MULTI-PLAYER PURSUIT-EVASION DIFFERENTIAL GAMES

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Dongxu Li, B.E., M.S.

The Ohio State University

2006

Dissertation Committee:

Jose B. Cruz, Jr., Adviser
Vadim Utkin
Philip Schniter

Approved by

Adviser
Graduate Program in Electrical Engineering
© Copyright by
Dongxu Li
2006
ABSTRACT

The increasing use of autonomous assets in modern military operations has led to renewed interest in (multi-player) Pursuit-Evasion (PE) differential games. However, the current differential game theory in the literature is inadequate for dealing with this newly emerging situation. The purpose of this dissertation is to study general PE differential games with multiple pursuers and multiple evaders in continuous time.

The current differential game theory is not applicable mainly because the terminal states of a multi-player PE game are difficult to specify. To circumvent this difficulty, we solve a deterministic problem by an indirect approach starting with a suboptimal solution based on “structured” controls of the pursuers. If the structure is set-time-consistent, the resulting suboptimal solution can be improved by the optimization based on limited look-ahead. When the performance enhancement is applied iteratively, an optimal solution can be approached in the limit. We provide a hierarchical method that can determine a valid initial point for this iterative process.

The method is also extended to the stochastic game case. For a problem where uncertainties only appear in the players’ dynamics and the states are perfectly measured, the iterative method is largely valid. For a more general problem where the players’s measurement is not perfect, only a special case is studied and a suboptimal approach based on one-step look-ahead is discussed.
In addition to the numerical justification of the iterative method, the theoretical soundness of the method is addressed for deterministic PE games under the framework of viscosity solution theory for Hamilton-Jacobi equations. Conditions are derived for the existence of solutions of a multi-player game. Some issues on capturability are also discussed for the stochastic game case.

The fundamental idea behind the iterative approach is attractive for complicated problems. When a direct solution is difficult, an alternative approach is usually to search for an approximate solution and the possibility of serial improvements based on it. The improvement can be systematic or random. It is expected that an optimal solution can be approached in the long term.
To my family.
I would like to express my deepest gratitude to my advisor, Professor Jose B. Cruz, Jr., for his guidance, encouragement and financial support throughout my Ph.D study. He has taught me how to find an interesting problem, given me a great degree of freedom to explore problems independently and provided valuable feedbacks. He always made himself available for me in his quite tight schedule. He has been my role model for scholarship and professionalism. I am also thankful to Professor Vadim Utkin and Philip Schniter for serving on my dissertation committee.

I am very grateful to Dr. Genshe Chen, who is not only a colleague, but also a friend. He has consistently shown his concern on my research as well as my academic growth over the years. He also provided me a precious opportunity of internship, which helped me find the topic of this dissertation, extended my school experience and made me aware of my long-term goal. I also want to thank Dr. Corey Schumacher for his insightful suggestions that inspire part of the fourth chapter of this dissertation.

Many thanks to those graduate students and postdoctoral scholars who have helped me throughout my Ph.D study to make this unique experience gratifying rather than stressful. I am especially thankful to our team Mo Wei, Dan Shen, Xiaohuan Tan, Xu Wang and Ziyi Sun. Thanks for sharing every step of bitterness and joy in our growth.
My special thanks goes to my girlfriend Ye Ye for her kindness, understanding and
tolerance for my absence at all times. Finally, I want to thank my parents Zhenming
Li and Zhonglan Liu, and my brother Chuntao Li for their unconditional love and
support all through my life.
VITA

December 15, 1977 .......................... Born - Nanjing, China

July 2000 ................................. B.E. in Engineering Physics
Department of Engineering Physics
Tsinghua University
Beijing, China

August 2002 ............................... M.S. in Nuclear Engineering
Department of Mechanical Engineering
The Ohio State University
Columbus, Ohio

September 2002 - Present .............. Graduate Research Associate
Graduate Teaching Associate
Dept. of Elec. & Computer Eng.
The Ohio State University
Columbus, Ohio

PUBLICATIONS

Research Publications


FIELDS OF STUDY

Major Field: Electrical and Computer Engineering
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapters:</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Background</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Motivation</td>
<td>10</td>
</tr>
<tr>
<td>1.3 Dissertation Outline</td>
<td>13</td>
</tr>
<tr>
<td>2. Deterministic Multi-Player Pursuit-Evasion Differential Games</td>
<td>17</td>
</tr>
<tr>
<td>2.1 Deterministic Game Formulation</td>
<td>18</td>
</tr>
<tr>
<td>2.2 Dilemma of the Current Differential Game Theory</td>
<td>24</td>
</tr>
<tr>
<td>2.3 A Structured Approach of Set-Time-Consistency</td>
<td>25</td>
</tr>
<tr>
<td>2.4 Iterative Improvement Based on Limited Look-ahead</td>
<td>29</td>
</tr>
<tr>
<td>2.5 Finite Convergence on a Compact Set</td>
<td>35</td>
</tr>
<tr>
<td>2.6 The Hierarchy Approach</td>
<td>38</td>
</tr>
<tr>
<td>2.6.1 Two-player Pursuit-Evasion Games at the Lower Level</td>
<td>39</td>
</tr>
<tr>
<td>2.6.2 Combinatorial Optimization at the Upper Level</td>
<td>39</td>
</tr>
<tr>
<td>2.6.3 Uniform Continuity of the Suboptimal Upper Value</td>
<td>42</td>
</tr>
</tbody>
</table>

Abstract .................................................................................. ii
Dedication ................................................................................. iv
Acknowledgments ......................................................................... v
Vita ......................................................................................... vii
List of Tables ........................................................................... xii
List of Figures .......................................................................... xiii
2.7 Theoretical Foundation of the Iterative Method ............... 44
   2.7.1 Viscosity Solution of Hamilton-Jacobi Equations ........... 44
   2.7.2 Existence of the Value Function .......................... 47
2.8 Summary and Discussion ........................................ 54

3. Simulation Results of Deterministic Pursuit-Evasion Games .......... 58
   3.1 Feasibility of the Iterative Method ............................ 59
      3.1.1 Analytical Solution of the Two-player Game Example ....... 59
      3.1.2 Numerical Solution by the Iteration Method ............ 61
   3.2 Suboptimal Solution by the Hierarchical Approach ............. 62
   3.3 Suboptimal Solution by Limited Look-ahead ..................... 67
      3.3.1 Performance Enhancement by Limited Look-ahead .......... 67
      3.3.2 Limited Look-ahead with a Heuristic Cost-to-go ........ 70

4. Stochastic Multi-Player Pursuit-Evasion Differential Games .......... 76
   4.1 Formulation of Stochastic Pursuit-Evasion Games with Perfect State
       Information ...................................................... 76
   4.2 The Iterative Approach to Stochastic Pursuit-Evasion Games ...... 81
   4.3 Solution to a Two-Player Pursuit-Evasion Game ................ 86
      4.3.1 Classical Theory of Stochastic Zero-sum Differential Games 86
      4.3.2 Solution to a Two-Player Pursuit-Evasion Game .......... 91
      4.3.3 On Finite Expectation of the Capture Time ............ 95
   4.4 On Existence of Solutions to Stochastic Pursuit-Evasion Games with
       Perfect State Information ....................................... 104
   4.5 Stochastic Pursuit-Evasion Games with Imperfect State Information
       4.5.1 Limited Look-ahead for Stochastic Control Problems ....... 106
       4.5.2 On Stochastic Pursuit-Evasion Games ..................... 110
       4.5.3 On Finite Expectation of the Capture Time ............ 112
   4.6 Simulation Results .............................................. 116

5. Dissertation Summary and Future Work ............................ 123
   5.1 Dissertation Summary ........................................... 123
   5.2 Future Work ..................................................... 126
      5.2.1 The Iterative Method for Nonzero-sum Games ............. 127
      5.2.2 On Efficient Algorithms .................................. 128
      5.2.3 On Decentralized Approaches .............................. 131

Appendices:
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Proof of Theorem 5</td>
<td>134</td>
</tr>
<tr>
<td>B. Proof of Theorem 6</td>
<td>140</td>
</tr>
<tr>
<td>Bibliography</td>
<td>145</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>The Necessary Parameters of the Players for PE Game Scenario 1</td>
<td>64</td>
</tr>
<tr>
<td>3.2</td>
<td>The Optimal Engagement Between the Pursuers and the Evaders with the First-Order Dubin’s Car Model</td>
<td>65</td>
</tr>
<tr>
<td>3.3</td>
<td>The Optimal Engagement Between the Pursuers and the Evaders with the Modified Dubin’s Car Model</td>
<td>66</td>
</tr>
<tr>
<td>3.4</td>
<td>The Necessary Parameters of the Players in the PE Game Scenario 2</td>
<td>69</td>
</tr>
<tr>
<td>3.5</td>
<td>The Necessary Parameters of the Players in the PE Game Scenario 3</td>
<td>72</td>
</tr>
<tr>
<td>4.1</td>
<td>The Necessary Parameters of the Pursuers and the Evaders in the Selected Stochastic PE Game Scenario</td>
<td>117</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Illustration of a Suboptimal Approach with Hierarchy</td>
<td>41</td>
</tr>
<tr>
<td>3.1</td>
<td>The Speed Vector under the Optimal Strategies of the Players</td>
<td>60</td>
</tr>
<tr>
<td>3.2</td>
<td>The Evolution of the Suboptimal Value $W_k$'s at (2,1) by Iteration</td>
<td>63</td>
</tr>
<tr>
<td>3.3</td>
<td>Pursuit Trajectories under the Best Engagement</td>
<td>68</td>
</tr>
<tr>
<td>3.4</td>
<td>Pursuit Trajectories under the Strategies by the Hierarchical Approach</td>
<td>69</td>
</tr>
<tr>
<td>3.5</td>
<td>Pursuit Trajectories under the Strategies Based on One-Step Limited Look-ahead</td>
<td>71</td>
</tr>
<tr>
<td>3.6</td>
<td>A Necessary Condition on the Capturability of the Superior Evader</td>
<td>73</td>
</tr>
<tr>
<td>3.7</td>
<td>Cooperative Pursuit Trajectories of the Superior Evader by the Limited Look-ahead Method Based on a Heuristic Cost-to-go</td>
<td>75</td>
</tr>
<tr>
<td>4.1</td>
<td>The Simplified Stochastic PE Game in a One Dimensional Space</td>
<td>95</td>
</tr>
<tr>
<td>4.2</td>
<td>The Probability Density of the Capture Time $p_T(t)$</td>
<td>97</td>
</tr>
<tr>
<td>4.3</td>
<td>Change of the $\tilde{x}$-$\tilde{y}$ Coordinates to the $x'$-$y'$ Coordinates</td>
<td>101</td>
</tr>
<tr>
<td>4.4</td>
<td>Illustration of the Pursuer's Control Based on a Noisy Measurement</td>
<td>114</td>
</tr>
<tr>
<td>4.5</td>
<td>Cooperative Pursuit Trajectories of All 4 Evaders by the 3 Pursuers</td>
<td>119</td>
</tr>
<tr>
<td>4.6</td>
<td>Cooperative Pursuit Trajectories of All 4 Evaders by Pursuer 1 and 2</td>
<td>120</td>
</tr>
</tbody>
</table>
4.7 Pursuit Trajectories of All 4 Evaders by Pursuer 1 . . . . . . . . . . . 121

4.8 Cooperative Pursuit Trajectories of the 4 Evaders by the 3 Pursuers
with the Imperfect Observation of the Evaders’ Strategies . . . . . . . 122

5.1 Improving Strategies of Pursuer $i$ in a “Cooperation Region” . . . . . 129
CHAPTER 1

INTRODUCTION

1.1 Background

Optimization is the kernel of decision-making in physical or organizational systems, where a choice is made among a set of possible alternatives with the minimum cost. There are three main components in an optimization problem: the decision-maker, the objective function and the information available to the decision-maker [1]. In traditional optimization theory, all these ingredients are singular, i.e., there is only one decision-maker, one objective function and one information set. Over the centuries, both the theory and the practice of optimization have gone in great depth [2].

The setup of optimization problems with a single decision-maker, criterion and information set is often not adequate for modelling complicated problems involving multiple decision-makers. In the applications of social science and economics, there are usually more than one decision-maker and each decision-maker has its own information set, control input to the system and personal preference of the outcome, which depends on the input from each decision-maker. In this type of problem, it is very unlikely that all decision-makers share a common objective and all available information. Therefore, optimization theory above may fail to model a problem where
(multiple) decision-makers who have conflicting personal interests. The study of decisions in such conflicting situations provided the original impetus for the research in games.

Game theory is the study of strategic interactions among rational players, where each player makes decisions according to its own preference. The history of game theory can go way back to the Babylonian Talmud\(^1\), where the so-called marriage contract problem may be regarded as anticipation of the theory of games [3]. The actual birth of the modern game theory was due to John Von Neumann, who was the first to prove the minimax theorem. After that, great progress has been made. In 1944, John Von Neumann and Oskar Morgenstern published their seminal work [4], which laid the groundwork of game theory including two-person zero-sum game and the notion of a cooperative game etc. They also gave the account of axiomatic utility theory, which has been widely accepted in decision theory. In 1950, John Nash proved the existence of a strategic equilibrium (now called as Nash equilibrium) for non-cooperative games [5]. Nowadays, it is the most widely adopted solution concept for problems in a game setting\(^2\). The last fifty years witnessed intensive research activities in game theory [3], among which is the establishment of the theory of dynamic games.

Dynamic games belong to game theory where the games evolve in (discrete or continuous) time. Their origin may be traced back to the mixed-strategy in a finite matrix game, which could be interpreted as the relative frequency of each move (pure

\(^1\)The compilation of ancient law and tradition set down during the first five centuries A.D..

\(^2\)It is worth mentioning that this is not the only solution concept for game problems. In a game where the decision-making of the players has a hierarchy, i.e., one group of the players (called leaders) can enforce the strategy of other players (called followers) who react (rationally) to the leader’s decisions, Stackelberg (hierarchical) equilibrium is used. This name is due to the first study of economic competition by Heinrich von Stackelberg [6].
strategy) by repeating the same game many times [7]. It has connections to a discrete-time dynamic game, which is played over (finite or infinite) stages. In a discrete-time dynamic game, each player has a payoff function associated with each stage (possibly different with stages) and the transition of the game from stage to stage is governed by a difference equation that depends on the inputs of all the players. If the evolution of a game is described by differential equations, the dynamic game is also referred to as “differential game” to emphasize the role of the differential equations.

Unlike the initial development of game theory which was inspired by the problems in social science and economics, the main motivation of differential games was the study of military problems such as Pursuit-Evasion (PE) games. In such a (two-player) PE game problem, the pursuer wants to capture the evader and the optimal strategies of both players are studied [11, 7]. Besides, another important motivation for zero-sum differential games is the worst-case controller design for a system with disturbance. In these design problems, the process to be controlled has two types of inputs, controlled and disturbance, and two types of outputs, regulated and measured. The goal is to suppress the effect from the disturbance on the regulated output. We refer the readers to [8] for linear systems and to Appendix B in [9] (and the references therein) for the nonlinear case.

Differential games were initiated by Rufus Isaacs in early 1950s when he studied the military PE type of problems in the Rand Corporation [10, 11]. The PE game he studied is a two-player zero-sum game, where the players have completely opposite interests. Isaacs used the method of “tenet of transition” [11] to resolve the strategic dynamic decision-making in this adversarial environment, and he analyzed many special cases of Partial Differential Equation (PDE), which is now called
Hamilton-Jacobi-Isaacs (HJI) equation. Although developed concurrently with Dynamic Programming (DP), his method is essentially a DP method based on backward analysis.

Following Isaacs, the research in differential games has been boosted. The theory of two-player zero-sum differential game has been generalized to the N-player non-cooperative (nonzero-sum) game case [7]. Nash equilibrium is adopted as the solution concept when the roles of all players are symmetric. The concept of Stackelberg solution that appeared in a hierarchical (static) optimization problem has also been extended into differential games due to the work by Chen, Simaan and Cruz [13, 14]. We refer to [7] for detailed introduction of this topic.

The development of differential game theory goes along with optimal control theory because their solution techniques are closely related. An optimal control problem can be viewed as a special case of a differential game with one player. The kernel of optimal control theory includes the minimum principle and DP. The minimum principle\(^3\) was developed by Pontryagin and his school in USSR in the 1960s [15]. It states that the optimal trajectories of the system satisfy the “Euler-Lagrange equations” and at each point along the trajectories, the optimal control must minimize the Hamiltonian. The minimum principle deals with one extremal associated with one state at a time, such that the resulting optimal control is open-loop.

The invention of DP was due to Richard Bellman and his school in the 1950s [12, 16]. Bellman is the first to apply DP to discrete-time optimal control problems, demonstrating that state rollback is a natural approach to solving optimal control

\(^3\)It is also refereed to as the maximum principle in the context of a maximization problem.
problems. More importantly, the DP procedure results in closed-loop, generally non-linear, feedback schemes, which makes it the only general method for stochastic control and differential games. DP principle (or principle of optimality) basically says that for an optimal system trajectory, its end portion starting from any point along the trajectory to the terminal must also be optimal. The optimal value of the performance index starting from any state at any time is defined as the value function. With additional assumptions on differentiability, the value function can be characterized as the unique solution of a PDE called Hamilton-Jacobi (HJ) equation. The notion of value function and HJ equations have roots in Calculus of Variations. Unlike the minimum principle, DP deals with a family of possible extremal trajectories at a time, which is associated with the value function at different states and times. Therefore, the control obtained is in the form of feedback, which makes it more useful in problems with uncertainty and games. We refer to [19] for connections of the minimum principle and DP, and to [20] for an overview of this subject.

Uncertainty is inherent in real-world problems, and it induces many disadvantages to optimal control problems (resp. differential games). If the information available to a controller (resp. player) is uncertain or if the evolution of the system process is governed by random processes, the performance criterion is no longer certain. Thus, the expected value may be the best to optimize, provided that the statistics of the uncertain factors is known. This leads to the problem of stochastic optimal control (resp. stochastic differential games).

4In the optimal control setting, it is called Hamilton-Jacobi-Bellman (HJB) equation; it is referred to as HJI equation in differential games.

5It can be dated back to the great mathematicians such as Bernoulli, Leibniz and Newton in connection with the brachistochrone problem [18].
The pursuit of stochastic optimal control as well as stochastic differential games can be traced back to the initial study of random process by Einstein on Brownian motion. Later, Norbert Wiener with G.I. Taylor made tremendous contributions [21], among which is Wiener’s theorem on optimal filtering based on minimizing the mean squared estimate error. It is during this period that stochastic techniques were introduced into control and communication theory. In 1961, Kalman and Bucy extended the Wiener filter to a time-varying linear system[22], and this state estimator is called the Kalman-Bucy filter\(^6\). At about the same time, optimal control of linear systems with a quadratic objective function was studied by Kalman [23] (see also Merriam [24]), and the problem was later called the Linear Quadratic Regulator (LQR) problem. The extension of this problem to linear dynamical systems with additive white noise\(^7\) was first studied by Florentin [25, 26], Joseph and Tou [27] and Kushner [28]. This problem can be handled by DP and stochastic calculus, where the HJB equation becomes a second-order PDE. The next extension was to consider the problem with noise-corrupted observation. Gunckel and Franklin [29] and Joseph and Tou [27] showed that the optimal linear feedback gains obtained in the deterministic LQR problem with the state estimate from the Kalman filter generates the optimal control, and this procedure was later called “the separation principle”. The general partially observable problem was first treated by Davis and Varaiya in [30], where they introduced the notion of information state to reduce the problem to a fully observable problem. The problem was also studied later by Fleming and Pardoux [31].

\(^6\)The discrete-time version is called the Kalman filter.

\(^7\)White noise driven systems have been used widely in modern stochastic optimal control and filtering theories. In fact, white noise, being a mathematical idealization, gives only an approximate description of real noise.
In general, a stochastic control problem with imperfect measurements is very hard, and the issue of the existence of solutions as well as solution techniques is still largely open [32].

DP can be used to solve a differential game problem provided that the value function exists. Although Isaacs resolved many differential game problems using the DP type of method, unfortunately, he did not provide a priori condition on the existence of value function, nor did he give an adequate definition of strategy. The first general and rigorous result about the existence of value in differential games was developed by Fleming. He proved the existence of value for finite-horizon problems and for special cases of PE games [33, 34]. Its main idea is to determine the strategies of players on small time intervals based on the game history. The value is defined by passing to the limit as the lengths of the time intervals tend to zero. Approaching the problem in a different way from Fleming, Varaiya [35], Roxin [36] and Elliott and Kalton [37, 38] introduced the notion of strategies and value of differential games, established the existence of value and demonstrated that the value satisfies the HJI equation in a generalized sense. Based on some of these results, Friedman developed his theory in the early 1970s and it is summarized in the books [39] and [40]. Furthermore, the theory of positional differential games was developed by Krasovskii and Subbotin in the 1970s, where the value function of the game is shown as the unique function that is stable with respect to both players [41]. Finally, the Berkovitz’s theory of existence of value combines features of the methods by Isaacs, Friedman and Krasovskii-Subbotin [42].

At that time, it was generally believed that in order to apply DP methods, not only the value function of a differential game (control) problem must exist, but also
it should be sufficiently smooth because the formulation of HJI (HJB) equations depends on its differentiability. This additional requirement makes the applicability of DP methods extremely limited because even for a simple game (optimal control) problem, its value function may be non-differentiable or even discontinuous. Under such circumstances, it would be difficult to justify the meaning of a HJ equation and whether its solution is the value function. From the 1960’s, many researchers worked on this subject [9]. The turning point is the introduction of viscosity solution for first-order HJ equations by M.G. Crandall and P.L. Lions in 1983 [43]. Viscosity solution is a weak (or generalized) solution for HJ equations, with which the rigorous mathematical justification of DP methods becomes possible. More importantly, the authors proved the uniqueness of the solution under relatively weak conditions. This concept was further simplified by Crandall, Evans and Lions in [44]. Viscosity solution was first extended to second-order HJ equations with convexity assumption by P.L. Lions via stochastic control arguments [45]. With purely analytical techniques, R. Jensen made a major breakthrough on more general (nonconvex) equations [46], and was further simplified by R. Jensen, P.L. Lions and P.E. Souganidis [47]. The most general results on second-order HJ equations were due to H. Ishii and P.L. Lions [48]. We refer to [45] and [49] for an introduction of this subject.

The value function of a differential game (optimal control) problem is closely connected to the viscosity solution of HJ equation, and the study on this relationship started from a significant observation made by P.L. Lions in his book [45] on page 53-54,

“value functions (of optimal control) do in fact satisfy the associated Hamilton-Jacobi equations in the viscosity sense”.

8
Afterwards, a considerable amount of research has been devoted to the applications of viscosity solution theory on optimal control as well as differential games. The connection between differential games and viscosity solution was perhaps first drawn by Souganidis in his Ph.D thesis in 1983 with respect to the notion of value of a game by Fleming and Friedman. Shortly after that, Evans and Souganidis proved that the value\(^8\) of a differential game is the viscosity solution of the associated HJI equation for problems with finite horizon. The connections of the work by Krasovskii and Subbotin, and Berkovitz with viscosity solutions have also been drawn \([9]\). Similar results have been obtained for differential game problems under various conditions. Finally, stochastic (two-player) zero-sum differential game has studied by Fleming and Souganidis under the framework of viscosity solutions of second-order HJI equations \([50]\).

Practically, DP suffers from the so-called “curse of dimensionality”, which stands for the exponential growth of the computational requirements with the number of state variables and time steps. The development of approximate DP methods on large-scale problems began with the policy iteration method for Markov Decision Process (MDP)\(^9\) by Ron Howard \([51]\). It provides all essential elements of underlying the theory and numerical algorithms of modern approximate DP (or reinforcement learning in computer science) \([52]\). Over the decades, the numerical DP methods for continuous-time optimal control problem and differential games have also been extensively studied \([53]\), and Appendix A in \([9]\) gives an excellent overview of this subject.

\(^8\)Here, the concept of value is due to Varaiya, Roxin and Elliott and Kalton.

\(^9\)MDP is a stochastic sequential decision problem with an infinite horizon, where the system evolution is described by a controlled Markov chain. It provides a mathematical framework for modelling decision-making in an uncertain environment.
Besides, since a closed-loop optimal solution for a stochastic control problem with an imperfect state information pattern is formidable, suboptimal optimal control methods such as Open-loop Feedback Control and rollout algorithms have been developed based on the DP principle [54]. Connections have been recently drawn between these methods and model predictive control that has been recently widely used in industry. A unified framework of those methods is introduced in [55], where readers can find a good survey of the recent developments in this area.

1.2 Motivation

PE game was the first impetus for the research of differential game theory. In such a problem, one or a group of players called pursuers go after one or more moving players called evaders. By application of differential game theory, a number of formal solutions regarding optimal strategies in particular PE problems can be achieved [11, 7, 56, 57]. In the literature, a PE problem is usually formulated as a zero-sum game in which the pursuers try to minimize a prescribed cost functional while the evaders try to maximize the same functional. Due to the development of LQ optimal control theory, a large portion of the literature focuses on PE differential games with a performance criterion in a quadratic form and linear dynamics [56, 57, 58]. Furthermore, most studies on PE games in the current literature concentrate on two-player games with a single pursuer and a single evader, and the results for general multi-player PE games are still largely sparse.

Intuitively, the strength of the pursuit by a group of pursuers is more than a simple summation of that from each individual pursuer. Cooperation among interactive pursuers plays an important role. To date, the literature in this field is very limited.
and only special cases are considered. In [64], Foley and Schmitendorf studied a problem with two pursuers and one evader using LQ formulation. They partitioned the state space according to the cases when the evader is pursued by either one or both of the pursuers. In the region that both of the pursuers are effective, an optimal strategy of the evader is determined against the nearest pursuer. Another special example was studied by Pashkov and Terekhov in [65], where once again a problem of one evader and two pursuers with a nonconvex payoff function was considered. More multi-player PE problems of special type can be found also in [66, 67] and the references therein.

Although differential game theory has extensive and potential applications that span from traffic control, engineering design, economics to behavioral biology, its primary importance lies in military type applications such as missile guidance, aircraft control and aerial tactics etc [11, 59, 58]. In recent years, Autonomous (Aerial/Ground) Vehicles AV (AAV/AGV) have shown a great potential value in reducing the human workload and have been increasingly used in military operations. A fleet of AV’s can work more efficiently and effectively than a single AV, but requires a complex high-level tactical planning tool. AV related problems have recently received a lot of attention, and one of which is the PE problems involving multiple pursuers and multiple evaders. A typical scenario is that a group of AVs is tasked to suppress a cluster of adversarial moving targets. This problem can naturally be modelled as a PE game with multiple pursuers and multiple evaders. However, the previous theoretical results that mainly focus on problems of a single pursuer and a single evader are not
sufficient to deal with the problem of cooperative pursuit of multiple evaders by multiple pursuers in the military situations. Hence, the study of multi-player PE games is very much needed for dealing with the newly emergent situations.

Recently, various researchers have reinitiated the study of this problem [59, 60, 61, 62, 63]. In [60, 61], Hespanha et al. formulated PE games in discrete time under a probabilistic framework, in which greedy and Nash equilibrium strategies with one-step look-ahead are solved respectively to maximize the probability of finding evaders. PE strategies were also studied by Antoniades, Kim and Sastry in [62], where several heuristic solutions were attempted and compared. Furthermore, the system structure and various issues of implementation of PE strategies by a team of autonomous vehicles are discussed in [59, 63]. These approaches all deal with problems in discrete time, in which the problem of search and pursuit are intertwined. The models they used are adopted from the implementation point of view; however, those pursuit strategies obtained are not saddle-point (equilibrium) solutions for PE dynamic games. Finally, the most recent related work was done by Mitchell, Bayen and Tomlin in [68], where a numerical algorithm to compute the reachable set of a dynamic game with two players was described. This is a complementary problem of a PE game; however, the algorithm is useful for solving PE game problems. Generally speaking, in spite of all these existing results, general PE differential game problems involving multiple pursuers and multiple evaders still deserve further research effort.
1.3 Dissertation Outline

In this dissertation, we study a general multi-player\textsuperscript{10} PE differential game problem with a cumulative cost functional and general nonlinear dynamics of the players. We use an indirect approach based on iterative improvement of a suboptimal solution, and then extend to the stochastic case. Existence of a value function under saddle-point equilibrium\textsuperscript{11} is examined under the framework of viscosity solution theory of HJ equations. Issues on implementation and approximation schemes are discussed, and the simulation results are presented. This dissertation is the first attempt to treat a general multi-player PE game in continuous time. The major objective is to extend the current theory of PE differential games (mainly on two-player PE games) to a general multi-player PE differential game. Moreover, the method we propose can solve not only multi-player PE game problems but also complex dynamic optimization problems such as cooperative control problems. The detailed outline of the dissertation is as follows.

In Chapter 2, the main idea of the this dissertation is presented. We first formulate a general deterministic multi-player PE differential game, where the strategy of the players is defined together with the (upper/lower) value function of a game. We show the deficiency of the current differential game theory in solving a multi-player PE differential game problem. To circumvent this difficulty, we first propose a suboptimal approach by only accounting for a subset of the pursuers’ controls under an additional structure. If the structure satisfies the property of set-time-consistency, the

\textsuperscript{10}The term “multi-player” means multiple pursuers and multiple evaders.

\textsuperscript{11}A saddle-point equilibrium is a Nash equilibrium in a zero-sum game. The rigorous definition in the context of multi-player PE games is provided in Chapter 2.
resulting suboptimal upper value function has an improving property, i.e., it can be improved by the optimization based on limited look-ahead. If it is applied iteratively, a convergent sequence of suboptimal upper value functions can be generated and the true upper value function can be approached in the limit. We further show that if the suboptimal upper value functions are continuous, the convergence can be achieved in a finite number of steps on a compact set of the state space. Regarding the structured approach, we propose a hierarchical method that is set-time-consistent, and the corresponding suboptimal solution is a valid starting point for the iterative process. Furthermore, we examine the feasibility of this iterative method under the framework of viscosity solution theory for HJ equations. Sufficient conditions are derived for a class of PE problems, under which the value function (saddle-point) of a multi-player PE game exists.

In chapter 3, several simulation results are presented to demonstrate the usefulness of the iterative method. We first apply the iterative method on a two-player PE game for which an analytical solution is also derived. The convergence result obtained coincides with the prediction from the analytical solution. For a multi-player PE game, we focus on suboptimal numerical solutions. First, we use the hierarchical method to determine a suboptimal solution. Then, we demonstrate that the performance of this suboptimal solution can be improved by the optimization based on one-step look-ahead. Finally, an advantage of the limited look-ahead in solving a more complicated problem is presented through a particular example involving a superior evader.

In Chapter 4, we extend the iterative method for deterministic PE games to stochastic problems. We formulate a stochastic problem where the players’ (pursuers/evaders) dynamics are governed by random processes and the game states are
perfectly accessible by all the players. Accordingly, the concept of the players’ strategies and value function are introduced. We show that the iterative method for deterministic problems is largely applicable to stochastic problems. Besides, some effort is devoted to two-player stochastic PE games. We derive an analytic solution to a specific game problem, and also, the issue on capturability is discussed. Finally, we briefly discuss a more general problem where the players no longer have perfect state information. It is a very hard problem and only a special case is investigated. A suboptimal method based on one-step look-ahead is proposed, and it is compared to the well-known suboptimal method called Open-Loop-Feedback-Control [54].

We summarize the dissertation in Chapter 5, and point out possible future research directions on this topic in the end.

**Basic Notations**

\[
\begin{align*}
\mathbb{R}_{\geq 0} & \quad \text{The set of nonnegative real numbers} \\
\mathbb{R}^+ & \quad \text{The set of positive real numbers} \\
\mathbb{Z}_{\geq 0} & \quad \text{The set of nonnegative integers} \\
\mathbb{Z}^+ & \quad \text{The set of positive integers} \\
\wedge & \quad r \wedge t \triangleq \min\{r, t\} \quad \text{for } r, t \in \mathbb{R} \\
\vee & \quad r \vee t \triangleq \max\{r, t\} \quad \text{for } r, t \in \mathbb{R} \\
T & \quad \text{Capture time of some evader} \\
\Gamma & \quad \text{The set of nonanticipative strategies of pursuers} \\
\Delta & \quad \text{The set of nonanticipative strategies of evaders} \\
\Gamma_j^i & \quad \text{The set of nonanticipative strategies of pursuer } i \text{ to evader } j \\
\Delta_j^i & \quad \text{The set of nonanticipative strategies of evader } j \text{ to pursuer } i \\
E & \quad \text{Engagement scheme between pursuers and evaders} \\
E_i & \quad \text{The set of evaders that are engaged with pursuer } i \\
|S| & \quad \text{The cardinal number of a finite set } S \\
|v| & \quad |v| \triangleq v^Tv \quad \text{for } v \in \mathbb{R}^n \\
\text{int}(S) & \quad \text{The interior of a set } S
\end{align*}
\]
\( \bar{S} \)  The closure of a set \( S \)
\( \partial S \)  The boundary of a set \( S \)
\( B(x, \delta) \)  Open ball of radius \( \delta \) centered at \( x \)
\( C^1(S) \)  Set of functions on \( S \) with a continuous first-order derivative
\( C^2(S) \)  Set of functions on \( S \) with a continuous second-order derivative
\( x_{\tau; x_t, a, b} \)  State \( x \) at \( \tau \) given the initial \( x_t \) under control \( a(\cdot) \) and \( b(\cdot) \) (\( \tau \geq t \))
\( z_{\tau; z_t, x} \)  State \( z \) at \( \tau \) given the initial \( z_t \) associated with \( x(\cdot) \) (\( \tau \geq t \))
CHAPTER 2

DETERMINISTIC MULTI-PLAYER PURSUIT-EVASION DIFFERENTIAL GAMES

In this chapter, we first formulate a deterministic multi-player PE differential game, and introduce the concepts of the players’ strategy and (upper/lower) value function. Then, the dilemma of the current differential game theory in solving multi-player games is discussed. To circumvent this difficulty, we adopt an indirect approach. We first solve the problem for a suboptimal upper value within a subset of the pursuers’ controls under some structure. If the control structure satisfies set-time-consistency, the suboptimal upper value obtained can be improved by the optimization based on limited look-ahead. When the improvement is applied iteratively, a convergent sequence of suboptimal upper value functions can be generated and the true upper value is approached in the limit. If the suboptimal upper value functions are continuous, this convergence is achieved in a finite number of steps on any compact subset of the state space. Regarding the structured approach, we propose a hierarchical decomposition method that is set-time-consistent, such that it can provide a valid starting point for the iterative process. At the end of this chapter, we study the theoretical soundness of the method under the framework of viscosity solution theory. Sufficient conditions are derived, under which the value function (saddle-point) exists.
2.1 Deterministic Game Formulation

Consider a general PE differential game with $N$ pursuers and $M$ evaders in a $n_0$ dimensional space $S$, $S \subseteq \mathbb{R}^{n_0}$. Denote by $x^i_p$ ($x^j_e$) the state variable associated with pursuer $i$, $i = 1, \cdots, N$ (evader $j$, $j = 1, \cdots, M$), where $x^i_p \in \mathbb{R}^{n^i_p}$ ($x^j_e \in \mathbb{R}^{n^j_e}$). Here, $n^i_p, n^i_e, n_0 \in \mathbb{Z}^+$ and $n^i_p, n^j_e \geq n_0$ depending on the dynamics of the players. Assume that the first $n_0$ elements in state $x^i_p$ ($x^j_e$) denote the physical position of pursuer $i$ (evader $j$) in $S$. The dynamics for pursuer $i$ and evader $j$ are

\[
\begin{align*}
\dot{x}^i_p(t) &= f^i_p(x^i_p(t), a_i(t)) \quad \text{with} \quad x^i_p(t_0) = x^i_p, \quad (2.1a) \\
\dot{x}^j_e(t) &= f^j_e(x^j_e(t), b_j(t)) \quad \text{with} \quad x^j_e(t_0) = x^j_e. \quad (2.1b)
\end{align*}
\]

In (2.1), $a_i$ and $b_j$ are control inputs of pursuer $i$ and evader $j$ respectively, which satisfy

\[
\begin{align*}
a_i(\cdot) &\in A^i(t) \triangleq \left\{ \phi : [t, T] \mapsto A^i_a | \phi(\cdot) \text{ is measurable} \right\}, \quad (2.2) \\
b_j(\cdot) &\in B^j(t) \triangleq \left\{ \varphi : [t, T] \mapsto B^j_a | \varphi(\cdot) \text{ is measurable} \right\}. \quad (2.3)
\end{align*}
\]

Here, $t \geq 0$; $A^i_a \in \mathbb{R}^{m^i_p}$ and $B^j_a \in \mathbb{R}^{m^j_e}$ are compact sets for $m^i_p, m^j_e \in \mathbb{Z}^+$; $T > 0$ may be the terminal time of the game or some intermediate time of the form $t + \Delta t$ that will be introduced shortly. We assume that functions $f^i_p : \mathbb{R}^{n^i_p} \times A^i_a \longrightarrow \mathbb{R}^{n^i_p}$ and $f^j_e : \mathbb{R}^{n^j_e} \times B^j_a \longrightarrow \mathbb{R}^{n^j_e}$ satisfy the following boundedness and Lipschitz conditions.

\[
\begin{align*}
\left\{ \begin{array}{l}
\| f^i_p(x^i_p, a_i) \| \leq C_i, \\
\| f^j_e(x^j_e, a_i) - f^j_e(x^j_e, b_j) \| \leq C_j \| x^j_e - x^j_e \|;
\end{array} \right. \quad (2.4a) \\
\left\{ \begin{array}{l}
\| f^j_e(x^j_e, b_j) \| \leq C_j, \\
\| f^j_e(x^j_e, b_j) - f^j_e(x^j_e, b_j) \| \leq C_j \| x^j_e - x^j_e \|.
\end{array} \right. \quad (2.4b)
\end{align*}
\]

for some $C_i, C_j > 0$ and all $x^i_p, x^j_e \in \mathbb{R}^{n^i_p}$, $x^j_e, x^j_e \in \mathbb{R}^{n^j_e}$, $a_i \in A^i_a$, $b_j \in B^j_a$. Here, $\| \cdot \|$ is the norm in the corresponding Euclidean space, and we adopt 2-norm in this
dissertation. For simplicity, let
\[
x_p = \left[ x_1^T, x_2^T, \ldots, x_N^T \right]^T, \quad x_e = \left[ x_1^eT, x_2^eT, \ldots, x_M^eT \right]^T,
\]
\[
a = \left[ a_1^T, a_2^T, \ldots, a_N^T \right]^T, \quad b = \left[ b_1^T, b_2^T, \ldots, b_M^T \right]^T,
\]
and similarly let \( f_p = \left[ f_1^pT, f_2^pT, \ldots, f_N^pT \right]^T \) and \( f_e' = \left[ f_1^eT, f_2^eT, \ldots, f_M^eT \right]^T \). Then, according to equation (2.1), the dynamic equation can be rewritten as
\[
\dot{x}_p(t) = f_p(x_p(t), a(t)) \quad \text{with} \quad x_p(t_0) = x_{p0}, \tag{2.5a}
\]
\[
\dot{x}_e(t) = f'_e(x_e(t), b(t)) \quad \text{with} \quad x_e(t_0) = x_{e0}. \tag{2.5b}
\]
Further, let \( x = \left[ x_p^T, x_e^T \right]^T \), and denote by \( X \) the state space, i.e.,
\[
X = \left( \prod_{i=1}^N \mathbb{R}^{n_i^p} \right) \times \left( \prod_{j=1}^M \mathbb{R}^{n_j^e} \right).
\]
In addition, denote by \( A_a \) the set of admissible controls of all the pursuers, i.e., \( A_a = \prod_{i=1}^N A_i^a \) and similarly for the evaders, \( B_a = \prod_{j=1}^M B_j^a \). Based on (2.5), the sets of the admissible control inputs of the pursuers and the evaders, \( A(t) \) and \( B(t) \) can be defined accordingly, which are composed of \( A^i(t) \) and \( B^j(t) \) respectively.

For pursuer \( i \), define the projection operator \( P : \mathbb{R}^{n_i^p} \to \mathbb{R}^{n_0} \) as
\[
P(x_p^i) = \begin{bmatrix} x_{p1}^i, x_{p2}^i, \ldots, x_{p_{n_0}}^i \end{bmatrix}^T. \tag{2.6}
\]
Clearly, \( P(x_p^i) \) stands for the physical position of pursuer \( i \). A similar operator is defined for every pursuer and every evader, and we use the same notation \( P(\cdot) \) for each of them. Evader \( j \) is considered captured if there exists pursuer \( i \), such that
\[
\| P(x_p^i(t)) - P(x_e^j(t)) \| \leq \varepsilon \quad \text{for} \quad t \geq 0.
\]
Here, \( \varepsilon \) is a predefined small positive number. The capture time of evader \( j \) is defined as
\[
T_j = \min \left\{ t \geq 0 \left| \exists \ i, 1 \leq i \leq N, \text{such that} \quad \| P(x_p^i(t)) - P(x_e^j(t)) \| \leq \varepsilon \right\}. \tag{2.7}
\]
In a PE game with multiple evaders, the evaders are generally not captured simultaneously. At the terminal of the game, all the pursuers are captured. The terminal time of the PE game, $T$, can be defined as

$$T = \max_j \{T_j\}. \quad (2.8)$$

If a PE game ends at a finite time and $N < M$, there must be some pursuer $i$ that captures more than one evader. To account for this situation, we assume that when pursuer $i$ captures an evader, it is free to go after other evaders. Thus, the number of “alive” evaders (evaders that have not been captured) in a PE game is crucial to the pursuit strategy. We augment the state $x$ with a discrete variable $z \in \mathbb{Z}^M = \mathbb{Z} \cdot \mathbb{Z} \cdots \mathbb{Z}$ with $Z = \{0, 1\}$. Here, $z_j$, the $j^{th}$ element of vector $z$, is a binary variable and $z_j = 1$ if evader $j$ is not captured while $z_j = 0$ if it is captured.

According to the definition of the capture, variable $z_j$ is governed by $g_j : Z \times X \to Z$ according to the algebraic equation as

$$g_j(0, x) = 0; \quad g_j(1, x) = \begin{cases} 0 & \left\|P(x_{p}^i(t)) - P(x_{e}^i(t))\right\| \leq \varepsilon \text{ for some } i, \ 0 \leq i \leq N, \\ 1 & \text{otherwise}. \end{cases}$$

Let $g \triangleq [g_1, \cdots, g_M]^T$, and the algebraic equation regarding $z$ can be put into a compact form as

$$z(t) = z(t^+) = g(z(t^-), x(t)). \quad (2.9)$$

Here, $z(t^+)$ and $z(t^-)$ denote the right and the left limit at time $t$. Define $|z| = z^T z$, indicating the number of “alive” evaders. Given any $z \in Z^M$, define the set

$$Z_z = \{ \tilde{z} \in Z^M | \tilde{z}_j = 0 \text{ if } z_j = 0 \}, \quad (2.10a)$$

$$I_z = \{ j \in \{1, \cdots, M\} | z_j \neq 0 \}. \quad (2.10b)$$
Clearly, the terminal time of a game, $T$, can also be described as $T = \min \{ t | z(t) = 0 \}$.

**Assumption 1** Every evader stops moving when it is captured.

With Assumption 1, the dynamics of evader $j$ in (2.1) becomes

$$
\dot{x}_e^j(t) = z_j(t) \cdot f_e^j(x_e^j(t), b_j(t)) \quad \text{with} \quad x_e^j(t_0) = x_{e0}^j.
$$

(2.11)

According to (2.11), control $b_j$ is meaningful only when $z_j(t) = 1$. Now, replace the dynamics $f'_e$ in (2.5) by $f = \left[ z_1 \cdot f_1^{T}, \cdots, z_M \cdot f_M^{MT} \right]^T$, and the dynamics of the players (pursuers/evaders) can be rewritten as

$$
\dot{x}(t) = f(x(t), a(t), b(t)) \quad \text{with} \quad x(t_0) = x_0,
$$

(2.12)

where $x = [x_p^T, x_e^T]^T$ and $f(x, a, b) = \left[ f_p^T(x, a), f_e^T(x, b) \right]^T$.

Now, We choose the objective functional as

$$
J(a, b; x_0, z_0) = \int_{t_0}^{T} G(x(t), z(t), a(t), b(t)) \, dt + Q(x(T))
$$

(2.13)

subject to (2.12) and (2.9).

In (2.13), the terminal cost $Q : X \rightarrow \mathbb{R}_{\geq 0}$ is bounded and Lipschitz, i.e.,

$$
|Q(x)| \leq C_Q \quad \text{and} \quad |Q(x) - Q(\bar{x})| \leq C_Q \| x - \bar{x} \|
$$

(2.14)

for some $C_Q > 0$ and all $x, \bar{x} \in X$; the cost rate $G : X \times Z^M \times A_a \times B_a \rightarrow \mathbb{R}_{\geq 0}$ satisfies

$$
\delta \leq G(x, z, a, b) \leq C_G, \quad \text{and} \quad |G(x, z, a, b) - G(\bar{x}, z, a, b)| \leq C_G \| x - \bar{x} \|
$$

(2.15)

for some $0 < \delta < C_G$ and all $x, \bar{x} \in X, z \in Z^M, a \in A_a, b \in B_a$. Since $\delta > 0$ in equation (2.15), the optimization with respect to $J$ is meaningful only when $T$ is finite.
We model a multi-player PE game as a zero-sum game, where the pursuers try to minimize the objective $J$, while the evaders are maximizing the same objective. With this formulation, the problem is aimed to solve for optimal “centralized” strategies of players (opposite to the decentralized scheme with respect to multiple agents). It is assumed that communications among the pursuers (evaders) are ideal and available information is shared by all the pursuers (evaders). Clearly, in contrast to a two-player PE game, both cooperation and competition are relevant in a multi-player game.

Before defining the value function of a multi-player PE differential game, we need to specify the information pattern of the players. We first define nonanticipative strategies as follow. Consider a strategy of the pursuers at time $t \geq 0$ as a map $\alpha : B(t) \mapsto A(t)$.

**Definition 1** The strategy $\alpha$ is called *nonanticipative strategy* if for any $t \leq s \leq T$ and any $b, \tilde{b} \in B(t)$, $b(\tau) = \tilde{b}(\tau)$ for $t \leq \tau \leq s$ almost everywhere implies $\alpha[b](\tau) = \alpha[\tilde{b}](\tau)$ for $t \leq \tau \leq s$ almost everywhere.

Denote by $\Gamma(t)$ the set of all nonanticipative strategies for the pursuers beginning at time $t$. Similarly, a nonanticipative strategy $\beta : A(t) \mapsto B(t)$ can be defined for the evaders and the set of all $\beta$’s is denoted by $\Delta(t)$. According to this definition, it can be seen that the players in a game choose their control inputs at each time instant only based on the current and past information. In the following, we use the notations $\alpha[b]_i$ and $\beta[a]_j$ to stand for the strategies of pursuer $i$ and evader $j$ respectively.
Remark 1 Note that by inspection of (2.11), at any \( x \in X, z \in Z^M \), the strategy of evader \( j \), \( \beta[a]_j \), is meaningful only if \( z_j \neq 0 \). Thus, different sets of “alive” evaders lead to different optimization problems depending on evader \( j \) for \( j \in I_z \), which is the case in (2.16)-(2.17) as follows.

Finally, we can define the lower and upper value of a multi-player PE game. We use \( x_t \) as a short notation for \( x(t) \), and similarly for \( z, a \) and \( b \). For any \( x_t \in X \) and \( z_t \in Z^M \), the lower Value of a game \( V(x_t, z_t) \) is defined as

\[
V(x_t, z_t) = \inf_{\alpha \in \Gamma(t)} \sup_{b \in B(t)} \left\{ J(\alpha[b], b; x_t, z_t) \right\}
= \inf_{\alpha \in \Gamma(t)} \sup_{b \in B(t)} \left\{ \int_t^T G(x_\tau, z_\tau, \alpha[b]_\tau, b_\tau) \, d\tau + Q(x_T) \right\}.
\]  

(2.16)

Similarly, the upper value \( \bar{V}(x_t, z_t) \) is

\[
\bar{V}(x_t, z_t) = \sup_{\beta \in \Delta(t)} \inf_{a \in A(t)} \left\{ J(a, \beta[a]; x_t, z_t) \right\}
= \sup_{\beta \in \Delta(t)} \inf_{a \in A(t)} \left\{ \int_t^T G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) \, d\tau + Q(x_T) \right\}.
\]  

(2.17)

Note that since the dynamics of the players are time invariant and the terminal time \( T \) is not fixed, \( \bar{V} \) and \( V \) are not a function of time \( t \). Thus, in the following discussion, unless otherwise explicitly stated, we use (general) time \( t \) to denote the initial time of a game. In (2.16), the pursuers have the informational advantage, i.e., at each \( t \geq 0 \) they know the decisions made by the evaders at \( t \), and similarly for the evaders in (2.17) [9]. In general, \( V(x_t, z_t) \leq \bar{V}(x_t, z_t) \). Furthermore, if

\[
V(x_t, z_t) = \bar{V}(x_t, z_t)
\]  

(2.18)

for any \( x_t \in X \) and \( z_t \in Z^M \), we say the value of a game exists, which is denoted by \( V(x_t, z_t) \). Condition (2.18) is called the Isaacs condition, under which \( V(x_t, z_t) \) is
the saddle-point equilibrium of the PE game. With these definitions, optimality can be interpreted according to $V$, $\nabla V$ or $V$. Henceforth, we use the capitalized “Value” to stand for the (lower/upper) Value functions of a multi-player PE game defined in (2.16)-(2.18). The definition of the Value above is due to Varaiya [35], Roxin [36] and Elliott and Kalton [37, 38].

2.2 Dilemma of the Current Differential Game Theory

The conventional methods for solving differential games are closely related to optimal control theory, which includes DP and the minimum principle. In DP methods, the Value of a differential game (optimal control) problem is characterized by a HJI (HJB) equation, in which initial conditions on the terminal states are generally needed. Note that Isaacs’ method of “tenet of transition” [11] is essentially the same as DP, which is based on the underlying idea of state rollback. Starting from the terminal, an optimal state trajectory is traced backwards under the optimal control strategies. However, in a multi-player PE game, the techniques based on state rollback encounter tremendous difficulty both in the problem formulation and in specifying terminal states. First, the treatment by the corresponding HJI equation for a multi-player game problem is not obvious because the additional state $z$ is discrete. Second, in contrast to a two-player PE game, the terminal state of a multi-player game is difficult to specify. In a multi-player PE game, if pursuer $i$ catches evader $j$, we say that both players are engaged. To conduct the analysis based on state rollback, we need to start with a specific engagement scheme between the pursuers and the evaders. However, the number of possible engagements between the pursuers and the evaders increases drastically with $N$ and $M$. This explosion makes the optimal
terminal state difficult to determine in the dynamic optimization problem. Moreover, the evaders are generally not captured at the same time, which makes it more difficult to find an optimal terminal state. Therefore, DP approaches cannot be directly applied to multi-player PE games. On the other hand, a similar obstacle will be encountered when the minimum principle is applied, because boundary conditions are also needed. Generally speaking, the current differential game theory is inadequate to solving a multi-player PE differential game, such that alternative techniques are greatly needed.

2.3 A Structured Approach of Set-Time-Consistency

We first clarify the problem of multi-player PE games to be solved. We do not assume the Isaacs condition, and our study is focused on the upper Value function $\bar{V}$ defined in (2.17). In this case, the evaders are endowed with the informational advantage, which corresponds to a worst case from the pursuers’ perspective. We confine our attention to the “capture region” [11, 7], the subset of the state space $X$ where the capture of all the evaders is ensured by certain strategy of the pursuers. By abuse of notation, we use $X$ to denote the “capture region” of interest. In other words, for all $x \in X$, $z \in Z^M$, there exists a strategy $a(\cdot) \in A(t)$ of the pursuers such that all the evaders are captured by a finite time for any $\beta \in \Delta(t)$. This restriction assures us that we are solving the problem of finding optimal strategies of the pursuers (or evaders) without worrying about capturability\textsuperscript{12}.

Since the current differential game theory is not applicable, we start with a sub-optimal solution for a multi-player PE differential game. We first specify a class of

\textsuperscript{12}Here, capturability is referred to the problem of determining whether or not all the evaders can be captured by the pursuers at certain states.
suboptimal methods, such that any member of the set solves a game problem similar to the original but with respect to a restricted set of the pursuers’ controls. Later in section 2.6, we will discuss a hierarchical approach that belongs to the set. Now, suppose some structure $S$ is imposed on the pursuers’ controls. Given any states $x \in X$ and $z \in Z^M$, we denote by $A^S_{x,z}(t)$ a nonempty restricted (or structured) control set of the pursuers, and $A^S_{x,z}(t) \subseteq A(t)$. Here, the superscript and the subscript in $A^S_{x,z}$ indicate the dependence on structure $S$ and on the states $x$ and $z$.

Assumption 2 For any $x \in X$, $z \in Z^M$, at time $t \geq 0$ there exists $a \in A^S_{x,z}(t)$ such that $T < \infty$ for all $b(\cdot) \in B(t)$.

Under Assumption 2, a minimax problem similar to (2.17) is formulated as

$$\tilde{V}(x, z) = \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A^S_{x,z}(t)} \left\{ \int_t^T G(x_\tau, z_\tau, a_\tau, \beta[a_\tau])d\tau + Q(x_T) \right\} \quad (2.19)$$

Suppose that with this restriction, the difficulty of the differential game theory in multi-player games can be circumvented such that $\tilde{V}(x, z)$ in (2.19) is solvable. Clearly, $\tilde{V}(x, z)$ is an upper-bound of the real upper Value in (2.17), i.e., $V(x, z) \leq \tilde{V}(x, z)$. Next, we define the property of Set-Time-Consistency (STC) of the structured controls under $S$, which specifies a class of suboptimal strategies that preserve certain property with time. Given any $\tilde{a}^t(\cdot) \in A^S_{x,t,z,t}(t)$, $x_t \in X$, $z_t \in Z^M$ and $\beta \in \Delta(t)$, denote by $x_{s|t}, \tilde{a}^t, \beta[\tilde{a}^t]$ the trajectory of $x$ for $s \geq t$ starting from $x_t$ under $\tilde{a}^t$ and $\beta[\tilde{a}^t]$. For short, let $\tilde{x}(s) = x_{s|t}, \tilde{a}^t, \beta[\tilde{a}^t]$. Denote by $z_{s|t}, \tilde{z}$ the trajectory of $z$ corresponding to $\tilde{x}$ and use $\tilde{z}(s) = z_{s|t}, \tilde{x}$ for a short notation.
Definition 2 (Set-Time-Consistency) Control structure $S$ is said to be set-time-consistent, or ST-consistent for short, if for any $\tilde{a}^t(\cdot) \in A^S_{\tilde{x},\tilde{z}}(t)$ at any $t$ for $0 \leq t \leq T$, there exists $\tilde{a}^s(\cdot) \in A^S_{\tilde{x},\tilde{z}}(s)$ associated with $\tilde{x}_s$ and $\tilde{z}_s$ at any $s$ for $t \leq s \leq T$ such that $\tilde{a}^t(\tau) = \tilde{a}^s(\tau)$ for $s \leq \tau \leq T$ under any $\beta \in \Delta(t)$.

Remark 2 STC indicates that at any state $\tilde{x}(s)$, $\tilde{z}(s)$ along the trajectories ($t \leq s \leq T$), the later portion of the structured control $\tilde{a}^t(\cdot) \in A^S_{\tilde{x},\tilde{z}}(t)$, determined at time $t$, from time $s$ to $T$, i.e., $\tilde{a}^t(\tau)$, $s \leq \tau \leq T$, belongs to the structured control set $A^S_{\tilde{x},\tilde{z}}(s)$ determined at time $s$ associated with $x_s$ and $z_s$.

Clearly, if the structure $S$ is independent of the state $x$, $z$ and time $t$, it is ST-consistent.

With STC, a suboptimal upper value in (2.19) has an improving property.

**Theorem 1** Under a ST-consistent control structure $S$, function $\tilde{V}(x, z)$ in (2.19) satisfies (2.20) for any $x \in X$, $z \in Z^M$ with $\Delta t$ for $0 \leq \Delta t \leq T - t$.

\[
\tilde{V}(x, z) \geq \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} \left\{ \int_t^{t+\Delta t} G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) d\tau + \tilde{V}(x_{t+\Delta t}; x, a, \beta[a], z_{t+\Delta t}; z) \right\}
\]

\[\text{(2.20)}\]

**Proof:** Given an arbitrary $\varepsilon > 0$ and for any $x \in X$ and $z \in Z^M$, there exists $\tilde{a}^t(\cdot) \in A^S_{\tilde{x},\tilde{z}}(t)$, such that

\[
\tilde{V}(x, z) \geq \sup_{\beta \in \Delta(t)} \left\{ \int_t^{t+\Delta t} G(x_\tau, z_\tau, \tilde{a}^t_\tau, \beta[\tilde{a}^t]_\tau) d\tau + \tilde{V}(x_{t+\Delta t}; x, \tilde{a}^t, \beta[\tilde{a}^t], z_{t+\Delta t}; z) \right\} - \varepsilon
\]

\[\text{(2.21)}\]
for any \( 0 \leq \Delta t \leq T - t \). In (2.21), the equality is due to the principle of DP. By the STC of the structure \( S \), the second term from \( t + \Delta t \) to \( T \) in (2.21) satisfies

\[
\sup_{\beta \in \Delta(t+\Delta t)} \left\{ \int_{t+\Delta t}^{T} G(x_\tau, z_\tau, \tilde{a}_\tau^t, \beta[\tilde{a}^t]_\tau) d\tau + Q(x_T) \right\} \\
\geq \sup_{\beta \in \Delta(t+\Delta t)} \left\{ \inf_{a(\cdot) \in A^{S}_{x_{t+\Delta t}, z_{t+\Delta t}}(t+\Delta t)} \left\{ \int_{t+\Delta t}^{T} G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) d\tau + Q(x_T) \right\} \right\} \\
= \tilde{V}(x_{t+\Delta t}; x, \tilde{a}_t^t, \beta[\tilde{a}^t], z_{t+\Delta t}; x, z), \tag{2.22}
\]

Substitute (2.22) into (2.21),

\[
\tilde{V}(x, z) \geq \sup_{\beta \in \Delta(t)} \left\{ \int_{t}^{t+\Delta t} G(x_\tau, z_\tau, \tilde{a}_\tau^t, \beta[\tilde{a}^t]_\tau) d\tau + \tilde{V}(x_{t+\Delta t}; x, \tilde{a}_t^t, \beta[\tilde{a}^t], z_{t+\Delta t}; x, z) \right\} - \varepsilon. \tag{2.23}
\]

Note that \( \tilde{a}^t(\cdot) \in A(t) \). It follows from (2.23) that

\[
\tilde{V}(x, z) \geq \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} \left\{ \int_{t}^{t+\Delta t} G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) d\tau + \tilde{V}(x_{t+\Delta t}; x, a, \beta[a], z_{t+\Delta t}; x, z) \right\} - \varepsilon \tag{2.24}
\]

Since \( \varepsilon \) is arbitrary, the proof is completed.

Remark 3: Note that the optimization of the pursuers over \([t, t + \Delta T]\) in (2.20) is taken over all possible strategies in \( A(t) \) instead of \( A^{S}_{x,z}(t) \). It indicates that this performance improvement results from the structure relaxation in the optimization.13

Theorem 1 states that \( \tilde{V} \) can be improved by the optimization based on limited look-ahead if \( \tilde{V} \) is associated with a control structure of STC. The following analysis is based on this limited look-ahead formulation, where the function \( \tilde{V} \) on the right side of (2.20) is referred to as “cost-to-go” function.

13We will further justify this observation by a specific example in Section 5.2.2 in Chapter 5.
2.4 Iterative Improvement Based on Limited Look-ahead

In this section, we discuss an iterative process based on a suboptimal solution that results from structured controls. Denote by \( W \) a function such that \( W \in \mathbb{W} \triangleq \{ \psi : X \times Z^M \rightarrow \mathbb{R} \} \). Let \( \Delta t > 0 \) be a (small) look-ahead interval. Define map \( \overline{H} : \mathbb{W} \rightarrow \mathbb{W} \) as

\[
\overline{H}[W](x, z) = \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} \left\{ \int_t^{t+\Delta t} G(x_\tau, z_\tau, a_\tau, \beta[\cdot]_\tau) d\tau \right. \\
+ W(x_{t+\Delta t}, z_{t+\Delta t}, \beta[\cdot]_{t+\Delta t}) \left. \right\} \quad \text{subject to (2.12) and (2.9)}, \quad (2.25)
\]

where \( z \neq 0 \), \( W \in \mathbb{W} \) and \( t + \Delta t = (t + \Delta t) \land T \triangleq \min\{t + \Delta t, T\} \)\(^{14}\). Note that transformation \( \overline{H} \) depends on \( \Delta t \); however, the dependence is suppressed here because the following study is independent of \( \Delta t \). In (2.25), a differential game problem with a shorter duration is solved with \( W \) as the terminal cost (or cost-to-go at time \( t + \Delta t \)).

Introduce a notation \( \tilde{z}^x \in Z^M \) to indicate the evaders that are captured according to \( x \in X \), i.e., the \( j \)th element of \( \tilde{z}^x \), \( \tilde{z}_j^x = 0 \) if \( \| P(x_i^x) - P(x_j^x) \| \leq \varepsilon \) for some \( i \), and \( \tilde{z}_j^x = 1 \) otherwise. For any state \( x \in X \), \( z \in Z^M \) at time \( t \), define \( z^x \) as follows.

\[
z_j^x = \tilde{z}_j^x \land z_j \quad \text{for } j = 1, \cdots, M \quad (2.26)
\]

Due to the instantaneous state transition of \( z \) along the corresponding trajectory \( x \) defined in (2.9) and the integral in the objective (2.25), clearly,

\[
\overline{H}[W](x_t, z_t) = \overline{H}[W](x_t, z_t^x) \quad (2.27)
\]

for any \( x_t \in X \) and \( z_t \in Z^M \).

\(^{14}\)Henceforth, without explicit indication, in an operation similar to \( \overline{H} \), the optimization is subject to the equations (2.12) and (2.9).
In the following, we focus on a subset of $\mathcal{W}$, i.e.,

$$
\mathcal{W}_\geq \triangleq \{ W \mid W \in \mathcal{W}, W(x,z) \geq L \text{ for some } L \in \mathbb{R};
W(x,z) \geq H[W](x,z) \text{ for any } x \in X, z \in Z^M, z \neq 0 \}.
$$

(2.28)

The set $\mathcal{W}_\geq$ contains all the functions that are bounded from below and decreasing with respect to the transformation $H$. Given any $W_0 \in \mathcal{W}_\geq$, a sequence of functions $\{W_k\}_{k=0}^\infty$ can be generated by $W_{k+1} = H[W_k]$. By (2.27),

$$
W_k(x_t, z_t) = W_k(x_{t-1}, z_{t-1}) \text{ for each } k \in \mathbb{Z}^+.
$$

(2.29)

The following theorem shows that $H[W] \in \mathcal{W}_\geq$ if $W \in \mathcal{W}_\geq$, which implies that $W_k \in \mathcal{W}_\geq$ for any $k \in \mathbb{Z}^+$ if $W_0 \in \mathcal{W}_\geq$. In addition, the sequence is decreasing so that it converges.

**Theorem 2**

(i) If $W_0(x,z) \in \mathcal{W}_\geq$, then the sequence $\{W_k\}_{k=0}^\infty$ converges pointwise; (ii) the limiting function $W_\infty(x,z) \triangleq \lim_{k \to \infty} W_k(x,z)$ satisfies $W_\infty(x,z) = H[W_\infty](x,z)$.

**Lemma 1**

Let $f : X \times Y \to \mathbb{R}$ and $g : X \times Y \to \mathbb{R}$ be two functions on $X$ and $Y$. If $f(x,y) \leq g(x,y)$ for any $x \in X, y \in Y$, then

$$
\sup_{y \in Y} \inf_{x \in X} \{ f(x,y) \} \leq \sup_{y \in Y} \inf_{x \in X} \{ g(x,y) \}, \text{ and }
\inf_{x \in X} \sup_{y \in Y} \{ f(x,y) \} \leq \inf_{x \in X} \sup_{y \in Y} \{ g(x,y) \}.
$$

for any $x \in X, y \in Y$. 30
Remark 4 The conclusion in Lemma 1 can be extended to a functional space.

Proof: [Proof of Theorem 2] (i) We first prove that $W_k \in \tilde{W}_\geq$ for $k \in \mathbb{Z}^+$ by induction. Suppose that $W_k \in \tilde{W}_\geq$ ($k \in \mathbb{Z}_{\geq 0}$), and we want to show that $W_{k+1} \in \tilde{W}_\geq$.

By definition,

$$W_{k+1}(x, z) = \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} \left\{ \int_t^{t+\Delta t} G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) \, d\tau + W_k(x_{t+\Delta t}, x, z_{t+\Delta t}, x, z) \right\}$$

Note that $W_k \in \tilde{W}_\geq$ implies $W_{k+1}(x, z) \leq W_k(x, z)$ for any $x \in X$, $z \in Z^M$. By Lemma 1,

$$W_{k+1}(x, z) = \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} \left\{ \int_t^{t+\Delta t} G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) \, d\tau + W_k(x_{t+\Delta t}, x, z_{t+\Delta t}, x, z) \right\} \geq W_k(x_{t+\Delta t}, x, z_{t+\Delta t}, x, z)$$

By (2.30), clearly, $W_{k+1} = \overline{W}_k \in \tilde{W}_\geq$. Since $W_0 \in \tilde{W}_\geq$, by induction, $W_k \in \tilde{W}_\geq$ for $k \in \mathbb{Z}^+$. In addition, the sequence $\{W_k\}$ is decreasing at any $x \in X$, $z \in Z^M$.

Therefore, $\{W_k\}$ converges point-wisely due to the fact that $W_k$ is bounded from below for any $k \in \mathbb{Z}^+$. Define $W_\infty(x, z) \triangleq \lim_{k \to \infty} W_k(x, z)$.

(ii) For any $k \in \mathbb{Z}^+$, $W_k(x, z) \geq W_\infty(x, z)$ and it follows that

$$W_k(x, z) \geq \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} \left\{ \int_t^{t+\Delta t} G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) \, d\tau + W_\infty(x_{t+\Delta t}, x, z_{t+\Delta t}, x, z) \right\}$$

Let $k \to \infty$, and then

$$W_\infty(x, z) \geq \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} \left\{ \int_t^{t+\Delta t} G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) \, d\tau + W_\infty(x_{t+\Delta t}, x, z_{t+\Delta t}, x, z) \right\}$$

(2.31)
On the other hand, for any $k \in \mathbb{Z}^+$,

$$W_\infty(x, z) \leq \sup_{\beta \in \Delta(t)} \inf_{a \in A(t)} \left\{ \int_t^{t + \Delta t} G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) d\tau + W_k(x_{\tau + \Delta t}, a, \beta[a]_\tau, z_{\tau + \Delta t}; x, z) \right\}$$

Similarly, let $k \to \infty$,

$$W_\infty(x, z) \leq \sup_{\beta \in \Delta(t)} \inf_{a \in A(t)} \left\{ \int_t^{t + \Delta t} G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) d\tau + \sup_{\beta \in \Delta(t)} \inf_{a \in A(t)} \left\{ \int_t^{t + \Delta t} G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) d\tau + W_\infty(x_{\tau + \Delta t}, a, \beta[a]_\tau, z_{\tau + \Delta t}; x, z) \right\}$$

(2.32)

From (2.31) and (2.32), $W_\infty(x, z) = \mathcal{H}[W_\infty](x, z)$.

Theorem 2 states that starting from some function $W_0$ in $\overline{W} \geq$, a fixed point of the transformation $\overline{H}$ in space $\overline{W} \geq$ can be approached by iteration of $\overline{H}$. Recall that the suboptimal upper Value function $\tilde{V}(x, z)$ determined from a ST-consistent suboptimal control satisfies (2.20) and bounded from below. Therefore, $\tilde{V} \in \overline{W} \geq$. Let $W_0 = \tilde{V}$, and the decreasing sequence $\{W_k\}_{k=0}^\infty$ can be generated by $\overline{H}$. Each $W_k$ in the sequence is an upper-bound of $V$ and $W_k \leq W_l$ for $0 \leq l < k$. In the following, we shall show that $W_\infty = \overline{V}$ if $W_0 = \tilde{V}$. Here, we focus on the upper Value in (2.17), and a similar process can also be applied to the lower Value $\underline{V}$.

**Lemma 2** The upper Value function $\overline{V}$ in (2.17) satisfies

$$\overline{V}(x, z) = \sup_{\beta \in \Delta(t)} \inf_{a \in A(t)} \left\{ \int_t^{t_m \wedge T} G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) d\tau + \overline{V}(x_{t_m \wedge T}, a, \beta[a]_\tau, z_{t_m \wedge T}, z) \right\}$$

for any $t_m \geq t$, $x \in X$ and $z \in Z^M$.

**Proof:** This is the principle of DP. If $t_m \geq T$, since

$$\overline{V}(x_{T}, a, \beta[a], z_{T}, z) = Q(x_T),$$

the equality in (2.33) is trivial. If $t_m < T$, the proof of Theorem 3.1 in [69] is valid.

32
Lemma 3  For a real-valued function $W(x, z)$, if there exists $\Delta t > 0$, such that for any $x \in X$ and $z \in Z^M$,

$$W(x, z) = \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} \left\{ \int_t^{t+\Delta t} \left( G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) d\tau + W(x_{t+\Delta t}; a_\tau, \beta[a]_\tau, z_{t+\Delta t}; z, x) \right) \right\},$$

(2.34)

then for any $K \in \mathbb{Z}^+$,

$$W(x, z) = \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} \left\{ \int_t^{t+K\Delta t} \left( G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) d\tau + W(x_{t+K\Delta t}; a_\tau, \beta[a]_\tau, z_{t+K\Delta t}; z, x) \right) \right\},$$

(2.35)

for any $x \in X$ and $z \in Z^M$.

Proof: Define $\tilde{W}(x, z; \lambda)$ as

$$\tilde{W}(x, z; \lambda) = \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} \left\{ \int_t^{t+\lambda} \left( G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) d\tau + W(x_{t+\lambda}; a_\tau, \beta[a]_\tau, z_{t+\lambda}; z, x) \right) \right\},$$

(2.36)

First, consider $K = 2$ and let $\lambda = 2\Delta t$. The equality in (2.35) reduces to (2.34) if $T \leq t + \Delta t$. If $T > t + \Delta t$, let $t_m = t + \Delta t$. By Lemma 2,

$$\tilde{W}(x; 2\Delta t) = \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} \left\{ \int_t^{t+\Delta t} \left( G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) d\tau + \tilde{W}(x_{t+\Delta t}; a_\tau, \beta[a]_\tau, z_{t+\Delta t}; z, x) \right) \right\}. \quad (2.37)$$

By (2.36) and the hypothesis in (2.34),

$$\tilde{W}(x_{t+\Delta t}; z_{t+\Delta t}; \Delta t) = W(x_{t+\Delta t}; a_\tau, \beta[a]_\tau, z_{t+\Delta t}; z, x).$$

Substitute it into (2.37) and by (2.34),

$$\tilde{W}(x; 2\Delta t) = \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} \left\{ \int_t^{t+\Delta t} \left( G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) d\tau + W(x_{t+\Delta t}; a_\tau, \beta[a]_\tau, z_{t+\Delta t}; z, x) \right) \right\} = W(x, z). \quad (2.38)$$
Thus, (2.35) holds for $K = 2$.

Now, consider $K = 3$. When $T \leq t + 2\Delta t$,

$$\tilde{W}(x, z; 3\Delta t) = \tilde{W}(x, z; 2\Delta t) = W(x, z).$$

If $T > t + 2\Delta t$, choose $t_m = t + 2\Delta t$, and by Lemma 2,

$$\tilde{W}(x, z; 3\Delta t) = \sup_{\beta \in \Delta(t)} \inf_A \left\{ \int_t^{t+2\Delta t} G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) d\tau + \tilde{W}(x_{t+2\Delta t}, a, \beta[a], z_{t+2\Delta t}; x, z; \Delta t) \right\}. \quad (2.39)$$

Note that $\tilde{W}(x, z; \Delta t) = W(x, z)$ for any $x \in X$ and $z \in Z^M$. According to (2.38), $\tilde{W}(x, z; 3\Delta t) = W(x, z)$. By induction, this statement is true for any $K \in \mathbb{Z}^+$. ■

**Theorem 3** Suppose that $W_0 = \tilde{V}$. For the sequence $\{W_k(x, z)\}_{k=1}^\infty$ with $W_{k+1} = \overline{W}[W_k]$, its limit $W_\infty(x, z) = \lim_{k \to \infty} W_k(x, z) = \overline{V}(x, z)$ for any $x \in X$, $z \in Z^M$.

**Proof:** It is trivial when $z^x = 0$, i.e., $W_k(x, z) = W_\infty(x, z) = Q(x) = \overline{V}(x, z)$ for $k \in \mathbb{Z}^+$. Now, we consider $z^x \neq 0$. By Theorem 1, $W_0 = \tilde{V} \in W_\geq$. Theorem 2 implies that $\{W_k(x, z)\}_{k=1}^\infty$ is decreasing and its limit satisfies

$$W_\infty(x, z) = \sup_{\beta \in \Delta(t)} \inf_A \left\{ \int_t^{t+\Delta t} G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) d\tau + W_\infty(x_{t+\Delta t}, a, \beta[a], z_{t+\Delta t}; x, z) \right\}. \quad (2.40)$$

By Lemma 3,

$$W_\infty(x, z) = \sup_{\beta \in \Delta(t)} \inf_A \left\{ \int_t^{t+K\Delta t} G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) d\tau + W_\infty(x_{t+K\Delta t}, a, \beta[a], z_{t+K\Delta t}; x, z) \right\}$$

for any $K \in \mathbb{Z}^+$. Furthermore, by Assumption 2, $W_0 < \infty$. Since $G \geq \delta > 0$ and $Q \geq 0$, the terminal time $T$ of a game starting from any $x$, $z$ and time $t$ satisfies that
\( T \leq t + W_0(x, z)/\delta \), which is finite. Therefore, there exists an integer \( K_0 \in \mathbb{Z}^+ \) such that \( W_0(x, z)/\delta \leq K_0 \Delta t \), i.e., \( T \leq t + K_0 \Delta t \). Let \( K = K_0 \) in (2.40), and then

\[
W_\infty(x, z) = \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} \left\{ \int_t^{t+K_0 \Delta t} G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) \, d\tau + W_\infty(x_{t+K_0 \Delta t}, z_{t+K_0 \Delta t}, a_{t+K_0 \Delta t}, \beta_{t+K_0 \Delta t}) \right\}
\]

for any \( x \in X \) and \( z \in Z^M \).

So far, we have provided a complete recipe for solving a deterministic multi-player PE differential game. By iteration of the optimization over limited look-ahead intervals based on a suboptimal solution determined by a ST-consistent control, the dilemma of the conventional differential game theory in a multi-player PE game is circumvented.

### 2.5 Finite Convergence on a Compact Set

In the previous section, we have shown that starting from certain initial suboptimal upper Value, the real upper Value can be approached iteratively. In this section, we show that the convergence can be achieved in a finite number of steps on a compact subset of \( X \). Let \( W_0 = \overline{V} \), and then \( W_k(x, z) \geq 0 \) for each \( k \in \mathbb{Z}_+ \). Based on the \( k^{th} \) function \( W_k(x, z) \) in the sequence, we can define a level set

\[
\Omega_r^k \triangleq \left\{ (x, z) \in X \times Z^M \mid W_k(x, z) < r \right\} \text{ for some } r \in \mathbb{R}_{\geq 0}.
\]

For now, we assume the continuity of \( W_k \) in \( x \), and later in Section 2.7.2 (Theorem 6), we will provide sufficient conditions, under which \( W_k \) is continuous in \( x \) given \( z \in Z^M \). If \( W_k \) is continuous, then \( \Omega_r^k \) is open. In addition, given any \( r \in \mathbb{R}_{\geq 0} \),
the sets $\Omega_k^r$ associated with different $k$ are nested, i.e., $\Omega_k^r \subseteq \Omega_{k+1}^r$ for any $k \in \mathbb{Z}_{\geq 0}$.
This is because that the sequence is decreasing. Define $\Omega_r = \cup_{k=0}^{\infty} \Omega_k^r$ and clearly, $\Omega_r = \left\{ (x,z) \in X \times Z^M | \mathcal{V}(x,z) < r \right\}$.

Lemma 4 For any $r \in \mathbb{R}_{\geq 0}$, there exists $N \in \mathbb{Z}^+$ such that for any $(x,z) \in \Omega_r$, $W_n(x,z) = \mathcal{V}(x,z)$ for any $n \geq N$, and thus $\Omega_r = \Omega_n^r$.

Proof: It is trivial if $\Omega_r = \emptyset$. Hence, we only consider $\Omega_r \neq \emptyset$. Since $G \geq \delta > 0$, given $r \in \mathbb{R}_{\geq 0}$, for any $(x,z) \in \Omega_r$, the game terminates by a finite time $T$ under certain strategy of the players. In addition, $T \leq t + r/\delta$. Next, we shall show that for any state $(x,z) \in X \times Z^M$, if the game ends by a finite time $T$, the sequence $\{W_k(x,z)\}_{k=0}^{\infty}$ converges to the upper Value in a finite number of steps. First, consider $r_1 = \delta \cdot \Delta t$. For any $(x,z) \in \Omega_{r_1}$, the game ends by $\Delta t$ and

$$W_1(x,z) = \sup_{\beta \in \Delta(t) a(\cdot) \in A(t)} \inf a(\cdot) \in A(t) \left\{ \int_t^{t+\Delta t} G(x, z, a, \beta | a \rangle \Delta t + W_0(x_{t+\Delta t}, a, \beta | a \rangle ) \right\}$$

$$= \sup_{\beta \in \Delta(t) a(\cdot) \in A(t)} \inf a(\cdot) \in A(t) \left\{ \int_t^{T} G(x, z, a, \beta | a \rangle \Delta t + Q(x_T) \right\}$$

$$= \mathcal{V}(x,z) \leq r_1. \quad (2.41)$$

The equality above is because that $W_0(x,z) = \mathcal{V}(x,z) = Q(x)$ at the terminal of a game for any $x \in X$, $z \in Z^M$ with $z^x = 0$. Thus, $(x,z) \in \Omega_{r_1}^1$, which implies $\Omega_{r_1} \subseteq \Omega_{r_1}^1$, and due to the nested property, i.e., $\Omega_{r_1}^1 \subseteq \Omega_{r_1}$, then $\Omega_{r_1} = \Omega_{r_1}^1$. Next, we consider $\Omega_{r_2}$ for $r_2 = \delta \cdot 2\Delta t$. For any $(x,z) \in \Omega_{r_2}$, by definition,

$$W_2(x,z) = \sup_{\beta \in \Delta(t) a(\cdot) \in A(t)} \inf a(\cdot) \in A(t) \left\{ \int_t^{t+\Delta t} G(x, z, a, \beta | a \rangle \Delta t + W_1(x_{t+2\Delta t}, a, \beta | a \rangle ) \right\} \quad (2.42)$$

36
If \( t + \Delta t \leq T \), (2.42) reduces to (2.41), i.e.,
\[
W_2(x, z) = \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} \left\{ \int_t^T G(x_\tau, z_\tau, a_\tau, \beta[a_\tau]) \, d\tau + Q(x_T) \right\} = \overline{v}(x, z) \leq r_2.
\]

If \( t + \Delta t > T \), by the principle of DP,
\[
W_2(x, z) = \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} \left\{ \int_t^{t+\Delta t} G(x_\tau, z_\tau, a_\tau, \beta[a_\tau]) \, d\tau + W_1(x_{t+\Delta t}, x, a, \beta[a], z_{t+\Delta t}; x) \right\} = \overline{v}(x, z) \leq r_2. \tag{2.43}
\]

In (2.43), the second equality is because the game must end by \( 2\Delta t \), and thus
\[
W_1(x_{t+\Delta t}, x, a, \beta[a], z_{t+\Delta t}; x) = \overline{v}(x_{t+\Delta t}, x, a, \beta[a], z_{t+\Delta t}; x).
\]

Therefore, \( \Omega_{r_2} \subseteq \Omega_{r_2}^2 \) and then \( \Omega_{r_2} = \Omega_{r_2}^2 \). This process can be continued for \( N \) steps until \( N \geq r/(\delta \cdot \Delta t) \), such that for any \((x, z) \in \Omega_r\), \( W_N(x, z) = \overline{v}(x, z) \), which implies \( \Omega_r = \Omega_r^N \). This is also true for any \( n \geq N \).

Now, construct any sequence \( \{r_k\}_{k=1}^\infty \) with \( r_k \in \mathbb{R}_{\geq 0} \), \( r_k < r_{k+1} \) and \( \lim_{k \to \infty} r_k = \infty \). Clearly, \( \Omega_{r_k} \subseteq \Omega_{r_{k+1}} \). By Assumption 2, \( X = \Omega \triangleq \bigcup_{k=1}^\infty \Omega_{r_k} \). The following theorem shows that if \( W_k \) is continuous in \( x \) for any \( k \in \mathbb{Z}^+ \), the \( \{W_k\}_{k=1}^\infty \) converges to \( \overline{v} \) in a finite number of steps on any compact subset \( \Theta \) of \( X \), and the limiting function \( W_\infty = \overline{v} \) is continuous in \( x \).

**Theorem 4** Suppose that \( W_k(x, z) \) is continuous in \( x \) for any \( k \in \mathbb{Z}^+ \). If \( \Theta \) is a compact subset of \( X \), then

(i) there exists \( N \in \mathbb{Z}_{\geq 0} \) such that \( \overline{v}(x, z) = W_n(x, z) \) for any \( n \geq N \) and \((x, z) \in \Theta\);
(ii) \( \overline{v}(x, z) \) is continuous in \( x \) on \( \Theta \).

**Proof:** (i) Since \( \Theta \) is compact and \( W_0 \) is continuous, there exists \((x_0, z_0) \in \Theta\) such that \( W_0(x_0, z_0) = \sup_{(x, z) \in \Theta} \{ W_0(x_0, z_0) \} = r_0 \). Note that \( \overline{v}(x, z) \leq W_0(x_0, z_0) = r_0 \)
for \((x, z) \in \Theta \subseteq \Omega_{n_0}\). By Lemma 4, there exists positive integer \(N \in \mathbb{Z}^+\) such that 
\(V(x, z) = W_n(x, z)\) for any \(n \geq N\) and \((x, z) \in \Theta\). (ii) The continuity of \(W_k\) implies the continuity of \(V\) on \(\Theta\).

### 2.6 The Hierarchy Approach

With the iterative method introduced earlier, a multi-player PE game reduces to a problem of finding a suboptimal upper Value with the improving property. In this section, a hierarchical approach of STC is proposed, and the resulting suboptimal solution is a valid starting point for the iterative process. We focus on a class of games with the players’ dynamics in (2.12) with (2.9) and the following objective.

\[
J(a, b; x, z) = \int_t^T \left[ \sum_{j=1}^M z_j(t) \right] dt \tag{2.44}
\]

The objective in (2.44) stands for the sum of the capture time of each evader.

To avoid the difficulty of determining the terminal states in a multi-player PE game and that from discrete \(z\), we use a two-level approach [70]. The upper level is to determine a proper engagement scheme between the pursuers and the evaders, such that a multi-player PE game can be decomposed into distributed two-player PE games. At the lower level, the decoupled two-player games are solved. Suppose that each evader can be engaged with no more than one pursuer; and each pursuer may be engaged with more than one evader, in which case the two-player games involving different evaders and the same pursuer are solved sequentially. The evaders that are captured in later stages are assumed to know the strategy that the pursuer exploits in chasing previous evaders. Clearly, this hierarchical approach depends on the solvability of the underlying two-player PE games and we have the following assumption.
Assumption 3 Each evader $j$ can be captured by at least one pursuer $i_j$. In this case, we say that evader $j$ is capturable by pursuer $i_j$.

Suppose that the two-player PE game between evader $j$ and pursuer $i_j$ is solvable. We denote the upper Value of the two-player game by $V_{i_jj}(x_{ij})$ with $x_{ij} = [x_{ip}^T, x_{je}^T]^T$.

The detailed procedure of the approach is described as follows.

2.6.1 Two-player Pursuit-Evasion Games at the Lower Level

We start with the lower level, where the problem is to solve the upper Value function of series of two-player PE games between pursuer $i$ ($i = 1, \ldots, N$) and evader $j$ ($j = 1, \ldots, M$). Let $x_{ij} = [x_{ip}^T, x_{je}^T]^T$. The upper Value is

$$\nabla_{ij}(x_{ij}) = \sup_{\beta_j \in \Delta^j_i(t)} \inf_{a_i(\cdot) \in A^i(t)} \{ J(a_i, \beta_j; x_{ij}) \} = \sup_{\beta_j \in \Delta^j_i(t)} \inf_{a_i(\cdot) \in A^i(t)} \{ \int_t^T d\tau \}. \quad (2.45)$$

In (2.45), the strategy set $\Delta^j_i(t)$ of evader $j$ is defined similarly as $\Delta(t)$, i.e.,

$$\Delta^j_i(t) = \{ \beta_j : A^i(t) \rightarrow B^j(t) | \beta_j \text{ is nonanticipative} \}.$$  

In the two-player game, if pursuer $i$ cannot catch evader $j$ at $x_{ij}$, then $\nabla_{ij}(x_{ij}) = \infty$.

2.6.2 Combinatorial Optimization at the Upper Level

Denote by $E_i = \{ e_{i1}, \ldots, e_{ni} \}$ the ordered set of the evaders that are engaged with pursuer $i$ ($i = 1, \ldots, N$). Denote by $|E_i|$ the cardinal number of set $E_i$. If $E_i = \emptyset$, $|E_i| = n_i = 0$. If $E_i$ is determined, pursuer $i$ goes after the evaders in $E_i$ sequentially. Denote by $T^i_k$ the capture time of the $k^{th}$ evader in $E_i$. Let $x_{ip}[k]$ be pursuer $i$’s state at the time when it is ready to go after the $k^{th}$ evader in $E_i$, i.e., $x_{ip}[k] = x_{ip}^i(T^i_{k-1})$, and similarly for $x_{je}[k]$, the state of evader $j$. Let the game start at time 0. Suppose that the $k^{th}$ evader in $E_i$ exploits the best control against pursuer
Time from 0 to $T_k$. Define $x_{ij}[k] = [x_p^T[k], x_e^T[k]]^T$. The problem at the upper level is to determine an optimal engagement scheme between the pursuers and the evaders, and the corresponding optimization problem can be formulated as

$$\tilde{V}^h(x, z) = \min_{\{\sigma_{ijk}\}} \left\{ \sum_{i=1}^{N} \sum_{k=1}^{K} \sum_{j=1}^{M} V_{ij}(x_{ij}[k]) \right\}$$

subject to $\sigma_{ijk} \in \{0, 1\}$, $\sum_{i=1}^{N} \sum_{k=1}^{K} \sigma_{ijk} = 1$, $\sum_{j=1}^{M} \sigma_{ijk} \leq 1$ and $\sigma_{ij(l+1)} \leq \sigma_{ijl}$ for $l = 1, \cdots, K - 1$.

In (2.46), the superscript $h$ in $\tilde{V}$ indicates that the suboptimal upper Value is determined by the hierarchical approach; $k$ is the index of stages, i.e., the $k^{th}$ evader in $E_i$ (if any) associated with pursuer $i$ and $K$ is the maximum number of evaders allowed to be engaged with one pursuer ($K \leq M$); the assignment variables $\sigma_{ijk}$'s are binary, where $\sigma_{ijk} = 1$ indicates that pursuer $i$ is engaged with evader $j$ as the $k^{th}$ evader in $E_i$ and $\sigma_{ijk} = 0$ if it is not. Denote by $E \triangleq \{\sigma_{ijk}\}$ an engagement scheme. To emphasize, the problem in (2.46) is to find proper engagement $E$ such that the sum of the capture time of all evaders is minimized under the hierarchical structure. By Assumption 3, $\tilde{V}^h (x, z) < \infty$ for any $x \in X$ and $z \in Z^{M}$

Note that the formulation in (2.46) is based on an implicit assumption that $V_{ij}$ is known from the lower level for any pursuer $i$ and evader $j$. Clearly, in this approach, a hierarchical structure is imposed on the pursuers’ controls. More importantly, the best strategy of the evaders is determined “locally” by each evader (at the lower level) against its engaged pursuer at proper stages. It is not hard to show that this (composite) strategy by the evaders is optimal with respect to the optimization problem in (2.19).

Assumption 3 implies that Assumption 2 holds in the hierarchical approach.
**Remark 5** Suppose that Assumption 3 holds between any pursuer $i$ and evader $j$. If $N \geq M$ and $K = 1$, the optimization problem in (2.46) can be formulated as a standard Mixed Integer Linear Programming problem [70], such that the commercial solvers such as CPLEX and LINDO can be utilized [71, 72].

![Figure 2.1: Illustration of a Suboptimal Approach with Hierarchy](image)

The hierarchical approach introduced above is illustrated in Figure 2.1. This method is often used in complex systems, such as problem of cooperative control of multiple autonomous mobiles [71, 73, 72]. Proper abstraction of coordination is used at upper levels to decompose the problem into simpler sub-problems that are easily solved at lower levels. In multi-player PE game problems, simplification results from an additional hierarchical structure $S$ imposed on the pursuers’ controls. More importantly, the hierarchical approach is ST-consistent because the “structure” (the set of possible engagements between the pursuers and the “alive” evaders) remains same at any state $x$ and time given any $z \in Z^M$. Therefore, the resulting suboptimal upper Value is valid starting point for the iterative approach.
2.6.3 Uniform Continuity of the Suboptimal Upper Value

It is shown in Section 2.5 that the finite convergence on compact sets depends on the continuity of the suboptimal upper Value functions. In the section, we study the (uniform) continuity of $\tilde{V}^h$ that is obtained from the hierarchical approach. Note that the hierarchical approach to a multi-player PE game depends on the solvability of simplified two-player games, and intuitively the continuity of the suboptimal upper Value also depends on the upper Value of a two-player game. First of all, we use the following definition to classify certain two-player PE games.

**Definition 3 (Strong Capturability)** In a two-player PE game, the evader is said to be strongly capturable by the pursuer if for any $x_p \in \mathbb{R}^{n_p}$, $x_e \in \mathbb{R}^{n_e}$, 1) there exists a pursuer’s control $a_p(\cdot) \in A(t)$ such that the capture time $T'(x_p, x_e; a_p, b_e) < \infty$ for any $b_e(\cdot) \in B(t)$, where $T'(x_p, x_e; a_p, b_e)$ depends on states $x_p$, $x_e$ and strategies $a_p$ and $b_e$; 2) Define the optimal capture time $T(x_p, x_e)$ as

$$T(x_p, x_e) = \sup_{\beta_e \in \Delta(t)} \inf_{a_p(\cdot) \in A(t)} T'(x_p, x_e; a_p, \beta_e[a_p]).$$

For any $\varepsilon > 0$, there exists $\delta > 0$ such that for any $\| [\tilde{x}_e^T, \tilde{x}_p^T]^T - [x_e^T, x_p^T]^T \| < \delta$, $\| T(\tilde{x}_p, \tilde{x}_e) - T(x_p, x_e) \| < \varepsilon$ \(^{16}\).

**Remark 6** (i) Strong capturability requires the (uniform) continuity of $T(x_p, x_e)$ in $x_p$ and $x_e$. Refer to [9] for more details (e.g., Theorem 1.12 on pp.231).

---

\(^{16}\)Here, $x_p$ and $x_e$ are the state variable of the pursuer and the evader respectively; $a_p$ and $b_e$ are the control inputs; $\beta_e$ is a nonanticipative strategy of the evader.
(ii) If the evader is strongly capturable by the pursuer in a two-player PE game, then for any \( \varepsilon > 0 \), there exists \( \delta > 0 \) such that for any \( \| [\hat{x}_e^T, \hat{x}_p^T]^T - [x_e^T, x_p^T]^T \| < \delta \), there exists a control \( a_p(\cdot) \in A(t) \) such that \( |T'(x_p, x_e; a_p, b_e) - T'(\hat{x}_p, \hat{x}_e; a_p, b_e)| \leq \varepsilon \) for any \( b_e(\cdot) \in B(t) \).

**Lemma 5** Suppose that \( M \leq N \) and \( K = 1 \) in (2.46). If every pursuer \( i \) has the strong capturability of every evader \( j \), then \( \tilde{V}^h(x, z) \) in (2.46) is uniformly continuous in \( x \).

**Proof:** Consider any \( x^0, x^1 \in X, z \in Z^M \). Let \( E^{(x, z)} \) be the optimal engagements at \( x \) and \( z \) according to (2.46). Denote by \( J^h(x, z; E) \) the value of the cost evaluated similarly as in (2.46) under engagement \( E \) at \( x \) and \( z \). Clearly, \( J^h(x^l, z; E^{(x^l, z)}) = \tilde{V}^h(x^l, z) \) for \( l \in \{0, 1\} \). By the strong capturability, the upper Value \( V_{ij} \) defined in (2.45) is uniformly continuous. Suppose that evader \( j \) is engaged with pursuer \( i_E \) under \( E \). By the continuity of \( V_{ij} \) in (2.45), given any \( \varepsilon > 0 \), there exists \( \delta(E) > 0 \) such that

\[
\| V_{i_E j}^E(x^0_E) - V_{i_E j}^E(x^1_E) \| \leq \varepsilon / M, \quad \text{for any} \quad x^0_E - x^1_E \leq \delta(E) \quad \text{and} \quad j = 1, \ldots, M.
\]  

(2.47)

Considering that there are finitely many engagements of \( E \), we choose \( \delta = \min_E \{ \delta(E) \} \) (uniform in \( x \)). Thus, equation (2.47) holds for \( \| x^0 - x^1 \| \leq \delta \) under any \( E \). Let \( E_0 = E^{(x^0, z)} \) and \( E_1 = E^{(x^1, z)} \). For \( E_1 \), we rewrite (2.47) as \( V_{i_E j}^E(x^0_E) \leq V_{i_E j}^E(x^1_E) + \varepsilon / M \) for all \( j \). Take summation of \( V_{i_E j}^E \) over all \( j \),

\[
\tilde{V}^h(x^0, z) \leq J^h(x^0, z; E_1) = \sum_{j=1}^M V_{i_E j}^E(x^0_E) \leq \sum_{j=1}^M V_{i_E j}^E(x^1_E) + \varepsilon = \tilde{V}^h(x^1, z) + \varepsilon
\]  

(2.48)
On the other hand, we can obtain

\[ \tilde{V}^h(x^1, z) \leq J^h(x^1, z; E_0) = \sum_{j=1}^{M} V_{i_j}^{E_{0_j}}(x_{i_j}^1, E_{0_j}) \leq \sum_{j=1}^{M} V_{i_j}^{E_{0_j}}(x_{i_j}^0, E_{0_j}) + \varepsilon = \tilde{V}^h(x^0, z) + \varepsilon \tag{2.49} \]

By (2.48) and (2.49), \( \tilde{V} \) is continuous in \( x \). Moreover, since \( \delta \) does not depend on \( x \), the continuity is uniform.

**Theorem 5** Suppose that if evader \( j \) is capturable by pursuer \( i \) at the current state \( x \), then evader \( j \) is strongly capturable by pursuer \( i \). The suboptimal upper Value \( \tilde{V}^h(x, z) \) in (2.46) is bounded and uniformly continuous in \( x \) on any bounded region \( \Omega \subseteq X \).

**Proof:** See Appendix A for a detailed proof.

**Remark 7** Theorem 5 differs from Lemma 5 in that the singular engagement constraint imposed on the pursuers is relaxed.

### 2.7 Theoretical Foundation of the Iterative Method

In this section, we study the theoretical soundness of the iterative method. The study is conducted under the framework of viscosity solution theory of HJ equations. We provide sufficient conditions for a class of PE games under which the Value function (a saddle-point) exists.

#### 2.7.1 Viscosity Solution of Hamilton-Jacobi Equations

We first introduce the concept of viscosity solution of a HJ equation in the context of differential games. Before the invention of this (viscosity solution) concept, the applicability of DP methods was limited due to the requirement of differentiability.
of a Value function. However, even for a simple differential game (optimal control) problem, the Value function may not be differentiable, or at least not everywhere. To overcome this difficulty, Crandall and Lions initiated the study of the viscosity solution for HJ equations [43], with which the Value function of a differential game (optimal control) problem that is merely continuous or even discontinuous can be characterized by a viscosity solution of the corresponding HJI (HJB) equation [9].

We first consider a (two-player) differential game problem with the following dynamic equation of the players.

\[ \dot{x}_\tau = f(x_\tau, a_\tau, b_\tau) \text{ with } x(t) = x_t. \]

Here, \( x \in \mathbb{R}^n \); player 1's (minimizer) control \( a \) and player 2's (maximizer) control \( b \) are maps from \([t, T]\) to compact sets \( A_a \) and \( B_a \), where \( 0 < T < \infty \). The terminal time \( T \) can be fixed or not. In the latter case, \( T \) is called the exit time, which is defined as the time when the state trajectory first enters some predefined (terminal) set \( \Lambda \). Suppose that \( f \) is bounded and Lipschitz in \( x \) uniformly with respect to control \( a \) and \( b \). Consider the following cost functional

\[ J(a, b; t, x) = \int_t^T g(x_\tau, a_\tau, b_\tau) d\tau + q(x(T)), \quad (2.50) \]

where \( g : \mathbb{R}^n \times A_a \times B_a \to \mathbb{R} \) is bounded and Lipschitz in \( x \) as well as \( q : \mathbb{R}^n \to \mathbb{R} \). Here, we adopt the definitions of (nonanticipative) strategy and the (upper/lower) Value function. The (upper/lower) Value can be shown as the unique viscosity solution of the corresponding HJI equation [69, 9]. First of all, viscosity solution of a HJ equation is defined as follows.
Definition 4 Suppose $T$ is fixed. Let $H : [0, T] \times \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}$ be a continuous map.

(i) A continuous function $u : [0, T] \times \mathbb{R}^n \mapsto \mathbb{R}$ is called a viscosity subsolution of the Hamilton-Jacobi (HJ) equation

$$\frac{\partial \psi}{\partial t} + H(t, x, \frac{\partial \psi}{\partial x}) = 0 \text{ on } (0, T) \times \mathbb{R}^n \text{ with } \psi(T, x) = q(x),$$

(2.51)

provided that for any continuously differentiable function $\phi : (0, T) \times \mathbb{R}^n \mapsto \mathbb{R}$, if $u - \phi$ attains a local maximum at $(t_0, x_0) \in (0, T) \times \mathbb{R}^n$, then

$$\frac{\partial \phi}{\partial t}(t_0, x_0) + H(t_0, x_0, \frac{\partial \phi}{\partial x}(x_0, t_0)) \geq 0$$

(2.52)

(ii) A continuous function $u : [0, T] \times \mathbb{R}^n \mapsto \mathbb{R}$ is called a viscosity supersolution of the HJ equation (2.51) provided that for any continuously differentiable function $\phi : (0, T) \times \mathbb{R}^n \mapsto \mathbb{R}$, if $u - \phi$ attains a local maximum at $(t_0, x_0) \in (0, T) \times \mathbb{R}^n$, then

$$\frac{\partial \phi}{\partial t}(t_0, x_0) + H(t_0, x_0, \frac{\partial \phi}{\partial x}(x_0, t_0)) \leq 0$$

(2.53)

(iii) A continuous function $u : [0, T] \times \mathbb{R}^n \mapsto \mathbb{R}$ is called a viscosity solution of the HJ equation (2.51) if it is both a viscosity subsolution and a viscosity supersolution.

Remark 8 Definition 4 is based on the case when $T$ is fixed, and an initial condition at time $T$ is provided in (2.51). For a problem with a flexible terminal time as a PE game, the definition is same except that an additional boundary condition on $\partial \Lambda$ (the boundary of the terminal set $\Lambda$) is needed.

Now, we define

$$H^-(x, p) \triangleq \max_{b \in B_a} \min_{a \in A_a} \left\{ f(x, a, b) \cdot p + g(x, a, b) \right\} \text{ and}$$

(2.54)

$$H^+(x, p) \triangleq \min_{a \in A_a} \max_{b \in B_a} \left\{ f(x, a, b) \cdot p + g(x, a, b) \right\},$$

(2.55)
for \(x, p \in \mathbb{R}^n\). In [69], Evans and Souganidis proved that the lower (resp. upper) Value of a differential game formulated above with the objective functional in (2.50) (where \(T\) is fixed) is the (unique) viscosity solution of the corresponding HJI equation as in (2.51) with \(H\) replaced by \(H^-\) in (2.54) (resp. \(H^+\) in (2.55)). For problems with flexible \(T\), refer to [9] and [45].

2.7.2 Existence of the Value Function

Transformation \(\overline{H}\) is crucial for the iterative method based on limited look-ahead for deterministic PE games. In this section, we aim to show that under \(\overline{H}\), the image of \(W_k\) (\(k \in \mathbb{Z}^+\)) defined in (2.25) is a viscosity solution of the corresponding HJI equation to be formulated.

**Uniform Continuity of Function \(W_k\)**

We first show the uniform continuity of function \(W_k\) (\(k \in \mathbb{Z}^+\)) in the sequence generated by \(\overline{H}\) starting from \(\widetilde{V}\). We start with the definition of set \(\Lambda_z\). Given any \(z \in \mathbb{Z}^M(z \neq 0)\), define

\[
\Lambda_z = \left\{ x \in X \mid \exists j \in I_z \text{ such that } \| P(x_p^i) - P(x_e^j) \| \leq \varepsilon \text{ for some } i = 1, \ldots, N \right\},
\]

(2.56)

Here, recall the definition of \(I_z\) in (2.10). Clearly, the set \(\Lambda_z\) is closed. Let \(\Lambda_z^c = X - \Lambda_z\).

**Lemma 6** For the sequence \(\{W_k(x, z)\}^\infty_{k=0}\) generated by \(\overline{H}\) with \(W_0 = \widetilde{V}\), \(W_k(k \in \mathbb{Z}_{\geq 0})\) is (uniformly) continuous in \(x\) on \(X\) if and only if \(W_k\) is (uniformly) continuous in \(x\) on \(\overline{\Lambda_z^c}\) for any \(z \in \mathbb{Z}^M\). Here, the set \(\overline{\Lambda_z^c}\) is the closure of the set \(\Lambda_z^c\).

**Proof:** Recall that the closed set \(\Lambda_z\) in (2.56) as

\[
\Lambda_z = \left\{ x \in X \mid \exists j \in I_z \text{ such that } \| P(x_p^i) - P(x_e^j) \| \leq \varepsilon \text{ for some } i = 1, \ldots, N \right\}.
\]

47
The necessity is trivial due to $\Lambda^c_z \subseteq X$. We only need to show the sufficiency. First, when $z = 0$, the $W_k$ is continuous because $Q(x) = W_k(x, z)$ is (uniform) continuous. Now, consider $z \neq 0$. Suppose that $W_k$ is continuous in $x$ on $\overline{\Lambda^c_z}$ for any $z \in Z^M$.

Considering the fact that $\Lambda^c_z$ is open such that by the property in (2.29), i.e.,

$$W_k(x, z_t) = W_k(x, z^x_t),$$

the continuity of $W_k$ on $\Lambda^c_z \cup \text{int}(\Lambda_z)$ (interior) can be easily proved. Note that

$$\Lambda^c_z \cup \text{int}(\Lambda_z) \subseteq \cup_{z' \in Z_z} \Lambda^c_{z'}. \tag{2.57}$$

Refer to (2.10) on page 20 for the definition of set $Z_z$. Thus, we only need to show the continuity on the boundary $\partial \Lambda_z$.

Let $|z| = m$, where $m = 1, \cdots, M$. For any $x_0 \in \partial \Lambda_z$, by the definition of set $\Lambda_z$, it is possible that $|z^{x_0}| = m - 1, m - 2, \cdots, 0$. Refer to (2.26) on page 29 for the definition of $z^x$. We first consider $|z^{x_0}| = m - 1$. In this case, there exists $\delta_1 > 0$ such that for any $x \in B(x_0; \delta_1)$ (open ball), $z^x$ must be one of the following cases: $z^x = z$ or $z^x = z^{x_0}$. By the continuity of $W_k$ on $\overline{\Lambda^c_z}$, given any $\varepsilon > 0$, there exists $\delta_2 > 0$ such that for any $x \in B(x_0; \delta_1 \wedge \delta_2)$ with $z^x = z$, $\|W_k(x, z) - W_k(x_0, z)\| < \varepsilon$. Note that in this case, $x_0 \in \partial \Lambda_z$. Similarly, there exists $\delta_3 > 0$ such that for any $x \in B(x_0; \delta_1 \wedge \delta_3)$ with $z^x = z^{x_0}$,

$$\|W_k(x, z) - W_k(x_0, z)\| = \|W_k(x, z^{x_0}) - W_k(x_0, z^{x_0})\| < \varepsilon$$

because $x, x_0 \in \Lambda^c_z$. Let $\delta = \delta_1 \wedge \delta_2 \wedge \delta_3$. Then, $\|W_k(x, z) - W_k(x_0, z)\| < \varepsilon$ for any $x \in B(x_0; \delta)$. If $|z^{x_0}| = m - 2$, there exists $\delta_1 > 0$ such that for any $x \in B(x_0; \delta_1)$, $z^x$ must be one of the following: $z^x = z$, $|z^x| = m - 1$ with $z^x \in Z_z$ and $z^{x_0} \in Z_{z^x}$ or $z^x = z^{x_0}$. Similarly as in the case when $|z^x| = m - 1$, the continuity can be proved.
in each of the different cases. In particular, in the first two cases, \( x_0 \in \Lambda_z \), and in the last case \( x_0 \in \Lambda_{z_0} \), in which the continuity can be easily proved. This can be extended to other cases, such that it is true for \( |z^{z_0}| = m - 1, m - 2, \ldots, 0 \). Finally, the conclusion is true for uniform continuity.

Next, we show the uniform continuity of \( W_k \) in \( x \) given the uniform continuity of \( W_0 \). Let us first introduce some necessary notations. Recall that the terminal time \( T \) of a PE game can be defined as \( T = \min \{ t | z(t) = 0 \} \). Suppose that \( |z(0)| = M^{17} \).

There exist \( M \) time instants \( T^m \) with \( T^m \triangleq \min \{ t | |z(t)| = m \} \) for \( m = M - 1, M - 2, \ldots, 0 \). Clearly, \( T^0 = T \). We assume that when \( |z(t)| = m \) and if \( n (n > 1) \) evaders are captured at the same time, we assume that \( T^{m-1} = \cdots = T^{m-n} \).

**Theorem 6** If every evader is strongly capturable by every pursuer, and \( W_0 = \tilde{V} \) is uniformly continuous in \( x \) on \( X \) for any \( z \in Z^M \), \( W_k(x, z) \) is uniformly continuous in \( x \) for any \( k \in Z^+ \).

**Proof:** Refer to Appendix B for a detailed proof.

**Remark 9** Note that the uniform continuity of the suboptimal upper Value \( \tilde{V}^h \) by the hierarchal approach is proved in Theorem 5. Under the conditions in Theorem 5 and Theorem 6, \( W_k \) is uniformly continuous in \( x \).

**Existence of the Value Function**

In the following, we show that the image of transformation \( \overline{H} \) in the iteration process is the (unique) viscosity solution of the corresponding HJI equation.

\(^{17}\)The game starts at time 0.
First of all, we formulate a series of proper HJI equations according to different $z$’s with proper boundary conditions. In this way, the difficulty of the discrete variable $z$ in a multi-player PE game can be circumvented, such that the viscosity theory is applicable. Since the players’ dynamics are time invariant, we consider time $t = 0$ as the lower limit of the integral in $\overline{H}$. Define

$$\overline{H}_W(t, x, z) = \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} \left\{ \int_t^{\Delta t} G(x, z, a, \beta) \, dt + W_k(x, \alpha, \beta) \right\}$$

(2.58)

for $0 \leq t \leq \Delta t$. Note that $\overline{H}_W$ is a function of time $t$, and clearly, $\overline{H}_W(0, x, z) = \overline{H}(W_k)(x, z)$ for all $x \in X$ and $z \in Z^M$. Given a function $W_k$ and for any $z \in Z^M (z \neq 0)$, the corresponding HJI equation is

$$\frac{\partial u^z}{\partial t} + H^+_z(x, \frac{\partial u^z}{\partial x}) = 0,$$

(2.59)

with

$$u^z(\Delta t, x) = W_k(x, z) \text{ for } x \in \Lambda^c_z,$$

$$u^z(\tau, x) = u^z(\Delta t - \tau, x) \text{ for } \tau \in [0, \Delta t) \text{ if } x \in \partial \Lambda_z.$$

In (2.59), $H^+_z$ is defined as

$$H^+_z(x, p) = \min_{a \in A, b \in B_a} \max \left\{ p \cdot f(x, a, b) + G(x, z, a, b) \right\}.$$

The boundary condition of the equation on $\partial \Lambda_z$ is specified by the solution $(u^{z^x})$ of the corresponding HJI equation according to $z^x$, which is defined for $x \in \partial \Lambda_z$ according to (2.26). This boundary condition is due to the observation in (2.29). If $z^{x^\tau} = 0$, the boundary condition becomes $u^0(\Delta t - \tau, x) = Q(x)$. The following theorem shows that $\overline{H}_W$ in (2.58) with a given $z$ is a viscosity solution of the corresponding HJI equation in (2.59) on $\Lambda^c_z$. If $x \in \Lambda_z$, $\overline{H}_W(t, x, z)$ can be transferred to $\overline{H}_W(t, x, z^x)$ by (2.29).
**Theorem 7** Suppose that $W_k$ is uniformly continuous. For any $z \in Z^M (z \neq 0)$, $\overline{H}_{W_k}(t, x, z)$ is a viscosity solution of the HJI equation in (2.59) on $\Lambda^c_z$.

**Proof:** We use Definition 4 to prove the theorem. Consider an arbitrary $\phi_z \in C^1((0, \Delta t) \times \Lambda^c_z)$. Suppose that $\overline{H}_{W_k}(t, x, z) - \phi_z(t, x)$ attains a local minimum at $(t_0, x_0) \in ((0, \Delta t) \times \Lambda^c_z)$, i.e.,

$$
\overline{H}_{W_k}(t_0, x_0, z) - \overline{H}_{W_k}(t, x, z) \leq \phi_z(t_0, x_0) - \phi_z(t, x)
$$

for all $\|x_0 - x\| \leq \delta$ and some $\delta > 0$. Since $\Lambda^c_z$ is open, there exists a $\delta > 0$ such that $x \in \Lambda^c_z$ when $\|x_0 - x\| \leq \delta$. Due to the boundedness of the dynamics $f$, there exists $\lambda > 0$ such that $\|x_{t_0 + \gamma; x_0, a, b} - x_0\| < \delta$ for $0 < \gamma < \lambda$. Since viscosity solution is a local property, the rest of the proof follows that of Theorem 4.1 in [69].

Next, we show that $\overline{H}_{W_k}$ is the unique viscosity solution of (2.59). Suppose that the pursuers and the evaders have independent dynamics, which are described by $\dot{x}_{pt} = f_p(x_{pt}, a_t)$ and $\dot{x}_{et} = f_e(x_{et}, b_t)$ respectively. We assume that the following is true.

$$
\|f_\xi(x_\xi, \sigma)\| \leq C(|x_\xi| + 1) \quad \text{for all} \quad x_\xi \in \mathbb{R}^{n_\xi}, \quad \sigma \in \Sigma \quad \text{and some constant} \quad C > 0,
$$

(2.60)

where the subscript $\xi \in \{p, e\}$ stands for the pursuers or the evaders; $\sigma \in \Sigma$ is the control input and $\Sigma \in \{A_a, B_a\}$ is the set of admissible controls. The uniqueness of the viscosity solution of a general HJ equation is provided by Theorem V.2 in [43].

**Theorem 8** Suppose that function $f$ satisfies (2.4) and (2.60) (if considered as $f_p$ and $f_e$ separately in (2.5)), and the cost rate $G$ satisfies (2.15). Function $\overline{H}_{W_k}(t, x, z)$ is the unique viscosity solution of the HJI equation in (2.59) on any bounded subset of $\Lambda^c_z$. 

51
Proof: The proof is a simple application of Theorem V.2 in [43]. It can be easily checked that all the conditions therein are satisfied. Particularly, the condition (2.60)\(^{18}\) yields the following equality.

\[
\limsup_{r \to 0} \left\{ |H^+_r(x_1, p) - H^+_r(x_2, p)| \left| |x_1 - x_2|(1 + \|p\|) \leq r \right\} = 0
\]

Therefore, the difference (by infinite-norm) between any two viscosity sup- or sub-solutions on the domain can be bounded by the difference between them on the boundary. Since the viscosity solution \(\overline{H}_{W_k}(t, x, z)\) satisfies the boundary condition in (2.59), they must be same, and \(\overline{H}_{W_k}(t, x, z)\) is the unique viscosity solution of (2.59).

With the uniqueness of the viscosity solution for (2.59), we can show that the Value of a multi-player PE game exists under additional conditions. Given a function \(W \in \overline{W}\), define a transformation \(\overline{H}\) similar to \(\overline{H}\) in (2.25) as

\[
\overline{H}[W](x, z) = \inf_{\alpha \in \Gamma(t)} \sup_{b \in B(t)} \left\{ \int_t^{t+\Delta t} G(x_\tau, z_\tau, \alpha[\cdot]_\tau, b_\tau) \, d\tau + W(x_{t+\Delta t}; x, \alpha[\cdot]_\tau, b_\tau, z_{t+\Delta t}) \right\}.
\]

Define the set \(\overline{W}_\geq\) as

\[
\overline{W}_\geq \triangleq \{ W | W(x, z) \geq L \text{ for some } L \in \mathbb{R} \text{ and } \}
\]

\[
W(x, z) \geq \overline{H}[W](x, z), \quad x \in X, \quad z \in Z^M \}.
\]

(2.61)

Given any \(W \in \overline{W}_\geq\) (refer to (2.28) for the definition),

\[
W(x, z) \geq \overline{H}[W](x, z) \geq \overline{H}[W](x, z),
\]

(2.62)

\(^{18}\)Refer to Theorem 3.17 on page 159 in [9] for a detailed proof of the equality of the limit.
which implies that $\overline{W}_\geq \subseteq \overline{W}_\geq$. By inspection of the definition of $\overline{H}$ and $H$, the second inequality in (2.62) holds because of the fact that the lower Value of a game is no larger than the upper Value of the same game, i.e., $\overline{V} \geq \underline{V}$ (c.f. pp.434 in [9]).

Starting from the same function $W_0 = \overline{V} \in W_\geq$, it can be shown similarly that the sequence $\{W'_k\}_{k=0}^\infty$ generated by $W'_{k+1} = H[W'_k]$ converges to the lower Value defined in (2.16). The corresponding HJI equation is

$$\frac{\partial u_z}{\partial t} + H^- (x, \frac{\partial u_z}{\partial x}) = 0 \quad (2.63)$$

with the same boundary conditions as in (2.59), where $H^-$ is defined as

$$H^- (x, p) = \max_{b \in B} \min_{a \in A} \{p \cdot f(x, a, b) + G(x, z, a, b)\}.$$

If $H^- = H^+$ for any $z \in Z^M$, i.e.,

$$\max_{b \in B} \min_{a \in A} \{p \cdot f(x, a, b) + G(x, z, a, b)\} = \min_{a \in A} \max_{b \in B} \{p \cdot f(x, a, b) + G(x, z, a, b)\} \quad (2.64)$$

equation (2.63) coincides with (2.59) for any $z \in Z^M$. By the uniqueness of the viscosity solution, $W'_k = W_k$ for any $k \in \mathbb{Z}^+$, which implies that $\overline{V} = \underline{V}$. Therefore, the Isaacs condition holds and the Value function of the multi-player PE game exists.

According to equation (2.64), the existence of the Value of a multi-player PE differential game reduces to the existence of saddle-point equilibrium of a corresponding static zero-sum game. This problem was originally studied by von Neumann and has been extensively studied afterwards. Readers can refer to the Minimax theorem for more details. The following Theorem provides a simple case when (2.64) holds.

**Theorem 9** If function $f(x, a, b) : \mathbb{R}^n \times A_a \times B_a \mapsto \mathbb{R}^n$ and $G(x, z, a, b) : \mathbb{R}^n \times Z^M \times A_a \times B_a \mapsto \mathbb{R}^n$ are separable, i.e.,

$$f(x, a, b) = f_a(x, a) + f_b(x, b) \text{ and } G(x, z, a, b) = G_a(x, z, a) + G_b(x, z, b), \quad (2.65)$$

53
then for any \( p \in \mathbb{R}^n \),

\[
\min_{a \in A_a} \max_{b \in B_a} \{ p^T f(x, a, b) + G(x, z, a, b) \} = \max_{b \in B_a} \min_{a \in A_a} \{ p^T f(x, a, b) + G(x, z, a, b) \} \quad (2.66)
\]

for any \( x \in \mathbb{R}^n \), \( a \in A_a \) and \( b \in B_a \).

Refer to [9] on pp.444 for a proof.

In a multi-player PE game, the separation property of \( f \) holds because the pursuers and the evaders have independent dynamics. Thus, a separable \( G \) implies the existence of Value, e.g., \( G = 1 \) when the capture time is the objective functional. In summary, we have shown the existence of solutions for the transformation \( \mathcal{P} \) and the existence of Value under the framework of the viscosity solution theory.

### 2.8 Summary and Discussion

In this chapter, we propose an indirect method for multi-player PE games, on which the conventional differential game theory is not applicable. We start from a suboptimal upper Value of a PE game, where only subset of the pursuers’ controls under certain structure \( S \) is considered. If the structure \( S \) satisfies STC, the suboptimal upper Value has an improving property, i.e., it can be improved by the optimization based on limited look-ahead. The performance enhancement here results from the structure relaxation. If this performance enhancement is applied iteratively, a true upper Value is approached in the limit. Furthermore, we show that a control structure \( S \) that is independent of the state and time is ST-consistent, such that a suboptimal solution under \( S \) can be improved by limited look-ahead. The hierarchical method provide a valid starting point for the iterative method.

The iterative method can be viewed as an extension of the DP-type methods to multi-player PE differential games. The related DP methods in the literature include
value iteration in Approximate DP (ADP) [52, 55], reinforcement learning [74] and rollout algorithms [55]. All these methods are aimed to improve an existing suboptimal solution based on some approximate strategies. The first two methods mainly deal with MDP problems on finite state spaces [52, 75]. Note that MDP is a discrete-time optimal control problem with an infinite horizon, such that a suboptimal Value function or state feedback strategies can be easily updated by the optimization based on look-ahead. However, a multi-player PE game has a finite horizon, and the improving property of an suboptimal Value function is not obvious with “receding horizons” in contrast to MDP problems. From this aspect, the iterative method is more related to the rollout algorithm for stochastic control problems. The rollout algorithm is a suboptimal control method, where the control input is calculated by optimization over look-ahead intervals based on some (heuristic) base strategy [54, 55]. The performance of the base strategy can be improved by the rollout if the base strategy satisfies “sequential consistency” [55]. However, the “sequential consistency” in [55] requires that given a trajectory generated by an existing base strategy, starting from any state along the trajectory, the base strategy generates the remaining state trajectory. Since the consistency is established along a particular trajectory, it may fail in stochastic problems. In our iterative method, STC is a property of the “base strategy”, which does not depend particular trajectories. In this sense, STC is more general and can be extended to stochastic situations. Furthermore, the hierarchical approach is ST-consistent and also useful in many large-scale problems. Finally, the DP methods mentioned above deal with discrete-time problems of finite dimensions; however, we are more interested in continuous-time problems in a (infinite dimensional) function space, such that existence of solutions may be an issue.
This dissertation is the first attempt to solve a general multi-player PE differential game and although we have demonstrated that the iterative method is feasible on such problems, it is worth mentioning that the iterative method here suffers from the so-called “curse of dimensionality” in all DP related methods. A practical algorithm needs further study. To this end, those extensively studied numerical methods for solving DP equations may benefit the implementation of the iterative method in many ways. First of all, a continuous-time problem can be approximated by a corresponding discrete DP equation [53] and [9](Appendix A). For a discrete-time problem, numerical methods such as ADP and reinforcement learning have been studied for decades by researchers in control, operations research and artificial intelligence. There are plenty of results such as the value function approximation and real-time DP etc, which can be useful for implementation of the iterative method. We refer the reader to [75, 76, 74] for good reviews and recent development in that area.

Besides, the iterative method has its practical value in improving suboptimal solutions. First, it can be used to determine a satisfactory solution rather than an optimal solution, and the process can stop at any step to provide the best upper-bound of the upper Value because the sequence of functions is decreasing. On the other hand, from the look-ahead point of view, the performance of a suboptimal strategy under a ST-consistent structured can be improved along the pursuit trajectories when the strategies by the optimization over look-ahead intervals are exploited at $t + k\Delta t$ for $k \in \mathbb{Z}^+$ [55, 77]. In practice, such a approach based on a carefully chosen cost-to-go function can usually perform very well. Finally, one way of simplification is to implement the iterative improvement only on subgroups of the pursuers and the evaders that closely interact with one another.
In summary, this iterative method may suggest potential uses in other research areas. It can be applicable to those complex dynamic optimization problems such as cooperative control problems including search, routing and allocation etc [77]. By the feedback implementation of the optimization based on look-ahead, the performance is enhanced compared to the hierarchical method that is often used, and also the results may be more robust against uncertainties [73, 72].
CHAPTER 3

SIMULATION RESULTS OF DETERMINISTIC PURSUIT-EVASION GAMES

In this chapter, several simulation results are presented to demonstrate the feasibility of the methods introduced in Chapter 2. We first apply the iterative method to a two-player PE game for which the analytical solution is also derived. The convergence result obtained from the numerical calculation coincides with the prediction from the analytical solution. Furthermore, we solve a multi-player PE game for a suboptimal solution by the hierarchical method. We show that the performance of the hierarchical method can be improved by the optimization based on one-step look-ahead. Finally, we demonstrate that the advantage of limited look-ahead in solving a more complicated problem when no suboptimal solution is available. Unfortunately, due to the so-called “curse of dimensionality”, we were unable to implement a complete iterative process for multi-player PE games, and an efficient algorithm is yet to be developed.
3.1 Feasibility of the Iterative Method

To our knowledge, analytical results for multi-player PE differential games as formulated in this paper are very limited. Here, we apply the iterative method numerically to a two-player PE game in a two dimensional space. An analytical solution is also derived, and to which the numerical solution can be compared. This example is created based on a problem similar to that on page 429 in [7]. Consider the following dynamics of the pursuer and the evader,

\[
\begin{align*}
\dot{x}_p(t) &= \cos(\theta(t)) \\
\dot{y}_p(t) &= \sin(\theta(t)) \\
\dot{x}_e(t) &= b(t) + 1 + \sqrt{2} \\
\dot{y}_e(t) &= -2
\end{align*}
\] (3.1)

Here, \(x_p\) and \(y_p\) and \(x_e\) and \(y_e\) are the real-valued state variables of the pursuer (evader); \(\theta\) is the control of the pursuer; \(b\) is the control of the evader with \(-1 \leq b(t) \leq 1\) for \(t \geq 0\). Let the initial condition be \([x_{p0}, y_{p0}]^T = [0, 0]^T\) and \([x_{e0}, y_{e0}]^T = [2, 1]^T\).

The terminal set \(\Lambda\) of the game is defined as \(\Lambda \triangleq \{(x_p, y_p), (x_e, y_e) \mid y_p \geq y_e\}\). Note that according to (3.1), \((\dot{y}_e - \dot{y}_p) \leq -1\), thus, the game terminates in finite time. Let \(\ddot{x} = [x_p, y_p, x_e, y_e]^T\) and the objective is

\[
J(\theta, b; \ddot{x}_0) = \int_0^T \left( b(t) + 1 + \sqrt{2} - \cos(\theta(t)) \right) dt
\] (3.2)

The pursuer minimizes the objective (3.2) subject to (3.1) while the evader maximizes it.

3.1.1 Analytical Solution of the Two-player Game Example

We first solve the PE game problem analytically. Change the variable as \(x = x_e - x_p\) and \(y = y_e - y_p\), such that the players’ dynamic equation in (3.1) becomes

\[
\begin{align*}
\dot{x}(t) &= b(t) + 1 + \sqrt{2} - \cos(\theta(t)) \\
\dot{y}(t) &= -2 - \sin(\theta(t))
\end{align*}
\] (3.3)
The initial condition is \((x_0, y_0)^T = (2, 1)^T\) and the terminal set becomes \(\Lambda' \triangleq \{(x, y) | y \leq 0\}\). Let \(\bar{x} = [x, y]^T\). By inspection of (3.2) and (3.3), the objective function can be simplified as

\[
J'(\theta, b; \bar{x}) = x(T) - x_0.
\] (3.4)

Note that \(x_0\) does not depend on the controls of the players. Thus, a game with the objective (3.4) is equivalent to that with the objective \(J''(\theta, b; \bar{x}) = x(T)\). In the following, we solve the game based on \(J''\).

Denote the Value function associated by \(J''\) by \(V''(\bar{x})\). According to the objective \(J''\), the Hamiltonian \(H\) of the game is

\[
H = V''_x \left( b(t) + 1 + \sqrt{2} - \cos(\theta(t)) \right) + V''_y \left( -2 - \sin(\theta(t)) \right),
\]

\(^{19}\)The Value \(V''\) exists.
Here, $V_x''$ and $V_y''$ are the partial derivative of $V''$ with respect to $x$ and $y$. Taking derivatives of $H$ with respect to $x$ and $y$ and by the minimum principle,

$$\dot{V}_x'' = -\frac{\partial H}{\partial x} = 0, \quad \text{and} \quad \dot{V}_y'' = -\frac{\partial H}{\partial y} = 0.$$ 

Thus, $V_x''$ and $V_y''$ are constants. At the terminal, $V_x''(x,y)|_{y=0} = 1$, which implies $V_x'' = 1$. To determine $V_y''$, consider the “main equation”\(^\text{20}\) in [11] (on page 67),

$$\min_{\theta_i} \max_{b_i} H = \min_{\theta_i} \max_{b_i} \left\{ V_x''(b_i + 1 + \sqrt{2} - \cos(\theta_i)) + V_y''(-2 - \sin(\theta_i)) \right\} = 0. \quad (3.5)$$

Substitute $V_x'' = 1$ into (3.5), and by (3.5) the optimal controls $\theta_i^*$ and $b_i^*$ satisfy

$$b_i^* = 1, \quad \begin{cases} \cos(\theta_i^*) = \frac{1}{\sqrt{1 + V_y''^2}}, \\ \sin(\theta_i^*) = \frac{V_y''}{\sqrt{1 + V_y''^2}}. \end{cases} \quad (3.6)$$

Substitute (3.6) into (3.5),

$$2 + \sqrt{2} - 2V_y'' - \sqrt{1 + V_y''^2} = 0. \quad (3.7)$$

Solving (3.7), we obtain $V_y'' = 1$, which yields that $V''(\bar{x}) = x + y$. Note that equation (3.5) remains the same if the order of the minimization and the maximization is switched. Thus, $V''$ is the Value of the two-player PE game. The solution is also illustrated in Figure 3.1. The circle stands for the set of possible controls of the pursuer and the vector $\eta$ is the velocity vector specified in (3.3) under optimal $\theta_i^*$ and $b_i^*$. Finally, the Value associated with $J'$ is $V'(\bar{x}) = V''(\bar{x}) - x = y$.

### 3.1.2 Numerical Solution by the Iteration Method

In this section, we solve the game numerically using the iterative method. Suppose that the Value of the game and the corresponding optimal controls of the players are

\(^{20}\)Here, the “main equation” is actually a Hamilton-Jacobi-Isaacs (HJI) equation.
unknown. Choose \( \hat{\theta}_t = \frac{\pi}{2} \) \((t \geq 0)\) as a suboptimal control of the pursuer and the corresponding optimal control of the evader is, \( b^*_t = 1 \) \((t \geq 0)\). Note that under \((\hat{\theta}, b^*)\), the PE game ends with a minimum time. The speed vector \( \gamma \) under \((\hat{\theta}, b^*)\) is shown by a dashed line in Figure 3.1. The suboptimal Value function \( \tilde{V}(\bar{x}) \) associated with \((\hat{\theta}, b^*)\) is \( \tilde{V}(\bar{x}_0) = \frac{2 + \sqrt{7}}{3} y_0 \). Since suboptimal control \( \hat{\theta} \) does not depend on state \( \bar{x} \) and time \( t \), it is ST-consistent. Therefore, \( \tilde{V} \) has the improving property. Given \((x_0, y_0) = (2, 1)\) and let \( W_0 = \tilde{V} \), we apply the iteration method. Here, we discretize the state space with a grid size of 0.05 and choose \( \Delta t = 0.03 \). Note that \( \Delta t \) is so small that dynamic optimization problems over limited look-ahead intervals as in (2.20) can be approximated by static optimization problems where the players’ strategies during \( \Delta t \) intervals are constant. The evolution of \( W_k(\bar{x}) \) at \( \bar{x}_0 = [2, 1]^T \) is illustrated in Figure 3.2. It can be seen that starting from \( \tilde{V}(2, 1) = \frac{2 + \sqrt{7}}{3} \approx 1.1381 \), \( W_k(2, 1) \) converges to the optimal Value \( V'(2, 1) = 1 \) at iteration 14. The convergence is achieved at about iteration 14. Here, the convergence rate clearly depends on the size of the look-ahead interval \( \Delta t \).

### 3.2 Suboptimal Solution by the Hierarchical Approach

In this section, a multi-player PE game is solved by the hierarchical approach introduced in Section 2.6. We select a PE scenario involving 3 pursuers and 5 evaders in a \( \mathbb{R}^2 \) space. The dynamics of the players are

\[
\begin{align*}
\dot{x}_\zeta &= v_\zeta \cos \theta_\zeta, \\
\dot{y}_\zeta &= v_\zeta \sin \theta_\zeta, \\
\text{with} & \quad x_\zeta(0) = x_{\zeta 0}, \\
y_\zeta(0) &= y_{\zeta 0}.
\end{align*}
\]

Equation (3.8) is the first-order Dubin’s car model, where the subscript \( \zeta \in \{p, e\} \) stands for a pursuer or an evader; \( x_\zeta \) and \( y_\zeta \) are state variables (displacement) along the \( x- \) and \( y- \) axis respectively; \( v_\zeta \) is the velocity; \( \theta_\zeta \) is the control variable. In
Figure 3.2: The Evolution of the Suboptimal Value \( W_k \)'s at (2,1) by Iteration

this model, the players are assumed to perform at their maximum speeds and have complete maneuverability, which can approximately describe the dynamics of small ground vehicles. The terminal set \( \Lambda \) of the game is chosen as

\[
\Lambda = \left\{ (x_p, y_p), (x_e, y_e) \mid \sqrt{(x_p - x_e)^2 + (y_p - y_e)^2} \leq \varepsilon \right\},
\]

(3.9) for some \( \varepsilon > 0 \). Here, we choose \( \varepsilon \) as 0.5. In this multi-player game, the objective function in (2.44) is adopted, i.e., the sum of the capture times of all evaders is to be optimized. The necessary parameters and the initial states are given in Table 3.1.

A complete solution of this problem is hard to obtain, and here we solve for a suboptimal solution by the hierarchical approach. Since the hierarchical approach is based on the solvability of the corresponding two-player PE games, here we first provide an analytical solution to a two-player game with the players’ dynamics in (3.8) for \( v_p > v_e \) and the capture time \( T \) is chosen as the objective function.
Consider a two-player PE game, where the state variables include $x_\varsigma, y_\varsigma$ for $\varsigma \in \{p, e\}$. Define $x = x_e - x_p$ and $y = y_e - y_p$, such that the dynamics in (3.8) becomes

$$
\begin{align*}
\dot{x} &= v_e \cos(\theta_e) - v_p \cos(\theta_p) \\
\dot{y} &= v_e \sin(\theta_e) - v_p \sin(\theta_p)
\end{align*}
$$

with $x_0 = x_{e0} - x_{p0}$ and $y_0 = y_{e0} - y_{p0}$. The terminal set of the game is $\Lambda' = \{(x, y) | \sqrt{x^2 + y^2} \leq \varepsilon\}$. The objective is, $J = \int_0^T dt$.

Denote by $V$ the Value function of such a two-player PE game. The Hamiltonian of the game is

$$
H = V_x(v_e \cos(\theta_e) - v_p \cos(\theta_p)) + V_y(v_e \sin(\theta_e) - v_p \sin(\theta_p)).
$$

Here, $V_x$ ($V_y$) is the partial derivative of $V$ with respect to $x$ ($y$). By the minimum principle, $\dot{V}_x = -\frac{\partial H}{\partial x} = 0$ and $\dot{V}_y = -\frac{\partial H}{\partial y} = 0$. Thus, $V_x$ and $V_y$ are constant. By the main equation [11], i.e., $\min_{\theta_p} \max_{\theta_e} H = 0$, such that we obtain that the optimal strategies $\theta_p^*$ and $\theta_e^*$ satisfies

$$
\cos \theta_p^* = \cos \theta_e^* = \frac{V_x}{\sqrt{V_x^2 + V_y^2}} \quad \text{and} \quad \sin \theta_p^* = \sin \theta_e^* = \frac{V_y}{\sqrt{V_x^2 + V_y^2}}.
$$

### Table 3.1: The Necessary Parameters of the Players for PE Game Scenario 1

<table>
<thead>
<tr>
<th>Pursuers</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(x_{p0}, y_{p0})$</td>
<td>(3.0)</td>
<td>(5.0)</td>
<td>(7.0)</td>
</tr>
<tr>
<td>$v_p$ (1/sec)</td>
<td>6</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>$\omega_p$ (rad/sec)</td>
<td>0.8$\pi$</td>
<td>0.8$\pi$</td>
<td>0.8$\pi$</td>
</tr>
<tr>
<td>$\theta_{p0}$</td>
<td>$\pi/2$</td>
<td>$\pi/2$</td>
<td>$\pi/2$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Evaders</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(x_{e0}, y_{e0})$</td>
<td>(1.5)</td>
<td>(4.5)</td>
<td>(6.5)</td>
<td>(7.5)</td>
<td>(9.5)</td>
</tr>
<tr>
<td>$v_e$ (1/sec)</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>
and \( \sqrt{V_x^2 + V_y^2} = 1/(v_p - v_e) \). By the definition of the terminal set \( \Lambda' \), the terminal state \((x_T, y_T)\) satisfies \( V_x \cdot (-y_T) + V_y \cdot x_T = 0 \). It follows that

\[
V_x/x = V_y/y = \frac{1}{(v_p - v_e)\sqrt{x^2 + y^2}}. \tag{3.11}
\]

Furthermore, it can be shown from (3.11) that

\[
V(x, y) = \left(\sqrt{x^2 + y^2} - \varepsilon\right)/(v_p - v_e). \tag{3.12}
\]

Clearly, the optimal strategy of the pursuer and the evader are

\[
\cos \theta^*_p = \cos \theta^*_e = -\frac{x}{\sqrt{x^2 + y^2}}, \quad \sin \theta^*_p = \sin \theta^*_e = -\frac{y}{\sqrt{x^2 + y^2}}. \tag{3.13}
\]

Based on the solution to a two-player game, we can use the hierarchical approach to solve the multi-player PE game for a suboptimal solution. Now, solve the (combinatorial) optimization problem at the upper level with the objective function (2.46) with \( K = 3 \), and the optimal engagement obtained is given in Table 3.2\(^{21}\). The performance index (the sum of the capture times) under such engagements is evaluated as 26.03 (sec).

<table>
<thead>
<tr>
<th>Engaged Evaders</th>
<th>Pursuer 1</th>
<th>Pursuer 2</th>
<th>Pursuer 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>E1</td>
<td>E2</td>
<td>E3</td>
</tr>
<tr>
<td>2nd</td>
<td>N/A</td>
<td>N/A</td>
<td>E4</td>
</tr>
<tr>
<td>3rd</td>
<td>N/A</td>
<td>N/A</td>
<td>E5</td>
</tr>
</tbody>
</table>

Table 3.2: The Optimal Engagement Between the Pursuers and the Evaders with the First-Order Dubin’s Car Model

Next, we show the possibility of using the hierarchical approach to solve a game where the analytical solutions of the underlying two-player PE games are unknown.\(^{21}\) Here, \( E_j \) stands for evader \( j \).
Consider the same game except that the players have more complicated dynamics as follows.

\[
\begin{align*}
\dot{x}_p &= v_p \cos \theta_p, \\
\dot{y}_p &= v_p \sin \theta_p, \\
\dot{\theta}_p &= \omega_p a_t, \\
\dot{x}_e &= v_e \cos \theta_e, \\
\dot{y}_e &= v_e \sin \theta_e
\end{align*}
\]

with \(x_p(0) = x_{p0}, y_p(0) = y_{p0}, \theta_p(0) = \theta_{p0}\) and \(x_e(0) = x_{e0}, y_e(0) = y_{e0}\). \hspace{1cm} (3.14)

In (3.14), \(a_t\) is the control the a pursuer with \(-1 \leq a_t \leq 1\) and \(\theta_e\) is the control of the evader. This model was originally used by Isaacs when he studied the so-called homicidal chauffeur game [11, 7]. The angular velocity \(\omega_p\) and the initial orientation of each pursuer are given in Table 3.1 (rows in shade). It can be verified that every evader in the game is in the “capture region” of some pursuer [11, 7]. That is, Assumption 3 on page 39 holds.

In a two-player PE game with the players’ dynamics given in (3.14), the optimal strategy of evader \(j\) (\(j = 1, \cdots, 5\)) can be shown as same as that in (3.13). Define \(x = x_e - x_p\) and \(y = y_e - y_p\). A heuristic strategy of the pursuer may be solved by considering the distance between the evader and the pursuer, \(D = \sqrt{x^2 + y^2}\). Suppose that the control of the evader is constant. Take the second order time derivative of \(D\), i.e., \(\ddot{D}\), and find \(a^*_t\) to minimize \(\ddot{D}\). It yields the optimal control of the pursuer as

\[a^*_t = -\text{sign}(x \sin \theta_p - y \cos \theta_p).\] \hspace{1cm} (3.15)

<table>
<thead>
<tr>
<th>Engaged Evaders</th>
<th>Pursuer 1</th>
<th>Pursuer 2</th>
<th>Pursuer 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1(^{st})</td>
<td>E1</td>
<td>E3</td>
<td>E5</td>
</tr>
<tr>
<td>2(^{nd})</td>
<td>N/A</td>
<td>E4</td>
<td>E2</td>
</tr>
</tbody>
</table>

Table 3.3: The Optimal Engagement Between the Pursuers and the Evaders with the Modified Dubin’s Car Model
In the hierarchical approach, given a possible engagement scheme, we simulate every decoupled game, in which the pursuers utilize the strategy in (3.15) and evaders play according to (3.13). All possible engagements are enumerated and the best engagement result is given in Table 3.3. In this case, the objective functional evaluated under the optimal engagement scheme is 34.54(sec), which is larger than 26.03(sec) in the previous case. The corresponding PE trajectories are shown in Figure 3.3, in which the trajectory of each pursuer along with that of its engaged evaders is plotted separately. Snapshots at the 1st and the 5th second are illustrated.

In this subsection, the simulation result demonstrates the cooperation of the pursuers by the hierarchical approach. The cooperation results from proper allocation of the pursuers among the evaders. In Figure 3.3, the optimal strategy is utilized by the evaders against their engaged pursuers. It implies that the evaders have the knowledge of the pursuers’ engagement (strategies). Although this may not be true in practice, the pursuers will be better off if the evaders use different strategies. This decision-making is conservative from the pursuers’ perspective.

3.3 Suboptimal Solution by Limited Look-ahead

3.3.1 Performance Enhancement by Limited Look-ahead

The iteration method has poor scalability; however, the appealing aspect of the method is its capability of improving the performance of the pursuers’ controls from the hierarchical approach. In this section, we illustrate the usefulness of limited look-ahead in the performance enhancement in cooperative pursuit problems. Consider a simple PE game involving 2 pursuers and 2 evaders with the dynamics in (3.8) and the objective as the sum of the capture times of all the evaders. Consider a specific
Figure 3.3: Pursuit Trajectories under the Best Engagement
scenario specified in Table 3.4. Let $\varepsilon = 0.5$. In Figure 3.4, a snapshot of the resulting cooperative pursuit trajectory is shown under the optimal engagement between the pursuers and the evaders by the hierarchical approach. Here, the arrows indicate the instantaneous moving directions of the players. Clearly, pursuer 1 (2) is engaged with evader 1 (2).

<table>
<thead>
<tr>
<th></th>
<th>Initial Position</th>
<th>Velocity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pursuer 1</td>
<td>(0,0)</td>
<td>6</td>
</tr>
<tr>
<td>Pursuer 2</td>
<td>(9,0)</td>
<td>6</td>
</tr>
<tr>
<td>Evader 1</td>
<td>(4,0)</td>
<td>3</td>
</tr>
<tr>
<td>Evader 2</td>
<td>(5,2)</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 3.4: The Necessary Parameters of the Players in the PE Game Scenario 2

Figure 3.4: Pursuit Trajectories under the Strategies by the Hierarchical Approach
Define \( \bar{x} = [x_p, y_p, x_e, y_e]^T \) and denote by \( \bar{V}^h(\bar{x}, z) \) as the suboptimal upper Value obtained by the hierarchical approach (see (2.46)), where \( z \) is the discrete variable. Suppose that the game starts at \( t_0 \). Given \( \Delta t > 0 \), the optimization problem based on limited look-ahead at a sample time \( t = t_0 + k\Delta t \) \( (k \in \mathbb{Z}_{\geq 0}) \) is

\[
\max_{\beta \in \Delta(t)} \min_{a(\cdot) \in A(t)} \left\{ \int_{t}^{t+\Delta t} \left( \sum_{j=1}^{2} z_j(\tau) \right) d\tau + \bar{V}^h(\bar{x}_{t+\Delta t}, \bar{x}_t, a, \beta[\cdot], z_{t+\Delta t}, z_t) \right\}.
\]

(3.16)

In this example, we choose \( \Delta t = 0.05 \), and approximate the “minimax” problem in (3.16) by a static optimization problem. The optimal strategies \( (a_t^*, \beta_t^*) \) solved at each sample time \( t_0 + k\Delta t \) are utilized to generate the trajectories of the pursuers and the evaders, which are illustrated in Figure 3.5. Clearly, the evolution of the game in Figure 3.5 can better resemble the reality compared to Figure 3.4. According to Figure 3.5, when the pursuers and the evader get closer, both pursuers can move cooperatively to force the evaders to change their escape directions. In this way, the pursuit performance is improved by 4 time steps and here we use the term “cooperation zone” to indicate the region where the extra cooperation between the two pursuers is possible.

### 3.3.2 Limited Look-ahead with a Heuristic Cost-to-go

In this subsection, we demonstrate a potential advantage of limited look-ahead in a problem when no obvious suboptimal solution is available. Consider a PE game with 3 pursuers and 1 evader with the dynamics specified in (3.8). The evader has higher speed than any of the pursuers. Choose the capture time of the evader \( T \) as the objective functional. Consider a specific scenario specified in Table 3.5. Note that since \( v_e > v_p^i \) for any \( i = 1, 2, 3 \), the hierarchical approach is no longer valid.
Figure 3.5: Pursuit Trajectories under the Strategies Based on One-Step Limited Look-ahead
In the following, we apply the optimization based on one-step limited look-ahead with a heuristic cost-to-go function. By inspection of the optimization based on limited look-ahead in (2.20), the performance of the resulting pursuers’ strategies depends largely on the cost-to-go function evaluated at time $t + \Delta t$. Note that the hierarchical approach is no longer applicable, and we construct a heuristic cost-to-go function as follows. We first analyze a necessary condition, under which the evader can be captured. An arbitrary pursuer $i$ ($i = 1, 2, 3$) and the evader are illustrated in Figure 3.6, where $v_p^i$ and $v_e$ are the speed vector of the pursuer and that of the evader respectively; $\beta_i$ ($\alpha_i$) is the angle between the line of sight from pursuer $i$ (the evader) to the evader (pursuer $i$) and the moving direction of the evader (pursuer $i$). Now, let $v_p^i$ and $v_e$ be fixed, i.e., $\alpha_i, \beta_i$ are constant. If the evader can be captured, $\alpha_i$ and $\beta_i$ must satisfies the following condition.

$$\sin \beta_i = \frac{|v_p^i|}{|v_e|} \sin \alpha_i \leq \frac{|v_p^i|}{|v_e|} \Rightarrow \bar{\beta}_i = \arcsin\left(\frac{|v_p^i|}{|v_e|}\right)$$

(3.17)

In (3.17), $|v_p^i|$ and $|v_e|$ are the norms of the speed vectors; $\bar{\beta}_i$ is a bound of $\beta_i$ when the capture is guaranteed, i.e., $|\beta_i| \leq \bar{\beta}_i$.

<table>
<thead>
<tr>
<th>Initial Position</th>
<th>Velocity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pursuer 1</td>
<td>(0,3)</td>
</tr>
<tr>
<td>Pursuer 2</td>
<td>(1.64, -1.15)</td>
</tr>
<tr>
<td>Pursuer 3</td>
<td>(-2.8,-1)</td>
</tr>
<tr>
<td>Evader</td>
<td>(0,0)</td>
</tr>
</tbody>
</table>

Table 3.5: The Necessary Parameters of the Players in the PE Game Scenario 3
Claim 1 If the evader is captured by a finite time $T > 0$, then at any time $t \geq 0$ and for any possible $v_e$ of the evader, there must exist a pursuer $i$ ($i = 1, 2, 3$), such that (3.17) holds.

It is clear that when there exists $v_e$ such that $|\beta_i|$ satisfies $|\beta_i| > \bar{\beta}_i$ for $i = 1, 2, 3$, the evader can escape from all the pursuers by moving along that direction.

Denote by $\bar{x}$ the composite state variable that includes the states of all the pursuers and the evader. Suppose that the pursuers can move cooperatively such that (3.17) holds at any time $t \geq 0$. We construct the heuristic cost-to-go function as follows. Define the set $\Pi = \{\bar{x} | (3.17) \text{ holds at } \bar{x}\}$. Consider the terms $\alpha_i$, $\beta_i$ and $\gamma_i$ ($\gamma_i = \pi - \alpha_i - \beta_i$) as shown in Figure 3.6, where $d_i$ is the distance between pursuer $i$ and the evader. All of these terms are functions of state $\bar{x}$. Define the function $\psi(\bar{x})$ as

$$\psi(\bar{x}) \triangleq \max_{\theta_e} \left\{ \min_i \left\{ \frac{d_i \sin \alpha_i}{|v_e| \sin \gamma_i} \right\} \right\}.$$  

Here, the maximization is taken with respect to the evader’s control $\theta_e$ in (3.8), which is the speed orientation. Let $D(\bar{x}) = \sum_i d_i$. Choose a heuristic cost-to-go function
\( \tilde{V}(\bar{x}) \) \((x \in \Pi)\) with the following properties.

\[
\begin{aligned}
\tilde{V}(\bar{x}_1) &< \tilde{V}(\bar{x}_2) \quad \text{if} \ \psi(\bar{x}_1) < \psi(\bar{x}_2), \\
\tilde{V}(\bar{x}_1) &< \tilde{V}(\bar{x}_2) \quad \text{if} \ \psi(\bar{x}_1) = \psi(\bar{x}_2) \ \text{and} \ D(\bar{x}_1) < D(\bar{x}_2).
\end{aligned}
\]

(3.18)

Choose \( \Delta t = 0.01 \) and \( \varepsilon = 0.5 \). Implement the look-ahead strategies of the pursuers and the evader repetitively determined from (3.16) with the cost-to-go function replaced by \( \tilde{V} \) that satisfies (3.18). The resulting cooperative pursuit trajectories are shown in Figure 3.7. It can be seen that with the guidance from \( \tilde{V} \), a proper formation of the pursuers is maintained and capture is ensured. This small example demonstrates a possible usefulness of the look-ahead approach in solving this complicated PE problem. A complete solution to this game needs further investigation.
Figure 3.7: Cooperative Pursuit Trajectories of the Superior Evader by the Limited Look-ahead Method Based on a Heuristic Cost-to-go
CHAPTER 4

STOCHASTIC MULTI-PLAYER PURSUIT-EVASION DIFFERENTIAL GAMES

In this chapter, we extend the iterative method for deterministic games to stochastic PE problems. For the case where the players’ (pursuers/evaders) dynamics are governed by random processes and the state variables are perfectly accessible all the players, the problem is rigorously defined with the concepts of the players’ strategies and Value function. We show that the iterative method is still applicable. Under the framework of the hierarchical method, we devote some effort to two-player stochastic PE games. An analytic solution for a specific stochastic two-player PE differential game is derived. Capturability of the evader in a two-player game is also examined. Finally, we extend the study to a more general case where the players no longer have perfect state information. It is a very hard problem and only a special case is addressed.

4.1 Formulation of Stochastic Pursuit-Evasion Games with Perfect State Information

We first formulate a stochastic multi-player PE differential game, where both the pursuers and the evaders can access the perfect state information at any time, based
on the deterministic case in Chapter 2. Suppose that there are $N$ pursuers and $M$ evaders in the same state space. Consider a noise corrupted version of the dynamics of the pursuers and the evaders with additive disturbances. The disturbance is modelled as a standard Wiener process (often called Brownian motion). The dynamic equations of pursuer $i$ and evader $j$ are given as follows.

\[
\begin{align*}
\frac{dx^i_{pt}}{dt} &= f_p^i(x^i_{pt}; a_{it}) dt + \sigma_p^i(x^i_{pt}) dw^i_{pt} \quad \text{with} \quad x^i_{pt}(t_0) = x^i_{p0}, \\
\frac{dx^j_{et}}{dt} &= f_e^j(x^j_{et}; b_{jt}) dt + \sigma_e^j(x^j_{et}) dw^j_{et} \quad \text{with} \quad x^j_{et}(t_0) = x^j_{e0}.
\end{align*}
\]  

In (4.1), $x^i_{pt} \in \mathbb{R}^{n_p^i}$, $x^j_{et} \in \mathbb{R}^{n_e^j}$ for $t \geq 0$, $n_p^i, n_e^j \in \mathbb{Z}^+$ and $a_{it} \in A^i_a$, $b_{jt} \in B^j_b$, where $A^i_a \subseteq \mathbb{R}^{n_p^i}$ and $B^j_b \subseteq \mathbb{R}^{n_e^j}$ are given compact sets with $m_p^i, m_e^j \in \mathbb{Z}^+$; $w^i_{pt} \subseteq \mathbb{R}^{k_p^i}$ ($w^j_{et} \subseteq \mathbb{R}^{k_e^j}$) is a standard Wiener process; $\sigma_p^i(\cdot)$ ($\sigma_e^j(\cdot)$) is the corresponding matrix with a proper dimension. Although this model is adopted for mathematical reasons, in practice, the additive disturbance may stand for a simple wind effect on small AAVs (Autonomous Aerial Vehicles).

Suppose that $f_p^i (f_e^j)$ is bounded and Lipschitz in $x^i_p (x^j_e)$, uniformly with respect to $a_{it}$ ($b_{jt}$) as in (2.1). Assume that $\sigma_p^i$ and $\sigma_e^j$ are bounded and satisfy

\[
\|\sigma_p^i(x^i_p) - \sigma_p^i(\tilde{x}^i_p)\| \leq \gamma^i_p\|x^i_p - \tilde{x}^i_p\|, \quad \|\sigma_e^j(x^j_e) - \sigma_e^j(\tilde{x}^j_e)\| \leq \gamma^j_e\|x^j_e - \tilde{x}^j_e\|
\]

for some $\gamma^i_p, \gamma^j_e > 0$ and any $x^i_p, \tilde{x}^i_p \in \mathbb{R}^{n_p^i}, x^j_e, \tilde{x}^j_e \in \mathbb{R}^{n_e^j}$ [50]. We use the same definition of the capture of evader $j$ in deterministic problems, and also augment the state $x$ with the discrete variable $z$. For simplicity, let

\[
\begin{align*}
x_p &= \left[ x^T_p, x^T_p, \cdots, x^T_p \right]^T, \quad x_e = \left[ x^T_e, x^T_e, \cdots, x^T_e \right]^T, \\
a &= \left[ a_1^T, a_2^T, \cdots, a_N^T \right]^T, \quad b = \left[ b_1^T, b_2^T, \cdots, b_M^T \right]^T, \\
f_p &= \left[ f^T_p, f^T_p, \cdots, f^T_p \right]^T, \quad f_e = \left[ z_1^T f_p^T, z_2^T f_p^T, \cdots, z_M^T f_p^T \right]^T, \\
w_p &= \left[ w^T_p, w^T_p, \cdots, w^T_p \right]^T, \quad w_e = \left[ w^T_e, w^T_e, \cdots, w^T_e \right]^T,
\end{align*}
\]
and define

\[ \sigma_p(x_p) = \text{diag}(\sigma_1^p(x_p^1), \ldots, \sigma_N^p(x_p^N)) \text{ and } \sigma_e(x_e) = \text{diag}(z_1\sigma_1^e(x_e^1), \ldots, z_M\sigma_M^e(x_e^M)) \]

as diagonal matrices with \( \sigma_i^p \)'s and \( z_j\sigma_i^e \)'s as the diagonal elements. Then, equation (4.1) can be rewritten as

\[
\begin{align*}
\dot{x}_p(t) &= f_p(x_p, a_t)dt + \sigma_p(x_p)dw_p(t) \quad \text{with } x_p(0) = x_{p0}, \\
\dot{x}_e(t) &= f_e(x_e, b_t)dt + \sigma_e(x_e)dw_e(t) \quad \text{with } x_e(0) = x_{e0}.
\end{align*}
\] 

Further,

\[
\dot{x}(t) = f(x, a_t, b_t)dt + \sigma(x)dw(t) \quad \text{with } x(0) = x_0,
\]

where \( x = [x_p^T, x_e^T]^T \), \( f(x, a, b) = [f_p^T(x, a), f_e^T(x, b)]^T \), \( \sigma(x) = \text{diag}(\sigma_p(x_p), \sigma_e(x_e)) \) and \( w = [w_p^T, w_e^T]^T \). Let \( X \triangleq \left( \prod_{i=1}^N \mathbb{R}^{n_p^i} \right) \times \left( \prod_{j=1}^M \mathbb{R}^{n_e^j} \right) \), and denote by \( A_a \) the set of possible controls of all pursuers, i.e., \( A_a = \prod_{i=1}^N A_a^i \) and similarly for evaders, \( B_a = \prod_{j=1}^M B_a^j \).

Due to the introduction of the random process, the strategies of the pursuers and the evaders should be defined properly such that (4.2) has a unique solution. Here, we adopt the definitions in [50].

**Definition 5** Let \((\Omega, \mathcal{F}, P)\) be a probability space, \(\{\mathcal{F}_t\}\) a monotone family of \(\sigma\)-fields \(\mathcal{F}_t \subseteq \mathcal{F}\) with \(\mathcal{F}_{t_1} \subseteq \mathcal{F}_{t_2}\) for any \(t_0 \leq t_1 \leq t_2\), and \(\mathcal{X}\) a complete separable metric space. A stochastic process \(\xi\) defined on the time interval \([t_0, T]\), \(\xi(t): \Omega \to \mathcal{X}\) with \(t_0 \leq t \leq T\), is said to be \(\{\mathcal{F}_t\}_{t \geq t_0}\)-progressively measurable if for all \(t \in [t_0, T]\), the map \((t, \omega) \mapsto \xi(t, \omega)\) is \(\mathcal{B}[t_0, t] \times \mathcal{F}_t / \mathcal{B}(\mathcal{X})\) measurable, where \(\mathcal{B}[t_0, t]\) is the Borel \(\sigma\)-field on \([t_0, t]\) and \(\mathcal{B}(\mathcal{X})\) is the Borel \(\sigma\)-field on \(\mathcal{X}\).
Define
\[ K = \sum_{i=1}^{N} k_p^i + \sum_{j=1}^{M} k_e^j. \]
The sample space for the stochastic process of (4.3) is defined as
\[ \Omega^\omega_t = \left\{ \omega \in C([t, T]), \omega : [t, T] \mapsto \mathbb{R}^K \text{ with } \omega_t = 0 \right\} \tag{4.4} \]
for every \( t \in [t_0, T] \). Denote by \( \mathcal{F}_{t,s}^\omega \subseteq \mathcal{F} \) the \( \sigma \)-field generated by paths from time \( t \) to \( s \) (for any \( t \leq s \leq T \)) in \( \Omega^\omega_t \). Provided with the Wiener measure \( P^\omega_t \) on \( \mathcal{F}_{t,T}^\omega \), \( \Omega^\omega_t \) becomes the canonical sample space for the stochastic process of (4.3). Note that \( T \) in the definitions above is related to the terminal time of a game defined in Chapter 2. In addition, we also use the space
\[ \Omega^\omega_{t,s} = \left\{ \omega \in C([t, s]), \omega : [t, s] \mapsto \mathbb{R}^K \text{ with } \omega_t = 0 \right\} \tag{4.5} \]
for \( t \leq s \leq T \). Next, we define dismissible controls and nonanticipative strategies of the pursuers and the evaders.

**Definition 6** In a stochastic multi-player PE game at any time \( t \geq 0 \), an admissible control of the pursuers and that of the evaders, \( a(\cdot) \) and \( b(\cdot) \) are defined respectively as
\[ a(\cdot) \in A(t) \triangleq \left\{ \phi : [t, T] \mapsto A_a \mid \phi(\cdot) \text{ is } \mathcal{F}_{t,s}^\omega \text{-progressively measurable} \right\}, \]
\[ b(\cdot) \in B(t) \triangleq \left\{ \varphi : [t, T] \mapsto B_a \mid \varphi(\cdot) \text{ is } \mathcal{F}_{t,s}^\omega \text{-progressively measurable} \right\} \]
for \( t \leq s \leq T \).

We say that \( a(\cdot), \tilde{a}(\cdot) \in A(t) \) are the same on interval \([t, s]\), denoted by \( a \approx \tilde{a} \) on \([t, s]\), if the probability \( P^\omega_t(a = \tilde{a} \text{ a.e. in } [t, s]) = 1 \). It is the same for \( b(\cdot), \tilde{b}(\cdot) \in B(t) \). A nonanticipative strategy of the pursuers (evaders) can be defined as follows.
Definition 7 A nonanticipative strategy $\alpha$ (resp. $\beta$) of the pursuers (resp. evaders) on $[t, T]$ is

$$\alpha \in \Gamma(t) \triangleq \{ \psi : B(t) \mapsto A(t) \mid \text{for any } b(\cdot), \tilde{b}(\cdot) \in B(t), b \approx \tilde{b} \text{ on } [t, s] \text{ implies } \alpha[b] \approx \alpha[\tilde{b}] \text{ on } [t, s] \text{ for every } s \in [t, T] \}$$

(resp. $\beta \in \Delta(t) \triangleq \{ \eta : A(t) \mapsto B(t) \mid \text{for any } a(\cdot), \tilde{a}(\cdot) \in A(t), a \approx \tilde{a} \text{ on } [t, s] \text{ implies } \beta[a] \approx \beta[\tilde{a}] \text{ on } [t, s] \text{ for every } s \in [t, T] \}$$.

As in the deterministic case, the definition of nonanticipative strategy implies that the decision-making of the players only depend on the information up to the current time. More importantly, under the definition of admissible control with that of nonanticipative strategy, the stochastic differential equation (4.3) has a unique solution [50]. More details about the setting-up in a stochastic problem can be found in [78, 19, 7].

Consider the following objective functional,

$$J(a, b; x_t, z_t) = \mathbb{E}_{\omega; t}\left\{ \int_t^T G(x_\tau, z_\tau, a_\tau, b_\tau) d\tau + Q(x_T) \right\} \tag{4.6}$$

subject to (4.3) and (2.9).

In (4.6), $\omega \in \Omega^{x_t, z_t}$; the terminal cost $Q : X \mapsto \mathbb{R}_{\geq 0}$ and the cost rate $G : X \times Z^M \times A_a \times B_a \mapsto \mathbb{R}_{\geq 0}$ satisfy the boundedness and the Lipschitz conditions that are similar to (2.15) [50]. Here, $\mathbb{E}_{\omega; t}$ denotes the expectation taken with respect to stochastic process $\omega$ given the information at time $t$. Henceforth, we will use $\mathbb{E}_\omega$ for a short notation.

A PE stochastic game is modelled as a zero-sum game. The lower and the upper Value can be defined similarly as in the deterministic case. For any $x_t \in X$ and
\[ z_t \in Z^M, \text{ the lower Value of a game } \mathcal{V}(x_t, z_t) \text{ is defined as} \]
\[ \mathcal{V}(x_t, z_t) = \inf_{a \in \Gamma(t)} \sup_{b(\cdot) \in B(t)} \left\{ \mathcal{J}(a[b], b; x_t, z_t) \right\} \]
\[ = \inf_{a \in \Gamma(t)} \sup_{b(\cdot) \in B(t)} \mathbb{E}_\omega \left\{ \int_t^T G(x_\tau, z_\tau, a[b]_\tau, b_\tau) d\tau + Q(x_T) \right\}. \quad (4.7) \]

Similarly, the upper Value \( \mathcal{V}(x_t, z_t) \) is
\[ \mathcal{V}(x_t, z_t) = \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} \left\{ \mathcal{J}(a, \beta[a]; x_t, z_t) \right\} \]
\[ = \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} \mathbb{E}_\omega \left\{ \int_t^T G(x_\tau, z_\tau, a[\beta]_\tau) d\tau + Q(x_T) \right\}. \quad (4.8) \]

To differentiate the objective and the (upper/lower) Value from those in the deterministic case, we use the notations \( \mathcal{J}, \mathcal{V} \) and \( \mathcal{V} \) in (4.7) and (4.8). The Value function \( \mathcal{V} \) exists if the Isaacs condition holds, i.e., \( \mathcal{V} = \mathcal{V} = \mathcal{V} \).

### 4.2 The Iterative Approach to Stochastic Pursuit-Evasion Games

A stochastic multi-player PE differential game is very difficult due to the uncertainty. However, for stochastic problems with the perfect state information pattern, the iterative approach proposed for deterministic PE games is still applicable. In the following, we summarize the iterative approach in the stochastic case, and again, we focus on the upper Value \( \mathcal{V} \).

First of all, we define the Sample-Set-Time-Consistency (SSTC) of a control structure imposed on the pursuers in the stochastic case in contrast to STC in deterministic problems. Given any \( \tilde{a}^t(\cdot) \in \mathcal{A}^S_{x_t, z_t}(t), x_t \in X, z_t \in Z^M, \beta \in \Delta(t) \) and \( \omega \in \Omega^\omega_{t,T} \), denote by \( x_{s; x_t, \tilde{a}^t, \beta[\tilde{a}^t], \omega} \) the trajectory of \( x \) for \( s \geq t \) starting from \( x_t \) under \( \tilde{a}^t \) and \( \beta[\tilde{a}^t] \) corresponding to \( \omega \). For short, let \( \hat{x}(s) = x_{s; x_t, \tilde{a}^t, \beta[\tilde{a}^t], \omega} \). Denote by \( z_{s; z_t, \tilde{x}} \) the trajectory of \( z \) corresponding to \( \tilde{x} \) and use \( \hat{z}(s) = z_{s; z_t, \tilde{x}} \) for a short notation.
Definition 8 (Sample-Set-Time-Consistency) A suboptimal control structure \( S \) is said to be \textbf{Sample-Set-Time-consistent}, or SST-consistent for short, if given any \( \omega \in \Omega_{t,T}^\omega \), for any \( \tilde{a}^i(\cdot) \in A_{\tilde{x}_i,\tilde{z}_i}^S(t) \) at any \( 0 \leq t \leq T \), there exists \( \tilde{a}^s(\cdot) \in A_{\tilde{x}_s,\tilde{z}_s}^S(s) \) associated with \( \tilde{x}_s \) and \( \tilde{z}_s \) at any \( s \) for \( t \leq s \leq T \) such that \( \tilde{a}^i(\tau) = \tilde{a}^s(\tau) \) for \( s \leq \tau \leq T \) under any \( \beta \in \Delta(t) \).

Suppose that there exists a feasible SST-consistent control structure such that a suboptimal upper \( V \)-value function can be determined as

\[
\tilde{V}(x,z) = \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A_{x,z}^S(t)} E_{\omega} \left\{ \int_t^T G(x_\tau, z_\tau, a_\tau, \beta[a_\tau]) d\tau + Q(x_T) \right\}, \quad (4.9)
\]

where \( \omega \in \Omega_{t,T}^\omega \) and \( A_{x,z}^S(t) \) is the restricted control set under structure \( S \) at time \( t \). If the control structure \( S \) is SST-consistent, it can be shown similarly as in Theorem 1 in Chapter 2 that

\[
\tilde{V}(x,z) \geq \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A_{x,z}^S(t)} E_{\omega} \left\{ \int_t^{t+\Delta t} G(x_\tau, z_\tau, a_\tau, \beta[a_\tau]) d\tau + \tilde{V}(x_{t+\Delta t}, x, a, \beta[a], \omega, z_{t+\Delta t}, x) \right\}. \quad (4.10)
\]

Let \( \mathcal{W} = \{ W : X \times Z^M \mapsto \mathbb{R} \} \). A transformation \( \overline{\mathcal{H}} \) on \( \mathcal{W} \) can be defined similarly as in (2.25)

\[
\overline{\mathcal{H}}[W](x,z) = \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A_{x,z}^S(t)} E_{\omega} \left\{ \int_t^{t+\Delta t} G(x_\tau, z_\tau, a_\tau, \beta[a_\tau]) d\tau + W(x_{t+\Delta t}, x, a, \beta[a], \omega, z_{t+\Delta t}, x) \right\}. \quad (4.11)
\]

subject to (4.3) and (2.9) for \( W \in \mathcal{W} \). Let \( W_0 = \tilde{V} \). A sequence of functions \( W_k \) can be generated by \( W_{k+1} = \overline{\mathcal{H}}[W_k] \). In the following, we will show that the sequence is decreasing, and its limit coincides with the upper \( V \)-value defined in (4.8).
Theorem 10  (i) If \( W_0(x,z) \) satisfies (4.10) for any \( x \in X \) and \( z \in Z^M \), then the sequence \( \{W_k\}_{k=0}^\infty \) converges point-wisely; (ii) The limiting function \( W_\infty(x,z) \triangleq \lim_{k \to \infty} W_k(x,z) \) satisfies \( W_\infty(x,z) = \overline{F}[W_\infty] \).

Proof: The proof of Theorem 2 in Chapter 2 is applicable.

Claim 2 (Principle of DP) The upper Value function \( \overline{V} \) satisfies

\[
\overline{V}(x,z) = \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} E_\omega \left\{ \int_t^{t_\land T} G(x_\tau, z_\tau, a_\tau, \beta[a_\tau]) \, d\tau + \overline{V}(x_{t_\land T}; x, a, \beta[a], \omega, z_{t_\land T}; z, x) \right\}
\]

(4.12)

for any \( t_\land \geq t_0 \), \( x \in X \) and \( z \in Z^M \).

Remark 10 Related work on DP principle includes Theorem 1.6 on page 298 in [50], Theorem 4.1 on page 777 in [69] and Theorem 3.3 on page 180 in [19].

Lemma 7 For a real-valued function \( W(x,z) \), if there exists \( \Delta t > 0 \), such that for any \( x \in X \) and \( z \in Z^M \),

\[
W(x,z) = \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} E_\omega \left\{ \int_t^{t+\Delta t} G(x_\tau, z_\tau, a_\tau, \beta[a_\tau]) \, d\tau + W(x_{t+\Delta t}; x, a, \beta[a], \omega, z_{t+\Delta t}; z, x) \right\},
\]

(4.13)

then for any \( K \in \mathbb{Z}^+ \)

\[
W(x,z) = \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} E_\omega \left\{ \int_t^{t+K\Delta t} G(x_\tau, z_\tau, a_\tau, \beta[a_\tau]) \, d\tau + W(x_{t+K\Delta t}; x, a, \beta[a], \omega, z_{t+K\Delta t}; z, x) \right\}
\]

(4.14)

for any \( x \in X \) and \( z \in Z^M \).

Proof: The proof of Lemma 3 in Chapter 2 is applicable.
Assumption 4 Define the set $S_G = \{ x \in X, z \in Z^M | z^x \neq 0 \}$\textsuperscript{22}. There exists a constant $C_E > 0$, and for any $(x, z) \in S_G$, any $\varepsilon > 0$ and $K \in \mathbb{Z}^+$, there exists $a(\cdot) \in A(t)$ such that (i)

$$V(x, z) + \varepsilon \geq \sup_{\beta \in \Delta(t)} E_\omega \left\{ \int_t^{t+K\Delta t} G(x_\tau, z_\tau, a_\tau, \beta[a_\tau])d\tau \right. $$

$$+ \left. V(x_{t+K\Delta t}, z_{t+K\Delta t}; z, x) \right\}; \quad (4.15)$$

(ii) for any $t \leq \tau \leq t + K\Delta t$, the conditional expectation

$$E_x \left\{ \tilde{V}(x_{t+K\Delta t}, a, \beta[a], \omega, z_{t+K\Delta t}; x, z) \right\} \left( x_{t+K\Delta t}, a, \beta[a], \omega, z_{t+K\Delta t}; x, z \right) \in S_G \right\} \leq C_E \text{\textsuperscript{23}} \quad (4.16)$$

for any $\beta \in \Delta(t)$. 

Theorem 11 Suppose that $W_0 = \tilde{V}$. If Assumption 6 is true, then for the sequence $\{ W_k(x, z) \}_{k=1}^{\infty}$ by $W_{k+1} = \mathcal{H}[W_k]$, its limit $W_\infty(x, z) = \lim_{k \to \infty} W_k(x, z) = V(x, z)$ for any $x \in X, z \in Z^M$.

Proof: Since $W_\infty(x, z) \geq V(x, z)$, we only need to show that $W_\infty(x, z) \leq V(x, z)$. For any $(x, z) \in S_G$ an arbitrary $\varepsilon > 0$, there exists $K \in \mathbb{Z}^+$ and $a(\cdot) \in A(t)$ such that

$$P\left( x_{t+K\Delta t}, z_{t+K\Delta t}; x, z \in S_G \right) < \varepsilon/C_E$$

due to the fact that $C_G \geq G(x, z, a, b) \geq \delta > 0$ for $z \neq 0$ in (2.15).

\textsuperscript{22}Set $S_G$ is interpreted as all the state $x$ and $z$ from where the game has not yet terminated.

\textsuperscript{23}It stands for the expectation of the suboptimal upper Value function at $x_{t+K\Delta t}, a, \beta[a], \omega, z_{t+K\Delta t}; x$ provided that $x_{t+K\Delta t}, a, \beta[a], \omega, z_{t+K\Delta t}; x$ is in $S_G$ at any $\tau$ for $t \leq \tau \leq t + K\Delta t$. 

84
Since (4.17) holds. For simplicity, let $\bar{x} = x_{t+K\Delta t}^\omega$ and $\bar{z} = z_{t+K\Delta t}^\omega$ for $\beta \in \Delta(t)$.

$\bar{V}(x, z) \geq \sup_{\beta \in \Delta(t)} \left\{ E_\omega \left[ \int_t^{t+K\Delta t} G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) d\tau \right] + E_\omega \bar{V}(\bar{x}, \bar{z}) \right\} - \varepsilon$

$\geq \left\{ E_\omega \left[ \int_t^{t+K\Delta t} G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) d\tau \right] + P\left( (\bar{x}, \bar{z}) \notin S_G \right) E_\omega \left\{ Q(\bar{x}) \big| (\bar{x}, \bar{z}) \notin S_G \right\}

\right\} - \varepsilon \quad (4.17)$

for any $\beta \in \Delta(t)$. Note that $\mathcal{W}_\infty = \mathcal{H}[\mathcal{W}_\infty]$ and by Lemma 7,

$\mathcal{W}_\infty(x, z) = \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} E_\omega \left\{ \int_t^{t+K\Delta t} \mathcal{W}_\infty(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) d\tau + \mathcal{W}_\infty(\bar{x}, \bar{z}) \right\}$.

Now, consider the same $a(\cdot)$ used in (4.17),

$\mathcal{W}_\infty(x, z) \leq \sup_{\beta \in \Delta(t)} E_\omega \left\{ \int_t^{t+K\Delta t} G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) d\tau + \mathcal{W}_\infty(\bar{x}, \bar{z}) \right\}$

$= \sup_{\beta \in \Delta(t)} \left\{ E_\omega \left[ \int_t^{t+K\Delta t} G(x_\tau, z_\tau, a_\tau, \beta[a]_\tau) d\tau \right] + P\left( (\bar{x}, \bar{z}) \notin S_G \right) \cdot E_\omega \left\{ Q(\bar{x}) \big| (\bar{x}, \bar{z}) \notin S_G \right\}

\right\} - \varepsilon \quad (4.18)$

Subtract (4.17) from (4.18),

$\mathcal{W}_\infty(x, z) - \bar{V}(x, z) \leq \sup_{\beta \in \Delta(t)} \left\{ P\left( (\bar{x}, \bar{z}) \in S_G \right) \cdot E_\omega \left\{ \mathcal{W}_\infty(\bar{x}, \bar{z}) - \bar{V}(\bar{x}, \bar{z}) \big| (\bar{x}, \bar{z}) \in S_G \right\} + \varepsilon \right\}$.

Since $P\left( (\bar{x}, \bar{z}) \in S_G \right) < \varepsilon/C_E$ and $\mathcal{W}_\infty(\bar{x}, \bar{z}) \leq \mathcal{W}_0 = \bar{V}$,

$\mathcal{W}_\infty(x, z) - \bar{V}(x, z) \leq \varepsilon/C_E \cdot E_\omega \left\{ \bar{V}(\bar{x}, \bar{z}) \big| (\bar{x}, \bar{z}) \in S_G \right\} + \varepsilon \leq 2\varepsilon \quad (4.19)$

The second inequality in (4.19) is because of Assumption 6 (ii). Since $\varepsilon$ is arbitrary, $\mathcal{W}_\infty(x, z) \leq \bar{V}(x, z)$. The proof is complete.
Finally, it is worth mentioning that the hierarchical approach is also applicable to stochastic games and it is SST-consistent because the hierarchy is independent of time $t$ and the states $x$ and $z$. Note that the hierarchical approach depends on the solvability of the distributed two-player PE games.

4.3 Solution to a Two-Player Pursuit-Evasion Game

Since the hierarchical approach relies on the solution of a two-player game, in the section, we focus on a two-player stochastic PE differential game. The results presented are based on the classical stochastic differential game (optimal control) theory.

4.3.1 Classical Theory of Stochastic Zero-sum Differential Games

We first introduce the necessary concepts and the classical theory of solving a stochastic differential game problem with perfect state information based on [78].

Consider a two-player stochastic differential game, where the dynamic equation of the players is

$$
\text{d}x_t = f(t, x_t, a_t, b_t)\text{d}t + \sigma(x_t)\text{d}\omega.
$$

In (4.20), $x \in \mathbb{R}^n$; $a$ and $b$ are the controls of player 1 (minimizer) and player 2 (maximizer) in compact sets $A_a$ and $B_b$; $\omega$ is a standard Wiener process with a proper dimension. Note that here the game model is time variant, which is more general than the model we use for multi-player PE games. Let the objective functional be

$$
\hat{J}(a, b; t, x_t) = \mathbb{E}_\omega \left\{ \int_t^T \mathcal{G}(t, x_\tau, a_\tau, b_\tau)\text{d}\tau + Q(x_T) \right\}.
$$
Here, the cost rate $\mathcal{G}$ is a map $\mathcal{G} : \mathbb{R} \times \mathbb{R}^n \times A_a \times B_a \mapsto \mathbb{R}$; terminal cost is $Q : \mathbb{R}^n \mapsto \mathbb{R}$; $T$ is the exit time, which is defined as the first time that $t$ or $x$ leaves some set $Q$ that will be defined shortly.

Suppose that $X$ is an open set in the state space. Denote by $C^2(X)$ a set of continuously differentiable functions on $X$. Let $\mathcal{T} = [0, T] \subseteq \mathbb{R}$, and define $Q = \mathcal{T} \times X$. Denote by $C^{1,2}(Q)$ a set of functions $\phi(t, x)$ that have continuous first- and second-order partial derivatives, $\phi_t$, $\phi_x$ and $\phi_{xx}$, where $t \in \mathcal{T}$ and $x \in X$. For short, we use $C^2$ and $C^{1,2}$ to denote the corresponding sets in the following discussion. Define the set $\Psi$ as

$$\Psi(X) \triangleq \{ \psi : X \mapsto \mathbb{R} | \psi \text{ is Borel measurable on } X \text{ and bounded} \}.$$  

Denote by $P(s, x_s, t, x_t)$ ($s < t$) the transition probability density of $x_t \in X$ at $t$ given the state $x_s$ at time $s$. For any $\psi \in \Psi(X)$, define operator $S_{s,t}$ as

$$S_{s,t}[\psi](x_s) = \mathbb{E}\{\psi(x_t)|x_s\} = \int_X \psi(x)P(s, x_s, t, x)d\mu.$$

Define the operator $\Xi(t)$ on $\Psi(X)$ as

$$\Xi(t)[\psi](x_t) = \lim_{h \to 0} h^{-1}(S_{t,t+h}[\psi](x_t) - \psi(x_t)).$$

Suppose that for any $\varphi \in \Psi(X)$, $\Xi(t)$ above is well defined. If $x$, is a stochastic process defined in (4.3), which is also called a diffusion process, $\Xi(t)$ is a second-order partial differential operator on $C^2(X)$, i.e.,

$$\Xi(t)[\psi] = \psi_x \cdot f(t, x_t, a, b) + \frac{1}{2} \text{tr} \left( \psi_{xx} \sigma(x_t) \sigma^T(x_t) \right). \quad (4.21)$$
Here, \( \text{tr}(M) \) stands for the trace of a square matrix \( M \). Consider any function \( \phi(t, x) \in C^{1,2} \), and by Ito’s stochastic differential rule [78, 19],

\[
\begin{align*}
\mathrm{d}\phi(t, x) &= \phi_t(t, x)\mathrm{d}t + \phi_x(t, x)\mathrm{d}x + \frac{1}{2} \text{tr}\left( \psi_{xx}(x)\sigma^T(x) \right)\mathrm{d}t \\
&= \phi_t(t, x)\mathrm{d}t + \Xi(t)\phi(t, x)\mathrm{d}t + \phi_x\sigma(x)\mathrm{d}t.
\end{align*}
\]

(4.22)

Integrate (4.22) on both sides and take the expectation,

\[
E\phi(t, x_t) - \phi(s, x_s) = E\left\{ \int_s^t \left( \phi_\tau(\tau, x_\tau) + \Xi(\tau)\phi(\tau, x_\tau) \right)\mathrm{d}\tau \right\}
\]

(4.23)

Define the exit time \( T \) as \( T = \min\{t | (t, x_t) \notin Q\} \). Suppose that the set \( Q \) is open, and the upper limit of the integral in (4.23) can be the exit time.

**Lemma 8** Assume that

(i) functions \( f \) and \( \sigma \) in (4.3) satisfy that

\[
\|f(t, a, b)\| \leq C(1 + \|x\|) \quad \text{and} \quad \|\sigma(x)\| \leq C(1 + \|x\|)
\]

for some constant \( C > 0 \), any \( x \in X, a \in A_a \) and \( b \in B_a \);

(ii) Let function \( V \in C^{1,2} \); there exist constants \( D \) and \( k \), such that

\[
\|V(t, x)\| \leq D(1 + \|x\|^k) \quad \text{for any} \ (t, x) \in Q;
\]

(iii) \( V \) is continuous on \( \bar{Q} \), the closure of \( Q \);

(iv) \( V_\tau + \Xi(\tau)V + G(\tau, x) \geq 0 \) for all \((t, x) \in Q\), where \( E\left\{ \int_s^T |G(\tau, x_\tau)|\mathrm{d}\tau \right\} < \infty \) for any \((s, x_s) \in Q\) and some function \( G : \mathbb{R} \times \mathbb{R}^n \mapsto \mathbb{R} \). Then,

\[
V(s, x_s) \leq E\left\{ \int_s^T G(\tau, x_\tau)\mathrm{d}\tau + V(T, x_T) \right\}.
\]

**Proof:** Refer to Theorem 5.1 on page 124 in [78].
We adopt the definition of nonanticipative strategy as well as the (lower/upper) Value function introduced in (4.7)-(4.8). Denote by \( \hat{V}, \hat{\nu}, \hat{v} \) the upper, the lower and the Value function in the following theorem.

**Theorem 12** Suppose that functions \( f, \sigma \) satisfy the conditions (i), and \( \hat{V} \) satisfies (ii)-(iii) in Lemma 8. Assume that there exists \( a(\cdot) \in A(s) \) such that

\[
\mathbb{E}_\omega \left\{ \int_s^T |G(\tau, x_\tau, a_\tau, \beta[a_\tau])|d\tau \right\} < \infty
\]

for any \((s, x_s) \in Q\) and \( \beta \in \Delta(s) \). Let \( \hat{V} \) be a solution of the following HJI equation,

\[
\hat{V}_t(t, x_t) + \min_{a_t \in A_a} \max_{b_t \in B_a} \left\{ \Xi(t) \hat{V}(t, x_t) + G(t, x_t, a_t, b_t) \right\} = 0
\]

(4.24)

for any \((t, x_t) \in Q \) with the boundary \( \hat{V}(T, x_T) = Q(x_T) \) where \((T, x_T) \in \partial Q \) (the boundary of \( Q \)), such that \( \hat{V} \in C^{1,2}(Q) \) and continuous on \( Q \) (the closure of \( Q \)).

Then, (i)

\[
\hat{V}(t, x_t) \leq \sup_{\beta \in \Delta(t)} \{ \hat{J}(a, \beta[a]; t, x_t) \}
\]

for any \( a(\cdot) \in A(t) \) and \( (t, x_t) \in Q \);

(ii) if there exists \( a^*(\cdot) \in A(s) \) such that \( a^*_t \in A_a \) and it satisfies

\[
\max_{b_t \in B_a} \{ \Xi(t) \hat{V}(t, x_t) + G(t, x_t, a^*_t, b_t) \} = \min_{a_t \in A_a} \max_{b_t \in B_a} \{ \Xi(t) \hat{V}(t, x_t) + G(t, x_t, a_t, b_t) \}
\]

(4.25)

for any \( (t, x_t) \in Q \), then

\[
\hat{V}(s, x_s) = \sup_{\beta \in \Delta(s)} \{ J(a^*, \beta[a^*]; s, x_s) \}
\]

for any \( (t, x_t) \in Q \).

**Proof:** The theorem can be easily proved by extending Theorem 4.1 on page 159 in [78] that is based on a minimization problem to a minimax problem. \( \square \)

89
Note that in (4.24) and (4.25), the minimum and the maximum are attainable due to the continuity of the functions in the brackets and the compactness of the control set \( A_a \) and \( B_a \). The saddle point equilibrium can be specified in the following corollary, where the Isaacs condition holds.

**Corollary 1** Suppose that function \( f, \sigma \) satisfy the conditions (i), and \( \hat{V} \) satisfies (ii)-(iii) in Lemma 8. Assume that there exists \( a(\cdot) \in A(s) \) such that

\[
E_a \left\{ \int_s^T |\hat{G}(\tau, x_\tau, a_\tau, \beta[a_\tau])| d\tau \right\} < \infty
\]

for any \((s, x_s) \in Q\) and \( \beta \in \Delta(s) \). Let \( \hat{V} \) be a solution of the following HJI equation,

\[
\hat{V}_t(t, x_t) + \min_{a_t \in A_a} \max_{b_t \in B_a} \{ \Xi(t) \hat{V}(t, x_t) + \mathcal{G}(t, x_t, a_t, b_t) \}
\]

\[
= \hat{V}_t(t, x_t) + \max_{b_t \in B_a} \min_{a_t \in A_a} \{ \Xi(t) \hat{V}(t, x_t) + \mathcal{G}(t, x_t, a_t, b_t) \} = 0 \quad (4.26)
\]

for any \((t, x_t) \in Q\) with the boundary \( \hat{V}(T, x_T) = \mathcal{Q}(x_T) \) where \((T, x_T) \in \partial Q\) (the boundary of \( Q \)), such that \( \hat{V} \in C^{1,2}(Q) \) and continuous on \( \overline{Q} \) (the closure of \( Q \)).

Then, (i)

\[
\inf_{a \in \Gamma(t)} \{ \hat{J}(a[b]; s, x_s) \} \leq \hat{V}(s, x_s) \leq \sup_{\beta \in \Delta(t)} \{ \hat{J}(\beta[a]; s, x_s) \}
\]

for any \( a(\cdot) \in A(s), b(\cdot) \in B(s) \) and \((s, x_s) \in Q\);

(ii) if there exists \( a^*(\cdot) \in A(s) \) such that \( a^*_t \in A_a \) and it satisfies

\[
\max_{b_t \in B_a} \{ \Xi(t) \hat{V}(t, x_t) + \mathcal{G}(t, x_t, a^*_t, b_t) \} = \min_{a_t \in A_a} \max_{b_t \in B_a} \{ \Xi(t) \hat{V}(t, x_t) + \mathcal{G}(t, x_t, a_t, b_t) \}
\]

for any \((t, x_t) \in Q\), and there exists \( b^*(\cdot) \in B(s) \) such that \( b^*_t \in B_a \) and it satisfies

\[
\min_{a_t \in A_a} \{ \Xi(t) \hat{V}(t, x_t) + \mathcal{G}(t, x_t, a_t, b^*_t) \} = \max_{b_t \in B_a} \min_{a_t \in A_a} \{ \Xi(t) \hat{V}(t, x_t) + \mathcal{G}(t, x_t, a_t, b_t) \}
\]
for any \((s, x_s) \in Q\), then

\[
\hat{V}(s, x_s) = \sup_{\beta \in \Delta(s)} \{J(a^*, \beta[a^*]; s, x_s)\} = \inf_{\alpha \in \Gamma(s)} \{J(\alpha[b^*], b^*; s, x_s)\} = J(a^*, b^*; s, x_s)
\]

for any \((s, x_s) \in Q\). This is a saddle-point equilibrium and \(\hat{V}\) is the Value of the game.

**Remark 11** In Corollary 1, the Value function of the game is characterized by a (classical) solution of the HJI equation. Under (4.26), the Isaacs condition holds.

### 4.3.2 Solution to a Two-Player Pursuit-Evasion Game

In this section, we solve a specific two-player PE game in a \(\mathbb{R}^2\) space with \(\bar{x}\) and \(\bar{y}\) as the coordinates. Consider the following dynamics of the players,

\[
\begin{align*}
\mathrm{d}\bar{x}_\varsigma &= v_\varsigma \cos \theta_\varsigma \mathrm{d}t + \sigma_\varsigma \mathrm{d}\omega_{\bar{x}_\varsigma} \quad \text{with} \quad \bar{x}_\varsigma(0) = \bar{x}_{\varsigma 0}, \\
\mathrm{d}\bar{y}_\varsigma &= v_\varsigma \sin \theta_\varsigma \mathrm{d}t + \sigma_\varsigma \mathrm{d}\omega_{\bar{y}_\varsigma} \quad \text{with} \quad \bar{y}_\varsigma(0) = \bar{y}_{\varsigma 0}.
\end{align*}
\]

Equation (4.27) is a noise-corrupted version of the first-order Dubin’s car model in (3.8), where the subscript \(\varsigma \in \{p, e\}\) stands for the pursuer or the evader; \(\bar{x}_\varsigma\) and \(\bar{y}_\varsigma\) are the state variables (displacement) along the \(\bar{x}\)- and \(\bar{y}\)-axis respectively; \(v_\varsigma\) is the velocity; \(\theta_\varsigma\) is the control input; \(\omega_{\bar{x}_\varsigma}\) (\(\omega_{\bar{y}_\varsigma}\)) is a standard Wiener process along the \(\bar{x}\)-axis (\(\bar{y}\)-axis); \(\sigma_\varsigma\) is a constant. Assume that \(\omega_{\bar{x}_\varsigma}\) and \(\omega_{\bar{y}_\varsigma}\) are independent, and so are \(\omega_{\bar{x}_p}\) (\(\omega_{\bar{y}_p}\)) and \(\omega_{\bar{y}_e}\) (\(\omega_{\bar{y}_e}\)). Choose the capture time of the evader as the objective, i.e., \(J = T\) \((G = 1\) and \(Q = 0\) in (4.6)). Define \(x_\varsigma \triangleq [\bar{x}_\varsigma, \bar{y}_\varsigma]^T\), and we rewrite the dynamics in (4.27) as \(\mathrm{d}x_\varsigma = f_\varsigma(x_\varsigma, \theta_\varsigma) \mathrm{d}t + \sigma_\varsigma \mathrm{d}\omega_\varsigma\). The game ends when \(\|x_p - x_e\| \leq \varepsilon\) for some \(\varepsilon > 0\).
Theorem 13 Assume that $E\{T\} < \infty$, $v_p > v_e$ with $\varepsilon > (\sigma_p^2 + \sigma_e^2)/2(v_p - v_e)$, the Value function of the PE game $V(x_p, x_e)$ is given by

$$V(x_p, x_e) = \sqrt{\frac{(\bar{x}_p - \bar{x}_e)^2 + (\bar{y}_p - \bar{y}_e)^2}{v_p - v_e}} + \frac{\sigma_p^2 + \sigma_e^2}{4(v_p - v_e)^2} \ln \left[ \frac{\bar{x}_p - \bar{x}_e}{v_p - v_e} + \frac{(\bar{y}_p - \bar{y}_e)^2}{2v_p - 2v_e} \right] + C(\varepsilon)$$

where the constant $C(\varepsilon) = -\frac{\varepsilon}{v_p - v_e} - \frac{(\sigma_p^2 + \sigma_e^2) \cdot \ln(\varepsilon^2)}{4(v_p - v_e)^2}$.

Proof: We apply Corollary 1 to prove the theorem. We need to show that $V$ is a solution of the corresponding HJI equation as in (4.26). It suffices to show that $\min_{\theta_p} \max_{\theta_e} \{\Xi(t)V(x(t)) + 1\} = 0$. First of all, it is easy to check that $f, \sigma$ in (4.27) and $V$ satisfy conditions (i)-(iii) in Lemma 8. The assumption that $E\{T\} < \infty$ implies that condition in (iv) holds, i.e.,

$$E\left\{ \int_s^t |G(\tau, x_\tau)| d\tau \right\} < \infty.$$

Note that here $G = 1$ and $V = 0$ when $\sqrt{(\bar{x}_p - \bar{x}_e)^2 + (\bar{y}_p - \bar{y}_e)^2} = \varepsilon$ (on the boundary of the capture range), and here $V$ is not an explicit function of time. By (4.22),

$$\Xi(t)V = \frac{\partial V}{\partial x_p} \cdot f_p + \frac{\partial V}{\partial x_e} \cdot f_e + \frac{1}{2} \sum_{i=1}^2 \frac{\partial^2 V}{\partial^2 x_{pi}} \sigma_p^2 + \frac{1}{2} \sum_{j=1}^2 \frac{\partial^2 V}{\partial^2 x_{ej}} \sigma_e^2.$$

(4.29)

Here, $i = 1, 2$ ($j = 1, 2$) stands for $\bar{x}$ and $\bar{y}$ respectively. Note that $\omega_{\bar{x}}$ and $\omega_{\bar{y}}$ are independent. Substitute (4.27) into (4.29) to replace $f_p$ and $f_e$. After a little
manipulation, the terms $D_1$ and $D_2$ in (4.29) become

\[
D_1 = \frac{1}{v_p - v_e} \left( \frac{(\bar{x}_p - \bar{x}_e)}{\sqrt{(\bar{x}_p - \bar{x}_e)^2 + (\bar{y}_p - \bar{y}_e)^2}} v_p \cos(\theta_p) \right)
+ \frac{(\bar{y}_p - \bar{y}_e)}{\sqrt{(\bar{x}_p - \bar{x}_e)^2 + (\bar{y}_p - \bar{y}_e)^2}} v_p \sin(\theta_p) + \frac{(\bar{x}_p - \bar{x}_e)}{\sqrt{(\bar{x}_p - \bar{x}_e)^2 + (\bar{y}_p - \bar{y}_e)^2}} v_e \cos(\theta_e)
+ \frac{\sigma_p^2 + \sigma_e^2}{2(v_p - v_e)^2} \left( \frac{2(\bar{x}_p - \bar{x}_e)}{(\bar{x}_p - \bar{x}_e)^2 + (\bar{y}_p - \bar{y}_e)^2} v_p \cos(\theta_p) \right)
+ \frac{2(\bar{x}_p - \bar{x}_e)}{(\bar{x}_p - \bar{x}_e)^2 + (\bar{y}_p - \bar{y}_e)^2} v_p \sin(\theta_p) + \frac{2(\bar{x}_p - \bar{x}_e)}{(\bar{x}_p - \bar{x}_e)^2 + (\bar{y}_p - \bar{y}_e)^2} v_e \cos(\theta_e)
+ \frac{2(\bar{x}_p - \bar{x}_e)}{(\bar{x}_p - \bar{x}_e)^2 + (\bar{y}_p - \bar{y}_e)^2} v_e \sin(\theta_e) \right),
\]

and,

\[
D_2 = \sigma_p^2 (v_{x_p \bar{x_p}} + v_{y_p \bar{y_p}}) + \sigma_e^2 (v_{x_e \bar{x_e}} + v_{y_e \bar{y_e}})
= \frac{1}{2(v_p - v_e)} \left( \frac{(\bar{x}_p - \bar{x}_e)^2}{[(\bar{x}_p - \bar{x}_e)^2 + (\bar{y}_p - \bar{y}_e)^2]^{3/2}} \sigma_p^2 + \frac{(\bar{y}_p - \bar{y}_e)^2}{[(\bar{x}_p - \bar{x}_e)^2 + (\bar{y}_p - \bar{y}_e)^2]^{3/2}} \sigma_e^2 \right)
+ \frac{(\bar{x}_p - \bar{x}_e)^2}{[(\bar{x}_p - \bar{x}_e)^2 + (\bar{y}_p - \bar{y}_e)^2]^{3/2}} \sigma_p^2 + \frac{(\bar{y}_p - \bar{y}_e)^2}{[(\bar{x}_p - \bar{x}_e)^2 + (\bar{y}_p - \bar{y}_e)^2]^{3/2}} \sigma_e^2
+ \frac{\sigma_p^2 + \sigma_e^2}{8(v_p - v_e)^2} \left( \frac{(\bar{y}_p - \bar{y}_e)^2 - (\bar{x}_p - \bar{x}_e)^2}{[(\bar{x}_p - \bar{x}_e)^2 + (\bar{y}_p - \bar{y}_e)^2]^{3/2}} \sigma_p^2 + \frac{(\bar{x}_p - \bar{x}_e)^2 - (\bar{y}_p - \bar{y}_e)^2}{[(\bar{x}_p - \bar{x}_e)^2 + (\bar{y}_p - \bar{y}_e)^2]^{3/2}} \sigma_e^2 \right)
+ \frac{(\bar{y}_p - \bar{y}_e)^2 - (\bar{x}_p - \bar{x}_e)^2}{[(\bar{x}_p - \bar{x}_e)^2 + (\bar{y}_p - \bar{y}_e)^2]^{3/2}} \sigma_p^2 + \frac{(\bar{x}_p - \bar{x}_e)^2 - (\bar{y}_p - \bar{y}_e)^2}{[(\bar{x}_p - \bar{x}_e)^2 + (\bar{y}_p - \bar{y}_e)^2]^{3/2}} \sigma_e^2 \right).
\]
By inspection of (4.29)-(4.31), only the term $D_1$ in (4.30) involves control $\theta_p$ and $\theta_e$. Clearly,

\[
\min_{\theta_p} \max_{\theta_e} \left\{ D_1 \right\} = \frac{1}{v_p - v_e} (-v_p + v_e) + \frac{\sigma_p^2 + \sigma_e^2}{4(v_p - v_e)^2} \left( -\frac{2(x_p - \bar{x}_e)^2}{[x_p - \bar{x}_e)^2 + (y_p - \bar{y}_e)^2]^{3/2}} v_p + \frac{2(y_p - \bar{y}_e)^2}{[x_p - \bar{x}_e)^2 + (y_p - \bar{y}_e)^2]^{3/2}} v_e \right) + \frac{2(\bar{y}_p - \bar{y}_e)^2}{[(x_p - \bar{x}_e)^2 + (y_p - \bar{y}_e)^2]^{3/2}} v_e
\]

\[
= -1 - \frac{1}{2} \frac{\sigma_p^2 + \sigma_e^2}{v_p - v_e} \frac{1}{\sqrt{(x_p - \bar{x}_e)^2 + (y_p - \bar{y}_e)^2}}.
\]

(4.32)

On the other hand, $D_2$ in (4.31) can be simplified as

\[
D_2 = \frac{1}{2} \frac{\sigma_p^2 + \sigma_e^2}{v_p - v_e} \frac{1}{\sqrt{(x_p - \bar{x}_e)^2 + (y_p - \bar{y}_e)^2}}.
\]

(4.33)

By (4.32)-(4.33), $\min_{\theta_p} \max_{\theta_e} \left\{ \Xi(t)\mathcal{V}(x(t)) + 1 \right\} = 0$. Thus, $\mathcal{V}$ is a solution of the HJI equation. Furthermore, since the dynamics of the pursuers and the evaders are independent, function $f$ is separable in the sense of Theorem 9. Due to the separability of $G$ ($G = 1$), by Theorem 9 and Corollary 1, $\mathcal{V}$ is the Value.

\[\blacksquare\]

**Remark 12** Compared with the Value function (3.12) of a corresponding deterministic game with the same objective, the Value (4.28) here has an additional term due to the disturbance.

With the Value derived in (4.28), a multi-player stochastic PE game that has the objective chosen as the sum of the capture times of all the evader and the dynamics in (4.27) can be decomposed into distributed two-player games. Then, the iterative method is applicable.
4.3.3 On Finite Expectation of the Capture Time

In this section, we study the condition of $E\{T\} < \infty$ in Theorem 12. We first introduce the following lemma.

**Lemma 9** For any $\lambda$, $0 < \lambda < 1$ and $k \in \mathbb{Z}_{\geq 0}$, $\sum_{n=1}^{\infty} n^k \lambda^{-n} < \infty$.

The proof is omitted and can be found in real analysis texts such as [79].

Let us first consider a simplified game in a one dimensional space as shown in Figure 4.1, where the dynamics of the players is similar to (4.27) but with states in $\mathbb{R}$, i.e.,

\[ dx_\varsigma = v_\varsigma u_\varsigma dt + \sigma_\varsigma d\omega_\varsigma \]

Here, $\varsigma \in \{p,e\}$; $v_\varsigma$ is the velocity; $u_\varsigma \in \{1,-1\}$ is the control variable; $\omega$ is a one dimensional standard Wiener process. Clearly, in order to capture the evader, the pursuer moves towards the evader and the evader follows the same direction, i.e., $u_p = u_e = 1$ as illustrated in Figure 4.1. Let $x = x_e - x_p$, and then, the dynamics of the game becomes

\[ dx = vdt + \sigma d\omega, \] (4.34)

where $v = v_e - v_p$ and $\sigma^2 = \sigma_p^2 + \sigma_e^2$.

\[ \begin{align*}
&\overset{v_p}{\longrightarrow} \quad \overset{v_e}{\longrightarrow} \\
&\text{Pursuer} \quad \quad \quad \quad \text{Evader} \\
&\quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad Mar 15, 2023
Theorem 14 If $v_p > v_e$ and $\varepsilon \geq 0$, then $E\{T\} < \infty$, where $T$ is the capture time, $T = \inf\{t \mid |x(t)| \leq \varepsilon\}$.

Proof: Suppose that the game starts at time 0 and $x(0) > \varepsilon$, otherwise $T = 0$. By the property of Wiener processes, the state $x_t$ at time $t > 0$ is a Gaussian random variable, i.e., $x_t \sim \mathcal{N}(\mu_{x_t}, \sigma^2_{x_t})$. Here, the mean is $\mu_{x_t} = x(0) + vt$ and the variance is $\sigma^2_{x_t} = (\sigma^2_p + \sigma^2_e)t$. Let $\bar{v} = -v = v_p - v_e > 0$. The fact that the evader has not been captured by time $t > 0$ implies that $x_t > \varepsilon$ at least time $t$. Denote by $P(T > t)$ the probability that evader has not been captured by time $t$. It satisfies

$$P(T > t) \leq P(x_t > \varepsilon) = \int_{x > \varepsilon} \frac{1}{\sigma \sqrt{2\pi t}} \exp \left( -\frac{(x - \mu_{x_t})^2}{2\sigma^2_{x_t}} \right) dx$$

$$= \int_{x > \varepsilon} \frac{1}{\sigma \sqrt{2\pi t}} \exp \left( -\frac{(x - x(0) - \bar{v}t)^2}{2\sigma^2_{x_t}} \right) dx. \quad (4.35)$$

Define $t_0 = (x(0) - \varepsilon)/\bar{v}$, $\tilde{t} = t - t_0$ and $r = x + \bar{v}\tilde{t}$, such that $P(x_t > \varepsilon)$ in (4.35) can be rewritten as

$$P(x_t > \varepsilon) = \int_{x > 0} \frac{1}{\sigma \sqrt{2\pi (t_0 + \tilde{t})}} \exp \left( -\frac{(x + \bar{v}\tilde{t})^2}{2\sigma^2(t_0 + \tilde{t})} \right) dx$$

$$= \int_{r > \bar{v}\tilde{t}} \frac{1}{\sigma \sqrt{2\pi (t_0 + \tilde{t})}} \exp \left( -\frac{r^2}{2\sigma^2(t_0 + \tilde{t})} \right) dr.\quad (4.35)$$

Choose some $t_1$ such that $\bar{v}(t_1 - t_0) > 1$, and define

$$t_2 = \max\{2t_0, t_1\}. \quad (4.36)$$

Consider the time $t$ when $t > t_2$, i.e., $\tilde{t} > t_0$ and $\bar{v}\tilde{t} > 1$, and then

$$P(x_t > \varepsilon) < \frac{1}{\sigma \sqrt{4\pi t_0}} \int_{r > \bar{v}\tilde{t}} \exp \left( -\frac{r^2}{4\sigma^2\tilde{t}} \right) dr$$

$$< \frac{1}{\sigma \sqrt{4\pi t_0} t_0} \int_{r > \bar{v}\tilde{t}} \exp \left( -\frac{\bar{v}^2}{4\sigma^2\tilde{t}} \right) dr$$

$$= \frac{2\sigma}{\bar{v} \sqrt{\pi t_0}} \exp \left( -\frac{\bar{v}^2}{4\sigma^2\tilde{t}} \right) = \frac{2\sigma}{\bar{v} \sqrt{\pi t_0}} \exp \left( -\frac{\bar{v}^2}{4\sigma^2(t - t_0)} \right). \quad (4.37)$$
Denote by $p_T(t)$ the probability density of the capture time $T$, the expectation $\mathbb{E}\{T\}$ is
\[
\mathbb{E}\{T\} = \int_0^\infty t \cdot p_T(t) dt = \int_0^{t_2} t \cdot p_T(t) dt + \int_{t_2}^\infty t \cdot p_T(t) dt, \tag{4.38}
\]
where $t_2$ is chosen as in (4.36). Next, we shall show that the second term on the right hand side of (4.38) is finite, which implies that $\mathbb{E}\{T\}$ is finite. Choose a small $\delta t > 0$, the second term can be written as follows, illustrated in Figure 4.2.

Figure 4.2: The Probability Density of the Capture Time $p_T(t)$

\[
\int_{t_2}^\infty t \cdot p_T(t) dt = \sum_{k=0}^{\infty} \left( \int_{t_2+k\delta t}^{t_2+(k+1)\delta t} t \cdot p_T(t) dt \right) \tag{4.39}
\]

Each term in the summation in (4.39) satisfies
\[
\int_{t_2+k\delta t}^{t_2+(k+1)\delta t} t \cdot p_T(t) dt < (t_2 + (k + 1)\delta t) \int_{t_2+k\delta t}^{t_2+(k+1)\delta t} p_T(t) dt < (t_2 + (k + 1)\delta t) \int_{t_2+k\delta t}^{\infty} p_T(t) dt \tag{4.40}
\]

Note that $P(T > t_2 + k\delta t) = \int_{t_2+k\delta t}^\infty p_T(t) dt$. By (4.35) and (4.37),
\[
\int_{t_2+k\delta t}^{\infty} p_T(t) dt < \frac{2\sigma}{\bar{v}\sqrt{\pi}t_0} \exp \left( -\frac{\bar{v}^2}{4\sigma^2}(t_2 + k\delta t - t_0) \right). \tag{4.41}
\]
Substitute (4.41) into (4.39), and by Lemma 9,
\[
\int_{t_2}^{\infty} t \cdot p_T(t) \, dt < \sum_{k=0}^{\infty} \left[ (t_2+(k+1)\delta t) \frac{2\sigma}{\sqrt{\pi t_0}} \exp \left( -\frac{\tilde{v}^2}{4\sigma^2} \right) \right] < \infty. \tag{4.42}
\]
By (4.38) and (4.42), \( E\{T\} < \infty \).

Now, we study the game in a \( \mathbb{R}^2 \) space specified in Theorem 13. First of all, change the variables as \( \tilde{x} = \bar{x} - \bar{x}_0 \) and \( \tilde{y} = \bar{y} - \bar{y}_0 \). According to (4.32), the optimal control of the pursuer coincides with that of the evader, namely, \( \theta_p^* = \theta_e^* \). Suppose that both the pursuer and the evader use the same optimal control and denoted by \( \theta \). Then the dynamic equation of players can be rewritten as
\[
d\tilde{x} = (v_p - v_e) \cos \theta(t) \, dt + \sigma_p \, d\omega_{p\tilde{x}} - \sigma_e \, d\omega_{e\tilde{x}} \quad \text{with} \quad \tilde{x}(0) = \tilde{x}_0, \tag{4.43a}
\]
\[
d\tilde{y} = (v_p - v_e) \sin \theta(t) \, dt + \sigma_p \, d\omega_{p\tilde{y}} - \sigma_e \, d\omega_{e\tilde{y}} \quad \text{with} \quad \tilde{y}(0) = \tilde{y}_0. \tag{4.43b}
\]
In (4.43a), \( \omega_{p\tilde{x}} \) and \( \omega_{e\tilde{x}} \) are independent, such that the term \( \sigma_p \, d\omega_{p\tilde{x}} - \sigma_e \, d\omega_{e\tilde{x}} \) is equivalent to a random process \( \sigma \, d\omega_{\tilde{x}} \) with \( \sigma = \sqrt{\sigma_p^2 + \sigma_e^2} \), where \( \omega_{\tilde{x}} \) is a standard Wiener process. Similarly, \( \sigma \, d\omega_{\tilde{y}} \) can be defined according to \( \sigma_p \, d\omega_{p\tilde{y}} - \sigma_e \, d\omega_{e\tilde{y}} \) in (4.43b). Let \( v = v_p - v_e \) and \( x = [\tilde{x}, \tilde{y}]^T \). Define \( \rho(x) = \sqrt{\tilde{x}^2 + \tilde{y}^2} \). The following theorem provides sufficient conditions under which the capture time \( T \) of the evader has a finite expectation.

**Theorem 15** In a PE game with the dynamics given in (4.27), if \( v_p > v_e, \varepsilon > (\sigma_p^2 + \sigma_e^2)/2(v_p - v_e) \) and the probability \( P(\rho(x_t) > \varepsilon) \) under the optimal control determined in (4.32) satisfies
\[
P(\rho(x_t) > \varepsilon) < \int_{\varepsilon}^{\infty} \frac{C \rho}{\sigma^2t} \exp \left( -\frac{(\rho - \rho_0 + \kappa^2t)^2}{2\sigma^2t} \right) \, d\rho \tag{4.44}
\]
for some \( C > 0, \kappa \in \mathbb{R} (\kappa \neq 0) \) and any \( t > 0 \), then \( E\{T\} < \infty \), where \( \rho_0 = \sqrt{\tilde{x}_0^2 + \tilde{y}_0^2} \).
Proof: The proof is similar to that of Theorem 14. Let $t_0 = (\rho_0 - \varepsilon)/\kappa^2$, $\bar{\ell} = t - t_0$ and $r = \rho + \kappa^2\bar{\ell}$. By (4.44),

\[
P(\rho(x_t) > \varepsilon) < \int_{\varepsilon}^{\infty} \frac{C\rho}{\sigma^2 t} \exp \left( - \frac{(\rho - \rho_0 + \kappa^2 t)^2}{2\sigma^2 t} \right) d\rho
\]

\[
= \int_{0}^{\infty} \frac{C\rho}{\sigma^2(t + t_0)} \exp \left( - \frac{(\rho + \kappa^2\bar{\ell})^2}{2\sigma^2(t + t_0)} \right) d\rho
\]

\[
\left(\text{because } r = \rho + \kappa^2\bar{\ell}\right)
\]

\[
= \int_{\kappa^2\bar{\ell}}^{\infty} \frac{C(r - \kappa^2\bar{\ell})}{\sigma^2(t + t_0)} \exp \left( - \frac{r^2}{2\sigma^2(t + t_0)} \right) dr.
\]

Choose some $t_1$ such that $\kappa^2(t_1 - t_0) > 1$ and let $t_2 = \max\{2t_0, t_1\}$. Consider that $t > t_2$, such that $\bar{\ell} > t_0$ and $\kappa^2\bar{\ell} > 1$.

\[
P(\rho(x_t) > \varepsilon) < \int_{\kappa^2\bar{\ell}}^{\infty} \frac{C(r - \kappa^2\bar{\ell})}{\sigma^2(t + t_0)} \exp \left( - \frac{r^2}{2\sigma^2(t + t_0)} \right) dr
\]

\[
< \int_{\kappa^2\bar{\ell}}^{\infty} \frac{C(r - \kappa^2\bar{\ell})}{2\sigma^2 t_0} \exp \left( - \frac{\kappa^2\bar{\ell} \cdot r}{4\sigma^2 t_0} \right) dr
\]

\[
= \int_{\kappa^2\bar{\ell}}^{\infty} \frac{C(r - \kappa^2\bar{\ell})}{2\sigma^2 t_0} \exp \left( - \frac{\kappa^2 r}{4\sigma^2} \right) dr
\]

\[
= \frac{8C\sigma^2}{\kappa^4 t_0} \exp \left( - \frac{\kappa^4(t - t_0)}{4\sigma^2} \right)
\]

for any $t \geq 0$. The rest of the proof follows the proof of Theorem 14 from (4.38).

\[\text{Remark 13 If the conditions in Theorem 15 hold, Pr}(T < \infty) = 1.\]

Justification of the Condition in (4.44)

In the following, we shall justify that the condition in (4.44). The discussion is rather heuristic than mathematically rigorous. The complete characterization of the distribution of the states requires an analytical solution of a PDE called Fokker
Planck Equation (FPE) [80] as
\[
\frac{\partial p(t,x)}{\partial t} = -\sum_{i=1}^{2} \frac{\partial}{\partial x_i} [p(t,x)f_i(x,\theta)] + \frac{1}{2} \sum_{i=1}^{2} \frac{\partial^2 p(t,x)}{\partial x_i^2} \sigma^2 \text{ with } p(0,x) = \delta(x - x_0),
\]
\[(4.46)\]

where \(\delta(\cdot)\) is Dirac-Delta function; \(f_i\) is the \(i^{th}\) element of function \(f\) in (4.43), where \(i = 1, 2\) stands for \(\hat{x}\) and \(\hat{y}\) respectively. An analytical solution of the equation (4.46) \(p(t,x)\) is formidable. In the following, we provide an upper-bound of \(p(t,x)\).

First, we analyze a simplified situation where the control variable \(\theta\) in (4.46) is fixed, i.e., \(\theta_t = \theta_0\) for \(t \geq 0\). Based on the property of the Wiener process, the probability distribution \(p^{\theta_0}_x(t,x)\) at time \(t\) is
\[
p^{\theta_0}_x(t,x) = \frac{1}{2\pi \sigma^2 t} \exp \left( -\frac{(\hat{x} - \mu_x(t))^2 + (\hat{y} - \mu_y(t))^2}{2\sigma^2 t} \right),
\]
\[(4.47)\]
where \(\mu_x(t) = \tilde{x}_0 + vt \cos \theta_0\) and \(\mu_y(t) = \tilde{y}_0 + vt \sin \theta_0\). Here in the notation of \(p^{\theta_0}_x\), the superscript indicates the fixed control \(\theta_0\) and the subscript \(x\) implies that the distribution is in the \(x\) coordinates. Now, we change the coordinates by a transformation \(G_t\) at time \(t\), which is illustrated in Figure 4.3.
\[
\hat{x} \triangleq \begin{bmatrix} x' \\ y' \end{bmatrix} = G_t \begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix}, \text{ with } G_t = \begin{bmatrix} \cos \beta_t & \sin \beta_t \\ -\sin \beta_t & \cos \beta_t \end{bmatrix},
\]
\[(4.48)\]
where \(\beta_t\) is the angle between the line of sight from the evader to the pursuer and the \(\hat{x}\)-axis at time \(t\).

Clearly, \(x'\) and \(y'\) are jointly Gaussian with the mean
\[
\begin{bmatrix} \mu_{x'} \\ \mu_{y'} \end{bmatrix} = \begin{bmatrix} \cos \beta_t & \sin \beta_t \\ -\sin \beta_t & \cos \beta_t \end{bmatrix} \begin{bmatrix} \mu_\hat{x} \\ \mu_\hat{y} \end{bmatrix} = \begin{bmatrix} \mu_\hat{x} \cos \beta_t + \mu_\hat{y} \sin \beta_t \\ -\mu_\hat{x} \sin \beta_t + \mu_\hat{y} \cos \beta_t \end{bmatrix} = \begin{bmatrix} \sqrt{\mu_\hat{x}^2 + \mu_\hat{y}^2} \\ 0 \end{bmatrix}
\]
\[(4.49)\]
and the covariance \(\text{Cov}_t\),
\[
\text{Cov}_t = G_t \begin{bmatrix} 2\sigma^2 t & 0 \\ 0 & 2\sigma^2 t \end{bmatrix} G_t^T = \begin{bmatrix} 2\sigma^2 t & 0 \\ 0 & 2\sigma^2 t \end{bmatrix}.
\]
Thus, \( x' \) and \( y' \) are independent. Note that \( \hat{x} = [x'^T, y'^T]^T \) and the distribution in the new coordinates \( p_{\hat{x}}^{\theta_0} \) becomes

\[
p_{\hat{x}}^{\theta_0}(t, \hat{x}) = \frac{1}{2\pi\sigma^2 t} \exp \left( - \frac{\left( x' - \sqrt{\mu_x^2(t) + \mu_y^2(t)} \right)^2 + y'^2}{2\sigma^2 t} \right). \tag{4.50}
\]

Next, we change the \( x'-y' \) coordinates to the \( \rho-\varphi \) (polar) coordinate system as

\[
\rho_{x',y'} = \sqrt{x'^2 + y'^2}, \quad \varphi_{x',y'} = \tan^{-1}(y'/x') \quad \text{with} \quad \varphi_{x',y'} \in [0, 2\pi).
\]

Here, \( \tan^{-1} \) is slightly different from the standard definition of the function \( \tan^{-1} \), and its range is the interval \( [0, 2\pi) \). At each \( (x', y') \), the absolute value of the Jacobian \( J_{G_t}(x', y') \) of the transformation \( G_t \) is

\[
|J_{G_t}(x', y')| = \det \begin{pmatrix}
\frac{x'}{\sqrt{x'^2+y'^2}} & \frac{y'}{\sqrt{x'^2+y'^2}} \\
\frac{y'}{x'^2+y'^2} & \frac{-x'}{x'^2+y'^2}
\end{pmatrix} = \frac{1}{\sqrt{x'^2+y'^2}} = \frac{1}{\rho_{x',y'}}.
\]
Now, we derive the probability density $p^\theta_{\rho, \varphi}(t, \rho, \varphi)$ in the $\rho$-$\varphi$ coordinates. By (4.47) and (4.48),
\[
p^\theta_{\rho, \varphi}(t, \rho, \varphi) = |J_{G_t}(x', y')|^{-1} p^\theta_2(t, x) = \frac{\rho}{2\pi\sigma^2 t} \exp \left( -\frac{(\rho \cos \varphi - \sqrt{\mu_x^2(t) + \mu_y^2(t)})^2 + (\rho \sin \varphi)^2}{2\sigma^2 t} \right) \tag{4.51}
\]
Let $\bar{\rho}_{\mu_x, \mu_y} = \sqrt{\mu_x^2 + \mu_y^2}$. If $\varphi = 0$,
\[
p^\theta_{\rho, \varphi}(t, \rho, \varphi = 0) = \frac{\rho}{2\pi\sigma^2 t} \exp \left( -\frac{(\rho - \bar{\rho}_{\mu_x, \mu_y})^2}{2\sigma^2 t} \right) \text{ for } \rho \geq 0. \tag{4.52}
\]
By inspection of (4.51) and (4.52), for any $\varphi \neq 0$, $p_{\rho, \varphi}(t, \rho, \varphi) < p_{\rho, \varphi}(t, \rho, \varphi = 0)$. Then, the probability that $\rho > r$ for some $r \geq 0$ under control $\theta_0$ at time $t$, $P^{\theta_0}(\rho > r; t)$ satisfies
\[
\begin{align*}
P^{\theta_0}(\rho > r; t) &< \int_0^{2\pi} d\varphi \int_r^\infty p_{\rho, \varphi}(t, \rho, \varphi = 0) d\rho \\
&= \int_0^{2\pi} d\varphi \int_r^\infty \frac{\rho}{2\pi\sigma^2 t} \exp \left( -\frac{(\rho - \bar{\rho}_{\mu_x, \mu_y})^2}{2\sigma^2 t} \right) d\rho. \tag{4.53}
\end{align*}
\]
Next, we study the probability $P^{\theta^*}(\rho > r; t)$ when both the pursuer and the evader exploit the optimal control $\theta^*$ determined in (4.39). Intuitively, optimal control $\theta^*$ (in state feedback) drives each state $(\tilde{x}, \tilde{y})$ towards the origin, such that it outperforms (static) $\theta_0$ with respect to minimizing the distance $\rho$ between the pursuer and the evader. On the other hand, suppose that the state $x_{\theta^*}(t)$ under $\theta^*$ has a smaller variance $\sigma^2$ at any time $t$ compared to that under a fixed control $\theta_0$. Similar to (4.53), we assume that the following inequality is true. Denote by $\mu_x^*(t)$ and $\mu_y^*(t)$ the mean of $\tilde{x}$ and $\tilde{y}$ at time $t \geq 0$ under $\theta^*$ and $\bar{\rho}_{\mu_x^*, \mu_y^*}(t) = \sqrt{\mu_{x, 1}^*(t)^2 + \mu_{y, 1}^*(t)^2}$. 

102
Assumption 5

\[ P^\theta_\omega (\rho > r; t) < \int_0^{2\pi} \, d\varphi \int_r^\infty \frac{C \rho}{2\pi\sigma^2 t} \exp \left( -\frac{(\rho - \bar{\rho}_\omega(t))^2}{2\sigma^2 t} \right) \, d\rho \quad (4.54) \]

for some constant \( C > 0 \).

Since function \( \rho = \sqrt{\tilde{x}^2 + \tilde{y}^2} \) is convex, by Jensen’s inequality,

\[ \bar{\rho}_\omega \mu_{\tilde{x}^*, \tilde{y}^*} = \sqrt{\mu_{\tilde{x}^*}^2 + \mu_{\tilde{y}^*}^2} \leq E \left\{ \sqrt{\tilde{x}_{\theta^*}^2 + \tilde{y}_{\theta^*}^2} \right\} = \bar{\rho}_{\theta^*}. \]

Thus, inequality (4.54) holds when \( \bar{\rho}_\omega \mu_{\tilde{x}^*, \tilde{y}^*} \) is replaced by \( \bar{\rho}_{\theta^*} \), i.e.,

\[ P^\theta_\omega (\rho > r; t) < \int_0^{2\pi} \, d\varphi \int_r^\infty \frac{C \rho}{2\pi\sigma^2 t} \exp \left( -\frac{(\rho - \bar{\rho}_{\theta^*}(t))^2}{2\sigma^2 t} \right) \, d\rho. \quad (4.55) \]

In the following, we derive the evolution of \( \bar{\rho}_{\theta^*}(t) \) under \( \theta^* \). Let \( \rho_{\theta^*} = \sqrt{\tilde{x}_{\theta^*}^2 + \tilde{y}_{\theta^*}^2} \) along the trajectory under optimal control \( \theta^* \). The evolution of \( \rho_{\theta^*}(t) \) can be derived based on the Ito’s rule of integral. According to the definition of \( \Xi(t) \) in (4.21), \( \rho_{\theta^*}(t) \) at any \((\tilde{x}, \tilde{y})\) under the control \( \theta^* \) is

\[
\dot{\rho}_{\theta^*}(t) = \frac{\partial \rho_{\theta^*}(t)}{\partial \tilde{x}} v \cos \theta^* + \frac{\partial \rho_{\theta^*}(t)}{\partial \tilde{y}} v \sin \theta^* + \frac{1}{2} \frac{\partial^2 \rho_{\theta^*}(t)}{\partial \tilde{x}^2} \sigma^2 + \frac{1}{2} \frac{\partial^2 \rho_{\theta^*}(t)}{\partial \tilde{y}^2} \sigma^2
\]

\[ = -v + \frac{1}{2\rho_{\theta^*}(t)} \sigma^2. \quad (4.56) \]

If the evader has not been captured by \( t \), then \( \rho(\tau) > \varepsilon \) for any \( \tau \) \((0 \leq \tau \leq t)\). Here, we focus on the set \( S = \{ (\tilde{x}, \tilde{y}) | \sqrt{\tilde{x}^2 + \tilde{y}^2} > \varepsilon \} \). Since \( \varepsilon > \sigma^2/2v \) (see Theorem 15), it follows from (4.56) that

\[
\dot{\rho}_{\theta^*}(t) = -v + \frac{1}{2\rho_{\theta^*}(t)} \sigma^2 < -v + \frac{1}{2\varepsilon} \sigma^2 = -\kappa^2 < 0 \quad (4.57)
\]

for any \( 0 \leq \tau \leq t \). Equation (4.57) indicates that the decreasing rate of \( \rho \) at any point in \( S \) is less than \( -\kappa^2 \). Since this holds at any \((\tilde{x}, \tilde{y})\) for \( \rho > \varepsilon \), the decreasing
rate of $\bar{\rho}_\theta^*$, the expectation of $\rho_\theta^*$, is less than $-\kappa^2$. Namely, $\bar{\rho}_\theta^*(t) < \rho_0 - \kappa^2 t$. This relationship holds as long as the state $(\tilde{x}, \tilde{y})$ remains in $S$. The inequality of (4.55) holds with $\bar{\rho}_\theta^*(t)$ replaced by $\rho_0 - \kappa^2 t$, which yields (4.44).

4.4 On Existence of Solutions to Stochastic Pursuit-Evasion Games with Perfect State Information

In this section, we outline the study on the existence of Value for the stochastic PE games with the perfect state information pattern, and general issues are discussed with suggestions for the future study. As in the deterministic case, this problem is discussed under the framework of the viscosity solution of (second-order) HJ equations.

Consider the players’ dynamics formulated in (4.3). Recall the second order differential operator defined for the stochastic process in (4.21). It is expected that the corresponding HJI equation (with a fixed $z$) for a stochastic game problem has the following form,

$$\frac{\partial u^z}{\partial t} + H_z(x, \frac{\partial u^z}{\partial x}, \frac{\partial^2 u^z}{\partial x^2}) = 0. \quad (4.58)$$

Here, the HJI equation in (4.58) is defined on $\Lambda^c_z$ with proper boundary conditions, where $H_z$ can be

$$H_z(x, p, D) = H_z^-(x, p, D) = \max_{b \in B_a} \min_{a \in A} \left\{ p \cdot f(x, a, b) + \frac{1}{2} \text{tr} (\Sigma(x) D) + G(x, z, a, b) \right\},$$

or,

$$H_z(x, p, D) = H_z^+(x, p, D) = \min_{a \in A} \max_{b \in B_a} \left\{ p \cdot f(x, a, b) + \frac{1}{2} \text{tr} (\Sigma(x) D) + G(x, z, a, b) \right\}.$$

Here, $\Sigma(x) = \sigma(x)\sigma^T(x)$, in which $\sigma(x)$ is given in (4.3). Note that (4.58) is a second-order PDE and its viscosity solution can be defined similarly according to Definition 4 with $H(t, x, p)$ replaced by $H_z(x, p, D)$. 

104
The method for a deterministic PE game can be used to show the existence of (upper/lower) Value of a Stochastic PE game. Suppose that an approximate upper Value function can be determined by a suboptimal control under a structure of SSTC.

The major step is to show that the image of transformation $\mathcal{H}$ of a continuous function $W_k$ is the (unique) viscosity solution of the corresponding HJI equation in (4.58), where $\mathcal{H}$ is defined as

$$
\mathcal{H}[W_k](x, z) = \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} \mathbb{E}_\omega \left\{ \int_t^{t+\Delta t} G(x_\tau, z_\tau, a_\tau, \beta[a]\tau) d\tau + W_k(x_{t+\Delta t}, x, a, [a], \omega, z, x_{t+\Delta t}) \right\}.
$$

This can be proved if $\mathcal{H}[W_k](x, z)$ in (4.59) satisfies the principle of DP [50], because the viscosity solution is a local property. In contrast to deterministic problems, the principle of DP for stochastic problems is non-trivial due to some measurability problem [50, 19]. Once $\mathcal{H}[W_k](x, z)$ is shown to be a viscosity solution of the corresponding HJI equation for any $k \in \mathbb{Z}_{\geq 0}$, by iteration, the upper Value function can be approached in the limit by Theorem 11. Furthermore, according to the most general uniqueness results on the viscosity solution of a second-order HJ equation by Ishii and Lions in [48] (Theorem VI.5 on pp.65), $\mathcal{H}[W_k](x, z)$ for $k \in \mathbb{Z}_{\geq 0}$ can be shown as the unique viscosity solution of the equation (4.58). Then, with a separable function $G$ as in (2.65), it follows that the upper and the lower Value coincide, such that the Value function of a stochastic game exists.

In this section, we only provide some guidelines to show the existence of the Value function of a stochastic PE game under the framework of the viscosity solution theory of HJ equations. A thorough investigation on this topic is a potential future research direction. The study may start with [50, 48, 19] and the references therein.
4.5 Stochastic Pursuit-Evasion Games with Imperfect State Information

Up to this point, we have discussed stochastic multi-player PE games with perfect state information. In this section, we deal with a more realistic model, where the measurements of the players are no longer perfect. This is a very difficult problem even for control (with one player) problems. Stochastic games may encounter additional difficulty since the fundamental assumption of common knowledge among the players may no longer be valid due to the distinct noisy measurements by the pursuers and the evaders respectively [81, 82, 83]. To avoid this kind of difficulties from the information structure, we only deal with a class of special problems where the pursuers have noisy measurements but the evaders can still measure the states perfectly. This presents a worst-case scenario from the pursuers’ perspective. In the following, we first propose a suboptimal approach to a stochastic control problem with limited look-ahead, based on which the results on PE game problems are presented afterwards.

4.5.1 Limited Look-ahead for Stochastic Control Problems

We consider a stochastic control problem with the following system dynamic and measurement equations.

\[
\begin{align*}
dx &= f(t,x,u)dt + \sigma(t,x)d\omega \\
y &= h(x,\xi)
\end{align*}
\]

(4.60a) (4.60b)

Here, \( x \) is the state; \( u \) is the control; \( \omega \) is a standard Wiener process; \( y \) is the measurement; \( \xi \) is the measurement disturbance, which is modelled as a Gaussian random variable. Assume that the state and the measurement are in some spaces with proper (finite) dimensions respectively. Suppose that \( f \) satisfies the regularity conditions in
t, x and u as well as σ, such that equation (4.60a) admits a unique solution under admissible controls (see Theorem 5.2 in [7] on pp. 230). Consider an integrable cost functional as

$$J(u; x_0, \omega) = \int_{t_0}^{T} g(t, x_t, u_t) dt + q(x_T). \quad (4.61)$$

In (4.61), $g \geq 0$ is the cost rate; $q \geq 0$ is the terminal cost. We refer to the information set as all the information that is available to a decision-maker (controller) up to time $t$, and denote it by $I_t$. Clearly, for the perfect state information case considered earlier, $I_t^{\text{perfect}} = \{x_\tau | t_0 \leq \tau \leq t\}$. In the case where the measurement is imperfect, the information set is $I_t = \{u_{\tau_u}, y_{\tau_y} | t_0 \leq \tau_u < t, \ t_0 \leq \tau_y \leq t\}$. Accordingly, the strategy at time $t$ is a map $\phi : I_t \mapsto U_a$ and denote by $\Phi$ the set of all $\phi$, under which equation (4.60a) admits a unique solution with $u(t) = \phi(I_t)$. The optimal (closed-loop) control problem is to find $\phi$ such that the following objective is minimized.

$$\min_{\phi \in \Phi} \mathbb{E}_{x_0, \omega, \xi} \{J(u; x_0) \mid I_0\}. \quad (4.62)$$

Here, the expectation is taken over $x_0$, $\omega$ and $\xi$ for all $t_0 \leq t \leq T$ given the information $I_0$. In general, the problem in (4.62) is very difficult, and a solution may not exist. Closed-form solutions to this kind of problems are mostly absent except for stochastic problems with linear dynamics, a quadratic objective functional and Gaussian noise, for which the underlying Riccati equations have solutions [7]. In most cases, suboptimal approaches are used, among which is the so-called Open-Loop Feedback Control (OLFC) [54]. OLFC is feedback implementation of open-loop optimal controls. At any sample time $t_0 + k\Delta t > 0$ ($k \in \mathbb{Z}^+$), an open-loop control problem is solved given the information set $I_{t_0 + k\Delta t}$. The open-loop control is implemented immediately afterwards from $t_0 + k\Delta t$ to $t_0 + (k + 1)\Delta t$. We denote by $V_{OL}$ the performance index.
of an open-loop optimal control as

\[
V_{OL}(I_{t_0 + k\Delta t}) \triangleq \min_{u(\cdot) \in U(t_0 + k\Delta t)} \mathbb{E}_{x_{t_0 + k\Delta t}, \omega} \left\{ J(u; t, x_t) \middle| I_{t_0 + k\Delta t} \right\}
\]

\[
= \min_{u(\cdot) \in U(t_0 + k\Delta t)} \mathbb{E}_{x_{t_0 + k\Delta t}, \omega} \left\{ \int_{I_{t_0 + k\Delta t}}^T g(\tau, x_\tau, u_\tau) d\tau + q(x_T) \middle| I_{t_0 + k\Delta t} \right\},
\]

(4.63)

where \( u(\cdot) \in U(t) \triangleq \left\{ \phi : [t, T] \mapsto U_a \middle| \phi(\cdot) \text{ is } \mathcal{F}_{t,T} - \text{progressively measurable} \right\} \) (see Definition 5) for \( t_0 \leq t \leq T \) and \( U_a \) is a compact set. Equation (4.63) is formulated based on an underlying assumption that no more measurement will be taken after time \( t_0 + k\Delta t \).

Now, taking one step further, we consider the control action based on limited look-ahead. Suppose that the current time is \( t_0 + k\Delta t \) and one more measurement \( y \) is allowed at \( t_0 + (k+1)\Delta t \) with \( t_0 + (k+1)\Delta t \leq T \). Clearly, the effect of the future measurement is taken into account in the decision-making at \( t_0 + k\Delta t \). The new optimization problem becomes

\[
V_{LA}(I_{t_0 + k\Delta t}) \triangleq \min_{u(\cdot) \in U(t_0 + k\Delta t)} \left\{ \mathbb{E}_{x_{t_0 + k\Delta t}, \omega} \left[ \int_{I_{t_0 + k\Delta t}}^T g(\tau, x_\tau, u_\tau) d\tau \right] \right\}
\]

\[
+ \mathbb{E}_{I_{t_0 + (k+1)\Delta t}} \left[ V_{OL}(I_{t_0 + (k+1)\Delta t}) \middle| I_{t_0 + k\Delta t} \right]\left| u_{t_0 + k\Delta t, t_0 + (k+1)\Delta t} \right| I_{t_0 + k\Delta t} \right\},
\]

(4.64)

where \( u_{t_0 + k\Delta t, t_0 + (k+1)\Delta t} \) stands for \( u(\tau) \) for \( t_0 + k\Delta t \leq \tau < t_0 + (k+1)\Delta t \). Note that the information set at \( t_0 + (k+1)\Delta t \) is

\[
I_{t_0 + (k+1)\Delta t} = \{ I_{t_0 + k\Delta t}, u_{t_0 + k\Delta t, t_0 + (k+1)\Delta t}, y_{t_0 + (k+1)\Delta t} \}.
\]

Therefore, in (4.64), the expectation taken with respect to \( I_{t_0 + (k+1)\Delta t} \) given \( I_{t_0 + k\Delta t} \) and \( u_{t_0 + k\Delta t, t_0 + (k+1)\Delta t} \) is equivalent to that taken with respect to \( y_{t_0 + (k+1)\Delta t} \).

In the following, we shall show that \( V_{LA} \leq V_{OL} \) at any sample time \( t_0 + k\Delta t \). First of all, we consider a static decision problem with the objective function \( J(u, \omega) \),
in which \( u \in U_a \) is the decision variable and \( \omega \in \Omega \) is a random variable in a finite dimensional space. If the decision-maker only knows the distribution \( p(\omega) \) of \( \omega \), the problem is to minimize the expected cost with a proper \( u \) as

\[
V = \min_{u \in U_a} E_\omega \{ J(u, \omega) \}. \tag{4.65}
\]

Intuitively, additional information to the decision-maker may lead to a better decision. Suppose that the decision-maker has extra knowledge of \( \xi \) in addition to \( p(\omega) \) before he makes the choice. Let \( \xi \in \Psi \) be a random variable and the joint distribution \( p(\omega, \xi) \) is known. A decision function \( \phi \) is a map, \( \phi : \Psi \mapsto U_a \). Denote by \( \Phi \) the set of all possible \( \phi \)'s. The expected cost for the decision-maker becomes

\[
V_\xi = \min_{\phi \in \Phi} E_\xi \{ E_\omega \{ J(\phi(\xi); \omega) \} \} = E_\xi \{ \min_{u \in U_a} E_\omega \{ J(u; \omega) \} \} \tag{4.66}
\]

**Lemma 10** \( V_\xi \leq V \).

Lemma 10 can be shown by using the fact that \( E_\omega \min \{ J(u, \omega) \} \leq \min \{ E_\omega J(u, \omega) \} \) and the details are omitted.

**Theorem 16** \( V_{OL}(I_{t_0 + k\Delta t}) \geq V_{LA}(I_{t_0 + k\Delta t}) \) for any \( k \in \mathbb{Z}_{\geq 0} \).

**Proof:** For a short notation, we use \( k \) to denote \( t_0 + k\Delta t \). By (4.63),

\[
V_{OL}(I_k) = \min_{u(\cdot) \in U(k)} \left\{ E_{x_k, \omega} \left[ \int_k^{k+1} g(\tau, x_\tau, u_\tau) d\tau \right] + \min_{u(\cdot) \in U(k+1)} E_{x_{k+1}, \omega} \left\{ \int_{k+1}^T g(\tau, x_\tau, u_\tau) d\tau + q(x_T) \right\} | I_k \right\}. \tag{4.67}
\]
Consider the measurement $y_{k+1}$ at $t_0 + (k + 1)\Delta t$, and view $y_{k+1}$ as the $\xi$ in Lemma 10. The second term in (4.67) satisfies

$$\min_{u(\cdot) \in \mathcal{U}(k+1)} \mathbb{E}_{x_{k+1}, \omega} \left\{ \int_{k+1}^{T} g(\tau, x_\tau, u_\tau) d\tau + q(x_T) \bigg| I_{k+1}, u_{[k,k+1]} \right\}$$

$$\geq \mathbb{E}_{y_{k+1}} \min_{u(\cdot) \in \mathcal{U}(k+1)} \left\{ \mathbb{E}_{x_{k+1}, \omega} \left[ \int_{k+1}^{T} g(\tau, x_\tau, u_\tau) d\tau + q(x_T) \bigg| I_{k}, u_{[k,k+1]}, y_{k+1} \right] \right\}$$

$$= \mathbb{E}_{y_{k+1}} \left\{ \mathbb{V}_{OL}(I_{k+1}) \bigg| I_{k}, u_{[k,k+1]} \right\}. \quad (4.68)$$

Note that $I_{k+1} = \{I_t, u_{[k,k+1]}, y_{k+1}\}$ and substitute (4.68) into (4.67), we obtain that

$$\mathbb{V}_{OL}(I_{t_0+k\Delta t}) \geq \mathbb{V}_{LA}(I_{t_0+k\Delta t}).$$

Both the OLFC scheme and the approach based limited look-ahead are suboptimal, and they both provide an upper-bound for the optimal Value function. Theorem 16 states that the controls by limited look-ahead provide a better (expected) performance than the controls by OLFC scheme at any sample time $t_0 + k\Delta t$. In other words, at each time $t_0 + k\Delta t (k \in \mathbb{Z}_{\geq 0})$, the problem in (4.64) is solved and the resulting control is exploited over the next $\Delta t$ interval. In this process, the upper-bound of the performance by limited look-ahead is continuously improved and this bound is better than that by the OLFC scheme.

### 4.5.2 On Stochastic Pursuit-Evasion Games

In this section, we consider a stochastic PE game with an imperfect state information pattern. We focus on a special game problem as follows. The scenario corresponds to a worst case from the pursuers’ perspective.
Assumption 6

(A1) The dynamics of both the pursuers and the evaders involve disturbance;

(A2) The pursuers measure the states subject to disturbances, while the evaders can access the perfect state information;

(A3) Information is shared by all the pursuers or the evaders due to sufficient communication on either side.

In the following, we shall use one-step look-ahead to solve a stochastic PE game where Assumption 6 holds. Consider stochastic PE games with the players’ dynamics in (4.3) with (2.9), and the objective functional given in (4.6). Suppose a decision is made at every sample time \( t_0 + k\Delta t \) for \( k \in \mathbb{Z}_{\geq 0} \), where \( t_0 \) is the starting time. Define the information set of the pursuers at time \( t \) for \( t_0 \leq t \leq T \) as

\[
I^p_t = \{ I^p_{t_0}, u_{\tau}, y_{t_0+k\Delta t} \mid t_0 \leq \tau \leq t, t_0 + k\Delta t \leq t \text{ for } k \in \mathbb{Z}^+ \},
\]

where \( I^p_{t_0} \) is the initial information set at \( t_0 \). Denote by \( \mathcal{I}^p_t \) the family of all possible information sets \( I^p_t \) at \( t \). In the approach based on limited look-ahead, given an cost-to-go function \( \tilde{V}_{C2G} : I^p_t \mapsto \mathbb{R} \), the pursuers solve the following optimization problem at each sample time \( t = t_0 + k\Delta t \).

\[
\sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} \left\{ E_{x_t,\omega} \left( \int_t^{t+\Delta t} G(\tau, x_\tau, z_\tau, a_\tau, \beta[\cdot]_\tau) \right) + E_{y_{t+\Delta t}} \tilde{V}_{C2G}(I^p_{t+\Delta t}) \right\}
\]

(4.69)

Here, the cost-to-go term is crucial to the performance. However, it is very difficult to determine an optimal cost-to-go function, which is essentially the true (upper/lower) Value function. Therefore, in practice, a heuristic or suboptimal cost-to-go \( \tilde{V}_{C2G} \) is often used instead in stochastic problems with imperfect state information patterns. For example, the performance index associated with open-loop strategies of the players can be used. However, in a game where the terminal time is not fixed, e.g., a
PE game, open-loop strategies may not exist. In this case, an alternative cost-to-go function needed to be selected. For instance, the cost-to-go function can be

\[
\tilde{\mathcal{V}}_{C2G}(I^p_t) = \mathbb{E}_{x_t}\left\{\tilde{\mathcal{V}}(x_t) \mid I^p_t\right\}.
\]

(4.70)

In (4.70), \( \tilde{\mathcal{V}} \) is an approximate upper Value function determined (possibly by the hierarchical approach) for corresponding stochastic problems with perfect state information patterns. To further reduce the computational complexity in (4.70), certainty equivalence can be used.

Generally speaking, a stochastic game (control) problem with imperfect state information is very hard, and the theoretical results in this area are still largely unavailable. In this section, only a special case of this type of problems is discussed. More general results entail further fundamental study.

4.5.3 On Finite Expectation of the Capture Time

This section is a counterpart of Section 4.3.3 and the theme is the capture time in a two-player stochastic PE game with an imperfect state information pattern. Particularly, we study conditions on the accuracy of the pursuer’s measurement, under which the capture time of the evader has a finite expectation. Consider a two-player PE game in a \( \mathbb{R}^2 \) space with the players’ dynamics given in (4.27) and the objective as the capture time. Recall that in (4.27), \( \bar{x}_\varsigma \) and \( \bar{y}_\varsigma \) are used to denote the state variables \( (\varsigma \in \{p, e\}) \), and also \( x_\varsigma \triangleq [\bar{x}_\varsigma, \bar{y}_\varsigma]^T \). We assume that the evader can measure the states \( (x_p, x_e) \) perfectly; while the pursuer knows its own state \( x_p \) but can only access a noisy measurement of \( x_e \). The study here is based on the analysis for the perfect state information case, but taking into account the effect of the measurement disturbance.

112
Since the evader can access the perfect states, we assume that the evader exploits the optimal control $\theta^*_e$ determined in (4.32). Suppose that the pursuer’s control $\tilde{\theta}_p$ is based on its noisy measurement of $x_e$. Let $\tilde{x} \triangleq \bar{x}_p - x_e$ and $\tilde{y} \triangleq \bar{y}_p - y_e$. Define $\rho = \sqrt{\tilde{x}^2 + \tilde{y}^2}$. According to (4.21), the evolution of $\rho$ with controls $\theta^*_e$ and $\tilde{\theta}_p$ can be described as follows, similarly as in (4.56).

$$\dot{\rho} = \left( \frac{\partial \rho}{\partial \bar{x}_p} v_p \cos \tilde{\theta}_p + \frac{\partial \rho}{\partial \bar{y}_p} v_p \sin \tilde{\theta}_p \right) + v_e + \frac{1}{2\rho} \sigma^2$$

(4.71)

Note that equation (4.71) reduces to (4.56) when the pursuer can measure $x_e$ perfectly and employs optimal $\theta^*_p$ determined in (4.32). The following analysis is conducted based on a specific model of the pursuer’s measurement and the result reveals how the measurement accuracy can affect the capturability of the evader.

Consider the following measurement equation.

$$y_e = x_e + \xi_e(x_e)$$

(4.72)

Here, $y_e$ is the measurement; $\xi_e(x_e)$ is a random vector representing the measurement disturbance, which may be state dependent. Suppose that the pursuer calculates its optimal control according to (4.32) but with the perfect state $x_e$ replaced by its measurement $y_e$. Assume that the $\xi_e$ is bounded, and we shall provide a bound $\Xi_e(\rho)$ of $\xi_e(x)$ that may depend on $\rho \triangleq \sqrt{\tilde{x}^2 + \tilde{y}^2}$, i.e., $||\xi_e(x)|| \leq \Xi_e(\rho)$, under which $\bar{\rho}$ has a positive decreasing rate. Here, $\bar{\rho} = E\{\rho\}$.

We consider a coordinate system whose origin is the current position of the pursuer. Denote by $\vartheta(x)$ the angle between the line of sight from the evader to the pursuer and the $\bar{x}$-axis as illustrated in Figure 4.4. Define $\gamma(\rho) = \cos^{-1} \left( \left( \frac{v_e + \sigma^2}{2\rho} / \sigma_p \right) / v_p \right)$, where $\sigma^2 = \sigma^2_p + \sigma^2_e$. Note that $\varepsilon > (\sigma^2_p + \sigma^2_e) / 2(v_p - v_e)$ in Theorem 15 (as in Theorem
Figure 4.4: Illustration of the Pursuer’s Control Based on a Noisy Measurement

13). Hence, $\gamma(\rho)$ is well defined when $\rho \geq \varepsilon$, and clearly $\gamma(\rho) < \pi/2$. Define the set $\mathcal{D}_\varepsilon \triangleq \{\delta|\gamma(\varepsilon) > \delta > 0\}$. Let $\gamma_\delta(\rho) = \gamma(\rho) - \delta$ for some $\delta \in \mathcal{D}_\varepsilon$.

**Proposition 1** If there exists a $\delta \in \mathcal{D}_\varepsilon$ such that $\Xi_e(\rho) \leq \rho \cdot \tan(\gamma_\delta(\rho))$ for $\rho \geq \varepsilon$ at any $x_e$ and any time $t \geq 0$, then $\dot{\rho}(t) < -\kappa^2$ at any $t \geq 0$ for some $\kappa \neq 0$.

**Proof**: Since $\|\xi_e(x)\| \leq \Xi_e(\rho)$, according to Figure 4.4, clearly, $y_e$ falls within a region $Y_e$, which is a circle of radius $\Xi_e(\rho)$ centered at $x_e$. Let $\tilde{x} = \bar{x}_p - x_{e1}$ and $\tilde{y} = \bar{y}_p - y_{e2}$. By (4.32), $\tilde{\theta}_p$ satisfies

$$
\cos \tilde{\theta}_p = -\frac{\tilde{x}}{\sqrt{\tilde{x}^2 + \tilde{y}^2}} \quad \text{and} \quad \sin \tilde{\theta}_p = -\frac{\tilde{y}}{\sqrt{\tilde{x}^2 + \tilde{y}^2}}.
$$

Namely, $\tilde{\theta}_p \in \Theta^\delta_x \triangleq \{\theta|\theta(x) + \pi - \gamma_\delta(\rho) \leq \theta \leq \vartheta(x) + \pi + \gamma_\delta(\rho)\}$. Under $\tilde{\theta}_p$, equation (4.71) becomes

$$
\dot{\rho} = \left(\frac{\partial \rho}{\partial \bar{x}_p} v_p \cos \tilde{\theta}_p + \frac{\partial \rho}{\partial \bar{y}_p} v_p \sin \tilde{\theta}_p\right) + v_e + \frac{1}{2\rho} \sigma^2. \quad (4.73)
$$

Note that in (4.73), $\frac{\partial \rho}{\partial \bar{x}_p} = \cos \vartheta(x)$ and $\frac{\partial \rho}{\partial \bar{y}_p} = \sin \vartheta(x)$. Thus, $\frac{\partial \rho}{\partial \bar{x}_p} \cos \tilde{\theta}_p + \frac{\partial \rho}{\partial \bar{y}_p} \sin \tilde{\theta}_p = \cos(\vartheta(x) - \tilde{\theta}_p)$. The fact that $\tilde{\theta}_p \in \Theta^\delta_x$ implies $|\tilde{\theta}_p - \vartheta(x) - \pi| \leq |\gamma_\delta(\rho)| < \pi/2$, i.e.,
\[
\cos(\bar{\theta}_p - \bar{\theta}(x) - \pi) \geq \cos(\gamma_\delta(\rho)).
\]
Thus,
\[
\cos(\bar{\theta}(x) - \bar{\theta}_p) = -\cos(\bar{\theta}(x) - \bar{\theta}_p - \pi) \leq -\cos(\gamma_\delta(\rho)) = \cos(\gamma(\rho) - \delta).
\] (4.74)

Substitute (4.74) into (4.73),
\[
\dot{\rho} \leq -v_p \cos(\gamma(\rho) - \delta) + v_e + \frac{1}{2\rho} \sigma^2 = -v_p \left[ \cos(\gamma(\rho) - \delta) - \left( v_e + \frac{1}{2\rho} \right) v_p \right]
\]
\[
= -v_p \left[ \cos(\gamma(\rho) - \delta) - \cos(\gamma(\rho)) \right] = -v_p \sin(\gamma(\rho) - \delta/2) \sin(\delta/2)
\]
\[
\leq -v_p \sin(\gamma(\varepsilon) - \delta/2) \sin(\delta/2) = -\kappa^2 < 0.
\] (4.75)

Since (4.75) holds for any \(\rho > \varepsilon\), thus it holds for \(\dot{\rho}\) if the evader is not captured, i.e.,
\[
\dot{\rho} < -\kappa^2.
\]

Proposition 1 states that if \(\|\xi_e\| \leq \rho \cdot \tan(\gamma_\delta(\rho))\), the distance between the pursuer and the evader decreases on average. Following the analysis in Section 4.3.3, it is expected that the capture time of the evader has a finite expectation.

Note that the measurement with a bounded disturbance considered in (4.72) is restrictive, and it is unlikely that the pursuer’s measurement has a bounded disturbance in a real-world problem. However, the assertion in Proposition 1 still has its practical value because it sheds light on the relationship between the measurement accuracy and the capturability as well as the capture range of the pursuer (indicated by \(\varepsilon\)). The results can be extended to a more general case where the measurement noise is unbounded but with a known probability distribution.

In this section, some fundamental issues of stochastic PE differential games have been discussed. The limited look-ahead approach is extended to stochastic games with both the perfect and the imperfect state information pattern. Most of the study is devoted to the perfect state information case. Problems with imperfect state
information needs further study on some fundamental issues. Except for some special cases, analytical solutions for most problems are still largely unavailable. Bayesian analysis used in Theorem 16 can be further extended to more general situations, e.g., in a situation when (useful) measurements are absent. In such a problem, a PE game can be used to model a search problem based on the probability maps of the detected and the hidden evaders [84].

4.6 Simulation Results

In this section, we implement the method based on limited look-ahead for stochastic PE games with an imperfect state information pattern. Consider a cooperative pursuit problem of multiple evaders in a $\mathbb{R}^2$ space. Suppose that there are 3 pursuers and 4 evaders, and the dynamics of the players are given in (4.27). The sum of the capture time of each evader is chosen as the objective. In addition to assumption 6, we also assume that the pursuers can measure their own states perfectly. The measurement equation (for evader $j$) of the pursuers is

$$y^j(t) = \begin{bmatrix} \bar{x}_j \\ \bar{y}_j \end{bmatrix} + \begin{bmatrix} \xi^j_x \\ \xi^j_y \end{bmatrix}, \tag{4.76}$$

where $\xi^j_x$ and $\xi^j_y$ are independent Gaussian distributed random variables, with zero-mean and a common variance, i.e., $\xi^j_x(\xi^j_y) \sim \mathcal{N}(0, \sigma^2_j)$. The necessary parameters of the pursuers and the evaders are given in Table 4.1. Let the variance of the measurement disturbance $\sigma^2_j$ be 0.5 for $j = 1, \cdots, 4$.

Since the dynamics in (4.27) is linear in the states and the disturbances both in the dynamics and in the measurement are Gaussian, the Kalman filter is applicable. The assumption here is that the evaders’ controls can be measured perfectly by the pursuers. We solve the stochastic game by the limited look-ahead method as in (4.69).
Suppose that the current time is \( t \). The cost-to-go function at time \( t + \Delta t \) is similar to (4.70) but with the state estimate used instead with the principle of certainty equivalence, i.e.,
\[
\tilde{V}_I(t+\Delta t) = \tilde{V}(\hat{x}_{t+\Delta t|I_p})
\]
where \( \hat{x}_{t+\Delta t|I_p} = \mathbb{E}\{\hat{x}_{t+\Delta t|I_p}\} \). Here, function \( \tilde{V} \) is determined by the hierarchical method for a corresponding problem with perfect state information.

The details of the simulation are introduced as follows. Note that the pursuers and the evaders have independent dynamics; thus, the dynamic equation is separable in the sense of Theorem 9. This property is useful to decouple the decision-making of the pursuers and that of the evaders. Since \( \tilde{V} \) results from the hierarchical method, we assume that under an engagement scheme between the pursuers and the evaders, each evader still exploits the optimal strategy that is calculated according to (4.32) in the perfect state information case against its engaged pursuer. The strategy of the evaders is optimal (against the pursuers) if the pursuers’ observation does not depend on the evaders’ control. With the evaders’ strategy obtained, a PE game

<table>
<thead>
<tr>
<th>Pursuers</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( (x_{p0}, y_{p0}) )</td>
<td>(0,-12)</td>
<td>(8,8)</td>
<td>(-8,8)</td>
</tr>
<tr>
<td>( v_p (1/sec) )</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>( \sigma^2_p )</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Evaders</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( (x_{e0}, y_{e0}) )</td>
<td>(0.6)</td>
<td>(-3,-3)</td>
<td>(-3,-3)</td>
<td>(0,0)</td>
</tr>
<tr>
<td>( v_e (1/sec) )</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>( \sigma^2_e )</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Table 4.1: The Necessary Parameters of the Pursuers and the Evaders in the Selected Stochastic PE Game Scenario

117
can be transformed into a stochastic optimal control problem with an imperfect state information pattern from the pursuers' perspective. The approach based on limited look-ahead is implemented in the following simulations.

Three games are simulated, which involve all four evaders and one, two or three pursuers respectively, i.e., the game between all the evaders and pursuer 1, pursuers (1,2) and pursuers (1, 2, 3). In the simulations, stochastic differential equations are simulated by using the Euler’s method with the time interval $\Delta t$ chosen as 0.1. This $\Delta t$ is also used as the size of limited look-ahead intervals. Since $\Delta t$ is small, the dynamic optimization problem based on limited look-ahead in (4.69) can be approximated by a static optimization problem with the controls of the players fixed during $\Delta t$ intervals.

The game involving four evaders and all three pursuers is illustrated in Figure 4.5 at various stages. The trajectories of the players are plotted and the circles in the stages 3 and 4 indicate the pursuers’ capture range as well as the capture of the corresponding evaders inside. An interesting phenomenon can be observed from Figure 4.5, as well as from the following figures, that the pursuers move cooperatively at the beginning stages to force the evaders to remain in a certain region. When the pursuers get close enough, they become engaged with specific evaders. This is because that according to the objective in (2.44), all the evaders are equally important. Thus, the evaders tend to avoid potential capture by each of the pursuers. When the pursuers move closer cooperatively, the evaders wander for a while trying to avoid each pursuer. From this aspect, the method based on limited look-ahead has an advantageous potential for concealing the true intent of the pursuers. Finally, Figure 4.5 only provides a sample run of the stochastic process.
Figure 4.5: Cooperative Pursuit Trajectories of All 4 Evaders by the 3 Pursuers

For comparison purposes, we simulate similar games with the pursuers 1 and 2 and pursuer 1 respectively. They share the same parameters and the necessary initial conditions with the game involving all three pursuers. The results are shown in Figure 4.6 and Figure 4.7 respectively.

Figure 4.6 illustrates the trajectories of the cooperative pursuit of the evaders by the pursuers 1 and 2. It is worth noting that pursuer 2 on the right is directly engaged with evader 1 (the fastest) at the top from the beginning, while pursuer 1 twists a little at the beginning stages and then goes after the rest of the evaders
systematically. In Figure 4.7, the cooperative pursuit trajectories between all the evaders and only pursuer 1 are illustrated. It can be seen that pursuer 1 goes after the 4 evaders systematically.

In the simulations above, the Kalman filter is used by the pursuers to estimate the evaders’ states. It requires the perfect knowledge of the evaders’ controls, which may be unrealistic. In the following, we relax this assumption and assume that the evaders’ controls can be measured subject to disturbance. Namely, for each evader $j$ ($j = 1, \cdots, 4$), the pursuers’ measurement of $\theta^j_e$ (evader $j$’s control) is based on
\[ \hat{\theta}_j^t = \theta_j^t + \eta_j, \] where \( \eta_j \) is a zero-mean Gaussian distributed random variable with the variance \( \sigma^2_{\theta_j^t} \). We choose \( \sigma^2_{\theta_j^t} = 0.5 \) for all \( j = 1, \cdots, 4 \), and the cooperative pursuit trajectories of all 4 evaders by the 3 pursuers are illustrated in Figure 4.8. With the noise-corrupted measurements of the controls used in the state estimator, it is expected that the pursuers’ performance is degraded.

From these results, one appealing factor of the limited look-ahead method is that it has a potential for concealing the true intent of the pursuers compared to the hierarchical approach. The reason is that the optimization at the upper level based
Figure 4.8: Cooperative Pursuit Trajectories of the 4 Evaders by the 3 Pursuers with the Imperfect Observation of the Evaders’ Strategies

on (2.44) only provides structured controls to the pursuers, where each pursuer has specific engaged evaders to go after; whereas in the optimization based on limited look-ahead in (4.11) as well as in (4.69), such a control structure is relaxed and the cooperative movement of the pursuers may result in a better outcome if the evaders are not aware of the pursuers’ strategy [85].
CHAPTER 5

DISSERTATION SUMMARY AND FUTURE WORK

5.1 Dissertation Summary

The increasing use of autonomous assets and robots in modern military operations has led to renewed interest in the subject of (multi-player) PE games. However, the current theory of differential games is inadequate for solving general multi-player PE differential games. This is the first attempt to solve general multi-player PE games in continuous time. The purpose of this dissertation is to extend the current differential game theory to the general multi-player PE game case. We approach the problem in an indirect way using an iterative method. The basic idea is to improve a suboptimal solution iteratively based on limited look-ahead, such that an optimal solution is approached in the limit. This approach shares insights with DP-type methods such as the rollout algorithm [54] and the value iteration method [51], which mainly deal with discrete-time problems. However, a multi-player PE differential game is more challenging because it involves multiple players and it is in continuous time. The major contribution of this dissertation is to provide a theoretical foundation for general multi-player PE games under the framework of the current differential
We adopt the concept of nonanticipative strategy and the (upper/lower) Value function due to Varaiya, Roxin and Elliott and Kalton in the game formulation, because this notion of Value can be characterized by a solution of the corresponding HJI equation in a general sense. Due to the multiplicity of players, the information about the “alive” evaders is naturally modelled by a discrete variable $z$. In a multiplayer PE game, the major difficulty of the current differential game techniques based on state rollback lies in the fact that the terminal states of the game cannot be easily specified. To circumvent this dilemma, we approach the multi-player PE problem in an indirect manner starting with a suboptimal solution by only considering a subset of the pursuers’ control inputs, which are called “structured” controls. A major breakthrough results from the improving property of the suboptimal solution that is determined by a ST-consistent structured control. The optimization based on limited look-ahead can be utilized to improve such a suboptimal solution. If it is applied iteratively, the resulting sequence of functions converges to a true upper Value function. Furthermore, if the functions in the sequence are continuous, the convergence can be achieved in a finite number of steps.

With this iterative method, a multi-player PE game reduces to a problem of finding a valid suboptimal solution with the improving property. Specifically, we show that the method based on hierarchical decomposition is applicable. With the hierarchical structure of STC, a valid initial point for the iterative process can be determined. Furthermore, the performance enhancement based on limited look-ahead is also applicable to many other complex dynamic optimization problems.
Uncertainty is inherent in real-world problems, especially in military applications when information sources are limited. For this reason, we extend our discussion to the stochastic PE game case. When uncertainties only appear in the players’ dynamics, the theoretical results of the iterative method are largely valid for stochastic problems. Under the framework of the hierarchical approach, the two-player stochastic PE game case is extensively examined. In particular, we derive an analytical solution to a two-player stochastic PE game with the players’ dynamics as the Dubin’s car model, and also investigate conditions under which the capture time has a finite expectation. In the case when the states are only measured subject to disturbance, a worst-case analysis is conducted from the pursuers’ perspective. Particularly, the pursuers’s measurement is assume to be noise-corrupted while the evaders can still access the perfect states. A suboptimal approach based on one-step look-ahead is studied. This method provides a better upper-bound of the performance index compared to the well-known suboptimal approach—OLFC.

In addition to the numerical justification of the iterative method based on look-ahead, we examine the theoretical soundness of the method for deterministic PE games, under the framework of viscosity solution theory for HJ equations. Conditions on the existence of the Value function of a multi-player PE game are derived in terms of the players’ dynamics and the objective functional. Besides, continuity of the suboptimal Value functions are studied.

Due to the close connection to the DP principle, the iterative method suffers from the “curse of dimensionality”. However, numerical methods of DP equations have been extensively studied over the years, and they can benefit the iterative approach to multi-player PE games in many ways. From a practical point of view, the iterative
process can stop at any step and provide the best upper-bound of the upper Value function due to the monotonicity of the sequence. Moreover, the suboptimal strategy that results from one-step look-ahead can usually perform well in many practical problems. We have demonstrated the practical usefulness and the advantage of the limited look-ahead through several selected multi-player PE game scenarios.

The fundamental idea behind the iterative approach introduced in this dissertation for multi-player PE games is attractive for general complex systems. When an optimal solution cannot be obtained directly, an alternative approach is usually to search for an approximate solution and the possibility of serial improvements based on that. The improvement can be systematic, as in the iterative method here, or random, as in genetic algorithm\textsuperscript{24} etc. It is often expected that an exact solution can be approached in the long term.

## 5.2 Future Work

We suggest several potential research directions both in theory and for practical applications. The first is to generalize the iterative method to the nonzero-sum game case, which can advance the theory of stochastic PE games. The second direction is on efficient algorithms, where an interesting observation of the optimization based on one-step limited look-ahead is discussed. A third is about decentralized approaches, which have better adaptability and robustness under most realistic circumstances.

\textsuperscript{24}A genetic algorithm is a numerical optimization (or search) technique that is to find true or approximate solutions to optimization (or search) problems.
5.2.1 The Iterative Method for Nonzero-sum Games

The iterative method studied is for zero-sum game problems. Here, an underlying assumption is that the information available for the decision-making of each player is a common knowledge [86], i.e., the information is shared by all the players and the fact that it is known is also known by all. However, in a general stochastic multi-player PE game with an imperfect state information pattern, the pursuers and the evaders actually access different sets of information because either the pursuers or the evaders have an independent measurement process. Therefore, the PE game is nonzero-sum for the following reason. Let $I_p^t$ and $I_e^t$ denote respectively the information sets of the pursuers and of the evaders at time $t$, and clearly $I_p^t \neq I_e^t$. Denote by $\mathcal{J}(a, b; x, z)$ the objective function associated with the corresponding deterministic game under the controls $a$ and $b$ at the states $x$ and $z$. Thus, the multi-player PE game problem can be described as

$$
\begin{align*}
\min_a \mathbb{E}\left( \mathcal{J}(a, b; x, z) \big| I_p^t \right) \\
\max_b \mathbb{E}\left( \mathcal{J}(a, b; x, z) \big| I_e^t \right)
\end{align*}
$$

(5.1)

Since $I_p^t \neq I_e^t$, the objective of the pursuers and that of the evaders are no longer opposite. Therefore, the game becomes a nonzero-sum game.

To address a stochastic game with an imperfect state information pattern in (5.1), we need to answer the question whether the iterative method can be extended to the nonzero-sum game case. In a general nonzero-sum game, there are $N$ ($N > 1$) players and each has an individual objective. Nash equilibrium is the solution concept. In contrast to a zero-sum game that has only one associated Value function, a nonzero-sum game problem is more complicated, where each player has its own Value evaluated at Nash equilibrium. Hence, in an iterative method for a nonzero-sum game (if any),

127
there are $N$ different sequences of suboptimal Value functions. Two issues need to be addressed: 1) the monotone property of the sequence in zero-sum game is no longer present, so is the property of convergence; 2) there is no obvious control structure or valid starting point. A relevant study can be found in [87].

5.2.2 On Efficient Algorithms

The iterative method is not scalable and efficient computation algorithms are needed. One possible remedy may resort to the methods in Approximate (or Neuro-) DP, such as functional approximation, temporal difference or Q-learning and simulation-based optimization etc. [76, 74, 52]. On the other hand, suboptimal approaches may be attractive in practice. In the following, we focus on a suboptimal approach based on one-step look-ahead. This is a brief introduction, and more details can be found in [85].

Consider a deterministic PE game involving $N$ pursuers and $M$ evaders with the objective functional in (2.44) and the players’ dynamics in (2.12). The following discussion is based on a specific $z \in Z^M$, i.e., $z$ does not change, so that $z$ is suppressed in the formulations. Let $x$ denote the composite state variable that contains the states of all the pursuers and the evaders. We adopt the definition of strategy given in Chapter 2. Let $x_{\tau;x,t,a,b}$ be the state $x_{\tau}$ at time $\tau$ for $t \leq \tau \leq t + \Delta t$ starting from $x_t$ under the controls $a(\cdot)$ and $b(\cdot)$ for some $\Delta t > 0$. Define

$$X_t = \{x_{\tau;x,t,a,b} \text{ for } a(\cdot) \in A(t), b(\cdot) \in B(t) \text{ and } t \leq \tau \leq t + \Delta t\}.$$  

Note that the dynamic in (2.12) is bounded, and so is the set $X_t$. Let $E^*(x)$ be the optimal engagement between the pursuers and the evaders at state $x$, i.e., $E^*(x)$ is the optimal solution of (2.46). Given a $\Delta t > 0$, suppose that there is no transition
of $z$ that occurs during $t$ to $t + \Delta t$ under any $a(\cdot) \in A(t)$ and $b(\cdot) \in B(t)$. Denote by $\tilde{V}^h(x)$\textsuperscript{25} the optimal performance index by the hierarchical approach according to (2.46). Let $\tilde{V}^E(x; E_0)$ be the performance index under an arbitrary engagement $E_0$. Clearly, $\tilde{V}^h(x) = \tilde{V}^E(x; E^*(x))$.

**Proposition 2** If $E^*(x_t) = E^*(x)$ for any $x \in \mathcal{X}_t$, then

$$
\tilde{V}^h(x_t) = \sup_{\beta \in \Delta(t)} \inf_{a(\cdot) \in A(t)} \left\{ \int_t^{t+\Delta t} M d\tau + \tilde{V}^h(x_{t+\Delta t}; x_t, a, \beta[a]) \right\}. \tag{5.2}
$$

Proposition 2 is easy to prove [85]. Define the set of all possible engagements over $\mathcal{X}_t$ as $\overline{E}_{xt} = \{ E^*(x) | x \in \mathcal{X}_t \}$. Proposition 2 implies that if $|\overline{E}_{xt}| = 1$ (the cardinal number), the optimal hierarchical structured control of the pursuers determined at $t$ is optimal. In other words, the optimization in (5.2) may result in a better performance (than the hierarchical approach) only if there exists $x \in \mathcal{X}_t$ such that $\tilde{V}^h(x) < \tilde{V}^E(x; E^*(x_t))$.

Here, we say that the state $x_t$ is in a cooperation region (of the pursuers) if $|\overline{E}_{xt}| > 1$.

---

\textsuperscript{25} Note that the argument $z$ is suppressed because no transition of $z$ occurs during $t$ to $t + \Delta t$.

\textsuperscript{26} Note that $G(x, z, a, b) = \sum_{j=1}^{M} z(\tau) = M$ if no transition of $z$ occurs.
Now, we assume that \(|E_{xi}| = m > 1\). Note that not every pursuer has \(m\) different engaged evaders based on the engagements in \(E_{xi}\). Suppose that pursuer \(i\) has two possible engaged evaders \(e_i^1\) and \(e_i^2\), as shown in Figure 5.1. Consider that the players’ dynamics are given in (3.8). According to (3.13) in Chapter 3, the optimal strategy of the pursuer in a two-player PE game is to move towards the evader. Thus, under the hierarchical structure, pursuer \(i\) has two potential orientations at time \(t\), as illustrated by the speed vector \(v_{p1}^i\) and \(v_{p2}^i\) in Figure 5.1. It is hard to believe that an optimal solution lies in this set with only two possible alternatives. Nevertheless, in the optimization over the look-ahead interval \(\Delta t\) in (5.2), the structure is relaxed such that all possible controls of pursuer \(i\) are considered. Intuitively, a better solution may exist within the relaxed control set. Let the arc \(\hat{12}\) in Figure 5.1 denote the control space spanned by structured controls \(v_{p1}^i\) and \(v_{p2}^i\), where the angle between the two vectors is no larger than \(\pi\). If \(|E_{xi}| = 2\), either \(v_{p1}^i\) or \(v_{p2}^i\) may not be optimal and pursuer \(i\) should account for both evader \(e_i^1\) and \(e_i^2\) simultaneously when making a decision. With the objective function given in (2.44) and by inspection of Figure 5.1, the following assertion is true.

**Claim 3** Any strategy \(v^i\) of pursuer \(i\) that falls outside of the region \(\hat{12}\) is dominated by \(v_{p1}^i\) or \(v_{p2}^i\) with respect to the pursuit of both \(e_i^1\) and \(e_i^2\). This is true in a general case when pursuer \(i\) \((i = 1, \cdots, N)\) has \(m\) engaged evaders according to the engagements in \(E_{xi}\).

In (3.8), \(\theta_p^i\) is the control of pursuer \(i\). Let \(\theta_{pk}^i\) be the \(k^{th}\) structured control of pursuer \(i\) for the \(k^{th}\) engagement in \(E_{xi}\). By Claim 3, the optimal control \(\theta_{p*}^i\) of pursuer \(i\) from the optimization problem based on one-step look-ahead in (5.2) can be written
in the following form.

\[
\theta^{i*}_p = \sum_{k=1}^{m} \lambda^i_k \theta^i_{pk} \text{ with } \lambda^i_k \geq 0 \text{ and } \sum_{k=1}^{m} \lambda^i_k = 1.
\]  

(5.3)

The analysis above on the pursuers is also valid for solving the optimal strategies of the evaders, i.e., \(\theta^{i*}_c\) for each evader \(j\) can also be written as linear combination of the structured controls with non-negative real coefficients that sum to 1.

This preliminary investigation reveals the fact that the performance enhancement by limited look-ahead based on the structured approach essentially results from the structure relaxation in the optimization over look-ahead intervals. This relaxation on the control structure enables a better coordination among the pursuers. More importantly, the linear structure of the improving strategy specified in (5.3) can be useful to reduce the computation. With proper approximation, the optimization problem in (5.2) may be formulated as a nonlinear programming problem, such that extensive numerical methods in the literature [2] are applicable.

### 5.2.3 On Decentralized Approaches

In this section, we point out the future research on decentralized methods. In general, a closed-form solution of a multi-player PE game is formidable and it relies on numerical calculation in most cases. However, most DP-type methods are computationally intensive, such that they may not be suitable for real-time implementation. On the other hand, in practical problems like military applications, uncertainty is usually an important factor, which entails on-line decisions. Moreover, considering possible failures in communication among the pursuers (evaders), the underlying assumption of the global information sharing among the pursuers (evaders) is invalid.
Thus, from a realistic point of view, a computationally efficient and scalable algorithm that is robust against practical uncertainties is in great need.

In contrast to a centralized method, a decentralized approach is more attractive for the decision-making in large-scale systems involving uncertainties. By “divide and conquer”, a decentralized approach has better adaptability, scalability and computational efficiency but with sacrifice on optimality. For the practical issues mentioned above, decentralized methods are certainly an important future research direction for multi-player PE games. Note that the hierarchical approach can be viewed as a preliminary decentralized approach. However, it is simple and not suitable for highly uncertain problems, e.g., some pursuers might be lost (destroyed in the battle-field) or new evaders may emerge before the end of the game.

A more sophisticated decentralized method is needed, where although decisions are made individually by each pursuer (evader) or by a smaller group of the pursuers (evaders) with local information, the coordination with other pursuers (evaders) should be taken into account. The ideal case is that a global optimum with respect to the original game objective criterion can be preserved by local optimization. From this aspect, team theory can play a role [88]. Another possible direction of centralized methods is to design a set of local negotiation rules among the pursuers (evaders) such that local decisions may lead to a desirable behavior at the group level. This idea is from the filed of multi-agent systems [89]. With the interconnections between the players, additional cooperation is possible compared to the hierarchical approach. Furthermore, cooperative game theory may be useful to determine a proper cooperation topology (“coalition”) among the pursuers (evaders) [86], with which centralized methods can be implemented on smaller groups. Intuitively, cooperative behavior
requires information sharing among the (cooperative) players. Thus, in decentralized approaches, the inter-player communication becomes an important issue. The information structure, communication efficiency and robustness are all of great importance.

Finally, throughout the study in this dissertation, we have encountered tremendous challenges in stochastic games. Not only the calculation becomes more difficult, but also the uncertainty raises fundamental problems on the information pattern in control as well as differential game problems. More additional assumptions have to be introduced in order to make the problem well defined and solvable. This type of problems have trapped mathematicians and engineers for decades and this area is still widely open. We wish some day more fundamental machineries are available to make a major breakthrough.
Proof: Suppose that the game starts at time $t_0$. For any $z \in Z^M$, $x \in X$, denote by $E_i$ the set of evaders that have been engaged with pursuer $i$ under an engagement $E$. Let $E^{(x,z)}$ be the optimal engagements at $x$ and $z$ according to (2.46). Denote by $J_i^h(x, z; E_i)$ the performance evaluation of the sub-game between pursuer $i$ and the evaders in $E_i$ with the optimal strategies exploited by pursuer $i$ and the evaders at the lower level. Clearly, $J_i^h = 0$ if $E_i = \emptyset$. If $|E_i| = n_i > 0$, let $E_i = \{j_1, \cdots, j_{n_i}\}$, where $j_k = 1, \cdots, M$. We say $E$ is feasible if $J^h(x, z; E) < \infty$. Under Assumption 3, there exists some $E$ that is feasible.

Note that

$$J^h(x, z; E) = \sum_{i=1}^{N} J_i^h(x, z; E_i).$$

Lemma 5 indicates that if $J_i^h(x, z; E_i)$ is uniformly continuous in $x^p$ and $x^j$ ($j \in E_i$) under any feasible $E$, then $J^h(x, z; E)$ is uniformly continuous in $x$ and so is $\tilde{V}^h(x, z)$ because there are finitely many $E'$s and

$$\tilde{V}^h(x, z) = \min_E \left\{ J^h(x, z; E) \right\} = \min_E \left\{ \sum_{j=1}^{N} J_j^h(x, z; E) \right\}.$$

In the following, we need to show the uniform continuity of $J_i^h(x, z; E_i)$ for each $E$. Given an arbitrary $z \in Z^M$, suppose that $E_i = \{1, \cdots, n_i\}$ for $1 \leq n_i \leq M$. Let
$x_1, x_2$ be arbitrary two point in $\Omega$. We denote by $x^1_p(\cdot)$ and $x^2_p(\cdot)$ be the trajectories of pursuer $i$ starting form $x^1_p$ and $x^2_p$ respectively under a common strategy that will be introduced shortly. According to the problem formulated in (2.46), pursuer $i$ proceeds to the evaders in $E_i$ sequentially, therefore,

$$J^h_i(x, z, E_i) = \sum_{j=1}^{n_i} \mathcal{V}_{ij}(x_{ij}[j]). \quad (A.1)$$

By (A.1), the uniform continuity of $J^h_i(x, z, E_i)$ depends on that of $\mathcal{V}_{ij}(x_{ij}[j])$ for every $j = 1, \cdots, n_i$. In the following, we shall show the uniform continuity of $\mathcal{V}_{ij}(x_{ij}[j])$ for each $j$.

For $x_l \in \Omega \ (l = 1, 2)$, denote by $x^i_{lp}$ and $x^j_{le}$ the state of pursuer $i$ and evader $j$ in $E_i$ associated with $x_l$. Also, define $x^i_{ij} \triangleq [x^i_{lp}^T, x^j_{le}^T]^T$. Denote by $J_{ij}(a, b; x_{ij})$ the objective functional of the two-player game between pursuer $i$ and evader $j$ in $E_i$, where $a_i$ and $b_j$ are the corresponding control inputs. Let $\varepsilon > 0$ be arbitrary. First of all, consider $\mathcal{V}_{i1}$, i.e., the upper Value of the sub-game between pursuer $i$ and the first evader in $E_i$. Recall that

$$\mathcal{V}_{i1}(x^2_{i1}) = \sup_{\beta_1 \in \Delta_1(t_0)} \inf_{a_i(\cdot) \in \mathcal{A}^i(t_0)} \{J_{i1}(a_i, b_1; x^2_{i1})\}. \quad (A.2)$$

For the given $\varepsilon > 0$, there exists a strategy of evader 1 , i.e., $\beta^2_1 \in \Delta_1(t)$\textsuperscript{27}, such that

$$\mathcal{V}_{i1}(x^2_{i1}) \leq J_{i1}(a^2_i, \beta^2_1; x^2_{i1}) + \varepsilon \text{ for any } a^2_i(\cdot) \in \mathcal{A}^i(t), \quad (A.3)$$

where $a^2_i(\cdot)$ is the control of pursuer $i$ starting from $x_2$. Now, consider the same sub-game (between pursuer $i$ and evader 1 in $E_i$) starting from $x_1$. Choose the strategy of evader 1, $\beta^1_1$, such that $\beta^1_1[a_i](t) = \beta^2_1[a^2_i](t)$ for $t_0 \leq t \leq T^1_1 \wedge T^2_p$ and any $a_i(\cdot) \in \mathcal{A}^i(t_0)$.

\textsuperscript{27}Here, the subscript of $\beta^2_1$ stands for the first evader and the superscript stands for the initial point $x_2 \in \Omega$. 

135
where \( T_1^1 \) and \( T_1^2 \) are the capture time of evader 1 in \( E_i \) associated with the initial states \( x_1 \) and \( x_2 \). Under this setting, by (A.2),

\[
\bar{V}_{i1}(x_{i1}) \geq \inf_{a_i(\cdot) \in \mathcal{A}^i(t_0)} J_{i1}(a_i, \beta_1^1; x_{i1}^1),
\]

and there exists \( a_i^1(\cdot) \in \mathcal{A}^i(t_0) \), such that

\[
\bar{V}_{i1}(x_{i1}) \geq J_{i1}(a_i^1, \beta_1^1; x_{i1}^1) - \varepsilon. \tag{A.4}
\]

Since (A.3) holds for any \( a_i^2 \in \mathcal{A}^i(t_0) \), it is true when \( a_i^2(t) = a_i^1(t) \) for \( t_0 \leq t \leq T_1^1 \wedge T_1^2 \).

Next, we prove by cases according to the capture time of the first evader in \( E_i \) associated with \( x_1 \) and \( x_2 \), i.e., \( T_1^1 \) and \( T_1^2 \).

**Case 1** \((T_1^1 \leq T_1^2)\): Since the dynamics of the players is Lipschitz according to (2.4) and \( \Omega \) is bounded, there exists a constant \( C_1 > 0 \) such that the trajectories starting from \( x_1 \) and \( x_2 \) under strategies \((a_i^1, \beta_1^1)\) and \((a_i^2, \beta_1^2)\) satisfies

\[
\left\| x_{i1}(T_1^1) - x_{i1}(T_1^1) \right\| \leq C_1 \left\| x_{i1}(t_0) - x_{i1}(t_0) \right\|. \tag{A.5}
\]

Here, by abuse of notation, we use \( x_{i1}(\cdot) \) to denote the trajectory of pursuers \( i \) and evader 1 starting from \( x_{i1} \) for \( l = 1, 2 \). Clearly, \( x_{i1}(t_0) = x_{i1} \). Because of the strong capturability, there exists \( \delta_1 > 0 \) such that if \( \left\| x_{i1}(t_0) - x_{i1}(t_0) \right\| \leq \delta_1 \) there exists \( a_i^2(\cdot) \in \mathcal{A}^i(T_1^1) \) such that \( T_1^2 \leq T_1^1 + \varepsilon \) for any \( \beta_1^2 \in \Delta_1(T_1^1) \). It follows by (A.3)-(A.4) that

\[
\bar{V}_{i1}(x_{i1}^2) - \bar{V}_{i1}(x_{i1}) \leq T_1^2 - T_1^1 + 2\varepsilon \leq 3\varepsilon. \tag{A.6}
\]

Choose a control input \( a_i^1(t) \) for \( T_1^1 \leq t \leq T_1^2 \leq T_1^1 + \varepsilon \), such that no second evader is captured along the trajectory \( x_{i1}(\cdot) \) during the time from \( T_1^1 \) to \( T_1^2 \). This is possible because the second evader in \( E_i \) tries to escape. Under this construction, we want to check the subsequent game (between pursuer \( i \) and the second evader in \( E_i \)) from
both $x_1$ and $x_2$ at the same time $T_1^2$. Due to the boundedness of the dynamics and by (A.5), regardless of the control of evader 2 in $E_i$ (will be examined shortly), there exists a constant $\delta'_1 > 0$ such that,

$$|\nabla_{i2}(x_{i2}^1(T_1^1)) - \nabla_{i2}(x_{i2}^1(T_1^2))| \leq \varepsilon$$  \hfill (A.7)

for any $\|x_{i1}^1(t_0) - x_{i1}^2(t_0)\| \leq \delta'_1$. Finally, by the boundedness of the dynamics in (2.4),

$$\|x_p^1(T_1^2) - x_p^2(T_1^2)\| \leq C_2\|x_{1p}^i - x_{2p}^i\| + C_3(T_1^2 - T_1^1) \leq C_2\|x_{1p}^i - x_{2p}^i\| + C_3\varepsilon \text{ for some } C_2, C_3 > 0. \hfill (A.8)$$

Therefore, $\|x_p^1(T_1^2) - x_p^2(T_1^2)\|$ can also be suppressed by the initial condition of the sub-game between pursuer $i$ and the evaders in $E_i$. It provides some sense of continuity.

**Case 2** ($T_1^1 > T_1^2$): By (A.3)-(A.4),

$$\nabla_{i1}(x_{i1}^2(T_1^1)) - \nabla_{i1}(x_{i1}^1) \leq T_1^2 - T_1^1 + 2\varepsilon \leq 2\varepsilon. \hfill (A.9)$$

By similar arguments as in **Case 1**, i.e., due to the boundedness and the Lipschitz condition of the players’ dynamics,

$$\|x_{i1}^1(T_1^2) - x_{i1}^2(T_1^1)\| \leq C_4\|x_{i1}^1(t_0) - x_{i1}^2(t_0)\| \text{ for some } C_4 > 0. \hfill (A.10)$$

Equation (A.10) is the counterpart of (A.5). Again, by the strong capturability, there exists $\delta_2 > 0$ such that for any $\|x_{i1}^1 - x_{i1}^2\| \leq \delta_2$, $T_1^1 \leq T_1^2 + \varepsilon$. Choose $a_2^i(t)$ for $T_1^2 \leq t \leq T_1^1$, such that no second evader is captured. Similar to the inequality (A.8) in **Case 1**,

$$\|x_p^1(T_1^1) - x_p^2(T_1^1)\| \leq C_4\|x_{1p}^i - x_{2p}^i\| + C_5(T_1^1 - T_1^2) \leq C_4\|x_{1p}^i - x_{2p}^i\| + C_5\varepsilon \text{ for some } C_4, C_5 > 0. \hfill (A.11)$$
Now, consider the second evader (2) in $E_i$. Denote by $T_1 = T_1^1 \lor T_1^2 \equiv \max\{T_1^1, T_1^2\}$. Choose the strategy of evader 2 associated with $x_2, \beta_2^2 \in \Delta_2(t_0)$, such that

$$\sup_{\beta \in \Delta_2(t_0)} V_{i2}(x_{i2}^2(T_1; x_{i2}^2(t_0), a_1^2, \beta(a_2^2))) \leq V_{i2}(x_{i2}^2(T_1; x_{i2}^2(t_0), a_1^2, \beta(a_2^2))) + \varepsilon. \tag{A.12}$$

Here, $x_{i2}^2(T_1; x_{i2}^2(t_0), a_1^2, \beta(a_2^2))$ is the state $x_{i2}^2$ at $T_1$ starting from $x_{i2}^2(t_0)$ under the controls $a_1^2$ and $\beta(a_2^2)$. Let $\beta_2^1[a_1^1](t) = \beta_2^2[a_2^2](t)$ for $0 \leq t \leq T_1$ and any $a_i \in A_i(t_0)$. Then,

$$\sup_{\beta \in \Delta_2(t_0)} V_{i2}(x_{i2}^1(T_1; x_{i2}^1(t_0), a_1^1, \beta(a_1^1))) \geq V_{i2}(x_{i2}^1(T_1; x_{i2}^1(t_0), a_1^1, \beta_2^1[a_1^1])). \tag{A.13}$$

By the Lipschitz condition of the dynamics, similarly,

$$\|x_{e2}^1(T_1) - x_{e2}^2(T_1)\| \leq C_0\|x_{e2}^1(t_0) - x_{e2}^2(t_0)\| \text{ for some } C_0 > 0. \tag{A.14}$$

Considering (A.5) and (A.8) in Case 1 as well as (A.10) and (A.11) in Case 2, the subsequent game between pursuer $i$ and evader 2 in $E_i$ at time $T_1$ is equivalent to the game between pursuer $i$ and evader 1 and its initial state $x_{i2}^l(T_l)$ ($l = 1, 2$) at time $T_l$ can be suppressed by the initial state $x_1$ and $x_2$. Since same conclusions as in (A.6) and (A.7) in Case 1 and (A.9) in Case 2 can be drawn for the game between pursuer $i$ and evader 2 in $E_i$.

By mathematical induction, the same conclusions can be drawn for the rest of the evaders in $E_i$. Note that $J_i^h(x, z, E_i) = \sum_{j=1}^{n_i} V_{ij}(x_{ij})$ in (A.1) and that (A.5), (A.8) in Case 1 and (A.10), (A.11) in Case 2 hold for any selected $a_1^1, a_2^1, \beta_1^1$ and $\beta_2^1$. By inspection of (A.12) and (A.13), it can be shown that for a given $\varepsilon > 0$, there exists $\delta > 0$ such that

$$J_i^h(x_2, z, E_i) - J_i^h(x_1, z, E_i) \leq \varepsilon$$

when $\|x_1 - x_2\| \leq \delta$ for some $\delta > 0$. Since $x_1$ and $x_2$ are arbitrary, it is true that $J_i^h(x_1, z, E_i) - J_i^h(x_2, z, E_i) \leq \varepsilon$ when $\|x_1 - x_2\| \leq \delta$ for some $\delta > 0$. Let
\( \delta^* = \min\{\delta, \bar{\delta}\} \). Then, \( \|J^h_i(x_1, z, E_i) - J^h_i(x_2, z, E_i)\| \leq \varepsilon \) for \( x_1 - x_2 \leq \delta \). Thus, \( J^h_i(x, z; E) \) is continuous in \( x \) with any feasible \( E \). Note that \( \delta^* \) does not depend on \( x \), such that the continuity is uniform with respect to \( x \).

Using the arguments similar to those in Lemma 5, it can be shown that \( \tilde{V}^h_i(x, z) \) is uniformly continuous. The boundedness of \( \tilde{V}^h \) is because \( \Omega \) is bounded.

**Remark 14** In the proof of Theorem 5, the argument in (A.10) is important because the arguments are constructed based on (A.3) and (A.4) and in (A.4) the “optimal” control input of pursuer \( i \) is considered. However, during the time from \( T^1_1 \) to \( T^2_1 \), the control \( a^1_i \) is chosen arbitrarily. Equation (A.10) shows that the sacrifice of Value due to \( a^1_i \) can be suppressed by the initial state, i.e., \( \|x_1 - x_2\| \).
APPENDIX B

PROOF OF THEOREM 6

Proof: Prove by induction. Assume that $W_n$ ($n \in \mathbb{Z}_{\geq 0}$) is uniformly continuous in $x$, we want to prove the uniform continuity of $W_{n+1}$. By Lemma 6, it suffices to show the uniform continuity on $\Lambda^c_z$.

Consider an arbitrary $z \in Z^M$ ($z \neq 0$). We first prove the continuity of $W_{n+1}$ on $\Lambda^c_z$. Choose any $x_0, x_1 \in \Lambda^c_z$, which is equivalent to

$$z = z^{x_0} = z^{x_1}. \quad (B.1)$$

Refer to (2.26) on page 29 for the definition of $z^x$. Without loss of generality, we assume that

$$W_{n+1}(x_0, z) \leq W_{n+1}(x_1, z). \quad (B.2)$$

Given any $\varepsilon > 0$, there exists a strategy $\beta_1(\cdot) \in \Delta(t)$, such that the following holds for any $a_1(\cdot) \in A(t)$.

$$W_{n+1}(x_1, z) = \overline{H}[W_n](x_1, z)$$

$$\leq \int_t^{t+\Delta t} G(x_\tau, z_\tau, a_1[\cdot]_\tau, \beta_1[\cdot]_\tau) d\tau + W_n\left(x_{t+\Delta t; x_1, a_1, \beta_1[\cdot]_1}, z_{t+\Delta t; x}\right) + \varepsilon \quad (B.3)$$
Here, we use the notation $\beta_1$ (or $\beta_0$) to stand for the strategy of the evaders associated with the initial state $x_1$ (or $x_0$), and similarly use $a_1$ (or $a_0$) for the pursuers. Select $\beta_0(\cdot)$ as

$$\beta_0[a_0](\tau) = \beta_1[a_1](\tau) \quad \text{for any } t \leq \tau \leq t + \Delta t \text{ and } a_0, a_1 \in A(t) \quad \text{(B.4)}$$

and specifically, $\beta_0[a_0]_j(\tau) = \beta_1[a_1]_j(\tau)$ (control of the $j^{th}$ evader) for $j \in I_{z_0(\tau)} \cup I_{z_1(\tau)}$, where $z_0(\tau)$ or $z_1(\tau)$ is the state $z$ at time $\tau$ associated with trajectory of $x$ with the initial point $x_0$ or $x_1$ under $(a_0, \beta_0)$ or $(a_1, \beta_1)$. By (B.4), clearly,

$$W_{n+1}(x_0, z) \geq \inf_{a(\cdot) \in A(t)} \left\{ \int_{t}^{t+\Delta t} G(x_\tau, z_\tau, a_\tau, \beta_0[a_\tau])d\tau + W_n(x_{t+\Delta t; x_0, a_0[\cdot], z_0[\cdot]}; z_{t+\Delta t; z, x}) \right\}$$

Further, there exists $a_0(\cdot) \in A(t)$, such that

$$W_{n+1}(x_0, z) \geq \int_{t}^{t+\Delta t} G(x_\tau, z_\tau, a_\tau, \beta_0[a_\tau])d\tau + W_n(x_{t+\Delta t; x_0, a_0[\cdot], z_0[\cdot]}; z_{t+\Delta t; z, x}) - \varepsilon. \quad \text{(B.5)}$$

Let the time when the transition of $z$ along $z_0(\tau)$ (or $z_1(\tau)$) occurs be $T^{[z]}_0$ (or $T^{[z]}_1$). Define $T^{[z]} = T^{[z]}_0 \wedge T^{[z]}_1$. Equation (B.4) completely defines strategy $\beta_0$ for each evader $j \in I_{z_0}$ for $t \leq \tau \leq T^{[z]}$. Denote by $x_1^\tau(\cdot)$ (resp. $x_0^\tau(\cdot)$) the players’ trajectories under $(a_1, \beta_1)$ (resp. $(a_0, \beta_0)$) starting from $x_1$ (resp. $x_0$). Clearly,

$$x_1^\tau(\tau) = x(\tau; x_1, a_1[\cdot])$$
$$x_0^\tau(\tau) = x(\tau; x_0, a_0[\cdot]) \quad \text{(B.6)}$$

for $t \leq \tau \leq T^{[z]}$. We prove by cases as follows.

**Case 1**) Suppose that $t + \Delta t \leq T^{[z]}$, i.e., there is no transition of $z$ occurring from $t$ to $t + \Delta t$ along both $z_0(\tau)$ and $z_1(\tau)$. Since (B.3) holds for any $a_1 \in A(t)$, it
is true if we choose $a_i = a_0$. By (B.3)-(B.5),

$$W_{n+1}(x_1, z) - W_{n+1}(x_0, z)$$

$$\leq \int_{t}^{t + \Delta t} \left[ G(x_{\tau}, z_{\tau}, a_1, \beta_1) - G(x_{\tau}, z_{\tau}, a_0, \beta_0) \right] d\tau$$

$$+ W_n \left( (x^i_{\tau}(t + \Delta t), z_1(t + \Delta t)) - W_n \left( (x_0^i(t + \Delta t), z_0(t + \Delta t)) \right) + 2\varepsilon. \right)$$

(B.7)

By the Lipschitz condition of the dynamics $f$, the boundedness and the Lipschitz condition of the cost rate $G$, and the uniform continuity of $W_n$, both terms on the right hand side of (B.7) can be suppressed by $\|x_0 - x_1\|$. Therefore, there exists $\delta_1 > 0$ such that

$$W_{n+1}(x_1, z) - W_{n+1}(x, z) \leq 3\varepsilon \text{ for } \|x_0 - x_1\| \leq \delta_1.$$

(B.8)

For the case $t + \Delta t > T^{i|z|-1}$, i.e., there is at least one evader that is captured along either the trajectory $x^i_1$ or $x^0_1$. We prove the case that only one evader is captured from $t$ to $t + \Delta t$, and it can be extended to the case when multiple evaders are captured during the time interval.

Case 2) $(t + \Delta t > T^{i|z|-1}; T^{|z|-1}_1 < T^{|z|-1}_0)$. Let $\tau_1 = T^{|z|-1}_1$. Suppose that evader $j$ that is captured by pursuer $i$ at $\tau_1$ along trajectory $x^i_1$. The following analysis is focused on pursuer $i$ and evader $j$. By the Lipschitz condition of the players’ dynamics, given any $\lambda > 0$, there is $\delta_2 > 0$ such that

$$\left\| x^{i\tau}_p(\tau_1) - x^{i\tau}_1 p(\tau_1) \right\| \leq \lambda \text{ and } \left\| x^{j\tau}_e(\tau_1) - x^{j\tau}_1 e(\tau_1) \right\| \leq \lambda$$

(B.9)

for any $\|x_1 - x_0\| \leq \delta_2$. Consider the two-player PE game between pursuer $i$ and evader $j$ starting from $x^{i\tau}_p(\tau_1)$ and $x^{j\tau}_e(\tau_1)$ along the trajectory $x^i_1$. By the assumption of $^{28}$Here, $x^{i\tau}_p(x^{i\tau}_0)$ denotes the state trajectory of pursuer $i$ (evader $j$) corresponding to the composite state trajectory of $x^0_0$. 

142
strong capturability, given any \( \gamma \) for \( 0 \leq \gamma \leq t + \Delta t - \tau_1 \), there exists \( \lambda_1 > 0 \) such that there exists a \( \tilde{a}_{0i}(\cdot) \in A^i(\tau_1) \) such that the capture time of evader \( j \) (associated with \( x_0^j \)), \( T_j^0 \leq \tau_1 + \gamma \) for any \( \|x_1^j(\tau_1) - x_0^j(\tau_1)\| \leq \lambda_1 \) for any \( b_j(\cdot) \in B^j(\tau_1) \). Furthermore, by (B.9), there exists a \( \delta_3 > 0 \) such that for any \( \|x_1 - x_0\| \leq \delta_3 \), it is true that \( \|x_1^j(\tau_1) - x_0^j(\tau_1)\| \leq \lambda_1 \) as well as \( T_j^0 \leq \tau_1 + \gamma < t + \Delta t \). Now, define a new strategy of the pursuers \( \tilde{a}_0 \) as

\[
\begin{cases}
\tilde{a}_0(\tau) = a_{0l}(\tau), & t \leq \tau < t + \Delta t \\
\tilde{a}_0(\tau) = a_{0l}(\tau) & t \leq \tau < \tau_1 \\
\tilde{a}_0(\tau) = \tilde{a}_{0i}(\tau) & \tau_1 \leq \tau < T_j^0
\end{cases}
\]  

(B.10)

where \( a_0 \) is defined in (B.5). Note that \( f \) in the dynamics is bounded and \( (T_j^0 - \tau_1) \) can be made arbitrarily small. Therefore, similar to (B.5), under the control \( \tilde{a}_0 \),

\[
W_{n+1}(x_0, z) \geq \int_t^{t+\Delta t} G(x_\tau, z_\tau, \tilde{a}_\tau, \beta_0[\tilde{a}]_\tau) d\tau + W_n(x_0^\tau(t + \Delta t), z_0(t + \Delta t)) - C_1 \varepsilon
\]

(B.11)

for some constant \( C_1 > 0 \). Now, let \( a_1(\tau) = \tilde{a}_0(\tau) \) for \( t \leq \tau < t + \Delta t \). Thus, (B.3) still holds. Due to (B.4) and by (B.11) and (B.3), there exists \( \delta_4 > 0 \) such that

\[
W_{n+1}(x_1, z) - W_{n+1}(x_0, z) \leq C_2 \varepsilon
\]

(B.12)

for any \( \|x_0 - x_1\| \leq \delta_4 \) and some constant \( C_2 > 0 \). Conclusion in (B.12) is drawn similarly as in (B.7)-(B.8).

**Case 3** \( (t + \Delta t > T^{[z]} - 1) \): \( T_1^{[z]} - 1 > T_0^{[z]} - 1 \). Similar to Case 2, let \( \tau_0 = T_0^{[z]} - 1 \), and suppose that evader \( j \) is captured by pursuer \( i \) at \( \tau_0 \) along the trajectory \( x_0^j \). By the strong capturability, there exists a strategy of pursuer \( i \), \( \tilde{a}_{1i}(\cdot) \in A^i(\tau_0) \), such that \( T_j \leq \tau_0 + \gamma \) for some \( \gamma \) \((0 < \gamma \leq t + \Delta t - \tau_0)\) that can be made arbitrarily small by suppressing \( \|x_0 - x_1\| \). In this case, we choose the \( a_0 \) found in (B.5), and construct
the strategy \( \bar{a}_1 \) of the pursuers along trajectory \( x_1^τ \) as

\[
\bar{a}_1(τ) = a_{0l}(τ), \quad t \leq τ < t + \Delta t \text{ for } l = 1, \cdots, N \text{ and } l \neq i;
\]

\[
\begin{cases}
\bar{a}_{1l}(τ) = a_{0l}(τ) & t \leq τ < τ_0 \text{ and } T_j^1 \leq τ < t + \Delta t, \\
\bar{a}_{1l}(τ) = \bar{a}_{1l}(τ) & τ_0 \leq τ < T_j^1,
\end{cases}
\]

(B.13)

Note that (B.3) holds for any \( a_1 \). Use the \( β_l \) found in (B.3), and consider (B.4), it can be proved similarly as in Case 2, i.e., there exists \( δ_5 > 0 \) such that

\[
W_{n+1}(x_1, z) - W_{n+1}(x_0, z) \leq C_3ε
\]

for some \( C_3 > 0 \) and any \( \|x_0 - x_1\| \leq δ_5 \).

Considering the (B.2), (B.8), (B.12) and (B.14), we can show the continuity of \( W_{n+1} \) in \( x \) on \( \Lambda_x^c \) if only one evader is captured during the \( Δt \) interval. Following the idea in the proof of Theorem 5, the conclusion can be extended the case where multiple evaders are captured during \( Δt \). Then, \( W_{n+1} \) is continuous in \( x \) on \( \Lambda_x^c \).

Next, we examine that if \( W_{n+1} \) is continuous in \( x \) on \( \overline{\Lambda}_x^c \), i.e., the cases other than that in (B.1). We first check the case of \( z^{x_0} = z^{x_1} \neq 0 \). Considering that \( \Lambda_x^c \) is open and the equality in (2.57), given \( x_0, x_1 \in \overline{\Lambda}_x^c \), if \( z^{x_0} = z^{x_1} \neq 0 \), the continuity of \( W_{n+1} \) on \( \overline{\Lambda}_x^c \) is equivalent to the continuity on \( \Lambda_x^c \), which has been proved. Furthermore, the cases where \( z^{x_0} \neq z^{x_1} \) can be transferred into those similar to Case 2 or Case 3 (if \( |z^{x_0}| \land |z^{x_1}| = |z| - 1 \) with \( τ_0 \) or \( τ_1 = t \) by checking if \( W_{n+1}(x_0, z) \leq W_{n+1}(x_1, z) \) or otherwise. More general cases when \( |z^{x_0}| \land |z^{x_1}| < |z| - 1 \) can be shown accordingly in the case with capture of multiple evaders during \( Δt \). Thus, \( W_{n+1}(x, z) \) is continuous in \( x \) on \( \overline{\Lambda}_x^c \). By Lemma 6, \( W_{n+1}(x, z) \) is continuous in \( x \) on \( X \). Note that the continuity does not depend on \( x \), so that it is uniform.

Finally, by the hypothesis that \( W_0 = \tilde{V} \) is uniformly continuous in \( x \) on \( X \). By mathematical induction, this is true for all \( W_k, k \in \mathbb{Z}^+ \).


150


