A Note on the Budgeted Maximization of Submodular Functions

Andreas Krause **Carlos Guestrin**

March 2005
CMU-CALD-05-103

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

Many set functions F in combinatorial optimization satisfy the diminishing returns property $F(\mathcal{A} \cup X) - F(A) \geq F(\mathcal{A}' \cup X) - F(\mathcal{A}')$ for $\mathcal{A} \subset \mathcal{A}'$. Such functions are called *submodular*. A result from Nemhauser et.al. states that the problem of selecting k-element subsets maximizing a nondecreasing submodular function can be approximated with a constant factor $(1 - 1/e)$ performance guarantee. Khuller et.al. showed that for the special submodular function involved in the MAX-COVER problem, this approximation result generalizes to a budgeted setting under linear nonnegative cost-functions. In this note, we extend this result to general submodular functions. Motivated by the problem of maximizing entropy in discrete graphical models, where the submodular objective cannot be evaluated exactly, we generalize our result to account for absolute errors.

Keywords: Submodular functions; Optimization; Constraints; Entropy maximization

1 Introduction

Many set functions F in combinatorial optimization satisfy the diminishing returns property $F(\mathcal{A} \cup X) - F(A) \geq F(\mathcal{A}' \cup X) - F(\mathcal{A}')$ for $\mathcal{A} \subset \mathcal{A}'$, i.e. adding an element to a smaller set helps more than adding it to a larger set. Such functions are called *submodular*. The submodular function motivating our research is the joint entropy $H(\mathcal{A})$ for a set of random variables $\mathcal{A}$. The entropy of a distribution $P : \{x_1, \ldots, x_d\} \to [0, 1]$ is defined as

$$H(P) = -\sum_k P(x_k) \log P(x_k),$$

measuring the number of bits required to encode $\{x_1, \ldots, x_d\}$ [1]. If $\mathcal{A}$ is a set of discrete random variables $\mathcal{A} = \{X_1, \ldots, X_n\}$, then their entropy $H(\mathcal{A})$ is defined as the entropy of their joint distribution. The conditional entropy $H(\mathcal{A} \mid \mathcal{B})$ for two subsets $\mathcal{A}, \mathcal{B} \subseteq \mathcal{V}$ is defined as

$$H(\mathcal{A} \mid \mathcal{B}) = -\sum_{\substack{\mathbf{a} \in \operatorname{dom} \mathcal{A} \\ \mathbf{b} \in \operatorname{dom} \mathcal{B}}} P(\mathbf{a}, \mathbf{b}) \log P(\mathbf{a} \mid \mathbf{b}),$$

measuring the expected uncertainty about variables $\mathcal{A}$ after variables $\mathcal{B}$ are observed.

Using the chain-rule of entropies [1], $H(\mathcal{A} \cup \mathcal{B}) = H(\mathcal{A} \mid \mathcal{B}) + H(\mathcal{B})$, we can compute $H(\mathcal{A} \cup X) - H(\mathcal{A}) = H(X|\mathcal{A})$. The *information never hurts* principle [1], $H(X \mid \mathcal{A}) \geq H(X \mid \mathcal{A}')$ for all $\mathcal{A} \subseteq \mathcal{A}'$, proves submodularity of the entropy. In the discrete setting, $H(X \mid \mathcal{A})$ is also always non-negative, hence the entropy is nondecreasing.

In practice, a commonly used algorithm for selecting a set of variables with maximum entropy is to greedily select the next variable to observe as the most uncertain variable given the ones observed thus far:

$$X_k := \operatorname*{argmax}_X H(X \mid \{X_1, \ldots X_{k-1}\}), \tag{1.1}$$

which is again motivated by the chain-rule.

It is no surprise that this problem has been tackled with heuristic approaches, since even the unit cost has been shown to be $\mathbf{NP}$-hard for multivariate Gaussian distributions [3], and a related formulation has been shown to be $\mathbf{NP}^{\mathbf{PP}}$-hard even for discrete distributions that can be represented by polytree graphical models [4].

Fortunately, a result from Nemhauser et.al. [5] states that the problem of selecting k-element subsets maximizing a nondecreasing submodular function can be approximated with a constant factor $(1-1/e)$ performance guarantee, using the greedy algorithm as mentioned above. Khuller et.al. [2] showed that for the special submodular function involved in the MAX-COVER problem, this approximation result generalizes to a budgeted setting under linear nonnegative cost-functions. In this note, we extend this result to general submodular functions.

Motivated by the problem of maximizing entropy in discrete graphical models, where the conditional entropies in (1.1) can in general not be evaluated both exactly and efficiently [4], we generalize our result to account for absolute errors. Our derivations in the following sections closely follows the analysis presented in [2].

2 Budgeted maximization of submodular functions

Let $\mathcal{V}$ be a finite set, and $F : \mathcal{V} \to \mathbb{R}$ be a set function with $F(\emptyset) = 0$. F is called *submodular* if $F(\mathcal{A} \cup X) - F(\mathcal{A}) \geq F(\mathcal{A}' \cup X) - F(\mathcal{A}')$ for all $\mathcal{A} \subset \mathcal{V}$ and $X \in \mathcal{V} \setminus \mathcal{A}$. F is called non-decreasing if $F(\mathcal{A} \cup X) - F(\mathcal{A}) \geq 0$ for all $\mathcal{A} \subset \mathcal{V}$ and $X \in \mathcal{V} \setminus \mathcal{A}$. The quantities $F'(\mathcal{A}; X) := F(\mathcal{A} \cup X) - F(\mathcal{A})$ are called *marginal increases* of F with respect to $\mathcal{A}$ and X. Furthermore define a cost function $c : \mathcal{V} \to \mathbb{R}^+$, associating a positive cost $c(X)$ with each element $X \in \mathcal{V}$. We extend c linearly to sets: For $\mathcal{A} \subset \mathcal{V}$ define

$$c(\mathcal{A}) = \sum_{X \in \mathcal{A}} c(X).$$

For a budget $B > 0$, the budgeted maximization problem is to maximize

$$OPT = \operatorname*{argmax}_{\mathcal{A} \subset \mathcal{V} : c(\mathcal{A}) \leq B} F(\mathcal{A}) \tag{2.1}$$

Note that the exclusion of zero cost does not incur loss of generality because since the submodular functions are nondecreasing. We refer to $c(\mathcal{A}) = |\mathcal{A}|$ as the unit-cost case.

3 A constant factor approximation

In analogy to the unit-cost case discussed in [5], we analyze the greedy algorithm, where the greedy rule adds to set $\mathcal{A}$ the element X^* such that

$$X^* = \max_{X \in \mathcal{W} \setminus \mathcal{G}_{i-1}} \frac{\hat{F}'(\mathcal{G}_{i-1}; X)}{c(X_i)}.$$

Khuller et.al. [2] prove that the simple greedy algorithm with this greedy selection rule has unbounded approximation ratio. They suggest a small modification, considering the best single element solution as alternative to the output of the naive greedy heuristic, which, as they prove, guarantees a constant factor approximation for the budgeted MAX-k-COVER problem. Their algorithm is stated here as Algorithm 1, and we extend their analysis to the case of general submodular functions. Motivated by the entropy maximization problem where we cannot efficiently evaluate the marginal increases $F'(\mathcal{A}; X)$ exactly [4], we only assume that we can evaluate $\hat{F}'(\mathcal{A}; X)$ such that $\left|\hat{F}'(\mathcal{A}; X) - F'(\mathcal{A}; X)\right| \leq \varepsilon$ for some $\varepsilon > 0$.

Input: $d > 0, B > 0, \mathcal{W} \subseteq \mathcal{V}$
Output: Selection $\mathcal{A} \subseteq \mathcal{W}$
begin
  $\mathcal{A}_1 := \text{argmax}\{F(\{X\}) : X \in \mathcal{W}, c(X) \leq B\}$;
  $\mathcal{A}_2 := \emptyset$;
  $\mathcal{W}' := \mathcal{W}$;
  while $\mathcal{W}' \neq \emptyset$ **do**
    foreach $X \in \mathcal{W}$ **do** $\Delta_X := \hat{F}'(\mathcal{A}_2; X)$;
    $X^* := \text{argmax}\{\Delta_X / c(X) : X \in \mathcal{W}'\}$;
1   **if** $c(\mathcal{A}_2) + c(X^*) \leq B$ **then** $\mathcal{A}_2 := \mathcal{A}_2 \cup X^*$;
    $\mathcal{W}' := \mathcal{W}' \setminus X^*$;
  end
  return $\underset{\mathcal{A} \in \{\mathcal{A}_1, \mathcal{A}_2\}}{\text{argmax}} F(\mathcal{A})$
end

Algorithm 1: Approximation algorithm for budgeted case.

Let us consider the computation of the set $\mathcal{A}_2$ in Algorithm 1. Renumber $\mathcal{V} = \{X_1, \ldots, X_n\}$ and define $\mathcal{G}_0 = \emptyset$ and $\mathcal{G}_i = \{X_1, \ldots, X_i\}$ such that

$$\frac{F(\mathcal{G}_i) - F(\mathcal{G}_{i-1}) + \varepsilon}{c(X_i)} \geq \max_Y \frac{F(\mathcal{G}_{i-1} \cup Y) - F(\mathcal{G}) - \varepsilon}{c(Y)}.$$

The sequence $(G_j)_j$ corresponds to the sequence of assignments to $\mathcal{A}_2$, and is motivated by the simple greedy rule, adding, for a prior selection $\mathcal{G}_{i-1}$, the element X_i such that

$$X_i = \max_{X \in \mathcal{W} \setminus \mathcal{G}_{i-1}} \frac{\hat{F}'(\mathcal{G}_{i-1}; X)}{c(X_i)}.$$

Let $l = \max\{i : c(\mathcal{G}_i) \leq B\}$ be the index corresponding to the iteration, where $\mathcal{A}_2$ is last augmented in Line 1, and let $c_m in = \min_X c(X)$. Hence $\mathcal{A}_2 = \mathcal{G}_l$. Let $L = c(OPT)$, and $w = |OPT|$. We first prove the following Theorem:

Theorem 1 (adapted from [2]). *Algorithm 1 achieves an*

$$\frac{1}{2}(1 - 1/e)F(OPT) - \frac{1}{2}\left(\frac{L}{c_{min}} + w\right)\varepsilon$$

approximation for (2.1)*, using* $\mathcal{O}(|\mathcal{W}|^2)$ *evaluations of* $\hat{F}'$.

To prove Theorem 1, we need two lemmas:

Lemma 2 (generalized from [2]). *For* $i = 1, \ldots, l+1$, *it holds that*

$$F(\mathcal{G}_i) - F(\mathcal{G}_{i-1}) \geq \frac{c(X_i)}{L}(F(OPT) - F(\mathcal{G}_{i-1})) - \varepsilon\left(1 + \frac{wc(X_i)}{L}\right)$$

Proof. Using monotonicity of F, we have

$$F(OPT) - F(\mathcal{G}_{i-1}) \leq F(OPT \cup \mathcal{G}_{i-1}) - F(\mathcal{G}_{i-1}) = F(OPT \setminus \mathcal{G}_{i-1} \cup \mathcal{G}_{i-1}) - F(\mathcal{G}_{i-1})$$

Assume $OPT \setminus \mathcal{G}_{i-1} = \{Y_1, \ldots, Y_m\}$, and let for $j = 1, \ldots, m$

$$Z_j = F(\mathcal{G}_{i-1} \cup \{Y_1, \ldots, Y_j\}) - F(\mathcal{G}_{i-1} \cup \{Y_1, \ldots, Y_{j-1}\}).$$

Then $F(OPT) - F(\mathcal{G}_{i-1}) \leq \sum_{j=1}^{m} Z_j$.

Now notice that

$$\frac{Z_j - \varepsilon}{c(Y_j)} \leq \frac{F(\mathcal{G}_{i-1} \cup Y_j) - F(\mathcal{G}_{i-1}) - \varepsilon}{c(Y_j)} \leq \frac{F(\mathcal{G}_i) - F(\mathcal{G}_{i-1}) + \varepsilon}{c(X_i)}$$

using submodularity in the first and the greedy rule in the second inequality. Since $\sum_{j=1}^{m} c(Y_j) \leq L$ it holds that

$$F(OPT) - F(\mathcal{G}_{i-1}) = \sum_{j=1}^{m} Z_j \leq L\frac{F(\mathcal{G}_i) - F(\mathcal{G}_{i-1}) + \varepsilon}{c(X_i)} + m\varepsilon$$

□

Lemma 3 (adapted from [2]). *For* $i = 1, \ldots, l+1$ *it holds that*

$$F(\mathcal{G}_i) \geq \left[1 - \prod_{k=1}^{i}\left(1 - \frac{c(X_k)}{L}\right)\right] F(OPT) - \left(\frac{L}{c(X_i)} + w\right)\varepsilon.$$

Proof. Let $i = 1$ for sake of induction. We need to prove that $F(\mathcal{G}_1) \geq \frac{c(X_1)}{L}F(OPT) - \left(\frac{L}{c(X_i)} + w\right)\varepsilon$. This follows from Lemma 2 and since

$$\frac{L}{c(X_i)} + w \geq 1 + \frac{wc(X_i)}{L}.$$

Now let $i > 1$. We have

$$\begin{aligned}
F(\mathcal{G}_i) &= F(\mathcal{G}_{i-1}) + [F(\mathcal{G}_i) - F(\mathcal{G}_{i-1})] \\
&\geq F(\mathcal{G}_{i-1}) + \frac{c(X_i)}{L}[F(OPT) - F(\mathcal{G}_{i-1})] - \varepsilon\left(1 + \frac{wc(X_i)}{L}\right) \\
&= \left(1 - \frac{c(X_i)}{L}\right)F(\mathcal{G}_{i-1}) + \frac{c(X_i)}{L}F(OPT) - \varepsilon\left(1 + \frac{wc(X_i)}{L}\right) \\
&\geq \left(1 - \frac{c(X_i)}{L}\right)\left[\left(1 - \prod_{k=1}^{i-1}\left(1 - \frac{c(X_k)}{L}\right)\right)F(OPT) - \left(\frac{L}{c(X_i)} + w\right)\varepsilon\right] + \frac{c(X_i)}{L}F(OPT) - \varepsilon\left(1 + \frac{wc(X_i)}{L}\right) \\
&= \left(1 - \prod_{k=1}^{i}\left(1 - \frac{c(X_k)}{L}\right)\right)F(OPT) - \varepsilon\left(1 + \frac{wc(X_i)}{L}\right) - \left(\frac{L}{c(X_i)} + w\right)\varepsilon\left(1 - \frac{c(X_i)}{L}\right) \\
&= \left(1 - \prod_{k=1}^{i}\left(1 - \frac{c(X_k)}{L}\right)\right)F(OPT) - \left(\frac{L}{c(X_i)} + w\right)\varepsilon
\end{aligned}$$

using Lemma 2 in the first and the induction hypothesis in the second inequality. □

From now on let $\beta = \frac{L}{c_{min}} + w$.

Proof of Theorem 1. Observe that for $a_1, \ldots, a_n \in \mathbb{R}^+$ such that $\sum a_i = A$, the function $(1 - \prod_{i=1}^{n}(1 - \frac{a_i}{A}))$ achieves its minimum at $a_1 = \cdots = a_n = \frac{A}{n}$.

We have

$$\begin{aligned} F(\mathcal{G}_{l+1}) &\geq \left[1 - \prod_{k=1}^{l+1}\left(1 - \frac{c(X_k)}{L}\right)\right] F(OPT) - \beta\varepsilon \\ &\geq \left[1 - \prod_{k=1}^{i}\left(1 - \frac{c(X_k)}{c(\mathcal{G}_{l+1})}\right)\right] F(OPT) - \beta\varepsilon \\ &\geq \left[1 - \left(1 - \frac{1}{l+1}\right)^{l+1}\right] F(OPT) - \beta\varepsilon \\ &\geq \left(1 - \frac{1}{e}\right) F(OPT) - \beta\varepsilon \end{aligned}$$

where the first inequality follows from Lemma 3 and the second inequality follows from the fact that $c(\mathcal{G}_{l+1}) > L$, since it violates the budget.

Furthermore note, that the violating increase $F(\mathcal{G}_{l+1}) - F(\mathcal{G}_l)$ is bounded by $F(X^*)$ for $X^* = \operatorname{argmax}_{X \in \mathcal{W}} F(X)$, i.e. the second candidate solution considered by the modified greedy algorithm. Hence

$$F(\mathcal{G}_l) + F(X^*) \geq F(\mathcal{G}_{l+1}) \geq (1 - 1/e)F(OPT) - \beta\varepsilon$$

and at least one of the values $F(X^*)$ or $F(\mathcal{G}_l)$ must be greater than or equal to $\frac{1}{2}((1 - 1/e)F(OPT) - \beta\varepsilon)$. □

4 An improved approximation guarantee

To achieve the same performance guarantee of $(1 - 1/e)$ which can be achieved for the unit-cost in the case of general submodular functions [5], Khuller et.al.[2] propose a partial enumeration heuristic which enumerates all subsets of up to d elements for some constant $d > 0$, and complements these subsets using the modified greedy algorithm Algorithm 1. They prove that this algorithm guarantees a $(1 - 1/e)$ approximation for the budgeted MAX-k-COVER problem. The algorithm is stated below for general nondecreasing submodular functions:

> **Input**: $d > 0, B > 0, \mathcal{W} \subseteq \mathcal{V}$
> **Output**: Selection $\mathcal{A} \subseteq \mathcal{W}$
> **begin**
> $\mathcal{A}_1 := \operatorname{argmax}\{F(\mathcal{A}) : \mathcal{A} \subseteq \mathcal{W}, |\mathcal{W}| < d, c(\mathcal{W}) \leq B\}$ $\mathcal{A}_2 := \emptyset$;
>   **foreach** $\mathcal{G} \subseteq \mathcal{W}, |\mathcal{G}| = d, c(\mathcal{G}) \leq B$ **do**
> $\mathcal{W}' := \mathcal{W} \setminus \mathcal{G}$;
>   **while** $\mathcal{W}' \neq \emptyset$ **do**
>   **foreach** $X \in \mathcal{W}'$ **do** $\Delta_X := \hat{F}'(\mathcal{G}; X)$;
> $X^* := \operatorname{argmax}\{\Delta_X / c(X) : X \in \mathcal{W}'\}$;
>   **if** $c(\mathcal{G}) + c(X^*) \leq B$ **then** $\mathcal{G} := \mathcal{G} \cup X^*$;
> $\mathcal{W}' := \mathcal{W}' \setminus X^*$;
>   **end**
>   **if** $F(\mathcal{G}) > F(\mathcal{A}_2)$ **then** $\mathcal{A}_2 := \mathcal{G}$
>   **end**
>   **return** $\operatorname{argmax}_{\mathcal{A} \in \{\mathcal{A}_1, \mathcal{A}_2\}} F(\mathcal{A})$
> **end**

Algorithm 2: Approximation algorithm for budgeted case.

Theorem 4 (adapted from [2]). *Algorithm 2 achieves an approximation guarantee of*

$$(1 - 1/e)F(OPT) - \left(\frac{L}{c_{min}} + w\right)\varepsilon$$

for (2.1) *if sets at least up to cardinality* $d = 3$ *are enumerated, using* $\mathcal{O}(|\mathcal{W}|^{d+2})$ *evaluations of* $\hat{F}'$.

Since we do not know L and w in general, the following corollary provides an explicitly computable bound on the absolute error:

Corollary 5. *Algorithm 2 achieves an approximation guarantee of*

$$(1-1/e)F(OPT) - \frac{2B}{c_{min}}\varepsilon$$

for (2.1) *if sets at least up to cardinality* $d = 3$ *are enumerated.*

Proof of Theorem 4. Assume that $|OPT| > k$, otherwise the algorithm finds the exact optimum. Renumber $OPT = \{Y_1, \ldots, Y_m\}$ such that

$$Y_{i+1} = \operatorname*{argmax}_{Y \in OPT} F(\{Y_1, \ldots, Y_i, Y\}) - F(\{Y_1, \ldots, Y_i\}),$$

and let $\mathcal{B} = \{Y_1, \ldots, Y_k\}$. Consider the iteration where the algorithm considers $\mathcal{B}$. Define the function

$$F'(\mathcal{A}) = F(\mathcal{A} \cup \mathcal{B}) - F(\mathcal{B}).$$

F' is a nondecreasing submodular set function with $F'(\emptyset) = 0$, hence we can apply the modified greedy algorithm to it. Let $\mathcal{A} = \{V_1, \ldots, V_l\}$ be the result of the algorithm, where V_i are chosen in sequence, let V_{l+1} be the first element from $OPT \setminus \mathcal{B}$ which could not be added due to budget constraints, and let $\mathcal{G} = \mathcal{A} \cup \mathcal{B}$. Per definition, $F(\mathcal{G}) = F'(\mathcal{A}) + F(\mathcal{B})$. Let $\Delta = F'(\mathcal{A} \cup V_{l+1}) - F'(\mathcal{A})$. Using Lemma 3, we find that

$$F'(\mathcal{A}) + \Delta \geq (1-1/e)F'(OPT \setminus \mathcal{B}) - \beta\varepsilon.$$

Furthermore observe, since the elements in OPT are ordered, that $F(\{Y_1, \ldots, Y_i\}) - F(\{Y_1, \ldots, Y_{i-1}\}) \geq \Delta$ for $1 \leq i \leq k$. Hence $F(\mathcal{B}) \geq k\Delta$. Now we get

$$\begin{aligned} F(\mathcal{G}) &= F(\mathcal{B}) + F'(\mathcal{A}) \\ &\geq F(\mathcal{B}) + (1-1/e)F'(OPT \setminus \mathcal{B}) - \Delta - \beta\varepsilon \\ &\geq F(\mathcal{B}) + (1-1/e)F'(OPT \setminus \mathcal{B}) - \frac{F(\mathcal{B})}{k} - \beta\varepsilon \\ &\geq (1-1/k)F(\mathcal{B}) + (1-1/e)F'(OPT \setminus \mathcal{B}) - \beta\varepsilon \end{aligned}$$

But by definition, $F(\mathcal{B}) + F'(OPT \setminus \mathcal{B}) = F(OPT)$, and hence for $k \geq 3$

$$F(\mathcal{G}) \geq (1-1/e)F(OPT) - \beta\varepsilon.$$

□

5 Conclusions

We presented an efficient approximation algorithm for the budgeted maximization of nondecreasing submodular set functions. We proved bounds on the absolute error which are incurred if the marginal increases can only be computed with an absolute error. We believe that our results are useful for the wide class of combinatorial optimization problems concerned with maximizing submodular functions.

References

[1] Thomas M. Cover and Joy A. Thomas, *Elements of information theory*, Wiley Interscience, 1991.

[2] Samir Khuller, Anna Moss, and Joseph Naor, *The budgeted maximum coverage problem*, Inf Proc Let **70** (1999), no. 1, 39–45.

[3] Chunwa Ko, Jon Lee, and Maurice Queyranne, *An exact algorithm for maximum entropy sampling*, OR **43** (1995), no. 4, 684–691.

[4] A. Krause and C. Guestrin, *Optimal nonmyopic value of information in graphical models - efficient algorithms and theoretical limits*, Proc. of IJCAI, 2005.

[5] G. Nemhauser, L. Wolsey, and M. Fisher, *An analysis of the approximations for maximizing submodular set functions*, Mathematical Programming **14** (1978), 265–294.