Maximizing Submodular Functions

David P. Williamson
Cornell University

August 17–21, 2015
ADFOCS

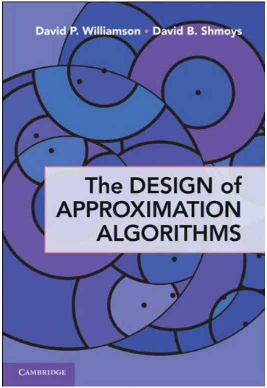
1

About me

Email: dpw@cs.cornell.edu
Web: www.davidpwilliamson.net/work

Feel free to send email to ask question, provide feedback, etc.

Book



Download `www.designofapproxalgs.com`

What is a Submodular Function?

Given a finite *ground set* of elements $E = \{e_1, \ldots, e_n\}$, consider any function $f : 2^E \rightarrow \Re^{\geq 0}$ that maps subsets of the ground set E to nonnegative reals. $f(\emptyset) = 0.$

Definition

The function f is *submodular* if for all $S, T \subseteq E$, $S \subseteq T$, and $\ell \in E - T$,

$$f(T \cup \{\ell\}) - f(T) \leq f(S \cup \{\ell\}) - f(S).$$

In other words, the function has decreasing marginal gains as the input set increases.

An Equivalent Definition

Equivalently, the function f is submodular if for all $A, B \subseteq E$,

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

Proof of Equivalence

$$f(A)+f(B) \geq f(A\cup B)+f(A\cap B) \Rightarrow f(T\cup\{\ell\})-f(T) \leq f(S\cup\{\ell\})-f(S)$$

$A = S \cup \{\ell\}$ $B = T$

$A \cup B = T \cup \{\ell\}$ $A \cap B = S$

Proof of Equivalence

$$f(T\cup\{\ell\})-f(T) \leq f(S\cup\{\ell\})-f(S) \Rightarrow f(A)+f(B) \geq f(A\cup B)+f(A\cap B)$$

If $B \subseteq A$, easy; let $B-A=\{b_1, b_2, \ldots, b_k\}$

$$f(A\cup B)-f(A)=\sum_{i=1}^{k}\left[f(A\cup\{b_1,\ldots b_i\})-f(A\cup\{b_1,\ldots,b_{i-1}\})\right]$$

$$\leq \sum_{i=1}^{k}\left[f((A\cap B)\cup\{b_1,\ldots,b_i\})-f((A\cap B)\cup\{b_1,\ldots,b_{i-1}\})\right]$$

$$= f(B)-f(A\cap B).$$

Some Terminology

Definition

f is *monotone* if for all $S, T \subseteq V$, $S \subseteq T$,

$$f(S) \leq f(T).$$

Otherwise, f is *nonmonotone*.

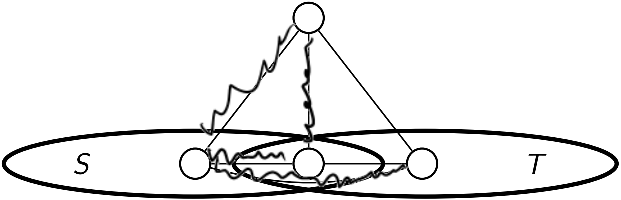
Some Examples

Given a graph $G = (V, E)$ with weights $w_{ij} \geq 0$ on the edges $(i, j) \in E$, let $w(S)$ be the total weight of the edges with one endpoint in $S \subseteq V$; that is, the weight of the *cut* induced by S.

Claim that w is submodular:

$w(S) = \sum_{(i,j) \in \delta(S)} w_{ij}$

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



For $S \subseteq T$, $\ell \notin T$

$$f(T \cup \{\ell\}) - f(T) \le f(S \cup \{\ell\}) - f(S)$$

Pf

$$f(S \cup \{\ell\}) - f(S) = \sum_{j \in P} \left(\max_{i \in S \cup \{\ell\}} v_{ij} - \max_{i \in S} v_{ij} \right)$$

$$= \sum_{j \in P} \max\left(0, v_{\ell j} - \max_{i \in S} v_{ij}\right)$$

$$f(T \cup \{\ell\}) - f(T) = \sum_{j \in P} \max\left(0, v_{\ell j} - \max_{i \in T} v_{ij}\right)$$

$$\max_{i \in T} v_{ij} \ge \max_{i \in S} v_{ij} \qquad \forall j \in P$$

$$\Rightarrow \max\left(v_{\ell j} - \max_{i \in T} v_{ij}, 0\right) \le \max\left(v_{\ell j} - \max_{i \in S} v_{ij}, 0\right)$$

Some Examples

Cornuejols, Fischer, and Nemhauser (1977) give an example of opening bank accounts to maximize *float*. Let B be a set of banks, P a set of payees, k the number of accounts to open, v_{ij} the value of float for paying $j \in P$ from $i \in B$.

Define $f(S) = \sum_{j \in P} \max_{i \in S} v_{ij}$.

Then goal is to find $S \subseteq B$, $|S| \leq k$, to maximize $f(S)$.

Lemma

f is submodular.

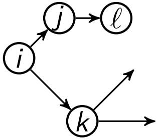
Some Examples

Kempe, Kleinberg, and Tardos (2003) give an example of selecting influential nodes in a social network. Input is directed graph $G = (V, A)$, probabilities p_{ij} for each arc $(i, j) \in A$.

Some Examples

Kempe, Kleinberg, and Tardos (2003) give an example of selecting influential nodes in a social network. Input is directed graph $G = (V, A)$, probabilities p_{ij} for each arc $(i, j) \in A$.

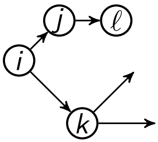
If node i is *activated*, then for all $j \in V$ such that $(i, j) \in A$, j becomes activated with probability p_{ij}.



Some Examples

Kempe, Kleinberg, and Tardos (2003) give an example of selecting influential nodes in a social network. Input is directed graph $G = (V, A)$, probabilities p_{ij} for each arc $(i, j) \in A$.

If node i is *activated*, then for all $j \in V$ such that $(i, j) \in A$, j becomes activated with probability p_{ij}.



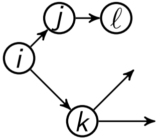
Let $f(S)$ be expected total number of vertices activated if we initially activate the vertices in S.

Goal is to maximize $f(S)$ subject to $|S| \leq k$.

Some Examples

Kempe, Kleinberg, and Tardos (2003) give an example of selecting influential nodes in a social network. Input is directed graph $G = (V, A)$, probabilities p_{ij} for each arc $(i, j) \in A$.

If node i is *activated*, then for all $j \in V$ such that $(i, j) \in A$, j becomes activated with probability p_{ij}.



Let $f(S)$ be expected total number of vertices activated if we initially activate the vertices in S.

Goal is to maximize $f(S)$ subject to $|S| \leq k$.

Claim

f is submodular.

An Issue

How do we give the function f as input?

An Issue

How do we give the function f as input?

If we list $f(S)$ for all $S \subseteq E$, then maximizing or minimizing f is easy in linear time.

An Issue

How do we give the function f as input?

If we list $f(S)$ for all $S \subseteq E$, then maximizing or minimizing f is easy in linear time.

Assume an *oracle model*: we have a subroutine that computes $f(S)$ for any $S \subseteq E$ in constant time.

NP-hardness and approximation algorithms

Some of these problems are NP-hard; for instance, finding a cut S to maximize $w(S)$ is the MAX CUT problem.

NP-hardness and approximation algorithms

Some of these problems are NP-hard; for instance, finding a cut S to maximize $w(S)$ is the MAX CUT problem.

Definition

An α-approximation algorithm for maximizing a submodular function f in the oracle model is a polynomial-time algorithm that finds a set S with $f(S) \geq \alpha \cdot OPT$, where $\alpha \leq 1$.

Greedy algorithm

What is a natural greedy algorithm for maximizing a monotone submodular function f subject to $|S| \leq k$?

$S \leftarrow \emptyset$

While $|S| < k$

$\quad i \leftarrow \underset{i \in E-S}{\operatorname{argmax}} \left[f(S \cup \{i\}) - f(S) \right]$

$\quad S \leftarrow S \cup \{i\}$

Greedy algorithm

What is a natural greedy algorithm for maximizing a monotone submodular function f subject to $|S| \leq k$?

Cornuejols, Fisher, Nemhauser (1977)

$S \leftarrow \emptyset$
while $|S| < k$ **do**
  $i \leftarrow \arg\max_{i \in E} f(S \cup \{i\}) - f(S)$
  $S \leftarrow S \cup \{i\}$
return S

The Result

Theorem (Cornuejols, Fisher, Nemhauser (1977))

The greedy algorithm is a $\left(1 - \frac{1}{e}\right)$*-approximation algorithm.*

The Result

Theorem (Cornuejols, Fisher, Nemhauser (1977))

The greedy algorithm is a $\left(1 - \frac{1}{e}\right)$-approximation algorithm.

The theorem is based on the following lemma.

Lemma

Pick any $S \subseteq E$, $|S| < k$. Let O be an optimal solution. Then

$$\max_{i \in E} [f(S \cup \{i\}) - f(S)] \geq \frac{1}{k}(f(O) - f(S)).$$

We'll assume the lemma and prove the theorem, then come back to the lemma.

Pf of thm   Let S^k = set chosen by alg at end of iter. k.

$$f(S^k) \geq \frac{1}{k} f(O) + \left(1 - \frac{1}{k}\right) f(S^{k-1})$$

$$\geq \frac{1}{k} f(O) + \left(1 - \frac{1}{k}\right)\left[\frac{1}{k} f(O) + \left(1 - \frac{1}{k}\right) f(S^{k-2})\right]$$

$$\geq \frac{f(O)}{k}\left[1 + \left(1 - \frac{1}{k}\right) + \left(1 - \frac{1}{k}\right)^2 + \dots + \left(1 - \frac{1}{k}\right)^{k-1}\right]$$

$$= \frac{f(O)}{k} \cdot \frac{1 - \left(1 - \frac{1}{k}\right)^k}{1 - \left(1 - \frac{1}{k}\right)}$$

$$= f(O) \cdot \left(1 - \left(1 - \frac{1}{k}\right)^k\right) \geq f(O) \cdot \left(1 - \frac{1}{e}\right)$$

using $\quad 1 - x \leq e^{-x}$

Pf of lemma: Let $O - S = \{i_1, \dots, i_p\}$, $p \le k$.

$$f(O) \le f(O \cup S) \qquad \text{(by monotonicity)}$$

$$= f(S) + \sum_{j=1}^{p} \left[f(S \cup \{i_1, \dots, i_j\}) - f(S \cup \{i_1, \dots, i_{j-1}\}) \right]$$

$$\le f(S) + \sum_{j=1}^{p} \left[f(S \cup \{i_j\}) - f(S) \right] \qquad \text{(by submodularity)}$$

$$\le f(S) + p \left[\max_{i \in E} f(S \cup \{i\}) - f(S) \right]$$

$$\le f(S) + k \left[\max_{i \in E} f(S \cup \{i\}) - f(S) \right]$$

Limits

Feige (1998) shows that we cannot have a $\left(1 - \frac{1}{e} + \epsilon\right)$-approximation algorithm for maximizing a monotone submodular function subject to a cardinality constraint unless $P = NP$. (An exercise, though a hard one).

Nonmonotone Submodular Functions

We now assume $E = \{1, \ldots, n\}$. Two natural greedy algorithms for maximizing a nonmonotone submodular function:

Nonmonotone Submodular Functions

We now assume $E = \{1, \ldots, n\}$. Two natural greedy algorithms for maximizing a nonmonotone submodular function:

- $X \leftarrow \emptyset$. For $i \leftarrow 1$ to n, if $f(X \cup \{i\}) > f(X)$, add i to X.

Nonmonotone Submodular Functions

We now assume $E = \{1, \ldots, n\}$. Two natural greedy algorithms for maximizing a nonmonotone submodular function:

- $X \leftarrow \emptyset$. For $i \leftarrow 1$ to n, if $f(X \cup \{i\}) > f(X)$, add i to X.
- $X \leftarrow E$. For $i \leftarrow 1$ to n, if $f(X - \{i\}) > f(X)$, remove i from X.

We'll look at an algorithm that in some sense randomizes between the two.

Notation

The algorithm will maintain a set $X_i \subseteq \{1, 2, \ldots, i\}$.

Define $\hat{X}_i \equiv X_i \cup \{i+1, \ldots, n\}$.

$\hat{X}_0 = E$

$X_0 = \emptyset$ $\quad \hat{X}_n =$ $\quad X_i \subseteq \hat{X}_i$ $\quad X_n = \hat{X}_n$.

Notation

The algorithm will maintain a set $X_i \subseteq \{1, 2, \ldots, i\}$.

Define $\hat{X}_i \equiv X_i \cup \{i+1, \ldots, n\}$.

$X_0 =$ $\hat{X}_n =$ X_i $\hat{X}_i$ X_n $\hat{X}_n$.

Each step of the algorithm will compute

$a_i \leftarrow f(X_{i-1} \cup \{i\}) - f(X_{i-1})$ value of adding i to X_{i-1}

$r_i \leftarrow f(\hat{X}_{i-1} - \{i\}) - f(\hat{X}_{i-1})$ value of removing i from $\hat{X}_{i-1}$

Lemma

$$a_i + r_i \geq 0.$$

Pf Note $X_{i-1} \subseteq \hat{X}_{i-1} - \{i\}$. By submodularity

$$\underbrace{f(\hat{X}_{i-1}) - f(\hat{X}_{i-1} - \{i\})}_{-r_i} \le \underbrace{f(X_{i-1} \cup \{i\}) - f(X_{i-1})}_{a_i}$$

The DoubleGreedy Algorithm

Buchbinder, Feldman, Naor, Schwartz (2012)

$X_0 \leftarrow \emptyset$
for $i \leftarrow 1$ **to** n **do**
  Compute a_i, r_i
  if $a_i \geq 0, r_i < 0$ **then**
    $X_i \leftarrow X_i \cup \{i\}$ // so $\hat{X}_i = \hat{X}_{i-1}$
  if $a_i < 0, r_i \geq 0$ **then**
    $X_i \leftarrow X_{i-1}$ // so $\hat{X}_i = \hat{X}_{i-1} - \{i\}$
  else

$$X_i \leftarrow \begin{cases} X_{i-1} \cup \{i\} & \text{w. prob } \frac{a_i}{a_i+r_i} \\ X_{i-1} & \text{w. prob } \frac{r_i}{a_i+r_i} \end{cases}$$

More Notation

Let OPT be an optimal set, and

$$OPT_i \equiv X_i \cup (OPT \cap \{i+1, \ldots, n\});$$

same as X_i on $\{1, 2, \ldots, i\}$ and same as OPT on $\{i+1, \ldots, n\}$.

$$OPT_0 = OPT \qquad OPT_n = X_n = \hat{X}_n$$

More Notation

Let OPT be an optimal set, and

$$OPT_i \equiv X_i \cup (OPT \cap \{i+1, \ldots, n\});$$

same as X_i on $\{1, 2, \ldots, i\}$ and same as OPT on $\{i+1, \ldots, n\}$.

$$OPT_0 = \qquad\qquad OPT_n =$$

Main Lemma

$$E[f(OPT_{i-1}) - f(OPT_i)] \leq \frac{1}{2} E[f(X_i) - f(X_{i-1}) + f(\hat{X}_i) - f(\hat{X}_{i-1})].$$

Main Result

Theorem (Buchbinder et al. (2012))

DoubleGreedy is a $\frac{1}{2}$-approximation algorithm for maximizing a nonmonotone submodular function.

Main Result

Theorem (Buchbinder et al. (2012))

DoubleGreedy is a $\frac{1}{2}$-approximation algorithm for maximizing a nonmonotone submodular function.

Lemma

$$E[f(OPT_{i-1}) - f(OPT_i)] \leq \frac{1}{2} E[f(X_i) - f(X_{i-1}) + f(\hat{X}_i) - f(\hat{X}_{i-1})].$$

Proof of the Lemma

Lemma

$$E[f(OPT_{i-1}) - f(OPT_i)] \leq \frac{1}{2} E[f(X_i) - f(X_{i-1}) + f(\hat{X}_i) - f(\hat{X}_{i-1})].$$

Pf $\sum_{i=1}^{n} E[f(OPT_{i-1}) - f(OPT_i)] \leq \frac{1}{2} \sum_{i=1}^{n} E[f(X_i) - f(X_{i-1}) + f(\hat{X}_i) - f(\hat{X}_{i-1})]$

$$E[f(OPT_0)] - E[f(OPT_n)] \leq \frac{1}{2} E[f(X_n) - f(X_0) + f(\hat{X}_n) - f(\hat{X}_0)]$$

$$f(OPT) - E[f(X_n)] \leq \frac{1}{2} E[f(X_n) + f(\hat{X}_n)] = E[f(X_n)]$$

$$f(OPT) \leq 2E[f(X_n)]$$

$$\rightarrow \quad E[f(X_n)] \geq \frac{1}{2} f(OPT)$$

Proof of the Lemma

Lemma

$$E[f(OPT_{i-1}) - f(OPT_i)] \leq \frac{1}{2} E[f(X_i) - f(X_{i-1}) + f(\hat{X}_i) - f(\hat{X}_{i-1})].$$

$X_0 \leftarrow \emptyset$

for $i \leftarrow 1$ **to** n **do**

$a_i \leftarrow f(X_{i-1} \cup \{i\}) - f(X_{i-1})$

$r_i \leftarrow f(\hat{X}_{i-1} - \{i\}) - f(\hat{X}_{i-1})$

if $a_i \geq 0, r_i < 0$ **then**

$X_i \leftarrow X_i \cup \{i\}$ // so $\hat{X}_i = \hat{X}_{i-1}$

if $a_i < 0, r_i \geq 0$ **then**

$X_i \leftarrow X_{i-1}$ // so $\hat{X}_i = \hat{X}_{i-1} - \{i\}$

else

$$X_i \leftarrow \begin{cases} X_{i-1} \cup \{i\} & \text{w. prob } \frac{a_i}{a_i + r_i} \\ X_{i-1} & \text{w. prob } \frac{r_i}{a_i + r_i} \end{cases}$$

Pf of lemma Assume $i \notin OPT$. Pf analogous $i \in OPT$.

Case (1): $a_i < 0$, $r_i \geq 0$, so i removed. $X_i = X_{i-1}$

$\hat{X}_i = \hat{X}_{i-1} - \{i\}$

Since $i \notin OPT$, $OPT_i = OPT_{i-1}$.

LHS 0

RHS is $\frac{1}{2}\left[f(\hat{X}_{i-1} - \{i\}) - f(\hat{X}_{i-1})\right] = \frac{1}{2} r_i \geq 0$

Case (2): $a_i \geq 0$, $r_i < 0$, so i added.

$X_i = X_{i-1} \cup \{i\}$, $\hat{X}_i = \hat{X}_{i-1}$, $OPT_i = OPT_{i-1} \cup \{i\}$

Note $OPT_{i-1} \subseteq \hat{X}_{i-1} - \{i\}$

By submodularity

$$f(\hat{X}_{i-1}) - f(\hat{X}_{i-1} - \{i\}) \leq f(OPT_{i-1} \cup \{i\}) - f(OPT_{i-1})$$
$$= f(OPT_i) - f(OPT_{i-1})$$

$-r_i$

LHS $f(OPT_{i-1}) - f(OPT_i) \leq r_i \leq 0.$

RHS $\frac{1}{2}\left(f(X_i) - f(X_{i-1})\right) = \frac{1}{2}\left(f(X_{i} \cup \{i\}) - f(X_{i-1})\right)$
$$= \frac{1}{2} a_i \geq 0.$$

Case (3): $a_i > 0$, $r_i > 0$.

LHS: $E[f(OPT_{i-1}) - f(OPT_i)] \leq 0 \cdot \frac{r_i}{a_i + r_i} + r_i \cdot \frac{a_i}{a_i + r_i}$

RHS: $\frac{1}{2} E[f(X_i) - f(X_{i-1}) + f(\hat{X}_i) - f(\hat{X}_{i-1})]$

$$= \frac{1}{2} r_i \frac{r_i}{a_i + r_i} + \frac{1}{2} a_i \frac{a_i}{a_i + r_i} = \frac{1}{2}\left(\frac{a_i^2 + r_i^2}{a_i + r_i}\right)$$

Want $\frac{a_i r_i}{a_i + r_i} \leq \frac{1}{2}\left(\frac{a_i^2 + r_i^2}{a_i + r_i}\right)$ since $a_i^2 + r_i^2 - 2a_i r_i \geq 0$

$(a_i - r_i)^2 \geq 0$. $\square$

Limits

Theorem (Feige, Mirrokni, Vondrák (2007))

There is no $\left(\frac{1}{2}+\epsilon\right)$-approximation algorithm for maximizing a nonmonotone submodular function in the oracle model for constant $\epsilon > 0$.

~~Open~~ Problem: Can you get a $\frac{1}{2}$-approx. alg. deterministically?

As of 8/10/15: Yes! (Buchbinder, Feldman)