Maximizing non-monotone submodular functions

Jean Pouget-Abadie

March 30, 2015

We recap the paper [1], which covers the maximisation of general submodular and non-negative functions without constraints. The main contribution of this paper is to show that non-monotone submodular functions can also be optimized (albeit not with the standard greedy algorithm) as well as providing interesting lower bounds. The paper also treats *symmetric* submodular functions as a special case. We recall the definition below:

Definition 1. *Let X be the universe of all elements. For any set $A \subseteq X$, we define $\bar{A}$ to be the complement of A in X. A* symmetric *function takes the same value on A and $\bar{A}$ for all $A \subseteq X$.*

1 Upper bounds

1.1 Random Set Algorithm

Let X be the universe of all elements. For any set $A \subseteq X$, we define $A(p) \equiv \{x \in A : rand(x) > p\}$ where $rand(x)$ is an independent biased coin of parameter p.

return uniformly random subset of X: $X(1/2)$

Algorithm 1: Random Set Algorithm

To prove guarantees on this algorithm, we are going to prove two lemmas.

Lemma 1. *Let f be a submodular function and $A \subseteq X$. Then:*

$$\mathbb{E}[f(A(p))] \geq (1-p)f(\emptyset) + pf(A)$$

Proof. This can be easily interpreted as the concavity of the multilinear extension of a submodular function along positive directions. It can also be proved by a simple induction on the size of A by taking the expectation of:

$$g(A \cup \{x\}(p)) - g(A(p)) \geq g(A \cup \{x\}(p) \cup A) - g(A)$$

Note that for this property to hold for non-monotone submodular functions, we have to couple $A \cup \{x\}(p)$ and $A(p)$. □

The above lemma can be generalized to union of sets by simple induction:

Lemma 2. *Let f be a submodular function and $A, B \subseteq X$ not necessarily disjoint, then:*

$$\mathbb{E}[f(A(p) \cup B(p))] \geq (1-p)(1-q)f(\emptyset) + p(1-q)f(A) + (1-q)f(B) + pqf(A \cup B)$$

Proof. Condition on $A(p) = R$, apply the previous lemma to the submodular function $g_R : B \mapsto g(R \cup B)$. Take the expectation of the previous expectations to remove the conditioning on $A(p) = R$ by iterated expectations. Apply the lemma again. □

Theorem 1. *The random set is a 1/2-approximation algorithm for unconstrainted maximisation of non-negative symmetric functions and a 1/4-approximation for general non-negative submodular functions.*

Proof. Let C be the optimal set. We can write $C = C(1/2) \cup \bar{C}(1/2)$. We apply the second lemma:

$$\mathbb{E}[f(R)] = 1/4 f(\emptyset) + 1/4 f(S) + 1/4 f(\bar{S}) + 1/4 f(X)$$

Since f is non-negative, we get the results. Observe that for symmetric functions, $f(S) = f(\bar{S})$ □

Note that the 1/2 approximation cannot be improved (though it can be made deterministic) for symmetric non-negative submodular functions.

1.2 Deterministic Local Search

Let $\epsilon > 0$. We define $\alpha \equiv \epsilon/n^2$.

$S = \{v\}$ where $v \in \arg\max_v f(\{v\})$;
while $\exists a \in \bar{S} : f(S \cup \{x\}) > (1 + \epsilon/n^2) f(S)$ **do**
  $S \leftarrow S \cup \{x\}$;
  while $\exists a \in S : f(S \backslash \{x\}) > (1 + \epsilon/n^2) f(S)$ **do**
    $S \leftarrow S \backslash \{x\}$;
  end
end
return $\arg\max(f(S), f(\bar{S}))$

Algorithm 2: Randomized Local Search

We can improve the approximation ratio for general functions to 1/3 and make the 1/2 approximation ratio deterministic. Recall the definition of $(1 + \alpha)$-approximate local optimum:

Definition 2. *S is a $(1 + \alpha)-$approximate local optimum if $\forall x \in X : (1 + \alpha) f(S) \geq f(S \cup \{x\})$ and $\forall x \in X : (1 + \alpha) f(S) \geq f(S \backslash \{x\})$*

We have the following interesting lemma for locally optimal sets:

Lemma 3. *If S is a $(1+\alpha)$-locally optimal set, then for all its subsets and supersets I: $f(I) \leq (1+n\alpha) f(S)$*

Proof. Consider a subset I of S and the collection of sets T_i such that $T_i \subset T_{i+1}$ and $T_0 = I$ and $T_l = S$. By submodularity and local approximality of S:

$$f(T_i) - f(T_{i-1}) \geq f(S) - f(S \backslash \{a_i\}) \geq -\alpha f(S)$$

By summing up these inequalities, we get $f(S) - f(I) \geq -l\alpha f(S) \geq -n\alpha f(S)$. The method for all supersets of S is identical. □

Theorem 2. *The local-search algorithm is a 1/3-approximation algorithm for maximising nonnegative submodular functions and a 1/2-approximation algorithm for maximising nonnegative symmetric submodular functions.*

Proof. Let C be the optimal solution and S the solution returned by the algorithm. Note that $S \cup C$ and $S \cap C$ are respectively subsets and supersets of S. Since S is $(1 + \alpha)$-locally optimal set, we have:

$$\begin{aligned} 2(1 + n\alpha) f(S) + f(\bar{S}) &\geq f(S \cap C) + f(S \cup C) + f(\bar{S}) \\ &\geq f(S \cap C) + f((S \cup C) \cap (\bar{S})) + f((S \cup C) \cup (\bar{S})) \\ &\geq f(S \cap C) + f(C \backslash S) \\ &\geq f(C) \end{aligned}$$

where we use submodularity and nonnegativity twice. We get the desired conclusion (recall that we return $\arg\max(f(S), f(\bar{S}))$). We can bound the running time of the algorithm by noticing that the value of the algorithm increases by a factor of at least $(1 + \epsilon/n^2)$. Note that $f(C) \geq nf(\{v\})$ and therefore if k is the number of iterations of the algorithm, then we have $(1 + \epsilon/n^2)^k < n \implies k < \frac{1}{\epsilon} n^2 \log n$ □

1.3 Smooth Local Search

Fix $\delta, \delta' \in [-1, 1]$. Let $p \equiv \frac{1+\delta}{2}$ and $q \equiv \frac{1-\delta}{2}$. In the algorithm, we are going to define an estimate of the following derivative, called the *weight* of an element:

$$w_{A,\delta} \equiv \mathbb{E}[f(A(p) \cup B(q) \cup \{x\})] - \mathbb{E}[f(A(p) \cup B(q) \backslash \{x\})]$$

In the algorithm, we suppose we can obtain an estimate of $w_{A,\delta}$ in polynomial time by random and uniform sampling. Furthermore, we will compare this estimate to OPT, which we will also estimate within a constant factor using local-search (for example).

$A \equiv \emptyset$;
while $\exists a \in \bar{S} : w_{A,\delta} > 2OPT/n^2$ **do**
  $S \leftarrow S \cup \{x\}$;
  while $\exists a \in S : w_{A,\delta} < -2OPT/n^2$ **do**
    $S \leftarrow S \backslash \{x\}$;
  end
end
return A random set from $A(\frac{1+\delta'}{2}) \cup B(\frac{1-\delta'}{2})$

Algorithm 3: Smooth Local Search

Theorem 3. *Smooth Local Search runs in polynomial time. If we run SLS for two choices of parameters and return the maximal set, then the better of the two solutions has expected value at least $2/5OPT$.*

The proof is a little bit hard to synthesize and does not give much intuition. See paper for full details.

2 Lower Bounds

2.1 NP-hardness result

Theorem 4. *There is no polynomial-time algorithm $(5/6+\epsilon)$-approximation algorithm to maximize a non-negative symmetric submodular function in succinct representation, unless $P = NP$.*

Proof. Consider an instance of MAx E4-Lin2, i.e a system of m equations on 4 boolean variables each, possibly shared amongst the equations, such that each equation is True if exactly half the variables are True. For each variable x_i, we define two values T_i and F_i (True or False) and we consider the universe of all possible truth values of each variable $\cup_{x_i}\{T_i, F_i\}$

Let e be one equation from the m considered on variables $\{x_i, x_j, x_k, x_l\}$. For any set of this universe S, we define $S' \equiv S \cap \{T_i, F_i, T_j, F_j, T_k, F_k, T_l, K_l\}$. We say that an assignment S' is valid if each variable has exactly one truth value T_i OR F_i. Consider the possible cases:

1. S' is not a valid assignment
   (a) $|S'| < 4$, $g_e(S) = |S'|$.
   (b) $|S'| > 4$, $g_e(S) = 8 - |S'|$
   (c) $|S'| = 4$, $g_e(S) = 10/3$
2. S' is a valid assignment.
   (a) Exactly half the variables are set to True. $g_e(S) = 4$
   (b) Not exactly half the variables are set to True. $g_e(S) = 8/3$

It can be verified that the function is submodular. Without loss of generality, we show that the optimal set S has exactly one truth value for each variable. Let $\tilde{S}$ be an optimal set with mutliple or no truth values for certain variables, which we call *undecided*. For each undecided variable, sample a truth value (from T_i, F_i) uniformly at random. Consider S the random set obtained. For each equation e, S will induce a valid assignment which satisfies e.

Consider $f(S) \equiv \sum_e g_e(S)$. f is submodular. For all equations with a valid assignment, the value will not change. For all equations without a valid assignment, the expected value of $g_e(S)$ will be $1/2(8/3+4) = 10/3$, which is at least the value fo any invalid assignment. Therefore, $\mathbb{E}[f(S)] \geq 10/3$ and by the probabilistic method, there exists an optimal set S which assigns a valid assignment to every equation. The optimal value of f is therefore: $f(S) = 8/3(m-r) + 4r = 8/3m + 4/3r$ where r is the maximum number of satisfiable equations.

Since it is NP-hard to distinguish whether $r > (1-\epsilon)m$ or $r < (1/2-\epsilon)m$, it is also NP-hard to distinguish whether $OPT \geq (4-\epsilon)m$ and $OPT \leq (10/3+\epsilon)m$

□

This result can be improved in the case of general submodular functions:

Theorem 5. *There is no polynomial time $(3/4+\epsilon)$-approximation algorithm to maximize a nonnegative submodular function in succinct representation unless $P = NP$.*

The proof relies on non-trivial properties of Hastad's PCP classifier, which we will not cover in these notes, but which will be most likely covered in later sections of this reading group. [2].

2.2 Value Query Complexity Results

We begin with the following proposition, which is not very interesting by itself, but the conclusion reached in the proof is interesting as explained below.

Proposition 1. *For any $\delta > 0$, there exists $\epsilon > 0$ such that for any (random) sequence of queries $\mathcal{Q} \subseteq 2^X, |\mathcal{Q}| \leq 2^{\epsilon n}$, there exists a nonnegative submodular function such that with high probability for all queries:*

$$f(Q) \leq (1/4+\delta)OPT$$

Proof. (sketch) We use the probabilistic method and consider cuts on the complete bipartite directed graph on (C,D). We fix the queries and choose a uniformly random $C \subseteq X$

$$f_C(S) = |S \cap C| \cdot |\bar{S} \cap D|$$

It is easy to see that n^2 is the optimum obtained for $S \equiv C$. However, most queries will intersect both sets and the value will be concentrated around 1/4OPT. We use Chernoff bounds followed by a union bound. For any set A of size a:

$$\mathbb{P}(|A \cap C| > 1/2(1+\delta)|A|) = \mathbb{P}(|A \cap C| < 1/2(1-\delta)|A|) < e^{-\delta^2 a/2}$$

Make the following observations:

- With very high probability, the value of C will be in $[(1/2-\delta)n, (1/2+\delta)n]$. If this is true:
- OPT $\geq (\frac{1}{4} - \delta^2)n^2$
- Q has to be of size $\in [\frac{1}{16}n, n - \frac{1}{16}n]$ for it to be "high enough".
- Both $|A \cap C| > 1/2(1+\delta)|A|$ and $|A \cap C| < 1/2(1-\delta)|A|$ have to occur for the query to be "high enough".
- This happens with very small probability. Conclusion by union bound on all queries.

□

As a result, we never observe a very high value. We could still learn enough from the queries to maximize the function correctly. The main conclusion from this example is that when considering cut functions of a complete directed graph, most sets are very "balanced" and we will never observe a high value. This provides the intuition for the following proof.

Theorem 6. *For any $\epsilon > 0$ there are instances of nonnegative symmetric submodular maximization such that there is no (adaptive, possibly randomized) algorithm using less than $e^{\epsilon^2 n/16}$ queries that always finds a solution of expected value at least $(1/2 + \epsilon)OPT$*

Proof. When unambiguous, let $k \equiv |S \cap C|$ and $l \equiv |S \cap D|$. Define f as follows:

1. $f(k, l) \equiv (k + l)(n - k - l) = |S|(n - |S|)$ when $|k - l| < \epsilon n$ (the set is said to be *balanced*).
2. $f(k, l) \equiv k(n - 2l) + (n - 2k)l - \mathcal{O}(\epsilon n^2)$ (the set is *unbalanced*). The precise value of $\mathcal{O}(\epsilon n^2)$ will be specified later, and will allow to make the function smooth and submodular.

What we note is that a balanced set provides no information about the sets C and D. Using the intuition developped during the previous proposition, we see that with high probability we will never get an unbalanced set. As such, we cannot distinguish our function f from the submodular function $S \mapsto |S|(n - |S|)$, which has an optimum of $\frac{1}{4}n^2$ whilst our function has an optimum of $\frac{1}{2}n^2$.

All that remains to show is that we can choose $\mathcal{O}(\epsilon n^2)$ such that our function f is submodular. See paper for full details. □

References

[1] Uriel Feige, Vahab S Mirrokni, and Jan Vondrak. Maximizing non-monotone submodular functions. *SIAM Journal on Computing*, 40(4):1133–1153, 2011.

[2] Johan Håstad. Some optimal inapproximability results. *Journal of the ACM (JACM)*, 48(4):798–859, 2001.