Approximation Algorithms

- NP-hard optimization problems (under the assumption $P \neq NP$) cannot be solved optimally in poly-time
- Relax optimality: solve with a guaranteed approximation

Definitions: An NP optimization problem Π consists of the following:

valid instances: each instance I is like an input

Feasible solution: For each I, $S_\Pi(I)$ is the set of all solutions for instance I

objective function: A poly-time computable func. $f(s,I)$ that gives a non-neg value for each solution s for I.

Definition: For a minimization problem Π, an algorithm A is an α-approximation if it is polytime & for each instance I produces a solution $s \in S_\Pi(I)$ s.t.

$$f(s,I) \leq \alpha(|I|) \cdot OPT(I),$$

where $OPT(I)$ is the smallest objective func. value among all feasible solutions of I.

if we use $A(I)$ to denote the value of solution returned by A then $A(I) \leq \alpha(|I|) \cdot OPT(I)$

We define approximation for maximization problems similarly

Asymptotic α-approx: if $\lim_{|I| \to \infty} \frac{A(I)}{opt(I)} \leq \alpha$.

PTAS (Polynomial time approximation Scheme): Is a series of algorithms s.t. for any given $\varepsilon > 0$ the algorithm A_ε (that is dependent on ε) runs in time polynomial in n and returns a $(1+\varepsilon)$-approximate solution.

(e.g. the time can be $O(n^{\frac{1}{\varepsilon^2}})$ or $O(2^{2^{1/\varepsilon}} \cdot n^{1/\varepsilon^{10}})$)

FPTAS (Fully PTAS): Same as PTAS but the time is polynomial in both n & $\frac{1}{\varepsilon}$ (e.g. $O(n^3/\varepsilon^5)$)

Set Cover & Max_Coverage

Set Cover

universe of n elements

$\mathcal{U} = \{e_1, e_2, \dots, e_n\}$

$S = \{S_1 \dots S_m\}$ collection of subsets of $\mathcal{U}$

$c : S \longrightarrow \mathbb{Q}^+$

Goal: find a min-cost subcollection of S that covers $\mathcal{U}$.

i.e. Find $I \subseteq \{1..m\}$ $\bigcup_{i \in I} S_i = \mathcal{U}$ & $\min \sum_{i \in I} c(S_i)$

Def: Frequency of an elmt e_i is # of sets S_j that contain e_i.

Max_Coverage: input is $\mathcal{U}$, S, and int k

Goal: Find a collectn of k set that max # of elements covered.

Natural Greedy for both S.C. & Max_Coverage

- At each step pick a set that covers the max # of uncovered elements at lowest cost.
- repeat until done: Set Cover: all elements are covered; Max_Coverage: k iterations

Theorem: Greedy is a $\underbrace{(\ln n + 1)}_{H_n}$-approx for Set Cover and a $1-(1-\frac{1}{k})^k \geq 1 - \frac{1}{e}$ approx for max-coverage.

Analysis for max_Coverage

Let opt^* be an opt sol with value opt.

Analysis for max coverage

let opt^* be an opt sol with value opt.

- let a_i be # of new elements covered at iteration i

and $b_i = \sum_{j=1}^{i} a_j$ and $\underbrace{c_i = opt - b_i}_{\text{how far from opt}}$

$a_0 = b_0 = 0$ & $c_0 = opt$

lemma 1: For all $0 \le i < k$: $\boxed{a_{i+1} \geq \frac{c_i}{k}}$

Proof: Consider iteration $i+1$. Since b_i elements are covered until end of iteration i, c_i elements of opt^* are left uncovered. let Z^* be the elements of opt^* that are not covered yet. $|Z^*| \geq c_i$

Since opt uses k sets, all elements of Z^* are covered (in opt) using k sets. $\longrightarrow$ $\exists$ a set that covers $\geq \frac{|Z^*|}{k}$ elements of Z^*

$\longrightarrow$ $\exists$ a set in iteration $i+1$ that can be chosen to cover $\geq \frac{|Z^*|}{k} \geq \frac{c_i}{k}$ $\longrightarrow$ $a_{i+1} \geq \frac{c_i}{k}$ ∎

lemma 2: For all $i \geq 0$: $c_i \leq (1-\frac{1}{k})^i \cdot opt$

(so $c_k \leq (1-\frac{1}{k})^k \cdot opt$)

proof: by induction on i

$\rightarrow$ $b_k \geq \left(1-(1-\frac{1}{k})^k\right)opt$

- $i=0$ base
- Ind step:

$$c_{i+1} = c_i - a_{i+1}$$

$$\leq c_i - \frac{c_i}{k} \quad \text{by lemma 1}$$

$$= c_i(1-\frac{1}{k})$$

$$\leq (1-\tfrac{1}{k})^{i} \cdot opt \cdot (1-\tfrac{1}{k})$$ Ind. Hyp.

$$= (1-\tfrac{1}{k})^{i+1} \cdot opt$$

Greedy SetCover SC1

$C \leftarrow \phi$
$S' \leftarrow \phi$
while $C \neq U$
  select $S_i \in S$ with min cost effectiveness $\alpha = \dfrac{c(S_i)}{|S_i - C|}$
  $S' \leftarrow S' \cup \{S_i\}$
  $C \leftarrow C \cup S_i$
  [for each newly covered $e \in S_i$ price (e) as α]

# of new elements covered

Lemma: $\sum_{e \in U} \text{price}(e) \leq H_n \cdot OPT$ $H_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \approx \ln n + 1$

order elements of U by the time they are covered by SC1
$e_1, e_2, \ldots, e_n$ (breaking ties arbit).

Consider the time just before e_k is covered.
So $n-k+1$ (uncovered) elements can be covered at a total price of opt. So we can pick some sets of total cost opt to cover $n-k+1$ elements.

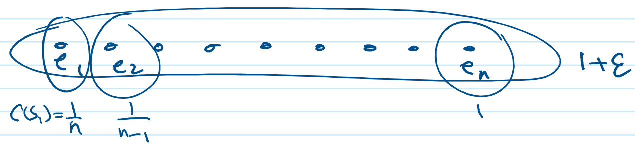
$\longrightarrow \exists$ a set which can cover elements at a price of $\leq \dfrac{opt}{n-k+1}$ So e_k pays no more than this to be covered.

$$\sum_{i=1}^{n} \text{Price}(e_i) \leq \sum_{i=1}^{n} \frac{opt}{n-i+1} = opt \cdot \sum_{i=1}^{n} \frac{1}{n-i+1} = H_n \cdot OPT$$

Unless P=NP there is no $(1-\varepsilon)\ln n$-approx for

Unless P=NP there is no $(1-\varepsilon)\ln n$-approx for Set Cover for any $\varepsilon > 0$.

Tight example:



opt $= 1+\varepsilon$

Greedy $= H_n$

Randomized Rounding: Set Cover

Rounding an LP solution to an integer one while bounding the change in obj value is a common way of designing approximation algorithms.

IP/LP formulation

$$\min \sum c_S . x_S$$

$$\forall e \in U: \sum_{e \in S} x_S \geq 1$$

$$x_S \geq 0$$

LP rounding for SC

Let x^* be an LP opt. solution

let $C \leftarrow \phi$

for each set S_i set $\tilde{x}_{S_i} = 1$ with prob. $x^*_{S_i}$

let $C = \{S_i \mid \tilde{x}_{S_i} = 1\}$

return C

First let's compute the expected cost of C,

$$E[\text{cost of } C] = \sum_{S_i \in S} \Pr[S_i \text{ is selected}] . c_{S_i} = \sum_{S_i} x^*_{S_i} . c_{S_i} = OPT_{LP}$$

The problem is that C may not be a set cover solution.

The problem is that C may not be a Set Cover solution.

let α be large enough s.t. $(\frac{1}{e})^{\alpha \ln n} \leq \frac{1}{4n}$ and suppose we repeat this algorithm α times.

let C_i be the solution for repeatition i and return

$$C = \bigcup_{i=1}^{\alpha \ln n} C_i$$

Easy to see: $E[\text{cost of } C] = \sum_{i=1}^{\alpha} E[\text{cost}(C_i)] = (\alpha \ln n)\, opt_{LP}$

We prove that w.h.p. C is a feasible solution.

– Consider some element e_j and suppose it belongs to sets $S_1 \ldots S_k$. So $x^*_{S_1} + x^*_{S_2} + \cdots + x^*_{S_k} \geq 1$

Lemma: Given $x^*_{S_1} + x^*_{S_2} + \cdots + x^*_{S_k} = 1$ if we select each S_i with prob. $x^*_{S_i}$ the prob. that none is selected is maximized when all $x^*_{S_i}$'s are equal to $\frac{1}{k}$

Proof: Calculus.

In this case: $\Pr[e_j \text{ is not covered in one iteration}] \leq (1 - \frac{1}{k})^k \leq 1/e$

So over all iterations: $\Pr[e_j \notin \bigcup_{i=1}^{\alpha \ln n} C_i] \leq (\frac{1}{e})^{4 \ln n} = \frac{1}{4n}$

Using union bound: $\Pr[C \text{ is not a S.C}] \leq n \cdot \frac{1}{4n} = \frac{1}{4}$ event E_1

Using Markov's inequality: $\Pr[\text{cost}(C) > 4 \ln n \, opt_{LP}] \leq \frac{1}{4}$ event E_2

→ with prob $\geq \frac{1}{2}$ E_1 & E_2 don't happen i.e.

With prob $\geq \frac{1}{2}$ (C is a S.C.) and ($\text{cost}(C) \leq 4 \ln n \cdot opt_{LP}$).