AD-A104 829 CARNEGIE-MELLON UNIV PITTSBURGH PA MANAGEMENT SCIENC--ETC F/G 12/1
OPTIMAL INTEGER AND FRACTIONAL COVERS: A SHARP BOUND ON THEIR R--ETC(U)
MAY 81 E BALAS
UNCLASSIFIED MSRR-475 N00014-75-C-0621
NL

1 OF 1
ADA
104829

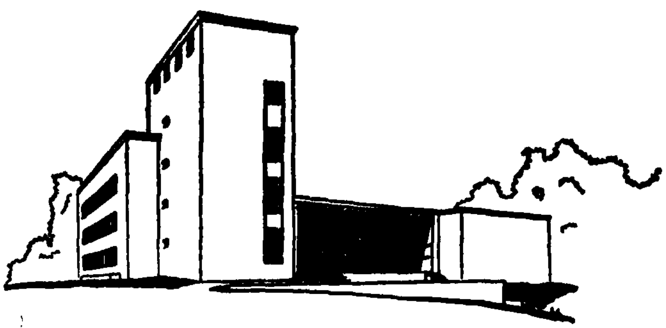
END
DATE
FILMED
10-81
DTIC

Carnegie-Mellon University

PITTSBURGH, PENNSYLVANIA 15213

GRADUATE SCHOOL OF INDUSTRIAL ADMINISTRATION

WILLIAM LARIMER MELLON, FOUNDER



W.P.# 53-80-81

Management Sciences Research Report No. 475*

OPTIMAL INTEGER AND FRACTIONAL COVERS: A SHARP BOUND ON THEIR RATIO

by

Egon Balas

May 1981

The research underlying this report was supported by Grant ECS-7902506 of the National Science Foundation and Contract N00014-75-C-0621 NR 047-048 with the U.S. Office of Naval Research. It was also supported by the Alexander von Humboldt Foundation of the Federal Republic of Germany through a Senior U.S. Scientist Award.

Management Science Research Group
Graduate School of Industrial Administration
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

*Issued simultaneously as Report 81-13 of the Mathematisches Institut, University of Cologne.

Abstract

The ratio of the values of optimal integer and fractional solutions to a set covering problem was shown by Johnson [5] and Lovász [6] to be bounded by $B(d) = 1 + \ln d$, where d is the largest column sum. We show that if n is the number of variables, $B(n) = \frac{1}{n}\left\lfloor\frac{n+1}{2}\right\rfloor\left\lceil\frac{n+1}{2}\right\rceil$ is a best possible bound on this ratio. Furthermore, for every $n \geq 20$ there are problems for which $B(n) \leq \frac{1}{2.5}B(d)$.

A

OPTIMAL INTEGER AND FRACTIONAL COVERS: A SHARP BOUND ON THEIR RATIO

by

Egon Balas

The simple (unweighted) set covering problem is

$$(C) \qquad z_C = \min\{e_n x \mid Ax \geq e_m,\ x \text{ binary}\},$$

where A is an $m \times n$ 0-1 matrix and for $k = m, n$, e_k is the k-vector whose components are all equal to 1, while x is an n-vector of variables.

If the 0-1 condition on the variables is relaxed to nonnegativity, we obtain the continuous or fractional set covering problem

$$(F) \qquad z_F = \min\{e_n x \mid Ax \geq e_m,\ x \geq 0\}.$$

A vector x that satisfies the constraints of (C) (of (F)) will be called a cover (fractional cover).

The set covering problem is known to be NP-complete. One of the best known procedures for finding a cover that approximates the optimum is the greedy heuristic, which consists of a sequence of steps, each of which assigns the value 1 to a variable whose column covers a maximal number of additional rows. The worst case behavior of the greedy heuristic for the (unweighted) set covering problem was shown by Johnson [5] and Lovász [6] to be given by the relation

$$(1) \qquad \frac{z_G}{z_F} \leq H(d) \ (< 1 + \ln d),$$

where z_G is the value of a cover obtained by the greedy heuristic,

$$d = \max_{j\epsilon\{1,\ldots,n\}} \sum_{i=1}^{m} a_{ij},$$

and

$$H(d) = \sum_{j=1}^{d} \frac{1}{j} .$$

Thus the ratio between the value of a "greedy" cover and that of an optimal fractional cover increases at most with the logarithm of the largest column sum.

Chvátal [2] has shown that the worst case bound given by (1) is also valid for the greedy heuristic when applied to the weighted set covering problem with arbitrary but positive cost coefficients c_j, $j = 1,\ldots,n$. If k_{jt} represents the number of new rows covered by column j at step t, the greedy heuristic for the weighted set covering problem assigns the value 1 at step t to a variable x_j whose choice maximizes k_{jt}/c_j. Furthermore, Ho [3] has shown that the bound given by (1) is best possible for any (weighted) set covering heuristic that assigns the value 1 at step t to a variable x_j whose choice maximizes some arbitrary function $f(c_j, k_{jt})$.

Another class of heuristics, which uses information (reduced costs) obtained from a (not necessarily optimal) solution to the dual linear program, has consistently outperformed in empirical tests the greedy heuristic and its above mentioned generalizations (see Balas and Ho [1]), but no worst case bound better than (or comparable to) (1) is known for it (see Hochbaum [4] for a discussion of bounds for this heuristic).

Sinze $z_G \geq z_C \geq z_F$, the relation (1) implies of course both

(2) $$\frac{z_G}{z_C} \leq H(d)$$

and

$$\frac{z_C}{z_F} \le H(d) \ . \tag{3}$$

However, while H(d) is a best possible bound for both z_G/z_F and z_G/z_C, it was until recently an open question whether it is also a best possible bound for z_C/z_F, since no better bound than H(d) was known for this latter ratio.

In this paper we give a best possible bound on the value of z_C/z_F for unweighted set covering problems, as a function of the number n of columns, for an arbitrary number of rows. For every value of n, there are problems for which this bound has a value of approximately $\frac{1}{2.5}$ H(d).

For an arbitrary 0-1 matrix A, we will denote by $z_{C(A)}$ and $z_{F(A)}$ the value of an optimal solution to the (unweighted) set covering problem defined by A, and to the fractional set covering problem defined by A, respectively.

Let $\mathcal{A}^n$ denote the class of 0-1 matrices with at most n columns, and let

$$\mathcal{A}^n(p) = \{A \in \mathcal{A}^n \mid z_{C(A)} = p\} \ .$$

<u>Theorem 1</u>. For any positive integer n and any $p \in \{1,\dots,n\}$,

$$\min_{A \in \mathcal{A}^n(p)} z_{F(A)} = \frac{n}{n-p+1} \ , \tag{4}$$

and the minimum in (4) is attained for the $\binom{n}{k} \times n$ matrix A* whose rows are all the distinct 0-1 n-vectors with exactly n-p+1 components equal to 1.

<u>Proof</u>. We first show that $A^* \in \mathcal{A}^n(p)$. A* has n columns by assumption. Any binary n-vector x having at least p components equal to 1 satisfies $A^*x \ge e_q$, where $q = \binom{n}{k}$, since no row of A* has more than p-1 entries equal to 0. Further, every binary n-vector $\bar{x}$ with at most p-1 components equal to 1 violates the

inequality corresponding to that particular row of A^*, whose p-1 entries equal to 0 include those positions where $\bar{x}_j = 0$. Thus $z_{C(A^*)} = p$, i.e., $A^* \in \mathscr{A}^n(p)$.

Next we show that $z_{F(A^*)} = n/(n-p+1)$. Let $k = n-p+1$, and let $\tilde{x}$ be defined by $\tilde{x}_j = 1/k$, $j = 1,\dots,n$. Let B be any $n \times n$ nonsingular submatrix of A^*, such that every column of B has exactly k entries equal to 1. The definition of A^* guarantees the existence of B. Now let $\tilde{u}$ be the q-vector defined by $\tilde{u}_i = 1/k$ if the i^{th} row of A^* is a row of B, $\tilde{u}_i = 0$ otherwise. Then $\tilde{x}$ and $\tilde{u}$ are feasible solutions to the linear program $\min\{e_n x \mid A^*x \geq e_q,\ x \geq 0\}$ and its dual, respectively, with value $e_n\tilde{x} = e_q\tilde{u} = n/k$. Hence $\tilde{x}$ is an optimal fractional cover, and $z_{F(A^*)} = e_n\tilde{x} = n/(n-p+1)$.

Finally, we show that A^* minimizes $z_{F(A)}$ over $\mathscr{A}^n(p)$. Assume this to be false, and let A^o be a matrix that minimizes $z_{F(A)}$ over $\mathscr{A}^n(p)$, with $z_{F(A^o)} < z_{F(A^*)}$. Also, let $A^* = (a^*_{ij})$, $A^o = (a^o_{ij})$. W.l.o.g., we may assume that A^o has n columns, since adding columns whose entries are all equal to 0 does not change either the integer or the fractional optimum. For every $S \subset \{1,\dots,n\}$ such that $|S| = p-1$, A^o has a row i such that $a^o_{ij} = 0$, $\forall\, j \in S$; or else $\hat{x}$ defined by $\hat{x}_j = 1$, $j \in S$, $\hat{x}_j = 0$, $j \notin S$, would be a cover with value p-1, contrary to the assumption that $A^o \in \mathscr{A}^n(p)$. Hence for every row i of A^*, A^o has a row k such that $a^o_{kj} \leq a^*_{ij}$, $j = 1,\dots,n$. But then $x \geq 0$, $A^o x \geq e_r$ implies $A^*x \geq e_q$ (where r is the number of rows of A^o), hence $z_{F(A^*)} \leq z_{F(A^o)}$, a contradiction.||

<u>Theorem 2</u>. For any $A \in \mathscr{A}^n$,

$$\text{(5)} \qquad \frac{z_{C(A)}}{z_{F(A)}} \leq \frac{1}{n}\left\lfloor\frac{n+1}{2}\right\rfloor\left\lceil\frac{n+1}{2}\right\rceil,$$

and this is a best possible bound.

<u>Proof</u>. For fixed $p \epsilon \{1,\ldots,n\}$, from Theorem 1

$$\max_{A \epsilon \mathscr{A}^n(p)} \frac{z_{C(A)}}{z_{F(A)}} = \frac{p}{n}(n-p+1). \tag{6}$$

If p is allowed to vary continuously in the interval [1, n], the right hand side of (6) is concave and attains its maximum for p = (n+1)/2. Since p has to be integer, the maximum is attained either for $p = \left\lfloor \frac{n+1}{2} \right\rfloor$, or for $p = \left\lceil \frac{n+1}{2} \right\rceil$; namely,

$$\max_{A \epsilon \mathscr{A}^n} \frac{z_{C(A)}}{z_{F(A)}} = \max \left\{ \frac{1}{n} \left\lfloor \frac{n+1}{2} \right\rfloor \left(n - \left\lfloor \frac{n+1}{2} \right\rfloor + 1\right), \frac{1}{n} \left\lceil \frac{n+1}{2} \right\rceil \left(n - \left\lceil \frac{n+1}{2} \right\rceil + 1\right) \right\}$$

$$= \frac{1}{n} \left\lfloor \frac{n+1}{2} \right\rfloor \left\lceil \frac{n+1}{2} \right\rceil . \quad \|$$

Another expression for the above bound is given by

$$\frac{1}{n} \left\lfloor \frac{n+1}{2} \right\rfloor \left\lceil \frac{n+1}{2} \right\rceil = \begin{cases} \frac{n}{4} + \frac{1}{2} & \text{if n is even} \\ \frac{n}{4} + \frac{1}{2} + \frac{1}{4n} & \text{if n is odd.} \end{cases} \tag{7}$$

Thus, the n variables set covering problem for which the ratio $z_{C(A)}/z_{F(A)}$ attains its maximum, is the one whose coefficient matrix has exactly $\left\lfloor \frac{n+1}{2} \right\rfloor$ 1's in every row, and contains as a row every binary n-vector with $\left\lfloor \frac{n+1}{2} \right\rfloor$ components equal to 1. For this problem, $z_{C(A)} = \left\lceil \frac{n+1}{2} \right\rceil$ and $z_{F(A)} = \frac{2n}{n+2-\delta}$, where $\delta = 0$ if n is even and $\delta = 1$ if n is odd.

Before concluding our paper, we compare the bound on $z_{C(A)}/z_{F(A)}$ given in Theorem 2, with the bound on $z_{G(A)}/z_{F(A)}$ given by (1). To do this, we note that when we consider the bound H(d) given by (1) for all set covering problems defined by matrices $A \epsilon \mathscr{A}^n$, the largest d that can occur (provided A has no componentwise equal rows), happens to occur for the matrix A* having as rows all possible 0-1 n-vectors with exactly $\left\lfloor \frac{n+1}{2} \right\rfloor$ components equal to 1. For this matrix, we denote $d(A^*) = d^*$, and we have

$$d^* = \binom{n-1}{\left\lfloor \frac{n+1}{2} \right\rfloor - 1} = \binom{n-1}{\left\lfloor \frac{n-1}{2} \right\rfloor} .$$

We want to assess the value of the ratio

$$R = \frac{1 + \ln d^*}{\frac{1}{n}\left\lfloor \frac{n+1}{2} \right\rfloor \left\lceil \frac{n+1}{2} \right\rceil} \tag{8}$$

Theorem 3. For $n \geq 2$,

$$R > 4\, \frac{n-1}{n+1} \ln\left(2\, \frac{n-1}{n}\right) . \tag{9}$$

Proof. From (8), we have

$$R = \frac{n}{\left\lfloor \frac{n+1}{2} \right\rfloor \left\lceil \frac{n+1}{2} \right\rceil} \left[1 + \ln \binom{n-1}{\left\lfloor \frac{n-1}{2} \right\rfloor} \right] . \tag{10}$$

Using Stirling's formula as refined by Robbins,

$$q^q e^{-q} (2\pi q)^{1/2} e^{1/(12q+1)} < q! < q^q e^{-q} (2\pi q)^{1/2} e^{1/12q} ,$$

we have

$$\binom{n-1}{\left\lfloor\frac{n-1}{2}\right\rfloor} = \frac{(n-1)!}{\left\lfloor\frac{n-1}{2}\right\rfloor!\left\lceil\frac{n-1}{2}\right\rceil!}$$

$$> \frac{(n-1)^{n-1}\cdot e^{1-n}\cdot[2\pi(n-1)]^{1/2}\cdot e^{\alpha}}{\left\lfloor\frac{n-1}{2}\right\rfloor^{\left\lfloor\frac{n-1}{2}\right\rfloor}\cdot\left\lceil\frac{n-1}{2}\right\rceil^{\left\lceil\frac{n-1}{2}\right\rceil}\cdot e^{-\left\lfloor\frac{n-1}{2}\right\rfloor-\left\lceil\frac{n-1}{2}\right\rceil}\cdot\left(4\pi^2\left\lfloor\frac{n-1}{2}\right\rfloor\left\lceil\frac{n-1}{2}\right\rceil\right)^{1/2}\cdot e^{\beta+\gamma}}$$

$$= \left(\frac{n-1}{\left\lfloor\frac{n-1}{2}\right\rfloor}\right)^{\left\lfloor\frac{n-1}{2}\right\rfloor}\cdot\left(\frac{n-1}{\left\lceil\frac{n-1}{2}\right\rceil}\right)^{\left\lceil\frac{n-1}{2}\right\rceil}\cdot\left(\frac{n-1}{2\pi\left\lfloor\frac{n-1}{2}\right\rfloor\left\lceil\frac{n-1}{2}\right\rceil}\right)^{1/2}\cdot e^{\alpha-\beta-\gamma}$$

where

$$\alpha = \frac{1}{12(n-1)+1}, \qquad \beta = \frac{1}{12\left\lfloor\frac{n-1}{2}\right\rfloor}, \qquad \gamma = \frac{1}{12\left\lceil\frac{n-1}{2}\right\rceil}.$$

Thus

$$\ln\binom{n-1}{\left\lfloor\frac{n-1}{2}\right\rfloor} > \left\lfloor\frac{n-1}{2}\right\rfloor \ln\frac{n-1}{\left\lfloor\frac{n-1}{2}\right\rfloor} + \left\lceil\frac{n-1}{2}\right\rceil \ln\frac{n-1}{\left\lceil\frac{n-1}{2}\right\rceil} + \frac{1}{2}\ln\frac{n-1}{2\pi\left\lfloor\frac{n-1}{2}\right\rfloor\left\lceil\frac{n-1}{2}\right\rceil} + \alpha-\beta-\gamma ,$$

and therefore, using (10),

$$R > \frac{n\left\lfloor\frac{n-1}{2}\right\rfloor}{\left\lfloor\frac{n+1}{2}\right\rfloor\left\lceil\frac{n+1}{2}\right\rceil}\ln\frac{n-1}{\left\lfloor\frac{n-1}{2}\right\rfloor} + \frac{n\left\lceil\frac{n-1}{2}\right\rceil}{\left\lfloor\frac{n+1}{2}\right\rfloor\left\lceil\frac{n+1}{2}\right\rceil}\ln\frac{n-1}{\left\lceil\frac{n-1}{2}\right\rceil} + \frac{n}{\left\lfloor\frac{n+1}{2}\right\rfloor\left\lceil\frac{n+1}{2}\right\rceil}\delta ,$$

where

$$\delta = 1 + \frac{1}{2}\ln\frac{2}{\pi(n-1)} + \alpha - \beta - \gamma ,$$

and we have used the fact that $n(n-2) < (n-1)^2$ for $n \geq 2$.

Using $\left\lfloor \frac{n-1}{2} \right\rfloor + \left\lceil \frac{n-1}{2} \right\rceil = n-1$ and $\ln \frac{n-1}{\left\lfloor \frac{n-1}{2} \right\rfloor} \geq \ln \frac{n-1}{\left\lceil \frac{n-1}{2} \right\rceil}$,

we obtain

$$(11) \qquad R > \frac{(n+1)(n-1)}{\left\lfloor \frac{n+1}{2} \right\rfloor \left\lceil \frac{n+1}{2} \right\rceil} \ln \frac{n-1}{\left\lceil \frac{n-1}{2} \right\rceil} + \frac{n}{\left\lfloor \frac{n+1}{2} \right\rfloor \left\lceil \frac{n+1}{2} \right\rceil} \left(\delta - \frac{n-1}{n} \ln \frac{n-1}{\left\lceil \frac{n-1}{2} \right\rceil} \right) .$$

As the last term is nonnegative for $n \geq 2$, and

$$\left\lfloor \frac{n+1}{2} \right\rfloor \left\lceil \frac{n+1}{2} \right\rceil \leq \frac{(n+1)^2}{4} , \qquad \left\lceil \frac{n-1}{2} \right\rceil \leq \frac{n}{2} ,$$

inequality (11) implies (9). ||

The value of the righthand side in (9) is 2.5 for n = 20, and it approaches the constant $4 \ln 2 \sim 2.769$ as n goes to infinity. Thus for the problems for which $d = d^*$, the bound on z_C/z_F is about 1/2.7 of the bound on z_G/z_F.

References

[1] E. Balas and A. Ho, "Set Covering Algorithms Using Cutting Planes, Heuristics and Subgradient Optimization." Mathematical Programming Study 12, 1980, p. 37-60.

[2] V. Chvátal, "A Greedy Heuristic for the Set Covering Problem." Mathematics of Operations Research, 4, 1979, p. 233-235.

[3] A. Ho, "Worst Case Analysis of a Class of Set Covering Heuristics." GSIA, Carnegie-Mellon University, June 1979.

[4] D. Hochbaum, "Approximation Algorithms for the Weighted Set Covering and Node Cover Problems." GSIA, Carnegie-Mellon University, April 1980.

[5] D. S. Johnson, "Approximation Algorithms for Combinatorial Problems." J. Comput. System Sci., 9, 1974, p. 256-278.

[6] L. Lovász, "On the Ratio of Optimal Integral and Fractional Covers." Discrete Mathematics, 13, 1975, p. 383-390.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: MSRR #475	2. GOVT ACCESSION NO.: AD-A104 839
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Optimal Integer and Fractional Covers: A Sharp Bound on Their Ratio	5. TYPE OF REPORT & PERIOD COVERED: Technical Report May 1981
	6. PERFORMING ORG. REPORT NUMBER: MSRR 475
7. AUTHOR(s): Egon Balas	8. CONTRACT OR GRANT NUMBER(s): N00014-75-C-0621 NR 047-048
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, Pa. 15213	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Personnel and Training Research Programs, Office of Naval Research (Code 434), Arlington, Va. 22217	12. REPORT DATE: May 1981
	13. NUMBER OF PAGES: 8
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Set covering, integer programming, worst case analysis

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The ratio of the values of optimal integer and fractional solutions to a set covering problem was shown by Johnson [5] and Lovász [6] to be bounded by $B(d) = 1 + \ln d$, where d is the largest column sum. We show that if n is the number of variables, $B(n) = \frac{1}{n}\left\lfloor\frac{n+1}{2}\right\rfloor\left\lceil\frac{n+1}{2}\right\rceil$ is a best possible bound on this ratio. Furthermore, for every $n \geq 20$ there are problems for which $B(n) \leq \frac{1}{2.5}B(d)$.