ARTICLE NO. 80-4045

Reprisal Strategies in Pursuit Games

Henry J. Kelley* and Eugene M. Cliff†
Optimization Incorporated, and Virginia Polytechnic Institute and State University, Blacksburg, Va.
and
Leon Lefton‡
Optimization Incorporated, Blacksburg, Va.

A preference-ordered discrete-gaming model, developed for pursuit-evasion studies in an earlier paper, is first reviewed and some refinements are described: graduation of draw outcomes via a threat-reciprocity concept and an improved cell buildup technique. A reprisal-strategy scheme, which exploits opponent's errors by extrapolation, is described and illustrated in an air-to-air combat example. The results suggest that the approach is quite promising.

Introduction

THE subject of differential games is still young,[1,2] but much progress has already been made on its theory, and enough done with simple examples to tempt engineers and operations analysts into applications work. Applications thinking has been dominated, quite properly, by the problem of obtaining approximate solutions, either by simplification of system models or of solution procedures. Little attention has been paid so far to the exploitation of solutions already in hand for closed-loop control purposes, i.e., real-time applictions. The present paper looks at the possibility of removing some of the conservatism inherent in gaming calculations in pursuit-evasion applications.

It is basic to differential gaming that each participant plays optimally, any departure by one's opponent being assumed momentary and not worth considering for planning purposes. Control policies based upon the extrapolation of an opponent's current departures from optimality are sometimes called "reprisal strategies," and have not been widely studied. One extreme is extrapolation of an opponent's trajectory, under some plausible assumption about his controls, for purposes of solving one's own pursuit (or evasion) problem as an optimal-control problem. The other extreme attempts to exploit a real or perceived advantage in system-delay time (the sum of information-processing and control delays) based upon short-term considerations. For this, there is an applicable theory of upper and lower games.[3] Attention is directed in the following mainly to the practically important cases intermediate between these extremes.

Discrete-Gaming Example

Some illustrative reprisal-strategy calculations will be presented for a preference-ordered discrete-gaming model of air combat employing 2-D constant-speed vehicle models with turn rate limits (Game-of-Two-Cars).[1,4] The discretization consists of restricted choice among a small number of closed-loop-guidance options for each vehicle with the possibility of the players switching choices at a preset interval—3 s in the examples of Ref. 4 and the present paper. Preference-ordering between win, lose, draw, disengagement, and mutual-capture outcomes is described in the reference cited. The system modeling presently employed is that described in Ref. 4. Some refinements have been made in the procedure for generating the active-cell structure, to be described in a later section.

Draw-Space Indices

Control logic for maneuvering in the draw region has received insufficient attention in the two-dimensional modeling of Ref. 4 and, for that matter, in air-combat modeling in general. The threat-reciprocity concept of Ref. 5 is of interest in this connection. An appropriate measure of generalized miss is the extended-weapon-envelope idea of Roberts and Montgomery[6] illustrated in Fig. 1. This offers miss measures suitable for use with the discrete-gaming model, the minima vs time of the generalized misses of the two combatants furnishing data for a scoring index in the draw region.

The functions Q_1 and Q_2 of the joint state describe extended-weapon-envelopes for the two players. $Q_1 \leq 0$ corresponds to the capture envelope of player #1 (Fig. 1). The function

$$\bar{Q}_1 = \min_{0 \leq t \leq t_f} Q_1(t)$$

is, for $\bar{Q}_1 \geq 0$, a sort of generalized closest-approach distance. Similarly,

$$\bar{Q}_2 = \min_{0 \leq t \leq t_f} Q_2(t)$$

The generalized misses $\bar{Q}_1$ and $\bar{Q}_2$ are defined for any control histories, not necessarily optimal. In particular, they are defined for the trajectory pairs corresponding to each matrix element for which the outcome is a draw.

The assumed preference-ordering is 1, 5, 3, 2 for player #1 and 2, 5, 3, 1 for player #2. A "1" outcome denotes capture by player #1, a "2" by #2, "3" denotes mutual capture, and "5" a draw. The special type of draw designated "4" in Ref. 4 will not figure in the example to be presented.

If $\bar{Q}_2 \gg 0$, player #1 can be aggressive and minimize $\bar{Q}_1$; however, if $\bar{Q}_2 \cong 0$, then player #1 must evade and maximize $\bar{Q}_2$. If $\bar{Q}_1 \gg 0$, then #2 can be aggressive and minimize $\bar{Q}_2$; however, if $\bar{Q}_1 \cong 0$, then player #2 must evade and maximize $\bar{Q}_1$. Thus, there are four possible qualitative "states" and associated desires within the draw region.

Received Dec. 14, 1978; revision received April 23, 1979.

Index categories: Guidance and Control; Analytical and Numerical Methods.

*President (OI) and Professor of Aerospace Engineering (VPI). Associate Fellow AIAA.

†Senior Programmer/Analyst (OI) and Associate Professor of Aerospace Engineering (VPI). Member AIAA.

‡Senior Programmer/Analyst.

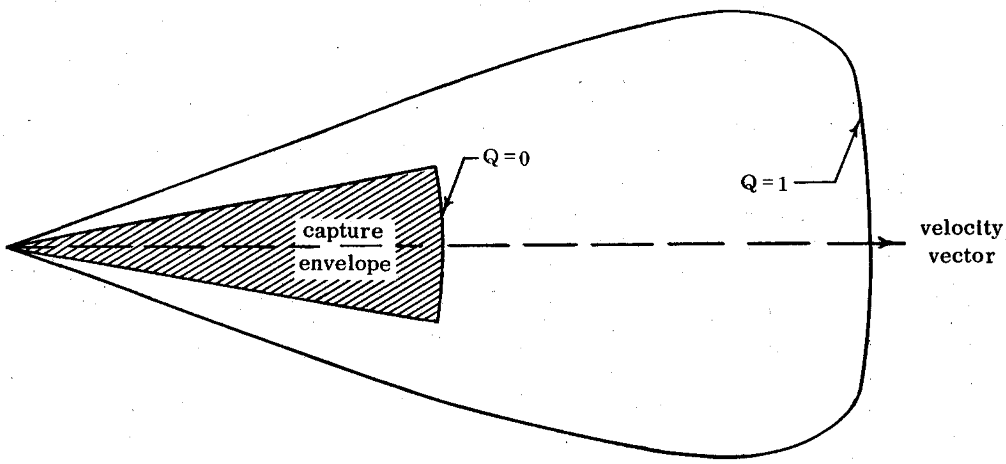

Fig. 1 Generalized miss.

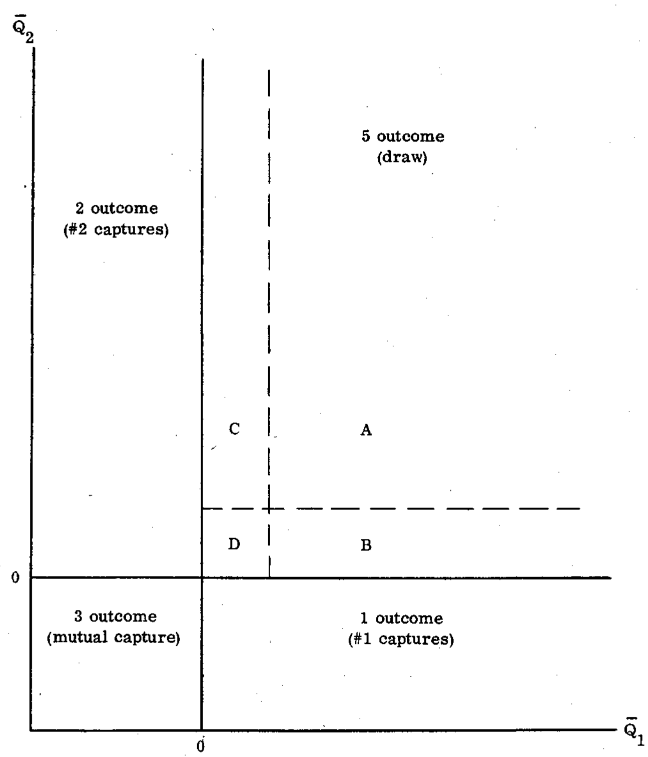

Fig. 2 Capture and draw regions in $\bar{Q}_1$, $\bar{Q}_2$ space.

A) $\bar{Q}_1 \gg 0$ and $\bar{Q}_2 \gg 0$

#1 will min $\bar{Q}_1$

#2 will min $\bar{Q}_2$

B) $\bar{Q}_1 \gg 0$ and $\bar{Q}_2 \cong 0$

#1 will max $\bar{Q}_2$

#2 will min $\bar{Q}_2$

C) $\bar{Q}_1 \cong 0$ and $\bar{Q}_2 \gg 0$

#1 will min $\bar{Q}_1$

#2 will max $\bar{Q}_1$

D) $\bar{Q}_1 \cong 0$ and $\bar{Q}_2 \cong 0$

#1 will max $\bar{Q}_2$

#2 will max $\bar{Q}_1$

This subdivision of the draw space is shown in Fig. 2, which also shows "1", "2", and "3" capture regions. In "states" B and C, zero-sum game theory is applicable in that the players agree upon an objective, each in opposition to the other.

The threat-reciprocity concept blends the control policies in the draw space smoothly between the regions in terms of the relative importance attached to $\bar{Q}_1$ and $\bar{Q}_2$ in each player's control choice. The implementation adopted in the presently reported first computational attempt, however, is much cruder than this. Essentially, it is the use of 0 and 1 weights in regions B, C, and D, as noted in the preceding listing, and the use of minimax $\bar{Q}_1 - \bar{Q}_2$ in region A. The thresholds were set somewhat arbitrarily at $\bar{Q} = 1.2$.

Scoring Illustration

For a given set of initial conditions, the trajectories are calculated twelve times, once for each combination of control parameters, and the results arranged in matrix form. Thus,

$$\begin{bmatrix} 3 & 3 & 2 & 2 \\ 3 & 3 & 2 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

The first player chooses the column, the second the row, each according to his preference order. In minimax, the first player is assumed to have chosen first; for each choice of column by #1, the choice by #2 is arrayed in a row as

$$[\ 3\ \ 3\ \ 2\ \ 2\ ]$$

The first player, anticipating this, would choose the outcome 3, the minimax score. In maximin, the second player is assumed to have chosen the row first. For each choice of row, the preference-ordered choice by the first player results in

$$\begin{bmatrix} 3 \\ 3 \\ 1 \end{bmatrix}$$

The second player, in anticipation of this, chooses the outcome 3, the maximin score. The terms minimax and maximin are employed loosely, as the game is not zero-sum unless player #1's preference order is the opposite of player #2's, and this does not arise naturally in familiar battle settings. In the interest of more nearly unique determination of guidance options, graduated preferences are adopted within each category. Thus, minimax time in 1; maximin time in 2; maximin time in 3; and reciprocity in 5 according to the scheme of the preceding section.

Cell-Buildup Technique

Preference-ordered-optimal guidance choices are stored in cellular subdivisions of the joint state space, whose variables are the separation distance r and the heading angles of the two vehicles measured from the line-of-sight. Each cell is divided into subcells according to the time remaining until attainment of the specified max time; thus, in the example of Ref. 4, there are ten subcells, each spanning 3 s of the 30-s max time. Development of cell-buildup technique, since the work of Ref. 4 was carried out, has resulted in significant gains in computational efficiency and considerable simplification. The following description applies to the improved procedure which employs a more rigid discretization than the former one: changes in guidance are implemented at 3-s intervals. A major effect of the modified technique is that trajectories generated in buildup calculations always find themselves in subcells already made active after the first 3-s interval, and none need be discarded for passing through neutral (i.e., not-yet-active) subcells. Captures are scored only at 3-s interval times.

The cell buildup starts with 3-s trajectories originating from the midpoints of target-set cells in order to determine which correspond to mutual captures. A dozen trajectories are generated from each, corresponding to various guidance combinations, as described in Ref. 4. The matrix scoring decides whether the subcell has a single-capture outcome or a mutual-capture outcome (determined by the occurrence of a second capture within 3-s). Target-set cells in a region of *overlap* of the two target sets can be scored without trajectory integrations, as the mutual-capture outcome is obvious at the outset. Such cells have the 3 outcome (mutual capture) in all ten subcells and *no* guidance choice stored, as the outcome is instantaneous and independent of guidance. For other target-set cells whose scoring results in a mutual-capture outcome within the 3-s, the subcells are activated with 3 outcomes plus whatever guidance combination emerged from the matrix scoring. Next, a dozen 3-s trajectories are run out of each cell, starting at time 27 s, and scoring done via the matrix procedure. The trajectory emerging from the matrix scoring is termed a "lead trajectory." Trajectory integrations are continued until the max duration, 30 s, if no capture has occurred by then, or until capture plus 3-s in order to determine whether there is a mutual capture. Thus, some nominally 3-s trajectories are actually 6, etc. The continuing buildup employs progressively longer trajectories: 6-s trajectories starting at 24 s, etc. When a subcell is activated with a capture outcome, the same outcome and guidance are used to generate all of the longer-time-to-go subcells of the particular cell. Active-cell data are ordinarily represented in several tabular arrays relating cell number, cell content, and the location of the cell in state space. The content of each subcell is represented by three digits, of which the first two are guidance (1-12) and the last is the outcome.

Reprisal Guidance

A "reprisal" guidance policy, i.e., one which attempts to exploit perceived departures from optimality by one's opponent by extrapolation of the error, may be synthesized via minor changes in the cell-buildup process. If one assumes that errors will persist for a time period equal to the basic buildup time increment, for example, the procedure is to erase the 3-s-to-go subcell and rebuild it as many times as the opponent has guidance options, in each instance assuming that the opponent's freedom is restricted to the particular option. This generates several top subcells, three or four in the example, to be used selectively depending upon the opponent's choice of control. If the opponent plays optimally, the guidance selected is optimal in the preference-ordered sense employed in the buildup. The risk lies entirely in the assumption that one's perception to the opponent's control is fast enough and accurate enough to trust.

Table 1 Effect of guidance options

Outcome	Both minimax	#2 hard-turning #1 reprisal (3 s)	#2 hard-turning #1 reprisal (30 s)
1 (#1 captures)	21	18	19
2 (#2 captures)	91	84	68
3 (mutual capture)	30	39	38
5 (draw)	108	109	128

There is an obvious additional margin to the informationally-advantaged player in a sequential game such as the present one, i.e., to the player who makes his choice with knowledge of his opponent's choice. In pursuit-evasion games, and in other differential games featuring a separable Hamiltonian function, this advantage is supposed to disappear in the limit as the discretization time-increment shrinks to zero (Ref. 3). It appears, however, that the effect of nonzero time-step may be substantial. There is also an important related effect, viz., that the informationally-advantaged player must play closed-loop to retain his advantage. His use of open-loop strategy against nonoptimal play may incur penalties, the effect being equivalent to successful deception by his opponent.

Extrapolation of opponent's errors for a time spanning several layers of subcells is also possible and attractive; these layers are simply rebuilt the requisite number of times, assuming that the opponent's guidance is locked into each choice in turn for the entire extrapolation time-span. In this case, the guidance generated will not generally be optimal against optimal play; one cannot have everything. However, a compromise suggests itself, viz., use preference-ordered-optimal guidance when one's opponent is playing optimally and extrapolate only in the face of nonoptimal play. Such a mix might be called a minimax-reprisal composite.

Illustrative Numerical Results

The example is identical to that of Ref. 4 except that the speed and turn-rate characteristics of the two aircraft are equal, in the present case, and the disengagement outcome, a special type of draw, does not appear as a result of this. Aircraft #1 has a narrow-angle (±10 deg) weapon effective out to 2 n. mi. Aircraft #2 has an identical weapon and, in addition, a wide-angle weapon (60 deg semiapex angle) with a reach of 1.5 n. mi. There are 250 cells each partitioned into 10 subcells of 3-s, time-remaining increments.

Table 1 summarizes results obtained for this family first under the assumption that both players make minimax control choices, the term being used here loosely to denote preference-ordered-optimal. These results indicate the sort of superiority for #2 that might be expected as a result of his weaponry advantage.

Results are shown for two cases in which #2 is locked into hard-turning guidance toward his opponent. There is a shift in favor of #1 reflected in reduction in #2 captures, dramatic for the case of long-term extrapolation to T_{max}, as much as 30 s. The reduction in #1 captures is believed attributable to coarseness-of-mesh in combination with the fact that many draws are near-captures. It should be borne in mind that 85 of the captures occur initially, in the capture set: 10 #1 captures, 52 #2 captures, and 23 #3 captures.

Concluding Remarks

The results of the present study and that of Ref. 4 suggest that departures from optimality in the details of air-combat maneuvers are not nearly as important as are mistakes in deciding whether to attack, to flee, or to maneuver for an improvement in the situation relative to one's opponent without committing one's self to an attack. A rationale for role-determination and for maneuvering in draw situations is furnished by the "threat-reciprocity" concept. The preference-ordered discrete-gaming computational approach in combination with the reprisal technique for exploiting the mistakes of one's opponent by extrapolation seems promising for applications work.

The short-term emphasis in future work should be on streamlining the computations and on systematic exploration of mesh-size effects and the effect of informational advantage. One direct development of interest is the blending of the two-dimensional cell-structure results with appropriate alt logic into a three-dimensional point-mass digital simulation.

Another quite exciting possibility for application is to two-on-one and, ultimately, to many-on-many. The advances presently reported fall precisely in the weakest area of existing air-combat-analysis comptuer programs (most of them simulations), viz., control logic. Serious enough for one-on-one, this weakness becomes overwhelming for the many-on-many case. The attractive approach is the use of one-on-one results for instantaneous evaluation of the threat posed by each vehicle in the fray to each opponent, and the rational assignment of roles and individual opponents with particular regard to weaknesses of and apparent tactical blunders by the opposition.

Acknowledgments

The research reported was performed under USAF Contract F33605-77-C-3144 with Analytical Mechanics Associates, Inc., Jericho, N.Y., almost all of it while H.J. Kelley and L. Lefton were with A.M.A. The Technical Monitors of USAF's Flight Dynamic Laboratory, Wright-Patterson AFB, Ohio, were L. Earl Miller and David Johnson, to whom the writers are grateful for helpful discussions and comments.

References

[1] Isaacs, R., *Differential Games,* Wiley, New York, 1965.

[2] Isaacs, R., "The Past and Some Bits of the Future," in *The Theory and Application of Differential Games,* edited by J.D. Grote, Reidel, Dordrecht, 1975.

[3] Friedman, A., *Differential Games,* Wiley-Interscience, New York, 1971.

[4] Kelley, H.J. and Lefton, L., "A Preference-Ordered Discrete-Gaming Approach to Air-Combat Analysis," *IEEE Transactions on Automatic Control,* Aug. 1978.

[5] Kelley, H.J., "A Threat-Reciprocity Concept for Pursuit-Evasion," in *Differential Games and Control Theory II,* edited by E.O. Roxin, P.T. Liu, and R.L. Sternberg, Marcel Dekker, Inc., New York, 1977.

[6] Roberts, D.A. and Montgomery, R.C., "The Use of Gradient Methods in Solving a Differential Game of Relevance in Aerial Combat," *Preprint Volume, 1971 Joint Automatic Control Conference,* Washington University, St. Louis, Mo., Aug. 11-13, 1971.