مجله کنترل
ISSN 2008-8345
جلد ۸، شماره ۴، زمستان ۱۳۹۳، صفحه ۳۰-۱۵

کنترل بهینه توزیع شده بازی های گرافی دیفرانسیلی غیر خطی به صورت برخط با استفاده از یادگیری تقویتی

فرزانه تاتاری[۱]، محمد باقر نقیبی سیستانی[۲]

[۱] دانشجوی دکتری مهندسی برق، گروه مهندسی برق، دانشگاه فردوسی مشهد، fa.tatari@stu-mail.um.ac.ir

[۲] استادیار دانشکده مهندسی، گروه مهندسی برق، دانشگاه فردوسی مشهد، mb-naghibi@um.ac.ir

(تاریخ دریافت مقاله ۱۳۹۳/۶/۱۸، تاریخ پذیرش مقاله ۱۳۹۳/۱۰/۲۹)

چکیده: این مقاله به معرفی بازی های گرافی دیفرانسیلی برای سیستم های چند عاملی غیر خطی زمان پیوسته می پردازد و یک روش بهینه توزیع شده برخط برای حل آنها پیشنهاد می کند. در بازی های گرافی دیفرانسیلی، دینامیک خطا و اندیس عملکرد هر بازیکن تنها بستگی به اطلاعات همسایگان محلی آن عامل دارد. الگوریتم تکرار سیاست توزیع شده پیشنهاد شده، حل تقریبی معادلات همیلتون-جاکوبی کوپل شده همکارانه متعلق به عامل های غیر خطی را به صورت برخط انجام می دهد. در این الگوریتم که بر مبنای یادگیری تقویتی طراحی شده، هر یک از بازیکنان از ساختار شبکه عصبی نقاد-کنترلر استفاده می کند و تنظیم وزن های شبکه های عصبی نقاد و کنترلر به صورت همزمان انجام می شود. در حالی که تمام شبکه های عصبی نقاد-کنترلر در حال یادگیری هستند، پایداری حلقه بسته و همگرایی به قوانین کنترل بهینه تضمین می گردد. در انتها، نتایج به دست آمده از شبیه سازی، عملکرد و صحت الگوریتم پیشنهادی را نشان می دهد.

کلمات کلیدی: بازی های گرافی دیفرانسیلی غیرخطی؛ شبکه های عصبی؛ کنترل بهینه؛ یادگیری تقویتی.

Distributed Optimal Control of Nonlinear Differential Graphical Games based on Reinforcement Learning

Farzaneh Tatari, Mohammad-Bagher Naghibi-Sistani

Abstract: This paper introduces continuous time nonlinear differential graphical games and proposes an online distributed optimal control algorithm to solve them. In differential graphical games, each agent error dynamics and performance index depend on its neighbors' information. The proposed online distributed policy iteration algorithm solves the cooperative coupled Hamilton-Jacobi equations. In this algorithm which is based on reinforcement learning, each agent uses an actor-critic neural network structure where the weights of these neural networks are tuned synchronously. While all actor-critic networks are learning, closed loop stability and convergence to optimal control laws are guaranteed. Finally simulation results demonstrate the validity and performance of the proposed algorithm.

Keywords: Artificial neural networks; Nonlinear differential graphical games; Optimal control; Reinforcement learning.

۱- مقدمه

شبکه های توزیع شده به علت انعطاف پذیری و عملکرد محاسباتیشان در سالهای گذشته مورد مطالعه قرار گرفته اند. این شبکه ها برای مدل سازی و حل مسائلی که نیاز به بیش از یک عامل دارند، استفاده می شوند. یکی از مسایل اساسی در کنترل توزیع شده سیستم های چند عاملی، طراحی قانون کنترلی ساده برای هر عامل با استفاده از اطلاعات همسایگانش است به طوری که سیستم توزیع شده بتواند به یک رفتار جمعی (اجماع) دست پیدا کند. معمولا برای نمایش سیستم های چند عاملی و نحوه ارتباط عامل ها با یکدیگر در شبکه های توزیع شده از تئوری گراف استفاده می شود.

در مهندسی کنترل، مسئله همزمان سازی یا اجماع را می توان به دو دسته تنظیم همکارانه و ردیابی همکارانه طبقه بندی کرد. در مسئله تنظیم همکارانه (بدون رهبر)، کنترل عامل ها به گونه ای طراحی می شود که همه عامل ها به یک مقدار مشترک که وابسته به شرایط اولیه آنهاست همگرا می شوند. در مسئله ردیابی همکارانه (اجماع پیرو-رهبر)، با طراحی کنترلرهای مناسب برای عامل ها، متغیر حالت همه عامل ها به حالت عامل کنترلی یا رهبر [۱،۲،۳] همگرا می شوند.

برای مدل سازی کنترل بهینه توزیع شده سیستم های چندعاملی، از تئوری بازی های دیفرانسیلی استفاده می کنیم. از آنجایی که در این بازی چند نفره دینامیک خطا و اندیس عملکرد هر عامل بستگی به عامل های همسایه اش دارد و یا به عبارتی بستگی به ساختار گراف ارتباطی شبکه دارد، این نوع از بازی ها را بازی های گرافی دیفرانسیلی[1] می نامند.

نظریه بازی های دیفرانسیلی شامل ترکیبی از مشخصه های تئوری بازی[۴] و تئوری کنترل بهینه می باشد [۵] . در بازی های گرافی دیفرانسیلی هدف یافتن مجموعه ای از سیاست های کنترلی قابل قبول است که منجر به همزمان سازی عامل ها، تضمین پایداری سیستم حلقه بسته و حداقل شدن تابع هزینه هر عامل در جهت رسیدن به تعادل نش می باشد. پیدا کردن حل های نش بازی، وابسته به حل معادلات پیوسته همیلتون-جاکوبی کوپل شده[2](CHJ) می باشد که در موارد خطی به معادلات پیوسته ریکاتی جبری کوپل شده[3] (CARE) تقلیل می یابد [۶،۷،۸] . حل این معادلات دیفرانسیل جزئی غیرخطی بسیار مشکل بوده و یا حتی در مواردی فاقد حل تحلیلی همه جایی می باشند.

مزایای استفاده از سیاست های کنترلی بهینه توزیع شده برخط به جای سیاست های کنترل بهینه توزیع شده برون خط این است که بستری برای ترکیب کنترل بهینه با کنترل تطبیقی فراهم می نماید. به این ترتیب، می توان به صورت زمان حقیقی به حل مسائل کنترل بهینه توزیع شده سیستم های چند عاملی دارای دینامیک معین، دارای عدم قطعیت پارامتری و یا حتی دارای دینامیک نامشخص پرداخت.

به منظور یافتن سیاست های کنترلی فیدبک بهینه به صورت برخط در مسائل کنترل بهینه همکارانه توزیع شده، می توان از روشهای یادگیری تقویتی[4] استفاده کرد. در یادگیری تقویتی یک یا چندین عامل به صورت زمان-حقیقی با محیطی که ممکن است برای عامل ها شناخته شده نباشد، تعامل نموده و بر اساس تجربیاتی که کسب میکنند، یاد میگیرند تا استراتژی های بهینه را برای رسیدن به یک هدف خاص و بیشینه کردن مجموع پاداش خود انتخاب نمایند.

تکنیک های به کار رفته برای حل مسئله اجماع در کنترل بهینه همکارانه سیستم های چندعاملی، برون خط می باشند [۹،۱۰،۱۱،۱۲،۱۳] و در اکثر این روش ها سعی می شود به نحوی از حل معادلات پیچیده CHJ اجتناب شود. علی رغم اینکه اکثر روش های کنترل بهینه برون خط می باشد، در این مقاله با استفاده از راهکارهای یادگیری تقویتی به طراحی کنترلر بهینه همکارانه برخط می پردازیم. کنترلر پیشنهاد شده به صورت زمان حقیقی با استفاده از داده های اندازه گیری شده در طی مسیر سیستم، به همزمان سازی عامل ها به رهبر می پردازد.

ساختار نقاد-کنترلر از انواع الگوریتم های یادگیری تقویتی پیشرو در زمان[5] می باشد و به صورت زمان حقیقی پیاده سازی می شود. مکانیزم یادگیری در این ساختار دارای دو گام ارزیابی سیاست و بهبود سیاست می باشد. در گام ارزیابی سیاست، کنترلر یک سیاست کنترلی به محیط اعمال می کند و نقاد ارزش این سیاست را برآورد می کند و در گام دوم با توجه به ارزش به دست آمده سیاست کنترلی بهبود می یابد. از الگوریتم های متعلق به کلاس نقاد-کنترلر می توان به الگوریتم های تکرار سیاست[6] و تکرار ارزش[7] اشاره کرد که اساس آنها حل معادله بلمن می باشد [۱۴] . با استفاده از برنامه ریزی پویای تقریبی[8] و تقریب تابع ارزش توسط یک تقریبگر مناسب، این روش ها را می توان به صورت برخط پیاده سازی نمود. برای یادگیری برخط ارزش بهینه و تخمین پارامترهای نامشخص تقریبگر، می توان از یادگیری تقویتی تفاوت زمانی[9] استفاده نمود. روش یادگیری تقویتی تفاوت زمانی در مقایسه با دیگر روش های موجود همچون محاسبه دقیق[10] و روش های مونت کارلو[11]، قابلیت پیاده سازی به صورت زمان حقیقی در طی مسیر سیستم را دارد و از جمله روش های مربوط به کنترل تطبیقی می باشد [۱۴] .

در رهیافت تکرار سیاست در هر تکرار، معادله بلمن متناظر با هر سیاست باید به صورت کامل حل شود، بنابراین این الگوریتم در مقایسه با الگوریتم تکرار ارزش که مبتنی بر معادلات بازگشتی است دارای محاسبات بیشتری می باشد. ولی با توجه به سرعت همگرایی بالای تکرار سیاست نسبت به تکرار ارزش و امکان انجام اثبات های همگرایی و

[1] Differential graphical games
[2] Coupled Hamilton-Jacobi
[3] Coupled Algebraic Riccati Equations

[4] Reinforcement Learning
[5] Forward in time
[6] Policy iteration
[7] Value iteration
[8] Approximate dynamic programming
[9] Temporal difference
[10] Exact computation
[11] Monte Carlo

پایداری برای روش های مبتنی بر این الگوریتم، در این مقاله از الگوریتم تکرار سیاست برای پیاده سازی برخط کنترل بهینه توزیع شده استفاده می شود.

به این ترتیب در این مقاله برای هر عامل در بازی گرافی دیفرانسیلی، یک ساختار مجزای نقاد-کنترلر مبتنی بر روش تکرار سیاست در نظر گرفته می شود.

از روش تکرار سیاست و ساختارهای نقاد-کنترلر برای حل برخط بازی های دیفرانسیلی مجموع غیر صفر نیز استفاده شده است. نویسندگان مرجع[۱۵] ، الگوریتمی بر اساس روش تکرار سیاست و در قالب نقاد-کنترلر برای حل معادلات CHJ بازی های دیفرانسیلی مجموع غیر صفر به صورت زمان-حقیقی برای سیستم های زمان پیوسته غیرخطی و دینامیک کاملاً معین ارائه کرده اند. همچنین حل بازی دونفره مجموع غیرصفر در مرجع [۱۶] مورد بررسی قرار گرفته است. نویسندگان در این مرجع از ساختار تک-شبکه برای هر بازیکن به جای ساختار شبکه عصبی دوتایی نقاد-کنترلر استفاده کرده اند، که در نتیجه منجر به کاهش پیچیدگی محاسبات نسبت به روش ارائه شده در مرجع[۱۵] شده است. در[۱۷] برای حل بازی دیفرانسیلی مجموع غیر صفر دو بازیکنه با دینامیک خطی، روش برنامه ریزی پویای تقریبی (ADP) برخط با استفاده از رهیافت یادگیری تقویتی انتگرالی (IRL) در غالب یک الگوریتم تکرار سیاست پیاده سازی شده است که در آن قسمتی از دینامیک سیستم مورد نیاز نمی باشد.

کنترل بهینه همکارانه برخط بازی های گرافی دیفرانسیلی چندعاملی برای سیستم های خطی زمان پیوسته با دینامیک کاملا معین در [۱۸،۱۹،۲۰] بررسی شده که با استفاده از روش تکرار سیاست و ساختارهای نقاد-کنترلر به حل تقریبی معادلات CHJ می پردازد. در [۲۰] اثبات پایداری و همگرایی به نقطه تعادل نش با استفاده از تکنیک لیاپانف انجام شده است. حل برخط بازی های گرافی دیفرانسیلی سیستم های خطی زمان پیوسته با دینامیک نیمه معین در [۲۱] انجام شده است که با استفاده از ایده [۲۲] به حل برخط این بازی ها با دینامیک نیمه معین (دینامیک داخلی نامعین و دینامیک ورودی معین) می پردازد.

با توجه به مطالعات انجام شده تا کنون هیچ گزارشی برای حل برخط بازی های گرافی دیفرانسیلی برای سیستم های غیرخطی ارایه نشده است. از اینرو هدف و نوآوری این مقاله ارایه الگوریتم برخط کنترل بهینه توزیع شده بازی های گرافی دیفرانسیلی چند نفره افق نامحدود بر اساس تکنیک تکرار سیاست برای سیستم های چندعاملی غیرخطی (خطی تبار[1]) با دینامیک معین است. از جمله نقاط قوت الگوریتم پیشنهاد شده این است که الگوریتم از تکنیک های برخط به جای روش های برون خط و از اطلاعات محلی برای همزمان سازی عامل ها به رهبر، استفاده می کند. در الگوریتم ارایه شده علاوه بر حل تقریبی معادلات CHJ، پایداری سیستم حلقه بسته و همگرایی به نقطه تعادل نش بازی ضمن اجرای الگوریتم به صورت برخط ارایه می گردد. جهت یافتن ارزش بهینه، آموزش و تنظیم وزن های شبکه های عصبی نقاد، با استفاده از قوانین تنظیم استاندارد گرادیان نزولی[2] انجام می شود که منجر به حداقل کردن خطای تفاوت زمانی و همگرایی به تعادل نش می شود. از طرفی برای آموزش و تنظیم وزن های شبکه عصبی کنترلگر از قوانین تنظیم غیراستاندارد استفاده می کنیم؛ به طوری که با انتخاب پارامترهای طراحی مناسب در این قوانین، اثبات پایداری حلقه بسته با استفاده از تکنیک لیاپانوف انجام شود.

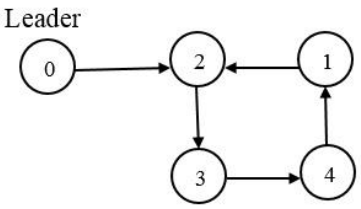
۲- مقدمه ای بر تئوری گراف

توپولوژی تعاملی تبادل اطلاعات بین N عامل، توسط گراف $Gr(V,\Sigma)$ توصیف می شود که در آن $V=\{1,2,...,N\}$ مجموعه گره های گراف است که نماینده N عامل می باشد، $\Sigma\subseteq V\times V$ مجموعه شاخه های گراف و $(i,j)\in\Sigma$ به معنی وجود یک شاخه از گره i به گره j می باشد. در این مقاله گراف ساده فرض می شود. توپولوژی یک گراف معمولا توسط ماتریس همسایگی آن $E=\left[e_{ij}\right]\in\Re^{N\times N}$ نمایش داده می شود به طوری که اگر $(j,i)\in\Sigma$ آنگاه $e_{ij}=1$ و در غیر این صورت $e_{ij}=0$ می باشد. $N_i=\{j:(j,i)\in\Sigma\}$ مجموعه همسایگان گره i است، به عبارت دیگر مجموعه گره ها با شاخه هایی است که به گره i وارد می شوند. $i_N=\{j:(i,j)\in\Sigma\}$ نیز نشان دهنده مجموعه ای از عامل ها هستند که عامل i در همسایگی آنها می باشد. اگر گره j همسایه گره i باشد، گره i می تواند از گره j اطلاعات دریافت کند ولی عکس آن الزاما در گرافهای جهتدار برقرار نمی باشد. اما در گرافهای بدون جهت، همسایگی یک رابطه متقابل می باشد. ماتریس درجه-واردشونده[3]، یک ماتریس قطری است $D=diag(d_i)\in\Re^{N\times N}$، با $d_i=\sum_{j\in N_i}e_{ij}$ که درجه-واردشونده گره i می باشد (یعنی مجموع عناصر سطر iام E). ماتریس لاپلاسین گراف به صورت $L=D-E$ نمایش داده می شود و مجموع عناصر هر سطر آن صفر می باشد. مسیر، دنباله ای از گره های به هم متصل در یک گراف است و یک گراف را متصل گویند اگر مسیری بین هر دو گره دلخواه آن وجود داشته باشد. معمولا گره رهبر توسط اندیس صفر نشان داده می شود و اطلاعات از رهبر به عامل هایی که رهبر در همسایگی آنهاست، فرستاده می شوند. شکل ۱ نمونه ای از گراف ارتباطی یک سیستم چندعاملی را نشان میدهد.

[1] Affine

[2] Gradient descent

[3] In-degree



شکل ۱- نمایشی از یک گراف ارتباطی

۳- بازی های گرافی دیفرانسیلی برای سیستم های چندعاملی غیرخطی خطی تبار

تعداد N سیستم یا عامل توزیع شده به همراه عامل رهبر را بر روی گراف ارتباطی Gr با دینامیک زیر در نظر بگیرید.

$$\dot{x}_i = f_i(x_i) + g_i(x_i)\, u_i \ , \quad i = 1,..., N. \qquad (۱)$$

$$\dot{x}_0 = f_0(x_0) \qquad (۲)$$

که $x_i \in \Re^n$ حالت عامل iام، $x_0 \in \Re^n$ حالت رهبر، $f_i(x_i) \in \Re^n$ دینامیک داخلی سیستم، $g_i(x_i) \in \Re^{n\times m}$ دینامیک ورودی سیستم و $u_i \in \Re^m$ ورودی کنترلی عامل iام می باشد. توجه شود که حالات عامل ها در دسترس هستند. توابع $f_i(x_i)$ و $g_i(x_i)$ $i = 1,..., N$ لیپشیتز محلی هستند و بر روی یک مجموعه فشرده تعریف شده اند. بعلاوه سیستم (۱) پایدارپذیر است و $f_0(0) = 0$. هدف مسئله همزمان سازی، طراحی پروتکل های کنترل محلی تنها با استفاده از اطلاعات عامل های همسایه است به طوری که حالت همه عامل ها به حالت رهبر همزمان سازی شود. یعنی $\lim_{t\to\infty} \|x_i(t) - x_0(t)\| = 0, \ \forall i$.

برای بیان اهداف همکارانه تیمی معمولا از خطای ردیابی همسایگان محلی $\delta_i \in \Re^n$ استفاده می شود که به صورت زیر محاسبه می شود[۲۳] :

$$\delta_i = \sum_{j\in N_i} e_{ij}(x_i - x_j) + e_{i0}(x_i - x_0) \qquad (۳)$$

e_{i0} را بهره اتصال[1] گویند که برای تعداد محدودی از عامل ها (گره ها) غیر صفر است. گره هایی که به صورت مستقیم با گره رهبر، x_0، در ارتباط اند دارای $e_{i0} = 1$ و در غیر این صورت دارای $e_{i0} = 0$ هستند. بردار خطای ردیابی کلی برای همه عامل ها به صورت زیر داده می شود

$$\delta = ((L + E_0)\otimes I_n)(x - \underline{x}_0) = ((L + E_0)\otimes I_n)\zeta \qquad (۴)$$

که در آن بردار حالت همه جایی $x = [x_1^T, x_2^T, ..., x_N^T]^T \in \Re^{nN}$ می باشد، بردار خطای ردیابی همه جایی $\delta = [\delta_1^T, \delta_2^T, ..., \delta_N^T]^T \in \Re^{nN}$ می باشد و $\underline{x}_0 = \underline{I}\, x_0 \in R^{nN}$ با $\underline{I} = \underline{1}\otimes I_n \in \Re^{nN\times n}$ و $\underline{1}$ بردار N تایی از ۱ است. $\otimes$ ضرب کرونکر می باشد [۲۴] . $E_0 \in \Re^{N\times N}$ ماتریس قطری با عناصر قطری برابر با بهره های اتصال e_{i0} می باشد. بعلاوه بردار عدم توافق همه جایی یا بردار خطای همزمان سازی به صورت $\zeta = (x - \underline{x}_0) \in \Re^{nN}$ تعریف می شود.

فرض می شود که گراف ارتباطی متصل است و برای حداقل یک i، $e_{i0} = 1$ است، آنگاه $(L + E_0)$ ناویژه خواهد بود و همه مقادیر ویژه آن دارای قسمت حقیقی مثبت هستند[۲۳] . اثبات می شود که اگر گراف ارتباطی متصل باشد و $E_0 \neq 0$ ، آنگاه خطای همزمان سازی به صورت زیر کراندار است[۲۱] .

$$\|\zeta\| \le \|\delta\| / \underline{\sigma}(L + E_0) \qquad (۵)$$

به این ترتیب هدف حداقل کردن خطای ردیابی همسایگان محلی $\delta_i(t)$ است، که با توجه به رابطه (۵)، همزمان سازی را تضمین می نماید. دینامیک خطای ردیابی همسایگان محلی به صورت زیر به دست می آید

$$\dot{\delta}_i = \sum_{j\in N_i} e_{ij}(\dot{x}_i - \dot{x}_j) + e_{i0}(\dot{x}_i - \dot{x}_0) \qquad (۶)$$

دینامیک خطای ردیابی همسایگان محلی با استفاده از روابط (۶)، (۱) و (۲) به صورت زیر به دست می آید.

$$\dot{\delta}_i = \sum_{j\in N_i} e_{ij}(f(x_i) - f(x_j)) + e_{i0}(f(x_i) - f(x_0)) + (d_i + e_{i0})g_i(x_i)u_i - \sum_{j\in N_i} e_{ij} g_j(x_j)u_j \qquad (۷)$$

دینامیک به دست آمده دارای ورودی های کنترلی از گره i و همسایگانش می باشد. اعمال کنترلی همسایگان عامل i و اعمال همه عامل های موجود در گراف منهای عامل i به ترتیب توسط $u_{-i} = \{u_j \mid j \in N_i\}$ و $u_{\bar{i}} = \{u_j \mid j \in N, j \neq i\}$ تعریف می شوند.

به منظور تعریف بازی گرافی دیفرانسیلی، اندیس های عملکرد محلی را برای هر عامل به صورت زیر تعریف می کنیم.

$$J_i(\delta_i(0), u_i, u_{-i}) = \frac{1}{2}\int_0^\infty \left(Q_i(\delta_i) + u_i^T R_{ii} u_i + \sum_{j\in N_i} u_j^T R_{ij} u_j \right) dt \qquad (۸)$$

که در آن $Q_i(\delta_i) > 0$ معمولا تابعی غیر خطی می باشد و ماتریس های وزن $R_{ii} > 0, R_{ij} > 0$ ، ثابت و متقارن هستند. لازم به توجه است که دینامیک (۷) و اندیس های عملکرد (۸) صریحا وابسته به توپولوژی گراف $Gr = (V, \Sigma)$ می باشند.

تعریف ۱- سیاست های کنترلی $\forall i, u_i$ بر روی مجموعه $\Omega \in \Re^n$ قابل قبول گفته می شوند اگر u_i بر روی Ω پیوسته باشد و $u_i(0) = 0$. همچنین u_i باید سیستم (۷) را به صورت محلی بر

[1] Pinning gain

روی مجموعه Ω پایدار سازد و (۸) را $\forall \delta_i(0) \in \Omega$ مقداری محدود کند [۲۰].

تعریف۲- استراتژی های $\{u_1^*, u_2^*, ..., u_N^*\}$، $u_i^* \in \Omega_i, i \in N$، حل تعادل نش[1] برای بازی با N بازیکن هستند، اگر نامساوی های $J_i^* = J_i(u_i^*, u_{\bar{i}}^*) \le J_i(u_i, u_{\bar{i}}^*)$ به ازای همه $u_i^* \in \Omega, i \in N$ برآورده شوند [۶].

تابع ارزش متناظر با گره i به صورت زیر می باشد

$$V_i(\delta_i(t)) = \frac{1}{2}\int_t^{\infty}\left(Q_i(\delta_i) + u_i^T R_{ii} u_i + \sum_{j\in N_i} u_j^T R_{ij} u_j\right) dt \quad (۹)$$

و هدف عامل i در بازی گرافی مورد نظر تعیین تابع ارزش زیر می باشد که منجر به همزمان سازی عامل ها به عامل رهبر و بهینه سازی ورودی های کنترلی قابل قبول می شود.

$$V_i^*(\delta_i(t)) = \min_{u_i}\int_t^{\infty}\frac{1}{2}\left(Q_i(\delta_i) + u_i^T R_{ii} u_i + \sum_{j\in N_i} u_j^T R_{ij} u_j\right) dt \quad (۱۰)$$

حداقل سازی (۱۰) با توجه به (۷) را یک بازی گرافی دیفرانسیلی برای سیستم های غیرخطی زمان پیوسته خطی تبار می نامند، که بستگی به توپولوژی گراف ارتباطی $Gr(V, \Sigma)$ دارد.

وقتی V_i محدود می شود، با مشتق گرفتن از (۹) با توجه به سیستم (۷)، معادله لیاپانفی به شکل معادله بلمن زیر با شرط اولیه $V_i(0) = 0$ به دست می آید.

$$\begin{aligned}&\nabla V_i^T[\sum_{j\in N_i} e_{ij}(f(x_i) - f(x_j)) + e_{i0}(f(x_i) - f(x_0))\\ &+(d_i + e_{i0})g_i(x_i)u_i - \sum_{j\in N_i} e_{ij} g_j(x_j)u_j]\\ &+\frac{1}{2}Q_i(\delta_i) + \frac{1}{2}u_i^T R_{ii} u_i + \frac{1}{2}\sum_{j\in N_i} u_j^T R_{ij} u_j = 0\end{aligned} \quad (۱۱)$$

حل رابطه (۱۱) به عنوان روشی دیگر برای ارزیابی انتگرال نامحدود (۹) برای پیدا کردن ارزش مربوط به سیاست های فیدبکی، موجود می باشد. برای دینامیک (۷) و اندیس عملکرد (۸)، همیلتونین مربوطه به صورت زیر به دست می آید.

$$\begin{aligned}H_i(\delta_i, \nabla V_i, u_i, u_{-i}) \equiv &\nabla V_i^T[\sum_{j\in N_i} e_{ij}(f(x_i)\\ &- f(x_j)) + e_{i0}(f(x_i) - f(x_0))\\ &+(d_i + e_{i0})g_i(x_i)u_i - \sum_{j\in N_i} e_{ij} g_j(x_j)u_j]\\ &+\frac{1}{2}Q_i(\delta_i) + \frac{1}{2}u_i^T R_{ii} u_i + \frac{1}{2}\sum_{j\in N_i} u_j^T R_{ij} u_j\end{aligned} \quad (۱۲)$$

که در آن $\nabla V_i = \partial V_i / \partial \delta_i \in \Re^n$، مشتق جزئی تابع ارزش نسبت به δ_i می باشد. با فرض مشتق پذیر بودن تابع معیار پیوسته و استفاده از اصل بهینگی بلمن، شرط بهینگی و کنترل فیدبک متناظر، u_i^*، به صورت زیر به دست می آید.

$$\begin{aligned}&0 = \min_{u_i}[H_i(\delta_i, \nabla V_i^*, u_i, u_{-i})] \rightarrow\\ &u_i^* = u_i^*(V_i^*) = -(d_i + e_{i0})R_{ii}^{-1} g_i^T(x_i)\nabla V_i^*\end{aligned} \quad (۱۳)$$

با جاگذاری سیاست های کنترلی (۱۳) در (۱۱)، معادلات CHJ پیوسته برای سیستم های غیرخطی به صورت زیر به دست می آید

$$\begin{aligned}&\nabla V_i^{*T}[\sum_{j\in N_i} e_{ij}(f(x_i) - f(x_j)) + e_{i0}(f(x_i)\\ &- f(x_0)) - (d_i + e_{i0})^2 g_i(x_i)R_{ii}^{-1} g_i^T(x_i)\nabla V_i^*\\ &- \sum_{j\in N_i} e_{ij}(d_j + e_{j0})g_j(x_j)R_{jj}^{-1} g_j^T(x_j)\nabla V_j^*]\\ &+\frac{1}{2}Q_i(\delta_i) + \frac{1}{2}(d_i + e_{i0})^2 \nabla V_i^{*T} g_i(x_i)\times\\ &R_{ii}^{-1} g_i^T(x_i)\nabla V_i^* + \frac{1}{2}\sum_{j\in N_i}(d_j + e_{j0})^2 \times\\ &\nabla V_j^{*T} g_j(x_j)R_{jj}^{-1} R_{ij} R_{jj}^{-1} g_j^T(x_j)\nabla V_j^* = 0\end{aligned} \quad (۱۴)$$

متناظر با هر گره یک معادله CHJ وجود دارد، بنابراین یکی از روش های حل مسئله بازی گرافی دیفرانسیلی N نفره با سیستم غیرخطی، مستلزم حل N معادله دیفرانسیل CHJ ذکر شده در رابطه (۱۴) می باشد. از آنجایی که حل معادلات دیفرانسیل CHJ (۱۴) در حالت کلی بسیار مشکل می باشد، از روش های تقریبی برای حل آنها استفاده می شود.

۴- الگوریتم تکرار سیاست برخط برای بازی های گرافی دیفرانسیلی غیرخطی

در این بخش الگوریتم تکرار سیاست برخط برای یادگیری حل کنترل بهینه بازی های گرافی دیفرانسیلی غیرخطی زمان پیوسته ارایه می گردد. ساختار یادگیری از دو شبکه عصبی نقاد-کنترلر برای هر عامل استفاده می کند، که پاسخ معادله بلمن و سیاست کنترلی را برای آن عامل تقریب می زند. در الگوریتم تکرار سیاست به کار رفته، هر دو شبکه عصبی نقاد و کنترلر به صورت همزمان در زمان حقیقی به روزرسانی می شوند.

فرض ۱- در این مقاله فرضیات استاندارد زیر برای تقریبگرهای شبکه های عصبی در نظر گرفته شده اند.

الف) خطای شبکه عصبی و گرادیان آن روی مجموعه فشرده Ω کراندار هستند.

ب) توابع فعالیت شبکه عصبی و گرادیان آنها کراندار هستند.

فرض۲- برای سیاست های کنترلی فیدبکی قابل قبول، معادلات لیاپانوف (۱۱) دارای حل هموار محلی $V_i^*(x) \ge 0$ می باشد.

[1]. Nash Equilibrium Solution

۴-۱ تقریب تابع هزینه توسط شبکه های عصبی نقاد

بر اساس قانون تقریب مرتبه بالای وایرشتراس [۲۵،۲۶،۲۷،۲۸] ، مجموعه های پایه مستقل وکامل $\sigma_i(\delta_i), i=1,...,N$ و وزن های ثابت شبکه عصبی $W_i,\ i=1,\ldots,N$ وجود دارند به گونه ای که حل V_i برای (۱۱) و ∇V_i به صورت یکنواخت بر روی مجموعه فشرده Ω به صورت زیر تقریب زده می شوند.

$$V_i = W_i^T \sigma_i(\delta_i) + \omega_i(\delta_i) \ , \ i=1,...,N \qquad (۱۵)$$

$$\nabla V_i = \nabla \sigma_i^T W_i + \nabla \omega_i \ , \ i=1,...,N \qquad (۱۶)$$

که در آن $\sigma_i(\delta_i) \in \Re^K$ بردارهای پایه توابع فعالیت شبکه عصبی هستند، K تعداد نرون های لایه مخفی، $\omega_i(\delta_i)$ خطای تقریب شبکه های عصبی می باشد، $\nabla V_i = \frac{\partial V_i}{\partial \delta_i}$ ، $\nabla \sigma_i = \frac{\partial \sigma_i}{\partial \delta_i}$ و $\nabla \omega_i = \frac{\partial \omega_i}{\partial \delta_i}$. اگر تعداد نرون های لایه مخفی $K \to \infty$، خطای تقریب به طور یکنواخت $\omega_i \to 0$ و $\nabla \omega_i \to 0$ [۲۵،۲۷] . بر مبنای فرض ۱، روی مجموعه فشرده Ω، داریم $\|\omega_i\| \le b_{\omega_i}$، $\|\sigma_i\| \le b_{\sigma_i}$، $\|\nabla \sigma_i\| \le b_{\nabla \sigma_i}, \forall i$ و $\|\nabla \omega_i\| \le b_{\nabla \omega_i}$.

با به کارگیری شبکه های عصبی تقریبگر توابع ارزش، که شبکه های عصبی نقاد نامیده می شوند (معادلات (۱۵) و (۱۶))، و سیاست های فیدبکی u_i و u_{-i}، هامیلتونین (۱۲) به صورت زیر به دست می آید.

$$\begin{aligned} H_i(\delta_i, W_i, u_i, u_{-i}) = W_i^T \nabla \sigma_i [\sum_{j\in N_i} e_{ij} (\\ f_i(x_i) - f_j(x_j)) + e_{i0}(f_i(x_i) - f_0(x_0)) + \\ (d_i + e_{i0}) g_i(x_i) u_i - \sum_{j\in N_i} e_{ij} g_j(x_j) u_j] + \\ \frac{1}{2} Q_i(\delta_i) + \frac{1}{2} u_i^T R_{ii} u_i + \frac{1}{2} \sum_{j\in N_i} u_j^T R_{ij} u_j = e_{B_i} \end{aligned} \qquad (۱۷)$$

$$\begin{aligned} e_{B_i} = -(\nabla \omega_i)^T [\sum_{j\in N_i} e_{ij}(f_i(x_i) - f_j(x_j)) \\ + e_{i0}(f_i(x_i) - f_0(x_0)) + (d_i + e_{i0}) g_i(x_i) u_i \\ - \sum_{j\in N_i} e_{ij} g_j(x_j) u_j] \end{aligned} \qquad (۱۸)$$

e_{B_i} خطای باقی مانده ناشی از تقریب شبکه های عصبی می باشند. تحت فرض ۱، این خطاهای باقی مانده روی مجموعه فشرده Ω کراندار هستند، به این معنی که $Sup_{x\in\Omega} \|e_{B_i}\| \le \bar{e}_i \ , \ i=1,...,N$.

فرض ۳- برای یک مجموعه فشرده $\Omega \subset \Re^n$ و $i=1,...,N$:

الف) $f_i(x_i) \le b_f \|x_i\|$

ب) $g_i(x_i)$ توسط یک ثابت مشخص کراندار است $\|g_i(x_i)\| \le b_{g_i}$.

ج) وزن های شبکه های عصبی نقاد توسط ثابت های مشخص کراندار هستند $\|W_i\| < W_{i\max}$.

در واقع وزن های ایده آل شبکه های عصبی نقاد $W_i, i=1,...,N$ که بهترین حل تقریبی برای (۱۷) هستند نامعلوم می باشند و باید به صورت زمان حقیقی تقریب زده شوند. بنابراین، خروجی شبکه های عصبی نقاد $\hat{V}_i$ و معادلات بلمن تقریبی (معادله (۲۰)) می توانند به ترتیب به صورت زیر نوشته شوند.

$$\hat{V}_i = \hat{W}_i^T \sigma_i(\delta_i) \qquad (۱۹)$$

$$\begin{aligned} e_{H_i} = \hat{W}_i^T \nabla \sigma_i [\sum_{j\in N_i} e_{ij}(f_i(x_i) - f_j(x_j)) \\ + e_{i0}(f_i(x_i) - f_0(x_0)) + (d_i + e_{i0}) g_i(x_i) u_i \\ - \sum_{j\in N_i} e_{ij} g_j(x_j) u_j] + \frac{1}{2} Q_i(\delta_i) + \frac{1}{2} u_i^T R_{ii} u_i \\ + \frac{1}{2} \sum_{j\in N_i} u_j^T R_{ij} u_j \end{aligned} \qquad (۲۰)$$

که در آن $\hat{W}_i$ ارزش تخمین زده کنونی از W_i می باشد. توجه شود که خطاهای e_{H_i} معادل زمان پیوسته خطای تفاوت زمانی می باشد [۲۹] . اکنون مسئله پیدا کردن تابع ارزش برای هر عامل به تنظیم پارامترهای شبکه های عصبی نقاد تبدیل شده است به گونه ای که خطای اختلاف زمانی e_{H_i} حداقل شود.

تابع هدف زیر را در نظر بگیرید.

$$E_i = \frac{1}{2} e_{H_i}^T e_{H_i} \qquad (۲۱)$$

برای $i=1,...,N$، با استفاده از (۲۰) و قوانین مشتق زنجیره ای داریم

$$\begin{aligned} \dot{\hat{W}}_i = -\alpha_i \frac{\partial E_i}{\partial \hat{W}_i} = -\alpha_i e_{H_i} \frac{\partial e_{H_i}}{\partial \hat{W}_i} = \\ -\alpha_i \frac{B_i}{(1 + B_i^T B_i)^2} e_{H_i} = -\alpha_i \frac{\bar{B}_i}{m_{s_i}} e_{H_i} \end{aligned} \qquad (۲۲)$$

که در آن

$$\begin{aligned} B_i = \nabla \sigma_i (\sum_{j\in N_i} e_{ij}(f_i(x_i) - f_j(x_j)) \\ + e_{i0}(f_i(x_i) - f_0(x_0)) + (d_i + e_{i0}) g_i(x_i) u_i \\ - \sum_{j\in N_i} e_{ij} g_j(x_j) u_j) , \end{aligned}$$

$m_{s_i} = 1 + B_i^T B_i$ ، $\bar{B}_i = \frac{B_i}{1 + B_i^T B_i}$. $\alpha_i > 0$ نرخ یادگیری می باشد و $(1 + B_i^T B_i)^2$ برای نرمال سازی به کار می رود.

تعریف ۳ (پایداری کراندار نهایی یکنواخت[1])- سیگنال وابسته به زمان $\vartheta(t)$ کراندار نهایی یکنواخت گفته می شود اگر مجموعه فشرده $S \subset R^n$ وجود داشته باشد به طوری که برای همه $\vartheta(0) \in S$ یک کران β و زمان $T(\beta, \vartheta(0))$ وجود داشته باشد به طوری که برای همه $t \geq t_0 + T$ داشته باشیم $\|\vartheta(t)\| \leq \beta$ [۳۰].

تعریف ۴ (تحریک پایا (PE[2]))- سیگنال برداری کراندار $\bar{B}_i(t), i = 1,...,N$ روی بازه $[t, t+T_i]$ تحریک پایا می باشد [۱۶] اگر وجود داشته باشند $T_i > 0, \gamma_i > 0$ و $\gamma_{i+N} > 0$ به طوری که برای هر t، $\gamma_i I \leq \int_t^{t+T_i} \bar{B}_i(\tau)\bar{B}_i^T(\tau)d\tau \leq \gamma_{i+N} I$.

لم ۱- مطابق با (۱۷) و با جایگذاری $V_i(t) = W_i^T \sigma_i(\delta_i)$، داریم

$$W_i^T B_i = -\frac{1}{2}Q_i(\delta_i) - \frac{1}{2}u_i^T R_{ii} u_i - \frac{1}{2}\sum_{j \in N_i} u_j^T R_{ij} u_j + e_{B_i}$$

لم ۲- $(u_i, u_{-i}), \forall i$ را به عنوان یک مجموعه سیاست فیدبکی کراندار قابل قبول در نظر بگیرید و (۲۲) را به عنوان تنظیم شبکه های عصبی نقاد در نظر گرفته و فرض کنید که $\bar{B}_i$ تحریک پایا باشد. خطای باقی مانده در (۱۸)، $Sup_{x \in \Omega} \|e_{B_i}\| \leq \bar{e}_i$, $i = 1,...,N$ می باشد. آنگاه خطای پارامترهای نقاد به صورت نمایی به یک مجموعه باقی مانده همگرا می شود.

اثبات لم ۲ – با استفاده از لم ۱

$$-W_i^T B_i + e_{B_i} = \frac{1}{2}Q_i(\delta_i) + \frac{1}{2}u_i^T R_{ii} u_i + \frac{1}{2}\sum_{j \in N_i} u_j^T R_{ij} u_j \quad (۲۳)$$

با قرار دادن (۲۳) در (۲۰) و انجام برخی محاسبات، خطای تفاوت زمانی به صورت زیر به دست می آید

$$e_{H_i} = -\tilde{W}_i^T B_i + e_{B_i}. \quad (۲۴)$$

با قرار دادن (۲۴) در (۲۲) و با در نظر گرفتن $\bar{B}_i = \frac{B_i}{(1 + B_i^T B_i)}$ و $m_{s_i} = 1 + B_i^T B_i$، دینامیک خطای وزن های نقاد (برای $i = 1,...,N$) به صورت زیر به دست می آید

$$\dot{\tilde{W}}_i = -\dot{W}_i = -\alpha_i \bar{B}_i \bar{B}_i^T \tilde{W}_i + \alpha_i \frac{\bar{B}_i}{m_{s_i}} e_{B_i}. \quad (۲۵)$$

فرض کنید که (۲۵) یک سیستم متغیر با زمان خطی با ورودی e_{B_i} می باشد. آنگاه حل $\tilde{W}_i$ به صورت زیر داده می شود

$$\tilde{W}_i(t) = \phi_i(t,t_0)\tilde{W}_i(0) + \alpha_i \int_{t_0}^{t} \phi_i(\tau,t_0) \frac{\bar{B}_i}{m_{s_i}} e_{B_i} d\tau \quad (۲۶)$$

که برای $i = 1,...,N$ ماتریس گذر حالت آن به صورت زیر تعریف می شود

$$\frac{\partial \phi_i(t,t_0)}{\partial t} = -\alpha_i \bar{B}_i \bar{B}_i^T \phi_i(t,t_0), \ \phi_i(t_0,t_0) = I \quad (۲۷)$$

ماتریس گذر حالت ϕ_i به صورت نمایی پایدار است به شرط آنکه $\bar{B}_i$ تحریک پایا باشد[۳۱] . از آنجایی که $\bar{B}_i$ تحریک پایا می باشد، $\|\bar{B}_i\| \leq 1$ و $\|e_{B_i}\| \leq \bar{e}_i$، برای $\eta_{i_1}, \eta_{i_2} > 0$، به دست می آوریم

$$\|\tilde{W}_i\| \leq \eta_{i_1} e^{-\eta_{i_2} t} + \frac{\alpha_i}{m_{s_i}\eta_{i_2}} \bar{e}_i, \quad i = 1,...,N. \quad (۲۸)$$

به این ترتیب اثبات کامل می گردد.

۵- حل برخط بازی های گرافی دیفرانسیلی برای سیستم های چندعاملی غیرخطی

اکنون الگوریتم کنترل بهینه که مسئله بازی های گرافی دیفرانسیلی غیرخطی را حل می نماید بر اساس الگوریتم تکرار سیاست برخط ارایه می گردد. پایداری حلقه بسته و همگرایی به حل تقریبی بازی نیز تضمین می گردد.

$V_i(x) = W_i^T \sigma_i, i = 1,...,N$ را به عنوان حل حداقل مربعات رابطه (۱۷) در نظر می گیریم و تعریف می کنیم ($i = 1,...,N$)

$$u_i = -(d_i + e_{i0}) R_{ii}^{-1} g_i^T(x_i) \nabla \sigma_i^T W_i \quad (۲۹)$$

(۲۹) به عنوان سیاست کنترلی بهینه ایده آل با W_i نامشخص در نظر گرفته می شود. با به کارگیری شبکه های عصبی کنترلر، تخمین سیاست های کنترلی بهینه به صورت زیر نوشته می شود

$$u_{i+N} = -(d_i + e_{i0}) R_{ii}^{-1} g_i^T(x_i) \nabla \sigma_i^T \hat{W}_{i+N} \quad (۳۰)$$

که $\hat{W}_{i+N}$ مقدار تخمین زده شده فعلی از وزن ایده آل شبکه عصبی W_i می باشد. خطاهای تخمین شبکه های عصبی نقاد و کنترلر به ترتیب در روابط (۳۱) و (۳۲) تعریف شده است

$$\tilde{W}_i = W_i - \hat{W}_i \quad (۳۱)$$

$$\tilde{W}_{i+N} = W_i - \hat{W}_{i+N} \quad (۳۲)$$

قضیه ۱- سیستم دینامیکی (۷) و تعاریف بازی گرافی دیفرانسیلی چندعاملی را در نظر بگیرید. ورودی های کنترلی از رابطه (۳۰) داده می شوند و قانون تنظیم شبکه عصبی نقاد برای عامل i به صورت زیر ارایه می شود.

[1] Uniformly Ultimately bounded
[2] Persistently Exciting

$$\dot{\hat{W}}_i = -\alpha_i e_{H_i} \frac{\partial e_{H_i}}{\partial \hat{W}_i} =$$
$$-\alpha_i \frac{B_{i+N}}{(1+B_{i+N}{}^T B_{i+N})^2}[B_{i+N}{}^T \hat{W}_i + \frac{1}{2}Q_i(\delta_i)$$
$$+\frac{1}{2}(d_i+e_{i0})^2 \hat{W}_{i+N}{}^T \bar{D}_i \hat{W}_{i+N} \quad (۳۳)$$
$$+\frac{1}{2}\sum_{j\in N_i}(d_j+e_{j0})^2 \hat{W}_{j+N}{}^T \nabla\sigma_j g_j(x_j)\times$$
$$R_{jj}{}^{-T} R_{ij} R_{jj}{}^{-1} g_j{}^T(x_j)\nabla\sigma_j{}^T \hat{W}_{j+N}]$$

و قانون تنظیم شبکه عصبی کنترلر متناظر با عامل i به صورت زیر داده می شود

$$\dot{\hat{W}}_{i+N} = -\alpha_{i+N}\{(S_i \hat{W}_{i+N} - F_i \hat{W}_i)$$
$$-\frac{1}{2}\bar{D}_i (d_i+e_{i0})^2 \hat{W}_{i+N} \frac{\bar{B}_{i+N}{}^T}{m_{s_i}} \hat{W}_i$$
$$-\frac{1}{2}\sum_{j\in i_N}(d_i+e_{i0})^2 \nabla\sigma_i g_i(x_i)\times \quad (۳۴)$$
$$R_{ii}{}^{-T} R_{ji} R_{ii}{}^{-1} g_i{}^T(x_i)\nabla\sigma_i{}^T \hat{W}_{i+N} \frac{\bar{B}_{j+N}{}^T}{m_{s_j}} \hat{W}_j$$

که در آن

$$\bar{D}_i = \nabla\sigma_i g_i(x_i) R_{ii}{}^{-1} g_i{}^T(x_i)\nabla\sigma_i{}^T \text{ ، } i=1,...,N \text{ ،}$$

$$\bar{B}_{i+N} = \frac{B_{i+N}}{1+B_{i+N}{}^T B_{i+N}} \text{ ، } m_{s_{i+N}} = 1+B_{i+N}{}^T B_{i+N} \text{ ،}$$

$$B_{i+N} = \nabla\sigma_i[\sum_{j\in N_i} e_{ij}(f(x_i)-f(x_j))$$
$$+e_{i0}(f(x_i)-f(x_0)) + (d_i+e_{i0})g_i(x_i)u_{i+N} .$$
$$-\sum_{j\in N_i} e_{ij} g_j(x_j) u_{j+N}]$$

S_i و F_i پارامترهای تنظیم مثبت معین قطری هستند. $\forall i, \bar{B}_{i+N}$ تحریک پایا می باشد و فرایض ۱ تا ۳ برقرار می باشد. آنگاه با به کارگیری تعداد کافی نرون برای شبکه های عصبی، حالت های سیستم حلقه بسته $\delta_i(t)$ و خطای شبکه عصبی نقاد $\tilde{W}_i$ و خطای شبکه عصبی کنترلر $\tilde{W}_{i+N}$ کراندار نهایی یکنواخت هستند.

اثبات قضیه ۱- تابع لیاپانوف زیر را در نظر بگیرید

$$L(t) = \sum_{i=1}^{N}\{V_i(t) + \quad (۳۵)$$
$$\frac{1}{2}\tilde{W}_i{}^T \alpha_i{}^{-1}\tilde{W}_i + \frac{1}{2}\tilde{W}_{i+N}{}^T \alpha_{i+N}{}^{-1}\tilde{W}_{i+N}\}$$

که در آن $V_i(t) = W_i^T \sigma_i, i=1,...,N$ حل تقریبی (۱۴) می باشد. مشتق زمانی تابع لیاپانوف به صورت زیر داده می شود.

$$\dot{L}(t) = \sum_{i=1}^{N}\{\dot{V}_i(t) - \quad (۳۶)$$
$$\tilde{W}_i^T \alpha_i{}^{-1}\dot{\hat{W}}_i - \tilde{W}_{i+N}{}^T \alpha_{i+N}{}^{-1}\dot{\hat{W}}_{i+N}\}$$

اولین جمله (۳۶) به صورت زیر می باشد

$$\sum_{i=1}^{N}\dot{V}_i(t) = \sum_{i=1}^{N}\{\frac{\partial V_i}{\partial\delta_i}(\dot{\delta}_i(t))\} =$$
$$\sum_{i=1}^{N}\{W_i^T\nabla\sigma_i(\sum_{j\in N_i} e_{ij}(f_i(x_i)-f_j(x_j)) \quad (۳۷)$$
$$+e_{i0}(f_i(x_i)-f_0(x_0)) + (d_i+e_{i0})g_i(x_i)u_{i+N}$$
$$-\sum_{j\in N_i} e_{ij} g_j(x_j)u_{j+N}) + \sum_{i=1}^{N}\omega_{i0}$$

که در آن

$$\sum_{i=1}^{N}\omega_{i0} = \sum_{i=1}^{N}\{\nabla\omega_i[\sum_{j\in N_i} e_{ij}(f_i(x_i)$$
$$-f_j(x_j)) + e_{i0}(f_i(x_i)-f_0(x_0)) -$$
$$(d_i+e_{i0})^2 g_i(x_i) R_{ii}{}^{-1} g_i^T(x_i)\nabla\sigma_i{}^T\hat{W}_{i+N} +$$
$$\sum_{j\in N_i} e_{ij}(d_j+e_{j0}) g_j(x_j) R_{jj}{}^{-1} g_j^T(x_j)\nabla\sigma_j{}^T\hat{W}_{j+N}]\}.$$

ثابت به اندازه بزرگ b_δ وجود دارد به طوری که

$$\sum_{i=1}^{N}\|\omega_{i0}\| \le \sum_{i=1}^{N}\{b_{\nabla\omega_i} b_f b_\delta \|\delta_i\| + b_{\nabla\omega_i}(d_i+e_{i0})^2 \times$$
$$\|g_i(x_i)\|^2 \sigma_{\min}(R_{ii}) b_{\nabla\sigma_i}(\|W_i\| + \|\tilde{W}_{i+N}\|) +$$
$$+b_{\nabla\omega_i}\sum_{j\in N_i} e_{ij}(d_j+e_{j0})\|g_j(x_j)\|^2 \times$$
$$\sigma_{\min}(R_{jj}) b_{\sigma_j}(\|W_j\| + \|\tilde{W}_{j+N}\|)\}.$$

با استفاده از لم ۱ و (۳۲)-(۲۹)، $\dot{V}_i$ به صورت زیر به دست می آید

$$\sum_{i=1}^{N}\dot{V}_i(t) =$$
$$\sum_{i=1}^{N}\{\dot{\bar{L}}_{V_i} + (d_i+e_{i0})^2 W_i^T \bar{D}_i \tilde{W}_{i+N} + \omega_{i0} \quad (۳۸)$$
$$-W_i^T\nabla\sigma_i\sum_{j\in N_i} e_{ij}(d_j+e_{j0}) g_j(x_j)\times$$
$$R_{jj}{}^{-1} g_j^T(x_j)\nabla\sigma_j{}^T\tilde{W}_{j+N}\}$$

که در آن

$$\dot{\bar{L}}_{V_i} = -\frac{1}{2}Q_i(\delta_i) - \frac{1}{2}(d_i+e_{i0})^2 W_i^T \bar{D}_i W_i + e_{B_i}$$
$$-\frac{1}{2}\sum_{j\in N_i}(d_j+e_{j0})^2 W_j{}^T\nabla\sigma_j g_j(x_j)\times$$
$$R_{jj}{}^{-T} R_{ij}\ R_{jj}{}^{-1} g_j{}^T(x_j)\nabla\sigma_j{}^T W_j .$$

برای جمله سوم از $\dot{L}$ ، $\sum_{i=1}^{N}\dot{L}_i=-\sum_{i=1}^{N}\tilde{W}_i^T\alpha_i^{-1}\dot{\hat{W}}_i$ ،

داریم

$$\dot{\tilde{W}}_i=-\dot{\hat{W}}_i=\alpha_i\frac{B_{i+N}}{(1+B_{i+N}{}^TB_{i+N})^2}\{B_{i+N}{}^T\hat{W}_i+\frac{1}{2}Q_i(\delta_i)+\frac{1}{2}(d_i+e_{i0})^2\hat{W}_{i+N}{}^T\bar{D}_i\hat{W}_{i+N}+\frac{1}{2}\sum_{j\in N_i}(d_j+e_{j0})^2\hat{W}_{j+N}{}^T\nabla\sigma_j g_j(x_j)\times R_{jj}{}^{-T}R_{ij}R_{jj}{}^{-1}g_j{}^T(x_j)\nabla\sigma_j{}^T\hat{W}_{j+N}\} \quad (۳۹)$$

با استفاده از (۳۱) و (۳۲) داریم

$$\begin{aligned}\dot{\tilde{W}}_i=\alpha_i\frac{B_{i+N}}{(1+B_{i+N}{}^TB_{i+N})^2}\{&B_{i+N}{}^TW_i-B_{i+N}{}^T\tilde{W}_i\\&+\frac{1}{2}Q_i(\delta_i)+\frac{1}{2}(d_i+e_{i0})^2W_i{}^T\bar{D}_iW_i\\&+\frac{1}{2}(d_i+e_{i0})^2\tilde{W}_{i+N}{}^T\bar{D}_i\tilde{W}_{i+N}\\&-\frac{1}{2}(d_i+e_{i0})^2W_i{}^T\bar{D}_i\tilde{W}_{i+N}\\&-\frac{1}{2}(d_i+e_{i0})^2\tilde{W}_{i+N}{}^T\bar{D}_iW_i\\&+\frac{1}{2}\sum_{j\in N_i}(d_j+e_{j0})^2W_j{}^T\nabla\sigma_jg_j(x_j)\times R_{jj}{}^{-T}R_{ij}R_{jj}{}^{-1}g_j{}^T(x_j)\nabla\sigma_j{}^TW_j\\&+\frac{1}{2}\sum_{j\in N_i}(d_j+e_{j0})^2\tilde{W}_{j+N}{}^T\nabla\sigma_jg_j(x_j)\times R_{jj}{}^{-T}R_{ij}R_{jj}{}^{-1}g_j{}^T(x_j)\nabla\sigma_j{}^T\tilde{W}_{j+N}\\&-\frac{1}{2}\sum_{j\in N_i}(d_j+e_{j0})^2W_j{}^T\nabla\sigma_jg_j(x_j)\times R_{jj}{}^{-T}R_{ij}R_{jj}{}^{-1}g_j{}^T(x_j)\nabla\sigma_j{}^T\tilde{W}_{j+N}\\&-\frac{1}{2}\sum_{j\in N_i}(d_j+e_{j0})^2\tilde{W}_{j+N}{}^T\nabla\sigma_jg_j(x_j)\times R_{jj}{}^{-T}R_{ij}R_{jj}{}^{-1}g_j{}^T(x_j)\nabla\sigma_j{}^TW_j\}\end{aligned} \quad (۴۰)$$

با استفاده از لم ۱ داریم

$$\begin{aligned}\dot{\tilde{W}}_i=\alpha_i\frac{B_{i+N}}{(1+B_{i+N}{}^TB_{i+N})^2}\{&-B_{i+N}{}^T\tilde{W}_i-\\&W_i^T\nabla\sigma_i\sum_{j\in N_i}e_{ij}(d_j+e_{j0})g_j(x_j)\times\\&R_{jj}{}^{-1}g_j{}^T(x_j)\nabla\sigma_j{}^T\tilde{W}_{j+N}\\&+e_{B_i}+\frac{1}{2}(d_i+e_{i0})^2\tilde{W}_{i+N}{}^T\bar{D}_i\tilde{W}_{i+N}\\&+\frac{1}{2}\sum_{j\in N_i}(d_j+e_{j0})^2\tilde{W}_{j+N}{}^T\nabla\sigma_jg_j(x_j)\times\\&R_{jj}{}^{-T}R_{ij}R_{jj}{}^{-1}g_j{}^T(x_j)\nabla\sigma_j{}^T\tilde{W}_{j+N}\\&-\sum_{j\in N_i}(d_j+e_{j0})^2W_j{}^T\nabla\sigma_jg_j(x_j)\times\\&R_{jj}{}^{-T}R_{ij}R_{jj}{}^{-1}g_j{}^T(x_j)\nabla\sigma_j{}^T\tilde{W}_{j+N}\}\end{aligned} \quad (۴۱)$$

به این ترتیب $\sum_{i=1}^{N}\dot{L}_i$ می تواند به صورت زیر نوشته شود

$$\begin{aligned}\sum_{i=1}^{N}\dot{L}_i=-\sum_{i=1}^{N}\tilde{W}_i^T\alpha_i^{-1}\dot{\hat{W}}_i=&\\\sum_{i=1}^{N}\{\dot{\bar{L}}_i+\frac{1}{2}\tilde{W}_i^T\frac{B_{i+N}}{(1+B_{i+N}{}^TB_{i+N})^2}[&\\(d_i+e_{i0})^2\tilde{W}_{i+N}{}^T\bar{D}_i\tilde{W}_{i+N}&\\+\sum_{j\in N_i}(d_j+e_{j0})^2\tilde{W}_{j+N}{}^T\nabla\sigma_jg_j(x_j)\times&\\R_{jj}{}^{-T}R_{ij}R_{jj}{}^{-1}g_j{}^T(x_j)\nabla\sigma_j{}^T\tilde{W}_{j+N}&\\-2(W_i^T\nabla\sigma_i\sum_{j\in N_i}e_{ij}(d_j+e_{j0})g_j(x_j)\times&\\R_{jj}{}^{-1}g_j{}^T(x_j)\nabla\sigma_j{}^T\tilde{W}_{j+N}&\\-\sum_{j\in N_i}(d_j+e_{j0})^2W_j{}^T\nabla\sigma_jg_j(x_j)\times&\\R_{jj}{}^{-T}R_{ij}R_{jj}{}^{-1}g_j{}^T(x_j)\nabla\sigma_j{}^T\tilde{W}_{j+N})]\}\end{aligned} \quad (۴۲)$$

که در آن $\dot{\bar{L}}_i=\tilde{W}_i^T\bar{B}_{i+N}(-B_{i+N}{}^T\tilde{W}_i+\frac{e_{B_i}}{m_{s_i}})$. با استفاده از جملات (۳۸) و (۴۲) و انجام یک سری تغییرات با به کارگیری (۳۱)-(۳۲)، (۳۶) به صورت زیر به دست می آید

$$\begin{aligned}\dot{L}(t)=\sum_{i=1}^{N}\ &\{\dot{\bar{L}}_{V_i}+\dot{\bar{L}}_i\\&+\omega_{i0}+(d_i+e_{i0})^2W_i^T\bar{D}_i\tilde{W}_{i+N}\\&-W_i^T\nabla\sigma_i\sum_{j\in N_i}e_{ij}(d_j+e_{j0})g_j(x_j)\times\\&R_{jj}^{-1}g_j^T(x_j)\nabla\sigma_j^T\tilde{W}_{j+N}\\&-\tilde{W}_i^T\frac{\bar{B}_{i+N}}{m_{s_i}}W_i^T\nabla\sigma_i\sum_{j\in N_i}e_{ij}(d_j+e_{j0})g_j(x_j)\times\\&R_{jj}^{-1}g_j^T(x_j)\nabla\sigma_j^T\tilde{W}_{j+N}\\&-\tilde{W}_i^T\frac{\bar{B}_{i+N}}{m_{s_i}}\sum_{j\in N_i}(d_j+e_{j0})^2W_j^T\nabla\sigma_jg_j(x_j)\times\\&R_{jj}^{-T}R_{ij}R_{jj}^{-1}g_j^T(x_j)\nabla\sigma_j^T\tilde{W}_{j+N}\\&+\frac{1}{2}(d_i+e_{i0})^2\tilde{W}_{i+N}^T\bar{D}_i\hat{W}_{i+N}\frac{\bar{B}_{i+N}^T}{m_{s_i}}\hat{W}_i\\&+\frac{1}{2}(d_i+e_{i0})^2\tilde{W}_{i+N}^T\bar{D}_i\tilde{W}_{i+N}\frac{\bar{B}_{i+N}^T}{m_{s_i}}W_i\\&-\frac{1}{2}(d_i+e_{i0})^2\tilde{W}_{i+N}^T\bar{D}_iW_i\frac{\bar{B}_{i+N}^T}{m_{s_i}}W_i\\&+\frac{1}{2}(d_i+e_{i0})^2\tilde{W}_{i+N}^T\bar{D}_iW_i\frac{\bar{B}_{i+N}^T}{m_{s_i}}\tilde{W}_i\\&+\frac{1}{2}\sum_{j\in N_i}(d_j+e_{j0})^2\tilde{W}_{j+N}^T\nabla\sigma_jg_j(x_j)\times\\&R_{jj}^{-T}R_{ij}R_{jj}^{-1}g_j^T(x_j)\nabla\sigma_j^T\hat{W}_{j+N}\frac{\bar{B}_{i+N}^T}{m_{s_i}}\hat{W}_i\\&+\frac{1}{2}\sum_{j\in N_i}(d_j+e_{j0})^2\tilde{W}_{j+N}^T\nabla\sigma_jg_j(x_j)\times\\&R_{jj}^{-T}R_{ij}R_{jj}^{-1}g_j^T(x_j)\nabla\sigma_j^T\tilde{W}_{j+N}\frac{\bar{B}_{i+N}^T}{m_{s_i}}W_i\\&-\frac{1}{2}\sum_{j\in N_i}(d_j+e_{j0})^2\tilde{W}_{j+N}^T\nabla\sigma_jg_j(x_j)\times\\&R_{jj}^{-T}R_{ij}R_{jj}^{-1}g_j^T(x_j)\nabla\sigma_j^TW_j\frac{\bar{B}_{i+N}^T}{m_{s_i}}W_i\\&+\frac{1}{2}\sum_{j\in N_i}(d_j+e_{j0})^2\tilde{W}_{j+N}^T\nabla\sigma_jg_j(x_j)\times\\&R_{jj}^{-T}R_{ij}R_{jj}^{-1}g_j^T(x_j)\nabla\sigma_j^TW_j\frac{\bar{B}_{i+N}^T}{m_{s_i}}\tilde{W}_i\\&-\tilde{W}_{i+N}^T\alpha_{i+N}^{-1}\dot{\hat{W}}_{i+N}\}\end{aligned}\qquad(۴۳)$$

برای به کارگیری جملات زیر که در (۴۳) ظاهر شده اند، جهت استخراج قانون تنظیم $\dot{\hat{W}}_{i+N}(t)$،

$$\sum_{i=1}^{N}\{\frac{1}{2}(d_i+e_{i0})^2\tilde{W}_{i+N}^T\bar{D}_i\hat{W}_{i+N}\frac{\bar{B}_{i+N}^T}{m_{s_i}}\hat{W}_i\},\qquad(۴۴)$$

$$\begin{aligned}&\sum_{i=1}^{N}\{\frac{1}{2}\sum_{j\in N_i}(d_j+e_{j0})^2\tilde{W}_{j+N}^T\nabla\sigma_jg_j(x_j)\times\\&R_{jj}^{-T}R_{ij}R_{jj}^{-1}g_j^T(x_j)\nabla\sigma_j^T\hat{W}_{j+N}\frac{\bar{B}_{i+N}^T}{m_{s_i}}\hat{W}_i\}\end{aligned}\qquad(۴۵)$$

ابتدا $\sum_{j\in N_i}$ در (۴۵) را باز نموده و آنگاه $\sum_{i=1}^{N}$ را بسط می دهیم. آنگاه از جملاتی که شامل $\tilde{W}_{i+N}(t)$ می باشند، $\tilde{W}_{i+N}(t)$ را فاکتورگیری می نماییم. انجام مراحل مذکور باعث تغییر (۴۵) به صورت زیر می گردد.

$$\begin{aligned}&\sum_{i=1}^{N}\tilde{W}_{i+N}^T\sum_{j\in i_N}\frac{1}{2}(d_i+e_{i0})^2\nabla\sigma_ig_i(x_i)\times\\&R_{ii}^{-T}R_{ji}R_{ii}^{-1}g_i^T(x_i)\nabla\sigma_i^T\hat{W}_{i+N}\frac{\bar{B}_{j+N}^T}{m_{s_j}}\hat{W}_j\end{aligned}\qquad(۴۶)$$

که در آن، i_N نشان دهنده مجموعه گره هایی می باشد که عامل i در همسایگی آنها می باشد. اکنون با به کارگیری (۴۴)، (۴۶) و $-\sum_{i=1}^{N}\tilde{W}_{i+N}^T\alpha_{i+N}^{-1}\dot{\hat{W}}_{i+N}$، $\dot{L}(t)$ می تواند به صورت زیر نوشته شود

$$\begin{aligned}\dot{L}(t)=\sum_{i=1}^{N}\ &\{\dot{\bar{L}}_{V_i}+\dot{\bar{L}}_i+\omega_{i0}+\dot{\bar{L}}_{i+N}\\&-\tilde{W}_{i+N}^T[\alpha_{i+N}^{-1}\dot{\hat{W}}_{i+N}\\&-\frac{1}{2}\bar{D}_i(d_i+e_{i0})^2\hat{W}_{i+N}\frac{\bar{B}_{i+N}^T}{m_{s_i}}\hat{W}_i\\&-\frac{1}{2}\sum_{j\in i_N}(d_i+e_{i0})^2\nabla\sigma_ig_i(x_i)\times\\&R_{ii}^{-T}R_{ji}R_{ii}^{-1}g_i^T(x_i)\nabla\sigma_i^T\hat{W}_{i+N}\frac{\bar{B}_{j+N}^T}{m_{s_j}}\hat{W}_j]\}\end{aligned}\qquad(۴۷)$$

که در آن

$$\dot{\tilde{L}}_{i+N} = (d_i + e_{i0})^2 W_i^T \bar{D}_i \tilde{W}_{i+N} -$$
$$\tilde{W}_i^T \frac{\bar{B}_{i+N}}{m_{s_i}} W_i^T \nabla\sigma_i \sum_{j\in N_i} e_{ij}(d_j + e_{j0}) \times$$
$$g_j(x_j) R_{jj}^{-1} g_j^T(x_j) \nabla\sigma_j^T \tilde{W}_{j+N}$$
$$-\tilde{W}_i^T \frac{\bar{B}_{i+N}}{m_{s_i}} (d_j + e_{j0})^2 W_j^T \nabla\sigma_j g_j(x_j) \times$$
$$R_{jj}^{-T} R_{ij} R_{jj}^{-1} g_j^T(x_j) \nabla\sigma_j^T \tilde{W}_{j+N}$$
$$+\frac{1}{2}(d_i + e_{i0})^2 \tilde{W}_{i+N}^T \bar{D}_i \tilde{W}_{i+N} \frac{\bar{B}_{i+N}^T}{m_{s_i}} W_i$$
$$-\frac{1}{2}(d_i + e_{i0})^2 \tilde{W}_{i+N}^T \bar{D}_i W_i \frac{\bar{B}_{i+N}^T}{m_{s_i}} W_i$$
$$+\frac{1}{2}(d_i + e_{i0})^2 \tilde{W}_{i+N}^T \bar{D}_i W_i \frac{\bar{B}_{i+N}^T}{m_{s_i}} \tilde{W}_i$$
$$-W_i^T \frac{\partial\sigma_i}{\partial\delta_i} \sum_{j\in N_i} e_{ij}(d_j + e_{j0}) \times$$
$$g_j(x_j) R_{jj}^{-1} B_j^T \frac{\partial\sigma_j}{\partial\delta_j}^T \tilde{W}_{j+N}$$
$$+\frac{1}{2} \sum_{j\in N_i} (d_j + e_{j0})^2 \tilde{W}_{j+N}^T \nabla\sigma_j g_j(x_j) \times$$
$$R_{jj}^{-T} R_{ij} R_{jj}^{-1} g_j^T(x_j) \nabla\sigma_j^T \tilde{W}_{j+N} \frac{\bar{B}_{i+N}^T}{m_{s_i}} W_i$$
$$-\frac{1}{2} \sum_{j\in N_i} (d_j + e_{j0})^2 \tilde{W}_{j+N}^T \nabla\sigma_j g_j(x_j) \times$$
$$R_{jj}^{-T} R_{ij} R_{jj}^{-1} g_j^T(x_j) \nabla\sigma_j^T W_j \frac{\bar{B}_{i+N}^T}{m_{s_i}} W_i$$
$$+\frac{1}{2} \sum_{j\in N_i} (d_j + e_{j0})^2 \tilde{W}_{j+N}^T \nabla\sigma_j g_j(x_j) \times$$
$$R_{jj}^{-T} R_{ij} R_{jj}^{-1} g_j^T(x_j) \nabla\sigma_j^T W_j \frac{\bar{B}_{i+N}^T}{m_{s_i}} \tilde{W}_i$$

براساس (۴۷)، $\dot{\hat{W}}_{i+N}$ به صورت زیر در نظر گرفته می شود

$$\dot{\hat{W}}_{i+N} = -\alpha_{i+N}\{(S_i \hat{W}_{i+N} - F_i \hat{W}_i)$$
$$-\frac{1}{2}\bar{D}_i (d_i + e_{i0})^2 \hat{W}_{i+N} \frac{\bar{B}_{i+N}^T}{m_{s_i}} \hat{W}_i$$
$$-\frac{1}{2} \sum_{j\in i_N} (d_i + e_{i0})^2 \nabla\sigma_i g_i(x_i) \times \qquad (۴۸)$$
$$R_{ii}^{-T} R_{ji} R_{ii}^{-1} g_i^T(x_i) \nabla\sigma_i^T \hat{W}_{i+N} \frac{\bar{B}_{j+N}^T}{m_{s_j}} \hat{W}_j$$

با به کارگیری (۴۸)، $\dot{\boldsymbol{L}}(\mathbf{t})$ به صورت زیر به دست می آید

$$\dot{L}(t) = \sum_{i=1}^{N} \{ \dot{\tilde{L}}_{V_i} + \dot{\tilde{L}}_i + \dot{\omega}_{i0} + \dot{\tilde{L}}_{i+N} + \tilde{W}_{i+N}^T S_i W_i \qquad (۴۹)$$
$$-\tilde{W}_{i+N}^T S_i \tilde{W}_{i+N} - \tilde{W}_{i+N}^T F_i W_i + \tilde{W}_{i+N}^T F_i \tilde{W}_i \}$$

از آنجایی که $Q_i(\delta_i) > 0,\ i = 1,...,N$، وجود دارد به طوری که $q_i > 0, \forall i$ $\delta_i^T q_i \delta_i < Q_i(\delta_i)$، بنابراین $-\delta_i^T q_i \delta_i > -Q_i(\delta_i)$.

با به کارگیری $\tilde{Z}_i = \left[\delta_i\ ,\ \tilde{W}_i\ ,\ \tilde{W}_{i+N}\ ,\ \tilde{W}_{j+N}\right]^T$، به $\dot{\boldsymbol{L}}$ صورت زیر نوشته می شود

$$\dot{L}(t) = \sum_{i=1}^{N} \{C_i - \tilde{Z}_i^T M_i \tilde{Z}_i + D_i \tilde{Z}_i \} \qquad (۵۰)$$

که در آن

$$M_i = \begin{bmatrix} m_{11}^i & m_{12}^i & m_{13}^i & m_{14}^i \\ m_{21}^i & m_{22}^i & m_{23}^i & m_{24}^i \\ m_{31}^i & m_{32}^i & m_{33}^i & m_{34}^i \\ m_{41}^i & m_{42}^i & m_{43}^i & m_{44}^i \end{bmatrix}, D_i = \left[d_1^i,\ d_2^i,\ d_3^i,\ d_4^i\right],$$

$$m_{11}^i = \frac{1}{2} q_i\ ,\ m_{22}^i = \bar{B}_{i+N} \bar{B}_{i+N}^T\ ,$$

$$m_{12}^i = m_{13}^i = m_{14}^i = m_{41}^i = m_{31}^i = m_{21}^i = 0,\ m_{34}^i = m_{43}^i = 0,$$

$$m_{23}^i = -\frac{1}{4m_{s_i}} (d_i + e_{i0})^2 \bar{D}_i W_i \bar{B}_{i+N}^T - \frac{1}{2} F_i = m_{32}^T\ ,$$

$$m_{24}^i = +\frac{1}{2m_{s_i}} \bar{B}_{i+N} W_i^T \nabla\sigma_i \times$$
$$\sum_{j\in N_i} e_{ij}(d_j + e_{j0}) g_j(x_j) R_{jj}^{-1} g_j^T(x_j) \nabla\sigma_j^T$$
$$+\frac{1}{2m_{s_i}} \bar{B}_{i+N} \sum_{j\in N_i} (d_j + e_{j0})^2 W_j^T \nabla\sigma_j \times$$
$$g_j(x_j) R_{jj}^{-T} R_{ij} R_{jj}^{-1} g_j^T(x_j) \nabla\sigma_j^T$$
$$-\frac{1}{4m_{s_i}} \sum_{j\in N_i} (d_j + e_{j0})^2 \nabla\sigma_j g_j(x_j) \times$$
$$R_{jj}^{-T} R_{ij} R_{jj}^{-1} g_j^T(x_j) \nabla\sigma_j^T W_j = m_{42}^T\ ,$$
$$m_{33}^i = -\frac{1}{2}(d_i + e_{i0})^2 \bar{D}_i \frac{\bar{B}_{i+N}^T}{m_{s_i}} W_i + S_i\ ,$$

$$m_{44}^i = -\frac{1}{4}(d_j + e_{j0})^2 \nabla\sigma_j g_j(x_j) \times$$
$$R_{jj}^{-T} R_{ij} R_{jj}^{-1} g_j^T(x_j) \nabla\sigma_j^T \frac{\bar{B}_{i+N}^T}{m_{s_i}} W_i ,$$

$$d_1^i = b_{\frac{\partial\omega_i}{\partial\delta_i}} b_f b_\delta \ , \quad d_2^i = \bar{B}_{i+N} \frac{e_{B_i}}{m_{s_i}} ,$$
$$d_3^i = -\nabla\omega_i (d_i + e_{i0})^2 g_i(\mathrm{x}_i) R_{ii}^{-1} g_i^T(\mathrm{x}_i) \nabla\sigma_i^T$$
$$-\frac{1}{2}(d_i + e_{i0})^2 \bar{D}_i W_i \frac{\bar{B}_{i+N}^T}{m_{s_i}} W_i + (d_i + e_{i0})^2 W_i^T \bar{D}_i$$
$$+W_i^T S_i - W_i^T F_i ,$$
$$d_4^i = \nabla\omega_i \sum_{j\in N_i} e_{ij}(d_j + e_{j0}) g_j(x_j) R_{jj}^{-1} g_j^T(x_j) \nabla\sigma_j^T$$
$$-W_i^T \nabla\sigma_i \sum_{i\in N_j} e_{ij}(d_j + e_{j0}) g_j(x_j) R_{jj}^{-1} g_j^T(x_j) \nabla\sigma_j^T$$
$$-\frac{1}{2}\sum_{j\in N_i} (d_j + e_{j0})^2 \nabla\sigma_j g_j(x_j) \times$$
$$R_{jj}^{-T} R_{ij} R_{jj}^{-1} g_j^T(x_j) \nabla\sigma_j^T W_j \frac{\bar{B}_{i+N}^T}{4m_{s_i}} W_i ,$$

$$C_i = \frac{1}{2}(d_i + e_{i0})^2 \|W_i\|^2 \left\|\bar{D}_i\right\| +$$
$$\frac{1}{2}(d_j + e_{j0})^2 \|W_j\|^2 \times$$
$$\left\|\nabla\sigma_j B_j R_{jj}^{-T} R_{ij} \ R_{jj}^{-1} B_j^T \nabla\sigma_j^T\right\| +$$
$$b_{\nabla\omega_i}(d_i + e_{i0})^2 b_{g_i}^2 \sigma_{\min}(R_{ii}) b_{\nabla\sigma_i} \|W_i\| +$$
$$b_{\nabla\omega_i} e_{ij}(d_j + e_{j0}) b_{g_j}^2 \sigma_{\min}(R_{jj}) b_{\nabla\sigma_j} \|W_j\| ,$$

$$C_i \le C_{i\max} \quad , \quad D_i \le D_{i\max}$$

پارامترهای تنظیم باید به گونه ای انتخاب شوند به طوری که $M_i > 0$. اکنون M_i در (۵۰)، به فرم فشرده زیر نوشته می شود

$$M_i = \begin{bmatrix} \frac{1}{2}q_i & 0 & 0 \\ 0 & I & M_{23}^i \\ 0 & M_{32}^i & M_{33}^i \end{bmatrix}, M_{23}^i = \begin{bmatrix} m_{23}^i & m_{24}^i \end{bmatrix}, \quad (۵۱)$$
$$M_{32}^i = \begin{bmatrix} m_{32}^i \\ m_{42}^i \end{bmatrix}, M_{33}^i = \begin{bmatrix} m_{33}^i & 0 \\ 0 & m_{44}^i \end{bmatrix}$$

برای مثبت معین بودن M_i باید ویژگی های زیر برقرار باشند

الف) $q_i > 0$.

ب) $I > 0$.

ج) مکمل شور[1] برای I، $D_{22}^i = I - M_{23}^i M_{33}^{i\,-1} M_{32}^i > 0$ می باشد که با انتخاب مناسب S_i و F_i می تواند برقرار باشد.

د) مکمل شور برای M_{33}^i، $D_{33}^i = M_{33}^i - M_{32}^i I^{-1} M_{23}^i > 0$ می باشد که با انتخاب مناسب S_i و F_i می تواند برقرار باشد.

اکنون برای (۵۰) داریم

$$\dot{L} < \sum_{i=1}^{N} \{-\left\|\tilde{Z}_i\right\|^2 \sigma_{\min}(M_i) + D_{i\max}\left\|\tilde{Z}_i\right\| + C_{i\max}\} \quad (۵۲)$$

با کامل کردن مربعات، مشتق لیاپانوف منفی خواهد بود اگر

$$\left\|\tilde{Z}_i\right\| > \frac{D_{i\max}}{2\sigma_{\min}(M_i)} + \sqrt{\frac{D_{i\max}^{\ 2}}{4\sigma_{\min}^{\ 2}(M_i)}\left\|\tilde{Z}_i\right\| + \frac{C_{i\max}}{\sigma_{\min}(M_i)}} \equiv B_{\tilde{Z}_i} \quad (۵۳)$$

واضح است که اگر نشان دهیم (۵۳) از یک کران مشخص فراتر رود، آنگاه $\dot{L}$ منفی خواهد بود. بنابراین مطابق با قضیه توسعه لیاپانوف استاندارد[2]، تحلیل های بالا نشان می دهد که حالت ها و وزن ها کراندار نهایی یکنواخت هستند [۳۲] .

شرط (۵۳) برقرار است اگر نرم هر قسمت از $\tilde{Z}_i$ از کران $B_{\tilde{Z}_i}$ فراتر رود، یعنی $\delta_i > B_{\tilde{Z}_i}$, $\tilde{W}_i > B_{\tilde{Z}_i}$, $\tilde{W}_{i+N} > B_{\tilde{Z}_i}$, $\tilde{W}_{j+N} > B_{\tilde{Z}_i}$. به این ترتیب اثبات کامل می گردد.

الگوریتم تکرار سیاست توزیع شده برخط:

شروع $(k=0)$: N سیاست کنترلی اولیه قابل قبول $u_i^0(x); i \in N$، را در نظر می گیریم.

۱. (ارزیابی سیاست): با در نظر گرفتن سیاست های کنترلی $u_i^k(x); i \in N$، معادلات لیاپانف (۱۱) برای تقریب ارزش های $\hat{V}_i^k = \hat{W}_i^{kT}\sigma_i^k(z_i); i \in N$ حل می شوند به عبارت دیگر به روزرسانی وزن های ارزش توسط شبکه های عصبی نقاد عامل ها با استفاده از قانون تنظیم (۳۳) به روزرسانی می شود.

(بهبود سیاست): N سیاست کنترلی را با استفاده از رابطه (۳۰) به روز می کنیم. به طوری که به روزرسانی وزن

[1] Schur Complement
[2] Standard Lyapunov Extension Theorem

های شبکه های عصبی کنترلر با استفاده از قانون تنظیم (۳۴) انجام می شود.

۲. $k=k+1$ و در صورت عدم همگرایی V_i^k, u_i^k برو به گام ۱ و در غیر این صورت پایان الگوریتم.

لازم به ذکر است که در الگوریتم تکرار سیاست استاندارد ارزیابی سیاست و بهبود سیاست به صورت پی در پی[1] و برون خط انجام می شود ولی در الگوریتم تکرار سیاست توزیع شده برخط پیشنهاد شده این دو گام به صورت همزمان و برخط انجام می شوند و جهت همزمان سازی عامل ها تنها از اطلاعات محلی هر عامل استفاده می شود.

۶- نتایج شبیه سازی

در این بخش برای نمایش عملکرد و صحت روش معرفی شده، نتایج روش معرفی شده به روی یک مثال ارایه می گردد. یک سیستم چندعاملی دارای ۴ گره، مطابق با شکل ۱ را در نظر بگیرید. وزن های اتصال و وزن یال ها، ۱ در نظر گرفته شده اند.

دینامیک عامل ها به صورت زیر می باشد.

$$\dot{x}_i = f_i(x_i) + g_i(x_i)\,u_i\ ,\quad x_i = \begin{bmatrix} x_{i1} \\ x_{i2} \end{bmatrix},$$

$$f_i(x_i) = \begin{pmatrix} x_{i2} \\ -x_{i1} + \varepsilon(1-{x_{i1}}^2)x_{i2} \end{pmatrix}, i = 1,2,3,4.$$

$$g_1(x_1) = \begin{bmatrix} 0 \\ -0.8 \end{bmatrix},\ g_2(x_2) = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

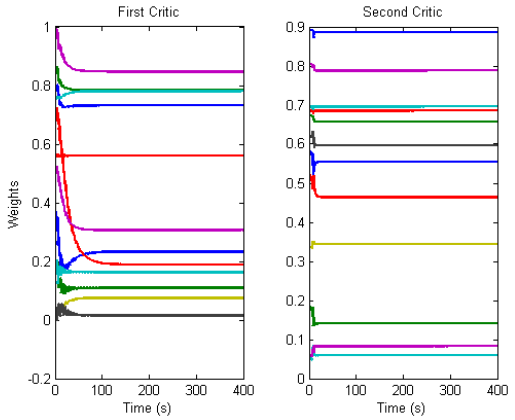
$$g_3(x_3) = \begin{bmatrix} 0 \\ 1.4 \end{bmatrix},\ g_4(x_4) = \begin{bmatrix} 0 \\ -0.2 \end{bmatrix},$$

که در آن $\varepsilon = 0.5$ و دینامیک عامل رهبر $f(x_0) = \begin{pmatrix} x_{02} \\ -x_{01} + \varepsilon(1-{x_{01}}^2)x_{02} \end{pmatrix}$ می باشد. برای $i, j = 1,2,3,4$، در نظر بگیرید $R_{ij} = 1,\ (\mathrm{i} \neq \mathrm{j},\ \mathrm{j} \in \mathrm{N}_i)$ در الگوریتم $Q_i(\delta_i) = \delta_i^T Q_{ii} \delta_i = \delta_i \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \delta_i$, $R_{ii} = 10,$ به کار رفته، پارامترهای طراحی به صورت $S_i = F_i = 5I$ انتخاب شده اند و $\alpha_i = 1$، $i = 1, 2, 3, 4$. بردار اطلاعات در دسترس برای هر عامل، $\delta_i = [\delta_{i1}, \delta_{i2}]^T$, $i = 1,\ldots,4$ می باشد که توسط گراف ارتباطی محدود شده است. توابع فعالیت شبکه های عصبی عامل ها به صورت

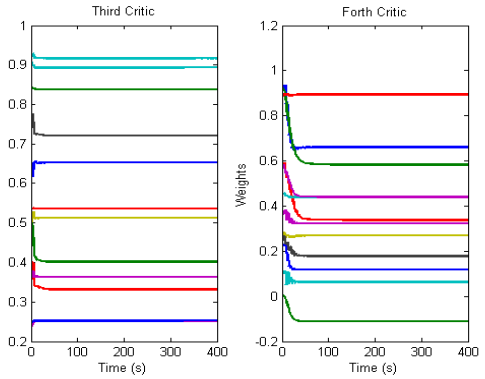
$$\sigma_i = [\delta_{i1}^{\ 2}, \delta_{i1}\delta_{i2}, \delta_{i2}^{\ 2}, \delta_{i1}^{\ 3}, \delta_{i1}^{\ 2}\delta_{i2}, \delta_{i1}\delta_{i2}^{\ 2}, \delta_{i2}^{\ 3}, \delta_{i1}^{\ 4},$$
$$\delta_{i1}^{\ 3}\delta_{i2}, \delta_{i1}^{\ 2}\delta_{i2}^{\ 2}, \delta_{i1}\delta_{i2}^{\ 3}, \delta_{i2}^{\ 4}], \qquad \mathrm{i} = 1,\ldots,4$$

انتخاب می شوند. هر $\sigma_i, i = 1,\ldots,4$ شامل توان های δ_{i1} و δ_{i2} از درجه ۴ می باشد. توجه شود که یک نویز نمایی کاهشی کوچک به ورودی های کنترلی اضافه می شود تا شرط تحریک پایا بودن تضمین شود.

الگوریتم بهینه معرفی شده برای حل مسئله اجماع، به بازی گرافی دیفرانسیلی غیرخطی مذکور اعمال شده است. شکل های ۲ و۳ همگرایی وزن های شبکه های عصبی نقاد همه عامل ها را نشان می دهد و شکل های ۴ و ۵ همگرایی وزن های شبکه های عصبی کنترلر برای عامل ها را با استفاده از الگوریتم پیشنهاد شده نشان می دهد.

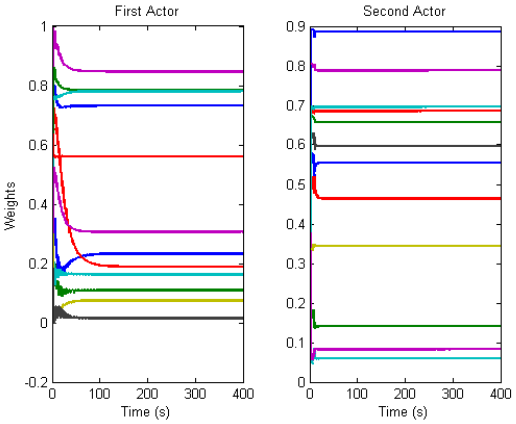


شکل ۲- همگرایی وزن های شبکه های عصبی نقاد عامل های ۱ و۲

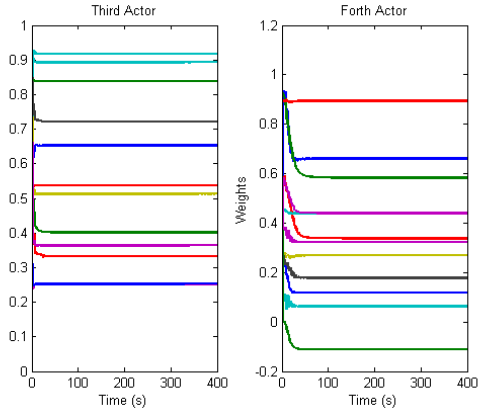


شکل ۳- همگرایی وزن های شبکه های عصبی نقاد عامل های ۳ و ۴

[1] Sequential

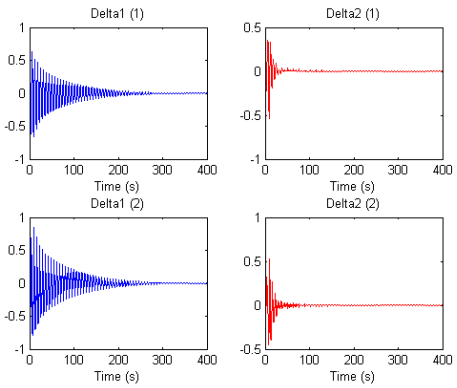


شکل ۴- همگرایی وزن های شبکه های عصبی کنترلر عامل های ۱ و ۲

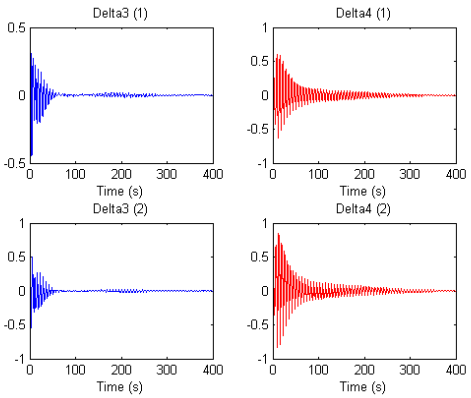


شکل ۵- همگرایی وزن های شبکه های عصبی کنترلر عامل های ۳ و ۴

شکل های ۶ و ۷ خطای ردیابی محلی برای عامل ها را نشان می دهند که تقریبا به صفر همگرا شده اند. با توجه به همگرایی خطا به صفر در شکل های مذکور، همزمان سازی همه عامل ها به رهبر نشان داده شده است.

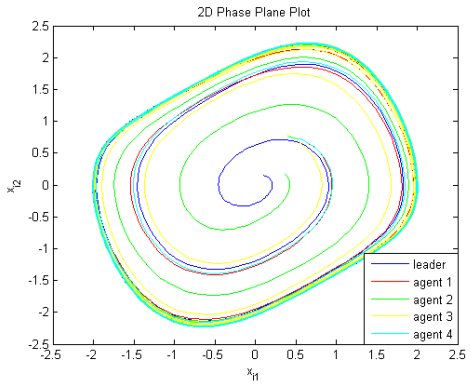


شکل ۶- خطای ردیابی عامل های ۱ و ۲



شکل ۷- خطای ردیابی عامل های ۳ و ۴

همزمان سازی و همگرایی حالت همه عامل ها به عامل رهبر در شکل ۸ (صفحه فاز ۲-بعدی) نشان داده شده اند.



شکل ۸: صفحه فاز ۲-بعدی حالت عامل ها

نتایج نشان می دهد که الگوریتم کنترل بهینه توزیع شده پیشنهادی برای بازی های گرافی دیفرانسیلی غیر خطی زمان پیوسته به حل تقریبی بهینه همگرا شده است.

۷- نتیجه‌گیری

این مقاله توانایی کنترل بهینه تطبیقی، بازی های دیفرانسیلی و سیستم های چند عاملی غیرخطی را به کار گرفته تا مسئله بازی های گرافی دیفرانسیلی غیرخطی چند عاملی پیرو-رهبر را معرفی نماید. یک الگوریتم کنترل بهینه تطبیقی توزیع شده، بر اساس تکنیک تکرار سیاست در یادگیری تقویتی برای حل بازی های گرافی دیفرانسیلی غیرخطی به صورت برخط ارایه شده است. هر عامل، از شبکه های عصبی نقاد و کنترلر به ترتیب برای یادگیری برخط ارزش بهینه و سیاست کنترلی بهینه استفاده می کند. پایداری سیستم حلقه بسته و کرانداری شبکه های عصبی نقاد و کنترلر بر اساس تکنیک لیاپانوف نشان داده شده است و موثر بودن روش پیشنهادی از طریق ارایه نتایج شبیه سازی بررسی گردیده است.

مراجع

[1] Hong Y., Hu J., Gao L., 2006, "Tracking control for multi-agent consensus with an active leader and variable topology" Automatica, 42 (7), pp. 1177–1182.

[2] Ren W., Moore K., Chen Y., 2007, "High-order and model reference consensus algorithms in cooperative control of multivehicle systems," J. Dynam. Syst., Meas., Control, 129(5), pp. 678–688.

[3] Wang X., Chen G., 2002, "Pinning control of scale-free dynamical networks," Physica A, 310(3-4), pp. 521–531.

[4] Tijs S., Introduction to Game Theory, India: Hindustan Book Agency, 2003.

[5] Isaacs R., Differential Games, New York, Wiley, 1965.

[6] Başar T., Olsder Geert J., Dynamic Non-cooperative Game Theory, 2nd edition, Classics in Applied Mathematics, SIAM: Philadelphia, 1999.

[7] Freiling G., Jank G., Abou-Kandil H., 2002, "On global existence of Solutions to Coupled Matrix Riccati equations in closed loop Nash Games", IEEE Transactions on Automatic Control, 41(2), pp. 264- 269.

[8] Gajic Z., Li T., 1988, "Simulation results for two new algorithms for solving coupled algebraic Riccati equations" Third Int. Symp. On Differential Games. Sophia, Antipolis, France.

[9] Wang J., Xin M., 2012, "Distributed optimal cooperative tracking control of multiple autonomous robots", Robotics and Autonomous systems, 60, 572-583.

[10] Lygeros J., Godbole D. N., Sastry S., 1996, "Multi-agent hybrid system design using game theory and optimal control", Proceeding of the 35th conference on decision and control, Kobe, Japan.

[11] Mao D., He Y., Ye X., Yu M., 2011, "Inverse optimal stabilization of cooperative control in networked multi-agent systems", Control and Decision Conference (CCDC), 1031 - 1037.

[12] Semsar-Kazerooni E., Khorasani K., 2009, "Multi-agent team cooperation: A game theory approach", Automatica, 45, 2205-2213.

[13] Shi G., Johansson K. H., Hong Y., 2011, "Multi-agent systems reaching optimal consensus with directed communication graphs", Proceedings of the American Control Conference.

[14] Lewis F. L., Vrabie D. L. and Syrmos V. L., 2012, Reinforcement Learning and Optimal Adaptive Control, in Optimal Control, Third Edition, John Wiley & Sons, Inc., Hoboken, NJ, USA.

[15] Vamvoudakis K., Lewis F.L., 2011, "Multi-player non-zero-sum games: online adaptive learning solution of coupled Hamilton-Jacobi equations", Automatica,47, 1556-1559.

[16] Zhang H., Cui L., Luo Y., 2013, "Near-optimal control for nonzero-sum differential games of continuous-time nonlinear systems using single-network ADP", IEEE Trans. Cybern., 43, 206-216.

[17] Vrabie D., Lewis F. L., 2010, "Integral Reinforcement Learning for Online Computation of Feedback Nash Strategies of Nonzero-Sum Differential Games", 49th IEEE Conference on Decision and Control, December 15-17, Hilton Atlanta Hotel, Atlanta, GA, USA.

[18] Vamvoudakis K. G., Mikulski D. G., Hudas G. R., Lewis F. L., Gu E. Y., 2010, "Distributed games for multi agent systems: games on communication graphs", 27th Army Science Conference Orlando, FL.

[19] Vamvoudakis K. G., Lewis F. L., 2011, "Multi-agent Differential Graphical Games", Proceedings of the 30th Chinese Control Conference, July 22-24.

[20] Vamvoudakis K. G., Lewis F. L., Hudas, G. R., 2012, "Multi-agent differential graphical games: Online adaptive learning solution for synchronization with optimality", Automatica, 48, 1598–1611.

[21] Abouheaf M., "Optimization and reinforcement learning techniques in multi agent graphical games and economic dispatch", Ph. D. thesis, The University of Texas at Arlington, 2012.

[22] Vrabie D., Lewis F. L., 2010, "Integral Reinforcement Learning for Online Computation of Feedback Nash Strategies of Nonzero-Sum Differential Games", 49th IEEE Conference on Decision and Control, December 15-17, Hilton Atlanta Hotel, Atlanta, GA, USA.

[23] Khoo S., Xie L., Man Z., 2009, "Robust Finite-Time Consensus Tracking Algorithm for Multi robot Systems," IEEE Transactions on Mechatronics, 14, pp. 219-228.

[24] Brewer J., 1978, "Kronecker products and matrix calculus in system theory", IEEE Transactions Circuits and Systems, 25(9), pp. 772-781.

[25] Abu-Khalaf M., Lewis F. L., 2005, "Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach", Automatica, 41, 779–791.

[26] Adams R. A., Fournier J. , Sobolev spaces, New York: Academic Press, 2003.

[27] Finlayson B. A., The method of weighted residuals and variational principles, New York: Academic Press, 1990.

[28] Hornik K., Stinchcombe M., White H., 1990, "Universal approximation of an unknown mapping and its derivatives using multi layer feedforward networks", Neural Networks, 3(5), p.551–560.

[29] Sutton R. S., Barto A. G., Reinforcement Learning–An Introduction. Massachusetts: Cambridge, MIT Press, 1998.

[30] Vamvoudakis K. G., Lewis F. L., 2010, "Online actor-critic algorithm to solve the continuous infinite-time horizon optimal control problem", Automatica, 46, p.878–888.

[31] Ioannou P., Sun J., Robust Adaptive Control, Prentice Hall, New Jersey, 1996.

[32] Khalil H. K., Nonlinear systems, Prentice-Hall, 1996.