Office of Naval Research
Contract N00014-67-A-0298-0006 NR-372-012

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
Grant NGR 22-007-068

ON A CLASS OF STOCHASTIC PURSUIT-EVASION GAMES

VE RI TAS

By

W. W. Willman

April 1969

Technical Report No. 585

Division of Engineering and Applied Physics
Harvard University • Cambridge, Massachusetts

Office of Naval Research

Contract N00014-67-A-0298-0006

NR-372-012

National Aeronautics and Space Administration

Grant NGR 22-007-068

ON A CLASS OF STOCHASTIC PURSUIT-EVASION GAMES

by

W. W. Willman

Technical Report No. 585

April 1969

The research reported in this document was made possible through support extended the Division of Engineering and Applied Physics, Harvard University by the U. S. Army Research Office, the U. S. Air Force Office of Scientific Research and the U. S. Office of Naval Research under the Joint Services Electronics Program by Contracts N00014-67-A-0298-0006, 0005, and 0008 and by the National Aeronautics and Space Administration under Grant NGR 22-007-068.

Division of Engineering and Applied Physics

Harvard University Cambridge, Massachusetts

ON A CLASS OF STOCHASTIC PURSUIT-EVASION GAMES

by

W. W. Willman

Division of Engineering and Applied Physics

Harvard University Cambridge, Massachusetts

ABSTRACT

This report continues the investigation of a class of stochastic differential games introduced in an earlier report by Willman [1]. Iterative algorithms for computing numerical solutions to games of this type are discussed. It is shown by numerical example that there exist multistage games analogous to this type of differential game for which minimax solutions do indeed exist. Several candidates for quasi-optimal strategies are presented which are simple to compute and easy to implement compared to the minimax strategies. A criterion is developed for evaluating the performance of non-optimal strategies. The performances of these quasi-optimal strategies are evaluated in terms of this criterion for a numerical example (an interception problem involving two second-order systems).

1. INTRODUCTION

This report is meant to be a continuation of an earlier report by this author [1]. The main subject of that earlier report is the stochastic differential game described by the following state vector transition equation:

$$\dot{x} = G_p u - G_e v + \xi , \tag{1:1}$$

where u is the control vector chosen by one player (called the pursuer) and v is the control vector chosen by the other player (called the evader). The players have a common prior probability distribution on the initial state which is Normal $(\bar{x}_o, P_o)$, and receive the following measurements:

$$z_p = H_p x + w_p \qquad \text{(pursuer's measurements)} \tag{1:2}$$

and

$$z_e = H_e x + w_e \qquad \text{(evader's measurements)}. \tag{1:3}$$

$\left[\begin{array}{c} \xi \\ \hline w_p \\ \hline w_e \end{array}\right]$ is a zero-mean Gaussian white noise process with spectral

density parameter $\left[\begin{array}{c|c|c} Q & 0 & 0 \\ \hline 0 & R_p & 0 \\ \hline 0 & 0 & R_e \end{array}\right]$. The criterion which the

pursuer wishes to minimize and the evader wishes to maximize is

$$J = \varepsilon\left[x_f^T S_f x_f + \int_0^{t_f} (u^T Bu - v^T Cv)\, dt\right]. \tag{1:4}$$

Pursuit and evasion strategies are sought to minimaximize the value of this criterion, making this a zero-sum game.

Some results are obtained there relating the solution of this game to that of the so-called "corresponding deterministic game." This corresponding deterministic game is described by the equations:

$$\dot{x} = G_p u - G_e v \qquad \text{(transition equation)} \tag{1:5}$$

$$J = x_f^T S_f x_f + \int_0^{t_f} (u^T Bu - v^T Cv)\, dt \qquad \text{(criterion)} \tag{1:6}$$

where both players can measure the state x exactly.

The solution to the corresponding deterministic game has been obtained by Ho, Bryson, and Baron [2]:

$$u = -B^{-1} G_p^T Sx \qquad \text{(pursuer's strategy)} \tag{1:7}$$

$$v = -C^{-1} G_e^T Sx \qquad \text{(evader's strategy)} \tag{1:8}$$

where the matrix $S(t)$ is defined by the differential equation

$$\dot{S} = S\left[G_p B^{-1} G_p^T - G_e C^{-1} G_e^T\right] S; \quad S(t_f) = S_f. \tag{1:9}$$

Two forms of the minimax strategies for the stochastic game are derived in Willman [1]. In one form, these strategies are expressed directly as linear functionals of the available measurements:

$$u(t) = -B^{-1}(t) G_p^T(t)\left[A_p(t)\, \bar{x}_o + \int_0^t \Lambda_p(t, r) z_p(r) dr\right] \tag{1:10}$$

and

$$v(t) = -C^{-1}(t) G_e^T(t)\left[A_e(t)\, \bar{x}_o + \int_0^{t_f} \Lambda_e(t, r) z_e(r) dr\right]. \tag{1:11}$$

Another form of these same minimax strategies is derived by utilizing the fact that the mean minimax sample path of this stochastic game coincides with the minimax path of the corresponding deterministic game (the so-called "certainty-coincidence property"). The minimax strategies are expressed in this alternate form, called Realization III, as the certainty-equivalent minimax strategies plus error terms which are zero on the mean sample path.

The stochastic differential game is solved (in terms of a set of implicit equations) in Willman [1] by considering it as the limiting form of a sequence of approximating multistage games. First, the solution to the analogous multistage game is characterized in terms of a complicated set of implicit difference equations. Then a sequence of multistage games of this type is considered, all of which arise from making successively fine time-discretizations of the differential game. A set of implicit integro-differential equations characterizing the solution to this differential game is then obtained formally by taking the limiting form of the equations for the multistage game sequence as the discretization interval becomes infinitesimal.

BLANK PAGE

2. COMPUTATIONAL CONSIDERATIONS

The solution in Willman [1] for the stochastic differential game was given in terms of a solution to the primary and subsidiary systems of implicit equations derived there. For the purpose of discussing the properties of such game solutions, it was assumed that a solution to these equation systems had been obtained. The questions of the existence and computation of such solutions to these implicit equations will be investigated in this section. As in the derivation of the game solution, this investigation will proceed by first considering the analogous multistage game and then treating the differential game as a limiting case of a sequence of multistage games.

A. The Multistage Game

The multistage analogue of the generic differential game considered here is examined in detail in Appendix A of Willman [1]. Restating the problem here for convenience, this multistage pursuit-evasion game is described by the equations

$$x(i+1) = x(i) + G_p(i)u(i) - G_e(i)v(i) + \xi(i)$$

$$J = \frac{1}{2}\,\varepsilon\left[x^T(N)\,S_f\,x(N) + \sum_{i=0}^{N-1}\Big(u^T(i)B(i)u(i) - v^T(i)C(i)v(i)\Big)\right]$$

$$z_p(i) = H_p(i)x(i) + w_p(i)$$

$$z_e(i) = H_e(i)x(i) + w_e(i)$$

$$\begin{bmatrix} \xi(i) \\ \hline \bar{w}_p(i) \\ \hline \bar{w}_e(i) \end{bmatrix} \text{ are independent Normal} \left(\begin{bmatrix} 0 \\ \hline 0 \\ \hline 0 \end{bmatrix}, \left[\begin{array}{c|c|c} Q(i) & 0 & 0 \\ \hline 0 & R_p(i) & 0 \\ \hline 0 & 0 & R_e(i) \end{array}\right] \right)$$

The common prior is independent of the noises and is Normal $(\bar{x}_o, P_o)$.

To briefly summarize the results of Appendix A of Willman [1], it is shown that if the pursuer uses a strategy of the form

$$U^*: u(i) = -a_p^*(i) - \sum_{j=0}^{i} \Lambda_p^*(i,j) z_p(j), \qquad i = 0, \ldots, N-1,$$

then the evasion strategy that maximizes against it, if such a strategy exists, is

$$V': v(i) = a_e'(i) + \sum_{j=0}^{i} \Lambda_e'(i,j) z_e(j), \qquad i = 0, \ldots, N-1,$$

where the $\Lambda_e'(i,j)$'s and the $a_e'(i)$'s are obtained from the $\Lambda_p^*(i,j)$'s and the $a_p^*(i)$'s by solving a stochastic optimal control problem. This dependence is expressed schematically as

$$\Lambda_e' = f_e(\Lambda_p^*) \tag{2:1}$$

and

$$a_e' = g_e(\Lambda_p^*, a_p^*). \tag{2:2}$$

Analogously, it is also shown that if the evader uses the generic strategy

$$V^*: v(i) = \alpha_e^*(i) + \sum_{j=0}^{i} \Lambda_e^*(i,j) z_e(j), \qquad i = 0, \ldots, N-1,$$

then the pursuit strategy that minimizes against it, if such exists, is

$$U': u(i) = -\alpha_p'(i) - \sum_{j=0}^{i} \Lambda_p'(i,j) z_p(j), \qquad i = 0, \ldots, N-1,$$

where

$$\Lambda'_p = f_p(\Lambda^*_e) \tag{2:3}$$

and

$$\alpha'_p = g_p(\Lambda^*_e, \alpha^*_e) . \tag{2:4}$$

Moreover, any evasion strategy satisfying the conditions (2:1) and (2:2) which also satisfies a certain convexity condition is shown to be optimal for the evader against the pursuit strategy U^*. Likewise, any pursuit strategy satisfying (2:3), (2:4), and the corresponding convexity condition minimizes against the evasion strategy V^*.

These results suggest the following kind of iterative procedure for determining a solution to this kind of game:

(i) Guess the values of $a^*_p(i)$, $\Lambda^*_p(i, j)$, $\alpha^*_e(i)$, and $\Lambda^*_e(i, j)$ for $i = 0, \ldots, N-1$; $j = 0, \ldots, i$.

(ii) Use the relations (2:1) - (2:4) to determine the values of $a'_e(i)$, $\Lambda'_e(i, j)$, $\alpha'_p(i)$, and $\Lambda'_p(i, j)$.

(iii) Repeat from step (i) with the values of $\alpha'_p(i)$, $\Lambda'_p(i, j)$, $a'_e(i)$, and $\Lambda'_e(i, j)$ substituted for $a^*_p(i)$, $\Lambda^*_p(i, j)$, $\alpha^*_e(i)$, and $\Lambda^*_e(i, j)$, respectively.

If it happens that the values being substituted in step (iii) on any iteration are all equal to the values that they are replacing, then these values satisfy the relations (2:1) - (2:4) by construction. That is, the convergence of this algorithm implies that it has converged at values of a'_p, Λ'_p, a'_e, and Λ'_e such that U' and V' are minimax pursuit and evasion strategies, provided that these values also satisfy the convexity conditions. These convexity conditions should be checked after the application of this algorithm, if it converges, to verify that the values to which it has converged is indeed a solution.

Notice that from a conceptual point of view this algorithm can be interpreted as the following procedure:

(i) Guess an initial pair of pursuit and evasion strategies U^* and V^*.

(ii) Calculate the pursuit strategy U' that minimizes against V^*, and the evasion strategy V' that maximizes against U^*.

(iii) Repeat from step (i) with U' and V' replacing U^* and V^*.

The steps here correspond to those listed in the preceding algorithm. This interpretation of the algorithm as a series of alternate optimizations will be useful at a later time when the extension of this algorithm to the differential game is discussed.

So far, this algorithm has only been presented in a vague outline form. Although there is no intention of examining all the details here, it is instructive to descend at least one level toward greater explicitness in order to appreciate the computational requirements of step (ii). It is apparent from the description of this step (ii) that its implementation requires the manipulation of several equation schemes from Appendix A of Willman [1]. These equation schemes involve some intermediate variables, which are either "enlarged" matrices or "enlarged" vectors. These intermediate variables are, for $i = 0, \dots, N$ and $j = 0, \dots, i$:

$$\left.\begin{matrix}\gamma_p(i)\\ \delta_p(i)\end{matrix}\right\}\quad (n+r)(N+1)\ \text{vectors}$$

$$\left.\begin{matrix}\gamma_e(i)\\ \delta_e(i)\end{matrix}\right\}\quad (n+q)(N+1)\ \text{vectors}$$

$$\left.\begin{matrix}S_p(i)\\ P_p(i)\end{matrix}\right\}\quad (n+r)(N+1)\,x\,(n+r)(N+1)\ \text{matrices}$$

$$\left.\begin{matrix}S_e(i)\\ P_e(i)\end{matrix}\right\}\quad (n+q)(N+1)\,x\,(n+q)(N+1)\ \text{matrices}$$

$K_p(i, j)$ $\quad (n+r)(N+1)\,xq$ matrices

$K_e(i, j)$ $\quad (n+q)(N+1)\,xr$ matrices

where N is the number of stages in the game, n is the dimension of the state variable x, and r and q are the respective dimensions of the evader's and pursuer's measurement vectors z_e and z_p.

With these intermediate variables in mind, it is possible to subdivide step (ii) of the preceding algorithm outline into a number of smaller steps. A flow diagram is shown in Figure 2-1 for the computation of the new values of the Λ_e's and a_e's. The procedure for computing successive values of the Λ_p's and a_p's is entirely analogous. Basically, each of the two parts of this implementation consists of a backward sweep followed by a forward sweep. The dominating storage requirement in this implementation is that of storing all N of the "S" matrices generated in the first backward sweep. From the manner in which the equation schemes generating the "S" matrices originate, it is well known that they are symmetric. This means that this algorithm requires the ability to store approximately $\frac{1}{2}(n+m)^2N^3$ numbers

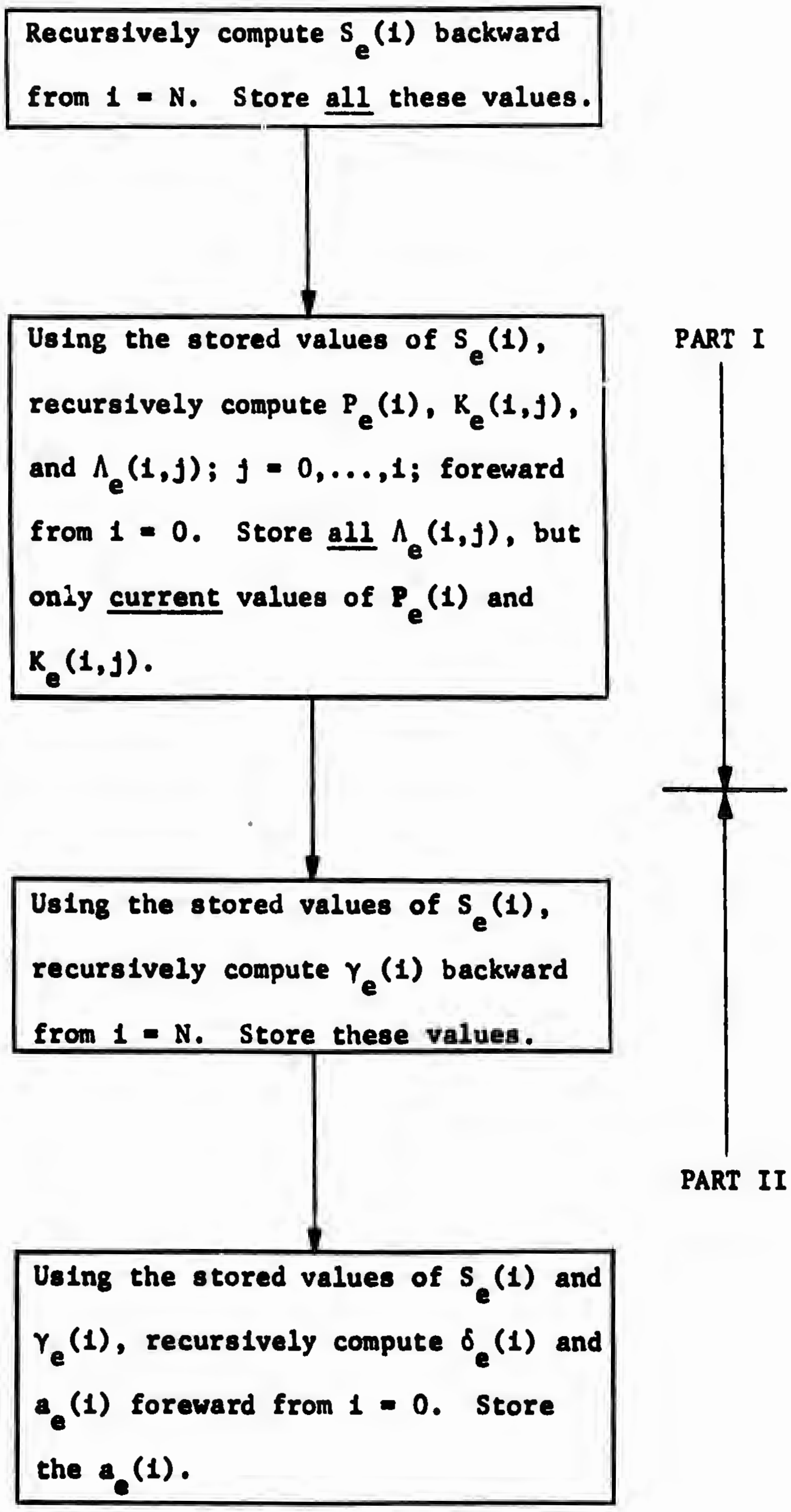

FIG. 2-1 MULTISTAGE GAME ALGORITHM

simultaneously, where $m = \max(q,r)$. Since each computation mentioned in the preceding implementation involves "enlarged" vectors and matrices, whose dimensions vary respectively as N and N^2, the amount of processing required also becomes very large as N (the number of stages) increases, although the extent of this requirement is hard to estimate without examining this algorithm in great detail. It is encouraging to note, however, that the only matrix inversions involved in these calculations are applied to matrices of dimension n or less.

A point of some minor interest is that the values of the Λ's can be computed independently of the a's with this type of algorithm, as is indicated by equations (2:1) - (2:4). This is important because the convexity conditions, which should be verified if the algorithm converges to a result, depend on the Λ's but <u>not</u> on the a's. If these conditions are satisfied, then the corresponding a's can be obtained later, using the previously calculated values of the Λ's. In fact, if $\bar{x}_o$, the initial estimate of the initial state, is zero, then it follows by inspection that a solution is obtained by taking all of the a's, γ's, and δ's to be zero.

B. An Illustrative Example

Since an effective procedure is known for determining whether or not a strategy pair produced by the above algorithm is actually a solution to the game, the question that arises next is whether the algorithm ever converges, and if so, under what circumstances. The only answers to these questions available at the present time are based on numerical evidence from the following two-stage scalar example.

This evidence indicates, to the extent that any numerical evidence can indicate, that this algorithm does indeed converge under a fairly wide range of circumstances and that solutions to the game exist under these circumstances. This last point is reassuring because it means that the theory presented here is not occurring in a vacuum.

Example

$x_{i+1} = x_i + u_i - v_i; \quad i = 0, 1$ (transition equation)

$J = \varepsilon\left[x_2^2 + \sum_{i=0}^{1}\left(bu_i^2 - cv_i^2\right)\right]; \quad b, c > 0$ (criterion)

$z_{pi} = x_i + w_{pi}$ (pursuer's measurements)

$z_{ei} = x_i + w_{ei}$ (evader's measurements)

prior on x_o is Normal(0, 1)

$\begin{bmatrix} w_{pi} \\ w_{ei} \end{bmatrix}$ are independent Normal$\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} q & 0 \\ 0 & r \end{bmatrix}\right)$ and independent of the prior.

The reason a two-stage example is chosen is that this is the minimum number of stages for which the interesting features of these stochastic games become apparent. The one-stage example solved in Willman [1] is misleadingly simple: the implicit equations could be solved explicitly by substitution and the estimation problem was independent of the control problem. Neither of these features is present in this example.

In this example the quantities b and c can be interpreted as indicators of the pursuer's and evader's energy capabilities, respectively. As b increases, the restriction on the pursuer's energy expenditure is greater, and hence his capability is lowered, and likewise for the evader. The quantities q and r, however, represent information capabilities. As q increases, the pursuer's measurements become less reliable and his information capability decreases. Similarly, an increase in r represents a decrease in the evader's information capability.

It can be shown that in the corresponding deterministic open-loop game (in which the players receive no measurements) a minimax solution exists if and only if $c \geq 2$. If this inequality is not satisfied, the evader can make J arbitrarily large, or "escape." In the corresponding deterministic closed-loop game (perfect measurements), on the other hand, the conditions under which a solution exists are $c \geq 1$ and $c \geq \frac{2b}{(1+b)}$. The first of these inequalities is a less stringent version of the existence condition for the open-loop game, less stringent because of the improved information capability of the pursuer. The second inequality represents a condition imposed on the relative energy capabilities of the two players. As $b \to \infty$, the operative inequality is the second, which approaches $c \geq 2$ (the open-loop condition); for $b \leq 1$, the first inequality is the operative one. These existence conditions are not important in themselves but rather as guideposts in examining the behavior of the above algorithm for various parameter values. These regions are shown in Figure 2-2.

Because of the small number of stages in this game, it is expedient to express the computations of step (ii) of this algorithm directly in terms of a single function evaluation for each Λ, bypassing the intermediate variables. Since the mean of the prior is zero in this example,

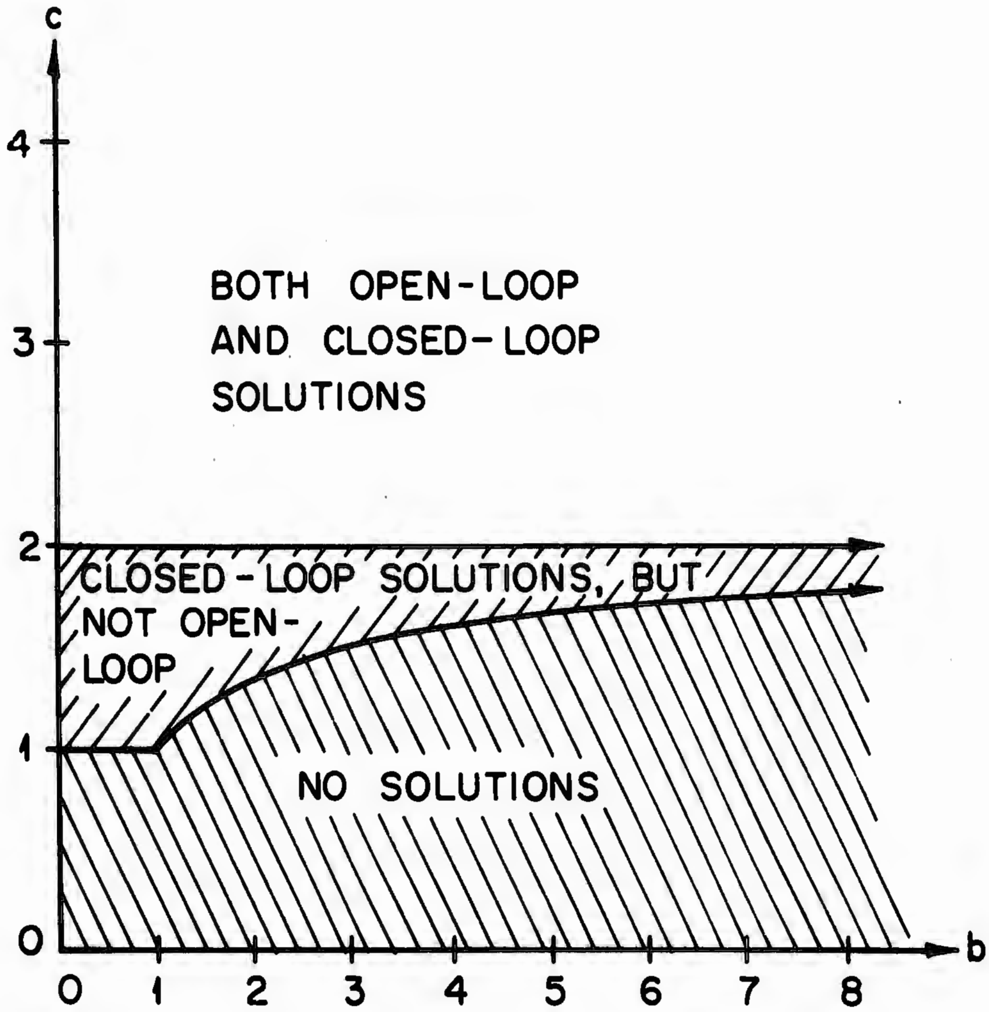

FIG. 2-2 REGIONS OF SOLUTION EXISTENCE

the values of the a's can be taken as zero, as was explained earlier. This set of functional relations, which constitute the "primary" system of implicit equations for this example, will be stated here because it is instructive to compare their complexity with that of the analogous equations in the one-stage example in Willman [1]. Before doing this, however, it is convenient to reduce the complexity of the notation by referring to the generic linear strategies as

$$U: \left\{ \begin{aligned} u_0 &= g_1 z_{p0} \\ u_1 &= g_2 z_{p0} + g_3 z_{p1} \end{aligned} \right\} \tag{2:5}$$

and

$$V: \left\{ \begin{aligned} v_0 &= h_1 z_{e0} \\ v_1 &= h_2 z_{e0} + h_3 z_{e1} \end{aligned} \right\} . \tag{2:6}$$

Defining

$$m_p \triangleq \left[\frac{(1-h_1)^2 q}{1+q} + h_1^2 r \right]$$

and

$$m_e \triangleq \left[\frac{(1+g_1)^2 r}{1+r} + g_1^2 q \right] ,$$

and denoting the "new" values of the g's and h's in expressions (2:5) and (2:6) with primes, the equations involved in the implementation of step (ii) of the iterative algorithm are:

$$g_1' = \frac{-(1+g_3-h_3+bg_3)g_2 - \left(\frac{1}{1+q}\right)\left[[(1+g_3-h_3)^2+bg_3^2-ch_3^2](1-h_1)-(1+g_3-h_3+ch_3)h_2\right]}{(1+g_3-h_3)^2 + b(1+g_3^2) - ch_3^2}$$

$$g_2' = \frac{1}{1+b}\left[\left(\frac{q}{m_p+q}\right)\left(g_1 + \frac{1-h_1}{1+q}\right)\left(h_3 - 1 + \frac{h_2(h_1-1)}{1+q} + \frac{rh_1h_2}{q}\right) + \frac{h_2}{1+q}\right]$$

$$g_3' = \frac{1}{(1+b)(m_p+q)}\left[(h_3-1)m_p + \frac{h_2(1-h_1)q}{1+q} - rh_1h_2\right]$$

$$h_1' = \frac{\left(\frac{1}{1+r}\right)\left[\left[(1+g_3-h_3)^2+bg_3^2-ch_3^2\right](1+g_1)+(1+g_3-h_3+bg_3)g_2\right] - (1+g_3-h_3+ch_3)h_2}{(1+g_3-h_3)^2 - c(1+h_3^2) + bg_3^2}$$

$$h_2' = \frac{1}{1-c}\left[\left(\frac{r}{m_e-r}\right)\left(\frac{1+g_1}{1+r} - h_1\right)\left(1 + g_3 - \frac{g_2(1+g_1)}{1+r} - \frac{qg_1g_2}{r}\right) + \frac{g_2}{1+r}\right]$$

$$h_3' = \frac{1}{(1-c)(m_e+r)}\left[(1+g_3)m_e + \frac{rg_2(1+g_1)}{1+r} + qg_1g_2\right].$$

This form of the algorithm was applied for several different parameter values. In all of the cases tried, convergence was apparently insensitive to the initial values chosen for the "g" and "h" parameters. The following results indicate the convergence properties of the algorithm for this example:

b	c	q	r	convergence in 100 steps?
1	2	.01	.01	YES
1.5	1.5	.01	.01	NO
1.5	1.5	1	1	YES
1.5	1.5	10	1	NO

In the first case, the parameters b and c are such that both the open-loop and closed-loop solutions exist; in the other three cases only the closed-loop solution exists. The apparent failure of the algorithm to converge, however, does not mean that the evader can make the criterion J arbitrarily large. As long as the closed-loop solution exists and the pursuer's measurement noise variance is finite, it follows from the certainty-equivalence principle (applied to the evader) that the pursuer can guarantee a finite upper bound on J by applying his closed-loop control law with z_{pi} substituted for x_i. Therefore, the pursuer can prevent the evader from "escaping" in all of the cases listed above, whereas this algorithm apparently converges in only two of them. In the cases where it did converge, the values were accurate to three significant figures after about twenty iterations.

The solution obtained for the first set of parameters listed above is interesting. For these values, the result was:

$$U^o: \begin{cases} u_o = -.4958\, z_{po} \\ u_1 = -.1997\, z_{po} - .3996\, z_{p1} \end{cases} \quad \text{(pursuer's stochastic minimax strategy)}$$

and

$$V^o: \begin{cases} v_o = -.2512\, z_{eo} \\ v_1 = -.1034\, z_{eo} - .1990\, z_{e1} \end{cases} \quad \text{(evader's stochastic minimax strategy)}$$

Because the standard deviation of each player's measurement noise is only one-tenth that of the prior on the initial state, it might be expected that the solution to this stochastic game is "close to" that of the corresponding deterministic closed-loop game, in which both players have perfect measurements. The solution to this deterministic game happens to be:

$$U^o: \begin{Bmatrix} u_o = -.5\, x_o \\ u_1 = -\frac{2}{3} x_1 \end{Bmatrix} \quad \text{and} \quad V^o: \begin{Bmatrix} v_o = -.25\, x_o \\ v_1 = -\frac{1}{3} x_1 \end{Bmatrix},$$

which does not appear to agree closely with the solution to the stochastic game under the assumption that $z_{pi} \cong z_{ei} \cong x_i$, $i = 1, 2$. This last assumption is based on the fact that the measurement noises are small. If the minimax strategies are substituted into the dynamics in the deterministic game, however, it is seen that

$$x_1 = .75\, x_o .$$

Using this last identity, the control histories produced by the minimax strategies in the deterministic game can be expressed as:

$$\left\{ \begin{aligned} u_o &= -.5\, x_o \\ u_1 &= -.2\, x_o - .4\, x_1 \end{aligned} \right\} \quad \text{and} \quad \left\{ \begin{aligned} v_o &= -.25\, x_o \\ v_1 &= -.1\, x_o - .2\, x_1 \end{aligned} \right\},$$

which do agree closely with the control histories produced by the stochastic minimax strategies under the assumption mentioned earlier. This means that the solutions to these two games are similar in the sense that the trajectories they generate are close to each other with high probability. They are not similar, however, in the straightforward sense that one might initially expect.

C. The Differential Game

The preceding algorithm for solving the multistage game immediately suggests the possibility of solving the differential game by the same method of alternate optimization. In the case of the differential game, the parameters being iteratively calculated are the functions $A_p(t)$, $A_e(t)$, $\Lambda_p(t, \gamma)$, and $\Lambda_e(t, \tau)$ occurring in expressions (1:10) and (1:11), and the equations used to calculate them are the primary and subsidiary equation systems listed in Section 4 of Willman [1].

As in the multistage case, it is instructive to examine in some detail the procedure used to calculate new values of these parameters from previous values. In the actual implementation of this procedure, of course, the independent variables t, τ, and σ will be discretized. These variables should therefore be considered in the following remarks as being discretized into N values of increment Δ in the range $[0, t_f]$. Under this assumption, the integrations called for in the primary and subsidiary equations should be interpreted as summations in the usual way.

Given the parameters characterizing the pursuit and evasion strategies at a particular iteration of this process, the suggested procedure for generating the values of the parameters for the pursuit strategy at the next iteration (i.e., the Λ_p's and the A_p's) is shown in Figure (2-3). The method for computing successive values of the Λ_e's and A_e's is entirely analogous.

An examination of this procedure shows that the storage requirement is approximately $5n^2N^2$ numbers, assuming that the dimensions of the measurement vectors z_p and z_e are less than n (the dimension of the state vector x). If the number of discretization steps N is large, this requirement represents a significant improvement over the storage capacity that would be required if the discretized game were treated as a general multistage game, which would require storage proportional to N^3. This saving is possible because the fact that the discretization step size Δ is small can be used to eliminate the computation of higher order terms. In other words, the preceding algorithm takes advantage

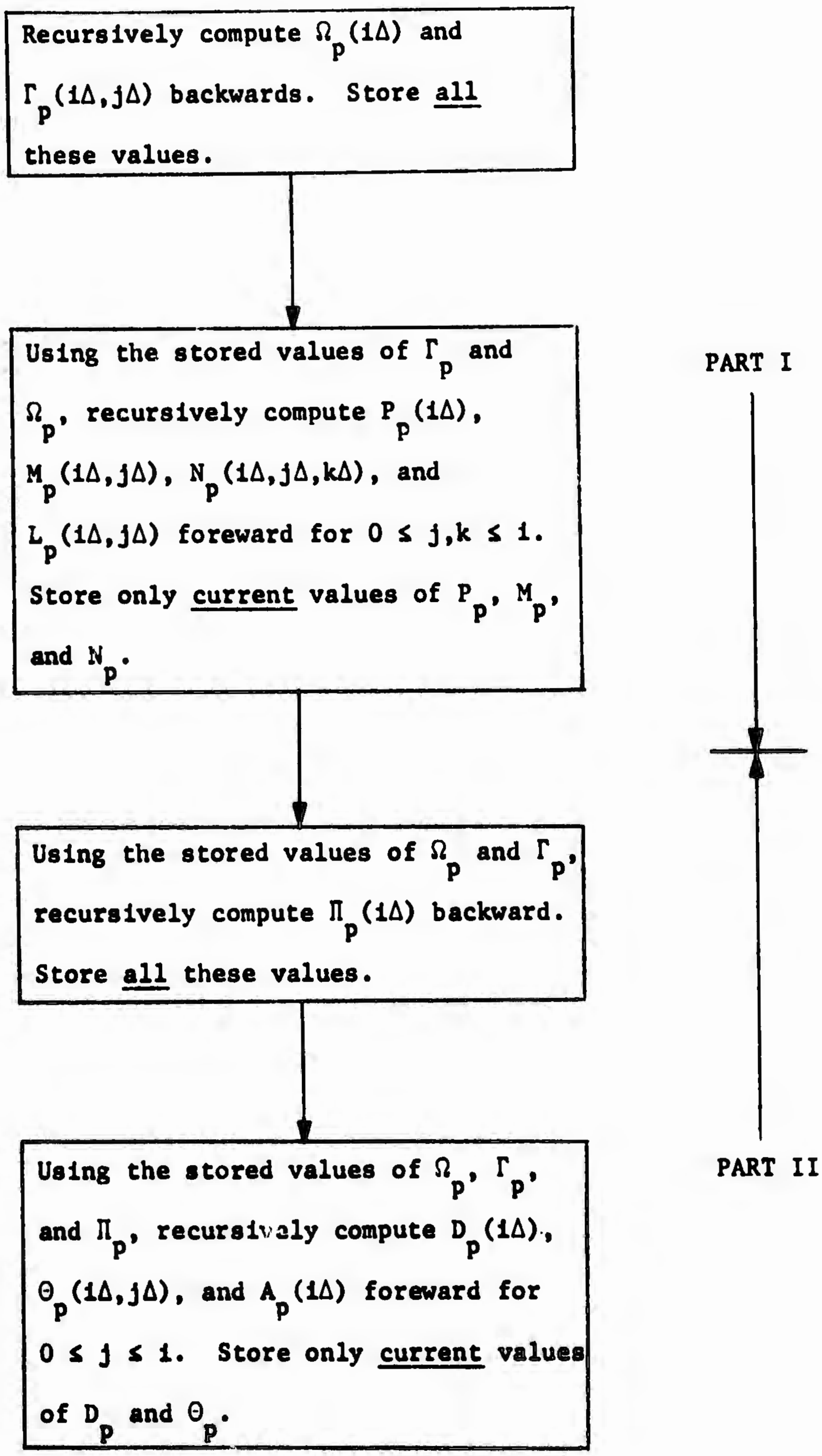

FIG. 2-3 DIFFERENTIAL GAME ALGORITHM

of the differential nature of the game.

If this differential game algorithm converges, the point to which it converges represents a solution by construction of the algorithm. It should be pointed out, however, that this algorithm has not been tested numerically and that nothing is known about its convergence properties. But it is equivalent, to within second-order terms, to the multistage game algorithm presented earlier, as applied to the discretized version of the differential game. Both algorithms proceed by computing alternately optimizing strategies. Since the available numerical evidence indicates that the multistage game algorithm converges under a fairly wide range of circumstances, it is reasonable to hope that the differential game algorithm would do likewise.

D. Summary

Algorithms were suggested for solving the two-point boundary value problems in terms of whose solutions the minimax strategies for the multistage and differential games are expressed. In both the multistage and differential cases, these algorithms consisted of successively computing alternately optimizing strategies. A significant reduction in the storage requirement was found possible in the case of the differential game if advantage were taken of the differential nature of the game.

The multistage game algorithm was tested numerically for a variety of two-stage scalar examples and was found to converge in several but not all cases. It was shown, however, that the apparent failure of this algorithm to converge does not necessarily mean that the evader can "escape." Since the inability of the evader to escape is not known to imply the existence of a minimax solution, it is still an open question whether the nonconvergence of this algorithm implies the nonexistence of a solution.

3. QUASI-OPTIMAL STRATEGIES

It is evident from the preceding chapter that the numerical determination of the minimax strategies for the type of stochastic game at hand here is fraught with formidable if not insuperable difficulties. Aside from these difficulties, moreover, the implementation of these strategies is not feasible by conventional means because they are infinite-dimensional. For these reasons, it is of considerable interest to examine the possibility of constructing strategies for these games which are simpler to determine and implement but which still perform reasonably well. Since this examination involves comparing the performances of optimal strategies and non-optimal ones, the first part of this section will be devoted to determining the performance of the minimax strategies in this type of game. The approach used in this determination generalizes naturally to the case of non-optimal strategies. The consideration of a special class of linear non-optimal strategies also leads to easily calculated upper and lower bounds for the minimax performance.

A. The Minimax Performance

Formulas for determining the minimax value of the criterion J can be obtained in a fairly straightforward way by considering the stochastic differential game as the limiting form of an approximating sequence of multistage games. In any multistage game of the type considered in Appendix A, the well-known (see Bryson and Ho [3] for example) results on the optimal performance of linear-quadratic-Gaussian control systems can be applied to either the pursuer's or evader's associated optimal control problem to compute the minimax performance, assuming that the minimax control laws have been found. That is, the minimax strategy of one of the two players is substituted into the transition and criterion equations, resulting in a linear-quadratic-Gaussian stochastic optimal control problem from the point of view of the other player. Since the other player's minimax strategy must be a solution to this associated optimal control problem, the minimax performance (or cost) of this game can be evaluated by applying the known control theoretic results to this control problem. Notice that two expressions can be obtained in this way for the minimax performance, since this procedure can be applied from either player's point of view.

The next step is to consider the particular subclass of such multistage games that arises in the discretization of differential games. As in the determination of the minimax strategies, the equations for determining the minimax cost can be considerably simplified for small discretization step sizes by eliminating second and higher order terms. The result for the differential game is then obtained by taking the limiting form of these simplified formulas as the discretization interval becomes infinitesimal.

Once the method used in the appendices for obtaining the minimax strategies is established, the application of the above procedure is essentially a tedious but straightforward repetition of these manipulations.

Therefore, the formulas for calculating the minimax value of the criterion J for the differential game will be presented here without the details of their derivation. One of the two expressions for this minimax value for the stochastic differential game defined in Section 1 is (suppressing the "t" argument for functions of t only):

$$J_{minimax} = \frac{1}{2}\,\bar{x}_o^T\left[Y_{po} + \Pi_{po} + \Pi_{po}^T + \int_0^{t_f} \Pi_p^T(T_eA_e - T_p\Pi_p)\,dt\right]\bar{x}_o$$

$$+ \frac{1}{2}\,\mathrm{Tr}\int_0^{t_f}\Big\{Y_pQ + Y_pT_pY_pP_p + S_{p3}(t,t,t)\,R_e$$

$$+ \int_0^t \Gamma_p^T(t,\xi)T_p\Gamma_p(t,\xi)R_e(\xi)d\xi + Y_pT_p\int_0^t \Gamma_p(t,\xi)M_p(t,\xi)d\xi$$

$$+ \int_0^t M_p^T(t,\xi)\Gamma_p^T(t,\xi)T_pY_p\,d\xi$$

$$+ \int_0^t\int_0^t \Gamma_p^T(t,\xi)T_p\Gamma_p(t,\sigma)N_p(t,\xi,\sigma)\,d\xi\,d\sigma\Big\}\,dt, \qquad (3:1)$$

where S_{p3} is determined by the equation

$$\begin{cases} \dfrac{\partial S_{p3}(t,\tau,\sigma)}{\partial t} = \Gamma_p^T(t,\tau)[T_p - T_e]\Gamma_p(t,\sigma) + [\Gamma_p^T(t,\tau) - \Lambda_e^T(t,\tau)]\,T_e[\Gamma_p(t,\sigma) - \Lambda_e(t,\sigma)] \\ S_{p3}(t_f,\tau,\sigma) = 0; \quad 0 \le \tau,\sigma \le t_f. \end{cases} \qquad (3:2)$$

This expression is obtained from the control-theoretic analysis of the pursuer's associated optimal control problem. The formulas for the minimax cost that arise from the evader's associated optimal control problem are exactly the same except that all the "p" and "e" subscripts are interchanged. By definition, the values of these two expressions are the same (the minimax value of J) if the parameters in them constitute a solution to the primary and subsidiary equation systems (i.e., represent a solution to the game).

Notice that the evaluation of either of these expressions for the minimax performance entails the solution of the primary and subsidiary equation systems. This means that it is in effect necessary to determine the optimal strategies in order to evaluate their performance, which is not surprising since it is also true of the corresponding type of stochastic optimal control problem. For this reason, these expressions are of practical importance only in cases where one is willing to calculate, if not implement, the minimax strategies for a game of this type.

B. Bounds on the Minimax Performance

Suppose now that the pursuer has decided upon some possible non-optimal admissible strategy U^*. If this strategy is substituted into the transition equation and criterion equation, a stochastic optimal control problem faces the evader. Defining the quantity

$$J^+(U^*) \triangleq \max_V J(U^*, V),$$

$J^+(U^*)$ has the interpretation of being the minimum value of J that the pursuer (who wishes to minimize J) can guarantee by using the strategy U^*. Similarly, the quantity

$$J^-(V^*) = \min_U J(U, V^*)$$

is the maximum value of the criterion that the evader can guarantee by adopting the evasion strategy V^*.

If a pair of minimax strategies U^o and V^o exist for this game, then for any U^* and V^*,

$$J^+(U^*) \geq J(U^*, V^o) \geq J^o \geq J(U^o, V^*) \geq J^-(V^*),$$

where J^o designates $J(U^o, V^o)$. The outer two inequalities hold as a consequence of the definitions of J^+ and J^-; the inner inequalities follow from the saddle point condition.

The utility of the concept of J^+ and J^- stems from the fact that these quantities are <u>easy to compute</u> for the type of game considered here if U^* and V^* are taken to be <u>finite-dimensional linear strategies</u>. This means that upper and lower bounds can be readily computed for the minimax value of the criterion <u>without finding the minimax strategies</u>. How tight these bounds are, of course, depends on the choice of U^* and V^*.

This concept also provides a way of evaluating non-optimal strategies which takes into account the element of <u>conflict</u> in a game. In the evaluation of a pursuit strategy U, for instance, $J^+(U)$ is the value of the criterion that would result if the evader knows that the pursuer is using this strategy and takes full advantage of this knowledge to achieve his own goal. Since the evader's goal is in complete conflict with the pursuer's, this advantage to the evader is a disadvantage to the pursuer. $J^-(V)$ is a similar evaluation of an evasion strategy V.

These evaluations of pursuit and evasion strategies are based on pessimistic assumptions about the action of the opposing player, but this pessimism is justified by the fact that the players' goals are in complete conflict. Moreover, this evaluation scheme has the appealing properties that, for minimax strategies U^o and V^o,

$$J^+(U^o) = J^o = J^-(V^o)$$

and

$$\left.\begin{array}{l} J^+(U^*) \geq J^o \\ J^-(V^*) \leq J^o \end{array}\right\} \text{ for any other admissible } U^* \text{ and } V^* .$$

These are properties that any reasonable evaluation scheme should certainly possess, and depend only on the fact that this is a zero-sum game. That is, this method of evaluating strategies and establishing bounds on the minimax value of the criterion (if such a value exists) is applicable to zero-sum games in general.

As a final remark, however, it should be pointed out that, for the class of games considered here, it is not known whether the existence (and finiteness) of $J^+(U)$ and $J^-(V)$ for some pursuit and evasion strategies U and V implies that minimax strategies exist. In other words, it might be in some games of this sort that pursuit and evasion strategies can be found such that J^+ and J^- are finite, but where no minimax value of the criterion J is defined. In such an eventuality, non-trivial performance bounds would exist, but no minimax performance value.

C. Reasonable Quasi-Optimal Strategies

Turning now to the question of finding simple but good strategies for the stochastic differential game defined in Section 1, it is convenient to begin with a consideration of the certainty-coincidence property possessed by this class of games. It is shown in Willman [1] that, as a consequence of this property, the minimax control laws for this game, if they exist, are of the form

$$U^o: \left\{ \begin{aligned} u &= -B^{-1}G_p^T[S\hat{x}_p + \text{error terms}] \\ \dot{\hat{x}}_p &= [T_e - T_p]S\hat{x}_p + P_pH_p^TR_p^{-1}[z_p - H_p\hat{x}_p] + \text{error terms} \\ \hat{x}_p(0) &= \bar{x}_o \end{aligned} \right\} \text{(pursuer)}$$

$$V^o: \left\{ \begin{aligned} v &= -C^{-1}G_e^T[S\hat{x}_e + \text{error terms}] \\ \dot{\hat{x}}_e &= [T_e - T_p]S\hat{x}_e + P_eH_e^TR_e^{-1}[z_e - H_e\hat{x}_e] + \text{error terms} \\ \hat{x}_e(0) &= \bar{x}_o \end{aligned} \right\} \text{(evader)}$$

where the matrix time function S is the same as in the corresponding deterministic game, and where P_p and P_e are respectively the covariances of the pursuer's and evader's estimates $\hat{x}_p$ and $\hat{x}_e$ of the current state, under the assumption that both players are using these strategies. The "error terms" have the property of being identically zero on the mean sample path (i.e., when $x(0) = \bar{x}_o$ and $w_p \equiv w_e \equiv \xi \equiv 0$) This means that these error terms represent noise-induced discrepancies between the behavior of the stochastic game and that of its deterministic counterpart.

The reason for considering the certainty-coincidence property at this point is that it makes it possible to immediately single out a special class of quasi-optimal strategies that can reasonably be expected to perform well. Since the error terms in the preceding equations only represent noise-induced deviations, it is reasonable for the pursuer to adopt a "certainty-equivalent" strategy of the form

$$U_{ce}: \left\{ \begin{array}{l} u = -B^{-1}G_p^T S\hat{x}_p \\ \dot{\hat{x}}_p = [T_e - T_p]S\hat{x}_p + P_p H_p^T R_p^{-1}[z_p - H_p\hat{x}_p] \\ \hat{x}_p(0) = \bar{x}_o \end{array} \right\}, \tag{3:3}$$

and similarly for the evader. These strategies are formed from the minimax strategies by dropping the error terms and considering P_p (and P_e in the corresponding evasion strategy) as undetermined parameters. In the minimax strategies, these parameters could be interpreted as the covariances of the pursuer's and evader's estimates of the state, but this interpretation does not extend to the present context.

In order to evaluate these certainty-equivalent strategies, it is necessary to calculate the extent to which the opposing players can optimize against them. Concentrating on the pursuer for the moment, it is easy to verify that, for any time function $P_p(t)$, the substitution of his certainty-equivalent strategy into the transition equation (1:1) and criterion equation (1:4) leads to the following situation from the evader's point of view:

$$\begin{bmatrix} \dot{x} \\ \hline \dot{x}-\dot{\hat{x}}_p \end{bmatrix} = \left[\begin{array}{c|c} -T_pS & T_pS \\ \hline -T_eS & T_eS - P_pQ_p \end{array}\right] \begin{bmatrix} x \\ \hline x-\hat{x}_p \end{bmatrix} - \begin{bmatrix} G_e \\ \hline G_e \end{bmatrix} v + \left[\begin{array}{c|c} I & 0 \\ \hline I & -P_pH_p^TR_p^{-1} \end{array}\right] \begin{bmatrix} \xi \\ \hline w_p \end{bmatrix} \quad \text{(transition equation)}$$

$$J = \frac{1}{2}\,\varepsilon \left\{ [x_f \,|\, x_f - \hat{x}_{pf}] \left[\begin{array}{c|c} S_f & 0 \\ \hline 0 & 0 \end{array}\right] \begin{bmatrix} x \\ \hline x-\hat{x}_p \end{bmatrix} + \int_0^{t_f} [x \,|\, x-\hat{x}_p] \left[\begin{array}{c|c} ST_pS & -ST_pS \\ \hline -ST_pS & ST_pS \end{array}\right] \begin{bmatrix} x \\ \hline x-\hat{x}_p \end{bmatrix} - v^TCv\,dt \right\} \quad \text{(criterion to be maximized)}$$

$$z_e = [H_e \,\vdots\, 0]\begin{bmatrix} x \\ --- \\ x-\hat{x}_p \end{bmatrix} + w_e \qquad \text{(available measurements)}$$

This represents a standard type of linear-quadratic-Gaussian stochastic optimal control problem, the solution to which is well known. In this particular case, the manipulations developed by Behn and Ho [4] can be used to express this solution in the form:

$$\left.\begin{aligned}
v &= -C^{-1}G_e^T[S\hat{x}_e+\Gamma_e\hat{\epsilon}_e] \\
\dot{\hat{x}}_e &= [T_e-T_p]S\hat{x}_e+[T_pS+T_e\Gamma_e]\hat{\epsilon}_e+P_{e1}H_e^TR_e^{-1}[z_e-H_e\hat{x}_e];\ \hat{x}_e(0)=\bar{x}_o \\
\dot{\hat{\epsilon}} &= [T_e\Gamma_e-P_pQ_p]\hat{\epsilon}_e+P_{e2}^TH_e^TR_e^{-1}[z_e-H_e\hat{x}_e];\ \hat{\epsilon}_e(0)=0
\end{aligned}\right\} \quad (3:4)$$

where $\hat{x}_e$ and $\hat{\epsilon}_e$ are the Kalman filter estimates of x and $\epsilon_e \triangleq x-\hat{x}_p$, and where

$$\dot{S} = S[T_p-T_e]S;\ S(t_f)=S_f \qquad \text{(i.e., this is the same "S" as before)}$$

$$\left.\begin{aligned}
\dot{P}_{e1} &= T_pS[P_{e2}^T-P_{e1}]+[P_{e2}-P_{e1}]ST_p-P_{e1}Q_eP_{e1}+Q;\ P_{e1}(0)=P_o \\
\dot{P}_{e2} &= T_pS[P_{e3}-P_{e2}^T]+[P_{e2}-P_{e1}]ST_e-P_{e2}Q_pP_p \\
&\quad -P_{e1}Q_eP_{e2}+Q;\ P_{e2}(0)=P_o \\
\dot{P}_{e3} &= T_eS[P_{e3}-P_{e2}]+[P_{e3}-P_{e2}^T]ST_e-P_pQ_pP_{e3} \\
&\quad -P_{e3}Q_pP_p-P_{e2}^TQ_eP_{e2}+Q+P_pQ_pP_p;\ P_{e3}(0)=P_o \\
\dot{\Gamma}_e &= -\Gamma_eT_e\Gamma_e+\Gamma_e[P_pQ_p-T_eS]+[Q_pP_p-ST_e]\Gamma_e-ST_pS;\ \Gamma_e(t_f)=0
\end{aligned}\right\} \quad (3:5)$$

The matrix Γ_e is the same feedback gain as that found in Behn and Ho [4] for the case where the evader has perfect measurements, which is explained by the certainty-equivalence principle.

Applying the standard formula given in Bryson and Ho [3] for the performance of this optimal opposing evasion strategy yields the following expression for $J^+(U_{ce})$:

$$J^+(U_{ce}) = \frac{1}{2}\,\bar{x}_0^T S_o \bar{x}_o + \frac{1}{2}\,\mathrm{Tr}\left\{ S_o P_o + \Gamma_{eo} P_o + \int_0^{t_f} (SQ + \Gamma_e Q + P_p Q_p P_p \Gamma_e)\,dt \right.$$

$$\left. - \int_0^{t_f} T_e (SP_{e1} S + SP_{e2}\Gamma_e + \Gamma_e P_{e2}^T S + \Gamma_e P_{e3} \Gamma_e)\,dt \right\} \qquad (3:6)$$

An entirely analogous analysis can be carried out for the optimal pursuit strategy opposing a certainty-equivalent evasion strategy V_{ce} for an arbitrary measurement weighting function $P_e(t)$.

All of the preceding formulas depend on the as yet undetermined matrix time function $P_p(t)$ (or $P_e(t)$ in the case of the evader). Several methods for choosing these quantities that suggest themselves are the following:

(1) Solve the nonlinear deterministic optimal control problem with Γ_e, P_{e1}, P_{e2}, and P_{e3} as state variables, P_p as the control variable, and $J^+(U_{ce})$ as the criterion to be minimized. This approach, although it gives the best quasi-optimal strategy of this form as evaluated by J^+, requires the solution of a matrix two-point boundary value problem.

(2) Since P_p is the covariance of the estimate in the minimax strategy, select P_p to be the covariance of the pursuer's estimate of the state <u>under the simplifying assumption</u> that the evader has perfect measurements and is using his <u>deterministic</u> minimax strategy. Under these assumptions, P_p is generated by the equation

$$\dot{P}_p = T_e S P_p + P_p S T_e - P_p Q_p P_p + Q ; \quad P_p(0) = P_o .$$

(3) Since this is only a quasi-optimal strategy anyway, assume that there is no great loss in restricting $P_p(t)$ to be a constant. Do a parameter search to determine the best (in terms of minimizing J^+) value of this constant.

Several of these methods are applied in the next section to a hopefully typical numerical example.

All of the quasi-optimal strategies considered thus far have been of the "certainty-equivalence" type, in which the presence of the "error terms" in the minimax strategies is ignored. There is, of course, an infinite variety of other types of possible approximations to the minimax strategies that could be applied. One of these other possibilities that immediately suggests itself is a refinement of a certainty-equivalent strategy in which the infinite-dimensional error term generation in a minimax strategy is approximated by a finite-dimensional approximation.

Such a higher order quasi-optimal approximation is also examined for the numerical example in the next section. For simplicity, the "second-order" approximation considered here is <u>taken</u> to be of the type that optimizes against the certainty-equivalent approximation. This means that, in the case of the evader, for example, this quasi-optimal strategy is of the form described by equation (3:4). Notice, however, that equation system (3:4) does not in itself specify the values of the parameters $P_p(t)$, $\Gamma_e(t)$, $P_{e1}(t)$, and $P_{e2}(t)$ to be used in this evasion strategy. Since this type of strategy is being considered as an approximation to the minimax evasion strategy instead of being used to optimize against a simpler pursuit strategy, it is not reasonable to use equation system (3:5) to determine these unspecified parameters. Instead, for the purposes of this approximation, P_p and Γ_e are <u>chosen arbitrarily</u> to be those values obtained in minimizing J^+ for the certainty-equivalent approximation discussed previously. The parameters P_{e1} and P_{e2} are considered free parameters to be selected to minimize J^- for this second-order quasi-optimal strategy.

In evaluating J^- for a second-order approximation of this type, it is again necessary to compute the extent to which the pursuer can minimize against this evasion strategy. If the above evasion strategy is substituted into the game dynamics and criterion functional, it is straightforward to find that the stochastic optimal control problem facing the pursuer is again of the standard "linear-quadratic-Gaussian" type. As in the case of the certainty-equivalent approximation, this control problem can be solved, and the optimal performance determined, in terms of the free parameters P_{e1} and P_{e2}, thereby determining J^- for this evasion strategy in terms of these parameters. The details of this optimization will not be carried out here since it is conceptually a repetition of the procedure used to determine equations (3:4) - (3:6).

Once the evader's "second-order" quasi-optimal strategy is chosen to be of this form, then, all that remains in finding the best strategy of this type is to determine $P_{e1}(t)$ and $P_{e2}(t)$ to maximize J^-. A second-order approximation of this kind can be defined similarly for the pursuer, in which two free parameters (actually time functions) need to be determined to minimize J^+. The effectiveness of these various quasi-optimal strategies is examined numerically in the next section.

4. A NUMERICAL EXAMPLE

A specific example of a stochastic differential game is introduced in this section for the purpose of testing some of the conjectures made earlier about quasi-optimal strategy. The particular example presented here, which hopefully is realistic enough to be typical and yet simple enough to be examined readily, is an interception problem involving two second-order dynamic systems.

A. Problem Formulation

The classical interception problem concerns the lateral maneuvers of a pursuer and an evader with acceleration control, approaching each other in space on a nominal collision course. A one-dimensional example of such a problem is considered here for simplicity, although it should be pointed out that a higher dimensional interception problem reduces to a set of independent one-dimensional problems for each lateral axis if there is no coupling between the dynamics and measurements in the various dimensions. This reduction is demonstrated in Behn and Ho [4].

Specifically, the interception game considered here is described in reduced state space (i.e., relative coordinates between the players) by the equations:

$$\left.\begin{aligned}\dot{x}_1 &= x_2\\ \dot{x}_2 &= u - v + \zeta\end{aligned}\right\}\ \text{(dynamics)}\ \left\{\begin{aligned}&x_1: && \text{relative position}\\ &x_2: && \text{relative velocity}\\ &u: && \text{pursuer's control}\\ &v: && \text{evader's control}\end{aligned}\right.$$

$$J = \frac{1}{2}\,\varepsilon\left\{x^2(t_f) + \int_0^{t_f} (bu^2 - cv^2)dt\right\} \quad \text{(criterion)}$$

The players have the following position and velocity measurements:

$$\begin{aligned}z_{p1} &= x_1 + w_{p1}\\ z_{p2} &= x_2 + w_{p2}\end{aligned} \quad \text{(pursuer's measurements)}$$

$$\begin{aligned}z_{e1} &= x_1 + w_{e1}\\ z_{e2} &= x_2 + w_{e2}\end{aligned} \quad \text{(evader's measurements)}$$

where

$$\begin{bmatrix}\zeta\\ \hline w_{p1}\\ \hline w_{p2}\\ \hline w_{e1}\\ \hline w_{e2}\end{bmatrix} \text{ is Gaussian white noise } \left(\begin{bmatrix}0\\ \hline 0\\ \hline 0\\ \hline 0\\ \hline 0\end{bmatrix},\ \left[\begin{array}{c|c|c|c|c} q & 0 & 0 & 0 & 0\\ \hline 0 & r_{p1} & 0 & 0 & 0\\ \hline 0 & 0 & r_{p2} & 0 & 0\\ \hline 0 & 0 & 0 & r_{e1} & 0\\ \hline 0 & 0 & 0 & 0 & r_{e2}\end{array}\right]\right)$$

and where the prior distribution of the initial state is the same for both players and is

$$\text{Normal}\left(\begin{bmatrix}\bar{x}_{10}\\ \hline \bar{x}_{20}\end{bmatrix},\ \left[\begin{array}{c|c} p_{11} & p_{12}\\ \hline p_{21} & p_{22}\end{array}\right]\right)$$

and independent of the process and measurement noises. The parameters q, r_{p1}, r_{p2}, r_{e1}, r_{e2}, b and c are constants.

It is possible to formulate this game in terms of a scalar-state variable by using the concept of "predicted terminal miss" developed by Ho, Bryson, and Baron [2]. Defining the variable

$$y \triangleq x_1 + (t_f - t)x_2 ,$$

the dynamics can be expressed as

$$\dot{y} = (t_f - t)(u - v) + \xi, \text{ where } \xi (\triangleq (t_f - t)\zeta) \text{ is a GWN}(0, q[t_f - t]^2) \text{ process,}$$

and the criterion as

$$J = \frac{1}{2}\, \varepsilon \left\{ y^2(t_f) + \int_0^{t_f} (bu^2 - cv^2)\, dt \right\} .$$

Since $[z_{p1} + (t_f - t)z_{p2}]$ is a sufficient statistic of the pursuer's measurements for y, and similarly for the evader, the measurements can be taken as

$$z_p = y + w_p \qquad \text{(pursuer)}$$

and

$$z_e = y + w_e \qquad \text{(evader)}$$

where

$$w_p = w_{p1} + (t_f - t)w_{p2}$$

and

$$w_e = w_{e1} + (t_f - t)w_{e2} .$$

Therefore, the vector $\begin{bmatrix} \xi \\ w_p \\ w_e \end{bmatrix}$ is a zero-mean Gaussian white noise

process with spectral density parameter

$$\left[\begin{array}{c|c|c} q(t_f-t)^2 & 0 & 0 \\ \hline 0 & r_{p1}+r_{p2}(t_f-t)^2 & 0 \\ \hline 0 & 0 & r_{e1}+r_{e2}(t_f-t)^2 \end{array}\right].$$

The prior on y(0) is Normal($\bar{y}_o$, p_o), where

$$\bar{y}_o = \bar{x}_{10} + \bar{x}_{20}t_f^2$$

and

$$p_o = p_{11} + 2t_f p_{12} + p_{22}t_f^2$$

by the standard rules for linear transformations of Normal random variables.

The numerical values chosen for the parameters of the game in this particular example are:

$t_f = 10$	
$b = .04$	pursuer's control penalty
$c = .1$	evader's control penalty
$q = 1$	process noise magnitude
$r_{p1} = 10$, $r_{p2} = .1$	pursuer's measurement noise magnitude
$r_{e1} = 100$, $r_{e2} = 1$	evader's measurement noise magnitude
$p_{11} = 100$, $p_{12} = 0$, $p_{22} = 1$	prior covariance of $\begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix}$

The mean of the prior is carried as an unspecified parameter. Notice that the above values of the parameters imply that the pursuer has a less stringent control penalty than the evader. Both players have velocity measurements which are better than their position measurements by the same factor, although the pursuer's measurements are more accurate than the evader's. When the game is transformed into a scalar problem in the manner indicated above, the prior variance of the initial "predicted terminal miss" becomes

$$p_o = 200 \; .$$

B. Certainty-Equivalent Quasi-Optimal Strategies

When this game is considered in its reduced scalar form, the various parameters introduced in earlier sections can be expressed explicitly as

$$G_p = G_e = 10 - t$$

$$T_p \triangleq G_p B^{-1} G_p^T = 25(10-t)^2$$

$$T_e \triangleq G_e C^{-1} G_e^T = 10(10-t)^2$$

$$H_p = H_e = 1$$

$$Q_p \triangleq H_p^T R_p^{-1} H_p = \frac{10}{100+(10-t)^2}$$

$$Q_e \triangleq H_e^T R_e^{-1} H_e = \frac{1}{100+(10-t)^2}$$

$$Q = (10-t)^2$$

In this example, it is also possible to calculate the quantity $S(t)$ analytically. Since S^{-1} obeys the differential equation

$$\frac{d}{dt}[S^{-1}] = T_e - T_p = -15(10-t)^2$$

and since $S^{-1}(t_f) = 1$,

$$S^{-1}(t) = 1 + 5(10-t)^3$$

and

$$S(t) = \frac{1}{1 + 5(10-t)^3} .$$

Therefore, it is possible to explicitly find the parts of J^+ and J^- for the certainty-equivalent quasi-optimal strategies which depend on the mean of the prior. From equation (3:6), this is just

$$\bar{x}_o^T S(0) \, x_o^T = \frac{\bar{x}_o^2}{5001}$$

for J^+ in this example. It turns out that the part of J^- depending on $\bar{x}_o$ is the same, as are also the parts of J^+ and J^- depending on this parameter for the second-order quasi-optimal strategies considered here. For this reason, the values of J^+ and J^- will henceforth be computed as if $\bar{x}_o$ were zero. For nonzero $\bar{x}_o$, this term can always be added to J^+ and J^-.

As mentioned in Section 3, the only latitude that exists in choosing a quasi-optimal strategy of the certainty-equivalent type (equations (3:5) and their counterparts for the evader) is the selection of the time functions $P_p(t)$ and $P_e(t)$. Of the three methods suggested in Section 3 for making this selection, only the second two are carried out here. The first method, which calls for the solution of a two-point boundary value problem, is not attempted because of its complexity.

The first of the remaining methods consists of calculating these parameters according to the equations:

$$\dot{P}_p = 2T_e S P_p - P_p^2 Q_p + Q; \quad P_p(0) = 200$$

and

$$\dot{P}_e = -2T_p S P_e - P_e^2 Q_e + Q; \quad P_e(0) = 200$$

These are the covariances of the players' estimates that would result under the simplifying assumption that their opponents have perfect measurements but are using their deterministic minimax strategies. The values of P_p and P_e determined by these equations are shown in Figure 4-1.

As it happens, the above simplifying assumptions lead to serious errors. The strategies determined on the basis of these assumptions perform extremely poorly when optimized against. The evader can make J arbitrarily large against the resulting pursuit strategy (i.e., $J^+ = \infty$) and $J^- = 2.7$. It will be shown shortly that much better bounds for the minimax performance, assuming it exists, can be obtained with certainty-equivalent approximations of this type.

The other method used here to obtain the parameters $P_p(t)$ and $P_e(t)$ is to consider, for simplicity, only constant functions. The best constant values are then determined empirically, which requires only a one-dimensional search in this case. For each value of P_p, equation systems (3:5) and (3:6) are used to evaluate $J^+(P_p)$, and similarly for P_e. The dependences of J^+ on P_p and J^- on P_e are shown in Figures 4-2 and 4-3.

As is apparent from these figures, P_p and P_e can be chosen respectively as 48 and 130 to bracket the minimax performance between a J^+ of 20.3 and a J^- of 8.56. The basic intuitive reasons that the

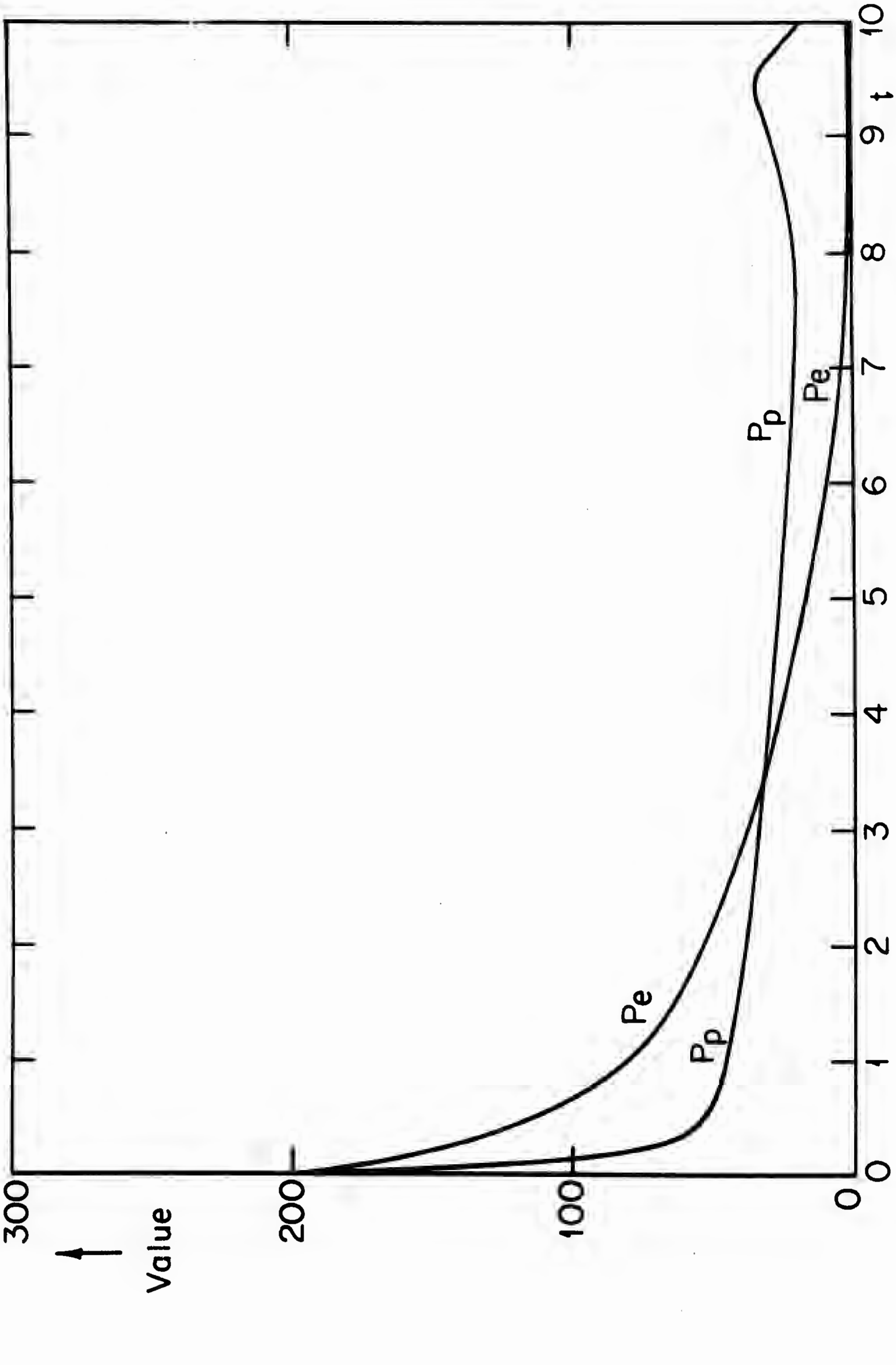

FIG. 4-1 "COVARIANCE" PARAMETERS

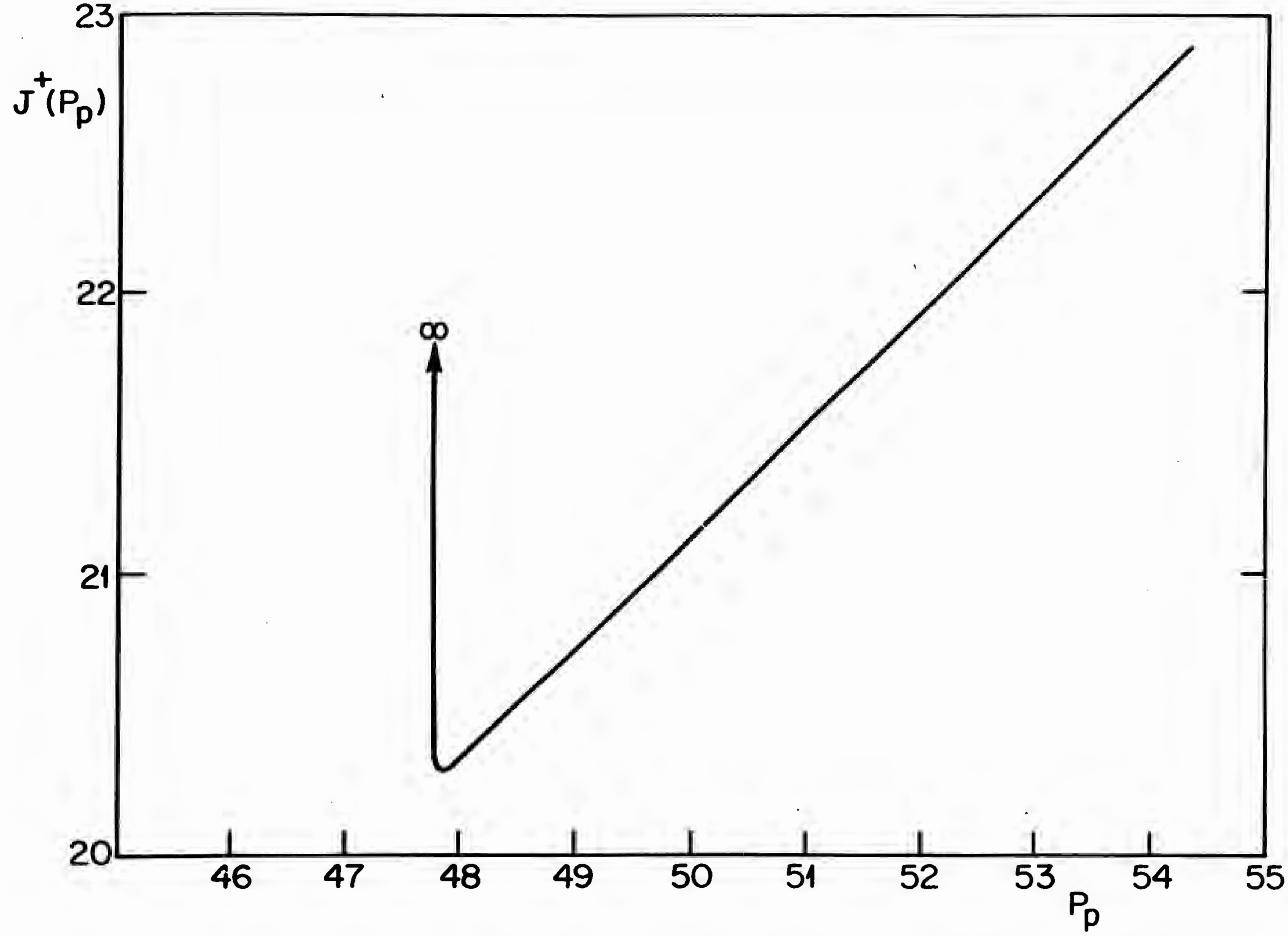

FIG. 4.2 PURSUER'S "CERTAINTY-EQUIVALENT" APPROXIMATION

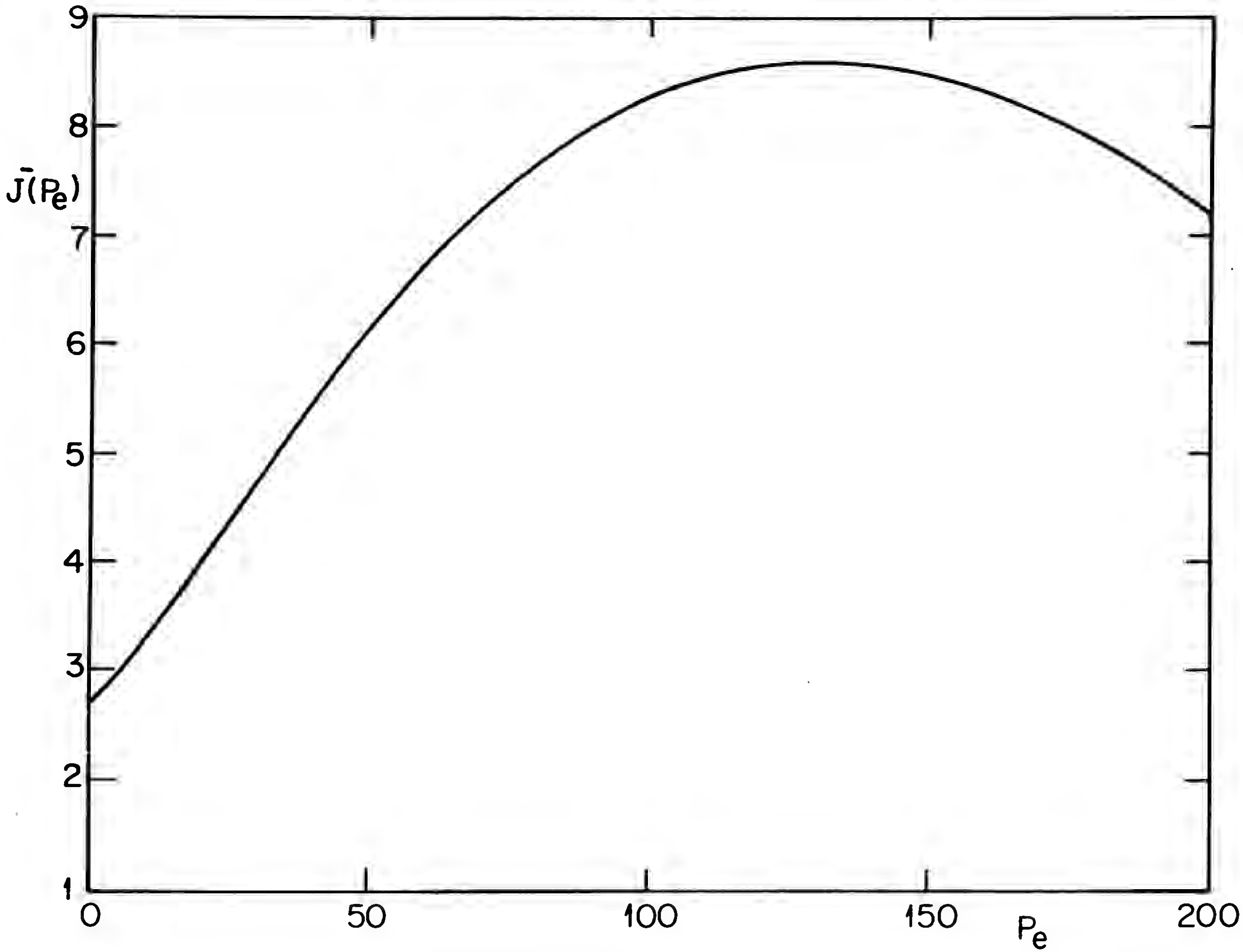

FIG. 4-3 EVADER'S "CERTAINTY-EQUIVALENT" APPROXIMATION

simplifying assumptions used earlier resulted in such inferior performance are:

(1) The pursuer underestimates the evader's ability to degrade his measurements and thus becomes overconfident in his estimate.

(2) The evader overestimates the pursuer's ability to estimate the state and hence does not make enough of an effort to avoid capture.

Comparing the parameter values obtained by these two methods, it is apparent that these assumptions lead to values of $P_p(t)$ and $P_e(t)$ which are too small on the average.

C. A More Complex Approximation

Another type of quasi-optimal strategy which was evaluated for this game is the second-order approximation described in Section 3. This approximation consists of an evasion strategy of the form described by equation system (3:4) and an analogous pursuit strategy. As explained in Section 3, the values of P_p and P_e used were $P_p(t) \equiv 48$ and $P_e(t) \equiv 130$. $\Gamma_e(t)$ was obtained from the $\dot{\Gamma}_e$ equation (3:5) with $P_p = 48$, and similarly for $\Gamma_p(t)$.

The parameters $P_{e1}(t)$, $P_{e2}(t)$, $P_{p1}(t)$, and $P_{p2}(t)$ in these quasi-optimal strategies are not determined, the idea being to select P_{e1} and P_{e2} to maximize J^-, and P_{p1} and P_{p2} to minimize J^+. Since these strategies are only quasi-optimal anyway, consideration was again limited for simplicity to constant values for these parameters.

The performances of these types of strategies are shown as a function of the free parameters in Figures 4-4 and 4-5. It can be seen from these figures that the best performances under these restrictions occur in the following circumstances:

$$\left.\begin{array}{l} P_{p1} = 31 \\ P_{p2} = -33 \end{array}\right\} J^{+} = 18.2 \quad \text{and} \quad \left.\begin{array}{l} P_{e1} = -25 \\ P_{e2} = 109 \end{array}\right\} J^{-} = 15.5$$

This means that with two-dimensional strategies, even under the above restrictions, it is possible to bound the minimax performance (for $\bar{x}_o = 0$) between 15.5 and 18.2.

D. Performance and Complexity

Referring to an "n-dimensional linear pursuit strategy" as one of the form

$$\begin{cases} u = k^T y \\ \dot{y} = mz_p + d \\ y(0) = y_o \end{cases}$$

where y is an n-vector, and similarly for an "n-dimensional linear evasion strategy," it follows that the certainty-equivalent strategies in this example are one-dimensional linear strategies and that the second-order quasi-optimal strategies are two-dimensional ones. Of course, the particular types of quasi-optimal strategies considered here do not exhaust the possible range of one- and two-dimensional linear strategies for this example. Therefore, the J^- to J^+ ranges obtained for these two types of quasi-optimal strategies contain, but are not necessarily equal

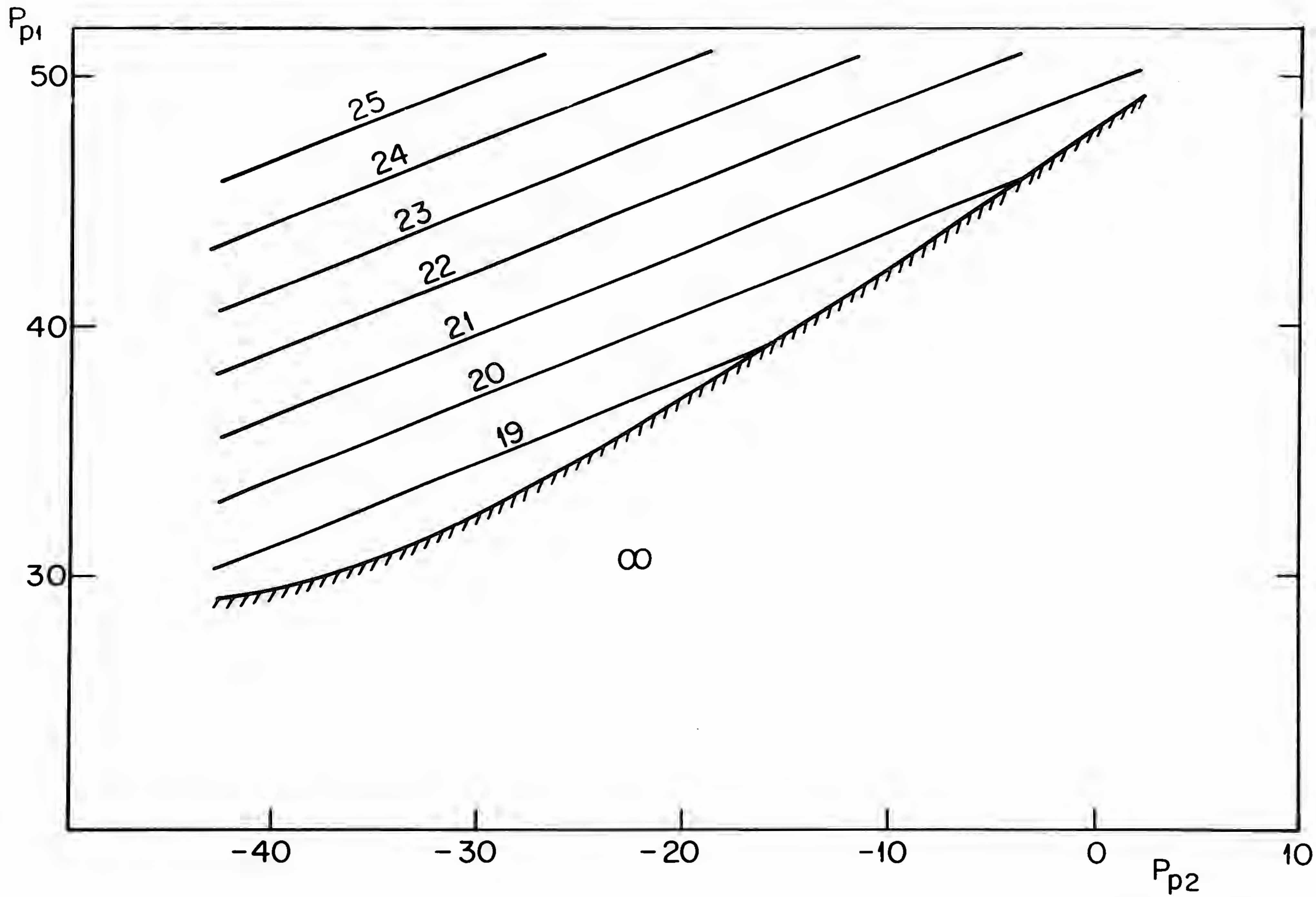

FIG. 4-4 PURSUER'S "SECOND ORDER" APPROXIMATION: $J^{+}(P_{p1}, P_{p2})$

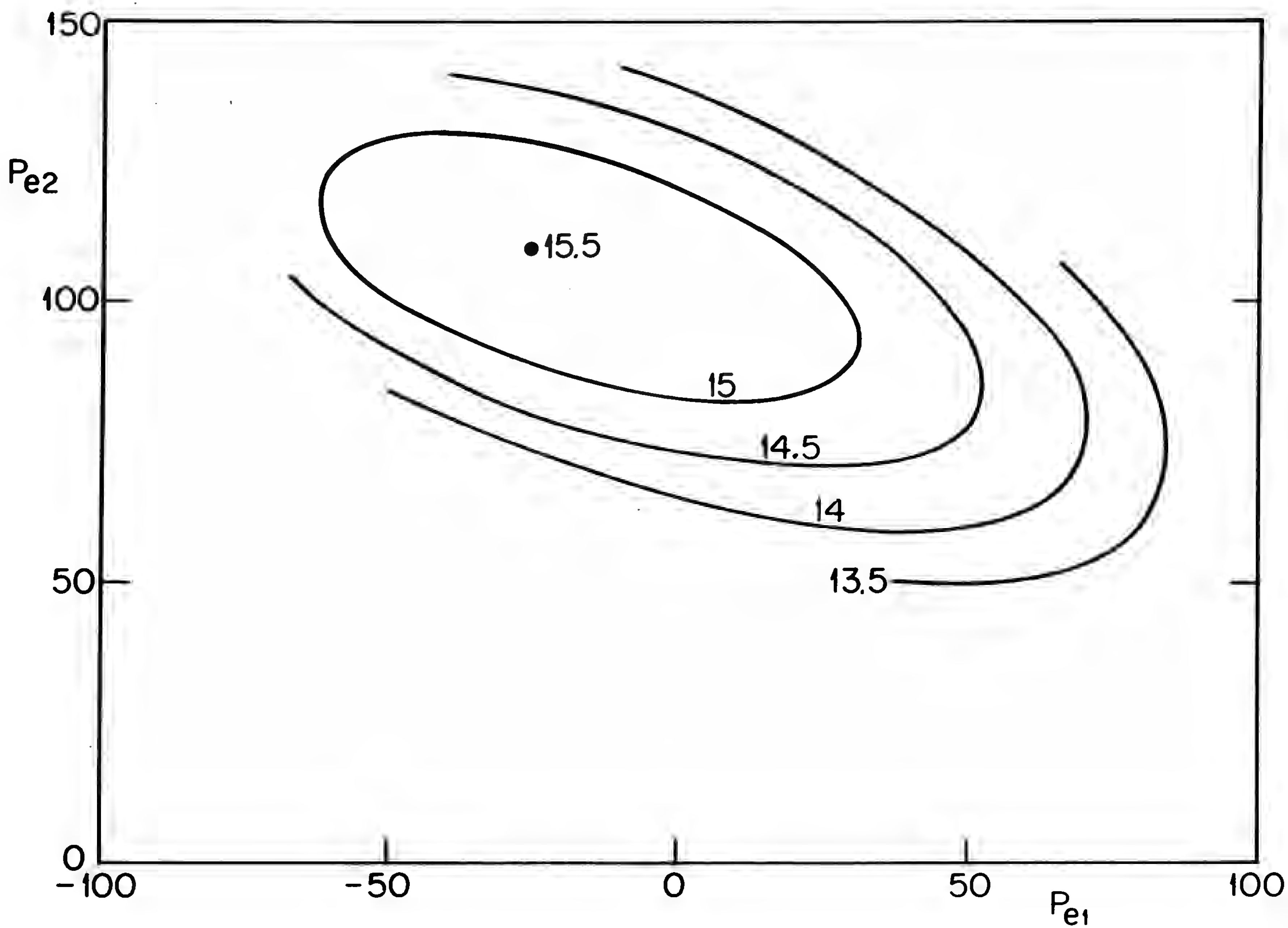

FIG. 4-5 EVADER'S "SECOND ORDER" APPROXIMATION: $J^{-}(P_{e1}, P_{e2})$

to, the <u>minimum</u> J^- to J^+ ranges obtainable with one- and two-dimensional strategies.

These containing ranges are shown in Figure 4-6 as a function of strategy dimension. A zero-dimensional range is also shown, a zero-dimensional strategy being one in which the available measurements are not used. This definition is made because an n-dimensional strategy is one in which an n-dimensional statistic of the available measurements is allowed to be used. The upper and lower end points of the ranges shown in this figure for one- and two-dimensional strategies are just the best values of J^+ and J^- found for the certainty-equivalent and second-order quasi-optimal strategies examined earlier. The range for zero-dimensional strategies was found by evaluating J^+ and J^- for the certainty-equivalent strategies with P_p and P_e equal to zero (so that the measurements are ignored). It is shown in Rhodes and Luenberger [5] that this is actually the best type of strategy to adopt if no measurements are available, so that the range shown for zero-dimensional strategies is the minimum range.

Using the dimensionality of linear strategies as a measure of their complexity, Figure 4-6 can be interpreted as an approximate depiction of the tradeoff between complexity and performance. Of course, these results are only for an isolated example. Furthermore, if the restricted classes of one- and two-dimensional quasi-optimal strategies examined here are not good choices, then the performances displayed in this figure are misleading in a conservative way. It is reassuring to know, however, that in this hopefully typical example a performance within fifteen percent of the minimax can be obtained by

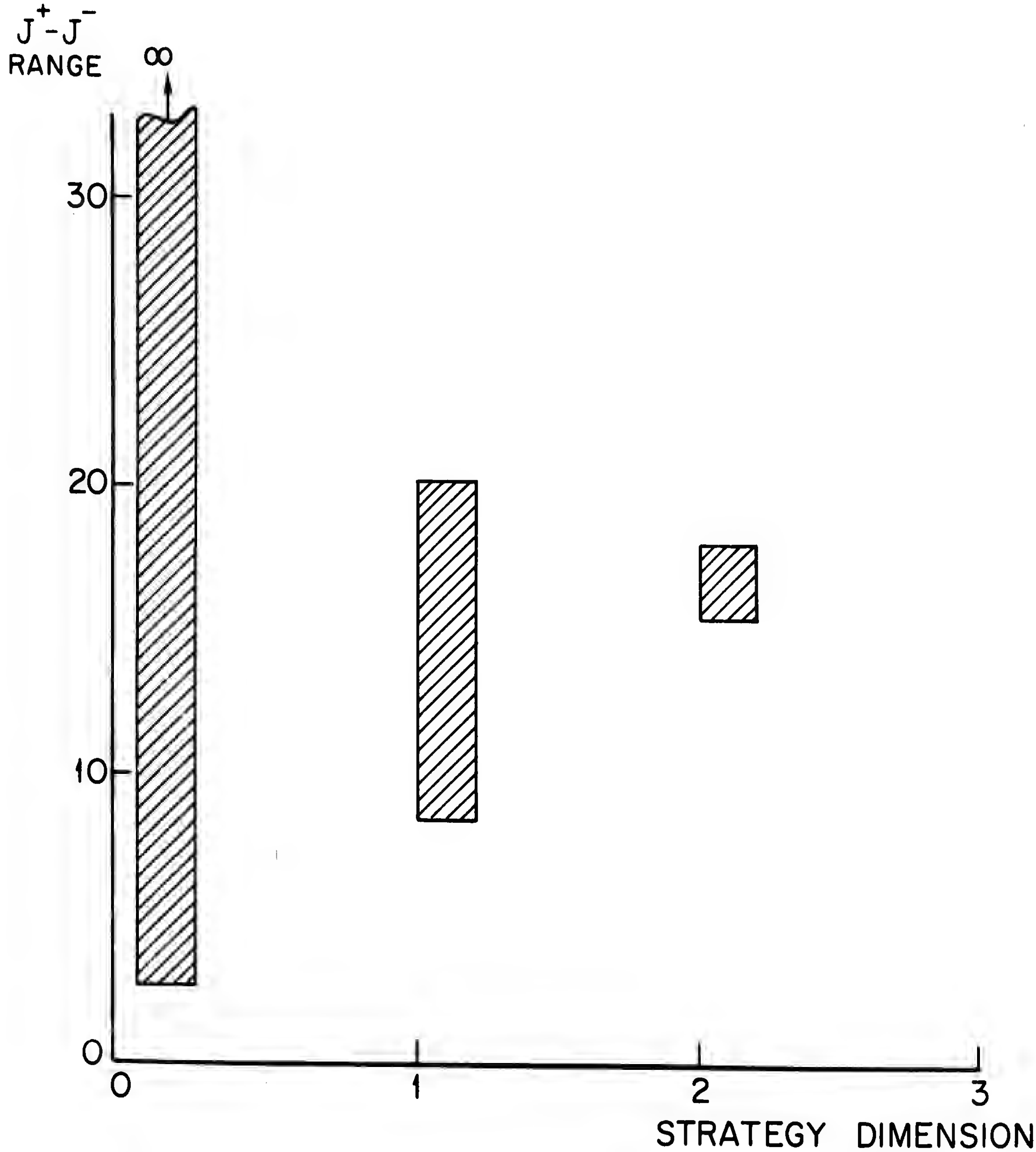

FIG. 4-6 PERFORMANCE VS. COMPLEXITY

either player with only a two-dimensional strategy. It is also interesting to note that this represents a significant improvement over the performance possible for the evader with only a one-dimensional strategy in this case, assuming that none of these is much better than the type of certainty-equivalent strategy evaluated here.

ACKNOWLEDGMENT

The author is very grateful to Professor Y. C. Ho of Harvard University for suggesting this course of research and for his generous help and encouragement during its development.

REFERENCES

[1] W. W. Willman, "Formal Solutions for a Class of Stochastic Pursuit-Evasion Games," Harvard University Technical Report TR 575, November 1968.

[2] Y. C. Ho, A. E. Bryson, Jr., and S. Baron, "Differential Games and Optimal Pursuit-Evasion Strategies," IEEE Transactions on Automatic Control, AC-10, No. 4, October 1965, pp. 385-389.

[3] A. E. Bryson, Jr. and Y. C. Ho, Optimization, Estimation and Control, Waltham, Mass., Blaisdell Company, to be published.

[4] R. D. Behn and Y. C. Ho, "On a Class of Linear Stochastic Differential Games," IEEE Transactions on Automatic Control, AC-13, No. 3, June 1968, pp. 227-239.

[5] I. B. Rhodes and D. G. Luenberger, "Differential Games with Imperfect State Information," to appear in IEEE Transactions on Automatic Control.

Unclassified
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) Division of Engineering and Applied Physics Harvard University Cambridge, Mass. 02138	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP
3. REPORT TITLE ON A CLASS OF STOCHASTIC PURSUIT-EVASION GAMES	
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Interim technical report	
5. AUTHOR(S) (First name, middle initial, last name) W. W. Willman	

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
April 1969	57	5

8a. CONTRACT OR GRANT NO. N00014-67-A-0298-0006 and NASA b. PROJECT NO. Grant NGR 22-007-068 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Technical Report No. 585 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted by the U. S. Government.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Office of Naval Research

13. ABSTRACT

This report continues the investigation of a class of stochastic differential games introduced in an earlier report by Willman [1]. Iterative algorithms for computing numerical solutions to games of this type are discussed. It is shown by numerical example that there exist multistage games analogous to this type of differential game for which minimax solutions do indeed exist. Several candidates for quasi-optimal strategies are presented which are simple to compute and easy to implement compared to the minimax strategies. A criterion is developed for evaluating the performance of non-optimal strategies. The performances of these quasi-optimal strategies are evaluated in terms of this criterion for a numerical example (an interception problem involving two second-order systems).

Unclassified
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Stochastic Differential Game Pursuit Evasion Quasi-optimal						

DD FORM 1 NOV 65 1473 (BACK)
S/N 0101-807-6821

Unclassified
Security Classification

A-31409

END