


Intelligent Robust Control of Flexible Satellite in Underactuated Conditions

Jamal zarei¹, Maryam malekzadeh², Mohammad Sayanjali³

¹Aerospace Engineering, Flight Mechanics and Control, University of Isfahan, Isfahan

²Aerospace Engineering, Sharif University of Technology, Tehran

³Satellite Systems Research Institute, Iranian Space Research Center, Tehran

ARTICLE INFO

Article Type

Original Research

Article History

Received: June 02, 2025

Revised: January 24, 2026

Accepted: January 25, 2026

ePublished: April 23, 2026

ABSTRACT

This study presents a robust and intelligent control method for flexible satellites operating under underactuated conditions. When the number of actuators is fewer than the system's degrees of freedom, issues like instability and vibrations arise. To address these problems, a combination of super-twisting sliding mode control and high-order adaptive sliding mode control, along with reinforcement learning, is used. Reinforcement learning helps to adaptively adjust the control gains, improving the system's performance in the presence of disturbances and actuator failures. Quaternion parameters are utilized to avoid singularity issues when modeling the satellite's angular orientation. In this approach, the control inputs for the first and second axes are adjusted to reduce the error in the third axis without requiring direct control. Various simulations have shown that the proposed method outperforms classical approaches in reducing errors, minimizing chattering, and enhancing system stability. Furthermore, the high-order adaptive sliding mode control demonstrates greater stability against model uncertainties, although with longer settling times. These results indicate the high potential of the proposed methods for use in sensitive space missions.

Keywords: Flexible satellite, Super-Twisting sliding mode control, High-order adaptive sliding mode control, Reinforcement learning, Underactuated satellite

How to cite this article

Zarei J, malekzadeh M, Sayanjali M, Intelligent Robust Control of Flexible Satellite in Underactuated Conditions. Modares Mechanical Engineering; 2026;26(06):495-515.

*Corresponding author's email: malekzadeh@sharif.edu

*Corresponding ORCID ID: 0000-0003-1922-4488



Copyright© 2025, TMU Press. This open-access article is published under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License which permits Share (copy and redistribute the material in any medium or format) and Adapt (remix, transform, and build upon the material) under the Attribution-NonCommercial terms.



کنترل مقاوم هوشمند ماهواره انعطاف‌پذیر در حالت کمبود عملگر

جمال زارعی^۱، مریم ملک زاده^۲، محمد سینجلی^۳

^۱ مهندسی هوافضا، مکانیک پرواز و کنترل، دانشگاه اصفهان، اصفهان

^۲ مهندسی هوافضا، دانشگاه صنعتی شریف، تهران

^۳ پژوهشکده سامانه های ماهواره، پژوهشگاه فضایی ایران، تهران

اطلاعات مقاله

نوع مقاله

مقاله پژوهشی

چکیده

در این پژوهش، یک روش کنترل مقاوم و هوشمند برای ماهواره‌های انعطاف‌پذیر طراحی شده است که با کمبود عملگر مواجه هستند. وقتی تعداد عملگرها از درجات آزادی سیستم کمتر باشد، مشکلاتی مانند ناپایداری و ارتعاشات به وجود می‌آید. برای حل این مشکلات، از ترکیب دو روش کنترل مود لغزشی فوق پیچشی و مود لغزشی تطبیقی مرتبه بالا همراه با یادگیری تقویتی استفاده شده است. یادگیری تقویتی به تنظیم ضرایب کنترلی کمک کرده و عملکرد سیستم را در برابر اغتشاشات و خرابی عملگرها بهبود می‌بخشد. برای جلوگیری از مشکلات تکینگی در مدل‌سازی وضعیت زاویه‌ای ماهواره، از پارامترهای کوآرنیون استفاده شده است. در این رویکرد، ورودی‌های کنترل محورهای اول و دوم به‌گونه‌ای تنظیم شده‌اند که خطای محور سوم کاهش یابد، بدون نیاز به ورودی مستقیم. شبیه‌سازی‌های مختلف نشان داده‌اند که روش پیشنهادی عملکرد بهتری نسبت به روش‌های کلاسیک در کاهش خطا و چترینگ و بهبود پایداری سیستم دارد. همچنین، کنترل مود لغزشی تطبیقی مرتبه بالا پایداری بیشتری در برابر عدم قطعیت‌های مدل دارد، هرچند زمان نشست بیشتری می‌طلبد. این نتایج نشان‌دهنده قابلیت بالای روش‌های پیشنهادی در کاربردهای فضایی حساس است.

کلیدواژه‌ها: ماهواره انعطاف‌پذیر، کنترل مود لغزشی فوق پیچشی، کنترل مود لغزشی تطبیقی مرتبه بالا، یادگیری تقویتی، کمبود عملگر

نحوه ارجاع به این مقاله

زارعی جمال، ملک زاده مریم، سینجلی محمد، کنترل مقاوم هوشمند ماهواره انعطاف‌پذیر در حالت کمبود عملگر، مهندسی مکانیک مدرس. ۱۴۰۵:۲۶(۰۷):۴۹۵-۵۱۵

* پست الکترونیکی نویسنده عهده‌دار مکاتبات: malekzadeh@sharif.edu

* شناسه ارکید نویسنده عهده‌دار مکاتبات: 0000-0003-1922-4488



Copyright© 2025, TMU Press. This open-access article is published under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License which permits Share (copy and redistribute the material in any medium or format) and Adapt (remix, transform, and build upon the material) under the Attribution-NonCommercial terms.



شکل ۱ ماهواره انعطاف پذیر

Fig. 1 Flexible Satellite

شکل ۱ نشان دهنده ماهواره‌ی انعطاف‌پذیر است که از دو بخش اصلی تشکیل شده‌اند:

قسمت صلب (Rigid Hub) و قسمت‌های انعطاف‌پذیر (Flexible Appendages). قسمت صلب به‌عنوان بخش مرکزی، شامل تجهیزات اصلی، سیستم‌های کنترلی و مخابراتی است که با طراحی مقاوم خود می‌تواند شرایط دشوار فضا را تحمل کند. قسمت‌های انعطاف‌پذیر، از جمله آرایه‌های خورشیدی یا بازوهای باقابلیت تغییر شکل یا ارتعاش هستند. طراحی این نوع ماهواره‌ها باهدف افزایش کارایی و کاهش وزن صورت می‌گیرد. آن‌ها برای مأموریت‌هایی که نیاز به مانورهای پیچیده یا سازگاری با شرایط محیطی غیرقابل پیش‌بینی دارند، بسیار مناسب هستند. استفاده از پنل‌های خورشیدی منجر به کاهش مصرف انرژی و بهینه‌سازی عملکرد ماهواره می‌شود.

در این مقاله، ترکیب روش‌های مود لغزشی با یادگیری تقویتی معرفی شده است که می‌تواند ضرایب کنترلی را به‌صورت تطبیقی تنظیم کند. در این رویکرد، سیستم به‌صورت هوشمند با تغییرات محیط سازگار شده و عملکرد آن در برابر نامعینی‌ها و اغتشاشات بهبود می‌یابد.

۲- مروری بر کارهای انجام شده

در زمینه کنترل فضاپیماها با درجه آزادی محدود و عملگرهای ناقص، تلاش‌های متعددی انجام شده است. مرجع شماره [۲] از رویکرد کنترل پیش‌بین مبتنی بر لوله (Tube-based MPC) برای کنترل وضعیت فضاپیماها با کمبود عملگر استفاده کرده و توانسته پایداری و مقاومت در برابر اغتشاشات را تضمین کند. در مرجع [۳]، یک کنترل‌کننده مقاوم ردیابی برای فضاپیماهایی که تنها داده‌های اندازه‌گیری وضعیت را در اختیار دارند و دچار خرابی عملگر هستند، طراحی شده است. مرجع شماره [۴] نیز به تحلیل رفتار فضاپیماهای کم‌کنترل‌پذیر تحت اثر اغتشاشات محیطی پرداخته و پایداری دینامیکی آن‌ها را مورد بررسی قرار داده است. در مرجع [۵]، کنترل‌کننده برای پایداری فضاپیماهای کم‌کنترل‌پذیر با ورودی‌های محدود معرفی شده که از مدل‌سازی دقیق دینامیکی بهره می‌برد. مرجع [۶] با بهره‌گیری از تکنیک‌های

۱- مقدمه

در عصر حاضر، کنترل وضعیت و پایداری ماهواره‌ها به‌عنوان یکی از مسائل اساسی در حوزه فضا موردتوجه قرار گرفته است. ماهواره‌ها در زمینه‌هایی مانند ارتباطات، سنجش‌ازدور، ناوبری مورد استفاده قرار می‌گیرند. با این حال، چالش‌های متعددی در طراحی و بهره‌برداری از این سیستم‌ها مشاهده می‌شود که از جمله آن‌ها می‌توان به خرابی عملگرها و اغتشاشات محیطی اشاره کرد. این عوامل می‌توانند منجر به اختلال در عملکرد کلی ماهواره شوند. همچنین ماهواره‌هایی که در وضعیت کمبود عملگر (Underactuated) قرار می‌گیرند و کنترل کامل درجات آزادی آن‌ها با دشواری همراه است. برای مقابله با این چالش‌ها، استفاده از روش‌های کنترلی مقاوم و هوشمند موردتوجه قرار گرفته است تا عملکرد و پایداری ماهواره‌ها در شرایط مختلف تضمین شود. کنترل مود لغزشی (Sliding Mode Control) به‌عنوان یکی از روش‌های مقاوم، در سیستم‌های فضایی به کار گرفته می‌شود. این روش با ایجاد یک سطح لغزش، سیستم را به سمت پایداری هدایت کرده و اغتشاشات خارجی و نامعینی‌ها را سرکوب می‌کند. با این حال، پدیده چترینگ (Chattering) که در روش کنترل مود لغزشی کلاسیک مشاهده می‌شود، می‌تواند عملکرد سیستم را تحت تأثیر قرار دهد. برای غلبه بر چترینگ، از روش‌هایی مانند کنترل مود لغزشی فوق پیچشی (Super-Twisting Algorithm) استفاده می‌شود که قادر است چترینگ را کاهش داده و دقت سیستم را افزایش دهد. علاوه بر این، کنترل مود لغزشی تطبیقی مرتبه بالا (Adaptive High-Order Sliding Mode Control) به‌عنوان یک روش پیشرفته‌تر معرفی شده است که با تعریف سطوح لغزش متعدد، عملکرد سیستم‌های پیچیده؛ مانند ماهواره‌های انعطاف‌پذیر را بهبود می‌بخشد. این روش، امکان مقابله با اغتشاشات و خرابی عملگرها را فراهم کرده و پایداری سیستم را حتی در شرایط بحرانی تضمین می‌کند.

در این مقاله، کنترل مود لغزشی فوق پیچشی و کنترل مود لغزشی تطبیقی مرتبه بالا به صورت جداگانه با یادگیری تقویتی (Reinforcement Learning) ترکیب شده‌اند. این ترکیب‌ها به صورت مستقل بررسی شده تا عملکرد هر روش در شرایط مختلف تحلیل شود و بهترین راهکار برای کنترل وضعیت ماهواره انعطاف‌پذیر در حالت کمبود عملگر ارائه گردد. [۱]

روش‌های پیشرفته‌تر مانند کنترل مود لغزشی فوق پیچشی یا تطبیقی استفاده شده است همچنین، در اکثر این پژوهش‌ها بهره‌گیری از الگوریتم‌های یادگیری تقویتی در طراحی کنترلر لحاظ نشده است.

در این مقاله، با ترکیب ساختارهای کنترلی مقاوم شامل کنترل مود لغزشی فوق پیچشی و کنترل مود لغزشی تطبیقی مرتبه بالا با الگوریتم‌های یادگیری تقویتی، یک چارچوب کنترلی هوشمند و تطبیقی ارائه شده که قادر است در برابر ازکارافتادن و کمبود عملگرها، اغتشاشات غیرمنتظره محیطی و عدم قطعیت‌های مدل، عملکرد بهینه‌ای از خود نشان دهد. این کنترلرها نه تنها قادر به کاهش قابل‌توجه چترینگ هستند، بلکه از طریق تنظیم خودکار ضرایب کنترلی به کمک یادگیری تقویتی، تطابق سیستم با شرایط متغیر محیطی را ممکن ساخته‌اند. نتایج شبیه‌سازی‌های انجام‌شده در سناریوهای مختلف نشان داده است که این ساختار پیشنهادی، برتری قابل‌ملاحظه‌ای نسبت به روش‌های کنترلی متداول داشته و می‌تواند به‌عنوان یک راهکار عملیاتی و قابل‌انکاب برای کنترل ماهواره‌های انعطاف‌پذیر در مأموریت‌های حساس فضایی مورداستفاده قرار گیرد.

۳- مدل دینامیکی ماهواره‌های انعطاف‌پذیر

در بررسی دینامیک ماهواره‌های انعطاف‌پذیر، بدنه اصلی ماهواره به‌عنوان یک بخش صلب مدل‌سازی می‌شود، درحالی‌که اجزای انعطاف‌پذیر آن نظیر پنل‌های خورشیدی با استفاده از روش‌های مودال تحلیل می‌گردند. رفتار دینامیکی کلی این سامانه از طریق فرمول‌بندی مبتنی بر مکانیک لاگرانژی به دست می‌آید که تعامل بین اجزای صلب و انعطاف‌پذیر را به طور دقیق توصیف می‌کند. حرکت زاویه‌ای بدنه صلب با استفاده از بردار سرعت زاویه‌ای ω مدل شده و ارتعاشات اجزای انعطاف‌پذیر توسط بردار جابه‌جایی مودهای انعطاف η توصیف می‌شود. معادلات حاکم بر دینامیک

این سیستم به‌صورت زیر هستند: [۱۸]

معادله حرکت چرخشی:

$$\dot{\omega} + \mathbf{d}^T \dot{\eta} = -\omega \times (J \omega + \mathbf{d}^T \eta) + \mathbf{u} \quad (1)$$

معادله حرکت ارتعاشی پنل‌ها:

$$\ddot{\eta} + C\dot{\eta} + K\eta = -\delta\dot{\omega} \quad (2)$$

در این معادله، $J^T \in \mathbb{R}^{3 \times 3}$ به عنوان ماتریس ممان اینرسی بدنه صلب تعریف شده است. بردار $\omega \in \mathbb{R}^3$ بیانگر سرعت زاویه‌ای بدنه و $\mathbf{u} \in \mathbb{R}^3$ نشان‌دهنده گشتاورهای کنترلی اعمال‌شده بر سیستم است.

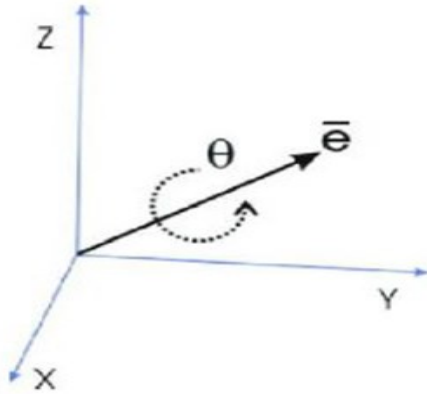
ماتریس $\delta \in \mathbb{R}^{N \times 3}$ کوپلینگ بین بدنه صلب و مودهای انعطاف‌پذیر را مدل می‌کند و بردار $\eta \in \mathbb{R}^N$ نیز مختصات مودال

غیرخطی، یک کنترل‌کننده ردیابی برای فضایی‌های کم‌کنترل‌پذیر توسعه داده و نشان داده که عملکرد مناسبی در پیروی از مسیر دارد. در مقاله [۷]، کنترل غیرخطی برای شناوری پایدار فضایی در مدار طراحی شده که پایداری وضعیت فضایی را در حضور ورودی‌های محدود تضمین می‌کند. مرجع [۸] از کنترل پیش‌بین مدل‌محور برای فضایی‌هایی با تنها دو چرخ عکس‌العملی استفاده کرده و نشان داده است که این روش قابلیت پایداری و کنترل‌پذیری خوبی دارد. در مرجع [۹]، راه‌حل‌های تحلیلی برای بهینه‌سازی پیکربندی مجدد تشکیل‌دهی فضایی‌های کم‌کنترل‌پذیر ارائه شده که به کاهش مصرف سوخت کمک می‌کند. مطالعه [۱۰]، کنترل مود لغزشی پایانه‌ای سریع را برای بازپیکربندی پیکربندی‌های فضایی کم‌کنترل‌پذیر پیشنهاد داده که باعث کاهش زمان پاسخ می‌شود. در مرجع [۱۱]، تحلیل دینامیک و کنترل در پیکربندی مجدد فضایی‌های کم‌کنترل‌پذیر در مدارهای بیضوی صورت گرفته که عملکرد سیستم در شرایط واقعی را شبیه‌سازی کرده است. در مرجع [۱۲]، کنترل مقاوم برای نگه‌داشت موقعیت و بازپیکربندی دسته‌جمعی فضایی‌های کم‌کنترل‌پذیر طراحی شده و بر پاسخ‌دهی سریع در برابر اغتشاشات تأکید دارد. مرجع [۱۳] یک کنترل تطبیقی بدون برخورد برای حرکت دسته‌جمعی فضایی‌های با درجه آزادی محدود ارائه داده که مانع از تصادف در مدارهای متراکم می‌شود. در مرجع [۱۴]، کنترل همزمان ملاقات (rendezvous) برای فضایی‌های کم‌کنترل‌پذیر توسعه یافته که نشان داده این روش در بازه‌های زمانی کوتاه قابل اجرا است. در نهایت، مرجع [۱۵]، کنترل تطبیقی توزیع‌شده برای وضعیت فضایی‌های شبکه‌ای انعطاف‌پذیر و کم‌کنترل‌پذیر معرفی کرده که از روش‌های مدرن کنترل دسته‌جمعی بهره می‌برد.

در همین زمینه، در پژوهش‌های صورت‌گرفته در نشریه مهندسی مکانیک مدرس نیز، تلاش‌هایی برای مدل‌سازی دقیق دینامیکی و کنترل ارتعاشات ماهواره‌های انعطاف‌پذیر ارائه شده است. در [۱۶]، مدل دینامیکی یک ماهواره با پنل‌های خورشیدی انعطاف‌پذیر ارائه شده و کنترل فعال ارتعاشات آن با استفاده از روش‌های تطبیقی بررسی شده است. همچنین در [۱۷]، کنترل مقاوم وضعیت برای یک ماهواره انعطاف‌پذیر با در نظر گرفتن پویایی عملگرها طراحی شده که عملکرد مناسبی در مقابله با اغتشاشات نشان داده است. این مطالعات به‌خوبی نشان‌دهنده ضرورت توسعه روش‌های کنترلی هوشمند در مواجهه با رفتار غیرخطی و انعطاف‌پذیر سامانه‌های فضایی هستند.

یکی از چالش‌های اساسی در کنترل وضعیت ماهواره‌های انعطاف‌پذیر، به‌ویژه در شرایط کمبود عملگر، عدم پاسخ‌گویی مناسب روش‌های کنترلی کلاسیک در برابر اغتشاشات محیطی، نامعینی‌های دینامیکی و خرابی عملگرها است. در بسیاری از پژوهش‌های پیشین، یا فرض بر وجود عملگرهای کامل بوده و یا به خرابی عملگرها پرداخته شده است. در موارد معدودی که از

بدون مشکلات نقاط تکین که در روش‌های دیگر مانند زوایای اویلر رخ می‌دهد، فراهم می‌کنند. با استفاده از کوآترنیون‌ها، حرکت زاویه‌ای سه‌بعدی به مسیری در فضای چهاربعدی نگاشت می‌شود. این نگاشت به‌گونه‌ای است که هر چرخش سه‌بعدی در فضای واقعی، معادلی در فضای چهاربعدی کوآترنیون دارد [۱۹].



شکل ۲ چرخش کوآترنیونها

Fig. 2 Quaternion Rotation

شکل ۲ نحوه نمایش چرخش‌های فضایی با استفاده از کوآترنیون‌ها را نشان می‌دهد. کوآترنیون‌ها با ارائه نمایش بدون تکینگی، جایگزینی مناسب برای زوایای اویلر در تحلیل‌های دینامیکی و سینماتیکی هستند.

رابطه کوآترنیون‌ها با محور و زوایای اویلر به شکل زیر است:

$$q^T = \text{esin} \frac{\theta}{2} \quad (6)$$

$$q_0 = |\text{e}| \cos \frac{\theta}{2} \quad (7)$$

در اینجا، e به‌عنوان بردار اویلر تعریف می‌شود. بردار ویژه مرتبط با این بردار، مبنای محاسبه روابط سینماتیکی پارامترهای کوآترنیون است که به‌صورت زیر ارائه می‌شود:

$$\dot{q} = \frac{1}{2} (q_0 I + q^x) \omega \quad (8)$$

$$\dot{q}_0 = -\frac{1}{2} q^T \omega \quad (9)$$

در این رابطه، J یک ماتریس همانی با ابعاد 3×3 است، ω سرعت زاویه‌ای ماهواره را نشان می‌دهد و q^x به‌عنوان ماتریس پادمتقارن بردار q تعریف می‌شود که به صورت زیر بیان می‌گردد:

$$q^x = \begin{bmatrix} 0 & -q_3 & q_2 \\ q_3 & 0 & -q_1 \\ -q_2 & q_1 & 0 \end{bmatrix} \quad (10)$$

اجزای انعطاف‌پذیر را شامل می‌شود. علاوه بر این، C به عنوان ماتریس دمپینگ مودال و K نیز ماتریس سختی مودال سیستم را تشکیل می‌دهند.

$$C = \text{diag}\{[2\xi_1 A_1 \quad \dots \quad 2\xi_N A_N]^T\} \quad (3)$$

$$K = \text{diag}\{[A_1^2 \quad \dots \quad A_N^2]^T\} \quad (4)$$

مدل فوق نشان‌دهنده تعامل میان دینامیک حرکت صلب و رفتار انعطاف‌پذیر اجزا است. سرعت زاویه‌ای بدنه موجب تحریک ارتعاشات شده و برعکس، جابه‌جایی پنل‌های انعطاف‌پذیر بر رفتار دینامیکی کل ماهواره تأثیرگذار هستند.

در مدل ساده‌شده، بدون در نظر گرفتن دمپینگ ساختاری، معادله دینامیکی کلی به‌صورت زیر است:

$$M\ddot{\eta} + K\eta = F^T \quad (5)$$

در این معادله، M به‌عنوان ماتریس جرم کلی سیستم در نظر گرفته شده است که ترکیبی از جرم بدنه صلب و جرم اجزای انعطاف‌پذیر را شامل می‌شود. ماتریس K نمایانگر سختی سیستم بوده و F به‌عنوان ماتریس نیروی ورودی ایفای نقش می‌کند. همچنین، بردار η شامل مجموعه‌ای از متغیرهای سیستم است که هم متغیرهای مرتبط با بخش صلب و هم متغیرهای مربوط به اجزای انعطاف‌پذیر را دربر می‌گیرد.

۴- سینماتیک ماهواره انعطاف‌پذیر

برای توصیف موقعیت زاویه‌ای ماهواره نسبت به چارچوب اینرسی، از پارامترهای دوران مختلفی استفاده می‌شود. روش‌های متنوعی برای نمایش این موقعیت زاویه‌ای وجود دارد که شامل زوایای اویلر، پارامترهای رودریگز و کوآترنیون‌ها است. با این حال، در وضعیت زاویه‌ای ماهواره، استفاده از زوایای اویلر با محدودیت‌هایی همراه است، چرا که این روش‌ها در برخی زوایای خاص مانند ۹۰ درجه دچار پدیده‌ای موسوم به تکینگی (Singularity) می‌شوند. این نقاط تکین باعث می‌شوند که محاسبات مربوط به موقعیت و تغییر وضعیت زاویه‌ای دچار اختلال شده و در نتیجه، عملکرد و پایداری سیستم کاهش یابد. در مقابل، پارامترهای کوآترنیون که مبتنی بر فضای چهاربعدی هستند، از این مشکل مصون بوده و حتی در زوایای بزرگ نیز دقت و پیوستگی لازم را حفظ می‌کنند. به همین دلیل، در این پژوهش برای توصیف موقعیت زاویه‌ای ماهواره، از پارامترهای کوآترنیون استفاده شده تا از بروز خطاهای ناشی از تکینگی جلوگیری شود و پایداری مدل سینماتیکی تضمین گردد.

پارامترهای کوآترنیون، مجموعه‌ای از بردارهای چهار مؤلفه‌ای هستند که به‌طور مؤثر برای نمایش دوران در فضای سه‌بعدی استفاده می‌شوند. این پارامترها امکان توصیف وضعیت زاویه‌ای را

۵- طراحی کنترل کننده

در این مقاله، برای کنترل ماهواره انعطاف پذیر از دو روش اصلی استفاده شده است: کنترل مود لغزشی فوق پیچشی و کنترل مود لغزشی تطبیقی مرتبه بالا ابتدا به طراحی کنترل مود لغزشی فوق پیچشی پرداخته می شود که به صورت زیر تعریف می شود:

$$u_1 = G^{-1}(\ddot{X}_d - \Lambda \dot{e} - K_1 \sqrt{|s|} \cdot \text{sign}(s) - 0.5K_2 \int s dt - f) \quad (12)$$

در این معادله، u_1 ورودی کنترلی است که باید به سیستم اعمال شود. بخش اول معادله، G^{-1} ، نمایانگر معکوس یک ماتریس است که معمولاً به ویژگی های مکانیکی سیستم مربوط می شود. این ماتریس به طور معمول برای تبدیل کنترل از فضای فیزیکی به فضای سیستم های کنترل استفاده می شود.

\ddot{X}_d شتاب مرجع است که سیستم باید آن را دنبال کند و به وضعیت مطلوب برسد $\Lambda \dot{e}$ بخش دیگری از کنترل است که مربوط به تنظیم خطا و سرعت تغییرات خطا می شود Λ . معمولاً یک ماتریس تنظیمی است که حساسیت سیستم نسبت به سرعت تغییرات خطا را تنظیم می کند. بخش $K_1 \sqrt{|s|} \cdot \text{sign}(s)$ به کنترل مود لغزشی مربوط است. این بخش برای کاهش چترینگ و بهبود پاسخ سیستم به تغییرات ناگهانی طراحی شده است. در اینجا، K_1 ضریب تنظیمی است که شدت این کنترل را تعیین می کند. $\sqrt{|s|}$ مربوط به سطح لغزش s است و $\text{sign}(s)$ برای تعیین جهت کنترل استفاده می شود. بخش $0.5K_2 \int s dt$ مربوط به یکپارچگی سطح لغزش است که اثرات تاریخی خطا را در نظر می گیرد K_2 . ضریب تنظیمی است که شدت اثر این یکپارچگی را تعیین می کند. در نهایت، f نمایانگر اغتشاشات یا نیروهای خارجی است که به سیستم وارد می شود. این اغتشاشات می توانند از محیط یا تغییرات خارجی ناشی شوند و در سیستم های پیچیده مانند ماهواره ها به طور مداوم وجود دارند.

$$f = G^{-1}(-S_w/\omega + \tau_d) \quad (13)$$

در این معادله، f نمایانگر نیرو یا اغتشاشات خارجی است که به سیستم وارد می شود. این اغتشاشات می توانند از تغییرات محیطی، خرابی ها یا نیروهای خارجی مانند جاذبه یا مقاومت هوا ناشی شوند. G^{-1} معکوس ماتریس G که معمولاً به ویژگی های مکانیکی یا دینامیکی سیستم مربوط می شود، مانند ممان اینرسی یا ماتریس جرم. این معکوس معمولاً برای تبدیل نیروی وارد بر سیستم به وضعیت فیزیکی آن به کار می رود. S_w این بخش نمایانگر سطح لغزش است که به طور معمول در کنترل مود لغزشی برای هدایت سیستم به سمت پایدار استفاده می شود. این پارامتر معمولاً با توجه به خطای سیستم و رفتار دینامیکی آن انتخاب می شود. ماتریس ممان اینرسی است که رفتار دینامیکی سیستم را توصیف می کند. این پارامتر برای شبیه سازی رفتار سیستم های چرخشی، مانند ماهواره ها، ضروری است. که در آن:

۵-۱- کنترل مود لغزشی فوق پیچشی

سطح لغزش S به عنوان یکی از اجزای کلیدی در طراحی کنترل کننده مود لغزشی، به گونه ای تعریف می شود که اختلاف میان سرعت زاویه ای واقعی و سرعت زاویه ای مرجع (مطلوب) را اصلاح کرده و سیستم را به سمت مسیر مطلوب هدایت کند. این سطح لغزش به صورت زیر بیان می شود:

$$S = \dot{e} + \Lambda e \quad (11)$$

در این معادله، S نمایانگر سطح لغزش است که هدف در کنترل مود لغزشی رسیدن سیستم به این سطح و ماندن در آن است. سطح لغزش به عنوان معیاری برای پایداری سیستم و هدایت آن به سمت وضعیت مطلوب در نظر گرفته می شود. در واقع، زمانی که سیستم به سطح لغزش وارد می شود، خطا و سرعت خطا به صفر می رسند و سیستم به حالت پایدار می رسد. e در اینجا نمایانگر خطای سیستم است. این خطا معمولاً تفاوت میان وضعیت فعلی سیستم (مثل زاویه یا موقعیت ماهواره) و وضعیت مطلوب (مرجع) است. هدف از کنترل مود لغزشی این است که این خطا را به حداقل برساند تا سیستم در حالت مطلوب قرار گیرد. به همین دلیل، به طور مداوم محاسبه می شود و باید به سمت صفر هدایت شود. \dot{e} نیز نمایانگر مشتق خطا است که سرعت تغییرات خطای سیستم را نشان می دهد. در واقع، این بخش به کمک سرعت تغییرات خطا به بهبود زمان همگرایی سیستم کمک می کند. به عبارت دیگر، مشتق خطا نشان می دهد که سیستم با چه سرعتی به سمت وضعیت پایدار حرکت می کند. Λ یک ماتریس است که برای تنظیم حساسیت سیستم نسبت به خطا و سرعت خطا استفاده می شود. این ماتریس معمولاً بر اساس دینامیک سیستم انتخاب می شود و به کنترل کننده کمک می کند تا سیستم به سرعت به سطح لغزش برسد و از آن عبور نکند. در این معادله، ترکیب خطا و سرعت خطا تحت تاثیر ماتریس Λ قرار می گیرد و هدف این است که مقدار سطح لغزش S به صفر برسد تا سیستم به سمت وضعیت پایدار حرکت کند.

پس از تعریف سطح لغزش، لازم است قانون کنترلی استخراج شود که سیستم را به سمت سطح لغزش هدایت کرده و در آن نگه دارد. قانون کنترلی پیشنهادی از دو بخش تشکیل می شود:

تأثیر ورودی کنترلی را تنظیم می‌کند. این ضریب به‌طور معمول برای تنظیم حساسیت کنترل نسبت به سطح لغزش به کار می‌رود. $sign(S)$ این بخش نمایانگر علامت سطح لغزش S است و تابعی است که به مقدار سطح لغزش S برچسب مثبت یا منفی می‌دهد. هدف از استفاده از این تابع تعیین جهت کنترل است. زمانی که سیستم از سطح لغزش عبور می‌کند، این ورودی کنترلی به‌طور خودکار جهت حرکت سیستم را تنظیم می‌کند.

۵-۲- اصلاح خطای محور سوم با استفاده از محورهای ۱ و ۲

ماهواره‌های انعطاف‌پذیر که دارای سه محور اصلی (۱، ۲ و ۳) هستند، معمولاً محور سوم ممکن است به دلیل کمبود عملگر و یا به دلایل دیگری از جمله اختلالات و نوسانات خارجی از مدار خارج شود. در این حالت، کنترل دقیق محور سوم (که باید صفر باقی بماند) نیاز به توجه ویژه دارد. با استفاده از کنترلی که طراحی شده است برای کنترل محور سوم استفاده از اطلاعات محورهای ۱ و ۲ نیاز است. اینکار از طریق اصلاح ورودی‌های کنترلی برای محورهای ۱ و ۲ انجام می‌شود تا اثرات مربوط به محور سوم را کاهش داده و خطای آن را کنترل کند.

برای این کار، یک سیگنال اصلاحی برای ورودی‌های محورهای ۱ و ۲ به شرح زیر تعریف می‌شود [۲۰]:

$$u_1(1) = u_1(1) - k_3(S_3 + 0.3 \cdot e_3) \cdot 0.05 \quad (18)$$

$$u_1(2) = u_1(2) - k_3(S_3 + 0.3 \cdot e_3) \cdot 0.05 \quad (19)$$

در این معادله، S_3 به عنوان سطح لغزش مربوط به محور سوم در نظر گرفته شده است. پارامتر e_3 بیانگر خطای وضعیت در همین محور می‌باشد و ضریب k_3 به عنوان یک پارامتر تنظیمی برای کاهش این خطا تعریف شده است. انتخاب مقدار مناسب برای k_3 از اهمیت ویژه‌ای برخوردار است، به‌گونه‌ای که ضمن جلوگیری از تحریک بیش از حد سیستم، موجب کاهش خطای محور سوم به میزان مطلوب گردد.

در اینجا، ورودی‌های کنترلی محورهای ۱ و ۲ با استفاده از این اصلاحات به‌طور غیرمستقیم روی محور سوم تأثیر می‌گذارند. هدف از این اصلاحات این است که محور سوم به مقدار صفر برسد یا در شرایط کمبود عملگر، آن را به حداقل خطا برساند. این روش باعث می‌شود که حتی بدون نیاز به کنترل مستقیم محور سوم، آن را در محدوده مناسب نگه دارد و پایداری سیستم را حفظ کند.

ورودی کلی u از ترکیب ورودی‌های محوری ۱ و ۲ و محور سوم به دست می‌آید. ورودی محور سوم که باید صفر باقی بماند، به صورت دستی تنظیم می‌شود. در نهایت، ورودی کلی به صورت زیر محاسبه می‌شود:

$$u = \begin{bmatrix} u_1(1) \\ u_1(2) \\ 0 \end{bmatrix} \quad (20)$$

S_w ماتریس سرعت زاویه‌ای است که به صورت زیر تعریف می‌شود:

$$S_w = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix} \quad (14)$$

τ_d گشتاور اغتشاشی است که به صورت زیر محاسبه می‌شود:

$$N = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (15)$$

$$\tau_d = -A\eta - D\dot{\eta} + 0.01 \sin(2\pi \cdot 0.5 \cdot t)N \quad (16)$$

عبارت τ_d بیانگر گشتاور اغتشاشی وارد بر سیستم است که با هدف ارزیابی پایداری و مقاومت کنترل‌کننده در برابر اغتشاشات مدل‌سازی شده است. این گشتاور از سه مؤلفه اصلی تشکیل شده است. عبارت $A\eta$ نشان‌دهنده تأثیر سختی مجازی سیستم بر پاسخ دینامیکی است. در اینجا، η به‌عنوان مختصات مودال ارتعاشی سیستم تعریف می‌شود که میزان انحراف ارتعاشی از حالت تعادل را توصیف می‌کند و پارامتر A یک ضریب سختی تنظیمی است که میزان مقاومت سیستم در برابر تغییر وضعیت را مدل می‌نماید. این بخش تمایل سیستم را برای بازگشت به موقعیت تعادلی افزایش می‌دهد. عبارت $D\dot{\eta}$ نمایانگر اثر میرایی مجازی در سیستم است. در این بخش، $\dot{\eta}$ مشتق زمانی η یا نرخ تغییر ارتعاش مودال می‌باشد و D یک ضریب میرایی تنظیمی است که به‌منظور کاهش نوسانات و جلوگیری از واگرایی طراحی شده است. این بخش نقش کلیدی در ایجاد همگرایی نرم و پایدار در سیستم دارد.

ترم آخر، یعنی $0.01 \sin(2\pi \cdot 0.5 \cdot t)$ ، یک اغتشاش خارجی با طبیعت سینوسی است که با فرکانس 0.5 هرتز به سیستم اعمال می‌شود. این مؤلفه برای شبیه‌سازی شرایط غیرایده‌آل محیطی همچون نوسانات خارجی یا اغتشاشات متناوب استفاده شده و به‌عنوان یک تست پایداری مهم در طراحی کنترلر مقاوم به شمار می‌آید.

به‌طور کلی، این معادله نمایانگر مدلی ترکیبی از رفتار درونی سیستم (شامل سختی و میرایی) و اغتشاش محیطی خارجی است که در شبیه‌سازی برای بررسی پاسخ کنترل‌کننده تحت شرایط واقعی‌تر استفاده می‌شود.

بخش کنترل مقاوم به‌منظور جبران اغتشاشات خارجی و نامعینی‌های سیستم طراحی شده است و به‌صورت زیر بیان می‌شود:

$$u_2 = -\lambda_5 \text{sign}(S) \quad (17)$$

در این معادله، u_2 نمایانگر ورودی کنترلی است که در سیستم اعمال می‌شود. این ورودی کنترلی از روش کنترل مود لغزشی ناشی می‌شود. λ_5 این پارامتر یک ضریب تنظیمی است که شدت و میزان

- τ_d گشتاور اغتشاشی ناشی از نامعینی‌ها یا عوامل محیطی است.

ورودی کنترل به‌گونه‌ای طراحی می‌شود که مشتق تابع لیاپانوف منفی شود. قانون کنترل پیشنهادی به شکل زیر است:

$$u = G^{-1} \left(\ddot{X}_d - \Lambda \dot{e} - K_1 \frac{S}{\sqrt{|S| + \epsilon}} - K_2 \int S dt - F \right) \quad (27)$$

که در آن:

- K_1, K_2 ماتریس‌های قطری و مثبت‌معین هستند.
- ϵ عدد مثبت کوچکی است برای جلوگیری از تقسیم بر صفر.
- \dot{X}_d شتاب مطلوب است.

با این طراحی، مشتق تابع لیاپانوف به شکل زیر درمی‌آید:

$$\dot{V}(S) = S^T \dot{S} < 0 \quad (28)$$

بنابراین، تابع لیاپانوف در طول زمان کاهش می‌یابد و پایداری سیستم طبق معیار لیاپانوف تضمین می‌شود.

۷- ساختار یادگیری تقویتی برای کنترل مود لغزشی [۲۱]

عامل یادگیرنده (Agent) بر اساس یک شبکه عصبی طراحی شده که وظیفه تنظیم پارامترهای کنترلی K_1 و K_2 و λ را بر عهده دارد. این عامل با استفاده از روش یادگیری تقویتی عمیق، ضرایب را بهینه می‌کند تا عملکرد کنترلر بهبود یابد.

فضای حالت شامل پارامترهای زاویه‌ای و سرعت زاویه‌ای ماهواره است که به صورت زیر تعریف می‌شود:

$$N = [\phi, \theta, \psi, \omega_x, \omega_y, \omega_z, \eta_x, \eta_y, \eta_z]^T \quad (29)$$

در این رابطه، ϕ, θ, ψ به ترتیب به‌عنوان زاویه‌های اوپلر مورد استفاده قرار گرفته‌اند که وضعیت زاویه‌ای ماهواره را توصیف می‌کنند. مؤلفه‌های $\omega_x, \omega_y, \omega_z$ بیانگر سرعت‌های زاویه‌ای در راستای محورهای اصلی سیستم هستند. همچنین، پارامترهای η_x, η_y, η_z ویژگی‌های انعطاف‌پذیری ماهواره را نمایش می‌دهند که نقش آن‌ها در تحلیل دینامیکی اجزای انعطاف‌پذیر حائز اهمیت است.

۷-۱ عمل (Action)

عمل کنترلی که توسط عامل یادگیرنده اعمال می‌شود، شامل تنظیم ضرایب K_1 و K_2 و λ است. این ضرایب به صورت پیوسته تغییر می‌کنند تا بهترین عملکرد حاصل شود.

۷-۲- پاداش (Reward)

محور سوم به صورت دستی صفر می‌شود، زیرا هدف این است که این محور هیچ‌گونه ورودی نداشته باشد.

۶- اثبات پایداری کنترلر با استفاده از تابع لیاپانوف

برای تحلیل پایداری سیستم، از تابع لیاپانوف زیر استفاده می‌شود:

$$V(S) = \frac{1}{2} S^T S \quad (21)$$

که در آن S سیگنال لغزش است و به صورت زیر تعریف می‌شود:

$$S = \dot{e} + \Lambda e \quad (22)$$

که در آن، e خطای زاویه‌ای است که اختلاف بین وضعیت زاویه‌ای واقعی ماهواره و وضعیت مطلوب را نشان می‌دهد. همچنین، $\dot{e} = \omega - \omega_d$ بیانگر نرخ تغییر خطای زاویه‌ای است که به صورت اختلاف بین بردار سرعت زاویه‌ای واقعی ω و سرعت زاویه‌ای مطلوب ω_d محاسبه می‌شود. ماتریس Λ به‌عنوان یک ماتریس وزنی مثبت تعریف شده است که برای تنظیم پاسخ دینامیکی سیستم و تضمین پایداری آن مورد استفاده قرار می‌گیرد. این ماتریس در تعریف سطح لغزش یا در قانون کنترل ظاهر شده و نقش مهمی در تعیین سرعت همگرایی و کاهش خطای وضعیت ایفا می‌کند. انتخاب مناسب مقادیر Λ می‌تواند عملکرد کنترل‌کننده را در برابر اغتشاشات و نامعینی‌ها بهبود بخشد.

۶-۱ محاسبه مشتق تابع لیاپانوف

مشتق تابع لیاپانوف نسبت به زمان به صورت زیر محاسبه می‌شود:

$$\dot{V}(S) = \frac{d}{dt} \left(\frac{1}{2} S^T S \right) = S^T \dot{S} \quad (23)$$

$\dot{S} = \dot{e} + \Lambda e$ مشتق S به صورت زیر محاسبه می‌شود:

$$\dot{S} = \ddot{e} + \Lambda \dot{e} \quad (24)$$

در سیستم، \ddot{e} تابعی از ورودی کنترلی و اغتشاشات دینامیکی است. با استفاده از معادلات حرکت جسم صلب، داریم:

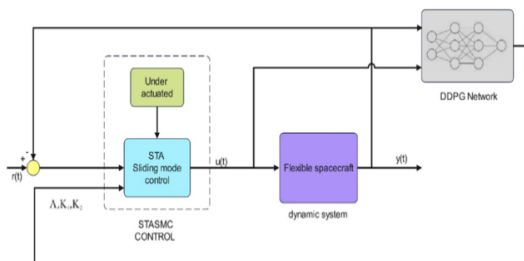
$$\ddot{e} = F + \tau_d \quad (25)$$

$$F = J^{-1}(-S(\omega)J\omega + u) \quad (26)$$

که در آن:

- $S(\omega)$ ماتریس ضرب خارجی بردار ω است (برای مدل سازی گشتاور ژيروسکوپ)

ارزش یک وضعیت-اقدام استفاده می‌شود. در اینجا، $Q(N, a)$ نمایانگر ارزش اقدام a در وضعیت N است که نشان‌دهنده مقدار پاداش است که انتظار می‌رود از انجام این اقدام در وضعیت مورد نظر به دست آید. مقدار β نرخ یادگیری است که به طور مستقیم تعیین می‌کند که چه میزان باید از تفاوت میان ارزش پیش‌بینی شده و ارزش واقعی برای به‌روزرسانی تابع استفاده شود. این پارامتر سرعت یادگیری را کنترل می‌کند و معمولاً بین ۰ و ۱ قرار دارد. در این معادله، R نمایانگر پاداش فوری است که عامل پس از انجام یک اقدام در وضعیت خاص دریافت می‌کند. فاکتور تخفیف است که اهمیت پاداش‌های آینده را مشخص می‌کند؛ این پارامتر معمولاً مقداری بین ۰ و ۱ دارد، به طوری که هرچه مقدار آن بیشتر باشد، آینده برای عامل مهم‌تر خواهد بود. بخش $\max Q(N', a')$ بیشترین ارزش تابع Q در وضعیت N' را برای بهترین اقدام a' در نظر می‌گیرد و این نشان می‌دهد که عامل از بیشترین پاداش ممکن که در آینده می‌تواند دریافت کند، آگاه است. با به‌کارگیری این معادله، عامل یادگیرنده قادر است ارزش هر وضعیت-اقدام را به‌طور تدریجی و با توجه به تجربیات خود به‌روزرسانی کند تا بهترین تصمیم را در وضعیت‌های مختلف اتخاذ کند و بیشترین پاداش را در طول زمان بدست آورد. این الگوریتم به سیستم این امکان را می‌دهد که با تغییرات محیطی و خرابی عملگرها به‌طور خودکار سازگار شده و ضرایب کنترلی بهینه‌تری پیدا کند.



شکل ۳ بلوک دیاگرام ماهواره انعطاف‌پذیر با کمبود عملگر و ترکیب کنترلر مود لغزشی با یادگیری تقویتی

Fig. 3 Block Diagram of a Flexible Satellite with Actuator Failure and the Combination of Sliding Mode Controller with Reinforcement Learning.

شکل ۳ نشان می‌دهد که بلوک دیاگرام ترکیبی از کنترلر مود لغزشی فوق پیچشی و یادگیری تقویتی عمیق برای کنترل یک ماهواره انعطاف‌پذیر را نمایش می‌دهد. در این ساختار ورودی مرجع سیستم، رفتار مطلوب را مشخص می‌کند. کنترلر مود لغزشی فوق پیچشی با استفاده از پارامترهایی نظیر Λ ، K_1 و K_2 تنظیم می‌شود تا پایداری سیستم را بهبود بخشد و اثر اغتشاشات را کاهش دهد. مازول کمبود عملگر برای مدیریت چالش‌های ناشی از کمبود عملگرها طراحی شده است، در حالی که دینامیک ماهواره

تابع پاداش برای هدایت عامل یادگیرنده به سمت بهبود عملکرد سیستم طراحی شده است. این تابع به‌گونه‌ای تنظیم شده که با کاهش خطاهای زاویه‌ای و سرعت زاویه‌ای و همچنین محدودکردن گشتاورهای کنترلی، پاداش بیشتری اختصاص داده شود. تابع پاداش به‌صورت زیر تعریف می‌شود:

$$R = -k_\phi \cdot \phi^2 - k_\omega \cdot \omega^2 - k_\tau \cdot \tau^2 \quad (30)$$

که در آن:

- k_ϕ ، k_ω و k_τ ضرایب تنظیمی هستند که شدت تاثیر هر یک از پارامترهای خطا، سرعت زاویه‌ای و گشتاور را کنترل می‌کنند.
- ϕ نمایانگر خطای زاویه‌ای ماهواره است (تفاوت بین وضعیت زاویه‌ای واقعی و وضعیت مطلوب).
- ω نمایانگر سرعت زاویه‌ای ماهواره است.
- τ نمایانگر گشتاور اعمال شده به سیستم است.

از نظر پیوستگی این نوع از توابع، به‌طور پیوسته و بدون گسستگی عمل می‌کنند، زیرا مقدار پاداش برای هر وضعیت سیستم به‌طور هموار تغییر می‌کند و هیچگونه پرش یا گسستگی در تغییرات آن وجود ندارد.

از لحاظ تشویق به کمینه‌سازی استفاده از توان دوم (مثل ϕ^2 به) این معنی است که هرچه خطا، سرعت زاویه‌ای یا گشتاور بیشتر باشد، پاداش منفی بیشتری دریافت می‌شود و این باعث می‌شود که سیستم به سمت کاهش این مقادیر حرکت کند و سازگاری با DRL این نوع تابع پاداش باعث می‌شود که عامل یادگیرنده بتواند به‌طور کارآمدتر و با ثبات بیشتری یاد بگیرد، زیرا مقادیر پاداش به‌طور مداوم و بدون نوسانات شدید تغییر می‌کنند.

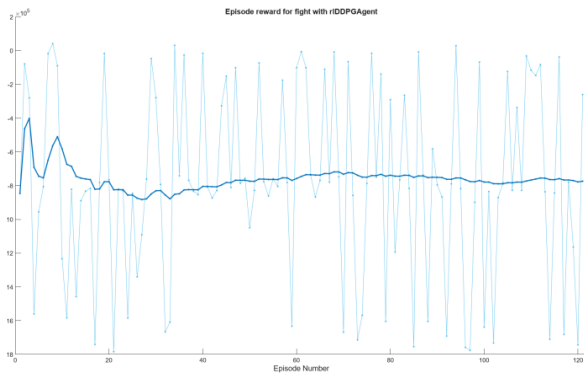
در نتیجه این تابع به‌گونه‌ای طراحی شده که در طول فرآیند یادگیری، عامل یادگیرنده به‌طور طبیعی به دنبال کاهش خطا، سرعت زاویه‌ای و گشتاور خواهد بود. در واقع، این تابع از عامل می‌خواهد که ماهواره به وضعیت پایدار و مطلوب برسد و از نوسانات زیاد و مصرف غیرضروری انرژی جلوگیری کند.

۳-۷- فرآیند به‌روزرسانی و یادگیری

عامل یادگیرنده با استفاده از روش یادگیری تقویتی عمیق، پارامترهای کنترلی را به‌صورت زیر به‌روزرسانی می‌کند:

$$Q(N, a) = Q(N, a) + \beta[R + \gamma \max_{a'} Q(N', a') - Q(N, a)] \quad (31)$$

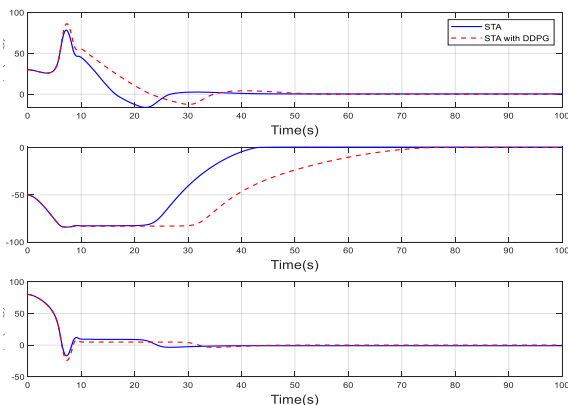
این معادله مربوط به الگوریتم یادگیری مبتنی بر تابع Q (الگوریتم Q-learning) در یادگیری تقویتی است و برای به‌روزرسانی تابع



شکل ۴ پاداش اپیزودیک برای کنترل مود لغزشی فوق پیچشی با یادگیری تقویتی

Fig. 4 Episodic Reward for Super-Twisting Sliding Mode Control with Reinforcement Learning

شکل ۴ نمودار پاداش اپیزودیک برای این کنترلرکه با ترکیب یادگیری تقویتی بدست آمده، عملکرد الگوریتم در طول زمان را نشان می‌دهد. این نمودار بیانگر موفقیت نسبی روش پیشنهادی در کنترل وضعیت سیستم با استفاده از رویکرد ترکیبی است. روند افزایشی میانگین پاداش در طول اپیزودها، توانایی یادگیری و بهبود سیاست‌های کنترلی توسط عامل را تأیید می‌کند.



شکل ۵ مقایسه تغییرات زوایای اویلر در شرایط عادی برای کنترل کننده فوق پیچشی برای ماهواره

Fig. 5 Comparison of Euler Angle Variations under Normal Conditions for the Super-Twisting Controller of the Satellite ring

شکل ۵، پاسخ زوایای اویلر (ψ, θ, ϕ) ، تحت تأثیر دو ساختار کنترلی مورد مقایسه قرار گرفته‌اند؛ یکی کنترل کننده مود لغزشی فوق پ (STA) و دیگری نسخه توسعه یافته آن با بهره‌گیری از الگوریتم یادگیری تقویتی (DDPG) مشاهده می‌شود که در محور ϕ ، کنترل کننده STA دارای نوسان شدیدی در شروع حرکت است، به گونه‌ای که مقدار زاویه ابتدا به حدود ۱۹۰ درجه افزایش یافته و سپس تا حدود منفی ۸۰ درجه کاهش می‌یابد. این نوسانات در نسخه تقویتی شده با DDPG به شکل محسوسی کاهش یافته و پیک اولیه تنها به حدود ۶۰ درجه محدود شده و نوسانات گذرا در کمتر از ۲۰ ثانیه میرا شده‌اند. این رفتار نشان‌دهنده کاهش

انعطاف پذیر را با در نظر گرفتن ویژگی‌های انعطاف‌پذیری سیستم نشان می‌دهد. یادگیری تقویتی عمیق خروجی سیستم را پردازش کرده و اطلاعات لازم را برای بهبود عملکرد کنترلر ارائه می‌دهد. ترکیب این کنترلر و یادگیری تقویتی عمیق باعث می‌شود سیستم در مقابله با اغتشاشات، عدم قطعیت‌ها و چالش‌های کمبود عملکرد عملکرد بهتری داشته و در شرایط مختلف بهینه عمل کند.

۸- شبیه‌سازی

در این پژوهش، برای آموزش کنترل کننده هوشمند، از الگوریتم یادگیری تقویتی (Deep Deterministic Policy Gradient) استفاده شده است. این الگوریتم به دلیل توانایی در یادگیری در فضاهای پیوسته و کنترل دقیق سیستم‌های غیرخطی، گزینه مناسبی برای مسئله کنترل وضعیت ماهواره انعطاف‌پذیر در شرایط کمبود عملکرد می‌باشد.

معماری شبکه عصبی شامل دو لایه پنهان با ۲۵۶ نرون برای عامل بازیگر (Actor) و منتقد (Critic) طراحی شده است. عامل یادگیرنده در طی ۵۰۰ اپیزود و در هر اپیزود به مدت ۴۰۰ گام زمانی آموزش دیده است. برای افزایش کیفیت همگرایی، از نرخ تخفیف $\gamma=0.99$ و نرخ یادگیری ۰/۰۰۱ استفاده شده و حجم حافظه بازپخش نیز برابر با 10^5 نمونه در نظر گرفته شده است.

در جدول زیر، پارامترهای کلیدی مورد استفاده در پیاده‌سازی الگوریتم DDPG به صورت خلاصه آورده شده‌اند:

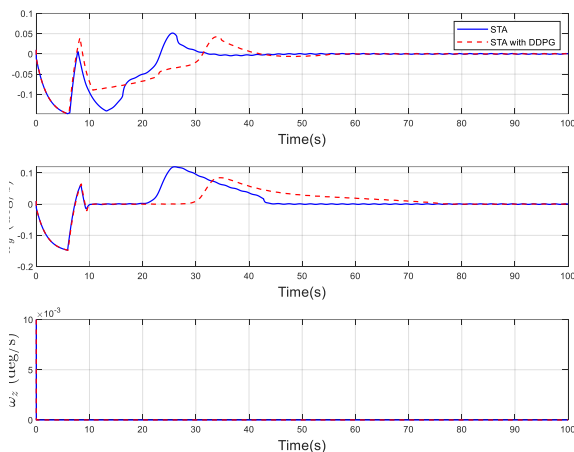
جدول ۱ پارامترهای کلیدی مورد استفاده در الگوریتم یادگیری تقویتی DDPG

Table 1 Sample table accepted by the journal

توضیحات	مقدار	پارامتر
Deep Deterministic Policy Gradient	DDPG	نوع الگوریتم یادگیری
و منتقد (Actor) برای بازیگر (Critic)	۲ لایه، هر کدام با ۲۵۶ نرون	ساختار شبکه عصبی
مقدار α	0.001	نرخ یادگیری (Learning Rate)
برای بهروزرسانی وزن‌ها	$\gamma = 0.99$	نرخ تخفیف (Discount Factor)
برای کاهش وزن پاداش‌های آینده	10^5	اندازه حافظه بازپخش
تعداد نمونه‌هایی که برای یادگیری ذخیره می‌شوند	500	تعداد اپیزودهای آموزشی
تعداد تکرار کامل آموزش	400	تعداد گام در هر اپیزود

پارامترهای نمایش داده شده در جدول ۱، تنظیمات کلیدی الگوریتم DDPG در فرآیند آموزش عامل یادگیرنده را شامل می‌شوند. این پارامترها با هدف دستیابی به تعادل میان دقت، پایداری، سرعت یادگیری و جلوگیری از واگرایی انتخاب شده‌اند. ساختار شبکه، نرخ یادگیری و نرخ تخفیف از عوامل اصلی در همگرایی صحیح عامل یادگیرنده هستند، در حالی که اندازه حافظه بازپخش و تعداد اپیزودها در کیفیت کلی آموزش و تعمیم‌پذیری الگوریتم نقش دارند.

پیک‌هایی با دامنه بیشتر از ۰.۲ درجه بر ثانیه است که به‌تدریج میرا می‌شوند، اما در STA-DDPG این دامنه‌ها به حدود ۰.۱ محدود شده‌اند و سیستم با نوسانی کمتر و رفتار تطبیقی سریع‌تری به حالت پایدار می‌رسد. این تفاوت در کنترل گذرای سرعت‌های زاویه‌ای، نشان‌دهنده تأثیر یادگیری تقویتی در کاهش پاسخ‌های ضربه‌ای و کنترل نرم سیستم است. در محور ω_z نیز هر دو روش کنترل رفتار مشابهی از خود نشان می‌دهند، اما کنترل‌کننده ترکیبی پاسخ سریع‌تر و بدون هیچ پیک اضافی دارد. اگرچه دامنه پاسخ در این محور بسیار کوچک است (در حد 10^{-3} درجه بر ثانیه)، اما تفاوت در نرخ همگرایی سیستم همچنان مشهود است. به‌طور کلی، این نتایج نشان می‌دهد که ترکیب کنترل لغزشی با یادگیری تقویتی باعث کاهش قابل توجه نوسانات در سرعت‌های زاویه‌ای شده، پایداری دینامیکی سیستم را ارتقا داده و امکان کنترل دقیق‌تری را در لحظات بحرانی و حالت گذرا فراهم کرده است. چنین ویژگی‌هایی به‌ویژه در کنترل وضعیت ماهواره‌های انعطاف‌پذیر که حساس به نوسانات و چترینگ هستند، از اهمیت حیاتی برخوردارند.

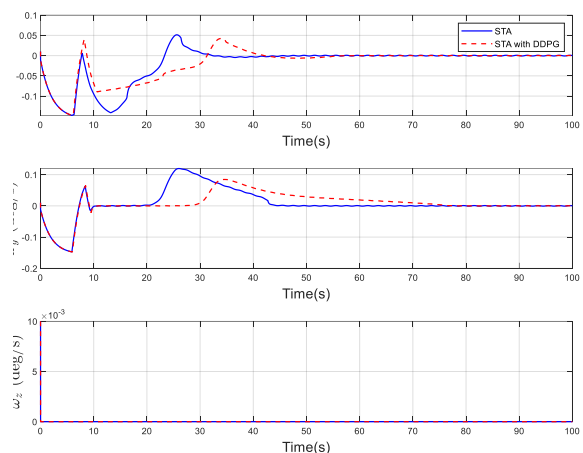


شکل ۷ مقایسه تغییرات گشتاورهای کنترلی در شرایط عادی برای کنترل‌کننده فوق پیچشی برای ماهواره

Fig.7 Comparison of Control Input Variations under Normal Conditions for the Super-Twisting Controller of the Satellite

در بررسی شکل ۷ پاسخ گشتاورهای کنترلی اعمال شده به سیستم نیز مزایای کنترل‌کننده ترکیبی به‌وضوح مشهود است. در محور τ_x ، کنترل‌کننده STA کلاسیک دارای نوسانات شدید و مداومی در کل بازه زمانی شبیه‌سازی است. این نوسانات با دامنه نسبتاً بالا و رفتار ضربه‌ای نه‌تنها نشان‌دهنده چترینگ شدید در خروجی کنترل است، بلکه ممکن است در عمل موجب آسیب به عملگرها و کاهش طول عمر سخت‌افزار شود. در مقابل، کنترل‌کننده STA-DDPG با وجود شروع اولیه نوسانی، خیلی سریع نوسانات را به‌صورت تطبیقی کاهش داده و گشتاور خروجی را در یک بازه محدود، یکنواخت و بدون چترینگ نگه داشته است. در محور τ_y نیز روندی

قابل توجه خطای اولیه و بهبود سرعت همگرایی است. در محور θ نیز کنترل‌کننده STA با دامنه نوسان بالاتر از ۲۰۰ درجه و زمان نشست نسبتاً طولانی عمل می‌کند. در حالی که نسخه STA-DDPG با کاهش دامنه نوسانات به حدود $50 \pm$ درجه، در کمتر از ۵۰ ثانیه به پایداری کامل می‌رسد. این کاهش زمان نشست در کنار کنترل دقیق‌تر دینامیک گذرا، دلیلی بر رفتار تطبیقی مؤثر یادگیری تقویتی است که منجر به کاهش چترینگ و کنترل نرم‌تر سیستم شده است. در محور ψ نیز، هرچند نوسانات کلی کمتر است، ولی نسخه کلاسیک دارای پیک اولیه در حدود ۱۰۰ درجه می‌باشد، در حالی که نسخه ترکیبی با DDPG نوسان اولیه را به زیر ۵۰ درجه کاهش داده و پاسخ به حالت پایدار با سرعت بیشتری انجام شده است. این تحلیل به روشنی نشان می‌دهد که ترکیب کنترل مود لغزشی با الگوریتم یادگیری تقویتی، توانسته است کنترلر را از منظر دقت، سرعت پاسخ، پایداری و کاهش نوسانات بهینه‌سازی کند و آن را برای استفاده در سیستم‌های حساس همچون ماهواره‌های انعطاف‌پذیر به گزینه‌ای بسیار کارآمدتر تبدیل سازد.



شکل ۶ مقایسه تغییرات سرعت‌های زاویه‌ای در شرایط عادی برای کنترل‌کننده فوق پیچشی برای ماهواره

Fig.6 Comparison of Angular Velocity Variations under Normal Conditions for the Super-Twisting Controller of the Satellite

شکل ۶ نشان می‌دهد در پاسخ سرعت‌های زاویه‌ای نیز روند بهبود عملکرد کنترل‌کننده ترکیبی به‌وضوح قابل مشاهده است. در محور ω_x ، کنترل‌کننده STA رفتار ناپایدار و پرنوسانی را از خود نشان می‌دهد به‌طوری که در بازه‌ی زمانی ابتدایی، چندین نوسان شدید با دامنه‌هایی نزدیک به ۰.۱ درجه بر ثانیه ایجاد می‌کند. این نوسانات با گذشت زمان به آرامی کاهش می‌یابند، در حالی که نسخه ترکیبی با یادگیری تقویتی از ابتدا نوسانات را محدود نگه داشته و بدون تولید پیک‌های شدید، سیستم را در مدت کوتاه‌تری به تعادل می‌رساند. این کنترلر ملایم‌تر در حضور شرایط گذرا باعث افزایش پایداری سیستم شده است. در محور ω_y نیز وضعیت مشابهی مشاهده می‌شود؛ کنترل‌کننده STA در چند ثانیه‌ی اول دارای

دینامیکی اشاره دارد و می‌تواند نمایانگر فاصله، زاویه یا هر ویژگی دیگری از وضعیت سیستم باشد. به‌طور کلی، معادله نشان می‌دهد که $S_{1,2}$ از ترکیب این دو بخش تشکیل شده است، که یکی به سرعت‌های زاویه‌ای و دیگری به یک کمیت خاص از سیستم مربوط می‌شود. ترکیب این دو بخش می‌تواند برای تحلیل یا کنترل رفتار دینامیکی سیستم استفاده شود. و برای محور سوم:

$$S_{1,3} = C_1 \omega_3 + C_2 q_4 \quad (33)$$

در این معادله، $S_{1,3}$ یک کمیت کنترلی یا معیار سیستم است که به ترکیب دو پارامتر مختلف وابسته است C_1 و C_2 ضرایب ثابت هستند که معمولاً برای تنظیم تأثیر سرعت زاویه‌ای ω_3 و کمیت q_4 بر روی $S_{1,3}$ به‌کار می‌روند. این ضرایب به‌طور معمول از طریق تحلیل سیستم یا شبیه‌سازی‌های مختلف تنظیم می‌شوند. معمولاً نمایانگر سرعت زاویه‌ای سیستم در محور سوم است که به‌طور معمول برای توصیف چرخش یا حرکت سیستم در یک محور خاص استفاده می‌شود q_4 . نیز به یک کمیت دیگر مربوط است که ممکن است نمایانگر زاویه، موقعیت، یا دیگر ویژگی‌های سیستم باشد. این کمیت معمولاً برای توصیف یکی از ویژگی‌های دینامیکی سیستم استفاده می‌شود. به‌طور کلی، معادله نشان می‌دهد که $S_{1,3}$ به‌طور ترکیبی از این دو بخش تشکیل شده است و می‌تواند برای مدل‌سازی یا کنترل رفتار سیستم‌های دینامیکی مختلف استفاده شود.

۹-۱-۲- کنترل معادل (Equivalent Control) برای محورهای اول و دوم

ابتدا زیرماتریس اینرسی برای محورهای ۱ و ۲ تعریف می‌شود:

$$J_{(t,12)} = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} \in \mathbb{R}^{2 \times 2} \quad (34)$$

که نشان‌دهنده‌ی زیرماتریس 2×2 از ماتریس اینرسی I_t مربوط به محورهای اول و دوم است. سپس کنترل معادل به صورت زیر بازنویسی می‌شود:

$$u_{eq_{12}} = -J_{(t,12)}^{-1} \left[(\omega \times (J\omega))_{1,2} \right] \quad (35)$$

در این رابطه، J ماتریس اینرسی ماهواره است که ویژگی‌های دینامیکی جرم ماهواره را نسبت به محورهای مختصات توصیف می‌کند. $(\cdot)_{1,2}$ مؤلفه‌های اول و دوم بردار مربوط به محورهای ۱ و ۲ است. عبارت $\omega \times (J\omega)$ نشان‌دهنده ضرب خارجی بین بردار

مشابه مشاهده می‌شود STA. کلاسیک دارای رفتار نوسانی با دامنه‌های بزرگ و ماندگار است، در حالی که کنترل‌کننده ترکیبی پس از گذشت زمان بسیار کوتاهی، نوسانات را به سطح حداقلی کاهش داده و ورودی‌های ملایم‌تری به عملگرها وارد می‌کند. این پایداری در ورودی کنترلی تأثیر مستقیمی بر پایداری کلی سیستم و عملکرد بلندمدت آن دارد. در محور τ_z نیز، اگرچه دامنه گشتاور در هر دو روش بسیار کوچک است، اما روش ترکیبی رفتار یکنواخت‌تری دارد و کوچک‌ترین نوسانات را حذف کرده است. در مجموع، کنترل‌کننده ترکیب‌شده با الگوریتم یادگیری تقویتی، با تولید سیگنال کنترلی نرم‌تر، کاهش چترینگ و محدود کردن نوسانات ورودی، کنترل دقیق‌تر و ایمن‌تری برای سیستم فراهم کرده است که این ویژگی برای سیستم‌های حساس فضایی مانند ماهواره‌های انعطاف‌پذیر بسیار حیاتی است.

۹- طراحی کنترل مود لغزشی تطبیقی مرتبه بالا

کنترل مود لغزشی تطبیقی مرتبه بالا به‌عنوان روشی مؤثر برای کنترل سیستم‌های انعطاف‌پذیر و تحت شرایط کمبود عملگر طراحی شده است. این روش با تعریف سطوح لغزش (Sliding Surface) در لایه‌های مختلف، به سیستم اجازه می‌دهد که به سمت پایداری هدایت شود. هدف از این کنترل‌کننده، افزایش دقت کنترل، کاهش چترینگ (Chattering)، و مقابله با اغتشاشات خارجی و نامعینی‌های مدل دینامیکی است. در این بخش، روابط و معادلات این کنترل‌کننده توضیح داده می‌شود.

۹-۱-۱- معادلات کنترل‌کننده مود لغزشی تطبیقی مرتبه بالا کنترل‌کننده به‌صورت زیر تعریف می‌شود:

۹-۱-۱- محاسبه سطح لغزش لایه اول

سطح لغزش برای محورهای اول و دوم به‌صورت زیر محاسبه می‌شود:

$$S_{1,2} = C_1 \omega_{1,2} + C_2 q_{2,3} \quad (32)$$

در این معادله، $S_{1,2}$ یک کمیت کنترلی است که از ترکیب دو پارامتر مختلف بدست می‌آید C_1 و C_2 ضرایب ثابت هستند که تأثیر سرعت زاویه‌ای $\omega_{1,2}$ و کمیت $q_{2,3}$ را بر روی $S_{1,2}$ تنظیم می‌کنند. این ضرایب معمولاً به‌طور دستی یا از طریق بهینه‌سازی تنظیم می‌شوند تا شدت تأثیر هر یک از این پارامترها بر کمیت کنترلی مورد نظر کنترل شود $\omega_{1,2}$. معمولاً نمایانگر سرعت‌های زاویه‌ای در دو محور از سیستم است که ممکن است مربوط به چرخش سیستم یا ماهواره در دو جهت مختلف باشد. این بخش به‌طور مستقیم به رفتار چرخشی سیستم وابسته است $q_{2,3}$. نیز می‌تواند نمایانگر یک پارامتر دیگر مانند زاویه بین دو جسم یا وضعیت نسبی سیستم باشد. این پارامتر معمولاً به ویژگی‌های خاص سیستم‌های

ورودی کنترلی برای محور سوم به طور مستقیم صفر خواهد ماند، زیرا تنها با استفاده از اصلاحات مربوط به محورهای اول و دوم این محور کنترل می‌شود.

$$u_3 = 0 \quad (۳۹)$$

در نهایت ورودی کلی به صورت زیر است: [۲۲]

$$U_{control} = \begin{bmatrix} U_{12} \\ u_3 \end{bmatrix} \quad (۴۰)$$

۱۰- اثبات پایداری کنترلر با استفاده از تابع لیپانوف

تابع لیپانوف پیشنهادی باید نمایانگر انرژی کل سیستم باشد. برای سیستم کنترل شده، یک تابع لیپانوف مناسب می‌تواند به صورت زیر تعریف شود:

$$V(S) = \frac{1}{2} S^T S \quad (۴۱)$$

که در آن $S = [S_1, S_2]^T$ سطح لغزش است.

$$S = \begin{bmatrix} S_1 \\ S_2 \end{bmatrix} \quad (۴۲)$$

$$S_1 = C_1 \omega_1 + C_2 q_2 \quad (۴۳)$$

$$S_2 = C_1 \omega_2 + C_2 q_3 \quad (۴۴)$$

این تابع لیپانوف نشان‌دهنده یک انرژی شبه مربعی در سیستم است که سطح لغزش را دنبال می‌کند.

۱۰-۱- محاسبه مشتق زمانی تابع لیپانوف

با مشتق‌گیری از تابع لیپانوف، داریم:

$$\dot{V}(S) = S^T \dot{S} \quad (۴۵)$$

با استفاده از تعریف سطح لغزش، مشتق آن برابر خواهد بود با:

$$\dot{S} = C_1 \dot{\omega} + C_2 \dot{q} \quad (۴۶)$$

از معادلات دینامیکی:

$$J \dot{\omega} = -\omega \times (J \omega) + u \quad (۴۷)$$

سرعت زاویه‌ای ω و حاصل ضرب ماتریس اینرسی در همان بردار سرعت زاویه‌ای است؛ که در مجموع، گشتاور دینامیکی ناشی از چرخش ماهواره را بیان می‌کند.

۹-۱-۳- کنترل سوئیچینگ (Switching Control) و کاهش چترینگ

برای کاهش چترینگ، از یک ضریب تطبیقی استفاده می‌شود که به صورت زیر تعریف می‌شود:

$$\rho_{12} = \rho_0 + \alpha \cdot e^{-\beta |s_{122}|} \quad (۳۶)$$

در این معادله، ρ_{12} ضریب تطبیقی مربوط به محورهای اول و دوم است که مقدار اولیه آن با ρ_0 مشخص می‌شود. پارامتر α برای افزایش تدریجی این ضریب به کار می‌رود و β نرخ کاهش چترینگ را تعیین می‌کند.

کنترل سوئیچینگ به صورت زیر اعمال می‌شود:

$$u_{sw12} = -k S_{12} - \rho_{12} \cdot \text{sign}(S_{12}) \quad (۳۷)$$

در این معادله، u_{sw12} ورودی کنترلی است که به طور خاص در کنترل مود لغزشی برای هدایت سیستم به سمت پایداری طراحی شده است. k یک ضریب تنظیمی است که شدت تاثیر سطح لغزش S_{12} را تنظیم می‌کند. این بخش از معادله به طور مستقیم مرتبط با کنترل خطای سیستم است که سیستم را به سمت وضعیت پایدار هدایت می‌کند S_{12} . سطح لغزش است که معمولاً از ترکیب خطاها و سرعت خطا به دست می‌آید و نشان‌دهنده میزان انحراف سیستم از وضعیت مطلوب است. بخش دوم معادله $\rho_{12} \cdot \text{sign}(S_{12})$ به کاهش چترینگ (نوسانات گشتاور) کمک می‌کند. در اینجا، ρ_{12} یک ضریب تنظیمی است که شدت تاثیر علامت سطح لغزش را بر ورودی کنترلی کنترل می‌کند. $\text{sign}(S_{12})$ تابعی است که علامت سطح لغزش را تعیین می‌کند و جهت حرکت سیستم را تنظیم می‌کند تا سیستم از سطح لغزش خارج نشود. به طور کلی، این معادله برای تنظیم ورودی کنترلی در سیستم‌های چرخشی و دینامیکی است که هدف آن‌ها کاهش نوسانات و هدایت سیستم به سمت پایداری است.

گشتاور کلی برای محورهای اول و دوم برابر است با:

$$U_{12} = u_{eq12} + u_{sw12} \quad (۳۸)$$

$$t \rightarrow \infty \quad S \rightarrow 0 \text{ در نهایت} \quad (59)$$

و رابطه $\dot{q} = \omega$ ، داریم:

$$\dot{S} = C_1 J_t^{-1} (-\omega \times (J_t \omega) + \tau) + C_2 \omega \quad (48)$$

این یعنی پایداری مجانبی سیستم تضمین شده است.

با جایگذاری کنترلر طراحی شده:

$$u = u_{eq_{12}} + u_{sw} \quad (49)$$

در نتیجه:

$$u_{eq_{12}} = -J_{(t,12)}^{-1} \left[(\omega \times (J_t \omega))_{1,2} \right] \quad (50)$$

$$u_{sw} = -kS - \rho_0 \text{sign}(S) \quad (51)$$

بنابراین، مشتق سطح لغزش:

$$\dot{S} = C_1 J_t^{-1} (-\omega \times (J_t \omega) + u_{eq} + u_{sw}) + C_2 \omega \quad (52)$$

با جایگذاری این مقدار در معادله مشتق لیاپانوف:

$$\dot{V}(S) = S^T (C_1 J_t^{-1} (-\omega \times (J_t \omega) + u_{eq} + u_{sw}) + C_2 \omega) \quad (53)$$

چون کنترلر معادل ($u_{eq_{12}}$) دقیقاً مقدار گشتاور مورد نیاز برای تعادل سیستم است، مقدار آن به گونه‌ای انتخاب شده که دینامیک غیرخطی را جبران کند. در نتیجه:

$$S^T u_{eq_{12}} \approx 0 \quad (54)$$

با جایگذاری این مقادیر در معادله مشتق لیاپانوف:

$$\dot{V}(S) = S^T (-kS - \rho_0 \text{sign}(S)) \quad (55)$$

با ساده‌سازی و جایگذاری کنترلر معادل و سوئیچینگ:

$$\dot{V}(S) = -kS^T S - \rho_0 S^T \text{sign}(S) \quad (56)$$

داریم: $S^T \text{sign}(S) = \|S\|$ و $S^T S = \|S\|^2$

$$\dot{V}(S) = -k\|S\|^2 - \rho_0 \|S\| \quad (57)$$

که همواره منفی است زیرا k و $\rho_0 > 0$. از معادله بالا نتیجه می‌گیریم که:

$$\dot{V}(S) \leq 0 \quad (58)$$

که نشان‌دهنده پایداری لیاپانوف است.

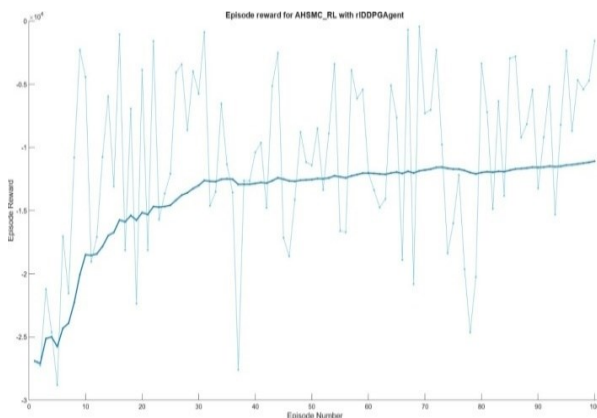
همچنین، از قضیه لیاپانوف، با توجه به اینکه $\dot{V}(S)$ منفی معین است و تابع لیاپانوف مثبت معین است، نتیجه می‌شود که سطح لغزش S در طول زمان به صفر همگرا می‌شود، یعنی:

۱۱- استفاده از یادگیری تقویتی برای بهبود عملکرد کنترلر مود لغزشی تطبیقی مرتبه بالا

در این بخش، از یادگیری تقویتی برای بهبود عملکرد کنترلر مود لغزشی تطبیقی مرتبه بالا استفاده شده است. این روش با الگوبرداری از ساختار و فرایند کلی به‌کارگیری یادگیری تقویتی که در کنترلر مود لغزشی فوق پیچشی ارائه شد، طراحی شده است. یادگیری تقویتی در هر دو مورد باهدف تنظیم تطبیقی ضرایب کنترلی و افزایش توانایی سیستم در مقابله با اغتشاشات و نامعینی‌های محیطی استفاده می‌شود. با این حال، تفاوت اساسی در اینجا، ساختار کنترلرکننده و نحوه تعریف سطوح لغزش است. کنترلر مود لغزشی تطبیقی مرتبه بالا با چندین سطح لغزش طراحی شده است. این سطوح لغزش به سیستم امکان می‌دهند که به‌صورت مرحله‌ای و با دقت بیشتر به سمت پایداری حرکت کند. در این ساختار، یادگیری تقویتی وظیفه تنظیم ضرایب تطبیقی هر سطح را بر عهده دارد و از طریق الگوریتم یادگیری تقویتی، سیستم را قادر می‌سازد به تغییرات دینامیکی محیط پاسخ داده و در برابر اغتشاشات پیچیده و خرابی عملگرها عملکرد مطلوبی داشته باشد

۱۲- شبیه‌سازی

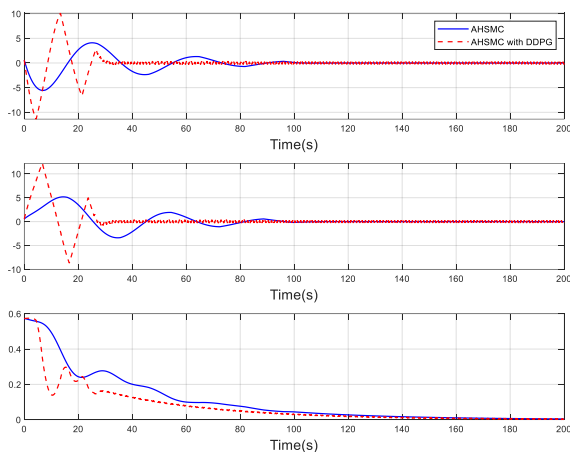
عامل یادگیرنده در محیط متلب شبیه‌سازی شده و کنترلر مود لغزشی تطبیقی مرتبه بالا را تنظیم می‌کند. شبیه‌سازی‌ها برای بررسی عملکرد سیستم در شرایط مختلف خرابی عملگر و اغتشاشات خارجی انجام می‌شود.



شکل ۸ پاداش اپیزودیک برای مود لغزشی تطبیقی مرتبه بالا با استفاده از یادگیری تقویتی

Fig. 8 Episodic Reward for Adaptive High-Order Sliding Mode Control Using Reinforcement Learning

به سمت وضعیت تعادل هدایت کرده است. این نکته تأیید می‌کند که حتی در محورهایی که کنترل نسبتاً ساده‌تر است، یادگیری تقویتی همچنان می‌تواند باعث کاهش پاسخ ضربه‌ای و هموارسازی رفتار سیستم شود. در مجموع، عملکرد ترکیبی AHSMC با یادگیری تقویتی در این سه محور حاکی از آن است که این ساختار نه تنها توانسته دقت و پایداری را افزایش دهد، بلکه رفتار کنترل را به سمت پاسخ نرم، بدون نوسان و مطمئن‌تر هدایت کرده است؛ ویژگی‌هایی که برای کنترل وضعیت ماهواره در فضا به شدت حیاتی هستند.

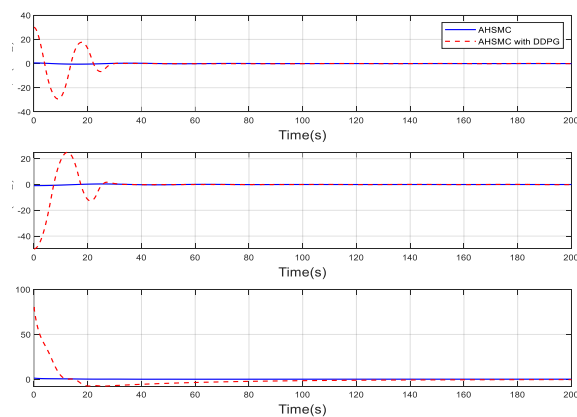


شکل ۱۰ مقایسه تغییرات سرعت های زاویه ای در شرایط عادی برای کنترل کننده مدل لغزشی تطبیقی برای ماهواره

Fig. 10 Comparison of Angular Velocity Variations under Normal Conditions for the Adaptive Sliding Mode Controller of the Satellite

در شکل ۱۰ بررسی پاسخ سرعت‌های زاویه‌ای نیز برتری قابل توجه کنترل‌کننده ترکیبی AHSMC-DDPG را نسبت به نسخه کلاسیک AHSMC تأیید می‌کند. در محور ω_x ، کنترل‌کننده کلاسیک رفتار نوسانی شدید و پریودی در ۵۰ ثانیه ابتدایی از خود نشان می‌دهد، به طوری که سیستم چندین بار از حالت تعادل خارج شده و به آرامی به سمت پایداری میل می‌کند. در مقابل، AHSMC ترکیب شده با DDPG دامنه نوسانات را به میزان چشمگیری کاهش داده و با رفتاری نرم‌تر و پایدارتر، سیستم را در زمانی کوتاه‌تر به تعادل رسانده است. در محور ω_y نیز رفتار مشابهی مشاهده می‌شود. کنترل‌کننده کلاسیک دارای پیک‌هایی متوالی در بازه‌ی زمانی ۰ تا ۶۰ ثانیه است که نشان‌دهنده تلاش کنترلر برای مهار اغتشاشات گذرا است، در حالی که نسخه یادگیرنده با دامنه نوسانات کوچکتر و سرعت همگرایی بالاتر، کنترل بهینه‌تری ارائه داده است. این ویژگی به ویژه در سناریوهای واقعی که در آن عملکرد نرم و پایدار در اولویت است، اهمیت دوچندان دارد. در محور ω_z نیز هرچند دامنه نوسانات در هر دو کنترل‌کننده بسیار کوچک است (در حد 10^{-3} درجه بر ثانیه)، اما تفاوت در کیفیت همگرایی کاملاً مشهود است. کنترل‌کننده AHSMC دارای نوسانات جزئی و رفت‌وبرگشتی در مراحل ابتدایی

در شکل ۸ بیانگر عملکرد ترکیب کنترلر با یادگیری تقویتی است که روند تغییرات پاداش اپیزودیک را در طول زمان نشان می‌دهد. که در مراحل اولیه، مقدار پاداش بسیار متغیر بوده و به تدریج افزایش یافته است. این امر به دلیل کاوش عامل در فضای تصمیم‌گیری و یادگیری سیاست‌های کنترلی مناسب است. پس از گذشت چند اپیزود، میانگین پاداش بهبود یافته و نوسانات تا حدودی کاهش یافته‌اند، که نشان‌دهنده پیشرفت عامل در یادگیری یک سیاست پایدارتر است. نوسانات پاداش همچنان در برخی اپیزودها مشاهده می‌شود، اما روند کلی حاکی از افزایش میانگین پاداش و تطبیق بهتر این کنترلر با شرایط محیطی است. این نوسانات ممکن است به دلیل عدم قطعیت‌های مدل، پیچیدگی محیط، یا چالش‌های مرتبط با همگرایی الگوریتم باشد.



شکل ۹ مقایسه تغییرات زوایای اوایلر در شرایط عادی برای کنترل کننده مدل لغزشی تطبیقی برای ماهواره

Fig. 9 Comparison of Euler Angle Variations under Normal Conditions for the Adaptive Sliding Mode Controller of the Satellite

شکل ۹ در ادامه‌ی بررسی ساختارهای کنترلی، نمودارهای زاویه‌های اوایلر برای کنترل‌کننده تطبیقی مرتبه بالا (AHSMC) و نسخه هوشمند آن با الگوریتم یادگیری تقویتی (DDPG) نشان می‌دهند که ترکیب یادگیری تقویتی بار دیگر به بهبود عملکرد سیستم منجر شده است. در محور ϕ ، کنترل‌کننده AHSMC کلاسیک دارای نوسانات ناپایدار و میرا نشده در بازه ابتدایی تا حدود ۶۰ ثانیه است که با چندین پیک رفت‌وبرگشتی همراه است. در مقابل، AHSMC-DDPG با وجود داشتن یک رفتار نوسانی اولیه ملایم، موفق شده در زمانی کمتر از ۵۰ ثانیه پاسخ را به حالت پایدار برساند و از نوسانات بزرگ جلوگیری کند. در محور θ نیز کنترل‌کننده کلاسیک در ابتدا دارای نوسانات پر دامنه و با روند کندتری نسبت به نسخه DDPG است. نسخه یادگیرنده، پیک اولیه را با شیب آهسته‌تری آغاز کرده و سپس در زمان حدود ۶۰ ثانیه به حالت پایدار رسیده است. این رفتار نشان‌دهنده کنترل بهتر در مرحله گذرا، کاهش احتمال چترینگ و افزایش نرمی در کنترل دینامیک است. در محور ψ نیز، AHSMC کلاسیک دارای نوسانی سریع است که در کمتر از ۲۰ ثانیه میرا می‌شود، در حالی که AHSMC-DDPG با رفتاری کاملاً نرم و یکنواخت، بدون پیک محسوس، سیستم را

در مجموع، استفاده از الگوریتم یادگیری تقویتی در ساختار AHSMC منجر به کاهش چترینگ، نرم‌سازی پاسخ‌های کنترلی و کاهش فشار روی عملگرها شده است؛ این موارد برای سیستم‌های حساس مانند ماهواره‌های انعطاف‌پذیر با منابع محدود، مزیت‌های حیاتی محسوب می‌شوند.

جدول ۲ مقایسه روش‌های کنترلی استفاده شده همراه با یادگیری تقویتی

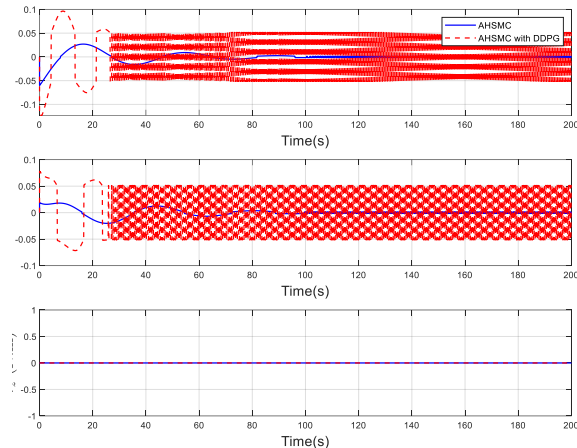
Table 2 Comparison of Control Methods Implemented with Reinforcement Learning

نوع کنترلر	مقاوم (نیوتن بر متر)	چترینگ (نیوتن بر متر)	زمان نشست (ثانیه)	میانگین مربعات خطا
STA	0.001	0.024	15	0.0001193
STA RL	0.01	0.009	15	5.691e-7
AHSMC	0.001	0.021	35	0.0001957
AHSMC RL	0.01	0.002	75	6.615e-6

باتوجه به جدول ابررسی عملکرد مقایسه عددی ارائه شده در جدول فوق به روشنی نشان می‌دهد که استفاده از الگوریتم یادگیری تقویتی در کنار ساختارهای کنترلی کلاسیک STA و AHSMC تأثیر قابل توجهی در بهبود عملکرد کلی سیستم داشته است. کنترل‌کننده‌های تقویتی STA_RL و AHSMC_RL با کاهش محسوس مقدار میانگین مربعات خطا (MSE) نشان داده‌اند که توانایی بالاتری در کاهش انحراف از مسیر مطلوب دارند. به‌ویژه کنترل‌کننده STA_RL با مقدار خطای بسیار کوچک در حد $e-7 \ 0/79$ عملکرد دقیقی از خود نشان داده است. از منظر چترینگ نیز ساختارهای ترکیب شده با یادگیری تقویتی موفق شده‌اند دامنه نوسانات پر فرکانس خروجی کنترلی را کاهش دهند، به طوری که میزان چترینگ در STA_RL به حدود $0/009$ و در AHSMC_RL به تنها $0/002$ نیوتن‌متر کاهش یافته است. این کاهش به معنای کاهش فشار مکانیکی بر عملگرها و بهبود پایداری دینامیکی سیستم است. در زمینه مقاومت نسبت به اغتشاش نیز نسخه‌های RL عملکرد قوی‌تری نشان داده‌اند؛ مقدار نیروی مقاوم در این ساختارها به $0/01$ نیوتن‌متر افزایش یافته که بیانگر افزایش تحمل سیستم در برابر اغتشاشات بیرونی و شرایط غیرخطی می‌باشد. هرچند در مورد AHSMC_RL زمان نشست نسبت به نسخه کلاسیک افزایش یافته و به ۷۵ ثانیه رسیده است، اما این افزایش با کاهش چشمگیر چترینگ و بهبود دقت همراه بوده که نشان از یک کنترل تطبیقی با رفتار نرم‌تر و ایمن‌تر دارد.

در مجموع، این نتایج عددی تأیید می‌کنند که کنترل‌کننده‌های ترکیب شده با یادگیری تقویتی در دستیابی به تعادل میان دقت بالا، پایداری کنترلی، کاهش نوسانات و افزایش مقاومت سیستم عملکردی برتر نسبت به نسخه‌های کلاسیک ارائه می‌دهند و برای کاربردهای فضایی حساس گزینه‌ای به مراتب قابل اعتمادتر هستند.

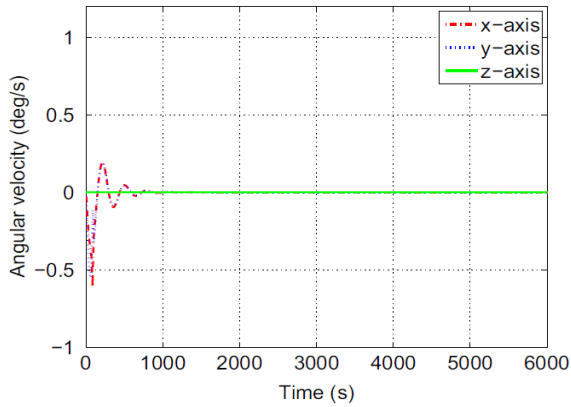
است، در حالی که نسخه DDPG به شکل یکنواخت و پیوسته به نقطه تعادل می‌رسد. در مجموع، استفاده از یادگیری تقویتی در ساختار AHSMC باعث کاهش قابل توجه نوسانات سرعت زاویه‌ای، بهبود پایداری دینامیکی و افزایش دقت در کنترل گذرا شده است. این رفتار تطبیقی در کنار کاهش چترینگ و مصرف انرژی، این ساختار را به گزینه‌ای قابل اعتماد برای کنترل وضعیت ماهواره در شرایط عملیاتی پیچیده تبدیل می‌کند.



شکل ۱۱ مقایسه تغییرات گشتاورهای کنترلی در شرایط عادی برای کنترل کننده مود لغزشی تطبیقی برای ماهواره

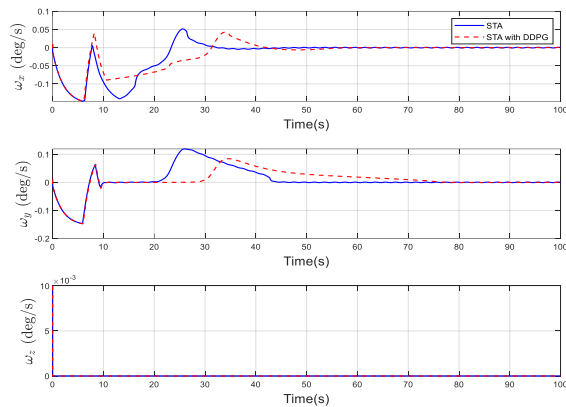
Fig. 11 Comparison of Control Input Variations under Normal Conditions for the Adaptive Sliding Mode Controller of the Satellite

شکل ۱۱ نمودارهای گشتاورهای کنترلی نیز بیانگر تفاوت بارز میان عملکرد کنترل‌کننده کلاسیک AHSMC و نسخه ترکیب شده با یادگیری تقویتی آن هستند. در محور τ_x ، کنترل‌کننده کلاسیک دچار نوسانات بسیار شدید و پر تکرار شده است، به طوری که سیگنال کنترلی دارای رفتار چترینگ مداوم با دامنه نسبتاً بالا در کل بازه زمانی مشاهده می‌شود. این رفتار برای عملگرهای واقعی خطرناک بوده و باعث کاهش دقت و استهلاک مکانیکی می‌گردد. در مقابل، کنترل‌کننده AHSMC-DDPG موفق شده با تعدیل ورودی و افزایش نرمی پاسخ، نوسانات را در کمتر از ۵۰ ثانیه مهار کرده و سپس گشتاور را در بازه‌ای پایدار و یکنواخت نگه دارد. در محور τ_y نیز رفتار مشابهی دیده می‌شود. کنترل‌کننده کلاسیک دارای نوسانات سینوسی ممتد با دامنه بالا است، در حالی که نسخه تقویتی با DDPG وجود شروع اولیه‌ی نوسانی، به سرعت چترینگ را کاهش داده و کنترل ملایم‌تری را اعمال کرده است. این کاهش نوسانات در عمل به معنای افزایش طول عمر عملگرها و کاهش مصرف انرژی است. در محور τ_z نیز، هر دو کنترل‌کننده رفتاری تقریباً مشابه و پایدار ارائه داده‌اند؛ اما تفاوت در دقت و نرخ همگرایی نسخه‌ی تقویتی قابل مشاهده است. این محور که گشتاورهای آن بسیار کوچک است، بیشتر تحت تأثیر نوسانات ساختارهای کنترلی نیست، اما همچنان دقت عملکرد AHSMC-DDPG از نظر کنترل یکنواخت برجسته است.



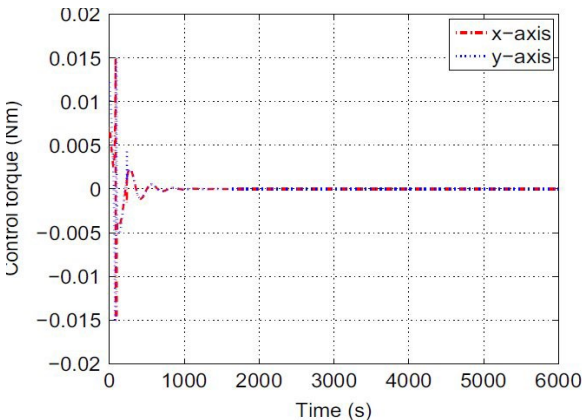
شکل ۱۴ مقایسه تغییرات سرعت زاویه ای از مقاله [۳]

Fig. 14 Comparison of angular velocity variations from [3].



شکل ۱۵ مقایسه تغییرات سرعت زاویه ای در شرایط عادی برای کنترل کننده فوق پیچشی ماهواره

Fig. 15 Comparison of angular velocity variations under normal conditions for the satellite super-twisting controller.

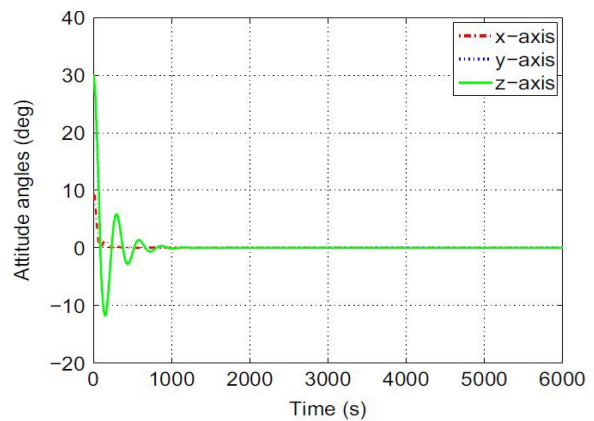


شکل ۱۶ مقایسه تغییرات گشتاور کنترلی از مقاله [۳]

Fig. 16 Comparison of control torque variations from [3]

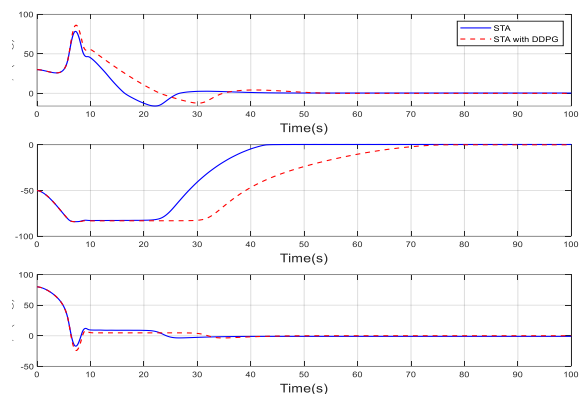
۱۳- صحت‌سنجی و مقایسه با مطالعات پیشین

در به دلیل نبود پژوهش‌های پیشین که به طور مستقیم به کنترل وضعیت ماهواره‌های انعطاف‌پذیر با دو چرخ عکس‌العملی پرداخته باشند، ناگزیر شدیم برای ارزیابی و مقایسه، از یک مقاله [۲۳] نزدیک در حوزه‌ی کنترل ماهواره‌های صلب با دو چرخ عکس‌العملی استفاده کنیم. مقاله‌ی مرجع انتخاب‌شده، طراحی و تحلیل یک کنترل‌کننده غیرخطی برای یک ماهواره صلب underactuated را ارائه می‌دهد و عملکرد آن را در شرایط مختلف بررسی کرده است.



شکل ۱۲ مقایسه تغییرات زوایای اویلر از مقاله [۳]

Fig. 12 Comparison of Euler angle variations from [3].



شکل ۱۳ مقایسه تغییرات زوایای اویلر در شرایط عادی برای کنترل کننده فوق پیچشی ماهواره

Fig. 13 Comparison of Euler angle variations under normal conditions for the satellite super-twisting controller.

یادگیری تقویتی، در جنبه‌هایی نظیر پایداری سریع‌تر، پاسخ نرم‌تر، دقت بالاتر و مدیریت بهتر ورودی‌های کنترلی، عملکرد بهتری نسبت به در مجموع، مقایسه کمی و کیفی انجام شده میان دو ساختار کنترلی بیانگر آن است که روش پیشنهادی این مقاله، به‌ویژه در ترکیب با یادگیری تقویتی، در جنبه‌هایی نظیر پایداری سریع‌تر، پاسخ نرم‌تر، دقت بالاتر و مدیریت بهتر ورودی‌های کنترلی، عملکرد بهتری نسبت به روش‌های موجود دارد. این نتایج مؤید صحت، کارایی و قابلیت اتکای ساختار کنترلی ارائه شده در این تحقیق هستند.

جدول ۳ مقایسه عملکرد کنترل کننده ها

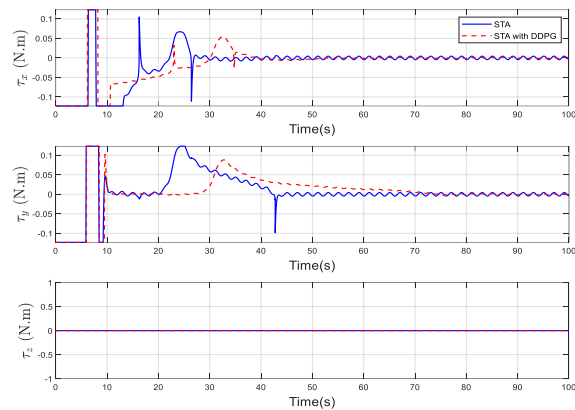
Table 3 Comparison of controller performance

شاخص عملکرد	مقاله مرجع	STA	STA with DDPG
Settling Time φ (s)	800-1000	35-40	20-25
Settling Time θ (s)	800-1000	45-50	30-35
Settling Time ψ (s)	700-900	20-25	12-15
Overshoot φ (%)	5-10	70-80	10-15
Overshoot θ (%)	≤ 10	≈ 100	≈ 50
Overshoot ψ (%)	≤ 10	10-15	5-8
Peak Torque ω_x	0.5-0.6	0.45-0.50	0.20-0.25
Peak Torque ω_y	0.25-0.30	0.35-0.40	0.18-0.22
Peak Torque ω_z	≈ 0	$\approx 10^{-3}$	$\approx 10^{-3}$
Peak Angular Velocity τ_x	0.015	0.04-0.05	0.018-0.022
Peak Angular Velocity τ_y	0.015	0.025-0.030	0.012-0.016
Peak Angular Velocity τ_z	0	0	0

جدول ۳ نشان می‌دهد مقایسه عملکرد نشان می‌دهد که ترکیب کنترل کننده فوق‌پیچشی (STA) با یادگیری تقویتی DDPG نسبت به هر دو روش مرجع و STA معمولی برتری قابل توجهی دارد. از نظر زمان نشست، این روش زمان رسیدن به پایداری را از چند صد ثانیه در کنترل کننده مرجع به تنها چند ده ثانیه کاهش داده و در تمامی زوایای وضعیت سریع‌ترین همگرایی را ارائه می‌دهد. میزان فراجاهش نیز نسبت به STA به‌تنهایی به شکل چشمگیری کاهش یافته و در مقایسه با روش مرجع در سطحی پایین‌تر قرار گرفته، در حالی که از فراجاهش بسیار زیاد مشاهده شده در STA معمولی جلوگیری شده است. در پاسخ‌های گشتاور و سرعت زاویه‌ای، روش STA همراه با DDPG مقادیر پیک بسیار کمتری ایجاد کرده که این موضوع نشان‌دهنده اعمال کنترل نرم‌تر و کاهش فشار مکانیکی بر عملگرها است. تلاش کنترلی نه تنها از نظر دامنه کوچک‌تر شده، بلکه به شکلی بهینه توزیع شده که به افزایش بهره‌وری انرژی کمک می‌کند. به طور کلی، ترکیب STA با یادگیری تقویتی (DDPG) پاسخی سریع‌تر، دقیق‌تر و پایدارتر همراه با کاهش فراجاهش و بهینه‌سازی استفاده از عملگرها ارائه داده و به وضوح عملکردی بهتر از هر دو روش مرجع و STA معمولی دارد.

۱۴- نتیجه‌گیری و پیشنهادها

ترکیب STA و AHSMC با DRL در این پژوهش عملکرد بهتری نسبت به روش‌های کلاسیک نشان داده است. این بهبود عملکرد



شکل ۱۷ مقایسه تغییرات گشتاور کنترلی در شرایط عادی برای کنترل کننده فوق پیچشی ماهواره

Fig. 17 Comparison of control torque variations under normal conditions for the satellite super-twisting controller.

شکل های ۱۲ و ۱۳ و ۱۴ نشان می‌دهند که مقایسه نتایج مقاله مرجع با داده‌های حاصل از این پژوهش، برتری روش پیشنهادی به‌ویژه در ترکیب کنترل کننده فوق‌پیچشی با یادگیری تقویتی به وضوح قابل مشاهده است. در نمودارهای مقاله مرجع، فرآیند رسیدن به وضعیت پایدار زمان بیشتری نیاز دارد و نوسانات گذرای قابل توجهی مشاهده می‌شود، در حالی که در این پژوهش و به خصوص در حالت استفاده از یادگیری تقویتی، پاسخ سیستم سریع‌تر به مقدار مرجع رسیده و دامنه نوسانات اولیه به شکل محسوسی کاهش یافته است. سرعت‌های زاویه‌ای نیز در روش پیشنهادی با میراشدن سریع‌تر و دامنه کمتر همراه بوده و خطای ماندگار تقریباً صفر به دست آمده که بیانگر پایداری بالاتر سیستم است. گشتاورهای کنترلی در مقاله مرجع دارای پیک‌های اولیه بزرگ‌تر و گذرای طولانی‌تر هستند، اما در این پژوهش استفاده از یادگیری تقویتی باعث ایجاد پروفایل گشتاور ملایم‌تر، کاهش پدیده چترینگ و توزیع بهینه تلاش کنترلی شده است. این امر علاوه بر کاهش فشار مکانیکی روی عملگرها، به بهره‌وری انرژی سیستم نیز کمک می‌کند. ترکیب یادگیری تقویتی با کنترل کننده فوق‌پیچشی امکان سازگاری تطبیقی با تغییرات دینامیکی و مقاومت در برابر نامعینی‌ها را فراهم کرده و بدون نیاز به تیونینگ مکرر پارامترها، عملکردی سریع، پایدار و مقاوم در حضور اغتشاشات ارائه می‌دهد. نتایج به دست آمده نشان می‌دهد که این روش در تمام شاخص‌های کلیدی، بر مقاله مرجع برتری دارد و می‌تواند به عنوان رویکردی کارآمد در کنترل وضعیت ماهواره مورد استفاده قرار گیرد.

در مجموع، مقایسه کمی و کیفی انجام شده میان دو ساختار کنترلی بیانگر آن است که روش پیشنهادی این مقاله، به‌ویژه در ترکیب با

همزمان بر روی چندین متغیر کنترلی نظارت کرده و تصمیمات بهینه‌ای برای کنترل بهتر وضعیت ماهواره اتخاذ کند.

در این مقاله، دو روش کنترل مود لغزشی فوق پیچشی و کنترل مود لغزشی تطبیقی مرتبه بالا برای کنترل ماهواره‌های انعطاف‌پذیر در شرایط کمبود عملگر طراحی و شبیه‌سازی شد. این روش‌ها به دلیل ویژگی‌های مقاوم در برابر اغتشاشات و خرابی عملگرها، عملکرد بهتری نسبت به روش‌های کلاسیک مانند کنترل فیدبک حالت ارائه دادند. با ترکیب این کنترلرها با یادگیری تقویتی، پایداری سیستم در شرایط مختلف بهبود یافت و خطاهای سیستم به طور قابل توجهی کاهش پیدا کرد. نتایج شبیه‌سازی نشان داد که ترکیب کنترل مود لغزشی فوق پیچشی و کنترل مود لغزشی تطبیقی مرتبه بالا با یادگیری تقویتی می‌تواند به طور مؤثر عملکرد ماهواره‌های انعطاف‌پذیر را در شرایط مختلف، از جمله خرابی عملگرها و اغتشاشات محیطی بهبود بخشد. این روش‌ها علاوه بر کاهش خطاها، به بهبود پایداری و کاهش پدیده چترینگ کمک کردند. در پایان، پیشنهاد می‌شود که در مقالات آینده، علاوه بر بهبود روش‌های کنترلی پیشنهادی، به موضوعاتی مانند بهینه‌سازی بیشتر ضرایب کنترلی و استفاده از سایر روش‌های هوش مصنوعی برای بهبود عملکرد در شرایط پیچیده‌تر پرداخته شود.

تأییدیه اخلاقی

از آنجا که این مطالعه صرفاً بر پایه شبیه‌سازی‌های عددی و مدل‌سازی رایانه‌ای انجام شده و شامل هیچ‌گونه آزمایش انسانی یا حیوانی نمی‌باشد، دریافت تأییدیه اخلاقی ضروری نبوده است.

تعارض منافع

نویسندگان بدین وسیله اعلام می‌کنند که هیچ‌گونه تعارض منافع اعم از مالی، تجاری، نهادی یا شخصی در تهیه، تحلیل و نگارش این مقاله نداشته‌اند.

منابع مالی/حمایت‌ها

نگارش و اجرای این تحقیق بدون بهره‌مندی از منابع مالی یا حمایت‌های مستقیم از سوی هیچ نهاد یا سازمانی صورت گرفته است.

مراجع

- [1] M. M. Alipour Shahraki, M. Malekzadeh, and A. Ariaei, "Active fractional-order sliding mode control of flexible spacecraft under actuators saturation," *Journal of Sound and Vibration*, vol. 535, Art. no. 117110, Sep. 2022. doi: 10.1016/j.jsv.2022.117110.
- [2] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction* (no. 1). MIT press Cambridge, 1998.

ناشی از چندین مکانیزم کلیدی است که در این بخش تحلیل می‌شود.

• کاهش چترینگ

یکی از مشکلات اصلی در کنترل مود لغزشی کلاسیک، پدیده چترینگ است که به نوسانات شدید و غیرمطلوب در ورودی‌های کنترلی منجر می‌شود. استفاده از STA در ترکیب با DRL موجب کاهش این نوسانات می‌شود. در این ترکیب، STA قادر است چترینگ را به میزان قابل توجهی کاهش دهد، زیرا با استفاده از الگوریتم‌های پیچیده‌تری که در STA به کار گرفته شده، ورودی‌های کنترلی به طور نرم‌تر و با دقت بیشتری تنظیم می‌شوند. DRL نیز به صورت پویا و تطبیقی به بهینه‌سازی ضرایب کنترلی کمک می‌کند و از تغییرات محیطی و وضعیت سیستم مطلع است، که به کاهش نوسانات و چترینگ بیشتر کمک می‌کند.

• بهبود سرعت همگرایی

یکی از مزایای DRL در این ترکیب، بهبود سرعت همگرایی سیستم است. در ترکیب STA با DRL، الگوریتم یادگیری تقویتی می‌تواند به طور مستمر ضرایب کنترلی را به روز کند تا کنترل‌گری به وضعیت پایدار برسد. در حالی که STA به تنهایی در برخی شرایط ممکن است زمان نشست طولانی‌تری داشته باشد، اضافه کردن DRL باعث می‌شود که سیستم بتواند سریع‌تر به هدف خود برسد. این بهبود در سرعت همگرایی بیشتر به دلیل قابلیت DRL در یادگیری تجربی و تطبیق با تغییرات شرایط سیستم است.

• تنظیم تطبیقی ضرایب کنترلی

یکی دیگر از عواملی که موجب بهبود عملکرد در ترکیب STA و AHSMC با DRL می‌شود، استفاده از قابلیت تطبیقی DRL در تنظیم ضرایب کنترلی است. به طور مستمر ضرایب کنترلی را بهینه‌سازی می‌کند تا سیستم بتواند عملکرد بهتری در برابر اغتشاشات محیطی، خرابی عملگرها و عدم قطعیت‌های مدل خود نشان دهد. این تطبیق خودکار و پویا باعث بهبود دقت و پایداری سیستم در طول زمان می‌شود.

• مقاومت بیشتر در برابر عدم قطعیت‌ها

در شرایطی که مدل سیستم دقیق نیست و یا در مواجهه با اغتشاشات غیرمنتظره، ترکیب STA و AHSMC با DRL می‌تواند عملکرد سیستم را به طور قابل توجهی بهبود بخشد. DRL می‌تواند خود را با شرایط محیطی مختلف تطبیق دهد و از طریق تجربه و یادگیری از محیط، راهکارهای بهینه‌تری را برای مقابله با این عدم قطعیت‌ها ارائه دهد. این ویژگی باعث می‌شود که سیستم در برابر شرایط غیر ایده‌آل و پر از تغییرات مقاوم‌تر باشد.

• کنترل نرم‌تر و بهینه‌تر در شرایط پیچیده

با ترکیب STA و AHSMC با DRL، سیستم توانایی بهتری در کنترل دقیق‌تر و نرم‌تر رفتارهای دینامیکی ماهواره دارد. این ویژگی به ویژه در شرایط بحرانی و در زمانی که کنترل دقیق و کم‌نوسان برای پایداری سیستم حیاتی است، اهمیت دارد. DRL می‌تواند به طور

doi: 10.1177/0954410017711722.

[17] H. Li, W. Yan, and Y. Shi, "Continuous-time model predictive control of under-actuated spacecraft with bounded control torques," *Automatica*, vol. 75, pp. 144–153, Jan. 2017. doi: 10.1016/j.automatica.2016.09.024.

[18] K. D. Kumar and A. Zou, "Robust stationkeeping and reconfiguration of underactuated spacecraft formations," *Acta Astronautica*, vol. 105, no. 2, pp. 495–510, Dec. 2014. doi: 10.1016/j.actaastro.2014.10.008.

[19] X. Liu, Z. Meng, and Z. You, "Adaptive collision-free formation control for under-actuated spacecraft," *Aerospace Science and Technology*, vol. 79, pp. 223–232, Jan. 2018.

doi: 10.1016/j.ast.2018.05.040.

[20] V. Muralidharan and M.R. Emami, "Concurrent rendezvous control of underactuated spacecraft," *Acta Astronautica*, vol. 138, pp. 28–42, May 2017.

doi: 10.1016/j.actaastro.2017.05.003.

[21] C.M. Pong, A. Saenz Otero, and D.W. Miller, "Autonomous thruster failure recovery on underactuated spacecraft using model predictive control," Tech. Rep., Massachusetts Institute of Technology, Cambridge, MA, Oct. 2011.

[22] T. Chen, J. Shan, and H. Wen, "Distributed adaptive attitude control for networked underactuated flexible spacecraft," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 1, pp. 215–225, Feb. 2019.

doi: 10.1109/TAES.2018.2849904.

[23] C. Han, J. Guo, and A. Pechev, "Nonlinear H^∞ based underactuated attitude control for small satellites with two reaction wheels," *Acta Astronautica*, vol. 104, no. 1, pp. 159–172, 2014.

[3] C. Han, J. Guo, and A. Pechev, "Nonlinear H^∞ based underactuated attitude control for small satellites with two reaction wheels," *Acta Astronautica*, vol. 104, no. 1, pp. 159–172, 2014.

[4] M. Mirshams and M. Khosrojerdi, "Attitude control of an underactuated spacecraft using tube-based MPC approach," *Aerospace Science and Technology*, vol. 48, pp. 140–145, Jan. 2016. doi: 10.1016/j.ast.2015.09.018.

[5] B. Xiao, Q. Hu, Y. Zhang, and X. Huo, "Fault-tolerant tracking control of spacecraft with attitude-only measurement under actuator failures," *Journal of Guidance, Control, and Dynamics*, vol. 37, no. 3, pp. 838–849, May 2014. doi: 10.2514/1.61369.

[6] D. Wang, Y. Jia, L. Jin, and S. Xu, "Control analysis of an underactuated spacecraft under disturbance," *Acta Astronautica*, vol. 83, pp. 44–53, Feb. 2013.

doi: 10.1016/j.actaastro.2012.10.029.

[7] P. Tsiotras and J. Luo, "Control of underactuated spacecraft with bounded inputs," *Automatica*, vol. 36, no. 8, pp. 1153–1169, Aug. 2000. doi: 10.1016/S0005-1098(00)00025-X.

[8] A. Behal, D. M. Dawson, E. Zergeroglu, and Y. Fang, "Nonlinear tracking control of an underactuated spacecraft," in *Proceedings of the 2002 American Control Conference (IEEE Cat. No. CH37301)*, vol. 6, pp. 4684–4689, May 2002, doi: 10.2514/2.4973.

[9] X. Huang, Y. Yan, and Y. Zhou, "Nonlinear control of underactuated spacecraft hovering," *Journal of Guidance, Control, and Dynamics*, vol. 39, no. 3, pp. 685–694, May 2016. doi: 10.2514/1.G001306.

[10] C. D. Petersen, F. Leve, and I. Kolmanovsky, "Model predictive control of an underactuated spacecraft with two reaction wheels," *Journal of Guidance, Control, and Dynamics*, vol. 40, no. 2, pp. 320–332, Feb. 2017. doi: 10.2514/1.G000320.

[11] X. Huang, Y. Yan, and Y. Zhou, "Analytical solutions to optimal underactuated spacecraft formation reconfiguration," *Advances in Space Research*, vol. 56, no. 10, pp. 2151–2166, Nov. 2015.

doi: 10.1016/j.asr.2014.10.014.

[12] J. Yu, Z. Li, L. Jia, and Y. Zhang, "Switching neural network control for underactuated spacecraft formation reconfiguration in elliptic orbits," *Applied Sciences*, vol. 12, no. 12, Art. no. 5792, Jun. 2022.

doi: 10.3390/app12125792.

[13] X. Huang and Y. Yan, "Saturated backstepping control of underactuated spacecraft hovering for formation flights," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 4, pp. 1988–2000, Aug. 2017.

doi: 10.1109/TAES.2017.2716799.

[14] X. Huang, Y. Yan, Y. Zhou, and D. Hao, "Fast terminal sliding mode control of underactuated spacecraft formation reconfiguration," *Journal of Aerospace Engineering*, vol. 29, no. 5, Art. no. 04016020, Sep. 2016.

doi: 10.1061/(ASCE)AS.1943-5525.0000610.

[15] X. Huang and L. Chen, "Indirect optimization of underactuated spacecraft formation reconfiguration in elliptic orbits," *Journal of Aerospace Engineering*, vol. 34, no. 1, Art. no. 04020086, Jan. 2021.

doi: 10.1061/(ASCE)AS.1943-5525.0001209.

[16] X. Huang, Y. Yan, and Y. Zhou, "Dynamics and control of underactuated spacecraft formation reconfiguration in elliptic orbits," *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, vol. 232, no. 12, pp. 2214–2230, Dec. 2018.