VARIABLE DELETION IN MULTI-DIMENSIONAL MULTIPLE-CHOICE KNAPSACK PROBLEMS

JOHN WALKER

Nanyang Business School

Nanyang Technological University, Singapore 639798

Abstract

The multi-dimensional multiple-choice knapsack problem is an integer programming problem with n zero-one variables, r resource constraints, and m mutually disjoint multiple-choice constraints. The importance of the problem stems from the ease with which the appropriate specification of its parameters allows for the formulation of several important integer programming problems having widespread application. The concept of dominance allows the identification of variables which can be deleted from the problem prior to the solution phase. The expected proportion of undominated variables is shown to be a function of r, the number of resource constraints, and c, the cardinality of the multiple-choice sets. Explicit closed form solutions are given for $r = 1, 2$, and an approximation, for fixed r and large c. For general r and c, a recursive formula is given and results tabulated for a region of interest. For small values of r the proportion of undominated variables is small.

Keywords: multi-dimensional multiple-choice knapsack problem, dominance, analytical analysis, proportion of undominated variables.

1. Introduction

The multi-dimensional multiple-choice knapsack (MDMCK) problem is an integer programming (IP) problem with n zero-one variables, r resource constraints, and m mutually disjoint multiple-choice (MC) constraints. It is defined in the following manner.

$$\text{Maximise} \quad \sum_{j \in N} a_{r+1j} x_j \quad ,$$

subject to:

$$\begin{aligned} \sum_{j \in N} a_{ij} x_j &\leq b_i \quad , \quad i = 1, \ldots , r \quad , \\ \sum_{j \in N_k} x_j &= 1 \quad , \quad k = 1, \ldots , m \quad , \\ x_j &\in \{0,1\} \quad , \quad j \in N \quad , \end{aligned}$$

where $N = \{1, \ldots , n\} = \bigcup_{k=1}^{m} N_k$ and $N_p \bigcap N_q = \emptyset$ for $1 \leq p < q \leq m$.

The importance of the MDMCK problem stems from the ease with which the appropriate specification of its parameters allows for the formulation of several important IP problems having widespread application. It is easily shown that the following IP problems are specialisations of the MDMCK problem: binary knapsack (BK) problem, multi-dimensional binary knapsack (MDBK) problem, bounded integer knapsack (BIK) problem, multi-dimensional bounded integer knapsack (MDBIK) problem, multiple-choice knapsack (MCK) problem, separable bounded integer programming (SBIP)problem and several variations of the generalised assignment problem (GAP). Rather than attempt to provide a comprehensive account of such problems and their associated applications the reader is referred to Nemhauser and Wolsey (1988). However, it will be convenient to note that the above problems can be classified into two categories. The first category contains problems in which the number of resource constraints, r, must be small. The BK, BIK and MCK problems, each having $r = 1$, fall into this category. The second category contains problems that may have, but do not necessarily have to have, a small number of resource constraints. The MDBK, MDBIK, SBIP and GAP problems fall in this category.

2. Dominance

There are, of course, many methods for solving MDMCK and its associated specialisations. The reader is again referred to Nemhauser and Wolsey (1988) for a comprehensive account. Whatever method is adopted, it is normally desirable to attempt to “reduce” the size of the problem by deleting appropriate variables prior to the solution phase. The concept of dominance allows the identification of those variables which can be deleted from the MDMCK problem. Johnston and Khan (1995) derived several results concerning the use of dominance in unbounded knapsack problems. The approach adopted in this paper is to derive expressions to determine the expected proportion of undominated variables within a MC set of a MDMCK problem. The approach will be justified under the assumption that objective function and resource constraint coefficients of the variables within a MC set are independently drawn from *any* continuous distribution function. The expected proportion of undominated variables is shown to be a function of r, the number of resource constraints, and c, the cardinality of the multiple-choice sets. Explicit closed form solutions are given for r = 1, 2, and an approximation, for fixed r and large c. For general r and c, a recursive formula is given and results tabulated for a region of interest. For small values of r the proportion of undominated variables is small.

Definition A variable x_s, $s \in N_k$, is said to be dominated if there exists a variable x_t, $t \in N_k$, such that:

$$\begin{aligned} &\text{(i)} \quad a_{it} \leq a_{is} \quad \text{, for all } i = 1, \ldots , r, \text{ and} \\ &\text{(ii)} \quad a_{r+1t} \geq a_{r+1s}. \end{aligned} \qquad (1)$$

The variable x_s, $s \in N_k$, is said to be undominated if and only if there does not exist a variable x_t, $t \in N_k$, such that relationship (1) holds.

Property The property of dominance which allows for the deletion of appropriate variables is given as follows. If variable x_s, is dominated by variable x_t then:

(i) if MDMCK is infeasible with $x_t = 1$, there exists no feasible solution with $x_s = 1$,

(ii) if MDMCK is feasible with $x_t = 1$, there exists an optimal solution with $x_s = 0$.

3. Expected Proportion Of Undominated Variables

In the following analysis the vector $(a_{1j}, \ldots , a_{rj}, a_{r+1j})$ of coefficients of x_j, $j \in N_k$, is regarded as being a point in R^{r+1}, and each coefficient a_{ij}, $j \in N_k$, is independently selected from a distribution having a density function f_i and distribution function F_i for $i = 1, \ldots , r$. The cardinality of the set N_k is denoted by $c = |N_k|$.

Corresponding to the definition of variable dominance, the point a_s, $s \in N_k$, is said to be undominated if and only if there does not exist a point a_t, $t \in N_k$, such that relationship (1) holds.

$$\text{Prob}[a_s \text{ is not dominated by } a_t] \quad = 1 - \text{Prob}[a_{1t} \leq a_{1s}]\ldots\text{Prob}[a_{rt} \leq a_{rs}].\text{Prob}[a_{r+1t} \geq a_{r+1s}]$$

Since the points in N_k are independently selected:

$$\text{Prob}[a_s \text{ is undominated in } N_k] \quad = [1 - F_1(a_{1s})\ldots F_r(a_{rs})(1-F_{r+1s}(a_{r+1s}))]^{c-1} \tag{2}$$

For $r \geq 0$ and $c \geq 1$, let $P_{r,c}$ denote the Prob[any point $(\mathsf{a}_1, \ldots, \mathsf{a}_r, \mathsf{a}_{r+1})$ in N_k is undominated]. $P_{r,c}$ can be interpreted as the proportion of undominated variables in N_k.

For $r \geq 0$ and $c = 1$, by definition:

$$\mathrm{P}_{\mathrm{r},1} \quad = \quad 1.0 \qquad , \mathrm{r} \geq 0 \tag{3}$$

For $r \geq 0$ and $c \geq 2$, the use of law of total probability on (2) gives:

$$P_{r,c} \quad = \quad \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [1 - F_1(\mathsf{a}_1)\ldots F_r(\mathsf{a}_r)(1-F_{r+1}(\mathsf{a}_{r+1}))]^{c-1} f_1(\mathsf{a}_1)$$

$$\ldots f_r(\mathsf{a}_r) f_{r+1}(\mathsf{a}_{r+1}) d\mathsf{a}_1 \ldots d\mathsf{a}_r d\mathsf{a}_{r+1} \tag{4}$$

Let $A_i = F_i(\mathsf{a}_i)$, $i = 1, \ldots, r$, $A_{r+1} = 1 - F_{r+1}(\mathsf{a}_{r+1})$, and substituting into (4):

$$P_{r,c} \quad = \quad \int_0^1 \ldots \int_0^1 \int_0^1 [1 - A_1 \ldots A_r A_{r+1}]^{c-1} dA_1 \ldots dA_r dA_{r+1} \qquad , r \geq 0,\ c \geq 2 \tag{5}$$

Integrating (5) with respect to A_r:

$$P_{r,c} \quad = \quad \frac{1}{c} \int_0^1 \ldots \int_0^1 \int_0^1 \left[\frac{1 - [1 - A_1 \ldots A_r A_{r+1}]^c}{-cA_1 \ldots A_{r-1} A_{r+1}} \right]_0^1 dA_1 \ldots dA_{r-1} dA_{r+1}$$

$$P_{r,c} \quad = \quad \frac{1}{c} \int_0^1 \ldots \int_0^1 \int_0^1 \frac{1 - [1 - A_1 \ldots A_{r-1} A_{r+1}]^c}{A_1 \ldots A_{r-1} A_{r+1}} dA_1 \ldots dA_{r-1} dA_{r+1} \tag{6}$$

Consider $1 - [1 - A_1 \ldots A_{r-1} A_{r+1}]^c \quad = \quad \sum_{t=0}^{c-1} (1 - A_1 \ldots A_{r-1} A_{r+1})^t (1 - (1 - A_1 \ldots A_{r-1} A_{r+1}))$ and substitute into (6):

$$\begin{aligned} P_{r,c} \quad = \quad \frac{1}{c} \int_0^1 \ldots \int_0^1 \int_0^1 & [1 + (1 - A_1 \ldots A_{r-1} A_{r+1}) + (1 - A_1 \ldots A_{r-1} A_{r+1})^2 + \ldots \\ & + (1 - A_1 \ldots A_{r-1} A_{r+1})^{c-1}] dA_1 \ldots dA_{r-1} dA_{r+1} \end{aligned}$$

Therefore, from (5)

$$P_{r,c} \quad = \quad \frac{1}{c} (P_{r-1,1} + P_{r-1,2} + \ldots + P_{r-1,c}) \text{ , and} \tag{7}$$

$$P_{r,c-1} = \quad \frac{1}{c-1} (P_{r-1,1} + P_{r-1,2} + \ldots + P_{r-1,c-1}) \tag{8}$$

Multiplying (7) and (8) by c and $-(c-1)$ respectively and adding:

$$P_{r,c} \quad = \quad \frac{1}{c}\,[(c\text{-}1)P_{r,c\text{-}1} + P_{r\text{-}1,c}] \qquad , r \geq 1,\ c \geq 2 \qquad (9)$$

Special Case 1: *r* = 1

From (5) when $r = 0$:

$$P_{0,c} \quad = \quad \int_0^1 \ [1 - A_{r+1}]^{c\text{-}1} dA_{r+1}$$

$$= \quad \left[\frac{(1-A_{r+1})^c}{-c}\right]_0^1$$

$$P_{0,c} \quad = \quad \frac{1}{c} \qquad , \text{c} \geq 2 \qquad (10)$$

From (7) when $r = 1$:

$$\text{P}_{1,c} \quad = \quad \frac{1}{c}\,(P_{0,1} + P_{0,2} + \ldots + P_{0,c})$$

Using (3) and (10):

$$P_{1,c} \quad = \quad \frac{1}{c}\,(1 + \frac{1}{2} + \ldots + \frac{1}{c})$$

$$P_{1,c} \quad = \quad \frac{1}{c}\,H_c \qquad , c \geq 2 \qquad (11)$$

H_c is the Harmonic number, see Greene and Knuth (1981),

$$H_c \quad = \quad \sum_{s=1}^{c} \frac{1}{s} \quad = \quad \log c + \gamma + \frac{1}{2c} - \frac{1}{12c^2} + O(c^{-4}) \text{ where } \gamma \quad \cong \quad 0.57722$$

Special Case 2: *r* = 2

From (7) when $r = 2$:

$$P_{2,c} \quad = \quad \frac{1}{c}\,(P_{1,1} + P_{1,2} + \ldots + P_{1,c})$$

Using (3) and (11):

$$P_{2,c} \quad = \quad \frac{1}{c}\,(1 + \frac{1}{2}\,(1+\frac{1}{2}) + \ldots + \frac{1}{c}\,(1+\frac{1}{2} + \ldots + \frac{1}{c}))$$

$$P_{2,c} \quad = \quad \frac{1}{c}(\sum_{s=1}^{c} \frac{1}{s^2} + \sum_{s=1}^{c-1} \sum_{t=s+1}^{c} \frac{1}{s}\frac{1}{t}) \tag{12}$$

Consider $H_c^2 \quad = \quad (\sum_{s=1}^{c} \frac{1}{s})^2 = \sum_{s=1}^{c} \frac{1}{s^2} + 2\sum_{s=1}^{c-1} \sum_{t=s+1}^{c} \frac{1}{s}\frac{1}{t}$.

Hence $\sum_{s=1}^{c-1} \sum_{t=s+1}^{c} \frac{1}{s}\frac{1}{t} = \frac{1}{2}(H_2^c - \sum_{s=1}^{c} \frac{1}{s^2})$ and substituting into (12):

$$P_{2,c} \quad = \quad \frac{1}{2c}(H_c^{[2]} + H_c^2) \qquad , c \geq 2 \tag{13}$$

where $H_c^{[2]} \quad = \quad \sum_{s=1}^{c} \frac{1}{s^2}$

Special Case 3: Fixed *r*, large *c*

Proposition For fixed r and large c:

$$P_{r,c} \cong \frac{(\log c\}^{r}}{cr!} \tag{14}$$

Proof The proposition can be justified heuristically as follows.

For $r = 0$

$$P_{0,c} \cong \frac{(\log c\}^{0}}{c0!} = \frac{1}{c}$$ which, from (3) and (10) is true.

Suppose the proposition is true for (r-1).

$$P_{r\text{-}1,c} \cong \frac{(\log c\}^{r-1}}{c(r-1)!}$$

To show it is also true for r, from (7)

$$P_{r,c} = \frac{1}{c}(P_{r\text{-}1,1} + P_{r\text{-}1,2} + \ldots + P_{r\text{-}1,c})$$

$$\cong \frac{1}{c}\int_{1}^{c} P_{r-1,s}\,ds$$

$$\cong \frac{1}{c}\int_{1}^{c} \frac{\{\log s)^{r-1}}{s(r-1)!}\,ds$$

$$\cong \frac{1}{c(r-1)!}\left[\frac{(\log s)^{r}}{r}\right]_{1}^{c}$$

$$\cong \frac{(\log c\}^{r}}{cr!}$$ as required

Tabulated Results For A Region of Interest

The recursive equation (9) and approximation (14) provide a simple means for evaluating $P_{r,c}$. for regions of interest. Such a range is displayed in Table 1. It can be seen that:

(i) for fixed *r*, the expected proportion of undominated variables decreases with increasing *c*; and

(ii) for fixed *c*; the expected proportion of undominated variables increases with increasing *r*.

		r										
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
c	**1**	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2	.500	.750	.875	.938	.969	.984	.992	.996	.998	.999	1.000
	3	.333	.611	.787	.887	.942	.970	.985	.992	.996	.998	.999
	4	.250	.521	.720	.846	.918	.957	.978	.989	.994	.997	.999
	5	.200	.457	.668	.810	.896	.945	.971	.985	.992	.996	.998
	6	.167	.408	.624	.779	.877	.933	.965	.982	.991	.995	.998
	7	.143	.370	.588	.752	.859	.923	.959	.979	.989	.994	.997
	8	.125	.340	.557	.728	.842	.913	.953	.975	.987	.993	.997
	9	.111	.314	.530	.706	.827	.903	.948	.972	.986	.993	.996
	10	.100	.293	.506	.686	.813	.894	.942	.969	.984	.992	.996
	12	.083	.259	.466	.651	.787	.877	.932	.964	.981	.990	.995
	14	.071	.232	.434	.621	.765	.862	.923	.958	.978	.988	.994
	16	.063	.211	.407	.595	.744	.848	.914	.953	.975	.987	.993
	18	.056	.194	.384	.572	.726	.835	.905	.948	.972	.985	.992
	20	.050	.180	.363	.552	.709	.823	.897	.943	.969	.984	.992
	30	.033	.133	.293	.475	.641	.772	.863	.921	.956	.976	.988
	40	.025	.107	.249	.423	.592	.732	.834	.903	.945	.970	.984
	50	.020	.090	.219	.385	.554	.700	.810	.886	.935	.964	.980
	60	.017	.078	.196	.355	.523	.673	.789	.872	.925	.958	.977
	70	.014	.069	.178	.331	.497	.649	.771	.858	.916	.953	.974
	80	.013	.062	.164	.311	.475	.629	.754	.846	.908	.948	.971
	90	.011	.056	.153	.294	.456	.610	.739	.835	.901	.943	.968
	100	.010	.052	.143	.279	.439	.594	.725	.825	.894	.938	.965
	120	.008	.045	.127	.255	.410	.565	.701	.806	.880	.930	.960
	140	.007	.039	.115	.236	.386	.541	.680	.789	.869	.922	.955
	160	.006	.035	.105	.220	.366	.521	.661	.774	.858	.914	.950
	180	.006	.032	.097	.207	.349	.502	.644	.761	.848	.907	.946
	200	.005	.029	.090	.195	.334	.486	.629	.748	.838	.901	.942

Table 1. The Proportion Of Undominated Variables

For additional emphasis, Table 1 is partitioned into regions where $P_{r,c} \leq 0.6$ and $P_{r,c} \leq 0.9$. For small values of r the proportion of undominated variables is small. For the category 1 MDMCK specialisations, i.e., the BK, BIK and MCK problems each having r =1, the proportion of undominated variables is very small. For the category 2 problems, Table 1 allows the analyst to determine, for given values of r and c, if the testing for dominance is likely to produce an appreciable reduction in the size of the MDMCK problem. Note that the results are valid when the coefficients are independently drawn from *any* continuous distribution.

References

Greene, D. H. and D. E. Knuth (1981), *Mathematics for the Analysis of Algorithms*, Birkhauser, Boston.

Johnston, R. E. and L .R. Khan (1995), A note on dominance in unbounded knapsack problems, *Asia Pacific Journal of Operational Research* 12, 145-160.

Nemhauser, G. L. and L. A. Wolsey (1988), *Integer and Combinatorial Optimisation*, Wiley, New York.