مکانیک هوافضا/ سال ۱۴۰۲/ دوره ۱۹/ شماره ۱/ صفحه ۶۱-۷۹

نشریه علمی مکانیک هوافضا

DOR: 20.1001.1.26455323.1402.19.1.5.0

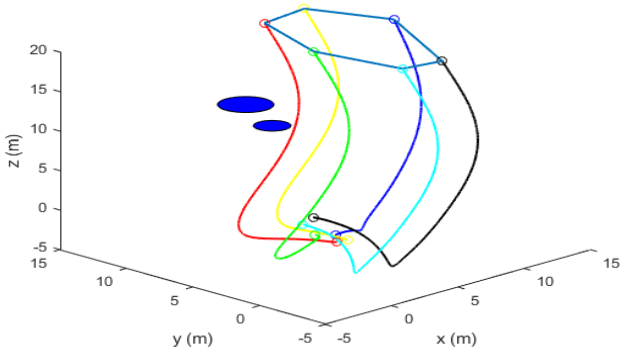
کنترل آرایش گروهی بهینه پرنده‌های بدون سرنشین با قید عدم برخورد و دینامیک ناشناخته

فاطمه مهدوی گلمیشه، سعید شمقدری[۲*]

[۱] دانشجوی دکتری، دانشکده مهندسی برق، دانشگاه علم و صنعت ایران، تهران، ایران

[۲] دانشیار، دانشکده مهندسی برق، دانشگاه علم و صنعت ایران، تهران، ایران

چکیده گرافیکی



برجسته‌ها

- کنترل آرایش گروهی توزیع‌شده سیستم چندپرنده بی‌سرنشین غیرخطی و ناهمگن
- ادغام CBF محلی با MARL برای تضمین قید عدم برخورد به روش داده محور
- ایجاد دو الگوریتم RL خارج از سیاست سری به ترتیب برای کنترل موقعیت و زاویه تا حصول آرایش گروهی بدون برخورد و مستقل از مدل

چکیده

این مقاله رویکرد آموزش توزیع‌شده‌ای را برای سیستمی با چند پرنده بدون سرنشین غیرخطی و ناهمگن، جهت حل مسئله کنترل آرایش گروهی ایمن و بهینه ارائه می‌نماید. هدف کنترل، تضمین ایمنی در حین دستیابی به عملکرد مطلوب است. برای این منظور دو کنترل‌کننده موقعیت و زاویه به‌صورت سری در نظر گرفته‌شده است. ابتدا، طراحی کنترل آرایش گروهی بهینه به‌عنوان عملکرد بهینه در کنترل موقعیت تعریف‌شده و توسط تابع هزینه مدل‌سازی می‌شود. در این مقاله، از طریق ادغام توابع هزینه با توابع کنترل مانع (Control Barrier Function (CBF)) محلی، مسائل بهینه‌سازی توزیع‌شده جدیدی معرفی می‌گردد. وجود CBF محلی در تابع هزینه افزوده موجب تضمین ایمنی در کنترل موقعیت شده و درنتیجه برخوردی در طول مسیر پرنده‌ها رخ نمی‌دهد. در روش ارائه‌شده، کنترل‌کننده‌های ایمن و بهینه موقعیت از حل مسائل بهینه‌سازی نامقید به‌جای مسائل بهینه‌سازی مقید به دست می‌آیند. در مرحله بعد، از کنترل موقعیت مجازی حاصل، زوایای مرجع به دست می‌آید. ردیابی بهینه این زوایا به‌عنوان عملکرد مطلوب در کنترل زاویه در نظر گرفته‌شده و با تابع هزینه مرتبط مدل‌سازی می‌شود. درنهایت، پایداری و ایمنی کنترل‌کننده‌های پیشنهادی اثبات می‌شود. این سیاست‌های بهینه و ایمن با استفاده از الگوریتم‌های یادگیری تقویتی چندعاملی (Multi-agent Reinforcement Learning)

مشخصات مقاله

تاریخچه مقاله:

نوع مقاله: علمی پژوهشی

دریافت: ۱۴۰۱/۰۶/۲۴

بازنگری: ۱۴۰۱/۰۷/۱۶

پذیرش: ۱۴۰۱/۰۹/۰۳

ارائه برخط: ۱۴۰۱/۰۹/۲۱

*نویسنده مسئول:

shamaghdari@iust.ac.ir

کلیدواژه‌ها:

پرنده بدون سرنشین

سیستم چندعاملی

کنترل آرایش گروهی

یادگیری تقویتی ایمن

یادگیری تقویتی مستقل از مدل

((MARL)) خارج از روال مرسوم، طراحی‌شده و به دانشی از دینامیک پرنده‌ها نیاز ندارد. الگوریتم‌های پیشنهادی از طریق شبیه‌سازی مسئله کنترل آرایش گروهی ۶ پرنده با قید عدم برخورد ارزیابی می‌شوند.

۱- مقدمه

پرنده‌های بدون سرنشین به دلیل اینکه کاربردهای مختلفی در امدادرسانی، شناسایی و نقشه‌برداری، صنعت کشاورزی و موارد دیگر می‌توانند داشته باشند، به‌طور گسترده‌ای در دهه‌های اخیر موردتوجه محققین قرارگرفته‌اند [۱]. علاوه بر این، استفاده از چند پرنده بدون سرنشین در کنار هم در انجام وظایف پیچیده موجب بهبود عملکرد جمعی می‌شود، به‌طور مثال، نظارت بر بلایای طبیعی، عملیات جستجوی گسترده و حمل‌ونقل‌های مشارکتی [۲ و ۳]. از مفاهیم کنترل سیستم چندعاملی و یا کنترل مشارکتی می‌توان برای پیاده‌سازی همکاری چند پرنده کمک گرفته و هر پرنده را به‌عنوان یک عامل فرض نمود. کنترل مشارکتی دارای زمینه‌های فعال تحقیقی متعددی مانند آرایش گروهی، اجماع [۴]، حرکت دسته‌جمعی [۵] و موارد دیگر می‌باشد که ازاین‌بین کنترل آرایش گروهی پژوهش‌های بسیاری را به خود اختصاص داده است. از کنترل آرایش گروهی برای حرکت عوامل در امتداد اشکال هندسی مشخص استفاده می‌شود [۶ و ۷]. ازجمله انواع کنترل آرایش گروهی می‌توان به آرایش گروهی متغیر بازمان [۸] و با روش رهبر-پیرو [۹] اشاره نمود. در کنترل آرایش گروهی با روش رهبر-پیرو یک رهبر دارای دینامیک موردنیاز است که همان‌طور که در [۹] نشان داده‌شده، می‌تواند مجازی باشد. در طراحی کنترل‌کننده آرایش گروهی با روش رهبر-پیرو [۱۰]، درحالی‌که دینامیک انتقال و چرخش برای گروهی از پرنده‌ها موردمطالعه قرارگرفته، درنهایت دینامیک غیرخطی آن‌ها ساده‌سازی شده است. در [۱۱]، هر پرنده بدون سرنشین یک سیستم غیرخطی با شش درجه آزادی در نظر گرفته‌شده است. به‌علاوه، آن مرجع مسئله کنترل آرایش گروهی را برای چند پرنده با دینامیک غیرخطی و جفت شده که در آن زیرسیستم‌های چرخشی و انتقالی اثر متقابلی بر دینامیک هم دارند، بررسی کرده است. بااین‌حال محیط در آن ایده‌آل و بدون مانع در نظر گرفته‌شده و لذا الگوریتم پیشنهادی آن ایمن نمی‌باشد؛ بنابراین مسئله کنترل آرایش گروهی مستقل از مدل چند پرنده با قید عدم برخورد، چالشی حل‌نشده است.

از طرفی چون ما در دنیایی پویا، پیچیده و غیرقابل‌پیش‌بینی زندگی می‌کنیم و دینامیک سیستم‌ها به‌طور فزاینده‌ای پیچیده می‌شوند، پیش طراحی رفتار یک عامل مستقل که بتواند با همه شرایط سازگار شود، اگر نگوییم غیرممکن، چالش‌برانگیز است. با توسعه هوش مصنوعی و نظریه کنترل، مطالعه بر هوش همکاری چندعاملی نیز شدت یافته است [۱۲]؛ بنابراین، محققان درحال‌توسعه الگوریتم‌های یادگیری تقویتی (RL) تک عاملی موجود به رویکردهای چندعاملی [۱۳] هستند. به‌علاوه، ظرفیت RL در مواجهه با طیف گسترده‌ای از مشکلات، با معماری ساده و بدون نیاز به دانش قبلی از دینامیک سیستم از مزایای اصلی آن می‌باشد

[۱۴ و ۱۵]. در صورت استفاده از الگوریتم RL در سیستم‌های چندعاملی، به آن الگوریتم، الگوریتم MARL می‌گویند. الگوریتم MARL در فرایند یادگیری، علاوه بر حالت خود عامل می‌تواند از حالت‌ها و ورودی‌های رهبر و همسایه‌های خود نیز استفاده نماید.

در مرور مراجع مرتبط به MARL با [۱۱، ۱۶ و ۱۷] روبه‌رو می‌شویم که دینامیک سیستم چندعاملی در آن‌ها غیرخطی در نظر گرفته‌شده است. در [۱۶] بدون نیاز به مدل سیستم، تابع مقدار و کنترل‌کننده تخمین زده‌شده و سیستم توسط سه شبکه شناساگر-منتقد-بازیگر شناسایی می‌گردد. یک الگوریتم تکرار سیاست چندعاملی برخط بر اساس شبکه منتقد در [۱۷] برای حل معادلات غیرخطی همیلتون-جاکوبی-جفت‌شده (HJC) استفاده‌شده است. باوجوداینکه در آن MARL برای سیستم‌های غیرخطی و ناهمگن بررسی گشته، الگوریتم پیشنهادی آن مبتنی بر مدل است. مسئله کنترل آرایش گروهی پرنده‌ها در [۱۱] توسط دو الگوریتم RL مستقل از مدل بررسی‌شده است. فرض ناهمگنی سیستم چندپرنده‌ای از این نظر حائز اهمیت است که در هر پرواز جرم محموله‌ها برای تیم پرنده ناشناخته بوده و بنابراین پارامترهای مرتبط با جرم در هرکدام از پرنده‌ها با یکدیگر متمایز و نامشخص است؛ بنابراین در این مقاله برای سیستم چند پرنده‌ای غیرخطی و ناهمگن، مسئله کنترل آرایش گروهی بهینه با تضمین قید عدم برخورد حل می‌گردد.

RL با چندین چالش روبه‌روست که یکی از آن‌ها برآورده شدن قیود ایمنی است. هنگام طراحی کنترل‌کننده برای سیستمی شامل چند پرنده بدون سرنشین در دنیای واقع که موانعی نیز در محیط آن وجود دارد، به دلیل احتمال برخورد پرنده‌ها با یکدیگر و با موانع، تضمین ایمنی و یا همان ارضای قید عدم برخورد در آن بسیار ضروری است. برای حل این مشکل، محققان اخیراً بر روی طراحی کنترل‌کننده‌هایی با استفاده از یادگیری تقویتی ایمن تمرکز نموده‌اند. برای این منظور از دو رویکرد کلی کنترل پیش‌بین (MPC) و CBF استفاده‌شده است. برای اطلاع بیشتر در مورد رویکرد MPC به [۱۸] مراجعه کنید. در رویکرد دیگر، CBF به‌عنوان یک تابع اضافه شونده به تابع هزینه یا یک قید ظاهر می‌شود که می‌تواند مبتنی بر مدل [۱۹] یا مستقل از مدل [۲۰] باشد. در [۲۰] با افزودن CBF به تابع مقدار، ایمنی به‌جای یک قید حالت به یک هدف کنترلی تبدیل می‌گردد. این کنترل‌کننده نه‌تنها می‌تواند بدون مدل دقیق طراحی شود، بلکه می‌تواند قیود ایمنی را نیز تضمین کند. مراجع متعددی در مورداستفاده از CBF در یادگیری تقویتی ایمن تک عاملی وجود دارد که اکثر آن‌ها مبتنی بر مدل هستند.

در کل، تحقیقات اندکی در زمینه MARL مقید وجود دارد. در [۲۱]، قید مسئله، قید اشباع ورودی است و برای ارضای آن یک عبارت غیر درجه دوم به تابع مقدار اضافه می‌شود. [۲۲] یک رویکرد کنترل آرایش گروهی متغیر بازمان برای یک سیستم چندعاملی خطی، با دینامیک ناشناخته و ناهمگن طراحی می‌کند. در آن مرجع، ورودی کنترل بهینه از حل یک مسئله بهینه‌سازی غیر درجه دوم با استفاده از RL خارج از سیاست با قید عدم برخورد به دست می‌آید. در مقایسه با [۲۲]، در این مقاله دینامیک عوامل غیرخطی بوده و از CBF برای تضمین ایمنی استفاده می‌گردد.

اهمیت موارد ذکرشده و جای کار زیادی که در کنترل سیستم‌های چندعاملی با استفاده از MARL ایمن وجود دارد، ما را برای انجام این پژوهش برانگیخت. یکی از مزیت‌های اصلی MARL ایمن، این است که نیازی به مدل دینامیکی دقیق عوامل وجود ندارد که این امر آن را برای سیستم‌های غیرخطی و پیچیده مانند سیستم چندپرنده بی‌سرنشین، بسیار کاربردی می‌سازد. به‌علاوه، این کنترل آرایش گروهی داده محور، ایمن نیز خواهد بود؛ بنابراین، سیاست کنترل بهینه داده محوری را برای سیستم چندپرنده بی‌سرنشین غیرخطی و ناهمگن، با تضمین قیود عدم برخورد و عملکرد آرایش گروهی پیشنهاد می‌دهیم. این مقاله ایمنی و پایداری رویکرد پیشنهادی را نیز بررسی می‌کند. آنچه این پژوهش را از سایر مراجع موجود در این زمینه متمایز می‌کند، مسائل اساسی موردمطالعه در آن است که به شرح ذیل می‌باشد:

۱) طراحی کنترل‌کننده موقعیت توزیع‌شده به‌منظور کنترل آرایش گروهی سیستم چندپرنده بی‌سرنشین غیرخطی و ناهمگن

۲) ادغام CBF محلی با MARL برای تضمین قید عدم برخورد به روش داده محور
۳) ایجاد دو الگوریتم MARL سری به ترتیب برای به دست آوردن کنترل‌کننده موقعیت و زاویه تا حصول کنترل آرایش گروهی بدون برخورد و مستقل از مدل

ساختار بقیه این مقاله به شرح زیر است. بخش ۲ اطلاعات اولیه‌ای را در مورد دینامیک مدل و تعریف مسئله ارائه می‌دهد. بخش ۳ نحوه طراحی کنترل‌کننده آرایش گروهی و تجزیه‌وتحلیل پایداری و ایمنی را نشان می‌دهد. در این بخش به ترتیب پس از طراحی کنترل بهینه و ایمن موقعیت با قید عدم برخورد و طراحی کنترل بهینه زاویه از الگوریتم MARL برای مستقل از مدل نمودن آن‌ها استفاده می‌شود. سپس، به تحلیل پایداری و تضمین قید عدم برخورد پرداخته می‌شود. نتایج شبیه‌سازی در بخش ۴ نشان می‌دهد که راهبرد کنترل پیشنهادی مؤثر است. بخش ۵ این مقاله مربوط به نتیجه‌گیری می‌باشد.

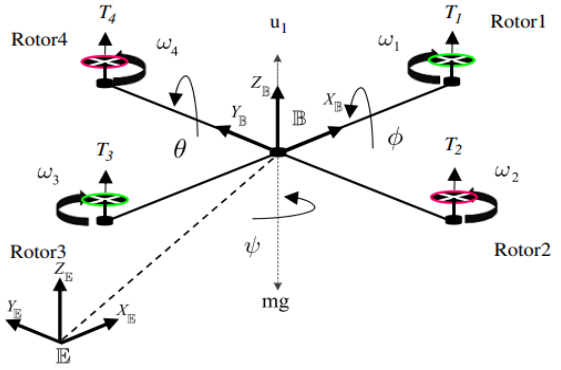
نمادهای استفاده‌شده در این مقاله استاندارد هستند. به‌طوری‌که $I_N \in \mathbb{R}^{N\times N}$ و $0_{m\times n} \in \mathbb{R}^{m\times n}$ به ترتیب نشان‌دهنده ماتریس $N \times N$ همانی و ماتریس صفر با ابعاد $m \times n$ است. بردارهای N بعدی 1_N و $e_{N,j}$ به ترتیب برداری با مقدار ۱ به ازای همه درایه‌ها و برداری با مقدار ۱ به ازای درایه jام و بقیه درایه‌های صفر را نشان می‌دهند. علامت $\otimes$ نشان‌دهنده ضرب کرونکر می‌باشد. تابع $\mathrm{vec}(X)$ عناصر موجود در ماتریس X را به یک بردار ستونی تغییرشکل می‌دهد. $\mathrm{int}(\Lambda)$ نمایانگر قسمت داخلی مجموعه Λ و $\partial(\Lambda)$ نشان‌دهنده مرز آن است. C^1 مجموعه توابع مشتق‌پذیر پیوسته را نشان می‌دهد.

۲- معرفی و فرمول‌بندی مسئله

در این بخش پس از ارائه دینامیک پرنده بدون سرنشین، به تعریف مسئله می‌پردازیم.

۲-۱- توصیف مدل دینامیکی پرنده بدون سرنشین

در این مقاله، مطابق شکل ۱، هر پرنده بدون سرنشین به‌صورت یک قاب ثابتی در نظر گرفته‌شده است که چهار روتور را به هم وصل می‌نماید. برای این منظور بردار موقعیت هر پرنده i به ازای $i \in \mathcal{I}$ در دستگاه مختصات اینرسی ثابت $\mathbb{E}$ به‌صورت $p_i = [x_i \ y_i \ z_i]^T \in \mathbb{R}^3$ تعریف‌شده و بردار زوایای اویلر آن، $\Theta_i = [\phi_i \ \theta_i \ \psi_i]^T \in \mathbb{R}^3$ می‌باشد، که در آن ϕ_i زاویه چرخش، θ_i زاویه فراز و ψ_i زاویه سمت هستند.



شکل (۱): پیکربندی پرنده بدون سرنشین [۲۳].

با استفاده از [۱۱]، برای توصیف مدل دینامیکی موقعیت و زاویه این پرنده داریم:

$$m_i\ddot{p}_i = R_{fi}(\Theta_i)f_i \quad (۱)$$

$$1J_i\ddot{\Theta}_i = -C_i(\Theta_i,\dot{\Theta}_i)\dot{\Theta}_i + \tau_i \quad (۲)$$

که در آن $C_i(\Theta_i,\dot{\Theta}_i) \in \mathbb{R}^{3\times3}$، $R_{fi} \in SO(3)$ و $J_i \in \mathbb{R}^{3\times3}$ به ترتیب ماتریس غیرخطی کوریولیس، ماتریس دوران و ماتریس اینرسی پرنده iام می‌باشند که در [۲۴] به‌تفصیل بیان‌شده‌اند. همچنین جرم پرنده iام با m_i نشان داده می‌شود. $f_i \in \mathbb{R}^3$ نشان‌دهنده برآیند نیروهای وارده به پرنده iام در دستگاه مختصات بدنه $\mathbb{B}$ و به‌صورت

$$f_i = [0 \ 0 \ T_{pi}]^T - R_{fi}^T[0 \ 0 \ m_i g]^T \quad (۳)$$

بوده و g در (۳) ثابت جاذبه زمین است. $T_{pi} \in \mathbb{R}$ نیروی بالابرنده کلی پرنده iام است که با توجه به شکل ۱ از رابطه $T_{pi} = T_1 + T_2 + T_3 + T_4$ و به ازای هر $i \in \mathcal{I}$ به دست می‌آید. $\tau_i = [\tau_{\phi_i} \ \tau_{\theta_i} \ \tau_{\psi_i}]^T$ بردار گشتاور پرنده iام در دستگاه مختصات بدنه $\mathbb{B}$ است. نیروی بالابرنده کلی T_{pi} و بردار گشتاور τ_i به ترتیب در روابط زیر صدق می‌کنند.

$$T_{pi} = k_\omega u_{z_i} \quad (۴)$$

$$\tau_i = [l_\tau k_\omega u_{\phi_i} \ l_\tau k_\omega u_{\theta_i} \ k_t u_{\psi_i}]^T \quad (۵)$$

در روابط (۴) و (۵)، k_ω، k_t و l_τ مقادیر مثبت ثابتی هستند. از (۱) و (۲) می‌توان مشاهده نمود که دینامیک موقعیت متأثر از زوایای اویلر بوده و لذا هر پرنده یک سیستم رباتیک غیرخطی و جفت شده با شش درجه آزادی و چهار ورودی است. درواقع به دلیل تعداد کمتر ورودی نسبت به درجه آزادی، این سیستم زیر تحریک می‌باشد. به‌علاوه در عمل می‌توان از یک سیستم توزیع توان برای توزیع ورودی‌های کنترلی u_{j_i} $(j = z, \phi, \theta, \psi)$ به چهار روتور استفاده نمود، به‌طوری‌که در آن‌ها روابط زیر به ازای سرعت چرخش روتور jام، ω_{j_i} $(j = 1,2,3,4)$ برآورده شوند.

$$\begin{aligned} u_{z_i} &= \omega_{1_i}^2 + \omega_{2_i}^2 + \omega_{3_i}^2 + \omega_{4_i}^2 \\ u_{\phi_i} &= \omega_{2_i}^2 - \omega_{4_i}^2 \\ u_{\theta_i} &= \omega_{1_i}^2 - \omega_{3_i}^2 \\ u_{\psi_i} &= \omega_{1_i}^2 - \omega_{2_i}^2 + \omega_{3_i}^2 - \omega_{4_i}^2 \end{aligned} \qquad (۶)$$

در رابطه (۶)، u_{z_i}، u_{ϕ_i}، u_{θ_i} و u_{ψ_i} ورودی‌های کنترلی پرنده iام هستند. بنابراین می‌توان سرعت‌های چرخشی موردنظر هرکدام از روتورها را طوری به دست آورد تا ورودی‌های کنترلی مطلوب یا طراحی‌شده را تولید کرده و حرکت پرنده را در شش درجه آزادی کنترل نماید.

برای تبدیل معادلات دینامیکی (۱) و (۲) به فرم معادلات حالت متناسب با آن‌ها، بردار حالت خطی و زاویه‌ای را به ترتیب به‌صورت $P_i = [p_i^T \ \dot{p}_i^T]^T \in \mathbb{R}^6$ و $\Sigma_i = [\Theta_i^T \ \dot{\Theta}_i^T]^T \in \mathbb{R}^6$ در نظر می‌گیریم. بنابراین از رابطه (۱) و (۲) داریم:

$$\dot{P}_i = F(P_i) + B_{P_i} u_{p_i} \qquad (۷)$$

$$\dot{\Sigma}_i = A_i(\Sigma_i)\Sigma_i + B_{\Sigma_i} u_{\Sigma_i} \qquad (۸)$$

که در آن‌ها

$$\begin{aligned} F(P_i) &= \begin{bmatrix} \dot{p}_i^T & -g e_{3,3}^T \end{bmatrix}^T \in \mathbb{R}^6 \\ B_{P_i} &= m_i^{-1} k_\omega [0_{3\times3} \quad I_3]^T \in \mathbb{R}^{6\times3} \\ u_{\Sigma_i} &= \begin{bmatrix} u_{\phi_i} & u_{\theta_i} & u_{\psi_i} \end{bmatrix}^T \\ B_{\Sigma_i} &= [0_{3\times3} \quad diag(l_\tau k_\omega, l_\tau k_\omega, k_t) J_i^{-1}]^T \in \mathbb{R}^{6\times3} \\ A_i(\Sigma_i) &= \begin{bmatrix} 0_{3\times3} & I_3 \\ 0_{3\times3} & -J_i^{-1} C_i(\Theta_i, \dot{\Theta}_i) \end{bmatrix} \in \mathbb{R}^{6\times6} \end{aligned}$$

می‌باشد. u_{p_i} ورودی کنترل موقعیت مجازی است که به شکل زیر در نظر گرفته می‌شود.

$$u_{p_i} = \begin{bmatrix} u_{p_i,x} & u_{p_i,y} & u_{p_i,z} \end{bmatrix}^T = R_{fi} e_{3,3} u_{z_i} \qquad (۹)$$

در عمل برای آرایش گروهی چند پرنده، جرم m_i و ماتریس اینرسی J_i متعلق به هرکدام از آن‌ها، به دلیل بارهای متفاوتی که حمل می‌کنند متفاوت است، که این مورد باعث ناهمگنی سیستم می‌شود. به‌علاوه، همان‌طور که در [۲۵] اشاره‌شده است، در هر پرواز اندازه محموله‌ها برای تیم پرنده ناشناخته می‌باشد، که این امر منجر به ناشناختگی مقادیر مرتبط با جرم (مانند m_i و J_i) در کاربردهای عملی می‌شود. بنابراین، سیستم چندعاملی را ناهمگن در نظر گرفته و از الگوریتم RL مستقل از مدل که نیاز به دانش دقیقی از پارامترهای دینامیکی پرنده‌ها ندارد، برای طراحی کنترل‌کننده استفاده می‌نماییم.

۲-۲- تعریف مسئله

هدف اصلی این مقاله طراحی یک سیاست کنترلی توزیع‌شده بهینه برای سیستم (۷)، به‌منظور پیروی از مسیر رهبر مجازی با حفظ آرایش گروهی معین، به‌گونه‌ای است که هیچ برخوردی بین پرنده‌ها با یکدیگر و با موانع اتفاق نیفتد.

آرایش گروهی مطلوب پرنده‌ها به ازای $(i, j \in \mathcal{I})$ با بردار $P_{ij,d} = \begin{bmatrix} p_{ij,d}^T & \dot{p}_{ij,d}^T \end{bmatrix}^T \in \mathbb{R}^6$ تعیین می‌گردد؛ که در آن $p_{ij,d} = \begin{bmatrix} x_{ij,d} & y_{ij,d} & z_{ij,d} \end{bmatrix}^T$ نشان‌دهنده موقعیت نسبی مطلوب بین پرنده iام و پرنده jام بوده و $\dot{p}_{ij,d} \in \mathbb{R}^3$ سرعت نسبی مطلوب بین پرنده iام و پرنده jام را نشان می‌دهد. چون در این مقاله شکل آرایش گروهی مطلوب ثابت فرض شده، در نتیجه $\dot{p}_{ij,d} = 0_{3\times1}$ است. $P_0 = [p_0^T \quad \dot{p}_0^T]^T \in \mathbb{R}^6$ بردار حالت رهبر مجازی است، به طوری‌که $p_0 = [x_0 \ y_0 \ z_0]^T$ نشان‌دهنده موقعیت مطلوب مرکز آرایش گروهی و یا همان مسیر حرکت رهبر مجازی در دستگاه مختصات $\mathbb{E}$ بوده و $\dot{p}_0 \in \mathbb{R}^3$ را به‌عنوان سرعت حرکت رهبر مجازی در آن دستگاه مختصات در نظر می‌گیریم. اگر به ازای $(i \in \mathcal{I})$، $P_{i,d} = \begin{bmatrix} p_{i,d}^T & \dot{p}_{i,d}^T \end{bmatrix}^T \in \mathbb{R}^6$ بردار حالت مطلوب پرنده iام نسبت به بردار حالت رهبر مجازی P_0 باشد، برای $P_{ij,d}$ داریم $P_{ij,d} = P_{i,d} - P_{j,d}$. خطای ردیابی آرایش گروهی به‌صورت زیر تعریف می‌شود:

$$z_i(t) = P_i(t) - P_0(t) - P_{i,d}(t) \qquad (۱۰)$$

تعریف ۱ [۲۶]: کنترل آرایش گروهی سیستم چندعاملی در صورت برآورده شدن معادلات زیر، حاصل می‌شود:

$$\lim_{t\to\infty}\|z_i(t)\| = \lim_{t\to\infty}\|P_i(t) - P_0(t) - P_{i,d}(t)\| = 0, \quad i = 1,2,\dots,N \tag{۱۱}$$

بنابراین موقعیت پرنده iام از $P_i(t) = P_0(t) + P_{i,d}(t)$ به دست می‌آید. از طرفی در عمل رهبر مجازی برای همه پرنده‌ها در دسترس نبوده و لذا $P_0(t)$ مستقیماً برای آن‌ها موجود نیست. در این مقاله برای غلبه بر این محدودیت از اطلاعات پرنده‌های همسایه و خود پرنده، برای به دست آوردن $P_i(t)$ استفاده‌شده است. همچنین ازآنجاکه مدل‌سازی تبادل اطلاعات بین عامل‌ها در سیستم‌های چندعاملی توسط گراف‌های جهت‌دار بسیار متداول است، می‌توان برای مطالعه بیشتر به [۲۷] مراجعه نمود. بنابراین، با استفاده از گراف ارتباطی، بردار خطای ردیابی آرایش گروهی پرنده iام $\delta_i(t) \in \mathbb{R}^6$ به‌صورت زیر تعریف می‌شود.

$$\delta_i = \sum_{j\in N_i} a_{ij}\left(P_i - P_j - P_{ij,d}\right) + b_i\left(P_i - P_0 - P_{i,d}\right) \tag{۱۲}$$

که در آن a_{ij} درایه مجاورت گراف ارتباطی را نشان داده و b_i نشان‌دهنده وضعیت ارتباط پرنده iام و رهبر مجازی است. این مقادیر در صورت برقراری اتصال مربوطه، مقداری مثبت داشته و در غیر این صورت مقدار آن‌ها صفر است. با جایگذاری (۱۰) در (۱۲) داریم:

$$\delta_i = \sum_{j\in N_i} a_{ij}\left(z_i - z_j\right) + b_i(z_i) \tag{۱۳}$$

با استفاده از تعریف ۱ نتیجه می‌گیریم که اگر و تنها اگر $\|z_i(t)\| \to 0$، پس $\|\delta_i(t)\| \to 0$. از (۷) و (۱۲) به دست می‌آید

$$\dot{\delta}_i = (d_i + b_i)\left(F(P_i) + B_{P_i}u_{p_i}\right) - \sum_{j\in N_i} a_{ij}\left(F(P_j) + B_{P_j}u_{p_j} + \dot{P}_{ij,d}\right) - b_i\left(\dot{P}_0 + \dot{P}_{i,d}\right) \tag{۱۴}$$

که در آن $d_i = \sum_{j\in N_i} a_{ij}$ درجه ورودی پرنده iام بوده و با فرض ثابت بودن الگوی آرایش گروهی، $\dot{P}_{ij,d} = \dot{P}_{i,d} = 0_{6\times 1}$ است. بنابراین داریم:

$$\dot{\delta}_i = (d_i + b_i)\left(F(P_i) + B_{P_i}u_{p_i}\right) - \sum_{j\in N_i} a_{ij}\left(F(P_j) + B_{P_j}u_{p_j}\right) - b_i\dot{P}_0 \tag{۱۵}$$

در ادامه به طراحی کنترل‌کننده آرایش گروهی خواهیم پرداخت.

۳- طراحی کنترل‌کننده آرایش گروهی و تحلیل پایداری و ایمنی

در این بخش، ابتدا از رویکرد بهینه و ایمنی برای استخراج پاسخ مناسب در معادلات دینامیک موقعیت (۷) استفاده می‌شود. سپس با استفاده از رویکرد بهینه در معادلات دینامیک زاویه (۸) راه‌حل بهینه به دست می‌آید. به دلیل ناشناخته بودن دینامیک پرنده‌ها در کاربردهای عملی، به دست آوردن مستقیم پاسخ بهینه دشوار بوده و درنتیجه با استفاده از داده‌های سیستم، الگوریتم‌های RL برای یادگیری کنترل‌کننده‌های بهینه در هر دو معادلات دینامیک موقعیت و زاویه به کار گرفته‌شده‌اند. در انتها نیز به تحلیل معیارهای پایداری و ایمنی می‌پردازیم.

۳-۱- طراحی کنترل‌کننده بهینه موقعیت با قید عدم برخورد

در این بخش ابتدا به بیان مسئله بهینه‌سازی مقید پرداخته، سپس با استفاده از CBF آن را به مسئله بهینه‌سازی بدون قید تبدیل می‌نماییم. به ازای ورودی‌های مرتبط با همسایه‌های پرنده iام $u_{p_{N_i}} = \left\{u_{p_j}\middle| j \in N_i\right\}$، تابع هزینه آن را داریم:

$$r_i\left(\delta_i, u_{p_i}, u_{p_{N_i}}\right) = \delta_i^T Q_i \delta_i + u_{p_i}^T R_{ii} u_{p_i} + \sum_{j\in N_i} u_{p_j}^T R_{ij} u_{p_j} \tag{۱۶}$$

که در آن ماتریس‌های وزنی $Q_i \geq 0$، $R_{ii} > 0$ و $R_{ij} > 0$ متقارن و معین می‌باشند. بنابراین شاخص عملکرد محلی پرنده iام عبارت است از:

$$J_i\left(\delta_i(t), u_{p_i}, u_{p_{N_i}}\right) = \int_t^\infty \left(r_i\left(\delta_i, u_{p_i}, u_{p_{N_i}}\right)\right) d\tau \tag{۱۷}$$

به ازای هر ورودی قابل‌قبول $u_{p_i} \in U_{p_i}$، تابع مقدار معادل تابع عملکرد بوده و $J_i\left(\delta_i, u_{p_i}, u_{p_{N_i}}\right) = V_i\left(\delta_i, u_{p_i}, u_{p_{N_i}}\right)$ است. موانع را به‌صورت کروی در نظر می‌گیریم، به‌طوری‌که $\mathbb{O} = \{o_1, o_2, \dots, o_l\}$ و $\mathbb{R}_o = \{r_{o_1}, r_{o_2}, \dots, r_{o_l}\}$ به ترتیب مجموعه موقعیت مرکز موانع و شعاع ایمن آن‌ها بوده و از پیش معین هستند. پس ورودی‌ای که مسئله بهینه‌سازی

مقید زیر را حل نماید، تابع مقدار را بهینه و از برخورد بین عوامل با یکدیگر و با موانع جلوگیری می‌کند.

$$\underset{u_{p_i}\in U_{p_i}}{\text{minimize}}\left(V_i\left(\delta_i(t),u_{p_i},u_{p_{N_i}}\right)=\int_t^{\infty} r_i\left(\delta_i,u_{p_i},u_{p_{N_i}}\right)d\tau\right) \quad (۱۸ \text{ الف})$$

$$s.t.\ (15),\delta_i(0)=\delta_{i0} \quad (۱۸ \text{ ب})$$

$$e_{io_k}\geq r_{si}+r_{o_k},\qquad o_k\in\mathbb{O} \quad (۱۸ \text{ ج})$$

$$e_{ij}\geq r_{si}+r_{sj},\qquad j\in N_i \quad (۱۸ \text{ د})$$

که در آن $e_{io_k}=\|p_i-o_k\|$ و $e_{ij}=\|p_i-p_j\|$ به ترتیب فاصله پرنده iام از مانع kام و پرنده jام را نشان می‌دهند. r_{s_i} نشان‌دهنده شعاع ایمن پرنده iام است. بنابراین، (۱۸ ج) و (۱۸ د) قیود ایمنی بوده و ازاین‌رو مجموعه ایمن با استفاده از این قیود نامساوی تشکیل می‌شود. مجموعه ایمن با حذف مناطق ناامن، ازنظر ریاضی به‌صورت زیر تعریف می‌شود:

$$S_i=\{p_i|\|p_i-p_j\|-(r_{si}+r_{sj})\geq 0, j\in N_i \quad \|p_i-o_k\|-(r_{si}+r_{o_k})\geq 0, o_k\in\mathbb{O}\} \quad (۱۹)$$

از [۲۰] می‌دانیم، که یک گزینه مناسب برای CBF، تابعی است که مقدارش در مجموعه ایمن S_i، مثبت بوده و با نزدیک شدن به مرز آن به بی‌نهایت نزدیک شود. به‌علاوه، به دلیل منفی بودن مقدار مشتق CBF در نزدیکی مرز ایمنی، مقدار CBF هرگز به بی‌نهایت میل نمی‌کند. بنابراین، به ازای حالت اولیه موجود در مجموعه S_i، وجود CBF به این معنی است که حالت آینده نیز در آن مجموعه قرار خواهد داشت. تعریف ریاضی خواص ارائه‌شده برای CBF به شرح زیر است.

تعریف ۲ [۲۰]: (ویژگی‌های CBF) برای یک سیستم کنترل، تابع $B_i: S_i\to\mathbb{R}$ یک تابع کنترل مانع برای مجموعه (۱۹) است، اگر توابع لیپشیتز محلی از کلاس $\mathcal{K}$، α_1، α_2 و α_3 وجود داشته باشند به‌طوری‌که

$$\frac{1}{\alpha_1(h(x))}\leq B_i(x)\leq\frac{1}{\alpha_2(h(x))},\forall x\in int(S_i) \quad (۲۰)$$

$$\dot{B}_i(x)\leq\alpha_3(h(x)),\qquad\forall x\in int(S_i) \quad (۲۱)$$

بوده و $h(x)\geq 0$ نشان‌دهنده قید ایمنی است.

برای ارزیابی نامساوی (۲۱) به‌عنوان قید، به‌طور کامل به دینامیک سیستم نیاز داریم، چراکه $\dot{B}_i(x)=\partial B_i/\partial x\,(\dot{x})$ بوده و اطلاعات $\dot{x}$ لازم است. برای رهایی از این الزام در طراحی سیاست کنترلی ایمن و بهینه، از ادغام RL با مفهوم CBF استفاده می‌شود.

با استفاده از ویژگی‌های تابع مانع توضیح داده‌شده، برای اطمینان از ایمنی و تضمین قیود عدم برخورد، با افزودن $B_i(p_i,p_{N_i})$ به‌عنوان تابع مانع به تابع هزینه، آن را گسترش می‌دهیم. به‌طوری‌که $p_{N_i}=\{p_j|j\in N_i\}$ مجموعه موقعیت همسایه‌های پرنده iام می‌باشد. بنابراین مسئله بهینه‌سازی جدید و بدون قید از بازنویسی رابطه (۱۸) به شکل زیر حاصل می‌شود.

$$\underset{u_{p_i}\in U_{p_i}}{\text{minimize}}\left(V_{ai}\left(\delta_i,u_{p_i},u_{p_{N_i}},p_i,p_{N_i}\right)=\int_t^{\infty} r_{ai}\left(\delta_i,u_{p_i},u_{p_{N_i}},p_i,p_{N_i},\tau\right)d\tau\right) \quad (۲۲ \text{ الف})$$

$$s.t.\ (15),\qquad\delta_i(0)=\delta_{i0} \quad (۲۲ \text{ ب})$$

که در آن تابع هزینه افزوده به‌صورت زیر تعریف‌شده

$$r_{ai}\left(\delta_i(t),u_{p_i},u_{p_{N_i}},p_i,p_{N_i}\right)=r_i\left(\delta_i,u_{p_i},u_{p_{N_i}}\right)+B_i(p_i,p_{N_i}) \quad (۲۳)$$

و

$$B_i(p_i,p_{N_i})=\sum_{o_k\in\mathbb{O}}B_{io_k}+\sum_{j\in N_i}B_{ij}. \quad (۲۴)$$

می‌باشد. که در آن B_{io_k} و B_{ij} توابع مانعی هستند که به ترتیب از برخورد بین پرنده iام و مانع o_k و پرنده jام جلوگیری می‌کنند. از مجموع این توابع مانع، $B_i(p_i,p_{N_i})$ به دست می‌آید که CBF محلی پرنده iام نامیده می‌شود. از کاندیداهای مناسب برای این CBFهای محلی داریم:

$$B_{io_k}(p_i)=-\log\left(\frac{\gamma_i\left(\|p_i-o_k\|-(r_{si}+r_{o_k})\right)}{\gamma_i\left(\|p_i-o_k\|-(r_{si}+r_{o_k})\right)+1}\right) \quad (۲۵)$$

$$B_{ij}(p_i,p_j)=-\log\left(\frac{\gamma_i\left(\|p_i-p_j\|-(r_{si}+r_{sj})\right)}{\gamma_i\left(\|p_i-p_j\|-(r_{si}+r_{sj})\right)+1}\right) \quad (۲۶)$$

که در آن پارامترهای $\gamma_i>0$ تسلط نسبی CBFهای محلی بر تابع هزینه و اهمیت ایمنی را نشان می‌دهد. به‌عبارت‌دیگر، این ضرایب تعیین می‌کنند که CBFهای محلی با نزدیک شدن به مرز ایمنی با چه سرعتی کاهش یابند. تابع مقدار افزوده بهینه توزیع‌شده و تابع HJC افزوده به ترتیب به شکل زیر توصیف می‌شوند.

$$V_{ai}^*\left(\delta(t),u_{p_i}^*,u_{p_{N_i}}^*,p_i,p_{N_i}\right) \quad (۲۷)$$

$$= \min_{u_{p_i}\in U_{p_i}} \int_t^{\infty} (\delta_i^T Q_i \delta_i + u_{p_i}^T R_{ii} u_{p_i} + \sum_{j\in N_i} u_{p_j}^{*T} R_{ij} u_{p_j}^* + B_i(p_i, p_{N_i}))d\tau$$

$$H_{ai}\left(\delta_i(t), \nabla V_{ai}, u_{p_i}, u_{p_{N_i}}, p_i, p_{N_i}\right) = r_{ai}\left(\delta_i, u_{p_i}, u_{p_{N_i}}, p_i, p_{N_i}\right) + \nabla V_{ai}^T \dot{\delta}_i = 0 \quad (۲۸)$$

که در آن $\nabla V_{ai} = (\partial V_{ai})/(\partial \delta_i)$ بردار گرادیان تابع مقدار افزوده بوده و $V_{ai}(0) = 0$. با توجه به نتایج بهینگی در کنترل کلاسیک، به ازای ورودی کنترل u_{p_i}ای که تابع مقدار افزوده را کمینه کند، تابع مقدار بهینه عامل iام به دست می‌آید. اگر گرادیان تابع مقدار بهینه را در تابع HJC (۲۸) قرار دهیم، نتیجه صفر حاصل می‌شود. لذا از آن رابطه می‌توان سیگنال کنترل بهینه را به دست آورد. پس H_{ai}^*، یعنی مقدار بهینه H_{ai}، به ازای ورودی کنترل بهینه

$$u_{p_i}^* = -0.5(d_i + b_i) R_{ii}^{-1} B_{p_i}^T \nabla V_{ai} \quad (۲۹)$$

و توسط تعریف زیر به دست می‌آید.

$$H_{ai}^* = H_{ai}\left(\delta_i(t), \nabla V_{ai}, u_{p_i}^*, u_{p_{N_i}}, p_i, p_{N_i}\right) \quad (۳۰)$$

نکته ۱: در مسئله بهینه‌سازی جدید (۲۲)، ایمنی به‌عنوان یک هدف کنترلی تعریف‌شده است. همان‌طور که از تابع هزینه افزوده (۲۳) برمی‌آید، B_i موجود در آن با نزدیک شدن عوامل به مرز ایمنی، ترم غالب می‌شود. درنتیجه B_i توسط کنترل‌کننده بهینه، کاهش‌یافته و با باقی ماندن عوامل در منطقه ایمن خود، درواقع بدون هیچ برخوردی مسیر موردنظر را دنبال می‌کنند. بنابراین، بدون اعمال هیچ‌گونه قیدی، $\dot{B}_i$ تمایل دارد که در نزدیکی مرز مقداری منفی داشته باشد. پس در حل مسئله بهینه‌سازی جدید می‌توان از روش‌های عددی مرسوم برای حل مسائل بهینه‌سازی نامقید استفاده نمود.

برخی از تعاریف و مفروضاتی که در بخش‌های بعد موردنیاز است، در ادامه مطرح می‌شود.

تعریف ۳: مجموعه ورودی‌های ایمن عامل iام به ازای موقعیت فعلی آن p_i، به‌صورت زیر تعریف‌شده است.

$$U_{c_i} = \left\{u_{p_i} \in \mathbb{R}^3 \middle| p_i^{u_{p_i}} \in int(S_i)\right\} \quad (۳۱)$$

که در آن $\text{int}(S_i)$ قسمت داخلی مجموعه ایمن تعریف‌شده در (۲۱) بوده و $p_i^{u_{p_i}}$ موقعیت عامل iام به ازای اعمال ورودی u_{p_i} است.

تعریف ۴: در مسئله کنترل بهینه (۲۲)، سیاست کنترلی عامل iام u_{p_i}، در صورتی قابل‌قبول است که سیستم (۱۵) را پایدار و تابع هزینه مربوطه را کران‌دار نماید.

نکته ۲: برای داشتن یک سیاست قابل‌قبول در تابع هزینه (۲۳)، هم r_i و هم B_i باید محدود باشند. به ازای ورودی کنترل شدنی $u_{p_i} \in U_i$، r_i محدود بوده و به ازای ورودی کنترل ایمن $u_{p_i} \in U_{c_i}$، B_i محدود است. بنابراین، در صورتی‌که $u_{p_i} \in U_i \cap U_{c_i}$ باشد، تابع هزینه (۲۳) کران‌دار خواهد بود. همچنین، هنگامی که $u_{p_i} \in U_i \cap U_{c_i}$، سپس $u_{p_i} \in U_i$ بوده و سیستم (۱۵) را پایدار می‌نماید. پس مجموعه ورودی‌های قابل‌قبول عامل iام حاصل از (۲۲) به‌صورت زیر تعریف شده است.

$$U_{a_i} = U_i \cap U_{c_i} \quad (۳۲)$$

فرض ۱: موقعیت اولیه هر پرنده i در داخل مجموعه ایمن S_i قرار داشته و

$$p_i(0) \in int(S_i). \quad (۳۳)$$

فرض ۲: برای (۲۲)، فرض بر این است که مجموعه ورودی‌های کنترلی قابل‌قبول هر پرنده i تهی نمی‌باشد، یعنی $U_{a_i} \neq \emptyset$. بنابراین یک سیاست کنترل اولیه $u_{p_i}(0) \in U_{a_i}$ به ازای هر موقعیت اولیه $p_i(0)$ وجود دارد.

از رابطه (۲۸) مشخص است، که به دلیل وجود دینامیک خطای ردیابی $\dot{\delta}_i$ در آن، نیاز به مشخص بودن دینامیک سیستم وجود دارد. از طرفی، فرض بر این است که دانش کاملی در مورد دینامیک سیستم در دسترس نیست، پس در ادامه از الگوریتم RL خارج از سیاست مستقل از مدل برای حل مسئله بهینه‌سازی (۲۲) استفاده می‌شود.

۲-۳- الگوریتم RL برای کنترل موقعیت

در این بخش یک الگوریتم RL با رویکرد تکرار سیاست، برای یافتن یک کنترل‌کننده بهینه و بدون نیاز به دانشی از دینامیک سیستم ارائه خواهد شد. در این رویکرد از دو سیاست مختلف، سیاست رفتار و سیاست هدف استفاده می‌شود. سیاست رفتار، سیاست ایمنی است که در ابتدا بر سیستم برای جمع‌آوری داده اعمال می‌شود و سیاست هدف، سیاستی است که با استفاده از داده‌های

جمع‌آوری‌شده در راستای سیاست بهینه به‌روز می‌شود. درنهایت، پس از اتمام یادگیری، سیاست ایمن و بهینه حاصل از سیاست هدف به سیستم اعمال می‌شود.

نسخه بسیار کوچک تابع هزینه افزوده، معادله بلمن زیر است،

$$0 = r_{ai} + J_{\delta_i}^T \dot{\delta}_i \tag{۳۴}$$

که در آن $J_{\delta_i}^T \dot{\delta}_i = \dot{J}_i$ است. برای استفاده از رویکرد خارج از سیاست، دینامیک سیستم (۱۵) باهدف تفکیک سیاست رفتار و سیاست هدف بازنویسی می‌شود. که داریم:

$$\begin{aligned}\dot{\delta}_i = &\sum_{j\in N_i} a_{ij}\left(F(P_i) - F(P_j) - B_{P_j}u_{p_j}\right)\\ &+b_i\left(F(P_i) - \dot{P}_0\right) + (d_i + b_i)B_{P_i}u_{p_i}^k\\ &+(d_i + b_i)B_{P_i}\left(u_{p_i} - u_{p_i}^k\right)\end{aligned} \tag{۳۵}$$

که در آن $u_{p_i}^k$ سیاست هدفی است که در الگوریتم به‌روز شده ولی به سیستم اعمال نمی‌شوند. سیاست پایدارساز و ایمن u_{p_i}، سیاست رفتار بوده و برای تولید داده یادگیری در الگوریتم RL به سیستم اعمال می‌شود. با استفاده از (۳۵) در (۳۴) داریم:

$$\begin{aligned}&J_{\delta_i}^{kT}[\sum_{j\in N_i} a_{ij}\left(F(P_i) - F(P_j) - B_{P_j}u_{p_j}\right)\\ &+b_i\left(F(P_i) - \dot{P}_0\right) + (d_i + b_i)B_{P_i}u_{p_i}^k\\ &+(d_i + b_i)B_{P_i}\left(u_{p_i} - u_{p_i}^k\right)]\\ &= -r_{ai}\left(\delta_i, u_{p_i}^k, u_{p_{N_i}}^*, p_i, p_{N_i}\right)\\ &+J_{\delta_i}^{kT}(d_i + b_i)B_{P_i}\left(u_{p_i} - u_{p_i}^k\right)\end{aligned} \tag{۳۶}$$

با انتگرال‌گیری از هر دو طرف رابطه فوق داریم:

$$\begin{aligned}&J_i^k\left(\delta_i(t)\right) - J_i^k\left(\delta_i(t-T)\right)\\ &= -\int_{t-T}^{t} r_{ai}\left(\delta_i, u_{p_i}^k, u_{p_{N_i}}^*, p_i, p_{N_i}\right)d\tau\\ &+(d_i + b_i)\int_{t-T}^{t} J_{\delta_i}^{kT}B_{P_i}\left(u_{p_i} - u_{p_i}^k\right)d\tau\end{aligned} \tag{۳۷}$$

ورودی کنترلی $u_{p_i}^k$ با بهینه‌سازی بر روی تابع HJC (۲۸) به شکل زیر به‌روز می‌شود.

$$u_{p_i}^{k+1} = -0.5(d_i + b_i)R_{ii}^{-1}B_{p_i}^T J_{\delta_i}^k \tag{۳۸}$$

که از این عبارت داریم:

$$(d_i + b_i)J_{\delta_i}^{kT}B_{p_i} = -2u_{p_i}^{k+1^T}R_{ii} \tag{۳۹}$$

با جایگذاری (۳۹) در (۳۷) داریم:

$$\begin{aligned}&J_i^k\left(\delta_i(t)\right) - J_i^k\left(\delta_i(t-T)\right)\\ &= -\int_{t-T}^{t} r_{ai}\left(\delta_i, u_{p_i}^k, u_{p_{N_i}}^*, p_i, p_{N_i}\right)d\tau\\ &-\int_{t-T}^{t} 2u_{p_i}^{k+1^T}R_{ii}\left(u_{p_i} - u_{p_i}^k\right)d\tau\end{aligned} \tag{۴۰}$$

در معادله بلمن خارج از سیاست (۴۰)، هم سیاست کنترل $u_{p_i}^{k+1}$ و هم تابع مقدار J_i^k، به‌طور هم‌زمان، به ازای سیاست رفتار u_{p_i} و با استفاده از داده‌های جمع‌آوری‌شده، به‌روز می‌شوند. در این راه‌حل، الگوریتم RL خارج از سیاست با استفاده از یک ساختار بازیگر-منتقد، آموزش داده می‌شود که نیاز به دانشی از دینامیک سیستم ندارد. شبکه منتقد، تابع مقدار J_i^k و شبکه بازیگر، ورودی کنترل موقعیت $u_{p_i}^{k+1}$ را به شکل زیر تخمین می‌زنند.

$$\hat{J}_i^k(\delta_i) = \widehat{W}_{1i}{}^{kT}\Phi(\delta_i) \tag{۴۱}$$

$$\hat{u}_{p_i}^{k+1}(\delta_i) = \widehat{W}_{2i}{}^{kT}\theta(\delta_i) \tag{۴۲}$$

که در آن $\Phi = [\Phi_1\ \Phi_2\ \dots\ \Phi_{l_\Phi}] \in \mathbb{R}^{l_\Phi}$ توابع پایه شبکه منتقد و $\theta = [\theta_1\ \theta_2\ \dots\ \theta_{l_\theta}] \in \mathbb{R}^{l_\theta}$ توابع پایه شبکه بازیگر هستند. به‌علاوه $\widehat{W}_{1i}{}^k \in \mathbb{R}^{l_\Phi}$ و $\widehat{W}_{2i}{}^k \in \mathbb{R}^{l_\theta\times 3}$ بردارهای وزنی شبکه‌های عصبی مربوطه می‌باشند. توجه داشته باشید که $u_{p_i}^{k+1}$ توسط شبکه عصبی (۴۲) تخمین زده‌شده و نیاز به هیچ‌دانشی در مورد دینامیک سیستم ندارند. خطای معادله بلمن (۴۰) با جایگذاری (۴۱) و (۴۲) در آن به شکل زیر تعریف می‌شود.

$$\begin{aligned}&e_{p_i}^k(t)\\ &= \widehat{W}_{1i}{}^{kT}\Phi\left(\delta_i(t)\right) - \widehat{W}_{1i}{}^{kT}\Phi\left(\delta_i(t-T)\right)\\ &-\int_{t-T}^{t} -(\delta_i^T Q_i\delta_i + u_{p_i}^{kT}R_{ii}u_{p_i}^k\\ &+\sum_{j\in N_i} u_{p_j}^T R_{ij}u_{p_j} + B_i\left(p_i, p_{N_i}\right))d\tau\\ &+\int_{t-T}^{t} 2\left(\widehat{W}_{2_i}{}^{kT}\theta(\delta_i)\right)^T R_{ii}v_i^k d\tau\end{aligned} \tag{۴۳}$$

که در آن $v_i^k = \left[v_{1i}^k\ v_{2i}^k\ v_{3i}^k\right]^T = u_{p_i} - u_{p_i}^k$ است. با استفاده از ویژگی‌های ضرب کرونکر داریم:

$$\begin{aligned}&\int_{t-T}^{t} -(\delta_i^T Q_i\delta_i + u_{p_i}^{kT}R_{ii}u_{p_i}^k + \sum_{j\in N_i} u_{p_j}^T R_{ij}u_{p_j}\\ &+B_i\left(p_i, p_{N_i}\right))d\tau + e_{p_i}^k(t)\\ &= \widehat{W}_{1i}{}^{kT}\left(\Phi\left(\delta_i(t)\right) - \Phi\left(\delta_i(t-T)\right)\right)\\ &+vec\left(\widehat{W}_{2i}{}^k\right)^T\int_{t-T}^{t} 2\left(R_{ii}v_i^k\otimes\theta(\delta_i)\right)d\tau\end{aligned} \tag{۴۴}$$

از روش حداقل مربعات برای به دست آوردن کمینه خطای تقریب بلمن (۴۴) استفاده می‌نماییم. برای این منظور (۴۴) به فرم رگرسیون زیر بازنویسی می‌گردد.

$$y_{p_i}^k(t) + e_{p_i}^k(t) = \widehat{W}_{p_i}^{kT} h_{p_i}^k(t) \quad (۴۵)$$

که در آن $\widehat{W}_{pi}^k$، بردار $l_\Phi + 3l_\theta$ بعدی و متشکل از بردارهای وزنی

$$\widehat{W}_{pi}^{kT} = \left[\widehat{W}_{1i}{}^{kT}, vec\left(\widehat{W}_{2i}{}^{k}\right)^T\right] \quad (۴۶)$$

بوده و $h_{p_i}^k(t)$ و $y_{p_i}^k(t)$ به فرم زیر در نظر گرفته شده‌اند.

$$h_{p_i}^k(t) = \begin{bmatrix} \Phi(\delta_i(t)) - \Phi\big(\delta_i(t-T)\big) \\ 2\int_{t-T}^{t} \left(R_{ii} v_i^k \otimes \theta(\delta_i)\right) d\tau \end{bmatrix} \quad (۴۷)$$

$$y_{p_i}^k(t) = -\int_{t-T}^{t} \big(\delta_i^T Q_i \delta_i + u_{p_i}^{kT} R_{ii} u_{p_i}^k + \sum_{j\in N_i} u_{p_j}^T R_{ij} u_{p_j} + B_i(p_i, p_{N_i})\big) d\tau \quad (۴۸)$$

داده‌های موردنیاز را در M نقطه و در فاصله زمانی T جمع‌آوری می‌کنیم، تا با حل (۴۵)، $\widehat{W}_{pi}^{kT}$ را به دست آوریم. اطلاعات جمع‌آوری‌شده را در ماتریس‌های $H_{p_i}^k$ و $Y_{p_i}^k$ ذخیره می‌نماییم.

$$H_{p_i}^k = \left[h_{p_i}^k(t_1), \dots, h_{p_i}^k(t_M)\right]$$
$$Y_{p_i}^k = \left[y_{p_i}^k(t_1), \dots, y_{p_i}^k(t_M)\right]$$

بنابراین فرم رگرسیون (۴۵) عبارت است از:

$$\widehat{W}_{pi}^{kT} H_{p_i}^k = Y_{p_i}^k \quad (۴۹)$$

که با استفاده از حداقل مربعات، پاسخ آن به فرم به‌دست‌آمده

$$\widehat{W}_{pi}^k = \left(H_{p_i}^k H_{p_i}^{kT}\right)^{-1} H_{p_i}^k Y_{p_i}^{kT} \quad (۵۰)$$

و به ازای

$$M > l_\Phi + 3l_\theta \quad (۵۱)$$

جواب دارد.

نکته ۳: سیاست پایدارساز و ایمن اولیه u_{p_i} برای سیستم چندپرنده، از مجموع یک کنترل‌کننده PID متعارف پایدارساز و ایمن و نویز جستجو به دست می‌آید. برای طراحی این کنترل‌کننده نیاز به دانش کاملی از دینامیک سیستم وجود نداشته و با اطلاعات کلی این امر صورت می‌گیرد.

همگرایی الگوریتم ۱ به سیاست بهینه را می‌توان به روشی مشابه [۲۰] اثبات نمود.

الگوریتم ۱: الگوریتم MARL ایمن کنترل موقعیت

۱: مقداردهی اولیه شبکه‌های بازیگر و منتقد، (۴۱) و (۴۲)
۲: **مرحله ۱:** جمع‌آوری داده
۳: از هر سیاست رفتار پایدارساز و ایمن دارای نویز $u_{p_i} \in U_{a_i}$ تا زمانی که (۵۱) برای هر عامل برآورده شود، استفاده می‌نماییم. برای یادگیری ایمن، باید با استفاده از ورودی‌ها سیستم به مرز ایمنی نزدیک گردد.
۴: **پایان مرحله ۱**
۵: **مرحله ۲:** یادگیری سیاست هدف بهینه توزیع‌شده با استفاده از داده‌های جمع‌آوری‌شده
۶: تولید ماتریس‌های $y_{p_i}^k(t)$ و$h_{p_i}^k(t)$ به‌صورت (۴۷)، (۴۸) برای همه $t = t_1, \dots, t_M$ و با استفاده از $u_{p_i}^k$
۷: به‌روزرسانی $u_{p_i}^k$ (۴۲) و J_i^k (۴۱) با استفاده از وزن‌های NN حاصل از (۵۰)
۸: اگر معیار توقف برآورده شد، الگوریتم متوقف شود، در غیر این صورت $k = k + 1$ و برو به ۵
۹: **پایان مرحله ۲**

اگر $\Theta_{r_i} = \left[\phi_{r_i}\ \theta_{r_i}\ \psi_{r_i}\right]^T$ را به‌عنوان بردار زاویه مرجع پرنده iام در نظر بگیریم، پس از تعیین ورودی کنترل موقعیت مجازی u_{p_i} از الگوریتم ۱، ورودی کنترل بالابرنده کل مطلوب u_{z_i}، زاویه چرخش مطلوب ϕ_{r_i} و زاویه فراز مطلوب θ_{r_i} را می‌توان از (۱۱) به‌صورت زیر به‌دست آورد [۱۱].

$$u_{z_i} = \frac{u_{p_i,z}}{\cos(\theta_i)\cos(\phi_i)} \quad (۵۲)$$

$$\phi_{r_i} = \arcsin\left(\frac{\cos(\phi_i)\sin(\theta_i)\sin(\psi_i) - \frac{u_{p_i,y}}{u_{z_i}}}{\cos(\psi_i)}\right) \quad (۵۳)$$

$$\theta_{r_i} = \arcsin\left(\frac{\frac{u_{p_i,x}}{u_{z_i}} - \sin(\phi_i)\sin(\psi_i)}{\cos(\phi_i)\cos(\psi_i)}\right) \quad (۵۴)$$

مقدار مرجع زاویه سمت ψ_{r_i} نیز به‌عنوان یک مقدار ثابت برای هماهنگی تیم پرنده‌ها در کاربردهای عملی فرض شده است. پس از به دست آمدن زوایای مطلوب از (۵۲-۵۴)،

طراحی یک کنترل‌کننده بهینه زاویه برای ردیابی زوایای مرجع مدنظر نیاز است.

۳-۳- طراحی کنترل‌کننده بهینه زاویه

در این بخش ابتدا مسئله رسیدن هر پرنده iام به زوایای مرجع حاصل از (۵۴-۵۲) را تعریف کرده، سپس مسئله بهینه‌سازی مطلوب را طراحی و حل می‌نماییم. بردار حالت زاویه‌ای مرجع پرنده iام را به‌صورت $\Sigma_{r_i} = \left[\Theta_{r_i}^T \ \dot{\Theta}_{r_i}^T\right]^T \in \mathbb{R}^6$ در نظر گرفته و خطای ردیابی زاویه را به‌صورت زیر تعریف می‌کنیم.

$$\varepsilon_i = \Sigma_i - \Sigma_{r_i} \tag{۵۵}$$

دینامیک خطای ردیابی زاویه‌ای پرنده iام با استفاده از (۸) به‌صورت زیر به دست می‌آید.

$$\dot{\varepsilon}_i = A_i(\Sigma_i)\Sigma_i + B_{\Sigma_i}u_{\Sigma_i} - \dot{\Sigma}_{r_i} \tag{۵۶}$$

همان‌طور که از (۵۶) مشخص است، خطای ردیابی زاویه‌ای هر پرنده تنها به خودش بستگی داشته و مستقل از دیگر پرنده‌ها می‌باشد. بنابراین، تابع هزینه پرنده iام را داریم:

$$\begin{aligned}\bar{r}_i(\varepsilon_i, u_{\Sigma_i}) &= \|\varsigma_i(\tau)\|^2_{\bar{Y}_i} \\ &= \varepsilon_i^T\bar{Q}_i\varepsilon_i + u_{\Sigma_i}^T\bar{R}_{ii}u_{\Sigma_i}\end{aligned} \tag{۵۷}$$

که در آن ماتریس‌های وزنی $\bar{Q}_i \geq 0$ و $\bar{R}_{ii} > 0$ متقارن و مثبت معین می‌باشند. بنابراین شاخص عملکرد زاویه‌ای پرنده iام عبارت است از:

$$\bar{J}_i(\varepsilon_i, u_{\Sigma_i}) = \int_t^\infty \left(\varepsilon_i^T\bar{Q}_i\varepsilon_i + u_{\Sigma_i}^T\bar{R}_{ii}u_{\Sigma_i}\right)d\tau \tag{۵۸}$$

به ازای هر ورودی قابل‌قبول $u_{\Sigma_i} \in U_{\Sigma_i}$، تابع مقدار معادل تابع عملکرد است یعنی $\bar{J}_i(\varepsilon_i, u_{\Sigma_i}) = \bar{V}_i(\varepsilon_i, u_{\Sigma_i})$. پس ورودی‌ای که مسئله بهینه‌سازی زیر را حل نماید، تابع مقدار را بهینه می‌کند.

$$\begin{aligned}&\underset{u_{\Sigma_i}\in U_{\Sigma_i}}{\text{minimize}}(\bar{V}_i(\varepsilon_i, u_{\Sigma_i}) \\ &\qquad = \int_t^\infty (\varepsilon_i^T\bar{Q}_i\varepsilon_i + u_{\Sigma_i}^T\bar{R}_{ii}u_{\Sigma_i})d\tau)\end{aligned} \tag{۵۹ الف}$$

$$s.t.\ (56), \varepsilon_i(0) = \varepsilon_{i0} \tag{۵۹ ب}$$

تابع مقدار بهینه و تابع همیلتون-جاکوبی (HJ) مرتبط به ترتیب به شکل زیر توصیف می‌شوند.

$$\begin{aligned}\bar{V}_i^*(\varepsilon_i, u_{\Sigma_i}^*) = \min_{u_{\Sigma_i}\in U_{\Sigma_i}} \int_t^\infty &\left(\varepsilon_i^T\bar{Q}_i\varepsilon_i\right. \\ &\left. + u_{\Sigma_i}^T\bar{R}_{ii}u_{\Sigma_i}\right)d\tau\end{aligned} \tag{۶۰}$$

$$H_i(\varepsilon_i, \nabla\bar{V}_i, u_{\Sigma_i}) = \bar{r}_i(\varepsilon_i, u_{\Sigma_i}) + \nabla\bar{V}_i^T\dot{\varepsilon}_i = 0 \tag{۶۱}$$

که در آن $\nabla\bar{V}_i = (\partial\bar{V}_i)/(\partial\varepsilon_i)$ بردار گرادیان تابع مقدار بوده و $\bar{V}_i(0) = 0$. با توجه به نتایج بهینگی در کنترل کلاسیک، به ازای ورودی کنترل u_{Σ_i}ای که تابع مقدار را کمینه کند، تابع مقدار بهینه عامل iام به دست می‌آید. اگر گرادیان تابع مقدار بهینه را در تابع HJ (۶۱) قرار دهیم، نتیجه صفر حاصل می‌شود. لذا از آن رابطه می‌توان سیگنال کنترل بهینه را به دست آورد. پس H_i^*، یعنی مقدار بهینه H_i، به ازای ورودی کنترل زاویه بهینه

$$u_{\Sigma_i}^* = -0.5\bar{R}_{ii}^{-1}B_{\Sigma_i}^{\mathrm{T}}\nabla\bar{V}_i \tag{۶۲}$$

و توسط تعریف زیر به دست می‌آید.

$$H_i^* = H_i\left(\varepsilon_i(t), \nabla\bar{V}_i, u_{\Sigma_i}^*\right) \tag{۶۳}$$

از رابطه (۶۱) مشخص است که به دلیل وجود دینامیک خطای ردیابی زاویه $\dot{\varepsilon}_i$ در آن، نیاز به مشخص بودن دینامیک سیستم وجود دارد. از طرفی، فرض بر این است که دانش کاملی در مورد دینامیک سیستم در دسترس نیست، پس در ادامه از الگوریتم RL خارج از سیاست مستقل از مدل برای حل مسئله بهینه‌سازی (۵۹) استفاده می‌شود.

۴-۳- الگوریتم RL برای کنترل زاویه

در این بخش نیز مانند بخش ۳-۲ یک الگوریتم RL با رویکرد تکرار سیاست، برای یافتن یک کنترل‌کننده زاویه بهینه و بدون نیاز به دانشی از دینامیک سیستم، ارائه خواهد شد.

نسخه بسیار کوچک تابع هزینه (۵۷)، معادله بلمن زیر است،

$$0 = \bar{r}_i + \bar{J}_{\varepsilon_i}^T\dot{\varepsilon}_i \tag{۶۴}$$

که در آن $\bar{J}_{\varepsilon_i}^T\dot{\varepsilon}_i = \dot{\bar{J}}_i$ است. برای استفاده از رویکرد خارج از سیاست، دینامیک خطای (۵۶) باهدف تفکیک سیاست رفتار و سیاست هدف به شکل زیر بازنویسی می‌شود.

$$\begin{aligned}\dot{\varepsilon}_i = A_i(\Sigma_i)\Sigma_i - \dot{\Sigma}_{r_i} + B_{\Sigma_i}u_{\Sigma_i}^k \\ + B_{\Sigma_i}\left(u_{\Sigma_i} - u_{\Sigma_i}^k\right)\end{aligned} \tag{۶۵}$$

که در آن $u_{\Sigma_i}^k$ سیاست هدفی است که در الگوریتم به‌روز شده ولی به سیستم اعمال نمی‌شوند. سیاست پایدارساز u_{Σ_i} سیاست رفتار بوده و برای تولید داده یادگیری در الگوریتم RL به سیستم اعمال می‌شود. با استفاده از (۶۵) در (۶۴) داریم:

$$\bar{J}_{\varepsilon_i}^{kT}(A_i(\Sigma_i)\Sigma_i - \dot{\Sigma}_{r_i} + B_{\Sigma_i}u_{\Sigma_i}^k \tag{۶۶}$$

$$+B_{\Sigma_i}\left(u_{\Sigma_i}-u_{\Sigma_i}^k\right))=-\bar{r}_i\left(\varepsilon_i,u_{\Sigma_i}^k\right)+\bar{J}_{\varepsilon_i}^{kT}B_{\Sigma_i}\left(u_{\Sigma_i}-u_{\Sigma_i}^k\right)$$

با انتگرال‌گیری از هر دو طرف رابطه (۶۶) داریم:

$$\bar{J}_i^k\left(\varepsilon_i(t)\right)-\bar{J}_i^k\left(\varepsilon_i(t-T)\right)=-\int_{t-T}^{t}\bar{r}_i\left(\varepsilon_i,u_{\Sigma_i}^k\right)d\tau+\int_{t-T}^{t}\bar{J}_{\varepsilon_i}^{kT}B_{\Sigma_i}\left(u_{\Sigma_i}-u_{\Sigma_i}^k\right)d\tau \quad (۶۷)$$

ورودی کنترلی $u_{\Sigma_i}^k$ با بهینه‌سازی بر روی تابع HJ (۶۱) به شکل زیر به‌روز می‌شود.

$$u_{\Sigma_i}^{k+1}=-0.5\bar{R}_{ii}^{-1}B_{\Sigma_i}^{\mathrm{T}}\bar{J}_{\varepsilon_i}^k \quad (۶۸)$$

که از این عبارات داریم:

$$\bar{J}_{\varepsilon_i}^{kT}B_{\Sigma_i}=-2{u_{\Sigma_i}^{k+1}}^T\bar{R}_{ii} \quad (۶۹)$$

با جایگذاری (۶۹) در (۶۷) داریم:

$$\bar{J}_i^k\left(\varepsilon_i(t)\right)-\bar{J}_i^k\left(\varepsilon_i(t-T)\right)=-\int_{t-T}^{t}(\varepsilon_i^T\bar{Q}_i\varepsilon_i+u_{\Sigma_i}^{kT}\bar{R}_{ii}u_{\Sigma_i}^k)d\tau-\int_{t-T}^{t}2{u_{\Sigma_i}^{k+1}}^T\bar{R}_{ii}\left(u_{\Sigma_i}-u_{\Sigma_i}^k\right)d\tau \quad (۷۰)$$

در معادله بلمن خارج از سیاست (۷۰)، هم سیاست کنترل $u_{\Sigma_i}^{k+1}$و هم تابع مقدار $\bar{J}_i^k$، به‌طور هم‌زمان، به ازای سیاست رفتار u_{Σ_i} و با استفاده از داده‌های جمع‌آوری‌شده، به‌روز می‌شوند.

در این راه‌حل، الگوریتم RL خارج از سیاست با استفاده از یک ساختار بازیگر-منتقد، آموزش داده می‌شود که نیاز به دانشی از دینامیک سیستم ندارد. شبکه منتقد، تابع مقدار $\bar{J}_i^k$ و شبکه بازیگر، ورودی کنترل زاویه $u_{\Sigma_i}^{k+1}$ را به شکل زیر تخمین می‌زنند.

$$\hat{\bar{J}}_i^k(\varepsilon_i)={\widehat{W}_{3i}}^{kT}\bar{\Phi}(\varepsilon_i) \quad (۷۱)$$

$$\hat{u}_{\Sigma_i}^{k+1}(\varepsilon_i)={\widehat{W}_{4i}}^{kT}\bar{\theta}(\varepsilon_i) \quad (۷۲)$$

که در آن $\bar{\Phi}=[\bar{\Phi}_1\ \bar{\Phi}_2\ \dots\ \bar{\Phi}_{l_{\bar{\Phi}}}]\in\mathbb{R}^{l_{\bar{\Phi}}}$ توابع پایه شبکه منتقد و $\bar{\theta}=[\bar{\theta}_1\ \bar{\theta}_2\ \dots\ \bar{\theta}_{l_{\bar{\theta}}}]\in\mathbb{R}^{l_{\bar{\theta}}}$ توابع پایه شبکه بازیگر هستند. علاوه بر این ${\widehat{W}_{3i}}^k\in\mathbb{R}^{l_{\bar{\Phi}}}$ و ${\widehat{W}_{4i}}^k\in\mathbb{R}^{l_{\bar{\theta}}\times 3}$ بردارهای وزنی شبکه‌های عصبی مربوطه هستند. توجه داشته باشید که $u_{\Sigma_i}^{k+1}$ توسط شبکه عصبی (۷۲) تخمین زده‌شده و نیاز به هیچ‌دانشی در مورد دینامیک سیستم ندارند. خطای معادله بلمن (۷۰) با جایگذاری (۷۱) و (۷۲) در آن و در نظر گرفتن $\bar{v}_i^k=\left[\bar{v}_{1i}^k\ \bar{v}_{2i}^k\ \bar{v}_{3i}^k\right]^T=u_{\Sigma_i}-u_{\Sigma_i}^k$ به شکل زیر تعریف می‌شود.

$$e_{\Sigma_i}^k(t)={\widehat{W}_{3i}}^{kT}\bar{\Phi}(\varepsilon_i(t))-{\widehat{W}_{3i}}^{kT}\bar{\Phi}(\varepsilon_i(t-T))-\int_{t-T}^{t}-(\varepsilon_i^T\bar{Q}_i\varepsilon_i+u_{\Sigma_i}^{kT}\bar{R}_{ii}u_{\Sigma_i}^k)d\tau+\int_{t-T}^{t}2\left({\widehat{W}_{4i}}^{kT}\bar{\theta}(\varepsilon_i)\right)^T\bar{R}_{ii}\bar{v}_i^k d\tau \quad (۷۳)$$

با استفاده از ویژگی‌های ضرب کرونکر داریم:

$$\int_{t-T}^{t}-(\varepsilon_i^T\bar{Q}_i\varepsilon_i+u_{\Sigma_i}^{kT}\bar{R}_{ii}u_{\Sigma_i}^k)d\tau+e_{\Sigma_i}^k(t)={\widehat{W}_{3i}}^{kT}\left(\bar{\Phi}(\varepsilon_i(t))-\bar{\Phi}(\varepsilon_i(t-T))\right)+vec\left({\widehat{W}_{4i}}^k\right)^T\int_{t-T}^{t}2\left(\bar{R}_{ii}\bar{v}_i^k\otimes\bar{\theta}(\varepsilon_i)\right)d\tau \quad (۷۴)$$

از روش حداقل مربعات برای به دست آوردن کمینه خطای تقریب بلمن (۷۴) استفاده می‌شود. برای این منظور (۷۴) به فرم رگرسیون زیر بازنویسی می‌گردد.

$$y_{\Sigma_i}^k(t)+e_{\Sigma_i}^k(t)=\widehat{W}_{\Sigma_i}^{kT}h_{\Sigma_i}^k(t) \quad (۷۵)$$

که در آن $\widehat{W}_{\Sigma i}^k$، بردار بعدی و متشکل از $l_{\bar{\Phi}}+3l_{\bar{\theta}}$ بردارهای وزنی

$$\widehat{W}_{\Sigma i}^{kT}=\left[{\widehat{W}_{3i}}^{kT},vec\left({\widehat{W}_{4i}}^k\right)^T\right] \quad (۷۶)$$

بوده و $h_{\Sigma_i}^k(t)$ و $y_{\Sigma_i}^k(t)$ به فرم زیر در نظر گرفته شده‌اند.

$$h_{\Sigma_i}^k(t)=\begin{bmatrix}\bar{\Phi}(\varepsilon_i(t))-\bar{\Phi}(\varepsilon_i(t-T))\\2\int_{t-T}^{t}\left(\bar{R}_{ii}\bar{v}_i^k\otimes\bar{\theta}(\varepsilon_i)\right)d\tau\end{bmatrix} \quad (۷۷)$$

$$y_{\Sigma_i}^k(t)=-\int_{t-T}^{t}(\varepsilon_i^T\bar{Q}_i\varepsilon_i+u_{\Sigma_i}^{kT}\bar{R}_{ii}u_{\Sigma_i}^k)d\tau \quad (۷۸)$$

داده‌های موردنیاز را در O نقطه و در فاصله زمانی T جمع‌آوری می‌کنیم تا با حل (۷۵)، $\widehat{W}_{\Sigma i}^{kT}$ را به دست آوریم. اطلاعات جمع‌آوری‌شده را در ماتریس‌های $H_{\Sigma_i}^k$ و $Y_{\Sigma_i}^k$ ذخیره می‌نماییم.

$$H_{\Sigma_i}^k=\left[h_{\Sigma_i}^k(t_1),\dots,h_{\Sigma_i}^k(t_O)\right]$$
$$Y_{\Sigma_i}^k=\left[y_{\Sigma_i}^k(t_1),\dots,y_{\Sigma_i}^k(t_O)\right]$$

بنابراین فرم رگرسیون (۷۵) عبارت است از:

$$\widehat{W}_{\Sigma i}^{kT}\ H_{\Sigma_i}^k=Y_{\Sigma_i}^k \quad (۷۹)$$

که با استفاده از حداقل مربعات، پاسخ آن به فرم

$$\widehat{W}_{\Sigma i}^k=\left(H_{\Sigma_i}^kH_{\Sigma_i}^{kT}\right)^{-1}H_{\Sigma_i}^kY_{\Sigma_i}^{kT} \quad (۸۰)$$

به‌دست‌آمده و به ازای

$$O>l_{\bar{\Phi}}+3l_{\bar{\theta}} \quad (۸۱)$$

جواب دارد. مقادیر بردارهای وزنی ${\widehat{W}_{3i}}^k$ و ${\widehat{W}_{4i}}^k$ را می‌توان با حل (۸۰) و با استفاده از الگوریتم ۲ در هر تکرار بهبود بخشید. همگرایی الگوریتم ۲ نیز مانند ۱ قابل‌اثبات است.

الگوریتم ۲: الگوریتم MARL خارج از سیاست کنترل زاویه

۱: مقداردهی اولیه شبکه‌های بازیگر و منتقد، (۷۱) و (۷۲)

۲: **مرحله ۱:** جمع‌آوری داده

۳: از هر سیاست رفتار پایدارساز دارای نویز $u_{\Sigma_i} \in U_{\Sigma_i}$ تا زمانی که (۸۱) برای هر عامل برآورده شود، استفاده می‌نماییم.

$$\nabla \bar{V}_i = -\varepsilon_i^T \bar{Q}_i \varepsilon_i - u_{\Sigma_i}^T \bar{R}_{ii} u_{\Sigma_i}$$

۴: **پایان مرحله ۱**

۵: **مرحله ۲:** یادگیری سیاست هدف بهینه توزیع‌شده با استفاده از داده‌های جمع‌آوری‌شده

$$\nabla \bar{V}_i = -\varepsilon_i^T \bar{Q}_i \varepsilon_i - \frac{1}{4} \nabla \bar{V}_i^T B_{\Sigma_i} \bar{R}_{ii}^{-1} B_{\Sigma_i}^T \nabla \bar{V}_i < 0$$

۶: تولید ماتریس‌های $y_{\Sigma_i}^k(t)$ و $h_{\Sigma_i}^k(t)$ به‌صورت (۷۷)، (۷۸) برای همه $t = t_1, \dots, t_O$ و با استفاده از $u_{\Sigma_i}^k$

۷: به‌روزرسانی $u_{\Sigma_i}^k$ (۷۲) و $\bar{J}_i^k$ (۷۱) با استفاده از وزن‌های NN حاصل از (۸۰)

۸: اگر معیار توقف برآورده شد، الگوریتم متوقف شود، در غیر این صورت $k = k + 1$ و برو به ۵

۹: **پایان مرحله ۲**

۵-۳- تحلیل پایداری

در این بخش، پایداری سیستم شامل چندپرنده ناهمگن با استفاده از کنترل‌کننده آرایش گروهی بهینه و ایمن پیشنهادی ازنظر پایداری مورد تجزیه‌وتحلیل قرار می‌گیرد.

قضیه ۱: با استفاده از کنترل‌کننده آرایش گروهی بهینه توزیع‌شده پیشنهادی، شامل سیاست کنترل موقعیت (۲۹) و سیاست کنترل زاویه (۶۲)، خطاهای ردیابی $e_i = [\delta_i^T \ \varepsilon_i^T]^T \in \mathbb{R}^{12}$ به صفر همگرا شده و سیستم پایدار مجانبی است.

اثبات: با مشتق‌گیری از تابع مقدار افزوده (۲۲a) داریم:

که با توجه به رابطه HJC (۲۸) داریم:

با جایگذاری (۲۹) در رابطه فوق داریم:

با توجه به مثبت بودن مقدار CBF درون مجموعه ایمن و مثبت معین بودن ماتریس‌های Q_i، R_{ii} و R_{ij} همواره $\dot{V}_{a_i} < 0$ می‌باشد. بنابراین خطای ردیابی موقعیت δ_i برای دستیابی به آرایش گروهی به صفر همگرا می‌شود. این امر نشان‌دهنده پایدار مجانبی بودن سیستم خطای ردیابی موقعیت برای پرنده iام به‌منظور رسیدن به آرایش گروهی مطلوب است. به‌طور مشابه برای تابع عملکرد زاویه از (۶۱) نیز داریم:

که با جایگذاری (۶۲) در رابطه فوق داریم:

از رابطه فوق نیز مشاهده می‌شود که خطای ردیابی زاویه ε_i به صفر همگرا شده و بنابراین سیستم خطای ردیابی زاویه نیز پایدار مجانبی است.

۶-۳- تضمین قید عدم برخورد

در این بخش ابتدا در لم زیر نشان داده می‌شود که یک تابع مقدار برای هر پرنده وجود دارد. سپس، کران‌دار بودن CBF محلی بررسی می‌شود. در پایان، با استفاده از این لم‌ها، ایمنی یا قید عدم برخورد در قضیه ۲ تضمین می‌شود.

لم ۱: $u_{p_i,1} \in U_{ai}$ را سیاست کنترل بازخورد قابل‌قبول برای پرنده iام در نظر می‌گیریم. اگر برای همه سیاست‌های کنترل بازخورد قابل‌قبول همسایه‌های پرنده iام $u_{p_{N_i}} \in U_{aN_i}$، یک تابع مثبت معین نامتغیر بازمان $W_i \in C^1$ وجود داشته باشد به‌طوری‌که

$$\frac{\partial W_i^T}{\partial \delta_i}\left(\dot{\delta}_i\right) + \delta_i^T Q_i \delta_i + u_{p_{i,1}}^T R_{ii} u_{p_{i,1}} + \sum_{j \in N_i} u_{p_j}^T R_{ij} u_{p_j} + B_i = 0 \quad (۸۲)$$

$$\dot{V}_{a_i} = \nabla V_{ai} \dot{\delta}_i = J_{ai}\left(\delta_i(0), u_{p_{i,1}}, u_{p_{N_i}}, p_i(0), p_{N_i}(0)\right) W_i\left(\delta_i(0), u_{p_{i,1}}, u_{p_{N_i}}, p_i(0), p_{N_i}(0)\right) \quad (۸۳)$$

باشد. W_i یک تابع مقدار افزوده پرنده iام به ازای $t \in [0, \infty)$ و هر $u_{p_j} \in U_{ai}$ بوده و داریم:

$$\dot{V}_{a_i} = -\delta_i^T Q_i \delta_i - u_{p_i}^T R_{ii} u_{p_i} - \sum_{j \in N_i} u_{p_j}^T R_{ij} u_{p_j} - B_i \le W_i\left(\delta_i, u_{p_i}, u_{p_{N_i}}, p_i, p_{N_i}\right) \quad (۸۴)$$

$$\dot{V}_{a_i} = -\delta_i^T Q_i \delta_i - \frac{1}{4}(J_{ai} + \delta_i)^2 [\nabla V_{a_i}^T B_{p_i} R_{ii}^{-1} B_{p_i}^T \nabla V_{a_i} + \sum_{j \in N_i} \nabla V_{aj}^T B_{p_j} R_{jj}^{-1} R_{ij} R_{jj} B_{p_j}^T \nabla V_{a_j}] - B_i$$

اثبات: فرض کنید $W_i\left(\delta_i,u_{p_i,1},u_{p_{N_i}},p_i,p_{N_i}\right)>0$ای وجود دارد، که به دلیل پیوسته مشتق‌پذیر بودن آن، داریم:

$$\begin{aligned}&W_i\left(\delta_i(t),u_{p_i,1},u_{p_{N_i}},p_i(t),p_{N_i}(t)\right)\\&-W_i\left(\delta_i(0),u_{p_i,1},u_{p_{N_i}},p_i(0),p_{N_i}(0)\right)\\&=\int_0^t\frac{\partial W_i^T}{\partial\delta_i}\dot{\delta}_i d\tau\end{aligned}\quad(۸۵)$$

با در نظر گرفتن (۲۲) و (۲۳)، داریم:

$$\begin{aligned}&J_i\left(\delta_i(t),u_{p_i,1},u_{p_{N_i}},p_i(t),p_{N_i}(t)\right)\\&-J_i\left(\delta_i(0),u_{p_i,1},u_{p_{N_i}},p_i(0),p_{N_i}(0)\right)\\&=-\int_0^t r_{ai}\left(\delta_i(\tau),u_{p_i,1},u_{p_{N_i}},p_i(\tau),p_{N_i}(\tau)\right)d\tau\end{aligned}\quad(۸۶)$$

اگر دو طرف روابط (۸۵) و (۸۶) را از هم کم کنیم، به‌دست‌آمده می‌آوریم:

$$\begin{aligned}&J_i\left(\delta_i(t),u_{p_i,1},u_{p_{N_i}},p_i(t),p_{N_i}(t)\right)\\&-W_i\left(\delta_i(t),u_{p_i,1},u_{p_{N_i}},p_i(t),p_{N_i}(t)\right)\\&=-\int_0^t\left[\frac{\partial W_i^T}{\partial\delta_i}\dot{\delta}_i\right.\\&\left.+r_{ai}\left(\delta_i(\tau),u_{p_i,1},u_{p_{N_i}},p_i(\tau),p_{N_i}(\tau)\right)\right]d\tau\\&+J_i\left(\delta_i(0),u_{p_i,1},u_{p_{N_i}},p_i(0),p_{N_i}(0)\right)\\&-W_i\left(\delta_i(0),u_{p_i,1},u_{p_{N_i}},p_i(0),p_{N_i}(0)\right)\end{aligned}\quad(۸۷)$$

با استفاده از (۸۲) و (۸۳) در (۸۷)،

$$\begin{aligned}&J_i\left(\delta_i(t),u_{p_i,1},u_{p_{N_i}},p_i(t),p_{N_i}(t)\right)\\&-W_i\left(\delta_i(t),u_{p_i,1},u_{p_{N_i}},p_i(t),p_{N_i}(t)\right)=0\end{aligned}$$

بوده و درنتیجه

$$\begin{aligned}&J_i\left(\delta_i(t),u_{p_i,1},u_{p_{N_i}},p_i(t),p_{N_i}(t)\right)=\\&W_i\left(\delta_i(t),u_{p_i,1},u_{p_{N_i}},p_i(t),p_{N_i}(t)\right).\end{aligned}$$

لم ۲: با توجه به‌توالی سیاست‌های کنترل توزیع‌شده قابل‌قبول $u_{p_i,1}(\delta_i,t),u_{p_i,2}(\delta_i,t),\dots,u_{p_i,n}(\delta_i,t)\in U_{ai}$، توابع مقدار مثبت معین پرنده iام را به‌صورت $W_i\left(\delta_i,t,u_{p_i,1},u_{p_{N_i}},p_i,p_{N_i}\right),W_i\left(\delta_i,t,u_{p_i,2},u_{p_{N_i}},p_i,p_{N_i}\right),\dots,W_i\left(\delta_i,t,u_{p_i,n},u_{p_{N_i}},p_i,p_{N_i}\right)$ در نظر می‌گیریم، که مخفف آن‌ها $W_{i1},W_{i2},\dots,W_{in}$ می‌باشد. اگر مقادیر همیلتونی کمینه تعریف‌شده در (۲۷)، عبارت زیر را ارضا نماید

$$H_{ai1_{min}}\le H_{ai2_{min}}\le\cdots\le H_{ain_{min}}\quad(۸۸)$$

پس مقدار کاندید CBF محلی متناظر با این سیاست‌های کنترلی یعنی $B_{i,j}$ به ازای $1\le j\le n$ در هر مرحله از این توالی کران‌دار می‌باشد.

اثبات: برای هر l و kای که $0\le l\le k\le n$ باشد، از (۸۸) ملاحظه می‌شود که $H_{ail_{min}}\le H_{aik_{min}}$. حال فرض کنید

$$W_{ik}=W_{il}+W_{id}\quad(۸۹)$$

به‌طوری‌که

$$W_{id}\triangleq W_{id}\left(\delta_i,t,u_{p_i,l},u_{p_{N_i}},p_i,p_{N_i}\right)$$

با جایگذاری $u_{p_i,k}^*=-0.5(d_i+b_i)R_{ii}^{-1}B_{p_i}^T\nabla W_{ik}$ در (۲۸)، داریم:

$$\begin{aligned}H_{aik_{min}}=&L_i\left(\delta_i,u_{p_{N_i}},p_i,p_{N_i}\right)\\&-\frac{(d_i+b_i)^2}{4}\nabla W_{ik}^TB_{p_i}R_{ii}^{-1}B_{p_i}^T\nabla W_{ik}\\&+\nabla W_{ik}^T\Delta_i\left(u_{p_{N_i}},P_i,P_{N_i},P_0\right)\end{aligned}\quad(۹۰)$$

که در آن $L_i\left(\delta_i,u_{p_{N_i}},p_i,p_{N_i}\right)=\delta_i^TQ_i\delta_i+B_i+\sum_{j\in N_i}u_{p_j}^{*T}R_{ij}u_{p_j}^*$ بوده و

$\Delta_i\left(u_{p_{N_i}},P_i,P_{N_i},P_0\right)=(d_i+b_i)F(P_i)-\sum_{j\in N_i}a_{ij}\left(F(P_j)+B_{P_j}u_{p_j}\right)-b_i\dot{P}_0.$

با استفاده از (۸۹) در (۹۰)، رابطه زیر به‌دست‌آمده

$$\begin{aligned}&H_{aik_{min}}\\&=L_i\left(\delta_i,u_{p_{N_i}},p_i,p_{N_i}\right)+\nabla W_{il}{}^T\Delta_i\left(u_{p_{N_i}},P_i,P_{N_i},P_0\right)\\&-\frac{(d_i+b_i)^2}{4}\nabla W_{il}{}^TB_{p_i}R_{ii}^{-1}B_{p_i}^T\nabla W_{il}\\&+\nabla W_{id}{}^T\Delta_i\left(u_{p_{N_i}},P_i,P_{N_i},P_0\right)\\&-\frac{(d_i+b_i)^2}{2}\nabla W_{id}{}^TB_{p_i}R_{ii}^{-1}B_{p_i}^T\nabla W_{il}\\&-\frac{(d_i+b_i)^2}{4}\nabla W_{id}{}^TB_{p_i}R_{ii}^{-1}B_{p_i}^T\nabla W_{id}\end{aligned}$$

و معادل آن داریم:

$$\begin{aligned}H_{aik_{min}}=&H_{ail_{min}}-u_{p_i,d}^*{}^TR_{ii}u_{p_i,d}^*\\&+\nabla W_{id}{}^T\dot{\delta}_i\left(u_{p_i,l}^*,u_{p_{N_i}},P_i,P_{N_i},P_0\right)\end{aligned}\quad(۹۱)$$

از رابطه (۸۸) می‌دانیم که $H_{aik_{min}}-H_{ail_{min}}+u_{p_i,d}^*{}^TR_{ii}u_{p_i,d}^*\ge0$ بوده و بنابراین

$$\frac{dW_{id}\left(\delta_i,t,u_{p_i,l},u_{p_{N_i}},p_i,p_{N_i}\right)}{dt}\ge0\quad(۹۲)$$

همچنین $\lim_{t\to\infty}W_{id}\left(\delta_i,t,u_{p_i,l},u_{p_{N_i}},p_i,p_{N_i}\right)=0$، برای حفظ پایداری لزوماً برقرار می‌باشد، درنتیجه به دلیل

صعودی بودن آن همواره $W_{id} \le 0$ بوده و لذا از رابطه (۸۹) داریم

$$W_{ik} \le W_{il} \quad (۹۳)$$

نسبت به‌توالی (۸۶)، به این نتیجه می‌رسیم که

$$\begin{aligned} & W_i\left(\delta_i, t, u_{p_{i,1}}, u_{p_{N_i}}, p_i, p_{N_i}\right) \\ & > W_i\left(\delta_i, t, u_{p_{i,2}}, u_{p_{N_i}}, p_i, p_{N_i}\right) > \cdots \\ & > W_i\left(\delta_i, t, u_{p_{i,n}}, u_{p_{N_i}}, p_i, p_{N_i}\right) \end{aligned} \quad (۹۴)$$

پس $\forall 1 \le l \le n$

$$W_{il} < W_{i1} \quad (۹۵)$$

بوده و از لم ۱ می‌دانیم که

$$\begin{aligned} & J_i\left(\delta_i(t), u_{p_{i,l}}, u_{p_{N_i}}, p_i(t), p_{N_i}(t)\right) \\ & < J_i\left(\delta_i(t), u_{p_{i,1}}, u_{p_{N_i}}, p_i(t), p_{N_i}(t)\right) \end{aligned} \quad (۹۶)$$

چون $J_i\left(\delta_i(t), u_{p_{i,l}}, u_{p_{N_i}}, p_i(t), p_{N_i}(t)\right)$ کران‌دار بوده و r_{ai}همواره مثبت معین است، پس r_{ai} و در نتیجه آن $B_{i,l}$ کران‌دار می‌باشد. درنهایت ثابت گردید که هر CBF محلی به ازای هر توالی‌ای کران‌دار است.

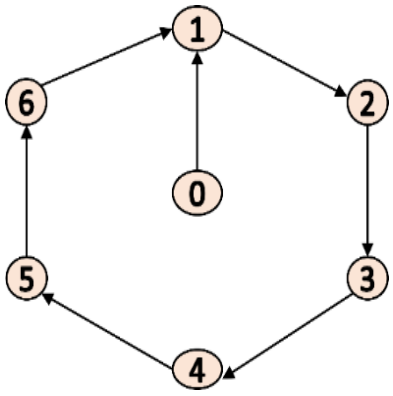
قضیه ۲: مسئله بهینه‌سازی توزیع‌شده (۲۲) را در نظر بگیرید و فرض کنید که فرضیات ۱ و ۲ برآورده شوند. بنابراین، سیستم چندعاملی برای همه عوامل و $t > 0$ ایمن می‌باشد، زیرا حالت هر عامل i از طریق بهبود متوالی ورودی کنترل (۲۹) تکامل‌یافته و در مجموعه ایمن باقی می‌ماند.

اثبات: همان‌طور که در لم ۲ نشان داده شد، پس از هر مرحله بهبود سیاست (۲۹)، CBF محلی B_i برای هر عامل i کران‌دار می‌باشد. از طرفی می‌دانیم که CBF محلی پیشنهادی تا زمانی که به مرز مجموعه ایمن نرسد به بی‌نهایت میل نمی‌کند. درنتیجه به دلیل کران‌دار ماندن CBF محلی پس از هر تکرار، حالت عامل iام هرگز از مرز مجموعه ایمن خود فراتر نمی‌رود. بدین ترتیب ایمنی سیستم چندعاملی و قید عدم برخورد عوامل تضمین می‌شود.

نکته ۴: در این مقاله قید ایمنی، قید عدم برخورد می‌باشد. قضیه ۲ ایمنی را به‌طورکلی ثابت نموده و بنابراین برخوردی در مسیر عامل‌ها رخ نمی‌دهد.

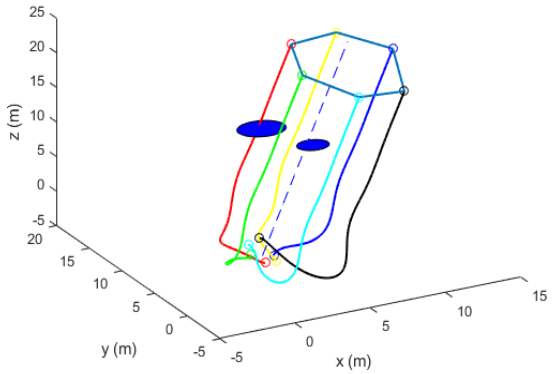
۴- نتایج شبیه‌سازی

در این بخش به بررسی نتایج شبیه‌سازی پرداخته و نتایج حاصل از [۱۱] را با رویکرد پیشنهادی مقایسه می‌نماییم. برای مقداردهی به پارامترهای رابطه (۷) و (۸) که در آن سیستم متشکل از ۶ پرنده ناهمگن می‌باشد، شرایط اولیه آن‌ها و مسیر رهبر مجازی از [۱۱] بهره جسته‌ایم. گراف ارتباطی این سیستم چند پرنده‌ای در شکل **۲** نشان داده‌شده است. این پرنده‌ها در حین دنبال کردن مسیر رهبر مجازی، مطلوب است که یک شش‌ضلعی منتظم با یکدیگر تشکیل دهند. با انتخاب ماتریس‌های وزنی مناسب در طراحی الگوریتم ۱ و ۲، توابع پایه شبکه‌های عصبی منتقد را به‌صورت چندجمله‌ای‌هایی با مرتبه چهار از خطای ردیابی موقعیت و موقعیت‌ها در (۴۱) و از خطای ردیابی زاویه در (۷۱) در نظر گرفته و توابع پایه شبکه‌های عصبی بازیگر در (۴۲) و (۷۲) به ترتیب به‌صورت $\theta(\delta_i) = \delta_i$ و $\bar{\theta}(\varepsilon_i) = \varepsilon_i$ می‌باشند.



شکل (۲): گراف ارتباطی.

ابتدا کنترل‌کننده بدون در نظر گرفتن موانع و با استفاده از رویکرد ارائه‌شده در [۱۱] طراحی‌شده است. شکل **۳** مسیر سه‌بعدی پرنده‌ها را با شش خط رنگی و رهبر مجازی را با خط‌چین نشان می‌دهد. با استفاده از کنترل‌کننده [۱۱]، پرنده‌ها رهبر مجازی را دنبال کرده و به آرایش گروهی موردنظر می‌رسند. این در حالی است که در شکل ۳ مشاهده می‌شود که یکی از پرنده‌ها با مانعی در طول مسیر خود برخورد کرده و بنابراین رویکرد موجود در [۱۱] ایمنی را تضمین نمی‌کند.

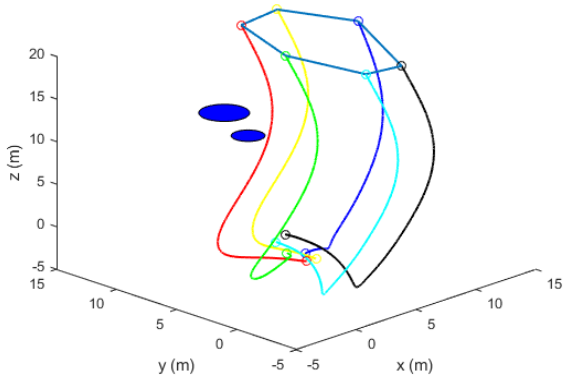


شکل (۳): مسیر سه‌بعدی بهینه و ناامن پرنده‌ها با استفاده از رویکرد [۱۱].

در گام بعد، کنترل‌کننده ایمن و بهینه را با استفاده از رویکرد پیشنهادی در این مقاله برای محیط شناخته‌شده متشکل از دو مانع دایره‌ای شکل زیر طراحی می‌کنیم،

$$C_i = \{(x_{o_i}, y_{o_i}, z_{o_i}) \mid z_{o_i} = c_{zi}, \\ (x_{o_i} - c_{xi})^2 + (y_{o_i} - c_{yi})^2 \le r_{o_i}^2\} \quad (۹۷)$$

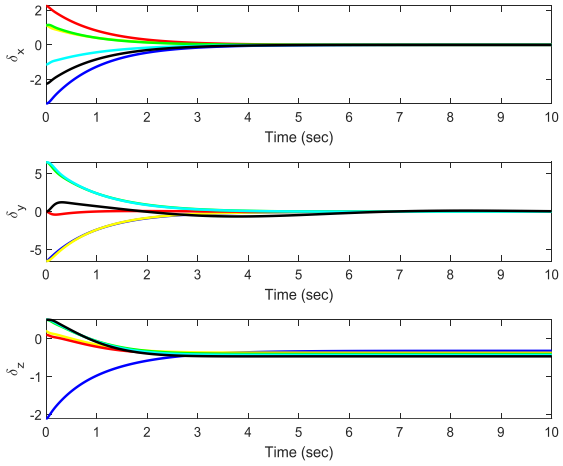
که در آن C_i به ازای $i = \{1,2\}$ مجموعه نقاطی را نشان می‌دهد که در داخل و روی مرز دایره‌های واقع در صفحه‌های $z_{o_i} = c_{zi}$، با مراکز (c_{xi}, c_{yi}, c_{zi}) و شعاع‌های r_{o_i} قرار دارند. بنابراین، موقعیت موانع با استفاده از موقعیت مرکز آن‌ها به صورت $O = \{o_1 = (4.5,10.5,11), o_2 = (9,13,5)\}$ و با شعاع $r_{o_1} = 1.5$ و $r_{o_2} = 1$ تعیین می‌شوند. شعاع ایمن پرنده‌ها نیز به صورت $r_{s_i} = 0.2$ تعریف می‌شود. با استفاده از کنترل‌کننده ایمن و بهینه پیشنهادی، مسیرهای سه‌بعدی ایمن پرنده‌ها در شکل **۴** نشان داده‌شده است. شکل **۴** صحت ادعای مطرح‌شده در این مقاله را نشان می‌دهد که در آن علاوه بر دستیابی به آرایش گروهی موردنظر، هیچ برخوردی رخ نداده است.



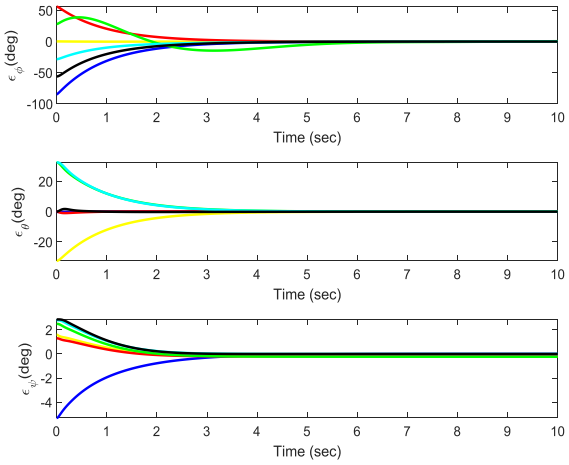
شکل (۴): مسیر سه‌بعدی بهینه و ایمن پرنده‌ها با استفاده از رویکرد پیشنهادی.

شکل **۳** و **۴** به ترتیب نتایج الگوریتم‌های مرجع [۱۱] و پیشنهادی را نشان می‌دهند. همان‌طور که در شکل ۴ نشان داده‌شده است، پرنده‌ها آرایش گروهی خود را در طول مسیر با بهینه‌سازی توابع هزینه افزوده حفظ می‌نمایند. شکل ۵ همچنین نشان می‌دهد که خطاهای ردیابی موقعیت به صفر یا نزدیک به صفر همگرا می‌شوند و بنابراین آرایش گروهی با خطای اندکی حفظ می‌شود. در مقایسه با شکل ۳ (جایی که وجود موانع مطرح نیست)، مسیرهای شکل ۴ بهینه نیستند. تعریف بهینگی بسته به اینکه محیط دارای مانع باشد یا نه، متفاوت است. درواقع، مسیرهای ایمن در شکل ۴ از مسئله بهینه‌سازی توزیع‌شده (۲۲) به‌دست‌آمده و در رابطه با توابع هزینه افزوده، بهینه هستند. این نتایج نشان‌دهنده بهینگی و ایمنی در رویکرد پیشنهادی است.

شکل **۵** خطاهای ردیابی موقعیت برای رسیدن به آرایش گروهی را برای سیستم چندپرنده‌ای با استفاده از کنترل‌کننده ایمن و بهینه آموزش‌دیده شده، نشان می‌دهد. همان‌طور که از شکل **۵** مشخص است، پرنده‌ها با نزدیک شدن δ_i به صفر به آرایش گروهی مطلوب میل می‌نمایند. شکل **۶** نیز نشان‌دهنده خطای ردیابی زاویه با استفاده از رویکرد پیشنهادی می‌باشد. همان‌طور که از این شکل مشخص است، ϵ_i در کمتر از ۵ ثانیه به صفر همگرا شده است. با همگرایی بردار خطای موقعیت و زاویه به صفر، نتیجه می‌گیریم که پرنده‌ها در یک راستای زاویه‌ای یکسان به آرایش گروهی مطلوب می‌رسند.

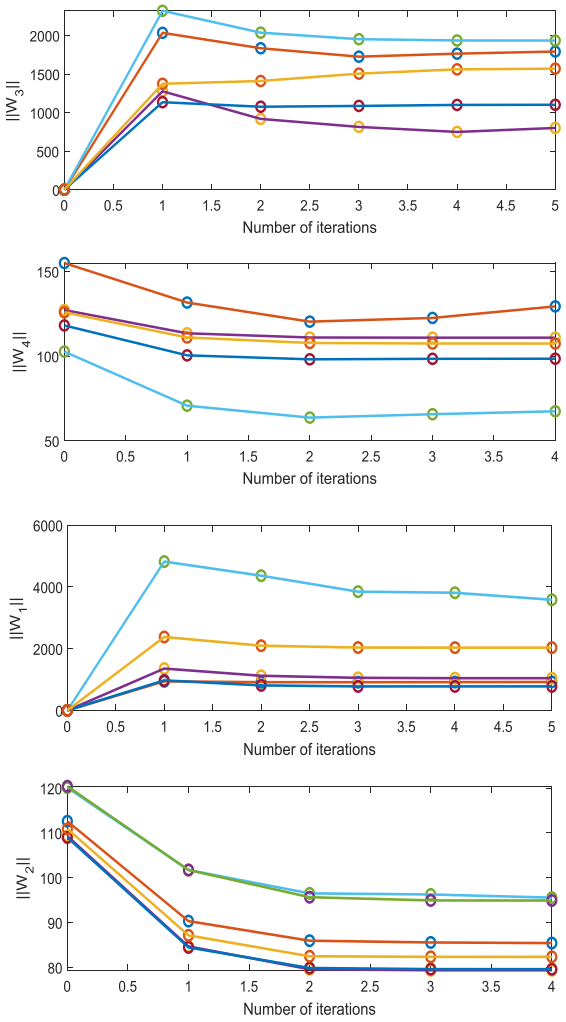


شکل (۵): خطای ردیابی آرایش گروهی با استفاده از رویکرد پیشنهادی.



شکل (۶): خطای ردیابی زاویه با استفاده از رویکرد پیشنهادی.

برای ارزیابی همگرایی وزن‌های NNهای W_i به ازای $(i = 1, \dots, 4)$، به‌منظور تخمین کنترل بهینه و تابع مقدار موجود در الگوریتم‌های ۱ و ۲ می‌توان به شکل **۷** مراجعه نمود. از این شکل مشخص است که این وزن‌ها در کمتر از ۴ تکرار به همگرایی می‌رسند. درنتیجه این همگرایی، سیاست‌های هدف پیشنهادی حاصل می‌شوند.

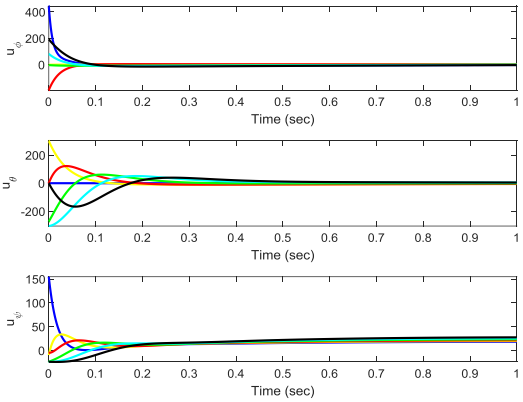


شکل (۷): همگرایی وزن‌های NNهای الگوریتم‌های ۱ و ۲.

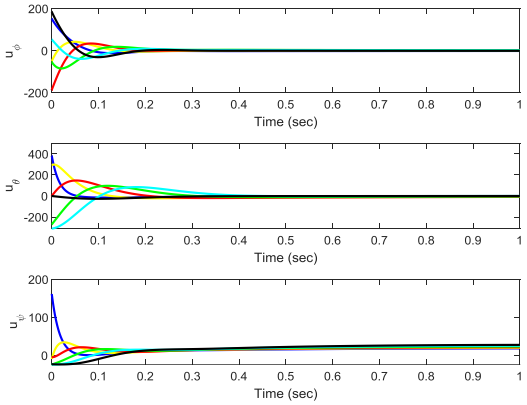
ورودی‌های کنترلی زاویه‌ای پرنده‌ها $(u_{\phi_i}, u_{\theta_i}, u_{\psi_i})$ با استفاده از رویکرد پیشنهادی و رویکرد مرجع [۱۱]، به ترتیب در شکل‌های **۸** و **۹** نشان داده‌شده است. همان‌طور که از مقایسه این ورودی‌های کنترلی ملاحظه می‌شود، هر دو در کمتر از ۱ ثانیه به مقداری ثابت همگرا شده و روند نسبتاً مشابهی از خود نشان می‌دهند. بنا به انتظاری که داشتیم و از شکل‌های **۸** و **۹** مشاهده می‌کنیم، هزینه کنترلی رویکرد ایمن پیشنهادی بیشتر از رویکرد بهینه [۱۱] است. این هزینه اضافی با توجه به تضمین ایمنی در این مقاله قابل‌اغماض می‌باشد.

از دیگر مواردی که می‌توان رویکرد پیشنهادی را با رویکرد [۱۱] مقایسه نمود، زمان لازم برای شبیه‌سازی است. در این مسئله خاص، شبیه‌سازی رویکرد [۱۱] پس از ۳۱ ثانیه به

جواب رسید و رویکرد پیشنهادی برای این منظور به ۳۴ ثانیه زمان نیاز داشت؛ بنابراین هزینه محاسباتی هر دو رویکرد تقریباً مشابه می‌باشد، با این تفاوت که رویکرد پیشنهادی از ویژگی ایمنی نیز برخوردار است.



شکل (۸): ورودی‌های کنترلی زاویه‌ای پرنده‌ها حاصل از رویکرد پیشنهادی.



شکل (۹): ورودی‌های کنترلی زاویه‌ای پرنده‌ها حاصل از رویکرد [۱۱].

۵- نتیجه‌گیری

در این مقاله یک سیاست کنترل آرایش گروهی بهینه و ایمن داده محور برای سیستم چند پرنده‌ای ناهمگن ارائه‌شده است، به‌طوری‌که با پیروی از رهبر مجازی به آرایش گروهی دلخواه دست یابد. با ادغام یک CBF محلی در تابع هزینه هر عامل، یک تابع مقدار افزوده توزیع‌شده برای طراحی کنترل‌کننده موقعیت مجازی ایجاد کرده و این راهبرد، ایمنی را برای جلوگیری از برخورد بین عوامل با یکدیگر و با موانع تضمین می‌نماید. از طریق کنترل‌کننده موقعیت مجازی ارائه‌شده زوایای مطلوب به‌دست‌آمده و کنترل‌کننده زاویه جهت رسیدن به این زوایای مطلوب طراحی می‌شود. درنهایت با اثبات پایداری و ایمنی، دستیابی به آرایش گروهی ایمن تضمین گردید. سیاست‌های کنترل موقعیت و زاویه‌ای پیشنهادی، ورودی کنترل آرایش گروهی بهینه و ایمن را با استفاده از الگوریتم RL خارج از سیاست و بدون نیاز به دانشی از دینامیک پرنده، ارائه می‌نماید. نتایج شبیه‌سازی برای سیستمی با ۶ پرنده ناهمگن، دستیابی به آرایش گروهی مطلوب و ایمن را با استفاده از رویکرد پیشنهادی، تأیید می‌کند. در ادامه این پژوهش محیط به‌صورت نیمه ناشناخته و دینامیک عوامل به‌صورت غیرخطی و ناشناخته در نظر گرفته خواهد شد.

۶- مراجع

[1] Barbastegan M, Bagheri A, Yazdani E, Chegini S. Optimal control of an aircraft pitch angle using pid and sliding mode control based on PSO algorithm. Journal of Aerospace Mechanics. 2020;15(4):49-66 (In Persian).

[2] Cao Y, Yu W, Ren W, Chen G. An overview of recent progress in the study of distributed multi-agent coordination. IEEE Transactions on Industrial informatics. 2012;9(1):427-38.

[3] Oh KK, Park MC, Ahn HS. A survey of multi-agent formation control. Automatica. 2015;53:424-40.

[4] Xu J, Wang L, Liu Y, Sun J, Pan Y. Finite-time adaptive optimal consensus control for multi-agent systems subject to time-varying output constraints. Applied Mathematics and Computation. 2022;427:127176.

[5] Zhou J, Zeng D, Lu X. Multi-agent trajectory-tracking flexible formation via generalized flocking and leader-average sliding mode control. IEEE Access. 2020;8:36089-99.

[6] Hua Y, Dong X, Li Q, Ren Z. Distributed adaptive formation tracking for heterogeneous multiagent systems with multiple nonidentical leaders and without well-informed follower. International Journal of Robust and Nonlinear Control. 2020;30(6):2131-51.

[7] Wang L, Xi J, He M, Liu G. Robust time-varying formation design for multiagent systems with disturbances: extended-state-observer method. International Journal of Robust and Nonlinear Control. 2020 May 10;30(7):2796-808.

[8] Amirani MZ, Bigdeli N, Haeri M. Time varying formation control of unmanned aerial vehicle multi-agent systems with unknown leader input. Journal of Aerospace Mechanics. 2021;17(2):53-69 (In Persian).

[9] Sayyaadi H, Mostafavi E. Formation control of unmanned helicopters by leader- follower method. Journal of Aerospace Mechanics. 2018;13(4): 59-69 (In Persian).

[10] E. Zhao, T. Chao, S. Wang, and M. Yang, "Finite-time Formation Control for Multiple Flight Vehicles with Accurate Linearization Model," Aerospace Science and Technology, vol. 71, pp. 90-98, 2017.

[11] Zhao E, Chao T, Wang S, Yang M. Finite-time formation control for multiple flight vehicles with accurate linearization model. Aerospace Science and Technology. 2017;71:90-8.

[12] Canese L, Cardarilli GC, Di Nunzio L, Fazzolari R, Giardino D, Re M, Spanò S. Multi-agent reinforcement learning: A review of challenges and applications. Applied Sciences. 2021;11(11):4948.

[13] Odekunle A, Gao W, Davari M, Jiang ZP. Reinforcement learning and non-zero-sum game output regulation for multi-player linear uncertain systems. Automatica. 2020;112:108672.

[14] Sutton RS, Barto AG. Reinforcement learning: An introduction. MIT press; 2018.

[15] Lewis FL, Vrabie D. Reinforcement learning and adaptive dynamic programming for feedback control. IEEE circuits and systems magazine. 2009;9(3):32-50.

[16] Wen G, Chen CP, Li B. Optimized formation control using simplified reinforcement learning for a class of multiagent systems with unknown dynamics. IEEE Transactions on Industrial Electronics. 2019;67(9):7879-88.

[17] Qu Q, Sun L, Li Z. Adaptive critic design-based robust cooperative tracking control for nonlinear multi-agent systems with disturbances. IEEE Access. 2021;9:34383-94.

[18] Bastani O. Safe reinforcement learning with nonlinear dynamics via model predictive shielding. In 2021 American Control Conference (ACC). 2021:3488-3494.

[19] Yazdani NM, Moghaddam RK, Kiumarsi B, Modares H. A Safety-Certified Policy Iteration Algorithm for Control of Constrained Nonlinear Systems. IEEE Control Systems Letters. 2020;4(3):686-91.

[20] Marvi Z, Kiumarsi B. Safe reinforcement learning: A control barrier function optimization approach. International Journal of Robust and Nonlinear Control. 2021;31(6):1923-40.

[21] Qin J, Li M, Shi Y, Ma Q, Zheng WX. Optimal synchronization control of multiagent systems with input saturation via off-policy reinforcement learning. IEEE transactions on neural networks and learning systems. 2018;30(1):85-96.

[22] B Yan B, Shi P, Lim CC, Shi Z. Optimal robust formation control for heterogeneous multi-agent systems based on reinforcement learning. International Journal of Robust and Nonlinear Control. 2022;32(5):2683-704.

[23] Labbadi M, Boudaraia K, Elakkary A, Djemai M, Cherkaoui M. A continuous nonlinear sliding mode control with fractional operators for quadrotor UAV systems in the presence of disturbances. Journal of Aerospace Engineering. 2022;35(1):04021122.

[24] Raffo GV, Ortega MG, Rubio FR. An integral predictive/nonlinear H∞ control structure for a quadrotor helicopter. Automatica. 2010;46(1):29-39.

[25] Lee H, Kim HJ. Constraint-based cooperative control of multiple aerial manipulators for handling an unknown payload. IEEE Transactions on Industrial Informatics. 2017;13(6):2780-90.

[26] Wang JL, Wu HN. Leader-following formation control of multi-agent systems under fixed and switching topologies. International Journal of Control. 2012;85(6):695-705.

[27] Olfati-Saber R, Murray RM. Consensus problems in networks of agents with switching topology and time-delays. IEEE Transactions on automatic control. 2004;49(9):1520-33.

Journal of Aerospace Mechanics/ 2023/ Vol.19/ No.1/ 61-79

Journal of Aerospace Mechanics

DOR: 20.1001.1.26455323.1402.19.1.5.0

Optimal Formation Control for Unmanned Aerial Vehicle Teams with Collision Avoidance Constraint and Unknown Dynamics

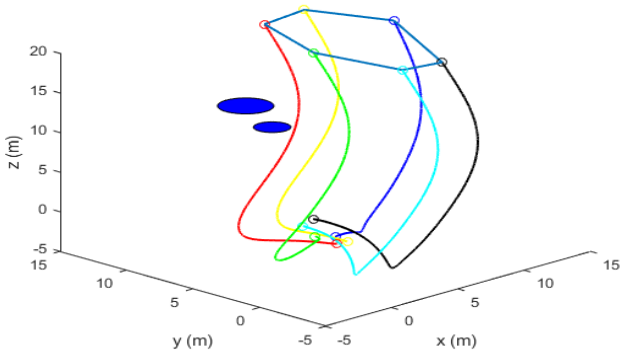
Fatemeh Mahdavi Golmisheh[1], Saeed Shamaghdari[2*]

[1] Ph.D. Student, Faculty of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran
[2] Associate Professor, Faculty of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran

HIGHLIGHTS

- Distributed formation control of a nonlinear and heterogeneous multi-UAV system.
- Integrating local CBF with MARL to guarantee collision-free constraints in a data-driven approach.
- Proposing two cascade off-policy RL algorithms for controlling position and attitude model-freely to achieve collision-free formation.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Article Type: Research paper
Received: 2 September 2022
Received in revised form: 3 October 2022
Accepted: 27 October 2022
Available online: 12 December 2022
*Correspondence:
shamaghdari@iust.ac.ir
How to cite this article:
M.M., S. Shamaghdari. Optimal formation control for unmanned aerial vehicle teams with collision avoidance constraint and unknown dynamics. Journal of Aerospace Mechanics. 2023; 19(1):61-79.

Keywords:
Unmanned Aerial Vehicles (UAV)
Multi-agent system
Formation control
Reinforcement Learning (RL)
Model-free RL

ABSTRACT

This paper presents distributed training approach for a nonlinear and heterogeneous multi-UAV system to solve a safe and optimal formation control problem. The objective of control is to ensure safety while achieving optimal performance. In this regard, the position and attitude controllers are considered in series. First, the optimal formation control design is defined as the optimal performance in position control and is modeled by the cost function. In this article, with the integration of cost functions and local control barrier functions (CBFs), a novel distributed optimization problems are introduced. Existing the local CBF in the augmented cost function ensures the safety of the position control, and as a result, collisions do not occur along the path of UAVs. The proposed method considers the safe and optimal position controllers by solving unconstrained optimization problems instead of constrained ones. In the next stage, the reference attitudes are driven by virtual position control. The attitude tracking optimal control is considered the optimal performance in the attitude control, and the related cost function models it. Finally, the stability and safety of the proposed controllers are proven. These optimal and safe policies are obtained sequentially using off-policy multi-agent reinforcement learning (MARL) algorithms which do not require knowledge of UAVs' dynamics. The proposed algorithms are validated by simulating the formation control problem of 6 UAVs with collision avoidance constraints.