Rewards, Costs and Flexibility in Dynamic Resource Allocation: A Stochastic Optimal Control Approach

X. Gao*, Y. Lu†, M. Sharma†, M.S. Squillante†, J.W. Bosman‡
* H. Milton Stewart School of ISyE, Georgia Institute of Technology, Atlanta, GA 30332, USA
† Mathematical Sciences Dept, IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA
‡ Centrum Wiskunde & Informatica, 1098 XG Amsterdam, The Netherlands

1. INTRODUCTION

Various canonical forms of general resource allocation problems arise naturally across a broad spectrum of computer systems and communication networks. As the complexities of these systems and networks continue to grow, together with ubiquitous advances in technology, new approaches and methods are required to effectively and efficiently solve these problems. Such environments often consist of different types of resources that are allocated in combination to serve demand whose behavior over time is characterized by different types of uncertainty and variability. Each type of resource has a different reward and cost structure that ranges from the best of a set of primary resource allocation options, having the highest reward, highest cost and highest net-benefit, to a secondary resource allocation option, having the lowest reward, lowest cost and lowest net-benefit. Each type of resource also has different structures for the flexibility and cost of making changes to the allocation capacity. The resource management optimization problem we consider consists of adaptively determining the primary and secondary resource allocation capacities that serve the uncertain demand and that maximize the expected net-benefit over a time horizon of interest based on the foregoing reward, cost and flexibility structural properties of the different types of resources.

The general class of resource allocation problems studied in this paper arises in a wide variety of application domains such as cloud computing and data center environments, computer and communication networks, and energy-aware and smart power grid environments, among many others. Across these and many other domain-specific resource allocation problems, there is a common need for the dynamic adjustment of allocations among multiple types of resources, each with different structural properties, to satisfy time-varying and uncertain demand. Taking a financial mathematics approach that hedges against future risks associated with resource allocation decisions and uncertain demand, we consider the underlying fundamental stochastic optimal control problem where the dynamic control policy that allocates primary resource capacities to serve uncertain demand is a variational stochastic process with conditions on its derivative, which in turn determines the secondary resource allocation capacity. The objective is to maximize the expected discounted net-benefit over time based on the structural properties of the different resources types, which we show to be equivalent to a minimization problem involving a piecewise-linear running cost and a proportional cost for making adjustments to the control policy process. Our solution approach is based on first deriving twice continuously differentiable properties of the value function at the optimal free boundary to determine a solution of the Hamilton-Jacobi-Bellman (HJB) equation, namely the smooth-fit principle. Our theoretical results also include an explicit characterization of the dynamic control policy, which is of threshold type, and then we verify that this control policy is optimal through a martingale argument. In contrast to an optimal static allocation strategy, our results establish that the optimal dynamic control policy adapts its allocation decisions in primary and secondary resources to hedge against the risks of under allocating primary resource capacity (resulting in lost reward opportunities) and over allocating primary resource capacity (resulting in incurred cost penalties).

The research literature covers a great diversity of resource allocation problems, with differing objectives, policies, rewards and costs. From a methodological perspective, the general resource allocation problem we consider in this paper is closely related to the vast financial mathematics literature on solving stochastic control problems for investment and capacity planning; refer to, e.g., [6]. For example, Beneš et al. [1] consider the so-called *bounded velocity follower problem* with a quadratic running cost objective function, where the authors propose a smooth-fit principle to characterize the optimal policy. In comparison with our study, however, the paper does not consider any costs associated with the actions taken by the control policy, and deals with a much simpler and smoother objective function. From an applications perspective, there is a growing interest in the computer system and communication network communities to address allocation problems involving various types of resources associated with computation, memory, bandwidth and/or energy. For example, Lin et al. [5] consider the problem of dynamically adjusting the number of active servers in a data center as a function of demand to minimize operating costs. In comparison with our study, however, the paper considers average demand over small intervals of time, subject to system constraints, and develops an online algorithm that is proven to be 3-competitive.

Our study provides important methodological contributions and new theoretical results by deriving the solution of a fundamental singular stochastic optimal control problem. This stochastic optimal control solution approach highlights the importance of timely and adaptive decision making in the allocation of a mixture of different resource options with distinct features in optimal proportions to satisfy time-varying and uncertain demand. Our study also provides important algorithmic contributions through a new class of online policies for dynamic resource allocation problems arising across a wide variety of application domains. Extensive numerical

A portion of this research was sponsored by the US Army Research Laboratory and the UK Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the US Army Research Laboratory, the US Government, the UK Ministry of Defense, or the UK Government. The US and UK Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

experiments quantify the effectiveness of our optimal online dynamic control algorithm over recent work in the research literature.

2. MODEL AND FORMULATION

We consider a singular stochastic optimal control problem underlying a general class of resource allocation problems in which different types of resources are allocated to serve uncertain demand. A primary resource allocation option has the highest net-benefit and bounded rate of change at any instant of time, whereas a secondary resource allocation option has the lowest net-benefit and more flexibility in its rate of change. The demand process $D(t)$ is modelled by the linear diffusion process $dD(t) = bdt + \sigma dW(t)$, where $b \in \mathbb{R}$ is the demand growth/decline rate, $\sigma > 0$ is the demand volatility, and $W(t)$ is a one-dimensional standard Brownian motion; it is well-known that the sample paths of Brownian motion are nondifferentiable [4]. A control policy defines at every time $t \in \mathbb{R}$ the level of primary resource allocation, denoted by $P(t)$. The level of secondary resource allocation, denoted by $S(t)$, is set to serve the remaining portion of the demand not served by primary resource capacity, namely $S(t) = [D(t) - P(t)]^+$.

Let $R_p(t)$ and $C_p(t)$ respectively denote the reward and cost associated with the primary resource allocation capacity $P(t)$ at time t. The rewards $R_p(t)$ are linear functions of the primary resource capacity and demand, whereas the costs $C_p(t)$ are linear functions of the primary resource capacity. We therefore have $R_p(t) = \mathcal{R}_p \times [P(t) \wedge D(t)]$ and $C_p(t) = \mathcal{C}_p \times P(t)$, where $\mathcal{R}_p \geq 0$ captures all per-unit rewards for serving demand with primary resource capacity, $\mathcal{C}_p \geq 0$ captures all per-unit costs for primary resource capacity, and $\mathcal{R}_p > \mathcal{C}_p$. Hence, from a risk hedging perspective, the risks associated with the primary resource allocation position at time t concern lost reward opportunities whenever $P(t) < D(t)$ on one hand and concern incurred cost penalties whenever $P(t) > D(t)$ on the other hand. Similarly, the reward function $R_s(t)$ and cost function $C_s(t)$ for the secondary resource allocation capacity are given by $R_s(t) = \mathcal{R}_s \times [D(t) - P(t)]^+$ and $C_s(t) = \mathcal{C}_s \times [D(t) - P(t)]^+$, where $\mathcal{R}_s \geq 0$ captures all per-unit rewards for serving demand with secondary resource capacity, $\mathcal{C}_s \geq 0$ captures all per-unit costs for secondary resource capacity, and $\mathcal{R}_s > \mathcal{C}_s$. Hence, the secondary resource allocation position at time t is riskless in the sense that rewards and costs are both linear in the resource capacity actually used.

The singular stochastic optimal control problem allows the dynamic control policy to adapt its allocation positions in primary and secondary resource capacities based on the demand realization observed up to the current time. More formally, the decision process $P(t)$ is adapted to the filtration $\mathcal{F}_t$ generated by $\{D(s) : s \leq t\}$. Any adjustments to the primary resource capacity have associated costs, where we write $\mathcal{I}_p$ and $\mathcal{D}_p$ to denote the per-unit costs of increasing and decreasing the decision process $P(t)$, respectively. Then the objective of the optimal dynamic control policy is to maximize the expected discounted net-benefit over an infinite horizon, where net-benefit at time t consists of the difference between rewards and costs from primary and secondary resource allocation capacities minus the additional costs for adjustments to $P(t)$.

In formulating the corresponding stochastic optimization problem, we impose additional conditions on the variational decision process $\{P(t) : t \geq 0\}$ based on practical considerations. The control policy cannot instantaneously change the primary resource allocation capacity in an attempt to directly follow the demand $D(t)$; i.e., some time is required to adjust $P(t)$. Moreover, the control policy cannot make unbounded adjustments in the primary resource allocation capacity at any instant in time; i.e., the amount of change in $P(t)$ at time t is restricted by various factors. Given these practical considerations, we assume that the rate of change in the primary resource allocation capacity by the control policy is bounded. More precisely, there are two finite constants $\theta_\ell < 0$ and $\theta_u > 0$ such that $\theta_\ell \leq \dot{P}(t) \leq \theta_u$, where $\dot{P}(t)$ denotes the derivative of the decision variable $P(t)$ with respect to time.

Now we can present the mathematical formulation of our stochastic optimization problem in which we formally seek to determine the optimal dynamic control policy that maximizes the objective

$$\begin{aligned}\max_{\dot{P}(t)} \quad & \mathbb{E} \int_0^\infty e^{-\alpha t} \Big\{ [R_p(t) - C_p(t) + R_s(t) - C_s(t)] dt \\ & -[\mathcal{I}_p \mathbb{1}_{\{\dot{P}(t)>0\}}] dP(t) - [\mathcal{D}_p \mathbb{1}_{\{\dot{P}(t)<0\}}] d(-P(t)) \Big\}, \qquad (2.1)\end{aligned}$$

where α is the discount factor and $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function. Define $X(t) := P(t) - D(t)$, $\mathcal{N}_p := \mathcal{R}_p - \mathcal{C}_p$, and $\mathcal{N}_s := \mathcal{R}_s - \mathcal{C}_s$. Then, through a sequence of algebraic steps, we simplify the first term of the objective function (2.1) and derive the following equivalent stochastic optimization problem:

$$\begin{aligned}\min_{\dot{P}(t)} \quad & \mathbb{E}_x \left[\int_0^\infty e^{-\alpha t} \left\{ \left(\mathcal{C}_+ X(t)^+ + \mathcal{C}_- X(t)^- \right) dt \right.\right. \\ & \left.\left. + \left(\mathcal{I}_p \mathbb{1}_{\{\dot{P}(t)>0\}} - \mathcal{D}_p \mathbb{1}_{\{\dot{P}(t)<0\}} \right) dP(t) \right\} \right] \qquad (2.2) \\ \text{s.t.} \quad & -\infty < \theta_\ell \leq \dot{P}(t) \leq \theta_u < \infty, \qquad (2.3) \\ & dX(t) = dP(t) - bdt - \sigma dW(t), \qquad (2.4) \\ & X(0) = x, \quad \mathcal{C}_+ = \mathcal{C}_p, \quad \mathcal{C}_- = \mathcal{N}_p - \mathcal{N}_s, \qquad (2.5)\end{aligned}$$

where $\mathbb{E}_x[\cdot]$ denotes expectation with respect to an initial state of x with probability one. The control variable is the rate of change in the primary resource capacity by the control policy at every time t subject to the lower and upper bound constraints in (2.3).

We use $V(x)$ to represent the optimal value of the objective function (2.2); i.e., $V(x)$ is the value function of the corresponding stochastic dynamic program. The focus of our analysis will be on on the stochastic dynamic program formulation in (2.2) – (2.5).

3. MAIN RESULTS

In this section we consider our main results on the optimal dynamic control policy for the stochastic optimization problem (2.2) – (2.5) in the case $\mathcal{D}_p < \mathcal{C}_+/\alpha$ and $\mathcal{I}_p < \mathcal{C}_-/\alpha$, which is likely to be the most relevant in practice. All other possible cases of our main results based on different parameter conditions, together with the rigorous proofs of our main results, are provided in [3].

Let us briefly interpret the conditions of this case of practical interest. Observe from (2.2) that $\mathcal{C}_+/\alpha$ reflects the discounted overage cost associated with the primary resource capacity and $\mathcal{C}_-/\alpha$ reflects the corresponding discounted shortage cost, recalling that α is the discount rate. In comparison, $\mathcal{D}_p$ represents the cost incurred for decreasing $P(t)$ when in an overage position while $\mathcal{I}_p$ represents the cost incurred for increasing $P(t)$ when in a shortage position. We now state our main result for this case.

THEOREM 1. *Suppose $\mathcal{D}_p < \mathcal{C}_+/\alpha$ and $\mathcal{I}_p < \mathcal{C}_-/\alpha$. Then there are two threshold values L and U with $L < U$ such that the optimal dynamic control policy is given by*

$$\dot{P}(t) = \begin{cases} \theta_u, & \text{if} \quad P(t) - D(t) < L, \\ 0, & \text{if} \quad P(t) - D(t) \in [L, U], \\ \theta_\ell, & \text{if} \quad P(t) - D(t) > U. \end{cases}$$

Moreover, the optimal threshold values L and U are uniquely determined by two explicit nonlinear equations that depend upon the relationships among the model parameters (as provided in [3]).

PROOF SKETCH. Our proof proceeds in three main steps. First, from the Bellman principle of optimality and Ito's formula we derive the HJB equation for the value function $V(x)$:

$$-\alpha V(x) + \frac{1}{2}\sigma^2 V''(x) - bV'(x) + \mathcal{C}_+ x^+ + \mathcal{C}_- x^- + \inf_{\theta_\ell \le \theta \le \theta_u} \mathcal{L}(\theta, x) = 0, \qquad (3.1)$$

where

$$\mathcal{L}(\theta, x) = \begin{cases} (V'(x) + \mathcal{I}_p)\theta & \text{if } \theta \ge 0, \\ (V'(x) - \mathcal{D}_p)\theta & \text{if } \theta < 0. \end{cases} \qquad (3.2)$$

Second, a solution to the HJB equation is constructed by solving the equation in different regions of the domain where the HJB equation is reduced to an ordinary differential equation whose parameterized solutions can be explicitly obtained, and by then determining the parameters to match the solutions up to the first order of derivatives at the boundaries of these regions. Finally, we exploit a martingale argument to confirm that this construction indeed provides an optimal solution to the stochastic singular control problem. □

Theorem 1 can be explained as follows. The optimal dynamic control policy seeks to maintain $X(t) = P(t) - D(t)$ within the risk-hedging interval $[L, U]$ at all time t, taking no action (i.e., $\dot{P}(t) = 0$) as long as $X(t) \in [L, U]$. Whenever $X(t)$ falls below L, the optimal dynamic control policy pushes toward the risk-hedging interval as fast as possible, namely at rate θ_u. Whenever $X(t)$ exceeds U, the optimal dynamic control policy pushes toward the risk-hedging interval as fast as possible, namely at rate θ_ℓ.

4. NUMERICAL EXPERIMENTS

Theorem 1 establishes the explicit optimal dynamic control policy among all admissible nonanticipatory control processes $dP(t)$ that maximizes the stochastic dynamic program (2.2) – (2.5). This optimal dynamic control policy renders a new class of online algorithms for general dynamic resource allocation problems that arise in a wide variety of application domains. The resulting online algorithm is easily implementable at runtime in real-world systems and consists of maintaining $X(t) = P(t) - D(t)$ within the risk-hedging interval $[L, U]$ at all time t, where L and U are easily obtained in terms of the system parameters. Extensive numerical experiments have been conducted based on real-world trace data taken from a broad spectrum of computer system/network environments. Once the average daily demand pattern $f(t)$ is extracted from these traces, we calibrate our optimal online algorithm by partitioning the corresponding drift function $b(t) = df(t)$ into piecewise linear segments and computing the values of L and U for each per-segment b and σ according to Theorem 1. This (fixed) version of our optimal online algorithm is applied to each daily sample path of the Brownian demand process and the time-average value of net-benefit is computed over a set of $10,000$ daily sample paths.

For comparison under the same set of Brownian demand process sample paths, we consider two recent alternative optimization approaches, starting with the optimal offline algorithm in [5]. This algorithm consists of making optimal provisioning decisions in a clairvoyant anticipatory manner based on the known average demand within each slot of a discrete-time model where the slot length is chosen to match the timescale at which the system can adjust its capacity and so that demand activity within a slot is nonnegligible in a statistical sense. Applying this offline algorithm within our framework, we partition the daily time horizon into T slots of length γ such that $h_i = (t_{i-1}, t_i]$, $\gamma = t_i - t_{i-1}$, $t_0 := 0$, and compute the average demand within each slot $g_i := \gamma^{-1} \int_{h_i} f(t)dt$. Define $\Delta(P_i) := P_i - P_{i-1}$, where P_i denotes the primary resource allocation capacity for slot i. The optimal solution under this offline algorithm is then obtained by solving the following LP:

$$\begin{aligned} \min_{\Delta(P_1),\ldots,\Delta(P_T)} \quad & \sum_{i=1}^{T} \mathcal{C}_+(P_i - g_i)^+ + \mathcal{C}_-(P_i - g_i)^- \\ & + \mathcal{I}_p(P_i - P_{i-1})^+ + \mathcal{D}_p(P_i - P_{i-1})^- \quad (4.1) \\ \text{s.t.} \quad & -\infty < \theta_\ell \le \Delta(P_i)/\gamma \le \theta_u < \infty, (4.2) \end{aligned}$$

where (4.2) corresponds to (2.3). In this deterministic optimization problem, the control variable is the rate of change in the primary resource allocation for each slot i over the daily horizon. We refer to this solution as the offline LP algorithm. The second algorithm for comparison is the optimal online algorithm in [2], which is based on a similar discrete-time model framework and a distinct stochastic optimization framework. We refer to this as the CF algorithm.

Our numerical experiments demonstrate that the optimal online dynamic control algorithm outperforms both alternative optimization approaches for all $\sigma > 0$, where the relative improvements in expected net-benefit grow in an exponential manner with respect to increasing values of σ. This includes relative improvements of 90% and 150% for one representative real-world workload in comparison with the offline LP and online CF algorithms, respectively. For another representative real-world workload, the relative improvements corresponding to the offline LP and online CF algorithms are respectively 130% and 230%. Note that the rewards and costs comprising the net-benefit associated with the primary and secondary resource capacities can be based on either performance or financial metrics, or a combination of both. The significant relative improvements under the optimal online dynamic control algorithm follow from our stochastic optimal control approach that directly addresses the volatility of the demand process in all primary and secondary resource allocation decisions. Moreover, our optimal online algorithm can exploit any consistent seasonal patterns for the drift b and volatility σ over time observed from historical traces in order to predetermine the threshold values L and U. The approach taken in [2] can also be used to adjust these threshold values in real-time based on any nonnegligible changes in the realized values for b and σ over time. This latter approach can be further used directly for system/network environments whose demand processes do not exhibit consistent seasonal patterns.

Our investigation of a general class of dynamic resource allocation problems strongly suggests that the stochastic optimal control approach taken in this study can provide an effective means to develop easily-implementable online algorithms for solving stochastic optimization problems in practice.

5. REFERENCES

[1] V. Beneš, L. Shepp, H. Witsenhausen. Some solvable stochastic control problems. *Stochastics*, 4(1):39–83, 1980.

[2] D. F. Ciocan, V. Farias. Model predictive control for dynamic resource allocation. *Math of OR*, 37(3):501–525, 2012.

[3] X. Gao, Y. Lu, M. Sharma, M. S. Squillante, J. W. Bosman. Stochastic optimal control for a class of dynamic resource allocation problems. Tech Rep, IBM Research, 2012.

[4] I. Karatzas, S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, Second edition, 1991.

[5] M. Lin, A. Wierman, L. L. H. Andrew, E. Thereska. Dynamic right-sizing for power-proportional data centers. In *Proceedings of IEEE INFOCOM*, 2011.

[6] J. Yong, X. Y. Zhou. *Stochastic Controls: Hamiltonian Systems and HJB Equations*. Springer-Verlag, 1999.