22nd International Congress on Modelling and Simulation, Hobart, Tasmania, Australia, 3 to 8 December 2017
mssanz.org.au/modsim2017

On the Set-Union Budget Scenario Problem

J Jagiello and R Taylor

Joint Warfare Mathematical Science
Joint and Operational Analysis Division
Defence Science and Technology Group
Canberra ACT
Email: jerzy.jagiello@dst.defence.gov.au

Abstract: At MODSIM 2015 (Taylor, 2015) formalised the *Budget Scenario Problem* as a simplified mathematical formulation of the problem presented by (Order, 2007, 2009). In the Budget Scenario Problem a list of initiatives is provided each with an anticipated cost. Each initiative is scored against a number of scenarios with a value indicating how useful the initiative is against that scenario. For a collection of initiatives the total score is calculated as a sum of best initiative scores within the collection for each scenario. In this paper we extend the mathematical model to take account of the dependency conditions expressed in words (see Table 2) in (Order, 2007, 2009) and to also accommodate the representation of synergies between initiatives. Since this extension is analogous to the Set-Union Knapsack Problem (Goldsmith, 1994) we call this formulation the Set-Union Budget Scenario Problem. In mathematical terms the Set-Union Budget Scenario Problem can be expressed as:

$$Maximise \sum_{i=1}^{m} s_i \max_{1\le j\le n}\left(v_{ij}(\alpha_j + \sum_{k=1}^{n} \beta_{jk}x_k)x_j \right), \tag{1}$$

$$\text{Subject to } \sum_{j=1}^{n} c_j x_j \le B, \tag{2}$$

where $x_j \in \{0,1\}$ is the decision to include initiative j or not, s_i is the probability of scenario i, v_{ij} is the value of score for scenario i and initiative j, α_j is the measure of independence for initiative j, β_{jk} is the value of interdependency between initiative j and k, c_j is the cost of initiative j and B is the cost bound.

The above formulation was tested on (Order, 2009) data but extended to different levels of dependency with results for dependency levels ranging from independent (Dependency Level = 0) to fully dependent (Dependency Level = 4) presented below in Figure 1.

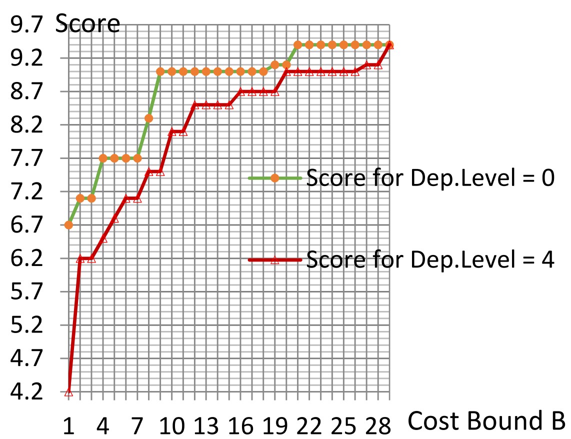

Figure 1. Score vs Cost Bound for Dependency Levels of 0 and 4

Results obtained from formulation (1) and (2) confirm the results produced by (Order, 2009) with some exceptions where (Order, 2009) appears to have not taken into account some dependencies (for example “B needs K”). The proposed formal approach has been applied to the interdependency context successfully and the obtained results are very encouraging and applicable to the force design domain.

Keywords: *Set-union knapsack problem, budget optimisation, decision making, interdependency*

1. INTRODUCTION

In two articles (Order, 2007, 2009), a problem is presented and analysed through a series of fictional discussions between clients and analysts. What appears at first to be a prioritization problem is eventually seen to be a budget-value problem in which the best value for a range of budgets is sought. A list of initiatives is provided each with an anticipated cost. Each initiative is scored against a number of scenarios with a value of 1-10 indicating how useful the initiative is against that scenario. As well each scenario is assigned a probability s_i indicating the likelihood that that scenario is occurring at any given time. Table 1 represents the Initiative Scenario Table (reproduced from (Order, 2009)). Conditions are also specified in the text (Order, 2009) whereby to be effective some initiatives require other initiatives to be present. This is shown here in Table 2.

Table 1. Order (2009) Initiative Scenario

Initiative	Cost	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Probability		0.3	0.1	0.4	0.2
O	1	10	1	8	2
M	1	7	1	3	4
N	3	6	3	5	6
D	4	8	7	1	2
J	7	4	3	7	9
I	6	8	7	4	1
B	8	8	8	3	10
L	11	10	5	6	9
E	13	8	4	9	7
A	12	7	2	9	1
H	14	1	5	9	9
G	17	9	5	8	4
C	14	5	7	5	7
K	20	6	4	9	4
F	16	1	7	9	1

Table 2. (Order, 2009) Requirements for Dependency

Scenario	Dependency
Scenario 1	O needs K B needs K
Scenario 3	J needs B,E,G L needs B,E,G H needs B,E,G

For a collection of initiatives the total value is calculated by summing the product of each scenario probability by the best value obtained by any initiative for that scenario in the collection. This initiative can be thought of as the best tool in the toolbox (collection) for the particular job (scenario), while the total value reflects the expected ability of the toolbox (initiative collection) to address any single job (scenario). Thus the initiatives O, M, and N have a total cost of 1 + 1 + 3 = 5 and a total value of 0.3 * 10 + 0.1 * 3 + 0.4 * 8 + 0.2 * 6 = 7.7. This situation can be modelled as a weighted bipartite graph G(X, Y) with initiatives X and scenarios Y in which there are weights c_j (costs) associated with each $j \in X$, and weights v_{ij} (values) associated with each edge *ij*; $i \in Y$; $j \in X$.

2. SET-UNION BUDGET SCENARIO (SUBS) PROBLEM DEFINITION

Given a budget bound *B* the Set-Union Budget Scenario Problem can be stated as:
Instance: A set *X* of items with a weight (cost) c_j associated with each item *j*; a collection of scenarios Y; weights v_{ij} (values) associated with each edge *ij*; $i \in Y$; $j \in X$; a cost bound *B*. *B* as well as all costs and weights are non-negative reals.

Optimization: Find a subset *S* of *X* with maximum total value, whose total cost is bounded by *B* taking into account dependencies.

Dependency is expressed for each scenario by the *n* x *n* dependency matrix *D* where *n* is the number of initiatives. The value of each element d_{lj} of the dependency matrix varies from zero to four, representing the level of dependency that initiative *l* has on initiative *j*. A value of zero implies that the initiative *l* is entirely independent of initiative *j*. An non-zero value of d_{lj} implies that initiative *l* has a certain non-zero level of dependency on initiative *j*. Table 3 shows the dependency values described by (Christensen, 2011), and matching independent capability values α with linear scaling between 0 and 1.

Table 3. Dependency and Independent Capability Levels

Dependency Name	Dependency Value (d)	Value α
Mandatory	4	0.0
Dependant	3	0.25
Associated	2	0.5
Tangential	1	0.75
Independent	0	1.0

Assuming that an initiative *l* can achieve its independent capability at the level α_l in the scale between 0 to 1 of its capability v_l then the remaining part of the full capability reflected in this scale is (1 - α_l) and will be shared between interdependent initiatives. If we assume proportional distribution of dependency for each initiative *j* from initiative *l* than the sharing formula proposed by (Wang, 2017) will be as follows:

$$\beta_{lj} = (1-\alpha_l)\left(\frac{d_{lj}}{\sum_{k=1}^{n} d_{lk}}\right). \tag{3}$$

If we express dependency as a continuous function instead of a discreet one than the sharing formula can be described as

$$\beta_{lj} = (1-\alpha_l)\left(\frac{\alpha_{lj}}{\sum_{k=1}^{n} \alpha_{lk}}\right) \tag{4}$$

where α_l , α_{lj} and α_{lk} are values obtained from a continuous function for appropriate $d_{\max}$, d_{lj} and d_{lk} values. The continuous function can be derived by any curve fitting method, for example least squares regression, interpolation, Fourier approximation etc.

In mathematical terms the Set-Union Budget Scenario Problem can be expressed to make explicit the independent and dependent capabilities of each initiative as follows:

$$Maximise \sum_{i=1}^{m} s_i \max_{1\le j\le n}\left(v_{ij}\left(\alpha_j + \sum_{k=1}^{n}\beta_{jk}x_k\right)x_j\right) \tag{5}$$

$$\text{Subject to } \sum_{j=1}^{n} c_j x_j \quad \le \boldsymbol{B}, \tag{6}$$

where:
i is the scenario index,
m is the number of scenarios,
j is the initiative index,
n is the number of initiatives,
$x_j \in \{0,1\}$ is the decision to include initiative *j* or not,
s_i is the probability of scenario *i*,
α_j is the measure of independence for initiative *j*,
v_{ij} is the value of score for scenario *i* and initiative *j*,
β_{jk} is the value of interdependency between initiative *j* and *k* (see Table 2),
c_j is the cost of initiative *j* (see Table 1),
B is the cost bound.

The implementation was based on exhaustive search of the initiative space supplemented by the Bellman optimization principle (Mulholland, 2016). This provides exact optimal solutions across the budget range which was compared to the results of (Order, 2009). As a generalization of the Budget Scenario Problem the Set-Union Budget Scenario Problem is NP-complete and so we should not expect there to be efficient algorithms to solve this problem exactly for large *n*. This is discussed in more detail in (Taylor, 2015) where approximation results for the Budget Scenario Problem are also provided.

3. SET-UNION BUDGET SCENARIO (SUBS) RESULTS

The above formulation was tested on (Order, 2009) data and the following results, showing score vs cost bound were obtained (see Figure 2). Results obtained from our formulation confirm the results produced by (Order, 2009) with some exceptions where (Order, 2009) has not taken into account some dependencies (e.g. “B needs K” see Table 2). Sensitivity analysis shows that there is a big difference between optimal solutions for dependency level 0 (initiatives are independent) and 4 (dependency is mandatory). Optimal solutions for dependency levels 1, 2 and 3 are the same. Testing the sensitivity of optimal solutions for different types of continuous functions (linear, exponential) did not alter the results of the optimization due to the fact that these functions were constructed based on discrete values from Table 3, although this may not always be the case.

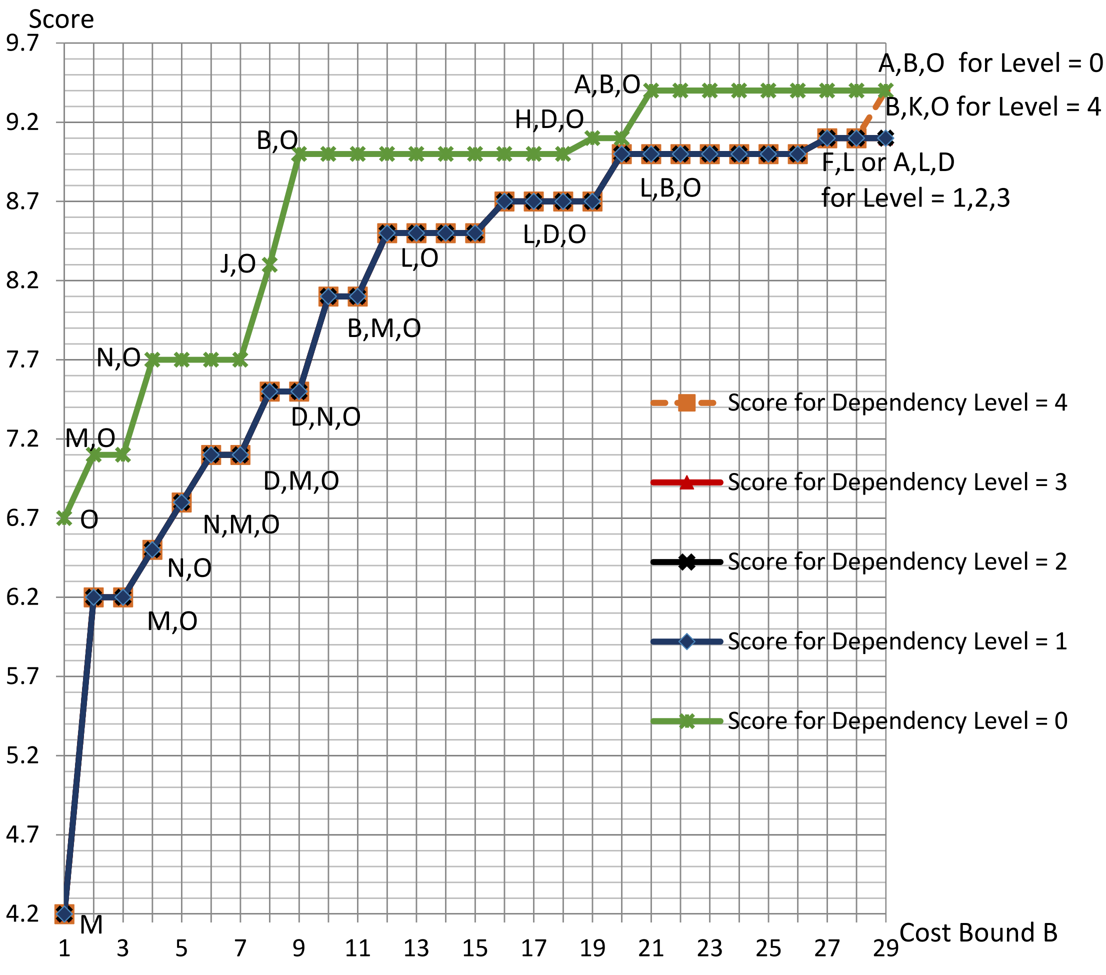

Figure 2. Score vs Cost Bound for Dependency Levels Ranging from 0 to 4

We note that there may be many initiative subsets with the same score within a given cost bound *B* – in which case we show a subset with the least cost for that score. This subset (or subsets) would correspond to the best choice of initiatives that achieve a given score. For example the subsets {F,L} and {A,L,D} produce equal scores of 9.1 with equal costs of 27, and there are no other subsets with scores at least 9.1 and costs at most 27.

If we compare Figures 2 and 3 we notice that the cost bound range in Figure 2 is a fraction of that of Figure 3. This is because optimal initiatives for cost bound 29 could not be improved further with increasing cost range, but the number of solutions with equal score grew and declined with increasing cost range up to 147. The explanation of this phenomenon is due to the fact that number of combinations is maximized for a cost of around a half of the total cost of all initiatives (in other words *n* choose *k* is maximized for *k=n/2*). Analysis of the number of local optimums as a function of the cost bound *B* presented below in Figure 3 shows that the peak is around the 70 – 80 cost bound and depends on the dependency level. Dependency level 0 and 4 produced the highest number of local optimums. We note that the number of optimums will grow exponentially with the number of initiatives and dependency levels which will impact the calculation time of this approach for very large *n*.

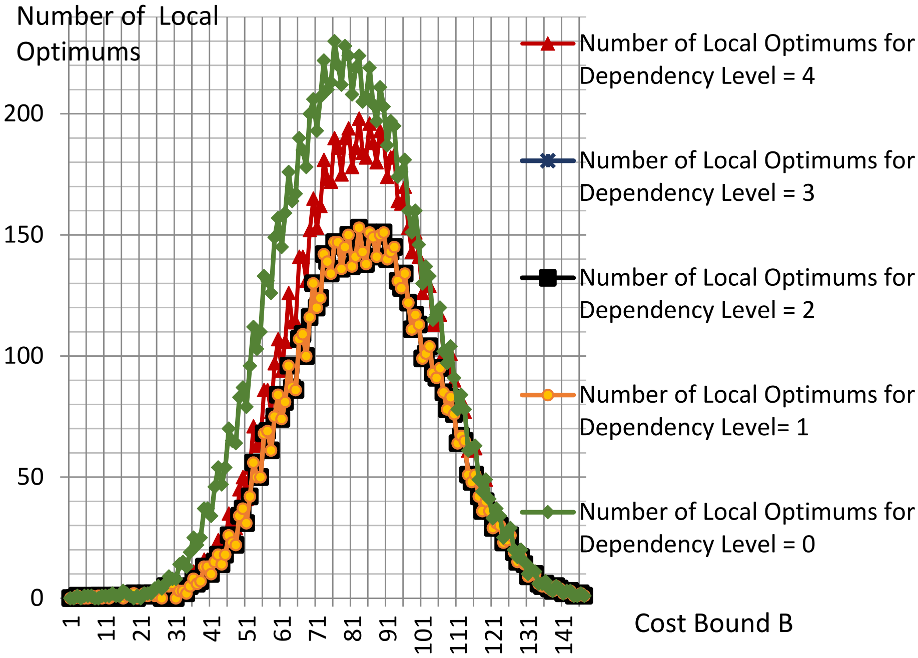

Figure 3. Number Local Optimums vs Cost Bound for Dependency Levels Ranging from 0 to 4

4. APPLICATION OF SET-UNION BUDGET SCENARIO (SUBS) PROBLEM TO THE SCME DATA

The implementation of a joint and integrated approach to the development of future Defence capability has led to an increasing interest in the dependencies and interdependencies across projects, products and programs (defined by sets of projects and products). Those dependencies and interdependencies should be addressed during the conceptual design phase in the Capability Life Cycle (CLC). In particular a project/product may depend on other projects/products to provide operational services. These may be categorised as Sensing(S), Understanding (U), Decision-making (D), Engagement (E), Physical Mobility (M), Information Mobility (I), and Logistics and support (L), or simply SUDEMIL (Lowe, 2015). These services can also be defined between programs from the projects/products they contain. Based on available dependency data for the subset of SUDEMIL services for Sensing (S), Command and Control (C), Mobility (M) and Engagement (E) (SCME) as well as cost and performance from (Order, 2009) we calculated optimum decisions for each cost bound *B*. If we compare these results with the original (Order, 2009) results we can see substantial differences due to the impact of the different dependency relationships (see Figure 4).

The score for the SCME data is generally smaller, and at most equal to the corresponding score for the (Order, 2009) data (dependency level 0). We also note the similarity between Figure 2 (Order, 2009) data and Figure 4 SCME data regarding different dependency levels.

Let the cost bound *B* range from B_0 to B_m. We can measure the difference between the two dependency curves as follows

$$integral\_distance = \int_{B_0}^{B_m} |\,(v_{dependency=0} - v_{dependency>0})\,|\,dB \tag{7}$$

$$average\_score\_distance = integral\_distance / (B_m - B_0) \tag{8}$$

Then the distance between the (Order, 2009) and SCME curves according to (7) and (8) is as follows:

- integral distance = 12.961 in score by cost units
- average score distance = 0.4469 in score units.

The integral distance 12.961 represents the area in Figure 4 between the SCME curve and Order curve and is quite significant. To see this we note that by comparison both the integral distance and the average score distance applied to Figure 2 for dependency levels equal 1, 2 and 3 are 0.

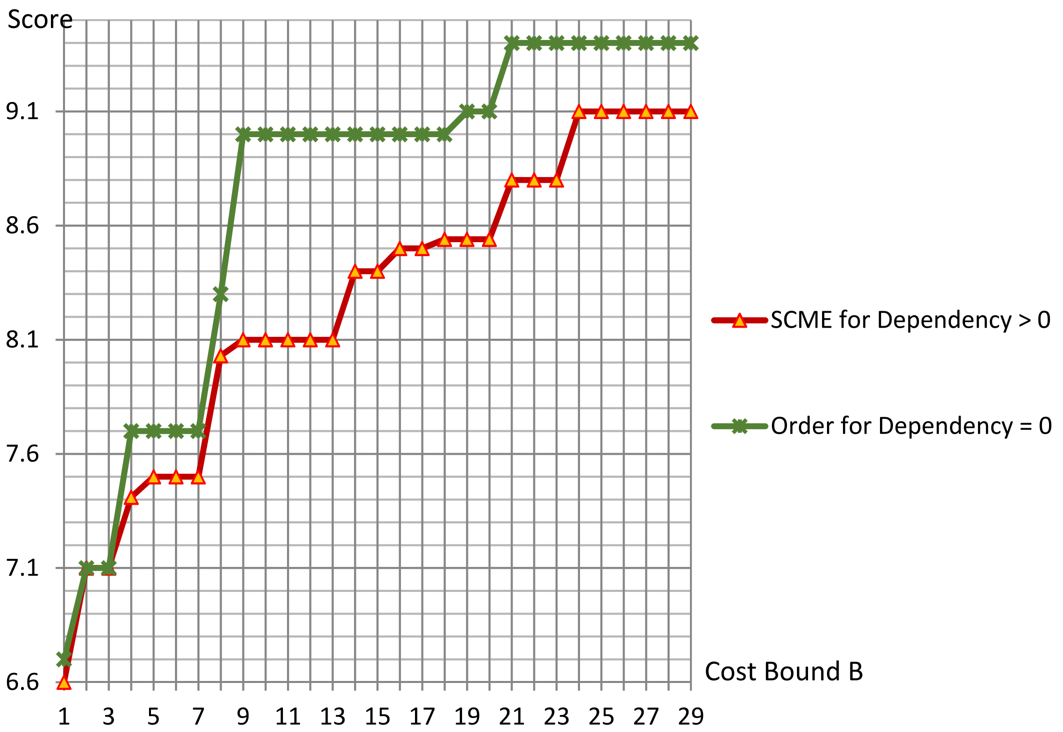

Figure 4. Comparison of SCME and Order(2009) Results

5. CONCLUSIONS

In this paper we extend the mathematical formalization given by (Taylor, 2015) of the problem described by (Order, 2007, 2009) to also encompass the dependencies described there in word form. We also show how this allows for more refined dependency data in the form of dependency levels proposed by (Christensen, 2001), as well as service based dependencies in the form of the SUDEMIL framework of (Lowe, 2015).

The novelty of this paper is the formal definition of the Budget Scenario Problem described by (Order, 2007, 2009) and the introduction of different dependency levels for both the data provided by (Order, 2009) and the SCME case study provided here.

Based on our experiments the time differences in obtaining optimal solutions for the (Order, 2009) and SCME data were negligible. Further testing might be required for more complex problems to improve the efficiency of our optimizer.

Further research will be carried out to extend the Order idea to more sophisticated optimization requirements as well as to provide alternative optimal solutions.

REFERENCES

Christensen M.B. (2011). A Method for Measuring Programmatic Dependency and Interdependency between DOD Acquisition Program, *Thesis, AFIT/GSE/ENV/11-D02DL*, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio.

Goldsmith O, Nehme D, and Yu G, (1994). Note: On the set-union knapsack problem, *Naval Research Logistics (NRL)*, 41(6), pp 833-842.

Lowe D. (2015). A quick overview of SCMILE Services and their evolution, personal correspondence, *DSTG, JOAD*.

Mulholland M. (2016), Applied Process Control, Publisher: John Wiley & Sons, ISBN: 9783527341184.

Order N. (2007). The priority list, PHALANX - *The Bulletin of Military Operations Research*, 40 (3), pp. 8-10.

Order N. (2009). The priority list - revisited, PHALANX - *The Bulletin of Military Operations Research*, 42 (4), pp. 23-34.

Taylor R. (2015). Algorithmic complexity of two defence budget problems, *MODSIM 2015*, Gold Coast, Queensland, Australia.

Wang Yue-Jin. (2017). Application of Mathematical Programming to Prioritising Interdependent Defence Investment Programs, *MODSIM2017*, Hobart, Tasmania, Australia.