Value Iteration for Linear Quadratic Optimal Control of Single-Input Single-Output Systems via Output Feedback

Corrado Possieri

***Abstract*—A value iteration approach based solely on input/output measurements is proposed to solve linear quadratic (LQ) optimal control problems for single-input, single-output (SISO) continuous-time systems. Such an algorithm is designed by coupling an adaptive Luenberger observer with an indirect value iteration architecture. The continuous-time implementation of this controller requires that the gathered estimates are strongly controllable. A hybrid adaptation mechanism is envisioned to overcome such a requirement. The effectiveness of the proposed approach is validated via numerical simulations.**

***Index Terms*—Reinforcement learning, linear systems, optimal control.**

I. Introduction

THANKS to their ease of use and adaptability in practical settings, optimal control problems associated to linear plants and quadratic cost functionals have received interest from both scholars and practitioners [1]. The solution to such linear quadratic (LQ) optimization problems is often described in state feedback form and is based on an exact model of the system dynamics [1]. A significant amount of research effort has been expended to bypass these requirements by using iterative, data-driven algorithms that may run online and require little a priori knowledge of the system [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14].

The main goal of this letter is to propose an adaptive, output feedback controller based on value iteration that is able to determine the solution to an LQ problem just by using input/output measurements [15], [16], [17]. Such an objective is pursued by coupling an adaptive Luenberger observer for the system in suitable coordinates [18] with the value iteration algorithm given in [19]. Differently from the methodologies given in [2], [3], [5], such a controller is built using a value iteration paradigm rather than a policy iteration architecture, and hence it does not require an initial stabilizing feedback gain. Further, differently from [6], [7], [19], it can be implemented by using just measurements of the input and of the output of the controlled plant, without having at one's disposal measurements of its state. Moreover, differently from [6], the algorithm proposed in this letter does not rely on delayed measurements of the state and the control input, but just on the current value of the input and output of the plant. Comparing the controller proposed in this letter with the one given in [10], it is an on-policy (rather than off-policy) adaptive control algorithm that, using value iteration (rather than policy iteration), reconstructs the optimal control policy and the value function for the considered LQ problem, without requiring an initial stabilizing gain. Differently from the approaches given in [13], [14] to deal with data-driven optimal control using linear programming, the proposed scheme does not require storing the trajectories of the system, but performs the adaptation online while controlling the plant. Finally, the proposed scheme has a reduced computational complexity with respect to the ones given in [18, Ch. 7] since it does not need to solve pole placement equations at running time, by delegating these computationally intensive tasks either to a boundary layer built using the forward-in-time differential Riccati equation or requiring these computations just at the sampling times.

The proposed controller can be framed as an on policy, indirect reinforcement learning architecture built on value iteration. In fact, it is based on *value iteration* since it continuously updates the estimate of the value function to determine the optimal feedback gain, it is *indirect* since the estimate of the value function is built upon estimates of the dynamical matrices of the system, and it is *on policy* since the current estimate of the optimal gain is used to control the plant. A preliminary version of this controller has been given in [20]. The main contributions of this letter with respect to [20] are: (i) herein all the proofs are reported to demonstrate the effectiveness of the proposed value iteration algorithms; (ii) a novel hybrid mechanism based on the hybrid adaptive Luenberger observer [18, Sec. 5.3.2] is provided to overcome the requirement made in [20] about strong controllability of the estimated system; (iii) new simulations are reported to asses the effectiveness of this new mechanism.

Manuscript received 13 February 2024; revised 3 April 2024; accepted 22 April 2024. Date of publication 25 April 2024; date of current version 7 May 2024. Recommended by Senior Editor V. Ugrinovskii.

The author is with the Dipartimento di Ingegneria Civile e Ingegneria Informatica, Università degli Studi di Roma Tor Vergata, 00133 Rome, Italy, and also with the Istituto di Analisi dei Sistemi ed Informatica "A. Ruberti," Consiglio Nazionale delle Ricerche, 00185 Rome, Italy (e-mail: corrado.possieri@uniroma2.it).

Digital Object Identifier 10.1109/LCSYS.2024.3393847

Notation: The symbol I denotes the identity matrix of suitable dimensions. Given $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, let $\operatorname{col}(x, y) = [\,x^\top \;\; y^\top\,]^\top$. The symbol s denotes the *differential operator* and $\frac{1}{s}$ its inverse. Given $n \in \mathbb{N}$, let $\alpha_{n-1} = [\,s^{n-1} \;\cdots\; s \;1\,]^\top$. Given $(\hat{a}, \hat{b}) \in \mathbb{R}^n \times \mathbb{R}^n$, let $\rho(\hat{a}, \hat{b})$ be the resultant [21, Chap. 3, §6] of the polynomials $s^n + \hat{a}^\top \alpha_{n-1}$ and $\hat{b}^\top \alpha_{n-1}$, so that $s^n + \hat{a}^\top \alpha_{n-1}$ and $\hat{b}^\top \alpha_{n-1}$ have a common factor if and only if $\rho(\hat{a}, \hat{b}) = 0$. The time-varying pair $(\hat{a}(t), \hat{b}(t))$ is *strongly controllable* if there exists $\chi^\star > 0$ such that $\rho^2(\hat{a}(t), \hat{b}(t)) \geq \chi^\star$, $\forall t \geq 0$, i.e., $s^n + \hat{a}^\top(t)\alpha_{n-1}$ and $\hat{b}^\top(t)\alpha_{n-1}$ do not have a common fact for all $t \geq 0$ [18, Ch. 7]. Let $|\cdot|_p$ denote the *p-norm* of the vector/matrix at argument. To simplify the notation, let $|\cdot|$ denote the *2-norm* of the argument. The $\mathcal{L}_p$ *norm* of the continuous-time vector signal x is $\|x\|_p = (\int_0^\infty |x(\tau)|_p^p \,\mathrm{d}\tau)^{\frac{1}{p}}$, whereas its $\mathcal{L}_\infty$ *norm* is $\|x\|_\infty = \sup_{t\geq 0} |x(t)|_\infty$ A signal x is in $\mathcal{L}_p$ if $\|x\|_p$ exists and is bounded, whereas it is in $\mathcal{L}_\infty$ if $\|x\|_\infty$ exists and is bounded. The $\mathcal{L}_{2\delta}$ *norm* of the signal x is $\|x\|_{2\delta} = (\int_0^t \mathrm{e}^{-\delta(t-\tau)} x^\top(\tau) x(\tau)\,\mathrm{d}\tau)^{\frac{1}{2}}$ where $\delta > 0$ is a constant [18, Sec. 3.3.2]. Given a sequence x_k, $k \in \mathbb{N}$, its ℓ_p norm is $\|x_k\|_p = (\sum_{k=0}^\infty |x_k|_p^p)^{\frac{1}{p}}$, whereas its ℓ_∞ *norm* is $\|x_k\|_\infty = \sup_{k\geq 0} |x_k|_\infty$. The sequence x_k is in ℓ_p (respectively, ℓ_∞) if $\|x_k\|_p$ (respectively, $\|x_k\|_\infty$) exists and is bounded. A signal ϕ is *persistently exciting* if there exist constants $\alpha_0 > 0$, $\alpha_1 > 0$, and $T_0 \geq 0$ such that $\alpha_0 I \leq \int_t^{t+T_0} \phi(\tau)\,\phi^\top(\tau)\,\mathrm{d}\tau \leq \alpha_1 I$, $\forall t \geq 0$. A signal d is *stationary* if the limit $\Xi_d(t) = \lim_{T\to+\infty} \frac{1}{T} \int_{t_0}^{t_0+T} d(\tau) d^\top(t+\tau)\,\mathrm{d}\tau$ exists uniformly with respect to t_0. A stationary signal is *sufficiently rich of order n* if the support of the Fourier transform of $\Xi_d(t)$ contains at least n points [18, Sec. 5.5.1], e.g., a signal d obtained as the sum of n sinusoidal signals at different frequencies is sufficiently rich of order $2n$.

II. LQ Optimal Control via Output Feedback

Consider the linear, time-invariant (LTI), single-input, single-output (SISO) continuous-time system

$$\dot{x} = A\,x + b\,u, \qquad y = C\,x, \tag{1}$$

where $x(t) \in \mathbb{R}^n$ is the state vector, $u(t) \in \mathbb{R}$ is the input, and $y(t) \in \mathbb{R}$ is the output, and the cost functional

$$J(x_0, u) = \int_0^\infty \Big(y^2(\tau, x_0, u) + r\,u^2(\tau)\Big)\,\mathrm{d}\tau, \tag{2}$$

where $r > 0$ is the *input weight*, $x_0 \in \mathbb{R}^n$ is the initial condition of system (1), and, with a slight abuse of notation, $y(\tau, x_0, u)$ is the output response of system (1) at time τ from the initial condition x_0 with input u. The next assumption is made all throughout this letter.

Assumption 1: The pair (A, b) is controllable and the pair (A, C) is observable.

Under Assumption 1, by classical results about LQ optimal control [1], letting $P^\star$ be the unique positive definite (PD) solution to the algebraic Riccati equation (ARE) $A^\top P + P A + C^\top C - r^{-1} P\,b\,b^\top P = 0$, and letting $K^\star = -r^{-1}\,b^\top P^\star$, the input u that minimizes the cost (2) subject to the dynamics (1) is $u^\star = K^\star x$. Furthermore, the optimal value function of the LQ optimization problem given by system (1) and the cost (2) is $V(x_0) = \min_u J(x_0, u) = x_0^\top P^\star x_0$.

Consider the following problem.

Problem 1: Suppose that the dimension n of the state of system (1) is known, but its dynamical matrices A, b, and C are unknown. Design an output feedback controller that recasts the optimal control $u^\star$ and the optimal value function just by using input/output measurements.

It is worth pointing out that the minimal requirement [22] about system (1) to guarantee that there exists a solution to the LQ optimization problem given by (1) and (2) that makes the closed loop asymptotically stable is that the pair (A, b) is stabilizable and the pair (A, C) is detectable. The stronger Assumption 1 is made here to guarantee that the system can be rewritten in coordinates that are suitable for identification and observer design and that no zero/pole cancellation occurs in the closed-loop transfer function. In fact, under Assumption 1, there exists a change of coordinates [23] $\bar{x} = T\,x$, with $T \in \mathbb{R}^{n\times n}$ being a non-singular matrix, such that the dynamics of system (1) in the $\bar{x}$-coordinates read as

$$\dot{\bar{x}} = \bar{A}\,\bar{x} + \bar{b}\,u, \qquad y = \bar{C}\,\bar{x}, \tag{3}$$

where $\bar{b} = T\,b$,

$$\bar{A} = \left[\,-\bar{a}\left|\begin{matrix} I \\ 0 \end{matrix}\right.\right], \quad \bar{C} = [\,1 \;\; 0 \;\; \cdots \;\; 0\,],$$

and $\bar{a}, \bar{b} \in \mathbb{R}^n$. The form (3) is usually referred to as the *observability canonical form* of system (1). Let $\bar{P}^\star$ be the unique PD solution to the ARE

$$\bar{A}^\top P + P\bar{A} + \bar{C}^\top\bar{C} - r^{-1}\,P\,\bar{b}\,\bar{b}^\top P = 0.$$

Since $x = T^{-1}\,\bar{x}$, such a matrix satisfies $\bar{P}^\star = T^{-\top} P^\star T^{-1}$. Further, the optimal control $u^\star$ can be equivalently obtained as $u^\star = \bar{K}^\star\,\bar{x}$, where $\bar{K}^\star = -r^{-1}\,\bar{b}^\top\,\bar{P}^\star$. Similarly, the optimal value function of the LQ optimization problem given by (1) and (2) can be equivalently obtained as $\bar{V}(\bar{x}_0) = \bar{x}_0^\top \bar{P}^\star\,\bar{x}_0$, where $\bar{x}_0 = T\,x_0$ is the initial condition of the system in the $\bar{x}$-coordinates.

Problem 1 is tackled by coupling an adaptive Luenberger observer for system (1) in the $\bar{x}$-coordinates [18, Sec. 5.3.2] with the value iteration algorithm presented in [19]. Namely, consider the output feedback controller

$$\phi_1 = \frac{\alpha_{n-1}(s)}{\Lambda^\star(s)}\,u,\; \phi_2 = -\frac{\alpha_{n-1}(s)}{\Lambda^\star(s)}\,y,\; z = \frac{s^n}{\Lambda^\star(s)}\,y, \tag{4a}$$

$$\dot{\hat{b}} = \gamma\,\frac{z - \hat{b}^\top\phi_1 - \hat{a}^\top\phi_2}{1 + \beta\left(\phi_1^\top\phi_1 + \phi_2^\top\phi_2\right)}\,\phi_1, \tag{4b}$$

$$\dot{\hat{a}} = \gamma\,\frac{z - \hat{b}^\top\phi_1 - \hat{a}^\top\phi_2}{1 + \beta\left(\phi_1^\top\phi_1 + \phi_2^\top\phi_2\right)}\,\phi_2, \tag{4c}$$

$$\hat{A} = \left[\,-\hat{a}\left|\begin{matrix} I \\ 0 \end{matrix}\right.\right], \tag{4d}$$

$$\dot{\hat{x}} = \hat{A}\,\hat{x} + \hat{b}\,u + \left(a^\star - \hat{a}\right)\left(y - \bar{C}\,\hat{x}\right), \tag{4e}$$

$$\varepsilon\,\dot{P} = \hat{A}^\top P + P\hat{A} + \bar{C}^\top\bar{C} - r^{-1}\,P\,\hat{b}\,\hat{b}^\top P, \tag{4f}$$

$$\hat{u}^\star = -r^{-1}\,\hat{b}^\top P\,\hat{x}, \qquad u = \hat{u}^\star + d, \tag{4g}$$

where $a^\star \in \mathbb{R}^n$ has to be designed so that

$$A^\star = \left[\,-a^\star\left|\begin{matrix} I \\ 0 \end{matrix}\right.\right]$$

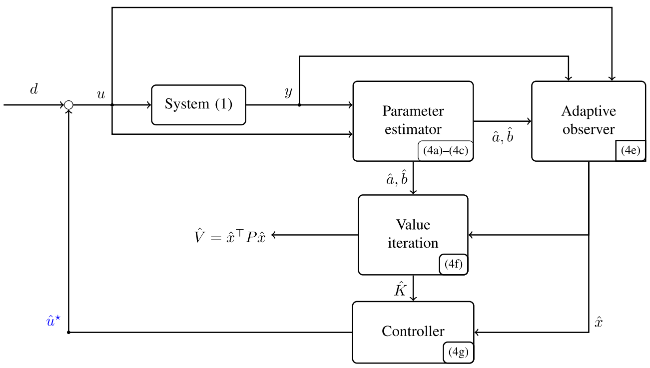

Fig. 1. Schematic representation of the controller (4).

is Hurwitz (i.e., all its eigenvalues have negative real part), $\Lambda^\star(s) = \det(sI - A^\star)$, $\gamma > 0$ is the *adaptation gain*, $\beta > 0$ is a normalizing constant, $\varepsilon > 0$ is a small parameter, $\phi_1 \in \mathbb{R}^n$, and $\phi_2 \in \mathbb{R}^n$, $z \in \mathbb{R}$ are filtrations of the input and output of system (1), respectively, $\hat{a} \in \mathbb{R}^n$, $\hat{b} \in \mathbb{R}^n$, and $P \in \mathbb{R}^{n\times n}$ are estimates of $\bar{a}$, $\bar{b}$, and $\bar{P}^\star$, respectively, $P(0) = P^\top(0) \succeq 0$, and d is the probing input, to be used in order to guarantee that the subsequent assumption about the persistence of excitation of the signal $\phi = \mathrm{col}(\phi_1, \phi_2)$ is met. Note that, by construction, if the pair $(\hat{a}(t), \hat{b}(t))$ is strongly controllable, then the pair $(\hat{A}(t), \hat{b}(t))$ is controllable for all $t \geq 0$. Figure 1 depicts a schematic representation of the controller (4).

With reference to Figure 1, system (4a)–(4c) is a parameter estimator for (3) based on *normalized gradient descent*, system (4e) is an *adaptive Luenberger observer* that estimates the state $\bar{x}$ of system (3) using the parameters estimated by system (4a)–(4c), system (4f) is a *value iteration algorithm* that builds estimates $\hat{V}$ and $\hat{K}$ of the value function $\bar{V}$ and of the optimal feedback gain $\bar{K}^\star$, respectively, on the basis of the current estimates $\hat{A}$ and $\hat{b}$ of $\bar{A}$ and $\bar{b}$, respectively, and (4g) is a controller that feeds back the optimal control corresponding to the current estimates of the parameters $\bar{a}$, $\bar{b}$ and of the state $\bar{x}$, provided by the parameter estimator and the adaptive observer, respectively.

The next theorem states that the output feedback controller (4) semi-globally solves Problem 1, provided that the estimated pair $(\hat{a}(t), \hat{b}(t))$ is strongly controllable.

Theorem 1: Let Assumption 1 hold, consider the interconnection of system (1) and the output feedback controller (4), let $d \in \mathcal{L}_\infty$, and suppose that the pair $(\hat{a}(t), \hat{b}(t))$ is strongly controllable. Then

1) for each $\Delta > 0$, there exists $\varepsilon^\star_\Delta > 0$ such that if $|\hat{a}(0) - \bar{a}| \leq \Delta$, $|\hat{b}(0) - \bar{b}| \leq \Delta$, $|\hat{x}(0) - \bar{x}(0)| < \Delta$, $|P(0) - \bar{P}^\star| < \Delta$ and $P(0) = P^\top(0) \succ 0$, then, for all $\varepsilon \in (0, \varepsilon^\star_\Delta)$, all the signals in the closed-loop are in $\mathcal{L}_\infty$;
2) if, additionally to the above conditions, the signal $\phi = \mathrm{col}(\phi_1, \phi_2)$ is persistently exciting, then

$$\lim_{t\to+\infty} \hat{a}(t) - \bar{a} = 0, \quad \lim_{t\to+\infty} \hat{b}(t) - \bar{b} = 0, \tag{5a}$$
$$\lim_{t\to+\infty} \hat{x}(t) - \bar{x}(t) = 0, \quad \lim_{t\to+\infty} P(t) - \bar{P}^\star = 0, \tag{5b}$$

with exponential convergence rate.

Proof: Letting $(\hat{A}, \hat{b})$ be a controllable estimate of (A, b), let $\hat{P}$ be the unique positive definite solution to the ARE

$$\hat{A}^\top P + P\hat{A} + \bar{C}^\top \bar{C} - r^{-1} P \hat{b} \hat{b}^\top P = 0. \tag{6}$$

Hence, considering the dynamics of the boundary layer (4f) and that, by (6), one has $\bar{C}^\top \bar{C} = -\hat{A}^\top \hat{P} - \hat{P}\hat{A} + r^{-1}\hat{P}\hat{b}\hat{b}^\top \hat{P}$, the dynamics of the error $\tilde{P} = P - \hat{P}$ are $\varepsilon\dot{\tilde{P}} = (\hat{A} - r^{-1}\hat{b}\hat{b}^\top \hat{P})^\top \tilde{P} + \tilde{P}(\hat{A} - r^{-1}\hat{b}\hat{b}^\top \hat{P}) - r^{-1}\tilde{P}\hat{b}\hat{b}^\top \tilde{P}$. Therefore, since the pair $(\hat{A}, \hat{b})$ is controllable, the matrix $\hat{A}_{\mathrm{cl}} = \hat{A} - r^{-1}\hat{b}\hat{b}^\top \hat{P}$ is Hurwitz, thus implying that $\tilde{P} = 0$ is exponentially stable. In fact, by classical results on the differential Riccati equation [1, pp. 21–23], if $P(0) \succeq 0$, then $P(t) \succeq 0$, $\forall t \geq 0$, and P converges to $\hat{P}$ exponentially.

Consider now the reduced system given by (4a)-(4e), with the control input provided by (4g) with $P = \hat{P}$. Since, by assumption, the pair $(\hat{a}(t), \hat{b}(t))$ is strongly controllable, one has that $\hat{P}(t)$ exists for all $t \geq 0$ and is such that the matrix $\hat{A}_{\mathrm{cl}}(t)$ is Hurwitz at each frozen $t \geq 0$ [1, p. 23]. By [18, Th. 4.3.2], letting $m^2 = 1 + \beta\phi^\top\phi$, $\tilde{a} = \bar{a} - \hat{a}$, $\tilde{b} = \bar{b} - \hat{b}$, $\tilde{\theta} = \mathrm{col}(\tilde{b}, \tilde{a})$, and $\psi = \frac{\tilde{\theta}^\top \phi}{m^2}$, one has $\psi, \psi\, m, \hat{b}, \hat{a}, \dot{\hat{b}}, \dot{\hat{a}}, \hat{A}, \hat{b} \in \mathcal{L}_\infty$ and $\psi, \psi\, m, \dot{\hat{b}}, \dot{\hat{a}}, \dot{\hat{A}}, \dot{\hat{b}} \in \mathcal{L}_2$, independently of the boundedness of ϕ. Thus, since $\hat{P}$ solves (6) and $\hat{A}_{\mathrm{cl}} = \hat{A} - r^{-1}\hat{b}\hat{b}^\top \hat{P}$, one has that $\hat{P}, \hat{A}_{\mathrm{cl}} \in \mathcal{L}_\infty$ and $\dot{\hat{P}}, \dot{\hat{A}}_{\mathrm{cl}} \in \mathcal{L}_2$.

The dynamics of the estimate $\hat{x}$ and of the estimation error $\tilde{x} = \bar{x} - \hat{x}$ are given by

$$\dot{\hat{x}} = \hat{A}_{\mathrm{cl}}\,\hat{x} + \hat{b}\,d + \left(a^\star - \hat{a}\right)\tilde{y}, \tag{7a}$$
$$\dot{\tilde{x}} = A^\star\,\tilde{x} + \tilde{b}\,u - \tilde{a}\,y, \tag{7b}$$

where $\tilde{y} = \bar{C}\tilde{x}$. Since $\hat{A}_{\mathrm{cl}} \in \mathcal{L}_\infty$, $\hat{A}_{\mathrm{cl}}(t)$ is Hurwitz for each $t \geq 0$, and $\dot{\hat{A}}_{\mathrm{cl}} \in \mathcal{L}_2$, by [18, Lemma 3.3.3, Th. 3.4.11], $\exists\delta_0 > 0$ such that, for all $\delta \in (0, \delta_0)$, one has $\|\hat{x}\|_{2\delta} \leq c_1 + c_1\|\tilde{y}\|_{2\delta}$, where $c_1 > 0$ is a sufficiently large constant. Thus, since $y = \bar{C}\hat{x} + \tilde{y}$ and $\hat{B}, \hat{P}, d \in \mathcal{L}_\infty$, one has $\|y\|_{2\delta} \leq \|\bar{C}\|\,\|\hat{x}\|_{2\delta} + \|\tilde{y}\|_{2\delta} \leq c_2 + c_2\|\tilde{y}\|_{2\delta}$ and $\|u\|_{2\delta} \leq \|\hat{B}\|\,\|P\|\,\|\hat{x}\|_{2\delta} + \|d\|_{2\delta} \leq c_2 + c_2\|\tilde{y}\|_{2\delta}$, where $c_2 > 0$ is a sufficiently large constant. Inspecting (7b), one has $\tilde{y} = \sum_{i=1}^n (\frac{s^{n-i}}{\Lambda^\star(s)}(\tilde{b}_i\,u) - \frac{s^{n-i}}{\Lambda^\star(s)}(\tilde{a}_i\,y)) + y_m$, where y_m is an exponentially vanishing term. By [18, Lemmaa A.1], one has that $\frac{s^{n-i}}{\Lambda^\star(s)}(\tilde{b}_i\,u) = \tilde{b}_i(\frac{s^{n-i}}{\Lambda^\star(s)}u) + W_{ci}(s)((W_{bi}(s)\,u)\ \dot{\tilde{b}}_i)$ and $\frac{s^{n-i}}{\Lambda^\star(s)}(\tilde{a}_i\,y) = \tilde{a}_i(\frac{s^{n-i}}{\Lambda^\star(s)}y) + W_{ci}(s)((W_{bi}(s)\,y)\ \dot{\tilde{a}}_i)$ where

$$W_{ci} = -\bar{C}\left(sI - A^\star\right)^{-1}, \quad W_{bi} = \left(sI - A^\star\right)^{-1} e_i, \tag{8}$$

and e_i is the ith column of the identity matrix. Therefore, since $\phi_{1i} = \frac{s^{n-i}}{\Lambda^\star(s)}u$ and $\phi_{2i} = -\frac{s^{n-i}}{\Lambda^\star(s)}y$, one has that $\tilde{y} = \psi\, m^2 + \nu + y_m$, where $\nu = \sum_{i=1}^n W_{ci}(s)((W_{bi}(s)\,u)\ \dot{\tilde{b}}_i - (W_{bi}(s)\,y)\ \dot{\tilde{a}}_i)$. Letting $\mu^2(t) = 1 + \|u\|^2_{2\delta} + \|y\|^2_{2\delta}$, by [18, Lemmaa 3.3.2], one has $\frac{\phi}{\mu}, \frac{m}{\mu} \in \mathcal{L}_\infty$. Thus, one has that $\|\tilde{y}\|_{2\delta} \leq c_3 + c_3\,\|(\psi\, m\, \mu)\|_{2\delta} + c_3\,\|(\dot{\tilde{\theta}}\,\mu)\|_{2\delta}$, where $c_3 > 0$ is a sufficiently large constant. Hence, it results that

$$\mu^2(t) \leq c_4 + c_4\,\|(\psi\, m\, \mu)\|^2_{2\delta} + \|\left(\dot{\tilde{\theta}}\,\mu\right)\|^2_{2\delta} \tag{9}$$

where $c_4 > 0$ is a sufficiently large constant. Therefore, since $\psi\, m, |\dot{\tilde{\theta}}| \in \mathcal{L}_2$, by the Bellman-Gronwall Lemma [18, Sec. 3.3.4], one has that $\mu \in \mathcal{L}_\infty$ and hence also $\phi \in \mathcal{L}_\infty$.

Furthermore, by (7b) and [18, Lemmaa 3.3.3], one has that $\tilde{x}, \tilde{y} \in \mathcal{L}_\infty$. Hence, by (7a), one has $\hat{x}, u, x, y, z \in \mathcal{L}_\infty$. Therefore, item (1) follows by [24, Sec. 11.4].

By [18, Th. 4.3.2], if ϕ is persistently exciting, then $\hat{a} - a$ and $\hat{b} - b$ converge exponentially to zero. Thus, since $A^\star$ in (7b) is Hurwitz and $u, y \in \mathcal{L}_\infty$, one has $\lim_{t \to +\infty} \tilde{x}(t) = 0$. Hence, item (2) follows by [24, Th. 11.4]. ∎

The assumption that the pair $(\hat{a}, \hat{b})$ is strongly controllable is crucial in Theorem 1 to guarantee that the solution $\hat{P}(t)$ to the ARE (6) exists for all times $t \geq 0$. Such an assumption is customary when dealing with schemes based on indirect adaptive control [18, Ch. 7]. In fact, loss of stabilizability of the pair $(\hat{A}(t), \hat{b}(t))$ may lead to instability of the closed-loop due to the fact that the state of the boundary layer (4f) may grow unbounded.

The next corollary shows that if the initial estimation errors $\hat{a}(0) - \bar{a}$, $\hat{b}(0) - \bar{b}$, $\hat{x}(0) - \bar{x}(0)$, and $P(0) - \bar{P}^\star$ are sufficiently small and the probing input d is small and sufficiently rich of order $2n$, then (5) holds.

Corollary 1: Let Assumption 1 hold, suppose that the pair $(A - r^{-1} b b^\top P^\star, C)$ is observable, and consider the interconnection of system (1) and the controller (4). There exist $\Delta^\star > 0$ and $\varepsilon^\star > 0$ such that if $|\hat{a}(0) - \bar{a}| \leq \Delta^\star$, $|\hat{b}(0) - \bar{b}| \leq \Delta^\star$, $|\hat{x}(0) - \bar{x}(0)| < \Delta^\star$, $|P(0) - \bar{P}^\star| < \Delta^\star$, $\varepsilon \in (0, \varepsilon^\star)$, $\|d\|_\infty < \Delta^\star$, and d is sufficiently rich of order $2n$, then (5) holds.

Proof: Letting $\Delta > 0$ be such that the bounds given in item (1) of Theorem 1 hold, by [18, Th. 4.3.2], one has that $|\tilde{\theta}(t)| \leq c_5 \Delta$, for all $t \geq 0$ and some $c_5 > 0$. Therefore, by (9), there exists $\Delta^\square > 0$ such that if $\Delta < \Delta^\square$, then the pair $(\hat{a}(t), \hat{b}(t))$ is strongly controllable and $\|\mu_t\|_{2\delta} \leq (1 + c_6 \Delta + c_6 \|d_t\|_{2\delta}) \mathrm{e}^{c_6 (\gamma + \gamma^2 \beta^{-\frac{1}{2}}) \Delta^2}$, for some $c_6 > 0$, thus implying that $\|\tilde{\theta}\|_\infty$ and $\|\tilde{x}\|_\infty$ can be made arbitrarily small by letting Δ and $\|d\|_\infty$ be sufficiently small. Since the pair (A, b) is controllable and the pair (A, C) is observable, one has that $\bar{A}_{\mathrm{cl}} = \bar{A} + \bar{b} \bar{K}^\star$ is Hurwitz. Further, since the pair $(\bar{A}_{\mathrm{cl}}, \bar{b})$ is controllable and the pair $(\bar{A}_{\mathrm{cl}}, \bar{C})$ is observable, the transfer function $\bar{C}(sI - \bar{A}_{\mathrm{cl}})^{-1} \bar{b}$ has no zero/pole cancellation. Hence, by [18, Th. 5.2.4] and [25, Th. 6.3], if d is sufficiently rich of order $2n$, then the signal ϕ is persistently exciting. ∎

III. Strong Stabilizability via Hybrid Adaptation

The main goal of this section is to modify the controller (4) to remove the assumption about strong controllability of the estimates $(\hat{a}, \hat{b})$ of the system parameters $(\bar{a}, \bar{b})$. This objective is pursued by updating $(\hat{a}, \hat{b})$ only at discrete time instants rather than continuously in time using an adaptation scheme based on the hybrid adaptive Luenberger observer [18, Sec. 5.3.2]. Letting $T_s > 0$ be a sampling time, let $t_k = kT_s$, $k \in \mathbb{N}$. The main goal of this section is to adapt the controller given in Section II to update the parameters just at times t_k enforcing that the pair $(\hat{a}_k, \hat{b}_k)$ is strongly controllable. To pursue this objective consider first the next lemma; for the definition of an exponentially stable in the large equilibrium see [18, Definition 3.4.15].

Lemma 1: Let $\mathcal{A}$ be a compact set such that all $A \in \mathcal{A}$ are Hurwitz, let $A_k \in \mathcal{A}$ be a sequence of matrices, and consider the hybrid system given by

$$\dot{x}(t) = A_k x(t), \tag{10}$$

for all $t \in (t_k, t_{k+1})$, $k \in \mathbb{N}$. If $A_{k+1} - A_k \in \ell_2$, then the origin is exponentially stable in the large for system (10).

Proof: Since the matrix A_k is Hurwitz, there exists a unique, positive definite solution Π_k to the Lyapunov equation $A_k^\top \Pi_k + \Pi_k A_k = -I$. Therefore, it result that $A_{k+1}^\top (\Pi_{k+1} - \Pi_k) + (\Pi_{k+1} - \Pi_k) A_{k+1} = -(A_{k+1} - A_k)^\top \Pi_k - \Pi_k (A_{k+1} - A_k)$, which, together with the fact that $A_k \in \mathcal{A}$, implies that $P_k \in \ell_\infty$ and $|\Pi_{k+1} - \Pi_k| \leq c |A_{k+1} - A_k|$, for some $c > 0$. Therefore, since by assumption $|A_{k+1} - A_k| \in \ell_2$, one has that $|\Pi_{k+1} - \Pi_k| \in \ell_2$. Furthermore, since $A_k \in \mathcal{A}$, there exist $\varkappa_1, \varkappa_2 > 0$ such that $\varkappa_1 I \prec \Pi_k \prec \varkappa_2 I$, for all $k \in \mathbb{N}$. Thus, letting $V = x^\top \Pi_k x$, it results that, for all,

$$\begin{aligned} \dot{V}(t) &\leq -x^\top(t) x(t) \leq -\varkappa_2^{-1} V(t), \quad t \in (t_k, t_{k+1}), \\ V(t_{k+1}^+) &= V(t_{k+1}^-) + x^\top(t_{k+1}^-)(\Pi_{k+1} - \Pi_k) x(t_{k+1}^-) \\ &\leq \left(1 + \varkappa_1^{-1} |\Pi_{k+1} - \Pi_k|\right) V(t_{k+1}^-), \quad k \in \mathbb{N}. \end{aligned}$$

Letting $V_k = V(t_k^+)$, this implies that, for all $k \in \mathbb{N}$, $V_{k+1} \leq (1 + \varkappa_1^{-1} |\Pi_{k+1} - \Pi_k|) \mathrm{e}^{-\varkappa_2^{-1} T_s} V_k$. Thus, since $\prod_{k=0}^{\tau} (1 + a_k) \leq \mathrm{e}^{\sum_{k=0}^{\tau} a_k}$, for $a_i > 0$, $i = 0, \ldots, \tau$, and by the Sedrakyan's inequality [26], one has $\prod_{k=0}^{\tau} (1 + \varkappa_1^{-1} |\Pi_{k+1} - \Pi_k|) \mathrm{e}^{-\varkappa_2^{-1} T_s} = \mathrm{e}^{-\varkappa_2^{-1} \tau T_s} \prod_{k=0}^{\tau} (1 + \varkappa_1^{-1} |\Pi_{k+1} - \Pi_k|) \leq \mathrm{e}^{-\varkappa_2^{-1} \tau T_s + \varkappa_1^{-1} \sum_{k=0}^{\tau} |\Pi_{k+1} - \Pi_k|} \leq \mathrm{e}^{-\varkappa_2^{-1} \tau T_s + \varkappa_1^{-1} \sqrt{\tau} \sqrt{\sum_{k=0}^{\tau} |\Pi_{k+1} - \Pi_k|^2}}$. Therefore, since $|\Pi_{k+1} - \Pi_k| \in \ell_2$, one has that $\lim_{k \to \infty} V_k = 0$, thus implying that x tends to 0 exponentially fast. ∎

Taking advantage of Lemma 1, consider the hybrid output feedback controller with continuous-time dynamics

$$\phi_1 = \frac{\alpha_{n-1}(s)}{\Lambda^\star(s)} u, \quad \phi_2 = -\frac{\alpha_{n-1}(s)}{\Lambda^\star(s)} y, \quad z = \frac{s^n}{\Lambda^\star(s)} y, \tag{11a}$$

$$\dot{\hat{x}} = \hat{A}_k \hat{x} + \hat{b}_k u + (a^* - \hat{a}_k)(y - \bar{C} \hat{x}), \tag{11b}$$

$$u = -r^{-1} \hat{b}_k \hat{P}_k \hat{x} + d, \tag{11c}$$

for all $t \in (t_k, t_{k+1})$, where d is a probing input, whose parameters are updated using the discrete-time dynamics

$$\hat{a}_{k+1} = \hat{a}_k + \gamma \int_{t_k}^{t_{k+1}} \frac{z(\tau) - \hat{\theta}_k^\top \phi(\tau)}{1 + \beta \phi^\top(\tau) \phi(\tau)} \phi_1(\tau) \mathrm{d}\tau, \tag{11d}$$

$$\hat{b}_{k+1} = \hat{b}_k + \gamma \int_{t_k}^{t_{k+1}} \frac{z(\tau) - \hat{\theta}_k^\top \phi(\tau)}{1 + \beta \phi^\top(\tau) \phi(\tau)} \phi_2(\tau) \mathrm{d}\tau, \tag{11e}$$

where $\gamma > 0$ is the *adaptation gain*, $\phi = \mathrm{col}(\phi_1, \phi_2)$,

$$\hat{A}_k = \left[-\hat{a}_k \middle| \begin{matrix} I \\ 0 \end{matrix} \right], \tag{11f}$$

$\hat{\theta}_k = \mathrm{col}(\hat{b}_k, \hat{a}_k)$, and $\hat{P}_k$ is the solution to the ARE

$$\hat{A}_k^\top P_k + P_k \hat{A}_k + \bar{C}^\top C - r^{-1} P_k \hat{b}_k \hat{b}_k^\top P_k = 0. \tag{11g}$$

The next theorem shows that if the estimates $(\hat{a}_k, \hat{b}_k)$ obtained using the hybrid output feedback controller (11) are strongly controllable, then it solves Problem 1.

Theorem 2: Let Assumption 1 hold and consider the interconnection of (1) and (11). If $|\rho(\hat{a}_k, \hat{b}_k)| > \chi$ for all $k \in \mathbb{N}$ and some $\chi > 0$, and $\gamma T_s < 2$, then

1) all the signals in the closed loop are in $\mathcal{L}_\infty$;

If, additionally, the signal ϕ is persistently exciting, then

2) the estimates $\hat{a}_k$, $\hat{b}_k$, $\hat{x}$, and $\hat{P}_k$ are such that

$$\lim_{k\to\infty} \hat{a}_k - \bar{a} = 0, \quad \lim_{k\to\infty} \hat{b}_k - \bar{b} = 0, \tag{12a}$$

$$\lim_{t\to\infty} \hat{x}(t) - \bar{x}(t) = 0, \quad \lim_{k\to\infty} \hat{P}_k - \bar{P}^\star = 0. \tag{12b}$$

Proof: The dynamics (11d) and (11e) implement the hybrid adaptive law given in [18, Th. 4.6.1]. Thus, if $\gamma T_s < 2$, then $\hat{\theta}_k \in \ell_\infty$, $|\hat{\theta}_{k+1} - \hat{\theta}_k| \in \ell_2 \cap \ell_\infty$, and $\epsilon, \epsilon m \in \mathcal{L}_2 \cap \mathcal{L}_\infty$.

Since $|\rho(\hat{a}_k, \hat{b}_k)| \geq \chi$ for all $k \in \mathbb{N}$, there exists an unique solution $\hat{P}_k$ to the ARE (11g), which is such that the matrix $\hat{A}_{\mathrm{cl},k} = \hat{A}_k - r^{-1}\hat{b}_k\hat{b}_k^\top \hat{P}_k$ is Hurwitz. Since $\hat{a}_k, \hat{b}_k \in \ell_\infty$ and $|\rho(\hat{a}_k, \hat{b}_k)| \geq \chi$ for all $k \in \mathbb{N}$, it results that $\hat{P}_k \in \mathcal{L}_\infty$. Additionally, since $\hat{P}_k$ and $\hat{P}_{k+1}$ solve (11g), one has $|\hat{P}_{k+1} - \hat{P}_k|, |\hat{A}_{\mathrm{cl},k+1} - \hat{A}_{\mathrm{cl},k}| \in \ell_2 \cap \ell_\infty$.

Consider the dynamics of $\hat{x}$, which are given by

$$\dot{\hat{x}} = \hat{A}_{\mathrm{cl},k}\hat{x} + \hat{b}_k d + (a^\star - \hat{a}_k)\tilde{y}, \tag{13}$$

for all $t \in (t_k, t_{k+1})$. Thus, by Lemma 1 and [18, Lemma 3.3.3], since $d, \hat{a}_k, \hat{b}_k \in \mathcal{L}_\infty$, one has $\|\hat{x}\|_{2\delta} \leq c + c\|\tilde{y}\|_{2\delta}$, for some $c > 0$. Defining the identification errors $\tilde{a}_k = \bar{a} - \hat{a}_k$ and $\tilde{b}_k = \bar{b} - \hat{b}_k$, the dynamics of $\tilde{x} = \bar{x} - \hat{x}$ are given by $\dot{\tilde{x}} = A^*\tilde{x} - \tilde{a}_k y + \tilde{b}_k u$. By [18, Lemma 5.3.1], there exist continuously differentiable signals $\tilde{\vartheta}_a, \tilde{\vartheta}_b$ such that $|\tilde{\vartheta}_a - \tilde{a}_k|, |\tilde{\vartheta}_b - \tilde{b}_k| \in \mathcal{L}_2$ and $\dot{\tilde{\vartheta}}_a, \dot{\tilde{\vartheta}}_b \in \mathcal{L}_2$. Therefore, by [18, Lemma A.1], one has $\tilde{y} = \tilde{\vartheta}^\top\phi + \sum_{i=1}^n W_{ci}(s)((W_{bi}(s)u)\,\dot{\tilde{\vartheta}}_{bi} - (W_{bi}(s)y)\,\dot{\tilde{\vartheta}}_{ai}) + \bar{C}(sI - A^\star)^{-1}e$, where W_{bi} and W_{ci} are defined in (8), $i = 1, \ldots, n$, and $e = (\tilde{b}_k - \tilde{\vartheta}_b)u - (\tilde{a}_k - \tilde{\vartheta}_a)y$. Therefore, letting μ be defined as in the proof of Theorem 1 and defining $\tilde{\vartheta} = \mathrm{col}(\tilde{\vartheta}_b, \tilde{\vartheta}_a)$, by [18, Lemma 3.3.3], one has that $\|\tilde{y}\|_{2\delta} \leq c + c\|\epsilon m\mu\|_{2\delta} + c\||\tilde{\vartheta} - \tilde{\theta}_k|\mu\|_{2\delta} + c\||\dot{\tilde{\vartheta}}|\mu\|_{2\delta}$, for some $c \geq 0$. Hence, considering that u is given by (11c) and that $y = \tilde{y} + \bar{C}\hat{x}$, since $\epsilon m, |\tilde{\vartheta} - \tilde{\theta}_k|, |\dot{\tilde{\vartheta}}| \in \mathcal{L}_2$ by the Bellman-Gronwall lemma, one has that $\mu, \phi \in \mathcal{L}_\infty$. Thus, following the same construction already used in the proof of Theorem 1, it can be derived that $x, \hat{x}, \tilde{x}, u, y \in \mathcal{L}_\infty$.

The proof of item (2) follows by reasoning similar to that used to prove the corresponding item in Theorem 1. ■

As stated in Theorem 2, the hybrid controller (11) still requires that the pair $(\hat{a}_k, \hat{b}_k)$ is strongly controllable in order to ensure that the signals in the closed-loop are bounded. However, such a controller can be easily adapted to overcome this requirement, provided that ϕ is persistently exciting. Namely, given $\chi > 0$, consider the modified controller

$$\phi_1 = \frac{\alpha_{n-1}(s)}{\Lambda^\star(s)}u, \quad \phi_2 = -\frac{\alpha_{n-1}(s)}{\Lambda^\star(s)}y, \quad z = \frac{s^n}{\Lambda^\star(s)}y, \tag{14a}$$

$$\dot{\hat{x}} = \check{A}_k\hat{x} + \check{b}_k u + \left(a^* - \check{a}_k\right)\left(y - \bar{C}\hat{x}\right), \tag{14b}$$

$$u = -r^{-1}\check{b}_k\check{P}_k\hat{x} + d, \tag{14c}$$

for all $t \in (t_k, t_{k+1})$, where d is a probing input, whose parameters are updated using the discrete-time dynamics

$$\hat{a}_{k+1} = \hat{a}_k + \gamma\int_{t_k}^{t_{k+1}} \frac{z(\tau) - \hat{\theta}_k^\top\phi(\tau)}{1 + \beta\phi^\top(\tau)\phi(\tau)}\phi_1(\tau)\mathrm{d}\tau, \tag{14d}$$

$$\hat{b}_{k+1} = \hat{b}_k + \gamma\int_{t_k}^{t_{k+1}} \frac{z(\tau) - \hat{\theta}_k^\top\phi(\tau)}{1 + \beta\phi^\top(\tau)\phi(\tau)}\phi_2(\tau)\mathrm{d}\tau, \tag{14e}$$

where $\gamma > 0$ is the *adaptation gain*, $\phi = \mathrm{col}(\phi_1, \phi_2)$, and

$$\check{A}_{k+1} = \begin{cases} \left[-\hat{a}_{k+1}\,\middle|\, \begin{smallmatrix} I \\ 0 \end{smallmatrix}\right], & \text{if } |\rho\left(\hat{a}_{k+1}, \hat{b}_{k+1}\right)| \geq \chi, \\ \check{A}_k, & \text{otherwise}, \end{cases} \tag{14f}$$

$$\check{b}_{k+1} = \begin{cases} \hat{b}_{k+1} & \text{if } |\rho\left(\hat{a}_{k+1}, \hat{b}_{k+1}\right)| \geq \chi, \\ \check{b}_k, & \text{otherwise}, \end{cases} \tag{14g}$$

and $\check{P}_k$ is the solution to the following ARE

$$\check{A}_k^\top P_k + P_k\check{A}_k + \bar{C}^\top C - r^{-1}P_k\check{b}_k\check{b}_k^\top P_k = 0. \tag{14h}$$

The following corollary shows that the output feedback, hybrid controller (14) solves Problem 1.

Corollary 2: Let Assumption 1 hold and consider the interconnection of (1) and (14). If $|\rho(\bar{a}, \bar{b})| > \chi$, $|\rho(\hat{a}(0), \hat{b}(0))| > \chi$, $\gamma T_s < 2$, and ϕ is persistently exciting, then items (1) and (2) of Theorem 2 hold.

Proof: Since the dynamics (14d) and (14e) implement the hybrid adaptive law given in [18, Th. 4.6.1], if ϕ is persistently exciting, then (12) holds. Therefore, if $|\rho(\bar{a}, \bar{b})| > \chi$, then there exists $K > 0$ such that $|\rho(\hat{a}_k, \hat{b}_k)| > \chi$ for all $k \geq K$, and hence $\check{A}_k = \hat{A}_k$ and $\check{b}_k = \hat{b}_k$ for all $k \geq K$. Hence, the proof follows directly by the construction employed to prove Theorem 2. ■

Remark 1: Algorithm (14) requires the solution to the ARE (14h) at the update times t_k. Therefore, it has a slightly increased computational complexity with respect to the scheme (4) wherein such a task is delegated to the boundary layer (4f). Nonetheless, the hybrid updates allow to overcome the requirement about strong controllability of the pair $(\hat{a}, \hat{b})$. Comparing such a scheme with the one given in [20, eq. (7)], note that it does not introduce time delays in the closed loop, thus reducing the memory burden.

IV. Numerical Simulations

Consider system (1) and the cost (2) with

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \end{bmatrix},$$

and $r = 10$, which is the same system considered in [10, Sec. IV-A]. Numerical simulations have been carried out to test the effectiveness of the hybrid controller (14) both in the presence and in the absence of the probing input d. Namely, letting $\gamma = 0.2$, $T_s = 0.5$, $\beta = 0.01$, $a^\star = [1.412 \;\; 1]^\top$, $\chi = 1$, and initializing the states of system (1) and of the controller (4) at random, a first simulation has been carried out letting $d = 0$. A second simulation has been carried out with the same parameters of above, but letting d be a band limited withe noise signal with power 5 and sampling time 0.1. Figure 2 depicts the results of such simulations.

As shown by Figure 2, the proposed adaptive output feedback controller guarantees that all the signals in the closed-loop are bounded even if the probing input does not sufficiently excite the plant. On the other hand, if the probing input guarantees that the signal ϕ is persistently exciting, then the proposed adaptive output feedback controller asymptotically estimates the state of the system in the $\bar{x}$-coordinates, its parameters, the optimal control, and the value function of the considered LQ problem.

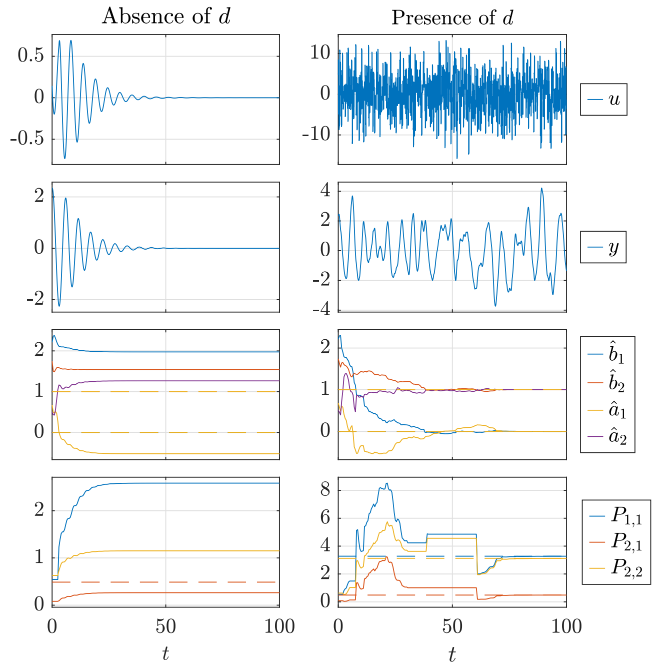

Fig. 2. Results of the numerical simulations. The dashed lines represent the true value of the parameters to be estimated.

Comparing the outcome of these simulations with those reported in [10, Sec. IV-A], note that the convergence rate of the proposed scheme is slower than the one given in [10, Sec. IV-A]. However, the proposed adaptive scheme does not require an initial stabilizing feedback. Further, it does not exploit time delays in the closed loop since the state of the system is reconstructed using a hybrid adaptive Luenberger observer. Finally, even if the probing input is absent, the proposed scheme guarantees boundedness of all the signals in the closed-loop although, in this case, convergence to the optimal feedback gain and value function is not guaranteed.

V. Concluding Remarks

An adaptive output feedback controller has been proposed to solve LQ optimal control problems for SISO systems by using just input/output measurements. Numerical simulations confirm that the state of the closed-loop is uniformly bounded even in the case that the regressor is not persistently exciting. If, additionally, such a signal is persistently exciting, then it further estimates the parameters, the optimal control, and the optimal value function.

References

[1] P. Dorato, V. Cerone, and C. Abdallah, *Linear-Quadratic Control: An Introduction*. Malabar, FL, USA: Krieger Publ., 2000.
[2] D. Vrabie, O. Pastravanu, M. Abu-Khalaf, and F. L. Lewis, "Adaptive optimal control for continuous-time linear systems based on policy iteration," *Automatica*, vol. 45, no. 2, pp. 477–484, 2009.
[3] 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.
[4] C. Possieri and M. Sassano, "Q-learning for continuous-time linear systems: A data-driven implementation of the Kleinman algorithm," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 10, pp. 6487–6497, Oct. 2022.
[5] 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.
[6] K. G. Vamvoudakis, "Q-learning for continuous-time linear systems: A model-free infinite horizon optimal control approach," *Syst. Control Lett.*, vol. 100, pp. 14–20, Feb. 2017.
[7] S. K. Jha, S. B. Roy, and S. Bhasin, "Direct adaptive optimal control for uncertain continuous-time LTI systems without persistence of excitation," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 65, no. 12, pp. 1993–1997, Dec. 2018.
[8] C. De Persis and P. Tesi, "Formulas for data-driven control: Stabilization, optimality, and robustness," *IEEE Trans. Autom. Control*, vol. 65, no. 3, pp. 909–924, Mar. 2020.
[9] G. Baggio, V. Katewa, and F. Pasqualetti, "Data-driven minimum-energy controls for linear systems," *IEEE Control Syst. Lett.*, vol. 3, no. 3, pp. 589–594, Jul. 2019.
[10] H. Modares, F. L. Lewis, and Z.-P. Jiang, "Optimal output-feedback control of unknown continuous-time linear systems using off-policy reinforcement learning," *IEEE Trans. Cybern.*, vol. 46, no. 11, pp. 2401–2410, Nov. 2016.
[11] K. G. Vamvoudakis and N.-M. T. Kokolakis, "Synchronous reinforcement learning-based control for cognitive autonomy," *Found. Trends®Syst. Control*, vol. 8, nos. 1–2, pp. 1–175, 2020.
[12] Z.-P. Jiang, T. Bian, and W. Gao, "Learning-based control: A tutorial and some recent results," *Found. Trends®Syst. Control*, vol. 8, no. 3, pp. 176–284, 2020.
[13] L. Falconi, A. Martinelli, and J. Lygeros, "Data-driven optimal control via linear programming: Boundedness guarantees," 2023, *arXiv:2310.19563*.
[14] A. Martinelli, M. Gargiani, M. Draskovic, and J. Lygeros, "Data-driven optimal control of affine systems: A linear programming perspective," *IEEE Control Syst. Lett.*, vol. 6, pp. 3092–3097, 2022.
[15] G. C. Calafiore, S. Gaubert, and C. Possieri, "Log-sum-exp neural networks and posynomial models for convex and log-log-convex data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 3, pp. 827–838, Mar. 2020.
[16] S. Brüggemann and C. Possieri, "On the use of difference of log-sum-exp neural networks to solve data-driven model predictive control tracking problems," *IEEE Control Syst. Lett.*, vol. 5, no. 4, pp. 1267–1272, Oct. 2021.
[17] L. Menini, C. Possieri, and A. Tornambe, "A 'practical' observer for nonlinear systems," in *Proc. 56th Annu. Conf. Decision Control*, 2017, pp. 3015–3020.
[18] P. A. Ioannou and J. Sun, *Robust adaptive control*. Mineola, NY, USA: Dover Publ., 2012.
[19] C. Possieri and M. Sassano, "Value iteration for continuous-time linear time-invariant systems," *IEEE Trans. Autom. Control*, vol. 68, no. 5, pp. 3070–3077, May 2023.
[20] C. Possieri, "Value iteration via output feedback for LQ optimal control of SISO systems," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 11861–11866, 2023. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85183652245&doi=10.1016
[21] D. Cox, J. Little, and D. O'Shea, *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*, 4th ed. New York, NY, USA: Springer, 2013.
[22] B. D. Anderson and J. B. Moore, *Optimal Control: Linear Quadratic Methods*. Hoboken, NJ, USA: Prentice Hall, 1990.
[23] P. J. Antsaklis and A. N. Michel, *Linear Systems*. Boston, MA, USA: Birkhäuser, 2006.
[24] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Englewood Cliffs, NJ, USA: Prentice Hall, 2002.
[25] I. M. Mareels and M. Gevers, "Persistency of excitation criteria for linear, multivariable, time-varying systems," *Math. Control, Signals Syst.*, vol. 1, no. 3, pp. 203–226, 1988.
[26] H. Sedrakyan and N. Sedrakyan, "A useful inequality," in *Algebraic Inequalities*. Cham, Switzerland: Springer Int., 2018, pp. 107–125.

Open Access funding provided by "Università degli Studi di Roma 'Tor Vergata"' within the CRUI CARE Agreement