International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:6, No:8, 2012

Periodic Storage Control Problem

Ru-Shuo Sheu, Han-Hsin Chou, Te-Shyang Tan

Abstract—Considering a reservoir with periodic states and different cost functions with penalty, its release rules can be modeled as a periodic Markov decision process (PMDP). First, we prove that policy- iteration algorithm also works for the PMDP. Then, with policy- iteration algorithm, we obtain the optimal policies for a special aperiodic reservoir model with two cost functions under large penalty and give a discussion when the penalty is small.

keywords—periodic Markov decision process, periodic state, policy-iteration algorithm.

I. Introduction

A state i in a Markov chain is said to have period d is that the transition probability of returning from state i to i equals to 0 whenever n is not divisible by d and d is the greatest integer with this property. Particularly, we say the state is aperiodic when d is equal to 1, otherwise it is periodic.

We consider a reservoir discretized into n blocks, i.e. $n+1$ levels of water, and check the water level at times with period d in it. Hence, if the time intervals are long, the transition probabilities of inflow can be treated as stationary or they form a non-stationary, cyclic, unichain Markov chain. At each decision point we can choose to sell up to any level of water in the reservoir with one of several different cost functions. Moreover, the selection of these cost functions is according to a finite-state Markov chain. A penalty M will be incurred if no water has been sold. Stochastic inflows are independent and identically distributed. If the inflows exceed the capacity of the reservoir, they will be spilt. Therefore, the release rules of this model can be considered as a periodic Markov decision process (PMDP).

In general, simulations, linear programming, non-linear programming and dynamic programming are available methods in dealing with release policies for reservoir systems. Though choice of methods depends on the characteristics of the system being considered, dynamic programming has the advantage of effectively decomposing intensively complex problems with a large number of variables into a series of subproblems which are solved recursively [8], [12], [13]. Nagy, Asante-Duan and Zsuffa [6] apply Moran's probability model [4] in reservoir storage designs. Heidari et al. [2] use incremental dynamic programming to reduce the requirements of computer capacities through water resources system. Nopmongcol and Askew [7] conclude that discrete differential dynamic programming is the generalization of incremental dynamic programming. However, these methods seek answers to probabilistic rather than optimization problems.

Considering optimal control problems, Turgeon [11] optimized a multireservoir hydroelectric power system with two dynamic programming methods to break up the original multivariables problem into a series of one-step variable subproblems. Goulter and Tai [1] found that a small number of storage states produces high skewness in the storage probability distribution functions and influences the optimal operation policy. Moran [5] determined a close-optimal operating policy for a multireservoir system.

For policy improvement techniques, Howard [3] introduced the Markov chain method of successive approximation and the policy iteration algorithm to solve reservoir operation problems. Tijms [10] pointed out that, for a Markov decision process (MDP) problem with a large number of states, policy iteration algorithm and value- iteration are usually the most efficient methods to solve it with quickly converging lower and upper bounds on the minimal costs. Finally, by applying policy iteration algorithm and value-iteration, Sheu et al. [9] induce the release policies for a reservoir model with single linear cost function.

In our approach, we first prove that the relative values and the average cost of a given policy R satisfy a simultaneous system of linear equations. Then, by constructing the average cost of a new policy $\overline{R}$ through the relative values of a given policy R, we imply that the average cost of a new policy is no more than that of the current policy under unichain assumption. That is, it proves that the policy-iteration algorithm also works for the PMDP problems. In the following section, we find the best policies for a modified reservoir model with large penalty and two different cost functions by policy-iteration algorithm. Finally, by the results of a simplified example, we give a short discussion and make a conjecture at the end.

II. Peridoic Model with Different Cost Functions

For the reservoir control model we considered in the previous section, let R be the stationary policy which depends on time within period, water levels and price function. If the period of R is d, define $t^* = t \bmod d$ and let x_n be the state of the system at the n-th decision epoch, $n \in N \cup 0$. Hence,

$$x_n = \{(i, t^*, k) \mid i : \text{reservoir level}, t^* : \text{time within period}, k : \text{cost function}\}.$$

Ru-Shuo Sheu is with the Hsin-Wu University of Technology, Taipei, Taiwan. He is now the head of the Department of Travel Management (e-mail: 080015@mail.hwc.edu.tw)

Han-Hsin Chou is with the Travel Management Department, Hsin-Wu University of Technology, Taipei, Taiwan.

Te-Shyang Tan is with the Hospitality Management Department, Hsin-Wu University of Technology, Taipei, Taiwan.

The transition probability is

$$\begin{aligned} &P[x_{n+1} = (j,(n+1)^*,c_m)|x_n = (i,n^*,c_l)] \\ &\quad = \; p_{(i,n^*,c_l)(j,(n+1)^*,c_m)}(R_{(i,n^*,c_l)}). \end{aligned}$$

A. Relative values associated with a given policy

For proving the relative values and the average cost of a given policy R satisfy a simultaneous system of linear equations, some notations need to be defined beforehand.

Let $\mathcal{V}_n(R,i,k,c_l)$ be the total expected cost over the first n decision epochs when the initial state is i-level of water, at time k with price function c_l and policy R. That is,

$$\begin{aligned} &\mathcal{V}_n(R,i,k,c_l) \\ &\quad = \sum_{t=k}^{n+k-1} \sum_{(j,q,c_m)\in I} p^{(t-k)}_{(i,k^*,c_l)(j,q^*,c_m)}(R)[c_m(R_{(j,t^*)})] \end{aligned}$$

and the recursion equation will be

$$\begin{aligned} \mathcal{V}_n(R,i,k,c_l) &= c_l(R_{(i,k^*)})+ \\ &\sum_{(j,k+1,c_m)\in I} p_{(i,k^*,c_l)(j,(k+1)^*,c_m)}(R)\mathcal{V}_{n-1}(R,j,k+1,c_m) \end{aligned}$$

where $n,k \geq 1$.

Since the cost functions $c_m(R_{(j,q^*)})$ are periodic and the state space is discrete, by Theorem 2.2.2 in Tijms' book[10], $g_{(i,k,c_l)}(R) = \lim_{n\to\infty} \frac{1}{n}\mathcal{V}_n(R,i,k,c_l)$ exists. So we define the average cost function as:

$$g_{(i,k,c_l)}(R) = \lim_{n\to\infty} \frac{1}{n}\mathcal{V}_n(R,i,k,c_l)$$

Morever, because the long-run average expected cost per unit of time is independent of the initial state (i,k,c_l) when it is assumed that x_n corresponding to policy R is unichained, we imply $g_{(i,k,c_l)}(R) = g(R)$.

Let $T_{(i,t^*,c_l)}(R)$ be the expected time until the first visit to the state (r,n^*,c_m) when starting in state (i,t^*,c_l) with policy R. That is, $T_{(r,n^*,c_m)}(R)$ is the expected length of a cycle. Let $K_{(i,t^*,c_l)}(R)$ be the expected cost incurred until the first visit to the state (r,n^*,c_m) when starting in state (i,t^*,c_l) with policy R. Note that $K_{(i,t^*,c_l)}(R)$ includes the cost incurred when starting in state (i,t^*,c_l) but excludes the cost incurred when returning to state (r,n^*,c_m).

Thus, we have

$$g(R) = K_{(r,n^*,c_m)}(R)/T_{(r,n^*,c_m)}(R)$$

and we define the relative value

$$w_{(i,k^*,c_l)}(R) = K_{(i,k^*,c_l)}(R) - g(R)T_{(i,k^*,c_l)}(R).$$

Theorem 1: Let R be given stationary policy such that the associated Markov chain x_n has no two disjoint closed set, then

(a) *the average cost function $g(R)$ and the relative value $w_{(i,k^*,c_l)}(R)$ satisfy the following system of linear equations in the unknowns g and $v_{(i,k,c_l)}$*

$$\begin{aligned} v_{(i,k,c_l)} &= c_l(R_{(i,k^*)}) - g+ \\ &\sum_{(j,k+1,c_m)\in I} p_{(i,k^*,c_l)(j,(k+1)^*,c_m)}(R)v_{(j,k+1,c_m)}. \end{aligned}$$

(b) *Let g and $v_{(i,k,c_l)}$ be any solution to (a), then $g = g(R)$ and for some constant c, we have*

$$v_{(i,k,c_l)} = w_{(i,k^*,c_l)}(R) + c, \ \forall (i,k,c_l) \in I.$$

(c) *Let s be an arbitrarily chosen state, then the linear equations in (a) together with the normalization equation $v_s(R) = 0$ have a unique solution.*

Proof. We omit the proof of the theorem, which follows in a very similar to that given by Tijms(1994) with minor modifications due to differences in the structure of the model. □

B. Policy-iteration algorithm in a periodic model

The intuitive idea behind the policy-iteration algorithm for improving a given policy is to make the difference in costs as negative as possible. To do this, we let $\triangle(i,k,c_l,a,R)$ be the difference in total expected cost over an infinitely long period of time by taking first action a, under time epoch k and price function c_l, and next using policy R rather than using policy R from the beginning onward when the initial state is (i,k^*,c_l). Provided results from previous section, we have

$$\begin{aligned} \triangle(i,k,c_l,a,R) &= c_l(a_{(i,k^*)}) - g(R) - v_{(i,k,c_l)}(R)+ \\ &\sum_{(j,k+1,c_m)\in I} p_{(i,k^*,c_l)(j,(k+1)^*,c_m)}(a_{(i,k^*)})v_{(j,k+1,c_m)}(R) \end{aligned}$$

So in each state (i,k^*,c_l), we look for an action a which makes

$$\begin{aligned} &c_l(a_{(i,k^*)}) - g(R)+ \\ &\sum_{(j,k+1,c_m)\in I} p_{(i,k^*,c_l)(j,(k+1)^*,c_m)}(a_{(i,k^*)})v_{(j,k+1,c_m)}(R) \end{aligned}$$

as small as possible.

Theorem 2: Let g and $v_{(i,k,c_l)}$, $(j,k,c_l) \in I$, be given. Suppose the policy $\bar{R}$, for each $(i,k,c_l) \in I$, follows :

$$\begin{aligned} &c_l(\bar{R}_{(i,k^*)}) - g+ \\ &\sum_{(j,k+1,c_m)\in I} p_{(i,k^*,c_l)(j,(k+1)^*,c_m)}(\bar{R})v_{(j,k+1,c_m)} \ \leq \ v_{(i,k,c_l)}. \end{aligned}$$

Then the long-run average cost of policy $\bar{R}$ satisfies

$$g_{(i,k,c_l)}(\bar{R}) \ \leq \ g, \quad (i,k,c_l) \ \in \ I.$$

It is also true when the inequality signs are reversed.

Proof. With a little algebra, we can induce the follolwing result bt induction;

$$v_{(i,k,c_l)} \geq \sum_{t=k}^{m+k-1} \sum_{(j,q,c_b)\in I} p^{(t-k)}_{(i,k^*,c_l)(j,q^*,c_b)}(\bar{R})[c_b(\bar{R}_{(j,t^*)})] \\ -mg + \sum_{(j,q,c_b)\in I} p^{(m)}_{(i,k^*,c_l)(j,q^*,c_b)}(\bar{R})v_{(j,k+m,c_b)};$$

where $m = 1, 2, 3, \ldots$ Because if

$$\mathcal{V}_n(R, i, k, c_l) = \\ \sum_{t=k}^{n+k-1} \sum_{(j,q,c_m)\in I} p^{(t-k)}_{(i,k^*,c_l)(j,q^*,c_m)}(R)[c_m(R_{(j,q^*)})]$$

then

$$\mathcal{V}_n(\bar{R}, i, k, c_l) = \\ \sum_{t=k}^{n+k-1} \sum_{(j,q,c_m)\in I} p^{(t-k)}_{(i,k^*,c_l)(j,q^*,c_m)}(\bar{R})[c_m(\bar{R}_{(j,q^*)})].$$

It implies

$$v_{(i,k,c_l)} \geq \mathcal{V}_m(\bar{R}, i, k, c_l) - mg + \\ \sum_{(j,q,c_b)\in I} p^{(m)}_{(i,k^*,c_l)(j,q^*,c_b)}(\bar{R})v_{(j,k+m,c_b)}.$$

Then dividing both sides by m and let $m \to \infty$, we have

$$\lim_{m\to\infty}[v_{(i,k,c_l)}/m] \\ \geq \lim_{m\to\infty}[\frac{\mathcal{V}_m(\bar{R}, i, k, c_l)}{m} - g + \\ \frac{1}{m}\sum_{(j,q,c_b)\in I} p^{(m)}_{(i,k^*,c_l)(j,q^*,c_b)}(\bar{R})v_{(j,k+m,c_b)}.]$$

Because

$$\lim_{m\to\infty} v_{(i,k,c_l)}/m = 0, \ \lim_{m\to\infty} \mathcal{V}_m(R, i, k, c_l)/m = g(\bar{R})$$

and

$$\lim_{m\to\infty} \frac{1}{m} \sum_{(j,q,c_b)\in I} [p^{(m)}_{(i,k^*,c_l)(j,q^*,c_b)}(\bar{R})v_{(j,k+m,c_b)}] = 0,$$

so $0 \geq g(\bar{R}) - g$. That is

$$g(\bar{R}) \leq g, \forall (i, k, c_l) \in I.$$

Similarly, the theorem remains true when the inequality signs are reversed. □

Hence, we prove that the policy-iteration algorithm works for the periodic MDP problems.

III. Control Policy of a Modified reservoir Model

We consider an aperiodic reservoir model with two different cost functions and penality of not releasing any water from the dam. The selection of these two cost functions is according to a finite-state Markov chain By applying the policy-iteration algorithm, we find out the best release policy in the modified aperiodic MDP problem with large penalty.

Let M be The penalty we receive if we do not release any water from the reservoir, p_i be the probability of i-unit of water comes into the reservoir with $0 \leq i \leq n$ and $\sum_{k=0}^{n} p(k) = 1$. Two cost functions, c_1 and c_2, are defined as :

$$c_j(r) = \begin{cases} M & r = 0 \\ -r\pi_j & r = 1, 2, \cdots, n \end{cases} ; \ \pi_1 \ \leq \ \pi_2$$

with constrants $\pi_1 \leq \pi_2$ and the transition matrix of them, P_c is

$$P_c = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}$$

where $p_{ij} \geq 0$, $p_{i1} + p_{i2} = 1$ and $i, j = 1, 2$.

The states of the model is $\mathcal{S} = \{(i,\ j) \mid i$: level of water, j : cost function; $i = 0, 1, \cdots, n$, $j = 1, 2.\}$. For sufficiently large penalty M, $M \gg 0$, the following theorem provides the optimal state-dependent policy under the unichain assumption.

Theorem 3: For sufficiently large M, the best state-dependent policy is to release 1 unit of water from the reservoir when it is possible.

Proof. Let $v_{ij}(R)$ be the related value with the $(i,\ j)$- starting state when policy R is used and for $l = 0, 1, 2, \cdots, n$,

$$A_{il} = \sum_{j=0}^{n-l-1} p_j(p_{i1}v_{(j+l)1} + p_{i2}v_{(j+l)2}) + \\ (p_{n-l} + p_{n-l+1} + \ldots + p_n)(p_{i1}v_{n1} + p_{i2}v_{n2}).$$

We prove the theorem by induction.

I. In state (1, 1) and (1, 2):
For $M \gg 0$, $v_{11} = -c_1(1) - g(R) + A_{10}$ and $v_{12} = -c_2(1) - g(R) + A_{20}$. That is, the best policy would be to release 1 unit of water from the reservoir.
Assume both in state (k, 1) and state (k, 2), the best policies be to release 1 unit of water from the reservoir. That is,

$$v_{(ij} = -c_j(1) - g(R) + A_{j(i-1)}$$

for $i = 1, 2, \cdots, k$ and $j = 1, 2$.

II. In state $(k + 1,\ 1)$, $k + 1 \leq n$:
By aplying contraction mapping theorem and with similar method to theorem2.1 in Sheu et al.'s[9], for $2 \leq s \leq k + 1$, we have

$$v_{(j+k-s+1)i} - v_{(j+k)i} \geq 0$$

and

$$v_{(j+k-s+1)i} - v_{ni} \geq 0$$

for $i = 1, 2$. Then

$$
\begin{aligned}
&[-c_1(1)-g(R)+A_{1k}]-[-c_1(s)-g(R)+A_{1(k-s+1)}] \\
&= \ (s-1)\pi_1 - \{ \sum_{j=0}^{n-k-1} p_j[p_{11}(v_{(j+k-s+1)1}-v_{(j+k)1}) \\
&\qquad + p_{12}(v_{(j+k-s+1)2} - v_{(j+k)2})] \\
&\qquad + \sum_{j=n-k}^{n-k+s} p_j[p_{11}(v_{(j+k-s+1)1} - v_{n1}) \\
&\qquad + p_{12}(v_{(j+k-s+1)2} - v_{n2})]\} \\
&\leq (s-1)\pi_1 - p_0p_{11}(v_{(k-s+1)1} - v_{k1}) \\
&= (s-1)\pi_1 - p_0p_{11}\sum_{i=1}^{s-1}[v_{(k-i)1} - v_{(k-i+1)1}] \\
&\leq (s-1)\pi_1 - p_0p_{11}(v_{(k-s+1)1} - v_{(k-s+2)1}) \\
&= (s-1)\pi_1 - p_0p_{11}[A_{1(k-s)} - A_{1(k-s+1)}] \\
&\leq (s-1)\pi_1 - p_0p_{11}[p_0p_{11}(v_{(k-s)1} - v_{(k-s+1)1})] \\
&= (s-1)\pi_1 - [p_0p_{11}]^2(v_{(k-s)1} - v_{(k-s+1)1}) \\
&\qquad\qquad \text{(repeat } k-s \text{ times)} \\
&\leq (s-1)\pi_1 - [p_0p_{11}]^{k-s+1}(v_{01} - v_{11}) \\
&= (s-1)\pi_1 - [p_0p_{11}]^{k-s+1}[M-(-c_1(1)-g(R)+A_{10})].
\end{aligned}
$$

Because $k \leq n-1$, $s \leq k+1$, $M \gg 0$ and $\pi_1 \leq \pi_2 < \infty$, then

$$(s-1)\pi_1 - [p_0p_{11}]^{k-s+1}[M-(-c_1(1)-g(R)+A_{10})] \leq 0.$$

Which implies

$$[-c_1(1)-g(R)+A_{1k}]-[-c_1(s)-g(R)+A_{1(k-s+1)}] \leq 0.$$

Similarly, for $2 \leq s' \leq k+1$,

$$[-c_2(1)-g(R)+A_{2k}]-[-c_2(s')-g(R)+A_{2(k-s'+1)}] \leq 0.$$

That is, for both state $(k+1,1)$ and state $(k+1,2)$, the best policies will be to release 1 unit of water from the reservoir.

By I and II, with $M \gg 0$, the best policy is to release 1 unit of water from the reservoir in all states. □

IV. Discussion

From previous section, the optimal state-dependent policies are obtained when the penalties are sufficiently large. For penalty $M < \infty$, we consider a reservoir whose capacity is discretized into 4 blocks, i.e. 0, 1, 2, and 3 levels of water in it. There are two different cost functions and the selection of them according to a finite-state Markov chain. By adding constraints $p_{11} = p_{21}$ and $p_{12} = p_{22}$ this simplified model, we are able to prove that the best policy for the higher cost function is always not less than the best policy for the lower cost function. The results also imply a random assignment every period, with no dependence on previous state, and hence, no periodicity. Hence, we make a conjecture and all of the optimal policies found here will be satisfied.

Conjecture. *The best policy for the higher cost function is always not less than the best policy for the lower cost function when the selection from these two different cost functions is according to a finite-state dependent Markov chain with constraints $p_{11} = p_{21}$ and $p_{12} = p_{22}$.*

Further work includes generalizing the model with more than two cost functions in it. Moreover, it is also interesting to discuss the necessities of adding constraints to the model.

References

[1] Goulter, I.C. and Tai, F-K.(1985) Practical Implications in the Use of Stochastic Dynamic Programing for Optimal Reservoir Operation. *Water Resources Bulletin* **21, No.1** 65-74.

[2] Heidari, M., Chow, V.T., Kokotovic, P.V.and Meredith, D.D.(1971) Discrete Differential Dynamic Programing Approach to Water Resources System Optimization. *Water Resources Research* **7, No.2** 273-282.

[3] Howard, R.A. *Dynamic Programming and Markov Processes* MIT Press, Cambridge, Mass.(1960).

[4] Moran, P.A.P.(1954) Probability Theory of Dams and Storage Systems. *Aust. J. Appl. Sci.* **5** 116-124.

[5] Moran, P.A.P.(1955) A Probability Theory of Dams and Storage Systems: Modifications of the Release Rules. *Aust. J. Appl. Sci.* **6** 117-130.

[6] Nagy, I.V. ,Asante-Duah, K. and Zsuffa, I., *Hydrological Dimensioning and Operation of Reservoirs* Kluwer Academic Publishers(2002).

[7] Nopmongcol, P. and Askew, A.J.(1976) Multilevel Incremental Dynamic Programing. *Water Resources Research* **12, No.6** 1291-1297.

[8] Reznicek, K. and Cheng, T.C.E.(1991) Stochastic Modelling of Reservoir Operations. *European Journal of Operation Research* **50** 235-248.

[9] Ru-Shuo Sheu, Pang-Ju Lou, and Han-Hsin Chou(2011) Reservoir Storage Release Rules Control Problem. *2011 International Symposium on Water Resource and Environmental Protection (ISWREP 2011)*(IEEE)

[10] Tijms, H.C. *Stochastic Models : An Algorithmic Approach* John Wiley & Sons, Inc. New York, (1994).

[11] Turgeon, A. (1980) Optimal Operation of Multireservoir Power System with Stochastic Inflows. *Water Resources Research* **16, No.2** 275-283.

[12] Wurbs, R.A. (1993) Reservoir-System Simulation and Optimization Models. *Journal of Water Resources Planning and Management* **119, No.4** 455-472.

[13] Yeh, W.W-G.(1985) Reservoir Management and Operations Models : A State-of-Art Review. *Water Resources Research* **21, No.12** 1797-1818.