

طراحی کنترل کننده تطبیقی بهینه بدون مدل برای پیکربندی مجدد آرایش پروازی ماهواره‌ها با یادگیری تقویتی

تاریخ دریافت: ۱۴۰۰/۱۲/۱۶

تاریخ پذیرش: ۱۴۰۱/۰۴/۰۹

محمد رسول کتکاشور^۱، حسین بلندی^{۲*}، ناصر مزینی^۳

۱- دانشجوی دکتری، دانشکده برق، دانشگاه علم و صنعت ایران، تهران

۲- دانشیار، دانشکده برق، دانشگاه علم و صنعت ایران، تهران. h_bolandi@iust.ac.ir

۳- دانشیار، دانشکده کامپیوتر، دانشگاه علم و صنعت ایران، تهران

چکیده

در این مقاله یک کنترل کننده تطبیقی بهینه بدون مدل برای پیکربندی مجدد آرایش پروازی ماهواره‌ها ارائه می‌شود. پیکربندی مجدد آرایش پروازی ماهواره‌ها، قابلیت مهمی برای دستیابی به اهداف مأموریت‌های آرایش پروازی ماهواره‌هاست. به دلیل پیچیدگی، استخراج یک مدل ریاضی دقیق و حضور اغتشاشات مداری و نامعینی‌ها طراحی یک کنترل کننده بهینه امری دشوار است. در این کار، ابتدا یک کنترل کننده بهینه بر مبنای یک تابع هزینه تعمیم یافته تخفیف یافته استخراج می‌شود. سپس پایداری آن با استفاده از روش لیاپانوف به اثبات می‌رسد. این کنترل کننده برای آنکه قابلیت پیاده‌سازی روی رایانه‌های ماهواره را داشته باشد، به صورت دیجیتالی طراحی شده است. سپس با استفاده از روش‌های یادگیری تقویتی، الگوریتمی برای حل مسئله ارائه می‌شود. این الگوریتم قادر است تا مسئله ردیابی پیکربندی مجدد آرایش پروازی ماهواره‌ها را به صورت برخط و بدون نیاز به مدل حل کند. در انتها، کارایی روش پیشنهادی در یک سناریو مأموریتی پیکربندی مجدد آرایش پروازی ماهواره‌ها، صحت‌گذاری می‌شود.

واژه‌های کلیدی: آرایش پروازی ماهواره‌ها، یادگیری تقویتی، کنترل تطبیقی بهینه، کنترل بدون مدل، یادگیری-کیو، سیستم چند عامله

Model-free adaptive optimal control of spacecraft formation flying reconfiguration using Q-Learning

MohammadRasoul Kankashvar¹, Hossein Bolandi², Naser Mozayani³

1 PhD Student, Electrical Engineering Department, Iran University of Science and Technology, Tehran

2 Associate Professor, Electrical Engineering Department, Iran University of Science and Technology, Tehran

3 Associate Professor, Computer Engineering Department, Iran University of Science and Technology, Tehran

Abstract

This paper investigates an optimal adaptive controller based on reinforcement learning while considering orbital perturbations. The controller can achieve mission goals, online without any model. Reconfiguration capabilities provide great flexibility in achieving formation flying mission goals. In reconfiguration, it is desired that spacecrafts migrate from the current formation to a new formation, thus achieving mission goals. Orbital perturbations, difficulties in extracting exact mathematical models, and unknown system dynamics make the optimal reconfiguration problem challenging. Due to the digital nature of spacecraft computer systems, controllers have to be implemented digitally. Accordingly, this paper introduces an adaptive optimal digital controller for a discounted generalized cost function. The stability of the proposed controller is proven by the Lyapunov method. Then, using the Q-learning method, an algorithm is presented so that the controller can find the optimal control gains in a model-free fashion. Finally, numerical simulations of a formation flying mission scenario, confirm the effectiveness of this method.

Keywords: Reinforcement learning; spacecraft formation flying; Q-learning; optimal adaptive control; multi-agent systems

۴۱

سال ۱۱- شماره ۱

بهار و تابستان ۱۴۰۱

نشریه علمی دانش و

فناوری هوا فضا



۱. مقدمه

آرایش پروازی ماهواره‌ها رویکردی بسیار جذاب برای طراحی مأموریت‌های فضایی است. در این رویکرد امکان اجرای مأموریت‌هایی با قابلیت بیشتر، منعطف‌تر (تعداد ماهواره‌ها، ابعاد و شکل آرایش، تغییر مأموریت) و مقاوم‌تر (قابلیت اطمینان بیشتر، تحمل خطا و خرابی ماهواره‌ها) نسبت به تک ماهواره وجود دارد [1]-[3]. پیکربندی مجدد^۱ آرایش ماهواره‌ها یکی از قابلیت‌های کلیدی است که انعطاف‌پذیری مأموریت را افزایش می‌دهد. پیکربندی مجدد آرایش ماهواره‌ها یعنی بازنشانی ماهواره‌ها در موقعیت‌های جدیدی که به برآورده شدن نیازهای مأموریت منجر خواهد شد.

برای پیکربندی مجدد ماهواره‌ها الزامات زیادی وجود دارد که باید برقرار شوند. یکی از این الزامات، می‌تواند بهینه‌سازی پیکربندی مجدد ماهواره‌ها باشد. این بهینه‌سازی باید در حضور عواملی مثل نامعینی‌ها، عدم قطعیت‌های پارامترهای مداری و اغتشاشات مداری صورت بپذیرد [2]، [3]. تاکنون دربارهٔ پیکربندی مجدد ماهواره‌ها مطالعات زیادی انجام شده است [4]-[9]. این مطالعات به‌صورت مجزا به مسئله بهینه‌سازی و مسئله نامعینی‌ها و اغتشاشات پرداخته‌اند.

روش‌های بهینه‌سازی مختلفی برای حل مسئله پیکربندی مجدد آرایش پروازی ماهواره‌ها ارائه شده است. در [10] با استفاده از روش برنامه‌ریزی-مربعی^۲ یک کنترل‌کننده سوخت بهینه طراحی شده است. روش برنامه‌ریزی مختلط-صحیح-خطی^۳ در [7] برای بهینه‌سازی مصرف سوخت بر اساس مدل‌سازی المان‌های

نسبی مداری^۴ موردبررسی قرار گرفته است. در [11] یک رویکرد کنترل بهینه مبتنی بر روش معادله‌های ریکاتی وابسته به حالت^۵ پیشنهاد شده است. با استفاده از روش بهینه‌سازی محدب^۶ در [12]، مسئله تغییر پیکربندی ماهواره‌ها در آرایش پروازی با قابلیت عدم برخورد^۷ مطالعه شده است. در [13] با استفاده از روش‌های تکاملی^۸ یک کنترل‌کننده زمان بهینه طراحی شده است. در بیشتر مقاله‌های مبتنی بر کنترل بهینه مسئله نامعینی‌ها، اغتشاشات و عدم قطعیت‌ها مورد مطالعه قرار نگرفته است. طراحی کنترل‌کننده بهینه با وجود این شرایط، چالش‌برانگیز است.

همچنین روش‌های زیادی برای غلبه بر اغتشاشات و نامعینی‌های مداری در چند سال اخیر پیشنهاد شده است [14]-[18]. در [14] یک کنترل‌کننده تطبیقی در حضور نامعینی در سرعت و جرم برای کنترل آرایش پروازی ماهواره‌ها در نقاط لاگرانژ^۹ طراحی شده است. برای کنترل شش درجه آزادی آرایش پروازی ماهواره‌ها در حضور اغتشاشات مداری و کشش اتمسفری^{۱۰} یک کنترل‌کننده تطبیقی مقاوم در [15] طراحی شده است. به‌طور مشابه نویسندگان در [16] توانستند یک کنترل‌کننده مقاوم توزیع‌یافته^{۱۱} در حضور نامعینی و اغتشاشات طراحی کنند. با در نظر گرفتن فرض عدم برخورد ماهواره‌ها، در [17] یک کنترل‌کننده مقاوم ردیاب حالت با کمک میدان پتانسیل مجازی^{۱۲} در حضور اغتشاشات خارجی طراحی شده است. در [18] با کمک یک رویکرد حالت، یک کنترل‌کننده ردیاب حالت با قابلیت دفع اغتشاش برای آرایش پروازی ماهواره‌ها طراحی شده است. با بررسی دقیق‌تر مقاله‌های منتشرشده در حوزه کنترل مقاوم و تطبیقی می‌توان مشاهده کرد که پاسخ‌های



به دست آمده بهینه نیستند.

رویکردی دیگر که برای پیکربندی مجدد ماهواره‌ها در آرایش پروازی استفاده می‌شود، رویکردهای تحلیلی و مدل‌سازی است. در [19] با استفاده از روش دینامیک معکوس^{۱۳}، یک راه‌حل تحلیلی برای پیکربندی مجدد آرایش پروازی ماهواره‌ها پیشنهاد شده است. با رویکردی مشابه، در [20] با استفاده از روش‌های تحلیلی و المان‌های نسبی مداری یک پاسخ تحلیلی برای مانورهای بهینه آرایش پروازی ماهواره پیشنهاد شده است. در [21] روشی تحلیلی برای پیدا کردن مسیرهایی با مصرف بهینه در حضور اغتشاشات مداری و اثر ناهمگونی زمین پیشنهاد شده است. در [22] محققان توانسته‌اند با استفاده از المان‌های نسبی مداری، معادله ریگاتی غیرخطی را برای تجزیه و تحلیل بهتر هدایت، کنترل و ناوبری آرایش پروازی ماهواره‌ها پیشنهاد دهند. مشکلی که در همه روش‌های تحلیلی وجود دارد آن است که هیچ تضمینی برای پایداری کنترل‌کننده وجود ندارد. همچنین روش‌های تحلیلی بسیار به مدل وابسته هستند و دقت راه‌حل‌های استخراج‌شده دقیقاً به دقت مدل در نظر گرفته شده است. در این روش‌ها هیچ تضمینی برای مقابله با نامعینی‌ها و اغتشاشات ارائه نمی‌شود. همچنین بهینه بودن پاسخ‌ها اثبات نشده باقی می‌ماند.

با مرور مطالب و منابع مطرح‌شده می‌توان این نتیجه را گرفت که طراحی کنترل‌کننده‌ای که بتواند در حضور نامعینی‌ها و اغتشاشات مداری، پاسخ‌های بهینه را ارائه کند، قدم بعدی در طراحی کنترل‌کننده مناسب برای حل مسئله آرایش پروازی ماهواره‌ها خواهد بود.

آموزش یادگیری-کیو^{۱۴}، یکی از روش‌های

یادگیری تقویتی^{۱۵} است که در [23] ارائه شده است. یادگیری تقویتی خود نیز یکی از روش‌های یادگیری ماشین^{۱۶} است [24]. با استفاده از این روش می‌توان پاسخ بهینه را بدون داشتن اطلاعات و مدل سیستم به دست آورد. این روش در جامعه علمی هوش مصنوعی بسیار مطالعه شده است [24]–[28]. از این روش برای کنترل تطبیقی بهینه سیستم‌های کنترلی نیز استفاده شده است [29]–[36]. تاکنون از روش‌های یادگیری ماشین برای کنترل انواع مأموریت‌های ماهواره‌ها استفاده شده است. در [37] یک کنترل‌کننده مبتنی بر یادگیری تقویتی برای کنترل وضعیت تک ماهواره در حضور نامعینی‌ها ارائه شده است. با رویکرد مشابهی در [38] یک کنترل‌کننده مبتنی بر یادگیری تقویتی برای تنظیم پارامترهای یک کنترل‌کننده تطبیقی خودتنظیم^{۱۷} استفاده شده است. برای حل مسئله متصل شدن^{۱۸} دو ماهواره و ملاقات^{۱۹} فضایی ماهواره‌ها، یک کنترل‌کننده یادگیری تقویتی در [39] ارائه شده است. به‌طور مشابه در [40] با استفاده از روش یادگیری-کیو به حل مسئله پهلوگیری و الحاق فضایی پرداخته شده است. در [41] با استفاده از روش مدل-مینا^{۲۰} یک کنترل‌کننده مبتنی بر یادگیری تقویتی برای کنترل زیر بهینه پیکربندی مجدد آرایش پروازی ماهواره‌ها پیشنهاد شده است.

متأسفانه، در اکثر مقاله‌های یادگیری تقویتی، محققان اغلب به مسئله آنالیز پایداری نمی‌پردازند [37], [39], [40], [42]. همچنین در طراحی کنترل‌کننده‌های ردیاب خود ضریب تخفیف^{۲۱} را در تابع هزینه در نظر نمی‌گیرند [37]–[41]. [43]. ضریب تخفیف به همگرایی سریع‌تر کمک می‌کند، همچنین یک پارامتر مهم در حل مسائل





ردیابی است [35]. در صورتی که این ضریب در نظر گرفته نشود، در صورت محدود نبودن سیگنال مسیر مرجع^{۲۲}، ممکن است سیگنال کنترلی نامحدود شود.

در این مقاله یک کنترل کننده تطبیقی بهینه مبتنی بر یادگیری تقویتی ارائه می شود که می تواند بر خلاف [41]، مسئله پیکربندی مجدد آرایش پروازی را بدون داشتن مدل سیستم به صورت بهینه حل کند. چون کنترل کننده بدون مدل طراحی شده است، اثرات نامعینی ها و اغتشاشات مداری را به صورت خودکار در نظر می گیرد. نویسندگان این مقاله با بررسی هایی که انجام داده اند، استفاده از چنین کنترل کننده ای را در حل مسئله پیکربندی مجدد آرایش پروازی ماهواره، تاکنون در جایی مشاهده نکرده اند.

از سوی دیگر، در این مقاله، ضریب تخفیف در تابع هزینه اعمال شده است. به این ترتیب ضعف هایی که در [43]، [41]-[37] ناشی از عدم اعمال ضریب تخفیف بوده است، پوشش داده می شود.

یکی دیگر از جنبه های در نظر گرفته شده در این مقاله، قابلیت پیاده سازی و عملیاتی کردن کنترل کننده طراحی شده است. با توجه به اینکه کامپیوترهای سیستم های ماهواره به صورت دیجیتال کار می کنند، کنترل کننده طراحی شده نیز به صورت دیجیتالی طراحی شده است. در حل مسئله کنترل بهینه به روش برنامه ریزی پویا، ضرایب کنترل کننده از پیش محاسبه شده و در حافظه ذخیره می شوند تا در حین کار به سیستم اعمال شوند. بنابراین اغلب روش های کنترل بهینه روشی هایی برون-خط^{۲۳} هستند [44]. در این مقاله کنترل بهینه به روشی بر-خط^{۲۴} استخراج می شود و کنترل کننده، اطلاعات مورد نیاز خود را

در طول فرایند پیکربندی مجدد به صورت زمان-حقیقی^{۲۵} در حین کار از سیستم جمع آوری می کند. به این ترتیب نیازی به اطلاعات از پیش ساخته شده یا دخالت ایستگاه زمینی نیست و کنترل کننده کاملاً خودبسنده خواهد بود.

بر اساس آنچه گفته شد، در این مقاله ابتدا مسئله پیکربندی مجدد ماهواره مطرح می شود. سپس یک تابع هزینه جدید تخفیف یافته^{۲۶} ارائه می شود. تابع هزینه تخفیف یافته ارائه شده برای نخستین بار به صورت تعمیم یافته^{۲۷} نیز در نظر گرفته شده است تا بتواند نیازهای طراح را به شکل عمومی تری در نظر بگیرد. سپس بر مبنای تابع هزینه طراحی شده، کنترل کننده بهینه معرفی می شود و اثبات پایداری آن ارائه می شود. در ادامه این روند، با استفاده از یادگیری-کیو ضرایب کنترل کننده بهینه به طوری استخراج می شوند تا دیگر نیازی به مدل نباشد. در نهایت برای بررسی اثربخشی این کنترل کننده یک سناریو پیکربندی مجدد از آرایش پروازی ماهواره ها در نظر گرفته می شود و اثربخشی کنترل کننده مورد بحث مورد بررسی قرار می گیرد.

۲. بیان مسئله

۲-۱. مدل سازی موقعیت نسبی ماهواره ها

روش های مختلفی برای مدل سازی آرایش پروازی ماهواره ها وجود دارد [45]. یکی از روش های مرسوم استفاده از موقعیت نسبی $\rho = [x, y, z]$ و سرعت نسبی $\dot{\rho} = [\dot{x}, \dot{y}, \dot{z}]$ ماهواره پیرو^{۲۸} در دستگاه چرخان منطبق بر مرکز ماهواره پیشرو^{۲۹} است. این دستگاه چرخان را با نام قائم-محلی-افق-محلی (LVLH^{۳۰})، یا دستگاه اوپلر-هیل^{۳۱} نیز می شناسند. در این دستگاه محور x در راستای شعاعی زمین، محور z در راستای

می‌توان معادله (۱) را در نهایت به شکل زیر خطی‌سازی کرد [46]:

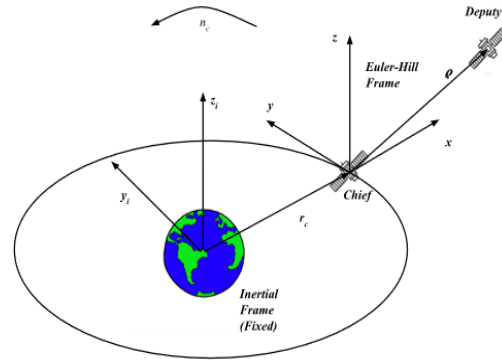
$$\begin{aligned} \ddot{x} - 2cn_c\dot{y} - (5c^2 - 2)n_c^2x & \quad (2) \\ & = -3n^2J_2\left(\frac{R_e^2}{r_c}\right) \\ & \times \left\{ \frac{1}{2} \right. \\ & - \left[\frac{3\sin^2 i_c \sin^2(kt)}{2} \right] \\ & \left. - \left[\frac{1 + 3\cos 2i_c}{8} \right] \right\} \end{aligned}$$

$$\begin{aligned} \ddot{y} + 2cn_c\dot{x} & \\ = -3n^2J_2\left(\frac{R_e^2}{r_c}\right)\sin^2 i_c \sin(kt) \cos(kt) & \quad \ddot{z} \\ + q^2z = 2lq\cos(qt + \phi) & \end{aligned}$$

که در آن شعاع زمین و i_c شیب مداری ماهواره پیشرو است. نحوه محاسبه ثابت‌های q, l, k, ϕ و c در [46] به تفصیل توضیح داده شده است. با توجه به اینکه در این مقاله نیازی نیست که این عوامل مستقیماً حساب شوند، از بیان جزئیات نحوه محاسبه این پارامترها اجتناب شده است. در ادامه این مقاله فرض می‌شود که مدار ماهواره پیشرو به صورت دایروی است.

برای آنکه دامنه مدار نسبی ماهواره پیرو نسبت به ماهواره پیشرو محدود باشد، لازم است تا روی شرایط اولیه محدودیت‌های برقرار کرد. در صورتی که این شروط برآورده نشوند، دامنه حرکت ماهواره پیرو نسبت به ماهواره پیشرو محدود نخواهد بود و با گذشت زمان ماهواره‌ها از یکدیگر فاصله می‌گیرند. این شرط، که به آن شرط تطبیق انرژی می‌گویند، در [3] به تفصیل بیان شده است. در صورت برآورده نشدن این شرط، به هنگام تغییر پیکربندی ماهواره‌ها در آرایش پروازی، ماهواره‌ها به تدریج از یکدیگر فاصله گرفته و برای بازنشانی مجدد ماهواره‌ها در پیکربندی باید انرژی/سوخت زیادی مصرف شود.

عمود بر صفحه مداری و محور y به گونه‌ای تعریف می‌شود که این دستگاه راستگرد باشد. شکل ۱ این دستگاه را در دستگاه اینرسی زمین مرکز (ECI^{۳۲}) نمایش می‌دهد.



شکل ۱. دستگاه LVLH و دستگاه ECI

معادله‌های دینامیکی موقعیت نسبی ماهواره را می‌توان به صورت زیر نوشت [3]:

$$\begin{aligned} \ddot{x} - 2\dot{f}_c\dot{y} - \dot{f}_c^2y - \dot{f}_c^2x & \\ = -\frac{\mu(r_c + x)}{[(r_c + x)^2 + y^2 + z^2]^{3/2}} + \frac{\mu}{r_c^2} & \\ + d_x & \\ \ddot{y} + 2\dot{f}_c\dot{x} + \dot{f}_c^2x - \dot{f}_c^2y & \\ = -\frac{\mu y}{[(r_c + x)^2 + y^2 + z^2]^{3/2}} + d_y & \quad (1) \\ \ddot{z} = -\frac{\mu z}{[(r_c + x)^2 + y^2 + z^2]^{3/2}} & \\ + d_z & \end{aligned}$$

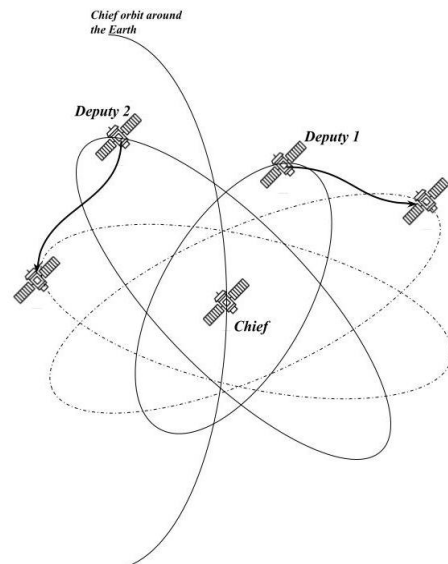
$$\ddot{r}_c = r_c\dot{f}_c^2 - \frac{\mu}{r_c^2}, \quad \dot{f}_c = -2\frac{\dot{r}_c\dot{f}_c}{r_c}$$

که در آن، f_c آنومالی حقیقی^{۳۳}، r_c اندازه شعاعی ماهواره پیشرو از مرکز زمین، μ ثابت گرانشی^{۳۴} و $d = [d_x, d_y, d_z]$ نشان‌دهنده اغتشاشات مداری است. معادله‌های غیرخطی (۱) را می‌توان با در نظر گرفتن فرض‌های مناسبی خطی‌سازی کرد [46]. با فرض اینکه ماهواره پیشرو در یک مدار دایروی قرار دارد، می‌توان تغییرات آنومالی حقیقی را ثابت در نظر گرفت و آن را معادل متوسط حرکت مداری^{۳۵}، n_c فرض کرد. با در نظر گرفتن اثر ناهمگونی زمین، J_2



۲-۲. پیکربندی مجدد بهینه آرایش پروازی

اگرچه موقعیت نسبی ماهواره پیرو نسبت به ماهواره پیشرو اطلاعات مفیدی را در اختیار مهندسان قرار می‌دهد، اما برای طراحی پیکربندی آرایش پروازی و پیکربندی مجدد آن‌ها، داشتن موقعیت نسبی ماهواره پیرو نسبت به مدار نسبی هدف در دستگاه ماهواره پیشرو الزامی است [46]. بر این مبنا فرض کنید که تعداد p ماهواره در یک مأموریت آرایش پروازی ماهواره‌ها حضور دارند و $X_i = [\rho_i \quad \dot{\rho}_i]$ بردار حالت توصیف‌کننده موقعیت نسبی و سرعت نسبی پیرو ماهواره i ام در آرایش پروازی است. در مسئله پیکربندی مجدد آرایش پروازی ماهواره لازم است تا ماهواره پیرو از مدار نسبی کنونی خود به مدار نسبی هدف مهاجرت کند و در آن باقی بماند. این پیکربندی مجدد در شکل ۲ نشان داده شده است.



شکل ۲. پیکربندی مجدد ماهواره‌ها، مدار کنونی با خط تیره و مدارهای هدف با خطچین نشان داده شده‌اند.

فرض کنید که حالت‌های مدار هدف با X_{i_d} نشان داده شده است. به‌این ترتیب، خطای بین

مدار کنونی با مدار هدف با استفاده از رابطه $e_i = X_i - X_{i_d}$ قابل محاسبه است. در نهایت می‌توان معادله‌های دینامیک خطای حاکم بر مسئله را در دستگاه هیل به صورت زیر نوشت:

$$\dot{e}_i(t) = A_c e_i(t) + B_{c_i} u_i(t) + D_i, \quad i = 1, 2, 3, \dots, N \quad (3)$$

که در آن،

$$A_c = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ (5c^2 - 2)n_c & 0 & 0 & 0 & 2cn_c & 0 \\ 0 & 0 & 0 & -2n_c & 0 & 0 \\ 0 & 0 & -q^2 & 0 & 0 & 0 \end{bmatrix}, \quad B_{c_i} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad D_i = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 2l_i q_i \cos(q_i t + \phi_i) \end{bmatrix}$$

در این معادله‌ها $u_i = [u_{x_i}, u_{y_i}, u_{z_i}]$ نیروی

کنترلی اعمالی بر ماهواره پیرو است. با فرض اینکه ماهواره‌ها به همدیگر نزدیک‌اند و تقریباً شیبی مداری دارند، می‌توان فرض کرد که ضریب l_i برابر با صفر است [46]. به‌این ترتیب معادله (۳) را می‌توان برای هر ماهواره حاضر در آرایش پروازی به شکل زیر نوشت:

$$\dot{e}_i(t) = A_c e_i(t) + B_{c_i} u_i(t), \quad i = 1, 2, \dots, N \quad (5)$$

که در آن N تعداد کل ماهواره‌های حاضر در آرایش است. در این معادله‌ها هدف کنترلی آن است که خطا به سمت صفر میل کند.

$$\lim_{t \rightarrow \infty} e_i(t) = 0 \quad (6)$$

۲-۳. تابع هزینه پیکربندی مجدد آرایش پروازی ماهواره‌ها

تابع هزینه در نظر گرفته شده برای ماهواره i ام به صورت زیر تعریف شده است:



۳-۱. تابع هزینه تعمیم یافته تخفیف یافته

برای حل مسئله کنترل پیکربندی مجدد آرایش پروازی ماهواره‌ها لازم است تا یک کنترل کننده ردیاب بهینه طراحی شود. برای این منظور برای اولین بار، تابع هزینه تعمیم یافته تخفیف یافته^{۴۰} یا تابع ارزش^{۴۱} برای ماهواره i ام به صورت زیر پیشنهاد می‌شود:

$$\begin{aligned} V_i(e_i(k)) &= \frac{1}{2} \sum_{j=k}^{\infty} \gamma_i^{j-k} U_i(e_i(j), u_i(j)) \\ &= \frac{1}{2} \sum_{j=k}^{\infty} \gamma_i^{j-k} [e_i(j)^T Q_i e_i(j) \\ &\quad + u_i(j)^T R_i u_i(j) \\ &\quad + 2e_i^T(j) N_i u_i(j)] \end{aligned} \quad (10)$$

در این معادله‌ها $0 < \gamma_i \leq 1$ ضریب تخفیف است. این ضریب هزینه را با گذشت زمان کاهش می‌دهد. ماتریس‌های وزنی R_i ، Q_i و N_i ماتریس‌هایی با ابعاد مناسب هستند. همچنین فرض می‌شود که $\det \begin{pmatrix} Q_i & N_i \\ N_i^T & R_i \end{pmatrix} > 0$ و $R_i > 0$ و $Q_i \geq 0$.

بر اساس قاعده بهینه‌سازی بلمن^{۴۲} می‌توان قانون کنترلی را به صورت زیر نوشت [47]:

$$u_i(k) = K_i e_i(k) \quad (11)$$

لم ۱- برای هر قانون کنترلی معادله (۱۱)، می‌توان نشان داد که تابع ارزش معادله (۱۰) را می‌توان به شکل مربعی زیر نوشت:

$$V_i(e_i(k), u_i(k)) = \frac{1}{2} e_i(k)^T P_i e_i(k) \quad (12)$$

که در آن P_i یک ماتریس مثبت معین است.

اثبات- با جاگذاری معادله (۱۱) در معادله (۱۰) می‌توان رابطه (۱۳) را به دست آورد. با استفاده از قانون کنترلی معادله (۱۱) در معادله (۸) می‌توان پاسخ را برحسب هر مقدار اولیه $e_i(k)$ به شکل رابطه (۱۴) نوشت.

$$\begin{aligned} J_i(t) &= \frac{1}{2} \int_{t_0}^{\infty} U_i(e_i(t), u_i(t)) dt \\ &= \frac{1}{2} \int_{t_0}^{\infty} e_i(t)^T Q_i e_i(t) \\ &\quad + u_i(t)^T R_i u_i(t) \\ &\quad + 2e_i^T(t) N_i u_i(t) dt \end{aligned} \quad (7)$$

این تابع هزینه به صورت عمومی و مربعی در نظر گرفته شده است. ماتریس‌های وزنی Q_i ، R_i و N_i ماتریس‌هایی با ابعاد مناسب هستند که به ترتیب وزن ورودی‌ها، حالت‌ها و ترکیب بین حالت و ورودی را در نظر می‌گیرد. از ماتریس‌های Q_i و N_i می‌توان برای هموار کردن^{۳۶} حرکت استفاده کرد. با R_i می‌توان میزان مصرف انرژی/سوخت را کنترل کرد. برخلاف مقاله‌های [37]-[41]، [43] این تابع هزینه به صورت عمومی تری در نظر گرفته شده است.

۳. حل مسئله کنترل ردیاب بهینه

برای حل مسئله کنترل بهینه می‌توان از روش برنامه‌ریزی پویا^{۳۷} [44] استفاده کرد. با فرض استفاده از فرمان‌های کنترلی ضربه‌ای^{۳۸} برای رانشگرها^{۳۹}، لازم است که ابتدا معادله‌ها را به صورت زمان گسسته نشان داد. به این ترتیب، می‌توان معادله‌ها را به شکل زیر گسسته کرد:

$$\begin{aligned} e_i(k+1) &= A e_i(k) + B_i u_i(k), \quad i \\ &= 1, 2, \dots, p \end{aligned} \quad (8)$$

که در آن ماتریس‌های A و B به ترتیب ماتریس سیستم گسسته و ورودی کنترلی است. Δt زمان نمونه برداری (یا زمان بین ضربه‌های رانشگر) است. فرم زمان گسسته تابع هزینه به صورت زیر قابل بیان است:

$$J_i(k) = \frac{1}{2} \sum_{j=k}^{\infty} U_i(e_i(j), u_i(j)). \quad (9)$$



$$\begin{aligned}
 & V_i(e_i(k)) \\
 &= \frac{1}{2} e_i(k)^T Q_i e_i(k) \\
 &+ \frac{1}{2} u_i(k)^T R_i u_i(k) + e_i(k)^T N_i u_i(k) \quad (18) \\
 &+ \frac{\gamma_i}{2} \sum_{j=k+1}^{\infty} \gamma_i^{j-(k+1)} (e_i(j)^T Q_i e_i(j) \\
 &+ u_i^T(j) R_i u_i(j) + 2e_i^T(j) N_i u_i(j))
 \end{aligned}$$

که برابر است با:

$$\begin{aligned}
 V_i(e_i(k)) &= \frac{1}{2} e_i(k)^T Q_i e_i(k) \\
 &+ \frac{1}{2} u_i(k)^T R_i u_i(k) \quad (19) \\
 &+ e_i(k)^T N_i u_i(k) \\
 &+ \gamma_i V_i(e_i(k+1))
 \end{aligned}$$

با جاگذاری معادله (۱۲) در معادله (۱۹)

می‌توان آن را برحسب ماتریس P_i بازنویسی کرد:

$$\begin{aligned}
 e_i(k)^T P_i e_i(k) &= e_i(k)^T Q_i e_i(k) \\
 &+ u_i(k)^T R_i u_i(k) \\
 &+ 2e_i(k)^T N_i u_i(k) \quad (20) \\
 &+ \gamma_i e(k) \\
 &+ 1)^T P_i e(k+1)
 \end{aligned}$$

در نهایت می‌توان معادله همیلتون^{۴۴} را به

شکل زیر نوشت:

$$\begin{aligned}
 H_i(e_i(k), u_i(k)) \\
 &= e_i(k)^T Q_i e_i(k) \\
 &+ u_i(k)^T R_i u_i(k) \quad (21) \\
 &+ 2e_i(k)^T N_i u_i(k) \\
 &+ \gamma_i V_i(e(k+1)) \\
 &- V_i(e_i(k))
 \end{aligned}$$

قضیه زیر نشان می‌دهد که چگونه می‌توان با

استفاده از برنامه‌ریزی پویا مسئله کنترل بهینه را حل کرد.

قضیه ۱- سیستم با معادله (۸) و تابع ارزش

(۱۰) در نظر بگیرید. می‌توان نشان داد که قانون

کنترلی به صورت

$$u_i(k) = -K_{1i} e_i(k) \quad (22)$$

قابل بیان است که در آن،

$$K_{1i} = (R_i + \gamma_i B_i^T P_i B_i)^{-1} (\gamma_i B_i^T P_i A + N_i^T) \quad (23)$$

$$\begin{aligned}
 & V_i(e_i(k)) \\
 &= \frac{1}{2} \sum_{j=k}^{\infty} \gamma_i^{j-k} (e_i(j)^T Q_i e_i(j) \\
 &+ e_i(j)^T K_i^T R_i K_i e_i(j) \\
 &+ 2e_i(j)^T N_i K_i e_i(j)) \quad (13) \\
 &= \frac{1}{2} \sum_{j=0}^{\infty} \gamma_i^j (e_i(j+k)^T Q_i e_i(j+k) \\
 &+ e_i(j+k)^T K_i^T R_i K_i e_i(j+k) \\
 &+ 2e_i(j+k)^T N_i K_i e_i(j+k))
 \end{aligned}$$

$$\begin{aligned}
 e_i(j+k) &= G^j e_i(k) \\
 &= (A + B_i K_i)^j e_i(k) \quad (14)
 \end{aligned}$$

با جاگذاری معادله (۱۴) در معادله (۱۳)

می‌توان رابطه زیر را به دست آورد:

$$\begin{aligned}
 & V_i(e_i(k)) \\
 &= \frac{1}{2} \sum_{j=0}^{\infty} \gamma_i^{j-k} \left(e_i(k)^T G_i^{jT} Q_i G_i^j e_i(k) \right. \\
 &+ e_i(k)^T G_i^{jT} K_i^T R_i K_i G_i^j e_i(k) \\
 &+ \left. 2e_i(k)^T G_i^{jT} N_i K_i G_i^j e_i(k) \right) \quad (15)
 \end{aligned}$$

که معادل است با:

$$V_i(e_i(k)) = \frac{1}{2} e_i(k)^T P_i e_i(k) \quad (16)$$

که در آن

$$\begin{aligned}
 P_i &= \sum_{j=0}^{\infty} \gamma_i^j \left(G_i^{jT} Q_i G_i^j \right. \\
 &+ G_i^{jT} K_i^T R_i K_i G_i^j \\
 &+ \left. 2G_i^{jT} N_i^T K_i G_i^j \right) \quad (17)
 \end{aligned}$$

به این ترتیب لم اثبات شد.

۲-۳. معادله بلمن و معادله جبری ریکاتی

در این بخش راه‌حلی برای حل مسئله کنترل

بهینه پیکربندی مجدد ماهواره‌ها با استفاده از

معادله‌های بلمن^{۴۳} ارائه می‌شود. برای این کار

معادله بلمن را به صورت رابطه (۱۸) می‌توان

نوشت.



و ماتریس P_i از حل معادله جبری ریکاتی تخفیف یافته زیر به دست خواهد آمد.

$$Q_i - P_i + \gamma_i A^T P_i A + ((2 - \gamma_i)L^T + N_i)S^{-1} \left(\left(2 - \frac{1}{\gamma_i} \right) L + \frac{1}{\gamma_i} N_i \right) = 0 \quad (24)$$

که در آن

$$L_i = \gamma_i B_i^T P_i A + N_i^T \\ S_i = R_i + \gamma_i B_i^T P_i B_i \text{ و}$$

اثبات - شرط لازم برای بهینگی به قرار زیر است [48]:

$$\frac{\partial H_i(e_i(k), u_i(k))}{\partial u_i(k)} = 2R_i u_i(k) + N_i^T e_i(k) + \frac{\gamma_i \partial e_i(k+1)}{\partial u_i(k)} \frac{\partial V_i(e_i(k+1))}{\partial e_i(k+1)} = 0. \quad (25)$$

بنابراین به راحتی می توان نشان داد که،

$$u_i(k) = -(R_i + \gamma_i B_i^T P_i B_i)^{-1} (\gamma_i B_i^T P_i A + N_i^T) e_i(k) \quad (26)$$

به این ترتیب قانون کنترل بهینه استخراج شد. با جاگذاری معادله (۸) و (۲۲) در معادله (۲۱)، در نهایت می توان معادله زیر را به دست آورد:

$$P_i - Q_i - K_{1i}^T R_i K_{1i} + 2N_i K_{1i} - \gamma_i A^T P_i A + \gamma_i K_{1i}^T B_i^T P_i A + \gamma_i A^T P_i B_i K_{1i} - \gamma_i K_{1i}^T B_i^T P_i B_i K_{1i} = 0 \quad (27)$$

با تعریف $L_i = \gamma_i B_i^T P_i A + N_i^T$ و

$S_i = R_i + \gamma_i B_i^T P_i B_i$ و با توجه به اینکه

$K_{1i} = S_i^{-1} L_i$ است، معادله ها را می توان

در نهایت به شکل ساده تر معادله (۲۴) نمایش داد. بنابراین قضیه اثبات شد.

در قضیه بعدی نشان داده خواهد شد که

قانون کنترلی معادله (۲۲) سیستم را پایدار می کند.

قضیه ۲- معادله (۸) و تابع ارزش معادله (۱۰) و قانون کنترلی معادله (۲۶) را در نظر بگیرید. می توان نشان داد که قانون کنترلی حاصل از حل معادله ریکاتی تخفیف یافته (۲۷) سیستم با معادله (۸) را پایدار می کند.

اثبات - فرض کنید که حالت های جدید را با

رابطه

$\bar{e}_i = \gamma_i^{\frac{1}{2}} e_i$ تعریف شده است. اگر حالت های جدید پایدار مجانبی باشند و به سمت صفر همگرا شوند، آنگاه می توان گفت که حالت های e_i نیز به سمت صفر همگرا خواهند شد. با جاگذاری تعریف جدید در معادله (۸) و استفاده از قانون کنترلی معادله (۲۲) می توان رابطه زیر را نوشت:

$$\bar{e}_i(k+1) = \gamma_i^{\frac{1}{2}} A \bar{e}_i(k) - \gamma_i^{\frac{1}{2}} B_i K_{1i} \bar{e}_i(k) = \gamma_i^{\frac{1}{2}} (A - B_i K_{1i}) \bar{e}_i(k) \quad (28)$$

که در آن K_{1i} مطابق معادله (۲۳) است. حال تابع لیاپانوف زیر را در نظر بگیرید:

$$V_i(\bar{e}_i(k)) = \frac{\gamma}{2} \bar{e}_i(k)^T P_i \bar{e}_i(k) \quad (29)$$

که در آن P_i از حل معادله جبری ریکاتی تخفیف یافته معادله (۲۴) به دست می آید. به این ترتیب می توان نوشت:

$$V_i(\bar{e}_i(k+1)) - V_i(\bar{e}_i(k)) = \frac{\gamma}{2} \bar{e}_i(k)^T (A - B_i K_{1i})^T P_i (A - B_i K_{1i}) \bar{e}_i(k) - \frac{\gamma}{2} \bar{e}_i(k)^T P_i \bar{e}_i(k) \quad (30)$$

با جاگذاری معادله (۲۸) در معادله (۳۰) می توان رابطه زیر را به دست آورد:



(36)

$$\begin{aligned}
& V_i(\bar{e}_i(k+1)) - V_i(\bar{e}_i(k)) \\
&= \frac{-1}{2} \left(\begin{bmatrix} \bar{e}_i(k) \\ -K_{1i}\bar{e}_i(k) \end{bmatrix}^T \begin{bmatrix} Q_i & N_i \\ N_i^T & R_i \end{bmatrix} \begin{bmatrix} \bar{e}_i(k) \\ -K_{1i}\bar{e}_i(k) \end{bmatrix} \right) \\
&= \frac{-1}{2} \left(\begin{bmatrix} \bar{e}_i(k) \\ u_i(k) \end{bmatrix}^T \begin{bmatrix} Q_i & N_i \\ N_i^T & R_i \end{bmatrix} \begin{bmatrix} \bar{e}_i(k) \\ u_i(k) \end{bmatrix} \right) \\
&= \frac{-1}{2} \left(\begin{bmatrix} \bar{e}_i(k) \\ u_i(k) \end{bmatrix}^T W_i \begin{bmatrix} \bar{e}_i(k) \\ u_i(k) \end{bmatrix} \right) < 0.
\end{aligned}$$

با توجه به اینکه فرض شده $Q_i \geq 0$ و $\det \begin{pmatrix} Q_i & N_i \\ N_i^T & R_i \end{pmatrix} > 0$ می‌توان استدلال کرد که ماتریس W_i نیز مثبت معین است. در نتیجه معادله تفاضلی لیپانوفی همواره منفی خواهد بود و سیستم پایدار است. با انتخاب مناسب ماتریس‌های وزنی Q_i ، R_i و N_i می‌توان میزان خطای ردیابی را به میزان دلخواه کوچک کرد.

در ادامه این مقاله، روشی ارائه خواهد شد تا بتوان به جای حل مستقیم معادله (۲۴)، آن را با استفاده از یادگیری تقویتی حل کرد. همچنین نشان داده خواهد شد که با استفاده از این روش می‌توان مسئله حل معادله ریکاتی تخفیف یافته را بدون نیاز به داشتن پارامترهای سیستم و تنها با اندازه‌گیری داده‌های حین کار به صورت برخط حل کرد.

۴. یادگیری-کیو برای حل مسئله

پیکربندی مجدد

همان‌طور که پیش‌ازین بیان شد، هدف از این مقاله طراحی یک کنترل‌کننده بدون مدل است که به صورت تطبیقی می‌تواند پاسخ‌های بهینه را به دست آورد. برای این کار ابتدا تابع-کیو^{۴۵} بر اساس معادله بلمن (۱۹) تعریف خواهد شد. سپس با استفاده از تابع-کیو نشان داده می‌شود که کنترل بهینه مستخرج بر اساس تابع-کیو

$$\begin{aligned}
& V_i(\bar{e}_i(k+1)) - V_i(\bar{e}_i(k)) \\
&= \frac{1}{2} \bar{e}_i(k)^T (\gamma_i A^T P_i A \\
&\quad - \gamma_i A^T P_i B_i K_{1i} \\
&\quad - \gamma_i K_{1i}^T B_i^T P_i A \\
&\quad + \gamma_i K_{1i}^T B_i^T P_i B_i K_{1i} \\
&\quad - P_i) \bar{e}_i(k)
\end{aligned} \quad (31)$$

با استفاده از معادله (۲۷) می‌توان رابطه زیر را نوشت:

$$\begin{aligned}
& \gamma_i A^T P_i A - \gamma_i K_{1i}^T B_i^T P_i A \\
&\quad - \gamma_i A^T P_i B_i K_{1i} \\
&\quad + \gamma_i K_{1i}^T B_i^T P_i B_i K_{1i} \\
&\quad - P_i \\
&= -Q_i - K_{1i}^T R_i K_{1i} \\
&\quad + 2N_i K_{1i}
\end{aligned} \quad (32)$$

با جاگذاری معادله (۳۲) در معادله (۳۱)

رابطه زیر به دست می‌آید:

$$\begin{aligned}
& V_i(\bar{e}_i(k+1)) - V_i(\bar{e}_i(k)) \\
&= \frac{1}{2} \bar{e}_i(k)^T (-Q_i \\
&\quad - K_{1i}^T R_i K_{1i} \\
&\quad + 2N_i K_{1i}) \bar{e}_i(k)
\end{aligned} \quad (33)$$

یا به طور مشابه،

$$\begin{aligned}
& V_i(\bar{e}_i(k+1)) - V_i(\bar{e}_i(k)) \\
&= \frac{-1}{2} (\bar{e}_i(k)^T (Q_i) \bar{e}_i(k) \\
&\quad + \bar{e}_i(k)^T K_{1i}^T R_i K_{1i} \bar{e}_i(k) \\
&\quad - 2\bar{e}_i(k)^T N_i K_{1i} \bar{e}_i(k))
\end{aligned} \quad (34)$$

از آنجا که هر جمله معادله (۳۴) یک عدد اسکالر است، می‌توان معادله (۳۴) را به شکل زیر بازنویسی کرد:

$$\begin{aligned}
& V_i(\bar{e}_i(k+1)) - V_i(\bar{e}_i(k)) \\
&= \frac{-1}{2} (\bar{e}_i(k)^T (Q_i) \bar{e}_i(k) \\
&\quad + \bar{e}_i(k)^T K_{1i}^T R_i K_{1i} \bar{e}_i(k) \\
&\quad - \bar{e}_i(k)^T N_i K_{1i} \bar{e}_i(k) \\
&\quad - \bar{e}_i(k)^T K_{1i}^T N_i^T \bar{e}_i(k))
\end{aligned} \quad (35)$$

یا به طور معادل



دقیقاً برابر با کنترل بهینه استخراج شده در معادله (۲۲) است. در نهایت الگوریتمی مبتنی بر یادگیری-کیو ارائه خواهد شد که می‌تواند با روشی بازگشتی، تابع-کیو را حل کند و بدون داشتن اطلاعات مدل و تنها با اندازه‌گیری اطلاعات سیستم حین کار سیستم، قانون کنترل بهینه را به صورت برخط پیدا کند. یادگیری-کیو یکی از روش‌های یادگیری تقویتی است [24]، [49]. در این روش سیگنال کنترلی و تابع ارزش با استفاده از اطلاعات استخراج شده از سیستم در حین حرکت قابل محاسبه است.

۴-۱. تابع-کیو برای مسئله ردیاب بهینه

برای طراحی کنترل کننده و حل معادله ریکاتی به روش یادگیری تقویتی ابتدا لازم است تا تابع-کیو را تعریف کنیم. با استفاده از معادله بلمن (۱۹)، می‌توان تابع-کیو را به شکل زیر تعریف کرد:

$$T_i(e_i(k), u_i(k)) = \frac{1}{2} e_i(k)^T Q_i e_i(k) + \frac{1}{2} u_i(k)^T R_i u_i(k) + e_i(k)^T N_i u_i(k) + \gamma_i e_i(k) + 1)^T P_i e_i(k+1) \quad (37)$$

با استفاده از معادله (۸)، معادله بالا را می‌توان به صورت زیر بازنویسی کرد:

$$T_i(e_i(k), u_i(k)) = \frac{1}{2} \begin{bmatrix} e_i(k) \\ u_i(k) \end{bmatrix}^T \begin{bmatrix} Q_i + \gamma_i A^T P_i A & N_i + \gamma_i A^T P_i B_i \\ N_i^T + \gamma_i B_i^T P_i A & R_i + \gamma_i B_i^T P_i B_i \end{bmatrix} \begin{bmatrix} e_i(k) \\ u_i(k) \end{bmatrix} \quad (38)$$

حال فرض کنید که تابع-کیو به شکل زیر تعریف شود:

$$T_i(e_i(k), u_i(k)) = \frac{1}{2} \begin{bmatrix} e_i(k) \\ u_i(k) \end{bmatrix}^T M_i \begin{bmatrix} e_i(k) \\ u_i(k) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} e_i(k) \\ u_i(k) \end{bmatrix}^T \begin{bmatrix} (M_{ee})_i & (M_{eu})_i \\ (M_{ue})_i & (M_{uu})_i \end{bmatrix} \begin{bmatrix} e_i(k) \\ u_i(k) \end{bmatrix} \quad (39)$$

که در آن ماتریس M_i یک ماتریس متقارن است ($M_i = M_i^T$). آنگاه سیگنال کنترل بهینه را می‌توان از رابطه زیر به دست آورد:

$$\frac{\partial T_i(e_i(k), u_i(k))}{\partial u_i(k)} = 0 \quad (40)$$

با حل این معادله می‌توان سیگنال کنترل بهینه را به شکل زیر محاسبه کرد:

$$u_i(k) = -(M_{uu}^{-1})_i (M_{ue})_i e_i(k) \quad (41)$$

با مقایسه این معادله با معادله (۳۹) می‌توان نوشت:

$$u_i(k) = -(R_i + \gamma_i B_i^T P_i B_i)^{-1} (N_i^T + \gamma_i B_i^T P_i A) e_i(k) \quad (42)$$

که دقیقاً مشابه معادله (۲۲) است. تنها تفاوت در اینجا این است که معادله (۴۲) تنها به اطلاعات ماتریس M_i نیاز دارد و به هیچ اطلاعاتی از دینامیک مسئله نیازی ندارد. اگر بتوان این ماتریس را به صورت مستقیم محاسبه کرد، آنگاه می‌توان کنترل کننده بهینه را بدون داشتن اطلاعات دینامیک سیستم یعنی ماتریس‌های A و B_i محاسبه کرد.

۴-۲. کنترل کننده بدون مدل مبتنی بر یادگیری-کیو

همان طور که اشاره شد، می‌توان سیگنال کنترلی را بدون داشتن اطلاعات سیستم به صورت مستقیم استخراج کرد. برای اینکه بتوان این کار را انجام داد، می‌دانیم که تابع-کیو معادله (۳۹) می‌تواند معادله بلمن (۱۹) را برآورده کند:

$$T_i(e_i(k), u_i(k)) = \frac{1}{2} e_i(k)^T Q_i e_i(k) + \frac{1}{2} u_i(k)^T R_i u_i(k) + e_i(k)^T N_i u_i(k) + \gamma_i T_i(e_i(k), u_i(k+1)) \quad (43)$$

که در آن قانون کنترل بهینه مطابق با معادله





(۴۱) است. حال بردار Z_i را به شکل زیر تعریف کنید:

$$Z_i(k) = \begin{bmatrix} e_i(k) \\ u_i(k) \end{bmatrix} \quad (44)$$

آنگاه می‌توان معادله (۳۹) را به شکل زیر بازنویسی کرد:

$$T_i(e_i(k), u_i(k)) = \frac{1}{2} Z_i^T M_i Z_i \quad (45)$$

با جاگذاری معادله (۴۵) در معادله (۴۳)، می‌توان معادله بلمن (۴۳) را به شکل زیر نوشت:

$$\begin{aligned} Z_i^T(k) M_i Z_i(k) &= e_i(k)^T Q_i e_i(k) \\ &+ u_i(k)^T R_i u_i(k) \\ &+ \gamma_i Z_i(k) \\ &+ 1)^T M_i Z_i(k+1) \end{aligned} \quad (46)$$

حال با استفاده از روش تکرار-ارزش^{۴۶} می‌توان سیگنال کنترل بهینه را با استفاده از الگوریتم ۱ استخراج کرد.

الگوریتم ۱. حل تابع-کیو با روش تکرار-

ارزش

گام اول	(مقداردهی اولیه) ابتدا مقدار اولیه بهره کنترلی $K_{1_i}^j$ را به دلخواه انتخاب کنید.
گام دوم	برای $j = 0, 1, 2, \dots$ گام‌های زیر را تکرار کنید.
گام سوم	در هر گام نمونه برداری اطلاعات سیستم $(e_i(k))$ و $(u_i(k))$ را جمع‌آوری کنید. آن قدر اطلاعات جمع کنید تا با استفاده از روش حداقل-مربعات ^{۴۷} ، ماتریس M_i^j را در مرحله j پیدا کنید: (47)
گام چهارم	پس از پیدا کردن ماتریس M_i^j ، سیگنال کنترلی را به شکل زیر به روزرسانی کنید: (48)
گام پنجم	اگر $\ M_i^{j+1} - M_i^j\ < \varepsilon$ شد، الگوریتم را متوقف کنید در غیر این صورت $j = j + 1$ قرار دهید و به گام سوم بازگردید.

باید اشاره کرد که در الگوریتم ۱ برای پیدا کردن سیگنال کنترلی بهینه از هیچ‌یک از اطلاعات سیستم A و B_i استفاده نشده است و با استفاده از محاسبه مستقیم ماتریس M_i ، سیگنال کنترلی بهینه به صورت تطبیقی و بدون نیاز به مدل، به صورت بر-خط به دست آمده است.

همگرایی الگوریتم ۱ در [50] به اثبات رسیده است. در این روش، سیگنال کنترلی به صورت برخط و بدون داشتن اطلاعات مدل محاسبه می‌شود. معادله (۴۷) را می‌توان با اطلاعات جمع‌آوری شده $Z_i(k)$ و $Z_i(k+1)$ و

$$\begin{aligned} r_i(k) &= e_i(k)^T Q_i e_i(k) \\ &+ (u_i^j(k))^T R_i u_i^j(k) \\ &+ 2e_i(k)^T N_i u_i^j(k) \end{aligned}$$

محاسبه کرد. در حقیقت معادله (۴۷) یک معادله عددی (اسکالر) است. بنابراین با نمونه برداری به تعداد درایه‌های ماتریس M_i^{j+1} می‌توان آن را به روش‌های حداقل-مربعات حل کرد. برای پیدا کردن ماتریس M_i^{j+1} می‌توان نشان داد که جمله $Z_i^T(k) M_i^{j+1} Z_i(k)$ را می‌توان به شکل زیر نوشت:

$$\begin{aligned} Z_i^T(k) M_i^{j+1} Z_i(k) &= (\text{vec}(M_i^{j+1}))^T (Z_i(k) \\ &\otimes Z_i(k)) \\ &= (\bar{m}_i^{j+1})^T \bar{Z}_i(k) \end{aligned}$$

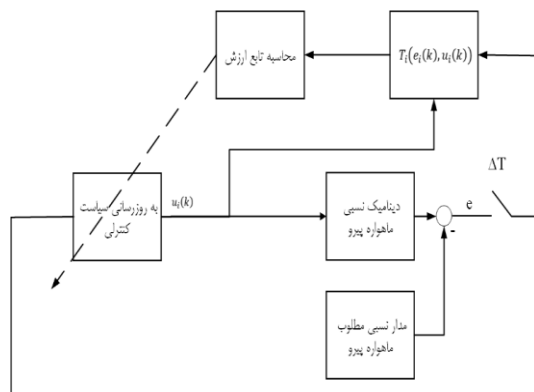
که در آن $\bar{m}_i^{j+1} = \text{vec}(M_i^{j+1})$ یک بردار ستونی است که در آن همه درایه‌های ماتریس M_i^{j+1} به صورت ستونی پشت سر هم قرار گرفته‌اند و علامت \otimes به معنی ضرب کرونگر^{۴۸} است. حال فرض کنید که ماتریس M_i^{j+1} یک ماتریس متقارن با ابعاد $n \times n$ است. پس حداقل به تعداد $n \times \frac{n+1}{2}$ گام نمونه برداری، باید از مسئله اطلاعات جمع شود تا بتوان معادله (۴۷) را

با روش حداقل-مربعات دسته‌ای^{۴۹} حل کرد. این روش را می‌توان به شکل زیر نوشت:

$$\begin{aligned} & (\bar{m}_i^{j+1})^T \bar{z}_i(k) \\ & = e_i(k)^T Q_i e_i(k) \\ & + (u_i^j(k))^T R_i u_i^j(k) \end{aligned} \quad (49)$$

$$+ 2e_i(k)^T N_i u_i^j(k) + (\bar{m}_i^j)^T \bar{z}_i(k) + 1$$

همان‌طور که از معادله (۴۹) پیداست، سمت راست معادله کاملاً معلوم است. پس از جمع تعداد کافی از اطلاعات می‌توان این معادله را حل کرد. لازم به اشاره است که برای اطمینان از همگرایی الگوریتم ۱ لازم است تا شرط تحریک مداوم^{۵۰} ورودی رعایت شود [50], [47][44]. بلوک دیاگرام کنترلی برای حل مسئله پیکربندی مجدد ماهواره‌ها در شکل ۳ نشان داده شده است.



شکل ۳. بلوک دیاگرام کنترلی سیستم حلقه-بسته پیکربندی مجدد ماهواره‌ها

۵. نتایج شبیه‌سازی

روش کنترلی تطبیقی بهینه بدون مدل برخط توسعه داده شده در بخش ۴ برای حل مسئله پیکربندی مجدد ماهواره‌ها در بخش ۲ مورد استفاده قرار می‌گیرد. برای این منظور یک سناریوی عملیاتی از پیکربندی مجدد ماهواره‌ها در آرایش پروازی ماهواره‌ها طراحی شده است. در ادامه، ابتدا سناریو شبیه‌سازی تشریح می‌شود،

سپس نتایج کنترل‌کننده طراحی شده ارائه خواهد شد.

در سناریوی طراحی شده، فرض شده است که ماهواره پیشرو در یک مدار دایروی قرار دارد. پارامترهای مداری ماهواره پیشرو در جدول ۱ گزارش شده است. در این سناریو لازم است تا ماهواره‌های پیرو از پیکربندی کنونی خود به پیکربندی مطلوب مهاجرت نمایند. شرایط اولیه مناسب (برقراری شرط تطبیق انرژی) برای تعریف مدار نسبی اولیه و مدار نسبی نهایی در جدول‌های ۲ و ۳ آورده شده است.

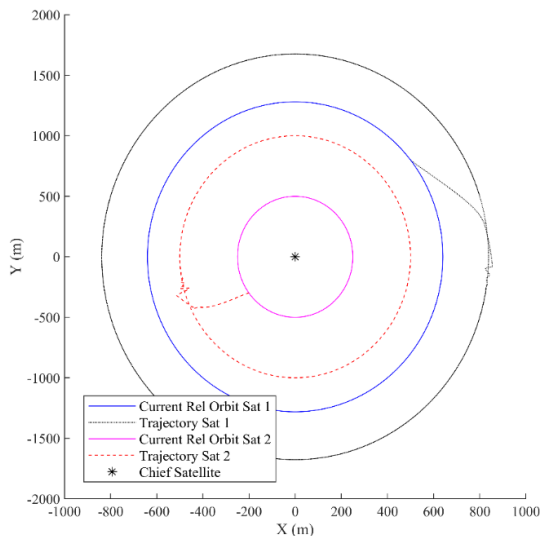
جدول ۱. پارامترهای مداری ماهواره پیشرو

پارامتر مداری	مقدار
طول نیم قطر بزرگ ^{۵۱}	۷۱۰۸ کیلومتر
خروج از مرکز ^{۵۲}	۰
شیب مداری ^{۵۳}	۷۰ درجه
آنومالی حقیقی	۰ درجه
حضیض مداری ^{۵۴}	۰ درجه
طول گره صعودی ^{۵۵}	۴۵ درجه

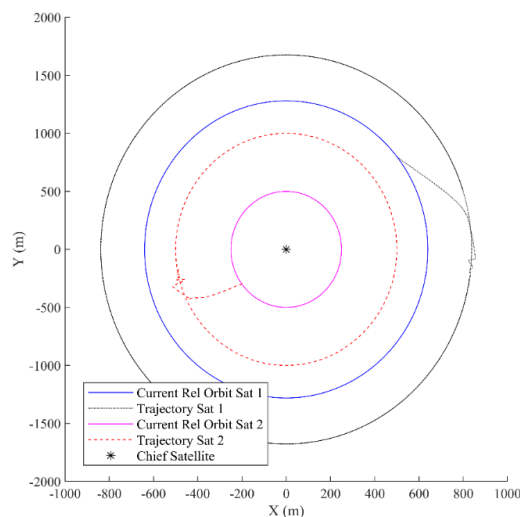
جدول ۲. شرایط اولیه تعریف مدارهای نسبی اولیه و نهایی ماهواره پیرو اول

مدار نسبی اولیه	مدار نسبی نهایی
x_0	-۲۰۰
y_0	-۳۰۰
z_0	-۵۰۰
v_{x_0}	-۰.۱۵۸
v_{y_0}	-۰.۴۲۱۴
v_{z_0}	-۰.۳۱۶

در این شبیه‌سازی فرض شده است که در هر راستای ماهواره، دو رانشگر نصب شده است. همچنین فرض شده است که اطلاعات دقیقی از دینامیک حرکت معادله (۳) وجود ندارد.



شکل ۴. پیکربندی مجدد ماهواره در صفحه x-y



شکل ۵. پیکربندی مجدد ماهواره در فضای x-y-z

همان طور که در شکل‌های ۴ و ۵ نشان داده شده است، ماهواره‌ها با موفقیت از مدار نسبی کنونی خود که با خط پر نشان داده شده است، به مدار نسبی هدف که با خط چین نشان داده شده است مهاجرت می‌کنند.

ضرایب کنترل بهینه با استفاده از الگوریتم ۱ بدون داشتن اطلاعات مدل به شکل زیر محاسبه شده است:

$$K_1 = \begin{bmatrix} 0.0628 & -0.0009 & -0.0006 & 0.6350 & 0.0038 & 0.0034 \\ -0.0003 & 0.0628 & -0.0006 & 0.0031 & 0.3651 & 0.0035 \\ -0.006 & -0.0006 & 0.0628 & 0.0034 & 0.0035 & 0.3650 \end{bmatrix}$$

$$K_2 = \begin{bmatrix} 0.0628 & -0.0009 & -0.0006 & 0.6350 & 0.0038 & 0.0034 \\ -0.0003 & 0.0628 & -0.0006 & 0.0031 & 0.3651 & 0.0035 \\ -0.006 & -0.0006 & 0.0628 & 0.0034 & 0.0035 & 0.3650 \end{bmatrix}$$

به این ترتیب اطلاعات ماتریس A_c و B_{c_i} برای کنترل کننده معلوم نیست.

جدول ۳. شرایط اولیه تعریف مدارهای نسبی اولیه و نهایی ماهواره پیرو دوم

	مدار نسبی اولیه	مدار نسبی نهایی
x_0	۸۰۰	-۴۰۰
y_0	۵۰۰	-۶۰۰
z_0	۵۰	-۳۵۰
v_{x_0}	۰.۲۶۳۴	-۰.۳۱۶
v_{y_0}	-۱.۶۸۵۶	۰.۸۴۲۸
v_{z_0}	۰.۵۲۶۷	-۰.۶۳۳۱

برای شبیه‌سازی فرض شده است که زمان نمونه‌برداری از سیستم ۲ ثانیه یکبار است. بنابراین فاصله زمانی بین هر ضربه رانشگر ۲ ثانیه خواهد بود. برای تعریف تابع هزینه معادله (۱۰) در نظر گرفته شده است. ماتریس‌های وزنی برای هر ماهواره

$$R_1 = R_2 = 10^2 I_3 N_1 = Q_1 = Q_2 = I_6$$

$$N_2 = -0.5 \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

فرض شده‌اند، که در آن I_n ماتریس واحد با ابعاد n است. ضریب تخفیف برابر با $\gamma_1 = \gamma_2 = 1$ در نظر گرفته شده است.

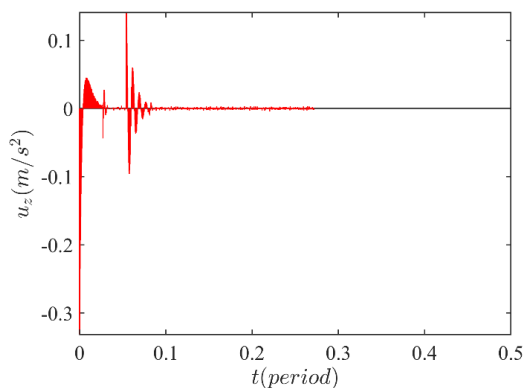
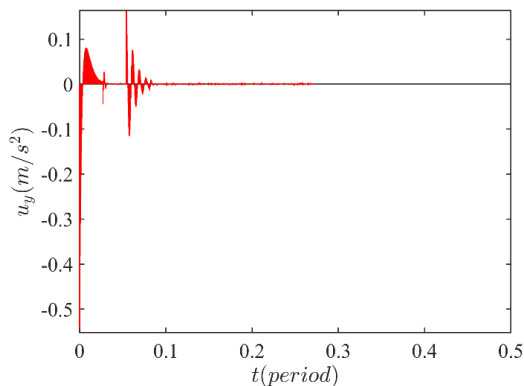
الگوریتم ۱ برای حل مسئله پیکربندی مجدد ماهواره‌ها مورداستفاده قرار گرفته است. نتایج شبیه‌سازی با استفاده از این الگوریتم در شکل‌های ۴ و ۵ نشان داده شده است.



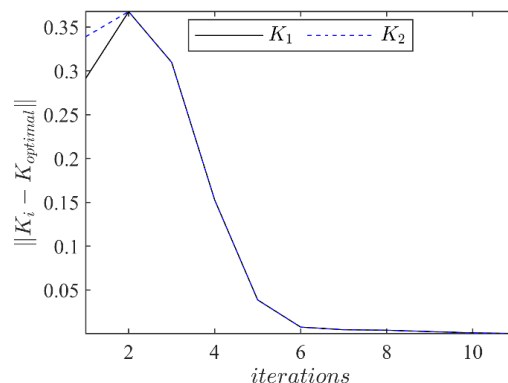
اگر فرض شود که اطلاعات مدل وجود دارد، می‌توان ضرایب کنترل‌کننده بهینه را به شکل زیر با استفاده از معادله **Error! Unknown switch argument.** به‌طور مستقیم حساب کرد:

$$K_{1optimal} = \begin{bmatrix} 0.0629 & -0.0009 & -0.0006 & 0.6353 & 0.0037 & 0.0033 \\ -0.0003 & 0.0629 & -0.0006 & 0.0030 & 0.3653 & 0.0033 \\ -0.006 & -0.0006 & 0.0629 & 0.0033 & 0.0033 & 0.3653 \end{bmatrix}$$

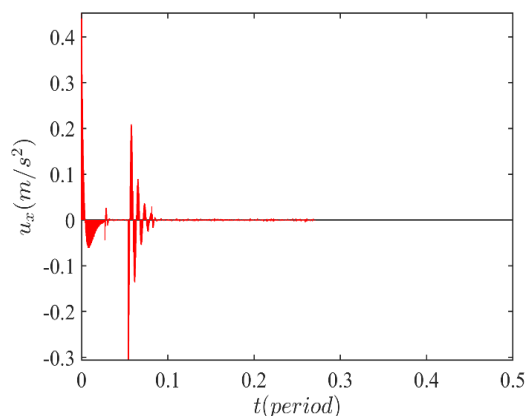
$$K_{2optimal} = \begin{bmatrix} 0.0629 & -0.0009 & -0.0006 & 0.6353 & 0.0037 & 0.0033 \\ -0.0003 & 0.0629 & -0.0006 & 0.0030 & 0.3653 & 0.0033 \\ -0.006 & -0.0006 & 0.0629 & 0.0033 & 0.0033 & 0.3653 \end{bmatrix}$$



همان‌طور که مشاهده می‌شود، ضرایب کنترلی حساب شده با استفاده از الگوریتم ۱ به ضرایب کنترل بهینه نزدیک می‌شوند. شکل ۶ نحوه همگرایی ضرایب محاسبه‌شده با الگوریتم ۱ را به ضرایب بهینه نشان می‌دهد. مشاهده می‌شود که تقریباً پس از ۱۰ تکرار الگوریتم ۱، ضرایب محاسبه‌شده به ضرایب کنترل‌کننده بهینه بسیار نزدیک شده است.



شکل ۶. اختلاف نرم ضرایب کنترل بهینه با ضرایب کنترلی محاسبه در الگوریتم ۱



شکل ۷. نیروی کنترلی اعمالی به ماهواره اول

ورودی کنترلی اعمال‌شده به هر یک از رانشگرها نیز در شکل‌های ۷ و ۸ نمایش داده شده است.

باید خاطرنشان کرد، هنگامی که ماهواره به مدار نسبی هدف خود می‌رسد و در آن قرار می‌گیرد، دیگر نیروی کنترلی به آن‌ها اعمال نمی‌شود و ماهواره بدون مصرف انرژی/سوخت می‌تواند در مدار هدف باقی بماند. همچنین، ضرایب کنترل‌کننده محاسبه‌شده با الگوریتم ۱ می‌تواند در پیکربندی مجدد بعدی مورد استفاده قرار گیرد، به این ترتیب می‌توان سرعت الگوریتم ۱



را بهبود بخشید.

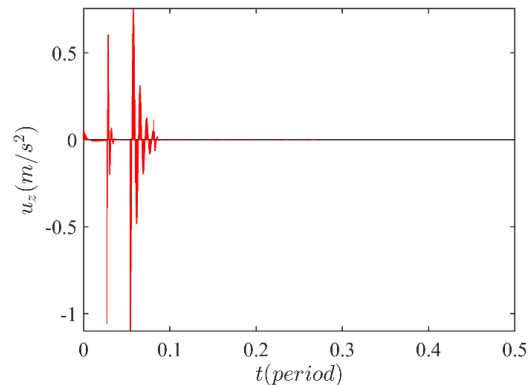
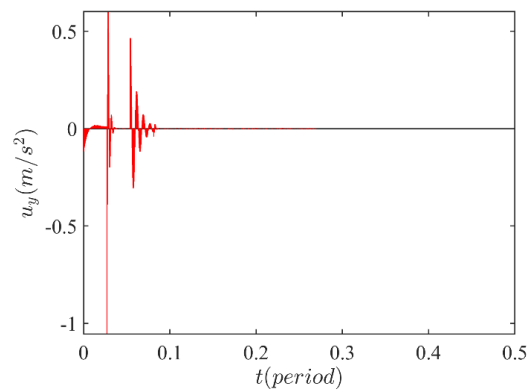
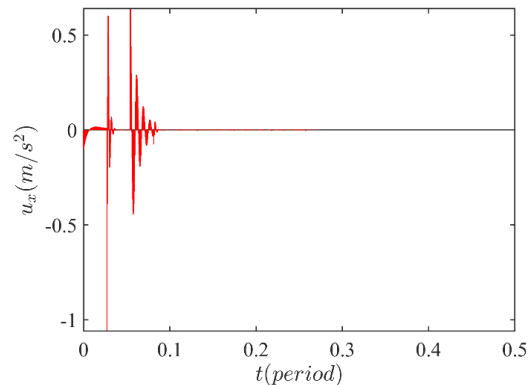
پیاده‌سازی می‌شود، و کنترل‌کننده خود اطلاعاتش را پیدا می‌کند و آموزش می‌بیند. به این ترتیب نیاز به دخالت ایستگاه‌های زمینی برای تعیین ضرایب کنترلی را به حداقل می‌رساند. علاوه بر این، حجم محاسبات در این روش زیاد نیست و از پیچیدگی کمی دارد [25].

۵. نتیجه‌گیری

در این مقاله یک کنترل‌کننده تطبیقی بهینه بدون مدل برای حل مسئله پیکربندی مجدد ماهواره‌ها ارائه شده است. در این راستا مسئله پیکربندی مجدد ماهواره‌ها مدل‌سازی و برای پیاده‌سازی آن روی رانشگرها با فرمان‌های ضربه‌ای به صورت زمان-گسسته بیان شد. سپس برای حل مسئله بهینه‌سازی یک تابع هزینه جدید تعمیم‌یافته تخفیف‌یافته برای بهینه‌سازی مسئله پیشنهاد شد و بر اساس آن یک کنترل‌کننده بهینه استخراج شد. پایداری این کنترل‌کننده به روش لیپانوف به اثبات رسید. سپس با استفاده از یادگیری-کیو، الگوریتمی برای حل معادله ریکاتی به صورت بر-خط و بدون داشتن اطلاعات دینامیکی سیستم ارائه شد. به دلیل آنکه کنترل‌کننده مذکور احتیاجی به مدل سیستم ندارد، می‌تواند خود را با نامعینی‌های مدل تطبیق دهد. برای نشان دادن کارایی الگوریتم پیشنهادی این مقاله، یک سناریو از پیکربندی مجدد آرایش پروازی ماهواره‌ها، با استفاده از این کنترل‌کننده شبیه‌سازی شد و مؤثر بودن صحت‌گذاری شد.

۶. مراجع

- [1] F. Y. Hadaegh, S. J. Chung, and H. M. Manohara, "On development of 100-gram-class spacecraft for swarm applications," *IEEE Syst. J.*, vol. 10,



شکل ۸. نیروی کنترلی اعمالی به ماهواره دوم

می‌توان با تغییر ماتریس‌های وزنی تابع هزینه، پاسخ‌های مختلف را متناسب با نیازهای مأموریت تولید کرد. با توجه به اینکه برای محاسبه ضرایب کنترل‌کننده بهینه از اطلاعات مدل استفاده نشده است، این روش می‌تواند در هنگامی که اطلاعات مدل موجود نیست و میزان دقیق پارامترها و اغتشاشات مشخص نیست، به خوبی نیازهای کنترلی را برآورده سازد. باید در نظر داشت که این روش به صورت بر-خط بر روی ماهواره



293–305, 2021.

- [10] D. Wang, B. Wu, and E. K. Poh, *Satellite Formation Flying*, vol. 87. Singapore: Springer Singapore, 2017.
- [11] H. Rouzegar, A. Khosravi, and P. Sarhadi, “Spacecraft formation flying control under orbital perturbations by state-dependent Riccati equation method in the presence of on-off actuators,” *Proc. Inst. Mech. Eng. Part G J. Aerosp. Eng.*, vol. 233, no. 8, pp. 2853–2867, Jun. 2019, doi: 10.1177/0954410018787417.
- [12] X. Liu, P. Lu, and B. Pan, “Survey of convex optimization for aerospace applications,” *Astrodynamics*, vol. 1, no. 1, pp. 23–40, Sep. 2017, doi: 10.1007/s42064-017-0003-8.
- [13] D. Parente, D. Spiller, and F. Curti, “Time-suboptimal satellite formation maneuvers using inverse dynamics and differential evolution,” *J. Guid. Control. Dyn.*, vol. 41, no. 5, pp. 1108–1121, 2018, doi: 10.2514/1.G003110.
- [14] W. Wang, G. Mengali, A. A. Quarta, and J. Yuan, “Distributed adaptive synchronization for multiple spacecraft formation flying around Lagrange point orbits,” *Aerosp. Sci. Technol.*, vol. 74, pp. 93–103, 2018, doi: 10.1016/j.ast.2018.01.007.
- [15] B. Shasti, A. Alasty, and N. Assadian, “Robust distributed control of spacecraft formation flying with adaptive network topology,” *Acta Astronaut.*, vol. 136, no. October 2016, pp. 281–296, Jul. 2017, doi: 10.1016/j.actaastro.2017.03.001.
- [16] H. Liu, Y. Tian, F. L. Lewis, Y. Wan, and K. P. Valavanis, “Robust formation flying control for a team of satellites subject to nonlinearities and uncertainties,” *Aerosp. Sci. Technol.*, vol. 95, p. 105455, Dec. 2019, doi: 10.1016/j.ast.2019.105455.
- [17] Y. Guo, J. Zhou, and Y. Liu, “Distributed RISE control for spacecraft formation reconfiguration with collision avoidance,” *J. Franklin Inst.*, vol. 356, no. 10, pp. 5332–5352, no. 2, pp. 673–684, 2016, doi: 10.1109/JSYST.2014.2327972.
- [2] D. Wang, B. Wu, and E. K. Poh, *Satellite Formation Flying Relative Dynamics, Formation Design, Fuel Optimal Maneuvers and Formation Maintenance*, vol. 87, no. 2001. 2017.
- [3] K. T. Alfriend, S. R. Vadali, P. Gurfil, J. P. How, and L. S. Breger, *Spacecraft formation flying: Dynamics, control and navigation*. 2009.
- [4] H. Cho, “Energy-optimal reconfiguration of satellite formation flying in the presence of uncertainties,” *Adv. Sp. Res.*, vol. 67, no. 5, pp. 1454–1467, Mar. 2021, doi: 10.1016/j.asr.2020.11.036.
- [5] K. Dharmarajan and G. B. Palmerini, “Optimal Reconfiguration Manoeuvres in Formation Flying Missions,” in *2021 IEEE Aerospace Conference (50100)*, Mar. 2021, vol. 2021-March, pp. 1–9, doi: 10.1109/AERO50100.2021.9438285.
- [6] X. Bai, Y. He, and M. Xu, “Low-Thrust Reconfiguration Strategy and Optimization for Formation Flying Using Jordan Normal Form,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 57, no. 5, pp. 3279–3295, Oct. 2021, doi: 10.1109/TAES.2021.3074204.
- [7] G. Di Mauro, D. Spiller, S. F. Rafano Carnà, and R. Bevilacqua, “Minimum-fuel control strategy for spacecraft formation reconfiguration via finite-time maneuvers,” *J. Guid. Control. Dyn.*, vol. 42, no. 4, pp. 752–768, 2019, doi: 10.2514/1.G003822.
- [8] G. Di Mauro, D. Spiller, R. Bevilacqua, and S. D’Amico, “Spacecraft formation flying reconfiguration with extended and impulsive maneuvers,” *J. Franklin Inst.*, vol. 356, no. 6, pp. 3474–3507, Apr. 2019, doi: 10.1016/j.jfranklin.2019.02.012.
- [9] H. M. PARI and H. Bolandi, “Discrete time multiple spacecraft formation flying attitude optimal control in presence of relative state constraints,” *Chinese J. Aeronaut.*, vol. 34, no. 4, pp.



- Lewis, T. Pipe, and C. Melhuish, "Reinforcement learning and optimal adaptive control: An overview and implementation examples," *Annu. Rev. Control*, vol. 36, no. 1, pp. 42–59, 2012, doi: 10.1016/j.arcontrol.2012.03.004.
- [27] D. Bertsekas, *Dynamic programming and optimal control: Volume I*, vol. 1. Athena scientific, 2012.
- [28] D. Bertsekas, *Reinforcement learning and optimal control*. Athena Scientific, 2019.
- [29] Y. Yang, Y. Wan, J. Zhu, and F. L. Lewis, "H ∞ Tracking Control for Linear Discrete-Time Systems: Model-Free Q-Learning Designs," *IEEE Control Syst. Lett.*, vol. 5, no. 1, pp. 175–180, Jan. 2021, doi: 10.1109/LCSYS.2020.3001241.
- [30] C. Chen, H. Modares, K. Xie, F. L. Lewis, Y. Wan, and S. Xie, "Reinforcement Learning-Based Adaptive Optimal Exponential Tracking Control of Linear Systems with Unknown Dynamics," *IEEE Trans. Automat. Contr.*, vol. 64, no. 11, pp. 4423–4438, Nov. 2019, doi: 10.1109/TAC.2019.2905215.
- [31] N. Li, I. Kolmanovsky, and A. Girard, "LQ control of unknown discrete-time linear systems—A novel approach and a comparison study," *Optim. Control Appl. Methods*, vol. 40, no. 2, pp. 265–291, 2019, doi: 10.1002/oca.2477.
- [32] S. A. A. Rizvi and Z. Lin, "Output Feedback Q-Learning Control for the Discrete-Time Linear Quadratic Regulator Problem," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 30, no. 5, pp. 1523–1536, 2019, doi: 10.1109/TNNLS.2018.2870075.
- [33] S. Ali Asad Rizvi and Z. Lin, "Model-Free Global Stabilization of Discrete-Time Linear Systems with Saturating Actuators Using Reinforcement Learning," *Proc. IEEE Conf. Decis. Control*, vol. 2018-Jul. 2019, doi: 10.1016/j.jfranklin.2019.05.003.
- [18] D. Lee, "Nonlinear disturbance observer-based robust control for spacecraft formation flying," *Aerosp. Sci. Technol.*, vol. 76, pp. 82–90, May 2018, doi: 10.1016/j.ast.2018.01.027.
- [19] G. Gaias and S. D'Amico, "Impulsive Maneuvers for Formation Reconfiguration Using Relative Orbital Elements," *J. Guid. Control. Dyn.*, vol. 38, no. 6, pp. 1036–1049, Jun. 2015, doi: 10.2514/1.G000189.
- [20] M. Chernick and S. D'Amico, "New Closed-Form Solutions for Optimal Impulsive Control of Spacecraft Relative Motion," *J. Guid. Control. Dyn.*, vol. 41, no. 2, pp. 301–319, Feb. 2018, doi: 10.2514/1.G002848.
- [21] B. L. Wu, D. W. Wang, and E. K. Poh, "Energy-optimal low-thrust satellite formation manoeuvre in presence of J2 perturbation," *Proc. Inst. Mech. Eng. Part G J. Aerosp. Eng.*, vol. 225, no. 9, pp. 961–968, 2011, doi: 10.1177/0954410011408659.
- [22] A. D. Ogundele, "Approximate analytic solution of nonlinear Riccati spacecraft formation flying dynamics in terms of orbit element differences," *Aerosp. Sci. Technol.*, vol. 113, p. 106686, 2021, doi: 10.1016/j.ast.2021.106686.
- [23] J. C. H. Christopher, "Watkins and peter dayan," *Q-Learning. Mach. Learn.*, vol. 8, no. 3, pp. 279–292, 1992.
- [24] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.
- [25] L. Buşoni, T. de Bruin, D. Tolić, J. Kober, and I. Palunko, "Reinforcement learning for control: Performance, stability, and deep approximators," *Annu. Rev. Control*, vol. 46, no. xxxx, pp. 8–28, 2018, doi: 10.1016/j.arcontrol.2018.09.005.
- [26] S. G. Khan, G. Herrmann, F. L.



10.2514/6.2018-0859.

- [41] S. Silvestrini and M. R. Lavagna, "Spacecraft Formation Relative Trajectories Identification for Collision-Free Maneuvers using Neural-Reconstructed Dynamics," no. January, pp. 1–14, 2020, doi: 10.2514/6.2020-1918.
- [42] M. Shirobokov, S. Trofimov, and M. Ovchinnikov, "Survey of machine learning techniques in spacecraft control design," *Acta Astronaut.*, vol. 186, no. April, pp. 87–97, Sep. 2021, doi: 10.1016/j.actaastro.2021.05.018.
- [43] A. Scorsoglio, A. D'Ambrosio, L. Ghilardi, B. Gaudet, F. Curti, and R. Furfaro, "Image-Based Deep Reinforcement Meta-Learning for Autonomous Lunar Landing," *J. Spacecr. Rockets*, pp. 1–13, 2021.
- [44] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal control*. John Wiley & Sons, 2012.
- [45] J. Sullivan, S. Grimberg, and S. D'Amico, "Comprehensive survey and assessment of spacecraft relative motion dynamics models," *J. Guid. Control. Dyn.*, vol. 40, no. 8, pp. 1837–1859, 2017, doi: 10.2514/1.G002309.
- [46] S. A. Schweighart and R. J. Sedwick, "High-Fidelity Linearized J Model for Satellite Formation Flight," *J. Guid. Control. Dyn.*, vol. 25, no. 6, pp. 1073–1080, Nov. 2002, doi: 10.2514/2.4986.
- [47] F. L. Lewis, D. Vrabie, and K. G. Vamvoudakis, "Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers," *IEEE Control Syst. Mag.*, vol. 32, no. 6, pp. 76–105, 2012.
- [48] D. Vrabie, K. G. Vamvoudakis, and F. L. Lewis, *Optimal adaptive control and differential games by reinforcement learning principles*. 2012.
- [49] P. Werbos, "Approximate dynamic programming for realtime control and neural modelling," *Handb. Intell. Decem*, no. Cdc, pp. 5276–5281, 2019, doi: 10.1109/CDC.2018.8618941.
- [34] S. Ali, A. Rizvi, and Z. Lin, "Output Feedback Optimal Tracking Control Using Reinforcement Q-Learning," *Proc. Am. Control Conf.*, vol. 2018-June, no. 2, pp. 3423–3428, 2018, doi: 10.23919/ACC.2018.8430997.
- [35] B. Kiumarsi, K. G. Vamvoudakis, H. Modares, and F. L. Lewis, "Optimal and Autonomous Control Using Reinforcement Learning: A Survey," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 29, no. 6, pp. 2042–2062, 2018, doi: 10.1109/TNNLS.2017.2773458.
- [36] X. Li, L. Xue, and C. Sun, "Linear quadratic tracking control of unknown discrete-time systems using value iteration algorithm," *Neurocomputing*, vol. 314, pp. 86–93, 2018, doi: 10.1016/j.neucom.2018.05.111.
- [37] M. Zheng, Y. Wu, and C. Li, "Reinforcement learning strategy for spacecraft attitude hyperagile tracking control with uncertainties," *Aerosp. Sci. Technol.*, vol. 119, p. 107126, Dec. 2021, doi: 10.1016/j.ast.2021.107126.
- [38] X. Wang, P. Shi, C. Wen, and Y. Zhao, "Design of Parameter-self-tuning Controller Based on Reinforcement Learning for Tracking Non-cooperative Targets in Space," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 9251, no. c, pp. 1–1, 2020, doi: 10.1109/taes.2020.2988170.
- [39] J. Broida and R. Linares, "Spacecraft rendezvous guidance in cluttered environments via reinforcement learning," *Adv. Astronaut. Sci.*, vol. 168, pp. 1777–1788, 2019.
- [40] F. Sun and K. Turkoglu, "Reinforcement learning based continuous-time on-line spacecraft dynamics control: Case study of NASA SPHERES spacecraft," *AIAA Guid. Navig. Control Conf. 2018*, no. 210039, pp. 1–11, Jan. 2018, doi:



-
36. Smoothness
 37. Dynamic Programming
 38. Impulsive control
 39. Thrusters
 40. Generalized discounted cost function
 41. Value Function
 42. Bellman's principle of optimality
 43. Bellman Equations
 44. Hamiltonian
 45. Q-Function
 46. Value Iteration
 47. Least-Square
 48. Kronecker Product
 49. Batch Least Square
 50. Persistent Excitation condition
 51. Semi-major axis
 52. Eccentricity
 53. Inclination
 54. Argument of Perigee
 55. of ascending node

Control neural, fuzzy Adapt. approaches, pp. 493–525, 1992.

- [50] A. Al-Tamimi, F. L. Lewis, and M. Abu-Khalaf, "Model-free Q-learning designs for linear discrete-time zero-sum games with application to H-infinity control," *Automatica*, vol. 43, no. 3, pp. 473–481, 2007, doi: 10.1016/j.automatica.2006.09.019.

پی نوشت:

-
1. Reconfiguration
 2. Quadratic-Programming
 3. Mixed-Integer-Linear Programming
 4. Relative orbit elements
 5. State-dependent-Riccati equations
 6. Convex optimization
 7. Collision avoidance
 8. Evolutionary
 9. Lagrange points
 10. Atmospheric drag
 11. Distributed
 12. Artificial Potential Function
 13. Inverse dynamics
 14. Q-Learning
 15. Reinforcement Learning
 16. Machine Learning
 17. Self-tuning
 18. Docking
 19. Rendezvous
 20. Model-based
 21. Discount factor
 22. Reference trajectory
 23. Offline
 24. Online
 25. Real-time
 26. Discounted
 27. Generalized
 28. Follower
 29. Leader
 30. Local-Vertical-Local-Horizontal
 31. Euler-Hill
 32. Earth-Centered Inertial
 33. True-Anomaly
 34. Gravitational constant
 35. Mean orbital motion

۶۰

سال ۱۱ - شماره ۱

بهار و تابستان ۱۴۰۱

نشریه علمی دانش و فناوری هوا فضا



طراحی کنترل کننده تطبیقی بهینه بدون مدل برای پیگردینی
مجدد آرایش پروازی ماهواره‌ها با یادگیری تقویتی