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. Reproduction in whole or in part is permitted for any purpose of the U.S. Government.

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 \( B(d) = 1 + 2n d \), where \( d \) is the largest column sum. We show that if \( n \) is the number of variables, \( B(n) = \frac{1}{n} \left\lfloor \frac{n+1}{2} \right\rfloor \) is the best possible bound on this ratio. Furthermore, for every \( n \geq 20 \) there are problems for which \( B(n) \leq \frac{1}{2.5} B(d) \).
OPTIMAL INTEGER AND FRACTIONAL COVERS: A SHARP BOUND ON THEIR RATIO

by

Egon Balas

The simple (unweighted) set covering problem is

\[ z_C = \min \{ e_n^\top x | Ax \geq e_m, \ x \text{ binary} \}, \]

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

\[ z_F = \min \{ e_n^\top x | Ax \geq e_m, x \geq 0 \}. \]

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

\[ \frac{z_G}{z_F} \leq H(d) (\leq 1 + 2n d), \]
where \( z_G \) is the value of a cover obtained by the greedy heuristic,

\[
d = \max_{j \in \{1, \ldots, m\}} \sum_{i=1}^{m} a_{ij},
\]

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).

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

\[
(2) \quad \frac{z_G}{z_C} \leq H(d)
\]
and

\[ \frac{z_C}{z_F} \leq H(d). \]

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

\[ \mathcal{A}^n(p) = \{ A \in \mathcal{A}^n | z_C(A) = p \}. \]

**Theorem 1.** For any positive integer \(n\) and any \(p \in \{1, \ldots, n\}\),

\[ \min_{A \in \mathcal{A}^n(p)} z_F(A) = \frac{n}{n-p+1}, \]

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.

**Proof.** 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 \(A^* x \geq e_q\), where \(q = \binom{n}{k}\), since no row of \(A^*\) has more than \(p-1\) entries equal to 0. Further, every binary \(n\)-vector \(x\) 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 $y_j = 0$. Thus $z_C(A^*) = p$, i.e., $A^* \in \mathcal{A}^n(p)$.

Next we show that $z_F(A^*) = n/(n-p+1)$. Let $k = n-p+1$, and let $\bar{x}$ be defined by $\bar{x}_j = 1/k$, $j = 1,...,n$. 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 $\bar{u}$ be the $q$-vector defined by $\bar{u}_i = 1/k$ if the $i$th row of $A^*$ is a row of $B$, $\bar{u}_i = 0$ otherwise. Then $\bar{x}$ and $\bar{u}$ are feasible solutions to the linear program $\min \{e^\top x | A^*x \geq e_q, x \geq 0\}$ and its dual, respectively, with value $e^\top \bar{x} = e^\top \bar{u} = n/k$. Hence $\bar{x}$ is an optimal fractional cover, and $z_F(A^*) = e^\top \bar{x} = n/(n-p+1)$.

Finally, we show that $A^*$ minimizes $z_F(A)$ over $\mathcal{A}^n(p)$. Assume this to be false, and let $A^0$ be a matrix that minimizes $z_F(A)$ over $\mathcal{A}^n(p)$, with $z_F(A^0) < z_F(A^*)$. Also, let $A^* = (a^*_i)$, $A^0 = (a^0_i)$. W.l.o.g., we may assume that $A^0$ 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 $S \subset \{1,...,n\}$ such that $|S| = p-1$, $A^0$ has a row $i$ such that $a^0_{ij} = 0$, $\forall j \in S$; or else $\bar{x}$ defined by $\bar{x}_j = 1$, $j \in S$, $\bar{x}_j = 0$, $j \notin S$, would be a cover with value $p-1$, contrary to the assumption that $A^0 \in \mathcal{A}^n(p)$. Hence for every row $i$ of $A^*$, $A^0$ has a row $k$ such that $a^0_{kj} = a^*_j$, $j = 1,...,n$. But then $x \geq 0$, $A^0x \geq e_r$ implies $A^*x \geq e_q$ (where $r$ is the number of rows of $A^0$), hence $z_F(A^*) \leq z_F(A^0)$, a contradiction.

Theorem 2. For any $A \in \mathcal{A}^n$,

$$\frac{z_C(A)}{z_F(A)} \leq \frac{1}{n-2} \left\lfloor \frac{n+1}{2} \right\rfloor,$$

and this is a best possible bound.
Proof. For fixed \( p \in \{1, \ldots, n\} \), from Theorem 1

\[
\max_{A \in \mathbb{A}^n(p)} \frac{\pi(A)}{\pi_F(A)} = \frac{n}{n} (n-p+1).
\]

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 \( p = \lfloor \frac{n+1}{2} \rfloor \), or for \( p = \lceil \frac{n+1}{2} \rceil \); namely,

\[
\max_{A \in \mathbb{A}^n} \frac{\pi(A)}{\pi_F(A)} = \max \left\{ \frac{1}{n} \lfloor \frac{n+1}{2} \rfloor \left( n - \lfloor \frac{n+1}{2} \rfloor + 1 \right), \frac{1}{n} \lceil \frac{n+1}{2} \rceil \left( n - \lceil \frac{n+1}{2} \rceil + 1 \right) \right\}
\]

\[
= \frac{1}{n} \lfloor \frac{n+1}{2} \rfloor \left( \frac{n+1}{2} \right).
\]

Another expression for the above bound is given by

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

Thus, the \( n \) variables set covering problem for which the ratio \( \frac{\pi(A)}{\pi_F(A)} \) 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 \( \left\lceil \frac{n+1}{2} \right\rceil \) components equal to 1. For this problem, \( \pi(A) = \left\lfloor \frac{n+1}{2} \right\rfloor \) and \( \pi_F(A) = \frac{2n}{n+2-\delta} \), where \( \delta = 0 \) if \( n \) is even and \( \delta = 1 \) if \( n \) is odd.

Before concluding our paper, we compare the bound on \( \frac{\pi(A)}{\pi_F(A)} \) given in Theorem 2, with the bound on \( \frac{\pi(G)}{\pi_F(A)} \) 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 \( A \in \mathbb{A}^n \), 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

\[
d^* = \begin{pmatrix}
    n-1 \\
    \left\lfloor \frac{n+1}{2} \right\rfloor - 1
\end{pmatrix} = \begin{pmatrix}
    n-1 \\
    \frac{n-1}{2}
\end{pmatrix}.
\]

We want to assess the value of the ratio

\[
R = \frac{1 + \ln d^*}{\frac{1}{n+1} \frac{n+1}{n} \frac{n+1}{2}}.
\]

**Theorem 3.** For \( n \geq 2 \),

\[
R > 4 \frac{n-1}{n+1} \ln \left( 2 \frac{n-1}{n} \right).
\]

**Proof.** From (8), we have

\[
R = \frac{n}{\left\lfloor \frac{n+1}{2} \right\rfloor} \left[ 1 + \ln \left( \frac{n-1}{2} \right) \right].
\]

Using Stirling's formula as refined by Robbins,

\[
q^{q-e-(2nq)^{1/2}}e^{1/(12q+1)} < q^1 < q^{q-e-(2nq)^{1/2}}e^{1/12q},
\]

we have
\[
\binom{n-1}{\frac{n-1}{2}} \cdot \frac{(n-1)!^2}{\left(\frac{n-1}{2}\right)!^2} = \frac{(n-1)^{n-1} \cdot e^{1-n} \cdot (2n(n-1))^{1/2} \cdot e^\alpha}{\left(\frac{n-1}{2}\right)!^2 \cdot \left(\frac{n-1}{2}\right)!^2 \cdot (4n^2 \cdot \frac{n-1}{2})^{1/2} \cdot e^\beta \cdot \gamma}
\]

\[
= \left(\frac{n-1}{\frac{n-1}{2}}\right)^2 \cdot \left(\frac{n-1}{\frac{n-1}{2}}\right)^2 \cdot \left(\frac{n-1}{\frac{n-1}{2}}\right)^2 \cdot e^{\alpha \beta \gamma}
\]

where

\[
\alpha = \frac{1}{12(n-1)+1}, \quad \beta = \frac{1}{12 \cdot \frac{n-1}{2}}, \quad \gamma = \frac{1}{12 \cdot \frac{n-1}{2}}.
\]

Thus

\[
\binom{n-1}{\frac{n-1}{2}} > \left[\binom{n-1}{\frac{n-1}{2}} \cdot \binom{n-1}{\frac{n-1}{2}} \cdot \binom{n-1}{\frac{n-1}{2}} + \binom{n}{\frac{n-1}{2}} \cdot \binom{n-1}{\frac{n-1}{2}} + \frac{1}{2} \binom{n}{\frac{n-1}{2}} \cdot \binom{n-1}{\frac{n-1}{2}} + \alpha \beta \gamma
\]

and therefore, using (10),

\[
R > \frac{n}{n+1} \cdot \frac{n}{n+1} \cdot \frac{n-1}{\frac{n-1}{2}} \cdot \frac{n-1}{\frac{n-1}{2}} + \frac{n}{n+1} \cdot \frac{n}{n+1} \cdot \frac{n}{\frac{n-1}{2}} \cdot \frac{n}{\frac{n-1}{2}} + \frac{n}{n+1} \cdot \frac{n}{n+1} \cdot \frac{n-1}{\frac{n-1}{2}} \cdot \frac{n-1}{\frac{n-1}{2}} + \delta
\]

where

\[
\delta = 1 + \frac{1}{2} \binom{n}{\frac{2}{n(n-1)}} + \alpha \beta \gamma
\]

and we have used the fact that \(n(n-2) < (n-1)^2\) for \(n \geq 2\).
Using $\left\lfloor \frac{n-1}{2} \right\rfloor + \left\lfloor \frac{n-1}{2} \right\rfloor = n-1$ and $\delta n \geq n-1$, we obtain

$$R > \left( \frac{n+1}{2} \right)^2 \frac{n}{n+1} \delta n \frac{n}{n+1} + \frac{n}{n+1} \left( \frac{n-1}{2} \right) \left( \frac{n-1}{2} \right).$$

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

$$\left( \frac{n+1}{2} \right)^2 \left( \frac{n}{n+1} \right) \leq \frac{(n+1)^2}{4}, \quad \left( \frac{n-1}{2} \right) \leq \frac{n}{2},$$

inequality (11) implies (9).

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

References


Optimal Integer and Fractional Covers: A Sharp Bound on Their Ratio

May 1981

Egon Balas

Graduate School of Industrial Administration
Carnegie-Mellon University
Pittsburgh, Pa. 15213

Personnel and Training Research Programs
Office of Naval Research (Code 434)
Arlington, Va. 22217

Set covering, integer programming, worst case analysis

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 $B(d) = 1 + 2\sqrt{d}$, where $d$ is the largest column sum. We show that if $n$ is the number of variables, $B(n) = \left\lfloor \frac{n+1}{2} \right\rfloor$ is a best possible bound on this ratio. Furthermore, for every $n \geq 20$ there are problems for which $B(n) \leq \frac{1}{2.5}B(d)$. 