외란관측기의 심층강화학습 기반 지능형 자동 동조 기법
Using Deep Reinforcement Learning for Intelligent Gain Adjustment of a Disturbance Observer

○이 효 찬[1], 최 경 환[1*]
[1)] 광주과학기술원 기계공학부 (TEL:062-715-2413; E-mail: hyochanlee@gm.gist.ac.kr, khchoi@gist.ac.kr)

Abstract This paper proposes a deep reinforcement learning (DRL) based dynamic gain disturbance observer (DOB) that intelligently adjusts its gain and thereby resolves the trade-off problem between increasing estimation accuracy and reducing noise sensitivity in designing DOBs. First, a variable gain DOB is designed by modifying the conventional DOB. 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 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.

Keywords disturbance observer, deep reinforcement learning, estimation accuracy, gain adjustment, noise sensitivity.

1. 서론

외란은 제어 시스템의 성능과 안정성에 치명적인 영향을 미치며 제어 시스템 설계에 이에 대응할 수 있는 기술 적용이 필요하다. 그 대표적인 기술인 외란관측기(disturbance observer, DOB)는 외란을 효과적으로 추정 및 보상할 수 있는 기술로 활발하게 연구 및 적용되어오고 있다. 하지만 추정 정확성과 노이즈 민감성 사이의 trade-off 는 여전히 해결이 필요한 문제이다[1].

이 문제를 해결하기 위한 몇 가지 사전연구들이 있었다. [2]에서는 Q 필터 설계 기반 강인성 개선을 통한 노이즈 저감 기술이 제안되었다. [3]에서는 심층강화학습 기반의 DOB 이득 자동 동조기술이 제안되었다. 이러한 연구들은 외란 추정 및 보상을 통한 제어성능 향상에 초점을 맞추었으나, 추정 정확성과 노이즈 민감성 사이의 trade-off 에 대해서는 엄밀한 고려가 없었다.

본 논문은 DOB 의 추정 정확성 향상과 노이즈 민감성 저감을 동시에 해결하는 심층강화학습 기반 지능형 자동 동조 기법을 제안한다. 이를 위해 가변이득 DOB 를 설계하고 이 이득이 두 목적(즉, 추정 정확성 향상과 노이즈 민감성 저감)을 모두 만족시키도록 심층강화학습 기반으로 최적화한다. 스칼라 시스템에서 학습된 심층강화학습 에이전트는 벡터 시스템으로 확장되어 검증된다.

2. 심층강화학습기반 지능형 자동 동조 기법

2.1 가변이득 DOB 설계와 노이즈 영향 분석

고려하는 시스템의 정의는 다음과 같다.

$$\frac{dx(t)}{dt} = f(x,u,t) + Fd(t), x(0) = x_0, \quad (1)$$

여기서 $x(t) \in R^n, u(t) \in R^m, d(t) \in R^p$ 는 각각 시스템의 상태, 제어 입력, 외란이며, $f(\cdot)$는 알 수 있는 시스템 함수, F는 $rank(F) = p$인 알 수 있는 행렬이다.

이를 기반으로 차수가 축소된 시스템은 다음과 같다.

$$\frac{dy(t)}{dt} = g(x,u,t) + d(t), \quad (2)$$

여기서 $y(t) = F^+x(t)$, $g(x,u,t) = F^+f(x,u,t)$, $y(t) \in R^p, g(x,u,t) \in R^p$ 이다. 가변이득 DOB 는 [4]에서 제안된 constant DOB 를 변형하여 다음과 같이 설계된다.

$$\hat{d}(t) = z(t) + L(t)y(t), \quad (3a)$$

$$\frac{dz(t)}{dt} = -L(t)\left(g(x,u,t) + \hat{d}(t)\right) - \frac{dL(t)}{dt}y(t), \quad (3b)$$

여기서 $\hat{d}(t) \in R^p$ 은 외란 추정치, $L(t) = diag\{l_1(t), \cdots, l_p(t)\}, l_i(t) > 0, (i = 1, \cdots, p)$는 DOB 의 이득 행렬, 초기값은 $z(0) = -LF^+x_0 \in R^p$이다.

DOB (3)의 추정 오차 $\epsilon(t) = d(t) - \hat{d}(t)$의 거동은 다음과 같다.

$$\frac{d\epsilon(t)}{dt} = -L\epsilon(t) + \frac{d(t)}{dt} + \epsilon_n(t), \quad (4)$$

여기서 $\epsilon_n(t)$ 는 노이즈의 영향 성분으로 $x_m(t) = x(t) + \omega(t)$ 와 같이 상태에 노이즈가 포함된 경우 $\epsilon_n(t) = -L(F^+\dot{\omega}(t) + g(x,u,y) - g(x_m,u,t))$ 와 같이 정의되며 $\omega(t) = 0$인 경우 $\epsilon_n(t) = 0$ 이다.

오차 거동 (4)에서 나타난 바와 같이 이득이 증가할 경우 수렴성 높아지지만 그와 동시에 노이즈 민감성이 증가하는 trade-off 문제가 발생한다.

2.2 심층강화학습 기반 학습

오차 거동 (4)는 $\dot{\epsilon}_i(t) = -l_i(t)\epsilon_i(t) + \dot{d}_i(t) + \epsilon_{n,i}(t), (i = 1, \cdots, p)$와 같은 스칼라 시스템으로 분리

할 수 있으며, 이를 기반으로 스칼라 시스템에서 학습한 뒤 벡터 시스템으로 확장한다.

학습 기법으로는 deep deterministic policy gradient (DDPG)를 선정하였으며 MATLAB/Simulink 환경에서 진행하였다. $f(x,u,t)=x(t)+u(t)$ 의 시스템에 대하여 band-limited 노이즈와 ± 50의 최대, 최소 값을 갖는 임의의 $d(t)=d_0+A_d\sin(2\pi f_d t+\phi_d)\,(t\geq T_d)$ 외란을 사용했다. 심층강화학습 설정은 다음과 같다: $T_s=1ms$, state: $s(t)=\{\hat{d}(t),l(t),a(t),\tilde{y}(t)/T_s\}$, observation: $o(t)=\{s(t),s(t-T_s),\cdots s(t-kT_s)\}$, action: $a(t)=\dot{l}(t)$, reward: $r(t)=-\epsilon(t)^2$.

3. 모의 실험

3.1 환경 설정

검증에 사용된 시스템은 다음과 같다.

$$\begin{bmatrix}\frac{dx_1(t)}{dt}\\ \frac{dx_2(t)}{dt}\end{bmatrix}=\begin{bmatrix}x_2(t)\\ -\sin\big(x_1(t)\big)-x_2(t)+u(t)\end{bmatrix}+Fd(t), \quad (5)$$

시스템 (5)는 $n=2, m=1, p=1, F=[1\quad 1]^T$ 인 비선형 2 차 시스템이며, 스칼라 시스템에서 학습된 에이전트를 (5)의 시스템으로 확장하여 검증하였다. l=100, 200, 400, 600, 800, 1000 의 상수 이득에 대해 다음과 같은 성능 검증 지표를 통해 비교되었다.

$$J_1=\int_0^T \epsilon(\tau)d\tau, \qquad J_2=\int_0^T \left(\hat{d}(\tau)-\hat{d}^*(\tau)\right)^2 d\tau, \quad (6)$$

여기서 $\hat{d}^*(t)$는 동일한 DOB 에서 노이즈가 포함되지 않은 추정 값으로 J_1, J_2는 각각 추정 정확성과 노이즈 민감성에 대한 정도를 수치적으로 나타낸다.

3.2 검증 결과

제안한 DOB 는 추정 정확성 노이즈 영향 저감에 대해 기존의 DOB 보다 모두 향상됨을 확인했다. 그림 1 은 50 번의 검증에 대한 J_1, J_2의 평균 수치를 나타내며, 그림 2 는 이중 한 시나리오에 대한 DOB 의 거동을 나타낸다.

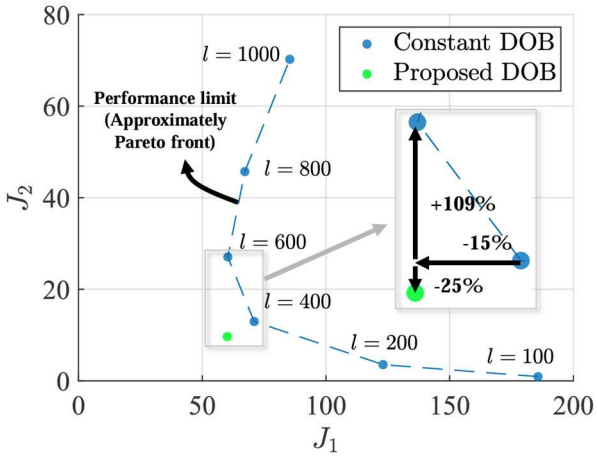

그림 1. Constant DOB 와 제안한 DOB 의 50 번의 검증에 대한 평균 성능

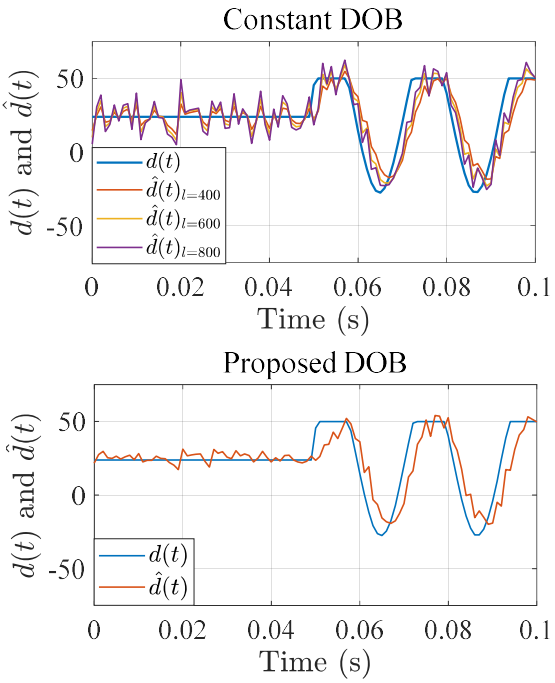

그림 2. 50 번 검증 중 한 시나리오에 대한 Constant DOB 와 제안한 DOB 의 추정 거동

4. 결론

본 논문에서는 DOB 의 심층강화학습 기반 지능형 자동 동조 기술을 제안하였고, 기존의 DOB 기술이 가지는 추정 정확성과 노이즈 민감성 간의 trade-off 문제를 해결하였다. 제안한 기술은 이득 값 변동을 위해 가변이득 DOB 를 설계하였고, DDPG 기반의 심층강화학습을 통한 최적의 이득 거동을 설계하였다. 스칼라 시스템에 대해 학습된 DRL 에이전트는 벡터 시스템으로 확장하여 검증되었다. 검증 결과 제안한 DOB 는 추정 정확성과 노이즈 영향 저감에 대하여 기존 DOB 의 튜닝으로 도달할 수 없는 성능을 가지는 것을 확인하였다. 이러한 결과는 제안한 지능형 자동 동조 기술이 DOB 의 이득 값이 필요한 경우 증가하고, 그 외의 경우 감소하여 노이즈의 영향을 낮춤으로써 달성하였다.

참고문헌

[1] D. Tena, I. P.-Alos, and R. Sanchis, “Performance, robustness and noise amplification trade-offs in Disturbance Observer Control design,” European Journal of Control, vol. 65, pp.100630, 2022.

[2] N. H. Jo, C. Jeon, and H. Shim, “Noise reduction disturbance observer for disturbance attenuation and noise suppression,” IEEE Trans. Ind. Electron., vol. 64, pp. 1381, 2016.

[3] D. Lee, H. Ahn, J. Lee and H. Bang, “Reinforcement Learning-based Nonlinear Disturbance Observer for UAV with Parametric Uncertainty and Unmodeled Dynamics”, AIAA SCITECH 2023 Forum, pp. 2457, 2023.

[4] 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, pp. 1905, 2010.