Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

The Dynamic and Stochastic Knapsack Problem with Random Sized Items

Anton J. Kleywegt, Jason D. Papastavrou,

THE DYNAMIC AND STOCHASTIC KNAPSACK PROBLEM WITH RANDOM SIZED ITEMS

ANTON J. KLEYWEGT

School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0205
anton.kleywegt@isye.gatech.edu

JASON D. PAPASTAVROU

School of Industrial Engineering, Purdue University, West Lafayette, Indiana 47907-1287
jdp@ecn.purdue.edu

(Received February 1997; revision received August 1998; accepted June 1999)

A resource allocation problem, called the dynamic and stochastic knapsack problem (DSKP), is studied. A known quantity of resource is available, and demands for the resource arrive randomly over time. Each demand requires an amount of resource and has an associated reward. The resource requirements and rewards are unknown before arrival and become known at the time of the demand's arrival. Demands can be either accepted or rejected. If a demand is accepted, the associated reward is received; if a demand is rejected, a penalty is incurred. The problem can be stopped at any time, at which time a terminal value is received that depends on the quantity of resource remaining. A holding cost that depends on the amount of resource allocated is incurred until the process is stopped. The objective is to determine an optimal policy for accepting demands and for stopping that maximizes the expected value (rewards minus costs) accumulated. The DSKP is analyzed for both the infinite horizon and the finite horizon cases. It is shown that the DSKP has an optimal policy that consists of an easily computed threshold acceptance rule and an optimal stopping rule. A number of monotonicity and convexity properties are studied. This problem is motivated by the issues facing a manager of an LTL transportation operation regarding the acceptance of loads and the dispatching of a vehicle. It also has applications in many other areas, such as the scheduling of batch processors, the selling of assets, the selection of investment projects, and yield management.

1. INTRODUCTION

Many practical problems involve the allocation of limited resources to competing demands that arrive randomly over time. In this paper we consider problems where a known amount of resource is available initially and demands arrive according to a Poisson process in time. Each demand requests some amount of the resource and offers a reward if the required amount of resource is allocated to it. The net reward associated with a demand also includes any costs that are incurred to service the demand. The resource requirements and rewards are unknown before arrival and become known at the arrival times of the demands. The joint distribution of the resource requirements and rewards is known, and the resource requirement and reward of a demand are independent of the arrival time of the demand and of the resource requirements and rewards of other demands. The resource requirement and the reward of a demand can be dependent. (In typical applications we would expect a positive correlation between the resource requirement and the reward of a demand.) Demands have to be either accepted or rejected upon arrival and cannot be recalled later. If the resource requirement of a demand is greater than the remaining amount of resource, the demand has to be rejected. If a demand is accepted, its associated reward is received; if it is rejected, a penalty is incurred.

There is a known deadline (possibly infinite) after which demands can no longer be accepted. There is a holding cost per unit time that depends on the total amount of resource already allocated, or equivalently, on the remaining amount of resource. It is allowed to stop waiting for arrivals at any time before or at the deadline, even if all the resource has not been allocated yet. In a freight consolidation application, this occurs when a vehicle is dispatched before it is filled to capacity and before the scheduled deadline is reached. A terminal value is earned that depends on the remaining amount of resource at the stopping time. Rewards and costs may be discounted. The objective is to determine a policy for accepting demands and for stopping that maximizes the expected total discounted value (rewards minus costs) accumulated, among the class of nonanticipatory policies. This problem is called the *dynamic and stochastic knapsack problem* (DSKP).

A special case of the DSKP, in which all demands require the same amount of resource, was studied in a recent paper by Kleywegt and Papastavrou (1998). Here we extend the study to the DSKP with random resource requirements. It is shown that many of the optimality results continue to hold. However, some of the intuitive structural characteristics of optimal solutions for the case with equal resource requirements do not generalize to the case with random resource requirements. Conditions are found under which these characteristics do continue to hold.

Subject classifications: Dynamic programming/optimal control, Markov: resources allocation. Transportation, scheduling: load acceptance dispatching. Production/scheduling, cutting stock: dynamic stock cutting.
Area of review: Transportation.

The investigation of the DSKP was motivated by a problem in the transportation industry, where common carriers carry loads for many different clients. The carriers receive transportation requests for different sized loads randomly over time, and prices are offered or negotiated for transporting the loads. If a load is accepted, costs are incurred for picking up and handling the load, and the negotiated price is received. If a load is rejected, some customer goodwill and possible future sales are lost, represented by a penalty for rejecting loads. Some carriers have a fixed schedule for moving vehicles and thus a deadline after which loads cannot be accepted for a specific shipment. There is also a holding cost per unit time, representing actual warehousing costs as well as the quality of customer service, which is incurred until the shipment is dispatched. The dispatcher can decide to dispatch a vehicle at any time before the deadline. There is also a dispatching and transportation cost that is incurred for the shipment as a whole, which depends on the total load in the shipment. A typical objective is to maximize the expected total profit (revenues minus costs) earned by the transportation operation. The DSKP has many other applications, such as the scheduling of batch processors, the selling of assets, the investment of funds, yield management, and the selling of tickets for events. Some of these were discussed in Kleywegt and Papastavrou (1998).

In contrast to the DSKP, the well-known knapsack problem is a static and deterministic resource allocation problem. Items to be loaded into a knapsack with fixed capacity are selected from a given set of items with known sizes and rewards. The objective is to maximize the total reward, subject to the capacity constraint. This static and deterministic knapsack problem has been studied extensively (Martello and Toth 1990).

Stochastic knapsack problems have also been studied; these can be classified as static or dynamic. In static stochastic knapsack problems the set of items is given, but the rewards and/or sizes are unknown. These problems were investigated by Sniedovich (1980), Carraway et al. (1993), and others. In dynamic stochastic knapsack problems the items arrive over time, and the rewards and/or sizes are unknown before arrival. Papastavrou et al. (1996) considered a finite horizon, discrete-time version of the DSKP, without holding costs. Kleywegt and Papastavrou (1998) studied the continuous-time DSKP with holding costs, for both the finite and infinite horizon cases, where all items have the same size.

The DSKP is related to some stopping time problems and optimal selection problems, such as the well-known secretary problem. This problem has been studied by Presman and Sonin (1972), Stewart (1981), Freeman (1983), and a multitude of others. The problem of selling a single asset has been analyzed by Rosenfield et al. (1983), Mamer (1986), and Albright (1977), among others. A more general problem is the sequential stochastic assignment problem, introduced by Derman et al. (1972). Many investment problems, such as the one studied by Prastacos (1983), are similar to the DSKP. Ross and Tsang (1989) introduced a problem called the stochastic knapsack problem for telecommunications applications. Perishable asset revenue management problems, also called yield management problems, are similar to the DSKP and have applications in the airline, hotel, car rental, cruise line, entertainment, and apparel industries. These problems have been studied widely, among others, by Kincaid and Darling (1963), Belobaba (1989), Stadje (1990), Weatherford and Bodily (1992), Brumelle and McGill (1993), and Gallego and Van Ryzin (1994).

The DSKP as well as a related Markov decision process are defined in §2. The DSKP without a deadline is studied in §3, and the DSKP with a deadline follows in §4. Concluding remarks are given in §5. Detailed proofs of results can be found in Kleywegt (1996) and in the *Operations Research* Online Collection, currently located at WWW URL ⟨http://or.pubs.informs.org⟩.

2. PROBLEM DEFINITION

First a formulation of the DSKP is given, then a Markov decision process model that is closely related to the DSKP is introduced.

2.1. Dynamic and Stochastic Knapsack Problem Formulation

Let N_0 denote the known initial amount of resource. Let $T \in (0, \infty]$ denote the deadline for accepting demands. Let $\{A_i\}_{i=1}^{\infty}$ denote the arrival times of a Poisson process on $(0, T)$ with rate λ. Let S_i denote the resource requirement of arrival i, and let R_i denote the reward of arrival i. Assume that $\{(R_i, S_i)\}_{i=1}^{\infty}$ is an i.i.d. sequence, independent of $\{A_i\}_{i=1}^{\infty}$, but R_i and S_i may be dependent. Let $F_{R,S}$ denote the joint probability distribution of R and S. Let F_S denote the marginal probability distribution of S, and let F_R denote the marginal probability distribution of R. Let $F_{R|S}$ denote the conditional probability distribution of R given S, and let $F_{S|R}$ denote the conditional probability distribution of S given R. Assume that $F_S(0) = 0$, and that $\int_{\Re \times [0, N_0]} |r| F_{R,S}(dr, ds) < \infty$.

Decision D_i denotes whether to accept or reject arrival i, defined as follows:

$$D_i \equiv \begin{cases} 1 & \text{if arrival } i \text{ is accepted,} \\ 0 & \text{if arrival } i \text{ is rejected.} \end{cases}$$

Decision $\mathcal{T} \in [0, T]$ denotes the chosen stopping time. Choosing a stopping time $\mathcal{T}$ is equivalent to choosing a unit step function

$$I(t) \equiv \begin{cases} 1 & \text{if } t \in (0, \mathcal{T}], \\ 0 & \text{if } t \in (\mathcal{T}, T], \end{cases} \tag{1}$$

where $I(t)$ indicates whether the process has been stopped (switched *off*, $I(t) = 0$) or not (switched *on*, $I(t) = 1$) at time t.

Let $\mathscr{A} \equiv \{\{A_i\}_{i=1}^{\infty} : 0 < A_1 < A_2 < \cdots < \infty\}$, let $\mathscr{R} \equiv \{\{R_i\}_{i=1}^{\infty} : R_i \in \Re\}$, let $\mathscr{S} \equiv \{\{S_i\}_{i=1}^{\infty} : S_i \in \Re_+\}$, where $\Re_+ \equiv [0, \infty)$, and let $\mathscr{D} \equiv \{\{D_i\}_{i=1}^{\infty} : D_i \in \{0, 1\}\}$. Let $\mathscr{I}_s$ denote the set of all unit step functions $I_s : (0, T] \mapsto \{0, 1\}$ of the form in (1).

The class $\prod_{\text{DSKP}}^{\text{HD}}$ of history-dependent deterministic policies for the DSKP is defined as the set of all Borel-measurable functions $\pi : \mathscr{A} \times \mathscr{R} \times \mathscr{S} \mapsto \mathscr{D} \times \mathscr{I}_s$, such that the nonanticipatory conditions are satisfied (which require that policies not be allowed to know exactly what is going to happen in the future), and such that the resource constraint

$$\sum_{\{i : A_i \leqslant T^{\pi}\}} D_i^{\pi} S_i \leqslant N_0$$

is satisfied, where $(\{D_i^{\pi}\}, I^{\pi}) \equiv \pi(\{A_i\}, \{R_i\}, \{S_i\})$.

Let $N^{\pi}(t)$ denote the remaining amount of resource under policy π at time t, where N^{π} is defined to be left-continuous, i.e.,

$$N^{\pi}(t) \equiv N_0 - \sum_{\{i : A_i < t\}} D_i^{\pi} I^{\pi}(A_i) S_i, \tag{2}$$

and let the remaining amount of resource just after time t be denoted by

$$N^{\pi}(t^+) \equiv N_0 - \sum_{\{i : A_i \leqslant t\}} D_i^{\pi} I^{\pi}(A_i) S_i. \tag{3}$$

Let $c : [0, N_0] \mapsto \Re$ be Borel-measurable and bounded, where $c(n)$ denotes the holding cost per unit time while the remaining amount of resource is n. Let $p : \Re \times \Re_+ \mapsto \Re$ be Borel-measurable, where $p(r, s)$ denotes the penalty that is incurred if a demand with reward r and resource requirement s is rejected. Assume that $\int_{\Re \times [0, N_0]} |p(r, s)| F_{R,S}(dr, ds) < \infty$. Let $v : [0, N_0] \mapsto \Re$ be Borel-measurable and bounded, where $v(n)$ denotes the terminal value that is earned at time $\mathscr{T}^{\pi}$ if the remaining amount of resource $N^{\pi}(\mathscr{T}^{\pi +}) = n$. Let $\alpha \geqslant 0$ be the discount rate; if $T = \infty$, we assume that $\alpha > 0$.

Let V_{DSKP}^{π} denote the expected total discounted value under policy $\pi \in \Pi_{\text{DSKP}}^{\text{HD}}$, i.e.,

$$\begin{aligned} V_{\text{DSKP}}^{\pi} \equiv E^{\pi}\Bigg[& \sum_{\{i : A_i \leqslant \mathscr{T}^{\pi}\}} e^{-\alpha A_i} [R_i D_i^{\pi} - p(R_i, S_i)(1 - D_i^{\pi})] \\ & - \int_0^{\mathscr{T}^{\pi}} e^{-\alpha \tau} c(N^{\pi}(\tau)) d\tau \\ & + e^{-\alpha \mathscr{T}^{\pi}} v(N^{\pi}(\mathscr{T}^{\pi +})) \Big| N^{\pi}(0^+) = N_0 \Bigg] \\ = E^{\pi}\Bigg[& \sum_{\{i : A_i \leqslant T\}} e^{-\alpha A_i} [R_i D_i^{\pi} - p(R_i, S_i)(1 - D_i^{\pi})] I^{\pi}(A_i) \\ & + \int_0^T e^{-\alpha \tau} \big[-c(N^{\pi}(\tau)) I^{\pi}(\tau) \\ & \qquad + \alpha v(N^{\pi}(\tau))(1 - I^{\pi}(\tau))\big] d\tau \\ & + e^{-\alpha T} v(N^{\pi}(T^+)) \Big| N^{\pi}(0^+) = N_0 \Bigg]. \end{aligned} \tag{4}$$

The equality results from the fact that a terminal value of $v(n)$ at time $\mathscr{T} \leqslant T$ has the same discounted value as a terminal value of $v(n)$ at time T plus interest of $\alpha v(n)$ per unit time from time $\mathscr{T}$ to time T.

The objective is to find the optimal expected value V_{DSKP}^*, i.e.,

$$V_{\text{DSKP}}^* \equiv \sup_{\pi \in \prod_{\text{DSKP}}^{\text{HD}}} V_{\text{DSKP}}^{\pi},$$

and to find an optimal policy $\pi^* \in \prod_{\text{DSKP}}^{\text{HD}}$ that achieves this optimal value, if such a policy exists.

2.2. A Related Markov Decision Process

If the DSKP did not include the option to stop at any time before the deadline, then the problem could be formulated as a continuous time Markov decision process (MDP). However, the option of choosing a stopping time in continuous time for the DSKP introduces a complexity into the problem that is not modeled in a straightforward way by an MDP. If a *stopped* state is introduced, then the process should make an infinite rate transition to this state as soon as the decision is made to stop. However, most results for MDPs assume that transition rates between states are bounded. Therefore an MDP is studied that is a relaxation of the DSKP, in that the MDP can switch *off* and *on* multiple (even an infinite number of) times, instead of stopping only once. This MDP has bounded transition rates between the states, because switching *off* and *on* involves choosing a different action only and not making a transition to a different state. It is shown in §3 for the infinite horizon case (no deadline) and in §4 for the finite horizon case (finite deadline) that there exist optimal policies for the MDP that switch *off* only once and remain switched *off* and, hence, are admissible and optimal for the DSKP.

The initial amount N_0 of resource, deadline T, arrival process $\{A_i\}$, rewards $\{R_i\}$, resource requirements $\{S_i\}$, and probability distributions are the same for the MDP as for the DSKP. The MDP has state space $[0, N_0]$. Similar to the DSKP, decision D_i denotes whether to accept $(D_i = 1)$ or reject $(D_i = 0)$ arrival i. Decision $I(t)$ denotes whether to be switched *on* $(I(t) = 1)$ or *off* $(I(t) = 0)$ at time t. In contrast to the DSKP, the MDP can switch *off* and *on* multiple times, and thus the function $I(t)$ can be any Borel-measurable indicator function on $(0, T]$, and not just a unit step function as in the case of the DSKP. Let $\mathscr{I}_I$ denote the set of all Borel-measurable indicator functions $I_I : (0, T] \mapsto \{0, 1\}$. Note that $\mathscr{I}_I \supset \mathscr{I}_s$.

The class $\prod_{\text{MDP}}^{\text{HD}}$ of history-dependent deterministic policies for the MDP is defined as the set of all Borel-measurable functions $\pi : \mathscr{A} \times \mathscr{R} \times \mathscr{S} \mapsto \mathscr{D} \times \mathscr{I}_I$, such that the nonanticipatory conditions and the resource constraint

$$\sum_{\{i : A_i \leqslant T\}} D_i^{\pi} I^{\pi}(A_i) S_i \leqslant N_0$$

are satisfied, where $(\{D_i^{\pi}\}, I^{\pi}) \equiv \pi(\{\mathscr{A}_i\}, \{\mathscr{R}_i\}, \{\mathscr{S}_i\})$. Because $\mathscr{I}_s \subset \mathscr{I}_I$, any admissible policy for the DSKP is also admissible for the MDP ($\prod_{\text{DSKP}}^{\text{HD}} \subset \prod_{\text{MDP}}^{\text{HD}}$). The state $N^\pi(t)$ (remaining amount of resource) under policy π at time t is given by (2) and (3).

The holding cost $c(n)$, rejection penalty $p(r, s)$, terminal value $v(n)$, and discount rate α are the same for the MDP as for the DSKP.

V_{MDP}^π denotes the expected total discounted value under policy $\pi \in \prod_{\text{MDP}}^{\text{HD}}$, given by

$$\begin{aligned} V_{\text{MDP}}^\pi \equiv E^\pi\Bigg[& \sum_{\{i: A_i \leqslant T\}} e^{-\alpha A_i}[R_i D_i^\pi - p(R_i, S_i)(1-D_i^\pi)] I^\pi(A_i) \\ & + \int_0^T e^{-\alpha\tau}[-c(N^\pi(\tau)) I^\pi(\tau) \\ & \qquad + \alpha v(N^\pi(\tau))(1-I^\pi(\tau))] d\tau \\ & + e^{-\alpha T} v(N^\pi(T^+)) \Bigg| N^\pi(0^+) = N_0 \Bigg]. \end{aligned} \tag{5}$$

The optimal expected value V_{MDP}^* is given by

$$V_{\text{MDP}}^* \equiv \sup_{\pi \in \prod_{\text{MDP}}^{\text{HD}}} V_{\text{MDP}}^\pi.$$

Note that when the MDP is switched *off* ($I^\pi(t) = 0$) at time t, not only are no arrivals accepted (which could be modeled by setting $D_i^\pi = 0$), but also no penalty is incurred for rejected demands, no holding cost is incurred, and interest of $\alpha v(N^\pi(t))$ per unit time is earned. This cost structure makes the objective function of the MDP similar to the objective function of the DSKP. Specifically, comparing (4) and (5), it follows that for any policy $\pi \in \prod_{\text{DSKP}}^{\text{HD}}$, $V_{\text{MDP}}^\pi = V_{\text{DSKP}}^\pi$. Hence the MDP is a relaxation of the DSKP. The optimal expected value of the MDP is therefore at least as good as that of the DSKP. This is the result of Lemma 1.

LEMMA 1. $V_{\text{MDP}}^* \geqslant V_{\text{DSKP}}^*$.

The class $\prod_{\text{MDP}}^{\text{MD}}$ of memoryless deterministic policies for the MDP is the subset of the history-dependent deterministic policies $\prod_{\text{MDP}}^{\text{HD}}$ for the MDP that depends only on the current state of the process and time, where the state is defined as the remaining amount of resource n. Specifically, let $\mathcal{I}_{RS}$ denote the set of all Borel-measurable functions $D_{RS}: \Re \times \Re_+ \mapsto \{0, 1\}$. Then the class $\prod_{\text{MDP}}^{\text{MD}}$ of memoryless deterministic policies for the MDP is the set of all Borel-measurable functions $\pi: [0, N_0] \times [0, T] \mapsto \mathcal{I}_{RS} \times \{0, 1\}$, where $\pi \equiv (D^\pi, I^\pi)$. $D^\pi(n, t, r, s)$ denotes the decision under policy π whether to accept or reject an arrival i at time $A_i = t$ with resource requirement $S_i = s$ and reward $R_i = r$ if the remaining amount of resource $N^\pi(t) = n$, as follows:

$$D^\pi(n, t, r, s) \equiv \begin{cases} 1 & \text{if } s \leqslant n \text{ and arrival } i \text{ is accepted,} \\ 0 & \text{if } s > n \text{ or arrival } i \text{ is rejected .} \end{cases}$$

Let the acceptance set for policy π be denoted by $\mathcal{R}_1^\pi(n, t) \equiv \{(r, s) \in \Re \times \Re_+ : D^\pi(n, t, r, s) = 1\}$, and let the rejection set be denoted by $\mathcal{R}_0^\pi(n, t) \equiv \{(r, s) \in \Re \times \Re_+ : D^\pi(n, t, r, s) = 0\}$. $I^\pi(n, t)$ denotes the decision under policy π whether to be switched *on* or *off* at time t if the remaining amount of resource $N^\pi(t) = n$, as follows:

$$I^\pi(n,\ t) \equiv \begin{cases} 1 & \text{if switched } on \text{ at time } t, \\ 0 & \text{if switched } off \text{ at time } t. \end{cases}$$

Clearly, policies $\pi \in \prod_{\text{MDP}}^{\text{MD}}$ satisfy the nonanticipatory conditions and the resource constraint imposed on policies in $\prod_{\text{MDP}}^{\text{HD}}$. As for policies in $\prod_{\text{MDP}}^{\text{HD}}$, policies $\pi \in \prod_{\text{MDP}}^{\text{MD}}$ are allowed to switch *on* and *off* multiple times.

Under a policy $\pi \in \prod_{\text{MDP}}^{\text{MD}}$, the MDP can be modeled with transition rates

$$\lambda(\pi(n, t)) \equiv \lambda I^\pi(n, t)$$

and Markov kernels

$$P[n | n, \pi(n, t)] \equiv \int_{\mathcal{R}_0^\pi(n, t)} F_{R,S}(dr, ds)$$

$$P[n - B | n, \pi(n, t)] \equiv \int_{\mathcal{R}_1^\pi(n, t) \cap (\Re \times B)} F_{R,S}(dr, ds)$$

for any Borel set $B \subseteq (0, \infty)$.

Let $V^\pi(n, t)$ be the expected total discounted value under policy $\pi \in \prod_{\text{MDP}}^{\text{MD}}$ from time t until time T, if the remaining capacity $N^\pi(t^+) = n$, i.e.,

$$\begin{aligned} V^\pi(n,t) \equiv E^\pi\Bigg[& \sum_{\{i: A_i \in (t, T]\}} e^{-\alpha(A_i - t)}[R_i D^\pi(N^\pi(A_i), A_i, R_i, S_i) \\ & - p(R_i, S_i)(1 - D^\pi(N^\pi(A_i), A_i, R_i, S_i))] \\ & \cdot I^\pi(N^\pi(A_i), A_i) \\ & + \int_t^T e^{-\alpha(\tau - t)}[-c(N^\pi(\tau)) I^\pi(N^\pi(\tau), \tau) \\ & \qquad + \alpha v(N^\pi(\tau))(1 - I^\pi(N^\pi(\tau), \tau))] d\tau \\ & + e^{-\alpha(T-t)} v(N^\pi(T^+)) \Bigg| N^\pi(t^+) = n \Bigg]. \end{aligned} \tag{6}$$

Note that if $I^\pi(n, t) = 0$ for all $t \in (t_1, T)$, then $V^\pi(n, t) = v(n)$ for all $t \in [t_1, T]$. Let $V^*(n, t)$ be the corresponding optimal expected value, i.e.,

$$V^*(n, t) \equiv \sup_{\pi \in \prod_{\text{MDP}}^{\text{MD}}} V^\pi(n, t).$$

Note that $V^*(n, t) \geqslant v(n)$ for all n and t, because the policy $\pi \in \prod_{\text{MDP}}^{\text{MD}}$ with $I^\pi = 0$ has $V^\pi(n, t) = v(n)$ for all n and t.

Kleywegt and Papastavrou (1998) showed that if all demands require the same amount of resource, then some intuitive properties hold (i.e., $V^*(n, t)$ decreases as the deadline approaches, and $V^*(n, t)$ decreases as the remaining amount of resource decreases), under conditions that typically hold in practice (i.e., the holding cost increases as the amount of allocated resource increases, and the terminal value decreases as the final remaining amount of resource decreases). These properties continue to hold if the demands have random resource requirements.

Proposition 1. *For any* $n \in [0, N_0]$, *and* $T \in (0, \infty]$, $V^*(n, t)$ *is nonincreasing in* t *on* $[0, T]$.

Proposition 2. *If* $c(n)$ *is nonincreasing and* $v(n)$ *is nondecreasing, then for any* $T \in (0, \infty]$, *and* $t \in [0, T]$, $V^*(n, t)$ *is nondecreasing in* n *on* $[0, N_0]$.

A policy $\pi \in \prod_{\text{MDP}}^{\text{MD}}$ is said to be a *threshold* policy if it has a threshold acceptance rule D^π with a reward threshold $x^\pi : [0, N_0] \times [0, T] \times [0, N_0] \mapsto \Re$ that works as follows. If a demand arrives at time t when the remaining amount of resource $N^\pi(t) = n$, then the demand is accepted if the resource requirement $s \leqslant n$ and the reward $r > x^\pi(n, t, s)$, and the demand is rejected if $s > n$ or $r < x^\pi(n, t, s)$. That is,

$$D^\pi(n, t, r, s) = \begin{cases} 1 & \text{if } s \leqslant n \text{ and } r > x^\pi(n, t, s), \\ 0 & \text{if } s > n \text{ or } r < x^\pi(n, t, s). \end{cases}$$

The decision if $s \leqslant n$ and $r = x^\pi(n, t, s)$ may depend on additional details. To develop some intuition, suppose a demand with resource requirement $s \leqslant n$ and reward r arrives at time t when the remaining amount of resource is n and $I^*(n, t) = 1$. If the demand is accepted, the optimal expected value from then on is $r + V^*(n - s, t)$. If the demand is rejected, the optimal expected value from then on is $V^*(n, t) - p(r, s)$. Hence, the demand is accepted if $r + V^*(n - s, t) \geqslant V^*(n, t) - p(r, s)$, i.e., if $r + p(r, s) \geqslant V^*(n, t) - V^*(n - s, t)$; otherwise, the demand is rejected. It is shown that under appropriate conditions there is a threshold policy π^* with reward threshold $x^*(n, t, s)$ and appropriate switching rule $I^*(n, t)$ that is optimal among all policies $\pi \in \prod_{\text{MDP}}^{\text{MD}}$. It also follows that such a memoryless policy π^* is optimal for the MDP among all history-dependent policies $\pi \in \prod_{\text{MDP}}^{\text{HD}}$, and that π^* is admissible for the DSKP, i.e., once policy π^* has switched off, it remains switched off. From this it follows that π^* is also an optimal policy for the DSKP among all policies $\pi \in \prod_{\text{DSKP}}^{\text{HD}}$.

Under typical conditions the reward threshold x^* is equivalent to a size threshold $z^* : [0, N_0] \times [0, T] \times \Re \mapsto [0, N_0]$, defined by

$$z^*(n, t, r) \equiv \sup\{s \in [0, n] : r + p(r, s) \geqslant V^*(n, t) - V^*(n - s, t)\}$$

(where $\sup \varnothing \equiv -\infty$). If a demand with reward r arrives at time t when the remaining amount of resource is n, then the demand is accepted if the resource requirement $s < z^*(n, t, r) \leqslant n$, and the demand is rejected if $s > z^*(n, t, r)$. That is,

$$D^*(n, t, r, s) \equiv \begin{cases} 1 & \text{if } s < z^*(n, t, r), \\ 0 & \text{if } s > z^*(n, t, r). \end{cases}$$

The decision if $s = z^*(n, t, r)$ may depend on additional details. It is shown that, under conditions that usually hold in applications, there is an optimal policy with size threshold z^*.

The class $\prod_{\text{MDP}}^{\text{SD}}$ of stationary deterministic policies for the MDP is the subset of $\prod_{\text{MDP}}^{\text{MD}}$ of policies π that do not depend on t. Clearly, stationary policies $\pi \in \prod_{\text{MDP}}^{\text{SD}}$ have unit step function stopping rules $I^\pi \in \mathcal{I}_s$. Therefore, any stationary policy $\pi \in \prod_{\text{MDP}}^{\text{SD}}$ is admissible for the DSKP, and $V_{\text{DSKP}}^\pi = V_{\text{MDP}}^\pi$.

3. THE INFINITE HORIZON DSKP

The dynamic and stochastic knapsack problem with no deadline is studied in this section. In §3.1 we show that under certain assumptions attention can be restricted to the class of stationary policies $\prod_{\text{MDP}}^{\text{SD}}$. In §3.2 an equation is given for the expected value V^π of a policy $\pi \in \prod_{\text{DSKP}}^{\text{SD}}$. In §3.3 the optimality equation is derived, giving the optimal expected value function V^*, and a stationary threshold policy π^* is obtained that is optimal for the MDP among all policies $\pi \in \prod_{\text{MDP}}^{\text{HD}}$. This policy π^* is also optimal for the DSKP among all policies $\pi \in \prod_{\text{DSKP}}^{\text{HD}}$. An intuitive recursive form that gives the optimal value function V^* is investigated in §3.4.

3.1. Optimality of Stationary Policies

It is known that an infinite horizon discounted MDP with a general state space may not be well behaved. For example, it was shown by Blackwell (1965) that there may not be an ε-optimal Borel-measurable policy, and the optimal value function $V^*(n)$ may not be Borel-measurable. However, it was shown by Yushkevich and Feinberg (1979) (Theorem 2) for an infinite horizon discounted MDP with a countable state space and a measurable action space that if the reward rate is bounded, then for any $\varepsilon > 0$ there exists a stationary deterministic policy $\pi_\varepsilon \in \prod_{\text{MDP}}^{\text{SD}}$ that is ε-optimal among all history-dependent deterministic policies $\prod_{\text{MDP}}^{\text{HD}}$. Therefore, attention is restricted to cases with a countable state space and the class of stationary policies $\prod_{\text{MDP}}^{\text{SD}}$, which are also admissible for the DSKP. To obtain a countable state space, assume that the resource requirement distribution is discrete with probability mass function f_S such that $\inf\{s \in \Re_+ : f_S(s) > 0\} > 0$. Then at most a finite number of demands with positive resource requirements can be accepted, because $N_0 < \infty$. Thus the set of feasible combinations of positive resource requirements accepted is a subset of a finite cartesian product of the countable set $\{s \in [0, N_0] : f_S(s) > 0\}$. Therefore, the state space can be restricted to a countable set, because any finite cartesian product of countable sets is countable.

For $\pi \in \prod_{\text{MDP}}^{\text{SD}}$, D^π is a function of n, r, and s only, and I^π and V^π are functions of n only. This also means that stopping times are restricted to the starting time and the times when the state (remaining amount of resource) changes, i.e., those arrival times when demands are accepted. For the case where all demands require the same (unit) amount of resource, this implies a stopping amount of resource $m^\pi \equiv \max\{n \in \{0, 1, \ldots, N_0\} : I^\pi(n) = 0\}$ for each policy $\pi \in \prod_{\text{MDP}}^{\text{SD}}$. This property does not extend in general to the case where demands require random amounts of resource.

3.2. The Value of a Policy

If $I^\pi(n) = 0$, then $V^\pi(n) = v(n)$. If $I^\pi(n) = 1$, then by conditioning on the arrival time, the resource requirement, and the reward of the next arrival, it follows that

$$\alpha V^\pi(n) = \lambda \int_{\mathcal{R}_1^\pi(n)} \{r + p(r,s) - [V^\pi(n) - V^\pi(n-s)]\} \cdot F_{R,S}(dr, ds) - [\lambda \bar{p} + c(n)], \tag{7}$$

and

$$V^\pi(n) = \frac{\lambda \int_{\mathcal{R}_1^\pi(n)} [r + p(r,s) + V^\pi(n-s)] F_{R,S}(dr,ds) - [\lambda \bar{p} + c(n)]}{\alpha + \lambda \int_{\mathcal{R}_1^\pi(n)} F_{R,S}(dr,ds)}$$

where $\bar{p} \equiv E[p(R,S)] = \int_{\Re \times \Re_+} p(r,s) F_{R,S}(dr,ds) < \infty$.

3.3. The Optimal Value Function and an Optimal Policy

From the results in Yushkevich and Feinberg (1979), it follows that under the conditions stated in §3.1, V^* is the unique bounded solution of

$$\alpha V^*(n) = \sup_{\mathcal{R}_1(n), I(n)} \left\{ \left[\lambda \int_{\mathcal{R}_1(n)} \{r - [V^*(n) - V^*(n-s) - p(r,s)]\} \cdot F_{R,S}(dr,ds) - [\lambda \bar{p} + c(n)] \right] I(n) + \alpha v(n)[1 - I(n)] \right\}. \tag{8}$$

This enables us to derive an equation for the optimal value function V^* and to show conditions under which there exists a stationary threshold policy that is optimal among all history-dependent deterministic policies for both the DSKP and the MDP.

THEOREM 1. *The optimal expected value function V^* is the unique bounded solution of*

$$\alpha V^*(n) = \max \left\{ \lambda \int_{\mathcal{R}_1^*(n)} \{r + p(r,s) - [V^*(n) - V^*(n-s)]\} \cdot F_{R,S}(dr,ds) - [\lambda \bar{p} + c(n)], \quad \alpha v(n) \right\},$$

where

$$\mathcal{R}_1^*(n) \equiv \{(r,s) \in \Re \times [0,n] : r + p(r,s) \geqslant V^*(n) - V^*(n-s)\}.$$

The following stationary deterministic policy π^* is an optimal policy among all history-dependent deterministic policies for the DSKP and the MDP. An optimal acceptance rule is

$$D^*(n,r,s) = \begin{cases} 1 & \text{if } s \leqslant n \text{ and } r + p(r,s) \geqslant V^*(n) - V^*(n-s), \\ 0 & \text{if } s > n \text{ or } r + p(r,s) < V^*(n) - V^*(n-s). \end{cases}$$

That is, $\mathcal{R}_1^*(n)$ is an optimal acceptance set. An optimal stopping rule is

$$I^*(n) = \begin{cases} 1 & \text{if } \lambda \int_{\mathcal{R}_1^*(n)} \{r + p(r,s) - [V^*(n) - V^*(n-s)]\} F_{R,S}(dr,ds) - [\lambda \bar{p} + c(n)] \geqslant \alpha v(n), \\ 0 & \text{if } \lambda \int_{\mathcal{R}_1^*(n)} \{r + p(r,s) - [V^*(n) - V^*(n-s)]\} F_{R,S}(dr,ds) - [\lambda \bar{p} + c(n)] < \alpha v(n). \end{cases}$$

Consider another MDP where the reward of demand i is given by $R_i + p(R_i, S_i)$, the holding cost is $\lambda \bar{p} + c(n)$ when the remaining amount of resource is n, and there is no penalty for rejecting a demand. From Theorem 1 it is seen that the optimal value function V_2^* and optimal policy π_2^* for the new MDP is the same as the optimal value function V_1^* and optimal policy π_1^* for the original MDP and DSKP. Thus for the purpose of finding the optimal value function V_1^* and optimal policy π_1^* for the original MDP and DSKP, it is sufficient to consider an MDP with no penalties, with the understanding that the new reward R_i represents $R_i + p(R_i, S_i)$, the distributions $F_{R,S}$, F_R, $F_{R|S}$, and $F_{S|R}$ are for the new reward R, and the holding cost $c(n)$ represents $\lambda \bar{p} + c(n)$. In the remainder of §3, the new MDP is considered.

Corollary 1 establishes the intuitive result that there exists an optimal policy with reward threshold $x^*(n,s) = V^*(n) - V^*(n-s)$.

COROLLARY 1. *The optimal expected value function V^* is the unique bounded solution of*

$$\alpha V^*(n) = \max \left\{ \lambda \int_0^n \int_{x^*(n,s)}^\infty \{r - [V^*(n) - V^*(n-s)]\} \cdot F_{R|S}(dr|s) F_S(ds) - c(n), \quad \alpha v(n) \right\},$$

where optimal reward threshold $x^*(n,s) = V^*(n) - V^*(n-s)$. The following stationary deterministic threshold policy π^* is an optimal policy among all history-dependent deterministic policies for the DSKP and the MDP. An optimal acceptance rule is

$$D^*(n,r,s) = \begin{cases} 1 & \text{if } s \leqslant n \text{ and } r \geqslant x^*(n,s), \\ 0 & \text{if } s > n \text{ or } r < x^*(n,s). \end{cases}$$

An optimal stopping rule is

$$I^*(n) = \begin{cases} 1 & \text{if } \lambda \int_0^n \int_{x^*(n,s)}^\infty \{r - [V^*(n) - V^*(n-s)]\} \cdot F_{R|S}(dr|s) F_S(ds) - c(n) \geqslant \alpha v(n), \\ 0 & \text{if } \lambda \int_0^n \int_{x^*(n,s)}^\infty \{r - [V^*(n) - V^*(n-s)]\} \cdot F_{R|S}(dr|s) F_S(ds) - c(n) < \alpha v(n). \end{cases}$$

It is also intuitive that there exists an optimal policy with size threshold

$$z^*(n,r) \equiv \sup\{s \in [0,n] : r \geqslant V^*(n) - V^*(n-s)\}$$

(where $\sup \varnothing \equiv -\infty$). Corollary 2 establishes that if $V^*(n)$ is nondecreasing, then there exists an optimal acceptance rule with size threshold $z^*(n, r)$. Note that Proposition 2 established that if the intuitive conditions hold that $c(n)$ is nonincreasing and $v(n)$ is nondecreasing, then $V^*(n)$ is nondecreasing.

Corollary 2. *Suppose $V^*(n)$ is nondecreasing. Then the optimal expected value function V^* is the unique bounded solution of*

$$\alpha V^*(n) = \max\left\{\lambda \int_{-\infty}^{\infty}\int_0^{z^*(n,r)} \{r - [V^*(n) - V^*(n-s)]\} \cdot F_{S|R}(ds|r)F_R(dr) - c(n), \quad \alpha v(n)\right\}.$$

The following stationary deterministic threshold policy π^* is an optimal policy among all history-dependent deterministic policies for the DSKP and the MDP. An optimal acceptance rule is

$$D^*(n, r, s) = \begin{cases} 1 & \text{if } s < z^*(n, r), \\ 0 & \text{if } s > z^*(n, r). \end{cases}$$

If $s = z^*(n, r)$, then $D^*(n, r, s) = 1$ if $r \geqslant V^*(n) - V^*(n - z^*(n, r))$, and $D^*(n, r, s) = 0$ otherwise. An optimal stopping rule is

$$I^*(n) = \begin{cases} 1 & \text{if } \lambda \int_{-\infty}^{\infty}\int_0^{z^*(n,r)} \{r - [V^*(n) - V^*(n-s)]\} \\ & \quad \cdot F_{S|R}(ds|r)F_R(dr) - c(n) \geqslant \alpha v(n), \\ 0 & \text{if } \lambda \int_{-\infty}^{\infty}\int_0^{z^*(n,r)} \{r - [V^*(n) - V^*(n-s)]\} \\ & \quad \cdot F_{S|R}(ds|r)F_R(dr) - c(n) < \alpha v(n). \end{cases}$$

It is easy to see that if $z^*(n, r) > -\infty$, and $V^*(n)$ is upper semicontinuous (and thus $V^*(n-s)$ is upper semicontinuous in s), then $r \geqslant V^*(n) - V^*(n - z^*(n, r))$, and $D^*(n, r, z^*(n, r)) = 1$ is an optimal decision. In turn, $V^*(n)$ is upper semicontinuous if $c(n)$ is lower semicontinuous and $v(n)$ is upper semicontinuous. Also, because $V^*(n)$ is nondecreasing, $V^*(n)$ being upper semicontinuous is equivalent to $V^*(n)$ being right-continuous. However, the converse does not necessarily hold, namely if $V^*(n)$ is lower semicontinuous, then it does not necessarily hold that $r \leqslant V^*(n) - V^*(n - z^*(n, r))$.

3.4. A Recursion for the Optimal Expected Value

In Kleywegt and Papastavrou (1998), for the case where all demands require the same amount of resource, a sequence $\{\psi(n)\}_{n=0}^{N_0}$ of policies was inductively defined with the property that $V^*(n) = \max\{v(n), V^{\psi(n)}(n)\}$ for all $n \in \{0, 1, \ldots, N_0\}$. This property gives an intuitive characterization of the optimal value function and an optimal policy and leads to an efficient computational method. In this section a similar sequence of policies $\{\psi(n)\}$ is defined that work as follows. While the remaining amount of resource is n, policy $\psi(n)$ continues to wait ($I^{\psi(n)}(n) = 1$), using a specific acceptance rule to be defined later. When the remaining amount of resource decreases to $n-s$ with the acceptance of a demand requiring s amount of resource, policy $\psi(n)$ stops if $V^{\psi(n-s)}(n-s) \leqslant v(n-s)$; otherwise, policy $\psi(n)$ continues, using the same acceptance rule as policy $\psi(n-s)$.

For such an inductive definition of policies to make sense, assume that the resource requirement distribution F_S is such that the possible values of the remaining amount of resource are $0 = n_0 < n_1 < n_2 < \cdots < N_0$. This is the case, for example, if resource requirements are measured in multiples of a unit size, such as if all resource requirements are positive integers. As the positive numbers that can be represented in a computer satisfy this property, and solutions are usually computed numerically, this assumption is not very restrictive. Without loss of generality, to simplify notation, assume that $n_k = k$ for all k.

Inductively define the sequence of threshold policies $\{\psi(n)\}_{n=0}^{N_0}$ as follows. Let $\mathscr{R}_1^{\psi(0)}(0) \equiv \varnothing$, and $I^{\psi(0)}(0) \equiv 1$. From (7), $\alpha V^{\psi(0)}(0) = -c(0)$. Let $\psi(n-s)$ and $V^{\psi(n-s)}(n') < \infty$ be defined for all $s \in \{1, \ldots, n\}$ and all $n' \in \{0, 1, \ldots, n-s\}$. Let $\widehat{V}(n)$ be the unique solution of

$$\alpha \widehat{V}(n) = \lambda \int_0^n \int_{\hat{x}(n,s)}^{\infty} \left\{r - \left[\widehat{V}(n) - \max\{v(n-s), V^{\psi(n-s)}(n-s)\}\right]\right\} F_{R|S}(dr|s) \cdot F_S(ds) - c(n), \tag{9}$$

where $\hat{x}(n, s) \equiv \widehat{V}(n) - \max\{v(n-s), V^{\psi(n-s)}(n-s)\}$. It is shown in Kleywegt (1996) that (9) has a unique solution $\widehat{V}(n)$. Let $I^{\psi(n)}(n) \equiv 1$, and $x^{\psi(n)}(n, s) \equiv \widehat{V}(n) - \max\{v(n-s), V^{\psi(n-s)}(n-s)\}$ for all $s \leqslant n$. Let

$$I^{\psi(n)}(n-s) \equiv \begin{cases} 1 & \text{if } V^{\psi(n-s)}(n-s) > v(n-s), \\ 0 & \text{if } V^{\psi(n-s)}(n-s) \leqslant v(n-s), \end{cases}$$

and $x^{\psi(n)}(n-s, s') \equiv x^{\psi(n-s)}(n-s, s')$ for all $s \leqslant n$ and $s' \leqslant n-s$. Then it follows by induction that $V^{\psi(n)}(n-s) = \max\{v(n-s), V^{\psi(n-s)}(n-s)\}$ for all $s \in \{1, \ldots, n\}$. It follows from (7) that $V^{\psi(n)}(n)$ satisfies

$$\alpha V^{\psi(n)}(n) = \lambda \int_0^n \int_{x^{\psi(n)}(n,s)}^{\infty} \left\{r - [V^{\psi(n)}(n) - \max\{v(n-s), V^{\psi(n-s)}(n-s)\}]\right\} F_{R|S}(dr|s) \cdot F_S(ds) - c(n) \tag{10}$$

From (9), (10) has a solution $V^{\psi(n)}(n) = \widehat{V}(n)$, and it is shown in Kleywegt (1996) that this is the unique solution of (10). We are now in a position to establish that $V^*(n) = \max\{v(n), V^{\psi(n)}(n)\}$ for all n.

Theorem 2. *The optimal expected value function V^* satisfies $V^*(n) = \max\{v(n), V^{\psi(n)}(n)\}$ for all n.*

Therefore, there exists an optimal policy with the following intuitive structure. For each remaining amount n of resource, if $V^{\psi(n)}(n) > v(n)$, then continue (i.e., $I^*(n) = 1$), using reward threshold $x^*(n, s) = x^{\psi(n)}(n, s) = V^{\psi(n)}(n) -$

$\max\{v(n-s), V^{\psi(n-s)}(n-s)\} = V^*(n) - V^*(n-s)$ for all $s \leqslant n$; else stop (i.e., $I^*(n) = 0$), and collect $v(n)$. This characterization leads to an algorithm, similar to that in Kleywegt and Papastavrou (1998), to compute the optimal value function V^* and optimal policy π^* by processing each state n only once (and thus in $\Theta(N_0)$ time, if solving (10) is counted as an operation for each value of n). This is more efficient than the iterative methods, such as value iteration (successive approximation) and policy iteration, that are typically used to solve infinite horizon MDPs and that approximate the value of each state n multiple times.

4. THE FINITE HORIZON DSKP

In this section the DSKP with a finite deadline is studied. First an equation is given in §4.1 for the value of a policy $\pi \in \prod_{\text{MDP}}^{\text{MD}}$. In §4.2 the optimality equation for the MDP is given; with it an optimal policy for the MDP is also found. In §4.3 the optimal policy for the MDP is shown to be admissible and optimal for the DSKP. A discretization method, which is important for computational purposes as well as for establishing certain properties of optimal solutions, is studied in §4.4, where it is shown that the discrete approximations converge uniformly to the optimal solution V^*. A number of structural characteristics of the optimal value function and optimal policy are established in §4.5. It is shown that under certain conditions the optimal value function and optimal policy are given by a recursion similar to that of the infinite horizon case. A number of monotonicity and convexity properties of the optimal value function and the optimal policy are also studied and are shown to deviate from the intuitive behavior found for the case where all demands require the same amount of resource, unless certain conditions are satisfied.

4.1. The Value of a Policy

The expected value function V^π under a policy $\pi \in \prod_{\text{MDP}}^{\text{MD}}$ satisfies (11), which can be derived by conditioning on whether an arrival takes place in the next Δt time units and on the resource requirement s and reward r of the demand if there is an arrival. It is shown in Kleywegt (1996) that (11) has a unique solution V^π.

THEOREM 3. *The expected value function V^π of policy $\pi \in \prod_{\text{MDP}}^{\text{MD}}$ is the unique solution of the differential equation*

$$\begin{aligned}\frac{\partial V^\pi}{\partial t}(n,t) = \Bigg[-\lambda \int_{\mathscr{R}_1^\pi(n,t)} &\{r + p(r,s) - [V^\pi(n,t) \\ &- V^\pi(n-s,t)]\} F_{R,S}(dr,ds) + \alpha V^\pi(n,t) \\ &+ \lambda \bar{p} - c(n)\Bigg] I^\pi(n,t) \\ &+ \big[\alpha V^\pi(n,t) - \alpha v(n)\big]\big[1 - I^\pi(n,t)\big], \qquad (11)\end{aligned}$$

with boundary condition $V^\pi(n,T) = v(n)$ for all $n \in [0, N_0]$.

4.2. The Optimal Value and an Optimal Policy for the MDP

From the results in Boel and Varaiya (1977), if the following differential equation has a solution $\widehat{V}^*$, and there is a policy $\pi^* \in \prod_{\text{MDP}}^{\text{MD}}$ that attains the supremum, then $\widehat{V}^* = V^*$, the optimal value function for the MDP, and π^* is an optimal policy for the MDP among all history-dependent deterministic policies $\pi \in \prod_{\text{MDP}}^{\text{HD}}$.

$$\begin{aligned}-\frac{\partial V}{\partial t}(n,t) &= \sup_{\mathscr{R}_1(n,t),\, I(n,t)} \Bigg\{ \Bigg[\lambda \int_{\mathscr{R}_1(n,t)} \{r + p(r,s) - [V(n,t) \\ &\qquad - V(n-s,t)]\} F_{R,S}(dr,ds) - \alpha V(n,t) \\ &\qquad - \lambda \bar{p} - c(n) \Bigg] I(n,t) + \big[-\alpha V(n,t) \\ &\qquad + \alpha v(n)\big]\big[1 - I(n,t)\big] \Bigg\}, \qquad (12)\end{aligned}$$

with boundary condition $V(n,T) = v(n)$ for all $n \in [0, N_0]$.

It is easily shown that

$$\begin{aligned}&\sup_{\mathscr{R}_1(n,t)} \lambda \int_{\mathscr{R}_1(n,t)} \{r + p(r,s) - [V(n,t) - V(n-s,t)]\} \\ &\quad \cdot F_{R,S}(dr,ds) \\ &\quad = \lambda \int_{\{(r,s)\in\Re\times[0,n]:\, r+p(r,s)\geqslant V(n,t)-V(n-s,t)\}} \\ &\quad \cdot \big\{r + p(r,s) - [V(n,t) - V(n-s,t)]\big\} F_{R,S}(dr,ds).\end{aligned}$$

Therefore, consider the differential equation

$$\begin{aligned}\frac{\partial V}{\partial t}(n,t) &= \min \Bigg\{ -\lambda \int_{\{(r,s)\in\Re\times[0,n]:\, r+p(r,s)\geqslant V(n,t)-V(n-s,t)\}} \\ &\qquad \cdot \{r + p(r,s) - [V(n,t) - V(n-s,t)]\} \\ &\qquad \cdot F_{R,S}(dr,ds) + \alpha V(n,t) + \lambda \bar{p} + c(n), \\ &\qquad \alpha V(n,t) - \alpha v(n) \Bigg\} \\ &\equiv \min \big\{ \lambda \Upsilon(n,t,V) + \alpha V(n,t) + \lambda \bar{p} + c(n), \\ &\qquad \alpha V(n,t) - \alpha v(n) \big\} \\ &\equiv f(n,t,V),\end{aligned}$$

where $\Upsilon : [0, N_0] \times [0, T] \times \mathscr{V} \mapsto \Re$ and $f : [0, N_0] \times [0, T] \times \mathscr{V} \mapsto \Re$ are defined above, and $\mathscr{V}$ denotes the set of bounded Borel-measurable functions $V : [0, N_0] \times [0, T] \mapsto \Re$. It is shown in Kleywegt (1996) that (12) has a unique solution $\widehat{V}^*$. Theorem 4 then follows from the results in Boel and Varaiya (1977).

THEOREM 4. *The optimal expected value function V^* for the MDP is the unique absolutely continuous solution of the differential equation*

$$\frac{\partial V^*}{\partial t}(n,t)$$

$$= \min\left\{ -\lambda \int_{\mathcal{R}_1^*(n,t)} \{r + p(r,s) - [V^*(n,t) - V^*(n-s,t)]\} \cdot F_{R,S}(dr, ds) + \alpha V^*(n,t) + \lambda \bar{p} + c(n), \quad \alpha V^*(n,t) - \alpha v(n) \right\}, \tag{13}$$

with boundary condition $V^*(n, T) = v(n)$ *for all* $n \in [0, N_0]$, *where*

$$\mathcal{R}_1^*(n,t) \equiv \{(r,s) \in \Re \times [0,n] : r + p(r,s) \geqslant V^*(n,t) - V^*(n-s,t)\}.$$

The following memoryless deterministic policy $\pi^* \in \prod_{\text{MDP}}^{\text{MD}}$ *is optimal for the MDP among all history-dependent deterministic policies* $\pi \in \prod_{\text{MDP}}^{\text{MD}}$. *An optimal acceptance rule is*

$$D^*(n,t,r,s) = \begin{cases} 1 & \text{if } s \leqslant n \text{ and } r + p(r,s) \geqslant V^*(n,t) - V^*(n-s,t), \\ 0 & \text{if } s > n \text{ or } r + p(r,s) < V^*(n,t) - V^*(n-s,t). \end{cases}$$

That is, $\mathcal{R}_1^*(n,t)$ *is an optimal acceptance set. An optimal stopping rule is*

$$I^*(n,t) = \begin{cases} 1 & \text{if } \lambda \int_{\mathcal{R}_1^*(n,t)} \{r + p(r,s) - [V^*(n,t) - V^*(n-s,t)]\} F_{R,S}(dr, ds) - \lambda \bar{p} - c(n) \geqslant \alpha v(n), \\ 0 & \text{if } \lambda \int_{\mathcal{R}_1^*(n,t)} \{r + p(r,s) - [V^*(n,t) - V^*(n-s,t)]\} F_{R,S}(dr, ds) - \lambda \bar{p} - c(n) < \alpha v(n). \end{cases}$$

4.3. The Optimal Value and an Optimal Policy for the DSKP

An optimal policy $\pi^* \in \Pi_{\text{MDP}}^{\text{MD}}$ for the MDP is not necessarily admissible for the DSKP because under policy π^* the process may be switched *on* and *off* multiple times. In this section we show that policy π^* given in Theorem 4 switches *off* only once and then remains switched *off*, i.e., $I^*(n, \cdot) \in \mathcal{I}_s$ for all $n \in [0, N_0]$ (Lemma 2), and therefore policy π^* is admissible and optimal for the DSKP among all policies $\pi \in \Pi_{\text{DSKP}}^{\text{HD}}$ (Theorem 5).

Lemma 2. *For each* $n \in [0, N_0]$, $I^*(n, \cdot)$, *is a unit step function of the form*

$$I^*(n,t) = \begin{cases} 1 & \text{if } t \in (0, \tau^*(n)], \\ 0 & \text{if } t \in (\tau^*(n), T], \end{cases}$$

where $\tau^*(n) \in [0, T]$.

Theorem 5. *The memoryless deterministic threshold policy* π^* *of Theorem 4 is an optimal policy among all history-dependent deterministic policies for the DSKP.*

Similar to the infinite horizon case, we can consider another MDP where the reward of demand i is given by $R_i + p(R_i, S_i)$, the holding cost is $\lambda \bar{p} + c(n)$ when the remaining amount of resource is n, and there is no penalty for rejecting a demand. From Theorem 4 it is seen that the optimal value function V_2^* and optimal policy π_2^* for the new MDP are the same as the optimal value function V_1^* and optimal policy π_1^* for the original MDP, and from Theorem 5 it follows that these are the same as the optimal value function V^* and optimal policy π^* for the original DSKP. Thus for the purpose of finding the optimal value function V^* and optimal policy π^* for the original MDP and DSKP, it is sufficient to consider an MDP with no penalties, with the understanding that the new reward R_i represents $R_i + p(R_i, S_i)$, the distributions $F_{R,S}$, F_R, $F_{R|S}$, and $F_{S|R}$ are for the new reward R, and the holding cost $c(n)$ represents $\lambda \bar{p} + c(n)$. In the remainder of §4, the new MDP is considered.

Corollary 3 establishes the intuitive result similar to the infinite horizon case, that the policy with reward threshold $x^*(n, t, s) = V^*(n, t) - V^*(n-s, t)$ is optimal.

Corollary 3. *The optimal expected value function* V^* *is the unique absolutely continuous solution of the differential equation*

$$\frac{\partial V^*}{\partial t}(n,t) = \min\left\{ -\lambda \int_0^n \int_{x^*(n,t,s)}^{\infty} \cdot \{r - [V^*(n,t) - V^*(n-s,t)]\} \cdot F_{R|S}(dr|s) F_S(ds) + \alpha V^*(n,t) + c(n), \quad \alpha V^*(n,t) - \alpha v(n) \right\},$$

where optimal reward threshold $x^*(n,t,s) = V^*(n,t) - V^*(n-s,t)$, *and boundary condition* $V^*(n,T) = v(n)$ *for all* $n \in [0, N_0]$. *The following memoryless deterministic threshold policy* π^* *is an optimal policy for the MDP and DSKP among all history-dependent deterministic policies. An optimal acceptance rule is*

$$D^*(n,t,r,s) = \begin{cases} 1 & \text{if } s \leqslant n \text{ and } r \geqslant x^*(n,t,s), \\ 0 & \text{if } s > n \text{ or } r < x^*(n,t,s). \end{cases}$$

An optimal stopping rule is

$$I^*(n,t) = \begin{cases} 1 & \text{if } \lambda \int_0^n \int_{x^*(n,t,s)}^{\infty} \{r - [V^*(n,t) - V^*(n-s,t)]\} \cdot F_{R|S}(dr|s) F_S(ds) - c(n) \geqslant \alpha v(n), \\ 0 & \text{if } \lambda \int_0^n \int_{x^*(n,t,s)}^{\infty} \{r - [V^*(n,t) - V^*(n-s,t)]\} \cdot F_{R|S}(dr|s) F_S(ds) - c(n) < \alpha v(n). \end{cases}$$

Similar to the infinite horizon case, consider the size threshold

$$z^*(n,t,r) \equiv \sup\{s \in [0,n] : r \geqslant V^*(n,t) - V^*(n-s,t)\}.$$

Corollary 4 establishes the intuitive result that if $V^*(n, t)$ is nondecreasing in n, then there exists an optimal acceptance rule with size threshold $z^*(n, t, r)$.

COROLLARY 4. *Suppose $V^*(n,t)$ is nondecreasing in n. Then the optimal expected value function V^* is the unique absolutely continuous solution of the differential equation*

$$\frac{\partial V^*}{\partial t}(n,t) = \min\Big\{-\lambda\int_{-\infty}^{\infty}\int_0^{z^*(n,t,r)} \cdot\{r-[V^*(n,t)-V^*(n-s,t)]\} \cdot F_{S|R}(ds|r)F_R(dr)+\alpha V^*(n,t) + c(n),\quad \alpha V^*(n,t)-\alpha v(n)\Big\}, \tag{14}$$

with boundary condition $V^*(n,T)=v(n)$ for all $n\in[0,N_0]$. The following memoryless deterministic threshold policy π^* is an optimal policy for the MDP and DSKP among all history-dependent determinstic policies. An optimal acceptance rule is

$$D^*(n,t,r,s) = \begin{cases} 1 & \text{if } s < z^*(n,t,r),\\ 0 & \text{if } s > z^*(n,t,r).\end{cases}$$

If $s = z^*(n,t,r)$, then $D^*(n,t,r,s)=1$ if $r \geqslant V^*(n,t) - V^*(n-z^*(n,t,r),t)$, and $D^*(n,t,r,s)=0$ otherwise. An optimal stopping rule is

$$I^*(n,t) = \begin{cases} 1 & \text{if } \lambda\int_{-\infty}^{\infty}\int_0^{z^*(n,t,r)}\{r-[V^*(n,t)-V^*(n-s,t)]\} \cdot F_{S|R}(ds|r)F_R(dr)-c(n)\geqslant \alpha v(n),\\ 0 & \text{if } \lambda\int_{-\infty}^{\infty}\int_0^{z^*(n,t,r)}\{r-[V^*(n,t)-V^*(n-s,t)]\} \cdot F_{S|R}(ds|r)F_R(dr)-c(n)< \alpha v(n).\end{cases}$$

Proposition 2 established that if $c(n)$ is nonincreasing and $v(n)$ is nondecreasing, then $V^*(n,t)$ is nondecreasing in n. Similar to the infinite horizon case, if $z^*(n,t,r) > -\infty$, and $V^*(n,t)$ is upper semicontinuous in n, then $r \geqslant V^*(n,t) - V^*(n-z^*(n,t,r),t)$, and $D^*(n,t,r,z^*(n,t,r))=1$ is an optimal decision. In turn, $V^*(n,t)$ is upper semicontinuous in n if $c(n)$ is lower semicontinuous and $v(n)$ is upper semicontinuous. Also, because $V^*(n,t)$ is nondecreasing in n, $V^*(n,t)$ being upper semicontinuous in n is equivalent to $V^*(n,t)$ being right-continuous in n. However, the converse does not necessarily hold; namely if $V^*(n,t)$ is lower semicontinuous in n, then it does not necessarily hold that $r \leqslant V^*(n,t) - V^*(n-z^*(n,t,r),t)$.

4.4. Convergence of Discrete-Time Approximations

To solve the DSKP numerically, we need to compute a solution to the optimality Equation (13). Usually a numerical method is used to solve such a differential equation. The existence proof of the optimal solution V^* suggests a convergent scheme, but usually this method is not efficient. Methods based on discretization tend to give better numerical performance. A desirable property for a numerical approximation method is that it should converge to the optimal solution as the discretization becomes finer. We show in Theorem 6 that the discretization method described in this section converges uniformly to the optimal value function V^*. This convergence result is also used to establish montonicity and convexity results in subsection 4.5.3.

Construct a discrete-time approximation to the differential equation (13) as follows. For $m=1,2,\ldots$, let $\Delta t = T/m$, and let $\{t_i = iT/m\}_{i=0}^m$ be a partition of $[0,T]$. Let V_m denote the solution of the following difference equation:

$$\begin{aligned} V_m(n,t_{i-1}) &= V_m(n,t_i) + \left(\frac{T}{m}\right) \cdot \max\Big\{\lambda\int_0^n\int_{x_m(n,t_i,s)}^{\infty} \cdot\{r-[V_m(n,t_i)-V_m(n-s,t_i)]\} \cdot F_{R|S}(dr|s)F_S(ds)-\alpha V_m(n,t_i) - c(n),\quad -\alpha V_m(n,t_i)+\alpha v(n)\Big\} \\ &\equiv V_m(n,t_i) - (\Delta t) f(n,t_i,V_m), \end{aligned} \tag{15}$$

where $x_m(n,t_i,s) \equiv V_m(n,t_i) - V_m(n-s,t_i)$. The boundary conditions are $V_m(n,t_m) = V_m(n,T) = v(n)$ for all $n\in[0,N_0]$, for $m=1,2,\ldots$. Complete $\{V_m(n,t_i)\}_{i=0}^m$ on $[0,T]$ for every $n\in[0,N_0]$ by linear interpolation, i.e., for any $t\in[t_{i-1},t_i]$, let

$$V_m(n,t) \equiv \frac{t_i - t}{t_i - t_{i-1}} V_m(n,t_{i-1}) + \frac{t-t_{i-1}}{t_i-t_{i-1}} V_m(n,t_i).$$

THEOREM 6. *The sequence of solutions $\{V_m\}_{m=1}^{\infty}$ of the sequence of difference equations* (15) *converges uniformly to the optimal value function V^*.*

4.5. Structural Characteristics

Conditions under which intuitive structural characteristics of the optimal value function V^* and optimal threshold x^* hold are established in this section. Subsection 4.5.1 discusses the fundamental roles played by the rate at which rewards are earned, the rate at which cost is incurred, and their relative magnitudes. A recursive form for the optimal expected value and an optimal policy is given in subsection 4.5.2 that holds under typical conditions. This characterization is useful because it gives a simple, intuitive recipe for following an optimal policy, and it simplifies computation of V^* and x^*. Thereafter some intuitive monotonicity and convexity properties are shown to hold, under certain conditions. Also interesting are the cases where counterintuitive properties are exhibited, some of which are illustrated by example.

4.5.1. A Basic Relationship. The properties of V^* depend to a large extent on the relative magnitudes of $\lambda\Upsilon(n,0,0)$ and $c(n)+\alpha v(n)$. An example of this is given in Proposition 3. The importance of these quantities makes intuitive sense by noting that $\lambda\Upsilon(n,0,0) = \lambda P[\{S\leqslant n\}\cap\{R\geqslant 0\}]E[R|\{S\leqslant n\}\cap\{R\geqslant 0\}]$ and by interpreting

$\lambda P[\{S \leqslant n\} \cap \{R \geqslant 0\}] E[R | \{S \leqslant n\} \cap \{R \geqslant 0\}]$ as the effective reward rate while one continues to wait, and comparing it with $c(n) + \alpha v(n)$, the rate at which (opportunity) cost is incurred while one continues to wait.

PROPOSITION 3. *If $c(\cdot)$ is nonincreasing, and $v(\cdot)$ is nondecreasing, then for each $n \in [0, N_0]$ such that $\lambda \Upsilon(n, 0, 0) \leqslant c(n) + \alpha v(n)$, it holds that $V^*(n, t) = v(n)$ for all $t \in [0, T]$.*

The conditions that $c(n)$ be nonincreasing and $v(n)$ be nondecreasing are the conditions that one would normally expect to hold in applications, namely that the holding cost increases as the amount of resource allocated increases and that the terminal value increases as the final remaining amount of resource increases. It is shown in Kleywegt (1996) why these conditions are needed.

4.5.2. A Recursion for the Optimal Expected Value. Similar to the infinite horizon case, a recursive form for the optimal expected value function V^* is derived. The conditions under which $V^*(n, t) = \max\{v(n), V^{\psi(n)}(n, t)\}$ for the finite horizon case are more restrictive than those for the infinite horizon case. It is shown in Kleywegt (1996) why these more restrictive conditions are needed.

Again assume that the size distribution F_S is such that $F_S(0) = 0$, and the possible values of the remaining capacity are $0 = n_0 < n_1 < n_2 < \cdots < N_0$. Without loss of generality, assume that $n_k = k$ for all k.

Inductively define the sequence of threshold policies $\{\psi(n)\}_{n=0}^{N_0}$ as follows. Let $\mathcal{R}_1^{\psi(0)}(0, t) \equiv \varnothing$, and $I^{\psi(0)}(0, t) \equiv 1$ for all $t \in [0, T]$. Then

$$V^{\psi(0)}(0, t) = e^{-\alpha(T-t)} v(0) - \frac{c(0)}{\alpha}(1 - e^{-\alpha(T-t)}),$$

if $\alpha > 0$, and $V^{\psi(0)}(0, t) = -c(0)(T-t) + v(0)$ if $\alpha = 0$. Let $\psi(n-s)$ and $V^{\psi(n-s)}(n', \cdot)$ be defined for all $s \in \{1, \ldots, n\}$ and all $n' \in \{0, 1, \ldots, n-s\}$. Let $\widehat{V}(n, \cdot)$ satisfy

$$\begin{aligned}\frac{\partial \widehat{V}}{\partial t}(n, t) = -\lambda \int_0^n \int_{\hat{x}(n,t,s)}^{\infty} \Big\{ r - \big[\widehat{V}(n, t) - \max\{v(n-s),& \\ V^{\psi(n-s)}(n-s, t)\}\big]\Big\} F_{R|S}(dr|s) F_s(ds)& \\ + \alpha \widehat{V}(n, t) + c(n)& \qquad (16)\end{aligned}$$

for $t \in (0, T)$, where $\hat{x}(n, t, s) \equiv \widehat{V}(n, t) - \max\{v(n-s), V^{\psi(n-s)}(n-s, t)\}$. The boundary condition is $\widehat{V}(n, T) = v(n)$. The right-hand side of (16) satisfies a Lipschitz condition with respect to $\widehat{V}$ with constant $\lambda + \alpha$. It follows that (16) has a unique solution $\widehat{V}(n, \cdot)$. Let $I^{\psi(n)}(n, t) \equiv 1$, and let $x^{\psi(n)}(n, t, s) \equiv \hat{x}(n, t, s)$ for all $s \leqslant n$ and all $t \in [0, T]$. Let

$$I^{\psi(n)}(n-s, t) \equiv \begin{cases} 1 & \text{if } V^{\psi(n-s)}(n-s, t) > v(n-s), \\ 0 & \text{if } V^{\psi(n-s)}(n-s, t) \leqslant v(n-s), \end{cases}$$

and $x^{\psi(n)}(n-s, t, s') \equiv x^{\psi(n-s)}(n-s, t, s')$ for all $s \leqslant n$, $s' \leqslant n-s$, and $t \in [0, T]$. From the definition of $\{\psi(n)\}$, for all $s \leqslant n$, $s' \leqslant n-s$, and $t \in [0, T]$, $\mathcal{R}_1^{\psi(n)}(n-s-s', t) = \mathcal{R}_1^{\psi(n-s-s')}(n-s-s', t) = \mathcal{R}_1^{\psi(n-s)}(n-s-s', t)$, and $I^{\psi(n)}(n-s-s', t) = I^{\psi(n-s)}(n-s-s', t)$. Hence $V^{\psi(n)}(n-s-s', t) = V^{\psi(n-s)}(n-s-s', t)$ for all $t \in [0, T]$. Then it follows from the definition of $\{\psi(n)\}$, Theorem 3, and (11) that $V^{\psi(n)}(n, t)$ is the unique solution of

$$\begin{aligned}&\frac{\partial V^{\psi(n)}}{\partial t}(n, t) \\ &\quad = -\lambda \int_0^n \int_{x^{\psi(n)}(n,t,s)}^{\infty} \big\{ r - \big[V^{\psi(n)}(n, t) - V^{\psi(n)}(n-s, t) \big] \big\} \\ &\qquad \cdot F_{R|S}(dr|s) F_S(ds) + \alpha V^{\psi(n)}(n, t) + c(n), \qquad (17)\end{aligned}$$

with boundary condition $V^{\psi(n)}(n, T) = v(n)$, and for any $s \leqslant n$, $V^{\psi(n)}(n-s, t)$ is the unique solution of

$$\begin{aligned}&\frac{\partial V^{\psi(n)}}{\partial t}(n-s, t) \\ &\quad = \Big[-\lambda \int_0^{n-s} \int_{x^{\psi(n)}(n-s,t,s')}^{\infty} \big\{ r - \big[V^{\psi(n)}(n-s, t) \\ &\qquad - V^{\psi(n)}(n-s-s', t) \big] \big\} F_{R|S}(dr|s') F_S(ds') \\ &\qquad + \alpha V^{\psi(n)}(n-s, t) + c(n-s) \Big] I^{\psi(n)}(n-s, t) \\ &\qquad + \big[\alpha V^{\psi(n)}(n-s, t) - \alpha v(n-s) \big] \big[1 - I^{\psi(n)}(n-s, t) \big] \\ &\qquad\qquad (18)\end{aligned}$$

with boundary condition $V^{\psi(n)}(n-s, T) = v(n-s)$.

LEMMA 3. *For all $n > 0$, all $t \in [0, T]$, and all $s \leqslant n$,*

$$V^{\psi(n)}(n-s, t) = \max\big\{ v(n-s), V^{\psi(n-s)}(n-s, t) \big\}.$$

It follows from (17) and Lemma 3 that $V^{\psi(n)}(n, \cdot)$ is the unique solution of

$$\begin{aligned}&\frac{\partial V^{\psi(n)}}{\partial t}(n, t) \\ &\quad = -\lambda \int_0^n \int_{\hat{x}(n,t,s)}^{\infty} \big\{ r - \big[V^{\psi(n)}(n, t) - \max\{v(n-s), \\ &\qquad\qquad V^{\psi(n-s)}(n-s, t)\} \big] \big\} F_{R|S}(dr|s) F_S(ds) \\ &\qquad + \alpha V^{\psi(n)}(n, t) + c(n) \qquad (19)\end{aligned}$$

with boundary condition $V^{\psi(n)}(n, T) = v(n)$. From (16), (19) has a unique solution $V^{\psi(n)}(n, \cdot) = \widehat{V}(n, \cdot)$. Therefore, $x^{\psi(n)}(n, t, s) = V^{\psi(n)}(n, t) - V^{\psi(n)}(n-s, t)$, and

$$\begin{aligned}\frac{\partial V^{\psi(n)}}{\partial t}(n, t) &= -\lambda \int_0^n \int_{x^{\psi(n)}(n,t,s)}^{\infty} \big\{ r - \big[V^{\psi(n)}(n, t) \\ &\qquad - V^{\psi(n)}(n-s, t) \big] \big\} F_{R|S}(dr|s) F_S(ds) \\ &\qquad + \alpha V^{\psi(n)}(n, t) + c(n) \\ &= -\lambda \Upsilon(n, t, V^{\psi(n)}) + \alpha V^{\psi(n)}(n, t) + c(n). \qquad (20)\end{aligned}$$

PROPOSITION 4. *If $v(\cdot)$ is nonincreasing, then for each $n \in [0, N_0]$ such that $\lambda \Upsilon(n, 0, 0) > c(n) + \alpha v(n)$, it holds that $V^{\psi(n)}(n, t) > v(n)$ for all $t \in [0, T)$.*

It is shown in Kleywegt (1996) why the condition that $v(n)$ be nonincreasing is needed.

By definition, $V^*(n,t) \geqslant \max\{v(n), V^{\psi(n)}(n,t)\}$ for all n and t. However, if $c(n)$ is nonincreasing and v is constant, then similar to the infinite horizon case, $V^*(n,t) = \max\{v, V^{\psi(n)}(n,t)\}$ for all n and t. This is the result of Theorem 7.

Theorem 7. *If $c(n)$ is nonincreasing and v is constant, then $V^*(n,t) = \max\{v, V^{\psi(n)}(n,t)\}$ for all n and t.*

It is shown in Kleywegt (1996) why the stated conditions are needed. The basic idea is as follows. It may be optimal to stop between the times that demands are accepted. However, policy $\psi(n)$ continues to wait while the remaining amount of resource is n. Thus the value $\max\{v(n), V^{\psi(n)}(n,t)\}$ is equal to the optimal value $V^*(n,t)$ if it is optimal to consider as candidate stopping times only the times when the remaining amount of resource changes. The conditions are imposed to ensure that it is sufficient to consider only these candidate stopping times.

Under the conditions of Theorem 7, an optimal policy π^* has the following convenient form. If $c(0)+\alpha v < 0$, then $V^*(n,t) = V^{\psi(n)}(n,t)$ and $I^*(n,t) = 1$ for all n and t. Else, if $c(0)+\alpha v \geqslant 0$, then let $n^* \equiv \max\{n \in \{0,1,\ldots,N_0\} : \lambda\Upsilon(n,0,0) \leqslant c(n)+\alpha v\}$. Then for all $n \leqslant n^*$, $V^*(n,t) = v$ and $I^*(n,t) = 0$ for all $t \in [0,T]$. Also, for all $n > n^*$, $V^*(n,t) = V^{\psi(n)}(n,t)$ and $I^*(n,t) = 1$ for all $t \in [0,T]$. Hence, as long as $t < T$ and $n > n^*$, it is optimal to continue, using reward threshold $x^*(n,t,s) = x^{\psi(n)}(n,t,s) = V^{\psi(n)}(n,t) - \max\{v, V^{\psi(n-s)}(n-s,t)\} = V^*(n,t) - V^*(n-s,t)$. It is optimal to stop and collect terminal value v as soon as n becomes less than or equal to n^*. It is interesting that n^* does not depend on t, i.e., it is optimal to continue if $n > n^*$ for all $t < T$, and it is optimal to stop if $n \leqslant n^*$ for all t. In other words, the optimal stopping decision does not depend on closeness to the deadline, and the only candidate stopping times that have to be considered are the initial time $t = 0$ and the times when the remaining amount of resource changes, that is, the times when demands are accepted. This result characterizes an optimal policy and the optimal value function in a simple, intuitive way. It also leads to an easy method for computing the optimal value function V^* and optimal policy π^*, by simply computing $V^{\psi(n)}(n,t)$ for $n = 0,1,\ldots,N_0$.

4.5.3. Monotonicity and Convexity Properties. In this section it is established which monotonicity and convexity properties hold if demands require random amounts of resource and which conditions are needed for these properties to hold.

It was shown in Proposition 1 that $V^*(n,t)$ is nonincreasing in t. In Proposition 2 it was shown that if $c(n)$ is nonincreasing and $v(n)$ is nondecreasing, then $V^*(n,t)$ is nondecreasing in n. It is intuitive to expect that an optimal policy becomes more lenient in accepting demands as the deadline approaches; in other words, the optimal reward threshold is decreasing in time for any given remaining amount of resource and resource requirement, or the optimal size threshold is increasing in time for any given remaining amount of resource and reward. It is also intuitive to expect that an optimal policy becomes more selective in accepting demands as the remaining amount of resource decreases; in other words, the optimal reward threshold is decreasing in the remaining amount of resource for any given time and resource requirement, or the optimal size threshold is increasing in the remaining amount of resource for any given time and reward. It also appeals to the intuition for the optimal expected value function to be concave in time and concave in the remaining amount of resource, that is, for an increasing amount of resource or time until the deadline to have decreasing marginal value.

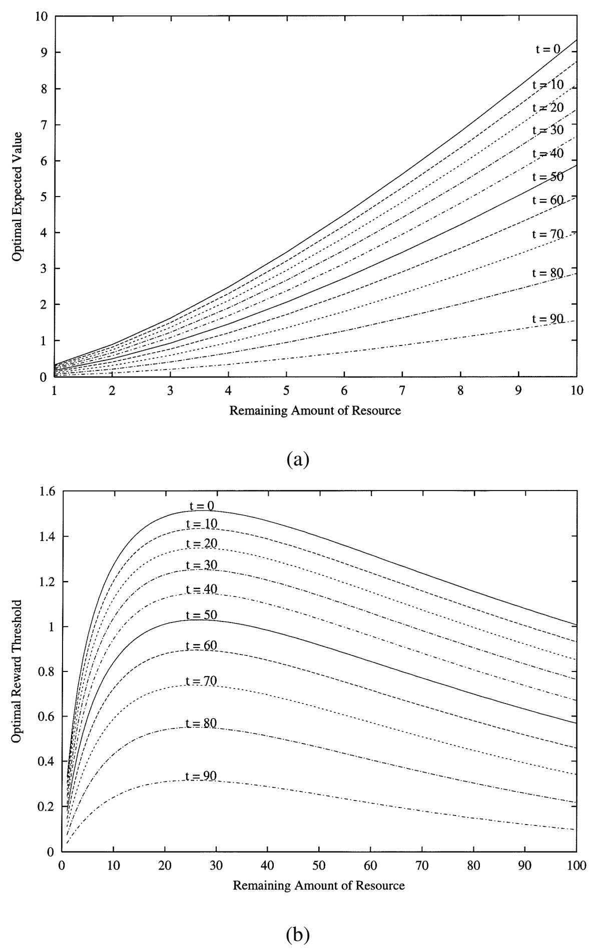
It was shown in Kleywegt and Papastavrou (1998) that these structural properties hold if all demands require the same amount of resource. This result is stated in Theorem 8.

Theorem 8. *If all demands require the same amount of resource, c is constant and $v(n)$ is concave nonincreasing, $\alpha = 0$, and $\lambda\Upsilon(1,0,0) \geqslant \lambda\bar{p} + c$, then the following conditions hold.*

- *(i) $V^*(n+1,t) - V^*(n,t) \leqslant V^*(n,t) - V^*(n-1,t)$ for all $n \in \{1,\ldots,N_0-1\}$ and all $t \in [0,T]$ (the optimal value function is concave in the remaining amount of resource).*
- *(ii) $x^*(n+1,t) \leqslant x^*(n,t)$ for all $n \in \{1,\ldots,N_0-1\}$ and all $t \in [0,T]$ (the optimal reward threshold is nonincreasing in the remaining amount of resource).*
- *(iii) $\partial V^*(n,t)/\partial t \leqslant \partial V^*(n-1,t)/\partial t$ for all $n \in \{1,\ldots,N_0\}$ and all $t \in (0,T)$ (the marginal optimal expected value of remaining time $-\partial V^*(n,t)/\partial t$ is nondecreasing in the remaining amount of resource).*
- *(iv) $\partial x^*(n,t)/\partial t \leqslant 0$ for all $n \in \{1,\ldots,N_0\}$, and all $t \in (0,T)$ (the optimal reward threshold is nonincreasing in time).*
- *(v) $\partial V^*(n,t_2)/\partial t \leqslant \partial V^*(n,t_1)/\partial t$ for all $n \in \{0,\ldots,N_0\}$ and all $0 < t_1 \leqslant t_2 < T$ ($\partial V^*(n,t)/\partial t$ is nonincreasing in time, or the optimal value function is concave in time).*

These properties do not hold in the same generality when demands require random amounts of resource. Figure 1 shows $V^*(n,t)$ and $x^*(n,t,1)$ as functions of n for different values of t, for an example with exponentially distributed sizes with mean 25, conditionally exponentially distributed rewards with mean equal to the size, the deadline $T = 100$, items arrive at rate $\lambda = 0.1$, the penalty $p = 0$, the holding cost per unit time $c = 0$, the terminal value $v = 0$, and the discount rate $\alpha = 0$. Note that $V^*(n,t)$ is not concave in n, and $x^*(n,t,1)$ is increasing in n for small values of n, although the conditions of Theorem 8 are satisfied. The packing of items with different sizes into the knapsack destroys some of the intuitive characteristics of the optimal value function and optimal policy. Therefore, a

Figure 1. Example to show that if items have random sizes, then an extension of Theorem 8 does not hold in general. (a) Optimal expected value $V^*(n, t)$ versus remaining amount n of resource for different points in time t. (b) Optimal reward threshold $x^*(n, t, 1)$ versus remaining amount n of resource for different points in time t.



straightforward extension of Theorem 8 does not hold for the case with random resource requirements.

In the remainder of this section we investigate which of the structural properties continue to hold in general if demands require random amounts of resource and what conditions are needed for the other structural properties to hold.

Lemma 4. *$V^*(n, t)$ is concave in n if and only if $x^*(n, t, s)$ is nonincreasing in n for each $s \in [0, n]$.*

That is, an optimal policy becomes more selective in accepting demands as the remaining amount of resource decreases.

Lemma 5. *If $c(n)$ is nonincreasing, $\alpha = 0$, and $x^*(n, t, s)$ is nonincreasing in n for each $s \in [0, n]$, then $\partial V^*(n, t)/\partial t$ is nonincreasing in n.*

That is, an additional unit of remaining time until the deadline is worth more if there is more resource remaining than if there is less resource remaining.

Lemma 6. *$\partial V^*(n, t)/\partial t$ is nonincreasing in n if and only if $x^*(n, t, s)$ is nonincreasing in t for each $s \in [0, n]$.*

That is, an optimal policy becomes more lenient in accepting demands as the deadline approaches.

Lemma 7. *If it is optimal to continue (i.e., $I^*(n, t) = 1$) for all $t \in (0, T)$, and $x^*(n, t, s)$ is nonincreasing in t for each $s \in [0, n]$, then $V^*(n, t)$ is concave in t on $[0, T]$.*

That is, an additional unit of remaining time until the deadline is worth more closer to the deadline than farther from the deadline.

Lemma 8. *If $V^*(n, t)$ is concave in n, then $z^*(n, t, r)$ is nondecreasing in n for all $r \in \Re$.*

This is another manifestation of the property that an optimal policy becomes more selective in accepting demands as the remaining amount of resource decreases.

Lemma 9. *If $c(n)$ is nonincreasing, $\alpha = 0$, and $V^*(n, t)$ is concave in n, then $z^*(n, t, r)$ is nondecreasing in t for all $n \in [0, N_0]$ and all $r \in \Re$.*

This is another manifestation of the property that an optimal policy becomes more lenient in accepting demands as the deadline approaches.

From Lemmas 4 through 9, it follows that the intuitive monotonicity and convexity properties hold for the case with random resource requirements if $V^*(n, t)$ is concave in n. Next it is established that $V^*(n, t)$ is concave in n under certain conditions. To do that, first consider a discrete-time approximation of (14), constructed similarly to the approximation in §4.4 as follows. For $m = 1, 2, \ldots$, let $\Delta t = T/m$, and let $\{t_i = iT/m\}_{i=0}^{m}$ be a partition of $[0, T]$. Let V_m be the solution of the following difference equation:

$$\begin{aligned}
&V_m(n, t_{i-1}) \\
&\quad = V_m(n, t_i) - (\Delta t) f(n, t_i, V_m) \\
&\quad = \max\Bigg\{ \lambda \Delta t \int_{-\infty}^{\infty} \Bigg[\int_0^{z_m(n, t_i, r)} [r + V_m(n-s, t_i)] F_{S|R}(ds|r) \\
&\qquad\qquad + V_m(n, t_i) \int_{z_m(n, t_i, r)}^{\infty} F_{S|R}(ds|r) \Bigg] F_R(dr) \\
&\qquad\qquad + [1 - (\alpha + \lambda)\Delta t] V_m(n, t_i) \\
&\qquad\qquad - c(n)\Delta t, \quad (1 - \alpha \Delta t) V_m(n, t_i) + \alpha v(n) \Delta t \Bigg\}. \quad (21)
\end{aligned}$$

The boundary conditions are $V_m(n, t_m) = V_m(n, T) = v(n)$ for all $n \in [0, N_0]$, for all $m > (\alpha + \lambda)T$. The size threshold is $z_m(n, t_i, r) \equiv \sup\{s \in [0, n] : r \geqslant V_m(n, t_i) - V_m(n - s, t_i)\}$. Complete $\{V_m(n, t_i)\}_{i=0}^{m}$ on $[0, T]$ for every $n \in$

$[0, N_0]$ by linear interpolation, i.e., for any $t \in [t_{i-1}, t_i]$, let

$$V_m(n,t) \equiv \frac{t_i - t}{t_i - t_{i-1}} V_m(n, t_{i-1}) + \frac{t - t_{i-1}}{t_i - t_{i-1}} V_m(n, t_i).$$

Lemma 10 gives the conditions under which the intuitive properties in Theorem 8 for the case where all demands require the same amount of resource extend to the case with random resource requirements. It is shown in Kleywegt (1996) why the additional conditions in Lemma 10 are needed.

LEMMA 10. *If $c = 0$, $v(n)$ is concave nondecreasing, $\alpha = 0$, and $F_{S|R}(s|r)$ is concave in s on $[0, N_0]$ for F_R-almost all $r \geqslant 0$, then $V_m(n,t)$ is concave in n on $[0, N_0]$ for all $t \in [0, T]$ and all $m > \lambda T$.*

LEMMA 11. *If $c = 0$, $v(n)$ is concave nondecreasing, $\alpha = 0$, and $F_{S|R}(s|r)$ is concave in s on $[0, N_0]$ for F_R-almost all $r \geqslant 0$, then $V^*(n,t)$ is concave in n for all $t \in [0, T]$.*

PROOF. From Lemma 10, $V_m(n,t)$ is concave in n on $[0, N_0]$ for all $t \in [0, T]$, for all $m > \lambda T$. From Corollary 4, under the stated conditions, difference equations (15) and (21) have the same solution. From Theorem 6, the sequence of discrete-time approximations $\{V_m(n,t)\}_{m=1}^{\infty}$ converges uniformly to the continuous-time optimal value function $V^*(n,t)$. Therefore, it follows from Rockafellar (1970, Theorem 10.8) that $V^*(n,t)$ is concave in n on $[0, N_0]$ for all $t \in [0, T]$. □

THEOREM 9. *If $c = 0$, $v(n)$ is concave nondecreasing, $\alpha = 0$, and $F_{S|R}(s|r)$ is concave in s on $[0, N_0]$ for F_R-almost all $r \geqslant 0$, then the following conditions hold.*

(i) *$V^*(n_1, t) - V^*(n_1 - s, t) \geqslant V^*(n_2, t) - V^*(n_2 - s, t)$ for all $0 \leqslant n_1 \leqslant n_2 \leqslant N_0$, all $s \in [0, n_1]$, and all $t \in [0, T]$ (the optimal value function is concave in the remaining amount of resource).*

(ii) *$x^*(n_1, t, s) \geqslant x^*(n_2, t, s)$ for all $0 \leqslant n_1 \leqslant n_2 \leqslant N_0$, all $s \in [0, n_1]$, and all $t \in [0, T]$ (the optimal reward threshold is nonincreasing in the remaining amount of resource).*

(iii) *$\partial V^*(n_1, t)/\partial t \geqslant \partial V^*(n_2, t)/\partial t$ for all $0 \leqslant n_1 \leqslant n_2 \leqslant N_0$ and all $t \in (0, T)$ (the marginal optimal expected value of remaining time $-\partial V^*(n,t)/\partial t$ is nondecreasing in the remaining amount of resource).*

(iv) *$\partial x^*(n, t, s)/\partial t \leqslant 0$ for all $n \in [0, N_0]$, all $s \in [0, n]$, and all $t \in (0, T)$ (the optimal reward threshold is nonincreasing in time).*

(v) *$\partial V^*(n, t_1)/\partial t \geqslant \partial V^*(n, t_2)/\partial t$ for all $n \in [0, N_0]$ and all $0 < t_1 \leqslant t_2 < T$ ($\partial V^*(n,t)/\partial t$ is nonincreasing in time, or the optimal value function is concave in time).*

(vi) *$z^*(n_1, t, r) \leqslant z^*(n_2, t, r)$ for all $0 \leqslant n_1 \leqslant n_2 \leqslant N_0$, all $t \in [0, T]$, and all $r \in \Re$ (the optimal size threshold is nondecreasing in the remaining amount of resource).*

(vii) *$z^*(n, t_1, r) \leqslant z^*(n, t_2, r)$ for all $n \in [0, N_0]$, all $0 \leqslant t_1 \leqslant t_2 \leqslant T$, and all $r \in \Re$ (the optimal size threshold is nondecreasing in time).*

Figure 2. Conditional probability density function $f_{S|R}(s|r)$ of amount S of resource required for different values of the reward r.

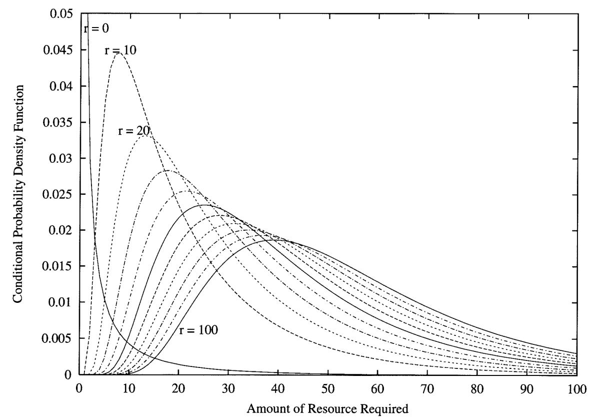


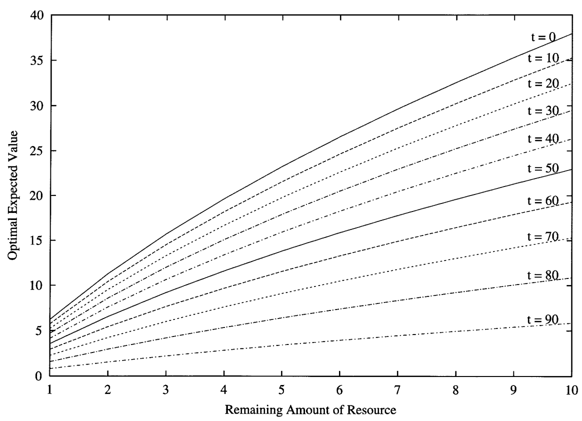
Figure 2 shows the conditional probability density function $f_{S|R}(s|r)$ for different values of the reward r for the example above. It shows that $f_{S|R}(s|r)$ is not nonincreasing in s, and thus $F_{S|R}(s|r)$ is not concave in s for all values of r, which causes $V^*(n,t)$ to be not concave in n and $x^*(n,t,1)$ to be increasing in n for small values of n. More examples are given in Kleywegt (1996) to show why the other conditions in Theorem 9 are needed.

Figure 3 shows $V^*(n,t)$ and $x^*(n,t,1)$ as functions of n for different values of t for an example with exponentially distributed rewards R with mean 25, the amount S of resource required conditionally exponentially distributed with mean equal to the reward, the deadline $T = 100$, demands arrive at rate $\lambda = 0.1$, the penalty $p = 0$, the holding cost per unit time $c = 0$, the terminal value $v = 0$, and the discount rate $\alpha = 0$. Note that $V^*(n,t)$ is concave in n, and $x^*(n,t,1)$ is decreasing in n, as stated in Theorem 9.

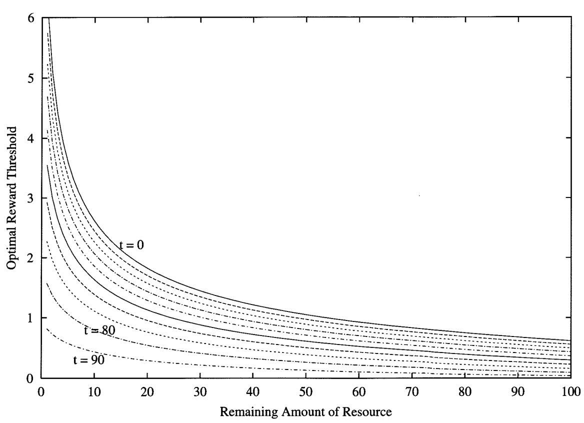
5. CONCLUSION

A dynamic and stochastic resource allocation problem with a large variety of applications was studied. This problem was introduced in a paper by Kleywegt and Papastavrou (1998), who investigated the case where all demands require the same amount of resource. We extended these results to the case where demands require random amounts of resource. Both the infinite horizon and finite horizon cases of the problem were studied. It was shown that an optimal acceptance rule is given by an easily computed threshold rule. It was also shown how the optimal stopping time is determined. For both the infinite horizon and finite horizon cases, structural results were obtained that lead to efficient algorithms for computing the optimal value functions and optimal policies. It was shown that the optimal value function and optimal policy may not exhibit the same intuitive characteristics that were shown to hold in the case where all demands require the same amount of resource. Conditions were obtained under which these characteristics continue to hold.

Figure 3. Example to show monotonicity and convexity properties according to Theorem 9. (a) Optimal expected value $V^*(n, t)$ versus remaining amount n of resource for different points in time t. (b) Optimal reward threshold $x^*(n, t, 1)$ versus remaining amount n of resource for different points in time t.



(a)



(b)

Many applications involve extensions of the DSKP. Often the arrival process of the demands is time-dependent, with daily, weekly, or annual cycles. The finite horizon MDP is easily extended to incorporate time-dependent parameters, but the resulting optimal policies may not be admissible for the DSKP any more. Solving dynamic and stochastic resource allocation problems with time-dependent parameters over finite and infinite time horizons requires additional research. The amount of resources available can itself be a stochastic process. For example, a portfolio manager manages a fund of which the size changes randomly. There may be a number of related knapsacks involved in the problem. For example, airline yield management addresses the selling of seats on a number of flights that interact through passenger transfers, aircraft assignments, timetables, and return flights. It was assumed that the joint distribution of reward and resource requirement is known. In most practical applications, the distributions are not known and have to be estimated with the available data. As the system develops and more data become available, these estimates can be updated, and with it the policies can be improved.

ADDENDUM

An addendum to this paper is included in the *Operations Research* Online Collection:

⟨http : //or.pubs.informs.org⟩.

The addendum contains some results and proofs not included here.

REFERENCES

Albright, S. C. 1977. A Bayesian approach to a general house selling problem. *Management Sci.* **24** 432–440.

Belobaba, P. P. 1989. Application of a probabilistic decision model to airline seat inventory control. *Oper. Res.* **37** 183–197.

Blackwell, D. 1965. Discounted dynamic programming. *Ann. Math. Statist.* **36** 226–235.

Boel, K., P. Varaiya. 1977. Optimal control of jump processes. *SIAM J. Control Optimiz.* **15** 92–119.

Brumelle, S. L., J. I. McGill. 1993. Airline seat allocation with multiple nested fare classes. *Oper. Res.* **41** 127–137.

Carraway, R. L., R. L. Schmidt, L. R. Weatherford. 1993. An algorithm for maximizing target achivement in the stochastic knapsack problem with normal returns. *Naval Res. Logist. Quart.* **40** 161–173.

Derman, C., G. J. Lieberman, S. M. Ross. 1972. A sequential stochastic assignment problem. *Management Sci.* **18** 349–355.

Freeman, P. R. 1983. The secretary problem and its extensions: a review. *Int. Statist. Rev.* **51** 189–206.

Gallego, G., G. Van Ryzin. 1994. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Sci.* **40** 999–1020.

Kincaid, W. M., D. A. Darling. 1963. An inventory pricing problem. *J. Math. Anal. Appl.* **7** 183–208.

Kleywegt, A. J. 1996. Dynamic and Stochastic Models with Freight Distribution Applications. Ph.D. Thesis, School of Industrial Engineering, Purdue University.

——, J. D. Papastavrou. 1998. The dynamic and stochastic knapsack problem. *Oper. Res.* **46** 17–35.

Mamer, J. W. 1986. Successive approximations for finite horizon, semi-Markov decision processes with application to asset liquidation. *Oper. Res.* **34** 638–644.

Martello, S., P. Toth. 1990. *Knapsack Problems. Algorithms and Computer Implementations*. John Wiley & Sons, West Sussex, England.

Papastavrou, J. D., S. Rajagopalan, A. J. Kleywegt. 1996. A stochastic model for the knapsack problem with a deadline. *Management Sci.* **42** 1706–1718.

Prastacos, G. P. 1983. Optimal sequential investment decisions under conditions of uncertainty. *Management Sci.* **29** 118–134.

Presman, E. L., I. M. Sonin. 1972. The best choice problem for a random number of objects. *Theory Probab. Its Appl.* **17** 657–668.

Rockafellar, R. T. 1970. *Convex Analysis*. Princeton University Press, Princeton, NJ.

Rosenfield, D. B., R. D. Shapiro, D. A. Butler. 1983. Optimal strategies for selling an asset. *Management Sci.* **29** 1051–1061.

Ross, K. W., D. H. K. Tsang. 1989. The stochastic knapsack problem. *IEEE Trans. Comm.* **37** 740–747.

Royden, H. L. 1988. *Real Analysis*. Macmillan, New York.

Sniedovich, M. 1980. Preference order stochastic knapsack problems: methodological issues. *J. Oper. Res. Soc.* **31** 1025–1032.

Stadje, W. 1990. A full information pricing problem for the sale of several identical commodities. *Zeitschrift für Oper. Res.* **34** 161–181.

Stewart, T. J. 1981. The secretary problem with an unknown number of options. *Oper. Res.* **29** 130–145.

Weatherford, L. R., S. E. Bodily. 1992. A taxonomy and research overview of perishable-asset revenue management: yield management, overbooking, and pricing. *Oper. Res.* **40** 831–844.

Yushkevich, A. A., E. A. Feinberg. 1979. On homogeneous Markov models with continuous time and finite or countable state space. *Theory Probab. Its Appl.* **24** 156–161.