This article has been accepted for publication in IEEE Control Systems Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2024.3382439

Policy Algebraic Equation for the LQR and the $\mathscr{H}_\infty$ control problems

Mario Sassano, *Senior Member, IEEE*

***Abstract*— The Linear Quadratic Regulator (LQR) and the $\mathscr{H}_\infty$ control problems for linear systems are revisited with the objective of deriving a novel algebraic (polynomial) equation alternative to the standard Algebraic Riccati Equation (ARE). Differently from the latter, the former is envisioned to involve the *policy alone*, in place of the *value function* as in the ARE. The resulting equation, referred to as the *Policy Algebraic Equation*, contains nm variables and equations, of order less than or equal to $2n$, where n and m denote the dimension of the state and the input, respectively.**

***Index Terms*— Optimal control, Linear systems, Robust control**

I. Introduction

IT is not surprising that, since their origin in the middle of the previous century, the LQR and the $\mathscr{H}_\infty$ control problems have been among the most studied topics in control theory (see e.g. [1], [2], [3]). Such a well-deserved attention stems from their importance in practice, as such techniques enable one to design control strategies that ensure particularly desirable properties to the resulting closed-loop plant. These encompass optimality with respect to prescribed criteria as well as robustness with respect to exogenous disturbances. Interestingly, their practical relevance is matched, if not even overshadowed, by the theoretical challenges originating from the abstract characterization of the underlying solution, which has proved to be an intriguing and interesting mathematical problem, see e.g. [4], [5], [6], [7]. Most of the existing approaches aim at characterizing *first* the value function, namely the cost of the optimal trajectory from a certain initial condition, and *subsequently* compute the policy that attains such a performance on the basis of the knowledge of its cost. Policies have recently regained a central role within the context of Reinforcement Learning (RL) [8], specifically in the framework of direct policy optimization.

The main contribution of this paper is twofold. First, the optimal policy and value function are simultaneously related by means of the observability matrix of the (state/costate) Hamiltonian system via measurements of the state alone. Similar constructions are discussed in [7], [9]. Differently from the latter, herein the conditions are extended to the setting of *matched* disturbance attenuation. Second, such an abstract property, which is interesting *per se* as it remains linear in the value function, is instrumental for deriving an algebraic equation that revolves around the policy alone. This equation, referred to as the Policy Algebraic Equation (PAE), permits the direct computation of the policy without any knowledge whatsoever of the corresponding cost, as it has been instead pursued hitherto in the literature involving dynamic optimization problems. The PAE may possess appealing computational features, since its dimension grows only linearly with the dimension of the state, whereas the standard ARE grows quadratically. The PAE contains nm equations of order $2n$, hence typically involving fewer unknown variables than the corresponding ARE. Moreover, while the ARE requires the underlying variable to be symmetric and positive definite, the PAE is such that its solution should then enforce asymptotic stability to the closed-loop system. While the latter is indeed a challenging constraint, the property may be verified *a posteriori* on the set of solutions to PAE (see Example 4). More importantly, the possibility of computing (or estimating and manipulating) the policy without resorting to the solution of any standard condition requiring the knowledge of the cost may be particularly valuable for strategies that aim at optimizing directly in the *policy space* rather than in the space of value functions. These methods have recently acquired a central role in the framework of RL.

The rest of the manuscript is organized as follows. First the (slightly more convoluted) $\mathscr{H}_\infty$ control problem is discussed in Section II. This choice allows to then immediately specialize similar claims to the LQR (sub-)case, which is tackled in Section V. The main objective of Section III consists in establishing an identity satisfied by the optimal actor/critic pair, i.e. policy and value function, respectively, which are simultaneously related via the observability matrix of the Hamiltonian dynamics. The latter identity is then shown in Section IV to be instrumental for deriving the Policy Algebraic Equation in terms of the policy alone.

II. Problem statement and preliminaries

Consider a perturbed LTI system described by

$$\begin{cases} \dot{x}(t) = Ax(t) + B_1 w(t) + B_2 u(t)\,, \quad x(0) = x_0 \\ y_p(t) = Cx(t)\,, \end{cases} \tag{1}$$

for $t \in \mathbb{R}_{\geqslant 0}$, with $x : \mathbb{R} \to \mathbb{R}^n$, $y_p : \mathbb{R} \to \mathbb{R}^q$, $w : \mathbb{R} \to \mathbb{R}^p$ and $u : \mathbb{R} \to \mathbb{R}^m$ denoting the state, (performance) output, (exogenous) disturbance and (controlled) input, respectively. To avoid trivialities, the matrices B_i, $i = 1, 2$ have full column rank. Moreover, $w \in \mathscr{L}_2(\mathbb{R}_{\geqslant 0})$, where the latter denotes the

M. Sassano is with the Dipartimento di Ingegneria Civile e Ingegneria Informatica, Università di Roma "Tor Vergata", Via del Politecnico, 1 00133 Roma, Italy (Email: mario.sassano@uniroma2.it).

space of functions $w : \mathbb{R}_{\geqslant 0} \to \mathbb{R}^p$ with the property that $\left(\int_0^\infty \|w(\tau)\|^2 d\tau\right) < \infty$ (see [10]).

Definition 1: Fix $\gamma \in \mathbb{R}_{>0}$. The (sub-optimal) $\mathcal{H}_\infty$ control problem consists in determining, if it exists, a linear feedback control law $u = K_2^\star x$ such that the $\mathcal{L}_2$-gain[1] of the closed-loop system from w to $\mathrm{col}(y_p, u)$ is smaller than γ. ◦

The classic solution to the $\mathcal{H}_\infty$ control problem revolves around the Algebraic Riccati Equation (see e.g. [2], [11])

$$0 = C^\top C + A^\top P + PA + P(\gamma^{-2} B_1 B_1^\top - B_2 B_2^\top)P \quad (2)$$

with respect to the unknown variable $P = P^\top \in \mathbb{R}^{n\times n}$. Furthermore, let the matrix $H \in \mathbb{R}^{2n\times 2n}$, defined as

$$H = \begin{bmatrix} A & \gamma^{-2} B_1 B_1^\top - B_2 B_2^\top \\ -C^\top C & -A^\top \end{bmatrix}, \quad (3)$$

denote the Hamiltonian matrix naturally associated with (2). As it has been elegantly established the solvability of the $\mathcal{H}_\infty$ control problem hinges upon certain properties of the Hamiltonian matrix (3). More precisely, H is said to belong to $\mathrm{dom}(\mathrm{Ric})$ (see [2, Ch. 13.2]) if (i) it does not possess eigenvalues on the imaginary axis and (ii) the subspaces $\mathcal{V}^-(H)$ and $\mathrm{im}\begin{bmatrix}0 & I\end{bmatrix}^\top$ are complementary, where $\mathcal{V}^-(H) \subset \mathbb{R}^{2n}$ denotes the (n-dimensional) stable invariant subspace of H. If $H \in \mathrm{dom}(\mathrm{Ric})$, then the matrix $P \triangleq P_2 P_1^{-1}$, with P_i such that $\mathrm{im}[P_1^\top,\ P_2^\top]^\top = \mathcal{V}^-(H)$, is uniquely determined from H, hence $H \mapsto P$ defines a function $\mathrm{Ric} : \mathrm{dom}(\mathrm{Ric}) \to \mathbb{R}^{n\times n}$.

Assumption 1: The pairs (A, B_2) and (A, C) are reachable and observable, respectively. The Hamiltonian matrix H in (3) belongs to $\mathrm{dom}(\mathrm{Ric})$ and $\mathrm{Ric}(H) \succeq 0$. ◦

Thus, provided Assumption 1 is satisfied, the ARE (2) admits a unique positive semi-definite solution, denoted by $P^\star := \mathrm{Ric}(H)$. Furthermore, the feedback policy solving the $\mathcal{H}_\infty$ control problem is obtained as

$$u^\star = -B_2^\top P^\star x =: K_2^\star x\,, \quad (4)$$

whereas the *worst-case disturbance*, as a function of the current value of the state, is defined (as a by-product) by

$$w^\star = \gamma^{-2} B_1^\top P^\star x =: K_1^\star x \quad (5)$$

provided that $\sigma(A + B_1 K_1^\star + B_2 K_2^\star) \subset \mathbb{C}^-$. A further ingredient towards the characterization of the $\mathcal{H}_\infty$ control problem is represented by the Hamiltonian dynamics associated with (3). Letting $\lambda : \mathbb{R}_{\geqslant 0} \to \mathbb{R}^n$ denote the *costate*, the time history of the underlying solution may be equivalently obtained by selecting $u^\star(t) = -B_2^\top \lambda^\star(t)$ and, similarly, $w^\star(t) = \gamma^{-2} B_1^\top \lambda^\star(t)$, for all $t \geqslant 0$, where $\lambda^\star$ describes the solution to the Hamiltonian dynamics

$$\begin{cases} \dot{x}(t) = Ax(t) + (\gamma^{-2} B_1 B_1^\top - B_2 B_2^\top)\lambda(t) \\ \dot{\lambda}(t) = -C^\top C x(t) - A^\top \lambda(t) \end{cases} \quad (6)$$

initialized according to $\mathrm{col}(x(0), \lambda(0)) = \mathrm{col}(x_0, P^\star x_0)$. Finally, as it is instrumental for providing concise statements of the subsequent results, some terminology is borrowed, and slightly extended, from the theory of RL (see e.g. [8]). To this end, the matrices $(K_1^\star, K_2^\star, P^\star)$ are referred to as the *actors/critic triplet*. The latter (compact) notation permits to denote the policies adopted by the *decision makers* involved in the underlying dynamic optimization problem together with the cost of the resulting trajectory. The former (cumulatively referred to as the *actors*) are in charge of selecting the worst disturbance and the corresponding best response, respectively, while the associated cost is encoded into the matrix $P^\star$ (namely the *critic*). In fact, the identity

$$\frac{1}{2}\|x_0\|^2_{P^\star} = \int_0^\infty (\|Cx(t)\|^2 + \|u^\star(t)\|^2 - \gamma^2 \|w^\star(t)\|^2) dt, \quad (7)$$

holds for all $x_0 \in \mathbb{R}^n$.

[1] Consider $\dot{x} = Ax + Bu$, $y = Cx$, with $\sigma(A) \subset \mathbb{C}^-$. The $\mathcal{L}_2$-gain from u to y is defined as $\sup_{u(\cdot)\in\mathcal{L}_2(\mathbb{R}_{\geqslant 0}), u(\cdot)\neq 0} \|y\|_2/\|u\|_2$, where $\|v\|_2^2 := \left(\int_0^\infty \|v(\tau)\|^2 d\tau\right)$ denotes the norm on $\mathcal{L}_2(\mathbb{R}_{\geqslant 0})$, see e.g. [10].

III. Solution via observability matrix

The aim of this section is to provide an alternative characterization of the solution to the $\mathcal{H}_\infty$ control problem that relies upon certain observability properties of the Hamiltonian dynamics, rather than on the standard ARE (2). Towards this end, a relevant role is played by the (virtual) output $\boldsymbol{y} := \Pi \mathrm{col}(x, \lambda) = x$, which is obtained by letting $\Pi = \begin{bmatrix} I_n & \mathbf{0}_{n\times n}\end{bmatrix}$, associated with the Hamiltonian system (6). First, a technical lemma is stated and proved.

Lemma 1: Suppose that Assumption 1 holds. Then the pair (H, Π) is observable. ◦

Proof: The observability properties of the pair (H, Π) are equivalent to those of the pair $(H + G\Pi, \Pi)$ for any matrix $G \in \mathbb{R}^{2n\times n}$. Thus, selecting $G = \begin{bmatrix}0 & C^\top C\end{bmatrix}^\top$ yields

$$H + G\Pi = \begin{bmatrix} A & \gamma^{-2} B_1 B_1^\top - B_2 B_2^\top \\ 0 & -A^\top \end{bmatrix}. \quad (8)$$

Furthermore, observability of (8) via Π can be studied by interpreting the former as the cascade interconnection of two subsystems described by the triplets $\Sigma_1 := (-A^\top, 0, \gamma^{-2} B_1 B_1^\top - B_2 B_2^\top)$ and $\Sigma_2 := (A, I_n, I_n)$. Since Σ_2 is (trivially) observable via the identity matrix, standard arguments concerning structural properties of interconnected systems ensure that the cascade of Σ_1 and Σ_2 is observable provided the pair $(-A^\top, \gamma^{-2} B_1 B_1^\top - B_2 B_2^\top)$ is observable or, via duality, that the pair $(A, \gamma^{-2} B_1 B_1^\top - B_2 B_2^\top)$ is reachable, hence proving the claim since the latter is implied by reachability of (A, B_2) and the rank properties of B_i. ■

Lemma 1 entails that the Hamiltonian dynamics (6) is observable via $\boldsymbol{y}$, namely whenever the state variable alone is measured. To provide a concise statement of the following result, let $\mathcal{O}(H, \Pi) \in \mathbb{R}^{2n^2\times 2n}$ denote the observability matrix associated to the pair (H, Π), namely

$$\mathcal{O}(H, \Pi) = \begin{bmatrix} \Pi^\top & (\Pi H)^\top & \dots & (\Pi H^{2n-1})^\top \end{bmatrix}^\top. \quad (9)$$

Let the matrix-valued operator $\mathcal{S} : \mathbb{R}^{p\times n} \times \mathbb{R}^{m\times n} \to \mathbb{R}^{2n^2\times n}$

This article has been accepted for publication in IEEE Control Systems Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2024.3382439

be defined according to

$$(K_1, K_2) \mapsto \mathcal{S}(K_1, K_2) := \begin{bmatrix} A + B_1K_1 + B_2K_2 \\ (A + B_1K_1 + B_2K_2)^2 \\ \vdots \\ (A + B_1K_1 + B_2K_2)^{2n} \end{bmatrix}. \tag{10}$$

Theorem 1: Fix $\gamma \in \mathbb{R}_{>0}$ and consider the $\mathscr{H}_\infty$ control problem for system (1). Suppose that Assumption 1 holds. The matrices $(K_1^\star, K_2^\star, P^\star)$, $P^\star = (P^\star)^\top \succeq 0$, constitute the actors/critic triplet if and only if

(i) $\sigma(A + B_1K_1^\star + B_2K_2^\star) \subset \mathbb{C}^-$;
(ii) $K_1^\star - \gamma^{-2}B_1^\top P^\star = 0$;
(iii) $K_2^\star + B_2^\top P^\star = 0$;
(iv) the algebraic equation

$$0 = \mathcal{O}(H, \Pi H) \begin{bmatrix} I \\ P^\star \end{bmatrix} - \mathcal{S}(K_1^\star, K_2^\star) \tag{11}$$

holds. ∘

Proof: The claim is proved by borrowing arguments similar to those employed in the proof of [7, Thm. 4] and of [9, Thm. 1] and [9, Thm. 3]. Towards this end, uniqueness of the actors/critic triplet, together with the property that the latter may be characterized via (2), (4), (5) as well as, equivalently, in terms of a certain trajectory of the Hamiltonian dynamics (6), ensure that the matrices $(K_1^\star, K_2^\star, P^\star)$ satisfy

$$t \mapsto \Pi e^{Ht} \begin{bmatrix} I \\ P^\star \end{bmatrix} - e^{(A+B_1K_1^\star+B_2K_2^\star)t} \equiv 0\,, \tag{12}$$

(see [7, Eq. (34)]). Since (12) is analytic, this implies, by Cayley-Hamilton Theorem, necessity of item (iv), whereas necessity of items (i)-(iii) follows by the properties of the underlying solution to the $\mathscr{H}_\infty$ control problem. Sufficiency of items (i)-(iv) is obtained by relying on constructions similar to those in the proof of [9, Thm. 1], although with the (full column) rank condition on $\begin{bmatrix} B_1 & B_2 \end{bmatrix}$ of [9, Thm. 1] replaced by reachability of the pair (A, B_2), and hence of $(A, \gamma^{-2}B_1B_1^\top - B_2B_2^\top)$, to ensure observability of the Hamiltonian dynamics, as guaranteed by Lemma 1. Finally, by observing that $\mathcal{O}(H, \Pi H) = \mathcal{O}(H, \Pi)H$, equivalence of *actors* $K_1^\star$, $K_2^\star$ satisfying items (i) and (iv) with those defined in (4), (5) is implied by items (ii) and (iii), along ideas inspired by item (iv) in [9, Thm. 3]. ■

The aim of the following three remarks consists in discussing the consequences of violating a few of the previous assumptions required for the applicability of Theorem 1.

Remark 1: Items (i)-(iii) individually possess an obvious counterpart in the set of classical conditions provided in Section II. Therefore, the conclusions of Theorem 1 reveal that (11) essentially replaces the ARE (2) by providing equivalent conditions, although by involving *simultaneously* the matrices P and K_i, $i = 1, 2$. These are jointly related via the observability matrix of the Hamiltonian dynamics (6). As a consequence, differently from (2), the algebraic equations (11) remain *linear* in the variable P – on which further constraints are imposed – although polynomial in the entries of the actors K_i, $i = 1, 2$. Furthermore, the identity (12) suggests an insightful interpretation of the actors/critic triplet that is revealed when Dynamic Programming and Pontryagin's Principle are combined: such a triplet is characterized by the property that the trajectories of the closed-loop system (1) are *immersed* into stable (output) trajectories of the Hamiltonian dynamics (6) for any initial condition. ▲

Remark 2: Within the framework of $\mathscr{H}_\infty$ control, the disturbance is said to be *unmatched* whenever it affects the state of the plant (1) via input directions that are linearly independent from those along which the controlled input may influence the state, namely whenever $\operatorname{rank}\left(\begin{bmatrix} B_1 & B_2 \end{bmatrix}\right) = p + m$. In such a class of problems, the latter rank condition implies items (ii) and (iii) of Theorem 1, provided item (iv) holds. In fact, the first block equation appearing in (11) yields

$$\begin{bmatrix} B_1 & B_2 \end{bmatrix} \begin{bmatrix} \gamma^{-2}B_1^\top \\ -B_2^\top \end{bmatrix} P = \begin{bmatrix} B_1 & B_2 \end{bmatrix} \begin{bmatrix} K_1 \\ K_2 \end{bmatrix}, \tag{13}$$

so that items (ii) and (iii) can be immediately obtained by relying on the full column-rank property of the input matrices. As a side effect, inspecting (13) permits visualizing the consequence of enforcing items (i) and (iv) (hence including (13)), although without items (ii) and (iii), when the disturbance is *matched* by the controlled input. In fact, the *stability* (i) and the *immersion* (iv) conditions alone identify an affine set of solutions described by $K_i = K_i^\star + W_i\Gamma$, $i = 1, 2$, for arbitrary $\Gamma \in \mathbb{R}^{\nu\times\nu}$ and where the matrices $W_1 \in \mathbb{R}^{p\times\nu}$ and $W_2 \in \mathbb{R}^{m\times\nu}$ are defined by the property

$$\operatorname{im} \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} = \begin{bmatrix} B_1 & B_2 \end{bmatrix}^\perp, \tag{14}$$

with $\nu = m + p - \operatorname{rank}(\begin{bmatrix} B_1 & B_2 \end{bmatrix})$. ▲

Remark 3: Whenever the pair (H, Π) is not observable, items (i)-(iv) of Theorem 1 are not enough to ensure that any solution (K_1, K_2, P) of (11) is the actors/critic triplet. In fact, by inspecting the identity (12), one has that the latter holds for any P with the property that

$$\operatorname{im} \begin{bmatrix} I \\ P \end{bmatrix} \subset \operatorname{im} \begin{bmatrix} I \\ P^\star \end{bmatrix} + \mathscr{I}, \tag{15}$$

where $\mathscr{I} \subset \mathbb{R}^{2n}$ denotes the unobservable subspace of (H, Π). Therefore, to identify a unique triplet, the item

(v′) $0 = C^\top C + A^\top P + P(A + B_1K_1 + B_2K_2)$,

must be added to the requirements of Theorem 1. This ensures that precisely n modes of the Hamiltonian matrix H are excited via the *input matrix* $\begin{bmatrix} I & P \end{bmatrix}^\top$ (see also [9, Thm. 4] for a similar condition in the case of dynamic games) ▲

IV. Policy Algebraic Equation

By further building on the characterization of the actors/critic triplet in terms of the observability matrix associated with the underlying Hamiltonian dynamics, illustrated in Section III, the purpose of this section is to envision *an algebraic equation that involves the actors K_i, $i = 1, 2$ alone*. This reverses all the existing strategies that rely upon (2), which

depends instead on the critic, i.e. P, alone. To provide a concise statement of the following result, let $\mathcal{G} \in \mathbb{R}^{2n^2 \times 2n}$ be defined as $\mathcal{G} := \mathcal{O}(H, \Pi H)$ and partitioned according to $\begin{bmatrix} \mathcal{G}_1 & \mathcal{G}_2 \end{bmatrix} = \mathcal{G}$, with $\mathcal{G}_i \in \mathbb{R}^{2n^2 \times n}$, namely containing the first and last n columns, respectively, of $\mathcal{G}$.

Lemma 2: Fix $\gamma \in \mathbb{R}_{>0}$ and consider the $\mathcal{H}_\infty$ control problem for system (1). Suppose that Assumption 1 holds. Let $K_i^\star$, $i = 1, 2$, denote the actors pair. Then the corresponding critic $P^\star$ is obtained as

$$P^\star = \mathcal{G}_2^\dagger (\mathcal{S}(K_1^\star, K_2^\star) - \mathcal{G}_1) \tag{16}$$

with $\mathcal{G}_2^\dagger = (\mathcal{G}_2^\top \mathcal{G}_2)^{-1} \mathcal{G}_2^\top$ and $\mathcal{S}(\cdot, \cdot)$ defined in (10). ∘

Proof: The claim is obtained immediately by relying on the partition of $\mathcal{G}$ and by observing that (11) may be written as $\mathcal{G}_1 + \mathcal{G}_2 P = \mathcal{S}(K_1, K_2)$. Lemma 1 ensures that $\mathcal{O}(H, \Pi)$ is full column-rank, while an identical property is possessed by $\mathcal{G}$ since $H \in \mathrm{dom}(\mathrm{Ric})$, and hence it is nonsingular. Therefore, the columns of $\mathcal{G}_2$ are linearly independent and (16) is established via standard pseudo-inversion. ■

Remark 4: The statement of Lemma 2 suggests that the critic $P^\star$ may be obtained as a *polynomial* function of the actors $K_i^\star$, $i = 1, 2$. It is worth observing that the standard strategy for computing the cost of a pair of stabilizing control policies consists in solving the Lyapunov equation

$$P A_{\mathrm{cl}} + A_{\mathrm{cl}}^\top P = -C^\top C - (K_2^\star)^\top K_2^\star + \gamma^2 (K_1^\star)^\top K_1^\star \tag{17}$$

with $A_{\mathrm{cl}} := A + B_1 K_1^\star + B_2 K_2^\star$. Note that (17) provides a characterization of $P^\star$ as a matrix containing *rational* functions of the entries of the actors $K_i^\star$, $i = 1, 2$. The key difference is that, while (17) relates *any* feedback to its cost by solving a linear system (*policy evaluation*), (16) relates directly the actor to the structure of the *optimal* critic. ▲

Example 1: To illustrate the intuition behind the comments in Remark 4, consider the scalar differential equation

$$\dot{x}(t) = a x(t) + b_1 w(t) + b_2 u(t), \tag{18}$$

with $a \in \mathbb{R}$, $b_i \in \mathbb{R}$, $i = 1, 2$ and with $c^2 > 0$. Provided γ is such that $\gamma^2 \neq b_1^2 b_2^{-2}$, $\mathcal{G} \in \mathbb{R}^{2 \times 2}$ is defined as

$$\mathcal{G} := \left[\begin{array}{c:c} a & \frac{b_1^2}{\gamma^2} - b_2^2 \\ a^2 - c^2 \left(\frac{b_1^2}{\gamma^2} - b_2^2 \right) & 0 \end{array} \right] \tag{19}$$

and hence $\mathcal{G}_2^\dagger = \begin{bmatrix} \gamma^2 (b_1^2 - \gamma^2 b_2^2)^{-1} & 0 \end{bmatrix}$. Therefore, while the Lyapunov equation (17) yields the rational function

$$p_{(17)} = \frac{-c^2 - k_2^2 + \gamma^2 k_1^2}{2(a + b_1 k_1 + b_2 k_2)}, \tag{20}$$

the formula (16) leads instead to the *linear* function

$$p_{(16)}^\star = \left(\frac{\gamma^2 b_1}{b_1^2 - \gamma^2 b_2^2} \right) k_1^\star + \left(\frac{\gamma^2 b_2}{b_1^2 - \gamma^2 b_2^2} \right) k_2^\star . \tag{21}$$

△

Theorem 2: Fix $\gamma \in \mathbb{R}_{>0}$ and consider the $\mathcal{H}_\infty$ control problem for system (1). Suppose that Assumption 1 holds. Consider the *Policy Algebraic Equation* (PAE)

$$0 = \gamma^2 K_1 - B_1^\top \mathcal{G}_2^\dagger (\mathcal{S}(K_1, K_2) - \mathcal{G}_1), \tag{22a}$$

$$0 = K_2 + B_2^\top \mathcal{G}_2^\dagger (\mathcal{S}(K_1, K_2) - \mathcal{G}_1). \tag{22b}$$

Then the actors $K_1^\star$, $K_2^\star$ solve (22). Moreover, if (22) admits a unique stabilizing solution, then this is the pair of optimal actors $(K_1^\star, K_2^\star)$. ∘

Proof: To begin with, replacing (16) into the definitions of the matrices $K_1^\star$ and $K_2^\star$ provided in (5) and (4), respectively, yields (22). As a consequence, the equations (22) constitute an identity satisfied by the actors $K_i^\star$, $i = 1, 2$. Conversely, provided (22) admit only one solution, denoted $(K_1^{\mathrm{s}}, K_2^{\mathrm{s}})$, that is stabilizing for the closed-loop system, then it follows that $(K_1^{\mathrm{s}}, K_2^{\mathrm{s}}) = (K_1^\star, K_2^\star)$, as the latter is stabilizing and belongs to the set of roots of (22). ■

It may be possible to replace the solution of (17) into (4), (5) to obtain a *rational* system of equations, each of which could be then multiplied by its denominator. Nonetheless, since A_{cl} depends on K_i, $i = 1, 2$, solving (the *vectorized* version of) (17) would require the symbolic inversion of a $n^2 \times n^2$ matrix, which could be a daunting computational obstacle. Computing $P^\star$ as a function of K_i, $i = 1, 2$ as in (16) instead requires only basic symbolic operations, such as sum and product. In fact, the inversion in $\mathcal{G}_2^\dagger$ involves only matrices of (known) numbers obtained from the problem data. Furthermore, (16) reveals an interesting structure, showing how the matrix $P^\star$ depends on the observability matrix of the Hamiltonian dynamics and on powers of the closed-loop plant, which would be destroyed by matrix inversion.

Example 2: To illustrate the results of this section, consider the perturbed linear system

$$\left\{ \begin{array}{rcl} \dot{x}_1 & = & -x_1 + x_2 + w \\ \dot{x}_2 & = & x_1 + x_2 + u \end{array} \right. \tag{23}$$

with $u(t) \in \mathbb{R}$, $w(t) \in \mathbb{R}$ and $y_p = x_1$. Suppose that $\gamma = 1$. By inspecting the equations in (23) it is immediate to observe that the control input and the disturbance are not matched. As a consequence, the coefficient of the quadratic term in (2), namely $\gamma^{-2} B_1 B_1^\top - B_2 B_2^\top$, consists of a sign-indefinite matrix. The Policy Algebraic Equation (22) instead comprises 4 polynomial equations, in the entries of the matrices $K_1 = \begin{bmatrix} k_{1,1} & k_{1,2} \end{bmatrix}$ and $K_2 = \begin{bmatrix} k_{2,1} & k_{2,2} \end{bmatrix}$, without any further structural constraint. More precisely, the closed-loop dynamic matrix is described by

$$A + B_1 K_1 + B_2 K_2 = \begin{bmatrix} k_{1,1} - 1 & k_{1,2} + 1 \\ k_{2,1} + 1 & k_{2,2} + 1 \end{bmatrix}, \tag{24}$$

from which the computation of $\mathcal{S}$ in (10) is straightforward. The corresponding equations (22) then admit only two solutions, which are in fact in one-to-one correspondence with the positive and negative solutions, respectively, of the underlying ARE (2) associated with (23). △

V. The case of the LQR problem

The purpose of this section consists in discussing how the Linear Quadratic Regulator problem can be approached as a specially-structured setting of the above results, hence extending the conclusions of Theorems 1 and 2. To this end, suppose that $w \equiv 0$ in (1) or – to avoid cumbersome notation in this section – consider instead the differential equation

$$\dot{x}(t) = Ax(t) + Bu(t), \qquad x(0) = x_0. \tag{25}$$

Moreover the associated cost functional (7) becomes

$$J_{x_0}(u(\cdot)) = \int_0^\infty (\|x(t)\|_Q^2 + \|u(t)\|_R^2)dt, \tag{26}$$

with $Q := C^\top C$ and $R = R^\top \succ 0$.

Assumption 2: The pairs (A, B) and (A, Q) are reachable and observable, respectively. ∘

As it has been elegantly established (see e.g. [2]), the solution to the optimal control problem described by (25) and the minimization of (26) is obtained by relying on (2) and (3) without any further requirement related to disturbance attenuation guarantees, namely (intuitively) considering the limiting case of γ that tends to infinity. Therefore, consider

$$0 = Q + A^\top P_{\rm o} + P_{\rm o}A - P_{\rm o}BR^{-1}B^\top P_{\rm o} \tag{27}$$

$P_{\rm o} \in \mathbb{R}^{n\times n}$, together with the Hamiltonian matrix

$$H_{\rm o} = \begin{bmatrix} A & -BR^{-1}B^\top \\ -Q & -A^\top \end{bmatrix} \tag{28}$$

respectively. Then the optimal policy is defined in terms of the feedback control law

$$u^\star = -R^{-1}B^\top P_{\rm o}^\star x =: K_{\rm o}^\star x, \tag{29}$$

where $P_{\rm o}^\star = (P_{\rm o}^\star)^\top \succ 0$ denotes the unique positive definite solution of (27). The existence of such special solution is ensured by the structural requirements of Assumption 2. The notation introduced in Sections III and IV may be inherited, within the optimal control framework, by letting

$$\mathcal{Z} = \begin{bmatrix} \mathcal{Z}_1 & \mathcal{Z}_2 \end{bmatrix} := \mathcal{O}(H_{\rm o}, \Pi H_{\rm o}), \tag{30}$$

with $\mathcal{Z}_i \in \mathbb{R}^{2n^2\times n}$, $i = 1, 2$, together with

$$K \mapsto \mathcal{S}_{\rm o}(K) := \begin{bmatrix} A + BK \\ (A + BK)^2 \\ \vdots \\ (A + BK)^{2n} \end{bmatrix}. \tag{31}$$

Note that the counterpart of Theorem 1 in Section III is stated in [7, Thm. 2] (although with slightly different notation). Thus, the following statements specialize the claims of the results in Sections III and IV to the case of the LQR problem.

Lemma 3: Consider the LQR problem described by (25) and (26). Suppose that Assumption 2 holds. Let $K_{\rm o}^\star$ denote the optimal actor. Then the critic $P_{\rm o}^\star$ is obtained as

$$P_{\rm o}^\star = \mathcal{Z}_2^\dagger(\mathcal{S}_{\rm o}(K_{\rm o}^\star) - \mathcal{Z}_1) \tag{32}$$

with $\mathcal{Z}_2^\dagger = (\mathcal{Z}_2^\top \mathcal{Z}_2)^{-1}\mathcal{Z}_2^\top$ and $\mathcal{S}_{\rm o}(\cdot)$ defined in (31). ∘

Proof: By [7, Thm. 2] one has that

$$\mathcal{Z}\begin{bmatrix} I & P_{\rm o} \end{bmatrix}^\top = \mathcal{Z}_1 + \mathcal{Z}_2 P_{\rm o} = \mathcal{S}_{\rm o}(K_{\rm o}^\star) \tag{33}$$

which is the optimal control counterpart of (11), while $\mathcal{Z}_2$ is full column rank by Assumption 2 and [7, Prop. 1]. ■

Theorem 3: Consider the LQR problem described by (25) and (26). Suppose that Assumption 2 holds. Consider the *Policy Algebraic Equation* (PAE)

$$0 = RK + B^\top \mathcal{Z}_2^\dagger(\mathcal{S}_{\rm o}(K) - \mathcal{Z}_1). \tag{34}$$

Then $K_{\rm o}^\star$ solves (34). Moreover, if (34) admits a unique stabilizing solution, then this is the optimal actor $K_{\rm o}^\star$. ∘

Remark 5: A few comments are in order about the *Policy Algebraic Equation* (34) (note that similar discussions may be immediately adapted to the case of (22)). The equation (34) provides a novel characterization of the optimal feedback gain alternative to the standard ARE (27). The former comprises mn variables and an equal number of equations of order less than or equal to $2n$ in the entries of the matrix K. This is different from the ARE (27), which contains $n(n+1)/2$ variables and equations of order 2 in the entries of the critic $P_{\rm o}$. Since both (34) and (27) consist of polynomial equations, one of the most computationally efficient approaches to address them relies upon the computation of *Gröbner bases*. According to [12, Thm. 6.2], whenever the equations admit a finite number of solutions, the computational complexity generically grows *polynomially* in the (maximal) order of the polynomials while *exponentially* with respect to the number of variables. Therefore, although several elegant and efficient techniques have been developed to tackle (27), reducing the number of unknowns, as in (34), may prove to be a desirable feature. ▲

Example 3: Consider a LTI system described by the following *chain of integrators*

$$\begin{cases} \dot{x}_i &= x_{i+1}, \quad i = 1, 2, 3 \\ \dot{x}_4 &= u \end{cases} \tag{35}$$

together with the cost functional (26) with $Q = I$ and $R = 1$. The standard Algebraic Riccati Equation (27) in this case contains 10 quadratic equations in the entries of the matrix $P_{\rm o}$ (10 variables), which must be positive definite. Conversely, the PAE comprises only 4 variables and equations, reported in (36) (overleaf). Note that, despite $n = 4$, the highest order appearing in (36) is 5. The latter system of equations admits 6 distinct solutions, only one of which is stabilizing, namely

$$K_{\rm o}^\star = \begin{bmatrix} -1 & -3.0777 & -4.2361 & -3.0777 \end{bmatrix} \tag{37}$$

which constitutes the optimal policy of the underlying control problem, i.e. the unique stabilizing solution of (27). △

The statement of Theorem 3 implicitly suggests that the PAE (34) is not equivalent to (33), which is instead equivalent to the standard ARE (27). In fact, the set of solutions of (33), as far as the variable K is concerned, is contained in that of (34). The key difference consists in the property that the latter is defined in terms of the feedback gain K alone, while the former jointly relates K and $P_{\rm o}$. Nonetheless, by further manipulating (33) it may be possible to derive an algebraic

$$\begin{bmatrix}0\\0\\0\\0\end{bmatrix}^\top = \begin{bmatrix} k_1^2+k_1k_3^2+3k_1k_3k_4^2-k_1k_3+k_1k_4^4-k_1k_4^2+2k_2k_1k_4+k_1+1 \\ k_2+2k_1k_2-k_1k_4-k_2k_3+k_2k_3^2+k_1k_4^3-k_2k_4^2+2k_2^2k_4+k_2k_4^4+2k_1k_3k_4+3k_2k_3k_4^2 \\ k_3-k_1+2k_1k_3-k_2k_4+k_1k_4^2+k_2k_4^3-k_3k_4^2+k_3k_4^4+k_2^2-k_3^2+k_3^3+3k_3^2k_4^2+4k_2k_3k_4-1 \\ k_4-k_2+2k_1k_4+2k_2k_3-2k_3k_4+3k_2k_4^2+3k_3^2k_4+4k_3k_4^3-k_4^3+k_4^5 \end{bmatrix}^\top \tag{36}$$

equation whose solutions are in *one-to-one* correspondence with those of (33), hence of (27), while still involving only the variable K, although at the price of larger dimension. This is discussed in the following remark.

Remark 6: Let $\mu = 2n^2 - n$ and $\mathcal{Z}_2^\perp \in \mathbb{R}^{\mu\times 2n^2}$ be a full row rank matrix with the property that $\mathcal{Z}_2^\perp \mathcal{Z}_2 = 0$. Then, since the matrix $\begin{bmatrix}(\mathcal{Z}_2^\perp)^\top & (\mathcal{Z}_2^\dagger)^\top\end{bmatrix}^\top$ is nonsingular by construction, one has that the equation

$$\begin{bmatrix}\mathcal{Z}_2^\perp \\ \mathcal{Z}_2^\dagger\end{bmatrix}(\mathcal{Z}_1 + \mathcal{Z}_2 P_{\rm o}) = \begin{bmatrix}\mathcal{Z}_2^\perp \mathcal{Z}_1 \\ \mathcal{Z}_2^\dagger \mathcal{Z}_1 + P_{\rm o}\end{bmatrix} = \begin{bmatrix}\mathcal{Z}_2^\perp \mathcal{S}_{\rm o}(K_{\rm o}) \\ \mathcal{Z}_2^\dagger \mathcal{S}_{\rm o}(K_{\rm o})\end{bmatrix} \tag{38}$$

is equivalent to (33), and hence to (27). Since the lower block is solved by selecting $P_{\rm o}$ as in (32), it follows that

$$0 = \mathcal{Z}_2^\perp(\mathcal{Z}_1 - \mathcal{S}_{\rm o}(K)) \tag{39}$$

constitutes an algebraic equation, in the variable K alone, whose set of solutions is in *one-to-one* correspondence with that of the standard ARE (27). ▲

Remark 7: The algebraic equation (39) may be employed to provide a straightforward *optimality certificate*. Namely, if one is given a candidate stabilizing policy described by $K_{\rm c}$, verifying whether $K_{\rm c}$ is indeed the optimal policy or not tantamounts to simply *plugging* $K_{\rm c}$ into (39) and *checking* that the equality holds. This avoids the need for explicitly computing the cost of $K_{\rm c}$ that should be then, in turn, employed to verify (27). The use of (39) envisioned herein may prove useful, for instance, in iterative methods akin to those currently developed within the RL framework. ▲

Example 4: Consider the LTI system described by

$$\begin{cases} \dot{x}_1 &= -x_2 - x_3 \\ \dot{x}_2 &= x_1 - x_2 \\ \dot{x}_3 &= -x_1 - x_2 + u \end{cases} \tag{40}$$

and the cost functional (26) with $Q = I$ and $R = 1$. The PAE (34) admits isolated (finitely many) solutions, one of which is indeed the optimal feedback gain $K_{\rm o}^\star = \begin{bmatrix}2.3034 & -0.0256 & -2.3679\end{bmatrix}$. Nonetheless, it is interesting to observe that (34) admits 6 distinct solutions, among which one can find a further stabilizing solution in addition to $K_{\rm o}^\star$, namely $\hat{K} = \begin{bmatrix}3.9334 & 0.5522 & -2.3806\end{bmatrix}$. More precisely, while the solution $K_{\rm o}^\star$ of (34) assigns the eigenvalues such that $\sigma(A+BK_{\rm o}^\star) = \{-1.3662, -1.0009 \pm 0.9669j\}$, consistently with the spectrum of the Hamiltonian matrix $H_{\rm o}$, i.e. $\sigma(H_{\rm o}) = \{-1.3662, -1.0009 \pm 0.9669j, 1.3662, 1.0009 \pm 0.9669j\}$, the solution $\hat{K}$ is such that $\sigma(A+B\hat{K}) = \{-1.3662, -1.0072 \pm 1.5961j\}$, so that only one eigenvalue simultaneously belongs to $\sigma(H_{\rm o})$. However, the right-hand side of (32) with $\hat{K}$ in place of $K_{\rm o}^\star$ (which should have been the *critic* associated with $\hat{K}$) is not symmetric, let alone positive semi-definite, hence ruling out $\hat{K}$ as a *candidate* solution. Finally, it is immediate to verify that instead the equation (39) specialized to (40) admits only 4 solutions, similarly to the ARE (27), which include $K_{\rm o}^\star$, whereas $\hat{K}$ is such that $\mathcal{Z}_2^\perp(\mathcal{Z}_1 - \mathcal{S}_{\rm o}(\hat{K})) \neq 0$. △

VI. Conclusions and Further Discussions

The optimal policy and value function of the LQR and $\mathcal{H}_\infty$ control problems have been related via the observability matrix of the Hamiltonian dynamics. The arising algebraic equation, equivalent to the underlying ARE, remains linear in the value function, although polynomial in the control policy. Furthermore, such a condition is instrumental for deriving an algebraic equation which involves the policy *alone*, i.e. deriving a Policy Algebraic Equation. As future work, it may be of interest to further investigate the nature (and the properties) of the set of solutions to (34) in comparison with that of (33). Moreover, envisioning an iterative and, possibly, data-driven strategy to address (22) or (34) may prove relevant in practice. This could be achieved for instance by adapting the *certainty equivalence* arguments of [13] to the PAE. Finally, the polynomial structure of (32) could be further leveraged, apart from its use in the construction of (34), e.g. for direct policy optimization.

References

[1] B. D. Anderson and J. B. Moore, *Optimal control: linear quadratic methods*. Prentice Hall, New Jersey, 1990.

[2] K. Zhou, J. Doyle, and K. Glover, *Robust and Optimal Control*, 1st ed. Prentice Hall, 1996.

[3] D. P. Bertsekas, *Dynamic programming and optimal control*. Athena scientific Belmont, MA, 2005.

[4] R. Bellman, *Dynamic Programming*. Princeton, NJ, USA: Princeton University Press, 1957.

[5] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko, *The Mathematical Theory of Optimal Processes*. John Wiley & sons, New York, 1962.

[6] A. Isidori and A. Astolfi, "Disturbance attenuation and $\mathcal{H}_\infty$-control via measurement feedback in nonlinear systems," *IEEE transactions on automatic control*, vol. 37, no. 9, pp. 1283–1293, 1992.

[7] M. Sassano and A. Astolfi, "Combining Pontryagin's principle and Dynamic Programming for linear and nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 65, no. 12, pp. 5312–5327, 2020.

[8] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An introduction*. MIT press, 2018.

[9] M. Sassano and A. Astolfi, "Constructive design of open-loop Nash equilibrium strategies that admit a feedback synthesis in LQ games," *Automatica*, vol. 133, p. 109840, 2021.

[10] A. J. van der Schaft, *L2-gain and passivity techniques in nonlinear control*. 2nd edition, Springer, 2000.

[11] J. C. Doyle, K. Glover, P. P. Khargonekar, and B. A. Francis, "State-space solutions to standard $\mathcal{H}_2$ and $\mathcal{H}_\infty$ control problems," *IEEE Transactions on Automatic Control*, vol. 34, no. 8, pp. 831–847, 1989.

[12] J.-C. Faugere, P. Gianni, D. Lazard, and T. Mora, "Efficient computation of zero-dimensional gröbner bases by change of ordering," *Journal of Symbolic Computation*, vol. 16, no. 4, pp. 329–344, 1993.

[13] H. Mania, S. Tu, and B. Recht, "Certainty equivalence is efficient for linear quadratic control," *Advances in Neural Information Processing Systems*, vol. 32, 2019.