

On a Class of Optimal Search Problems

W. W. WILLMAN

**Operations Research Group
Report 71-4**

Mathematics and Information Sciences Division

September 28, 1971



NAVAL RESEARCH LABORATORY
Washington, D.C.

CONTENTS

Abstract.....	1
Problem Status.....	1
Authorization.....	1
INTRODUCTION.....	1
A SEARCH PROBLEM.....	1
ANOTHER FORMULATION.....	2
STATE ESTIMATION.....	3
THE VALUE FUNCTION.....	3
THE BELLMAN EQUATION.....	3
A SIMPLIFICATION.....	5
OPTIMAL POLICIES.....	7
REMOVAL OF POLICY RESTRICTION.....	10
DISCUSSION.....	11
REFERENCES.....	11

On a Class of Optimal Search Problems

WARREN W. WILLMAN

*Operations Research Group
Mathematics and Information Sciences Division*

Abstract: Optimal policies 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 policies are described explicitly for a large subclass of these problems. Results are obtained by formulating the search as a multistage decision process and using a dynamic programming approach.

INTRODUCTION

Optimal policies 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 sort are encountered in the analysis of mining operations. They share some features of the problems studied by Derman and Ignall (1) but are basically different because the main question is where to search, not when to stop. 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 , where $i = 0, 1, \dots$, 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 . The term m_i represents the desired search point, whereas y_i is the actual search location which is unknown to the searcher. The y 's 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

$$J = \sup\{0\} \cup \{G(y_i) : i < N\} - N,$$

where G represents the gain from the search and has the form:

$$G(x) \triangleq \begin{cases} kx, & \text{if } x \leq b; k > 0 \\ 0, & \text{if } x > b. \end{cases}$$

The cost of a single search step has been taken as unity without loss of generality. The quantity b represents a random boundary location and has a symmetric trapezoidal probability density of the class shown in Fig. 1, such that the "lower" and "upper midpoints" are 0 and s_0 , where $s_0 > T$. Also, b is statistically independent of the y 's.

NRL Problem B01-10; Project RR 003-02-41-6152. This is a final report one phase of the problem; work is continuing on other phases. Manuscript submitted May 20, 1971.

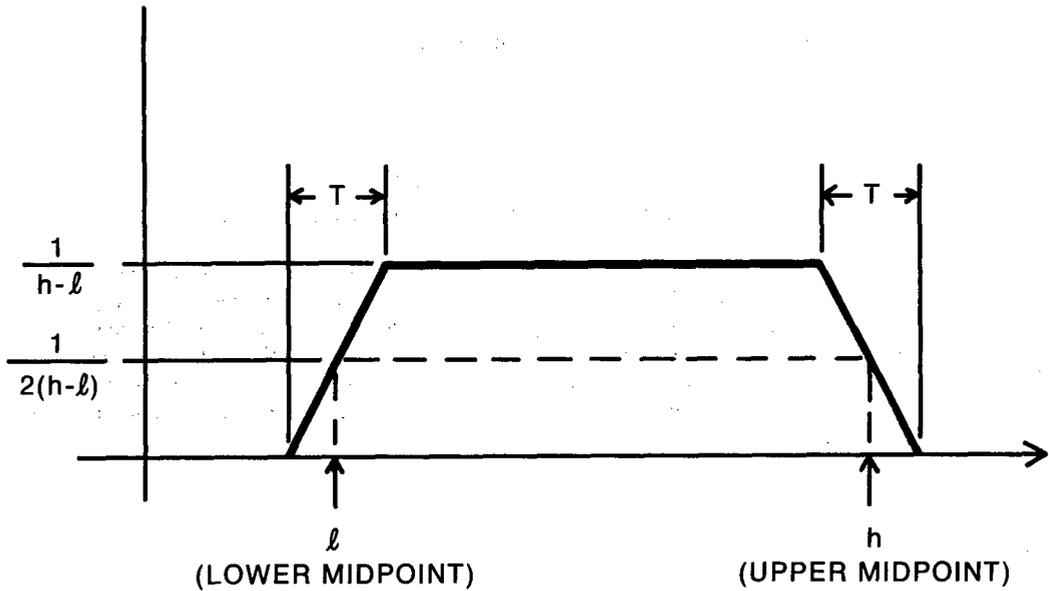


Figure 1 — A class of symmetric trapezoidal probability densities. The term T is the distribution width of the y 's, where $T < h - l$.

At decision time i , the searcher knows the parameters s_0, T, k, i , the previous decisions m_j , where $j < i$, and the values of $\text{sgn}(b - y_j)$ for all $j < i$ (i.e., the side of the boundary on which the previous actual search points were). The problem investigated here is finding search policies which maximize the (prior) *expected value* of the return. As usual, a policy 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 always terminates.

ANOTHER FORMULATION

At this point the following three sequences of random variables can conveniently be defined:

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

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

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

It is immediately apparent that a better alternative than choosing m_i outside the interval $[l_i - T, h_i + T]$ always exists. 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 $[l_i + T, h_i - T]$. Conditions are established later under which this restriction can be removed with no effect on the optimal search policies. This additional restriction makes it possible to express the return as

$$J = \sum_{i=0}^{N-1} (k(y_i - \lambda_i) [1/2 + 1/2 \text{sgn}(b - y_i)] - 1), \text{sgn}(0) \triangleq 1,$$

where N is the epoch at which the decision is made to terminate the search.

This alternative expression for the return makes this search process amenable to a dynamic programming analysis. The boundary location b and λ_i serve as the state variables in this analysis; the search points m_i are the control variables, and the search "results" $\text{sgn}(b - y_i)$ are noisy measurements of the state. The b component of the state is static; neither component is known exactly.

STATE ESTIMATION

The temporary policy restriction ensures that the points 0 and s_0 and the m 's are all separated by a distance of at least T . By using this restriction and the statistical independence of the random variables y_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 shown in Fig. 1. 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 , namely by the parameter l_i . Therefore, h_i and l_i are sufficient statistics for the joint conditional distribution of the state variables given the data.

THE VALUE FUNCTION

Let \mathcal{U} be the class of search policies which satisfy the restriction imposed previously and for which the functional dependence of the action at epoch i on the previous search points and results is determined by h_i and l_i for all values of i . Since the joint conditional distribution of b and λ_i is also determined by h_i and l_i , and since the values of y_i are statistically independent, the following definition is unambiguous for a policy $\pi \in \mathcal{U}$:

Definition. For $h - T > l \geq T$, $\pi \in \mathcal{U}$, $L(i, l, h, \pi)$ is defined as the conditional expected future return at epoch i from policy π given that $l_i = l$ and $h_i = h$, 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 $\pi \in \mathcal{U}$, the notation $\pi(i, l, h)$ is used to denote the action specified by π at epoch i for $l_i = l$ and $h_i = h$. The value function Q is now defined as follows:

Definition. $Q(i, l, h) \triangleq \sup_{\pi \in \mathcal{U}} L(i, l, h, \pi)$.

The results of Stratonovich (2) 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, if $l_i = l$ and $h_i = h$, then $Q(i, l, h)$ is the supremum of the conditional expected future returns for all policies satisfying the restriction imposed in the section "Another Formulation." In particular, Q is the optimal value function if optimal policies exist.

THE BELLMAN EQUATION

For $\pi \in \mathcal{U}$, the additive expression for J and the statistical independence of the y 's imply the recursion

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

where

$$f(i, l, h, m, \pi) \triangleq \mathbb{E}_{\substack{l_{i+1}=l \\ h_{i+1}=h \\ m_i=m}} L(i+1, l_{i+1}, h_{i+1}, \pi) - 1 + k(y_i - \lambda_i) [1/2 + 1/2 \text{sgn}(b - y_i)].$$

Using the state estimation results derived earlier, f can be evaluated as

$$f(i, \ell, h, m, \pi) = P_r(y_i \leq b) [k \mathfrak{E}(y_i - \lambda_i) + L(i+1, m, h, \pi)] \\ + P_r(y_i > b) L(i+1, \ell, m, \pi) - 1,$$

where the expectation is also conditioned on the event $y_i \leq b$. Since λ_i depends only on $\{y_j: j < i\}$ and since the y 's are statistically independent,

$$\mathfrak{E}_{\substack{y_i \leq b \\ \ell_i = \ell \\ h_i = h \\ m_i = m}}(\lambda_i) = \mathfrak{E}_{\substack{\ell_i = \ell \\ h_i = h \\ m_i = m}}(\lambda_i) = \ell.$$

Also, it is straightforward to show that

$$P_r(y_i \leq b | \ell_i = \ell, h_i = h, m_i = m) = \frac{h - m}{h - \ell}.$$

Finally, the Bayes Rule implies that the conditional density of y_i , given $y_i \leq b$, is

$$p_{y_i | y_i \leq b}(t) = \frac{P_r(y_i \leq b | y_i = t)}{P_r(y_i \leq b)} p_{y_i}(t) \\ = \frac{P_r(b \geq t)}{P_r(y_i \leq b)} p_{y_i}(t),$$

where all probabilities are conditioned on $\ell_i = \ell$, $h_i = h$, and $m_i = m$. So

$$p_{y_i | y_i \leq b}(t) = \begin{cases} \frac{h - t}{h - m} \cdot \frac{1}{T}, & \text{if } |t - m| \leq 1/2T \\ 0 & \text{otherwise.} \end{cases}$$

Therefore,

$$\mathfrak{E}_{\substack{y_i \leq b \\ \ell_i = \ell \\ h_i = h \\ m_i = m}}(y_i) = \int_{m-1/2T}^{m+1/2T} t \cdot \frac{h - t}{h - m} \cdot \frac{1}{T} dt = m - \frac{T^2}{12(h - m)}.$$

From the additivity of the expectation operator, it follows that

$$f(i, \ell, h, m, \pi) = \frac{h - m}{h - \ell} \left[L(i+1, m, h, \pi) + k \left(m - \ell - \frac{T^2}{12(h - m)} \right) \right] \\ + \frac{m - \ell}{h - \ell} L(i+1, \ell, m, \pi) - 1.$$

From the dynamic programming argument explained in Bellman (1), an optimal policy for a multistage decision process must be optimal at every intermediate epoch (the Principle of Optimality). It follows by a standard backward induction argument (3) and by an appropriate limiting procedure if an optimal policy does not exist, that the value function satisfies the following Bellman Equation:

$$Q(i, l, h) = \max \left\{ 0, \sup_{l+T \leq m \leq h-T} \sup_{\pi \in \mathcal{U}} f(i, l, h, m, \pi) \right\}.$$

If π° is an optimal search policy (assumed in \mathcal{U} with no loss of generality), it is a further consequence of this argument that

- (i) $Q(i, l, h) > 0 \Rightarrow Q(i, l, h) = f(i, l, h, \pi^\circ(i, l, h), \pi^\circ)$
- (ii) $Q(i, l, h) = 0 = \pi^\circ(i, l, h) = \text{"terminate"} \text{ or } f(i, l, h, \pi^\circ(i, l, h), \pi^\circ) = 0.$

A SIMPLIFICATION

Defining the two new variables

$$s \triangleq h - l$$

and

$$u \triangleq m - l,$$

the Bellman equation can be written as

$$Q(i, l, h) = \max \left\{ 0, \sup_{T \leq u \leq s-T} \left\{ \frac{s-u}{s} \left[Q(i+1, l+u, l+s) + k \left(u - \frac{T^2}{12(s-u)} \right) \right] + \frac{u}{s} Q(i+1, l, l+u) - 1 \right\} \right\}.$$

Because of the current policy restriction, the search must terminate at epoch i if $h_i - l_i < 2T$, which implies that

$$Q(i, l, l) \equiv 0.$$

To avoid a contradiction, therefore, $Q(i, l, l+s)$ must depend only on s . This makes the following definition unambiguous:

Definition. For $s \geq 0$,

$$V(s) \triangleq Q(i, l, l+s).$$

Furthermore, the Bellman equation and its boundary conditions for this search process can be expressed as

$$V(s) = \max \left\{ 0, \sup_{T \leq u \leq s-T} \left\{ \frac{s-u}{s} \left[V(s-u) + k \left(u - \frac{T^2}{12(s-u)} \right) \right] + \frac{u}{s} V(u) - 1 \right\} \right\}.$$

$$V(0) = 0$$

Henceforth, only this simplified equation will be referred to as the Bellman equation for the search process, and V as its value function.

LEMMA. *If $V(s)$ is a solution to the Bellman equation,*

$$V(s) = 0 \Leftrightarrow s < 2T \text{ or } \frac{ks}{4} - \frac{kT^2}{12s} \leq 1.$$

Proof. (\Leftarrow).

By definition, $V(s) \geq 0$ and $s < 2T \Rightarrow V(s) = 0$.

If

$$\frac{ks}{4} - \frac{kT^2}{12s} \leq 1, V(s) > 0, \text{ and } s \geq 2T,$$

then

$$\sup_{T \leq u \leq s-T} \left\{ \frac{s-u}{s} V(s-u) + \frac{u}{s} V(u) \right\} > 0,$$

because

$$\frac{s-u}{s} ku \leq \frac{ks}{4} \text{ for all } u \in (T, s-T).$$

Therefore, $\exists u^* \in [T, s-T]$ such that $V(u^*) > 0$. This argument can be repeated, replacing s by u^* , because $(ks/4) - (kT^2/12s)$ is monotonic in s . After a number of such repetitions not exceeding s/T , this argument implies that $\exists u^* \in [T, 2T]$ such that $V(u^*) > 0$. If $u^* = 2T$ in the last repetition, the Bellman equation and monotonicity of $(ks/4) - (kT^2/12s)$ would imply that

$$\frac{ks}{4} - \frac{kT^2}{12s} \geq \frac{kT}{2} - \frac{kT}{24} > 1,$$

a contradiction. Any other possibility contradicts the Bellman equation.

(\Rightarrow)

If $s \geq 2T$ and $(ks/4) - (kT^2/12s) > 1$, then it follows from the Bellman equation (using $u = s/2$) that

$$V(s) \geq \frac{ks}{4} - \frac{kT^2}{12s} - 1 + V\left(\frac{s}{2}\right) > V\left(\frac{s}{2}\right) \geq 0. \quad \square$$

COROLLARY. *A unique solution to the Bellman equations exists for this search process.*

Proof. From the preceding lemma, the Bellman equation is equivalent to

$$V(s) = \sup_{T \leq u \leq s-T} \left\{ I(s) \left[\frac{s-u}{s} \left(ku - \frac{kT^2}{12(s-u)} \right) - 1 \right] + I(s) \frac{s-u}{s} V(s-u) + I(s) \frac{u}{s} V(u) \right\};$$

$$V(0) = 0,$$

where

$$I(s) \triangleq \begin{cases} 1, & \text{if } \frac{ks}{4} - \frac{kT^2}{12s} > 1 \text{ and } s \geq 2T \\ 0 & \text{otherwise.} \end{cases}$$

An extension of Theorem 1 in Chapter IV of Ref. 3 can be applied to this equation to give the desired result. \square

OPTIMAL POLICIES

If $\pi^* \in \mathcal{Q}$ is a search policy such that $\pi^*(i, l, h) = \text{“terminate”} \Leftrightarrow h_i - l_i < s_{min}$ and if

$$V(s) \equiv \frac{s - u^*}{s} \left(V(s - u^*) + k \left[u^* - \frac{T^2}{12(s - u^*)} \right] \right) + \frac{u^*}{s} V(u^*) - 1;$$

otherwise, where $u^* = \pi^*(i, l, h) - l$, $s = h - l$, and where

$$s_{min} \triangleq \inf \left\{ s \geq 2T : \frac{ks}{4} - \frac{kT^2}{12s} \geq 1 \right\} = \max \left\{ 2T, \frac{2}{k} + \sqrt{\left(\frac{2}{k}\right)^2 + \frac{T^2}{3}} \right\},$$

then π^* is optimal since the solution to the Bellman equation is unique. Formal differentiation of the right-hand side of this equation with respect to u gives the following expression for $s > s_{min}$:

$$\frac{k}{s} (s - 2u) + \frac{1}{s} [V(u) - V(s - u)] + \left(\frac{u}{s} [V'(u) + V'(s - u)] - V'(s - u) \right).$$

All three terms are zero if $u = s/2$, suggesting that the following policy, referred to here as π^- , is optimal:

$$\pi^- : \begin{cases} \text{Terminate search if } h_i - l_i < s_{min} \\ \text{Choose } m_i = 1/2(h_i + l_i) \text{ otherwise.} \end{cases}$$

This policy is not always optimal, however. For example, if $s_0 = 8$, $T = 0$, and $k = 1$, the expected return from π^- is 1, whereas an expected return of 25/24 is given by the policy:

$$\begin{cases} m_0 = 3-1/3. \\ \text{Terminate search at epoch 1 if } h_1 = 3-1/3, \text{ (i.e., if } y_0 > b). \\ \text{Choose } m_1 = 5-2/3 \text{ and terminate at epoch 2 otherwise.} \end{cases}$$

Nevertheless, the conditional expected future return from π^- happens to give an extremely accurate lower bound for the value function. Denoting $L(i, l, l + s, \pi^-)$ by $V^-(s)$ for $s \geq 0$, it can be shown by induction on n that

$$V^-(s) = \frac{ks}{2} + \frac{kT^2}{12s} - \frac{k}{2} \left(\frac{s}{2^n} + \frac{2^n T^2}{6s} \right) - n; 2^{n-1} s_{min} \leq s < 2^n s_{min},$$

for $n = 0, 1, 2, \dots$

An upper bound for V can also be established by noting that if

$$V(s) \leq \frac{ks}{2} + \frac{kT^2}{12s} - \log_2 s + d \text{ for all } s \in [T, r-T],$$

where $r \geq s_{min}$, then from the Bellman equation,

$$\begin{aligned} V(r) &= \sup_{T \leq u \leq r-T} \left\{ \frac{k}{r} u(r-u) + \frac{1}{r} [(r-u)V(r-u) + uV(u)] - \frac{kT^2}{12r} - 1 \right\} \\ &\leq \sup_{T \leq u \leq r-T} \left\{ \frac{k}{r} u(r-u) + \frac{1}{r} \left[(r-u) \frac{k(r-u)}{2} + \frac{ku^2}{2} - (r-u) \log_2 (r-u) \right. \right. \\ &\quad \left. \left. - u \log_2 u + du + d(r-u) + \frac{(r-u)kT^2}{12(r-u)} + \frac{ukT^2}{12u} \right] - \frac{kT^2}{12r} - 1 \right\} \\ &= \frac{kr}{2} + \frac{kT^2}{12r} - 1 + d - \frac{1}{r} \inf_{T \leq u \leq r-T} \{ (r-u) \log_2 (r-u) + u \log_2 u \}. \end{aligned}$$

Since $u \log_2 u$ is concave in u , the infimum is attained by $u = r/2$, and

$$V(r) \leq \frac{kr}{2} + \frac{kT^2}{12r} - \left[1 + \log_2 \left(\frac{r}{2} \right) \right] + d = \frac{kr}{2} + \frac{kT^2}{12r} - \log_2 r + d.$$

It therefore follows from an obvious contradiction argument that if

$$\frac{ks}{2} + \frac{kT^2}{12s} - \log_2 s + d \geq 0, \text{ for all } s \in [T, s_{min}],$$

then

$$V(s) \leq \frac{ks}{2} + \frac{kT^2}{12s} - \log_2 s + d, \text{ for all } s \geq T.$$

To find a value of d which makes this upper bound as accurate as possible, we first evaluate the zero, \bar{s} , of the bound's s derivative, giving

$$\bar{s} = \frac{\log_2 e}{k} + \sqrt{\frac{\log_2^2 e}{k^2} + \frac{T^2}{6}}.$$

It also happens that this derivative is positive for $s = s_{min}$, implying that $\bar{s} < s_{min}$. Since the upper bound is a concave function of s , our purpose is achieved by choosing d such that its value at \bar{s} is zero, or equivalently,

$$d = \log_2 \bar{s} - \frac{k\bar{s}}{2} - \frac{kT^2}{12\bar{s}}.$$

Denoting the resulting upper bound by $V^+(s)$, we have $V^+(\bar{s}) = 0 = V(\bar{s}) = V^-(\bar{s})$.

It follows from the formulas for V^+ and V^- that for $2^{n-1}s_{min} \leq s < 2^n s_{min}$, $n = 0, 1, 2, \dots$,

$$V^+(s) - V^-(s) = \frac{k}{2} \left(\frac{s}{2^n} - \bar{s} \right) + \frac{kT^2}{12} \left(\frac{2^n}{s} - \frac{1}{\bar{s}} \right) - \log_2 \left(\frac{s}{\bar{s}} \right) + n.$$

Therefore, $V^+(2s) - V^-(2s) = V^+(s) - V^-(s)$ if $s \geq s_{min}/2$. One important consequence of this result is that $V^+(2^n \bar{s}) = V^-(2^n \bar{s}) = V^-(2^n \bar{s})$ for any positive integer n , and hence that the search policy π^- is optimal for $s_0 = 2^n \bar{s}$, if $\bar{s} \geq s_{min}/2$. It is apparent from the formulas for \bar{s} and s_{min} that

$$\bar{s} < \frac{s_{min}}{2} \Rightarrow s_{min} = 2T$$

and therefore that

$$\bar{s} \geq \frac{s_{min}}{2} \Leftrightarrow T \leq \frac{12 \log_2 e}{5k}$$

Also, since for $2^{n-1}s_{min} \leq s < 2^n s_{min}$, $n = 0, 1, \dots$,

$$[V^+(s) - V^-(s)]'' = \frac{\log_2 e}{s^2} + \frac{2^n k T^2}{6s^3} > 0,$$

it follows in any case that $V^+(s) - V^-(s) \leq V^+(s_{min})$ if $s \geq s_{min}/2$.

The results for optimal search policies under the constraint $l_i + T \leq m_i \leq h_i - T$ can be summarized as:

1. The value function $V(s)$ is zero if and only if

$$s \leq \max \left\{ 2T, \frac{2}{k} + \sqrt{\left(\frac{2}{k}\right)^2 + \frac{T^2}{3}} \right\} \triangleq s_{min}.$$

2. It is optimal to continue the search if and only if $s \geq s_{min}$.

3. If $s \in [2^{n-1}s_{min}, 2^n s_{min}]$, $n = 0, 1, 2, \dots$,

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

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

4. $V^+(s) - V^-(s) \leq V^+(s_{min})$ for $s \geq s_{min}/2$.
5. $V^+(2^n \bar{s}_{min}) = V^-(2^n \bar{s}_{min}) = V^-(2^n \bar{s}_{min})$ for any positive integer n if

$$T \leq \frac{12 \log_2 e}{5k}.$$

The policy π^- (where $m_i = h_i/2 + l_i/2$) is optimal if $s_0 = 2^n \bar{s}_{min}$.

6. Cases exist for which $T < (12 \log_2 e)/5k$ and π^- is not optimal.

It is also interesting to note that if $T = 0$, the maximum error of the bounds for $V(s)$ for $s \geq s_{min}/2$ is given by the expression

$$V^+(s) - V^-(s) \leq V^+(s_{min}) - V^-(s_{min}) = \log_2 \log_2 e - \log_2 e + 1 = 0.0862,$$

which is independent of k .

REMOVAL OF POLICY RESTRICTION

The derivation of the preceding results depended on an artificial restriction imposed on the class of admissible search policies. This section shows that this restriction is superfluous if $2Tk < 1$, in the sense that for every policy not satisfying the restriction $l_i + T \leq m_i \leq m_i - T$ there is a policy which does satisfy it and for which the expected return is greater. This condition obviously implies that $T \leq (12 \log_2 e)/5k$ (or equivalently, $\bar{s} \geq s_{min}/2$), which is the case of major interest.

Assuming the contrary, the principle of optimality implies the existence of a case for which $2kT < 1$, a policy and a realization such that $m_j < l_j + T$ ($m_j > h_j - T$ is basically a symmetrical case) for some epoch j , and such that the conditional expected future return at that point is greater than that of any policy which terminates then or for which $l_j + T \leq m_j \leq h_j - T$. Let s_0 , T , and k be fixed such that this possibility exists and let A be the set of all such triples. Define the set B as

$$B \triangleq \{x \in R: (x, T, k) \in A\}.$$

Let σ be an element of B , such that

$$\sigma < \inf B + T.$$

Let π denote a policy for which this possibility exists for the above values of T and k and for $s_0 = \sigma$. Consider any such realization and denote by i the index of the first epoch for which $m_i < l_i + T$. Denote $m_i - l_i$ by u_i .

Since termination is not optimal at epoch i and $2kT < 1$, $s_i > T$, where $s_i \triangleq h_i - l_i$. Since, in addition, $l_j + T < m_j < h_j - T$ for all $j < i$, the conditional density of b at epoch i is symmetric trapezoidal with upper and lower midpoints h_i and l_i .

Given $y_i > b$ (and u_i), the future return at epoch i under these circumstances is less than

$$\max \{-1, k(T + u_i) - 2\}$$

for any realization and policy. Therefore, the conditional expected future return under π is also less than this value, which is negative since $u_i < T$. Given $y_i \leq b$, this conditional expected future return is less than

$$k(u_i + T) - 1 + \epsilon [\text{future return at epoch } (i + 1) \text{ using } \pi \text{ given } y_i \leq b].$$

Now let $z = (1/2) - (1/2)\text{sgn}(b - \alpha + \zeta)$, where $\alpha = \max(T, u_i + T)$ and ζ is a random variable independent of b and the y 's, with rectangular density of median zero and width T . (A knowledge of z at epoch $(i + 1)$ is "free" extra information.) The preceding expectation is less than or equal to

$$\begin{aligned} & \epsilon \{ \text{future return at epoch } (i + 1) \text{ under } \pi' \text{ given } z = 1 \cdot \text{Prob } \{z = 1 \text{ given } y_i \leq b\} \\ & + \epsilon \{ \text{future return under } \pi' \text{ given } z = 0 \} \cdot \text{Prob } \{z = 0 \text{ given } y_i \leq b\}, \end{aligned}$$

where π' is an optimal (or ϵ -optimal) policy for the altered search problem in which z is also known at epoch $(i + 1)$. The first expectation is bounded by $V(s_i - \alpha)$ because $s_i - \alpha < \sigma$ and the conditional density of b at epoch $(i + 1)$ given $z = 1$ is symmetric trapezoidal, or because $s_i - \alpha < T$. The future return in the second case is bounded by $2kT - 1$ in any event; hence, so is the second term.

Denoting the conditional expected future return at epoch i under π by F , it follows that

$$F < 0 + k(u_i + T) - 1 + V(s_i - \alpha) + 2kT - 1$$

by bounding all probabilities by 1. Since the existence of a policy constraint cannot increase the maximum expected return,

$$V(s_i) < F.$$

Therefore, since $V(s_i - \alpha) < V(s_i)$,

$$0 < 3kT + ku_i - 2 < 4kT - 2.$$

This contradicts the assumption of $2kT < 1$, which verifies the desired result.

DISCUSSION

The type of search problem described in this report can be formulated in the usual context of stochastic optimal control problems, as outlined in the section "Another Formulation." As such, they are discrete-time control problems with piecewise linear dynamics and criterion, nonlinear measurements, nonadditive rectangular measurement and process noises, and unspecified terminal time. Both the linear form of the gain function G and the rectangular form of the measurement noise distributions were essential in deriving specific results here. Adding a constant to $G(x)$ for $x \leq b$ also changes the fundamental character of the problem.

The optimal state estimator can be implemented by recursively updating l_i and h_i from the measurement $\text{sgn}(b - y_i) \triangleq z_i$ as

$$l_{i+1} = \max \{ l_i, m_i z_i \}$$

and

$$h_{i+1} = \max \{ h_i z_i, m_i \}.$$

The search policy π^- is not always optimal, but it is for arbitrarily large values of s_0 , the initial range of uncertainty of the boundary location, if $2kT \leq 1$. This fact indicates that π^- is a good search policy in practice for such cases, even when it is not strictly optimal. Also, the π^- stopping rule is always optimal.

Except for the stopping rule, π^- is a certainty equivalent policy. Its search points are the optimal search points for the sequence of deterministic problems in which all random variables are replaced by their current conditional means.

The practical significance of this work is that it adds to the repertory of decision process models for which there is specific knowledge about optimal performance and policies. This type of model can be useful, for example, in the analysis of mining operations. These results are of little practical interest if $2kT > 1$, however, because they depend on an unrealistic policy restriction in this case.

REFERENCES

1. C. Derman and E. Ignall, "On Getting Close to but not Beyond a Boundary," *J. Math. Anal. Appl.* **28**, 128-143 (1969).
2. R. L. Stratonovich, "On the Theory of Optimal Control: Sufficient Coordinates," *Automation and Remote Control* **23**, 847-854 (1963).
3. R. Bellman, *Dynamic Programming*, Princeton U. Press, Princeton, New Jersey, 1957.

Security Classification		DOCUMENT CONTROL DATA - R & D	
<i>(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)</i>			
1. ORIGINATING ACTIVITY <i>(Corporate author)</i> Naval Research Laboratory Washington, D.C. 20390		2a. REPORT SECURITY CLASSIFICATION Unclassified	
		2b. GROUP	
3. REPORT TITLE ON A CLASS OF OPTIMAL SEARCH PROBLEMS			
4. DESCRIPTIVE NOTES <i>(Type of report and inclusive dates)</i> Final report on one phase of the problem; work is continuing on other phases.			
5. AUTHOR(S) <i>(First name, middle initial, last name)</i> W. W. Willman			
6. REPORT DATE September 28, 1971		7a. TOTAL NO. OF PAGES 15	7b. NO. OF REFS 3
8a. CONTRACT OR GRANT NO. NRL Problem B01-10		9a. ORIGINATOR'S REPORT NUMBER(S) NRL Report 7309	
b. PROJECT NO. RR 003-02-41-6152		9b. OTHER REPORT NO(S) <i>(Any other numbers that may be assigned this report)</i> Operations Research Group Report 71-4	
c.			
d.			
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 policies 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 policies are described explicitly for a large subclass of these problems. 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 Dynamic programming Optimal control Stopping rule						