〈研究論文〉

Journal of the Korean Institute of Industrial Engineers Vol. 12, No. 1, June 1986

A Shortest Path Dynamic Programming for Expansion Sequencing Problems

Sheung K. Kim*

Abstract

A shortest path dynamic programming formulation is proposed and attemped to solve an uncapacitated expansion sequencing problem. It is also compared with the Extended Binary State Space approach with total capacity. Difficulties and merits associated with the formulation are discussed. The shortest path dynamic programming lacks the separability condition and an optimal solution is not guaranteed. However it has other merits and seems to be the practical solution procedure for the expansion sequencing problem in a sense that it finds near optimal solution with less state evaluations.

1. Introduction

Expansion sequencing problem is the determination of the optimal sequencing and timing of a set of capacity expansion projects that minimize the total discounted investment cost while meeting demands projected over a planning horizon. Since several projects are available at different points of time to meet growing demand, there is a possible advantage in delaying a portion of the total investment outlays. From an economic point of view, economies of scale favor the building of large capital projects whereas a positive discount rate encourages the building of a series of small projects at earlier time stages. The expansion sequencing problem is in fact a trade-off analysis between these conflicting factors.

Recently developed solution methods for capacitated expansion sequencing problem in water resources systems include: Butcher, Haimes and Hall (1969), Morin and Esogbue (1971), Erlenkotter (1973), Tsou, Mitten and Russel (1973). Since the development of a dynamic programming (DP) formulation by Butcher, et.al. (1969), a number of studies have been made to examine the expansion sequencing problem in water supply projects. Morin and Esogbue (1971) developed the imbedded state space approach, where all the permutation schedules are considered through the state space of all pos-

* Department of Industrial Engineering, Korea University

sible cumulative capacity levels. Erlenkotter (1973) pointed out the danger of having a non-optimal solution by the conventional DP formulation and proposed the binary status vector for a project status set, in which each vector element represents the condition of the project establishment. Morin (1975) also discussed the case of yielding non-optimal solution and extended the imbedded state space approach to get solutions for more general scheduling problems. Tsou, Mitten and Russel (1973) have developed a heuristic searching method with a ranking index which is similar to the equivalent annual cost factor of Erlenkotter (1973).

Many of the previous studies dealt with capacitated expansion sequencing problems, in which capacities of individual projects are already specified and as a result, the optimum sequence and timing could be reached for a finite planning horizon. The capacitated expansion sequencing problem, however, is in fact a particular case of a general uncapacitated expansion sequencing problem with upper and lower bounds for each project capacity. For an uncapacitated expansion-sequencing problem, the project scale is another decision to be made as well as project sequencing and timing. In general, however, the project sequencing is not independent of sizing decisions. With a finite planning horizon, the interdependency between sequencing and sizing requires maintaining all the permutation schedules until the last stage has been reached. In other words, a solution procedure which is not developed over an entire policy space composed of all permutation schedules cannot be complete. The difficulties of incorporating scale decision into sequencing decision are well described by Erlenkotter (1976). Associated with a water supply expansion sequencing problem, Becker and Yeh (1974) pointed out that the firm water production of a river basin not only depends upon reservoir capacities but also upon stream configuration and reservoir operation. By incorporating reservoir operation into the spectrum, capacities of projects cannot be determined independent of previous projects chosen, which excludes the use of backward dynamic programming (DP). Hence, they incorporated the capacity decision into a forward dynamic programming formulation in conjunction with basin-wide reservoir operation simulation. Erlenkotter (1975) pointed out that the forward dynamic programming formulation may not guarantee a global optimum and proposed to include total capacity as another state variable with the binary state vector for a dynamic programming formulation. However, the proposed solution scheme still does not resolve the issue. It is because we still have to discretize the level of the total capcity for solution, and the sequencing decision depends on how the levels are discritized, and the discritized level is in fact the scale decision that we also want to find.

2. General Formulation for the Expansion Sequencing Problem

The statement of the problem which defines the multiple projects expansion sequencing problem over a finite planning horizon, T may be represented by the following general mathematical expression:

$$\underset{\substack{n \in I,\ s(i) \in S,\ t_{s(i)},\ y_{s(i)}, \\ \text{for all } i \leqslant n}}{\text{Minimize}}\ Z = \sum_{i=1}^{n} C_{s(i)}(y_{s(i)}) e^{-dt_{s(i)}} \tag{1}$$

subject to

$$t_{s(1)} = 0 \tag{2}$$

$$t_{s(i)} = D^{-1}[M(y_{s(1)}, y_{s(2)}, \ldots, y_{s(i-1)})]$$

$$D(T) = M[y_{s(1)}, y_{s(2)}, \ldots, y_{s(n)}] \quad (3)$$

$$M^{-1}_{s(i)} [D(T), y_j, j = 1, 2, \ldots n, j \neq s(i)] \leq y_{s(i)} \leq MAX\, y_{s(i)} \quad (4)$$

where

$s(i)$ = a specific project in the ith order of a permutation sequence s, which consists of n projects,

$Z[n, s(i), t_{s(i)}, y_{s(i)}]$ = total discounted investment cost of developing n projects as a function of project capacity $y_{s(i)}$, and investment timing $t_{s(i)}$, according to a project expansion schedule s(i),

S = total set of permutation schedules,

I = total set of candidate projects,

$y_{s(i)}$ = capacity of the project in the ith order of the sequence s(i),

$t_{s(i)}$ = expansion timing of the project in the ith order of the sequence s(i).

$e^{-dt_{s(i)}}$ = continuous time discount factor for discount rate d,

$C_{s(i)}[y_{s(i)}]$ = prespecified cost function of the project in the ith order of the sequence s(i) as a function of its project size $y_{s(i)}$,

$D^{-1}(.)$ = inverse demand function which shows timing,

$M[y_{s(1)}, y_{s(2)}, \ldots, y_{s(k)}]$ = a set of constraints or a simulation model operator which finds a firm water yield of the system with capacities $(y_{s(1)}, y_{s(2)}, \ldots, y_{s(i)})$.

D(T) = target demand at the year of the planning horizon T,

$M^{-1}_{s(i)} [D(T), y_j, j=1, 2, \ldots n, j \neq s(i)]$ = a model operator which provides capacity of project s(i) with given target demand and capacities of other projects given by y_j,

MAX y_i = a physical maximum capacity limit of project i,

i = index = 1, . . . , n.

Every permutation sequence determines a unique functional relationship for the total discounted investment cost. The sequence s(i), i = 1, . . . , n and the number of projects n are decisions to be made together with $t_{s(i)}$, $y_{s(i)}$, for each i. Depending upon the type of problem, the constraint set (3) - (4) could either be a resource allocation model or an operation simulation model with M as a modeling operator. In both cases, the order in which projects are undertaken and the timing of each project establishment are to be determined such that the total investment cost over a specific planning horizon is minimized. An expansion policy is defined as an expansion schedule that consists of development sequence, timing, and the corresponding project capacities.

The formulation is developed based on the following assumptions.

1. A deterministic non-decreasing demand is projected over a finite planning horizon.
2. Project costs are incurred as lump sums at the time of construction. The construction period is taken as being small. Operating costs are assumed to be small and are not taken into account.
3. Costs are discounted at a constant discount rate.
4. Each project is unique and its capacity cannot be expanded economically once the project is selected.
5. A project once constructed is operable over the finite planning horizon and only one project at a time will be constructed.

3. DP Solution Procedure with Time-Stages and the Separability.

An optimum expansion without any backstop supply would take place on the demand curve. This can easily be proved by contradiction. Suppose there is an optimum expansion path of which expansions do not take place on the demand curve but occurrs before the demand exceeds the capacity. By construction, we can find an expansion policy of which expansion takes place where the demand equals the capacity with less total discounted project cost, which is contradictory to the optimum path.

3.1 A Shortest Path Dynamic Programming

Since the optimum expansions take place on the demand curve, a simple Shortest Path Acyclic Network formulation of Bellman (1958) and Dijkstra (1959) based on time-stages can be proposed. For the network presentation, nodes are instants of capacity expansion on the predicted demand curve and the length of an arc represents the cost of capacity expansions that extends from the initial node to a terminal node of which network presentation is similar to the One Shot Dynamic Programming (OSDP) of Jacobsen (1977) for a dynamic plant location problem. Figure 1 shows the acyclic network projection of the problem. The corresponding Shortest Path Dynamic Programming (SPDP) has the recursive equation (5) - (13).

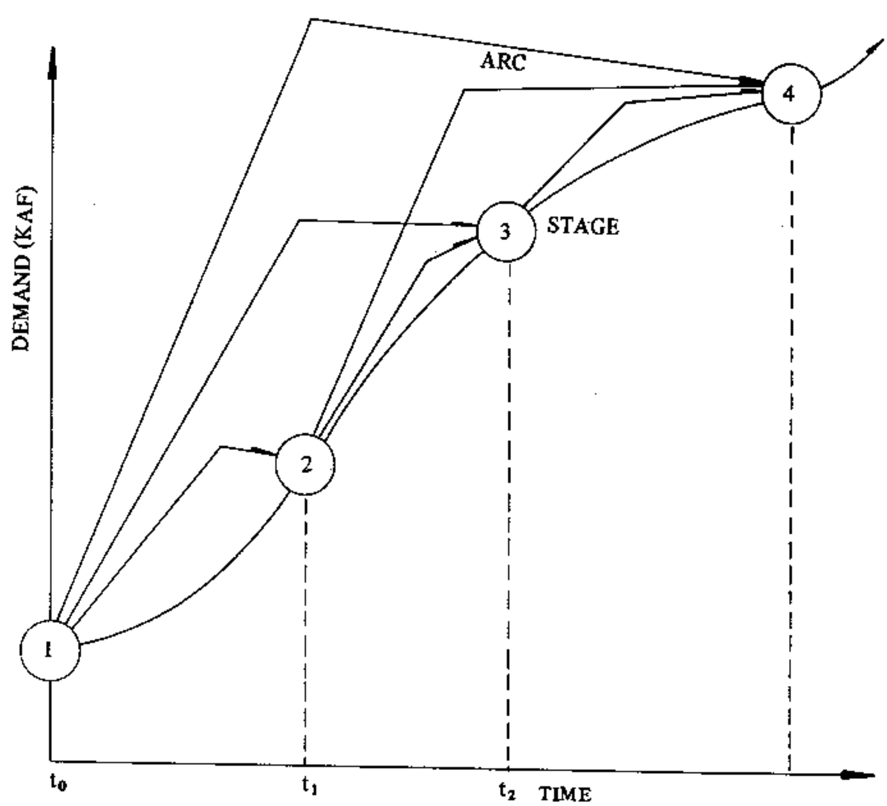

Figure 1. Acyclic Network Representation over Demand Projection

$$f_j(q_j, X_j, N_j) = \underset{i \in I}{\mathrm{Min}} \left[\underset{k(i) \in X_i}{\min} \left\{ C^{k(i)}(y_{ij}) \right\} \cdot (1+d)^{-t(i)} + f_i(q_i, X_i, N_i) \right] \quad (5)$$

subject to:

$$q_r = D[t(r)], \text{ for } r=1,2,\ldots n, \quad (6)$$

$$y_{ij} = q_j - q_i, \quad (7)$$

$$N_j = N_i + 1, \; N_i = \text{total number of projects established at stage i}, \quad (8)$$

$$X_j = X_i - k(i) \quad (\text{exclusive of } k(i) \text{ from } X_i), \quad (9)$$

$$\mathrm{MIN}\, k(i) \leq y_{ij} \leq \mathrm{MAX}\, k(i), \text{ for } k(i) \in X_i, \quad (10)$$

$$f_1(q_1, X_1, N_1) = 0, \quad (11)$$

$$t(1) = 0, \quad (12)$$

$$I = [1,2,3,\ldots j-1], \; j = 2,3,\ldots n, \quad (13)$$

$j > i$, where i, j are stages,

where

I = set of stage indices,

$f_k(q_i\ X_i, N_i)$ = Minimum total discounted investment cost of providing the demand requirement of q_i at the time stage i, with N_i number of available projects in the set of X_i.

y_{ij} = additional capacity required to meet the demand increase from time stage i to j,

q_j = total demand requirement up to the time stage j,

$k(i)$ = a specific project site considered at the time stage i,

N_i = the number of projects established at stage i,

X_i = the total set of available projects at stage i,

$t(i)$ = time of stage i,

$C^{k(i)}(y_{ij})$ = prespecified investment cost function for the project at site k(i) as a function of capacity increase from i to j,

d = discrete time discount rate,

MAX k(i) = maximum capacity that can be considered for project site k(i),

MIN k(i) = minimum capacity that can be considered for project site k(i),

3.2. Separability and Expansion Sequencing Problem.

Since optimal expansions would take place on the demand curve, it seems natural to choose time instants as expansion stages of DP formulation. Nevertheless by choosing time instants as stages for DP formulation, we are violating "Separability Condition", which is one of important conditions to fulfill the "Principle of Optimality". As a result, the sequence found by the specific formulation may not be the optimal expansion sequence.

In order to demonstrate the point, let me borrow an example from Hastings(1973). Suppose we want to climb a mountain top from a base camp and come back to the starting point. We want to find the fastest way to come back to the base camp after taking a picture at the snow covered summit. It takes 4 hours to climb with ski equipment from the base camp to the summit and 3 hours without ski. On our return, it takes half an hour from the summit to the base camp by skiing and 2 hours on foot. Figure 2 represents network diagram of the decision paths.

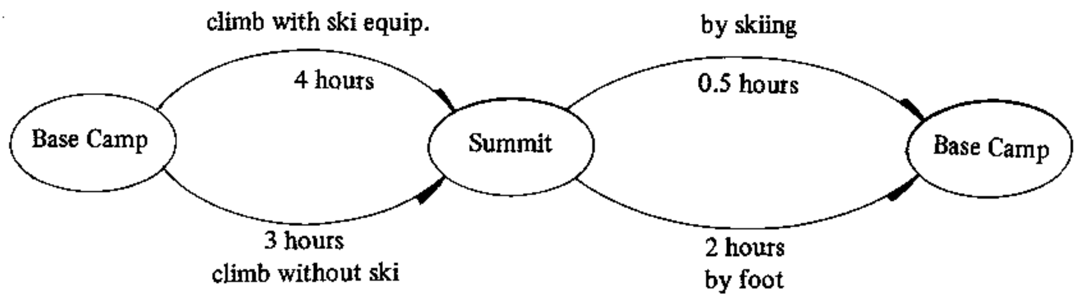

Figure 2. A Network Diagram of the Mountain Climbing Example

Applying a forward recursion, utilizing the "principle of optimality" to the network representation, the solution is to climb without ski and come back on foot, which is not the optimum solution. Climbing with ski and back by skiing takes half an hour less than the solution found. Why does it happen? This is resulted from incorrect identification of states and stages. The identification of states should be separated into independent groups so that independent decisions can be made on each state at different stages. Figure 3 shows another net-work representation that corrects the problem.

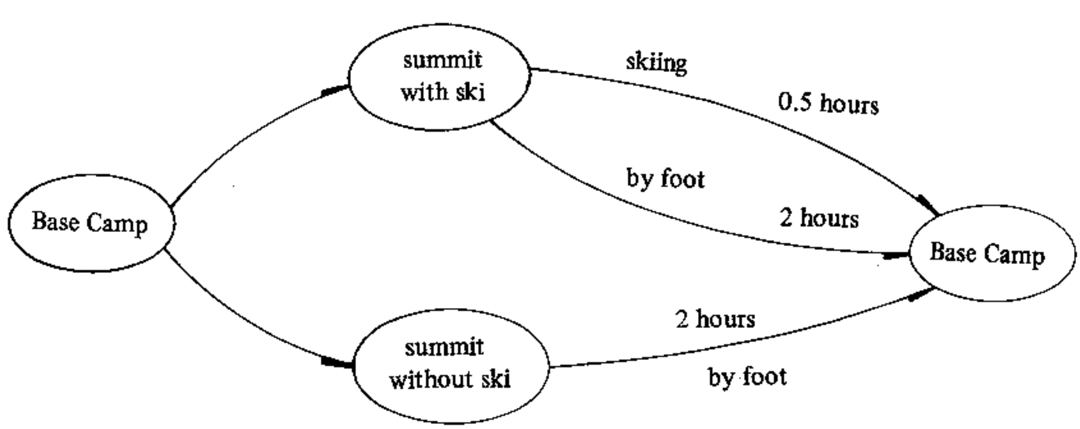

Figure 3. A Correct Network Diagram for the Mountain Climbing Example

The point we made here applies to the expansion sequencing problem whenever a DP solution procedure is formulated based on time-stages. Since an event of project selection happens at the same time axis as the progress of DP recursive frame work, if we define stages directly associating with time instants as in Figure 4, we would violate the "Separability Condition". With time-stages, a decision on project selection at one time stage excludes other permutation schedules that could have been considered if other project had been selected. This is the reason behind the Shortest Path DP formulation fails to find the optimum sequence.

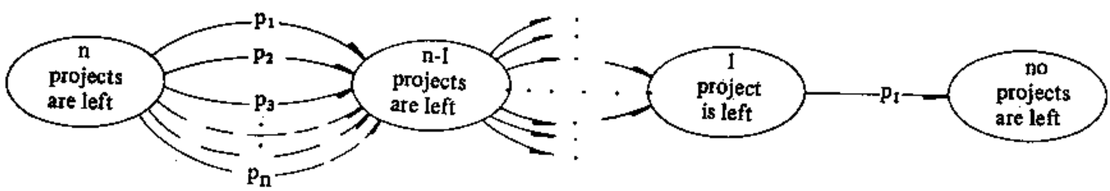

Figure 4. Stages on Instants of Selection Process

In order to reveal the characteristics of the SPDP formulation for the capacity expansion problems, an uncapacitated expansion sequencing problem of Kim (1985) was solved and compared with two other selected solution procedures. Demand is considered to be strictly increasing as was assumed by Morin (1975), and is given in Figure 5. An effective annual discount rate of 5% was arbitrarily used for the present worth calculation.

3.3 An Uncapacitated Expansion Sequencing Problem

Suppose we have only three projects to choose to meet the demand projection shown in Figure 5. Project investment costs are given as a function of project capacities and illustrated in Figure 6. In order to make the problem simple, linear cost functions are assumed. Other details of each project are given in Table 1. Since we assumed a deterministic demand projection, the timing t(.) is determined by the inverse demand function assuming no backstop supply is allowed.

Table 1. Cost Functions of Potential Projects

Project	1(A)	2(B)	3(C)
Upper Bound	35	50	50
Lower Bound	5	15	10
Cost Function	1.3Q + 9	Q + 10	1.25Q

* Q represents the capacity of each project.

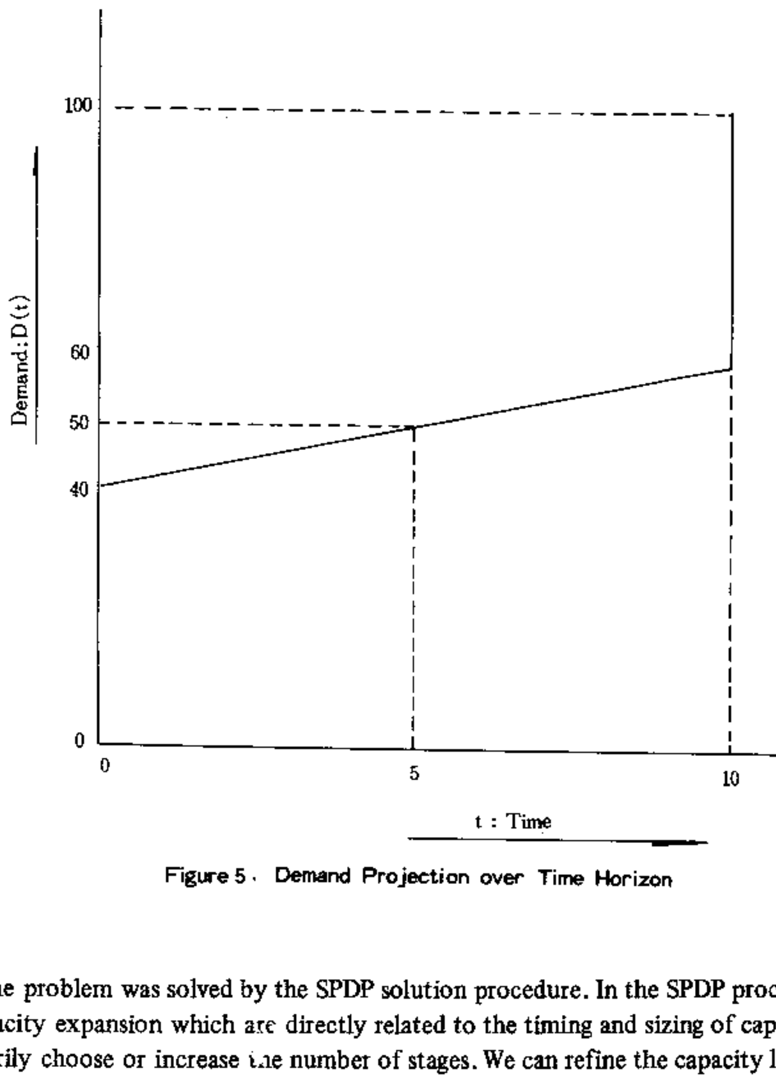

Figure 5. Demand Projection over Time Horizon

The problem was solved by the SPDP solution procedure. In the SPDP procedure, stages are instants of capacity expansion which are directly related to the timing and sizing of capacity expansions. We can arbitrarily choose or increase the number of stages. We can refine the capacity levels and timings of each project by increasing the number of stages considered. If we take stages as equal increment of time within the finite planning horizon, we may evaluate the economic cost of capacity expansion schedule as a function of the total number of stages. For the sake of fairness to the EBSS-DP for later comparison, 10 and 20 stages are considered in Table 2, so that stages in multiples of 5 are evaluated.

Table 2. Solutions by the Shortest Path Dynamic Programming (SPDP)

Item	10-stages	20-stages
Project Sequence	2 - 3	2 - 3 - 1
Timing	0 5	0 5 10
Sizing	50 50	50 15 35
Discounted Total Cost	108.97	108.15

The SPDP has capability of finding optimum sizes provided that optimum sequence and enough stage evaluations are allowed. However, the optimum sequence is not always guaranteed. In order to find an ultimate optimal solution, the uncapacitated expansion sequencing problem was solved by solving 3! (= 6) nonlinear programming problems subject to linear constraints as was defined in equation (1), (2), (3), (4). Solutions are described in Table 3. Due to the non-convexity of objective function for each permutation schedule, the optimality of each solution has been tested for various initial solutions. And the results in Table 3 are believed to be the optimum solutions for each given sequence. For a small number of projects (number of project ≤ 5) the total enumeration of all permutation schedules seems to be a viable solution procedure.

Table 3. Solutions by Total Enumeration
* indicates the selected sequence

NO.	Sequence Timing Capacity			Total Discounted Cost
1	1 0 10	2 0 50	3 10 40	112.696
2	1 0 19.94	3 0 40.06	2 10 40.0	115.693
3	2 0 50	1 5 10	3 10 40	107.934*
4	2 0 50	3 5 15	1 10 35	108.15
5	3 0 50	1 5 10	2 10 40	110.433
6	3 0 20	2 0 50	1 10 30	114.468

According to this result, it is obvious that the SPDP formulation fails to find an optimum sequence. Let me present another solution approach.

Erlenkotter (1975) proposed using total capacity as another state variable with the Binary State Vector, R to include selection of project scale as in equation (14).

$$Z(R, Y\text{-}y_i) = \underset{0 \leqslant y_i \leqslant Y}{\text{Min}} \quad [C_i(y_i)e^{-rt(Y\text{-}y_i)} + Z(RUi, Y)] \tag{14}$$

$$\text{subject to} \quad t(x) = D^{-1}(x) \tag{15}$$

$$D^{-1}(x) = \begin{cases} 0 & \text{for } 0 \leqslant x \leqslant 40 \\ 0.5(x\text{-}40) & \text{for } 40 \leqslant x \leqslant 60 \\ 10 & \text{for } 60 \leqslant x \leqslant 100 \end{cases}$$

where, $Z(R,Y)$: total discounted cost functional at stage (R U Y),
R : subset of projects whose elements are the remaining candidate projects,
Y : set of cumulative capacities,
y_i : capacity of project i,
$C_i(y_i)$: prespecified cost function as a function of project size y_i,
$t(x)$: the timing for project expansion when the cumulative capacity level is at x,
$D^{-1}(x)$: inverse demand function for capacity level x.

Numerically the recursive equation (14) may be solved by discretizing the capacity state variable Y to a finite set of values. However, the selection of appropriate aggregated capacity levels for the capacity state variable Y is, in fact, another decision that has to be made for the uncapacitated problem. Since we have no idea about the optimal capacity levels, we have to select them arbitrarily and the optimal solution should be found by trial and error. As a first trial, let the capacity state variable Y be discretized to a set of values, (100, 65, 55, 50, 45, 35). A network diagram representation of the Extended Binary State Space Dynamic Programming (EBSS-DP) formulation of equation (14) is depicted in Figure 7. The optimal solution constructs project 2 first, followed by project 3 and then project 1 with corresponding capacity levels of 50, 15, and 35 at timing 0, 5th year, and 10th year respectively.

If the same recursive equation had been applied to a set of total capacity values (100, 80, 60, 50, 40, 20), it would have found a totally different solution, which is better than the previous one. For comparison, the solution of both applications are shown in Table 4. Although the difference in the total discounted cost is not very significant in this example, it shows that the sequencing decisions are dependent upon the scale decision. The different order of project establishment could mean different system operation rule during the interim periods for practical application, which would likely result in different solution if interdependencies among projects exist.

Table 4 shows that the optimal expansion sequence is dependent upon the sizing decisions. Also the scale decisions are dependent upon the sequencing decision. It can easily be seen from the fact that once a sequence is given, the problem turns into nonlinear programming problems for scale decision as was represented in equation (1). In other words, the selection of the optimum sequence may not be attained simultaneously unless the capacity levels are fixed at certain discretized values. Likewise the optimum scale decisions cannot be determined until the optimum sequence is predetermined. Hence either scales or sequence should arbitrarily be fixed to get the optimum solution. Accordingly the total enumeration of all N! permutation sequences with nonlinear capacity optimization is indeed the only measure that

Table 4. Solutions for the Uncapacitated Example by EBSS - DP

Item	Solution Obtained For a Set of Capacity Levels (100, 65, 55, 50, 45, 35)	Alternate Solution For a Set of Capacity Levels (100, 80, 60, 50, 40, 20)
Project Sequence Timing Sizing	 2 – 3 – 1 0 5 10 50 15 35	 2 – 1 – 3 0 5 10 50 10 40
Discounted Total Cost	108.15	107.93

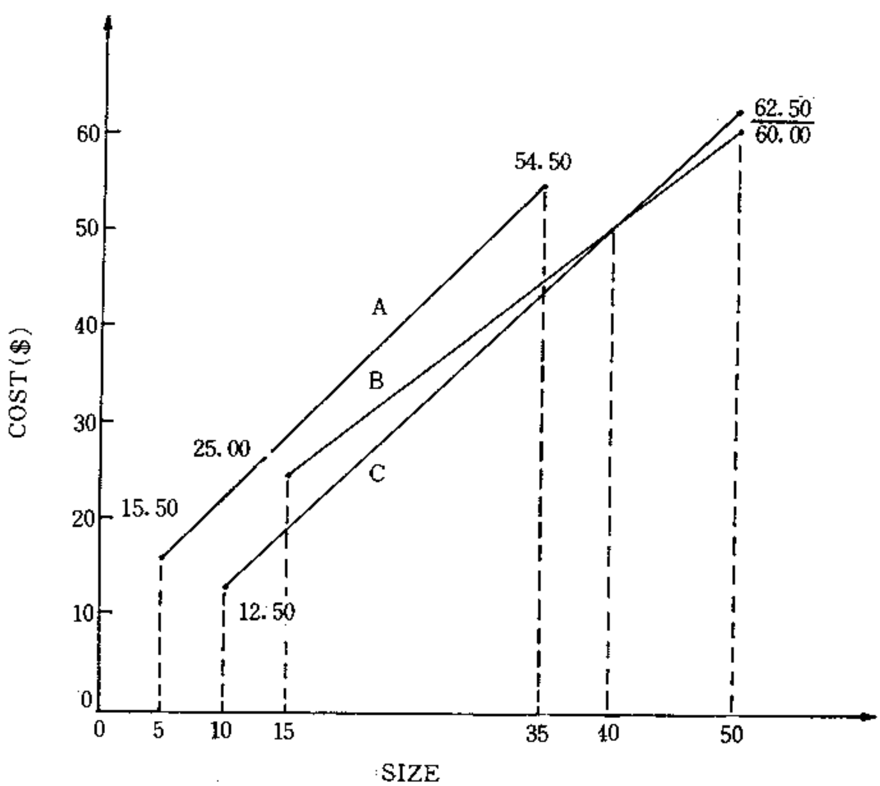

Figure 6. Cost Functions Assumed in the Example

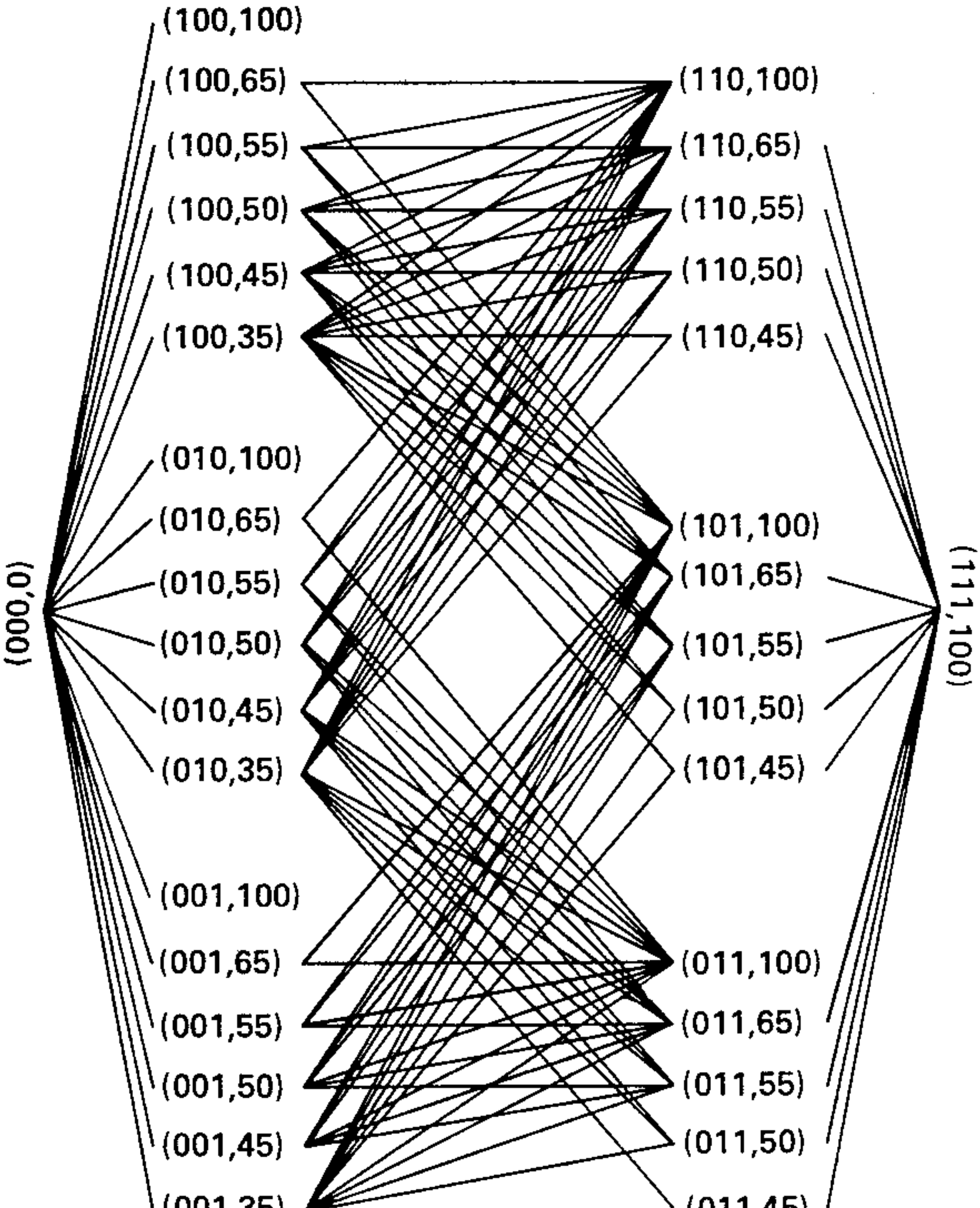

Figure 7. Network Diagram for the EBSS-DP with Total Capacity

can guarantee the global optimum for the uncapacitated expansion sequencing problem.

By the total enumeration in Table 3, we may conclude that the set of capacity levels, (100, 80, 60, 50, 40, 20) was indeed the appropriate set that would have lead the EBSS-DP approach to the optimum solution. Hence, we know that the EBSS-DP has capability of finding an optimum sequence if a correct set of total capacities is provided. By choosing arbitrary capacity levels, the EBSS-DP sets states. On the other hand, the SPDP sets stages arbitrarily and as a result, it indirectly sets potential expansion timings and sizes of capacity expansions. The SPDP may not guarantee the optimum sequence but a "good sequence" because of inappropriate separability condition from the standpoint of project sequence.

And yet the proposed SPDP solution procedure is theoretically comparable with a nonlinear solution procedure that would find local optimum capacity level, provided that the selection of a project sequence is fixed and that the number of time stages can be varied to sufficiently large number within a planning horizon. As long as the objective function has the property of additive separability, we may solve the nonlinear optimization problem by the SPDP. It can be achieved by considering each additive separable cost term as the cost function of individual expansion candidate.

On the other hand, the SPDP has capability of finding optimum capacity levels provided that optimum sequence and enough stage evaluations are allowed. Even if the optimum sequence is not guaranteed always, the SPDP has the tendency of finding a "good sequence" with intrinsical flexibility in choosing specific projects at specific states that the EBSS-DP lacks. Besides, a simulation submodel can be easily incorporated into the SPDP solution framework.

4. Conclusion

The Extended Binary State Space approach would be effective in finding an optimal expansion sequence provided that a correct set of total capacity levels is found.

Since the shortest Path Dynamic Programming (SPDP) formulation is based on the time-staged nodes with incomplete separability condition, it may not guarantee the optimal sequence. It is because specific decisions of the capacity and project selection at each time stage exclude other permutation schedules by restricting the scope of states to the attributes resulted from decisions on previous stages under the acyclic network framework.

In order to have a complete separability for DP solution procedure, different states must be established for each permutation as in the example of mountain climbing in the previous discussion. The correct realization of the separability condition would lead us to have an expanding tree associated with each permutation sequences. Practically, however, it is equivalent to searching all permutation schedules. Usually the separability condition can be achieved by increasing the number of states, which might induce the "curse of dimensionality".

As the number of projects increases, the number of states to be evaluated becomes more critical and it would be better to keep the number of states from increasing. In this regard, the SPDP solution procedure would be a practical solution approach to find a "good sequence" with less state evaluation. Besides, it is comparable with a nonlinear solution procedure that would find local optimum capacity level if the selection of a project sequence is fixed. To overcome the inability of covering all the permutation schedule to some extent, a heuristic procedure of sequence generation may be appropriate. Since the SPDP solution framework is flexible to incorporate exogenous conditions that fix a portion of a sequence string, the SPDP may be used to evaluate the economic consequence of expansion policies as a 'return function'.

References

1. Becker, L., and W. W-G. Yeh, "Optimal Timing, Sequencing, and Sizing of Multiple Reservoir Surface Water Suply Facilities," *Water Resources Research,* Vol. 10, No. 1 (February 1974), pp. 57-62.
2. Bellman, R., "On a Routing Problem," *Quarterly of Applied Mathematics,* Vol. 16, No. 1, (April 1958), pp. 87-90.
3. Butcher, W. S., Y. Y. Haimes, and W. A. Hall, "Dynamic Programming for the Optimal Sequencing of Water Supply Projects," *Water Resources Reasearch,* Vol. 5, No. 6 (December 1969), pp. 1196-1204.
4. Dijkstra, E. W., "A Note on Two Problems in Connection with Graphs," *Numerische Mathematik,* Vol. 1, (1959), pp. 269-271.
5. Erlenkotter, D., "Sequencing Expansion Projects," *Operations Research,* Vol. 21, No. 2 (March-April 1973), pp. 542-553.
6. Erlenkotter, D., "Comments on 'Optimal Timing, Sequencing, and Sizing of Multiple Reservoir Surface Water Supply Facilities' by L. Becker and W. W-G. Yeh," *Water Resources Reasearch,* Vol. 11, No. 2 (April 1975), pp. 380-381.
7. Erlenkotter, D., "Coordinating Scale and Sequencing Decisions for Water Resources Projects," *Economic Modeling for Water Policy Evaluation,* North-Holland/TIMS Studies in the Management Sciences. Vol. 3, (1976), pp. 97-112.
8. Hastings N. A. J. *Dynamic Programming with Management Applications,* Butterworths, London, 1973, pp. 28-29.
9. Jacobsen, S., "Heuristic Procedures for Dynamic Plant Location," The Institute of Mathematical Statistics and Operations Research, The Technical University of Denmark DK-2800 Lyngby, Denmark, Working Paper (August 1977).
10. Kim, S. K., "Capacity Expansion Models for Optimum Investment Planning," *Journal of the Korean Institute of Industrial Engineers,* Vol. 11, No. 2, (Dec. 1985), pp. 141-154.
11. Morin, T. L., and A. M. O. Esogbue, "Some Efficient Dynamic Programming Algorithms for the Optimal Sequencing and Scheduling of Water Supply Projects," *Water Resources Research,* Vol. 7, No. 3 (June 1971), pp. 479-484.
12. Morin, T. L., "Multidimensional Sequencing Rule," *Operations Research,* Vol. 23, No. 3 (May-June 1975), pp. 576-580.
13. Tsou, C. A., L. G. Mitten, S. O. Russell, "Search Technique for Project Sequencing," *Journal of the Hydraulics Division, Proceedings of the American Society of Civil Engineers,* Vol. 99, No. HY5 (May 1973), pp. 833-839.