برنامه‌ریزی پویا برای حل مسئله کنترل بهینه درجه دوم خطی تصادفی

محمود محمودی[۱] ، مدینه فرنام[۲] *

[۱] گروه ریاضی، دانشکده علوم، دانشگاه قم، قم، ایران

Mmahmoodi_ir@yahoo.com

[۲] دانشجوی دوره دکتری گروه ریاضی، دانشکده علوم، دانشگاه قم، قم، ایران

M.Farnam@stu.qom.ac.ir

چکیده: در این مقاله الگوریتم پویای تطبیقی (ADP) برپایه روشی تکراری مقدار برای حل مسئله کنترل بهینه تصادفی (SLQ) در زمان نامتناهی برای سیستم‌های گسسته زمان با دینامیکی کاملاً ناشناخته به کار رفته است. به طور کلی حل مسئله کنترل بهینه درجه دوم نیاز به اطلاعات کاملی از دینامیک سیستم دارد. ابتدا مسئله SLQ با استفاده از تبدیل مناسب، به سیستمی قطعی تبدیل شده است. سپس الگوریتم تکراری مقدار برای حل مسئله کنترل بهینه با تحلیل همگرایی معرفی شده است. در ادامه به عنوان کاربردی از الگوریتم تکراری، یک شبکه عصبی (NN) برای شناسایی سیستم ناشناخته بکار رفته است و آنگاه از دو NNs برای تقریب تابع هزینه و ماتریس بهره کنترل استفاده شده است. در نهایت مثالی شبیه‌سازی شده برای توضیح اثربخشی الگوریتم ارائه می‌شود.

۰۱ پیش‌گفتار

مسئله کنترل بهینه SLQ با پیشگامی ونهام [۱] شروع شده و سرعت بسیاری در نظریه و کاربرد یافته است[۳-۲]. این نوع مسائل نقش مهمی در نظریه کنترل مدرن می‌توانند داشته باشند و در مقایسه با حالت قطعی بسیار پیچیده‌تر از معادله ریکاتی می‌باشند[٤]. برخی روابط اصلی بین معادله جبری تصادفی و معادله ریکاتی در [٥] بحث شده است. اما با توجه به ویژگی‌های غیرخطی SAE، یافتن جواب تحلیلی مشکل می‌باشد. به ویژه وقتی دینامیک سیستم ناشناخته باشد بسیاری از روش‌های مطرح شناخته شده معتبر نخواهند بود. بنابراین اگر بخواهیم مسئله کنترل بهینه SLQ را بدون اطلاعی از مدل سیستم حل کنیم با مسئله‌ای چالش برانگیز مواجه خواهیم بود. اما با معرفی برخی از روشهای ریاضی، حل مسئله SLQ سهولت بیشتری می‌یابد. برنامه‌ریزی پویای تطبیقی (ADP) به عنوان ابزاری قدرتمند در حل مسئله‌های کنترل بهینه قطعی توجهات بسیاری را به خود جلب کرده [۷-٦] و اخیراً توسعه‌های بسیاری یافته‌است[۸]. در این مقاله می‌خواهیم روش را برای کنترل بهینه SLQ در سیستم گسسته زمانی ناشناخته به کار ببریم. به منظور جلوگیری از حل مستقیم معادله جبری تصادفی، الگوریتم ADP تکراری مقدار و همگرایی آن ارائه می‌شود. به عنوان کاربردی از الگوریتم برای سیستم‌های ناشناخته سه شبکه‌ی عصبی برای تقریب مدل سیستم، تابع هزینه و ماتریس بهره کنترل به کار رفته است. در ادامه تعریف مسئله، الگوریتم ADP تکراری و همگرایی آن به علاوه کاربرد شبکه‌های عصبی برای طرح تکراری آورده شده است. در انتها نتیجه‌گیری مباحث ارائه می‌شود.

2010 Mathematics Subject Classification. Primary 47A55; Secondary 39B52, 34K20, 39B82.

واژگان کلیدی: کنترل بهینه درجه دوم خطی تصادفی، شبکه‌های عصبی، برنامه‌ریزی پویای تطبیقی، ماتریس بهره کنترل.

* سخنران

۲. بیان مسئله

در این مقاله سیستم گسسته زمانی خطی تصادفی زیر را در نظر می‌گیریم:

$$x_{t+1} = (Ax_t + Cx_t w_t^x) + (Bu_t + Du_t w_t^u) \qquad (1)$$

که $x_t \in R^n$ ، $u_t \in R^m$ به ترتیب متناظر با بردار وضعیت سیستم و بردار ورودی کنترل می‌باشند.

x_0 وضعیت اولیه، $A, C \epsilon R^{n\times n}$ و $B, D \epsilon R^{n\times m}$ ماتریس‌های حقیقی قطعی داده شده‌اند. دنباله ورودی اغتشاش تصادفی$\{w_t^x, w_t^u;\ t = 0, 1, 2, \dots\}$ روی فضای احتمالاتی کامل (Ω, F, ρ)می‌باشند که به صورت متغیرهای تصادفی عددی هستند. در این مطالعه، فرض می‌شود که دنباله ورودی اغتشاش تصادفی در شرایط زیر صدق می‌کند:

۱) $w_0^x = 0$ و $w_0^u = 0$

۲) $E(w_0^x) = 0$ و $E(w_0^u) = 0$

۳) $E(x_0 w_0^x) = 0$ و $E(x_0 w_0^u) = 0$

٤) $E(w_t^u w_s^u) = \delta_{st}$ و $E(w_t^x w_s^x) = \delta_{st}$ و $E(w_t^x w_s^u) = 0$

هر کنترل نیاز به معیاری برای اندازه‌گیری بهینگی دارد. تابع کمکی هزینه (۱) به صورت زیر می‌باشد:

$$J(x_0, u) = E\sum_{t=0}^{\infty}(x_t^T Q x_t + u_t^T R u_t) \qquad (2)$$

در این مقاله کنترل بهینه در گروهی از فیدیک‌های خطی به فرم زیر به دست می‌آید:

$$u_t = Kx_t \quad \text{و} \quad K \in R^{m\times n} \qquad (3)$$

اکنون می‌خواهیم مسئله SLQ را به نوع قطعی نظیر آن تبدیل نماییم. برای این منظور فرض می‌کنیم، $X_t = E(x_t x_t^T)$ از اینرو سیستم (۱) می تواند تبدیل شود به:

$$X_{t+1} = E(x_{t+1} x_{t+1}^T) \qquad (4)$$

$$= E([(A + BK) + (Cw_t^x + DKw_t^u)]x_t x_t^T[(A + BK) + (Cw_t^x + DKw_t^u)]^T \qquad (5)$$

که در آن $u_t = Kx_t$ کنترل می باشد. بعد از محاسباتی ساده، سیستم (۵) به صورت زیر ساده می‌شود:

$$X_{t+1} = (A + BK)X_t(A + BK)^T + CX_t C^T + DKX_t K^T D^T \qquad (6)$$

که $X_t \in R^{n\times n}$. به این طریق مسئله کنترل بهینه SLQ به صورت زیر محاسبه می‌شود:

$$V(X_0) = min\ J(X_0, K) \qquad \text{که} \qquad J(X_0, K) = tr\{\sum_{t=0}^{\infty}(F + K^T R K)X_t\}$$

۳. کنترل بهینه برپایه الگوریتم تکراری ADP

۳. ۱ اقتباس از الگوریتم یادگیری ADP

فرض کنید $u_t = Kx_t$ کنترل قابل قبول باشد داریم:

$$J(X_k,K) = tr\{(Q + K^T RK)X_k\} + tr\{\sum_{t=k+1}^{\infty}(Q + K^T RK)X_t\} \qquad (۷)$$

براساس اصل بهینگی بلمن تابع هزینه باید رابطه هامیلتون، ژاکوبین، بلمن را برقرار سازد:

$$V^*(X_k) = \min_K\{tr\{(Q + K^T RK)X_k\} + V^*(X_{k+1}) \qquad (۸)$$

در نهایت برای i =۱,۲,۳,...، الگوریتم تکراری می‌تواند بین روابط زیر به کار رود:

$$K_i = \arg\min_K\{(tr\{(Q + K^T RK)X_k\}) + V_i(X_{k+1})\} \qquad (۹)$$

و

$$V_{i+1}(X_k) = \min_K\{(tr\{(Q + K^T RK)X_k\}) + V_i(X_{k+1})\}$$

$$= tr\left\{\left((Q + {K_i}^T RK_i)X_k\right)\right\} + V(\tilde{X}_{k+1}) \qquad (۱۰)$$

که در آن $\tilde{X}_{k+1} = (A + BK_i)X_k(A + BK_i)^T + (C + DK_i)X_k(C + DK_i)^T$

و در آن i اندیس تکرار و k اندیس زمان است. تابع هزینه و ماتریس بهره کنترل با تکرار بازگشتی، هنگامی که i از صفر تا بی‌نهایت افزایش می‌یابد به روزرسانی می‌شوند.

۲.۳ همگرایی الگوریتم تکراری *ADP*

در این قسمت قضیه همگرایی الگوریتم *ADP* آورده شده است.

قضیه ۱: دنباله تابع هزینه $\{V_i\}$ در رابطه (۱۰) در نظر می‌گیریم، در این صورت داریم:

$$V_\infty(X_k) = \min_K\{(tr\{(Q + K^T RK)X_k\}) + V_\infty(X_{k+1})\}$$

قضیه۲: دنباله $\{K_i\}$ و $\{V_i\}$ تعریف شده در روابط (۹) و (۱۰) را در نظر بگیرید، در این صورت داریم:

$$V_\infty = V^* \text{ و } K_\infty = K^*$$

٤. کاربرد NN در الگوریتم تکراری

برای اجرای الگوریتم تکراری ADP بین (۹) و (۱۰) شبکه‌های عصبی را برای تقریب تابع هزینه $V_i(X_k)$ و ماتریس بهره کنترل K_iدر هر تکرار به کار می‌بریم. تعداد نورون‌های لایه پنهان را با l نشان می‌دهیم. ماتریس وزن بین لایه ورودی و لایه پنهان v و بین لایه پنهان و خروجی w است. در این صورت خروجی NN عبارتست از:

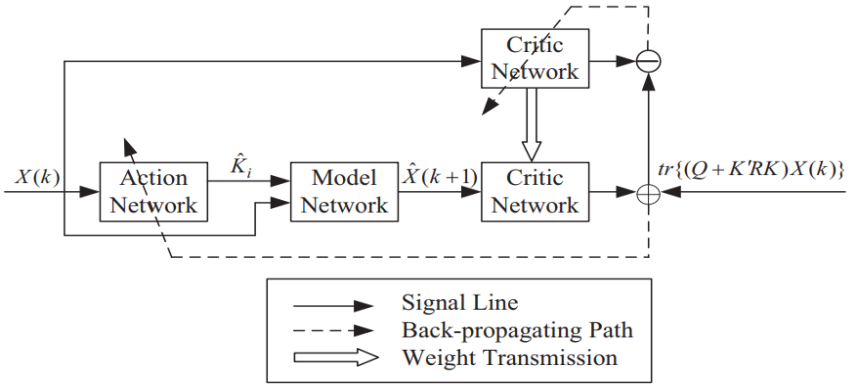
$$\hat{F}(x,v,w) = w'\sigma(v'x) \qquad \text{و} \qquad [\sigma(z)]_i = \frac{e^{z_i} - e^{-z_i}}{e^{z_i} + e^{-z_i}} \qquad i = ۱.۲.۳ \dots . l$$

که در آن $\|\sigma(v'x)\| < \sigma_c$یعنی تابع فعالسازی کران‌دار است. تخمین NN می‌تواند به صورت

$$\hat{F}(x,v,w) = w^{*'}\sigma(v^{*'}x) + \varepsilon(x)$$

نوشته شود، که w^* و v^* ماتریس‌های وزن ایده‌ال و $\varepsilon(x)$ خطای دوباره سازی می‌باشند. به منظور به کارگیری الگوریتم تکراری ADP، سه لایه پیشرو NNsانتخاب شده است که به صورت زیر می‌باشند:

شبکه مدل: برای شناسایی سیستم ناشناخته
شبکه نقاد: برای تقریب تابع هزینه
شبکه عملکرد: برای تقریب ماتریس بهره کنترل



شکل ۱: ساختار دیاگرام الگوریتم مقداری ADP

برای سیستم ناشناخته (۱) شبکه مدل را قبل از اجرای الگوریتم ADP آموزش می‌دهیم. هنگامی که یادگیری شبکه مدل پایان یافت، وزن‌ها ثابت و بدون تغییر می‌ماند. براساس شبکه مدل، شبکه نقاد برای تقریب تابع هزینه استفاده می‌شود.

نتیجه گیری و پیشنهادات آتی: حل مسئله کنترل بهینه SLQ از طریق معادله جبری تصادفی نیاز به شناسایی و اطلاعات کاملی از دینامیک سیستم دارد. وقتی سیستم ناشناخته است اغلب الگوریتم‌های حل ناکارآمد می‌شوند. الگوریتم تکراری ADP در این مقاله برای حل مسئله کنترل بهینه SLQ بکار رفته است که برای حل با این روش نیاز به اطلاعاتی از دینامیک سیستم نمی‌باشد. علاوه براین دنباله‌های ماتریس بهره کنترل و ماتریس هزینه به مقادیر بهینه خود همگرا می‌شوند. تحقیقات بسیار کمی بین محققین داخلی در این حوزه ثبت شده است. استفاده از توابع فعالسازی دیگر و به کارگیری روشهای ترکیباتی جهت افزایش سرعت همگرایی یا بررسی این روش حل در سایر مسائل کنترل بهینه می‌تواند موضوع تحقیقات آتی باشد.

مراجع

۱, *W.M. Wonham, On a matrix Riccati equation of stochastic control, SIAM J. Control 6 (1968) 681–697 .*

۲, *X.J. Su, P. Shi, L.G. Wu, S.K. Nguang, Induced filtering of fuzzy stochastic systems with time-varying delays, IEEE Trans. Cybern. 43 (2013) 1251–1264.*

۳, *X.J. Su, L.G. Wu, P. Shi, Y.D. Song, A novel approach to output feedback control of fuzzy stochastic systems, Automatica 50 (2014) 3268–3275.*

٤, *A. AI-Tamimi, F.L. Lewis, M. Abu-Khalaf, Discrete-time nonlinear HJB solution using approximate dynamic programming: convergence proof, IEEE Trans Syst. Man Cybern. Part B 38 (2008) 943–949.*

٥, *S. Chen, J. Yong, Stochastic linear quadratic optimal control problems, Appl.Math. Optim. 43 (2001) ٢١–٤٥.*

٦, *P.J. Werbos, Approximate dynamic programming for real-time control and neural modeling, in: Handbook of Intelligent Control: Neural Fuzzy and Adaptive Approaches , vol. 15, 1992, pp. 493–525.*

٧, *H.G. Zhang, Y.H. Luo, D.R. Liu, Neural-network-based near-optimal control for a class of discrete-time affine nonlinear systems with control constraints, IEEE Trans. Neural Netw. 20 (2009) 1490–1503.*

٨, *H.G. Zhang, D.R. Liu, Y.H. Luo, D. Wang, Adaptive Dynamic Programming for Control Algorithms and Stability, Springer-Verlag, London, (٢٠١٣).*