

VU Research Portal

Optimal Quality of Service Control in Communication Systems

Bosman, J.W.

2014

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Bosman, J. W. (2014). *Optimal Quality of Service Control in Communication Systems*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam].

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Stochastic Optimal Control for a General Class of Dynamic Resource Allocation Problems

5

Thus far, the considered models and corresponding analyses were focused on optimal usage of available resources. In this chapter we consider a general class of dynamic resource allocation problems within a stochastic optimal control framework. This class of problems arises in a wide variety of applications, each of which intrinsically involves resources of different types and demand with uncertainty and/or variability. The goal is to dynamically allocate capacity for every resource type in order to serve the uncertain/variable demand and maximize the expected net-benefit over a time horizon of interest based on the rewards and costs associated with the different resources. In [48, 49] Xuefeng Gao, Yingdong Lu, Mayank Sharma, and Mark Squillante derived the optimal control policy within a singular control setting, which includes easily implementable algorithms for governing the dynamic adjustments to resource allocation capacities over time.

Based on the singular control policy, we have performed benchmarks to other methods in literature in a realistic setting. Therefore we developed a simulation environment in which experiments are constructed to demonstrate that this control policy is working extremely well. To make the setting more realistic, we analyze Internet-traffic traces and fit these to a demand model that is used in the simulation experiments. Numerical experiments investigate various issues of both theoretical and practical interest, quantifying the significant benefits of our approach over alternative optimization approaches.

5.1 Background

Various canonical forms of general dynamic 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 canonical forms of general dynamic resource allocation problems in such complex system and network environments. These environments often consist of different types of resources that are allocated in combination to serve demand whose behavior over time includes different types of uncertainty and variability. Each type of resource has a different reward and cost structure that

⁵This chapter is based on [49].

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 degrees of flexibility and different cost structures with respect to making changes to the allocation capacity. The resource allocation optimization problem we consider consists of adaptively determining the primary and secondary resource capacities that serve the uncertain/variable demand and that maximize the expected net-benefit over a time horizon of interest based on the foregoing structural properties of the different types of resources.

The general class of resource allocation problems studied in this chapter arises in a wide variety of application domains such as cloud computing and data center environments, computer and communication networks, and power-aware (energy-aware) and smart power grid environments, among many others. For example, large-scale cloud computing and data center environments often involve resource allocation over different server options (from fastest performance and most expensive to slowest performance and least expensive) and different network bandwidth options (from guaranteed performance at a cost to opportunistic options at no cost, such as the Internet); e.g., refer to [39, 88, 33, 3, 67]. An additional critical issue in large-scale cloud computing and data center environments concerns the effective and efficient management of the consumption of power by resources in the face of time-varying uncertain system demand and energy prices; e.g., see [55, 83]. Related issues arise in smart power grids where resource allocation is required across a diversity of available energy sources, including a long-term market (stable and less expensive, but rather inflexible), local generation (significant operating constraints and limited capacity), a real-time spot market (readily available and responsive, but at a premium price), and renewables such as wind and solar (less expensive and green, but with high volatility); e.g., refer to [50, 104]. 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.

Motivated by this general class of resource allocation problems, we take a financial mathematics approach that hedges against future risks associated with resource allocation decisions and uncertain demand. Specifically, we consider the underlying fundamental stochastic optimal control problem where the dynamic control policy that allocates primary resource capacity to serve uncertain/variable demand is a variational stochastic process [110] with conditions on its rate of change with respect to time, 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 equation, i.e., the so-called 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, in which a single primary allocation capacity is determined to maximize expected net-benefit over the entire time horizon, our theoretical 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 objective functions, control policies, and rewards, costs and flexibility structures. A wide variety of approaches and methods have been developed and applied to address this diversity of resource allocation problems including, for example, online algorithms and dynamic programming. It is therefore important to compare and contrast our problem formulation and solution approach with some prominent and closely related alternatives. One classical instance of a dynamic resource allocation problem is the multi-armed bandit problem [81] where the rewards are associated with tasks and the goal is to determine under uncertainty which tasks the resource should work on, rather than the other way around. Another widely studied problem is the *ski-rental* or *lease-or-buy* problem [44] where there is demand for a resource, but it is initially not known as to how long the resource would be required. In each decision epoch, the choice is between two options: either lease the resource for a fee, or purchase the resource for a price much higher than the leasing fee. Our resource allocation problem differs from this situation in that there are multiple types of resources each with an associated reward and cost per unit time of allocation, since the resources cannot be purchased outright.

From a methodological perspective, the general resource allocation problem we consider in this chapter is closely related to the vast financial mathematics literature on solving stochastic control problems for investment and capacity planning; refer to, e.g., [68, 110, 91]. For example, Beneš et al. [12] 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 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 power. For example, Lin et al. [83] 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 shown to be within a constant factor worse than the corresponding optimal offline policy.

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 experiments quantify the effectiveness of our optimal online dynamic control algorithm over recent work in the area, including comparisons demonstrating how our optimal online algorithm significantly outperforms the type of optimal offline algorithm within a discrete-time framework recently proposed in [83], which turns out to be related to the optimal online algorithm proposed in [35] within a different discrete-time stochastic optimization framework. This includes relative improvements up to 90% and 130% in comparison with the optimal offline algorithm considered in [83], and significantly larger relative improvements in comparison with the optimal online algorithm in [35].

As a specific application example used for illustrative purposes throughout the chapter, which includes our representative numerical experiments, we shall focus on a basic power-aware resource allocation problem that arises in data center environments. In particular, we consider the problem of dynamically adjusting the allocation of high-performance, high-power servers (primary resources) to serve the uncertain/variable demand within a data center, where any remaining demand is served by low-performance, low-power servers (secondary resources), with the objective of maximizing expected profit (expected rewards minus expected costs). Here, the rewards are based on the performance properties of the type of resources allocated to serve demand over time and the costs are based on the power properties of the type of resources allocated to serve demand over time, together with the costs incurred for making adjustments to the allocation of primary resources over time.

The remainder of this chapter is organized as follows. Section 5.2 defines our mathematical model and formulation of the resource allocation optimization problem. The optimal control policy and corresponding results are presented in Section 5.3, with proofs provided in Section 5.A. A representative sample of numerous numerical experiments are discussed in Section 5.4, followed by some concluding remarks.

5.2 Model

5.2.1 System model

We investigate a general class of resource allocation problems in which different types of resources are allocated to satisfy demand whose behavior over time includes uncertainty and/or variability. To simplify the presentation, we focus on two types of resources: a *primary* resource allocation option that has the highest net-benefit and a *secondary* resource allocation option that has the lowest net-benefit. In terms of our representative application example, the primary resource option consists of high-performance/power servers and the secondary resource option consists of low-performance/power servers. Moreover, high-performance/power server capacity is somewhat less flexible in the sense that its rate of change at any instant of time is bounded, whereas low-performance/power server capacity is more flexible in this regard, each of which is made more precise below. Beyond these differences, both types of resources are capable of serving the demand and all of this demand needs to be served (i.e., no loss of demand). A control policy defines at every time $t \in \mathbb{R}$ the level of primary (high-performance/power) resource allocation, denoted by $P(t)$, and the level of secondary (low-performance/power) resource allocation, denoted by $S(t)$, that are used in combination to satisfy the uncertain/variable demand, denoted by $D(t)$.

Our mathematical resource allocation model generalizes to multiple primary resource allocation options with an analogous net-benefit ordering. Namely, the first primary resource option (highest performance/power servers) has the highest net-benefit, followed by the second primary resource option (next highest performance/power servers) having the next highest net-benefit, and so on, with the (single) secondary resource option (lowest performance/power servers) having the lowest net-benefit. In addition, we have extended our mathematical analysis presented herein to address various forms of this general resource allocation model under certain conditions. However, the single primary and secondary instance of our general resource allocation model captures the key aspects of the fundamental trade-offs among the net-benefits of the various resource allocation options together with their associated risks. We have also shown that the optimal dynamic control policy for various instances of the general model under certain conditions has a very similar structure to that of the single primary and secondary resource model instance. In contrast, the general model, the additional notation, and the technical arguments used to establish these more general results all require much more space than is available to us here. Hence, our focus in this chapter shall be on the canonical single primary (high-performance/power) and secondary (low-performance/power) resource allocation model. The interested reader is referred to [48] for these additional technical details.

We consider the singular stochastic optimal control problem underlying our resource allocation model in which uncertain and/or variable demand needs to be served by primary (high-performance/ power) and secondary (low-performance/power) resource allocation capacities. The demand process $D(t)$ is given by the linear diffusion model

$$dD(t) = bdt + \sigma dW(t),$$

where $b \in \mathbb{R}$ is the demand growth/decline rate (which can be extended to a deterministic function of time, but we do not consider this further in the present chapter), $\sigma > 0$ is the demand volatility/variability, and $W(t)$ is a one-dimensional standard Brownian motion, whose sample paths are nondifferentiable [69, 68]. This demand process is served by the combination of primary (high-performance/power) and secondary (low-performance/power) resource allocation capacities $P(t) + S(t)$. Given the higher net-benefit structure of the primary resource option, the optimal dynamic control policy seeks to determine at every time $t \in \mathbb{R}$ the high-performance/power server allocation capacity $P(t)$ to serve the demand $D(t)$ such that any remaining demand is served by the low-performance/power server allocation capacity $S(t)$.

Let $R_p(t)$ and $C_p(t)$ respectively denote the reward and cost associated with the primary (high-performance/power) 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. Therefore, we have

$$R_p(t) = \mathcal{R}_p \times [P(t) \wedge D(t)], \quad (5.1)$$

$$C_p(t) = \mathcal{C}_p \times P(t), \quad (5.2)$$

where $x \wedge y := \min\{x, y\}$, $\mathcal{R}_p \geq 0$ captures all per-unit rewards for serving demand with high-performance/power server capacity, $\mathcal{C}_p \geq 0$ captures all per-unit costs for high-performance/power server capacity, and $\mathcal{R}_p > \mathcal{C}_p$. Observe that the rewards are linear in $P(t)$ as long as $P(t) \leq D(t)$, otherwise any primary resource capacity exceeding demand solely incurs costs without rendering rewards. Hence, from a risk hedging perspective, the risks associated with the primary (high-performance/power) resource allocation position at time t , $P(t)$, 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.

Since the optimal dynamic control policy serves all remaining demand with secondary (low-performance/power) resource allocation capacity, we therefore have

$$S(t) = [D(t) - P(t)]^+.$$

The corresponding reward function $R_s(t)$ and cost function $C_s(t)$ are then given by

$$R_s(t) = \mathcal{R}_s \times [D(t) - P(t)]^+, \quad (5.3)$$

$$C_s(t) = \mathcal{C}_s \times [D(t) - P(t)]^+, \quad (5.4)$$

where $x^+ := \max\{x, 0\}$, $\mathcal{R}_s \geq 0$ captures all per-unit rewards for serving demand with low-performance/power server capacity, $\mathcal{C}_s \geq 0$ captures all per-unit costs for low-performance/power server capacity, and $\mathcal{R}_s > \mathcal{C}_s$. Hence, from a risk hedging perspective, the secondary (low-performance/power) resource allocation position at time t , $S(t)$, is riskless in the sense that rewards and costs are both linear in the resource capacity actually used.

5.2.2 Problem formulation

The singular stochastic optimal control problem of the previous section 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, which we call the risk-hedging position of the dynamic control policy. More formally, the decision process $P(t)$ is adapted to the filtration \mathcal{F}_t generated by $\{D(s) : s \leq t\}$. Furthermore, any adjustments to the primary (high-performance/power) resource allocation 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; namely, \mathcal{I}_p represents the per-unit cost for increasing the allocation of high-performance/power servers while \mathcal{D}_p represents the per-unit cost for decreasing the allocation of high-performance/power servers. 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 (high-performance/power) and secondary (low-performance/power) resource allocation capacities minus the additional costs for adjustments to $P(t)$.

In formulating the corresponding stochastic optimization problem, we impose a couple of additional conditions on the variational decision process $\{P(t) : t \geq 0\}$ based on practical aspects of the diverse application domains motivating our study. The control policy cannot instantaneously change the primary (high-performance/power) resource allocation capacity in an attempt to directly follow the demand $D(t)$; i.e., some time is required (even if only a very small amount of time) to adjust $P(t)$. Moreover, the control policy cannot make unbounded adjustments in the primary (high-performance/power) resource allocation capacity at any instant in time; i.e., the amount of change in $P(t)$ at time t is restricted (even if only to a very small extent) 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. Defining

$$\begin{aligned} N_p(t) &:= R_p(t) - C_p(t), \\ N_s(t) &:= R_s(t) - C_s(t), \end{aligned}$$

we seek to determine the optimal dynamic control policy that solves the problem (SC-OPT)

$$\begin{aligned} \max_{\dot{P}(t)} \quad & \mathbb{E} \int_0^\infty e^{-\alpha t} [N_p(t) + N_s(t)] dt \\ & - \mathbb{E} \int_0^\infty e^{-\alpha t} [\mathcal{I}_p \cdot \mathbb{1}_{\{\dot{P}(t) > 0\}}] dP(t) \\ & - \mathbb{E} \int_0^\infty e^{-\alpha t} [\mathcal{D}_p \cdot \mathbb{1}_{\{\dot{P}(t) < 0\}}] d(-P(t)) \quad (5.5) \\ \text{s.t.} \quad & -\infty < \theta_\ell \leq \dot{P}(t) \leq \theta_u < \infty, \quad (5.6) \\ & dD(t) = bdt + \sigma dW(t), \quad (5.7) \end{aligned}$$

where α is the discount factor and $\mathbb{1}_{\{A\}}$ denotes the indicator function returning 1 if A is true and 0 otherwise. The control variable is the rate of change in the primary (high-performance/power) resource capacity by the control policy at every time t subject to the lower and upper bound constraints on $\dot{P}(t)$ in (5.6). Note that the second (third) expectation in (5.5) causes a decrease with rate \mathcal{I}_p (\mathcal{D}_p) in the value of the objective function whenever the control policy increases (decreases) $P(t)$.

The first expectation in the objective function of the stochastic optimization problem (SC-OPT) can be simplified as follows. Define

$$\begin{aligned} X(t) &:= P(t) - D(t), \\ \mathcal{N}_p &:= \mathcal{R}_p - \mathcal{C}_p, \\ \mathcal{N}_s &:= \mathcal{R}_s - \mathcal{C}_s, \end{aligned}$$

and $x^- := -\min\{x, 0\}$. Upon substituting (5.1), (5.2), (5.3) and (5.4) into the first expectation in (5.5), and making use of the fact that

$$[P(t) \wedge D(t)] = D(t) - [D(t) - P(t)]^+,$$

we obtain

$$\mathbb{E} \left[\int_0^\infty e^{-\alpha t} [-\mathcal{C}_p X(t) + (\mathcal{N}_s - \mathcal{R}_p) X(t)^-] dt \right] + \mathcal{N}_p \mathbb{E} \left[\int_0^\infty e^{-\alpha t} D(t) dt \right]. \quad (5.8)$$

Since the second expectation in (5.8) does not depend on the control variable $\dot{P}(t)$, this term plays no role in determining the optimal dynamic control policy. Together with the above results, we derive the following stochastic optimization problem which is equivalent to the original optimization problem formulation (SC-OPT):

$$\min_{\dot{P}(t)} \mathbb{E}_x \left[\int_0^\infty e^{-\alpha t} \left\{ (\mathcal{C}_+ X(t)^+ + \mathcal{C}_- X(t)^-) dt + \left(\mathcal{I}_p \mathbb{1}_{\{\dot{P}(t) > 0\}} - \mathcal{D}_p \mathbb{1}_{\{\dot{P}(t) < 0\}} \right) dP(t) \right\} \right] \quad (5.9)$$

$$\text{s.t.} \quad -\infty < \theta_\ell \leq \dot{P}(t) \leq \theta_u < \infty, \quad (5.10)$$

$$dX(t) = dP(t) - bdt - \sigma dW(t), \quad (5.11)$$

$$X(0) = x, \quad (5.12)$$

$$\mathcal{C}_+ = \mathcal{C}_p, \quad (5.13)$$

$$\mathcal{C}_- = \mathcal{N}_p - \mathcal{N}_s, \quad (5.14)$$

where $\mathbb{E}_x[\cdot]$ denotes expectation with respect to the initial state distribution (i.e., state at time $t = 0$) being x with probability one.

We use $V(x)$ to represent the optimal value of the objective function (5.9); namely, $V(x)$ is the value function of the corresponding stochastic dynamic program. Given its equivalence with the original optimization problem (SC-OPT), the remainder of this chapter will focus on the stochastic dynamic program formulation in (5.9) - (5.14).

5.3 Optimal control policy

In this section we consider our main results on the optimal dynamic control policy for the stochastic optimization problem (5.9) - (5.14). After some technical preliminaries, we present our main results under the conditions $\mathcal{I}_p \geq 0$ and $\mathcal{D}_p \geq 0$, which are likely to be the most interesting case in practice. All other cases of our main results are covered in [48]. Consideration of the proofs of our main results is postponed until the next section.

5.3.1 Preliminaries

To elucidate the exposition, we henceforth assume $b \geq 0$ without loss of generality as one can readily verify that our main results hold when $b < 0$. For notational convenience, we next define the constants

$$r_1 := \frac{b + \sqrt{b^2 + 2\alpha\sigma^2}}{\sigma^2} > 0, \quad (5.15)$$

$$r_2 := \frac{b - \sqrt{b^2 + 2\alpha\sigma^2}}{\sigma^2} < 0, \quad (5.16)$$

$$s_1 := \frac{b - \theta_u + \sqrt{(b - \theta_u)^2 + 2\alpha\sigma^2}}{\sigma^2} > 0, \quad (5.17)$$

$$s_2 := \frac{b - \theta_u - \sqrt{(b - \theta_u)^2 + 2\alpha\sigma^2}}{\sigma^2} < 0, \quad (5.18)$$

$$t_1 := \frac{b - \theta_\ell + \sqrt{(b - \theta_\ell)^2 + 2\alpha\sigma^2}}{\sigma^2} > 0, \quad (5.19)$$

$$t_2 := \frac{b - \theta_\ell - \sqrt{(b - \theta_\ell)^2 + 2\alpha\sigma^2}}{\sigma^2} < 0. \quad (5.20)$$

These quantities are the roots of the quadratic equation

$$\frac{\sigma^2}{2}y^2 + (\theta - b)y - \alpha = 0,$$

when θ takes on the values of θ_ℓ , 0 or θ_u .

Finally, for additional convenience in stating our main results, we further define the following constants

$$\begin{aligned} B_1 &= (C_+ - \alpha\mathcal{D}_\rho)(t_2 - r_2), \\ B_2 &= (C_- - \alpha\mathcal{I}_\rho)(s_1 - r_2), \\ B_3 &= (C_+ + C_-)(-r_2), \\ A &= (C_+ + \alpha\mathcal{I}_\rho)(r_2 - r_1), \\ J_1 &= (C_+ - \alpha\mathcal{D}_\rho)(r_1 - t_2), \\ J_2 &= (C_- - \alpha\mathcal{I}_\rho)(r_1 - s_1), \\ J_3 &= (C_+ + C_-)r_1, \\ K &= (C_- + \alpha\mathcal{D}_\rho)(r_2 - r_1), \end{aligned}$$

in terms of $r_1, r_2, s_1, s_2, t_1, t_2$ given in (5.15) - (5.20). Since $\theta_\ell < 0$ and $\theta_u > 0$, we conclude that B_i and J_i are all positive for $i = 1, 2, 3$, and that A and K are both negative.

5.3.2 Case 1: $\mathcal{D}_p < \mathcal{C}_+/\alpha$ and $\mathcal{I}_p < \mathcal{C}_-/\alpha$

Let us first briefly interpret the conditions of this section. Observe from the objective function (5.9) that \mathcal{C}_+/α reflects the discounted overage cost associated with the primary resource capacity and \mathcal{C}_-/α 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.

5.3.1. Theorem. *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*

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

Moreover, the values of L and U can be characterized by the following three cases.

1. If

$$0 < \frac{B_3 - B_2}{B_1} < 1$$

and

$$\left(\frac{B_3 - B_2}{B_1} \right)^{\frac{r_2}{r_1}} \geq \frac{J_3 - J_2}{J_1},$$

or

$$B_3 \leq B_2,$$

then we have

$$U > L \geq 0,$$

where L and U are uniquely determined by the two equations:

$$B_1 e^{r_1(L-U)} + J_1 e^{r_2(L-U)} + A = 0, \quad (5.21)$$

$$\begin{aligned} \frac{B_1 r_2}{r_1 - r_2} e^{r_1(L-U)} + \frac{J_1 r_1}{r_1 - r_2} e^{r_2(L-U)} = \\ (r_1 + r_2 - s_1)(\alpha \mathcal{I}_p + \mathcal{C}_+) + (\mathcal{C}_+ + \mathcal{C}_-) s_1 \cdot e^{s_2 L}. \end{aligned} \quad (5.22)$$

II. If

$$\frac{B_3 - B_1}{B_2} > 1$$

and

$$\left(\frac{B_3 - B_1}{B_2} \right)^{\frac{r_2}{r_1}} \geq \frac{J_3 - J_1}{J_2},$$

then we have

$$L < U \leq 0,$$

where L and U are uniquely solved by the two equations:

$$B_2 e^{r_1(U-L)} + J_2 e^{r_2(U-L)} + K = 0, \quad (5.23)$$

$$\begin{aligned} B_2 \frac{r_2}{r_1 - r_2} e^{r_1(U-L)} + J_2 \frac{r_1}{r_1 - r_2} e^{r_2(U-L)} = \\ (r_1 + r_2 - t_2)(\alpha \mathcal{D}_p + \mathcal{C}_-) + (\mathcal{C}_+ + \mathcal{C}_-) t_2 \cdot e^{t_1 U}. \end{aligned} \quad (5.24)$$

III. If none of the above conditions hold, we then have

$$U \geq 0 \geq L,$$

where L and U are uniquely determined by

$$B_1 e^{-r_1 U} + B_2 e^{-r_1 L} = B_3, \quad (5.25)$$

$$J_1 e^{-r_2 U} + J_2 e^{-r_2 L} = J_3. \quad (5.26)$$

Theorem 5.3.1 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., making no change to $P(t)$) 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 , thus increasing the primary (high-performance/power) resource capacity allocation. Similarly, 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 , thus decreasing the primary (high-performance/power) resource capacity allocation. In each of the cases I, II and III, the optimal threshold values L and U are uniquely determined by two non-linear equations.

5.3.3 Remaining cases

We further establish our main results for all remaining possible conditions on the adjustment costs \mathcal{D}_p and \mathcal{I}_p . However, this is not in the scope of this dissertation. Therefore we refer the interested reader to [48] for these additional technical details.

5.4 Numerical experiments

The foregoing sections establish the explicit optimal dynamic control policy among all admissible nonanticipatory control processes $dP(t)$ within a singular stochastic optimal control setting that maximizes the original stochastic dynamic program (SC-OPT) given in (5.5) – (5.7). This optimal dynamic control policy renders a new class of practical online algorithms for general dynamic resource allocation problems that arise in a wide variety of application domains. The resulting online algorithm is easily implementable in computer systems and communication networks (among others) at runtime 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 system/network parameters. Extensive numerical experiments have been conducted across a broad spectrum of system/network environments to investigate various issues of both theoretical and practical interest by comparing our online optimal dynamic control algorithm against alternative optimization approaches from recent work in the research literature. In this section, we present a representative sample of these numerical experiments.

The characteristics of the demand process can differ significantly from one system/network environment to the next. However, within a particular environment as well as across a class of similar environments, one often finds consistent seasonal patterns in the average demand process over time, including consistent seasonal effects at daily, weekly, monthly and yearly time scales. We confirm this to be the case for the environments motivating our study through detailed analyses of real-world trace data from a wide variety of proprietary and publicly-available system/network environments. This includes our detailed analysis of request and packet traces from commercial web server environments, university web site proxy servers, e-commerce Internet server environments, and information technology service delivery centers. Based on such detailed analyses of real-world traces, we accurately fitted the average demand process for each environment of interest by a smooth function $f(t)$. Figure 5.1 depicts representative examples of two of these average daily demand patterns $f^1(t)$ and $f^2(t)$, where time t reflects the time zone of

the system/network environment which may be different from that of the demand source. In addition to the average daily demand process, our detailed analyses of real-world traces reveal common seasonal patterns in the volatility of the demand process over time. These demand process volatility patterns tend to be fairly consistent within each system/network environment, whereas they tend to vary much more significantly over an expansive range of values from one environment to the next. Although a great diversity of daily average and volatility demand patterns were discovered throughout our detailed analyses of trace data, the results of numerous numerical experiments comparing our optimal dynamic control policy with alternative optimization approaches under these diverse demand patterns exhibit very similar performance trends, both quantitatively and qualitatively, for a given level of volatility σ . We therefore focus in the remainder of this section on the average daily demand patterns $f^1(t)$ and $f^2(t)$ in Figure 5.1 while varying the volatility parameter σ , noting that the corresponding numerical results are representative of a broad spectrum of system/network environments.

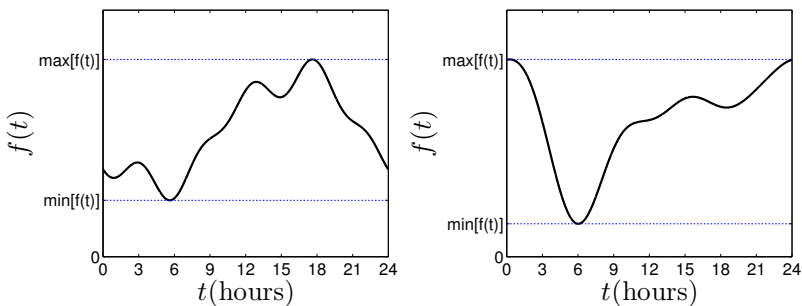


Figure 5.1: Representative average daily demand patterns $f^1(t)$ (left) and $f^2(t)$ (right).

To evaluate the benefits of our optimal online dynamic control algorithm in realistic system/network environments, we consider an alternative optimization approach that has recently appeared in the research literature. As previously noted, Lin et al. [83] study a particular optimal offline algorithm for dynamically adjusting the number of active, power-aware servers in a data center to minimize operating costs, and then develop an optimal online algorithm which is proven to be 3-competitive. Our interest for comparison here is in the offline algorithm of [83], which consists of making optimal provisioning decisions in a clairvoyant anticipatory manner based on the known demand within each slot of a discrete-time model where the slot length is chosen to match the timescale at which the data center can adjust its capacity and so that demand activity within a slot is sufficiently nonnegligible in a statistical

sense. Applying this particular optimal offline algorithm within our mathematical framework, we partition the daily time horizon into T slots of length γ such that

$$\begin{aligned} h_i &= (t_{i-1}, t_i], \\ \gamma &= t_i - t_{i-1}, \end{aligned}$$

$i = 1, \dots, T$, $t_0 := 0$, and we compute the average demand

$$g_i := \gamma^{-1} \int_{h_i} f(t) dt$$

within each slot i yielding the average demand vector

$$(g_1, g_2, \dots, g_T).$$

Define

$$\Delta(P_i) := P_i - P_{i-1},$$

where P_i denotes the primary (high-performance/power) resource allocation capacity for slot i . The optimal solution under this offline algorithm is then obtained by solving the following linear program (LP):

$$\begin{aligned} \min_{\Delta(P_1), \dots, \Delta(P_T)} \quad & \sum_{i=1}^T \mathcal{C}_+(P_i - g_i)^+ + \mathcal{C}_-(P_i - g_i)^- + \\ & \mathcal{I}_p(P_i - P_{i-1})^+ + \mathcal{D}_p(P_i - P_{i-1})^- \end{aligned} \quad (5.27)$$

$$\begin{aligned} \text{s.t.} \quad & -\infty < \theta_\ell \leq \Delta(P_i)/\gamma \leq \theta_u < \infty, \\ & \forall i = 1, \dots, T, \end{aligned} \quad (5.28)$$

where the constraints on $\Delta(P_i)$ in (5.28) correspond to (5.10). In this deterministic optimization problem, the control variable is the rate of change in the primary (high-performance/power) resource allocation for each slot i over the daily horizon. We refer to this solution as the offline LP algorithm. This alternative optimization approach is used for comparison in evaluating the benefits of our optimal dynamic control policy.

We also consider a related optimal online algorithm proposed by Ciocan and Farias [35], which is based on a discrete-time model framework similar to that above though within a distinct stochastic optimization framework. We refer to this as the online CF algorithm. Our numerical experiments indicate that the primary (high-performance/power) resource allocation capacity decisions under the online CF algorithm are identical to those under the offline LP algorithm shifted by one time slot.

Hence, the benefits of our optimal online dynamic control algorithm over the online CF algorithm are considerably more significant than those presented here for the offline LP algorithm. We refer the interested reader to [48] for these results and other technical details.

The workloads used for our numerical experiments are generated from real-world trace data taken from various system/network environments. Specifically, once the average daily demand pattern $f(t)$ and the volatility pattern $\sigma(t)$ are extracted from the traces through our detailed analyses, as described above, we construct a linear diffusion process for the entire time horizon such that the drift of the demand process is obtained as the derivative of $f(t)$ (i.e., $b(t) = df(t)$) and the corresponding volatility term is set to match $\sigma(t)$. Since the volatility pattern $\sigma(t)$ tended to be fairly consistent with respect to time within each daily real-world trace for a specific environment and since the volatility pattern tended to vary considerably from one daily real-world trace to another, our linear diffusion demand process is assumed to be governed by the following dynamic model

$$dD(t) = b(t)dt + \sigma dW(t),$$

where we vary the volatility term σ to investigate different system/network environments. The workload for each system/network environment then consists of a set of sample paths generated from the Brownian demand process $D(t)$ defined in this manner. Given such a workload demand process for a specific system/network environment of interest, we calibrate our optimal online dynamic control algorithm by first partitioning the drift function $b(t)$ of the demand process $D(t)$ into piecewise linear segments and then computing the threshold values L and U for each per-segment drift and σ according to Theorem 5.3.1. This (fixed) version of our optimal online dynamic control algorithm is applied to every daily sample path of the Brownian demand process $D(t)$ and the time-average value of net-benefit is computed over this set of daily sample paths. Based on detailed analyses of real-world traces, such applications of our optimal dynamic control policy are easily realized in practice. For comparison under the same set of Brownian demand process sample paths, we compute the average demand vector (g_1, \dots, g_T) and the corresponding solution under the offline LP algorithm for each daily sample path by solving the linear program (5.27),(5.28) with respect to (g_1, \dots, g_T) , and then we calculate the time-average value of net-benefit over the set of daily sample paths. All of our numerical experiments were implemented in Matlab using, among other functionality, the econometrics toolbox.

We now present a representative sample of our extensive numerical experiments, starting with a first collection of workloads based on the average daily demand pattern $f^1(t)$ illustrated in the top plot of Figure 5.1. Define $f_{\min} := \min_t \{f(t)\}$, $f_{\max} := \max_t \{f(t)\}$, and $f_{\text{avg}} := T^{-1} \int_0^T f(t)dt$. The base parameter settings for this

first set of workloads are: $\alpha = 0.02$, $\sigma = 0.4$, $\theta_l = -10$, $\theta_u = 10$, $\mathcal{C}_+ = 20$, $\mathcal{C}_- = 2$, $\mathcal{D}_p = 0.5$, $\mathcal{I}_p = 0.5$, $f_{\min}^1 = 2$, $f_{\max}^1 = 7$, $f_{\text{avg}}^1 = 4.5$, $x = X(0) = P(0) - D(0) = 0$, and $P_0 = D(0)$. In addition to these base settings, we vary certain parameter values to investigate the impact and sensitivity of these parameters on the performance of both optimization algorithms. This includes conducting numerical experiments under the base parameter settings while varying one of $\sigma \in [0.01, 1.0]$, $\mathcal{C}_+ \in [10, 40]$, $\mathcal{C}_- \in [1, 10]$, $f_{\min}^1 \in [1, 5]$, and $f_{\max}^1 \in [4, 25]$, all representing parameter values exhibited in real-world system/network environments. For each numerical experiment comprised of a specific workload, we generate $N = 10,000$ daily sample paths using a timescale of a couple of seconds and a γ setting of five minutes, noting that a wide variety of experiments with different timescale and γ settings provided the same performance trends as those presented herein. We then apply our optimal dynamic control policy and the alternative optimization approach to this set of N daily sample paths as described above, where our performance evaluation comparison is based on the expectation of net-benefit realized under each of the two algorithms, also as described above. In particular, the expected net-benefit is computed as the time-average value of the rewards minus the costs from the primary (high-performance/power) and secondary (low-performance/power) resource allocation capacities and minus the costs for adjustments to the primary resource allocation capacity, taken over all N daily sample paths under each of our optimal online dynamic control algorithm and the offline LP algorithm.

Figure 5.2 presents a representative sample of some of our numerical results for the first set of workloads based on $f^1(t)$. The top graph provides performance comparisons of our optimal online dynamic control algorithm against the alternative offline LP algorithm, where the comparisons are based on the relative improvements in expected net-benefit under our optimal control policy as a function of σ ; the relative improvement is defined as the difference in expected net-benefit under our optimal dynamic control policy and under the alternative offline LP approach, divided by the expected net-benefit of the latter. For the purpose of comparison across sets of workloads with very different f_{avg} values, we plot this graph as a function of the coefficient of variation $\text{CoV} = \sigma/f_{\text{avg}}$. The bottom graph provides similar comparisons of relative improvement in expected net-benefit between our optimal dynamic control policy and the alternative offline LP approach as a function of \mathcal{C}_+ , both with σ fixed to be 0.4.

We first observe from the top graph in Figure 5.2 that our optimal online dynamic control algorithm outperforms the alternative optimization approach for all $\sigma > 0$. The relative improvements in expected net-benefit under our optimal dynamic control policy grow in an exponential manner with respect to increasing values of σ over the range of CoV values considered, with relative improvements up to 90% in comparison with the offline LP algorithm. The rewards and costs associated with the primary (high-performance/power) and secondary (low-performance/power) resource capacities can be based on either performance or financial metrics, or a

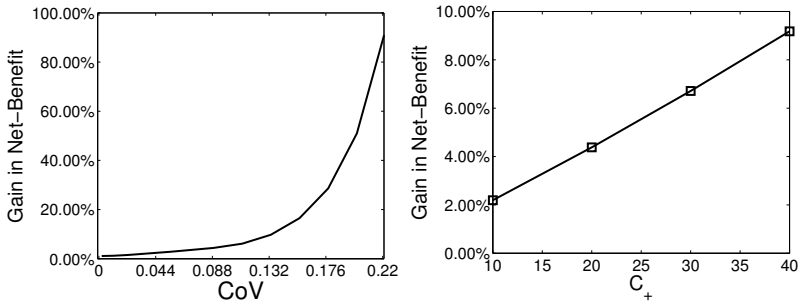


Figure 5.2: Improvement in expected net-benefit under our optimal dynamic control policy relative to the alternative offline LP algorithm for the first set of workloads based on $f^1(t)$ and for varying values of σ and C_+ .

combination of both. Our results illustrate and quantify the fact that, even in discrete-time models with small time slot lengths γ , nonnegligible volatility plays a critical role in the expected net-benefit of any given resource allocation policy. The significant relative improvements under the optimal online dynamic control algorithm then follow from our stochastic optimal control approach that directly addresses the volatility of the demand process in all primary (high-performance/power) and secondary (low-performance/power) resource allocation decisions. Since our detailed analysis of real-world traces exhibited relatively large CoV values, the results in Figure 5.2 suggest that these net-benefit improvements under our optimal dynamic control policy can be very significant in practice. While the offline LP algorithm based on (5.27),(5.28) would eventually outperform our optimal online dynamic control algorithm as the time slot length γ decreases, we note that the choice for γ in our numerical experiments is considerably smaller than the 10-minute intervals suggested as examples in the literature [83]. Moreover, as discussed in [35], the optimal choice of γ is a complex issue in and of itself and it may need to vary over time depending upon the statistical properties of the demand process $D(t)$. A key advantage of our optimal online dynamic control algorithm is that such parameters are not needed. As explained above, our algorithm can exploit any consistent seasonal patterns for $b(t)$ and $\sigma(t)$ observed from historical traces in order to predetermine the threshold values L and U . In addition, approaches similar to those taken in [35] can be used to adjust these threshold values in real-time based on any nonnegligible changes in the realized values for $b(t)$ and $\sigma(t)$. Furthermore, this latter approach can be used directly for system/network environments whose demand processes do not exhibit consistent seasonal patterns.

We next observe from the bottom graph in Figure 5.2 that the relative improvements in expected net-benefit under our optimal online dynamic control algorithm similarly increases with respect to increasing values of C_+ , though in a more lin-

ear fashion. We also note that very similar trends were observed with respect to varying the value of \mathcal{C}_- , though the magnitude of the relative improvement in expected net-benefit is smaller. Our numerical experiments suggest that the relative improvements in net-benefit under our optimal dynamic control policy can be more sensitive to \mathcal{C}_+ than to \mathcal{C}_- . Recall that $\mathcal{C}_+ = \mathcal{C}_p$ is the cost for the primary resource allocation capacity, whereas $\mathcal{C}_- = \mathcal{N}_p - \mathcal{N}_s$ is the difference in net-benefit between the primary (high-performance/power) and secondary (low-performance/power) resource allocation capacities.

We next present a representative sample of our numerical experiments for a second collection of workloads based on the average daily demand pattern $f^2(t)$ illustrated in the bottom plot of Figure 5.1. The base parameter settings for this second set of workloads are: $\alpha = 0.02$, $\sigma = 7.0$, $\theta_l = -100$, $\theta_u = 100$, $\mathcal{C}_+ = 20$, $\mathcal{C}_- = 2$, $\mathcal{D}_p = 0.5$, $\mathcal{I}_p = 0.5$, $f_{\min}^2 = 15$, $f_{\max}^2 = 90$, $f_{\text{avg}}^2 = 61$, $x = X(0) = P(0) - D(0) = 0$ and $P_0 = D(0)$. In addition, analogous to the first set of workloads, we conducted numerical experiments under the base parameter settings while varying one of $\sigma \in [0.01, 15]$, $\mathcal{C}_+ \in [10, 40]$, $\mathcal{C}_- \in [1, 10]$, $f_{\min}^2 \in [1, 20]$ and $f_{\max}^2 \in [9, 120]$, all representing parameter values exhibited in real-world system/network environments. Once again, for each experiment comprised of a specific workload, we generate $N = 10,000$ sample paths using a timescale of a couple of seconds and a γ setting of five minutes, noting that a wide variety of experiments with different timescale and γ settings provided performance trends that are identical to those presented herein. We then apply our optimal dynamic control policy and the alternative optimization approach to this set of N sample paths as described above. Our performance evaluation comparisons are based on the expectation of net-benefit realized under each of the two algorithms, also as described above.

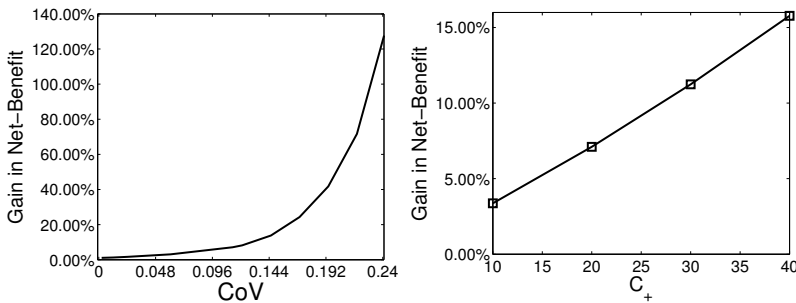


Figure 5.3: Improvement in expected net-benefit under our optimal dynamic control policy relative to the alternative offline LP algorithm for the second set of workloads based on $f^2(t)$ and for varying values of σ and \mathcal{C}_+ .

Figure 5.3 presents a representative sample of some of our numerical results for the second set of workloads based on $f^2(t)$, providing the analogous results that cor-

respond to those in Figure 5.2. We note that the larger range $[f_{\min}^2, f_{\max}^2]$ exhibited in the second average daily demand pattern as well as a higher value of f_{avg}^2 lead to both a higher relative net-benefit for fixed σ and a higher sensitivity to changes in σ . Hence, our optimal online dynamic control algorithm has enhanced gains over the alternative offline LP algorithm in terms of expected net-benefit. This relative improvement in expected net-benefit as compared to the set of experiments for the average daily demand pattern $f^1(t)$ can be understood to be caused by the sharp drop in average demand from the maximum value of 90 to a minimum of 15 within a fairly short time span, thus contributing to an increased effective volatility over and above that represented by σ . Therefore, the fact that the relative improvement exhibited by our optimal online dynamic control algorithm is larger under the average daily demand pattern $f^2(t)$, up to 130% in comparison with the offline LP algorithm, can be viewed in a sense to be very much an extension of the finding that the relative improvement provided by our optimal online algorithm increases with an increase in CoV. A similar gain in performance improvement can be seen in the bottom plot when we vary \mathcal{C}_+ with a fixed value of σ .

In comparing the results in Figures 5.2 and 5.3 as well as those from our many other numerical experiments, we observe that all of these net-benefit results are quite consistent in terms of qualitative and quantitative trends, where our optimal online dynamic control algorithm consistently provides superior results even when the demand uncertainty/volatility is very small. As an additional observation from our numerical experiments, we find that, upon applying our optimal online dynamic control algorithm and the offline LP algorithm directly to the actual trace data, consistent performance trends among the two optimization approaches were observed in comparison with those shown above for the time-average over many sample paths from the Brownian motion demand process fitted to the same trace data. Finally, though not presented herein, the magnitude of the relative improvements in expected net-benefits exhibited in Figures 5.2 and 5.3 are significantly larger when comparing our optimal online dynamic control algorithm and the online CF algorithm.

5.5 Discussion

In this chapter we investigated a general class of dynamic resource allocation problems arising across a broad spectrum of applications that involve different types of resources and uncertain/variable demand. With a goal of maximizing expected net-benefit based on rewards and costs from the different resources, Xuefeng Gao, Yingdong Lu, Mayank Sharma, and Mark Squillante derived the provably optimal dynamic control policy within a singular stochastic optimal control setting. Their mathematical analysis includes obtaining simple expressions that govern the dy-

namic adjustments to resource allocation capacities over time under the optimal control policy. Based on this analysis, we constructed a wide variety of extensive numerical experiments. The results demonstrate and quantify significant benefits of the optimal dynamic control policy over recently proposed alternative optimization approaches in addressing a general class of resource allocation problems across a diverse range of application domains, including cloud computing and data center environments, computer and communication networks, and power-aware (energy-aware) and smart power grid environments. Moreover, our results strongly suggest that the stochastic optimal control approach taken in this chapter can provide an effective means to develop easily-implementable online algorithms for solving stochastic optimization problems.

Appendix 5.A Proofs

In this section we consider the proofs of the main results. We focus on some aspects of our rigorous proof of Theorem 5.3.1, with all remaining technical details as well as the proofs of the other main results provided in [48]. Our proof proceeds in three main steps. First, we express the optimality conditions for the stochastic dynamic program, i.e., the Bellman equation corresponding to (5.9) – (5.14). We then derive a solution of the Bellman equation and determine the corresponding candidate value function and dynamic control policy. Finally, we verify that this dynamic control policy is indeed optimal through a martingale argument. Each of these main steps is presented in turn.

5.A.1 Proof of Theorem 5.3.1: Step 1

From the Bellman principle of optimality, we deduce that the value function V satisfies for each $t \geq 0$

$$V(x) = \min_{\theta_t \leq P(t) \leq \theta_u} \mathbb{E}_x \left[\int_0^t e^{-\alpha s} \left[(C_+ X(s)^+ + C_- X(s)^-) ds + (\mathcal{I}_p \mathbb{1}\{\dot{P}(s) > 0\} - \mathcal{D}_p \mathbb{1}\{\dot{P}(s) < 0\}) dP(s) \right] + e^{-\alpha t} V(X(t)) \right], \quad (5.29)$$

refer to [110, Chapter 4]. Suppose the value function V is smooth, belonging to the set C^2 (i.e., the set of twice continuously differentiable functions) except for a finite number of points, which will be established in step 2. Then, based on a standard

application of Ito's formula as in [73], we derive that the desired Bellman equation for the value function V has the form

$$-\alpha V(x) + \frac{1}{2}\sigma^2 V''(x) - bV'(x) + C_+x^+ + C_-x^- + \inf_{\theta_l \leq \theta \leq \theta_u} \mathcal{L}(\theta, x) = 0, \quad (5.30)$$

where

$$\mathcal{L}(\theta, x) = \begin{cases} (V'(x) + \mathcal{I}_p)\theta & \text{if } \theta \geq 0, \\ (V'(x) - \mathcal{D}_p)\theta & \text{if } \theta < 0. \end{cases} \quad (5.31)$$

5.A.2 Proof of Theorem 5.3.1: Step 2

Our next goal is to construct a convex function Y that satisfies the Bellman equation (5.30) and show that the threshold values L and U are uniquely determined by the corresponding pair of nonlinear equations in Theorem 5.3.1. Suppose a candidate value function $Y(x)$ satisfies (5.30). We then seek to find L and U such that

$$Y'(x) = \begin{cases} \geq \mathcal{D}_p, & \text{if } x \geq U, \\ \in (-\mathcal{I}_p, \mathcal{D}_p) & \text{if } L < x < U, \\ \leq -\mathcal{I}_p & \text{if } x \leq L. \end{cases} \quad (5.32)$$

Moreover, $Y(x) = O(|x|)$ when $|x|$ goes to ∞ , and Y meets smoothly at the points L , 0 and U to order one.

For each of the three cases in Theorem 5.3.1, reflecting different relationships among L , U and 0 based on model parameters, we first establish the desired convexity properties together with the corresponding pair of threshold equations. Case II of Theorem 5.3.1 is considered in the next subsection, noting that Cases I and III and other technical details can be found in [48]. Then, in Section 5.A.2.2, we show that the thresholds L and U can be uniquely determined through these equations.

5.A.2.1 Case II: $0 \geq U > L$

Focusing on the case $0 \geq U > L$ where L and U satisfy (5.23) and (5.24) subject to (5.32), we proceed to solve the Bellman equation (5.30) depending on the value of x in relation to U , 0 and L . There are four subcases to consider as follows.

(i). If $x \geq 0$, we have

$$Y'(x) \geq \mathcal{D}_p$$

and

$$\inf_{\theta_\ell \leq \theta \leq \theta_u} \mathcal{L}(\theta, x) = \mathcal{L}(\theta_\ell, x).$$

Then the Bellman equation yields

$$-\alpha Y(x) + \frac{1}{2} \sigma^2 Y''(x) - bY'(x) + \mathcal{C}_+ x^+ + \mathcal{C}_- x^- + \mathcal{L}(\theta_\ell, x) = 0,$$

or equivalently

$$-\alpha Y(x) + \frac{1}{2} \sigma^2 Y''(x) - bY'(x) + \mathcal{C}_+ x + (Y'(x) - \mathcal{D}_p) \theta_\ell = 0.$$

Solving this equation, we derive for $x \geq 0$

$$Y(x) = \frac{\mathcal{C}_+}{\alpha} x + \frac{1}{\alpha} \left(\frac{\mathcal{C}_+}{\alpha} (\theta_\ell - b) - \mathcal{D}_p \theta_\ell \right) + l_3 e^{t_2 x}, \quad (5.33)$$

where l_3 is a generic constant to be determined.

(ii). If

$$U < x < 0,$$

we have

$$Y'(x) \geq \mathcal{D}_p,$$

which yields

$$\inf_{\theta_\ell \leq \theta \leq \theta_u} \mathcal{L}(\theta, x) = \mathcal{L}(\theta_\ell, x),$$

and thus we obtain

$$-\alpha Y(x) + \frac{1}{2} \sigma^2 Y''(x) - bY'(x) - \mathcal{C}_- x + (Y'(x) - \mathcal{D}_p) \theta_\ell = 0.$$

This implies for $U < x < 0$

$$Y(x) = -\frac{\mathcal{C}_-}{\alpha} x + \frac{1}{\alpha} \left(-\frac{\mathcal{C}_-}{\alpha} (\theta_\ell - b) - \mathcal{D}_p \theta_\ell \right) + \lambda_1 e^{t_1 x} + \lambda_2 e^{t_2 x}, \quad (5.34)$$

where λ_1 and λ_2 are generic constants to be determined.

(iii). If

$$L < x < U,$$

we have

$$-\mathcal{I}_p \leq Y'(x) \leq \mathcal{D}_p$$

and the Bellman equation (5.30) renders

$$-\alpha Y(x) + \frac{1}{2}\sigma^2 Y''(x) - bY'(x) - \mathcal{C}_-x = 0.$$

Solving this equation, we deduce for $U > x > L$

$$Y(x) = -\frac{\mathcal{C}_-}{\alpha}x + \frac{b\mathcal{C}_-}{\alpha^2} + \tilde{\lambda}_1 e^{r_1 x} + \tilde{\lambda}_2 e^{r_2 x}, \quad (5.35)$$

where $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ are generic constants to be determined.

(iv). If $x \leq L$, we have

$$Y'(x) \leq -\mathcal{I}_p$$

and the Bellman equation (5.30) becomes

$$-\alpha Y(x) + \frac{1}{2}\sigma^2 Y''(x) - bY'(x) - \mathcal{C}_-x + (Y'(x) + \mathcal{I}_p)\theta_u = 0.$$

The solution is then given by

$$Y(x) = -\frac{\mathcal{C}_-}{\alpha}x + \frac{1}{\alpha}\left(-\frac{\mathcal{C}_-}{\alpha}(\theta_u - b) + \mathcal{I}_p\theta_u\right) + l_1 e^{s_1 x}, \quad (5.36)$$

where l_1 is a generic constant to be determined.

Now we match the value and the first-order derivative of Y at the points U , 0 and L . Hence, the function Y that we construct will be twice continuously differentiable

with the exception of at most three points. Let us first consider such matchings at the point U . From (5.34) and (5.35), we derive

$$Y'(U+) = -\frac{C_-}{\alpha} + \lambda_1 t_1 e^{t_1 U} + \lambda_2 t_2 e^{t_2 U} = \mathcal{D}_p, \quad (5.37)$$

$$Y'(U-) = -\frac{C_-}{\alpha} + \tilde{\lambda}_1 r_1 e^{r_1 U} + \tilde{\lambda}_2 r_2 e^{r_2 U} = \mathcal{D}_p, \quad (5.38)$$

$$Y(U+) = -\frac{C_-}{\alpha} U + \frac{1}{\alpha} \left(-\frac{C_-}{\alpha} (\theta_\ell - b) - \mathcal{D}_p \theta_\ell \right) + \lambda_1 e^{t_1 U} + \lambda_2 e^{t_2 U},$$

$$Y(U-) = -\frac{C_-}{\alpha} U + \frac{bC_-}{\alpha^2} + \tilde{\lambda}_1 e^{r_1 U} + \tilde{\lambda}_2 e^{r_2 U}.$$

Since

$$Y(U+) = Y(U-),$$

we immediately obtain

$$\tilde{\lambda}_1 e^{r_1 U} + \tilde{\lambda}_2 e^{r_2 U} = \lambda_1 e^{t_1 U} + \lambda_2 e^{t_2 U} - \left(\mathcal{D}_p + \frac{C_-}{\alpha} \right) \frac{\theta_\ell}{\alpha}. \quad (5.39)$$

Matching at the point L , we deduce [48] from (5.35) and (5.36) expressions for $Y'(L+)$, $Y'(L-)$, $Y(L+)$, $Y(L-)$, and then solve for $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ in terms of L and the model parameters as follows:

$$\tilde{\lambda}_1 = \frac{\sigma^2}{2\alpha^2} e^{-r_1 L} \left(1 - \frac{r_1}{r_2} \right)^{-1} B_2, \quad (5.40)$$

$$\tilde{\lambda}_2 = \frac{\sigma^2}{2\alpha^2} e^{-r_2 L} \left(1 - \frac{r_2}{r_1} \right)^{-1} (-J_2). \quad (5.41)$$

Upon substituting these expressions into (5.38), we obtain (5.23). To establish (5.24), matching Y at the point 0 to order one renders [48] expressions for $Y'(0+)$, $Y'(0-)$, $Y(0+)$, $Y(0-)$, and thus we can derive

$$\lambda_1 = \frac{\sigma^2}{2\alpha^2} (C_+ + C_-) \frac{t_2^2}{t_1 - t_2}, \quad (5.42)$$

$$\lambda_2 - l_3 = \frac{C_+ + C_-}{\alpha} \left(\frac{1}{t_2} + \frac{1}{t_1 - t_2} \right). \quad (5.43)$$

Upon substituting the expressions for λ_1 , $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ into (5.37) and (5.39), cancelling λ_2 , and simplifying the resulting expressions, we can conclude that L and U satisfy (5.24).

Next, we verify that the candidate value function Y satisfies the required first-order properties in (5.32). Since we have constructed the function Y with

$$\begin{aligned} Y'(U) &= \mathcal{D}_p, \\ Y'(L) &= -\mathcal{I}_p, \end{aligned}$$

then to establish (5.32) it suffices to verify the convexity of the function Y . To this end, one first readily confirms that for $x < L$

$$Y''(x) = I_1 s_1^2 e^{s_1 x} = \left(\frac{\mathcal{C}_-}{\alpha} - \mathcal{I}_p \right) s_1 e^{s_1(x-L)}.$$

Given that $\mathcal{I}_p < \mathcal{C}_-/\alpha$ and that $s_1 > 0$ in (5.17), we can conclude

$$Y''(x) > 0 \quad \text{for } x < L.$$

For $x \in [L, U]$, we substitute (5.40) and (5.41) for $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ into (5.35), from which we deduce [48]

$$Y''(x) > 0.$$

Turning to $x \in (U, 0]$, we derive [48] from (5.37)

$$\begin{aligned} Y''(x) &= \lambda_1 t_1^2 e^{t_1 x} + \lambda_2 t_2^2 e^{t_2 x}, \\ &= \lambda_1 t_1^2 e^{t_1 x} + \left(\mathcal{D}_p + \frac{\mathcal{C}_-}{\alpha} - \lambda_1 t_1 e^{t_1 U} \right) t_2 e^{t_2(x-U)}, \\ &= \left(\mathcal{D}_p + \frac{\mathcal{C}_-}{\alpha} \right) t_2 e^{t_2(x-U)} + e^{t_1 U} (-t_2) \frac{\mathcal{C}_+ + \mathcal{C}_-}{\alpha} \times \\ &\quad \frac{t_1 e^{t_1(x-U)} - t_2 e^{t_2(x-U)}}{t_1 - t_2}, \\ &= \left[\left(\mathcal{D}_p + \frac{\mathcal{C}_-}{\alpha} \right) t_2 + e^{t_1 U} (-t_2) \frac{\mathcal{C}_+ + \mathcal{C}_-}{\alpha} \right] \cdot e^{t_2(x-U)} + \\ &\quad e^{t_1 U} (-t_2 t_1) \frac{\mathcal{C}_+ + \mathcal{C}_-}{\alpha} \cdot \frac{e^{t_1(x-U)} - e^{t_2(x-U)}}{t_1 - t_2}. \end{aligned} \tag{5.44}$$

Now suppose we have

$$\left(\mathcal{D}_p + \frac{\mathcal{C}_-}{\alpha} \right) t_2 + e^{t_1 U} (-t_2) \frac{\mathcal{C}_+ + \mathcal{C}_-}{\alpha} > 0. \tag{5.45}$$

Then we readily obtain from (5.44) that

$$Y''(x) \geq 0, \quad \text{for all } x \in (U, 0).$$

Therefore, to show that Y is convex on $(U, 0)$, it remains for us to establish (5.45). We first note from (5.24) that

$$\begin{aligned} & t_2(\alpha\mathcal{D}_p + \mathcal{C}_-) + (\mathcal{C}_+ + \mathcal{C}_-)(-t_2) \cdot e^{t_1 U} \\ &= (r_1 + r_2)(\alpha\mathcal{D}_p + \mathcal{C}_-) - \left[B_2 \frac{r_2}{r_1 - r_2} e^{r_1(U-L)} \right. \\ & \quad \left. + J_2 \frac{r_1}{r_1 - r_2} e^{r_2(U-L)} \right], \end{aligned} \quad (5.46)$$

and thus (5.45) holds if and only if

$$(r_1 + r_2)K + r_2 B_2 e^{r_1(U-L)} + r_1 J_2 e^{r_2(U-L)} < 0,$$

which due to (5.23) is equivalent to

$$r_1 B_2 e^{r_1(U-L)} + r_2 J_2 e^{r_2(U-L)} > 0. \quad (5.47)$$

Define

$$f(x) = B_2 e^{r_1 x} + J_2 e^{r_2 x} + K.$$

We deduce that

$$f'(0) = r_1 B_2 + r_2 J_2 = (r_1 - r_2)s_1(\mathcal{C}_- - \alpha\mathcal{I}_p) > 0. \quad (5.48)$$

One can readily verify that f is convex. Upon combining this with (5.48), we obtain

$$f'(U - L) > 0,$$

which is exactly (5.47). Finally, we show that for $x > 0$

$$Y''(x) = l_3 t_2^2 e^{t_2 x} \geq 0.$$

It suffices to show $l_3 \geq 0$, which by (5.43) is equivalent to showing

$$\lambda_2 \geq \frac{\mathcal{C}_+ + \mathcal{C}_-}{\alpha} \left(\frac{1}{t_2} + \frac{1}{t_1 - t_2} \right). \quad (5.49)$$

From (5.37) and (5.42), we know

$$\lambda_2 = (\mathcal{D}_p + \frac{\mathcal{C}_-}{\alpha}) \frac{e^{-t_2 U}}{t_2} - \frac{\sigma^2}{2\alpha^2} (\mathcal{C}_+ + \mathcal{C}_-) \frac{t_1 t_2}{t_1 - t_2} e^{(t_1 - t_2)U}.$$

Upon substituting this expression into (5.49) and multiplying both sides by $t_2(t_1 - t_2)$, we simply need to show

$$(\mathcal{D}_p + \frac{\mathcal{C}_-}{\alpha})(t_1 - t_2)e^{-t_2 U} - \frac{\sigma^2}{2\alpha^2}(\mathcal{C}_+ + \mathcal{C}_-)t_1 t_2^2 e^{(t_1 - t_2)U} \leq \frac{\mathcal{C}_+ + \mathcal{C}_-}{\alpha} t_1. \quad (5.50)$$

Given that $\mathcal{D}_p < \mathcal{C}_+/\alpha$ and using the fact that

$$t_1 t_2 = \frac{-2\alpha}{\sigma^2},$$

then establishing the inequality (5.50) is equivalent to showing [48]

$$e^{-t_2 U}(t_1 - t_2) + t_2 e^{(t_1 - t_2)U} \leq t_1. \quad (5.51)$$

To this end, we set

$$g(y) = e^{-t_2 y}(t_1 - t_2) + t_2 e^{(t_1 - t_2)y} - t_1,$$

and verify

$$g'(y) = -t_2(t_1 - t_2)e^{-t_2 y}(1 - e^{t_1 y}),$$

which implies that

$$g'(y) \geq 0 \quad \text{for } y \leq 0.$$

Combining this with the fact that $g(0) = 0$, we deduce for $U \leq 0$

$$g(U) \leq g(0) = 0,$$

thus showing (5.51). We have therefore established the convexity of the candidate value function Y , which implies (5.32).

5.A.2.2 Uniqueness of the threshold values

Finally, we show that L and U are uniquely determined by two nonlinear equations and provide necessary and sufficient conditions under which the two equations are both negative, both positive, or of different signs. Once again, we focus here on

proving these results for Case II of Theorem 5.3.1, noting that Cases I and III and other technical details can be found in [48]. Since

$$\mathcal{I}_p + \mathcal{D}_p > 0,$$

we obtain

$$B_2 + J_2 + K = \alpha(r_1 - r_2) \cdot (-\mathcal{I}_p - \mathcal{D}_p) < 0.$$

Defining

$$f(x) = B_2 e^{r_1 x} + J_2 e^{r_2 x} + K,$$

one can readily verify that f is convex,

$$f(0) = B_2 + J_2 + K < 0$$

and

$$\lim_{|x| \rightarrow \infty} f(x) = \infty,$$

which implies that $f(x) = 0$ has only one positive solution. We therefore conclude that (5.23) uniquely determines $U - L$ and that (5.24) uniquely determines U . To establish necessary and sufficient conditions under which $U \leq 0$, observe that [48]

$$\begin{aligned} U \leq 0 &\Leftrightarrow e^{t_1 U} \leq 1, \\ &\Leftrightarrow \text{RHS (5.24)} \geq (r_1 + r_2 - t_2)(\alpha \mathcal{D}_p + \mathcal{C}_-) + \\ &\quad (\mathcal{C}_+ + \mathcal{C}_-) t_2, \\ &\Leftrightarrow \text{LHS (5.24)} \geq (r_1 + r_2 - t_2)(\alpha \mathcal{D}_p + \mathcal{C}_-) + \\ &\quad (\mathcal{C}_+ + \mathcal{C}_-) t_2, \\ &\Leftrightarrow e^{r_1(U-L)} \leq \frac{B_3 - B_1}{B_2}. \end{aligned}$$

Considering (5.23) and letting

$$h(y) = B_2 y + J_2 y^{\frac{r_2}{r_1}} + K$$

for $y > 0$, one can readily verify that h is strictly increasing on $[1, \infty)$ with $h(1) < 0$. Guaranteeing that (5.23) has a solution is equivalent to showing that $h(y) = 0$ has a solution in the interval

$$\left(1, \frac{B_3 - B_1}{B_2}\right].$$

Hence, it suffices to show

$$\frac{B_3 - B_1}{B_2} > 1 \quad \text{and} \quad h\left(\frac{B_3 - B_1}{B_2}\right) \geq 0, \quad (5.52)$$

which can be confirmed by simple algebraic manipulations establishing that the conditions (5.52) are the same as the second set of conditions in Theorem 5.3.1.

5.A.3 Proof of Theorem 5.3.1: Step 3

The final step of our proof of Theorem 5.3.1 consists of verifying that the proposed two-threshold dynamic control policy is optimal and that $Y(x) = V(x)$ for all x . We take a martingale argument approach where the key idea is to construct a submartingale to prove that the candidate value function Y is a lower bound for the optimization problem (5.9) – (5.14). To this end, first consider an admissible process $dP(t) = \theta dt$, where $P(t)$ is adapted to the filtration \mathcal{F}_t generated by $\{D(s) : 0 \leq s \leq t\}$ and $\theta \in [\theta_\ell, \theta_u]$. Recalling $X(t) = P(t) - D(t)$, we define the process

$$\begin{aligned} M(t) = & e^{-\alpha t} Y(X(t)) + \int_0^t e^{-\alpha s} (\mathcal{C}_+ X(s)^+ + \mathcal{C}_- X(s)^- \\ & + \mathcal{I}_\rho \theta \mathbb{1}_{\{\theta \geq 0\}} - \mathcal{D}_\rho \theta \mathbb{1}_{\{\theta < 0\}}) ds, \end{aligned}$$

with our goal being to show that $M(t)$ is a submartingale.

Applying Ito's formula to $e^{-\alpha t} Y(X(t))$ renders for any $t_1 \leq t_2$

$$\begin{aligned} M(t_2) - M(t_1) = & \int_{t_1}^{t_2} e^{-\alpha s} \left(-\alpha Y(X(s)) + \frac{1}{2} \sigma^2 Y''(X(s)) \right. \\ & + (\theta - b) Y'(X(s)) + \mathcal{C}_+ X(s)^+ + \mathcal{C}_- X(s)^- \\ & \left. + \mathcal{I}_\rho \theta \mathbb{1}_{\{\theta \geq 0\}} - \mathcal{D}_\rho \theta \mathbb{1}_{\{\theta < 0\}} \right) ds \\ & - \int_{t_1}^{t_2} e^{-\alpha s} Y'(X(s)) \sigma dW(s). \end{aligned} \quad (5.53)$$

We have established in Section 5.A.2 that Y satisfies the Bellman equation

$$-\alpha Y(x) + \frac{1}{2}\sigma^2 Y''(x) - bY'(x) + \mathcal{C}_+ x^+ + \mathcal{C}_- x^- + \inf_{\theta_\ell \leq \theta \leq \theta_u} \mathcal{L}(\theta, x) = 0,$$

where

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

This implies, for any given x and any $\theta \in [\theta_\ell, \theta_u]$, that

$$-\alpha Y(x) + \frac{1}{2}\sigma^2 Y''(x) - bY'(x) + \mathcal{C}_+ x^+ + \mathcal{C}_- x^- + \mathcal{L}(\theta, x) \geq 0.$$

Since $Y'(\cdot)$ is bounded, upon taking the conditional expectation in (5.53) with respect to the filtration \mathcal{F}_{t_1} , we deduce for any $t_1 \leq t_2$ that

$$\mathbb{E}_x[M(t_2)|\mathcal{F}_{t_1}] - M(t_1) \geq 0.$$

Namely, $M(t)$ is a submartingale and therefore we have

$$\mathbb{E}_x[M(t)] \geq M(0) = Y(x), \quad \text{for any } t \geq 0.$$

Letting t go to ∞ , we can conclude that Y is a lower bound for the optimal value of the optimization problem (5.9) - (5.14), and thus $Y(x) = V(x)$ for all x . Hence, the dynamic control policy characterized by the two threshold values L and U is indeed optimal, and our proof of Theorem 5.3.1 is complete.

