AD-753 809

MORE ON A CLASS OF OPTIMAL SEARCH PROBLEMS

Warren W. Willman

Naval Research Laboratory
Washington, D. C.

1 December 1972

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

AD753809

NRL Report 7487

More on a Class of Optimal Search Problems

WARREN W. WILLMAN

Operations Research Branch
Report 72-6
Mathematics and Information Sciences Division

December 1, 1972

DDC
JAN 11 1973
C

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U S Department of Commerce
Springfield VA 22151

NAVAL RESEARCH LABORATORY
Washington, D.C.

Approved for public release; distribution unlimited.

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Naval Research Laboratory Washington, D.C. 20390	Unclassified
	2b. GROUP

3 REPORT TITLE

MORE ON A CLASS OF OPTIMAL SEARCH PROBLEMS

4 DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

A final report on one phase of problem. Work continues on other phases of the problem.

5 AUTHOR(S) *(First name, middle initial, last name)*

Warren W. Willman

6 REPORT DATE	7a. TOTAL NO OF PAGES	7b. NO OF REFS
December 1, 1972	15	5

8a. CONTRACT OR GRANT NO	9a ORIGINATOR'S REPORT NUMBER(S)
NRL Problem No. B01-10	NRL Report 7487
b. PROJECT NO RR003-02-41-6152	
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d.	Operations Research Branch Report 72-6

10 DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11 SUPPLEMENTARY NOTES	12 SPONSORING MILITARY ACTIVITY
	Department of the Navy (Office of Naval Research) Arlington, Virginia 22217

13 ABSTRACT

Optimal strategies are investigated for a class of one-dimensional search processes in which the objective is to find a point which is near, but not beyond, a boundary of uncertain location. Problems of this type are encountered in the analysis of mining operations. Upper and lower bounds for the optimal expected payoff are derived, and the optimal search strategies are described explicitly for a large subclass of these processes. Results are obtained by formulating the search as a multistage decision process and using a dynamic programming approach.

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Search Theory Dynamic programming Optimal control Stopping rule						

CONTENTS

MORE ON A CLASS OF OPTIMAL SEARCH PROBLEMS

Warren W Willman

Operations Research Branch
Mathematics and Information Sciences Division

Abstract: Optimal strategies are investigated for a class of one-dimensional search processes in which the objective is to find a point which is near, but not beyond, a boundary of uncertain location. Problems of this type are encountered in the analysis of mining operations. Upper and lower bounds for the optimal expected payoff are derived, and the optimal search strategies are described explicitly for a large subclass of these processes. Results are obtained by formulating the search as a multistage decision process and using a dynamic programming approach.

INTRODUCTION

Optimal policies are investigated here for a class of one-dimensional adaptive search processes in which the objective is to find a point which is near, but not beyond, a boundary of uncertain location. This class is an extension of a class of similar search processes examined previously by the author (1). Aside from theoretical considerations, this extension is important because of its applications to certain mining operations. These problems share some features of those studied by Derman and Ignall (2), but are basically different because the main question here is where to search, not when to stop. They are also basically different from the classical search problem described in Koopman (3), where the objective is to locate a small object, at least approximately, within a large planar region of uncertainty. The only applications of the results of this report to planar searches would be to the location of the boundary of a large planar region, where the uncertainty of the boundary location is small compared to the size of the region. The results here are obtained by formulating the search as a multistage decision process and using a dynamic programming approach.

A SEARCH PROBLEM

The search process considered here proceeds sequentially. At epoch i ($i = 0, 1, \ldots$) a searcher has the choice of terminating the search or selecting the median m_i of a random variable y_i whose distribution is rectangular with width $T \geq 0$. The term m_i represents the desired search point, whereas y_i is the actual search location, which is unknown to the searcher. The random variables $(y_i - m_i)$ are statistically independent, but each has the same distribution width T.

If the search is terminated at epoch $N \geq 0$, the searcher receives a *return* J such that

$$J = \begin{cases} 0 & , \quad \text{if} \quad N = 0 \\ G(y_N) - N\,, & \text{if} \quad N > 0 \end{cases}$$

NRL Problem B01-10; Project RR003-02-41-6152. This is a final report on one phase of the problem. Work on other phases of this problem is continuing. Manuscript submitted August 18, 1972.

where

$$G(x) = \begin{cases} a + kx, & \text{if} \quad x \leq b \\ 0, & \text{if} \quad x > b. \end{cases}$$

Also, $k > 0$, $a > 2 + (1/3k)$, and b is a random variable with a symmetric trapezoidal probability density of the class shown in Fig. 1 such that the lower and upper midpoints are 0 and s_0, respectively,

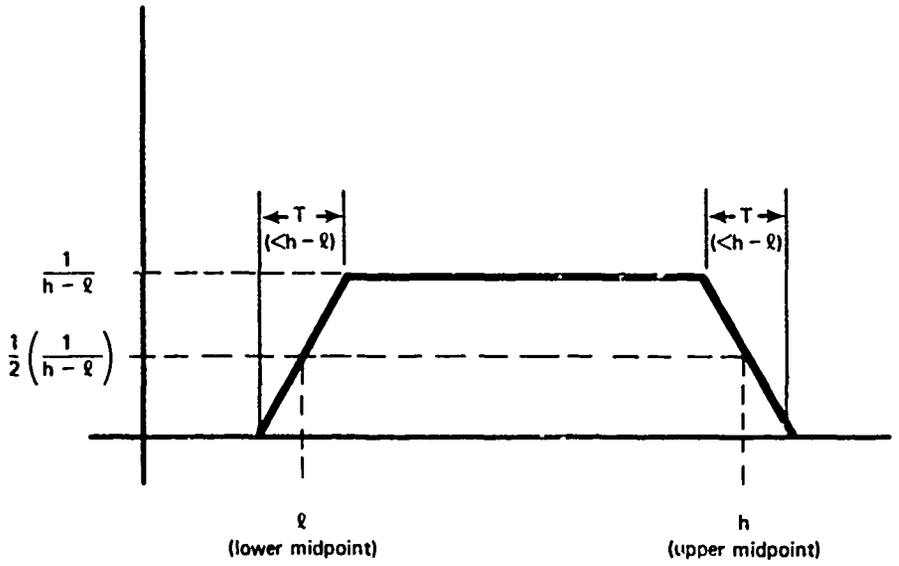

Fig. 1 – A class of symmetric trapezoidal probability densities b. The quantity $T \geq 0$ is the distribution width of the random variables y.

where $s_0 > T$ and $T < 1/3k$. Also, b is statistically independent of the y's. The quantity G represents the gain from the search. The cost of a single search step has been taken as unity, without loss of generality. The random variable b represents a random boundary location. A rectangular density for b on the interval $[0, s_0]$ would serve equally well here, but that would entail more complicated formulae in the following analysis.

At decision time i, the searcher knows the values of s_0, T, k, a, i, and, for all $j < i$, the search decisions m_j and the corresponding values of $\text{sgn}(b - y_j)$. This last sequence of values represents a knowledge of the side of the boundary b on which the previous actual search locations were. The problem investigated in this report is that of finding search policies which maximize the (prior) *expected value* of the return J. As usual, a policy, or strategy, is defined as a decision rule which determines the searcher's action as a function of the information available to him, for any possible realization of the search process, and for which the search terminates with a probability equal to 1. This search is adaptive in the sense that the searcher's actions depend on previous search results.

The search problem treated here is a modification of one studied in a previous work by the present author (1); it differs in only two details. In the previous work $a = 0$, and the value of the return J is

$$\sup_{i<N} G(y_i)$$

if $N > 0$. The present modifications make the search process a more accurate approximation to certain kinds of searches actually conducted in mining operations. Because of its similarity, however, much of the analysis of the previous problem can be carried over to this one. The results are also of a similar sort and will be compared with those obtained for the other search problem at the end of this report.

ANOTHER FORMULATION

It is convenient at this point to define the following four sequences of random variables:

$$h_i = \min\{s_0\} \cup \{m_j : y_j > b,\ j < i\}$$

$$\ell_i = \max\{0\} \cup \{m_j : y_j \leq b,\ j < i\}$$

$$\lambda_i = \max\{0\} \cup \{y_j : y_j \leq b,\ j < i\}$$

$$z_i = \frac{1}{2} + \frac{1}{2}\,\mathrm{sgn}(b - y_i);\ i = 0,1,\ldots;\ z_{-1} = 0;$$

where

$$\mathrm{sgn}(0) \triangleq 1 .$$

It is immediately apparent that there is always a better alternative than choosing m_i outside the interval $[\ell_i - T, h_i + T]$. Search policies for which such a choice is possible will not be considered further. In addition, we temporarily admit only policies for which m_i is always in the interval $[\ell_i + (1/3k), h_i - (1/3k)]$ if $h_i - \ell_i \geq 4/k$.

It can be shown by induction that it is possible to express the return as

$$J = \sum_{i=0}^{N-1} \left[a(z_i - z_{i-1}) + k\left(y_i z_i - \lambda_i (1 - z_i) z_{i-1}\right) - 1 \right]$$

where N is the epoch at which termination occurs. This alternative expression for the return makes the search process amenable to a dynamic programming analysis. The boundary location b, and the quantities λ_i and z_{i-1} serve as the state variables at epoch i in this analysis; the intended search points m_i are the control variables and the search "results" z_i are noisy measurements of the state. The b component of the state is static; the z_{i-1} component is known exactly.

STATE ESTIMATION

The temporary policy restriction ensures that the points 0, s_0, and the m's are all separated by a distance of at least T as long as $h_i - \ell_i \geq 4/k$. By using this fact and the statistical independence of the random variables $(y_i - m_i)$, the usual inductive use of the Bayes Rule shows that the posterior probability density of b at epoch i (given the data available to the searcher at that time) is also a symmetric trapezoidal density of the class previously shown in Fig. 1, whenever this condition is

satisfied. The upper and lower midpoints of this conditional density are h_i and ℓ_i, respectively. The conditional density of λ_i, given b and the data at epoch i, is also determined by the posterior distribution of b under these circumstances, namely by the parameter ℓ_i. Since the state variable z_{i-1} is known exactly from the data at epoch i, it follows that h_i, ℓ_i, and z_{i-1} are sufficient statistics for the joint conditional distribution of the state variables for the portion of the search where $h_i - \ell_i \geq 4/k$. It is important to note that these estimation results depend on the fact that, for any $i \geq j \geq 0$, the relation $h_i - \ell_i \geq 4/k \Rightarrow h_j - \ell_j \geq 4/k$, which is an immediate consequence of the definitions of h_i and ℓ_i.

THE VALUE FUNCTION

Let $\mathfrak{U}$ be the class of search policies which satisfy the temporary restriction imposed previously and for which the functional dependence of the action at epoch i on the available data is determined uniquely by the statistics h_i, ℓ_i, and z_{i-1} for all i such that $h_i - \ell_i \geq 4/k$. Since the joint conditional distribution of the state variables is also determined by these statistics in this case, and since the values of $(y_i - m_i)$ are statistically independent, the following definition is unambiguous for such a policy.

Definition: For $h - (4/k) \geq \ell \geq 0$, $z = 0$ or 1, and $\pi \in \mathfrak{U}$, the quantity $L(i, \ell, h, z, \pi)$ is defined as the conditional expected future return at epoch i from policy π given that $\ell_i = \ell$, $h_i = h$, and $z_{i-1} = z$, where the future return at epoch i is the total return minus the return that would result from terminating the search at that epoch.

For $\pi \in \mathfrak{U}$, the notation $\pi(i, \ell, h, z)$ is used to denote the action specified by π at epoch i for $\ell_i = \ell$, $h_i = h$, and $z_{i-1} = z$. The *value function* is now defined as:

Definition:

$$\left.\begin{aligned} Q(i,\ell,h) &= \sup_{\pi \in \mathfrak{U}} L(i,\ell,h,1,\pi) \\ R(i,\ell,h) &= \sup_{\pi \in \mathfrak{U}} L(i,\ell,h,0,\pi) \end{aligned}\right\} \text{for } h-(4/k) \geq \ell \geq 0;\ i = 0,1,\ldots.$$

The value function is defined in terms of the two partial functions Q and R for conceptual convenience. Intuitively, Q is the optimal expected future return if the last search point was below the boundary, and R is the optimal expected future return if it was above the boundary.

The results of Stratonovich (4) imply that the conditional expected future return for an optimal policy at a given epoch of any realization is determined by the conditional probability distribution of the state under those conditions. Therefore, this value function is the supremum of the conditional expected future returns for *all* search policies satisfying the restriction imposed in an earlier section titled "Another Formulation" (for the domain of definition of this function). In particular, it is the optimal value function in the sense of Bellman (5) if optimal policies exist within this restricted set of policies. Furthermore, it will be shown later that no optimal policies are excluded by this restriction, so this value function has these properties with respect to the class of all admissible search policies.

The situation is more complicated if $h_i - \ell_i < 4/k$ because the statistics h_i, ℓ_i, and z_{i-1} are in general no longer sufficient to determine the conditional probability distribution of the state variables. This case will be treated separately.

THE BELLMAN EQUATION

For $\pi \in \Pi$ and $h-(4/k) \geq \ell \geq 0$, the additive expression for J and the statistical independence of the $(y_i - m_i)$'s imply the recursion

$$L(i,\ell,h,z,\pi) = \begin{cases} f(i,\ell,h,z,m,\pi), & \text{if} \quad \pi(i,\ell,h,z) = \text{"search at } m\text{"} \\ 0 \quad , & \text{if} \quad \pi(i,\ell,h,z) = \text{"terminate search"} \end{cases}$$

where

$$f(i,\ell,h,z,m,\pi) \triangleq \mathop{\mathcal{E}}_{\substack{j\ell_i=\ell \\ h_i=h \\ z_{i-1}=z \\ m_i=m}} \left\{ L(i+1,\ell_{i+1},h_{i+1},z_i,\pi) + a(z_i - z_{i-1}) + k\left(y_i z_i - \lambda_i (1-z_i) z_{i-1}\right) - 1 \right\} .$$

Repeating the manipulations performed in Ref. 1, it follows from the Principle of Optimality developed by Bellman (5) that the value function satisfies the equations (together constituting the Bellman equation)

$$Q(i,\ell,h) = \max \begin{cases} 0 , \\ \sup\limits_{\ell+(1/3k)<m<h-(1/3k)} \left\{ \dfrac{h-m}{h-\ell} \left[k(m-\ell) + Q(i+1,m,h) \right] \right. \\ \qquad + \dfrac{m-\ell}{h-\ell} \left[R(i+1,\ell,m) - a - k\ell \right] \\ \qquad \left. - \dfrac{kT^2}{12(h-\ell)} - 1 \right\} \end{cases}$$

$$R(i,\ell,h) = \sup_{\ell+(1/3k)<m<h-(1/3k)} \left\{ \frac{h-m}{h-\ell} \left[k(m-\ell) + Q(i+1,m,h) \right] \right.$$
$$+ \frac{m-\ell}{h-\ell} \left[R(i+1,m,\ell) - a - k\ell \right]$$
$$\left. + a + k\ell - \frac{kT^2}{12(h-\ell)} - 1 \right\}$$

for $h-(4/k) \geq \ell \geq 0$. The reason that R cannot be zero is that searching at $m = 1/3k$, then at $m = -1/3k$ (guaranteeing that $z_{i+1} = 1$), is always preferable to terminating the search at epoch i if $z_{i-1} = 0$ because $a > 2 + (1/3k)$. From these two equations, it follows that

$$Q(i,\ell,h) \equiv \max\{0, R(i,\ell,h) - a - k\ell\} \tag{1}$$

in this range of ℓ and h.

ADDITIONAL PROPERTIES OF THE OPTIMAL VALUE FUNCTION

The next step here is to establish some additional properties of the partial value function Q which will be useful in the analysis of this search problem. These properties are proved in this section as a series of lemmata. In the context of this entire section, Q is the partial value function as defined in an earlier section titled "The Value Function", and not any other solution to the Bellman equation, if such exist.

LEMMA 1. $h - \ell = 4/k \Rightarrow Q(i, \ell, h) = 0$.

Proof. For any admissible policy and any realization of the random variables of the search process, the future return at any epoch i, such that $z_{i-1} = 1$ and $h_i - \ell_i \geqslant 4/k$, is bounded above by the future return from this policy in the search process analyzed in Ref. 1 with the same values of s_0, T, k, and i, because the future return is the same if $z_N = 1$ and exceeds the future return of the present process by a if $z_N = 0$. Since every policy which is admissible here is also admissible in this other search process, and since the optimal expected future return is zero in the other process if $h_i - \ell_i = 4/k$, the lemma is verified.

LEMMA 2. *$Q(i, \ell, h)$ is monotonic in $(h-\ell)$.*

Proof. Suppose that $h-\ell \geqslant 4/k$ is increased by a factor $c > 1$ for some value of i. If the problem is changed so that the value of T is also increased by this factor, the remainder of the search process at epoch i for the corresponding realization is merely scaled up by the same factor. Thus, for any policy $\pi \in \Pi$ in the original problem giving conditional expected future return M at this point, the scaled-up policy (such that cm_j always replaces m_j) is admissible in the scaled-up problem and gives a return greater than M for each realization, and hence a greater conditional expected future return at the corresponding point. Therefore,

$$Q(i, \ell, h) \leqslant Q_1(i, c\ell, ch)$$

where Q_1 is the corresponding partial value function for the scaled-up search problem. Finally, since an increase of T is a degradation of search data, it can never increase the supremum of a conditional expected future return, so that

$$Q_1(i, c\ell, ch) \leqslant Q(i, c\ell, ch) .$$

Definition: $s^* = \inf\{s: \exists\, i, h, \ell \ni s = h - \ell \quad \text{and} \quad Q(i, \ell, h) > 0\}$.

LEMMA 3. $Q(i, \ell, \ell + s^*) \equiv 0$.

Proof. If $s^* = 4/k$, this is true by Lemma 1; if not, $s^* > 4/k$. In this case, assume that $Q(i, \ell, \ell + s^*) = g > 0$. Let $m^* \in [\ell + (1/3k), \ell + s^* - (1/3k)]$ be such that

$$\frac{k}{s^*}(\ell + s^* - m^*)(m^* - \ell) + \frac{m^* - \ell}{s^*}[R(i+1, \ell, m^*) - a - k\ell] - \frac{kT^2}{12s^*} - 1 > \frac{g}{2} .$$

Such an m^* exists by virtue of the Bellman equation and Lemma 1. It also follows from Lemma 1 that for any ϵ such that $0 < \epsilon < s^* - 4/k$,

$$\left(\frac{\ell+s^*-m-\epsilon}{s^*-\epsilon}\right)k(m^*-\ell)+\left(\frac{m^*-\ell}{s^*-\epsilon}\right)[R(i+1,\ell,m^*)-a-k\ell]-\left(\frac{kT^2}{12(s^*-\epsilon)}\right)-1$$

$$\leqslant Q(i,\ell,\ell+s^*-\epsilon)=0\,.$$

From the definition of m^*, however, this is impossible for sufficiently small ϵ because all of the terms in the large parenthesis are continuous functions of ϵ.

LEMMA 4. *$Q(i,\ell,\ell+s)$ depends only on s.*

Proof. Lemmata 1 to 3 imply that Q satisfies the following modified Bellman equation:

$$Q(i,\ell,h)=\sup_{1/3k\leqslant u\leqslant s-(1/3k)}\left\{I(s)\left[\frac{s-u}{s}\left(ku-\frac{kT^2}{12(s-u)}\right)-1\right]\cdot I(s)\,\frac{s-u}{s}\,Q(i+1,\ell+u,\ell+s)\right.$$

$$\left.+I(s)\,\frac{u}{s}\,Q(i+1,\ell,\ell+u)\right\}$$

where

$$I(s)\triangleq\begin{cases}0\,, & \text{for} \quad s\leqslant s^*\\ 1\,, & \text{for} \quad s>s^*\end{cases}$$

and where $u=m-\ell$. Since $Q(i,\ell,\ell)\equiv 0$ by Lemma 1, $Q(i,\ell,\ell+s)$ must depend only on s to avoid a contradiction.

A SIMPLIFICATION

By Lemma 4, the following definition is unambiguous:

Definition:

$$V(s)\triangleq\begin{cases}Q(i,\ell,\ell+s)\,, & \text{for} \quad s\geqslant 4/k\\ 0\qquad\quad\;\;, & \text{for} \quad 0\leqslant s\leqslant 4/k\,.\end{cases}$$

Also, by the proof of Lemma 4, V satisfies the equation and boundary condition given by

$$V(s)=\sup_{1/3k\leqslant u\leqslant s-(1/3k)}\left\{I(s)\left[\frac{s-u}{s}\left(ku-\frac{kT^2}{12(s-u)}\right)-1\right]+I(s)\,\frac{s-u}{s}\,V(s-u)+I(s)\,\frac{u}{s}\,V(u)\right\} \tag{2}$$

and

$$V(0)=0$$

where

$$I(s) \triangleq \begin{cases} 0, & \text{for} \quad s \leqslant s^* \\ 1, & \text{for} \quad s > s^*. \end{cases}$$

LEMMA 5. *There exists a unique solution V to Eq. 2.*

Proof. By an extension of Theorem 1 in Chapter IV of Bellman's book (5), there exists one and only one solution V to this equation such that $V(0) = 0$ and $V(s)$ is continuous at $s = 0$. Since all solutions clearly have these two properties, the lemma follows.

The problem of finding the partial value function Q has now been simplified to finding s^* and finding a solution to Eq. 2. The following result is helpful in determining s^*:

LEMMA 6. $s^* = \sup\{s \geqslant 0 : R(0,0,s) < a\}$.

Proof. By Lemma 4 and Eq. 1,

$$Q(i,\ell,h) = Q(0,0,s) = \max\{0, R(0,0,h-\ell) - a\}.$$

Therefore,

$$Q(i,\ell,h) > 0 \Leftrightarrow R(0,0,s) - a > 0 \quad \text{and} \quad s = h - \ell.$$

By definition,

$$s^* = \inf\{s \geqslant 0 : \exists\, i,\ell,h \ni s = h - \ell \quad \text{and} \quad Q(i,\ell,h) > 0\}.$$

Hence,

$$\begin{aligned} s^* &= \inf\{s \geqslant 0 : \exists\, i,\ell,h \ni s = h - \ell \quad \text{and} \quad R(0,0,s) > a\} \\ &= \inf\{s \geqslant 0 : R(0,0,s) > a\} \\ &= \sup\{s \geqslant 0 : R(0,0,s) < a\}, \end{aligned}$$

since termination is nonoptimal, $Q(i,\ell,h)$ is monotonic in $(h-\ell)$, and $R = Q$ when $Q \geqslant 0$.

OPTIMAL POLICIES

The simplified Bellman Eq. 2 is exactly the same as the one derived for the search process studied in Ref. 1, except that the parameter s^* here replaces $s_{\min}$ there. The results there, and the fact that $Q = R$ for $Q \geqslant 0$, imply the following results for $\pi \in \Pi$ (this policy restriction is shown to be superfluous in the next section):

a. If $z_{i-1} = 1$ at epoch i, it is optimal to continue the search if and only if $s_i \geqslant s^*$.

b. It is always optimal to continue the search at epoch i if $z_{i-1} = 0$.

c. If $s \in [2^{n-1}s^*, 2^n s^*]$, $n = 0, 1, 2, \ldots$,

then

$$V^-(s) \triangleq \frac{k}{2}\left(s + \frac{T^2}{6s} - \frac{s}{2^n} - \frac{2^n T^2}{6s}\right) - n \leq V(s) \leq \frac{k}{2}\left(s - \bar{s} + \frac{T^2}{6}\left[\frac{1}{s} - \frac{1}{\bar{s}}\right]\right) - \log_2\left(\frac{s}{\bar{s}}\right) \triangleq V^+(s)$$

where

$$\bar{s} \triangleq \frac{\log_2 e}{k} + \sqrt{\left(\frac{\log_2 e}{k}\right)^2 + \frac{T^2}{6}}\,.$$

d. $V^+(s) - V^-(s) \leq V^+(s^*)$ for $s \geq s^*/2$.

e. $V^+(2^n\bar{s}) = V^-(2^n\bar{s}) = V(2^n\bar{s})$ for $n = 0, 1, 2, \ldots$.

f. The policy π^-, where $m_i = (h_i/2) + (\ell_i/2)$, is optimal if $s_0 = 2^n\,\bar{s}$, $n = 1, 2, \ldots$, and $h_i - \ell_i > s^*$.

g. Cases exist where $h_i - \ell_i > s^*$ and $m_i = (h_i/2) + (\ell_i/2)$ are not optimal.

Furthermore, if $T = 0$ the conditional probability density of b given the search results is rectangular for any policy in $\mathfrak{U}$, so the Bellman equation can be extended to entire search process. It is then straightforward, but tedious, to show by direct substitution in the Bellman equation that the policy

$$u_i = \begin{cases} 0 & , \quad \text{if} \quad 0 \leq s_i \leq 1/k \\ \dfrac{s_i}{2} - \dfrac{1}{2k} & , \quad \text{if} \quad 1/k \leq s_i \leq 3/k \\ \dfrac{2}{3}\,s_i - 1/k\,, & \quad \text{if} \quad 3/k \leq s_i \leq (3+\sqrt{6})/k \end{cases}$$

is optimal for $s_i \leq s^*$ and, by Lemma 6, that $s^* = (3+\sqrt{6})/k$ in this case. If $T > 0$ it is possible that admissible policies lead to posterior probability distributions for b which are not symmetric trapezoidal; so s^* and optimal policies for $s < s^*$ cannot be found in this way. Since an increase of T from zero to a positive value represents a degradation of information, however, the quantity $(3+\sqrt{6})/k$ is a lower bound for s^* in this case. An upper bound can be established by evaluating the expected return from the admissible but nonoptimal policy

$$u_i = \begin{cases} -T & , \quad \text{if} \quad -T \leq h_i - \ell_i < 1/k \\ \dfrac{s_i}{2} - \dfrac{1}{2k} - T & , \quad \text{if} \quad 1/k \leq h_i - \ell_i < 3/k \\ \dfrac{2}{3}\,s_i - \dfrac{1}{k} - T\,, & \quad \text{if} \quad 3/k \leq h_i - \ell_i \leq 6/k\,. \end{cases}$$

It is not important to evaluate this return exactly; it is positive if

$$s_i = \frac{3+\sqrt{6}}{k} + \frac{7}{6}\,kT\,.$$

Hence, the results about s^* obtained here can be summarized as

$$\frac{3+\sqrt{6}}{k} \leqslant s^* \leqslant \frac{3+\sqrt{6}}{k} + \frac{7}{6} kT .$$

REMOVAL OF POLICY RESTRICTION

The analysis in this report has heretofore been based on a restriction of admissible search policies to those for which $m_i \in [\ell_i + 1/3k, h_i - 1/3k]$ always. It is the purpose of the present section to show that this restriction, although convenient for analytical reasons, is superfluous with regard to optimal policies. In particular, it is shown that for any policy not satisfying this restriction there is a policy that does, and one which gives an expected return at least as great. Consequently, the results pertaining to optimal search policies in this report can be regarded as applying to all search policies without restriction.

LEMMA 7. *If $s_i \geqslant 4/k$ and $T < 1/3k$ for any epoch i, it is not optimal to choose m_i such that $m_i \notin [\ell_i + (1/3k), h_i - (1/3k)]$.*

Proof. Assuming the contrary, the Principle of Optimality (5) implies the existence of a case where $s_0 \geqslant 4/k$, $T < 1/3k$, a policy, and a realization of the search process such that $m_j < \ell_j + 1/3k$, or $m_j > h_j - 1/3k$, for some epoch j, and such that this policy's conditional expected future return at that point is greater than that of any policy which terminates then or for which $\ell_j + 1/3k \leqslant m_j \leqslant h_j - 1/3k$. Let the variable triple (s_0, T, k) be fixed such that this possibility exists and let A be the set of all such triples with this property. Define the set B as

$$B = \{x \geqslant 4/k : (x, T, k) \in A\} .$$

Let σ be an element of B such that $\sigma < \inf B + 1/3k$, and let Π denote an admissible policy for which the possibility described above exists for the triple (σ, T, k), the existence of which is guaranteed by the construction of σ. Consider the corresponding search process and realization and denote by i the first epoch for which $m_i < \ell_i + 1/3k$ or $m_i > h_i - 1/3k$. For convenience, denote $h_i - \ell_i$ by s_i and $m_i - \ell_i$ by u_i. All probabilities and expectations in the following computations are meant to be conditioned on the data available to the searcher at epoch i.

Since $\ell_j + T < \ell_j + 1/3k \leqslant m_j \leqslant h_j - 1/3k < h_j - T$ for all $j < i$, the conditional density of b at epoch i is symmetric trapezoidal with upper and lower midpoints at h_i and ℓ_i. By construction, $s_i \geqslant 4/k > 12T$.

Case 1: $m_i < \ell_i + 1/3k$. In this case, the searcher's conditional expected future return is not decreased by giving him free knowledge at epoch $i+1$ of the random variable $(1/2) - (1/2)\operatorname{sgn}(b - \alpha + \zeta)$, where $\alpha = \max(1/3k, u_i + 1/3k)$ and ζ is a random variable independent of b and the y's with rectangular density of median zero and width T, and allowing him to proceed optimally with this extra knowledge. At this point (epoch $i+1$) the conditional density of b is symmetric trapezoidal by construction with midpoints less than $\inf(B)$ apart. Hence, the optimal expected future return is given by the Bellman equation for policies in $\mathfrak{U}$. Some tedious computations similar to those in Ref. 1 then show that the conditional expected future return from policy Π at epoch i in this case is less than or equal to

$$a - 1 + k(\alpha + T) + Q(i, \ell_i + \alpha, h_i)$$

if $z_{i-1} = 0$, and

$$-1 + k(\alpha + T) + Q(i, \ell_i + \alpha, h_i)$$

if $z_{i-1} = 1$. Since $Q(i,\ell,k)$ is monotonic in $h-\ell$ and is greater than or equal to $R(i,\ell,h)$, either of these possibilities implies that

$$0 < -1 + k(\alpha + T - \ell_i) \leqslant kT - 1/3,$$

contradicting the original assumption.

Case 2: $m_i > h_i - 1/3k$. The proof in this case is similar, except that the "free" extra information given to the searcher at epoch $i+1$ is the random variable $(1/2) - (1/2)\,\mathrm{sgn}(b-\beta+\xi)$, where $\beta = \min(h_i - 1/3k,\ h_i, -u_i - 1/3k)$.

DISCUSSION

As was mentioned in the section titled "Another Formulation," the search problem investigated in this report is similar to the one described in Ref. 1. The main differences in the results here are the inclusion of a new state variable z_{i-1} in the formulation of the search as an optimal control problem, a new stopping rule, the searching strategy for $s_i < s^*$, and the difficulty in finding the value of s^* (which corresponds to $s_{\min}$ in the previous problem) for $T \neq 0$. If $s_0 > s^*$, the optimal expected return in this search process is $a + V(s_0)$, where V is defined by Eq. 2. Although this function is slightly different from the function called V in Ref. 1, it has many similar properties (see section titled "Optimal Policies") and obeys the same equation, except that $s^* \neq s_{\min}$. It is for this reason that the certainty-equivalent policy π^- is optimal for $s_i = 2^n\bar{s}$ in this search process, as well as in the one investigated in Ref. 1.

REFERENCES

1. Willman, W.W., "On a Class of Optimal Search Problems," NRL Report 7309, Sept. 1971.
2. Derman, C., and Ignall, E., "On Getting Close to But Not Beyond a Boundary," *J. Math. Anal. & Appl.* 28:128-143 (1969).
3. Koopman, B.O., "Search and Screening," OEG Report 56, 1946.
4. Stratonovich, R.L., "On the Theory of Optimal Control: Sufficient Coordinates," Automation and Remote Control, 23:847-854 (1963).
5. Bellman, R., "Dynamic Programming," Princeton: Princeton Univ. Press, 1957.