

Maximizing k -Submodular Functions and Beyond

JUSTIN WARD, EPFL, Switzerland

STANISLAV ŽIVNÝ, University of Oxford, United Kingdom

We consider the maximization problem in the value oracle model of functions defined on k -tuples of sets that are submodular in every orthant and r -wise monotone, where $k \geq 2$ and $1 \leq r \leq k$. We give an analysis of a deterministic greedy algorithm that shows that any such function can be approximated to a factor of $1/(1+r)$. For $r = k$, we give an analysis of a randomised greedy algorithm that shows that any such function can be approximated to a factor of $1/(1 + \sqrt{k/2})$.

In the case of $k = r = 2$, the considered functions correspond precisely to bisubmodular functions, in which case we obtain an approximation guarantee of $1/2$. We show that, as in the case of submodular functions, this result is the best possible in both the value query model, and under the assumption that $NP \neq RP$.

Extending a result of Ando et al., we show that for any $k \geq 3$ submodularity in every orthant and pairwise monotonicity (i.e. $r = 2$) precisely characterize k -submodular functions. Consequently, we obtain an approximation guarantee of $1/3$ (and thus independent of k) for the maximization problem of k -submodular functions.

CCS Concepts: • **Theory of computation** → **Approximation algorithms analysis**;

Additional Key Words and Phrases: submodularity, bisubmodularity, k -submodularity

ACM Reference Format:

Justin Ward and Stanislav Živný. 2015. Maximizing k -Submodular Functions and Beyond. *ACM Trans. Algor.* V, N, Article A (January YYYY), 26 pages.
DOI: <http://dx.doi.org/10.1145/0000000.0000000>

1. INTRODUCTION

Given a finite nonempty set U , a set function $f : 2^U \rightarrow \mathbb{R}_+$ defined on subsets of U is called *submodular* if for all $S, T \subseteq U$,

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

Submodular functions are a key concept in operations research and combinatorial optimization [Nemhauser and Wolsey 1988; Narayanan 1997; Topkis 1998; Schrijver 2003; Fujishige 2005; Korte and Vygen 2007; Iwata 2008]. Examples of submodular functions include cut capacity functions, matroid rank functions, and entropy functions. Submodular functions are often considered to be a discrete analogue of convex functions [Lovász 1983].

Both minimizing and maximizing submodular functions have been considered extensively in the literature, in both constrained and unconstrained settings. Submodular

Justin Ward was supported by EPSRC grants EP/J021814/1 and EP/D063191/1 (work performed while the author was at the University of Warwick, United Kingdom). Stanislav Živný was supported by a Royal Society University Research Fellowship. An extended abstract of part of this work appeared in the *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2014 [Ward and Živný 2014]. Authors' addresses: J. Ward, EPFL IC IIF THL2, INJ 131 (Bâtiment INJ), Station 14, CH-1015 Lausanne, Switzerland. S. Živný, Department of Computer Science, University of Oxford, Oxford OX1 3QD, United Kingdom.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© YYYY ACM. 1549-6325/YYYY/01-ARTA \$15.00

DOI: <http://dx.doi.org/10.1145/0000000.0000000>

function maximization is easily shown to be NP-hard [Schrijver 2003] since it generalizes many standard NP-hard problems such as the maximum cut problem [Garey and Johnson 1979; Feige et al. 2011]. In contrast, the problem of minimizing a submodular function can be solved efficiently with only polynomially many evaluations of the function [Iwata 2008] either by using the ellipsoid algorithm [Grötschel et al. 1981; 1988], or by using one of several combinatorial algorithms that have been obtained in the last decade [Schrijver 2000; Iwata et al. 2001; Iwata 2002; 2003; Orlin 2009; Iwata and Orlin 2009].

Following a question by Lovász [Lovász 1983], a generalization of submodularity to biset functions has been introduced. Given a finite nonempty set U , a function $f : 3^U \rightarrow \mathbb{R}_+$ defined on pairs of disjoint subsets of U is called *bisubmodular* if for all pairs (S_1, S_2) and (T_1, T_2) of disjoint subsets of U ,

$$f(S_1, S_2) + f(T_1, T_2) \geq f((S_1, S_2) \sqcap (T_1, T_2)) + f((S_1, S_2) \sqcup (T_1, T_2)),$$

where we define

$$(S_1, S_2) \sqcap (T_1, T_2) = (S_1 \cap T_1, S_2 \cap T_2),$$

and

$$(S_1, S_2) \sqcup (T_1, T_2) = ((S_1 \cup T_1) \setminus (S_2 \cup T_2), (S_2 \cup T_2) \setminus (S_1 \cup T_1)).$$

Bisubmodular functions were originally studied in the context of rank functions of delta-matroids [Bouchet 1987; Chandrasekaran and Kabadi 1988]. Bisubmodularity also arises in bicooperative games [Bilbao et al. 2008] as well as variants of sensor placement problems and coupled feature selection problems [Singh et al. 2012]. The minimization problem for bisubmodular functions using the ellipsoid method was solved in [Qi 1988]. More recently, combinatorial [Fujishige and Iwata 2005] and strongly combinatorial [McCormick and Fujishige 2010] algorithms for maximizing bisubmodular functions have been developed.

In this paper, we study the natural generalization of submodular and bisubmodular functions: given a natural number $k \geq 1$ and a finite nonempty set U , a function $f : (k+1)^U \rightarrow \mathbb{R}_+$ defined on k -tuples of pairwise disjoint subsets of U is called *k -submodular* if for all k -tuples $S = (S_1, \dots, S_k)$ and $T = (T_1, \dots, T_k)$ of pairwise disjoint subsets of U ,

$$f(S) + f(T) \geq f(S \sqcap T) + f(S \sqcup T),$$

where we define

$$S \sqcap T = (S_1 \cap T_1, \dots, S_k \cap T_k),$$

and

$$S \sqcup T = ((S_1 \cup T_1) \setminus \bigcup_{i \in \{2, \dots, k\}} (S_i \cup T_i), \dots, (S_k \cup T_k) \setminus \bigcup_{i \in \{1, \dots, k-1\}} (S_i \cup T_i)).$$

Under this definition, 1-submodularity corresponds exactly to the standard notion of submodularity for set functions, and similarly 2-submodularity corresponds to bisubmodularity. (We note that Ando has used the term k -submodular to study a different class of functions [Ando 2002].)

1.1. Related work

The terminology for k -submodular functions was first introduced in [Huber and Kolmogorov 2012] but the concept has been studied previously in [Cohen et al. 2006]. The concept of k -submodularity is a special case of strong tree submodularity [Kolmogorov 2011] with the tree being a star on $k+1$ vertices.

To the best of our knowledge, it is not known whether the ellipsoid method can be employed for minimizing k -submodular functions for $k \geq 3$ (some partial results can be found in [Huber and Kolmogorov 2012]), let alone whether there is a (fully) combinatorial algorithm for minimizing k -submodular functions for $k \geq 3$. However, it has recently been shown that explicitly given k -submodular functions can be minimized in polynomial time [Thapper and Živný 2012]¹, and these results have proved useful in the design of fixed-parameter algorithms [Wahlström 2014].

Some results on maximizing special cases of bisubmodular functions have appeared in Singh, Guillery, and Bilmes [Singh et al. 2012], who showed that simple bisubmodular function can be represented as a matroid constraint and a single submodular function, thus enabling the use of existing algorithms in some special cases. Unfortunately they show that this approach may require that the submodular function take negative values and so the approach does not work in general. (We note that our definition of bisubmodularity corresponds to directed bisubmodularity in [Singh et al. 2012].)

A different generalization of bisubmodularity, called skew bisubmodularity, has proved important in classifying finite-valued CSPs on domains with three elements [Huber et al. 2014]; this result was then generalized by a complexity classification of finite-valued CSPs on domains of arbitrary size [Thapper and Živný 2013]. Explicitly given skew bisubmodular functions can be minimized efficiently by results of Thapper and Živný [Thapper and Živný 2012]. The general question of whether all bisubmodular, and, more generally, k -submodular functions can be approximately maximized was left open.

1.2. Contributions

Following the question by Lovász [Lovász 1983] of whether there are generalizations of submodularity that preserve some nice properties such as efficient optimization algorithms, we consider the class of functions that are submodular in every orthant and r -wise monotone (the precise definition is given in Section 2), which includes as special cases bisubmodular and k -submodular functions.

Specifically, we consider the problem of *maximizing* bisubmodular and, more generally, k -submodular functions in the *value oracle model*. We provide the first approximation guarantees for maximizing a general bisubmodular or k -submodular function.

In Section 3, we prove that for any $k \geq 2$, k -submodular functions are precisely the k -set functions that are submodular in every orthant and pairwise monotone, thus extending the result from [Ando et al. 1996] that showed this result for $k = 2$.

In Section 4, we show that the naive random algorithm that simply returns a random partition of the ground set U is $1/4$ -approximation for maximizing any bisubmodular function and a $1/k$ -approximation for maximizing a k -submodular function with $k \geq 3$. We also show that our analysis is tight.

In Section 5, we show that a simple greedy algorithm for maximizing k -set functions that are submodular in every orthant and r -wise monotone for some $1 \leq r \leq k$ achieves a factor of $1/(1+r)$. We also show that our analysis is tight. Consequently, this algorithm achieves a factor of $1/3$ for maximizing k -submodular functions.

In Section 6, we develop a randomized greedy algorithm for maximizing k -set functions that are submodular in every orthant and k -wise monotone. The algorithm is inspired by the algorithm of Buchbinder et al. [Buchbinder et al. 2012] for unconstrained

¹In fact, results in [Thapper and Živný 2012] imply that much larger classes of functions can be minimized in polynomial time, including as one special case functions that are (strong) tree submodular, which in turn includes k -submodular functions.

submodular maximization. We show that this algorithm approximates any such k -set function to a factor of $1/(1 + \sqrt{k/2})$.

Finally, in Section 7, we relate our results on bisubmodular functions and existing results on submodular functions via a known embedding of submodular functions into bisubmodular functions. Using this embedding we can translate inapproximability results for submodular function into analogous results for bisubmodular functions. Moreover, we show that the algorithm of Buchbinder et al. [Buchbinder et al. 2012] may be viewed as a special case of our algorithm applied to this embedding.

Recently, Iwata, Tanigawa, and Yoshida [Iwata et al. 2013] have independently obtained a $1/k$ -approximation algorithm for maximizing k -submodular functions. Here we improve this factor to $1/3$, while also considering several other algorithms and generalizations of k -submodular functions.

2. PRELIMINARIES

We denote by \mathbb{R}_+ the set of all non-negative real numbers. Let U be a ground set containing n elements and $k \geq 1$ be a fixed integer. We consider functions that assign a value in \mathbb{R}_+ to each partial assignment of the values $\{1, \dots, k\}$ to the elements of U . We can represent each such partial assignments as vectors \mathbf{x} in $\{0, \dots, k\}^U$, where we have $x_e = 0$ if element $e \in U$ is not assigned any value in $\{1, \dots, k\}$, and otherwise have x_e equal to the value assigned to e . It will be useful to consider the partial assignment obtained from another (possibly partial) assignment \mathbf{x} by “forgetting” the values assigned to all elements except for some specified set $S \subseteq U$. We represent this as the vector $\mathbf{x}|_S$ whose coordinates are given by $(\mathbf{x}|_S)_e = x_e$, for all $e \in S$ and $(\mathbf{x}|_S)_e = 0$ for all $e \in U \setminus S$. Note that $\mathbf{x}|_S$ is similar to the projection of \mathbf{x} onto S , but we here require that all coordinates $e \notin S$ be set to 0, while the standard notion of projection would remove these coordinates from the resulting vector. In particular, this means that $\mathbf{x}|_S$ and \mathbf{x} both have n coordinates.

In order to relate our results to existing work on submodular functions, we shall also use terminology from set functions. In this setting, we consider k -set functions, which assign a value to each tuple of k disjoint sets $S = (S_1, \dots, S_k)$, where $S_i \subseteq U$ and $S_i \cap S_j = \emptyset$ for all $i \neq j$. It is straightforward to check that the two notions are equivalent by having $e \in S_i$ if and only if $x_e = i$. Then, we have $x_e = 0$ if and only if e does not appear in any of the sets S_1, \dots, S_k .

The solution space over which we optimize our functions is thus the set of partitions of some subset $U' \subseteq U$ into k disjoint sets, where in our vector notation U' is equivalent to the set of coordinates in \mathbf{x} that are non-zero. We shall refer to a partition of the entire ground set U as an *orthant* of U , and use the word *partial solution* to refer to a partition of some subset of U , to emphasize that they may not necessarily assign every element in U to a set. Given a partial solution \mathbf{s} and an orthant \mathbf{t} , we say that \mathbf{s} is in orthant \mathbf{t} if $\mathbf{s} = \mathbf{t}|_A$ for some set $A \subseteq U$. That is, \mathbf{s} is in orthant \mathbf{t} if and only if \mathbf{s} agrees with \mathbf{t} on all non-zero values.

Consider the operations \min_0 and \max_0 given by

$$\min_0(s, t) \stackrel{\text{def}}{=} \begin{cases} 0, & s \neq 0, t \neq 0, s \neq t \\ \min(s, t), & \text{otherwise} \end{cases}$$

and

$$\max_0(s, t) \stackrel{\text{def}}{=} \begin{cases} 0, & s \neq 0, t \neq 0, s \neq t \\ \max(s, t), & \text{otherwise,} \end{cases}$$

where $\min(s, t)$ (respectively, $\max(s, t)$) returns the smaller (respectively, the larger) of s and t with respect to the usual order on the integers. Then, for vectors \mathbf{s} and \mathbf{t} in $\{0, \dots, k\}^U$ we let $\min_0(\mathbf{s}, \mathbf{t})$ (respectively, $\max_0(\mathbf{s}, \mathbf{t})$) denote the vector obtained from applying \min_0 (respectively, \max_0) to \mathbf{s} and \mathbf{t} coordinate-wise. Using these operations we can define the general class of k -submodular functions:

Definition 2.1. Given a natural number $k \geq 1$ and a finite nonempty set U , a function $f : \{0, \dots, k\}^U \rightarrow \mathbb{R}_+$ is called k -submodular if for all \mathbf{s} and \mathbf{t} in $\{0, \dots, k\}^U$,

$$f(\mathbf{s}) + f(\mathbf{t}) \geq f(\min_0(\mathbf{s}, \mathbf{t})) + f(\max_0(\mathbf{s}, \mathbf{t})). \quad (1)$$

Note that if \mathbf{s} and \mathbf{t} are both orthants, then we have $\min_0(\mathbf{s}, \mathbf{t}) = \max_0(\mathbf{s}, \mathbf{t}) = \text{id}_0(\mathbf{s}, \mathbf{t})$, where the operation id_0 on each coordinate of \mathbf{s} and \mathbf{t} is given by $\text{id}_0(s, t) = s = t$ if $s = t$, and $\text{id}_0(s, t) = 0$ otherwise. Thus, if f is a k -submodular function, we have

$$f(\mathbf{s}) + f(\mathbf{t}) \geq 2f(\text{id}_0(\mathbf{s}, \mathbf{t})) \quad (2)$$

for any two orthants \mathbf{s} and \mathbf{t} of U .

Example 2.2. The well-known Max-Cut problem demonstrates that maximizing (1)-submodular functions is NP-hard, even if the objective function is given explicitly [Garey and Johnson 1979]. We show that the same hardness result holds for any $k \geq 1$. Consider the function $f^{(u,v)} : \{0, \dots, k\}^{\{u,v\}} \rightarrow \mathbb{R}_+$ given by² $f^{(u,v)}(x_u, x_v) = \llbracket x_u \neq x_v \rrbracket$. It is easy to check that $f^{(u,v)}$ is k -submodular. Given a graph (V, E) with $V = \{1, \dots, n\}$, we consider the function $f(\mathbf{x}) = \sum_{\{i,j\} \in E} f^{(i,j)}(x_i, x_j)$. Because f is a positive combination of k -submodular functions, it is also k -submodular. Moreover, maximizing f amounts to solving the Max- k -Cut problem, which is NP-hard [Papadimitriou and Yannakakis 1991].

While concise, Definition 2.1 gives little intuition in the traditional setting of set functions. We now consider this setting in order to provide some intuition. Consider two partial solutions $S = (S_1, \dots, S_k)$ and $T = (T_1, \dots, T_k)$ and let \mathbf{s} and \mathbf{t} be the vectors in $\{0, \dots, k\}^U$ representing S and T , respectively. Consider some element $e \in U$. We have $\min_0(s_e, t_e) = i \neq 0$ precisely when $s_e = t_e = i \neq 0$. Thus, the vector $\min_0(\mathbf{s}, \mathbf{t})$ in Definition 2.1 corresponds exactly to the coordinate-wise intersection $(S_1 \cap T_1, \dots, S_k \cap T_k)$ of S and T . Similarly, $\max_0(s_e, t_e) = i \neq 0$ precisely when either $s_e = t_e \neq 0$ or when one of s_e, t_e is $i \neq 0$ and the other is 0. Thus, the vector $\max_0(\mathbf{s}, \mathbf{t})$ corresponds exactly to the coordinate-wise union of S and T after we have removed any element e occurring in two different sets in S and T . That is, if we set $X_{-i} = \bigcup_{j \neq i} (S_j \cup T_j)$, then $\max_0(\mathbf{s}, \mathbf{t})$ corresponds to $((S_1 \cup T_1) \setminus X_{-1}, \dots, (S_k \cup T_k) \setminus X_{-k})$. The removal of X_{-i} from the i th union effectively enforces the condition that no element occurs in two different sets in the resulting partial solution.

The following equivalences, first observed by Cohen et al. [Cohen et al. 2006], allow us to relate k -submodular functions to existing families of set functions. When $k = 2$, Definition 2.1 requires that

$$f(S_1, S_2) + f(T_1, T_2) \geq f(S_1 \cap T_1, S_2 \cap T_2) + f((S_1 \cup T_1) \setminus (S_2 \cup T_2), (S_2 \cup T_2) \setminus (S_1 \cup T_2)),$$

which agrees exactly with the definition of bisubmodular functions given in [Fujishige 2005]. When $k = 1$, there is only a single set in each partial solution, and hence a single non-zero value in each corresponding vector, and so $X_{-1} = \emptyset$. Thus, Definition 2.1 requires that

$$f(S_1) + f(T_1) \geq f(S_1 \cap T_1) + f(S_1 \cup T_1),$$

²Here and throughout, we employ the Iverson bracket notation $\llbracket p \rrbracket$ to denote a value that is 1 when statement p is true and 0 when p is false.

which agrees exactly with the standard definition of submodular functions [Nemhauser and Wolsey 1988].

It is well-known that for standard set functions submodularity is equivalent to the property of *diminishing marginal returns*. Let $f : 2^U \rightarrow \mathbb{R}_+$ be a set function on U and define the marginal value of e with respect to S as $f_e(S) \stackrel{\text{def}}{=} f(S \cup \{e\}) - f(S)$ for all $S \subseteq U$ and $e \notin S$. Then, f is submodular if and only if

$$f_e(A) \geq f_e(B)$$

for all $A \subseteq B$ and $e \notin B$.

We shall see that marginal returns also play an important role in characterizing k -submodular functions. In this setting, however, we must specify not only which element we are adding to the solution, but which set in the partition it is being added to. For a k -set function $f : \{0, \dots, k\}^U \rightarrow \mathbb{R}_+$, an element $e \in U$, and a value $i \in \{1, \dots, k\}$, we define the marginal value $f_{i,e}(S)$ by

$$f_{i,e}(S) \stackrel{\text{def}}{=} f(S_1, \dots, S_{i-1}, S_i \cup \{e\}, S_{i+1}, \dots, S_k) - f(S_1, \dots, S_k)$$

for any partial solution $S = (S_1, \dots, S_k)$ such that $e \notin S_i$ for any i . Equivalently, in vector notation, we have

$$f_{i,e}(s) \stackrel{\text{def}}{=} f(s + i \cdot \mathbf{1}_e) - f(s),$$

where s is any partial solution satisfying $s_e = 0$, and $\mathbf{1}_e$ denotes the unit vector that is 1 in coordinate e and 0 in all other coordinates.

Definition 2.3. Let $k \geq 1$, and $1 \leq r \leq k$. We say that a function $f : \{0, \dots, k\}^U \rightarrow \mathbb{R}_+$ is:

- *submodular in every orthant*, if for any two partial solutions \mathbf{a} and \mathbf{b} in the same orthant of U , $f(\mathbf{a}) + f(\mathbf{b}) \geq f(\min_0(\mathbf{a}, \mathbf{b})) + f(\max_0(\mathbf{a}, \mathbf{b}))$.
- *r -wise monotone*, if for any element e , any partial solution s with $s_e = 0$, and any set of r distinct values $I \in \binom{\{1, \dots, k\}}{r}$:

$$\sum_{i \in I} f_{i,e}(s) \geq 0.$$

We remark that the case of $k = r = 1$ corresponds to monotone submodular functions. In the case of $k = r = 2$, Ando, Fujishige, and Naito [Ando et al. 1996] have shown that these two properties give an exact characterization of the class of bisubmodular functions. In Section 3, we extend their result by showing that submodularity in every orthant and pairwise monotonicity in fact precisely characterize k -submodular functions for all $k \geq 2$.

Let us now give some justification for the terminology “submodular in every orthant.” Let \mathbf{x} be an orthant of U . Given a k -submodular function f , we call set function $h : 2^U \rightarrow \mathbb{R}_+$ defined for any $S \subseteq U$ by

$$h(S) \stackrel{\text{def}}{=} f(\mathbf{x}|_S)$$

the function *induced by \mathbf{x} and f* . In the language of set functions, the function h is obtained by first assigning each element e in U to a single set X_i (where $i = x_e$). Then, $h(S)$ is simply the value of $f(S \cap X_1, \dots, S \cap X_k)$. We now show f is k -submodular in an orthant (in the sense of Definition 2.3) if and only if the function h induced by this orthant and f is submodular.

LEMMA 2.4. *Let (X_1, \dots, X_k) be an orthant of U , with vector representation \mathbf{x} . Then, f is k -submodular in the orthant \mathbf{x} if and only if the function h induced by \mathbf{x} and f is submodular.*

PROOF. Let A and B be two subsets of U , with associated partial solutions $\mathbf{a} = \mathbf{x}|_A$ and $\mathbf{b} = \mathbf{x}|_B$ in orthant \mathbf{x} . Then, note that $e \in A \cap B$ if and only if $\min(a_e, b_e)$ is non-zero, and $e \in A \cup B$ if and only if $\max(a_e, b_e)$ is non-zero. Moreover, since \mathbf{a} and \mathbf{b} agree on all non-zero coordinates, we have $\min_0(\mathbf{a}, \mathbf{b}) = \min(\mathbf{a}, \mathbf{b})$ and $\max_0(\mathbf{a}, \mathbf{b}) = \max(\mathbf{a}, \mathbf{b})$. Hence,

$$\begin{aligned} h(A \cup B) &= f(\mathbf{x}|_{A \cup B}) = f(\max(\mathbf{x}|_A, \mathbf{x}|_B)) = f(\max_0(\mathbf{x}|_A, \mathbf{x}|_B)) = f(\max_0(\mathbf{a}, \mathbf{b})), \\ h(A \cap B) &= f(\mathbf{x}|_{A \cap B}) = f(\min(\mathbf{x}|_A, \mathbf{x}|_B)) = f(\min_0(\mathbf{x}|_A, \mathbf{x}|_B)) = f(\min_0(\mathbf{a}, \mathbf{b})). \end{aligned}$$

Thus, we have

$$h(A) + h(B) \geq h(A \cap B) + h(A \cup B)$$

for any $A, B \subseteq U$ if and only if

$$f(\mathbf{a}) + f(\mathbf{b}) \geq f(\min_0(\mathbf{a}, \mathbf{b})) + f(\max_0(\mathbf{a}, \mathbf{b}))$$

for the associated partial solutions \mathbf{a}, \mathbf{b} in orthant \mathbf{x} . \square

Many of our proofs will use this connection between the standard notion of submodularity and the k -set functions in Definition 2.1. Specifically, we shall make use of the following result from Lee, Sviridenko, and Vondrák [Lee et al. 2010], which we restate here.

LEMMA 2.5 ([LEE ET AL. 2010, LEMMA 1.1]). *Let f be a non-negative submodular function on U . Let $S, C \subseteq U$ and let $\{T_\ell\}_{\ell=1}^t$ be a collection of subsets of $C \setminus S$ such that each element of $C \setminus S$ appears in exactly p of these subsets. Then*

$$\sum_{\ell=1}^t [f(S \cup T_\ell) - f(S)] \geq p[f(S \cup C) - f(S)].$$

In fact, the following weaker statement will be sufficient for our purposes:

COROLLARY 2.6 (OF LEMMA 2.5). *Let f be a non-negative submodular function on U . Let $S, C \subseteq U$ and let $\{T_\ell\}_{\ell=1}^t$ be a collection of subsets of $C \setminus S$ such that each element of $C \setminus S$ appears in exactly p of these subsets. Then*

$$\sum_{\ell=1}^t f(S \cup T_\ell) \geq pf(S \cup C).$$

PROOF. Add $\sum_{\ell=1}^t f(S)$ to each side of the inequality in Lemma 2.5. This gives

$$\begin{aligned} \sum_{\ell=1}^t f(S \cup T_\ell) &\geq p \cdot f(S \cup C) - p \cdot f(S) + \sum_{\ell=1}^t f(S) \\ &= p \cdot f(S \cup C) + (t - p) \cdot f(S) \\ &\geq p \cdot f(S \cup C), \end{aligned}$$

since $p \leq t$. \square

3. CHARACTERIZATION OF K -SUBMODULARITY

THEOREM 3.1. *Let $f : \{0, \dots, k\}^U \rightarrow \mathbb{R}_+$ be a k -set function, where $k \geq 2$. Then, f is k -submodular if and only if f is submodular in every orthant and pairwise monotone.*

In order to prove Theorem 3.1, we shall make use of the following lemma, which allows us to generalize pairwise monotonicity to solutions that disagree on the placement of multiple elements e .

LEMMA 3.2. *Let $k \geq 2$ and suppose that $f : \{0, \dots, k\}^U \rightarrow \mathbb{R}_+$ is submodular in every orthant and pairwise monotone. Let \mathbf{a} and \mathbf{b} in $\{0, \dots, k\}^U$ satisfy $0 \neq a_e \neq b_e \neq 0$ for all $e \in I$ and $a_e = b_e$ for all $e \in U \setminus I$, and define $\mathbf{c} = \mathbf{a}|_{U \setminus I} = \mathbf{b}|_{U \setminus I}$. Then, $f(\mathbf{a}) + f(\mathbf{b}) \geq 2f(\mathbf{c})$.*

PROOF. The proof is by induction on the size of I . In the case that $|I| = 0$, the claim is trivial. Suppose, then, that $|I| = p > 0$ and so I contains at least 1 element e . We can represent \mathbf{a} and \mathbf{b} as $\mathbf{a} = \mathbf{c} + \mathbf{x}$, and $\mathbf{b} = \mathbf{c} + \mathbf{y}$ where \mathbf{x} and \mathbf{y} are vectors in $\{0, \dots, 1\}^U$ satisfying $0 \neq x_e \neq y_e \neq 0$ for all $e \in I$, and $x_e = y_e = 0$ for all $e \in U \setminus I$.

Let $e \in I$ be some element on which \mathbf{a} and \mathbf{b} disagree. We define $\bar{\mathbf{x}} = \mathbf{x}|_{I \setminus \{e\}}$, $\bar{\mathbf{y}} = \mathbf{y}|_{\{e\}}$, and $\mathbf{z} = \bar{\mathbf{x}} + \bar{\mathbf{y}}$. Then, we have

$$f(\mathbf{a}) + f(\mathbf{b}) = f(\mathbf{c} + \mathbf{x}) + f(\mathbf{c} + \mathbf{y}) = [f(\mathbf{c} + \mathbf{x}) + f(\mathbf{c} + \mathbf{z})] + [f(\mathbf{c} + \mathbf{y}) + f(\mathbf{c} + \mathbf{z})] - 2f(\mathbf{c} + \mathbf{z}). \quad (3)$$

The solutions $\mathbf{c} + \mathbf{x}$ and $\mathbf{c} + \mathbf{z}$ disagree on precisely the single element e in I and are non-zero for this element. Thus, by the induction hypothesis

$$f(\mathbf{c} + \mathbf{x}) + f(\mathbf{c} + \mathbf{z}) \geq 2f(\mathbf{c} + \bar{\mathbf{x}}). \quad (4)$$

Similarly, $\mathbf{c} + \mathbf{y}$ and $\mathbf{c} + \mathbf{z}$ disagree on precisely those $p - 1$ elements in $I \setminus \{e\}$ and are non-zero for these elements. Thus, by the induction hypothesis

$$f(\mathbf{c} + \mathbf{y}) + f(\mathbf{c} + \mathbf{z}) \geq 2f(\mathbf{c} + \bar{\mathbf{y}}). \quad (5)$$

Combining (3), (4), and (5) we obtain

$$f(\mathbf{a}) + f(\mathbf{b}) \geq 2f(\mathbf{c} + \bar{\mathbf{x}}) + 2f(\mathbf{c} + \bar{\mathbf{y}}) - 2f(\mathbf{c} + \mathbf{z}). \quad (6)$$

Now, we note that $\mathbf{c} + \bar{\mathbf{x}}$ and $\mathbf{c} + \bar{\mathbf{y}}$ are both in the orthant $\mathbf{c} + \mathbf{z}$. Thus, from submodularity in every orthant,

$$f(\mathbf{c} + \bar{\mathbf{x}}) + f(\mathbf{c} + \bar{\mathbf{y}}) \geq f(\min_0(\mathbf{c} + \bar{\mathbf{x}}, \mathbf{c} + \bar{\mathbf{y}})) + f(\max_0(\mathbf{c} + \bar{\mathbf{x}}, \mathbf{c} + \bar{\mathbf{y}})) = f(\mathbf{c}) + f(\mathbf{c} + \mathbf{z}). \quad (7)$$

Combining (6) and (7) we obtain

$$f(\mathbf{a}) + f(\mathbf{b}) \geq 2f(\mathbf{c}) + 2f(\mathbf{c} + \mathbf{z}) - 2f(\mathbf{c} + \mathbf{z}) = 2f(\mathbf{c}).$$

□

We now return to the proof of the Theorem 3.1.

PROOF OF THEOREM 3.1. We begin by showing that necessity of the two properties. Suppose that f is k -submodular. Then, submodularity in every orthant follows directly from (1). For pairwise monotonicity, let \mathbf{s} satisfy $s_e = 0$. Consider any pair of distinct values i, j from $\{1, \dots, k\}$, and let $\mathbf{s}^i = \mathbf{s} + i \cdot \mathbf{1}_e$ and $\mathbf{s}^j = \mathbf{s} + j \cdot \mathbf{1}_e$. Then,

$$\begin{aligned} f_{i,e}(\mathbf{s}) + f_{j,e}(\mathbf{s}) &= f(\mathbf{s}^i) - f(\mathbf{s}) + f(\mathbf{s}^j) - f(\mathbf{s}) \\ &\geq f(\min_0(\mathbf{s}^i, \mathbf{s}^j)) + f(\max_0(\mathbf{s}^i, \mathbf{s}^j)) - 2f(\mathbf{s}) \\ &= f(\mathbf{s}) + f(\mathbf{s}) - 2f(\mathbf{s}). \end{aligned}$$

We now show that submodularity in every orthant and pairwise monotonicity imply k -submodularity. Let f be a function that is submodular in every orthant and pairwise monotone, and consider two arbitrary vectors \mathbf{x} and \mathbf{y} in $\{0, \dots, k\}^U$. Let I be the set of all elements $e \in U$ for which $x_e \neq 0$, $y_e \neq 0$ and $x_e \neq y_e$. We can write

$$f(\mathbf{x}) + f(\mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y}|_{U \setminus I}) + f(\mathbf{y}) + f(\max_0(\mathbf{x}, \mathbf{y})) - f(\mathbf{y}|_{U \setminus I}) - f(\max_0(\mathbf{x}, \mathbf{y})). \quad (8)$$

We note that \mathbf{x} and $\mathbf{y}|_{U \setminus I}$ are in the same orthant, since they agree on all non-zero coordinates. Thus,

$$\begin{aligned} f(\mathbf{x}) + f(\mathbf{y}|_{U \setminus I}) &\geq f(\min_0(\mathbf{x}, \mathbf{y}|_{U \setminus I})) + f(\max_0(\mathbf{x}, \mathbf{y}|_{U \setminus I})) \\ &= f(\min_0(\mathbf{x}, \mathbf{y})) + f(\max_0(\mathbf{x}, \mathbf{y}) + \mathbf{x}|_I), \end{aligned} \quad (9)$$

where in the final equation we have used the fact that for all $e \in I$, $x_e \neq 0$, $y_e \neq 0$ and $x_e \neq y_e$ and so $\min_0(x_i, y_i) = \max_0(x_i, y_i) = 0$. Similarly, we have \mathbf{y} and $\max_0(\mathbf{x}, \mathbf{y})$ in the same orthant, and so

$$\begin{aligned} f(\mathbf{y}) + f(\max_0(\mathbf{x}, \mathbf{y})) &\geq f(\min_0(\mathbf{y}, \max_0(\mathbf{x}, \mathbf{y}))) + f(\max_0(\mathbf{y}, \max_0(\mathbf{x}, \mathbf{y}))) \\ &= f(\mathbf{y}|_{U \setminus I}) + f(\max_0(\mathbf{x}, \mathbf{y}) + \mathbf{y}|_I). \end{aligned} \quad (10)$$

Combining (8), (9), and (10), we obtain

$$f(\mathbf{x}) + f(\mathbf{y}) \geq f(\min_0(\mathbf{x}, \mathbf{y})) + f(\max_0(\mathbf{x}, \mathbf{y}) + \mathbf{x}|_I) + f(\max_0(\mathbf{x}, \mathbf{y}) + \mathbf{y}|_I) - f(\max_0(\mathbf{x}, \mathbf{y})). \quad (11)$$

Finally, from Lemma 3.2 we have:

$$f(\max_0(\mathbf{x}, \mathbf{y}) + \mathbf{x}|_I) + f(\max_0(\mathbf{x}, \mathbf{y}) + \mathbf{y}|_I) \geq 2f(\max_0(\mathbf{x}, \mathbf{y})). \quad (12)$$

Combining (11) and (12) then gives

$$f(\mathbf{x}) + f(\mathbf{y}) \geq f(\min_0(\mathbf{x}, \mathbf{y})) + f(\max_0(\mathbf{x}, \mathbf{y})).$$

□

We now provide an example of a natural class of k -set functions which are submodular in every orthant and k -wise monotone but not k -submodular.

Example 3.3. Let $f^{(u,v)} : \{0, \dots, k\}^{\{u,v\}} \rightarrow \mathbb{R}_+$ be given by:

$$f^{(u,v)}(x_u, x_v) = \begin{cases} 0, & x_u = x_v = 0 \\ \frac{1}{k} \sum_{i=1}^k \llbracket x_u < i \rrbracket = \frac{k-x_u}{k}, & x_u \neq 0, x_v = 0 \\ \frac{1}{k} \sum_{i=1}^k \llbracket i < x_v \rrbracket = \frac{x_v-1}{k}, & x_u = 0, x_v \neq 0 \\ \llbracket x_u < x_v \rrbracket, & \text{otherwise.} \end{cases}$$

The function $f^{(u,v)}$ has the following intuitive interpretation: we begin with the valued constraint $\llbracket x_u < x_v \rrbracket$, where x_u and x_v range over $\{1, \dots, k\}$. This gives a function that is defined on all orthants. We extend the function to partial assignments by setting $f^{(u,v)}(0, 0) = 0$, and otherwise assigning $f^{(u,v)}(x_u, 0)$ and $f^{(u,v)}(0, x_v)$ the probability that $x_u > i$ and $i > x_v$, respectively, when i is chosen uniformly at random from $\{1, \dots, k\}$.

The function $f^{(u,v)}$ arises in the following graph layout problem: we are given a directed graph $G = (V, E)$ and a number k , and we wish to partition V into k layers so that as many directed edges as possible travel from a lower- to a higher-numbered layer. This problem is equivalent to maximizing the function $f(\mathbf{x}) : \{0, \dots, k\}^V \rightarrow \mathbb{R}_+$ given by $f(\mathbf{x}) = \sum_{(u,v) \in E} f^{(u,v)}(x_u, x_v)$ Although this function allows some vertices to

remain unassigned, k -wise monotonicity implies that there is always a maximizer of f that is an orthant.

We now show that $f^{(u,v)}$ is submodular in every orthant and k -wise monotone. Fix an orthant $(x_u = i, x_v = j)$, where $i, j \in \{1, \dots, k\}$, and let h be the submodular function induced by $f^{(u,v)}$ and this orthant. If $i \geq j$, we have

$$\begin{aligned} h_u(\emptyset) &= h(\{u\}) - h(\emptyset) = \frac{k-i}{k} & h_v(\emptyset) &= h(\{v\}) - h(\emptyset) = \frac{j-1}{k} \\ h_u(\{v\}) &= h(\{u, v\}) - h(\{v\}) = -\frac{j-1}{k} & h_v(\{u\}) &= h(\{u, v\}) - h(\emptyset) = -\frac{k-i}{k}, \end{aligned}$$

while if $i < j$ (and hence $i \leq j-1$), we have:

$$\begin{aligned} h_u(\emptyset) &= h(\{u\}) - h(\emptyset) = \frac{k-i}{k} = 1 - \frac{i}{k} & h_v(\emptyset) &= h(\{v\}) - h(\emptyset) = \frac{j-1}{k} \geq \frac{i}{k} \\ h_u(\{v\}) &= h(\{u, v\}) - h(\{v\}) = 1 - \frac{j-1}{k} \leq 1 - \frac{i}{k} \\ h_v(\{u\}) &= h(\{u, v\}) - h(\emptyset) = 1 - \frac{k-i}{k} = \frac{i}{k}. \end{aligned}$$

In all cases, we observe that the marginals of h are decreasing, and so h is a submodular function.

In order to show that $f^{(u,v)}$ is k -wise monotone, we note that $f_{i,e}^{(u,v)}(0,0)$ is non-negative for all values of i and e , and so $\sum_{i=1}^k f_{i,e}^{(u,v)}(0,0) \geq 0$ for all $e \in \{u, v\}$. For the remaining marginals, suppose that $j \neq 0$. Then, for we have

$$\begin{aligned} \sum_{i=1}^k f_{i,u}^{(u,v)}(0, j) &= \sum_{i=1}^k \left[\mathbb{I}[i < j] - \frac{1}{k} \sum_{p=1}^j \mathbb{I}[p < j] \right] = \sum_{i=1}^k \mathbb{I}[i < j] - \sum_{p=1}^k \mathbb{I}[p < j] = 0, \\ \sum_{i=1}^k f_{i,v}^{(u,v)}(j, 0) &= \sum_{i=1}^k \left[\mathbb{I}[j < i] - \frac{1}{k} \sum_{p=1}^j \mathbb{I}[j < p] \right] = \sum_{i=1}^k \mathbb{I}[j < i] - \sum_{p=1}^k \mathbb{I}[j < p] = 0. \end{aligned}$$

4. THE NAIVE RANDOM ALGORITHM

We now consider the performance of the naive random algorithm for maximizing a k -submodular function $f : \{0, \dots, k\}^U \rightarrow \mathbb{R}_+$. Note that pairwise monotonicity of f , guaranteed by Theorem 3.1, implies that any partial solution $S \in \{0, \dots, k\}^U$ can be extended greedily to an orthant of U without any loss in the value of f , since for every element $e \notin S$, we must have $f_{i,e}(S) \geq 0$ for some $i \in \{1, \dots, k\}$. Thus, we may assume without loss of generality that f takes its maximum value on some orthant \mathfrak{o} . We now consider the expected performance of a random algorithm that simply selects an orthant of U uniformly at random.

THEOREM 4.1. *Let $f : \{0, \dots, k\}^U \rightarrow \mathbb{R}_+$ be a k -submodular function attaining its maximum value on orthant \mathfrak{o} , and let \mathbf{x} be an orthant of U selected uniformly at random. Then, $\mathbb{E}[f(\mathbf{x})] \geq \frac{1}{4}f(\mathfrak{o})$ if $k = 2$, and $\mathbb{E}[f(\mathbf{x})] \geq \frac{1}{k}f(\mathfrak{o})$ if $k \geq 3$.*

We present the analysis for the case in which $k \geq 3$ first, as it is simpler and will aid in motivating some of the constructions used for the case $k = 2$.

4.1. Analysis for $k \geq 3$

Let $h : 2^U \rightarrow \mathbb{R}_+$ be the submodular function induced by \mathfrak{o} and f . For each $e \in U$ we consider a fixed permutation π_e on the set $\{1, \dots, k\}$ with the property that $\pi_e(o_e) =$

o_e and $\pi_e(z) \neq z$ for all $z \in \{1, \dots, k\} \setminus \{o_e\}$.³ Then, we denote by $\pi(\mathbf{x})$ the vector $(\pi_e(x_e))_{e \in U}$.

Let $P(A)$ be the set of orthants of U that agree with \mathbf{o} on exactly those coordinates $e \in A$. The following lemma allows us to relate the sum of the values of all partitions in $P(A)$ to the value of \mathbf{o} .

LEMMA 4.2. *For each set $A \subseteq U$,*

$$\sum_{\mathbf{x} \in P(A)} f(\mathbf{x}) \geq (k-1)^{n-|A|} h(A).$$

PROOF. Consider the sum $\sum_{\mathbf{x} \in P(A)} f(\pi(\mathbf{x}))$. Because $\pi_e(x_e) = o_e$ if and only if $x_e = o_e$ already, we have $\pi(\mathbf{x}) \in P(A)$ if and only if $\mathbf{x} \in P(A)$. Then, because each π_e is a bijection, we have

$$\sum_{\mathbf{x} \in P(A)} f(\mathbf{x}) = \sum_{\mathbf{x} \in P(A)} f(\pi(\mathbf{x})),$$

and so,

$$\sum_{\mathbf{x} \in P(A)} f(\mathbf{x}) = \frac{1}{2} \left[\sum_{\mathbf{x} \in P(A)} f(\mathbf{x}) + \sum_{\mathbf{x} \in P(A)} f(\pi(\mathbf{x})) \right] = \frac{1}{2} \sum_{\mathbf{x} \in P(A)} [f(\mathbf{x}) + f(\pi(\mathbf{x}))]. \quad (13)$$

Now, we note that \mathbf{x} and $\pi(\mathbf{x})$ are both orthants. Thus, from (2) we have

$$f(\mathbf{x}) + f(\pi(\mathbf{x})) \geq 2\text{id}_0(\mathbf{x}, \pi(\mathbf{x})).$$

Consider an arbitrary coordinate $e \in U$. If $e \in A$ we have $x_e = o_e$ and so $\pi_e(x_e) = x_e$ and hence $\text{id}_0(x_e, \pi_e(x_e)) = x_e$. If $e \notin A$, then we have $x_e \neq o_e$ and so $\pi_e(x_e) \neq x_e$ and hence $\text{id}_0(x_e, \pi_e(x_e)) = 0$. Thus,

$$2\text{id}_0(\mathbf{x}, \pi(\mathbf{x})) = 2f(\mathbf{o}|_A) = 2h(A).$$

Combining this with (13) we have,

$$\sum_{\mathbf{x} \in P(A)} f(\mathbf{x}) = \frac{1}{2} \sum_{\mathbf{x} \in P(A)} [f(\mathbf{x}) + f(\pi(\mathbf{x}))] \geq \sum_{\mathbf{x} \in P(A)} h(A) = (k-1)^{n-|A|} h(A),$$

since there are precisely $k-1$ choices $i \neq o_e$ for x_e for each of the $n-|A|$ coordinates $e \notin A$. \square

We now complete the proof of Theorem 4.1 in the case $k \geq 3$. We formulate the expectation as

$$\mathbb{E}[f(\mathbf{x})] = \frac{1}{k^n} \sum_{i=0}^n \sum_{A \in \binom{U}{i}} \sum_{\mathbf{x} \in P(A)} f(\mathbf{x}).$$

Using Lemma 4.2 we obtain

$$\sum_{i=0}^n \sum_{A \in \binom{U}{i}} \sum_{\mathbf{x} \in P(A)} f(\mathbf{x}) \geq \sum_{i=0}^n \sum_{A \in \binom{U}{i}} (k-1)^{n-i} h(A). \quad (14)$$

³Such a permutation can be obtained by taking, for example, $\pi_e(o_e) = o_e$, $\pi_e(o_e-1) = o_e+1$, and $\pi(z) = z+1 \pmod k$ for all other $z \in \{1, \dots, k\}$.

Consider a fixed value $i \in \{0, \dots, n\}$. Each element $e \in U$ appears in exactly $\binom{n-1}{i-1}$ of the $\binom{n}{i}$ sets $A \in \binom{U}{i}$. Because h is submodular, Corollary 2.6 then implies that

$$\sum_{A \in \binom{U}{i}} h(A) \geq \binom{n-1}{i-1} h(U) = \binom{n-1}{i-1} f(\mathbf{o}). \quad (15)$$

Combining (14) and (15) with our formulation of $\mathbb{E}[f(\mathbf{x})]$ we obtain:

$$\begin{aligned} \mathbb{E}[f(\mathbf{x})] &\geq \frac{1}{k^n} \sum_{i=0}^n \binom{n-1}{i-1} (k-1)^{n-i} f(\mathbf{o}) \\ &= \frac{(k-1)^{n-1}}{k^n} \sum_{i=0}^n \binom{n-1}{i-1} (k-1)^{-(i-1)} f(\mathbf{o}) \\ &= \frac{(k-1)^{n-1}}{k^n} \sum_{i=0}^{n-1} \binom{n-1}{i} (k-1)^{-i} f(\mathbf{o}) \\ &= \frac{(k-1)^{n-1}}{k^n} \cdot \left(1 + \frac{1}{k-1}\right)^{n-1} \cdot f(\mathbf{o}) \\ &= \frac{(k-1)^{n-1}}{k^n} \cdot \frac{k^{n-1}}{(k-1)^{n-1}} \cdot f(\mathbf{o}) \\ &= \frac{1}{k} \cdot f(\mathbf{o}). \end{aligned}$$

4.2. Analysis for $k = 2$

Now we consider the case in which f is a bisubmodular function, i.e. the case of $k = 2$. In the previous analysis of k -submodular functions for $k \geq 3$ we used a bijection π_e on $\{1, \dots, k\}$ with the property that $\pi_e(o_e) = o_e$ and $\pi_e(z) \neq z$ for all $z \neq o_e$. However, when $k = 2$, no such bijection exists and we must adopt a different approach.

Suppose again that f attains its maximum on orthant $\mathbf{o} \in \{1, 2\}^U$. For a value $v \in \{1, 2\}$ we let $\bar{v} \stackrel{\text{def}}{=} (v \bmod 2) + 1$ (i.e. the other value in $\{1, 2\}$). Then, for any disjoint subsets A and B of U we define the (partial) solution $T(A, B)$ by

$$T(A, B)_i = \begin{cases} o_i, & i \in A \\ \bar{o}_i, & i \in B \\ 0, & \text{otherwise} \end{cases}.$$

It will simplify our analysis to work with with symmetrized values, which depend only on the sizes of the sets A and B chosen. We define

$$F_{i,j} = \binom{n}{i}^{-1} \binom{n-i}{j}^{-1} \sum_{A \in \binom{U}{i}} \sum_{B \in \binom{U \setminus A}{j}} [f(T(A, B))].$$

Then, $F_{i,j}$ gives the average value of f over all partial solutions on $i + j$ elements that agree with \mathbf{o} on exactly i and disagree with it on exactly j elements. In particular, we have $F_{n,0} = f(\mathbf{o})$, and $F_{i,n-i} = \binom{n}{i}^{-1} \sum_{A \in \binom{U}{i}} f(T(A, U \setminus A))$. Our next lemma relates these two values.

LEMMA 4.3. *For all i such that $0 \leq i \leq n$,*

$$F_{i,n-i} \geq \frac{i(i-1)}{n(n-1)} F_{n,0}. \quad (16)$$

PROOF. We prove 2 separate inequalities which together imply the lemma. First, we shall show that for all $1 \leq i \leq n-1$,

$$F_{i,n-i} \geq F_{i-1,n-i-1}. \quad (17)$$

We do this by showing that a related inequality holds for arbitrary sets of the appropriate size, and then average over all possible sets to obtain (17). Fix $1 \leq i \leq n-1$ and let A be any subset of U of size $i+1$. Set $B = U \setminus A$ and let x and y any two distinct elements in A . Consider the solutions $T(A-x, B+x)$ and $T(A-y, B+y)$ ⁴. They are both orthants and agree on all elements except x and y . Thus, from (2), the inequality

$$\begin{aligned} f(T(A-x, B+x)) + f(T(A-y, B+y)) &\geq 2\text{id}_0(T(A-x, B+x), T(A-y, B+y)) \\ &= 2f(T(A-x-y, B)) \end{aligned}$$

holds for any such choice of A , x , and y , where $|A| = i+1$ and $|B| = |U \setminus A| = n-i-1$. Averaging the resulting inequalities over all possible choices for A , $B = U \setminus A$, x , and y and dividing both sides by 2 then gives (17).

Next, we show that for any $1 \leq i \leq n-1$,

$$F_{i-1,n-i-1} \geq \frac{i-1}{i+1} F_{i+1,n-i-1}. \quad (18)$$

Again fix $i \geq 1$, let A be any subset of U of size $i+1$ and set $B = U \setminus A$. Let h be the submodular function induced by the orthant $T(A, B)$ and f . Note then, that we can express h as $h(X) = T(A \cap X, B \cap X)$. We consider the sum:

$$\sum_{C \in \binom{A}{2}} [f(T(A \setminus C, B)) - T(\emptyset, B)] = \sum_{C \in \binom{A}{2}} [h(U \setminus C) - h(B)]$$

Each element of A appears in exactly $\binom{|A|-1}{2} = \binom{i}{2}$ of the sets $U \setminus C$ above (one for each way to choose a two element set C from the remaining $|A|-1$ elements). Applying Corollary 2.6 we then obtain

$$\sum_{C \in \binom{A}{2}} h(U \setminus C) \geq \binom{i}{2} h(U) = \binom{i}{2} T(A, B).$$

Altogether, we obtain the inequality

$$\sum_{C \in \binom{A}{2}} f(T \setminus C, B) \geq \binom{i}{2} T(A, B),$$

valid for any choice of A , with $|A| = i+1$, and $|B| = |U \setminus A| = n-i-1$. Averaging the resulting inequalities over all possible choices for A , we obtain

$$\binom{i+1}{2} F_{i-1,n-i-1} \geq \binom{i}{2} F_{i+1,n-i-1},$$

which is equivalent to (18).

⁴Here, we employ the shorthand $A+x$ for $A \cup \{x\}$ and $A-x$ for $A \setminus \{x\}$.

Combining (17) and (18) then gives the symmetrized inequality

$$F_{i,n-i} \geq \frac{i-1}{i+1} F_{i+1,n-i-1}. \quad (19)$$

The desired inequality (16) then follows from reverse induction on i . If $i = n$, then (16) is trivial. For the inductive step, we suppose that $1 \leq i \leq n-1$. Then, applying (19) followed by the induction hypothesis gives

$$F_{i,n-i} \geq \frac{i-1}{i+1} F_{i+1,n-i-1} \geq \frac{i-1}{i+1} \cdot \frac{(i+1)i}{n(n-1)} F_{n,0} = \frac{i(i-1)}{n(n-1)} F_{n,0}.$$

If $i = 0$, we cannot apply (19). In this case, however, (16) follows directly from non-negativity of f . \square

We now complete the proof of Theorem 4.1 in the case that $k = 2$. We can formulate the expectation in terms of our symmetric notation as

$$\mathbb{E}[f(\mathbf{x})] = 2^{-n} \sum_{i=0}^n \sum_{A \in \binom{U}{i}} T(A, U \setminus A) = 2^{-n} \sum_{i=0}^n \binom{n}{i} F_{i,n-i}.$$

Then, we have

$$\begin{aligned} 2^{-n} \sum_{i=0}^n \binom{n}{i} F_{i,n-i} &\geq 2^{-n} \sum_{i=2}^n \binom{n}{i} F_{i,n-i} \\ &\geq 2^{-n} \sum_{i=2}^n \binom{n}{i} \frac{i(i-1)}{n(n-1)} F_{n,0} \\ &= 2^{-n} \sum_{i=2}^n \binom{n-2}{i-2} F_{n,0} \\ &= 2^{-n} \sum_{i=0}^{n-2} \binom{n-2}{i} F_{n,0} \\ &= 2^{-n} \cdot 2^{n-2} F_{n,0} \\ &= \frac{1}{4} f(\mathbf{o}), \end{aligned}$$

where the first inequality follows from non-negativity of f (and hence of F) and the second inequality follows from Lemma 4.3.

Example 4.4. As a tight example for $k = 2$, we consider the function $f^{(u,v)}$ defined as in Example 3.3 for the special case in which $k = 2$. Then, the resulting function is submodular in every orthant and 2-wise monotone and hence must be bisubmodular. Moreover, the probability that a random orthant will set $x_u = 1$, and $x_v = 2$ is $\frac{1}{4}$, and the function has value 0 for all other orthants. Thus, $\mathbb{E}[f^{(u,v)}(\mathbf{x})] = \frac{1}{4}$, whereas the maximum value is 1.

This example is easily extended to ground sets $U = \{u\} \cup V$ of arbitrary size, by setting $f(\mathbf{x}) = \sum_{v \in V} f^{(u,v)}(x_u, x_v)$. This function is also bisubmodular as it is a positive combination of bisubmodular functions. Moreover, the assignment setting $x_u = 1$ and $x_v = 2$ for all $v \in V$ has value $|V|$, but by linearity of expectation a uniform random assignment has expected value only $\frac{1}{4}|V|$.

Example 4.5. As a tight example for $k \geq 3$, we consider the single-argument k -submodular function $f^{(e)} : \{0, \dots, k\}^{\{e\}}$ given by $f(x_e) = \llbracket x_e = 1 \rrbracket$. It is easy to verify that this function is indeed k -submodular. Moreover, a uniform random assignment sets $x_e = 1$ with probability only $\frac{1}{k}$, and so $\mathbb{E}[f^{(e)}(x_e)] = \frac{1}{k}$. Similar to the previous example, we can generalize to an arbitrary ground set U by setting $f(\mathbf{x}) = \sum_{e \in U} f^{(e)}(x_e)$. We note also that the value 1 in the definition of each $f^{(e)}$ can be replaced by any value $p \in \{1, \dots, k\}$.

5. A DETERMINISTIC GREEDY ALGORITHM

In this section we consider a deterministic greedy algorithm for maximizing a k -set function $f : \{0, \dots, k\}^U \rightarrow \mathbb{R}_+$, that is submodular in every orthant and r -wise monotone for some $1 \leq r \leq k$, where $k \geq 2$. As a special case, we obtain an approximation algorithm for k -submodular functions.

The algorithm begins with the initial solution $\mathbf{s} = \mathbf{0}$ and considers elements of the ground set U in some arbitrary order, permanently setting $s_e = i$ for each element e , based on the increase that this gives in f . Specifically, the algorithm sets s_e to the value i that yields the largest marginal increase $f_{i,e}(S)$ in f with respect to the current solution \mathbf{s} . If there is more than one option we set s_e the smallest such i giving the maximal increase.

Algorithm 1 Deterministic Greedy

```

 $\mathbf{s} \leftarrow \mathbf{0}$ 
for each  $e \in U$  do
  for  $i = 1$  to  $k$  do
     $y_i \leftarrow f_{i,e}(\mathbf{s})$ 
   $y = \max(y_1, \dots, y_k)$ 
  Let  $q$  be the smallest value from  $\{1, \dots, k\}$  so that  $y_i = y$ .
   $s_e \leftarrow q$ 
return  $\mathbf{s}$ 

```

THEOREM 5.1. *Let \mathbf{s} be the solution produced by the deterministic greedy algorithm on some instance $f : \{0, \dots, k\}^U \rightarrow \mathbb{R}_+$ that is submodular in every orthant and r -wise monotone for some $1 \leq r \leq k$, and let \mathbf{o} be the optimal solution for this instance. Then,*

$$(1+r)f(\mathbf{s}) \geq f(\mathbf{o}).$$

PROOF. Our analysis considers 2 sequences of n solutions. First let, $\mathbf{s}^{(j)}$ be the algorithm's solution after j elements of U have been considered, and let $U^{(j)}$ be the set of elements that have been considered. Let $\mathbf{o}^{(j)} = \mathbf{o}|_{U \setminus U^{(j)}} + \mathbf{s}^{(j)}$ be a partial solution that agrees with $\mathbf{s}^{(j)}$ on the placement of the elements considered by the greedy algorithm in its first j phases and with \mathbf{o} on the placement of all other elements. Note that in particular we have $\mathbf{o}^{(0)} = \mathbf{o}$ and $\mathbf{o}^{(n)} = \mathbf{s}$. Our analysis of the greedy algorithm will bound the loss in $f(\mathbf{o}^{(j)})$ incurred at the each stage by the improvement in $\mathbf{s}^{(j)}$ made by the algorithm. In Lemma 5.2, we show that for every $0 \leq j \leq n$, $f(\mathbf{o}^{(j)}) - f(\mathbf{o}^{(j+1)}) \leq r[f(\mathbf{s}^{(j+1)}) - f(\mathbf{s}^{(j)})]$.

Summing this inequality from $j = 0$ to $n - 1$, we obtain

$$\sum_{j=0}^{n-1} [f(\mathbf{o}^{(j)}) - f(\mathbf{o}^{(j+1)})] \leq r \sum_{j=0}^{n-1} [f(\mathbf{s}^{(j+1)}) - f(\mathbf{s}^{(j)})].$$

Telescoping the summations on each side, we then have

$$f(\mathbf{o}^{(0)}) - f(\mathbf{o}^{(n)}) \leq r \left[f(\mathbf{s}^{(n)}) - f(\mathbf{s}^{(0)}) \right].$$

The theorem then follows immediately from the facts $\mathbf{o}^{(0)} = \mathbf{o}$, $\mathbf{o}^{(n)} = \mathbf{s}^{(n)} = \mathbf{s}$, and $\mathbf{s}^{(0)} \geq 0$. \square

It remains to show the following inequality.

LEMMA 5.2. *For $0 \leq j \leq n - 1$,*

$$f(\mathbf{o}^{(j)}) - f(\mathbf{o}^{(j+1)}) \leq r \left[f(\mathbf{s}^{(j+1)}) - f(\mathbf{s}^{(j)}) \right].$$

PROOF. Let e be the element considered in the $(j + 1)$ th phase of the algorithm.

We define the solution $\mathbf{t} = \mathbf{o}|_{U \setminus U^{(j+1)}} + \mathbf{s}^{(j)}$, and let $a_i = f_{i,e}(\mathbf{t})$ for $1 \leq i \leq k$. Then, we note that for any value i , $\mathbf{t} + i \cdot \mathbf{1}_e$ and $\mathbf{s}^{(j)} + i \cdot \mathbf{1}_e$ are in the same orthant. For some value i , let h be the submodular function induced by this orthant and f . Then h must be submodular, and so

$$y_i = f_{i,e}(\mathbf{s}^{(j)}) = h_e(U^{(j)}) \geq h_e(U \setminus \{e\}) = f_{i,e}(\mathbf{t}) = a_i.$$

Suppose that in the optimal solution we have $o_e = p$ but the greedy algorithm sets $s_e \leftarrow q$. Then, we observe that $f(\mathbf{o}^{(j)}) = f(\mathbf{t}) + f_{p,e}(\mathbf{t})$ and $f(\mathbf{o}^{(j+1)}) = f(\mathbf{t}) + f_{q,e}(\mathbf{t})$, and so

$$f(\mathbf{o}^{(j)}) - f(\mathbf{o}^{(j+1)}) = f_{p,e}(\mathbf{t}) - f_{q,e}(\mathbf{t}) = a_p - a_q.$$

Similarly,

$$f(\mathbf{s}^{(j+1)}) - f(\mathbf{s}^{(j)}) = f_{p,e}(\mathbf{s}^{(j)}) = y_j.$$

By r -wise monotonicity, for any $I \subseteq \{1, \dots, k\}$ with $|I| = r$ we have $\sum_{\ell \in I} a_\ell \geq 0$ and thus $-a_q \leq \sum_{\ell \in I \setminus \{q\}} a_\ell$. Therefore, $a_p - a_q \leq a_p + \sum_{\ell \in I \setminus \{q\}} a_\ell \leq r \cdot y_q$ as $a_i \leq y_i$ for every $1 \leq i \leq k$ and $y_q = \max(y_1, \dots, y_k)$. \square

Combining Theorem 3.1 and Theorem 5.1 gives us the following.

COROLLARY 5.3. *Let \mathbf{s} be the solution produced by the deterministic greedy algorithm for some k -submodular function $f : \{0, \dots, k\}^U \rightarrow \mathbb{R}_+$, and let \mathbf{o} be an optimal solution for this instance. Then,*

$$\frac{1}{3}f(\mathbf{s}) \geq f(\mathbf{o}).$$

The following is a tight example for Theorem 5.1.

Example 5.4. Let $0 \leq r \leq k$ and consider the function $f^{(u,v)} : \{0, \dots, k\}^{\{u,v\}} \rightarrow \mathbb{R}_+$ given by $f^{(u,v)}(x_u, x_v) = \frac{1}{r+1} \llbracket x_u \neq 0 \rrbracket + \frac{r}{r+1} \llbracket x_u \neq 1 \wedge x_v = 2 \rrbracket$. We shall first show that $f^{(u,v)}$ is submodular in every orthant and r -wise monotone.

Fix an orthant $(x_u = i, x_v = j)$ with $j \neq 2$, and let h be the function induced by $f^{(u,v)}$ and this orthant. Then, the marginals of h are given by:

$$\begin{aligned} h_u(\emptyset) &= h(\{u\}) - h(\emptyset) = \frac{1}{r+1} & h_v(\emptyset) &= h(\{v\}) - h(\emptyset) = 0 \\ h_u(\{v\}) &= h(\{u, v\}) - h(\{v\}) = \frac{1}{r+1} & h_v(\{u\}) &= h(\{u, v\}) - h(\{u\}) = 0. \end{aligned}$$

Now, fix an orthant ($x_u = i, x_v = 2$), and let h be the function induced by $f^{(u,v)}$ and this orthant. We have

$$\begin{aligned} h_u(\emptyset) &= h(\{u\}) - h(\emptyset) = \frac{1}{r+1} & h_v(\emptyset) &= h(\{v\}) - h(\emptyset) = \frac{r}{r+1} \\ h_u(\{v\}) &= h(\{u, v\}) - h(\{v\}) = \frac{1}{r+1} - \frac{r}{r+1} \mathbb{1}[i = 1] & h_v(\{u\}) &= h(\{u, v\}) - h(\{u\}) = \frac{r}{r+1} \mathbb{1}[i \neq 1]. \end{aligned}$$

In all cases, the marginals of h are decreasing, and so $f^{(u,v)}$ is submodular in every orthant. We now show that $f^{(u,v)}$ is r -wise monotone. The marginals of $f^{(u,v)}$ are non-negative, except the one obtained by setting x_u from 0 to 1 in the case that $x_v = 2$. Thus, the only non-trivial case is that in which $x_v = 2$, and I is a set of r distinct values with $1 \in I$. In this case,

$$\begin{aligned} \sum_{i \in I} [f_{i,u}^{(u,v)}(0, 2)] &= f^{(u,v)}(1, 2) - f^{(u,v)}(0, 2) + \sum_{i \in I \setminus \{1\}} [f^{(u,v)}(i, 2) - f^{(u,v)}(0, 2)] \\ &= \frac{1}{r+1} - \frac{r}{r+1} + (r-1) \cdot \frac{1}{r+1} = 0. \end{aligned}$$

Now, we analyze the performance of the deterministic greedy algorithm on $f^{(u,v)}$. We suppose, without loss of generality, that the algorithm considers u before v . When u is considered, we have $\mathbf{s} = \mathbf{0}$ and $f_{i,u}^{(u,v)}(0, 0) = \frac{1}{r+1}$ for all $i \in \{1, \dots, k\}$, and so the algorithm sets $s_u = 1$. In the next iteration, we have $f_{i,v}^{(u,v)}(1, 0) = 0$ for all values $i \in \{1, \dots, k\}$, and so the algorithm set $s_v = 1$ and returns $\mathbf{s} = (1, 1)$. We then have $f^{(u,v)}(\mathbf{s}) = \frac{1}{r+1}$, but $f^{(u,v)}(2, 2) = 1$. As in previous examples, we can easily obtain a function over ground sets of arbitrary size by summing the values of several different functions $f^{(u,v)}$.

6. A RANDOMIZED GREEDY ALGORITHM

In this section we consider the performance of a simple randomized greedy algorithm for maximizing a k -set function that is submodular in every orthant and k -wise monotone. Our algorithm is inspired by the algorithm of Buchbinder et al. [Buchbinder et al. 2012] for unconstrained submodular maximization. It begins with the initial solution $\mathbf{s} = \mathbf{0}$ and considers elements of the ground set U in some arbitrary order, permanently setting s_e to some value $i \in \{1, \dots, k\}$, based on the marginal increase in f that this yields. Specifically, the algorithm randomly sets $s_e = i$ with probability proportional to the resulting marginal increase $f_{i,e}(\mathbf{s})$ in f with respect to the current solution \mathbf{s} . If $f_{i,e}(\mathbf{s}) < 0$, we set $s_e = i$ with probability 0. Note that Theorem 3.1 shows that we cannot have $f_{i,e}(\mathbf{s}) < 0$ for all i , but it may be the case that $f_{i,e}(\mathbf{s}) = 0$ for all i . In this case, we set $s_e = 1$.

THEOREM 6.1. *Let $f : \{0, \dots, k\}^U$ be a k -set function that is submodular in every orthant and k -wise monotone, where $k \geq 2$. Let \mathbf{o} be orthant of U that maximizes f and let \mathbf{s} be the orthant produced by the randomized greedy algorithm. Then,*

$$\left(1 + \sqrt{\frac{k}{2}}\right) \mathbb{E}[f(\mathbf{s})] \geq f(\mathbf{o}).$$

PROOF. As in the analysis of the deterministic greedy algorithm, we consider 2 sequences of n solutions. Let $\mathbf{s}^{(j)}$, and $\mathbf{o}^{(j)}$ be defined as in the proof of Theorem 5.1, and note that \mathbf{s} (and hence each $\mathbf{s}^{(j)}$) is now a random variable depending on the random choices made by the algorithm. In Lemma 6.2, we bound the expected decrease $\mathbb{E}[f(\mathbf{o}^{(j)}) - f(\mathbf{o}^{(j+1)})]$ relative to the increase $\mathbb{E}[f(\mathbf{s}^{(j+1)}) - f(\mathbf{s}^{(j)})]$ in each iteration.

Algorithm 2 Randomized Greedy

```

s  $\leftarrow$  0
for each  $e \in U$  do
  for  $i = 1$  to  $k$  do
     $y_i \leftarrow \max(0, f_{i,e}(\mathbf{s}))$ 
   $\beta = \sum_{i=1}^k y_i$ 
  if  $\beta \neq 0$  then
    Let  $q \in \{1, \dots, k\}$  be chosen randomly, with  $\Pr[i = \ell] = \frac{y_\ell}{\beta}$  for all  $\ell \in \{1, \dots, k\}$ .
     $s_e \leftarrow q$ 
  else
     $s_e \leftarrow 1$ 
return  $\mathbf{s}$ 

```

Specifically, we show that

$$\mathbb{E}[f(\mathbf{o}^{(j)}) - f(\mathbf{o}^{(j+1)})] \leq \sqrt{\frac{k}{2}} \mathbb{E}[f(\mathbf{s}^{(j+1)}) - f(\mathbf{s}^{(j)})] \quad (20)$$

for all j . Summing the resulting inequalities for $j = 0$ to n , we then obtain

$$\sum_{j=0}^n \mathbb{E}[f(\mathbf{o}^{(j)}) - f(\mathbf{o}^{(j+1)})] \leq \sqrt{\frac{k}{2}} \sum_{j=0}^n \mathbb{E}[f(\mathbf{s}^{(j+1)}) - f(\mathbf{s}^{(j)})],$$

which simplifies to

$$\mathbb{E}[f(\mathbf{o}^{(0)})] - \mathbb{E}[f(\mathbf{o}^{(n)})] \leq \sqrt{\frac{k}{2}} (\mathbb{E}[f(\mathbf{s}^{(n)})] - \mathbb{E}[f(\mathbf{s}^{(0)})]) \leq \sqrt{\frac{k}{2}} \mathbb{E}[f(\mathbf{s}^{(n)})].$$

The theorem then follows from the definitions $\mathbf{o}^{(0)} = \mathbf{o}$, and $\mathbf{s}^{(n)} = \mathbf{o}^{(n)} = \mathbf{s}$. \square

We now show that inequality (20) must hold.

LEMMA 6.2. *For any $0 \leq j \leq n$,*

$$\mathbb{E}[f(\mathbf{o}^{(j)}) - f(\mathbf{o}^{(j+1)})] \leq \sqrt{\frac{k}{2}} \mathbb{E}[f(\mathbf{s}^{(j+1)}) - f(\mathbf{s}^{(j)})].$$

PROOF. Let e be the element of U considered by the randomized greedy algorithm in the $(j+1)$ th phase, and let $U^{(j)}$ and $\mathbf{o}^{(j)}$ be defined as in the proof of Theorems 6.1 and 5.1. We condition on an arbitrary, fixed value for both $\mathbf{s}^{(j)}$, $\mathbf{o}^{(j)}$, and consider the expectation over choices the algorithm makes for e . Because our result will hold for an arbitrary $\mathbf{s}^{(j)}$ or $\mathbf{o}^{(j)}$ it then extends to the expectation over the first j choices made by the algorithm.

As in the proof of Lemma 5.2, we define the solution $\mathbf{t} = \mathbf{o}|_{U^{(j-1)}} + \mathbf{s}^{(j-1)}$, and set $a_i = f_{i,e}(\mathbf{t})$ for $1 \leq i \leq k$. Let the values y_i be defined as in the algorithm. Then, as in the proof of Lemma 5.2, submodularity of f in every orthant implies that

$$a_i \leq y_i \text{ for every } i \in \{1, \dots, k\}. \quad (21)$$

Moreover, r -wise monotonicity of f implies that

$$\sum_{i \in I} a_i \geq 0 \text{ for all } I \in \binom{\{1, \dots, k\}}{r}. \quad (22)$$

Finally, by the construction of Algorithm 2, we have $y_i \geq 0$ for each $1 \leq i \leq k$.

Now, let suppose that in the optimal solution $o_e = p$ but the greedy algorithm sets $s_e \leftarrow q$. Then, we have $f(\mathbf{o}^{(j)}) = f(\mathbf{t}) + f_{p,e}(\mathbf{t})$ and $f(\mathbf{o}^{(j+1)}) = f(\mathbf{t}) + f_{q,e}(\mathbf{t})$, and so, as in the proof of Lemma 5.2,

$$f(\mathbf{o}^{(j)}) - f(\mathbf{o}^{(j+1)}) = f_{p,e}(\mathbf{t}) - f_{q,e}(\mathbf{t}) = a_p - a_q,$$

and

$$f(\mathbf{s}^{(j+1)}) - f(\mathbf{s}^{(j)}) = f_{q,e}(\mathbf{s}^{(j)}) = y_q.$$

For any given value q , the probability that the greedy algorithm makes such a choice is precisely y_q/β , and so

$$\mathbb{E}[f(\mathbf{s}^{(j+1)}) - f(\mathbf{s}^{(j)})] = \frac{1}{\beta} \sum_i y_i^2,$$

and

$$\mathbb{E}[f(\mathbf{o}^{(j)}) - f(\mathbf{o}^{(j+1)})] = \frac{1}{\beta} \sum_i y_i(a_p - a_i) = \frac{1}{\beta} \sum_{i \neq p} y_i(a_p - a_i).$$

In order to prove the lemma it is thus sufficient to show that

$$\sum_{i \neq p} y_i(a_p - a_i) \leq \sqrt{\frac{k}{2}} \sum_i y_i^2. \quad (23)$$

For any value of y_1, \dots, y_k , the left hand side of (23) is upper bounded by the optimal value of the following linear program in a_1, \dots, a_k , whose constraints are given by (21) and (22):

$$\begin{aligned} & \text{maximize} && \sum_{i \neq p} y_i(a_p - a_i) \\ & \text{subject to} && a_i \leq y_i, && \text{for } 1 \leq i \leq k \\ & && \sum_{i \in I} a_i \geq 0, && \text{for all } I \in \binom{\{1, \dots, k\}}{r} \end{aligned}$$

We consider an optimal, extreme-point solution a_1^*, \dots, a_k^* for this program. We first note that by increasing a_p we cannot violate the final constraint and can only increase the objective, and so we may assume that $a_p^* = y_p$. Of the remaining k constraints, $k - 1$ must be tight, of which $k - 2$ must be of the first type. Hence, for all i except at most 1 value $\ell \neq p$, we in fact have $a_i^* = y_i$. This accounts for $k - 1$ total tight constraints. The final tight constraint must imply either $a_\ell^* = y_\ell$ or $\sum_i a_i^* = 0$. Because $a_i^* = y_i$ for all $i \neq \ell$, the latter is equivalent to $a_\ell^* = -\sum_{i \neq \ell} y_i$. Moreover, because $y_i \geq 0$ for all i , setting $a_\ell^* = -\sum_{i \neq \ell} y_i$ always gives an objective value at least as large as setting $a_\ell^* = y_\ell$. Thus, we can characterize the optimal solution to this linear program by $a_i^* = y_i$ for all $i \neq \ell$, and $a_\ell^* = -\sum_{i \neq \ell} y_i$, where ℓ is some value distinct from p .

Returning to (23), we have

$$\begin{aligned}
\sum_{i \neq p} y_i (a_p - a_i) &\leq \sum_{i \neq p} y_i (a_p^* - a_i^*) \\
&= \sum_{i \neq p, \ell} y_i (y_p - y_i) + y_\ell \left(y_p + \sum_{i \neq \ell} y_i \right) \\
&= 2y_\ell y_p + \sum_{i \neq p, \ell} [y_\ell y_i + y_p y_i - y_i^2],
\end{aligned}$$

for any $y_1, \dots, y_k \geq 0$. In order to prove (23) it then suffices to show that

$$0 \leq \alpha \sum_i y_i^2 - 2y_\ell y_p - \sum_{i \neq p, \ell} [y_\ell y_i + y_p y_i - y_i^2], \quad (24)$$

where $\alpha = \sqrt{\frac{k}{2}}$. This follows directly from the fact that the right hand side of (24) can be written as the following sum of squares:

$$(y_\ell - y_p)^2 + \sum_{j \neq o, \ell} \left(\sqrt{\frac{\alpha-1}{k-2}} y_\ell - \sqrt{\frac{\alpha+1}{2}} y_j \right)^2 + \sum_{j \neq o, \ell} \left(\sqrt{\frac{\alpha-1}{k-2}} y_p - \sqrt{\frac{\alpha+1}{2}} y_j \right)^2. \quad (25)$$

In order to verify that this is the case, we note that

$$(y_\ell - y_p)^2 = y_\ell^2 - 2y_\ell y_p + y_p^2$$

and

$$\begin{aligned}
\left(\sqrt{\frac{\alpha-1}{k-2}} y_\ell - \sqrt{\frac{\alpha+1}{2}} y_i \right)^2 &= \frac{\alpha-1}{k-2} y_\ell^2 - 2\sqrt{\frac{(\alpha-1)(\alpha+1)}{2(k-2)}} y_\ell y_i + \frac{\alpha+1}{2} y_i^2 \\
&= \frac{\alpha-1}{k-2} y_\ell^2 - 2\sqrt{\frac{\alpha^2-1}{2(k-2)}} y_\ell y_i + \frac{\alpha+1}{2} y_i^2 \\
&= \frac{\alpha-1}{k-2} y_\ell^2 - 2\sqrt{\frac{\frac{k}{2}-1}{2(k-2)}} y_\ell y_i + \frac{\alpha+1}{2} y_i^2 \\
&= \frac{\alpha-1}{k-2} y_\ell^2 - 2\sqrt{\frac{\frac{k-2}{2}}{2(k-2)}} y_\ell y_i + \frac{\alpha+1}{2} y_i^2 \\
&= \frac{\alpha-1}{k-2} y_\ell^2 - 2\sqrt{\frac{1}{4}} y_\ell y_i + \frac{\alpha+1}{2} y_i^2 \\
&= \frac{\alpha-1}{k-2} y_\ell^2 - y_\ell y_i + \frac{\alpha+1}{2} y_i^2,
\end{aligned}$$

and, similarly,

$$\left(\sqrt{\frac{\alpha-1}{k-2}} y_p - \sqrt{\frac{\alpha+1}{2}} y_i \right)^2 = \frac{\alpha-1}{k-2} y_p^2 - y_p y_i + \frac{\alpha+1}{2} y_i^2$$

Thus, (25) is equal to

$$\begin{aligned}
& y_\ell^2 - 2y_\ell y_p + y_p^2 + \sum_{i \neq p, \ell} \left[\frac{\alpha - 1}{k - 2} y_\ell^2 - y_\ell y_i + \frac{\alpha + 1}{2} y_i^2 \right] + \sum_{i \neq p, \ell} \left[\frac{\alpha - 1}{k - 2} y_p^2 - y_p y_i + \frac{\alpha + 1}{2} y_i^2 \right] \\
&= y_\ell^2 - 2y_\ell y_p + y_p^2 + (\alpha - 1)y_\ell^2 + (\alpha - 1)y_p^2 - \sum_{i \neq p, \ell} \left[y_\ell y_i - \frac{\alpha + 1}{2} y_i^2 \right] - \sum_{i \neq p, \ell} \left[y_p y_i - \frac{\alpha + 1}{2} y_i^2 \right] \\
&= y_\ell^2 - 2y_\ell y_p + y_p^2 + (\alpha - 1)y_\ell^2 + (\alpha - 1)y_p^2 - \sum_{i \neq p, \ell} [y_\ell y_i + y_p y_i - (\alpha + 1)y_i^2] \\
&= \alpha y_\ell^2 + \alpha y_p^2 - 2y_\ell y_p + \alpha \sum_{i \neq p, \ell} y_i^2 - \sum_{i \neq p, \ell} [y_\ell y_i + y_p y_i - y_i^2] \\
&= \alpha \sum_i y_i^2 - 2y_\ell y_p - \sum_{i \neq p, \ell} [y_\ell y_i + y_p y_i - y_i^2].
\end{aligned}$$

□

The guarantees we obtain for the randomized greedy algorithm are better than for the deterministic greedy algorithm on r -wise monotone k -set functions only when k is small or r is large. While we do not have a tight example for the randomized greedy algorithm on r -wise monotone k -set functions for every fixed value of r and k , the following example confirms that the randomized algorithm can indeed perform worse than the deterministic algorithm for k -submodular (i.e. pairwise monotone) functions, once k grows large enough. This behavior is somewhat unintuitive, as the randomized greedy algorithm has an expected approximation ratio of $1/2$ for bisubmodular functions, while the deterministic greedy algorithm has an approximation ratio of only $1/3$.

Example 6.3. Consider the weighted set-coverage function $f^{(u,v)} : \{0, \dots, k\}^{\{u,v\}} \rightarrow \mathbb{R}_+$ given as follows. We have a universe $\{a, b\}$ where a has weight 1 and b has weight $\gamma = \frac{1}{\sqrt{k-1}}$. Additionally, we have sets $S_1 = \{a\}$ and $S_i = \{b\}$ for every $2 \leq i \leq k$, and $T_i = \{b\}$, for every $1 \leq i \leq k$. The value of $f^{(u,v)}(x_u, x_v)$ is then simply the total weight of all elements in $S_u \cup T_v$. The function induced by $f^{(u,v)}$ and any orthant is then a weighted set coverage function, and so is submodular. Moreover, all marginals of $f^{(u,v)}$ are non-negative and so $f^{(u,v)}$ is trivially r -wise monotone for any r .

We now consider the performance of the randomized greedy algorithm on $f^{(u,v)}$. We suppose, without loss of generality, that the greedy algorithm considers u before v . Initially we have $\mathbf{s} = \mathbf{0}$, and in the first phase, the algorithm sets $s_u \leftarrow 1$ with probability $\frac{1}{1+(k-1)\gamma}$ and for each $2 \leq i \leq k$, sets $s_u \leftarrow i$ with probability $\frac{\gamma}{1+(k-1)\gamma}$. In the next step, the algorithm considers v . We note that all the sets T_i are identical, and so the algorithm's particular choice in this phase does not affect the final value of the function. The solution \mathbf{s} produced by the algorithm has value $1 + \gamma$ if $s_u = 1$ and γ otherwise. Thus, the expected value of solution produced by the algorithm is:

$$\frac{1 + \gamma + (k - 1)\gamma^2}{1 + (k - 1)\gamma} = \frac{2 + \gamma}{1 + (k - 1)\gamma}.$$

The optimal value of $f^{(u,v)}$ is $1 + \gamma$ and so the expected approximation ratio of the randomized greedy algorithm on $f^{(u,v)}$ is

$$\alpha = \frac{2 + \gamma}{1 + (k - 1)\gamma} \cdot \frac{1}{1 + \gamma} = \frac{2 + \gamma}{1 + (k - 1)\gamma + \gamma + (k - 1)\gamma^2} = \frac{2 + \gamma}{2 + k\gamma}.$$

In particular, for all $k \geq 21$, we have $\alpha < 1/3$. For large k , α is approximately $1/\left(1 + \sqrt{\frac{k}{4}}\right)$. In the appendix, we show that the randomized greedy algorithm does indeed attain a similar, improved ratio for k -submodular functions.

7. CONCLUSION

In the preceding sections we have considered the problem of maximizing k -submodular functions by both a random partition and two simple greedy algorithms. In the case of maximizing a bisubmodular function, we obtained the same approximation ratios as those already known in the submodular case: $1/4$ for the naive random solution [Feige et al. 2011] and $1/2$ via a randomized greedy approach [Buchbinder et al. 2012]. We can make this correspondence more explicit by considering the following embedding of a submodular function into a bisubmodular function. Given a submodular function $g : 2^U \rightarrow \mathbb{R}_+$, we consider the biset function $f : 3^U \rightarrow \mathbb{R}_+$ defined by

$$f(S, T) \stackrel{\text{def}}{=} g(S) + g(U \setminus T) - g(U). \quad (26)$$

This embedding has been studied by Fujishige and Iwata, who show that the function f is bisubmodular and has the following property: if (S, T) is a minimizer (maximizer) of f then both S and $U \setminus T$ are minimizers (maximizers) of g [Fujishige and Iwata 2005]. Thus, exact 2-submodular function minimization (maximization) is a generalization of 1-submodular function minimization (maximization). We can in fact show a stronger result: that this embedding preserves approximability.

Suppose that some algorithm gives a α -approximation for bisubmodular maximization. Then, consider an arbitrary submodular function g and let f be the embedding of g defined as in (26). Let $O = (O_1, O_2)$ be a maximizer f , and suppose that the algorithm returns a solution $S = (S_1, S_2)$. Then, since f is pairwise monotone, we can greedily extend S to a partition $S' = (S'_1, S'_2)$ of U . Similarly, we can assume without loss of generality that O is a partition of U . Then, we have $f(U \setminus S'_2) = f(S'_1)$ and $f(U \setminus O_2) = f(O_1)$, and so

$$\begin{aligned} g(S'_1) &= \frac{1}{2} (g(S'_1) + g(U \setminus S'_1)) \\ &= \frac{1}{2} (f(S'_1, S'_2) + g(U)) \\ &\geq \frac{1}{2} (\alpha f(O_1, O_2) + g(U)) \\ &= \frac{1}{2} (\alpha g(O_1) + \alpha g(U \setminus O_2) + (1 - \alpha)g(U)) \\ &\geq \frac{1}{2} (\alpha g(O_1) + \alpha g(U \setminus O_2)) \\ &= \alpha g(O_1). \end{aligned}$$

Since O_1 is a maximizer of g , the resulting algorithm is an α -approximation for maximizing g . Hence, the $1/2 + \epsilon$ inapproximability results of [Feige et al. 2011; Dobzinski and Vondrák 2012] hold for bisubmodular maximization as well, in both the value oracle setting and under the assumption that $NP \neq RP$.

The embedding (26) also allows us to provide new intuition for the performance of the randomized greedy algorithm for submodular maximization considered by Buchbinder et al. [Buchbinder et al. 2012]. This algorithm maintains 2 solutions, S_1 and S_2 which are initially \emptyset and U . At each step, it considers an element e , and either adds e to S_1

or removes e from S_2 , with probability proportional to the resulting increase in the submodular function in either case.

In comparison, we consider the case in which we embed a submodular function g into a bisubmodular function f using (26) and then run the greedy algorithm of Section 6 on f . Suppose at some step we have a current solution $T = (T_1, T_2)$ and we consider element e , and define $S_1 = T_1$ and $S_2 = U \setminus T_2$. The algorithm will add e to either T_1 or T_2 with probability proportional to the resulting increase in f . In the first case, this increase is precisely $g(T_1 + e) - g(T_1) = g(S_1 + e) - g(S_1)$, and adding e to T_1 corresponds to adding e to S_1 . In the second case this increase is precisely $g(U \setminus (T_2 + e)) - g(U \setminus (T_2 - e)) = g(S_2) - g(S_2 - e)$ and adding e to T_1 corresponds to removing e from S_1 . Thus, the operation of the algorithm of Buchbinder et al. [Buchbinder et al. 2012] may be viewed as that of the natural, straightforward randomized greedy algorithm presented in Section 6, viewed through the lens of the embedding (26).

An interesting open question is whether the symmetry gap technique from [Vondrák 2009; Dobzinski and Vondrák 2012] can be generalized to obtain hardness results for k -submodular maximization for $k \geq 3$, and, more generally, for maximizing k -set functions that are submodular in every orthant and r -wise monotone for some $1 \leq r \leq k$.

Acknowledgments

We are grateful to Maxim Sviridenko for many insightful conversations.

References

- Kazutoshi Ando. 2002. K -submodular functions and convexity of their Lovász extension. *Discrete Applied Mathematics* 122, 1-3 (2002), 1–12. DOI: [http://dx.doi.org/10.1016/S0166-218X\(01\)00318-3](http://dx.doi.org/10.1016/S0166-218X(01)00318-3)
- Kazutoshi Ando, Satoru Fujishige, and Takeshi Naitoh. 1996. A characterization of bisubmodular functions. *Discrete Mathematics* 148, 1-3 (1996), 299–303. DOI: [http://dx.doi.org/10.1016/0012-365X\(94\)00246-F](http://dx.doi.org/10.1016/0012-365X(94)00246-F)
- Jesús M. Bilbao, Julio R. Fernández, Nieves Jiménez, and Jorge J. López. 2008. Survey of Bicooperative Games. In *Pareto Optimality, Game Theory and Equilibria*, Altannar Chinchuluun, Panos M. Pardalos, Athanasios Migdalas, and Leonidas Pitsoulis (Eds.). Springer. DOI: http://dx.doi.org/10.1007/978-0-387-77247-9_8
- André Bouchet. 1987. Greedy algorithm and symmetric matroids. *Mathematical Programming* 38, 2 (1987), 147–159. DOI: <http://dx.doi.org/10.1007/BF02604639>
- Niv Buchbinder, Moran Feldman, Joseph Naor, and Roy Schwartz. 2012. A Tight Linear Time (1/2)-Approximation for Unconstrained Submodular Maximization. In *Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS'12)*. IEEE, 649–658. DOI: <http://dx.doi.org/10.1109/FOCS.2012.73>
- Ramaswamy Chandrasekaran and Santosh N. Kabadi. 1988. Pseudomatroids. *Discrete Mathematics* 71, 3 (1988), 205–217. DOI: [http://dx.doi.org/10.1016/0012-365X\(88\)90101-X](http://dx.doi.org/10.1016/0012-365X(88)90101-X)
- David A. Cohen, Martin C. Cooper, Peter G. Jeavons, and Andrei A. Krokhin. 2006. The Complexity of Soft Constraint Satisfaction. *Artificial Intelligence* 170, 11 (2006), 983–1016. DOI: <http://dx.doi.org/10.1016/j.artint.2006.04.002>
- Shahar Dobzinski and Jan Vondrák. 2012. From Query Complexity to Computational Complexity. In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing (STOC'12)*. ACM, 1107–1116. DOI: <http://dx.doi.org/10.1145/2213977.2214076>
- Uriel Feige, Vahab S. Mirrokni, and Jan Vondrák. 2011. Maximizing Non-monotone Submodular Functions. *SIAM J. Comput.* 40, 4 (2011), 1133–1153. DOI: <http://dx.doi.org/10.1137/090779346>
- Satoru Fujishige. 2005. *Submodular Functions and Optimization* (2nd ed.). Annals of Discrete Mathematics, Vol. 58. North-Holland, Amsterdam.
- Satoru Fujishige and Satoru Iwata. 2005. Bisubmodular Function Minimization. *SIAM Journal on Discrete Mathematics* 19, 4 (2005), 1065–1073. DOI: <http://dx.doi.org/10.1137/S0895480103426339>
- Michael R. Garey and David S. Johnson. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman.
- M. Grötschel, L. Lovasz, and A. Schrijver. 1981. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica* 1, 2 (1981), 169–198. DOI: <http://dx.doi.org/10.1007/BF02579273>

- M. Grötschel, L. Lovasz, and A. Schrijver. 1988. *Geometric Algorithms and Combinatorial Optimization*. Algorithms and Combinatorics, Vol. 2. Springer.
- Anna Huber and Vladimir Kolmogorov. 2012. Towards Minimizing k -Submodular Functions. In *Proceedings of the 2nd International Symposium on Combinatorial Optimization (ISCO'12) (Lecture Notes in Computer Science)*, Vol. 7422. Springer, 451–462. DOI: http://dx.doi.org/10.1007/978-3-642-32147-4_40
- Anna Huber, Andrei Krokhin, and Robert Powell. 2014. Skew bisubmodularity and valued CSPs. *SIAM J. Comput.* 43, 3 (2014), 1064–1084. DOI: <http://dx.doi.org/10.1137/120893549>
- S. Iwata. 2002. A fully combinatorial algorithm for submodular function minimization. *Journal of Combinatorial Theory, Series B* 84, 2 (2002), 203–212. DOI: <http://dx.doi.org/10.1006/jctb.2001.2072>
- S. Iwata. 2003. A faster scaling algorithm for minimizing submodular functions. *SIAM J. Comput.* 32, 4 (2003), 833–840. DOI: <http://dx.doi.org/10.1137/S0097539701397813>
- Satoru Iwata. 2008. Submodular Function Minimization. *Mathematical Programming* 112, 1 (2008), 45–64. DOI: <http://dx.doi.org/10.1007/s10107-006-0084-2>
- Satoru Iwata, Lisa Fleischer, and Satoru Fujishige. 2001. A combinatorial strongly polynomial algorithm for minimizing submodular functions. *Journal of the ACM* 48, 4 (2001), 761–777. DOI: <http://dx.doi.org/10.1145/502090.502096>
- Satoru Iwata, Shin ichi Tanigawa, and Yuichi Yoshida. 2013. *Bisubmodular Function Maximization and Extensions*. Technical Report METR 2013-16. The University of Tokyo. <http://www.keisu.t.u-tokyo.ac.jp/research/techrep/data/2013/METR13-16.pdf>
- Satoru Iwata and James B. Orlin. 2009. A Simple Combinatorial Algorithm for Submodular Function Minimization. In *Proceedings of the 20th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'09)*. 1230–1237. DOI: <http://dx.doi.org/10.1145/1496770.1496903>
- Vladimir Kolmogorov. 2011. Submodularity on a tree: Unifying $L^\#$ -convex and bisubmodular functions. In *Proceedings of the 36th International Symposium on Mathematical Foundations of Computer Science (MFCS'11) (Lecture Notes in Computer Science)*, Vol. 6907. Springer, 400–411. DOI: http://dx.doi.org/10.1007/978-3-642-22993-0_37
- Bernhard Korte and Jens Vygen. 2007. *Combinatorial Optimization* (4th ed.). Algorithms and Combinatorics, Vol. 21. Springer.
- Jon Lee, Maxim Sviridenko, and Jan Vondrák. 2010. Submodular Maximization over Multiple Matroids via Generalized Exchange Properties. *Mathematics of Operations Research* 35, 4 (2010), 795–806. DOI: <http://dx.doi.org/10.1287/moor.1100.0463>
- László Lovász. 1983. Submodular Functions and Convexity. In *Mathematical Programming – The State of the Art*, A. Bachem, M. Grötschel, and B. Korte (Eds.). Springer, Berlin, 235–257.
- S. Thomas McCormick and Satoru Fujishige. 2010. Strongly polynomial and fully combinatorial algorithms for bisubmodular function minimization. *Mathematical Programming* 122, 1 (2010), 87–120. DOI: <http://dx.doi.org/10.1007/s10107-008-0242-9>
- H. Narayanan. 1997. *Submodular Functions and Electrical Networks*. North-Holland, Amsterdam.
- George L. Nemhauser and Laurence A. Wolsey. 1988. *Integer and Combinatorial Optimization*. John Wiley & Sons.
- James B. Orlin. 2009. A Faster Strongly Polynomial Time Algorithm for Submodular Function Minimization. *Mathematical Programming* 118, 2 (2009), 237–251. DOI: <http://dx.doi.org/10.1007/s10107-007-0189-2>
- Christos H. Papadimitriou and Mihalis Yannakakis. 1991. Optimization, Approximation, and Complexity Classes. *J. Comput. System Sci.* 43, 3 (1991), 425–440. DOI: [http://dx.doi.org/10.1016/0022-0000\(91\)90023-X](http://dx.doi.org/10.1016/0022-0000(91)90023-X)
- Liqun Qi. 1988. Directed submodularity, ditroids and directed submodular flows. *Mathematical Programming* 42, 1-3 (1988), 579–599. DOI: <http://dx.doi.org/10.1007/BF01589420>
- Alexander Schrijver. 2000. A Combinatorial Algorithm Minimizing Submodular Functions in Strongly Polynomial Time. *Journal of Combinatorial Theory, Series B* 80, 2 (2000), 346–355. DOI: <http://dx.doi.org/10.1006/jctb.2000.1989>
- Alexander Schrijver. 2003. *Combinatorial Optimization: Polyhedra and Efficiency*. Algorithms and Combinatorics, Vol. 24. Springer.
- Ajit P. Singh, Andrew Guillory, and Jeff Bilmes. 2012. On Bisubmodular Maximization. In *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics (AISTATS'12) (JLMLR Workshop and Conference Proceedings)*, Vol. 22. 1055–1063. <http://jmlr.csail.mit.edu/proceedings/papers/v22/singh12/singh12.pdf>
- Johan Thapper and Stanislav Živný. 2012. The power of linear programming for valued CSPs. In *Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS'12)*. IEEE, 669–678. DOI: <http://dx.doi.org/10.1109/FOCS.2012.25>

- Johan Thapper and Stanislav Živný. 2013. The complexity of finite-valued CSPs. In *Proceedings of the 45th ACM Symposium on the Theory of Computing (STOC'13)*. ACM, 695–704. DOI: <http://dx.doi.org/10.1145/2488608.2488697>
- Donald Topkis. 1998. *Supermodularity and Complementarity*. Princeton University Press.
- Jan Vondrák. 2009. Symmetry and Approximability of Submodular Maximization Problems. In *Proceedings of the 50th IEEE Symposium on Foundations of Computer Science (FOCS '09)*. IEEE Computer Society, 651–670. DOI: <http://dx.doi.org/10.1109/FOCS.2009.24>
- Magnus Wahlström. 2014. Half-integrality, LP-branching and FPT Algorithms. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'14)*. SIAM, 1762–1781. DOI: <http://dx.doi.org/10.1137/1.9781611973402.128>
- Justin Ward and Stanislav Živný. 2014. Maximizing Bisubmodular and k -Submodular Functions. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'14)*. SIAM, 1468–1481. DOI: <http://dx.doi.org/10.1137/1.9781611973402.108>

A. IMPROVED ANALYSIS OF ALGORITHM 2 FOR K -SUBMODULAR FUNCTIONS

In the case that f is in fact pairwise monotone (and, hence, k -submodular), we can prove the following stronger form of Lemma 6.2.

LEMMA A.1. *Suppose that f is k -submodular. Then, for any $0 \leq j \leq n$,*

$$\mathbb{E}[f(\mathbf{o}^{(j)}) - f(\mathbf{o}^{(j+1)})] \leq \alpha \mathbb{E}[f(\mathbf{s}^{(j+1)}) - f(\mathbf{s}^{(j)})].$$

where $\alpha = \max(1, \sqrt{\frac{k-1}{4}})$.

PROOF. Using the same notation as in the proof of Lemma 6.2, we shall now show

$$\sum_{i \neq p} y_i (a_p - a_i) \leq \alpha \sum_i y_i^2, \quad (27)$$

where $\alpha = \max(1, \sqrt{\frac{k-1}{4}})$. As in the proof of Lemma 6.2, we note that for any value of y_1, \dots, y_k , the left hand side of (27) is upper bounded by the optimal value of a linear program in a_1, \dots, a_k . Now, however, because f is pairwise monotone, we replace the (22) with $\binom{k}{2}$ constraints of the form of $a_i + a_\ell \geq 0$. This gives the program

$$\begin{aligned} & \text{maximize} && \sum_{i \neq p} y_j (a_p - a_i) \\ & \text{subject to} && a_i \leq y_i && 1 \leq i \leq k \\ & && a_i + a_\ell \geq 0 && \forall \{i, \ell\} \in \binom{\{1, \dots, k\}}{2}. \end{aligned}$$

Consider an optimal solution for this program. We note that increasing a_p cannot violate any constraint $a_p + a_\ell \geq 0$, and will increase the objective. Thus, we may assume that $a_p^* = y_p \geq 0$. We now consider 2 cases.

First, suppose that we have $a_\ell^* = -t < 0$ for some $\ell \in \{1, \dots, k\}$ and some value $t > 0$. Because $a_i^* + a_\ell^* \geq 0$ for all $i \neq \ell$, there can be at most one such ℓ . Moreover, we must have $a_i^* \geq t$ for all $i \neq \ell$. For any value $i \notin \{\ell, p\}$, we note that decreasing a_i^* can only increase the objective of our linear program. Thus, in this case, we may assume that $a_i^* = t$ for all $i \notin \{\ell, p\}$, $a_\ell = -t$ and $a_p = y_p$. We can then rewrite our objective as:

$$\sum_{i \neq p} y_i y_p + t \left(y_\ell - \sum_{j \neq \ell, p} y_j \right). \quad (28)$$

Because $t > 0$, we must have $y_\ell \geq \sum_{j \neq \ell, p} y_j$ (otherwise, we could increase (28) by decreasing t). Moreover, we must have $t \leq y_p$, since otherwise we would have $a_p^* + a_\ell^* = y_p - t < 0$. Hence, we have:

$$\sum_{i \neq p} y_i y_p + t \left(y_\ell - \sum_{j \neq \ell, p} y_j \right) \leq \sum_{i \neq p} y_i y_p + y_p y_\ell - y_p \sum_{j \neq \ell, p} y_j = 2y_p y_\ell \leq y_p^2 + y_\ell^2 \leq \sum_i y_i^2,$$

and we have proved (27) with $\alpha = 1$.

Next, suppose that $a_i \geq 0$ for all $i \in \{1, \dots, k\}$. Then, the objective of our program satisfies

$$\begin{aligned} \sum_{i \neq p} y_i (a_p - a_i) &\leq \sum_{i \neq p} y_i a_p \\ &= \sum_{i \neq p} y_i y_p \\ &= \frac{1}{2\sqrt{k-1}} \cdot 2\sqrt{k-1} y_p \sum_{i \neq p} y_i \\ &\leq \frac{1}{2\sqrt{k-1}} \left[(k-1) y_p^2 + \left(\sum_{i \neq p} y_i \right)^2 \right] \\ &\leq \frac{1}{2\sqrt{k-1}} \left[(k-1) y_p^2 + (k-1) \sum_{i \neq p} y_i^2 \right] \\ &= \frac{\sqrt{k-1}}{2} \sum_i y_i^2, \end{aligned}$$

where the second inequality follows from $a^2 + b^2 \geq 2ab$ for any real numbers a and b , and third inequality follows from the Cauchy-Schwarz inequality. Thus, we have proved (27) with $\alpha = \sqrt{\frac{k-1}{4}}$. \square

By replacing Lemma 6.2 with Lemma A.1, in the proof of Theorem 6.1, we obtain the following result.

THEOREM A.2. *Let $f : \{0, \dots, k\}^U$ be a k -submodular set function. Let \mathfrak{o} be an orthant of U that maximizes f and let \mathfrak{s} be the orthant of U produced by the randomized greedy algorithm. Then,*

$$(1 + \alpha) \mathbb{E}[f(\mathfrak{s})] \geq f(\mathfrak{o}),$$

for $\alpha = \max(1, \sqrt{\frac{k-1}{4}})$.