2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Derivative Feedback Control Using Reinforcement Learning

Muhammad Hamad Zaheer[1], Se Young Yoon[1] and Syed Ali Asad Rizvi[2]

Abstract— **In this paper, the model-free state-derivative and output-derivative feedback control of continuous-time dynamic systems are considered. First, an online iterative algorithm is designed following the reinforcement learning framework, in which measurements of the system state derivatives are collected to synthesize an optimal state-derivative feedback control law. The proposed approach employs the integral reinforcement learning technique to iteratively solve a derivative feedback algebraic Riccati equation. The iterative algorithm is extended to the output-derivative feedback case by introducing a state-parametrization scheme that reconstructs the state-derivative signal from the output derivatives and input data. Based on this parametrization, we develop an online iterative algorithm based on the reinforcement learning framework to determine the optimal output-derivative feedback controller. Convergence of the iterative algorithms to the analytical optimal control solutions is demonstrated. Numerical simulations motivated by practical applications demonstrate the benefits of our method compared to the standard output feedback reinforcement learning algorithms.**

I. Introduction

Many practical applications are restricted to measurements of state derivatives and output derivatives for feedback control. For instance, applications in vibration suppression [25], [8], [9], vehicles' active suspension systems and robotic manipulators [19] use accelerometers as the only sensing device. While accelerometer measurements can be integrated to obtain velocities and displacements, these signals are susceptible to large integration drift during continued operation due to the accumulation of noise. For such applications, directly utilizing the derivative signals of the state or output from the accelerometer can result in a simpler and more robust feedback control solution [1], [2].

Derivative feedback in control is also used for applications where the equilibrium point is unknown. For example, in the control of compressor surge [34], the equilibrium flow states are determined from empirically mapped compressor characteristic curves, and the resulting surge control solutions are highly sensitive to errors in the estimated equilibrium flow curve [4]. State-derivative feedback control is proposed in [29] to stabilize these types of systems. These results are extended in [5], [35] for improved robustness to model uncertainties and actuator saturation.

[1]Muhammad Hamad Zaheer and Se Young Yoon are with the Department of Electrical and Computer Engineering, University of New Hampshire, Durham, NH 03824, USA mz1038@usnh.edu, seyoung.yoon@unh.edu

[2]Syed Ali Asad Rizvi is with the Department of Electrical and Computer Engineering, Tennessee Technological University, Cookeville, TN 38505, USA srizvi@tntech.edu

This work is partially supported by the National Science Foundation (NSF 2218063)

The above discussed control methods for derivative feedback systems are model-based in their design, and they require a good understanding of the system dynamics. While robustness to dynamic uncertainties has been considered for derivative feedback control in [5], model-free derivative feedback control solutions have not been treated in the literature to the best of our knowledge.

Reinforcement Learning (RL) [31] enables learning of the optimal state-feedback and output-feedback controllers from the measurements of the systems' inputs, states and/or outputs to minimize a prescribed cost function. These model-free control solutions are thus reminiscent of adaptive optimal control methods [20], [36] and have been used to find optimal controllers for linear [32], [18], [14] and nonlinear continuous-time systems [16], [11], [30]. Reinforcement learning control has so far been successful in solving a wide range of classical control problems, including output regulation [6], [7], robust control [15], [33], [23], time delay control [12], and even inverse learning problems [22], with applications such as in robotics [13]. When the full states are not available for feedback control, output feedback solutions are presented in [21], [10], [11] for discrete-time systems. In [26], an output feedback RL approach is proposed for continuous-time systems, which overcomes the estimation bias issue due to the exploration signal [21] without the limitation of discounted cost functions [24].

In view of the need for model-free solutions in the control of systems with state-derivative or output-derivative measurements, and the potential demonstrated by RL-based solutions in control applications, this paper investigates RL-based methods for learning optimal control laws for state-derivative and output-derivative feedback systems. As part of the contributions of this work, an iterative algorithm is formulated to determine the optimal derivative-feedback control law that minimizes a prescribed quadratic cost function. The formulated algorithm extends the standard Policy Iteration (PI) algorithm to the state-derivative feedback case, and the convergence of the iterative solution to the optimal control law is demonstrated. An online implementation of the proposed iterative algorithm is then provided, which uses measurements of the state derivatives and inputs to determine the optimal derivative-feedback controller. The convergence of the proposed online algorithm to the optimal solution is guaranteed under a full-rank condition. Next, an output feedback iterative algorithm is presented to provide an optimal output-derivative feedback controller trained from the output-derivative measurements.

The remainder of this paper is structured as follows. The control problem considered in this work is introduced in

Section II. Iterative algorithms for the synthesis of the model-free optimal state-derivative and output-derivative feedback controllers are given in Section III and Section IV, respectively. Section V presents simulation results for the proposed solutions.

II. Problem Description and Preliminaries

Consider a linear system given by,

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t), \end{aligned} \tag{1}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$ and $y \in \mathbb{R}^p$ are the state, input, and output vectors, respectively. The objective is to design a state-derivative and output-derivative feedback control law such that the closed-loop system is asymptotically stable and the quadratic cost,

$$V(x(t)) = \int_t^\infty \left(\dot{y}^{\mathrm{T}}(\tau)Q_y\dot{y}(\tau) + u^{\mathrm{T}}(\tau)Ru(\tau)\right)d\tau, \tag{2}$$

is minimized. If the system states or outputs are available for feedback control, then iterative algorithms can be used to design a model-free optimal state feedback [32], [14] or output feedback controllers [26].

When only derivatives of states or outputs are available for feedback, they can be integrated to obtain the system states or outputs. However, this leads to drift in the estimated states, driving the system away from equilibrium, and causing potential instability [3]. Therefore, it is preferable to use the derivatives of the states or outputs as the feedback signal in such control applications.

A. State-Derivative Feedback Controller

The state-derivative feedback controller is given by,

$$u(t) = -K\dot{x}(t), \tag{3}$$

where $K \in \mathbb{R}^{m\times n}$ is the controller gain.

The dynamics of the closed-loop system is given by,

$$(I + BK)\dot{x} = Ax. \tag{4}$$

The optimal state-derivative feedback controller (3), which minimizes the cost function (2), can be obtained by solving an Algebraic Riccati Equation.

Proposition 2.1: ([2]) For system (1) with full-rank A, controllable pair (A, B) and detectable pair $(\sqrt{Q_y}C, A)$, the state-derivative feedback controller gain that minimizes the cost function (2) is given by,

$$K^* = -R^{-1}B^{\mathrm{T}}{A^{-1}}^{\mathrm{T}}P^*, \tag{5}$$

where $P^* > 0$ is the solution of the ARE,

$$P^*A^{-1} + {A^{-1}}^{\mathrm{T}}P^* - P^*A^{-1}BR^{-1}B^{\mathrm{T}}{A^{-1}}^{\mathrm{T}}P^* + Q_x = 0. \tag{6}$$

where ${\sqrt{Q_y}}^{\mathrm{T}}\sqrt{Q_y} = Q_y$ and $Q_x = C^{\mathrm{T}}Q_yC$.

III. Iterative Solution for Optimal State-Derivative Feedback Control

In this section, model-based and model-free iterative algorithms are proposed to solve the ARE (6) for estimating the optimal state-derivative feedback controller gain (5).

A. Model-Based Policy Iteration

First, an extension to the classical policy iteration (PI) algorithm [17] is presented here to account for the ARE (6) associated with the optimal state-derivative feedback control. We propose a PI algorithm to provide an iterative solution to (5) and (6).

Theorem 3.1: Let K_0 be any stabilizing state-derivative feedback controller gain and $P_i > 0$ be the solution of the Lyapunov equation,

$$P_iA_i^{-1} + {A_i^{-1}}^{\mathrm{T}}P_i + Q_x + K_i^{\mathrm{T}}RK_i = 0, \tag{7}$$

where $A_i^{-1} = A^{-1}(I + BK_i)$. For K_{i+1} calculated as,

$$K_{i+1} = -R^{-1}B^{\mathrm{T}}{A_i^{-1}}^{\mathrm{T}}P_i, \tag{8}$$

with $i = 0, 1, 2, \ldots$, the following hold,

1) $(I + BK_{i+1})^{-1}A$ is Hurwitz,
2) $P^* \le P_{i+1} \le P_i$,
3) $\lim_{i\to\infty} P_i = P^*$, $\lim_{i\to\infty} K_i = K^*$.

Proof: 1) For a stabilizing state-derivative feedback controller gain K_i, let $V_i(x(t))$ be the cost function of the form (2) associated with gain K_i,

$$V_i(x(t)) = \int_t^\infty \dot{x}^{\mathrm{T}}(t)(Q_x + K_i^{\mathrm{T}}RK_i)\dot{x}(t)d\tau. \tag{9}$$

As K_i is a stabilizing controller gain, the integral (9) converges to a finite value that is quadratic in x [2],

$$V_i(x(t)) = x^{\mathrm{T}}(t)P_ix(t). \tag{10}$$

Differentiating both sides of (9) and (10) along the state trajectories generated by K_i, and equating the right-hand side of the resulting equations, yield (7).

Now we show that (10) is also a Lyapunov function of the state trajectories generated by the controller K_{i+1}. Taking the derivative of $V_i(x(t))$ along the state trajectories generated by the controller K_{i+1}, results in,

$$\dot{V}_i(x(t)) = \dot{x}^{\mathrm{T}}(t)(P_iA_{i+1}^{-1} + {A_{i+1}^{-1}}^{\mathrm{T}}P_i)\dot{x}(t). \tag{11}$$

Substituting $A_{i+1}^{-1} = A_i^{-1} - A^{-1}B(K_i - K_{i+1})$ in (11) we get,

$$\begin{aligned} \dot{V}_i(x(t)) = \dot{x}^{\mathrm{T}}(t)\Big(&P_iA_i^{-1} + {A_i^{-1}}^{\mathrm{T}}P_i \\ &- P_iA^{-1}B(K_i - K_{i+1}) - (K_i - K_{i+1})^{\mathrm{T}}B^{\mathrm{T}}{A^{-1}}^{\mathrm{T}}P_i\Big)\dot{x}(t). \end{aligned}$$

Now using the controller update (8) and (7) we get,

$$\begin{aligned} \dot{V}_i(x(t)) = \dot{x}^{\mathrm{T}}(t)\Big(&P_iA_i^{-1} + {A_i^{-1}}^{\mathrm{T}}P_i \\ &+ K_{i+1}^{\mathrm{T}}R(K_i - K_{i+1}) + (K_i - K_{i+1})^{\mathrm{T}}RK_{i+1}\Big)\dot{x}(t), \end{aligned}$$

which can be further simplified to,

$$\begin{aligned} \dot{V}_i(x(t)) = \dot{x}^{\mathrm{T}}(t)\Big(&P_iA_i^{-1} + {A_i^{-1}}^{\mathrm{T}}P_i + K_i^{\mathrm{T}}RK_i \\ &- K_{i+1}^{\mathrm{T}}RK_{i+1} - (K_i - K_{i+1})^{\mathrm{T}}R(K_i - K_{i+1})\Big)\dot{x}(t). \end{aligned}$$

Using (7) we get,

$$\dot{V}_i(x(t)) = -\dot{x}^{\mathrm{T}}(t)(Q_x + K_{i+1}^{\mathrm{T}} R K_{i+1})\dot{x}(t) - \dot{x}^{\mathrm{T}}(t)\left((K_i - K_{i+1})^{\mathrm{T}} R (K_i - K_{i+1})\right)\dot{x}(t). \quad (12)$$

Therefore, $\dot{V}_i(x(t)) \leq 0$. Then, according to the LaSalle's invariance principle, all state trajectories of (4) converge to the set $\mathcal{I} = \{x | \dot{V}_i(x(t)) = 0\} = \{x | \dot{x} = 0\}$. As A in (4) is full rank, the set $\mathcal{I}$ contains only the trivial trajectory $x(t) = 0$. Therefore, K_{i+1} is a stabilizing state-derivative feedback controller gain.

2) Let $V_{i+1}(x(t))$ be the value function associated with the controller gain K_{i+1},

$$V_{i+1}(x(t)) = \int_t^\infty \dot{x}^{\mathrm{T}}(t)(Q_x + K_{i+1}^{\mathrm{T}} R K_{i+1})\dot{x}(t). \quad (13)$$

As K_{i+1} is stabilizing due to (12), the above integral is bounded. Let,

$$V_{i+1}(x(t)) = x^{\mathrm{T}}(t) P_{i+1} x(t). \quad (14)$$

Taking the derivative of V_{i+1} along the trajectories generated by K_{i+1} and using (13) and (14) we get,

$$P_{i+1} A_{i+1}^{-1} + {A_{i+1}^{-1}}^{\mathrm{T}} P_{i+1} = -(Q_x + K_{i+1}^{\mathrm{T}} R K_{i+1}). \quad (15)$$

By subtracting (15) from (7), and then using $A_i^{-1} = A_{i+1}^{-1} + A^{-1}B(K_i - K_{i+1})$ and the controller update equation (8) we get,

$$(P_i - P_{i+1})A_{i+1}^{-1} + {A_{i+1}^{-1}}^{\mathrm{T}}(P_i - P_{i+1}) + (K_i - K_{i+1})^{\mathrm{T}} R (K_i - K_{i+1})^{\mathrm{T}} = 0. \quad (16)$$

As A_{i+1} is Hurwitz due to (12), there exists a positive definite solution of the above Lyapunov equation,

$$P_i - P_{i+1} \geq 0. \quad (17)$$

Therefore, $P_{i+1} \leq P_i$. It can similarly be shown that $P^* \leq P_{i+1}$

3) As the sequence P_i is monotonic and all P_i are positive operators, there exists a $P_\infty > 0$ and a corresponding K_∞ such that, $\lim_{i\to\infty} P_i = P_\infty$ from the theorem on monotonic convergence of positive operators [17], and K_∞ satisfies (8). Substitute this K_∞ into (7) to get,

$$P_\infty A^{-1} + {A^{-1}}^{\mathrm{T}} P_\infty - P_\infty A^{-1} B R^{-1} B^{\mathrm{T}} {A^{-1}}^{\mathrm{T}} P_\infty + Q_x = 0, \quad (18)$$

which is the same ARE as in (6). Because P^* is a unique solution to (6), $K_\infty = K^*$ and $P_\infty = P^*$ must hold. ■

B. Model-free Online Policy Iteration

This section presents an online model-free algorithm based on reinforcement learning techniques that computes the iterations in Theorem 3.1 using perturbed measurements of the state $\bar{x}$, the state derivative $\dot{x}$ and the input u. These results extend existing model-free [14] and partially model-free [32] RL-based PI methods to the online estimation of optimal state-derivative feedback control.

A perturbed measurement $\bar{x} = x + x_{\mathrm{b}}$ of the system state x and an unknown static perturbation x_{b} is considered for the training of the proposed controller. This is motivated by the assumption that only measurements of the state derivatives are available for control, and integrating these measurements to determine the system states adds a slow-changing perturbation caused by the accumulation of noise and uncertainty in the initial conditions. The presence of the perturbation x_{b} in the state information also aligns with the challenges of controlling dynamic systems with uncertain equilibrium states, as discussed in Section I.

Under the control law (3) and for some positive real value $T > 0$, (10) and (7) yields,

$$x^{\mathrm{T}}(t)P_i x(t) - x^{\mathrm{T}}(t+T)P_i x(t+T) = \int_t^{t+T} \dot{x}^{\mathrm{T}}(\tau)(Q_x + K_i^{\mathrm{T}} R K_i)\dot{x}(\tau)d\tau - 2\int_t^{t+T} \left(u(\tau) + K_i\dot{x}(\tau)\right)^{\mathrm{T}} R K_{i+1}\dot{x}(\tau)d\tau. \quad (19)$$

The right-hand-side of (19) is rewritten as,

$$\int_t^{t+T} \dot{x}^{\mathrm{T}}(\tau)(Q_x + K_i^{\mathrm{T}} R K_i)\dot{x}(\tau)d\tau = I_{xx}^{t,t+T}\mathrm{vec}(Q_x + K_i^{\mathrm{T}} R K_i), \quad (20)$$

and

$$\int_t^{t+T} \left(u(\tau) + K_i\dot{x}(\tau)\right)^{\mathrm{T}} R K_{i+1}\dot{x}(\tau)d\tau = \left(I_{xx}^{t,t+T}(I_n \otimes K_i^{\mathrm{T}} R) + I_{xu}^{t,t+T}(I_n \otimes R)\right)\mathrm{vec}(K_{i+1}), \quad (21)$$

where

$$I_{xx}^{t,t+T} = \int_t^{t+T} \dot{x}^{\mathrm{T}}(\tau) \otimes \dot{x}^{\mathrm{T}}(\tau)d\tau,$$
$$I_{xu}^{t,t+T} = \int_t^{t+T} \dot{x}^{\mathrm{T}}(\tau) \otimes u^{\mathrm{T}}(\tau)d\tau,$$

and $\mathrm{vec}(\cdot)$ is the vectorization of a given matrix. It is also noted that $x = \bar{x} - x_{\mathrm{b}}$ and,

$$x^{\mathrm{T}}(t)P_i x(t) = \bar{x}^{\mathrm{T}}(t)P_i\bar{x}(t) + \epsilon^{\mathrm{T}}\bar{x} + x_{\mathrm{b}}^{\mathrm{T}}P_i x_{\mathrm{b}},$$

for $\epsilon = -2P_i x_{\mathrm{b}}$. Using this relation, we can write the left-hand-side of (19) as,

$$x^{\mathrm{T}}(t)P_i x(t) - x^{\mathrm{T}}(t+T)P_i x(t+T) = \mathrm{vec}(P_i)^{\mathrm{T}}\left(x_\kappa(t+T) - x_\kappa(t)\right) + \epsilon^{\mathrm{T}}\left(\bar{x}(t+T) - \bar{x}(t)\right), \quad (22)$$

where x_κ is the polynomial basis of the measured system states calculated using the Kronecker product $x_\kappa = \bar{x} \otimes \bar{x}$.

For N data samples collected over time intervals $[(j-1)T, jT]$, $j = 1, 2, \cdots, N$, (19), (20), (21) and (22) yield the following matrix equation,

$$X_i \begin{bmatrix} \mathrm{vec}(P_i) \\ \epsilon \\ \mathrm{vec}(K_{i+1}) \end{bmatrix} = Y_i, \quad (23)$$

where

$$X_i = \begin{bmatrix}\Delta_{x_\kappa},\ \Delta_x,\ -2I_{xx}(I_n \otimes K_i^{\mathrm{T}} R) - 2I_{xu}(I_n \otimes R)\end{bmatrix},$$
$$\Delta_{x_\kappa} = \begin{bmatrix}x_\kappa(T) - x_\kappa(0),\ \ldots,\ x_\kappa(NT) - x_\kappa((N-1)T)\end{bmatrix}^{\mathrm{T}},$$
$$\Delta_x = \begin{bmatrix}\bar{x}(T) - \bar{x}(0),\ \ldots,\ \bar{x}(NT) - \bar{x}((N-1)T)\end{bmatrix}^{\mathrm{T}},$$
$$Y_i = -I_{xx}\mathrm{vec}(Q_x + K_i^{\mathrm{T}} R K_i),$$
$$I_{xx} = \begin{bmatrix}I_{xx}^{0,T},\ I_{xx}^{T,2T},\ \ldots,\ I_{xx}^{(N-1)T,NT}\end{bmatrix}^{\mathrm{T}},$$
$$I_{xu} = \begin{bmatrix}I_{xu}^{0,T},\ I_{xu}^{T,2T},\ \ldots,\ I_{xu}^{(N-1)T,NT}\end{bmatrix}^{\mathrm{T}}.$$

If X_i is full column rank, then (23) has a solution

$$\begin{bmatrix}\mathrm{vec}(\hat{P}_i)\\ \hat{\epsilon}\\ \mathrm{vec}(\hat{K}_{i+1})\end{bmatrix} = (X_i^{\mathrm{T}} X_i)^{-1} X_i^{\mathrm{T}} Y_i. \tag{24}$$

Theorem 3.2: If X_i has full column rank, then $\hat{P}_i$ and $\hat{K}_{i+1}$ evaluated using (24) are equivalent to the solution of (7) and (8). Furthermore, if K_0 is a stabilizing gain for (4) then $\hat{P}_i$ and $\hat{K}_{i+1}$ converges to P^* and K^*, respectively, as $i \to \infty$.

Proof: Because P_i and K_{i+1} satisfy (23), if X_i has full column rank, then $\hat{P}_i = P_i$ and $\hat{K}_{i+1} = K_{i+1}$ are unique solutions given by (24). Therefore, the solution of (24) is equivalent to the evaluation of (7) and (8). According to Theorem 3.1, it then implies that $\hat{P}_i$ and $\hat{K}_{i+1}$ must converge to the optimal values. ■

Algorithm 1: Online State-Derivative Feedback PI

Data: Initial stabilizing controller K_0, threshold $\bar{\eta}$
Result: Optimal controller gain K^*
begin
  $i = 0$
  while $\|P_i - P_{i-1}\| > \bar{\eta}$ **do**
    Set $u = -K_i \dot{x}(t)$ as the input
    Collect training data during the interval $[iNT, (i+1)NT]$, where N is selected such that X_i is full rank
    Calculate P_i and K_{i+1} using (24)
    Increment counter $i \leftarrow i + 1$
  end
end

IV. Iterative Solution for Optimal Output-Derivative Feedback Control

To obtain an optimal output-derivative feedback controller, a parametrization of the system state will be used to obtain a model-free output-derivative feedback solution.

Theorem 4.1: There exists a state parametrization,

$$\eta(t) = \Gamma_u \alpha(t) + \Gamma_y \beta(t) \tag{25}$$

that converges exponentially to the state x for an observable system, where Γ_u and Γ_y are system-dependent matrices containing the system's transfer function coefficients, and

$$\alpha(t) = \begin{bmatrix}(\alpha^1)^{\mathrm{T}}(t) & (\alpha^2)^{\mathrm{T}}(t) & \ldots & (\alpha^m)^{\mathrm{T}}(t)\end{bmatrix}^{\mathrm{T}}, \tag{26}$$

$$\beta(t) = \begin{bmatrix}(\beta^1)^{\mathrm{T}}(t) & (\beta^2)^{\mathrm{T}}(t) & \ldots & (\beta^p)^{\mathrm{T}}(t)\end{bmatrix}^{\mathrm{T}}, \tag{27}$$

are given by,

$$\dot{\alpha}^i(t) = \mathcal{A}\alpha^i(t) + \mathcal{B}u_i(t), \forall i = 1, 2, \ldots, m \tag{28}$$

and,

$$\dot{\beta}^i(t) = \mathcal{A}\beta^i(t) + \mathcal{B}\dot{y}_i(t), \forall i = 1, 2, \ldots, p \tag{29}$$

where u_i and $\dot{y}_i$ are the i^{th} input and output derivatives, respectively. The matrix $\mathcal{A}$ is any user-defined Hurwitz matrix and $\mathcal{B} = \begin{bmatrix}0 & 0 & \ldots & 1\end{bmatrix}^{\mathrm{T}}$.

Due to space constraints, the proof of the above theorem has been omitted.

A. Model-Based Policy Iteration

Using the state parametrization (25), a revised description of the cost function (10) is given as

$$V(x(t)) = z^{\mathrm{T}}(t) \bar{P} z(t), \tag{30}$$

where $z(t) = \begin{bmatrix}\alpha^{\mathrm{T}}(t) & \beta^{\mathrm{T}}(t)\end{bmatrix}^{\mathrm{T}}$ and

$$\bar{P} = \begin{bmatrix}\Gamma_u^{\mathrm{T}} P \Gamma_u & \Gamma_u^{\mathrm{T}} P \Gamma_y\\ \Gamma_y^{\mathrm{T}} P \Gamma_u & \Gamma_y^{\mathrm{T}} P \Gamma_y\end{bmatrix}. \tag{31}$$

The output feedback controller takes the form,

$$u(t) = -\bar{K}\dot{z}(t), \tag{32}$$

where $\bar{K} = K\begin{bmatrix}\Gamma_u & \Gamma_y\end{bmatrix}$ and $\dot{z}(t) = \begin{bmatrix}\dot{\alpha}^{\mathrm{T}}(t) & \dot{\beta}^{\mathrm{T}}(t)\end{bmatrix}^{\mathrm{T}}$. Matrices $\bar{K}^*$ and $\bar{P}^*$ correspond to the optimal values K^* and P^*, respectively.

B. Model-Free Online Policy Iteration

Similar to the procedure presented in Sect. III-B, using (2), (30) and (32) lead to the following Bellman equation,

$$\begin{aligned} z^{\mathrm{T}}(t)\bar{P}_i z(t) - z^{\mathrm{T}}(t+T)\bar{P}_i z(t+T) \\ = \int_t^{t+T} \left(\dot{y}^{\mathrm{T}}(\tau) Q_y \dot{y}(\tau) + \dot{z}^{\mathrm{T}}(\tau)\bar{K}_i^{\mathrm{T}} R \bar{K}_i \dot{z}(\tau)\right) d\tau \\ - 2\int_t^{t+T} \left(u(\tau) + \bar{K}_i \dot{z}(\tau)\right)^{\mathrm{T}} R \bar{K}_{i+1} \dot{z}(\tau) d\tau. \end{aligned} \tag{33}$$

Algorithm 2: Online Output-Derivative Feedback PI

Data: Initial stabilizing controller $\bar{K}_0$, threshold $\bar{\eta}$
Result: Optimal controller gain $\bar{K}^*$
begin
  $i = 0$
  while $\|\bar{P}_i - \bar{P}_{i-1}\| > \bar{\eta}$ **do**
    Set $u = -\bar{K}_i z(t)$ as the input
    Collect measurements of the system's output derivatives and inputs
    Calculate $\bar{P}_i$ and $\bar{K}_{i+1}$ using (33)
    Increment counter $i \leftarrow i + 1$
  end
end

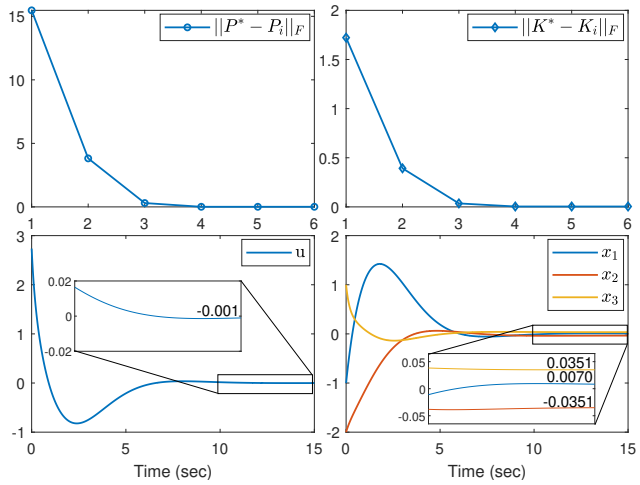

Fig. 1. Top Left and Right: Frobenius norm of error in the estimates of the value function matrix P and controller gain K. Bottom Left: Control input. Bottom Right: System states.

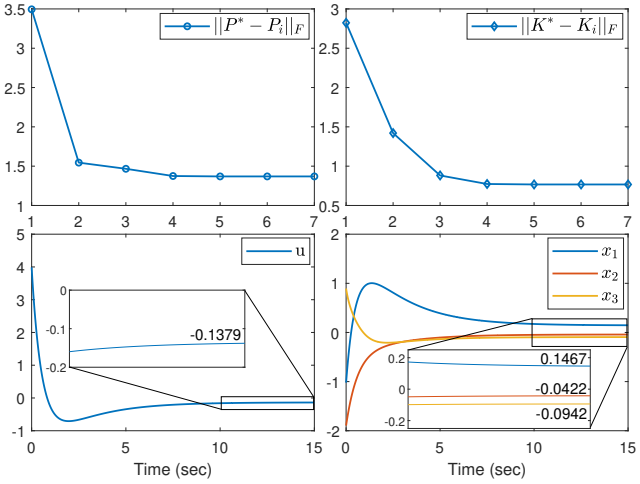

Fig. 2. Top Left and Right: Frobenius norm of error in the estimates of P and K. Bottom Left: Control input. Bottom Right: System states.

V. Numerical Simulation

A. Stabilization of Chaotic System

Algorithm 1 is used to find a local optimal controller for a Rössler attractor [27] given by the following equations,

$$\begin{aligned} \dot{x_1} &= -x_2 - x_3 + u, \\ \dot{x_2} &= x_1 + ax_2 + u, \\ \dot{x_3} &= b + x_3(x_1 - c) + u. \end{aligned} \tag{34}$$

We consider the values of the uncertain system parameters to be $\{a, b, c\} = \{0.2, 0.2, 5.7\}$, resulting in equilibrium at $x_r = \begin{bmatrix} 0.007 & -0.035 & 0.035 \end{bmatrix}^{\mathrm{T}}$. It is assumed in the example that the system dynamics are unknown and thus the equilibrium state can't be determined. On the other hand, with an initial local stabilizing control gain K_0 obtained from [5] and nominal values of a, b and c, a local optimal control law can be trained using Algorithm 1. An exploratory input signal $e(t) = \sum_{\omega=1}^{100} 10^{-3} \sin(\omega t)$ is used during training and is removed after the rank condition is satisfied. Figure 1 shows the response of the trained controller, and we observe that the closed-loop system converges to the equilibrium states and the control input converges to zero. The controller gain K_i also converges to the optimal value $K^* = \begin{bmatrix} -0.769 & 1.335 & 0.434 \end{bmatrix}$ given by the ARE (6).

If above results are compared to the standard PI algorithm in [14], we note that the latter requires information about the actual equilibrium states. Using nominal values of the system parameters, a nominal equilibrium point may be determined as $x_r^0 = \begin{bmatrix} 0.021 & -0.105 & 0.105 \end{bmatrix}^{\mathrm{T}}$. Figure 2 shows that because of the uncertain equilibrium point the estimated controller gain K_s and value function matrix P_s obtained by the standard PI method do not converge to the optimal values. Moreover, the trained controller does not stabilize the system at the true equilibrium resulting in non-zero steady-steady state control input.

B. Vibration Suppression Control

In this example, Algorithm 2 is used to find an optimal output-derivative controller to suppress vibrations in a car suspension system. The system matrices of the quarter-car dynamical model [28] are given as,

$$A = \begin{bmatrix} 0 & 1 & 0 & -1 \\ -58 & -3.4 & 0 & 3.4 \\ 0 & 0 & 0 & 1 \\ 284.9 & 16.9 & -3220.3 & -16.9 \end{bmatrix}, \tag{35}$$

$$B = \begin{bmatrix} 0 \\ 0.0034 \\ 0 \\ -0.0169 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{36}$$

Algorithm 2 is used to learn the output-derivative feedback controller using output derivatives in the reward function with $Q_y = 10^6$ and $R = 1$. The controller gain $\bar{K}_i$ converges to the optimal output feedback controller gain $\bar{K}^*$ as shown in Fig. 3. On the other hand, PI algorithm, which uses outputs during the training process, does not converge to the optimal controller due to bias caused by integrating the output-derivative signals (Fig. 3).

VI. Conclusion

In this paper, a new scheme was presented for the online iterative synthesis of optimal state-derivative and output-derivative feedback controllers under unknown system dynamics. The motivation of the derivative-feedback control scheme is to mitigate the effect of measurement drift in control applications, and it is demonstrated that the solutions of the proposed data-driven iterative algorithms converge to the analytically derived optimal control laws. Two numerical examples are offered to illustrate the theoretical results.

References

[1] T. H. S. Abdelaziz and M. Valášek, "Pole-placement for siso linear systems by state-derivative feedback," *Proceedings of the Institution of Electrical Engineers*, vol. 151, no. 4, pp. 377–385, Jul. 2004.

[2] T. H. S. Abdelaziz and M. Valášek, "State derivative feedback by lqr for time-invariant systems," *IFAC Proceedings Volumesl*, vol. 38, no. 1, pp. 435–440, 2005.

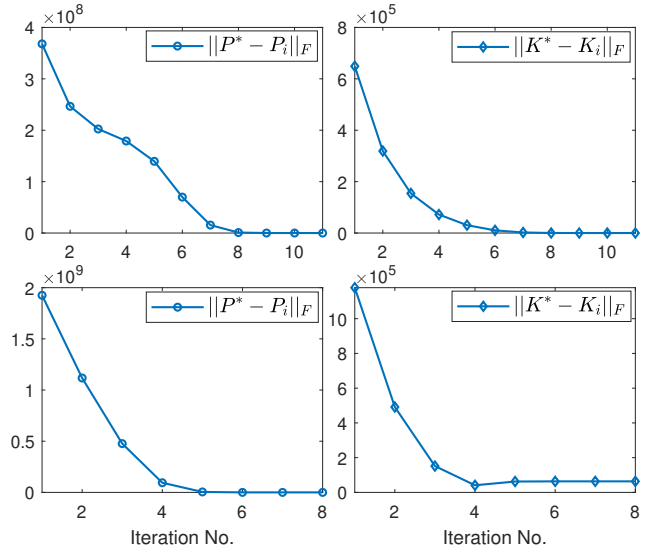

Fig. 3. Top left and right: Frobenius norm of error in the value function matrix P_i and controller gain K_i estimated using Algorithm 2. Bottom left and right: Frobenius norm of error in the value function matrix P_i and controller gain K_i estimated using output feedback PI Algorithm [26].

[3] K. M. Arthur, H. Basu, and S. Y. Yoon, "Control of compressor surge in systems with uncertain equilibrium states," in *2018 Annual American Control Conference (ACC)*, 2018, pp. 1758–1763.

[4] K. M. Arthur, H. Basu, and S. Y. Yoon, "Stabilization of compressor surge in systems with uncertain equilibrium flow," *ISA Transactions*, vol. 93, pp. 115–124, 2019.

[5] K. M. Arthur and S. Y. Yoon, "Robust stabilization at uncertain equilibrium by output derivative feedback control," *ISA Transactions*, vol. 107, pp. 40–51, 2020.

[6] C. Chen, H. Modares, K. Xie, F. L. Lewis, Y. Wan, and S. Xie, "Reinforcement learning-based adaptive optimal exponential tracking control of linear systems with unknown dynamics," *IEEE Transactions on Automatic Control*, vol. 64, no. 11, pp. 4423–4438, 2019.

[7] C. Chen, L. Xie, Y. Jiang, K. Xie, and S. Xie, "Robust output regulation and reinforcement learning-based output tracking design for unknown linear discrete-time systems," *IEEE Transactions on Automatic Control*, 2022.

[8] M. P. B. de Noyer and S. V. Hanagud, "Single actuator and multi-mode acceleration feedback control," *Journal of Intelligent Material Systems and Structures*, vol. 9, no. 7, pp. 522–533, 1998.

[9] J. Deur and N. Peric, "A comparative study of servosystems with acceleration feedback," in *Conference Record of the 2000 IEEE Industry Applications Conference. Thirty-Fifth IAS Annual Meeting and World Conference on Industrial Applications of Electrical Energy (Cat. No.00CH37129)*, vol. 3, 2000, pp. 1533–1540 vol.3.

[10] W. Gao, Y. Jiang, Z.-P. Jiang, and T. Chai, "Output-feedback adaptive optimal control of interconnected systems based on robust adaptive dynamic programming," *Automatica*, vol. 72, pp. 37–45, 2016.

[11] W. Gao and Z.-P. Jiang, "Learning-based adaptive optimal tracking control of strict-feedback nonlinear systems," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 6, pp. 2614–2624, 2017.

[12] W. Gao and Z.-P. Jiang, "Adaptive optimal output regulation of time-delay systems via measurement feedback," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 3, pp. 938–945, 2018.

[13] C. He, Y. Wan, Y. Gu, and F. L. Lewis, "Integral reinforcement learning-based multi-robot minimum time-energy path planning subject to collision avoidance and unknown environmental disturbances," *IEEE Control Systems Letters*, vol. 5, no. 3, pp. 983–988, 2020.

[14] Y. Jiang and Z.-P. Jiang, "Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics," *Automatica*, vol. 48, no. 10, pp. 2699–2704, 2012.

[15] Y. Jiang and Z.-P. Jiang, "Robust adaptive dynamic programming and feedback stabilization of nonlinear systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 5, pp. 882–893, 2014.

[16] Y. Jiang and Z.-P. Jiang, "Global adaptive dynamic programming for continuous-time nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 60, no. 11, pp. 2917–2929, 2015.

[17] D. Kleinman, "On an iterative technique for riccati equation computations," *IEEE Transactions on Automatic Control*, vol. 12, no. 1, pp. 114–115, 1968.

[18] J. Y. Lee, J. B. Park, and Y. H. Choi, "Integral q-learning and explorized policy iteration for adaptive optimal control of continuous-time linear systems," *Automatica*, vol. 48, no. 11, pp. 2850–2859, 2012.

[19] J. Y. Lew and S.-M. Moon, "Acceleration feedback control of compliant base manipulators," in *Proceedings of the 1999 American Control Conference (Cat. No. 99CH36251)*, vol. 3, 1999, pp. 1955–1959 vol.3.

[20] F. L. Lewis, D. Vrabie, and V. Syrmos, *Optimal Control*. New York, NY, USA: Wiley, 2012.

[21] F. L. Lewis and K. G. Vamvoudakis, "Reinforcement learning for partially observable dynamic processes: Adaptive dynamic programming using measured output data," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, no. 1, pp. 14–25, 2010.

[22] B. Lian, W. Xue, F. L. Lewis, and T. Chai, "Inverse reinforcement learning for adversarial apprentice games," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[23] B. Lian, W. Xue, F. L. Lewis, and T. Chai, "Robust inverse Q-learning for continuous-time linear systems in adversarial environments," *IEEE Transactions on Cybernetics*, 2021.

[24] R. Postoyan, L. Buşoniu, D. Nešić, and J. Daafouz, "Stability analysis of discrete-time infinite-horizon optimal control with discounted cost," *IEEE Transactions on Automatic Control*, vol. 62, no. 6, pp. 2736–2749, 2016.

[25] A. Preumont and N. Loix, "Active damping of a stiff beam-like structure with acceleration feedback," vol. 34, 1994, p. 23–26.

[26] S. A. A. Rizvi and Z. Lin, "Reinforcement learning-based linear quadratic regulation of continuous-time systems using dynamic output feedback," *IEEE Transactions on Cybernetics*, vol. 50, no. 11, pp. 4670–4679, 2020.

[27] O. Rössler, "An equation for continuous chaos," *Physics Letters A*, vol. 57, no. 5, pp. 397–398, 1976.

[28] Y. M. Sam, J. H. Osman, and M. A. Ghani, "A class of proportional-integral sliding mode control with application to active suspension system," *Systems & Control Letters*, vol. 51, no. 3, pp. 217–223, 2004.

[29] T. Shigekuni and T. Takimoto, "Stabilization of uncertain equilibrium points by dynamic state-derivative feedback control," in *13th International Conference on Control, Automation and Systems*, 2013, pp. 23–27.

[30] K. G. Vamvoudakis, D. Vrabie, and F. L. Lewis, "Online adaptive algorithm for optimal control with integral reinforcement learning," *International Journal of Robust and Nonlinear Control*, vol. 24, no. 17, pp. 2686–2710, 2014.

[31] K. G. Vamvoudakis, Y. Wan, F. L. Lewis, and D. Cansever, *Handbook of Reinforcement Learning and Control*. Springer, 2021.

[32] D. Vrabie, O. Pastravanu, M. Abu-Khalaf, and F. Lewis, "Adaptive optimal control for continuous-time linear systems based on policy iteration," *Automatica*, vol. 45, no. 2, pp. 477–484, 2009.

[33] Y. Yang, Y. Wan, J. Zhu, and F. L. Lewis, "h_∞ tracking control for linear discrete-time systems: model-free q-learning designs," *IEEE Control Systems Letters*, vol. 5, no. 1, pp. 175–180, 2020.

[34] S. Y. Yoon, Z. Lin, and P. Allaire, *Control of surge in centrifugal compressors by active magnetic bearings: Theory and implementation*. London: Springer, 2012.

[35] M. H. Zaheer, K. M. Arthur, and S. Y. Yoon, "Derivative feedback control of nonlinear systems with uncertain equilibrium states and actuator constraints," *Automatica*, vol. 127, p. 109495, 2021.

[36] H. Zhang, D. Liu, Y. Luo, and D. Wang, *Adaptive dynamic programming for control: algorithms and stability*. Springer Science & Business Media, 2012.