Diminishing Returns and Beyond

Andreas Haupt

June 2, 2015

Abstract

In several areas of application (i.e. problems of facility location planning, database management and document summarization) besides quality of the set of chosen objects obeying given constraints also maximizing their diversity is an objective. A model for this is the constrained maximization of a sum of a relevance- (usually submodular) and a diversity-measuring set function.

We present existing variants of the function class of submodular functions that include such sums. We prove, that the standard greedy algorithm is a $\frac{1}{5.95}+o(1)$-approximation algorithm for the maximization of weakly submodular functions (see [BLY14]) subject to cardinality constraints, completing a proof ibidem. We define a generalization of monotone submodular functions and prove constant-factor approximations for the standard greedy algorithm.

1 Introduction

As set function maximization is unbearably hard and most combinatorial optimization problems may be reduced to it, special function classes have been studied, that admit constant factor approximations. One of these is the class of *submodular functions*, functions that exhibit a marginal gains property: The added value of an element does not increase when added to a larger set. This property seems intuitive for many real-world quality and relevance measures. Nevertheless, for some applications, the class of submodular functions is too narrow:

Assume, one sends the query “Swiss Cheese Old” to a food database. Assume further, that the database software is clever enough to figure out that “Swiss” means the origin of the product, “Cheese” is the product class and “Old” is the stage of maturity of the cheese. A maximization with the only objective to maximize *relevance* might output the most mature cheeses of all Switzerland, which might not be convenient for the consumer, as he might be looking for a mature version of some popular mild cheese (the database cannot know).

In this example, it might be better as a decision rule to select an output, that has high relevance and *at the same time* is also diverse. One intuitive diversity measures are the following set functions:

$A \mapsto \sum_{a,b\in A} d(a,b)$, for d a metric returns the sum of distances inside a set. It is called *sum-dispersion*. As this function is monotone, when maximizing subject to a cardinality constraint k, this is equivalent to the maximization of the average distance between points in the set (multiplying sum-dispersion with $1/\binom{k}{2}$).

Then the optimization problem for the database is to solve could be

$$\max_{|S|\le k} f(S) + \lambda d(S)$$

where f is a relevance measure, $\lambda \ge 0$ a payoff parameter, d sum-dispersion and k a cardinality constraint (e.g. the number of entries that fit on the first page of the query answer).

This article is devoted to studying set function classes that include $f(S) + \lambda d(S)$. Subsequently, we will call those functions *target functions*.

The structure of the article is as follows: In section 2 we give fundamental definitions and notations. Section 3 is devoted to two variants of subdmodular functions studied in the literature. In this section, we prove our main result, Theorem 3.14. In section 4, we define another variant of submodularity and obtain a constant-factor approximation.

2 Preliminaries

As conventions, throughout the article, $V = \{v_1, \dots, v_n\}$ denotes a finite set, $\mathbf{R}_+$ the nonnegative reals and $\binom{V}{k}$ the set of k-subsets of V. $S(V)$ denotes the set of permutations of V and $\mathcal{P}(V)$ its power set. Furthermore, we mean by a conic combination of functions a linear combination with non-negative coefficients. Unif_A denotes the uniform distribution on the set A. As last convention, if we set, e.g. in pseudocode, an element to be the argumentum maximum of some function, then we mean that the element is defined to be *any* element in the function's argumentum maximum.

2.1 Definition (Submodularity) $f: \mathcal{P}(V) \to \mathbf{R}_+^0$ *is called submodular if and only if for all $A, B \subseteq V$*

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B). \tag{1}$$

We furthermore use for $e \in V \supseteq B$ the notations

$$\Delta^e f(A) \coloneqq f(B \cup \{e\}) - f(B)$$
$$\Delta_e f(A) \coloneqq f(B \setminus \{e\}) - f(B),$$

which stand in the relationship

$$\Delta^e f(A) = -\Delta_e f(A \cup \{e\}). \tag{2}$$

2.2 Proposition (Local Formulation) *Submodularity is equivalently characterized by*

$$\Delta^e(D) \leq \Delta^e(C) \tag{3}$$

for all $C \subseteq D$, $e \in V \setminus D$.

Proof: One direction is obtained by taking (1) with

$$A \coloneqq C \cup \{e\} \qquad B \coloneqq D.$$

For the other direction, let $\{b_1, \dots, b_{|B \setminus A|}\} \coloneqq B \setminus A$. Then, with a telescopic sum, (3) yields

$$f(B) - f(A \cap B) = \sum_{i=1}^{|B \setminus A|} \Delta^{b_i}((A \cap B) \cup \{b_1, \dots, b_{i-1}\})$$
$$\geq \sum_{i=1}^{|B \setminus A|} \Delta^{b_i}(A \cup \{b_1, \dots, b_{i-1}\}) = f(A \cup B) - f(A). \quad \blacksquare$$

2.3 Definition (Monotone) $f: \mathcal{P}(V) \to \mathbf{R}$ *is monotone if and only if $\Delta^e f(A) \geq 0$ for all $e \in V \supseteq A$.*

2.4 Remark (Value Oracle) For algorithms, the efficient computability of the values of a submodular function is critical. In the following, we implicitly assume that all algorithms are given a *value oracle* for the input function.

The maximization of submodular functions was extensively studied in the literature (e.g. [Buc+14; NWF78; Buc+12; Svi04]). In Appendix A we present some of the most important results, which is postponed for better readibility. We continue with the discussion of variants of submodular functions studied in the literature.

3 Generalizations of Submodular Functions

Many generalizations or variants of submodularity studied in recent years may be categorized into the following: those that allow the violation of (3) a *certain number of times* and those that relax *each* of these inequalities. The first category, that we call the *"how often"-approach* category, will be discussed in subsection 3.1 by means of the MPH-hierarchy from [Fei+14]. On the other hand, subsection 3.2 is devoted to the second, "how much"-category, represented by the *weakly submodular functions* of [BLY14].

3.1 "How often"-approaches

[Fei+14] give a hierarchy of monotone, nonnegative set functions. It depends largely on the notion of a *hypergraph representation*:

3.1 Proposition (Hypergraph Representation) *Every function $f: \mathcal{P}(V) \to \mathbf{R}_+$ has a unique hypergraph representation f^H, i.e. a function $f^H: \mathcal{P}(V) \to \mathbf{R}_+$, such that for all $S \subseteq V$*

$$f(S) = \sum_{T \subseteq S} f^H(T)$$

Proof: Define f^{H} inductively on k on $\binom{V}{k}$. For $k = 0$, existence is clear. Now suppose f^{H} is defined on $\bigcup_{j=0}^{k-1}\binom{V}{j}$. Then we can define any $A \in \binom{V}{k}$ by

$$f^{\mathrm{H}}(A) = f(A) - \sum_{B \subsetneq A} f^{\mathrm{H}}(A)$$

which will not interfere with any other set from $\binom{V}{k}$. The uniqueness is easily obtained. ∎

Therefore

$$\operatorname{rank} f := \max\{|S| | f_H(S) \neq 0\}$$

is well-defined for any f. The MPH-hierarchy is defined as:

3.2 Definition (MPH-hierarchy) *Define the rth level in the* maximum over positive hypergraph *hierarchy.*

$$\begin{aligned}\mathrm{MPH}(r) := \{f\colon \mathcal{P}(V) \to \mathbf{R}^{+} | \exists (f_i)_{i\in I}\colon \\ f = \max_{i\in I} f_i \wedge (\forall i \in I\colon f_i^H \geq 0 \wedge \operatorname{rank} f_i \leq r)\}\end{aligned}$$

All monotone functions are $\mathrm{MPH}(n)$, as for any monotone function f,

$$f(A) = \max_{B\in\mathcal{P}(V)} \sum_{C\subseteq A} \delta_{AB}(B),$$

where we chose $I := \mathcal{P}(V)$.

$$\delta_{AB}(B) = \begin{cases} 1 & A = B \\ 0 & \text{else.}\end{cases}$$

which obviously is non-negative and of rank n.

Furthermore, MPH is *monotone*, i.e. $\mathrm{MPH}(i) \subseteq \mathrm{MPH}(i+1)$, $i \in \mathbf{N}$ which follows by definition.

3.3 Theorem ([see LLN01, section 3.4.3]) *All submodular, monotone functions are* $\mathrm{MPH}(1)$.

Proof: Let $f\colon \mathcal{P}(V) \to \mathbf{R}_{+}$ be monotone, submodular. We choose $I := S(V)$, the set of permutations of V and $f_\pi^{\mathrm{H}}(\{v_i\}) := \Delta^{\pi(v_i)} f(\{\pi(v_1), \ldots, \pi(v_{i-1})\})$. This uniquely defines f_π by Proposition 3.1

$$\begin{aligned}\max_{\pi\in S(V)} f_\pi(A) &= \max_{\pi\in S(V)} \sum_{v_i\in A} f_\pi^{\mathrm{H}}(\{v_i\}) \\ &= \max_{\pi\in S(V)} \sum_{v_i\in A} \underbrace{\Delta^{\pi(v_i)} f(\{\pi(v_1), \ldots, \pi(v_{i-1})\})}_{\leq \Delta^{\pi(v_i)}(\{\pi(v_1),\ldots,\pi(v_{i-1})\}\cap A)}\end{aligned}$$

which means, that the maximum is taken, if all $\pi(v_1), \ldots, \pi(v_{i-1}) \in A$ if $\pi(v_i) \in A$. Thus, for a maximizing π^*: $\pi(A) = \{1, \ldots, |A|\}$ and a numbering $A = \{a_1, \ldots, a_{|A|}\}$ such that $\pi(v_i) = a_i$

$$\max_{\pi\in S(V)} f_\pi(A) = \sum_{i=1}^{|A|} \Delta^{a_i}(\{a_1, \ldots, a_{i-1}\}) = f(A) - f(\varnothing)$$

So, $f = \max_{\pi\in S(V)}(f(\varnothing) + f_\pi)$. The positivity and rank assumption is easily observed. ∎

We also have

3.4 Proposition *Sum-dispersion is* $\mathrm{MPH}(2)$.

Proof: We take $I := \{1\}$ and

$$f_1^{\mathrm{H}}(B) = \begin{cases} d(a,b) & B = \{a,b\}, a \neq b \\ 0 & \text{else.}\end{cases}$$

∎

The following theorem shows, that $\mathrm{MPH}(k)$ is a cone, i.e. closed under conic combinations. Hence, all conic combinations of a submodular function and sum-dispersion are $\mathrm{MPH}(2)$ (the target functions).

3.5 Theorem $\mathrm{MPH}(k)$ *is a cone.*

Proof: Let $\max_{i\in I} f_i, \max_{j\in J} g_j \in \mathrm{MPH}(k)$, $\alpha, \beta > 0$. Then

$$\alpha \max_{i\in I} f_i + \beta \max_{j\in J} g_j = \max_{(i,j)\in I\times J} \alpha f_i + \beta g_j$$

∎

Thus, this function class includes our target function, unfortunately, one should not hope for a constant-factor approximation for $\mathrm{MPH}(2)$-functions. Consider

3.6 Problem (k-Dense Subgraph) *Given $G = (V,E)$ undirected, $k \in \mathbf{N}$ find $A \subseteq \binom{V}{k}$ that maximizes $|E(G[A])|$.*

3.7 Proposition *There is an approximation-preserving reduction of k-dense subgraph to the cardinality-constrained maximization of* $\mathrm{MPH}(2)$*-functions.*

Proof: In fact, it is just a special case: We let $V(G)$ be the universe, $I = \{1\}$. Then

$$|E(G[A])| = \sum_{u,v \in A} \mathbb{1}_{E(G)}(B)$$

where $\mathbb{1}_{E(G)}(B) = 1$ if and only if $B \in E(G)$. ■

This is not a hardness result, as it has not yet been proved that k-dense subgraph is not in `APX`, but the fastest known algorithm is a $O(n^{1/4})$-approximation from [Bha+10]; yet it is considered unlikely, that there is a constant-factor approximation for this problem.

3.8 Remark (Supermodularity Degree) [FI14] studied another hierarchy of monotone submodular functions. They proved, that a variant of the Greedy algorithm obtains an approximation ratio of $1 - \exp(\frac{1}{d(f)+1})$ in $\mathrm{Poly}(n, 2^{d(f)})$ for $d(f)$ the so called *supermodularity degree* of f,

$$d(f) \coloneqq \max_{u \in V} |\{v \in V | \exists S \subseteq V : \Delta^u(S \cup \{v\}) > \Delta^u(S)\}.$$

It can be shown, that functions that have supermodularity degree d are $\mathrm{MPH}(d+1)$ and that this inclusion is strict. Still, even this definition is not appropriate to deal with the target functions, as, without making assumptions about the distances in the given metric, the supermodularity degree of target functions may be arbitrarily high. Furthermore, the time dependency on $d(f)$ might cause the algorithm analyzed in [FI14] to be inefficient.

3.2 "How much"-approach

[BLY14] proposed a generalization of the notion of a monotone submodular function that, as $\mathrm{MPH}(2)$ contains the target functions, as we will prove in subsubsection 3.2.1. Contrary to $\mathrm{MPH}(2)$ prior, a local search algorithm for maximizing such functions to a cardinality constraint (even to an arbitrary matroid constraint) was analyzed in [BLY14, section 6] to yield a 14.5-approximation for any size of the cardinality constraint and asymptotically a 10.22-approximation. Interestingly, when the paper was released, an incomplete proof that Algorithm 1 yields a constant-factor approximation for

Algorithm 1: Greedy

```
Input: f: P(V) → R+
Output: P_k ⊆ V, |P_k| = k
P_0 ← ∅
for i = 1 to k do
    e ← arg max_{e' ∈ V ∖ P_{i-1}} Δ^{e'} f(P_{i-1})
    P_i ← P_{i-1} ∪ {e}
return P_k
```

the cardinality-constrained maximization of weakly submodular functions subject to cardinality constraints was part of the paper: The approximation ratio was proved to be bounded by the limit of a complicated product for that only numerical calculations *suggest* that it converges. In subsubsection 3.2.2, we complete this proof and show that Greedy yields an asymptotic constant-factor approximation. But first, we define weakly submodular functions and give fundamental properties in subsubsection 3.2.1.

3.2.1 Definitions and Properties of Weakly Submodular Functions

3.9 Definition (Weakly Submodular) *$f : \mathcal{P}(V) \to \mathbf{R}_+$ is said to be weakly submodular if and only if it is* normalized, *i.e. $f(\varnothing) = 0$ and for all $A, B \subseteq V$*

$$|A|f(B) + |B|f(A) \geq |A \cap B| f(A \cup B) + |A \cup B| f(A \cap B). \quad (4)$$

In all of the following, we do not need $f(\varnothing) = 0$. In [BLY14], it was, to our best knowledge, only needed implicitly for the proof of [BLY14, Proposition 3.3], a reflection principle we do not consider so important to justify strengthening the definition. We therefore propose, to omit the assumption $f(\varnothing) = 0$ from Definition 3.9.

3.10 Remark (Local Formulation) We make a calculation as in the second part of Proposition 2.2:

For $C \subseteq D$ and $e \in V \setminus D$, Definition 3.9 implies

$$\begin{aligned}
&|C \cup \{e\}| f(D) + |D| f(C \cup \{e\}) \\
&\qquad \geq |(C \cup \{e\}) \cap D| f(C \cup \{e\} \cup D) \\
&\qquad\qquad + |C \cup \{e\} \cup D| f((C \cup \{e\}) \cap D).
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
&(|C| + 1) f(D) + |D| f(C \cup \{e\}) \\
&\qquad\qquad \geq |C| f(D \cup \{e\}) + (|D| + 1) f(C)
\end{aligned}$$

which is again equivalent to

$$\frac{\Delta^e f(C)}{|C| + 1} \geq \frac{\Delta^e f(D)}{|D| + 1} + \frac{f(C \cup \{e\}) - f(D \cup \{e\})}{(|C| + 1)(|D| + 1)} \quad (5)$$

or

$$|D| \Delta^e f(C) \geq |C| \Delta^e f(D) + f(C) - f(D). \quad (6)$$

We remark, that in the following, we only need the inequalities of (6) for $C \subseteq D$ and not (4)'s full power. It is an interesting question whether (6) for $C \subseteq D$ characterize the same function class as (4).

(5) demonstrates, that thanks to $|\cdot|$'s monotonicity, we relax the submodularity in the case of a *monotone* f by comparing *relative marginal gains* (see section 4) and adding another term.

It is readily proved by an example with two very distant points in a metric space as universe, that this additive term is needed for the function class of weakly submodular (see section 4) functions to include max-sum dispersion. That this term is also sufficient for this, shows the following:

3.11 Theorem ([BLY14, Proposition 3.4]) *Sum-dispersion is weakly submodular.*

Proof: Denote $d(A, B) = \sum_{\substack{a \in A \\ b \in B}} d(A, B)$.

Observe first, that for any disjoint $A, B, C \subseteq V$

$$\begin{aligned}
&|C| d(A, B) = \sum_{\substack{a \in A \\ b \in B}} \sum_{c \in C} d(a, b) \\
&\leq \sum_{\substack{a \in A \\ b \in B}} \sum_{c \in C} d(a, c) + d(c, b) = |B| d(A, C) + |A| d(B, C).
\end{aligned}$$

Now take $F, G \subseteq V$. Then for

$$A = F \setminus G \qquad B = G \setminus F \qquad C = G \cap F$$

we have

$$\begin{aligned}
&|F \cap G| d(F \cup G) + |F \cup G| d(F \cap G) \\
&= |C| (d(A) + d(B) + d(C) + d(A, C) + d(B, C) \\
&\quad + d(A, B)) + |F \cup G| d(F \cap G) \\
&\leq |C| (d(A) + d(B) + d(C) + d(A, C) + d(B, C)) \\
&\quad + |B| d(A, C) + |A| d(B, C) + |F \cup G| d(F \cap G) \\
&= |F| d(G) + |G| d(F)
\end{aligned}$$

■

Furthermore, we also have

3.12 Proposition ([BLY14, Proposition 3.1]) *Monotone, submodular functions are weakly submodular.*

Proof: Let $F, G \subseteq V$, without loss of generality $|F| \leq |G|$. Then

$$\begin{aligned}
&|F| f(G) + |G| f(F) \\
&= |F| \underbrace{(f(F) + f(G))}_{\geq f(F \cup G) + f(F \cap G)} + (|G| - |F|) \underbrace{f(F)}_{\geq f(F \cap G)} \\
&= |F| f(F \cup G) + |G| f(F \cap G) \\
&= |F \cap G| f(F \cup G) + (|F| - |F \cap G|) f(F \cup G) \\
&\quad + |F| f(F \cap G) \\
&\geq |F \cap G| f(F \cup G) + |F \cup G| f(F \cap G)
\end{aligned}$$

■

As also, (4) is preserved under conic combinations (as well as normalization), our target functions are indeed weakly submodular.

3.13 Remark In fact, the proof of Proposition 3.12 indeed can be generalized. If we substitute for $|\cdot|$ in (4) some function g, we arrive at another variant of submodularity, that one could call g-submodularity. Then if f is monotone and g_1-modular, and g_2 is modular, then f is also $g_1 g_2$-modular. Using this, one may define another hierarchy of monotone, nonnegative functions. Unfortunately, we were not able to prove interesting results for such function classes.

3.2.2 Constant-Factor Greedy Approximation

Now, we prove, that Algorithm 1 yields a constant-factor approximation. For this, we will need the results, from Appendix B and Appendix C, that where postponed into the appendix for better readibility.

3.14 Theorem *If f is monotone and weakly submodular, Algorithm 1 yields an output P_k such that $f(P_k) \geq (\frac{1}{5.95} + o(1)) \max_{S \subseteq V} f(S)$.*

Proof (only using (6)): Denote P^* an optimal solution and let $\{p_1, \dots, p_{|P^* \setminus P_i|}\} := P^* \setminus P_i$. Then, we have the following equations by (6):

$$\begin{aligned}(i + |P^* \setminus P_i| - j)\Delta^{p_j} f(P_i) &\geq i\Delta^{p_j} f(P_i \cup \{p_{j+1}, \dots, p_{|P^* \setminus P_i|}\}) \\ &\quad + f(P_i) - f(P_i \cup \{p_{j+1}, \dots, p_{|P^* \setminus P_i|}\}) \qquad (7_{ij})\end{aligned}$$

for $1 \leq i \leq k$ and $1 \leq j \leq |P^* \setminus P_i|$. Now $\sum_{j=1}^{|P^* \setminus P_j|} (\frac{i+1}{i})^{j-1} (7_{ij})$ takes the form

$$\begin{aligned}&\sum_{j=1}^{|P^* \setminus P_j|} \left(\frac{i+1}{i}\right)^{j-1} (i + |P^* \setminus P_i| - j)\Delta^{p_j} f(P_i) \\ &\geq i f(P^* \cup P_i) - (i+1)\left(\frac{i+1}{i}\right)^{|P^* \setminus P_i| - 1} f(P_i) \\ &\qquad \sum_{j=1}^{|P^* \setminus P_i|} \left(\frac{i+1}{i}\right)^{j-1} f(P_i).\end{aligned}$$

Observe, that on the right hand side is a telescopic sum. Using inequalities

$$f(P_{i+1}) - f(P_i) \geq \Delta^{p_j} f(P_i) \quad f(P^* \cup P_i) \geq f(P^*),$$

which are due to the choice in line 3 and the monotonicity of f, this implies (after rearranging)

$$\begin{aligned}&\sum_{j=1}^{|P^* \setminus P_i|} \left(\frac{i+1}{i}\right)^{j-1} (i + |P^* \setminus P_i| - j) f(P_{i+1}) \\ &\geq i f(P^*) + \sum_{j=1}^{|P^* \setminus P_i| - 1} \left(\frac{i+1}{i}\right)^{j-1} (i + |P^* \setminus P_i| - j + 1) f(P_i),\end{aligned}$$

Calling the coefficient of $f(P_{i+1})$ $a^i_{|P^* \setminus P_i|}$ and the coefficient of $f(P_i)$ $b^i_{|P^* \setminus P_i|}$, we notice, that for any $l \in \mathbf{N}$

$$\begin{aligned}a^i_l - b^i_l &= \left(\frac{i+1}{i}\right)^{l-1} i - \sum_{j=1}^{l-1} \left(\frac{i+1}{i}\right)^{j-1} \\ &= \left(\frac{i+1}{i}\right)^{l-1} i - \left(\left(\frac{i+1}{i}\right)^{l-1} i - i\right) = i\end{aligned}$$

having made use of Lemma C.1. Now, we can simplify the term for a^i_l using Lemma C.2

$$\begin{aligned}a^i_l &= (i + l)\sum_{j=1}^{l} \left(\frac{i+1}{i}\right)^{j-1} - \sum_{j=1}^{l} j \left(\frac{i+1}{i}\right)^{j-1} \\ &= (i^2 + il)\left(\left(\frac{i+1}{i}\right)^{l} - 1\right) - i^2 + (i^2 - il)\left(\frac{i+1}{i}\right)^{l} \\ &= -2i^2 - il + 2i^2 \left(\frac{i+1}{i}\right)^{k}.\end{aligned}$$

Where for the second-last equality, one observes for the second term $(1 - \frac{i+1}{i})^{-2} = i^2$ and

$$\begin{aligned}&\sum_{j=1}^{l} j \left(\frac{i+1}{i}\right)^{j-1} \\ &= i^2 \left(l \left(\frac{i+1}{i}\right)^{l} \left(1 + \frac{1}{i}\right) - l\left(\frac{i+1}{i}\right)^{l} - \left(\frac{i+1}{i}\right)^{l} + 1\right) \\ &= i^2 + il\left(\frac{i+1}{i}\right)^{l} - i^2 \left(\frac{i+1}{i}\right)^{l}\end{aligned}$$

Up to this point, we followed the proof of [BLY14, Theorem 5.1]. Here we choose to Now, as b^i_l is increasing in l

$$\begin{aligned}i f(P^*) &\leq a^i_{|P^* \setminus P_i|} f(P_{i+1}) - b^i_{|P^* \setminus P_i|} f(P_i) \\ &\overset{(9)}{=} i f(P_{i+1}) + b^i_{|P^* \setminus P_i|}(f(P_{i+1}) - f(P_i)) \\ &\leq i f(P_{i+1}) + b^i_k (f(P_{i+1}) - f(P_i))\end{aligned}$$

which implies, using a zero addition,

$$\begin{aligned}f(P^*) - f(P_{i+1}) &\leq \frac{b^i_k}{i}(f(P_{i+1}) - f(P^*)) \\ &= \frac{b^i_k}{i}((f(P^*) - f(P_i)) - (f(P^*) - f(P_{i+1}))).\end{aligned}$$

which is equivalent to

$$
\begin{aligned}
f(P^*) - f(P_{i+1}) &\leq \frac{b_k^i}{i} \frac{1}{1 + \frac{b_k^i}{i}} (f(P^*) - f(P_i)) \\
&= \frac{b_k^i}{a_k^i} (f(P^*) - f(P_i))
\end{aligned}
$$

and yields by iterative application of this inequality and $f(P_1) \geq 0$

$$
\begin{aligned}
f(P^*) - f(P_k) &\leq \frac{b_k^{k-1}}{a_k^{k-1}} (f(P^*) - f(P_{k-1})) \\
&\leq \prod_{i=1}^{k-1} \frac{b_k^i}{a_k^i} f(P^*)
\end{aligned}
$$

We need another inequality to finally bound $\prod_{i=1}^{k-1} \frac{b_k^i}{a_k^i}$. For this, denote $h(x) \coloneqq 2x(\exp(\frac{1}{x}) - 1) - 1$.

$$
\begin{aligned}
2i\left(1 + \frac{1}{i}\right)^k - k - 2i &\leq 2i\left(\exp\left(\frac{k}{i}\right) - 1\right) - k \\
&= k\left(\frac{\exp(\frac{k}{i}) - 1}{\frac{k}{i}} - 1\right) = kh\left(\frac{i}{k}\right) \qquad (8)
\end{aligned}
$$

which follows from Lemma C.3. We finally arrive at

$$
\begin{aligned}
\prod_{i=1}^{k-1} \frac{b_k^i}{a_k^i} &= \prod_{i=1}^{k-1} 1 - \frac{i}{a_k^i} \\
&= \prod_{i=1}^{k-1} 1 - \frac{1}{2i(1 + \frac{1}{i})^k - 2i - k} \\
&\overset{(8)}{\leq} \prod_{i=1}^{k-1} 1 - \frac{1}{k} \frac{1}{h(\frac{i}{k})} \\
&\xrightarrow{k \to \infty} \prod_0^1 (1 - \frac{1}{h(x)}) = \exp\left(\int_0^1 -\frac{1}{h(x)} \,\mathrm{d}x\right)
\end{aligned}
$$

Where we used Theorem B.2 for the limit (observe, that h may be continuously extended to the point 0 by 0 (compare Figure 1). Now, $\exp\left(\int_0^1 -\frac{1}{h(x)} \,\mathrm{d}x\right) \approx \frac{1}{5.95}$. This concludes the proof. ∎

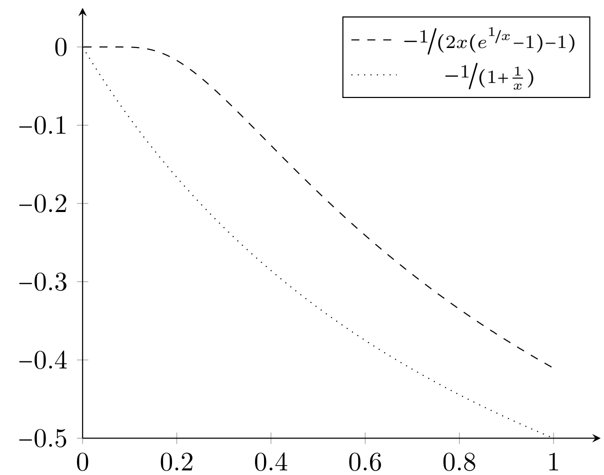

Figure 1: Plot of integrands from end of proofs of Theorem 3.14 respectively Theorem 4.3 ($\zeta = 2$).

4 Diminishing Relative Gains

The Volterra product integral proved useful in the proof of Theorem 3.14. We show in the following, that it is as well useful for the analysis of other function classes: A definition that is very convincing in the framework of valuation functions is the following: Instead of requiring diminishing marginal gains, we only stipulate that *relative* marginal gains decrease, i.e.

4.1 Definition (Diminishing g-Relative Gains) *$f : \mathcal{P}(V) \to \mathbf{R}_+$ is said to have diminishing g-relative marginal gains if and only if for all $C \subseteq D \subseteq V \setminus \{e\}$, $g(C) \neq 0$*

$$
\frac{\Delta^e f(D)}{g(D)} \leq \frac{\Delta^e f(C)}{g(C)}, \qquad (9)
$$

where $g : \mathcal{P}(V) \to \mathbf{R}_+$.

Obviously, if g is monotone, this relaxes the definition of subdmodularity. For modular functions g that take more than two different values, this is a strict generalization, as then g^2 is not-submodular: Let $g(\{e\}) > 0$ and $g(D) > g(C)$. Then

$$
\begin{aligned}
\Delta^e g^2(D) &= g(\{e\})^2 + 2g(\{e\})g(D) \\
&> g(\{e\})^2 + 2g(\{e\})g(C) = \Delta^e g^2(C)
\end{aligned}
$$

On the other hand for any $C, D \subseteq V$

$$\frac{\Delta^e g^2(D)}{g(D)} = \frac{g(e)^2}{g(D)} + 2g(\{e\}) \leq \frac{g(e)^2}{g(C)} + 2g(\{e\}) = \frac{\Delta^e g^2(C)}{g(C)}$$

4.2 Remark (Sum-Dispersion) One easily constructs examples of a metric, such that for any nonzero function g sum-dispersion does not exhibit diminishing g-relative marginal gains: Let A be with $g(A) > 0$, let $e, b \in V \setminus A$. Then for $e \notin D \supseteq C$

$$\frac{\Delta^e f(A)}{g(A)} = \frac{d(e,A)}{g(A)} < \frac{d(e,A) + d(e,b)}{g(A \cup \{b\})} = \frac{\Delta^e f(A \cup \{b\})}{g(A)}$$

if and only if $\Delta^b g(A) < g(A)\frac{d(e,b)}{d(e,A)}$, which is possible by a big enough choice of $d(e,b)$.

The following theorem establishes a constant-factor approximation for functions exhibiting diminishing g-relative marginal gains, with some assumption on g. This assumption reads

$$\Delta^e g(B) \leq \zeta \frac{g(A)}{|A|}, A \subseteq B,$$

i.e. the marginal gain of an element is bounded by a constant times the average value of an element in any subset. Interestingly, (9) is, for sum-dispersion, equivalent to

$$\frac{g(A)}{|A|}\frac{|A|}{d(e,A)}\frac{d(e, B \setminus A)}{|B \setminus A|} \leq \frac{1}{|B \setminus A|}\sum_{i=1}^{|B \setminus A|} \Delta^{b_i} g(A \cup \{b_1, \dots, b_{i-1}\})$$

for $\{b_1, \dots, b_{|B \setminus A|}\} \coloneqq B \setminus A$, which is implied by and equivalent to (for an appropiate metric)

$$\frac{g(A)}{|A|}\rho \leq \frac{1}{|B \setminus A|}\sum_{i=1}^{|B \setminus A|} \Delta^{b_i} g(A \cup \{b_1, \dots, b_{i-1}\})$$

for ρ the ratio of biggest and smallest distance within V. This means, that one cannot (for a general metric with a bounded ratio of biggest and smallest distance) choose ζ in the following theorem to be smaller than ρ. As the following result, even if we choose $\rho = \zeta$, is only better than the bound of Theorem 3.14 if $\rho \leq 0.51$, which it cannot take by definition (it is greater or equal to 1), the following result is of no use for the sum-dispersion function, but interesting for other function, as it generalized (set $\zeta = 0$) Theorem A.1.

4.3 Theorem *Let $f: \mathcal{P}(V) \to \mathbf{R}_+$ be monotone and exhibit diminishing g-relative marginal gains. Let furthermore $\Delta^e g(B) \leq \zeta \frac{g(A)}{|A|}$, $\zeta \geq 0$. Then Algorithm 1 yields a*

$$\left(1 - \frac{(1+\frac{2}{\zeta})^{\zeta}}{e} + o(1)\right)\text{-}approximation.$$

Proof: Let P^* be an optimal solution and $\{p_1, \dots, p_{|P^* \setminus P_i|}\} \coloneqq P^* \setminus P_i$. Then we have

$$\begin{aligned}
&f(P^*) - f(P_i) \leq f(P^* \cup P_i) - f(P_i) \\
&= \sum_{j=1}^{|P^* \setminus P_i|} \Delta^{p_j} f(P_i \cup \{p_1, \dots, p_{j-1}\}) \\
&\leq \sum_{j=1}^{|P^* \setminus P_i|} \Delta^{p_j} f(P_i)\frac{g(P_i \cup \{p_1, \dots, p_{j-1}\})}{g(P_i)} \\
&= \sum_{j=1}^{|P^* \setminus P_i|} \Delta^{p_j} f(P_i)\left(1 + \sum_{k=1}^{j-1}\frac{\Delta^{p_k} g(P_i \cup \{p_1, \dots, p_k\})}{g(P_i)}\right) \\
&\leq (f(P_{i+1}) - f(P_i))\sum_{j=1}^{k}\left(1 + \frac{\zeta(j-1)}{i}\right) \\
&= (f(P_{i+1}) - f(P_i))k\left(1 + \frac{\zeta(k-1)}{2i}\right) \\
&\leq (f(P^*) - f(P_i) - (f(P^*) - f(P_{i+1})))k\left(1 + \frac{\zeta k}{2i}\right)
\end{aligned}$$

Thus, after rearranging, and using $f(P_1) \geq 0$

$$f(P^*) - f(P_{i+1}) \leq \frac{k(1+\frac{\zeta k}{2i}) - 1}{k(1+\frac{\zeta k}{2i})}(f(P^*) - f(P_i))$$

Thus, we may invoke Theorem B.2 and get

$$f(P^*) - f(P_k) \le f(P^*) \prod_{i=1}^{k-1} \left(1 - \frac{1}{k} \frac{1}{1 - \frac{\zeta k}{2i}}\right)$$

$$\xrightarrow{k\to\infty} \prod_0^1 (1 - h(x)\,\mathrm{d}x) = \exp(\int_0^1 -h(x)\,\mathrm{d}x)$$

$$= \frac{\left(1 + \frac{2}{\zeta}\right)^{\zeta}}{e},$$

which yields the claim after rearranging. ■

See also Figure 1 for a plot of the integrand.

4.4 Remark (Open Problem) [BLY14] gave bound on the approximation ratio for Theorem 3.14 that *seemed* to be monotone with the same limit we proved here. Unfortunately, as the integrands in both of our proofs are monotone *decreasing*, we cannot use Proposition B.4, which would yield, for monotone integrands, immediate bounds. We believe, that the proofs may be altered to get non-asymptotic approximation ratios.

References

[Bha+10] Aditya Bhaskara et al. "Detecting high log-densities: an O (n $\frac{1}{4}$) approximation for densest k -subgraph". In: *Proceedings of the forty-second ACM symposium on theory of computing*. STOC '10. 2010.

[BLY14] Allan Borodin, Dai Le, and Yuli Ye. "Weakly Submodular Functions". In: *CoRR* abs/1401.6697 (2014). URL: `http://arxiv.org/abs/1401.6697`.

[Buc+12] N. Buchbinder et al. "A Tight Linear Time (1/2)-Approximation for Unconstrained Submodular Maximization". In: *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*. Oct. 2012, pp. 649–658. DOI: `10.1109/FOCS.2012.73`.

[Buc+14] Niv Buchbinder et al. "Submodular Maximization with Cardinality Constraints". In: *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms*. 2014, pp. 1433–1452.

[Fei+14] Uriel Feige et al. "A Unifying Hierarchy of Valuations with Complements and Substitutes". In: *Electronic Colloquium on Computational Complexity (ECCC)* 21.103 (2014).

[Fei98] Uriel Feige. "A Threshold of ln n for Approximating Set Cover". In: *Journal of the ACM* 45 (1998), pp. 314–318.

[FI14] Moran Feldman and Rani Izsak. "Constrained Monotone Function Maximization and the Supermodular Degree". In: *CoRR* abs/1407.6328 (2014). URL: `http://arxiv.org/abs/1407.6328`.

[FM07] Uriel Feige and Vahab S. Mirrokni. "Maximizing non-monotone submodular functions". In: *In Proceedings of 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS*. 2007, p. 2007.

[GKar] Daniel Golovin and Andreas Krause. "Submodular Function Maximization". In: *Tractability: Practical Approaches to Hard Problems*. to appear.

[KV12] Bernhard Korte and Jens Vygen. *Kombinatorische Optimierung*. 2nd ed. Springer-Verlag, 2012.

[Lee+09] Jon Lee et al. "Non-monotone submodular maximization under matroid and knapsack constraints". In: *CoRR* abs/0902.0353 (2009). URL: `http://arxiv.org/abs/0902.0353`.

[LLN01] Benny Lehmann, Daniel Lehmann, and Noam Nisan. "Combinatorial Auctions with Decreasing Marginal Utilities". In: *EC'01* (Oct. 2001).

[NWF78] G.L. Nemhauser, L.A. Wolsey, and M.L. Fisher. “An Analysis of Approximation for Maximizing Submodular Set Functions”. In: *Mathematical Programming* 14.1 (1978), pp. 265–294.

[Sla07] Antonín Slavik. *Product Integration, its History and Applications*. Vol. 29. History of Mathematics. Prague: MATFYZPRESS, 2007.

[Svi04] Maxim Sviridenko. “A note on maximizing a submodular set function subject to a knapsack constraint”. In: *Operations Research Letters* 32 (2004), pp. 41–43.

[Vol87] Vito Volterra. “Sui fondamenti della teoria delle equazioni differenziali lineari”. In: *Memorie della Società Italiana della Scienze* VI.3 (1887).

[Wol82] Laurence A. Wolsey. “Maximising Real-valued Submodular Functions: Primal and Dual Heuristics for Location Problems”. In: *Mathematics of Operations Research* 7.3 (1982), pp. 410–425.

A Submodular Function Maximization

In this section, we present known algorithms for the maximization of submodular functions. For several special cases of constraints on the maximizing set, greedy-like algorithms have been proposed ([Buc+14; NWF78; Buc+12; Lee+09]). The algorithms presented here are summarized in Table 1. In subsection A.1 we present two approximation algorithms for the maximization of monotone submodular functions subject to cardinality respectively knapsack constraints, in subsection A.2 we present two algorithms for unconstrained respectively cardinality-constrained maximization of submodular functions.

A.1 Monotone Objectives

The first approximation ratio we prove is

	$\max_{S\subseteq V} f(S)$	$\max_{\substack{S\subseteq V\\ \lvert S\rvert\le k}} f(S)$	$\max_{\substack{S\subseteq V\\ g(S)\le k}} f(S)$
MS	1	$1-e^{-1}$, 1	$1-e^{-1}$, 2
GS	$\frac{1}{2}$, 3	$\frac{1}{e}$, 4	$\frac{1}{5}-\varepsilon$, [Lee+09][1]

Table 1: Approximation ratios for algorithms presented in Appendix A for the case that f is monotone submodular (MS) or general submodular (GS). The last column has g modular, i.e. means maximization subject to knapsack constraints. After commas, the number of the algorithm that is proved to yield the approximation ratio written before the comma is given.

A.1 Theorem ([GKar, Theorem 1.3])
Algorithm 1 can be implemented in $O(kn)$ time with the same number of value oracle calls for f. If f is monotone and submodular it yields an output $P_k \subseteq V$ such that $f(P) \geq (1-\frac{1}{e}) \max_{|S|\leq k} f(S)$.

Proof: Since we can find the maximum of a set of size at most n in $O(n)$, the running time is clear.

Let P^* be an optimal solution and $\{p_1,\dots,p_{|P^*\setminus P_i|}\} := P^* \setminus P_i$. We, show, that for any $i \in [k]$

$$f(P_i) \geq \left(1-e^{-\frac{i}{k}}\right) f(P^*),$$

which yields the claim for $i = k$. Observe that

$$\begin{aligned}
&f(P^*) - f(P_i) \leq f(P^*\cup P_i) - f(P_i)\\
&= \sum_{j=1}^{|P^*\setminus P_i|} \underbrace{f(P_i\cup\{p_1,\dots,p_j\}) - f(P_i\cup\{p_1,\dots,p_{j-1}\})}_{\Delta^{p_j}(P_i\cup\{p_1,\dots,p_{j-1}\})}\\
&\leq \sum_{j=1}^{|P^*\setminus P_i|} \Delta^{p_j}(P_i)\\
&\leq \sum_{j=1}^{|P^*\setminus P_i|} (f(P_{i+1}) - f(P_i))\\
&\leq k(f(P_{i+1}) - f(P_i))\\
&= k(f(P^*) - f(P_i) - (f(P^*) - f(P_{i+1}))),
\end{aligned}$$

having used—for the first inequality—monotonicity of f, for the second its submodularity and for the third the choice in line 3 and for the fourth $|P^*| \le k$. We obtain

$$f(P^*) - f(P_{i+1}) \le \left(1 - \frac{1}{k}\right)(f(P^*) - f(P_i))$$
$$\le \left(1 - \frac{1}{k}\right)^{i+1}(f(P^*) - f(P_0)) \le \left(1 - \frac{1}{k}\right)^{i+1} f(P^*).$$

The right hand side is bounded by $e^{-\frac{i+1}{k}}$ (Lemma C.3). Rearranging yields the claim. ■

Interestingly, the same approximation ratio may be obtained for the more general knapsack constraints ([Svi04]), i.e. for the problem:

> **A.2 Problem (Knapsack)** *Given* $f{:}\mathcal{P}(V) \to \mathbf{R}_+$ *submodular,* $g{:}\mathcal{P}(V) \to \mathbf{Q}_+$ *modular,* $k \in \mathbf{R}$*, compute* $\max_{g(S)\le k} f(S)$

For this, Algorithm 1 is varied to include g in the Greedy choice (line 5), small sets are enumerated (line 1), and elements that make the current set infeasible are discarded from the universe (line 8).

A.3 Theorem ([Svi04, Theorem 1])
Algorithm 2 is correct and yields in $O(n^5)$ *oracle calls (for* g *and* f*) and* $O(n^5)$ *time a* $(1 - e^{-1})$*-approximation for Problem A.2.*

Proof: Correctness is easily observed, as $g(P_1) \le k$ by definition and each P_i^S is feasible as an invariant of the algorithm.

For the running time, observe, that line 1 needs $O(n^2)$ oracle calls and time. The For-loop in line 2 has $O(n^3)$ iterations, each consists of a while-loop in line 4 that is repeated $O(n)$ times, where a maximum over a set of size $O(n)$ has to be found. Altogether, the running time is obtained.

For the approximation ratio: First, we may, by scaling the function g appropiately, assume that it is integer-valued. Let P^* be an optimal solution. We may assume further, that $|P^*| > 3$, as otherwise, the optimal solution will be found by the algorithm. Consider an ordering $S = \{w_i\}_{i=1}^{|S|}$, such that

$$w_l = \arg\max_{w\in P^*\setminus\{w_1,\dots,w_{l-1}\}} \Delta^w f(\{w_1,\dots,w_{l-1}\})$$

Algorithm 2: Knapsack-Greedy

Input: $f{:}\mathcal{P}(V) \to \mathbf{R}_+^0$ submodular, $g{:}\mathcal{P}(V) \to \mathbf{Q}_+$ modular, $k \in \mathbf{R}$
Output: $P \subseteq V$, $g(P) \le k$

1 $P_1 \leftarrow \arg\max_{\substack{|S|\le 2\\ g(S)\le C}} f(S)$
2 **foreach** $S \in \binom{V}{3}$ **do**
3 $\quad V_0^S \leftarrow V \setminus S$, $P_0^S \leftarrow S$, $i \leftarrow 1$
4 $\quad$**while** $V_{i-1}^S \ne \varnothing$ **do**
5 $\quad\quad e \leftarrow \arg\max_{e'\in V_{i-1}^S \setminus P_{i-1}^S} \frac{\Delta^{e'} f(P_{i-1}^S)}{g(e')}$
6 $\quad\quad$**if** $g(P_{i-1}^S \cup \{e\}) \le k$ **then**
7 $\quad\quad\quad P_i^S \leftarrow P_{i-1}^S \cup \{e\}$
8 $\quad\quad$**else**
9 $\quad\quad\quad P_i^S \leftarrow P_{i-1}^S, V_i^S \leftarrow V_{i-1}^S \setminus \{e\}$
10 $\quad\quad i \leftarrow i + 1$
11 $\quad P^S \leftarrow P_i^S$
12 $P_2 \leftarrow \arg\max_{S\in\binom{V}{3}} f(P^S)$
13 **return** $\arg\max\{f(P_1), f(P_2)\}$

We claim, that $P_i^{\{w_1,w_2,w_3\}}$ for a high enough i gives a $(1-e^{-1})$-approximation, which establishes the claim. We consider only this iteration of line 2 and suppress therefore the superscript. Observe, that by submodularity and the choice of the ordering $\{w_i\}_{i=1}^{|S|}$ we have for any $l \ge 4$ and $Z \in V \setminus \{w_1, w_2, w_3, w_l\}$

$$\Delta^{w_l} f(\{w_1, w_2, w_3\} \cup Z)$$
$$\le \begin{cases} \Delta^{w_l} f(\varnothing) \le f(\{w_1\}) \\ \Delta^{w_l} f(\{w_1\}) \le \Delta^{w_2} f(\{w_1\}) \\ \Delta^{w_3} f(\{w_1, w_2\}), \end{cases}$$

where in the first inequality $f(\varnothing) \ge 0$ was used. Summing these inequalities, this yields

$$\Delta^{w_l} f(\{w_1, w_2, w_3\} \cup Z) \le \frac{1}{3} f(\{w_1, w_2, w_3\}) \quad (10)$$

We now consider the first iteration, in which line 8 is executed. Let $y_1, \dots, y_{t^*}$ be the elements added to w_1, w_2, w_3 before and y_{t^*+1} be the element not added to the set (so $P_{t^*} = \{w_1, w_2, w_3, y_1, \dots, y_{t^*}\}$). Without loss of generality $y_{t^*+1} = w_k$, as another element not being added to an approximate solution when considered cannot be added to it later and may thus be expelled from V as it cannot change the approximation ratio. We have

$$g(P_{t^*} \cup \{w_k\}) > k \tag{11}$$

and for $t \in [t^*]$ by submodularity and $\{w_1, w_2, w_3\} \subseteq P^* \cap P_{t^*}$

$$\begin{aligned} f(P^*) &\leq f(P_t) + \sum_{i \in P^* \setminus P_t} \Delta^i f(P_t) \\ &\leq f(P_t) + \left(\max_{j \in P^* \setminus P_t} \frac{\Delta^j f(P_t)}{g(\{j\})} \right) \sum_{i \in P^* \setminus P_t} g(i) \\ &\leq f(P_t) + \left(\max_{j \in P^* \setminus P_t} \frac{\Delta^j f(P_t)}{g(\{j\})} \right) (k - g(\{w_1, w_2, w_3\})), \end{aligned} \tag{12}$$

Now define

$$\rho_k = \max_{j \in P^* \setminus P_{t-1}} \frac{\Delta^j f(P_{t-1})}{g(\{j\})}$$

if $k \in \{g(\{y_1, \ldots, y_{t-1}\}) + 1, \ldots, g(\{y_1, \ldots, y_t\})\}$. This means, taking the minimum over bounds (12)

$$\begin{aligned} &\frac{f(P_{t^*} \cup \{y_{t^*+1}\}) - f(\{w_1, w_2, w_3\})}{f(P^*) - f(\{w_1, w_2, w_3\})} \\ &\geq \frac{f(\{y_1, \ldots, y_{t^*+1}\})}{\left(\min\limits_{t \in [t^*]} f(\{y_1, \ldots y_t\}) + \max\limits_{j \in P^* \setminus P_t} \frac{\Delta^j f(P_t)}{g(\{j\})} (k - g(\{w_1, w_2, w_3\})) \right)} \\ &= \frac{\sum_{j=1}^{g(\{y_1, \ldots, y_{t^*+1}\})} \rho_j}{\min\limits_{s \in [g(\{y_1, \ldots, y_{t^*+1}\})]} \sum_{j=1}^{s-1} \rho_j + (k - g(\{w_1, w_2, w_3\})) \rho_s}. \end{aligned}$$

Now, we may use Lemma C.5 , as $g(\{y_1, \ldots, y_{t^*+1}\}) > B - g(\{w_1, w_2, w_3\})$ by (11) to bound this by

$$1 - \exp\left(-\frac{g(\{y_1, \ldots, y_{t^*+1}\})}{B - g(\{y_1, y_2, y_3\})} \right) > 1 - \exp(-1) \tag{13}$$

Putting together (10) and (13), we get

$$\begin{aligned} f(P_{t^*}) &= f(\{w_1, w_2, w_3\}) + f(P_{t^*} \cup \{w_k\}) - \Delta^{w_k} f(S_{t^*}) \\ &\geq f(\{w_1, w_2, w_3\}) + (1 - \exp(-1)) \\ &\quad (f(P^*) - f(\{w_1, w_2, w_3\})) - \frac{f(\{w_1, w_2, w_3\})}{3} \\ &\geq (1 - \exp(-1)) f(P^*), \end{aligned}$$

which concludes the proof. ∎

A.2 Non-Monotone Objectives

This section is devoted to the maximization of *non*-monotone submodular functions. In contrast to the monotone case, here even unconstrained maximization is non-trivial.The next algorithm presented, algorithm 3, deals with this. For the cardinality-constrained maximization, we

We present a randomized algorithm from [Buc+12] that yield an expected $\frac{1}{2}$–approximation for the unconstrained maximization of general submodular functions (see algorithm 3).

A.4 Remark In fact, [Buc+12, section II] also prove a $\frac{1}{3}$-approximation ratio for a deterministic algorithm, in which line 3 is executed if and only if $\Delta^{v_i} f(P_{i-1}) \geq \Delta_{v_i} f(Q_{i-1})$. The proof of this result is shorter than the proof given here.

Algorithm 3: Double Greedy

Input: $f\colon \mathcal{P}(V) \to \mathbf{R}_+^0$ submodular
Output: $P_n \subseteq V$
1 $P_0 \leftarrow \varnothing$, $Q_0 \leftarrow V$
2 **for** $i = 1$ **to** n **do**
3   **with probability** $\frac{\Delta^{v_i}(P_{i-1})^+}{\Delta_{v_i}(Q_{i-1})^+ + \Delta_{v_i}(Q_{i-1})^+}$ **do**
4         $P_i \leftarrow P_{i-1} \cup \{v_i\}$, $Q_i \leftarrow Q_{i-1}$
5   **otherwise do**
6         $P_i \leftarrow P_{i-1}$, $Q_i \leftarrow Q_{i-1} \setminus \{v_i\}$
7 **return** P_n

Obviously, the algorithm is correct and can be implemented in linear time and a linear number of calls of the value oracle for f. The approximation ratio is as follows:

A.5 Theorem ([Buc+12, Section III]) *algorithm 3 yields an expected $\frac{1}{2}$-approximation for unconstrained submodular maximization.*

Proof: Let P^* be an optimal solution. We define the random sets $P_i^* \coloneqq (P^* \cup P_i) \cap Q_i$. Then $P_0^* = P^*$, $P_n^* = P_n = Q_n$ (observe that P_i and Q_i are equal on the elements $v_1, \dots, v_i$). Thus, $f(P_0^*) - f(P_i^*)$ reflects the loss of value of an optimal solution throughout the algorithm. We claim:

$$\begin{aligned}&\mathbf{E}[f(P_{i-1}^*) - f(P_i^*)]\\ &\quad\le \frac{\mathbf{E}[f(P_i) - f(P_{i-1}) + f(Q_i) - f(Q_{i-1})]}{2};\end{aligned} \tag{14_i}$$

This readily implies the thesis of the theorem, since $\sum_{i=1}^n (14_i)$ takes the form (observe telescopic sums)

$$\begin{aligned}&\mathbf{E}[f(P_0^*) - f(P_n^*)]\\ &\quad\le \frac{\mathbf{E}[f(P_n) - f(P_0) + f(Q_n) - f(Q_0)]}{2}\\ &\quad\le \frac{\mathbf{E}[f(P_n) + f(Q_n)]}{2}.\end{aligned}$$

This yields, after rearranging, $\mathbf{E}[f(P_n)] \ge \frac{1}{2} f(P^*)$, as $P_n^* = P_n = Q_n$, $P_0^* = P^*$. To prove (14_i), it suffices to show the inequalities for the conditional expectations given the events $A_i = \{P_{i-1} = S_{i-1}\}$, $S_{i-1} \subseteq \{v_1, \dots, v_{i-1}\}$ (for, taking the expectation over the values of P_{i-1}, these inequalities yield the very inequality (14_i)). Conditional on A_i, $Q_{i-1} = S_{i-1} \cup \{v_i, \dots, v_n\}$ almost surely and, by their respective definitions, P_{i-1}^* and $\Delta^{v_i} f(P_{i-1})$ and $\Delta_{v_i} f(Q_{i-1})$ are deterministic. We consider the signs of these last two values in line 3:

If $\Delta^{v_i} f(P_{i-1}), \Delta_{v_i} f(Q_{i-1}) > 0$, we have

$$\begin{aligned}&\mathbf{E}[f(P_i) - f(P_{i-1}) + f(Q_i) - f(Q_{i-1}) | A_i]\\ &\quad= \frac{\Delta^{v_i} f(P_{i-1})}{\Delta^{v_i} f(P_{i-1}) + \Delta_{v_i} f(Q_{i-1})} \Delta^{v_i} f(P_{i-1})\\ &\quad+ \frac{\Delta_{v_i} f(Q_{i-1})}{\Delta^{v_i} f(P_{i-1}) + \Delta_{v_i} f(Q_{i-1})} \Delta_{v_i} f(Q_{i-1})\\ &\quad= \frac{(\Delta^{v_i} f(P_{i-1}))^2 + (\Delta_{v_i} f(Q_{i-1}))^2}{\Delta^{v_i} f(P_{i-1}) + \Delta_{v_i} f(Q_{i-1})},\end{aligned} \tag{15}$$

where the left ratios give respectively the probability of the event that an element is added to P_{i-1} or that one is deleted from Q_{i-1}. Furthermore,

$$\begin{aligned}&\mathbf{E}[f(P_{i-1}^*) - f(P_i^*) | A_i]\\ &\quad= \frac{\Delta^{v_i}(P_{i-1})}{\Delta^{v_i}(P_{i-1}) + \Delta_{v_i}(Q_{i-1})} (-\Delta^{v_i}(P_{i-1}^*))\\ &\quad+ \frac{\Delta_{v_i}(Q_{i-1})}{\Delta^{v_i}(P_{i-1}) + \Delta_{v_i}(Q_{i-1})} (-\Delta_{v_i}(P_{i-1}^*)).\end{aligned} \tag{16}$$

If $v_i \in P_{i-1}^*$, $\Delta^{v_i} f(P_{i-1}^*) = 0$ and

$$-\Delta_{v_i}(P_{i-1}^*) \overset{(2)}{=} \Delta^{v_i}(P_{i-1}^* \setminus \{v_i\}) \le \Delta^{v_i}(P_{i-1})$$

as $P_{i-1} \subseteq P_{i-1}^* \setminus \{v_i\}$ by submodularity. Otherwise, $\Delta_{v_i} f(P_{i-1}^*) = 0$ and

$$-\Delta^{v_i} f(P_{i-1}^*) \overset{(2)}{=} \Delta_{v_i} f(P_{i-1}^* \cup \{v_i\}) \le \Delta_{v_i} f(Q_{i-1})$$

as $P_{i-1}^* \cup \{v_i\} \subseteq Q_{i-1}$ (note $\{v_i, \dots, v_n\} \subseteq Q_{i-1}$). Consequently, (16) can be bounded by

$$\frac{\Delta^{v_i} f(P_{i-1}) \Delta_{v_i} f(Q_{i-1})}{\Delta^{v_i} f(P_{i-1}) + \Delta_{v_i} f(Q_{i-1})}$$

which is smaller or equal than $\frac{1}{2}(15)$ by Lemma C.4. Thus, in this case, (14_i) is proved.

We remark that $\Delta^{v_i} f(P_{i-1}) + \Delta_{v_i} f(Q_{i-1}) = f(P_{i-1} \cup \{v_i\}) - f(P_{i-1}) + f(Q_{i-1} \setminus \{v_i\}) - f(Q_{i-1}) \ge 0$ by the definition of submodularity (1) with $A \coloneqq Q_{i-1} \setminus \{v_i\}$ and $B \coloneqq P_{i-1} \cup \{v_i\}$. Therefore, we may now assume one of the values to be nonnegative.

In this case, the choice in line 3 is deterministic. For the case $\Delta^{v_i} f(P_{i-1}) \ge 0$, (14_i) simplifies to

$$\begin{aligned}&f(P_{i-1}^*) - f(P_{i-1}^* \cup \{v_i\})\\ &\quad\le \frac{f(P_{i-1} \cup \{v_i\}) - f(P_{i-1})}{2}\end{aligned}$$

or equivalently

$$\Delta_{v_i} f(P_{i-1}^* \cup \{v_i\}) \le \frac{\Delta^{v_i} f(P_i)}{2}.$$

But since $P_{i-1}^* \cup \{v_i\} \subseteq Q_{i-1}$, by assumption,

$$\Delta_{v_i} f(P_{i-1}^* \cup \{v_i\}) \le \Delta_{v_i} f(Q_{i-1}) \le 0 \le \Delta^{v_i} f(P_{i-1})$$

The case $\Delta_{v_i} f(Q_{i-1}) > 0$ is similar.[2]

This gives finally (14_i) and concludes the proof. ■

The last algorithm in this section, algorithm 4, yields a e^{-1}-approximation for cardinality-constrained maximization for general, i.e. not necessarily monotone, submodular functions. For this, we need use the following notation: For a submodular function $f\colon \mathcal{P}(V) \to \mathbf{R}_+$ we let

$$f'\colon \mathcal{P}(V') \to \mathbf{R}_+,$$
$$f'(R) = f(R \cap V), V' \coloneqq V \cup \{w_1, \dots, w_{2k}\}$$

which clearly is then also submodular.

Algorithm 4: Random Greedy

Input: $f\colon \mathcal{P}(V) \to \mathbf{R}_+^0$ submodular
Output: $P \subseteq V$, $\mathbf{E}[f(P)] \geq e^{-1} \max_{\substack{S\subseteq V\\|S|\leq k}} f(S)$

```
1 P_0 ← ∅
2 for i = 1 to k do
3     Let M_i ∈ arg max_{T⊆V'∖P_{i-1}, |T|=k} ∑_{e∈T} Δ^e f'(P_{i-1})
4     if M_i ≠ ∅ then
5         e_i ~ Unif_{M_i}
6         P_i ← P_{i-1} ∪ {e_i}
7 return P_n ∩ V
```

A.6 Remark Instead of considering f', one may also change the algorithm, such M_i is chosen as a set of size *at most* k maximinizing line 3 and in line line 5, an element is added only with a probability $\frac{|M_i|}{k}$.

A.7 Theorem ([Buc+14, Theorem 1.3]) *algorithm 4 is correct and can be implemented in running time $O(nk)$ and the same number of value oracle calls for f.*

Proof: The choice in line 3 consists of finding the k best elements in $V \setminus P_{i-1}$ with respect to the maximization of $f(P_{i-1} \cup \cdot\,)$ that do not yield negative change in function value. This choice can be made in $O(n)$. The rest of a for loop is $O(\log n)$ and there are k iterations. This yields the desired running time.

Let P^* denote an optimal solution. We first prove two inequalities. The first is: For any $S_{i-1} \subseteq V$

$$\begin{aligned}\mathbf{E}[\Delta^{e_i} f(P_{i-1}) | P_{i-1} = S_{i-1}] &= \sum_{e\in M_i} \frac{1}{k} \Delta^e(S_{i-1})\\ &\geq \frac{1}{k} \sum_{e\in P^*\setminus S_{i-1}} \Delta^e f(S_{i-1})\\ &\geq \frac{1}{k}(f(P^* \cup S_{i-1}) - f(S_{i-1})), \quad (17)\end{aligned}$$

where the first inequality is due to the fact, that $H_i \coloneqq P^* \setminus S_{i-1} \cup \{w_1, \dots, w_j\}$ with j such $|H_i| = k$ is also object to the maximization in line 5 and the second due to submodularity using a telescopic sum. Taking the expectation over all values of S_{i-1}, we get an unconditional inequality. For the second,

$$\begin{aligned}\mathbf{P}[e \in P_i \setminus P^*] &\leq \mathbf{P}[e \in P_i]\\ &= 1 - \prod_{j=1}^{i} \mathbf{P}[e \notin P_i \setminus P_{i-1}] \leq 1 - \left(1 - \frac{1}{k}\right)^i, \quad (18)\end{aligned}$$

as $e \in V \setminus P_{i-1}$ is *not* chosen in line 5 in the ith iteration with probability at least $1 - \frac{1}{k}$. Let $\{p_1, \dots, p_{|V'\setminus P^*|}\} \coloneqq V' \setminus P^*$ with $(\mathbf{P}[p_i \in P_i])_{i=1}^{|V'\setminus P^*|}$ decreasing.[3] Then, using

$$\begin{aligned}&f'(P^* \cup P_i) = f'(P^* \uplus (P_i \cap \{p_1, \dots, p_{|V\setminus P^*|}\}))\\ &= f(P^*) + \sum_{j=1}^{|V'\setminus P^*|} \mathbb{1}_{\{p_j \in P_i\}} \Delta^{p_j} f'(P^* \uplus (P_i \cap \{p_1, \dots, p_{j-1}\}))\end{aligned}$$

we have

$$\begin{aligned}&\mathbf{E}[f'(P^* \cup P_i)] = \mathbf{E}[f(P^*)\\ &\quad + \sum_{j=1}^{|V'\setminus P^*|} \mathbb{1}_{\{p_j \in P_i\}} \Delta^{p_j} f'(P^* \uplus (P_i \cap \{p_1, \dots, p_{j-1}\}))]\\ &\geq \mathbf{E}[f(P^*) + \sum_{j=1}^{|V'\setminus P^*|} \mathbb{1}_{\{p_j \in P_i\}} \Delta^{p_j} f'(P^* \uplus \{p_1, \dots, p_{j-1}\})]\\ &= f(P^*) + \sum_{j=1}^{|V'\setminus P^*|} \mathbf{P}[p_j \in P_i] \Delta^{p_j} f'(P^* \uplus \{p_1, \dots, p_{j-1}\})\end{aligned}$$

[2] Then is suffices to prove $\Delta^{v_i}(P_{i-1}^* \setminus \{v_i\}) \leq \Delta^{v_i} f(P_{i-1})$, which holds by $P_{i-1}^* \setminus \{v_i\} \supseteq P_{i-1}$.

[3] Note $\mathbf{P}[p_i \in P_i \cup P^*] = \mathbf{P}[p_i \in P_i]$ for any $p_i \in V' \setminus P^*$.

$$
\begin{aligned}
&= (1 - \mathbf{P}[p_1 \in P_i]) f(P^*) \\
&\quad + \sum_{j=1}^{|V' \setminus P^*|-1} (\mathbf{P}[p_{j-1} \in P_i \cup P^*] \\
&\quad - \mathbf{P}[p_j \in P_i \cup P^*]) f'(P^* \uplus \{p_1, \dots, p_{j-1}\}) \\
&\quad + \mathbf{P}[p_{|V' \setminus P^*|} \in P_i] f'(P^* \cup P_i) \\
&\geq (1 - \mathbf{P}[p_1 \in P^* \cup P_i]) f(P^*) \\
&\geq \left(1 - \frac{1}{k}\right)^i f(P^*).
\end{aligned} \tag{19}
$$

where the first inequality is a consequence of f's submodularity, the second one of the choice of the ordering of $(p_i)_{i=1}^{|V \setminus P^*|}$ and the third of (18). Consequently,

$$
\begin{aligned}
\mathbf{E}[f(P_i)] &= \mathbf{E}[f(P_{i-1})] + \mathbf{E}[\Delta^{e_i} f(P_{i-1})] \\
&\overset{(17)}{\geq} \mathbf{E}[f(P_{i-1})] + \frac{1}{k}(f(P^* \cup P_{i-1}) - f(P_{i-1})) \\
&\overset{(19)}{\geq} \left(1 - \frac{1}{k}\right) \mathbf{E}[f(P_{i-1})] + \frac{(1 - \frac{1}{k})^i}{k} f(P^*) \\
&= \left(1 - \frac{1}{k}\right) f(P_0) + \frac{k}{k}\left(1 - \frac{1}{k}\right)^{k-1} f(P^*) \\
&\geq e^{-1} f(P^*).
\end{aligned}
$$

The last inequality is due to $f(P_0) = f(\emptyset) \geq 0$ and by Lemma C.3. This concludes the proof. ∎

A.3 Concluding Remarks

The approximation ratio proved obtainable in polynomial time in Theorem A.7 is not best possible, as [Buc+14] show by perturbing the size of the set M_i in line 3 of algorithm 4 (this yields a minor improvement of the approximation of 0.004). However, Theorem A.1 and Theorem A.5 yields (assuming $\mathtt{P} \neq \mathtt{NP}$) the best approximation ratio obtainable in polynomial time, as [Fei98] respectively [FM07, Theorem 4.5] show. Clearly, as Theorem A.3 implies the same bound for the special case of cardinality constraints, this is also best possible.

A.8 Remark (Nonnegativity) We only consider non-negative functions. This is not restrictive with respect to exact computations, as addition of a constant pertains (1) and a lower bound on the function values of a submodular functions can be computed efficiently as one easily deduces from [KV12, Proposition 14.16]).

But, since we only aim to give approximation algorithms, this addition of a constant is restrictive as it changes the quality of approximation.

B Product Integration

Italian Mathematician Vito Volterra (1860-1940) developed in the late 1880s a multiplicative calculus for matrix-valued functions (most relevant to us is [Vol87]) for the solution of integral equations. We will only state very few of the properties of the product integral. For a more thorough treatment of the history of product integration [Sla07]. In the following, denote by $\int \cdot \, \mathrm{d}x$ the Riemann integral. We fix $f\colon [0,1] \to \mathbf{R}$ Riemann integrable. One easily obtains analogous results for intervals with arbitrary endpoints.

B.1 Definition (Product Integral) *Let $f\colon[0,1]$ be Riemann-integrable. Then we define*

$$
\prod_0^1 (1 + f(x)\,\mathrm{d}x) \coloneqq \lim_{k \to \infty} \prod_{i=1}^k \left(1 + \frac{1}{k} f\left(\frac{i}{k}\right)\right)
$$

if the limit exists (it does: Theorem B.2).

B.2 Theorem *$\prod_0^1 (1 + f(x)\,\mathrm{d}x)$ exists for any Riemann-integrable f and is equal to*

$$
\exp(\int_0^1 f(x)\,\mathrm{d}x)
$$

Proof: We prove

$$
\prod_0^1 (1 + f(x)\,\mathrm{d}x) = \lim_{k \to \infty} \prod_{i=1}^m e^{\frac{f(\frac{i}{k})}{k}}
$$

which implies the claim by using the functional equation for exp and its continuity, and furthermore, that

$$
\lim_{k \to \infty} \sum_{i=1}^k \frac{f(\frac{i}{k})}{k} = \int_0^1 f(x)\,\mathrm{d}x.
$$

Now let $|f| \leq M$ (properly Riemann-integrable functions are bounded). Then by regarding exp's series expansion

$$\left|e^{\frac{f(\frac{i}{k})}{k}} - \left(1 + \frac{f(\frac{i}{k})}{k}\right)\right| \le \left|\frac{f(\frac{i}{k})}{k}\right|^2 \exp\left(\frac{f(\frac{i}{k})}{k}\right) \le \left(\frac{M}{k}\right)^2 e^{\frac{M}{k}}. \quad (20)$$

Furthermore,

$$\begin{aligned}
&\left|\prod_{i=1}^{k} e^{\frac{f(\frac{i}{k})}{k}} - \prod_{i=1}^{k}\left(1 + \frac{f(\frac{i}{k})}{k}\right)\right| \\
&\overset{(23)}{=} \left|\sum_{j=1}^{k}\left(\prod_{i=1}^{j-1}\left(1 + \frac{f(\frac{i}{k})}{k}\right)\right)\left(e^{\frac{f(\frac{i}{k})}{k}} - 1 - \frac{f(\frac{i}{k})}{k}\right)\prod_{i=j+1}^{k} e^{\frac{f(\frac{i}{k})}{k}}\right)\right| \\
&\le \left|\sum_{j=1}^{k}\left(\prod_{i=1}^{j-1} e^{\frac{f(\frac{i}{k})}{k}}\left(e^{\frac{f(\frac{i}{k})}{k}} - 1 - \frac{f(\frac{i}{k})}{k}\right)\prod_{i=j+1}^{k} e^{\frac{f(\frac{i}{k})}{k}}\right)\right| \\
&\le \left|\sum_{j=1}^{k}\left(e^{\frac{f(\frac{i}{k})}{k}} - 1 - \frac{f(\frac{i}{k})}{k}\right)e^{M}\right| \\
&\overset{(20)}{=} e^{M} M^2 \sum_{i=1}^{k} \frac{e^{\frac{M}{k}}}{k^2} = e^{M} M^2 \frac{e^{M}}{k} \xrightarrow{k\to\infty} 0
\end{aligned}$$

where we used for the first inequality Lemma C.3 and for the second the functional equation for exp and the definition of M. ■

We therefore get immediately from the properties of the Riemann integral:

B.3 Proposition *Let f be Riemann-integrable*

- $\prod_0^1(1 + \cdot \,\mathrm{d}x)$ *is monotone*
- *for any $a < b < c$:*

$$\prod_a^c(1+f(x)\,\mathrm{d}x) = \prod_a^b(1+f(x)\,\mathrm{d}x)\prod_b^c(1+f(x)\,\mathrm{d}x) \quad (21)$$

B.4 Proposition *Let f be Riemann-integrable and monotone increasing. Then for any k*

$$\prod_{i=0}^{k-1}\left(1 + \frac{1}{k}f\left(\frac{i}{k}\right)\right) \le \prod_0^1(1 + f(x)\,\mathrm{d}x)$$

Proof: t suffices to observe with Lemma C.3

$$\begin{aligned}
\prod_0^1(1 + f(x)\,\mathrm{d}x) &= \prod_{i=1}^{k}\prod_{\frac{i-1}{k}}^{\frac{i}{k}}(1 + f(x)\,\mathrm{d}x) \\
&= \prod_{i=0}^{k-1}\exp\Big(\int_{\frac{i}{k}}^{\frac{i+1}{k}} f(x)\,\mathrm{d}x\Big) \\
&\ge \prod_{i=0}^{k-1x}\Big(1 + \int_{\frac{i}{k}}^{\frac{i+1}{k}} f(x)\,\mathrm{d}x\Big) \\
&\ge \prod_{i=0}^{k-1}\Big(1 + \int_{\frac{i}{k}}^{\frac{i+1}{k}} f\left(\frac{i}{k}\right)\mathrm{d}x\Big) \\
&= \prod_{i=0}^{k-1}\Big(1 + \frac{1}{k}f\left(\frac{i}{k}\right)\mathrm{d}x\Big),
\end{aligned}$$

where for the first inequality Lemma C.3, and for the second f's monotonicity was used. ■

One might wonder, whether the product integral may be upper bounded by the "upper products", in analogy to the Riemann case. This is not true. Take $f(x) = 1$. Then

$$\prod_0^1(1 + 1\,\mathrm{d}x) = \exp(1) > \prod_{i=1}^{1}(1+1) = 2$$

C Mathematical Tools

In this section we state and prove lemmas that are used throughout the article.

C.1 Lemma *Let $x \in \mathbf{R}$. Then*

$$\sum_{i=0}^{n} x^i = \frac{1 - x^{n+1}}{1 - x} \quad (22)$$

Proof: Denote $S_n(x) \coloneqq \sum_{i=0}^{n} x^i$. Then

$$xS_n(x) - S_n(x) = x^{n+1} - 1 \Leftrightarrow S_n(x) = \frac{1 - x^{n+1}}{1 - x}$$ ■

C.2 Lemma *Let $x \in \mathbf{R}$. Then*

$$\sum_{i=1}^{n}(i+1)x^{i-1} = \frac{nx^{n+1} - (n+1)x^n + 1}{(1-x)^2}$$

Proof: Let $S_n(x)$ be as in the proof of Lemma C.1.

$$\frac{\partial}{\partial x}\sum_{i=0}^{n} x^i \overset{(22)}{=} \frac{\partial}{\partial x}\frac{1-x^{n+1}}{1-x}$$
$$\Leftrightarrow \sum_{i=1}^{n} ix^{i-1} = \frac{nx^{n+1}-(n+1)x^n+1}{(1-x)^2} \qquad \blacksquare$$

C.3 Lemma *For any* $z \in \mathbf{R}$ $1+z \le \exp(z)$.

Proof: For $z \le -1$: $1-z \le 0 \le e^z$. Otherwise equivalent to: for all $x \ge 0$

$$x \le \exp(x-1) \Leftrightarrow \ln x \le x-1$$

which is proved using Taylor's theorem (taylor expansion around 1 of ln to degree 1). $\blacksquare$

C.4 Lemma *Let* $x, y \ge 0$. *Then* $\frac{x+y}{2} \ge \sqrt{xy}$

Proof: $(\sqrt{x}+\sqrt{y})^2 \ge 0 \Rightarrow x+y \ge 2\sqrt{xy}$ $\blacksquare$

C.5 Lemma ([Wol82, Theorem 0]) *Let* $B,\ D \in \mathbf{Z}_+$, $(\rho_i)_{i=1}^P \subseteq \mathbf{R}_+$. *If* $\min_{t\in[P]}\sum_{i=1}^{t-1}\rho_i + D\rho_t \ge 0$, *we have*

$$\frac{\sum_{i=1}^{P}\rho_i}{\min_{t\in[P]}\sum_{i=1}^{t-1}\rho_i + D\rho_t} \ge 1 - e^{-\frac{P}{D}}$$

Proof: We can rescale all ρ such that

$$\min_{t\in[P]}\rho_i\sum_{i=1}^{t-1}\rho_i + D\rho_t = 1$$

as this leaves the quotient we want to bound unchanged. This quotient must be bounded from below by optimal solution value of the linear program

$$\min\sum_{i=1}^{P}\rho_i \text{ s.t. } \sum_{i=1}^{t-1}\rho_i + D\rho_t \ge 1, \quad \forall t \in [P]$$

whose dual is equivalent to

$$\min\sum_{i=1}^{P}\eta_{p-i} \text{ s.t. } \sum_{i=1}^{t-1}\eta_{p-i} + D\eta_{p-t} \ge 1, \quad \forall t \in [P]$$

$\rho_t = \eta_{P-t} = \frac{1}{D}\left(\frac{D-1}{D}\right)^{t-1}$ is primal respectively dual feasible with identical objective value $1-\left(\frac{D-1}{D}\right)^P$. $\blacksquare$

C.6 Lemma *Let* $(a_i)_{i=1}^n$, $(b_i)_{i=1}^n \subset \mathbf{R}$. *Then*

$$\prod_{i=1}^{n} a_i - \prod_{i=1}^{n} b_i \le \sum_{i=1}^{n}\left(\prod_{j=1}^{i-1} a_j(a_i-b_i)\prod_{j=i+1}^{n} b_j\right) \tag{23}$$

Proof: Observe that the right hand side of (23) is a telescopic sum. $\blacksquare$