49th IEEE Conference on Decision and Control
December 15-17, 2010
Hilton Atlanta Hotel, Atlanta, GA, USA

Dynamic Water Resources Planning with Locational Release and Annual Consumption Constraints

Yanjia Zhao, *Student Member*, *IEEE*, Xi Chen, *Member*, *IEEE*, Qing-Shan Jia, *Member*, *IEEE*, Xiaohong Guan, *Fellow*, *IEEE*, and Hao Wang

***Abstract*—Water resources planning is very important for water resource management and electric energy production. The problem is challenging in view of the stochastic system dynamics, nonlinear rewards, coupled hydraulic constraints, and large problem size. Existing methods based on discretization are facing the dilemma of solution accuracy and computational efforts. This paper formulates the dynamic water resources planning problem with locational release and annual consumption constraints as a finite-horizon Markov decision process (MDP) with continuous variables, and develops a sensitivity-based approach to optimize the policies. Numerical results based on a practical system on Yellow River in north China demonstrate the effectiveness of the formulation and the efficiency of the new algorithm.**

I. Introduction

Water resources planning is very important for water resource management and electric energy production. The planning goal discussed in this paper is to maximize the total reward from hydro generation by determining the discharge and spillage to downstream and the water consumption for irrigation and urban supply of all reservoirs in each stage, subject to various operating and hydraulic coupling constraints, and government regulations for daily life, agricultural and ecological requirements. Typically, the planning horizon is one year, and the decision stage is one week or two weeks. This planning problem has huge economical and environmental impact with the potential to improve the water resources allocation and to alleviate the negative ecological effect.

The dynamic water resources planning has been an active research area over the past decades due to its significant economic impact [1][8][12][27][30]. Markov Decision Process (MDP) [13]-[15] is widely used to formulate the dynamic water resources planning due to its ability to cope with nonlinear and stochastic characteristics of such problems [2][8]. However, it faces the well known challenge of huge decision or policy space, often referred to as "curse of dimensionality" as surveyed in [8]. Various methods were developed to deal with this challenge. Approximation for value functions, such as piecewise linear function or neural network, are investigated in [7] and [8]. Approximation in policy space, such as various operation rules and heuristics, are studied in [5] and [6]. Problem approximations, like aggregation and decomposition of reservoirs, are considered in [9]-[11]. In our previous work, the water resources planning problem is formulated as a constrained Markov decision problem and a "rollout" based approximate dynamic programming method is developed to solve the problem with quantified approximation [28]. The major issue of the existing methods is that discretization in decision space is generally needed and we have to deal with the dilemma of approximation precision and computational efforts.

This paper studies a water resource planning problem with cascaded reservoirs located in Yellow River, the second largest river in China. The problem considered in the paper has the following features:

1) the locational release constraints which designate a lower bound of water release for each reservoir, are considered to prevent water cutout in the peak demand seasons;
2) the annual consumption constraint, which designate an upperbound of water consumption of the province in one year, is considered to rationalize the usage of limited water resources among different provinces;
3) the seasonal water demands for the irrigation and urban supplies are considered to guarantee the basic water consumption;
4) the water recession, i.e., portion of water consumption becomes a part of inflows to downstream reservoirs is considered.

This dynamic water resource planning problem with the above considerations and uncertainties in natural inflows is generally very difficult. For many methods that can only handle discrete decision variables, trade-offs must be considered between computational accuracy and efficiency. The framework of Markov Decision Process (MDP) with continuous states and actions is introduced in this paper to capture the uncertainties, and complicated constraints and regulations without discretizing the resource volumes. Based on the conceptual framework of perturbation analysis [16][21], a new algorithm is developed to solve the MDP problem with continuous variables and finite-horizon. The new algorithm does not rely on the ergodic and stationary assumptions and can deal with both continuous and discrete

This work is supported in part by NSFC (60736027, 60704008, 60921003), the 111 International Collaboration Project of China (B06002), and the National New Faculty Funding for Universities with Doctoral Program (20070003110).

Yanjia Zhao, Xi Chen, Qing-Shan Jia and Xiaohong Guan are with Center for Intelligent and Networked Systems, Department of Automation, TNList, Tsinghua University, Beijing 100084, China.

Xiaohong Guan is also with SKLMS Lab and MOE KLINNS Lab, Xian Jiaotong University, Xian China.

Hao Wang is with Chinese Institute of Water Resources and Hydropower Research, Beijing 100038, China.

The corresponding author is Xi Chen (bjchenxi@tsinghua.edu.cn).

variables, since the performance derivative is derived for the finite-horizon total-cost MDPs with continuous state and action variables. There is no need to determine the difficult trade-off between computational accuracy and efficiency. Numerical results illustrate that the MDP formulation is effective and the new sensitivity-based algorithm is efficient to solve the water resources planning problem with locational release and annual consumption constraints considered in this paper.

II. Problem Formulation

Assume a watershed with I hydro plants in one province and each hydro plant has its own fore-bay reservoir. It is required to determine the water discharge, spillage, and water consumption of all reservoirs over a specified horizon T subject to the operating constraints of individual reservoirs and hydraulic coupling. The goal is to maximize the total hydro generation reward. The decision stage is for one or two weeks and the planning horizon is usually one year.

The following assumptions are made to simplify the discussions without losing generality.

1) The water flow time delay between reservoirs is no more than a decision stage in the long-term planning problem;
2) The water released from a reservoir directly enters only one reservoir.
3) The reservoir network is acyclic.

In fact, these assumptions are typically made in the literature [8] and the assumption 2) and 3) can be removed by adding more variables in the model.

The *s*ystem states, actions, and dynamics are modeled as follows. For any t = 0, 1, …, T-1, the state $X(t)$ at stage t involves the storage $x_i(t)$ and the accumulated water consumption $m_i^d(t)$ in period d of reservoir i, d = 0, 1, …, D-1, i = 1, 2, …, I. The control action $A(t)$ at stage t includes discharge $w_i(t)$, spillage $s_i(t)$, and consumption $u_i(t)$ of reservoir i, i = 1, 2, …, I. Note that the spillage does not generate energy and has no reward and the consumed water is used for agricultural irrigation and industrial and urban supplies, etc.

Given the system state $X(t)$ and control action $A(t)$, the system dynamics at stage t is determined as follows: $\forall\ t$ = 0, 1, …, T-1, i = 1, 2, …, I

$$x_i(t+1) = x_i(t) - u_i(t) - r_i(t) + \sum_{j\in\mathcal{U}_i}[r_j(t) + \lambda_j u_j(t)] + \xi_i(t) \quad (1)$$

$$m_i^d(t+1) = \begin{cases} m_i^d(t) + u_i(t), & \text{if } t_d \le t < t_{d+1} \\ m_i^d(t), & \text{otherwise}, \end{cases} \quad (2)$$

$$\forall\ d = 0, 1, \ldots, D\text{-}1.$$

Eq. (1) depicts the water balance of the cascaded reservoirs, where $x_i(t)$ is the storage of reservoir i at time t and the initial storage $x_i(0)$ is given; $r_i(t)$ is the water release defined as

$$r_i(t) = w_i(t) + s_i(t). \quad (3)$$

$\mathcal{U}_i$ is the set of direct upstream reservoirs of reservoir i**.** The inflows to reservoir i involves three parts: the water discharge and the water recession from consumption of upstream reservoirs and the natural inflows. λ_j is the recession ratio of reservoir j which indicates the portion of consumption coming back to the water system. $\xi_i(t)$ is the natural inflow of reservoir i at time t, which is a random variable with given distribution.

Eq. (2) shows the accumulated consumption of each period, and we set the initial state $m_i^d(0) = 0$ for d = 0, 1, …, D-1.

Remark 1: The water recession, i.e., portion of water consumption comes back and becomes a part of inflows to downstream reservoirs, is considered as shown in Eq. (1). □

The feasible action set at stage t with state $X(t)$ is constrained by the following:

$$\underline{x}_i \le x_i(t) \le \overline{x}_i, \quad (4)$$

$$0 \le w_i(t) \le \overline{w}_i, \quad (5)$$

$$s_i(t) \ge 0, \quad (6)$$

$$\sum_{j\in\mathcal{U}_i} r_j(t) \ge \phi_i, \quad (7)$$

for i = 1, 2, …, I. Eq. (4), (5) and (6) are the physical limits for the storage, discharge, spillage, and consumption, and (7) is required for the minimum inflows to all the hydro plants to prevent water cutout, where ϕ_i is a given lower bound of controlled inflows to reservoir i every stage.

Remark 2: The locational release constraints are considered in order to prevent water cutout, as (7) shows. □

Reward structure and objective function are considered as follows. One step reward function at stage t is the benefit from hydro generation for t = 0, 1, …, T-1:

$$f_t(X_t, A_t) = a_t P_t^2 + b_t P_t + c_t \quad (8)$$

where the benefit function is a quadratic polynomial with respect to the total hydro generation p_t in stage t [31]. The hydro generation function is

$$p(t) = \sum_{i=1}^{I} \rho_i w_i(t), \quad (9)$$

where $w_i(t)$ is the water discharge from reservoir i in stage t; ρ_i is a coefficient representing the efficiency of reservoir i.

The terminal reward at stage T considers the penalty cost for the violations of seasonal demands for water irrigation and urban supply and the annual consumption constraint:

$$f_T(X(T)) = -M \cdot \sum_{i=1}^{I}\sum_{d=0}^{D-1} \mathbf{1}\left(m_i^d(T) < \underline{u}_i^d\right) \quad (10)$$

$$- M \cdot \mathbf{1}\left(\sum_{i=1}^{I}\sum_{d=0}^{D-1} m_i^d(T) > \overline{u}\right), \quad (11)$$

where M is a sufficient large number indicating the penalties; $\mathbf{1}(\bullet)$ is an indicator function which equals to 1 (or 0) if logical expression $(\bullet)$ is true (or false); $m_i^d(T)$ depicts the total water consumption of reservoir i within season d according to (2), i.e.,

$$m_i^d(T) = \sum_{t=t_d}^{t_{d+1}-1} u_i(t), \tag{12}$$

where i = 1, 2, …, I, d = 0, 1, …, D-1. Eq. (10) indicates the penalty for the violation of seasonal demands for water irrigation and urban supply, where $\underline{u}_i^d$ is a constant denoting the seasonal demand for water consumption at reservoir i in demand-period d. Eq. (11) is the penalty for violation of the annual consumption constraint, where $\overline{u}$ is a given upper bound of the annual water consumption in the province within horizon T. The value of $\underline{u}_i^d$ and $\overline{u}$ are given based on the governmental regulations and we have

$$\sum_{i=1}^{I}\sum_{d=0}^{D-1} \underline{u}_i^d \le \overline{u}\ . \tag{13}$$

Remark 3: This paper considers the seasonal demands of water in each reservoir for the irrigation and urban supplies, so that the requirement for water consumption is guaranteed. Note that this constraint is reflected by the penalty in (10). □

Remark 4: This paper considers the annual consumption constraint by setting up an upperbound of water consumption of the province in one year, in order to balance the usage of limited water resources among different provinces. Note that this constraint is reflected by the penalty in (11). This constraint is practical and important nowadays. For example, Chinese government introduces this constraint for Yellow River in north China, in consideration of the limited water resources there. □

The objective considered in this paper is to maximize the expected total reward over finite horizon T

$$\max_{\pi} : \eta(X_0, \pi) = E\left\{\sum_{t=0}^{T-1} f_t(X(t), A(t)) + f_T(X(T))\right\}, \tag{14}$$

where initial state X_0 is given; scheduling policy π consists of a series of decision rules which are mappings from the state space to the action space, i.e.,

$$\pi = (\pi_0, \pi_1, \cdots, \pi_{T-1}); \tag{15}$$

$$A(t) = \pi_t(X(t)), \forall t = 0, \cdots, T-1. \tag{16}$$

To summarize, in this section, we formulated the water resources planning with locational release and annual consumption constraints as MDP with continuous states and actions. In this way, we not only capture the nonlinear and stochastic characteristics of this problem, but also avoid the curse of dimensionality of traditional discrete formulation.

III. Sensitivity based Approach for Continuous MDP

Sensitivity-based approaches are introduced to optimize MDPs with discrete state and action spaces in [19][20] and MDPs with continuous state and action spaces in [23], combining with ideas of perturbation analysis (PA) [14][21] and reinforcement learning [22]. However, existing results limit to MDPs with infinite-horizon average-cost and rely on the ergodic assumption of stationary policies [16][17], so that they are not applicable to problems with the finite-horizon and the total-cost.

We have developed a sensitivity based approach for finite horizon MDPs with discrete state and action spaces in our previous work [24][25][28]. In this section, we will extend the main result to the finite-horizon MDPs with continuous state and action spaces in order to solve the water resources planning problem formulated in the previous section.

A. Finite-horizon Markov Chain with Continuous States

Consider a discrete-time Markov chain

$$\mathbf{X} := \{X(0), X(1), \cdots, X(T-1), X(T)\}, \tag{17}$$

with finite horizon T and a continuous state space $\mathcal{X} = \mathbb{R}^n$. Let $\mathcal{B}$ be the σ -field of $\mathbb{R}^n$ containing all the Lebesgue measureable sets. Given state $X(t) = x \in \mathbb{R}^n$ at time t, $t = 0, 1, \cdots, T-1$, the probability that the next state lie in a set $B \in \mathcal{B}$ at time t+1 can be denoted as a state transition function $P_t(B \mid x)$ which satisfies

$$P_t(\mathbb{R}^n \mid x) = \int_{\mathbb{R}^n} P_t(dy \mid x) = 1, \forall x \in \mathbb{R}^n\ . \tag{18}$$

Without further specification, we assume that all sets and functions discussed in this paper are Lebesgue measurable. Define a linear right operator $\mathbf{P}_t$ corresponding to $P_t(B \mid x)$ on the function space:

$$\mathbf{P}_t h(x) := \int_{\mathbb{R}^n} h(y) P_t(dy \mid x), \tag{19}$$

where $h(y)$ is any measurable function. The product of any two operators $\mathbf{P}_t$ and $\mathbf{P}_t^{'}$ is defined as: $\forall x \in \mathcal{X}, B \in \mathcal{B}$,

$$(\mathbf{P}_t \mathbf{P}_t^{'})(B \mid x) = \int_{\mathbb{R}^n} P_t^{'}(B \mid y) P_t(dy \mid x). \tag{20}$$

A probability measure $\nu(B)$ itself can be viewed as a special state transition function $\nu(B \mid x)$ which takes the same value $\nu(B)$ for all $x \in \mathbb{R}^n$. Thus, any probability measure $\nu(B)$ can be viewed as an operator ν. The state distribution at time t is denoted as β_t. With given initial state distribution β_0, it can be obtained as

$$\beta_t = \beta_{t-1}\mathbf{P}_{t-1}, \forall t = 1, 2, \cdots, T\ . \tag{21}$$

Let $f_t(x)$ be a cost function at time t with respect to state x. The total cost of the Markov chain (17) over finite horizon T is

$$\eta(x) = E\left\{\sum_{t=0}^{T} f_t(X(t)) \mid X(0) = x\right\}. \tag{22}$$

B. Potentials and Performance Sensitivity Formula

The potential (also known as relative value function [13] or cost-to-go function [15]) for state x at time t is obtained by

$$\begin{aligned} &g_T(x) = f_T(x), \\ &g_t(x) = f_t(x) + \mathbf{P}_t g_{t+1}(x), \forall t = 0, \cdots, T-1. \end{aligned} \tag{23}$$

Let (P_t, f_t) and $(P_t^{'}, f_t^{'})$ be the transition functions and performance functions at time t of two Markov chains with the same state space $\mathcal{X} = \mathbb{R}^n$. Let η, g_t, β_t and $\eta', g_t^{'}, \beta_t^{'}$ be their corresponding total performances, potential functions, and the state distribution at time t, respectively. Then we have

the following results.

Lemma 1*: The total performance difference formula for the above two Markov chains is*

$$\eta' - \eta = \sum_{t=0}^{T-1} \beta_t' \left[\left(f_t' - f_t \right) + \left(\mathbf{P}_t' - \mathbf{P}_t \right) g_{t+1} \right]. \quad (24)$$

Proof: In an extended journal version of this paper due to the page limits. □

Lemma 1 provides a neat way to calculate the difference of the total performance between two Markov chains. It has a salient feature to separate the efforts from these two chains so that one chain places its influence on the total difference through state distribution β_t' while the other places its influence on the potential function g_{t+1}.

From Lemma 1, the total performance derivative formula can be derived as follows. Suppose the transition function and performance function depend on parameters $\theta(t)$, $t = 0, 1, \ldots, T\text{-}1$, and are denoted as P_t^θ and f_t^θ, respectively. Denoting the operator as $\mathbf{P}_t^\theta$, we have the following result.

Theorem 1*: The total performance derivative with respect to* $\theta(t)$, $t = 0, 1, \ldots, T\text{-}1$, *is*

$$\frac{\partial \eta}{\partial \theta(t)} = \beta_t^\theta \left[\left(\frac{\partial \mathbf{P}_t^\theta}{\partial \theta(t)} \right) g_{t+1} + \left(\frac{\partial f_t^\theta}{\partial \theta(t)} \right) \right]. \quad (25)$$

Proof: From (24) in Lemma 1, (25) is easily proved. □

In Theorem 1, the performance derivative can be obtained by simulating the system since it only depends on the original parameters $\theta(t)$, $t = 0, 1, \ldots, T\text{-}1$. This is consistent with the theory of perturbation analysis [21][16] that we can obtain local information including the gradient by observing and analyzing a sample path of the original system. Therefore, gradient based policy optimization approach (explained in the next subsection) can be carried out for the finite horizon total cost MDPs with continuous states and action spaces.

C. Gradient-based Performance Optimization

With performance derivative formula (25), gradient-based policy optimization approach is developed in this subsection.

Algorithm 1*: (The gradient-based optimization approach).*

Step 1: *(Initialization)*. Randomly generate an initial policy with parameter $\theta^0(t)$ for $t = 0, 1, \ldots, T\text{-}1$, (or use the practical policy or heuristic policy as the starting point), and set the iteration number $k = 0$.

Step 2: *(Policy evaluation)*. Let $\theta^k(t)$, $t = 0, 1, \ldots, T\text{-}1$, be the parameters in iteration k. Do the simulation under parameter $\theta^k(t)$ and obtain the potential functions according to (23).

Step 3: *(Policy improvement)*. Obtain gradient $\nabla \eta(\theta^k(t))$ through the performance derivative formula (25). Update the policy according to the gradient-based methods in [26] with steepest descent direction and diminishing step size:

$$\theta^{k+1}(t) = \theta^k(t) - \alpha^k \nabla \eta(\theta^k(t)), \quad (26)$$

where step size $\alpha^k = 1/k$.

Step 4: *(Stopping criteria)*. If $\| \nabla \eta(\theta^k(t)) \| \le \epsilon$ for $t = 0, 1, \ldots, T\text{-}1$, stop; otherwise, let $k = k + 1$ and go to step 2.

Algorithm 1 has the following property.

Proposition 1*: Algorithm 1 converges to a stationary point of the total performance.*

Proof: Since the step size $\alpha^k = 1/k$, we have $\lim_{k\to\infty} \alpha^k = 0$ and $\sum_{k=0}^{\infty} \alpha^k = \infty$. From [26], Proposition 1 is proved. □

Algorithm 1 provides a systematic and general way to optimize the parameterized policies and it can be carried out online. In online version, to evaluate a policy, we estimate potentials by averaging samples over multiple sample paths.

IV. Numerical Examples

A. Optimization for Water Resources Planning Policies

The approach derived in Section III is applied to solve the water resources planning problem. Suppose the natural inflow to the reservoir i at time t, $\xi_i(t)$, has a normal distribution with a mean of $\mu_i(t)$ and a standard deviation of $\sigma_i(t)$, and truncated within $[\underline{\xi_i}(t), \overline{\xi_i}(t)]$. To simplify the discussion, we assume that the natural inflows to different reservoirs are independent (this might not be true in consideration of the geography and climate relations among reservoirs. But it is beyond our specialties and we use this assumption as a starting point to simplify the discussion here). However, our results are not limited to the independency and the normal distribution. In fact, the approach explained in this section is applicable as long as the PDF (probability density function) is known. We let

$$\mu(t) = (\mu_1(t), \mu_2(t), \cdots, \mu_I(t))^T, \quad (27)$$

and

$$\Sigma(t) = \begin{pmatrix} \sigma_1^2(t) & 0 & \cdots & 0 \\ 0 & \sigma_2^2(t) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_I^2(t) \end{pmatrix}. \quad (28)$$

Given system state $X(t)$ and action $A(t)$ at stage t, we can obtain the distribution of state $X(t+1)$ at stage t+1 as follows. According to the dynamic in (1), given system state $x_i(t)$ and actions $u_i(t)$, $w_i(t)$, and $s_i(t)$ for $i = 1, 2, \ldots, I$, $x_i(t+1)$ is also a normal distribution with a mean of $\mu_i^x(t)$, a standard deviation of $\sigma_i(t)$, and truncated within the interval of $[\Delta_i(t) + \underline{\xi_i}(t), \Delta_i(t) + \overline{\xi_i}(t)]$, where

$$\mu_i^x(t) = \Delta_i(t) + \mu_i(t), \quad (29)$$

$$\Delta_i(t) = x_i(t) - u_i(t) - r_i(t) + \sum_{j \in \mathcal{U}_i} [r_j(t) + \lambda_j u_j(t)]. \quad (30)$$

According to the dynamic in (2), $m_i^d(t+1)$ is deterministic given state $m_i^d(t)$ and action $u_i(t)$ for $i = 1, 2, \ldots, I$. Then, according to (19), we have

$$\mathbf{P}_t g_{t+1}(X(t)) = \int_{\Psi^I} g_{t+1}(Y) \frac{1}{(2\pi)^{\frac{I}{2}} |\Sigma(t)|^{\frac{1}{2}}} \cdot \exp\left\{-\frac{1}{2}\left(y-\mu^x(t)\right)^T \Sigma(t)^{-1}\left(y-\mu^x(t)\right)\right\} dy \quad (31)$$

where

$$\mu^x(t) = \left(\mu_1^x(t), \mu_2^x(t), \cdots, \mu_I^x(t)\right)^T, \quad (32)$$

$$y = \left(x_1(t+1), x_2(t+1), \cdots, x_I(t+1)\right)^T, \quad (33)$$

$$Y = X(t+1) = \left(y^T, m_1^0(t+1), \cdots, m_I^{D-1}(t+1)\right)^T, \quad (34)$$

$$\Psi^I \subset \mathbb{R}^I : \Delta_i(t) + \underline{\xi}_i(t) \le x_i(t+1) \le \Delta_i(t) + \overline{\xi}_i(t), \quad \forall i = 1, 2, \cdots, I. \quad (35)$$

Based on (31), for i = 1, 2, …, I, we obtain

$$\frac{\partial \mathbf{P}_t g_{t+1}(X(t))}{\partial u_i(t)} = \int_{\Psi^I} g_{t+1}(Y) \frac{1}{(2\pi)^{\frac{I}{2}} |\Sigma(t)|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{I} \frac{(y_i - \mu_i^x(t))^2}{(\sigma_i(t))^2}\right\} \cdot \left(\frac{-(y_i - \mu_i^x(t))}{(\sigma_i(t))^2} + \sum_{j \in \mathcal{D}_i} \frac{\lambda_i (y_j - \mu_j^x(t))}{(\sigma_j(t))^2}\right) dy \quad (36)$$

and

$$\frac{\partial \mathbf{P}_t g_{t+1}(X(t))}{\partial r_i(t)} = \int_{\Psi^I} g_{t+1}(Y) \frac{1}{(2\pi)^{\frac{I}{2}} |\Sigma(t)|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{I} \frac{(y_i - \mu_i^x(t))^2}{(\sigma_i(t))^2}\right\} \cdot \left(\frac{-(y_i - \mu_i^x(t))}{(\sigma_i(t))^2} + \sum_{j \in \mathcal{D}_i} \frac{(y_j - \mu_j^x(t))}{(\sigma_j(t))^2}\right) dy \quad (37)$$

where $y_i = x_i(t+1)$; $\mathcal{D}_i$ is the set of direct downstream plants of hydro plant i, for i = 1, 2, …, I.

Given the release r_t^i for i = 1, 2, …, I, t = 0, 1, …, T-1, the optimal allocation between discharge and spillage is

$$w_t^i = \min\{r_t^i, \overline{w}^i\}, \ s_t^i = r_t^i - w_t^i, \quad (38)$$

which can be proved according to the reward structure in (8) and (9). Therefore, we only need to derive the performance derivatives with respect to u_t^i and r_t^i for i = 1, 2, …, I, t = 0, 1, …, T-1. According to (8) and (9), we have

$$\frac{\partial f_t(X(t), A(t))}{\partial u_i(t)} = 0 ; \quad (39)$$

and

$$\frac{\partial f_t(X_t, A_t)}{\partial r_t^i} = \rho^i (2a_t P_t + b_t) \quad (40)$$

for $r_t^i \le \overline{w}^i$, and

$$\frac{\partial f_t(X_t, A_t)}{\partial r_t^i} = 0 \quad (41)$$

for $r_t^i > \overline{w}^i$.

Therefore, from these sensitivity formulas (36) - (41), together with Theorem 1, we can apply the sensitivity-based optimization approach, i.e., Algorithm 1, to solve this water resources planning problem.

B. Numerical Tests

We compare the results from the sensitivity-based optimization with the traditional backward induction method based on discretization of the state space and action space. Here we randomly give an initial policy as the starting point, and then use Algorithm 1 developed in Section III to optimize it. The comparison results of the expected total rewards and CPU time are shown in Table I. All the models and algorithms are implemented in Matlab and all runs are performed on a Windows PC with 2.93 GHz CPU and 2.00 GB RAM.

TABLE I. COMPARISON WITH DISCRETIZED BACKWARD INDUCTION

Case #	Sensitivity Reward	Sensitivity CPU (s)	Backward Induction Reward	Backward Induction CPU (s)
1	9805.63	20.52	4205.50	441.35
2	3904.69	20.75	2709.94	782.89
3	17319.31	277.95	11752.77	823.64
4	5265.79	177.45	2374.13	912.50
5	6384.18	281.40	3213.27	11104.58
6	10899.75	769.70	9663.22	11689.85
7	3928.16	360.32	3032.81	19496.53
8	4259.87	86.28	2263.57	19748.62
9	8252.87	17.12	6471.12	22629.96
10	14664.88	17.52	12809.29	22762.62

The testing results illustrate that the sensitivity-based optimization approach is effective and efficient when solving the water resources planning problem. Firstly, the sensitivity-based optimization approach is more efficient than the backward induction method, and the tests show that it needs less than 1/10 computational efforts on average. The reason for the time-consuming backward induction is due to the curse of dimensionality when discretizing the state and action spaces. Secondly, the results obtained from the sensitivity-based optimization can on average improve the expected total rewards by more than 3/5, and in some cases, it can double the total rewards, in comparison with the backward induction. The backward induction does not perform well since it has to deal with the dilemma of the approximation precision and computational efforts. Only in cases #6 and #10, the results of these two methods are close, since the discretization points in the backward induction happens to be near the optimal solutions.

TABLE II. COMPARISON WITH THE SDDP METHOD

Case #	Sensitivity Reward	Sensitivity CPU (s)	SDDP Reward	SDDP CPU (s)
1	9805.63	20.52	9054.76	332.29
2	3904.69	20.75	4328.65	637.38
3	17319.31	277.95	16211.89	643.94
4	5265.79	177.45	3582.81	709.68
5	6384.18	281.40	6889.84	3692.24
6	10899.75	769.70	9458.86	1604.28
7	3928.16	360.32	4176.30	1665.66
8	4259.87	86.28	4615.28	1685.61
9	8252.87	17.12	7933.20	3656.18
10	14664.88	17.52	14296.42	3843.95

We also compare our approach to stochastic dual dynamic programming (SDDP), an important class of methods for solving hydro-thermal planning problems [29]. The results are compared in Table II, and they show that the SDDP method needs more computational efforts on average than the sensitivity-based optimization approach. The reason is that the discretization of the probability space is still necessary in SDDP because it captures the problem randomness with scenarios [29], and the total number of scenarios grows exponentially with respect to the problem scale.

V. Conclusions

In this paper, the finite-horizon MDP with continuous state and action spaces is introduced to depict the stochastic nature of system dynamics and the coupling of hydraulic constraints of the long term water resource planning problem. The governmental regulations on the annual water consumption and release are formulated by introducing the accumulated state variables in the MDP. A sensitivity based optimization approach is developed for solving the finite-horizon MDP with continuous state and action spaces. A performance derivative for the finite horizon continuous MDP is derived and a gradient-based algorithm is developed to solve the MDP problem. The salient feature of the new method is that the performance is improved systematically and there is no need to determine the difficult trade-off between the computational accuracy and efficiency. It is demonstrated that the model proposed in the paper can effectively capture the governmental regulations on annual water consumption and release. Numerical testing results show that the sensitivity based approach is effective and efficient for solving the water resources planning problem. This approach is also applicable for solving other MDP problems with practical sizes. Stochastic system demands and the time dependency of natural inflows will be considered in our future work.

Acknowledgment

The authors would like to thank Prof. Yu-Chi Ho, Prof. Xi-Ren Cao, Prof. Qianchuan Zhao, and the anonymous reviewers for their valuable advices and suggestions.

References

[1] M. Christoforidis, M. Aganagic, B. Awobamise, S. Tong, and A. F. Rahimi, "Long-Term/Mid-Term Resource Optimization of A Hydrodominant Power System Using Interior Point Method," *IEEE Transactions on Power Systems*, Vol. 11(1), pp. 287-294, 1996.

[2] L. Martinez and S. Soares, "Comparison Between Closed-Loop and Partial Open-Loop Feedback Control Policies in Long Term Hydrothermal Scheduling," *IEEE Transactions On Power Systems*, Vol. 17(2), pp. 330-336, 2002.

[3] A.H. Mantawy, S.A. Soliman, and M.E. El-Hawary, "The Long-Term Hydro-Scheduling Problem - A New Algorithm," *Electric Power Systems Research*, Vol. 64(1), pp. 67-72, 2003.

[4] R. Fuentes-Loyola and V. H. Quintana, "Medium–Term Hydrothermal Coordination by Semidefinite Programming," *IEEE Transactions on Power Systems*, Vol. 18, No. 4, November 2003.

[5] M.-Y. Tu, N.-S. Hsu, and W. W.-G. Yeh, "Optimization of Reservoir Management and Operation with Hedging Rules," *Journal of Water Resources Planning and Management*, Vol. 129(2), pp. 86-97, 2003.

[6] R. Oliveira and D.P. Loucks, "Operating rules for multi-reservoir systems," *Water Resources Research*, Vol. 33(4), pp. 839-852, 1997.

[7] B.F. Lamond and A. Boukhtouta, "Neural Approximation for the Optimal Control of a Hydroplant with Random Inflows and Concave Revenues," *Journal of Energy Engineering*, Vol. 131(1), pp. 72-95, 2005.

[8] B.F. Lamond and A. Boukhtouta, "Optimizing Long-term Hydro-power Production Using Markov Decision Process," *International Transactions in Operational Research*, Vol. 3, 1996.

[9] T.W. Archibald, C.S. Buchanan, L.C. Thomas, and K.I.M. McKinnon, "Controlling multi-reservoir systems," *European Journal of Operational Research*, Vol. 129(3), pp. 619-626 , 2001.

[10] T.W. Archibald, K.I.M. McKinnon, and L.C. Thomas, "An aggregate stochastic dynamic programming model of multi-reservoir systems", working paper, University of Edinburgh Department of Business, 1996.

[11] A. Turgeon, "Decomposition Method for the Long-Term Scheduling of Reservoirs in Series," *Water Resources Res*. Vol. 17(6), pp. 1565-1570, 1981.

[12] A.J. Wood and B.F. Wollenberg, *Power Generation Operation and Control,* New York : J. Wiley & Sons, 1996.

[13] M.L. Puterman, *Markov Decision Process: Discrete Stochastic Dynamic Programming*, John Wiley & Sons, Inc., New York, 1994.

[14] X.-R. Cao, "A Unified Approach to Markov Decision Problems and Performance Sensitivity Analysis," *Automatica*, Vol. 36, 771-774, 2000.

[15] D.P. Bertsekas, *Dynamic Programming and Optimal Control*, Belmont, Mass. : Athena Scientific, 2005.

[16] X.-R. Cao, "*Stochastic Learning and Optimization - a Sensitivity-Based Approach*". Springer, 2007.

[17] Y. Zhao, Q. Zhao, Q.-S. Jia, X. Guan, X.-R. Cao, "Event-Based Optimization for Dispatching Policies in Material Handling Systems of General Assembly Lines", In: *Proceeding of the 47th IEEE Conference on Decision and Control*, pp. 2173-2178, 2008.

[18] E. Ni, X. Guan, R. Li, "Scheduling Hydrothermal Power Systems with Cascaded and Head-Dependent Reservoirs," *IEEE Transactions on Power Systems*, Vol. 14(3), pp. 1127-1132, 1999.

[19] X.-R. Cao and H. Chen, "Perturbation realization, potentials, and sensitivity analysis of Markov processes," *IEEE Transactions on Automatic Control*, Vol. 42, no. 10, pp. 1382–1393, 1997.

[20] X.-R. Cao, "From perturbation analysis to Markov decision processes and reinforcement learning," Discrete Event Dynamic Systems, vol. 13, no. 1, pp. 9–39, 2003.

[21] Y.-C. Ho and X.-R. Cao, *Perturbation analysis of discrete event dynamic systems*. Boston : Kluwer Academic Publishers, 1991.

[22] R. Sutton and A. Barto, *Reinforcement learning: An introduction*. MIT press, 1998.

[23] K.-J. Zhang, Y.-K. Xu, X. Chen and X.-R. Cao, Policy Iteration Based Feedback Control, *Automatica*, Vol. 44, pp. 1055-1061, 2008.

[24] Y. Zhao, Q. Zhao, and X. Guan, "Stochastic Optimal Control for A Class of Manufacturing Systems Based on Event-Based Optimization," the *3rd Japan-China Joint Workshop on Control*, Fukuoka, Japan, 18 August, 2009.

[25] Y. Zhao, Q. Zhao, and X. Guan, "Event-Based Optimization for Finite-Horizon Total-Cost Markov Decision Processes," under review.

[26] D.P. Bertsekas, *Nonlinear programming*, Athena Scientific, 1999.

[27] S. Yakowitz, "Dyanmic Programming Applications in Water Resources," *Water Resources Research*, Vol. 18(4), pp. 673-696, 1982.

[28] Y. Zhao, X. Chen, Q.S. Jia, X. Guan, S. Zhang, Y. Jiang, "Long-term Scheduling for Cascaded Hydro Energy Systems with Annual Water Consumption and Release Constraints," *IEEE Transactions on Automation Science and Engineering*, to appear, 2010.

[29] M.V.F. Pereira, and L.M.V.G.Pinto, "Multi-stage Stochastic Optimization Applied to Energy Planning," Mathematical Programming, Vol. 52(1), pp. 359-375, 1991.

[30] Z. Yu, F. T. Sparrow, and B. H. Bowen, "A New Long-Term Hydro Production Scheduling Method for Maximizing the Profit of Hydroelectic Systems," *IEEE Transactions on Power Systems*, Vol. 13(1), pp. 66-71, 1998.

[31] F.R. Forsund, *Hydropower economics*, International Series in Operations Research & Management Science, Springer, 2007.