OPERATIONS RESEARCH
Vol. 52, No. 3, May–June 2004, pp. 454–463
ISSN 0030-364X | EISSN 1526-5463 | 04 | 5203 | 0454

DOI 10.1287/opre.1030.0086

Shoot-Look-Shoot: A Review and Extension

Kevin Glazebrook
School of Management, University of Edinburgh, Edinburgh EH8 9JY, United Kingdom, kevin.glazebrook@ed.ac.uk

Alan Washburn
Operations Research Department, Naval Postgraduate School, Monterey, California 93943, washburn@nps.navy.mil

We consider the optimal use of information in shooting at a collection of targets, generally with the object of maximizing the average number (or value) of targets killed. The shooting problem is viewed as a Markov decision process, and the modal solution technique is stochastic dynamic programming. Information obtained about target status may or may not be perfect, and there may or may not be constraints on the number of shots. Previous results are reviewed, and some new results are obtained.

Subject classifications: decision analysis: sequential; dynamic programming/optimal control: Markov finite state, Markov infinite state; military: targeting; probability: Markov processes, stochastic model applications.
Area of review: Military.
History: Received September 2002; accepted February 2003.

Introduction

The subject of this paper is the manner in which information that is gradually acquired about the status of a target ought to influence the process of shooting at it. It is a military subject that has become increasingly important with the advent of long-range, accurate, but expensive, weapons. For example, suppose that the single-shot probability of killing a target is 0.9, but that the target is so important that even a 0.1 survival probability is not acceptable. One could shoot independently twice at the target, thereby achieving a kill probability of 0.99 at the expense of two shots. Alternatively, one could shoot at the target, look to see if it has been killed, and then shoot again, if necessary. The kill probability is still 0.99 with this shoot-look-shoot policy, but the average expenditure of shots is only 1.1—just over half of the two-shot expenditure. The payoff for acquiring and using information optimally can be significant.

Use of the term "shoot-look-shoot" sometimes implies that only a single look is contemplated, but not here. In general, we will consider multiple error-prone looks and shots in stages. We will consistently use the term "kill" because of the problem's military heritage, but terms such as "damage," "identify," or "find" could also be substituted. The essential requirements are that the "target" be in one of two states, one of which is desirable and the other not, and that the marksman have a succession of opportunities for altering and discovering the target's state.

The general problem considered here involves a matrix $P = (P_{ij})$, where P_{ij} is the probability that a shot of type i is effective against target j. All attempts to kill a target are assumed to be independent, as in the two-shot example above. There may also be data associated with the discovery process. Because the targets are not always identical, it is necessary to attribute a weight or value v_j to target j, and the firing problem usually takes the form of an optimization where the object is to maximize the expected value of the total target value killed.

The expensiveness of shots can take three forms: one (§1) where the number of shots of type i is constrained to not exceed some given level, one (§2) where shots are available in unlimited quantities, as long as they are paid for, and a third form (§3) where shots are constrained as well as costly. The *time* at which a target is killed is usually irrelevant; §4 is an exception where rewards are discounted.

1. The Constrained Case

We assume in this section that the problem is to assign a given set of shots to a given set of targets.

1.1. Perfect Information, Infinite Time Horizon

Assume that the results of each shot are revealed immediately after each shot is made, and that shots can be made one at a time because of the infinite time horizon. The import of these assumptions is that every shot is made with exact knowledge of the state of the target set. We take the state of the firing process to be (S, T), where S and T are the sets of remaining shots and live targets, respectively. The object is to find $V(S, T)$, the largest amount of target value that can be killed with all remaining shots, on the average, together with the shot $i \in S$ that should be taken against target $j \in T$. If either S or T is empty, then of course $V(S, T) = 0$.

$V(S, T)$ can be found by a dynamic programming recursion. Because there are only two possibilities for each shot, and because all shots are independent, by the conditional

expectation theorem we have

$$V(S,T) = \max_{i\in S,\, j\in T} \{P_{ij}(v_j + V(S-i, T-j)) + (1-P_{ij})V(S-i,T)\}. \quad (1)$$

If S has m shots left in it, and if T has more than m targets, then (1) will have to be evaluated 2^m times in the process of computing $V(S,T)$. Computational difficulty can be expected as m becomes large.

If $v_j = v$ and $P_{ij} = p$ for all i and j, then (1) simplifies considerably because the sets S and T merely need to be counted. Let s and t be the numbers of shots and targets, and let X be a binomial random variable with parameters s and p. X can be interpreted as the number of effective shots in the set S. The number of kills will be X unless the marksman runs out of targets. Therefore $V(S,T) = E(\min(X,t))$, a relatively simple computation. Anderson (1989) notes that the same formula applies as long as P_{ij} does not depend on j; X still has the same interpretation, although it is no longer binomial. See also Przemieniecki (1990).

Because (1) is computationally challenging for large problems, we next develop some bounds on $V(S,T)$.

A simple lower bound $V\_(S,T)$ can be constructed by computing the optimal pair $(i^*, j^*) = \arg\max_{i\in S,\, j\in T} v_j P_{ij}$. This is the "myopic" firing policy—every shot is taken without regard to future consequences. Equation (1) (with (i^*, j^*) substituted for (i,j) and $V\_()$ replacing $V()$ on both sides) must still be employed to evaluate $V\_(S,T)$, so exact evaluation is still difficult as m becomes large. However, the myopic policy is trivial to implement because knowledge of $V\_(S,T)$ is not needed.

The myopic policy is not always optimal. For a counterexample, consider two shots and two targets, with $P = \left[\begin{smallmatrix} 1 & 0.9 \\ 0.9 & 0 \end{smallmatrix}\right]$ and $v = (1,1)$. The myopic policy will assign shot 1 to target 1, after which shot 2 is useless, and the total score is 1. The optimal policy is to assign shot 2 to target 1, and then shot 1 to target 1 in case of failure, or shot 1 to target 2 in case of success. The optimized total score is $0.9(1+0.9) + 0.1(0+1) = 1.81$—a substantial improvement over the myopic score.

The reader may wish to download an Excel workbook *SLS.xls* from *http://diana.gl.nps.navy.mil/~washburn/*. On sheet "Optimal," *SLS.xls* computes solutions for small problems according to the method described above.

$V(S,T)$ can also be usefully bounded from above. See §1.3 below.

Although the time horizon in this section has been assumed to be infinite, it should be obvious that enough time to make m shots is all that is required. Even smaller amounts of time may suffice. Call a shot "unitary" if it has a zero kill probability against all targets but one. Because such shots have no flexibility, they can all be assigned at once without jeopardizing the maximum score $V(S,T)$. There may also be less obvious opportunities for shortening the time horizon. Let $H(S,T)$ be the shortest horizon within which there still exists a shooting policy that will guarantee $V(S,T)$. Computation of $H(S,T)$ is itself an interesting topic, but we will not pursue it further here.

1.2. Perfect Information, Finite Horizon, Identical Shots and Targets

The analysis in §1.1 is comparatively simple because every shot is taken in perfect knowledge of the status of the target set. This is no longer true if either time is constrained or if information is imperfect. One reason for time to be constrained is that the "targets" might actually be incoming missiles, in which case the problem is one of self-defense. The general problem appears to be difficult, although various specializations such as those in this section can still be solved. We assume that all shots and targets are identical, so that the sets S and T can be replaced by simple counts of the number of shots and targets remaining.

Suppose that $P_{ij} = 1-q$ for all i, j, where q is the miss probability for all shots, and that $v_j = 1$ for all j. Because all targets are identical, each salvo should treat the remaining targets as evenly as possible, and the firing question reduces to determining how many shots to spend in each salvo. Let random variable X be the number of targets killed out of t when x shots are allocated, and let $F_n(s,t)$ be the maximum expected number that can be killed with s shots, t targets, and n salvos remaining. Then $F_0(s,t) = 0$, and for $n > 0$ we have the dynamic programming recursion

$$F_n(s,t) = \max_{0<x\leqslant s} E(X + F_{n-1}(s-x, t-X)). \quad (2)$$

Calculating the expected value in (2) requires a distribution for X. The distribution is binomial if x is less than t or an integer multiple of t, or otherwise the convolution of two binomial distributions. In either case, the numerical solution of (2) is not difficult as long as s and t are not too large.

Suppose that the "targets" are actually missiles that are attacking home base, with the shots being defending interceptors. The number of salvos is limited because of the speed of the attackers. Let ρ be the probability that any given attacker will kill home base if not intercepted, and let $G_n(s,t)$ be the largest home base survival probability when there are s shots, t targets, and n salvos remaining. Then $G_0(s,t) = (1-\rho)^t$, and for $n > 0$ we have

$$G_n(s,t) = \max_{0<x\leqslant s} E(G_{n-1}(s-x, t-X)). \quad (3)$$

Computational issues are the same as with (2). Eckler and Burr (1972) allude to (3) and give some results for the case $n = 2$. Wilkening (1999) considers the case where $n = 2$ and $\rho = 1$ in the context of ballistic missile defense. A reader who searches the Internet for "shoot-look-shoot" will discover that this scenario is frequently referred to.

The aforementioned workbook *SLS.xls* includes an implementation of (3) on sheet "DynProg" for a three-stage

problem ($n = 3$). The optimal policy with only one stage remaining is to use all remaining interceptors, but with more stages the optimal policy uses fewer interceptors in order to avoid wasting them on dead targets. Of course, at least one shot at every stage is always made at every surviving target, as long as sufficient shots remain to do so. The spreadsheet also permits a minor generalization: The interceptor kill probability can depend on the number of stages remaining.

1.3. Imperfect Information

It is possible that when information is imperfect, shots will be made against targets that are already dead. The probability that shot i is effective against target j remains P_{ij}, but an effective shot will kill its target only if the target is alive when the shot is taken.

We first develop an upper bound on the best target value killed that is valid regardless of the information available to the marksman. Fixed sets of shots and targets are, as usual, given. Define a collection of indicator random variables that can be associated with any firing policy that assigns shots to individual targets:

$Y_{ij} = 1$ if target j is killed by shot i (for each j, at most one of these can be 1),

$Y_j = 1$ if target j is killed, and

$X_{ij} = 1$ if shot i is assigned to target j (for each i, at most one of these can be 1).

The firing policy will induce many correlations between these random variables, so independence assumptions among them are not appropriate, but the collection is still useful for formulating the problem of finding the optimal policy. We first note that $Y_j = \sum_i Y_{ij}$, and that the total value killed is $Z = \sum_j v_j Y_j$. The problem (call it P1) of finding the optimal policy can therefore be posed as maximizing z, the expected value of Z, subject to the following constraints:

(a) $\sum_j X_{ij} \leqslant 1$ for all i with certainty;
(b) $Y_j = \sum_i Y_{ij}$ for all j with certainty;
(c) $Y_j \leqslant 1$ for all j with certainty; and
(d) other constraints.

The other constraints in P1 include the crucial relationship between X_{ij} and Y_{ij}, the essence of the firing policy. For example, the (probably foolish) policy of ignoring all information and simply making $X_{i1} = 1$ for all i would probably kill target 1, but would also result in $Y_j = 0$ for $j > 1$. There are potentially an astronomical number of firing policies, because the decision about what to do next can depend in many ways on the information available. Nonetheless, regardless of the policy employed, we assume that $E(Y_{ij}) \leqslant P_{ij}E(X_{ij})$, with strict inequality being possible on account of the possibility that target j is already dead when weapon i attacks it. Now, using lowercase letters for expected values of random variables, we can construct a relaxation of P1 that we name P2:

$$\text{maximize } z = \sum_j v_j y_j,$$

subject to

(a) $\sum_j x_{ij} \leqslant 1$ for all i;
(b) $y_j \leqslant \sum_i x_{ij}P_{ij}$ for all j;
(c) $y_j \leqslant 1$ for all j; and
(d) all variables nonnegative.

P2 is a relaxation of P1 because a relationship that is true with certainty will also be true on the average, because sums and expected values can be interchanged, because $E(Y_{ij}) \leqslant x_{ij}P_{ij}$, and because the other constraints of P1 have simply been omitted. P2 is a simple linear program that provides an upper bound on what is achievable with any shoot-look-shoot policy. P2 has a direct interpretation where x_{ij} is the probability of using shot i on target j, and y_j is the probability that target j is killed. In P2, (a) requires that shot i not be used more than once, (b) requires that the effect of each shot not exceed P_{ij}, and (c) requires that target j be killed at most once.

If b_i shots of type i are actually identical, then constraints (a) of P2 can be changed to $\sum_j x_{ij} \leqslant b_i$, a simple consequence of collecting terms with identical coefficients in P2. In that case the interpretation of x_{ij} is "average number of shots of type i used on target j."

Because perfection is a special case of imperfection, the P2 upper bound also applies to problems of the type considered in §1.1. In fact, the aforementioned workbook *SLS.xls* also computes the upper bound for problems defined on sheet "Optimal," with the upper bound computations being carried out on sheet "LP_Upper" when the command button on sheet "Optimal" is pressed. Using it, the reader can verify that the upper bound for the small example introduced in §1.1 is actually exact, and that the upper bound is usually sharp for problems not designed to make it look bad. For one of the latter, consider a problem with one target and two shots, each of which will kill the target with probability 0.5. The optimal, indeed the only, allocation of weapons to targets produces a kill probability of 0.75, while the upper bound is 1.

We next turn to the construction of optimal firing policies in specific circumstances where information is imperfect.

Manor and Kress (1997) consider a firing problem in which all shots are identical, with each shot having kill probability p_j against target j. If target j is not killed, there is no feedback to the marksman. If the target is killed, that fact is confirmed with probability q_j, or otherwise there is no feedback. Information would be perfect if $q_j = 1$, because the state of the target could be inferred from feedback or lack thereof. If $q_j < 1$, however, lack of feedback is possible from live targets, as well as dead ones, so the marksman will not be certain of the state of a target unless he has a confirmed kill. Manor and Kress argue that such a model applies to weapon systems such as fiber-optic guided missiles (FOG-M). The marksman's object is to kill as many targets as possible, on the average, with a fixed inventory of shots. Manor and Kress demonstrate that the "greedy" policy of always shooting at the target for which the immediate gain is largest is actually optimal.

The targets need not all be identical, but, if they are, the policy reduces to shooting at the least-shot-at target among those not known to be dead. This policy results in a lot of switching from one target to another, a regrettable tendency when speed is important. Aviv and Kress (1997) explore other policies that are almost as good but that do not switch targets so frequently.

One might expect the greedy policy to be optimal under more general status reporting. Suppose then that there are $s > 0$ shots and $t > 0$ targets, and let random variable X_{ik} indicate whether target i is alive at time k ($X_{ik} = 1$) or not ($X_{ik} = 0$). Let $\pi_{ik} \equiv P(X_{ik} = 1)$, with π_{i0} known, $i = 1, \ldots, t$. Shots are made one at a time. The effect of a shot at target i at time $k-1$ is to kill it with probability $1-q$, if it is not already dead, and to produce a report Y_k in some countable set. It is assumed that Y_k depends only on X_{ik}, and that $P(Y_k = y \mid X_{ik} = x)$ is known and independent of i and k for all y, and for $x = 0$ or 1. The effect of these assumptions is that, if the shot is aimed at target i and $Y_k = y_k$, by Bayes' theorem,

$$\pi_{ik} = \frac{q\pi_{i,k-1}P(Y_k = y_k \mid X_{ik} = 1)}{q\pi_{i,k-1}P(Y_k = y_k \mid X_{ik} = 1) + (1 - q\pi_{i,k-1})P(Y_k = y_k \mid X_{ik} = 0)}. \quad (4)$$

Because π_{ik} is a conditional probability, it is not—and need not be—defined if the probability of the conditioning event (denominator of (4)) is 0. There is no effect on other targets; that is, $\pi_{ik} = \pi_{i,k-1}$ if target i is not shot at.

It is sometimes assumed that Y_k can have only two values, typically "Live" and "Dead" reports. However, Y_k might also be some physical measurement such as the temperature of the target or the output of a digital filter that measures the extent to which the target optically resembles the thing that was initially shot at. As long as the conditional probabilities in (4) are known for all (i, k), Bayes' theorem applies.

Let a policy be called myopic if, at every time $k > 0$, it chooses a target i for which π_{ik} is largest.

THEOREM 1. *Any myopic firing policy will maximize the expected number of targets killed in total.*

PROOF. The proof of this theorem is long, so we defer it to the appendix.

If M is the total number of targets killed using a myopic policy, and N is the total number of targets killed under an arbitrary policy, then one might expect the stronger result that M stochastically dominates N. However, this is not the case. A counterexample is $t = 2$, $s = 2$, and $q = 0.5$, with no information between shots and with the first target being alive and the second very likely dead. The myopic policy will fire both shots at the first target, thus making it impossible to kill both of them, whereas the policy of splitting the two shots has at least a slight chance of killing both. Therefore, $P(M \leqslant 1) > P(N \leqslant 1)$; that is, M does not stochastically dominate N.

It is also not true that only myopic policies can maximize the expected number of targets killed. A counterexample is $t = 3$, $s = 2$ and $q = 0.5$, with no information between shots and both targets alive. The policy of shooting twice at target 1 then once at target 2 is optimal, but not myopic.

There is no reason to expect myopic policies to be optimal in more general circumstances where the targets or shots differ, nor is there any known way of efficiently computing what the optimal policy is in those circumstances. Oddly, the situation can be simpler if the number of shots is random rather than fixed. If the number of shots has a geometric distribution, the firing problem may be *indexable* (see §4).

2. The Unconstrained Case

We now assume that shots are available in unlimited quantities, as long as they are paid for at the cost of c_i for each shot of type i. We will assume that looks are also expensive. The idea that looks are costly or in short supply was not incorporated in §1 but is actually an important part of some shoot-look-shoot problems.

The great analytical advantage of not having explicit constraints on shots and looks is that the firing processes at the several targets *decouple*. As a result, problems with many different targets are much easier to solve than in the constrained case. Therefore, we will suppress reference to the target subscript j in this section.

The target value v and the various costs must of course be measured in commensurate units. This may seem to make the model unwieldy in a practical sense, because target values are often just relative valuations of importance, while shot costs are monetary. The prospects for application are not as bad as one might suppose. Section 2.4 includes a discussion of this issue.

2.1. Perfect Information, Infinite Horizon

Let q_i be the miss probability for one shot of type i, and assume each look reveals correctly whether the target is dead or alive. To avoid trivialities we assume that the look cost d is positive. The marksman must pay the costs of all looks and shots and must also pay the target's value v if the target is still alive after all shots have finished. The marksman's objective is to minimize the expected sum of all costs.

In general, the marksman will not know whether the target is dead or alive, so we let p be the probability of the target being alive, a state in the interval [0,1]. To this we add, for convenience, one additional state R, in which the marksman has retired and can take no further action. Except in state R, the marksman has his choice of looking or shooting at the target in various ways, or of choosing the action E (for "End"), which costs nothing but sends the state to R.

Using the code that L stands for looking and that positive integer i stands for a shot of type i, a typical firing policy might be $1123L44L11E$. It should be understood that firing ends immediately after any look that reveals that the

target is dead. Thus, after shooting four times, the marksman looks, shoots two more times if the target is still alive, looks again, shoots two more times if the target is still alive, and then quits. Any policy can be encoded in this manner and then evaluated; for example, the miss probability for the named policy is $q_1^4 q_2 q_3 q_4^2$. It should be obvious, however, that there is no hope of determining the optimal policy by exhaustion. Instead, we employ the theory of partially observable Markov decision processes (Monahan 1982) to first determine the structure of the optimal policy.

A POMDP requires the specification of two functions: $c(s, a)$ and $P(s|s, a)$. The first is the immediate (average) cost of taking action a when the target is in state s, given in the present case by

$$c(s,a)=\begin{cases}0 & \text{if } s=R,\\ c_i & \text{if } s=p \text{ and } a=i,\\ d & \text{if } s=p \text{ and } a=L,\\ pv & \text{if } s=p \text{ and } a=E.\end{cases}$$

The second function gives the probability that s' will be the next state, given that the current state and action are s and a. It is given by

$$P(s'\,|\,s,a)=\begin{cases}1 & \text{if } s'=R \text{ and either } s=R \text{ or } a=E,\\ 1 & \text{if } s'=pq_i, s=p \text{ and } a=i,\\ p & \text{if } s'=1, s=p \text{ and } a=L,\\ 1-p & \text{if } s'=0, s=p \text{ and } a=L.\end{cases}$$

The last two lines correspond to looking, with the two possibilities after a look being that the next state is 1 or 0.

Let $V(s)$ be the minimal total cost over all stages, given that the initial state is s. Clearly, $V(R)=0$, because subsequent action is impossible from state R. Otherwise, for p in the interval [0,1], the function $V(p)$ is concave and satisfies the functional equation

$$V(p)=\min\left\{vp;\, d+pV(1);\, \min_{i>0}(c_i+V(pq_i))\right\}, \tag{5}$$

where the three expressions correspond to stopping (action E), looking (action L), or shooting with shot i (Strauch 1966, Theorem 9.1). Furthermore, there is an optimal stationary policy that, in state p, chooses the action corresponding to the minimal term.

It would be equivalent to let $U(p)=vp-V(p)$, and deal with the functional equation

$$U(p)=\max\left\{0;\, -d+pU(1);\, \max_i(-c_i+pv(1-q_i)+U(pq_i))\right\}. \tag{6}$$

$U(p)$ can be described as the marksman's gain relative to quitting, with the gain for choosing E being 0 and the immediate gain from shooting being the expected target value killed minus the cost of the shot. The sum of $U(p)$ and $V(p)$ is in all cases vp. While the two formulations are equivalent, we prefer (5).

We have the following structural theorem.

THEOREM 2. *Action E is optimal in an interval $[0, p_1]$, with $p_1 \geqslant 0$. Action L is optimal over a possibly empty interval $[p_2, p_3]$, with $p_1 \leqslant p_2$.*

PROOF. Action E is optimal at 0, so it suffices to show that the optimality set is convex. Suppose that E is also optimal at p_1, and let $U(p)=vp-V(p)$. The function $U(p)$ is convex because $V(p)$ is concave, and $U(0)=U(p_1)=0$. Therefore, $U(p)\leqslant 0$ on the interval $[0, p_1]$. But $U(p)\geqslant 0$ for all p by the definition of $V(p)$, so actually $U(p)=0$ for $0\leqslant p\leqslant p_1$; that is, action E is optimal over some convex set. The proof for action L (or any action whose penalty is a linear function of p) is similar. The only state p^* where E and L have the same penalty is the one where $vp^*=d+p^*V(1)$. The optimality sets for E and L therefore cannot overlap except perhaps at a single point, so p_2 cannot be smaller than p_1. □

A stationary policy will always choose the same action when in state 1, which happens after every look unless the target is revealed to be dead. Therefore, there is an optimal stationary policy in state 1 that repeats itself after every look unless the target is revealed to be dead, unlike the nonstationary example given earlier. If such a policy begins $1123L$, then it must continue $1123L1123L1123L\ldots$ indefinitely until the target is finally revealed to be dead. Let x be a shot sequence of the form $(i_1, i_2, \ldots, i_k)$, where i_j is a shot type for $1\leqslant j\leqslant k$. Because the order of shots is immaterial in x, we will list them in nondecreasing order. Adopt the notation xL for stationary policies that shoot x, look, and repeat the cycle until the target is killed. Another possibility is that the policy shoots x and then quits, for which we adopt the notation xE. The final possibility is that the policy simply chooses E at the first opportunity, a special case of xE where $k=0$. There are no other possibilities, so we have Theorem 3.

THEOREM 3. *Let $c_x \equiv \sum_{j=1}^k c_{i_j}$ and $q_x \equiv \prod_{j=1}^k q_{i_j}$, with $c_x \equiv 0$ and $q_x \equiv 1$ if $k=0$. Then*

$$V(1)=\min\left\{\min_x\left(\frac{d+c_x}{1-q_x}\right);\, \min_x(vq_x+c_x)\right\}. \tag{7}$$

PROOF. Stationary policies of the xL type lead to the first term. The cost $d+c_x$ is paid a geometric number of times, $1/(1-q_x)$ on the average, before the target is finally killed. Stationary policies of type xE lead to the second term, with $k=0$ corresponding to ending immediately. No other policies need be considered because there is guaranteed to be an optimal policy that is either xL or xE. □

A similar expression could be proved for $V(p)$ in general, but Theorem 3 will suffice because the target is normally assumed to start in the live state.

Even though Theorem 3 provides a comparatively simple way to determine $V(1)$, the amount of computation might still be significant if there were many shot types. The following corollary concerns a lower bound that is easier to compute.

COROLLARY 1. *Let i^* be the shot type that minimizes $c_i/(-\ln(q_i))$, let $\alpha = -\ln(q_{i^*})$, and let $c = c_{i^*}$. Then*

$$V(1) \geqslant \min\left\{\min_{u>0}\left(\frac{d+cu}{1-\exp(-\alpha u)}\right); \min_{u\geqslant 0}(v\exp(-\alpha u)+cu)\right\}.$$

PROOF. Let $\alpha_i = -\ln(q_i)$, and let u_i be the number of shots of type i in the shot sequence x. Then, $q_x = \exp(-\sum_i \alpha_i u_i)$ and $c_x = \sum_i c_i u_i$, where both sums extend over all shot types. By relaxing the minimization (7) to permit noninteger values for u_i, we obtain the bound

$$V(1) \geqslant \min\left\{\min\left(\frac{d+\sum_i c_i u_i}{1-\exp(-\sum_i \alpha_i u_i)}\right); \min\left(v\exp\left(-\sum_i \alpha_i u_i\right)+\sum_i c_i u_i\right)\right\},$$

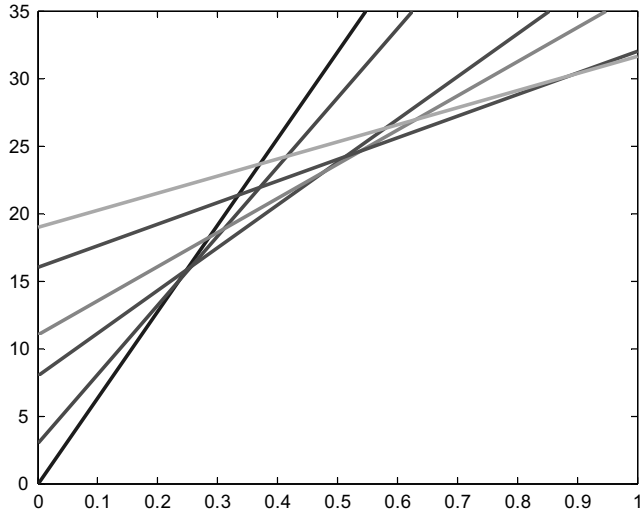
where both inner minimizations are with respect to the vector (u_i). If $u_i > 0$ for $i \neq i^*$, then increase u_{i^*} by $u_i(c_i/c_{i^*})$ and set u_i to 0. The resulting change will leave c_x unchanged while not increasing q_x, so it suffices to consider only the single weapon type i^*. □

COROLLARY 2. *Let $r = \min_i c_i/(1-q_i)$ and suppose $d > 0$. If $r > v$, then the optimal action in state 1 is E. If $r < v$, then the optimal action in state 1 is to shoot.*

PROOF. We first note by inspection of (5) that action L cannot be optimal in state 1, regardless of r. Let $p_i = 1 - q_i$. More generally, let $p_x \equiv \sum_{j=1}^{k} p_{i_j}$ for any shot sequence x, and note that $p_x + q_x \geqslant 1$ for all x, as can be easily proved by induction on the length of x. By assumption, $p_i r \leqslant c_i$ for all shot types i, and therefore, $p_x r \leqslant c_x$ for all shot sequences x. It follows that $(1-q_x)r \leqslant c_x$ for all x. Suppose xE is an optimal stationary policy for some nonnull shot sequence x. Then, $v \geqslant vq_x + c_x$, hence, $v \geqslant c_x/(1-q_x)$, and hence, $v \geqslant r$. If actually $v < r$, the contradiction shows that x can only be null and establishes that the optimal action in state 1 is E. If $v > r$, then there is some shot type i for which $vq_i + c_i < v$, and therefore E cannot be optimal. □

EXAMPLE. Suppose $v = 64$, $d = 8$, $c_1 = 8$, $c_2 = 3$, $q_1 = 0.5$, $q_2 = 0.8$. It can be shown by exhaustion that the optimal policy when $p = 1$ is $12L$. The target is always killed, and the average cost of doing so is $V(1) = (8+3+8)/(1-(0.5)(0.8)) = 31.67$. The minimal cost with a pure shot type is 32, achieved with any of policy $1L$, $11L$, $11E$, or $111E$. This shows that it is not in general possible to confine attention to "pure" shooting policies. The lower bound is 30.912, obtained by shooting 1.421 times with shot $i^* = 1$, looking, and repeating the action until the target is killed. According to Corollary 2, the optimal action in state 1 will be to shoot (either with shot 1 or shot 2) as long as $v > 15$.

Figure 1. Penalties for six policies (E, $2E$, $L12L$, $2L12L$, $11E$, $12L$) in the example problem of §2.1 are shown by six lines, listed in increasing order of the intersection with the y-axis.



Note. The function $V(p)$ is the lower envelope (concave hull).

Figure 1 shows the penalties for each of six stationary policies as linear functions of the state p. All other policies are dominated by one of these six. The first character of each policy shows the (first) optimal action in that state. Note that there is only one policy that begins with E and one that begins with L, consistent with Theorem 2.

Solving the equation $64p = 8 + pV(1)$ for p, we find that the E and L actions are tied in (5) when $p = 0.247$, at which point both penalties are 15.84. However, the optimal policy in that state is $2E$, which produces a smaller penalty of $3 + 0.8(0.247)64 = 15.65$. In fact, action 2 is the optimal action in the interval [0.234, 0.256], which separates the interval where E is the optimal action from the interval where L is the optimal action. Thus, we have a counterintuitive situation where increasing the probability of being alive can cause the optimal action to change from shooting to looking. In other words, the set of states where shooting is optimal is not convex. This phenomenon is also noted by Ross (1971) in a different Markov decision process. States in the interval [0.234, 0.256]—the interval where $2E$ is the optimal policy—will never arise if the initial state is 1 and the optimal policy $12L$ is followed.

2.2. Imperfect Information, Infinite Horizon

The looks and shots of §2.1 can each be regarded as special cases of a generalized shot. Generalized shot i (hereafter, simply shot i) has the effect of first killing the target with probability $(1-q_i)$, and then providing a report about the new state of the target. The effect of such a shot is to first change the state from p to pq_i, and then to provide a report about the transformed target state. Let $b_i(pq_i, k)$ be the posterior probability that the target is alive, given that

the prior probability (before the shot) is p and that report k is generated. We assume this function to be known, typically through an application of Bayes' theorem. If shot i is selected as the next action, the resulting report will be a random variable K with a known distribution, and the next state will be a random variable $b_i(pq_i, K)$. Shots with no information output can be modeled by providing only one possibility for K (in which case $b_i(pq_i, K) = pq_i$), and shots that provide only information ("sensors") can be modeled by setting $q_i = 1$.

Let $V(p)$ be the marksman's minimal total loss, including all shooting costs and the cost of the target's survival. The generalized version of (5) is then

$$V(p) = \min\left\{vp; \min_i(c_i + E(V(b_i(pq_i, K))))\right\}. \tag{8}$$

The expected value in (8) is with respect to the distribution of K, the sensor report.

The same existence results apply to (8) as to (5). However, there is no counterpart to (7) in this case because of the complicated structure associated with looks. Solution of (8) must be by an approximation method such as policy or value iteration (Thomas et al. 1983, Puterman 1994).

2.3. Finite Horizon

Stationary policies can no longer be expected when the time horizon is finite. The best action will depend on time, as well as on the probability that the target is still alive, with the marksman sometimes resorting to increasingly expensive shots when time is almost used up. Obtaining POMDP solutions is more difficult and essentially numerical, but at least the problem of firing at many targets still decouples; there is no need to consider a joint firing problem as long as shots have costs, rather than constraints. See Monahan (1982) or Lovejoy (1991) for surveys of solution methods.

Models of this sort are particularly attractive if there are many targets of the same type, since the policy found in solving one POMDP applies to every target. Exactly such a model lies at the heart of Yost (1998), who considers problems where the number of target types is small, even though the number of targets is large.

2.4. Connection to the Constrained Case

In the example of §2.1, the target is always eventually killed after expending 1.667 looks, 1.667 shots of type 1, and 1.667 shots of type 2, on the average. If there were $n = 600$ targets, the expenditures to kill all of them would scale up to 1,000 for each resource. Now, consider the constrained problem of killing as many of 600 identical targets as possible, subject to constraints of 1,000 on each of the three resources, a problem that is much too hard to solve using the techniques of §1. We have seemingly almost stumbled onto a solution, because the application of the optimal policy to each of the targets independently will consume exactly the right resources in total. Everett (1963) shows that there can be no better solution to the constrained problem, but there are still two obstacles to application.

One obstacle is that the constraints have to be interpreted as constraints holding only on the average, because the consumption of the three resources in the joint POMDP is 1,000 on the average. On the other hand, constraints in the actual firing problem are more likely to be required to hold with certainty. This problem is least objectionable as the number n of targets grows large. The standard deviation of the total allocation in each case is proportional to the square root of n, while the mean is proportional to n, so the ratio of the two (the coefficient of variation) will approach zero as n approaches infinity. In this sense, the unconstrained case can be thought of as a solution method for constrained problems where n is large because resource consumption fluctuations become of less and less concern as n increases.

The other obstacle is that the constraints are unlikely to be met exactly, even on the average, because the coincidence that the shot and look costs happen to be chosen so that the constraints are all exactly satisfied is too much to hope for. There needs to be some mechanism for adjusting the shot costs to make that happen. Yost and Washburn (2000) propose to do this in an iterative scheme where a linear program produces dual variables that play the role of costs, while the POMDP uses the costs to produce new policies that end up being columns in the linear program. This mechanism for deriving resource costs from constraints makes it unnecessary to make a priori judgments about the comparative values of targets and shots.

The Yost-Washburn method is a column-generation technique for finding the correct costs and associated optimal policies. It is characteristic of such schemes that feasible solutions show rapid improvements at first, but that convergence in the tails is very slow. Because upper and lower bounds are available at all times when using that method, a practical implementation will incorporate a termination rule that accepts nearly optimal solutions.

3. The Semi-Constrained Case with Perfect Information

Assume that each shot i in some set S has a cost c_i and a kill probability $p_i = 1 - q_i$, and the object is to kill the single target as cheaply as possible. The target is presumed to be so valuable that shooting will not stop until either shots are exhausted or the target is killed. The probability of killing the target is therefore $1 - \prod_{i \in S} q_i$—the usual "powering up" formula that applies to independent shots. Because shooting will stop if the target is killed, it makes intuitive sense to first use shots with high kill probabilities and low costs. The following theorem makes this idea precise.

THEOREM 4. *Rank the shots according to increasing values of the cost/effectiveness ratio c_i/p_i. The minimal average cost of firing is achieved by making the shots in that order.*

PROOF. Let $Q_i = \prod_{j<i} q_i$. If the shots are made in index order, Q_i is the probability that all shots before the ith shot fail to kill the target, and therefore also the probability that the ith shot will be required. The expected cost of firing is therefore $C = \sum_{i \in S} c_i Q_i$. Let C_k be the expected cost of firing if the kth shot is interchanged with the $k+1$st in the firing order, and let $d \equiv C - C_k$. All but two terms from each sum cancel in the subtraction, so $d = Q_k((c_k + c_{k+1}q_k) - (c_{k+1} + c_k q_{k+1})) = Q_k(c_k p_{k+1} - c_{k+1}p_k)$. The difference d will be positive if and only if $c_k/p_k > c_{k+1}/p_{k+1}$. Because k is arbitrary, this shows that the shots should be ordered as stated. □

4. Applicability of Bandit Processes

A "bandit" process is one where the decision maker must at each time choose one of a fixed number of Markovian bandits, thus changing the state of the selected bandit, receiving a reward, and receiving some information about the changed state of the selected bandit. Bandits other than the one chosen are unaffected. Rewards are time-discounted by the factor $0 \leqslant \beta < 1$, and the decision maker's object is to maximize the expected value of the total discounted reward. The original application was to slot machines (one-armed bandits) with unknown payout statistics, hence the name. In the present context, the decision maker is a marksman who must at each time choose a bandit target to shoot at. The attractive feature of bandit processes is that they are indexable; that is, there exists a Gittins index, computable for each bandit separately from the others, such that choosing the bandit with the largest index at all times is an optimal policy. The Gittins index depends on the state of the bandit and is generally difficult to compute. Even so, the separability of each bandit from the others is an attractive feature.

An alternative interpretation is that rewards are not discounted but the bandit process may be terminated at any time, after which there are no further rewards. With this interpretation, β is the probability that the bandit process will not terminate after each decision, in which case the number of decision opportunities is a gradually revealed geometric random variable with mean $1/\beta$. In the present shooting context, $1/\beta$ is the average number of shots available to the marksman.

Bandit processes have occasionally been applied to shooting problems. For example, Glazebrook et al. (2001) consider a two-sided problem where each bandit is an attacker of unknown type, and where the defending marksman's problem is to decide which bandit to shoot at next, given that the bandit will return fire if not killed. Barkdoll et al. (2002) consider a related problem in suppression of enemy air defenses (SEAD) where the decision maker is the enemy air defense supposedly being suppressed. The marksman must not only decide which bandit (i.e., attacker) to shoot at next, but also how his engagement radar should operate in support of the shot, whether in continuous or intermittent mode. Here, the decisions concern not only the choice of bandit, but also how the chosen bandit should be dealt with. This additional feature formally takes the decision process out of the class for which index strategies are known to be optimal. Even so, Barkdoll et al. (2002) propose an index-based heuristic for their shooting problem that performs well in numerical experiments. See Glazebrook and Fay (1990) for a theoretical discussion of such developments in bandit problems.

In the interest of formulating a problem where the Gittins index can be derived explicitly, we consider next a simple, one-sided problem similar to those considered in previous sections.

The decision maker cannot have any choices other than bandit selection, so our marksman cannot have his choice of several shot types once he selects a target. Therefore, we introduce the following modified version of Equations (6) and (8), the modifications being to delete the shot subscript i, and replace the option of retiring with 0 payoff by the option of retiring with a payoff of M:

$$U(p, M) = \max\big(M; -c + pv(1-q) + \beta E(U(b(pq, K), M))\big). \quad (9)$$

The Gittins index $M(p)$ is the smallest value of M for which the options of retiring and continuing are equivalent (Whittle 1982). In general, $M(p)$ must be computed by an iterative numerical procedure, but the case where information is perfect is an exception. In that case (9) reduces to

$$U(p, M) = \max\big(M; -c + pv(1-q) + \beta\{pqU(1, M) + (1-pq)U(0, M)\}\big). \quad (10)$$

We will first find $U(0, M)$, then $U(1, M)$, and then deal with the general case. Because $U(0, M) = \max(M, -c + \beta U(0, M))$, it follows that $U(0, M) = \max(M, -c/(1-\beta))$, and therefore that $M(0) = -c/(1-\beta)$.

For $M \geqslant M(0)$, $U(0, M) = M$, and therefore,

$$U(1, M) = \max\big(M; -c + v(1-q) + \beta\{qU(1, M) + (1-q)M\}\big). \quad (11)$$

Equating $U(1, M)$ to the continuation expression in (10) and solving for $U(1, M)$, we see that (11) is equivalent to

$$U(1, M) = \max\left(M; \frac{-c + v(1-q) + \beta(1-q)M}{1-\beta q}\right). \quad (12)$$

For values of M between $M(0)$ and $M(1)$, $U(0, M) = M$ and $U(1, M)$ is the continuation expression in (12), so for the general case $0 \leqslant p \leqslant 1$ we have Equation (13)

$$U(p, M) = \max\left(M; -c + pv(1-q) + \beta\left\{pq\frac{-c + v(1-q) + \beta(1-q)M}{1-\beta q} + (1-pq)M\right\}\right). \quad (13)$$

The smallest (in fact only) value of M for which the two expressions in (13) are equal is the Gittins index,

$$M(p)=\frac{pv(1-q)}{(1-\beta)(1-\beta q(1-p))}-\frac{c}{1-\beta};\qquad 0\leqslant p\leqslant 1.(14)$$

If there are actually several targets, then we have only to add subscripts to p, c, v, and q to have a shooting index—a number that determines the next shot in all circumstances.

The index policy is myopic when $\beta=0$ because the target with the largest immediate gain $pv(1-q)-c$ is always chosen. However, in general $M(p)$ is an increasing, concave nonlinear function of p, and the best policy is not necessarily myopic.

Example. Suppose $\beta=0.9$, and that there are two targets with $q=0.5$ and $c=0$ for each. If $v_1=1$ and $p_1=1$, the index for target 1 is 5. If $v_2=4$ and $p_2=0.2$, the index for target 2 is 6.25. The myopic rule would prefer to engage target 1 first, but the optimal rule prefers target 2. Valuable targets with a low probability of being alive have a surprisingly large index on account of the concavity of $M(p)$.

5. Summary

We have reviewed the shoot-look-shoot state of the art and introduced several new results.

As often happens, the difficult constrained problems are in the middle. The problems of §1 are tractable when the numbers of shots and targets are small. When all numbers are large, the methods of subsequent sections may produce approximate solutions by decoupling the targets. Problems with moderate numbers of targets and shots remain difficult, particularly if information is not perfect.

Appendix. Proof of Theorem 1

Proof (refer to §1.3 for notation and exact statement of Theorem 1). For y an observation and π a probability, we first introduce the function

$$T(\pi,y)\equiv\frac{q\pi P(Y=y\mid X=1)}{q\pi P(Y=y\mid X=1)+(1-q\pi)P(Y=y\mid X=0)}$$

(recall that the conditional probabilities are assumed known). If $(y_1,\ldots,y_k)$ is a sequence of observations, we generalize by defining

$$T(\pi,(y_1,\ldots,y_k))\equiv T(T(\pi,(y_1,\ldots,y_{k-1})),y_k).$$

With this notation, Equation (4) is $\pi_{ik}=T(\pi_{i,k-1},\,y_k)$. Equivalently, $\pi_{ik}=T(\pi_{i,0},\,(y_1,\ldots,y_k))$. We first observe that $T(\pi,y)$ is a nondecreasing function of π for any observation or sequence of observations, as is easily proved by induction.

Let $\pi_k^*\equiv\max_i\{\pi_{ik}\}$ be the largest probability of being alive at time k. The theorem states that any myopic policy is optimal, a myopic policy being any policy that always chooses a target i for which $\pi_{ik}=\pi_k^*$ at all times $k\geqslant 0$. Our proof is by induction on s, the number of shots. The proof is trivial if $s=1$. The problem is to show that there is some optimal policy that makes the myopic choice at $k=0$ because it can be assumed by the induction principle that the myopic choice is optimal for $k>0$. Suppose that some optimal policy P first chooses target 2, where $\pi_{20}<\pi_0^*$. We can assume that P proceeds myopically for $k>0$, because myopic continuations are optimal by induction. We will show that there is a different policy Q that first chooses myopically and that is at least as good as P.

Because P is myopic for $k>0$, it will shoot at target 2 until $\pi_{2K}<\pi_0^*$ at some random stopping time K, or until shots are exhausted, whichever comes first. Let a sequence of observations $\sigma\equiv(y_1,\ldots,y_k)$ be called "critical" if observation of that sequence of reports is included in the event $(K=k)$, let $\Sigma(k)$ be the set of all critical sequences of length k, and let $\Sigma\equiv\bigcup_{k=1}^{s}\Sigma(k)$. P will switch from target 2 to some other target immediately after some critical sequence in Σ is observed, unless shots are exhausted first.

Suppose that P switches from target 2 to target 1 after time K, and note that target 1 is necessarily a myopic target at time 0 as well as time K, because target 1's probability of being alive is π_0^* at both times. Suppose further that P next switches from target 1 to some other target (possibly target 2) after some observation sequence $\sigma=(y_1,\ldots,y_k)$ associated with target 1 is observed. Then it must be true that $T(\pi_{10},\sigma)<\pi_0^*$. Because $\pi_{20}\leqslant\pi_{10}$, and because $T(\pi,\sigma)$ is a nondecreasing function of π, it is also true that $T(\pi_{20},\sigma)<\pi_0^*$; that is, either σ or some subsequence of σ must be critical. In other words, for any observation history for which P continues to shoot at target 2, P will also continue to shoot at target 1, given the same reports about target 1. In fact, P can be described as the policy that first shoots at target 2 according to Σ, then shoots at target 1 according to Σ, and then proceeds myopically (possibly by continuing to shoot at target 1) as long as shots remain. We refer to the first two parts of P as being Σ-controlled. Let Q be the policy that simply switches targets 1 and 2 in this description. We show that Q is at least as good as P, thus proving the theorem.

In comparing P and Q, we need only compare the chances of killing targets 1 and 2 under the Σ-controlled phase. This is because the policies differ only in the order in which targets 1 and 2 are engaged, so the distribution of the system state at the end of Σ-controlled shooting, including the possibility that all shots are exhausted, is the same in either case.

We now consider the four possibilities for whether targets 1 and 2 are actually alive at time 0. If both are dead, clearly P and Q will be equally effective under Σ-controlled shooting. The same is true if both are initially alive, because the same stopping rule is applied to both targets, so consider the case where only one target is alive. Let $D(t)$ be the probability of killing a target that is

shot at until shooting is stopped by either Σ or the horizon t, whichever comes first, given that the target is initially live, and note that $D(t)$ is a nondecreasing function of t. If the live target is engaged first, then the probability of killing it under Σ-controlled fire is $D(s)$. If it is engaged second, then let X be the number of shots used (wasted) by first shooting at the one that is dead. The probability of killing the live target is then $D^* \equiv E(D(s-X))$, which cannot exceed $D(s)$ on account of the monotonic nature of $D(t)$. Considering only targets 1 and 2 under Σ-controlled fire, the average number of targets killed by policy P is, therefore, $\pi_{20}(1-\pi_{10})D(s)+\pi_{10}(1-\pi_{20})D^*$, with a similar expression with targets 1 and 2 reversed for policy Q. Because $\pi_{20} \leqslant \pi_{10}$ and $D^* \leqslant D(s)$, it is trivial to show that policy Q will kill at least as many targets as policy P.

Because Policy Q has been shown to be at least as good as policy P, and because policy Q begins myopically, this completes the inductive proof that all myopic policies are optimal. □

References

Anderson, L. 1989. A heterogeneous shoot-look-shoot attrition process. Paper P-2250, Institute for Defense Analysis, Alexandria, VA, 10–11.

Aviv, Y., M. Kress. 1997. Evaluating the effectiveness of shoot-look-shoot tactics in the presence of incomplete damage information. *Military Oper. Res.* **3**(1) 79–89.

Barkdoll, T., D. Gaver, K. Glazebrook, P. Jacobs, S. Posadas. 2002. Suppression of enemy air defenses (SEAD) as an information duel. *Naval Res. Logist.* **49** 723–742.

Blackwell, D. 1965. Discounted dynamic programming. *Ann. Math. Statist.* **36**(1) 226–235.

Eckler, A., S. Burr. 1972. *Mathematical Models of Target Coverage and Missile Allocation*. Military Operations Research Society, Washington, DC, 109–112.

Everett, H. 1963. Generalized Lagrange multiplier method for solving problems of optimal allocation of resources. *Oper. Res.* **11** 399–417.

Glazebrook, K., N. Fay. 1990. Evaluating strategies for Markov decision processes in parallel. *Math. Oper. Res.* **15** 17–32.

Glazebrook, K., D. Gaver, P. Jacobs. 2001. On a military scheduling problem. Technical Report NPS-OR-01-010, Naval Postgraduate School, Monterey, CA.

Lovejoy, W. 1991. A survey of algorithmic methods for partially observed Markov decision processes. *Ann. Oper. Res.* **28** 47–65.

Manor, G., M. Kress. 1997. Optimality of the greedy shooting strategy in the presence of incomplete damage information. *Naval Res. Logist.* **44** 613–622.

Monahan, G. 1982. A survey of partially observable Markov decision processes. *Management Sci.* **28** 1–16.

Przemieniecki, J. 1990. *Introduction to Mathematical Methods in Defense Analysis*. American Institute of Aeronautics and Astronautics, Washington, DC, 134–136.

Puterman, M. 1994. *Markov Decision Processes*. Wiley, New York.

Ross, S. 1970. *Applied Probability Models with Optimization Applications*. Holden-Day, San Francisco, CA, 132–141.

Ross, S. 1971. Quality control under Markov deterioration. *Management Sci.* **17** 587–596.

Strauch, R. 1966. Negative dynamic programming. *Ann. Math. Statist.* **37** 871–890.

Thomas, L., R. Hartley, A. Lavercombe. 1983. Computational comparison of value iteration algorithms for discounted Markov decision processes. *Oper. Res. Lett.* **2** 72–76.

Yost, K. 1998. Solution of large scale allocation problems with partially observable outcomes. Ph.D. dissertation, Naval Postgraduate School, Monterey, CA.

Yost, K., A. Washburn. 2000. Optimizing assignments of air-to-ground assets and BDA sensors. *Military Oper. Res.* **5**(2) 77–91.

Whittle, P. 1982. *Optimization Over Time*, Ch. 14. Wiley, NewYork.

Wilkening, D. 1999. A simple model for calculating ballistic missile defense effectiveness. Working paper, Center for International Security and Cooperation, Stanford, CA, 12–17.