(1 - e^{-\alpha})$ approximation when maximizing a monotone submodular function under any matroid constraint. Thus, when $\alpha \leq 1.58933$ the the continuous greedy outperforms the algorithm of (Buchbinder, Feldman, and Garg 2018) and when $\alpha \leq 1$ the continuous greedy outperforms the algorithm of (Călinescu et al. 2011). Finally, (Bian et al. 2017) show that the deterministic greedy algorithm achieves a $\frac{1}{\alpha}(1-e^{-\alpha})$ approximation when maximizing submodular monotone functions of curvature α, but only under cardinality constraints.

All of the aforementioned approximation results rely on the fact that f is monotone. In practice submodular functions such as maximum cut, combinatorial auctions, sensor placement and experimental design the functions need not be monotone. To solve such problems using simple greedy algorithms, often assumptions are made that the function f is monotone or that f is under some sense "close" to being monotone. Practical problems that are solved using greedy algorithms under such assumptions can be found in many articles such as (Bian et al. 2017; Das and Kempe 2011; Lawrence, Seeger, and Herbrich 2002; Singh et al. 2009).

In this article we show that the greedy algorithm finds a $\frac{1}{\alpha}(1-e^{-\alpha})$-approximation in $\mathcal{O}(dn)$ oracle evaluations, for the problem of maximizing a submodular function subject to partition matroid constraints (Theorem 1). Our result extends previous known approximation ratios on monotone submodular functions to all submodular functions under partition matroid constraints.

Additionally, we extend the results on monotone submodular functions to another direction, to the class of monotone subadditive functions. Subadditivity is a natural property assumed to hold for functions evaluating items sold in combinatorial acutions (Bhawalkar and Roughgarden 2011; Assadi 2017). Formally, we say that a set function $f\colon 2^V \rightarrow \mathbb{R}$ is *subadditive* if for all $U, W \subseteq V$, $f(U) + f(W) \geq f(U \cup W)$. We show (Theorem 2) that the greedy algorithm achieves a $\frac{1}{\alpha}(1-e^{\alpha^2-\alpha})$ approximation ratio when optimizing monotone subadditive functions with curvature α under partition matroid constraints.

We motivate our results by considering three real world applications. The first application we consider is to maximize the logarithm of determinant functions. In this setting we are given a matrix $\mathcal{P}$ and we want to find the submatrix $\mathcal{A}$ of $\mathcal{D}$ with the largest determinant, where $\mathcal{A}$ satisfies matroid partition constraints. This problem appears in a variety of real world settings. In this article, as a real world example of this application, we compute the sensor (thermometer) placement across the world that maximizes entropy, subject to a cardinality constraint and subject to a partition matroid constrain where the partitions of the data sets are countries.

Our second application is the problem of finding the maximum directed cut of a graph, under partition matroid constraints. The cut function of a graph is known to be submodular and non-monotone in general (Feige, Mirrokni, and Vondrák 2011). We show how to bound the curvature of the cut function with respect to the maximum degree. We also run experiments on this setting, showing that in most graphs of our dataset the deterministic greedy algorithm finds the actual optimal solution. Thus GREEDY seems to perform well on non-monotone submodular functions in practice.

Finally, the third application is computing the social welfare of a subadditive combinatorial auction. We show that the social welfare is also a subadditive function and its curvature is bounded by the maximum curvature of the utility functions.

Preliminary Definitions and Algorithms

Problem description.

Let $f\colon 2^V \longrightarrow \mathbb{R}_{\geq 0}$ be a non-negative function over a set V of size n, let $B_1, \dots, B_k$ be a collection of disjoint subsets of V, and let d_i integers s.t. $1 \leq d_i \leq |B_i|$, $\forall i \in [k]$. We consider the maximization problem

$$\max_{S \subseteq V} \{f(S) : |S \cap B_i| \leq d_i,\ \forall i \in [k]\}. \tag{1}$$

Note that the problem of maximizing f under a cardinality constrain is a special case of the above, where $k = 1$ and $B_1 = V$.

We evaluate the quality of an approximation of a global maximum as follows. Let $U \subseteq V$ be a feasible solution to Problem (1). We say that U is an ε-approximation if $f(\text{OPT})/f(U) \geq \varepsilon$, where OPT is the optimal solution set. We often refer to the value $f(U)$ as the f-value of U.

In this paper, we evaluate run time in the black-box oracle model: We assume that there exists an oracle that returns the corresponding f-value of a solution candidate, and we estimate the run time, by counting the total number of calls to the valuation oracle.

To simplify the exposition, throughout our analyses, we always assume that the following reduction holds.

Reduction 1. *For Problem* (1) *we may assume $cd_i \leq |B_i|$ for all $i = 1, \dots, k$, for an arbitrary constant $c > 0$. Moreover, we may assume that there exists a set $D_i \subseteq B_i$ of size d_i s.t. $f(S) = f(S \setminus D_i)$ for all $S \subseteq V$, for all $i = 1, \dots, k$.*

Algorithms.

GREEDY is the simple discrete greedy algorithm that appears in Algorithm 1. Starting with the empty set, GREEDY iteratively adds points that maximize the marginal values with respect to the already found solution. This algorithm is a mild generalization of the simple deterministic greedy due to Nemhauser and Wolsey (Nemhauser and Wolsey 1978).

Notation.

For any non-negative function $f\colon 2^V \longrightarrow \mathbb{R}_{\geq 0}$ and any two subsets $S, \Omega \subseteq V$, we define the *marginal value* of S with respect to Ω as $\rho_\Omega(S) = f(S \cup \Omega) - f(S)$.

We denote with $B_1, \dots, B_k$ disjoint subsets of V and with $d_1, \dots, d_k$ their respective sizes, as in the problem description section. We denote with d the sum $\sum_{j=1}^{k} d_j$. We denote with D the subset of "dummy" elements as in Reduction 1, and we denote with OPT any solution to Problem (1), such that $\text{OPT} \cap D = \emptyset$.

We let S_t be a solution found by GREEDY at time step t and we denote with ρ_t the marginal value $\rho_t = f(S_t) - f(S_{t-1})$. We use the convention $\rho_0 = f(\emptyset)$.

Curvature

In this paper we give approximation guarantees in terms of the *curvature*. Intuitively, the curvature is a parameter that bounds the maximum rate with which a function changes.

Algorithm 1: The GREEDY algorithm.

input: a function $f: V \longrightarrow \mathbb{R}_{\geq 0}$;
    disjoint subsets $B_1, \dots, B_k \subseteq V$;
    integers $d_1, \dots, d_k$ s.t. $0 \leq d_i \leq |B_i|$, $\forall i \in [k]$;
output: an approximate global maximum S of f s.t. $|S \cap B_i| \leq d_i$, $\forall i \in [k]$;
$S \leftarrow \emptyset$;
while $|S| \leq \sum_{i=1}^{k} d_i$ **do**
    let $\omega \in V$ maximizing $f(S \cup \{\omega\}) - f(S)$ and s.t. $|(S \cup \{\omega\}) \cap B_i| \leq d_i$, $\forall i \in [k]$;
    $S \leftarrow S \cup \{\omega\}$;
return S;

As our functions f map sets to positive reals, i.e. $f: 2^V \longrightarrow \mathbb{R}_{\geq 0}$, we say that f has curvature α if the value $f(S \cup \{\omega\}) - f(S)$ does not change by a factor larger than $1 - \alpha$ when varying S. This parameter was first introduced by (Conforti and Cornuéjols 1984) to beat the $(1 - e^{-1})$-approximation barrier of monotone submodular functions. Formally we use the following definition of curvature, relaxing the definition of *greedy* curvature (Bian et al. 2017).

Definition 1 (Curvature)**.** *Consider a non-negative function $f: 2^V \longrightarrow \mathbb{R}$ as in Problem* (1)*. The curvature is the smallest scalar α s.t.*

$$\rho_\omega((S \cup \Omega) \setminus \{\omega\}) \geq (1 - \alpha)\rho_\omega(S \setminus \{\omega\}),$$

for all $S, \Omega \subseteq V$ of size at most $\sum_i d_i$ and $\omega \in S \setminus \Omega$.

Note that $\alpha \geq 0$. We say that a function f has *positive curvature* if $\alpha \leq 1$. We further observe that the curvature of a monotone function is always $\alpha \leq 1$. Hence, all functions with negative curvature are non-monotone. We remark that the curvature is invariant under multiplication by a positive scalar. In other words, if a function f has curvature α, then any function cf has curvature α, for all $c > 0$. Moreover, the following simple result holds.

Proposition 1. *Let $f, g: 2^V \longrightarrow \mathbb{R}_{\geq 0}$ be non-negative functions with curvature α_1, α_2 respectively. Then the curvature α of the function $f + g$ is upper-bounded as $\alpha \leq \sup_i \alpha_i$.*

In the case of a submodular function, it is possible to give a simple characterization of Definition 1. In fact, one can easily prove the following.

Proposition 2. *Let $f: 2^V \longrightarrow \mathbb{R}_{\geq 0}$ be a submodular function with curvature α, as in Problem* (1)*. Then,*

$$\alpha \leq 1 - \min_{\{S \subseteq V,\, \omega \in S\}} \frac{f(S) - f(S \setminus \{\omega\})}{f(\omega) - f(\emptyset)},$$

for all subsets $S \subseteq V$ s.t. $|S| \leq \sum_i d_i$

Proposition 2 gives the original definition of curvature for submodular functions (Conforti and Cornuéjols 1984). However, the inequality of Proposition 2 does not necessarily hold for non-submodular functions.

Approximation Guarantees

We give approximation guarantees for GREEDY on Problem (1), when optimizing a (non-monotone) submodular function with bounded curvature α. Our proof technique generalizes the results of (Conforti and Cornuéjols 1984) to non-monotone functions f by utilizing the notion of curvature. We have the following theorem.

Theorem 1. *Let f be a submodular function with curvature α.* GREEDY *is a $\frac{1}{\alpha}(1 - e^{-\alpha})$-approximation algorithm for Problem* (1) *with run-time in $\mathcal{O}(dn)$.*

Note that if f is monotone, then our approximation guarantee matches the approximation guarantee of Conforti and Cornuéjols, which is known to be nearly optimal (Conforti and Cornuéjols 1984; Vondrák 2010). Furthermore, in the non-monotone case our lower-bound may yield significant improvement over state-of-the-art known bounds (Buchbinder et al. 2014; Buchbinder and Feldman 2018). Particularly, we beat the $1/e$-approximation barrier on functions with curvature $\alpha \leq 2.49375$ and the $1/2$-approximation barrier on functions with curvature $\alpha \leq 1.59362$.

Observation 1. *There exists a submodular function f with curvature $\alpha = n$, such that* GREEDY *does not find a solution which is better than a $1/n$ approximation of* OPT*. Therefore, the approximation guarantee of Theorem 1 is asymptotically tight.*

We give a formal description of Observation 1 when discussing the constrained maximum directed cut problem on a graph as in Figure 1.

We give some approximation guarantee for GREEDY, assuming that the function f is monotone subadditive. Our proof method further generalizes the proof of (Conforti and Cornuéjols 1984). The following theorem holds.

Theorem 2. *Let f be a monotone subadditive function with curvature $\alpha \in [0, 1]$, and suppose that $f(\emptyset) = 0$. Then* GREEDY *is a $\frac{1}{\alpha}(1 - e^{\alpha^2 - \alpha})$-approximation algorithm for Problem* (1) *with run-time in $\mathcal{O}(dn)$.*

To our knowledge, this is the first approximation guarantee for the simple GREEDY maximizing a monotone subadditive function under partition matroid constraints.

Applications

Maximizing the logarithm of determinant functions.

An $n \times n$ matrix $\mathcal{P}$ is positive definite if $\mathcal{P}$ is symmetric and all its eigenvalues $\lambda_1, \dots, \lambda_n$ are strictly greater than

0. Consider such an $n \times n$ positive definite matrix $\mathcal{P}$. The *determinant function* $\det_{\mathcal{P}} : \{0,1\}^n \to \mathbb{R}_{\geq 0}$, with input an array $\mathbf{x} \in \{0,1\}^n$, returns the determinant of the square sub-matrix of $\mathcal{P}$ indexed by $\mathbf{x}$. We search for a sub-matrix of $\mathcal{P}$ that satisfies a partition or a cardinality constraint, and such that $\log\det$ is maximal.

Variations of this setting can be found in informative vector machines (Lawrence, Seeger, and Herbrich 2002) and in maximum entropy sampling problems (Krause, Singh, and Guestrin 2008).

The constrained problem of maximizing $\log\det_{\mathcal{P}}$ studied in the context of maximizing submodular functions under a single matroid constraint with a continuous greedy and non-oblivious local search (Sviridenko, Vondrák, and Ward 2017).

The problem of maximizing $\det_{\mathcal{P}}$ under a cardinality constraint is studied in (Bian et al. 2017), when $\mathcal{P}$ is a matrix of the form $\mathcal{P} = \mathcal{I} + \sigma\Sigma$, with $\mathcal{I}$ the $n \times n$ identity matrix, Σ a positive semidefinite matrix, and $\sigma > 0$ a scalar. In this case, the function $\det_{\mathcal{P}}$ is monotone, supermodular, and the submodularity ratio can be estimated in terms of the eigenvalues. Note that a matrix of the form $\mathcal{I} + \sigma\Sigma$ always has eigenvalues $\lambda_j \geq 1$.

We study the problem of maximizing $\det_{\mathcal{P}}$ under a partition matroid constraint, assuming that $\mathcal{P}$ is positive definite with eigenvalues $\lambda_j \geq 1$. We show that in this case the simple greedy algorithm is sufficient to obtain a nearly-optimal approximation guarantee. If $\log\det_{\mathcal{P}}$ is non-constant, using Proposition 2 we can upper bound the activity by $\alpha \leq 1 - 1/\lambda$, where λ the largest eigenvalue of $\mathcal{P}$ (Sviridenko, Vondrák, and Ward 2017). Thus, GREEDY gives a $(1 - 1/\lambda)\left(1 - e^{1/\lambda - 1}\right)$-approximation for Problem (1) when $f = \log\det_{\mathcal{P}}$ with runtime in $\mathcal{O}(nd)$. We do not assume that the eigenvalues are such that $\lambda_j > 1$, so our analysis applies to monotone as well as non-monotone functions. For instance, consider the function $\log\det_{\mathcal{A}}$ with

$$\mathcal{A} = \begin{pmatrix} \delta & \sqrt{\delta - 1} \\ \sqrt{\delta - 1} & 1 \end{pmatrix}$$

for all $\delta > 1$. In this case, the function $\log\det_{\mathcal{A}}$ is neither monotone, nor approximately monotone (Lee et al. 2009; Krause, Singh, and Guestrin 2008). GREEDY, nevertheless, finds a $(1 - 1/\delta)\left(1 - e^{1/\delta - 1}\right)$-approximation of the global optimum under partition matroid constraints.

We can further generalize this result to more complex function, by means of Proposition 1. For instance, let f be the entropy function of a Gaussian process, as defined in (4). Then the function f is the sum of a linear term $((1 + \ln(2\pi))/2)\,|S|$ and $1/2 \ln\det_{\Sigma}(S)$, for a positive semidefinite matrix Σ with eigenvalues $\lambda_j \geq 1$. This function is submodular, because both terms are submodular. Moreover, the linear term has curvature $\alpha = 0$, and the function $1/2\ln\det_{\Sigma}(S)$ has curvature $1 - 1/\lambda$, with λ the largest eigenvalue of Σ. Hence, we combine Theorem 1 with Proposition 1 to conclude that GREEDY is a $(1-1/\lambda)(1-e^{1/\lambda-1})$-approximation algorithm for Problem (1), with f the entropy as in (4). Note that our analysis does not require monotonicity, and it holds for matrices such as $\Sigma = \mathcal{A}$.

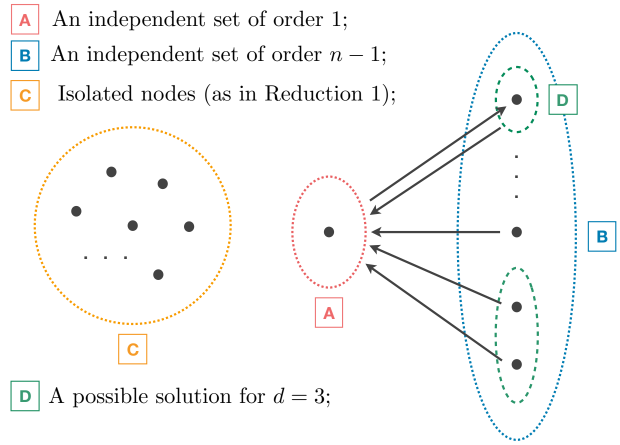

Figure 1: We consider a bipartite graph $G = (V, E)$ of order n and size n, with partitions labeled as A and B. In this example, there's only one node in A, and $n - 1$ nodes in B. Moreover, there's only one edge from A to B, whereas there is one edge from each node in B to A. Since all nodes in A and B have equal f-value, then GREEDY may output A as a solution to the maximum cut under uniform constrain of size d. This yields a $1/d$-approximation of the global optimum.

Finding the maximum directed cut of a graph.

Let $G = (V, E)$ be a graph with n vertices and m edges, together with a non-negative weight function $w \colon E \to \mathbb{R}_{\geq 0}$. We consider the problem of finding a subset $U \subseteq V$ of nodes such that the sum of the weights on the outgoing edges of U is maximal. This problem is the maximum directed cut problem known to be NP-complete. We consider a constrained version of this problem, as in Problem (1). We consider both directed and undirected graphs. We first define the cut function as follows.

Let $G = (V, E)$ and w be as above. For each subset of nodes $U \subseteq V$, consider the set $C(U) = \{(e_1, e_2) \in E \colon e_1 \in U \text{ and } e_2 \notin U\}$ of the edges leaving U. We define the *cut function* $f \colon 2^V \to \mathbb{R}_{\geq 0}$ with $f(U) = \sum_{e \in C(U)} w(e)$.

The constrained maximum directed cut problem can be approached by maximizing the cut function under a uniform cardinality constraint. Since we require the weights to be non-negative, this function is also non-negative. As noted in (Feige, Mirrokni, and Vondrák 2011), the cut function is always submodular and, in general, non-monotone.

Denote with Δ^+ the maximum out-degree of G, i.e. the maximum degree when counting *outgoing* edges and denote with Δ^- the maximum in-degree of G, obtained by counting the *incoming* edges only. Then from Proposition 2 the curvature of the corresponding cut function is upper-bounded as

$$\alpha \leq 1 + \frac{\inf\{d, \Delta^-\}}{\Delta^+}. \tag{2}$$

When G is undirected, $\Delta^- = \Delta^+$ and, therefore, $\alpha \leq 2$. Thus, Theorem 1 yields that GREEDY is a $1/2(1 - e^{-2})$-approximation algorithm for the constrained maximum cut problem. This approximation guarantee improves as d decreases.

When G is a directed graph the approximation guarantee can drop to $1/d$. Consider a bipartite graph $G = (V, E)$ with n vertices and n edges of weight 1 described as follows (see Figure 1). Let A, B be the partitions of V. A contains exactly one node and B contains $n-1$ nodes. The unique vertex of A has exactly one outgoing edge to a vertex in B. Each vertex in B has an outgoing edge to the only vertex of A. When maximizing the cut function of this graph G under the special case of cardinality constraint d, the optimal solution consists of d nodes in B. GREEDY though, may output A as a possible solution, which yields only a $1/d$-approximation of the optimal solution. In this case the curvature is $\alpha \geq d$. However, we show experimentally that the GREEDY performs well on a variety of real-world networks. We remark that in real-world networks the degree Δ_+ is expected to grow in the problem size (Newman 2003; Albert and Barabasi 2001).

Social welfare in combinatorial auctions.

We consider combinatorial auctions with n players competing for m items, where the items can have different values for each player. Moreover, the value of each item for a player may depend on the particular combination of items allocated to that player. For any given player $i = 1, \dots, n$, the value of a combination of items is expressed by the *utility function* $u_i \colon 2^{[m]} \to \mathbb{R}_{\geq 0}$. The objective of the social welfare problem (SW) is to find disjoint sets $S_1, \dots, S_n$ maximizing the total welfare $\sum_{i=1}^{n} u_i(S_i)$. Following (Bhawalkar and Roughgarden 2011), we make following natural assumptions on all utility functions:

1. $u_i(\emptyset) = 0$;
2. $u_i(U \cup T) \leq u_i(U) + u_i(T)$ for all $U, T \subseteq M$;
3. $u_i(U) \leq u_i(T)$ for all $U \subseteq T \subseteq M$.

Since an explicit description of a utility function may require exponential space, we assume the existence of an *oracle* returns the values of u_i for sets of items. In the literature, various oracle models have been considered (Dobzinski, Nisan, and Schapira 2010). We study the case where for each utility function u_i and any set of items $S \subseteq M$ there exists an oracle that returns the value $u_i(S)$. We refer to this setting as *value oracle model*. We remark that in the context of combinatorial auctions, the utility function u_i of a player is unknown to other players. Thus players may choose not to reveal the true value of the cost functions. In this setting, however, we assume all players to be *truthful*.

We formalize SW as a maximization problem under a partition matroid constraint, following (Feige and Vondrák 2010). For a given set of items M and n players, we define a ground set $X = [n] \times M$. The elements of X are copies of the items in M. For each player we require a copy of each item in M. For each player i we define a mapping $\pi_i : 2^X \longmapsto 2^N$ that assigns copies of items to respective players. In other words, for each set $I \times S \subseteq X$ it holds

$$\pi_i(I \times S) = \{\omega \in M \colon (i, \omega) \in I \times S\}.$$

Given utility functions $u_1, \dots, u_n$, the social welfare problem (SW) consists then of maximizing the following function

$$f(S) = \sum_{i=1}^{n} u_i(\pi_i(S)).$$

We note that the function f is subadditive, monotone and such that $f(\emptyset) = 0$. In this setting a feasible solution S can only assign one item per player. Thus, if we define $B_m = [n] \times \{m\}$, for all items $m \in M$, then a feasible solution S must fulfill the constrain $|S \cup B_m| \leq 1$ for all $m \in M$. Thus, maximizing f in the above setting is equivalent to maximizing a monotone function under a partition matroid constraint.

Consider SW with n players, m items, and utility functions $u_1, \dots u_n$. Denote with α_i the curvature of each utility function u_i. Then the function f has curvature $\alpha \leq \max_i \alpha_i$, by iteratively applying Proposition 1. We can now apply Theorem 2 and conclude that GREEDY is a $1/\alpha(1 - e^{\alpha^2 - \alpha})$-approximation algorithm for SW in the value oracle model.

Experiments

The maximum entropy sampling problem.

In this set of experiments we study the following problem: Given a set of random variables (RVs), find the most informative subset of variables, subject to a side constraint as in Problem (1). This setting finds a broad spectrum of applications, from Bayesian experimental design (Sebastiani and Wynn 2002), to monitoring spatio-temporal dynamics (Singh et al. 2009).

We consider the Berkley Earth climate dataset [1]. This dataset combines 1.6 billion temperature reports from 16 preexisting data archives, for over 39.000 unique stations. For each station, we consider a unique time series for the *average* monthly temperature. We always consider time series that span between years 2015-2017. This gives us a total of 2736 time series, for unique corresponding stations. The code is available at [removed for review].

We study the problem of searching for the most informative sets of time series under various constraints, based on these observations. Given a time series $\mathbf{X} = \{X_t\}_t$ we study the corresponding variation series $\overline{\mathbf{X}} = \{\overline{X}_t\}_t$ defined as $\overline{X}_t = X_t - X_{t-1}$. A visualization of time series $\overline{\mathbf{X}}$ is given in Figure 3(a).

We compute the covariance matrix Σ between series $\overline{\mathbf{X}}$, $\overline{\mathbf{Y}}$, the entries of which are defined as

$$\mathsf{cov}(\overline{\mathbf{X}}, \overline{\mathbf{Y}}) = \frac{1}{m-1} \sum_{t=1}^{m} (\overline{X}_t - \mathbb{E}\left[\overline{\mathbf{X}}\right])(\overline{Y}_t - \mathbb{E}\left[\overline{\mathbf{Y}}\right]), \quad (3)$$

with $m = 35$ the length of each series. A visualization of the covariance the matrix Σ is given in Figure 4.

Assuming that the joint probability distribution is Gaussian, we proceed by maximizing the *entropy*, defined as

$$f(S) = \frac{1 + \ln(2\pi)}{2} |S| + \frac{1}{2} \ln \det_{\Sigma}(S) \quad (4)$$

[1] http://berkeleyearth.org/data/

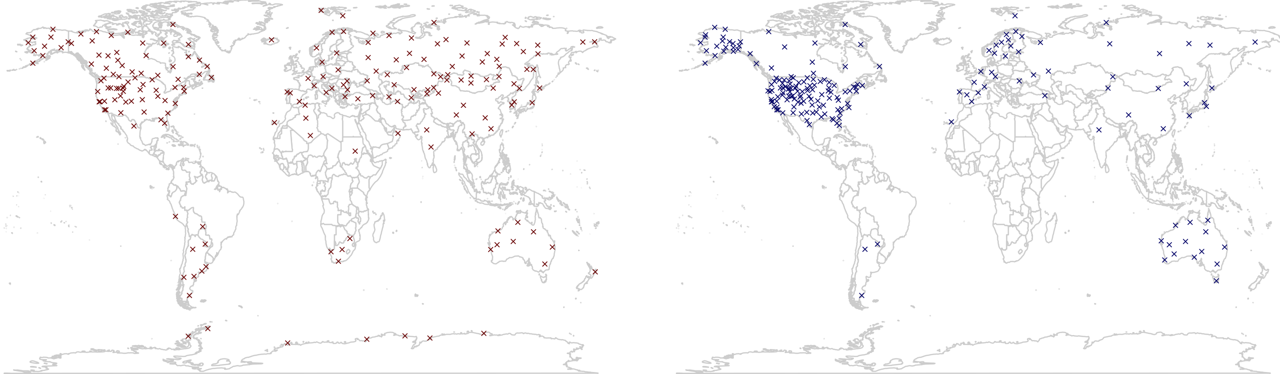

Figure 2: A visualization of the solution found by GREEDY for $d = 10\%$ in the case of a uniform constraint (left), and a partition constraint by countries (right). In both case, a solution is obtained by maximizing the entropy as given in (4). The covariance matrix Σ for all possible locations is displayed in Figure 4. We observe that in the case of a cardinality constraint, the informative stations tend to be spread out, whereas in the partition constraint by countries they tend to be grouped in a few areas. We remark that in the original dataset stations are not distributed uniformly among countries.

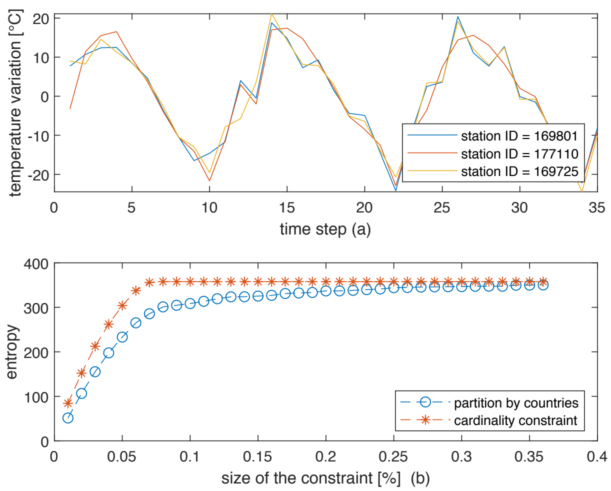

Figure 3: (a) A visualization of the monthly temperature variations of three time series, with particularly high variance. Each series corresponds to a unique station ID. We model each variation series as a Gaussian distribution.
(b) Optimal solution found by GREEDY for a uniform constraint and a partition matroid constraint by countries. The f-value of each set of stations is the entropy (4), with Σ the covariance matrix of variation series as in (a) (see Figure 4).

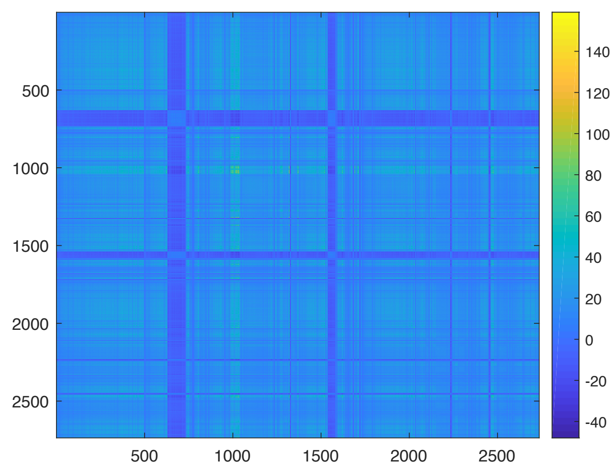

Figure 4: A visualization of the covariance matrix Σ of time series available in the Berkley Earth climate dataset. We consider stations that have full available reports between years 2015-2017, for a total of 2736 stations. We consider the variation between average monthly temperatures of each time series. Each entry of this matrix is computed by taking the sample covariance as in (3).

for any indexing set $S \subseteq \{0, 1\}^n$.

We consider two types of constraints. In a first set of experiments we consider the problem of maximizing the entropy as in (4), under a cardinal constraint only. Specifically, given a parameter d, the goal is to find a subsets of time series that maximizes the entropy, of size at most d of all available data. We also consider a more complex constraint: Find a subset of time series that maximizes the entropy, and s.t. it contains at most d of all available data of each country. The latter constraint is a partition matroid constraints, where each subset B_i consists of all data series measured by stations in a given country.

A summary of the results is displayed in Figure 3(b). We observe that in both cases the entropy quickly evolves to a stationary local optimum, indicating that a relatively small subset of stations is sufficient to explain the random variations between monthly observations in the model. We observe that the GREEDY reaches similar approximation guarantees in both cases. We remark that the GREEDY finds a nearly optimal solution under a cardinality constraint, assuming that the entropy is (approximately) mono-

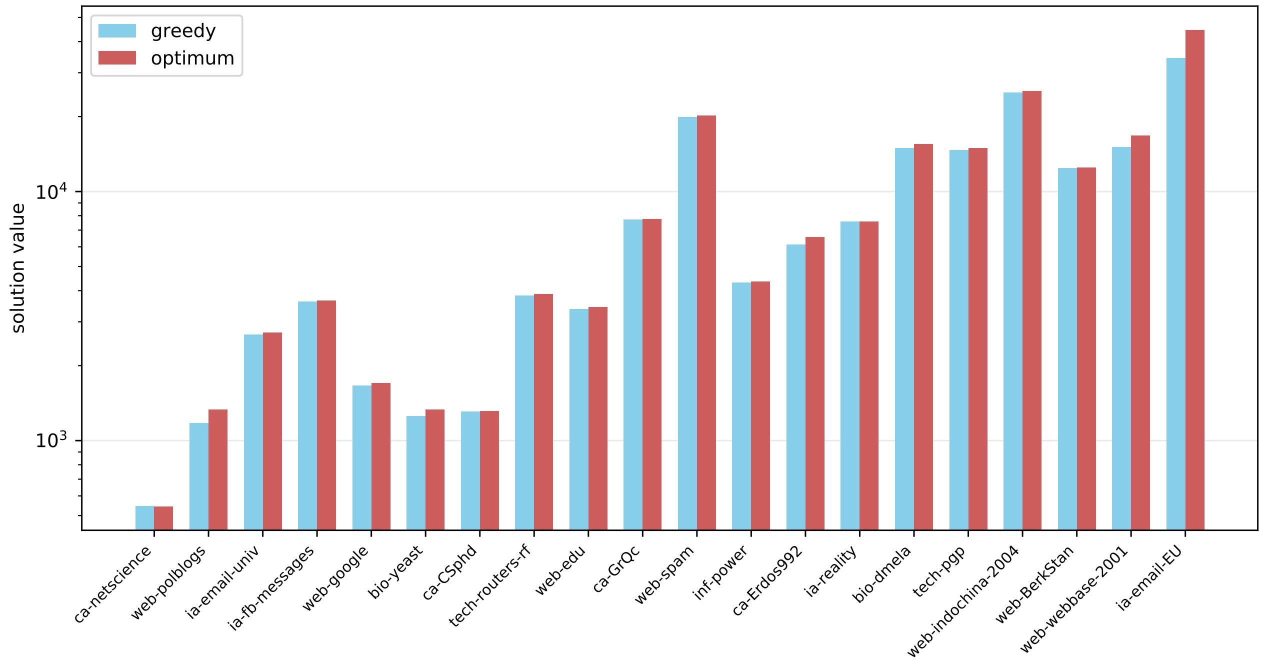

Figure 5: Visualization of the optimal solution and the solution found by GREEDY for the maximum directed cut problem. The input graphs are ordered by increasing number of vertices from left to right.

tone (Krause, Singh, and Guestrin 2008).

In Figure 2 we display solutions found by GREEDY for the cardinality and partition matroid constraint, with $d = 10\%$.

We observe that in the case of a cardinality constraint, the sensors spread across the map; in the case of a partition matroid constraint sensors tend to be placed unevenly. We remark that in the original data set, some countries have a much higher density of stations than others.

Finding the maximum directed cut of a graph.

In this set of experiments we study the performance of GREEDY for maximum directed cut in unweighted graphs. We compare these results with the optimal solutions, which we found via an Integer Linear Program solved with the state-of-the-art solver Gurobi (Gurobi Optimization 2018). The experiments were conducted on 20 instances from Network Repository (Rossi and Ahmed 2016).

Figure 5 displays the quality of the solution found by GREEDY compared to the optimal solution, in the unconstrained case. One can see that in most cases the greedy solution is very close to the optimum. This suggests that GREEDY might perform well on real-world instances. We remark that the solution quality is expected to increase as the size of a possible constraint lowers (see (2)). Thus, GREEDY is expected to perform even better in the constrained case.

Theorem 1 implies that this might be due to the curvature α of these graphs. However, we find that the solution quality of GREEDY is much better than the upper bound on the curvature due to inequality (2) suggests.

Conclusion

In this paper we consider the problem of maximizing a function with bounded curvature under a single partition matroid constraint.

We derive approximation guarantees for the simple GREEDY algorithm (see Algorithm 1) on those problems, in the case of a (non-monotone) submodular function, and a monotone subadditive function (see Theorem 1 and Theorem 2). We observe that the lower bound on the approximation guarantee is asymptotically tight in the case of a submodular function.

We discuss three applications. We first consider the problem of maximizing a determinant function of a positive semidefinite matrix, and related maximum entropy sampling problem. We then consider the constrained maximum cut problem on directed graphs. We conclude discussing an application to game theory.

We observe that the simple GREEDY algorithm is suitable to approach Problem (1) it the function f has bounded curvature. Particularly, we need not search for (approximate) monotonicity or perform a local search to obtain good approximate solutions.

Acknowledgement

This research has been supported by the Australian Research Council (ARC) through grant DP160102401 and the German Science Foundation (DFG) through grant FR2988 (TOSU).

References

[Albert and Barabasi 2001] Albert, R., and Barabasi, A.-L. 2001. Statistical mechanics of complex networks. *Reviews of Modern Physics* 74.

[Assadi 2017] Assadi, S. 2017. Combinatorial auctions do need modest interaction. In *EC*, 145–162.

[Bhawalkar and Roughgarden 2011] Bhawalkar, K., and Roughgarden, T. 2011. Welfare guarantees for combinatorial auctions with item bidding. In *Proc. of SODA*, 700–709.

[Bian et al. 2017] Bian, A. A.; Buhmann, J. M.; Krause, A.; and Tschiatschek, S. 2017. Guarantees for greedy maximization of non-submodular functions with applications. In *Proc. of ICML*, 498–507.

[Buchbinder and Feldman 2018] Buchbinder, N., and Feldman, M. 2018. Deterministic algorithms for submodular maximization problems. *ACM Transactions on Algorithms* 14(3):32:1–32:20.

[Buchbinder et al. 2014] Buchbinder, N.; Feldman, M.; Naor, J.; and Schwartz, R. 2014. Submodular maximization with cardinality constraints. In *Proc. of SODA*, 1433–1452.

[Buchbinder, Feldman, and Garg 2018] Buchbinder, N.; Feldman, M.; and Garg, M. 2018. Deterministic (1/2 + ϵ)-approximation for submodular maximization over a matroid. *CoRR* abs/1807.05532.

[Călinescu et al. 2011] Călinescu, G.; Chekuri, C.; Pál, M.; and Vondrák, J. 2011. Maximizing a monotone submodular function subject to a matroid constraint. *SIAM Journal of Computing* 40(6):1740–1766.

[Conforti and Cornuéjols 1984] Conforti, M., and Cornuéjols, G. 1984. Submodular set functions, matroids and the greedy algorithm: Tight worst-case bounds and some generalizations of the rado-edmonds theorem. *Discrete Applied Mathematics* 7(3):251–274.

[Cornuejols, Fisher, and Nemhauser 1977] Cornuejols, G.; Fisher, M. L.; and Nemhauser, G. L. 1977. Location of Bank Accounts to Optimize Float: An Analytic Study of Exact and Approximate Algorithms. *Management Science* 23(8):789–810.

[Das and Kempe 2011] Das, A., and Kempe, D. 2011. Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection. In *Proc. of ICML*, 1057–1064.

[Dobzinski, Nisan, and Schapira 2010] Dobzinski, S.; Nisan, N.; and Schapira, M. 2010. Approximation algorithms for combinatorial auctions with complement-free bidders. *Mathematics of Operations Research* 35(1):1–13.

[Elenberg et al. 2017] Elenberg, E. R.; Dimakis, A. G.; Feldman, M.; and Karbasi, A. 2017. Streaming weak submodularity: Interpreting neural networks on the fly. In *Advances in Neural Information Processing Systems*, 4047–4057.

[Feige and Vondrák 2010] Feige, U., and Vondrák, J. 2010. The submodular welfare problem with demand queries. *Theory of Computing* 6(1):247–290.

[Feige, Mirrokni, and Vondrák 2011] Feige, U.; Mirrokni, V. S.; and Vondrák, J. 2011. Maximizing non-monotone submodular functions. *SIAM Journal of Computing* 40(4):1133–1153.

[Goemans and Williamson 1995] Goemans, M. X., and Williamson, D. P. 1995. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM* 42(6):1115–1145.

[Gurobi Optimization 2018] Gurobi Optimization, L. 2018. Gurobi optimizer reference manual.

[Jegelka and Bilmes 2011] Jegelka, S., and Bilmes, J. 2011. Submodularity beyond submodular energies: Coupling edges in graph cuts. In *Proc. of CVPR*, 1897–1904.

[Krause, Singh, and Guestrin 2008] Krause, A.; Singh, A. P.; and Guestrin, C. 2008. Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research* 9:235–284.

[Lawrence, Seeger, and Herbrich 2002] Lawrence, N. D.; Seeger, M. W.; and Herbrich, R. 2002. Fast sparse gaussian process methods: The informative vector machine. In *Proc. of NIPS*, 609–616.

[Lee et al. 2009] Lee, J.; Mirrokni, V. S.; Nagarajan, V.; and Sviridenko, M. 2009. Non-monotone submodular maximization under matroid and knapsack constraints. In *Proc. of STOC*, 323–332.

[Lin and Bilmes 2010] Lin, H., and Bilmes, J. 2010. Multi-document summarization via budgeted maximization of submodular functions. In *Proc. of HLT*, 912–920.

[Maehara et al. 2017] Maehara, T.; Kawase, Y.; Sumita, H.; Tono, K.; and Kawarabayashi, K. 2017. Optimal pricing for submodular valuations with bounded curvature. In *Proc. of AAAI*, 622–628.

[Nemhauser and Wolsey 1978] Nemhauser, G. L., and Wolsey, L. A. 1978. Best algorithms for approximating the maximum of a submodular set function. *Mathematics of Operations Research* 3(3):177–188.

[Newman 2003] Newman, M. E. J. 2003. The structure and function of complex networks. *SIAM Review* 45(2):167–256.

[Rossi and Ahmed 2016] Rossi, R. A., and Ahmed, N. K. 2016. An interactive data repository with visual analytics. *SIGKDD Explorations* 17(2):37–41.

[Sebastiani and Wynn 2002] Sebastiani, P., and Wynn, H. P. 2002. Maximum entropy sampling and optimal bayesian experimental design. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 62(1):145–157.

[Singh et al. 2009] Singh, A.; Krause, A.; Guestrin, C.; and Kaiser, W. J. 2009. Efficient informative sensing using multiple robots. *Journal of Artificial Intelligence Research* 34:707–755.

[Sviridenko, Vondrák, and Ward 2017] Sviridenko, M.; Vondrák, J.; and Ward, J. 2017. Optimal approximation for submodular and supermodular optimization with bounded curvature. *Mathematics of Operations Research* 42(4).

[Tschiatschek, Singla, and Krause 2017] Tschiatschek, S.; Singla, A.; and Krause, A. 2017. Selecting sequences of items via submodular maximization. In *Proc. of AAAI*, 2667–2673.

[Vondrák 2010] Vondrák, J. 2010. Submodularity and curvature: the optimal algorithm. *RIMS Kokyuroku Bessatsu* B23:253–266.

Appendix (Missing Proofs)

Proof of Reduction 1. Fix a constant $c > 0$.We observe that if the condition of the statement does not hold, then it is sufficient to add a set D of $\sum_i cd_i$ " dummy" elements that do not have any effect on the f-values, and remove them from the output of the algorithm, for all $i = 1, \ldots, k$. Denote with $D_1, \ldots, D_k$ a partition of D with $|D_i| = cd_i$. This only increases the size of the instance by a multiplicative constant factor. We define new subsets $\overline{B}_i = B_i \cup D_i$ for all $i = 1, \ldots, k$. Thus, we can maximize the function f on the newly-defined partition constraint without affecting neither the global optimum, nor the value of the algorithm's output. □

Proof of Proposition 1. Fix two subsets $S, \Omega \subseteq V$ of size at most d, and a point $\omega \in S \setminus \Omega$. From the definition of curvature we have

$$f(S \cup \Omega) - f((S \cup \omega) \setminus \{\omega\}) \geq (1 - \alpha_1)(f(S) - f(S \setminus \{\omega\})), \tag{5}$$
$$g(S \cup \Omega) - g((S \cup \omega) \setminus \{\omega\}) \geq (1 - \alpha_2)(g(S) - g(S \setminus \{\omega\})). \tag{6}$$

Thus, we have that it holds

$$\begin{aligned}
(f + g)(S \cup \Omega) - (f + g)((S \cup \omega) \setminus \{\omega\}) &= f(S \cup \Omega) - f((S \cup \omega) \setminus \{\omega\}) + g(S \cup \Omega) - g((S \cup \omega) \setminus \{\omega\}) && (7)\\
&\geq (1 - \alpha_1)(f(S) - f(S \setminus \{\omega\})) + (1 - \alpha_2)(g(S) - g(S \setminus \{\omega\})) && (8)\\
&\geq (1 - \sup_i \alpha_i)\left(f(S) - f(S \setminus \{\omega\}) + g(S) - g(S \setminus \{\omega\})\right) && (9)\\
&= (1 - \sup_i \alpha_i)\left((f + g)(S) - (f + g)(S \setminus \{\omega\})\right), && (10)\\
& && (11)
\end{aligned}$$

where (7) holds from the definition of $f + g$; (8) follows from (7) by means of (5) and (6); (9)-(10) follow by means of standard calculations, and from the definition of $f + g$. The claim follows. □

Proof of Proposition 2. Fix any two subsets $S, \Omega \subseteq V$ of size at most $\sum_i d_i$, and let $\omega \in S \setminus \Omega$. Then it holds

$$\alpha \leq 1 - \frac{f(S \cup \Omega) - f((S \cup \Omega) \setminus \{\omega\})}{f(\omega) - f(\emptyset)} \leq 1 - \frac{f(S \cup \Omega) - f((S \cup \Omega) \setminus \{\omega\})}{f(S) - f(S \setminus \{\omega\})},$$

where the last inequality follows from the definition of submodular function. The claim follows. □

Proof of Theorem 1. We assume without loss of generality that f is non-constant. Moreover, we may assume that $d > \alpha$ (see Reduction 1). Let $M_t \subseteq V$ be a set of size d that maximizes the sum $\sum_{\omega \in M_t} \rho_\omega(S_{t-1})$, and consider a set $\overline{M}_t \subseteq V$ of size k of the form $\overline{M}_t = (\text{OPT} \setminus S_{t-1}) \cup D_t$ with $D_t \subseteq D$ s.t. $\left|\overline{M}_t \cap B_i\right| \leq d_i$ for all $i = 1, \ldots, k$. Thus, we have that

$$\rho_t \geq \frac{1}{d} \sum_{\omega \in M_t} \rho_\omega(S_{t-1}) \geq \frac{1}{d} \sum_{\omega \in \overline{M}_t} \rho_\omega(S_{t-1}) \geq \frac{1}{d}\left(f(S_{t-1} \cup \text{OPT}) - f(S_{t-1})\right), \tag{12}$$

where the first inequality follows from the maximality of M_t over all sets of the same size, and the second inequality in 12 follows from the definition of approximate monotonicity. To continue with the proof we consider the following lemma.

Lemma 1. *For any subset $S \subseteq V$ of size at most d it holds $f(S \cup \text{OPT}) \geq f(\text{OPT}) + (1 - \alpha)(f(S) - f(\emptyset))$.*

Proof. Denote with ℓ the size of S, and let $\omega_1, \ldots, \omega_\ell \in S$ be the points of S in any given order. For any subset $S \subseteq V$ it holds $f(S \cup \text{OPT}) \geq f(\text{OPT}) + (1 - \alpha)\left(f(S) - f(\emptyset)\right)$. Define the sets $U_i = \{\omega_1, \ldots, \omega_i\}$, for all $i = 1, \ldots, \ell$. Then the U_i are s.t. $U_i \subseteq U_j$ for all $i \leq j$, and $U_i \setminus U_{i-1} = \omega_i$, for all $i = 1, \ldots, \ell$. Note also that it holds $U_\ell = S$. From the definition of curvature we have that

$$f(U_i \cup \text{OPT}) - f(U_{i-1} \cup \text{OPT}) \geq (1 - \alpha)\left(f(U_i) - f(U_{i-1})\right).$$

for all $i = 1, \ldots, \ell$. It follows that

$$\begin{aligned}
f(U_\ell \cup \text{OPT}) &\geq f(U_{\ell-1} \cup \text{OPT}) + (1 - \alpha)\left(f(U_\ell) - f(U_{\ell-1})\right) && (13)\\
&\geq f(\emptyset \cup \text{OPT}) + \sum_{j=1}^{\ell}(1 - \alpha)\left(f(U_j) - f(U_{j-1})\right) && (14)\\
&= f(\text{OPT}) + (1 - \alpha)\left(f(U_\ell) - f(\emptyset)\right), && (15)
\end{aligned}$$

where (14) follows by iteratively applying (13) to the $f(U_{j-1} \cup \text{OPT})$, and (15) follows by taking the telescopic sum. □

We combine Lemma 1 by setting $S = S_{t-1}$ with (12) to obtain that

$$\rho_t \geq \frac{1}{d}\left(f(\text{OPT}) - \alpha f(S_{t-1})\right) - \frac{1-\alpha}{d} f(\emptyset). \tag{16}$$

for all $t = 1, \ldots, d$. Since it holds $\sum_{j=1}^{t-1} \rho_j = f(S_{t-1}) - f(\emptyset)$, and we can combine this observation with (16), to obtain the following LP

$$\begin{bmatrix} d & 0 & \ldots\ldots\ldots\ldots\ldots & & & 0 \\ \alpha & d & 0 & \ldots\ldots\ldots\ldots & & 0 \\ \alpha & \alpha & d & 0 & \ldots\ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ \alpha & \ldots\ldots & \alpha & d & 0 & 0 \\ \alpha & \ldots\ldots\ldots & & \alpha & d & 0 \\ \alpha & \ldots\ldots\ldots\ldots\ldots & & & \alpha & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{d-2} \\ x_{d-1} \\ x_d \end{bmatrix} \geq \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \\ 1 \end{bmatrix} \tag{17}$$

with $x_t := \rho_t / \left(f(\text{OPT}) - f(\emptyset)\right)$, for all $t = 1, \ldots, d$. To continue with the proof, we consider the following lemma.

Lemma 2. *Let $(y_1, \ldots, y_d) \in \mathbb{R}^d_{\geq 0}$ be a solution to the LP given in* (17). *Then it holds*

$$y_t \geq \frac{1}{d}\left(1 - \frac{\alpha}{d}\right)^{t-1}$$

for all $t = 1, \ldots, d$.

Proof. We first show by induction that any solution $(z_1, \ldots, z_k)$ that fulfills the constrains

$$\begin{bmatrix} d & 0 & \ldots\ldots\ldots\ldots\ldots & & & 0 \\ \alpha & d & 0 & \ldots\ldots\ldots\ldots & & 0 \\ \alpha & \alpha & d & 0 & \ldots\ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ \alpha & \ldots\ldots & \alpha & d & 0 & 0 \\ \alpha & \ldots\ldots\ldots & & \alpha & d & 0 \\ \alpha & \ldots\ldots\ldots\ldots\ldots & & & \alpha & d \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_{d-2} \\ z_{d-1} \\ z_d \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \\ 1 \end{bmatrix} \tag{18}$$

yields

$$z_t = \frac{1}{d}\left(1 - \frac{\alpha}{d}\right)^{t-1}. \tag{19}$$

The base case with $t = 1$ is trivially true. Suppose now that the claim holds for all $z_1, \ldots, z_{t-1}$. Then it holds

$$z_t = \frac{1}{d}\left(1 - \alpha \sum_{j=1}^{t-1} z_j\right) \tag{20}$$

$$= \frac{1}{d}\left(1 - \alpha \sum_{j=1}^{t-1} \frac{1}{d}\left(1 - \frac{\alpha}{d}\right)^{j-1}\right) \tag{21}$$

$$= \frac{1}{d}\left(1 - \frac{\alpha}{d}\right)^{t-1} \tag{22}$$

where (20) follows from (18); (21) follows by applying the inductive hypothesis to (20), and (22) follows by means of standard calculations. Thus, (19) holds. In particular, since any other solution $(y_1, \ldots, y_k)$ to (17) is s.t. $y_t \geq z_t$ for all $t = 1, \ldots, d$, then the claim follows. □

Thus, from Lemma 2 it holds

$$\rho_t \geq \frac{1}{d}\left(1 - \frac{\alpha}{d}\right)^{t-1} \left(f(\text{OPT}) - f(\emptyset)\right) \tag{23}$$

for all $t = 1, \ldots, d$. Therefore, using the linearity of expectation we have that

$$\begin{aligned} f(S_d) &\geq \sum_{t=1}^{d} \rho_t + f(\emptyset) && (24) \\ &\geq \sum_{t=1}^{d} \frac{1}{d}\left(1 - \frac{\alpha}{d}\right)^{t-1} (f(\text{OPT}) - f(\emptyset)) + f(\emptyset) && (25) \\ &= \frac{1}{\alpha}\left(1 - \left(1 - \frac{\alpha}{d}\right)^{d}\right)(f(\text{OPT}) - f(\emptyset)) + f(\emptyset) && (26) \\ &\geq \frac{1}{\alpha}\left(1 - e^{-\alpha}\right) f(\text{OPT}), && (27) \end{aligned}$$

where (24) follows from the definition of ρ_t; (25) follows from (23); (26) and (27) follow by means of standard calculations. □

Proof of Theorem 2. This proof is similar to that of Theorem 1. Again, we assume without loss of generality that f is non-constant. We first observe that the following holds. Let $S \subseteq V$ be any subset of size at most d such that $\rho_{\text{OPT}}(S) \neq 0$. Then it holds

$$\begin{aligned} \sum_{\omega \in \text{OPT}} \frac{\rho_\omega(S)}{\rho_{\text{OPT}}(S)} &\geq \sum_{\omega \in \text{OPT}} \frac{\rho_\omega(S)}{f(\text{OPT})} && (28) \\ &\geq (1-\alpha) \sum_{\omega \in \text{OPT}} \frac{f(\{\omega\}) - f(\emptyset)}{f(\text{OPT})} && (29) \\ &\geq (1-\alpha), && (30) \end{aligned}$$

where (28) follows applying subadditivity to the denominator; (29) follows applying the definition of curvature to the numerator; (30) follows applying the definition of subadditivity to the numerator, together with the assumption that $f(\emptyset) = 0$. Let $M_t \subseteq V$ be a set of size d that maximizes the sum $\sum_{\omega \in M_t} \rho_\omega(S_{t-1})$, and consider a set $\overline{M}_t \subseteq V$ of size k of the form $\overline{M}_t = (\text{OPT} \setminus S_{t-1}) \cup D_t$ with $D_t \subseteq D$ s.t. $\left|\overline{M}_t \cap B_i\right| \leq d_i$ for all $i = 1, \ldots, k$. Thus, from (28)-(30) we have that

$$\rho_t \geq \frac{1}{d} \sum_{\omega \in M_t} \rho_\omega(S_{t-1}) \geq \frac{1}{d} \sum_{\omega \in \overline{M}_t} \rho_\omega(S_{t-1}) \geq \frac{1-\alpha}{d} \left(f(S_{t-1} \cup \text{OPT}) - f(S_{t-1})\right), \tag{31}$$

where the first inequality follows from the maximality of M_t over all sets of the same size, and the second inequality in 31 follows from the definition of approximate monotonicity. We combine Lemma 1 by setting $S = S_{t-1}$ with (31) to obtain that

$$\rho_t \geq \frac{1-\alpha}{d} \left(f(\text{OPT}) - \alpha f(S_{t-1})\right). \tag{32}$$

for all $t = 1, \ldots, k$. Since it holds $\sum_{j=1}^{t-1} \rho_j = f(S_{t-1})$, and we can combine this observation with (32), to obtain the following linear program

$$\begin{bmatrix} \frac{d}{1-\alpha} & 0 & \cdots\cdots & & & & 0 \\ \alpha & \frac{d}{1-\alpha} & 0 & \cdots & & & 0 \\ \alpha & \alpha & \frac{d}{1-\alpha} & 0 & \cdots & & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & & \vdots \\ \alpha & \cdots & & \alpha & \frac{d}{1-\alpha} & 0 & 0 \\ \alpha & \cdots & & & \alpha & \frac{d}{1-\alpha} & 0 \\ \alpha & \cdots & & & & \alpha & \frac{d}{1-\alpha} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{d-2} \\ x_{d-1} \\ x_d \end{bmatrix} \geq \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \\ 1 \end{bmatrix} \tag{33}$$

with $x_t := \rho_t / f(\text{OPT})$, for all $t = 1, \ldots, d$.

Lemma 3. *Let $(y_1, \ldots, y_d) \in \mathbb{R}^d_{\geq 0}$ be a solution to the LP given in* (33). *Then it holds*

$$y_t \geq \frac{1-\alpha}{d}\left(1 - \frac{\alpha(1-\alpha)}{d}\right)^{t-1}$$

for all $t = 1, \ldots, d$.

Proof. We first show by induction that any solution $(z_1, \ldots, z_k)$ that fulfills the constrains

$$\begin{bmatrix} \frac{d}{1-\alpha} & 0 & \ldots\ldots\ldots\ldots\ldots\ldots\ldots & & & 0 \\ \alpha & \frac{d}{1-\alpha} & 0 & \ldots\ldots\ldots\ldots\ldots & & 0 \\ \alpha & \alpha & \frac{d}{1-\alpha} & 0 & \ldots\ldots\ldots\ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ \alpha & \ldots\ldots\ldots & \alpha & \frac{d}{1-\alpha} & 0 & 0 \\ \alpha & \ldots\ldots\ldots\ldots\ldots & & \alpha & \frac{d}{1-\alpha} & 0 \\ \alpha & \ldots\ldots\ldots\ldots\ldots\ldots\ldots & & & \alpha & \frac{d}{1-\alpha} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_{d-2} \\ z_{d-1} \\ z_d \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \\ 1 \end{bmatrix} \tag{34}$$

yields

$$z_t = \frac{1-\alpha}{d} \left(1 - \frac{\alpha(1-\alpha)}{d}\right)^{t-1}. \tag{35}$$

The base case with $t = 1$ is trivially true. Suppose now that the claim holds for all $z_1, \ldots, z_{t-1}$. Then it holds

$$z_t = \frac{1-\alpha}{d} \left(1 - \alpha \sum_{j=1}^{t-1} z_j\right) \tag{36}$$

$$= \frac{1-\alpha}{d} \left(1 - \alpha \sum_{j=1}^{t-1} \frac{1-\alpha}{d} \left(1 - \frac{\alpha(1-\alpha)}{d}\right)^{j-1}\right) \tag{37}$$

$$= \frac{1-\alpha}{d} \left(1 - \frac{\alpha(1-\alpha)}{d}\right)^{t-1} \tag{38}$$

where (36) follows from (34); (37) follows by applying the inductive hypothesis to (36), and (38) follows by means of standard calculations. Thus, (35) holds. In particular, since any other solution $(y_1, \ldots, y_k)$ to (33) is s.t. $y_t \geq z_t$ for all $t = 1, \ldots, d$, then the claim follows. □

Thus, from Lemma 3 it holds

$$\rho_t \geq \frac{1-\alpha}{d} \left(1 - \frac{\alpha(1-\alpha)}{d}\right)^{t-1} f(\text{OPT}) \tag{39}$$

for all $t = 1, \ldots, d$. Therefore, using the linearity of expectation we have that

$$f(S_d) \geq \sum_{t=1}^{d} \rho_t \tag{40}$$

$$\geq \sum_{t=1}^{d} \frac{1-\alpha}{d} \left(1 - \frac{\alpha(1-\alpha)}{d}\right)^{t-1} f(\text{OPT}) \tag{41}$$

$$= \frac{1}{\alpha} \left(1 - \left(1 - \frac{\alpha(1-\alpha)}{d}\right)^{d}\right) f(\text{OPT}) \tag{42}$$

$$\geq \frac{1}{1-\alpha} \left(1 - e^{-(1-\alpha)\alpha}\right) f(\text{OPT}), \tag{43}$$

where (40) follows from the definition of ρ_t; (41) follows from (39); (42) and (43) follow by means of standard calculations. □