مجله کنترل
ISSN (print) 2008-8345
ISSN (online) 2538-3752
مقاله پژوهشی
جلد۱۷، شماره ۴، زمستان ۱۴۰۲، صفحه ۸۷-۷۵

کنترل بهینه تطبیقی برخط سیستم‌های دوخطی زمان پیوسته با دینامیک ناشناخته

سیده نفیسه منوچهری رهبر [۱]، ناصر پریز [۲]، محمد رضا رمضانی آل [۳]، عقیله حیدری [۴]

[۱] دانشجوی دکتری، گروه ریاضی ، دانشگاه پیام نور، ص.پ. ۱۹۳۹۵-۴۶۹۷، تهران ، ایران sn.manoochehri@student.pnu.ac.ir

[۲] استاد، گروه مهندسی برق، دانشکده فنی و مهندسی، دانشگاه فردوسی مشهد،مشهد، ایران n-pariz@um.ac.ir

[۳] استادیار، گروه مهندسی برق، دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی قوچان،قوچان، ایران m-ramezani@qiet.ac.ir

[۴] استاد، گروه ریاضی، دانشگاه پیام نور، ص.پ. ۱۹۳۹۵-۴۶۹۷، تهران ، ایران. A_heidari@pnu.ac.ir

دریافت: ۱۴۰۲/۰۷/۱۶ ویرایش: ۱۴۰۲/۱۰/۱۵ پذیرش: ۱۴۰۲/۱۱/۱۲

چکیده: طراحی کنترل‌کننده‌ی بهینه برای سیستم‌های دوخطی زمان پیوسته با معلوم بودن دینامیک سیستم طبق اصل بهینگی بلمن پیچیدگی محاسباتی بالایی دارد و عموماً از روش‌های تقریبی وابسته به دانستن دینامیک سیستم برای طراحی کنترل‌کننده استفاده می شود. هنگامی‌که دینامیک سیستم نامعلوم است این مسئله بسیار پیچیده‌تر می‌شود. اولین چیزی که برای حل این مشکل به نظر می‌رسد شناسایی سیستم دوخطی به کمک روش‌های شناسایی سیستم است. همان‌طور که می‌دانیم روش‌های شناسایی مدلی خطی شده بر اساس داده‌های ورودی و خروجی سیستم در اختیار طراح قرار می‌دهد تا به سراغ طراحی کنترل‌کننده برود. در این مقاله با استفاده از رویه‌ای برخط و تطبیقی، یک روش تکراری جدید به‌منظور طراحی کنترل‌کننده بهینه برای یک سیستم دوخطی که دینامیک آن نامعلوم است پیشنهاد می‌گردد. در روش تکرای پیشنهادی و به صورتی تطبیقی، به‌جای دانستن دینامیک سیستم دوخطی با استفاده از اطلاعات برخط ورودی و اندازه‌گیری حالت‌ها، کنترل‌کننده‌ی بهینه طراحی می‌گردد. همچنین با اعمال نویز به‌منزله ورودی به سیستم در یک بازه‌ی زمانی خاص، نیاز به اندازه‌گیری مجدد حالت‌ها برای تکرارهای بعدی برطرف می‌گردد. همگرایی روش تکراری تطبیقی به کنترل‌کننده بهینه به‌صورت قضیه ارائه و اثبات شده است.

کلمات کلیدی: کنترل بهینه، سیستم‌های دوخطی، دینامیک ناشناخته، تطبیقی، سیاست تکرار.

An online policy iteration for adaptive optimal control of unknown bilinear systems

Seyyede Nafiseh Manoochehri rahbar, Naser Pariz, Mohammad Reza Ramezani-al, Aghileh Heydari

Abstract: Bellman's optimality principle states that designing an optimal controller for continuous-time bilinear systems with known system dynamics has a high computational complexity. As a result, controller design typically uses approximation techniques that depend on system dynamics knowledge. This problem will become more challenging when the system dynamics are unknown. Identifying the bilinear system dynamics through identification techniques is the first step toward overcoming this. It is well known that the identification methods give the designer a linear model to use in the controller design, based on the input and output data of the system. This paper proposes a new iterative method to design an optimal controller for a bilinear system whose dynamics are unknown, using an online adaptive policy iteration. In the proposed iterative method, instead of knowing the dynamics of the bilinear system, the optimal controller is designed by using the online input information and measurement of states. Also, by applying noise as an input for the system in a certain time interval, the need to measure the states for the next iterations is eliminated. The convergence of the adaptive iterative process to the optimal controller has been presented and proved in a theorem.

Keywords: Optimal control, Bilinear systems, Unknown dynamics, Adaptive policy iteration (PI)

[DOI: 10.61186/joc.17.4.75]

۱- مقدمه

به‌عنوان یک دسته خاص از سیستم‌های غیرخطی، سیستم‌های دوخطی از اوایل دهه ۱۹۶۰ به‌عنوان واسطی بین سیستم‌های خطی و سیستم‌های غیرخطی در نظر گرفته شده‌اند. اهمیت این سیستم‌ها در این واقعیت نهفته است که علاوه براینکه بسیاری از سیستم‌ها که دارای مدل طبیعی هستند دوخطی‌اند، بسیاری از فرآیندهای مهم نه‌تنها در مهندسی [۲،۱]، بلکه در زیست‌شناسی [۴،۳]، شیمی [۶،۵]، اجتماعی و اقتصادی [۷] با سیستم‌های دوخطی مدل می‌شوند. همچنین اکثر سیستم‌های غیرخطی را نیز می‌توان با یک سیستم دوخطی تقریب زدکه نسبت به‌تقریب خطی سیستم‌ها دقت بهتری دارند [۸]. بنابراین این حوزه یک زمینه جالب و کاربردی برای مطالعات و تحقیقات کنترلی است.

دستیابی به بهترین عملکرد و رفتار در همه فرآیندهای دینامیکی کنترل‌پذیر، هدف اصلی مهندسان و طراحان سیستم‌های کنترل است. طراحی کنترل‌کننده‌ی بهینه برای سیستم‌ها وابسته به جواب مثبت تعریف‌شده‌ی معادله‌ی همیلتونی-ژاکوبی-بلمن[1] HJB است [۹]. هنگامی‌که سیستم خطی است، حل معادله HJB به حل معادله ریکاتی[2] ARE کاهش می‌یابد ولی برای سیستم‌های دوخطی به معادله ریکاتی وابسته به حالت[3] SDRE تبدیل می‌شود که حل تحلیلی ندارد و عموماً از روش‌های برنامه‌ریزی پویا[4] DP [۱۰] برای حل آن استفاده می‌شود. لذا مراجع [۱۲،۱۱] با تولید دنباله‌ای از مسائل خطی، رساله [۱۳] با جایگزینی بسط سری توانی به‌جای ماتریس‌های وابسته به حالت، مرجع [۱۴] یک روش کنترل بهینه چندجمله ای بر اساس مدل سیستم چند جمله ای ، مرجع [۱۵] به کمک توابع بلاک پالس[5] به‌عنوان یک‌پایه متعامد برای مسئله کمینه زمان، مرجع [۴] به کمک روش عددی تفاضل متناهی و مرجع [۱۶] با خطی سازی بازخورد و تبدیل کردن سیستم به مدل شبه خطی، کنترل بهینه سیستم دوخطی را به‌طور وابسته به دینامیک سیستم به روش حل تقریبی معادله SDRE طراحی می‌کنند.

این روش‌ها برون‌خط هستند و نسبت به تغییرات در سیستم حساس نیستند لذا نمی‌توانند برای کنترل برخط استفاده شوند. همچنین در مواجهه با سیستم‌های پیچیده و با ابعاد بالا با مشکل مواجه‌اند. برای حل این مشکل اخیراً از یادگیری تقویتی[6] RL به دلیل کاربردهای موفق آن بیشتر مورد توجه محققین قرار گرفته است [۱۷-۱۹]. مشابه ساختار منطقی کنترل بهینه، در RL کنترل‌کننده به‌طور پیوسته با محیط ناشناخته تعامل می‌کند تا راه‌حل بهینه را با اندازه‌گیری‌های برخط حالت‌ها به دست آورد. این ویژگی در مواقعی که قسمتی یا همه‌ی دینامیک سیستم برای طراح مجهول است و یا مواقعی که احتمال بروز تغییراتی در روند فرآیند کنترل وجود دارد بسیار سودمند است. برای این‌گونه طراحی‌های برخط، مرجع [۲۰] با اعمال ورودی پایدارساز به سیستم خطی، حالت و مشتق حالت را در یک روش تکراری جدید جایگزین ماتریس حالت در معادله ریکاتی می‌نماید. مراجع [۲۲،۲۱] با الهام از آن به کمک نظریه پایداری لیاپانوف و در نظر گرفتن تابع لیاپانوف به‌عنوان تابع هزینه سیستم، رابطه تطبیقی تکراری و برخطی تولید می‌کنند که به‌اندازه گیری مشتق حالت در هر مرحله تکرار روش نیازی ندارد. به‌عنوان کاربرد عملی این روش در سیستم تنظیم‌کننده خودکار ولتاژ[7]، مرجع [۲۳] از این روش جهت کمینه کردن تابع هزینه افق بی‌نهایت تنظیم‌کننده مربعی استفاده می کند. برای سیستم‌های غیرخطی تبار[8]، به طور مشابه مرجع [۲۴] با در نظر گرفتن تابع لیاپانوف به‌عنوان تابع هزینه سیستم و ایجاد رابطه تطبیقی تکراری برخط، نیاز به دانستن دینامیک تابع غیرخطی حالت سیستم تبار برای محاسبه کنترل‌کننده بهینه را با حالت‌های اندازه‌گیری شده از سیستم در گام‌های تکرار روش پیشنهادی خود برطرف می‌نماید و از شبکه عصبی برای حل معادلات در هرگام استفاده می‌نماید. درحالی‌که مرجع [۲۵] برای دسته‌ای از سیستم های تبار، نخست با خطی سازی به کمک شبکه عصبی چندلایه همان روش برخط تطبیقی [۲۲] را برای این سیستم‌ها و با سرعت همگرایی بالاتر نسبت به [۲۴] معرفی می‌نماید.

برای حذف ماتریس ورودی علاوه بر ماتریس حالت در مورد سیستم های خطی، مرجع [۲۶] ضمن در نظر گرفتن تابع لیاپانوف به‌عنوان تابع هزینه سیستم و ایجاد رابطه تکراری تطبیقی برخط، اعمال نویز شناسایی به یک ورودی پایدارساز اولیه را پیشنهاد می‌دهد. سرعت همگرایی این روش نیز در مرجع [۲۷] به کمک تنظیم‌کننده با درجه مشخصی از پایداری افزایش یافته است. مرجع [۲۸] به کمک بازخورد خروجی بر نامعینی‌های سیستم خطی به‌صورت برخط غلبه می‌کند و مرجع [۲۹] برای کمینه کردن تابع هزینه مربعی شامل بخش غیرخطی ضرب حالت و ورودی (تابع هزینه توسعه داده شده) در سیستم خطی، ساخت معادله ریکاتی تعمیم‌یافته و جایگذاری حالت و ورودی را پیشنهاد می‌دهد. همچنین با تکیه بر روش‌های حذف ماتریس حالت از محاسبات کنترل‌کننده بهینه، مراجع [۳۲-۳۰] با اضافه کردن دینامیک پیش جبران ساز[9] به سیستم و تولید سیستم افزوده و با تقریب به کمک شبکه های عصبی توانستند نیاز به دانستن اطلاعات ماتریس ورودی را محاسبات خود مرتفع سازند.

اخیراً و برای حالت افق زمانی محدود در [۳۳] ، کنترل‌کننده بهینه موضعی سیستم‌های دوخطی زمان گسسته با دینامیک ناشناخته بر مبنای روش‌های داده-محور[10] و استفاده از روش تکراری و به کمک اصول بهینه‌سازی محدب-مقعر ارائه شده است. همچنین کنترل‌کننده‌های بازخورد حالت بر اساس توابع بلاک-پالس برای مسئله ردیابی این سیستم‌ها با دینامیک ناشناخته به همراه تأخیر زمانی متعدد در رفتار سیستم و تحت ورودی کنترل مقید در [۳۴] بررسی شده است. در [۳۵] از

[1] Hamiltonian Jacobi Bellman (HJB)
[2] Algebraic Rickety equation (ARE)
[3] State-dependent Riccati equation (SDRE)
[4] Dynamic Programing (DP)
[5] Block-pulse functions
[6] Reinforcement learning (RL)
[7] Automatic voltage regulator (AVR)
[8] Affine
[9] Pre compensator
[10] Data-driven

[DOI: 10.61186/joc.17.4.75]

تبدیل‌های کوپمن[1] و توابع ویژه کوپمن[2] برای تبدیل یک سیستم غیرخطی با دینامیک ناشناخته به یک سیستم دوخطی جهت حل یک مسئله بهینه‌سازی مربعی در افق زمانی محدود استفاده شده است. برای طراحی کنترل‌کننده بهینه افق نامتناهی سیستم‌های دوخطی زمان پیوسته، مرجع [۳۶] و با الهام از روش [۲۲] به‌صورت تطبیقی و برخط توانسته حالت‌ها را جایگزین ماتریس حالت نماید. از طرفی برقراری پایداری سیستم‌های دوخطی به‌راحتی پایداری سیستم‌های خطی نیست و عموماً طراحی کنترل‌کننده‌ی پایدارساز بر اساس نظریه لیاپانوف نیازمند تشکیل تابع لیاپانوف مربعی و بررسی شرط‌های آن با کنترل‌کننده مدنظر است. کارهای انجام‌شده به شناخت کامل دینامیک سیستم تکیه دارد، به‌طوری‌که با معرفی دسته‌ای از توابع لیاپانوف مرجع [۳۷] با یک کنترل بنگ بنگ[3] خاص شرایط پایدار مجانبی سراسری سیستم دوخطی را بررسی می کند و مرجع [۳۸] با یک بازخورد خطی حالت شرایط پایدار مجانبی محلی و مرجع [۳۹] نیز با یک بازخورد غیرخطی شرایط پایداری سراسری این سیستم ها را طبق نظریه لیاپانوف برقرار می‌سازند. در [۴۴] یک روش یادگیری بر خط برای حل مسئله H_∞ سیستم های زمان پیوسته با ورودی های کران دار ارائه شده است به طوری که از روش های نوین مبتنی بر نظریه بازی و شبکه های عصبی برای حل معادله همیلتون-ژاکوبی-ایزاکس[4] HJI استفاده می نماید.

هدف اصلی این مقاله این است که درصورتی‌که ماتریس حالت و ماتریس ورودی نامعلوم باشند کنترل بهینه تنظیم‌کننده مربعی افق بی‌نهایت را بصورت برخط برای چنین سیستم‌های دوخطی ناشناخته ای که پاسخ ورودی صفر آن پایدار است به دست آوریم. درواقع قصد داریم به کمک بازخورد خطی حالت، یک روش تکراری تطبیقی برخط جدید ارائه دهیم که پاسخ SDRE منطبق با اصل بهینگی بلمن را تقریب بزند. در اینجا ما با ذخیره ورودی‌های اعمال‌شده به سیستم تنها در یک بازه‌ی زمانی خاص، رفتار حالات سیستم را اندازه‌گیری می‌کنیم و این اطلاعات را جایگزین ماتریس حالت و ماتریس ورودی در روند محاسبات بهینه‌سازی می‌نماییم.

در ادامه، مقاله به‌صورت زیر سامان‌دهی شده است. بخش دو ضمن معرفی سیستم و فرض‌های مسئله به بیان کنترل‌کننده‌ی بهینه‌ی سیستم دوخطی طبق اصل بهینگی بلمن در حالت معلوم بودن دینامیک سیستم می‌پردازد. در این بخش ضمن اشاره به تقریبی بودن کنترل‌کننده‌های بهینه برای این سیستم‌ها حتی در حالت معلوم بودن دینامیک، به‌مرور سریع روش طراحی کنترل بهینه تطبیقی برخط[۳۶] و بیان کاستی‌های آن می‌پردازد. در بخش سه، کنترل‌کننده بهینه دینامیک خطی سیستم دوخطی را تقریب موضعی کنترل‌کننده بهینه این سیستم در نظر می‌گیرد و پایداری سیستم با بازخوردهای خطی پایدارساز مربوط به دینامیک خطی سیستم بررسی می‌گردد. همچنین به کمک تابع لیاپانوف پایداری سیستم با بازخوردهای خطی پایدارساز مطرح‌شده در این بخش، یک روش تکراری تطبیقی برخط جدید بر پایه کنترل بازخورد خطی حالت ارائه می‌شود. روش جدید به کنترل‌کننده بهینه تقریبی مفروض بدون نیاز به دانستن دینامیک خطی سیستم همگراست که اثبات آن در غالب قضیه مطرح می‌شود. پیاده‌سازی برخط روش پیشنهادی به کمک ضرب‌های کرونکر[5] [۴۰] ارائه می‌شود. در بخش چهار برای نشان دادن کارایی روش پیشنهادی، مسئله طراحی کنترل‌کننده بهینه یک راکتور مخزن هم زن‌دار با واکنش گرمازا و طراحی کنترل‌کننده بهینه ماشین کاغذسازی و یک مسئله‌ی طراحی کنترل‌کننده‌ی بهینه‌ی پرتکرار برای یک سیستم دوخطی دو ورودی با روش جدید و روش مورد اشاره در بخش دو ارائه و شبیه‌سازی می‌شود. در انتها نتایج پژوهش ارائه می‌گردد.

نمادگذاری: در طول این مقاله نماد $\|.\|$ نرم اقلیدسی بردارها یا ماتریس‌ها را نشان می‌دهد، نماد $\otimes$ برای نمایش ضرب کرونر و $vec(A)$ برای نمایش یک بردار mn-عضوی حاصل از پشت‌هم قرار دادن ستون‌های ماتریس $A\epsilon\mathbb{R}^{m\times n}$ به کار می‌روند. یک قانون کنترلی نیز یک سیاست نامیده می‌شود. $K_0 \in R^{m\times n}$ یک ماتریس بهره‌ی پایدارساز سیستم $\dot{x} = Ax + Bu + \{xN\}u$ است هرگاه $A - BK_0$ هورویتس باشد.

۲- بیان مسئله

سیستم دوخطی (۱) و تابعی هزینه (۲) را در نظر بگیرید

$$\dot{x} = Ax + Bu + \{xN\}u \qquad (۱)$$

$$J_\infty = \int_0^{+\infty} [x^T Qx + u^T Ru]dt \qquad (۲)$$

که درآن زوج (A,B) کنترل‌پذیر کامل، $\{xN\} = \sum_{i=1}^{n} x_i N_i$، $x \in \mathbb{R}^n$، $u \in \mathbb{R}^m$ و $A \in \mathbb{R}^{n\times n}$ ماتریس حالت و $B \in \mathbb{R}^{m\times n}$ ماتریس ورودی، برای $k = 1,2,\dots,n$ ، $N_k \in \mathbb{R}^{n\times m}$ ماتریس ورودی - حالت نامیده می شوند. همچنین $Q \in \mathbb{R}^{n\times n}$ ماتریس مثبت نیمه معین، $R \in \mathbb{R}^{m\times m}$ ماتریس مثبت معین و زوج $(A, Q^{\frac{1}{2}})$ آشکارسازپذیر است. فرض می‌شود که تمام حالت‌های سیستم در دسترس هستند و مقادیر مشخصه ماتریس A منفی بوده و پاسخ ورودی صفر سیستم (۱) پایدار است. از طرفی می دانیم با استفاده از روش ژاکوبی خطی[6]، می توان دینامیک سیستم(۱) راحول نقطع تعادل به طور تقریبی با دینامیک بخش خطی آن معادل کرد.

$$\dot{x} = Ax + Bu \qquad (۳)$$

[1] Koopman canonical transform
[2] Koopman eigen functions
[3] Bang-Bang control
[4] Hamilton–Jacobi–Isaacs (HJI)
[5] Kronecker product
[6] Jacobian linearization

[DOI: 10.61186/joc.17.4.75]

هدف، طراحی برخط یک قانون کنترل بهینه و تطبیقی برای سیستم (۱) است که در آن ماتریس‌های A و B نامشخص و بقیه پارامترها معلوم هستند.

در مرجع [۱۳] مسئله کنترل بهینه سیستم دوخطی (۱) با معلوم بودن ماتریس‌های ضرایب حالت و ورودی طبق اصل بهینگی بلمن حل شده است که در لم ۱ بیان می‌شود.

لم ۱ ([۱۳]): قانون کنترل (۴) قانون کنترل بهینه سیستم دوخطی (۱) است به طوریکه $P(x)$ تابع ماتریسی متقارن و جواب منحصربه‌فرد معادله جبری ریکاتی وابسته به حالت (۵) باشد.

$$u = -R^{-1}(B + \{xN\})^T P(x)x \quad (۴)$$

$$Q + P(x)A + A^T P(x) - P(x)(B + xN)R^{-1}(B + xN)^T P(x) = 0, \quad (۵)$$

تبصره ۱: معادله ریکاتی وابسته به حالت (۵) راه‌حل تحلیلی ندارد و روش‌های ارائه‌شده در مراجع [۱۱-۱۵] بر پایه حل عددی معادله (۵) استوار است. همچنین روش‌های این مراجع برون‌خط هستند و به دانستن کامل ماتریس‌های ضرایب سیستم (۱) نیازمندند.

در [۳۶] یک روش تکراری تطبیقی و برخط برای تقریب جواب معادله (۵) با فرض این‌که ماتریس حالت در (۱) معلوم نیست و بقیه‌ی پارامترها معلوم هستند ارائه و اثبات شده است که روش به ماتریس P مثبت معین و ثابت همگرا می گردد. این روش با اضافه کردن شرط توقف به‌صورت زیر و در ۴ گام آمده است.

روش۱ [۳۶]:

گام۱- نخست مقدار زمانی مناسب δt برای گام شبیه‌سازی و ε شرط توقف در نظر بگیرید. قرار دهید $k = 0$ ، ماتریس متقارن و مثبت معین P_0 را طوری انتخاب کنید که $u_0 = -R^{-1}B^T P_0 x$ یک قانون کنترل پایدارساز برای سیستم (۶) باشد.

$$\dot{x} = Ax + Bu_0 \quad (۶)$$

گام۲- با سیاست کنترلی u_k، ماتریس مثبت معین و متقارن P_{k+1} را با استفاده از رابطه (۷) تعیین کنید.

$$x(t)^T P_{k+1} x(t) = x(t+\delta t)^T P_{k+1} x(t+\delta t) + \int_t^{t+\delta t} x(\tau)^T Q x(\tau) + u_k^T(\tau) R u_k(\tau) d\tau \quad (۷)$$

گام ۳- قانون کنترل را با استفاده از رابطه‌ی (۸) به‌روز نمایید.

$$u_{k+1} = -R^{-1}(B + \{xN\})^T P_{k+1} x \quad (۸)$$

گام٤- قرار دهید $k = k + 1$ و $t = t + \delta t$ ، اگر شرط توقف $\|P_k - P_{k-1}\| < \varepsilon$ برقرار باشد قرار دهید $P_{k+1} = P_k$ به گام ۳ بروید ، در غیر این صورت به گام ۲ بروید.

در مرجع [۳۶] از روش کمترین مجموع مربعات خطای دسته ای [1] جهت محاسبه درایه های مجهول ماتریس P که به تعداد $D = n(n+1)/2$ است استفاده شده است که نیازمند اعمال قانون کنترل و معکوس سازی یک ماتریس مربعی با بعد D از حالت‌های سیستم در هر گام تکرار است. همچنین جهت اعمال قانون کنترل (۸) نیاز به معلوم بودن ماتریس B است. درصورتی‌که ماتریسB مجهول باشد روش ارائه‌شده کارایی ندارد لذا در بخش بعدی به ارائه روشی برای حل مسئله در حالتی که ماتریس B هم مجهول است پرداخته می‌شود.

۳- طراحی کنترل بهینه تطبیقی برخط برای سیستم‌های دوخطی با دینامیک ناشناخته

در این بخش سعی داریم با اعمال یک ورودی بازخورد خطی حالت به‌صورت (۹) با ماتریس بهره‌ی (۱۰) یک روش تکراری تطبیقی و برخط جدید جهت کنترل بهینه تطبیقی سیستم (۱) ارائه دهیم.

$$u = -Kx \quad (۹)$$

$$K = R^{-1}B^T P \quad (۱۰)$$

در ادامه برای ماتریس $\boldsymbol{B}$ جایگزینی بر اساس اطلاعات ارائه خواهد شد. هنگامی‌که $\mathbf{A}$ و $\boldsymbol{B}$ معلوم باشد ماتریس $\boldsymbol{P}$ از معادله ریکاتی جبری (۱۱) به دست می‌آید.

$$\mathrm{A}^{\mathrm{T}}P + PA + Q - PBR^{-1}B^T P = 0 \quad (۱۱)$$

لازم به ذکر است که طبق اصل بهینگی بلمن پاسخ معادله (۱۱) به همراه روابط (۹-۱۰) کنترل‌کننده‌ی بهینه برای سیستم (۳) است. لذا این کنترل کننده در صورت وجود به عنوان کنترل کننده بهینه موضعی سیستم (۱) در نظر گرفته می شود. با توجه به غیرخطی بودن معادله (۱۱) نسبت به P ، عموماً حل آن برای سیستم‌های با ابعاد بالا با مشکل مواجه است لذا قضیه زیر مطرح می‌شود.

قضیه ۱ ([۳۸]): اگر $K_0 \in R^{m\times n}$ یک ماتریس بهره‌ی پایدارساز اولیه برای سیستم (۳) با فرض معلوم بودن A و B باشد، آنگاه ماتریس P_k جواب متقارن و مثبت معین معادله‌ی لیاپانوف (۱۲) است به‌طوری‌که با تعریف رابطه بازگشتی (۱۳) برای $k = 1,2,\ldots$ شرایط زیر برقرار است :

۱) ماتریس $A - BK_{k-1}$ هورویتس است،

۲) $P^* \le P_{k+1} \le P_k \le \cdots$ ،

۳) $\lim_{k\to\infty} P_k = P^*$ و $\lim_{k\to\infty} K_k = K^*$ که در آن P^* جواب (۱۱) و K^* بهره متناظر آن با رابطه (۱۰) است.

$$(\mathrm{A} - \mathrm{BK_k})^{\mathrm{T}}\mathrm{P_k} + \mathrm{P_k}(\mathrm{A} - \mathrm{BK_k}) + \mathrm{Q} + \mathrm{K_k^T R K_k} = 0 \quad (۱۲)$$

$$\mathrm{K_k} = \mathrm{R^{-1}B^T P_{k-1}} \quad (۱۳)$$

[1] Batch least square error

[DOI: 10.61186/joc.17.4.75]

همانطور که مشاهده می شود روابط (۱۲) و (۱۳) یک روش تکراری برون خط و وابسته به دینامیک سیستم را جهت محاسبه کنترل کننده بهینه موضعی سیستم (۱) ارائه می کند.

در لم۲ نشان می دهیم کنترل کننده های خطی حاصل از این روابط سیستم (۱) را به طور موضعی پایدار می نماید.

لم۲ : اگر $K_0 \in R^{m\times n}$ یک ماتریس بهره‌ی پایدارساز اولیه برای سیستم(۱) با فرض معلوم بودن A و B باشد، آنگاه برای $k = 1,2, \dots$ بازخورد های خطی با ماتریس بهره ی (۱۳)، سیستم (۱) را بطور مجانبی حول نقطه ی کار پایدار می نماید.

اثبات: در سیستم (۱) و بنا بر نظریه لیاپانوف برای سیستم های غیرخطی نامتغیر با زمان اگر (A,B) کنترل پذیر باشد، پایداری مجانبی سیستم حلقه بسته (۱) با قرار دادن مقادیر مشخصه $A-BK$ در نیمه سمت چپ صفحه مختلط تضمین می شود. بر طبق قضیه ۱ برای ماتریس های بهره ی (۱۳) ، ماتریس $A-BK_{k-1}$ هورویتس است . □

جهت محاسبه‌ی کنترل‌کننده‌ی بهینه تطبیقی برای سیستم (۱) با ماتریس حالت و ماتریس ورودی نامعلوم، مطابق با لم ۲ سیستم (۱) توسط کنترل کننده بازخورد خطی با بهره ی (۱۳) پایداراست بنابراین سیستم (۱) به‌صورت (۱۴) بازنویسی می‌شود.

$$\dot{x} = (A-BK_k)x + B(u+K_kx) + \sum_{i=1}^{n} x_i N_i u \qquad (۱۴)$$

حال با انتخاب $V(x) = x^T P_k x$ به‌عنوان تابع لیاپانوف داریم:

$$\begin{aligned}\dot{V}(x) &= \dot{x}^T P_k x + x^T P_k \dot{x}\\ &= x^T[(A-BK_k)^T P_k + P_k(A-BK_k)]x\\ &\quad +2(u+K_kx)^T B^T P_k x + 2x^T P_k \sum_{i=1}^{n} x_i N_i u\\ &= -x^T(Q+K_k^T R K_k)x + 2(u+K_kx)^T R K_{k+1} x\\ &\quad +2x^T P_k \sum_{i=1}^{n} x_i N_i u\end{aligned} \qquad (۱۵)$$

تبصره۲: طبق لم۲ شرایط قضیه ۱ برای سیستم (۱) برقرار است و عبارت‌های $[(A-BK_k)^T P_k + P_k(A-BK_k)]$ و $B^T P_k$ در جمله اول و دوم (۱۵) به ماتریس‌های A و B وابسته هستند که به ترتیب طبق رابطه (۱۲) و (۱۳) با $Q+K_k^T R K_k$ و RK_{k+1} جایگزین شده اند.

با انتگرال‌گیری از طرفین رابطه (۱۵) در بازه زمانی $[t, t+\delta t]$ داریم:

$$\begin{aligned}&x(t+\delta t)^T P_k x(t+\delta t) - x(t)^T P_k x(t) =\\ &-\int_t^{t+\delta t} x^T Q_k x d\tau + 2\int_t^{t+\delta t} (u+K_kx)^T\\ &RK_{k+1} x d\tau + 2\int_t^{t+\delta t} x^T P_k \sum_{i=1}^{n} x_i N_i u d\tau\end{aligned} \qquad (۱۶)$$

تبصره۳: قابل توجه است که (۱۶) یک روش سیاست تکرار جدید را فرموله می کند. به طوریکه با معلوم بودن ماتریس بهره‌ی مرحله اول بدون نیاز به راه‌اندازی مجدد سیستم و نیاز به ماتریس B به کمک جمله‌ی دوم طرف راست آن، ماتریس بهره‌ی مرحله بعد را محاسبه می‌نماید. بنابراین به‌منظور حفظ این جمله با فرض یک K_0 پایدارساز معلوم برای سیستم (۱)، در گام نخست ورودی $u = -K_0x$ همراه با نویز به‌صورت $u = -K_0x + e(t)$ در نظر گرفته می‌شود. درواقع با در نظر گرفتن نویز $e(t)$ در نقش نویز شناسایی به‌عنوان سیگنال ورودی برای یادگیری، نیازی به ماتریس B و راه اندازه مجدد سیستم و اندازه‌گیری مجدد حالت‌ها به‌منظور محاسبه‌ی ماتریس بهره‌ی مرحله بعد نیست.

در ادامه رویکرد جدیدی بر اساس ضرب کرونکر جهت ساده‌سازی حل معادله ماتریسی غیرخطی (۱۶) و محاسبه برخط K_{k+1} و P_k ارائه می‌شود. نمایش عبارت‌های $\bar{x}$ ، $vec(P)$ ، $x^T Q_k x$ ، $(u+K_k)^T R K_{k+1} x$ و $x^T P \sum_{i=1}^{n} x_i N_i u$ به‌صورت ضرب کرونکر با الهام از مرجع [۲۸] در روابط (۱۷-۲۱) آمده است.

$$\bar{x} = (x \otimes x) \epsilon \mathbb{R}^{n^2} \qquad (۱۷)$$

$$vec(P) \epsilon \mathbb{R}^{n^2\times 1} \qquad (۱۸)$$

$$x^T Q_k x = (x^T \otimes x^T) vec(Q_k) \qquad (۱۹)$$

$$\begin{aligned}&(u+K_k)^T R K_{k+1} x =\\ &[(x^T \otimes x^T)(I_n \otimes K_k^T R)\\ &\quad +(x^T \otimes u^T)(I_n \otimes R)] vec(K_{k+1})\end{aligned} \qquad (۲۰)$$

$$\begin{aligned}&x^T P \sum_{i=1}^{n} x_i N_i u =\\ &\left(\left(\sum_{i=1}^{n} x_i N_i u\right)^T \otimes x^T\right) vec(P)\end{aligned} \qquad (۲۱)$$

به‌علاوه برای عدد ثابت L که در ادامه به آن پرداخته می‌شود، ماتریس‌های (۲۲ – ۲۵) تعریف می‌شوند.

$$\begin{aligned}\delta_{xx} = [&\bar{x}(t_1) - \bar{x}(t_0), \bar{x}(t_2) - \bar{x}(t_1), \dots\\ &, \bar{x}(t_L) - \bar{x}(t_{L-1})]^T \in R^{L\times n^2}\end{aligned} \qquad (۲۲)$$

$$\begin{aligned}\Lambda_{xx} = [&\int_{t_0}^{t_1} \left(\left(\sum_{i=1}^{n} x_i N_i u\right)^T \otimes x^T\right) d\tau,\\ &\int_{t_1}^{t_2} x \left(\left(\sum_{i=1}^{n} x_i N_i u\right)^T \otimes x^T\right) d\tau, \dots,\\ &\int_{t_{L-1}}^{t_L} \left(\left(\sum_{i=1}^{n} x_i N_i u\right)^T \otimes x^T\right) d\tau]^T\\ &\in \mathbb{R}^{L\times n^2}\end{aligned} \qquad (۲۳)$$

$$\begin{aligned}I_{xx} = [&\int_{t_0}^{t_1} \bar{x}(\tau) d\tau, \int_{t_1}^{t_2} \bar{x}(\tau) d\tau, \dots\\ &, \int_{t_{L-1}}^{t_L} \bar{x}(\tau)\, d\tau]^T \in \mathbb{R}^{L\times n^2}\end{aligned} \qquad (۲۴)$$

$$\begin{aligned}I_{xu} = [&\int_{t_0}^{t_1} x \otimes u\, d\tau, \int_{t_1}^{t_2} x \otimes u\, d\tau, \dots\\ &, \int_{t_{L-1}}^{t_L} x \otimes u\, d\tau]^T \in \mathbb{R}^{L\times mn}\end{aligned} \qquad (۲۵)$$

[DOI: 10.61186/joc.17.4.75]

در این روابط لحظات زمانی به‌صورت $t_0 < t_1 < \cdots < t_L$ مرتب هستند. به کمک این ماتریس‌ها معادله ماتریسی (۲۶) جایگزین معادله جبری غیرخطی (۱۶) می‌شود.

$$\Theta_k \begin{bmatrix} vec(P_k) \\ vec(K_{k+1}) \end{bmatrix} = \Xi_k \qquad (۲۶)$$

که در آن $\Theta_k \in R^{L\times[n^2+mn]}$ و $\Xi_k \in R^L$ به‌صورت (۲۷) و (۲۸) تعریف می‌شوند:

$$\Theta_k = [\ \delta_{xx} - 2\Lambda_{xx},\ -2I_{xx}(I_n \otimes K_k^T R) - 2I_{xu}(I_n \otimes R)] \qquad (۲۷)$$

$$\Xi_k = -I_{xx} vec(Q_k)$$

برای حل رابطه (۲۶) می‌توان از تخمین کمترین مجموع مربعات خطا استفاده و بردار پارمتر $\begin{bmatrix} vec(P_k) \\ vec(K_{k+1}) \end{bmatrix}$ را محاسبه کرد. اگر $L \geq [n^2 + mn]$ و ماتریس $\Theta_k^T\Theta_k$ معکوس‌پذیر باشد براساس تخمین کمترین مجموع مربعات خطا می توان معادله (۲۶) به‌صورت (۲۸) بازنویسی کرد.

$$\begin{bmatrix} \mathrm{vec(P_k)} \\ \mathrm{vec(K_{k+1})} \end{bmatrix} = (\ \Theta_k^T\Theta_k)^{-1}\Theta_k^T\Xi_k. \qquad (۲۸)$$

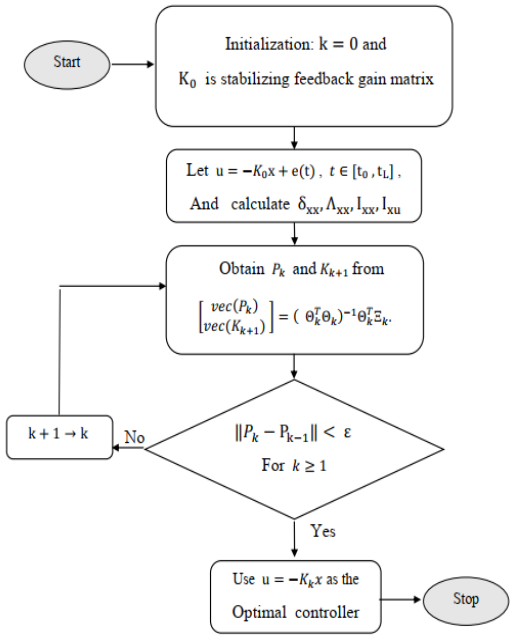
تبصره ٤: انتخاب نویز جهت شناسایی سیستم به‌ویژه برای سیستم‌های با ابعاد بالا، موضوع مهمی است. در [۴۲] انواع مختلفی از نویز معرفی و بکار گرفته‌شده است. با انتخاب نویز شناسایی برای $L \geq [n^2 + mn]$ از کامل بودن رتبه ماتریس Θ_k طبق اصل رتبه کاملی ماتریس با درایه‌های تصادفی اطمینان حاصل می شود [۴۳].

همانطور که مشاهده می شود رابطه (۲۸) برای $k = 0,1,2, \ldots$ یک دنباله از $\{P_k\}_{k=0}^{\infty}$ و $\{K_k\}_{k=0}^{\infty}$ تولید می کند. از این رو همگرا بودن این دنباله به کنترل کننده بهینه موضعی (۹) با بهره ی (۱۰) را در قضیه ۲ بیان و اثبات می کنیم.

قضیه ۲ : اگر $K_0 \in R^{m\times n}$ بهره پایدارساز اولیه برای برای سیستم (۱) باشد دنباله $\{P_i\}_{i=0}^{\infty}$ و $\{K_j\}_{j=0}^{\infty}$ حاصل از رابطه (۲۹) به ترتیب به P^* و K^* پاسخ معادله (۱۱) و بهره (۱۰) همگراست.

اثبات: اگر K_k بهره‌ی پایدارساز سیستم (۱) باشد و P_k پاسخ معادله‌ی (۱۲)، آنگاه بنا به قضیه۱ $K_{k+1} = R^{-1}B^T P_k$ به‌صورت یکتا تعیین می‌شود و در (۱۶) و (۲۸) نیز صدق می‌کند. همچنین اگر $P_q \in R^{n\times n}$ و $K_{q+1} \in R^{m\times n}$ جواب دیگری برای (۲۸) باشد بطوریکه $\Theta_k \begin{bmatrix} vec(P_q) \\ vec(K_{q+1}) \end{bmatrix} = \Xi_k$، چون Θ_k رتبه کامل دارد لذا $vec(P_q) = vec(P_k)$ و $vec(K_{q+1}) = vec(K_{k+1})$. بنابراین سیاست تکرار (۲۹) مطابق سیاست تکرار ارائه‌شده در رابطه (۱۲) به ترتیب به P^* پاسخ معادله (۱۱)و $K^* = R^{-1}B^T P^*$ به بهره (۱۰) همگرا است. در نتیجه $P_\infty = P^*,\ K_\infty = K^*$. □

همانطور که مشاهده شد روش تکراری ایجاد شده با رابطه (۲۸) به کنترل کننده بهینه تقریبی همگرا می شود. شکل ۱ نمودار بلوکی روش جدید پیشنهادی را نشان می دهد.



شکل۱: نمودار بلوکی روش جدید

روش تطبیقی برخط جدید:

روش پیشنهادی به‌طور خلاصه بر اساس استفاده از تخمین حداقل مربعات دسته ای (۲۸) در گام‌های زیر ارائه شده است:

گام۱: قرار دهید $k = 0$ ، بهره‌ی K_0 اولیه طوری در نظر بگیرید که سیستم (۱) پایدار باشد. با نویز مناسب $e(t)$، ورودی $u = -K_0x + e(t)$ را در بازه‌ی زمانی $[t_0, t_L]$ به سیستم اعمال کنید و I_{xu} ، I_{xx} ، Λ_{xx} ، δ_{xx} و Θ_kو Ξ_k محاسبه کنید.

گام۲: K_{k+1} و P_k را از حل رابطه‌ی (۲۸) به دست آورید.

گام۳: قرار دهید $k + 1 \to k$. اگر $k = 1$ باشد ماتریس های Θ_k و Ξ_k را با بهره‌ی گام ۲ بروز نمایید و به گام ۲ بروید. در غیر اینصورت به گام ۴ بروید.

گام٤: اگر $\|P_k - P_{k-1}\| < \varepsilon$ و به گام ۵ بروید. در غیر این صورت Θ_k و Ξ_k را با بهره‌ی گام ۲ بروز نمایید و به گام ۲ بروید.

گام۵: از $u_k = -K_k x$ به‌عنوان سیاست کنترل بهینه تطبیقی تقریبی استفاده نمایید.

٤- پیاده‌سازی کنترل بهینه تطبیقی برخط پیشنهادی

در این بخش جهت بررسی کارایی روش پیشنهادی به شبیه سازی سه سیستم کاربردی پرداخته می شود. در این شبیه‌سازی‌ها ازاطلاعات ماتریس‌های A و B جهت محاسبه کنترل کننده بهینه باروش پیشنهادی

[DOI: 10.61186/joc.17.4.75]

استفاده نشده است و شرط توقف $\|P_k - P_{k-1}\| < \varepsilon,$ با مقدار $(\varepsilon = 0.03)$ انتخاب شده است.

سیستم شبیه سای شده۱ – سیستم دوخطی راکتور تانک هم زن دار با واکنش گرمازا را در نظر بگیرید[۳۶].

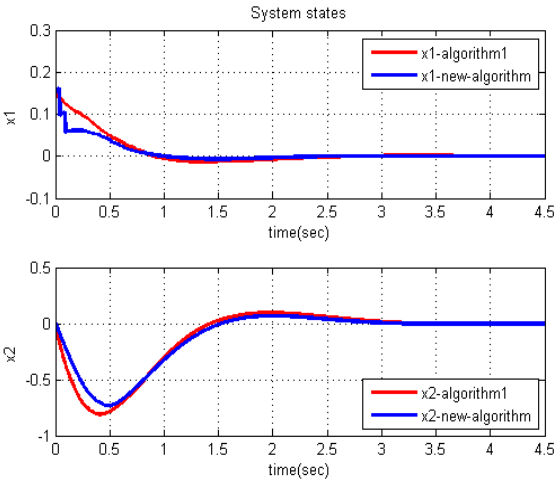
$$\dot{x} = \begin{bmatrix} \frac{13}{6} & \frac{5}{12} \\ -\frac{50}{3} & -\frac{8}{3} \end{bmatrix} x + \begin{bmatrix} -1 \\ 0 \end{bmatrix} x_1 u + \begin{bmatrix} -\frac{1}{8} \\ 0 \end{bmatrix} u$$

$$x_0 = \begin{bmatrix} 0.15 \\ 0 \end{bmatrix} \qquad (۲۹)$$

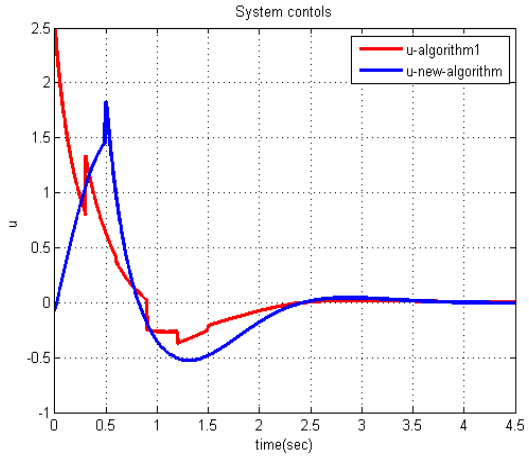
ماتریس‌های وزنی تابع هزینه به‌صورت زیر هستند:

$$Q = \begin{bmatrix} 6.02 & -0.6 \\ -0.6 & 5.11 \end{bmatrix} \qquad R = 1,$$

ازآن‌جا که سیستم حلقه باز پایدار است، $K_0 = 0$ در نظر گرفته شده است. برای اجرای روش جدید، نخست برای مدت ۰.۵ ثانیه نویز شناسایی $e(t) = 0.8 \sum_{i=1}^{100} \sin(w_i t)$ به سیستم اعمال می شود به‌طوری که w_i ها برای $i = 1,\ldots,100$ به‌صورت تصادفی در بازه $[-2,2]$ انتخاب و حالت‌های (۲۹) در هر ۰.۰۱ثانیه اندازه‌گیری شدند. همچنین روش۱ با گام های زمانی مشابه و بروزرسانی ماتریس P هر ۰.۵ ثانیه شبیه سازی شده است. نمودارهای رفتار مسیر متغیرهای حالت و سیگنال کنترلی سیستم برای هر دو روش در طول دوره یادگیری و پس از اعمال کنترل کننده بهینه به ترتیب در شکل۲ و ۳ به‌صورت زیر رسم شده است.



شکل۲: مسیرهای حالت سیستم دوخطی در طول فرآیند یادگیری و بعد ازآن

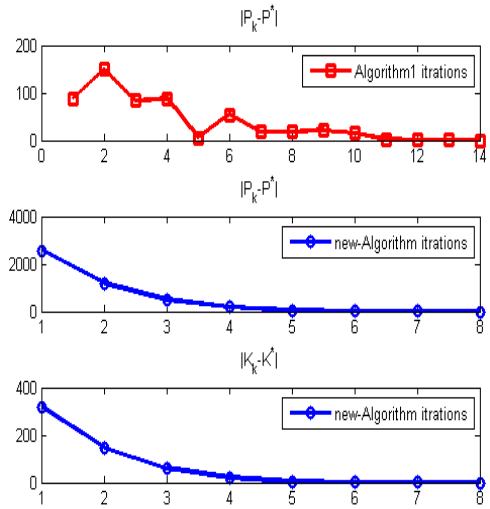


شکل۳: سیگنال کنترلی واردشده به سیستم دوخطی در طول فرآیند یادگیری و بعد ازآن

شکل۲ و شکل۳ زمان میرایی سیستم برای هر دو روش را تقریبا یکسان نشان می دهد. در روش جدید بر مبنای حالات اندازه‌گیری شده و نویز اعمال‌شده به سیستم در بازه‌ی زمانی $[0,0.5]$ ثانیه، همگرایی پس از ۸ بار تکرار حاصل شده است.

$$P^* = P_8 = \begin{bmatrix} 149.2359 & 9.1416 \\ 9.1416 & 2.1417 \end{bmatrix}$$
$$K^* = K_8 = \begin{bmatrix} 2.1417 \\ -1.1427 \end{bmatrix}$$

در حالی که با معلوم بودن دینامیک B روش۱ با انجام محاسبات مشابه بعد از ۱۴ تکرار همگرا شده است. نحوه ی همگرایی P_k و K_k به مقادیر بهینه‌ی خود در شکل ۴ نشان داده‌شده است.



شکل۴: تعداد تکرارها و همگرایی $P_k \to P^*$ و $K_k \to K^*$

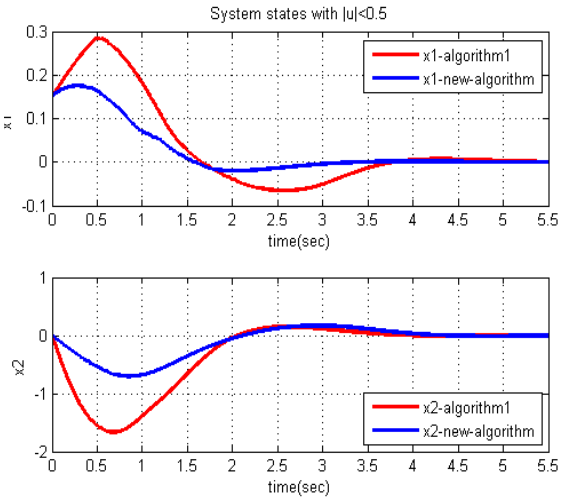
در جدول زیر مقایسه ای بین عملکرد روش جدید و روش۱ صورت گرفته است.

[DOI: 10.61186/joc.17.4.75]

جدول۱. مقایسه روش جدید و روش۱ برای سیستم۱

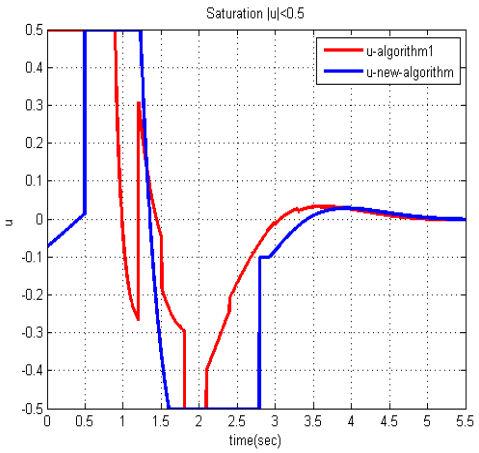
روش	عدم وابستگی به	تعداد تکرار روش	تعداد اجرای سیستم	زمان محاسبات (ثانیه)
روش ۱	ماتریس **A**	۱۴	۱۴	۸.۱۴۳
روش جدید	ماتریسهای **A** و **B**	۸	۲	۰.۹۷۶

جهت بررسی تاثیر ورودی های اشباع بر عملکرد روش پیشنهادی و روش ۱ ، سیگنال ورودی را بصورت کران دار با شرط $|u| < 0.5$ به سیستم (۲۹) اعمال شد. شکل های ۵ و۶ به ترتیب رفتار حالت سیستم و سیگنال های کنترلی تحت اشباع را نشان می دهد.



شکل۵: مسیرهای حالت سیستم دوخطی در طول فرآیند یادگیری و بعد از آن با در نظر گرفتن اشباع در ورودی

همانطور که مشاهده می شود اشباع زمان نشست حالت های سیستم را در هر دو روش به یک میزان افزایش می دهد. همچنین روش پیشنهادی عملکرد بهتری را نسبت به روش۱ تحت اشباع نشان می دهد.



شکل۶: سیگنال کنترلی واردشده به سیستم دوخطی در طول فرآیند یادگیری و بعد از آن با در نظر گرفتن اشباع در ورودی

هر دو روش به کنترل کننده بهینه خود همگرا شدند که جدول ۲ مقایسه ای را از لحاظ تعداد دفعات تکرار و زمان سپری سیستم جهت محاسبه کنترل کننده بهینه با محدودیت اشباع مدنظر را نشان می دهد.

جدول۲. مقایسه روش جدید و روش۱ برای سیستم۱ تحت اشباع در ورودی

روش	عدم وابستگی به	تعداد تکرار روش	تعداد اجرای سیستم	زمان محاسبات (ثانیه)
روش ۱	ماتریس **A**	۲۴	۲۴	۱۲.۶۳۴
روش جدید	ماتریسهای **A** و **B**	۸	۲	۰.۹۷۶

سیستم شبیه سازی شده۲- سیستم دارای دو ورودی زیر را در نظر بگیرید [۱۱،۱۲،۳۶].

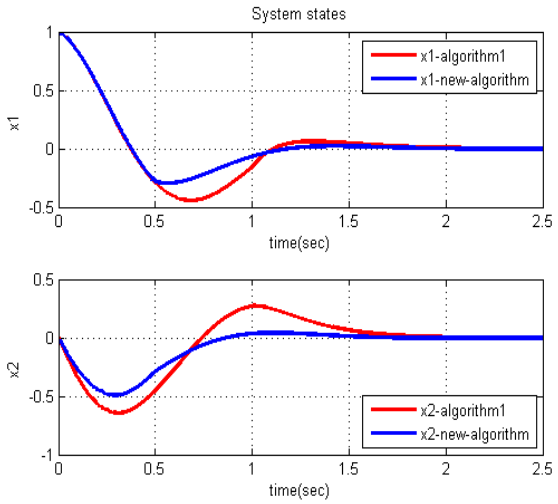
$$\dot{x} = \begin{bmatrix} -1 & 4 \\ -4 & -1 \end{bmatrix} x + \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + x_1 \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + x_2 \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$x_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad (۳۰)$$

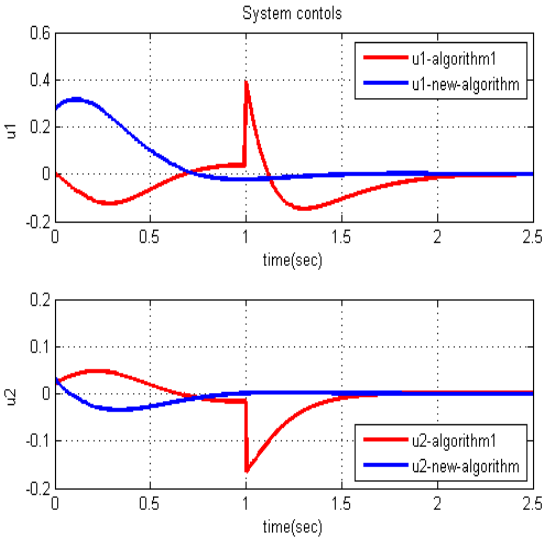
ماتریس‌های وزنی تابع هزینه صورت زیر هستند:

$$Q = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \qquad R = \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix},$$

با انتخاب $K_0 = 0$ نخست برای مدت نیم ثانیه نویز شناسایی $e(t) = 0.5 \sum_{i=1}^{100} \sin(w_i t)$ به سیستم اعمال شده است به‌طوری‌که w_i ها برای i =1,…,100 به‌صورت تصادفی در بازه [-2,2] انتخاب شده‌اند. حالت‌های سیستم (۳۰) در هر ۰.۰۱ثانیه اندازه‌گیری شده اند. همچنین روش۱ با گام های زمانی مشابه و بروزرسانی ماتریس P هر ۰.۵ ثانیه شبیه سازی شده است. حالتها و سیگنال کنترلی سیستم در هر دو روش در طول دوه یادگیری و پس از آن به ترتیب در شکل های۷ و۸ رسم شده اند.



شکل۷: مسیرهای حالت سیستم دوخطی در طول فرآیند یادگیری و بعد ازآن



شکل۸: سیگنال های کنترلی واردشده به سیستم دوخطی در طول فرآیند یادگیری و بعد ازآن

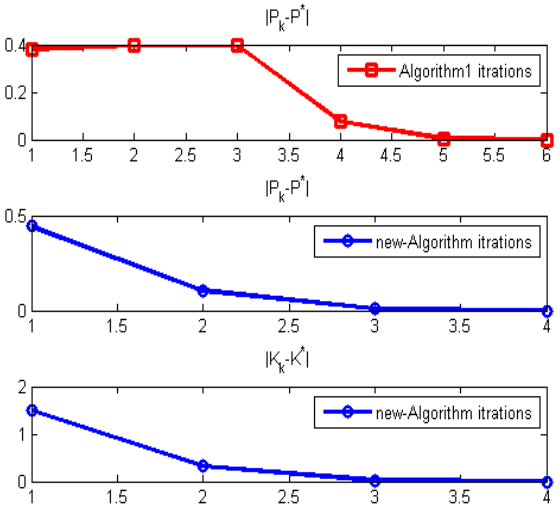
شکل۷ و۸ زمان میرایی سیستم برای هر دو روش را تقریبا یکسان نشان می دهد. در روش جدید بر مبنای حالات اندازه‌گیری شده و نویز اعمال‌شده به سیستم در بازه‌ی زمانی $[0,0.5]$ ثانیه، همگرایی پس از ۴ بار تکرار حاصل شده است.

روش پیشنهادی ماتریس P^* و ماتریس بهره‌ی بازخورد K^* را به‌صورت زیر نتیجه می‌دهد.

$$P^* = P_4 = \begin{bmatrix} 0.3530 & 0.0009 \\ 0.0009 & 0.3985 \end{bmatrix}$$

$$K^* = K_4 = \begin{bmatrix} 1.2936 & -0.2622 \\ -0.2347 & 0.2650 \end{bmatrix}$$

در حالی که با معلوم بودن دینامیک B روش ۱ با انجام محاسبات مشابه بعد از ۶ تکرار همگرا شده است. نحوه ی همگرایی P_k و K_k به مقادیر بهینه‌ی خود در شکل۹ نشان داده‌شده است.



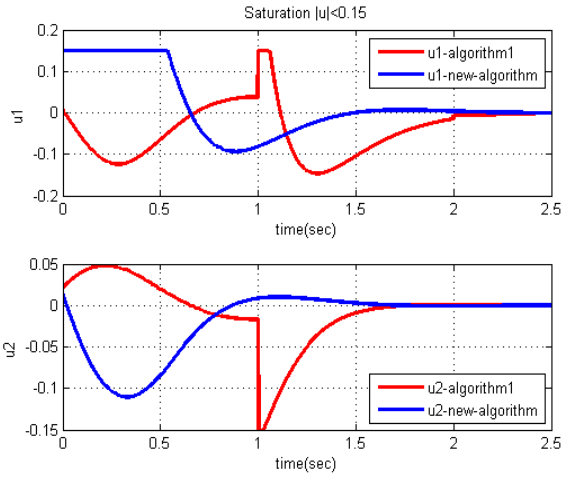
شکل۹. تعداد تکرارها و همگرایی $P_k \to P^*$ و $K_k \to K^*$

در جدول زیر مقایسه ای بین عملکرد روش جدید و روش۱ صورت گرفته است.

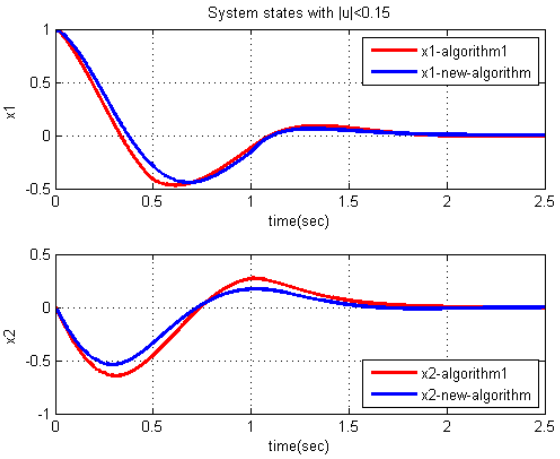
جدول۳. مقایسه روش جدید و روش۱ برای سیستم۲

روش	عدم وابستگی به	تعداد تکرار روش	تعداد اجرای سیستم	زمان محاسبات (ثانیه)
روش ۱	ماتریس **A**	۶	۶	۷.۵۱۹ ثانیه
روش جدید	ماتریسهای **A** و **B**	۴	۲	۰.۷۹۳ثانیه

جهت بررسی تاثیر ورودی با در نظر گرفتن اشباع بر عملکرد روش پیشنهادی و روش ۱ ، سیگنال ورودی را به صورت کران دار با شرط $|u| < 0.15$ به سیستم اعمال شد. شکل های ۱۰ و۱۱ به ترتیب رفتار حالت سیستم و سیگنال های کنترلی تحت اشباع را نشان می دهد.



شکل ۱۰: سیگنال های کنترلی واردشده به سیستم دوخطی در طول فرآیند یادگیری و بعدازآن در نظر گرفتن اشباع در ورودی



شکل ۱۱: مسیرهای حالت سیستم دوخطی در طول فرآیند یادگیری و بعد ازآن در نظر گرفتن اشباع در ورودی

همانطور که مشاهده می شود اشباع زمان نشست حالت های سیستم را در هر دو روش به یک میزان افزایش می دهد. همچنین برای این سیستم نیز روش پیشنهادی عملکرد بهتری را نسبت به روش۱ تحت اشباع کنترل کننده نشان می دهد. هر دو روش به کنترل کننده بهینه خود همگرا شدند که جدول ۴ مقایسه ای را از لحاظ تعداد دفعات تکرار و زمان سپری سیستم

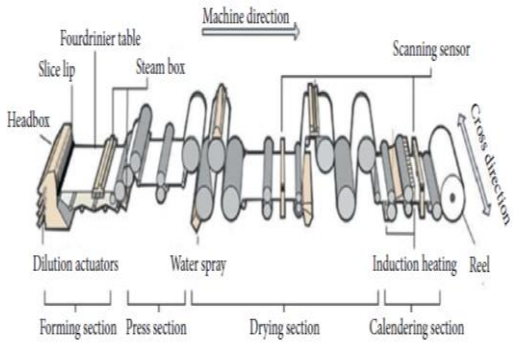
[DOI: 10.61186/joc.17.4.75]

جهت محاسبه کنترل کننده بهینه با محدودیت اشباع مدنظر را نشان می دهد.

جدول۴. مقایسه روش جدید و روش۱ برای سیستم۲ تحت اشباع در ورودی

روش	عدم وابستگی به	تعداد تکرار روش	تعداد اجرای سیستم	زمان محاسبات (ثانیه)
روش ۱	ماتریس **A**	۶	۶	۸.۵۶۱
روش جدید	ماتریسهای **A** و **B**	۴	۲	۱.۰۳۲

سیستم شبیه سازی شده۳- مسئله کنترل ماشین کاغذسازی توسط یک مدل دوخطی توصیف‌شده است [۱۱،۱۲،۳۴] .



شکل۱۲: بخش هایی از فرآیند تولید کاغذی [۳۴]

مدل ریاضی دوخطی این سیستم بر اساس پارامتر های داده به صورت روابط (۱)-(۲) است. ماتریس های این سیستم بصورت ذیل بیان می شود:

$$A=\begin{bmatrix}-1.930 & 0 & 0 & 0\\ 0.394 & -0.426 & 0 & 0\\ 0 & 0 & -0.630 & 0\\ 0.095 & -0.103 & 0.413 & -0.426\end{bmatrix}$$

$$B=\begin{bmatrix}1.274 & 1.274\\ 0 & 0\\ 1.34 & -0.65\\ 0 & 0\end{bmatrix} \qquad X_0=\begin{bmatrix}4\\ 3.2\\ 4\\ 2.8\end{bmatrix}$$

$$N1=\begin{bmatrix}0 & 0\\ 0 & 0\\ 0.755 & -0.366\\ 0 & 0\end{bmatrix} \qquad N2=N4=\begin{bmatrix}0 & 0\\ 0 & 0\\ 0 & 0\\ 0 & 0\end{bmatrix}$$

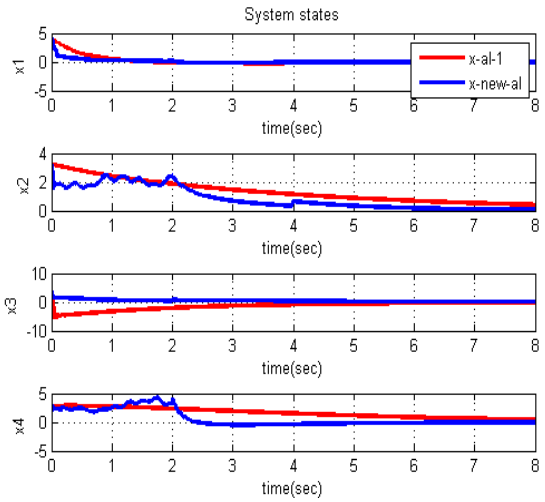
$$,N3=\begin{bmatrix}0 & 0\\ 0 & 0\\ -0.718 & -0.718\\ 0 & 0\end{bmatrix}.$$

ماتریس‌های وزنی تابع هزینه صورت زیر هستند:

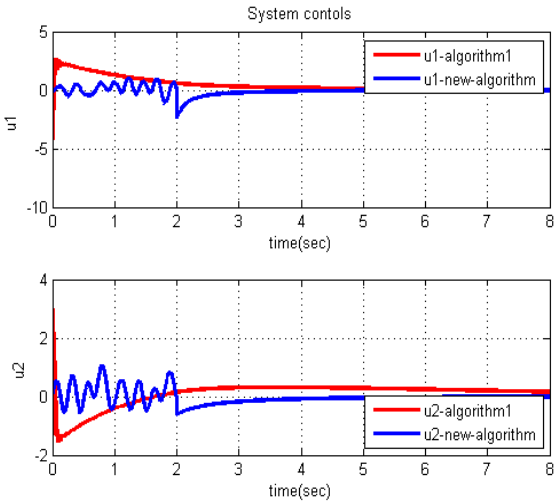
$$, \quad Q=\begin{bmatrix}1 & 0 & 0.13 & 0\\ 0 & 1 & 0 & 0.09\\ 0.13 & 0 & 0.1 & 0\\ 0 & 0.09 & 0 & 0.2\end{bmatrix}$$

$$R=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

با انتخاب $K_0=0$، نخست برای مدت ۱.۵ ثانیه نویز شناسایی $e(t)=0.5\sum_{i=1}^{100}sin\,(w_i t)$ به سیستم اعمال شده است به‌طوری‌که w_i ها برای i =1,...,100 به‌صورت تصادفی در بازه $[-10,10]$ انتخاب‌شدند. حالت‌های سیستم در هر ۰.۰۱ثانیه اندازه‌گیری شده اند . همچنین روش۱ را با تقسیم بندی مشابه و بروزرسانی ماتریس P هر ۰.۵ ثانیه شبیه سازی شده است. حالت ها و سیگنال های کنترلی سیستم هر دو روش در طول دوه یادگیری و پس از آن به ترتیب در شکل های ۱۳ و ۱۴ رسم شده است.



شکل۱۳: مسیرهای حالت سیستم دوخطی در طول فرآیند یادگیری و بعدازآن



شکل۱۴: سیگنال های کنترلی واردشده به سیستم دوخطی در طول فرآیند یادگیری و بعد ازآن

شکل۱۳ همگرایی حالت ها را با روش جدید، سریع تر از روش۱ نشان می دهد. در روش جدید بر مبنای حالات اندازه‌گیری شده و نویز اعمال‌شده به سیستم در بازه‌ی زمانی $[0,1.5]$ ثانیه، همگرایی پس از ۳ بار تکرار حاصل شده است.

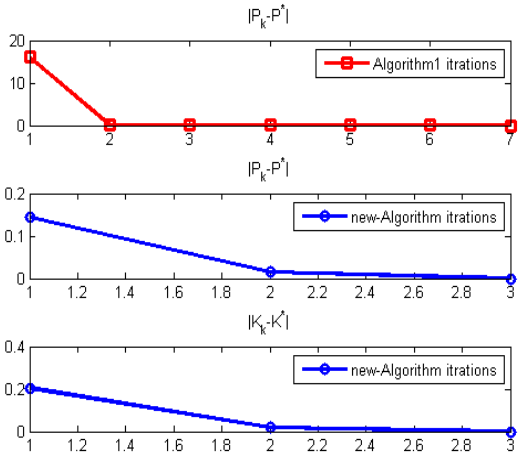
روش پیشنهادی ماتریس P^* و ماتریس بهره‌ی بازخورد K^* را به‌صورت زیر نتیجه می‌دهد.

$$P^*=\begin{bmatrix}0.2365 & 0.1380 & 0.0328 & 0.0087\\ 0.1380 & 1.0867 & -0.0052 & 0.655\\ 0.0328 & -0.0052 & 0.0996 & 0.0686\\ 0.0087 & 0.0655 & 0.0686 & 0.2210\end{bmatrix}$$

$$K^*=\begin{bmatrix}0.3452 & 0.1688 & 0.1752 & 0.1031\\ 0.2799 & 0.1791 & -0.0229 & -0.0334\end{bmatrix}$$

[DOI: 10.61186/joc.17.4.75]

در حالی که با معلوم بودن دینامیک B و انجام محاسبات مشابه روش۱ بعد از ۷ تکرار همگرا شده است.نحوه ی همگرایی P_k و K_k به مقادیر بهینه‌ی خود در شکل۱۵ نشان داده‌شده است.



شکل۱۵: تعداد تکرارها و همگرایی $P_k \to P^*$ و $K_k \to K^*$

در جدول زیر مقایسه ای بین دو عملکرد روش جدید و روش۱ صورت گرفته است.

جدول۵. مقایسه روش جدید و روش۱ برای سیستم۳

روش	عدم وابستگی به	تعداد تکرار روش	تعداد اجرای سیستم	زمان محاسبات (ثانیه)
روش ۱	ماتریس **A**	۶	۶	۹.۹۱۰ ثانیه
روش جدید	ماتریسهای **A** و **B**	۳	۲	۲.۱۸۰ ثامیه

۵- نتیجه‌گیری

روش‌های طراحی کنترل‌کننده‌ی بهینه برای سیستم‌های دوخطی روش‌هایی تقریبی و وابسته به دانستن دینامیک کامل سیستم یا حداقل وابسته به دینامیک ماتریس ورودی هستند. در این مقاله با استفاده از بازخورد خطی حالت بر پایه سیاست یادگیری، یک روش کنترل بهینه تطبیقی بر خط جدید ارائه شد. روش پیشنهادی پاسخ تقریبی معادله HJB برای این سیستم‌ها را در یک فرآیند تکراری تقریبی-تطبیقی با اندازه‌گیری برخط حالت و ورودی اعمال‌شده به سیستم در یک بازه‌ی زمانی ثابت و بدون نیاز به دانستن ماتریس حالت و ماتریس ورودی در اختیار قرار می دهد. همگرایی روش تکرای و برخط جدید به کنترل کننده بهینه به صورت قضیه بیان و اثبات شده است. در پایان نیز برای نشان دادن کارایی روش پیشنهادی سه سیستم فیزیکی مورد استفاده پژوهشگران پیشین توسط روش جدید بدون استفاده از ماتریس های حالت و ماتریس ورودی سیستم شبیه سازی شدند. شبیه سازی ها نشان می دهد از نظر پاسخ زمانی و تعداد دفعات تکرار روش وزمان اجرای سیستم روش جدید عملکرد قابل مقایسه ای نسبت به روشی که ماتریس ورودی را در محاسبات خود اعمال می کند دارد. همچنین روش پیشنهادی تحت ورودی اشباع شبیه سازی و تحلیل شد. در این حالت نیز عملکرد قابل قبولی را علی رغم عدم دسترسی به ماتریس های سیستم نسبت به روش ذکر شده نشان می دهد. روش پیشنهادی می‌تواند به‌عنوان یک ابزار مفید و کارای محاسباتی برای مطالعه‌ی کنترل‌کننده‌ی بهینه‌ی تطبیقی سیستم‌های غیرخطی با دینامیک نامعلوم باشد .

مراجع

[1] M. Ven, "Input-to-State Stability for bilinear systems," MS thesis. University of Twente, 2020.

[2] R.R Mohler, And A.Y. Khapalov, "Bilinear control and application to flexible ac transmission systems. Journal of Optimization Theory and Applications," 105, pp. 621-637, 2000.

[3] D. Williamson, "Observation of bilinear systems with application to biological control," Automatica, 13(3), pp. 243-254, 1977.

[4] O. Balatif, I. Abdelbaki, M. Rachik, and Z. Rachik, "Optimal control for multi-input bilinear systems with an application in cancer chemotherapy," International Journal of Scientific and Innovative Mathematical Research (IJSIMR), 3(2), pp. 22-31, 2015.

[5] D. Gao, Q. Yang, M. Wang and Y. Yu, "Feedback linearization optimal control approach for bilinear systems in CSTR chemical reactor," Intelligent Control and Automation, 3(03), p. 274, 2012.

[6] M.V. Basin and M.A.A. García, "Optimal filtering for bilinear system states and its application to terpolymerization process identification'. Applied Mathematics E-Notes, 4, pp. 7-15, 2004.

[7] T. Naik, "Uncertainty propagation in bilinear and polynomial system for probabilistic threshold detection," Master Thesis, Delf University of Technology, 2021.

[8] P.M.S. Burt and J.H. de Morais Goulart, "Efficient computation of bilinear approximations and volterra models of nonlinear systems," IEEE Transactions on Signal Processing, 66(3), pp. 804-816, 2017.

[9] F.L. Lewis, D.L. Vrabie, and V.L. Syrmos, "Reinforcement learning and optimal adaptive control. Optimal Control," Third Edition, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2012.

[10] D.E. Kirk, Optimal control theory: An introduction. Courier Corporation, 2004.

[11] W.A. Cebuhar, and V. Costanza, "Approximation procedures for the optimal control of bilinear and nonlinear systems," Journal of Optimization Theory and Applications, 43, pp. 615-627,1984.

[12] Z. Aganovic, and Z. Gajic, "Successive approximation procedure for steady-state optimal

[DOI: 10.61186/joc.17.4.75]

[25] S. He, H. Fang, M. Zhang, F. Liu, and Z. Ding, "Adaptive optimal control for a class of nonlinear systems: The online policy iteration approach," IEEE transactions on neural networks and learning systems, 31(2), pp. 549-558, 2019.

[26] Y. Jiang, and Z.P. Jiang, "Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics," Automatica, 48(10), pp. 2699-2704, 2012.

[27] K. Zhang, and S.L. Ge, "Adaptive optimal control with guaranteed convergence rate for continuous-time linear systems with completely unknown dynamics," IEEE Access, 7, pp. 11526-11532, 2019.

[28] Z. Shi, and Z. Wang, "Adaptive output-feedback optimal control for continuous-time linear systems based on adaptive dynamic programming approach" Neurocomputing, 438, pp. 334-344, 2021.

[29] M. Gan, J. and C. Zhang, "Extended adaptive optimal Zhao, control of linear systems with unknown dynamics using adaptive dynamic programming," Asian Journal of Control, 23(2), pp. 1097-1106, 2021.

[30] Q. Wei, L. Zhu , R. Song, P. Zhang, D. Liu, and J. Xiao, "Model-free adaptive optimal control for unknown nonlinear multiplayer nonzero-sum game," IEEE Transactions on Neural Networks and Learning Systems, 33(2), pp. 879-892, 2020.

[31] D. Xu, Q.Wang and Y. Li, "Adaptive optimal control approach to robust tracking of uncertain linear systems based on policy iteration," Measurement and Control, 54(5-6), pp. 668-680, 2021.

[32] J. Zhang, H. Zhang, Z.Liu, and Y. Wang, "Model-free optimal controller design for continuous-time nonlinear systems by adaptive dynamic programming based on a pre-compensator," ISA Transactions, 57, pp. 63-70, 2015.

[33] Z.Yuan, and J. Cortés. "Data-driven optimal control of bilinear systems," IEEE Control Systems Letters, 6, pp. 2479-2484, 2022.

[34] B. Iben Warrad, M.K. Bouafoura and N.Benhadj Braiek, "Combined constrained robust least squares approach and block-pulse functions technique for tracking control synthesis of uncertain bilinear systems with multiple time-delayed states under bounded input control," Mathematical Problems in Engineering, 2020, pp. 1-28, 2020.

[35] D. Goswami, and D.A. Paley, "Bilinearization, reachability, and optimal control of control-affine nonlinear systems: A Koopman spectral approach," IEEE Transactions on Automatic Control, 67(6), pp. 2715-2728, 2021.

control of bilinear systems," Journal of optimization theory and applications, 84, pp. 273-291. 1995.

[13] M. Ekman, Modeling and control of bilinear systems: application to the activated sludge process. Diss. Acta Universitatis Upsaliensis, 2005.

[14] H.Wang, M. Zhu, W. Hong, C. Wang, W. Li, G.Tao, and Y. Wang, "Network-wide traffic signal control using bilinear system modeling and adaptive optimization," IEEE Transactions on Intelligent Transportation Systems, 24(1), pp.79-91, 2022.

[15] S. Bichiou, M.K. Bouafoura, and N. Benhadj Braiek, "Time optimal control laws for bilinear systems," Mathematical Problems in Engineering, 2018.

[16] D. Gao, Q. Yang, M. Wang, and Y. Yu, "Feedback linearization optimal control approach for bilinear systems in CSTR chemical reactor," Intelligent Control and Automation, 3(03), pp. 274-277, 2012.

[17] X. Yang, H. He, D. Liu, and Y. Zhu, "Adaptive dynamic programming for robust neural control of unknown continuous-time non-linear systems," IET Control Theory & Applications, 11(14), pp. 2307-2316, 2017.

[18] Y. Wen, J. Si, A. Brandt, X. Gao, and H.H. Huang, "Online reinforcement learning control for the personalization of a robotic knee prosthesis," IEEE Transactions on Cybernetics, 50(6), pp. 2346-2356, 2019.

[19] T. Tan, F. Bao, Y. Deng, A. Jin, Q. Dai, and J. Wang, Cooperative deep reinforcement learning for large-scale traffic grid signal control. IEEE transactions on cybernetics, 50(6), pp. 2687-2700, 2019.

[20] J.J. Murray, C.J. Cox, G.G. Lendaris, and R. Saeks, "Adaptive dynamic programming," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 32(2), pp. 140-153, 2002.

[21] D. Vrabie, "Online adaptive optimal control for continuous-time systems", 2010.

[22] D. Vrabie, O. Pastravanu, M. Abu-Khalaf, and F.L. Lewis, "Adaptive optimal control for continuous-time linear systems based on policy iteration," Automatica, 45(2), pp. 477-484, 2009.

[23] L.B. Prasad, H.O. Gupta, and B. Tyagi, "Application of policy iteration technique based adaptive optimal control design for automatic voltage regulator of power system," International Journal of Electrical Power & Energy Systems, 63, pp. 940-949, 2014.

[24] D. Vrabie and F.L. Lewis, "Adaptive optimal control algorithm for continuous-time nonlinear systems based on policy iteration," In 2008 47th IEEE Conference on Decision and Control, pp. 73-79, IEEE, 2008.

[DOI: 10.61186/joc.17.4.75]

[36] B. Luo, and H.N. Wu, "Online adaptive optimal control for bilinear systems," In 2012 American Control Conference (ACC), pp. 5507-5512, IEEE, June 2012.

[37] R. Longchamp, "Controller design for bilinear systems," IEEE Transactions on Automatic Control, 25(3), pp. 547-548.1980.

[38] I. Derese, and E.Noldus, "Design of linear feedback laws for bilinear systems," International Journal of Control, 31(2), pp. 219-237. 1980.

[39] A. Benallou, D.A Mellichamp, and D.E. Seborg, "Optimal stabilizing controllers for bilinear systems," International Journal of Control, *48*(4), pp. 1487-1501, 1988.

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

[41] D. Kleinman, "On an iterative technique for Riccati equation computations," IEEE Transactions on Automatic Control, 13(1), pp. 114-115. 1968

[42] A. Al-Tamimi, F.L. Lewis and M. Abu-Khalaf, "Model-free Q-learning designs for linear discrete-time zero-sum games with application to H-infinity control," Automatica, 43(3), pp. 473-481. 2007.

[43] X. Feng, and Z. Zhang, "The rank of a random matrix," Applied mathematics and computation, 185(1), pp. 689-694. 2007.

[44] H. Modares, F.L. Lewis, and M.B.N. Sistani, "Online solution of nonquadratic two-player zero-sum games arising in the H∞ control of constrained input systems," International Journal of Adaptive Control and Signal Processing, 28(3-5), pp. 232-254. 2014.

[DOI: 10.61186/joc.17.4.75]