Using Deep Reinforcement Learning for Dynamic Gain Adjustment of a Disturbance Observer

Kyunghwan Choi[1], Hyochan Lee[1], and Wooyong Kim[1]

[1]Affiliation not available

March 29, 2024

Using Deep Reinforcement Learning for Dynamic Gain Adjustment of a Disturbance Observer

Kyunghwan Choi[1], Hyochan Lee[1], and Wooyong Kim[2]

Abstract— **Increasing estimation accuracy and reducing noise sensitivity are challenging trade-offs in designing disturbance observers (DOBs). The DOB gain tuning process for overcoming this trade-off is not straightforward, nor does it guarantee optimal performance for the resulting DOBs. This paper presents a dynamic gain DOB that intelligently adjusts its gain based on deep reinforcement learning (DRL) to overcome this trade-off. First, a variable gain DOB is designed by modifying the conventional DOB. The variable gain DOB can exponentially estimate a constant disturbance with a varying gain. Then, DRL is used to train a dynamic gain adjuster for the variable gain DOB. A case study demonstrated that the proposed dynamic gain DOB increases its gain only when needed (i.e., when the estimation error is significant) and otherwise decreases the gain to reduce noise. Comparison with the conventional DOB of various constant gains shows that the proposed DOB achieves superior performance.**

I. INTRODUCTION

The disturbance observer (DOB) has been widely acknowledged for its ability to compensate for plant uncertainties and to reject disturbances across various control applications. However, a persistent challenge has been overcoming the trade-off between disturbance estimation accuracy and noise sensitivity [1]. High-gain DOB settings improve disturbance compensation but increase noise sensitivity, potentially degrading the overall system performance.

While there has been considerable research on disturbance compensation by DOB approaches, less attention has been given to suppressing the sensor noise effect, excluding [2]–[5]. A DOB design method based on the H_∞ synthesis technique was presented in [2]; Q filter design methods for noise reduction were presented in [3], [4]; and a combination of DOB and a Kalman filter was discussed in [5]. In contrast to DOB design methods [2], [3], which present only sufficient conditions for robust stability, the DOB design method [4] requires a necessary and sufficient condition.

There have been a few attempts to improve DOB performance using optimization or deep reinforcement learning (DRL), which differ from the conventional methods described above. In [6], receding-horizon optimization-based gain tuning of nonlinear DOB was proposed to balance disturbance estimation accuracy and noise suppression. DRL was utilized to optimize the gains of a nonlinear DOB and an active disturbance rejection controller in [7] and [8], respectively, to improve the disturbance rejection performance. A novel paradigm was presented in [9], where a DOB was designed by recurrent neural networks (RNNs) and DRL was utilized to optimize the RNNs for a target environment.

Conventional DOB design methods [2]–[4] have endeavored to reduce the sensor noise effect and derive robust stability conditions. DRL-based methods [7]–[9] have focused on improving disturbance rejection performance. However, no prior report has investigated a method to simultaneously resolve this trade-off by improving disturbance estimation accuracy and reducing noise sensitivity.

In this context, this paper presents a dynamic gain DOB that intelligently adjusts its gain based on DRL. The key concept of the proposed DOB is to increase the gain only when the estimation error is significant and to otherwise decrease gain to reduce sensor noise effects. Thus, estimation accuracy is improved and while noise sensitivity is simultaneously reduced. The dynamic gain DOB is designed by modifying the conventional DOB presented in [10], which was denoted the constant DOB due to its exponential convergence to constant disturbances. A case study using disturbances of various waveforms demonstrates that the proposed DOB achieves superior performance to that of the conventional DOB.

The remainder of this paper is organized as follows. Section II revisits the constant DOB as a motivational work. Section III introduces the proposed DOB consisting of a variable gain DOB and a dynamic gain adjuster based on DRL. Case study results are reported in Section IV. Finally, Section V offers conclusions and outlooks on future work.

II. MOTIVATION: A REVISIT TO CONSTANT DOB

A constant DOB is presented in [10] and used as the basis for designing the proposed DOB. The constant DOB is reviewed in this section.

A. Constant DOB

The system is defined as:

$$\dot{x}(t) = f(x, u, t) + Fd(t), x(0) = x_0, \tag{1}$$

where $x(t) \in \mathbb{R}^n$ is the state, $u(t) \in \mathbb{R}^m$ is the control input, $d(t) \in \mathbb{R}^p$ is the disturbance, $f(\cdot)$ is a known function, and matrix F is known and with $rank(F) = p$. The state is

*This work was supported by a Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE) (P0020535, The Competency Development Program for Industry Specialist)

[1]Kyunghwan Choi and Hyochan Lee are with the School of Mechanical Engineering, Gwangju Institute of Science and Technology, Gwangju 61005, Republic of Korea khchoi@gist.ac.kr; hyochanlee@gm.gist.ac.kr

[2]Wooyong Kim is with the Department of Biomedical & Robotics Engineering, Incheon National University, Incheon 22012, Republic of Korea wooyongkim@inu.ac.kr

assumed to be measured, and the initial state condition x_0 is known. The reduced-order system can be expressed as

$$\dot{y}(t) = g(x, u, t) + d(t). \tag{2}$$

where $y(t) = F^+x(t) \in \mathbb{R}^p$ and $g(x, u, t) = F^+f(x, u, t) \in \mathbb{R}^p$.

The constant DOB is designed for the system (2) as follows:

$$\hat{d}(t) = z(t) + Ly(t), \tag{3a}$$

$$\dot{z}(t) = -L(g(x, u, t) + \hat{d}(t)), \tag{3b}$$

with the disturbance estimate $\hat{d}(t) \in \mathbb{R}^p$, the DOB gain matrix $L = diag\{l_1, \cdots, l_p\}$, $l_i > 0$ $(i = 1, \cdots, p)$, and $z(0) = -LF^+x_0 \in \mathbb{R}^p$.

B. Error Dynamics

The error dynamics of the constant DOB are given by:

$$\dot{\epsilon}(t) = -L\epsilon(t) + \dot{d}(t), \tag{4}$$

where $\epsilon_i(t) = d_i(t) - \hat{d}_i(t)$ is the estimation error. Therefore, it follows that:

$$|\epsilon_i(t)| \leq e^{-l_i t}|\epsilon_i(0)| + (1/l_i) \sup_{0 \leq \tau \leq t} |\dot{d}_i(\tau)|, \tag{5}$$

where $(\cdot)_i$ denotes the ith component of the argument vector. The estimation characteristics are asymptotically and exponentially stable for the initial error, while the steady-state error depends on the envelope of the time derivative of the disturbance. When the disturbance is constant (i.e., $\dot{d}_i(t) = 0$), the disturbance observer exactly estimates the disturbance in the steady state.

The DOB gain l_i affects the estimation error in two ways:

- Determining the exponential decay rate of the initial error (consider the first term on the right side of inequality (5))
- Determining the suppression level of the steady-state error (consider the second term on the right side of inequality (5))

Therefore, in principle, the larger the DOB gain l_i is, the faster the estimation error $\epsilon_i(t)$ decays and the lower the steady-state error.

C. Noise Effect

In practice, state measurements contain noise as follows:

$$x_m(t) = x(t) + w(t), \tag{6}$$

where $x_m(t)$ is the measured state and $w(t)$ is the noise signal. When using $x_m(t)$ instead of $x(t)$, the error dynamics of the constant DOB are rewritten as follows:

$$\begin{aligned} \dot{\epsilon}(t) = &-L\epsilon(t) + \dot{d}(t) \\ &- L(F^+\dot{w}(t) + g(x, u, t) - g(x_m, u, t)). \end{aligned} \tag{7}$$

Clearly, the noise $w(t)$ affects the estimation error $\epsilon(t)$ even when $\epsilon(t)$ and $\dot{d}(t)$ are close to zero. The larger the DOB gain matrix L is, the greater the effect of noise on the estimation error $\epsilon(t)$. When the disturbance estimate $\hat{d}(t)$ is used for control, the noise component in $\hat{d}(t)$ is directly transferred to the control input $u(t)$, which is undesirable.

D. Trade-off Between Improving Estimation Accuracy and Reducing Noise Sensitivity

Sections II-B and II-C demonstrate the trade-off between improving the estimation accuracy and reducing the noise sensitivity of a constant DOB. Selecting a large value for the DOB gain l_i could reduce the estimation error while simultaneously increasing the noise power in the disturbance estimate. Many previous studies have confirmed this trade-off exists in various forms of DOBs [1]–[4].

Therefore, finding an optimal value for the constant gain l_i, which can reduce both the estimation error and noise component, seems technically impossible. This limitation motivated us to answer the following question: Can we design a DOB that guarantees the minimization of both the estimation error and the noise component?

III. Dynamic Gain DOB Based on DRL

Addressing the above question, this study proposes a dynamic gain DOB, which is based on a simple concept: designing a variable gain DOB and dynamically adjusting the DOB gain to minimize the estimation error or the noise component depending on the operating conditions. The variable gain DOB is designed by modifying the constant DOB (3), as presented in Section III-A. However, designing a gain adjustment policy is challenging; thus, the policy is designed based on DRL, as described in Section III-B.

A. Design of Variable Gain DOB

The variable gain DOB is designed as follows:

$$\hat{d}(t) = z(t) + L(t)y(t), \tag{8a}$$

$$\dot{z}(t) = -L(t)(g(x, u, t) + \hat{d}(t)) - \dot{L}(t)y(t), \tag{8b}$$

where the DOB gain matrix $L(t) = diag\{l_1(t), \cdots, l_p(t)\}$ is time-varying with $l_i(t) > 0 (i = 1, \cdots, p)$, $l_i(0) = l_{i,0}$, and $z(0) = -L(0)F^+x_0 \in \mathbb{R}^p$. The DOB gain $l_i(t)$ is bound by its maximum ($l_{\max}$) and minimum ($l_{\min}$) values. The variable gain DOB provides the following error dynamics:

$$\dot{\epsilon}(t) = -L(t)\epsilon(t) + \dot{d}(t), \tag{9}$$

which is the same error dynamics as the constant DOB (4) except that the gain $L(t)$ is now a variable. In this context, it is possible to adjust the gain to reduce the estimation error or noise component. Note that the measured state $x_m(t)$ is used instead of the actual state $x(t)$ when implementing the DOB; thus, the noise component will be included in the error dynamics (9) as in (7).

Consider a tracking control problem such that $x(t) \to x^*(t)$, where $x^*(t)$ is the reference signal. A tracking control law $u(t)$ can be designed leveraging the information from $\hat{d}(t)$ as an estimate of the true disturbance $d(t)$. Then, an approximate state estimate at the next time step is given by:

$$\hat{y}(t + T_s) = y(t) + T_s(g(x(t), u(t), t) + \hat{d}(t)), \tag{10}$$

where T_s is the sampling time. On the other hand, the actual state at the next time step is approximated as:

$$y(t + T_s) \approx y(t) + T_s(g(x(t), u(t), t) + d(t)), \tag{11}$$

which is measured at the next time step. Accordingly, an approximation of the disturbance estimation error is calculated by subtracting (10) from (11) as follows:

$$\epsilon(t) \approx \tilde{y}(t+T_s)/T_s, \tag{12}$$

where $\tilde{y}(t+T_s) = y(t+T_s) - \hat{y}(t+T_s)$. Equation (12) is a noncausal estimation due to the use of $y(t+T_s)$, which is the state measurement at the next time step. Nonetheless, a time-delayed calculation of (12) can be a useful criterion for adjusting the DOB gain matrix $L(t)$, as is explained in detail in the following section.

B. Dynamic Gain Adjustment Based on DRL

The error dynamics (9) yield the following scalar dynamics:

$$\dot{\epsilon}_i(t) = -l_i(t)\epsilon_i(t) + \dot{d}_i(t), i = 1, \cdots, p, \tag{13}$$

which implies that the variable gain DOB (8) operates independently across each component of the disturbance. In essence, the variable gain DOB comprises p scalar DOBs, each equipped with a variable gain mechanism. As a result, a gain adjustment policy optimized for any scalar dynamics system can be effectively extended to the vector dynamics system. This approach is depicted in Fig. 1. In this section, the training system is defined as $f(x,u,t) = -x(t) + u(t), F = 1$ with $n = 1, m = 1$, and $p = 1$ but can be any other scalar dynamics. The training controller is designed as $u(t) = K_p(x^*(t) - x_m(t)) + x_m(t) - \hat{d}(t)$ so that the state $x(t)$ tracked the reference signal $x^*(t)$ with a bandwidth that equals the control gain K_p. Any stabilizing controller is acceptable for training.

The amount of gain adjustment is determined by the time derivative of the DOB gain, $\dot{l}(t)$, as given by the gain adjustment policy. This policy is trained to be optimal by maximizing the cumulative reward utilizing the state information contained in the observation. The deep deterministic policy gradient (DDPG) method is adopted for training, whereby the model learns a state-action value function (critic) and a continuous and deterministic policy function (actor), with each function represented by a neural network. The DRL agent is detailed in the following.

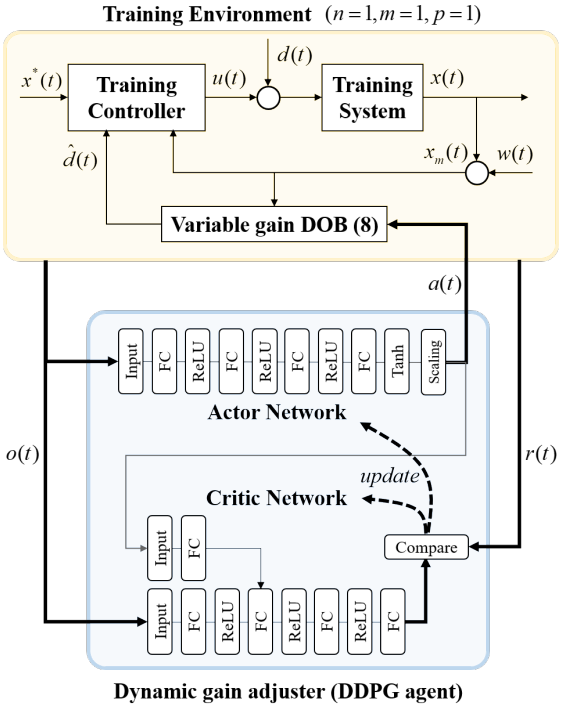
1) Action: The action $a(t)$ is selected as $\dot{l}(t)$ instead of $l(t)$. This is because $\dot{l}(t)$ is used in the variable gain DOB (8) and $l(t)$ can be easily obtained by integrating $\dot{l}(t)$ as $l(t) = \int_0^t \dot{l}(\tau)d\tau + l_0$. Obtaining $\dot{l}(t)$ from $l(t)$ is much more difficult.

The action space is defined by $A = \{a | \underline{a} \le a \le \overline{a}\}$ with $\overline{a} = (l_{max} - l_{min})/(NT_s)$ and $\underline{a} = -\overline{a}$, where N is a positive integer. The lower and upper bounds are designed as above such that $l(t)$ reaches the maximum (or minimum) from the minimum (or maximum) values in N time steps.

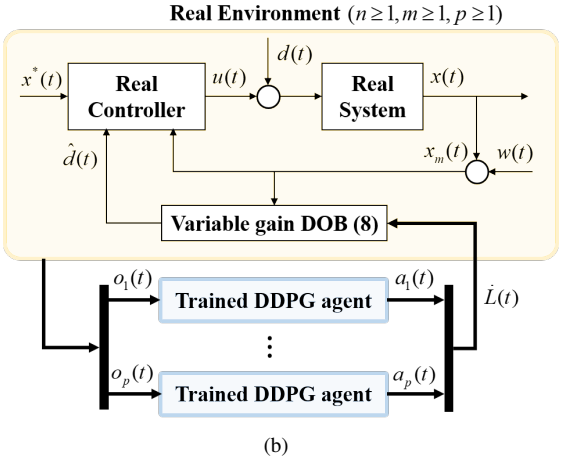
2) Reward Function: The reward function is defined by

$$r(t) = -\epsilon^2(t). \tag{14}$$

to reduce the disturbance estimation error.



(a)



(b)

Fig. 1. (a) Training and (b) implementation of dynamic gain DOB based on DRL.

3) State and Observation: The error dynamics of the variable gain DOB (9) are rewritten as:

$$\dot{\hat{d}}(t) = -l(t)\epsilon(t) \tag{15}$$

with the following DOB gain dynamics:

$$\dot{l}(t) = a(t), \tag{16}$$

which indicates that the state of the DOB comprises $\hat{d}(t)$, $l(t)$, $a(t)$, and $\epsilon(t)$. Because $\epsilon(t)$ is not observable to the agent, the term $\tilde{y}(t)/T_s$ in (12) is used instead. Accordingly, the state is defined by:

$$s(t) = \{\hat{d}(t), l(t), a(t), \tilde{y}(t)/T_s\}. \tag{17}$$

The observation is defined as follows, including the current and past states to enrich information:

$$o(t) = \{s(t), s(t - T_s), \cdots, s(t - kT_s)\}, \tag{18}$$

where k is the number of past states.

4) Neural Networks: Neural networks for the DDPG agent are depicted in Fig. 1 along with the control system as the environment, in which Input, FC, ReLU, and Tanh represent a feature input layer, a fully connected layer, a rectified linear unit function, and a hyperbolic tangent layer, respectively. Each input layer has the same number of features as its input and normalizes the input within -1 and 1. Each fully connected layer has M neurons, except for the last layers of the actor and critic networks, whose number of neurons equals 1. The hyperbolic tangent layer is used to limit its output value to within -1 and 1. The scaling layer amplifies the input value by $\overline{a}$ so that the action $a_j(t)$ has a range defined by A.

5) Training: The proposed dynamic gain DOB shown in Fig. 1 was implemented in the MATLAB/Simulink environment for training. The noise signal $w(t)$ was modeled by a band-limited white noise block with a power of 10^{-7}. The control gain K_p was set to 100. The parameters of the variable gain DOB and the dynamic gain adjuster were set as follows: $T_s = 0.001$ s, $l_{\max} = 1000$, $l_{\max} = 10$, $N = 5$, $k = 9$, and $M = 20$.

The disturbance $d(t)$ was designed as follows to describe various waveforms:

$$d(t) = \begin{cases} d_{\min} & (d_{us}(t) < d_{\min}) \\ d_{\max} & (d_{us}(t) > d_{\max}) \\ d_{us}(t) & (d_{\min} \le d_{us}(t) \le d_{\max}) \end{cases}, \tag{19}$$

where $d_{\min}(= -50)$ and $d_{\max}(= 50)$ represent the minimum and maximum values of the disturbance, respectively, and $d_{us}(t)$ is the unsaturated disturbance given by:

$$d_{us}(t) = \begin{cases} d_0 & (t < T_d) \\ d_0 + A_d \sin(2\pi f_d t + \phi_d) & (t \ge T_d) \end{cases}, \tag{20}$$

with disturbance parameters of d_0, A_d, f_d, ϕ_d and T_d. The reference signal was defined by:

$$x^*(t) = \begin{cases} x_0 & (t < T_r) \\ x_r & (t \ge T_r) \end{cases}, \tag{21}$$

where x_r and T_r are the final value and step time of the reference signal, respectively.

The dynamic gain adjuster (i.e., DDPG agent) was implemented by the RL agent block and trained for 10,000 episodes with an episode length of $T = 0.1$ s. At the start of every episode, parameters not predefined above were randomly initialized from uniform distributions. These parameters and their uniform distributions are listed in Table I. Note that the seed of the noise block was also initialized to implement a quasirandom noise signal. The hyperparameters of the DDPG agent were selected as shown in Table II.

TABLE I
RANDOMLY INITIALIZED PARAMETERS FROM UNIFORM DISTRIBUTIONS

	Parameter	Uniform distribution
Plant	x_0	$[-1, 1]$
Variable gain DOB	l_0	$[10, 600]$
Disturbance	d_0	$[-50, 50]$
	A_d	$[-100, 100]$
	f_d	$[0, 50]$
	ϕ_d	$[0, 2\pi]$
	T_d	$[0, T/2]$
Reference signal	x_r	$[-1, 1]$
	T_r	$[0, T_s]$
Band-limited noise box	Seed	$[0, 30000]$

TABLE II
HYPERPARAMETERS OF DDPG AGENT

Hyperparameter	Value
Actor learning rate	10^{-4}
Critic learning rate	10^{-3}
Discount factor	0.99
Experience buffer length	10^6
Minibatch size	64
Noise variance	7
Noise variance decay rate	10^{-4}

IV. CASE STUDY

A. Setup

A case study was conducted to validate the proposed DOB in the real environment with the pretrained agent. The randomly initialized parameters listed in Table I were used identically in the real environment, with the exceptions of l_0 and T_d, which were set to 100 and $T/2$, respectively. The real system was selected by the pendulum equation, which includes uncertain terms, as follows:

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} x_2(t) \\ -\sin(x_1(t)) - x_2(t) + u(t) \end{bmatrix} + Fd(t) \tag{22}$$

with $n = 2$, $m = 1$, $p = 1$, and $F = \begin{bmatrix} 1 & 1 \end{bmatrix}^\top$. The real controller was designed based on the backstepping approach as follows:

$$u(t) = \sin(x_1(t)) + x_2(t) + e(t) - k_2\phi(t) - \hat{d}(t), \tag{23}$$

where

$$e(t) = x^*(t) - x_1(t), \tag{24}$$

$$\phi(t) = x_2(t) - (k_1 e(t) - \hat{d}(t)). \tag{25}$$

with control gains $k_1 > 0$ and $k_2 > 0$. The new variables $\phi(t)$ denotes the error of $x_2(t)$ relative to its desirable value. The Lyapunov function is defined as follows to demonstrate the system's stability:

$$V(t) := \frac{1}{2}e^2(t) + \frac{1}{2}\phi^2(t) + \frac{1}{2}\epsilon^2(t). \tag{26}$$

Taking the time-derivative of $V(t)$ yields

$$\dot{V}(t) = e(t)\dot{e}(t) + \phi(t)\dot{\phi}(t) + \epsilon(t)\dot{\epsilon}(t) \tag{27}$$

$$= -\left(k_1 - \frac{1}{2}\right)\left(e(t) - \frac{k_1^2}{2k_1 - 1}\phi(t)\right)^2 - k_3\phi^2(t) - \left(l(t) - \frac{1}{2}\right)\left(\epsilon(t) - \frac{1 + l(t) + k_1}{2l(t) - 1}\phi(t)\right)^2 + \alpha(t)^\top\beta(t), \tag{28}$$

where

$$k_3 = k_2 - k_1 - \frac{k_1^4}{4k_1 - 2} - \frac{(1 + k_1 + l(t))^2}{4l(t) - 2},$$
$$\alpha(t) = [\dot{x}^*(t), -k_1\dot{x}^*(t), \dot{d}(t)]^\top,$$
$$\beta(t) = [e(t), \phi(t), \epsilon(t)]^\top.$$

If k_1 and $l(t)$ are greater than $1/2$, and k_1 and k_2 are selected for k_3 to be positive, the following holds:

$$\dot{V}(t) \leq \alpha(t)^\top\beta(t), \tag{29}$$

which indicates L_2-stability for the input–output mapping of $\alpha(t) \mapsto \beta(t)$. Therefore, the system is stable if the input $\alpha(t)$ is bounded. The control gains k_1 and k_2 were selected as 1 and 300, respectively, to satisfy the above condition in the range of $l(t)$.

The proposed DOB was simulated 50 times, as was the conventional constant DOB (3) for comparison. Six different constant gains, $l = 100, 200, 400, 600, 800$ and 1000, were used for the constant DOB. Two performance indices were defined as follows:

$$J_1 = \int_0^T \epsilon^2(\tau)d\tau, \quad J_2 = \int_0^T ((\hat{d}(\tau) - \hat{d}^*(\tau))^2 d\tau,$$

where $\hat{d}^*(t)$ is the disturbance estimate obtained with the same DOB, while assuming that the state measurement does not contain noise (i.e., $x_m(t) = x(t)$); thus, $\hat{d}^*(t)$ does not contain any noise components. J_1 and J_2 correspond to the disturbance estimation error and the noise component, respectively.

B. Results and Discussion

Figure 2(a) shows the distributions of the performance indices of the proposed DOB and constant DOB for a total of 50 simulations. The performance indices of the proposed DOB were distributed at small values. In contrast, depending on the DOB gain l, the constant DOB exhibits a clear trade-off between the two performance indices: the higher the gain is, the lower the estimation error variance, but the greater the noise component. In the constant DOB, no gain value guarantees a distribution of lower performance indices, as achieved by the proposed DOB. This statement is clearly validated by examining the average point of the performance indices distribution, as shown in Fig. 2(b). The average point of the proposed DOB is lower in both performance indices than that of the constant DOB with $l = 400, 600, 800$ and 1000. Although one performance index (J_2) of the constant DOB with $l = 100$ and 200 is

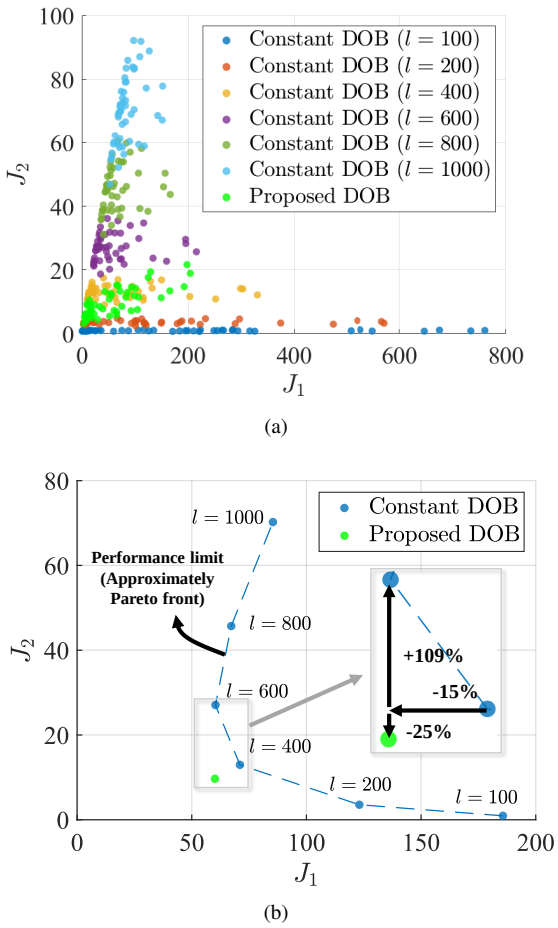
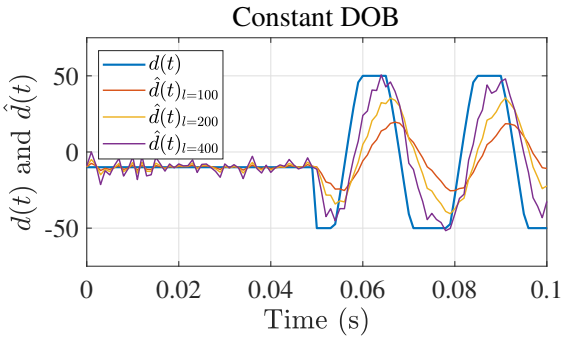


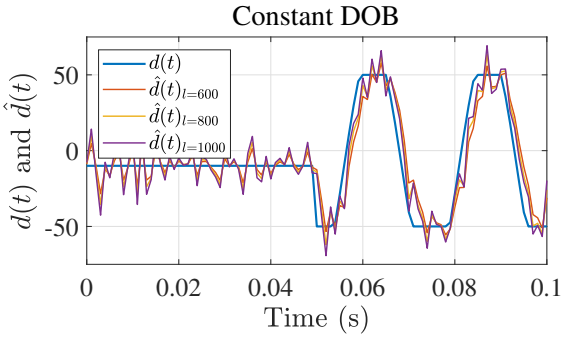
Fig. 2. Distribution of the performance indices of the proposed DOB and constant DOB (a) for a total of 50 simulations, and (b) their average points.

lower than that of the proposed DOB on average, the other average performance index (J_1) is much greater. The blue dashed line can be regarded as the performance limit of the constant DOB, which is approximately Pareto optimal performance. Notably, the average point of the proposed DOB is located below the performance limit, moving toward the utopia point off the Pareto front. The average J_1 values for both the proposed DOB and the constant DOB with $l = 600$ were nearly identical, showing a reduction of 15% compared to the constant DOB with $l = 400$. Conversely, the average J_2 for the constant DOB with $l = 600$ increased by 109% over the constant DOB with $l = 400$, whereas the proposed DOB saw a 25% reduction in J_2 compared to the constant DOB with $l = 400$. This indicates that the proposed DOB effectively lowers both performance indices by intelligently adjusting the DOB gain, in contrast to the constant DOB, which demonstrates a significant trade-off between the two indices when setting the DOB gain. Additionally, it is important to note that the average J_1 value does not continue to decrease with higher gains, attributed to the noise component's amplification within the disturbance estimate.

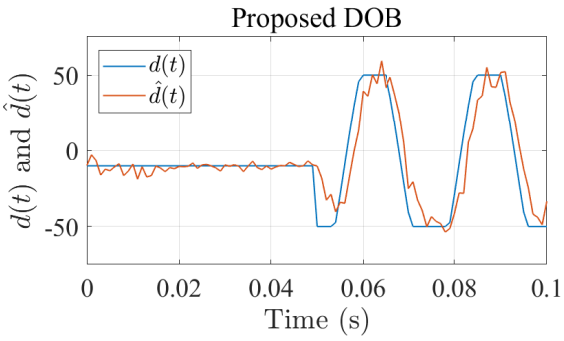
Among the total 50 simulations, two simulations were



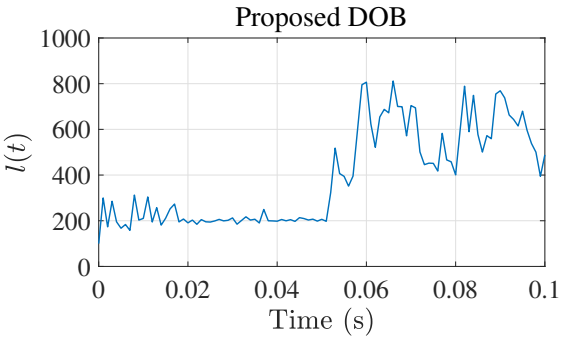
(a)



(b)

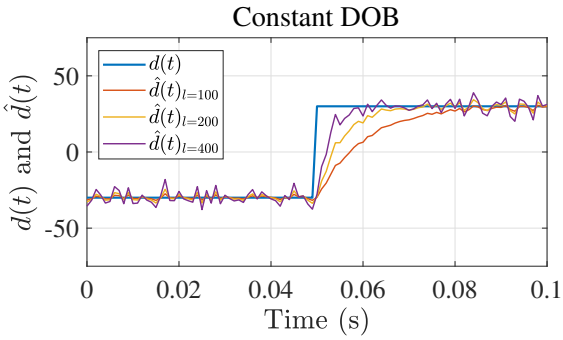


(c)

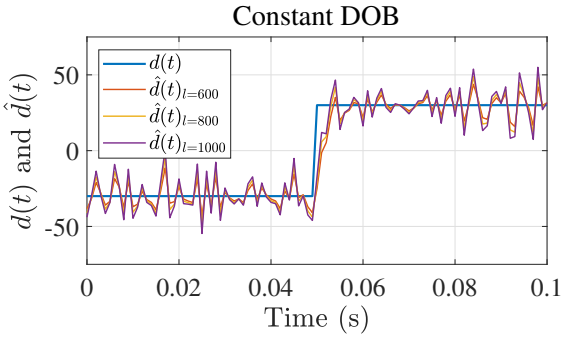


(d)

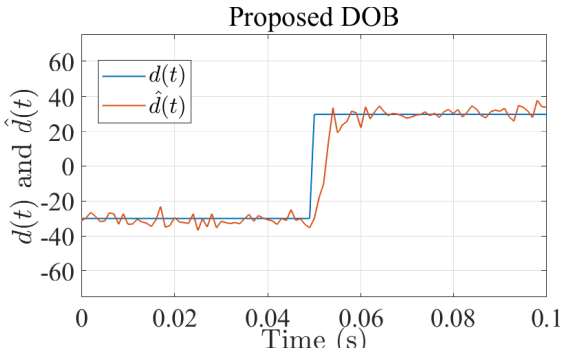
Fig. 3. Comparison of the proposed DOB and constant DOB in one simulation. The disturbance estimates $\hat{d}(t)$ of the constant DOB obtained with $l = 100, 200$, and 400 and with $l = 600, 800$, and 1000 are presented in (a) and (b), respectively, along with the true disturbance $d(t)$. The disturbance estimate $\hat{d}(t)$ of the proposed DOB is presented in (c) along with the true disturbance $d(t)$. The corresponding DOB gain $l(t)$ of the proposed DOB is shown in (d).



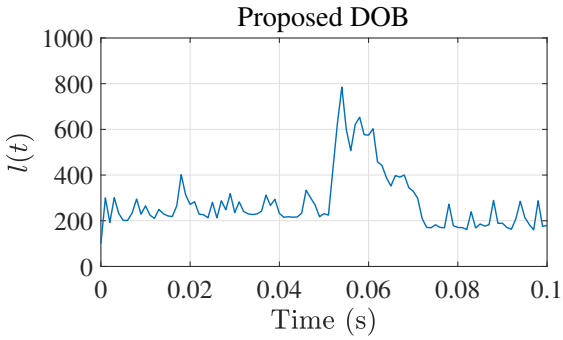
(a)



(b)



(c)



(d)

Fig. 4. Comparison of the proposed DOB and constant DOB in another simulation. The disturbance estimates $\hat{d}(t)$ of the constant DOB obtained with $l = 100, 200$, and 400 and with $l = 600, 800$, and 1000 are presented in (a) and (b), respectively, along with the true disturbance $d(t)$. The disturbance estimate $\hat{d}(t)$ of the proposed DOB is presented in (c) along with the true disturbance $d(t)$. The corresponding DOB gain $l(t)$ of the proposed DOB is shown in (d).

selected to qualitatively compare the proposed DOB and constant DOB. The two simulation results are shown in Figs. 3 and 4, respectively. For one simulation, the disturbance estimates $\hat{d}(t)$ of the constant DOB obtained with $l = 100, 200$ and 400 and with $l = 600, 800$ and 1000 are presented in Figs. 3(a) and 3(b), respectively, along with the true disturbance $d(t)$. The data show that the higher the gain l is, the faster the estimation speed, but the more significant the noise component. The disturbance estimate obtained with $l = 400$ provided tolerable estimation performance and a tolerable noise component, while that obtained with $l = 600$ provided better estimation performance but a worse noise component. In contrast, as shown in Fig. 3(c), the disturbance estimate of the proposed DOB had good estimation performance, similar to the constant DOB with $l = 600$ and a lower noise component than the constant DOB with $l = 400$. This desirable estimation behavior, which cannot be attained using a constant DOB, could be obtained due to the dynamically varying gain $l(t)$. As shown in Fig. 3(d), the gain was approximately 200 for the constant disturbance from 0 s to 0.05 s and then increased and adjusted appropriately for the time-varying disturbance from 0.05 s. A similar analysis can be performed for another simulation result shown in Fig. 4. Interestingly, as shown in Fig. 4(d), the gain rapidly increased to 800 when a step change in the disturbance occurred and remained low otherwise.

The qualitative comparison has demonstrated that the proposed dynamic gain DOB increases its gain only when necessary (i.e., when the estimation error is significant) and decreases it otherwise to reduce noise effects. This desirable behavior explains how the proposed DOB can operate close to the optimal point in terms of performance indices by overcoming the performance limit of the constant DOB (see Fig. 2(b)).

V. Conclusion and Future Work

This paper presents a dynamic gain DOB based on DRL to solve the trade-off of conventional DOBs simultaneously increasing estimation accuracy and reducing noise sensitivity. The proposed DOB comprised a variable gain DOB, which was formulated by modifying the conventional constant DOB, and a dynamic gain adjuster, which determined the gain of the variable gain DOB. The dynamic gain adjuster was based on a DDPG agent, and the training framework of this agent was also presented. A case study using disturbances of various waveforms demonstrated that the proposed DOB achieves superior performance to that of the conventional DOB in terms of estimation accuracy and noise sensitivity. This ideal performance was achieved because the proposed dynamic gain DOB increases its gain only when necessary and otherwise decreases the gain to reduce noise effects.

The proposed DOB is anticipated to be used as a universal DOB with plug-and-play integration capabilities. For instance, it will be possible to attach the proposed DOB to any control system and easily enjoy its superior functionality. The following issues need to be addressed in a future study to realize this capability:

- Generalization of the DOB formulation to be applicable to output-feedback systems;
- Adaptation to various levels of noise and various ranges of sampling time.

References

[1] D. Tena, I. Peñarrocha-Alós, and R. Sanchis, "Performance, robustness and noise amplification trade-offs in disturbance observer control design," *European Journal of Control*, vol. 65, p. 100630, 2022.

[2] J. Su, L. Wang, and J. Yun, "A design of disturbance observer in standard h_∞ control framework," *International Journal of Robust and Nonlinear Control*, vol. 25, no. 16, pp. 2894–2910, 2015.

[3] W. Xie, "High frequency measurement noise rejection based on disturbance observer," *Journal of the Franklin Institute*, vol. 347, no. 10, pp. 1825–1836, 2010.

[4] N. H. Jo, C. Jeon, and H. Shim, "Noise reduction disturbance observer for disturbance attenuation and noise suppression," *IEEE Trans. Ind. Electron.*, vol. 64, no. 2, pp. 1381–1391, 2016.

[5] C. Mitsantisuk, K. Ohishi, and S. Katsura, "Estimation of action/reaction forces for the bilateral control using kalman filter," *IEEE Trans. Ind. Electron.*, vol. 59, no. 11, pp. 4383–4393, 2011.

[6] Y. Yan, Z. Sun, J. Yang, and S. Li, "A guide for gain tuning of disturbance observer: Balancing disturbance estimation and noise suppression," in *2018 IEEE Conference on Control Technology and Applications (CCTA)*. IEEE, 2018, pp. 1558–1563.

[7] D. Lee, H. Ahn, J. Lee, and H. Bang, "Reinforcement learning-based nonlinear disturbance observer for uav with parametric uncertainty and unmodeled dynamics," in *AIAA SCITECH 2023 Forum*, 2023, p. 2357.

[8] Y. Wang, S. Fang, and J. Hu, "Active disturbance rejection control based on deep reinforcement learning of pmsm for more electric aircraft," *IEEE Trans. Power Electron.*, vol. 38, no. 1, pp. 406–416, 2022.

[9] T. Wang, W. Lu, Z. Yan, and D. Liu, "Dob-net: Actively rejecting unknown excessive time-varying disturbances," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 1881–1887.

[10] K.-S. Kim, K.-H. Rew, and S. Kim, "Disturbance observer for estimating higher order disturbances in time series expansion," *IEEE Trans. Autom.*, vol. 55, no. 8, pp. 1905–1911, 2010.