

HIGHER-ORDER QUADRATIC CONTROL FOR SPACECRAFT PURSUIT-EVASION GAMES

Jacob H. Wolf* and Keith A. LeGrand†

The proliferation of spacecraft and advances in maneuverability have made spacecraft pursuit-evasion games increasingly relevant. In non-cooperative spacecraft scenarios, zero-sum pursuit-evasion games provide a framework to find optimal control policies for each agent. Due to the nonlinear dynamics encountered in typical orbital regimes, linear control laws rapidly become ineffective as the system evolution differs significantly from its linear approximation. Standard nonlinear methods offer superior performance but are often intractable and computationally intensive. Higher-order quadratic regulators maintain strong performance over larger neighborhoods than their linear counterparts without requiring significantly more computational load. This work presents a tensor-theoretic derivation and application of higher-order quadratic regulators to spacecraft pursuit-evasion games. The methodology described in this paper provides a second-order state-dependent nonlinear control law that outperforms the state-dependent Riccati equation method with substantially less online computation cost than iterative nonlinear solvers and data-driven approaches.

INTRODUCTION

The growing number of active satellites in Earth orbit increases the demand for improved autonomy to avoid intentional and non-intentional collisions. This can take the form of collision with nearby active spacecraft or debris. Spacecraft operating on independent objectives with limited-to-no communication with nearby vehicles and objects introduces scenarios where they must act without coordination or awareness of intent. In these environments, spacecraft must autonomously maneuver under nonlinear dynamics in response to the behavior of an external agent whose future movements are unpredictable.

To approach this problem, methods have been developed for non-cooperative relative navigation,¹⁻⁸ rendezvous,^{9,10} and control.^{11,12} A special case of this problem involves strategic competing agents. This situation is naturally modeled with differential game theory as introduced by Isaacs.¹³ Specifically, the problem is framed as a zero-sum pursuit-evasion game where a pursuer spacecraft intends to intercept an evader spacecraft. The optimal policy is characterized by Hamilton-Jacobi-Isaacs (HJI) partial differential equation (PDE), which is difficult to solve, leading to approximate approaches being used to find near-optimal solutions instead. Prior work includes the state-dependent Riccati equation (SDRE) method for near-optimal solutions to nonlinear differential games,¹⁴ direct numerical methods,¹⁵ reinforcement learning methods,¹⁶ and analytical solutions for simplified systems.¹⁷

*Graduate Research Assistant, School of Aeronautics and Astronautics, Purdue University, 701 W. Stadium Ave., West Lafayette, IN 47907.

†Assistant Professor, School of Aeronautics and Astronautics, Purdue University, 701 W. Stadium Ave., West Lafayette, IN 47907.

Due to the high nonlinearity in the relevant dynamics, methods which make use of local linear approximations face steep declines in precision as the game evolves and methods which use simplified dynamics are of limited utility. Near-optimal nonlinear approaches, like the SDRE method, mitigate this issue by periodically solving for state-dependent gains, giving a nonlinear control law. However, each instance of this solution is valid only in quasi-linear regions about the expansion point.

This paper improves upon existing linear control and SDRE methods by employing higher-order Taylor expansions of the dynamics vector field and second-order state feedback control. Compared to existing techniques, this higher-order approach increases the region about which the dynamics approximations are valid and the control law near-optimal. Inspired by the SDRE approach, the control gains are recomputed periodically about the current state to update the expansion point, producing a locally near-optimal solution with a higher-order feedback law compared to first-order SDRE methods. After solving the SDRE, the higher-order contributions are given as the solution to a tensor Sylvester equation. This higher-order feedback approach provides more degrees-of-freedom in the control vector and incorporates quadratic nonlinearities of the dynamics, resulting in improved local performance. Depending on the system dynamics, this approach may require less-frequent gain recalculations than in quasi-linear counterparts while achieving superior performance within a quasi-quadratic region about the update point. The contribution of this paper is a second-order extension of the SDRE method that is referred to as the state-dependent Riccati Sylvester equation (SDRSE) method.

PROBLEM FORMULATION

Notation

In this paper, scalars are denoted by lowercase characters and vectors by boldface lowercase characters. Matrices are represented by uppercase characters and third-order tensors by boldface uppercase characters. The functional J is exempt from these conventions to align with standard naming conventions. Superscripts are used to denote contravariant indices whereas subscripts denote covariant indices. Einstein summation convention is used throughout this document to imply summation over repeated indices.

Dynamics

The problem is formulated with control-affine two-body dynamics for both agents with respect to the Earth. The equations of motions (EOMs) are given by

$$\ddot{\mathbf{r}}_{(\cdot)} = -\mu \frac{\mathbf{r}_{(\cdot)}}{\|\mathbf{r}_{(\cdot)}\|^3} + B_{(\cdot)} \mathbf{u}_{(\cdot)} \quad (1)$$

where

$$\mathbf{r}_{(\cdot)} = x_{(\cdot)} \hat{\mathbf{e}}_1 + y_{(\cdot)} \hat{\mathbf{e}}_2 + z_{(\cdot)} \hat{\mathbf{e}}_3 \quad (2)$$

and $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3)$ are Earth-centered inertial (ECI) unit vectors. The subscript (\cdot) may indicate either the pursuer, p , or the evader, e . A diagram of the frame is shown in Figure 1. Assuming that the mass of the agents is negligible, μ is the standard gravitational parameter for Earth.

In state-space representation, the uncontrolled state dynamics of a single agent under two-body dynamics (1) are given by

$$\dot{\mathbf{y}} = \begin{bmatrix} \dot{\mathbf{r}}(\cdot) \\ -\mu \frac{\mathbf{r}(\cdot)}{\|\mathbf{r}(\cdot)\|^3} \end{bmatrix} \quad (3)$$

where $\mathbf{y}(\cdot) = [\mathbf{r}(\cdot)^\top \quad \dot{\mathbf{r}}(\cdot)^\top]^\top$. In a multi-agent scenario, the uncontrolled system is modeled as the relative state dynamics under (1) as

$$\dot{\mathbf{x}} = \dot{\mathbf{y}}_p - \dot{\mathbf{y}}_e = \begin{bmatrix} \dot{\mathbf{r}}_p - \dot{\mathbf{r}}_e \\ -\mu \left(\frac{\mathbf{r}_p}{\|\mathbf{r}_p\|^3} - \frac{\mathbf{r}_e}{\|\mathbf{r}_e\|^3} \right) \end{bmatrix} \quad (4)$$

where the relative state is defined as $\mathbf{x} = \mathbf{y}_p - \mathbf{y}_e$, allowing derivatives with respect to \mathbf{x} to be used to approximate the relative dynamics of the two agents using a Taylor expansion. While the control law discussed in this paper is applicable to a wide range of other dynamical systems, including higher fidelity models, the two-body formulation is chosen here for simplicity.

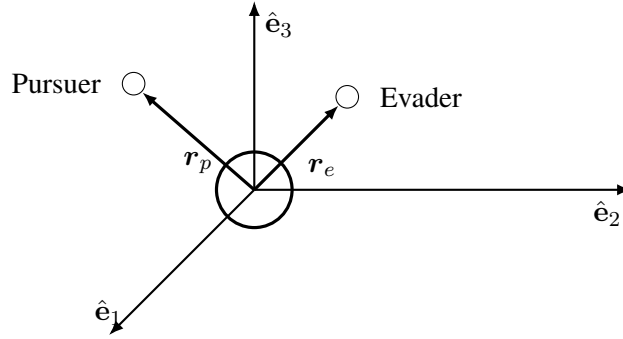


Figure 1 Reference frame

Control Problem Formulation

The payoff J of a differential game is the value that agents seek to minimize or maximize with their respective control vector. A Lagrange form integral payoff is one of the form

$$J(\cdot) = \int_0^{t_f} \mathcal{L}(\cdot) dt \quad (5)$$

where \mathcal{L} is the Lagrangian of the optimization problem and t_f is the terminal time. For the formulation considered in this paper, the pursuer applies the control $\mathbf{u}_p \in \mathbb{R}^3$ and seeks to *minimize* J . Similarly, the evader's control input, $\mathbf{u}_e \in \mathbb{R}^3$, is selected to *maximize* J . The optimal policy pair $(\mathbf{u}_p^*, \mathbf{u}_e^*)$, where $(\cdot)^*$ denotes optimality, is the policy pair that satisfies the zero-sum optimization problem

$$J^*(\cdot) = \min_{\mathbf{u}_p} \max_{\mathbf{u}_e} J(\cdot) \quad (6)$$

For convenience, we take the pursuer's objective function to be $J_p = J$ and the evader's as $J_e = -J$. The evader minimizing $-J$ is equivalent to maximizing J . This allows for the minimax problem

to be viewed as standard minimization problems. The optimal policy pair has conditions given by [18, pp. 297]

$$J_p(\mathbf{u}_p^*, \mathbf{u}_e^*) \leq J_p(\mathbf{u}_p, \mathbf{u}_e^*) \quad (7)$$

$$J_e(\mathbf{u}_p^*, \mathbf{u}_e^*) \leq J_e(\mathbf{u}_p^*, \mathbf{u}_e) \quad (8)$$

In other words, no agent can outperform the payoff under the policy pair $(\mathbf{u}_p^*, \mathbf{u}_e^*)$ by altering only their control. Any deviation from the optimal policy pair is unilaterally harmful to the deviating agent. Solving this optimization problem is equivalent to solving the HJI PDE and gives a saddle point solution to the optimal control problem, as illustrated in Figure 2.

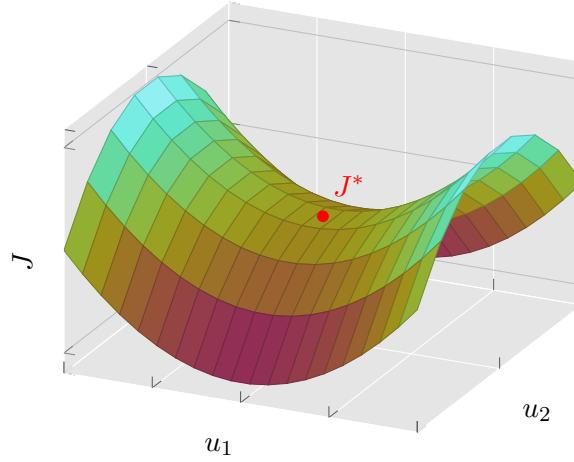


Figure 2 Saddle point solution of a zero-sum differential game.

An infinite horizon quadratic form payoff is chosen for this problem as an analog to the standard linear-quadratic regulator (LQR) problem. The state vectors for the pursuer and the evader are defined as

$$\mathbf{y}_{(\cdot)} = \begin{bmatrix} \mathbf{r}_{(\cdot)}^\top & \dot{\mathbf{r}}_{(\cdot)}^\top \end{bmatrix}^\top = [x_{(\cdot)} \ y_{(\cdot)} \ z_{(\cdot)} \ \dot{x}_{(\cdot)} \ \dot{y}_{(\cdot)} \ \dot{z}_{(\cdot)}]^\top \quad (9)$$

The payoff function of the game is given by¹⁹

$$J = \frac{1}{2} \int_0^\infty (\mathbf{x}^\top Q \mathbf{x} + \mathbf{u}_p^\top R \mathbf{u}_p - \gamma^2 \mathbf{u}_e^\top R \mathbf{u}_e) dt \quad (10)$$

where the relative state is given by $\mathbf{x} = \mathbf{y}_p - \mathbf{y}_e \in \mathbb{R}^6$. The constant $\gamma > 0$ is a fixed scalar design parameter that scales the evader's control effort weight. The payoff motivates the pursuer to rendezvous and drive the relative position and velocities to zero. Conversely, the evader attempts to maximize the relative position and velocity. The symmetric matrix $Q \succeq 0$, $Q \in \mathbb{R}^{6 \times 6}$ weights the state separation of the agents. Control effort is penalized with the symmetric matrix $R \succ 0$, $R \in \mathbb{R}^{3 \times 3}$.

For ease of exposition, we develop our higher-order control solution for the standard quadratic form before generalizing it to the game-theoretic payoff in (10). The infinite-horizon linear-quadratic objective function is

$$J_p = \frac{1}{2} \int_0^\infty (\mathbf{x}^\top Q \mathbf{x} + \mathbf{u}^\top R \mathbf{u}) dt \quad (11)$$

The single-agent objective function is different from the game theoretic objective as it does not converge to a saddle point but rather a minima. Solving this optimization problem is equivalent to solving the Hamilton-Jacobi-Bellman (HJB) PDE since there is no minimax structure.

BACKGROUND

Partial Derivative Tensors

The nonlinear function $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ represents the uncontrolled dynamics of the system and is assumed to be twice differentiable in a neighborhood about a point \tilde{x} . Considering small deviations, δ , from the expansion point the Taylor expansion of the function about \tilde{x} is

$$f(\tilde{x} + \delta) = f(\tilde{x}) + A\delta + \frac{1}{2}\mathbf{A}^{(2)}\delta^2 + \mathcal{O}(\delta^3) \quad (12)$$

where the Jacobian and second-order partial derivative tensor are respectively given by

$$A = \left. \frac{\partial f}{\partial x} \right|_{x=\tilde{x}} \in \mathbb{R}^{n \times n}, \quad \mathbf{A}^{(2)} = \left. \frac{\partial^2 f}{\partial x^2} \right|_{x=\tilde{x}} \in \mathbb{R}^{n \times n \times n} \quad (13)$$

The shorthand $\mathbf{A}^{(2)}\delta^2$ is used to denote the double contraction of third-order tensor $\mathbf{A}^{(2)}$ with two copies of vector δ . Explicitly,

$$(\mathbf{A}^{(2)}\delta^2)^i = A_{j,k}^i \delta^j \delta^k \quad (14)$$

As a partial derivative tensor, $\mathbf{A}^{(2)}$ is symmetric across its covariant indices, meaning that $A_{j,k}^i = A_{k,j}^i$.

Matrix-Tensor Contraction

For higher-order tensors, there is no unique definition for left- and right-multiplication because there are multiple indices along which contractions can occur. For the derivations in this paper, however, all tensors are third-order and symmetric across covariant indices. Because of this, the order of contraction over covariant indices does not affect the results. For notational convenience, let matrix $M \in \mathbb{R}^{n \times n}$ and third-order tensor $\mathbf{T} \in \mathbb{R}^{n \times n \times n}$ have two modes of contraction denoted by the shorthand

$$(M\mathbf{T})_{j,k}^i = M_m^i T_{j,k}^m \quad (15)$$

$$(\mathbf{T}M)_{j,k}^i = T_{j,m}^i M_k^m \quad (16)$$

where $M\mathbf{T}$ is a contraction over the contravariant index of the 3-tensor and $\mathbf{T}M$ is a contraction over one of the covariant indices of the 3-tensor.

State-Dependent Methods

The SDRE method provides an effective approach to solving nonlinear control problems. Given that the nonlinear function $f(\cdot) \in \mathcal{C}^1$ satisfies $f(\mathbf{0}) = \mathbf{0}$, a control-affine form may always be written as²⁰

$$\dot{x}(t) = A(x)x(t) + B(x)u, \quad x(0) = x_0 \quad (17)$$

for some continuous nonlinear matrix map $A(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$. Similar to LQR formulations, the state feedback control is given by

$$\mathbf{u}(\mathbf{x}) = -R^{-1}(\mathbf{x})B^\top(\mathbf{x})P(\mathbf{x})\mathbf{x} \quad (18)$$

where $P(\mathbf{x})$ is the unique positive-definite solution to the algebraic SDRE

$$P(\mathbf{x})A(\mathbf{x}) + A^\top(\mathbf{x})P(\mathbf{x}) - P(\mathbf{x})B(\mathbf{x})R^{-1}(\mathbf{x})B^\top(\mathbf{x}) + Q(\mathbf{x}) = \mathbf{0} \quad (19)$$

At each timestep, the *state-dependent Riccati equation*, (19), is solved with the current values of A , B , R , and Q to obtain the matrix P , allowing for the feedback gain to be computed.

METHODOLOGY

Single-Agent Quadratic-Quadratic Regulator

We first present the non-game theoretic quadratic-quadratic regulator (QQR) derivation for an optimal control problem. We then consider the game-theoretic problem and derive the optimal policy pair solution.

An optimal control problem with an integral payoff of form (5) satisfies necessary conditions for optimality given by the Euler-Lagrange equations. Solving the Euler-Lagrange equations is equivalent to solving the HJB equation along the optimal trajectory.

Consider nonlinear control-affine system

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}) + B\mathbf{u}, \quad \mathbf{x}(0) = \mathbf{x}_0 \quad (20)$$

where $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and is Lipschitz continuous on $t = [0, \infty)$, $B \in \mathbb{R}^{n \times m}$, and $\mathbf{u} \in \mathbb{R}^m$. The system (20) is expanded about the point $\tilde{\mathbf{x}}$. It is assumed, without loss of generality, that the function \mathbf{f} satisfies $\mathbf{f}(\tilde{\mathbf{x}}) = \mathbf{0}$. The second-order Taylor approximation about $\tilde{\mathbf{x}}$, which is valid for small deviations $\delta = \mathbf{x} - \tilde{\mathbf{x}}$, is then

$$\dot{\delta} \approx A\delta + \frac{1}{2}\mathbf{A}^{(2)}\delta^2 + B\mathbf{u} \quad (21)$$

The costate is of the system is denoted by $\lambda \in \mathbb{R}^n$ and the Hamiltonian is defined as

$$\mathcal{H} \triangleq \mathcal{L} + \lambda^\top \mathbf{f} = \frac{1}{2}\delta^\top Q\delta + \frac{1}{2}\mathbf{u}^\top R\mathbf{u} + \lambda^\top [A\delta + \frac{1}{2}\mathbf{A}^{(2)}\delta^2 + B\mathbf{u}] \quad (22)$$

where the Lagrangian, \mathcal{L} , is the integrand of (11). The necessary conditions of optimality are given by the Euler-Lagrange equations [21, pp. 188] as

$$\dot{\lambda}^\top(t) = -\frac{\partial \mathcal{H}}{\partial \delta} \quad (23)$$

$$\lambda^\top(t_f) = \frac{\partial \phi(\delta(t_f), t_f)}{\partial \delta} \quad (24)$$

$$\mathbf{0} = \frac{\partial \mathcal{H}}{\partial \mathbf{u}} \quad (25)$$

where the arguments for $\mathcal{H}(\delta(t), \mathbf{u}(t), \boldsymbol{\lambda}(t), t)$ are dropped for brevity. The terminal cost ϕ is zero because the optimal control problem is formulated with an infinite horizon. Substituting (22) and ϕ into the Euler-Lagrange equations and evaluating the derivatives results in

$$\dot{\boldsymbol{\lambda}} = -(\boldsymbol{\delta}^\top Q + \boldsymbol{\lambda}^\top [A + \frac{1}{2}S])^\top \quad (26)$$

$$\boldsymbol{\lambda}(t_f) = \mathbf{0} \quad (27)$$

$$\mathbf{u} = -R^{-1}B^\top \boldsymbol{\lambda} \quad (28)$$

where

$$S_m^i = 2A_{m,k}^i \delta^k \quad (29)$$

Finding the optimal feedback policy has been reduced to the problem of solving for the costate $\boldsymbol{\lambda}$. For nonlinear systems, the ordinary differential equation (ODE) (26) generally has no closed-form solution leading to a generally unavailable closed-form optimal control law. To make the problem tractable, we impose structure on the costate and truncate it as a polynomial in $\boldsymbol{\delta}$.

In order to incorporate the second-order contributions of (26), the costate must be at least quadratic in $\boldsymbol{\delta}$. The inclusion of higher polynomial orders in the costate form drastically increases the number of unknowns due to the *curse of dimensionality* and, without approximation, the costate would form an infinite series. Consequently, we restrict the space of admissible control laws to be quadratic in $\boldsymbol{\delta}$ and approximate the costate as

$$\boldsymbol{\lambda}(t) = P(t)\boldsymbol{\delta}(t) + \mathbf{L}(t)\boldsymbol{\delta}^2 \quad (30)$$

where $P \in \mathbb{R}^{n \times n}$ and $\mathbf{L} \in \mathbb{R}^{n \times n \times n}$ are symmetric in all indices. Given this costate form, the control policy (28) is known but the values of $P(t)$ and $\mathbf{L}(t)$ are undetermined. The remainder of this section is dedicated to obtaining a solution of P and \mathbf{L} .

To do so, the costate (30) can be differentiated with respect to time and equated by homogeneous polynomial orders with (26). The time derivative of the costate is given by

$$\dot{\boldsymbol{\lambda}} = \dot{P}\boldsymbol{\delta} + P\dot{\boldsymbol{\delta}} + \frac{d}{dt}(\mathbf{L}\boldsymbol{\delta}^2) \quad (31)$$

Evaluating the time derivative results in

$$\frac{d}{dt}L_{j,k}^i \delta^j \delta^k = \dot{L}_{j,k}^i \delta^j \delta^k + 2L_{j,k}^i \dot{\delta}^j \delta^k \quad (32)$$

$$= \dot{\mathbf{L}}\boldsymbol{\delta}^2 + \mathbf{L}\boldsymbol{\delta}\dot{\boldsymbol{\delta}} \quad (33)$$

where $\mathbf{L}\boldsymbol{\delta}\dot{\boldsymbol{\delta}}$ is shorthand for $(\mathbf{L}\boldsymbol{\delta}\dot{\boldsymbol{\delta}})^i = 2L_{j,k}^i \dot{\delta}^j \delta^k$. With this shorthand, (31) can be written as

$$\dot{\boldsymbol{\lambda}} = \dot{P}\boldsymbol{\delta} + P\dot{\boldsymbol{\delta}} + \dot{\mathbf{L}}\boldsymbol{\delta}^2 + \mathbf{L}\boldsymbol{\delta}\dot{\boldsymbol{\delta}} \quad (34)$$

Substituting (21) and (28) into $\mathbf{L}\boldsymbol{\delta}\dot{\boldsymbol{\delta}}$ and dropping higher-order terms gives

$$(\mathbf{L}\boldsymbol{\delta}\dot{\boldsymbol{\delta}})^i \approx 2L_{j,k}^i (A_c)_m^j \delta^m \delta^k \quad (35)$$

where $A_c \triangleq A - BR^{-1}B^\top P$ represents the linearized closed loop dynamics of the system. Substituting (21) and (28) into (34) gives

$$\dot{\lambda} = (\dot{P} + PA - PBR^{-1}B^\top P)\delta + \left(\frac{1}{2}PA^{(2)} - PBR^{-1}B^\top L + \dot{L} + 2LA_c\right)\delta^2 \quad (36)$$

Equation (26) can similarly be simplified to

$$\dot{\lambda} = -(Q\delta + A^\top \lambda + \frac{1}{2}S\lambda) \quad (37)$$

Inspecting $S\lambda$ individually,

$$S_m^i \lambda^m = 2A_{m,k}^i \delta^k P_n^m \delta^n + \mathcal{O}(\delta^3) \quad (38)$$

$$\approx 2A_{m,k}^i P_j^m \delta^j \delta^k \quad (39)$$

A truncation of monomials—in this case, those of $\mathcal{O}(\delta^3)$ —is unavoidable when the dynamics are nonlinear and the costate, and corresponding feedback policy, are restricted to finite degree polynomials. For example, if the costate was assumed to be cubic in δ , terms quartic in δ would appear in (26) and likewise require truncation.

Expanding $S\delta$ and grouping monomials gives

$$\dot{\lambda} = (-Q - A^\top P)\delta + (-A^\top L - A^{(2)}P)\delta^2 \quad (40)$$

Setting the δ and δ^2 terms in (36) and (40) equal and simplifying yields a matrix and tensor ODE

$$\dot{P} = -PG - A^\top P + PBR^{-1}B^\top P - Q \quad (41)$$

$$\dot{L} = -\frac{1}{2}PA^{(2)} - A^{(2)}P - A_c^\top L - 2LA_c \quad (42)$$

For finite-horizon problems, $P(t)$ and $L(t)$ satisfy the differential Riccati and Sylvester equations (41), (42), respectively and must be integrated backwards in time. Under an infinite horizon, $\lim_{t \rightarrow \infty} \dot{P} = 0$ and $\lim_{t \rightarrow \infty} \dot{L} = 0$ if the system is time-invariant.²² By this assumption, finding P and L reduces to solving the following algebraic Riccati equation (ARE) and tensor Sylvester equation

$$0 = A^\top P + PG - PBR^{-1}B^\top P + Q \quad (43)$$

$$0 = \frac{1}{2}PA^{(2)} + A^{(2)}P + A_c^\top L + 2LA_c \quad (44)$$

Having found P and L , the costate λ can be substituted into (28) to obtain the optimal control law

$$\mathbf{u}(t) = -R^{-1}B^\top P\delta(t) - R^{-1}B^\top L\delta^2(t) \quad (45)$$

The first-order component of a QQR controller is identical to an infinite-horizon LQR controller. The added tensor component of the control law provides a unique weight to each pair $\delta^i \delta^j$. Considering these cross-terms allows for an approximation of local nonlinearities, giving an improved performance over a first-order controller.

Remark: (Connection to Al'Brekht's Method) Al'brekht's method²³ is a recursive algorithm for calculating the Taylor series of the value function and accompanying state feedback to an arbitrary degree. Notably, it can be shown that the result obtained here is equivalent to that obtained using Al'brekht's dynamic programming based method.²⁴

Pursuit-Evasion Game Quadratic-Quadratic Regulator

Consider a pursuer and an evader whose motion are approximated by (21) which seek to minimize and maximize the objective (10) respectively. The second-order approximation of the agents' relative dynamics is given as

$$\dot{\mathbf{x}} = A\mathbf{x} + \frac{1}{2}\mathbf{A}^{(2)}\mathbf{x}^2 + B_p\mathbf{u}_p + B_e\mathbf{u}_e \quad (46)$$

where $B_e = -B_p$. The pursuer's objective function is given by (10). The terms A and $\mathbf{A}^{(2)}$ represent the Jacobian and second-order partial derivative tensor, respectively, of the relative orbital dynamics (4). Since the agents have opposing objectives, the evader's objective function is given by $J_e = -J_p$. Substitution into the first-order necessary conditions for the pursuer yields

$$\dot{\boldsymbol{\lambda}} = -(\mathbf{x}^\top Q + \boldsymbol{\lambda}^\top [A + \frac{1}{2}S])^\top \quad (47)$$

$$\boldsymbol{\lambda}(t_f) = \mathbf{0} \quad (48)$$

$$\mathbf{u}_p = -R^{-1}B_p^\top \boldsymbol{\lambda} \quad (49)$$

where S is defined by (29). The evader's optimal policy is then

$$\mathbf{u}_e = \frac{1}{\gamma^2}R^{-1}B_e^\top \boldsymbol{\lambda} \quad (50)$$

The costate again takes the form $\boldsymbol{\lambda} = P\boldsymbol{\delta} + \mathbf{L}\boldsymbol{\delta}^2$. Assuming steady state and following the method used to obtain the single-agent optimal control policy gives the terms P and \mathbf{L} as the solutions to

$$0 = PA + A^\top P - PB_p R^{-1} B_p^\top P + \frac{1}{\gamma^2} P B_e R^{-1} B_e^\top P + Q \quad (51)$$

$$\mathbf{0} = \frac{1}{2} P \mathbf{A}^{(2)} + A_c^\top \mathbf{L} + 2\mathbf{L} A_c + \mathbf{A}^{(2)} P \quad (52)$$

where $A_c = A - B_p R^{-1} B_p^\top P + \frac{1}{\gamma^2} B_e R^{-1} B_e^\top P$ is the linearized closed loop dynamics. The differential Riccati equation (DRE) (51) matches the standard form for a zero-sum linear-quadratic differential game.¹⁹ The DRE, or accompanying ARE if steady state, may have a conjugate point in which the solution does not exist. If γ is chosen to be sufficiently large, the game admits a unique saddle point solution.¹⁹

The second-order control component varies if the adversarial agent is using a first- or second-order policy. If the pursuer is known to be using a first-order controller, the evader's second-order equation (52) must reflect this by being rewritten as

$$\mathbf{0} = -\frac{1}{2} P \mathbf{A}^{(2)} - \frac{1}{\gamma^2} P B_e R^{-1} B_e^\top \mathbf{L} - 2\mathbf{L} A_c - A^\top \mathbf{L} - \mathbf{A}^{(2)} P \quad (53)$$

Similarly, if the evader is known to be using a first-order controller, the pursuer must alter (52) to instead be

$$\mathbf{0} = -\frac{1}{2} P \mathbf{A}^{(2)} + P B_p R^{-1} B_p^\top \mathbf{L} - 2\mathbf{L} A_c - A^\top \mathbf{L} - \mathbf{A}^{(2)} P \quad (54)$$

The first-order component given by (51) is unchanged so long as the adversary is using at least a first-order expansion.

Practical Considerations

In the context of orbital PE games, second-order expansions of the relative dynamics do not provide a suitable approximation over a large enough area in the state space to allow for static gains. Motivated by prior work on state-dependent methods for H_∞ and differential game applications,^{14,25} the state-dependent approach is taken where the partial derivative tensors are periodically evaluated at state x and (51) and (52) are solved periodically. We introduce this method as the *state-dependent Riccati Sylvester Equation* (SDRSE), an extension of the SDRE method to quadratic feedback. A comparison between the the SDRE and SDRSE methods is depicted in Figure 3.

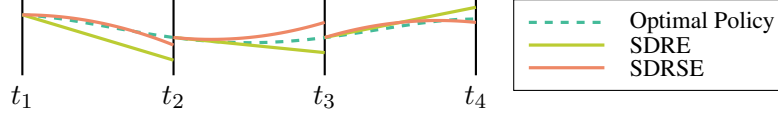


Figure 3 Trajectories of control policies obtained from state-dependent methods updated periodically at time t_j compared to the globally optimal solution.

The ARE is solved with typical QZ decomposition methods, however in regions of weak controllability is integrated backwards from a large final to approximate an infinite horizon.

$$P \approx \int_0^{t_f} \dot{P} dt, \quad P(t_f) = 0 \quad (55)$$

where $t_f \gg 0$ to approximate the terminal condition.

The state-dependent terms $P(x)$ and $L(x)$ are recalculated every 20 seconds in simulation time, consistent with prior SDRE work for spacecraft pursuit-evasion games.¹⁴ This interval is small enough to capture changes in the relative dynamics without performing unnecessary recomputation at each ODE step, which is dynamically determined to meet integration tolerances with the initial step being two seconds.

The equation for L , (52), is a tensor Sylvester equation of the form $AX + XB = C$. It may be explicitly written in this form by defining

$$A = A_c^\top \quad (56)$$

$$B = 2A_c \quad (57)$$

$$C = -\frac{1}{2}PA^{(2)} - A^{(2)}P \quad (58)$$

$$X = L \quad (59)$$

The tensor Sylvester equation is linear in X and can be solved by rewriting the system as

$$(I \otimes A + B^\top \otimes I)\text{vec}(X^i) = \text{vec}(C^i) \quad (60)$$

where \otimes denotes the Kronecker product

$$(F \otimes A)_j^i = f_{ij}A = \begin{bmatrix} f_{11}A & f_{12}A & \dots \\ f_{21}A & f_{22}A & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \quad (61)$$

and $\text{vec}(X^i)$ is a n^2 length vector of the elements of X^i listed in column major order. The third-order tensor \mathbf{X} may be solved explicitly slice-wise along the third index n times because the only tensor term, $C_{j,k}^i = -\frac{1}{2}P_m^i A_{j,k}^m - A_{m,k}^i P_j^m$, is decoupled along the i^{th} index. For different agent control form permutations (53), (54), the factorization into standard Sylvester equation form must be altered to reflect belief.

APPLICATIONS TO SPACECRAFT PURSUIT-EVASION GAMES

In this section, we demonstrate that SDRSE policies and SDRE policies find similar terminal states—yet, SDRSE policies produce trajectories that outperform their first-order counterparts. The following analysis considers nine unique game permutations arising from the first- and second-order controllers and their associated beliefs about the opponent’s controller form. The different game configurations are shown in Table 1. These scenarios are of interest as it can give the relative performance of first- and second-order controllers in competition. Furthermore, because higher-order controllers require a belief about the opponents form, there can be a mismatch between belief and reality. When deciding to use a first- or second-order controller in spaceflight operations, it is imperative to note that a belief may be incorrect. Thus, studying the effects of incorrect beliefs are critical to analyzing the effectiveness of a second-order controller.

Each scenario is simulated using the simulation parameters adopted from Jagat and Sinclair,¹⁴ listed in Table 2. The initial states are given in terms of the classical orbital elements ($a, e, i, \Omega, \omega, \nu$) which represent semi-major axis, eccentricity, inclination, right ascension of the ascending node, argument of periapsis, and true anomaly, respectively. The large initial differences in orbital state, while drastic, represent a stressing case that clearly reveals the effects of nonlinear dynamics on controller performance. Figure 4 shows the original orbits specified by the initial conditions prior to control application. Simulations are propagated for a timespan equivalent to two periods of the evader’s initial orbit. The scaling parameter γ^2 gives the evader twice as much control penalty weighting as the evader, meaning that it is more costly for the evader to maneuver than the pursuer. This scaling puts the pursuer at a direct advantage in the simulation since it is more willing to perform maneuvers.

To analyze outcomes across game permutations, rendezvous time and orbital elements are used to compare the responses and ΔV expenditure is used as measures of control effort. The total payoffs for each scenario are considered to determine which agent achieves their objective more effectively.

Table 1 Unique game permutations.

Evader (form, belief)	Pursuer (form, belief)		
	SDRE	(SDRSE, SDRE)	(SDRSE, SDRSE)
SDRE	Game 1	Game 2	Game 3
(SDRSE, SDRE)	Game 4	Game 5	Game 6
(SDRSE, SDRSE)	Game 7	Game 8	Game 9

Due to system dynamics and the evader’s effort, exact convergence to $\mathbf{x} = \mathbf{0}$ may be slow in some scenarios. To mitigate this issue, we seek to find the time after which the system deviation is bounded by some design constant for the remainder of the simulation. We define rendezvous as

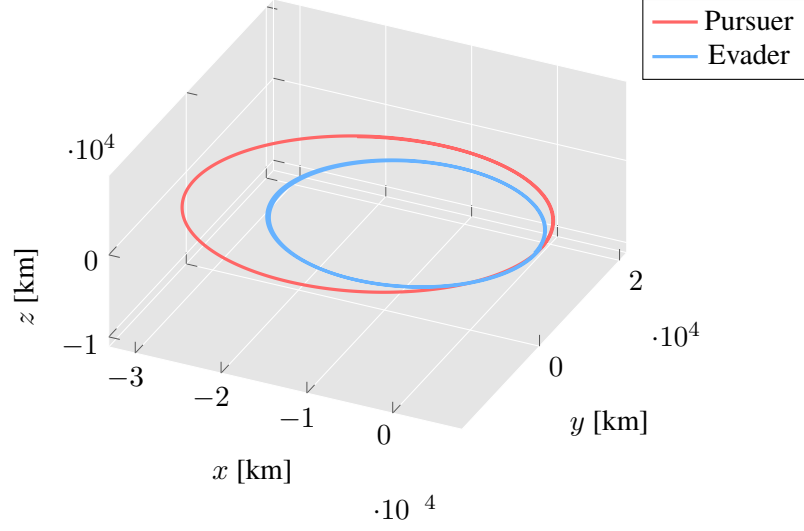


Figure 4 Uncontrolled initial orbits for pursuer and evader.

Table 2 Simulation parameters and initial conditions.

Orbital Elements	Pursuer Values	Evader Values
a [km]	20584	15000
e	0.612	0.5
i [deg]	3.573	0
Ω [deg]	270.963	0
ω [deg]	83.020	0
ν [deg]	9.588	0
Q	$\begin{bmatrix} I_{3 \times 3} \text{ km}^{-2} & \mathbf{0}_{3 \times 3} \text{ s km}^{-2} \\ \mathbf{0}_{3 \times 3} \text{ s km}^{-2} & I_{3 \times 3} \text{ s}^2 \text{ km}^{-2} \end{bmatrix}$	
R	$I_{3 \times 3} \times 10^{13} \text{ s}^4 \text{ km}^{-2}$	
γ^2	2	

when agents sufficiently close to each other and the convergence of the relative state has decayed to a negligible amount. In all scenarios inspected, rendezvous occurs with the relative distance being less than 200 m and the relative velocity less than 0.35 m/s. The rendezvous bounds are set as a percentage of the state’s norm at the end of simulation to identify when convergence slows. In an analog to the settling-time metric used in frequency domain analysis, we choose a 2% margin as the bounds for rendezvous. Formally, the rendezvous time, t_r , is given by

$$t_r = \min\{t \mid \left| \frac{\|\mathbf{x}(\tau)\|_Q - \|\mathbf{x}(t_f)\|_Q}{\|\mathbf{x}(t_f)\|_Q} \right| \leq 0.02 \forall \tau \geq t\} \quad (62)$$

where t_f is the final value of simulation time and $\|\mathbf{x}\|_Q$ denotes norm of vector \mathbf{x} induced by the weight matrix Q , which is constant across all scenarios. By this induced norm, the state is nondimensionalized allowing for both position and velocity components to be accounted for in the scalar. All rendezvous times, listed in Table 3, occur within 140 seconds of each other with the fastest occurring when the pursuer is second-order and mistakenly thinks the evader is second-order. The slowest rendezvous occurs when the evader is second-order regardless of belief. In both of these scenarios, the agent with a higher-order controller is able to alter the rendezvous time in its favor. Notably, the second-order pursuer with correct belief obtains a slower rendezvous with a first-order evader than if it had an incorrect belief.

Table 3 Rendezvous time (2%) [s].

Evader (form, belief)	Pursuer (form, belief)		
	SDRE	(SDRSE, SDRE)	(SDRSE, SDRSE)
SDRE	18600	18520	18480
(SDRSE, SDRE)	18620	18560	18520
(SDRSE, SDRSE)	18620	18560	18520

The trajectories of the Keplerian elements for each scenario are shown in Figure 5 where the red curves indicates a pursuer and the blue an evader. While there is some deviation in the terminal elements, primarily eccentricity, all game permutations converge to a similar terminal orbit. The ECI trajectories taken are shown in Figure 6. The figure shows that all controller pairs take different paths to reach a similar terminal orbit. The varied trajectories are further demonstrated by the relative trajectory in Figure 7. The trajectories deviate more in the beginning of the simulation, before eventually coalescing at rendezvous.

The total ΔV expenditure for the agents are listed for each scenario in Table 4. It is worth noting that these values are not representative of SDRE and SDRSE controllers in realistic scenarios, and careful tuning would further reduce the values. They do, however, show a trend between the various controller forms and belief of of opponent controller forms. In Game 1, the ΔV ratio is 0.5, indicating that the evader spend expended half as much ΔV as the pursuer over the course of the game. This ratio is a product of the weighting parameter γ^2 which weights the evader’s fuel expenditure twice as much as the pursuers. Game 1 is taken as a baseline to compare the other scenarios against. A game with a favorable ΔV exchange for the pursuer has a larger value of $\Delta V_e/\Delta V_p$ and a smaller value is favorable for the evader. Comparing the pursuer’s and the

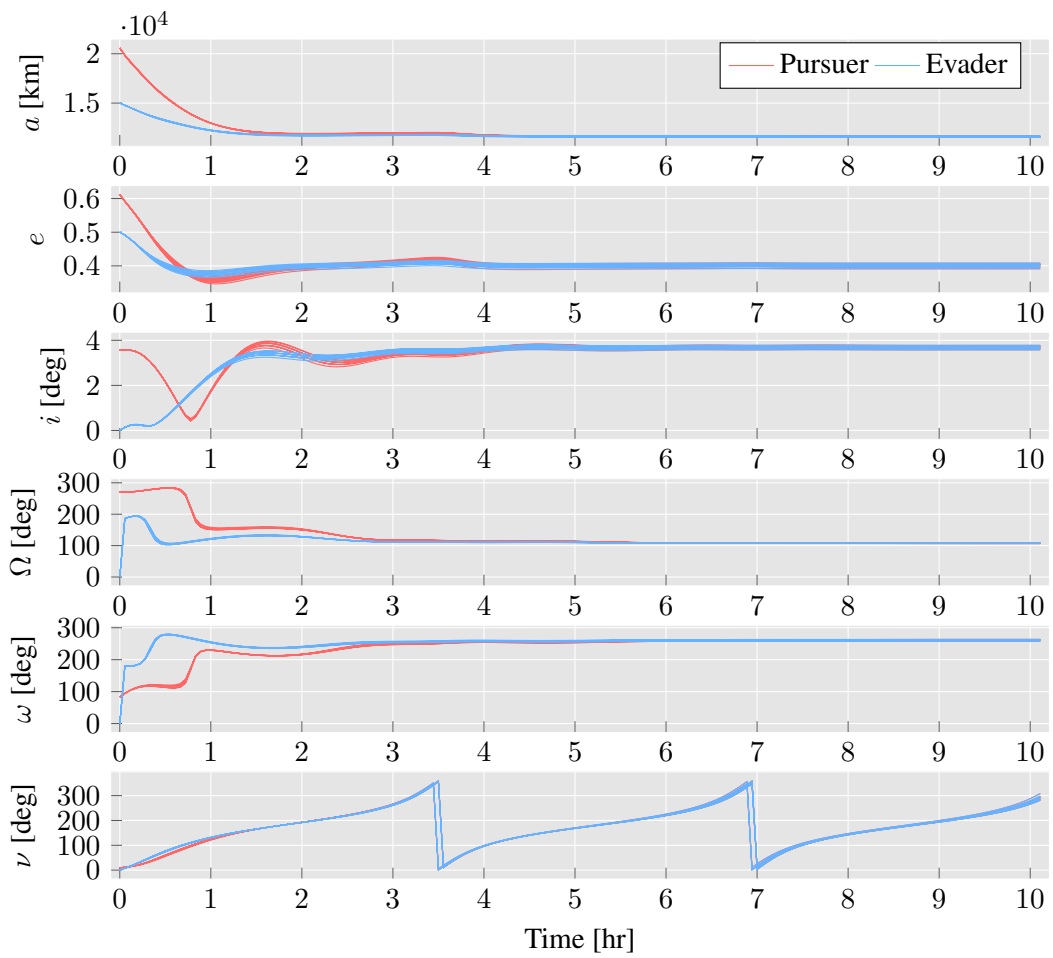


Figure 5 Orbital elements trajectories for configurations with different controller form and different belief of opponent controller form.

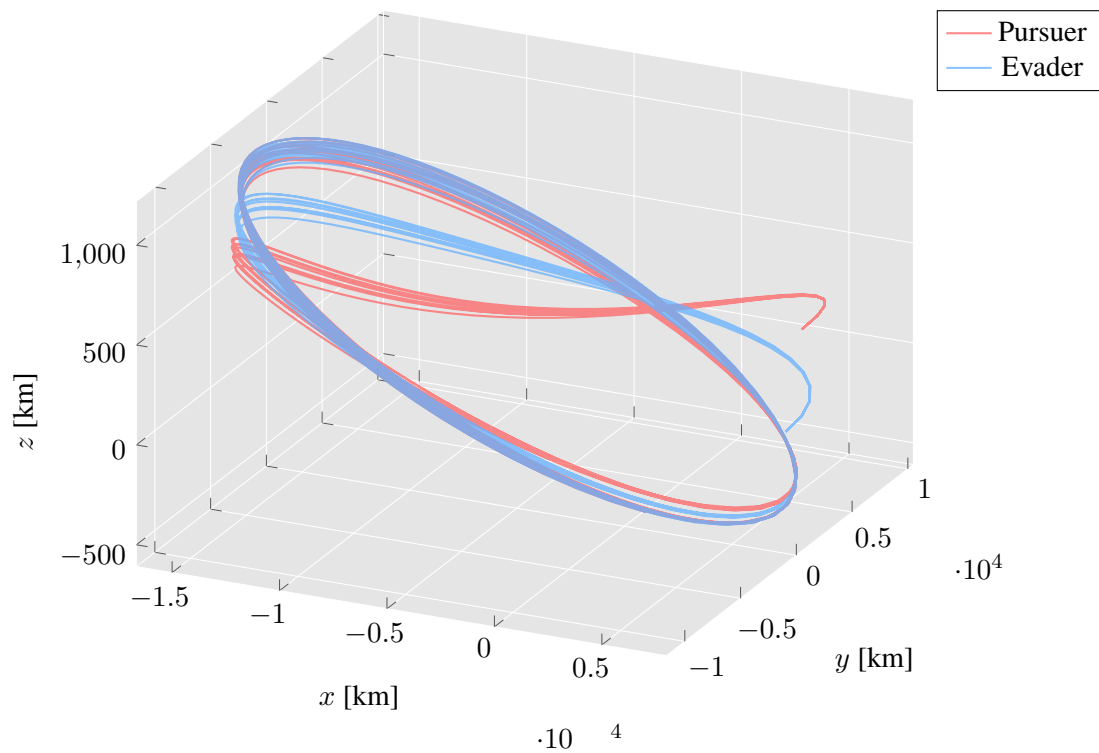


Figure 6 Agent ECI trajectories for different configurations of controller form and belief of opponent's controller form. Non-equal axis scalings are used to better show differences in z motion.

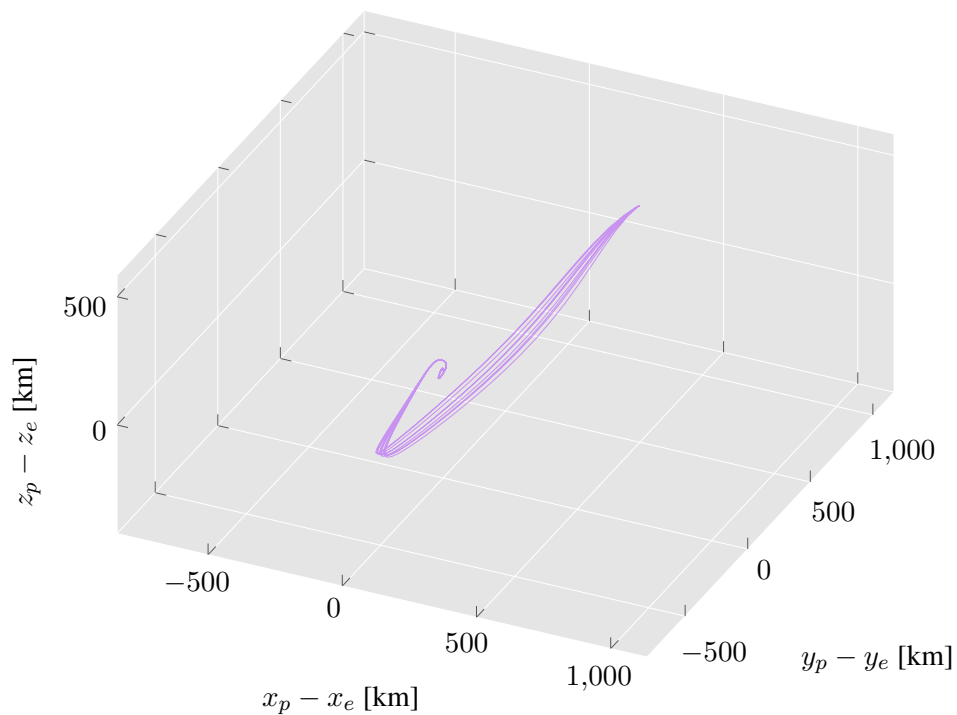


Figure 7 Relative trajectories, $r_p - r_e$, for all scenarios.

evader's ΔV values with respect to each other is more meaningful than comparing the absolute values because the payoff structure (10) rewards the adversary using more control effort in the same way it penalizes using control effort. In Games 2 and 3, the pursuer is using a second-order controller against a first-order opponent. As expected, in both of these scenarios the pursuer achieves a ΔV quotient greater than the baseline. Games 4 and 7 involve a second-order evader competing against a first-order pursuer. In these games, the ΔV quotient is smaller than the baseline, indicating that the pursuer is achieving a favorable effort trade-off. Both agents use second-order controllers in Games 5, 6, 8, and 9. In Game 5, both agents have incorrect beliefs about the other agent's controller, and the evader expends slightly less than half as much ΔV as the pursuer. Games 6 and 8 involve one agent with a correct belief and one agent with an incorrect belief. In both scenarios, the agent with the correct belief achieves a favorable outcome. It should be noted that the ΔV quotient in Game 8 is only slightly to the evader's advantage. In Game 9, both agents use second-order controllers with correct beliefs. The pursuer achieves a favorable outcome in terms of ΔV for this game.

The maximum control vector was found to be an uninformative metric for the controllers. Across all scenarios, the maximum control vector for the pursuer ranges between 3.4 and 3.5 [m/s²], whereas the evader's is between 1.7 and 1.8 [m/s²].

Table 4 Agent ΔV for configurations with different controller form and different belief of opponent controller form [km/ s].

Game Number	Pursuer Controller	Pursuer Belief	Evader Controller	Evader Belief	ΔV_p	ΔV_e	$\Delta V_e/\Delta V_p$
Game 1	SDRE	N/A	SDRE	N/A	4.52	2.26	0.500
Game 2	SDRSE	SDRE	SDRE	N/A	4.59	2.33	0.508
Game 3	SDRSE	SDRSE	SDRE	N/A	4.74	2.46	0.519
Game 4	SDRE	N/A	SDRSE	SDRE	4.40	2.16	0.491
Game 5	SDRSE	SDRE	SDRSE	SDRE	4.44	2.21	0.498
Game 6	SDRSE	SDRSE	SDRSE	SDRE	4.56	2.31	0.507
Game 7	SDRE	N/A	SDRSE	SDRSE	4.40	2.16	0.491
Game 8	SDRSE	SDRE	SDRSE	SDRSE	4.46	2.23	0.499
Game 9	SDRSE	SDRSE	SDRSE	SDRSE	4.60	2.35	0.510

The payoffs for each game are listed in Table 5. The values are normalized with respect to the payoff Game 1 to compare relative performance to engagements where both agents are first-order. The pursuer is able to increase the payoff compared to the baseline in the scenarios where it is second-order and the pursuer is first-order. It should be noted that the payoff is larger when the evader's belief is incorrect. In all other scenarios, the pursuer achieves a favorable payoff. The lowest payoff is obtained when the pursuer is second-order and has an incorrect belief about the evader's form.

The trend shown in rendezvous time, ΔV , and payoff demonstrate that second-order controllers consistently outperform first-order ones. Incorrect beliefs of the opponents form do not degenerate the controllers effectiveness, and can even improve it in some scenarios.

Table 5 Payoff normalized against the payoff for a first-order pursuer and a first-order evader. A higher value is desired for the evader and a lower value for the pursuer.

Evader (form, belief)	Pursuer (form, belief)		
	SDRE	(SDRSE, SDRE)	(SDRSE, SDRSE)
SDRE	1.000	0.967	0.931
(SDRSE, SDRE)	1.008	0.979	0.950
(SDRSE, SDRSE)	1.010	0.980	0.948

CONCLUSION

This work presents higher-order approximate optimal control solutions for non-cooperative spacecraft pursuit-evasion games. The impact of the new approach on game outcomes is studied in simulations across a wide variety of player controller order and opponent belief accuracy configurations. With identical initial conditions, these permutations showed little impact to the time of interception despite yielding different trajectories in the relative state-space. Higher-order controllers achieved favorable relative ΔV expenditure than lower-order controllers compared to the baseline case. The inclusion of higher-order terms also improved payoffs against lower-order opponents. Notably, incorrect belief structures did not meaningfully degenerate the performance of higher-order controllers and at times improved it. This insensitivity to opponent belief uncertainty, in tandem with periodic gain updates, may reasonably offer robust performance against a wide array of unanticipated opponent strategies. Future work will involve analysis of the performance of higher-order control policies with varying state-dependent update cadences. The integration of probabilistic state- and intent-estimation with the proposed control strategy will also be considered.

ACKNOWLEDGMENT

This material is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-25-1-0101

REFERENCES

- [1] J. O. Woods and J. A. Christian, "Lidar-Based Relative Navigation with Respect to Non-Cooperative Objects," *Acta Astronautica*, Vol. 126, Sept. 2016, pp. 298–311.
- [2] M. Maestrini, M. A. De Luca, and P. Di Lizia, "Relative Navigation Strategy About Unknown and Uncooperative Targets," *Journal of Guidance, Control, and Dynamics*, Vol. 46, No. 9, 2023, pp. 1708–1725.
- [3] A. Wu, S. Zhang, L. Shu, C. Si, and X. Wan, "Non-cooperative Spacecraft Relative Navigation Based on Monocular Camera in Space On-orbit Servicing," *2022 International Conference on Service Robotics (ICoSR)*, June 2022, pp. 133–138.
- [4] K. A. LeGrand, K. J. DeMars, and H. J. Pernicka, "Bearings-Only Initial Relative Orbit Determination," *Journal of Guidance, Control, and Dynamics*, Vol. 38, No. 9, 2015, pp. 1699–1713.
- [5] K. A. LeGrand and K. J. DeMars, "Relative Multiple Space Object Tracking Using Intensity Filters," *Proceedings of the 18th International Conference on Information Fusion (2015, Washington, DC)*, Sept. 2015, pp. 1253–1261.
- [6] J. Iannamorelli, S. Semeraro, K. A. LeGrand, and C. Frueh, "Adaptive Filtering for Multi-Satellite Relative Navigation with Quantized Measurements," *AIAA SCITECH 2024 Forum*, AIAA SciTech Forum, American Institute of Aeronautics and Astronautics, Jan. 2024.
- [7] K. A. LeGrand and K. J. DeMars, "Analysis of a Second-order Relative Lambert Solver," *Proceedings of the 2014 AIAA/AAS Astrodynamics Specialists Conference*, 2014.

- [8] J. Boh, Z. Funke, and M. Akella, "A Game-Theoretical Exploration of L1/L2 Cislunar Space Situational Awareness Using Bayesian Games," *The Advanced Maui Optical and Space Surveillance (AMOS) Technologies Conference*, 2025, p. 63.
- [9] D. Zhang, S. Song, and R. Pei, "Safe Guidance for Autonomous Rendezvous and Docking with a Non-Cooperative Target," *AIAA Guidance, Navigation, and Control Conference*, Toronto, Ontario, Canada, American Institute of Aeronautics and Astronautics, Aug. 2010.
- [10] M. Dor and P. Tsiotras, "ORB-SLAM Applied to Spacecraft Non-Cooperative Rendezvous," *2018 Space Flight Mechanics Meeting*, Kissimmee, Florida, American Institute of Aeronautics and Astronautics, Jan. 2018.
- [11] G. Sun, M. Zhou, and X. Jiang, "Non-Cooperative Spacecraft Proximity Control Considering Target Behavior Uncertainty," *Astrodynamics*, Vol. 6, Dec. 2022, pp. 399–411.
- [12] J. Kulik, W. Clark, and D. Savransky, "State Transition Tensors for Continuous-Thrust Control of Three-Body Relative Motion," *Journal of Guidance, Control, and Dynamics*, Vol. 46, No. 8, 2023, pp. 1610–1619.
- [13] R. Isaacs, *Differential Games*. Dover, 1965.
- [14] A. Jagat and A. J. Sinclair, "Nonlinear Control for Spacecraft Pursuit-Evasion Game Using the State-Dependent Riccati Equation Method," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 53, Dec. 2017, pp. 3032–3042.
- [15] M. Pontani and B. A. Conway, "Numerical Solution of the Three-Dimensional Orbital Pursuit-Evasion Game," *Journal of Guidance, Control, and Dynamics*, Vol. 32, Mar. 2009, pp. 474–487.
- [16] Z. Zheng, P. Zhang, and J. Yuan, "Nonzero-Sum Pursuit-Evasion Game Control for Spacecraft Systems: A Q-Learning Method," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 59, Aug. 2023, pp. 3971–3981.
- [17] Z.-y. Li, H. Zhu, Z. Yang, and Y.-z. Luo, "A Dimension-Reduction Solution of Free-Time Differential Games for Spacecraft Pursuit-Evasion," *Acta Astronautica*, Vol. 163, Oct. 2019, pp. 201–210.
- [18] T. Başar and G. Olsder, *Dynamic Noncooperative Game Theory*. SIAM, 2. ed ed., 1999.
- [19] T. Başar and P. Bernhard, "Continuous-Time Systems with Perfect-State Measurements," *H-infinity-Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach* (T. Başar and P. Bernhard, eds.), pp. 107–188, Boston, MA: Birkhäuser, 2008.
- [20] T. Çimen, "Systematic and Effective Design of Nonlinear Feedback Controllers via the State-Dependent Riccati Equation (SDRE) Method," *Annual Reviews in Control*, Vol. 34, Apr. 2010, pp. 32–51.
- [21] D. Kirk, *Optimal Control Theory*. Dover, 1970.
- [22] B. Friedland, *Control System Design: An Introduction to State-Space Methods*. New York : McGraw-Hill, 1986.
- [23] E. G. Al'brekht, "On the Optimal Stabilization of Nonlinear Systems," *Journal of Applied Mathematics and Mechanics*, Vol. 25, Jan. 1961, pp. 1254–1266.
- [24] J. Borggaard and L. Zietsman, "The Quadratic-Quadratic Regulator Problem: Approximating feedback controls for quadratic-in-state nonlinear systems," *2020 American Control Conference (ACC)*, July 2020, pp. 818–823.
- [25] A. J. v. d. Schaft, "On a State Space Approach to Nonlinear H-Infinity Control," *Systems & Control Letters*, Vol. 16, Jan. 1991, pp. 1–8.