Using Homogeneous Weights for Approximating the Partial Cover Problem *

Reuven Bar-Yehuda
Computer Science Department
Technion - IIT, Haifa 32000 Israel

e-mail: reuven@cs.technion.ac.il
http://www.cs.technion.ac.il/~reuven

October 27, 2000

Abstract

In this paper we consider the following natural generalization of two fundamental problems: the Set-Cover problem and the Min-Knapsack problem. We are given an hypergraph, each vertex has a nonnegative weight and each edge has a nonnegative length. For a given threshold $\widehat{\ell}$, our objective is to find a subset of the vertices with minimum total cost, such that at least a length of $\widehat{\ell}$ of the edges are covered. This problem is called the *partial set cover* problem. We present an $O(|V|^2+|H|)$ time, Δ_E-approximation algorithm for this problem, where $\Delta_E \geq 2$ is an upper bound on the edge cardinality of the hypergraph, and $|H|$ is the size of the hypergraph (i.e. the sum of all its edges cardinalities). The special case where $\Delta_E = 2$ is obviously the partial vertex cover problem. For this problem a 2-approximation was previously known, however, the time complexity of our solution, i.e. $O(|V|^2)$, is a dramatic improvement.

We show that if the weights are *homogeneous* (i.e., proportional to the potential coverage of the sets) then any minimal cover is a good approximation. Now, using the local-ratio technique, it is sufficient to repeatedly subtract a homogeneous weight function from the given weight function.

Keywords: Approximation Algorithm, Local Ratio, Covering Problems, Vertex Cover, Set Cover, Partial Covering, Knapsack.

*This research was supported by the U.D. Erteschik fund for practical research.

1 Introduction

1.1 Definitions and Notations

We are given a hypergraph $H = (V, E)$, a weight function $\omega : V \to I\!R^+$, a length function $\ell : E \to I\!R^+$, and a covering bound $\widehat{\ell} \in I\!R^+$. Define:

$\omega(V') = \sum_{v \in V'} \omega(v)$ for $V' \subseteq V$.

$\ell(E') = \sum_{e \in E'} \ell(e)$ for $E' \subseteq E$.

$E(V') = \{e \in E : e \cap V' \neq \phi\}$.

$E(v) = E(\{v\})$ for $v \in V$.

$d(v) = |E(v)|$ for $v \in V$.

$|H| = \sum_{v \in V} d(v) = \sum_{e \in E} |e|$ (the size of the hypergraph)

$\Delta_V = \max_{v \in V} d(v)$.

$\Delta_E = \max_{e \in E} |e|$. We assume $\Delta_E \geq 2$, unless stated otherwise.

We say that C is a *feasible solution* (or just an *$\widehat{\ell}$-cover*) if $C \subseteq V$ and $\ell(E(C)) \geq \widehat{\ell}$. An $\widehat{\ell}$-cover C^* is optimal if $\omega(C^*) \leq \omega(C)$ for all feasible solutions C. A feasible solution C is called an *r-approximate cover*, if $\omega(C) \leq r \cdot \omega(C^*)$.

An algorithm A is called an r-approximation algorithm if for all instances $H, \omega, \ell, \widehat{\ell}$, A returns an r-approximate cover.

Notice that there is a feasible solution iff $\ell(E) \geq \widehat{\ell}$. When $\ell(E) = \widehat{\ell}$ and $\ell(\cdot) > 0$ the length function ℓ is redundant and we get the well known classic problem, the Set Cover problem.

1.2 Some History of the Vertex Cover and the Set Cover Problems

A set $C \subseteq V$ is called a *vertex cover* of a graph $G = (V, E)$ if every edge has at least one endpoint in C, i.e., $\forall_{e \in E}\ e \cap C \neq \emptyset$. The vertex cover problem (VC) is: given a graph $G = (V, E)$ and a weight function $\omega : V \to I\!R^+$, find a vertex cover C with minimum total weight. The vertex cover problem (VC) is NP-hard even for planar cubic graphs with unit weights [7]. For unit weight vertex cover, Gavril (see [7]) suggested a linear time 2-approximation algorithm. For the general vertex cover problem, Nemhauser and Trotter [11] developed a local optima algorithm that implies a 2-approximation.

Currently, the best ratio known is 2. The first linear time algorithm was found by Bar-Yehuda and Even [2]. Their proof uses the primal-dual approach. It took a few more years to find a different kind of proof – the local-ratio theorem [3]. Recently, Bar-Yehuda [1] has presented a unified approach and a generic approximation algorithm for a family of covering problems. The algorithm in this paper uses this approach.

The *set-cover* problem is a generalization of the vertex cover problem. A set $C \subseteq V$ is called a *set cover* of a hypergraph $H = (V, E)$ if every edge has at least one endpoint in C, i.e., $\forall_{e \in E}\ e \cap C \neq \phi$. The set cover problem (SC) is: given a hypergraph $H = (V, E)$ and a weight function $\omega : V \to I\!R^+$ find a set cover C with minimum total weight. The best known approximation algorithms for SC are a $(\ln \Delta_V)$-approximation algorithm by Chvátal [6] and a Δ_E-approximation linear time algorithm by Bar-Yehuda and Even [2].

1.3 The Unit Length Partial Covering Problems

The partial covering case in which all the edges have unit length, called the $\widehat{\ell}$-Set-Cover problem, was first studied by Kearns [9] in relation to learning. In this excellent Ph.D. dissertation, the Chvátal greedy approach [6] is also studied. Later Slavík [12] showed that the $(\ln \Delta_V)$-approximation of set cover can be extended to the case in which $\widehat{\ell} = p \cdot |E|$ for any given constant $0 \le p \le 1$. Only recently, Burroughs in her master's thesis [5] extended this result to any $\widehat{\ell} \le |E|$. The special case in which all the edges have unit length and the edge cardinality is exactly 2, called the $\widehat{\ell}$-Vertex-Cover problem, was studied by Bshouty and Burroughs [4]. They have presented the first polynomial-time 2-approximation algorithm, which is based on solving a linear program. They also showed that improving the ratio to a constant smaller than 2 would be a breakthrough. They have shown that such an improvement on many families of instances would imply improving the ratio for the classic VC problem as well. Recently Hochbaum [8] presented an $O(|E|\,|V|\log(\frac{|V|^2}{|E|})\log|V|)$ time 2-approximation algorithm for the $\widehat{\ell}$-Vertex-Cover problem. Note that the time complexity of our 2-approximation algorithm for this problem is $O(|V|^2)$.

1.4 A Special Case: The Min Knapsack Problem

The *Knapsack* decision problem is defined as follows:

Instance: A set $V = \{1, 2, ..., n\}$, weight function $\omega : V \to I\!R^+$, length function $\ell : V \to I\!R^+$, $\widehat{\ell} \in I\!R^+$ and $\widehat{\omega} \in I\!R^+$.
Question: Is there a set $C \subseteq V$ s.t $\omega(C) \le \widehat{\omega}$ and $\ell(C) \ge \widehat{\ell}$?

There are two related optimization problems to the Knapsack decision problem, one [1] is the *Maximum Knapsack Problem*: $\max\{\ell(C) : C \subseteq V, \omega(C) \le \widehat{\omega}\}$ and the other is the *Minimum Knapsack Problem*: $\min\{\omega(C) : C \subseteq V, \ell(C) \ge \widehat{\ell}\}$. The later is obviously a special case of our problem, when $\Delta_E = 1$. The traditional approach to attacking the knapsack problem (min and max) is to greedily select the vertex with the smallest ratio of weight per length. A simple modification of this is known to be a 2-approximation. This approach can be generalized to the partial covering problem and, as we observed, this is exactly the approach of Chvátal's greedy algorithm [6] for set cover, and of its generalizations to partial set cover, [12, 4]. However, the greedy approach can be generalized in another direction. A knapsack is *homogeneous* if $\omega = \ell$. Homogeneous knapsacks are interesting in themselves, see e.g Kearns [9]. In this case, every selection order is greedy, and it can be proved that

[1] See Khuller, Moss, and Naor [10] for a $(1 - e^{-1})$-approximation.

any minimal set C is a 2-approximation. Our generalized approach is to repeatedly subtract from ω a homogeneous weight, e.g., $\epsilon \cdot \ell$, for some $\epsilon > 0$. So, implementing our algorithm for the special case of min-knapsack gives a 2-approximation. It is, however, still interesting to observe that the case $\Delta_E = 1$ (i.e. the min-knapsack) is known to have a full approximation scheme (which is an 'almost 1-approximation') and this is in accordance with our intuition that a Δ_E-approximation should be available for all $\Delta_E \geq 1$.

1.5 Overview

In section 2 we define the homogeneous weight function and prove its useful properties. In section 3 we present the Δ_E-approximation algorithm. The main step of this algorithm is a subtraction of homogeneous weight function from the given function weight. Using the properties of the homogeneous function and the local ratio technique we prove its correctness.

2 The Homogeneous Weight Function

We are given a hypergraph $H = (V, E)$, a length function ℓ, and a covering bound $\widehat{\ell}$. For every $v \in V$ we define the *effective degree* of v by: $\delta(v) = \min\{\widehat{\ell}, \ell(E(v))\}$.

An $\widehat{\ell}$-cover C is minimal if $\forall_{v \in V} C \setminus \{v\}$ is not an $\widehat{\ell}$-cover.

With respect to $H, \ell, \widehat{\ell}$, the weight function δ is called homogeneous. The main valuable property of the homogeneous weight function is that any minimal cover is a "good" approximation. In the next two lemmas we show this fact by proving that the total weight of any minimal $\widehat{\ell}$-cover is in the interval: $[\widehat{\ell}, \Delta_E \widehat{\ell}]$.

Lemma 1 *Let δ be an homogeneous weight function with respect to $H, \ell, \widehat{\ell}$. If C is any $\widehat{\ell}$-cover of H then $\delta(C) \geq \widehat{\ell}$*

Proof. If $\delta(v) = \widehat{\ell}$ for some $v \in C$ then obviously $\delta \geq \widehat{\ell}$. Otherwise,

$$\begin{array}{rcll} \delta(C) & = & \sum_{v \in C} \delta(v) & \text{[by definition]} \\ & = & \sum_{v \in C} \ell(E(v)) & [\delta(v) < \widehat{\ell}] \\ & \geq & \ell(E(C)) & [E(C) = \bigcup_{v \in C} E(v)] \\ & \geq & \widehat{\ell} & \text{[by definition of } \widehat{\ell}\text{-cover]} \end{array}$$

□

Lemma 2 *Let δ be an homogeneous weight function with respect to $H, \ell, \widehat{\ell}$. If C is any minimal $\widehat{\ell}$-cover, then $\delta(C) \leq \Delta_E \cdot \widehat{\ell}$.*

Proof. We distinguish between three cases:

Case 1: $|C| \leq \Delta_E$:

If $|C| \leq \Delta_E$ then $\delta(C) = \sum_{v \in C} \delta(v) \leq |C| \cdot \max_{v \in C} \delta(v) \leq |C| \cdot \widehat{\ell} \leq \Delta_E \cdot \widehat{\ell}$.

Case 2: $\exists_{v \in C} \delta(v) = \widehat{\ell}$:

If $\delta(v) = \widehat{\ell}$ for some $v \in C$ then minimality implies $C = \{v\}$ and therefore $|C| = 1 < \Delta_E$, which is case 1.

Case 3: $|C| > \Delta_E$ and $\forall_{v \in C} \delta(v) < \widehat{\ell}$:

Partition $E(C)$ into two sets. The set of edges 'covered' by exactly one endpoint in C; $E_1 = \{e : |e \cap C| = 1\}$, and the others that are 'covered' by more endpoints: $E_2 = \{e : |e \cap C| > 1\}$.

Now,

$$\begin{array}{rcll} \delta(C) & = & \sum_{v \in C} \delta(v) & \text{[by definition]} \\ & = & \sum_{v \in C} \ell(E(v)) & [\forall_{v \in C} \delta(v) < \widehat{\ell}] \\ & \leq & \Delta_E \cdot \ell(E_2) + \ell(E_1) & \text{[by definition of } E_1, E_2] \\ & = & \Delta_E \cdot \ell(E(C)) - (\Delta_E - 1) \cdot \ell(E_1) & \\ & = & \Delta_E \cdot (\widehat{\ell} + \widehat{\ell}') - (\Delta_E - 1) \cdot \ell(E_1) & \text{[define: } \widehat{\ell}' = \ell(E(C)) - \widehat{\ell}] \\ & \leq & \Delta_E \cdot (\widehat{\ell} + \widehat{\ell}') - \ell(E_1) & \text{[since } \Delta_E \geq 2] \\ & = & \Delta_E \cdot \widehat{\ell} + (\Delta_E \cdot \widehat{\ell}' - \ell(E_1)) & \end{array}$$

To complete the proof we have to show that $\ell(E_1) \geq \Delta_E \cdot \widehat{\ell}'$. Let $v \in C$, and let $E_1(v) = E_1 \cap E(v)$. Since C is a minimal $\widehat{\ell}$-cover, it follows that $\ell(E(C \backslash \{v\})) < \widehat{\ell}$. The contribution of $E_1(v)$ to the total length must therefore satisfy: $\ell(E_1(v)) > \widehat{\ell}'$. The same is true for all the vertices in C, and therefore $\ell(E_1) > |C| \cdot \widehat{\ell}'$. Since $|C| > \Delta_E$, it follows that $\ell(E_1) > \Delta_E \cdot \widehat{\ell}'$. □

Now, let us extend the definition to the $\epsilon-$homogeneous weight function. ϖ is $\epsilon-$homogeneous for a given $\epsilon > 0$ and with respect to $H, \ell, \widehat{\ell}$ if $\forall_{v \in V} \varpi(v) = \epsilon \cdot \delta(v)$. We conclude this section with the following useful theorem:

Theorem 2.1 *Let ϖ be an $\epsilon-$homogeneous weight function with respect to $H, \ell, \widehat{\ell}$. If C is a minimal $\widehat{\ell}$-cover and C^* is an optimal one, then $\varpi(C) \leq \Delta_E \varpi(C^*)$.*

Proof.

$$\begin{array}{rcll} \varpi(C) & \leq & \Delta_E \cdot \epsilon \cdot \widehat{\ell} & \text{[by lemma 2]} \\ & \leq & \Delta_E \cdot \varpi(C^*) & \text{[by lemma 1]} \end{array}$$

□

3 The Δ_E-approximation algorithm

Our algorithm is written in recursive fashion. It is based on iterative subtraction of an $\epsilon-$homogeneous weight function, from the current weight function to obtain a new weight function. The algorithm **Cover** $(V, E, \omega, \ell, \widehat{\ell})$ is described in Figure 1.

Lemma 3 *Algorithm **Cover** has at most $2|V|$ iterations.*

```
Algorithm Cover (V, E, ω, ℓ, ℓ̂)
  If ℓ̂ ≤ 0
    return φ
  If E = φ
    return 'no solution'
  V_0 = {v : δ(v) = 0}
  If V_0 ≠ φ
    return Cover(V\V_0, E\E(V_0), ω, ℓ, ℓ̂)
  V_0 = {v : ω(v) = 0}
  If V_0 ≠ φ
    return V_0 ∪ Cover(V\V_0, E\E(V_0), ω, ℓ, ℓ̂ − |ℓ(E(V_0))|)

  /*At this point ℓ̂ > 0, E ≠ φ, ω(·) > 0, and δ(·) > 0*/
  /*we find an ε−homogeneous weight function*/
  ε = min_{v∈V} ω(v)/δ(v)
  for each v ∈ V
    ω_1(v) = ε · δ(v) /*the ε−homogeneous weight function*/
    ω_2(v) = ω(v) − ω_1(v) /*the residual weight function*/

  /*Main recursion call*/
  C = Cover(V, E, ω_2, ℓ, ℓ̂)
  /*Minimality loop*/
  for each v ∈ C
    if C\{v} is a ℓ̂-cover
      C = C\{v}
  return C
```

Figure 1: Algorithm **Cover**

Proof. Since in each of the first two recursive calls, at least one vertex is removed from V, it follows that these two calls are performed at most $|V|$ times. When ϵ is computed for all $v \in V$, we have $\omega(v) > 0$. Clearly, by the computation of ϵ, $\omega_1(\cdot)$, and $\omega_2(\cdot)$, at least one vertex $v \in V$ has the value $\omega_2(v) = 0$. Thus, the third recursive call is performed at most $|V|$ times, and we have total of at most $2|V|$ iterations. $\square$

Theorem 3.1 *Algorithm **Cover** is a Δ_E-approximation for the $\widehat{\ell}$-cover problem.*

Proof. By lemma 3, we have at most $2|V|$ iterations. The proof proceeds by induction on the number of iterations.
Base: 0 iterations implies that, $\widehat{\ell} \leq 0$, and the empty set which is returned by the algorithm is an optimal $\widehat{\ell}$-cover.

Step: We assume $\widehat{\ell} > 0$, $E \neq \phi$, $\omega(\cdot) > 0$, and $\delta(\cdot) > 0$, since, otherwise, the induction step is trivial. Let C be the minimal $\widehat{\ell}$-cover obtained after the minimality loop. Let C^*, C_1^*, and C_2^* be optimal $\widehat{\ell}$-covers with respect to ω, ω_1, and ω_2.

$$
\begin{array}{rcll}
\omega(C) & = & \omega_1(C) + \omega_2(C) & \text{[by definition]} \\
& \leq & \Delta_E \cdot \omega_1(C_1^*) + \Delta_E \omega_2(C_2^*) & \text{[by theorem 2.1 and induction hyp]} \\
& \leq & \Delta_E \cdot \omega_1(C^*) + \Delta_E \omega_2(C^*) & [C_1^* \text{ and } C_2^* \text{ are } \widehat{\ell}\text{-cover of } \omega_1 \text{ and } \omega_2] \\
& \leq & \Delta_E \cdot \omega(C^*) & [\omega = \omega_1 + \omega_2]
\end{array}
$$

□

Lemma 4 *Algorithm* ***Cover*** *can be implemented in time* $O(|V|^2 + |H|)$.

Proof. To update the value of δ, we have to update the value of $\ell(E(v))$ for each non-deleted vertex $v \in V$. This value is changed only when a vertex is deleted, and all its adjacent edges should be deleted. For each deleted edge $e \in E$, $\ell(e)$ is subtracted from all $|e|$ vertices contained in e. Therefore, the total time devoted to update δ for all vertices is bounded by $O(\sum_{e \in E} |e|) = O(|H|)$

By lemma 3 the total number of iterations is bounded by $2|V|$. At each iteration we need $O(|V|)$ other operations, and the total time for all other operations is $O(|V|^2)$.

□

Corollary 1 *For simple graphs the* $\widehat{\ell}$*-Vertex Cover problem has an* $O(|V|^2)$ *time 2-approximation algorithm.*

Acknowledgment
We would like to thank Lynn Burroughs, Tuvi Etzion and Dror Rawitz for helpful discussions, and Hadas Heier and Avigail Orni for their careful reading and suggestions.

References

[1] R. Bar-Yehuda. One for the price of two: A unified approach for approximating covering problems. In *APPROX'98 1st International Workshop on Approximation Algorithms for Combinatorial Optimization Problems*, 7 1998. To appear in Algorithmica.

[2] R. Bar-Yehuda and S. Even. A linear time approximation algorithm for the weighted vertex cover problem. *Journal of Algorithms*, 2:198–203, 1981.

[3] R. Bar-Yehuda and S. Even. A local-ratio theorem for approximating the weighted vertex cover problem. *Annals of Discrete Mathematics*, 25:27–46, 1985.

[4] N. Bshouty and L. Burroughs. Massaging a linear programming solution to give a 2-approximation for a generalization of the vertex cover problem. In *The Proceedings of the Fifteenth Annual Symposium on the Theoretical Aspects of Computer Science*, pages 298–308, 1998.

[5] L. Burroughs. Approximation algorithms for covering problems. Master's thesis, University of Calgary, February 1998.

[6] V. Chvátal. A greedy heuristic for the set-covering problem. *Mathematics of Operations Research*, 4(3):233–235, 1979.

[7] M. Garey and D. Johnson. *Computers and Intractability; A Guide to the Theory of NP-Completeness*. W.H. Freeman and Company, 1979.

[8] D. S. Hochbaum. The t - vertex cover problem: Extending the half integrality framework with budget constraints. *APPROX'98 1st International Workshop on Approximation Algorithms for Combinatorial Optimization Problems*, pages 111–122, 1998.

[9] M. Kearns. *The Computational Complexity of Machine Learning*. M.I.T. Press, 1990.

[10] S. Khuller, A. Moss, and J. Naor. The budgeted maximum coverage problem. *Submited*, 1998.

[11] G. L. Nemhauser and J. L. E. Trotter. Vertex packings: structural properties and algorithms. *Mathematical Programming*, 8:232–248, 1975.

[12] P. Slavík. Improved performance of the greedy algorithm for partial cover. *Inf. Process. Lett.*, 64(5):251–254, 15 Dec. 1997.