مجله کنترل
ISSN 2008-8345
جلد ۱۰، شماره ۲، تابستان ۱۳۹۵، صفحه ۴۳-۳۵

یک الگوریتم تکراری برای حل مسایل کنترل بهینه تصادفی با استفاده از زنجیر مارکوف

بهزاد کفاش [۱]، زهرا نیکویی‌نژاد [۲]، علی دلاورخلفی [۳]

[۱] استادیار، دانشکده فنی و مهندسی، گروه علوم مهندسی، دانشگاه اردکان Bkafash@ardakan.ac.ir
[۲] دانشجوی دکتری ریاضی کاربردی، دانشکده ریاضی، گروه ریاضی کاربردی، دانشگاه یزد Nikoueinezhad@yahoo.com
[۳] دانشیار، دانشکده ریاضی، گروه ریاضی کاربردی، دانشگاه یزد Delavarkh@yazd.ac.ir

(تاریخ دریافت مقاله ۱۳۹۵/۲/۲۵، تاریخ پذیرش مقاله ۱۳۹۵/۵/۱۳)

چکیده: در این مقاله، یک روش عددی برای حل مساله کنترل بهینه تصادفی با استفاده از زنجیرهای مارکوف ارائه شده‌است. بدین ترتیب که، ابتدا فرایند پخش کنترلی وضعیت سیستم با استفاده از یک زنجیر مارکوف کنترلی روی یک فضای وضعیت متناهی تقریب زده می‌شود. سپس تقریبی از تابع هزینه اولیه با استفاده از این زنجیر مارکوف تقریبی، به‌دست می‌آید. برای اثبات همگرایی روش و یافتن یک زنجیر مارکوف تقریبی مناسب برای فرایند پخش، باید دو معیار مورد بررسی قرار گیرد. به عبارتی باید امید ریاضی و کوواریانس شرطی تغییرات وضعیت زنجیر مارکوف با میانگین و کوواریانس موضعی فرایند پخش اولیه متناسب باشند. با استفاده از تقریبات تفاضلات متناهی می‌توان احتمالات انتقال و بازه‌های زمانی تغییر وضعیت زنجیر مارکوف را به گونه‌ای تعیین کرد که زنجیر مارکوف در دو ویژگی سازگاری موضعی فوق صدق کند. در ادامه معادله برنامه‌ریزی پویا با زنجیر مارکوفی که بدین ترتیب به‌دست آمده و دارای این ویژگی های می‌باشد، تقریب زده می‌شود. نهایتاً، با میل دادن پارامتر گسسته سازی زنجیر مارکوف به صفر، مشاهده می‌شود که جواب مسئله کنترل تصادفی تقریبی زنجیر مارکوف، به جواب مسئله کنترل بهینه تصادفی اولیه همگرا می‌باشد. در پایان یک الگوریتم تکراری برای حل مساله کنترل بهینه تصادفی پیشنهاد شده و از آن برای حل یک مثال استفاده شده است.

کلمات کلیدی: مساله کنترل بهینه تصادفی، زنجیر مارکوف، روش عددی، الگوریتم تکراری.

An Iterative Algorithm for Solving Stochastic Optimal Control via the Markov Chain Approximation

Behzad Kafash, Zahra Nikoeenezhad, Ali Delavarkhalafi

Abstract: In this paper, a numerical method for solving stochastic optimal control problem by using Markov chain approximation method has presented. The basic idea of the Markov chain approximation method is to approximate the original controlled process by an appropriate controlled Markov chain on a finite state space. Also, we need to approximate the original cost function by one which is appropriate for the approximating chain. These approximations should be chosen such that a good numerical approximation to the associated optimal control problem can be obtained, which means the conditional mean and covariance of the changes in state of the chain are proportional to the local mean drift and covariance for the original process. The finite difference approximations are used to the construction of locally consistent approximating Markov chain, the coefficients of the resulting discrete equation can serve as the desired transition probabilities and interpolation interval.

Keywords: Stochastic optimal control problem, Markov chain approximation, Numerical method, iterative algorithm.

۱- مقدمه

در مهندسی و ریاضیات نظریه کنترل با رفتار سیستم‌های دینامیکی سر و کار دارد. سیستم‌های دینامیکی به سامانه‌هایی اطلاق می‌گردد که حالات آن‌ها با زمان تغییر می‌کند که اغلب در صورت وجود عدم قطعیت و اغتشاش در سیستم رفتار آن‌ها به وسیله یک معادله دیفرانسیل تصادفی توصیف می‌شود. مفهوم کنترل را می‌توان به عنوان فرایندی که رفتار یک سیستم دینامیکی را برای رسیدن به یک هدف خاص تحت تاثیر قرار می‌دهد، در نظر گرفت. اگر هدف بهینه‌سازی تابع عملکردی باشد که به متغیر کنترل یا ورودی از سیستم وابسته است، آن‌گاه مساله را کنترل بهینه تصادفی می‌نامند. نظریه کنترل بهینه، یک روش بهینه سازی ریاضی است که برای به دست آوردن سیاست‌های کنترلی مورد استفاده قرار می‌گیرد و تعمیمی بر حساب تغییرات می‌باشد. این روش به وسیله‌ی لو پانتریاگین و همکارانش در شوروی و نیز ریچارد بلمن در آمریکا ابداع شد. درواقع نظریه کنترل زمینه‌ای است، که راه‌های کنترل و تحت سیطره درآوردن عملکرد سیستم‌های دینامیکی را مورد مطالعه قرار می‌دهد. هرچند، تولد و پیدایش نظریه کنترل بهینه از مهندسی و ریاضیات آغاز شد اما به‌مرور، نظریه‌ی کنترل راه خود را به سمت کاربردهای نوین در عرصه‌های پیچیده‌تری مثل علوم اجتماعی و در زمینه‌هایی از آن نظیر روان‌شناسی و جامعه‌شناسی و بخصوص در زمینه ریاضیات مالی و اقتصاد پیدا کرده‌است. به‌طورخلاصه می‌توان گفت، منظور از کنترل یک پدیده، دخالت در رفتار آن به‌گونه‌ای است که، نتایج مطلوب حاصل گردد. مساله کنترل بهینه در کتاب‌ها [۲-۱۱] و مقاله‌های مختلف [۱۲-۱۴] معرفی و مورد بررسی قرار گرفته است.

به طور کلی، برای حل مسائل کنترل بهینه راه‌های متفاوتی وجود دارد. نظریه کنترل بهینه در پی یافتن قانون کنترل برای یک سیستم معین است به شکلی که ضابطه بهینگی خاصی به دست آید. در واقع، کنترل بهینه مسیری از متغیرهای کنترل که تابع هزینه را کمینه می‌کند، نشان می‌دهند. کنترل بهینه می‌تواند با استفاده از اصل ماکزیمم پانتریاگین، یا حل معادله همیلتن-ژاکوبی-بلمن به دست آید. اصل ماکزیمم پانتریاگین شرایط لازم برای یک اکسترمم را بیان می‌کند، که اغلب برای حل مسائل با کنترل‌های حلقه باز مورد استفاده قرار می‌گیرد[۱۰]. روش دیگری که با یک دیدگاه متفاوت عمل می‌کند و در این مقاله ما از این روش استفاده خواهیم کرد به روش برنامه‌ریزی پویا معروف است که برای اولین بار در سال ۱۹۵۰ توسط بلمن معرفی گردیده است[۲].

اصل بهینگی بلمن بیان می‌کند که در یک مسیر بهینه، متغیرهای حالت و تصمیم اولیه هر چه باشند، همه تصمیمات باقی مانده باز یک روش بهینه را در خصوص حالتی که از تصمیم اول نتیجه می‌شود را تشکیل خواهند داد. اصل بهینگی وابسته به مفهوم زیرساخت بهینه است، و مسائلی که زیرساخت بهینه در آنها دخیل است را اغلب می‌توان به روش برنامه‌ریزی پویا حل کرد. در این روش مقدار بهینه عملکرد سیستم را به عنوان تابعی از وضعیت اولیه سیستم در نظر گرفته و آن را تابع ارزش می‌نامند. این روش برای حل مسائل کنترل قطعی و تصادفی مورد استفاده قرار می‌گیرد. تابع ارزش $V(.)$ جوابی از معادله دیفرانسیل جزئی که معادله همیلتن-ژاکوبی-بلمن نامیده می‌شود، می‌باشد. هدف حل این معادله تابعی به دست آوردن تابع ارزش $V(.)$ و در نتیجه کنترل بهینه است.

در حالت کلی حل معادلات تابعی چندان ساده نمی‌باشد و هیچ روش کلی برای حل این دسته از معادلات وجود ندارد[۱۵]، با این وجود می‌توان با برخی روش‌های ذکر شده معادله همیلتن-ژاکوبی-بلمن را حل کرد. روش اول، روش ضرایب نامعین یا همان روش حدس و بررسی می‌باشد. در این روش ابتدا تابع ارزش به صورت یک تابع با ضرایب مجهول در نظر گرفته می‌شود. این تابع پیشنهادی در معادله تابعی جایگذاری شده و با حل دستگاه معادلات دیفرانسیل حاصل، ضرایب مجهول به دست می‌آید. حدس تابع ارزش، کار چندان ساده‌ای نمی‌باشد و جز در موارد خاص مورد استفاده قرار نمی‌گیرد. روش دوم استقرای وارونه است، که خود یک روش عددی قابل استفاده در بسیاری از مسائل است. حل مساله با این روش، در مواردی که تعداد متغیرهای حالت زیاد شود، با توجه به مشکل بعد پذیری، غیر ممکن می شود. با این وجود این روش، در مسائل گسسته زمان و متناهی وضعیت روش بسیار کارآمدی می‌باشد. به علت عدم وجود روش تحلیلی معینی برای حل معادله همیلتن-ژاکوبی-بلمن، استفاده از روش‌های عددی منطقی به نظر می‌رسد، ارائه روشهای عددی برای حل این دسته مسائل یک زمینه فعال تحقیقاتی را برای محققان علوم ریاضی فراهم نموده و باعث ظهور روش‌ها و الگوریتم‌های مختلف برای حل این دسته مسائل شده است (برای نمونه مراجع [۱۶] تا [۲۰] را ببینید.) یکی از این روش ها، روش تقریب با تفاضلات متناهی است. در این روش، مشتقات جزیی معادله همیلتن-ژاکوبی-بلمن با استفاده از تفاضلات متناهی تقریب زده می‌شود. بدین ترتیب این معادله به یک معادله تفاضلی تبدیل می‌شود و می‌توان روش‌های موجود برای حل معادلات تفاضلی را برای محاسبه متغیرهای حالت و کنترلی مساله را بکار برد. اگر فضای وضعیت متناهی باشد می‌توان با استفاده از زنجیر مارکوف تقریبی و ماتریس احتمالات انتقال زنجیر مارکوف، معادله برنامه‌ریزی پویا را به فرم برداری نوشت. در این صورت می‌توان تابع ارزش را با استفاده از روش‌های تکراری برای حل دستگاه معادلات، مانند ژاکوبی یا گوس‌سایدل، به دست آورد[۲۱]. نظریه شناسایی و تخمین پارامترها و متغیرهای حالت یک سیستم دینامیکی و کنترل تصادفی از مباحث مهم و پرکاربرد در سیستم های صنعتی می‌باشند[۱]، خالوزاده و محمدیان در این مقاله موضوعات انتساب کواریانس خاص به متغیرهای حالت یک سیستم دینامیکی، سیستم‌های کوانتومی، مکان‌یابی، شناسایی و کنترل تصادفی ربات‌ها، کاربرد نظریه تخمین در سیستم‌های ناوبری، تخمین، شناسایی و کنترل تصادفی موتورهای الکتریکی و تخمین ترافیک در سیستم‌های حمل و نقل هوشمند را بررسی نموده‌اند.

۲- مساله کنترل بهینه تصادفی و معادله هامیلتون-ژاکوبی-بلمن

فرض کنید معادله وضعیت سیستم در یک مدل پخش، به صورت معادله‌ی دیفرانسیل تصادفی زیر باشد:

$$dx(t) = b(x(t),u(t))dt + \sigma(x(t))d\omega(t) \qquad (۱)$$

و تابع هزینه مساله کنترل بهینه تصادفی پیوسته زمان به فرم زیر بیان شده باشد:

$$W(x,u) = E_x^u\{\int_0^\tau k(x(t),u(t))dt + g(x(\tau))\} \qquad (۲)$$

که در آن $x(.)$ و $u(.)$ به ترتیب نشان‌دهنده متغیر وضعیت و کنترل می‌باشند. متغیر وضعیت $x(.)$ روی فضای فشرده G با درون G° تعریف می‌شود. متغیر کنترل $u(.)$ یک فرایند اندازه‌پذیر نسبت به فیلتر تولید شده توسط حرکت براونی $\omega(.)$ است که مقادیرش را از یک مجموعه فشرده U اختیار می‌کند. بعلاوه توابع پیوسته و کراندار $k(.,.)$ و $g(.)$ به ترتیب هزینه جاری (عملکرد) و هزینه توقف (مرزی) نامیده می‌شوند. هم‌چنین $\tau = \inf\{t : x(t) \notin G^\circ\}$ اولین زمان خروج فرایند $x(.)$ از درون ناحیه G باشد.

آنچنان که قبلاً نیز اشاره شد، برای حل مساله کنترل بهینه روش‌های متفاوتی وجود دارد. یکی از این روش‌ها، روش برنامه‌ریزی پویا است. در این روش مقدار بهینه‌ی عملکرد به عنوان تابعی از شرط اولیه در نظر گرفته شده و تابع ارزش نامیده می‌شود. به عبارتی $V(x) = \inf_{u(.)} W(x,u)$ و هدف پیدا کردن تابع ارزش $V(.)$ یعنی تابع مینیمم کننده هزینه و فرایند کنترل $u^*(.)$ به گونه‌ای است که تابع هزینه مقدار مینیمم خود را اختیار کند. از آنجا که روش برنامه‌ریزی پویا برای حل مساله کنترل بهینه تصادفی بر اصل بهینگی بلمن استوار است. در قضیه بعد، این اصل بیان شده‌است[۶و۲۲].

قضیه ۲-۱(مرجع [۲۲] را ببینید): فرایند کنترل $u^*(.)$ یک جواب بهینه برای مساله کنترل بهینه تصادفی است، اگر تابع پیوسته مشق‌پذیر $V(t,x):[0,\tau]\times R^k \to R$ که به‌صورت زیر تعریف می‌شود، وجود داشته باشد:

$$\begin{cases} \min\limits_u \{k(x,u) + V_x(t,x)b(x,u)\} \\ \qquad = -V_t(t,x) - \dfrac{1}{2}\sum\limits_{i,j}^{m} a^{ij}(t,x)V_{x^i x^j}(t,x), & x \in G^\circ, \\ g(x), & x \notin G^\circ, \end{cases}$$

یافتن مقدار دقیق تابع ارزش $V(.)$ و کنترل بهینه $u^*(.)$ که در معادله دیفرانسیل جزیی فوق صدق کند کار چندان ساده‌ای نیست. در ادامه یک روش عددی معرفی می‌شود که با استفاده از آن تابع ارزش و کنترل بهینه تقریب زده می‌شود.

۳- تقریب مساله کنترل بهینه تصادفی با زنجیر مارکوف

در این روش، ابتدا زنجیر مارکوف کنترلی پارامتر گسسته را به صورت زیر تعریف می‌کنیم:

$$\{\xi_n^h, n < \infty, h > 0\} \qquad (۳)$$

زنجیر مارکوف (۳) روی فضای وضعیت گسسته $S_h = \{0, \pm h, \pm 2h, ...\}$ با احتمالات انتقال $p^h(x,y|\alpha)$ تعریف می‌شود. احتمال انتقال $p^h(x,y|\alpha)$ تابعی از مقادیر کنترل $\alpha \in U$ است و احتمال انتقال زنجیر، از وضعیت فعلی x به وضعیت y تابعی از مقدار کنترل α در زمان فعلی است. برای این‌که زنجیر مارکوف (۳) تقریب مناسبی از فرایند پخش (۳) باشد باید تغییرات از وضعیت زنجیر مارکوف $\Delta\xi_n^h$ در دو ویژگی سازگاری موضعی زیر صدق کند[۲۱]:

$$\begin{aligned} &E_{x,n}^{h,\alpha}\{\Delta\xi_n^h\} \equiv b_h(x,\alpha)\Delta t(\xi_n^h, u_n^h) \\ &= b(x,\alpha)\Delta t(x,\alpha) + o(\Delta t^h(x,\alpha)), \\ &E_{x,n}^{h,\alpha}\{[\Delta\xi_n^h - E_{x,n}^{h,\alpha}\Delta\xi_n^h][\Delta\xi_n^h - E_{x,n}^{h,\alpha}\Delta\xi_n^h]'\} \\ &\equiv a_h(x,\alpha)\Delta t(\xi_n^h, u_n^h) + o(\Delta t^h(x,\alpha)), \end{aligned} \qquad (۴)$$

که درآن $a(x) = \sigma(x)\sigma'(x)$ و باید:

$$\sup|\xi_{n+1}^h - \xi_n^h| \to 0, \qquad h \to 0. \qquad (۵)$$

همچنین u_n^h نشان‌دهنده یک متغیر تصادفی کنترل برای زنجیر زمان گسسته در زمان n، $\Delta t_n^h = \Delta t(\xi_n^h, u_n^h)$ نشان‌دهنده طول بازه زمانی است که زنجیر از وضعیت ξ_n و با کنترل u_n^h به وضعیت بعدی در گام $n+1$ انتقال می‌یابد. G_h° نشان‌دهنده اشتراک از مولفه‌های فضای وضعیت گسسته زنجیر S_h با دورن فضای وضعیت G° از فرایند پخش $x(.)$ است، یعنی $G_h^\circ = S_h \cap G^\circ$. بعلاوه تفاضلات $\Delta\xi_n^h = \xi_{n+1}^h - \xi_n^h$ نشان‌دهنده تغییرات وضعیت از زنجیر مارکوف از گام n به گام $n+1$ می‌باشد. همچنین $E_{x,n}^{h,\alpha}$ نشان‌دهنده امید ریاضی شرطی نسبت به شرط $\{\xi_i^h, u_i^h, i \le n, \xi_n^h = x, u_n^h = \alpha\}$ است. در ادامه، با استفاده از تقریبات تفاضلات متناهی، احتمالات انتقال زنجیر مارکوف را به گونه‌ای خواهیم یافت که زنجیر مارکوف در شرایط (۴) و (۵) صدق کند. به عبارتی باید امید ریاضی و کوواریانس شرطی تغییرات وضعیت زنجیر مارکوف با میانگین و کوواریانس موضعی فرایند پخش اولیه متناسب باشند. با استفاده از تقریبات تفاضلات متناهی می‌توان احتمالات انتقال و بازه‌های زمانی تغییر وضعیت زنجیر مارکوف را به گونه‌ای تعیین کرد که زنجیر مارکوف در دو ویژگی سازگاری موضعی فوق صدق کند. ذکر این نکته ضروری است که یک دنباله از سیاست‌های کنترل $u^h = \{u_n^h, n < \infty\}$، یک کنترل مجاز برای زنجیر مارکوف به شمار می‌رود در صورتی‌که زنجیر مارکوف با احتمال انتقال شرطی نسبت به کنترل u_n^h، همچنان در شرط ویژگی مارکوف $P\{\xi_{n+1}^h = y | \xi_i^h, u_i^h, i \le n\} = P^h(\xi_n^h, y | u_n^h)$ صدق کند.

ویژگی مارکوف

$P\{\xi^h_{n+1}=y\,|\,\xi^h_i,u^h_i,i\le n\}=P^h(\xi^h_n,y\,|\,u^h_n)$ صدق کند.

۴- درونیابی پیوسته زمان، زنجیر مارکوف گسسته زمان

زنجیر مارکوف تقریبی معرفی شده در (۳) یک فرایند پارامتری گسسته زمان است. برای اینکه فرایند پخش پیوسته زمان $x(.)$ از مسئله اصلی با این زنجیر مارکوف تقریب زده شود، به یک درونیابی پیوسته زمان از زنجیر مارکوف گسسته نیاز است. فرض کنید $\{\xi^h_n,n<\infty\}$ زنجیر مارکوف تقریبی گسسته زمان و $\{u^h_n,n<\infty\}$ یک کنترل مجاز برای زنجیر و $t^h_n=\sum_{i=0}^{n-1}\Delta t^h_i$. درونیابی پارامتری پیوسته زمان از $\xi^h(.)$ و $u^h(.)$ را برای هر $t\in[t^h_n,t^h_{n+1}]$ به صورت زیر تعریف می‌شود:

$$\xi^h(t)=\xi^h_n,\qquad u^h(t)=u^h_n\qquad t\in[t^h_n,t^h_{n+1}).\qquad (۶)$$

در این صورت فرایندهای درونیابی شده تکه‌ای ثابت تعریف شده در (۶) یک تقریب از فرایند پخش پیوسته زمان $x(.)$ از معادله‌ی (۱) است. همچنین، این فرایندها در ویژگی‌های سازگاری موضعی (۴) و (۵) صدق می‌کند. اگرچه بازه‌های درونیابی $\Delta t^h(x,\alpha)$ را می‌توان همواره برابر با مقدار ثابتی انتخاب کرد ولی در بعضی مواقع با محدود کردن بازه‌های درونیابی می‌توان سرعت همگرایی را افزایش داد. برای مثال اگر سرعت نوسانات موضعی $b(.,\alpha)$ در مقایسه با x بزرگ باشد آن‌گاه می‌توان با در نظر گرفتن یک بازه درونیابی کوچک‌تر، سرعت همگرایی روش عددی را افزایش داد. بدین منظور شرط انعطاف پذیری بازه‌های درونیابی متغیر در نظر گرفته می‌شود. درادامه با استفاده از تقریبات تفاضلات متناهی بازه‌های درونیابی مناسب برای زنجیر را به دست می آوریم. پس از آن با استفاده از زنجیر مارکوف تقریبی پارامتری پیوسته زمان، تابع هزینه $W(x,u)$ تقریب زده می‌شود. بدین منظور، ابتدا فرض می‌کنیم N_h اولین زمان خروج زنجیر مارکوف $\{\xi^h_n,n<\infty\}$ از G°_h باشد .پس تابع هزینه تقریبی متناظر با زنجیر مارکوف تولید شده به صورت زیر تعریف می‌شود:

$$W^h(x,u^h)=$$
$$E^{u^h}_x\{\sum_{n=0}^{N_h-1}\int_{t^h_n}^{t^h_{n+1}}K(\xi^h(t),u^h(t))dt+g(\xi^h_{N_h})\}.$$

با توجه به تکه‌ای ثابت بودن فرایندهای $u^h(t),\xi^h(t)$ روی بازه‌های درونیابی Δt^h می‌توان مقدار هزینه جاری در طول بازه درونیابی را تابعی از وضعیت اولیه زنجیر در نظر گرفت. به عبارتی تابع هزینه کلی تقریبی را می‌توان به فرم ساده‌تر زیر نوشت:

$$W^h(x,u^h)=$$
$$E^{u^h}_x\{\sum_{n=0}^{N_h-1}K(\xi^h_n(t),u^h_n(t))\Delta t^h_n+g(\xi^h_{N_h})\}.\qquad (۷)$$

تابع ارزش پارامتری وابسته به وضعیت اولیه زنجیر، متناظر با مقدارمینیمم هزینه کلی در بین تمام کنترل‌های مجاز به صورت زیر بیان می‌شوند:

$$V^h(x)=\inf_u W^h(x,u).\qquad (۷)$$

با استفاده از روش برنامه‌ریزی پویا و اصل بهینگی بلمن می‌توان معادله برنامه‌ریزی پویا را به صورت زیر نوشت که تابع ارزش در آن صدق می‌کند.

$$V^h(x)=\min_{\alpha\in U}\{K(x,\alpha)\Delta t+E^\alpha_x V^h(\xi_1)\}$$

با داشتن احتمالات انتقال و بازه‌های درونیابی می‌توان معادله برنامه‌ریزی پویا $V^h(x)$ را به فرم ساده‌تر زیر تبدیل نمود:

$$\begin{cases}\min\limits_{\alpha\in U}\{K(x,\alpha)\Delta t^h(x,\alpha)\\ \qquad\qquad +\sum\limits_y P^h(x,y\,|\,\alpha)V^h(y)\}, & x\in G^\circ_h,\\ g(x), & x\notin G^\circ_h,\end{cases}\qquad (۸)$$

با توجه به تشابه تابع هزینه تقریبی (۷) با تابع هزینه اولیه (۲)، همچنین تشابه ویژگی‌های موضعی درونیابی $\xi^h(.)$ به فرایند پخش اولیه $x(.)$، انتظار می‌رود که تابع ارزش $V^h(x)$ برای مقادیر h به اندازه کافی کوچک، تقریب مناسبی برای تابع ارزش $V(x)$ مساله اصلی باشد. این موضوع در ادامه ثابت خواهد شد. از طرفی تحت شرایط مناسب، هر دنباله $\xi^h(.)$ در یک شرایط مناسب دارای زیردنباله‌ای همگرا به فرایند پخش کنترل شده (۱) است. در این قسمت، فرض می کنیم که دنباله $\xi^h(.)$ همگرا به یک حد فرایند پخش $x(.)$ با کنترل مجاز $u(.)$ باشد. تحت شرایط مناسب، با میل کردن پارامتر h به صفر، دنباله τ_h از زمان‌هایی که زنجیرها برای اولین بار از G°_h خارج می شوند به زمانی که حد فرایند پخش $x(.)$ برای اولین بار از G° خارج می‌شود، همگرا می‌باشد. دنباله‌ی توابع ارزش $V^h(x)$ برای دنباله‌ای از زنجیرهای $\xi^h(.)$ همگرا به تابع ارزش $V(x)$ برای فرایند پخش $x(.)$ می‌باشد. برای اثبات همگرایی، نیاز داریم که کلاس کنترل‌های مجاز را به کلاس کنترل‌های ریلکس شده توسیع دهیم، ولی اینفیمم تابع هزینه، روی این دو کلاس با یکدیگر برابر خواهد بود.

۵- روش تفاضلات متناهی برای محاسبه احتمالات انتقال و بازه‌های درونیابی زنجیر مارکوف

روش تقریبات تفاضلات متناهی یک راه مناسب برای محاسبه احتمالات انتقال و بازه‌های درونیابی از زنجیر مارکوف به گونه‌ای است که در شرط ویژگی‌های سازگاری موضعی (۴) و (۵) صدق می‌کند. در ادامه این روش را با استفاده از یک مثال بیان می‌کنیم.

مثال۵-۱ (مرجع [۲۱] را ببینید): مساله کنترل بهینه تصادفی با دینامیک وضعیت (۹) و تابع هزینه (۱۰) روی فضای وضعیت $G=[0,B]$ را در نظر بگیرید:

$$dx=b(x,u(x))dt+\sigma(x)d\omega. \qquad (۹)$$

$$W(x,u)=\begin{cases}E_x^u\{\int_0^\tau k(x(s),u(x(s))dt & x\in(0,B),\\ \qquad +g(x(\tau))\}, & \\ g(x), & x\in\{0,B\},\end{cases} \qquad (۱۰)$$

با استفاده از اصل بهینگی و فرمول ایتو، تابع $W(x,u(x))$ در معادله با شرایط مرزی زیر صدق می‌کند:

$$\begin{cases}L^{u(x)}W(x,u)+K(x,u(x))=0, & x\in(0,B),\\ W(0,u)=g(0), \quad W(B,u)=g(B).\end{cases} \qquad (۱۱)$$

که در آن $L^\alpha=b(x,\alpha)\frac{d}{dx}+\frac{\sigma^2(x)}{2}\frac{d^2}{dx^2}$ یک عملگر دیفرانسیلی از فرایند پخش (۹) است ومتغیر کنترل، مقدار ثابت α را اختیار کرده است. با در نظر گرفتن (۱۱) و عملگر دیفرانسیلی L^α و برای $x\in(0,B)$ داریم:

$$b(x,\alpha)W_x(x,\alpha)+\frac{\sigma^2(x)}{2}W_{xx}(x,\alpha)+K(x,\alpha)=0, \qquad (۱۲)$$

که شرایط مرزی آن به‌صورت $W(0,u)=g(0)$ و $W(B,u)=g(B)$ می‌باشد. تقریبات تفاضلات متناهی به صورت زیر تعریف می‌شود:

$$W_x(x,\alpha)=\frac{W^h(x+h,\alpha)-W^h(x,\alpha)}{h}, \qquad b(x,\alpha)\ge 0,$$
$$W_x(x,\alpha)=\frac{W^h(x,\alpha)-W^h(x-h,\alpha)}{h}, \qquad b(x,\alpha)<0,$$
$$W_{xx}(x,\alpha)=\frac{W^h(x+h,\alpha)-2W^h(x,\alpha)+W^h(x-h,\alpha)}{h^2}, \qquad (۱۳)$$

پس از جایگذاری تقریبات تفاضلات متناهی (۱۳) در معادله (۱۲) و مرتب کردن ضرایب و تقسیم طرفین بر ضریب $W^h(x,\alpha)$ و قرار دادن $|b(x,\alpha)|=b^+(x,\alpha)+b^-(x,\alpha)$ داریم:

$$W^h(x,\alpha)=\frac{hb^+(x,\alpha)+\frac{\sigma^2(x)}{2}}{h|b(x,\alpha)|+\sigma^2(x)}W^h(x+h,\alpha)$$
$$+\frac{hb^-(x,\alpha)+\frac{\sigma^2(x)}{2}}{h|b(x,\alpha)|+\sigma^2(x)}W^h(x-h,\alpha)$$
$$+\frac{h^2}{h|b(x,\alpha)|+\sigma^2(x)}K(x,\alpha). \qquad (۱۴)$$

با مقایسه معادله (۱۴) و رابطه‌ی بازگشتی تابع هزینه

$$W^h(x,u)=$$
$$K(x,u(x))\Delta t^h(x,u(x))+\sum_{y\neq x}P^h(x,y\,|\,u(x))W^h(y,u)$$

در می‌یابیم که:

$$P^h(x,x+h\,|\,\alpha)=\frac{hb^+(x,\alpha)+\frac{\sigma^2(x)}{2}}{h|b(x,\alpha)|+\sigma^2(x)}, \qquad (۱۵)$$

$$P^h(x,x-h\,|\,\alpha)=\frac{hb^-(x,\alpha)+\frac{\sigma^2(x)}{2}}{h|b(x,\alpha)|+\sigma^2(x)}, \qquad (۱۶)$$

$$\Delta t^h(x,\alpha)=\frac{h^2}{h|b(x,\alpha)|+\sigma^2(x)}. \qquad (۱۷)$$

برای هر $y\neq x\pm h$ داریم $P^h(x,y\,|\,\alpha)=0$. معادلات (۱۵) و (۱۶) نشان‌دهنده احتمالات انتقال و معادله (۱۷) بازه درونیابی از زنجیر مارکوف تقریبی مورد نظر می‌باشد. حال با داشتن احتمالات انتقال و بازه درونیابی می‌توان نشان داد که تغییرات وضعیت از زنجیر مارکوف تقریبی با این احتمالات انتقال در ویژگی سازگاری موضعی نیز صدق می‌کند:

$$E_{x,n}^{h,\alpha}\{\Delta\xi_n^h\}$$
$$=h\{\frac{hb^+(x,\alpha)+\frac{\sigma^2(x)}{2}}{h|b(x,\alpha)|+\sigma^2(x)}\}-h\{\frac{hb^-(x,\alpha)+\frac{\sigma^2(x)}{2}}{h|b(x,\alpha)|+\sigma^2(x)}\}$$
$$=b(x,\alpha)\Delta t^h(x,\alpha)+o(\Delta t^h(x,\alpha)),$$

و

$$E_{x,n}^{h,\alpha}\{[\Delta\xi_n^h-E_{x,n}^{h,\alpha}\Delta\xi_n^h][\Delta\xi_n^h-E_{x,n}^{h,\alpha}\Delta\xi_n^h]'\}$$
$$=h^2\{\frac{hb^+(x,\alpha)+\frac{\sigma^2(x)}{2}}{h|b(x,\alpha)|+\sigma^2(x)}\}$$
$$+(-h)^2\{\frac{hb^-(x,\alpha)+\frac{\sigma^2(x)}{2}}{h|b(x,\alpha)|+\sigma^2(x)}\}$$
$$=a(x,\alpha)\Delta t^h(x,\alpha)+o(\Delta t^h(x,\alpha)),$$

۶- تقریب در روش فضای سیاست

روش محاسباتی برای حل معادله برنامه‌ریزی پویا (۸) روش تقریب در فضای سیاست است، که آن را با جزئیات بیشتر توضیح می‌دهیم. در این روش به صورت متوالی یک دنباله مینیمم کننده از سیاست‌های کنترل فیدبک $\{u_n(.)\}$ برای زنجیر محاسبه می‌شود، به گونه‌ای که کنترل $u_{n+1}(.)$ از یک تقریب جواب برای تابع هزینه با کنترل $u_n(.)$ به دست می‌آید. به عبارتی فرض می‌کنیم که $u_n(.)$ یک کنترل بهینه تجربی (آزمایشی) برای زنجیر و تابع هزینه داده شده باشد و هدف

محاسبه کنترل $u_{n+1}(.)$ با استفاده از این روش تکراری است. با استفاده از این روش، می‌توان تقریبی از کنترل بهینه اصلی را با میل دادن n به بینهایت $(n \to \infty)$ به دست آورد. توجه داشته باشید که با استفاده از روش برنامه‌ریزی پویا می‌توان رابطه بازگشتی زیر را برای تابع هزینه $W(x,u)$ نوشت، اگر احتمالات انتقال و بازه‌های درونیابی زنجیر را به دست آورده باشیم:

$$W^h(x,u) = K(x,u(x))\Delta t^h(x,u(x)) + \sum_{y\neq x} P^h(x,y \mid u(x))W^h(y,u) \quad x \in G_h^\circ, \tag{۱۸}$$

از طرفی، طبق رابطه‌ی بازگشتی (۱۸) یک مقدار تقریبی برای تابع هزینه $W^h(x,.)$ برای کنترل u_n (یعنی $W^h(x,u_n)$) به صورت زیر محاسبه می‌شود:

$$\begin{cases} K(x,u_n(x))\Delta t^h(x,u_n(x)) + \sum_{y\neq x} P^h(x,y \mid u_n(x))W^h(y,u_n), & x \in G_h^\circ, \\ g(x), & x \notin G_h^\circ, \end{cases} \tag{۱۹}$$

با استفاده از مقدار تقریبی به‌دست آمده برای تابع هزینه(۱۹) ،کنترل $u_{n+1}(.)$، از رابطه‌ی زیر محاسبه می‌شود:

$$u_{n+1}(x) = \arg\min_{\alpha\in U} \{K(x,\alpha)\Delta t^h(x,\alpha) + \sum_{y\neq x} P^h(x,y \mid \alpha)W^h(y,u_n)\}.$$

با در نظر گرفتن یکسری شرایط اضافی، مشاهده می‌شود که با انتخاب یک کنترل فیدبک مجاز $u_0(.)$ و محاسبه‌ی تابع هزینه متناظر آن $W(u_0)$، برای $n \geq 1$, دنباله‌ی کنترل‌های فیدبک $u_n(.)$ و هزینه های $W(u_n)$ که از این روش بازگشتی محاسبه می‌شوند، به گونه‌ای است که $W(u_n) \to V$.

۷- الگوریتم تکراری برای حل مساله کنترل بهینه تصادفی

با در نظر گرفتن زنجیر مارکوف با احتمالات انتقال و بازه‌های درونیابی بدست آمده در بخش ۵، که در در شرط سازگاری موضعی نیز صدق می‌کند، می‌توان تقریب مناسبی از فرایند پخش $x(.)$ بدست آورد .در نهایت کافی است با استفاده روش تقریب در فضای سیاست که در بخش۶ معرفی شد، کنترل بهینه و تابع ارزش تقریبی از مساله کنترل بهینه تصادفی را به دست آورد. مطالب مطرح شده در قسمت‌های قبل را در قالب الگوریتم زیر خلاصه می‌کنیم.

ورودی: یک $\varepsilon > 0$ ، وضعیت اولیه $x \in G_0^h$ و مقدار دلخواه کنترل $u_0 \in U$.

خروجی : مقدار تقریبی کنترل u و تابع ارزش W.

گام اول: برای $n = 0$ مقدار اولیه‌ی کنترل u_0 را در معادله‌ی زیر جایگذاری کن:

$$W_0^h(x,u_0) = \sum_y p^h(x,y \mid u_0)W^h(y,u_0) + k(x,u_n)\Delta t^h(x,u_0). \tag{۱۹}$$

گام دوم: قرار دهید $n \to n+1$ و u_n را از رابطه‌ی زیر حساب کن:

$$u_n(x) = \arg\min_{\alpha\in U} \{\sum_y p^h(x,y \mid \alpha)W_{n-1}^h(y,u_{n-1}) + k(x,\alpha)\Delta t^h(x,\alpha)\}. \tag{۲۱}$$

و متناظر با آن، حاصل $W_n^h(x,u_n)$ را از رابطه‌ی زیر حساب کن:

$$W_n^h(x,u_n) = \sum_y p^h(x,y \mid u_{n-1})W_{n-1}^h(y,u_{n-1}) + k(x,u_{n-1})\Delta t^h(x,u_{n-1}). \tag{۲۲}$$

گام سوم: اگر $|W_n^h(x,u_n) - W_{n-1}^h(x,u_{n-1})| < \varepsilon$ باشد، توقف نموده و u_n را مقدار تقریبی کنترل و W_n^h را مقدار تقریبی تابع ارزش در نظر بگیر، در غیر این صورت به گام۲ بر گردد.

۸- مثال کاربردی: مدل سرمایه‌گذاری مرتون

در این قسمت یک مثال کاربردی از ریاضیات مالی، مدل سرمایه‌گذاری مرتون، را بیان و حل می‌نماییم. در این مدل فرض بر این است که سرمایه‌گذار نسبت $u(t)$ از ثروت X خود را بر روی سهام با تغییرات قیمت $p(t)$ تحت معادله‌ی دیفرانسیل تصادفی زیر سرمایه‌گذاری می‌کند:

$$dp(t) = p(t)(bdt + \sigma dw(t)),$$

نرخ بازدهی برای دارایی با ریسک، b بوده و قیمت هر سهم از این نوع دارایی طبق معادله‌ی دیفرانسیل تصادفی فوق تغییر می‌کند. هم‌چنین، سرمایه‌گذاری نسبت $1-u(t)$ از ثروت خود را بر روی دارایی بدون ریسک با تغییرات قیمت $q(t)$ تحت معادله‌ی زیر یعنی با نرخ بهره‌ی ثابت r سرمایه‌گذاری می‌کند:

$$dq(t) = rq(t)dt,$$

در این تعاریف، w حرکت برآونی استاندارد یک بعدی است و فرایند پخش b و نوسان‌پذیری σ پیوسته و دارای شرایط $b > r$ و $\sigma > 0$ می‌باشند. با این شرایط طبق مدل بلک-شولز معادله‌ی مربوط به ثروت سرمایه‌گذار به صورت زیر می‌باشد:

$$dX = (r + (b-r)u(t))Xdt + Xu(t)\sigma dw(t)$$

در این مدل مطلوبیت سرمایه‌گذار به‌صورت تابع صعودی و مقعر F می‌باشد. در واقع مطلوبیت انتخاب سهام، با امید ریاضی ثروت در زمان

پایانی سنجیده می‌شود. به عبارت دیگر به‌دنبال بیشینه نمودن عملکردی به‌صورت زیر می‌باشیم:

$$J(t,x;u)=E_{tx}\{F(X(t_1))\}.$$

در این‌جا تابع مطلوبیت $F(x)=\frac{1}{\gamma}x^{\gamma}$ که $0<\gamma<1$ درنظر گرفته شده‌است. کنترل بهینه به‌صورت $u^*=\frac{(r-b)V_X}{x\sigma^2V_{XX}}$ و معادله‌ی همیلتون-ژاکوبی-بلمن برای این مدل به‌صورت زیر به‌دست می‌آید:

$$\begin{cases} V_t-\frac{(r-b)^2V_X^2}{2\sigma^2V_{XX}}+xrV_X=0, \\ V(t_1,X)=\frac{1}{\gamma}X^{\gamma}. \end{cases}$$

جواب دقیق این معادله برای تابع ارزش $V(t,x)$ به‌صورت زیر می‌باشد:

$$V(t,X)=e^{\rho(t_1-t)}X^{\gamma},\ \rho=\frac{(b-r)^2}{2\sigma^2}\frac{\gamma}{1-\gamma}+r\gamma,$$

پس از حل معادله‌ی همیلتون-ژاکوبی-بلمن و به‌دست آوردن $V(t,x)$ مقدار دقیق کنترل بهینه به‌صورت زیر به‌دست خواهد آمد:

$$u^*=\frac{b-r}{\sigma^2(1-\gamma)}.$$

وجود جواب دقیق برای تابع ارزش و کنترل بهینه، امکان مقایسه جواب‌های تقریبی به‌دست آمده با جواب دقیق را فراهم می‌نماید.

روش به‌کار رفته در مرجع [۲۰]، رابطه‌ی بازگشتی زیر را برای یافتن مقدار تقریبی شاخص عملکرد نتیجه می‌دهد و مقدار عددی این شاخص از تساوی $\hat{J}=V_n(0,1)$ حاصل می‌شود:

$$\begin{cases} V_0(t,X)=\frac{1}{\gamma}X^{\gamma}, \\ V_{n+1}(t,X)=V_n(t,X)+\int_t^{t_1}\Bigg(\frac{\partial V_n(\xi,X)}{\partial\xi} \\ \qquad -\frac{(r-b)^2(\frac{\partial V_n(\xi,X)}{\partial X})^2}{2\sigma^2\frac{\partial^2V_n(\xi,X)}{\partial X^2}}+rX\frac{\partial V_n(\xi,X)}{\partial X}\Bigg)\partial\xi \end{cases}$$

که با درنظرگرفتن مقدار عددی متغیرهای $b=0.1$، $r=0.055$، $\sigma=0.45$ و $\gamma=\frac{1}{2}$ و متناظر با $n=2$ نتایج تقریبی گزارش شده در جدول‌های ۱-۵ به‌دست می‌آید.

برای حل این مثال با الگوریتم پیشنهادی، پارامترهای گسسته سازی حالت $h=0.2$ و زمان $\delta=0.0001$ همچنین کنترل $u_0=1$ را به دلخواه انتخاب و نتایج را در جدول‌های ۱-۵ خلاصه می‌کنیم:

جدول ۱: نتایج متناظر با $t=0.0$ مقادیر مختلف متغیرحالت X در مقایسه با نتایج مرجع [۲۰] و جواب دقیق.

X	جواب دقیق	نتایج مرجع[۲۰]	الگوریتم پیشنهادی
0	0.0	0.0	0.0
0.2	0.923973600	0.92396844 3	0.89463098 7
0.4	1.30669600	1.30668870	1.26522382
0.6	1.60036922	1.60036029	1.54958883
0.8	1.84794720	1.84793689	1.78931242
1.0	2.06606778	2.06605625	1.99450748

جدول ۲: نتایج متناظر با $t=0.25$ مقادیر مختلف متغیرحالت X در مقایسه با نتایج مرجع [۲۰] و جواب دقیق.

X	جواب دقیق	نتایج مرجع[۲۰]	الگوریتم پیشنهادی
0	0.0	0.0	0.0
0.2	0.91649673 6	0.916494560	0.894580032
0.4	1.29612211	1.29611904	1.26514562
0.6	1.58741892	1.58741514	1.54948995
0.8	1.83299347	1.83298912	1.78919914
1.0	2.04934900	2.04934414	1.99587919

جدول ۳: نتایج متناظر با $t=0.5$ مقادیر مختلف متغیرحالت X در مقایسه با نتایج مرجع [۲۰] و جواب دقیق.

X	جواب دقیق	نتایج مرجع[۲۰]	الگوریتم پیشنهادی
0	0.0	0.0	0.0
0.2	0.909080368	0.909079723	0.894529081
0.4	1.28563379	1.28563287	1.26506743
0.6	1.57457339	1.57457227	1.54939107
0.8	1.81816074	1.83298912	1.78908504
1.0	2.03276550	2.03276406	1.99725185

جدول ۴: نتایج متناظر با $t=0.75$ مقادیر مختلف متغیرحالت X در مقایسه با نتایج مرجع [۲۰] و جواب دقیق.

X	جواب دقیق	نتایج مرجع[۲۰]	الگوریتم پیشنهادی
0	0.0	0.0	0.0
0.2	0.90172401 6	0.901723936	0.894478134
0.4	1.27523033	1.27523022	1.26498924
0.6	1.56183181	1.56183167	1.54929220
0.8	1.80344803	1.80344787	1.78897012
1.0	2.01631620	2.01631602	1.99862545

جدول ۵: نتایج متناظر با $t = 1.0$ مقادیر مختلف متغیرحالت X در مقایسه با نتایج مرجع [۵] و جواب دقیق.

X	جواب دقیق	نتایج مرجع[۲۰]	الگوریتم پیشنهادی
0	0.0	0.0	0.0
0.2	0.894427190	0.894427190	0.894427191
0.4	1.26491106	1.26491106	1.26491106
0.6	1.54919334	1.54919334	1.54919334
0.8	1.78885438	1.78885438	1.78885438
1.0	2.0	2.0	2.0

از آن‌جا که هدف این مساله مینیمم سازی تابع هزینه است، پیش‌بینی می‌شود که با کوچک‌تر شدن گام‌های گسسته سازی مقدار تابع ارزش نیز کاهش می‌یابد، همچنین نتایج حاصل از کنترل $u_0 = 1$ همگرایی سریعی را نتیجه می‌دهد، شایان ذکراست که مقدار دقیق کنترل در این مثال برابر با $u^* = 0.\overline{4}$ می‌باشد.

۹- نتیجه‌گیری

هدف این مقاله، ارائه یک روش عددی برای حل مساله کنترل بهینه تصادفی با استفاده از زنجیرهای مارکوف تقریبی است. مساله کنترل تصادفی را می‌توان فرایند تاثیرگذاری روی رفتار یک سیستم دینامیک تصادفی، برای رسیدن به هدف خاصی درنظر گرفت .اگر هدف بهینه سازی معیار عملکردی که به این کنترل و وضعیت سیستم وابسته است باشد، مساله موردنظر کنترل بهینه تصادفی نامیده می‌شود. فرض کنید مدل تصادفی دینامیک وضعیت سیستم، فرایند پخش و هدف کنترل کننده، یافتن تابع کنترل بهینه و تابع ارزش(مقدار بهینه عملکرد سیستم) باشد به گونه‌ای که تابع هزینه را مینیمم نماید. متداول‌ترین روش برای حل مساله کنترل بهینه تصادفی، روش برنامه‌ریزی پویا است. این روش به حل معادله دیفرانسیل جزئی همیلتن-ژاکوبی-بلمن منجر می‌شود، که تابع ارزش جوابی از آن است. شایان ذکراست که، در حالت کلی حل این معادله‌ی تابعی یعنی بدست آوردن جواب تحلیلی آن کار چندان ساده‌ای نیست. به همین دلیل روش‌های عددی که بتوان با استفاده از آن‌ها تابع ارزش و کنترل بهینه را با دقت خوبی تقریب زد، اهمیت ویژه‌ای دارند. ایده کلی روش تقریب با زنجیره مارکوف بدین صورت است که ابتدا فرایند پخش کنترل شده مربوط به مساله کنترل اولیه، با یک زنجیر مارکوف کنترل شده، روی یک فضای وضعیت متناهی تقریب زده می شود. سپس تقریبی از تابع هزینه مربوط به مساله اولیه با استفاده از زنجیر مارکوف تقریبی، بدست می‌آید. باید در نظر داشت که این تقریبات باید به گونه‌ای در نظر گرفته شوند تا در ویژگی سازگاری موضعی صدق کنند. به عبارت دیگر باید میانگین و واریانس شرط تغییرات وضعیت زنجیر مارکوف با میانگین و کوواریانس موضعی فرایند پخش اولیه متناسب باشند. روشی که با استفاده از آن احتمالات انتقال زنجیر مارکوف را بدست می‌آوریم، روش تفاضلات متناهی است. جواب‌های این روش در دو ویژگی سازگاری موضعی صدق می‌کنند. حال با در نظر گرفتن یک معادله بازگشتی برای تابع هزینه و انتخاب یک کنترل مجاز دلخواه اولیه، می‌توان با استفاده از روش تقریب در فضای سیاست، دنباله ای از مقادیر تقریبی توابع هزینه را بدست آورد که تحت این شرایط این دنباله با میل کردن n به بینهایت، به تابع ارزش مساله کنترل بهینه همگرا می‌شوند. در پایان یک الگوریتم تکراری ارائه گردیده و از آن برای حل یک مساله کنترل بهینه تصادفی که کاربردی از ریاضیات مالی می‌باشد، استفاده شده است.

مراجع

[۱] حمید خالوزاده ، عطیه کشاورز محمدیان، مروری بر کاربردهای نظریه تخمین، شناسایی و کنترل تصادفی در سیستم‌های صنعتی، مجله کنترل، جلد ۸، شماره ۳، پاییز۱۳۹۳.

[2] R.E. Bellman, *Dynamic Programming,* Princeton University Press, Princeton N J, 1957.

[3] R.E. Bellman, S.E. Dreyfus *Applied Dynamic Programming,* Princeton University Press, Princeton N J, 1962.

[4] A.E. Bryson, Y.C. Ho, Applied Optimal Control, Hemisphere Publishing Corporation, Washangton D.C, 1975.

[5] W.H. Fleming, C.J. Rishel, Deterministic and Stochastic Optimal Control, Springer-Verlag, New York, NY, 1975.

[6] W.H. Fleming, H.M. Soner, Controlled Markov Processes and Viscosity Solutions, Springer, 2006.

[7] D.E. Kirk, Optimal Control Theory, an Introduction, Prentice-Hall, Englewood Cliffs, 1970.

[8] A.B. Pantelev, A.C. Bortakovski, T.A. Letova, Some issues and examples in optimal control, Moscow, MAI Press, 1996, (In Russian).

[9] E.R. Pinch, Optimal Control and the Calculus of Variations, Oxford University Press, London, 1993.

[10] L.S. Pontryagin, The Mathematical Theory of Optimal Processes, In Interscience, John Wiley and Sons, 1962.

[11] J.E. Rubio, Control and Optimization: The Linear Treatment of Non-linear Problems, Manchester University Press, Manchester, 1986.

[12] M. Athans, The status of optimal control theory and application for deterministic systems, IEEE Transactions on Automatic Control, 11 (1699) 580-596.

[13] A.E. Bryson, Optimal control-1950 to 1985, IEEE Control System Magazine, 16 (1996) 26-33.

[14] H.J. Sussmann, J.C. Willems, 300 years of optimal control: From the Brachystochrone to the

[15] H.M. Jaddu, Numerical Methods for solving optimal control problems using chebyshev polynomials, PhD thesis, School of Information Science, Japan Advanced Institute of Science and Technology, (1998).

[16] H. Saberi Nik, S. Effati, M. Shirazian, An approximate-analytical solution for the Hamilton-Jacobi-Bellman equation via homotopy perturbation method, Appl. Math. Model. 36 (2012) 5614–5623.

[17] B. Kafash, A. Delavarkhalafi, S. M. Karbassi, Application of Chebyshev polynomials to derive efficient algorithms for the solution of optimal control problems, Sci. Iran. D, Comput. Sci. Eng. Electr. Eng. 19 (3) (2012) 795–805.

[18] B. Kafash, A. Delavarkhalafi, S. M. Karbassi, Application of variational iteration method for Hamilton-Jacobi-Bellman equations, Applied Mathematical Modelling (2012), 37 (2013), 3917-3928.

[19] B. Kafash, A. Delavarkhalafi, S. M. Karbassi, Numerical solution of nonlinear optimal control problems based on state parameterizations, Iranian J. Sci. Technol., 36 (A3) (2012), 331-340.

[20] B. Kafash, A. Delavarkhalafi, S. M. Karbassi, A Computational Method for Stochastic Optimal Control Problems in Financial Mathematics, Asian Journal of Control, Vol. 18, No. 4, pp. 1–12, July 2016.

[21] H. J. Kushner, P. Dupuis, Numerical methods for stochastic control problems in continuous time, Springer, New York, 1992.

[22] W.K.Y. David, Leon A. Petrosyan, Cooperative Stochastic Differential Games, Springer, 2005.