46/2002

Raport Badawczy

Research Report

RB/55/2002

Preliminary Results on the Average Case Analysis of the Set Packing Problem

K. Szkatuła

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 8373578

fax: (+48) (22) 8372772

Kierownik Pracowni zgłaszający pracę:
Prof. dr hab. inż. Krzysztof Kiwiel

Warszawa 2002

Preliminary results on the average case analysis of the set packing problem

Krzysztof SZKATUŁA
Systems Research Institute, Polish Academy of Sciences
ul. Newelska 6, 01-447 Warszawa, Poland
E-mail: Krzysztof.Szkatula@ibspan.waw.pl

January 6, 2003

Abstract

In the paper we deal with the well known set packing problem. It is assumed that some of the problem coefficients are realizations of mutually independent random variables. Certain probablistic properties of selected problem characteristics are investigated for the variety of possible instances of the problem.

1 Introduction

Let us consider a set packing problem formulated as the binary multiconstraint knapsack problem, see Nemhauser and Wolsey [5]:

$$\begin{array}{ll} z_{OPT}(n) = \max \sum_{i=1}^{n} c_i \cdot x_i & \\ \text{subject to} \quad \sum_{i=1}^{n} a_{ji} \cdot x_i \leqslant 1 & \\ \text{where} \quad j = 1, ..., m, \quad x_i = 0 \text{ or } 1 & \end{array} \tag{1}$$

It is assumed that:

$$c_i > 0,\ a_{ji} = 0 \text{ or } 1,\ i = 1, \ldots, n,\ j = 1, \ldots, m.$$

Set packing problem (1) is well known to be is well known to be $\mathcal{NP}$ hard, see Garey and Johnson [2]. Although set packing problem may be formulated as the binary multiconstraint knapsack problem, it is rather special case of it, see Martello and Toth [3]. Its peculiarity consists in 2 facts:

- All the constraints left hand sides coefficients are equal either to 1 or to 0, i.e.

$$a_{ji} = 0 \text{ or } 1,\ i = 1, \ldots, n,\ j = 1, \ldots, m.$$

- All of the constraints right hand sides coefficients are equal to 1.

In the general formulation of the binary multiconstraint knapsack problem it is only required that all of the knapsack problem coefficients, i.e. goal function, constraints left and right hand sides, are non-negative or, in order to avoid unclear interpretations, strictly positive. It especially applies to goal function and constraints right hand sides coefficients.

2 Definitions

The following definitions are necessary for the further presentation:

Definition 1 *We denote $V_n \approx Y_n$, where $n \to \infty$, if*

$$Y_n \cdot (1 - o(1)) \leqslant V_n \leqslant Y_n \cdot (1 + o(1))$$

when V_n, Y_n are sequences of numbers, or

$$\lim_{n \to \infty} P\{Y_n \cdot (1 - o(1)) \leqslant V_n \leqslant Y_n \cdot (1 + o(1))\} = 1$$

when V_n is a sequence of random variables and Y_n is a sequence of numbers or random variables, where $\lim_{n \to \infty} o(1) = 0$ as usual.

Definition 2 *We denote $V_n \preceq Y_n (V_n \succeq W_n)$ if*

$$V_n \leqslant (1 + o(1)) \cdot Y_n \ (V_n \geqslant (1 - o(1)) \cdot W_n)$$

when V_n, Y_n (W_n) are sequences of numbers, or

$$\lim_{n \to \infty} P\{V_n \leqslant (1 + o(1)) \cdot Y_n\} = 1 \ (\lim_{n \to \infty} P\{V_n \geqslant (1 - o(1)) \cdot W_n\} = 1)$$

when V_n is a sequence of random variables and Y_n (W_n) is a sequence of numbers or random variables, where $\lim_{n \to \infty} o(1) = 0$.

Definition 3 *We denote $V_n \cong Y_n$ if there exist constants $c'' \geqslant c' > 0$ such that*

$$c' \cdot Y_n \preceq V_n \preceq c'' \cdot Y_n$$

where Y_n, V_n are sequences of numbers or random variables.

The following random model of (1) will be considered in the paper:

- m, n are arbitrary positive integers, $n \to \infty$, $i = 1, \ldots, n$, $j = 1, \ldots, m$.
- c_i, a_{ji} are realizations of mutually independent random variables and moreover c_i, are uniformly distributed over $(0, 1]$ and $P\{a_{ji} = 1\} = p$, where $0 < p \leq 1$.

Under the assumptions made about c_i, a_{ji}, and taking into account (??) the following always hold

$$0 \leqslant z_{OPT}(n) \leqslant \sum_{i=1}^{n} c_i \leqslant n, \tag{2}$$

Moreover, from the strong law of large numbers it follows that

$$\sum_{i=1}^{n} c_i \approx E(c_1) \cdot n = n/2, \ \sum_{i=1}^{n} a_{ji} \approx p \cdot n.$$

Therefore, it is justified to enhance formula (2) in the following way:

$$0 \leqslant z_{OPT}(n) \preceq n/2, \ \sum_{i=1}^{n} a_{ji} \preceq 1, \text{ if } p < \frac{1}{n} \text{ or } \sum_{i=1}^{n} a_{ji} \succeq 1 \text{ when } p > \frac{1}{n}. \quad (3)$$

Formula (3) shows that random model of set packing problem (1) is complete in the sense that nearly all possible instances of the problem are considered.

The growth of $z_{OPT}(n)$ - value of the optimal solution of the problem (1) may be influenced by the problem coefficients, namely:

$$n, \ m, \ c_i, \ a_{ji}, \text{ where } i = 1, \dots, n, \ j = 1, \dots, m.$$

We have assumed that c_i, a_{ji} are realizations of the random variables and therefore their impact on the $z_{OPT}(n)$ growth is in this case indirect. Moreover, we have assumed that m, n are arbitrary fixed positive integers and $n \to \infty$. The aim of the probabilistic analysis is to investigate asymptotic behaviour of $z_{OPT}(n)$ when $n \to \infty$.

3 Lagrange and dual estimations

When we consider the knapsack problem, with on or many constraints, the Lagrange function and the problem dual to, see Averbakh [1], Meanti, Rinnooy Kan, Stougie and Vercellis [4], Szkatuła [6] and [7] is very useful tool to perform various kind of analyses. In the case of set packing problem Lagrange function of the problem (1) may be formulated as follows:

$$\begin{aligned} L_n(x) &= \sum_{i=1}^{n} c_i \cdot x_i + \sum_{j=1}^{m} \lambda_j \cdot \left(1 - \sum_{i=1}^{n} a_{ji} \cdot x_i\right) = \\ &= \sum_{j=1}^{m} \lambda_j + \sum_{i=1}^{n} \left(c_i - \sum_{j=1}^{m} \lambda_j \cdot a_{ji}\right) \cdot x_i \end{aligned}$$

where $x = [x_1, \dots, x_n]$ and $\Lambda = [\lambda_1, \dots, \lambda_m]$ - vector of Lagrange multipliers. Moreover, let for every Λ, $\lambda_j \geq 0$, $j = 1, \dots, m$:

$$\phi_n(\Lambda) = \max_{x \in \{0,1\}^n} L_n(x, \Lambda) = \max_{x \in \{0,1\}^n} \left\{ \sum_{j=1}^{m} \lambda_j + \sum_{i=1}^{n} \left(c_i - \sum_{j=1}^{m} \lambda_j a_{ji}\right) x_i \right\}.$$

Taking the following notation:

$$x_i(\Lambda) = \begin{cases} 1 & \text{if } c_i - \sum\limits_{j=1}^{m} \lambda_j \cdot a_{ji} > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

$$c_i(\Lambda) = \begin{cases} c_i & \text{if } c_i - \sum\limits_{j=1}^{m} \lambda_j \cdot a_{ji} > 0 \\ 0 & \text{otherwise.} \end{cases}$$

$$a_{ji}(\Lambda) = \begin{cases} a_{ji} & \text{if } c_i - \sum\limits_{j=1}^{m} \lambda_j \cdot a_{ji} > 0 \\ 0 & \text{otherwise.} \end{cases}$$

we have for every Λ, $\lambda_j \geq 0$, $j = 1, \dots, m$:

$$\begin{aligned} \phi_n(\Lambda) &= \sum_{j=1}^{m} \lambda_j + \sum_{i=1}^{n} \left(c_i - \sum_{j=1}^{m} \lambda_j \cdot a_{ji} \right) \cdot x_i(\Lambda) = \\ &= \sum_{j=1}^{m} \lambda_j + \sum_{i=1}^{n} \left(c_i(\Lambda) - \sum_{j=1}^{m} \lambda_j \cdot a_{ji}(\Lambda) \right) \end{aligned}$$

Obviously

$$c_i(\Lambda) = c_i \cdot x_i(\Lambda), \quad a_{ji}(\Lambda) = a_{ji} \cdot x_i(\Lambda).$$

Problem dual to set packing problem (1) maybe formulated as follows:

$$\Phi_n^* = \min_{\Lambda \geq 0} \phi_n(\Lambda). \tag{5}$$

For every $\Lambda \geq 0$ the following holds:

$$z_{OPT}(n) \leq \Phi_n^* \leq \phi_n(\Lambda).$$

Let us denote:

$$\begin{aligned} z_n(\Lambda) &= \sum_{i=1}^{n} c_i \cdot x_i(\Lambda) = \sum_{i=1}^{n} c_i(\Lambda), s_j(\Lambda) = \sum_{i=1}^{n} a_{ji} \cdot x_i(\Lambda) = \sum_{i=1}^{n} a_{ji}(\Lambda), \\ S_{nm}(\Lambda) &= \sum_{j=1}^{m} \lambda_j \cdot s_j(\Lambda), \; \tilde{\Lambda}(m) = \sum_{j=1}^{m} \lambda_j. \end{aligned}$$

By definition of $c_i(\Lambda)$ and $a_{ji}(\Lambda)$, see also (4), we have:

$$c_i(\Lambda) \geq \sum_{j=1}^{m} \lambda_j \cdot a_{ji}(\Lambda)$$

and therefore

$$z_n(\Lambda) \geq S_{nm}(\Lambda). \tag{6}$$

For certain Λ, $x_i(\Lambda)$ given by (4) may provide feasible solution of (1), i.e.:

$$s_j(\Lambda) \leq 1 \quad \text{for every} \quad j = 1, \ldots, m. \tag{7}$$

Then:

$$z_n(\Lambda) \leq z_{OPT}(n) \leq \Phi_n^* \leq \phi_n(\Lambda) = z_n(\Lambda) + \tilde{\Lambda}(m) - S_{nm}(\Lambda).$$

If (7) holds, then the below inequality also holds:

$$\tilde{\Lambda}(m) - S_{nm}(\Lambda) \geq 0.$$

From (6) we get:

$$\frac{\phi_n(\Lambda)}{z_n(\Lambda)} = \frac{z_n(\Lambda)}{z_n(\Lambda)} + \frac{\tilde{\Lambda}(m) - S_{nm}(\Lambda)}{z_n(\Lambda)} \leq 1 + \frac{\tilde{\Lambda}(m) - S_{nm}(\Lambda)}{S_{nm}(\Lambda)}.$$

Therefore if (7) holds, then the following inequality also holds:

$$1 \leq \frac{z_{OPT}(n)}{z_n(\Lambda)} \leq \frac{\Phi_n^*}{z_n(\Lambda)} \leq \frac{\phi_n(\Lambda)}{z_n(\Lambda)} \leq \frac{\tilde{\Lambda}(m)}{S_{nm}(\Lambda)}. \tag{8}$$

Formula (8) shows, that if there exits such a set of Lagrange multipliers $\Lambda(n)$, fulfilling the formula (7) and if the formula below holds:

$$\lim_{n \to \infty} \frac{\tilde{\Lambda}(m)}{S_{nm}(\Lambda(n))} = 1 \tag{9}$$

then $x_i(\Lambda(n))$, $i = 1, \ldots, n$, given by (4), is the asymtotically sub-optimal solution of the set packing problem (1). Moreover the value of $z_n(\Lambda(n))$ is an asymptotical approximation of the optimal solution value of the set packing problem i.e. $z_{OPT}(n)$.

4 Probabilistic analysis

In the present section of the paper some probablistic properties of the set packing problem (1) will be investigated. Let us observe that due to the assumptions made the following holds, for $i = 1, \ldots, n$, $j = 1, \ldots, m$:

$$\begin{aligned} P\{a_{ji} &= 1\} = p, \; P\{a_{ji} = 0\} = 1 - p, \; P\{a_{ji}(\Lambda) = 1\} = 1 - P\{a_{ji}(\Lambda) = 0\}, \\ P(c_i &< x) = \begin{cases} 0 & \text{when } x \leqslant 0 \\ x & \text{when } 0 < x \leqslant 1 \\ 1 & \text{when } x \geqslant 1 \end{cases}. \end{aligned} \tag{10}$$

Moreover for the random variable $\sum_{k=1, k \neq j}^{m} a_{ji}$, due to the binomial distribution, the following holds for every r - integer, $0 \leqslant r \leqslant m - 1$:

$$P\left\{ \sum_{k=1, k \neq j}^{m} a_{ki} = r \right\} = \binom{m-1}{r} \cdot p^r \cdot (1-p)^{m-r-1}. \tag{11}$$

Let us also assume that

$$\Lambda = \{\lambda, \cdots, \lambda\}, \text{ i.e. } \lambda_j = \lambda, \; j = 1, \cdots, m.$$

Lemma 1 *If a_{ji} are realizations of mutually independent random variables where $P\{a_{ji}=1\}=p$, $0<p\leq 1$, then*

$$P\{a_{ji}(\Lambda)=1\}=p-p\sum_{r=0}^{m-1}\binom{m-1}{r}\cdot p^r\cdot(1-p)^{m-r-1}\min\{1,\lambda(r+1)\}.$$

If, moreover, $\lambda\leqslant 1/m$ then:

$$P\{a_{ji}(\Lambda)=1\}=p\cdot(1-\lambda\cdot(m\cdot p+1-p)).$$

Proof. From (4), (10) and (11) and taking into account that random variable $\sum_{k=1,k\neq j}^{m}a_{ji}$ may take any integer value r from the range $[0,m-1]$ with the probability given in (11) it follows that:

$$\begin{aligned}
P\{a_{ji}(\Lambda)=0\} &= P\left\{a_{ji}=0\cup a_{ji}=1\cap c_i<\lambda\cdot\left(\sum_{k=1,k\neq j}^{m}a_{ji}+1\right)\right\}= \\
&= 1-p+p\cdot P\left\{c_i<\lambda\cdot\left(\sum_{k=1,k\neq j}^{m}a_{ji}+1\right)\right\}= \\
&= 1-p+p\sum_{r=0}^{m-1}\binom{m-1}{r}\cdot p^r\cdot(1-p)^{m-r-1}\min\{1,\lambda(r+1)\}.
\end{aligned}$$

Due to the (10) it proves the first formula of the Lemma. When $\lambda\leqslant 1/m$ then the following holds

$$P\{a_{ji}(\Lambda)=0\}=1-p+\lambda\sum_{r=0}^{m-1}\frac{(m-1)!\cdot(r+1)}{r!\cdot(m-1-r)!}\cdot p^{r+1}\cdot(1-p)^{m-r-1} \tag{12}$$

Let us observe that for every integers l, m, $l,>1$, $m\geqslant 2$, and $0\leqslant p\leqslant 1$ the following hold

$$\begin{aligned}
\sum_{k=0}^{l}\binom{l}{k}\cdot p^k\cdot(1-p)^{l-k} &= (p+1-p)^l=1 \\
r+1 &= m-(m-1-r).
\end{aligned}$$

Using the above mentioned formulas (12) may be rewritten as:

$$\begin{aligned}
P\{a_{ji}(\Lambda)=0\} &= 1-p+\lambda\cdot p\left(\sum_{r=0}^{m-1}\frac{(m-1)!\cdot m}{r!\cdot(m-1-r)!}\cdot p^r\cdot(1-p)^{m-1-r}-\right. \\
&\quad \left.-\sum_{r=0}^{m-1}\frac{(m-1)!\cdot(m-1-r)}{r!\cdot(m-1-r)!}\cdot p^r\cdot(1-p)^{m-1-r}\right)= \\
&= 1-p+\lambda\cdot p\left(m\sum_{r=0}^{m-1}\binom{m-1}{r}\cdot p^r\cdot(1-p)^{m-1-r}-\right. \\
&\quad \left.-p\cdot(m-1)\cdot(1-p)\sum_{r=0}^{m-2}\binom{m-2}{r}\cdot p^r\cdot(1-p)^{m-2-r}\right)= \\
&= 1-p+\lambda\cdot p\cdot(m-(m-1)\cdot(1-p))= \\
&= 1-p+\lambda\cdot p\cdot(m\cdot p+1-p).
\end{aligned}$$

Finally above formulas can be summarized as:

$$P\{a_{ji}(\Lambda)=0\}=1-p+\lambda\cdot p\cdot(m\cdot p+1-p). \tag{13}$$

Due to the formulas (10) and (13) we have

$$\begin{aligned} P\{a_{ji}(\Lambda)=1\} &= 1-P\{a_{ji}(\Lambda)=0\}= \\ &= p-\lambda\cdot p\cdot(m\cdot p+1-p)=p\cdot(1-\lambda\cdot(m\cdot p+1-p)). \end{aligned}$$

■

As the direct consequence of the above formulas we have

$$E(a_{ji}(\Lambda))=1\cdot P\{a_{ji}(\Lambda)=1\}+0\cdot P\{a_{ji}(\Lambda)=0\}=P\{a_{ji}(\Lambda)=1\}. \tag{14}$$

Now instead of Λ we will consider $\Lambda(n)$. It does mean that for every value of integer n, we may consider different vector $\Lambda(n)=\{\lambda(n),\cdots,\lambda(n)\}$.
For every j, $j=1,\cdots,m$, we have:

$$\begin{aligned} E(s_j(\Lambda(n))) &= \sum_{i=1}^{n} E(a_{ji}(\Lambda(n)))=n\cdot P\{a_{ji}(\Lambda(n))=1\}= \\ &= n\cdot p(1-\lambda(n)\cdot(m\cdot p+1-p)). \end{aligned} \tag{15}$$

Lemma 2 *The following choice of* $\lambda(n)$, *where* $\alpha>0$:

$$\lambda(n)=\frac{1-\alpha/(n\cdot p)}{m+1-p} \text{ is solving the equation } E(s_j(\Lambda(n)))=\alpha.$$

Corollary 1 *If* $E(s_j(\Lambda(n)))=\alpha$, *then* $P\{a_{ji}(\Lambda(n))=1\}=\alpha/n$.

Proof. Proof of Lemma and Corollary follows immediately from formulas (14) and (15). ■

Solution of the set packing problem (1) given by formula (4) is feasible if and only if the formula (7) holds.

Proposition 1 *For the* $\Lambda(n)$, *providing* $E(s_j(\Lambda(n)))=\alpha$, $\alpha>0$, *the following hold*

$$P\{s_j(\Lambda(n))\leqslant 1\}=\left(1-\frac{\alpha}{n}\right)^{n-1}\cdot(2-\frac{\alpha}{n})$$

Proof. As it was already mentioned solution of problem (1) given by formula (4) is feasible if and only if formula (7) holds i.e. $s_j(\Lambda(n))=0$ or $s_j(\Lambda(n))=1$. For every $\Lambda(n)$, random variable $s_j(\Lambda(n))=\sum_{i=1}^{n}a_{ji}(\Lambda(n))$ may take any integer value r from the range $[0,n]$ with the probability given by the following formula:

$$P\left\{\sum_{i=1}^{n}a_{ji}(\Lambda(n))=r\right\}=\binom{n}{r}\cdot\tilde{p}^{r}\cdot(1-\tilde{p})^{n-r}, \text{ where } \tilde{p}=P\{a_{ji}(\Lambda(n))=1\}.$$

From the above formula and Corollary 1 it follows that

$$\begin{aligned} P\{s_j(\Lambda(n))\leqslant 1\} &= P\left\{\sum_{i=1}^{n}a_{ji}(\Lambda(n))=0\cup\sum_{i=1}^{n}a_{ji}(\Lambda(n))=1\right\}= \\ &= \left(1-\frac{\alpha}{n}\right)^{n}+\alpha\left(1-\frac{\alpha}{n}\right)^{n-1}=\left(1-\frac{\alpha}{n}\right)^{n-1}\cdot(1+\alpha-\frac{\alpha}{n}) \end{aligned} \tag{16}$$

■

Corollary 2 *If* $\alpha = 1$ *then*

$$P\{s_j(\Lambda(n)) \leqslant 1\} \approx \frac{2}{e} \tag{17}$$

Proof. Formula (17) follows immediately from the (16) and from the fact that $\left(1 - \frac{\alpha}{n}\right)^{n-1} \approx \frac{1}{e}$. ∎

5 Concluding remarks

In the present report some very preliminary results describing probablistic properties of the set packing problem (1) are summarized.

In the paper distribution functions of the various random variables representing important problems characteristics are presented. Moreover some results concerning the feasibility of the received solutions are obtained.

Important hints for the future research is convergence of the approximate solutions to the optimal solution and possibility of investigating their values.

References

[1] I. Averbakh. Probabilistic properties of the dual structure of the multidimensional knapsack problem and fast statistically effcient algorithms. *Mathematical Programming*, 65:311–330, 1994.

[2] M. Garey and D. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* Freeman, San Francisco, 1979.

[3] S. Martello and P. Toth. *Knapsack Problems: Algorithms and Computer Implementations.* Wiley & Sons, 1990.

[4] M. Meanti, A. R. Kan, L. Stougie, and C. Vercellis. A probabilistic analysis of the multiknapsack value function. *Mathematical Programming*, 46:237–247, 1990.

[5] G. Nemhauser and L. Wolsey. *Integer and Combinatorial Optimization.* John Wiley & Sons Inc., New York, 1988.

[6] K. Szkatuła. On the growth of multi-constraint random knapsacks with various right-hand sides of the constraints. *European Journal of Operational Reserch*, 73:199–204, 1994.

[7] K. Szkatuła. The growth of multi-constraint random knapsacks with large right-hand sides of the constraints. *Operations Research Letters*, 21:25–30, 1997.