Computers & Operations Research 40 (2013) 2760–2769

Contents lists available at SciVerse ScienceDirect

Computers & Operations Research

journal homepage: www.elsevier.com/locate/caor

A simulation-and-regression approach for stochastic dynamic programs with endogenous state variables ☆

Michel Denault [a,b,*], Jean-Guy Simonato [c], Lars Stentoft [c]

[a] *HEC Montréal (Management Sciences), Canada*
[b] *GERAD, Canada*
[c] *HEC Montréal (Finance), Canada*

ARTICLE INFO

Available online 25 April 2013

Keywords:
Stochastic control
Approximate dynamic programming
Simulation and regression
Least-squares Monte Carlo
Hydropower management

ABSTRACT

We investigate the optimum control of a stochastic system, in the presence of both exogenous (control-independent) stochastic state variables and endogenous (control-dependent) state variables. Our solution approach relies on simulations and regressions with respect to the state variables, but also grafts the endogenous state variable into the simulation paths. That is, unlike most other simulation approaches found in the literature, no discretization of the endogenous variable is required. The approach is meant to handle several stochastic variables, offers a high level of flexibility in their modeling, and should be at its best in non time-homogenous cases, when the optimal policy structure changes with time. We provide numerical results for a dam-based hydropower application, where the exogenous variable is the stochastic spot price of power, and the endogenous variable is the water level in the reservoir.

1. Introduction

We propose a new solution approach for stochastic dynamic control problems involving both exogenous and endogenous state variables, i.e. control-independent and control-dependent variables. The approach relies on simulations and regressions (also known as LSMC, for Least Squares and Monte Carlo) to approximate the value function and the underlying conditional expectations.

Stochastic control problems with endogenous and exogenous variables are found in a wide variety of engineering and management applications. For example, in finance, the portfolio optimization problem is characterized by the endogenous wealth of the investors which depends on the choices about the weights of the exogenous random stock returns. Another example is nonrenewable resource management, in which the remaining endogenous level of the resource depends on the extraction or production decisions driven by an exogenous random price. Hydro power production is another example in which the endogenous level of the water reservoir depends on the electricity production driven by an exogenous random price level.

The main features of the type of problems we consider, are explained in terms of the control (the decisions) and the two types of state variables:

- the evolution of the endogenous variable through time depends on the decisions;
- there is great value in making the right decisions at the right time;
- the current decisions have an impact on the availability of decisions (options) later;
- the decisions are made with incomplete information about the future, given the stochastic, exogenous state variables.

Recently, gas storage has been an important development vector for stochastic dynamic programs with endogenous and exogenous variables (most authors explicitly recognize the applicability of their methods beyond gas storage). The endogenous variable is the level of gas in storage and the exogenous variable is the gas price. We cite for now Chen and Forsythe [5,6], and Thompson et al. [20], who solve the control problems through partial differential equations techniques. Such methods are typically efficient but less flexible than simulation-based techniques.

The use of simulations and regressions to solve stochastic dynamic programming (DP) problems goes back at least to Keane and Wolpin [11]. The approach became very popular in financial engineering after the papers of Longstaff and Schwartz [14] and Tsitsiklis and Van Roy [22], who set option pricing as an optimal stopping time problem.

Generalizing from financial option pricing, the method of simulations and regressions was extended to gas storage by Boogert and de Jong [2] and Carmona and Ludkovski [4] while Felix and Weber [10] use simulations and trees. In contrast to our approach, Boogert and de Jong discretize the gas level, and run a

☆The authors are thankful to Michèle Breton, Matt Davison, Pierre L'Écuyer and Ilya Ryshov for their helpful comments, and to Mathieu Rousseau and Sofiane Tafat for their programming help.

* Corresponding author. Tel.: +1 1 514 340 7161.
E-mail address: michel.denault@hec.ca (M. Denault).

0305-0548/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cor.2013.04.008

separate regression for each gas level. Carmona and Ludkovski introduce a "quasi-simulation" of the endogenous variable and regress on both the exogenous and the endogenous variables; this is also the starting point of our approach. Our algorithm differs from that of [4] on three main counts. First, we build forward-optimal endogenous variable paths by relying on existing policies, while the authors do a guessing step but provide little detail on guiding the guess. Second, we propose a method to account for the impact of upper and lower bounds of the endogenous variable on the regressions. Finally, we suggest a rather natural approach to solve the problem of *clustering*, cf. [4, p. 366].

Stochastic dynamic programming methods based on simulations and regressions fall in the broad area of Approximate Dynamic Programming (ADP), that is, dynamic programming methods that rely on approximations of the value function or the policies. Two main references for ADP are the books of Bertsekas [1] and Powell [17]. An approximate dynamic programming method without either parametric function fitting nor simulation is proposed by Lai, Margot and Secomandi [12] for the gas storage problem; see also the related paper by Secomandi [18]. Nascimento and Powell [16] bring another approach, using simulations and parametric function approximations but still quite different from the papers mentioned above. Indeed, the authors do not work on a full set of simulation paths, but rather rely on iterative improvements of the value function using the forward ADP approach (see [17]). Our focus is on stochastic problems, but it is interesting to note that our time-discretization strategy shares some similarities with the deterministic control parametrization methods of Loxton et al. [15] and Lin et al. [13].

Although the solution method we propose applies in general to stochastic control problem with exogenous variables and (bounded) endogenous variables, we refer throughout the paper to the management of a dam-based hydropower system to support the intuition. Our motivation to depart from gas storage is that multi-scale (short and long term) storage and supplementary stochastic variables such as water inflows are natural features of the hydro problem, to be explored in subsequent work.

The approach that we propose can be applied to a broad number of problems, from supply chain management to oil and mine exploitation. Moving away from the storage of physical assets, one could take capital as the endogenous variable, and tackle portfolio management optimization, or underground exploration rights acquisition.

In the next section, we formulate the problem. We describe our approach in Section 3, and briefly discuss convergence results from the literature in Section 4. Numerical evidence is presented in Section 5; the paper concludes with Section 6. Appendix A provides the details of our Markov-chain benchmark approach.

2. Problem description

In this section, we describe the problem's main characteristics, indicate its underlying hypotheses and set the notation up.

We consider a stochastic control problem where a state variable varies (deterministically) with time according to the control choices that are made; we call such a variable "endogenous". The payoff during each period of time is determined by the control choice and a stochastic, exogenous state variable. Although the set up is completely general, we believe that it helps the intuition to refer to specific variables in describing the method. We will thus consider throughout the paper the case of a dam-based hydropower producer: the endogenous variable is the amount of energy (as water) behind the dam, the exogenous variable is the spot price of power, and the control is the amount of power to produce and sell at each moment. Improvements to this bare-bones case are discussed later.

We shall frame our problem within a finite horizon, with a discrete number of time periods during which neither the exogenous variable nor the control decision can change. The control space is also finitely discretized.

The components of our model can be defined as follows:

Time as a moment is represented by $t \in [0, T]$, though t is also used to represent the *period* from (time point) t to $t+1$, a duration we simply call "one time period". Context should relieve any possible confusion between time point and time period.

The exogenous state variable S_t, which can be thought of as a spot power price during period t. It is revealed at the beginning of each period.

The endogenous state variable L_t, which can be thought of as the amount of power (contained as water) in the reservoir. Simple operational constraints are given by bounds $L_{\min} \le L_t \le L_{\max}$: control decisions that would bring L_t beyond those bounds are not allowed.

The above are the state variables of our dynamic programming formulation. The control variable is u, defined immediately; with it, we can also define the payoff associated to control decisions.

Control variable u_t directly influences the endogenous variable level, and partly determines the payoff. As mentioned, not all controls may be available, depending on the endogenous variable level. We denote by $\mathcal{U}(S, L, t)$ the set of admissible regimes when starting with state values (S,L) and for period t. The control can be thought of as the amount of power produced and sold. In the simplest case, we consider only three regimes and $u_t \in \{+1, 0, -1\}$. Depending on the application, u_t may or may not be sign-constrained; we come back later to the meaning of a negative u_t in our dam context.

Payoff $\pi_t(u; S, L)$ is the cash-flow obtained when running regime u during period t, starting with state variables values (S,L). (Throughout the paper, we leave aside time discounting, without loss of generality). A typical payoff function would be

$$\pi_t(u_t; S_t, L_t) = q u_t S_t \phi(L_t)$$

where function ϕ can be used to account for operational constraints on the endogenous variable, and fixed scalar q is used to scale the unitless decision value into an amount of energy.

We can now set an objective for the control problem, which is to find the *value function*

$$V_t(S, L) \equiv \sup_{u \in \mathcal{U}(S,L,t)} \mathbb{E}\left\{ \sum_{s=t}^{T-1} \pi_s(u; S_s, L_s) + W(S_T, L_T) | (S_t, L_t) = (S, L) \right\} \quad (1)$$

where $W(S, L)$ is the terminal value of the system with the given values of the state variables. The deterministic state variable L_t abides the state equation

$$L_{t+1} = h(u_t; L_t); \quad (2)$$

note that given time-t information and decision, the level of L_{t+1} is known. The stochastic state variable follows a Markov transition probability distribution $p_S(S_{t+1} | S_t)$.

The underlying optionality of the problem is as follows. The level of the endogenous variable influences which control decisions, and thus which payoffs, are available. But the control decision at time t influences the endogenous variable level for all future time periods, making some future payoffs unavailable. For example, water not used now may be used later, but at a power price unknown at the present time.