

Different from the state reconstruction given in (8) which is specific to the discrete-time systems, an alternative way to obtain the state information is to use a state observer to estimate the real state. Concretely, the state of system (1) can be estimated by using the following Luenberger observer:

$$\hat{x}(k+1) = (A - LC)\hat{x}(k) + Bu(k) + Ly(k). \quad (10)$$

It is known that the Luenberger observer (10) has the property that  $\lim_{k \rightarrow \infty} \hat{x}(k) = x(k)$  when all the eigenvalues of  $(A - LC)$  are set within the unit circle under the observability condition. As given in Rizvi & Lin (2019), by setting  $\hat{x}(k) = \mathbf{0}$ , the estimated state  $\hat{x}(k)$  can be parametrized by

$$\hat{x}(k) = W_y \omega(k) + W_u \sigma(k) = \begin{bmatrix} W_y & W_u \end{bmatrix} \begin{bmatrix} \omega(k) \\ \sigma(k) \end{bmatrix} := \bar{W} \varpi(k) \quad (11)$$

where the parameterization matrices  $W_y \in \mathbb{R}^{n \times np}$  and  $W_u \in \mathbb{R}^{n \times nm}$  are determined by the coefficients of the numerators in the transfer function matrix of the Luenberger observer system (10) with  $y(k)$  and  $u(k)$  being the inputs to the observer system, and the dynamics of  $\omega(k)$  and  $\sigma(k)$  are given as

$$\begin{aligned} \omega(k+1) &= \bar{\mathcal{A}}_y \omega(k) + \bar{\mathcal{B}}_y y(k), \quad \omega(0) = \mathbf{0} \\ \sigma(k+1) &= \bar{\mathcal{A}}_u \sigma(k) + \bar{\mathcal{B}}_u u(k), \quad \sigma(0) = \mathbf{0} \end{aligned}$$

where  $\bar{\mathcal{A}}_y \in \mathbb{R}^{np \times np}$  and  $\bar{\mathcal{A}}_u \in \mathbb{R}^{nm \times nm}$  are user-defined matrices determined by the eigenvalues of  $(A - LC)$ , and  $\bar{\mathcal{B}}_y \in \mathbb{R}^{np \times p}$  and  $\bar{\mathcal{B}}_u \in \mathbb{R}^{nm \times m}$  are also user-defined matrices which can make  $(\bar{\mathcal{A}}_y, \bar{\mathcal{B}}_y)$  and  $(\bar{\mathcal{A}}_u, \bar{\mathcal{B}}_u)$  controllable. It is easy to see that the data of  $\omega(k)$  and  $\sigma(k)$  are available since their system matrices  $(\bar{\mathcal{A}}_y, \bar{\mathcal{B}}_y)$  and  $(\bar{\mathcal{A}}_u, \bar{\mathcal{B}}_u)$  are user-defined.

Using the state parameterization in (11), the LQR problem can be solved by

$$u(k) = -K^* \bar{W} \varpi(k) := -K_{\bar{W}} \varpi(k) \quad (12)$$

when the state of the observer system converges to the real state. It is worth mentioning that, if all the eigenvalues of  $(A - LC)$  are chosen as 0, the observer state  $\hat{x}(k)$  converges to the real state  $x(k)$  when  $k \geq n$ .

Thus, by using the state reconstruction in (8) or the state parameterization in (11), the LQR problem can be solved by the output feedback controller (9) or (12) when the accurate knowledge of system matrices are known.

### 2.3. Problem Formulation

In this paper, we aim to solve the LQR problem of linear discrete-time systems with completely unknown system matrices and unmeasurable state, which can be formulated as follows:

**Problem 1.** For system (1) where the system matrices  $A$ ,  $B$ , and  $C$  are unknown, and the state  $x$  is unmeasurable, find an optimal control policy sequence  $u(k)$  to satisfy

$$V(x(0)) = \sum_{k=0}^{\infty} y^T(k) Q_y y(k) + u^T(k) R u(k) = x^T(0) P^* x(0)$$

where  $P^*$  is the solution to ARE (6). ■

The challenges of Problem 1 arise from the solution of ARE (6) and the design of the parameterization matrix  $\bar{M}$  or  $\bar{W}$ , which becomes difficult in the absence of knowledge of the system matrices  $A$ ,  $B$ , and  $C$ .

By combining the ADP method with a series of historical data of input and output, some ADP-based output feedback learning approaches focused on estimating the optimal control gain  $K_{\bar{M}}$  and  $K_{\bar{W}}$  directly, for instance, Chen et al. (2023); Gao et al. (2016); Lewis & Vamvoudakis (2011); Rizvi & Lin (2019, 2020, 2023). Particularly, it is proven in Rizvi & Lin (2023) that the requirement on the full row rank of  $\bar{W}$  is necessary to guarantee the convergence performance of the aforementioned ADP-based output feedback learning approaches, so is the same requirement on  $\bar{M}$  along the similar proof. However, as we can see from (8), under Assumption 1, the parameterization matrix  $\bar{M}$  may not be of full row rank when the system matrix  $A$  has the eigenvalue 0. It follows from Rizvi & Lin (2023, Theorem 4) that the parameterization matrix  $\bar{W}$  may not be of full row rank when the matrices  $(A - LC)$  and  $A$  have common eigenvalues. On the other hand, as stated in Postoyan et al. (2016), the stability of the closed-loop system under the learning control approach proposed in Lewis & Vamvoudakis (2011) cannot be guaranteed. Although this issue was eliminated in Chen et al. (2023); Rizvi & Lin (2019, 2023), the requirement of the convergence of state observer is needed to ensure the equivalence of the dynamic output feedback controller with the state feedback controller. This implies that the observer error will influence the convergence and optimality performance of the ADP-based output feedback learning control approaches in Chen et al. (2023); Rizvi & Lin (2019, 2023).

To deal with the above issues, the objective of this paper is to provide a generalized learning based output feedback solution to Problem 1 and a detailed analysis of convergence, stability, and optimality for the proposed output feedback learning control approach.

## 3. Main Results

In this section, we propose a novel output feedback learning control approach to solve the LQR problem. In particular, a new dynamic output feedback controller is designed that is equal to a static state feedback controller. Then, a data-driven learning algorithm is established to estimate the optimal control gain without prior knowledge of system matrices. Finally, the convergence, stability, and optimality analyses of the proposed learning control approach are given.

### 3.1. Dynamic Output Feedback Controller Design

Now we present a generalized dynamic output feedback controller design method, where the internal model is established without using any prior knowledge of the system dynamics. Moreover, the equivalence relationship between the proposed output feedback controller and the state feedback controller always holds even in the presence of observer error.