فصلنامه علمی فناوری در مهندسی هوافضا

دریافت مقاله: ۱۳۹۹/۰۵/۲۰
پذیرش مقاله: ۱۴۰۰/۰۳/۲۰
DOI:10.22034/JTAE.2022.140118

(یادداشت فنی)

طراحی مسیر بهینه مانورهای پهلوگیری خودکاربراساس یادگیری تقویتی Qو شبکه‌بندی مکعبی

ایمان شفیعی‌نژاد*[۱]، محمد صیامی عراقی[۲]، علیرضا سخاوت بنیس***[۲]، علی میرزایی***[۲] و ایمان فزونی تلوکی**[۳]**

۱- پژوهشگاه هوافضا، وزارت علوم تحقیقات و فناوری، تهران، ایران، کدپستی: ۱۴۶۶۵۸۳۴.

۲- گروه مهندسی هوافضا، دانشکده فنی و مهندسی، دانشگاه آزاد اسلامی واحد علوم و تحقیقات، تهران، ایران.

۳- دانشگاه آزاد اسلامی واحد قائم شهر، قائم شهر، ایران.

* استادیار (نویسنده پاسخگو)، ایمیل:
shafieenejad@ari.ac.ir

** دانشجوی کارشناسی ارشد

*** دانشجوی دکتری

در این مقاله با بهره‌گیری از الگوریتم یادگیری کیو، طراحی مسیر بهینه مانورهای فضایی خودکار برای یک ربات فضایی محاسبه و انجام شده است. افزایش روزافزون تعداد ماهواره‌های ارسال شده برای قرار گیری در مدارهای مختلف حول زمین، طراحی و ساخت سرویس‌های پشتیبان فضایی را مورد توجه محققین قرار داده است. در این راست، ابا توجه به پیشرفت‌های انجام گرفته در علوم رباتیک بهره‌گیری از ربات‌های فضایی خودکار به جهت تعمیر و سرویس‌دهی به ماهواره‌های آسیب دیده گزینه‌ای مناسب تلقی می‌شود. هدایت، کنترل و ناوبری یک ربات فضایی در فازهای پهلوگیری و اتصال به ماهواره سرویس‌گیرنده نیازمند دقت بالا یی است. به همین جهت در این مقاله با رویکرد سنجش نحوه کارکرد الگوریتم یادگیری کیو در مانورهای فضایی پهلوگیری و اتصال در فضا به وسیله شبیه‌سازی‌های کامپیوتری مورد بررسی قرار گرفته است. نتایج این تحقیق بیانگر دقت بالا الگوریتم یادگیری کیو است.

واژه‌های کلیدی: یادگیری تقویتی،یادگیری کیو،بهینه‌سازی مسیر حرکت، سرویس درون‌مداری

Optimal Path Planning for Autonomous Space Maneuvers Based on Reinforcement Q-learning and Cubic Network

I. Shafieenejad[1*], M. Siami Araghi[2], A. Sekhavat Benis[2***], A. Mirzaee[2***], and I. Fozouni Taloki[3**]**

1- Aerospace Research Institute, Ministry of Science, Research and Technology, Postal Code: 1465774111, Tehran, IRAN

2- Department of Aerospace Engineering, Science and Research Branch, Islamic Azad University, Tehran, IRAN

3- Ghaemshahr Branch, Islamic Azad University, Ghaemshahr, IRAN

* Assistant Professor (Corresponding Author): Email:
shafieenejad@ari.ac.ir

** M.Sc. Student

*** Ph.D. Student

This paper proposes a method based on the reinforcement Q-learning to solve the problem of fully autonomous optimal path planning of a space robot by increasing the number of available satellites in earth orbits, designation, and implementation of satellite orbital servicing stations considered by researchers. Nowadays, by considering the advancement in robotics science, space robots could be chosen as a part of solution for maintaining the damaged satellites in earth's orbits. Guidance, control, and navigation of space robots throughout docking and joint maneuvers need a high degree of precision. In this paper, reinforcement Q-learning algorithm functionality in path planning is analyzed through various computational simulations. The finding results from computational simulations demonstrate the usefulness of the mentioned approach.

Keywords: Reinforcement Learning, Q-learning, Path-planning Optimization, Orbital Maneuver, On-orbit Servicing

(یادداشت فنی)

ایمان شفیعی نژاد، محمد صیامی عراقی، علیرضا سخاوت، علی میرزایی و ایمان فزونی تلوکی

۱- مقدمه

پس از پرتاب موفقیت‌آمیز اولین ماهواره مصنوعی ساخته دست بشر در تاریخ چهارم اکتبر سال ۱۹۵۷ توسط اتحاد جماهیر شوروی به نام اسپوتینک ۱ بیش از ۸۹۰۰ ماهواره در ۶ دهه اخیر به مدارهای کرهٔ زمین ارسال شده است [۱]. درصد بالایی از ماهواره‌های ارسال شده توانسته‌اند بدون مواجه با مشکلات فنی مأموریت خود را به طور کامل به اتمام برسانند، اما در این بین مشکلات ناشی ازخرابی باتری، عملکرد نادرست سیستم‌های مکانیکی مرتبط با صفحه‌های خورشیدی، اختلال در سیستم کنترل موقعیت ماهواره موجب از کار افتادن ماهواره و شکست مأموریت‌های فضایی متعددی شده است [۲].

از کار افتادگی سیستم‌های تأمین‌کننده و ذخیره‌کنندهٔ انرژی مانند صفحه‌های خورشیدی و باتری‌ها در ماهواره‌ها از جمله مهمترین دلایل از کارافتادگی ناگهانی ماهواره‌های در حال گردش در مدار و سرویساست. با توجه به پرهزینه بودن فرآیندهای طراحی و ارسال سامانه‌های ماهواره‌ای به فضا، دو عامل عمر عملیاتی کوتاه و بروز نقص‌های فنیناگهانی در سیستم‌های ماهواره‌ای در حالگردش در مدار سبب اتلاف هزینه‌های میلیارد دلاری از سرمایه‌های شرکت‌های سازنده فناوری‌های ماهواره‌ای و در نتیجه افزایش هزینه‌های عرضهٔ این فناوری‌ها گشته است [۳, ۴].

سازمان‌های فضایی کشورهای ایالات متحده امریکا (ناسا[1]) و آلمان (دی.ال.ار[2]) از دهه ۸۰ میلادی شروع به تحقیق و توسعه به جهت عملی‌سازی انجام مأموریت‌های سرویس درون‌مداری نموده‌اند. سرویس درون‌مداری یا او.اس.اس[3]به سیستم فضایی طراحی شده برای قرارگیری در مدارهای مختلف حول زمین اشاره می‌شود که قابلیت تعمیر، سوار کردن قطعات، ارتقاء و سوخت‌رسانی را برای ماهواره‌ها و پایگاه‌های فضایی فراهم می‌سازد. ماموریت‌های سرویسدرون‌مداری می‌توانند شامل به‌کارگیری هم زمان نیروی انسانی و ربات‌های فضایی و یا تنها به‌کارگیری سیستم‌های رباتیک خودکار فضایی باشد. در گذشته تعداد زیادی از سیستم‌های فضایی، تعمیر و ارتقاء یافته‌اند. از برجسته‌ترین آنها می‌توان به مأموریت تعمیر و ارتقا تلسکوپ فضایی هابل[4] اشاره نمود که طی سال‌های ۱۹۹۳ تا ۲۰۰۹ میلادی در ۵ مأموریت فضایی توسط ناسا انجام شده است. در این مأموریت‌ها با به‌کارگیری هم زمان فضانوردان و ربات‌های فضایی، سیستم‌های فنی تلسکوپ فضایی هابل مورد تعمیر و ارتقا قرار گرفته است [۵].

تمامی مأموریت‌های سرویس درون‌مداری از سه بخش اصلی فضاپیما یا ربات فضایی، بازوی‌های رباتیک متصل به ربات فضایی و ماهواره سرویس‌گیرنده تشکیل می‌شود. فضاپیمایا ربات فضایی وظیفه دارد تا پس از تکمیل فرآیند قرارگیری در مدار، ماهواره سرویس‌گیرنده را رهگیری نموده و با استفاده سیستم‌های پیشران، خود را به موقعیت هدف‌گذاری شده برساند. پس از تکمیل مانور رهگیری، ربات فضایی می‌بایست با تعیین دقیق مسیر پهلوگیری[5] خود را به فاصله قابل دسترسی برای بازوی رباتیک متصل به خود برساند تا بازوی رباتیک بتواند عملیات اتصال[6] نهایی را تکمیل و عملیات سرویس‌دهی به ماهواره سرویس‌گیرنده را آغاز نماید. با توجه به پیچیده بودن فرآیندهای هدایت، کنترل و ناوبری به واسطه اجرا توسط انسان، طراحی و پیاده‌سازی سیستم‌های خودکار که توانایی انجام فرآیندهای ذکر شده را بدون نیاز به دخالت انسان داشته باشند امری ضروری است. بر همین اساس، در سال‌های اخیر تحقیقات بر روی الگوریتم‌های کنترلی کارآمد جهت محاسبه هوشمند مسیر بهینه اتصال فضاپیماها و کپسول‌های فضایی مورد توجه محققان این حوزه قرار گرفته است [۶].

به طور معمول، در مأموریت‌های سرویس درون‌مداری با توجه به وجود عدم قطعیت در مشخصه‌های ساختاری ماهواره سرویس‌گیرنده مانند وزن سوخت مصرف شده توسط سیستم‌های پیشرانیا تغییرات ایجاد شده در محل‌های قرارگیری سلول‌های خورشیدی، همچنین اغتشاش و غیر خطی بودن رفتار جرم فضاپیما، محاسبهٔ تخمینی متغیرهای دینامیکی بسیار دشوار است. با توجه به دشواری در محاسبهٔ متغیرهای دینامیکی ماهواره، استفاده از الگوریتم‌های کنترلی سنتی مانند PD که براساس متغیرهای دینامیکی ماهواره طراحی می‌شوند غیر ممکن است. بر این اساس، بهره‌گیری از الگوریتم‌های بدون ناظر مانند الگوریتم‌های یادگیری تقویتی که می‌توانند بدون نیاز به مدل از پیش تعیین شده از رفتار سیستم، وضعیت فضاپیما یا ربات فضایی را به جهت تکمیل مانورهای پهلوگیری و اتصال کنترل نمایند انتخابی مناسب است [۷].

در طراحی الگوریتم‌های یادگیری تقویتی از منطق رفتاری حیوانات الگوبرداری شده است. در الگوریتم‌های یادگیری تقویتی برای تحلیل رفتار مدل‌های دینامیکی پیچیده از شیوه تعامل

1. NASA (National Aeronautics and SpaceAdministration)
2. DLR (DeutschesZentrumfürLuft- und Raumfahrte.V.)
3. OSS (On-Orbit Satellite Servicing)
4. Hubble Space Telescope
5. Rendezvous
6. Docking

مستقیم با محیط استفاده می‌شود. در این روش، الگوریتم به طور مستقیم با محیط تعامل دارد و با تفسیر مستمر بازخورد تصمیم‌گیری‌های خود در هر زمان بهترین راه حل را براساس تجارب پیشین خود بر می‌گزیند [۸].

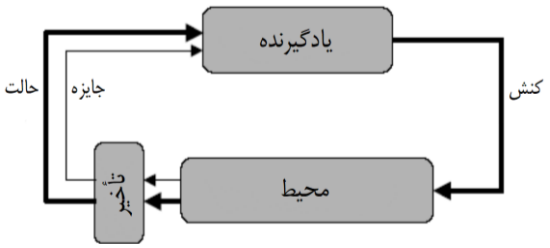
الگوریتم یادگیری کیو[1] با توجه به سادگی در پیاده‌سازی یکی از پر استفاده‌ترین الگوریتم‌های بر پایه یادگیری تقویتی است. از الگوریتم یادگیری کیو برای حل مسائل مرتبط با کنترل وضعیت و ناوبری سیستم‌های رباتیکبه خصوص در محیط‌های مغشوش استفاده شده است [۹]. در مرجع [۱۰] نویسنده از مدل ساده‌سازی شده الگوریتم یادگیری کیو برای ناوبری خودمختار یک ربات هولونومیک[2]در محیط ناشناخته بهره گرفته است. درمراجع [۱۳-۱۱] نویسندگان برای ارتقا عملکرد الگوریتم یادگیری کیو در حل مسائل پیچیده از ترکیب الگوریتم یادگیری کیو با منطق فازی استفاده کرده‌اند. اگرچه ترکیب الگوریتم‌های یادگیری کیو با منطق فازی سبب بهبود عملکرد الگوریتم یادگیری کیو در حل مسائل مطرح شده در حالت پیوسته می‌شود، اما با توجه به پیچیدگی‌های حاصل شده از ترکیب این دو الگوریتم پیاده‌سازی آن بر روی سیستم‌های رباتیک حقیقی با توجه به محدودیت‌های سخت افزاری این سیستم‌ها با دشواری هایی همراه است. بر این اساس، تلاش برای بهبود مدل پایه الگوریتمیادگیری کیو برای به‌کارگیری در مسائل محیط پیوسته که قابلیت پیاده‌سازی بر روی سیستم‌های رباتیک حقیقی را نیز داشته باشد، امری ضروری می‌باشد.با توجه به کارکرد مناسب الگوریتم یادگیری کیو در مدل‌های دینامیکی سیستم‌های مغشوش غیر خطی، در این مقاله سعی شده است با مدلسازی مناسب، نحوه تعیین مسیر خودکار ربات فضایی در مرحله پهلوگیری و اتصال نهایی به ماهواره سرویس‌گیرنده با بهره‌گیری از الگوریتم یادگیری کیو، شبیه سازی لازم صورت گرفته و کارایی آن مورد سنجش قرار گیرد.

۲- مانورهای اتصال و عوامل اغتشاشی

به طور کلی، مراحل برقراری اتصال و جدایش از ماهوار سرویس‌گیرنده توسط ربات فضایی می‌تواند به چهار فاز جداگانه پهلوگیری، اتصال، قفل شدن و جدایش تقسیم شود. در فاز پهلوگیری فاصله میان ربات فضایی و ماهواره سرویس‌گیرنده بین ۱۰ کیلومتر تا ۱۰۰ متر می‌باشد [۱۴]. در این فاز با توجه به فاصله بسیار طولانی میان ربات فضایی و ماهواره سرویس‌گیرنده، حسگرهای موجود بر روی ربات فضایی که وظیفه دریافت اطلاعات از هدف را برعهده دارند به سبب وجود خطای ذاتی در اندازه گیری بیشترین میزان اغتشاشات را وارد سیستم محاسباتی رایانه پردازشگر ربات فضایی می‌گردانند. بنابراین، به جهت تعیین دقیق مسیر حرکت نیاز به انجام محاسبات حلقه بسته است. پس از پایان مانور پهلوگیری، در فاز اتصال ربات فضایی به فاصله ۱۰ متری ماهواره سرویس‌گیرنده رسیده است [۱۵]. اگر چه با توجه به فاصله نزدیک ربات فضایی و ماهواره سرویس‌گیرنده در این فاز میزان اغتشاش وارد شونده به سیستم ناوبری ربات فضایی در مقایسه با فاز پهلوگیری بسیار کم تر است، اما به دلیل دقت بالای مورد نیاز به جهت برقراری اتصال میان دو جسم که به طور متوسط ۰,۱ متر خواهد بود، همچنان اغتشاش ناشی از عوامل خارجی و داخلی می‌تواند موجب ایجاد خطا در سیستم‌های هدایت کنترل و ناوبری ربات فضاییشده و درنتیجه فرآیندهای مسیریابی و اتصال را دچار اختلال نماید. یکی از عوامل اصلی ایجاد اغتشاش در این فاز که در خارج از سیستم‌های هدایت، کنترل و ناوبری ربات فضایی قرار دارد، رفتار غیر مشارکتی[3]ماهواره سرویس‌گیرنده است. به طور معمول در مأموریت‌های سرویس درون مداری، ماهواره سرویس‌گیرنده هیچ سیگنالی از موقعیت قرارگیری دقیق خود منتشر نمی‌کند. به همین سبب، تحلیل و واردسازی معادلات حرکت اغتشاشی ماهواره سرویس‌گیرنده به رایانه ربات فضایی می‌تواند موجب پیچیدگی فرآیند پردازش به جهت هدایت، کنترل و ناوبری ربات فضایی شود. بنابراین، استفاده از روش ارائه شده در این مقاله می‌تواند این مشکب را برطرف نماید.

۳- کاربرد الگوریتم‌های یادگیری تقویتی

الگوریتم‌های یادگیری تقویتی به جهت یافتن پاسخ بهینه، به طور مستقیم با محیط ارتباط برقرار کرده و با تجربه اندوزی از طریق تکرار متناوب فرآیند سعی و خطا، راه حل‌های مناسب را به وسیله امتیازدهی دسته‌بندی می‌نمایند (شکل **۱**). در هر مرحله از فرآیند سعی و خطا میزان بهینگی راه حل براساس میزان مثبت و یا منفی بودن امتیاز پاداش توسط هسته مرکزی الگوریتم تحلیل می‌شود. با ادامه روند حل و افزایش تجربه الگوریتم می‌تواند با سنجش راه حل‌های قابل دسترس، راه حل دارای بالاترین امتیاز را به عنوان پاسخ نهایی مسئله اعلام نماید [۱۶].



شکل (۱): ساختار الگوریتم یادگیری تقویتی.

1. Q-learning
2. Holonomic
3. Non-cooperative

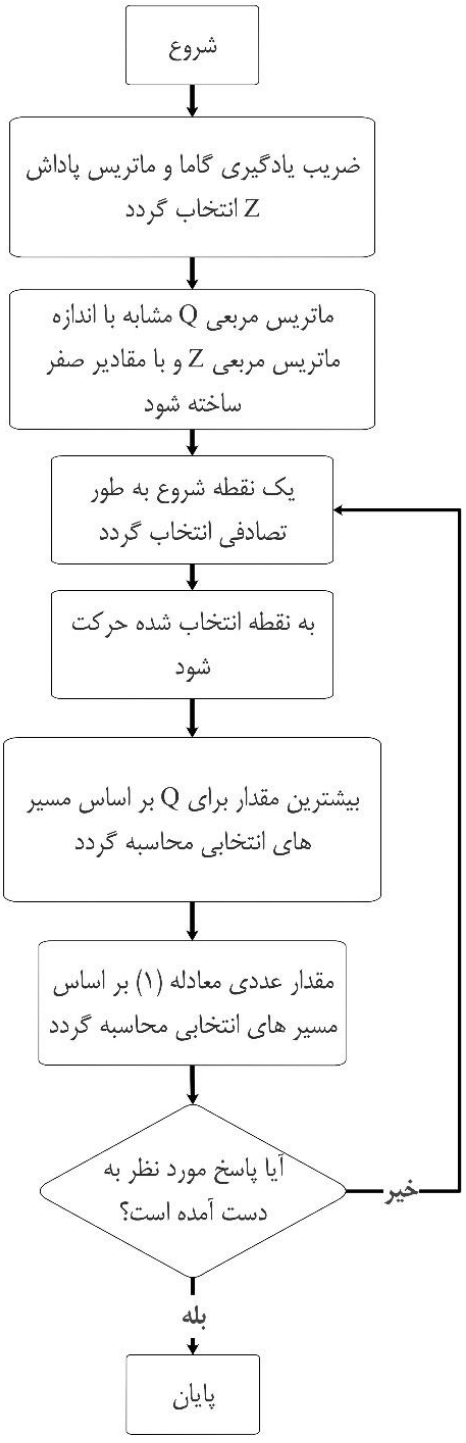
روش‌های یادیگری تقویتی می‌توانند بدون نیاز به مدل از پیش تعیین شده از محیط و تنها با تکیه بر تکرار متناوب فرآیند سعی و خطا و بدون دخالت ناظر خارجی، راه حل بهینه را در محیط‌های غیر خطی و مغشوش ارائه نمایند. با توجه به وجود منابع تولید اغتشاش در مسئله مورد بحث در این مقاله، استفاده از این روش می‌تواند موجب ساده‌سازی و افزایش دقت در مسیرهای بهینه تولید شده برخط[1] به جهت انجام صحیح مانورهای پهلوگیری و متصال شود.

۳-۱- جذب انتشار تک زاویه

الگوریتم یادیگیری کیویکی از روش‌های یادیگری تقویتی بدون ناظر است که می‌تواند بدون نیاز به مدل از پیش تعیین شده از محیط و هدایت ناظر خارجی، راه حل نهایی را به روش کسب تجربه از طریق تکرار متناوب سعی و خطا و امتیازدهی به دست آورد. با توجه به عملکرد برخط این الگوریتم و سرعت بالای آن در همگرا شدن به پاسخ نهایی، می‌توان از این الگوریتم به جهت کنترل و مسیریابی حرکات سیستم‌های رباتیک بهره گرفت. الگوریتم یادگیری کیو به هر زوج حالت–کنش[2] مقدار Q(s,a) را مرتبط می‌کند که این مقدار برابر مجموع پاداش‌های دریافتی در زمانی است که متغیر حالت s عمل کرده و کنش a انجام شده باشد. سیاست عملکرد الگوریتم به جهت همگرایی به پاسخ نهایی براساس معادلهٔ (۱) تنظیم می‌شود که در این رابطه، گاما ضریب یادگیری بوده که بین اعداد ۰ تا ۱ است. همچنین، Z ماتریس مربعی قانون می‌باشد که چهارچوب رفتارهای مجاز را برای هر زوج حالت–کنش Q(s,a) تعریف می‌نماید.

$$Q\big(s(i),a(i)\big) = Z\big(s(i),a(i)\big) + \gamma ma\,x\big[Q\big(s(i+1),a(i\in[1,N])\big)\big] \quad (۱)$$

در معادلهٔ (۱) متغیر i مرتبهٔ یادگیری است و N تعداد مرتبهٔ نهایی تکرار الگوریتم یادگیری است. روند عملکرد الگوریتم یادگیری کیو در شکل **۲** مورد بررسی قرار گرفته شده است.



شکل (۲): فلوچارت الگوریتم یادگیری کیو.

۴- طراحی مسیر بهینه حرکت در فاز اتصال براساس الگوریتم یادگیری کیو

در این بخش به طراحی مسیر بهینه حرکت در فاز اتصال براساس الگوریتم یادگیری کیو پرداخته خواهد شد. بدین منظور، ابتدا برقراری ارتباط و تجربه اندوزی و سپس نحوه عملکرد الگوریتم یادگیری کیو در ضرایب یادگیری ثابت و متفاوت عنوان می‌شود.

۴-۱- برقراری ارتباط و تجربه اندوزی

در مسئله مورد بررسی در این مقاله ماهواره سرویس‌گیرنده در مدار دایره‌ای شکل با شعاع R_0 و پارامتر گرانشی μ و حرکت متوسط n براساس معادله (۲) و با سرعت زاویه ای $\vec{\omega}$ براساس

1. On-line

2. State-Action

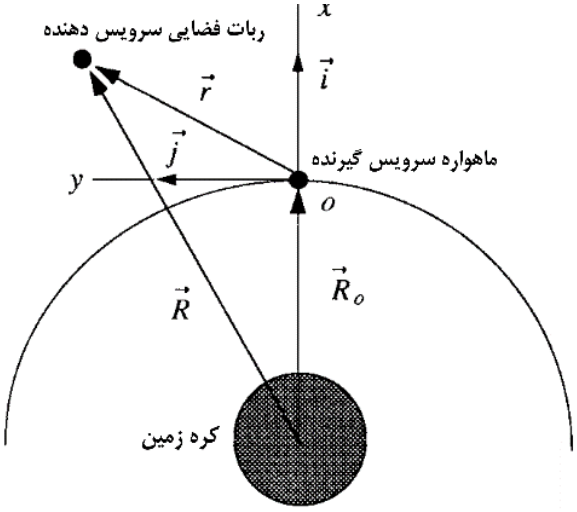
(یادداشت فنی)

طراحی مسیر بهینه مانورهای پهلوگیری خودکاربراساس یادگیری تقویتی Q و شبکه‌بندی مکعبی

معادله (۳) در دستگاه رفرنس (x, y, z) و بردارهای یکه $\{\vec{i}, \vec{j}, \vec{k}\}$ که در مرکز جرم آن تعریف شده در حال چرخش حول کرهٔ زمین است (شکل **۳**).

$$n = \sqrt{\mu / R_0^3} \quad (۲)$$

$$\vec{\omega} = n\vec{k} \quad (۳)$$



شکل (۳): موقعیت قرار گیری مکانی ربات فضایی سرویس‌دهنده و ماهواره سرویس‌گیرنده.

برای طراحی مسیر حرکت بهینه مانورهای پهلوگیری و اتصال میان دو جسم فضایی، حرکت نسبی ربات فضایی سرویس‌دهنده نسبت به ماهواره سرویس‌گیرنده براساس معادلات کلاسی–ویلشریا معادله هیل[1]تحلیل می‌شود که فرم خطی‌سازی شده آن در معادلات (۴) الی (۶) نمایش داده شده است [۱۷].

$$\ddot{x} - 2n\dot{y} - 3n^2 x = 0 \quad (۴)$$

$$\ddot{y} + 2n\dot{x} = 0 \quad (۵)$$

$$\ddot{z} + n^2 z = 0 \quad (۶)$$

که معادله خارج از صفحه (۶) از دو معادلهٔ دیگر گسسته شده و حالت ماتریسی پاسخ آن به صورت معادله (۷) نمایش داده می‌شود.

$$\begin{bmatrix} z(t) \\ \dot{z}(t) \end{bmatrix} = \begin{bmatrix} \cos n\,t & \sin nt/n \\ -n \sin nt & \cos nt \end{bmatrix} \begin{bmatrix} z_0 \\ \dot{z}_0 \end{bmatrix} \quad (۷)$$

همچنین، پاسخ کامل معادلات درون صفحه‌ای ((۵)–(۴)) نیز در معادله (۸) براساس ماتریس انتقال حالت، نمایش داده می‌شود.

$$\begin{bmatrix} x(t) \\ y(t) \\ \dot{x}(t) \\ \dot{y}(t) \end{bmatrix} = \begin{bmatrix} 4-3\cos nt & 0 & \sin nt/n & 2(1-cosnt)/n \\ 6\sin nt - 6nt & 1 & 2(-1+\cos nt)/n & 4\sin nt/n - 3t \\ 3n \sin nt & 0 & \cos nt & 2\sin nt \\ 6n(-1+\cos nt) & 0 & -2\sin nt & -3+4\cos nt \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ \dot{x}_0 \\ \dot{y}_0 \end{bmatrix} \quad (۸)$$

با واردسازی مقادیر اولیه مکان و سرعت ربات فضایی سرویس‌دهنده در معادلات معرفی شده به سادگی می‌توان مقادیر پارامترهای سرعت و مکان نسبی ربات فضایی را نسبت به ماهواره سرویس‌گیرنده براساس زمان طی شده محاسبه نمود. باتوجه به فاصله نزدیک میان ماهواره سرویس‌گیرنده و ربات فضایی سرویس‌دهنده در فاز اتصال، ربات فضایی سرویس‌دهنده می‌تواند با بهره‌گیری از دوربین‌های تصویر برداری متصل به خود به طور بر خط تصاویر دریافتی را پردازش نموده و محل قرارگیری ماهواره سرویس‌گیرنده را با سرعت و دقت بالا محاسبه نماید. براساس معادلات کلاسی–ویلشر موقعیت مکانی ماهواره سرویس‌گیرنده و ربات فضایی محاسبه می‌شود. در شکل **۳** تصویر بردار مکانی ماهواره سرویس‌گیرنده نسبت به کرهٔ زمین با $\vec{R}_0$ در معادله (۹) تعریف شده و بردار مکانی نسبی ربات فضایی سرویس‌دهنده نسبت به ماهواره سرویس‌گیرنده با $\vec{r}$ نمایش داده شده و به صورت معادله (۱۰) تعریف می‌شود.

$$\vec{R}_0 = R_0\vec{i} \quad (۹)$$

$$\vec{r} = x\vec{i} + y\vec{j} + z\vec{k} \quad (۱۰)$$

همچنین، بردار موقیت ربات فضایی سرویس‌دهنده نسبت به کرهٔ زمین نیز به صورت معادله (۱۱) تعریف می‌شود.

$$\vec{R} = (R_0 + x)\vec{i} + y\vec{j} + z\vec{k} \quad (۱۱)$$

پس از تعیین پارامتر مکان نسبی ربات فضایی سرویس‌دهنده نسبت به ماهواره سرویس‌گیرنده، می‌توان چارچوب قرار گیری دوربین متصل به بدنه ربات فضایی را با متغیر α نمایش داد که نقطه آغازین آن مرکز ربات فضایی بوده و توسط بردار موقعیت نسبی $\vec{r}$ به مرکز چارچوب قرار گیری ماهواره سرویس‌گیرنده β متصل شده است. در مسئله مورد بررسی در این مقاله فرض بر آن است که دو ماهواره در مرحله اتصال دارای سرعت زاویه‌ای و سرعت خطی برابری نسبت به یکدیگر بوده و همچنین هر دو جسم نسبت به یکدیگر درموقعیت موازی قرار گرفته‌اند. مختصات محل قرار گیری ربات فضایی (۰، ۰، ۰) می‌باشد و مختصات محل قرار گیری ماهواره سرویس‌گیرنده (۳، ۳، ۳) در نظر گرفته شده است. بنابراین، به جهت تکمیل مانور اتصال ربات فضایی تنها به صورت خطی و با سرعت ثابت در مختصات دستگاه محلی کارتزین (x, y, z) حرکت می‌نماید.

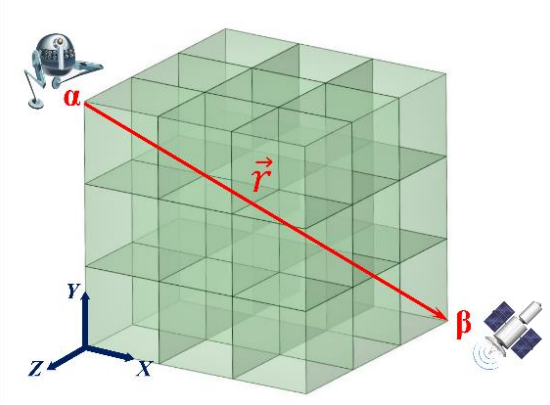
به جهت تعیین مسیرحرکت بهینه ربات فضایی، پس از محاسبهٔ مختصات نسبی محل قرار گیری ماهواره سرویس‌گیرنده، نقشه‌ای سه بعدی از مسیرهای قابل دسترسی براساس شبکه‌بندی مکعبی (شکل **۴**) محاسبه و طراحی می‌شود. بر این

1. Clohessy–Wiltshire or Hill's Equation

(یادداشت فنی)

ایمان شفیعی نژاد، محمد صیامی عراقی، علیرضا سخاوت، علی میرزایی و ایمان فزونی تلوکی

اساس راس و یال‌های مکعب‌های تشکیل شده میان ربات فضایی و ماهواره سرویس‌گیرنده مجموعه‌ای از راه‌های پیش‌بینی شده‌ای می‌باشند که ربات فضایی با طی کردن فاصله میان هر دو راس می‌تواند خود را به محل قرار گیری ماهواره سرویس‌گیرنده برساند. تبدیل مسیرهای قابل دسترس به شبکه‌بندی مکعبی سبب می‌شود تا مسئله به حالتی قابل حل توسط الگوریتم یادگیری کیو تبدیل شود. افزایش تراکم تعداد شبکه‌های مکعبی سبب افزایش دقت مسیر بهینه طراحی شده خواهد شد. اما از طرف دیگر سبب افزایش بار محاسباتی بر روی رایانه پردازشگر ربات فضایی می‌شود. به همین جهت، میزان تراکم شبکه‌های مکعبی می‌بایست با در نظر گرفتن دو عامل توان سخت افزاری در دسترس و دقت مورد نیاز به جهت طراحی مسیر بهینه انتخاب شود. میزان تراکم شبکه مکعبی با محاسبه میزان انحناهای لازم به جهت جلوگیری از برخورد ربات فضایی تعیین شود. توان سخت افزاری شامل حافظهٔ ذخیره سازی تصادفی[1](رم) و توان پردازشی محاسبه گر مرکزی[2](سی. پی. یو) می‌باشد.



شکل (۴): شبکه‌بندی مسیرهای اتصال.

۴-۲- نحوه عملکرد الگوریتم یادگیری کیو با ضریب یادگیری ثابت

به جهت شروع به کار الگوریتم یادگیری کیو پس از شبکه‌بندی مکعبی و تعیین مسیرهای در دسترس، ماتریس مربعی Z که رفتار الگوریتم را شکل می‌دهد امتیازبندی می‌شود. در این ماتریس مربعی ارتباط هر سطر و ستون نمایانگر نحوه دسترسی هر راس به راس‌های دیگر است که با توجه به سیاست حل مسئله امتیازبندی می‌شوند. در مسئله مورد بررسی در این مقاله هدف طی نمودن بهینه‌ترین مسیر از نظر مقدار مسافت براساس مسیرهای تعیین شدهٔ در دسترس است. بدین سبب امتیاز حرکت از هر راس به راس‌های قابل دسترس دیگر صفر در نظر گرفته شده است. همچنین، به جهت ممنوع کردن حرکت در جا در یک راس و نیز حرکت غیر مجاز به راس‌های تعریف نشده، امتیاز ۱- برای این حرکات مد نظر خواهد بود. علاوه‌براین، به جهت تشویق نمودن الگوریتم به پیدا نمودن مسیر بهینه به سمت ماهواره سرویس‌گیرنده، کلیهٔ راس‌های منتهی به آن دارای امتیاز مثبت ۱۰۰ هستند. مدت زمان و تعداد مراحل لازم به جهت همگرایی به پاسخ نهایی براساس میزان پیچیدگی در هر مسئله متفاوت می‌باشد، به همین دلیل الگوریتم یادگیری تقویتی با محاسبهٔ تعداد قدم‌های طی شده در هر مرحله تعداد قدم‌های طی شده به جهت حرکت ربات فضایی سرویس‌گیرنده به سمت ماهواره سرویس‌گیرنده را مورد شمارش قرار داده و در صورت تکراری شدن تعداد مسیرهای بهینه در مراحل قبل روند حل مسئله توسط الگوریتم متوقف شده و مسیر بهینه انتخاب می‌شود.

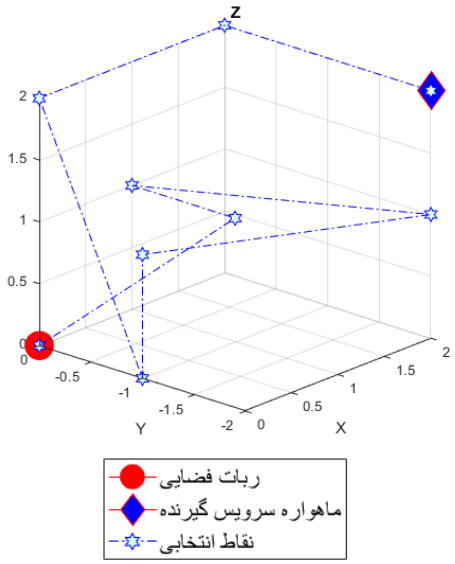
پس از تکمیل فرآیند امتیازدهی ماتریس مربعی Z، الگوریتم یادگیری کیو براساس مختصات به دست آمده از مبدا و مقصد، شروع به کسب تجربه براساس فرآیند سعی و خطا برای امتیازبندی و دسته‌بندی مسیرهای قابل حرکت براساس شبکه‌بندی‌های مکعبی انجام شده خواهد نمود. در ادامه امتیازهای به دست آمده از هر مرحله به جهت شکل دهی منطق الگوریتم در ماتریس مربعیکیو ذخیره می‌شود. یکی از عوامل مهم شکل‌دهی سیاست انتخاب مسیر بهینه در الگوریتم یادگیری کیو، تعیین ضریب یادگیری گاما در معادله (۱) است. این ضریب می‌تواند دارای مقداری بین ۰ تا ۱ باشد. در صورت انتخاب مقادیر نزدیک به ۱ الگوریتم بر روییافتن پاداش‌های لحظه‌ای تمرکز خواهد نمود و در صورت انتخاب مقادیر نزدیک به صفر الگوریتم با به تأخیر انداختن انتخاب پاداش‌های لحظه‌ای، روی یافتن پاداش‌های بزرگتر تمرکز می‌نماید. معیار بهینگی و تکمیل فرآیند الگوریتم ارضای شرایط پایانی و رسیدن به هدف است. بنابراین، می‌توان معادله (۱۲) را برای معیار بهینگی معرفی کرد.

$$J = min\{(\vec{R}_\alpha - \vec{R}_\beta) + (\vec{V}_\alpha - \vec{V}_\beta)\} \qquad (۱۲)$$

شکل ۵ نمایش‌دهنده نحوهٔ رفتار الگوریتم در یافتن مسیر بهینه قبل از شروع یادگیری یا مرحلهٔ صفر است. همان‌طورکه مشاهده می‌شود، مسیر انتخاب شده توسط الگوریتم دارای ۸ حرکت بوده که از منظر مقدار مسافت طی شده برای رسیدن به هدف بهینه نمی‌باشد. همچنین، با توجه به طی نشدن مراحل یادگیری توسط الگوریتم، قوائد حرکتی موجود در ماتریس Z که در ماتریس حافظه Q اعمال نگشته موجب انتخاب مسیرهای حرکتی غیر مجاز شده است.

1. Random Access Memory (RAM)

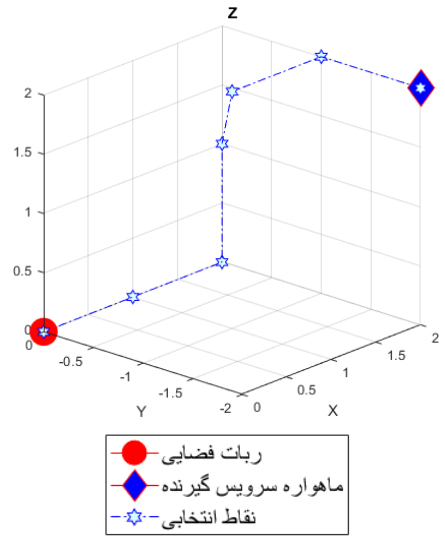
2. Central Processing Unit (CPU)



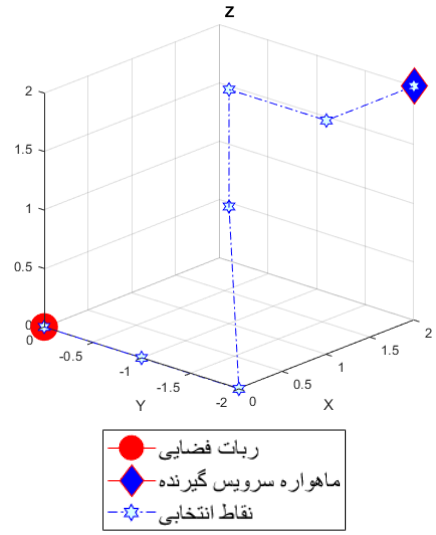
شکل (۵): نحوه اتصال ربات فضایی و ماهواره سرویس‌گیرنده بدون در نظرگرفتن یادگیری در مرحله صفر.

۴-۳- نحوه عملکرد الگوریتم یادگیری کیو با ضریب یادگیری متفاوت

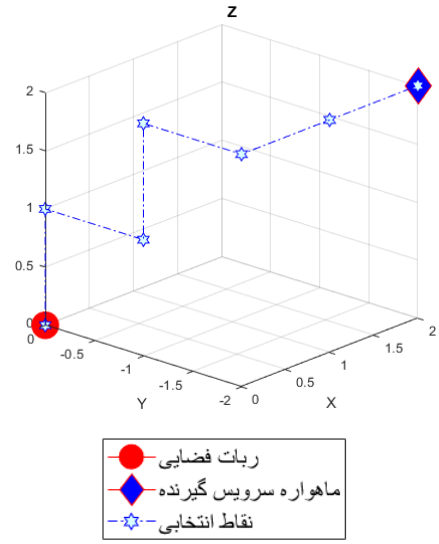
همان‌طورکه در بخش قبل اشاره شد، ضریب یادگیری گاما به طور مستقیم بر روی سیاست‌های تصمیم‌گیری الگوریتم یادگیری کیو تاثیرگذار است. به جهت بررسی نحوه رفتار الگوریتم یادگیری کیودر ضرایب یادگیری ۰.۱، ۰.۵ و ۰.۹ فرآیند تعیین مسیر بهینه برای ۳۰۰ مرحله یادگیری شبیه‌سازی شده و به ترتیب در شکل‌های **۶** الی **۸** نمایش داده شده است.



شکل (۶): نحوه اتصال ربات فضایی و ماهواره سرویس‌گیرنده با در نظرگرفتن یادگیری در ۳۰۰ مرحله و ضریب یادگیری ۰.۱.



شکل (۷): نحوه اتصال ربات فضایی و ماهواره سرویس‌گیرنده با در نظرگرفتن یادگیری در ۳۰۰ مرحله و ضریب یادگیری ۰.۵.



شکل (۸): نحوه اتصال ربات فضایی و ماهواره سرویس‌گیرنده با در نظرگرفتن یادگیری در ۳۰۰ مرحله و ضریب یادگیری ۰.۹.

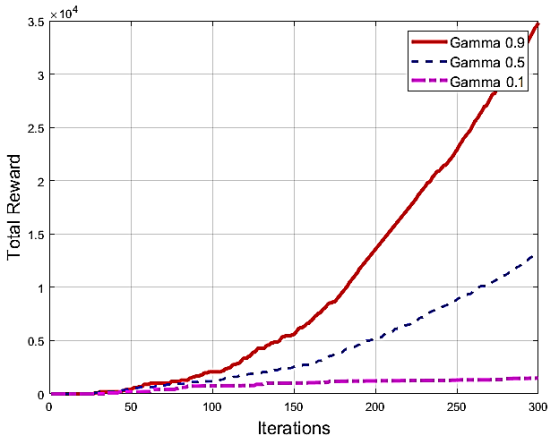
همان طور که در جدول **۱** نمایش داده شده است، الگوریتم یادگیری کیو در ضریب یادگیری ۰.۱ و ۰.۵ عملکرد مناسبی نداشته است. زیرا، اگر چه حداقل تعداد قدم برای رسیدن به ماهوار سرویس‌گیرنده طی شده است، اما در هر دو حالت یک حرکت خلاف قوانین نگارش شده در ماتریس Z بوده است که نشانگر کافی نبودن تعداد مراحل یادگیری به جهت همگرایی الگوریتم به پاسخ بهینه بوده است. در ضریب یادگیری ۰.۹ تعداد قدم‌های طی شده همانند دو ضریب دیگر در حالت بهینه‌است، اما به سبب نزدیک تر بودن مقدار ضریب یادگیری انتخاب شده به عدد ۱ در مجموع میزان پاداش بیشتری در طول مسیریادگیر جمع‌آوری شده است که منجربه همگرایی الگوریتم به پاسخ بهینه و انتخاب

قدم‌های صحیح براساس قوانین تدوین شده در ماتریس Z بوده است.

جدول(۱): مقایسه ضرایب مختلف یادگیری.

تعداد قدم‌های خلاف قانون	تعداد قدم‌های صحیح	تعداد قدم‌های طی شده	ضریب یادگیری
۱	۶	۷	۱.۰
۱	۶	۷	۵.۰
۰	۷	۷	۹.۰

در شکل **۹** می‌توان مشاهده نمود که ضریب یادگیری به طور مستقیم روی عملکرد الگوریتم در کسب پاداش تاثیر مستقیم داشته است. بر این اساس، نمودار مربوط به ضریب یادگیری ۰.۹ بیشترین میزان پاداش را در طی ۳۰۰ مرحله یادگیری خود کسب نموده است. در حالی الگوریتم یادگیری تقویتی در ضرایب یادگیری ۰.۵ و ۱.۰ به ترتیب میزان پاداش کم تری را در طول مراحل یادگیری کسب نموده‌اند که نشانگر بازدهی کمتر آنها در یافتن مسیر بهینه در مسئله مورد بررسی است.



شکل(۹): مقایسه تاثیر ضرایب یادگیری در جمع پاداش‌های کسب شده در طول دفعات یادگیری.

۵- نتیجه‌گیری

در این مقاله نحوه عملکرد الگوریتم یادگیری کیو در طراحی مسیر بهینه در فاز پهلوگیری و اتصال میان ربات فضایی و ماهواره سرویس‌گیرنده تحت عنوان سرویس درون‌مداری مورد بررسی قرار گفته است. براساس نتایج مدلسازی‌های انجام شده می‌توان نتیجه‌گیری کرد که استفاده از روش شبکه‌بندی مکعبی به جهت مدل سازی مسیرهای دردسترس میان ربات فضایی و ماهواره سرویس‌گیرنده، موجب ساده‌سازی و قانونمند نمودن فرآیند طراحی ماتریس Z یا ماتریسپاداش شده است. همچنین، با استفاده از روش شبکه‌بندی مکعبی می‌توان از الگوریتم یادگیری کیو به جهت یافتن مسیرهای بهینه حرکت در فضای سه بعدی استفاده نمود. از طرف دیگر، استفاده از الگوریتم یادگیری کیو سبب می‌شود تا ربات فضایی بدون راهنمایی ایستگاه زمینی و تنها براساس تعامل با محیط اطراف بتواند تصمیمات لازم را به جهت تکمیل مانورهای اتصال و پهلوگیری اتخاذ نماید. علاوه بر این، می‌توان بیان کرد که ضریب یادگیری تاثیر مستقیمی بر کنترل سیاست عملکردی الگوریتم یادگیری کیو دارد و می‌بایست براساس اهداف مأموریت تنظیم شود. براساس نتایج این مقاله، مقدار ضریب یادگیری می‌بایست با انجام فرایندهای شبیه‌سازی و به صورت تجربی انتخاب شود.

۶- مراجع

[1] *Satellite Database | Union of Concerned Scientists* Available: https://www. ucsusa. org/resources/satellite-database

[2] G. A. Landis, S. G. Bailey, and R. Tischler, "Causes of power-related satellite failures," in *2006 IEEE 4th World Conference on Photovoltaic Energy Conference*, 2006, pp. 1943-1945.

[3] A. Ellery, J. Kreisel, and B. Sommer, "The case for robotic on-orbit servicing of spacecraft: Spacecraft reliability is a myth," *Acta Astronautica*, vol. 63, pp. 632-648, 2008.

[4] M. Tafazoli, "A study of on-orbit spacecraft failures," *Acta Astronautica*, vol. 64, pp. 195-205, 2009.

[5] F. Sellmaier, T. Boge, J. Spurmann, S. Gully, T. Rupp, and F. Huber, "On-orbit servicing missions: Challenges and solutions for spacecraft operations," in *SpaceOps 2010 Conference Delivering on the Dream Hosted by NASA Marshall Space Flight Center and Organized by AIAA*, 2010, p. 2159.

[6] E. Stoll, U. Walter, J. Artigas, C. Preusche, P. Kremer, G. Hirzinger, *et al.* , "Ground verification of the feasibility of telepresent on-orbit servicing," *Journal of Field Robotics*, vol. 26, pp. 287-307, 2009.

[7] Y. Wang, Z. Ma, Y. Yang, Z. Wang, and L. Tang, "A new spacecraft attitude stabilization mechanism using deep reinforcement learning method," in *8th European Conference for Aeronautics and Space Sciences (EUCASS)*, 2019.

[8] S. Willis, D. Izzo, and D. Hennes, "Reinforcement learning for spacecraft maneuvering near small bodies," in *AAS/AIAA Space Flight Mechanics Meeting*, 2016, pp. 14-18.

[9] M. Hatem and F. Abdessemed, "Simulation of the Navigation of a Mobile Robot by the QLearning using Artificial Neuron Networks," in *CIIA*, 2009.

[10] W. Adiprawita, A. S. Ahmad, J. Sembiring, and B. R. Trilaksono, "Simplified Q-learning for holonomic mobile robot navigation," in *2011 2nd International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering*, 2011, pp. 64-68.

[11] H. Wicaksono, K. Anam, P. Prihastono, I. A. Sulistijono, and S. Kuswadi, "COMPACT FUZZY Q LEARNING FOR AUTONOMOUS MOBILE ROBOT NAVIGATION," 2014.

[12] L. Khriji, F. Touati, K. Benhmed, and A. Al-Yahmedi, "Mobile robot navigation based on Q-learning technique," *International Journal of Advanced Robotic Systems*, vol. 8, p. 4, 2011.

[13] Y. Duan, "Fuzzy reinforcement learning and its application in robot navigation," in *2005 International Conference on Machine Learning and Cybernetics*, 2005, pp. 899-904.

[14] J. R. Wertz and R. Bell, "Autonomous rendezvous and docking technologies: status and prospects," *Space Systems Technology and Operations*, vol. 5088, pp. 20-30, 2003.

[15] C. J. Dennehy and J. R. Carpenter, "A summary of the rendezvous, proximity operations, docking, and undocking (rpodu) lessons learned from the defense advanced research project agency (darpa) orbital express (oe) demonstration system mission," 2011.

[16] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*: MIT press, 2018.

[17] G. Pollock, J. Gangestad, and J. Longuski, "Analysis of Lorentz spacecraft motion about Earth using the Hill-Clohessy-Wiltshire equations," in *AIAA/AAS astrodynamics specialist conference and exhibit*, 2008, p. 6762.