Air Combat Maneuvering via Operations Research and Artificial Intelligence Methods

THESIS

James B. Crumpacker, Capt, USAF

AFIT-ENS-MS-21-M-152

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENS-MS-21-M-152

AIR COMBAT MANEUVERING VIA OPERATIONS RESEARCH AND ARTIFICIAL INTELLIGENCE METHODS

THESIS

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Operations Research

James B. Crumpacker, BS

Capt, USAF

March 25, 2021

AFIT-ENS-MS-21-M-152

AIR COMBAT MANEUVERING VIA OPERATIONS RESEARCH AND ARTIFICIAL INTELLIGENCE METHODS

THESIS

James B. Crumpacker, BS
Capt, USAF

Committee Membership:

Dr. Matthew J. Robbins
Chair

Capt Phillip R. Jenkins, PhD
Member

AFIT-ENS-MS-21-M-152

Abstract

Within visual range air combat involves execution of highly complex and dynamic activities, requiring rapid, sequential decision-making to survive and defeat the adversary. Fighter pilots spend years perfecting tactics and maneuvers for these types of combat engagements, yet the ongoing emergence of unmanned, autonomous vehicle technologies elicits a natural question – can an autonomous unmanned combat aerial vehicle (AUCAV) be imbued with the necessary artificial intelligence to perform challenging air combat maneuvering tasks independently? We formulate and solve the air combat maneuvering problem (ACMP) to examine this important question, developing a Markov decision process (MDP) model to control an AUCAV seeking to destroy a maneuvering adversarial vehicle. The MDP model includes a 5-degree-of-freedom, point-mass aircraft state transition model to accurately represent both kinematics and energy while maneuvering. The high dimensional and continuous nature of the state space within the ACMP precludes the implementation of classical solution approaches. Instead, an approximate dynamic programming (ADP) approach is proposed wherein we develop and test an approximate policy iteration algorithm that implements neural network regression to attain high-quality maneuver policies for the AUCAV. A representative intercept scenario is specified for the purposes of computational testing wherein an AUCAV is tasked with defending an area of responsibility and must engage and destroy an adversary aircraft attempting to penetrate the defended airspace. Several designed experiments are conducted to determine how aircraft characteristics and adversary maneuvering tactics impact the efficacy of the proposed ADP solution approach. Moreover, designed experiments enable efficient algorithmic hyperparameter tuning. ADP-generated policies are compared to

two accepted benchmark maneuver policies found in the contemporary ACMP literature, one considering position-only and one considering both position and energy. Across the 18 problem instances investigated, ADP policies outperform position-only benchmark policies in 15 of 18 instances and outperform position-energy benchmark policies in 9 of 18 instances, attaining improved probabilities of kill among problem instances most representative of typical air intercept engagements. As an interesting excursion, and for qualitative validation of our approach, maneuvers generated by the ADP policies are compared to standard, basic fighter maneuvers and common aerobatic maneuvers. Results indicate that our proposed ADP solution approach produces policies that imitate known flying maneuvers.

Acknowledgements

Foremost, I would like to express my sincere gratitude to my advisor, Dr. Matthew J. Robbins, for his mentorship and support. I also want to thank the Air Force Research Lab and the Air Force Strategic Development Planning and Experimentation Office for sponsoring this research. Lastly, I want to thank my family for their unwavering support.

James B. Crumpacker

Table of Contents

Page

List of Figures

List of Tables

AIR COMBAT MANEUVERING VIA OPERATIONS RESEARCH AND ARTIFICIAL INTELLIGENCE METHODS

I. Introduction

> "Fighter aircraft exist to destroy other aircraft. The airplane itself may be considered only a weapons platform designed to bring the weapons system into position for firing." (Shaw, 1985)

The primary mission of the United States Air Force (USAF) is air superiority. Achieving air superiority is often viewed as a necessary precursor to any military operation (United State Air Force, 2015). The USAF conducts counterair operations continuously using fighter aircraft (i.e. aerial vehicles) to achieve the desired level of control over an area of responsibility (AOR). The joint counterair framework comprises both offensive counterair (OCA) and defensive counterair (DCA) operations (Joint Chiefs of Staff, 2018). OCA operations include the following combat activities: attacking airfields, suppressing enemy air defenses, and performing fighter escort and sweep missions. DCA operations include executing active air and missile defense missions as well as performing many passive activities including camouflage, concealment, dispersion, and hardening (United States Air Force, 2019).

Active DCA operations primarily focus on air and missile defense wherein direct defensive actions are taken to destroy, or reduce the effectiveness of, attacking air and missile threats. Although air-to-air combat has historically been performed by manned fighter aircraft, the United States Department of Defense (DOD) recognizes the opportunity presented by the ongoing development of autonomous robotic systems. The DOD predicts the emergence of these systems working together in

groups to perform complex military tasks, including making autonomous decisions and delivering lethal force (Joint Chiefs of Staff, 2016). Indeed, many countries, including the United States and Australia, are flight testing semi-autonomous aircraft (Gregg, 2019; Insinna, 2020). Byrnes (2014) contends that the convergence of these new technologies indicate an impending revolutionary change to air warfare tactics. More recently, at the 2020 Air Force Association's Air Warfare Symposium, Elon Musk, when asked how aerial combat could be revolutionized, declared that locally autonomous drone warfare is the future and that the fighter jet era has passed. In response, retired USAF Lieutenant General David Deptula writes that manned fighters will dominate for decades, but does concede that it might be possible one day to have one-versus-one (1v1) within visual range (WVR) air combat that is fully autonomous (Deptula, 2020). Although the specific timeline is debatable, most agree that autonomous combat aerial vehicles (AUCAVs) are on the horizon.

Aircraft have performed combat operations since the early days of aviation. The Wright brothers took the first powered flight in December of 1903 and, shortly after, the United States military purchased its first aircraft in August of 1909. Early air combat consisted of pilots employing small arms fire from their cockpits in an attempt to shoot down enemy aircraft. The first recorded instance of aerial combat occurred in November of 1913 during the Mexican Revolution. Over the Arizona-Mexico border, two pilots fired revolvers while circling each other for about 15 minutes, with neither achieving any hits (Woodman, 1989).

As aviation became a mature science, the military formalized its use with the National Defense Act of 1920, creating a permanent air service within the United States Army. From the speed and high altitude emphasis of the 1950s, to the stealth and maneuverability of the 1990s, as technology advanced so, too, did tactics. With each new revolutionary change in air combat technology, tactics were developed,

evaluated, modified, tested, and passed down to the next generation of fighter pilot. This "human" evolutionary optimization process has proven successful over the years. However, it loses applicability in the context of AUCAVs. If the human pilot is removed from the aircraft, the natural question becomes: how do we imbue an AUCAV with the necessary artificial intelligence (AI) to accomplish complex tasks such as air combat?

One possible solution is to program the AUCAV with so called rules-based logic, developed from current tactics, wherein the aircraft executes a prescribed maneuver based on some combination of features of the air combat situation such as the current geometry. This approach suffers from two major flaws. First, it precludes the ability to learn new tactics. For example, if scientists and engineers develop a new air-to-air missile, human pilots must somehow develop employment tactics for both the new weapon and the autonomous aircraft they cannot fly themselves. Second, it limits AUCAV performance based on a constraint it no longer has: a human pilot.

To date, WVR air combat tactics have been driven by human limitations. Indeed, contemporary air combat only involves those maneuvers that human pilots can physically withstand. Without the physical limitation of a human pilot, an AUCAV has a much wider range of maneuvers available to it. It can freely pull higher positive and negative G loads for longer periods of time compared to human pilots. These higher performance maneuvers yield tighter turns and faster turn rates, providing a significant advantage in air combat. As an example, Shaw (1985) describes an out-of-plane lag pursuit roll maneuver used to decrease closure rate and avoid an overshoot when approaching a target with low angle off tail (AOT) and high overtake. In this maneuver, the defending aircraft is executing a horizontal turn. The attacking aircraft pulls up out of the defender's plane of turn, executes a slow roll as it passes over the defender, matches turn rate, and dives back down into the defender's six o'clock

position. The purpose of the attacker's slow roll maneuver is to allow the attacking pilot to keep eyes on the defender (i.e., maintain situational awareness) throughout the maneuver. For an AUCAV not limited by human eye sight, would this roll be necessary? Could the AUCAV execute a modified maneuver that puts it in a better position to counter the defender's reaction while still allowing it to employ weapons effectively? Existing fighter aircraft typically employ fixed forward-firing guns, which drives the desire to maneuver into the defender's six o'clock position. What if the AUCAV carried directed-energy weapons capable of firing at targets abeam of, or directly below, the attacker?

A better solution for generating intelligent AUCAV behavior is to use approximate dynamic programming (ADP) and reinforcement learning techniques to discover and optimize maneuvers based on the AUCAV's capabilities, unhindered by human limitations. This AI approach does not suffer from the flaws of the rules-based methodology because, as new capabilities are delivered, new near-optimal maneuvers can be computed without bias from past experiences.

AI-based air combat maneuver (ACM) policies can inform weapon concept development, improve testing and evaluation, and enhance training. Incorporating AI-based ACM policies (i.e., intelligent behavior models) into existing simulations such as the Advanced Framework for Simulation, Integration, and Modeling (AFSIM) (West and Birkmire, 2019) and the Joint Simulation Environment (JSE) (Casem, 2019) increases the realism of the scenarios and usefulness of the results. For example, when performing simulation-based studies to evaluate effectiveness of futuristic weapons, the inclusion of intelligent entities yields more meaningful results and better informs concept development decisions. Moreover, including intelligent entities in a man-in-the-loop simulator significantly increases its flexibility as both a test and training tool by allowing for the development of more complex and realistic scenarios (Toubman,

2020).

This research examines the 1v1 WVR ACM problem (WVR-ACMP) wherein an AUCAV must make maneuver decisions in the context of a DCA mission. A patrolling friendly (i.e., blue) AUCAV must defend an AOR by destroying an attacking adversary (i.e., red) aircraft. The resulting WVR-ACMP is formulated as an infinite-horizon, discounted Markov decision process (MDP) wherein the sequential decisions consider changes in pitch, roll, and throttle setting. The MDP model determines how to maneuver the blue AUCAV with the goal of attaining and maintaining a position of advantage, relative to the red aircraft, from which weapons can be employed.

The MDP formulation is used to develop maneuver policies for a simple yet representative intercept scenario. We develop 18 problem instances to explore the efficacy and robustness of the ADP-generated ACM policies. Moreover, we compare our ADP-generated policies against two benchmark ACM policies derived from common decision functions found in the ACMP literature. We also comment on the ability of the ADP generated policies to replicate common aerobatic maneuvers and basic combat tactics as described by (Shaw, 1985).

The remainder of this research is organized as follows. Chapter II provides a review of research relating to the ACMP and aircraft trajectory optimization. Chapter III presents a description of the 1v1 ACMP, presents the MDP formulation, and describes the ADP solution approach used to determine the blue AUCAV maneuvers. Chapter IV presents a quantitative analysis of the resulting maneuver policies as compared to the benchmark policies and provides qualitative comments on the ability of the ADP-generated ACM policies to generate common maneuvers. Chapter V concludes the research and proposes several directions for future research.

II. Literature Review

Air combat is a highly complex activity that requires the consideration of many disciplines. Although the goal of this research is to find high-quality air combat maneuver (ACM) policies via stochastic optimization, it is important to understand air combat and its current issues in general so that appropriate features can be included in the model. Four areas of literature inform the development and analysis of the one-versus-one (1v1) within visual range (WVR) air combat maneuvering problem (ACMP). The first concerns the dynamics of flight and air combat geometry. The second concerns current and emerging air combat weapons technology. The third concerns the physiological effect of acceleration on human pilots. The fourth concerns the modeling approaches taken by different communities.

2.1 Aircraft Dynamics

Consider an aircraft in flight. One way to model the motion of the aircraft is to treat it as a point mass moving through space. Figure 1 shows the angles associated with this type of point mass model where γ is the flight path angle, χ is the heading angle, and V is the velocity. The aircraft in this model is free to translate along the x, y, and z axes. This model is considered a three degree of freedom (3DOF) point mass aircraft model and represents the lowest level of fidelity for modeling aircraft motion in three-dimensional space.

Figure 1 shows two frames of reference: the Earth frame and the body frame. The Earth frame of reference is that of an observer on the ground and is denoted with an e subscript. It is important to note the Earth frame is centered at a fixed arbitrary point relative to the Earth and does not follow the center of gravity of the aircraft. The Earth frame axes are drawn from the aircraft center of gravity to highlight flight

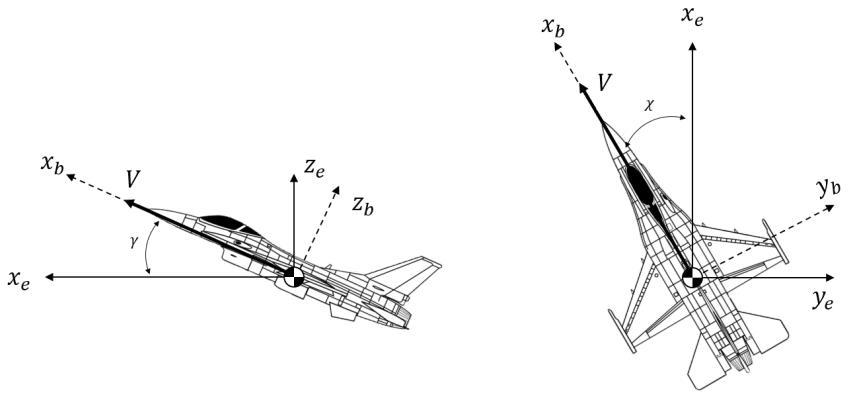


Figure 1. 3DOF Point Mass Aircraft Model

angles. The body frame of reference is fixed to the aircraft with its origin at the center of gravity. From Figure 1 it is straightforward to derive the equations of motion for this 3DOF point mass model.

$$\frac{dx}{dt} = V \cos\gamma \cos\chi \tag{1a}$$

$$\frac{dy}{dt} = V \cos\gamma \sin\chi \tag{1b}$$

$$\frac{dz}{dt} = V \sin\gamma \tag{1c}$$

The 3DOF point mass aircraft model can be described as flying a velocity vector though space where the control inputs are changes in flight path angle, heading angle, and velocity. This model is popular primarily due to its simplicity and has been used with ADP techniques to develop reasonable maneuver policies (Wang et al., 2020). One problematic feature of the 3DOF point mass model is its lack of forces. For example, intuitively the aircraft should slow down due to drag. However, the equations of motion as written (i.e., Equations (1a)-(1c)) allow the aircraft to maintain velocity indefinitely with a commanded change in velocity of zero. Moreover, the equations of motion do not include the influence of gravity. The model implies a flight condition with no drag wherein the only means of increasing or decreasing

altitude is by commanding a positive or negative flight path angle γ.

Consider the same aircraft with all forces acting through the center of gravity while angular momentum of the rigid body is ignored. Figure 2 shows the forces and angles associated with this point mass model. The aircraft is free to translate along the x, y, and z axes. Additionally, the aircraft is free to rotate (i.e., roll, pitch, yaw) about the x, y, and z axes. This model is considered a six degree of freedom (6DOF) point mass aircraft model and represents a medium level of fidelity for modeling aircraft motion in three-dimensional space.

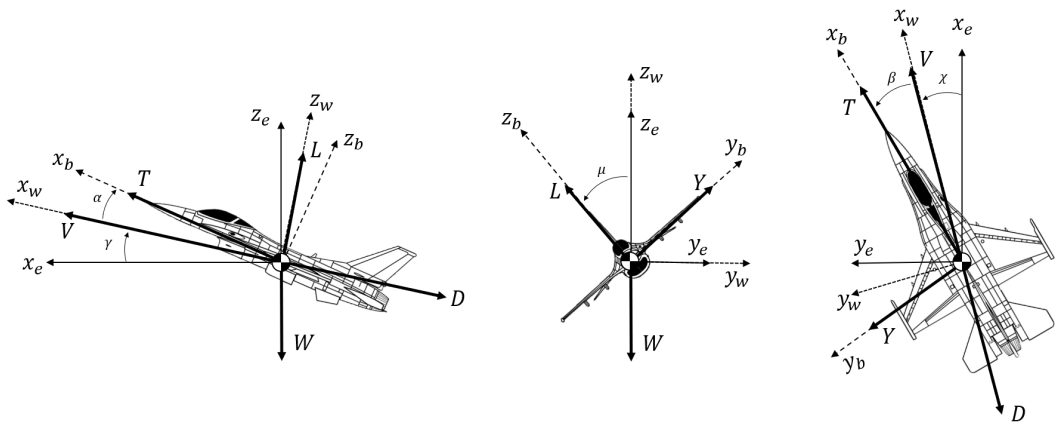


Figure 2. 6DOF Point Mass Aircraft Model

The additional degrees of freedom (i.e., rotation about each axis) necessitate the addition of a third frame of reference: the wind frame, indicated with a subscript w. The positive x_w axis is defined to be aligned with the velocity vector. Figure 2 shows the forces acting on the aircraft in flight are lift (L), weight due to gravity (W), thrust (T), drag (D), and side-force (Y). The forces on the aircraft result in the velocity vector (V). Five angles are shown, four of which relate to the velocity vector with the last relating the two frames of reference. The angle of attack (α) and sideslip angle (β) measure the angle between the velocity vector and the body frame. The flight path angles (γ, χ) measure the angle between the velocity vector and the Earth frame. The bank angle (μ) measures the angle between the body frame and

the Earth frame. The angles α, β, and μ are referred to as aerodynamic angles while γ and χ are called flight path angles.

The four primary aerodynamic forces are lift, which counters weight, and thrust, which counters drag. The side-force is the result of non-zero sideslip and/or direct side-force control (DSFC), which induces lateral accelerations independent of the aircraft roll and yaw. DSFC can be useful for terminal area tasks (i.e., approach and landing) and weapons delivery (Binnie and Stengelf, 1979; Watson and McAllister, 1977). Lift always acts along the y_w axis, drag always acts along the negative x_w axis, and the velocity vector always acts along the positive x_w axis. We assume thrust always acts along the x_b axis. Although not considered in this research, for completeness, it should be noted that advanced maneuverability systems such as thrust vectoring violate this assumption. The force of gravity always acts along the z_e axis. Lift, drag, and thrust are functions of velocity, altitude, and aircraft specific parameters.

$$L = \frac{1}{2}\rho V^2 S C_L \tag{2a}$$

$$D = \frac{1}{2}\rho V^2 S C_D \tag{2b}$$

$$Y = \frac{1}{2}\rho V^2 S C_Y \tag{2c}$$

$$W = mg \tag{2d}$$

$$T = f(V, h, E) \tag{2e}$$

Equations (2a) - (2c) can be found in any entry level aerodynamics text, where ρ is the density of the atmosphere at the current altitude h; S is the wing planform area; and C_L, C_D, and C_Y are the non-dimensional coefficients of lift, drag, and side-force, respectively. In general, the non-dimensional coefficients are functions of the Mach number, angle of attack, and sideslip angle. Equation (2d) is the force of gravity

wherein m is the mass of the aircraft and g is the constant acceleration due to gravity. Equation (2e) indicates thrust, which is generally a function of velocity, altitude, and other engine parameters E, such as the fuel-air ratio and exhaust velocity (Hill and Peterson, 1992).

Consider the sideslip angle β in Figure 2. When β is commanded to be zero, the aircraft is said to be in coordinated flight. When β is non-zero, drag is increased, and for some aircraft, there exists the possibility of entering dangerous flight conditions such as a spin. For these reasons, β is typically commanded to zero during normal flight operations. Fixing β to be zero allows for some simplification in the dynamic model (i.e., thrust is always inline with velocity, and side-force is always zero). Moreover, it reduces the action space in the subsequent MDP formulation, which improves our ability to find high-quality solutions via ADP. For these reasons, we set aside a 6DOF aircraft model and instead construct a 5DOF model wherein the sideslip angle is assumed fixed at zero. This constriction represents an assumption of *coordinated flight*. Consider that, in a military context, it may be useful to induce non-zero sideslip for the purpose of pointing the nose of the aircraft towards a target to employ weapons. However, this is not a common tactic and is not considered in this research.

Although more complex, the equations of motion can be derived from Figure 2 using classical mechanics and the coordinated flight assumption (i.e., $\beta = 0$ and $Y = 0$).

$$\frac{dV_{wx}}{dt} = \left(\frac{1}{m}\right)(T\cos\alpha - D - W\sin\gamma) \tag{3a}$$

$$\frac{dV_{wy}}{dt} = \left(\frac{1}{m}\right) L\sin\mu \tag{3b}$$

$$\frac{dV_{wz}}{dt} = \left(\frac{1}{m}\right)(T\sin\alpha + L\cos\mu - W\cos\gamma) \tag{3c}$$

$$\frac{dV}{dt} = \frac{dV_{wx}}{dt} \tag{3d}$$

$$d\gamma = \arctan \frac{dV_{wz}}{V} \tag{3e}$$

$$d\chi = \arctan \frac{dV_{wy}}{V} \tag{3f}$$

$$\frac{dx}{dt} = V \cos\gamma \cos\chi \tag{3g}$$

$$\frac{dy}{dt} = V \cos\gamma \sin\chi \tag{3h}$$

$$\frac{dz}{dt} = V \sin\gamma \tag{3i}$$

$$\frac{dm}{dt} = -cT - \sum_{w \in \mathcal{W}} I_w m_w \tag{3j}$$

wherein V_{wx}, V_{wy}, and V_{wz} are the decomposed changes in velocity in the wind frame, α is the angle of attack, and μ is the bank angle. Equations (3a) - (3c) give the decomposed change in velocity in the wind frame due to the forces of thrust, lift, drag, and weight. Equation (3d) gives the change aircraft velocity along the x_w axis. Equations (3e) and (3f) give the change in flight path and heading angle, respectively. Equations (3g) - (3i) are carried forward from the 3DOF model. Equation (3j) gives the change in aircraft mass over time, wherein c is the specific fuel consumption, I_w is an indicator functions that returns 1 if the wth weapon in the set of weapons $\mathcal{W}$ has been fired, and m_w is the mass of a projectile for the wth weapon. During an air combat engagement mass is lost in two ways: fuel burn and weapon expenditure. Mass loss due to fuel burn is a function of the specific fuel consumption and current thrust while the mass lost due to weapon expenditure is a function of the total number of weapons expended. For simplicity many aircraft models assume changes in aircraft mass are negligible over the short duration of combat maneuvering, which Shaw (1985) notes can be five minutes or less. However, for aircraft that carry many weapons or have large internal fuel tanks, the change in mass may be influential and thus should be considered. If the engine and weapons parameters are known, there

is no reason to neglect the change of mass over time and risk neglecting a potentially significant term.

Control surfaces (e.g., ailerons, rudder, elevators, and canards) are aerodynamic surfaces used to control an aircraft in flight. To execute a maneuver such as a roll, the control surfaces are deflected, which causes moments about the center of gravity of the aircraft. These moments translate to changes in the angle of attack and bank angle, which in turn cause a change in the flight path of the aircraft. Due to mechanical and electrical efficiencies, the deflections of control surfaces are limited to specified rates (e.g., the ailerons can be deflected at a maximum rate of 90 degrees per second). Based on the design of the aircraft, this control surface rate limit corresponds to aerodynamic angle rate limits (e.g., angle of attack can be changed at a maximum rate of 45 degrees per second), which in turn corresponds to flight path angle rate limits (e.g., vertical flight path angle can be changed at a maximum rate of 30 degrees per second). In terms of fidelity, using control surface deflections as the control variable represents the highest level of fidelity in aircraft path modeling while using flight path angles (i.e. γ, χ) represents the lowest. Specifying angle rates at any level implicitly assumes the modeled aircraft can achieve sufficient rates at lower levels to support the assumed rate. For example, if we specify that an aircraft can change its vertical flight path angle γ by 20 degrees per second, we implicitly assume that the aerodynamic angle rates can support this specification and that the design of the aircraft and control deflection rates can in turn support those aerodynamic angle rates. In general, the angular rate limits at any level (i.e., flight path, aerodynamic, control surface) are specific to the aircraft under consideration.

The models presented thus far have considered the aircraft as a point mass. The highest fidelity model considers the aircraft as a rigid body, accounts for angular momentum, and allows for 6DOF motion. The derivation of a 6DOF rigid body

aircraft model is beyond the scope of this research. Interested readers should consult the text by Stevens et al. (2015).

At a higher conceptual level, air combat can largely be described as a non-cooperative geometry problem. That is, a major consideration in air combat is the relative positions of the aircraft. The primary angles and vectors of concern are shown in Figure 3 wherein λ is the radar angle, or antenna train angle, ϵ is the aspect angle, and R is the range between aircraft. The geometric position is defined by these values. For example, when the attacker's radar angle and aspect angle are both zero (i.e., $\lambda_A = 0°$, $\epsilon_A = 0°$), the attacker is directly behind and pointed at the target. If the range R is between the minimum and maximum gun range, then the attacker is able to employ guns on the target. The aspect angles are related and represent important features in a zero-sum game, which lends itself to a game theoretic view of air combat.

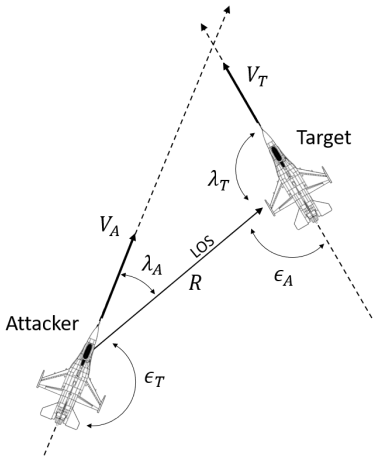
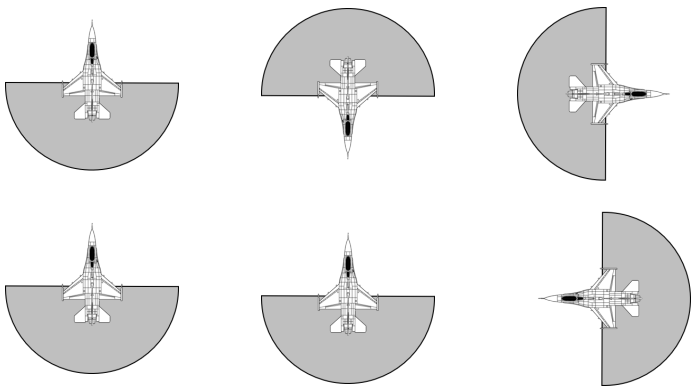


Figure 3. Air Combat Geometry

Taken from one aircraft's perspective, the position situation of air combat can generally be classified as either: advantage, disadvantage, or neutral. Position classifications are based on regions of vulnerability. For fighter aircraft specifically, the traditional region of vulnerability is the rear hemisphere as both guns and missile

are most easily employed against a defender from the rear. For example, when the attacker's radar angle and aspect angle are both zero, the attacker is said to have position advantage. Figure 4a shows this situation wherein the aircraft at the bottom of the figure is the attacker with a position advantage. The aircraft at the top of the figure is the defender with a position disadvantage. The zero-sum relationship between aspect angles results in a mutually exclusive classification of position. When one aircraft has a position advantage, the other aircraft is at a position disadvantage. Figures 4b and 4c show two neutral orientations wherein neither aircraft has the advantage.

During air combat both aircraft maneuver to obtain a position advantage for the purpose of weapons employment. From any geometric position, this involves turning both to deny an attacker position advantage and to gain position advantage. Intuitively, the more maneuverable aircraft in terms of turn rate has a better chance of winning the engagement.



(a) Advantage **(b) Neutral** **(c) Neutral**

Figure 4. Air Combat Position Classifications

Another consideration in air combat is aircraft energy. Energy maneuverability theory was developed by John Boyd and Thomas Christie in the early 1960s as a way to relate the energy-state of the aircraft with agility (David Aronstein, 1997). At any given time, the total specific energy of an aircraft is given by the sum of its potential

energy and kinetic energy, as shown in Equation (4).

$$E_s = h + \frac{V^2}{2g} \tag{4}$$

An aircraft can increase its energy state anytime excess power is available (i.e., thrust is greater than drag). This excess power may be used to gain altitude or increase velocity. Shaw (1985) notes that, in general, being in a higher energy-state relative to the adversary is advantageous. The pilot with more energy has more options available. For example, an aircraft with ample potential energy can quickly gain speed by diving. Conversely, an aircraft with sufficient kinetic energy can climb easily and quickly.

Aircraft energy is related to the geometric position through turn rate. For example, at a constant flight velocity the turn rate can be increased by decreasing altitude (i.e., energy). That is, energy can be exchanged for increased maneuverability and thus a better position. The optimal balance between position and energy is a critical aspect of air combat that is difficult to quantify. Experienced fighter pilots learn to trade effectively between energy and position to maneuver into and out of positions of advantage and disadvantage, respectively.

2.2 Air Combat Weapons

2.2.1 Gun

The most ubiquitous air combat weapon is the gun. First used in 1913, the machine gun was the primary weapon for air combat throughout World War I with most fighters employing two fixed forward-firing .30 caliber machine guns (Shaw, 1985). Aircraft performance and armament improved dramatically during the interwar period. Increased aircraft performance gave pilots the ability to carry heavier weapons

such as larger (in size and number) machine guns and cannons (guns that fired explosive shells). Popular fighters during World War II included the P-51 Mustang armed with six or eight .50 caliber machine guns and the German Me-262 armed with four 30 mm cannons. With the introduction of missiles in the 1950s, focus shifted away from the gun as the fighter's primary weapon. As a result many aircraft, such as the F-4 Phantom II, were designed without a gun. F-4 combat experience in Vietnam proved the value of the gun, and it was reintroduced by the late 1960s (Dwyer, 2014). All modern fighters in the USAF inventory are designed with an internally carried gun although the missile continues to be the primary weapon in air combat (USAF, 2015; Host, 2019). Table 1 lists internal gun parameters for several modern fighter aircraft (Jackson, 2019; Hunter, 2019).

2.2.2 Air-to-Air Missiles

Air to air missiles (AAMs), first introduced in the mid 1950s, consist of four main subsystems: seeker, guidance module, propulsion, and warhead (fuze). AAMs can generally be categorized by employment range and guidance type. Short range AAMs (SRAAMs) are designed for WVR engagements and typically employ infrared (IR) or heat seekers and solid rocket motors. Medium and long range AAMs (MRAAMs and LRAAMs) are designed for beyond visual range (BVR) engagements and gener-

Table 1. Gun Capabilities for Selected Modern Fighters

Aircraft	Gun	Ammunition	Fire Rate (RPM)	1s Bursts
F-15C	M61A1 20mm	940	6000	9.4
F-16	M61A1 20mm	511	6000	5.1
F/A-18	M61A2 20mm	400	6000	4
F-22	M61A2 20mm	480	6000	4.8
F-35A	GAU-22/A 25mm	–	3300	–
J-10B	GSh-23 23mm	–	3500	–
J-20	GSh-301 30mm	150	1800	5
Su-35	GSh-301 30mm	150	1800	5

ally employ one of several radar guidance methods. To maximize range, MRAAMs and LRAAMs use a variety of motors including liquid fuel or even ramjet engines. SRAAMs are the primary AAM of interest for the 1v1 WVR ACMP.

SRAAMs primarily employ IR seekers, which track and guide the missile toward IR sources such as a target aircraft's engine exhaust or, to a lesser extent, its skin, which radiates in the IR band due to aerodynamic heating. Since the mid 1970s, IR seekers have been sufficiently sensitive to lock and track a target from all aspect angles. The range at which a lock is achieved depends on the aspect angle. Engine exhaust provides a stronger IR return than the aircraft's heated skin; thus an IR lock on the rear of a target can be achieved at a much greater distance than from a high aspect approach. For example, the Russian made RVV-MD IR missile is advertised to have a maximum range of 40 km against strong IR targets and a 10 km range against a fighter sized target approaching head on (Udoshi, 2017).

Arguably the most limiting factor of SRAAMs is the seeker field of view (FOV). Generally a target must be contained within the seeker FOV in order for a launch to occur. Early missiles had narrow FOVs that required the aircraft to be pointed at the target long enough for the missile to acquire and track the target. Advances in seeker technology led to improved missiles with FOVs encompassing the entire forward hemisphere. Table 2 contains parameters for selected SRAAMs (Udoshi, 2017).

Table 2. Capabilities for Selected Short Range Missile

		Range (mi)		Seeker FOV			
Missile	Weight (slug)	Min	Max	Launch	Flight	Max Speed (M)	Max G
AIM-9X	5.8	–	6.2	±90°	±90°	–	–
IRIS-T	6.2	–	15.5	±90°	±90°	–	–
A-Darter	6.4	–	12.4	±90°	±90°	–	100
PL-5E	5.7	0.5	9.9	±25°	±40°	2.5	40
PL-9C	7.9	0.5	13.7	±30°	±40°	3.5	40
TY-90	1.4	0.5	3.7	±40°	±40°	2	20
R-60	3.0	–	5.0	±20°	±20°	2	–
RVV-MD	7.3	0.3	24.9	±75°	±75°	–	–

Among the various guidance laws for interceptor missiles, proportional navigation (PN) is the most well known. First developed in the 1950s, PN has seen widespread use among missile systems and is the basis of many modern guidance laws. PN is a straightforward method for homing guidance wherein the commanded missile acceleration is proportional to the rotation rate of the line of sight (LOS) vector between the missile and target. Equations (5a)-(5d) give the commanded acceleration $\vec{a}$ of the missile wherein $\vec{\Omega}$ is the LOS rotation vector, $\vec{V}_r$ is the relative velocity between the target and missile, $\vec{V}_t$ is the velocity of the target, $\vec{V}_m$ is the velocity of the missile, $\vec{R}$ is the range between the target and missile, $\vec{R}_t$ is the range of the target, $\vec{R}_m$ is the range of the missile, and N is the proportionality constant.

$$\vec{a} = N\vec{V}_r \times \vec{\Omega} \tag{5a}$$

$$\vec{\Omega} = \frac{\vec{R} \times \vec{V}_r}{\vec{R} \cdot \vec{R}} \tag{5b}$$

$$\vec{V}_r = \vec{V}_t - \vec{V}_m \tag{5c}$$

$$\vec{R} = \vec{R}_t - \vec{R}_m \tag{5d}$$

Equation (5a)-(5d) provide a simple form of PN that is optimal for a non-maneuvering target but is sub-optimal for maneuvering targets. Several modern guidance laws have augmented simple PN to achieve better results against maneuvering targets (Murtaugh and Criel, 1966).

2.2.3 Directed Energy Weapons

The world's first optical laser was introduced in 1960 by researchers at the Hughes Aircraft Company (Maiman, 1960). Light detection and ranging (LIDAR) systems were developed in the 1970s and 80s and proved useful for mapping and target designation. These systems require very low power and were thus easier to develop. They have since been widely proliferated (Wilson, 2018).

Directed energy weapons (DEWs) are generally more complex systems and require significantly more power. During the 1970s, the first DEW, Airborne Laser Laboratory (ALL), was tested and demonstrated the ability to shoot down missiles in flight. The ALL system utilized a 400 kW chemical laser installed in a KC-135 aircraft (Sabatini et al., 2015). The follow on effort to ALL was the Airborne Laser (ABL) system (shown in Figure 5a), which used a chemical laser mounted in a Boeing 747-400F aircraft. The ABL was capable of firing at most 40 3-5 second laser bursts. The ABL successfully demonstrated the ability to shoot down multiple targets in 2011 (Sabatini et al., 2015).

After the cancellation of the ABL program in 2012 the advanced tactical laser (ATL) program was initiated to adapt the ABL technology for use by smaller aircraft such as the C-130H (shown in Figure 5b) with the goal of developing a 100 kW class laser with range greater than 10 km (Hambling, 2020). Other projects include (1)

DARPA's Aero-Adaptive/Aero-Optic Beam Control (ABC) program, which seeks to develop turrets (shown in Figure 5c) that give 360 degree coverage around an aircraft to enable high-energy lasers to engage other aircraft and incoming missiles, and (2) the Air Force Research Lab's (AFRL) Self-Protect High-Energy Laser Demonstrator (SHiELD) program, which seeks to provide a high power laser within the size, weight, and power confines of an F-15 pod by FY2021 (Keller, 2014; Cohen, 2019). Air Force Special Operations Command (AFSOC) announced that it intends to flight test a 60 kW laser installed in an AC-130J gunship by 2022 (Hambling, 2020).

(a) ABL **(b) ATL** **(c) ABC**

Figure 5. Progression of Aircraft DEW Technology

Clearly, since their introduction in 1960, lasers have become smaller, more available, and more powerful at an increasingly rapid pace. Although not possible today, it is reasonable to assume that a laser weapon similar to that installed on the AC-130J scheduled for test in 2022 could be further developed to fit a fighter aircraft internally in the near future.

For the purpose of this research, we consider aircraft with only an internally carried gun. The inclusion of additional weapons into the MDP model is discussed in Chapter III.

2.3 Physiological Effects of Acceleration

Acceleration is one of the many physical stresses that occur during air combat. Loss of consciousness (LOC) during certain aerobatic flight maneuvers was first reported in 1918. As aircraft performance improved dramatically during the 1980s,

acceleration-induced loss of consciousness (G-LOC) due to a lack of cerebral perfusion (i.e., lack of blood and oxygen to the brain) became a reoccurring issue for fighter pilots that persists today (Newman, 2016).

G-LOC occurs during turning flight when the pilot experiences centripetal acceleration due to the additional force required to turn the aircraft. Consider an aircraft in a coordinated horizontal turn (i.e., zero sideslip angle and altitude remains constant), as shown in Figure 6, where R is the turning radius and $n = {}^{L}/_{W}$ is the load factor. To maintain the horizontal turn, lift L must be increased such that $L\cos\mu = W$. The vector nW opposes the lift and is the apparent weight felt by the pilot, along the z_b axis, where the load factor n is the number of Gs pulled. As an example, a coordinated horizontal turn with bank angle $\mu = 60°$ is a +2 G maneuver. By convention positive Gs (+Gz) occur when the pilot feels the apparent weight pulling down (i.e., when their head points to the inside of the turn) and negative Gs (-Gz) occur when the pilot feels the apparent weight lifting them from the seat (i.e., when their head points to the outside of the turn) such as in a steep dive. G loads along the transverse (Gx) and lateral (Gy) are possible but uncommon during typical flight. All following references to G loads refer to the positive vertical (+Gz) axis unless otherwise specified.

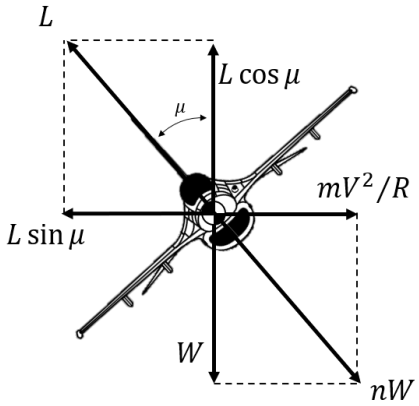


Figure 6. Coordinated Horizontal Turning Flight

Balldin (2002) provides a thorough description of G-LOC. When a pilot experiences a positive acceleration, the apparent weight of their blood is increased, and it becomes difficult for the heart to sustain adequate perfusion to the brain (i.e., blood pressure above the heart drops). Sustained G loads can lead to brain ischemia, or a lack of blood flow to the brain, which results in grayouts (dimming of vision and narrowing of visual fields), blackouts (loss of vision), and after 10 seconds, LOC. Although not as common, negative accelerations (-G) have the opposite effect with regard to arterial blood pressure, inducing a redout or reddish fogging of the vision, increased ocular pressure, and in sever cases, compression of brain tissue. A combination of negative G loads followed by positive G loads, known as the push-pull effect, can significantly reduce a pilot's G tolerance causing them to become more susceptible to G-LOC.

The effect of G loads on pilots has been studied for many years and continues to be an active area of research (Tripp et al., 2006; Whinnery and Forster, 2013; Whinnery et al., 2014; Newman, 2016). Researchers have found that, depending on the rate of G load onset, pilots may experience degraded memory and tracking abilities about 10 seconds prior to the G-LOC event. Pilots remain unconscious for about 12 seconds (absolute incapacitation) and then experience a 12-second period of confusion while regaining consciousness (relative incapacitation). Full task performance recovery from the onset of G-LOC symptoms could take upwards of 87 seconds. In terms of air combat, a G-LOC event would likely prove fatal as even a 30-second period of incapacitation followed by instantaneous recovery would provide an adversary ample time to obtain an advantageous position.

An unprotected human can tolerate about 4G before losing consciousness. Several devices and techniques have been introduced in an effort to improve this tolerance level. The anti-G straining maneuver (AGSM) and anti-G suits (AGS) were intro-

duced in the 1940s and allowed pilots to withstand limited periods of 7G loads. Aircraft of the 1980s and 90s were structurally capable of sustaining 9G maneuvers; however, their maneuverability was limited to about 7G due to the maximum human tolerance with pilot AGSM and early model AGSs. Skilled use of the AGSM paired with modern AGSs allow experienced pilots to operate at levels up to 9G. Performing the AGSM is particularly fatiguing at high G (Newman, 2016). Subsequent advancements in anti-G equipment include the extended coverage AGS (ECGS), balance pressure breathing vests, and assisted positive pressure breathing for G protection (PBG) systems.

ECGSs provide a slight improvement in peak G tolerance. As noted by Newman (2016), researchers have found the ECGS increases a pilots peak G tolerance from about 9.4G with the standard ASG and AGSM to 10.6G with a full ECGS and AGSM (Paul, 1996). The PBG system works to eliminate the fatigue associated with the AGSM at high G. Although it does not eliminate the need to execute an AGSM, it reduces the workload and allows the pilot to tolerate the high G environment for longer periods of time. In simulated air combat maneuver profiles, experienced pilots have successfully endured alternating 10 second periods of 5G and 9G loads for more than 12 minutes (Balldin, 2002). However, even with these advancements G-LOC continues to be an issue in modern fighters. Within the USAF alone, 18 mishaps between 1982 and 1990 were attributed to G-LOC events (Lyons et al., 1992). More recently, G-LOC was determined to be the cause of two F-16 mishaps (Bacon, 2011; Cunningham, 2018), and symptoms of G-LOC onset (known as almost LOC or A-LOC) were determined to cause an F-22 mishap (Eidsaune, 2009).

All modern fighter aircraft have been developed for human pilots and thus have been structurally designed for flight up to about 9Gs, which coincides with the current acceleration tolerance levels of pilots. Advances in autonomy have made possible

the concept of autonomous fighter aircraft that could be designed with structural tolerances greater than 9Gs. Although limited information is available regarding the upper acceleration limit for purpose-built unmanned fighter aircraft, Balldin (2002) notes that recent aircraft improvements have necessitated study of human acceleration up to 12G. Many human centrifuges have been built to withstand sustained 30G loads, and, although not a direct comparison, many missiles can withstand commanded accelerations of 40G or more (Bates and Santucci, 1990; Udoshi, 2017). It is conceivable that an AUCAV may be designed to withstand sustained loads of 12 to 20G or more for an indefinite period of time. This increased loading greatly expands the design space and flight envelope for such aircraft, which could allow for much more aggressive tactics (Anderson, 2008).

2.4 Modeling Approaches

The ACM literature can generally be divided into three broad categories defined by the specific modeling approach employed: operations research methods, control theory, and game theory. Each of these approaches are rooted in different communities: operations research, engineering, and economics, respectively. Aircraft maneuvering has been studied extensively by the engineering community since the 1940s. Until recently, operations research and game theory approaches saw limited application due to the computational complexity of the ACMP and limitations in computer resources.

2.4.1 Operations Research

The field of operations research is concerned with applying analytical methods to decision making problems. The 1v1 WVR ACMP can be viewed as a sequential decision making problem wherein the pilot must make maneuver decisions with the goal of maneuvering to a position of advantage from which weapons can be employed against

the adversary. At the time, dynamic programming (Bellman, 1952) received little attention as a solution procedure for this type of problem due to the computational complexity of relevant air combat scenarios.

McGrew et al. (2010) appear to be the first to demonstrate the feasibility of applying an approximate dynamic programming (ADP) solution procedure to an air combat problem. The authors consider a two-dimensional, altitude restricted, fixed velocity, 1v1 WVR air combat scenario and formulate the problem as an MDP. The system state is defined by the absolute position and orientation of each aircraft while the action space is limited to basic movements across the two dimensional plane: roll-left, maintain current bank angle, or roll-right. The aircraft motion is governed by two dimensional point mass motion equations. The reward structure incentivizes maneuvering to a position aft of and pointed towards the adversary aircraft. Unique to the work is a flight test of the policy with micro unmanned arial systems (UAS) in a lab environment to compare computational trajectories with flight test data.

The reward structure of an MDP formulation of this type of problem defines the weapon engagement zones (WEZs), or the areas from which weapons can be employed on the target. The reward structure used by McGrew et al. (2010) corresponds to a desire to maneuver into a six o'clock position relative to the adversary aircraft (i.e., directly behind the adversary aircraft). Although this is generally considered the most advantageous position in air combat, it is not the only position from which weapons, such as fixed forward-firing guns, can be employed. For example, Shaw (1985) notes that the gun can be employed at almost any aspect although some are more effective than others. In particular, gun shots on the nose are more lethal because of the increased projectile kinetic energy. The reward structure defined by McGrew et al. (2010) is limiting because it defines only a portion of the true gun employment envelope and precludes policies that would employ the gun at high aspect

angles. As an example, several behavior models participating in the Defense Advanced Research Projects Agency (DARPA) AlphaDogfight Trials exploited uncommon, but feasible, gun shots and were able to defeat a trained fighter pilot in simulated combat (Everstine, 2020).

Fang et al. (2016) extend McGrew et al. (2010) by considering the third dimension, altitude. In their model, the state of the system is defined by the absolute position and orientation of each aircraft while the action space consists of seven standard maneuvers: maintain steady flight and six maximum G loading maneuvers including pull-up, dive, left bank, right bank, acceleration, and deceleration. The authors incorporate a 4DOF point mass aircraft model by adding roll control to a 3DOF point mass model. The reward structure borrows from McGrew et al. (2010) but includes an additional term for aircraft energy. Although an improvement of McGrew et al. (2010), two limitations are evident. First, the energy-position tradeoff is defined explicitly in the reward function. Second, the authors use primarily maximum G loading maneuvers as action choices.

The reward structure used by Fang et al. (2016) is similar to McGrew et al. (2010)'s and is limiting for the same reasons. At its core, air combat is a game of position and geometry. The concept of energy maneuverability and its trade off with position was developed from air combat experience and is influenced by existing aircraft capabilities and human experience. This experience has been passed forward to younger generations, and pilots are taught when to maneuver for energy and when to trade that energy for position. By explicitly including a weighted energy score into the reward function and tuning the weights, the behavior of the agent (i.e., computer pilot) is being influenced. Much like the human experience, the agent in these solution approaches is being instructed on a particular trade off based on human experience with existing aircraft. This forced preference mechanism is undesirable for

the purpose of exploring novel AUCAV configurations and maneuver policies that may not yet exist. It is possible that new configurations are best utilized under a different trade off scheme than what is currently taught to human pilots. It is preferable for the aircraft dynamic model to be of sufficient fidelity to implicitly capture energy effects and for the concept of energy maneuverability and energy-position trade-off to be learned by the agent.

Bang-bang guidance describes a system wherein controls are driven to their limits regardless of the magnitude of correction required. Considering only maximum G loading maneuvers in the action space is a form of bang-bang guidance and is a strong assumption for air combat maneuvering as there are scenarios where fine control maneuvers are preferred. For example, an out of plane lag-roll maneuver, as described by Shaw (1985), is used to reduce the closure rate between two aircraft. In this situation the attacking aircraft should pull-up just enough (not necessarily to maximum G loading) to reduce airspeed to facilitate maneuvering into the six o'clock position. Moreover, during a simulation with a time step of 0.01 seconds, sudden large changes in command inputs such as *maximum loading left bank* to *maximum loading right bank* might not actually be possible in an aircraft due to control rate or overall G limits. An action space defined in terms of changes in control inputs, rather than absolute control positions, would alleviate these issues. It should be noted that this type of action space is incompatible with the 4DOF model used by the authors.

Wang et al. (2020) contribute the most recent work in this area. The authors consider the three dimensional 1v1 WVR ACMP with the system state defined by the absolute position, orientation, and velocity of each aircraft, similar to Fang et al. (2016). The action space consists of sets of discrete incremental changes to the flight path and velocity, addressing the bang-bang guidance issues of Fang et al. (2016). The authors incorporate a 3DOF point mass aircraft model. The reward function is

identical to that used by McGrew et al. (2010) but parameterized to represent ranges applicable to full size fighters rather than micro-UASs.

As with any modeling choice, there are advantages and disadvantages with the 3DOF point mass approach. The primary advantage (beyond simplicity) is that it is generally aircraft agnostic. In other words, by specifying flight path rates that are representative of a particular class of aircraft (e.g., fighter aircraft), one can find a high-quality flight path. Utilizing the maneuver generator process described by Snell et al. (1989) and a 6DOF rigid body aircraft model, Wang et al. (2020) are able to generate high-fidelity simulations of a particular aircraft attempting to execute the the high-quality flight path. This approach requires additional analysis to determine how well the aircraft of interest follows the high-quality flight path. It is possible that a high-quality flight path for a 3DOF point mass model is actually a lower-quality path for the 6DOF model or that the 6DOF model cannot reasonably follow the prescribed flight path at all due to features of the 6DOF model that are not accounted for in the lower fidelity 3DOF representation (e.g., limits on control surface rates).

Being platform agnostic is also a disadvantage. Depending on the difference between the flight path rate limits selected for modeling and the flight path rates of the specific aircraft of interest, a 6DOF rigid body simulation may not be able to follow the prescribed flight path exactly. In the worst case, the simulated aircraft will be unable to follow the prescribed path and instead exhibit much lower-quality behavior. In other words, by generalizing the model, we risk finding high-quality maneuver paths, as intended, for a 3DOF point mass model that are actually of low-quality when executed by a subsequent higher-fidelity model, such as a 6DOF rigid body model in the case of Wang et al. (2020). Moreover, this approach does not necessarily capitalize on the full capabilities of any given aircraft because it does not account for specific aircraft parameters.

A secondary disadvantage of the 3DOF point mass model is its implicit restricting of the reward function. For the ACMP, the stochastic optimization process involves construction of reward functions that represent WEZs. The 3DOF point mass model does not control any of the aerodynamic angles shown in Figure 2 and thus precludes the use of reward structures that are functions of these angles.

A simple albeit non-fighter example of this issue is the AC-130 gunship. The gunship contains three weapons: 25mm, 40mm, and 105mm cannons. These cannons are mounted to fire out of the port (left as viewed by the pilot) side of the aircraft at targets on the ground. As depicted in Figure 7, the WEZs for these weapons are clearly a function of the bank angle μ, altitude h_e, and weapon FOV. A 3DOF point mass model that does not specify bank angle is insufficient for the reward function that describes the WEZs of these weapons. For reward structures that utilize the aerodynamic angles, a model with additional DOFs must be used. In the context of air combat, a laser weapon could be an example of a WEZ that is a function of aerodynamic angles. The WEZ of the turret shown in Figure 5a is independent of μ, whereas the WEZs of the turrets shown in Figure 5b and 5c are both functions of μ.

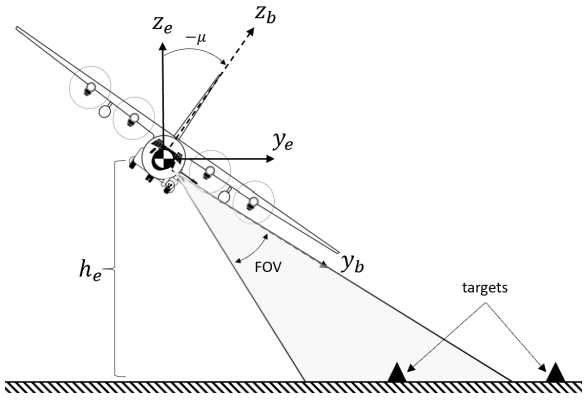


Figure 7. Notional μ Dependent WEZ

2.4.2 Control Theory

The field of control theory (see Frank, 2018) is rooted in the engineering community and is concerned with the control of dynamic systems. The optimization of aircraft trajectories with control theory has been studied for many years. The highly coupled and nonlinear nature of the differential equations governing aircraft dynamics makes the problem of optimizing general maneuvers difficult. Optimization criteria typically involve time to complete a maneuver; linear approximation methods, such as perturbation theory (see Naidu, 2002), have been used to simplify the problem although other numerical methods exist (Lachner et al., 1995). These methods have been applied to small fighter combat problems such as a 180 degree turn to fire a missile, the minimum time to intercept a target, minimum time heading reversal, and a head-on missile duel (Jarmark and Hillberg, 1984; Rajan and Ardema, 1985; Bocvarov et al., 1994; Jarmark, 1985). Another difficulty with control theory solutions to the ACMP is capturing the uncertainty regarding the opposing aircraft's actions. Although these methods are useful for studying maneuvers in idealized scenarios, they have not proven to be well suited for developing full maneuver policies in a 1v1 ACMP scenario.

One possible approach utilizing this work is to develop a number of individualized scenarios for study, find optimal maneuvers for each, and then compile a maneuver library that can be indexed when a particular situation arises in an air combat simulation. A similar process, based on basic fighter maneuvers (BFM) rather than optimized maneuvers, has been implemented in several different air-to-air models with varying results (Burgin and Sidor, 1988). This rule based approach uses a decision matrix to choose the maneuver that maximizes the pilot's objective function (e.g., reduce the distance between the attacker and defender). Two major limitations with early implementations of this approach are the inability to allow for adaptation be-

tween maneuvers and the heavy influence of existing tactics. Each maneuver from the library is treated as an event that, once selected, is executed fully. The resulting state of the air combat is determined, and a new maneuver is selected based on a pre-defined set of engagement phases. In reality, pilots in combat rarely complete these basic maneuvers fully because of the constant changes in the situation as each aircraft respond to the other. To address this issue, Burgin and Sidor (1988) apply a so called trial-maneuver approach that considers smaller "tactical" maneuvers (i.e., continuation of the current flight path, or maximum G turn) rather than complete BFM maneuvers. By shrinking the decision horizon and the scope of a maneuver, the ability to adapt between maneuvers can be achieved. However, the system still suffers from the second limitation.

2.4.3 Game Theory

The field of game theory is concerned with modeling self-interested, rational decision makers and is rooted in the economic community. The analysis of ACMPs can generally be traced back to the homicidal chauffeur pursuit-evasion game originally described by Isaacs (1951). Although pure pursuit-evasion formulations are useful for certain types of air combat problems (e.g., an aircraft evading a missile), they are inappropriate for the 1v1 WVR ACMP. The primary shortfall is that pursuit-evasion assumes distinct roles, that one aircraft is the pursuer and the other is the evader. In other words, the evader never becomes the pursuer. In reality, each aircraft is pursuing the other and both have the opportunity to "win" the engagement.

From the game theoretic perspective, the 1v1 WVR ACMP is a two-target differential game, as discussed by Blaquière et al. (1969) and Getz and Leitmann (1979). The game is commonly viewed as zero-sum with each aircraft using the same target sets (i.e., WEZs), which corresponds to each aircraft possessing the same weapons. The

literature distinguishes between two solution methodologies: qualitative and quantitative (Grimm and Well, 1991). A qualitative approach considers the four possible outcomes of the game (i.e., single kill of either aircraft, mutual kill, or draw) and computes the regions in space from which a particular outcome is guaranteed to occur. The qualitative approach does not yield optimal flight paths to achieve the identified outcomes. Qualitative solutions have been found for variations of the 1v1 ACMP by following the methodology first developed by Isaac (1965) and expanded by Shinar and Davidovitz (1987) (see Merz, 1985; Davidovitz and Shinar, 1985, for examples).

Quantitative approaches tend to use solution procedures common in the control theory field. Austin et al. (1990) use a game-matrix approach to determine maneuver policies for helicopters. Virtanen et al. (2006) use an influence diagram over a limited, moving horizon to determine maneuvers in a 1v1 air combat scenario. As noted by McGrew et al. (2010), longer planning horizons than those implemented by Virtanen et al. (2006) are desired so as to avoid greedy maneuver decisions.

III. Methodology

3.1 Problem Description

Air combat is the primary mission of the United States Air Force (USAF). The most basic form of air combat is the one-versus-one (1v1) within visual range (WVR) engagement. During the course of a typical air combat engagement each aircraft attempts to maneuver into a position from which weapons can be employed. Simultaneously, each aircraft maneuvers to deny the adversary an opportunity to employ weapons. As a first step toward analyzing the general 1v1 WVR air combat maneuvering problem (ACMP) with higher fidelity aircraft and weapons models, we consider the special case wherein the attacking aircraft is a high fast flyer (e.g., an interceptor aircraft or cruise missile) attempting to penetrate a defensive zone and attack high value assets. The defending aircraft is an autonomous unmanned combat aerial vehicle (AUCAV) with limited armament (i.e., only an internally carried gun representative of currently fielded technologies).

In a defensive counter air (DCA) or airborne missile defense (AMD) operation, a friendly (i.e., blue) AUCAV patrols a bounded area of responsibility (AOR) forward of friendly assets that may be vulnerable to attack, such as tanker, mobility, or intelligence, surveillance, and reconnaissance aircraft. The blue AUCAV loiters in the AOR for the duration of the assigned mission, which may be limited by time or fuel status. We assume the blue AUCAV carries external fuel tanks when patrolling the AOR and immediately drops them upon an incursion by an attacking (i.e., red) aircraft. Hence, we assume the blue AUCAV begins all engagements with a full fuel tank.

When a red aircraft enters the AOR, the blue AUCAV moves to engage the threat. We assume the red aircraft's mission is to pass through the AOR, employing one of

two evasive maneuver schemes. The first evasion tactic is employed as follows. Upon entry into the AOR, the red aircraft is flying at cruising speed. If the blue AUCAV approaches within an alert distance d_1 of the the red aircraft, it evades by increasing its velocity by an amount proportional to the relative velocity between itself and the blue AUCAV (i.e., the closing speed of the blue AUCAV). For example, suppose the red aircraft is cruising at 530 $^{\text{ft}}/_{\text{s}}$ with an alert distance of $d_1 = 5,280$ ft (i.e., one mile). As the blue AUCAV approaches within one mile of the red aircraft at a velocity of 1,000 $^{\text{ft}}/_{\text{s}}$, the red aircraft increases its velocity by $p(1,000 - 530)$ wherein p is the proportion of the relative velocity by which the red aircraft increases its own velocity. In this example, if $p = {}^{1}/_{10}$ the red aircraft increases its velocity by 47 $^{\text{ft}}/_{\text{s}}$. We call this policy the "Burn" evasive maneuver because the red aircraft attempts to escape by only increasing its speed.

The second evasion tactic is a modification of the Burn maneuver. The red aircraft increases its speed as in the Burn maneuver. Additionally, if the blue AUCAV approaches within a second alert distance d_2 of the red aircraft, it evades by increasing its flight path angle and executing a right turn. The red aircraft continues its climbing turn until the blue AUCAV is no longer within the first alert distance d_1, at which point the red aircraft turns back to a heading of 0° to continue its effort to maneuver through the AOR. We call this policy the "Turn & Burn" evasive maneuver because the red aircraft attempts to escape by increasing its speed while simultaneously executing a climbing turn.

The engagement terminates due to one of three conditions: (1) the red aircraft is destroyed by the blue AUCAV, (2) the red aircraft leaves the AOR, (3) the blue AUCAV departs the AOR. The blue AUCAV is considered to have departed the AOR when its fuel reserves are depleted (i.e., it crashes due to fuel). If the red aircraft is destroyed before departing the AOR, the blue AUCAV is credited with a victory. If

the red aircraft departs the AOR, the blue AUCAV is credited with a defeat. If the blue AUCAV departs the AOR, or crashes, it is credited with a loss (i.e., worse than a defeat).

AUCAV aerodynamic and engine performance data are derived from National Aeronautics and Space Administration (NASA) technical reports (Smith et al., 1979; Gilbert et al., 1976; Fox and Forrest, 1993) on several developmental variants of the F-16. While representative of modern jet fighters, the data used in this research is not intended to represent any specific aircraft. The AUCAV is limited to a maximum acceleration load of 18G (i.e., twice the approximate human threshold) for an indefinite amount of time. We assume the AUCAV structure can handle the sustained load with no other limitation on time.

We assume the blue AUCAV and red aircraft have perfect information. That is, the AUCAV knows the exact position and attitude of the target red aircraft and the red aircraft knows the speed of the blue AUCAV at all times. In an operational context this corresponds to having perfect sensors or receiving perfect data from an outside source such as an airborne early warning and control (AWACS) aircraft.

3.2 MDP Formulation

The objective of the Markov decision process (MDP) model is to determine high-quality maneuver policies for the blue AUCAV given a set of weapons and acceleration limitations in order to maximize expected total discounted reward over an infinite horizon. Sequential maneuver decisions are made according to a fixed time step δ^t. Let $\mathcal{T} = \{0, \delta^t, 2\delta^t, ...\}$ be the set of decision epochs over which the engagement occurs and maneuver decisions are made.

3.2.1 State space

We extend Wang et al. (2020) when developing the state space for our MDP model. Let $S_t = (B_t, R_t) \in \mathcal{S}$ denote the state of the system wherein B_t is the status tuple for the blue AUCAV and R_t is the status tuple of the red aircraft at time t.

The state of the AUCAV is given as a tuple

$$B_t = (K_{tB}, G_t), \tag{6}$$

wherein K_{tB} is a tuple describing the kinematic status and G_t is a tuple describing the status of the internally carried gun at time t.

The kinematic status of the AUCAV at time t is given as a tuple

$$K_{tB} = (x_t, y_t, z_t, V_t, \gamma_t, \chi_t, \alpha_t, \mu_t, T_t^{thl}, m_t, f_t), \tag{7}$$

wherein x_t, y_t, z_t is the three dimensional position of the aircraft in the fixed Earth reference frame; V_t is the velocity of the aircraft along the x axis of the wind reference frame; and γ_t and χ_t are the flight path and heading angles in the fixed Earth frame, respectively. The variables α_t and μ_t are respectively the angle of attack and roll angle of the aircraft; T_t^{thl} is the throttle setting, m_t is the mass of the aircraft; and f_t is the amount of fuel remaining.

The status of the internally carried gun at time t is given as a tuple

$$G_t = (a_t^g), \tag{8}$$

wherein a_t^g is the amount of gun ammunition remaining.

The state of the red aircraft is given as a tuple

$$R_t = (K_{tR}), \tag{9}$$

wherein K_{tR} is a tuple describing the kinematic status of the red aircraft at time t.

The kinematic status of the red aircraft at time t is given by

$$K_{tR} = (x_t, y_t, z_t, V_t, \gamma_t, \chi_t), \tag{10}$$

wherein x_t, y_t, z_t is the three dimensional position of the cruise missile in the fixed Earth reference frame; V_t is the velocity of the missile along the x axis of the wind reference frame; and γ_t and χ_t are the flight path and heading angles in the fixed Earth frame respectively.

3.2.2 Action space

The decision of the blue AUCAV at time t is given as a vector

$$x_t^B = (x_t^\alpha, x_t^\mu, x_t^{thl}, x_t^g), \tag{11}$$

wherein x_t^α is the change in angle of attack; x_t^μ is the change in roll angle; x_t^{thl} is the throttle setting, and x_t^g is the binary decision to fire the gun.

Let $\mathcal{Z}^B$ be the set of states for which the blue AUCAV is in the weapon engagement zone (WEZ), which is given by

$$\mathcal{Z}^B = \{S_t : \lambda_{t,B} < 30^\circ, 500 \leq R(B_t, R_t) \leq 3000\} \tag{12}$$

wherein $\lambda_{t,B}$ is the radar angle of the blue AUCAV and $R(B_t, R_t)$ denotes the range between the blue AUCAV and the red aircraft at time t.

For the firing decision x_t^g, 0 indicates a decision to not fire the gun and 1 indicates a decision to fire the gun. When the blue AUCAV is in the WEZ (i.e., $S_t \in \mathcal{Z}^B$) it can employ weapons against the target, i.e., $x_t^g \in \{0, 1\}$. When the blue AUCAV is not in the WEZ (i.e., $S_t \notin \mathcal{Z}^B$) it cannot employ weapons against the target, i.e., $x_t^g = 0$. The aerodynamic angle rates are a function of the specific aircraft being modeled and the selected time step δ^t. Let α_i^L and α_i^U denote the lower and upper bound on the angle of attack.

The discrete sets of allowable state dependent actions for each decision variable are given by

$$\mathcal{X}^{\alpha}(S_t) = \left\{x_t^{\alpha} \in \{-\delta^{\alpha}, {}^{-\delta^{\alpha}}/_2, {}^{-\delta^{\alpha}}/_5, 0, {}^{\delta^{\alpha}}/_5, {}^{\delta^{\alpha}}/_2, \delta^{\alpha}\} : \alpha^L \leq \alpha_t + x_t^{\alpha} \leq \alpha^U\right\}, \quad (13a)$$

$$\mathcal{X}^{\mu}(S_t) = \{-\delta^{\mu}, {}^{-\delta^{\mu}}/_2, {}^{-\delta^{\mu}}/_5, 0, {}^{\delta^{\mu}}/_5, {}^{\delta^{\mu}}/_2, \delta^{\mu}\}, \quad (13b)$$

$$\mathcal{X}^{thl}(S_t) = \{0, 0.2, 0.4, \ldots, 1\}, \quad (13c)$$

$$\mathcal{X}^{g}(S_t) = \begin{cases} \{0, 1\} & \text{if } a_t^g > 0, S_t \in \mathcal{Z}^B \\ \{0\} & \text{otherwise} \end{cases}, \quad (13d)$$

Equation (13a) indicates that the AUCAV can increase, decrease, or leave unchanged the angle of attack by a proportion (i.e., $1, {}^1/_2, {}^1/_5$) of the discrete amount δ^{α}, up to the lower or upper bound. For example, the F-16 is limited to a maximum angle of attack of $\alpha^U = 25°$. If the aircraft is currently at this condition, the set of allowable actions for the angle of attack control is $\mathcal{X}_t^{\alpha} = \{-\delta^{\alpha}, {}^{-\delta^{\alpha}}/_2, {}^{-\delta^{\alpha}}/_5, 0\}$ (i.e., the aircraft is not allowed to exceed the limit $\alpha^U = 25°$). Equation (13b) indicates that the AUCAV can increase, decrease, or leave unchanged the roll angle by a proportion (i.e., $1, {}^1/_2, {}^1/_5$) of the discrete amount δ^{μ}. Equation (13c) indicates that any discrete throttle position can be selected in any given time step. Equation (13d) indicates that the AUCAV can choose to either fire or not fire the internally carried gun as long as ammunition

is available and the target is in the WEZ.

Let $\mathcal{J} = \{1, 2, \ldots, 4\}$ be the index set of the decision variables in the decision vector x_t^B (i.e., $x_{t,1}^B = x_t^\alpha, x_{t,2}^B = x_t^\mu$, and so on) and their corresponding decision spaces (i.e., $\mathcal{X}_1 = \mathcal{X}^\alpha(S_t), \mathcal{X}_2 = \mathcal{X}^\mu(S_t)$, and so on). Let L_t^G be the G load on the AUCAV at time t and L_{max}^g be the maximum allowable G load on the AUCAV. The set of feasible decisions for the blue AUCAV at time t can be written as

$$\mathcal{X}^B(S_t) = \left\{(x_{tj}^B)_{j\in\mathcal{J}} : x_{tj} \in \mathcal{X}_j(S_t), L_t^G(S_t, x_t) \leq L_{max}^g\right\}. \tag{14}$$

Equation (14) indicates that allowable decisions include those that do not exceed the aerodynamic angle and G load limits of the aircraft.

Let $\mathcal{I} = \{B, R\}$ denote the set of aircraft in the air combat scenario wherein B indicates the blue AUCAV and R indicates the red aircraft. Let $X^{\pi^i} : \mathcal{S} \rightarrow \mathcal{X}^i(S_t), S_t \in \mathcal{S}$ denote the decision function based on policy π^i that maps the state space to the action space for the ith aircraft. That is, $X^{\pi^B}(S_t)$ returns the maneuver decision x_t^B the blue AUCAV takes when in state S_t at time t. Similarly, $X^{\pi^R}(S_t)$ returns the maneuver decision x_t^R of the red aircraft when in state S_t at time t. In this research we train the blue AUCAV against a red aircraft employing fixed, high-performing benchmark policies from the literature.

3.2.3 Transitions

Let $\Delta_{o\in\mathcal{O}}$ denote the terminal state for the outcome $o \in \mathcal{O} = \{R_K, B_D, R_D\}$ of the engagement, wherein R_K indicates the red aircraft was killed by the blue AUCAV and R_D (B_D) indicates the red (blue) aircraft departed the AOR.

The AUCAV makes maneuver decisions during each time step of the engagement. Manuever and weapon firing decisions occur deterministically. Once a weapon is fired, the outcome of the shot (i.e., whether or not the shot results in a kill) occurs

stochastically. For example, if the AUCAV chooses to fire the gun the system may transition to a terminal state with some probability that is a function of the aspect angle of the AUCAV to the target. Let S^M denote the system model. The general state transition function is written as

$$S^M(S_t, x_t^B, W_{t+1}) = S_{t+1}, \tag{15}$$

wherein W_{t+1} indicates the stochastic exogenous information of whether or not the weapon shot destroyed the target.

Following the previously defined state decomposition, the state of the system at time $t+1$ is given by

$$S_{t+1} = (B_{t+1}, R_{t+1}). \tag{16}$$

The state of the blue AUCAV at time $t+1$ is given as a tuple

$$B_{t+1} = (K_{t+1,B}, G_{t+1}). \tag{17}$$

Let K_B^M denote the kinematic model of the blue AUCAV, which governs the deterministic evolution of the kinematic state. Hence

$$K_B^M(S_t, x_t^B) = K_{t+1,B}, \tag{18}$$

wherein $K_{t+1,B}$ is computed from the 5DOF kinematic model as previously introduced. The model is re-expressed as follows.

$$x_{t+1} = x_t + (V_{t+1} \cos\gamma_{t+1} \cos\chi_{t+1})\, \delta^t \tag{19a}$$

$$y_{t+1} = y_t + (V_{t+1} \cos\gamma_{t+1} \sin\chi_{t+1})\, \delta^t \tag{19b}$$

$$z_{t+1} = z_t + (V_{t+1} \sin\gamma_{t+1})\, \delta^t \tag{19c}$$

$$V_{t+1} = V_t + \frac{\delta^t}{m_t}\left(T\cos\alpha_{t+1} - D - W\sin\gamma_t\right) \tag{19d}$$

$$\gamma_{t+1} = \arctan\left(\frac{{}^{\delta^t}\!/_{m_t}\left(T\sin\alpha_{t+1} + L\cos\mu_{t+1} - W\cos\gamma_t\right)}{V_{t+1}}\right) \tag{19e}$$

$$\chi_{t+1} = \arctan\left(\frac{{}^{\delta^t}\!/_{m_t}\left(L\sin\mu_{t+1}\right)}{V_{t+1}}\right) \tag{19f}$$

$$\alpha_{t+1} = \alpha_t + x_t^{\alpha} \tag{19g}$$

$$\mu_{t+1} = \mu_t + x_t^{\mu} \tag{19h}$$

$$T_{t+1}^{thl} = x_t^{thl} \tag{19i}$$

$$m_{t+1} = m_t - \left(x_t^g r^g m^g\right)\delta^t \tag{19j}$$

$$f_{t+1} = f_t - \left(cT\right)\delta^t \tag{19k}$$

The above equations are presented in the order of appearance in the tuple $K_{t+1,B}$, not in the order of calculation. For example, Equation (19d) and (19e) must be computed before Equation (19c).

Let G^M denote the deterministic gun ammunition transition model of the AUCAV.

$$G^M(S_t, x_t^B) = G_{t+1}, \tag{20}$$

wherein $G_{t+1} = a_{t+1}^g$ is computed as follows

$$a_{t+1}^g = \begin{cases} a_t^g - r^g\delta^t & \text{if } x_t^g = 1 \\ a_t^g & \text{otherwise} \end{cases}, \tag{21a}$$

wherein r^g is the rate of fire of the gun. Equation (21a) indicates the ammunition should be reduced if the decision is made to fire the gun.

Let K_R^M denote the kinematic model of the red aircraft, which governs the deter-

ministic evolution of the kinematic state. Hence

$$K_R^M(S_t, \pi^R) = K_{t+1,R}, \tag{22}$$

wherein π^R is the predetermined policy of the red aircraft. The red aircraft actions are deterministic reactions to the decisions of the blue AUCAV and thus are included in the red aircraft transition function. The tuple $K_{t+1,R}$ is computed from the 3DOF kinematic model previously introduced. The model is re-expressed as follows.

$$x_{t+1} = x_t + (V_{t+1} \cos\gamma_{t+1} \cos\chi_{t+1})\,\delta^t \tag{23a}$$

$$y_{t+1} = y_t + (V_{t+1} \cos\gamma_{t+1} \sin\chi_{t+1})\,\delta^t \tag{23b}$$

$$z_{t+1} = z_t + (V_{t+1} \sin\gamma_{t+1})\,\delta^t \tag{23c}$$

$$V_{t+1} = V_t + x^V \tag{23d}$$

$$\gamma_{t+1} = \gamma_t + x^\gamma \tag{23e}$$

$$\chi_{t+1} = \chi_t + x^\chi \tag{23f}$$

Note the above equations are presented in the order of appearance in the tuple $K_{t+1,R}$, not in the order of calculation. For example, Equation (23d) and (23e) must be computed before Equation (23c).

The actions of the red aircraft are given as a vector $x_t^R = (x_t^\gamma, x_t^\chi, x_t^V)$ wherein x_t^γ is incremental change in flight path angle, x_t^χ is the incremental change in heading, and x_t^V is the incremental change in velocity. In this research the red aircraft employs one of two fixed, predetermined policies: $X^{\pi^R_{\text{Burn}}}(S_t)$ and $X^{\pi^R_{\text{Turn \& Burn}}}(S_t)$. The policy $X^{\pi^R_{\text{Burn}}}(S_t)$ describes a *Burn* evasion maneuver and is given by

$$x_t^\gamma = 0 \tag{24a}$$

$$x_t^{\chi} = 0 \tag{24b}$$

$$x_t^{V} = \begin{cases} p(V_t^B - V_t) & \text{if } R(B_t, R_t) \leq d_1, V_t^B \geq V_t \\ V^{\text{cruise}} - V_t & \text{if } R(B_t, R_t) > d_1, V_t \geq V^{\text{cruise}} \\ 0 & \text{otherwise} \end{cases}, \tag{24c}$$

wherein $p \in [0, 1]$ is the proportion of the relative velocity by which the red aircraft increases its speed by (i.e., the *Burn maneuver parameter*); V_t^B is the velocity of the blue AUCAV; the function $R(B_t, R_t)$ indicates the range between the blue AUCAV and the red aircraft; d_1 is the alert distance of the red aircraft; and V^{cruise} is the cruise speed of the red aircraft. Equations (24a) and (24b) indicate the red aircraft does not change its flight path angle or heading in any case. Equation (24c) indicates the red aircraft increases its speed proportionally to the closing speed of the approaching aircraft if it is within its alert radius d_1. If the approaching aircraft maneuvers outside of the alert radius, the red aircraft immediately returns to its cruising speed. If the red aircraft is not threatened, it remains at its cruise velocity. If the Burn maneuver parameter $p = 0$, the red aircraft is a non-maneuvering, constant velocity target.

The second policy $X^{\pi^R_{\text{Turn \& Burn}}}(S_t)$ describes a *Turn & Burn* evasion maneuver and is given by

$$x_t^{\gamma} = \begin{cases} 0.1 & \text{if } R(B_t, R_t) \leq d_2, \gamma_t = 0 \\ -0.1 & \text{if } R(B_t, R_t) > d_1, \gamma_t > 0 \\ 0 & \text{otherwise} \end{cases} \tag{25a}$$

$$x_t^{\chi} = \begin{cases} 10 & \text{if } R(B_t, R_t) \leq d_2 \\ -10 & \text{if } 180 \geq \chi_t > 0, R(B_t, R_t) > d_1 \\ 10 & \text{if } 180 \leq \chi_t < 360, R(B_t, R_t) > d_1 \\ 0 & \text{otherwise} \end{cases} \tag{25b}$$

$$x_t^{V} = \begin{cases} p(V_t^B - V_t) & \text{if } R(B_t, R_t) \leq d_1, V_t^B \geq V_t \\ V^{\text{cruise}} - V_t & \text{if } R(B_t, R_t) > d_1, V_t \geq V^{\text{cruise}} \\ 0 & \text{otherwise} \end{cases}, \tag{25c}$$

wherein $d_2 < d_1$ is a second alert distance dictating when the red aircraft should execute a climbing turn away from the incoming AUCAV. Equation (25a) indicates the red aircraft should begin a 0.1° climb when the AUCAV is within its alert radius d_2 and return to level flight once the AUCAV is at least a distance d_1 away (i.e., it has fallen outside of the first alert radius). Equation (25b) indicates the red aircraft should turn away from the incoming AUCAV at a rate of 20°s of heading per second. If the AUCAV maneuvers outside of the first alert radius d_1, the red aircraft begins a turn back toward a heading of 0° to continue its effort to maneuver through the AOR. Equation (25c) indicates the red aircraft should speed up using the same procedure as in the Burn maneuver policy. In general, this policy forces the red aircraft to increase its speed when the AUCAV closes to within a distance d_1 of the red aircraft. If the AUCAV closes to within a distance d_2, the red aircraft continues to speed up and begins a climbing right turn away from the AUCAV.

3.2.4 Contributions

Contributions (i.e., rewards) are earned each time step. The contribution function is defined as follows

$$C(S_t, x_t^B, S_{t+1}) = \begin{cases} \nu & \text{if } S_t \notin \{\Delta_o\}_{o\in\mathcal{O}}, S_{t+1} = \Delta_{R_K} \\ -\nu & \text{if } S_t \notin \{\Delta_o\}_{o\in\mathcal{O}}, S_{t+1} = \Delta_{B_D} \\ -\zeta & \text{if } S_t \notin \{\Delta_o\}_{o\in\mathcal{O}}, S_{t+1} \notin \{\Delta_o\}_{o\in\mathcal{O}} \\ 0 & \text{otherwise} \end{cases}, \tag{26}$$

wherein ν is a large positive contribution and ζ is a small contribution relative to ν (i.e., $\zeta \ll \nu$). Equation (26) indicates that, if the blue AUCAV destroys the red aircraft, it earns a large positive contribution. A transition to this terminal state is considered a victory for the blue AUCAV. If the red aircraft departs the AOR, the blue AUCAV earns no contribution and the engagement terminates. A transition to this terminal state is considered a defeat for the blue AUCAV because it allowed the target to pass through the AOR. If the blue AUCAV departs the AOR, it earns a large negative contribution. A transition to this terminal state is considered a total loss. For each time step in the engagement (i.e., when the system has not yet reached a terminal state), the blue AUCAV earns a small negative contribution to incentivize the AUCAV to seek the positive contribution. After transitioning to a terminal state, the blue AUCAV earns no further contributions. For compactness, let $C(S_t, x_t^B) = \mathbb{E}\left[C(S_t, x_t^B, S_{t+1})\right]$.

3.2.5 Objective function and optimality equation

The objective of the MDP model is to determine the optimal maneuver policy π^{*B} that maximizes the expected total discounted contribution, given by

$$\max_{\pi^B \in \Pi^B} \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t C\left(S_t, X^{\pi^B}(S_t)\right)\right], \tag{27}$$

wherein $\gamma \in (0, 1]$ is the fixed discount factor. For clarity, note that the discount factor γ is different than the flight path angle γ_t. The optimal blue AUCAV maneuver policy π^{*B} satisfies the following optimality equation

$$V(S_t) = \max_{x_t^B \in \mathcal{X}^B(S_t)} \left(C(S_t, x_t^B) + \gamma \mathbb{E}\left[V(S_{t+1})|S_t\right]\right), \tag{28}$$

wherein $V(S_t)$ denotes the value of being in state S_t at time t.

3.3 ADP Solution Procedure

The high dimensionality and continuous nature of the state space make solving for π^{*B} using exact dynamic programming intractable. Instead, we employ an approximate policy iteration (API) algorithmic strategy to find high-quality maneuver policies for the blue AUCAV based on approximate value functions. This research uses basis functions to capture important features of the ACMP and replaces the true value function with a statistical approximation based on neural network (NN) learning within the API framework. Our development follows that of Jenkins et al. (2020) with two differences. Our implementation (1) approximates the value function around the pre-decision state and (2) utilizes a parameterized rectified linear unit (PReLU) activation function within the NN architecture.

Let $\phi_f(S_t)$ be a basis function where $f \in \mathcal{F}$ is a feature in the set of features

$\mathcal{F}$. The selection of basis functions and features can be difficult, is highly problem dependent, and has a direct impact on solution quality. Our development of basis functions is informed by McGrew et al. (2010) and Wang et al. (2020). The first four basis functions capture the position and orientation of the blue AUCAV.

$$\phi_1(S_t) = x_{tB} \tag{29a}$$

$$\phi_2(S_t) = y_{tB} \tag{29b}$$

$$\phi_3(S_t) = z_{tB} \tag{29c}$$

$$\phi_4(S_t) = \gamma_{tB} \tag{29d}$$

Equations (29a)-(29c) describe the spatial position of the blue AUCAV at time t. Equation (29d) is the flight path angle of the blue AUCAV.

The next six basis functions capture the relative kinematics between the blue AUCAV and the red aircraft.

$$\phi_5(S_t) = \gamma_{tR} \tag{30a}$$

$$\phi_6(S_t) = \chi_{tB} - \chi_{tR} \tag{30b}$$

$$\phi_7(S_t) = V_{tB} - V_{tR} \tag{30c}$$

$$\phi_8(S_t) = x_{tB} - x_{tR} \tag{30d}$$

$$\phi_9(S_t) = y_{tB} - y_{tR} \tag{30e}$$

$$\phi_{10}(S_t) = z_{tB} - z_{tR} \tag{30f}$$

Equation (30a) is the flight path angle of the red aircraft. Equations (30b)-(30f) are the relative heading, velocity, downrange position, crossrange position, and altitude, respectively.

The next four basis functions capture key features of the air combat geometry.

$$\phi_{11}(S_t) = e^{-\frac{|R(B_t,R_t)-R_d|}{\kappa_r}} \tag{31a}$$

$$\phi_{12}(S_t) = 1 - \frac{\lambda_{t,B}}{180} \tag{31b}$$

$$\phi_{13}(S_t) = \frac{a_t^g}{a_0^g} \tag{31c}$$

$$\phi_{14}(S_t) = \epsilon_{t,B} \tag{31d}$$

$$\phi_{15}(S_t) = \dot{\lambda}_{t,B} \tag{31e}$$

The function $R(B_t, R_t)$ indicates the range between the blue AUCAV and the red aircraft; the range parameter R_d indicates the desired weapon employment range; the tunable parameter κ_r controls the rate of decay of the range scaling, and the parameter a_0^g indicates the initial amount of gun ammunition with which the AUCAV begins the engagement. Equation (31a) scales range information to the interval $(0, 1]$ with the highest value achieved when $R(B_t, R_t) = R_d$. After initial testing, the parameter values were fixed at $R_d = 1,000$ and $\kappa_r = 3,000$. Equation (31b) scales radar angle information to the interval $[0, 1]$ with the highest value achieved when $\lambda_{t,B} = 0$, which corresponds to the target being directly ahead of the aircraft. Equation (31c) scales ammunition information to the interval $[0, 1]$ with the highest value achieved when the AUCAV has the maximum amount of ammunition. Equations (31d)-(31e) are the blue AUCAV's aspect angle and the rate of change of the radar angle, respectively.

The next two basis functions capture the energy state of the AUCAV over the course of the engagement. Let

$$\phi_{16}(S_t) = (m_{tB} + f_{tB}) g z_{tB}, \tag{32a}$$

$$\phi_{17}(S_t) = \frac{1}{2}(m_{tB} + f_{tB}) V_{tB}^2, \tag{32b}$$

wherein g is the standard acceleration due to gravity (i.e., 32.174 $^{\text{ft}}/_{\text{s}^2}$) and the mass of the AUCAV is defined as the mass of the aircraft hardware (m_{tB}) and the mass of the remaining fuel (f_{tB}) at time t. Equation (32a) is the potential energy and Equation (32b) is the kinetic energy of the blue AUCAV.

The final two basis functions capture interactions between variables related to weapon employment.

$$\phi_{18}(S_t) = \phi_{11}(S_t)\phi_{12}(S_t)\phi_{13}(S_t) \tag{33a}$$

$$\phi_{19}(S_t) = \lambda_{t,B}\dot{\lambda}_{t,B} \tag{33b}$$

Equation (33a) captures the interaction among the range between the aircraft, the radar angle of the blue AUCAV to the red aircraft, and the amount of ammunition remaining. These three variables constrain weapon employment. Equation (33b) captures the interaction between the radar angle and its rate of change.

Similar to McGrew et al., we explore utilizing all first order interactions and second order terms. This approach leads to NN convergence issues; thus, only selected interaction terms are included in the final basis function vector. The final basis function vector $\phi(S_t)$, with elements $\{\phi_f(S_t)\}_{f\in\mathcal{F}}$, is scaled to the range $[-1, 1]$.

We employ a three layer (i.e., input, hidden, output) feed-forward NN, wherein the hidden layer contains a set of $\mathcal{H} = \{1, 2, \ldots, |\mathcal{H}|\}$ nonlinear perceptron nodes. The input layer produces $|\mathcal{H}|$ outputs, given by

$$\Upsilon_h^{(2)}(S_t) = \sum_{f\in\mathcal{F}} \Theta_{f,h}^{(1)}\phi_f(S_t), \quad \forall h \in \mathcal{H} \tag{34}$$

wherein $\Theta^{(1)} \equiv [\Theta_{f,h}^{(1)}]_{f\in\mathcal{F},h\in\mathcal{H}}$ is an $|\mathcal{F}| \times |\mathcal{H}|$ matrix of weights from the input layer to the hidden layer. Our NN implementation applies a PReLU activation function with

fixed parameter $\alpha = 0.05$,

$$\text{PReLU}(z) = \begin{cases} z & \text{if } z \geq 0 \\ \alpha z & \text{if } z < 0 \end{cases}, \tag{35}$$

which is applied at each perceptron to produce the inputs to the hidden layer

$$Z_h^{(2)}(S_t) = \text{PReLU}\left(\Upsilon_h^{(2)}(S_t)\right), \quad \forall h \in \mathcal{H}. \tag{36}$$

The hidden layer produces a single output given by

$$\Upsilon^{(3)}(S_t) = \sum_{h \in \mathcal{H}} \Theta_h^{(2)} Z_h^{(2)}(S_t), \tag{37}$$

wherein $\Theta^{(2)} \equiv [\Theta_h^{(2)}]_{h \in \mathcal{H}}$ is an $|\mathcal{H}| \times 1$ matrix of weights from the hidden layer to the output layer. The output layer produces a single output by applying the PReLU activation function, which yields a state value function approximation

$$\bar{V}(S_t|\Theta) = \text{PReLU}\left(\Upsilon^{(3)}(S_t)\right), \tag{38}$$

wherein $\Theta = (\Theta^{(1)}, \Theta^{(2)})$ is the compactly represented NN parameter tuple.

For a given set of NN parameters Θ, the NN-API algorithm makes decisions utilizing the policy

$$X_{\text{ADP}}^{\pi^B}(S_t|\Theta) = \underset{x_t^B \in \mathcal{X}^B(S_t)}{\arg\max} \left\{C(S_t, x_t^B) + \gamma \bar{V}(S_{t+1}|\Theta)\right\}. \tag{39}$$

Replacing the value function with the NN statistical approximation in Equation 38 and substituting Equation 39 into Equation 28, we obtain the approximate state value

function

$$\bar{V}(S_t|\Theta) = C(S_t, X^{\pi^B}_{\text{ADP}}(S_t|\Theta)) + \gamma \mathbb{E}[\bar{V}(S_{t+1}|\Theta)|S_t]. \tag{40}$$

Having defined the NN-based approximate value function of the pre-decision state, we proceed with the presentation the NN-API algorithm, shown in Algorithm 1. Our algorithm is a modified version of the algorithm presented by Jenkins et al. (2020), formulated around the pre-decision state.

Algorithm 1 NN-API

Initialize Θ^0 ▷ small, random values near zero
for m=1:M **do** ▷ policy-improvement loop
  for n=1:N **do** ▷ policy-evaluation loop
    Generate random state S_t^n ▷ using state sampling scheme
    Record basis function evaluation $\phi^n = \phi(S_t^n)$
    Record a sample realization of the value $\hat{v}_t^n$ utilizing Equation (41)
  end for
  Compute $\hat{\Theta}^m$ using nonlinear programming iterative solution procedure
  Update Θ^m utilizing Equation (42)
end for
return decision function $X^{\pi^B}_{\text{ADP}}(\cdot|\Theta^M)$ for policy π^B

The algorithm begins by initializing Θ with small, random values near zero. It then enters a policy-evaluation phase wherein, for each iteration $n = 1, 2, \ldots, N$, it generates a random state S_t^n utilizing a deliberately designed state sampling scheme; records the scaled, basis function representation $\phi^n(S_t)$; and computes a sample realization of the value utilizing the current policy

$$\hat{v}_t^n = \max_{x_t^B \in \mathcal{X}^B(S_t)} C(S_t, x_t^B) + \gamma \bar{V}(S_{t+1}|\theta). \tag{41}$$

At the end of the policy-evaluation loop, the algorithm has collected N samples of the value of following the current policy. It then executes a policy-improvement procedure wherein, for each iteration $m = 1, 2, ..., M$, it computes updated NN weights $\hat{\Theta}^m$ using a quasi-Newton nonlinear programming (NLP) optimization procedure described in

detail by Jenkins et al. (2020). It then updates the NN parameter tuple as follows

$$\Theta^m = \alpha_m \hat{\Theta}^m + (1 - \alpha_m)\Theta^{m-1}, \tag{42}$$

wherein Θ^m is the tuple of NN parameters after the mth policy-improvement loop and $\alpha_m \in [0, 1]$ is a learning rate parameter that controls how much emphasis is placed on recent samples. In this research, we adopt a polynomial learning rate given by

$$\alpha_m = \frac{1}{m^\beta}, \tag{43}$$

wherein m is the current policy-improvement loop and β is a tunable parameter that controls the rate of decay of the learning rate.

The general ACMP suffers from a sparse reward problem. Positive rewards are only earned when the blue AUCAV successfully destroys the red aircraft. However, there are many states from which the AUCAV cannot destroy the target (i.e., out of range) and few states from which the AUCAV can destroy the target (i.e., the WEZ). When choosing states randomly across the entire AOR, it is difficult to sample enough rewardable states (i.e., states wherein the target is in the AUCAV's WEZ) to construct a useful statistical approximation of the value. For example, recall the blue AUCAV's WEZ $\mathcal{Z}^B$ is defined as $S_t \in \mathcal{Z}^B \iff \lambda_{t,B} \leq 30°, 500 \leq R(B_t, R_t) \leq 3000$. That is, the blue AUCAV is in a rewardable state when it is at least 500 feet, but no more than 3,000 feet, away from the target and the target is within a 60° cone off the nose. Initial testing indicates that, for this WEZ, less than 1 in 100,000 random state samples is a rewardable state. A statistical approximation of state value based on this type of sample might be $\bar{V}(S_t) = -\zeta$. That is, based on the sample, we expect to receive a small negative reward regardless of the state observed. This is not a useful value function as it indicates that all actions are equal and that there is no possibility

of achieving positive rewards.

A common approach to addressing this issue is reward shaping, modifying the small negative reward to incentivize the agent to move towards rewardable states. For the ACMP in particular, reward shaping usually involves a heuristic measure of relative position (McGrew et al., 2010; Wang et al., 2020; Yang et al., 2019) or an empirically weighted combination of heuristic measures for position and energy (Fang et al., 2016). Although these heuristic measures can improve algorithm performance, they risk biasing the resulting maneuver policies. For example, one common reward shaping heuristic incentivizes maneuvering to a position behind the target aircraft. This heuristic ignores the possibility of snapshots (i.e., high aspect shots) on the target that are sometimes taken in combat (Shaw, 1985). Additionally, in the context of an ACMP, it is more desirable for our algorithm to learn an appropriate position-energy trade-off given the problem parameters rather than specifying it a priori.

To address this issue we propose using a deliberately designed sampling scheme. We note that one of the conditions for a rewardable state is range, defined entirely by the position variables x, y, x of both aircraft. Instead of sampling randomly from the large AOR for which there is little chance of choosing a state such that the range meets the WEZ criteria (a necessary condition for a rewardable state), we sample randomly from a much smaller region in the AOR. We begin with a region small enough such that all sampled states meet the range criteria. This approach gives a higher chance of choosing a rewardable state. The smaller sample region is progressively increased as a function of the policy-improvement loop counter. As the number of policy improvements increases, the sampling region eventually becomes the entire AOR. We call samples drawn from this expanding region *high-quality samples* as they are more likely to be rewardable states. The value function approximation should be valid across the entire AOR, so we supplement our high-quality set with

random samples over the entire AOR. These random samples over the entire AOR are referred to as *low-quality samples* as they are unlikely to be rewardable states.

Our complete sampling scheme selects a combination of high and low-quality samples, given by

$$\hat{N}_m^H = \lceil p_m N \rceil, \tag{44a}$$

$$\hat{N}_m^L = N - \hat{N}_m^H, \tag{44b}$$

wherein $\hat{N}_m^H$ is the number of high-quality points to use in the mth policy-evaluation loop, $\hat{N}_m^L$ is the number of low-quality points, and p_m is the percentage of high quality points needed. We adopt a polynomial rule for the parameter p_m given by

$$p_m = \frac{1}{m^{\beta_p}}, \tag{45}$$

wherein β_p is a tunable parameter that controls the rate of decay of the percentage of high-quality samples required. That is, over time we require less high-quality samples to be deliberately selected because the value of being in these states has been defused to a larger number of surrounding states.

This technique ensures early sample sets have a higher rate of rewardable states to support construction of useful statistical approximations of the value function. Using our gun WEZ, this technique results in the algorithm observing approximately 8% rewardable states when sampling during the first policy-improvement iteration, a significant improvement over the less than 1 in 100,000 rate achieved with random sampling over the entire AOR. Finding useful value function approximations early helps diffuse the value from the few rewardable states to the surrounding states, which provide a gradient of increasing reward the AUCAV can follow to find the large reward. Our reward envelope expansion state sampling scheme implicitly diffuses the

value from the few rewardable states to avoid explicitly diffusing the value with reward shaping.

IV. Computational Results and Analysis

This chapter presents a simplified one-versus-one (1v1) within visual range (WVR) air combat scenario to demonstrate the applicability of our Markov decision process (MDP) model, with its embedded 5 degree of freedom (DOF) point mass aircraft dynamic model, to solving the air combat maneuvering problem (ACMP). We first describe a simple yet representative intercept scenario that is used for the ensuing analysis. We then discuss two benchmark policies, derived from reward shaping functions found in the ACMP literature, that are subsequently compared to policies generated by our approximate dynamic programming (ADP) solution approach. We discuss our evaluation strategy and design of our computational experiments, the first used for efficiently tuning algorithm parameters and the second used for exploring the efficacy and robustness of the resulting policies. Finally, we present and discuss the results of our computational experiments and make a numerical comparison of the ADP policies to the benchmark policies. As an interesting excursion, we provide qualitative comments on the observed maneuvers of the ADP policies as they compare to known aerobatic maneuvers and combat tactics described by Shaw (1985).

4.1 Scenario Description

This examination utilizes a notional scenario wherein a friendly (i.e., blue) autonomous combat aerial vehicle (AUCAV) tasked with a defensive counterair (DCA) mission is patrolling a bounded area of responsibility (AOR). Upon an incursion by an enemy (i.e., red) aircraft (or cruise missile), the blue AUCAV moves to neutralize the target. Once close to the target, the blue AUCAV must destroy the target before it departs the AOR.

Recall the red aircraft employs one of two evasive maneuver policies: "Burn" or

"Turn & Burn". The Burn evasive maneuver policy behaves as follows. If the blue AUCAV is outside of its alert distance, the red aircraft continues along its path, flying at its cruising speed. If the blue AUCAV is within its alert distance, it speeds up by an amount proportional to the relative velocity between the two aircraft. The red aircraft does not alter its initial heading or flight path angle at any time. The Turn & Burn evasive maneuver policy behaves as follows. If the blue AUCAV is outside the first alert distance, the red aircraft continues along its path, flying at its cruising speed. If the blue AUCAV is within the first alert distance, the red aircraft speeds up by an amount proportional to the relative velocity between the two aircraft. If the blue AUCAV is within the second alert distance the red aircraft begins a shallow, turning climb to escape the AUCAV. Once the AUCAV falls back outside the first alert distance, the red aircraft levels off and turns back toward its original heading of 0° in an effort to exit the far side of the AOR.

The blue AUCAV is equipped with an internally carried gun that is capable of firing on targets that are at least 500 feet, but no more than 3,000 feet forward of the aircraft and within 30° of the centerline. These constraints define the AUCAV weapon engagement zone (WEZ), shown as the shaded region in Figure 8. When

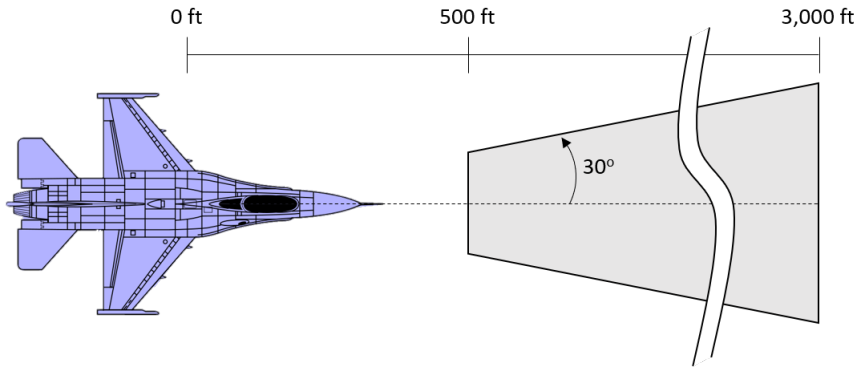


Figure 8. AUCAV Internal Gun WEZ

the AUCAV chooses to fire on the red aircraft it expends ammunition and may or may not destroy the target. The effectiveness of a weapon shot is determined by

the probability of kill (P_k) of the weapon (i.e., the internally carried gun) against the target. For our scenario, we fix the weapon effectiveness (i.e., the P_k values). Figure 9 shows the P_k map for the gun against the red aircraft. For simplicity, we limit the P_k to be a function of the aspect angle only. In general, the P_k could be a more complicated function of many other variables such as range, flight path angle, speed, or G load.

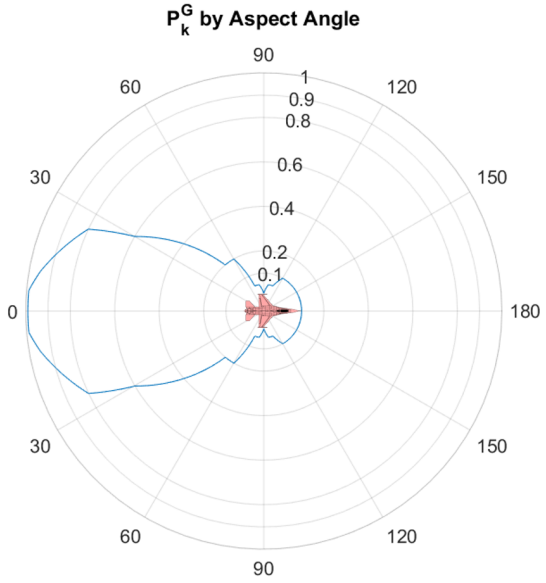


Figure 9. AUCAV Gun P_k Map

4.2 Benchmark Policies

To determine the quality of our resulting ADP policies, it would be ideal to compare them to current air combat maneuvering (ACM) policies. However, due to the highly complex nature of the 1v1 WVR ACMP, there are no simple benchmark policies that accurately capture current practices because the decision making process of military fighter pilots is learned over many years of training. The ideal method for developing a benchmark policy for the 1v1 WVR ACMP is to record pilots flying a series of prescribed engagements in a simulator. One could then directly compare an

ADP policy to the recorded pilot performance. Unfortunately, this type of effort is beyond the scope of this research.

In lieu of a benchmark policy that accurately represents the complete nature of current ACM practices, we construct two benchmark policies that makes decisions myopically according to several reward shaping functions found in the ACMP literature. The first reward shaping function is known as the range score because it scales range information into the interval $(0, 1]$ and is given by

$$S_R = e^{-\frac{|R(B_t,R_t)-R_d|}{\kappa_r}}, \tag{46}$$

wherein $R(B_t, R_t)$ is the range between the blue AUCAV and red aircraft, R_d is a tunable parameter related to the desired weapon employment range, and κ_r is a tunable parameter that controls the decay of the range score as the range increases.

The second reward shaping function is known as the angle score because it scales angle information into the interval $[0, 1]$ and is given by

$$S_A = 1 - \frac{\lambda_B + \epsilon_B}{360}, \tag{47}$$

wherein λ_B is the radar angle and ϵ_B is the aspect angle of the blue AUCAV to the red aircraft.

The third reward shaping function is known as the energy score, given by

$$S_E = \begin{cases} 1 & \text{if } k > 2 \\ 1/2 + \frac{\left(k-k^{-1}\right)}{3} & \text{if } 1/2 \leq k \leq 2 \\ 0 & \text{otherwise} \end{cases}, \tag{48}$$

wherein $k = {}^{H_{E,B}}/_{H_{E,R}}$ is the ratio of specific energy for the blue AUCAV to the red

aircraft and $H_E = H + {}^{V^2}/_{2g}$ is the specific energy, wherein H is the altitude of the aircraft, V is the velocity, and g is the standard acceleration due to gravity. This score attempts to capture the relative energy state between the aircraft and presumes that a higher energy state relative to the adversary is advantageous, a sentiment supported by Shaw (1985).

The first benchmark policy considers only position information and makes myopic decisions according to

$$\underset{x_t^B \in \mathcal{X}^B(S_t)}{\arg\max} \{S_R S_A\}. \tag{49}$$

Equation (49) produces actions that minimize the range and aspect angle, which results in a policy that generally drives the aircraft to a position that is a distance R_d aft of the red aircraft. This position-only benchmark policy has no notion of firing decisions, so we assume it takes all available shots against the target. Experimentation indicates that range score parameter values of $R_d = 1,000$ and $\kappa_r = 2,000$ provide good performance for our problem instances.

The second benchmark policy is a modification of the first that includes the energy score. This position-energy benchmark policy makes decisions according to

$$\underset{x_t^B \in \mathcal{X}^B(S_t)}{\arg\max} \{\omega S_R S_A + (1-\omega) S_E\}, \tag{50}$$

wherein $\omega \in [0,1]$ is an experimentally determined weighting coefficient. Equation (50) seeks to achieve some balance between aircraft position and energy state while still driving towards a position that is a distance R_d aft of the red aircraft. This position-energy benchmark policy has no notion of firing decisions, so we assume it takes all available shots against the target. Experimentation indicates that an energy score parameter weight $w = 0.95$ provides good performance for our problem instances.

4.3 Evaluation Strategy and Experimental Design

We are primarily interested in how the AUCAV maneuvers to destroy the target and not how it transits from its loiter position in the AOR to the target. Hence, we evaluate policies over 132 starting configurations as shown in Figure 10. The blue AUCAV is arranged on the surface of an *engagement sphere* (i.e., the engagement begins when the AUCAV first crosses into the sphere), centered on the red aircraft, with a radius of two miles (10,560 feet). We evaluate how the blue AUCAV intercepts the red aircraft from positions every 30° of heading angle (azimuth) in the interval $[0°, 360°]$ and every 15° of flight path angle (elevation) in the interval $[-75°, 75°]$. The starting positions are selected to determine the best flight path and aspect angle for the blue AUCAV to approach the target. We evaluate the average performance of each policy over all 132 initial states, across 18 problem instances.

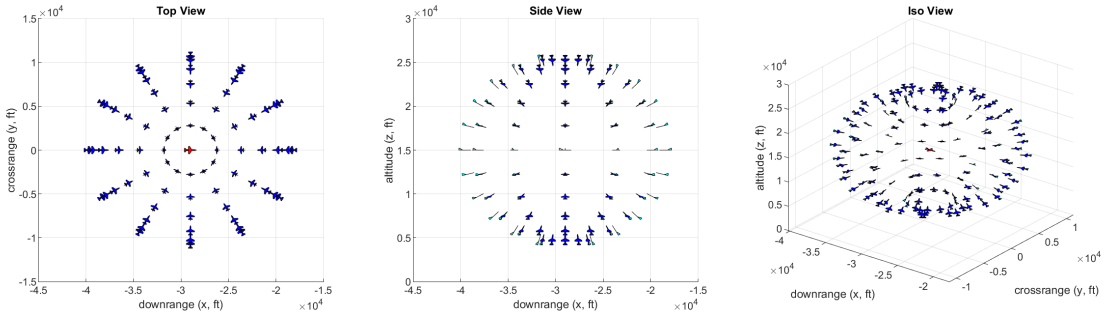


Figure 10. Evaluation Configurations

Problem instances are characterized by the blue AUCAV's initial velocity ($V_{0,B}$) as it crosses the engagement sphere, the cruise speed of the red aircraft (V_R^{cruise}) when not threatened, the Burn maneuver parameter (p), the red aircraft's alert radii (d_1, d_2), the red aircraft's initial altitude ($z_{0,R}$), and the evasion tactic of the red aircraft. Table 3 gives the full factorial experimental design for the problem instance parameters considered in this research. Although there are many more instances that could be examined, we are interested primarily in how the 5DOF aircraft model allows

the ADP policy to manage aircraft energy over the course of the engagement. Hence, we choose to vary the initial velocity of the AUCAV, the burn maneuver parameter, and the red aircraft evasion tactic. All other parameters are fixed. Specifically, the cruise speed is fixed at 530 ft/s, the first alert radius is fixed at 5,280 feet, and the second alert radius is fixed at 4,000 feet. The altitude is fixed relatively low at 15,000 feet to allow room for climbing maneuvers when the Turn & Burn evasion tactic is used.

Table 3. Experimental Design for Problem Instances

Instance	$V_{0,B}$ (ft/s)	p	Evasion Tactic
1	300	0	Burn
2	300	0.05	Burn
3	300	0.1	Burn
4	600	0	Burn
5	600	0.05	Burn
6	600	0.1	Burn
7	900	0	Burn
8	900	0.05	Burn
9	900	0.1	Burn
10	300	0	Turn & Burn
11	300	0.05	Turn & Burn
12	300	0.1	Turn & Burn
13	600	0	Turn & Burn
14	600	0.05	Turn & Burn
15	600	0.1	Turn & Burn
16	900	0	Turn & Burn
17	900	0.05	Turn & Burn
18	900	0.1	Turn & Burn

In our formulation, aircraft dynamics and maneuver decisions occur deterministically whereas weapon shots destroy the target stochastically as specified by P_k data. We leverage these features to achieve computational savings when evaluating policies by computing the expected reward exactly. We track each shot the policy takes and use the P_k data to compute the exact expected reward, expected engagement time, and probability of kill when the AUCAV engages the target from a particular starting

point on the engagement sphere. Our ADP algorithm seeks to maximize expected total discounted reward, which is highly correlated with the average probability of kill over all 132 starting states. Hence, we use the more readily interpretable average probability of kill as our primary metric for reporting results.

Table 4. Experimental Design for ADP Hyperparameter Tuning

Factor	Parameter Setting	Description
M	$\{4, 8, \ldots, 24\}$	Number of policy improvement iterations
N	$\{15000, 20000\}$	Number of policy evaluation samples
$\lvert\mathcal{H}\rvert$	$\{35\}$	Number of NN hidden layer nodes
λ	$\{0.001\}$	NN regularization parameter
β	$\{0.4, 0.5, 0.6\}$	Learning rate parameter
β_p	$\{0.2, 0.4\}$	High-quality sample rate parameter
p	$\{0.05, 0.1\}$	Burn maneuver parameter

Table 4 displays the algorithmic parameters of interest. Initial testing informed our selection of factor levels. Note that the Burn maneuver parameter shown in Table 4 is the parameter used for training the blue AUCAV during the ADP algorithm execution. A full factorial computational experiment is designed and executed on an Intel Xeon E5-2680v3 workstation with 192 gigabytes of RAM. A high discount factor of $\gamma = 0.99$ is used to incentivize the AUCAV to position itself to take advantage of future opportunities.

We now present the results of our computational experiments. We first present and discuss the performance of the policies generated from the hyperparameter tuning experimental design, shown in Table 4, over the first nine problem instances (i.e., against a target performing the Burn evasive maneuver), shown in Table 3. We identify the three superlative hyperparameter settings (i.e., ADP policies) among this initial set of problem instances and then present and discuss the performance of these policies over the last nine problem instances (i.e., against a target performing the Turn & Burn evasion maneuver). Finally, we perform an interesting excursion and give qualitative comments on observed maneuvers.

4.4 Case 1: Red Aircraft Employing Burn Evasion Tactic

Table 5 gives the results of the hyperparameter tuning computational experiment across the first nine problem instances, when the red aircraft is performing the Burn evasive maneuver. The first column gives the ADP policy number. The next six columns give the algorithmic and problem feature parameters. The last nine columns give the superlative mean P_k obtained by the best performing ADP policy generated from the corresponding algorithmic and problem features. Note the problem feature p in Column 7 is the Burn maneuver parameter used during ADP algorithm execution and is not the parameter used when evaluating problem instances. For example, ADP Policy 1 is generated by experiencing a target that behaved with $p = 0.05$ whereas Problem Instance 1 evaluates the resulting policy against a target that behaves with $p = 0$.

The superlative ADP policies across the first nine problem instances are shown in bold font. It is clear that ADP policy 15 performs the best in Problem Instances 1 through 3 whereas ADP policy 19 performs the best in Problem Instances 4 through 9. The top three performing ADP policies across the first nine problem instances are indicated in Column 1 with a dagger symbol. That is, ADP policy 2, 15, and 19 are the best performing policies across the first nine problem instances, when the red aircraft employs the Burn evasion tactic.

Table 6 summarizes the experimental results reported in Table 5 by comparing the top performing parameter combinations (i.e., bolded ADP policies in Table 5) to each benchmark policy, for each problem instance. The three left most columns provide the problem instance number, AUCAV initial velocity, and red aircraft Burn maneuver parameter. For compactness we do not include the evasion tactic employed by the red aircraft in the table because the first nine problem instances use the same evasion tactic, Burn. The next five columns provide the algorithm and problem

feature parameters that generate the superlative policy for the problem instance. The next two columns report the solution quality in terms of percent improvement over the average probability of kill of the benchmark policies. For example, the superlative ADP policy for Problem Instance 1 attains a mean probability of kill that is 16.6% higher than Benchmark Policy 1, and 9% lower than Benchmark Policy 2. The remaining two columns report the computational effort required to obtain the listed policy and the mean P_k achieved by the policy in each instance.

The results from Table 6 indicate that, after relatively few policy iterations, the ADP algorithm is able to generate policies that significantly outperform Benchmark Policy 1 in the first three problem instances when the initial engagement speed is 300 ft/s. Benchmark Policy 1 only considers position information and has no notion of aircraft energy. The benchmark policy has difficulty maintaining airspeed from some starting locations whereas the ADP policy manages its airspeed to give itself a chance to destroy the red aircraft before it exits the AOR. Benchmark Policy 2 considers both position and energy through the inclusion of the energy score. Benchmark Policy 2 outperforms the ADP policy by nearly 9% in all three of the 300 ft/s initial velocity instances. The differences between each policy are displayed in Figure 11, which shows a sample trajectory from a particular starting location (i.e., Entry Point 107, (270,30)) on the engagement sphere for the first problem instance. Figure 12 shows the angle of attack (AOA) setting and velocity for each policy over the duration of the engagements shown in Figure 11.

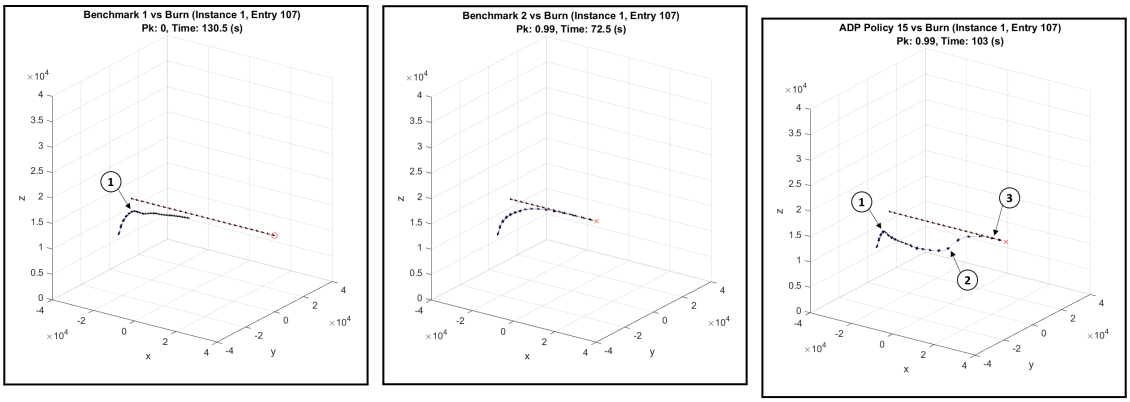
Figure 11a shows the AUCAV trajectory when executing Benchmark Policy 1. Recall that it only considers position information when making decisions and that actions taken during each decision epoch are changes in angle of attack (AOA), roll angle, and throttle setting. The AUCAV begins the engagement in a relatively low energy state (i.e., low altitude and low speed). Between the initial position and Point

Table 5. Hyperparameter Tuning Results

	API-NNR parameters						Superlative Mean P_k by instance								
	N	$\mathcal{H}$	λ	β	β_p	p	1	2	3	4	5	6	7	8	9
1	15,000	35	0.001	0.4	0.2	0.05	0.460	0.463	0.456	0.787	0.761	0.759	0.823	0.831	0.829
2†	15,000	35	0.001	0.4	0.2	0.1	0.803	0.811	0.803	0.940	0.948	0.945	0.923	0.946	0.952
3	15,000	35	0.001	0.4	0.4	0.05	0.678	0.678	0.670	0.873	0.869	0.853	0.892	0.906	0.887
4	15,000	35	0.001	0.4	0.4	0.1	0.599	0.592	0.578	0.805	0.791	0.811	0.825	0.801	0.827
5	15,000	35	0.001	0.5	0.2	0.05	0.473	0.465	0.450	0.821	0.819	0.821	0.861	0.856	0.860
6	15,000	35	0.001	0.5	0.2	0.1	0.673	0.673	0.645	0.911	0.926	0.911	0.919	0.924	0.915
7	15,000	35	0.001	0.5	0.4	0.05	0.612	0.589	0.575	0.866	0.866	0.843	0.836	0.875	0.856
8	15,000	35	0.001	0.5	0.4	0.1	0.729	0.752	0.697	0.850	0.846	0.821	0.804	0.863	0.869
9	15,000	35	0.001	0.6	0.2	0.05	0.048	0.025	0.025	0.505	0.330	0.209	0.716	0.592	0.499
10	15,000	35	0.001	0.6	0.2	0.1	0.441	0.441	0.426	0.824	0.805	0.617	0.800	0.804	0.704
11	15,000	35	0.001	0.6	0.4	0.05	0.630	0.624	0.553	0.845	0.886	0.866	0.858	0.872	0.894
12	15,000	35	0.001	0.6	0.4	0.1	0.653	0.622	0.591	0.843	0.841	0.861	0.847	0.812	0.821
13	20,000	35	0.001	0.4	0.2	0.05	0.023	0.023	0.023	0.311	0.132	0.132	0.567	0.449	0.272
14	20,000	35	0.001	0.4	0.2	0.1	0.023	0.023	0.023	0.301	0.123	0.122	0.581	0.456	0.282
15†	20,000	35	0.001	0.4	0.4	0.05	**0.838**	**0.838**	**0.826**	0.942	0.944	0.939	0.960	0.968	0.967
16	20,000	35	0.001	0.4	0.4	0.1	0.500	0.505	0.497	0.888	0.881	0.875	0.914	0.937	0.936
17	20,000	35	0.001	0.5	0.2	0.05	0.008	0.008	0.008	0.221	0.055	0.055	0.475	0.368	0.220
18	20,000	35	0.001	0.5	0.2	0.1	0.015	0.015	0.015	0.227	0.089	0.088	0.499	0.393	0.214
19†	20,000	35	0.001	0.5	0.4	0.05	0.798	0.788	0.765	**0.961**	**0.969**	**0.949**	**0.982**	**0.978**	**0.989**
20	20,000	35	0.001	0.5	0.4	0.1	0.466	0.466	0.460	0.886	0.886	0.881	0.967	0.960	0.976
21	20,000	35	0.001	0.6	0.2	0.05	0.010	0.010	0.010	0.161	0.044	0.045	0.429	0.341	0.134
22	20,000	35	0.001	0.6	0.2	0.1	0.007	0.007	0.007	0.190	0.038	0.038	0.442	0.332	0.146
23	20,000	35	0.001	0.6	0.4	0.05	0.530	0.530	0.523	0.817	0.824	0.827	0.911	0.931	0.954
24	20,000	35	0.001	0.6	0.4	0.1	0.379	0.366	0.296	0.817	0.811	0.753	0.894	0.929	0.895

Table 6. Superlative ADP Policies vs Burn Evasion Tactic

Instance parameters			API-NNR parameters					Relative (%) improvement over		Computational	ADP
	$V_{0,B}$ (ft/s)	p	M	N	β	β_p	p	Benchmark 1	Benchmark 2	effort (hrs)	Policy P_k
1	300	0	20	20,000	0.4	0.4	0.05	16.5	-9.0	1.99	0.838
2	300	0.05	20	20,000	0.4	0.4	0.05	17.5	-9.1	1.99	0.838
3	300	0.1	20	20,000	0.4	0.4	0.05	19.0	-8.9	1.99	0.826
4	600	0	20	20,000	0.5	0.4	0.05	-3.3	-0.9	2.47	0.961
5	600	0.05	20	20,000	0.5	0.4	0.05	-3.1	-0.3	1.52	0.969
6	600	0.1	20	20,000	0.5	0.4	0.05	-5.1	-2.0	2.47	0.949
7	900	0	20	20,000	0.5	0.4	0.05	0.5	0.9	1.98	0.982
8	900	0.05	12	20,000	0.5	0.4	0.05	0.4	0.3	1.98	0.978
9	900	0.1	16	20,000	0.5	0.4	0.05	1.6	2.8	1.98	0.989



(a) Benchmark Policy 1 (position-only) **(b) Benchmark Policy 2 (position-energy)** **(c) ADP Policy 15**

Figure 11. Sample Trajectories versus Burn Evasion Tactic, Instance 1

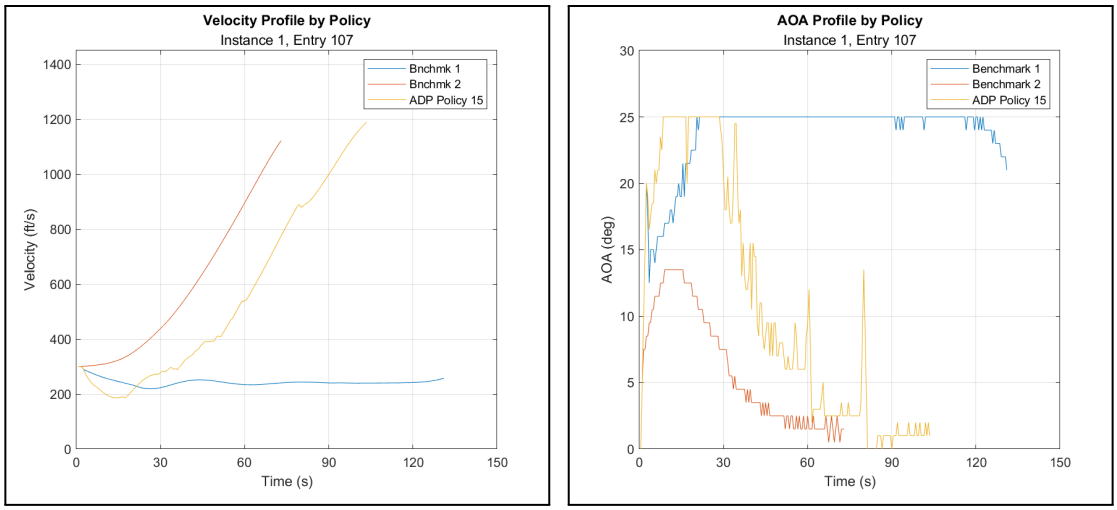
1, it attempts to reduce the distance to the target by climbing at full throttle and commanding increased AOA (i.e., generating the maximum amount of lift). At Point 1 the AUCAV's airspeed (i.e., kinetic energy) is reduced to the point that it can no longer climb, even with maximum throttle (i.e., the AUCAV only has potential energy). The AUCAV begins to level off but still commands maximum AOA and maximum throttle. The AUCAV is unable to increase speed (i.e., kinetic energy) and match the target altitude because it still commands maximum AOA (see Figure 12b). Recall the energy considerations discussed in Chapter II; an aircraft can increase its energy state (i.e., the sum of potential and kinetic energy) anytime excess power is available (i.e., thrust is greater than drag). Although maximum AOA maximizes lift, it also maximizes drag. The AUCAV is unable to increase its energy state, which is required to climb, because the thrust produced by the engines at max throttle is entirely applied to overcoming the increased drag from commanding maximum AOA. The AUCAV continues in a stable trajectory toward the red aircraft unable to increase its speed effectively. The engagement ends when the red aircraft departs the AOR as indicated by the red circle at the end of its trajectory.

Figure 11b shows the AUCAV trajectory when following Benchmark Policy 2, which considers both position and energy information. The AUCAV begins the en-

gagement in the same condition as for Benchmark Policy 1. It commands increased, but not maximum AOA and maximum throttle to climb towards the target. After a short time, the AUCAV begins to reduce AOA while retaining maximum throttle. Recall that the force of lift is a function of both the AOA, through the coefficient of lift, and velocity. A reduction in AOA reduces the lift but also reduces the drag, which allows the AUCAV to increase speed. Following Benchmark Policy 2, the AUCAV is able to continue its climb because it successfully transitioned from generating lift through AOA to lift through velocity (i.e., it efficiently increased its kinetic energy to support a continued climb towards the target). The AUCAV is able to match altitude, chase down, and destroy the target with a high P_k shot.

Figure 11c shows the AUCAV trajectory when following ADP Policy 15. The ADP policy exhibits a type of hybrid behavior with aspects of both benchmark policies. The AUCAV initially commands maximum AOA at a faster rate than Benchmark Policy 1, forcing maximum throttle. At Point 1 in Figure 11c, the AUCAV has expended all of its kinetic energy, faster than Benchmark Policy 1 (see Figure 12a), and it begins to level off, unable to climb. This is evident in Figure 12a as the velocity profile for ADP Policy 15 reaches a minimum (around 15 seconds into the engagement) much quicker than the velocity profile for the Benchmark Policy 1, which reaches a minimum around 28 seconds into the engagement. +

Between Points 1 and 2, the AUCAV reduces AOA to gain speed (i.e., gain kinetic energy). At Point 2 the AUCAV is at a much higher energy state and is able to easily climb up to the target altitude, chase down, and destroy the red aircraft with a high P_k shot about 27 seconds before it departs the AOR. From the initial position to Point 1, the AUCAV behaves similar to Benchmark Policy 1. From Point 2 to Point 3, it behaves more like Benchmark Policy 2 in that it maneuvers to increase its energy state. From Point 3 to the end of the engagement, the AUCAV successfully



(a) Velocity Profile **(b) Angle of Attack Profile**

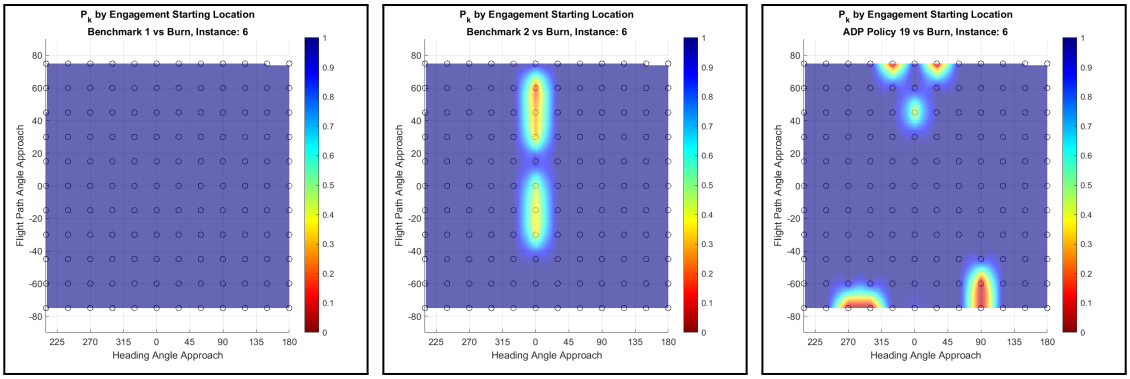
Figure 12. Selected Controls versus Burn Evasion Tactic, Instance 1

converts the excess energy gained during the previous leg into a position advantage in order to employ weapons on the target. The moment the AUCAV initiates the conversion of energy into position advantage is evident in Figure 12a as the slight reduction in velocity about 80 seconds into the engagement and in Figure 12b as the rapid increase in AOA at the same time. It is also clear from Figure 12 how the performance of ADP Policy 15 falls between that of Benchmark Policy 1, which has relatively poor performance for this trajectory, and Benchmark Policy 2, which has high-quality performance for this trajectory.

The ADP policy performance is similar to both benchmark policies for the next three problem instances, wherein the initial engagement velocity is 600 ft/s. Benchmark Policy 1 performs slightly better than both the ADP and Benchmark 2 Policies, achieving a greater than 0.994 mean P_k across these three instances and leaving very little room for improvement. We note that, due to the initial velocity, the initial energy state of the AUCAV is sufficiently higher for these instances such that the position only policy does not have to manage the energy in order to successfully destroy the red aircraft. With an adequate amount of energy from the entry point, it

is possible, and extremely efficient, to consider only position.

Figure 13 shows the expected P_k when approaching the target from any point on the engagement sphere, achieved by following the listed policy. Note the horizontal axis gives the approach heading and the vertical axis gives the approach flight path angle, as seen by the red aircraft. The reader has the perspective of the AUCAV where the red aircraft would be at (0,0) flying out of the page. For example, along the x-axis, 180° indicates approaching the red aircraft from the rear. Along the y-axis, the top row at 75° indicates approaching the red aircraft from the top of the sphere. Figure 13 is a projection of the engagement sphere onto a two dimensional surface where each black circle corresponds to an aircraft shown in Figure 10. The primary utility of these figures is to compactly show from what approach vectors each policy is successful and from which approaches the policies perform poorly. Hence, we call these type of figures P_k *approach maps*. In Figure 13b we observe that Benchmark Policy 2 has difficulty destroying the target when approaching head-on (i.e., 0 degrees heading) from several flight path angles. In contrast, Figure 13a shows that Benchmark Policy 1 achieves a high P_k from any approach, outperforming both Benchmark Policy 2, which considers energy information, and ADP Policy 19. Figure 13c shows the areas where ADP Policy 19 struggles to destroy the target. Figure 13 visually indicates



(a) Benchmark Policy 1 **(b) Benchmark Policy 2** **(c) ADP Policy 19**

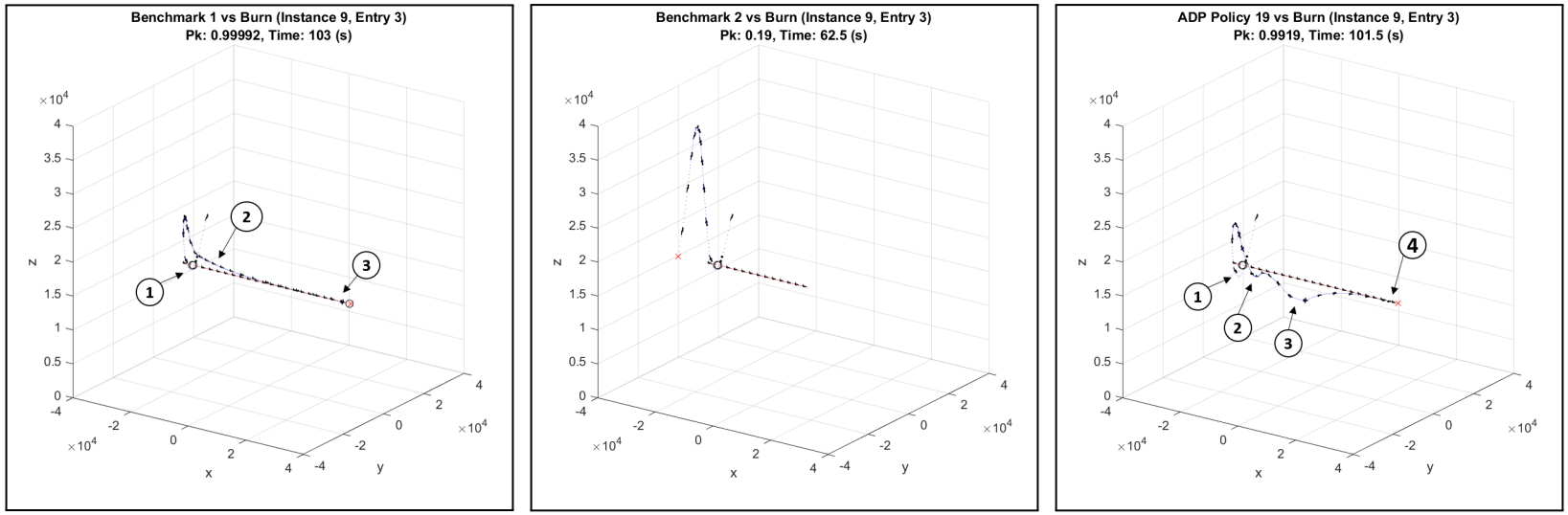
Figure 13. P_k Maps versus Burn Evasion Tactic, Instance 6

that, for this problem instance, there is very little room for improvement over the benchmark policies.

ADP Policy 19 marginally outperforms both benchmarks in the next three problem instances (i.e., Problem Instances 6 through 9) when the initial engagement speed is high, at 900 ft/s. The largest improvement occurs in the final instance against the most stressing Burn evasion tactic. Figure 14 compares a sample trajectory from a particular starting location (i.e., Entry Point 3, (0,60), on Figure 16) on the engagement sphere for this Problem Instance 9. In this case the AUCAV approaches the red aircraft head-on at a steep downward trajectory (i.e., from the top of the sphere). Figure 14b shows that Benchmark Policy 2 actually performs the worst from this position, carrying too much energy through the first pass and maneuvering out of the AOR ending the engagement. On its only pass by the target, Benchmark Policy 2 does get one high-aspect, low P_k snapshot, but misses, as indicated by the open black circle over the red aircraft trajectory.

Benchmark Policy 1 and the ADP policy are both able to destroy the target with $P_k > 0.99$ but have very different trajectories. Figure 14a shows the trajectory for the AUCAV following Benchmark Policy 1. Starting the engagement in a high energy state (due to the high initial velocity) the AUCAV races past the target after getting a single (missed) snapshot. As soon as the AUCAV passes the target (Point 1), the AUCAV pulls over 10 G to pitch up and maneuver back towards the target. Although not apparent in the figure, all of the AUCAV maneuvering takes place in the vertical plane along the x-axis. After pulling over the top, the AUCAV quickly descends to the target altitude (Point 2) and begins a sprint to overtake and destroy the target with a high P_k shot.

The ADP policy marginally outperforms Benchmark Policy 1 in this trajectory as it is able to destroy the target with a similar P_k but nearly 3 seconds faster. Figure 15



(a) Benchmark Policy 1 (b) Benchmark Policy 2 (c) ADP Policy 19

Figure 14. Sample Trajectories versus Burn Evasion Tactic, Instance 9

shows the trajectory for the AUCAV following ADP Policy 19. The AUCAV races past the target, taking a low P_k shot along the way, similar to Benchmark Policy 1. Instead of pulling up immediately upon passing the target, the AUCAV holds the downward trajectory slightly longer, reaching a lower altitude before pulling up. This delayed pull allows the AUCAV to gain slightly more kinetic energy (speed) than when following Benchmark Policy 1. Figure 15 shows the slight difference in kinetic energy and velocity around the 10 second mark. Although a minor difference, this allows the AUCAV to carry more speed at the top of the trajectory (i.e., 20 second mark) when it pitches over, back toward the target. At the 45 second mark, the AUCAV reaches Point 2, a position lower than the target altitude. This maneuver past the target altitude increases the kinetic energy further, allowing the AUCAV to begin closing on the target earlier than with Benchmark Policy 1 (see Figure 15). The AUCAV uses some of the increased kinetic energy to gain altitude and executes

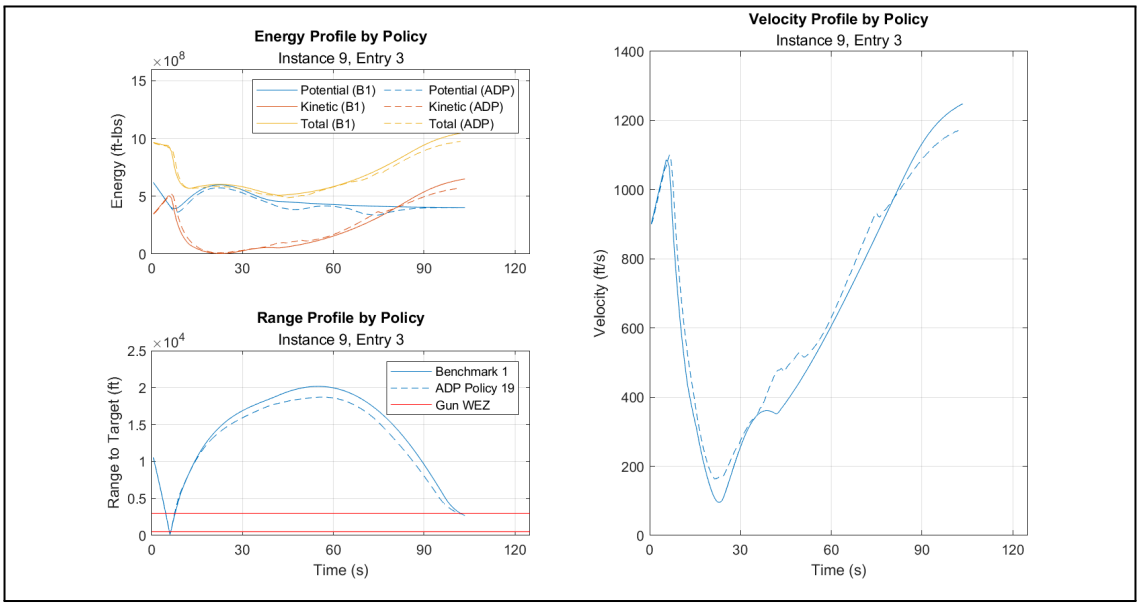
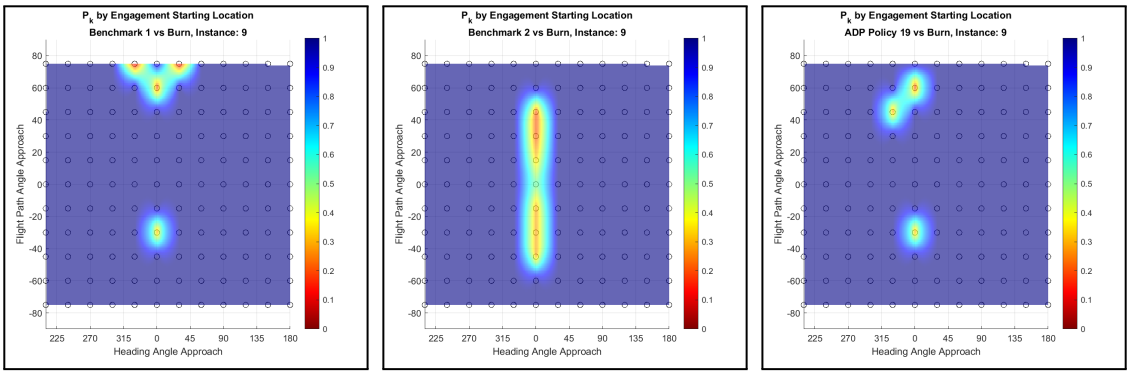


Figure 15. ADP Policy 19 Energy-Range Profile, versus Burn Evasion Tactic, Instance 9

another energy maneuver (Point 3) at the 75 second mark, increasing its velocity further and allowing the AUCAV to reach the target 3 seconds earlier than under

Benchmark Policy 1. In contrast to Benchmark Policy 2, which considers energy information through an a priori weighting scheme, this sample trajectory highlights the energy-maneuver behaviors possible with our 5 DOF MDP formulation that are difficult to capture with simple scores. This sample trajectory also highlights how small changes in maneuvers early in the engagement (i.e., Points 1 and 2) can result in better performance near the endgame.

Figure 16 shows the P_k approach maps for each policy. We note that Benchmark Policy 2 does not perform well in the same area as in the medium speed (600 ft/s) instances. Benchmark Policy 1 and ADP Policy 19 perform similarly.



(a) Benchmark Policy 1 **(b) Benchmark Policy 2** **(c) ADP Policy 19**

Figure 16. P_k Maps versus Burn Evasion Tactic, Instance 9

Although our ADP policies do not outperform the benchmarks in all of the first nine problem instances, the difference in terms of average probability of kill are small for most instances. The tuned benchmarks perform very well for the selected problem instances and do not leave much room for improvement in many problem instances. We consider the marginal performance of the ADP policies relative to the benchmarks an indication of the high-quality nature of the benchmarks for handling a scenario wherein the red aircraft executes the simple *Burn* evasive maneuver. These results illustrate that our reward envelope expansion sampling scheme is successful in implicitly diffusing the rewards from the small number of rewardable states in our model

development. That is, our ADP algorithm generated policies that achieve performance similar to common ACMP reward shaping functions without explicitly biasing the reward structure. Moreover, as the results indicate, the 5DOF MDP formulation can successfully generate subtle yet complex energy maneuvers not captured by reward shaping functions.

Having presented the results over the first nine problem instances given in Table 3 and identified the top three performing ADP policies in Table 5 against a simply maneuvering target, we now examine the robustness of the ADP policies. The next section presents the results for the three superlative ADP policies, as indicated with a dagger in Column 1 of Table 5, against a more maneuverable adversary employing the Turn & Burn evasive manuever.

4.5 Case 2: Red Aircraft Employing Turn & Burn Evasion Tactic

To explore the robustness of the ADP policies, we compare the performance of the three superlative ADP policies from the previous section against the benchmarks when facing a more evasive target (i.e., Problem Instances 10 through 18). In the following problem instances, the red aircraft employs the Turn & Burn evasion tactic, which extends the Burn tactic. The red aircraft accelerates as before if the AUCAV is within a distance d_1 of the red aircraft. When the AUCAV approaches within a distance d_2, the red aircraft begins a shallow climb and turns to the right at a rate of 20° of heading per second as long as the AUCAV is within distance d_1. Once the AUCAV is outside of the first alert distance d_1, the red aircraft levels off at the new altitude and turns back to a heading of 0° to continue on toward its targets, outside of the immediate AOR.

We select the top three policies from Table 5, indicated with a dagger next to the policy number, and evaluate those against a red aircraft employing the Turn & Burn

evasion tactic. Table 7 summarizes these results. The first column gives the problem instance, which corresponds to the instances listed in Table 3. The final nine columns are divided into three groups of three columns each. Each group of three columns gives the relative improvement over the benchmarks for each of the three superlative ADP policies. For example, Column 2 indicates that ADP Policy 19 attains a mean probability of kill that is 42.9% greater than Benchmark Policy 1. ADP Policy 15 attains a 47% increase over the same benchmark whereas ADP Policy 2 achieves a 48.6% increase.

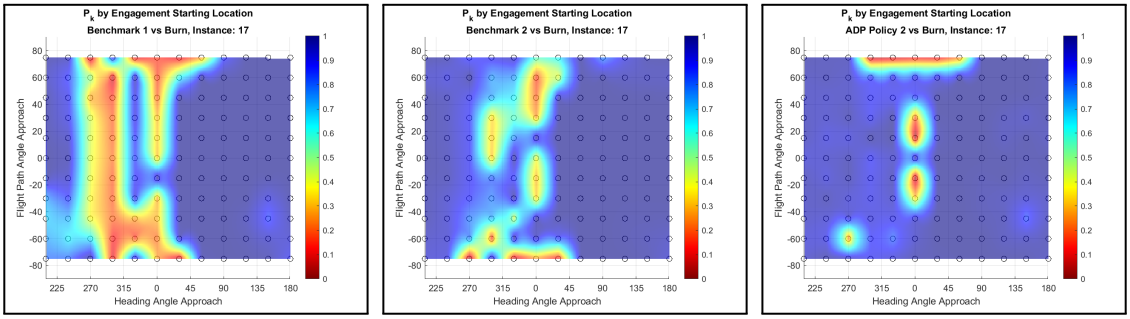
The top performing policies against the simply maneuvering target of Problem Instances 1 through 9 do not necessarily outperform the benchmark policies against a more evasive target. However, ADP policy 2 exhibits superior performance, outperforming Benchmark Policy 1 in all 9 Turn & Burn instances and outperforming Benchmark Policy 2 in 6 of the 9 instances. Consider the column in Table 7 that reports ADP Policy 2's relative improvement over Benchmark Policy 2. For instances with low energy initial conditions (i.e., Problem Instances 10 through 12) the benchmark is superior. For instances with an intermediate level of energy (i.e., speed) at engagement start, the ADP policy is superior. For the final instances with a high initial energy state, the ADP policy remains superior, but by a smaller margin than in the intermediate instances.

Figure 17 shows the P_k approach maps for each policy from every approach vector for Problem Instance 17. Benchmark Policy 1 has difficulty destroying the target when approaching from the left side of the engagement sphere, as viewed by the AUCAV. Benchmark Policy 2 performs better than Benchmark Policy 1 although it still achieves a moderately low P_k, about 0.7, when approaching from the left. ADP Policy 2 clearly performs better than both benchmarks for this instance. The ADP policy obtains high P_k values when approaching from the left. We also note the ADP

Table 7. Superlative ADP Policies vs Turn & Burn Target Maneuvering

	Relative (%) improvement over		P_k ADP	Relative (%) improvement over		P_k ADP	Relative (%) improvement over		P_k ADP
Instance	Benchmark 1	Benchmark 2	Policy 19	Benchmark 1	Benchmark 2	Policy 15	Benchmark 1	Benchmark 2	Policy 2
10	42.9	-5.5	0.755	47.0	-2.7	0.777	48.6	-5.9	0.752
11	36.1	-7.9	0.744	42.1	-3.9	0.776	42.5	-7.6	0.746
12	35.0	-7.2	0.726	42.1	-2.3	0.764	38.2	-8.1	0.719
13	25.3	8.6	0.844	13.1	-2.0	0.762	37.1	12.2	0.873
14	20.0	5.8	0.833	13.0	-0.4	0.784	28.3	9.1	0.859
15	13.9	-1.9	0.795	9.9	-5.4	0.766	21.1	3.4	0.838
16	26.7	6.4	0.897	14.6	-3.7	0.812	28.3	7.4	0.906
17	18.5	4.5	0.899	5.7	-6.7	0.802	19.2	4.7	0.900
18	13.8	4.2	0.880	7.0	-2.0	0.828	15.9	2.5	0.866

policy has smaller regions of low P_k when approaching head-on (i.e., from a heading of 0°) but does perform worse when approaching head-on from above.

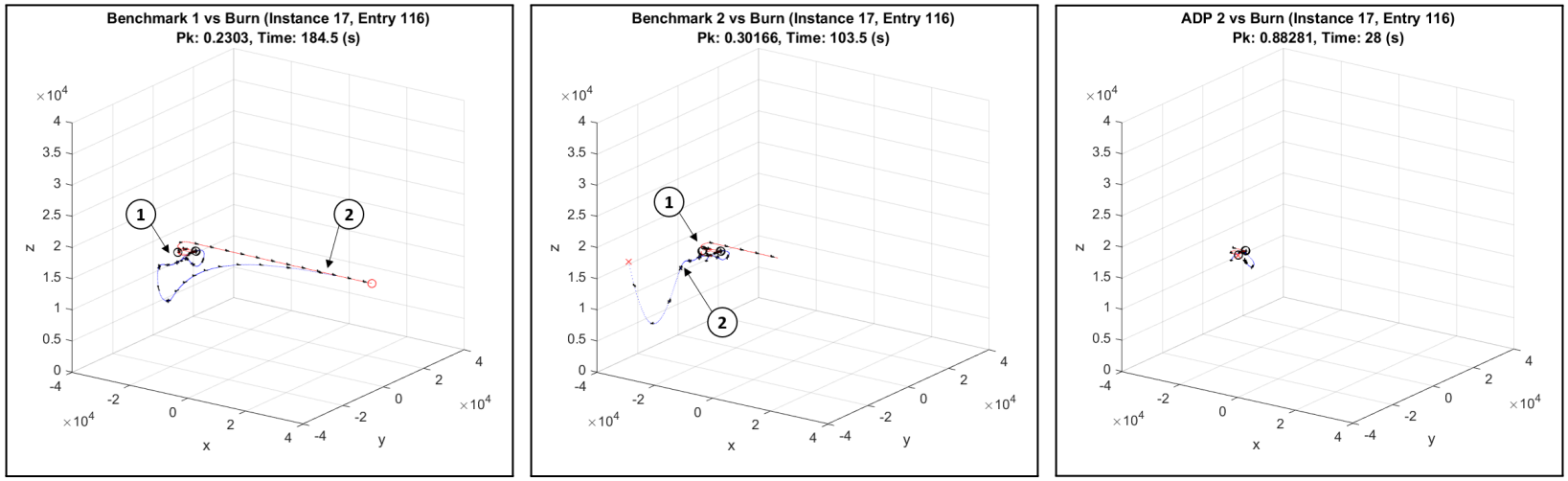


(a) Benchmark 1 **(b) Benchmark 2** **(c) ADP Policy 19**

Figure 17. P_k Maps versus Turn & Burn Evasion Tactic, Instance 17

Figure 18 shows a wide view of sample trajectories for the AUCAV following each policy from a particular starting location (i.e., Entry Point 116, (292.5,15), on Figure 17). Figure 18a shows the AUCAV following Benchmark Policy 1. Beginning the engagement with a high energy level, the AUCAV is able to reach the target quickly. As the AUCAV gets into position to fire (Point 1) the red aircraft begins the Turn & Burn evasive maneuvers. The red aircraft is able to deny the AUCAV high P_k shots and is able to escape the initial phase of the engagement, turning back toward a heading of 0°. The more aggressive maneuvering around Point 1 leaves the AUCAV in a lower energy state (i.e., it has lost speed due to maneuvering) and pointed away from the target. The AUCAV maneuvers around and begins a slow climb back to the target. The reduced energy state of the AUCAV after the initial pass results in slow recovery maneuvers, allowing the red aircraft to penetrate farther into the AOR. By the time the AUCAV recovers altitude and speed (Point 3), it is too late; the red aircraft is able to build enough of a lead that it successfully exits the AOR as indicated by the red circle at the end of its trajectory.

Figure 18b shows a sample trajectory of the AUCAV following Benchmark Policy 2. Similar to Benchmark Policy 1, the AUCAV is able to reach the target quickly and



(a) Benchmark Policy 1 (b) Benchmark Policy 2 (c) ADP Policy 2

Figure 18. Sample Trajectories versus Turn & Burn Evasion Tactic, Instance 17

get off a few low P_k shots. Once the red aircraft begins the Turn & Burn maneuvers, the AUCAV is denied additional shots and unable to turn with the red aircraft. Once the AUCAV is far away from the target, it appears to lose interest in the target and begins a steep dive manuever and departs the AOR ending the engagement. On initial inspection it appears Benchmark Policy 2 lost control and executed a wild maneuver. However, the steep dive is a product of the Benchmark Policy 2 decision function and highlights the limitations of such functions. Recall that Benchmark Policy 2 is a weighted combination of position information and energy information. It makes decisions according to

$$\underset{x_t^B \in \mathcal{X}^B(S_t)}{\arg\max} \left\{ \omega S_R S_A + (1-\omega) S_E \right\}. \tag{51}$$

For position, the decision function considers the product of a range and angle score. The range score decays exponentially while the angle score decays linearly. That is, an AUCAV following benchmark 2 seeks to maximize the weighted sum of the position and energy score. The range score S_R decays exponentially. Hence, when the range between the AUCAV and target is large, the first term in Equation (51) is near zero. Moreover, decision making based on changes in range and angles are significantly reduced because the slope of S_R is near zero. This results in a decision rule that seeks to maximize energy. For energy, the decision function considers the ratio of specific energy of the AUCAV to the red aircraft. The Turn & Burn evasive maneuvering scheme of the red aircraft is relatively stable and does not result in significant changes in its energy state (i.e., it never changes altitude or speed drastically). Hence, maximizing S_E results in maximizing the specific energy of the AUCAV. Recall the specific energy is given by

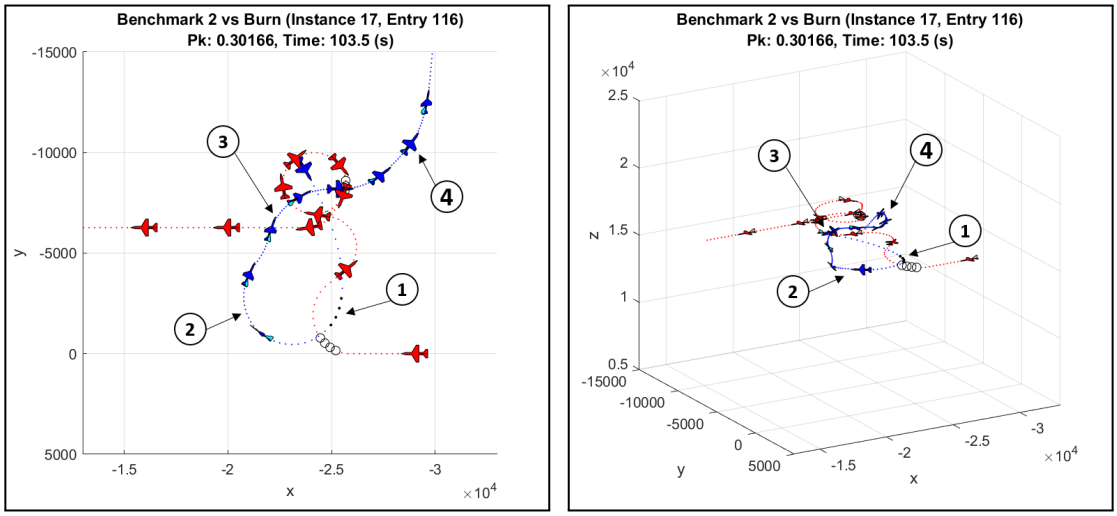
$$E_s = h + \frac{V^2}{2g}. \tag{52}$$

The best way to maximize Equation (52) is to increase kinetic energy (i.e., velocity). At Point 2 in Figure 18b, the AUCAV transitions to a state where the second term in the decision function dominates the first. The result is a steep dive to increase kinetic energy and then a pull up to increase potential energy (i.e., altitude). In pursuit of maximizing its total energy, AUCAV flies out of the AOR because it has no concept of AOR boundaries. Similar behavior is seen in Figure 14b. A detailed discussion of the early phase maneuvering follows.

Figure 18c shows a sample trajectory of the AUCAV following ADP Policy 2. Similar to the benchmarks, the AUCAV is able to reach the target quickly and, through effective energy management, it is able to achieve better shots (i.e., higher P_k) on the red aircraft, which results in successfully destroying the target. We examine the differences in the close-in-maneuvering between Benchmark Policy 2 and ADP Policy 2 to better understand why the benchmark failed and the ADP policy succeeded.

Figure 19 provides a detailed view of the initial engagement depicted in Figure 18b. As the AUCAV approaches the target (Point 1), it fires four low P_k shots, which all miss (i.e., black dots on the AUCAV trajectory indicate taking shots, black circles on the red aircraft trajectory indicate missing shots). At Point 1 the red aircraft begins the Turn & Burn evasive maneuvers, turning into the AUCAV and denying high P_k shots. As the AUCAV passes the red aircraft and approaches Point 2, the red aircraft begins to turn back towards a 0° heading. At Point 2 the AUCAV conducts a horizontal turn (as seen in Figure 19b) toward the red aircraft. Its speed at Point 2 is significantly lower, having executed the tight turn moving from Point 1 to Point 2. As the AUCAV approaches the red aircraft for the second pass (Point 3), the red aircraft is able to speed up and maneuver away from the AUCAV. The AUCAV does not have enough speed to close on the target before it maneuvers away. The AUCAV is left in a low energy state, flying away from the target (Point 4). As the red aircraft increases

the distance between itself and the AUCAV, the AUCAV enters a state wherein its decision making becomes dominated by energy considerations, and the AUCAV takes a steep dive seeking to increases its kinetic energy, ultimately maneuvering out the the AOR and ending the engagement.



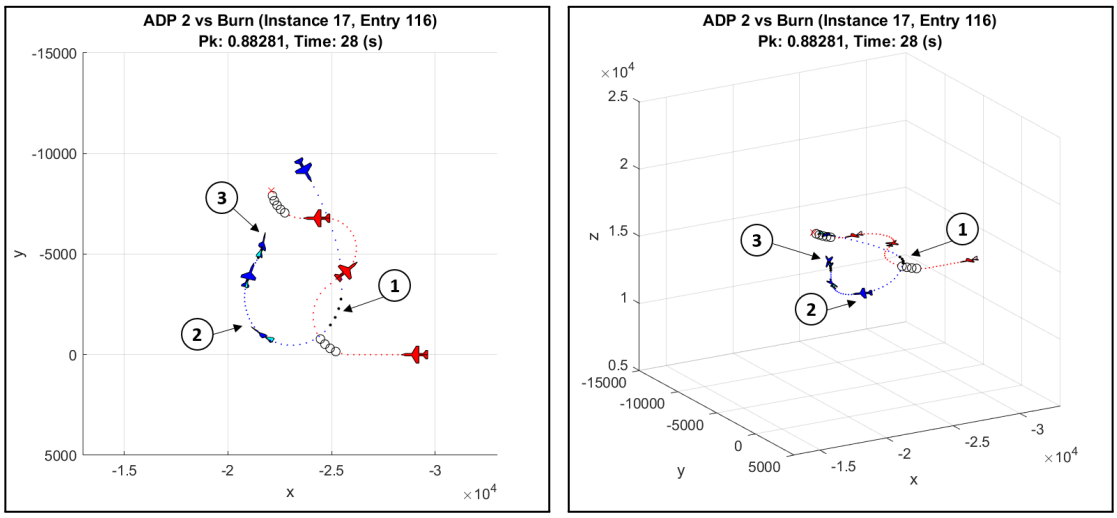
(a) Top View **(b) Isometric View**

Figure 19. Benchmark Policy 2 versus Turn & Burn Evasion Tactic, Instance 17

Figure 20 shows a detailed view of the initial engagement of the AUCAV following ADP Policy 2, depicted in Figure 18c. The AUCAV takes nearly the same approach to the target as with Benchmark Policy 2. The AUCAV takes, and misses, the same low P_k shots at Point 1. The difference between Benchmark Policy 2 and ADP Policy 2 is that here the AUCAV makes a descending turn (as seen in Figure 20b) as it maneuvers from Point 1 to Point 2. As with Benchmark Policy 2, as the AUCAV comes out of Point 1 it has lost a significant amount of energy. The descending turn through point allows the AUCAV to recover some airspeed as it approaches the red aircraft, which has turned back towards a 0° heading. As the AUCAV approaches for the second pass it now has enough energy to close to firing range and has the opportunity to take several shots that eventually down the red aircraft.

The difference at Point 2, a horizontal turn for Benchmark Policy 2 and a descend-

ing turn for ADP Policy 2, results in radically different outcomes for both policies. Figure 21 shows a detailed energy and position comparison between the two sample trajectories. The top row shows the energies (potential, kinetic, and total) for the Benchmark Policy 2 (left) and ADP Policy 2 (right) over the course of the initial engagement. The second row gives the range to the red aircraft over time with the gun WEZ indicated as the region between the red horizontal lines. The last row gives the radar angle to the red aircraft over time with the gun WEZ indicated as the region below the red horizontal line at 30°. When both the range and radar angle gun WEZ constraints are satisfied (i.e., both curves are between or below the red lines), the AUCAV can fire the gun at the target, as indicated by the black dots. The AUCAV flying the Benchmark Policy 2 takes four initial shots on the target at



(a) Top View **(b) Isometric View**

Figure 20. ADP Policy 2 versus Turn & Burn Evasion Tactic, Instance 17

around 8 seconds into the engagement. The AUCAV executes a horizontal turn back towards the target and almost earns a second shot opportunity around 25 seconds but is just out of maximum gun range. The AUCAV flying ADP Policy 2 gets the second opportunity because it executed the descending turn instead of the horizontal turn of Benchmark Policy 2. This gives the AUCAV just enough extra energy (i.e., speed) to

get inside of firing rage of the second shot at 25 seconds. The blue potential energy curves in the top row shows the difference in potential energy between the two aircraft between 10 and 25 seconds. Benchmark Policy 2 maintains potential energy whereas ADP Policy 2 trades the potential energy for a slight increase in kinetic energy that gives it the second shot opportunity which results in destruction of the red aircraft.

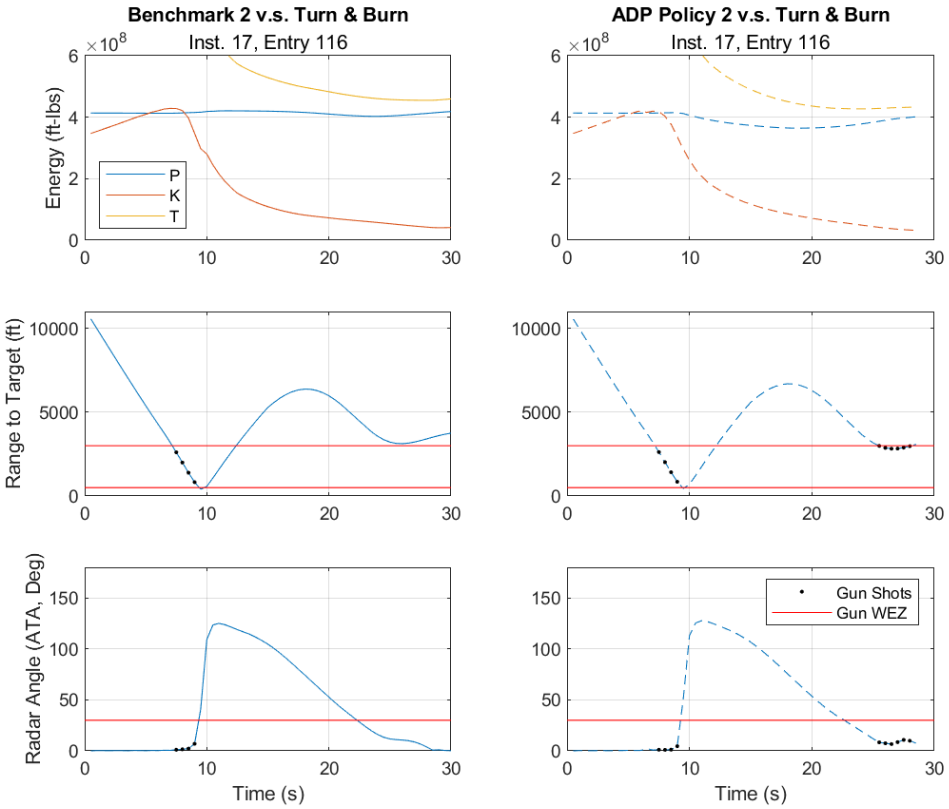


Figure 21. Shot Opportunities: Benchmark 2 versus ADP Policy 2

This sample trajectory demonstrates the importance of energy management throughout the entire engagement and the advantage of our 5DOF model that implicitly allows such energy management. Benchmark Policy 2 attempts to capture the energy concepts that are critical to air combat but fall short. The flaw in Equation 51 is that it represents a desire to increase total energy. Energy maneuverability considerations are more nuanced. There are times when one should gain energy, times to

trade kinetic energy for potential energy, times to trade potential energy for kinetic, and times to trade total energy for position. Benchmark Policy 2 cannot capture the interplay of potential and kinetic energy that is critical to effective and efficient maneuvering. As demonstrated, ADP policies generated from our 5DOF model can.

4.6 Excursion: Maneuvers and Tactics

One objective of this research is to determine if the ADP approach can generate policies that exhibit realistic behaviors. This section examines four aerobatic maneuvers and two basic fighter maneuvers (BFMs) that are observed from the top performing ADP policies (i.e., Policies 2, 15, and 19). We note that no single ADP policy exhibits all of the listed maneuvers. Moreover, the engagement scenario and problem instances used in the previous analyses do not lead to situations where all of the following maneuvers are expected to be used. Therefore, we construct situations for which itis reasonable to use each of the manuevers to examine how well the ADP policies replicate common aerobatic maneuvers and combat tactics. All aircraft icons are spaced 10 seconds apart in the following figures.

4.6.1 Wingover

A wingover is an energy efficient maneuver used to quickly turn 180° with a small turn radius. The basic wingover is executed as follows. The aircraft pulls up into a vertical climb, converting nearly all of its kinetic energy into potential energy (i.e., conserving it). At the top of the climb, when airspeed is low and the aircraft nose is pointed straight up, the aircraft turns in the vertical plane 180° until the nose is pointed straight down. The aircraft descends, regains its kinetic energy, and pulls level at the desired altitude. This results in the aircraft following nearly the same trajectory as on the way up, but going the opposite direction (i.e., a 180° degree

turn).

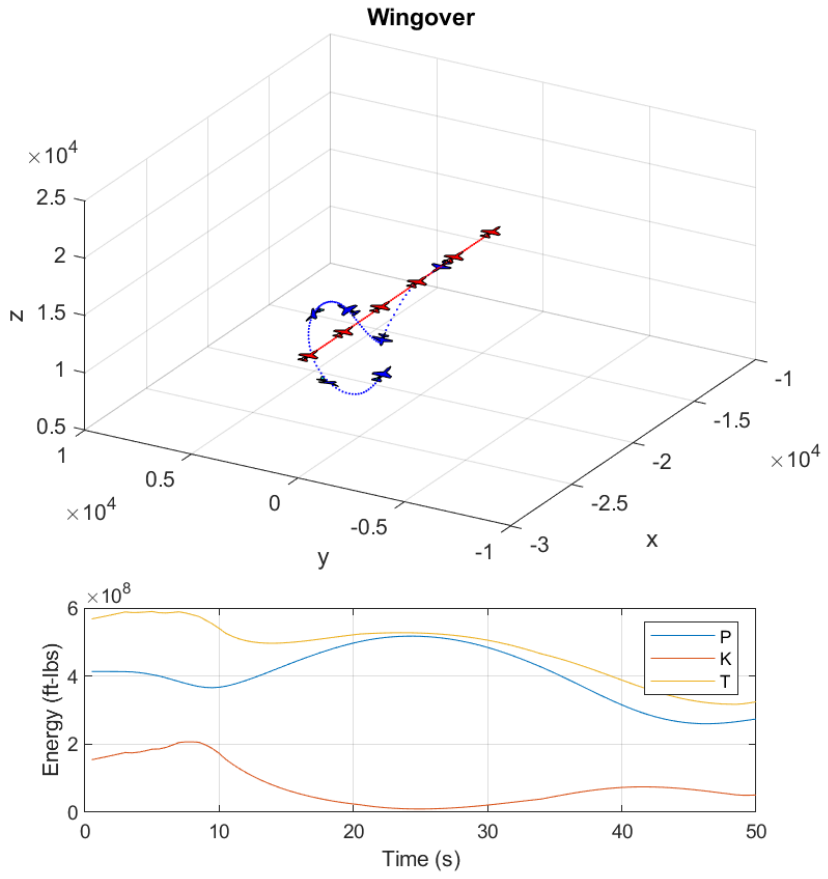


Figure 22. Wingover Maneuver

Figure 22 shows the AUCAV executing a wingover maneuver to quickly change direction to follow the red aircraft after making a head-on pass. The AUCAV does not end the wingover at the same altitude as it enters the maneuver because it needs to increase its speed to catch up to the red aircraft. Note the bulge between the potential and kinetic energy lines 25 seconds into the maneuver. This bulge is indicative of an energy efficient maneuver where much of the kinetic energy is conserved as potential energy during the maneuver and then converted back to kinetic energy at the end of the maneuver.

4.6.2 Immelmann

An Immelmann (i.e., half-loop) is a simple maneuver used to turn an aircraft 180° while also increasing altitude. An Immelmann is executed as follows. From a level flight condition, pull up through the vertical until the aircraft is inverted and pointed in the opposite direction (i.e., at the top of a half-loop), then roll 180° back to a wings level condition. The aircraft is now on a reverse heading and at a higher altitude.

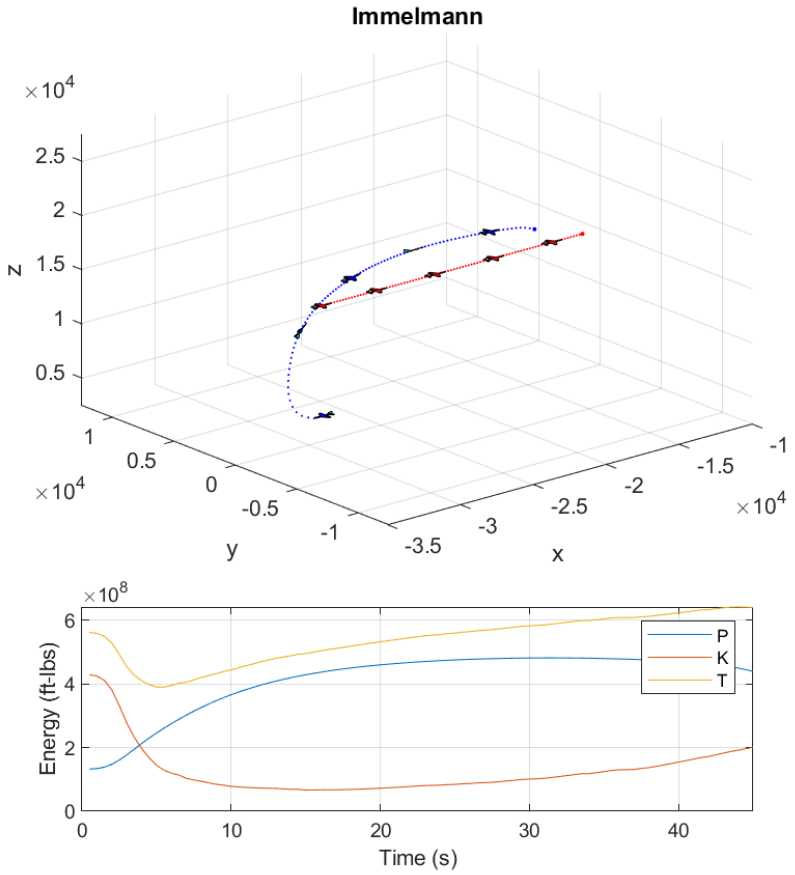


Figure 23. Immelmann Maneuver

Figure 23 shows the AUCAV executing an Immelmann to change direction and increase altitude in pursuit of the red aircraft. In contrast to the wingover, this maneuver is used to trade kinetic energy for potential energy (i.e., altitude) and

position in the form of a heading reversal. The aircraft ends the maneuver with increased potential energy and reduced kinetic energy.

4.6.3 Split-S

The split-S (i.e., half-loop) is used to reverse heading 180° while also descending in altitude. The split-S is the opposite of the Immelmann and is executed as follows. From a level flight condition, roll 180° inverted, then pull down. The aircraft passes through the downward vertical and up into a level flight condition. The aircraft is now on a reverse heading at a lower altitude.

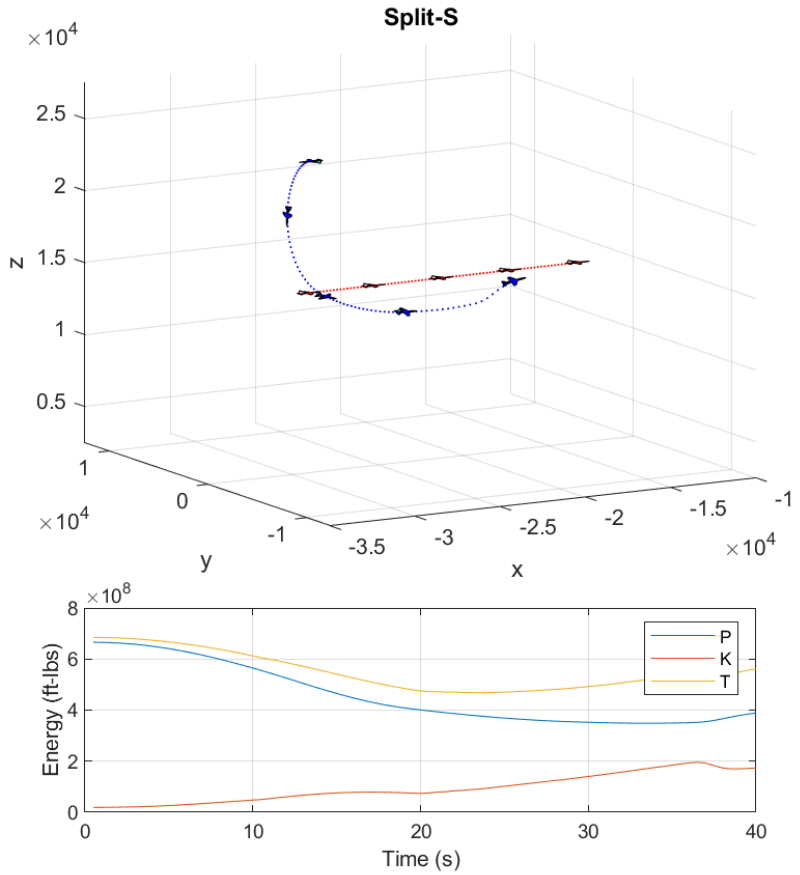


Figure 24. Split-S Maneuver

Figure 24 shows the AUCAV executing a split-S maneuver to change direction

and decrease altitude in pursuit of the red aircraft. Like the Immelmann, the split-S is used to trade between kinetic and potential energy. At the end of a split-S, the aircraft has increased its kinetic energy and reduced its potential energy.

4.6.4 Vertical Loop

The vertical loop is less commonly used but effective in situations with a high rate of closure against a non-maneuvering target. The maneuver is executed by pulling up, past the vertical, continuing to pull until the nose of the aircraft has completed a full 360° loop and the aircraft is back on its original heading.

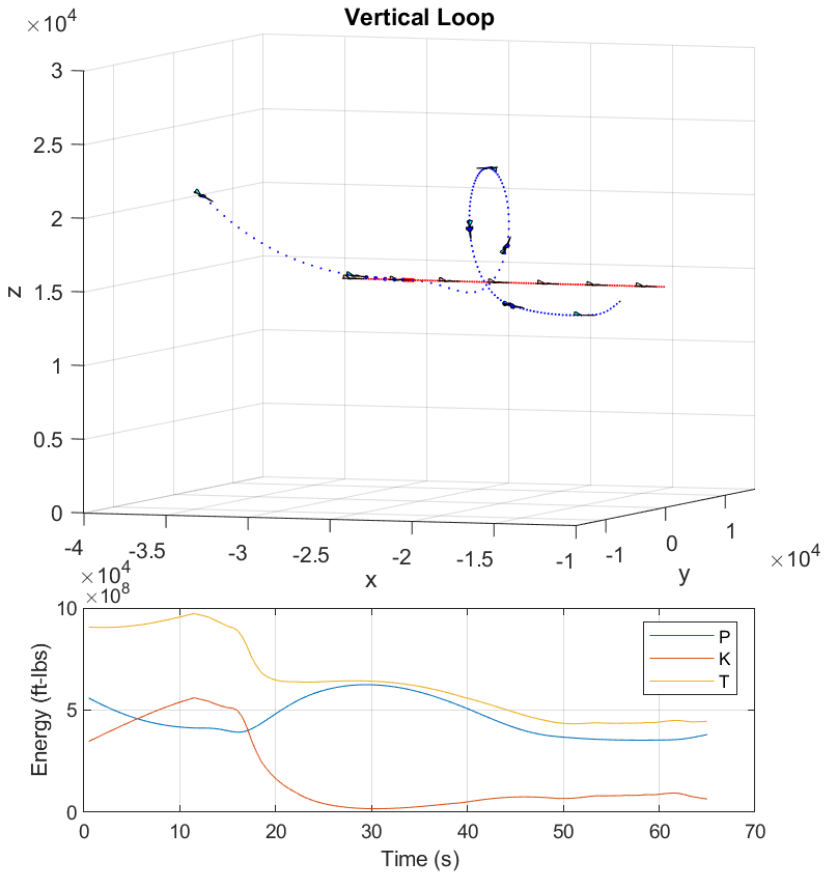


Figure 25. Vertical Loop Maneuver

Figure 25 shows a fast flying AUCAV in pursuit of a slow flying red aircraft. The

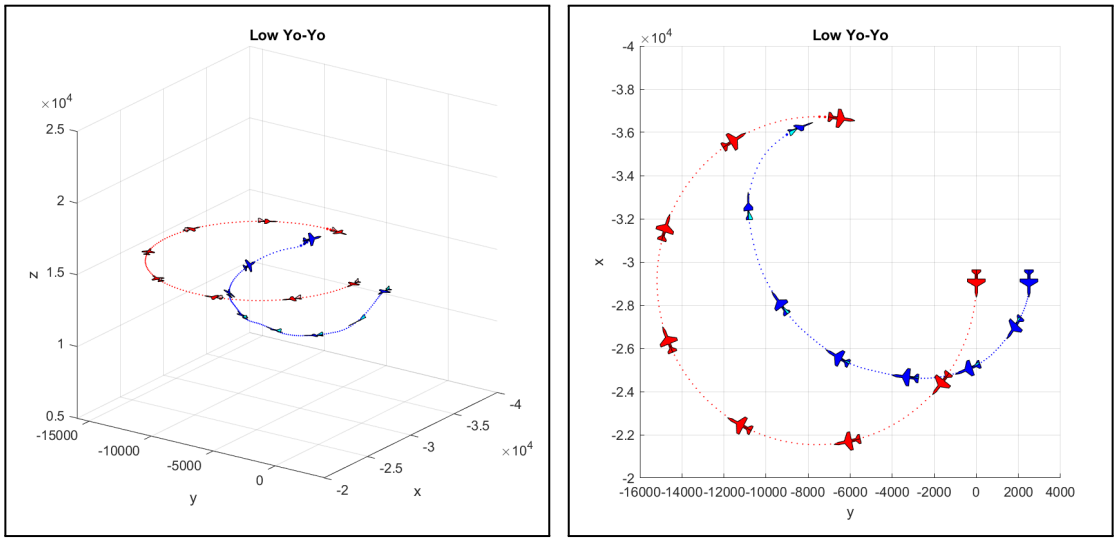
AUCAV's high closing speed causes the aircraft to overshoot the target. The AUCAV pulls up sharply to reduce its velocity along the x-axis. About 30 seconds into the engagement, the AUCAV is inverted at the top of the vertical loop. The red aircraft continues along its trajectory flying past the AUCAV. As the AUCAV pitches down and completes the second half of the loop, it is now aft of the red aircraft and in position to make another high P_k gun pass. As with the wingover, the vertical loop is an energy efficient maneuver that stores kinetic energy as potential energy as the AUCAV waits for a better position relative to the red aircraft. At the end of the maneuver the AUCAV uses the potential energy to attain the desired speed.

4.6.5 Low Yo-Yo

BFMs are a group of standard maneuvers, taught to all United States Air Force pilots, that form the building blocks of fighter tactics. Many advanced tactics build from the concepts of BFMs. The low yo-yo maneuver is a BFM used to increase the closure rate between the pursuer and target when the pursuing aircraft cannot turn its nose to point at the target either because of turn rate limitations or low speed. The pursuing aircraft establishes a lead-pursuit curve (see Shaw (1985) for a discussion of pursuit curves) while descending in a nose low attitude. The descent allows the aircraft to increase its speed while the lead-pursuit cuts the corner of the defender's turn allowing the pursuing aircraft to close the distance to the target and achieve a good firing position.

Figure 26 shows the AUCAV executing a low yo-yo maneuver from a speed disadvantage with less than half the speed of the red aircraft. Recall that aircraft icons are spaced 10 seconds apart. The AUCAV immediately takes a descending turns toward the target and achieves lead-pursuit (i.e., its velocity vector is pointed ahead of the red aircraft) 20 seconds into the engagement. As the AUCAV increases its speed it

reduces the distance to the red aircraft. 60 second into the engagement the AUCAV transitions to pure-pursuit (i.e., its velocity vector is pointed at the red aircraft) and pitches up to take a high P_k shot on the target.



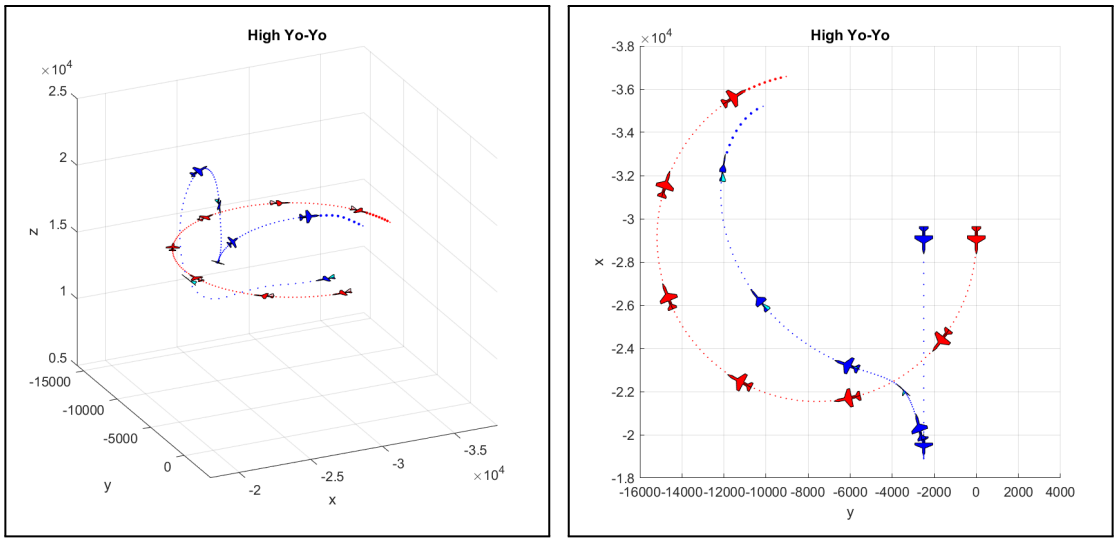
(a) Isometric View **(b) Top View**

Figure 26. Low Yo-Yo Air Combat Maneuver

4.6.6 High Yo-Yo

The high yo-yo maneuver is a BFM used to prevent overshoots (i.e., flying past the target) and reduce the aspect angle when the pursuer is at the same or higher speed than the target. The pursuer pitches up out of the plane of the targets maneuver in order to reduce speed and achieve a smaller turn radius while preventing an overshoot. Once the desired separation is achieve, the pursuer dives down after the target.

Figure 27 shows what could be considered a high yo-yo. The AUCAV begins the engagement with a speed advantage. It overshoots the red aircraft and pitches up to reduce speed. About 20 seconds into the engagement, the AUCAV is at the top of the maneuver and has regained a position aft of the red aircraft. The AUCAV dives back down into the red aircraft's plane of maneuver and lines up for a high P_k gun shot.



(a) Isometric View **(b) Top View**

Figure 27. High Yo-Yo Air Combat Maneuver

V. Conclusions

This research examines the one-versus-one (1v1) within visual range (WVR) air combat maneuvering problem (ACMP) with the intent of finding high-quality maneuver policies that reasonably replicate known tactics and maneuvers while improving the expected probability of kill for a generic intercept scenario. We develop a discounted, infinite horizon Markov decision process (MDP) model with several unique features that allow for examination of additional problem features not possible with models present in the existing literature. For example, the inclusion of a 5 degree of freedom (DOF) point-mass aircraft model allows for energy maneuverability concepts to be represented implicitly within the system dynamics rather than specified explicitly in the reward function. Moreover, the 5DOF model controls the aircraft roll angle, which allows for the modeling of weapons with engagement zones that depend on the roll angle, such as side firing weapons.

The high dimensionality and continuous nature of our MDP model preclude solving for the optimal policy exactly using dynamic programming. To address this curse of dimensionality, we utilize approximate dynamic programming (ADP) techniques. We design and employ an approximate policy iteration algorithm with a value function approximation based on neural network learning. Moreover, we develop a novel reward envelope expansion state sampling scheme to address the the sparse reward problem and avoid explicit reward shaping. We constructed a generic intercept scenario wherein an autonomous unmanned combat aerial vehicle (AUCAV) tasked with defending an area of responsibility must engage and destroy an adversary aircraft attempting to penetrate the airspace and attack high value assets. In lieu of a status quo benchmark policy for air combat, we compare maneuver policies generated from our solution procedure to two benchmarks derived from common reward shaping functions found in the ACMP literature. The first benchmark considers only position

information whereas the second considers both position and the aircraft energy state. We compare policies across 18 problem instances and against an adversary aircraft exhibiting two different evasion tactics.

The results of our computational experiments indicate that the ADP policies outperform position-only benchmark policies in 15 of 18 instances and outperform position-energy benchmark policies in 9 of 18 instances, attaining improved probabilities of kill among problem instances most representative of typical air intercept engagements. We observe that the selected benchmarks perform well in simple air combat situations such as ones with a non-maneuvering target, or ones with a target executing simple evasion maneuvers such as the Burn tactic. Under these conditions the benchmarks do not leave much room for improvement with respect to mean probability of kill. The benchmark policies do not perform as well against targets that exhibit slightly more aggressive evasive maneuvers such as the Turn & Burn tactic. The ADP generated policies are more robust, outperforming the benchmark policies in 15 of the 18 problem instances against a red aircraft employing the Turn & Burn evasion tactic. The improvement of the ADP maneuver policies over the benchmark policies results from their ability to manage the AUCAV kinetic and potential energy more effectively than either benchmark. The first benchmark has no notion of energy state and as a result cannot manage the AUCAV energy. The second benchmark considers the total energy state (i.e., the sum of the kinetic and potential energy) of the AUCAV and is incapable of managing energy at the component level (i.e., it cannot convert kinetic energy to potential energy and vice versa).

The qualitative comparison of the ADP maneuver policies ability to perform common aerobatic maneuvers indicate that our solution approach can successfully generate policies that reasonably replicate known behaviors. From an air combat perspective, we observe the ADP policies performing common fighter tactics such as

utilizing lead, pure, and lag pursuit curves to maneuver into a a position from which weapons can be employed. Moreover, we observe what appear to be specific basic fighter maneuvers (BFMs), low yo-yo and high yo-yo, against a turning target. These maneuvers are part of a subset of maneuvers (i.e., BFM) that form the building blocks of fighter tactics and are taught to all Air Force fighter pilots.

This research is of interest to the analytical modeling and simulation community. Analysts can apply our solution procedure to develop maneuver policies for 1v1 WVR air combat with guns. These maneuver policies can be incorporated into air combat simulations such as the Analytic Framework for Simulation, Integration, and Modeling (AFSIM) (West and Birkmire, 2019) to improve the behaviors of constructive entities. Our model formulation is flexible in that both the weapon effectiveness map (i.e., gun p_k map) and aircraft aerodynamic data can be easily changed which allows for modeling of maneuvers based on specific aircraft-weapon combinations.

Extensions to our approach include modeling additional weapons such as those discussed in Chapter II. The internally carried gun has historically been a fixture of fighter aircraft and has driven the development of common tactics. However, all modern fighters carry short range infrared air-to-air missiles that have a much larger and less restrictive weapon engagement zone (WEZ) and are likely to be the primary weapon of choice over the gun in any engagement. Recent advances in directed energy weapons (DEWs) increase the likelihood of these types of weapons being fielded on a fighter sized aircraft in the near future. The nature of DEWs (i.e., the firing of coherent light versus projectiles) lend themselves to non-standard WEZs as compared to the traditional forward firing gun and missiles. The addition of these types of weapons could dramatically change the tactics associated with the 1v1 WVR ACMP and should be investigated.

Bibliography

Anderson, J. D. (2008), 'Introduction to flight'.

Austin, F., Carbone, G., Falco, M., Hinz, H. and Lewis, M. (1990), 'Game theory for automated maneuvering during air-to-air combat', *Journal of Guidance, Control, and Dynamics* **13**(6), 1143–1149.

Bacon, D. J. (2011), 'United States Air Force Aircraft Accident Investigation Board Report'.

Balldin, U. I. (2002), *Acceleration Effects on Fighter Pilots*, Vol. 2, Office of the Surgeon General, United States Army, chapter 33.

Bates, C. and Santucci, G., eds (1990), *High G Physiological Protection Training*, Aerospace Medical Panel Working Group 14, NATO AGARD.

Bellman, R. (1952), On the theory of dynamic programming, Technical report, Rand Corp.

Binnie, W. B. and Stengelf, R. F. (1979), 'Flight investigation and theory of direct side-force control', *Journal of Guidance and Control* **2**(6), 471–478.

Blaquière, A., Gérard, F. and Leitmann, G. (1969), *Quantitative and Qualitative Games by Austin Blaquiere, Francoise Gerard and George Leitmann*, Academic Press.

Bocvarov, S., Cliff, E. and Lutze, F. (1994), Aircraft time-optimal heading reversal maneuvers, *in* 'Guidance, Navigation, and Control Conference', pp. 146–153.

Burgin, G. H. and Sidor, L. (1988), Rule-based air combat simulation, Technical report, Titan Systems Inc.

Byrnes, M. W. (2014), Nightfall: Machine autonomy in air-to-air combat, Technical report, Air University Maxwell AFB, Air Force Research Institute.

Casem, G. (2019), 'Joint simulation environment inches closer to reality', *Air Force News* .

Cohen, R. (2019), 'Experimental laser weapon downs multiple missiles in test', *Air Force Magazine* .

Cunningham, C. A. (2018), 'United States Air Force Aircraft Accident Investigation Board Report'.

David Aronstein, A. P. (1997), *The F-16 Lightweight Fighter*, pp. 203–229.

Davidovitz, A. and Shinar, J. (1985), 'Eccentric two-target model for qualitative air combat game analysis', *Journal of Guidance, Control, and Dynamics* **8**(3), 325–331.

Deptula, D. (2020), 'No, Elon Musk: The era of the manned fighter jet won't be over anytime soon', *Forbes* .

Dwyer, L. (2014), 'McDonnell Douglas F-4 Phantom II'.
URL: *http://www.aviation-history.com/mcdonnell/f4.html*

Eidsaune, D. W. (2009), 'United States Air Force Aircraft Accident Investigation Board Report'.

Everstine, B. (2020), 'Artificial intelligence easily beats human fighter pilot in DARPA trial', *Air Force Magazine* .

Fang, J., Zhang, L., Fang, W. and Xu, T. (2016), Approximate dynamic programming for CGF air combat maneuvering decision, *in* '2016 2nd IEEE International Conference on Computer and Communications (ICCC)', IEEE, pp. 1386–1390.

Fox, M. C. and Forrest, D. K. (1993), Supersonic aerodynamic characteristics of an advanced F-16 derivative aircraft configuration, Technical Report 3355, NASA.

Frank, S. A. (2018), *Control theory tutorial: basic concepts illustrated by software examples*, Springer Nature.

Getz, W. and Leitmann, G. (1979), 'Qualitative differential games with two targets', *Journal of Mathematical Analysis and Applications* **68**(2), 421–430.

Gilbert, W. P., Nguyen, L. T. and Vangunst, R. W. (1976), Simulator study of the effectiveness of an automatic control system designed to improve the high-angle-of-attack characteristics of a fighter airplane, Technical Report TN D-8176, NASA.

Gregg, A. (2019), 'Air force completes first flight test of valkyrie unmanned fighter jet', *Washington Post* .

Grimm, W. and Well, K. (1991), Modelling air combat as differential game, *in* 'Differential Games—Developments in Modelling and Computation', Springer, pp. 1–13.

Hambling, D. (2020), 'U.S. special forces test laser gunship for covert strikes', *Forbes.com* .

Hill, P. G. and Peterson, C. R. (1992), *Mechanics and thermodynamics of propulsion*, Addison-Wesley.

Host, P. (2019), 'DOT&E says F-35A gun accuracy issues remain', *Janes* .

Hunter, J. (2019), *Jane's All The Worlds Aircraft*.

Insinna, V. (2020), 'Boeing rolls out Australia's first 'Loyal Wingman' combat drone', *Defense News* .

Isaac, R. (1965), *Differential Games*, Kriger Publisher Company.

Isaacs, R. (1951), 'Games of pursuit'.

Jackson, P. (2019), *Jane's All The Worlds Aircraft*.

Jarmark, B. (1985), 'A missile duel between two aircraft', *Journal of Guidance, Control, and Dynamics* **8**(4), 508–513.

Jarmark, B. and Hillberg, C. (1984), 'Pursuit-evasion between two realistic aircraft', *Journal of Guidance, Control, and Dynamics* **7**(6), 690–694.

Jenkins, P. R., Robbins, M. J. and Lunday, B. J. (2020), 'Approximate dynamic programming for military medical evacuation dispatching policies', *INFORMS Journal on Computing* .

Joint Chiefs of Staff (2016), 'The joint force in a contested and disordered world'.

Joint Chiefs of Staff (2018), 'Joint publication 3-01: Countering air and missile threats'.

Keller, J. (2014), 'Lockheed Martin tests laser weapons to protect aircraft from enemy fighters and missiles', *Military & Aerospace Electronics* pp. 26,33.

Lachner, R., Breitner, M. H. and Pesch, H. J. (1995), Three-dimensional air combat: Numerical solution of complex differential games, *in* 'New trends in dynamic games and applications', Springer, pp. 165–190.

Lyons, T. J., Harding, R., Freeman, J. and Oakley, C. (1992), 'G-induced loss of consciousness accidents: USAF experience 1982-1990.', *Aviation, space, and environmental medicine* **63**(1), 60–66.

Maiman, T. (1960), 'Speech by Dr. Theodore Maiman', press conference.
URL: *https://www.hrl.com/lasers/pdfs/maiman_60.07.07.pdfTheodore%20Maiman*

McGrew, J. S., How, J. P., Williams, B. and Roy, N. (2010), 'Air-combat strategy using approximate dynamic programming', *Journal of guidance, control, and dynamics* **33**(5), 1641–1654.

Merz, A. (1985), 'To pursue or to evade-that is the question', *Journal of guidance, control, and dynamics* **8**(2), 161–166.

Murtaugh, S. A. and Criel, H. E. (1966), 'Fundamentals of proportional navigation', *IEEE spectrum* **3**(12), 75–85.

Naidu, D. (2002), 'Singular perturbations and time scales in control theory and applications: an overview', *Dynamics of Continuous Discrete and Impulsive Systems Series B* **9**, 233–278.

Newman, D. (2016), *High G flight*, Routledge.

Paul, M. A. (1996), 'Extended-coverage-bladder g-suits can provide improved g-tolerance and high gz foot pain.', *Aviation, space, and environmental medicine* **67**(3), 253–255.

Rajan, N. and Ardema, M. (1985), 'Interception in three dimensions-an energy formulation', *Journal of Guidance, Control, and Dynamics* **8**(1), 23–30.

Sabatini, R., Richardson, M. A., Gardi, A. and Ramasamy, S. (2015), 'Airborne laser sensors and integrated systems', *Progress in Aerospace Sciences* **79**, 15–63.

Shaw, R. L. (1985), *Fighter Combat*, US Naval Institute Press.

Shinar, J. and Davidovitz, A. (1987), 'Unified approach for two-target game analysis', *IFAC Proceedings Volumes* **20**(5), 65–71.

Smith, C. W., Ralston, J. and Mann, H. (1979), Aerodynamic characteristics of forebody and nose strakes based on F-16 wind tunnel test experience, Technical Report CR-3053, NASA.

Snell, S., Garrard, W. and Enns, D. (1989), Nonlinear control of a supermaneuverable aircraft, *in* 'Guidance, Navigation and Control Conference', p. 3486.

Stevens, B. L., Lewis, F. L. and Johnson, E. N. (2015), *Aircraft control and simulation: dynamics, controls design, and autonomous systems*, John Wiley & Sons.

Toubman, A. (2020), Calculated Moves, PhD thesis, Leiden University.

Tripp, L. D., Warm, J. S., Matthews, G., Chiu, P., Werchan, P. and Deaton, J. E. (2006), '+ gz acceleration loss of consciousness: time course of performance deficits with repeated experience', *Human factors* **48**(1), 109–120.

Udoshi, R. (2017), *IHS Jane's Weapons*.

United State Air Force (2015), 'Basic doctrine'.

United States Air Force (2019), 'Counterair operations'.

USAF (2015), 'Fact sheets'.
URL: *https://www.af.mil/About-Us/Fact-Sheets/Indextitle/F/*

Virtanen, K., Karelahti, J. and Raivio, T. (2006), 'Modeling air combat by a moving horizon influence diagram game', *Journal of guidance, control, and dynamics* **29**(5), 1080–1091.

Wang, M., Wang, L., Yue, T. and Liu, H. (2020), 'Influence of unmanned combat aerial vehicle agility on short-range aerial combat effectiveness', *Aerospace Science and Technology* **96**, 105534.

Watson, J. and McAllister, J. (1977), Direct-force flight-path control-the new way to fly, *in* 'Guidance and Control Conference', p. 1119.

West, T. D. and Birkmire, B. (2019), 'Afsim', *CSIAC Journal* **7**(3), 50.

Whinnery, T. and Forster, E. M. (2013), 'The+ gz-induced loss of consciousness curve', *Extreme physiology & medicine* **2**(1), 19.

Whinnery, T., Forster, E. M. and Rogers, P. B. (2014), 'The+ gz recovery of consciousness curve', *Extreme physiology & medicine* **3**(1), 9.

Wilson, J. (2018), 'Laser weapons show their stuff in real-world conditions', *Military & Aerospace Electronics* .

Woodman, H. (1989), *Early Aircraft Armament*, Smithsonian Institute Press.

Yang, Q., Zhang, J., Shi, G., Hu, J. and Wu, Y. (2019), 'Maneuver decision of UAV in short-range air combat based on deep reinforcement learning', *IEEE Access* **8**, 363–378.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
25–03–2021	Master's Thesis	August 2019 — March 2021

4. TITLE AND SUBTITLE

Air Combat Maneuvering via Operations Research and Artificial Intelligence Methods

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Crumpacker, James B., Capt, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENS-MS-21-M-152

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Strategic Development Planning & Experimentation Office
Mr. David P. Panson
1864 4th Street
Wright-Patterson AFB, OH 45433
(937) 904-6539

10. SPONSOR/MONITOR'S ACRONYM(S)

SDPE

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution Statement A: Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Within visual range air combat requires rapid, sequential decision-making to survive and defeat the adversary. Fighter pilots spend years perfecting maneuvers for these types of engagements, yet the emergence of unmanned, autonomous vehicle technologies elicits a natural question – can an autonomous unmanned combat aerial vehicle (AUCAV) be imbued with the necessary artificial intelligence to perform air combat maneuvering tasks independently? We formulate and solve the air combat maneuvering problem to examine this question, developing a Markov decision process model to control an AUCAV seeking to destroy a maneuvering adversarial vehicle. An approximate dynamic programming (ADP) approach implementing neural network regression is used to attain high-quality maneuver policies for the AUCAV. ADP policies attain improved probabilities of kill among problem instances most representative of typical air intercept engagements. Maneuvers generated by the ADP policies are compared to basic fighter maneuvers and common aerobatic maneuvers. Results indicate that our proposed ADP solution approach produces policies that imitate known flying maneuvers.

15. SUBJECT TERMS

markov decision process, approximate dynamic programming, artificial intelligence

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Matthew J. Robbins, AFIT/ENS
U	U	U	UU	114	19b. TELEPHONE NUMBER *(include area code)* (937) 255-3636, x4539; Matthew.Robbins@afit.edu