دانشگاه صنعتی شریف
دانشکده‌ی مهندسی هوافضا

پروژه کارشناسی ارشد
مهندسی فضا

عنوان:

هدایت یادگیری تقویتی مقاوم مبتنی بر بازی دیفرانسیلی در محیط‌های پویای چندجسمی با پیشران کم

نگارش:

علی بنی اسد

استاد راهنما:

دکتر هادی نوبهاری

دی ۱۴۰۳

بسم الله الرحمن الرحیم

به نام خدا

دانشگاه صنعتی شریف

دانشکده‌ی مهندسی هوافضا

پروژه کارشناسی ارشد

عنوان: هدایت یادگیری تقویتی مقاوم مبتنی بر بازی دیفرانسیلی در محیط‌های پویای چندجسمی با پیشران کم

نگارش: علی بنی اسد

کمیته‌ی ممتحنین

استاد راهنما: دکتر هادی نوبهاری امضاء:

استاد مشاور: استاد مشاور امضاء:

استاد مدعو: استاد ممتحن امضاء:

تاریخ:

سپاس

از استاد بزرگوارم جناب آقای دکتر نوبهاری که با کمک‌ها و راهنمایی‌های بی‌دریغشان، بنده را در انجام این پروژه یاری داده‌اند، تشکر و قدردانی می‌کنم. از پدر دلسوزم ممنونم که در انجام این پروژه مرا یاری نمود. در نهایت در کمال تواضع، با تمام وجود بر دستان مادرم بوسه می‌زنم که اگر حمایت بی‌دریغش، نگاه مهربانش و دستان گرمش نبود برگ برگ این دست نوشته و پروژه وجود نداشت.

چکیده

در این پژوهش، یک چارچوب هدایت مقاوم برای فضاپیماهای کم‌پیشران در محیط‌های دینامیکی چندجسمی (مدل CRTBP زمین–ماه) ارائه شده است. مسئله به‌صورت بازی دیفرانسیلی مجموع‌صفر بین عامل هدایت (فضاپیما) و عامل مزاحم (عدم قطعیت‌های محیطی) فرمول‌بندی شده و با رویکرد آموزش متمرکز–اجرای توزیع‌شده پیاده‌سازی گردیده است. در این راستا، چهار الگوریتم یادگیری تقویتی پیوسته DDPG، TD3، SAC و PPO به نسخه‌های چندعاملی مجموع‌صفر گسترش یافته‌اند (MA-DDPG، MATD3، MASAC و MAPPO) و جریان آموزش آن‌ها همراه با ساختار شبکه‌ها در قالب ارزش–سیاست مشترک تشریح شده است.

ارزیابی الگوریتم‌ها در سناریوهای متنوع عدم قطعیت شامل شرایط اولیه تصادفی، اغتشاش عملگر، نویز حسگر، تأخیر زمانی و عدم تطابق مدل روی مسیر مدار لیاپانوف زمین–ماه انجام گرفت. نتایج به‌وضوح نشان می‌دهد که نسخه‌های مجموع‌صفر در تمامی معیارهای ارزیابی بر نسخه‌های تک‌عاملی برتری دارند. به‌ویژه الگوریتم MATD3 با حفظ پایداری سیستم، کمترین انحراف مسیر و مصرف سوخت بهینه را حتی در سخت‌ترین سناریوهای آزمون از خود نشان داد.

به‌منظور تسهیل استقرار عملی، سیاست‌های آموخته‌شده روی بستر ROS 2 با بهره‌گیری از کوانتیزاسیون INT8 و تبدیل به فرمت ONNX پیاده‌سازی شدند. این بهینه‌سازی‌ها زمان استنتاج را به ۵٫۸ میلی‌ثانیه و مصرف حافظه را به ۹٫۲ مگابایت کاهش داد که به‌ترتیب بهبود ۴۷ درصدی و ۵۳ درصدی نسبت به مدل FP32 را نشان می‌دهد، در حالی‌که چرخه کنترل ۱۰۰ هرتز بدون هیچ‌گونه نقض زمانی حفظ شد.

در مجموع، چارچوب پیشنهادی نشان می‌دهد که یادگیری تقویتی چندعاملی مبتنی بر بازی دیفرانسیلی می‌تواند بدون نیاز به مدل‌سازی دقیق، هدایت تطبیقی و مقاوم فضاپیماهای کم‌پیشران را در نواحی ذاتاً ناپایدار سیستم‌های سه‌جسمی تضمین کند و برای پیاده‌سازی روی سخت‌افزار در حلقه آماده باشد.

کلیدواژه‌ها: یادگیری تقویتی عمیق، بازی دیفرانسیلی، سیستم‌های چندعاملی، هدایت کم‌پیشران، مسئله محدود سه‌جسمی، کنترل مقاوم.

فهرست مطالب

فهرست جداول

فهرست تصاویر

فهرست الگوریتم‌ها

فصل ۱

مدل‌سازی محیط یادگیری سه جسمی

مسیرهای فضایی در بسیاری از مأموریت‌ها نه‌تنها تحت تأثیر گرانش یک جسم مرکزی (مانند خورشید یا زمین)، بلکه به‌طور هم‌زمان تحت نفوذ دست‌کم یک جسم دیگر نیز هستند. در این وضعیت، مدل‌های دوجسمی با اختلالات جسم سوم دقت کافی ندارند و باید دینامیک دو جسمِ اصلی و اثرات آن‌ها به‌صورت هم‌زمان در نظر گرفته شود. مسئله‌ی سه‌جسمیِ محدود با دو جسم اصلی و یک جسم سوم با جرم ناچیز (فضاپیما) چارچوبی طبیعی برای مطالعه‌ی این پدیده‌ها و نیز یک محیط مناسب برای به‌کارگیری روش‌های یادگیری تقویتی است؛ زیرا دینامیک غیرخطی و پیچیده‌ی آن ویژگی‌های غنی (مانند نقاط تعادلِ لاگرانژ) ایجاد می‌کند.

در این فصل ابتدا در بخش ۱-۱ دستگاهِ بی‌بُعد و چارچوبِ چرخان تعریف شده و در بخش ۱-۲ معادلاتِ حرکت در مسئله‌ی سه‌جسمیِ محدودِ دایره‌ای استخراج می‌شود.

۱-۱ مسئله‌ی سه‌جسمیِ محدودِ دایره‌ای (CRTBP)

دو جرمِ اصلی (زمین با جرم m_1 و ماه با جرم m_2) روی مدارهایی دایره‌ای و هم‌صفحه پیرامونِ مرکزِ جرمِ مشترک حرکت می‌کنند. جرمِ سوم (فضاپیما با جرمِ ناچیز m_3) چنان کوچک فرض می‌شود که تأثیرِ گرانشیِ آن بر حرکتِ دو جسمِ اصلی قابلِ صرفِ نظر است؛ بدین ترتیب، مسئله‌ی سه‌جسمیِ محدودِ دایره‌ای شکل می‌گیرد.

جدول ۱-۱: مقادیر عددی برای مسئله سه‌جسمی محدود (سیستم زمین–ماه)

پارامتر	توصیف	مقدار عددی
m_1	جرم زمین	$5.972 \times 10^{24}\,\mathrm{kg}$
m_2	جرم ماه	$7.348 \times 10^{22}\,\mathrm{kg}$
μ	نسبت جرمی	0.0121505856
ω	سرعت زاویه‌ای سیستم	$2.6617 \times 10^{-6}\,\mathrm{rad/s}$

دستگاهِ مختصاتِ چرخانی هم‌دوران با دو جرم اصلی انتخاب می‌شود؛ مبدأ در مرکزِ جرمِ سامانه است، محور x خطِ واصلِ دو جرم و محور y بر آن عمود (در صفحه‌ی مدارها) است. واحدِ طول برابر فاصله‌ی ثابتِ میان دو جرم و واحدِ زمان چنان تعریف می‌شود که دوره‌ی مداریِ سامانه 2π (و در نتیجه $\omega = 1$) گردد. همچنین جرم‌ها به‌گونه‌ای مقیاس می‌شود که مجموع دو جرم برابر با یک شود:

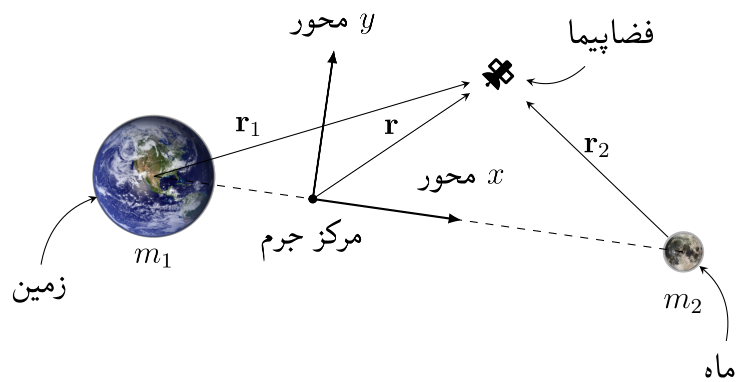
$$m_1 + m_2 = 1. \tag{۱-۱}$$

با نسبتِ جرمی

$$\mu \equiv \frac{m_2}{m_1 + m_2}, \tag{۱-۲}$$

داریم $m_1 = 1 - \mu$ و $m_2 = \mu$ و مکانِ دو جرم در دستگاهِ بی‌بُعد به صورت

$$\mathbf{r}_{\text{Earth}} = (-\mu, 0), \qquad \mathbf{r}_{\text{Moon}} = (1 - \mu, 0). \tag{۱-۳}$$



شکل ۱-۱: هندسه‌ی مسئله‌ی سه‌جسمیِ محدود در چارچوبِ چرخان

۱-۱-۱ لاگرانژ و معادلات حرکت

با در نظر گرفتن $G=1$ در حالت بی‌بُعد، تابعِ لاگرانژِ جرمِ سوم در دستگاهِ چرخان برابر است با [۱]

$$L=\tfrac{1}{2}\big(\dot{x}^2+\dot{y}^2+\dot{z}^2\big)+(1-\mu)\,\frac{1}{r_1}+\mu\,\frac{1}{r_2}+\tfrac{1}{2}\big(x^2+y^2\big), \tag{۱-۴}$$

که در آن

$$r_1=\sqrt{(x+\mu)^2+y^2+z^2},\qquad r_2=\sqrt{\big(x-1+\mu\big)^2+y^2+z^2}. \tag{۱-۵}$$

با به‌کارگیری رابطه‌ی اویلر–لاگرانژ

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_i}-\frac{\partial L}{\partial q_i}=0,\qquad q_i\in\{x,y,z\},$$

معادلاتِ بی‌بُعدِ حرکتِ جرمِ سوم به دست می‌آید:

$$\ddot{x}-2\dot{y}=x-\frac{1-\mu}{r_1^3}(x+\mu)-\frac{\mu}{r_2^3}\big(x-1+\mu\big), \tag{۱-۶}$$

$$\ddot{y}+2\dot{x}=y-\frac{1-\mu}{r_1^3}\,y-\frac{\mu}{r_2^3}\,y, \tag{۱-۷}$$

$$\ddot{z}=-\frac{1-\mu}{r_1^3}\,z-\frac{\mu}{r_2^3}\,z. \tag{۱-۸}$$

یا به نگاشتِ برداری به‌صورت زیر است.

$$\ddot{\mathbf{r}}+2\,\boldsymbol{\omega}\times\dot{\mathbf{r}}=\nabla\Omega(\mathbf{r}),\qquad \Omega(x,y,z)=\tfrac{1}{2}\big(x^2+y^2\big)+\frac{1-\mu}{r_1}+\frac{\mu}{r_2}. \tag{۱-۹}$$

که در آن Ω پتانسیلِ مؤثر است و در بخش ۱-۲ برای یافتنِ نقاطِ تعادل از شرطِ $\nabla\Omega=0$ استفاده می‌شود.

۱-۲ نقاط تعادلِ لاگرانژ

نقطه‌ی تعادل مکانی است که در چارچوبِ چرخان، جرمِ سوم بی‌حرکت می‌ماند. این شرط با صفر شدن مؤلفه‌های سرعت و شتاب حاصل می‌شود؛ ازاین‌رو در معادلاتِ حرکتِ بخش ۱-۱ قرار می‌دهیم $\dot{x}=\dot{y}=\dot{z}=\ddot{x}=\ddot{y}=\ddot{z}=0$. در نتیجه دستگاهِ جبریِ زیر برای مختصاتِ نقطه‌ی تعادل به‌دست می‌آید:

$$0=x-\frac{1-\mu}{r_1^3}(x+\mu)-\frac{\mu}{r_2^3}\big(x-1+\mu\big), \tag{۱-۱۰}$$

$$0=y\Big[1-\frac{1-\mu}{r_1^3}-\frac{\mu}{r_2^3}\Big], \tag{۱-۱۱}$$

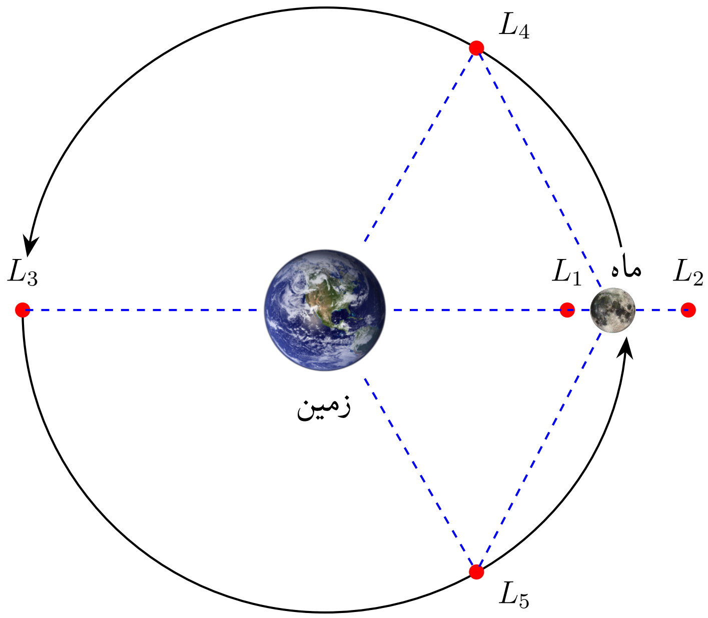
$$0=-\frac{1-\mu}{r_1^3}\,z-\frac{\mu}{r_2^3}\,z. \tag{۱-۱۲}$$

معادله‌ی سوم نشان می‌دهد در حالتِ عمومی باید $z = 0$ باشد؛ بنابراین، نقاطِ تعادل همگی در صفحه‌ی مدار قرار می‌گیرند.

دسته‌بندی کلی

۱. نقاطِ هم‌خط[1]: سه نقطه‌ی L_1، L_2 و L_3 روی خطِ واصلِ دو جرم قرار دارند و لذا $y = 0$ است.

۲. نقاطِ سه‌گوش[2]: . دو نقطه‌ی L_4 و L_5 رأس‌های مثلثِ متساوی‌الاضلاع با دو جرمِ اصلی را تشکیل می‌دهند و در آن‌ها $y \neq 0$.



شکل ۱-۲: نقاطِ لاگرانژ در سامانه‌ی زمین–ماه

نقاطِ هم‌خط (L_1, L_2, L_3)

با اعمال $y = 0$، تنها معادله‌ی زیر باقی می‌ماند:

$$x - \frac{1-\mu}{|x+\mu|^3}(x+\mu) - \frac{\mu}{|x-1+\mu|^3}(x-1+\mu) = 0. \tag{۱-۱۳}$$

این معادله در سه ناحیه‌ی مجزا—بین دو جرم، بیرونِ جرمِ کوچک و بیرونِ جرمِ بزرگ—دارای یک ریشه است که به‌ترتیب نقاطِ L_1، L_2 و L_3 را تعیین می‌کند.

---
[1] Collinear
[2] Triangular

برای $\mu \ll 1$ (همچون سامانه‌ی خورشید–زمین یا زمین–ماه) می‌توان تقریب‌های شناخته‌شده را نوشت:

$$x_{L_1} \simeq (1-\mu) - \left(\tfrac{\mu}{3}\right)^{1/3},$$
$$x_{L_2} \simeq (1-\mu) + \left(\tfrac{\mu}{3}\right)^{1/3},$$
$$x_{L_3} \simeq -1 - \tfrac{5}{12}\,\mu; \qquad y_{L_i} = 0.$$

در عمل، ریشه‌ی دقیقِ معادله‌ی (۱-۱۳) با یک روشِ عددی (نیوتن–رافسون) محاسبه می‌شود.

نقاطِ سه‌گوش (L_4, L_5)

در این نقاط $r_1 = r_2 = 1$ و شرطِ $1-(1-\mu)/r_1^3 - \mu/r_2^3 = 0$ به‌طور طبیعی برقرار است. مختصات به عبارت‌اند از

$$x_{L_4} = x_{L_5} = \tfrac{1}{2} - \mu, \qquad y_{L_4} = +\tfrac{\sqrt{3}}{2}, \qquad y_{L_5} = -\tfrac{\sqrt{3}}{2}. \tag{۱-۱۴}$$

پایداری این نقاط مستلزم نسبتِ جرمِ کافی است؛ شرطِ کلاسیک $m_1/m_2 > 24.96$ (یا معادل آن $\mu < \mu_{\rm R} \approx 0.03852$) در سامانه‌های خورشید–سیاره یا زمین–ماه برقرار است و سببِ وجودِ خانواده‌ی سیارک‌های تروجان حول L_4 و L_5 می‌شود. در مقابل، نقاطِ هم‌خط ناپایدارند و معمولاً مأموریت‌های فضایی روی مدارهای هاله‌ای یا لیساژور در پیرامونِ آن‌ها قرار می‌گیرند.

برای سامانه‌ی زمین–ماه، $\mu \simeq 0.01215$ است. جدولِ زیر مختصاتِ بی‌بُعدِ هر پنج نقطه را نشان می‌دهد (واحدِ طول: فاصله‌ی زمین–ماه). موقعیتِ زمین در $(-\mu, 0)$ و ماه در $(1-\mu, 0)$ است.

جدول ۱-۲: مقادیر عددی نقاط لاگرانژ برای مسئله‌ی سه‌جسمیِ محدودِ سیستمِ زمین–ماه

نقطه‌ی لاگرانژ	x (بی‌بعد)	y (بی‌بعد)
L_1	$+0.83692$	0
L_2	$+1.15568$	0
L_3	-1.00506	0
L_4	$+0.48785$	$+0.86603$
L_5	$+0.48785$	-0.86603

این نتایج نشان می‌دهد که L_1 در حدودِ 0.84 فاصله‌ی زمین–ماه از زمین قرار دارد (فاصله‌ی آن تا ماه در حدود 0.16 واحد طول است) و L_2 بیرونِ مدارِ ماه است. نقطه‌ی L_3 تقریباً یک واحدِ طول در سوی مقابلِ ماه نسبت به زمین قرار دارد. دو نقطه‌ی L_4 و L_5 در مختصات $(0.488, \pm 0.866)$ قرارگرفته و با زمین و ماه مثلثِ متساوی‌الاضلاع می‌سازند.

این نتایج نشان می‌دهد که L_1 در حدودِ 0.84 فاصله‌ی زمین–ماه از زمین قرار دارد (فاصله‌ی آن تا ماه در حدود 0.16 واحد طول است) و L_2 بیرونِ مدارِ ماه است. نقطه‌ی L_3 تقریباً یک واحدِ طول در سوی مقابلِ ماه نسبت به زمین قرار دارد. دو نقطه‌ی L_4 و L_5 در مختصات $(0.488, \pm 0.866)$ قرارگرفته و با زمین و ماه مثلثِ متساوی‌الاضلاع می‌سازند.

Bibliography

[1] D. Vallado and W. McClain. *Fundamentals of Astrodynamics and Applications.* Fundamentals of Astrodynamics and Applications. Microcosm Press, 2001.

Abstract

This thesis proposes a robust guidance framework for low-thrust spacecraft operating in multi-body dynamical environments modeled by the Earth–Moon circular restricted three-body problem (CRTBP). The guidance task is cast as a zero-sum differential game between a controller agent (spacecraft) and an adversary agent (environmental disturbances), implemented under a centralized-training/ decentralized-execution paradigm. Four continuous-control reinforcement-learning algorithms—DDPG, TD3, SAC, and PPO—are extended to their multi-agent zero-sum counterparts (MA-DDPG, MATD3, MASAC, MAPPO); their actor–critic network structures and training pipelines are detailed.

The policies are trained and evaluated on transfers to the Earth–Moon lyapunov orbit under five uncertainty scenarios: random initial states, actuator perturbations, sensor noise, communication delays, and model mismatch. Zero-sum variants consistently outperform their single-agent baselines, with MATD3 delivering the best trade-off between trajectory accuracy and propellant consumption while maintaining stability in the harshest conditions.

The results demonstrate that the proposed multi-agent, game-theoretic reinforcement-learning framework enables adaptive and robust low-thrust guidance in unstable three-body regions without reliance on precise dynamics models, and is ready for hardware-in-the-loop implementation.

Keywords: Deep Reinforcement Learning; Differential Game; Multi-Agent; Low-Thrust Guidance; Three-Body Problem; Robustness.

Sharif University of Technology
Department of Aerospace Engineering

Master Thesis

Robust Reinforcement Learning Differential Game Guidance in Low-Thrust, Multi-Body Dynamical Environments

By:

Ali BaniAsad

Supervisor:

Dr.Hadi Nobahari

December 2024