



**Diseño y simulación de controladores por aprendizaje reforzado
para convertidores de potencia utilizados en para micro redes de corriente directa.**

John Fredy Franco Martínez

Código: 11291726282

Brayan Sebastián Laserna

Código: 11291726286

Universidad Antonio Nariño

Programa Ingeniería Control y Automatización Industrial.

Facultad de Ingeniería Mecánica, Electrónica y Biomédica

Bogotá, Colombia

Año 2023

Diseño y simulación de controladores por aprendizaje reforzado para convertidores de potencia utilizados en micro redes de corriente directa.

John Fredy Franco Martínez

Brayan Sebastián Laserna

Proyecto de grado presentado como requisito para optar al título de:

Ingeniero en control y automatización industrial.

Director (a):

Ing. Christian Erazo Ordoñez PhD

Línea de Investigación:

Control.

Universidad Antonio Nariño

Programa Ingeniería Control y Automatización Industrial

Facultad de Ingeniería Mecánica, Electrónica y Biomédica

Bogotá, Colombia

2023

NOTA DE ACEPTACIÓN

El trabajo de grado titulado

_____.

Cumple con los requisitos para optar

Al título de _____.

Firma del Tutor

Firma Jurado

Firma Jurado

Contenido

Resumen.....	11
Abstract.....	12
1. Introducción	13
1.1 Planteamiento del problema.....	16
1.2 Justificación	16
1.3 Objetivos	17
1.3.1 General.....	17
1.3.2 Específicos.....	17
1.4 Metodología.....	18
2. Marco teórico	20
2.1 Convertidor de potencia reductor (Buck).....	20
2.2 Convertidor de potencia elevador (Boost)	23
2.3 Aprendizaje por refuerzo	26
2.3.1 Descripción del aprendizaje por refuerzo.....	27
2.3.2 Agente.....	27
2.3.3 Entorno	28
2.3.4 Observación del entorno.....	29
2.3.5 Función de recompensa	30
2.3.6 Política de estrategia.....	31
3. Desarrollo Metodológico	32
3.1 Definición del entorno de los convertidores Buck.....	33
3.1.1 Definición del entorno de los convertidores Boost.....	34
3.2 Diseño de las funciones de recompensa para convertidor reductor elevador 35	
3.3 Entrenamiento del agente DDGP para un convertidor reductor (Buck).....	37
3.3.1 Entrenamiento convertidor elevador (Boost).....	39
3.4 Validación del agente reductor (Buck).....	41
3.4.1 Resultados convertidores Reductor (Buck) con PID	44
3.4.2 Validación del agente elevador (Boost).....	46
3.4.3 Resultados convertidores elevador (Boost) con PID.....	50

4.	Analisis de resultados	52
5.	Conclusiones	55
6.	Referencias Bibliográficas	63

Lista de Figuras

Figura 2-1 Diagrama circuito del convertidor reductor (Buck). Fuente	21
Figura 2-2 Convertidor reductor (Buck) Lazo abierto Voltaje de entrada Vs Voltaje de salida. Fuente Propia.....	22
Figura 2-3 Diagrama circuito del convertidor elevador (Boost) Fuente. [propia].....	23
Figura 2-4 Clasificación de algoritmos RL.....	26
Figura 2-5 interacciones agente-ambiente en el aprendizaje por refuerzo	28
Figura 2-6 Entorno ambiente [11].....	29
Figura 2-7 Aprendizaje por refuerzo basado en modelos []	30
Figura 2-8 Ejemplo acción de recompensa [10]	30
Figura 3-1 entorno convertidor reductor (Buck) fuente propia	33
Figura 3-2 Entorno controlador Buck Fuente. Elaboración propia.....	34
Figura 3-3 Entorno controlador elevador (Boost) Fuente. Elaboración propia	35
Figura 3-4 Bloques recompensa y penalización Fuente. Elaboración propia.....	35
Figura 3-5 Número de episodios vs el promedio de recompensas convertidor reductor (Buck) Fuente. (Tearle, n.d.).....	38
Figura 3-6 Entrenamiento convertidor elevador (Boost), ajuste Vref vs Vboost. Fuente. Elaboración Propia.....	40
Figura 3-7 Graficas de desarrollo control aprendizaje por refuerzo controlador reductor (Buck) Fuente. Elaboración Propia.....	41
Figura 3-8 Recompensa del agente. Fuente. Elaboración Propia.	42
Figura 3-9 Ajuste del ciclo de trabajo. Fuente. Elaboración Propia.	43
Figura 3-10 Diagrama convertidor reductor (Buck) con PID. Fuente. Elaboración Propia.	44
Figura 3-11 Grafica convertidor reductor (Buck) con PID entrada 24 V salida 6 V. Fuente. Elaboración Propia.....	45
Figura 3-12 Graficas de desarrollo control aprendizaje por refuerzo controlador elevador (Boost) Fuente. Elaboración Propia.....	47
Figura 3-13 Recompensa del agente. Fuente. Elaboración Propia.	48
Figura 3-14 interacción con el entorno en búsqueda de la señal de referencia en el control de aprendizaje por refuerzo controlador elevador (Boost) Fuente. Elaboración Propia.....	49
Figura 3-15 Convertidor elevador (Boost) Con control PID Fuente. Elaboración Propia.	50
Figura 3-16 Grafica convertidor elevador (Boost) con Control PID entrada 12 V, salida 24. Fuente Propia	52
Figura 4-1 Grafica convertidor reductor (Buck) con PID entrada 24 V salida 6 V. Fuente. Elaboración Propia.....	53

Figura 4-2 Grafica convertidor elevador (Boost) con Control PID entrada 12 V, salida 24. Fuente Propia	53
Figura 4-3 Graficas de desarrollo control aprendizaje por refuerzo controlador reductor (Buck) Fuente. Elaboración Propia.....	54
Figura 4-4 Graficas de desarrollo control aprendizaje por refuerzo controlador elevador (Boost) con variación de la carga. Fuente. Elaboración Propia.....	54

Lista de tablas

Tabla 1 Hiperparametros de entrenamiento convertidor Buck.....	37
Tabla 2 Hiperparametros entrenamiento convertidor Boost.....	39
Tabla 3 Diseño convertidor reductor (Buck). Fuente. Elaboración Propia.....	41
Tabla 4 Valores Convertidor reductor con control PID.....	45
Tabla 5 Tabla parámetros convertidor elevador (Boost)	46
Tabla 6 Parámetros del diseño convertidor elevador (Boost) con PID.....	51
..	

(Dedicatoria)

*Dedicado principalmente a Dios
a la vida misma y a la familia
por ser conspiradores de este sueño
son los más grandes y fieles testigos
de que el trabajo, el esfuerzo, la lealtad
y la honorabilidad construyen
personas integrales.*

Agradecimientos

El principal agradecimiento es a Dios por brindarnos la sabiduría, la fe y compromiso para terminar una carrera profesional, a nuestras familias que nos han brindado su apoyo incondicional.

Al equipo docente que con su sabiduría nos guiaron por este largo trayecto, expresaron voz de apoyo y creyeron en nuestro proceso de formación.

Al director del proyecto el Ingeniero Christian Erazo, agradecemos mucho el orientarnos, su compromiso, dedicación, paciencia desde el momento de la búsqueda del anteproyecto hasta la culminación del proyecto.

A la comunidad UAN, a todas aquellas personas que con sus labores amenizan y hacen un lugar cálido y pulcro.

Resumen

En este proyecto se presenta una propuesta la cual está asociada al proyecto de investigación “Gestión de energía basada en computación en la nube para la interoperabilidad entre un grupo de microrredes aisladas” financiado por Min Ciencias. Este proyecto tiene como objetivo principal diseñar, implementar y validar técnicas de control inteligente para convertidor de potencia reductor (Buck) y convertidor de potencia elevador (Boost) utilizados en micro redes de corriente directa.

El diseño de los convertidores de potencia se realiza utilizando el entorno de diagrama de bloques multidominio Simulink, lo que permite parametrizar los valores de los componentes del circuito.

Este trabajo presenta los resultados obtenidos tras la implementación de un agente DDGP (Deep Deterministic Policy Gradient) como controlador del ciclo de trabajo, permitiendo la variación del voltaje de salida de los convertidores reductores (Buck) y elevador (Boost). Se lleva a cabo una evaluación exhaustiva de rendimiento y se comparan los resultados con métodos de control convencionales. Las conclusiones de este estudio demuestran que el enfoque de control de aprendizaje reforzado mediante el agente DDGP es altamente eficaz en la regulación del voltaje de salida de los convertidores de potencia. Se obtienen mejoras significativas en términos de precisión y respuesta dinámica en comparación con los métodos convencionales. Estos hallazgos respaldan la viabilidad y utilidad de la implementación de técnicas de control inteligente en micro redes de corriente directa.

Los resultados obtenidos confirman la efectividad de esta aproximación y su potencial para la aplicación práctica en sistemas de gestión de energía basados en micro redes aisladas acopladas a sistemas de energías renovables.

Palabras Clave: Control, Convertidores, Interfaz, Microrredes, Simulación, Potencia.

Abstract

This project presents a proposal which is associated with the research project "Energy management based on cloud computing for interoperability between a group of isolated microgrids" financed by Min Ciencias. The main objective of this project is to design, implement and validate intelligent control techniques for step-down (Buck) power converters and boost (Boost) power converters used in direct current microgrids.

The design of the power converters is done using the Simulink multi-domain block diagram environment, which allows the values of the circuit components to be parameterized.

This paper presents the results obtained after the implementation of a DDGP (Deep Deterministic Policy Gradient) agent as duty cycle controller, allowing the variation of the output voltage of the buck (buck) and boost (boost) converters. A comprehensive performance evaluation is carried out and the results are compared with conventional control methods.

The conclusions of this study show that the reinforcement learning control approach using the DDGP agent is highly effective in regulating the output voltage of power converters. Significant improvements are obtained in terms of precision and dynamic response compared to conventional methods. These findings support the feasibility and usefulness of the implementation of intelligent control techniques in direct current microgrids.

The results obtained confirm the effectiveness of this approach and its potential for practical application in energy management systems based on isolated microgrids coupled to renewable energy systems.

Keywords: Control, Converters, Interface, Microgrids, Simulation, Power

1. Introducción

Las micro redes de corriente directa (MG) están conformadas por generadores de energía, usualmente fuentes renovables como son paneles solares, turbinas eólicas, baterías, etc., y cargas domésticas o industriales, Además, las micro redes MG también incluyen un sistema de distribución que permite el flujo eficiente de energía desde los generadores hacia las cargas conectadas. [1]. Cada vez más, estas soluciones energéticas, flexibles y eficientes, se están volviendo comunes en lugares apartados, de difícil acceso donde la red de distribución general no es capaz de llegar, [8], [9].

Las características más importantes que aseguran la operatividad y estabilidad de la micro red son el controlador y la topología de la red [10]. Los autores realizan un análisis comparativo de las topologías de MGs más comunes: tipo bus sencilla, multiterminal, anillo y zonal. También presentan diversas arquitecturas de control centralizado, descentralizado, resaltando las características del tipo maestro esclavo y droop. Dentro de las micro redes de corriente directa, los convertidores de potencia juegan un papel importante en la interfaz entre cada uno de los dispositivos. Los convertidores más utilizados son los reductores - elevadores (Buck - Boost) [2]

El convertidor elevador (Boost) utiliza una configuración diferente para la ubicación del interruptor y la inductancia, lo que resulta en un voltaje de salida mayor que el voltaje de entrada. Sin embargo, esta configuración presenta un desafío adicional en la reducción de los efectos de perturbaciones externas, ya que tienden a amplificarse [4]. En este tipo de convertidor, si el interruptor permanece abierto, el voltaje de salida sería igual al voltaje de entrada. Por otro lado, si el interruptor se mantiene cerrado, se produciría una corriente muy alta en el inductor, lo que podría ocasionar un cortocircuito en el circuito de corriente directa (DC) [6]. En este trabajo, se realiza el control de un convertidor reductor (Buck) y de un Convertidor elevador (Boost) en el cual el voltaje de salida puede ser de mayor o menor amplitud que la entrada, presentando el mismo problema que el convertidor elevador (Boost) si el interruptor se mantiene cerrado. En el caso de un convertidor Buck el circuito puede interpretarse como un filtro LC aplicado a una señal de

PWM [3].

Para los otros dos, el inductor es el encargado de entregar la corriente necesaria a la carga por lo que es importante evitar saturar su núcleo, para este fin, el inductor se carga cuando el interruptor está cerrado y entrega su energía a la carga cuando se abre el interruptor, el rizado de voltaje es filtrado por el capacitor en paralelo con la carga.

En la literatura se han propuesto diferentes técnicas de control de convertidores siendo las más robustas a variaciones en la carga las técnicas en modos deslizantes (SMC) [10] y el control predictivo basado en modelo [11].

Estas técnicas si bien muestran excelentes prestaciones en cuanto a la regulación del voltaje y variaciones en la carga tienen una desventaja y es que la síntesis del controlador se basa en el modelo de la planta, de lo cual muchas veces no se dispone. Como solución al problema anteriormente comentado surge la teoría del control en modo deslizante. El control en modo deslizante (VSC) aplicado a sistemas de estructura variable (VSS) fue introducido en los años 50 en la antigua Unión Soviética por Emelyanov y otros colaboradores Markov [25].

Aquí un algoritmo DDPG es utilizado para generar la señal del ciclo de trabajo estable, a partir del ajuste de la señal PWM. En comparación con los resultados encontrados en la literatura, la técnica propuesta es capaz de regular de forma que se ajusta y mantiene el voltaje del bus DC a un valor deseado, sin un conocimiento a priori del modelo [23]; desde entonces, ha sido desarrollado en sistemas no lineales, multi entrada-salida, sistemas discretos, sistemas estocásticos, etc., en aplicaciones en campos tan diversos como la electrónica de potencia, problemas de diseño aeroespacial, control de brazos robot, etc.

Las características más importantes se deben a la adición de robustez al sistema respecto a variaciones de sus parámetros y perturbaciones externas; mediante la técnica de control en modo deslizante cuando el sistema se encuentra en régimen deslizante el orden de la dinámica se reduce en un grado y las características de la superficie de deslizamiento se "imponen" al sistema. Según la definición de Sira-Ramírez [24]

Una superficie en el espacio de estado de un sistema dinámico representa una relación entre las variables de estado que describen el comportamiento del sistema.

Si éste es forzado a evolucionar sobre esta superficie, las relaciones estáticas de la dinámica resultante quedan determinadas por los parámetros y ecuaciones que definen la superficie. Por otra parte, la acción de control ha de ser la apropiada para que las trayectorias del sistema se "dirijan" hacia la superficie. Cuando esto ocurre se dice que el sistema se encuentra en régimen o dinámica deslizando y la superficie se denomina superficie de deslizamiento ambiente para calcular la recompensa y los estados a partir de un proceso de decisión, recientemente las técnicas de control inteligente como el aprendizaje por refuerzo profundo (DRL) y aquellas basadas en redes neuronales, han llamado la atención de la comunidad científica [12].

En un algoritmo basado en una política determinística se utilizó para compensar la señal del ciclo de trabajo, la cual es capaz de regular el voltaje de salida de un convertidor reductor (Buck) que alimenta una carga no lineal variante en el tiempo. Los resultados experimentales mostraron un rendimiento superior con respecto al control MPC, disminuye el tiempo de respuesta y el sobre impulso [13]. Se diseñó un controlador basado en aprendizaje por refuerzo, donde se establece el modelo del ambiente para calcular la recompensa y los estados a partir de un proceso de decisión de Márkov. Aquí un algoritmo DDPG es utilizado para generar la señal del ciclo de trabajo estable, a partir del ajuste de la señal PWM.

En comparación con los resultados encontrados en la literatura, la técnica propuesta en este proyecto demuestra su capacidad para regular de manera precisa y eficiente el voltaje DC, manteniéndolo en un valor deseado de acuerdo con un voltaje de referencia establecido. Esto se logra mediante el empleo de técnicas de control inteligente. Esta aproximación representa un avance significativo en el campo de los convertidores de potencia, ya que permite un control preciso y adaptativo del voltaje DC sin la necesidad de conocimientos previos detallados sobre el modelo de la planta.

1.1 Planteamiento del problema

La implementación y validación de sistemas complejos como los convertidores de potencia presentan desafíos prácticos y económicos. Uno de ellos se encuentra en el control de los convertidores de potencia, como se mencionó anteriormente. En la literatura, la mayoría de las técnicas de control propuestas dependen de un modelo del sistema.

Por lo tanto, este trabajo tiene como objetivo desarrollar técnicas de control inteligente para los convertidores de potencia reductores (Buck) y elevadores (Boost). La falta de estas técnicas de control inteligente en micro redes representa un obstáculo significativo para lograr un funcionamiento eficiente, estable y confiable de estos sistemas. Las técnicas de control convencionales basadas en modelos requieren un conocimiento detallado del sistema y no pueden adaptarse fácilmente a las condiciones cambiantes de la micro red. Además, la implementación de estas técnicas puede ser costosa y compleja desde el punto de vista económico.

1.2 Justificación

Más del 13 % de la población mundial no tiene acceso a la electricidad, de los cuales el 84 % habitan en zonas rurales, mayormente en países en vía de desarrollo [14]. En Colombia según datos del Instituto de Planificación y Promoción de Soluciones Energéticas para las Zonas no Interconectadas (Ipse), existen 1.710 localidades rurales en donde se calcula que 128.587 personas solo acceden al servicio entre cuatro y doce horas al día [15]. En este sentido, el control y operación de los dispositivos que componen una micro red es de gran interés en la comunidad científica que busca mejorar la disponibilidad del recurso energético para muchas de estas poblaciones no interconectadas.

Ante los desafíos mencionados, es necesario desarrollar técnicas de control inteligente que superen las limitaciones de los enfoques basados en modelos existentes. Estas técnicas deben ser capaces de adaptarse de forma autónoma a las fluctuaciones en la carga, las fuentes de energía y las condiciones operativas de la micro red, así mismo deben permitir una implementación económica y práctica en entornos reales.

1.3 Objetivos

1.3.1 General

Desarrollar técnicas de control de aprendizaje por refuerzo para convertidores de potencia reductor (Buck) y elevadores (Boost) acoplados a fuentes de energía renovables y sistemas de almacenamiento de energía.

1.3.2 Específicos.

- Establecer el ambiente o entorno de acción para un convertidor reductor (Buck) y un convertidor elevador (Boost), considerando modelos de fuentes de energía renovables.
- Definir las funciones de recompensa para un convertidor reductor (Buck) y un convertidor elevador (Boost), que incentive a los agentes a cumplir con los objetivos de control.
- Entrenar iterativamente los agentes para que aprendan una política de seguimiento a medida que interactúan con el entorno.
- Realizar la validación de la regulación de los voltajes en los convertidores de potencia en el software Matlab, teniendo en cuenta variaciones en la carga. Métricas de rendimiento de variaciones en la carga tiempo de respuesta y error en estado estacionario.

1.4 Metodología

Para el desarrollo del proyecto se contará con 4 fases, las cuales permitirán alcanzar los objetivos propuestos.

Fase1: Esta etapa comprende la construcción de los circuitos reductor y elevador de potencia utilizando las librerías de componentes eléctricos de Simulink. Se realizarán dos simulaciones independientes, una para cada convertidor. En el caso del convertidor reductor (Buck), se empleará un voltaje de entrada de 24V y se buscará obtener un voltaje de salida de 12V, siguiendo las métricas de diseño establecidas. Para el convertidor elevador (Boost), se utilizará un voltaje de entrada de 12V y se procurará alcanzar un voltaje de salida de 30V, también en consonancia con las especificaciones de diseño.

Estos valores de voltaje representan los requerimientos necesarios para el adecuado funcionamiento y rendimiento de cada convertidor. El circuito construido será utilizado como la "planta" durante el proceso de entrenamiento del agente de control.

Fase 2: Deep Deterministic Policy Gradient DDPG, Actor Critico, dependiendo del rendimiento del algoritmo. Posteriormente se establecerán las funciones de recompensa que permitirán a los agentes "controladores" llevar el voltaje de salida a un valor deseado. Para ello se buscarán funciones que logren hacer esto de manera independiente en cada variable de estado. Dichas recompensas se traducen en una suma de expresiones matemáticas p. ej. exponenciales, en la que por cada eje (x, y) se busca una función que tenga un valor máximo en la posición objetivo respecto a un plano cartesiano dispuesto sobre el entorno. Para el proceso de diseño del agente y de las funciones de recompensa se utilizará el conjunto de funciones de la librería de aprendizaje por refuerzo de Matlab (Reinforcement Learning Toolbox - Matlab) [17].

Fase 3. En el proceso de entrenamiento de los agentes, se deben ajustar los parámetros del algoritmo como por ejemplo las condiciones iniciales, las perturbaciones, variaciones en la carga, señales de actuación y observaciones, serán tenidas en cuenta para la generación del ambiente de las variables de estado y sus

limitaciones, así como también el número máximo de episodios, dado que esto reducirá considerablemente el tiempo de entrenamiento. Es importante monitorear el proceso de entrenamiento por si es necesario ajustar las funciones de recompensa o penalización o inclusive seleccionar otro tipo de agente. En esta fase se realiza el proceso de diseño del controlador, por lo cual las métricas de desempeño no están estandarizadas tipológicamente.

Fase 4. Para la validación de los algoritmos, se plantearán escenarios donde exista una variación en la carga, esperando que el voltaje de salida de los convertidores sea la deseada, aumentando el voltaje por encima del voltaje de alimentación (elevador 12V -30 V) o la funcionalidad reductora $V_o < 24$ V.

2. Marco teórico

En este capítulo se establecen los parámetros para la aplicación de un método de aprendizaje por refuerzo (Reinforcement Learning) para un convertidor de potencia reductor (Buck) y convertidor elevador (Boost) [51]. El aprendizaje por refuerzo es un subconjunto del aprendizaje automático (Machine Learning) consistente en optimizar la función de recompensa, las observaciones producen acciones de bajo nivel, así, en un entorno dinámico, no estático, donde el agente mediante la exploración interactúa, mejorando la secuencia de acciones, aprendiendo del entorno [51].

Los convertidores reductores (Buck) son de entrada de corriente continua (DC) y producen una salida de corriente continua a una tensión menor. Son muy comunes en aplicaciones de electrónica de potencia, como en fuentes de alimentación de computadoras, cargadores de teléfonos móviles, convertidores solares y muchos otros dispositivos que requieren una tensión de salida más baja que la entrada [52].

Por otra parte, el convertidor Elevador (Boost) es aquel que por construcción se usa para convertir una fuente de voltaje de entrada en una salida de voltaje mayor. El convertidor elevador (Boost) se emplea en aplicaciones de potencia, como fuentes de alimentación de iluminación LED, sistemas de carga de baterías y muchos otros dispositivos que requieren una tensión de salida más alta que la entrada [52]. Sus aplicaciones en el campo de energía renovable, como en sistemas de paneles solares, son excepcionales para aumentar la tensión de salida de los paneles y alimentar cargas de mayor voltaje [52].

2.1 Convertidor de potencia reductor (Buck)

Los convertidores DC-DC a menudo se utilizan como fuentes de alimentación en una variedad de aplicaciones, incluidos automóviles eléctricos [18,19], aeroespaciales aplicaciones [20], sistemas de energía renovable [21] y equipo electrónico [22]. Este dispositivo tiene como función regular el voltaje en la carga a un valor deseado, el cual es inferior al voltaje de entrada.

En la figura 2.1 se muestra el circuito del convertidor reductor, que se compone de un dispositivo de conmutación (Mosfet), voltaje de entrada V_i , un diodo (D), un inductor

(L), un capacitor (C) y una carga (R). Las variables del sistema corresponden a el voltaje en el capacitor V_c y corriente en la inductancia I_L .

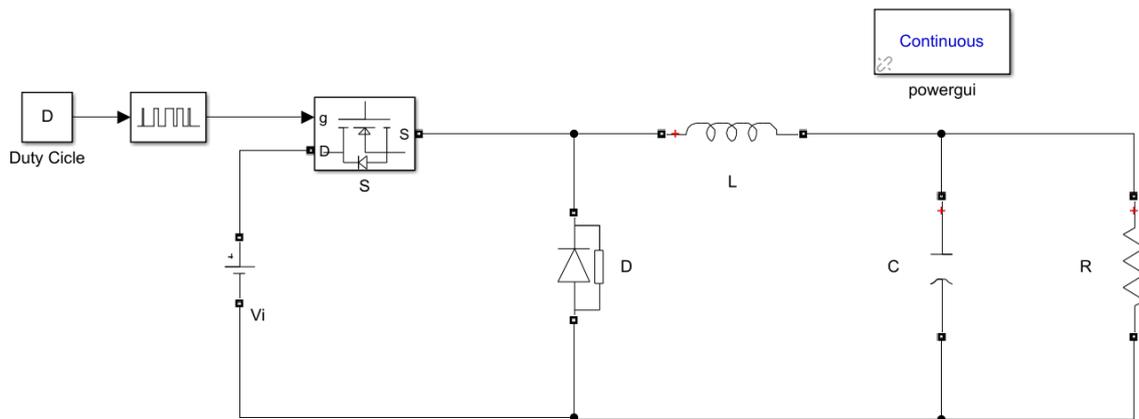


Figura 2-1 Diagrama circuito del convertidor reductor (Buck). Fuente

Utilizando las leyes de Kirchoff, las ecuaciones que describen el comportamiento del convertidor reductor (Buck) son:

$$\begin{bmatrix} \dot{i}_L \\ \dot{V}_c \end{bmatrix} = \begin{bmatrix} 0 & -\frac{1}{L} \\ \frac{1}{C} & -\frac{1}{RC} \end{bmatrix} \begin{bmatrix} i_L \\ V_c \end{bmatrix} + \begin{bmatrix} \frac{1}{L} \\ 0 \end{bmatrix} v_{in} \quad (1)$$

$$\begin{bmatrix} \dot{i}_L \\ \dot{V}_c \end{bmatrix} = \begin{bmatrix} 0 & -\frac{1}{L} \\ \frac{1}{C} & -\frac{1}{RC} \end{bmatrix} \begin{bmatrix} i_L \\ V_c \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} v_{in} \quad (2)$$

Los parámetros del convertidor R, L y C se pueden determinar a partir de especificaciones de diseño: voltaje de entrada, voltaje de salida (V_o), potencia (P), frecuencia de conmutación f_s y porcentaje de rizado r , a partir de las siguientes expresiones [28]:

$$I_{out\ max} = \frac{P}{V_o} \quad (3)$$

$$\Delta I_L = 0.01 * I_{out\ max} \quad (4) \quad \Delta V_o = 0.01 * v_o \quad (5)$$

$$L = \frac{V_o(v_i - v_o)}{\Delta I_L * f_s * v_i} \quad (6)$$

$$C = \frac{\Delta I_L}{8 * \Delta V_o * f_s} \quad (7)$$

$$R = \frac{V_o}{I_{out\ max}} \quad D = \frac{v_o}{v_i} \quad (8)$$

A continuación, se presenta el diseño de un convertidor reductor con las siguientes características: $v_i = 70V$, $P = 950W$, $f_s = 10kHz$, $r = 1\%$ y $v_o = 50V$.

Utilizando las ecuaciones (3-8) se tiene: $L=0.0075H$; $C=0.475\mu F$ $R=2.6316\Omega$ y $D = 0.7143$. El voltaje de salida del convertidor se muestra en la figura 2-2, donde se observa que llega al valor de salida establecido.

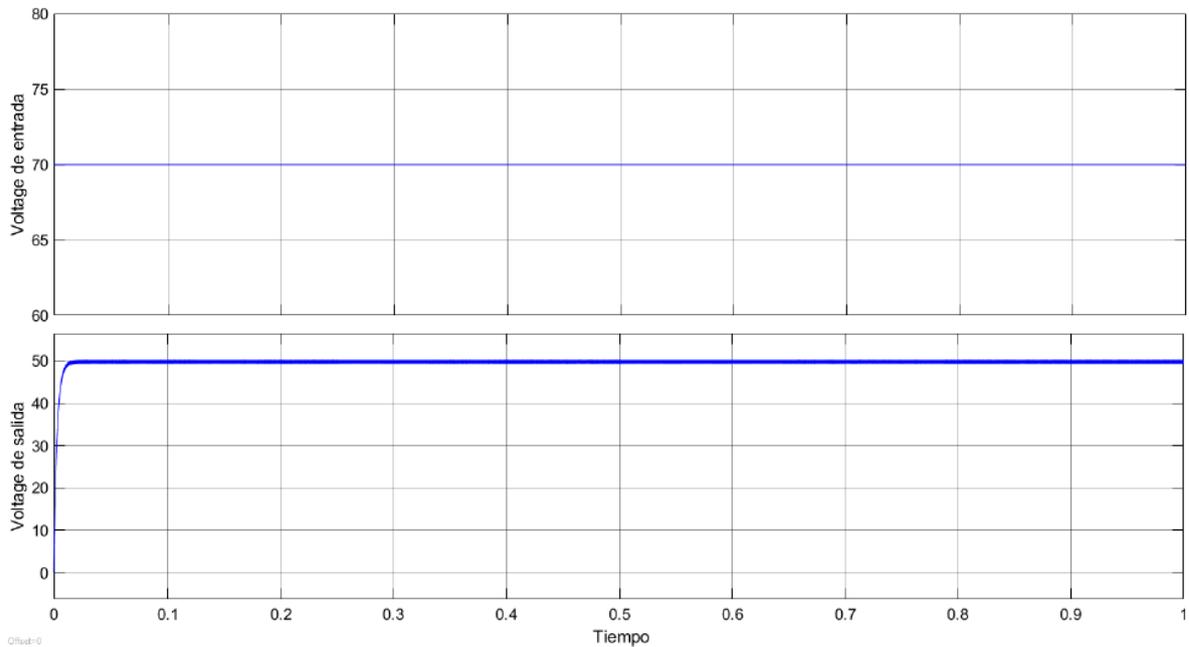


Figura 2-2 Evolución temporal de los voltajes de entrada y salida del convertidor reductor, para los parámetros de diseño especificados. Fuente Propia.

2.2 Convertidor de potencia elevador (Boost)

En el convertidor elevador (Boost), se logra obtener un voltaje de salida mayor que el voltaje de entrada. Este voltaje de salida es continuo, pero puede presentar una pequeña fluctuación conocida como rizado, el cual está influenciado por los valores de inductancia y capacitancia, así como la frecuencia de conmutación [26]. Como resultado, el convertidor elevador produce niveles de voltaje más altos en la salida.

Las aplicaciones más comunes del convertidor elevador (Boost) se encuentran en el ámbito de fuentes de potencia y sistemas fotovoltaicos, donde se utilizan como etapas iniciales de acondicionamiento [27]. Estos convertidores desempeñan un papel crucial en la adaptación de la energía generada por fuentes renovables, como la energía solar, y permiten elevar el voltaje de entrada para alimentar cargas de mayor voltaje de manera eficiente.

En la figura 2.3 se muestra el circuito del convertidor elevador (Boost) Los parámetros del convertidor R , L y C se pueden determinar a partir de especificaciones de diseño: voltaje de entrada, voltaje de salida (V_o), potencia (P), frecuencia de conmutación f_s y porcentaje de rizado r , a partir de las siguientes expresiones [29]:

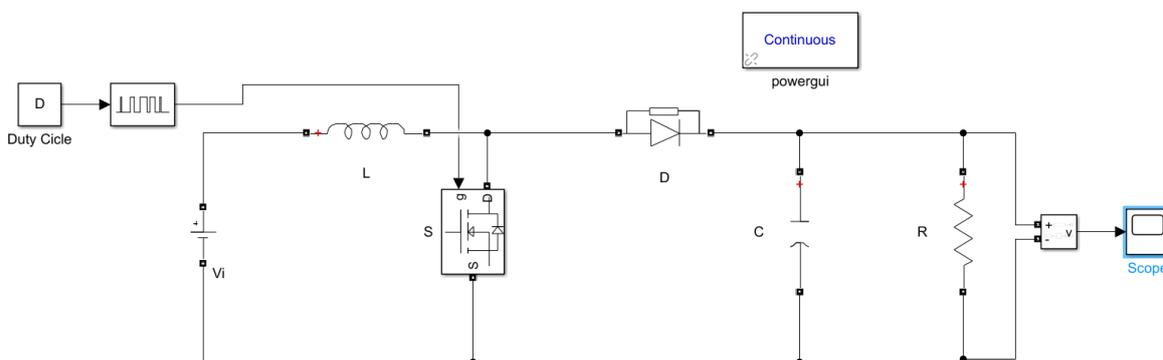


Figura 2-3 Diagrama circuito del convertidor elevador (Boost) Fuente. [propia]

Utilizando las leyes de Kirchhoff, las ecuaciones que describen el comportamiento del convertidor elevador (Boost) son:

$$\begin{bmatrix} \dot{v} \\ \dot{i}_L \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} V_c \\ i_L \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ -\frac{1}{RC} & \frac{1}{C} \end{bmatrix} \begin{bmatrix} V_c \\ i_L \end{bmatrix} \quad (9)$$

$$\begin{bmatrix} \dot{v} \\ \dot{i}_L \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} V_c \\ i_L \end{bmatrix} + \begin{bmatrix} 0 \\ -\frac{1}{RC} \end{bmatrix} V_c + \begin{bmatrix} 0 \\ \frac{1}{C} \end{bmatrix} \quad (10)$$

Los parámetros del convertidor R, L y C se pueden determinar a partir de especificaciones de diseño: voltaje de entrada, voltaje de salida (V_o), potencia (P), frecuencia de conmutación f_s y porcentaje de rizado r , a partir de las siguientes expresiones [29]:

$$I_{out\ max} = \frac{P}{V_o} \quad (11) \quad \Delta I_L = 0.01 * I_{out\ max} * \frac{V_o}{V_i} \quad (12) \quad \Delta V_o = 0.01 * v_o \quad (13)$$

$$L = \frac{V_i(V_o - V_i)}{\Delta I_L * f_s * V_o} \quad (14) \quad C = \frac{I_{out\ max} \left(1 - \frac{V_i}{V_o}\right)}{\Delta V_o * f_s} \quad (15) \quad R = \frac{V_o}{I_{out\ max}} \quad (16)$$

A continuación, se presenta el diseño de un convertidor elevador (Boost) con las siguientes características:

$V_i = 12V$, $P = 5009\ W$, $f_s = 10\ kHz$, $r = 1\%$ y $V_o = 24$ Utilizando las ecuaciones (11-16) se tiene: $L = 0.014\ H$; $C = 0.0043\ F$, $R = 1.1520\ \Omega$ y $D = 0.5$ El voltaje de salida del convertidor se muestra en la figura 2-4, donde se observa que llega al valor de salida establecido.

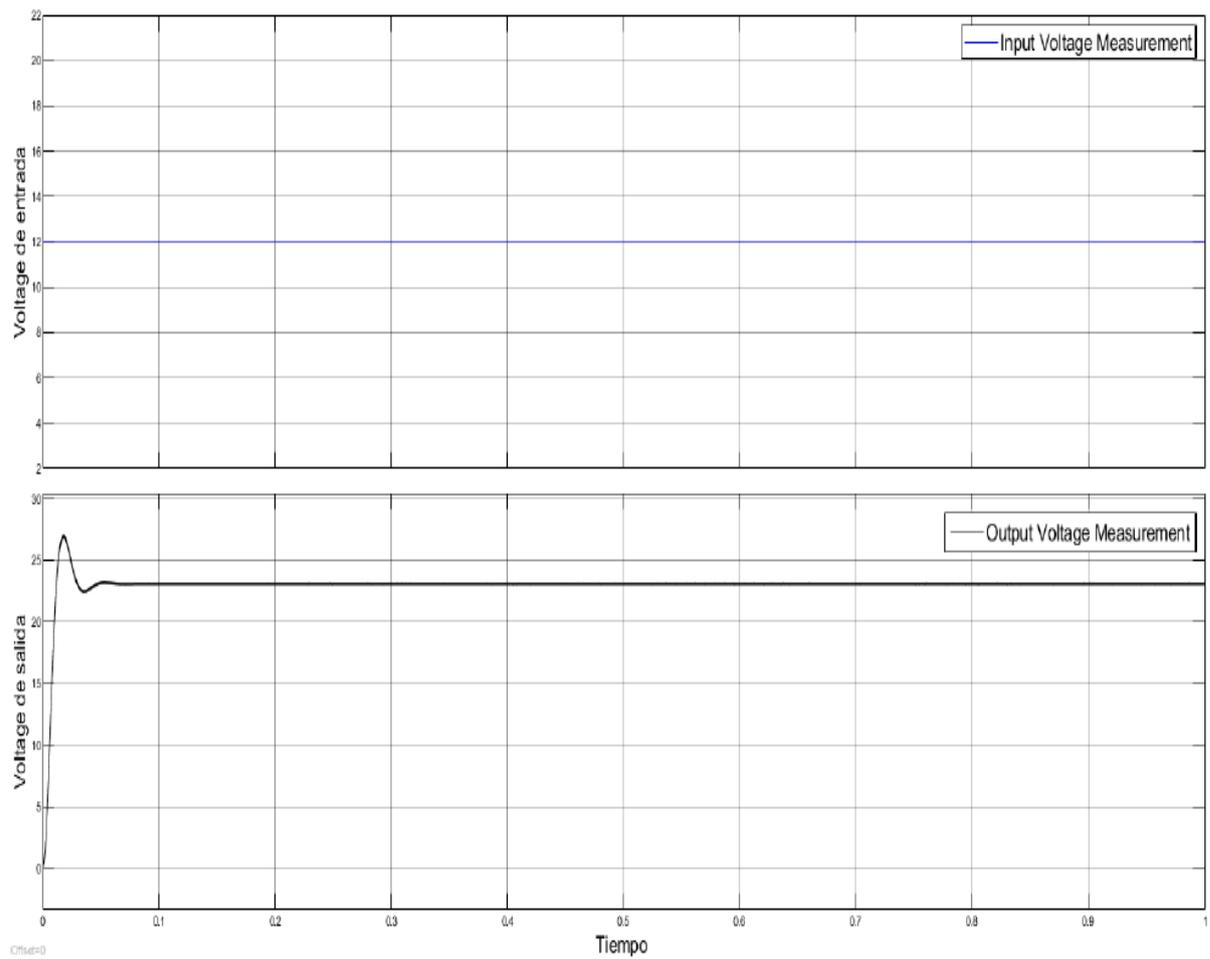


Figura 2.4 Convertidor elevador (Boost) Lazo abierto Voltaje de entrada Vs Voltaje de salida. Fuente propia.

2.3 Aprendizaje por refuerzo

El aprendizaje por refuerzo es un subconjunto de aprendizaje automático (machine learning) en la que un agente toma decisiones basadas en la retroalimentación que recibe del ambiente en el que se encuentra. Este proceso se basa en la idea de que el agente aprende a maximizar una recompensa a través de la exploración de diferentes acciones. En la figura 1-1 se ilustra esta clasificación [44].

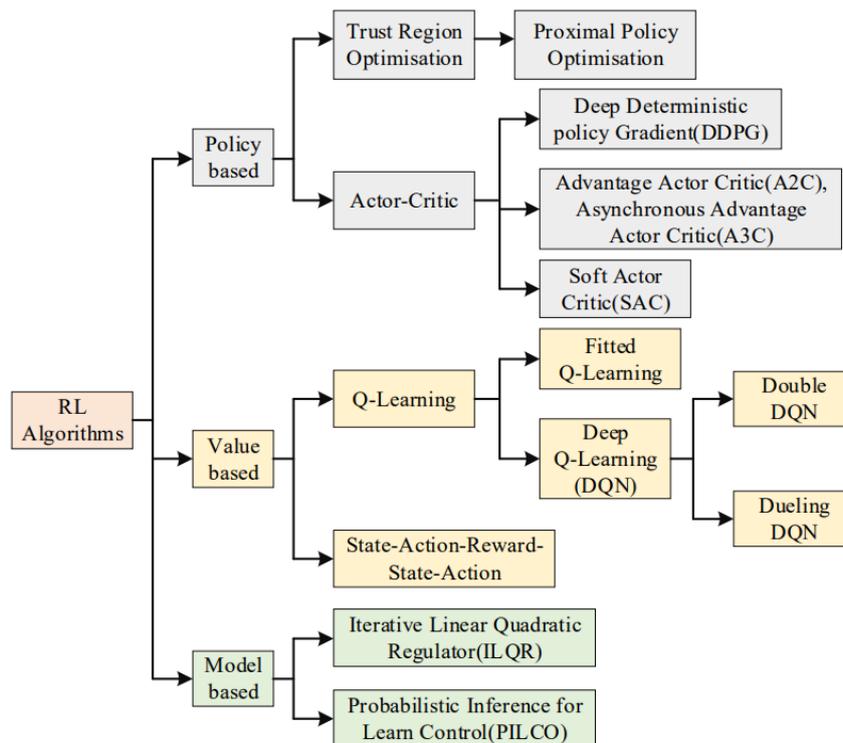


Figura 2-4 Clasificación de algoritmos RL

El agente recibe observaciones y recompensas del entorno y responde enviando sus acciones. La recompensa indica qué tan exitosa es una actividad en términos de lograr el objetivo. Este proceso de interacción continua permite al agente aprender a tomar las decisiones correctas para maximizar su recompensa.

El aprendizaje por refuerzo tiene una amplia variedad de aplicaciones en campos como la robótica, los videojuegos, la automatización de procesos, la optimización de recursos energéticos, el control de tráfico, la gestión de carteras, entre otros. Por ejemplo, se han desarrollado robots capaces de aprender a caminar, correr y trepar mediante técnicas de aprendizaje por refuerzo [31]. En el ámbito de los videojuegos, se han utilizado algoritmos de aprendizaje por refuerzo para entrenar agentes que pueden jugar a juegos de estrategia, como el ajedrez y el go [32]. También se ha utilizado el aprendizaje por refuerzo en la optimización de procesos industriales, como la fabricación de semiconductores [33], y en la gestión de tráfico aéreo, para mejorar la eficiencia del sistema de control de tráfico aéreo [34]

2.3.1 Descripción del aprendizaje por refuerzo

Los elementos clave, para el aprendizaje por refuerzo (Reinforcement Learning) son: el agente, el entorno, las acciones, las observaciones, la política y la función de recompensa. El agente es el sistema que toma decisiones y aprende a partir de las interacciones con el entorno. Existen diferentes algoritmos y métodos para el aprendizaje por refuerzo, como el Q-learning, la aproximación de funciones de valor y la búsqueda de políticas. Estos algoritmos se basan en matemáticas y estadísticas para aprender a partir de las interacciones del agente con el entorno.

2.3.2 Agente

El agente es un componente fundamental en el aprendizaje por refuerzo, cuyo objetivo es maximizar la recompensa obtenida en el tiempo mediante la elección de acciones apropiadas en función del estado actual del entorno y la política asignada. El agente aprende a través de la teoría de los procesos de decisión de Markov y la ecuación de Bellman, que permiten modelar la interacción con el entorno y calcular el valor esperado de la recompensa. Además, técnicas como TD Learning y Q-Learning se utilizan para mejorar el rendimiento del agente en la exploración y elección de acciones. El proceso de entrenamiento del agente implica la exploración y prueba y error para

maximizar la recompensa [38]. El diagrama de flujo de las interacciones agente-ambiente se puede encontrar en la Figura 2-6.

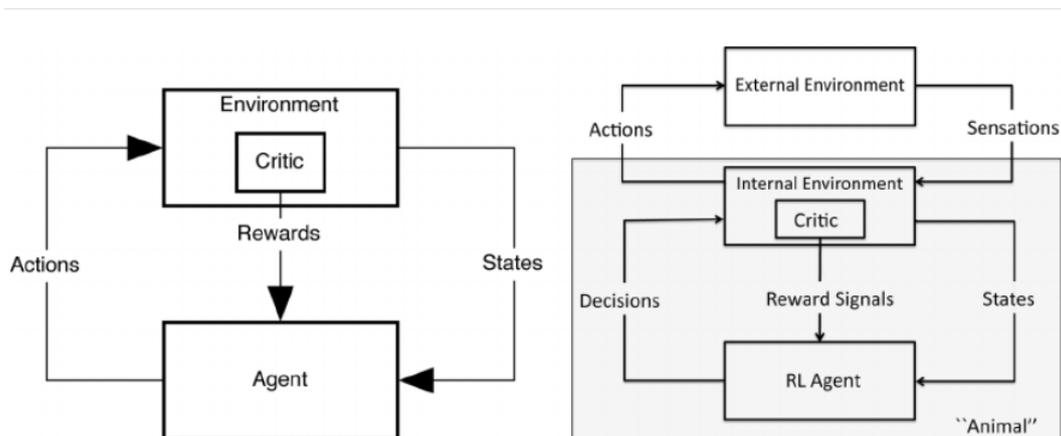


Figura 2-5 interacciones agente-ambiente en el aprendizaje por refuerzo

2.3.3 Entorno

El entorno es todo lo externo del sistema, donde actúa el agente envía acciones y obtiene recompensas o penalizaciones, también es donde se entrena al agente, su comportamiento suele ser desconocido y estocástico [35]. Dentro del entrenamiento del agente es importante resaltar del entorno su funcionamiento partiendo de estado-selección de algoritmo-mejor, acción-recompensa, la dinámica del entorno en el aprendizaje por refuerzo (Reinforcement Learning) representa la mayor parte del sistema, el agente es una parte del software que genera las acciones y actualiza constantemente la política a través del aprendizaje [37]. A partir de la definición del entorno, en el cual va a operar el agente, sumando la interfaz de la relación agente y el entorno basado en un entorno de simulación, su seguridad y espacio a experimentar representan una ventaja en el diseño [37].

El entorno cambia de estado en respuesta a la acción del agente y produce una recompensa por esa acción, que es recibida por el agente [38]. Usando esta nueva información, el agente puede determinar si esa acción fue buena y debe repetirse, o si fue mala y debe evitarse. El ciclo observación-acción-recompensa continúa hasta que se completa el aprendizaje [40]Figura 2-7.



Figura 2-6 Entorno ambiente [11]

Cuando crea un objeto de entorno, debe especificar las señales de acción y observación que utiliza el agente para interactuar con el entorno. Las señales que seleccione como acciones y observaciones dependen de su aplicación. Por ejemplo, para aplicaciones de sistemas de control, las integrales (ya veces las derivadas) de las señales de error suelen ser observaciones útiles. Además, para aplicaciones de seguimiento de referencia, es útil tener una señal de referencia variable en el tiempo como observación [37].

2.3.4 Observación del entorno

En el aprendizaje por refuerzo (Reinforcement Learning), El agente es capaz de observar el estado actual del entorno a través de una señal de observación, y a partir de ese estado decide qué acción tomar que será enviada al entorno a través de una señal de acción [34]. Esta información se conoce como observaciones y puede tomar muchas formas, dependiendo del dominio en el que se esté aplicando el aprendizaje por refuerzo. Por ejemplo, en el caso de un robot que aprende a caminar, las observaciones pueden incluir la posición de las articulaciones del robot y la velocidad a la que se mueve. En un juego de ajedrez, las observaciones pueden ser el estado actual del tablero y las posibles jugadas que el agente puede realizar [36].

En algunos casos, las observaciones pueden ser proporcionadas directamente por sensores físicos, como cámaras o micrófonos [37]. En otros casos, las observaciones pueden ser generadas por un modelo del entorno, que se actualiza continuamente a medida que el

agente interactúa con él [38]. En la figura 2.7 describe un agente, y las observaciones en un modelo basado en aprendizaje por refuerzo(Reinforcement Learning)

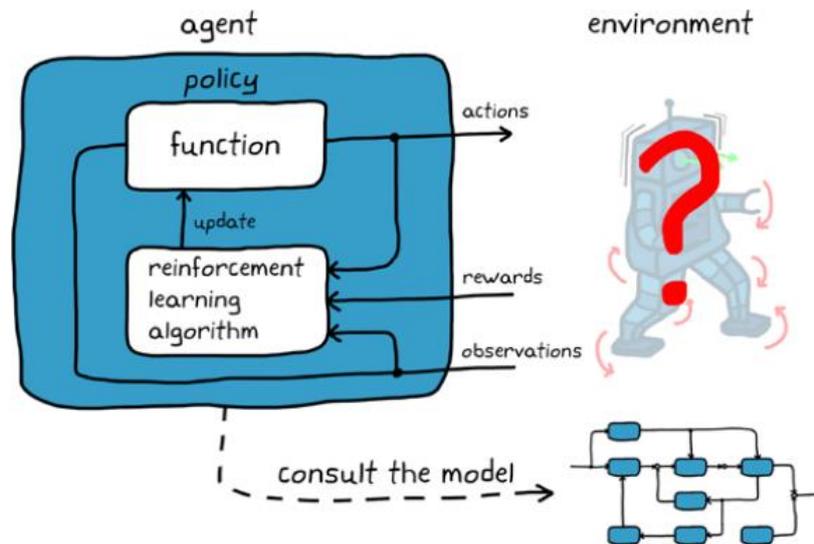


Figura 2-7 Aprendizaje por refuerzo basado en modelos []

2.3.5 Función de recompensa

La función de recompensa es fundamental para que el agente pueda medir su rendimiento y lograr sus objetivos en el entorno de trabajo. Esta señal es proporcionada por el entorno de trabajo, y puede consistir en recompensas numéricas o señales que indican si el agente ha realizado una acción correcta o incorrecta. [41], Figura 2-8.

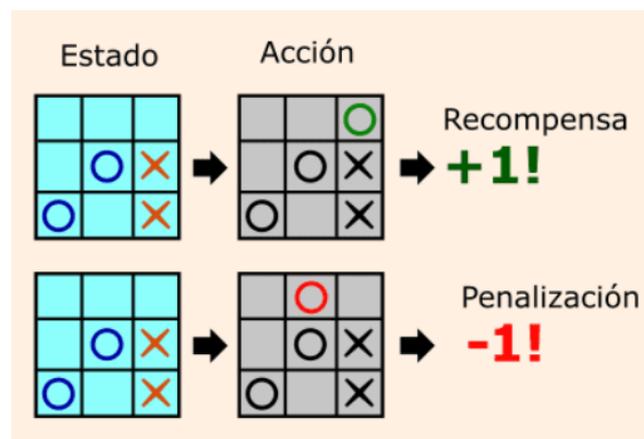


Figura 2-8 Ejemplo acción de recompensa [10]

Es importante señalar que la forma de dar forma a la señal de recompensa puede ser complicada y requerir varias iteraciones hasta lograr el resultado deseado. Además, al establecer la recompensa, es necesario considerar si hay más de una señal de entrada, en cuyo caso se debe parametrizar el peso de cada señal de manera individual. Esto también se aplica al establecer los pesos entre las conexiones de una red neuronal para determinar la incidencia en el resultado acumulado según la señal. [42].

2.3.6 Política de estrategia

La política de asignación de comportamiento o reglas para el agente es fundamental para lograr que éste busque maximizar la recompensa recibida por realizar satisfactoriamente las tareas asignadas. Para encontrar la política más acertada se hace uso de algoritmos de aprendizaje que actualizan constantemente los parámetros de la función de política, a fin de maximizar la recompensa recibida a largo plazo. [41]

La función de política se compone de lógica y parámetros ajustables dentro de una estructura adecuada, que conducen a un conjunto de parámetros que producen un desarrollo óptimo en dicha política, lo cual se logra mediante un mapeo de estados, a las acciones que producen la mayor recompensa a largo plazo. Debido a que el entorno puede cambiar con el tiempo, la política no es un mapeo estático, sino que debe ser constantemente actualizada y ajustada. [42]

Matemáticamente, la política se puede expresar como una función que toma como entrada la observación del estado y produce una acción:

$$\text{Acciones} = \text{función} (\text{Observación del estado})$$

Sobre los parámetros ajustables de esta función se realiza el aprendizaje, que consiste en ajustar sistemáticamente dichos parámetros hasta obtener la mejor política posible. [43]

3. Desarrollo Metodológico

En este capítulo se desarrolla los objetivos del control inteligente mediante el uso de una red neuronal, utilizando un agente DDPG (Deep Deterministic Policy Gradient) para controlar un convertidor Buck y un convertidor Boost. La implementación de un algoritmo de control de aprendizaje por refuerzo (Reinforcement Learning) como DDPG, ya que se utiliza comúnmente en entornos de control continuos, el agente DDPG utiliza dos redes neuronales: una red actor que toma el estado actual del sistema como entrada y genera una acción de control de salida, y una red crítica que estima el valor Q de la acción de control en el estado actual. Estas redes se entrenan simultáneamente mediante retroalimentación de error de RL para mejorar la política de control del agente.

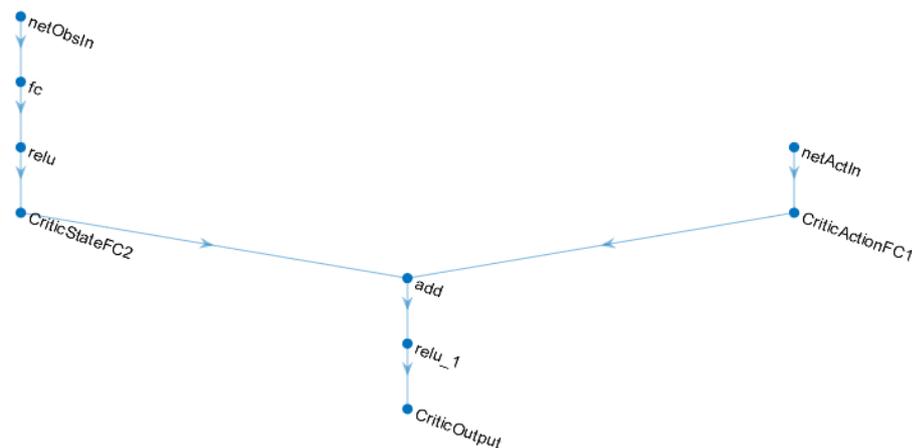


Figura 3.1 Red Neuronal Actor – Crítico Fuente Propia

El objetivo del agente DDPG es maximizar una función de recompensa que se define en función del comportamiento deseado del sistema de convertidor elevador (Boost) y convertidor reductor (Buck), en la figura 3-2. se muestra el escenario general donde se encuentran los bloques por subsecciones, con sus respectivas configuraciones y parámetros, cálculos de recompensa, penalización y el entorno donde realiza entrenamiento el agente.

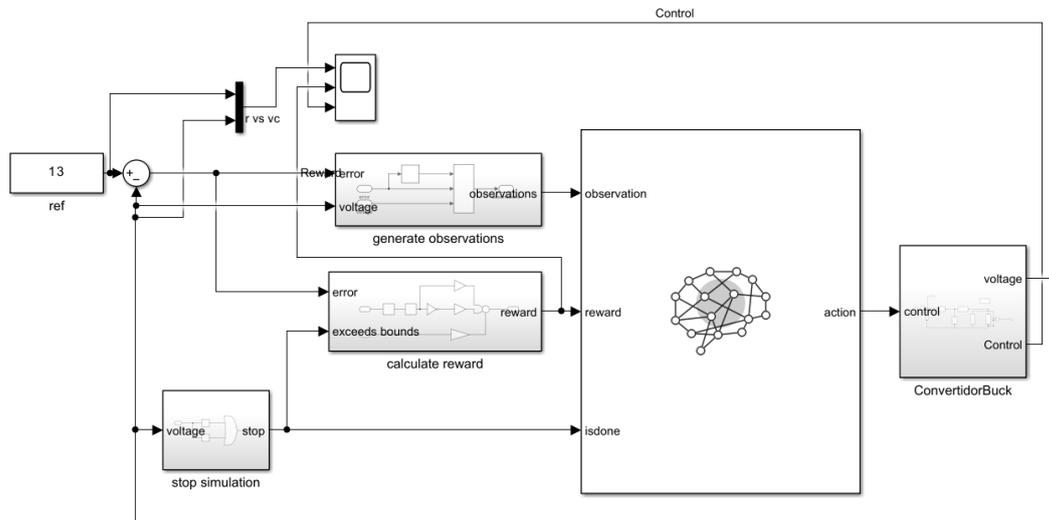


Figura 3-1 entorno convertidor reductor (Buck) fuente propia

Para lograr este objetivo se plantean estrategias de asignación de recompensas y las penalizaciones para el correcto aprendizaje del agente en un entorno de aprendizaje por refuerzo para los convertidores de potencia reductor (Buck) y convertidor elevador (Boost) respectivamente creado en Simulink, así como el proceso de validación de la señal de referencia mediante observación y acciones con base a los resultados del entrenamiento. Finalmente, la respectiva validación del agente donde se compara al controlador inteligente con técnicas de control clásicas PID.

3.1 Definición del entorno de los convertidores Buck

Para el diseño del entorno se crea un modelo del convertidor reductor con los elementos de circuito de la librería de Simscape de Simulink, que representa los elementos principales del circuito que son, una fuente de entrada tipo panel solar, un transistor de potencia tipo MOSFET, un capacitor, un inductor, una resistencia y un bloque PWM para controlar el ciclo de trabajo, la frecuencia de operación del convertidor se mantendrá constante, se ajustará únicamente el ciclo de trabajo para controlar el voltaje de salida del convertidor como se observa en la figura 3.2.

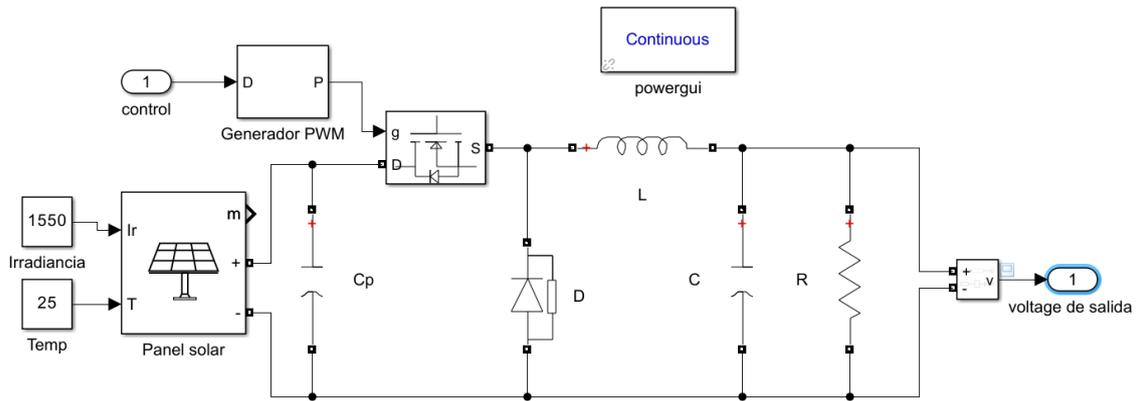


Figura 3-2 Diagrama del circuito del convertidor reductor – entorno. Elaboración propia

El control del voltaje se fundamenta en la medición del error de voltaje, y para lograrlo, se implementará un agente DDPG (Deep Deterministic Policy Gradient) que se encargará de ajustar el ciclo de trabajo del bloque PWM con el objetivo de obtener un voltaje de salida ajustado. Para llevar a cabo este proceso, se utilizará un entorno de simulación que permitirá al agente entrenarse y evaluar su desempeño en la tarea de controlar el voltaje deseado.

Dentro del desarrollo para la definición de un entorno se debe incluir una señal de acción y una señal de observación mediante la cual el agente en el circuito del convertidor Buck y convertidor Boost, interactúa con los estados de conmutación ON y OFF, donde definida la acción de recompensa el agente la utiliza para medir el correcto aprendizaje dentro del entorno y sus especificaciones de datos de observación o acción continua para entornos de aprendizaje por refuerzo.

3.1.1 Definición del entorno de los convertidores Boost

El entorno del convertidor elevador (Boost), su configuración utiliza los mismos parámetros en los elementos de circuito, (Fuente de energía renovable, Resistencia, Diodo, inductor, capacitor e IGBT) pero de una configuración diferente ya que como se ha mencionado anteriormente el convertidor DC-DC elevador (Boost), su funcionamiento es elevar el voltaje mediante la carga y descarga de la inductancia de la bobina. De igual forma que en el convertidor Buck, este convertidor tendrá un bloque de control de ciclo de trabajo por medio de un PWM realizado en Simulink. Como se observa en la figura 3.3.

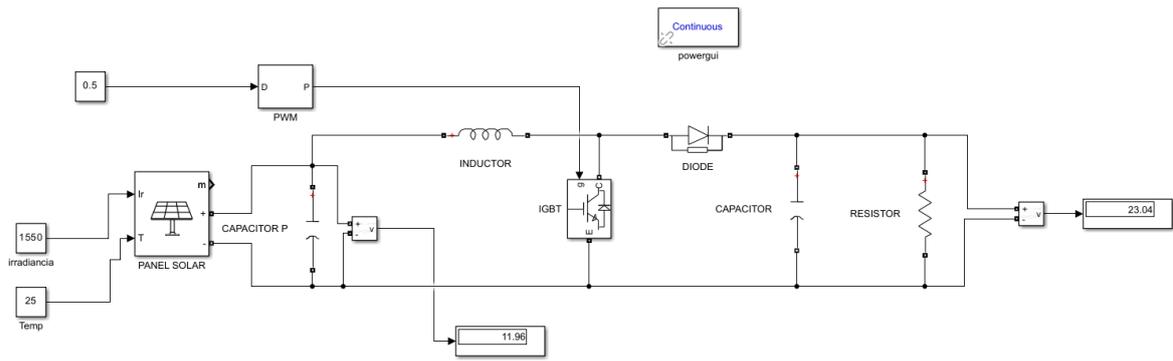


Figura 3-3 Entorno controlador elevador (Boost) Fuente. Elaboración propia

3.2 Diseño de las funciones de recompensa para convertidor reductor elevador

Con el entorno establecido de los convertidores, el siguiente paso es pensar en lo que quiere que haga su agente y cómo lo recompensará por hacer lo que quiere, Esto requiere crear una función de recompensa para que el algoritmo de aprendizaje "entienda" cuándo la política está mejorando y, en última instancia, converja. en el resultado buscado. Para el ajuste de la recompensa del controlador se utiliza parámetros de recompensa de 10, y una penalización de -1, estos parámetros se irán ajustando al agente según el observador de estado para que dicho agente vaya aprendiendo y se ajuste a la señal de referencia, para esto se realiza una configuración enlazada por diagramas de bloques de Simulink en la red neuronal ajustando así las penalizaciones y recompensas como se muestra en la figura 3.4.

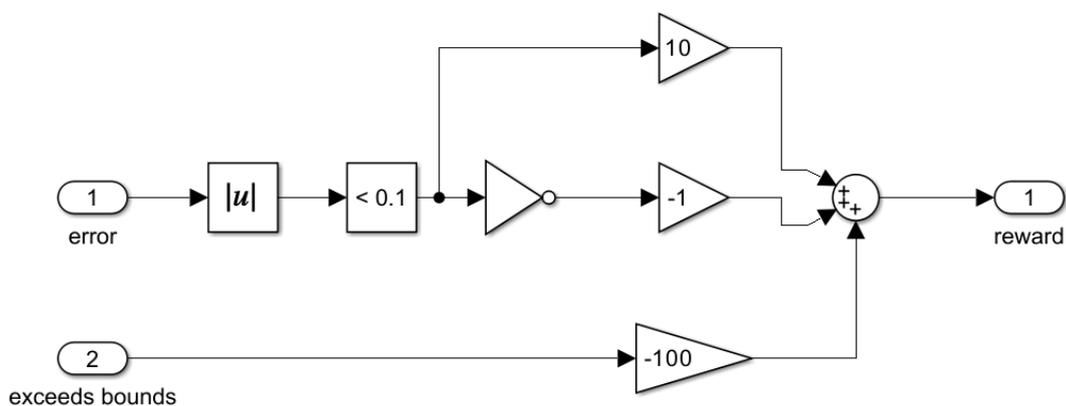


Figura 3-4 Bloques recompensa y penalización Fuente. Elaboración propia

En el bloque para la definición de la recompensa, se utiliza un algoritmo de penalización y recompensa basando en un vector de observación descrito matemáticamente mediante la ecuación:

$$[\int e dt \quad y \quad h]^T \quad \text{ecuación 3.1}$$

El vector de observación utilizado incluye la integral del error, el error y el voltaje medido. Estos elementos se seleccionaron para proporcionar al agente DDPG información completa sobre el estado del sistema. La integral del error permite considerar la acumulación de errores pasados, mientras que el error actual y el voltaje medido brindan información instantánea sobre la diferencia entre la señal de referencia y el valor deseado. Se estableció una penalización numérica de 10 cuando el error del voltaje está por debajo de 0.1. Esta elección se basa en la importancia de mantener el error del voltaje dentro de un rango aceptable. Al asignar una penalización alta, se motiva al agente a reducir el error del voltaje y a tomar acciones que lo mantengan cerca del valor deseado. Cuando la señal de error de voltaje es mayor a 0.1, se otorga una penalización de -1. Esta decisión se basa en el deseo de evitar errores significativos en el control del voltaje. Al proporcionar una penalización negativa, el agente aprenderá a evitar acciones que conduzcan a un aumento del error y, en cambio, buscará soluciones que reduzcan la discrepancia entre la señal de referencia y el valor deseado.

Si el voltaje excede los límites configurados en el bloque de stop, (límite superior 26, límite inferior 0). Se le otorgara una penalización de -100 como se muestra en la ecuación 3.2. esto con el fin de controlar los límites de exploración del agente.

$$Reward \begin{cases} 10 \text{ si el error es } < 0.1 \\ -1 \text{ si el error es } > 0.1 \\ -100 \text{ si el voltaje exede los limites} \end{cases} \quad \text{ecuación 3.2}$$

3.3 Entrenamiento del agente DDGP para un convertidor reductor (Buck)

Para lograr un entrenamiento deseado y que el agente aprendiera se ajustaron hiperparametros del agente como fueron la tasa de aprendizaje, exploracion y demas, haciendo que el agente en medio de diferentes iteraciones de episodios explore y encuentre la señal deseada, dando así un aprendizaje optimo, para el entrenamiento se utilizaron los siguientes hiperparametros tabla 1.

HIPER PARAMETROS	VALORES
TARGETSMOOTHFACTOR	1e-3
SAMPLETIME	0.1
DISCOUNTFACTOR	0.4
MINIBATCHSIZE	256
EXPERIENCEBUFFERLENGTH	1e6
NOISEOPTIONS.VARIANCE	0.8
NOISEOPTIONS.VARIANCEDECAYRATE	1e-4
CRITICOPTIMIZEROPTIONS.	1
CRITICOPTIMIZEROPTIONS.LEARNRATE	1e-3

Tabla 1 Hiperparametros de entrenamiento convertidor Buck

los hiperparámetros fueron seleccionados considerando aspectos como la estabilidad del entrenamiento, el equilibrio entre la exploración y la explotación, y la convergencia óptima del agente hacia los objetivos deseados. Cada valor fue ajustado de manera cuidadosa para lograr un entrenamiento eficiente y un aprendizaje óptimo del agente en el convertidor Reductor (Buck). Estos hiperparametros se fueron ajustando respecto a la respuesta del agente con base al entrenamiento y la búsqueda de la señal de referencia.

El agente aprenderá a través del cálculo del promedio de recompensas obtenidas en cada episodio, con el objetivo de maximizar dicho promedio. Este proceso de aprendizaje se ilustra en la Figura 3.5.

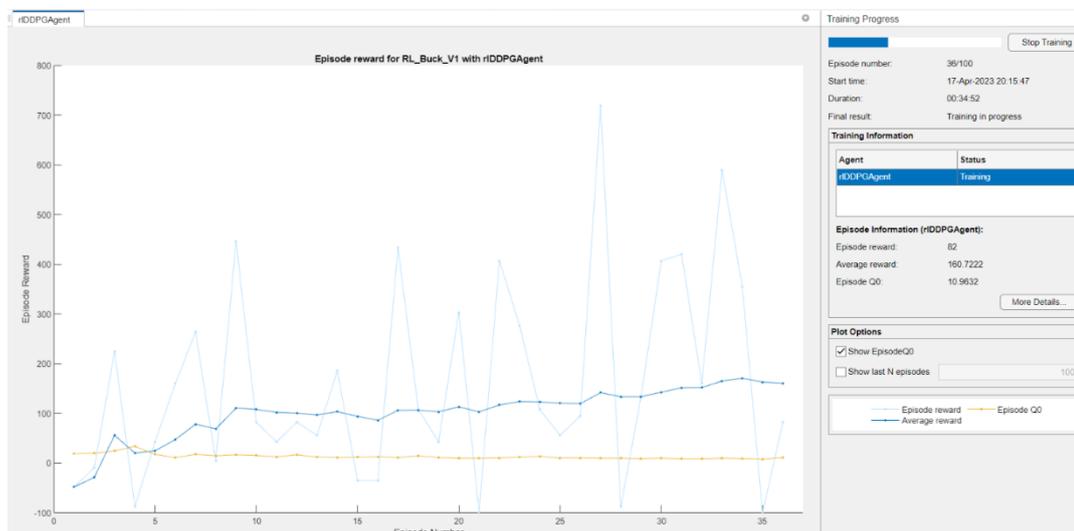


Figura 3-5 Número de episodios vs el promedio de recompensas convertidor reductor (Buck) Fuente. (Tearle, n.d.)

En la figura se puede observar cómo el agente continúa iterando a lo largo de múltiples episodios, actualizando sus acciones y mejorando su desempeño para maximizar las recompensas acumuladas. El promedio de recompensas obtenido en cada episodio es utilizado como una medida de evaluación para monitorear el progreso del aprendizaje del agente y su capacidad para alcanzar los objetivos establecidos.

Durante el entrenamiento del agente DDPG en el convertidor reductor (Buck), se utilizan diferentes métricas para evaluar su desempeño. Estas métricas incluyen el "Episode Reward" (recompensa del episodio), que representa la cantidad de recompensa obtenida en cada episodio; el "Average Reward" (recompensa promedio), que calcula el promedio de las recompensas obtenidas en episodios anteriores; y "Episode Q0", que es la recompensa esperada a largo plazo.

En el caso específico del convertidor reductor (Buck), se estableció un objetivo de entrenamiento con un promedio máximo de recompensas de 450. Una vez que el agente alcanza y mantiene este promedio durante el entrenamiento, se detiene el proceso de aprendizaje. El propósito de esta configuración es obtener un agente DDPG entrenado, consolidado y robusto, capaz de controlar eficientemente el convertidor reductor (Buck) en diversas condiciones de operación, adaptándose al voltaje de referencia asignado.

De esta manera, el agente es entrenado con el objetivo de lograr un control preciso y estable del convertidor reductor (Buck), optimizando su desempeño y asegurando que pueda operar de manera confiable en diferentes situaciones.

3.3.1 Entrenamiento convertidor elevador (Boost).

El entrenamiento del agente DDGP del convertidor elevador (Boost) para controlar el ciclo de trabajo, demanda mayor modificación con los Hiperparametros correspondientes tales como la tasa de aprendizaje, el factor de descuento y la varianza del ruido de exploración, para mejorar su capacidad de aprendizaje y generalización.

Tabla 2

HIPER PARAMETROS	VALORES
TARGETSMOOTHFACTOR	1e-3
SAMPLETIME	Ts
DISCOUNTFACTOR	0.4
MINIBATCHSIZE	64
EXPERIENCEBUFFERLENGTH	1e6
NOISEOPTIONS.VARIANCE	0.9
NOISEOPTIONS.VARIANCEDECAYRATE	1e-6
CRITICOPTIMIZEROPTIONS.	1
CRITICOPTIMIZEROPTIONS.LEARNRATE	1e-3

Tabla 2 Hiperparametros entrenamiento convertidor Boost

El agente ira aprendiendo por medio del promedio de recompensa de los episodios y asi busca maximisar el promedio de recompensa y logre un buen entrenamiento, así eleva el voltaje de salida y lo mantiene dentro de los limites deseados, mientras se minimiza el ciclo de trabajo, en la Figura 3.6 logra minimizar la diferencia de voltaje o error entre el voltaje de referencia y lograr el objetivo de elevar el voltaje de salida y ajustarse a la señal de referencia previamente configurada.

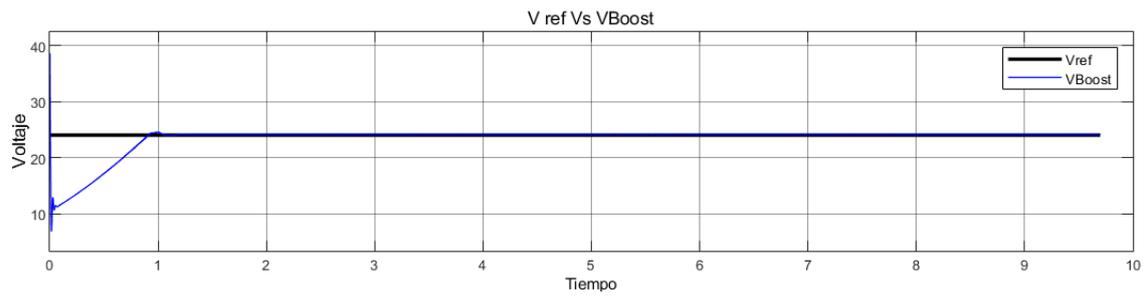


Figura 3-6 Entrenamiento convertidor elevador (Boost), ajuste V_{ref} vs V_{Boost} . Fuente. Elaboración Propia.

Dentro de las consideraciones en el proceso de entrenamiento del agente DDGP, resalta el tiempo de muestreo (T_s) y la función de transferencia (TF) del sistema.

El tiempo de muestreo es el tiempo de cada acción que está tomando el agente, y en el caso del (TF) indica cuánto tiempo lleva para que los cambios en el sistema surtan efecto después de que se han realizado las acciones del controlador.

3.4 Validación del agente reductor (Buck)

Una vez ha finalizado el entrenamiento del agente DDGP y haber logrado maximizar su recompensa promedio de 450, la validación correspondiente es consistente en variar el set point del voltaje de referencia para que así, el agente ajuste el ciclo de trabajo y logre ese voltaje deseado. para este ejemplo se utilizó un convertidor reductor (Buck) con los siguientes parámetros:

Convertidor reductor (Buck)	Valores
Voltaje entrada (V_{in})	24 V
Frecuencia	10 KHz
Inductor	10 H
Capacitor	200 μ F
Resistor	10 Ω

Tabla 3 Diseño convertidor reductor (Buck). Fuente. Elaboración Propia.

Con estos valores la validación del agente respecto al control por aprendizaje por refuerzo (Reinforcement Learning) para un convertidor reductor (Buck) obtiene una señal de voltaje que se va atenuando frente a la señal deseada en este caso, se validó el agente del controlador reductor Buck con un voltaje de entrada de 24 V, y una señal deseada de 6 V. como se muestra en la siguiente figura 3.7.

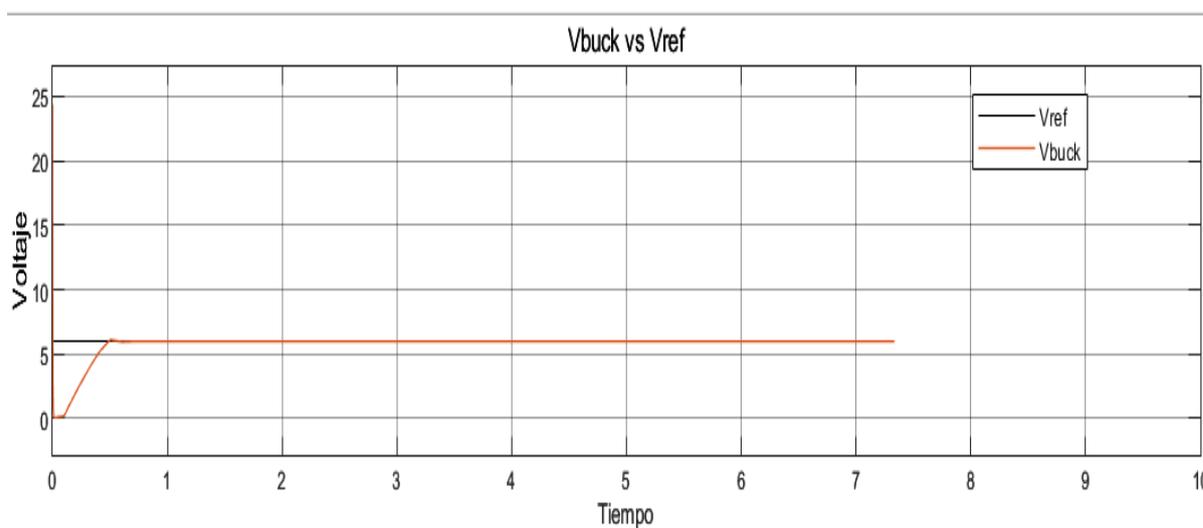


Figura 3-7 Graficas de desarrollo control aprendizaje por refuerzo controlador reductor (Buck) Fuente. Elaboración Propia.

En esta figura se representa la comparación entre la señal de voltaje de referencia deseado y el voltaje de salida generado por el convertidor. Se observa cómo el agente DDPG fue capaz de ajustar el ciclo de trabajo del bloque PWM para obtener un voltaje de salida que se aproxima al voltaje de referencia deseado.

Para la figura 3-8 siguiendo los parámetros del convertidor y la definición de recompensa establecida. Esta se asigna iterando la señal de referencia deseada con el voltaje de salida generado por el convertidor, en este contexto, se observó que cuando la señal de referencia y el voltaje de salida están en concordancia, se otorga una bonificación positiva de 10 al agente. Esto indica que el agente ha logrado un buen ajuste y control del convertidor, lo que se refleja en una recompensa favorable.

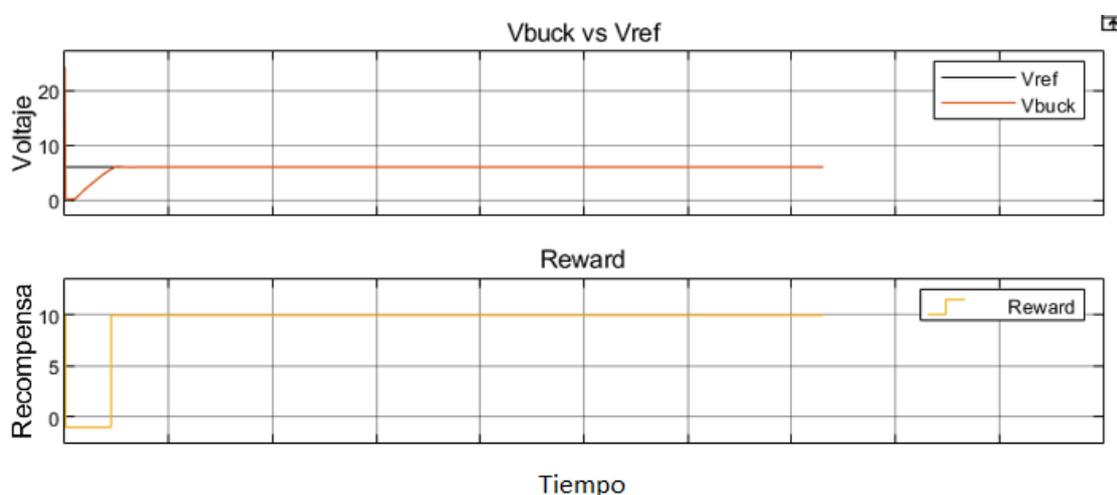


Figura 3-8 Recompensa del agente. Fuente. Elaboración Propia.

Es importante destacar que el diseño de la señal de recompensa tiene un impacto significativo en el aprendizaje del agente. Proporcionar una recompensa positiva incentiva al agente a realizar acciones que son beneficiosas para el control del convertidor. Por otro lado, la asignación de una recompensa negativa o penalización desalienta al agente a tomar acciones que podrían conducir a un desempeño deficiente.

Por último, en la figura 3-9 se presenta el ciclo de trabajo ajustado por el agente durante el proceso de control del convertidor. El ciclo de trabajo representa la proporción del período de conmutación en el cual el interruptor del convertidor está cerrado. Se espera que el agente sea capaz de ajustar el ciclo de trabajo de manera óptima para obtener el voltaje de salida deseado.

Teniendo en cuenta la figurada mencionada anteriormente se observa mientras la señal se aproxima al voltaje deseado (6 V) el actor ira ajustando el ciclo de trabajo (Duty Cycle) así logrando variar el voltaje para ir dando recompensa, al agente cada vez que cumpla la condición de los 6 V. o se alinea al voltaje de referencia.

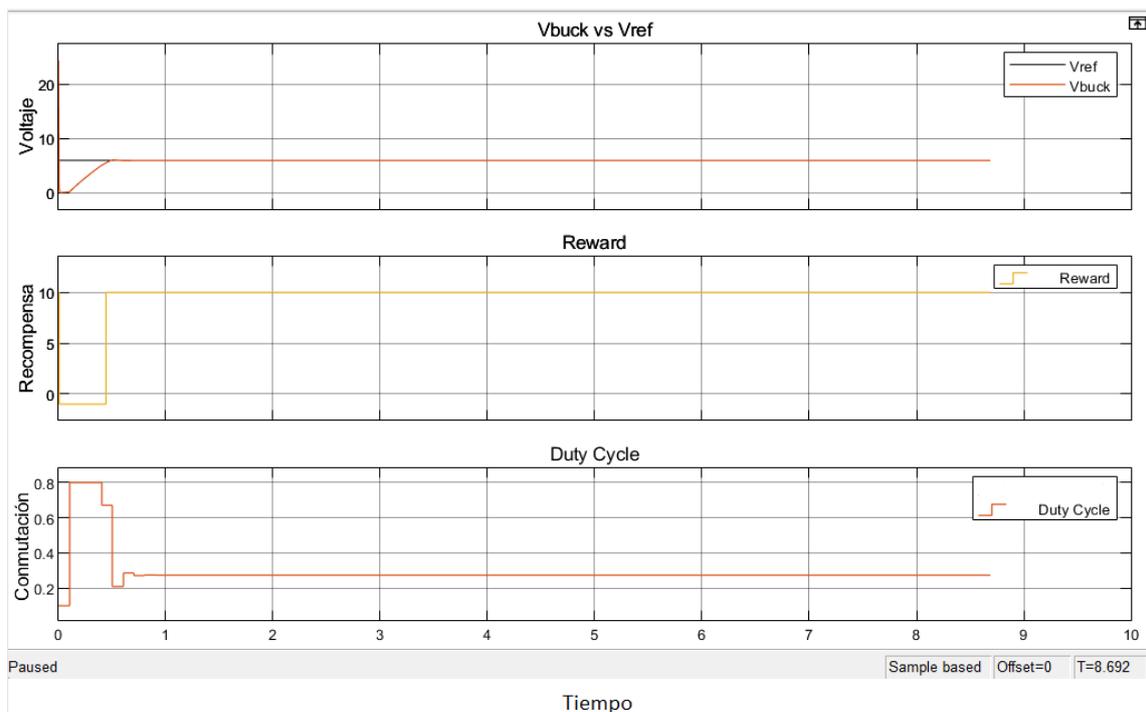


Figura 3-9 Ajuste del ciclo de trabajo. Fuente. Elaboración Propia.

En este caso, el agente realizó un ajuste del ciclo de trabajo al valor de 0.273 o 27.3%, con el objetivo de obtener una salida de voltaje de 6 V.

3.4.1 Resultados convertidores Reductor (Buck) con PID

Para evaluar la eficiencia del control mediante aprendizaje reforzado, se realizó una comparación con un controlador PID en Simulink para el convertidor reductor (Buck). En la Figura 3-11 se muestra la simulación realizada con el controlador PID, con el objetivo de comparar su desempeño con el control basado en aprendizaje reforzado.

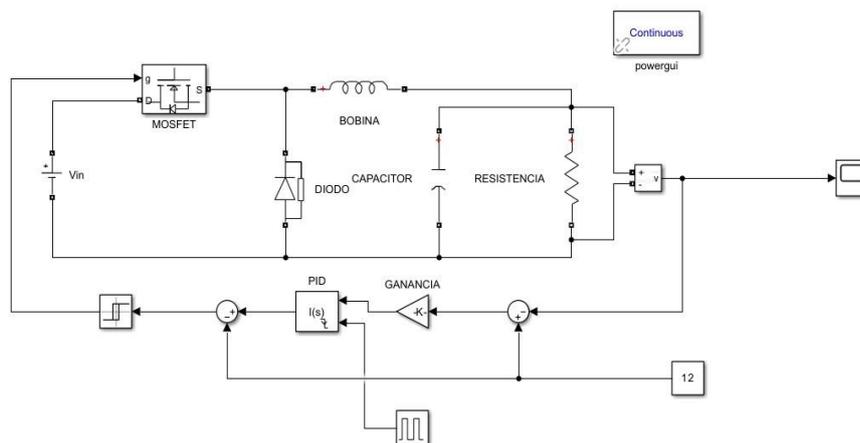


Figura 3-10 Diagrama convertidor reductor (Buck) con PID. Fuente. Elaboración Propia.

A continuación, se describen los parámetros que se utilizaron de la Tabla 3. Para el diseño de un convertidor reductor (Buck) con técnica de control PID. Resaltando que para la sintonización de las ganancias del controlador, se obtuvo un modelo lineal, a través de la herramienta System Identification, la cual utiliza métodos como Hammerstein-Wiener y ARX no lineales, procesos gaussianos (GP), máquina de vectores de soporte (SVM) y otras representaciones para estimar los modelos. También, se pueden crear modelos de ecuaciones diferenciales ordinarias (EDO) utilizando Deep Learning para capturar la dinámica de sistemas no lineales [53].

Convertidor reductor (Buck) PID	Valores
Voltaje entrada (Vin)	24 V
Frecuencia	10 KHz
Inductor	10 H
Capacitor	200 μ F
Resistor	10 Ω
Ganancia KI	200

Tabla 4 Valores Convertidor reductor con control PID

Los valores que se establecieron en el convertidor Buck reductor con aprendizaje por refuerzo son los mismos que se utilizaron para el diseño del convertidor reductor Buck con PID.

En la figura 3-12 se observa el comportamiento de voltaje de salida del convertidor reductor (Buck) con la técnica de control PID.

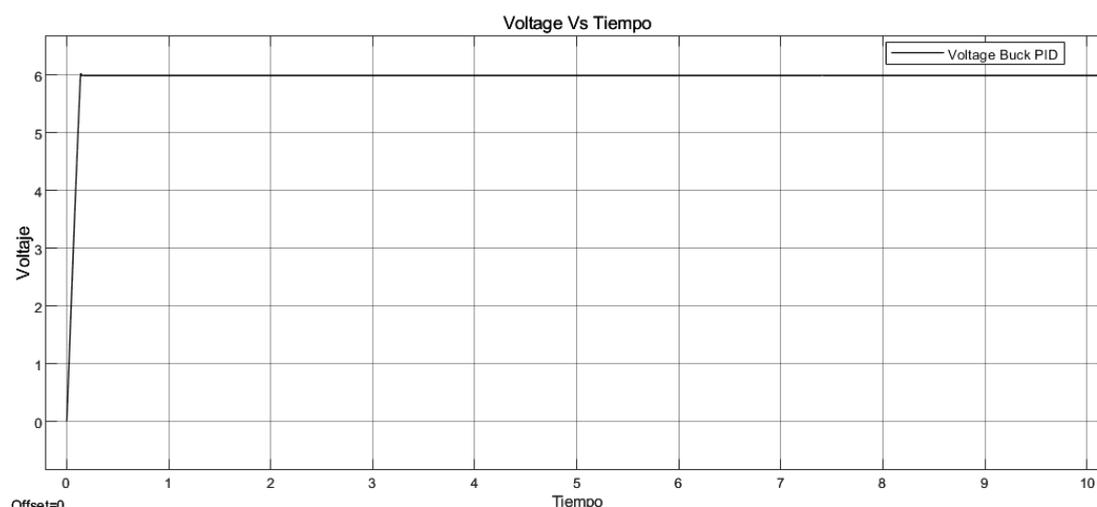


Figura 3-11 Grafica convertidor reductor (Buck) con PID entrada 24 V salida 6 V.
Fuente. Elaboración Propia

Esta comparación permite evaluar la efectividad del controlador de aprendizaje reforzado en términos de precisión, rapidez y robustez frente a perturbaciones. Con la comparación se busca demostrar las ventajas del avance de técnicas de aprendizaje por refuerzo para mejorar el control de convertidores de potencia, en particular del convertidor reductor (Buck).

3.4.2 Validación del agente elevador (Boost)

El agente se ejecuta en el entorno de prueba y la evaluación del desempeño de un agente de aprendizaje por refuerzo (Reinforcement Learning) en un entorno se evalúa su desempeño en función de su capacidad para maximizar la recompensa o minimizar la penalización Después de haber finalizado el entrenamiento del agente DDGP y haber logrado maximizar su recompensa, que fue mencionada anteriormente como 450, la asignación del valor del Setpoint de referencia y comprobar que el agente sea capaz de ajustar correctamente el ciclo de trabajo para lograr el voltaje deseado. Los valores de referencia del voltaje. La observación del agente al ajustar automáticamente el ciclo de trabajo para lograr el voltaje deseado de manera precisa y eficiente. Este resultado comprueba que el agente DDGP ha aprendido correctamente y podía aplicarse de manera efectiva en este tipo de sistemas de control.

Para validar el funcionamiento del convertidor elevador (Boost), la variación de la señal de referencia o setpoints. Estas señales son valores de voltaje que se establecen para lograr un cambio de voltaje específico en la salida del convertidor. El objetivo de utilizar diferentes Setpoints para validar el comportamiento del convertidor bajo diferentes condiciones de carga y frecuencia. En este proceso, el convertidor utiliza una inductancia y un interruptor para almacenar energía en la inductancia durante un período de tiempo determinado y luego liberarla en un ciclo de trabajo específico.

Convertidor Boost	Valores
Voltaje entrada (V_{in})	12 V
Frecuencia	10 KHz
Inductor	0.0058 H
Capacitor	0.0026 F
Resistor	2.5920 Ω

Tabla 5 Tabla parámetros convertidor elevador (Boost)

Con estos valores la validación del agente respecto al control por aprendizaje por refuerzo (Reinforcement Learning) para un convertidor elevador (Boost) obtiene una señal de

voltaje que se va atenuando frente a la señal deseada en este caso, se validó el agente del controlador elevador (Boost) con un voltaje de entrada de 12 V, y una señal deseada de 24 V. como se muestra en la siguiente figura 3-13

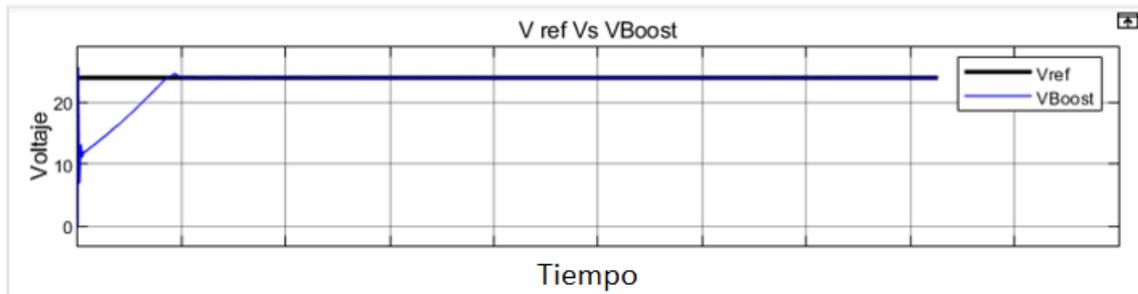


Figura 3-12 Graficas de desarrollo control aprendizaje por refuerzo controlador elevador (Boost) Fuente. Elaboración Propia.

En esta figura se representa la comparación entre la señal de voltaje de referencia deseado y el voltaje de salida generado por el convertidor. Se observa cómo el agente DDPG fue capaz de ajustar el ciclo de trabajo del bloque PWM para obtener un voltaje de salida que se aproxima al voltaje de referencia deseado.

Para la figura 3-13 siguiendo los parámetros del convertidor y la definición de recompensa establecida. Esta se asigna iterando la señal de referencia deseada con el voltaje de salida generado por el convertidor, en este contexto, se observó que cuando la señal de referencia y el voltaje de salida están en concordancia, se otorga una bonificación positiva de 10 al agente. Esto indica que el agente ha logrado un buen ajuste y control del convertidor, lo que se refleja en una recompensa favorable.

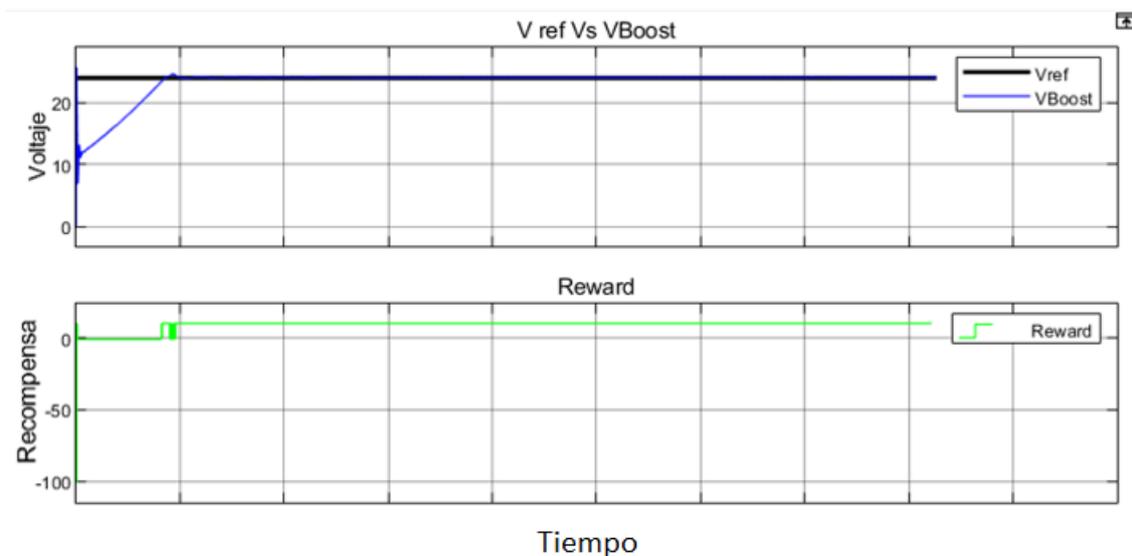


Figura 3-13 Recompensa del agente. Fuente. Elaboración Propia.

Es importante destacar que el diseño de la señal de recompensa tiene un impacto significativo en el aprendizaje del agente. Proporcionar una recompensa positiva incentiva al agente a realizar acciones que son beneficiosas para el control del convertidor. Por otro lado, la asignación de una recompensa negativa o penalización desalienta al agente a tomar acciones que podrían conducir a un desempeño deficiente.

Por último, en la figura 3-14 se presenta el ciclo de trabajo ajustado por el agente durante el proceso de control del convertidor. El ciclo de trabajo representa la proporción del período de conmutación en el cual el interruptor del convertidor está cerrado. Se espera que el agente sea capaz de ajustar el ciclo de trabajo de manera óptima para obtener el voltaje de salida deseado

Teniendo en cuenta la figurada mencionada anteriormente se observa mientras la señal se aproxima al voltaje deseado (24 V) el actor ira ajustando el ciclo de trabajo (Duty Cicle) así logrando variar el voltaje para ir dando recompensa,

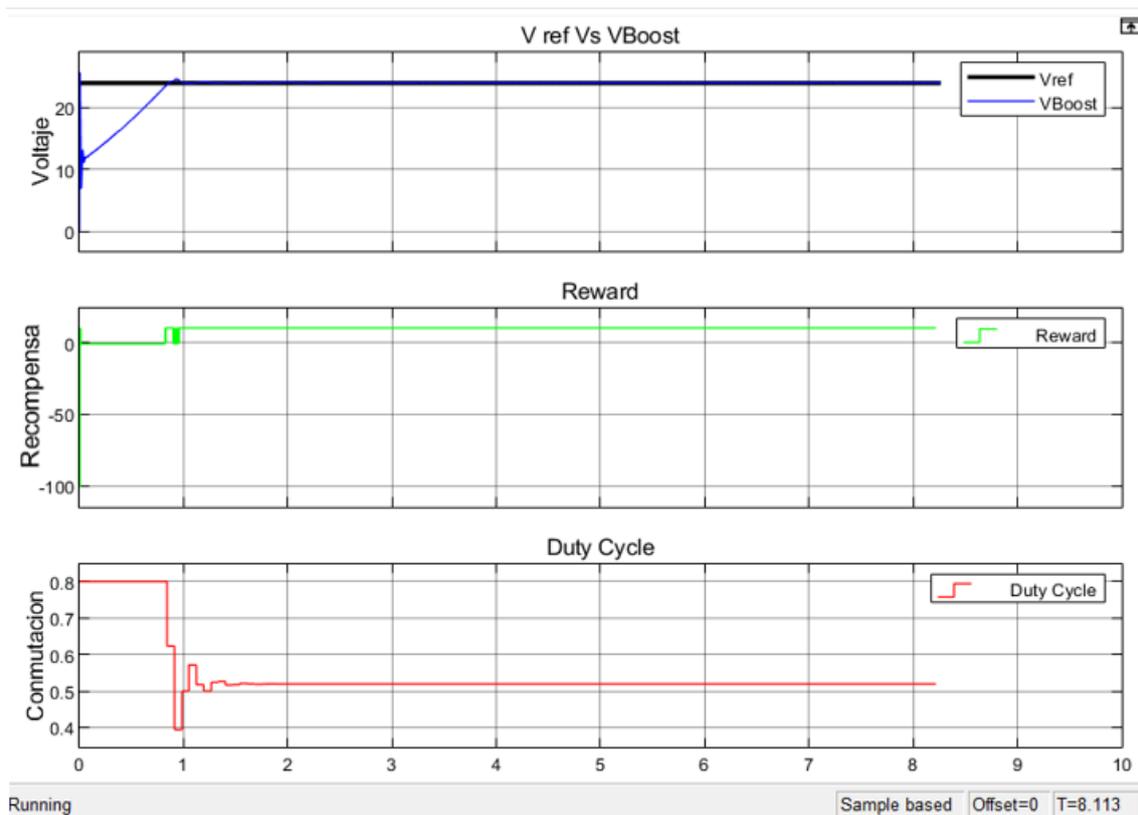


Figura 3-14 interacción con el entorno en búsqueda de la señal de referencia en el control de aprendizaje por refuerzo controlador elevador (Boost) Fuente. Elaboración Propia

la validación del agente DDPG en el convertidor elevador (Boost), se pudo observar un ajuste exitoso del ciclo de trabajo para lograr una elevación del voltaje. El agente logró ajustar el ciclo de trabajo en un valor de 0.51 para elevar el voltaje de entrada de 12V a 24 V, tal como se deseaba.

Este resultado demuestra la efectividad del enfoque de aprendizaje por refuerzo utilizando el agente DDPG en el control del convertidor elevador. A través de iteraciones y experiencias de entrenamiento, el agente fue capaz de aprender y mejorar su rendimiento, encontrando el equilibrio adecuado en el ciclo de trabajo para lograr la elevación del voltaje objetivo.

El ciclo de trabajo desempeña un papel fundamental en el control del convertidor elevador, ya que determina la proporción de tiempo durante el cual se enciende el interruptor. Al ajustar el ciclo de trabajo de manera precisa, el agente pudo elevar el voltaje de entrada de 12V a 24V, cumpliendo así con el objetivo establecido.

3.4.3 Resultados convertidores elevador (Boost) con PID

Para validar la eficiencia del control por aprendizaje reforzado se emplea un circuito siguiendo los parámetros de los elementos de un convertidor elevador (Boost) la comparación con un control PID en Simulink. En la Figura 3.9 se muestra la simulación realizada con control PID en Simulink, con el objetivo de comparar la eficiencia del control de aprendizaje reforzado con el control clásico.

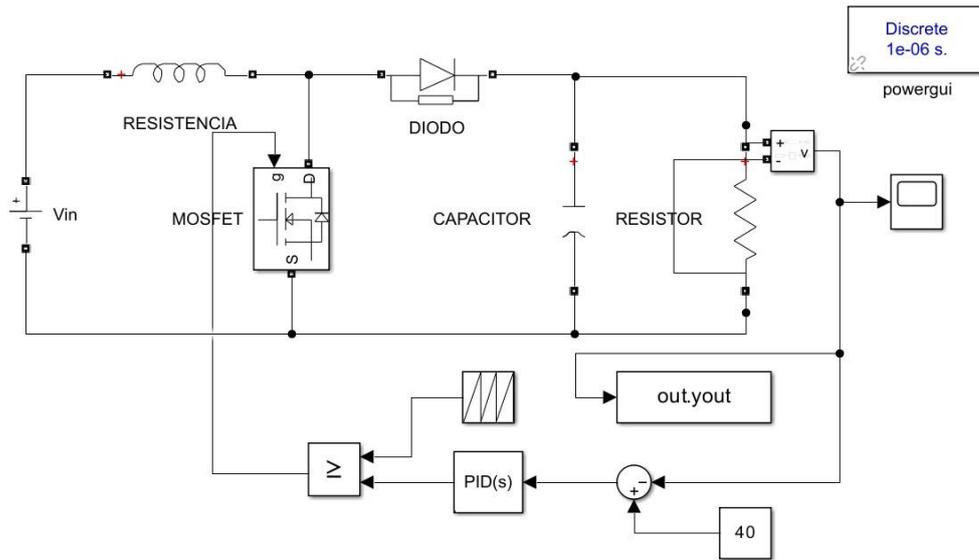


Figura 3-15 Convertidor elevador (Boost) Con control PID Fuente. Elaboración Propia

A continuación, se describen los parámetros que se utilizaron de la Tabla 6. Para el diseño de un convertidor elevador (Boost) con técnica de control PID. Resaltando que para la consecución de estos valores se generaron mediante System Identification Toolbox Cree modelos de sistemas dinámicos lineales y no lineales a partir de datos de E/S La toolbox permite estimar la dinámica de sistemas no lineales utilizando modelos Hammerstein-Wiener y ARX no lineales con técnicas de Machine Learning tales como proceso gaussiano (GP), máquina de vectores de soporte (SVM) y otras representaciones. Si prefiere, puede crear modelos de ecuaciones diferenciales ordinarias (EDO) neuronales utilizando Deep Learning para capturar la dinámica de sistemas no lineales. Con esta toolbox[54]

Convertidor elevador (Boost) PID	Valores
Voltaje entrada (Vin)	12 V
Frecuencia	10 KHz
Inductor	0.045 H
Capacitor	150 uF
Resistor	10 Ω
Ganancia KP	0.1
Ganancia KI	0.035

Tabla 6 Parámetros del diseño convertidor elevador (Boost) con PID

Se obtienen los datos mostrados en la gráfica de la figura 3.12 del convertidor elevador (Boost) con PID donde con un voltaje de entrada de 12 V se obtiene un voltaje de salida de 24 V.

En la figura 3-16 la gráfica que representa el comportamiento de este circuito al utilizar un voltaje de entrada de 12 V. Se puede observar que, al aplicar un voltaje de entrada se logra obtener un voltaje de salida de 24 V a través del circuito convertidor elevador (Boost) con técnica de control PID. el controlador utilizado en este circuito permite ajustar y estabilizar la tensión de salida en función de las variaciones en la carga y en la tensión de entrada. Esto significa que el controlador ajusta automáticamente los valores de ganancia, tiempo integral y tiempo derivativo para mantener una tensión de salida estable y precisa en todo momento.



Figura 3-16 Grafica convertidor elevador (Boost) con Control PID entrada 12 V, salida 24. Fuente Propia

4. Analisis de resultados

El análisis comparativo entre el control de aprendizaje reforzado y el control PID para los convertidores Buck y Boost reveló importantes diferencias en términos de adaptabilidad y capacidad de respuesta a las perturbaciones. Estas características son especialmente relevantes en el contexto de las energías renovables y las microredes, donde la variabilidad de la generación de energía y las fluctuaciones en la carga representan desafíos significativos.

En este contexto, se realizaron simulaciones comparativas utilizando tanto el control PID como el control de aprendizaje reforzado en los convertidores Buck y Boost, con el objetivo de evaluar su desempeño en situaciones de variación de carga y generación intermitente de energía renovable.

Las Figuras 4-1 y 4-2 muestran las respuestas de los convertidores Buck con control PID y control de aprendizaje reforzado, respectivamente. En el caso del control PID, se observa una respuesta inicial adecuada, pero con oscilaciones ligeras en la tensión de salida. En contraste, el control de aprendizaje reforzado logra una respuesta más suave y

estable, adaptándose de manera efectiva a las perturbaciones presentes en el sistema. Esta capacidad de adaptación es esencial en entornos de microrredes y energías renovables, donde la generación de energía puede ser variable y la carga puede experimentar cambios rápidos.

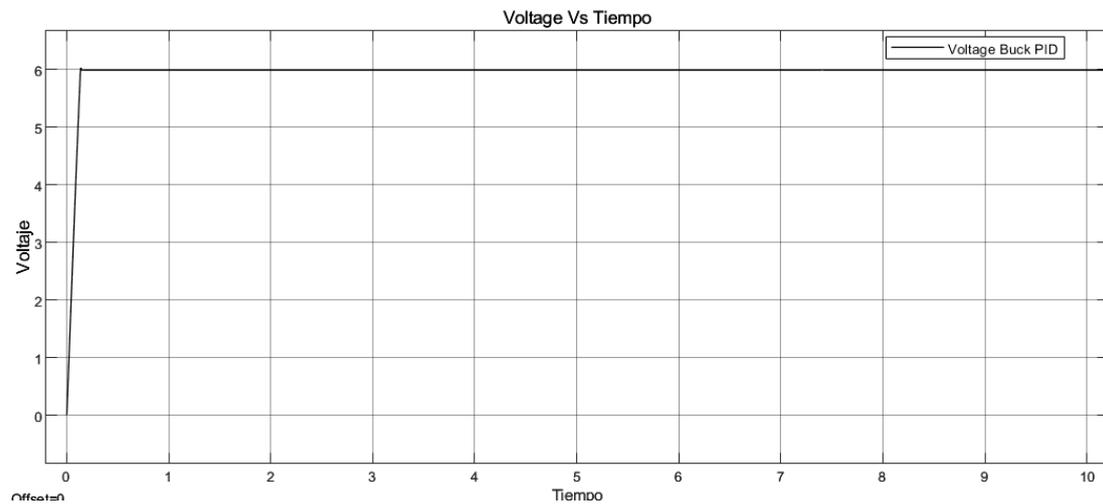


Figura 4-1 Grafica convertidor reductor (Buck) con PID entrada 24 V salida 6 V.
Fuente. Elaboración Propia



Figura 4-2 Grafica convertidor elevador (Boost) con Control PID entrada 12 V, salida 24. Fuente Propia

Las Figuras 4-3 y 4-4 presentan las respuestas de los convertidores Boost con control PID y control de aprendizaje reforzado, respectivamente. Al igual que en el caso del

convertidor Buck, el control de aprendizaje reforzado muestra una respuesta más precisa y estable, sin oscilaciones notables en la tensión de salida. Esto es particularmente relevante en aplicaciones de energías renovables, donde el objetivo es maximizar la eficiencia y la estabilidad del sistema eléctrico.

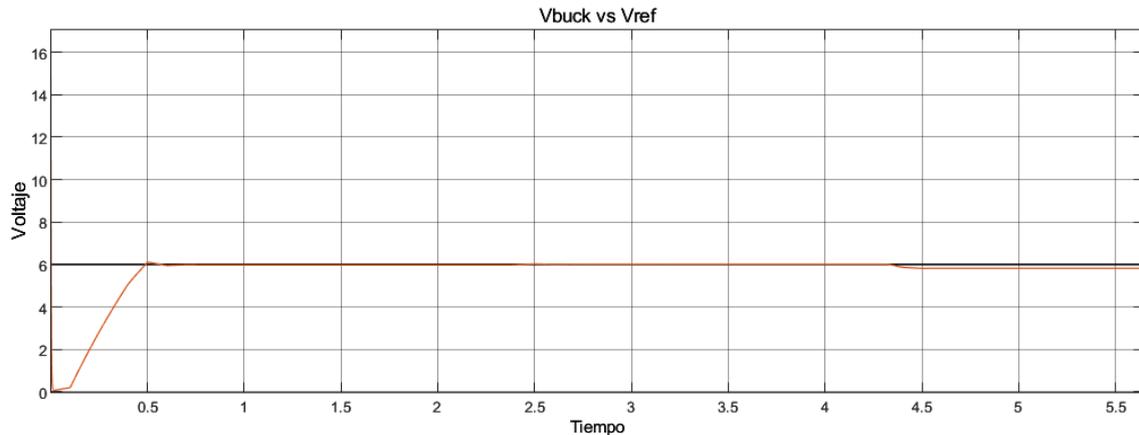


Figura 4-3 Graficas de desarrollo control aprendizaje por refuerzo controlador reductor (Buck) con variación de la carga Fuente. Elaboración Propia.

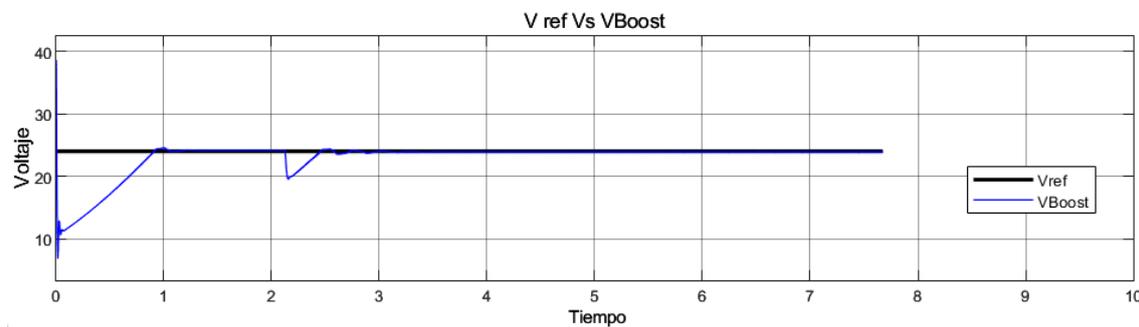


Figura 4-4 Graficas de desarrollo control aprendizaje por refuerzo controlador elevador (Boost) con variación de la carga. Fuente. Elaboración Propia.

Estas figuras comparativas ilustran claramente las diferencias entre el control PID y el control de aprendizaje reforzado en los convertidores Buck y Boost. El control de aprendizaje reforzado muestra una respuesta más precisa, estable y adaptativa, lo que lo convierte en un potencial de profundización para aplicaciones en energías renovables y situaciones de variación de carga.

5. Conclusiones

A lo largo del proyecto, se logró entrenar un agente de aprendizaje reforzado capaz de controlar de manera efectiva los convertidores Boost y Buck. El agente aprendió a ajustar el ciclo de trabajo con una gente DDGP de forma óptima para mantener el voltaje en línea con el voltaje de referencia, logrando un control preciso y estable del sistema.

La aplicación de este enfoque de control en convertidores Boost y Buck tiene un gran potencial en el campo de las energías renovables. Estos convertidores son ampliamente utilizados en sistemas de generación de energía solar y eólica, donde se requiere una conversión eficiente y controlada de voltaje. El control de aprendizaje reforzado permite optimizar la eficiencia y el rendimiento de estos sistemas, maximizando la generación de energía renovable y minimizando el uso de energía no renovable.

El enfoque de control de aprendizaje reforzado ofrece ventajas significativas, como la capacidad de adaptarse a condiciones cambiantes, mejorar la eficiencia energética y reducir los costos operativos. En el contexto de las energías renovables, esto puede conducir a un mayor aprovechamiento de fuentes de energía limpias y una menor dependencia de combustibles fósiles.

En conclusión, este proyecto demuestra la viabilidad y el potencial del control de aprendizaje reforzado aplicado a convertidores Boost y Buck. El enfoque puede contribuir significativamente a la mejora de sistemas de energías renovables, permitiendo un control preciso y eficiente de la generación y distribución de energía limpia. Estos resultados respaldan la importancia de continuar investigando y desarrollando nuevas técnicas de control inteligente para promover la transición hacia un futuro energético sostenible y respetuoso con el medio ambiente.

$$L \frac{dI_L}{dt} = -V_C + u(V_i) \quad \text{ecuación 6-1}$$

$$C \frac{dV_C}{dt} = I_L - \frac{V_C}{R} \quad \text{ecuación 6-2}$$

6.2 Modelo matemático convertidor elevador (Boost)

Un modelo matemático para un controlador elevador (Boost) es una representación simplificada del comportamiento del convertidor en forma de ecuaciones y relaciones matemáticas. Este modelo se utiliza para comprender y analizar el funcionamiento del controlador y para diseñar y optimizar su rendimiento.

El modelo matemático de un controlador elevador (Boost) se basa en una serie de suposiciones y simplificaciones, como el uso de componentes ideales (sin pérdidas) y condiciones ideales de conmutación. Estas simplificaciones permiten expresar el comportamiento del controlador en términos de ecuaciones algebraicas y diferenciales que describen las relaciones entre las variables eléctricas, como las tensiones y corrientes de entrada y salida. Figura 6-2

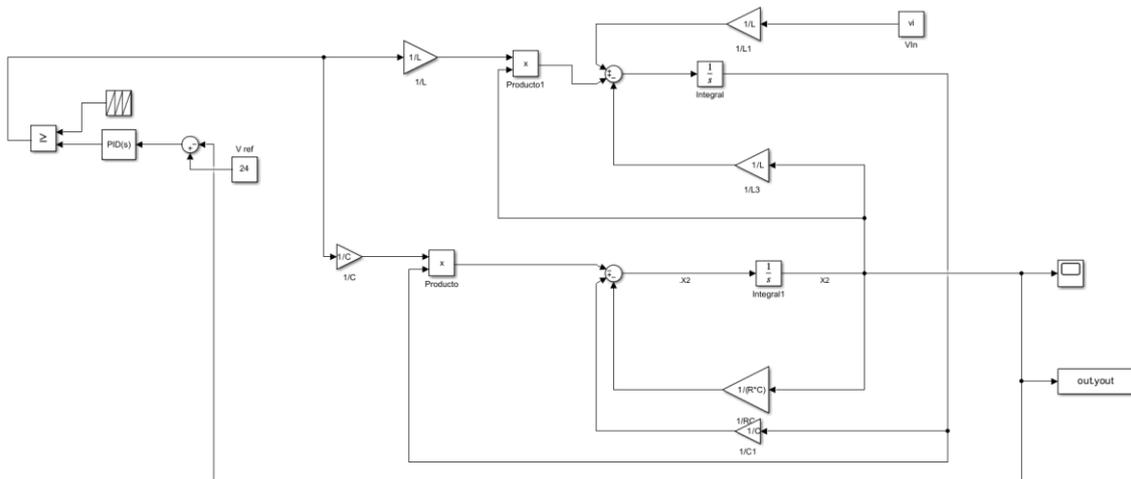


Figura 6-2 modelo matemático convertidor elevador (Boost) PID Simulink fuente propia.

Para el modelo matemático se basó con las ecuaciones 6-4 6-5

$$L \frac{dI_L}{dt} = -(1 - u)v_c + v_i \quad \text{ecuación 6-4}$$

$$C \frac{dv_c}{dt} = (1 - u)(I_L) - \frac{v_c}{R} \quad \text{ecuación 6-5}$$

6.3 Error del estado estacionario convertidor reductor (Buck)

El comportamiento deseado en estado estacionario del buck es una señal periódica de pequeña amplitud $g_i(t)$, más una señal constante \hat{z}_i . Este comportamiento se presenta si la posición del interruptor es controlada por un tren de pulsos de frecuencia constante. Al considerar el modelo promedio se tiene un único punto de equilibrio, que depende del valor de los parámetros (Agrawal (2001); Mohan et al. (2002)). Sin embargo, al considerar el modelo discontinuo los puntos de equilibrio dependen de la estrategia de conmutación utilizada

Para calcular la diferencia entre los dos controladores se corre las simulaciones y se compra figura

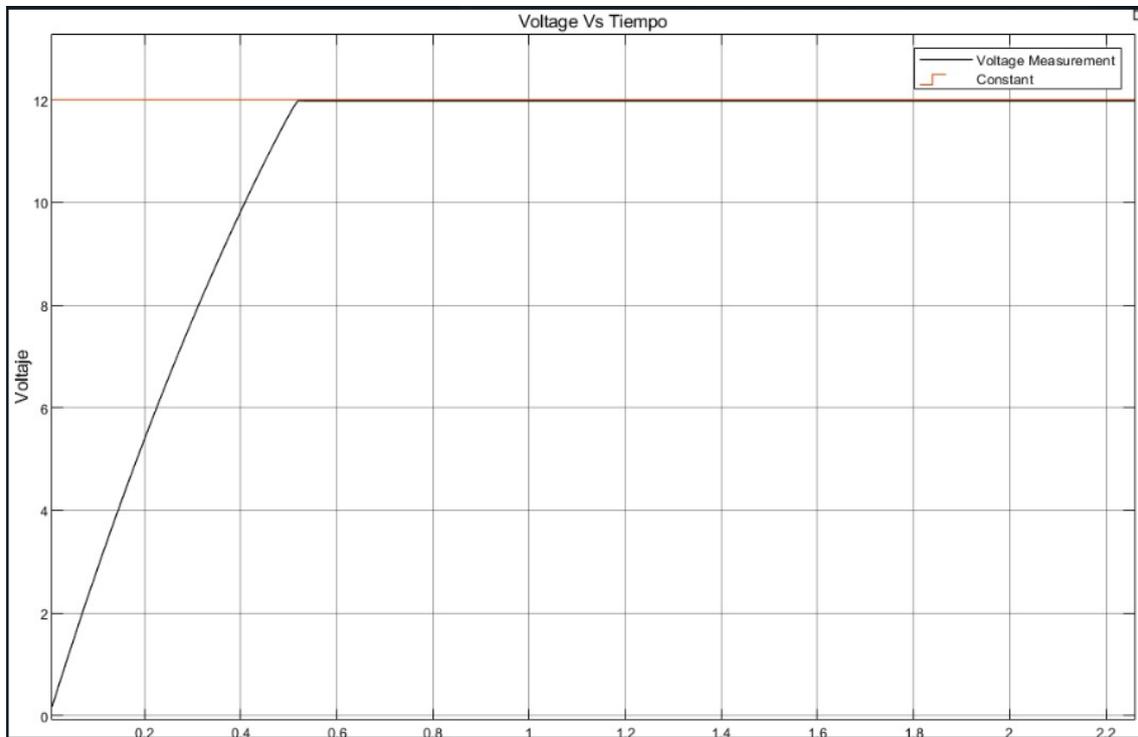


Ilustración 1 Figura 6-5 – 6-6 error en estado estacionario convertidor reductor (Buck) PID fuente propia

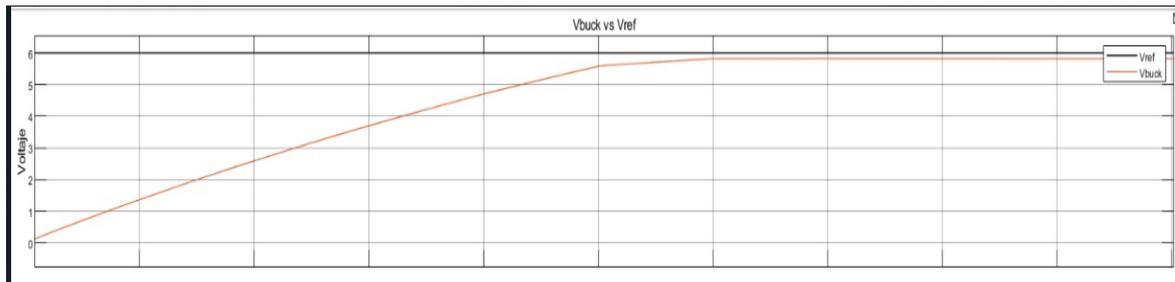


Figura 6 -5 error en estado estacionario convertidor reductor (Buck) Aprendizaje reforzado fuente propia.

El controlador Buck con PID en estado estacionario, además de proporcionar un control estable, también busca minimizar el error en estado estacionario. Sin embargo, es importante destacar que el controlador PID no puede eliminar completamente el error en estado estacionario en todas las situaciones. Esto se debe a que el término proporcional solo puede corregir el error proporcionalmente a la señal de error, el término integral corrige el error acumulado a lo largo del tiempo y el término derivativo mejora la respuesta transitoria.

En el caso del controlador Buck con PID, para lograr un rendimiento óptimo y minimizar el error en estado estacionario, es necesario ajustar adecuadamente los parámetros K_p , K_i y K_d . El ajuste de estos parámetros implica considerar las características específicas del sistema y la respuesta deseada. Si los parámetros PID no se ajustan correctamente, puede haber un error residual en estado estacionario.

En contraste, el controlador Buck en estado estacionario con aprendizaje por refuerzo también busca reducir el error en estado estacionario, pero a través de un enfoque diferente. Al utilizar algoritmos de aprendizaje por refuerzo, el controlador puede aprender y ajustar los parámetros de forma automática. Estos algoritmos se basan en la retroalimentación del rendimiento del sistema, lo que permite mejorar el control con el tiempo y reducir el error en estado estacionario de manera continua.

El controlador con aprendizaje por refuerzo tiene la capacidad de adaptarse a cambios en las condiciones de funcionamiento y perturbaciones en tiempo real. Ajusta sus acciones según las señales de entrada y salida, lo que le permite adaptarse a situaciones cambiantes y mejorar el rendimiento del sistema de control Buck. Al optimizar continuamente el rendimiento del sistema, el controlador con aprendizaje por refuerzo puede lograr una mejor eficiencia energética y reducir el error en estado estacionario en comparación con el controlador PID.

Sin embargo, es importante tener en cuenta que el controlador con aprendizaje por refuerzo puede tener una estructura más compleja que el controlador PID. Requiere una configuración inicial adecuada y un entrenamiento continuo para lograr un rendimiento óptimo y reducir el error en estado estacionario de manera efectiva.

6.4 Error del estado estacionario convertidor elevador (Boost)

Para el cálculo del error del estado estacionario se debe tener en cuenta la ecuación 6-6.

$$EEE = V \text{ referencia} - V \text{ salida en estado estacionario} \quad \text{ecuación 6-6}$$

Para calcular la diferencia entre los dos controladores se corre las simulaciones y se compra figura 6-3 6-4

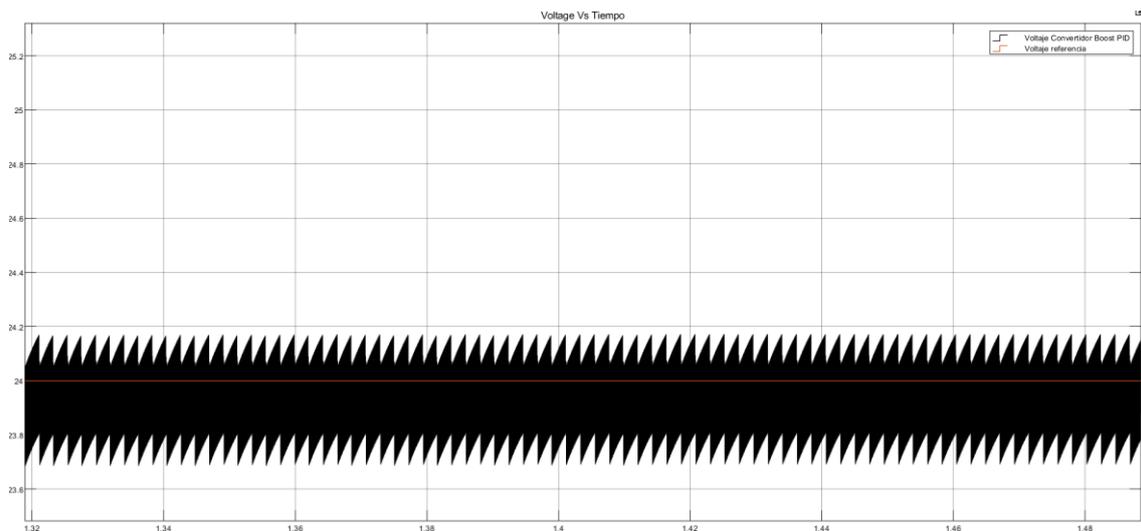


Figura 6-3 error en estado estacionario convertidor elevador (Boost) PID fuente propia

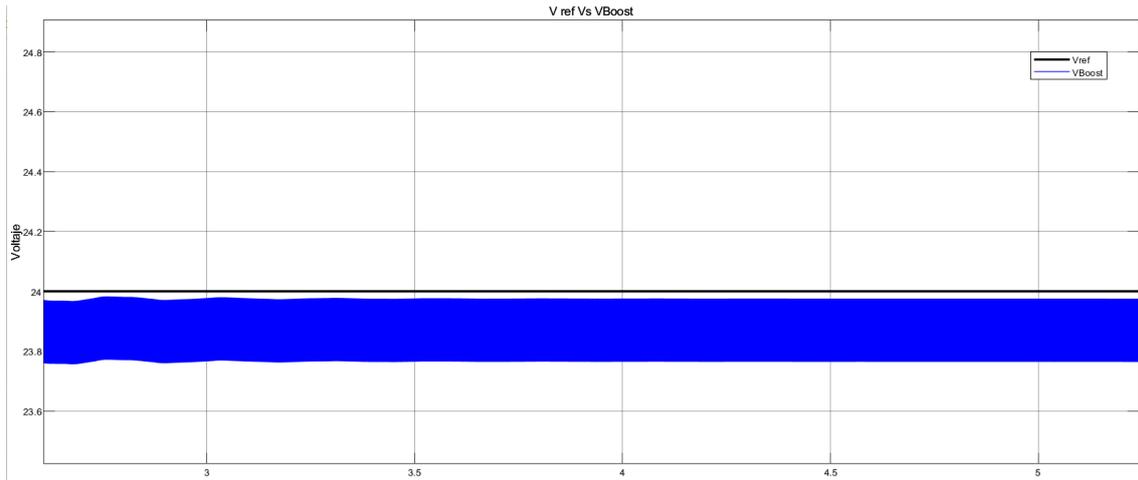


Figura 6-4 error en estado estacionario convertidor elevador (Boost) Aprendizaje reforzado fuente propia.

Como se observa en las dos graficas (Figura 6-3,6-4) el error en estado estacionario es un poco mayor en el control PID. Además, que mientras que un control PID se basa en ajustar los coeficientes del controlador para reducir el error en estado estacionario, el aprendizaje reforzado puede aprender a través de la interacción con el entorno y adaptarse de manera autónoma para minimizar dicho error. El aprendizaje reforzado tiene el potencial de lograr un mejor rendimiento en términos de error en estado estacionario, pero también puede requerir más recursos y tiempo de entrenamiento en comparación con un controlador PID tradicional.

7. Referencias Bibliográficas

- [1] Ricardo M., “Microrredes basadas en electrónica de potencia características, operación y estabilidad, Ingenius: Revista de Ciencia y Tecnología, vol. 12, 2014.
- [2] Mario Centeli, convertidores electrónicos de potencia, (2013)
- [3] Hart, D.W.: Introduction to Power Electronics (1996)
- [4] Lawton, B.W.: Damage to human hearing by airborne sound of very high frequency or ultrasonic frequency. Institute of Sound and Vibration Research (2001)
- [5] Mohan, N., Undeland, T.M., Robbins, W.P.: Power Electronics: Converters, Applications and Design (2009)
- [6] Ballester, E., Piqué, R.: Electrónica de Potencia: principios fundamentales y estructuras básicas (2012)
- [8] T. Dragičević, X. Lu, J. C. Vasquez, and J. M. Guerrero, “DC Microgrids - Part II: A Review of Power Architectures, Applications, and Standardization Issues,” IEEE Trans. Power Electron., vol. 31, no. 5, pp. 3528– 3549, 2016.
- [9] G. Prinsloo, A. Mammoli, and R. Dobson, “Customer domain supply and load coordination: A case for Smart villages and transactive control in rural off-grid microgrids,” Energy, vol. 135, pp. 430–441, 2017.
- [10] Z. Wang, S. Li, and Q. Li, “Discrete-time fast terminal sliding mode control design for dc–dc Buck converters with mismatched disturbances,” IEEE Transactions on Industrial Informatics, vol. 16, no. 2, pp. 1204– 1213, 2019.

- [11] Q. Xu, Y. Yan, C. Zhang, T. Dragicevic, and F. Blaabjerg, "An offset-free composite model predictive control strategy for DC/DC buck converter feeding constant power loads," *IEEE Transactions on Power Electronics*, vol. 35, no. 5, pp. 5331–5342, 2019.
- [12] M. Gheisarnejad, H. Farsizadeh, and M. H. Khooban, "A Novel Nonlinear Deep Reinforcement Learning Controller for DC–DC Power Buck Converters," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 8, pp 6849–6858, 2020.
- [13] Chenggang Cui, Member, Nan Yan, Baixiang Huangfu, Tianxiao Yang, Chuanlin Zhang, "Voltage Regulation of DC-DC Buck Converters Feeding CPLs via Deep Reinforcement Learning" *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2021.
- [14] Manuel Planelles y Cristian Delgado. "El 13% de la población mundial aún no tiene acceso a la electricidad". EL PAIS, Madrid 2018.
- [15] Julian Vivas, "Los 1.710 poblados que aún se alumbran con velas en el país" *EL TIEMPO*, 2019.
- [16] J. Conti, P. Holtberg, J. Diefenderfer, A. LaRose, J. T. Turnure y L. Westfall, «International Energy Outlook 2016 With Projections to 2040,» 2016.
- [17] MathWorks Reinforcement Learning eBook
<https://la.mathworks.com/campaigns/offers/reinforcementlearning-with-matlab-eb>
- [18] T. Longcheng, W. Bin, S. Rivera, V. Yaramasu, "Comprehensive DC Power Balance Management in High-Power Three-Level DC – DC Converter for Electric Vehicle Fast Charging," *IEEE Transactions on Power Electronics*, vol. 31, no. 1, pp. 89–100, Feb. 2015.
- [19] L. Young-Joo, A. Khaligh. A. Emadi, "Advanced Integrated Bidirectional AC/DC and DC/DC Converter for Plug-In Hybrid Electric Vehicles," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 8, pp. 3970–3980, Oct. 2009.
- [20] R. T. Naayagi, A. J. Forsyth, R. Shuttleworth, "High-Power Bidirectional DC–DC Converter for Aerospace Applications," *IEEE Transactions on Power Electronics*, vol. 27, no. 11, pp. 4366 – 4379, Jan. 2012.

- [21] A. Nakajima, S. Masukawa, “Study of Boost type DC-DC Converter for Single Solar Cell,” Annual Conference of the IEEE Industrial Electronics Society IECON, Washington, USA, Dec. 2018. [ook.html%0d](#).
- [22] T. Weinert, W. Oberschelp, G. Schroder, “ A current-fed DC/DC onverter for the efficient charging of HV capacitors in mobile applications”, IEEE European Conference on Power Electronics and Applications EPE, Karlsruhe, Germany, Oct. 2016.
- [23] MathWorks. (n.d.-a). Agentes de Q-Learning. Retrieved March 29, 2022, from <https://la.mathworks.com/help/reinforcement-learning/ug/q-agents.html> MathWorks. (n.d.-b). Train Multiple Agents for Path Following Control. Retrieved May 5, 2022,
- [24] Sira Ramirez, Franklin Rivas. Control de sistemas no lineales Herbertl. 2005.
- [25] E.I.Gordienkoy J.A.Minj´arez-Sosa, Adaptive control for discrete-time Markov processes with unbounded costs: average criterion, ZOR-Math. Methods of Oper. Res., vol.48, 1998, 37–55.
- [26] Garcera Sanfeliu, G.; Figueres Amoros, E.; Abellan Garcia, A. 1998. “Conversores conmutados: Circuitos de Potencia y Control”. Universidad Polit´ecnica de Valencia, Servicio de Publicaciones.
- [27] L. A. D’Alessio¹, L. G. Gonz´alez^{1,2*}, R. O. C´aceres¹ Avances en Ciencias e Ingenierías, 2014, Vol. 6, No. 2, Pags. C15-C20
- [28] A. H. Nafie and M. Shafique, Performance Enhancement of Buck Converter using Reinforcement Learning Control. Cham, Switzerland: Springer International Publishing
- [29] A. H. Nafie and M. Shafique, Performance Enhancement of Buck Converter using Reinforcement Learning Control. Cham, Switzerland: Springer International Publishing
- [30] Liu, X.; Yang, P.; Liu, Y.; Deng, J. 2008. “Modeling and simulation of Parallel Current Mode controlled boost converter”. Industrial Electronics and Applications, 2008. ICIEA 2008. 3rd IEEE Conference: 2199.
- [31] Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018). Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. arXiv preprint arXiv:1801.01290.

- [32] Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., ... & Tassa, Y. (2017). Mastering the game of Go without human knowledge. *Nature*, 550(7676), 354-359.
- [33] Gu, S., Holly, E., Lillicrap, T., & Levine, S. (2017). Deep reinforcement learning for robotic assembly planning with symbolic perception. In *Robotics: Science and Systems*.
- [34] Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine*, 34(6), 26-38.
- [35] Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.
- [36] Schulman, J., Levine, S., Abbeel, P., Jordan, M., & Moritz, P. (2015). Trust region policy optimization. In *International Conference on Machine Learning* (pp. 1889-1897).
- [37] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Petersen, S. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529-533.
- [38] Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4, 237-285.
- [39] Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of artificial intelligence research*, 4, 237-285.
- [40] Wiering, M. (2012). *A survey of reinforcement learning*. PhD thesis, University of Groningen.
- [41]. Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.
- [42] Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2019). Continuous control with deep reinforcement learning. In *Proceedings of the international conference on learning representations*.

- [43] Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of artificial intelligence research*, 4, 237-285.
- [44] Di Cao, Weihao Hu, Junbo Zhao, Guozhou Zhang, Bin Zhang, Zhou Liu, Zhe Chen, and Frede Blaabjerg, "Reinforcement Learning and Its Applications in Modern Power and Energy Systems: A Review", *Journal of modern power systems and clean energy*, vol. 8, no. 6, November 2020.
- [50] Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.
- [51] R. S. Sutton and A. G. Barto, "Reinforcement Learning: An Introduction," MIT Press, 2018.
- [52] M. H. Rashid, "Power Electronics: Circuits, Devices, and Applications," Pearson Education.
- [53] MathWorks. (n.d.-a) 29, 2023, from <https://la.mathworks.com/help/reinforcement-learning/ug/q-agents.html> MathWorks. (n.d.-b). Train Multiple System Identification ToolboxAgents for Path Following Control. Retrieved May 5, 2023
- [54] [53] MathWorks. (n.d.-a) 29, 2023, from <https://la.mathworks.com/help/reinforcement-learning/ug/q-agents.html> MathWorks. (n.d.-b). Train Multiple System Identification ToolboxAgents for Path Following Control. Retrieved May 5, 2023