



## نهمین کنگره مشترک سیستم‌های فازی و هوشمند ایران

۱۱-۱۳ اسفندماه ۱۴۰۰، مجتمع آموزش عالی بوم

# پیاده سازی یادگیری تقویتی عمیق برای کنترل هوشمند مبدل افزایشده

سبا عسکری نوغانی<sup>۱</sup>، ناصر پریز<sup>۲</sup>، محمد باقر نقیبی سیستانی<sup>۳</sup>

<sup>۱</sup> دانشجوی کارشناسی ارشد، گروه برق، دانشگاه فردوسی، مشهد،  
saba.askari1992@gmail.com

<sup>۲</sup> استاد، گروه برق، دانشگاه فردوسی، مشهد،  
n-pariz@um.ac.ir

<sup>۳</sup> دانشیار، گروه برق، دانشگاه فردوسی، مشهد.  
mb-naghbi@um.ac.ir

### چکیده

طراحی کنترل گر برای مبدل‌های افزایشده DC یکی از چالش‌های مهم در کنترل غیرخطی است. در این مقاله، سعی شده است تا نحوه پیاده‌سازی کنترل گر غیرخطی بر اساس یادگیری تقویتی عمیق برای کنترل مبدل افزایشده DC شرح داده می‌شود. در این پژوهش، روند کنترل مبدل افزایشده DC به عنوان یک محیط برای عامل در نظر گرفته شده است. عامل‌های مورد استفاده در این پژوهش شامل روش PPO و DQN هستند. پس از بیان چگونگی پیاده‌سازی روش‌ها در محیط سیمولینک، برای پیاده‌سازی مناسب روش‌های یادگیری تقویتی عمیق، عملیات بهینه یابی ابرپارامترهای شبکه عصبی عمیق انجام شده است. مقایسه روش‌ها نشان می‌دهد که تکنیک‌های مبتنی بر یادگیری تقویتی عمیق عملکرد مناسب‌تری نسبت به روش PI در کنترل مبدل‌های افزایشده DC دارند. بررسی حساسیت روش‌ها نسبت به نویز سفید نشانگر توانمندی روش PPO در کنترل غیرخطی مبدل افزایشده DC است.

### کلمات کلیدی

مبدل افزایشده، یادگیری تقویتی عمیق، یادگیری ماشین، کنترل غیرخطی، یادگیری تقویتی

### ۱- مقدمه

غیرخطی بودن سیستم مبدل افزایشده عملکرد رضایت بخشی را ارائه نمی‌دهند [۵]. نحوه عملکرد مبدل افزایشده به دو صورت کلید باز و کلید بسته است [۶]. در پژوهش پرادیپ و همکاران [۷] ابتدا مسئله کنترل مبدل افزایشده به عنوان یک فرایند تصمیم‌گیری متوالی بهینه در نظر گرفته شده است. باید توجه داشت که با توجه به رشد روزافزون استفاده از یادگیری تقویتی عمیق، از این روش‌ها در کنترل مسائل متنوعی استفاده شده است. قیصر نژاد و همکاران [۸] روش کنترل مدل فوق محلی را بر اساس یادگیری تقویتی DDPG برای کنترل مبدل کاهنده استفاده کردند. همچنین آن‌ها در پژوهشی دیگر این روش ابداعی را برای کنترل مبدل کاهشی-افزایشی نیز استفاده کردند [۹]. حاجی حسینی و همکاران [۱۰] از روش یادگیری تقویتی PPO برای تنظیم پارامترهای کنترل گر PID در کنترل مبدل کاهشی-افزایشی استفاده کردند. البته باید توجه داشت که در زمینه‌ی کنترل مبدل افزایشده، توجه کمی به استفاده از روش‌های یادگیری عمیق شده است و هنوز بسیاری از روش‌های یادگیری تقویتی عمیق به طور مستقیم برای کنترل مبدل افزایشده مورد آزمون قرار نگرفته‌اند.

طراحی کنترل گر برای مبدل افزایشده یکی از چالش‌های مهم در کنترل غیرخطی محسوب می‌شود. نکته قابل توجه این است که تکنیک‌های کنترل خطی کلاسیک عملکرد ضعیفی در کنترل مبدل‌های افزایشده دارند. از طرفی باید توجه داشت که مسئله کنترل مبدل‌های افزایشده در یک فضای حالت پیوسته اتفاق می‌افتد. از این رو در این پژوهش سعی در پیاده‌سازی روش یادگیری تقویتی عمیق در کنترل مبدل افزایشده می‌گردد. باید توجه داشت که تا به امروز، روشی جامع که توانایی کنترل مبدل افزایشده در یک فضای حالت پیوسته را داشته باشد، پدید نیامده است.

استفاده از روش‌های کنترل خطی مانند کنترل کننده‌های (PID) برای کنترل مبدل افزایشده به طور گسترده‌ای در مقالات گزارش شده است. روش‌های سنتی طراحی کنترل گر با هدف تنظیم و ثابت نگه‌داشتن کنترل کننده PID انجام شده است. تا مبدل افزایشده، ولتاژ خروجی ثابتی داشته باشد [۱-۵]. لازم به ذکر است، تکنیک‌های کنترل خطی توصیف شده در مقالات به دلیل

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} \frac{-r_L + R_{Load} // r_C}{L} & \frac{-R_{Load}}{L(R_{Load} + r_C)} \\ \frac{-R_{Load}}{L(R_{Load} + r_C)} & \frac{-1}{C(R_{Load} + r_C)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \frac{1}{L} \\ 0 \end{bmatrix} V_{in} \quad (3)$$

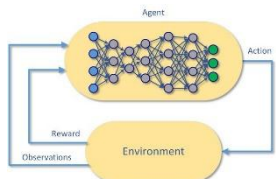
$$V_O(x) = \begin{bmatrix} R_{Load} // r_C & \frac{R_{Load}}{R_{Load} + r_C} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad (4)$$

که در این رابطه‌ها، ولتاژ ورودی،  $V_{in}$  ولتاژ ورودی،  $L$  ضریب خودالقایی،  $r_L$  مقاومت معادل القاگر،  $R_{Load}$  مقاومت در برابر بار،  $C$  ظرفیت خازن و  $r_C$  مقاومت معادل خازن هستند.

روابط ۱ تا ۴ در حقیقت معادلات دیفرانسیل خطی هستند که جداگانه بیان شده‌اند. مدل فضای حالت خطی که توسط رابطه‌های (۱) و (۲) ارائه شده است، سیستم را زمانی که کلید روشن است، توصیف می‌کند. رابطه‌های (۳) و (۴) توصیف‌کننده سیستم به هنگامی بسته بودن کلید هستند. از آنجاکه مدل کلی سیستم به وضعیت المان سوئیچینگ بستگی دارد، سیستم کلی رفتاری غیرخطی دارد. در حقیقت، یک سیستم خطی طبق تعریف، سیستمی است که توسط یک معادله فضای حالت واحد به صورت  $\frac{dx}{dt} = f(x, u) = Ax + Bu$  توصیف می‌شود. در این رابطه تابع  $f$  باید خطی باشد. در مبدل افزایشده، تابع  $f$  با یک معادله خطی واحد توصیف نمی‌شود و در واقع بین دو معادله حالت متفاوت به طور مداوم در حال تغییر است؛ بنابراین سیستم مبدل افزایشده به طور کلی غیرخطی است.

### ۳- یادگیری تقویتی عمیق

یادگیری تقویتی عمیق زیرشاخه‌ای از یادگیری ماشین است که ترکیبی از یادگیری تقویتی و یادگیری عمیق است. در یادگیری تقویتی در واقع یادگیرنده به دنبال هدفی خاص و مشخص هست و با سعی و خطا در رسیدن به هدف تلاش دارد. در حقیقت می‌توان گفت که با استفاده از الگوریتم‌هایی موجب کمک به عامل در یافتن عمل می‌شود. در یادگیری تقویتی عمیق از شبکه‌های عصبی عمیق برای حل مسئله‌های یادگیری تقویتی استفاده می‌شود، به همین دلیل در نام آن از کلمه عمیق استفاده شده است. الگوریتم‌های یادگیری تقویتی عمیق توانایی این را دارد که در فضای پیوسته عمل کنند. همچنین تصمیم بگیرند که چه اقدام‌هایی را برای پیدا کردن جواب بهینه انجام دهند. یادگیری تقویتی عمیق برای مجموعه متنوعی از پژوهش‌ها و تحقیقات مثل روباتیک، بازی‌های ویدئویی، پردازش زبان طبیعی، بینایی رایانه‌ای، آموزش، حمل‌ونقل، امور مالی و مراقبت‌های بهداشتی مورد استفاده قرار گرفته است [۱۳-۱۴]. شکل ۲ نشان‌دهنده روند کار یادگیری تقویتی عمیق است. در ادامه به طور جداگانه، روش‌های یادگیری تقویتی عمیقی که در این پژوهش استفاده شده است، ذکر خواهد شد.



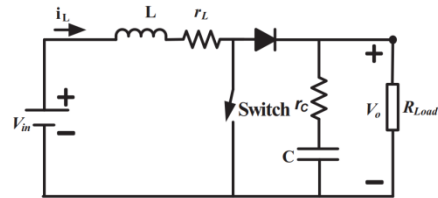
شکل ۲: روند یادگیری تقویتی عمیق [۱۳].

هدف اصلی این پژوهش، پیاده‌سازی یادگیری تقویتی عمیق در کنترل مبدل افزایشده است. به طور کلی می‌توان موارد زیر را به عنوان نوآوری در این پژوهش در نظر گرفت:

- مقایسه عملکرد کنترل خطی PID با نتایج حاصل از کنترل بر اساس یادگیری تقویتی عمیق.
- بررسی عملکرد روش‌های PPO و DQN در کنترل بهینه غیرخطی مبدل‌های افزایشده.
- دستیابی به ابر پارامترهای شبکه عمیق در فرایند یادگیری تقویتی برای یافتن بهترین مشخصات شبکه عمیق در کنترل بهینه غیرخطی مبدل افزایشده.

## ۲- مبدل افزایشده

مبدل افزایشده<sup>۱</sup> [۱۱ و ۱۲] یک مبدل DC به DC است که کاربردهای زیادی در استفاده از انرژی خورشیدی، پیل‌های سوختی، خودروهای برقی و نورهای فلورسنت دارد. در حقیقت، مبدل‌های افزایشده برای دستیابی به ولتاژهای بالاتر در منابع ولتاژ DC استفاده می‌شوند. همه مبدل‌های افزایشده به حداقل دو دستگاه سوئیچینگ و همچنین یک عنصر ذخیره انرژی نیاز دارند. کنترل ولتاژ خروجی مبدل افزایشده یک مسئله کنترل غیرخطی چالش برانگیز است. در واقع، پاسخ مبدل به حالت عناصر سوئیچینگ بستگی دارد. باید توجه داشت که استراتژی‌های معمول طراحی کنترل‌کننده با استفاده از مدل‌های خطی تقریبی، عملکرد خوبی ندارند؛ بنابراین دستیابی به روش‌های جایگزین بهینه و غیرخطی بسیار مهم است. شکل ۱ مدار مبدل افزایشده را به نمایش گذاشته است. رفتار مبدل افزایشده را می‌توان با دو مدل خطی فضای حالت در نظر گرفت. اولین مدل در حالتی است که سوئیچ مبدل افزایشده باز است و حالت دوم سوئیچ مبدل افزایشده بسته است.



شکل ۱: مدار مبدل افزایشده [۱۱].

مدل فضای حالت مبدل افزایشده با المان سوئیچینگ در حالت باز در رابطه‌های (۱) و (۲) به نمایش گذاشته شده است.

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} \frac{-r_L}{L} & 0 \\ 0 & \frac{-1}{C(R_{Load} + r_C)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \frac{1}{L} \\ 0 \end{bmatrix} V_{in} \quad (1)$$

$$V_O(x) = \begin{bmatrix} 0 & \frac{R_{Load}}{R_{Load} + r_C} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad (2)$$

مدل فضای حالت مبدل افزایشده با المان سوئیچینگ در حالت بسته در رابطه‌های (۳) و (۴) به نمایش گذاشته شده است.

<sup>1</sup> Boost Converter

## ۴- روش شبکه عمیق Q

ادامه، عملکرد روش‌های کنترلی با جزییات موردبررسی قرار گرفته است. نتایج عددی بررسی عملکرد کنترل در محیط سیمولینک ارائه می‌گردد. لازم به ذکر است که در این پژوهش ولتاژ مطلوب ۸۰ ولت در نظر گرفته شده است؛ همچنین عملکرد روش‌های مورد بحث در برابر نویز سفید نیز موردبررسی قرار می‌گیرد. برای سنجش و مقایسه روش‌ها از معیارهای خطا استفاده شده است. میانگین مربعات خطا (MSE)<sup>۴</sup>، میانگین خطای مطلق (MAE)<sup>۵</sup> و انتگرال خطای مطلق (IAE)<sup>۶</sup> به‌عنوان معیار خطا مورد استفاده قرار گرفته‌اند.

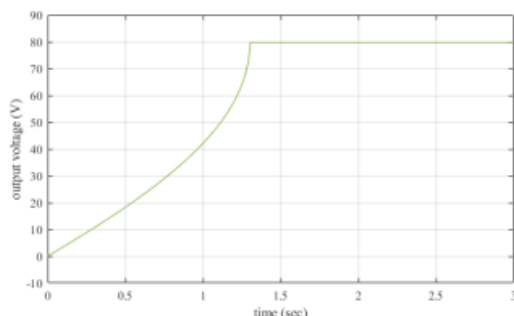
الگوریتم شبکه عمیق Q<sup>۱</sup> یک روش یادگیری تقویتی بدون مدل و آنلاین است. عامل شبکه عمیق Q، یک عامل یادگیری تقویتی مبتنی بر ارزش است که منتقد را برای تخمین بازده یا پاداش‌های آینده تربیت می‌کند. شبکه عمیق Q یک نوع آموزش Q است. عامل‌های شبکه عمیق Q را می‌توان در محیط‌هایی با فضا‌های مشاهده و عمل جدول ۱ آموزش داد.

جدول ۱: فضا‌های مشاهده و عمل در آموزش عامل شبکه عمیق Q

فضای مشاهده	فضای عمل
پیوسته یا گسسته	گسسته

## ۷- کنترل کننده PI

با اعمال کنترل کننده PI بر سیستم نتایج کنترل سیستم برای ولتاژ مرجع موردبررسی قرار گرفته است. هدف از استفاده از کنترل گر PI مقایسه نتایج آن با عملکرد کنترل گر یادگیری تقویتی عمیق است. شکل ۳ نتیجه خروجی از کنترل گر PI را برای ولتاژ مرجع ۸۰ ارائه می‌دهند. زمان کل شبیه‌سازی کنترل ۳ ثانیه است.



شکل ۳: خروجی کنترل گر PI برای ولتاژ مرجع ۸۰ ولت.

## ۵- روش سیاست بهینه‌سازی مبدأ<sup>۳</sup> (PPO)

الگوریتم سیاست بهینه‌سازی مبدأ یک روش یادگیری تقویتی بدون مدل و آنلاین است. این الگوریتم مستلزم استفاده از تجربیات کوچک از تعامل با محیط برای به‌روزرسانی سیاست تصمیم‌گیری است. پس از به‌روزرسانی خط‌مشی، تجربیات گذشته کنار گذاشته می‌شوند و دسته جدیدی برای به‌روزرسانی خط‌مشی ایجاد می‌شود. الگوریتم سیاست بهینه‌سازی مبدأ یک کلاس از روش‌های گرادیان سیاست است که سعی می‌کند واریانس گرادیان را نسبت به سیاست‌های بهتر کاهش دهد؛ که باعث پیشرفت مداوم و اطمینان از عدم تغییر شدید سیاست نسبت به سیاست قبلی و یا کاهش مسیرهای غیرقابل برگشت شود. الگوریتم سیاست بهینه‌سازی مبدأ دارای دو تابع تقریب، شبکه بازیگر و شبکه منتقد است.

### ۱-۵- شبکه بازیگر

این شبکه انتخاب‌های عمل را مستقیماً به حالت مشاهده شده نگاهت می‌کند. در هر زمان خاص بازیگر حالت مشاهده شده را به دست می‌آورد و احتمال انجام عمل در فضای عمل را پیش‌بینی می‌کند. باین‌حال، این تابع میزان عمل مناسب را در مقایسه با سایر اقدامات موجود اندازه‌گیری نمی‌کند. از این رو، شبکه منتقد برای نقد اقدامات بازگشت داده شده توسط شبکه بازیگر به کار گرفته می‌شود.

### ۲-۵- شبکه منتقد

این شبکه حالت مشاهده شده و عمل برگشت داده شده توسط شبکه بازیگر را به‌عنوان ورودی در نظر می‌گیرد و انتظار می‌رود پاداش طولانی‌مدت تخفیف داده شده را برآورده کند.

## ۶- روش پژوهش

همان‌طور که اشاره گردید، هدف اصلی این پژوهش پیاده‌سازی روش‌های یادگیری عمیق در کنترل غیرخطی مبدل افزایشده است. برای این منظور از دو روش یادگیری تقویتی عمیق که شامل DQN و PPO می‌شود استفاده شده است؛ همچنین این روش‌ها در محیط سیمولینک پیاده‌سازی شده است. در

## ۷-۱- بهینه‌یابی ابرپارامترهای شبکه عصبی

باید توجه داشت که برای اعمال روش‌های یادگیری تقویتی عمیق لازم است پیش از دستیابی به نتایج نهایی با استفاده از روش‌های جستجو و سعی و خطا بهترین پارامترهای شبکه عصبی بهینه‌یابی گردند. مهم‌ترین ابرپارامترهای شبکه عصبی شامل تعداد لایه‌ها و تعداد نورون‌ها در هر لایه از شبکه هستند. برای دستیابی به بهترین تعداد لایه‌ها و نورون‌ها در این پژوهش به تأثیر این پارامترها بر میزان خطا و هزینه‌های محاسباتی توجه شده است. در حقیقت برای یافتن بهترین تعداد لایه سعی شده است تا تأثیر افزایش لایه‌ها بر میزان خطا و هزینه‌های محاسباتی سنجیده شود. برای این منظور تعداد لایه‌ها از تعداد ۳ لایه تا ۱۲ لایه موردبررسی قرار گرفته‌اند. باید توجه داشت که برای همه محاسبات تعداد نورون‌ها در تمام لایه‌ها برابر با ۲۵۶ در نظر گرفته شده است تا فقط تأثیر افزایش تعداد لایه‌ها در عملکرد روش‌ها سنجیده شود. با توجه به نتایج موجود واضح است که افزایش تعداد لایه‌ها موجب افزایش چشمگیر هزینه محاسباتی در فرایند یادگیری است. همچنین با توجه به تغییرات اندک در MSE می‌توان نتیجه گرفت که افزایش تعداد لایه‌ها تأثیر چندانی در بهبود نتایج کنترل در این مسئله ندارد. بنابراین واضح است که استفاده از تعداد ۳ لایه برای شبکه عصبی عمیق در فرایند یادگیری امری

<sup>4</sup> Mean squared error

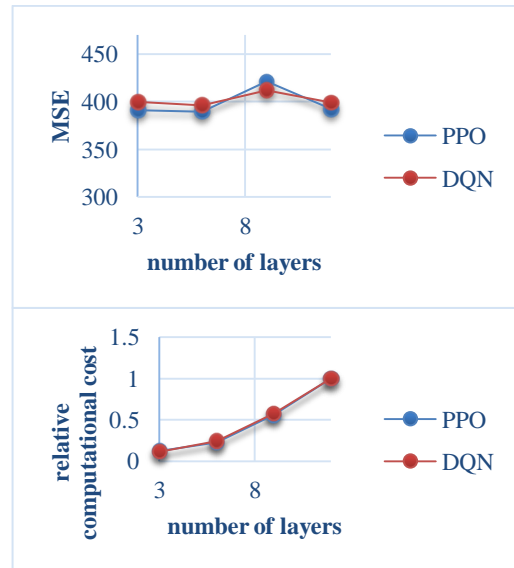
<sup>5</sup> Mean absolute error

<sup>6</sup> Integral absolute error

<sup>2</sup> Deep Q Network

<sup>3</sup> Proximal policy optimization

منطقی است. در ادامه پژوهش کلیه نتایج بر اساس روش‌های یادگیری تقویتی عمیق با سه لایه ارائه می‌گردد.



شکل ۴: میزان خطا (MSE) و هزینه محاسبات کنترل گر یادگیری عمیق با استفاده از لایه‌های مختلف برای ولتاژ مرجع ۸۰ ولت.

یکی دیگر از پارامترهای بسیار مهم و تأثیرگذار بر عملکرد شبکه عصبی تعداد نورون‌ها در هر لایه است. هرچند که واضح است که افزایش تعداد نورون‌ها در بیشتر حالات موجب عملکرد بهتر شبکه عصبی می‌گردد، لازم به ذکر است تعداد بالای نورون موجب افزایش چشمگیر حجم محاسبات می‌گردد. بنابراین در ادامه حالت‌های مختلف تعداد نورون مورد بررسی قرار گرفته است. جدول ۲ تأثیر تعداد نورون را بر میزان MSE نشان می‌دهد.

جدول ۲: مقایسه عملکرد کنترل گر یادگیری تقویتی عمیق با توجه به تعداد نورون‌های مختلف.

80 V		
	neurons	MSE
PPO	32	585.3651
	64	452.6352
	128	412.253
	256	<b>391.2658</b>
	512	392.6850
DQN	32	623.365
	64	433.658
	128	<b>394.9941</b>
	256	395.6352
	512	398.8526

کمترین میزان خطا برای هر یک از روش‌ها در جدول ۲ به صورت برجسته نشان داده شده است. با توجه به نتایج موجود در جدول ۲ می‌توان نتیجه گرفته که روش PPO برای عملکرد مناسب نیازمند به شبکه‌ای عصبی با ۲۵۶ نورون است، در حالی که روش DQN بهترین عملکردش را با ۱۲۸ نورون نشان می‌دهد. هرچند باید توجه داشت که هنگامی که تعداد نورون‌ها از ۱۲۸ عدد در هر لایه بیشتر می‌شود، در همه حالات، هر دو روش PPO و DQN عملکرد مناسبی نشان می‌دهند. در حقیقت افزایش تعداد نورون‌ها بیشتر از تعداد ۱۲۸ تغییر شگرفی در نتایج ایجاد نمی‌کند. در ادامه این پژوهش برای اطمینان از عملکرد مناسب روش‌ها از ۲۵۶ نورون در هر لایه استفاده شده است.

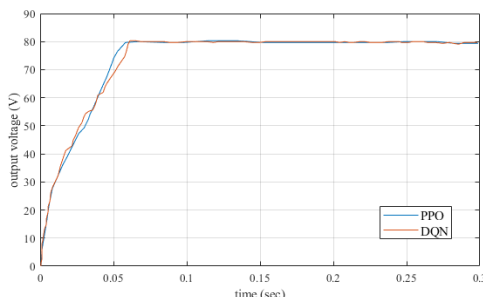
لازم به ذکر است که در حین فرایند یادگیری، لازم است تا دائماً وزن‌های شبکه بهینه گردند. در حقیقت روش‌های بهینه‌سازی تأثیر فراوانی در نتایج یادگیری دارند. یکی دیگر از اهداف این تحقیق بررسی عملکرد ۳ روش بهینه‌سازی در روند یادگیری تقویتی در کنترل غیرخطی مبدل‌های افزایشنده است. روش‌های بهینه‌سازی  $sdgm^7$ ،  $ame^8$  و  $rmsp^9$  در این پژوهش مورد ارزیابی قرار گرفتند. لازم به ذکر است که طبق جدول ۳، روش  $ame$  نسبت به سایر روش‌ها نتیجه بهتری را به نمایش گذاشت. بنابراین در ادامه پژوهش از روش  $ame$  برای بهینه‌سازی وزن‌های شبکه استفاده می‌شود.

جدول ۳: مقایسه تأثیر روش‌های مختلف بهینه‌سازی در عملکرد کنترل گر یادگیری تقویتی عمیق.

V <sub>ref</sub>	RL method optimization method	PPO			DQN		
		rmsp	ame	sdgm	rmsp	ame	sdgm
80 V	MSE	398.6587	<b>391.2658</b>	392.3697	395.3686	<b>394.9941</b>	401.3695

## ۸- پیاده‌سازی کنترل گر بر اساس یادگیری تقویتی عمیق

هدف اصلی این پژوهش مقایسه عملکرد کنترلی مبتنی بر یادگیری تقویتی عمیق با عملکرد کنترل گر PI است. خروجی کنترل گر بر اساس یادگیری تقویتی عمیق برای ولتاژ مرجع ۸۰ ولت در شکل ۵ نشان داده شده است. با مقایسه نتایج حاصل از کنترل بر اساس یادگیری تقویتی عمیق و داده‌های حاصل از کنترل گر PI واضح است که روش‌های یادگیری تقویتی عمیق عملکرد بسیار خوبی به لحاظ کاهش زمان نشست دارند. باید توجه داشت که با توجه به عملکرد روش‌های یادگیری تقویتی و کاهش چشمگیر زمان نشست، کل فرآیند شبیه‌سازی 0.3 ثانیه انتخاب شده است.



شکل ۵: خروجی کنترل گر یادگیری تقویتی عمیق برای ولتاژ مرجع ۸۰ ولت.

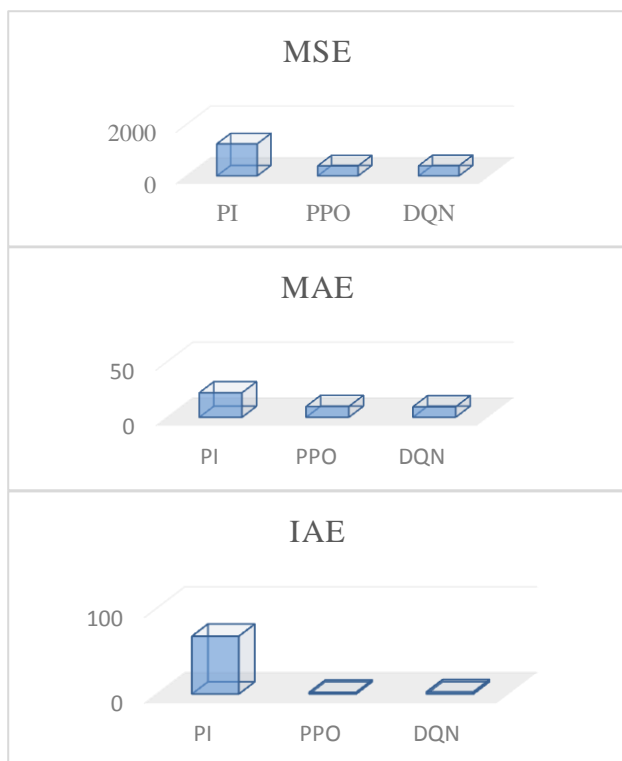
نتایج عددی ارائه شده در شکل ۶ و شکل ۷ شامل زمان نشست و میزان خطا است. باید توجه داشت که برای بررسی دقیق‌تر رفتار کنترل گر، میزان خطا در کل زمان شبیه‌سازی گزارش شده است. نتایج ارائه شده برای روش‌های یادگیری تقویتی عمیق در حقیقت میانگین نتایج اعتبارسنجی روش‌های کنترلی در ۱۰۰ آزمایش است.

<sup>7</sup> stochastic gradient descent with momentum

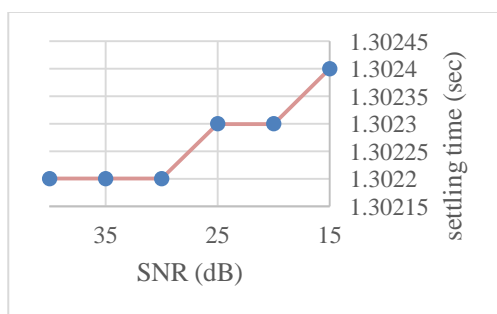
<sup>8</sup> adaptive moment estimation

<sup>9</sup> root mean square propagation

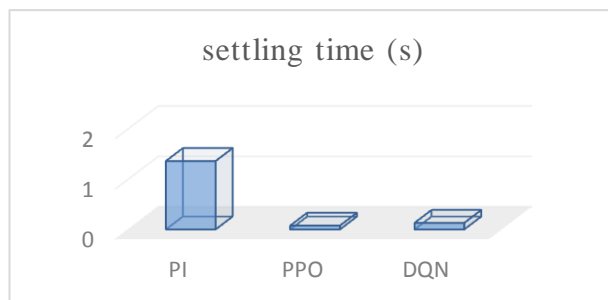
شکل ۹ میزان عملیات موفقیت‌آمیز با استفاده از روش‌های PPO و DQN را در برابر ۱۰۰ آزمایش دارای نویز سفید نشان می‌دهد. نتایج نشان می‌دهد که روش PPO در تمامی حالات برای SNR برابر با ۲۵ و ۳۰ دسی‌بل موفق به کنترل سیستم دارای نویز شده است. لیکن روش DQN عملکرد مطلوبی ندارد و حتی تا مقدار ۹۰ درصد آزمایشات را شکست می‌خورد. یعنی در حقیقت در بیشتر حالات دارای نویز با SNR کمتر از ۳۰ اصلاً قادر به دستیابی به ولتاژ مرجع موردنظر نیست. با توجه به نتایج در شکل ۹ روش PPO بسیار نیرومندتر از روش DQN است.



شکل ۷: مقایسه میزان خطا در تمام فرایندهای کنترلی برای روش‌های کنترلی مختلف در ولتاژ مرجع برابر با ۸۰ ولت.



شکل ۸: تأثیر افزایش میزان نویز در افزایش زمان نشست در کنترل‌گر PI ولتاژ مرجع برابر با ۸۰ ولت.



شکل ۶: مقایسه میزان زمان نشست برای روش‌های کنترلی مختلف در ولتاژ مرجع برابر با ۸۰ ولت.

شکل ۶ مقادیر زمان نشست روش‌های مختلف را با هم مورد مقایسه قرار می‌دهد. همان‌طور که واضح است زمان نشست روش‌های یادگیری تقویتی عمیق نسبت به روش PI بسیار کمتر است. این امر عملکرد بسیار مناسب روش‌های یادگیری تقویتی عمیق را در دستیابی سریع به ولتاژ مرجع در کنترل مبدل‌های افزایش‌دهنده DC نشان می‌دهد. همچنین با مقایسه میزان زمان نشست روش‌های یادگیری تقویتی، می‌توان نتیجه گرفت که عملکرد روش PPO به لحاظ سرعت کنترلی بهتر از روش DQN است.

شکل ۷ نتایج مربوط به فرایند کنترل برای ولتاژ مرجع ۸۰ ولت را نشان می‌دهند. این نتایج عملکرد بسیار خوب روش‌های PPO و DQN را نسبت به روش PI نشان می‌دهند. باید توجه داشت که در بیشتر حالات و همچنین بر اساس اکثر معیارهای خطا، عملکرد روش PPO بهتر از روش DQN فرایند کنترل غیرخطی مبدل‌های افزایش‌دهنده DC است.

## ۹- حساسیت کنترل‌گرها به نویز سفید

برای سنجش نیرومندی روش‌های کنترلی در این بخش عملکرد این روش‌ها در برابر اعمال نویز گوسی سفید<sup>۱۰</sup> سنجیده شده است. نویز گوسی سفید یک نویز خطی است که به سیگنال ارسالی سیستم اعمال می‌گردد. این نویز دارای توزیع گوسی نسبت به زمان است. برای تعریف نویز گوسی سفید معمولاً از نسبت سیگنال به نویز (SNR)<sup>۱۱</sup> استفاده می‌شود. این نسبت بر اساس رابطه-۵ تعریف می‌شود.

$$SNR = \frac{P_{signal}}{P_{noise}} \quad (5)$$

که در این رابطه  $P_{signal}$  توان اطلاعات موردنظر و  $P_{noise}$  توان سیگنال ناخواسته است.

برای سنجش اثر نویز ابتدا نویز سفید به کنترل‌گر PI اعمال شده است. شکل ۸ تأثیر افزایش نویز سفید را بر عملکرد کنترل‌گر PI به لحاظ افزایش زمان نشست نشان می‌دهد. همان‌طور که در شکل ۸ نشان داده شده است، با کاهش SNR (افزایش نویز) مقدار زمان نشست افزایش می‌یابد. با توجه به شکل می‌توان نتیجه گرفت که مقدار SNR برابر با ۲۵ و ۳۰ دسی‌بل در حقیقت مرز ایجاد تغییرات در فرایند کنترلی است. از این‌رو، رفتار روش‌های DQN و PPO نسبت به اعمال نویز با مقادیر ۲۵ و ۳۰ دسی‌بل سنجیده می‌شود.

<sup>10</sup> Additive White Gaussian Noise

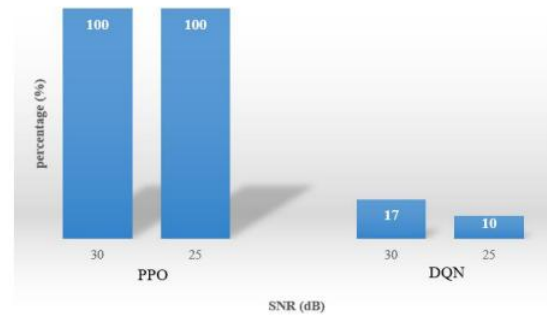
<sup>11</sup> Signal to noise ratio



- مقایسه عملکرد روش‌های یادگیری تقویتی به لحاظ زمان نشست و میزان خطا نشان داد که روش PPO نسبت به روش DQN در کنترل غیرخطی مبدل افزایشده بهتر عمل می‌کند.
- بررسی اعمال نویز سفید بر کنترل‌گرهای مبتنی بر یادگیری تقویتی نشان داد که روش DQN حساسیت بالایی به اعمال نویز سفید در فرایند کنترل مبدل افزایشده DC دارد. لیکن روش PPO در تمامی حالات مورد آزمایش، عملکرد خوبی در برابر اعمال نویز داشت.

## مراجع

- [1] Hung JY, Gao W, Hung JC. *Variable structure control: A survey*. IEEE transactions on industrial electronics. 1993 Feb;40(1):2-2.
- [2] Cominos P, Munro N. *PID controllers: recent tuning methods and design to specification*. IEE Proceedings-Control Theory and Applications. 2002 Jan 1;149(1):46-53.
- [3] Guo L, Hung JY, Nelms RM. *Digital controller design for buck and boost converters using root locus techniques*. In IECON'03. 29th Annual Conference of the IEEE Industrial Electronics Society (IEEE Cat. No. 03CH37468) 2003 Nov 2 (Vol. 2, pp. 1864-1869). IEEE.
- [4] Balestrino A, Corsanini D, Landi A, Sani L. *Circle-based criteria for performance evaluation of controlled dc-dc switching converters*. IEEE Transactions on Industrial Electronics. 2006 Nov 30;53(6):1862-9.
- [5] Perry AG, Feng G, Liu YF, Sen PC. *A new design method for PI-like fuzzy logic controllers for DC-to-DC converters*. In 2004 IEEE 35th Annual Power Electronics Specialists Conference (IEEE Cat. No. 04CH37551) 2004 Jun 20 (Vol. 5, pp. 3751-3757). IEEE.
- [6] Andalibi M, Hajihosseini M, Teymoori S, Kargar M, Gheisarnejad M. *A Time-Varying Deep Reinforcement Model Predictive Control for DC Power Converter Systems*. In 2021 IEEE 12th International Symposium on Power Electronics for Distributed Generation Systems (PEDG) 2021 Jun 28 (pp. 1-6). IEEE.
- [7] Pradeep DJ, Noel MM, Arun N. *Nonlinear control of a boost converter using a robust regression based reinforcement learning algorithm*. Engineering Applications of Artificial Intelligence. 2016 Jun 1;52:1-9.
- [8] Gheisarnejad M, Farsizadeh H, Khooban MH. *A Novel Nonlinear Deep Reinforcement Learning Controller for DC-DC Power Buck Converters*. IEEE Transactions on Industrial Electronics. 2020 Jul 1;68(8):6849-58.
- [9] Gheisarnejad M, Farsizadeh H, Tavana MR, Khooban MH. *A Novel Deep Learning Controller for DC-DC Buck-Boost Converters in Wireless Power Transfer Feeding CPLs*. IEEE Transactions on Industrial Electronics. 2020 May 20;68(7):6379-84.
- [10] Hajihosseini M, Andalibi M, Gheisarnejad M, Farsizadeh H, Khooban MH. *DC/DC power converter control-based deep machine learning techniques: Real-time implementation*. IEEE Transactions on Power Electronics. 2020 Mar 2;35(10):9971-7.
- [11] Sundareswaran K, Sreedevi VT. *Boost converter controller design using queen-bee-assisted GA*. IEEE Transactions on industrial electronics. 2008 Oct 31;56(3):778-83.
- [12] Rashid MH. *Power electronics: circuits, devices, and applications*. Pearson Education India; 2009.
- [13] Li Y. *Deep reinforcement learning: An overview*. arXiv preprint arXiv:1701.07274. 2017 Jan 25.



شکل ۹: مقایسه عملکرد روش‌های PPO و DQN در برابر اعمال نویز سفید برای ولتاژ مرجع برابر با ۸۰ ولت.

## ۱۰- نتیجه‌گیری

پژوهش حاضر شامل چگونگی پیاده‌سازی روش‌های یادگیری تقویتی عمیق در کنترل غیرخطی مبدل افزایشده DC است. نتایج عددی این پیاده‌سازی و همچنین مقایسه عملکرد روش‌ها موجب دستیابی به نتایج مفیدی گردید. نتیجه‌های به‌دست‌آمده از پیاده‌سازی و مقایسه روش‌ها به‌طور خلاصه در ادامه بیان می‌گردد.

- بهینه‌یابی تعداد لایه‌های شبکه عصبی عمیق در عملکرد روش‌های یادگیری عمیق نشان داد که افزایش تعداد لایه‌ها تأثیر چندانی در میزان خطا ندارد. باید توجه داشت که این افزایش تعداد لایه‌ها هرچند ممکن است که تأثیری کم در کاهش خطای عملکرد کنترل‌گر داشته باشد، لیکن موجب افزایش چشمگیر هزینه محاسباتی می‌گردد. از این‌رو، برای کنترل غیرخطی مبدل افزایشده DC در این پژوهش از ۳ لایه بهره برد شد.
- بررسی تعداد نورون‌ها در میزان خطای کنترل‌گر نشان داد که پس از رسیدن تعداد نورون‌ها به عدد ۲۵۶، عامل عملکرد مناسبی از خود نشان می‌دهد. واضح است که افزایش تعداد نورون‌ها موجب افزایش هزینه محاسباتی می‌گردد. لیکن عملکرد عامل‌های یادگیری تقویتی عمیق با تعداد نورون کمتر خیلی مناسب نیست. همچنین باید توجه داشت که عامل با تعداد نورون بالاتر از ۲۵۶ نتایج مشابه ارائه می‌کرد. بنابراین در این پژوهش تعداد نورون‌ها ۲۵۶ انتخاب گردید.
- یکی دیگر از عوامل مهم در عملکرد روش‌های یادگیری عمیق، انتخاب روش بهینه‌سازی برای شبکه عصبی است. در این پژوهش عملکرد سه روش معروف بهینه‌سازی مورد بررسی قرار گرفت. نتایج مقایسه‌ای بیانگر برتری روش adam نسبت به سایر روش‌ها بود. باید توجه داشت که هر سه روش بهینه‌سازی توانایی خوبی داشتند.
- بررسی مقایسه‌ای روش‌های مختلف کنترلی نشان داد که کنترل‌گر مبتنی بر یادگیری تقویتی عمیق دارای زمان نشست بسیار کمتری نسبت به کنترل‌گر PI است. این امر موجب می‌شود که خطا در فرایند کنترلی کاهش چشمگیری نسبت به کنترل‌گر PI داشته باشد.